N° d'ordre NNT : 2025ISAL0105

THÈSE de DOCTORAT DE L'INSA LYON, membre de l'Université de Lyon

École Doctorale N° 341
Évolution, Écosystèmes, Microbiologie, Modélisation
Discipline de doctorat
Biologie Évolutive, Biologie des populations, Écophysiologie

Soutenue publiquement le 28/11/2025, par
Anaïs Larue

The contribution of transposable elements to the phenotypic plasticity of *Drosophila melanogaster*

Devant le jury composé de :

Cameron GHALAMBOR	**Rapporteur**
Professeur, Norwegian University of Science and Technology (NTNU)	
Christoph GRUNAU	**Rapporteur**
Professeur, Université de Perpignan *Via Domitia*	
Pascale LESAGE	**Rapportrice**
Directrice de Recherche, Centre National de la recherche scientifique (CNRS)	
Patricia GIBERT	**Examinatrice**
Directrice de Recherche, Centre National de la recherche scientifique (CNRS)	
Abdelaziz HEDDI	**Examinateur**
Professeur, Institut National des Sciences Appliquées (INSA)	
François SABOT	**Examinateur**
Directeur de Recherche, Institut de Recherche pour le Développement (IRD)	
Rita REBOLLO	**Directrice de thèse**
Chercheuse, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE)	
Cristina VIEIRA	**Co-encadrante de thèse**
Professeure, Université Claude Bernard Lyon 1 (UCBL1)	
Séverine CHAMBEYRON	**Invitée**
Directrice de Recherche, Centre national de la recherche scientifique (CNRS)	

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

Département de la Formation par la Recherche et des Études Doctorales (**FEDORA**)

Bâtiment Charlotte Perriand 3e étage
17, av. des Arts
69621 Villeurbanne Cédex
fedora@insa-lyon.fr

Référence : TH1299_LARUE Anais

L'INSA Lyon a mis en place une procédure de contrôle systématique via un outil de détection de similitudes (logiciel Compilatio). Après le dépôt du manuscrit de thèse, celui-ci est analysé par l'outil. Pour tout taux de similarité supérieur à 10%, le manuscrit est vérifié par l'équipe de FEDORA. Il s'agit notamment d'exclure les auto-citations, à condition qu'elles soient correctement référencées avec citation expresse dans le manuscrit.

Par ce document, il est attesté que ce manuscrit, dans la forme communiquée par la personne doctorante à l'INSA Lyon, satisfait aux exigences de l'Etablissement concernant le taux maximal de similitude admissible.

INSA LYON
Campus LyonTech La Doua
20, avenue Albert Einstein - 69621 Villeurbanne cedex - France
Tél. +33 [0]4 72 43 83 83 - Fax +33 [0]4 72 43 85 00
www.insa-lyon.fr

membre de UNIVERSITÉ DE LYON

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Département FEDORA – INSA Lyon - Ecoles Doctorales

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
ED 206 CHIMIE	**CHIMIE DE LYON** https://www.edchimie-lyon.fr Sec. : Renée EL MELHEM Bâtiment Blaise PASCAL 7, Avenue Jean Capelle, 69621 Villeurbanne CEDEX secretariat@edchimie-lyon.fr	**M. Stéphane DANIELE** C2P2-CPE LYON-UMR 5265 Bâtiment F308, BP 2077 43 Boulevard du 11 novembre 1918, 69616 Villeurbanne CEDEX directeur@edchimie-lyon.fr
ED 341 E2M2	**ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION** http://e2m2.universite-lyon.fr Sec. : Bénédicte LANZA Bâtiment Atrium, 43, Boulevard du 11 Novembre 1918, 69622 Villeurbanne CEDEX Tél : 04.72.44.83.62 secretariat.e2m2@univ-lyon1.fr	**Mme Sandrine CHARLES** Université Claude Bernard Lyon 1 UFR Biosciences, Bâtiment Mendel 43, Boulevard Du 11 Novembre 1918, 69622 Villeurbanne CEDEX e2m2.codir@listes.univ-lyon1
ED 205 EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTÉ** http://ediss.universite-lyon.fr Sec. : Bénédicte LANZA Bâtiment Atrium, Université Claude Bernard Lyon 1 43, Boulevard du 11 Novembre 1918, 69622 Villeurbanne CEDEX Tél : 04.72.44.83.62 secretariat.ediss@univ-lyon1.fr	**M. Jérôme LAMARTINE** Université Claude Bernard Lyon 1 Equipe Intégrité Fonctionnelle du Tissu Cutané LBTI - UMR5305 CNRS IBCP - 7 passage du Vercors, 69367 Lyon CEDEX 07 Tel : 04.72.72.26.66 jerome.lamartine@univ-lyon1.fr
ED 34 EDML	**MATÉRIAUX DE LYON** http://ed34.universite-lyon.fr Sec. : Yann DE ORDENANA Tél : 04.72.18.62.51 yann.de-ordenana@ec-lyon.fr	**M. Stéphane BENAYOUN** Ecole Centrale de Lyon Laboratoire LTDS 36 Avenue Guy de Collongue, 69134 Ecully CEDEX Tél : 04.72.18.64.37 stephane.benayoun@ec-lyon.fr
ED 160 EEA	**ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE** https://edeea.universite-lyon.fr Sec. : Philomène TRECOURT Bâtiment Charlotte PERRIAND, 69621 Villeurbanne CEDEX Tél : 04.72.43.71.70 secretariat.edeea@insa-lyon.fr	**M. Philippe DELACHARTRE** INSA LYON Laboratoire CREATIS, Bâtiment Blaise Pascal, 7 Avenue Jean Capelle, 69621 Villeurbanne CEDEX Tél : 04.72.43.88.63 philippe.delachartre@insa-lyon.fr
ED 512 INFOMATHS	**INFORMATIQUE ET MATHÉMATIQUES** http://edinfomaths.universite-lyon.fr Sec. : Renée EL MELHEM Bâtiment Blaise PASCAL 7, Avenue Jean CAPELLE, 69621 Villeurbanne CEDEX Tél : 04.72.43.80.46 infomaths@univ-lyon1.fr	**M. Dragos IFTIMIE** Université Claude Bernard Lyon 1 Bâtiment Braconnier 21 Avenue Claude BERNARD, 69 622 Villeurbanne CEDEX Tél : 04.72.44.79.58 direction.infomaths@listes.univ-lyon1.fr
ED 162 MEGA	**MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE** http://edmega.universite-lyon.fr Sec. : Philomène TRECOURT Tél : 04.72.43.71.70 Bâtiment Charlotte PERRIAND, 69621 Villeurbanne CEDEX mega@insa-lyon.fr	**M. Etienne PARIZET** INSA Lyon Laboratoire LVA Bâtiment St. Exupéry 25 bis Avenue Jean CAPELLE, 69621 Villeurbanne CEDEX etienne.parizet@insa-lyon.fr
ED 483 ScSo	**ScSo**[1] https://edsciencessociales.universite-lyon.fr Sec. : Mélina FAVETON Tél : 04.78.69.72.76 Université Lyon 2 Campus Berges du Rhône, Bureau BEL 309 18, Quai Claude BERNARD, 69365 LYON CEDEX 07 melina.faveton@univ-lyon2.fr	**M. Bruno MILLY** Université Lyon 2 Campus Berges du Rhône Bureau BEL 204 18, Quai Claude BERNARD 69365 LYON CEDEX 07 bruno.milly@univ-lyon2.fr

[1] ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Contribution of Transposable Elements to the Phenotypic Plasticity of *Drosophila melanogaster*

Anaïs Larue

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Acknowledgements

Some may easily write this section, but for me, it almost feels like the hardest task. Where to even start? How not to forget to mention dear friends and colleagues? I guess I'll begin by expressing how immensely grateful I am to have been given the opportunity to carry out such a challenging yet gratifying experience as completing a PhD. For that, I want to deeply thank both Aziz and Cristina, without whom I wouldn't be here writing these words at all. Cristina, if it weren't for you recruiting me as an assistant engineer, I would never have had the chance to second-guess my decision and embark on a PhD. So, thank you, I am forever grateful! And Aziz, if it weren't for your support and dedication, this whole journey would have remained but a distant dream. From the bottom of my heart, thank you!

Of course, some had to initiate this whole crazy project, and someone equally crazy had to make it happen on a daily basis while coping with my deranged brain! And that would be you, Rita! Rita, there are no words in this language, and perhaps in any other, that can truly convey how grateful I am to you. You have been a delight to work with since day one. You are not only an incredibly talented researcher but also a beautiful human being. Working alongside you, I truly got the best of both worlds. So much so that it was sometimes hard for me not to intertwine work and friendship… "Rich people's problems," as you would say, am I right? If I had to do it all over again, I wouldn't hesitate for a second, simply because you made this journey an enjoyable ride from start to finish. Muito obrigado!

There are many more people related to this project who deserve acknowledgements. I especially want to thank Séverine Chambeyron, for she has been of invaluable scientific input, having both helped come up with the project in collaboration with Cristina and provided the flies without which none of this would have been possible. I also want to thank her on a more personal level. Thank you, Séverine, for being there on that dreadful day in Paris, and thank you for believing in this project when I was unsteady! I would also like to thank the members of my thesis committee, of which Séverine was part, who have all demonstrated great kindness: Patricia Gibert, Francesca Merlin, and Emilien Luquet.

The last people I would like to acknowledge in relation to this project are the members of my thesis jury. I sincerely thank the "rapporteur.trice.s" for taking the time to read this work thoroughly and for contributing with their own expertise to the discussion: Cameron Ghalambor, Christoph Grunau and Pascale Lesage. I wish also to express my gratitude to Patricia Gibert and François Sabot for kindly accepting to serve as "examinateur.trice.s" for my PhD defence.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

These three years spent mostly at work also gave me the chance to meet wonderful people, some of whom I am proud to call friends today. I will start by expressing my immense gratitude to the Pôle Biotechnologie and the Pôle Administration for their warm welcome when I first joined the LBBE lab as an assistant engineer, and for their unfailing help ever since. There are so many people in these two poles, and I truly want to thank each and every one of them! A special mention goes to Sonia Janillon who greatly contributed to collecting data for my PhD project, and who has always been so willingly helpful to me!

I also want to warmly acknowledge the incredible bond I have built with my former office mate, Zaïnab Belgaïdi. Zaïnab, I have cherished every moment with you! Your brightness and joy are contagious, and working alongside you has always been both deeply enjoyable and enlightening. Thank you, my dear friend, I wish you all the very best!

Now for my other laboratory, because yes, I was also lucky to be part of two labs, within both of which many great people deserve a demonstration of my respect and acknowledgment. Hence, I want to thank Agnès Vallier for all the technical help and the softness she has always shown me. It is so soothing to work with you, thanks both to your immense experience – experiments hardly ever fail with you – and also to your kindness. I would also like to thank Auréline Lalouette for her dynamism and for showing me that it's more than okay to be unapologetically oneself! You are a great friend! I wish to gratify Sylvain Benhamou, for you are not only a bright colleague who has helped me a ton with statistical issues, but also because, in Lyon, you are one of my oldest friends! Having experienced the thesis before me, you were always an incredible source of support and advice. Thank you so very much!

And I could not forget the Norwegian fellows I met during my many visits to Norway! Thanks to all of them, and particularly to Alexander Mauro for generously sharing his precious professional and personal time with me. I truly hope to see you grow into an established researcher, and hopefully, one day we may even collaborate! I also wish to express my most sincere thanks to Cameron Ghalambor for hosting me in his laboratory and for the many incredibly insightful conversations.

Now of course I could finish with none other than the TE team for the acknowledgments of the work-related people, although you all know you are so much more than that to me! You are like a science family. You have always been so caring, and this has made work so much more enjoyable, being so wonderfully surrounded. I want to thank the former members of the team who knew me back when I was still part of the technical staff: Miriam Meranciano and Daniel Siqueira de Oliveira. Shout out to William Vilas Boas Nunes, Chloé Garambois, and Julie Delescluse, you are incredible people and I will miss you dearly. Sharing daily life with you has been truly treasurable! I am also deeply grateful to have shared an office with you, Tomás Carrasco-Valenzuela. You have all been so generous in helping me with

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

experimental and bioinformatic questions. I truly had amazing moments with you both in and out of the lab! I hope our friendship will defy time and distance!

I wish to finish by thanking the people who have been with me all along and supported me through this adventure, even without fully understanding what the fuss was all about! I want to thank my family for always believing in me and giving me the best chances in life. My parents have been a constant source of support and have never pushed me into anything I didn't want to do. I have always made unconventional decisions, and my career path has never been linear, yet they have always trusted that I was doing what suited me at the time. It took me four years after my master's degree to start a PhD, and they were always so understanding of my choices. I am truly fortunate!

Of course, I want to thank my brother, Rémy, and my sister, Cloé, because one could hardly dream of a better sibling bond than the one we share! I am especially grateful to my sister, who has been my roommate in Lyon for the past two years. Living with you has made my PhD journey far more manageable and has been an immense source of emotional support. I cannot thank you enough for that! Sharing this time with you has been, and will remain, one of the most treasured parts of this period in my life. And because they are also my family, and as an unabashed cat lady, I would also like to mention my two little balls of fluff, who have been a constant source of emotional support throughout this journey and whom I deeply cherish.

Finally, I want to thank my partner, whom I wouldn't even have had the chance to meet if it weren't for this PhD. Dag, you have been an incredible presence during this time. My storm-like mind needed your peaceful and steady one. I will keep it contained here, but know just how much you mean to me and how deeply grateful I am for the emotional support you have given me. I could always count on you to be the best listener when I was ranting, and to be present in exactly the way I needed when I didn't even know what that was. Every cell in my being wants to thank you for who you are and for all that you are to me!

If you feel your name should be here and I forgot to mention you, I might have simply forgot, and if so, I am truly sorry.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Abstract

Phenotypic plasticity refers to the ability of an organism with a given genotype to produce a range of phenotypes in response to environmental changes. In an adaptive context, this plasticity can enhance an organism's ability to face environmental fluctuations. Transposable elements (TEs) are ubiquitous repetitive DNA sequences that are capable of moving within genomes. Although often overlooked in the context of phenotypic plasticity, TEs may play a significant role through their ability to modulate gene expression in an environmentally-dependent manner. The epigenetic control machinery that regulates TEs, as well as their intrinsic regulatory components, can be tuned to the environmental fluctuations. These fluctuations may lead to differential TE expression or mobilization, potentially influencing gene expression and inducing epigenomic, genomic, and transcriptomic plasticity. Such plasticity at the molecular level could ultimately lead to inter-individual diversity and phenotypic plasticity. In order to investigate this hypothesis in controlled conditions, we used genetically engineered *Drosophila melanogaster* flies that have different TE content but share a common genetic background. Phenotypic screening revealed significant differences between populations. Further investigation of the plastic response under different environment conditions revealed variations in plastic responses and an increase in trait variance for the TE-burdened populations. These findings suggest that TEs contribute to the heterogeneity in environmental response and therefore can be an additional underlying mechanism to phenotypic plasticity.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Table of contents

Acknowledgements 6
Abstract 10
Table of contents 12
Abbreviations 16
Table of figures 18
Table of tables 20
Foreword 22
Introduction 24
I. A brief conceptual context: the biggest picture 24
II. Plasticity, or genomes' responsiveness 26
1. A brief historical overview 26
2. An analytical framework 27
3. Evolutionary relevance: phenotypic plasticity as a source of variation 29
4. Mechanisms underlying plasticity 31
a. Hormones 32
b. Transcription factors 33
c. Heatshock proteins 34
d. Epigenetic mechanisms 35
1. DNA methylation 36
2. Histone post-translational modifications 37
3. microRNA 40
e. Cryptic genetic variants 41
III. Transposable elements, between disruption and innovation 45
1. An introduction to transposable elements 45
a. Classification 46
1. Class I retrotransposons 47
2. Class II DNA transposons 49
b. The regulatory mechanisms 51
1. siRNA pathway: somatic regulation of TEs 51
2. piRNA pathway: a predominantly gonadal regulation of TEs 52
3. Epigenetic regulation 54

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

2. Transposable element, an evolutionary perspective 55
a. TE contribution to adaptative evolution 55
b. TEs as molecular drivers of variation 57
c. Evolutionary/phenotypic capacitance from TEs 59
d. Evidence of TE-driven phenotypic variability 60
1. TE-driven phenotypic variation 61
2. TEs are environment-responsive 63
3. The ultimate example: TE-driven plasticity-led evolution 65
IV. Plastic flies with variable TE content: a mean to an end 69
Chapter 1. On the trail of Transposable Elements: exploring their impact on phenotypes, phenotypic variability, and plasticity in Drosophila melanogaster 72
Foreword 72
I. Introduction 72
II. Material and Methods 74
1. Fly husbandry 74
2. DNA extraction and sequencing 75
3. Genome assembly 75
4. Transposable element analysis 76
5. Viability and development time 77
6. Embryonic hatchability 77
7. Dry weight 78
8. Negative geotaxis assays 78
9. Thermal tolerance 79
10. Longevity 79
11. Dietary restriction 80
12. Oxidative stress 80
13. Metabolic rate 80
14. Evaluation of the phenotypic variation 81
15. Linking TE insertions and phenotypes 81
III. Results and Discussion 82
1. TE content increase with generation 82
2. TE composition varies in frequencies and position between the populations 85
3. Variable TE contents influence the phenotypic profiles across the populations 88

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

4. High TE content increases the observed phenotypic variability in females 95
5. Variation in TE content impacts the responses to environmental changes 97
6. Investigation into TEs reveals candidate insertions 103
IV. Conclusions 104
Chapter 2. Caught between TEs and Wolbachia, a facultative endosymbiont: investigating the phenotypic effects in Drosophila melanogaster 108
Foreword 108
I. A brief introduction to symbiosis 108
1. Facultative endosymbiosis and contribution to the host's phenotype 109
2. Wolbachia in D. melanogaster 110
3. Wolbachia's interaction with TEs 111
II. Results and discussion 113
1. Wolbachia density does not differ between populations 113
2. TE content changes following Wolbachia removal 114
3. The populations devoid of Wolbachia display distinct phenotypic profiles 118
4. The impact of Wolbachia on the phenotypes of populations with variable TE content 122
5. Sexual dimorphism in longevity and dry weight is amplified by Wolbachia infection status 125
6. Environment-independent effects of Wolbachia on metabolic rate in populations with varying TE content 129
III. Conclusion 132
IV. Material and Methods 133
1. Fly husbandry 133
2. Wolbachia treatment and fly maintenance 134
3. Wolbachia relative density 134
4. DNA extraction and sequencing 135
5. Transposable element analysis 135
6. Viability and development time 136
7. Embryonic hatchability 136
8. Dry weight 137
9. Negative geotaxis essays 137
10. Longevity 137
11. Metabolic rate 138
General discussion 140
Conclusion 152

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

A personal take-away *152*
Women in Science: when McClintock meets West-Eberhard and Margulis *154*
Other publications *156*
Résumé *162*
I. Introduction *162*
II. Matériels et méthodes *166*
III. Résultats et discussion *166*
IV. Conclusion *167*
S2. Supplementary material for Chapter 1 ..... *168*
S3. Supplementary material for Chapter 2 ..... *192*
Annexe 1 ..... *202*
Annexe 2 ..... *236*
Annexe 3 ..... *258*
Annexe 4 ..... *276*
Annexe 5 ..... *298*
References *348*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Abbreviations

Ac activator element
ADH alcohol dehydrogenase
Ago argonaute protein
Aub aubergine protein
bp base pairs
CAGE cap analysis of gene expression
CGV cryptic genetic variation/variant
CI cytoplasmic incompatibility
DCV *Drosophila* C virus
DIRS dictyostelium intermediate repeat sequence
DNA deoxyribonucleic acid
DNMT DNA methyltransferase
Ds dissociation element
DSB double stranded DNA break
dsDNA double stranded DNA
eccDNA extrachromosomal circular DNA
EN endonuclease
ERV endogenous retrovirus
EV ectopic vein
FHV Flock House virus
FLC flowering locus C
GWAS genome-wide association study
HP1 heterochromatin protein 1
HS heatshock
HSF heatshock factor
HSP heatshock protein
IFN interferon response
IIS insulin/insulin-like signalling pathway
JH juvenile hormone
JHEH juvenile hormone epoxy hydrolase
kb kilo base pairs
KD knockdown
LINE1 long interspersed nuclear element-1
LTR long terminal repeat
m6A N6-methyladenosine modifications
miRNA micro RNA
MITE miniature inverted repeat transposable element
mRNA messenger RNA
ncRNA non-coding RNA
ORF open reading frame
PcG Polycomb group
PCR polymerase chain reaction
PEV position effect variegation
piRNA piwi-interacting RNA
RIP repeat-induced point mutation
RISC RNA-induced silencing complex
RNA ribonucleic acid
RNAi RNA interference
RNAse ribonuclease
ROS reactive oxygen species
RT reverse transcriptase
sh short hairpin
siRNA small interfering RNA
sRNA small RNA
TE transposable element
TF transcription factor
TFBS transcription factor binding site
TIRs terminal inverted repeats
tj traffic jam
TPRT target primed reverse transcriptase
TrxG Trithorax group
TSD target site duplication
TSS transcriptional start site
VE environment induced variation
Vε variance due to stochasticity
VG genetic variation
VGE the variation induced by genome-by-environment interaction
VLP virus-like particle
VP phenotypic variation
VTE transposable element induced variation
YR tyrosine recombinase

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Table of figures

Introduction

Figure 1. 1: Schematic representation of the three reaction norm scenarios (adapted from Kelly et al. 2012). 28
Figure 1. 2: Illustration of the ways phenotypic plasticity may contribute to evolution. 30
Figure 1. 3: Schematic representation of all the histone tail modifications that have been reported in the literature (Liu et al. 2023). 38
Figure 1. 4: Representation of contribution of cryptic genetic variants (CGVs) to genotype-by-genotype and genotype-by-environment interactions (adapted from Pfennig 2021c). 42
Figure 1. 5: Schematic representation of the siRNA pathway (adapted from Merenciano et al. 2025). 52
Figure 1. 6: Schematic representation of the somatic and germinal piRNA pathways (adapted from Merenciano et al. 2025). 54
Figure 1. 7: The impact of TEs on gene expression (adapted from Nishihara 2019). 58
Figure 1. 8: Representation of the position effect variegation (PEV) phenotype in *D. melanogaster* (from Slotkin and Martiessen 2007). 61
Figure 1. 9: Graphical summary of the effect of the roo solo-LTR in *D. melanogaster* upon infection with *Pseudomonas entomophila* (adapted from Merenciano and Gonzáles 2023). 63
Figure 1. 10: Graphical summary of the histone post-translational modifications resulting from the presence of *Bari-Jheh* (adapted from Guio et al 2018). 65
Figure 1. 11: Experimental design following Waddington's genetic assimilation evolutionary experiment protocol for the ectopic vein phenotype (adapted from Sabarís et al 2025). 67
Figure 1. 12: Schematic representation of the molecular mechanisms involved in promoting transcriptional plasticity thereby underlying phenotypic plasticity. 68
Figure 1. 13: **Schematic representation of the genetic modification allowing the accumulation of TEs across generations in *D. melanogaster* (from Barckmann et al. 2018).** 70
Figure 1. 14: Schematic of the experimental setup to establish the fly populations with increased TE content (adapted from Varoqui et al. 2025). 70

Results

Chapter 1

Figure 2. 1: TE content increases in the populations. 85
Figure 2. 2: TE composition varies between the populations. 87
Figure 2. 3: Phenotypic characterization of the populations with increasing TE content. 92
Figure 2. 4: Dimensional and correlative data analysis of the phenotypes of the population with increasing TE content. 94
Figure 2. 5: Summary of the effect sizes of the contribution of the "population" variable to the data. 95
Figure 2. 6: High TE content and insertional polymorphism increases phenotypic variation in females. 96
Figure 2. 7: Longevity varies under variable environment demonstrating a plastic response. 99
Figure 2. 8: Metabolic rate and resistance to oxidative stress under variable conditions. 101
Figure 2. 9: TE content contributes to differences in plasticity responses of various traits in a range of environments. 103

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Chapter 2

Figure 3. 1: Relative *Wolbachia* density in the Wb+ population. 114
Figure 3. 2: TE composition of the new insertions compared to G0-*Wb+* in the populations studied after tetracycline treatment against the endosymbiont *Wolbachia*. 117
Figure 3. 3: Phenotypes of the populations devoid of *Wolbachia* with varying TE composition 120
Figure 3. 4: Phenotypic profiles of the populations devoid of *Wolbachia*. 122
Figure 3. 5: Reaction norms of the phenotypic measurement under variable *Wolbachia* infection status. 125
Figure 3. 6: Reaction norm of the dry weight and LT80 according to the *Wolbachia* infection status. 126
Figure 3. 7: Reaction norm of the metabolic rate according to *Wolbachia* infection status. 130

General discussion

Figure 4. 1: Synteny analysis across populations and assemblies for the *Wolbachia*-infected populations. 145
Figure 4. 2: Thermal tolerance and oxidative stress resistance across populations. 146
Figure 4. 3: Structure of the Myosin 81F gene in S31-*Wb+*. 150

S2. Supplementary Chapter 1

Figure S2. 1: Rarefaction curves of new TE insertions detected by TrEMOLO across populations. 170
Figure S2. 2: Experimental design for the viability and hatchability assays. 171
Figure S2. 3: Newly detected LTR retrotransposon insertions relative to G0-*Wb+* across consensus library families. 174
Figure S2. 4: Newly detected LINE transposable element insertions relative to G0-*Wb+* across consensus library families. 175
Figure S2. 5: Newly detected DNA transposon insertions relative to G0-*Wb+* across consensus library families. 176
Figure S2. 6: Genomic distribution of transposable element insertions detected relative to the G0-*Wb+* reference. 177
Figure S2. 7: Transposable elements (TE) abundance as estimated by DeviaTE in the *Wolbachia*-infected populations. 182

S3. Supplementary Chapter 2

Figure S3. 1: Newly detected LTR retrotransposon insertions relative to G0-*Wb+* across consensus library families for the *Wolbachia*-treated populations. 192
Figure S3. 2: Newly detected LINE transposable element insertions relative to G0-*Wb+* across consensus library families for the *Wolbachia*-treated populations. 193
Figure S3. 3: Newly detected DNA transposon insertions relative to G0-*Wb+* across consensus library families for the *Wolbachia*-treated populations. 194
Figure S3. 4: Shared transposable element (TE) insertions between *Wolbachia* infection status populations. 195

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Table of tables

Chapter 1

Table 2. 1: Summary statistics of the insertion frequencies at the population scale. 83
Table 2. 2: Summary of phenotypic measurements across populations, sexes, and environmental conditions. 89
Table 2. 3: Summary table of statistical analysis and results of the phenotypic measurement under variable environment. 102
Table 2. 4: Summary table of candidate insertions detected through both manual and genome-wide scan methods and their potential phenotypic effects. 107

Chapter 2

Table 3. 1: Summary table of the effects of population, environment, and their interaction on the measured phenotypes. 121
Table 3. 2: Summary table of the effects of population (tested separately for each *Wolbachia* infection status) and *Wolbachia* infection on the measured phenotypes. 124
Table 3. 3: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, and *Wolbachia* infection, as well as their interactions, on females' longevity. 127
Table 3. 4: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, and *Wolbachia* infection, as well as their interactions, on males' longevity. 128
Table 3. 5: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, and *Wolbachia* infection, as well as their interactions, on females' metabolic rate. 131
Table 3. 6: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, and *Wolbachia* infection, as well as their interactions, on males' metabolic rate. 132

S2. Supplementary Chapter 1

Table S2. 1: Assembly completeness and quality metrics for G0-*Wb+* reference genomes. 169
Table S2. 2: Number of flies assayed for longevity in the Wolbachia-infected populations under different temperature. 171
Table S2. 3: Number of flies assayed for longevity in the *Wolbachia*-infected populations on different rearing media. 172
Table S2. 4: Number of flies assayed for longevity in the *Wolbachia*-infected populations exposed to variable concentration of paraquat. 173
Table S2. 5: Distribution of the insertion across genomic regions. 181
Table S2. 6: Ranking of the standard deviation values (SD) of female traits across populations. 181
Table S2. 7: Slope of the reaction norm of LT80 under temperature fluctuation. 182
Table S2. 8: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, as well as their interactions, on females' metabolic rate. 183
Table S2. 9: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, as well as their interactions, on males' metabolic rate. 183
Table S2. 10: Results of a three-way analysis of variance (ANOVA) from the LMER model of the effect of population, paraquat concentration, and their interaction on females' longevity. 184

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Table S2. 11: Results of a three-way analysis of variance (ANOVA) from the LMER model of the effect of population, paraquat concentration, and their interaction on males' longevity. 184
Table S2. 12: Slope of the reaction norm of LT80 in response to changes in paraquat concentration. 185
Table S2. 13: Estimates of LT80 under variable paraquat concentrations in *Wolbachia*-infected populations. 185
Table S2. 14: Manually curated candidate transposable element (TE) insertions. 186
Table S2. 15: Genome-scan candidate transposable element (TE) insertions with manual curation and integration of female phenotypic data. 188
Table S2. 16: Genome-scan candidate transposable element (TE) insertions with manual curation and integration of male phenotypic data. 190
Table S2. 17 199

S3. Supplementary Chapter 2

Table S3. 1: Total transposable element (TE) abundance across five populations under different Wolbachia infection status. 195
Table S3. 2: Estimated marginal means (Emmeans) contrasts from the LMER model for males' longevity. 196
Table S3. 3: Estimated marginal means (Emmeans) contrasts from the LMER model for females' longevity. 197
Table S3. 4: Estimated marginal means (Emmeans) contrasts from the LMER model for females' metabolic rate. 198
Table S3. 5: Estimated marginal means (Emmeans) contrasts from the LMER model for females' metabolic rate. 199
Table S3. 6: Primer sequences used for the qPCR amplification. 200
Table S3. 7: Number of flies assayed for longevity in the *Wolbachia*-treated populations under different temperature. 201

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Foreword

Some of you readers may not be totally aware of the novelty that this project represents, so I would like with this foreword to explain a little bit about the structure of the following manuscript. This project consists in bridging two never-before-connected topics, offering empirical support for what could become a new field of study. While that may sound overly ambitious, that is by no means the intention of these few sentences. What I mean to convey is simply that very little existing literature addresses this question, and as such, I felt it necessary to build a strong argument, perhaps, for some, through a rather extensive introduction. My view is that, to lay the foundation for a new research direction, one must clearly understand the origin of the reasoning. I sincerely hope that, although long, you will enjoy reading this introduction as much as I enjoyed writing it, and I very much look forward to hearing your thoughts and perspectives.

Following the introduction, the first substantial chapter compiles nearly the entirety of my thesis work into a complete research article. In contrast, the second chapter serves more as an opening, presenting work that is still in progress. I recognize that this may not be a conventional thesis format, but then again, I've never been one to follow the ordinary. I hope this work lays the foundation of a new interface between ecological and genomic research, and that it inspires many exciting future studies.

Through this piece of work, I also hope to convey how much this project has meant to me, and how it has profoundly changed me for the better. If you are able to grasp even a glimpse of what this has represented to me, then, my mission is accomplished! There were moments of joy and of doubt, but it is precisely the combination of both that gave birth to this manuscript. Hopefully, this marks only the beginning of a life of research, driven by an unsatiating thirst for knowledge in Evolution.

From the broad to the intricate, you are now invited to take a journey through my thoughts!

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Introduction

I. A brief conceptual context: the biggest picture

Evolution explains biodiversity on Earth, potentially tracing back to the initial self-replicating molecules (Cech 2012; Pressman et al. 2015). Formalization of evolution, however, is relatively recent in human history. Darwin's book "On the Origin of Species" released in 1859, was highly controversial in its days but came to transform the field of biology (Darwin 1859). Darwin implied that the environment favours the individuals better at reproducing (*fittest*), while the maladapted ones become sparse. Ignoring all about genetics or mutations, his theory was the foundation for an entirely new research paradigm.

Over time, numerous evolutionary biologists extended his ideas. With new developments in genetics – such as the discovery of mutations, genetic drift, and population genetics – Darwin's original theory had to be updated, evolving into the modern synthesis of evolutionary biology (Julian Huxley 1942). Now, the then-popular statement of Theodosius Dobzhansky, "Nothing in biology makes sense except in the light of evolution," is accepted as common sense, since it highlights the central role of evolution in understanding the *bios* (Dobzhansky 1973). Following the discovery of genetics and heredity, the focus in evolutionary theory later shifted towards the gene-centred view with genes being the unit of selection rather than the phenotypes (Roe and Simpson 1958; Hamilton 1964; Williams 1966; Dawkins 1976). In this context, organisms are, to some scientists, the "vehicles" for their genomes, or as Dawkins puts it, the "survival machines" of genes (Roe and Simpson 1958; Hamilton 1964; Williams 1966; Dawkins 1976). Following the publication of "The Selfish Gene", Dawkins published "The Extended Phenotype" in 1982 (Dawkins 1982). In this later book, he proposed that while selection acts on phenotypes, it ultimately acts through genes, wherein, genes are the true units of selection, but their effects are expressed into phenotypes (Dawkins 1982). This perspective reinforces the idea that genetic variation underlies phenotypic variation. The necessity of phenotypic variation lies at the heart of many proponents of the phenotype-first view of evolution (West-Eberhard 2003; Rust 2022). Many have since argued that phenotypic variation does not arise exclusively from the genome, but that the environment can also play a crucial role in promoting variation (Pigliucci 2001; Schlichting and Smith 2002; West-Eberhard 2003; Ghalambor et al. 2007; Pfennig 2021b). This supports the view that variation originates from two primary sources: genetically induced and environmentally induced variation (Falconer 1996; W. G. Hill and Mackay 2004).

Indeed, in her landmark book *Developmental Plasticity and Evolution* (2003), Mary Jane West-Eberhard expands on this idea through the *plasticity-first hypothesis* (West-Eberhard 2003). This theory supports the notion that environmentally induced variation (*plasticity*) can precede and facilitate evolutionary change. She thus

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

establishes a new theoretical framework for understanding evolution through the lens of plasticity. She specifically argues against the idea that genes have intrinsic fitness, contending that selection acts upon genes through their expression into phenotypes, making phenotypes the true units of selection and thus preceding adaptation (West-Eberhard 2003). Unlike Dawkins, who places genes at the origin of evolutionary processes, West-Eberhard positions genes as followers, responding to, rather than initiating adaptive change (West-Eberhard 2003; Ghalambor et al. 2007). Notably, West-Eberhard's *plasticity-first theory* echoes Darwin's original ideas, wherein, lacking knowledge of genetics, he proposed that organisms were selected based on their phenotypic fitness.

Nowadays, the biological and molecular mechanisms that underlie environmentally-induced variability, *i.e.* plasticity, remain compelling questions that biologists continue to investigate. It is within this scope that this project is situated. Here, we seek to investigate a putative mechanism that may contribute to environmentally driven changes in phenotypes.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

II. Plasticity, or genomes' responsiveness

1. A brief historical overview

Among the foundational principles of evolutionary biology and ecology is the concept of phenotypic plasticity. Perhaps before introducing the concept, the term phenotype needs to be defined. Phenotypes are the observable or measurable properties and characteristics of an organism. An early study, now assimilated to the concept of phenotypic plasticity – though the term plasticity had not yet been coined – was conducted by Richard Woltereck in 1909 (Woltereck 1909). Investigating phenotypic expression in *Daphnia* water flea across a range of environmental conditions (a model organism for plasticity studies to this day), Woltereck introduced the concept of reaction norms (Woltereck 1909). Today, reaction norms remain a key approach for examining plasticity, a statistical tool that will be discussed further. Although earlier studies, such as Woltereck's, hinted at the idea of phenotypic plasticity, it was Anthony David Bradshaw (1965) who formally introduced the concept (Bradshaw 1965). While growing plants in extreme environmental conditions, he observed that genetically similar seedlings exhibited strikingly different phenotypes, providing clear evidence of developmental plasticity in action (Bradshaw 1965). Bradshaw's study introduced the term *phenotypic plasticity*, marking a significant milestone in the evolution of this concept (Bradshaw 1965).

While the discovery and conceptual establishment of plasticity have historically emerged from the study of developmental stages (developmental plasticity), the present study specifically investigates phenotypic plasticity, thereby addressing environmentally induced phenotypic changes experienced within an organism's lifetime (although, for the same reason, some mention of developmental plasticity examples will be made). However, both share a common definition. The definition of phenotypic plasticity can vary slightly, but the commonly accepted one is that a single genotype can yield multiple phenotypes under environmental variation (Bradshaw 1965; Stearns 1989; Scheiner et al. 1989; West-Eberhard 2003; Pigliucci and Hayden 2001; Pfennig 2021b). Perhaps the most significant nuance in this definition arises from opposing perspectives: for some, the phenotype is an environmental response, while to others, it is environmentally induced (Stearns 1989). This represents a debated dichotomy in the scientific community between plasticity being active *vs.* passive (Pfennig 2021b). This is also reflected in the evolutionary significance of plasticity, whether adaptive, non-adaptive, or even maladaptive. Plasticity may partially be a passive phenomenon, as phenotypic changes may arise from fundamental biochemical and physical laws, where factors such as temperature and nutrient availability inevitably influence an organism's

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

physiology (Pfennig 2021a). In this sense, plasticity is an inherent property of all living organisms, making phenotypic plasticity an ubiquitous phenomenon and thus particularly relevant to study (Nijhout 2003; Palacio-López et al. 2015).

2. An analytical framework

To discuss phenotypic plasticity, one first needs to measure it. As a reminder, here we focused on plasticity during an organism adulthood. Since phenotypic plasticity was generally accepted to be induced by a combination of multiple genes (pleiotropy), an approach to assessing it emerged within the field of quantitative genetics. In 1960, Falconer published a book on quantitative genetics with Chapter 8 specifically addressing phenotypic variance (Falconer 1960; 1996; W. G. Hill and Mackay 2004). In describing the factors contributing to the global phenotypic variance, Falconer identifies two key components that reflect plasticity: one representing the portion of variance explained by the environment's influence, and the other accounting for the genetic variation for plasticity also referred to as the interaction between the genome and the environment.

$$VP = VG + VE + VG \times E + V\varepsilon$$

Where VP represents the total phenotypic variance, VG the total genotypic-induced variance, VE the variance explained by the environment, VGxE the variance explained by the interaction between the genotype and the environment, and Vε represents the variance due to stochasticity.

In this equation, both VE (environmental variance) and VGxE (the variance attributable to the interaction between the genotype and the environment) reflect the plasticity of the phenotypic variability. A special mention should be made here: GxE is the condition under which plasticity can be heritable and therefore under selection (Bowman 1972; Via and Lande 1985; Scheiner 1993; Baythavong 2011; Goldstein and Ehrenreich 2021). The study of variance and its partitioning can be conducted under laboratory conditions by controlling one component and estimating the other. There are thus two types of source of phenotypic variation, the one ensuing from genetic diversity (VG) and one ensuing from environmental effects (VE). To address one, the other must be removed or reduced to the minimum. To address the effect of VE, one must compare identical genotypes. Typically, the genetic component is fixed by comparing highly genetically similar individuals, such as those from isogenic, full-sibling lines or even clones (Falconer 1960; 1996; W. G. Hill and Mackay 2004).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Falconer's approach, combined with the reaction norm framework developed by Woltereck provides a method to quantify plasticity (Woltereck 1909; Falconer 1960). Continuous traits, as opposed to discrete traits, which represent a special case of phenotypic plasticity known as polyphenism, can follow three possible reaction norm scenarios when measured across different environments. The first scenario occurs when the measured trait remains constant across different environments, indicating that the trait is non-plastic under these conditions (as observed for genotype 1, 2 and 3 in Figure 1.1A). In the second scenario, the phenotype of each of the three genotypes changes similarly with the environment, *i.e.* plasticity (Figure 1.1B). The third scenario occurs if as in the figure below, in a range of environments, the responses of different genotypes vary (Figure 1.1C). This represents the genotype-by-environment interaction (G×E) where the genome influences how an organism respond to its environment (Figure 1.1C).

An important clarification is required to distinguish between the terms: variation, variance, and variability. Variation is the existence of differences between individuals of a population or species (Hallgrímsson and Hall 2005). Variance is a statistical measure of variation, expressing the departure of the data from the mean (Hallgrímsson and Hall 2005). As for variability, it describes the propensity/capacity for variation (Hallgrímsson and Hall 2005).

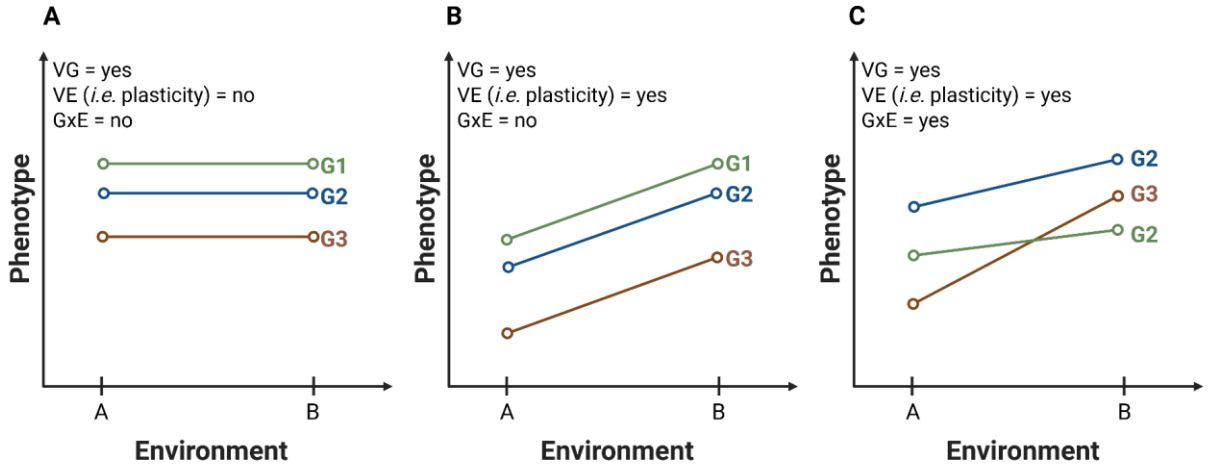


Figure 1. 1: Schematic representation of the three reaction norm scenarios (adapted from Kelly et al. 2012).

G1, G2 and G3 represent three different genotypes. A) This panel illustrates the case where the phenotype does not display plasticity in the tested environments, but there is an effect of the genotype as each genotype is different. B) This panel represents the scenario where the phenotype is plastic only in an environmental-dependent manner resulting in parallel plastic responses. In this scenario an effect of the genotype is also observable as each genotype is different, however their responses to environmental change are identical, resulting in identical reaction norms' slopes. Finally, panel C shows the GxE

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

interaction wherein different genotypes display distinct phenotypes in response to the environment, hence demonstrating various plasticities.

3. Evolutionary relevance: phenotypic plasticity as a source of variation

Arguably, the most significant question regarding phenotypic plasticity concerns its evolutionary implications. Given that all organisms may be subjected to changes in the expression of their genomes, the evolutionary implications of plasticity were greatly debated during the course of the establishment of the concept (West-Eberhard 2003; Ghalambor et al. 2007; Pfennig 2021c). Extensive literature exists on the subject, and in this section we provide an overview of the evolutionary implications of plasticity through the *plasticity-led evolution hypothesis* (West-Eberhard 2003). A landmark example of the evolution of environmentally-induced phenotype, known as the process of genetic assimilation, was first stated and demonstrated by Conrad Waddington in 1953 (Waddington 1953). With the assistance of his collaborator, Evelyn Paton, Waddington provided evidence for the genetic assimilation of an environmentally induced trait in *D. melanogaster* (Waddington 1953). When fruit flies were exposed to a heatshock at 40°C during the pupation stage, they developed an altered crossvein phenotype in their wings (Waddington 1953). From these treated flies, two experimental selection lines were established: one selected for the crossveinless phenotype and the other for its absence, both under normal temperature conditions. Remarkably, within 23 generations, the crossvein phenotype became nearly fixed in all replicates of the selected population, even without further environmental induction (Waddington 1953). This study illustrates how an environmentally induced phenotype can later become genetically assimilated through selection, highlighting the crucial role of plasticity in evolutionary processes. This experiment also testifies of the inheritable nature of the plastic response, through the fixation of environmentally induced phenotypic traits.

Genetic assimilation may, in some cases, represent the final outcome of evolution driven by phenotypic plasticity, however, it is not the only way plasticity can contribute to adaptation. Pfenning, in his chapter on "*Key questions about phenotypic plasticity*", highlights three non-mutually exclusive hypotheses on how plasticity can facilitate adaptive evolution (Pfennig 2021a; 2021b). The first is the *buying time* hypothesis, which postulates that plasticity provides organisms with a within-generation buffering safety net against environmental perturbations, allowing them to survive long enough for genetic adaptations to arise (Figure 1.2) (Diamond and Martin 2016; Scheiner et al. 2017; Fox et al. 2019; Diamond and Martin 2021). Phenotypic plasticity may provide a short-term solution, allowing organisms to adjust to new conditions – even if the response is maladaptive – by changing their traits in response to

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

environmental cues. The second is the *non-genetic inheritance* hypothesis, which states that plasticity may result in adaptation through inheritance that is independent of the DNA sequence change. This includes parental effects such as the transmission of environmentally induced epigenetic marks (Jones 2012; Aliaga et al. 2019). Environmental factors can shape these epigenetic modifications, which in rare cases may be transmitted to the offspring (Figure 1.2) (Dominguez-Salas et al. 2014; Heard and Martienssen 2014; Ryu et al. 2018; Fitz-James and Cavalli 2022; Husby 2022). The third, and for this project the most relevant one, is the *plasticity-led evolution hypothesis* (also *plasticity-first theory*) (West-Eberhard 2003; Levis and Pfennig 2019a). While advocating for the *plasticity-led evolution*, West-Eberhard outlined the process in a series of sequential steps (West-Eberhard 2003; Levis and Pfennig 2019a). First, new environmental conditions expose pre-existing genetic variants, leading to changes in the phenotype (Ledón-Rettig et al. 2014; Levis and Pfennig 2019b; Pfennig 2021a). Second, if a variant provides an adaptive advantage compared with the original phenotype in the new environment, it may spread through the population, facilitated by its initial cryptic presence (Levis and Pfennig 2019b; Pfennig 2021a). Finally, genetic assimilation occurs when the adaptive variant becomes fixed within the population, strengthening the adaptation (West-Eberhard 2003). This framework position genes as “followers” in the evolutionary process as described before (West-Eberhard 2003; Ghalambor et al. 2007).

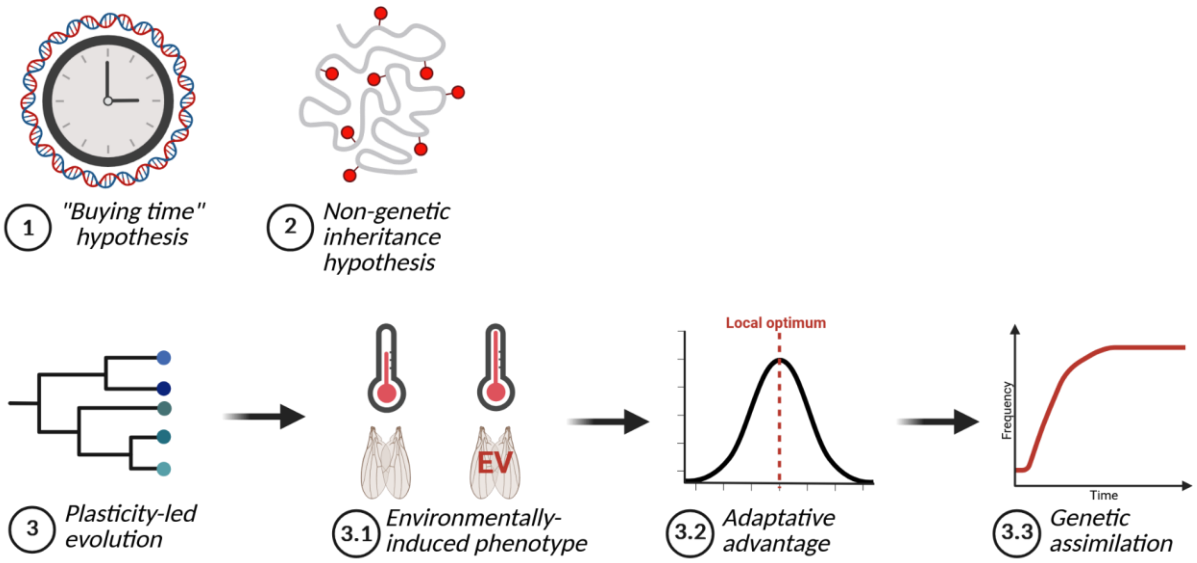


Figure 1. 2: Illustration of the ways phenotypic plasticity may contribute to evolution.
1) The buying time hypothesis depicts how plasticity may contribute as a buffering mechanism until selection of genetic adaptation. 2) The non-genetic inheritance hypothesis stipulates plasticity may occur from the emergence of inheritance of epigenetic marks. 3) The plasticity-led hypothesis occurs in three steps, from emergence of an environmentally-induced phenotype to the fixation of said phenotype in the population, provided it is selected (panels 3.1 to 3.3).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

4. Mechanisms underlying plasticity

In the following section, I will address the different known mechanisms that induce plasticity when targeted by environmental stimuli. Fortunately, with advancements in molecular biology and sequencing technologies, scientists can now assess genome-wide variation in gene expression and gene regulatory networks, which has made it possible to study phenotypic plasticity on a mechanistic level (Lewontin 2000; Nijhout 2003; Carroll et al. 2005; Gilbert 2012; Stinchcombe and Kirkpatrick 2012). Trudy Mackay is a pioneer in quantitative trait genetics and has contributed significantly to the understanding of phenotypic response to environmental changes (Vieira et al. 2000; W. G. Hill and Mackay 2004; Mackay and Lyman 2005; Mackay et al. 2009; 2009; Zhou et al. 2012; Morgante et al. 2015; Mackay and Anholt 2024). Her research has been particularly influential in the study of *D. melanogaster* quantitative traits and plasticity (Vieira et al. 2000; Mackay and Lyman 2005; Carbone et al. 2006; Jordan et al. 2006; Mackay et al. 2012; Zhou et al. 2012; Morgante et al. 2015; Huang et al. 2020). In her most recent work, Mackay assumes that phenotypic plasticity is an expression of transcriptomic plasticity, and considering the transcriptome as a molecular phenotype she investigates how gene expression varies under different environment or stresses (Zhou et al. 2012; Morgante et al. 2015; Huang et al. 2020). This allows to capture insights at the molecular level and interrogate the underlying mechanisms involved in phenotypic changes (Zhou et al. 2012; Morgante et al. 2015; Huang et al. 2020). One notable study in an outbred *D. melanogaster* population, investigated transcriptomic plasticity across a diverse range of 20 environmental treatments, from heat shock to caffeine exposure (Zhou et al. 2012). The study found that 15% of the transcriptome showed differential expression across conditions. Among them, 25 transcription factors displayed a modulation of their expression, with enrichment in detoxification, metabolism, proteolysis, and heat shock response pathways (Zhou et al. 2012). Additionally, four key fitness traits – development time, lifespan, starvation resistance, and chill coma recovery – were measured to evaluate the extent to which transcriptomic plasticity influences organismal plasticity. Each phenotype was significantly affected by environmental conditions, with 426, 186, 320, and 440 transcripts differentially expressed for the four traits respectively (Zhou et al. 2012). Notably, there were distinct differences between males and females in their transcriptomic responses (Zhou et al. 2012). This transcriptomic plasticity raises the question of the underlying mechanisms that initiate these responses.

I will now introduce a wide array of mechanisms that may play a part in generating such gene expression plasticity, and that are not mutually exclusive. Indeed, the interplay between the different

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

molecular players described below (hormones, transcription factors, heatshock proteins, epigenetic mechanisms and microRNAs) is complex and most likely intertwined. We need to keep in mind that determining the mechanism by which the cascade of events leading to phenotypic plasticity is initiated is a challenge, as these mechanisms are often activated simultaneously and influence one another. Rather than acting in a strict linear pathway, they are each representing a stepping stone adding to the organism's response to environmental variation.

a. Hormones

Hormones are molecular messengers secreted by the endocrine glands of an organism (Heyland et al. 2005; Tata 2005; Nelson 2010). They are highly sensitive to environmental perturbations, whether internal (cellular environment) or external (organism's surroundings) and are therefore highly relevant to plasticity and adaptation (Dufty 2002). Various environmental signals, such as photoperiod, temperature, and extrinsic pheromones, are involved in controlling hormonal pathways (Wingfield 2008; Lema and Kitano 2013; Lema 2014; 2020; Little and Seebacher 2024). Some hormones bind to cell receptors to activate a physiological response, whereas others bind to nuclear receptors to modulate gene expression (Aranda and Pascual 2001; Bassett et al. 2003; Zhu et al. 2003; Hammes and Levin 2007; Lema and Kitano 2013). The binding of the hormone can induce a change in physiological or behavioural state (Aranda and Pascual 2001; Dufty 2002; Nelson 2010; Lema 2014). Some have described hormones as "molecular connectors" of the organism to its environment (Pigliucci 2001; West-Eberhard 2003). In many insects, including *D. melanogaster*, well-known endocrine systems such as the insulin/insulin-like signalling (IIS) pathway, ecdysones, and the juvenile hormone (JH) play crucial roles in the response to environmental changes (Peric-Mataruga et al. 2006; Corona et al. 2016; Miura 2019; Bobrovskikh and Gruntenko 2023). The IIS system in *D. melanogaster* is known to influence body weight, organ size, and overall growth rate by regulating translational initiation, ribosome biogenesis, and transcription, among other processes (Taniguchi et al. 2006; Shingleton 2010; Green and Extavour 2014). Beyond its role in growth and development, the IIS system also plays a crucial role in behavioural responses. It has been implicated in regulating feeding behaviour, courtship and mate choice, and responses to various stressors, including temperature fluctuations, dietary restrictions, and infections in adult flies (Nässel 2012; Weger and Rittschof 2024). Likewise, ecdysones can act as gene regulators, coactivators and corepressors, controlling metamorphosis in insects (Riddiford et al. 2000). Hormones also play a central role in an extreme form of phenotypic plasticity, known as polyphenism, where environmental cues lead to the development of discrete alternative phenotypes (Miura 2019). Well-documented examples of polyphenism are found in

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

social insects (Verma 2008; Miura 2019; Weger and Rittschof 2024). For instance, in termites, the titer of the JH influences caste differentiation (one given genome expressed into different castes): higher levels of JH can guide an individual's developmental fate toward becoming a worker or an alate, the latter being the precursor to reproductive forms such as queens or kings (Cornette et al. 2008; Miura 2019). Similarly, in aphids, wing polyphenism is also regulated by the JH. Elevated levels of this hormone inhibit the development of the wings, leading to the production of wingless aphids under certain environmental conditions (Verma 2008; Ishikawa et al. 2013; Miura 2019).

These are a few studies demonstrating the importance of hormones in modulating the phenotypic expression. However, it is worth noting that the majority of the studies, and thus most of the empirical data, focus on the role of hormones during developmental stages as well as in polyphenism, which falls outside the scope of this project. Hormones' implication in mediating plastic responses to environmental changes during adulthood are comparatively less explored.

b. Transcription factors

Transcription factors (TFs) are proteins, with DNA-binding properties, that regulate the activation or repression of gene transcription (Latchman 1997; Lee and Young 2000; O'Malley 2007). TFs are liable to the environment, and as such, they serve as the maestros of the gene expression symphony, orchestrating physiological and behavioural changes in organisms and playing a crucial role in plastic responses (Nijhout 2003; O'Malley 2007). Some TFs, known as master regulators, sit at the top of regulatory cascades, controlling the expression of other TFs and genes. These master regulators can initiate entire pathways, profoundly influencing organismal development and responses (Graba et al. 1997; Arnone and Davidson 1997; Chen and Zhu 2004; O'Malley 2007; Kin Chan 2013; Song et al. 2016). Following are a few examples of the implication of TF in promoting plastic responses in *D. melanogaster*.

For example, it was demonstrated that the complex of the NrF2 TF with its associated repressor Keap1 (Nrf2/Keap1), is involved in the response to oxidative stress by activating antioxidant and detoxification responses, ultimately regulating longevity (Sykiotis and Bohmann 2008). Another study investigated the role of the Mondo/ChREBP-Mlx transcriptional network in modulating adaptation to changes in sugar-rich diets (Havula et al. 2013). In the latter study, Mlx mutants displayed an impaired energy metabolism and a significantly reduced tolerance to sugars, including sucrose, glucose, and fructose (Havula et al. 2013). These metabolic deficiencies were associated with the disruption of the transcriptional response that is normally induced in high-sugar conditions (Havula et al. 2013). The study

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

emphasizes the essential role of the Mondo/ChREBP-Mlx transcriptional network in enabling *D. melanogaster* to adapt and exploit various sugar-rich diets, highlighting how transcriptional flexibility supports metabolic flexibility in response to environmental variation (Havula et al. 2013). Still in *D. melanogaster*, the TF FOXO, has been described to be activated whenever nutrition and the IIS are low, resulting in the inhibition of cell growth and division, thereby regulating phenotypic plasticity (Tang et al. 2011; Shingleton and Tang 2012). This last example exposes the interplay between different levels of mechanisms responsible for the expression of plasticity, as for example TF and hormones in this case.

This highlights the pivotal functions of TFs in controlling gene expression and acting as mediators of environmental responses. TFs can respond immediately to environmental signals or act as intermediate regulatory molecules in other molecular processes, such as hormonal or epigenetic pathways. As master regulators, their function allows them to respond to several environmental inputs and orchestrate physiological and behavioural responses, thereby playing a central role in the organism's plasticity.

c. Heatshock proteins

Due to their poikilothermic nature, insects are highly sensitive to temperature, making it one of the primary environmental factors driving plastic responses (Kerkut and Taylor 1958; Colinet et al. 2010; Wojda 2017; González-Tokman et al. 2020; Perez and Aron 2020; Lahondère 2023). Heatshock proteins (HSPs) are key actors in temperature responses as well as in general phenotypic plasticity and overall stress responses (Craig and Schlesinger 1985; Feige and Polla 1994; Koštál and Tollarová-Borovanská 2009; Sawarkar et al. 2012; Mahat et al. 2016; Sørensen et al. 2017; Chen et al. 2018a). Their role is indeed varied, as they can also be activated by other types of environmental stresses, such as chemical exposure (Planelló et al. 2008; 2010), hypoxia (Michaud et al. 2011), or drought and dehydration (Gusev et al. 2011). While their expression is regulated by heatshock factors (HSFs), a class of TF (Wu 1995; Jedlicka et al. 1997), their functional role is mechanical, acting as molecular chaperones. By assisting in the folding, refolding, and stabilization of proteins under stress conditions, HSPs help maintain cellular homeostasis (Lindquist 1986; Feder and Hofmann 1999; Zhao and Jones 2012). This chaperone activity also makes them an essential component of an organism’s response to environmental stress and changes, bridging gene regulation with direct physiological response (Zhao and Jones 2012; Feder and Hofmann 1999; King and MacRae 2015; Chen et al. 2018a). It is worth mentioning that variation in HSP expression may not solely represent an *active* regulatory response to environmental stimuli, but can also arise as a *by-product* of thermodynamic principles and their inherent sensitivity to environmental stressors (Hofmann 2005; Chen

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

et al. 2018b). This duality makes HSPs particularly valuable as indicators of stress, effectively acting as environmental "sensors" due to their inducibility. Interestingly, even when the induction is passive, it may still trigger downstream physiological or regulatory responses, activating broader stress-response pathways (Kingsolver and Woods 2016; Lima et al. 2016; Chen et al. 2018b).

Azad and collaborators reported the role of Hsp70 in the hemocytes of adult *D. melanogaster* in hypoxia and oxidative stress response (Azad et al. 2011). The overexpression of Hsp70 in hemocytes led to increased resistance to both stressors (Azad et al. 2011). More precisely, Hsp70 overexpression resulted in lower levels of reactive oxygen species (ROS) in the whole fly, a typical consequence of both hypoxic and oxidative stress (Azad et al. 2011). HSPs were also shown to participate in the response to cold stress in *D. melanogaster*. In a study by Colinet and collaborators, flies were subjected to prolonged cold exposure (9h at 0°C) followed by a 8h recovery period at 25°C (Colinet et al. 2010). Using RT-qPCR, the expression of 11 HSP genes (comprised of four conserved HSPs and seven small HSP of size below 30 kDa) was monitored (Colinet et al. 2010). Amongst these, only three genes did not respond to cold, while two, namely *Hsp70Aa* and *Hsp68,* were highly responsive and significantly upregulated during cold stress (Colinet et al. 2010). Notably, recovery time varied substantially amongst individuals and was positively correlated with the duration of cold exposure (Colinet et al. 2010). The longer the stress, the longer the recovery, and the fewer flies ultimately recovered, suggesting that cold stress causes damage beyond repair (Colinet et al. 2010). This study highlights the environment-inducible aspect of HSPs and positions them as sensors that may initiate a cascade of pathways in response to environmental cues.

These couple examples support the implication of HSPs in the response to environmental perturbations, whether passive, active, or both. Through their "sensor" and chaperone function, HSPs may influence many downstream pathways, making them an important hinge in phenotypic plasticity.

d. Epigenetic mechanisms

Epigenetic was introduced by Waddington in 1942, almost contemporarily to Bradshaw's introduction of the concept of phenotypic plasticity (Waddington 1942; Noble 2015; Henikoff 2023). As previously discussed, Waddington used *D. melanogaster* as a model organism to understand the variation in development upon temperature exposure, and doing so he developed the concept of the epigenetic landscape, which he proposed as the basis for genetic assimilation (Waddington 1953). Today, epigenetic mechanisms are recognized as mediators of not only developmental plasticity but also of phenotypic plasticity (Kelly et al. 2012; Duncan et al. 2014; Frías-Lasserre and Villagra 2017b; Villagra and Frías-

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Lasserre 2020; Duncan et al. 2022; Du and Goodisman 2024). Changes in gene expression that underlie plasticity are indeed tightly linked to epigenetic mechanisms and are the focus of this section.

First a definition is in order: epigenetic mechanisms involve non-DNA mediated changes in gene expression. Epigenetic mechanisms can be influenced by the environment and therefore mediate a physiological response (Jaenisch and Bird 2003; Feil and Fraga 2012; Kelly et al. 2012; Schrey et al. 2016; Du and Goodisman 2024; Gupta and Nair 2025). Various pathways contribute to the organism epigenetic landscape, such as DNA methylation, histone tail modifications and non-coding RNA (ncRNA) (Jaenisch and Bird 2003; Holliday 2006; Bewick et al. 2017; Frías-Lasserre and Villagra 2017b; Villagra and Frías-Lasserre 2020; Henikoff 2023; Chatterjee 2024). These mechanisms operate together to control gene expression transcriptionally or post-transcriptionally (Jaenisch and Bird 2003; Holliday 2006; Henikoff 2023). Here, we will describe the different epigenetic mechanisms known to impact gene expression under environmental influence. One way to look at it is to consider it as epigenetic plasticity, wherein the DNA undergoes epigenetic reprogramming upon stress exposure or environmental changes.

1. DNA methylation

DNA methylation refers to the modification of the DNA molecule through the addition of a methyl group to a cytosine (Razin and Cedar 1991; Singal and Ginder 1999; Moore et al. 2013; Mattei et al. 2022). Although we do not discuss in detail – since *D. melanogaster*, the model organism in this study, does not feature DNA methylation (Lyko et al. 2006) – here is a brief overview. DNA methylation is found in a wide range of organisms, from prokaryotes to eukaryotes (Martienssen and Richards 1995; He et al. 2011; Schmitz et al. 2019; Seong et al. 2021; Casadesús and Sánchez-Romero 2022). Depending on the organism, methylation can occur in a symmetric pattern, referred to as CpG methylation, where a cytosine (C) followed by a guanine (G) on the 5'-3' strand is methylated, and the cytosine on the complementary strand is also methylated. This pattern is common in mammals and is mostly associated with gene silencing and transcriptional repression when located at promoter sites (Razin and Cedar 1991; Singal and Ginder 1999; Mattei et al. 2022). Methylation can also occur in a non-symmetrical way, known as non-CpG methylation, where any C on either DNA strand can be methylated. This form is more commonly observed in plants (Zhang et al. 2018). In insects, the abundance and function of CpG methylation is highly variable across taxa (Bewick et al. 2017).

DNA methylation is liable to environment making it a relevant actor of the environmental response (Schrey et al. 2016; Duncan et al. 2022). A remarkable example of the implication of DNA methylation in generating phenotypic plasticity is found in bees. In eusocial bees (*Apis mellifera*), a

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

totipotent egg can develop into either a worker or a queen, depending on environmental inputs. Specifically, the type of food provided to the larvae is associated with DNA methylation changes (Maleszka 2008; Foret et al. 2009; Lyko et al. 2010; Alhosin 2023; Maleszka 2024). Indeed, the feeding of a larva with royal jelly will induce its development into a queen, while feeding it standard honey will yield a worker bee (Maleszka 2008; Foret et al. 2009; Lyko et al. 2010; Alhosin 2023; Maleszka 2024). Thus, the same genome can give rise to two distinct phenotypes. To investigate the mechanisms driving this caste differentiation, a study silenced DNMT3, a DNA methyltransferase (DNMT) in charge of the deposition of methyl groups in C residues, by using RNA interference (RNAi) in larvae (Kucharski et al. 2008). This manipulation resulted in 72% of the treated individuals to develop into queens, compared to a mere 23% "queen-like" individuals in the control group (Kucharski et al. 2008). These results suggest that DNMT3 inhibition has a developmental effect comparable to that of royal jelly, supporting the idea that DNA methylation is diet-responsive (Kucharski et al. 2008). Further investigation of the DNA methylation patterns in the brains of larvae and adult honey bees identified differentially methylated genes, further supporting that each caste possess a unique methylation landscape (Lyko et al. 2010). Interestingly, a correlation was observed between methylation levels and alternative splicing (Lyko et al. 2010). For example, the gene *GB18602* shows one splicing variant that is equally expressed in both castes, while a second variant is only upregulated in queen brains (Lyko et al. 2010). This alternative splicing is associated with increased methylation near a stop codon in workers, which may interfere with the splicing process (Lyko et al. 2010). In conclusion, DNA methylation in the context of phenotypic plasticity it is mostly studied as a process of caste differentiation, which, unlike the focus of this project, often pairs with developmental plasticity.

2. Histone post-translational modifications

Histone tail modifications also participate in transcriptional activation or repression of genes' expression. DNA is wrapped around a nucleosome (146 bp), which consists of two of each of the following histones: H2A, H2B, H3 and H4, forming an octamer (Marmorstein 2001; Zhang and Reinberg 2001; Zheng and Hayes 2003; Peterson and Laniel 2004). Each histone possess a tail, that can be post-transcriptionally modified in different positions (Figure 1.3) (Marmorstein 2001; Zhang and Reinberg 2001; Zheng and Hayes 2003; Peterson and Laniel 2004). Tails can be methylated (addition of a methyl group), acetylated (addition of an acetyl group), phosphorylated (phosphorylation of the position), ubiquitinated (ubiquitination of the position), or SUMOylated (covalent binding of a SUMO protein to the position) (Figure 1.3) (Marmorstein 2001; Zhang and Reinberg 2001; Zheng and Hayes 2003; Peterson and Laniel

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

2004). Different types of modifications at different positions can result in transcription activation (permissive mark) or repression (repressive mark) (Marmorstein 2001; Zhang and Reinberg 2001; Zheng and Hayes 2003; Peterson and Laniel 2004). As an example, H3K4me3 refers to the addition of a trimethyl group (me3) to the fourth lysine residue (K4) of histone tail (H3), and is a permissive mark, while H3K27me3 has a repressive function.

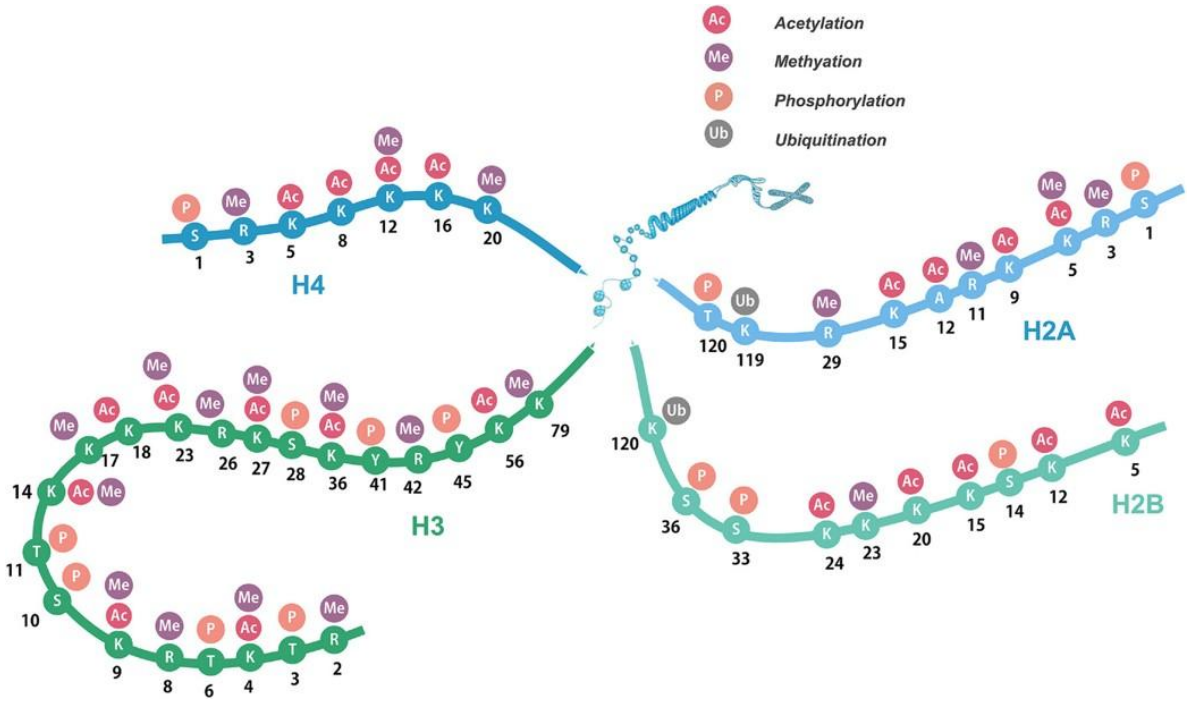


Figure 1. 3: Schematic representation of all the histone tail modifications that have been reported in the literature (Liu et al. 2023).

Several post-translational modifications can be added to histone tails, including acetylation (Ac), methylation (Me), phosphorylation (P), and ubiquitination (Ub). Methylation can occur as mono-, di-, or tri-methylation, corresponding to the addition of one, two, or three methyl groups, respectively.

A study showed that additional to DNA methylation, histone tail modifications were also involved in the plasticity of eusocial bees (Spannhoff et al. 2011; Polsinelli and Yu 2018). Indeed, it was discovered that royal jelly acts as an inhibitor of the histone deacetylase HDAC3, causing treated larvae to develop into queens (Spannhoff et al. 2011). Further investigation into the royal jelly composition revealed that it was in fact one of the primary ingredients, the fatty acid 10-Hydroxy-2-decenoic acid (10HDA), responsible for the HDAC3 inhibition, and was taking part in the induction of caste differentiation (Polsinelli and Yu 2018). Additionally, significant genome-wide differences in H3K4me3, H3K27ac and H3K36me3 were detected and highly correlated with caste-specific transcription patterns (Wojciechowski et al. 2018).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Specifically, H3K27ac, typically associated with active transcription, seem particularly pivotal in workers' caste differentiation when present in intronic regions (Wojciechowski et al. 2018).

In *D. melanogaster*, a study investigated the histone marks present in ovaries of diapausing individuals *vs.* non-diapausing ones, all genetically identical (Evans et al. 2023). None of the studied repressive marks showed any significant differences between the conditions, however, permissive marks H3K4me3 and H3K36me1 were significantly depleted in diapausing ovaries. To further correlate this decrease to the diapausing phenotype, genetically modified strains were knocked down (KD) for enzymes that either deposit or remove histone marks (*Set1*, *Set2*, methyltransferases of H3K4me3 and H3K36me1, respectively, and *lid* a H3K4me3 eraser). The two methyltransferase-mutants displayed an increase in diapausing individuals, while the eraser-mutant showed a decrease, hinting at a potential role of epigenetic marks in reproductive diapose (Evans et al. 2023). Additionally, the results were consistent across different genotypes, suggesting that the mechanism is similar within the *D. melanogaster* species (Evans et al. 2023). These are however examples illustrating plasticity of discrete trait which are very well studied as striking examples of integration of environmental signal into distinct phenotypes.

Additionally, histone mark modifications have also been linked to quantitative traits, such as the response to desiccation. A study in *D. melanogaster* investigated the transcriptomic response to desiccation and the epigenetic modifications associated with changes in gene expression (Sharma et al. 2017). Out of a panel of 60 stress-responsive genes, 75% were upregulated and 15% were downregulated in response to desiccation in the laboratory strain Canton-S. The desiccation treatment was applied to two heterozygous mutant lines: one for a Polycomb group (PcG) gene, associated with the deposition of the repressive mark H3K27me3, and another for a Trithorax group (TrxG) gene, associated with the deposition of the permissive mark H3K4me3 (Sharma et al. 2017). The PcG heterozygous mutants possessed reduced survival when exposed to desiccation stress, while TrxG heterozygotes were more resistant (Sharma et al. 2017). Building on these results, the authors found that the changes in expression of stress-responsive genes in Canton-S were correlated with changes in the expression of genes from the PcG and TrxG groups in whole flies exposed to desiccation (Sharma et al. 2017). They also reported significant differences in the H3K4me3/H3K27me3 ratio between control and desiccation conditions upstream of stress-responsive genes. For instance, in the case of gene *CG7069*, this ratio increased from 1.60 to 12.37 indicating a strong enrichment in permissive histone marks upon exposure to stress (Sharma et al. 2017). This study highlights the sensitivity of the histone mark deposition mechanism to the environment.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

The following study offers an example of the complex interplay between various mechanisms that contribute to phenotypic plasticity. Abdomen colouration is a highly plastic trait in *D. melanogaster* and varies with temperature (Gibert et al. 1998; 1999; 2016; 2017; Lafuente et al. 2024). In a study by Gibert and colleagues, it was demonstrated that the melanisation level of the abdomen is under the control of the TF *tan*, itself regulated by the upstream enhancer *t-MSE* (Gibert et al. 2016). The enhancer is suspected to be temperature sensitive and to induce the expression of *tan* under lower temperature conditions, here 18°C (Gibert et al. 2016). Additionally, they investigated the histone marks at the enhancer and at the promoter of *tan*, and showed an enrichment of the permissive H3K4me3 at the *tan* promoter at 18°C (Gibert et al. 2016). Although this is an example of developmental plasticity rather than phenotypic plasticity, this study still corroborates the intertwined relationship of the different actors of plasticity, here TFs and histone post-translational modifications.

Together, these examples represent only a handful of the most recent findings in the field of epigenetic plasticity. Among them, bees stand out as a particularly remarkable example, demonstrating how closely interconnected epigenetic mechanisms and hormonal pathways can be in mediating environmental responses. Altogether, these findings emphasize the central role of chromatin remodelling in shaping organismal responses to environmental change.

3. microRNA

MicroRNAs (miRNAs) are small ncRNA molecules that regulate gene expression in eukaryotes and, as such, they may generate phenotypic plasticity (Eddy 2001; Cai et al. 2009; Fu 2014; Frías-Lasserre and Villagra 2017a; Lu and Rothenberg 2018; Palazzo et al. 2024). MiRNAs are 18-25 nucleotide long and, once mature, are loaded onto the RNA-induced silencing complex (RISC) (Cai et al. 2009; Huntzinger and Izaurralde 2011; Lu and Rothenberg 2018). When sequence complementarity is achieved, the RISC-miRNA complex targets messenger RNA (mRNA) transcripts to promote translational repression. MiRNA-mediated repression may occur through cleavage of the mRNA, shortening of its polyadenylation tail, and reduction of its translation into proteins (Cai et al. 2009; Huntzinger and Izaurralde 2011; Bartel 2018; Lu and Rothenberg 2018). Here are few examples of the interplay between miRNA and the response to environmental perturbations and their contribution to phenotypic plasticity.

In 2014, Yang and colleagues demonstrated that *mir-133* inhibits aggressive behaviour in gregarious locusts by regulating dopamine synthesis (Yang et al. 2014). They found that *mir-133* plays a role in the post-transcriptional control of two genes involved in the dopamine pathway: *henna* and *pale* (*pal*) (Yang et al. 2014). Most interestingly, the overexpression of *mir-133* using the agomir technique

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

caused a behavioural shift from solitary and aggressive, to gregarious, a shift usually induced by environmental factors such as temperature, food availability, and population density (Yang et al. 2014). Empirical evidence for the involvement of miRNAs in plasticity is also well documented in caste differentiation in social insects. A study conducted in the ant *Solenopsis Invicta* – a species that undergoes age-dependent behavioural changes during adulthood triggered by a combination of genetic and environmental factors – revealed a crosstalk between *miR-279c-5p* and the insulin synthesis pathway (Chen et al. 2023). Specifically, inhibition of this miRNA by an antagomir led to an overexpression of the *Rab8A* gene, resulting in a reduction of insulin production and a behavioural shift of workers toward a nurse-like phenotype (Chen et al. 2023). Conversely, overexpression of *mir-279c-5p* increased insulin levels and promoted a shift from nurse-like to forager-like behaviour (Chen et al. 2023). This study highlights the role of *mir-279c-5p* as a key regulator of labour division plasticity in *S. invicta*, further underscoring the intricate interplay of multiple mechanisms contributing to environmental responsiveness.

We have highlighted two examples of miRNAs contribution to plasticity in insects in the context of caste and behavioural differentiation. However, miRNAs have also been studied in the context of responses to other stresses. Lyons and collaborators introduced the term *cryomiRs* to refer to miRNAs involved in the response to cold stress (Lyons et al. 2013). They support that miRNAs are impacted by extremely low temperatures, making them potential players in the organism's response.

In general, miRNAs appear to contribute to phenotypic changes, as seen in cases of behavioural shifts and stress responses, through their role as regulators of gene expression. However, establishing causation remains an open field of research, requiring further investigation to clarify their contribution to physiological and phenotypic changes in response to environmental perturbations.

e. Cryptic genetic variants

As previously explained, to assess phenotypic variation attributed to the environment (VE and VGxE), one must reduce cofounding effects. While studying fully clonal organisms allows the assessment of environmentally induced variation, such organisms can be very difficult to obtain for certain species and may not even exist. Moreover, clonal or isogenic lines are typically established under standard laboratory conditions. Thus, although we may assume we are working with clonal individuals, in reality the residual genetic variants are likely controlled by the standardized environment, resulting in uniform phenotypes across individuals (Schlichting 2008; Paaby and Rockman 2014). But upon exposure to

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

environmental perturbations, the mechanisms that buffered these variants may be disrupted (Schlichting 2008; Paaby and Rockman 2014). As a result, the then cryptic genetic variants (CGVs) become expressed, and genetic differences result in phenotypic diversity (Figure 1.4). These CGVs have a two-fold effect on phenotypic variation as they are subjected to both GxE and GxG interactions, the latter referring to their pleiotropic effect (Schlichting 2008; Paaby and Rockman 2014). Waddington research was already illustrating the role of CGVs in heritable phenotypic variation, by showing that when "identical" organisms are subjected to environmental conditions that disrupt their buffering systems, they exhibit heritable phenotypic variation (Waddington 1953; Rutherford and Lindquist 1998; Lahti et al. 2009; Paaby and Rockman 2014). In other words, CGVs are standing genetic variation that can be revealed by the environment (Paaby and Rockman 2014). Thereby, CGVs represent a reservoir of unexpressed variation that may promote rapid adaptive evolution by supplying matter for selection to act upon (Lewontin 1974; Gibson and Dworkin 2004; Ledón-Rettig et al. 2014; Schneider and Meyer 2017; Pfennig 2021a). The environment participates in inducing the phenotypic variation that will ultimately be selected (Pfennig 2021c).

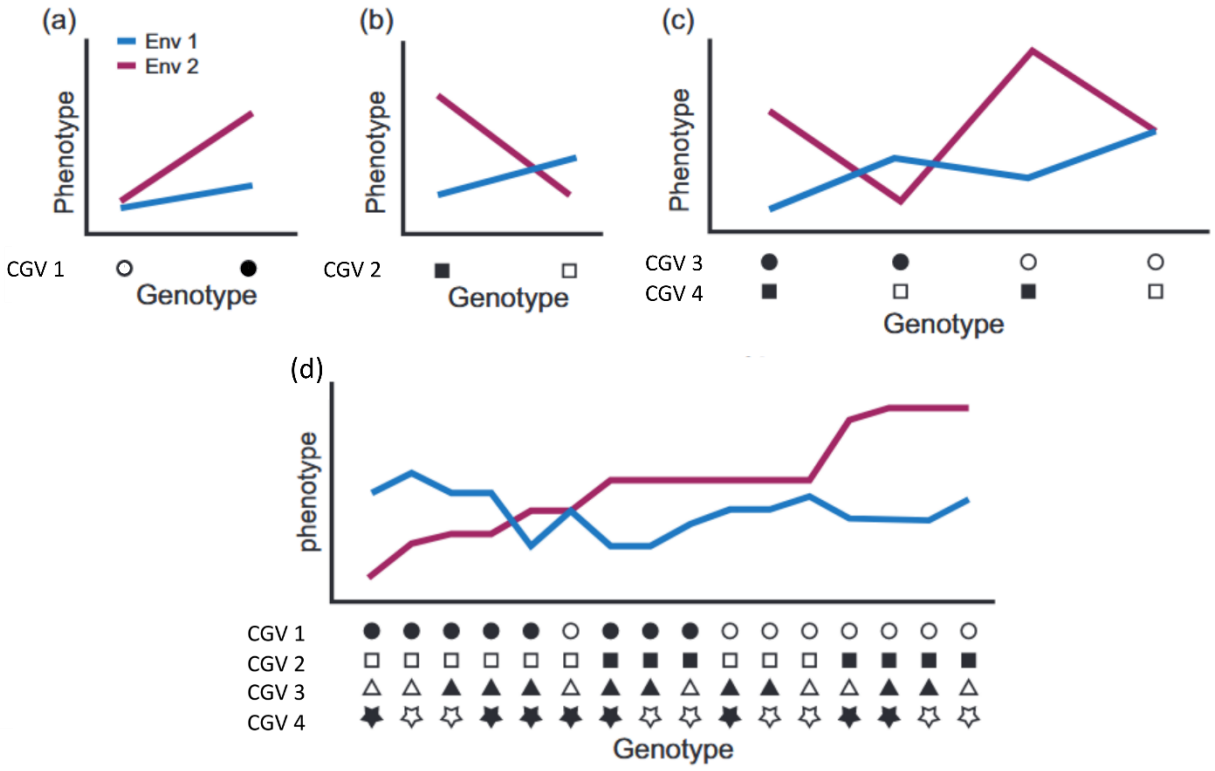


Figure 1. 4: Representation of contribution of cryptic genetic variants (CGVs) to genotype-by-genotype and genotype-by-environment interactions (adapted from Pfennig 2021c).
Each CGV presented in the figure above can be either present (filled shape) or absent (empty shape) from a genotype and have a differential contribution to the phenotype according to the environment. a) The

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

presence of CGV 1 has a stronger phenotypic effect in environment 2 (Env 2) compared to environment 1 (Env 1). b) The presence of CGV 2 has an antagonistic effect on the phenotype depending on the environment, whereas in its absence the environment has little influence. This illustrates a case of genome-by-environment (GxE) interaction, or more specifically CGV-by-environment (CGVxE) interaction. c) CGV 1 and CGV 2 contribute additively to the phenotype across environments. d) Combined additive and epistatic effects of four CGVs across environments, illustrating both CGVxE and gene-by-gene (GxG) interactions, or more specifically CGVxCGV interaction. Even in this simplified model, the combined effects of CGVxE and CGVxCGV interactions can shape phenotypes in complex ways.

The release of CGVs may occur through changes in epigenetic regulation or through the destabilization of buffering systems, such as molecular chaperones like HSPs. For instance, a recent simulation-based study by Romero-Mujalii and collaborators demonstrated that epigenetic mutations can hide alleles that are conditionally unfavourable until environmental conditions shift in their favour (Romero-Mujalli et al. 2024). This finding highlights how epigenetic regulation can act as a buffer against deleterious effects while preserving hidden genetic potential, that can be unveiled when advantageous, further reinforcing the evolutionary relevance of epigenetically mediated plasticity. It is also worth mentioning HSP90, which has been extensively studied in the context of stress-induced phenotypic changes (Rutherford and Lindquist 1998; Wong and Houry 2006; Sangster et al. 2008; Jarosz and Lindquist 2010; Chen and Wagner 2012; Rohner et al. 2013; Schrader and Schmitz 2019). Indeed, the role of HSP90 in leading to phenotypic variation, specifically through the release of CGVs during development, has been reported various times in insects and other taxa (Rutherford and Lindquist 1998; Wong and Houry 2006; Sangster et al. 2008; Jarosz and Lindquist 2010; Chen and Wagner 2012; Rohner et al. 2013; Schrader and Schmitz 2019). While screening for mutants of the *Hsp83* gene (encoding for HSP90) across 42 natural populations of *D. melanogaster*, Chen and Wagner (2012) selected three mutants for which the expression of *Hsp83* was reduced due to an insertion of a *P-element* in the proximal promoter region (Chen and Wagner 2012). These three populations displayed a reduction in fitness in a variety of traits, regardless of the genetic background. The populations were further inbred to expose cryptic recessive mutations, and by doing so, further decreased their fitness (Chen and Wagner 2012). Upon exposure to even mild thermal stress (28 °C), these inbred mutant lines went extinct due to the reduced expression of *Hsp83* impairing the buffering of deleterious CGVs (Chen and Wagner 2012). This highlights the role of HSP90 in keeping CGVs under control, but also its function as a driver of phenotypic variation when its chaperone capacity is compromised, whether due to a mutation or to environmental stresses.

The accumulation of mutations that remain cryptic until they are revealed by phenotype-altering conditions also falls in the concept of phenotypic or evolutionary capacitance. Evolutionary capacitance refers to an increased evolvability by the presence of selectable genetic variants (Bergman and Siegal

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

2003; Masel 2006; 2013). This also echoes the first step of the proposed *plasticity-led evolution* where "new environmental conditions expose pre-existing genetic variants, leading to changes in the phenotype" (West-Eberhard 2003).

Throughout the section II.4 describing the "Mechanisms underlying plasticity", many mechanisms have been reported. These are not mutually exclusive and participate in the plastic response in unison. Likely other mechanisms have been unwillingly omitted, and probably many more remain to be described. One of the key objectives of this project is the identification of a new mechanism linking environmental perturbations to phenotypic effects. Among the many features of genomes, transposable elements (TEs) stand out as particularly compelling drivers of plasticity due to their ubiquity and repetitive nature. Yet, TEs have been largely overlooked in studies focused on phenotypic plasticity. As we will explore in the next section, TEs encapsulate within a single genomic component many of the features underlying plasticity. Furthermore, TEs represent a form of CGV, thereby capable of generating novel phenotypes upon environmental perturbations. This makes them not only contributors to immediate plastic responses, but also potential sources of heritable variation that can facilitate genetic assimilation and long-term adaptation. In many respects, TEs fulfil many aspects associated with plasticity and *plasticity-led evolution*, making them relevant candidates for further investigation into the molecular basis of environment-driven phenotypic change. The following section will introduce TEs in more detail, covering their regulation, phenotypic impacts, and role in adaptive evolution.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

III. Transposable elements, between disruption and innovation

1. An introduction to transposable elements

TEs are genetic elements capable of changing position along the chromosomes, and in some rarer cases, even moving between genomes through horizontal transfer (Finnegan 1992; Bourque et al. 2018; Hayward and Gilbert 2022). TEs are found in virtually all organisms, though they vary greatly in both abundance and form (Bourque et al. 2018; Wells and Feschotte 2020). TEs' discovery initially sparked controversy, as is often the case with pioneering and disruptive discoveries. And, as it also frequently goes, it was only later – thanks to technological advances – that TEs came to be fully recognized as part of organisms' genomes and as an integral part of their evolution.

In the 1950s, Barbara McClintock was studying the colour patterns of corn kernels, particularly the intriguing variegated patterns (McClintock 1950; 1953; 1956). Some kernels displayed a distinct colour pattern where different cells expressed different colours. Curious about this phenomenon, McClintock performed crosses to better understand these peculiar patterns. The wild-type phenotype resulted in dark purple kernels, while the alternative phenotype produced light yellow kernels. The variegated phenotype however, featured purple spots on a yellow background. This phenotype appeared to illustrate the case of a reversible mutation. Delving deeper into the genetics behind these observations, McClintock developed a model to explain the emergence of the variegated phenotype. The model relies on the presence of two key elements: *Ds* (dissociation element) and *Ac* (activator element) (McClintock 1950). The *Ds* element could change the phenotype only in the presence of *Ac*, but *Ac* alone could also modify the phenotype (McClintock 1950). During kernel development, the *Ds* element could be excised from the gene controlling kernel colour, and cells derived from this excision event would express a different phenotype than the original zygote cells (McClintock 1950). Through this discovery, McClintock not only uncovered these moving genomic elements but also identified two distinct types: autonomous and non-autonomous elements. The autonomous *Ac* could move independently, while the non-autonomous *Ds* depended on the machinery provided by *Ac* to transpose within the genome. McClintock's pioneering work revolutionized our perception of genetic material by showing that genomes are not solely composed of genes (McClintock 1950; 1953; 1956).

Long after McClintock's discovery, TEs were generally viewed as genomic parasites, exploiting what was referred to as the host's transcriptional machinery to transpose, and often having deleterious effects on the organism (Shapiro 2000; Biémont 2010a; Feschotte 2023). Indeed, if one were to overlook

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

all evidence of their adaptive contributions, it would be easy to be overwhelmed by the many ways in which TEs negatively impact both the genomes and the organisms (Shapiro 2000; Biémont 2010a; Feschotte 2023). For one, TEs can disrupt gene function by inserting into coding sequences, modifying open reading frames (ORFs), introducing premature stop codons, interfering with proper transcript splicing, or supplying alternative polyadenylation signals (Dubin et al. 2018). TEs can modify genes' expression by providing *trans*- and *cis*- regulatory elements, thus altering gene expression timing and/or levels, sometimes with tissue-specific effect (Chung et al. 2007; Rebollo et al. 2012; Chuong et al. 2017; Moschetti et al. 2020; Sundaram and Wysocka 2020; Gebrie 2023). Additionally, their repetitive nature can facilitate ectopic recombination events, leading to harmful intra- and inter-chromosomal rearrangements (Colonna Romano and Fanti 2022). Some TEs may also cause double-stranded DNA (dsDNA) breaks, which trigger repair mechanisms that are often error-prone and therefore adding to their mutagenic properties (Wells and Feschotte 2020; Colonna Romano and Fanti 2022).

With advancements in sequencing technologies, step-by-step, these previously exclusively deleterious effects were associated with the discovery of adaptive mutations. TEs were therefore considered through a more nuanced and ambivalent lens, being recognized as components of genomes with deleterious effects, but also provide context-dependent, beneficial, and even adaptive functions (Biémont and Vieira 2006; Biémont 2010b; Stapley et al. 2015; Rey et al. 2016a; Nishihara 2019; Colonna Romano and Fanti 2022; Feschotte 2023). In the following sections, we will explore their classification, the regulatory mechanisms that keep their activity in check, their role as genomic and transcriptomic drivers, and ultimately, their contribution to evolution as CGVs and their environment responsiveness, with support from examples and theoretical insights.

a. Classification

TEs are highly diverse, exhibiting a variety of transposition mechanisms, structures, and spanning a wide range of lengths, from a few hundred base pairs to as much as 700 kilobases (for the biggest *Starship* transposons) (Arkhipova and Yushenova 2019; Gluck-Thaler et al. 2022). The current classification was proposed by Wicker and collaborators in 2007 and first separated TEs into two main categories that are Class I retrotransposons, and Class II DNA transposons (Wicker et al. 2007). These classes are further subdivided into subclasses, superfamilies and families according to the mechanism of transposition as well as the phylogenetic resemblance. Families are defined by sequence similarity which requires setting thresholds of homology requirements, commonly based on the 80-80-80 rule (Wicker et al. 2007). This

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

rule implies that a TE family includes all insertions that are at least 80 bp long, share at least 80% of identity over (you guessed it!) 80% of their length. Moreover, upon transposition most TEs leave target site duplications (TSD), generally considered a hallmark of transposition (Wicker et al. 2007; Wells and Feschotte 2020). TSDs are by-products of transposition events and are generated by nearly all TEs (Wicker et al. 2007; Wells and Feschotte 2020). TSDs are distinctive to each TE family, due to the specificity in the activity of the transposases and integrases, and can serve as signatures to identify past insertion events (Wicker et al. 2007; Wells and Feschotte 2020).

1. Class I retrotransposons

Retrotransposons are characterized by a transposition mechanism relying on the reverse transcription of an RNA molecule (Eickbush and Malik 2007; Goodier and Kazazian 2008). The TE mRNA is then translated, part of which code for a reverse transcriptase, which allows for the reverse transcription of the RNA product back into the DNA. Within the retrotransposon class, some TEs may also encode for an integrase and/or a Ribonuclease H (RNase H) (Eickbush and Malik 2007; Goodier and Kazazian 2008; Wells and Feschotte 2020). The former mediates the reintegration into the genome (Eickbush and Malik 2007; Goodier and Kazazian 2008; Wells and Feschotte 2020). Upon transposition, retrotransposons leave the template insertion in its original place while also inducing the insertion of a new copy into a new position along the genome, owing them the *copy-and-paste* alias. Retrotransposons are divided into three categories according to Wells and Feschotte recent review of TEs (Wicker et al. 2007; Wells and Feschotte 2020).

First, Long Terminal Repeat (LTR) retrotransposons are characterized by the presence of (as the name indicates) long repeated DNA sequences flanking the TEs. LTR elements are characterized by the presence of at least a set of two ORFs inherited from their retroviral origin, that are a *gag* and a *pol* coding sequences (Eickbush and Malik 2007; Chandler et al. 2020). The *pol* ORF contains the coding sequences for the reverse transcriptase, the integrase and, for some TE families, a RNAse H. Some of the retrotransposons have maintained the *env* coding sequence, providing them with fusogenic properties (cell-to-cell fusion) allowing them to infect other cells, through the formation of virus-like particles (VLPs) (Eickbush and Malik 2007; Chandler et al. 2020). These specific retrotransposons are often referred to as endogenous retroviruses (ERVs). LTR retrotransposons are transcribed into RNA, which is then exported to the cytoplasm, where the *gag* capsid protein is synthesized. Within this capsid, the RNA is reverse-transcribed into a circular dsDNA molecule. The capsid can either target the producing cell or invade neighbouring cells, and, if successful, the synthesized dsDNA may be integrated into a new position in the

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

genome by the integrase (Eickbush and Malik 2007; Chandler et al. 2020). Due to their LTR sequences, LTR retrotransposons are often subject to homologous recombination between the two LTR, which can lead to the excision of the sequence within, resulting in the formation of a solo-LTR insertion (Vitte and Panaud 2003). LTR retrotransposons have been very successful at invading certain genomes, as it is the case of *D. melanogaster*, representing around 62% of all TE insertions (Mérel et al. 2020).

Non-LTR retrotransposons typically encode two ORFs: the ORF1 is often dispensable depending on the TE family, while ORF2 encodes the endonuclease (EN) and reverse transcriptase (RT) (Eickbush 1992; Wicker et al. 2007; Grandi and An 2013; Wells and Feschotte 2020). Unlike LTR elements, non-LTR retrotransposons lack retroviral features (Jiang et al. 2014). Their integration into the genome occurs through a mechanism called Target-Primed Reverse Transcription (TPRT) mediated by the RT, in which the RNA is directly reverse-transcribed into DNA at the target site after an endonuclease (EN) has nicked the DNA at a specific recognition site (Luan et al. 1993). Often leading to the 5' truncation of the new copy resulting in dead-on-arrival insertions, this process can lead to the accumulation of non-autonomous non-LTR retrotransposons, alongside autonomous ones, that can hijack their molecular actors for their own transposition (Wells and Feschotte 2020). Long interspersed nuclear element-1 (LINE1) are abundant non-LTR retrotransposons that constitute approximately 17% of the human genome and represent the last known active autonomous TE family in humans (Lander et al. 2001). By extension, *Alu* elements, which are primate-specific SINE elements (short interspersed elements) relying on the molecular machinery of LINE1, also represent one of the remaining active element of the human genome and account for nearly 11% of it (Deininger 2011).

One particular mention needs to be made for the *Penelope* non-LTR element for its unique pseudo-LTR and for its GIY-YIG (amino acid sequence motif) endonuclease domain, that make it a stand-alone TE family (Evgen'ev et al. 1997; Evgen'ev and Arkhipova 2005; Arkhipova 2006; Schostak et al. 2008; Wells and Feschotte 2020). *Penelope* has only one ORF that encodes both the GIY-YIG EN and an RT (Schostak et al. 2008). They transpose with a TPRT-like mechanism and have been previously described to have a slight target preference for AT-rich regions of the genome (Pyatkov et al. 2004). They are found across the entire tree of life, from *Drosophila virilis,* where they were first described, to other eukaryotes, such as protist, fungi and plants (Arkhipova 2006).

The third and last categories of Class I retrotransposons are the YR elements, also known as DIRS (Dictyostelium Intermediate Repeat Sequence). DIRS resemble LTR retrotransposons except they encode a tyrosine recombinase (YR) instead of an integrase (Wells and Feschotte 2020). DIRS are characterized by terminal inverted repeats (TIRs) (Poulter and Butler 2015). However, unlike LTR elements, DIRS

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

integrate into the genome without generating a TSD footprint (Cappello 1985; Goodwin and Poulter 2004). For reference, DIRS account for 78% of all repeats of the caecilian *Ichthyophis bannanicus* genome (J. Wang et al. 2021).

2. Class II DNA transposons

DNA transposons comprise elements that use a single-stranded or dsDNA intermediate for their transposition into the genome (Feschotte and Pritham 2007; Wells and Feschotte 2020). Unlike Class I retroelements, most Class II DNA transposons mobilize *via* a *cut-and-paste* mechanism, meaning they typically do not duplicate but rather move from one genomic position to another (Feschotte and Pritham 2007; Wells and Feschotte 2020). Class II elements are divided into four categories: 1) *cut-and-paste* transposons that transpose *via* DDE transposases (Doak et al. 1994), 2) *Cryptons,* that transpose by means of a double-stranded break (Kojima and Jurka 2011), 3) *peel-and-paste Helitrons*, which transpose through a rolling-circle mechanism (Grabundzija et al. 2016), and 4) *Mavericks/Polintons*, which are self-synthesizing transposons as they encode their own polymerase, unlike all other TEs (including Class I) (Kapitonov and Jurka 2006).

DDE-transposase-mediated transposons represent a diverse family that includes some well-studied TEs, such as the *P-element* – a model transposon in TE research in *D. melanogaster* (Coen et al. 1994; Ryder and Russell 2003; Majumdar and Rio 2015; Ghanim et al. 2020) – as well as *PiggyBac*, a TE that has been repurposed as a genetic engineering tool (Kim and Pyykko 2011; Gregory et al. 2016). This category of TEs encodes a transposase that mediates dsDNA breaks at the terminal inverted repeats (TIRs) flanking the TE coding sequence (Nesmelova and Hackett 2010). DDE transposons are rather simple, as they typically encode for a single ORF (Nesmelova and Hackett 2010; Wells and Feschotte 2020). The cleaved DNA is then circularized and reinserted at a new genomic location by the same transposase. The excision site also results in TSDs (Curcio and Derbyshire 2003). Additionally, some DDE transposons have lost their coding sequences over time, rendering them non-autonomous and dependent on the transposition machinery of other autonomous elements (Nesmelova and Hackett 2010; Wells and Feschotte 2020). A well-known example of such elements are the Miniature Inverted-Repeat Transposable Elements (MITEs), which rely on related transposases for their mobilization (Fattash et al. 2013).

Cryptons are DNA transposons that, like YR retrotransposons, transpose using a tyrosine recombinase. *Cryptons* are relatively simple transposon in that they encode a single ORF (Kojima and Jurka 2011). Their transposition mechanism involves dsDNA breaks, circularization of the excised sequence, and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

reintegration into a new genomic position, without leaving a trace at the original insertion site (Kojima and Jurka 2011). *Cryptons* were first described in basidiomycete and the ascomycete fungi, but they have since been described in a few other species, including animals such as in the mosquito *Aedes aegypti*, however they remain largely unknown (Goodwin et al. 2003; Kojima and Jurka 2011).

Recently a new tyrosine recombinase DNA element family, known as *Starship*, was described in filamentous Ascomycetes (Gluck-Thaler et al. 2022). These very large TEs (110 kb) are often referred to as cargo-carrying elements and carry along many host accessory genes, called *Cargos,* tailing a *Captain* sequence, and are known to encode for traits that can either be deleterious or adaptive, depending on the ecological context (Cheeseman et al. 2014; Brockhurst et al. 2019; McDonald et al. 2019; Gluck-Thaler et al. 2022; Urquhart et al. 2022). *Starships* pushes further the symbiotic nature of TEs providing substantial host fitness benefits associating them to potential mutualistic genetic elements that may cooperate with the host, rather than parasitic sequences (Gluck-Thaler et al. 2022).

As for *Helitrons*, they are DNA transposons using a rolling-circle mechanism to transpose, also known as *peel-and-paste* mechanism (Thomas and Pritham 2015). To that end, they encode a rolling-circle replication endonuclease (Thomas and Pritham 2015). Autonomous *Helitrons* encode three ORFs (Thomas and Pritham 2015). Unlike other Class II elements, *Helitrons* duplicate themselves, leaving the original template at the original insertion site as well as inserting a new copy into the landing position (Thomas and Pritham 2015). Like *Crypton* elements, *Helitrons* are characterized by the absence of TSDs. Most *Helitrons* discovered thus far are non-autonomous elements (Barro-Trastoy and Köhler 2024). *Helitrons* have often been associated with gene capture mechanisms, referring to the process by which these TEs can lead to the accidental hitchhiking of nearby genes or gene fragments during transposition (Barro-Trastoy and Köhler 2024). As such, *Helitrons* have been repeatedly implicated in generating genomic novelties (Barro-Trastoy and Köhler 2024).

Finally, *Mavericks/Politons* – described as 'the most enigmatic' by Wells and Feschotte (2020) – are notable for their large size (15 to 20 kb) and high structural complexity (Haapa-Paananen et al. 2014; Wells and Feschotte 2020). These elements encode their own polymerase and DDE recombinase, and may contain up to 20 protein-coding sequences, all flanked by long TIRs (Kapitonov and Jurka 2006). *Mavericks/Politons* share similarities with free-living dsDNA viruses (Kapitonov and Jurka 2006). Upon insertion, they generate TSDs at the original site (Kapitonov and Jurka 2006; Haapa-Paananen et al. 2014). Interestingly, some *Mavericks/Politons* also encode capsid-like proteins, similar to some LTR retrotransposons, further reinforcing their viral homology (Krupovic et al. 2014).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

The classification of TEs highlight the remarkable diversity of TEs, stemming from their diverse origins and distinct transposition mechanisms. It also points out the multiple ways in which TEs can influence genome architecture, transcriptomic composition, and ultimately, the biology and physiology of organisms. However, it is important to note that this classification does not reflect the prevalence of these different TE subclasses across the tree of life. Indeed, not all organisms carry the same types of TEs, and both overall TE abundance and the relative proportions of TE subclasses and superfamilies can vary drastically between species (Wells and Feschotte 2020).

b. The regulatory mechanisms

To counteract TEs deleterious effects, evolution has favoured the emergence of regulatory mechanisms. Some of these mechanisms will sound familiar, as they have already been introduced in the section II.4.d "Epigenetic mechanisms" as a plasticity mechanism. We will however expand on the role of these mechanisms in TE control, with a specific focus in insects and particularly in *D. melanogaster*, where many of these mechanisms have been extensively described.

1. siRNA pathway: somatic regulation of TEs

SiRNAs are small ncRNAs that mediate sequence-specific degradation of RNA transcripts (Dana et al. 2017). The siRNA pathway is involved in the regulation of TE expression in both somatic and germinal tissues, although it is not its primary function, as it is mostly acknowledged for its role in viral infection defence (Czech et al. 2008; Rij and Berezikov 2009; Hyun 2017; Onishi et al. 2021). siRNA biogenesis depends on the presence of a double-stranded RNA (dsRNA) molecule, which, in the case of TEs, can originate from bidirectional transcription of TE copies, homologous pairing of transcripts from distinct insertions, and potentially from self-hybridization of LTR or TIR regions of TE transcripts (Okamura et al. 2008; Miyoshi et al. 2010; Claycomb 2014). The dsRNA is processed by the Dicer-2 protein and its cofactor Loquacious-PD to cleave it into 21-nucleotide-long fragments (Czech et al. 2008; Rij and Berezikov 2009; Hyun 2017; Onishi et al. 2021; Wei et al. 2023). These 21-nucleotide siRNAs are then loaded onto the Argonaute-2 protein with the help of its partner protein, R2D2 (Czech et al. 2008; Rij and Berezikov 2009; Hyun 2017; Onishi et al. 2021; Wei et al. 2023). Upon loading, the complementary strand is removed, forming the RISC (Gebert and Rosenkranz 2015) (Figure 1.5). When this complex encounters a complementary TE RNA target, the siRNA guides the RISC to slice it (Figure 1.5).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

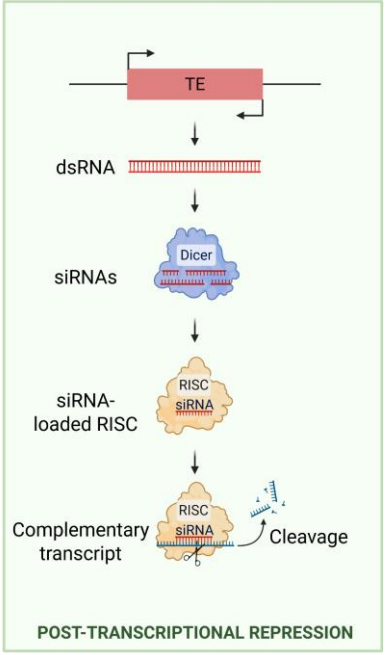


Figure 1. 5: Schematic representation of the siRNA pathway (adapted from Merenciano et al. 2025).
Fragments of dsRNA that may ensue from bidirectional transcription of TEs are processed by Dicer into 21-nt-long fragments and then loaded onto Ago2 to form an RNA-induced silencing complex (RISC) and cleave complementary transcripts, thereby controlling TE expression post-transcriptionally.

2. piRNA pathway: a predominantly gonadal regulation of TEs

The Piwi-interacting sRNA pathway represents one of the first line of defence against TEs in many organisms (Zamore 2010; Czech and Hannon 2016; Czech et al. 2018; Onishi et al. 2021; X. Wang et al. 2023). In *D. melanogaster*, the piRNA pathway is active in the female and male germline and in the ovarian somatic cells. It has also been demonstrated to be active in the fat body of both larvae and adults, where it plays an important role in the proper establishment of the IIS system (Jones et al. 2016; Yun et al. 2024).

Somatic piRNA biogenesis happens in the somatic ovarian cells, and begins with the transcription of a long precursor RNA molecule from specific genomic loci known as piRNA clusters (Zamore 2010) (Figure 1.6). These clusters are composed of TE sequences, some of which may be degenerated, that once processed into mature piRNAs enable the silencing of TE transcripts through antisense complementarity (Czech and Hannon 2016). One of the most well-characterized piRNA clusters is the *flamenco* locus, which is unidirectionally transcribed by RNA polymerase II in the ovarian soma (Goriaux et al. 2014). The early steps of piRNA biogenesis occur in the cytoplasmic Yb bodies, where the precursor RNA is first bound by *Piwi* and *Armitage*. These proteins facilitate the transfer of the precursor RNA to *Zucchini*, an endonuclease anchored to the outer mitochondrial membrane. *Zucchini* processes the precursor RNA into 23-31-nucleotide-long sRNAs. These processed piRNAs are then loaded onto *Piwi* proteins to form phased

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Piwi-associated piRNAs and lead to TE transcriptional silencing by sequence complementarity binding of nascent TE transcripts, and the recruitment effectors of heterochromatinization (Zamore 2010; Czech and Hannon 2016; Czech et al. 2018; Onishi et al. 2021; X. Wang et al. 2023; Choi 2024) (Figure 1.6). For one, the protein Asterix (Arx) binds to the piRNA and, through the recruitment of multiple effectors of which *Eggless* (also known as SETDB1), a H3K9 methyltransferase, and Heterochromatin Protein 1 (HP1), can, respectively, induce and maintain transcriptional silencing (Ninova et al. 2020; Onishi et al. 2021; K. Wang et al. 2021). Indeed, this complex promotes the heterochromatinization of TE loci through the deposition of H3K9me3 marks.

Germline piRNAs are predominantly derived from bidirectionally transcribed piRNA clusters (Czech and Hannon 2016; Czech et al. 2018; X. Wang et al. 2023; Choi 2024) (Figure 1.6). The resulting precursor RNAs are exported from the nucleus to the cytoplasmic Nuage granules, where they enter the *ping-pong cycle*, a self-reinforcing loop in which piRNAs are generated both from precursor and TE transcripts, thereby amplifying post-transcriptional TE control. In this process, TE-derived RNAs are initially cleaved by maternally inherited *initiator piRNAs* that are loaded onto *Aubergine* (Aub) proteins (Brennecke et al. 2008) (Figure 1.6). These Aub-bound piRNAs guide cleavage to 10 to 11 nucleotides downstream from the 5’ end of the targeting piRNA, of both piRNA precursors transcripts and TE transcripts. The newly generated antisense sRNAs are then loaded onto *Argonaute-3* (Ago3) proteins, forming the so-called *responder piRNAs* (Figure 1.6). These responder piRNAs then target complementary TE mRNAs, which are cleaved and subsequently loaded onto Ago3 proteins, forming *sense responder piRNAs* (Figure 1.6). These responder piRNAs, in turn, can target and cleave more precursor RNAs (typically rich in antisense degenerated TE fragments), thereby perpetuating the ping-pong cycle (Czech and Hannon 2016; Czech et al. 2018; X. Wang et al. 2023; Choi 2024). This feedforward cleaving pattern results in a 10 nucleotide overlap between sense and antisense piRNAs, known as the ping-pong signature (Brennecke et al. 2007). Additionally, a fraction of the Aub-bound precursor RNAs is carried outside the Nuage granule and handled by *Armitage*, which relocates them to the mitochondria. There, *Zucchini* cleaves them, and they are subsequently loaded onto Piwi, forming *phased piRNAs*. These phased Piwi-piRNA complexes, similar to somatic piRNAs, mediate the transcriptional repression of active TE insertions (Czech and Hannon 2016; Czech et al. 2018; X. Wang et al. 2023; Choi 2024).

In conclusion, the piRNA pathway not only functions post-transcriptionally but also contributes to transcriptional silencing of TEs.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

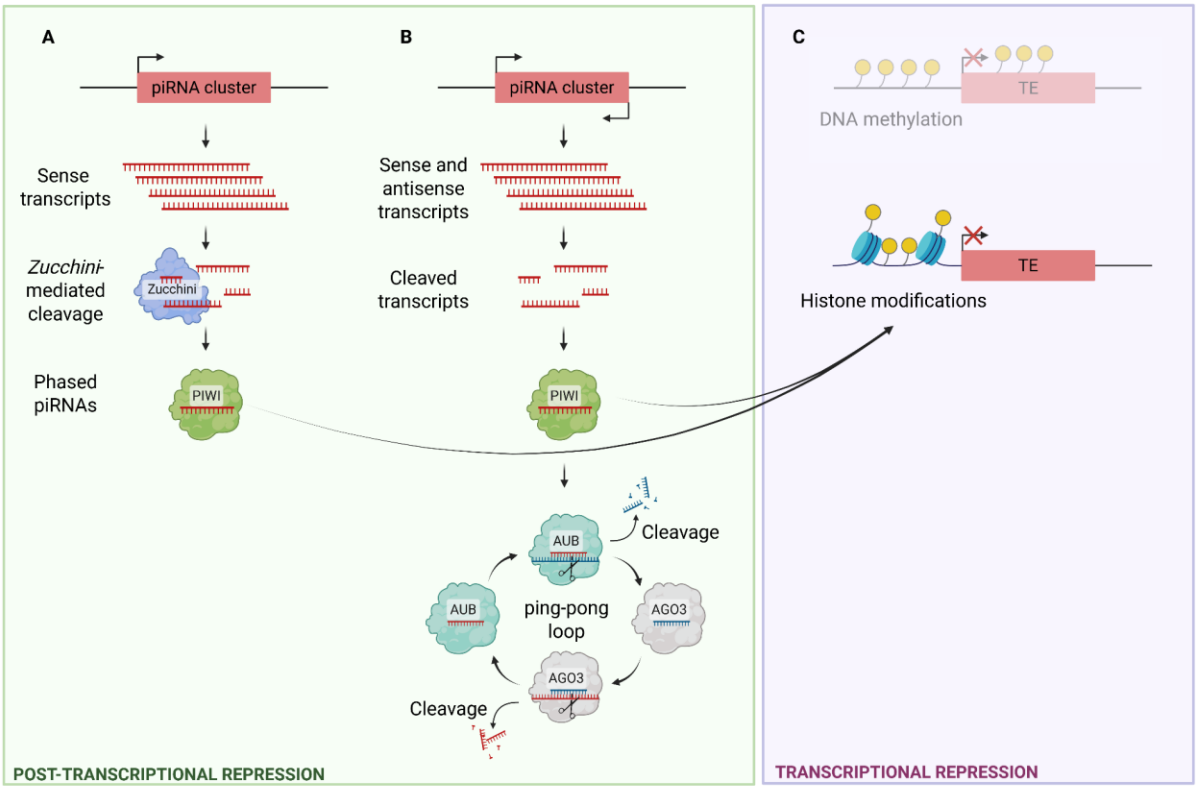


Figure 1. 6: Schematic representation of the somatic and germinal piRNA pathways (adapted from Merenciano et al. 2025).

A) Production of piRNAs in the somatic cells debuts with the unidirectional transcription of piRNA clusters, such as Flamenco. Precursor piRNAs are exported to the nucleus where they are cleaved into 23-31 nt-long fragments by Zucchini, anchored in the outer membrane of the mitochondria. After entering the nucleus, Piwi-loaded piRNAs can then recognize complementary nascent transcripts and initiate transcriptional silencing of TE insertions. B) Production of piRNAs in the germinal cells starts with sense and antisense precursor RNAs from bidirectional transcription of piRNA clusters. Precursors are cleaved by maternally inherited piRNAs to initiate the ping pong cycle, where both precursors and TE transcripts are cleaved and re-purposed into sense and antisense piRNAs. C) Both somatic and germinal piwi-bound piRNAs can re-enter the nucleus and by binding to nascent TE transcripts can promote deposition of heterochromatin marks at the TE loci.

3. Epigenetic regulation

While in many eukaryotes, DNA methylation is deposited on TEs as a mean to control their expression and prevent further transposition events, in insects, it is not the primary mechanism responsible for TE silencing (Field et al. 2004; Glastad et al. 2011). In fact, levels and patterns of DNA methylation vary greatly across insect orders: it is notably high in Hemiptera and Blattodea, somewhat present in Hymenoptera, and nearly to completely absent in Diptera, such as in *D. melanogaster* (Glastad et al. 2011; Bewick et al. 2017). Given the focus of this manuscript revolves around *D. melanogaster*, this

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

is be the only mention of DNA methylation's role in the control of TEs. Instead, we will now briefly focus on post-transcriptional histone tail modifications as epigenetic control of TEs.

Indeed, post-transcriptional histone tail modifications, have been well documented as playing a central role in TE control in insects (Slotkin and Martienssen 2007; Onishi et al. 2021). In *Drosophila*, TE silencing is also regulated by post-transcriptional histone tail modifications, notably through the deposition of H3K9me3, a repressive histone mark. This mark can be deposited by three methyltransferases: SETDB1, Su(var)3-9, and G9a (Yin et al. 2011; Boros 2012). As described previously, SETDB1 functions in a *piwi*-dependent manner to promote transcriptional repression of nascent TE transcripts by the deposition of H3K9me3 (Sienski et al. 2015; Czech and Hannon 2016; Onishi et al. 2021). In non-gonadal somatic tissues, TE silencing mainly depends on the maintenance of constitutive heterochromatin by Su(var)3-9 (Yin et al. 2011; Boros 2012). The enzyme HP1 binds to the methylated H3K9 and helps maintain the chromatin in a compacted conformation, thereby repressing transcription (Slotkin and Martienssen 2007; Onishi et al. 2021).

Together transcriptional and post-transcriptional silencing mechanisms are essential to protect organisms from the deleterious effects of TEs. Note that other silencing systems have been reported in various organisms, such as defence through the induction of point mutations of TE transcripts prior to their reintegration, by Repeat-Induced Point mutation (RIP) in fungi (Galagan and Selker 2004), or cytosine-to-uracil (C-to-U) deamination by APOBEC3 mutators in primates (Modenini et al. 2022). In eukaryotes, N6-methyladenosine (m6A) modifications of mRNAs also mediates the transcriptional silencing of genomic repeats, such as LINE1 elements in humans (Hwang et al. 2021; Barter and Cho 2025).

2. Transposable element, an evolutionary perspective

a. TE contribution to adaptative evolution

In addition to their neutral and deleterious effects, TE can also have beneficial effects on the organisms. The effect of a given insertion, or a set of insertions, may induce adaptive phenotypes and become an integral part of a species biology. As we will see now, the ways that TEs can contribute to evolution are diverse (Schrader and Schmitz 2019). For the sake of illustration, and given that examples are scarcer than in the case of phenotypic plasticity, we will broaden the scope in this section to include examples from various organisms, while keeping a focus on *D. melanogaster* whenever possible.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

TE sequences may be domesticated, meaning that TE-derived functions can be selected to contribute to an organism's biological functions (Capy 2021). A great example of this can be found in the evolution of *Drosophila*'s telomeres, which indeed stem from the domestication of TEs (*HeT-A*, *TART*, and *TAHRE*) (Pardue and DeBaryshe 2008). Another remarkable example of domestication can be found in the convergent evolution of the TE-derived *syncytin* gene in placental mammals, which plays an essential role in cell-to-cell fusion at the foetal-maternal interface (Lavialle et al. 2013). TEs can also be exapted, a process that refers to the acquisition of new functions derived from TE sequences (Capy 2021). For instance, exonized TEs can be exapted, wherein a TE inserted into a gene becomes a functional part of it (Joly-Lopez and Bureau 2018).

Importantly, even without full domestication, the sole presence and activity of TEs can drive evolutionary change by facilitating the co-option of their sequences for beneficial functions. For one, TEs may also promote the formation of retrogenes, a process in which a gene transcript is reintegrated into the genome *via* the TE machinery (Ding et al. 2006; Sciamanna et al. 2009; Kubiak and Makałowska 2017). In *D. yakuba* and *D. teissieri*, two African *Drosophila* species, *jingwei* was shown to be a functional retrogene of the *alcohol dehydrogenase* (ADH) gene, capable of efficiently utilizing long-chain primary alcohols found in hormone and pheromone metabolism (Zhang et al. 2004).

Because TEs carry the regulatory components required for their transcription, TE-driven adaptation has often been acquired through their *cis*-/*trans*- activity on nearby genes (Rebollo et al. 2012; Chuong et al. 2017; Moschetti et al. 2020). For example, the industrial melanism morph of the peppered moth (*Biston betularia*), known as *carbonaria*, was revealed to be induced by the insertion of a transposon in the first intron of the *cortex* gene. This insertion induces an overexpression of *cortex* and results in hypermelanisation (Van't Hof et al. 2016). This morph, during the industrial era, experienced an increased fitness as it was protected from predation (Van't Hof et al. 2016). In *D. melanogaster,* an example of *cis*-acting TE was described as being responsible for the emergence of insecticide resistance. Indeed, it was observed that the cytochrome P450 gene, *Cyp6g1,* was significantly more expressed in the insecticide-resistant strain, Hikone-R, compared to the non-resistant one (Chung et al. 2007). As it turns out, the Malpighian tubules, the midgut and the fat body display a differential expression of 10-40 fold higher in Hikone-R larvae (Chung et al. 2007). The presence of an *Accord* retrotransposon LTR sequence, 291 bp upstream *Cyp6g1,* was demonstrated to act as a tissue-specific enhancer, resulting in an increased expression and a more efficient detoxification leading to the resistance to insecticide (Chung et al. 2007).

Additionally, TEs may interfere with gene regulatory networks, resulting in changes in gene expression (Feschotte 2008). For example, in humans, the MER41 primate-specific endogenous retrovirus

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

family has contributed to the emergence of interferon-sensitive enhancers across the genome, providing biding sites for effectors of the interferon response (IFN), and thereby taking part in the evolution of the innate immune response to viral infection (Chuong et al. 2016). Likewise, the RLTR30B murid-specific ERV has significantly contributed to providing binding sites to the signal transducer and activator of transcription STAT1, a crucial *cis*-regulatory element of the interferon response, within the murine genome (Chuong et al. 2016).

Finally, TEs may also drive evolutionary innovation by contributing to alternative splicing events. For instance, it has been proposed that the insertion of an Alu element into an intron of the *TBXT* gene in the hominoid lineage caused the loss of the tail. In this scenario, the hominoid-specific Alu pairs with an ancestral Alu inserted in the opposite orientation, resulting in exon skipping and the production of an alternative mRNA isoform (Xia et al. 2024). Using CRISPR to generate transgenic mice lacking *TBXT* exon 6 led to a reduction or complete absence of the tail (Xia et al. 2024).

These are a handful of examples underscoring the impact of the activity and presence of TEs on evolutionary trajectories. While domestication and exaptation represent remarkable evolutionary outcomes, the majority of TE-related research suggests that their predominant evolutionary impact lies in altering gene expression, regulatory networks and in promoting ectopic recombination. These examples illustrated the effects of a single, or a small set of insertions, fixed in a species. However, even before an adaptive TE becomes fully retained into an organism's biology, TEs can already foster within-species variability through their diverse impacts on genome integrity and gene expression.

b. TEs as molecular drivers of variation

TEs are regarded as notorious drivers of genotypic, epigenetic, and transcriptomic variation (Chuong et al. 2017; Drongitis et al. 2019; Sundaram and Wysocka 2020; Gebrie 2023). The fact that individuals within the same species or population may not all carry the same TE insertions, a phenomenon known as insertional polymorphism, serves as standing genetic diversity (Bennett et al. 2004; Grzebelus 2006). Therefore, even in the context of otherwise highly similar genetic backgrounds, TEs serve as sources of genetic variation. These insertional differences can constitute CGVs, potentially leading to differential phenocopies across individuals (Piacentini et al. 2014; Marin et al. 2020; Catlin and Josephs 2022b).

TEs, as drivers of epigenetic diversity, contribute to the emergence of epigenetic variants known as epialleles (Marin et al. 2020). Indeed, through the mechanisms involved in their control, they may

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

impact the epigenetic landscape surrounding nearby genes. It has been repeatedly reported that epigenetic marks associated with TEs may spread to nearby genes, resulting in gene expression modulation (Rebollo et al. 2012; Chuong et al. 2017; Moschetti et al. 2020). A study reported that, in *D. melanogaster*, spreading of the repressive mark H3K9me2 occurs around more than 50% of euchromatic TEs, and can even reach up to 20 kb (Lee and Karpen 2017). As such, TE-derived epigenetic variants can be a source of adaptation, provided they have a beneficial impact on the organism's fitness. In the Chapter 6 of the book *Epigenetics and Evolution,* to which I contributed with the "Evolutionary consequences of epigenetic regulation of TEs" subsection, we cover both the deleterious and beneficial outcomes of TE-driven epigenetic diversity (Annexe 1).

On top of genetic and epigenetic diversity, TEs may also promote transcriptomic diversity (Figure 1.7). Indeed, TEs may influence the expression of adjacent genes, induce alternative splicing or contribute to the formation of chimeric transcripts composed of TE and gene sequences (Figure 1.7). Recently, we published the ChimeraTE pipeline to detect such chimeric transcripts, to which I contributed by confirming that the chimeric transcripts identified by ChimeraTE, through short-read RNA sequencing, were indeed present and could be amplified using RT-PCR (Oliveira et al. 2023) (Annexe 2). Approximately 1% of *D. melanogaster* genes were involved in gene-TE chimeras, with nearly 89% of these cases involving exonized TEs, meaning that the TE sequence is transcribed as an exon (Schmitz and Brosius 2011; Oliveira et al. 2023).

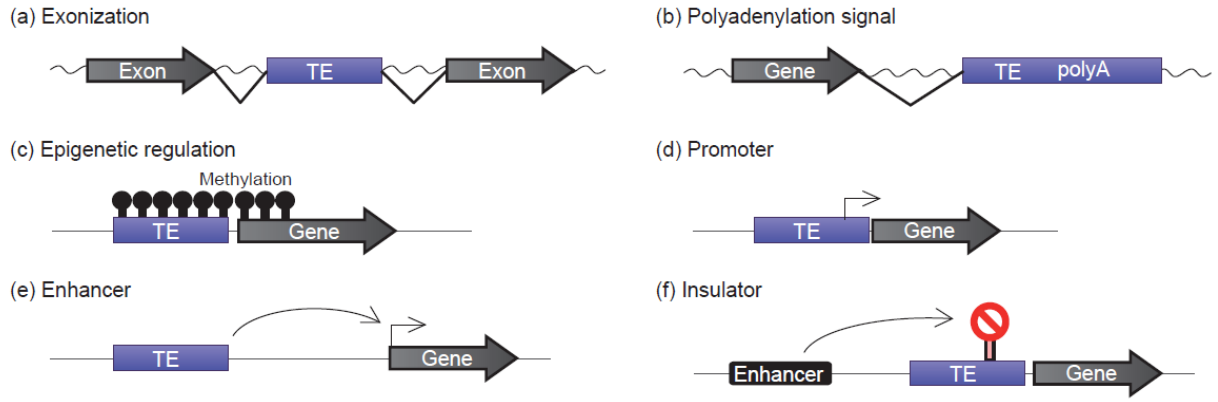


Figure 1. 7: The impact of TEs on gene expression (adapted from Nishihara 2019).
a) A TE can contribute to a protein coding sequence through exonization, either due to TE inserted directly into exons or by exonization of intronic TEs. b) A TE can provide a polyadenylation signal downstream of a gene, either truncating the gene if such insertion is present at the 3' end of the gene, or by elongating the gene transcript. c) The epigenetic control of TEs can spread onto nearby genes, resulting in gene expression changes. d) TEs can act as gene promoters when inserted upstream the canonical 5' promoter of the gene, or even when present within introns. e) TEs can also serve as enhancers, providing transcription factor binding sites to host genes either enhancing the gene expression or in some instances, functioning as

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

insulator elements (f). These illustrate the variety of mechanisms through which TEs can contribute to changes in transcriptomic landscapes.

In turn, these TE-induced genetic, epigenetic and transcriptomic changes may be reflected in the physiology and phenotype of the organism. With a special focus on ageing and longevity, we reviewed the implications of TEs on ageing, and, conversely, how ageing may influence TE activity (Merenciano et al. 2025). This work contributes to highlighting the role of TEs in generating variability, and how such variability may impact an organism's phenotype and, ultimately, its fitness. This review article was the result of a successful collaborative effort between one post-doctorate (Miriam Merenciano) and two other PhD students (Chloé Garambois and William Vilas Boas) under the supervision of Pr. Cristina Vieira. I have been involved every step of the way, with special emphasis on the "TEs and Their Association With Longevity in Long-Lived Organisms" section (Annexe 3). This work was featured on the Genome Biology and Evolution journal cover and highlighted by Andrade (Andrade 2025), a great achievement for all five of us.

Hence, TEs serve as a reservoir of cryptic genetic diversity which, through insertional polymorphism, contributes to epigenetic and transcriptomic variability, both across individuals and between populations.

c. Evolutionary/phenotypic capacitance from TEs

Evolutionary or phenotypic capacitance refers to the storing of genetic variation that has no immediate phenotypic effects as mentioned in the section II.4.e. on "Cryptic genetic variants". This concept encompasses the accumulation of CGVs within a genome, whose effects are typically neutral or silent under standard conditions, acting as standing genetic variation (Gibson and Reed 2008; Masel and Trotter 2010; Paaby and Rockman 2014; Díaz-Castillo 2017).

In a conceptual paper, Carlos Diaz-Castillo addresses the correlation between repeated DNA content and evolutionary capacitance (Díaz-Castillo 2017). He argues that repeated DNA, particularly in the form of TEs, can impact gene expression through its mere presence in the genome, without necessarily altering gene regulatory networks or inducing epigenetic changes through TE activity for instance. According to him, during gametogenesis, the limited pool of maternally provided heterochromatin-forming components can be "sunk" by "junk" DNA, leading to differential heterochromatin formation across the genome (Díaz-Castillo 2017; Murphy and Berger 2023). This process could result in variable epigenetic landscapes between gametes, eventually resulting in transcriptomic and phenotypic

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

heterogeneity amongst offspring. This heterogeneity could provide populations with greater resilience to environmental variation, potentially improving survival and dispersal to diverse habitats. On the other hand, genomes with lower repeated DNA content might lack such buffering effect and would rely more on standing genetic variation, making them more susceptible to selective pressures and promoting genetic diversification. To test this hypothesis, Diaz-Castillo used genome size as a proxy for junk DNA content within species of the Drosophilidae family (Díaz-Castillo 2017). He found a negative correlation between genome size and various proxies for genetic diversification, suggesting that repeated DNA may enhance phenotypic heterogeneity while reducing the need for genetic change (Díaz-Castillo 2017). Based on these findings, he proposes that "heterochromatic junk DNA could also act as genomic tuning knobs, and, [...], phenotypic heterogeneity-promoting capacitors" (Díaz-Castillo 2017). This hypothesis echoes the *buying-time* concept, whereby phenotypic plasticity can delay the need for genetic adaptation, and is also reminiscent of the idea that plasticity could hinder selection. It also connects with the concept of *diversifying bet-hedging strategies*, where the production of phenotypically diverse offspring increases the chances of some surviving under changing environmental conditions (Hopper 1999).

d. Evidence of TE-driven phenotypic variability

Evidence of TE-driven evolution and molecular diversity is now fairly abundant in the literature (Whitelaw and Martin 2001; Kidwell 2002; Brookfield 2004; Biémont and Vieira 2006; Casacuberta and González 2013; Mateo et al. 2014; Chuong et al. 2016; Van't Hof et al. 2016; Wei and Cao 2016; Broecker and Moelling 2019; Cranz-Mileva et al. 2024; Cabral-de-Mello and Palacios-Gimenez 2024; Huang et al. 2024; Xia et al. 2024). However, the impact of this TE-driven molecular diversity on organismal phenotypes remains elusive. In this section, we review how TE-induced epiallelic and transcriptomic variability can lead to phenotypic variation. Many of the mechanisms discussed in the section II.4 on the "Mechanisms underlying plasticity" also apply to TE-driven phenotypic variability, including epigenetic pathways that regulate TE activity and respond to environmental conditions. As a starting point, we will review the few existing studies that report TE-driven phenotypic variation (VP). We will then emphasize the sensitivity of TE impacts to environmental changes and explore their contribution to the environmental components of phenotypic variation (*i.e.* VE and VG×E), a largely unexplored area. Finally, we conclude with an example of genetic assimilation involving a TE responsible for the evolution of an environmentally induced phenotype. This section aims to illustrate how TEs may contribute to overall phenotypic variation, to environmentally responsive traits, and ultimately to the final stage of *plasticity-led evolution*: genetic assimilation.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

1. TE-driven phenotypic variation

One of the earliest examples of a TE-induced phenotypic variation is known as position effect variegation (PEV), first observed in *D. melanogaster* eye colour. Muller (1930) originally described PEV as the expression of a variegated phenotype later described to be caused by the spreading of heterochromatin into nearby euchromatic genes (Muller 1930; Lippman et al. 2004; Elgin and Reuter 2013). Later on, transgene insertions at various locations across the *D. melanogaster* genome revealed that the same suppressors and silencers responsible for mediating PEV were also involved in TE silencing (Slotkin and Martienssen 2007) (Figure 1.8). Indeed, it was shown that when a transgene was inserted near a TE, its expression could vary, resulting in PEV (Slotkin and Martienssen 2007) (Figure 1.8). Similar PEV observations were made later in mammals, with the presence of an IAP transposon upstream of the *Agouti* gene in the mouse, as well as in plants, with for example, the insertion of a MuLE element upstream the *Dfr-B* gene causing colour-streaked petals in the morning glory flower (Lander et al. 2001; Iida et al. 2004).

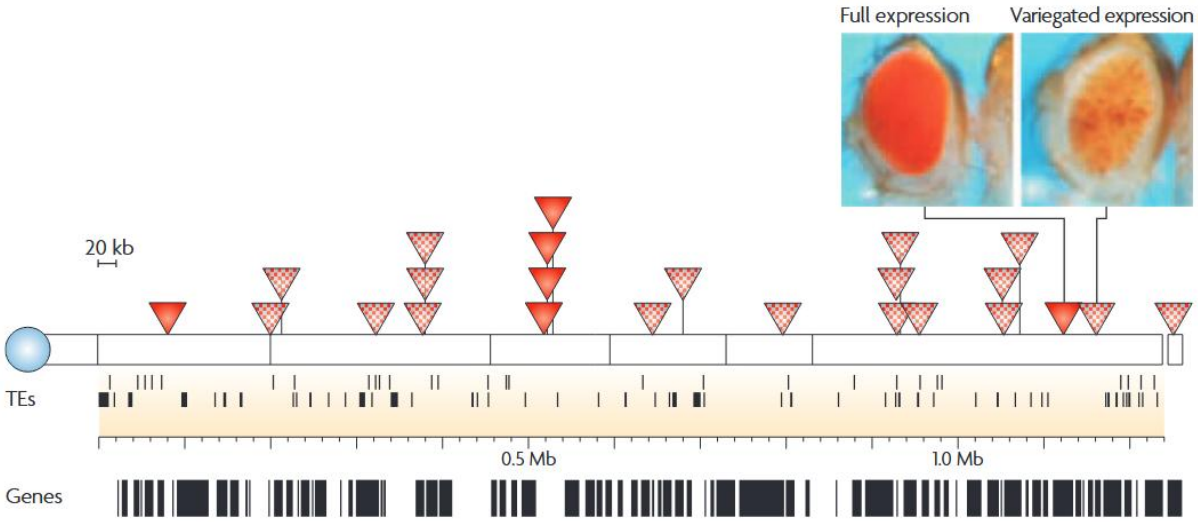


Figure 1. 8: Representation of the position effect variegation (PEV) phenotype in* D. melanogaster *(from Slotkin and Martiessen 2007).
Triangles represent the positions of transgene insertions, with the colour indicating the resulting phenotype (red: no PEV, chequered: PEV). The representative phenotypes are depicted with D. melanogaster *eye photos in the top right corner. The position of TE insertions and genes are shown below as black rectangles. The TE track is shaded in orange.*

Another very early mention of TE-induced increase in phenotypic variation was made by Mackay (Mackay 1986). In *P-element*-contaminated strains, the variance in traits related to fitness (measured as competitive fitness over multiple generations), viability (egg-to-adult survival), and fertility was observed

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

to be higher than in the control strain (Mackay 1986). By increasing trait variance even in genetically similar individuals, TEs may serve as hidden sources of phenotypic diversification, as discussed above. However, the strains were generated through reciprocal *Drosophila* crosses between a strain carrying approximately 50 *P-element* insertions and a strain completely devoid of functional *P-elements*. Although the crosses were designed to minimize variation in the genetic background, part of the observed phenotypic variation may still be attributed to residual genetic differences rather than solely to TE effects, which were not fully isolated.

Another example of TE-driven phenotypic variation comes from the comparison of *Capsella rubella* and its sister species *C. grandiflora* (Niu et al. 2019). Although *C. rubella* exhibits a reduced genetic variation due to a severe bottleneck event following speciation, it shows remarkable TE polymorphism compared to *C. grandiflora* (Guo et al. 2009). By analysing TE distribution in approximately 20,000 orthologous genes of 27 *C. rubella* accessions, the authors found a clear enrichment of TEs in promoter and downstream regions in *C. rubella* (Niu et al. 2019). Further examination also uncovered 3,808 polymorphic insertions between the accessions (Niu et al. 2019). Transcriptomic analysis of three of these accessions revealed that 4.2% of TE polymorphisms were associated with changes in the expression of nearby genes (Niu et al. 2019). These findings suggest that TEs can diversify gene expression, potentially shaping phenotypic variation in *C. rubella* (Niu et al. 2019). Notably, the study also identified a TE insertion hotspot near the *FLOWERING LOCUS C (FLC)* gene that could explain up to 12.5% of the observed variation in flowering time in natural populations (Niu et al. 2019). This result mostly highlights the potential for TE polymorphism at a specific locus to drive phenotypic divergence, and does not address the broader, transcriptome-wide impact of TEs on overall phenotypic variation.

Although adaptive contributions of TEs and TE-induced phenotypic variation have been sparsely reported across taxa as just seen in the previous paragraphs, establishing causation remains challenging. Hence, further investigation into TE-driven phenotypic variability and whether they may even provide organisms with plasticity remains to be fully explored. But first, for plasticity to be observed, environment-dependent responses must occur, a feature that TEs inherently possess. Although examples of TEs having conditional impacts on phenotypes are still relatively rare, the ones that exist are both compelling and insightful.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

2. *TEs are environment-responsive*

In this section, we report some of the ways in which the impact of TEs may be influenced by the environment. For one, molecular chaperones, like HSPs, have also been implicated in the stress-induced activation of TEs. For instance, Hsp70 and Hsp90 proteins are involved in the proper biogenesis of piRNAs, which are crucial for TE silencing (Karam et al. 2017). When *D. melanogaster* is subjected to heat shock, TE expression increases in both ovaries and testes (Cappucci et al. 2019). The authors suggest that under stressful conditions, HSPs may be diverted from their role in the piRNA pathway toward stabilizing other essential molecules, thereby ensuring survival of the offspring. As a consequence of this reduction in the piRNA-mediated TE silencing, TEs are activated, potentially increasing the genetic diversity of germ cells and, ultimately, the offspring (Cappucci et al. 2019). The authors speculate that this increase of genetic variation in the population may potentiate evolutionary plasticity and evolvability (Cappucci et al. 2019).

Not only are TEs responsive to the environment, but their impact on the host phenotype may also vary according to the environment. A notable example of stress-induced TE-driven phenotypic impact comes from a study by Merenciano and González (Merenciano and González 2023). In this work, the authors identified an insertion of a solo-LTR of a *roo* element upstream of the *Lime* TF gene (Merenciano and González 2023). The presence of this insertion introduces a functional alternative promoter to *Lime* and impacts the responses to both cold and immune stresses (Merenciano and González 2023) (Figure 1.9). For example, upon infection to *Pseudomonas entomophila,* flies carrying the insertions experience an increased survival, driven by a significant upregulation of the *Lime* gene, mediated through *roo*'s dual role as an enhancer and a TFBS (Merenciano and González 2023) (Figure 1.9). This study highlights how TEs' impact can act in an environmental-dependent manner.

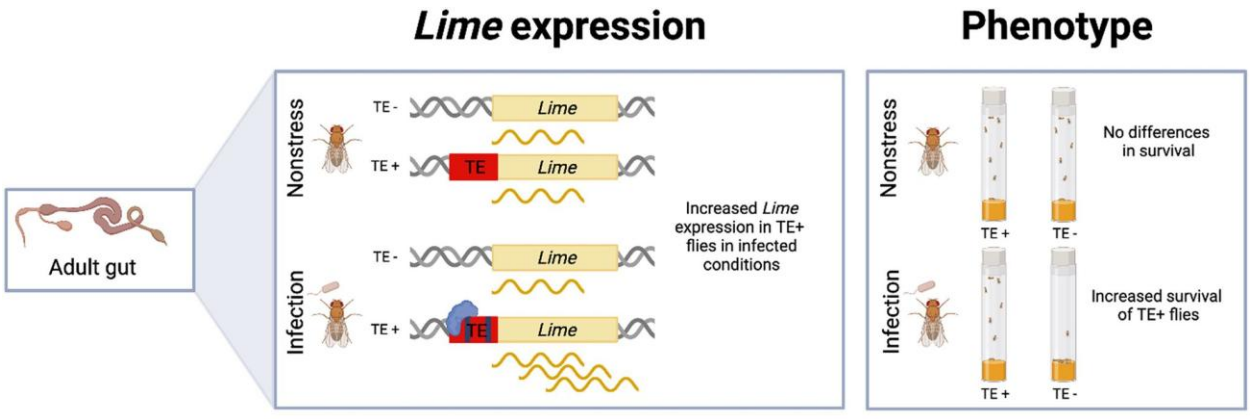
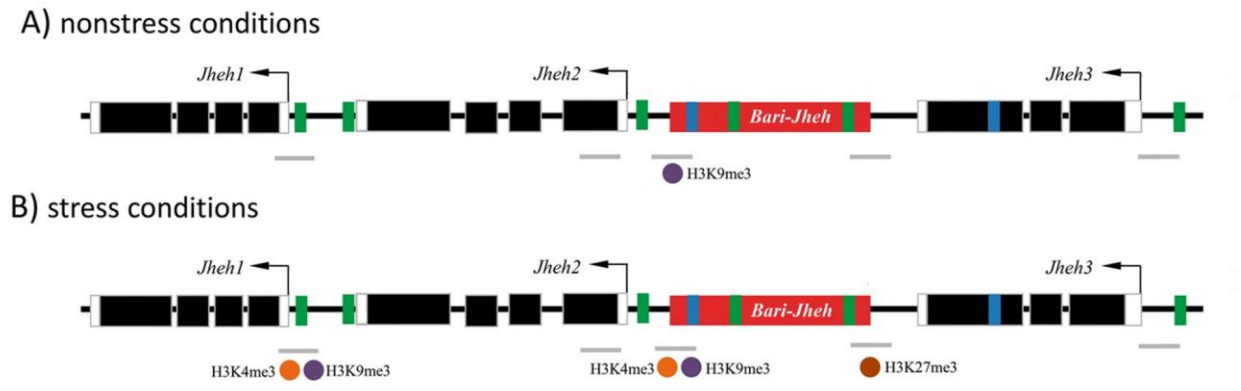


Figure 1. 9: Graphical summary of the effect of the* roo *solo-LTR in* D. melanogaster *upon infection with* Pseudomonas entomophila *(adapted from Merenciano and Gonzáles 2023).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

The insertion of a roo *solo-LTR upstream of the* Lime *gene alters its expression under infection* by P. entomophila*, conferring a survival advantage to* D. melanogaster. Lime *expression (yellow rectangle in the middle panel) was assayed in adult guts upon infection and compared to non-infected conditions. Upon infection,* Lime *becomes overexpressed due to the presence of the upstream TE insertion (red box in the middle panel), which introduces immune-related transcription factor binding sites (blue line within the TE). These sites enable transcription factors to bind (light blue shape in the middle panel) and initiate the transcription of* Lime*.* D. melanogaster *flies carrying the TE exhibit increased survival upon infection, as illustrated in the phenotype panel.*

Another example can be found in the change in chromatin state triggered by the *Bari-Jheh* transposon under oxidative stress. This insertion, a *Bari1* element, is known to mediate resistance to oxidative stress through its impact on the *Juvenile hormone epoxy hydrolase* (*Jheh*) genes, hence the name *Bari-Jheh* (González et al. 2008). Under oxidative stress conditions, this *Bari1* insertion alone induces the upregulation of *Jheh1* and *Jheh2* and the downregulation of *Jheh3* genes (González and Petrov 2009). A follow-up study suggested that *Bari-Jheh*'s action is mediated through differential enrichment of histone marks depending on the environmental condition (Guio et al. 2018) (Figure 1.10). Specifically, *Bari-Jheh* promotes heterochromatin formation in non-stress conditions, with an enrichment in the repressive H3K9me3 mark (Guio et al. 2018) (Figure 1.10). Under oxidative stress induced by paraquat exposure, however, it becomes associated with the permissive mark H3K4me3, leading to a differential expression of *Bari-Jheh* (Guio et al. 2018) (Figure 1.10). Additionally, the *Jheh1* gene, located near the original *Bari-Jheh*, shows enrichment for both H3K4me3 and H3K9me3 of its promoter region under stress, suggesting a spreading of histone post-translational modifications (Guio et al. 2018) (Figure 1.10). The presence of such antagonistic marks suggest that cells might respond differently to such stress, and a deeper analysis would be need to disentangle the effect of oxidative stress on these specific loci. Meanwhile, *Jheh3*, yet another nearby *Jheh* gene, experiences an increase in the repressive H3K27me3 mark in response to paraquat (Guio et al. 2018) (Figure 1.10). Altogether, these findings underscore how TEs can reshape the epigenetic landscape of nearby genes in an environment-dependent manner.



Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Figure 1. 10: Graphical summary of the histone post-translational modifications resulting from the presence of* Bari-Jheh *(adapted from Guio et al 2018).
A) Under non-stress conditions, H3K9me3 repressive marks silence Bari-Jheh*. B) Under oxidative stress induced by paraquat exposure, active H3K4me3 marks are detected at both* Bari-Jheh *and* Jheh1*, while repressive H3K27me3 marks are enriched near* Jheh3*. Panels A and B illustrate the structure of the* Bari-Jheh *and* Jheh1-3 *loci. Exons are shown as black boxes, and arrows indicate transcriptional start sites. The* Bari-Jheh *locus (red box) contains the TE and* Jheh *gene. Additional sites discussed in the original study but omitted here are shown in green and blue. Histone modifications are represented by coloured circles below their positions along the locus: orange for H3K4me3 and purple for H3K9me3.*

TEs, as drivers of genomic variation with environment-dependent effects, effectively tick all the boxes for being potent mechanisms promoting plasticity. In fact, their influence could be mediated by three effects already recognized as being mechanisms generating plastic responses. First, TEs can function as both *cis*- and *trans*-regulatory elements, thereby altering gene expression patterns (Chuong et al. 2017; Moschetti et al. 2020). Second, the mechanisms that regulate TE activity are also key players in plastic responses, and these regulatory marks, as we have seen, can spread from TEs to adjacent genes resulting in the emergence of epialleles (Slotkin and Martienssen 2007; Rey et al. 2016a; Marin et al. 2020). Lastly, because both regulatory effects and TE activity are sensitive to environmental cues, the genomic plasticity they induce is inherently environment-responsive (Capy et al. 2000; Casacuberta and González 2013; Piacentini et al. 2014; Horváth et al. 2017; Lanciano and Mirouze 2018; Pimpinelli and Piacentini 2020). As such, TEs emerge as remarkably relevant candidates as plasticity-inducing mechanisms. In a recent paper, Hunter (2024) puts forward the compelling idea that TEs may have evolved as "environmental stress detectors" (Hunter 2024). Described by McClintock (1956) as "particularly responsive to stresses and shocks", TEs have consistently been associated with genome responses to environmental challenges (McClintock 1956). These points were thoroughly explored by Pimpinelli and Piacentini (2020), where they proposed an extension of the concept of phenocopy (environmentally induced phenotypic variation) to encompass TE-induced phenotypic variation (Pimpinelli and Piacentini 2020).

3. The ultimate example: TE-driven plasticity-led evolution

Finally, to illustrate the evolutionary relevance of TEs and their putative role in plasticity, here is an example of TE-driven genetic assimilation tying the whole reasoning together.

In a very recent study, Sabarís and colleagues (2025) investigated the mechanism hidden behind the genetic assimilation of the ectopic veins (EV) phenotype in *D. melanogaster*, all following an experimental setup similar to that of Waddington (Figure 1.11) (Sabarís et al. 2025). The evolutionary experiment was performed on five lines: four inbred lines and a 5th mixed line (mix of the four inbred lines). Four experimental selection paths were conducted: (1) a control group with neither heatshock

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

(environmental induction) nor selection for the EV phenotype, (2) a group with selection for EV without induction (non-assimilated selection), (3) a group subjected to both heatshock and selection (induced selection), and (4) a group initially exposed to heatshock followed by selection for EV under normal conditions (assimilated selection) (Figure 1.11) (Sabarís et al. 2025). In the mixed population, naturally carrying significant genetic diversity, both non-assimilated and assimilated selection routes led to the total fixation of EV in the population (Sabarís et al. 2025). Functional analysis was conducted to investigate the underlying genetic mechanism explaining the observations (Sabarís et al. 2025). Candidate genes shared between the assimilated and non-assimilated lines were functionally tested using knockdown, overexpression, and CRISPR/Cas9 techniques, and the resulting EV phenotypes were quantified (Sabarís et al. 2025). Amongst the candidate genes, *Cad96Ca* appears to contribute significantly to the evolution of EV. Interestingly, an insertion of an *I-element* (a LINE element) was found in the 3'UTR of *Cad96Ca*, which they demonstrate was present in low frequency in the population prior to selection and reaches fixation by the end of the experiment (Sabarís et al. 2025). Rapid Amplification of cDNA Ends at the 3' end (3'RACE) was performed and results show that the *I-element* induces the production of different mRNA isoforms, likely tittering the expression of *Cad96Ca* and contributing in generating the EV (Sabarís et al. 2025).

This remarkable study first demonstrates that TEs can contribute to phenotypes in an environment-dependent manner, as previously shown. Secondly, it shows that these environmentally induced TE-driven phenotypes can undergo genetic assimilation. Altogether, the study provides compelling evidence of the evolutionary relevance of TEs, highlighting their potential to generate selectable phenotypic variation in response to environmental cues.

In this section, I have presented evidence supporting TEs as potential contributors to phenotypic plasticity. This includes their contribution to overall phenotypic variability, their inducibility by the environment, and their possible implications in genetic assimilation. Building upon these observations, we now seek to investigate whether during the process leading to fixation, TEs may influence an organism's response to environmental change, not only through the effects of specific insertions but also *via* the cumulative impact of the overall TE content. This has virtually never been done before, at least in animals, mainly because it is extremely challenging to disentangle TE effects from those of other genetic variants and mechanisms. To investigate the contribution of TEs to an organism's plasticity, one needs to rule out all other variants that could lead to confounding effects.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

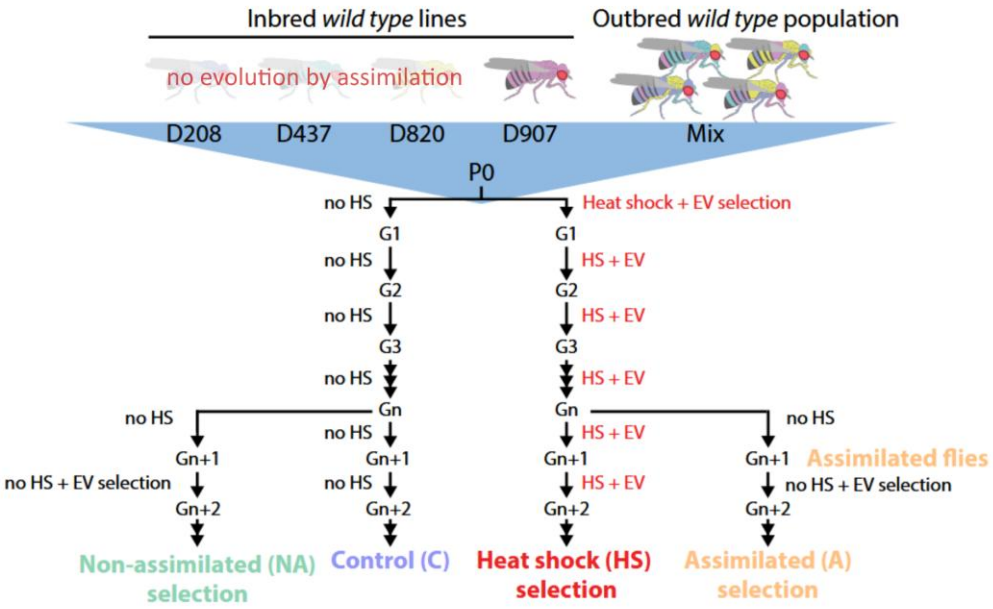


Figure 1. 11: Experimental design following Waddington's genetic assimilation evolutionary experiment protocol for the ectopic vein phenotype (adapted from Sabarís et al 2025).
Four inbred lines (D208, D437, D820, and D907) were used, three of which did not reach genetic assimilation. A mixed population, representing an outbred wild type background, was generated by genetic admixture of the four inbred lines. P0 refers to the ancestral lines prior to experimental evolution. From G1 to G3, all lines underwent either no heat shock, or heatshock exposure followed by EV selection. After this initial phase, lines were separated into four distinct experimental evolution routes: (i) the control group, evolved without heat-shock induction or EV selection; (ii) the non-assimilated group, subjected only to EV selection without induction; (iii) the heatshock-selection group, subjected to both heatshock and EV selection; and (iv) the assimilated-selection group, initially exposed to heatshock and EV selection, followed by selection for EV under control conditions. A: assimilate, NA: non-assimilated, C: control, HS: heatshock, EV: ectopic vein, Gn: number of generations.

In the following section, I will introduce the model organism and the particular system used in this project to bypass that limitation. As a final note, I would like to clarify that I do not suggest that TEs are almighty molecular actors. My supervisors, collaborators, and I simply want to highlight the possibility of their contribution to yet another aspect of organismal biology, namely plasticity. We have not investigated whether any potentially TE-induced plasticity is adaptive, maladaptive, or neutral. Our aim is simply to offer empirical evidence that could contribute to the emergence of a new view within the mechanistic study of phenotypic plasticity.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

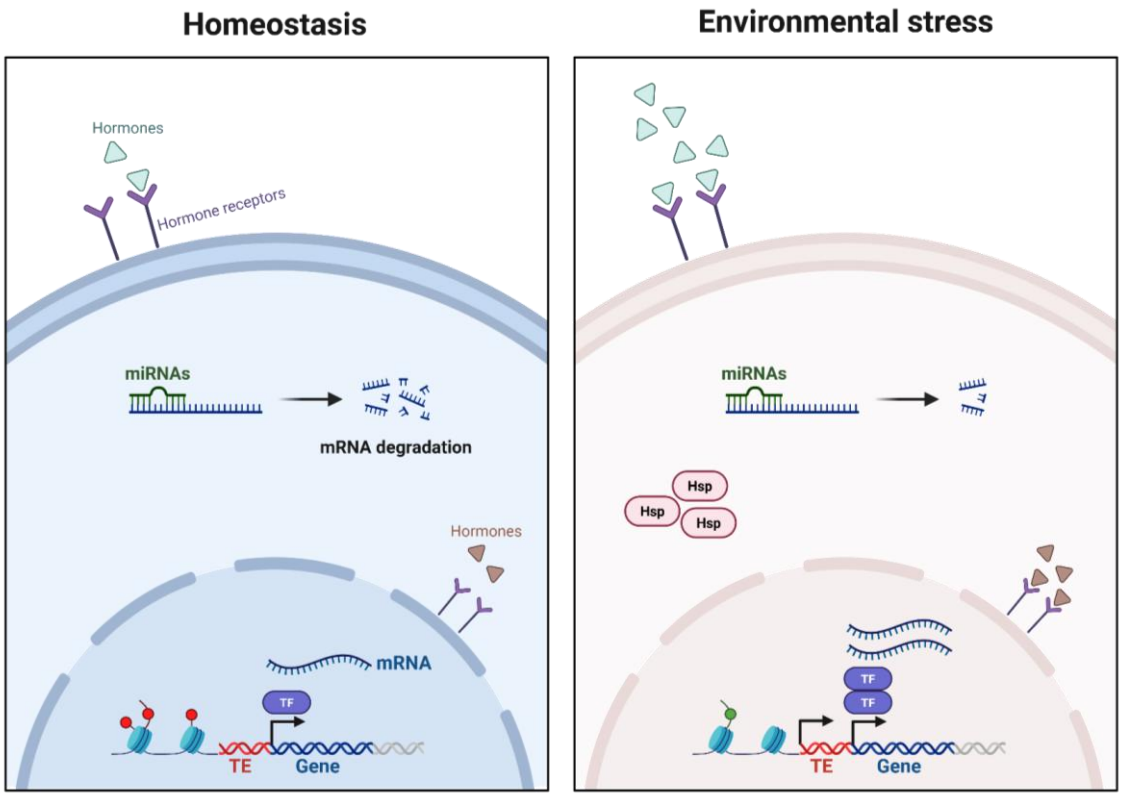


Figure 1. 12: Schematic representation of the molecular mechanisms involved in promoting transcriptional plasticity thereby underlying phenotypic plasticity.
Upon stress, a number of molecular mechanisms may be disturbed. Hormone levels may increase and membrane and nuclear receptors may be more solicited, resulting in a change in cellular response. miRNA expression can fluctuate with the environment, leading to differential expression of protein-coding genes. Epigenetic mechanisms are labile to the environment, potentially leading to gene expression changes. Histone tail modifications are depicted as red and green dots on the schematic histones, and represent the change in chromatin accessibility. Heatshock protein expression can be triggered by environmental stimuli and can initiate a cascade of molecular responses, including inducing the upregulation of transcription factors. Finally, TE expression and their regulatory mechanisms may also be modified by the environment and cause the modulation in the expression of genes in their vicinity.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

IV. Plastic flies with variable TE content: a mean to an end

D. melanogaster has been a model organism for studying TEs for many decades, and with reason. First, TEs in *D. melanogaster* remain relatively dynamic, with potentially full-length, active copies (Biémont et al. 1994; Vieira et al. 2012; Mérel et al. 2020). This contributes to a high level of insertional polymorphism, to the point where even clustering individuals by geographic location or climate zone becomes challenging (Barrón et al. 2014; Rech et al. 2022a). Interestingly, in *D. melanogaster*, the transposition rate (8.99×10^{-5} copy^{-1} gen^{-1}) is remarkably higher than the point mutation rate (1.67×10^{-9} site^{-1} gen^{-1}), positioning transposition as a substantial driver of genetic diversity (Y. Wang et al. 2023). Additionally, *D. melanogaster* is a remarkable study system for investigating phenotypic plasticity (Shingleton and Tang 2012; Zhou et al. 2012; Morgante et al. 2015; Huang et al. 2020; Flatt 2020). Finally, both the genome and the TEs of this species are well annotated, making sequencing analyses more reliable.

So, in order to investigate TE's contribution to phenotypic plasticity, we need to free ourselves from any confounding effects and to isolate the variable we want to test, *i.e.* TEs. Fortunately, thanks to our collaborator Séverine Chambeyron based at the Institute of Human Genetics, we have access to a powerful and well-suited tool that allows us to carry out controlled plasticity assays. Indeed, a *D. melanogaster* strain was previously genetically modified to suppress the piRNA pathway in charge of TE silencing in the somatic follicular cells (Barckmann et al. 2018). Since a constitutive piRNA knockdown is lethal (Cox et al. 1998), a system was engineered with a piRNA knockdown inducible by a change in temperature. The fly line carries three components: (1) an ubiquitously expressed thermo-sensitive GAL4-inhibitor, GAL80ts, (2) a GAL4 UAS-activator driven by the follicle cell-specific traffic jam (tj)-promoter (tj-GAL4) and (3) a UAS-short-hairpin(sh)-piwi, that induces Piwi RNAi (Figure 1.13) (Barckmann et al. 2018). Since some TE families are capable to cross the barrier from the follicular cells to the germline, namely ERVs, TEs can accumulate from one generation to the next, such that generation after generation, this induced piRNA depletion allows for the establishment of populations with varying TE content in a constant genetic background (G0) (Barckmann et al. 2018). In total, five populations harbouring different TE content and insertions sites were generated (G0, S10, S31, S73, S100) (Figure 1.14). G0 is the parental population, and all subsequent shifts are named SX, where X is the number of generations the piRNA has been knocked down. Hence, S10 corresponds to 10 generations of knockdowns and therefore of TE accumulation (Figure 1.14).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

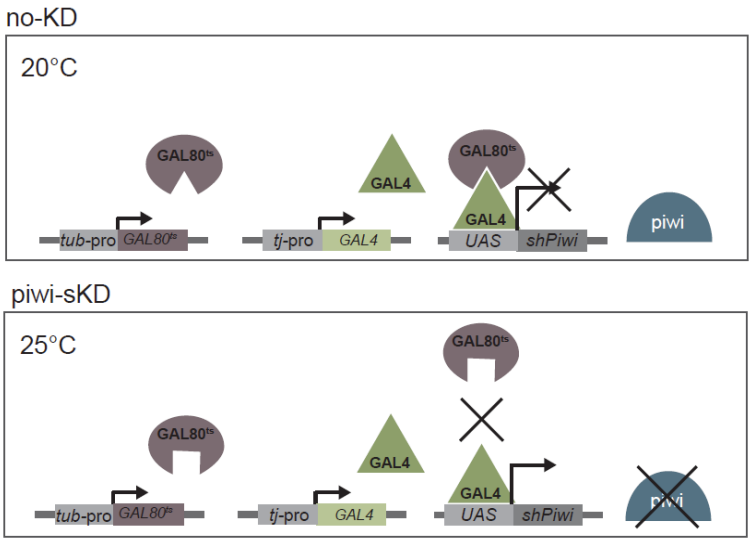


Figure 1. 13**: **Schematic representation of the genetic modification allowing the accumulation of TEs across generations in* D. melanogaster *(from Barckmann et al. 2018).
At 20°C the thermo-sensitive GAL80ts is in the right conformation to sequester GAL4 resulting in the inhibition of the expression of a short hairpin directed at Piwi (shPiwi), hence the somatic piRNA is functional in follicular cells. At 25°C, the thermo-sensitive GAL80ts changes conformation and is no longer capable of binding GAL4, the shPiwi is thus expressed inhibiting Piwi and in consequence disrupting the somatic piRNA pathway. sKD: piwi somatic knockdown, shPiwi: short hairpin Piwi, tub: tubulin, tj: traffic jam, UAS: upstream activator sequence.

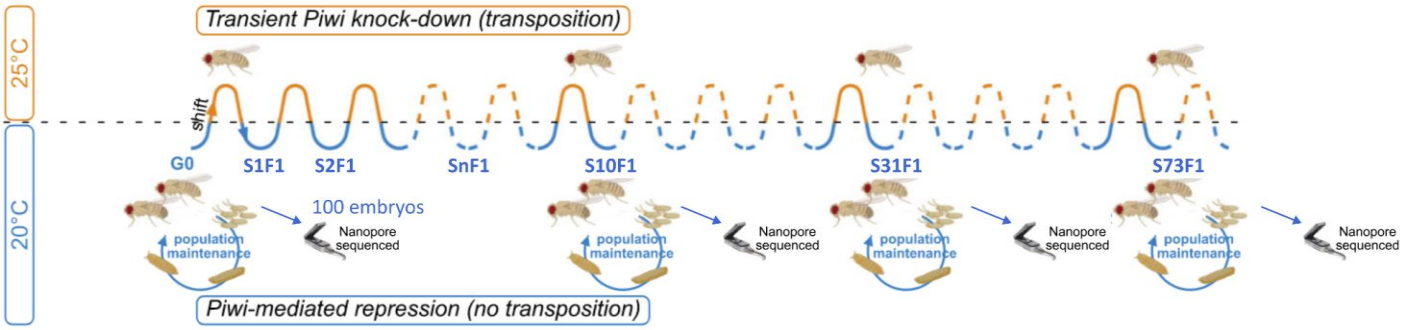


Figure 1. 14: Schematic of the experimental setup to establish the fly populations with increased TE content (adapted from Varoqui et al. 2025).
Successive changes in rearing temperature were applied across generations to induce a transient Piwi knockdown, leading to TE derepression and accumulation. During the induction phase (for five days at 25°C, represented by the orange lines and denoted as S for shift followed by the generation number), populations underwent progressive TE accumulation. Following induction, selected populations were maintained under constant conditions at 20°C (blue lines) to prevent further TE accumulation, thereby "freezing" their TE content. F1 corresponds to the first generation following the last shift, from which populations were either maintained at 20°C (life cycle represented by the circular arrows) or subjected to additional shifts for continued TE accumulation. For reference, the study includes the S100 population, in which TE accumulation was extended over 100 generations. The G0 population, continuously reared at 20 °C, served as the control.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

It is essential to highlight that, thanks to these genetically modified flies, we are in a unique position where the populations share a common genetic background (the parental G0), meaning that genetic variation (VG) is nearly absent, while the variation ensuing from the TE content and composition differs (VTE) (S10, S31, S73, and S100). This experimental setup provides a unique opportunity to evaluate the contribution of overall TE content to the environmentally induced phenotypic changes, *i.e.*, plasticity.

In this context, we can establish the following framework:

$$VP = VG + VE + VG \times E + V\varepsilon, \text{ with } VG = VTE + VI$$

Where VTE refers to the genetic variance caused by the sum of the effect of the insertions at different loci, and VI to the interaction between insertions at different loci as well as their interaction with the surrounding genes.

The novelty of this study lies in the assessment of plasticity which was made possible by comparing the different populations in variable environments. However, in doing so, we also observed that TEs, independently of environmental changes, contribute to increased phenotypic variability. As previously discussed, few studies have previously highlighted this aspect. Therefore, while our primary focus is on plasticity (VE and VGxE), we will also discuss the relationship between increased TE content and the overall rise in phenotypic variability (VP), acknowledging TEs as a genetic component contributing to this variation.

Thank you for making it to the end of this introduction. I hope it has brought you as much clarity as it has brought me. Trying to bridge these two very distinct biology topics has not been an easy task but, then again, easy is rarely exciting! I truly hope I have made a compelling case for this study and have succeeded in convincing you of its relevance.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Chapter 1. On the trail of Transposable Elements: exploring their impact on phenotypes, phenotypic variability, and plasticity in *Drosophila melanogaster*

Foreword

This chapter may appear as a single unified section, but it reflects three years of work and includes material that could have been divided into multiple chapters. We have taken the stand to tell the story as a whole, to clearly connect the analysis of TE content with its impact on traits and phenotypic plasticity. We acknowledge that this is a dense piece of work! Moreover, all the phenotypes presented in this chapter are used as measures of differentiation of our populations. Each one could definitely be discussed in far greater biological detail, and I'm sure that specialists in thermotolerance, fecundity, or longevity might find our analysis is somewhat shallow, and I'll give them that. However, the goal of this project was not to fully dissect each trait individually. Rather, it was to use them cumulatively to investigate a broader question: the contribution of TEs to phenotypic plasticity. And surprisingly, maybe even unexpectedly, that task has been successfully initiated!

I. Introduction

Transposable elements (TEs) are ubiquitous repetitive DNA sequences capable of mobilizing within most eukaryotic genomes (Wells and Feschotte 2020). Although TEs are most often neutral, they also have deleterious effects. TEs may compromise genome integrity and are inherently mutagenic, occasionally being the cause of disease (Burns 2020). More rarely, TEs have contributed to the emergence of adaptive phenotypes, as illustrated by several well-documented cases (Rey et al. 2016a; Platt et al. 2018; Schrader and Schmitz 2019; Catlin and Josephs 2022b; Galbraith and Hayward 2023). For example, the industrial hypermelanism of the peppered moth (*Biston betularia*) is linked to a TE insertion in the first intron of the *cortex* gene that induces its overexpression (Van't Hof et al. 2016). Another classical example has been reported in *Drosophila melanogaster* where an *Accord* element inserted upstream of the detoxification cytochrome P450 gene *Cyp6g1* acts as an enhancer, increasing resistance to various insecticides (Chung et al. 2007). The potential of TEs to drive adaptation and generate phenotypic changes stems from two key properties. First, TEs contribute to genetic diversity by carrying regulatory elements capable of influencing the expression of nearby genes, acting as promoters or enhancers, or supplying transcription factor binding sites (TBFS) (Chuong et al. 2017; Dubin et al. 2018; Drongitis et al. 2019; Moschetti et al. 2020; Sundaram and Wysocka 2020; Gebrie 2023). TE insertions can also generate new polyadenylation signals, introduce premature stop codons, or give rise to chimeric transcripts (Oliveira et

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

al. 2023; Huang and Lee 2024). Secondly, TEs also contribute to epigenetic variation in DNA methylation, histone tail modifications, and small RNAs (sRNAs) (Rebollo et al. 2012; Rey et al. 2016a; Fueyo et al. 2022). These epigenetic mechanisms not only regulate TE activity but can also impact the expression of nearby genes (Jaenisch and Bird 2003; Schrey et al. 2016). Notably, these regulatory mechanisms can be influenced by environmental changes (Jaenisch and Bird 2003; Feil and Fraga 2012; Rey et al. 2016b; Murray et al. 2022). For example, in *D. melanogaster*, oxidative stress induces an upregulation of the *Bari-Jheh* locus, consisting of a *Bari1* element inserted upstream of the *Juvenile hormone epoxy hydrolase* (*Jheh*) genes. This upregulation is associated with an enrichment of the permissive H3K4me3 mark, and increases the survival of the flies. Under control conditions, however, *Bari-Jheh* is associated with the repressive H3K9me3 mark (Guio and González 2019; Guio et al. 2018).

Environmental changes can modulate TEs' activity and epigenetic control, which may influence the response of the organism to environmental variation. As a result, even individuals with identical genomes may exhibit distinct TE effects. Such environment-induced changes are the definition of the concept of phenotypic plasticity, wherein the same genotype can yield distinct phenotypes under different environmental conditions (West-Eberhard 2003). The assessment of plasticity is commonly performed by partitioning the total phenotypic variance (VP) into its components: 1. the variance induced by the genotype (VG), 2. the one attributed to the environment (VE) (*i.e.* plasticity) and 3. the genetic variation for plasticity which is the genome-by-environment interaction (G×E) (Falconer 1960; W. Hill and Mackay 2004). GxE represents the condition under which plasticity becomes subject to natural selection and can drive evolutionary changes (Pfennig 2021a). Yet, the role TEs play in generating GxE remains unexplored.

To address this challenge, we first hypothesized that the overall TE composition, and its variation across populations sharing an otherwise nearly identical genetic background, could contribute to differences in phenotypes and in the amount of phenotypic variability. Secondly, we hypothesized that, upon exposure to environmental changes, differences in TE content and position, may lead to different plasticity responses. Testing these hypotheses requires a specific experimental design in which sources of genetic variation, other than TE insertions, are reduced. This approach maximizes the exclusion of confounding factors and enables the estimation of the contribution of TEs. We used a genetically modified isogenic *D. melanogaster* line (G0), from which four populations were derived (S10, S31, S73, and S100). Each of the populations is the result of a varying number of generations of inducible *Piwi* knockdown (Barckmann et al. 2018; Varoqui et al. 2025). Indeed, the *Piwi* protein plays a central role in the piRNA pathway that regulates TE activity in *Drosophila*, and its successive knockdown (KD) resulted in TE

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

accumulation, all within an otherwise constant genetic background (Barckmann et al. 2018; Varoqui et al. 2025). Within this framework, S10 corresponds to ten successive generations of inducible *piwi* KD in a population of around 500 individuals, resulting in polymorphic TE accumulation, while S31 corresponds to 31 generations, and so on. This provided us with a controlled experimental framework to evaluate the impact of TEs on phenotypes and plastic responses. Under these conditions, the partitioning of the variance can be described as: VP = VTE + VE + VTE×E, where VTE is the variation induced by TEs.

The phenotypic screening of the five populations revealed significant differences, including a notable decrease in fertility in the S73 population. This reduction was, however, seemingly accompanied by a general increase in phenotypic variability, suggesting that a greater insertional polymorphism results in dispersion of the phenotypes around the population mean. Additionally, plasticity assays supported the contribution of TEs to genotype-by-environment (G×E) plasticity, or in our case TE×E plasticity. Finally, while overall TE content contributes to phenotypic differences between populations, we also identified candidate insertions through a genome-wide scan approach. To our knowledge, this is the first study to provide evidence for the contribution of TEs to phenotypic plasticity under such a tightly controlled experimental framework in insects.

II. Material and Methods

1. Fly husbandry

A *D. melanogaster* line was previously modified to suppress the piRNA pathway in charge of TE regulation (Barckmann et al. 2018). Since a constitutive piRNA knock-down is lethal, a system was engineered to target ovarian somatic cells with a heat-induced piRNA knock-down. Briefly, the fly line carries three components: (i) a GAL4 UAS-activator driven by the follicle cell-specific traffic jam (tj)-promoter (tj-GAL4), (ii) a UAS-short-hairpin(sh)-piwi that induces *Piwi* RNAi and (iii) an ubiquitously expressed thermo-sensitive GAL4-inhibitor, GAL80ts (Barckmann et al. 2018). The initial isogenic population (G0) was repeatedly exposed to temperature changes while maintaining a population size of around 500 progenitors in standard medium-filled bottles, to reduce genetic drift. The consequence is the occurrence of independent transposition events in each individual of the offspring, leading to insertion polymorphism in a constant genetic background (Barckmann et al. 2018). In total, five populations harbouring different TE content and insertions were generated (G0, S10, S31, S73, S100). G0 is the parent population, and all subsequent shifts are named SX, where X is the number of generations that have been shifted. Hence, S10 corresponds to 10 generations of shifts (10 generations of piRNA knock-down) and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

therefore of TE accumulation. Gifted flies were all infected by *Wolbachia.* Flies were maintained at 18°C in 12-hour light cycles, with 65% relative humidity, in vials with standard medium (yeast, sugar, wheat flour and nipagin as a preservative), with a density of approximately 100 progenitors at each generation. Each population was maintained in five vials per generation, which were frequently shuffled to avoid vial effects.

2. DNA extraction and sequencing

DNA extractions were performed on 200 embryos per population using a homemade protocol designed to yield very long DNA fragments. Briefly, tissues are homogenized in STE buffer (chloride-Tris-EDTA), to which is added 10% SDS (sodium dodecyl sulfate). The mix is incubated for 2 h 30 min with proteinase K at 55°C. The samples are then treated with RNase for 1 h at 37°C. The DNA is then precipitated with phenol/chloroform/isoamyl alcohol. The aqueous phase is recovered, and the DNA is purified by isopropanol and ethanol washing steps. The DNA is eluted in 35 µL of elution buffer and is stored at 4°C until library construction and sequencing, performed at the PSI in house sequencing platform at the Institut de Génomique Fonctionnelle de Lyon (IGFL), on a PromethION 2 Solo using Oxford Nanopore Sequencing Technology. The data received was filtered for reads with quality above Q10.

3. Genome assembly

Raw reads were analysed to infer coverage and general statistics using the following tools: Nanoplot (v 1.42.0) (De Coster and Rademakers 2023), Cramino (v 0.14.5) (De Coster and Rademakers 2023), and Qualimap (v 2.3) (Okonechnikov et al. 2016). G0 was assembled from an average coverage nearing 50 X. The genomic assembly was performed with Flye (v 2.9.3-b1797) using standard parameters except for the --nano-raw option (Kolmogorov et al. 2020), followed by four rounds of Racon (v 1.5.0) with minimap2 (v 2.29-r1283) conducted using the parameters for ONT data (Vaser et al. 2017; Li 2021). A step of purging was performed with Purge Haplotigs (v1.1.2) and Samtools (v 1.6) (Roach et al. 2018; Danecek et al. 2021). Finally, scaffolding was performed using RagTag (v 2.1.0) (Alonge et al. 2022) with default parameters using the dm6 version of *D. melanogaster* as the reference genome (https://hgdownload.soe.ucsc.edu/goldenPath/dm6/bigZips/). The Benchmarking Universal Single-Copy Orthologs (BUSCO, v 5.7.1) analysis was performed on the assembly using the Arthropoda reference gene set (Manni et al. 2021) (Table S2.1). The Quast pipeline (v 5.3.0) was also ran providing assembly parameters to evaluate assembly quality (Mikheenko et al. 2018) (Table S2.1). The G0 assembly was

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

annotated using Liftoff (v 1.6.3) with Dm6 annotation as reference (Shumate and Salzberg 2021). The resulting G0 GFF files was extracted for 3'UTR, 5'UTR, exon, intron, promoter (2000 kb) and genes using Gencode Regions for subsequent analysis (https://github.com/saketkc/gencode_regions). RepeatMasker (v 4.1.5) was run on the G0 assembly using the *D. melanogaster* Manually Curated TE library (MCTE) previously published (Smit et al. 2021; Rech et al. 2022b).

4. Transposable element analysis

Rarefaction curves were generated to assess TE detection power using a logarithmic prediction model using the built-in Stat package R 4.4.0 (Bolar 2019; R Core Team 2021). TrEMOLO was first run on subsets of reads, with coverage ranging from 5x to the maximum coverage available for each population with 10x increments (Mohamed et al. 2023). The number of TE insertions detected at each coverage level was used to model the expected number of insertions using a logarithmic function, allowing prediction of TE discovery saturation as coverage approaches the sequencing of all haplotypes. Downstream analyses were computed on 35 X coverage subsamples of the raw reads for all populations generated using Rasusa (v 2.0.0) (Hall 2022). TE variants and frequencies were obtained with TrEMOLO using default parameters of the OUTSIDER module and G0 as reference genome and the *D. melanogaster* MCTE TE library previously published (Rech et al. 2022b; Mohamed et al. 2023). Using the logarithmic extrapolation of the TrEMOLO output, we estimate that at 35X coverage, approximately 50% of TE variants are detected in each of the five populations (S10=49.74%, S31=53.238%, S73=52.38%, and S100 = 51.60%) (Figure S2.1). TE insertional variants were filtered from the "TE_INFOS.bed", and no_name contigs were removed for downstream analysis. Differences in average TE insertion frequency among populations were tested using the Kruskal-Wallis test. A shared insertion matrix was generated using a custom script, based on 100 bp windows. Cases of ambiguous sharing were manually curated: insertions were merged if the same TE was present in more than one population, or marked as unresolved if no consistent pattern could be identified. Additionally, when multiple insertions from the same population were found at the same position, the same criteria were applied, and their frequencies were summed. Plotting of the insertions per class and families was performed in R 4.4.0 (R Core Team 2021). The insertions were categorized into the following frequency classes: <20%,]20-40%],]40-60%],]60-80%], and >80%. These categories, along with the full set of insertions, were subjected to permutation tests based on 1,000 simulations of insertion distributions using *bedtools shuffle* (v2.30.0) (Quinlan and Hall 2010). The simulated distributions were then compared in a two-sided test with the observed number of insertions across the genomic compartments previously obtained from the G0 GTF file. DeviaTE (v2.2.1) was run on the subsampled data

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

using default settings and three normalization single-copy genes: *Act5*, *Piwi*, and *Rpl32* (Weilguny and Kofler, 2019). The software aligns long reads to the TE library and estimates the TE insertion number for each TE family per haploid genome. The sum of high-quality (>Q15) coverage estimates was used as a proxy for TE load in each population.

5. Viability and development time

To measure viability, 150 three-day-old female-male couples per population were placed in a small Dutscher egg-laying chamber (sexed and counted under CO2 anaesthesia). Flies were allowed to lay eggs for half a day. From these egg laying chambers, five replicates of 40 eggs were collected on small regular fly-food caps positioned in vials containing the standard media (Figure S2.2A). Eggs were collected under binoculars using the tip of a scalpel to scoop the eggs out of the media. Hatched flies were counted daily, from the first emergence to the last. Viability was computed as the proportion of adults emerging from the initial number of eggs (40). Data was analysed under R software (v 4.4.0) using generalized linear mixed-effects model (GLMM) with binomial distribution using the Lme4 package (v 1.1-36) (R Core Team 2021; Bates et al. 2025). ANOVA Chi square test was applied to assess the fixed effect. McFadden's R^2 was estimated with the package Jtools (v 2.3.0) (Long 2024). From these data development time (development time) was also assessed, for each egg counted for its viability was also recorded the time until emergence of the adults. Using this data, a cumulative percentage of flies emerged each day since oviposition was calculated from the total number of eggs tested per replicate (five replicates of 40 eggs per conditions). The comparative statistical analysis of the emergence curves between populations was performed using a Log Rank test in R (v 4.4.0) (R Core Team 2021). Development time were also analysed using ANOVA on the time at which each fly has emerged in R 4.4.0 (Stat built-in package) (Bolar 2019; R Core Team 2021). Partial Eta^2 was calculated using package Effectsize (v 1.0.1) (Ben Shachar et al. 2020).

6. Embryonic hatchability

To measure hatchability, 150 three-day-old female-male couples per population were placed in a small Dutscher egg-laying chamber. Flies were allowed to lay eggs for half a day. Eggs were collected using a fine brush and water, and filtered through a fine mesh. Eggs remained dipped in water while five replicates of 50 eggs were collected and aligned on a black piece of Canson 160 g/m² paper (Figure S2.2B). Two days after sampling, the number of emerged eggs were counted under binoculars. Data were analysed under R software (v 4.4.0) using generalized linear mixed-effects model (GLMM) with binomial

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

distribution using the Lme4 package (v 1.1-36) (R Core Team 2021; Bates et al. 2025). ANOVA Chi square test was applied to assess the fixed effect. McFadden's R^2 was estimated with the package Jtools (v 2.3.0) (Long 2024).

7. Dry weight

To conduct dry weight measurements, density-controlled vials were prepared. Specifically, 150 three-day-old female-male pairs per population were placed in a small Dutscher egg-laying chamber, with flies sexed and counted under CO_2 anaesthesia. Flies were allowed to lay eggs for half a day. From each chamber, five replicates of 40 eggs were collected and transferred into vials containing standard fly medium. Emerging adults were aged for five days, then sexed and pooled into groups of five. The flies were frozen at -20 °C overnight and subsequently dried in an oven at 75 °C for 24 hours. The dry weight data were analysed using ANOVA with R version 4.4.0 (Stat built-in package) (Bolar 2019; R Core Team 2021). Partial Eta^2 was calculated using package Effectsize (v 1.0.1) (Ben Shachar et al. 2020).

8. Negative geotaxis assays

A pool of ten 3-to-5-days old flies was placed in pre-washed vials (2 cm in diameter, 9.5 cm in length), which had been dried and left to air out to eliminate any residual plastic odour. The vial was divided into three predefined zones: the first zone extended from the bottom to a height of 2 cm, the second from 2 cm above the bottom to 2 cm below the top, and the third from 2 cm below the top to the top of the vial. Prior to the trial, flies were placed in the tubes using a mouth aspirator and allowed to acclimate for 30 minutes. The flies were gently tapped three times to ensure they all fell to the bottom. After a two-minute trial, the number of flies in each zone was recorded. This procedure was repeated in batches of three experimental sessions, with each population and sex represented in every batch, resulting in six biological replicates (totalling 60 flies per condition). Taking the count of flies reaching the top of the tube as success, and the ones in the rest of the tubes as failure (middle and bottom zones), the statistical analysis was conducted in R (v 4.4.0) using a generalized linear mixed-effects model (GLMM) with a binomial distribution (R Core Team 2021; Bates et al. 2025). ANOVA was applied to assess fixed effects, using a Chi-square test. McFadden's R^2 was estimated with the package Jtools (v 2.3.0) (Long 2024).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

9. Thermal tolerance

Critical thermal maximum (CT_{max}) was assessed for females under controlled temperature ramping in a water bath, increasing at a rate of 1°C per minute from 20°C until all flies entered a coma, typically around 40°C. A precision thermometer continuously recorded the temperature to ensure consistency across trials. Data were collected from six pools of five female flies each, totalling 30 individuals per condition. Statistical analysis was performed in R 4.4.0 using a linear mixed-effects model to account for variation between trials and arenas (R Core Team 2021; Christensen [2018] 2025). Post-hoc pairwise comparisons of CT_{max} across populations were conducted using estimated marginal means (EMMs), with Tukey adjustment for multiple comparisons. Partial Eta^2 was calculated using package Effectsize (v 1.0.1) (Ben Shachar et al. 2020).

10. Longevity

Ten replicates of pools of around 10 virgin females or males were placed in vials containing standard fly food, for each of the five populations (G0, S10, S31, S73, S100) (Table S2.2). All replicates for all populations were carried within a week and therefore stem from the same generation of flies. In addition, flies were placed on the same batch of medium. Virgin flies were collected in the afternoon, after vials were emptied in the morning, and are sexed under CO2 anaesthesia. Vials were changed once every week to prevent low-quality media from impacting the experimental outcome. The cumulative number of dead flies was counted every 24 hours on weekdays. Survival curves and Log Rank test pairwise comparisons were generated with the Survminer (v 0.5.0) and Survival (v 3.5-8) packages in R 4.4.0 (Kassambara et al. 2021; Therneau et al. 2024). This experiment was conducted at 18° C and 22°C. The time for 80% of the population to have died (LT80) was estimated using the Ecotox (v 1.4.4) package in R 4.4.0 (Wheeler et al. 2006; R Core Team 2021). The slopes were calculated from a linear model of LT80 vs. temperature. Analysis of the survival chances was performed using generalized linear model (GLMER) and probit distribution with replicates as random effects. Type III ANOVA was applied to assess the fixed effects, and EMMs were adjusted using the Bonferroni adjustment method to account for multiple comparisons. Additionally, a linear comparative analysis of longevity at 18°C *vs.* 22°C was performed using a linear mixed-effects model with the lmerTest (v 3.1-3) package (Christensen [2018] 2025). Type III ANOVA was also applied to assess the fixed effects, and estimated marginal means (EMMs) were adjusted using the Bonferroni adjustment method to account for multiple comparisons. Partial Eta^2 were calculated using package Effectsize (v 1.0.1) (Ben Shachar et al. 2020).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

11. Dietary restriction

Ten replicates of pools of approximately ten virgin females or males were prepared for each of the diet-specific longevity tests, except for the complete medium condition, for which five replicates were conducted. Replicates were conducted in two separate batches of five, ensuring that each population was represented in both experimental batches (Table S2.3). Virgin females and males were transferred to the experimental media after ageing for four days on standard medium. All experimental media contained nipagin (Methyl 4-hydroxybenzoate) as a preservative. The diets were derived from the standard rearing medium, maintaining a sugar-to-yeast ratio of 5:4 in the complete medium, with corn flour removed. The yeast-deprived medium was based on the complete medium formulation with yeast omitted, and the sugar-depleted medium was similarly prepared by omitting sugar. The starvation medium consisted of agar and nipagin only as a preservative. Mortality was recorded three times per day (at 9 am, 1 pm, and 5 pm) every day until all flies had died. Time of death was then converted into days for analysis. Statistical analysis was performed using the same work flow as for longevity assays, only temperature was replaced by medium. Additionally, the starvation data were analysed separately as an independent trait, using a similar statistical approach.

12. Oxidative stress

Resistance to oxidative stress was assessed using 15 mM and 25 mM concentrations of paraquat. Ten replicates of pools of approximately ten virgin females or males were first aged for four days on standard rearing medium, and then transferred to media containing either concentration of paraquat (Table S2.4). Additionally, five replicate pools of approximately ten virgin females or males were maintained on standard medium as controls. The number of dead flies was recorded three times per day (at 9 am, 1 pm, and 5 pm) every day until all flies died. Time of death was then converted into days for analysis. Statistical analysis was performed using the same work flow as the longevity assays, only temperature was replaced by concentration. Control data was analysed and indicated that no confounding effect have occurred.

13. Metabolic rate

Metabolic rate measurements were obtained with a MAVEn™ device from Sable System International. Five flies for each sex and population were placed in arenas adapted to the device. Five replicates per condition were conducted, with measures taken for three cycles (three repeated measures

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

for each arena). Experimental temperatures were 18°C, 25°C and 30°C, in the darkness to prevent differences in activity due to unequal exposure to light. Volume of carbon dioxide (VCO2) emission and infrared-light activity detection were extracted and subsequently used as a proxy to metabolic rate and activity, respectively. Each pool of flies was frozen at -20°C and dried for 24 h at 75°C for subsequent weighing. VCO2 log-transformed data were analysed with linear mixed effect models in R (v 4.4.0), with dry weight and activity as covariants (R Core Team 2021; Christensen [2018] 2025). Type II ANOVA was also applied to assess the fixed effects, and estimated marginal means (EMMs) were adjusted using the Bonferroni adjustment method to account for multiple comparisons. Partial Eta^2 were calculated using package Effectsize (v 1.0.1) (Ben Shachar et al. 2020).

14. Evaluation of the phenotypic variation

To assess differences in phenotypic trait variability across populations, two complementary methods were employed. First, the raw standard deviation was calculated for each trait within each population and used as a proxy to phenotypic variation. These values were analysed using Friedman's rank test to evaluate whether variation in TE content explains phenotypic variability. As a complementary approach, the coefficient of variation (CV) was also computed for each trait (Colwell 1974; Burgess and Marshall 2014; Morgante et al. 2015). CV values were log-transformed (ln(CV)), and a linear mixed-effects model (LMER) was applied using the lmerTest package (v3.1-3), with the trait included as a random effect (R Core Team 2021; Christensen [2018] 2025). EMMs were adjusted for multiple comparisons using the Bonferroni correction, enabling population-level pairwise comparisons.

15. Linking TE insertions and phenotypes

The first approach consisted in investigating insertions present in high frequency in only one or a few populations experiencing specific phenotypes, so called *a priori*-approach. This method was reinforced by the investigation of insertions positioned in candidate regions, such as insertions found in 3' and 5' UTRs, exons and promoters. Using the insertion frequencies from TrEMOLO OUTSIDER module output (Mohamed et al. 2023), TE insertions were filtered based on frequency, using a threshold of ≥60% prevalence to identify variants likely to have a phenotypic impact at the population level. Insertions were then further filtered to retain only those present in a single population (within 100 bp windows). These were then intersected with the different gene regions of G0. TE insertion positions were cross-referenced with Flybase (https://flybase.org/) and manually curated to retain those with reported phenotypes

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

relevant to the traits under investigation in this study. For the second approach, a genome-wide association analysis was conducted using the Baypass tool, employed following the Bayes Factor (BF) (expressed in deciban (dB)) methods from (Mérel et al. 2021). Indeed, Baypass (v 2.41) was used to identify TE insertions involved in insertion-phenotype associations (Gautier 2015). By leveraging insertion frequencies obtained from TrEMOLO, we inferred the allelic frequencies of each insertion for each population and analysed the data using the Baypass pipeline. We then identified loci (*i.e.* insertions) with high BF values, indicating strong associations with phenotypic variation within a Bayesian framework. Since sequencing was conducted on pooled embryos with an unknown female-to-male ratio, we excluded insertions on sex chromosomes and only analysed autosomal TE insertions. Five iterations of Baypass were conducted, and the median of the five BF was taken as the indicator of strong associations with phenotypic variation within a Bayesian framework. Insertions with BF values greater than 14 were considered strong candidate insertions in accordance with Jeffrey's rule (Jeffreys 1948). In addition, an *a priori* manual curation was conducted to refine candidate selection for insertions with BF values below this threshold, taking into account their frequencies, genomic positions, and associated phenotypes as reported in FlyBase. Since Baypass utilizes phenotypic data, male and female datasets were analysed separately.

III. Results and Discussion

1. TE content increase with generation

We have taken advantage of a genetically modified line of *D. melanogaster* that was subjected to a temperature-inducible knockdown of the *piwi* protein in the somatic ovarian follicular cells (Barckmann et al. 2018), leading to an increase in TE insertions across generations and a high within population TE insertional polymorphism (Barckmann et al. 2018; Mohamed et al. 2020; Varoqui et al. 2025). Indeed, Varoqui and collaborators sequenced DNA at a high coverage (~60 to 140 X) from 100 males belonging to G11, G31 and G73 populations corresponding to 11, 31 and 73 generations of TE reactivation respectively (see Material and Methods for more information on these populations). TrEMOLO (Mohamed et al. 2023) was employed to assess the number of potentially full-length TE insertions detected, hereafter called TE variants, compared to the initial population (G0), detecting 280, 514, and 798 novel LTR retroelement insertions compared to the assembled G0 genome in each population respectively (Varoqui et al. 2025). We have acquired a subset of these populations (G0, S10, S31, S73 and S100) and confirmed the accumulation of TE insertions along with high insertional polymorphism by DNA sequencing (35X

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

coverage). Indeed, TrEMOLO identified 373, 414, 496 and 518 additional insertional TE variants in the S10, S31, S73 and S100 populations respectively compared to the G0 (Figure 2.1A). The average variant frequencies estimated with TrEMOLO (Mohamed et al. 2023) range from approximately 13% to 17% across the populations (Table 2.1), suggesting high TE insertion polymorphism in agreement with Varoqui (Varoqui et al. 2025). It is worth noting that global frequency distributions did not differ significantly between populations (Kruskal-Wallis p-value = 0.0659).

Population	Frequency mean	Frequency median	Standard deviation	Minimum frequency	Maximum frequency
S10	13.921	5.88	21.460	0.0135	100
S31	17.048	6.667	26.342	0.028	100
S73	13.034	6.927	18.942	0.014	100
S100	12.733	6.452	18.945	0.0294	100

Table 2. 1: Summary statistics of the insertion frequencies at the population scale.
Full-length TE insertions absent from G0 were detected using TrEMOLO, which also inferred insertion frequencies for each population. The table reports the mean, median, standard deviation, and the minimum and maximum frequency of TE insertions for each population.

The increase in TE content observed by Varoqui and colleagues mainly involves two different LTR families, *Zam* and *gtwin* (Varoqui et al. 2025). Indeed, in the populations included in this work, LTR retroelement insertions increase by respectively 247, 267, 281 and 295 in populations S10, S31, S73 and S100 (Figure 2.1A, Figure S2.3). Within this TE subclass, *Zam* and *gtwin* display significant increase in copy number compared to G0, confirming previous results (Barckmann et al. 2018; Varoqui et al. 2025) (Figure 2.1A, 2.1B and 2.1C). In addition to the 128 *Zam* and 116 *gtwin* initial copies present in G0, detected by RepeatMasker in the assembly, *Zam* full-length insertions increase up to population S73 (S10=5, S31=13, S73=21) to finally plateau in the S100 with a total number of 20 additional insertions (Figure 2.1A). As for *gtwin*, the number of detected full-length insertions continuously increases until generation 100 (S10=2, S31=2, S73=11 and S100=22) (Figure 2.1B). In addition to *Zam* and *gtwin* retrotransposons, we also observe an increase in a few other LTR superfamilies. Notably, *mdg4* (formerly *Gypsy* as seen in the figure)

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

full-length insertions increase up to generation 73 and then slowly plateau in S100 (S10=97, S31=121, S73=128 and S100=135) (Figure 2.1D). On the other hand, *Copia* elements remained relatively stable throughout the generations except for a slight decrease in the S100 (Figure 2.1D).

While the consecutive TE derepression of G0 only triggers new insertions of LTR retroelements, the subsequent populations also display an increase in both LINE and DNA transposons (Figure 2.1A, Figure S2.4 and S2.5). LINE elements increased by 101 in S10, and by 117, 190 and 198 in S31, S73 and S100, respectively (Figure 2.1A). For instance, *Jockey* superfamily continuously increase until generation 100 (S10=78, S31=90, S73=165 and S100=175) (Figure 2.1D). More specifically, two LINE elements, *F-element* and *Doc,* display an important increase in copy number across generations (Figure S2.4). As *F-element* transcripts were previously observed in the nurse cells of stage 2 and 3 up to the end of oogenesis in *D. melanogaster* (Minchiotti et al. 1994), the *F-element* could have hitchhiked with the virus-like particles (VLPs) produced by *Zam* and *gtwin*, or inherited through maternal contribution of RNA and proteins (Gutzeit and Koppa 1982; Brennecke et al. 2008; Park et al. 2019). In addition, the *F-element* displays a similar temporal and spatial expression pattern to *Zam* (Leblanc et al. 1997; Brasset et al. 2006). *Doc*, on the other hand, was reported to be expressed in all tested tissues in *D. melanogaster* (Vaury et al. 1994). The evolution of a ubiquitous expression pattern, as observed for the *Doc* element, likely underlies its successful invasion. Moreover, it may have benefited from the derepression of ERVs, and like *F-element* insertions, *Doc* could have bypassed the somatic-germinal barrier and hence increase in copy number.

In conclusion, with our experimental approach we were able to recapitulate the increase in TE insertions, within a G0 background, across the shifted population as previously observed (Varoqui et al. 2025).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

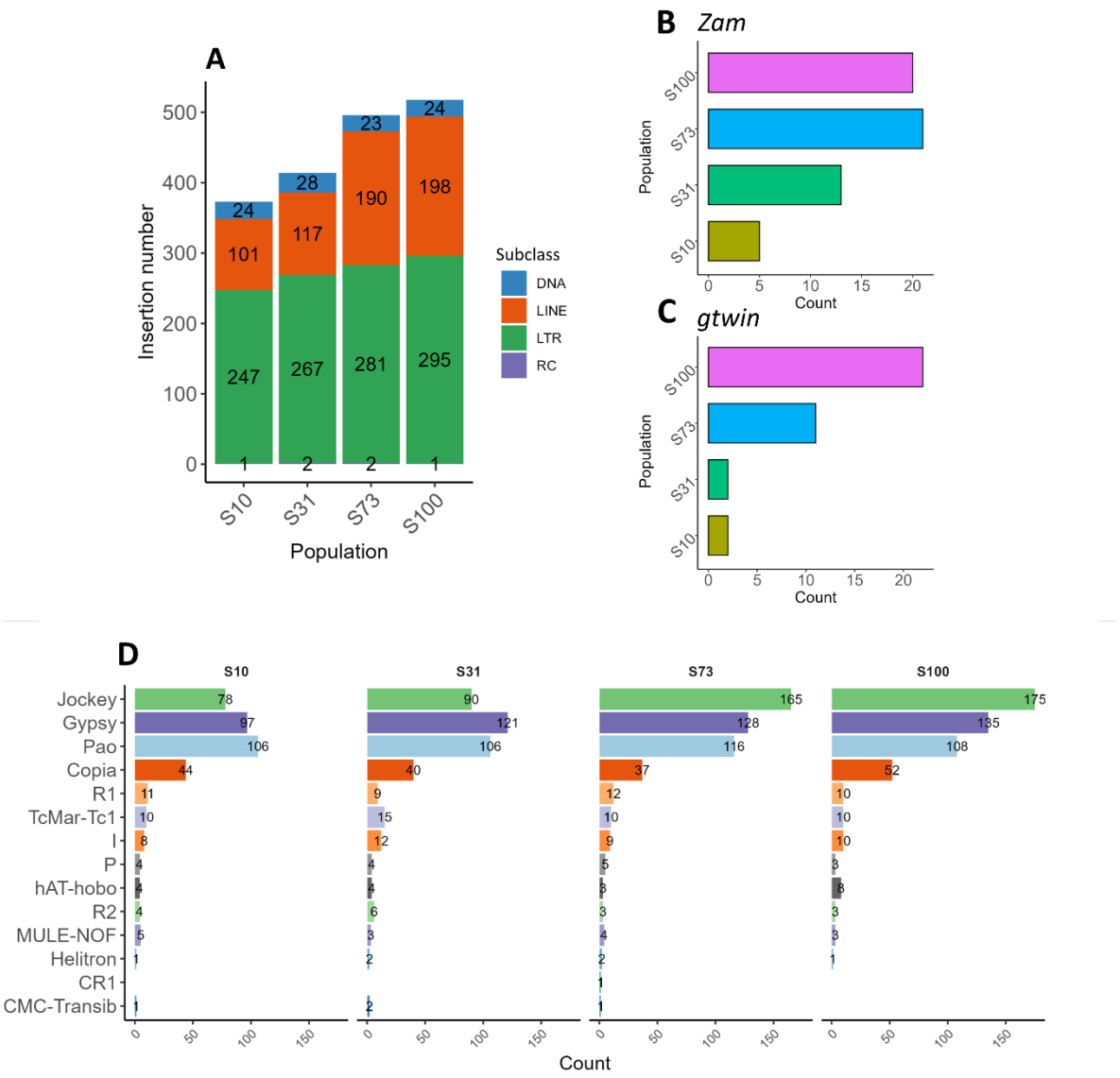


Figure 2. 1: TE content increases in the populations.

A) Distribution of the newly detected insertions compared to G0 per classes detected by TrEMOLO. The number in each of the colour category corresponds to the number of insertions for the class. New TE insertions increase with the generations, specifically for LTR and LINE elements. The number of new insertions in each population for the LTR Zam (B) and for gtwin (C). D) TE composition at the superfamily level. Likewise, to the plot A), the number indicate the insertion detected by TrEMOLO. Jockey and gypsy superfamilies increase significantly across the generations.

2. TE composition varies in frequencies and position between the populations

We used the insertion frequencies and positions to infer the variation in TE composition of the four populations derived from G0. We first performed a principal component analysis (PCA) on the insertion frequencies (Figure 2.2A). The PCA cluster the populations into three groups. PC1 explains 36.64% of the variance and clusters S10 and S31 together, followed by S73 and S100 (Figure 2.2A) while

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

PC2 explains 25.43% of the variance and mostly excludes S73 from the other populations (Figure 2.2A). This PCA analysis highlights the insertional polymorphism between populations, adding a layer of variability beyond TE load. We further investigated the genomic position of the insertions to evaluate if newly inserted TE were displaying specific pattern of enrichments (Figure 2.2B and 2.2C, Figure S2.6).

Permutation tests were first performed on all insertions regardless of their frequencies. These tests suggest that the TEs are depleted from exons and 2kb promoter regions in all five populations, as well as in 3'UTRs for S73 (Table S2.5). Such results are expected, considering that insertions in exons and promoter regions are likely to impair gene function and expression (Hedges and Deininger 2007; Chuong et al. 2017; Dubin et al. 2018). Permutation tests were then conducted on the five defined frequency categories to refine the investigation (see Material and Methods). Insertions <20% are depleted in promoter, 3'UTR regions, in introns and exons for S10 and S31 (Table S2.5). This pattern, except for the 3'UTRs, where the number of observed insertions is not significantly different from expectation, is also consistent in S73 (Table S2.5). In S100, promoters, 3'UTRs, and exons are depleted, while introns and 5'UTRs are not. Insertions in the 20-40% frequency category show neither depletion nor enrichment in any genomic compartment in S10 and S31 (Table S2.5). However, permutation tests indicate a depletion in promoter regions for both S73 and S100, as well as in exons for S100 (Table S2.5). For the 40-60% frequency class, S73 harbours a depletion of new TE insertions in promoters. Notably, in this frequency class, S10 shows a considerable enrichment of new TE insertions in introns, with 11 insertions observed compared to the number of expected insertions (p-value = 0.038) (Table S2.5). Interestingly, for the 60-80% frequency category, representing insertions relatively common in the populations, S100 shows an enrichment of new TE insertions in introns, while none of the other populations display either a depletion nor an enrichment pattern (Figure 2.2C). This may suggest that S100 retains TEs preferentially in introns rather than in other genomic regions (Table S2.5). Finally, for insertions with a prevalence above 80%, only S31 displays a remarkable pattern. While S31 carries fewer new TE insertions in promoter regions than expected (6 detected versus 13.04 expected), it shows more in introns (10.3 expected versus 20 observed). Thus, both S100 and S31 harbour considerably more medium- to high-frequency insertions in introns compared to the other populations. Notably, we already reported a plateau in TE content in S100, which may indicate that its pattern of medium- to high-frequency insertions reflects a true biological response to the inducible *Piwi* knockdown-driven TE accumulation.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Here, we demonstrate that not only TE content varies between populations, but that their frequency profiles and distributions across genomic compartments are also distinct. Taken together, the variation in TE content and positioning, along with their potential regulatory and structural implications, makes these populations an ideal experimental system for evaluating the impact of TEs on phenotypic variation and plasticity.

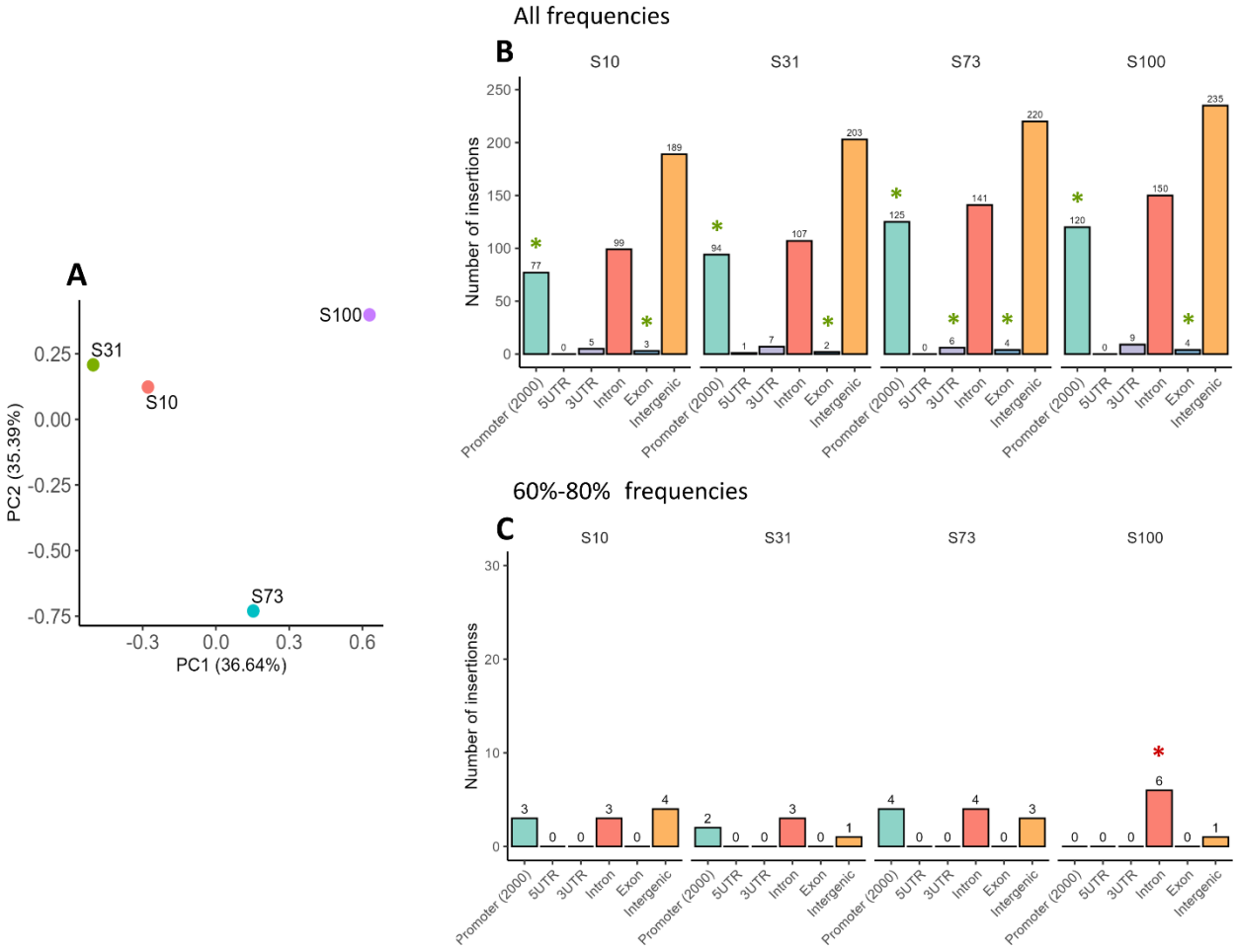


Figure 2. 2: TE composition varies between the populations.

A) PCA of the TE frequencies. Both PC1 and PC2 explain a considerable proportion of the variation between populations, separating S73 and S100 from S10 and S31. S10 and S31 cluster together suggesting their TE landscape is similar. S73 and S100 are separated both along PC1 and PC2, suggesting distinctive TE compositions. B) Genomic distribution of TE insertions, including all insertions of all frequencies. C) Genomic distribution of TE insertions with frequencies in the 60%-80% range. S100 displays an enrichment in additional insertion compare to G0 in the intronic regions. Green Asterix indicated a significant depletion, and the red one the significant enrichment in TE insertions, based on a two-sided test of 1000 permutations.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

3. Variable TE contents influence the phenotypic profiles across the populations

To evaluate TEs' impact on phenotypes, a broad range of traits were measured, aiming at skimming a panel of biological functions. Out of the 17 traits measured, a population effect, *i.e.* the effect of TE composition, was observed in six female traits, seven male traits and two traits with undetermined sex (mixed) (p-value < 0.001) (Table 2.2).

Phenotype	Measure	Environment	Sex	Population effect	Statistical test
Hatchability	Egg-to-first-instar success rate	18°C	Mixed	***	GLM binomial
Viability	Egg-to-adult success rate	18°C	Mixed	***	GLM binomial
Developmental time	Egg-to-adult duration	18°C	Mixed	***	Log-Rank test
Dry weight	Proxy for size	18°C	F	**	ANOVA
			M	**	
Climbing	Negative geotaxis assay	18°C	F	ns	GLMER binomial
			M	*	
Thermal resistance	CTmax		F	**	LMER
Longevity		18°C	F	**	LMER
		22°C	F	***	
		18°C	M	***	
		22°C	M	***	
Metabolic rate	CO2 emission	18°C	F	ns	LMER (covariants: dry weight and activity)
		25°C	F	ns	
		30°C	F	*	
		18°C	M	ns	

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

		25°C	M	*	
		30°C	M	ns	
Longevity		Complete	F	***	LMER
		-Sugar	F	***	
		-Yeast	F	***	
		Starvation	F	***	
		Complete	M	***	
		-Sugar	M	***	
		-Yeast	M	**	
		Starvation	M	***	
Longevity		15mM paraquat	F	***	LMER
		25mM paraquat	F	.	
		15mM paraquat	M	***	
		25mM paraquat	M	***	

Table 2. 2: Summary of phenotypic measurements across populations, sexes, and environmental conditions.
*The "Phenotype" column indicates the trait assessed, "Measure" specifies how the trait was quantified, and "Environment" indicates the conditions under which the measurement was performed. "Sex" identifies whether the measurement was made on females (F), males (M), or mixed. "Population effect" denotes the statistical significance of differences among populations. "Statistical test" specifies the model used to assess the population effect, including generalized linear models (GLM), linear mixed-effects models (LMER), generalized linear mixed-effects models (GLMER), and Log-Rank tests for time-to-event data. For the metabolic rate, the LMER analyses accounted for dry weight and activity as covariates. -Sugar and -Yeast designate the nutritional source depleted from the media. Signification codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 'ns' 1*

Amongst the most striking differences was the significant decrease in hatchability (the proportion of eggs from which a larva emerges (David et al. 1975)) in the S31 and S73 compared to the G0 and S10

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

(p-values = 0.0002, 0.0032, respectively) (Figure 2.3A). The decrease in hatchability observed is in agreement with Pasyukova et al. (2004) where a *D. melanogaster* strain showed a deficiency in hatchability (Pasyukova et al. 2004). This line had up to 90 additional TE copies of *copia*, *roo* and *P-element* compared to the other lines. They propose that the effect of TEs on hatchability may be a consequence of higher rates of ectopic exchanges resulting in deleterious chromosomal rearrangements and in an increased embryonic death (Pasyukova et al. 2004). These, together with our results, testify of the impact of the TE load on hatchability.

In addition, a decrease in viability (measure of the egg-to-adult survival (Flatt 2020)) in the S73 compared to all other populations suggest that not only hatchability is impacted by TE load, but later developmental stages are also affected in this population (Figure 2.3B). Viability in optimal condition is expected to range from 70% to 80% in *D. melanogaster* (Flatt 2020). Coherent with our observation, previous fitness measurements conducted on three lines with varying *P-element* content revealed that viability decreased in strains contaminated by *P-elements* (Mackay 1986).

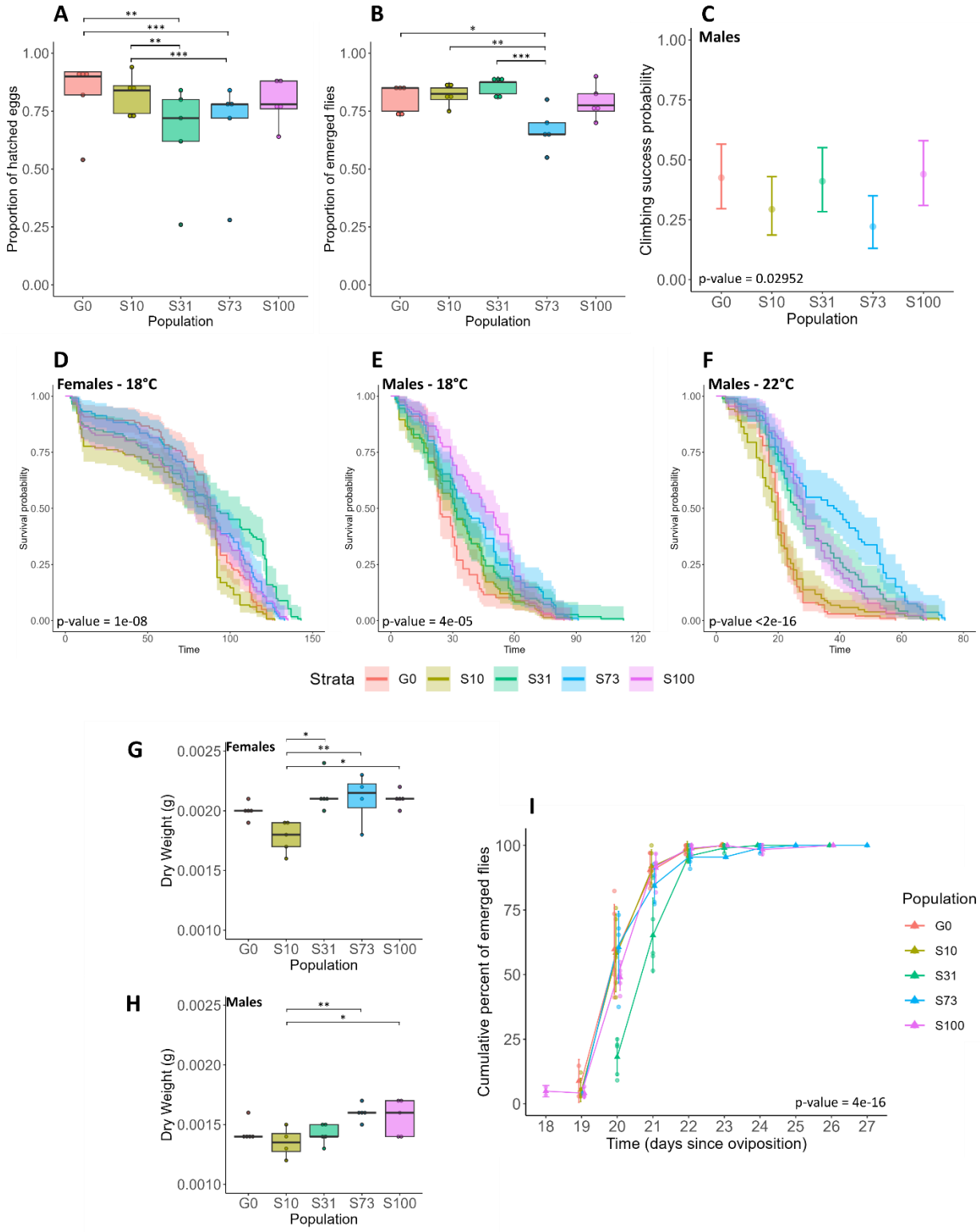
Finally, the negative geotaxis assays show a significant population effect (p-value = 0.02952) with marginally significant lower propensity to reach the upper part of the test vial for the males S73 compared to the males S100 (p-value = 0.0556) (Figure 2.3C). It was previously reported that males with somatically active *P-elements* have shown lower locomotion than their *P-element*-free counterparts (Woodruff et al. 2000). We hypothesized that the marked increase in TE content from S31 to S73 (414 and 469 insertions, respectively), followed by a plateau at S100 (518 insertions), could reflect a specific turnover in TE insertions at S73, acting as a burden and potentially manifesting as an impairment in males' locomotion.

While the three previously described phenotypes appear to reflect the influence of overall TE load, without showing a strict correlation with its increase *per se*, the subsequent measurements seem to be driven by population-specific insertions, highlighting the potential targeted effects of particular TE insertions on phenotypic differences between populations.

First, longevity measurements were carried out knowing that TEs are linked to ageing, by disturbing overall biological processes (Copley and Shorter 2023; Wood and Helfand 2013; Yushkova and Moskalev 2023; Merenciano et al. 2025), however no general pattern of correlation emerges from this data. For example, at 18°C the females S10 experience a decrease in longevity compared to S31, S73 and S100 (Log-Rank test p-values = 7.7e-07, 0.0031 and 0.01804, respectively) (Figure 2.3D). The lifespans of G0 and S10 males at 18 °C are also reduced compared to those of S73 and S100 (Log-Rank test p-values: 0.03968 and 0.00275 for G0 *vs.* S73 and S100, respectively; 0.00482 and 0.00045 for S10 *vs.* S73 and S100,

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

respectively) (Figure 2.3E). The decrease in lifespan of the males G0 and S10 appears to be emphasized at 22°C, while S73's is drastically increased (Figure 2.3F).



Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Figure 2. 3: Phenotypic characterization of the populations with increasing TE content.
A) Hatchability expressed in proportion of hatched eggs into larvae. Each condition is composed of five replicates of 50 eggs. S31 and S73' hatchability appears reduced compared to both G0 and S10. B) Viability, representing the proportion of survival from egg to adult. Each condition comprised of five replicates of 40 initial eggs. S73 suffers from a significant reduction in viability. C) The fitted climbing success probability for the flies of each population to have reached the top of the tube at the end of the trial. Each condition was replicated six times with pool of 10 flies. S73 expresses a slightly lower propensity to reach the upper part of the tube than S100. D)-F) Longevity measured as the survival curves of the females at 18°C (D) and the males at 18°C (E) and at 22°C (F). Females S31 have a longer lifespan, while G0 and S10 males have a lower longevity, which is exacerbated at 22°C. The dry weight of the females (G) and males (H). Each of the conditions is represented by five replicates of pool of five flies. S10 females and males alike appear to be lighter than the populations with higher TE contents. I) Cumulative percent of emergent flies as a function of the day since oviposition. Each condition comprised five replicates of 40 initial eggs. S31 experiences a consistent delay in development time.

The dry weight appears as a potential predictor of the S10's reduced lifespan (Figure 2.3G and 2.3H). In *D. melanogaster*, the relationship between body size and longevity is highly variable, however a positive correlation has been previously described for males of a wide panel of tested strains (Khazaeli et al. 2005). And indeed, in males S10, individuals are generally smaller (the dry weight is used as a proxy for size) than both S73 and S100 (p-values = 0.0073757 and 0.0201497, respectively) (Figure 2.3H). Similarly, females S10 appear to be smaller than S31, S73 and S100 (p-values = 0.0184431, 0.0051339 and 0.0418102, respectively) (Figure 2.3G).

It is also worth noting the delay in development time of the S31 from the other populations (Figure 2.3I). Longer development time is often correlated with longer lifespan (Lints and Lints 1971; Mayer and Baker 1985; Stearns et al. 2000). Therefore, it is not unexpected to also observe an increase in longevity in the females of the S31 population compared to G0, S10 and S100 at 18°C (Log-Rank test p-values = 0.00052, 7.7e-07 and 0.02607, respectively) (Figure 2.3D). Indeed, the average emergence time for S31 is 21.23 days, while it lays around 20 days for the other populations as expected at 18°C (Pulver and Berni 2012) (Figure 2.3I). Therefore, S31 consistently takes one more day to reach adulthood. The difference in development time being specifically bound to one population may suggest that this is the result of one or more isolated TE variants' effect, rather than related to the global TE content.

Generally, the longevity measurements reveal that the impact of TEs on longevity is not as straightforward as a simple correlation between higher TE content and reduced longevity. Instead, the few significant differences observed across populations appear to reflect population-specific effects rather than a general correlation with TE abundance. This is further supported by correlations between these population-specific effects across different traits. These findings suggest that phenotypic variation

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

may not solely arise from total TE load, but also from the presence of specific insertions contributing to phenotypic differentiation.

In an attempt to see if the phenotypes can explain differences between the populations, PCAs were conducted to explore if a pattern emerges. Pairwise Pearson's correlation coefficients were also computed to further investigate the correlations between traits. In females, G0, S10 and S100 cluster together while S31 and S73 separate along PC1 and PC2 respectively (Figure 2.4A). Pearson's coefficients reveal six significant positive correlations between phenotypes as well as six negative correlations (Figure 2.4C). Notably, development time seems to be positively correlated with longevity at 18°C (0.96), measured as the LT80 (the estimated time for 80% of the individuals to have died (see Material and Methods), as previously discussed (Figure 2.4C). In males, PCA analysis shows a slight clustering of G0 and S10 on one side, and of S73 and S100 and the other side, leaving S31 separated along PC2 (Figure 2.4B). Pearson's coefficients reveal six positive correlations for one negative (Figure 2.4D). Additionally, for each phenotype, partial Eta^2 and R^2 metrics were extracted to estimate the proportion of variance explained by the TE content (according to the statistical model) (Figure 2.5). One phenotype for males and six for females fall below the threshold of mild effect (Figure 2.5). In females, three phenotypes are mildly to strongly impacted by the TE content, while four for males and one for mixed (Figure 2.5). Altogether, 14 phenotypes are strongly impacted by the TE content (Figure 2.5).

These results suggest that differences in TE composition in a nearly identical genetic background influences diverse phenotypes, highlighting the importance of considering cumulative TE effects in *D. melanogaster.*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

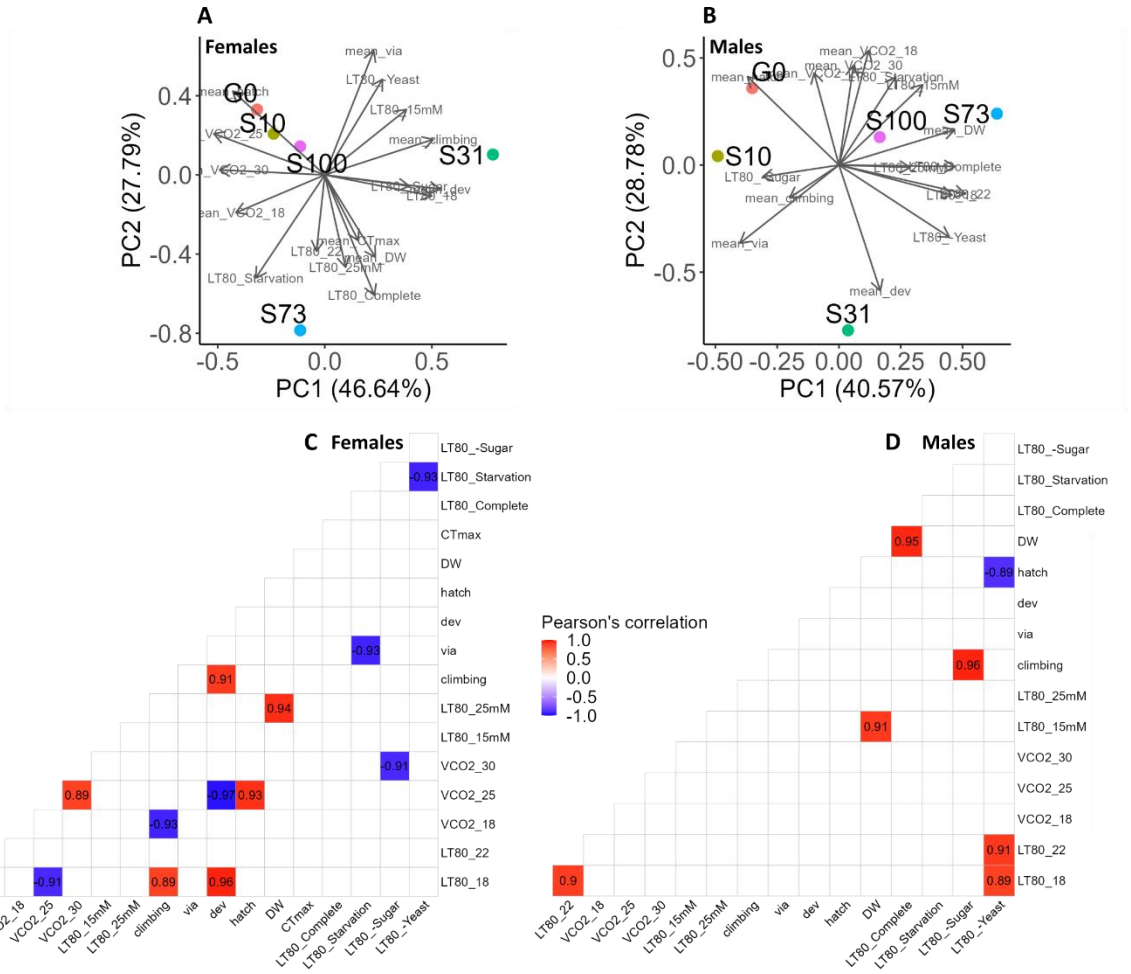


Figure 2. 4: Dimensional and correlative data analysis of the phenotypes of the population with increasing TE content.

A) and B) PCA of the means and LT80 of the females and males, respectively. C) and D) Pearson's correlation matrixes of the means and LT80 for the females and the males respectively. "VCO2" refers to the metabolic rate measurements, "hatch", to the hatchability, "via", to the viability, "dev", to the development time, "climbing", to the negative geotaxis assay, "15mM" and "25mM" apply to the concentrations of paraquat, "DW" is the dry weight, "18", "22", "25", "30" are the environmental temperature, and "Complete", "Starvation", "-Yeast", "-Sugar", the environmental media.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

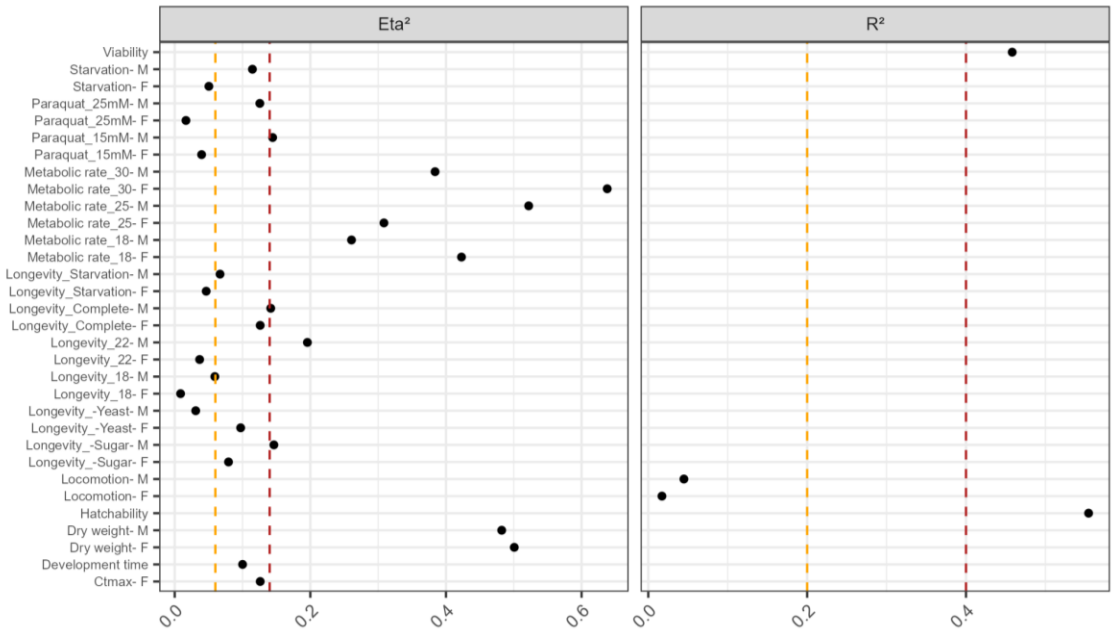


Figure 2. 5: Summary of the effect sizes of the contribution of the "population" variable to the data. *The contribution of the population variable to the observed data are expressed as the Eta² and the R² according to the nature of the data and statistical tests. The yellow vertical dashed line represents the threshold of medium effect and the red the threshold of strong effect. "F" refers to measures in females and "M" in males, "15mM" and "25mM" are the concentrations of paraquat, "DW", the dry weight, "18", "22", "25", "30" are the environmental temperature, and "Complete", "Starvation", "-Yeast", "-Sugar", the environmental media.*

4. High TE content increases the observed phenotypic variability in females

While the previous section has shown that TE content explains part of the differences in the phenotypes that were observed, the data also indicate differences in the variance of the measured phenotypes depending on the TE content. Compiling the standard deviations of all the phenotypic experiments for the females (a proxy for phenotypic variation as a measure of deviation from the mean), a Friedman's rank test was performed showing that S73 variation was consistently greater than that of G0 and S10 (Friedman Rank test p-values = 0.022 and 0.017, respectively) (Figure 2.6A, Table S2.6). Additionally, we computed the coefficient of variation (CV) for all the phenotypes, a measure of dispersion that expresses the standard deviation as a percentage of the mean and hence frees itself from the relationship to the mean (Morgante et al. 2015). CV was again higher for the S73 compared to the G0 and S10 (ANOVA p-values = 0.023 and 0.0062, respectively) (Figure 2.6B). Both methods led to similar results supporting the hypothesis that indeed S73 having a greater TE content associated with higher insertion polymorphism in the population generally expresses a wider range of variation of its phenotypes. S31 may not yet have accumulated enough TEs for this pattern to be observed. As for S100, although it carries a

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

slightly higher number of insertions than S73, its phenotypic profile appears closer to that of G0 (Figure 2.6A). The particular profile of S100 may be explained by two aspects of its TE content. First, S100 shows a distinctive enrichment pattern for insertions in of high frequencies, particularly for the 60%-80% frequency range. Specifically, it shows an enrichment of TE insertions in the introns compared to the other populations. TE insertions within introns may have relatively minor effects on gene function and expression compared to other genomic compartments, possibly making their accumulation in introns a way to safeguard the genome from transcriptomic variation that could otherwise be reflected at the phenotypic level. An analysis of the overall distribution of TE insertions within the transcriptome reported that 60% of the TEs are inserted in introns in *D. melanogaster* and specifically 52% of LTRs occurred in introns, corroborating the bias in accumulation across genomic regions (Sela et al. 2010). Second, an analysis of TE abundance per haplotype using a normalized coverage method with DeviaTE (Weilguny and Kofler 2019) revealed that S100 shares a similar overall TE abundance with S10 and S31, with values of 1263, 1257, and 1303, respectively. This result was consistent across three replicates of 35 X S100 subsampling. For reference, TE abundance in S73 was estimated at 1466, which is considerably higher than in the other four populations (Figure S2.7). Hence, this may explain why only S73 exhibits greater phenotypic variation. While such a hypothesis might partly explain the behaviour of S100, our understanding remains limited by the lack of transcriptomic data on TE and gene expression interactions, and on the overall epigenetic landscape.

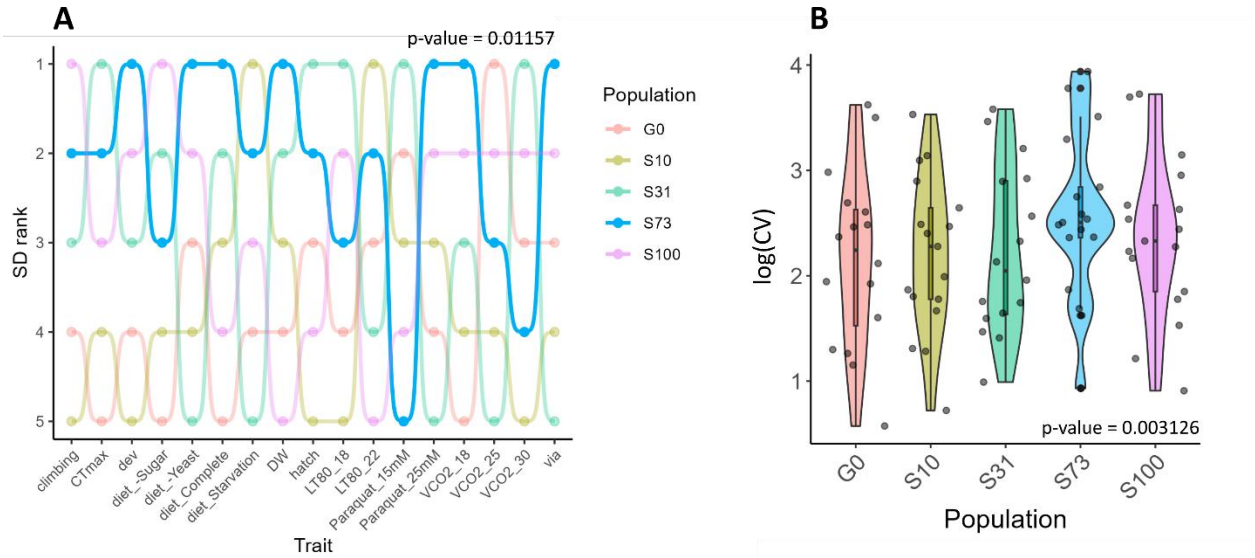


Figure 2. 6: High TE content and insertional polymorphism increases phenotypic variation in females. *A) Ranking of standard deviations (SD) across traits. Population S73 consistently occupies the first and second ranks for many measured phenotypes, indicating elevated variability. Ranking differences were assessed using a Friedman test. B) Logarithm of the coefficient of variation of the trait (CV) across the*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

measured phenotypes, showing higher values for S73 compared to the other populations. Each dot represents the CV for a single phenotype. Statistical comparisons between populations were performed using a linear mixed-effects model (LMER). "VCO2" refers to the metabolic rate measurements, "hatch" to the hatchability, "via" to the viability, "dev", the development time, "climbing", the negative geotaxis assay. "15mM" and "25mM" are the concentrations of paraquat, "DW", the dry weight, "18", "22", "25", "30" are the environmental temperature, and "Complete", "Starvation", "-Yeast", "-Sugar", the environmental media.

This suggests that variability in TE insertions, even within a given population and thus within an otherwise identical genetic background, can yield significant phenotypic variability. It also indicates that a high TE load alone may not induce such variation; rather, it likely depends on the regions where the TE insertions occur, as seen with S73 and S100, which exhibit different accumulation patterns despite carrying a similar TE load. Few studies have previously highlighted the contribution of TEs to phenotypic variability. For instance, in *P-element*-burdened strains, the variance in traits related to fitness, viability, and fertility was observed to be higher than in the control strain (Mackay 1986). Alternatively, in a study comparing *Capsella rubella*, commonly called the pink shepherd's-purse, and its sister species *C. grandiflora*, it was reported that although *C. rubella* exhibits reduced genetic variation due to a severe bottleneck event following speciation (Guo et al. 2009), it carries remarkable TE polymorphism compared to *C. grandiflora.* This turned out to be associated with a significant increase in transcriptomic variability near the *FLOWERING LOCUS C (FLC)* gene, a hotspot for *Helitron* insertions, which may explain the phenotypic divergence seen across natural populations (Niu et al. 2019). Although rare examples of TEs' contribution to phenotypic variability are available, and especially in animals, there has been extensive theoretical work on the topic (Whitelaw and Martin 2001; Slotkin and Martienssen 2007; Rey et al. 2016a; Chuong et al. 2017; Pimpinelli and Piacentini 2020; Catlin and Josephs 2022b). This study provide evidence that TEs influence phenotypic variance and, for the first time, demonstrates this effect across an extensive diversity of phenotypes.

5. Variation in TE content impacts the responses to environmental changes

Phenotypic variability depends on multiple components, one of which is the impact of the environment. Typically, to assess the environment-induced effect, the genetic component is minimized. Taking advantage of multiple populations sharing the same genotype but differing in TE content, we were able to isolate and quantify the contribution of TEs to the phenotypic plasticity of *D. melanogaster.* Four measures of environmental responses were performed on the five populations with increasing TE content: a measure of longevity under two temperatures and under four diets, a measure of oxidative stress

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

resistance at two different concentrations, and finally, a measure of the metabolic rate under three different temperatures. All four experiments presented a plastic response of the phenotype, and three displayed a significant interaction between genotype (*i.e.* TE content) and environment.

First, longevity was assessed at 18°C and 22°C as it is well established that poikilothermic organisms, such as *D. melanogaster*, experience significant modulation of lifespan in response to temperature changes, making lifespan a plastic trait (David and Capy 1988; Zwaan et al. 1992; Vermeulen and Bijlsma 2004). The measures were conducted in females and males. The longevity decreased significantly with the increase in temperature (Figure 2.7A). A two-step approach was used to analyse longevity. First, the estimated time for 80% of the flies to have died (LT80) was computed along with survival estimates under a probit distribution. Second, the data were analysed using a linear model (LMER) based on the time of death of each individual fly (see Material and Methods). In males and females, both LT80 and LMER analysis support a strong effect of the interaction population by environment, suggesting that different populations and hence TE content, experience different responses to temperature (TExE interaction). The S31 shows a rather drastic decrease in longevity with temperature increase (slope = -12.9), while S10's reaction norm is nearly flat (slope = -1.5) (Table S2.7). These two cases illustrate the differences in the response to temperature changes. As for the males, the reaction norms are overall less steep than the females (average slope = -3.5 in males for -7.8 in females) (Table S2.7), however they still reveal variation in responses (Figure 2.7A). In both males and females, the TExE interaction appears to be more influenced by the TE composition rather than the overall TE load. Specifically, longevity does not follow a consistent pattern across generations of TE accumulation, suggesting that the composition of TEs, rather than their sheer number, plays a role in shaping the response to environmental changes. These findings suggest the presence of a TE-driven plasticity in *D. melanogaster*.

Measures of longevity on different diets were also conducted at 18°C (Figure 2.7B). Complete medium contained a 5:4 ratio of proteins to carbohydrates, in the form of yeast and sugar. Medium was alternatively deprived of one or the other nutritional source, and finally medium was deprived of all nutritional sources to test starvation resistance. In both males and females, we observed a very strong population-by-environment effect, *i.e.* TE-by-environment effect (all p-values < 2.2e-16). LT80 highlights the gradual decrease in lifespan as medium is deprived of nutritional value (Figure 2.7B). Interestingly, deprivation of proteins from the diet appears to have a greater impact on females than on most male populations, except for S73 (Figure 2.7B). In fact, where the decrease in lifespan is gradual with depletion in nutritional values in females, in males the deprivation in protein intake, by removal of yeast from the media, seems to have a smaller effect (Figure 2.7B). Instead, lifespan drops significantly from the complete

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

medium to the sugar deprived one (Figure 2.7B). The differences between males' and females' responses to dietary restrictions are an indicator of the sexual dimorphism and the variation in the nutritional needs likely tied to their distinct energy expenditure pattern. Indeed, it has been previously demonstrated that males' lifespan is maximized on lower protein regimen (Wu et al. 2020; Carey et al. 2022), whereas females require significant input of proteins to allow proper egg production (Wu et al. 2020; Carey et al. 2022). Additionally, strong interactions between populations and the environment are seen, suggesting the importance of TE content in the response to changes in diet. This has never been observed before, and hence, reasons for the differences between the populations remain to be investigated.

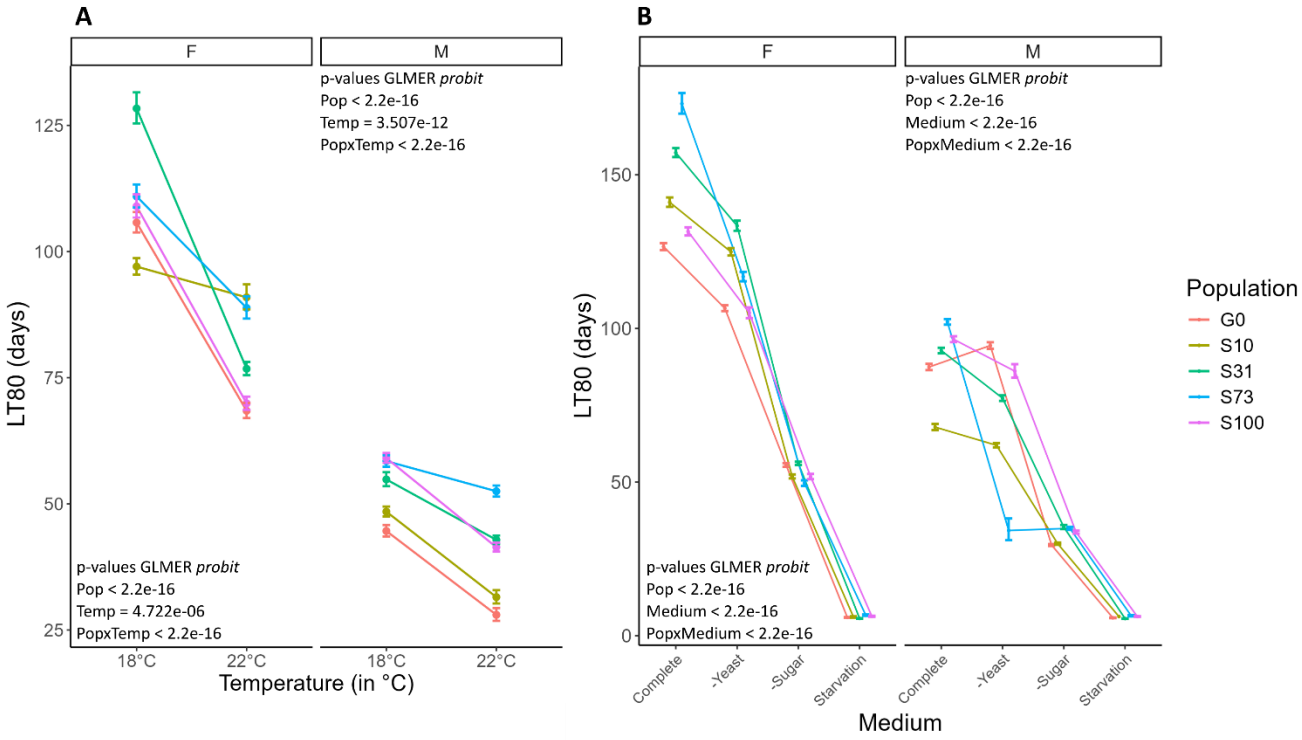


Figure 2. 7: Longevity varies under variable environment demonstrating a plastic response.
A) Reaction norms of the LT80 for the females on the left and males on the right at variable temperature, namely 18°C and 22°C. Higher temperature drastically reduces longevity, more so in females than in males. However, this decrease is population-dependent. B) Reaction norms of the LT80 for females and males, fed on different media. The complete media represent a standard media in which the corn flour was removed, "-Yeast" is the complete media without the yeast, "-Sugar" is the complete media without the sugar and "Starvation" is a media with only water agar. F refers to the females and M to the males. Females express a gradual decrease in lifespan with each diet depletion. As for the males, apart from the S73, the depletion in protein by removal of the yeast from the medium only slightly impact lifespan. Both environmental variations are supported by a significant GxE interaction.

Next, we investigated the metabolic rate as an umbrella measure providing insight into a broad range of processes that may otherwise be challenging to measure directly. Metabolic rate can be influenced by various biotic factors, such as body weight or age, as well as abiotic factors, including

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

temperature, as tested here (Arking et al. 1996; Berrigan 1997; Van Voorhies 2009; Alruiz et al. 2023). Both the metabolic rate of the females and the males were significantly influenced by the temperature of the assay and by the population variable. This suggests that the metabolic rate is indeed a plastic trait. On the other hand, the data shows no TExE interaction, suggesting that all populations reacted in a similar fashion to the temperature change (Figure 2.8A and 2.8B). Notably, the dry weight covariate explained some of the variance in females but none in males, whereas the opposite trend was observed for the activity covariate (Table S2.8 and S2.9). Videlier and collaborators observed a similar pattern when measuring the basal metabolic rate (BMR) in *D. melanogaster* (Videlier et al. 2019). They reported a positive correlation between activity covariate and BMR in males, whereas in females, BMR was either uncorrelated or even slightly negatively correlated with activity. Furthermore, a follow-up study on the genetic components highlighted sex-specific differences in the genetic architecture of locomotor activity and its impact on BMR (Videlier et al. 2021). This reflects the differences in energy investment between the sexes, as it is the case for gamete energy allocation, where females across all animal kingdoms invest more than males (Hayward and Gillooly 2011). Females are likely investing energy in reproduction, particularly in egg production, with larger females typically producing more eggs (Lefranc and Bundgaard 2000). In contrast, males' reproductive investment relies more on their ability to find and court a mate (Asahina 2018). Therefore, males likely invest more energy in activity than females. These covariates reflect the sexual dimorphism in energy allocation suggesting that TE content, and in turn the environment, influence the phenotypes. However, no interaction between TE content and the environment is observed, as the reaction norms follow a comparable increase across the generations. This also suggests that, the effect of TEs may occur on very specific traits (as previous observed), rather than having an overall effect on the metabolism of the organism.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

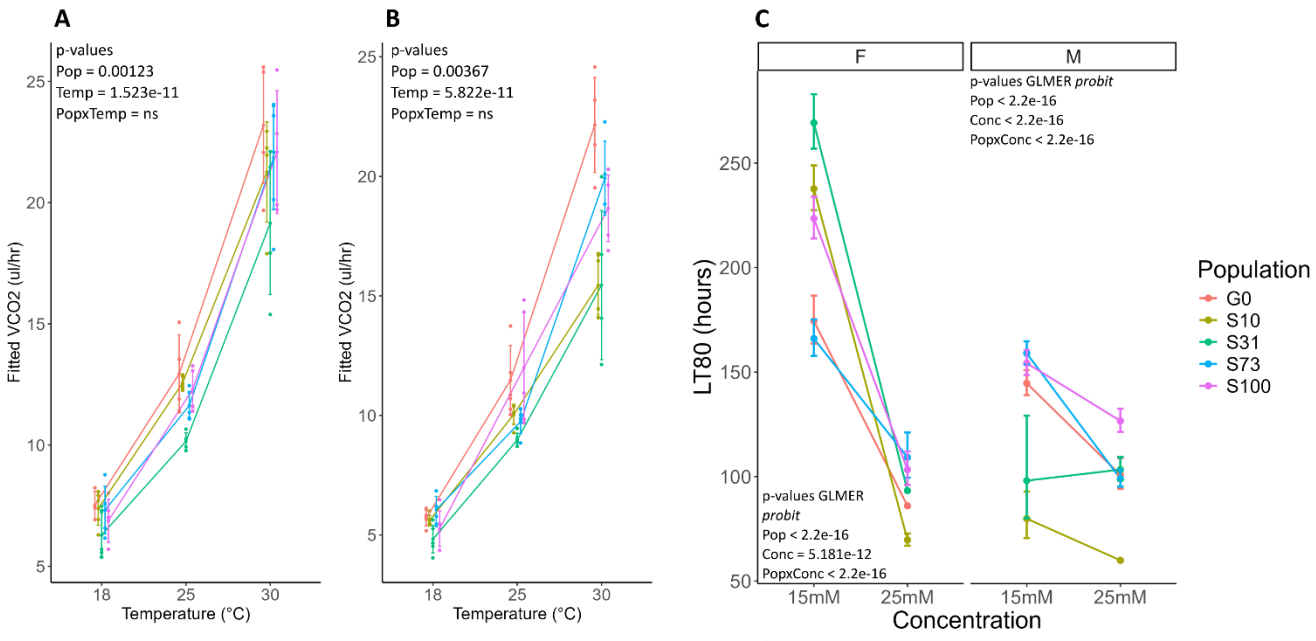


Figure 2. 8: Metabolic rate and resistance to oxidative stress under variable conditions.
A) and B) Plot of the fitted values of the metabolic rate for the females and males, respectively. Each condition is represented by five replicates of pool of five flies. The fitted values account for the dry weight and the activity as covariants. C) Plot of the LT80 under variable paraquat concentrations, for the females on the left and the males on the right. F refers to the females and M to the males.

Finally, the populations were exposed to two different paraquat concentrations, to measure the effect of different intensity of oxidative stress. Exposure to paraquat increases the formation of reactive oxygen species (ROS), known for their impact on longevity by causing cellular damage (Oleson et al. 2021; Shields et al. 2021; Costantini 2024). In both males and females, we report a significant interaction between TE content and the environment, *i.e.* the paraquat concentration (Figure 2.8C) (Table S2.10 and S2.11). Female's longevity is significantly reduced by the increase from 15 to 25 mM of paraquat (average slope = -12.2), whereas males' hardly drops (average slope = -2.9) (Figure 2.8C, Table S2.12). Interestingly, females' S73 longevity difference from 15 to 25mM is significantly reduced compared to the other populations with a slope of -5.7 (Figure 2.8C). This last observation seems to be tied to the fact that S73's resistance to the lowest concentration of paraquat is significantly lower than that of the other populations, except for that of G0 (Figure 2.8C). Estimated survival probabilities from the GLMER model suggest that only females G0 and S100 chances of survival from one time point to the next do not change with paraquat concentrations. Additionally, at 25 mM, females G0 and S10 have significant lower probability of survival than the other populations, echoing the results obtained for the longevities under different temperatures. Males S10 have a strikingly low resistance to paraquat and especially at 25 mM concentration, with an LT80 around 60 h, for 100 h to 150 h in the other populations (Table S2.13).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Summarizing the plasticity assays, all measured traits showed plasticity, of which only one is not under the influence of an interaction between TE content and environment (Table 2.3). Partial Eta^2 were computed for the population, environment and interaction variable, to estimate the effect size of each to the observed phenotypes (Figure 2.9). Despite significant variable effects, effect sizes are in some instance quite subtle, while the opposite is also observed, as it is the case for MR. The discrepancies may lay in the fact that while the effect is statistically supported, the amplitude of the effect can be low (Sullivan and Feinn 2012; Dunkler et al. 2020). This can be explained by large sample sizes (as it is the case for the longevity measurements), whereas the reciprocal situation can be explained by small sample size.

Phenotype	Measure	Environment	Sex	Population effect	Interaction term	Statistical test
Longevity	Variable temperatures	Temperature	F	*	***	LMER
			M	***	**	
Metabolic rate	CO2 emission	Temperature	F	**	ns	LMER (covariants: dry weight and activity)
			M	**	ns	
Longevity		Diet	F	***	***	LMER
			M	***	***	
Oxidative stress		Paraquat concentration	F	**	***	LMER
			M	***	***	

Table 2. 3: Summary table of statistical analysis and results of the phenotypic measurement under variable environment.

*The "Phenotype" column indicates the trait assessed, "Measure" specifies how the trait was quantified, and "Environment" indicates the conditions under which the measurement was performed. "Sex" identifies whether the measurement was made on females (F) or males (M). "Population effect" denotes the statistical significance of differences among populations. "Interaction term" refers to the statistical significance of effect of the interaction of the population with the environment. "Statistical test" specifies the model used to assess the population effect including the linear mixed-effects models (LMER) and generalized linear mixed-effects models (GLMER). For the metabolic rate, the LMER model accounted for dry weight and activity as covariates. Signification codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 'ns' 1*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

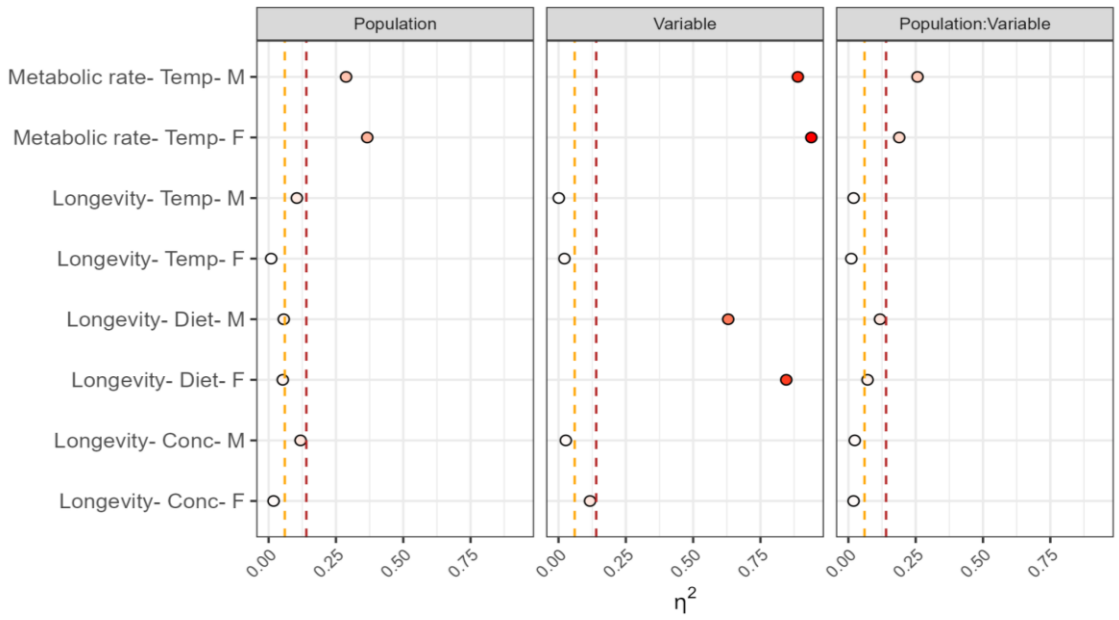


Figure 2. 9: TE content contributes to differences in plasticity responses of various traits in a range of environments.

This plot displays the effect size of the population, environmental variable and their interaction on the phenotypes. The yellow vertical dashed line represents the threshold of medium effect and the red, the threshold of strong effect. "F": females, "M": males, "Temp": temperature, "Conc" refers to the concentration of paraquat and "Diet" to the different media on which longevity was measured.

The strength of this study lies in the diversity of plasticity assays, which provides robust support for the role of TEs in environmentally-induced phenotypic variation. To our knowledge, this is the first study to investigate the contribution of TEs to phenotypic plasticity across such a broad range of phenotypes and environmental conditions in insects, offering empirical evidence that bridges longstanding theoretical predictions with experimental validation. Overall, these findings underscore the role of TEs in shaping organismal responses to environmental changes and highlight the importance of considering global TE composition as a potential driver of environmentally-induced phenotypic variation, particularly in the context of otherwise identical genetic backgrounds.

6. Investigation into TEs reveals candidate insertions

Acknowledging that global TE composition contributes both to phenotypic variation and phenotypic plasticity does not exclude the possibility of phenotypic effects due to discrete TE insertions. Hence, an examination of TEs at the insertion level was performed to detect candidate insertions involved in the occurrence of given phenotypes. To investigate this, several methods were employed to narrow down the list of candidate insertions. First, we filtered TE insertions of frequency ≥60% and unique to each

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

population, selecting insertions likely to have a phenotypic impact at the population level. Second, a genome-wide association analysis was conducted using Baypass (see Material and Methods).

From the first method, a total of 12 putative candidate insertions were identified through manual curation (see Material and Methods) (Table S2.14). Three candidate insertions were due to *Doc* elements, consistent with its observed increase across generations. Notably, three candidate insertions belonged to the *roo* family which is in accordance with Varoqui and colleagues (2025), that have reported this family to be especially active in these populations (Varoqui et al. 2025). Additionally, both *Zam* and *gtwin* families were each represented by one candidate insertion. These insertions are of particular interest, as they are likely recent, potentially fully functional, thus retaining key regulatory or coding components.

From the female Baypass analysis, comprising 843 markers (*i.e.* insertions) and 17 phenotypic values, seven markers came out with median Bayes Factor (BF) values above 14 thereby being strong candidates (see Material and Methods). Of these candidates, three corresponded to a *roo* element, and six to the insertion of a *Doc* element. These are once again rather expected considering the dynamic nature of *roo*, as well as the notable increase in insertion of *Doc* in the present populations. Four candidates were associated with a *Zam* insertion, and one with a *gtwin* (Fabry et al. 2021; Varoqui et al. 2025) (Table S2.15). The complementary manual curation raised the list of candidates to 19 (see Material and Methods), with BF ranging from 9 to 17 (note that the same markers can have different BF values for each phenotype) (Table S2.15 and S2.16). Markers with high frequencies and with association to multiple phenotypes were part of the manually curated selection.

By integrating results from multiple methods, we compiled a comprehensive set of 13 insertions spread across the derived populations (S10, S31, S73 and S100) for further investigation (Table 2.4). In particular, insertions identified by both methods and located near functionally relevant genes represent especially strong candidate loci. This work is still in progress, and a deeper analysis may be required to further refine these candidate insertions.

IV. Conclusions

Understanding the sources of phenotypic variation remains a central question in evolutionary biology, as does elucidating the impact of environmental changes on such variation. While the contribution of TEs to environmental responses is widely studied in plants (Negi et al. 2016; Joly-Lopez et al. 2017; Hirsch and Springer 2017; Ito 2022; Latzel et al. 2023; Pozzi et al. 2025), the field remains comparatively underexplored in metazoans. In insects in particular, TEs have been extensively investigated in the context of adaptive evolution (Casacuberta and González 2013; Gilbert et al. 2021a;

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Cabral-de-Mello and Palacios-Gimenez 2024), yet their role in immediate environmental responses remains largely elusive and mostly theoretical (Capy et al. 2000; Piacentini et al. 2014; Horváth et al. 2017; Lanciano and Mirouze 2018; Marin et al. 2020; Pimpinelli and Piacentini 2020). This gap is not without reason: assessing the contribution of TEs to an organism's response to environmental perturbation requires comparisons between individuals that are genetically identical except for their TE content, a condition that is exceedingly rare, if not unattainable, in natural populations.

We took advantage of previously established *D. melanogaster* populations in which an increase in TE content was confirmed within a stable genetic background, in order to evaluate the contribution of TEs to phenotypes and environmental responses (Barckmann et al. 2018). This study system provides an unprecedented opportunity to assess the role of TEs in shaping phenotypic variation and plasticity. We first demonstrate that variation in TE content, both in terms of load and composition, influences a broad range of phenotypic traits, including fertility-related traits, longevity, and behavioural traits such as climbing ability.

This study also suggests that TE not only drive phenotypic divergence, but that higher TE content broaden phenotypic variation, as observed for the S73 populations. However, while a higher TE load increase phenotypic variability, it appears to come with a trade-off in fitness, as observed with the S73's impaired hatchability and viability. This highlights the duality of TEs, which can be both deleterious and beneficial depending on the environmental context. Such degraded fitness may have been detrimental to the point of selecting for an alternative TE control mechanism during the induction of TE derepression, resulting in a decrease in TE accumulation as observed in S100. While the underlying mechanism remain unresolved, the TE content of the S100 seems to have reached a plateau, as evidenced by its phenotypic profile highly resembling that of G0's. While these findings support the notion that an excessive TE load can be detrimental, it has also been shown that a complete absence of TEs may be disadvantageous. Cranz-Mileva and colleagues (2024) demonstrated that in *Schizosaccharomyces pombe*, strains completely depleted of the long terminal repeat retrotransposon Tf2 exhibited reduced fitness, whereas strains retaining TEs conferred a positive contribution to the host (Cranz-Mileva et al. 2024).

While the genetic background of the populations was not explicitly verified in this study, it is unlikely that a substantial number of other structural variants (SVs) have arisen during their establishment. Even if such variants had occurred, their number would likely remain limited compared to the abundance of TE variants driven by their derepression. As reported by the TE abundance estimation with DeviaTE, these populations show considerable differences in TE load. This corroborates the notion that TE content is the main differentiating factor between the populations. Moreover, given the extensive

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

phenotyping performed here, it is improbable that all the observed trait differences could be explained by rare non-TE SVs. The strength of this study lies in the breadth of the dataset: it is difficult to dismiss the contribution of TEs when so many traits differ between populations that all derive from an initial pair of inbred flies with severely reduced genetic diversity.

Most interestingly, the assessment of the populations' responses to environmental changes revealed that three of the four traits, namely the longevity under different temperature and diet and the oxidative stress resistance, were characterized by diverging genetic variation for plasticity, *i.e.* GxE interaction, thereby positioning TEs as contributors to the plastic response. To our knowledge, such an effect has essentially never been observed under comparably strict experimental conditions in any metazoan. In conclusion, this study broadens our understanding of the functional reach of TEs, and opens the possibility of a new field of study in the mechanistic aspects underlying plasticity.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Population	Region	Chromosome	start	end	TE	Frequency	ChrX correted frequency	Approach	G0 annotation	Reported phenotypes (flybase)	Phenotype in the present study
S10	Intron	chr2R	18047870	18047871	con19_Quasimodo#LTR/Gypsy	60	NA	Manual	5-HT1A	short lived	Longevity
S10	Promoter (2kb)	chr2R	9478907	19478908	BDGP_ZAM#LTR/Gypsy	45.4545	NA	Baypass	Obp57c		Longevity under 25mM paraquat*, dry weight*
S10	Intron	chrX	11594589	11594590	con48_roo#LTR/Pao	80	60	Manual	Ptp10D		
S31	Promoter (2kb)	chr2R	12409450	12409451	con48_roo#LTR/Pao	75	NA	Baypass	CG17048		Metabolic rate at 30°C*
S31	Intron	chr3R	875331	875332	con12_Doc#LINE/Jockey	100	NA	Manual/Baypass	Myo81F		
S31	Intron	chr3R	939426	939427	con12_Doc#LINE/Jockey	90.9091	NA	Manual/Baypass	Myo81F		
S31	Intron	chrX	219907	219908	con12_Doc#LINE/Jockey	80	60	Manual	CG3777		
S31	3'UTR	chrX	4169729	4169730	con48_roo#LTR/Pao	80	60	Manual	CG15570		
S31	Intron	chr4	1014820	1014826	con16_blood#LTR/Gypsy	90	NA	Baypass	PlexA		Development time*
S73	Promoter (2kb)	chr2L	17674467	17674468	con14_F-element#LINE/Jockey	71.4286	NA	Manual	lncRNA:CR43802	involved in spermatogenesis	Hatchability
S73	Promoter (2kb)	chr3R	21186346	21186347	BDGP_ZAM#LTR/Gypsy	66.6667	NA	Manual	CG34377		
S100	Intron	chr3L	20902911	20902912	con2_gtwin#LTR/Gypsy	69.2308	NA	Manual/Baypass	pip		
S100	Intron	chr3R	27577719	27577720	con48_roo#LTR/Pao	64.2857	NA	Manual/Baypass	sima	abnormal stress response	CTmax

Table 2. 4: Summary table of candidate insertions detected through both manual and genome-wide scan methods and their potential phenotypic effects.

*The "Population" column indicate the population concerned by the TE insertion, "Region" the genomic location of the TE insertion, "Chromosome", "Start", "End" the genomic coordinates of the insertion, "TE" the TE identity, "Frequency" the insertion frequency within the population, "ChrX corrected frequency" the adjusted frequency for X-linked insertions, "Approach" the method used for the identification, "G0 annotation" the reference gene annotation in the G0 genome, "Reported phenotypes (FlyBase)" the phenotypes previously associated with the gene in FlyBase, and "Phenotype in the present study" the phenotypes observed in this study that may relate to the insertion. *association detected by Baypass.*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Chapter 2. Caught between TEs and *Wolbachia,* a facultative endosymbiont: investigating the phenotypic effects in *Drosophila melanogaster*

Foreword

As a side note, I would like to emphasize that although the following data were produced at the same time as the data from the previous chapter, we have thus far focused our efforts on analysing the data from the initially gifted flies (populations carrying the endosymbiont *Wolbachia*, that is). The results in Chapter 2 would definitely deserve more effort... This chapter addresses a subject central to the research of both laboratories in which I have pursued my PhD, namely symbiosis, in particular within the Laboratoire de Biologie Fonctionnelle, Insectes et Interactions (BF2I). While I am not myself a specialist in symbiosis, I have been regularly exposed to this topic throughout my PhD, which was one of the motivations for undertaking this research direction. I do apologize in advance if this chapter is a bit of work to chew on! I appreciate your patience, and thank you immensely for the time you will spend reading this!

I. A brief introduction to symbiosis

Among the various factors that can impact an organism's phenotype and its response to the environment, one of considerable importance is symbiosis, a widespread phenomenon that plays a major role in species evolution (Klepzig et al. 2009). Symbiosis describes an interspecies interaction that exists along a continuum, ranging from parasitism to commensalism and finally to mutualism (Margulis 1938; Paracer and Ahmadjian 2000; Moran 2006; Leung and Poulin 2008). Parasitism implies one organism benefits from the interaction at the expense of the other (Dobson et al. 2008; Klepzig et al. 2009; Poulin and Randhawa 2015; Tipton et al. 2019). Commensalism describes an interaction where one organism benefits from it without significantly affecting the other (Leung and Poulin 2008; Mathis and Bronstein 2020), while a mutualistic symbiosis implies that all partners benefit from the interaction (Leung and Poulin 2008; Bronstein 2015; Chacón et al. 2021). A symbiont can navigate through these interactions, such that the same symbiont species may be parasitic in one host and mutualistic in another (Werren et al. 2008; Drew et al. 2021). In addition, such interactions may vary depending on the development of the host and symbiont, and even on the ecological context, in that one beneficial symbiont may become an unwanted partner in a different environment (Ferrari and Vavre 2011). Symbiotic interactions are categorized not only by the nature of the interaction (*i.e.* parasitic, commensalist or mutualist) but also

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

by their integration into the host biology. Primary symbionts are obligatory, meaning that the host is not viable when deprived of it, and their interaction is generally mutualistic (Moran 2006; Su et al. 2013). Secondary symbionts are, as the name may indicate, additional facultative symbionts (Ferrari and Vavre 2011; Su et al. 2013). Multiple secondary symbionts can coexist with a host, and their presence can vary between individuals and populations (Ferrari and Vavre 2011; McLean et al. 2018). Finally, symbionts can be extraorganismal (ectosymbionts), or present within the organism, either extracellularly or intracellularly (endosymbionts).

The combination of the host and its symbionts is sometimes referred to as the holobiont (Margulis and Fester 1991; Rosenberg and Zilber-Rosenberg 2011; 2018; Morris 2018), and a holobiont combined genetic content is referred to as the hologenome (Margulis and Fester 1991; Roughgarden 2023). Evolutionary theories predict that, much like the holobiont, both the individual genomes and the hologenome are subjected to selection (Bordenstein and Theis 2015; Rosenberg and Zilber-Rosenberg 2018; Roughgarden 2023). Hence, variation in the symbiotic composition and symbiont prevalence within the same host can be perceived as a source of hologenotypic variation (Bordenstein and Theis 2015; Roughgarden 2023).

1. Facultative endosymbiosis and contribution to the host's phenotype

While vertical transmission is the primary transmission mode of obligate endosymbionts, facultative endosymbionts are often transmitted both vertically and horizontally, increasing the potential for their spread as well as their loss (Vautrin and Vavre 2009; Bright and Bulgheresi 2010; Sudakaran et al. 2017; Haag 2018). Unlike obligate endosymbionts, which are often confined to bacteriocytes – specialized bacteria-bearing host cells (José Gosalbes et al. 2010; Baumann et al. 2013) –, secondary endosymbionts can be found in a variety of tissues, sometimes even showing a preference for specific tissues over others, a phenomenon known as tissue tropism (López-Madrigal and Duarte 2019). Secondary endosymbiotic interactions are especially interesting to study in that, not only do they provide hologenotypic variation, since their prevalence can vary within and between populations, but also because this variation, as we will see now, can induce intra- and inter-specific phenotypic variation. Secondary endosymbiotic relationships in insects are well studied, and their contribution to hosts' functions has been consistently reported (Ferrari and Vavre 2011; Su et al. 2013).

Facultative endosymbiosis has been shown to impact behaviour, with for example the negative relationship between the presence of endosymbiont *Rickettsia* and the propensity for the money spider, *Erigone atra,* to initiate airborne long distance travelling (Goodacre et al. 2009). The meta-population

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

analysis suggests variable prevalence of *Rickettsia*, suggesting that this endosymbiont-mediated behavioural trait varies within and between populations. The presence of certain facultative endosymbionts has also been associated with insecticide resistance. For instance, in the bean bug *Riptortus pedestris,* mutualistic gut bacteria of the genus *Burkholderia*, of which some strains are known to degrade fenitrothion, a widely used organophosphorus insecticide, provide resistance to the insecticide (Kikuchi et al. 2012). Facultative endosymbionts have also been described to provide insects with additional defence mechanisms. For instance, *Hamiltonella defensa* is a heritable facultative endosymbiont present in around 40% of aphid species and has been demonstrated to increase aphid resistance to parasitoids (Moran 2006; Rothacher et al. 2016; Oliver and Higashi 2019). *H. defensa* prevalence also varies within aphid species and it has been demonstrated that this could confer variable levels of protection (Vorburger and Rouchet 2016; Oliver and Higashi 2019). These examples demonstrate that symbiotic relationships can influence the host's response to environmental stressors, and that variation in infection status can lead to differences in host phenotype and its ability to cope with environmental challenges (Bordenstein and Theis 2015). Therefore, a single host genotype may give rise to distinct phenotypes through its hologenome (Bordenstein and Theis 2015). This supports the idea that symbiosis not only promotes phenotypic variation, but may also shape the responses of an organism to its environment.

2. Wolbachia in D. melanogaster

Wolbachia pipientis is a gram-negative Alphaproteobacteria and a common endosymbiont in arthropods and nematodes (Werren et al. 2008). Notably, it is estimated to infect at least 65% of insect species (Hilgenboecker et al. 2008; Werren et al. 2008). *Wolbachia* is vertically transmitted through the infected female germline (Werren et al. 2008). The nature of *Wolbachia*'s endosymbiosis varies greatly depending on both the bacterial strain and the host species (Fry and Rand 2002; Chrostek et al. 2013; Gruntenko et al. 2017; Martinez et al. 2017; Strunov et al. 2022). On one hand, *Wolbachia* can behave parasitically, manipulating host reproduction through mechanisms such as parthenogenesis, male killing and feminization (Landmann 2019). It is also known to induce cytoplasmic incompatibility (CI), as in *D. melanogaster*, where infected males can only successfully reproduce with infected females, thus promoting its own spread (Merçot and Charlat 2004; Landmann 2019). On the other end of the symbiotic spectrum, *Wolbachia* functions as an obligate mutualist in the bedbug *Cimex lectularius*, providing both biotin and riboflavin B vitamins necessary for moulting and successful reproduction (Hosokawa et al. 2010; Poulain et al. 2024). In *D. melanogaster*, apart from the previously cited CI, *Wolbachia* has also been

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

described to have various other influences: positive impacts on fecundity have been reported, although sometimes associated with a cost of reduced lifespan (Fry et al. 2004; Serga et al. 2021); an increase in recombination rate (Singh 2019; Bryant and Newton 2020; Mostoufi and Singh 2022); a reduction of the gut microbiome diversity (Ye et al. 2017); and even affecting the host thermal preferences, in favour of lower temperatures when the flies are infected (Truitt et al. 2019). And finally, a well described effect of *Wolbachia* in *D. melanogaster* is its ability to provide the host with protection against RNA viruses (Pimentel et al. 2021; Yu and Iatsenko 2025). This has been shown for commonly used laboratory viruses such as *Drosophila C virus* (DCV), a natural pathogen of *Drosophila*, and *Flock House virus* (FHV), initially isolated from a species of beetle (Asselin and Johnson 2025). More recently, this protection has also been demonstrated against viruses isolated from wild *D. melanogaster* populations, and even against the Zika virus, a human pathogen transmitted by mosquitos, demonstrating the versatility of the *Wolbachia*-mediated viral protection (Bruner-Montero and Jiggins 2023; Karavioti 2024). However, this protection sometimes appears to come at a cost of shortened lifespan (Chrostek et al. 2013).

In conclusion, *Wolbachia* adds a layer of phenotypic variation that may vary within populations, and contribute to individual differentiation beyond their genomes.

3. Wolbachia's interaction with TEs

Not much is known on the interaction between *Wolbachia* and TEs, and here we describe the few studies discussing the subject. As many TEs share relatedness to RNA viruses, it may not come as a surprise that *Wolbachia*, although drastically understudied in this context, has been reported to interact with TEs (Xiong and Eickbush 1990; Eickbush and Malik 2007; Mustafin 2018). It is plausible that *Wolbachia* can modulate TE activity, potentially providing a protective effect by influencing their expression and mobilization. Indeed, TE expression can be significantly impacted by the presence of *Wolbachia*. A study from 2023 by Eugénio and colleagues measured the expression of 14 TE families (of both retrotransposons and DNA transposons) in whole-bodies of *D. melanogaster* females of 25 different genotypes (Eugénio et al. 2023). Although no consistent pattern was observed, every studied TE family showed at least one occurrence of change in expression upon *Wolbachia* removal in at least one of the genomes (Eugénio et al. 2023). Some genotypes exhibited a unidirectional trend in the effect of *Wolbachia* on TE expression. For instance, the RAL-21 genotype consistently showed an increase in TE expression upon removal of the endosymbiont, whereas RAL-181 appeared to display the opposite trend (Eugénio et al. 2023). In a more recent study, the presence of *Wolbachia* was associated with a general decrease in TE expression, with approximately 61% of differentially expressed TEs showing a reduction in expression (Mostoufi and Singh

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

2024). Many of these differentially expressed TEs were LTR elements, the most prevalent class in *D. melanogaster*, such as *Copia* and *Ty3* (Mostoufi and Singh 2024). Other TE families were also displaying a similar trend, although not reflected at the class level, such as *TcMar-Pogo* or *Helitrons* (Mostoufi and Singh 2024). These studies highlight the correlation between TE expression and infection by *Wolbachia,* however, the underlying mechanisms remain to be investigated. Notably, it remains to be seen if the impact of *Wolbachia* on TE expression, and potential activity, is a direct or indirect effect.

One common aspect of *Wolbachia* and of some TE families, specifically endogenous retroviruses (EVRs) producing virus like particles (VLPs), is their inheritance by maternal transmission (Chalvet et al. 1999; Brasset et al. 2006; Touret et al. 2014). It was theorized that *Wolbachia* may have played a part in the invasion of *D. melanogaster* by the *P-element*, considering the temporal and spatial correlation with the recent *Wolbachia* strain replacement (Riegler et al. 2005). Indeed, Riegler and collaborators, while analysing the genomes of *Wolbachia* from *D. melanogaster* lines maintained in laboratories, representing both historical and more recent natural populations, detected the complete replacement of one *Wolbachia* strain by another (Riegler et al. 2005). This constitutes one of the few documented cases of natural endosymbiont replacement, a phenomenon now believed to be more common than initially assumed, and reflective of the dynamic nature of symbiotic relationships (Riegler et al. 2005; Sudakaran et al. 2017). In line with this theory, Touret and collaborators unveiled that *Wolbachia* changes the pattern of *mdg4* transmission from follicular cells to the germline, without affecting the TE's expression level (Touret et al. 2014). The authors proposed that since *Wolbachia* and *mdg4* employ similar germline invasion systems, *Wolbachia* could impact maternal transmission of TEs, and in that case, that of *mdg4* (Touret et al. 2014). Additionally, the authors found that *Wolbachia* significantly reduced the rate of insertions of *mdg4* in the *ovo* gene, which is usually a hotspot for *mdg4* insertions (Dej et al. 1998; Touret et al. 2014). Lastly, despite the following example not being reported in *D. melanogaster* but instead in *Aedes aegypti* mosquito cell lines, it was found that *Wolbachia* affects the abundance of both miRNAs and piRNAs (Mayoral et al. 2014). It may only be a matter of time before similar patterns are observed in *D. melanogaster*, and considering the crucial role of piRNAs in TE regulation, this interaction could be highly relevant.

Altogether, the potential source of genetic diversity provided by the hologenome, and the impact of *Wolbachia* on both phenotypes and TE activity begs the question whether *Wolbachia* adds an additional layer to our understanding of TEs' contribution to phenotypic variability and plasticity. In our previous work, we reported that populations with increasing TE content exhibited differences in phenotype and variable responses to environment changes, thereby highlighting the role of TEs in phenotypic plasticity

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

(Larue et al. in prep.). These populations were established through successive inducible *piwi* knockdowns in the somatic ovarian follicle cells, which play a central role in regulating TEs via the piRNA pathway (Barckmann et al. 2018; Czech et al. 2018). When the piRNA pathway is non-functional, VLPs derived from ERVs may invade the germline and accumulate across generations (Chalvet et al. 1999; Brasset et al. 2006). The populations used in that study were all reported carrying the endosymbiont *Wolbachia*. Since *Wolbachia* has been shown to influence both phenotypes and TE expression and localization, we sought to investigate whether *Wolbachia* may be taking part in the response to environmental changes in interaction with TE content, and thus complement the results described in Chapter 1 (Larue et al. in prep.). For that, we have taken advantage of previous tetracycline-treated subpopulations (Mayeux et al. 2024) derived from the five populations with different TE content (G0, S10, S31, S73 and S100) (Barckmann et al. 2018; Varoqui et al. 2025; Larue et al. in prep.). While the methods applied post-treatment are identical to those used in Chapter 1, following the removal of *Wolbachia* from the populations, the resulting data have only been recently analysed. Therefore, the chapter that follows presents very preliminary findings, which would undoubtedly benefit from deeper analysis.

II. Results and discussion

1. Wolbachia density does not differ between populations

Before investigating the influence of the presence of *Wolbachia* on the phenotypes and their response to environmental changes, it was necessary to evaluate its prevalence in the previously studied populations (Larue et al. in prep.). Therefore, we quantified *Wolbachia* load in the G0, S10*,* S31*,* S73 and S100 populations by quantitive PCR in whole individual flies. The relative density of *Wolbachia* varies according to the sex (p-value = 0.00026). In *Drosophila simulans*, a difference in *Wolbachia* density in gonadal tissues of females and males was previously reported in agreement with our observation (Correa and Ballard 2012). However, no significant differences are observed between populations (p-value = 0.16198) (Figure 3.1). This suggests that the observations made for these populations are unlikely to be the result of varying *Wolbachia* quantities (Larue et al. in prep.). Moreover, among the eight replicates per condition, we observed some variation in *Wolbachia* presence. For example, one male in the G0 and S10 populations (M3 and M4, respectively) appears not to carry the endosymbiont, as well as one female in the S73 population (F2) (Figure 3.1). Within-population variation in infection status may arise from the incomplete cytoplasmic imcompatiblity (CI) observed in *D. melanogaster* compared to its sister species *D. simulans*. Indeed, some embryos can evade the *Wolbachia*-induced mortality following crosses between

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

infected males and uninfected females (as reviewed in (Merçot and Charlat 2004). Variation in *Wolbachia* prevalence was also previously reported for natural population of both *D. melanogaster* and *D. simulans* (Verspoor and Haddrill 2011). This highlights the variability in infection status further emphasizing the importance of assessing the potential influence of the endosymbiont in the phenotypes.

Eugénio et al. (2023) reported that distinct TE families were differentially expressed depending on the presence or absence of the endosymbiont, and that these relationships were further influenced by the strain genetic background (Eugénio et al. 2023). While the present populations were established in such a way that the genetic background is nearly identical (Barckmann et al. 2018), TE content may serve as genetic variants through which *Wolbachia* can influence phenotypes. We obtained populations devoid of *Wolbachia* by tetracycline treatment and monitored the absence of the endosymbiont by qPCR (see Material and Methods). The establishment of these *Wb+* populations (presence of *Wolbachia*) and *Wb-* populations (absence of *Wolbachia*) enables us to investigate the effect of the possible interplay between TE content and *Wolbachia* on phenotypes.

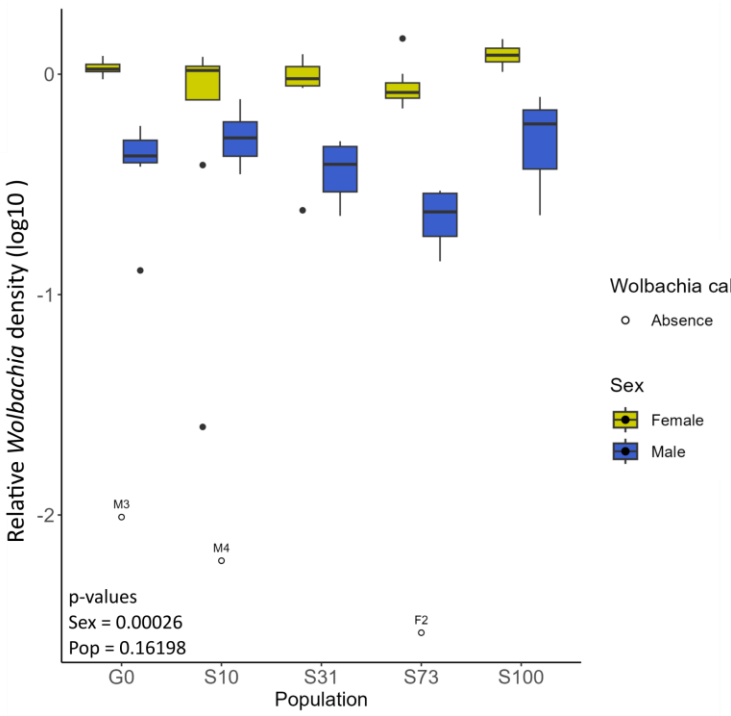


Figure 3. 1: Relative* Wolbachia *density in the Wb+ population.
The relative density is expressed as the log10 of the ratio computed as indicated by Pffafl (2001) (see Material and Methods). Each condition is represented by eight individual flies. The flies devoid of Wolbachia *are indicated by an empty circle and tagged with the sex and replicate number (for instance, M3 corresponds to the third male fly assessed). The statistical analysis was performed using a two-way ANOVA test.*

2. TE content changes following Wolbachia *removal*

The TE genotyping of the *Wb-* populations was performed using TrEMOLO on pool sequencing of 200 embryos, following the same protocol as previously described to ensure comparability (Larue et al. in

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

prep.). As in the latter study, TE content was addressed using the G0-*Wb+* assembled genome as the reference. While we reported a gradual increase in new TE insertions from S10-*Wb+* to S100-*Wb+* compared to G0-*Wb*+, the pattern in the tetracycline-treated populations was less straightforward (Figure 3.2A) (Larue et al. in prep.). Specifically, the total number of new TE insertions increased by 458 in S10-*Wb*-, followed by 484, 477, and 548 new insertions in S31-*Wb*-, S73-*Wb*-, and S100-*Wb*-, respectively, relative to G0-*Wb+* (for all TE results refer to Figures S3.1, S3.2 and S3.3). Thus, in the present study, the populations harbour slightly higher overall TE content than the *Wolbachia*-infected original populations except for S73-*Wb*- (Figure 3.2A).

At the TE subclass level, LINE elements gradually increase from S10-*Wb*- to S100-*Wb*- reaching 205 new insertions (Figure 3.2A). Notably, the *Jockey* superfamilies display a global increase across the populations (Figure 3.2E). Furthermore, within the *Jockey* superfamily, both *F-element* and *Doc* show a gradual accumulation of insertions with the generations as reported in the *Wb+* populations (Larue et al. in prep.) (Figure S3.2). *Doc* insertion numbers are stable between S10-*Wb*- and S31-*Wb*- with 42 and 43 additional insertions compared to G0-*Wb+* respectively, and increase to 103 new insertions in S100-*Wb*- (Figure S3.2).

As for LTR retrotransposons, they increase from 301 in S10-*Wb*- with 301 to 314 in S31-*Wb*- and S100-*Wb*- (Figure 3.2A). Nevertheless, in S73-*Wb*-, new LTR elements detected are less abundant, with only 189 new insertions. The *mdg4* (formerly *gypsy*, as indicated in Figure 3.2E) show a global increase in new insertions across the populations. Specifically, both *Zam* and *gtwin* previously reported to show high transposition activity exhibit significant accumulation, consistent with findings in *Wb+* populations (Barckmann et al. 2018; Varoqui et al. 2025; Larue et al. in prep.). Indeed, *Zam* accounts for three additional insertions in the S10-*Wb*-, and reaches 17 new insertions in both S31-*Wb*- and S73-*Wb*-, and 22 in S100-*Wb*- (Figure 3.2B). Interestingly, the number of new *Zam* insertions detected in S73-*Wb*- seems to reflect the global TE accumulation pattern and further testify of its singularity compared to S73-*Wb+* (Figure 3.5B). As for *gtwin*, it gradually increases with the generations reaching 17 new insertions in the S100-*Wb*- (Figure 3.5C).

Insertion frequencies for all populations were obtained from TrEMOLO and range from 13.33% to 15.42%, with no significant differences between the populations (p-value = 0.7381) (Figure 3.2D). This suggests considerable within-population insertional polymorphism, similar to *Wb+* flies, although with slightly lower average frequencies, with *Wb+* populations' frequencies ranging from 13% to 17%. In agreement with such low frequencies, the shared number of insertions is relatively low between the *Wb*- populations, with only 150 insertions shared out of 1,967 (Figure 3.2F). Furthermore, the number of

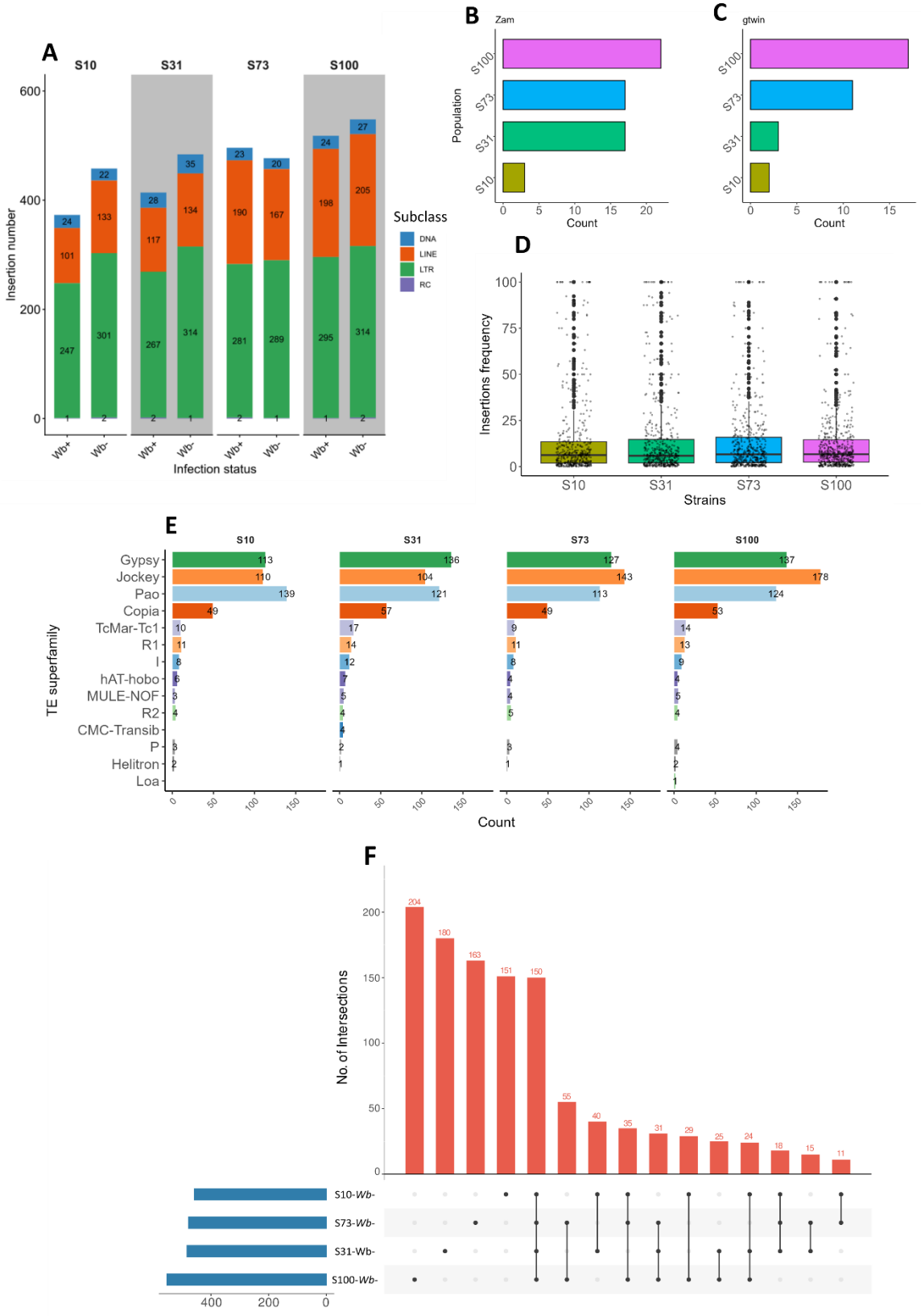
Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

shared insertions between one *Wb-* population and its *Wb+* counterpart is around 60%, suggesting that around 40% of their insertions were absent, or rather undetected, from the other dataset (Figure S3.4). This observation suggests that nearly half of the insertions in the *Wb-* populations likely originate from very low frequency insertions in the *Wb+* populations, which may explain the slightly lower average TE frequencies observed in *Wb-*. To further corroborate these results, an analysis of TE abundance with an approach using normalized coverage with DeviaTE was performed (Weilguny and Kofler 2019). S10-*Wb-* is characterized by a slightly higher TE abundance than its *Wb+* related population. On the other hand, S31-*Wb-*, S73-*Wb-* and S100-*Wb-* have lower abundances than their *Wb+* counterparts (Figure S3.4). However, where TrEMOLO indicates an increase for S31-*Wb-*, DeviaTE detects a decrease compared to S31-*Wb+.* Most interestingly, DeviaTE abundance estimates still position S73-*Wb-* as the population with the highest number of insertions with 1389, while S10-*Wb-*, S31-*Wb-* and S100-*Wb-* have respectively, 1326, 1214 and 1183 (Table S3.1).

Finally, while for the *Wb+* populations, the number of new TE insertions uniquely detected by TrEMOLO in a specific population followed the generations of inducible *piwi* knockdown, in the *Wb-* populations, the pattern is not as linear (Figure 3.2F). This is mostly due to the higher number of insertions detected in the S10-*Wb-* and S31-*Wb-* compared to their *Wb+* counterparts, which results in a distinct TE accumulation generational timeline. The antibacterial treatment may have acted as a selective force, potentially shaping the observed TE landscape. However, this hypothesis remains to be explored.

Although the expected increase across generations is not as marked as for the *Wb+* populations, we still see a similar increase for the TE families previously described to have accumulated in these populations (*i.e. Zam* and *gtwin*). Therefore, this system still allows us to understand the impact of TE load on phenotypes, replicating the previous experiments, but also testing if the presence of *Wolbachia* might have an influence on the TE-by-environment interaction reported we previously reported in Chapter 1 (Larue et al. in prep.).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés



Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

***Figure 3. 2: TE composition of the new insertions compared to G0-Wb+ in the populations studied after tetracycline treatment against the endosymbiont* Wolbachia.**
A) TE composition across the different Wb+ and Wb- populations. This plot illustrates the number of insertions detected by TrEMOLO by subclass of TEs. The number in each of the colour categories corresponds to the number of insertions for the given class. Distribution of the newly detected Zam (B) and gtwin (C) copies within the populations. D) Insertion frequencies computed by TrEMOLO. All the insertions detected are displayed for each of the populations. E) TE composition at the superfamily level. The numbers depicted indicate the insertions detected by TrEMOLO. Gypsy and Jockey gradually increase from S10-Wb- to S100-Wb-. F) Distribution of shared and unique insertions across Wb- populations. The intersection was performed on a 100 bp window (see Material and Methods). The number of unique insertions in each population follows the same general trend than observed for the total number of insertions in each population.

3. The populations devoid of Wolbachia *display distinct phenotypic profiles*

Phenotypic assays identical to those conducted on *Wb+* populations were performed after tetracycline treatment (Larue et al. in prep.). The rate of hatchability (also known as embryonic survival (Flatt 2020)) shows a significant population effect (p-value = 0.0121), however not supported by any significant pairwise comparisons, likely caused by the considerable spread of the data for G0-*Wb-*, S10-*Wb-*, S31-*Wb-* as well as S100-*Wb-* (Figure 3.3A). On the other hand, the viability, a measure of egg-to-adult survival (Flatt 2020), is significantly lower in S73-*Wb-* than S10-*Wb-* (p-value = 0.0259) (Figure 3.3B).

As for the locomotion data, two noticeable results are to be highlighted. For one, the S73-*Wb-* probability to reach the top of the tube at the end of the trial is significantly reduced compared to G0-*Wb-* (p-value = 0.0008) (Figure 3.3C). Conversely, S31-*Wb-*'s probability is significantly higher than S10-*Wb-*, S73-*Wb-* and S100-*Wb-* (p-values = 0.0001, <0.0001 and <0.0001, respectively) (Figure 3.3C).

As for the longevity assays, across both temperature conditions (18°C and 22°C) and sexes, only the males at 22°C show no effect of the population on the lifespan from the linear model analysis (Table 3.1). The Kaplan-Meier analysis reveals that pairwise comparisons of the survival curves are only significantly different at 18°C for females and males S31-*Wb-* compared to S100-*Wb-* (p-values = 0.0021, 0.042, respectively) (Figure 3.3D and 3.3E).

Finally, the metabolic rate follows a consistent trend across all experimental conditions, with the population factor showing no significant effect (Table 3.1). The temperature variable on the other hand drastically influences the metabolic rate of both males and females (Figure 3.3F and 3.3G). However, similar to what has been previously observed in *Wb+* populations, no significant interaction between the environment and the population is observed (Figure 3.3F and 3.3G) (Larue et al. in prep.). These results

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

further support the view that variation in TE content is not reflected in this general measure of metabolic performance.

Principal component analyses were conducted to explore the phenotypic profiling of the *Wb*- populations. The females' PCA cluster G0-*Wb*-, S10-*Wb*- and S31-*Wb*- together and separates S73-*Wb*- and S100-*Wb*- (Figure 3.4A). These results are contrasting to what we previously observed, wherein G0-*Wb+*, S10-*Wb+* and S100-*Wb+* would cluster excluding S31-*Wb+* and S73-*Wb+* (Larue et al. in prep.). This may be due to the different number of traits measured between the studies. Indeed, for females, 17 traits were assessed in the *Wb+* populations, whereas in the present study only 10 traits were evaluated. Moreover, some of the traits contributing to the unique phenotypes of S31-*Wb+* females were not measured in the *Wb*- populations, such as longevity on different media and under varying paraquat concentrations (Larue et al. in prep.). As for the males, no clustering pattern is observed in the present study in agreement with our previous observations (Figure 3.4B) (Larue et al. in prep.). Interestingly, the *Wb*- males' PCA suggests that G0-*Wb*- express a phenotypic profile averaging that of the other populations (Figure 3.4B).

Globally, most of the traits measured appear to be influenced by the population variable and hence by their TE content except for the metabolic rate (Table 3.1). While this may not exactly reflect the increase in TE load, due to the loss of a clear pattern of accumulation, it still reflects the variation in TE composition.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

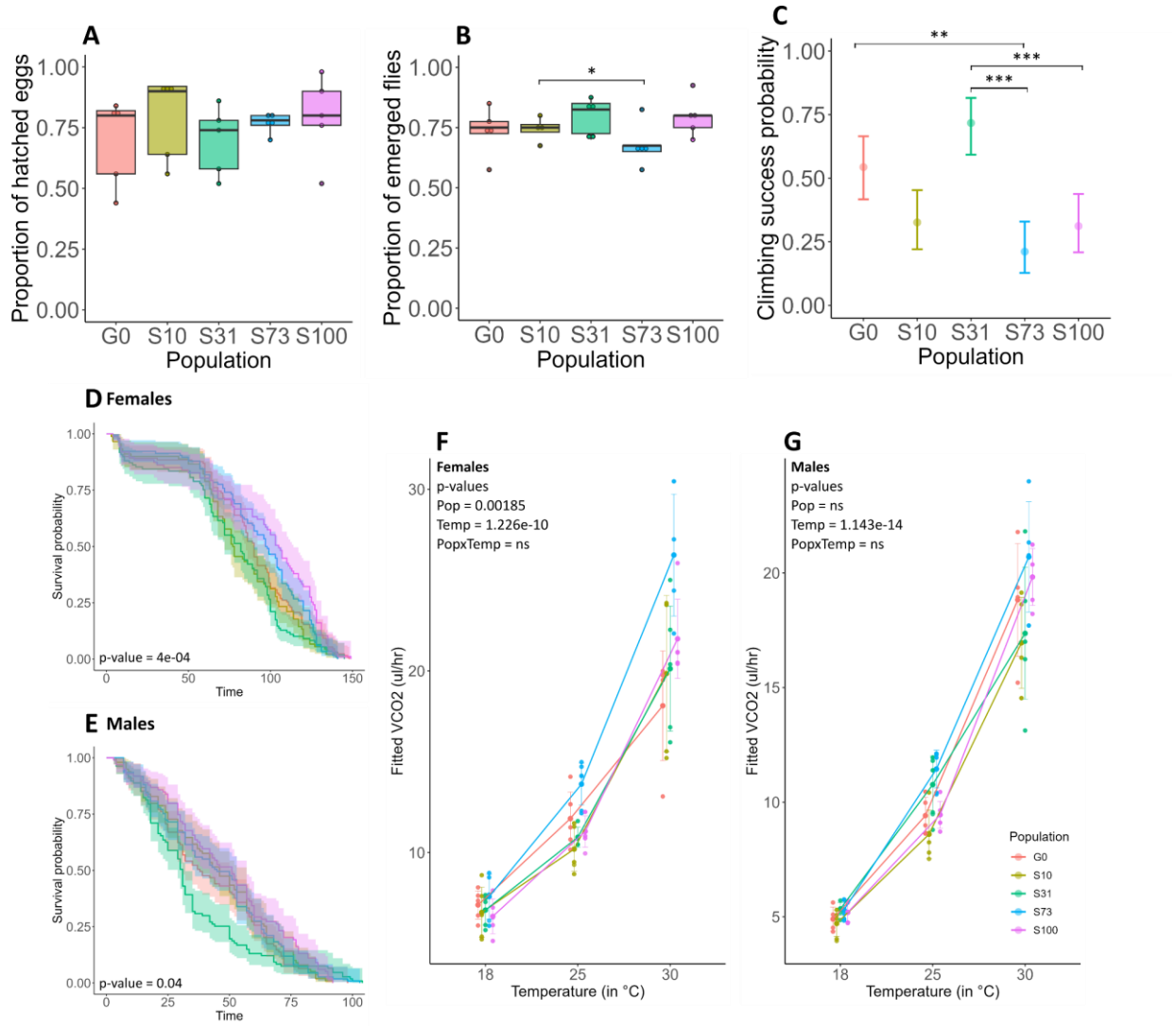


Figure 3. 3: Phenotypes of the populations devoid of* Wolbachia *with varying TE composition.
A) Hatchability of the Wb- populations measured as the proportion of eggs hatched into larvae. B) Viability of the Wb- populations expressed as the proportion of egg-to-adult survival. For both A) and B), each dot represents one of the five replicates per conditions and are expressed in proportion. C) Climbing success probability for the males Wb-. Upper and lower confidence intervals are depicted around the estimated marginal means. Longevity at 18°C displayed as survival curves with confidence intervals for the females D) and males E). Reaction norm plot of the fitted metabolic rate measurements for the females (F) and for the males (G). Each dot represents the fitted estimate of the replicate across multiple repeated measurements. Five replicates of pools of five flies were conducted for each of the experimental conditions. Dry weight and activity were accounted for in the models. In both females and males, a significant influence of the temperature on the metabolic rate is observed, however only for the females does the population variable show a significant impact.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Phenotype	Sex	Environment	Pop effect	Env effect	EnvxPop effect	Statistical test
Hatchability	Mix	NA	0.0121	NA	NA	GLMER *binomial*
Viability	Mix	NA	0.01196	NA	NA	GLMER *binomial*
Dry Weight	F	NA	0.0001	NA	NA	ANOVA
	M	NA	2.23e-07	NA	NA	
Locomotion	F	NA	ns	NA	NA	GLMER *binomial*
	M	NA	1.02e-08	NA	NA	
Longevity	F	22°C	0.0075	< 2.2e-16	0.0013	LMER
		18°C	0.0046			
	M	22°C	ns	4.93e-10	ns	
		18°C	0.0010			
Metabolic rate	F	18°C	ns	1.16e-10	ns	LMER
		25°C	ns			
		30°C	ns			
	M	18°C	ns	1.03e-14	ns	
		25°C	ns			
		30°C	ns			

Table 3. 1: Summary table of the effects of population, environment, and their interaction on the measured phenotypes.

The p-values were obtained by sequentially complexifying the model with the environmental variable when applicable. However, the p-values within an environmental condition were obtained by subsampling the target data. Pop: population, Env: environment. The "Phenotype" column indicate, the measured trait, "Sex" whether measured in females (F), males (M), or if the sex was undetermined (Mix), "Environment" the experimental condition under which the trait was measured, "Pop effect" the statistical significance of the population effect, "Env effect" the statistical significance of the environmental variable, "EnvxPop effect" the statistical significance of the population-by-environment interaction and "Statistical test" the test used for the analysis. NA: not applicable, ns: non-significant.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

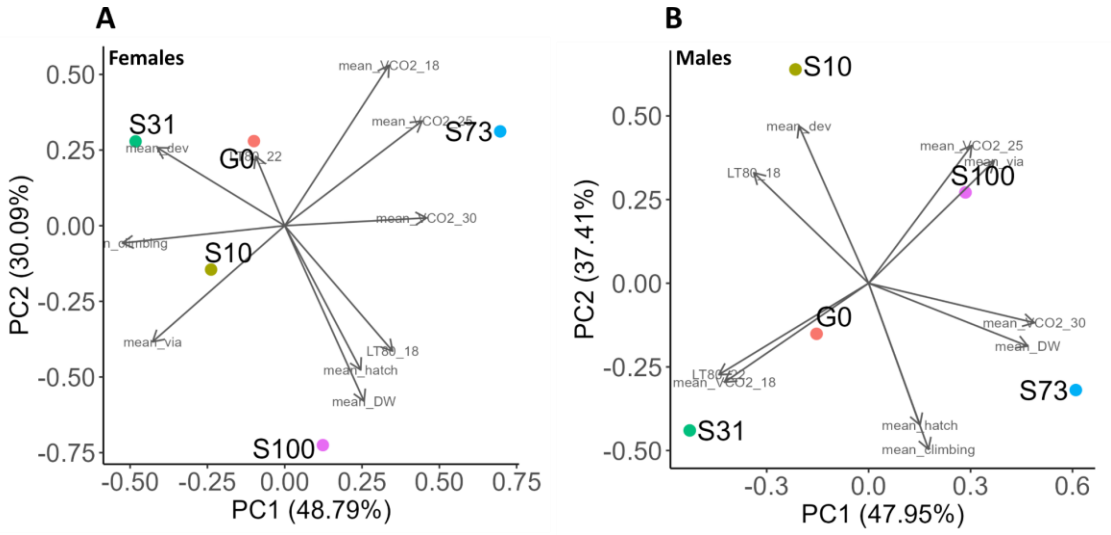


***Figure 3. 4: Phenotypic profiles of the populations devoid of* Wolbachia.**
PCA analysis of the phenotypes for the females (A) and males (B). Each phenotype is expressed either as its mean or as its LT80 estimate. VCO2 refers to the metabolic rate measurements, "hatch" to the hatchability, "via", the viability, "dev", the development time, "climbing", the negative geotaxis assay, "DW" is the dry weight, and "18", "22", "25", "30" are the environmental temperature.

4. The impact of Wolbachia on the phenotypes of populations with variable TE content

To investigate the interplay between *Wolbachia* and TE content, and their impact on phenotypes, we compared the data from the *Wb-* populations with the previously analysed *Wb+* ones (Larue et al. in prep.) (Table 3.2). First, the comparison of the hatchability rate under variable *Wolbachia* infection status was assessed across the populations. Our data revealed a significant interaction between population and infection status (p-value = 0.001157), indicating that *Wolbachia* affects hatchability depending on the population. The hatchability of G0-*Wb-* is significantly decreased following *Wolbachia* removal (p-value = 0.001) (Figure 3.5A). Previous studies have shown that *Wolbachia* can enhance fertility (in the form of fecundity), which aligns with our observation in G0-*Wb-* (Fry et al. 2004; Serga et al. 2021). Curiously, S73-*Wb-* shows the opposite trend, with hatchability significantly increasing in the absence of *Wolbachia* (p-value = 0.0283) (Figure 3.5A). S73-*Wb+* is heavily TE-burdened compared to S73-*Wb-*, supported by a reduced DeviaTE estimated TE abundance from 1466 in S73-*Wb+* to 1389 in S73-*Wb-* (Table S3.1). This reduction in TE burden could explain the recovery in hatchability observed in S73-*Wb-,* rather than the removal of *Wolbachia*. This would further support that S73-*Wb+*'s hatchability is indeed strongly impacted by the global TE content as advanced in Chapter 1 (Larue et al. in prep.).

As for viability, no differences are detected upon *Wolbachia* removal (Figure 3.5B). This may suggest that, since both S73-*Wb-* and S73-*Wb+* display reduced viability – even though TE content

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

decreased considerably upon *Wolbachia* removal – viability is not influenced by overall TE content but rather by specific insertions with targeted effects that may have been carried into the S73-*Wb*- population after treatment. Of the approximately 60% of shared insertions, those with higher insertion frequencies may serve as promising candidates for investigation in future studies (Figure S3.4).

Locomotion measurements, in the form of negative geotaxis assays, were also conducted and compared between infection status. The probability to have reached the top of the assay tube by the end of the trial shows no difference upon removal of *Wolbachia* for the females (Figure 3.5C). However, males' data suggest an increase in their natural behaviour to climb up the tube for population S31 upon *Wolbachia* removal (p-value = 0.0004) (Figure 3.5D). This is supported by a significant interaction between population and infection status (p-value = 0.006374). It was previously reported that *Wolbachia* can decrease olfactory-cued locomotion in *D. melanogaster* (Peng et al. 2008). However, in our experimental setup, all odour cues were controlled to avoid any bias in the results, making it difficult to directly compare our findings with those of the literature. This is an interesting result given that studies on the effect of *Wolbachia* on locomotion mainly report an increase in locomotion when infected (Fleury et al. 2000; Evans et al. 2009; Peng and Wang 2009; Bi and Wang 2020). Although we are not yet able to explain this observation, it nonetheless hints at the idea that *Wolbachia* may impact some phenotypes, and more specifically, in a TE-content-dependent manner.

Phenotype	**Sex**	**Environment**	**Pop effect**		**Wolbachia effect**	**Statistical test**
			Wb+	**Wb-**		
Hatchability	Mix	NA	4.50E-07	0.0121	**0.001157**	GLMER *binomial*
Viability	Mix	NA	0.0002	0.01196	ns	GLMER *binomial*
Development time	Mix	NA	4.00E-16	<2e-16		Log-rank test
Dry Weight	F	NA	0.0058	0.0001	**0.0073432**	ANOVA
	M	NA	0.0081	2.23E-07	ns	
Locomotion	F	NA	ns	ns	ns	GLMER *binomial*
	M	NA	0.0295	1.02E-08	**0.006374**	
Longevity	F	22°C	0.0001	0.0075	ns	LMER

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

		18°C	0.0033	0.0046	**0.00503**	
	M	22°C	**< 2.2E-16**	**ns**	**2.97E-07**	
		18°C	3.80E-06	0.001	**0.01471**	
Metabolic rate	F	18°C	ns	ns	ns	LMER
		25°C	ns	ns	ns	
		30°C	**0.0302**	**ns**	ns	
	M	18°C	ns	ns	**0.006579**	
		25°C	**0.0378**	**ns**	**0.046**	
		30°C	ns	ns	ns	

Table 3. 2: Summary table of the effects of population (tested separately for each* Wolbachia *infection status) and Wolbachia infection on the measured phenotypes.

*The "Phenotype" column refers to the measured trait, "Sex" for whether the females (F), males (M), or undetermined sex (Mix) were tested, "Environment" for the experimental condition under which the trait was measured, "Pop effect" to the statistical significance of the population variable, "*Wolbachia *effect" to the statistical significance of the effect of* Wolbachia *infection, shown separately for Wb+ and Wb- populations and "Statistical test" to the test applied for analysis. Reported values are p-values from the corresponding statistical tests (as indicated in the last column). Models were adjusted to include* Wolbachia *infection as an explanatory variable within each sex. NA: not applicable, ns: non-significant.*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

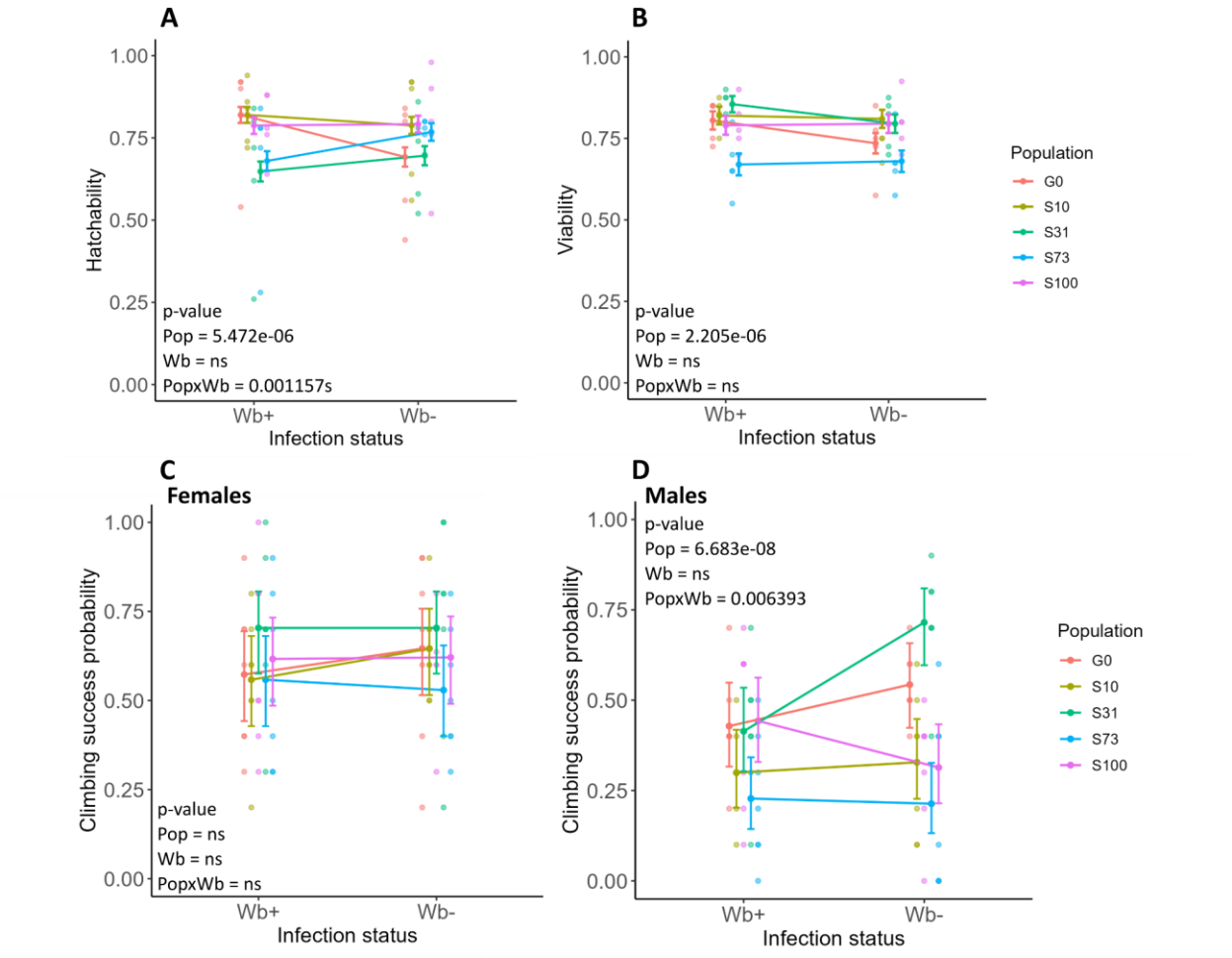


Figure 3. 5: Reaction norms of the phenotypic measurement under variable* Wolbachia *infection status. *A) and B) represent the reaction norms of the hatchability and viability, respectively. C) and D) show the reaction norms of the probability for the flies to have climbed all the way up the tube at the end of the trial, upon* Wolbachia *removal, for the females and males respectively. In every panel the dots represent the raw data while the confidence intervals are the fitted data. These datasets were analysed using a GLMER model with binomial distribution. Pop: population, Wb: Wolbachia.*

5. Sexual dimorphism in longevity and dry weight is amplified by Wolbachia *infection status*

We previously report that the lower dry weight observed in G0-*Wb+* and S10-*Wb+* males could underlie their shorter lifespans observed both at 18°C and 22°C. Both G0-*Wb*- and S10-*Wb*- males also exhibited significantly lower dry weight compared to S73-*Wb*- (p-values = 0.0041 and 0.001, respectively), and S10-*Wb*- to S100-*Wb*- (p-value = 0.0317) (Figure 3.6B) (Larue et al. in prep.). *Wolbachia* removal had no observable effect on the dry weight of males from any population. However, both G0-*Wb*- and S10-*Wb*- males, despite maintaining lower dry weight, displayed lifespans comparable to the other populations, a contrasting result from G0-*Wb+* and S10-*Wb+* (Figure 3.6C, Table S3.2). The fact that males'

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

longevity is rescued following *Wolbachia* removal without any corresponding change in dry weight suggests that, rather than the dry weight being the causal factor for reduced longevity in G0-*Wb+* and S10-*Wb+*, it may in fact be the presence of *Wolbachia* itself that contributes to their shortened lifespan (Figure 3.6C). It was previously reported that uninfected males live longer than infected ones both at 24°C and 28°C (Fry and Rand 2002; Werren et al. 2008; Strunov et al. 2022). The observation that *Wolbachia* infection appears to affect only the longevity of G0-*Wb*- and S10-*Wb*- males, populations with the lowest TE contents, is supported by a significant Population-by-*Wolbachia* interaction (p-value = 2.394e-07) (Table 3.3). However, these results remain puzzling despite supporting the idea that TE content and *Wolbachia* may interact in shaping this phenotype.

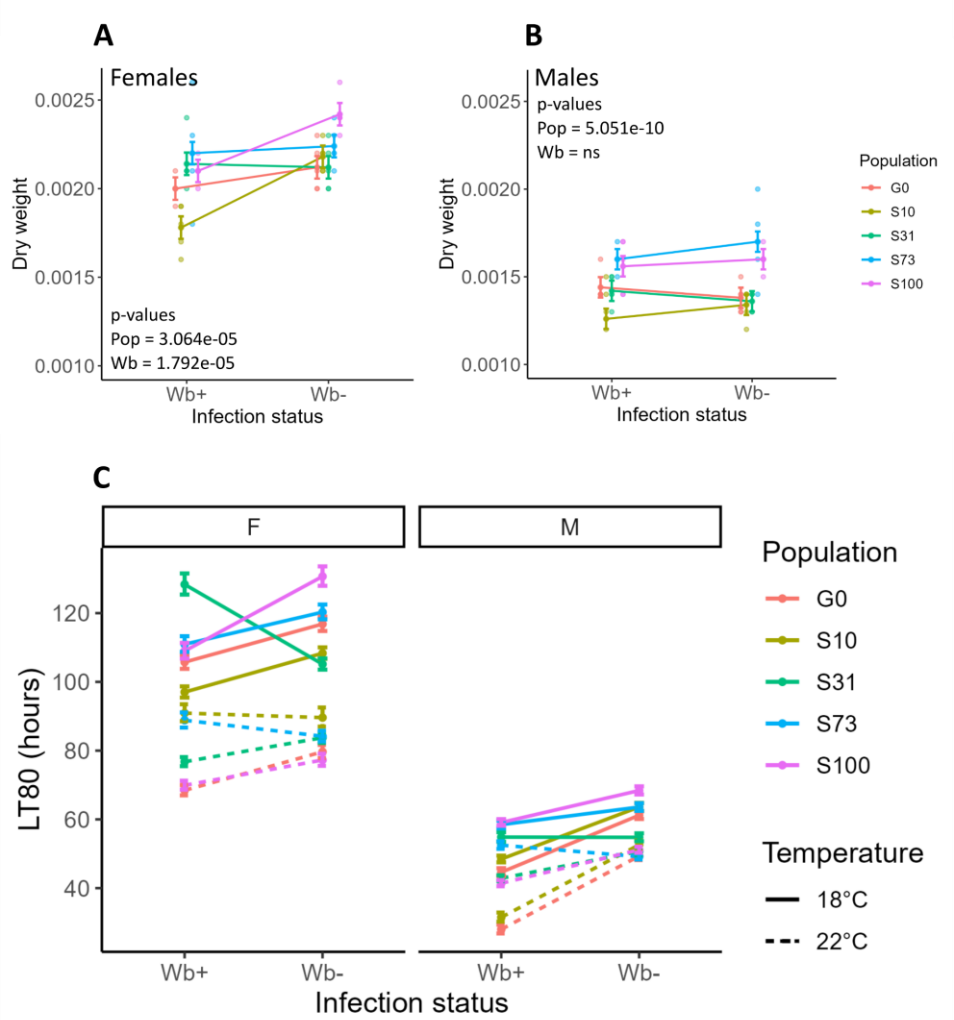


Figure 3. 6: Reaction norm of the dry weight and LT80 according to the* Wolbachia *infection status.
A) The reaction norm of the dry weight in females (A) and males (B), upon removal of Wolbachia*. Upper and lower confidence intervals represent the fitted data, while the dots represent the raw data (in grams). C) The reaction norm of the LT80 depending on the infection status. Data for each temperature is*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

represented with full lines for longevity at 18°C and dashed lines for measures at 22°C. LT80, upper and lower confidence intervals are estimated from the raw data. Pop: population, Wb: Wolbachia*.*

In females, only at 18°C does *Wolbachia* have an impact on females' longevity (Figure 3.6C). Although the LT80 estimates suggest a strong decrease in longevity for the S31-*Wb*- population upon *Wolbachia* removal (Figure 3.6C), the linear model analysis of individual lifespan (time until death per fly) does not support a significant difference (p-value = 0.0765) (Table 3.3, Table S3.3). This discrepancy is likely due to the high variability in the data, which the LT80 function may not fully account for. In contrast, both females S10-*Wb*- and S100-*Wb*- populations show a significant increase in longevity at 18°C following *Wolbachia* removal (LMER p-values = 0.0026 and 0.0004, respectively) (Figure 3.6C). Interestingly, these two populations also exhibit a significant increase in dry weight after treatment against *Wolbachia* (p-values = 0.0001 for S10 and 0.001 for S100) (Figure 3.6A). While *Wolbachia* had no effect on male dry weight, in females, there is a significant interaction between population and dry weight (p-value = 0.0073) (Figure 3.6A). These are opposite results to the males that may be explained by the differences in physiological determinants of lifespan between sexes (as reviewed in Lushchak et al. 2023).

Term	Sum of Square	Mean Square	F-statistic value	Pr(>F)
Population	12166.3364	3041.5841	2.8689	**0.0219**
Temp	341180.6821	341180.6821	321.8082	**3.391e-67**
Wolbachia	12942.7107	12942.7107	12.2078	**0.0005**
Population:Temp	35623.1341	8905.7835	8.4001	**1.009e-06**
Population:Wolbachia	11591.1939	2897.7985	2.7333	**0.02765**
Temp:Wolbachia	2609.0551	2609.0551	2.4609	0.1169
Population:Temp:Wolbachia	9232.0353	2308.0088	2.177	0.0692

Table 3. 3: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, and* Wolbachia *infection, as well as their interactions, on females' longevity.

Term indicates the main effects or interaction terms included in the model. The table reports the sum of squares (SS), mean squares (MS), F-statistic values, and associated probabilities (Pr(>F)). Significant effects

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

(p < 0.05) are shown in bold, indicating main and interaction terms contributing to variation in the response. Temp = temperature.

Term	Sum of Square	Mean Squared	F-statistic value	Pr(>F)
Population	21909.1484	5477.2871	13.7355	**4.729e-11**
Temp	36421.4313	36421.4313	91.335	**3.443e-21**
Wolbachia	22067.6836	22067.6836	55.3397	**1.502e-13**
Population:Temp	5592.7708	1398.1927	3.5063	**0.0073**
Population:Wolbachia	14641.7469	3660.4367	9.1794	**2.394e-07**
Temp:Wolbachia	11.4303	11.4303	0.0287	0.8656
Population:Temp:Wolbachia	2096.9423	524.2356	1.3146	0.2622

Table 3. 4: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, and* Wolbachia *infection, as well as their interactions, on males' longevity.
"Term" indicates the main effects or interaction terms included in the model. The table reports the sum of squares (SS), mean squares (MS), F-statistic values, and associated probabilities (Pr(>F)). Significant effects (p < 0.05) are shown in bold, indicating main and interaction terms contributing to variation in the response. Temp: temperature.

Overall, the relationship between longevity, dry weight, and *Wolbachia* infection status here appears to require further investigation. While dry weight appears as a correlative factor of the longevity in females, it is not supported by the data from the males. In males, the interplay may even extend to the interaction of the TE content with the presence of *Wolbachia*. Furthermore, it has been repeatedly reported that *Wolbachia* exerts sex-specific effects, differentially impacting male and female *D. melanogaster* (Fry and Rand 2002; Fry et al. 2004; Werren et al. 2008; Strunov et al. 2022), as we also observe here. Beyond the mere presence of the endosymbiont, our current findings suggest an additional layer to this sexual dimorphism: the interaction between *Wolbachia* and TEs may further shape male-specific patterns of ageing and longevity.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

6. Environment-independent effects of Wolbachia *on metabolic rate in populations with varying TE content*

The metabolic rate was measured for both males and females at three different temperatures (18°C, 25°C and 30°C) after the removal of *Wolbachia* and compared to the *Wb+* data (Larue et al. in prep.). In males, *Wolbachia* infection status has an overall effect on the metabolic rate (p-value = 0.0058) (Table 3.6), with it being generally higher in the presence of the endosymbiont (Figure 3.7B). This trend is however not observed in the females' dataset. A general observation is that G0-*Wb*- responds with a decrease in metabolic rate when uninfected. If we consider that G0-*Wb*- is the baseline reaction to *Wolbachia* removal, the departure from the trend for nearly all the other populations suggests that TE content and *Wolbachia* interact resulting in a modification of the metabolic rate (Figure 3.7A and 3.7B). Additionally, a general trend in the data is that, as temperature increases, the effect of *Wolbachia* removal in females seems to lead to more differentiation between populations, while in males the opposite trend is observed. Supporting these, in both females and males, the statistical analysis shows a significant interaction between the infection status and the population variables (p-values = 0.0005 and 0.0128, respectively) (Tables 3.5 and 3.6). More specifically, in males at 18°C both G0-*Wb*- and S73-*Wb*- show a significant decrease in metabolic rate upon *Wolbachia* treatment (p-values = 0.0188 and 0.0007, respectively) (Table S3.4). At 25°C, only G0-*Wb*- keeps this trend (p-value = 0.0378). In females, on the other hand, no significant differences are observed at 18°C (Figure 3.7A, Table S3.5). However, at 25°C, S73-*Wb*- experiences an increase in metabolic rate upon *Wolbachia* removal (p-value = 0.0156), an effect that is accentuated at 30°C (p-value = 0.0005) (Figure 3.7A). Meanwhile, females S100-*Wb*- display a significant decrease in metabolic rate upon *Wolbachia* removal at 25°C (p-value = 0.0364) (Figure 3.7A).

Activity, which may be influence by the presence of *Wolbachia* (Fleury et al. 2000; Peng and Wang 2009; Evans et al. 2009; Bi and Wang 2020), can drastically influence the metabolic rate (Berrigan and Partridge 1997; Lehmann and Schützner 2010), and although our statistical model accounts for activity, we cannot entirely rule out that it may contribute to the observed differences in metabolic rate, however unlikely. Additionally, it has been reported that tetracycline treatment can reduce mitochondrial efficiency and density two generations after treatment in *D. simulans*, which would in turn affect the metabolic rate (Ballard and Melvin 2007). However, this explanation is also unlikely, as the treatment occurred many tens of generations ago, and mitochondrial numbers should have since recovered (Ballard and Melvin 2007). Lastly, it has been observed that bacterial infection can lead to more homogeneous patterns of activity in *D. melanogaster* (Shirasu-Hiza et al. 2007). However, unlike *Wolbachia*, the bacteria

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

studied were exogenous pathogens, namely *Streptococcus pneumoniae* and *Listeria monocytogenes*, and this is therefore only marginally informative for interpreting the observations in the present study (Shirasu-Hiza et al. 2007). Overall, few studies have explored the effect of *Wolbachia* on the metabolic rate, and to our knowledge, none have investigated the combined influence of TE content, *Wolbachia*, and the metabolic rate. This analysis would greatly benefit from a more in-depth investigation to better understand the underlying mechanisms.

The results, however, further support the possibility of an interaction between the presence of the endosymbiont and its influence on the organism's TE dynamics. In this context, the metabolic rate serves as a useful indicator of overall energy expenditure and metabolic state. It is also noteworthy that the effects once again differ between males and females, aligning with the longevity results and further highlighting the sexual dimorphism in *Wolbachia*'s impact.

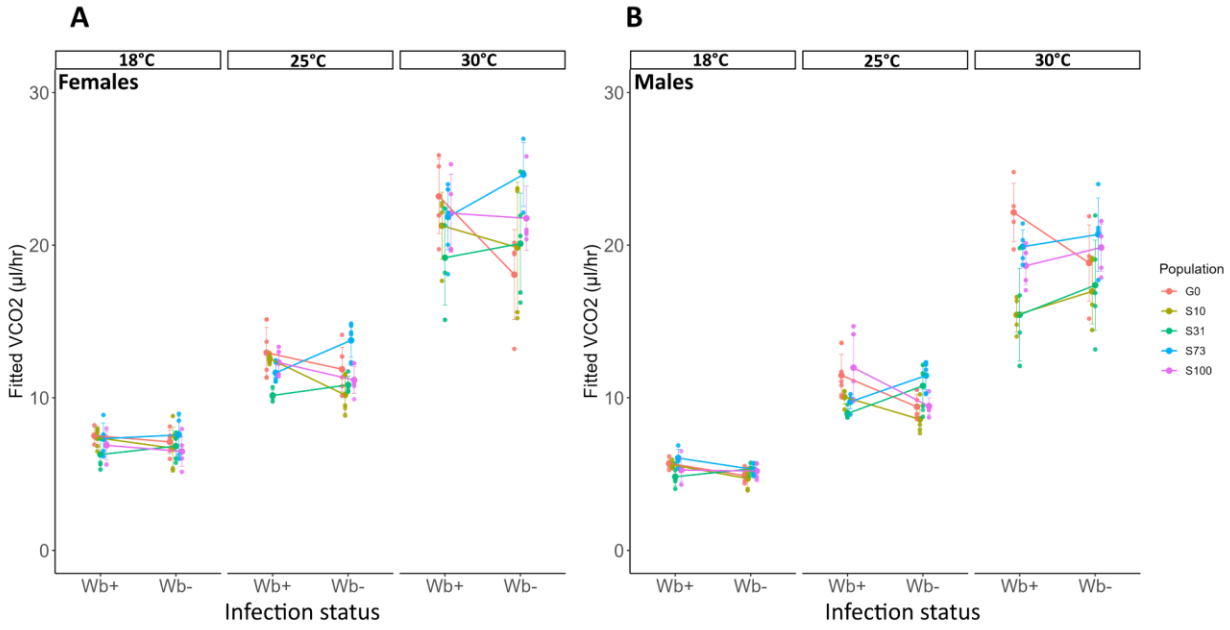


Figure 3. 7: Reaction norm of the metabolic rate according to* Wolbachia *infection status.
The reaction norm of the metabolic rate in females (A) and in males (B) upon removal of Wolbachia*, at the three tested temperatures (18°C, 25°C and 30°C). Data points, upper and lower confidence intervals all represent the fitted data. Wb+: presence of* Wolbachia*, Wb-: treated against* Wolbachia*.*

Term	Sum of Square	Mean Squared	F-statistic value	Pr(>F)
Temp	1.8056	0.9028	162.869	**4.347e-12**

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Population	0.1110	0.0277	5.0089	**0.0011**
Wolbachia	0.0001	0.0001	0.0212	0.8847
log(DW_Fly)	0.3155	0.3155	56.9335	**2.096e-11**
log(Act_Fly + 1)	0.0277	0.0277	5.0098	**0.0273**
Temp:Population	0.0593	0.0074	1.3366	0.2353
Temp:Wolbachia	0.0089	0.0045	0.8073	0.4491
Population:Wolbachia	0.1229	0.0307	5.5468	**0.0005**
Temp:Population:Wolbachia	0.0395	0.0049	0.8921	0.5263

Table 3. 5: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, and* Wolbachia *infection, as well as their interactions, on females' metabolic rate.

"Term" indicates the main effects or interaction terms included in the model. The table reports the sum of squares (SS), mean squares (MS), F-statistic values, and associated probabilities (Pr(>F)). Significant effects ($p < 0.05$) are shown in bold, indicating main and interaction terms contributing to variation in the response. Temp: temperature, DW: dry weight, Act: activity.

Term	Sum of Square	Mean Squared	F-statistic value	Pr(>F)
Temp	2.8259	1.413	2	**1.348e-13**
Population	0.1008	0.0252	4	**0.0262**
Wolbachia	0.0695	0.0695	1	**0.0058**
log(DW_Fly)	0.1116	0.1116	1	**0.0005**
log(Act_Fly + 1)	0.1329	0.1329	1	**0.0002**

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Temp:Population	0.0424	0.0053	8	0.7691
Temp:Wolbachia	0.0460	0.0230	2	0.0765
Population:Wolbachia	0.1172	0.0293	4	**0.0128**
Temp:Population:Wolbachia	0.0725	0.0090	8	0.4131

Table 3. 6: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, and* Wolbachia *infection, as well as their interactions, on males' metabolic rate.

"Term" indicates the main effects or interaction terms included in the model. The table reports the sum of squares (SS), mean squares (MS), F-statistic values, and associated probabilities (Pr(>F)). Significant effects ($p < 0.05$) are shown in bold, indicating main and interaction terms contributing to variation in the response. Temp: temperature, DW: dry weight, Act: activity.

III. Conclusion

Here, we revisited populations with variable TE content from which the endosymbiont *Wolbachia* had been removed, and we investigated the resulting phenotypic effects, given that *Wolbachia* alone – or in interaction with TEs – could induce phenotypic modifications. It is important to note that the antibiotic treatment subjected these populations to a bottleneck, which appears to be reflected in the distinctive TE content observed following treatment. This treatment also seems to have caused a modification of TE landscape, both in terms of which insertions were transmitted to subsequent generations – as a *Wb-* population and its *Wb+* counterpart share approximately 60% of insertions – and in their frequencies, as evidenced by the slight decrease in overall TE frequencies. A detailed analysis to identify candidate insertions in relation to phenotypic traits remains to be conducted.

While the impact of endosymbionts on host phenotypes is well documented, the role of TEs, as well as the interaction between TEs and endosymbionts, remains largely understudied. This project therefore aims to bridge this gap, albeit with considerable limitations, both technical and due to the scarcity of available literature. Our results support an interaction between *Wolbachia* infection and the host's TE content, with consequences observed across several phenotypic traits. Notably, infection status appears to enhance sexual dimorphism, with males and females often showing distinct responses to the removal of *Wolbachia*, as previously reported by a few studies (Fry and Rand 2002; Fry et al. 2004; Werren et al. 2008; Strunov et al. 2022). This suggests that the impact of *Wolbachia* cannot be generalized and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

may partly depend on the host's TE landscape. Additionally, we show that this interaction is sensitive to environmental conditions, as illustrated by the complex relationship between longevity, dry weight, and infection status, as well as the responses of metabolic rate under varying temperatures and infection status. These results suggest a potential interaction between hologenotypic variability and TE content, influencing both phenotypes and their responses to environmental changes. Overall this study testifies of the complexity of the interaction, all the more so as TE content along with *Wolbachia* infection status varies, introducing confounding factors that make it difficult to draw conclusions.

Although there is some evidence that *Wolbachia* can influence TE expression, the relationship between TEs, *Wolbachia* and environmental response remains to be explored (Mayoral et al. 2014; Eugénio et al. 2023; Mostoufi and Singh 2024). While a few mechanisms have been proposed to explain the endosymbiont's influence on TE activity, we are far from understanding how this interaction could translate into phenotypic outcomes, something we have only begun to touch upon here. Future work should aim to fill this gap by explicitly investigating how TE-*Wolbachia* interactions influence host phenotypes and responses to environmental variation.

IV. Material and Methods

1. Fly husbandry

A *D. melanogaster* line was previously modified to suppress the piRNA pathway in charge of TE regulation (Barckmann et al. 2018). Since a constitutive piRNA knock-down is lethal, a system was engineered to target ovarian somatic cells with a heat-induced piRNA knock-down. Briefly, the fly line carries three components: (i) a GAL4 UAS-activator driven by the follicle cell-specific traffic jam (tj)-promoter (tj-GAL4), (ii) a UAS-short-hairpin(sh)-piwi that induces *Piwi* RNAi and (iii) an ubiquitously expressed thermo-sensitive GAL4-inhibitor, GAL80ts (Barckmann et al. 2018). The initial isogenic population (G0) was repeatedly exposed to temperature changes while maintaining a population size of around 500 progenitors in standard medium-filled bottles, to reduce genetic drift. The consequence is the occurrence of independent transposition events in each individual of the offspring, leading to insertion polymorphism in a constant genetic background (Barckmann et al. 2018). In total, five populations harbouring different TE content and insertions were generated (G0, S10, S31, S73, S100). G0 is the parent population, and all subsequent shifts are named SX, where X is the number of generations that have been shifted. Hence, S10 corresponds to 10 generations of shifts (10 generations of piRNA knock-down) and therefore of TE accumulation. Gifted flies were all infected by *Wolbachia.* Flies were maintained at 18°C

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

in 12-hour light cycles, with 65% relative humidity, in vials with standard medium (yeast, sugar, wheat flour and nipagin as a preservative), with a density of approximately 100 progenitors at each generation. Each population was maintained in five vials per generation, which were frequently shuffled to avoid vial effects.

2. Wolbachia treatment and fly maintenance

To treat the populations and eliminate *Wolbachia*, five replicates of 50 3- to 6-hours-old embryos were collected and placed on standard fly media supplied (yeast, sugar, wheat flour and nipagin as a preservative) with tetracycline hypochloride (Sigma-Aldrich) to 0.25mg/mL, as described in Mayeux and colleagues (Mayeux et al. 2024). The flies were frequently checked for the effective absence, as well as presence, of *Wolbachia* with polymerase chain reaction (PCR) experiments on the following genes: FtsZ (5'-CGAGATGGGCAAAGCGATGA-3' and 5'-ATTCCTTGCGCACCTTTCAT-3') (Monnin et al. 2016), wsp (5'-TGGTCCAATAAGTGATGAAGAAAC-3' and 5'-AAAAATTAAACGCTACTCCA-3') (Riegler et al. 2005) for *Wolbachia* detection, and RP49 as *D. melanogaster*'s control gene (5'-CGGATCGATATGCTAAGCTGT-3' and 5'-GCGCTTGTTCGATCCGTA-3') (Akkouche et al. 2012). Flies were maintained at 18°C in 12-hour light cycles, with 65% relative humidity, in vials with standard medium, with a density of approximately 50 couples at each generation. A substantial number of generations have elapsed between the antibiotic treatment and the subsequent experiments, making it unlikely that the treatment had any confounding effects.

3. Wolbachia relative density

Relative *Wolbachia* titers were obtained by quantitative PCR (qPCR). Duplex qPCR reactions were performed simultaneously for *RP49* (*D. melanogaster* housekeeping gene) and *FtsZ* (*Wolbachia* gene) with a ratio 2:1 for primers and probes to ensure maximum amplification efficiency for both genes as routinely performed (Monnin et al. 2021). Primers are listed in table S3.6. Reactions are set up using SsoAdvanced Universal Probe Supermix (Biorad) following manufacturer's instructions, with the following thermocycler's parameters: 2 min of initial denaturation at 95°C followed by 40 cycles of 10 sec of denaturation at 95°C and 30 sec of combined hybridation and elongation at 60°C. The quantification cycle (Cq) were analysed using R software (v 4.4.0) to compute the relative quantification of *Wolbachia* following the mathematical method proposed by Pfaffl (Pfaffl 2001). The ratio was log10 transformed and statistically tested with a two-way ANOVA. The *Wolbachia* call status (presence/absence) were obtained

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

with Design&Analysis (v 2.6.0) Real-Time PCR systems from appliedbiosystems by Thermo Fisher Scientific. Briefly, the normalized intensity of the reporter dye (Rn) is measured both prior and after qPCR amplification and the change in fluorescence (ΔRn) is computed to obtain end point amplification. Samples lacking detectable amplification of the target Wolbachia sequence were classified as *Wb-*.

4. DNA extraction and sequencing

DNA extractions were performed on 200 embryos per population using a homemade protocol designed to yield very long DNA fragments. Briefly, tissues are homogenized in STE buffer (chloride-Tris-EDTA), to which is added 10% SDS (sodium dodecyl sulfate). The mix is incubated for 2 h 30 min with proteinase K at 55°C. The samples are then treated with RNase for 1 h at 37°C. The DNA is then precipitated with phenol/chloroform/isoamyl alcohol. The aqueous phase is recovered, and the DNA is purified by isopropanol and ethanol washing steps. The DNA is eluted in 35 µL of elution buffer and is stored at 4°C until library construction and sequencing, performed at the PSI in house sequencing platform at the Institut de Génomique Fonctionnelle de Lyon (IGFL), on a PromethION 2 Solo using Oxford Nanopore Sequencing Technology. The data received was filtered for reads with quality above Q10.

5. Transposable element analysis

Analyses were computed on 35 X coverage subsamples of the raw reads for all populations generated with Rasusa (v 2.0.0) (Hall 2022) using *D. melanogaster* dm6 genome version to guide the subsampling. TE variants and frequencies were obtained with TrEMOLO using default parameters of the OUTSIDER module and G0-*Wb+* as reference genome and the *D. melanogaster* MCTE TE library previously published (Rech et al. 2022; Mohamed et al. 2023). TE insertional variant were filtered from the "TE_INFOS.bed", and no_name contigs were removed for downstream analysis. Differences in average TE insertion frequency among populations were tested using the Kruskal-Wallis test. A shared insertion matrix was generated using a custom script, based on 100 bp windows. Cases of ambiguous sharing were manually curated: insertions were merged if the same TE was present in more than one population, or marked as unresolved if no consistent pattern could be identified. Additionally, when multiple insertions from the same population were found at the same position, the same criteria were applied, and their frequencies were summed. Plotting of the insertions per class and families was performed in R 4.4.0. Venn diagrams were generated using Intervene (v0.5.8) (Khan and Mathelier 2017) to estimate the number of shared insertions within a 100bp window between *Wb+* and *Wb-* populations. DeviaTE (v2.2.1) was run

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

on the subsampled data using default settings and three normalization single-copy genes: *Act5*, *Piwi*, and *Rpl32* (Weilguny and Kofler, 2019). The software aligns long reads to the TE library and estimates the TE insertion number for each TE family per haploid genome. The sum of high-quality (>Q15) coverage estimates was used as a proxy for TE load in each population.

6. Viability and development time

To measure viability, 150 three-day-old female-male couples per population were placed in a small Dutscher egg-laying chamber (sexed and counted under CO2 anaesthesia). Flies were allowed to lay eggs for half a day. From these egg laying chambers, five replicates of 40 eggs were collected on small regular fly-food caps positioned in vials containing the standard media. Eggs were collected under binoculars using the tip of a scalpel to scoop the eggs out of the media. Viability was computed as the proportion of adults emerging from the initial number of eggs (40). Data is analysed under R software (v 4.4.0) using generalized linear mixed-effects model (GLMM) with binomial distribution using the Lme4 package (v 1.1-36) (R Core Team 2021; Bates et al. 2025). ANOVA Chi square test was applied to assess the fixed effect. McFadden's R^2 was estimated with the package Jtools (v 2.3.0) (Long 2024). From these data development time (development time) was also assessed, for each egg counted for its viability was also recorded the time until emergence of the adults. Using this data, a cumulative percentage of flies emerged each day since oviposition was calculated from the total number of eggs tested per replicate (five replicates of 40 eggs per conditions). The comparative statistical analysis of the emergence curves between populations was performed using a Log Rank test in R (v 4.4.0) (R Core Team 2021). Development time were also analysed using ANOVA on the time at which each fly has emerged in R 4.4.0 (Stat built-in package) (Bolar 2019; R Core Team 2021).

7. Embryonic hatchability

To measure hatchability, 150 three-day-old female-male couples per population were placed in a small Dutscher egg-laying chamber. Flies were allowed to lay eggs for half a day. Eggs were collected using a fine brush and water, and filtered through a fine mesh. Eggs remained dipped in water while five replicates of 50 eggs were collected and aligned on a black piece of Canson 160 g/m² paper. Two days after sampling, the number of emerged eggs were counted under binoculars. Data were analysed under R software (v 4.4.0) using generalized linear mixed-effects model (GLMM) with binomial distribution using

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

the Lme4 package (v 1.1-36) (Bates et al. 2025). ANOVA Chi square test was applied to assess the fixed effect.

8. Dry weight

To measure hatchability, 150 three-day-old female-male couples per population were placed in a small Dutscher egg-laying chamber. Flies were allowed to lay eggs for half a day. Eggs were collected using a fine brush and water, and filtered through a fine mesh. Eggs remained dipped in water while five replicates of 50 eggs were collected and aligned on a black piece of Canson 160 g/m² paper. Two days after sampling, the number of emerged eggs were counted under binoculars. Data were analysed under R software (v 4.4.0) using generalized linear mixed-effects model (GLMM) with binomial distribution using the Lme4 package (v 1.1-36) (R Core Team 2021; Bates et al. 2025). ANOVA Chi square test was applied to assess the fixed effect.

9. Negative geotaxis essays

A pool of ten 3-to-5-days old flies was placed in pre-washed vials (2 cm in diameter, 9.5 cm in length), which had been dried and left to air out to eliminate any residual plastic odour. The vial was divided into three predefined zones: the first zone extended from the bottom to a height of 2 cm, the second from 2 cm above the bottom to 2 cm below the top, and the third from 2 cm below the top to the top of the vial. Prior to the trial, flies were placed in the tubes using a mouth aspirator and allowed to acclimate for 30 minutes. The flies were gently tapped three times to ensure they all fell to the bottom. After a two-minute trial, the number of flies in each zone was recorded. This procedure was repeated in batches of three experimental sessions, with each population and sex represented in every batch, resulting in six biological replicates (totalling 60 flies per condition). Taking the count of flies reaching the top of the tube as success, and the ones in the rest of the tubes as failure (middle and bottom zones), the statistical analysis was conducted in R (v 4.4.0) using a generalized linear mixed-effects model (GLMM) with a binomial distribution (R Core Team 2021; Bates et al. 2025). ANOVA was applied to assess fixed effects, using a Chi-square test.

10. Longevity

Ten replicates of pools of around 10 virgin females or males were placed in vials containing standard fly food, for each of the five populations (G0, S10, S31, S73, S100) (Table S3.7). All replicates for

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

all populations were carried within a week and therefore stem from the same generation of flies. In addition, flies were placed on the same batch of medium. Virgin flies were collected in the afternoon, after vials were emptied in the morning, and are sexed under CO2 anaesthesia. Vials were changed once every week to prevent low-quality media from impacting the experimental setup. The cumulative number of dead flies was counted every 24 hours on weekdays. Survival curves and Log Rank test pairwise comparisons were generated with the Survminer (v 0.5.0) and Survival (v 3.5-8) packages in R 4.4.0 (Kassambara et al. 2021; Therneau et al. 2024). This experiment was conducted at 18° C and 22°C. The time for 80% of the population to have died (LT80) was estimated using the Ecotox (v 1.4.4) package in R 4.4.0 (Wheeler et al. 2006; R Core Team 2021). Analysis of the survival chances was performed using generalized linear model (GLMER) and probit distribution with replicates as random effects. Type III ANOVA was applied to assess the fixed effects, and EMMs were adjusted using the Bonferroni adjustment method to account for multiple comparisons. Additionally, a linear comparative analysis of longevity at 18°C *vs.* 22°C was performed using a linear mixed-effects model with the lmerTest (v 3.1-3) package (Christensen [2018] 2025). Type III ANOVA was also applied to assess the fixed effects, and estimated marginal means (EMMs) were adjusted using the Bonferroni adjustment method to account for multiple comparisons.

11. Metabolic rate

Metabolic rate measurements were obtained with a MAVEn™ device from Sable System International. Five flies for each sex and population were placed in arenas adapted to the device. Five replicates per condition were conducted, with measures taken for three cycles (three repeated measures for each arena). Experimental temperatures were 18°C, 25°C and 30°C, in the darkness to prevent differences in activity due to unequal exposure to light. Volume of carbon dioxide (VCO2) emission and infrared-light activity detection were extracted and subsequently used as a proxy to metabolic rate and activity, respectively. Each pool of flies was frozen at -20°C and dried for 24 h at 75°C for subsequent weighing. VCO2 log-transformed data were analysed with linear mixed effect models in R (v 4.4.0), with dry weight and activity as covariants (R Core Team 2021; Christensen [2018] 2025). Type II ANOVA was also applied to assess the fixed effects, and estimated marginal means (EMMs) were adjusted using the Bonferroni adjustment method to account for multiple comparisons.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

General discussion

The understanding of where phenotypic variation arises from is the foundation of the theory of evolution, for if no variation exists, there cannot be differences in fitness on which selection can act upon. All along this project we have tried, and perhaps even succeeded, in bringing further evidence that TEs contribute significantly in promoting phenotypic variability. We even went a step further and collected empirical evidence that support that TEs also contribute to environmentally induced phenotypic changes. While this study certainly sets a new stone in the understanding of the contribution of the overall TE load to phenotypes, and particularly to the phenotypic response to environmental changes, a great achievement in itself, it also comes with its share of limitations. In this discussion, I would like to emphasize the aspects that could have made this study even stronger and highlight the issues we faced along the way, so that if someone were ever to embark on a similar project, they might learn from our mistakes and realizations.

Arguably, one of the most debatable aspects of this study is the use of a model system with within-population TE polymorphism. Paradoxically, this characteristic is essential to the results presented in the previous two chapters, yet it also greatly complicates our understanding of the relationship between TEs and phenotypes. This difficulty stems mainly from the bioinformatic limitations of working with pooled sequencing. In this context, not only it is challenging to establish correlations between TE insertions and phenotypes, but it is also difficult to clearly assess the TE burden of an individual within a population. This remains a technical hurdle to overcome. TrEMOLO and DeviaTE were both useful tools (Weilguny and Kofler 2019; Mohamed et al. 2023); however, the fact that they yield different results is, to say the least, puzzling. On the one hand, TrEMOLO was developed as a dedicated tool for the analysis of TE insertions in these populations (Mohamed et al. 2023; Varoqui et al. 2025). TrEMOLO provides insight into the number of newly detected insertions, spanning 80% of the TE consensus length, relative to a reference genome. This, however, means that the choice of reference genome is critical and can significantly influence the outcome, as we experienced ourselves. We initially ran the tool against the *D. melanogaster* reference genome dm6, expecting its high-quality annotation to be an advantage. Yet, using this reference, the anticipated TE accumulation was nearly undetected (data not shown). We therefore re-ran TrEMOLO using the G0 assembly, which revealed an undeniable increase in TE content across generations. Still, one key limitation of this tool is the inability to estimate TE content per individual. As for DeviaTE, which estimates TE abundance per haplotype by comparing TE consensus coverage with that of single-copy genes, it was originally optimized for short-read data (Weilguny and Kofler 2019). While theoretically

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

applicable to long reads, in practice this results is a highly biased estimates of TE abundance per haplotype (Weilguny and Kofler 2019). Altogether, these experiences underscore the inherent complexity and technical challenges of detecting and annotating TEs in genomic data, which is further complicated by studying populations.

So as a first piece of advice, I would recommend moving away from population-level studies and instead focusing on isogenic lines with varying TE contents. If we were to establish ten isogenic lines for each of the five populations – representing ten replicates with identical TE load but differing TE composition – we would have an experimental approach that would allow us to test not only the effect of TE load, but also the differences between varying TE cohorts. Additionally, this approach would enable the investigation of the effects of individual insertions but also provide insight into potential interactions between sets of insertions. A logical starting point for evaluating these isogenic lines would be a phenotypic screening to determine whether they show divergence. Subsequently, long-read sequencing could establish the genotype of each line. With this biological framework, it would then be possible to perform a genome-wide association study (GWAS), an approach successfully and repeatedly employed by Mackay's laboratory (Mackay et al. 2012; Montgomery et al. 2014; Ivanov et al. 2015; Vonesch et al. 2016), which could help generate more precise TE-phenotype associations.

Echoing this particular aspect of the project, investigating the underlying causes of the phenotypic differences and the increased variability observed in S73 would be particularly challenging at the population level. Indeed, our hypothesis rests on the idea that TEs promote transcriptomic variability and plasticity, which then translates into phenotypic plasticity. If we were to explore transcriptomic profiles in these populations, both gene- and TE-wise, we would likely face substantial variability due to micro-environmental plasticity, making it difficult, if not impossible, to detect genuine biological signals. Indeed, several studies have demonstrated that even genetically identical individuals from isogenic lines can display markedly different transcriptomic profiles (Huang et al. 2015; Lin et al. 2016; García-Nieto et al. 2022). An additional layer of complexity comes from the difficulty of deciding on the tissue and developmental stage at which to perform transcriptomic analyses. For some traits, such as hatchability, the answer seems obvious: RNA sequencing should be performed on eggs. However, for most other traits, selecting the appropriate sample would require in-depth literature research to make an informed decision. To this challenge, one can add the complication of choosing samples according to where the focal genes and TEs are actually expressed. Hence, transcriptomic studies may be difficult to conduct on these populations, as would attempts to investigate whether the phenotypic variability and plasticity observed are due to changes in TE and gene expression. Only if the signal is strong would we be able to

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

detect it, provided that the appropriate assessment method is used. Some ecological diversity indexes have been adapted to assess transcriptomic diversity, such as the Shannon entropy index, and may offer a methodological approach to overcome this challenge (Ameri and Lewis 2021; García-Nieto et al. 2022; Jones et al. 2024).

Transcriptomic variability may not only be studied in terms of overall diversity, but TE contributions to transcriptomic profiles may also arise from the formation of chimeric transcripts, in which TE sequences and genes are merged. These chimeric transcripts may be more abundant in S73 than in other populations, under control conditions and/or environmental stress, potentially leading to an increased phenotypic variation. Applying ChimeraTE (Oliveira et al. 2023) to investigate whether an increased TE load results in more chimeras, and how this relationship changes under environmental variation, would greatly enhance our understanding of the mechanisms underlying TE-driven phenotypic plasticity. Similarly, ParasiTE was designed to detect chimeric TE-gene transcripts using long read RNA sequencing. Unlike ChimeraTE, ParasiTE also allows the detection and the relative quantification of different chimeric isoforms, as shown in *Arabidopsis thaliana* (Berthelier et al. 2023). To investigate the mechanisms controlling isoform expression, Berthelier and colleagues analysed isoform usage in DNA methylation and H3K9 methylation mutants, reporting that 10-20% of genes associated with alternative TE-gene isoforms displayed significant differential isoform usage. Notably, changes in isoform usage were associated with reduced DNA methylation at TE loci (Berthelier et al. 2023). Furthermore, studying datasets with *A. thaliana* submitted to variable environmental stresses, they also uncovered changes in isoform usage and differential expression upon exposure (Berthelier et al. 2023). This further corroborates the relevance to investigate chimeric transcripts in response to environmental variation in the present *D. melanogaster* populations, as well as the epigenetic mechanism regulating TE activity.

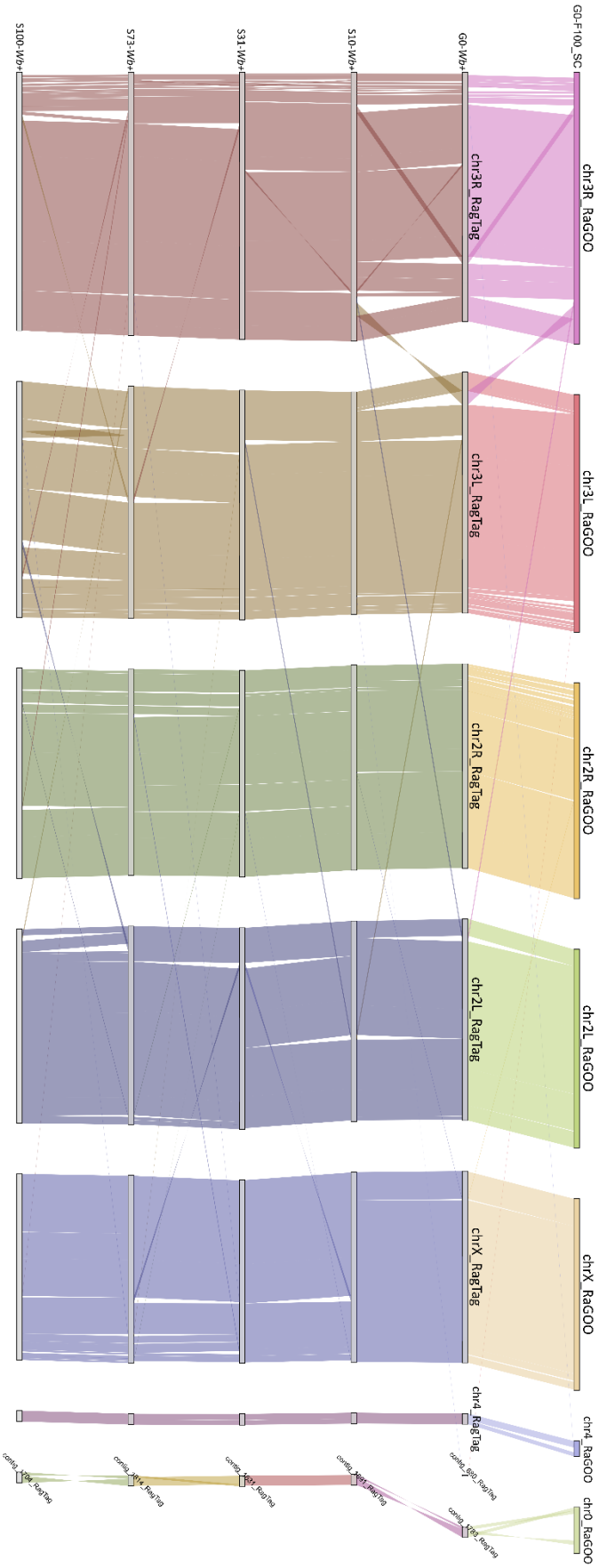
Our hypothesis also implies that TE-induced transcriptomic variability is mediated by mechanisms that are themselves sensitive to environmental fluctuations (Jaenisch and Bird 2003; Slotkin and Martienssen 2007). The investigation of chromatin accessibility through the study of repressive histone tail modifications, such as H3K9me3 and H3K27me3, could shed light on how TE regulation contributes to the phenotypic profile, variability, and plasticity of the five populations with increasing TE content. However, as shown by Guio et al. (2018), both the repressive H3K9me3 and the permissive H3K4me3 marks were observed to co-occur at the *Bari-Jheh* locus upon exposure to paraquat, even though these marks are mutually exclusive, illustrating the complexity of studying epigenetic marks (Guio et al. 2018). Hence, working at the population level remains a challenge in understanding how plasticity may be TE-driven.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Additionally, although *a priori* we do not expect a significant number of non-TE SV, genetic drift is unpredictable, and may have muddled our genotypes. Therefore, a non-TE SV analysis is required to remedy this current loophole in our project. To evaluate the extent to which the populations diverge beyond their TE content, a preliminary synteny analysis was conducted, yielding promising results, mostly highlighting assembly errors rather than true sequence divergence (Figure 4.1). We are also currently performing a comparative analysis to further assess divergence beyond TE content. However, once again, conducting such analyses at the population level renders interpretation somewhat challenging.

While it may seem that I am being overly critical of this study, I want to emphasize the value of having undertaken it in the first place. Only by going through this step, I am now able to see its caveats. But also, although the design may no longer be the ideal experimental setup, it was the best available option at the time, and it nonetheless allowed us to obtain results to test our initial hypothesis.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés



Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Figure 4. 1: Synteny analysis across populations and assemblies for the* Wolbachia*-infected populations. *Synteny relationships are shown between the reference assemblies (G0-F100_SC, assembled genome from our collaborator (Barckmann et al. 2018; Mohamed et al. 2023) and our G0-Wb+ assembly) and the TE accumulation populations (S10-Wb+, S31-Wb+, S73-Wb+, S100-Wb+). Grey ribbons connect homologous genomic regions, illustrating conserved collinearity and rearrangements across scaffolds and contigs. Chromosome arms and contigs are indicated at the top and aligned for comparison. The synteny suggest some assembly errors that are corrected in the following population as displayed by the symmetrical translocations and inversions.*

My second piece of advice is: choose your phenotypic measurements and environmental fluctuations wisely. We hypothesized that variable TE content could differentially burden cellular homeostasis, by impacting transcriptomic profiles, piRNA amounts, and epigenetic profiles, thereby affecting gene expression and leading to phenotypic variations (Fablet and Vieira 2011; Rebollo et al. 2012; Chuong et al. 2017; Ecker et al. 2018; Choi and Lee 2020; Moschetti et al. 2020). Thus, it was crucial to test our hypothesis with phenotypic data, the substrate for selection. Instead of precision, we opted for breadth, resulting in the measure of 17 phenotypes in the females and 16 in males *Wb+*. This enabled us to span a wide array of phenotypes, however in doing so, we reduced the extent to which we explore the details of each trait. While this multitude of traits helped reinforce our initial hypothesis, we remain far from fully understanding the intricacies of each individual phenotype.

For example, the CT_{max} results were not discussed while the females S100-*Wb+* display a reduction in resistance to heat stress compared to the other populations (Figure 4.2A). Indeed, the resistance to heat stress was measured by computing the CT_{max}, a metric used to assess the temperature at which an organism can no longer perform essential biological processes (Lutterschmidt, 1997, Santos et al, 2011, MacLean et al 2019). CT_{max} was measured exclusively in females, as previous studies have shown there were no significant differences between sexes (Rolandi et al. 2018). On average, S100-*Wb+* flies entered heat-induced coma at 38.81°C, compared to 39.40°C for G0-*Wb+*, 39.76°C for S10-*Wb+*, 39.80°C for S31-*Wb+*, and 39.92°C for S73-*Wb+* (Figure 4.2A). Additionally, remarkable differences in survival to starvation are observed between the populations (Figure 4.2B). For instance, the S31-*Wb+* shows a dramatic reduction in survival, measured as the LT80, in both males and females compared to every other population (Figure 4.2B). At the same time, both S73-*Wb+* and S100-*Wb+* appear to have increased resistance to starvation compared to G0-*Wb+* (Figure 4.2B). Moreover, an overall trend is observed for S31-*Wb+* to exhibit a lower metabolic rate across both temperature and sex. Although this trend is not always statistically significant, likely due to p-value corrections for multiple testing, it remains consistent across all experimental conditions. The fact that S31-*Wb+* stands out from the other populations, as it is the case for development time, longevity and starvation resistance, supports the hypothesis that S31-

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Wb+ might indeed carry one or more TE insertions responsible for the observed differences. Interestingly, Raymond Pearl (1928), after experimenting on *D. melanogaster*, stated that "the duration of life varies inversely as the rate of energy expenditure during life" (Pearl 1928). This concept, now widely known as the rate-of-living hypothesis, proposes that metabolism is inversely proportional to longevity, which could explain the pattern of the S31-*Wb+* population. Additionally, TE control via the piRNA pathway is directly linked to mitochondria, as the *Zucchini* protein, which plays a key role in somatic piRNA biogenesis, is anchored to the outer mitochondrial membrane (Nishimasu et al. 2012; Czech and Hannon 2016; Czech et al. 2018; Pippadpally and Venkatesh 2020; Choi 2024). Further investigation into the mitochondrial efficiency could also provide information on the underlying cause for S31-*Wb+*'s phenotypes. These are only a few examples of traits that were not examined in depth, yet they support the idea that TE effects may act on very specific traits, further underscoring the importance of carefully selecting phenotypic traits.

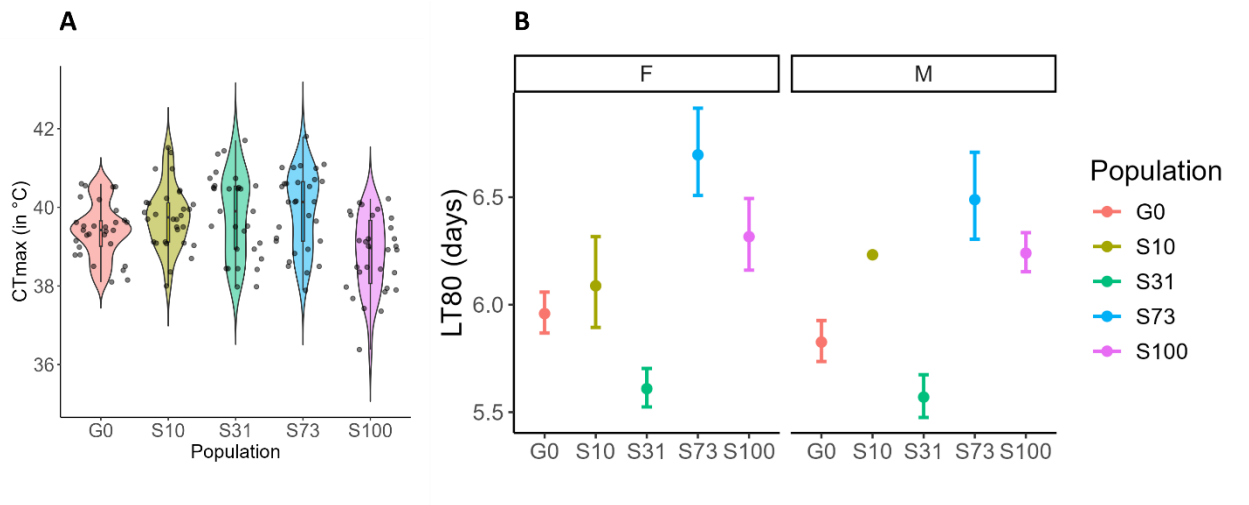


Figure 4. 2: Thermal tolerance and oxidative stress resistance across populations.
A) Critical thermal maximum (CTmax) measured in females from five populations (G0-Wb+, S10-Wb+, S31-Wb+, S73-Wb+, S100-Wb+). Violin plots show the distribution of individual values, with medians and interquartile ranges indicated. B) LT80 under paraquat treatment (time until 80% mortality) in females (F) and males (M) across the same populations. Dots represent estimated means and error bars the 95% confidence intervals.

Not only were we measuring traits, but we did so under variable conditions, which also required careful selection of the environments to test. While we included measurements under different conditions such as temperature changes, diet depletion, and oxidative stress intensity, many other environmental perturbations could also impact *D. melanogaster* physiology (Gebhardt and Stearns 1993; Shingleton and Tang 2012; Zhou et al. 2012; Bauerfeind et al. 2014; Behrman et al. 2015; Klepsatel et al. 2020; Kim et al. 2020; Flatt 2020). An environmental challenge that would be interesting to test is an

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

immune challenge. Indeed, as we saw in the introduction with the study by Merenciano and González (2023), TE can introduce immune-related TFBS, and the phenotypic impact of these insertions only becomes apparent under infection conditions (Merenciano and González 2023). In fact, prior investigations from the same laboratory revealed 12 insertions associated with expression changes in immune-related genes, three of which were functionally validated (Ullastres et al. 2021). And as we have seen in Mayeux et al. (2024), the siRNA pathway operates at the interface of both viral and TE control (Mayeux et al. 2024) (Annexe 4). Hence, studying the viral defence response is generally relevant to understanding TE dynamics. Additionally, Schmidleithner and colleagues make a compelling case for considering TEs in the context of immunity (Schmidleithner et al. 2025). They highlight the role of TEs as enhancers in adaptive immunity and note the broader observation that "ERV TEs serve as *cis*-regulatory elements that have a critical role in the control of T H[elper] cell development and plasticity" (Schmidleithner et al. 2025). While this perspective paper focuses on mammalian immunity, it goes to show that TEs and immunity are not independent systems. Altogether, these reinforce the relevance of immune challenges, especially considering that different pathogens (viruses, bacteria) activate different immune pathways in *D. melanogaster* (siRNA, Toll and IMD pathways) (Hoffmann 2003; Leclerc and Reichhart 2004; Czech et al. 2008), each of which may be affected in distinct ways by TEs (Roy et al. 2020; Mayeux et al. 2024).

Finally, we decided early on to investigate adults' responses to environmental changes, yet environmental fluctuations experienced during development can have a profound and lasting impact on the adult phenotype (Nijhout 2003; West-Eberhard 2003; Lafuente and Beldade 2019). This raised a challenge that has been consistently difficult to navigate throughout the writing of this manuscript: distinguishing between phenotypic plasticity and developmental plasticity. This challenge comes from the fact that the theoretical foundations of phenotypic plasticity have often been established through the study of developmental plasticity. Developmental plasticity is defined as "the property by which the same genotype produces distinct phenotypes depending on the environmental conditions under which development takes place" (Lafuente and Beldade 2019). In *D. melanogaster*, conditions encountered during development – such as temperature, crowding, or diet – have remarkably long-term effects on individuals (David et al. 1997; Tu and Tatar 2003; Shingleton and Tang 2012; Klepsatel et al. 2018; Flatt 2020). On top of that, the numerous effects of TEs on genome structure could further alter developmental processes, making variation all the more likely (Whitelaw and Martin 2001; Pimpinelli and Piacentini 2020; Marin et al. 2020; Catlin and Josephs 2022a; Merenciano et al. 2024). We already have evidence that TEs impact developmental stages, as shown by our results on hatchability, viability, and development time.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

When TE-driven phenotypic variation is induced during development and carried into adulthood, its evolutionary implications may align with the ecological concept of *diversifying bet-hedging*, briefly mentioned in the introduction. This strategy involves a trade-off in which a lower overall individual fitness is offset by an increase in genetic and/or phenotypic variance among offspring, thereby increasing the likelihood that at least a few will survive under unpredictable environmental conditions (Hopper 1999). Given this evolutionary relevance, I believe it would be particularly valuable to subject these populations to developmental plasticity assays. For example, we reported differences in dry body mass between the populations, making it particularly relevant to investigate how larval crowding, known to influence adult size (Miller and Thomas 1958; Bakker 1959; Klepsatel et al. 2018), may contribute to these results. Larval crowding has also been associated with variation in development time and viability (Horváth and Kalinka 2016). Likewise, we observed that diet, and specifically the depletion of different nutritional sources, drives differences in adult lifespan. This raises the question of how diet experienced during larval development may exert long-term effects on adult survival and development time in these populations with increasing TE content, given that larval diet has been demonstrated to influence a wide array of traits in the adults (Tu and Tatar 2003; Piper et al. 2005; Shingleton et al. 2017; Grangeteau et al. 2018; Poças et al. 2022). For instance, it has been reported that adults emerging from yeast-deprived larvae exhibit smaller body size, delayed hatching, reduced ovariole number, and decreased age-specific fecundity (Tu and Tatar 2003).

To conclude on this aspect, selecting which traits and environments to test can dictate how impactful the results and conclusions will be. In our case, many other relevant traits may have been investigated. *D. melanogaster* being a model organism, much literature was available to help us decide, and we were fortunate to observe distinct phenotypic profiles between the populations. We thus now have evidence that, with a nearly constant genetic background but variable TE contents, changes in phenotypes and responses to environmental variation can occur. However, the fact that this study was conducted on an artificial biological system allowed us to isolate the variables we wanted to test. But by doing so, we obviously magnified the relationship between TEs and phenotypic plasticity. Arguably, for researchers working in ecology with natural populations, or any non-model system where controlling TE content is virtually impossible, the extent to which our results will be impactful is probably close to none. If we even get a mention from that community, I'll take it as a win. In line with this, the TE community and the plasticity community rarely ever overlap. Naturally, these topics differ greatly. Will the TE folks be sceptical of the plasticity angle in this research? And the other way around? I sometimes day-nightmare

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

about this project not quite belonging to any research community… Being at the crossroads can be both a motivation and a curse! But that's of course just me overthinking it!

Lastly, while we were able to provide evidence for the effect of the global TE landscape on phenotypic variability and plasticity, the role of specific insertions on phenotypic traits still remain to be addressed. Due to time constraints, the candidate insertion analysis is limited to a list of TE insertions that have not yet been functionally investigated. Some of these insertions could be tested for their impact on gene expression using RT-qPCR, or further explored with mutant lines or CRISPR-Cas9 to remove the insertions and assess their effects at the phenotypic level. There are a few promising candidates, some of which are discussed below.

You may remember the delay in development time of the S31-*Wb+*. The candidate insertions analysis revealed the presence of two high frequency S31 *Wb+*-specific *Doc* insertions (100% and 90.9%) in the Myosin81F gene, which encodes a myosin ATPase (CG45784) (Figure 4.3). Myosin ATPases are enzymes responsible for ATP hydrolysis, providing energy for actomyosin contraction (Rayment 1996). These two *Doc* insertions are located in opposite orientation and are positioned two exons apart within the Myosin81F gene: one sense *Doc* insertion occurs before exon 6, while one antisense *Doc* insertion is located after exon 7 (Figure 4.3). Ectopic recombination induced by inverted *Doc* insertions has previously been reported to disrupt the structure and expression of the *Antp* gene in *Antp73b* mutants (Schneuwly et al. 1987). Schneuwly and colleagues provided evidence for homologous recombination between inverted Doc insertions in close vicinity. Interestingly, the preliminary analysis of non-TE SV using GraffiTE (Groza et al. 2024) (comparing all *Wb+* populations against G0-*Wb+*) indicates that, out of nine <100 bp SVs detected in Myosin81F, six are indels specific to S31-*Wb+* (data not shown). This may suggest that the two oppositely oriented *Doc* elements could indeed affect the integrity of this genomic region, or it may simply reflect that the assembly of this region is more difficult to resolve due to the presence of these insertions. A more thorough sequencing may be required to clarify this genomic region. Furthermore, the 3D RNA conformation resulting from homologous recombination between the two inverted *Doc* elements could potentially affect RNA expression (Winter et al. 2018; Haws et al. 2022; Marie and Symington 2022). Myosin81F is expressed during developmental stages L3 and pupae as seen in Flybase (https://flybase.org/). The delay in development time of the S31 population, may be due to impaired motor activity. This impairment could hinder the larvae's ability to extract themselves from the substrate and initiate pupation, contributing to the delay. These insertions represent an intriguing subject for further study, given the correlation with the observed phenotype.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

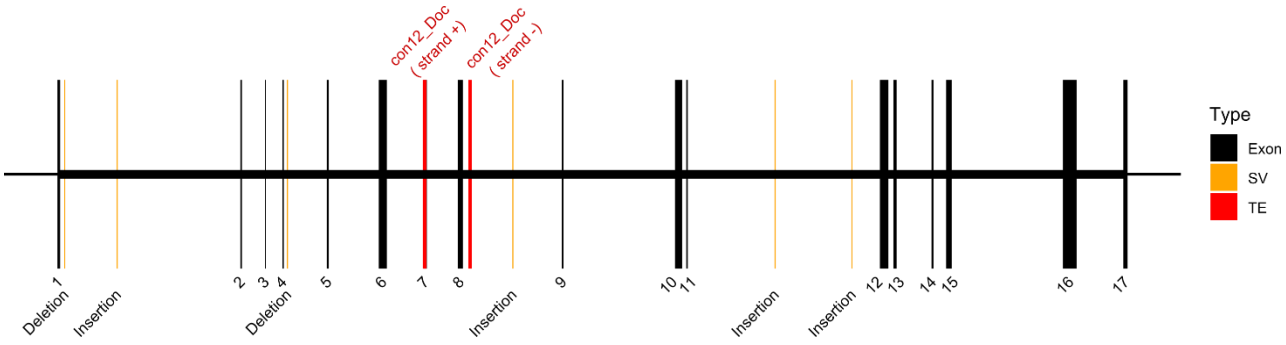


Figure 4. 3: Structure of the Myosin 81F gene in S31-Wb+.
The gene is approximatively 1.5 Mb long. Exons are indicated by black boxes, with their numbers labelled underneath. Only the S31-Wb+-specific TEs are shown (red lines), with their names and strand orientation indicated. The Doc element upstream of exon 6 is fixed in the population, while the second Doc element downstream of exon 7 is nearly fixed (90.91%). The insertion sizes are respectively 4737bp and 4725bp. Non-TE structural variants (SVs) are shown in orange, with labels indicating the type of SV.

Additionally, S31-*Wb+* also carries a *roo* insertion in the 5'UTR of gene CG2999, also known as Unc-13. The insertion is fixed in this population while it is detected at 6.89% in S73-*Wb+* and not detected-to-absent in S10-*Wb+* and S100-*Wb+*. Unc-13 encodes a protein that plays a central role in the assembly of the release sites of the vesicles at the apex of synapses, a process known as synaptic vesicle exocytosis (Dittman 2019). Therefore, it is essential to the proper functioning of synaptic transmission (Aravamudan et al. 1999). Synaptic transmission is the mechanism by which neuron-to-neuron or neuron-to-cell communication occurs, and as such is a vital process (Cotman and McGaugh 1980). If impaired, it could have deleterious consequences. *D. melanogaster* Unc-13 mutants often expressed defect in neurophysiology and larval locomotor behaviour (Böhme et al. 2016; Ramesh et al. 2021). Neither of these phenotypes were assessed in larvae in the present study, however, as hypothesized for the *Doc* insertions in the Myosin81F gene, perhaps an impaired larval locomotion could lead to an apparent increase in developmental rate. Larval locomotion may need to be investigated to elucidate this phenomenon.

In S73-*Wb+*, the candidate TE examination revealed an insertion in the 2Kb promoter region of a lncRNA involved in spermatogenesis (CR43802), reaching a frequency of 71.43% while absent from the other populations. Knockout of CR43802 using CRISPR-Cas9 results in a decrease in male fertility by up to 30% (Wen et al. 2016). As a reminder, the S73 population exhibits a decrease in egg hatchability, but unfortunately, we did not test for male fertility. A potential defect in spermatogenesis could lead to fewer functional spermatozoa, and females could lay both fertilized and unfertilized eggs, resulting in an overall decrease in hatching success.

It could be interesting to go a step further and investigate whether, beyond overall TE load and candidate insertions, specific insertions can be directly linked to plasticity, suggesting that phenotypic

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

plasticity may arise not only from cumulative TE content but also from the action of a single insertion acting as an enhancer of variation.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Conclusion

From genomic parasites or other selfish elements to integral components of the genome, ranging from deleterious to catalysts of evolution, the perception of TEs has come a long way. Our understanding of the various levels at which TEs interact with the genome and the organism continues to be deepened. In this study, we provide empirical support for yet another, previously hypothesized but largely untested, contribution of TEs to host phenotypes. First, our results support the idea that TEs broaden phenotypic expression, known as phenotypic variance, across a wide range of traits. But perhaps most importantly, and arguably the true achievement of this project, is our demonstration that TEs contribute to variation in phenotypic responses to environmental change, that is, plasticity.

We have also touched upon the impact that the endosymbiont *Wolbachia* may have on TEs and does have on the host phenotype. By doing so, we are contributing to extending our understanding of *Wolbachia*'s interaction with the host's TE dynamics, a field of research that currently has very little literature. Hopefully, this will spark some interest in the community to better study the interplay between symbiosis and TEs.

The living is composed of multiple interacting layers that occur at many scales, from the molecular to the cellular, to the multicellular level and its interaction with the environment. Only altogether do these layers and their connections, forming a unique alloy, enable organisms to integrate environmental cues into their responses, whether adaptive or not. In this study, we focus on how the molecular scale influences the phenotypic scale, notably also in relation to the presence or absence of symbionts.

A personal take-away

TEs, when adaptive and under selection, are often associated with strong one-insertion-one-phenotype effects. Observing their contribution at shorter evolutionary timescales is far more challenging. In many ways, we are digging in hopes of finding gold nuggets, biased by our focus on TEs, while in reality, most of the time there is only mud. Chapter one of this work may appear very assertive regarding the revolutionary role of TEs in phenotypic variation and plasticity. The measures we have performed likely do reflect the effects of TEs; however, these are also likely to be mostly deleterious when transposed to natural populations. This is an important reminder that highlighting the contribution of TEs does not mean reinforcing the view that they are solely beneficial. Rather, as emphasized earlier, TEs occur along a parasitic-to-symbiotic spectrum. The observation that S73 shows both reduced fertility and increased phenotypic variability is a strong example of this duality.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Women in Science: when McClintock meets West-Eberhard and Margulis

You thought I was done rambling! I promise this part will be brief. I just wanted to make a point in bringing up something that is close to my heart. I am a fervent defender of inclusivity and diversity, and I strongly believe that we cannot fully understand science (or anything as a matter of fact!) if research is conducted only by a subset of the society. Diversity in perspective enriches science and helps reduce inherent biases.

I am incredibly grateful to have been walking in the footsteps of three pioneering women in science, as to know Barbara McClintock, Mary Jane West-Eberhard and Lynn Margulis. I have also been incredibly lucky to grow as a scientist in an environment surrounded by women and colleagues from various backgrounds. While this may not be the place for personal acknowledgments, I want to use this place to encourage all women and underrepresented groups to embrace science and show that your place also is in the lab. Diversity makes science richer, fairer, and wholesome.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Other publications

Daniel S Oliveira, Marie Fablet, Anaïs Larue, Agnès Vallier, Claudia M A Carareto, Rita Rebollo, Cristina Vieira, ***ChimeraTE: a pipeline to detect chimeric transcripts derived from genes and transposable elements****, Nucleic Acids Research, Volume 51, Issue 18, 13 October 2023, Pages 9764–9784* https://doi.org/10.1093/nar/gkad671 (Annexe 2)

ChimeraTE is a tool designed to fill the computational gap in detecting the contribution of TEs to transcriptomic variability, specifically by assessing the number of transcripts containing sequences derived from both TEs and genes, *i.e.* chimeric transcripts. The pipeline was designed to address this question using paired-end short-read sequencing in two ways: one reference genome-guided (mode 1), and one genome-blind approach (mode 2). Each mode has its own strengths and caveats.

Mode 1 allows for the precise detection of where in the transcript the TE inserts and categorizes chimeric transcripts into TE-initiated, TE-exonized, and TE-terminated, while also providing both TE and gene identification. While this mode offers valuable information for downstream analyses – such as investigating the biological implications for the organism – it requires high-quality input data, particularly a well-annotated genome. Using this mode on four wild-type strains of *D. melanogaster* (two from Brazil and two from France), 327 genes were found to produce 408 chimeric transcripts. Once standardized across strains, this corresponds to 1.12% of all annotated genes (approximately 17,278 genes). This represents a non-negligible proportion of TEs being co-transcribed with genes. Most of these chimeras fell into the TE-exonized category (89%), with the majority of those insertions located within exons (61.84%), suggesting that TEs inserted in exons are more likely to be transcribed.

Mode 2 provides an alternative, though not mutually exclusive, approach to Mode 1 for quantifying and identifying TE-gene chimeras. Without relying on genome information, and thus lacking genomic context, this mode does not categorize chimeric transcripts into TE-initiated, TE-exonized, or TE-terminated. However, it enables the detection of chimeric transcripts that may result from polymorphic TE insertions: insertions absent from the reference genome. Using the same four wild-type *D. melanogaster* strains, 324 genes were found to produce 339 chimeric transcripts. When normalized, this corresponds to 1.07% of all genes. After aligning these detected chimeras to the *dm6* reference genome, 11 out of the 324 genes produced chimeric transcripts not present in dm6. Additionally, aligning these transcripts to the genomes of the wild-type strains revealed polymorphic chimeras between the Brazilian and French populations.

I participated in this work by conducting the experimental benchmarking of the pipeline, notably by designing primers to verify 10 polymorphic chimeras, and performing RT-PCR to validate their

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

presence. For five out of the 10 polymorphic chimeras, the experimental results were in complete agreement with the output of Mode 2. For the remaining five, the pipeline predicted presence in only one of the two populations from the same location, while experimental results supported their presence in both. This discrepancy is likely due to insufficient coverage, which may have prevented the insertion from being detected by the pipeline. Overall, these results indicate that Mode 2 can effectively predict polymorphic chimeric transcripts, so long as the genome assembly coverage is sufficient.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Mayeux, C.A., Larue, A., Oliveira, D.S., Varoqui, M., Henri, H., Rebollo, R., Kremer, N., Chambeyron, S. and Fablet, M., 2024. ***Intricate interactions between antiviral immunity and transposable element control in Drosophila.*** *BioRxiv, pp.2024-03* *https://www.biorxiv.org/content/10.1101/2024.03.18.585529v1* (Annexe 4)

In *Drosophila*, the regulation of TEs, as we have seen, is performed by small RNAs, including the siRNA pathway. The primary role of siRNA is the regulation of viral infections. This dual role raises the question of how these two regulatory functions interact, and whether there is a potential trade-off or a feedforward loop between them. To investigate this interaction, viral infection challenges can be performed to shed light on the relationship between TE regulation and antiviral defence.

In this work, the impact of viral infection with the *Drosophila C Virus* (DCV) was investigated in both *D. melanogaster* and *D. simulans*. Survival curves following infection revealed that *D. melanogaster* was more resistant than its sister species, *D. simulans*. RT-qPCR confirmed that, at 4 days post-infection (dpi), DCV levels were significantly higher in *D. simulans* (2.6-fold), likely explaining its reduced survival. To explore the underlying causes of this disparity in viral load, RNA and small RNA sequencing were conducted, focusing on the ovaries and carcasses of each species. RNA sequencing showed that DCV RNA was highly abundant in the carcasses, while only low levels were detected in the ovaries, possibly reflecting an experimental artefact rather than a true biological signal.

From the RNA sequencing data, TE expression was analysed for 237 TE families. This analysis revealed that TE expression undergoes a global increase upon DCV infection in *D. melanogaster* carcasses, while virtually no change in expression was detected in *D. simulans*. Interestingly, the opposite trend was observed in the ovaries of *D. melanogaster*, where TE expression decreased following infection. In contrast, once again, almost no modulation was detected in *D. simulans*.

To get to the bottom of the TE expression changes, small RNA (sRNA) sequencing was analysed. First, the piRNAs showed a signature of the ping-pong cycle, indicating that the piRNA pathway remains fully functional even under viral infection by DCV. Contrary to expectations, TE-derived sRNAs increased in *D. melanogaster* carcasses upon infection, an intriguing result, considering that TE transcripts also increased. We would indeed expect that an increase in sRNAs would correlate with a decrease in TE transcript levels. However, a relevant characteristic of DCV is that it encodes a suppressor of RNA interference (protein 1A) which protects its own genetic material from the host's antiviral response. Mayeux and collaborators hypothesize that the increase in TE expression, despite the rise in sRNAs, could result from TE transcripts hijacking 1A. To investigate this hypothesis, similar experiments were conducted on none other than G0-*Wb*- and S73-*Wb*-. If TEs indeed hijack 1A to evade sRNA-mediated silencing,

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

possibly to the detriment of DCV, then one would expect that the more TEs present, the lower the DCV titer. And indeed, S73, which carries more TE, exhibits a lower DCV titer than G0.

Currently, Julies Delescluse, a postdoctoral researcher in Marie Fablet's laboratory, is conducting immunoprecipitation of the 1A protein in an effort to investigate the proportions of TE *vs*. DCV transcripts bound to it. I contributed to the experimental aspects of this work, including viral amplification, survival monitoring, and RNA/sRNA extractions.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Oliveira, D.S., Larue, A., Nunes, W.V.B., Sabot, F., Bodelón, A., García Guerreiro, M.P., Vieira, C. and Carareto, C.M.A., 2024. ***Transposable elements contribute to the evolution of host shift-related genes in cactophilic Drosophila species.*** *BioRxiv, pp.2024-03* https://doi.org/10.1101/2024.03.27.587021 (Annexe 5)

Using Nanopore long-read sequencing, high-resolution genome assemblies were obtained for several species of the *Drosophila repleta* group that exploit different cactus hosts, spanning a gradient from less toxic (*Opuntia sp.*) to more toxic columnar cacti. These included *D. buzzatii* (*Opuntia* dweller), *D. koepferae* (columnar cacti dweller), *D. arizonae* (*Opuntia* or columnar cacti dweller), *D. mojavensis mojavensis* (*Ferocactus cylindraceus* dweller), *D. mojavensis wrigleyi* (*Opuntia* dweller), and *D. mojavensis sonorensis* (*Stenocereus thurberi* dweller). *De novo* gene annotation was performed for *D. buzzatii* and *D. koepferae*, and orthologous relationships were computed across species. Genes involved in host usage, including host localization, acceptance, and utilization, were the main focus of the study.

To investigate whether these genes were under different selective trajectory (relaxed, positive, negative), a branch-length method was applied, comparing *repleta* group species specialized on toxic cacti to those feeding on the ancestral host (*Opuntia* sp.). This analysis, conducted on orthologous genes shared across all species (n = 130), identified six genes under relaxed negative selection: *Obp56d, Obp19d*, *Gr5a, Ir25a, Cyp4s3*, and *esterase S.* This evolutionary signature might allow for diversification to occur.

From a TE perspective: the TE libraries were created for *D. buzzatii* (also used for *D. koepferae*) and *D. m mojavensis* (used for *D. arizonae* and all *D. mojavensis* subspecies). These libraries enabled quantification and annotation of TEs in the focal genomes. LTRs and *Helitrons* were the most abundant TE families in the *mojavensis-arizonae* group. Notably, *D. m mojavensis* carried the highest TE content (22.01%), driven by expansion of LTRs and *Helitrons*, as reflected by copy number. Analysis of TE insertions in host-acceptance and localization genes revealed enrichment in specific cases. For example, *D. koepferae* showed enrichment of TEs in the promoters of gustatory receptor (Gr) genes. By contrast, no such enrichment was observed in promoters of olfactory receptor (OR) genes compared to other TE-rich regions, suggesting that OR promoters are not preferential targets of TE insertions or are not retained. However, further investigation of TEs located near OR genes identified TFBS corresponding to previously described OR regulators. Comparison with other gene categories revealed that OR genes harbour more TE-derived TFBS, with ≈50% of these originating from *Helitrons*. Expression analysis of host-acceptance and localization genes showed differential expression between the *mojavensis-arizonae* group and the *buzzatii-koepferae* group. However, these expression changes were not directly associated with the TE enrichment observed in OR genes. The chimeric gene-TE transcript analysis (based on paired-end reads)

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

revealed that the *mojavensis-arizonae* group expresses 11 times more chimeric transcripts than the *buzzatii-koepferae* group. This suggests either that the former is more prone to producing chimeras or that the latter is under stronger selection against TE-driven chimeras. Interestingly, the *Opuntia* users (*D. buzzatii* and *D. m. wrigleyi*) showed the highest number of species-specific chimeric transcripts, with 468 from host localization and usage genes.

Although this study did not pinpoint a single mode of TE-driven adaptation underlying the host shift, the results highlight multiple ways in which TEs may have contributed to evolutionary divergence among these species. First, TEs were enriched in promoters of OR genes containing OR-specific TFBS. Second, most of these TEs were *Helitrons*, a novel observation in *Drosophila*. Third, while TE insertions did not systematically alter gene expression, they were frequently involved in the formation of chimeric transcripts, accounting for 1.31% of the transcriptome, including in genes involved in host localization and usage. Altogether, these findings support a role for TEs in driving divergence in the *repleta* group during cactus host specialization.

I have contributed to this work on the technical aspects, including maintaining the fly stock, performing the dissections and the RNA extractions.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Résumé

La plasticité phénotypique désigne la capacité d'un organisme doté d'un génotype donné à produire des phénotypes distincts en réponse aux changements environnementaux. Dans un contexte adaptatif, cette plasticité peut améliorer la potentialité d'un organisme à faire face aux fluctuations environnementales. Les éléments transposables (ET), bien que souvent négligés dans ce contexte, peuvent jouer un rôle important dans la plasticité phénotypique grâce à leur capacité à moduler l'expression génétique en fonction de l'environnement. Les ET sont des séquences d'ADN mobiles et répétitives. Leur activité est régulée par des mécanismes épigénétiques, mais aussi par leurs propres séquences régulatrices, tous deux sensibles aux variations environnementales. Ces fluctuations peuvent entraîner une expression ou une mobilisation différentielle des ET, influençant l'expression génique et induisant une plasticité épigénomique, génomique et transcriptomique. Une telle plasticité au niveau moléculaire pourrait conduire à une diversité interindividuelle et promouvoir une réponse plastique aux changements environnementaux. Afin d'étudier cette hypothèse dans des conditions contrôlées, nous avons étudié des mouches *Drosophila melanogaster* génétiquement modifiées ayant un contenu en ET différent, mais partageant un patrimoine génétique commun. Le criblage phénotypique a révélé des différences significatives entre les populations. Une étude des phénotypes dans divers environnements a révélé des variations dans les réponses plastiques et une augmentation de la variance des traits pour la population la plus chargée en ET. Ces résultats suggèrent que les ET contribuent à l'hétérogénéité de la réponse environnementale et peuvent donc constituer un mécanisme sous-jacent à la plasticité phénotypique.

I. Introduction

La plasticité phénotypique désigne la capacité d'un même génome à produire différents phénotypes, c'est-à-dire l'ensemble des traits observables ou mesurables d'un organisme, en réponse à des variations environnementales (Bradshaw 1965; West-Eberhard 2003; Ghalambor et al. 2007). Ce phénomène peut conférer un bénéfice adaptatif en diversifiant l'éventail des phénotypes exprimables et donc soumis à la sélection naturelle (Pigliucci 2001; Pfennig 2021c). La plasticité phénotypique est ainsi considérée comme un catalyseur de l'évolution (Pigliucci 2001; Pfennig 2021c). Toutefois, les mécanismes sous-jacents permettant ces fluctuations phénotypiques demeurent une question centrale en biologie et en écologie. Une grande partie de cette plasticité peut être attribuée à des modifications de l'expression génique, souvent impliquant plusieurs gènes simultanément (pléiotropisme) (Scheiner 1993; W. G. Hill

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

and Mackay 2004). Les outils utilisés pour l'étudier combinent l'analyse des normes de réaction (Woltereck 1909) et l'étude de la variation phénotypique par le biais de la génétique quantitative, comme proposé dès 1960 par Falconer (Falconer 1960; 1996; W. G. Hill and Mackay 2004). Dans ce cadre, la variation phénotypique totale (VP) est décomposée en trois composantes : la variation d'origine génétique (VG), la variation due à l'environnement (VE) et la variation issue de l'interaction génotype-environnement (VGxE). Cette dernière composante, VGxE, est la condition pour que la plasticité soit héritable et donc puisse contribuer à l'évolution des organismes (Via and Lande 1985). Elle correspond à la variabilité génétique de la plasticité elle-même. Ainsi, étudier la plasticité (VE et VGxE) nécessite de réduire l'effet de VG afin d'isoler et de mieux comprendre l'influence de l'environnement et des interactions génotype-environnement.

Ce projet vise à comprendre quels mécanismes peuvent induire des modifications de l'expression génique et conduire à l'émergence de variations phénotypiques en réponse à des changements environnementaux. Dans ce cadre, les éléments transposables (ET) sont souvent cités comme de potentiels acteurs de la plasticité phénotypique, bien que les preuves empiriques restent limitées. Les ET sont des séquences d'ADN présentes dans le génome de la plupart des organismes et se distinguent par leur capacité à se déplacer et à se multiplier au sein du génome, un phénomène mutagène appelé transposition. La découverte de ces séquences mobiles remonte aux travaux pionniers de Barbara McClintock dans les années 1950 sur le maïs (McClintock 1950; 1953; 1956).

Les ET sont classés en deux grandes catégories selon leurs mécanismes de transposition, et se subdivisent en superfamilles puis en familles, en fonction de leur relation phylogénétique, de leur structure et de leurs composants moléculaires (Wicker et al. 2007; Wells and Feschotte 2020). La classe I, ou rétroéléments, se caractérise par l'utilisation d'une molécule d'ARN comme intermédiaire à la transposition. Le mécanisme est de type "copier-coller", produisant une nouvelle copie qui s'intègre ailleurs dans le génome (Eickbush and Malik 2007). Parmi ces rétroéléments, on distingue notamment les LTR rétrotransposons, dont certains partagent une forte similitude avec les rétrovirus, allant parfois jusqu'à la capacité de former des particules de type virale (VLPs) capables d'envahir des cellules voisines (Eickbush and Malik 2007; Goodier and Kazazian 2008). Les non-LTR rétrotransposons constituent l'autre grande catégorie de rétroéléments, représentée principalement par les LINEs (Long Interspersed Nuclear Elements). Ces derniers sont particulièrement abondants dans le génome humain. Leur transposition s'effectue par un mécanisme appelé TPRT (Target-Primed Reverse Transcription), caractérisé par une tendance fréquente à la troncation en 5' lors de l'intégration de la nouvelle copie dans l'ADN (Eickbush 1992; Eickbush and Malik 2007; Grandi and An 2013). La seconde catégorie, dite classe II ou éléments

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

transposables à ADN, repose principalement sur un mécanisme de type “couper-coller”, au cours duquel l’élément est excisé du site donneur pour être réinséré dans un nouveau site cible (Feschotte and Pritham 2007). Ces éléments sont souvent caractérisés par des inverted terminal repeats (TIRs) à leurs extrémités. Outre les transposons à TIR, on retrouve également des familles particulières comme les *Helitrons*, qui se déplacent par un mécanisme de réplication en “cercle roulant”, ou encore les *Mavericks/Polintons*, de grande taille, et ayant la particularité d’encodé pour leur propre polymérase (Kapitonov and Jurka 2006; Thomas and Pritham 2015).

La majorité des événements de transposition sont neutres ou délétères pour l’organisme, ce qui a conduit à l’évolution de divers mécanismes de régulation. Parmi ces mécanismes épigénétiques, la méthylation de l’ADN est un acteur majeur, permettant de réguler l’activité des éléments transposables dans de nombreux organismes (Martienssen and Richards 1995; Lippman et al. 2004). Cette modification chimique bloque la transcription des ET en favorisant la compaction de la chromatine, et constitue ainsi une barrière efficace contre leur mobilisation (Martienssen and Richards 1995; Lippman et al. 2004). Toutefois, chez *Drosophila melanogaster*, la méthylation de l’ADN est quasi absente, et la régulation repose principalement sur un autre mécanisme : la voie des piARN (PIWI-interacting RNAs) (Lyko et al. 2006; Zamore 2010; Czech and Hannon 2016; Czech et al. 2018). Ces petits ARN non codants dérivent de régions génomiques enrichies en copies d’ET, appelées *piRNA clusters* (Brennecke et al. 2007). Associés aux protéines de la famille PIWI, ils assurent une répression des ET à la fois post-transcriptionnelle, en dégradant leurs ARN, et transcriptionnelle, en induisant la formation d’hétérochromatine répressive par des modifications d’histones, notamment par la déposition de marques H3K9me3 (Czech and Hannon 2016; Czech et al. 2018; Choi 2024).

Néanmoins, les ET ne se limitent pas à des effets délétères. De nombreuses études ont montré qu’ils peuvent également jouer un rôle dans la régulation de l’expression génique de l’hôte et, dans certains cas, contribuer à l’émergence de caractères adaptatifs (Casacuberta and González 2013; Rey et al. 2016a; Gilbert et al. 2021b; Cabral-de-Mello and Palacios-Gimenez 2024). Un exemple emblématique est celui de la phalène du bouleau (*Biston betularia*), chez qui l’insertion d’un ET dans un gène contrôlant la pigmentation alaire entraîne une surexpression conduisant à un morphotype mélanique, mieux adapté aux environnements recouverts de suie durant la révolution industrielle (Van’t Hof et al. 2016). Ce cas illustre comment l’insertion d’un ET peut remodeler l’expression génique et générer une variabilité phénotypique soumise à la sélection naturelle. Mais cela ne se limite pas à l’effet direct des ET, mais peut aussi être due à la propagation des marques épigénétiques, en charge de leur contrôle, qui peut s’étendre aux gènes environnants, modifiant ainsi leur expression (Muller 1930; Lippman et al. 2004; Elgin and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Reuter 2013; Lee and Karpen 2017). De plus, ces régulations épigénétiques sont dynamiques et sensibles aux conditions environnementales, ce qui implique que l'activité des ET, par leur répression et leurs effets sur les gènes hôtes, peut varier en fonction du contexte environnemental (Jaenisch and Bird 2003; Slotkin and Martienssen 2007; Chuong et al. 2017). Pourtant, peu d'études se sont attachées à analyser l'ensemble des ET dans un génome et leur influence sur l'expression des gènes en réponse aux changements environnementaux et les phénotypes pouvant en résulter. Notre objectif est d'apporter la première preuve du rôle des ET dans la plasticité phénotypique chez les insectes. Dans ce contexte, l'étude des ET apparaît particulièrement pertinente pour mieux comprendre les mécanismes moléculaires qui sous-tendent leur remarquable adaptabilité.

Un autre facteur pouvant influencer l'expression phénotypique et la réponse des organismes à leur environnement est la présence de symbiotes (Klepzig et al. 2009). La symbiose est un phénomène biologique complexe que l'on peut décrire comme un continuum allant du parasitisme au mutualisme, en passant par le commensalisme, selon la nature de l'interaction entre les partenaires (Margulis 1938; Moran 2006; Paracer and Ahmadjian 2000; Dobson et al. 2008). Dans ce cadre, l'hôte correspond à l'entité la plus grande de l'interaction, tandis que le symbiote est souvent un organisme unicellulaire vivant à la surface ou à l'intérieur de l'hôte (Margulis 1938; Moran 2006; Paracer and Ahmadjian 2000; Dobson et al. 2008). Lorsque le symbiote tire profit de l'interaction au détriment de l'hôte, on parle de parasitisme, alors que si les deux partenaires en retirent un bénéfice réciproque, on parle de mutualisme (Margulis 1938; Moran 2006; Paracer and Ahmadjian 2000; Dobson et al. 2008).

Les symbiotes peuvent contribuer de manière significative à la biologie de leur hôte, notamment par la complémentation nutritionnelle, l'amélioration de la résistance aux pathogènes ou encore l'influence sur certains comportements, des effets qui peuvent eux-mêmes varier en fonction de l'environnement (Goodacre et al. 2009; Bordenstein and Theis 2015; Oliver and Higashi 2019; Duron and Gottlieb 2020). Chez les insectes, l'un des symbiotes les plus étudiés est *Wolbachia*, une alphaprotéobactérie intracellulaire (endosymbiote) facultative très répandue (Werren et al. 2008; Landmann 2019). *Wolbachia* est connue pour induire des phénomènes de stérilité cytoplasmique, caractérisés par l'incompatibilité entre individus infectés et non infectés, mais aussi pour conférer une résistance accrue aux infections virales (Zheng et al. 2011; Chrostek et al. 2013; Pimentel et al. 2021; Bruner-Montero and Jiggins 2023).

Plus récemment, il a été rapporté que *Wolbachia* peut également influencer la dynamique des ET, en modulant leur expression, affectant leur transmission maternelle et altérant la quantité de piARN impliqués dans leur régulation (Mayoral et al. 2014; Touret et al. 2014; Eugénio et al. 2023; Mostoufi and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Singh 2024). Compte tenu de ces observations, nous avons cherché à évaluer l'influence de la présence de *Wolbachia* sur l'interaction entre les ET et l'environnement, afin de mieux comprendre son rôle potentiel dans la plasticité phénotypique des insectes.

II. Matériels et méthodes

Pour répondre à cette problématique, nous travaillons avec un organisme modèle bien connu : la mouche du vinaigre, *Drosophila melanogaster*. Les populations de *D. melanogaster* sur lesquelles nous travaillons sont génétiquement modifiées et maintenues en conditions contrôlées. Elles se caractérisent par des contenus en ET distincts, tout en partageant un fond génétique identique (Barckmann et al. 2018; Varoqui et al. 2025). De la population G0, représentant le contenu d'ET initial, s'ensuivent les populations S10, S31, S73 et s100, ayant subi respectivement 10, 31, 73 and 100 générations de dérépression et donc d'accumulation des ET. Cette approche nous garantit que les différences observées lors des expérimentations peuvent être exclusivement attribuées aux ET. Un génotypage approfondi de cinq populations a été réalisé à l'aide de la technologie Oxford Nanopore, qui génère des séquences de fragments d'ADN de grande longueur. Grâce à cette approche, nous avons pu quantifier et localiser précisément les ET dans les populations d'intérêt avec l'aide de l'outil TrEMOLO (Mohamed et al. 2023). Ces populations ont été soumises à des tests phénotypiques, incluant des mesures de fertilité, de longévité et de taux métabolique, certains sous différentes conditions environnementales (température, régime alimentaire, stress oxydatif). Les données génétiques et phénotypiques sont analysées afin de tester une corrélation entre le contenu en ET et la plasticité des différentes populations. De même manière, les cinq populations, après avoir subies un traitement à la tétracycline contre l'endosymbiote *Wolbachia*, ont été séquencées et soumis à des tests phénotypiques en conditions contrôle et sous différents environnements.

III. Résultats et discussion

Les résultats de séquençage confirment l'accumulation d'ET dans un fond génétique identique, permettant l'évaluation de la variation de la plasticité en condition contrôlée. Une analyse des caractéristiques phénotypiques révèle que les populations, bien que ne divergeant que par leur contenu en ET, arborent des profils phénotypiques distincts. Par exemple, les populations S31 et S73 sont sujettes à une réduction de l'éclosion des œufs, ce qui est aussi accompagné chez la S73 par une réduction de sa viabilité. Les tests de plasticité réalisés montrent que certains traits varient en fonction de

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

l'environnement et du contenu en ET, démontrant une interaction GxE. Par exemple, les données de longévités sur différents milieux nutritifs témoignent d'une interaction génotype-environnement. Les tests de survie à différentes concentrations de stress oxydatif induite par la présence de paraquat suivent une tendance similaire. Nous pouvons donc avancer que ces expérimentations témoignent d'une interaction TE-environnement. Un autre trait, le taux métabolique, présente une plasticité qui dépend uniquement du facteur environnemental température (VE), sans lien direct avec le contenu en ET. Cette étude montre aussi que la population ayant la plus forte charge en ET (S73) présente une surprenante variabilité individuelle dans la plupart des traits mesurés, supporté par une moyenne de ces coefficients environnementaux (CV) à la hausse. Ces résultats illustrent que les ET peuvent induire une variabilité phénotypique, qui plus est peut être influencé par l'environnement.

L'investigation de l'influence de *Wolbachia* sur cette relation est encore très préliminaire, mais les résultats suggèrent que certains phénotypes et certaines réponses à l'environnement sont effectivement sous l'effet de la triple interaction TE-environnement-*Wolbachia*. Ceci porte à approfondir ces observations.

IV. Conclusion

Cette étude supporte de la contribution de l'ensemble des ET d'un génome, dans la réponse d'un organisme confronté à des changements environnementaux. La variabilité générée par les ET pourrait, dans certain contexte, représenter un avantage évolutif, en modulant l'expression des gènes et en contribuant à la plasticité phénotypique. Ce projet permet de mieux comprendre le rôle des ET dans l'adaptabilité des organismes. Par ailleurs, l'interaction avec *Wolbachia* souligne la complexité des mécanismes qui permettent aux organismes de répondre aux variations environnementales, soulignant l'importance des facteurs génomiques et symbiotiques dans la régulation phénotypique.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

S2. Supplementary material for Chapter 1

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

	Complete	Single Copy	Duplicated	Fragmented	Missing	Scaffolds	Contigs	Total length	% gaps	N° de contigs	N50	GC %
G0_Wb+	99.70%	99.20%	0.50%	0.10%	0.20%			134182872		256	24220491	4205.00%
G0-F100_SC	99.80%	99.20%	0.50%	0.10%	0.10%	77	652	142523898	0.04	77	25540886	42.14

Table S2. 1: Assembly completeness and quality metrics for G0-Wb+ reference genomes.
Assembly completeness was assessed using BUSCO, reporting percentages of complete, single-copy, duplicated, fragmented, and missing genes. Assembly contiguity and quality are summarized with the number of scaffolds and contigs, total assembly length, percentage of gaps. The grey boxes indicate that BUSCO did not provide with the assembly statistics. The contig count, N50, and GC content were assessed with Quast. Results are shown for G0_Wb+ and G0-F100_SC assemblies (Mohamed et al. 2023).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

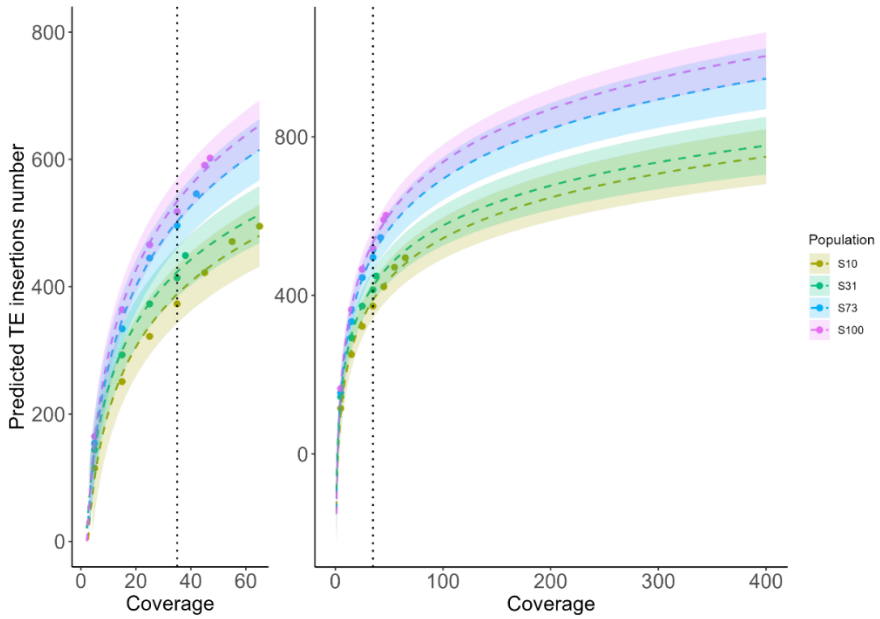
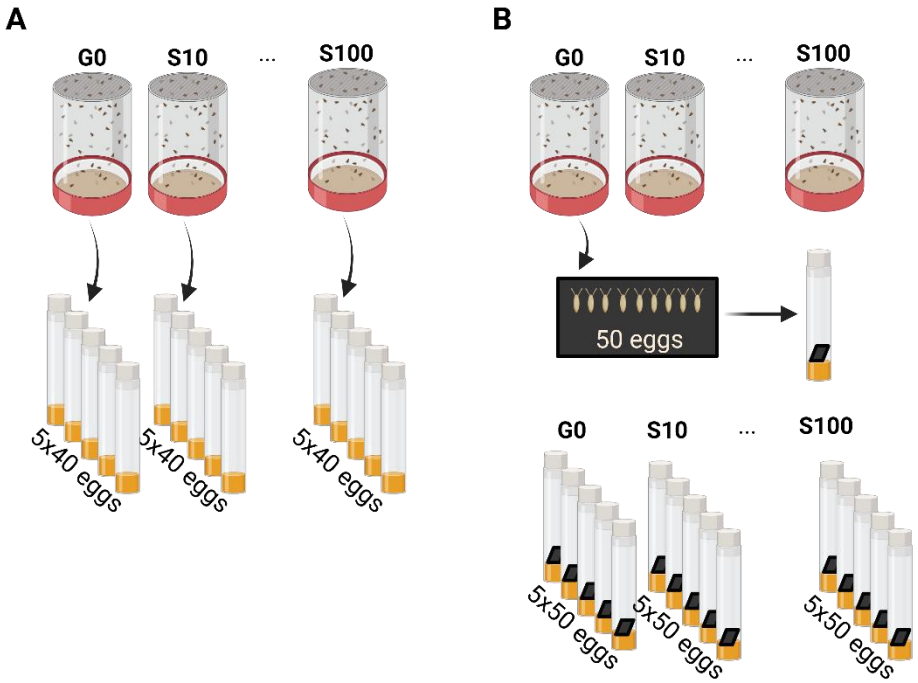


Figure S2. 1: Rarefaction curves of new TE insertions detected by TrEMOLO across populations.
The number of predicted TE insertions is plotted as a function of sequencing coverage using a logarithmic prediction function. Each dashed line represents the fitted rarefaction curve for one population, with shaded areas indicating confidence intervals. Dots correspond to observed values. The vertical dotted lines indicate selected coverage thresholds, namely 35 X. The left panel provides a zoomed-in view of the curves at lower coverage values, corresponding to the initial region of the right panel.



Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Figure S2. 2: Experimental design for the viability and hatchability assays.
A) For the viability experiment, adults were allowed to oviposit for a few hours in the morning, and the collected progeny were distributed into five replicate vials with 40 eggs for each condition. The number of emerging adults per vial was recorded to estimate viability. B) For the hatchability experiment, 50 eggs were collected from each cage, transferred into replicate vials, and monitored for hatching success by observation of the eggs laid on a black piece of paper. Hatchability was estimated as the proportion of eggs that successfully hatched across replicates.

Population	Sex	Temperature	Effective number
G0	F	18	120
G0	F	22	111
G0	M	18	78
G0	M	22	100
S10	F	18	130
S10	F	22	103
S10	M	18	134
S10	M	22	102
S31	F	18	113
S31	F	22	137
S31	M	18	111
S31	M	22	93
S73	F	18	103
S73	F	22	100
S73	M	18	101
S73	M	22	80
S100	F	18	115
S100	F	22	115
S100	M	18	120
S100	M	22	90

Table S2. 2: Number of flies assayed for longevity in the* Wolbachia*-infected populations under different temperature.
For each population, the table reports the sex of the flies (F: female, M: male), the rearing temperature (°C), and the effective number of individuals for which longevity data were collected.

Population	Sex	Medium	Effective number
G0	F	Agar	104
G0	F	-Yeast	150
G0	F	Complete	117
G0	F	-Sugar	52
G0	M	Agar	100
G0	M	-Yeast	108
G0	M	Complete	116

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

G0	M	-Sugar	53
S10	F	Agar	105
S10	F	-Yeast	100
S10	F	Complete	120
S10	F	-Sugar	57
S10	M	Agar	117
S10	M	-Yeast	119
S10	M	Complete	112
S10	M	-Sugar	59
S31	F	Agar	107
S31	F	-Yeast	150
S31	F	Complete	149
S31	F	-Sugar	62
S31	M	Agar	104
S31	M	-Yeast	120
S31	M	Complete	102
S31	M	-Sugar	65
S73	F	Agar	102
S73	F	-Yeast	129
S73	F	Complete	135
S73	F	-Sugar	56
S73	M	Agar	103
S73	M	-Yeast	124
S73	M	Complete	110
S73	M	-Sugar	51
S100	F	Agar	106
S100	F	-Yeast	121
S100	F	Complete	131
S100	F	-Sugar	56
S100	M	Agar	117
S100	M	-Yeast	104
S100	M	Complete	100
S100	M	-Sugar	60

Table S2. 3: Number of flies assayed for longevity in the Wolbachia-infected populations on different rearing media.

For each population, the table reports the sex of the flies (F: female, M: male), the rearing media, and the effective number of individuals for which longevity data were collected. The "Complete" medium contains both sugar and yeast. The "-Yeast" medium was identical to the complete medium but without yeast, while the "-Sugar" medium lacked sugar. The "Agar" medium consists only of water and preservative, and thus provided no nutritional value.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Population	Sex	Concentration	Effective number
G0	F	15mM	98
G0	F	25mM	103
G0	M	15mM	99
G0	M	25mM	105
S10	F	15mM	101
S10	F	25mM	106
S10	M	15mM	101
S10	M	25mM	100
S31	F	15mM	100
S31	F	25mM	102
S31	M	15mM	97
S31	M	25mM	106
S73	F	15mM	102
S73	F	25mM	103
S73	M	15mM	92
S73	M	25mM	101
S100	F	15mM	104
S100	F	25mM	103
S100	M	15mM	111
S100	M	25mM	106

Table S2. 4: Number of flies assayed for longevity in the Wolbachia-infected populations exposed to variable concentration of paraquat.

For each population, the table reports the sex of the flies (F: female, M: male), the paraquat concentration (15mM and 25mM), and the effective number of individuals for which longevity data were collected.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

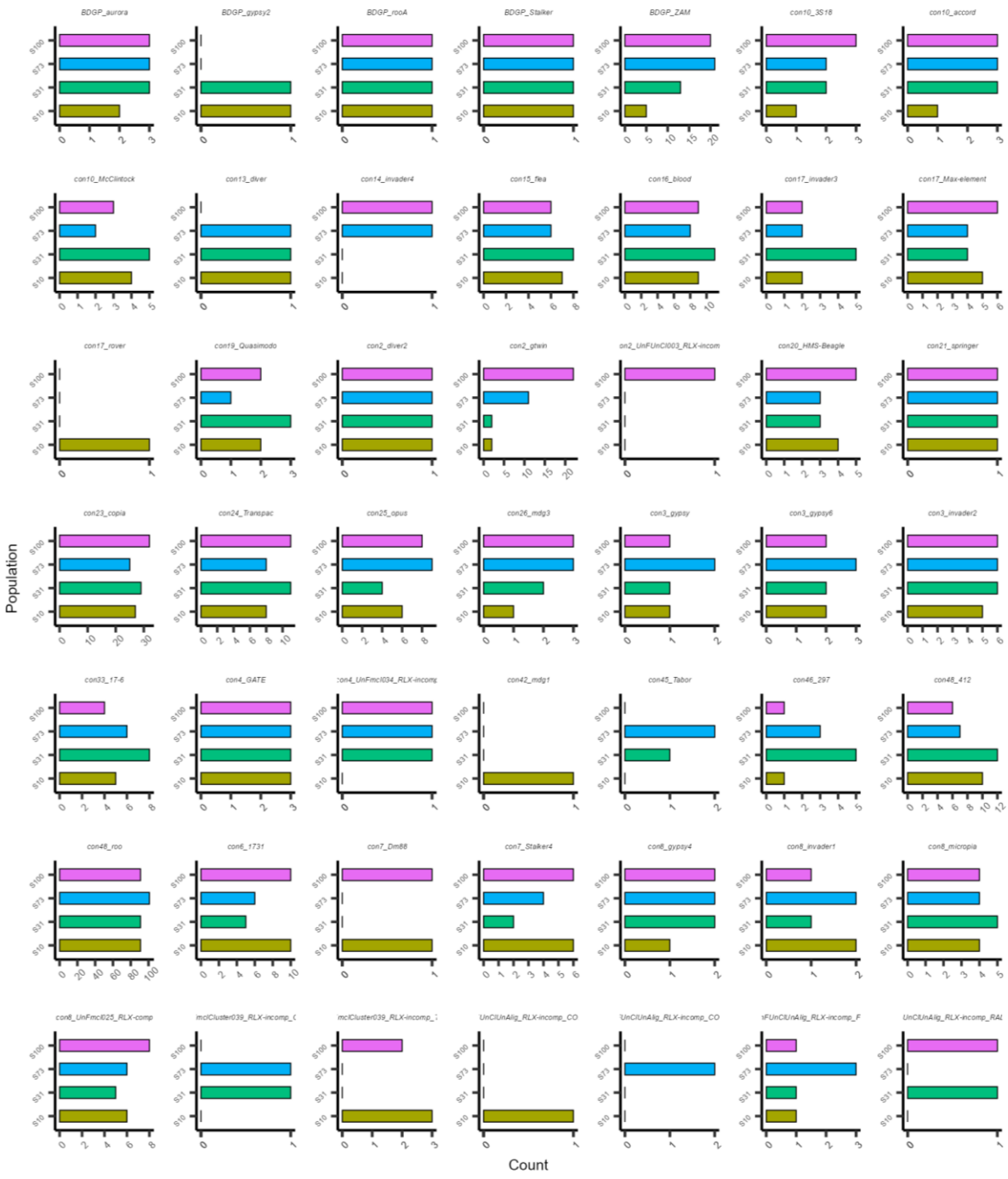


Figure S2. 3: Newly detected LTR retrotransposon insertions relative to G0-Wb+ across consensus library families.

Each bar represents the count of insertions identified in the sample that are absent in the G0-Wb+ reference, as detected by TrEMOLO. Insertions are categorized by family according to the consensus library.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

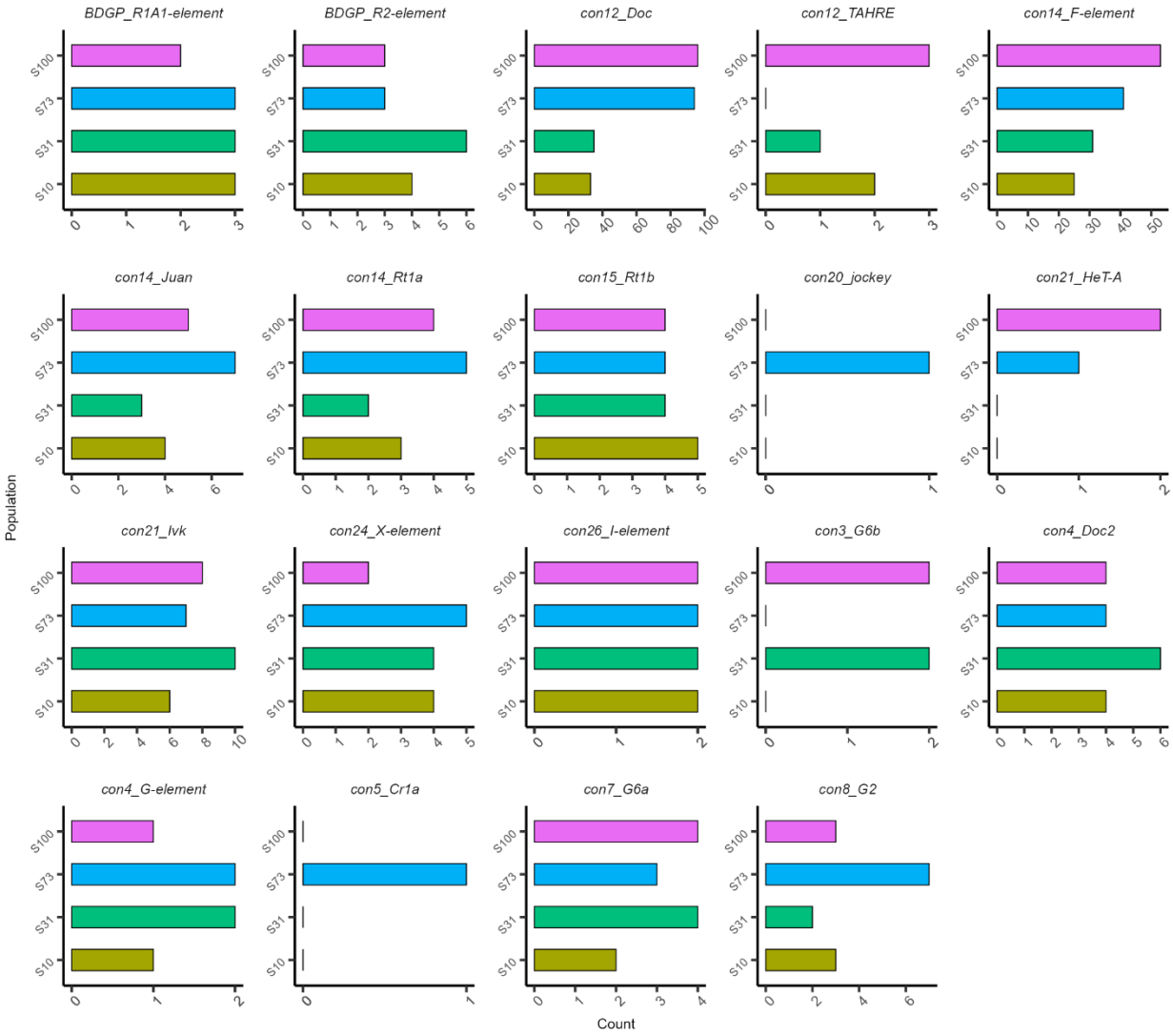


Figure S2. 4: Newly detected LINE transposable element insertions relative to G0-Wb+ across consensus library families.

Each bar represents the count of insertions identified in the sample that are absent in the G0-Wb+ reference, as detected by TrEMOLO. Insertions are categorized by family according to the consensus library.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

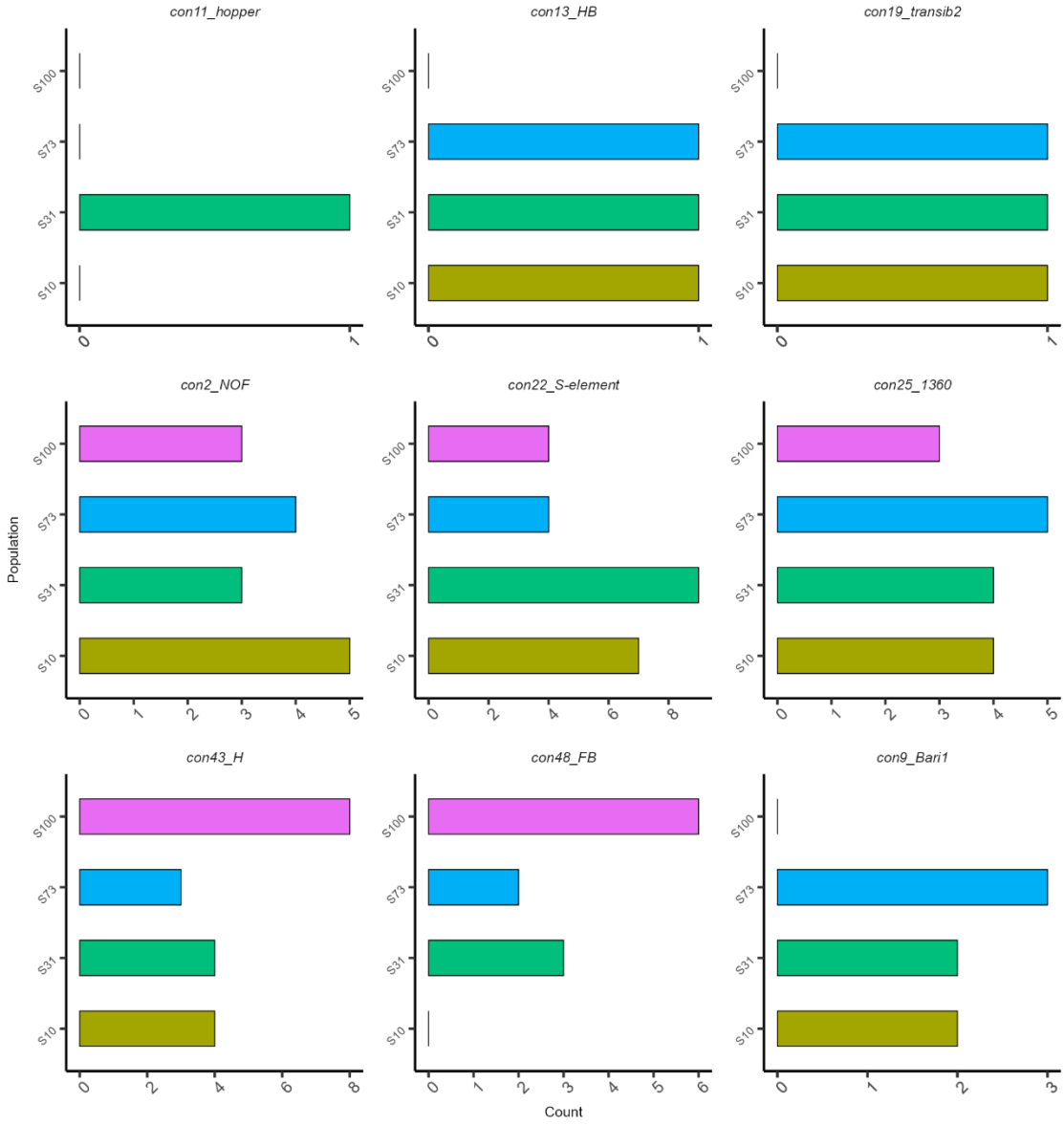
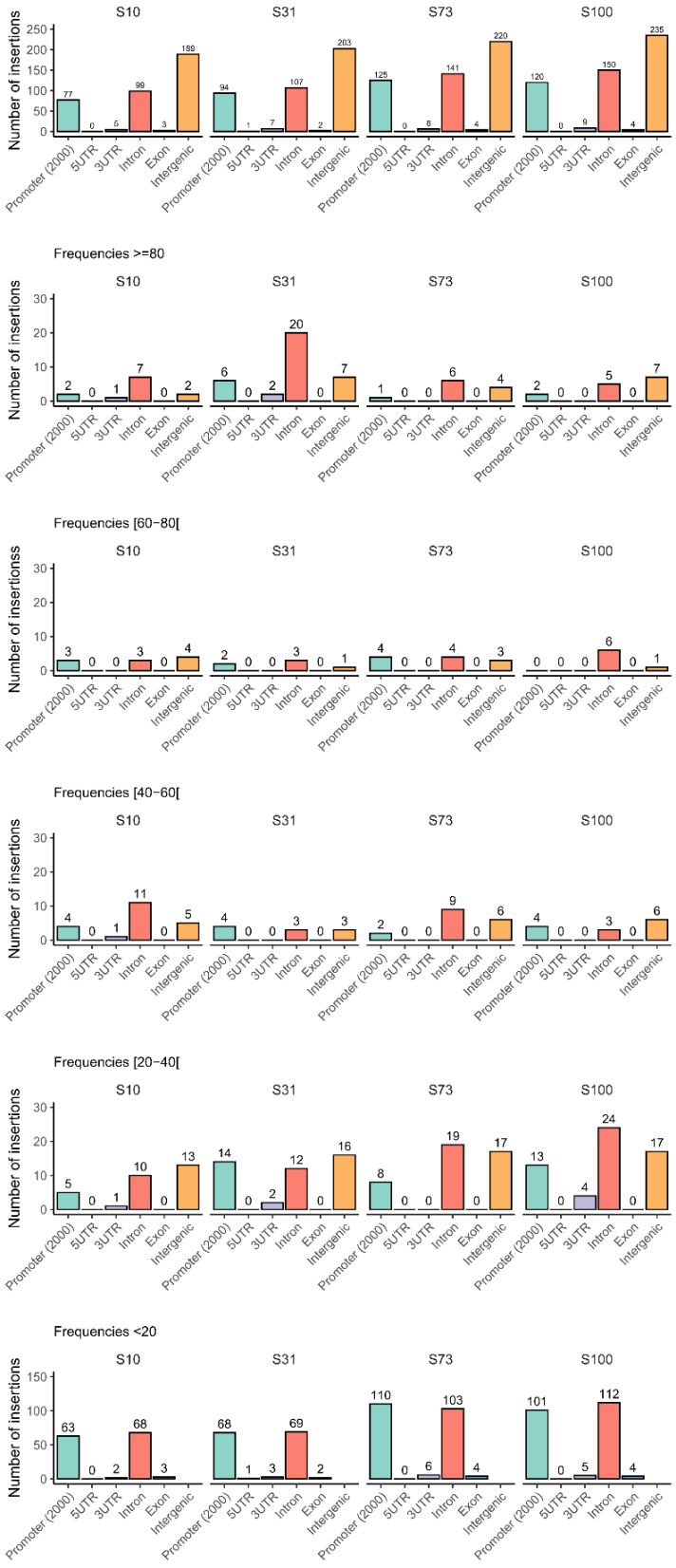


Figure S2. 5: Newly detected DNA transposon insertions relative to G0-Wb+ across consensus library families.

Each bar represents the count of insertions identified in the sample that are absent in the G0-Wb+ reference, as detected by TrEMOLO. Insertions are categorized by family according to the consensus library.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés



Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Figure S2. 6: Genomic distribution of transposable element insertions detected relative to the G0-Wb+ reference.

Bars represent the number of insertions identified in different genomic regions (promoters (2000 bp upstream); UTRs; exons; introns; intergenic) that are absent in the G0-Wb+ reference, as detected by TrEMOLO. The insertions are grouped by frequency categories: the top panel shows all insertions combined, and subsequent panels (from top to bottom) display insertions from the most abundant to the rarest categories (see panel titles for details).

Genomic region	Population	Insertion frequency category	Expected count	Observed count	Bidirectional test p-value
Promoter (2000)	S10	-20	112.98	63	0.002
5UTR	S10	-20	0.86	0	0.8392
3UTR	S10	-20	8.81	2	0.02
Intron	S10	-20	88.64	68	0.014
Exon	S10	-20	19.45	3	0.002
Promoter (2000)	S100	-20	160.09	101	0.002
5UTR	S100	-20	1.23	0	0.5654
3UTR	S100	-20	12.34	5	0.036
Intron	S100	-20	125.22	112	0.1558
Exon	S100	-20	27.51	4	0.002
Promoter (2000)	S31	-20	119.75	68	0.002
5UTR	S31	-20	0.93	1	1
3UTR	S31	-20	9.36	3	0.046
Intron	S31	-20	93.82	69	0.008
Exon	S31	-20	20.7	2	0.002
Promoter (2000)	S73	-20	155.15	110	0.002
5UTR	S73	-20	1.2	0	0.5814
3UTR	S73	-20	11.94	6	0.1119
Intron	S73	-20	121.47	103	0.048
Exon	S73	-20	26.64	4	0.002
Promoter (2000)	S10	80	4.45	2	0.2478
5UTR	S10	80	0.04	0	1
3UTR	S10	80	0.35	1	0.6054
Intron	S10	80	3.49	7	0.0799
Exon	S10	80	0.85	0	0.8511
Promoter (2000)	S100	80	5.22	2	0.1259
5UTR	S100	80	0.04	0	1
3UTR	S100	80	0.4	0	1
Intron	S100	80	4.09	5	0.7572
Exon	S100	80	0.96	0	0.7612
Promoter (2000)	S31	80	13.04	6	0.022

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

5UTR	S31	80	0.12	0	1
3UTR	S31	80	1.03	2	0.5674
Intron	S31	80	10.3	20	0.004
Exon	S31	80	2.32	0	0.2118
Promoter (2000)	S73	80	4.11	1	0.0899
5UTR	S73	80	0.04	0	1
3UTR	S73	80	0.31	0	1
Intron	S73	80	3.17	6	0.1598
Exon	S73	80	0.77	0	0.9051
Promoter (2000)	S10	20-40	10.8	5	0.0539
5UTR	S10	20-40	0.1	0	1
3UTR	S10	20-40	0.85	1	1
Intron	S10	20-40	8.52	10	0.6753
Exon	S10	20-40	1.92	0	0.3017
Promoter (2000)	S100	20-40	21.71	13	0.03
5UTR	S100	20-40	0.18	0	1
3UTR	S100	20-40	1.71	4	0.1838
Intron	S100	20-40	17.01	24	0.0759
Exon	S100	20-40	3.8	0	0.046
Promoter (2000)	S31	20-40	16.41	14	0.5355
5UTR	S31	20-40	0.14	0	1
3UTR	S31	20-40	1.31	2	0.7832
Intron	S31	20-40	12.96	12	0.8831
Exon	S31	20-40	2.87	0	0.1099
Promoter (2000)	S73	20-40	16.41	8	0.014
5UTR	S73	20-40	0.14	0	1
3UTR	S73	20-40	1.31	0	0.5654
Intron	S73	20-40	12.96	19	0.0919
Exon	S73	20-40	2.87	0	0.1099
Promoter (2000)	S10	40-60	7.81	4	0.1518
5UTR	S10	40-60	0.07	0	1
3UTR	S10	40-60	0.62	1	0.9271
Intron	S10	40-60	6.13	11	0.038
Exon	S10	40-60	1.44	0	0.4755
Promoter (2000)	S100	40-60	4.84	4	0.8691
5UTR	S100	40-60	0.04	0	1
3UTR	S100	40-60	0.38	0	1
Intron	S100	40-60	3.79	3	0.9371
Exon	S100	40-60	0.91	0	0.8052
Promoter (2000)	S31	40-60	3.75	4	1
5UTR	S31	40-60	0.04	0	1

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

3UTR	S31	40-60	0.29	0	1
Intron	S31	40-60	2.86	3	1
Exon	S31	40-60	0.69	0	0.993
Promoter (2000)	S73	40-60	6.29	2	0.048
5UTR	S73	40-60	0.05	0	1
3UTR	S73	40-60	0.5	0	1
Intron	S73	40-60	4.97	9	0.0719
Exon	S73	40-60	1.15	0	0.6394
Promoter (2000)	S10	60-80	3.75	3	0.8911
5UTR	S10	60-80	0.04	0	1
3UTR	S10	60-80	0.29	0	1
Intron	S10	60-80	2.86	3	1
Exon	S10	60-80	0.69	0	0.993
Promoter (2000)	S100	60-80	2.62	0	0.0819
5UTR	S100	60-80	0.03	0	1
3UTR	S100	60-80	0.2	0	1
Intron	S100	60-80	2	6	0.004
Exon	S100	60-80	0.47	0	1
Promoter (2000)	S31	60-80	2.25	2	1
5UTR	S31	60-80	0.02	0	1
3UTR	S31	60-80	0.17	0	1
Intron	S31	60-80	1.72	3	0.4456
Exon	S31	60-80	0.41	0	1
Promoter (2000)	S73	60-80	4.11	4	1
5UTR	S73	60-80	0.04	0	1
3UTR	S73	60-80	0.31	0	1
Intron	S73	60-80	3.17	4	0.7672
Exon	S73	60-80	0.77	0	0.9051
Promoter (2000)	S10	all	139.9	77	0.002
5UTR	S10	all	1.07	0	0.6753
3UTR	S10	all	10.84	5	0.0939
Intron	S10	all	109.84	99	0.2298
Exon	S10	all	24.14	3	0.002
Promoter (2000)	S100	all	194.34	120	0.002
5UTR	S100	all	1.46	0	0.4535
3UTR	S100	all	14.95	9	0.1578
Intron	S100	all	152.77	150	0.8571
Exon	S100	all	33.43	4	0.002
Promoter (2000)	S31	all	155.61	94	0.002
5UTR	S31	all	1.21	1	1
3UTR	S31	all	12.01	7	0.1938

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Intron	S31	all	121.67	107	0.1199
Exon	**S31**	**all**	**26.71**	**2**	**0.002**
Promoter (2000)	**S73**	**all**	**186.19**	**125**	**0.002**
5UTR	**S73**	**all**	**1.38**	**0**	**0.4995**
3UTR	**S73**	**all**	**14.31**	**6**	**0.02**
Intron	**S73**	**all**	**146.26**	**141**	**0.6134**
Exon	**S73**	**all**	**32.03**	**4**	**0.002**

Table S2. 5: Distribution of the insertion across genomic regions.
The insertion frequencies were divided into five categories, in addition to considering the complete set of insertions (see Material and Methods for details). Expected counts were calculated by simulating a random genome-wide insertion distribution, while observed counts represent the actual number of detected insertions. Statistical significance was assessed using a bidirectional test, with p-values indicating whether the observed number of insertions deviates significantly from expectation (p-values < 0.05 are highlighted in red).

Trait	G0	S10	S31	S73	S100
LT80_18	4	5	1	3	2
LT80_22	3	1	4	2	5
VCO2_18	5	4	3	1	2
VCO2_25	1	4	5	3	2
VCO2_30	3	5	1	4	2
Paraquat_15mM	2	3	1	5	4
Paraquat_25mM	4	3	5	1	2
Climbing	4	5	3	2	1
Viability	3	4	5	1	2
Development time	4	5	3	1	2
Hatchability	3	5	1	2	4
Dry weight	4	3	2	1	5
CTmax	5	4	1	2	3
Diet_Complete	5	3	2	1	4
Diet_Starvation	4	1	5	2	3
Diet_-Sugar	5	4	2	3	1
Diet_-Yeast	3	4	5	1	2

Table S2. 6: Ranking of the standard deviation values (SD) of female traits across populations.
The ranks for the 17 measured traits are reported for females in five populations. A rank of 1 corresponds to the highest SD for a given trait, indicating the greatest variability among individuals.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

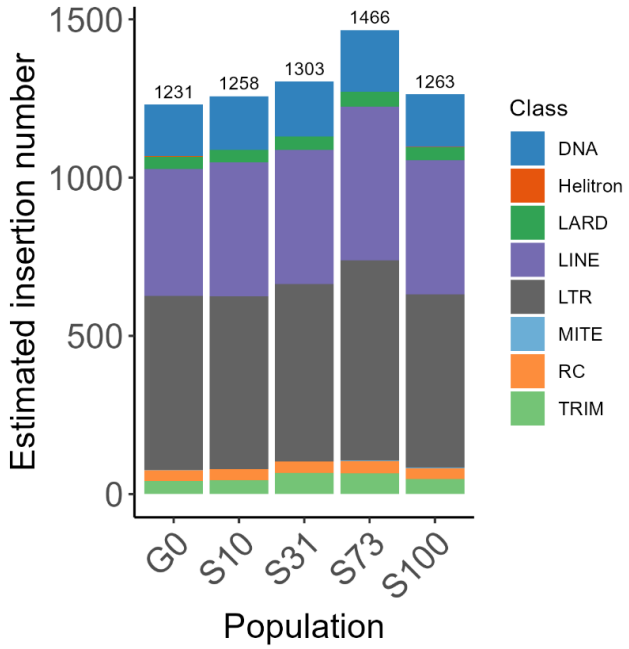


Figure S2. 7: Transposable elements (TE) abundance as estimated by DeviaTE in the Wolbachia-infected populations.

DeviaTE enables to estimate the overall TE quantification independent of a reference genome. In the bar plot, each colour represents a different TE subclass, and the total TE abundance estimate for each population is shown at the top of the corresponding bar.

Population	Sex	Slope	Average
G0	F	-9.325964	-7.8085056
S10	F	-1.526493	
S31	F	-12.896336	
S73	F	-5.52649	
S100	F	-9.767245	
G0	M	-4.154817	-3.4593196
S10	M	-4.234348	
S31	M	-2.992948	
S73	M	-1.494491	
S100	M	-4.419994	

Table S2. 7: Slope of the reaction norm of LT80 under temperature fluctuation.

The slopes were calculated with a linear model of LT80 versus the temperature. Separate slopes are reported for each sex (F: females, M: males) for each population, along with the average slope per sex across the populations.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Term	Sum Sq	Mean Sq	F value	Pr(>F)
Temp	1.50672	0.75336	137.99419	0.00000
Population	0.12215	0.03054	5.59580	0.00123
log(DW_Fly)	0.09698	0.09698	17.77512	0.00016
log(Act_Fly + 1)	0.00826	0.00826	1.51433	0.22450
Temp:Population	0.04442	0.00555	1.01753	0.43963

Table S2. 8: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, as well as their interactions, on females' metabolic rate.
Term indicates the main effects or interaction terms included in the model. The table reports the sum of squares (Sum Sq), mean squares (Mean Sq), F-statistic values (F-value), and associated probabilities (Pr(>F)). Dry weight (DW) and activity (Act) were included as covariates in the model. All data were log-transformed prior to analysis. Temp: temperature, DW: dry weight, Act: activity.

Term	Sum Sq	Mean Sq	F value	Pr(>F)
Temp	1.75513	0.87756	95.63844	0.00000
Population	0.16860	0.04215	4.60432	0.00368
log(DW_Fly)	0.01375	0.01375	1.50250	0.22630
log(Act_Fly + 1)	0.07339	0.07339	8.01722	0.00704
Temp:Population	0.12220	0.01528	1.66850	0.13581

Table S2. 9: Results of a three-way analysis of variance (ANOVA) from the LMER model testing the effects of the population, temperature, as well as their interactions, on males' metabolic rate.
Term indicates the main effects or interaction terms included in the model. The table reports the sum of squares (Sum Sq), mean squares (Mean Sq), F-statistic values (F-value), and associated probabilities (Pr(>F)). Dry weight (DW) and activity (Act) were included as covariates in the model. All data were log-transformed prior to analysis. Temp: temperature, DW: dry weight, Act: activity.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Term	Sum Sq	Mean Sq	F value	Pr(>F)
Population	134990.07997	33747.51999	4.69642	0.00093
Concentration	955136.40284	955136.40284	132.92009	0.00000
Population:Concentration	141569.48155	35392.37039	4.92532	0.00062

Table S2. 10: Results of a three-way analysis of variance (ANOVA) from the LMER model of the effect of population, paraquat concentration, and their interaction on females' longevity.
Term indicates the main effects or interaction terms included in the model. The table reports the sum of squares (Sum Sq), mean squares (Mean Sq), F-statistic values (F-value), and associated probabilities (Pr(>F)). Dry weight (DW) and activity (Act) were included as covariates in the model.

Term	Sum Sq	Mean Sq	F value	Pr(>F)
Population	344220.36261	86055.09065	33.86607	0.00000
Concentration	71343.98476	71343.98476	28.07667	0.00000
Population:Concentration	61832.96141	15458.24035	6.08343	0.00008

Table S2. 11: Results of a three-way analysis of variance (ANOVA) from the LMER model of the effect of population, paraquat concentration, and their interaction on males' longevity.
Term indicates the main effects or interaction terms included in the model. The table reports the sum of squares (Sum Sq), mean squares (Mean Sq), F-statistic values (F-value), and associated probabilities (Pr(>F)). Dry weight (DW) and activity (Act) were included as covariates in the model.

Population	Sex	Slope	Average
G0	F	-8.8393279	-12.1881913
S10	F	-16.8011068	
S31	F	-17.599134	
S73	F	-5.6830703	
S100	F	-12.0183175	
G0	M	-4.384834	-2.92268918
S10	M	-1.9949077	
S31	M	0.5274522	

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

S73	M	-5.9973389	
S100	M	-2.7638175	

Table S2. 12: Slope of the reaction norm of LT80 in response to changes in paraquat concentration.
The slopes were calculated with a linear model of LT80 versus the concentration. Separate slopes are reported for each sex (F: females, M: males) for each population, along with the average slope per sex across the populations.

Population	Sex	Concentration	LT80	LCL	UCL
G0	F	25mM	86.0104075	NA	NA
G0	F	15mM	174.403687	163.660409	186.569396
G0	M	25mM	100.789266	94.1138279	108.906886
G0	M	15mM	144.637605	138.920145	150.928207
S10	F	25mM	69.6729217	66.9074091	72.7663758
S10	F	15mM	237.683989	227.490759	248.905344
S10	M	25mM	59.9235176	NA	NA
S10	M	15mM	79.8725942	70.5686047	92.8409987
S31	F	25mM	93.2764749	NA	NA
S31	F	15mM	269.267815	256.88378	282.973953
S31	M	25mM	103.326411	98.1128562	109.347036
S31	M	15mM	98.0518891	80.0546353	129.143436
S73	F	25mM	109.201941	99.4769627	121.123447
S73	F	15mM	166.032644	157.769222	175.161052
S73	M	25mM	98.9805433	95.1878346	103.211389
S73	M	15mM	158.953932	153.632469	164.744728
S100	F	25mM	103.306844	95.9727987	112.053601
S100	F	15mM	223.490019	213.923413	233.908189
S100	M	25mM	126.63997	121.390001	132.504481
S100	M	15mM	154.278145	148.570153	160.553743

Table S2. 13: Estimates of LT80 under variable paraquat concentrations in* Wolbachia*-infected populations.
The table reports the estimated time until 80% mortality is reached (LT80) along with upper and lower confidence intervals, calculated using the LT_probit function (see Material and Methods). Estimates are shown separately for females (F) and males (M) in each population. NA: not applicable.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

TE_ID	Strand	TSD	pident	psize_TE	SIZE_TE	NEW_POS	FREQ	FREQ_WITH_CLIPPED	SV_SIZE	G0 annotation
con19_Quasimodo#LTR/Gypsy	-	TATATATAT	97.346	98.02	7379	18047875	60	90.9091	7443	5-HT1A
con48_roo#LTR/Pao	+	CTCACT	98.527	95.94	9088	11594593	80	97.8723	9161	Ptp10D
con2_NOF#DNA/MULE-NOF	-	ATTC	97.723	139.86	7385	14182869	75	97.7273	9392	CG32599
con12_Doc#LINE/Jockey	+	AGAA	98.614	98.68	4737	875331	100	100	4788	Myo81F
con12_Doc#LINE/Jockey	-	ATCAATAAA	98.678	98.77	4725	939437	90.9091	97.2973	4795	Myo81F
con12_Doc#LINE/Jockey	+	GAACTGG	98.844	98.6	4717	219916	80	95.7447	4795	CG3777
con19_Quasimodo#LTR/Gypsy	-	ATTTTTTTT	97.558	98.06	7395	12750052	62.5	92.5	7449	Pde9
con48_roo#LTR/Pao	+	CTAAG	98.799	95.79	9070	4169725	80	97.4359	9136	CG15570
con14_F-element#LINE/Jockey	+	TGGCC	98.418	99.28	4706	17674475	71.4286	90.6977	4773	lncRNA:CR43802
BDGP_ZAM#LTR/Gypsy	+	CGCG	98.76	100.08	8442	21186350	66.6667	92.1569	8540	CG34377
con2_gtwin#LTR/Gypsy	+	AGAA	99.057	99.8	7424	20902907	69.2308	93.75	7487	pip
con48_roo#LTR/Pao|sniffles.INS.22216	-	GGCATTT	98.514	96.09	9122	27577710	64.2857	90.1961	9164	sima

Table S2. 14: Manually curated candidate transposable element (TE) insertions.

These candidate TE insertions were selected based on their frequency (≤ 60%), their genomic location, and their insertion site according to the G0-Wb+ gene annotation. The table reports the percent identity of each insertion with the TE consensus sequence (pident), the percentage of its size relative to the TE in the database (psize_TE), the estimated TE size (SIZE_TE), and the corrected insertion position accounting for the target site duplication (TSD) (NEW_POS). Additional information includes the insertion frequency supported by clipped reads (FREQ_WITH_CLIPPED), the corresponding gene annotation from the G0 reference, and the size of the structural variant (SV) (SV_SIZE), which can exceed the TE size.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Phenotype	MRK	Median BF	Insertion position	S10-*Wb+*	S31-*Wb+*	S73-*Wb+*	S100-*Wb+*	TE_ID	G0 annotation	Genomic region
LT80_25mM	40	13.1003713	chr2L:19750653-19750654	50	0	0	0	con12_Doc#LINE/Jockey	Lim3	intron
mean_DW	40	15.7822985								
LT80_Starvation	62	7.6696387	chr2L:2009581-2009582	0	30.4348	0	0	UnFUnClUnAlig_RLX-incomp_FIN#LTR/Gypsy	CG42789	promoter
LT80_-Yeast	62	8.41516451								
LT80_Starvation	66	7.4740983	chr2L:5132542-5132543	0	33.3333	0	0	con16_blood#LTR/Gypsy	CG15431	intron
LT80_-Yeast	66	9.50387632							Sfp24F	promoter
mean_VCO2_30	72	15.4999989	chr2L:11572307-11572308	0	30.7692	0	0	con48_roo#LTR/Pao	CG14070	downstream promoter?
LT80_Starvation	72	9.33723813								
LT80_-Sugar	72	17.5406512								
LT80_-Yeast	72	11.5781495								
mean_VCO2_18	210	16.1896231	chr2R:3589872-3589874	31.25	0	12.5	0	con13_HB#DNA/TcMar-Tc1	Gprk1	intron
LT80_25mM	236	9.57176045	chr2R:18047870-18047871	60	0	0	0	con19_Quasimodo#LTR/Gypsy	5-HT1A	intron
mean_DW	236	10.6097564								
LT80_25mM	238	10.444542	chr2R:19478907-19478908	45.4545	0	0	0	BDGP_ZAM#LTR/Gypsy	Obp57c	promoter
mean_DW	238	12.4827633								
mean_VCO2_18	275	8.8649673	chr2R:1557008-1557009	0	80	0	7.14286	con12_Doc#LINE/Jockey	CG17684	intron
mean_climbing	275	11.0989049								
mean_dev	275	9.86725776								
LT80_-Yeast	275	8.69753861								
mean_climbing	289	10.4244136	chr2R:17692679-17692680	0	33.3333	0	0	BDGP_ZAM#LTR/Gypsy	Hs3st-A	intron
LT80_Starvation	289	7.64229466								
LT80_-Yeast	289	12.6366491								
mean_VCO2_30	483	12.2926804	chr3L:12131294-12131295	0	25	0	0	con12_Doc#LINE/Jockey	CG43127	promoter
mean_CTmax	589	9.72709716	chr3L:20902911-20902912	0	0	0	69.2308	con2_gtwin#LTR/Gypsy	pip	promoter

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

mean_CTmax	600	14.1557797	chr3L:27834387-27834388	0	0	0	33.3333	con12_Doc#LINE/Jockey	intergenic	
LT80_25mM	634	15.8461623	chr3R:7665555-7665556	16	0	0	0	con48_roo#LTR/Pao	pyd	intron
mean_VCO2_25	673	10.349935	chr3R:875331-875332	0	100	0	0	con12_Doc#LINE/Jockey	Myo81F	intron
mean_climbing	673	9.94998743								
mean_dev	673	11.1856632								
LT80_-Yeast	673	8.7954663								
mean_VCO2_25	674	9.32762686	chr3R:939426-939427	0	90.9091	0	0	con12_Doc#LINE/Jockey	Myo81F	intron
mean_climbing	674	8.97597385								
mean_dev	674	10.0673228								
LT80_-Yeast	674	8.53014668								
mean_VCO2_25	691	10.1570024	chr3R:12688519-12688520	0	45.4545	0	0	BDGP_ZAM#LTR/Gypsy	E5	intron
mean_climbing	691	10.3280374								
mean_dev	691	11.609184								
LT80_-Yeast	691	9.57558657								
mean_CTmax	768	14.3058921	chr3R:2826547-2826553	0	0	0	38.4615	BDGP_ZAM#LTR/Gypsy	Pzl	intron
mean_VCO2_25	800	11.1229102	chr4:886098-886099	0	100	6.89655	0	con48_roo#LTR/Pao	unc-13	exon/5'UTR
mean_dev	800	11.2087868							CR43361	promoter
mean_VCO2_18	801	7.96473777	chr4:1014820-1014826	0	90	0	2.5641	con16_blood#LTR/Gypsy	PlexA	intron
mean_dev	801	10.1412845								
LT80_-Yeast	801	9.25659856								

Table S2. 15: Genome-scan candidate transposable element (TE) insertions with manual curation and integration of female phenotypic data. *The candidate TE insertions were initially selected based on the median of five iterations of Baypass, retaining those with a Bayes Factor (BF) above 14, indicating strong association. Further manual curation was performed to select insertions associated with multiple traits (see phenotypes in the first column), occurring at higher frequencies (as reported for each populations), and located in genomic regions other than intergenic. The table reports the genomic position of each insertion, the TE identity, the nearest or overlapping gene based on the G0-Wb+ annotation, and the genomic region of insertion.*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Phenotype	MRK	Median BF	Insertion position	S10-*Wb+*	S31-*Wb+*	S73-*Wb+*	S100-*Wb+*	TE_ID	G0 annotation	Genomic region
LT80_25mM	40	13.1003713	chr2L:19750653-19750654	50	0	0	0	con12_Doc#LINE/Jockey	Lim3	intron
mean_DW	40	15.7822985								
mean_VCO2_30	72	15.4999989	chr2L:11572307-11572308	0	30.7692	0	0	con48_roo#LTR/Pao	CG14070	downstream promoter?
LT80_Starvation	72	9.33723813								
LT80_-Sugar	72	17.5406512								
LT80_-Yeast	72	11.5781495								
mean_VCO2_18	210	16.1896231	chr2R:3589872-3589874	31.25	0	12.5	0	con13_HB#DNA/TcMar-Tc1	Gprk1	intron
LT80_25mM	238	10.444542	chr2R:19478907-19478908	45.4545	0	0	0	BDGP_ZAM#LTR/Gypsy	Obp57c	promoter
mean_DW	238	12.4827633								
mean_VCO2_18	275	8.8649673	chr2R:1557008-1557009	0	80	0	7.14286	con12_Doc#LINE/Jockey	CG17684	intron
mean_climbing	275	11.0989049								
mean_dev	275	9.86725776								
LT80_-Yeast	275	8.69753861								
mean_climbing	289	10.4244136	chr2R:17692679-17692680	0	33.3333	0	0	BDGP_ZAM#LTR/Gypsy	Hs3st-A	intron
LT80_Starvation	289	7.64229466								
LT80_-Yeast	289	12.6366491								
mean_CTmax	600	14.1557797	chr3L:27834387-27834388	0	0	0	33.3333	con12_Doc#LINE/Jockey	intergenic	
LT80_25mM	634	15.8461623	chr3R:7665555-7665556	16	0	0	0	con48_roo#LTR/Pao	pyd	intron
mean_VCO2_25	673	10.349935	chr3R:875331-875332	0	100	0	0	con12_Doc#LINE/Jockey	Myo81F	intron
mean_climbing	673	9.94998743								
mean_dev	673	11.1856632								
LT80_-Yeast	673	8.7954663								
mean_VCO2_25	674	9.32762686	chr3R:939426-939427	0	90.9091	0	0	con12_Doc#LINE/Jockey	Myo81F	intron
mean_climbing	674	8.97597385								
mean_dev	674	10.0673228								
LT80_-Yeast	674	8.53014668								
mean_VCO2_25	691	10.1570024		0	45.4545	0	0	BDGP_ZAM#LTR/Gypsy	E5	intron

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

mean_climbing	691	10.3280374	chr3R:12688519-12688520							
mean_dev	691	11.609184								
LT80_-Yeast	691	9.57558657								
mean_CTmax	768	14.3058921	chr3R:2826547-2826553	0	0	0	38.4615	BDGP_ZAM#LTR/Gypsy	Pzl	intron
mean_CTmax	786	9.70566132	chr3R:27577719-27577720	0	0	0	64.2857	con48_roo#LTR/Pao	sima	intron
mean_VCO2_25	800	11.1229102	chr4:886098-886099	0	100	6.89655	0	con48_roo#LTR/Pao	unc-13	exon/5'UTR
mean_dev	800	11.2087868							CR43361	promoter
mean_VCO2_18	801	7.96473777	chr4:1014820-1014826	0	90	0	2.5641	con16_blood#LTR/Gypsy	PlexA	intron
mean_dev	801	10.1412845								
LT80_-Yeast	801	9.25659856								

Table S2. 16: Genome-scan candidate transposable element (TE) insertions with manual curation and integration of male phenotypic data.
The candidate TE insertions were initially selected based on the median of five iterations of Baypass, retaining those with a Bayes Factor (BF) above 14, indicating strong association. Further manual curation was performed to select insertions associated with multiple traits (see phenotypes in the first column), occurring at higher frequencies, and located in genomic regions other than intergenic. The table reports the genomic position of each insertion, its frequency across Wolbachia-infected populations, the TE identity, the nearest or overlapping gene based on the G0-Wb+ annotation, and the genomic region of insertion.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

S3. Supplementary material for Chapter 2

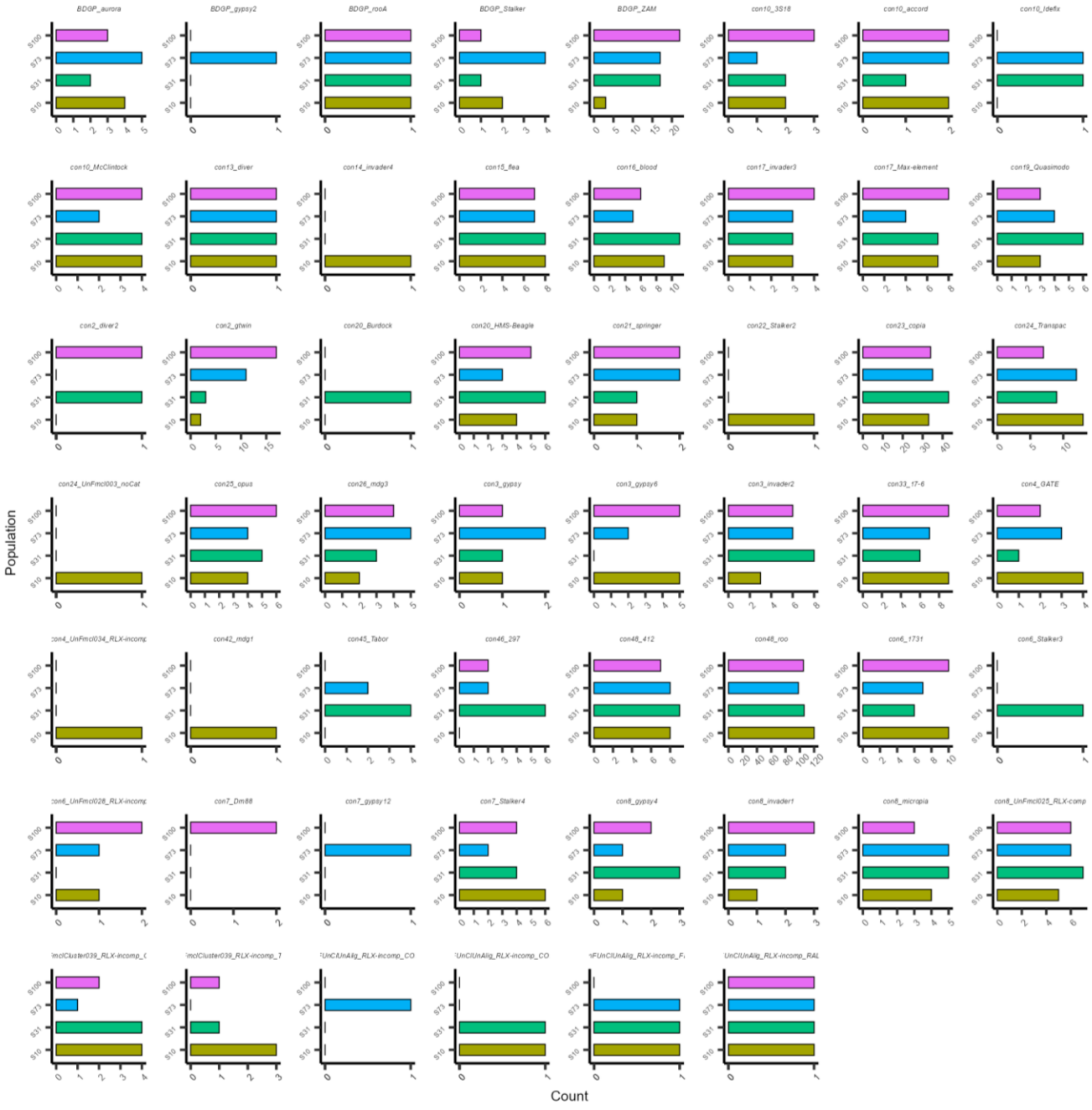


Figure S3. 1: Newly detected LTR retrotransposon insertions relative to G0-Wb+ across consensus library families for the* Wolbachia*-treated populations.

Each bar represents the count of insertions identified in the sample that are absent in the G0-Wb+ reference, as detected by TrEMOLO. Insertions are categorized by family according to the consensus library.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

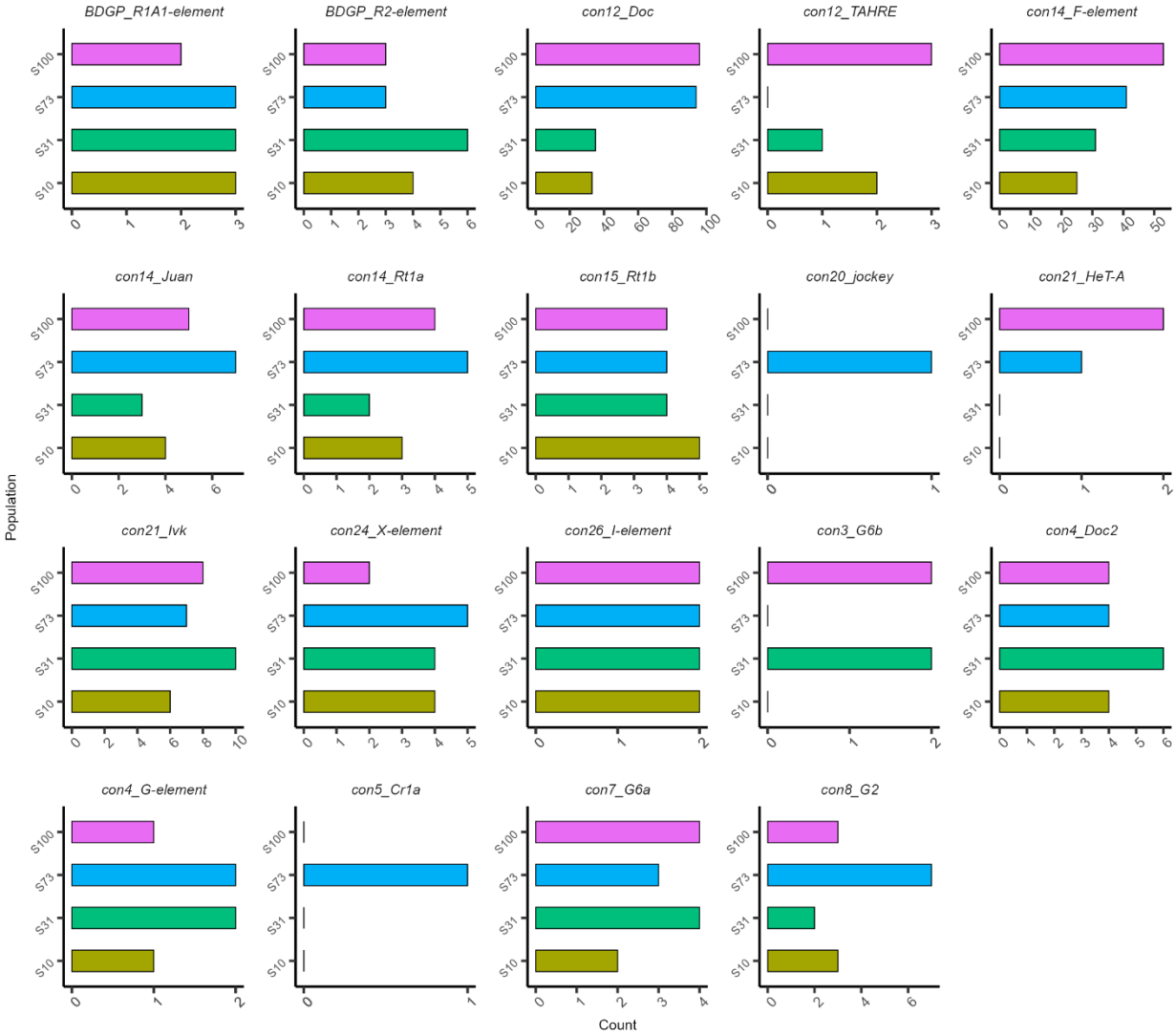


Figure S3. 2: Newly detected LINE transposable element insertions relative to G0-Wb+ across consensus library families for the* Wolbachia*-treated populations.

Each bar represents the count of insertions identified in the sample that are absent in the G0-Wb+ reference, as detected by TrEMOLO. Insertions are categorized by family according to the consensus library.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

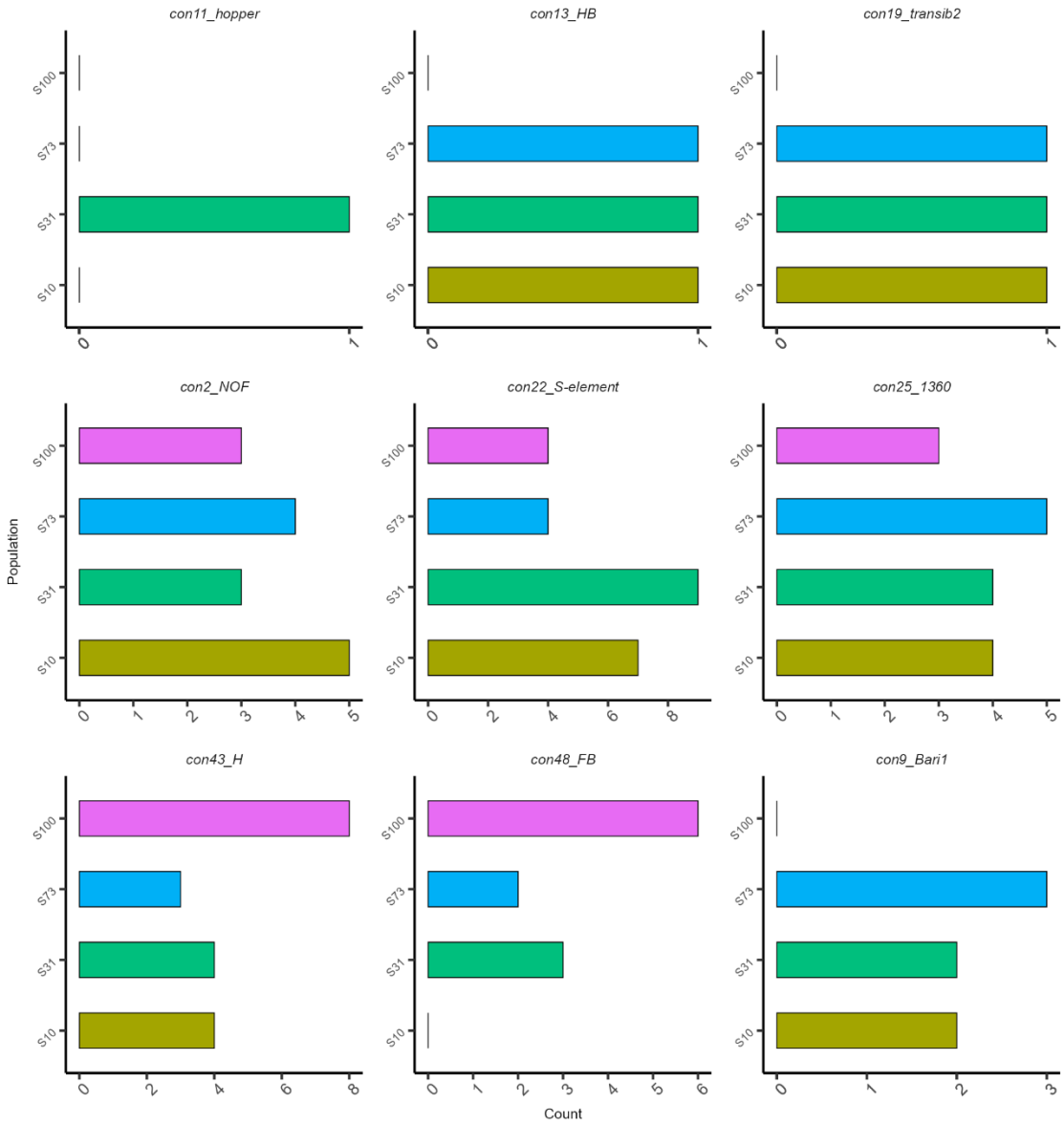


Figure S3. 3: Newly detected DNA transposon insertions relative to G0-Wb+ across consensus library families for the* Wolbachia*-treated populations.

Each bar represents the count of insertions identified in the sample that are absent in the G0-Wb+ reference, as detected by TrEMOLO. Insertions are categorized by family according to the consensus library.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés



Figure S3. 4: Shared transposable element (TE) insertions between* Wolbachia *infection status populations.

The Venn diagrams show the number of TE insertions shared within an extended window of 100 bp between a pair of populations (see Material and Methods). Approximately 60% of TE insertions are shared between the Wolbachia*-treated populations (Wb-, yellow) and their* Wolbachia*-infected counterparts (Wb+, blue).*

	TE abundance (HQ reads only)	
	Wb+	**Wb-**
G0_35x	1230.690813	1273.785882
S10_35x	1257.53623	1326.18502
S31_35x	1303.319749	1214.098021
S73_35x	1466.012079	1389.40346
S100_35x	1263.210063	1183.176642

Table S3. 1: Total transposable element (TE) abundance across five populations under different* Wolbachia *infection status.

The TE abundance was estimated using DeviaTE, based on the 35 X subsampled data. The total abundance corresponds to the sum of estimated abundances across all TE families, calculated using only high-quality reads (HQ, >Q15).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Population	Temperature	contrast	emmean	*SE*	*df*	lower.CL	upper.CL
G0	18	(Wb-) - (Wb+)	11.74	2.97	1,976.51	2.40	21.07
S10	18	(Wb-) - (Wb+)	12.92	2.65	1,964.39	4.58	21.25
S31	18	(Wb-) - (Wb+)	0.06	2.71	1,972.60	-8.46	8.58
S73	18	(Wb-) - (Wb+)	4.57	2.76	1,979.93	-4.12	13.25
S100	18	(Wb-) - (Wb+)	4.99	2.71	1,982.00	-3.55	13.53
G0	22	(Wb-) - (Wb+)	12.76	2.99	1,979.51	3.35	22.17
S10	22	(Wb-) - (Wb+)	14.59	2.95	1,979.39	5.32	23.87
S31	22	(Wb-) - (Wb+)	2.92	2.95	1,972.39	-6.38	12.21
S73	22	(Wb-) - (Wb+)	-3.88	2.93	1,970.25	-13.10	5.34
S100	22	(Wb-) - (Wb+)	6.36	2.88	1,980.12	-2.71	15.44

Table S3. 2: Estimated marginal means (Emmeans) contrasts from the LMER model for males' longevity. *The table reports contrasts of* Wolbachia *infection status (Wb- vs. Wb+) across populations and temperatures, obtained from the LMER model. Values include the estimated marginal mean difference (Emmean), standard error (SE), degrees of freedom (df), and the 95% confidence interval (lower.CL, upper.CL).*

Population	Temperature	Contrast	Emmean	*SE*	*df*	lower.CL	upper.CL
G0	18	(Wb-) - (Wb+)	5.14	4.13	2,162.73	-7.85	18.13
S10	18	(Wb-) - (Wb+)	13.53	4.49	2,159.11	-0.59	27.66
S31	18	(Wb-) - (Wb+)	-7.75	4.37	2,152.81	-21.51	6.02
S73	18	(Wb-) - (Wb+)	8.17	4.41	2,161.25	-5.72	22.07

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Population	Temperature	Contrast	Emmean	*SE*	*df*	lower.CL	upper.CL
S100	18	(Wb-) - (Wb+)	16.56	4.64	2,163.85	1.94	31.18
G0	22	(Wb-) - (Wb+)	1.93	4.44	2,161.00	-12.05	15.91
S10	22	(Wb-) - (Wb+)	1.32	4.83	2,161.24	-13.88	16.53
S31	22	(Wb-) - (Wb+)	2.99	4.15	2,152.17	-10.06	16.04
S73	22	(Wb-) - (Wb+)	0.81	4.59	2,152.59	-13.65	15.27
S100	22	(Wb-) - (Wb+)	6.55	4.43	2,162.74	-7.38	20.49

Table S3. 3: Estimated marginal means (Emmeans) contrasts from the LMER model for females' longevity.

The table reports contrasts of Wolbachia *infection status (Wb- vs. Wb+) across populations and temperatures, obtained from the LMER model. Values include the estimated marginal mean difference (Emmean), standard error (SE), degrees of freedom (df), and the 95% confidence interval (lower.CL, upper.CL).*

Population	Temp	contrast	emmean	*SE*	*df*	lower.CL	upper.CL
G0	18	(Wb+) - (Wb-)	0.14	0.06	100.04	-0.06	0.35
S10	18	(Wb+) - (Wb-)	0.09	0.06	100.70	-0.12	0.29
S31	18	(Wb+) - (Wb-)	-0.02	0.06	103.08	-0.22	0.18
S73	18	(Wb+) - (Wb-)	0.21	0.06	97.64	0.01	0.42
S100	18	(Wb+) - (Wb-)	0.05	0.06	101.33	-0.16	0.25
G0	25	(Wb+) - (Wb-)	0.12	0.05	93.65	-0.07	0.30
S10	25	(Wb+) - (Wb-)	0.06	0.06	99.39	-0.15	0.26

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Population	Temp	contrast	emmean	*SE*	*df*	lower.CL	upper.CL
S31	25	(Wb+) - (Wb-)	-0.07	0.06	96.40	-0.26	0.12
S73	25	(Wb+) - (Wb-)	-0.03	0.06	96.53	-0.22	0.16
S100	25	(Wb+) - (Wb-)	0.08	0.05	93.58	-0.11	0.26
G0	30	(Wb+) - (Wb-)	0.11	0.06	98.83	-0.11	0.33
S10	30	(Wb+) - (Wb-)	-0.07	0.07	95.61	-0.30	0.15
S31	30	(Wb+) - (Wb-)	-0.07	0.06	96.36	-0.28	0.15
S73	30	(Wb+) - (Wb-)	-0.00	0.07	97.24	-0.24	0.24
S100	30	(Wb+) - (Wb-)	0.03	0.06	86.55	-0.19	0.24

Table S3. 4: Estimated marginal means (Emmeans) contrasts from the LMER model for females' metabolic rate.

The table reports contrasts of Wolbachia *infection status (Wb- vs. Wb+) across populations and temperatures, obtained from the LMER model. Values include the estimated marginal mean difference (Emmean), standard error (SE), degrees of freedom (df), and the 95% confidence interval (lower.CL, upper.CL).*

Population	Temp	contrast	emmean	*SE*	*df*	lower.CL	upper.CL
G0	18	(Wb+) - (Wb-)	0.08	0.06	93.53	-0.12	0.28
S10	18	(Wb+) - (Wb-)	-0.01	0.05	95.00	-0.19	0.17
S31	18	(Wb+) - (Wb-)	0.02	0.05	92.94	-0.15	0.19
S73	18	(Wb+) - (Wb-)	-0.03	0.05	97.12	-0.20	0.14
S100	18	(Wb+) - (Wb-)	0.04	0.05	100.38	-0.14	0.22

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Population	Temp	contrast	emmean	*SE*	*df*	lower.CL	upper.CL
G0	25	(Wb+) - (Wb-)	0.00	0.05	93.63	-0.18	0.18
S10	25	(Wb+) - (Wb-)	0.02	0.06	97.75	-0.16	0.21
S31	25	(Wb+) - (Wb-)	0.05	0.06	93.67	-0.14	0.24
S73	25	(Wb+) - (Wb-)	-0.12	0.05	92.30	-0.28	0.04
S100	25	(Wb+) - (Wb-)	0.11	0.05	94.97	-0.07	0.29
G0	30	(Wb+) - (Wb-)	0.06	0.06	94.17	-0.14	0.26
S10	30	(Wb+) - (Wb-)	-0.05	0.06	90.67	-0.24	0.14
S31	30	(Wb+) - (Wb-)	0.02	0.05	88.94	-0.16	0.20
S73	30	(Wb+) - (Wb-)	-0.21	0.06	92.15	-0.40	-0.01
Table S2. 17 S100	30	(Wb+) - (Wb-)	0.08	0.05	89.42	-0.10	0.25

Table S3. 5: Estimated marginal means (Emmeans) contrasts from the LMER model for females' metabolic rate.

The table reports contrasts of Wolbachia *infection status (Wb- vs. Wb+) across populations and temperatures, obtained from the LMER model. Values include the estimated marginal mean difference (Emmean), standard error (SE), degrees of freedom (df), and the 95% confidence interval (lower.CL, upper.CL).*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Gene	Organism	Sense	Sequence (3'-5)	Annealing temperature	Amplicon size	Probe (3'-5')
RP49	*D. melanogaster*	Forward	CTGCCCACCGGATTCAAG	60	91	5'(HEX)-CCTCCAGCTCGCCGACGTTG-(BHQ-1)3'
	D. melanogaster	Reverse	CGATCTCGCCGCAGTAAA			
FtsZ	*Wolbachia*	Forward	CGAGATGGGCAAAGCGATGA	60	126	5'(FAM)-ACCGGAGAGGCAGAAGGAGA-(BHQ-1)3'
	Wolbachia	Reverse	ATTCCTTGCGCACCTTTCAT			

Table S3. 6: Primer sequences used for the qPCR amplification.
Primer pairs are listed for D. melanogaster *reference gene (Rp49) and the* Wolbachia*-specific gene (FtsZ). For each gene, the table provides the forward and reverse primer sequences, annealing temperature (°C), the amplicon size and the associated fluorescent probe sequence.*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Population	Sex	Temperature	Effectif number
G0	F	18	130
G0	F	22	204
G0	M	18	109
G0	M	22	162
S10	F	18	90
S10	F	22	163
S10	M	18	99
S10	M	22	166
S31	F	18	109
S31	F	22	224
S31	M	18	107
S31	M	22	179
S73	F	18	116
S73	F	22	201
S73	M	18	109
S73	M	22	220
S100	F	18	87
S100	F	22	203
S100	M	18	99
S100	M	22	203

Table S3. 7*: *Number of flies assayed for longevity in the* Wolbachia*-treated populations under different temperature.

For each population, the table reports the sex of the flies (F: female, M: male), the rearing temperature (°C), and the effective number of individuals for which longevity data were collected.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Annexe 1

CHAPTER

Epigenetics and genotypic variation: A transposable elements' perspective

6

Miriam Merenciano[1], Anaïs Larue[1,2], Cristian Groza[3], Cristina Vieira[1], Rita Rebollo[2] and Clément Goubert[3,4,5]

[1]*Universite Claude Bernard Lyon 1, CNRS, Laboratoire de Biometrie et Biologie Evolutive, UMR 5558, Villeurbanne, France;* [2]*INRAE, INSA Lyon, BF2I, UMR203, Villeurbanne, France;* [3]*Human Genetics, McGill University, Montreal, QC, Canada;* [4]*McGill Genome Centre, Montreal, QC, Canada;* [5]*R. Ken Coit College of Pharmacy, University of Arizona, Tucson, AZ, United States*

Introduction

Transposable elements (TEs) are repeated DNA sequences, many of which are able to move along the genome through a process called transposition. TEs are extremely successful in invading the germline, frequently representing more than 50% of the DNA of eukaryotic genomes (e.g., in humans, arthropods, and crops such as barley and maize) (de Koning et al., 2011; Parisot et al., 2021; Stitzer et al., 2021; Wicker et al., 2007), up to an astonishing 90% in the lungfish (Meyer et al., 2021). TEs are classified into two major classes depending on their mechanisms of transposition (Fig. 6.1). Class I TEs, or retrotransposons, move through an RNA intermediate, consequently increasing their copy number in the genome ("copy and paste"). Retrotransposons can be further classified by the presence of long terminal repeats (LTRs) into LTR/endogenous retrovirus (ERV) elements, highlighting their relationship with exogenous retroviruses. Contrary to LTR/ERV elements that transpose through a typical retroviral mechanism, non-LTR retrotransposons use target-primed reverse transcription to insert new copies into the genome. On the other hand, class II elements or DNA transposons transpose without an RNA intermediate, usually taking advantage of their own DNA sequences. Terminal inverted repeat (TIR) elements encode a transposase capable of cutting and directing the excised DNA copy into another location of the genome ("cut-and-paste"). Other DNA elements include groups transposing via rolling circle (*Helitrons*), those using tyrosine recombinases (*DIRS*, *Cryptons*) and the large (>20 kb) and complex "self-synthetizing" *Mavericks*/*Polintons* (Wells & Feschotte, 2020). Each TE class encompasses superfamilies and families based on their sequence divergence and structural characteristics. At a finer scale, transposition generates a homogeneous group of copies spread throughout the genome and stemming from the same original copy. These copies can be grouped into subfamilies, ideally based on their monophyletic origin, though most often based on an empirical homology rule referred to as the

On Epigenetics and Evolution. https://doi.org/10.1016/B978-0-443-19051-3.00006-1
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

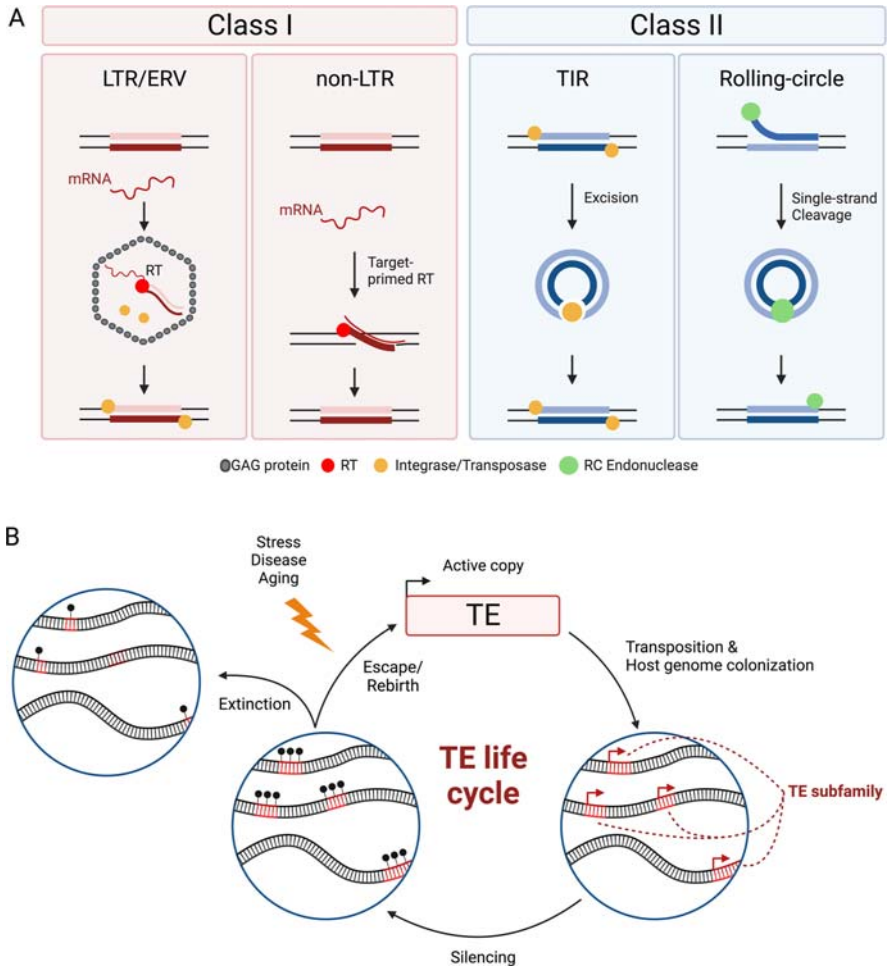


FIGURE 6.1

Schematic representation of the different classes of transposable elements (TEs) and their general life cycle. (A) TE classification. Depending on their transposition mechanism, TEs can be divided into class I (which transpose through an mRNA intermediate) or class II (which use a DNA intermediate). Within each class, TEs are further grouped depending on their structure, open reading frames and transposition mode. Retrotransposons, or class I elements, are composed of long terminal repeat (LTR) or endogenous retrovirus (ERV) elements, along with non-LTR elements. All class I elements mobilize via a "copy-and-paste" mechanism using a reverse transcriptase (RT), however, non-LTR elements follow a target-primed reverse transcription (TPRT) mechanism while LTR/ERVs follow a typical retroviral integration. Within DNA elements (class II), terminal inverted repeat elements (TIR) transpose via a "cut-and-paste" mechanism using a DDE transposase, while *Helitrons* follow a rolling circle, "peel-and-paste" transposition process using a Rep/Helicase protein with endonuclease activity. (B) Schematic representation of the TE life cycle, including transposition of active, functional

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

"80-80-80" rule (copies must align for at least 80 bp, with a minimum of 80% identity along 80% of their sequences (Wicker et al., 2007).

Inherent to their replication mode, TEs constitute an important source of genetic variation, causing large-scale structural variants (SVs), such as insertions, deletions, segmental duplications, inversions, and translocations. Homologous recombination within TE copies or nonallelic (ectopic) recombination between copies, also contributes to generating TE-derived SVs. Active TE families in humans may account for more than ∼2000 SVs observed between a given individual and the reference genome (Gardner et al., 2017; Sudmant et al., 2015). In maize, where up to 70% of the genome can be identified as TEs, ∼30% of these annotations can be classified as polymorphic (differentially present or absent) between the B73 reference and 26 inbred lines (Munasinghe et al., 2023).

TEs are a major evolutionary force, not only for their mutation-causing jumping mechanisms but also thanks to their inherent regulatory sequences and coding potential. Indeed, TEs can contain all the sequences necessary for producing the transpositional machinery, including promoter sequences, transcription factor binding sites, splicing, and poly(A) sites along with functional open reading frames (Bourque et al., 2018; Chuong et al., 2017). The presence of such sequences has been recurrently involved in the modulation of host pathways. For instance, syncytins, proteins promoting cell-to-cell fusion in placental mammals, have been independently co-opted multiple times from ERV envelope genes (*env*) (Lavialle et al., 2013). In *Drosophila*, three TE families have been domesticated for telomere maintenance (McGurk et al., 2021). Finally, the widespread colonization of TEs within a genome has been associated with gene regulation, in some instances effectively rewiring gene regulatory networks through the exaptation of their *cis*-regulatory sequences (Chuong et al., 2017; Giménez-Orenga & Oltra, 2023, pp. 323−355; Klein & Anderson, 2022; Sundaram & Wysocka, 2020).

While it is now undeniable that many TE-derived sequences play a beneficial role in host biology, most new TE insertions are either neutral or deleterious (Arkhipova, 2018). Indeed, TEs are often the target of purifying selection, and this is illustrated by the fact that several insertions in humans are linked to genetic diseases (Payer & Burns, 2019). Hence, regulation of TE sequences is crucial in maintaining genome stability and integrity. The major regulatory mechanisms controlling TE activity are epigenetic, acting at the transcriptional and post-transcriptional levels, and involve small RNAs, chromatin alterations such as post-translational modifications of histone tails, and DNA methylation ((Slotkin & Martienssen, 2007), Box 6.1).

elements, followed by rapid epigenetic host silencing, eventually leading to degeneration and effective inactivation of TE copies. In some instances, reactivation can occur (e.g., through epigenetic derepression during development, stress or aging). The TE life cycle demonstrates that TEs are able to colonize genomes, are the target of epigenetic marks (represented by dark lollipops) and can suffer deletions and mutations, to the point that they are no longer recognized as repeat sequences. TEs able to escape epigenetic silencing can also give birth to novel subfamilies.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Box 6.1 Transposable element regulation through epigenetic pathways

While TE copies are enriched within heterochromatic regions and are therefore embedded in a compact repressed chromatin structure, there are many copies present in euchromatic regions that are the target of repressive epigenetic marks. Repressive histone post-translational modifications as di- or trimethylation of lysine 9 of histone 3 (H3K9me2/3) are targeted to euchromatic TE copies by zinc finger proteins and histone methyltransferase complexes (Choi & Lee, 2020; Lisch, 2009; Slotkin & Martienssen, 2007). In plants and vertebrates, DNA methylation also targets TE copies through DNA methyltransferases, while in arthropods, the presence of DNA methylation is phylogenetically variable and not necessarily present at TE copies (Rollins et al., 2006; Smith et al., 2012; Walsh et al., 1998; Wang & Baulcombe, 2020). Finally, small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs) drive transcriptional silencing of TE sequences by slicing TE mRNAs. Post-transcriptionally silencing can also be mediated through small RNAs. Indeed, in plants, RNA-directed DNA methylation (RdDM) involves silencing of active TE copies via siRNAs (Panda et al., 2016). piRNAs also contribute to TE transcriptional silencing through the mediation of H3K9me3 deposition at TE loci (Le Thomas et al., 2013, 2014). piRNAs stem from piRNA clusters, that are segments of the genome where TE copies have accumulated and are nested within each other. Such clusters are able to produce single or dual-stranded capped polyA mRNAs that are processed by a complex biogenesis pathway involving Argonaute and Piwi proteins in the so-called ping-pong cycle (Czech & Hannon, 2016). Notably, a TE copy needs to produce a mRNA in order to activate the ping-pong cycle and consequently silence the same copy. These piRNA clusters are considered immune primers against TE activity and may differ between populations and closely related species (Duc et al., 2019).

While usually TEs are kept under tight epigenetic control, through development, aging, disease, and stress, TEs are able to escape silencing and in some cases, produce novel insertions. In turn, new TE-derived variants can trigger changes in their neighboring epigenetic landscape while being targeted by epigenetic defense mechanisms, mutating or providing regulatory elements. While the consequence of TE-derived epigenetic effects on gene regulation remains elusive, the layering of genomic, transcriptomic, and epigenetic data at the scale of populations and species provides key insights into the relationship between TEs and epigenetic variation.

In this chapter, we discuss the intricate interactions between the epigenome and genetic variation through the lens of TEs. We first describe exciting new methods—including pan-genomic approaches—aimed to alleviate the bioinformatic challenges faced by epigenetic studies of TEs. Next, we review how epigenetic changes can affect genomic variation by unleashing TEs and the consequences of these new variants for the epigenetic landscape and gene regulatory networks. Finally, we discuss how evolution shapes the interplay between TEs and epigenomes.

Studying TE epigenetics: Between challenges and breakthroughs

The bioinformatic analysis of TEs, and in particular, the epigenetic changes they induce, is utterly challenging. Indeed, their repetitive nature has been a long-time puzzle: when TE copies within different genomic locations are identical, how can

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

one assert with confidence the origin of short genomic, transcriptomic or epigenomic readouts—the "short-reads"—that have dominated the next-generation sequencing era? To make matters worse, TEs and other repeated DNA sequences are to blame for the lack of completion in genomic assemblies. Furthermore, structural variation (including those caused by TEs) between individual samples is another source of uncertainty, which may confound the underlying biological processes if relevant sequences are differentially present between a sample and a given reference genome (insertion polymorphism).

The first strategy to mitigate such problems is to study epigenetic and transcriptional activation of TEs at the subfamily level. Indeed, there are many tools that have been developed to study TE transcription at this scale, including TEtranscripts (Jin et al., 2015) and TEtools (Lerat et al., 2017). These tools aggregate the reads mapping to TE copies belonging to the same subfamily to yield a global readout. It is important to note that mappability issues arise when studying young, potentially active copies. For instance, *D. melanogaster* harbors many young, transpositional competent TE families (Mérel et al., 2020), rendering estimation of copy-specific expression complicated with short reads. Recently, two different methods have emerged in order to reassign multimapping short reads based on uniquely mapping read coverage to specific TE copies (Squire (Yang et al., 2019) and Telescope (Bendall et al., 2019)). It is important to note that family or copy-level analysis of TE expression also suffers from prior biases of the TE location, structure, and integrity. For instance, a TE copy present within an intron might originate from read-through transcription independently of the canonical TE promoter activation. This and other considerations on TE RNA-seq analysis have thoroughly been reviewed by (Lanciano & Cristofari, 2020).

Regarding histone post-translational modifications, chromatin immunoprecipitation sequencing (Kaufmann et al., 2010) or the newest Cut&Tag or Cut&Run methods allows sequencing of DNA fragments associated with the chromatin mark of interest (Kaya-Okur et al., 2019; Skene & Henikoff, 2017). While many authors adopted a subfamily level analysis when dealing with such methods (Almeida da Paz & Taher, 2022; Criscione et al., 2014), it is also possible to take advantage of TE's flanking sequences, which are usually unique, to extrapolate the enrichment of a specific histone mark within a TE copy. DNA methylation analyses are usually performed using bisulfite-treated DNA, which allows the conversion of unmethylated cytosines into uracils and provides cytosine-level information when fragments are amplified and sequenced (Statham et al., 2012). While bisulfite-seq holds the same short-read bias described above, the amplicons are usually longer, around 500 bp, allowing 5′ and 3′ TE regions to be properly anchored within their flanking sequences. Bisulfite can also be followed by PCR and Sanger sequencing, allowing the study of single copies instead of genome-wide readouts. Finally, small RNA sequencing also involves short-reads allowing mostly for TE subfamily analysis.

Advances in long-read sequencing technologies have enabled a new era in bioinformatics, making available a larger number of high-quality genomes for many species. Recently, complete or near-complete genome sequences, spanning complex

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

heterochromatic regions, have been published in humans (Hoyt et al., 2022; Nurk et al., 2022) or *Drosophila* (Bracewell et al., 2019; Chang et al., 2019; Mahajan et al., 2018; Mohamed et al., 2020). Fueled by these technological improvements, the Human Pangenome Reference Consortium recently published high-quality assemblies for 94 human genomes (Liao et al., 2023). Such datasets can now directly access the sequence of structural variation in most repetitive regions, a task that was previously difficult or impossible using short reads. For example, long-read sequencing of 35 human genomes provided 9453 nonreference mobile elements and 107,590 SVs, of which 68% were not known in short-read datasets (Ebert et al., 2021). At the same time, data structures such as genome graphs have been developed to represent large-scale SVs in addition to single nucleotide variants (SNVs) and indels in populations of genomes, *a.k.a.* pangenomes (Garrison et al., 2023; Li et al., 2020). Therefore, it is now possible to sequence and assemble a set of genomes and compile them into a graph genome reference that summarizes many possible haplotypes, improving read alignment (Garrison et al., 2018). Interestingly, these richer references contain novel insertions that are not present in the linear reference genomes of the same species. These novel insertions may include additional copies of genes, variable number of tandem repeats (VNTRs) and TEs (Bayer et al., 2020; Della Coletta et al., 2021; Li et al., 2022; Zhou et al., 2022). Given that these sequences were hidden in single reference genomes, their impact on the overall epigenome and consequent modulation of host gene expression was unknown and can now be explored.

Fortuitously, the same long-read technologies also allow simultaneous measuring of the epigenetic state of the DNA fragment that is being sequenced (Lee et al., 2020; Rand et al., 2017; Simpson et al., 2017), enabling new bioinformatic analyses of TEs and their associated epigenetic states (Daron & Slotkin, 2017; Ewing et al., 2020). Therefore, a possible approach is to sequence multiple genomes to create a genome graph and then superimpose the methylation data to reveal the epigenetic state of structural variation in the graph. This approach conveniently facilitates differential methylation analysis of SVs such as TE insertions. For instance, pangenome assembly and methylation analysis of 12 Medaka genomes revealed 4459 differentially methylated regions in this species, which are enriched near transcription start sites (Leger et al., 2022). In humans, data such as ChIP-seq and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) can be aligned to genome graphs to further characterize histone modifications and chromatin openness overlapping the SVs (Groza et al., 2020, 2023). If gene annotation and RNA-seq data are available, spliced genome graphs can also detect SVs involved in haplotype-specific transcripts (Sibbesen et al., 2023). Therefore, genome graphs are a promising tool to integrate epigenetic and expression data in order to find functional and polymorphic SVs such as TEs that might have been co-opted by the host genome.

Finally, it is important to acknowledge the bioinformatic revolution fostered by machine learning (ML) models. ML resolves complex problems such as the detection, annotation, and classification of genomic and epigenomic features, with

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

minimal assumptions about the underlying causal mechanisms. This is particularly relevant to identify and label TE-derived sequences (Bell et al., 2022; Orozco-Arias et al., 2019), resolve complex SVs (Cai et al., 2019; Chu et al., 2021) and infer epigenetic states (Cook et al., 2020; Giassa & Alexiou, 2021). Nevertheless, the accuracy of ML models relies on large and highly curated datasets, which remain limited in the field of epigenomics. Thus, the empirical acquisition of experimental data based on reliable molecular biology assays remains a central asset in the quest to understand the impacts of TEs on their hosts' evolution mediated by epigenetic mechanisms.

TEs as a source of genotypic variation through modulation of their epigenetic control

Aging promotes TE mobilization through relaxation of their epigenetic control

The TE theory of aging proposes that increased activation of TEs in somatic tissues can lead to a shortening of the lifespan (Driver & McKechnie, 1992) (Fig. 6.2). In fact, aging has been associated with chronic deterioration of chromatin integrity (López-Otín et al., 2013), thus promoting the expression of TEs in different species, from *Drosophila* (Chen et al., 2016; Fabian et al., 2021; Jones et al., 2016; Li et al., 2013; Nirala et al., 2021; Schneider et al., 2023; Wood et al., 2016; Yang et al., 2022), to rodents (Gamdzyk et al., 2020; Ramirez et al., 2022; Simon et al., 2019; Stow et al., 2021), and humans (De Cecco et al., 2013; LaRocca et al., 2020; Sur et al., 2017). Specifically, in *Drosophila*, modifications in components of the small RNA-dependent TE silencing pathways lead to TE expression and to age-related phenotypes (Li et al., 2013; Wood et al., 2016; Yang et al., 2022). For instance, mutations in the *Drosophila Argonaute 2* (*Ago2*) gene, which interacts with siRNAs to form RNA-induced silencing complexes (RISCs), caused elevated TE transcript amounts and de novo insertions in a similar way as aging (Li et al., 2013). *Ago2* mutations also resulted in progressive and age-dependent memory impairment and shortened lifespan, thus suggesting that TE activation may contribute to neuronal decline (Li et al., 2013). Similar results were obtained by (Yang et al., 2022), where many TEs were upregulated and some changed in copy number during aging. When genetically boosting RNAi activity in aged flies via knocking down the conserved transcription elongation factor PAF1 that antagonizes RNAi pathways, TE transcripts no longer accumulate and an extension of lifespan is observed (Yang et al., 2022). Furthermore, overexpression in old flies of the *piwi, aub*, and *Ago2* genes involved in RNAi pathways mitigated the aging-related increase of TE transcripts (Yang et al., 2022). A link between aging and TE activation has also been reported in termites due to the down-regulation of homologous genes of *Drosophila* piRNA-mediated TE silencing mechanisms (Elsner et al., 2018). Indeed, termite queens and kings can live longer without showing a significant increase in TE expression, while

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

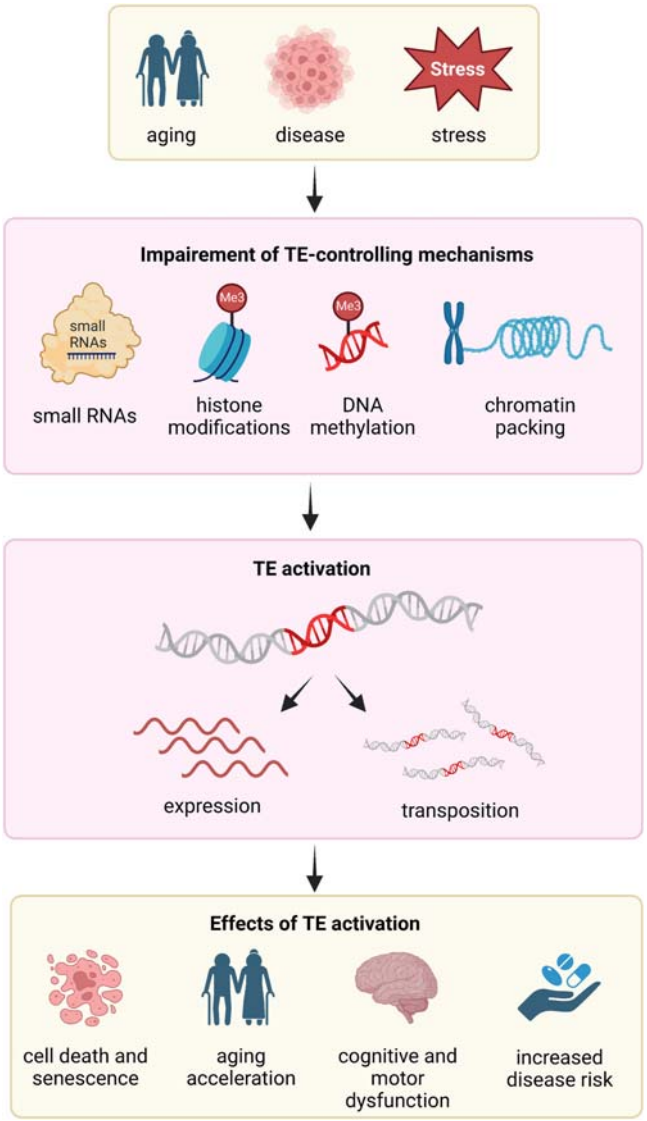


FIGURE 6.2

Potential causes and effects of transposable element (TE) activation through the modulation of their controlling mechanisms. Aging, disease, and stress can impair all types of TE-controlling epigenetic mechanisms leading to an increased TE activation (in terms of expression and/or transposition). The increased TE activation can in turn impact cell death and senescence, aging, cognitive and motor capacities or increase disease risk.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

major workers, which have reduced lifespan, show upregulation of TEs upon aging (Elsner et al., 2018).

Besides small RNA-mediated silencing mechanisms, the alteration of chromatin structure through histone modifications or chromatin packing has also been shown to have an impact on the age-related activation of TEs (Chen et al., 2016; Morgunova et al., 2022; Nirala et al., 2021; Wood et al., 2016). In Wood et al. (2016), authors genetically overexpressed *Sir2*, and *Su(var)3-9* genes to correlate changes in heterochromatin structure with the previously found increased expression and mobilization of TEs with age in both fly heads and fat bodies. These genes play important roles in the repressive heterochromatin establishment and maintenance in somatic cells. They found a reduction in the overexpression of the TEs previously shown to be activated with age, suggesting a common link between heterochromatin factors and proper maintenance of TE silencing during aging (Wood et al., 2016). A similar mechanism has been found in mice, where SIRT6-deficient cells accumulate LINE-1 (L1) transcripts in the same way as in aged mice, resulting in pathological inflammation (Simon et al., 2019). SIRT6 is a mono-ADP-ribosylase/deacetylase protein that silences L1 interacting with HP1 and forming transcriptionally silent heterochromatin (Van Meter et al., 2014). Moreover, in humans, the use of formaldehyde-assisted isolation of regulatory elements (FAIRE) to map genome-wide chromatin conformations revealed that senescence cells have an overall opening of chromatin in heterochromatic gene-poor regions (De Cecco et al., 2013). This loss of heterochromatin caused an increase in TE transcription and ultimately transposition of the major active retrotransposon classes *Alu*, SVA, and L1 (De Cecco et al., 2013). Furthermore, expression of RB1, a factor that promotes heterochromatinization of repetitive elements including L1s, was found to be reduced in the same senescent cells, and this was associated with the loss of H3K9me3 and H3K27me3 at L1 loci resulting in their expression (De Cecco et al., 2019). More recently, the loss of function of the Hinfp transcription factor was found to be associated with the derepression of many TE families in *Drosophila* promoting aging and cancer-related phenotypes (Nirala et al., 2021). During this process, it was shown that Hinfp interacts with Histone 1, needed for the condensation of the nucleosomes and essential for chromatin assembly (Nirala et al., 2021).

Nuclear lamins, major structural components of the nuclear lamina, are thought to contribute to the establishment and maintenance of heterochromatin. In *Drosophila*, aging has been associated with a gradual reduction of lamin-B (LAM) (Chen et al., 2014). Disruption of LAM was correlated with an increased expression of retrotransposons (Chen et al., 2016; Morgunova et al., 2022). Indeed, some of these retrotransposons showed decreased levels of H3K9me3 upon both aging or LAM depletion, suggesting that LAM reduction may cause TE activation during aging (Chen et al., 2016; Morgunova et al., 2022). In humans, PIN1, whose loss has been associated with premature aging, plays a key role in the stabilization of the HP1a in complex with LaminB (Napoletano et al., 2021). Indeed, its loss causes heterochromatin degradation and TE derepression (Napoletano et al., 2021).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Finally, in many animal species, males and females often differ in their lifespan, and this has been correlated with TE abundance (Lemaître et al., 2020; Marais et al., 2018). The amount of TE sequences can differ substantially between males and females due to the presence of a repeat-rich Y or W chromosome in the heterogametic sex. With the disruption of TE-silencing mechanisms with age, more TEs may become active in old individuals from the heterogametic sex, generating more somatic mutations, accelerating aging, and reducing longevity in these individuals (toxic Y effect) (Lemaître et al., 2020; Marais et al., 2018). This hypothesis has been tested in *D. melanogaster*, where aged males showed a more rapid loss of H3K9me2 histone modification at repetitive regions compared to females. Hence, females maintained an efficient TE repression during aging, while males showed an overexpression of TEs, especially on repeats located in the Y chromosome (Brown et al., 2020). This sex-specific TE activation in old flies can lead to mutagenic insertions and genomic instability, resulting in sex-specific cellular aging. However, in *Drosophila miranda*, although the age-associated loss in heterochromatin is more pronounced in males than females, the overall level of TE activation seems to be increased in males relative to females even in young stages (Nguyen & Bachtrog, 2021). The toxic effect of the heterogametic sex has also been studied in avian genomes with the Z/W chromosome system. For instance, in crows, TEs located in the W chromosome were found to be more expressed than insertions located in other chromosomes. Indeed, the W chromosome might be a more permissive environment for TE transcription due to ineffective TE silencing (Warmuth et al., 2022).

Overall, aging is associated with the deregulation and consequent activity of TEs, potentially leading to mutations and DNA damage that can accelerate aging and age-related diseases. As such, longitudinal studies investigating the progressive activity of TEs throughout their lifetime would provide more information about their role in aging and associated pathologies.

Epigenetic changes in diseases are associated with TE activity

Many diseases are associated with genome-wide chromatin changes, potentially leading to an increase in TE activity (Fig. 6.2). Most of these studies have focused on multiple cancer types (Burns, 2017; Grundy et al., 2022) and neurodegenerative disorders (reviewed in Copley & Shorter, 2023). Due to the global epigenetic dysregulation in cancer, that is, changes in the DNA silencing methylation or the alteration of chromatin marks, TEs can be reactivated in malignant cells and have oncogenic potential (Grundy et al., 2022). ERVs are one of the classes of TEs that frequently escape silencing in different types of cancers including colorectal, breast, and prostate cancer, among others (Grundy et al., 2022). Furthermore, L1 elements, the only functional retrotransposon in humans with active autonomous copies, are also recurrently derepressed during oncogenesis and its expression is considered a hallmark of cancer (Ardeljan et al., 2017; McKerrow et al., 2022; Rodić et al., 2014; Shah et al., 2023; Tubio et al., 2014). When being transcribed, L1s code

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

for two proteins, named ORF1p and ORF2p, respectively. One of these proteins, ORF1p has been found in nearly half of all human cancers, especially in high-grade malignant cells, thus being considered an oncogenic marker (Ardeljan et al., 2017; McKerrow et al., 2022; Rodić et al., 2014).

The observed activation of TEs in malignant cells can have in turn detrimental effects, altering the expression of nearby oncogenes, serving as a source of transcription factor binding sites or promoters or even acting as enhancers (Grundy et al., 2022). While human ERVs have been shown to promote ectopic expression of nearby genes in cancer including potential proto-oncogenes (Babaian & Mager, 2016; Cohen et al., 2009; Deniz et al., 2020), L1s have been associated with copy number alteration in breast, ovarian, endometrial, and colon cancers (McKerrow et al., 2022; Rodriguez-Martin et al., 2020). Indeed, L1 elements are part of the most frequent type of somatic structural variation in colorectal cancers and are often responsible for L1-induced deletions, including megabase-scale events, potentially promoting oncogenesis through the loss of tumor-suppressor genes (Rodriguez-Martin et al., 2020).

In the case of tauopathies such as Alzheimer's disease, it has been suggested that derepressed TEs can be a causal factor driving neurodegeneration (Grundman et al., 2021; Guo et al., 2018; Ochoa et al., 2023; Ramirez et al., 2022; Sun et al., 2018). In a *D. melanogaster* model of tauopathy, pathogenic tau expression in the adult brain was correlated with altered levels of TE transcripts and subsequent neuronal death, shortened lifespan, and decreased locomotor activity (Sun et al., 2018). The authors found a similar number of TE families with increased and decreased expression levels compared to controls. Mechanistically, both tau-induced heterochromatin decondensation and *Piwi* and piRNA depletion were the factors associated with the aberrant TE transcription in the adult brain (Sun et al., 2018). While in *Drosophila,* TE transcripts were both increased and decreased in the context of tauopathy, in mice, the effect of tauopathy was mainly associated with an increased expression of TEs. In fact, ERVs were particularly enriched in tau-pathogenic brains (Ramirez et al., 2022).

Both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) can be characterized by a nuclear depletion of TDP-43 in degenerating neurons. TDP-43 is an RNA-binding protein with a prion-like domain that functions as a transcriptional repressor and it is also a key element in the DNA repair mechanisms. An increased expression of TEs has been found in different models of ALS/FTD, including *Drosophila* (Chang et al., 2019; Krug et al., 2017; Romano et al., 2020), mice (Li et al., 2012; Zhang et al., 2019) and humans (Liu et al., 2019). In fact, in humans, ATAC-seq experiments showed that there is a depletion in chromatin accessibility of LINE elements in neurons with TDP-43, further suggesting a role of this protein in TE silencing (Liu et al., 2019). Moreover, in *Drosophila,* the depletion of TDP-43 has been associated with a reduced expression of Dicer-2, an important protein in siRNA processing, thus leading to an increased expression of *Ty3*/Metaviridae (formerly known as *Gypsy* (Wei et al., 2022)) elements (Romano et al., 2020). This was in turn associated with locomotion impairment, a phenotype

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

that was mitigated with the overexpression of Dicer-2 in the TDP-43 null model (Romano et al., 2020). It still remains unclear if therapeutic silencing of TEs could help mitigate neurodegenerative disorders.

Environmental and genomic stress contribute to TE mobilization

In addition to aging and disease, environmental stress has long been proposed to have an impact on TE expression and transposition (Galindo-González et al., 2017; Horváth et al., 2017; Ito, 2022; Lanciano & Mirouze, 2018; Negi et al., 2016) (Fig. 6.2). Stress can activate TEs due to the presence of cis-regulatory elements in their promoter sequences or through epigenetic changes (Horváth et al., 2017; Ito, 2022; Villanueva-Cañas et al., 2019). For instance, under nonstress conditions, L1 elements are repressed by the SIRT6 protein, which coordinates the packaging of these elements into transcriptionally silenced heterochromatin (Van Meter et al., 2014). However, under oxidative stress conditions, SIRT6 is required at DNA-damaged sites and as a consequence, L1s are no longer kept in a heterochromatic state and are reactivated (Van Meter et al., 2014).

Hybridization between different species or within species may contribute to changes in TE activity and is often considered a genomic shock (García Guerreiro, 2012; Michalak, 2008; Rebollo et al., 2010). A thorough case study involves interspecific hybrids between two Australian wallabies where an ERV element (KERV-1), hypomethylated in hybrids, caused centromere expansion in just one generation (Waugh O'Neill et al., 1998). In whitefish, there is a clear link between the DNA demethylation of TEs observed in hybrids between two closely related species and TE overexpression (Laporte et al., 2019). In other systems, such as the *Drosophila buzzatii/Drosophila koepferae*, the analysis of the F1 hybrids did not reveal any changes in the chromatin landscape, despite an increased rate of transposition (Romero-Soriano et al., 2017; Vela et al., 2014). In crosses between *Drosophila mojavensis* and *Drosophila arizonae*, a well-known system to study speciation, an increase of TE expression was observed in F1 hybrids but only for a very limited number of families (Banho et al., 2021). Intraspecific crosses may also lead to massive rearrangements of the genomes due to TE activity. Hybrid dysgenesis, first described in *D. melanogaster*, is a phenomenon occurring when crossing TE unbalanced strains, leading to female sterility at the F1 (Hill et al., 2016; Kidwell, 1983; Mérel et al., 2020). More specifically, males harboring active TE copies are mated with females lacking such copies, producing offspring that are unable to control such active copies due to the absence of maternal piRNAs (Brennecke et al., 2007). The aberrant embryonic activity of such TE copies leads to female sterility and is usually due to one or a couple of TE families (Brennecke et al., 2007; García Guerreiro, 2012; Petrov et al., 1995). The observed transposition rates may be as high as 10^{-3}, having a clear impact on genome structure (Nuzhdin & Mackay, 1994; Vieira & Biémont, 1997).

Stress-mediated activation of TEs is thus suggested to have an evolutionary role as transposition activation generates a higher mutation rate, thus increasing genetic

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

variability upon which natural selection can act (Horváth et al., 2017; Lanciano & Mirouze, 2018). However, the effect of stress on TEs seems to depend on both the type of stress and the type of TEs (Horváth et al., 2017). Overall, modifications in the epigenetic landscape can alter TE-silencing mechanisms, modify its activity and have an impact at the phenotypic level.

How does TE variation affect the epigenetic landscape?

New TE insertions have been documented to affect the epigenetic landscape in many ways. These include the spread of heterochromatin from a TE to neighboring genes and the addition of cis-regulatory regions (Fablet et al., 2023; Hirsch & Springer, 2017; Huang et al., 2022; Rebollo et al., 2011), anchoring of metastable epialleles (Baduel & Colot, 2021; Masson et al., 1987; Morgan et al., 1999; Pignatta et al., 2018), providing pioneering factor binding sites (Taube et al., 2010), as well as providing critical motifs for the establishment and maintenance of the 3D structure of the genome (Choudhary et al., 2023; Golicz et al., 2020; Wang et al., 2021; Yang et al., 2020).

Intraspecific TE insertion polymorphisms can be linked to variation in chromatin accessibility

To appreciate the breadth of impacts TE variation has on the chromatin landscape, including the complex interactions between the aforementioned mechanisms (sometimes antagonistic), we review in this section recent studies combining population genomics and regulatory variation conducted in a growing range of species.

The comparison of epigenetic readouts at or near genomic loci where TEs are either present or absent between individuals allows direct inference of the TE impact on the epigenetic state of the studied region (Fig. 6.3). For instance, in human lymphoblastoid cell lines (LCL), polymorphic insertions of Alu, L1, and SVA retrotransposons can be associated with variation in chromatin accessibility at neighboring loci (monitored with ATAC-seq), with, for a subset of them, correlated effects on gene expression (Goubert et al., 2020). Interestingly, although the effect of TE insertions can correlate with either an increase or decrease in chromatin accessibility, TEs inserted directly within ATAC peaks (active chromatin regions) are systematically associated with repressed states. In mice, TE-specific effects have been reported in embryonic stem cells (mESC) as L1 elements carrying intact 5′ promoter regions are associated with increased chromatin accessibility, while IAP (ERVs) most often correlate with repressive chromatin (Ferraj et al., 2023). Similarly to human LCL, only a subset of these TEs can be linked to variation in gene expression, highlighting the complex interplay between chromatin state and gene expression. In maize, which displays extreme values both regarding TE content and insertional polymorphisms, layering methylation levels and chromatin accessibility readout via ATAC-seq also depict a two-faced influence of TE polymorphisms. On the one hand, TEs can

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

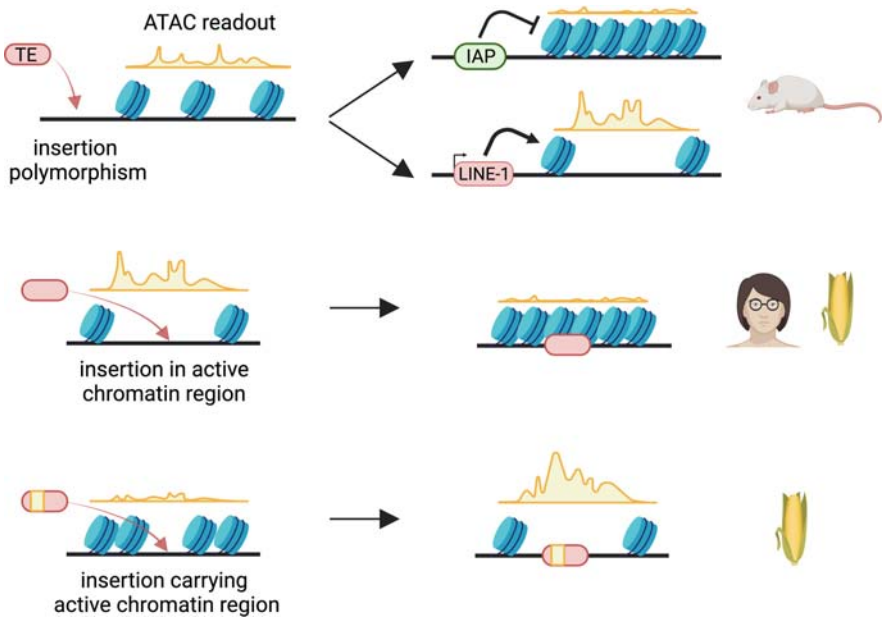


FIGURE 6.3

Effects of transposable element (TE) insertion polymorphism on the epigenetic landscapes of their host species. ATAC-seq readouts can be used to detect associations between TE insertion and chromatin accessibility. In mice, IAP elements (class I retrotransposons) can be associated with repressed chromatin states, while L1 elements with intact 5′ promoters correlate with accessible chromatin regions. In humans and maize, TE insertions within active chromatin regions (ACR) generally correlate with decreased accessibility. Alternatively, TEs carrying ACR in maize can lead to increased accessibility at insertion sites.

interrupt active chromatin regions, leading to changes in the chromatin conformation (Noshay et al., 2020). On the other hand, several TE insertions harbor such active chromatin regions that can be conserved upon insertion. In turn, and similarly to findings in animal models, a subset of these genomic variants affecting chromatin state also correlate with nearby gene expression.

Lineage-specific TE insertions highlight the complex relationship between genetic variation, epigenetic changes and genomic functions

At larger evolutionary scales, the combined effect of drift and natural selection leads a subset of TE insertions to fixation, in a species- or lineage-specific fashion. Lineage-specific insertions have been shown to provide the building blocks for evolutionary innovation through co-option (often dubbed as "domestication") of

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

TE-derived sequences in a wide array of contexts such as development and cell identity (Barnada et al., 2022; Etchegaray et al., 2021; Hsu et al., 2021; Ye et al., 2020), immune (Chuong et al., 2016; Kelly et al., 2022) and stress response (Garrigues et al., 2019; Qiu & Köhler, 2020; Villanueva-Cañas et al., 2019), as well as the establishment of adaptive phenotypes (Ruggieri et al., 2022). With the increased availability of genome-wide epigenome sequencing, new comparative epigenomic approaches have emerged as a powerful tool to understand how lineage-specific TEs potentiate epi-variation and shape gene regulatory networks (Venuto & Bourque, 2018).

Recent studies focused on the spread of repressive histone marks from TEs in *Drosophila* depict significant yet complex relationships between TE abundance and heterochromatin spreading (Fig. 6.4). Specifically, several authors have reported a stronger influence of euchromatic TE insertions on the enrichment in

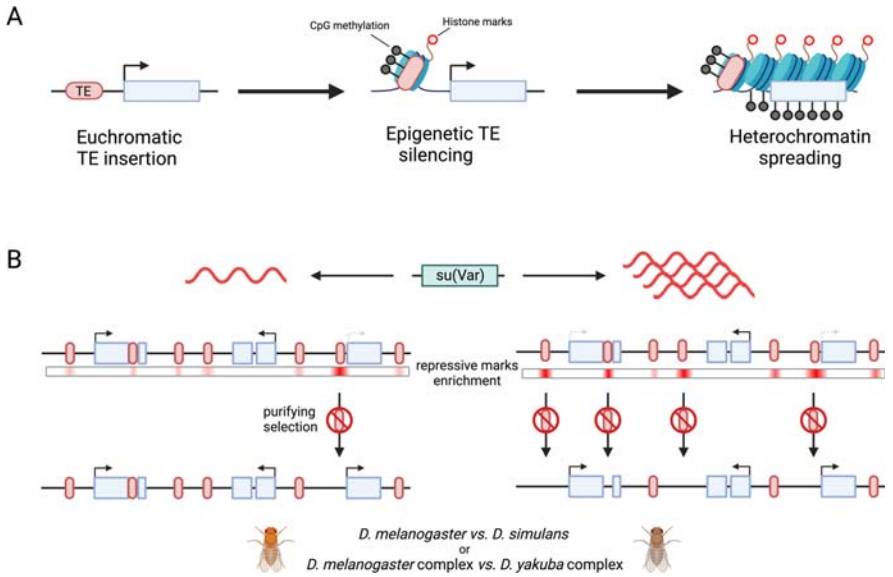


FIGURE 6.4

Complex relationships between transposable element (TE) insertions and spread of heterochromatin. (A) New TE insertions are rapidly targeted by epigenetic host defense mechanisms such as repressive histone marks, eventually leading to the spread of heterochromatin to nearby regions. (B) Recent studies in *Drosophila* depict a complex relationship between the abundance of euchromatic TE insertions and the magnitude of their epigenetic effects. Species with relatively higher euchromatic TE content (left) show lower expression of *Su(var)* genes, involved in heterochromatin maintenance, relative to species harboring lower densities of euchromatic TEs. As the magnitude of TE-derived epigenetic effect correlates with *Su(var)* gene expression, natural selection might be stronger against euchromatic insertions in species with higher levels of *Su(var)* (right), in order to mitigate deleterious consequences of such effect.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

repressive marks at nearby loci in *D. simulans* compared to *D. melanogaster*, the former harboring significantly fewer TEs than the latter (Fablet et al., 2023; Lee & Karpen, 2017). Using larger species samples, this inverse correlation between euchromatic TEs and epigenetic effects holds within species of the *D. melanogaster* or *D. yakuba* complexes (Huang et al., 2022). Based on population frequency estimates of different families of TEs, the same authors suggest that purifying selection is stronger against TE insertions associated with larger epigenetic effects. Furthermore, Huang et al. (2022) observed that the expression of *Su(var)* genes, involved in heterochromatin maintenance in *Drosophila*, correlates with the average magnitude of H3K9me2 enrichment at TEs across species. These results indicate that species-specific levels of *Su(var)* expression (influenced by the density of heterochromatin repeats) correlate with the epigenetic effect of euchromatic TEs. While the effect on neighboring gene expression, and eventually the individuals' fitness remains elusive (Fablet et al., 2023; Huang et al., 2022), species with stronger TE epigenetic effects may reinforce selection against their euchromatic TEs.

Another significant contribution of lineage-specific TEs to epigenetic landscape variation stems from their contribution to the 3D structure of the genome. Thanks to chromosome conformation capture techniques such as Hi-C, topologically associating domains (TADs) have been discovered in a wide array of species including vertebrates (D. Li et al., 2022), invertebrates (Dixon, 2019; Szabo et al., 2018), plants (Dong et al., 2017), and fungi (Winter et al., 2018). TADs are typically delineated by insulator elements, which prevent the spread of *cis*-regulatory factors across domains (Fig. 6.5A). In mammals, TAD boundaries and chromatin loops anchored within TADs are enriched in CTCF-binding motifs, for which between 8% and 37% are contributed by identifiable repeated elements (Fig. 6.5B). Note that CTCF-independent insulator elements present in mammalian TEs such as MIR can also contribute to TADs (Wang et al., 2015). Comparing human, mouse, dog, and rhesus macaque, Choudhary et al. (2023) observed that conserved TAD across species do not necessarily harbor similar TEs, suggesting that lineage-specific insertions may provide functional redundancy, making up for the eventual loss of binding domain via point mutations. On the other hand, the same authors showed that species-specific TADs harbor significantly more TEs than conserved ones, highlighting their role in the functional diversification of genomes. TEs may also contribute to lineage-specific compartmentalization of the genome in plants; however, further research is needed to elucidate their precise implication (Wang et al., 2021). It is worth mentioning that the influence of the 3D structure of the genome on gene expression remains elusive and certainly involves mechanisms yet to be discovered.

The examples reported in this section demonstrate the pervasive influence of TE variation in shaping the epigenetic context. Nevertheless, causality—addressed by the means of candidate-specific molecular assays—still remains to be assessed in the vast majority of cases, in order to expand our knowledge about proximal mechanisms and their interactions.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

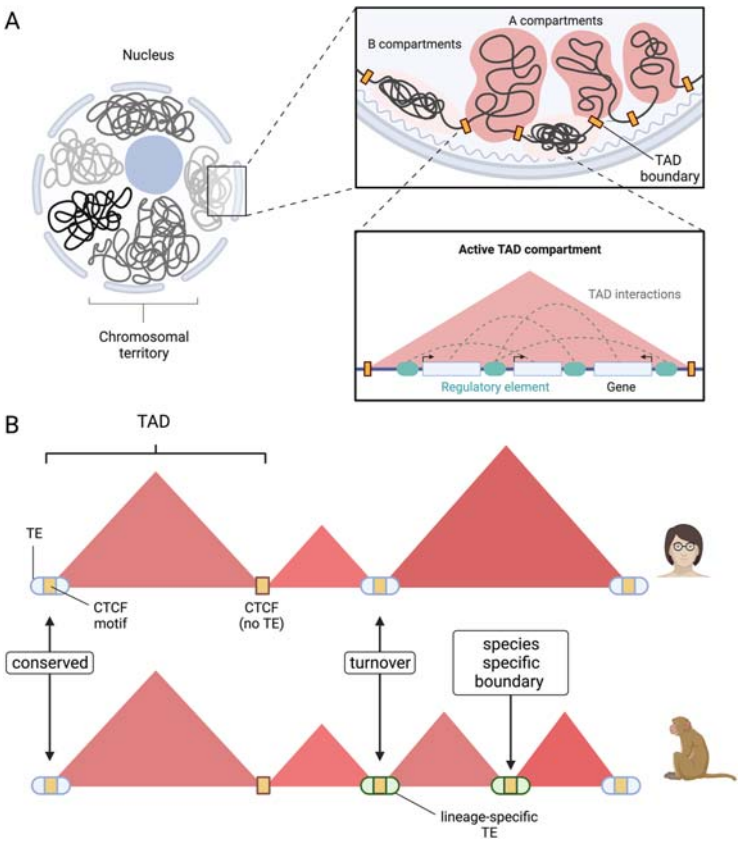


FIGURE 6.5

Transposable element (TE) contribution to the 3D genome. (A) Eukaryotes' DNA is organized in chromosomal territories within the cell nucleus. Within chromosomal territories, active (A) and inactive (B) compartments can be distinguished. Compartments are delineated by insulator elements defining topologically associating domains (TADs). Active TADs include genes and their regulatory elements, connected spatially through chromatin loops exposed to transcription factors, enzymes and other active molecules. (B) In mammals, TAD boundaries are enriched in CTCF-binding domains, which often occur in TEs. Comparative epigenomic studies have shown that CTCF-harboring TEs can contribute to the conservation of TADs across species, through the maintenance of orthologous insertions (conserved) or by replacing a potentially mutated site with the insertion of a species-specific TE containing a new CTCF binding site (turnover). Additionally, species-specific TE insertions are also responsible for the establishment of species-specific TADs, suggesting that they may contribute to functional divergence between species via reorganization of the 3D structure of the genome.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Evolutionary consequences of epigenetic regulation of TEs

A growing number of studies have directed their focus on the contribution of TEs to phenotypes and more recently to phenotypic plasticity and adaptation (Casacuberta & González, 2013; Marin et al., 2020). Phenotypic plasticity is defined by the ability of one genotype to produce a range of phenotypes in response to changes in the environment (Bradshaw, 1965). While a wide array of mechanisms underlies phenotypic plasticity (Kelly et al., 2012; Schlichting & Smith, 2002), TEs can be involved as a source of epigenetic variation. One of the earliest mentions of the role of TEs on phenotypic plasticity is the differential expression of the *agouti* gene in mice, due to different levels of methylation of a TE inserted upstream of the gene, resulting in striking contrast in the agouti pigmentation phenotype (Morgan et al., 1999). In. *D. melanogaster,* Guio and collaborators (Guio et al., 2018) investigated the histone marks associated with the *Bari-Jheh* transposon under oxidative stress (via paraquat exposure) revealing locus-specific enrichment for H3K4me3, H3K9me3, and H3K27me3. This study showed a consistent pattern with the expression of the nearby genes, demonstrating an epigenetic effect of TEs on genes under stressful conditions. Niu and collaborators (Niu et al., 2019) recently investigated the phenotypic variation in *Capsella rubella*, the pink shepherd's-purse, arguing that despite a diminished genetic variation due to a recent bottleneck, *C. rubella* harbors a high phenotypic diversity associated with a TE-rich genome, compared to its sister species *Capsella grandiflora*.

Adaptive phenotypes (i.e., underlined by genetic variation) can also involve the epigenetic machinery responsive to the environment. In maize, Sun and collaborators (Sun et al., 2023) discovered a trade-off between yield and drought tolerance, modulated by the presence or absence of TE-derived inverted repeats (IRs) in the major expression QTL *DRESH8*. Insertions from several *Ty3*/Metaviridae elements within the third intron of the gene *ZmPP2C16* formed an IR, which presence in some populations is strongly correlated with higher expression of small RNAs derived from this locus (small RNA eQTL). Indeed, CRISPR-Cas9 directed deletion of *DRESH8* shows a significant reduction in expression of *Ty3*-derived IRs' sRNAs, coupled with higher drought tolerance. The target for these TE-derived sRNAs, *ZmMYBR38*, encodes the *MYB* transcription factor, which experimental overexpression in maize and *Arabidopsis* showed higher tolerance to drought. Nevertheless, the $DRESH8^{+}$ allele, carrying the TE-derived IR, shows evidence of selective sweep (recent positive selection) when compared to teosinte (wild relative of maize), and a significant association with kernel size, suggesting that this TE-derived allele was selected during domestication. The authors also observed that the expression of these inverted-TE-derived-sRNAs varies with other abiotic factors and suggest that inverted-TE-derived-sRNAs might mediate adaptation to various environments. This example illustrates the implication of agriculture in the selection of strong TE-derived phenotypes, and the evolutionary trade-off between yield and drought orchestrated by this particular TE.

The scientific community has long drifted from the concept of "junk DNA". Nowadays, more and more studies emerge on the evolutionary role of TEs in the adaptation of species to new environments and particularly when it comes to invasive species (Mounger et al., 2021). It is worth mentioning that TE-derived phenotypes, influenced

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

by the epigenetic landscape in which they reside, may simply remain neutral, as one can expect from evolutionary novelties (Guerrero-Bosagna, 2017). To paraphrase Dobzhansky, there is good reason to believe that "*Nothing about mobile elements makes sense except in the light of evolution.*" (Kidwell & Lisch, 1997).

Conclusion

The intricate relationship between TE mobilization and hosts' epigenetic landscapes is a fundamental characteristic of eukaryotic genomes. In this chapter, we explored the vital epigenetic countermeasures that continuously evolve to control the unchecked spread of TEs within and between genomes. Stress (environmental and genomic), diseases, or aging can impair the epigenetic machinery, creating an opportunity for some TEs to proliferate and introduce new genomic variants. As new TEs invade individuals and species genomes, these copies can suffer epigenetic regulation, while also providing new or disrupting existing regulatory elements. TE insertional polymorphism, whether within or between species, can lead to epigenetic changes and, in various cases, rewire genes' regulatory networks. Understanding the consequences for the host requires considering evolution, where TE-derived phenotypes and their underlying genomic variants (TEs themselves or mutations affecting their epigenetic regulation) undergo selection-drift balance at the population level.

Assessing the epigenome-wide effects of TEs remains challenging due to the heterogeneity of TE types, transposition mechanisms, and epigenetic regulations. Specific effects may exist in a locus-specific manner, while existing methods are mostly designed to study TEs at the family or subfamily levels. Fortunately, the proliferation of long-read omics and associated statistical methods promises better insights into the interplay between TEs and epigenetic variation, with the potential to understand their elusive connection with gene expression. Comparative epigenomics approaches, combining population-wide genomic and epigenetic variation, are ideal for identifying candidates, but functional validation through molecular assays remains essential to elucidate proximal mechanisms.

Acknowledgments

We thank Marie Fablet for her valuable comments. All figures were created with BioRender.com.

References

Almeida da Paz, M., & Taher, L. (2022). T3E: A tool for characterising the epigenetic profile of transposable. *Mobile DNA, 13*(1), 29. https://doi.org/10.1186/s13100-022-00285-z

Ardeljan, D., Taylor, M. S., Ting, D. T., & Burns, K. H. (2017). The human long interspersed element-1 retrotransposon: An emerging biomarker of Neoplasia. *Clinical Chemistry, 63*(4), 816−822. https://doi.org/10.1373/clinchem.2016.257444

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Arkhipova, I. R. (2018). Neutral theory, transposable elements, and eukaryotic genome evolution. *Molecular Biology and Evolution, 35*(6), 1332−1337. https://doi.org/10.1093/molbev/msy083

Babaian, A., & Mager, D. L. (2016). Endogenous retroviral promoter exaptation in human cancer. *Mobile DNA, 7*(1), 1−21. https://doi.org/10.1186/s13100-016-0080-x

Baduel, P., & Colot, V. (2021). The epiallelic potential of transposable elements and its evolutionary. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 376*(1826), 20200123. https://doi.org/10.1098/rstb.2020.0123

Banho, C. A., Oliveira, D. S., Haudry, A., Fablet, M., Vieira, C., & Carareto, C. M. A. (2021). Transposable element expression and regulation profile in gonads of. *Cells, 10*(12). https://doi.org/10.3390/cells10123574

Barnada, S. M., Isopi, A., Tejada-Martinez, D., Goubert, C., Patoori, S., Pagliaroli, L., Tracewell, M., Trizzino, M., & Feschotte, C. (2022). Genomic features underlie the co-option of SVA transposons as cis-regulatory elements in human pluripotent stem cells. *PLoS Genetics, 18*(6), e1010225. https://doi.org/10.1371/journal.pgen.1010225

Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J., & Edwards, D. (2020). Plant pan-genomes are the new reference. *Nature Plants, 6*(8), 914−920. https://doi.org/10.1038/s41477-020-0733-0

Bell, E. A., Butler, C. L., Oliveira, C., Marburger, S., Yant, L., & Taylor, M. I. (2022). Transposable element annotation in non-model species: The benefits of species-specific repeat libraries using semi-automated EDTA and DeepTE de novo pipelines. *Molecular Ecology Resources, 22*(2), 823−833. https://doi.org/10.1111/1755-0998.13489

Bendall, M. L., de Mulder, M., Iñiguez, L. P., Lecanda-Sánchez, A., Pérez-Losada, M., Ostrowski, M. A., Jones, R. B., Mulder, L. C. F., Reyes-Terán, G., Crandall, K. A., Ormsby, C. E., & Nixon, D. F. (2019). Telescope: Characterization of the retrotranscriptome by accurate. *PLoS Computational Biology, 15*(9), e1006453. https://doi.org/10.1371/journal.pcbi.1006453

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology, 19*(1), 199. https://doi.org/10.1186/s13059-018-1577-z

Bracewell, R., Chatla, K., Nalley, M. J., & Bachtrog, D. (2019). Dynamic turnover of centromeres drives karyotype evolution in Drosophila. *Elife, 8*. https://doi.org/10.7554/eLife.49002

Bradshaw, A. D. (1965). Evolutionary significance of phenotypic plasticity in plants. In E. W. Caspari, & J. M. Thoday (Eds.), *Advances in genetics* (Vol. 13, pp. 115−155). Academic Press. Article ISSN 0065-2660, ISBN 9780120176137. https://doi.org/10.1016/S0065-2660(08)60048-6

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon. *Cell, 128*(6), 1089−1103. https://doi.org/10.1016/j.cell.2007.01.043

Brown, E. J., Nguyen, A. H., & Bachtrog, D. (2020). The Y chromosome may contribute to sex-specific ageing in Drosophila. *Nature Ecology and Evolution, 4*(6), 853−862. https://doi.org/10.1038/s41559-020-1179-5

Burns, K. H. (2017). Transposable elements in cancer. *Nature Reviews Cancer, 17*(7), 415−424. https://doi.org/10.1038/nrc.2017.35

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Cai, L., Wu, Y., & Gao, J. (2019). DeepSV: Accurate calling of genomic deletions from high-throughput sequencing data using deep convolutional neural network. *BMC Bioinformatics, 20*(1). https://doi.org/10.1186/s12859-019-3299-y

Casacuberta, E., & González, J. (2013). The impact of transposable elements in environmental adaptation. *Molecular Ecology, 22*(6), 1503−1517. https://doi.org/10.1111/mec.12170

Chang, C.-H., Chavan, A., Palladino, J., Wei, X., Martins, N. M. C., Santinello, B., Chen, C.-C., Erceg, J., Beliveau, B. J., Wu, C.-T., Larracuente, A. M., & Mellone, B. G. (2019). Islands of retroelements are major components of Drosophila centromeres. *PLoS Biology, 17*(5), e3000241. https://doi.org/10.1371/journal.pbio.3000241

Chen, H., Zheng, X., Xiao, D., & Zheng, Y. (2016). Age-associated de-repression of retrotransposons in the Drosophila fat. *Aging Cell, 15*(3). https://doi.org/10.1111/acel.12465

Chen, H., Zheng, X., & Zheng, Y. (2014). Age-associated loss of lamin-b leads to systemic inflammation and gut hyperplasia. *Cell, 159*(4), 829−843. https://doi.org/10.1016/j.cell.2014.10.028

Choi, J. Y., & Lee, Y. C. G. (2020). Double-edged sword: The evolutionary consequences of the epigenetic. *PLoS Genetics, 16*(7), e1008872. https://doi.org/10.1371/journal.pgen.1008872

Choudhary, M. N. K., Quaid, K., Xing, X., Schmidt, H., & Wang, T. (2023). Widespread contribution of transposable elements to the rewiring of mammalian 3D genomes. *Nature Communications, 14*(1). https://doi.org/10.1038/s41467-023-36364-9

Chu, C., Borges-Monroy, R., Viswanadham, V. V., Lee, S., Li, H., Lee, E. A., & Park, P. J. (2021). Comprehensive identification of transposable element insertions using. *Nature Communications, 12*(1), 1−12. https://doi.org/10.1038/s41467-021-24041-8

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. *Nature Reviews Genetics, 18*(2), 71−86. https://doi.org/10.1038/nrg.2016.139

Chuong, E. B., Elde, N. C., & Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science, 351*(6277), 1083−1087. https://doi.org/10.1126/science.aad5497

Cohen, C. J., Lock, W. M., & Mager, D. L. (2009). Endogenous retroviral LTRs as promoters for human genes: A critical assessment. *Gene, 448*(2), 105−114. https://doi.org/10.1016/j.gene.2009.06.020

Cook, D. E., Kramer, H. M., Torres, D. E., Seidl, M. F., & Thomma, B. P. H. J. (2020). A unique chromatin profile defines adaptive genomic regions in a fungal plant pathogen. *Elife, 9*. https://doi.org/10.7554/elife.62208

Copley, K. E., & Shorter, J. (2023). Repetitive elements in aging and neurodegeneration. *Trends in Genetics, 39*(5), 381−400. https://doi.org/10.1016/j.tig.2023.02.008

Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M., & Neretti, N. (2014). Transcriptional landscape of repetitive elements in normal and cancer. *BMC Genomics, 15*, 583. https://doi.org/10.1186/1471-2164-15-583

Czech, B., & Hannon, G. J. (2016). One loop to rule them all: The ping-pong cycle and piRNA-guided silencing. *Trends in Biochemical Sciences, 41*(4), 324−337. https://doi.org/10.1016/j.tibs.2015.12.008

Daron, J., & Slotkin, R. K. (2017). EpiTEome: Simultaneous detection of transposable element insertion sites and their DNA methylation levels. *Genome Biology, 18*(1). https://doi.org/10.1186/s13059-017-1232-0

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

De Cecco, M., Criscione, S. W., Peterson, A. L., Neretti, N., Sedivy, J. M., & Kreiling, J. A. (2013). Transposable elements become active and mobile in the genomes of aging. *Aging, 5*(12). https://doi.org/10.18632/aging.100621

De Cecco, M., Ito, T., Petrashen, A. P., Elias, A. E., Skvir, N. J., Criscione, S. W., Caligiana, A., Brocculi, G., Adney, E. M., Boeke, J. D., Le, O., Beauséjour, C., Ambati, J., Ambati, K., Simon, M., Seluanov, A., Gorbunova, V., Slagboom, P. E., Helfand, S. L., Neretti, N., & Sedivy, J. M. (2019). L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature, 566*(7742), 73−78. https://doi.org/10.1038/s41586-018-0784-9

de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genetics, 7*(12), e1002384. https://doi.org/10.1371/journal.pgen.1002384

Della Coletta, R., Qiu, Y., Ou, S., Hufford, M. B., & Hirsch, C. N. (2021). How the pan-genome is changing crop genomics and improvement. *Genome Biology, 22*(1), 1−19. https://doi.org/10.1186/s13059-020-02224-8

Deniz, Ö., Ahmed, M., Todd, C. D., Rio-Machin, A., Dawson, M. A., & Branco, M. R. (2020). Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. *Nature Communications, 11*(1). https://doi.org/10.1038/s41467-020-17206-4

Dixon, J. R. (2019). TADs for life: Chromatin domain organization regulates lifespan in *C. elegans*. *Developmental Cell, 51*(2), 131−132. https://doi.org/10.1016/j.devcel.2019.09.021

Dong, P., Tu, X., Chu, P. Y., Zhu, N., Grierson, D., Du, B., Li, P., & Zhong, S. (2017). 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Molecular Plant, 10*(12), 1497−1509. https://doi.org/10.1016/j.molp.2017.11.005

Driver, C. J., & McKechnie, S. W. (1992). Transposable elements as a factor in the aging of *Drosophila melanogaster*. *Annals of the New York Academy of Sciences, 673*. https://doi.org/10.1111/j.1749-6632.1992.tb27439.x

Duc, C., Yoth, M., Jensen, S., Mouniée, N., Bergman, C. M., Vaury, C., & Brasset, E. (2019). Trapping a somatic endogenous retrovirus into a germline piRNA cluster. *Genome Biology, 20*(1), 1−14. https://doi.org/10.1186/s13059-019-1736-x

Ebert, P., Audano, P. A., Zhu, Q., Rodriguez-Martin, B., Porubsky, D., Bonder, M. J., Sulovari, A., Ebler, J., Zhou, W., Serra Mari, R., Yilmaz, F., Zhao, X., Hsieh, P. H., Lee, J., Kumar, S., Lin, J., Rausch, T., Chen, Y., Ren, J., … Eichler, E. E. (2021). Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science, 372*(6537). https://doi.org/10.1126/science.abf7117

Elsner, D., Meusemann, K., & Korb, J. (2018). Longevity and transposon defense, the case of termite reproductives. *Proceedings of the National Academy of Sciences of the United States of America, 115*(21), 5504−5509. https://doi.org/10.1073/pnas.1804046115

Etchegaray, E., Naville, M., Volff, J.-N., & Haftek-Terreau, Z. (2021). Transposable element-derived sequences in vertebrate development. *Mobile DNA, 12*(1), 1−24. https://doi.org/10.1186/s13100-020-00229-5

Ewing, A. D., Smits, N., Sanchez-Luque, F. J., Faivre, J., Brennan, P. M., Richardson, S. R., Cheetham, S. W., & Faulkner, G. J. (2020). Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Molecular Cell, 80*(5), 915−928. https://doi.org/10.1016/j.molcel.2020.10.024

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Fabian, D. K., Dönertaş, H. M., Fuentealba, M., Partridge, L., & Thornton, J. M. (2021). Transposable element landscape in Drosophila populations selected for. *Genome Biology and Evolution, 13*(4). https://doi.org/10.1093/gbe/evab031

Fablet, M., Salces-Ortiz, J., Jacquet, A., Menezes, B. F., Dechaud, C., Veber, P., Rebollo, R., & Vieira, C. (2023). A quantitative, genome-wide analysis in *Drosophila* reveals transposable elements' influence on gene expression is species-specific. *Genome Biology and Evolution, 15*(9), evad160. https://doi.org/10.1093/gbe/evad160

Ferraj, A., Audano, P. A., Balachandran, P., Czechanski, A., Flores, J. I., Radecki, A. A., Mosur, V., Gordon, D. S., Walawalkar, I. A., Eichler, E. E., Reinholdt, L. G., & Beck, C. R. (2023). Resolution of structural variation in diverse mouse genomes reveals chromatin remodeling due to transposable elements. *Cell Genomics, 3*(5). https://doi.org/10.1016/j.xgen.2023.100291

Galindo-González, L., Mhiri, C., Deyholos, M. K., & Grandbastien, M. A. (2017). LTR-retrotransposons in plants: Engines of evolution. *Gene, 626*, 14−25. https://doi.org/10.1016/j.gene.2017.04.051

Gamdzyk, M., Doycheva, D. M., Araujo, C., Ocak, U., Luo, Y., Tang, J., & Zhang, J. H. (2020). cGAS/STING pathway activation contributes to delayed neurodegeneration in. *Molecular Neurobiology, 57*(6). https://doi.org/10.1007/s12035-020-01904-7

García Guerreiro, M. P. (2012). What makes transposable elements move in the Drosophila genome? *Heredity, 108*(5), 461−468. https://doi.org/10.1038/hdy.2011.89

Gardner, E. J., Lam, V. K., Harris, D., Chuang, N. T., Scott, E. C., Pittard, W. S., Mills, R. E., 1000 Genomes Project Consortium, Devine, S. E. (2017). The mobile element locator tool (MELT): Population-scale mobile element. *Genome Research, 27*(11), 1916−1929. https://doi.org/10.1101/gr.218032.116

Garrigues, J. M., Tsu, B. V., Daugherty, M. D., & Pasquinelli, A. E. (2019). Diversification of the Caenorhabditis heat shock response by Helitron transposable elements. *Elife, 8*. https://doi.org/10.7554/elife.51139

Garrison, E., Guarracino, A., Heumos, S., Villani, F., Bao, Z., Tattini, L., Hagmann, J., Vorbrugg, S., Marco-Sola, S., Kubica, C., Ashbrook, D. G., Thorell, K., Rusholme-Pilcher, R. L., Liti, G., Rudbeck, E., Nahnsen, S., Yang, Z., Moses, M. N., Nobrega, F. L., … Prins, P. (2023). Building pangenome graphs. *bioRxiv*. https://doi.org/10.1101/2023.04.05.535718

Garrison, E., Sirén, J., Novak, A. M., Hickey, G., Eizenga, J. M., Dawson, E. T., Jones, W., Garg, S., Markello, C., Lin, M. F., Paten, B., & Durbin, R. (2018). Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nature Biotechnology, 36*(9), 875−879. https://doi.org/10.1038/nbt.4227

Giassa, I.-C., & Alexiou, P. (2021). Bioinformatics and machine learning approaches to understand the regulation of mobile genetic elements. *Biology, 10*(9), 896. https://doi.org/10.3390/biology10090896

Giménez-Orenga, K., & Oltra, E. (2023). *Transposable elements shaping the epigenome*. Elsevier BV. https://doi.org/10.1016/b978-0-323-91909-8.00035-9

Golicz, A. A., Bhalla, P. L., Edwards, D., & Singh, M. B. (2020). Rice 3D chromatin structure correlates with sequence variation and meiotic recombination rate. *Communications Biology, 3*(1). https://doi.org/10.1038/s42003-020-0932-2

Goubert, C., Zevallos, N. A., & Feschotte, C. (2020). Contribution of unfixed transposable element insertions to human. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 375*(1795), 20190331. https://doi.org/10.1098/rstb.2019.0331

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Groza, C., Chen, X., Pacis, A., Simon, M. M., Pramatarova, A., Aracena, K. A., Pastinen, T., Barreiro, L. B., & Bourque, G. (2023). Genome graphs detect human polymorphisms in active epigenomic state during influenza infection. *Cell Genomics, 3*(5). https://doi.org/10.1016/j.xgen.2023.100294

Groza, C., Kwan, T., Soranzo, N., Pastinen, T., & Bourque, G. (2020). Personalized and graph genomes reveal missing signal in epigenomic data. *Genome Biology, 21*(1), 1−22. https://doi.org/10.1186/s13059-020-02038-8

Grundman, J., Spencer, B., Sarsoza, F., & Rissman, R. A. (2021). Transcriptome analyses reveal tau isoformdriven changes in transposable element and gene expression. *PLoS One, 16*(9). https://doi.org/10.1371/journal.pone.0251611

Grundy, E. E., Diab, N., & Chiappinelli, K. B. (2022). Transposable element regulation and expression in cancer. *FEBS Journal, 289*(5), 1160−1179. https://doi.org/10.1111/febs.15722

Guerrero-Bosagna, C. (2017). Evolution with No reason: A neutral view on epigenetic changes, genomic. *BioScience, 67*(5), 469−476. https://doi.org/10.1093/biosci/bix021

Guio, L., Vieira, C., & González, J. (2018). Stress affects the epigenetic marks added by natural transposable element insertions in *Drosophila melanogaster*. *Scientific Reports, 8*(1). https://doi.org/10.1038/s41598-018-30491-w

Guo, C., Jeong, H. H., Hsieh, Y. C., Klein, H. U., Bennett, D. A., De Jager, P. L., Liu, Z., & Shulman, J. M. (2018). Tau activates transposable elements in Alzheimer's disease. *Cell Reports, 23*(10), 2874−2880. https://doi.org/10.1016/j.celrep.2018.05.004

Hill, T., Schlötterer, C., & Betancourt, A. J. (2016). Hybrid dysgenesis in *Drosophila simulans* associated with a rapid invasion of the P-element. *PLoS Genetics, 12*(3). https://doi.org/10.1371/journal.pgen.1005920

Hirsch, C. D., & Springer, N. M. (2017). Transposable element influences on gene expression in plants. *Biochimica et Biophysica Acta-Gene Regulatory Mechanisms, 1860*(1), 157−165. https://doi.org/10.1016/j.bbagrm.2016.05.010

Horváth, V., Merenciano, M., & González, J. (2017). Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends in Genetics, 33*(11), 832−841. https://doi.org/10.1016/j.tig.2017.08.007

Hoyt, S. J., Storer, J. M., Hartley, G. A., Grady, P. G. S., Gershman, A., de Lima, L. G., Limouse, C., Halabian, R., Wojenski, L., Rodriguez, M., Altemose, N., Rhie, A., Core, L. J., Gerton, J. L., Makalowski, W., Olson, D., Rosen, J., Smit, A. F. A., Straight, A. F., … O'Neill, R. J. (2022). From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science, 376*(6588). https://doi.org/10.1126/science.abk3112

Hsu, P.-S., Yu, S.-H., Tsai, Y.-T., Chang, J.-Y., Tsai, L.-K., Ye, C.-H., Song, N.-Y., Yau, L.-C., & Lin, S.-P. (2021). More than causing (epi)genomic instability: Emerging physiological implications of transposable element modulation. *Journal of Biomedical Science, 28*(1). https://doi.org/10.1186/s12929-021-00754-2

Huang, Y., Shukla, H., & Lee, Y. C. G. (2022). Species-specific chromatin landscape determines how transposable elements. *Elife, 11*. https://doi.org/10.7554/eLife.81567

Ito, H. (2022). Environmental stress and transposons in plants. *Genes & Genetic Systems, 97*(4), 169−175. https://doi.org/10.1266/ggs.22-00045

Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). TEtranscripts: A package for including transposable elements in. *Bioinformatics, 31*(22), 3593−3599. https://doi.org/10.1093/bioinformatics/btv422

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Jones, B. C., Wood, J. G., Chang, C., Tam, A. D., Franklin, M. J., Siegel, E. R., & Helfand, S. L. (2016). A somatic piRNA pathway in the Drosophila fat body ensures metabolic. *Nature Communications, 7*. https://doi.org/10.1038/ncomms13856

Kaufmann, K., Muiño, J. M., Østerås, M., Farinelli, L., Krajewski, P., & Angenent, G. C. (2010). Chromatin immunoprecipitation (ChIP) of plant transcription factors. *Nature Protocols, 5*(3), 457−472. https://doi.org/10.1038/nprot.2009.244

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single. *Nature Communications, 10*(1), 1930. https://doi.org/10.1038/s41467-019-09982-5

Kelly, C. J., Chitko-McKown, C. G., & Chuong, E. B. (2022). Ruminant-specific retrotransposons shape regulatory evolution of bovine immunity. *Genome Research, 32*(8), 1474−1486. https://doi.org/10.1101/gr.276241.121

Kelly, S. A., Panhuis, T. M., & Stoehr, A. M. (2012). Phenotypic plasticity: Molecular mechanisms and adaptive significance. *Comprehensive Physiology, 2*(2), 1417−1439. https://doi.org/10.1002/cphy.c110008

Kidwell, M. G. (1983). Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America, 80*(6), 1655−1659. https://doi.org/10.1073/pnas.80.6.1655

Kidwell, M. G., & Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Sciences of the United States of America, 94*(15), 7704−7711. https://doi.org/10.1073/pnas.94.15.7704

Klein, S. P., & Anderson, S. N. (2022). The evolution and function of transposons in epigenetic regulation in response to the environment. *Current Opinion in Plant Biology, 69*, 102277. https://doi.org/10.1016/j.pbi.2022.102277

Krug, L., Chatterjee, N., Borges-Monroy, R., Hearn, S., Liao, W. W., Morrill, K., Prazak, L., Rozhkov, N., Theodorou, D., Hammell, M., & Dubnau, J. (2017). Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS. *PLoS Genetics, 13*(3). https://doi.org/10.1371/journal.pgen.1006635

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell, 153*(6), 1194. https://doi.org/10.1016/j.cell.2013.05.039

Lanciano, S., & Cristofari, G. (2020). Measuring and interpreting transposable element expression. *Nature Reviews Genetics, 21*(12). https://doi.org/10.1038/s41576-020-0251-y

Lanciano, S., & Mirouze, M. (2018). Transposable elements: All mobile, all different, some stress responsive, some adaptive? *Current Opinion in Genetics & Development, 49*, 106−114. https://doi.org/10.1016/j.gde.2018.04.002

Laporte, M., Le Luyer, J., Rougeux, C., Dion-Côté, A. M., Krick, M., & Bernatchez, L. (2019). DNA methylation reprogramming, TE derepression, and postzygotic isolation of nascent animal species. *Science Advances, 5*(10). https://doi.org/10.1126/sciadv.aaw1644

LaRocca, T. J., Cavalier, A. N., & Wahl, D. (2020). Repetitive elements as a transcriptomic marker of aging: Evidence in. *Aging Cell, 19*(7). https://doi.org/10.1111/acel.13167

Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., & Heidmann, T. (2013). Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 368*(1626), 20120507. https://doi.org/10.1098/rstb.2012.0507

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., Hur, J. K., Aravin, A. A., & Tóth, K. F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a. *Genes & Development, 27*(4), 390−399. https://doi.org/10.1101/gad.209841.112

Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y.-C. A., Luo, Y., Sachidanandam, R., Toth, K. F., Patel, D., & Aravin, A. A. (2014). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing. *Genes & Development, 28*(15), 1667. https://doi.org/10.1101/gad.245514.114

Lee, I., Razaghi, R., Gilpatrick, T., Molnar, M., Gershman, A., Sadowski, N., Sedlazeck, F. J., Hansen, K. D., Simpson, J. T., & Timp, W. (2020). Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nature Methods, 17*(12), 1191−1199. https://doi.org/10.1038/s41592-020-01000-7

Lee, Y. C. G., & Karpen, G. H. (2017). Pervasive epigenetic effects of Drosophila euchromatic transposable elements impact their evolution. *Elife, 6*. https://doi.org/10.7554/elife.25762

Leger, A., Brettell, I., Monahan, J., Barton, C., Wolf, N., Kusminski, N., Herder, C., Aadepu, N., Becker, C., Gierten, J., Hammouda, O. T., Hasel, E., Lischik, C., Lust, K., Sokolova, N., Suzuki, R., Tavhelidse, T., Thumberger, T., Tsingos, E., … Fitzgerald, T. (2022). Genomic variations and epigenomic landscape of the Medaka Inbred. *Genome Biology, 23*(1), 1−24. https://doi.org/10.1186/s13059-022-02602-4

Lemaître, J. F., Ronget, V., Tidière, M., Allainé, D., Berger, V., Cohas, A., Colchero, F., Conde, D. A., Garratt, M., Liker, A., Marais, G. A. B., Scheuerlein, A., Székely, T., & Gaillard, J. M. (2020). Sex differences in adult lifespan and aging rates of mortality across wild mammals. *Proceedings of the National Academy of Sciences of the United States of America, 117*(15), 8546−8553. https://doi.org/10.1073/pnas.1911999117

Lerat, E., Fablet, M., Modolo, L., Lopez-Maestre, H., & Vieira, C. (2017). TEtools facilitates big data expression analysis of transposable elements. *Nucleic Acids Research, 45*(4), e17. https://doi.org/10.1093/nar/gkw953

Li, D., He, M., Tang, Q., Tian, S., Zhang, J., Li, Y., Wang, D., Jin, L., Ning, C., Zhu, W., Hu, S., Long, K., Ma, J., Liu, J., Zhang, Z., & Li, M. (2022). Comparative 3D genome architecture in vertebrates. *BMC Biology, 20*(1), 1−21. https://doi.org/10.1186/s12915-022-01301-7

Li, H., Feng, X., & Chu, C. (2020). The design and construction of reference pangenome graphs with minigraph. *Genome Biology, 21*(1), 1−19. https://doi.org/10.1186/s13059-020-02168-z

Li, W., Prazak, L., Chatterjee, N., Grüninger, S., Krug, L., Theodorou, D., & Dubnau, J. (2013). Activation of transposable elements during aging and neuronal decline in. *Nature Neuroscience, 16*(5). https://doi.org/10.1038/nn.3368

Li, W., Liu, J., Zhang, H., Liu, Z., Wang, Y., Xing, L., He, Q., & Du, H. (2022). Plant pangenomics: Recent advances, new challenges, and roads ahead. *Journal of Genetics and Genomics, 49*(9), 833−846. https://doi.org/10.1016/j.jgg.2022.06.004

Li, W., Jin, Y., Prazak, L., Hammell, M., & Dubnau, J. (2012). Transposable elements in TDP-43-mediated neurodegenerative disorders. *PLoS One, 7*(9). https://doi.org/10.1371/journal.pone.0044099

Liao, W. W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J. K., Monlong, J., Abel, H. J., Buonaiuto, S., Chang, X. H., Cheng, H., Chu, J., Colonna, V., Eizenga, J. M., Feng, X., Fischer, C., Fulton, R. S., … Paten, B. (2023). A draft human pangenome reference. *Nature, 617*(7960), 312−324. https://doi.org/10.1038/s41586-023-05896-x

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Lisch, Damon (2009). Epigenetic regulation of transposable elements in plants. *Annual Review of Plant Biology, 60*, 43−66. https://doi.org/10.1146/annurev.arplant.59.032607.092744

Liu, E. Y., Russ, J., Cali, C. P., Phan, J. M., Amlie-Wolf, A., & Lee, E. B. (2019). Loss of nuclear TDP-43 is associated with decondensation of LINE retrotransposons. *Cell Reports, 27*(5), 1409−1421. https://doi.org/10.1016/j.celrep.2019.04.003

Mérel, V., Boulesteix, M., Fablet, M., & Vieira, C. (2020). Transposable elements in Drosophila. *Mobile DNA, 11*(1), 23. https://doi.org/10.1186/s13100-020-00213-z

Mahajan, S., Wei, K. H.-C., Nalley, M. J., Gibilisco, L., & Bachtrog, D. (2018). De novo assembly of a young Drosophila Y chromosome using single-molecule. *PLoS Biology, 16*(7), e2006348. https://doi.org/10.1371/journal.pbio.2006348

Marais, G. A. B., Gaillard, J. M., Vieira, C., Plotton, I., Sanlaville, D., Gueyffier, F., & Lemaitre, J. F. (2018). Sex gap in aging and longevity: Can sex chromosomes play a role? *Biology of Sex Differences, 9*, 1−14. https://doi.org/10.1186/s13293-018-0181-y

Marin, P., Genitoni, J., Barloy, D., Maury, S., Gibert, P., Ghalambor, C. K., & Vieira, C. (2020). Biological invasion: The influence of the hidden side of the (epi)genome. *Functional Ecology, 34*(2), 385−400. https://doi.org/10.1111/1365-2435.13317

Masson, P., Surosky, R., Kingsbury, J. A., & Fedoroff, N. V. (1987). Genetic and molecular analysis of the spm-dependent a-m2 alleles of the maize a locus. *Genetics, 117*(1), 117−137. https://doi.org/10.1093/genetics/117.1.117

McGurk, M. P., Dion-Côté, A.-M., & Barbash, D. A. (2021). Rapid evolution at the Drosophila telomere: Transposable element dynamics. *Genetics, 217*(2). https://doi.org/10.1093/genetics/iyaa027

McKerrow, W., Wang, X., Mendez-Dorantes, C., Mita, P., Cao, S., Grivainis, M., Ding, L., LaCava, J., Burns, K. H., Boeke, J. D., & Fenyö, D. (2022). LINE-1 expression in cancer correlates with p53 mutation, copy number alteration, and S phase checkpoint. *Proceedings of the National Academy of Sciences of the United States of America, 119*(8). https://doi.org/10.1073/pnas.2115999119

Meyer, A., Schloissnig, S., Franchini, P., Du, K., Woltering, J. M., Irisarri, I., Wong, W. Y., Nowoshilow, S., Kneitz, S., Kawaguchi, A., Fabrizius, A., Xiong, P., Dechaud, C., Spaink, H. P., Volff, J.-N., Simakov, O., Burmester, T., Tanaka, E. M., & Schartl, M. (2021). Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature, 590*(7845), 284−289. https://doi.org/10.1038/s41586-021-03198-8

Michalak, P. (2008). Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity, 102*(1), 45−50. https://doi.org/10.1038/hdy.2008.48

Mohamed, M., Dang, N. T. M., Ogyama, Y., Burlet, N., Mugat, B., Boulesteix, M., Mérel, V., Veber, P., Salces-Ortiz, J., Severac, D., Pélisson, A., Vieira, C., Sabot, F., Fablet, M., & Chambeyron, S. (2020). A transposon story: From TE content to TE dynamic invasion of Drosophila. *Cells, 9*(8). https://doi.org/10.3390/cells9081776

Morgan, Hugh D., Sutherland, Heidi G. E., Martin, David I. K., & Whitelaw, Emma (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics, 23*(3), 314−318. https://doi.org/10.1038/15490

Morgunova, V. V., Sokolova, O. A., Sizova, T. V., Malaev, L. G., Babaev, D. S., Kwon, D. A., & Kalmykova, A. I. (2022). Dysfunction of lamin B and physiological aging cause telomere instability in Drosophila germline. *Biochemistry (Moscow), 87*(12−13), 1600−1610. https://doi.org/10.1134/S000629792212015X

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Mounger, J., Ainouche, M. L., Bossdorf, O., Cavé-Radet, A., Li, B., Parepa, M., Salmon, A., Yang, J., & Richards, C. L. (2021). Epigenetics and the success of invasive plants. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 376*(1826). https://doi.org/10.1098/rstb.2020.0117

Munasinghe, M., Read, A., Stitzer, M. C., Song, B., Menard, C., Ma, K. Y., Brandvain, Y., Hirsch, C. N., & Springer, N. (2023). Combined analysis of transposable elements and structural variation in. *bioRxiv*. https://doi.org/10.1101/2023.03.02.530873

Napoletano, F., Ferrari Bravo, G., Voto, I. A. P., Santin, A., Celora, L., Campaner, E., Dezi, C., Bertossi, A., Valentino, E., Santorsola, M., Rustighi, A., Fajner, V., Maspero, E., Ansaloni, F., Cancila, V., Valenti, C. F., Santo, M., Artimagnella, O. B., Finaurini, S., … Del Sal, G. (2021). The prolyl-isomerase PIN1 is essential for nuclear Lamin-B structure and function and protects heterochromatin under mechanical stress. *Cell Reports, 36*(11). https://doi.org/10.1016/j.celrep.2021.109694

Negi, P., Rai, A. N., & Suprasanna, P. (2016). Moving through the stressed genome: Emerging regulatory roles for transposons in plant stress response. *Frontiers in Plant Science, 7*(2016). https://doi.org/10.3389/fpls.2016.01448

Nguyen, A. H., & Bachtrog, D. (2021). Toxic Y chromosome: Increased repeat expression and age-associated heterochromatin loss in male Drosophila with a young Y chromosome. *PLoS Genetics, 17*(4). https://doi.org/10.1371/journal.pgen.1009438

Nirala, N. K., Li, Q., Ghule, P. N., Chen, H. J., Li, R., Zhu, L. J., Wang, R., Rice, N. P., Mao, J., Stein, J. L., Stein, G. S., van Wijnen, A. J., & Ip, Y. T. (2021). Hinfp is a guardian of the somatic genome by repressing transposable. *Proceedings of the National Academy of Sciences of the United States of America, 118*(41). https://doi.org/10.1073/pnas.2100839118

Niu, X. M., Xu, Y. C., Li, Z. W., Bian, Y. T., Hou, X. H., Chen, J. F., Zou, Y. P., Jiang, J., Wu, Q., Ge, S., Balasubramanian, S., & Guo, Y. L. (2019). Transposable elements drive rapid phenotypic variation in Capsella rubella. *Proceedings of the National Academy of Sciences of the United States of America, 116*(14), 6908−6913. https://doi.org/10.1073/pnas.1811498116

Noshay, J. M., Marand, A. P., Anderson, S. N., Zhou, P., Mejia Guerra, M. K., Lu, Z., O'Connor, C. H., Crisp, P. A., Hirsch, C. N., Schmitz, R. J., & Springer, N. M. (2020). Assessing the regulatory potential of transposable elements using. *Genetics, 217*(1), 1−13. https://doi.org/10.1093/genetics/iyaa003

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., … Phillippy, A. M. (2022). The complete sequence of a human genome. *Science, 376*(6588), 44−53. https://doi.org/10.1126/science.abj6987

Nuzhdin, S. V., & Mackay, T. F. C. (1994). Direct determination of retrotransposon transposition rates in *Drosophila melanogaster*. *Genetical Research, 63*(2), 139−144. https://doi.org/10.1017/S0016672300032249

Ochoa, E., Ramirez, P., Gonzalez, E., De Mange, J., Ray, W. J., Bieniek, K. F., & Frost, B. (2023). Pathogenic tau−induced transposable element−derived dsRNA drives neuroinflammation. *Science Advances, 9*(1). https://doi.org/10.1126/sciadv.abq5423

Orozco-Arias, S., Isaza, G., Guyot, R., & Tabares-Soto, R. (2019). A systematic review of the application of machine learning in the. *PeerJ, 7*, e8311. https://doi.org/10.7717/peerj.8311

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Panda, K., Ji, L., Neumann, D. A., Daron, J., Schmitz, R. J., & Slotkin, R. K. (2016). Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation. *Genome Biology, 17*(1). https://doi.org/10.1186/s13059-016-1032-y

Parisot, N., Vargas-Chávez, C., Goubert, C., Baa-Puyoulet, P., Balmand, S., Beranger, L., Blanc, C., Bonnamour, A., Boulesteix, M., Burlet, N., Calevro, F., Callaerts, P., Chancy, T., Charles, H., Colella, S., Da Silva Barbosa, A., Dell'Aglio, E., Di Genova, A., Febvay, G., … Heddi, A. (2021). The transposable element-rich genome of the cereal pest Sitophilus oryzae. *BMC Biology, 19*(1), 241. https://doi.org/10.1186/s12915-021-01158-2

Payer, L. M., & Burns, K. H. (2019). Transposable elements in human genetic disease. *Nature Reviews Genetics, 20*(12), 760−772. https://doi.org/10.1038/s41576-019-0165-8

Petrov, D. A., Schutzman, J. L., Hartl, D. L., & Lozovskaya, E. R. (1995). Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. *Proceedings of the National Academy of Sciences of the United States of America, 92*(17), 8050−8054. https://doi.org/10.1073/pnas.92.17.8050

Pignatta, D., Novitzky, K., Satyaki, P. R. V., & Gehring, M. (2018). A variably imprinted epiallele impacts seed development. *PLoS Genetics, 14*(11), e1007469. https://doi.org/10.1371/journal.pgen.1007469

Qiu, Y., & Köhler, C. (2020). Mobility connects: Transposable elements wire new transcriptional networks by transferring transcription factor binding motifs. *Biochemical Society Transactions, 48*(3), 1005−1017. https://doi.org/10.1042/bst20190937

Ramirez, P., Zuniga, G., Sun, W., Beckmann, A., Ochoa, E., DeVos, S. L., Hyman, B., Chiu, G., Roy, E. R., Cao, W., Orr, M., Buggia-Prevot, V., Ray, W. J., & Frost, B. (2022). Pathogenic tau accelerates aging-associated activation of transposable. *Progress in Neurobiology, 208*. https://doi.org/10.1016/j.pneurobio.2021.102181

Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., & Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods, 14*(4), 411−413. https://doi.org/10.1038/nmeth.4189

Rebollo, R., Horard, B., Hubert, B., & Vieira, C. (2010). Jumping genes and epigenetics: Towards new species. *Gene, 454*(1−2), 1−7. https://doi.org/10.1016/j.gene.2010.01.003

Rebollo, R., Karimi, M. M., Bilenky, M., Gagnier, L., Miceli-Royer, K., Zhang, Y., Goyal, P., Keane, T. M., Jones, S., Hirst, M., Lorincz, M. C., & Mager, D. L. (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genetics, 7*(9). https://doi.org/10.1371/journal.pgen.1002301

Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M. S., Hruban, R. H., Iacobuzio-Donahue, C. A., Maitra, A., Torbenson, M. S., Goggins, M., Shih, I. M., Duffield, A. S., Montgomery, E. A., Gabrielson, E., Netto, G. J., Lotan, T. L., De Marzo, A. M., Westra, W., … Burns, K. H. (2014). Long interspersed element-1 protein expression is a hallmark of many human cancers. *American Journal of Pathology, 184*(5), 1280−1286. https://doi.org/10.1016/j.ajpath.2014.01.007

Rodriguez-Martin, B., Alvarez, E. G., Baez-Ortega, A., Zamora, J., Supek, F., Demeulemeester, J., Santamarina, M., Ju, Y. S., Temes, J., Garcia-Souto, D., Detering, H., Li, Y., Rodriguez-Castro, J., Dueso-Barroso, A., Bruzos, A. L., Dentro, S. C., Blanco, M. G., Contino, G., Ardeljan, D., … Rodriguez-Martin, B. (2020). Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nature Genetics, 52*(3), 306−319. https://doi.org/10.1038/s41588-019-0562-0

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Rollins, R. A., Haghighi, F., Edwards, J. R., Das, R., Zhang, M. Q., Ju, J., & Bestor, T. H. (2006). Large-scale structure of genomic methylation patterns. *Genome Research, 16*(2), 157−163. https://doi.org/10.1101/gr.4362006

Romano, G., Klima, R., & Feiguin, F. (2020). TDP-43 prevents retrotransposon activation in the Drosophila motor system through regulation of Dicer-2 activity. *BMC Biology, 18*(1). https://doi.org/10.1186/s12915-020-00816-1

Romero-Soriano, V., Modolo, L., Lopez-Maestre, H., Mugat, B., Pessia, E., Chambeyron, S., Vieira, C., & Garcia Guerreiro, M. P. (2017). Transposable element misregulation is linked to the divergence between. *Genome Biology and Evolution, 9*(6), 1450−1470. https://doi.org/10.1093/gbe/evx091

Ruggieri, A. A., Livraghi, L., Lewis, J. J., Evans, E., Cicconardi, F., Hebberecht, L., Ortiz-Ruiz, Y., Montgomery, S. H., Ghezzi, A., Rodriguez-Martinez, J. A., Jiggins, C. D., McMillan, W. O., Counterman, B. A., Papa, R., & Van Belleghem, S. M. (2022). A butterfly pan-genome reveals that a large amount of structural variation underlies the evolution of chromatin accessibility. *Genome Research, 32*(10), 1862−1875. https://doi.org/10.1101/gr.276839.122

Schlichting, C. D., & Smith, H. (2002). Phenotypic plasticity: Linking molecular mechanisms with evolutionary outcomes. *Evolutionary Ecology, 16*(3), 189−211. https://doi.org/10.1023/A:1019624425971

Schneider, B. K., Sun, S., Lee, M., Li, W., Skvir, N., Neretti, N., Vijg, J., & Secombe, J. (2023). Expression of retrotransposons contributes to aging in Drosophila. *Genetics, 224*(2). https://doi.org/10.1093/genetics/iyad073

Shah, N. M., Jang, H. J., Liang, Y., Maeng, J. H., Tzeng, S. C., Wu, A., Basri, N. L., Qu, X., Fan, C., Li, A., Katz, B., Li, D., Xing, X., Evans, B. S., & Wang, T. (2023). Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements. *Nature Genetics, 55*(4), 631−639. https://doi.org/10.1038/s41588-023-01349-3

Sibbesen, J. A., Eizenga, J. M., Novak, A. M., Sirén, J., Chang, X., Garrison, E., & Paten, B. (2023). Haplotype-aware pantranscriptome analyses using spliced pangenome graphs. *Nature Methods, 20*(2). https://doi.org/10.1038/s41592-022-01731-9

Simon, M., Van Meter, M., Ablaeva, J., Ke, Z., Gonzalez, R. S., Taguchi, T., De Cecco, M., Leonova, K. I., Kogan, V., Helfand, S. L., Neretti, N., Roichman, A., Cohen, H. Y., Meer, M. V., Gladyshev, V. N., Antoch, M. P., Gudkov, A. V., Sedivy, J. M., Seluanov, A., & Gorbunova, V. (2019). LINE1 derepression in aged wild-type and SIRT6-deficient mice drives. *Cell Metabolism, 29*(4). https://doi.org/10.1016/j.cmet.2019.02.014

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods, 14*(4), 407−410. https://doi.org/10.1038/nmeth.4184

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA. *Elife, 6*. https://doi.org/10.7554/eLife.21856

Slotkin, R Keith, & Martienssen, Robert (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics, 8*(4), 272−285. https://doi.org/10.1038/nrg2072

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., & Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature, 484*(7394), 339−344. https://doi.org/10.1038/nature10960

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Statham, A. L., Robinson, M. D., Song, J. Z., Coolen, M. W., Stirzaker, C., & Clark, S. J. (2012). Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq). *Genome Research, 22*(6), 1120−1127. https://doi.org/10.1101/gr.132076.111

Stitzer, M. C., Anderson, S. N., Springer, N. M., & Ross-Ibarra, J. (2021). The genomic ecosystem of transposable elements in maize. *PLoS Genetics, 17*(10), e1009768. https://doi.org/10.1371/journal.pgen.1009768

Stow, E. C., Kaul, T., deHaro, D. L., Dem, M. R., Beletsky, A. G., Morales, M. E., Du, Q., LaRosa, A. J., Yang, H., Smither, E., Baddoo, M., Ungerleider, N., Deininger, P., & Belancio, V. P. (2021). Organ-, sex- and age-dependent patterns of endogenous L1 mRNA expression. *Nucleic Acids Research, 49*(10). https://doi.org/10.1093/nar/gkab369

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M. H. Y., Konkel, M. K., Malhotra, A., Stütz, A. M., Shi, X., Casale, F. P., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., … Korbel, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature, 526*(7571), 75−81. https://doi.org/10.1038/nature15394

Sun, W., Samimi, H., Gamez, M., Zare, H., & Frost, B. (2018). Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. *Nature Neuroscience, 21*(8), 1038−1048. https://doi.org/10.1038/s41593-018-0194-1

Sun, X., Xiang, Y., Dou, N., Zhang, H., Pei, S., Franco, A. V., Menon, M., Monier, B., Ferebee, T., Liu, T., Liu, S., Gao, Y., Wang, J., Terzaghi, W., Yan, J., Hearne, S., Li, L., Li, F., & Dai, M. (2023). The role of transposon inverted repeats in balancing drought tolerance and yield-related traits in maize. *Nature Biotechnology, 41*(1), 120−127. https://doi.org/10.1038/s41587-022-01470-4

Sundaram, V., & Wysocka, J. (2020). Transposable elements as a potent source of diverse cis-regulatory. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 375*(1795), 20190347. https://doi.org/10.1098/rstb.2019.0347

Sur, D., Kustwar, R. K., Budania, S., Mahadevan, A., Hancks, D. C., Yadav, V., Shankar, S. K., & Mandal, P. K. (2017). Detection of the LINE-1 retrotransposon RNA-binding protein ORF1p in. *Mobile DNA, 8*. https://doi.org/10.1186/s13100-017-0101-4

Szabo, Q., Jost, D., Chang, J. M., Cattoni, D. I., Papadopoulos, G. L., Bonev, B., Sexton, T., Gurgo, J., Jacquier, C., Nollmann, M., Bantignies, F., & Cavalli, G. (2018). TADs are 3D structural units of higher-order chromosome organization in Drosophila. *Science Advances, 4*(2). https://doi.org/10.1126/sciadv.aar8082

Taube, J. H., Allton, K., Duncan, S. A., Shen, L., & Barton, M. C. (2010). Foxa1 functions as a pioneer transcription factor at transposable elements to activate Afp during differentiation of embryonic stem cells. *Journal of Biological Chemistry, 285*(21), 16135−16144. https://doi.org/10.1074/jbc.m109.088096

Tubio, J. M. C., Li, Y., Ju, Y. S., Martincorena, I., Cooke, S. L., Tojo, M., Gundem, G., Pipinikas, C. P., Zamora, J., Raine, K., Menzies, A., Roman-Garcia, P., Fullam, A., Gerstung, M., Shlien, A., Tarpey, P. S., Papaemmanuil, E., Knappskog, S., Van Loo, P., … Campbell, P. J. (2014). Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science, 345*(6196). https://doi.org/10.1126/science.1251343

Van Meter, M., Kashyap, M., Rezazadeh, S., Geneva, A. J., Morello, T. D., Seluanov, A., & Gorbunova, V. (2014). SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this. *Nature Communications, 5*, 5011. https://doi.org/10.1038/ncomms6011

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Vela, D., Fontdevila, A., Vieira, C., & García Guerreiro, M. P. (2014). A genome-wide survey of genetic instability by transposition in Drosophila. *PLoS One, 9*(2), e88992. https://doi.org/10.1371/journal.pone.0088992

Venuto, D., & Bourque, G. (2018). Identifying co-opted transposable elements using comparative epigenomics. *Development Growth and Differentiation, 60*(1), 53−62. https://doi.org/10.1111/dgd.12423

Vieira, C., & Biémont, C. (1997). Transposition rate of the 412 retrotransposable element is independent of copy number in natural populations of *Drosophila simulans*. *Molecular Biology and Evolution, 14*(2), 185−188. https://doi.org/10.1093/oxfordjournals.molbev.a025751

Villanueva-Cañas, J. L., Horvath, V., Aguilera, L., & González, J. (2019). Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*. *Nucleic Acids Research, 47*(13). https://doi.org/10.1093/nar/gkz490

Walsh, C. P., Chaillet, J. R., & Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine. *Nature Genetics, 20*(2), 116−117. https://doi.org/10.1038/2413

Wang, J., Vicente-García, C., Seruggia, D., Moltó, E., Fernandez-Miñán, A., Neto, A., Lee, E., Gómez-Skarmeta, J. L., Montoliu, L., Lunyak, V. V., & King Jordan, I. (2015). MIR retrotransposon sequences provide insulators to the human genome. *Proceedings of the National Academy of Sciences of the United States of America, 112*(32), E4428. https://doi.org/10.1073/pnas.1507253112

Wang, M., Li, J., Wang, P., Liu, F., Liu, Z., Zhao, G., Xu, Z., Pei, L., Grover, C. E., Wendel, J. F., Wang, K., & Zhang, X. (2021). Comparative genome analyses highlight transposon-mediated genome expansion. *Molecular Biology and Evolution, 38*(9), 3621−3636. https://doi.org/10.1093/molbev/msab128

Wang, Z., & Baulcombe, D. C. (2020). Transposon age and non-CG methylation. *Nature Communications, 11*(1), 1−9. https://doi.org/10.1038/s41467-020-14995-6

Warmuth, V. M., Weissensteiner, M. H., & Wolf, J. B. W. (2022). Accumulation and ineffective silencing of transposable elements on an avian W Chromosome. *Genome Research, 32*(4), 671−681. https://doi.org/10.1101/gr.275465.121

Waugh O'Neill, R. J., O'Neill, M. J., & Marshall Graves, J. A. (1998). Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature, 393*(6680), 68−72. https://doi.org/10.1038/29985

Wei, K., Aldaimalani, R., Mai, D., Zinshteyn, D., Prv, S., Blumenstiel, J. P., Kelleher, E. S., & Brooks, E. (2022). *Rethinking the "gypsy" retrotransposon: A roadmap for community-driven*. https://doi.org/10.31219/osf.io/fma57

Wells, J. N., & Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annual Review of Genetics, 54*, 539−561. https://doi.org/10.1146/annurev-genet-040620-022145

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics, 8*(12), 973−982. https://doi.org/10.1038/nrg2165

Winter, D. J., Ganley, A. R. D., Young, C. A., Liachko, I., Schardl, C. L., Dupont, P.-Y., Berry, D., Ram, A., Scott, B., & Cox, M. P. (2018). Repeat elements organise 3D genome structure and mediate transcription in. *PLoS Genetics, 14*(10), e1007467. https://doi.org/10.1371/journal.pgen.1007467

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Wood, J. G., Jones, B. C., Jiang, N., Chang, C., Hosier, S., Wickremesinghe, P., Garcia, M., Hartnett, D. A., Burhenn, L., Neretti, N., & Helfand, S. L. (2016). Chromatin-modifying genetic interventions suppress age-associated. *Proceedings of the National Academy of Sciences of the United States of America, 113*(40). https://doi.org/10.1073/pnas.1604621113

Yang, H., Luan, Y., Liu, T., Lee, H. J., Fang, L., Wang, Y., Wang, X., Zhang, B., Jin, Q., Ang, K. C., Xing, X., Wang, J., Xu, J., Song, F., Sriranga, I., Khunsriraksakul, C., Salameh, T., Li, D., Choudhary, M. N. K., … Yue, F. (2020). A map of cis-regulatory elements and 3D genome structures in zebrafish. *Nature, 588*(7837), 337−343. https://doi.org/10.1038/s41586-020-2962-9

Yang, N., Srivastav, S. P., Rahman, R., Ma, Q., Dayama, G., Li, S., Chinen, M., Lei, E. P., Rosbash, M., & Lau, N. C. (2022). Transposable element landscapes in aging Drosophila. *PLoS Genetics, 18*(3). https://doi.org/10.1371/journal.pgen.1010024

Yang, W. R., Ardeljan, D. L., Pacyna, C. N., Payer, L. M., & Burns, K. H. (2019). SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Research, 47*(5), e27. https://doi.org/10.1093/nar/gky1301

Ye, M., Goudot, C., Hoyler, T., Lemoine, B., Amigorena, S., & Zueva, E. (2020). Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. *Proceedings of the National Academy of Sciences of the United States of America, 117*(14), 7905−7916. https://doi.org/10.1073/pnas.1912008117

Zhang, Y. J., Guo, L., Gonzales, P. K., Gendron, T. F., Wu, Y., Jansen-West, K., O'Raw, A. D., Pickles, S. R., Prudencio, M., Carlomagno, Y., Gachechiladze, M. A., Ludwig, C., Tian, R., Chew, J., DeTure, M., Lin, W. L., Tong, J., Daughrity, L. M., Yue, M., … Petrucelli, L. (2019). Heterochromatin anomalies and double-stranded RNA accumulation underlie C9orf72 poly(PR) toxicity. *Science, 363*(6428). https://doi.org/10.1126/science.aav2606

Zhou, Y., Yang, L., Han, X., Han, J., Hu, Y., Li, F., Xia, H., Peng, L., Boschiero, C., Rosen, B. D., Bickhart, D. M., Zhang, S., Guo, A., Van Tassell, C. P., Smith, T. P. L., Yang, L., & Liu, G. E. (2022). Assembly of a pangenome for global cattle reveals missing sequences and novel structural variations, providing new insights into their diversity and evolutionary history. *Genome Research, 32*, 1585−1601. https://doi.org/10.1101/gr.276550.122

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous d

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Annexe 2

9764–9784 *Nucleic Acids Research, 2023, Vol. 51, No. 18* *Published online 24 August 2023*
https://doi.org/10.1093/nar/gkad671

ChimeraTE: a pipeline to detect chimeric transcripts derived from genes and transposable elements

Daniel S. Oliveira[1,2], Marie Fablet[2,3], Anaïs Larue[2,4], Agnès Vallier[4], Claudia M.A. Carareto[1], Rita Rebollo[4,*] and Cristina Vieira[2,*]

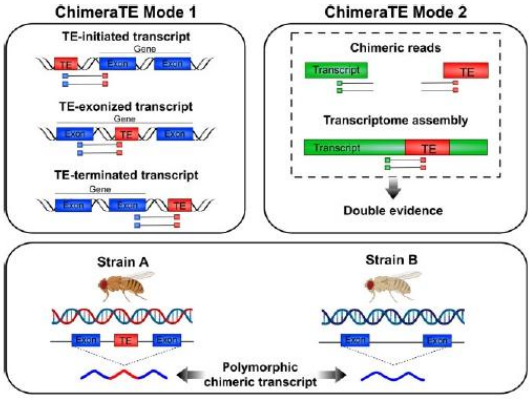
[1]São Paulo State University (Unesp), Institute of Biosciences, Humanities and Exact Sciences, São José do Rio Preto, SP, Brazil, [2]Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, CNRS, UMR5558, Villeurbanne, Rhone-Alpes, 69100, France, [3]Institut Universitaire de France (IUF), Paris, Île-de-FranceF-75231, France and [4]Univ Lyon, INRAE, INSA-Lyon, BF2I, UMR 203, 69621 Villeurbanne, France

Received November 03, 2022; Revised July 25, 2023; Editorial Decision July 27, 2023; Accepted August 09, 2023

ABSTRACT

Transposable elements (TEs) produce structural variants and are considered an important source of genetic diversity. Notably, TE-gene fusion transcripts, i.e. chimeric transcripts, have been associated with adaptation in several species. However, the identification of these chimeras remains hindered due to the lack of detection tools at a transcriptome-wide scale, and to the reliance on a reference genome, even though different individuals/cells/strains have different TE insertions. Therefore, we developed ChimeraTE, a pipeline that uses paired-end RNA-seq reads to identify chimeric transcripts through two different modes. Mode 1 is the reference-guided approach that employs canonical genome alignment, and Mode 2 identifies chimeras derived from fixed or insertionally polymorphic TEs without any reference genome. We have validated both modes using RNA-seq data from four *Drosophila melanogaster* wild-type strains. We found ∼1.12% of all genes generating chimeric transcripts, most of them from TE-exonized sequences. Approximately ∼23% of all detected chimeras were absent from the reference genome, indicating that TEs belonging to chimeric transcripts may be recent, polymorphic insertions. ChimeraTE is the first pipeline able to automatically uncover chimeric transcripts without a reference genome, consisting of two running Modes that can be used as a tool to investigate the contribution of TEs to transcriptome plasticity.

GRAPHICAL ABSTRACT



INTRODUCTION

Transposable elements (TEs) are mobile DNA sequences that comprise a large fraction of eukaryotic genomes, from 15% in *Drosophila melanogaster* (1), 45% in humans (2), to 85% in maize (3). Many TE copies have lost their ability to transpose as a result of accumulated mutations and recombination throughout evolution (4). Despite their lack of mobility, such ancient TE insertions may still harbor functional protein domains, alternative splice sites, and *cis*-acting regulatory sequences, as transcription factor binding sites (TFBSs) and polyadenylation (PolyA) sites. Therefore, TEs are a major source of genetic diversity, not only due to their mobilization, but also because they donate protein domains to genes (5–8), and regulatory sequences that modify the expression of nearby genes (9–13). The participation of TE-derived sequences in the host biology is a process called domestication or exaptation (14). The ancestral role of the TE sequence can be domesticated into an essential host

*To whom correspondence should be addressed. Tel: +33 4 72 44 82 16; Email: cristina.vieira@univ-lyon1.fr
Correspondence may also be addressed to Rita Rebollo. Tel: +33 4 72 43 62 46; Email: rita.rebollo@inrae.fr

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

function, but it can also be modified, and adapted, into a new exapted function that may also be essential for the host species (14).

Chimeric transcripts are RNAs stemming from two sequences of different origins (15). Hereafter we define chimeric transcripts as mature transcripts that have both gene and TE-derived sequences. These transcripts can be divided into three types: (1) TE-initiated transcripts: chimeric transcripts with a TE transcription start site (TSS) (16,17); (2) TE-exonized transcripts: TE sequences are incorporated into the transcript either partially or as full-length exons (18–20); and (3) TE-terminated transcripts: chimeric transcripts with a TE transcription termination site (21,22). TE-initiated and TE-terminated transcripts might modulate gene expression levels either by the presence of TFBSs, PolyA sites, or chromatin changes; while TE-exonized transcripts may alter the protein sequence of coding genes and have a direct effect on the protein function. Regardless of the TE position, such events of TE exaptation and domestication have been associated with many biological roles and are widespread among eukaryotic species (14). In *D. melanogaster*, the *CHKov1* gene produces a chimeric transcript with a truncated mRNA resulting in resistance to insecticide and viral infection (23). In humans, the *SETMAR* gene produces a chimeric transcript containing a *Hsmar1* copy, involved in non-homologous end-joining DNA repair (24). In cancer, TEs become active due to a global hypomethylation state (25) and such activation may generate new chimeric transcripts with detrimental outcomes (26), a process called onco-exaptation (9). For example, in large B-cell lymphoma, the *FABP7* gene has an endogenous retrovirus LTR co-opted as a promoter, generating a novel protein involved in abnormal cell proliferation (27). Recently, such phenomenon has been observed in a larger scale. A pan-cancer study revealed 1,068 tumor-specific TE-initiated transcripts from genes coding antigens, showing the high prevalence of TE exaptation in cancer (28). Therefore, chimeric transcripts have a large impact on host biology, but their study remains hindered by the ubiquitous repetitive nature of TE copies, and lack of methods to identify chimeric transcripts.

Previous studies with Cap Analysis Gene Expression (CAGE) revealed a significant percentage of genes producing TE-initiated transcripts, ranging from 3–14% in humans and mice, depending on the tissue (29). More specifically, in human pluripotent stem cells, chimeric transcripts comprise 26% of coding and 65% of noncoding transcripts (30). In *D. melanogaster*, a study using expressed sequence tags (ESTs) showed that the proportion of genes with chimeric transcripts is only ∼1% (31). Another study used an approach based on cap-enriched mRNA sequencing but with a preliminary transcript elongation step allowing not only to detect TSSs but also to estimate mRNA expression rates (RAMPAGE) (32). This study detected TE-initiated chimeric transcripts in 36 stages of *D. melanogaster* life cycle, and observed that up to 1.6% of all transcripts were chimeras, representing up to ∼1% of all genes (33). More recently, a tissue-specific study has shown that 264 genes produce chimeric transcripts in the midbrain of *D. melanogaster*, corresponding to ∼1.5% of all genes (34).

Several bioinformatic methods have been developed to take advantage of RNA sequencing (RNA-seq) to identify chimeric transcripts, such as CLIFinder (35) and LIONS (36). The former is designed to identify chimeric transcripts derived from long interspersed nuclear elements (LINEs) in the human genome, whereas the latter identifies only TE-initiated transcripts. Both methods need a reference genome and they only detect chimeric transcripts derived from TE insertions present in the reference genome. Therefore, it is impossible to identify chimeric transcripts derived from polymorphic TE insertions that may exist in other populations, strains, or individuals. Finally, the latest addition to chimeric transcript detection, TEchim (34), can detect chimeras with TEs that are polymorphic and absent from the reference genome of *D. melanogaster*, but it is not a pipeline designed to run automatically with any other genome.

In this study, we have developed ChimeraTE, a pipeline that uses paired-end RNA-seq reads to identify chimeric transcripts. The pipeline has two Modes: Mode 1 can predict chimeric transcripts through genome alignment, whereas Mode 2 performs chimeric transcript searches without a reference genome, being able to identify chimeras derived from fixed or polymorphic TE insertions. To benchmark the pipeline, we have used RNA-seq from ovaries of four *D. melanogaster* wild-type strains, for which we have assembled and annotated genomes. We found that ∼1.12% of genes have chimeric transcripts in the ovarian transcriptome, of which ∼88.97% are TE-exonized transcripts. Our results also reveal that the retrotransposon *roo* is the most frequent exonized TE family. In addition, with Mode 2, we found 11 polymorphic chimeric transcripts deriving from TE insertions that are absent from the *D. melanogaster* reference genome. Therefore, this work provides a new strategy to identify chimeric transcripts with or without the reference genome, in a transcriptome-wide manner.

MATERIALS AND METHODS

ChimeraTE: the pipeline

ChimeraTE was developed to detect chimeric transcripts with paired-end RNA-seq reads. It is developed in *python3 v.3.6.15* language, and it is able to fully automate the process in only one command line. The pipeline has two detection Modes: (1) genome-guided, the reference genome is provided and chimeric transcripts are detected aligning reads against it; and (2) genome-blind, the reference genome is not provided and chimeric transcripts are predicted for fixed or polymorphic TEs. These Modes have distinct approaches that may be used for different purposes. In Mode 1, chimeric transcripts will be detected considering the genomic location of TE insertions and exons. Chimeras from this Mode can be classified as TE-initiated transcripts (TE located upstream of the gene), TE-exonized transcripts (TE within introns or embedded in gene exons), and TE-terminated transcripts (TE located downstream of the gene). In addition, results from Mode 1 can be visualized in a genome browser, which allows a manual curation of chimeric transcripts in the reference genome. Mode 1 does not detect chimeric transcripts derived from TE insertions absent from the provided reference genome. Mode 2 predicts chimeric

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

transcripts considering the alignment of reads against transcripts and TE insertions, in addition to a transcriptome assembly (user optional). Hence, Mode 2 detects chimeric transcripts from *de novo* TE insertions and an assembled genome is not necessary. In this Mode, two alignments are performed: (i) transcript alignment and (ii) TE alignment. Then, based on both alignments, the pipeline identifies chimeric reads that support chimeric transcripts, regardless of the TE genomic location. In Mode 2, since there is no alignment against an annotated genome, it is not possible to classify chimeric transcripts considering the TE position as in Mode 1.

Both ChimeraTE Modes use chimeric reads, which are defined by paired-end reads that contain both TE and exon sequences, as evidence of chimeric transcripts. This method has been widely demonstrated by other authors as a potential source of artifactual reads, mainly due to the occurrence of mixed clusters (*index hopping*) on the Illumina's flow cell that may be too close to each other, as well as jumping PCR (*in vitro* crossover artifact), generating read pairs that are connecting two cDNA portions that are not joined in the sample (37–40). Indeed, it has been shown that up to 1.56% of all reads produced by Illumina multiplexed approaches might generate chimeric reads (41), including cases that may support chimeric transcripts derived from different genes. These artifactual reads originate more likely from highly expressed genes since there are more molecules on the Illumina's cell. Conversely, because TE-derived sequences might comprise a low proportion of the transcriptome, artifactual reads from TEs should be produced at a low frequency. Furthermore, it is unlikely that artifactual reads from the same gene and TE family among RNA-seq replicates would be produced. Nonetheless, to avoid including false chimeric reads, both Modes of ChimeraTE only call chimeric transcripts that are detected in at least two RNA-seq replicates (user optional for more).

ChimeraTE mode 1: genome-guided approach

In ChimeraTE Mode 1, paired-end RNA-seq reads, a whole genome assembly, and its respective gene/TE annotation are used to predict chimeric transcripts (Figure 1A). We highlight the need for robust gene and TE annotations to submit an analysis to Mode 1, since each TE insertion and exon will be used to identify chimeric reads. Firstly, RNA-seq alignment to the genome is performed with *STAR v.2.7.6a* with default parameters (Figure 1B), and transcript expression is assessed with *Cufflinks v.2.2.1* (42) for each RNA-seq replicate. We consider FPKM $\geq$1 as an expressed gene by default, but it can be changed by the user with the *–fpkm* parameter. Only concordant reads (both reads from the pair are aligned) with unique alignment (*samtools -q 255*) are selected and converted to BED format with *samtools v.1.10* (43) and *bedtools v.2.30.0* (44), implemented in *python3* with *pybedtools v.0.9.0* (45). Split reads are also considered as chimeric reads when one mate of the pair stems from the TE to the exon, or vice-versa (Supplementary Figure 1). The alignment position of these reads are considered independently with *bedtools bamtobed -split*. Finally, reads aligned into the forward and reverse strands are separated with *samtools* (43).

The IDs from reads that have aligned against exons are identified with *bedtools intersect* (Figure 1B). Then, reads with at least 50% of their length (*–overlap 0.50*) aligned against reference TE copies have their IDs selected (Figure 1B), and TE copies without aligned reads are removed from the downstream analysis. These two lists of IDs are intersected between each other and reads that have aligned to both exons and TEs are identified, generating a raw list of chimeric reads (Figure 1C). TE-initiated transcripts are the first to be searched by ChimeraTE Mode 1. Expressed genes that have TE copies within their 3 kb upstream region (default but adjustable with *–window* parameter) are identified with *bedtools* (44). For each gene, the pipeline will check whether the TE located upstream has chimeric reads shared with the gene exons (Figure 1D). If there are no chimeric reads, both gene and TE are discarded. The same method is applied to identify TE-terminated transcripts, but with genes harbouring downstream TEs up to 3 kb (user adjustable value) (Figure 11D). Finally, TE-exonized transcripts are also identified thanks to chimeric reads aligned to exons and TEs located within genes. Based on the TE position towards exons, they are classified as embedded, intronic and overlapped. Embedded TEs are located entirely within an exon. In these cases, reads aligned only to the TE sequence are not considered chimeric reads (Supplementary Figure 1). Indeed, the exon can be artificially divided into two portions, and chimeric reads must have a read (or a split read) aligned to the TE, whereas its mate (or the extension from the split read) aligned to at least one exon portion (Supplementary Figure 1). This method avoids autonomous TE expression as evidence of chimeric transcript expression. The same is performed to identify TE-exonized transcripts derived from TEs that are partially overlapping exons but are also located in introns (Supplementary Figure 1). Finally, TE-exonized transcripts with TEs located in introns are identified with *bedtools intersect*, and chimeric reads are detected to support chimeric transcripts.

These steps are repeated for all RNA-seq replicates provided in the input. Then, the raw results from replicates are compared, and all chimeric transcripts that were found in $\geq$2 replicates and have been identified with at least $\geq$2 chimeric reads on average between replicates are considered true chimeric transcripts (Figure 1C). These thresholds may be changed by the user with *–cutoff* and *–replicate* parameters. Mode 1 output is a table with a list of predicted chimeric transcripts categorized by TE position, with gene ID, TE family, chimeric read coverage, TE location, gene location, and gene expression (Figure 1D).

ChimeraTE mode 2: a genome-blind approach to uncover chimeric transcripts

ChimeraTE Mode 2 is the genome-blind approach of the pipeline. The input data are stranded RNA-seq reads, gene transcripts, and reference TE insertions. TE consensus can be used, but it will decrease the alignment rate, and therefore it should be used only when a fasta file with TE insertions is not available (Figure 2A). The data will be used to perform two alignments with *bowtie2 v.2.4.5* (46), one against all transcripts and another against all TE sequences, both with parameters: *-D 20 -R 3 -N 1 -L 20 -i S,1,0.50* (Figure 2A). To

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

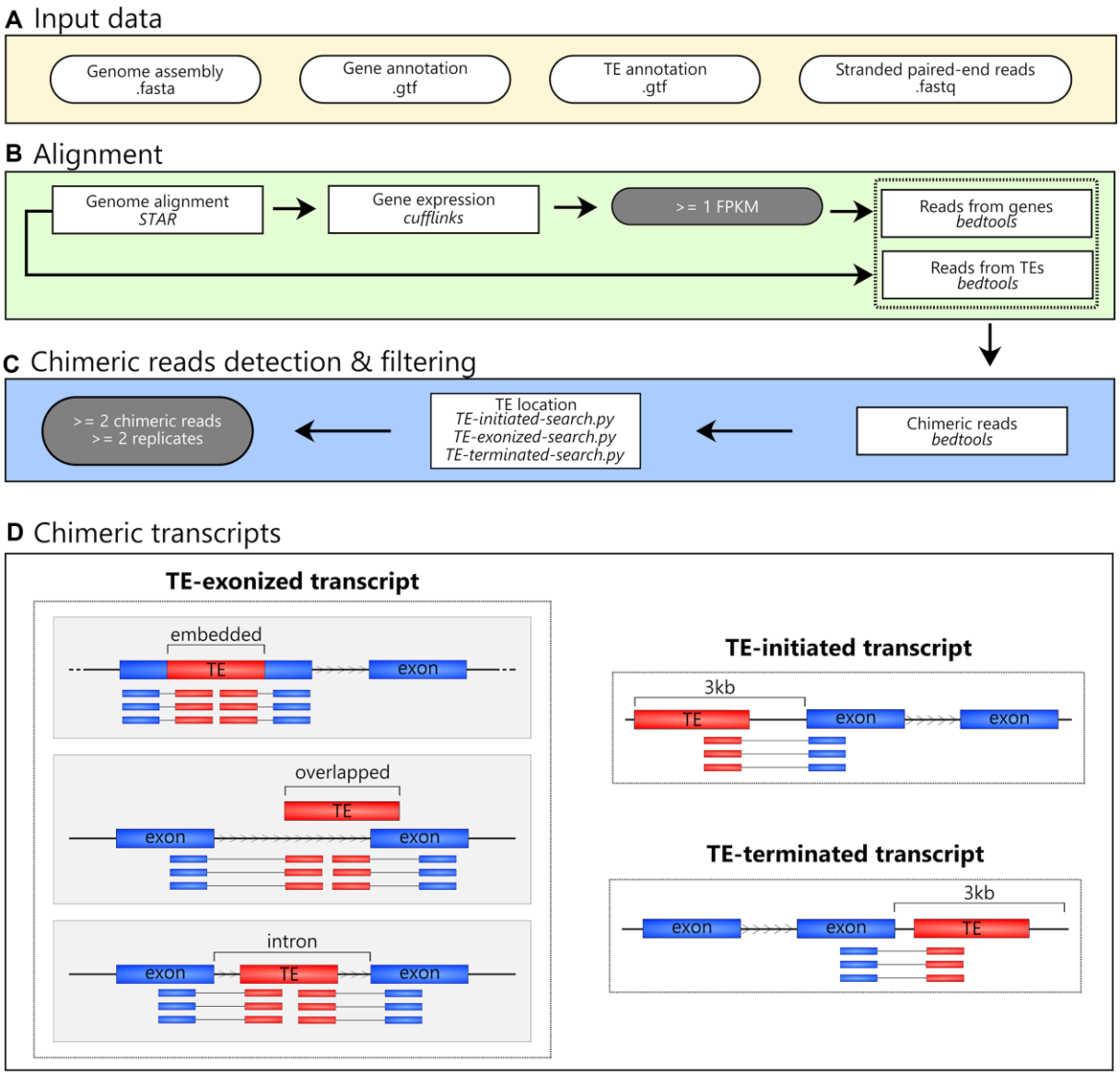


Figure 1. ChimeraTE Mode 1 (genome-guided) workflow. Round white boxes: input data; square boxes: pipeline step; round gray boxes: thresholds that can be modified. (**A**) Input data: fasta file with the genome assembly, gtf files with gene and TE annotations, as well as stranded paired-end reads from RNA-seq (fastq). (**B**) Alignment: RNA-seq alignment to the genome is used to calculate gene expression levels. Genes with FPKM < 1 are removed from downstream analyses. A subsequent list of reads that have aligned against genes or TE insertions is created. (**C**) Chimeric read detection & filtering: both read lists are then compared and read pairs that have common reads between the two lists are named chimeric reads, *i.e.* paired-end reads mapping to a gene and a TE copy. The average of these reads between replicates is used as chimeric read coverage for each putative chimeric transcript. All putative chimeras are then processed with three ChimeraTE scripts to categorize them into TE-initiated, TE-exonized, and TE-terminated transcripts. These steps are performed for all RNA-seq replicates. Finally, all chimeric transcripts present in at least 2 replicates and with at least 2 chimeric reads on average between replicates are maintained. (**D**) Chimeric transcripts: Three predictions obtained from Mode 1. Blue boxes: exons; red boxes: TEs; arrowhead in between TE and exon boxes: transcription sense; blue and red boxes linked by a line: chimeric reads. The ChimeraTE mode 1 output is divided into three predictions: (i) TE-initiated transcript: the TE insertion is located upstream of the gene region; (ii) TE-exonized transcript: the TE insertion is present within exons (embedded), overlapping exons (overlapped), or introns (intronic); (iii) TE-terminated transcript: the TE insertion is located downstream of the gene region.

avoid very low-expressed transcripts predicted as chimeras and to ensure reasonable processing time, the SAM alignment is converted to BAM with *samtools v.1.10* (43), and FPKMs are computed for the reference transcripts provided in the input using *eXpress v.1.5.1* (47). Then, all genes with average FPKM <1 are removed from the downstream analysis (Figure 2B). To identify chimeric reads between TEs and gene transcripts, both alignments are converted to BED with *bedtools v.2.30.0* (44). Among all aligned paired-end reads, the pipeline considers as chimeric transcripts the ones that have at least one read aligned to the TE sequence (singleton mapping) and its mate to the gene transcript, or when both reads have aligned (concordant mapping) to the TE and gene transcript. In order to identify these reads, the TE alignment output is used to create a list with all read 1 IDs that have aligned against TEs, and another list with all

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

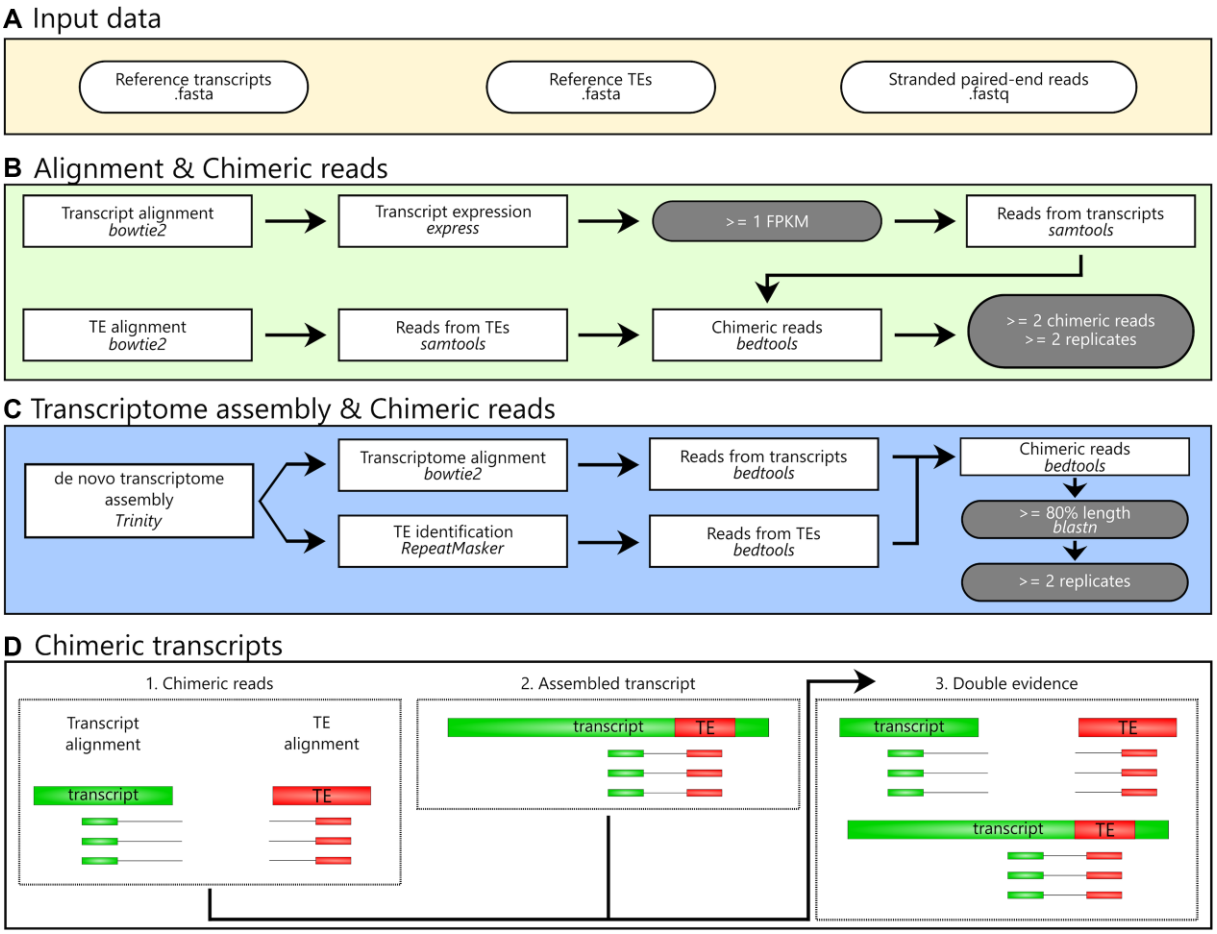


Figure 2. ChimeraTE Mode 2 (genome-blind)workflow. Round white boxes: input data; square boxes: pipeline step; round gray boxes: thresholds that can be modified. (**A**) Input data: two fasta files containing reference transcripts and TE insertions, as well as stranded paired-end reads from RNA-seq (fastq). (**B**) Alignment and chimeric reads: The alignment against transcripts is performed and their expression is calculated. Transcripts with FPKM < 1 are removed from the downstream analysis. Next, a list of reads aligned against transcripts is created. Through the alignment of reads against TE insertions, a second list with reads stemming from TEs is also created. Then, mapped paired-end reads and singletons are identified, generating the list of chimeric reads, for all replicates. All chimeric transcripts that have an average of chimeric reads $>= 2$ and are present in $>= 2$ replicates are maintained as true chimeras. (**C**) Transcriptome assembly and chimeric reads: The de novo transcriptome assembly is a non-default option of ChimeraTE Mode 2. It performs a transcriptome assembly and aligns reads against the assembled transcripts. Then, TE insertions in the assembled transcripts are identified with RepeatMasker and the TE reads are recovered. Using the two lists of reads (transcripts and TEs), the chimeric read list is generated and the putative assembled chimeric transcripts are predicted. Next, a blastn is performed between these transcripts and the reference transcripts provided in the input. All transcripts with length $>= 80\%$ are selected. The process is repeated for all RNA-seq replicates and chimeric transcripts assembled $>= 2$ replicates are maintained as true chimeras. (**D**) Chimeric transcripts: if the assembly is activated, ChimeraTE mode 2 provides three outputs: (1) Chimeric reads: predicted only based on the method depicted in B; (2) Assembled transcripts: predicted only based on the transcriptome assembly method depicted in C; and (3) Double evidence: predicted by both methods -B and C-.

read 2 IDs, regardless if their mates have also aligned (concordant mappings or singleton mappings). The same lists are created by using the transcript BED file: (1) read 1 IDs of transcript mapping reads and (2) mate 2 IDs of all mate 2 reads, regardless of mate mapping. All mate 2 IDs that have a TE-aligned read 1 are searched in the list of transcript-aligned mate 2. The same is performed in the opposite direction (TE-aligned read 2, transcript-aligned mate 1). These read pairs will therefore be comprised of two mates from the same pair that were singletons in the alignments, *i.e.* pairs comprised of one read that has aligned against a TE, and its mate against a gene transcript. The cases in which the chimeric transcript does not have the TE insertion in the reference transcript will be supported only by these singleton chimeric reads (Figure 2D). For cases in which the TE insertion is present inside the reference transcript, chimeric reads supporting it may either be singleton or concordant reads. Therefore, chimeric reads can be concordant reads in both alignments (TEs and genes), or they may be concordant only in the gene transcript alignment and singleton in the TE alignment. Due to the repetitive nature of TEs, short-read alignment methods provide very few unique aligned reads against loci-specific TE copies as most reads align ambiguously between similar TE insertions. Therefore, when a chimeric transcript has been identified involving more than one TE family, the TE family with the highest coverage of chimeric reads is maintained. Subsequently, ChimeraTE uses two chimeric reads as a threshold for calling a chimeric transcript, that can be modified by the user with the *–cutoff* parameter. Such value does not represent

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

transcript expression nor TE expression, but it represents the coverage supporting the junction between a gene transcript (CDSs/UTRs) and a TE sequence. Finally, the output tables show the list of genes and the respective TE families detected as chimeras, reference transcript ID and the total coverage of chimeric reads supporting it.

The support of chimeric transcripts performed by ChimeraTE Mode 2 is from chimeric reads aligned by an *end-to-end* approach. Such alignment may reduce alignment sensitivity, since exon/TE junctions may be covered by split reads. To mitigate the loss of detection power due to the alignment method with Mode 2, alongside with chimeric read detection using alignments against transcripts and TEs, there is an option to run Mode 2 with a transcriptome assembly approach, which can be activated with *–assembly* parameter (Figure 2C). This approach will use RNA-seq reads to perform a *de novo* transcriptome assembly with *Trinity v2.9.1* (48). *RepeatMasker v4.1.2* (49) is used to identify assembled transcripts that may have TE-derived sequences providing *–ref_TEs*, a custom TE library, or predefined TE consensus sequences from *Dfam v3.3* (50), according to the taxonomy level, i.e.: flies, mouse, humans. Then, RNA-seq reads are aligned with *bowtie2 v.2.4.5* (46) against the assembled transcripts. Subsequently, the alignment is used to identify whether transcripts containing TE-derived sequences have chimeric reads, including split reads. All assembled transcripts with chimeric transcripts are selected as candidate chimeric transcripts. Next, these candidates are submitted to a homology analysis with reference transcripts using *blastn v2.11.0* (51). Finally, all assembled transcripts with masked TEs that have at least 80% of similarity with reference transcripts across 80% of their length (can be modified with *–min_length* parameter) are considered chimeric transcripts. All these steps are repeated for all RNA-seq replicates provided in the input. Finally, the list of chimeric transcripts obtained from all replicates with the transcriptome assembly approach is compared, and all chimeras that have been identified with at least ≥2 chimeric reads and were found in ≥2 replicates, are considered as true chimeric transcripts. By activating the *–assembly* option in Mode 2, the output table will provide chimeric transcripts that have been predicted based on different evidences (Figure 2D): (1) only based on chimeric reads; (2) only based on transcript assembly; (3) based on chimeric reads and transcript assembly.

D. melanogaster wild-type strains: genome assemblies and RNA-seq

To assess ChimeraTE's performance, as well as the efficiency in the identification of chimeric transcripts derived from polymorphic TE insertions, we have mined previously available RNA-seq data from ovaries of four *D. melanogaster* wild-type strains (52), two from Gotheron, France (44_56′0″N 04_53′30″E), named dmgoth101 and dmgoth63; and two from São José do Rio Preto, Brazil (20_41′04.3″S 49_21′26.1″W), named dmsj23 and dmsj7. Such RNA-seq data was produced with mRNA libraries, which is strongly advised to use with ChimeraTE, since total RNA libraries contain pre-mRNA sequences and might introduce many false positives of TE-exonized transcripts. Sequencing was performed on an Illumina Hiseq (125 bp paired-end reads), with two biological replicates. All RNA-seq libraries were trimmed by quality, and adapters were removed with Trimmomatic (53). Each strain had also its genome previously sequenced by Nanopore long reads and assembled (54). The high-quality assemblies allowed us to manually check whether chimeric transcripts predicted by both ChimeraTE Modes have the predicted TE insertions inside/near genes, as well as manually curate the presence of chimeric reads.

Running ChimeraTE with *D. melanogaster* data

To run ChimeraTE Mode 1 on the available RNA-seq data, we performed gene annotation in the four *D. melanogaster* genomes with *Liftoff v.1.6.3* (55) using *-s 0.5 -s 0.75 -exclude_partial* parameters and the r6.49 *D. melanogaster* genome (*dm6* strain) available in Flybase as reference. TE annotation was performed with *RepeatMasker v4.1.2* (49), with parameters: *-nolow*; *-norna*; *-s;* and *-lib* with the TE sequence library for *D. melanogaster* provided by Bergman's lab *v.10.2* (https://github.com/bergmanlab/drosophila-transposons/blob/master/current/D_mel_transposon_sequence_set.fa). Small TE insertions with length < 80 bp were discarded due to the lack of robustness to define them as TEs, following the 80–80–80 rule (56). In addition, due to the presence of simple sequence repeats (SSRs) in many TEs from *D. melanogaster*, such as *roo*, *412*, *tirant* (57), several insertions from *RepeatMasker* might be indistinguishable from SSRs if the repeat region comprises a large proportion of the TE sequence (58,59). Therefore, to increase the robustness of the TE annotation, we identified the presence of SSRs with *TRF v.4.09* (60) across all insertions annotated on each strain, with parameters *2*, *5*, *6*, *75*, *20*, *50*, *500*, *-m*, *-h* and removed TEs that presented SSRs over >50% of their sequences. We used ChimeraTE Mode 1 with default parameters, setting *–strand rf-stranded*, since our paired-end reads have the orientation as reverse and forward for the first and second reads, respectively. In case the stranded paired-end reads were produced as forward and reserve (first and second reads, respectively), the parameter would be *–strand fwd-stranded*. For ChimeraTE Mode 2, we demonstrate its potential in detecting chimeric transcripts derived from TE insertions that are not present in a reference genome, even though the transcript sequences and TE copies provided as input stemmed from the *dm6* reference genome. Therefore, we have used ChimeraTE mode 2 with RNA-seq from the four wild-type strains listed previously, with reference transcripts from *D. melanogaster* (*dm6* strain) available in Flybase r6.49 (http://ftp.flybase.net/genomes/Drosophila_melanogaster/current/fasta/dmel-all-transcript-r6.49.fasta.gz) (61). The TE annotation for the *dm6* genome was assessed with the same protocol used on the four wild-type strains, with *RepeatMasker* (49). We have used default parameters, except by *–assembly*.

Additional species data

In order to demonstrate that ChimeraTE can be used with other species than *Drosophila* spp., we applied it to

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Arabidopsis thaliana, *Homo sapiens* and the fish *Poecilia reticulata* (guppy) RNA-seq datasets. The latter was selected as an example of non-model species. The genome and gene annotation of *A. thaliana* was obtained from NCBI (GCF_000001735.14), version TAIR10.1. The TE annotation was assessed with *RepeatMasker v.4.1.2* (49), with *parameters -species arabidopsis -cutoff 225 -nolow -norna -a -s*. We downloaded two replicates of paired-end RNA-seq from *A. thaliana* leaf, available on NCBI (SRR21230172, SRR21230173). Regarding human data, the genome, gene and TE annotations were retrieved from NCBI (62), version GRCh38.p14. Two paired-end RNA-seq replicates from the human leukemic cell-line K562 were downloaded from NCBI (SRR521457, SRR521461). Finally, *P. reticulata*'s genome and gene annotation were also retrieved from NCBI (GCF_000633615.1), and TE annotation was assessed with *RepeatMasker v.4.1.2*, parameters*: -species Actinopteri -cutoff 225 -nolow -norna -a -s*. Two RNA-seq replicates from ovary follicular tissue were downloaded from NCBI (SRR17332506, SRR17332508). Both ChimeraTE Mode 1 and Mode 2 were used with default parameters, except by *–threads 32*, and *–ram 64*.

Benchmarking polymorphic chimeric transcripts with nanopore genomes

Once chimeric transcripts were identified, we used the high-quality Nanopore assemblies for dmgoth101, dmgoth63, dmsj23 and dmsj7 previously published (54) to confirm whether genes predicted by Mode 1 as chimeric transcripts have indeed the respective TE insertion located near or within them. To do so, we have used an *ad-hoc* bash script to create three bed files from genes: 3 kb upstream; 3 kb downstream, and gene region. Then, we used *bedtools intersect* (44) to identify genes with TEs located in the three regions. For Mode 1 we have randomly sampled 100 chimeric transcripts of each wild-type strain to visualize the alignments performed by Mode 1 on the *IGV* genome browser (63). For Mode 2, all genes not found by *bedtools intersect* with the predicted TE insertion were visualized in *IGV*. In both manual curations, we considered false positives those cases in which we did not find the TE insertion, or we found the TE insertion, but without chimeric reads.

In order to assess the number of chimeric transcripts found by Mode 2 in wild-type strains derived from TE insertions absent in the *dm6* genome, we also used the *ad-hoc* bash script to create the bed files with 3 kb ± and the gene regions for *dm6*. Then, we used *bedtools intersect* (44) with TE annotation and the gene regions. By using this method, we generated a list of genes with TEs located 3 kb upstream, inside genes (introns/exons), and 3 kb downstream for the *dm6* genome. Then, the polymorphic chimeric transcripts were identified with the comparison of genes with TEs inside/nearby in *dm6* and the list of chimeric transcripts in the wild-type strains. In addition, all chimeric transcripts derived from TEs that were not found in *dm6* were manually curated with the IGV genome browser (63).

Validation of chimeric transcripts by RT-PCR

D. melanogaster strains were kept at 24°C in standard laboratory conditions on cornmeal–sugar–yeast–agar medium. For sampling, 3–7 day old *D. melanogaster* females were immersed in Phosphate-buffered saline solution (PBS) and dissected under a stereomicroscope. Three biological samples were collected in buffer TA (35 mM Tris/HCl, 25 mM KCl, 10 mM $MgCl_2$, 250 mM sucrose, pH 7.5), each composed of 30 pairs of ovaries. All samples were collected on ice in 1.5 ml RNAse free tubes and stored at −80°C until use. Total RNA was extracted using QIAGEN AllPrep DNA/RNA Micro extraction kit (Qiagen 80284) following the manufacturer's guideline. cDNA synthesis was performed on 1 μg of total RNA or water (RT negative control) with BIORAD iScript cDNA systhesis (Biorad 1708891). Primers were designed considering the TE and the exon location with aligned chimeric reads (Supplementary Table 1) and PCRs were performed with Taq'Ozyme (Ozya001-1000).

Sequence and protein analysis of TE-exonized elements

The sequences of TE-exonized elements were extracted from wild-type genomes with *bedtools getfasta* (44), parameter *-s,* using BED files created by ChimeraTE Mode 1. Since the TE reading frame incorporated into the gene transcript is unknown, we recovered all open reading frames (ORFs) with *EMBOSS v.6.6.0 getorf* (64) in the three coding frames, from both strands. Next, the protein domains in these sequences were assessed with *pfam-scan v.1.6* (65), using *Pfam-A.hmm* database from InterPro (66), with default parameters. We applied a filtering step for protein domains with multiple overlapped matches in the same TE insertion, keeping in the downstream analysis only the longest match. We also removed protein domains that had no association with any TE function, following the description from InterPro. Finally, to analyze whether *roo* elements generating TE-exonized embedded transcripts provided a specific motif to the chimeric exons, we extracted their sequences with *bedtools getfasta*, and aligned with *roo* consensus sequence with *MUSCLE v.5.1* (67). Alignment plots were performed with *MIToS v.2.11.1* (68), using the package *Plots*.

Differential expression analysis

The read count for differential gene expression analysis for the four wild-type strains was obtained from Fablet *et al.* (52). The data were normalized with *DEseq2 v.3.10* (69), with a designed model (∼genotype), assuming that the only variable to identify differentially expressed genes between strains must be the genotype. We performed pairwise comparisons with the four wild-type strains, considering differentially expressed genes the ones with adjusted p-value < 0.05.

RESULTS AND DISCUSSION

ChimeraTE predicts chimeric transcripts derived from genes and TEs using two different strategies. Mode 1 is a genome-guided approach that will predict chimeric transcripts from paired-end RNA-seq through chimeric read pair detection. The main advantages of Mode 1 in comparison to Mode 2 are that the first one can detect split reads between exons and TEs, capture chimeric transcripts with low

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

coverage/expression and classify chimeric transcripts according to the TE position: TE-initiated, TE-exonized, TE-terminated transcripts. However, Mode 1 misses chimeric transcripts derived from TE insertions that are absent from a reference genome. Mode 2, which is a genome-blind approach, performs two alignments against reference transcripts and TE copies and then, similar to Mode 1, predicts chimeric reads between these two alignments. In addition, Mode 2 can optionally perform a *de novo* transcriptome assembly able to detect chimeric transcripts and chimeric reads through split read alignment, improving the sensitivity. Despite the similarities between both Modes, the results from Mode 1 and 2 had an overlap of 50.46% (Supplementary Data, note 1). We have shown that such differences are likely to be associated with the alignment used by the Modes, and read length (Supplementary Data, note 1, Fig S1.1) (70,71). We also observed that increasing the minimum number of chimeric reads (default = 2) causes a decrease of 48.45% for Mode 1 and 33.48% for Mode 2 in the total number of chimeras found in two RNA-seq replicates (Supplementary Data, note 2). Therefore, a substantial part of the results is detected with low coverage of chimeric reads. Finally, we demonstrated that the use of RNA-seq replicates is crucial to reduce the frequency of potential chimeras supported by artifactual chimeric reads (Supplementary Data, note 3) (37,72–77).

Setting up the datasets for ChimeraTE

Each ChimeraTE Mode requires different input datasets. To run Mode 1 (genome-guided), gene and TE annotations, along with a genome fasta file are necessary. We took advantage of available paired-end RNA-seq datasets from the ovaries of four wild-type strains of *D. melanogaster* (52), for which high-quality genome assemblies were also available (54). We performed gene annotation in the new assemblies using *D. melanogaster*'s genome (*dm6*) from Flybase r6.49 as a reference and obtained ∼17,278 genes per genome (Supplementary Table 2). Regarding TE annotation, we found ∼10.48% of TE content in the four wild-type strains, similar to our previous estimates for these strains (54). Filtering out TE insertions smaller than 80 bp, removed ∼5,549 TEs (20.59%) across the four wild-type strains. In addition, to remove potential mapping artifacts, we discarded ∼752 TE copies (2.79%) per strain that harbored SSRs longer than half the TE length. On average across strains, 85% of the TEs filtered out due to the presence of SSRs were *roo* elements (Supplementary Figure 2). These *roo* copies were around 112 bp (Supplementary Figure 3), which is the same length as the known SSR present in the roo consensus sequence (57). In total, 123 TE families in the four genomes were uncovered, comprising a mean of ∼21,402 TE insertions (*standard deviation*= 883). In all genomes the TE content in bp is higher for LTR, then LINE elements, followed by DNA and Helitron families (Supplementary Table 2) (78). In order to run Mode 2 (genome-blind), we used reference transcripts from Flybase r6.49 and performed TE annotation on the reference *dm6* genome. We have obtained 29,086 TE insertions, representing 16.37% of the genome content, following similar TE family proportions as seen for the four wild-type genomes (Supplementary Table 2). Nevertheless, the *dm6* genome TE annotation shows an extra ∼6% of TE content compared to the wild-type genomes, mainly due to LTRs and LINEs (Supplementary Table 2). The *D. melanogaster's* heterochromatin corresponds to ∼32.5% of its genome (79), and includes repeat sequences, as TE copies (80). The wild-type assemblies have a high coverage for euchromatin regions, since 96% of gene content was recovered, but a poor heterochromatin coverage, corresponding to the lack of TEs compared to the *dm6* genome. However, the search for chimeric transcripts with ChimeraTE Mode 1 is restricted to the gene region and its 3 kb boundaries, hence, the lack of heterochromatic TE insertions does not impact the chimeric transcript search.

ChimeraTE mode 1 reveals that 1.12% of genes produce chimeric transcripts in *D.melanogaster* wild-type strains

ChimeraTE mode 1 was run on the four wild-type strain genomes and their respective ovarian RNA-seq data (52). Across all strains, we found 327 genes producing 408 chimeric transcripts, accounting for 1.83% of the total genes in *D. melanogaster*, 4.04% of the expressed genes (FPKM > 1), and representing 1.12% of all genes when normalized by the four strains. Previous studies have found similar proportion of genes generating chimeras in *D. melanogaster*. Among them, Coronado-Zamora et al. is the only study that performed chimeric transcripts search in multiple strains and different tissues, including ovaries, for which the authors found 1.41% of genes with chimeric transcripts when normalized by all strains (81). To verify whether ChimeraTE identified the correct TE family for each chimera, we have compared the genomic coordinates of TEs and genes (3 kb upstream/downstream and inside genes) with *bedtools intersect* (44) and predicted genes with TE copies in these regions. All chimeric transcripts in the four wild-type strains had at least one TE insertion in the expected position from the predicted TE family (Supplementary Tables 3–6). In addition, we randomly selected 100 chimeric transcripts in each wild-type strain to visualize in IGV (63) and confirm the presence of chimeric reads as expected (Figure 3A). All the 400 manually inspected chimeras were correctly found in the genome browser. Among all genes generating chimeric transcripts, around 89% are TE-exonized (61.84% correspond to TE-exonized embedded, 17.18% to TE-exonized intronic, 9.95% to TE-exonized overlapped), 8.64% TE-terminated and 2.36% TE-initiated transcripts (Figure 3B). It is important to note that ChimeraTE does not classify an internal TE copy as either a TE-initiated or a TE-terminated transcript, even if the transcript does indeed start, or end, at the TE sequence. TE-initiated and TE-terminated repeats are necessarily outside of gene regions as *per* ChimeraTE categories (see Materials and Methods). Therefore, one should assume that the high prevalence of TE-exonized transcripts (∼89%) might be associated with potential cases of misannotated TE-initiated and TE-terminated transcripts. Indeed, a previous study in *D. melanogaster* has shown that ∼2.8% of genes have TE-derived promoters in embryos, pupae, and larvae tissues (33).

A recent analysis in *D. melanogaster* ovaries has found 884 chimeric transcripts derived from 549 genes, among

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

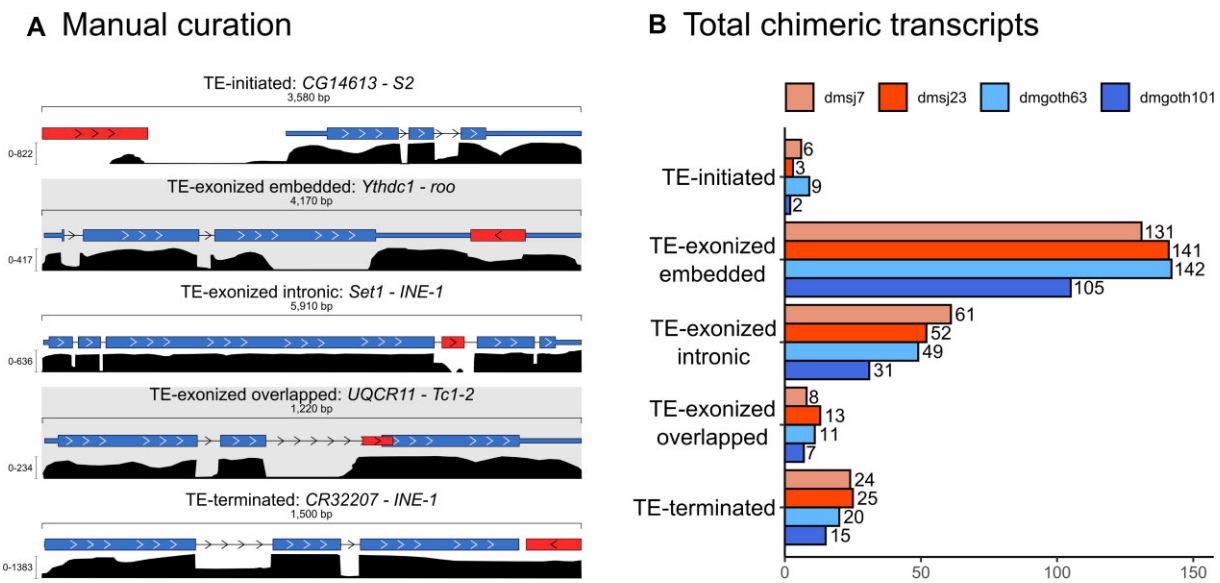


Figure 3. General results from Chimera Mode 1. (**A**) Five examples of chimeric transcripts manually curated with the IGV genome browser. Red boxes: TE insertion; blue boxes: exons and UTRs; black density graphs: coverage of RNA-seq reads; head arrows: transcription sense, blue and red boxes linked by a line: chimeric reads. (**B**) Total genes generating chimeric transcripts, following the TE position classification in the four wild-type strains.

five wild-type strains (81), indicating that a gene is capable of producing multiple chimeric transcripts involving different TE copies. In agreement, TE-exonized chimeric genes detected by ChimeraTE, show evidence of multiple TE sequences acting as chimeric transcripts (22.6% of TE-embedded genes, 30.95% for overlapped, and 30.34% for intronic). Nevertheless, TE-initiated transcripts do not involve more than one TE *per* gene. Among 46 TE-terminated transcripts found in all strains, only five genes were reported with more than one TE copy: (1): *NADH dehydrogenase* (*CG40002*) in dmgoth63; (2,3) *Max* and *rudimentary-like* in dmsj23; and (4,5) *six-banded*, *CG14464* in dmsj23. All of them, except *CG14464*, were associated with two TE insertions from the same TE family, located close to each other (Supplementary Figure 4), suggesting they might be multiple hits from *RepeatMasker* corresponding to only one TE copy (82). Thus, multiple TE copies producing chimeric transcripts exist only when TE copies are within genes. We then tested whether the number of TEs within genes has a correlation with the presence of multiple TE-exonized chimeras. By using all the genes expressed in ovaries (FPKM $>$ 1) with at least one TE insertion within, we found weak positive correlation between them (Supplementary Figure 5), among which the highest was found in dmsj23 (Pearson; $r = 0.22$; $P = 2.6\text{E-}08$), and non-significant for dmgoth101 (Pearson; $r = 0.06$; $P = 0.08$). However, when we take into account the number of TE copies per TE family present within genes, and the respective number of chimeras generated by TE family, we observed a positive correlation (Pearson; $r = 0.87$; $P < 2.2\text{E-}16$) (Supplementary Figure 6). Therefore, TE-rich genes are not the most frequent among the ones generating chimeras, but TE families with high copy number within genes are the ones likely to produce TE-exonized transcripts.

Among all chimeric transcripts predicted by ChimeraTE Mode 1, 11 chimeras have been previously described in *D. melanogaster* using EST data (31). It is important to note that the authors did not consider heterochromatic regions, as well as chimeric transcripts derived from INE-1 elements. From these 11 chimeras, five were found in all strains: *Ssdp* (gene) - *HMS-Beagle* (TE family); *Agpat1-1360*; *anne-1360*; *Atf6-Xelement* and *ctp-HMS-Beagle*. The other six chimeric transcripts are absent from some strains: *CG6191-jockey*; *CG15347-HB*; *CG3164-McClintock; Svil-roo; CHKov1-Doc* and *Kmn1-pogo* (Supplementary Table 7). We manually checked these chimeric transcripts in the IGV genome browser and we confirmed the presence of the TE families predicted by ChimeraTE (Supplementary Table 7). In addition, we also found three chimeras reported previously with RAMPAGE: *CG31999-Tc1-2*; *CkIIα-1366*; *Svil-roo* (33). Finally, in another study on *D. melanogaster* midbrain, 264 genes were shown to produce chimeric transcripts using single-cell RNA-seq data (34). The authors demonstrate that retrotransposons have splice donor and acceptor sites that generate new chimeric isoforms through alternative splicing (34). Despite the differences between tissues and methods, we found 22 chimeric transcripts identified previously using scRNA-seq (34) (Supplementary Table 7), of which two were previously found in ESTs (31). Taken together, the genome-wide analysis performed by ChimeraTE Mode 1 has uncovered 296 genes with chimeric transcripts in the *D. melanogaster* ovarian tissue for the first time.

ChimeraTE mode 2: a method to uncover chimeric transcripts without genome alignment

D. melanogaster has a high rate of TE insertion polymorphism across worldwide population (83–85). To test the ability of ChimeraTE Mode 2 in detecting chimeric transcripts derived from TE insertions absent from a reference genome, we used RNA-seq from the four wild-type strains with transcript and TE annotations from the *D. melanogaster* reference genome (*dm6*). ChimeraTE

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

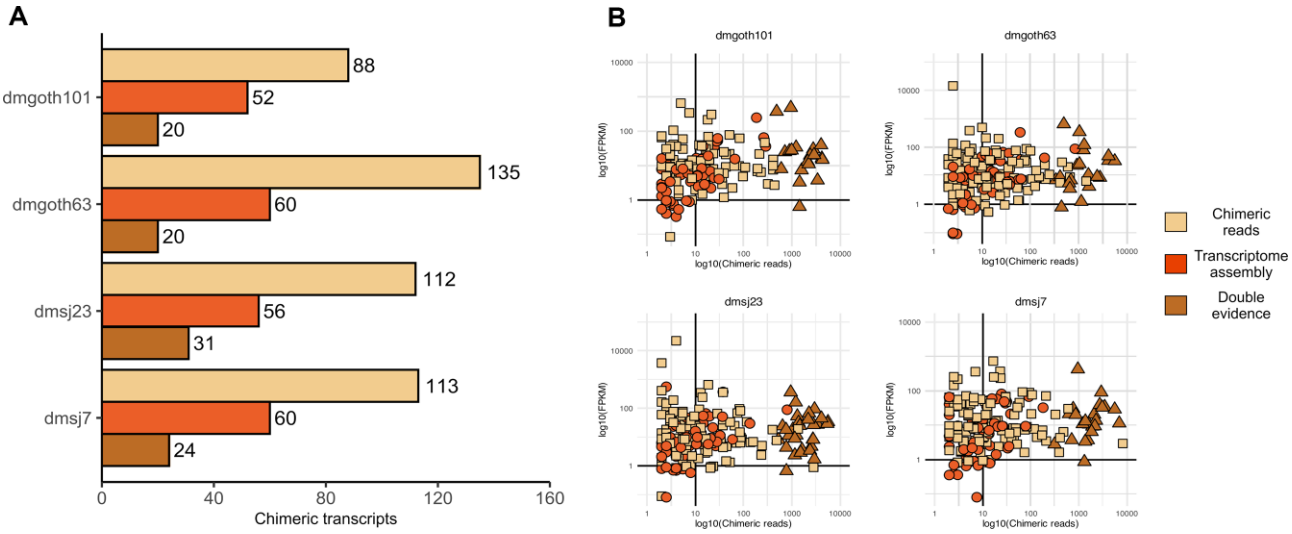


Figure 4. (**A**) Total number of chimeric transcripts found by ChimeraTE mode 2. "Assembled transcripts": chimeric transcripts detected only by the method of transcriptome assembly (Figure 2C). "Chimeric reads": chimeric transcripts detected only by the method of chimeric reads (Figure 2B). "Double evidence": chimeric transcripts detected by both methods. (**B**) Correlation between chimeric read coverage and gene expression of chimeric transcripts found by Mode 2. In all strains, we did not observe correlation between gene expression (FPKM) and the coverage of chimeric reads between the three categories of evidence. All double evidence chimeras were found with high coverage of chimeric reads, suggesting high reliability of chimeras when both methods (chimeric reads and transcriptome) are considered together.

Mode 2 may predict chimeric transcripts based on two types of evidence: either using chimeric reads only or by taking advantage of *de novo* transcriptome assembly. Chimeric transcripts with both evidences are named 'double-evidence' (chimeric reads and transcript assembly - Figure 2B and C). Among the four wild-type strains, ChimeraTE Mode 2 has identified 324 genes (Supplementary Table 8–11) producing 339 chimeric transcripts (Figure 4A), representing 1.81% of the total *D. melanogaster* genes, and 1.07% when normalized by the four wild-type strains. Comparing the 'chimeric read' approach with the 'transcript assembly' one, the former method found ∼51% as many chimeric transcripts than the detection by transcriptome assembly (Figure 4A). This is probably due to the technical limits of *de novo* transcriptome assembly of repetitive-rich transcripts (86), which may lead to the lack of chimeric transcript isoforms. Between both methods, we did not find any correlation with chimeric read coverage and gene expression (FPKM), indicating that even low expressed genes (FPKM < 10) can be detected by both methods independently (Figure 4B). Double evidence chimeras represents ∼14% of all chimeras detected by Mode 2, and they have been found only with high chimeric read coverage (Figure 4B).

As each chimeric transcript was detected based on reference transcripts and TE insertions, we used the Nanopore assemblies to manually inspect the presence of the predicted TE family inside and near (±3 kb) genes, with the intersection of genomic coordinates from genes and TEs with *bedtools intersect* (44). We considered true chimeric transcripts the cases in which we found in the assembly the presence of the predicted TE insertion inside/near the gene. The alignment of RNA-seq libraries against the genome sequence was also used to confirm the presence of the TE insertion, as well as the presence of chimeric reads, with IGV genome browser (63). The manual curation was performed with the three groups of results from Mode 2 ('chimeric reads', 'transcriptome assembly' and 'double-evidence'). We found that 96.30% of all chimeric transcripts predicted by double evidence were true, whereas we observed 90.59% from transcriptome assembly and 73.13% from chimeric reads. Therefore, taken together, ChimeraTE Mode 2 has provided a reliable inference with 81.2% sensitivity (weighted average), based on genomic manual curation.

The main difference of ChimeraTE Mode 2 from the previous pipelines is its ability to detect chimeric transcripts derived from TEs that are absent in a reference genome, using RNA-seq from non-reference individuals/cells/strains. We first identified in the *dm6* reference genome the genes with TEs located upstream, inside, and downstream. We found that 2,239 genes in the *dm6* genome have TEs located 3 kb upstream, 1,863 inside (introns and exons), and 2,320 downstream. These genes were selected as candidates to produce chimeras in the *dm6* genome and then compared with the list of chimeric transcripts generated by ChimeraTE Mode 2 in the four wild-type strains. In addition to the comparison with *dm6* genes harboring TEs inside/near, we have manually curated the TE insertions and the presence of chimeric transcripts with the IGV genome browser (63). Such manual curation was performed with the wild-type genomes, by checking the presence of the TE and chimeric read alignments. For instance, the *Mps1-FB* chimera is an interesting case, since it is the only chimeric transcript specific to French populations, as we found it in both dmgoth101 and dmgoth63. In the *dm6* reference genome, *Mps1* has an overlap between its 3′ UTR and the 3′ UTR of the *alt* gene, located on the other strand (Figure 5A). The same distribution of both genes is found in the Brazilian strains, dmsj23, and dmsj7 (Figure 5B). Conversely, in dmgoth101 and dmgoth63, there is a gap of ∼9,500 bp between *Mps1* and *alt*,

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

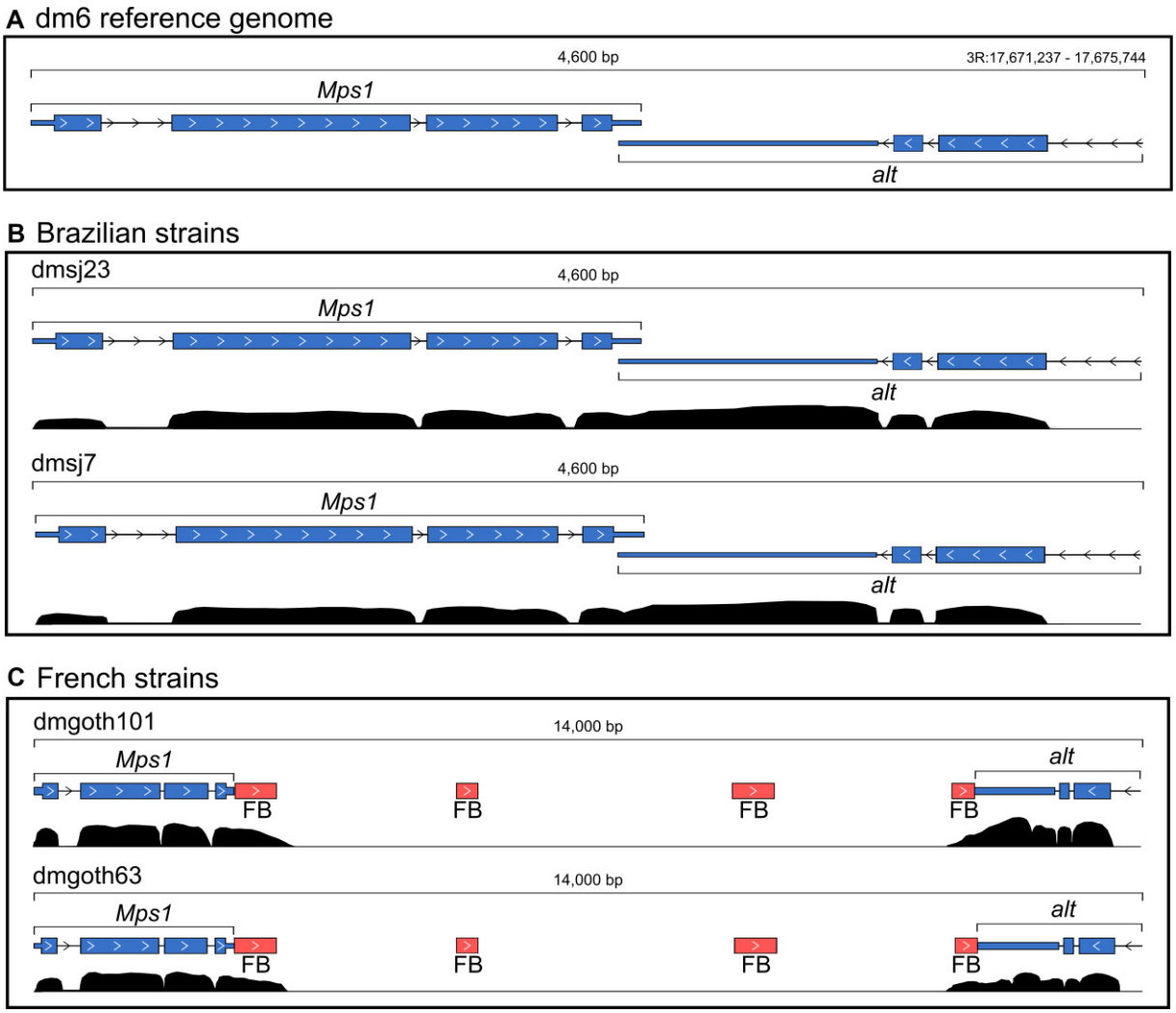


Figure 5. *Mps1* gene and its downstream region in the genomes of the four wild-type strains. (**A**) *Mps1* in the *dm6* reference genome and the *alt* gene located downstream to it, on opposite strands and with overlapped 3' UTRs. (**B**) In the dmsj23 and dmsj7, *Mps1* and *alt* are distributed as found in the reference genome. (**C**) In dmgoth101 and dmgoth63, there is a *FB* insertion located downstream to *Mps1*, which has chimeric reads supporting a TE-terminated downstream in both strains.

with four small *FB* copies of ∼412 bp in length (Figure 5C), indicating that it may be an old insertion. Furthermore, one of the *FB* copies is located downstream of the *alt* gene, which has also been identified as TE-terminated downstream in dmgoth63 and dmgoth101. This result shows that ChimeraTE Mode 2 can detect chimeric transcripts derived from TEs that are not present in the reference genome.

Taking into account all chimeric transcripts detected by Mode 2, we found 11 genes (3.40%) generating chimeric transcripts derived from TEs that are absent from the *dm6* genome (Table 1). There are specific chimeras from French strains: *Mps1-FB*, *CG1358-S* and *CG46280-POGON1* genes, but only *Mps1* had chimeric transcripts in both French dmgoth63 and dmgoth101 strains, whereas *CG1358* was found as chimera only in the dmgoth101, and *CG46280* only in dmgoth63. The *Ythdc1-roo* chimera was observed with a strain-specific polymorphic *roo* element inside an exon in the dmgoth63 strain. Regarding the Brazilian strains, we found two TE insertions present in both dmsj23 and dmsj7, involving the genes *cic* and *TrpRS-m*, but they were found as chimeras only in the dmsj23 strain. We also found *rb*, *r-1* and *ArfGAP1* with dmsj23-specific TE insertions, whereas in the dmsj7 we found *caps* and *Ubi-p5E* with specific TE insertions giving rise to chimeric transcripts.

To evaluate Mode 2 accuracy in detecting chimeric transcripts derived from TEs absent in the reference genome, we have performed RT-PCRs for 10 out of 11 cases predicted (it was not possible to design specific primers for the *TrpRSm-protop_a* chimera). The primers were designed considering the TE and the exon location with aligned chimeric reads (Supplementary table 1). For *Mps1-FB*, *caps-pogon1*, *Ythdc1-roo*, *Ubip5E-gypsy*, *CG46280-pogon1*, the amplified fragment sizes were in agreement with ChimeraTE Mode 2 predictions (Supplementary Figure 7A–E). The *r1-stalker2* chimera was amplified in both Brazilian strains, even though the TE insertion generating this chimeric transcript was absent in dmsj7 genome assembly

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Table 1. 11 polymorphic chimeric transcripts from TE insertions that are not present in the *dm6* reference genome. The red triangle in a line represents the presence of a chimeric transcript; the white triangle in a line represents the presence of the TE insertion but without a chimeric transcript; the line without triangles represents the absence of the TE insertion

Gene	TE	TE location	Strains			
			dmgoth101	dmgoth63	dmsj23	dmsj7
FBgn0000063 - Mps1	FB4	TE-terminated	▲	▲	—	—
FBgn0003210 - rb	MDG3	TE-terminated	—	—	▲	—
FBgn0003257 - r-l	Stalker2	TE-terminated	—	—	▲	—
FBgn0020655 - ArfGAP1	POGO	TE-terminated	—	—	▲	—
FBgn0023095 - caps	POGON1	TE-exonized	—	—	—	▲
FBgn0027616 - Ythdc1	ROO	TE-exonized	—	▲	—	—
FBgn0033196 - CG1358	S	TE-terminated	▲	△	—	—
FBgn0036763 - TrpRS-m	PROTOP_A	TE-exonized	—	—	▲	△
FBgn0086558 - Ubi-p5E	Gypsy	TE-initiated	—	—	—	▲
FBgn0262582 - cic	POGON1	TE-exonized	—	—	▲	△
FBgn0283438 - CG46280	POGON1	TE-exonized	△	▲	—	—

(Supplementary Figure 8A). We hypothesized that this might be due to a low-frequency TE in the dmsj7, absent from the assembly and without enough read coverage in the RNA-seq to predict it. Indeed, we confirmed the presence of this TE insertion by PCR in both Brazilian strains (Supplementary Figure 8B). The chimeric transcript from *rb-mdg3* was detected only in dmsj23 by ChimeraTE Mode 2, but the RT-PCR assay was successful only in the dmsj7 strain (Supplementary Figure 9). The chimera derived from *CG1358-S* was predicted only in dmgoth101 (Supplementary Figure 10), despite the TE insertion being also present in dmgoth63 genome. By RT-PCR, we detected this chimeric transcript in both strains, suggesting that it was not detected in dmgoth63 by ChimeraTE due to Illumina coverage issues. A similar result was found for the *cic-pogon1* chimera, for which we amplified the chimeric transcript in both Brazilian strains, but it was detected only in dmsj23 by ChimeraTE (Supplementary Figure 11). Altogether, these results indicate that Mode 2 can predict chimeric transcripts from TEs absent from the reference genome, but this prediction depends upon the coverage of chimeric reads in the RNA-seq data.

Evaluation of chimeric transcripts in dmgoth101 through nanopore RNA-seq

Short-read RNA-seq may introduce biases during library preparation (e.g. reverse transcription, second-strand synthesis, PCR amplification) (87), or even during transcriptome assembly, with loss of information regarding isoform diversity, especially for those with low-expression and repetitive regions (88), which might be the case of many chimeric transcripts. In the past years, the third generation of long-read sequencing has been used to mitigate such biases, because there is no need to perform transcriptome assembly. Thus, we validated chimeric transcripts predicted by ChimeraTE Mode 1 with short reads using long reads produced with Oxford Nanopore Technology (ONT) from dmgoth101 ovaries (89). We did not perform the same validation with RNA long reads for chimeras from Mode 2 predicted with short reads, due to the lack of information regarding the genomic position of TEs generating them. The long reads were aligned to the dmgoth101 genome and then checked for the presence of all chimeric transcripts detected by ChimeraTE Mode 1 in dmgoth101. The presence of the chimeric transcripts was confirmed when the long read was aligned into the gene exons, and it had the TE-derived sequence as part of the reads. From two TE-initiated transcripts, we found only *CG14613-S2* chimera by ONT reads. In TE-terminated transcripts, six (40%) out of 16 were also observed by ONT data. Finally, in TE-exonized transcripts, we identified 89 (62.24%) out of 143 embedded TEs, 12 (52.17%) out of 23 overlapped TEs, and 13 (46.43%) out of 28 intronic. The undetected chimeras through ONT reads might be associated with the long-read library construction protocol, TeloPrime, which was used to generate full-length cDNA libraries. Such a protocol might miss weakly expressed isoforms, as well as truncated 3′ ends due to RNA degradation *in vivo* or during RNA manipulation (90). We investigated whether short-read coverage could explain the lack of detection of certain chimeric transcripts with ONT data. Taking all dmgoth101 chimeras from Mode 1, chimeric transcripts found by both ChimeraTE and ONT data had an average chimeric read coverage of 69.44, whereas the undetected ones had 10.83. Therefore, these results indicate that the detection of chimeras with ONT reads is likely to be more efficient with highly expressed chimeric transcripts. Overall, our results demonstrate that 50.16% of all chimeric transcripts detected in the ovary transcriptome of *D. melanogaster* with ChimeraTE Mode 1 can also be found by ONT RNA-seq.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

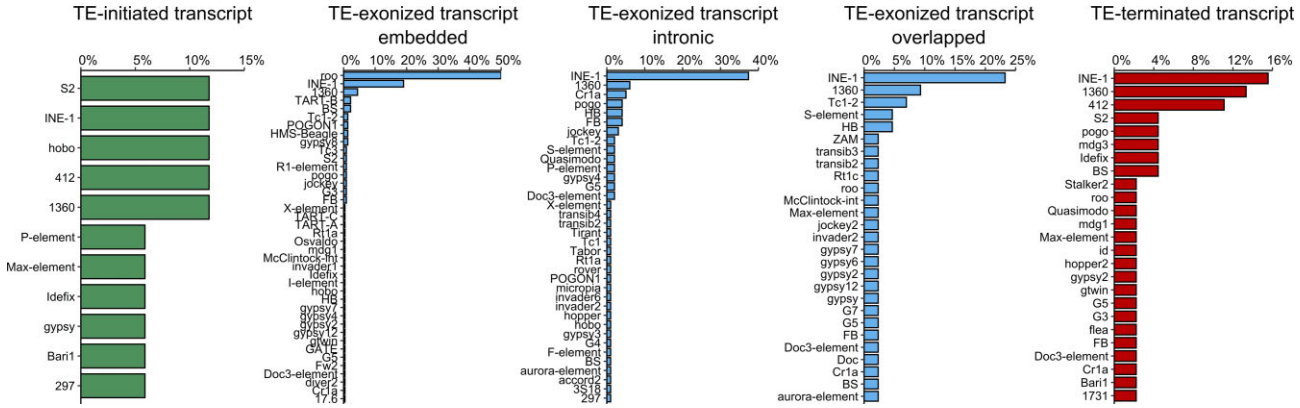


Figure 6. The frequency of the 76 TE families generating chimeric transcripts in the four wild-type strains. In chimeric transcripts derived from TEs near genes, *INE-1* was the most frequent (15%) in TE-terminated downstream, whereas for TE-initiated upstream, *S2*, *INE-1*, *hobo*, *1360* and *412* had the same frequency (11%). Regarding TEs inside genes, the *roo* element has the highest frequency of TE-exonized embedded, representing 50% of all chimeras. In TE-exonized intronic and overlapped, *INE-1* was the most frequent TE family.

The high prevalence of *roo* elements in chimeric transcripts is due to ∼132 nt TE insertions embedded into exons

Several TE families have been associated with chimeric transcripts in *D. melanogaster* (31,34). Here, we found 76 TE families producing at least one chimeric transcript in the four wild-type strains, representing 61.79% of all TE families annotated. We found a positive correlation between the frequency of TE families in chimeric transcripts and their respective abundance in the *D. melanogaster* genomes (Pearson; $r = 0.53$; $P < 2.2E\text{-}16$). In the TE-initiated transcripts, *S2*, *INE-1*, *hobo*, *1360* and *412* had the highest frequencies (Figure 6). In TE-exonized transcripts, the prevalence of *roo* elements is more pronounced in the embedded TEs group, comprising 50% of all chimeras, followed by *INE-1* elements, with 19%.

Interestingly, for both TE-exonized intronic and overlapped, *INE-1* had the highest frequencies, with 37.37% and 23.25%, respectively (Figure 6); whereas *roo* element was observed in only one chimera (*Lk6 - roo*, in dmgoth63 and dmsj23) as a TE-exonized overlapped. Notably, the *1360 element* is among the most frequent TE families in all TE-exonized categories. The other 38 families have an average frequency of 2%. Finally, in TE-terminated transcripts, *INE-1* is the most frequent TE, with 15.5% of all cases, followed by *1360* and *412*, with 13% and 11%, respectively. Taken together, these results suggest that *INE-1, S2, 412* and *1360* are candidate TE families to investigate the exaptation of regulatory motifs since they were the most abundant in TE-initiated and TE-terminated transcripts. On the other hand, in TE-exonized transcripts, *roo* might have a potential for exaptation/domestication of both regulatory and protein domain motifs, since it is observed with high prevalence within exons (UTRs and CDSs). Regarding TE-exonized transcripts derived from TEs at intronic regions and overlapping with exons, *INE-1* and *1360* are the families with the highest frequency of intron retention, suggesting they might harbour splice sites allowing incorporation into the gene transcripts.

We then investigated whether exonized TE insertions generating TE-exonized transcripts could contribute to protein domains, especially *roo* due to its high prevalence in TE-exonized embedded chimeras. Through our analysis with *pfam-scan*, we found in total 28 genes with TE-exonized transcripts from TE insertions harbouring a TE protein domain, corresponding to an average of ∼16 per strain (Supplementary Table 12). We assessed the completeness of these protein domains conserved on these insertions, assuming domains with >75% of completeness as complete, and <75% as partial. Our analysis revealed that ∼45% of all domains are likely to be conserved (Figure 7A), for which 92% correspond to catalytic domains, and 8% to zinc finger binding domains (PRE_C2HC – PF07530.14; zf-CCHC – PF00098.26). In addition, 17.69% of all TE-derived protein domains stem from embedded TE insertions, 27.43% from overlapped with exon, and 54.86% from intron regions. Furthermore, only one *roo* insertion had protein domains detected (Peptidase A17; DUF5641), which is embedded within CR46472 long noncoding RNA gene. Taking into account that 50% of the TE-exonized transcripts are derived from embedded *roo* insertions, this result suggests that despite their high frequency, they are old insertions without catalytic potential.

Since we predict these chimeric transcripts only based on chimeric reads, we are not able to confirm whether these domains have been fully incorporated into gene transcripts, limiting our biological conclusions regarding catalytic or binding effects. There are only three chimeric transcripts with TE-derived protein domains for which we found the whole TE sequence incorporated into the transcript with ONT RNA-seq data (Supplementary Figure 12). Among them, *CG4004 – hobo*, and *nxf2 – TART-A* are present into CDS regions, suggesting a putative functional role of the TE sequences. The third chimera, *CR42653 – Doc*, is a noncoding pseudo-gene, which suggests that the TE-derived protein domain does not have a functional role.

TE insertions disrupting coding regions and promoters are more likely to be deleterious, being consequently eliminated by purifying selection (91), although several exaptation events have been documented (23,92). The presence of *roo* elements in 50% of all TE-exonized embedded transcripts suggests a neutral or adaptive role of this family when incorporated into gene transcripts. *Roo* is an LTR

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

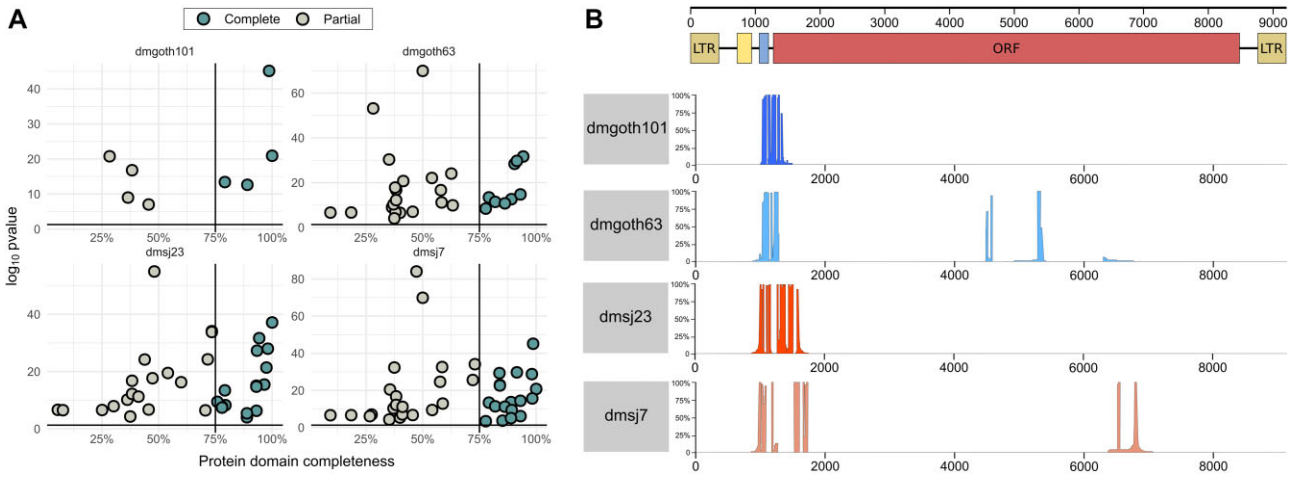


Figure 7. (**A**) Completeness of protein domains identified in TE insertions generating TE-exonized transcripts, with p-value < 0.05. (**B**) Alignment depth with *roo* consensus and embedded *roo* insertions generating TE-exonized transcripts. At the top, the scheme of the full-length roo element: brown boxes: LTRs; yellow box: first tandem repeats at 5' UTR; blue box: second tandem repeat at 5' UTR; red box: Open reading frame (ORF). The coverage depth of the multiple alignments between embedded roo insertions and the consensus is separated by strain.

retrotransposon, encoding three proteins: *gag*, *pol*, and *env*, which have been through domestication events from retroelements in many species, including *Drosophila* (8,93–94). It is the most abundant euchromatic TE family of *D. melanogaster* (1,95,96), and its insertions have been associated with modifications in the expression level of stress response genes due to the presence of TFBSs (97). They have also low enrichment of repressive histone marks (98,99), potentially explaining their high transposition rates (100,101). We analyzed the *roo* sequences that are embedded into exons of TE-exonized transcripts. The length of these chimeric insertions is ∼132 bp, confirming that they are old insertions since the full-length *roo* consensus sequence is 9,250 bp (Figure 7B). Subsequently, we analyzed whether these exonized *roo* insertions are donors of preferential motifs to the chimeric transcripts. All *roo* insertions stem from a specific region, between the 5′ UTR and the *roo* open-reading frame (Figure 7B). Compared to a *roo* consensus sequence, in dmgoth63 and dmsj7, most insertions show a deletion from the 5′ UTR up to the middle of the ORF. Despite the low nucleotide diversity of *roo* insertions in *the D. melanogaster* genome, the 5′ UTR region has a hypervariable region, including deletions and repeats, with several copies missing a tandem repeat of 99 bp (57,102). It has been proposed that this region may have a role in *roo* transposition, by heterochromatinization, recruitment of RNA pol II, and interaction with other enzymes (102). Curiously, a study assessed the nucleotide diversity between *roo* insertions, and characterized the same region as a deletion hotspot (78). A recent work also observed this region as part of TE-exonized transcripts in a transcriptome-wide manner in *D. melanogaster* (81). Why this region has been maintained through evolution, and whether it could be adaptive or neutral, remains unclear.

Polymorphic TE insertions in the wild-type strains generate 76 chimeric transcripts

Polymorphic TE insertions are common across *D. melanogaster* strains (103,104). To quantify how many chimeric transcripts in the wild-type strains are derived from polymorphic TE insertions compared to the reference genome, we used the list of genes with TEs located 3 kb upstream, inside (introns and exons), and 3 kb downstream generated previously for the dm6 genome. These genes were selected as potential sources of chimeras and then compared with the list of chimeric transcripts generated by ChimeraTE Mode 1 in the four strains. We found that 76 genes (23.24%) with chimeric transcripts in the wild-derived strains were generated by TE insertions that are absent in the reference genome (Supplementary Table 13). The number of genes generating chimeras from polymorphic TEs found by Mode 1 is higher than Mode 2 (Table 1; 11 genes). Such result is likely due to the divergence of TE insertions from the reference genome that is used with Mode 2, decreasing the alignment rate, whereas with Mode 1 we use strain-specific insertions. From the 76 genes, we found nine out of the 11 chimeras derived from polymorphic insertions identified by Mode 2 (Table 1). The chimeric transcripts from *CG1358* and *caps* genes were not found by Mode 1, despite the TEs presence on the genomes. Considering the 76 chimeras, we observed twenty-seven TE families generating polymorphic chimeric transcripts: *roo* (44.15%), *412* (6.49%), *pogo* (6.49%), *POGON1* (6.49%), among others. Except for *INE-1*, for which we found two polymorphic chimeras, most of the other TE families are known to be active in *D. melanogaster* (57). *INE-1* chimeras are unexpected since it is an old and inactive TE family in *D. melanogaster* (105), however, *INE-1* polymorphism among *D. melanogaster* populations has previously been shown (106).

The study of TEs in wild-type strains offers new insights into their ability to provide genetic variability. Depending upon the position where TEs are inserted regarding genes, they can contribute to gene expression or protein sequence variation. We then compared whether polymorphic TE insertions generating chimeric transcripts are more likely inserted near or inside genes. We found that 5.19% correspond to TEs located upstream, 74.03% inside (intron/exons) and 20.78% downstream (Figure 8). In the

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

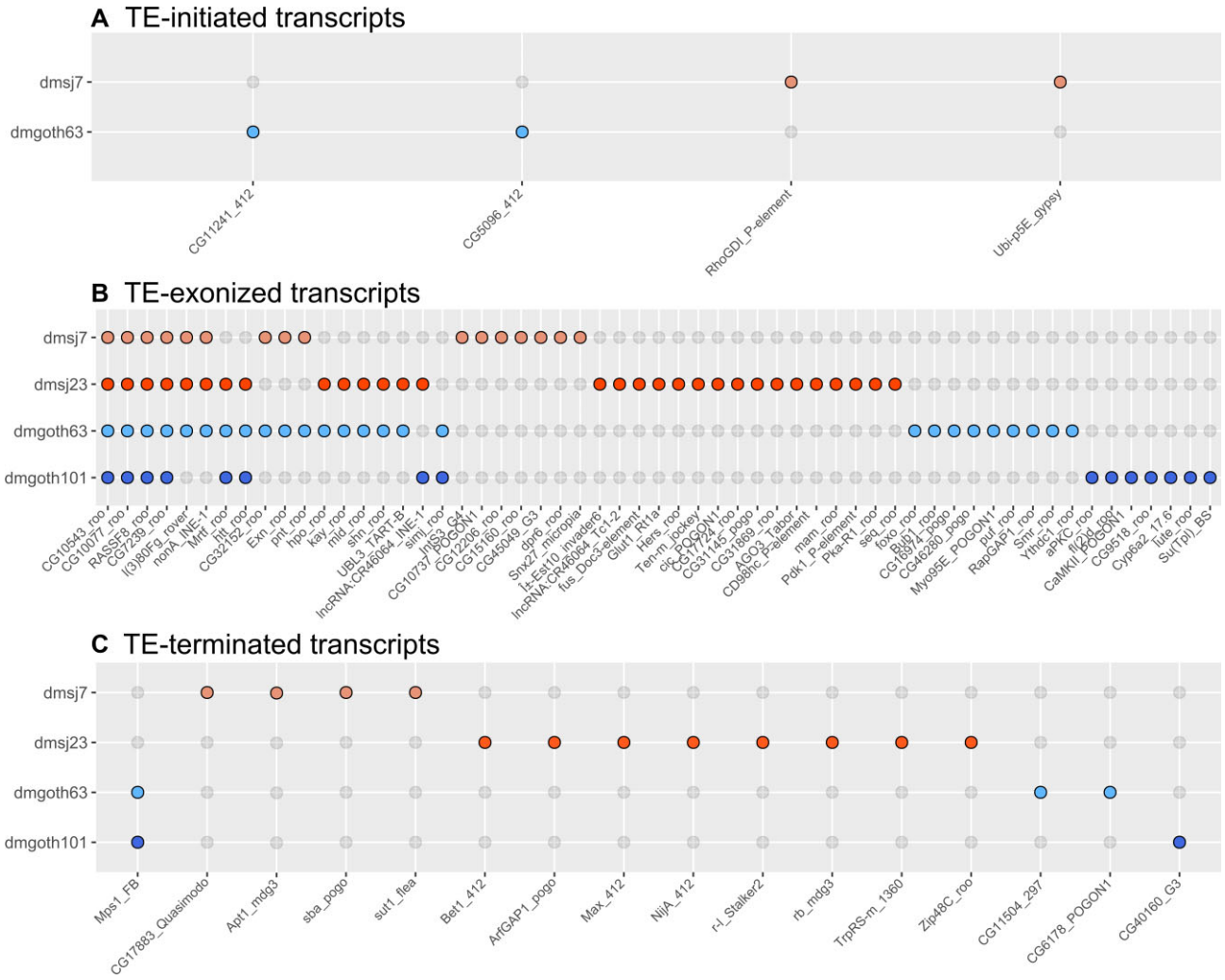


Figure 8. The 76 chimeric transcripts derived from TE insertionsthat are absent in the dm6 reference genome, 5.19% of them correspond to TEs located upstream, 74.03% to TEs located inside genes (introns and exons), and 20.78% to TEs located downstream. (**A**) TE upstream: Chimeric transcripts in which the TE is located up to 3kb upstream of the gene. (**B**) TE inside: Chimeric transcripts with TE insertions located inside the gene region (exons and introns). There are four chimeric transcripts found in all strains, and one specific to French strains. (**C**) TE downstream: Chimeric transcripts in which the TE is located up to 3kb downstream of the gene. Only Mps1-FB was specific to French strains, whereas all the other 15 chimeras are strain-specific.

genomic context, we observed ∼1,767 polymorphic TE insertions in these three regions regardless of the presence of chimeric transcripts, for which we found an equal distribution between TEs upstream (∼32%), TEs inside genes (∼35%) and TEs downstream (∼32%). Such result suggests that polymorphic insertions are not preferentially maintained by selection in specific gene locations, but chimeric transcripts from polymorphic TEs are more likely to be generated when they are within gene regions. In addition, we investigated whether the TE insertions absent from the reference genome could either be population-specific (Brazil or France) or strain-specific. The four TE-initiated transcripts derived from polymorphic TE insertions were strain-specific, two from dmgoth63 and two from dmsj7 (Figure 8A). For TEs inserted inside genes, 66.67% were found only in one strain. Only four chimeric transcripts are common to all strains *CG10543 - roo*, *CG10077 - roo*, *RASSF8 - roo* and *CG7239 - roo* (Figure 8B). We did not find a TE-exonized derived from a TE insertion present only in the Brazilian strains, but we did find one for the French strains, the *simj – roo* chimera. Interestingly, nine chimeras were found between pairs of Brazilian and French strains (Figure 8B), indicating retention of ancestral polymorphisms in these populations due do incomplete lineage sorting (107). Finally, for TEs located downstream genes (Figure 8C), the chimera *Mps1–FB* is the only one found in both French strains, and absent from Brazilian strains, as we observed by Mode 2 (Figure 5). Taken together, these results reinforce the potential of TEs in generating genetic novelty between *D. melanogaster* strains.

Modification in gene expression is not a general rule for chimeric transcripts

TEs can modify gene expression when inserted inside and in their close vicinity. We hypothesized that genes with

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

strain-specific chimeric transcripts may have differences in their expression levels in comparison to strains without the same chimeras. We first identified 67 strain-specific chimeras per strain, through pairwise comparisons. Then, through differential expression analysis between the four strains, we found that most genes have a similar expression level regardless of the presence of chimeric transcripts (Figure 9). In average, only 8 genes per strain were differentially expressed (Figure 9). Taking all these genes together, these chimeras are composed of 11.34% TE-initiated, 76.28% TE-exonized and 12.37% TE-terminated.

Among differentially expressed genes generating strain-specific chimeras, we observed that ∼47% had $\log_2$ fold-change <|2|, representing a mild transcription modulation. The other ∼53% (36 genes) had on average a log2 fold-change of ∼4.5 indicating higher expression when the chimera is present. Nevertheless, the gene expression quantification does not take into account differences between chimeric and non-chimeric isoforms. Therefore, we cannot differentiate gene expression differences due to chimeric isoform expression. We then focused on differentially expressed genes that are expressed in a given strain, but silent in the other (log2 normalized counts < 1). Only two chimeric cases fell into such category: TE-exonized CR45600 – *Tc1-2; gypsy7* in dmgoth101 compared to dmsj7 (Figure 9A); and the TE-terminated CG30428 – *INE-1* in dmgoth101, dmsj23, and dmsj7 compared to dmgoth63 (Figure 9A, C, D). To verify whether the chimeric transcript has a relatively high contribution to the gene expression, we compared the depth of uniquely mapped reads between exons and the TE insertion. For the CR45600 – *Tc1-2; gypsy7* chimera, we found a similar depth (Supplementary Figure 13A), indicating that the chimeric transcript isoform represents a major contribution to the gene expression; whereas for CG30428 – *INE-1* chimera, the depth in *INE-1* is half of CG30428 exons (Supplementary Figure 13B), suggesting that the TE sequence contributes with only ∼50% of gene expression.

In dmgoth101, we found the gene *Cyp6a2* differentially expressed and involved in a TE-exonized transcript with a *17.6* insertion. Such chimera is absent from dmsj23, dmsj7, and dmgoth63 transcriptome. We checked whether *17.6* could be present on the genome of the other strains and we observed it only in dmgoth63, but undetected by ChimeraTE because *Cyp6a2* is not expressed. The presence of the *17.6* insertion within the *Cyp6a2* 3′UTR generating a chimeric transcript was initially proposed to cause the overexpression of this gene in flies resistant to xenobiotics (108), but later the opposite has been shown (109). Furthermore, regardless of *17.6* presence, the overexpression of this gene has been associated with xenobiotic resistance in *D. melanogaster* strains (110). Here, the differential expression analysis reveals that *Cyp6a2* is up-regulated in dmgoth101 in comparison to dmgoth63 (adj. *P*-value: 0.005; $\log_2$ FC: 4.38), but no significant differences were found when comparing with strains without the *17.6* insertion, dmsj23 (adj. *P*-value: 0.29), and dmsj7 (adj. *P* value: 0.15), reinforcing the lack of contribution of *17.6* insertion to *Cyp6a2* expression (109). Finally, the chimera derived from *Cyp6a14*, with a *1360* insertion embedded into the 3′ UTR has been found only in dmsj7, in comparison to the other three strains. This gene is associated with xenobiotic resistance (111), as *Cyp6a2*. Interestingly, our results reinforce previous findings showing that *Cyp* genes have accumulated more TE insertions when compared to a random sample of genes (112), potentially as a source of alternative regulatory motifs.

Collectively, the presence of chimeric transcripts might not drive gene differential expression. Despite eight differentially expressed genes per strain, the transcriptional activation of specific genes in one genotype may be determined by its genetic background, which can be associated with structural variants as TEs, but also includes pleiotropic effects from regulatory networks, cell physiology and epigenetics (113). Nevertheless, given that TEs might provide TFBSs and epigenetic marks to chromatin accessibility, we propose these differentially expressed genes as candidates to investigate the domestication of regulatory networks in further studies, especially the ones with TEs inserted at regulatory regions.

ChimeraTE uncovers chimeric transcripts in model and non-model species

We sought to demonstrate that both Modes of ChimeraTE are able to identify chimeric transcripts regardless of the species. Therefore, we used ChimeraTE with human, *A. thaliana* and *P. reticulata* (guppy fish) data. In Mode 1, the RNA-seq of the human K562 cell line revealed 67 TE-initiated, 14,999 TE-exonized, and 166 TE-terminated transcripts, derived from 5,170 genes (10.45% of total). Furthermore, Mode 2 revealed 4,372 chimeric transcripts. From these, 3,405 chimeras were detected from chimeric reads, 342 from transcriptome assembly, and 625 from double evidence. The overlap from chimeras found by both Modes was 2,997 (68.55%), suggesting a high amount of chimeric transcripts derived from TE insertions that are not present in the reference genome, or from TEs located farther away than the chosen 3 kb gene flanking regions. In *A. thaliana*, we used ChimeraTE with RNA-seq from leaf tissue. Our results from Mode 1 demonstrated that 266 genes (0.7% out of 38,319) generate 397 chimeras, corresponding to 24 TE-initiated, 338 TE-exonized, and 35 TE-terminated transcripts. Mode 2 has detected 141 genes generating chimeras, for which 63.12% were from chimeric reads, 24.11% from transcriptome assembly, and 12.77% from double evidence. Similar to human, Mode 2 found 39.72% of chimeras detected by Mode 1. Finally, the guppy transcriptome was analyzed with RNA-seq produced from ovary follicular tissue. For the first time, we identified 1,151 genes (5.85% of total genes) generating chimeric transcripts in this species. From these, 39 were TE-initiated, 1,507 TE-exonized, and 136 TE-terminated transcripts. Mode 2 revealed 640 genes with chimeras, having an overlap with 318 genes (49.69%) detected by Mode 1. Taking together our results from human, *A. thaliana*, guppy and *D. melanogaster* data, we demonstrated that ChimeraTE can identify chimeras in different species. We then analyzed how the genome size could affect ChimeraTE Mode 1 and Mode 2 in the processing time, and we have observed a positive correlation between them in

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

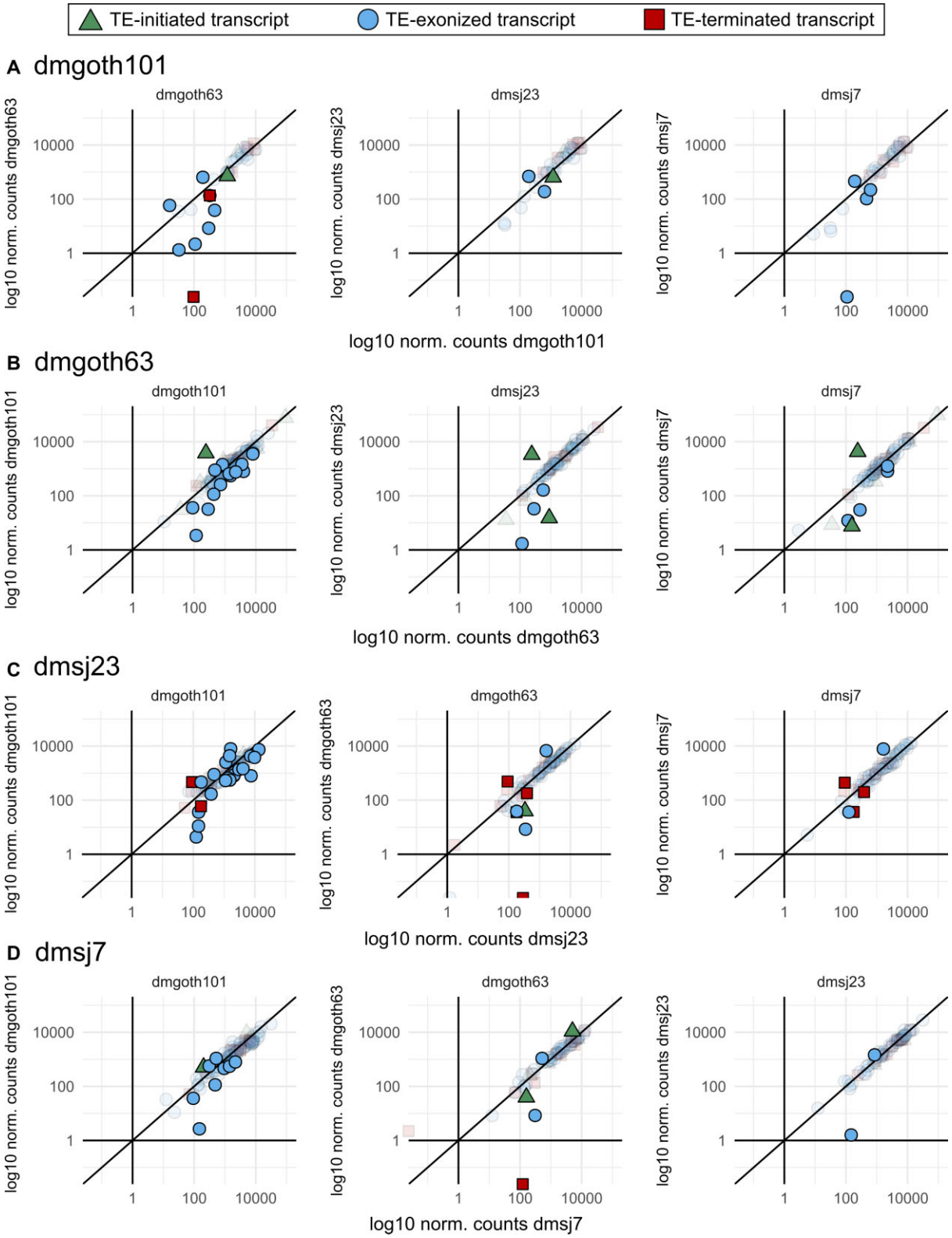


Figure 9. Normalized counts from DEseq2 indicating gene expression of genes generating polymorphic chimeric transcripts among the four wild-type strains. Colorful forms represent differentially expressed genes (adj. p-value < 0.05); transparent forms represent non-differentially expressed genes (adj. p-value). (**A**) Expression level of genes producing chimeric transcripts only in dmgoth101, in comparison to dmgoth63, dmsj23, and dmsj7. (**B**) Expression level of genes producing chimeric transcripts only in dmgoth63, in comparison to dmgoth101, dmsj23, and dmsj7. (**C**) Expression level of genes producing chimeric transcripts only in dmsj23, in comparison to dmgoth101, dmgoth63, and dmsj7. (**D**) Expression level of genes producing chimeric transcripts only in dmsj7, in comparison to dmgoth101, dmgoth63, and dmsj23. Overall, gene expression does not change when comparing strains with and without chimeric transcripts.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Mode 1 (Pearson; $r = 0.99$; $P = 0.01$), as well as for in Mode 2 (Pearson; $r = 0.99$; $P = 0.009$) (Supplementary Figure 14).

CONCLUSION

In the last decades, RNA-seq has provided the opportunity to understand transcriptome plasticity, which can lead to phenotypic divergence, from related species or individuals from a population (114). Among several sources of modification in gene expression and novel isoform transcripts, TEs have been considered fundamental suppliers of transcriptome plasticity, participating in gene networks and incorporating either regulatory sequences or protein domains into gene transcripts. The identification of chimeric transcripts is an important step to the understanding of transcriptome plasticity, since they may be triggered by ectopic conditions, such as cancer, oxidative stress, and heat shock (26,115,116), which may lead to both detrimental and advantageous outcomes (117). Therefore, uncovering the extent of chimeric transcripts between individual/cell/strain transcriptomes is a crucial first step to investigate potential exaptation/domestication events, or gene disruption and loss of function.

Chimeric transcripts have been identified more recently by different methods exploiting RNA-seq data (34–36,81), but none of them provided the possibility to predict chimeras from TEs that are absent from reference genomes. Here, we developed ChimeraTE, a pipeline able to identify chimeric transcripts from TEs that are absent from the reference genome. Mode 1 is a genome-guided approach and may be used either when the user is not interested in chimeras derived from TEs absent from the reference genome, or when the user has a high-quality genome assembly for each individual/cell/strain; whereas Mode 2 is a genome-blind approach, with the ability to predict chimeric transcripts without the assembled genome, but missing chimeras where the TE is smaller than the length of the reads.

We analyzed ovarian RNA-seq from four *D. melanogaster* wild-type strains, for which we have genome assemblies. Altogether, we found that ∼1.12% of all genes are generating chimeric transcripts in ovaries, following the proportion from previous studies obtained with ESTs, and RNA-seq in midbrain tissue (31,34). Furthermore, our results revealed that 50% of all TE-exonized transcripts with TEs embedded derive from *roo* elements, and most specifically, a small region between tandem repeats in the 5′ UTR and the beginning of the *roo* ORF. These results suggest that these *roo* insertions have neutral or advantageous effects as they are maintained within these gene transcripts. However, we did not provide enough support to claim these *roo*-exonized transcripts as exaptation or domestication events, due to the lack of evidence regarding the functional role of these chimeras. Further studies must be performed to clarify this subject, mainly because chimeric transcripts detected by ChimeraTE can be degraded by surveillance pathways that degrade aberrant mRNA, such as non-sense-mediated mRNA decay (118), no-go decay (119), and non-stop decay (120).

Altogether, this new approach allows studying the impact of new mobilization events between populations or between treatment conditions, providing insights into biological questions from a broad community of researchers, ranging from cancer research, to population transcriptomics, and adaptation studies. ChimeraTE implementation will be useful for the next discoveries regarding the evolutionary role of TEs and their impact on the host transcriptome.

DATA AVAILABILITY

ChimeraTE is an open-source collaborative initiative available in the Zenodo repository (https://doi.org/10.5281/zenodo.8143045). The RNA-seq data used in this study are available in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/), under PRJNA795668 accession, while the long-read ONT data is available under PRJNA956863.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Josefa González, Marta Coronado-Zamora, Dixie Mager and Vincent Lacroix for useful discussions and advice. This work was performed using the computing facilities of the CC LBBE/PRABI.

FUNDING

Agence Nationale de la Recherche [Exhyb ANR-14-CE19-0016-01 and LongevitY ANR-20-CE02-0015 to C.V.]; Fondation pour la Recherche Médicale [DEP20131128536 to C.V.]; Idex Lyon fellowship to D.S.O.; Campus France Eiffel [Eiffel Fellowship P769649C to D.S.O.]; TIGER [H2020-MSCA-IF-2014-658726 to R.R.]; National Council for Scientific and Technological Development [308020/2021-9 to C.M.A.C.]; São Paulo Research Foundation [2020/06238-2 to C.M.A.C.]. Funding for open access charge: Fondation pour la Recherche Médicale [DEP20131128536] and Agence Nationale de la Recherche [ExHyb, ANR-14-CE19-0016-01].

Conflict of interest statement. None declared.

REFERENCES

1. Quesneville,H., Bergman,C.M., Andrieu,O., Autard,D., Nouaud,D., Ashburner,M. and Anxolabehere,D. (2005) Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput. Biol.*, **1**, e22.
2. International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, Center for Genome Research, Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
3. Schnable,P.S., Ware,D., Fulton,R.S., Stein,J.C., Wei,F., Pasternak,S., Liang,C., Zhang,J., Fulton,L., Graves,T.A. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.
4. Sotero-Caio,C.G., Platt,R.N., Suh,A. and Ray,D.A. (2017) Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evolut.*, **9**, 161–177.
5. Danilevskaya,O.N., Arkhipova,I.R., Pardue,M.L. and Traverse,K.L. (1997) Promoting in tandem: the promoter for telomere transposon HeT-A and implications for the evolution of retroviral ltrs. *Cell*, **88**, 647–655.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

6. Kapitonov,V.V. and Jurka,J. (2005) RAG1 Core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.*, **3**, e181.
7. Kapitonov,V.V. and Koonin,E.V. (2015) Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. *Biol. Direct*, **10**, 20.
8. Volff,J.-N. (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays*, **28**, 913–922.
9. Babaian,A., Romanish,M.T., Gagnier,L., Kuo,L.Y., Karimi,M.M., Steidl,C. and Mager,D.L. (2016) Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. *Oncogene*, **35**, 2542–2546.
10. Daborn,P.J., Yen,J.L., Bogwitz,M.R., Le Goff,G., Feil,E., Jeffers,S., Tijet,N., Perry,T., Heckel,D., Batterham,P. *et al.* (2002) A single P450 allele associated with insecticide resistance in *Drosophila*. *Science*, **297**, 2253–2256.
11. Jordan,I.K., Rogozin,I.B., Glazko,G.V. and Koonin,E.V. (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.*, **19**, 68–72.
12. Mateo,L., Ullastres,A. and González,J. (2014) A transposable element insertion confers xenobiotic resistance in *Drosophila*. *PLos Genet.*, **10**, e1004560.
13. Modzelewski,A.J., Shao,W., Chen,J., Lee,A., Qi,X., Noon,M., Tjokro,K., Sales,G., Biton,A., Anand,A. *et al.* (2021) A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. *Cell*, **184**, 5541–5558.
14. Capy,P. (2021) Taming, domestication and exaptation: trajectories of transposable elements in genomes. *Cells*, **10**, 3590.
15. Fueyo,R., Judd,J., Feschotte,C. and Wysocka,J. (2022) Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.*, **23**, 481–497.
16. Lamprecht,B., Walter,K., Kreher,S., Kumar,R., Hummel,M., Lenze,D., Köchert,K., Bouhlel,M.A., Richter,J., Soler,E. *et al.* (2010) Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.*, **16**, 571–579.
17. McGinnis,W., Shermoen,A.W. and Beckendorf,S.K. (1983) A transposable element inserted just 5′ to a *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell*, **34**, 75–84.
18. Almeida,L.M., Amaral,M.E.J., Silva,I.T., Silva,W.A. Jr, Riggs,P.K. and Carareto,C.M. (2008) Report of a chimeric origin of transposable elements in a bovine-coding gene. *Genet. Mol. Res.*, **7**, 107–116.
19. Sela,N., Mersch,B., Hotz-Wagenblatt,A. and Ast,G. (2010) Characteristics of transposable element exonization within human and mouse. *PLoS One*, **5**, e10907.
20. Sorek,R. (2007) The birth of new exons: mechanisms and evolutionary consequences. *RNA*, **13**, 1603–1608.
21. Bogwitz,M.R., Chung,H., Magoc,L., Rigby,S., Wong,W., O'Keefe,M., McKenzie,J.A., Batterham,P. and Daborn,P.J. (2005) Cyp12a4 confers lufenuron resistance in a natural population of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 12807–12812.
22. Farré,D., Engel,P. and Angulo,A. (2016) Novel role of 3'UTR-embedded Alu elements as facilitators of processed pseudogene genesis and host gene capture by viral genomes. *PLoS One*, **11**, e0169196.
23. Magwire,M.M., Bayer,F., Webster,C.L., Cao,C. and Jiggins,F.M. (2011) Successive increases in the resistance of *Drosophila* to viral infection through a transposon insertion followed by a duplication. *PLoS Genet.*, **7**, e1002337.
24. Cordaux,R., Udit,S., Batzer,M.A. and Feschotte,C. (2006) Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 8101–8106.
25. Ehrlich,M. (2009) DNA hypomethylation in cancer cells. *Epigenomics*, **1**, 239–259.
26. Babaian,A. and Mager,D.L. (2016) Endogenous retroviral promoter exaptation in human cancer. *Mobile DNA*, **7**, 24.
27. Lock,F.E., Rebollo,R., Miceli-Royer,K., Gagnier,L., Kuah,S., Babaian,A., Sistiaga-Poveda,M., Lai,C.B., Nemirovsky,O., Serrano,I. *et al.* (2014) Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E3534–E3543.
28. Shah,N.M., Jang,H.J., Liang,Y., Maeng,J.H., Tzeng,S.-C., Wu,A., Basri,N.L., Qu,X., Fan,C., Li,A. *et al.* (2023) Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements. *Nat. Genet.*, **55**, 631–639.
29. Faulkner,G.J., Kimura,Y., Daub,C.O., Wani,S., Plessy,C., Irvine,K.M., Schroder,K., Cloonan,N., Steptoe,A.L., Lassmann,T. *et al.* (2009) The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.*, **41**, 563–571.
30. Babarinde,I.A., Ma,G., Li,Y., Deng,B., Luo,Z., Liu,H., Abdul,M.M., Ward,C., Chen,M., Fu,X. *et al.* (2021) Transposable element sequence fragments incorporated into coding and noncoding transcripts modulate the transcriptome of human pluripotent stem cells. *Nucleic Acids Res.*, **49**, 9132–9153.
31. Lipatov,M., Lenkov,K., Petrov,D.A. and Bergman,C.M. (2005) Paucity of chimeric gene-transposable element transcripts in the *Drosophila melanogaster* genome. *BMC Biol.*, **3**, 24.
32. Batut,P. and Gingeras,T.R. (2013) RAMPAGE: promoter activity profiling by paired-end sequencing of 5′-complete cDNAs. *Curr. Protoc. Mol. Biol.*, **104**, 25.
33. Batut,P., Dobin,A., Plessy,C., Carninci,P. and Gingeras,T.R. (2013) High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res.*, **23**, 169–180.
34. Treiber,C.D. and Waddell,S. (2020) Transposon expression in the *Drosophila* brain is driven by neighboring genes and diversifies the neural transcriptome. *Genome Res.*, **30**, 1559–1569.
35. Pinson,M.-E., Pogorelcnik,R., Court,F., Arnaud,P. and Vaurs-Barrière,C. (2018) CLIFinder: identification of LINE-1 chimeric transcripts in RNA-seq data. *Bioinformatics*, **34**, 688–690.
36. Babaian,A., Thompson,I.R., Lever,J. and Gagnier,L. (2019) LIONS: analysis suite for detecting and quantifying transposable element initiated transcription from RNA-seq. *Bioinformatics*, **35**, 3839–3841.
37. Kircher,M., Sawyer,S. and Meyer,M. (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.*, **40**, e3.
38. Evrony,G.D., Lee,E., Park,P.J. and Walsh,C.A. (2016) Resolving rates of mutation in the brain using single-neuron genomics. *eLife*, **5**, e12966.
39. Quail,M.A., Kozarewa,I., Smith,F., Scally,A., Stephens,P.J., Durbin,R., Swerdlow,H. and Turner,D.J. (2008) A large genome center's improvements to the Illumina sequencing system. *Nat. Methods*, **5**, 1005–1010.
40. Treiber,C.D. and Waddell,S. (2017) Resolving the prevalence of somatic transposition in *Drosophila*. *eLife*, **6**, e28297.
41. Martin Cerezo,M.L., Raval,R., de Haro Reyes,B., Kucka,M., Chan,F.Y. and Bryk,J. (2022) Identification and quantification of chimeric sequencing reads in a highly multiplexed RAD-seq protocol. *Mol. Ecol. Resour.*, **22**, 2860–2870.
42. Trapnell,C., Roberts,A., Goff,L., Pertea,G., Kim,D., Kelley,D.R., Pimentel,H., Salzberg,S.L., Rinn,J.L. and Pachter,L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.*, **7**, 562–578.
43. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and and 1000 Genome Project Data Processing Subgroupand 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and samtools. *Bioinformatics*, **25**, 2078–2079.
44. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
45. Dale,R.K., Pedersen,B.S. and Quinlan,A.R. (2011) Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics*, **27**, 3423–3424.
46. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
47. Roberts,A. and Pachter,L. (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat. Methods*, **10**, 71–73.
48. Grabherr,M.G., Haas,B.J., Yassour,M., Levin,J.Z., Thompson,D.A., Amit,I., Adiconis,X., Fan,L., Raychowdhury,R., Zeng,Q. *et al.* (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.*, **29**, 644–652.
49. SMIT, Arian FA (2004) RepeatMasker Open 3.0.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

50. Storer,J., Hubley,R., Rosen,J., Wheeler,T.J. and Smit,A.F. (2021) The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA*, **12**, 2.
51. Altschul,S. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
52. Fablet,M., Salcez-Ortiz,J., Jacquet,A., Menezes,B.F., Dechaud,C., Veber,P., Noûs,C., Rebollo,R. and Vieira,C. (2022) A quantitative, genome-wide analysis in Drosophila reveals transposable elements' influence on gene expression is species-specific. bioRxiv doi: https://doi.org/10.1101/2022.01.20.477049, 22 January 2022, preprint: not peer reviewed.
53. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
54. Mohamed,M., Dang,N.T.-M., Ogyama,Y., Burlet,N., Mugat,B., Boulesteix,M., Mérel,V., Veber,P., Salces-Ortiz,J., Severac,D. *et al.* (2020) A transposon story: from TE content to TE dynamic invasion of *drosophila* genomes using the single-molecule sequencing technology from Oxford Nanopore. *Cells*, **9**, 1776.
55. Shumate,A. and Salzberg,S.L. (2021) Liftoff: accurate mapping of gene annotations. *Bioinformatics*, **37**, 1639–1643.
56. Wicker,T., Sabot,F., Hua-Van,A., Bennetzen,J.L., Capy,P., Chalhoub,B., Flavell,A., Leroy,P., Morgante,M., Panaud,O. *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.*, **8**, 973–982.
57. Lerat,E., Rizzon,C. and Biémont,C. (2003) Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. *Genome Res.*, **13**, 1889–1896.
58. Flutre,T., Duprat,E., Feuillet,C. and Quesneville,H. (2011) Considering transposable element diversification in de novo annotation approaches. *PLoS One*, **6**, e16526.
59. Flutre,T., Permal,E. and Quesneville,H. (2012) Transposable element annotation in completely sequenced eukaryote genomes. In: Grandbastien,M-.A. and Casacuberta,J.M. (eds.) *Plant Transposable Elements, Topics in Current Genetics*. Springer, Berlin Heidelberg, Vol. **24**, pp. 17–39.
60. Benson,G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, **27**, 573–580.
61. dos Santos,G., Schroeder,A.J., Goodman,J.L., Strelets,V.B., Crosby,M.A., Thurmond,J., Emmert,D.B., Gelbart,W.M. and the FlyBase Consortiumthe FlyBase Consortium (2015) FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.*, **43**, D690–D697.
62. Lee,C.M., Barber,G.P., Casper,J., Clawson,H., Diekhans,M., Gonzalez,J.N., Hinrichs,A.S., Lee,B.T., Nassar,L.R., Powell,C.C. *et al.* (2019) UCSC Genome Browser enters 20th year. *Nucleic Acids Res.*, **48**, D756–D761.
63. Thorvaldsdottir,H., Robinson,J.T. and Mesirov,J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings Bioinf.*, **14**, 178–192.
64. Rice,P., Longden,I. and Bleasby,A.EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
65. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L.L., Tosatto,S.C.E., Paladin,L., Raj,S., Richardson,L.J. *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
66. Paysan-Lafosse,T., Blum,M., Chuguransky,S., Grego,T., Pinto,B.L., Salazar,G.A., Bileschi,M.L., Bork,P., Bridge,A., Colwell,L. *et al.* (2023) InterPro in 2022. *Nucleic Acids Res.*, **51**, D418–D427.
67. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
68. Zea,D.J., Anfossi,D., Nielsen,M. and Marino-Buslje,C. (2017) MIToS.Jl: mutual information tools for protein sequence analysis in the Julia language. *Bioinformatics*, **33**, 564–565.
69. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
70. Lu,Z., Marand,A.P., Ricci,W.A., Ethridge,C.L., Zhang,X. and Schmitz,R.J. (2019) The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat. Plants*, **5**, 1250–1259.
71. Bakoulis,S., Krautz,R., Alcaraz,N., Salvatore,M. and Andersson,R. (2022) Endogenous retroviruses co-opted as divergently transcribed regulatory elements shape the regulatory landscape of embryonic stem cells. *Nucleic Acids Res.*, **50**, 2111–2127.
72. Annala,M.J., Parker,B.C., Zhang,W. and Nykter,M. (2013) Fusion genes and their discovery using high throughput sequencing. *Cancer Lett.*, **340**, 192–200.
73. Kanagawa,T. (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.*, **96**, 317–323.
74. Houseley,J. and Tollervey,D. (2010) Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. *PLoS One*, **5**, e12271.
75. Van Der Valk,T., Vezzi,F., Ormestad,M., Dalén,L. and Guschanski,K. (2020) Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. *Mol. Ecol. Resour.*, **20**, 1171–1181.
76. Mitelman,F., Johansson,B. and Mertens,F. (2007) The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer*, **7**, 233–245.
77. Zhou,Y., Zhang,C., Zhang,L., Ye,Q., Liu,N., Wang,M., Long,G., Fan,W., Long,M. and Wing,R.A. (2022) Gene fusion as an important mechanism to generate new genes in the genus Oryza. *Genome Biol.*, **23**, 130.
78. Kaminker,J.S., Bergman,C.M., Kronmiller,B., Svirskas,R., Patel,S., Frise,E., Lewis,S.E., Rubin,G.M., Ashburner,M. and Celniker,S.E. (2002) The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.*, **3**, RESEARCH0084.
79. Gatti,M. and Pimpinelli,S. (1992) Functional elements in *Drosophila melanogaster* heterochromatin. *Annu. Rev. Genet.*, **26**, 239–276.
80. Carmena,M. and Gonzfilez,C. (1995) Transposable elements map in a conserved pattern of distribution extending from beta-heterochromatin to centromeres in *Drosophila melanogaster*. *Chromosoma*, **103**, 676–684.
81. Coronado-Zamora,M. and González,J. (2022) Transposons contribute to the functional diversification of the head, gut, and ovary transcriptomes across *Drosophila* natural strains. bioRxiv doi: https://doi.org/10.1101/2022.12.02.518890, 04 December 2022, preprint: not peer reviewed.
82. Bailly-Bechet,M., Haudry,A. and Lerat,E. (2014) "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. *Mobile DNA*, **5**, 13.
83. Kapun,M., Barrón,M.G., Staubach,F., Obbard,D.J., Wiberg,R.A.W., Vieira,J., Goubert,C., Rota-Stabelli,O., Kankare,M., Bogaerts-Márquez,M. *et al.* (2020) Genomic analysis of european *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. *Mol. Biol. Evol.*, **37**, 2661–2678.
84. Rech,G.E., Bogaerts-Márquez,M., Barrón,M.G., Merenciano,M., Villanueva-Cañas,J.L., Horváth,V., Fiston-Lavier,A.-S., Luyten,I., Venkataram,S., Quesneville,H. *et al.* (2019) Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila*. *PLoS Genet.*, **15**, e1007900.
85. Rech,G.E., Radío,S., Guirao-Rico,S., Aguilera,L., Horvath,V., Green,L., Lindstadt,H., Jamilloux,V., Quesneville,H. and González,J. (2022) Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. *Nat. Commun.*, **13**, 1948.
86. Lima,L., Sinaimeri,B., Sacomoto,G., Lopez-Maestre,H., Marchet,C., Miele,V., Sagot,M.F. and Lacroix,V. (2017) Playing hide and seek with repeats in local and global de novo transcriptome assembly of short RNA-seq reads. *Algorith. Mol. Biol.*, **12**, 2.
87. Lahens,N.F., Kavakli,I.H., Zhang,R., Hayer,K., Black,M.B., Dueck,H., Pizarro,A., Kim,J., Irizarry,R., Thomas,R.S. *et al.* (2014) IVT-seq reveals extreme bias in RNA sequencing. *Genome Biol.*, **15**, 15.
88. Wang,S. and Gribskov,M. (2016) Comprehensive evaluation of *de novo* transcriptome assembly programs and their effects on differential gene expression analysis. *Bioinformatics*, **33**, 327–333.
89. Rebollo,R., Cumunel,E., Mary,A., Burlet,N., Gillet,B., Hughes,S., Oliveira,D.S., Goubert,C., Fablet,M., Vieira,C. *et al.* (2023) Detection and identification of transposable element transcripts using Long Read RNA-seq in Drosophila germline tissues. bioRxiv

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

doi: https://doi.org/10.1101/2023.05.27.542554, 30 May 2023, preprint: not peer reviewed.
90. Ura,H., Togi,S. and Niida,Y. (2022) A comparison of mRNA sequencing (RNA-Seq) library preparation methods for transcriptome analysis. *Bmc Genomics [Electronic Resource]*, **23**, 303.
91. Cridland,J.M., Macdonald,S.J., Long,A.D. and Thornton,K.R. (2013) Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources. *Mol. Biol. Evol.*, **30**, 2311–2327.
92. Lerman,D.N. and Feder,M.E. (2005) Naturally occurring transposable elements disrupt hsp70 promoter function in *Drosophila melanogaster*. *Mol. Biol. Evol.*, **22**, 776–783.
93. Nefedova,L.N., Kuzmin,I.V., Makhnovskii,P.A. and Kim,A.I. (2014) Domesticated retroviral GAG gene in *Drosophila*: new functions for an old gene. *Virology*, **450–451**, 196–204.
94. Malik,H.S. and Henikoff,S. (2005) Positive selection of Iris, a retroviral envelope–Derived host gene in *Drosophila melanogaster*. *PLoS Genet.*, **1**, e44.
95. Rahman,R., Chirn,G., Kanodia,A., Sytnikova,Y.A., Bergman,C.M. and Lau,N.C. (2015) Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes. *Nucleic Acids Res.*, **43**, 10655–10672.
96. Vieira,C. and Biemont,C. (2004) Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *drosophila simulans*. *Genetica*, **120**, 115–123.
97. Merenciano,M., Ullastres,A., de Cara,M.A.R., Barrón,M.G. and González,J. (2016) Multiple independent retroelement insertions in the promoter of a stress response gene have variable molecular and functional effects in *Drosophila*. *PLoS Genet.*, **12**, e1006249.
98. Rebollo,R., Horard,B., Begeot,F., Delattre,M., Gilson,E. and Vieira,C. (2012) A snapshot of histone modifications within transposable elements in *Drosophila* wild type strains. *PLoS One*, **7**, e44253.
99. Yasuhara,J.C. and Wakimoto,B.T. (2008) Molecular landscape of modified histones in *Drosophila* heterochromatic genes and euchromatin-heterochromatin transition zones. *PLoS Genet.*, **4**, e16.
100. Díaz-González,J., Vázquez,J.F., Albornoz,J. and Domínguez,A. (2011) Long-term evolution of the *roo* transposable element copy number in mutation accumulation lines of *Drosophila melanogaster*. *Genetics Res.*, **93**, 181–187.
101. Díaz-González,J., Domínguez,A. and Albornoz,J. (2010) Genomic distribution of retrotransposons 297, 1731, copia, mdg1 and roo in the *Drosophila melanogaster* species subgroup. *Genetica*, **138**, 579–586.
102. Díaz-González,J. and Domínguez,A. (2020) Different structural variants of roo retrotransposon are active in *Drosophila melanogaster*. *Gene*, **741**, 144546.
103. Vieira,C., Lepetit,D., Dumont,S. and Biemont,C. (1999) Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol. Biol. Evol.*, **16**, 1251–1255.
104. Barrón,M.G., Fiston-Lavier,A.-S., Petrov,D.A. and González,J. (2014) Population genomics of transposable elements in *Drosophila*. *Annu. Rev. Genet.*, **48**, 561–581.
105. Kapitonov,V.V. and Jurka,J. (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 6569–6574.
106. Kofler,R., Betancourt,A.J. and Schlötterer,C. (2012) Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet.*, **8**, e1002487.
107. Maddison,W.P. and Knowles,L.L. (2006) Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.*, **55**, 21–30.
108. Waters,L.C. (1992) Possible involvement of the long terminal repeat of transposable element 17.6 in regulating expression of an insecticide resistance-associated P450 gene in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 4855–4859.
109. Delpuech,J.M., Aquadro,C.F. and Roush,R.T. (1993) Noninvolvement of the long terminal repeat of transposable element 17.6 in insecticide resistance in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 5643–5647.
110. Brun,A., Cuany,A., Le Mouel,T., Berge,J. and Amichot,M. (1996) Inducibility of the *Drosophila melanogaster* cytochrome P450 gene, CYP6A2, by phenobarbital in insecticide susceptible or resistant strains. *Insect Biochem. Mol. Biol.*, **26**, 697–703.
111. Seong,K.M., Coates,B.S., Berenbaum,M.R., Clark,J.M. and Pittendrigh,B.R. (2018) Comparative CYP-omic analysis between the DDT-susceptible and -resistant *Drosophila melanogaster* strains *91-C* and *91-R*. *Pest Manage. Sci.*, **74**, 2530–2543.
112. Carareto,C.M.A., Hernandez,E.H. and Vieira,C. (2014) Genomic regions harboring insecticide resistance-associated Cyp genes are enriched by transposable element fragments carrying putative transcription factor binding sites in two sibling *Drosophila* species. *Gene*, **537**, 93–99.
113. Hill,M.S., Vande Zande,P. and Wittkopp,P.J. (2021) Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.*, **22**, 203–215.
114. Marguerat,S. and Bähler,J. (2010) RNA-seq: from technology to biology. *Cell. Mol. Life Sci.*, **67**, 569–579.
115. Oliveira,D.S., Rosa,M.T., Vieira,C. and Loreto,E.L.S. (2021) Oxidative and radiation stress induces transposable element transcription in *Drosophila melanogaster*. *J. Evol. Biol.*, **34**, 628–638.
116. Horváth,V., Merenciano,M. and González,J. (2017) Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends Genet.*, **33**, 832–841.
117. Nicolau,M., Picault,N. and Moissiard,G. (2021) The evolutionary volte-face of transposable elements: from harmful jumping genes to major drivers of genetic innovation. *Cells*, **10**, 2952.
118. Hug,N., Longman,D. and Cáceres,J.F. (2016) Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.*, **44**, 1483–1495.
119. Harigaya,Y. and Parker,R. (2010) No-go decay: a quality control mechanism for RNA in translation. *WIREs RNA*, **1**, 132–141.
120. Vasudevan,S., Peltz,S.W. and Wilusz,C.J. (2002) Non-stop decay—a new mRNA surveillance pathway. *Bioessays*, **24**, 785–788.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad671/7249921 by guest on 13 October 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

GBE

Exploring the Relationship of Transposable Elements and Ageing: Causes and Consequences

Miriam Merenciano [1], Anaïs Larue [1,2], Chloé Garambois [1], William Vilas Boas Nunes [1], Cristina Vieira [1,*]

[1]Laboratoire de Biométrie et Biologie Evolutive, CNRS, UMR5558, Université Claude Bernard Lyon 1, Villeurbanne 69100, France
[2]INRAE, BF2I, UMR203, INSA Lyon, Villeurbanne 69621, France

*Corresponding author: E-mail: cristina.vieira@univ-lyon1.fr.

Accepted: May 02, 2025

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Abstract

Ageing is a gradual biological process marked by a decline in physiological function, increasing susceptibility to disease, and mortality. Transposable elements (TEs) are repetitive DNA sequences capable of moving within the genome and thus potentially inducing mutations and disrupting normal cellular functions. Their mobile nature contributes to genomic variation, as transposition events can alter gene expression, chromosome structure, and the epigenetic landscape. To mitigate TE-induced damage, cells rely on epigenetic mechanisms, such as DNA methylation, histone modifications, and small RNAs, to repress TE activity. However, these silencing mechanisms become less effective with age, leading to increased TE activation. This review explores the dual role of TEs as both a cause and consequence of ageing, suggesting a complex relationship between TEs and the ageing process.

Key words: senescence, transposable elements, chromatin, age-related diseases, long-lived organisms.

Significance

Transposable elements (TEs) are highly mutagenic repetitive DNA sequences that have long been understudied due to technical challenges in detecting and analyzing them. However, the emergence of long-read sequencing techniques has greatly advanced TE research, providing deeper insights into their genomic impact. Growing evidence now links TE activity to ageing and age-related diseases, suggesting that TEs can influence lifespan. In this review, we highlight recent discoveries showing that TE activation can be both a cause and a consequence of ageing, emphasizing the importance of studying these genomic elements.

Introduction

Ageing is of particular interest in biology, especially in understanding the underlying mechanisms that help elucidating patterns of longevity. Ageing is a gradual biological process characterized by a decline in physiological function and an increased susceptibility to disease and mortality over time (Kirkwood and Austad 2000; López-Otín et al. 2013). As organisms age, the accumulation of senescent cells (cells that have lost the ability to divide) contributes to progressive dysfunction at the organismal level, ultimately impairing survival and fertility.

The evolutionary theory of ageing suggests that the strength of natural selection, a measure of how effectively selection influences survival or reproduction, diminishes with age (Haldane 1941; Medawar 1946, 1952; Williams 1957). This decline forms the basis for hypotheses such as mutation accumulation and antagonistic pleiotropy. The mutation accumulation hypothesis posits that deleterious

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

mutations expressed later in life persist in populations because natural selection is less effective at removing them when reproductive success is no longer significantly impacted (Medawar 1946, 1952). In contrast, the antagonistic pleiotropy hypothesis suggests that some genes have dual effects, offering benefits during early life, such as enhanced reproductive fitness, but causing negative effects in later life, contributing to ageing (Williams 1957). This trade-off arises because natural selection favours traits that enhance early-life fitness, even at the cost of late-life detriments.

Modern theories of ageing, which seek to explain its underlying mechanisms, are divided into two main categories: the error/damage and the programmed perspective (Jin 2010). The error/damage model proposes that ageing results primarily from the accumulation of cellular and molecular damage over time. This theory emphasises that environmental factors, lifestyle choices, and metabolic processes contribute to this damage (Weismann 1882; Davidovic et al. 2010; Jin 2010). In contrast, the programmed model views ageing as an inherent and essential part of the life cycle, driven by genetic and hormonal mechanisms rather than simply being a consequence of accumulated damage over time (Davidovic et al. 2010; Jin 2010).

Advances in whole-genome sequencing techniques have enabled the study of the genetic mechanisms involved in ageing. Among various genomic components, transposable element (TE) effects have been repeatedly linked to ageing due to their capacity to generate mutations with potential to disrupt normal cellular functions (Wood and Helfand 2013; Copley and Shorter 2023; Yushkova and Moskalev 2023). TEs are repetitive DNA sequences capable of moving (transpose) within the genome, which are commonly classified into two main classes based on their mechanism of transposition (Bourque et al. 2018). Class I elements, or retrotransposons, transpose *via* an RNA intermediate through a "copy and paste" mechanism. In contrast, Class II elements, or DNA transposons, move using a DNA intermediate and typically follow a "cut and paste" mechanism (Bourque et al. 2018). To capture the diversity within these broad categories, TEs are further classified in subclasses, orders, and superfamilies based on mechanistic and enzymatic criteria (Wicker et al. 2007). TEs are present in virtually all eukaryotic and prokaryotic genomes (Hua-Van et al. 2011; Kissinger and DeBarry 2011), and they typically represent a considerable fraction of the genomes, although their abundance is highly variable from one species to another (Guio and González 2019; Mérel et al. 2020). Due to their mobile and repetitive nature, TEs are a source of genomic variation. The DNA breaks and insertions associated with transposition events lead to obvious alterations to the genome (Bourque et al. 2018). The consequences of TE expression and mobilization can also have widespread effects, altering gene expression and structure, chromosome dynamics, as well as the epigenetic landscape of the genome (Casacuberta and González 2013; Bourque et al. 2018; Bourgeois and Boissinot 2019). The idea that TEs can contribute to ageing processes through mutations (associated with the error/damage theory of ageing) was first proposed by Shmookler Reis and Goldstein (1983) and Vijg et al. (1985). Building up on the same idea, the transposon ageing model, introduced a few years later by Murray (1990), postulates that an exponential increase of TE copy number with time could eventually kill the cell or the organism by inactivating essential genes. Indeed, the activation of TEs has been demonstrated to affect lifespan associated with DNA damage in several organisms like fruit flies, and mice (Wood et al. 2016; Simon et al. 2019). Similarly, TE activation has been recently associated with neurodegenerative, autoimmune, and cancer diseases, which can in turn affect organismal lifespan (Gorbunova et al. 2021).

To mitigate detrimental TE-related effects, TE activity (expression and/or transposition) is normally repressed by epigenetic mechanisms that can involve DNA methylation, histone modifications, and/or the production of small RNAs (reviewed in Slotkin and Martienssen 2007). Ageing disrupts these TE silencing mechanisms, increasing their activity. Examples of that have been documented in several organisms, where TE expression and sometimes TE transposition increased with age in different somatic tissues (Li et al. 2013; De Cecco et al. 2013a; Chen et al. 2016; Yang et al. 2022).

In this review, we explore current literature demonstrating that TE activity can be associated with both the causes and consequences of ageing, leading to a more complex hypothesis regarding the role of TEs in ageing processes.

TEs as a Source of Genomic Variation

TEs were first described by Barbara McClintock in the late 1940s in maize (*Zea mays*). Originally referred to as "controlling elements", TEs were thought to move in response to environmental changes, thereby influencing gene expression (McClintock 1950, 1953). However, these ideas were initially rejected, and for many years, TEs were classified as "selfish" or "junk" genetic elements thought to cause only harmful effects (Strobel et al. 1979; Doolittle and Sapienza 1980; Orgel and Crick 1980; Hickey 1982). Due to the emergence of sequencing techniques, it is now understood that TEs are a significant component of genomes across various species (Bourque et al. 2018; Guio and González 2019; Wells and Feschotte 2020). Over the last decades, the importance of TEs has grown as their roles in influencing recombination rates, chromosomal rearrangements, mutagenesis, and gene regulation have been confirmed (Biémont 2010; Bourque et al. 2018). Additionally, TEs have been implicated in genome evolution and recent

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

adaptation (Biémont and Vieira 2006; González et al. 2010; Arkhipova 2018; Xia et al. 2024). Therefore, the scientific community now acknowledges that while TEs can have detrimental effects, they have also played a significant role in shaping the structure, function, and evolution of genomes.

However, this same mobility that underlies their evolutionary impact also poses a risk to genome integrity. The ability of TEs to move within the genome is central to their mutagenic potential. Over time, the accumulation of transposition events within cells can lead to genomic instability at the organismal level, a phenomenon linked to both the transposon and the error/damage theories of ageing (Weismann 1882; Murray 1990; Davidovic et al. 2010; Jin 2010).

In addition to the effects directly caused by transposition events, TEs can alter both gene structure and expression. The impact of TEs on gene structure is determined by the specific location of their insertion. When inserted into introns, TEs can be incorporated as new exons (exonization), provide alternative splicing sites, or introduce alternative polyadenylation (poly(A)) signals (Schmitz and Brosius 2011; Chénais et al. 2012; Casacuberta and González 2013; Warren et al. 2015; Oliveira et al. 2023; Xia et al. 2024). TEs can also alter gene structure when inserted in other non-coding regions such as 5′ untranslated (5′ UTR) regions. There, they can, for example, introduce alternative transcription start sites, hence modulating gene transcript lengths (Faulkner et al. 2009; Batut et al. 2013; Coronado-Zamora and González 2023; Oliveira et al. 2023).

Furthermore, TEs can serve as a source of regulatory material for the modulation of gene expression (Bourque et al. 2018). They can harbour regulatory sequences, such as promoters, enhancers, or repressive elements, that fine-tune nearby gene expression, contribute to heterochromatin formation, and act as insulators to prevent the spread of heterochromatin (Chuong et al. 2017; Trizzino et al. 2017; Villanueva-Cañas et al. 2019). Additionally, TEs influence gene expression through post-transcriptional mechanisms, affecting mRNA decay, localization, and translation efficiency (Elbarbary et al. 2016). They also help define topological domains within the nucleus, facilitating chromatin loop formation (Mamillapalli et al. 2013).

These diverse effects highlight the role of TEs in shaping genomic evolution and in influencing gene regulation.

Regulatory Mechanisms of TEs

There are multiple epigenetic mechanisms regulating TE activity, including transcriptional and post-transcriptional repression (Slotkin and Martienssen 2007; Gebert and Rosenkranz 2015).

One of the mechanisms for silencing TEs involves RNA interference pathways, which depend on the production of small non-coding RNA molecules that complementary target and regulate TE transcripts (Slotkin and Martienssen 2007).

First, the small-interfering RNA (siRNA) pathway involves the cleavage of double-stranded RNA by Dicer proteins into short siRNAs. These siRNAs are then loaded into Argonaute (AGO) proteins to form the RNA-induced silencing complex (RISC), guiding the degrading complex to a complementary transcript for cleavage (Slotkin and Martienssen 2007; Gebert and Rosenkranz 2015; Fig. 1).

Second, the Piwi-interacting RNA pathway (piRNA) was first described as the main defence mechanism against TE transcripts in germ cells. But recently, this pathway has also been identified in the somatic cells of arthropods and molluscs (Martens et al. 2005; Ghildiyal et al. 2008; Lee et al. 2011; Perrat et al. 2013; Jones et al. 2016). The mechanisms of the piRNA pathway are well understood in mice and *Drosophila*, where extensive research has been conducted. In *Drosophila*, initiator piRNAs are transcribed from piRNA clusters, which consist of loci enriched in TE sequences and processed into 23- to 30-nucleotide fragments. These piRNAs bind to PIWI proteins and guide them to complementary TE loci in the genome, where they induce repressive chromatin modifications leading to transcriptional silencing of TEs (Hyun 2017; Fig. 1). Additionally, piRNAs can target TE transcripts for post-transcriptional cleavage, further inhibiting TE expression. In that case, piRNAs bind to the Argonaute protein Aubergine (AUB) and guide the cleavage of complementary TE transcripts. This cleavage releases responder piRNAs that can interact with Argonaute3 proteins (Hyun 2017; Fig. 1). These responder piRNAs target the original piRNA precursor, leading to further processing and amplification of piRNAs, which can subsequently silence additional TE transcripts. This process, known as the "ping-pong loop", allows the amplification of piRNAs derived from TE transcripts (Hyun 2017; Fig. 1).

Finally, transcriptional regulation of TEs can also be achieved by suppressing their expression at the DNA level by inducing heterochromatisation through DNA methylation and/or histone modifications (Fig. 1). While not conserved across all eukaryotes, the methylation of cytosine residues in TE promoters by DNA methyltransferases is commonly associated with their transcriptional repression (Slotkin and Martienssen 2007). Similarly, modifications to the amino (N)-terminal tails of histones, such as deacetylation and methylation, create a repressive chromatin state around TEs. Indeed, nucleosomes associated with TEs are usually enriched in H3K9 methylation, a mark indicative of transcriptionally repressive and inactive chromatin (Martens et al. 2005).

Impact of Ageing on TE Expression

There is substantial evidence in the literature indicating that TEs can be activated with age leading to increased TE transcription in fruit flies (Table 1; Li et al. 2013; Chen et al.

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

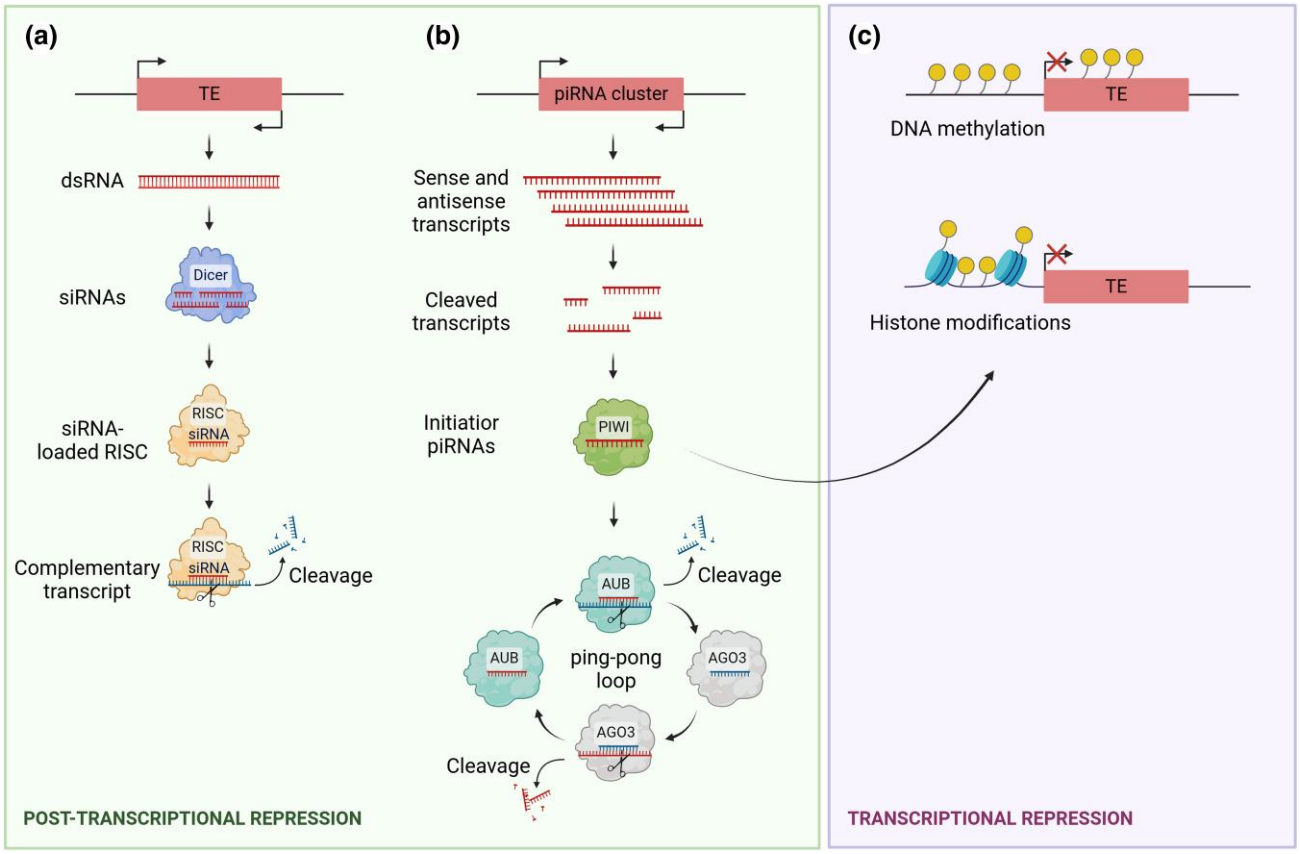


Fig. 1. Schematic representation of the main regulatory mechanisms of TE activity. a) The small-interfering RNA (siRNA) pathway: double-stranded RNA (dsRNA) is cleaved by Dicer into siRNAs, which are loaded into Argonaute (AGO) proteins to form the RNA-induced silencing complex (RISC). This complex guides the cleavage of complementary TE transcripts. b) The Piwi-interacting RNA (piRNA) pathway: piRNAs, transcribed from piRNA clusters, bind to PIWI proteins, guiding them to TE loci where they induce repressive chromatin modifications, silencing TEs at the transcriptional level. Additionally, piRNAs can target TE transcripts for post-transcriptional cleavage via the Argonaute protein Aubergine (AUB), triggering the "ping-pong loop" to amplify piRNA production. c) Transcriptional regulation through chromatin modification: TEs can be silenced at the DNA level by heterochromatinization, mediated by DNA methylation and histone modifications such as deacetylation and methylation.

Table 1 Summary table of the impact of ageing on TE expression in different organisms

Organism	Tissue	Mechanism	Impact on TE expression	References
D. melanogaster	Head	Loss of H3K9me2	Increase	Brown et al. (2020a)
D. melanogaster	Head	*Ago2* impairment in young flies	Increase	Li et al. (2013)
D. melanogaster	Fat body	LAM depletion (H3K9me3 decrease and H3K4me3 increase)	Increase	Chen et al. (2016)
D. melanogaster	Fat body, head	Loss of heterochromatin	Increase	Wood et al. (2016)
D. melanogaster	Fat body	piRNA pathway impairment	Increase	Jones et al. (2016)
D. melanogaster	Whole body, brain	ND	Increase	Yang et al. (2022)
M. musculus	Liver, muscle	Chromatin decondensation	Increase	De Cecco et al. (2013a)
M. musculus	Senescent cells	Chromatin decondensation	Increase	De Cecco et al. (2013b)
M. musculus	Cell lines	Loss of SIRT6 protein	Increase	Van Meter et al. (2014)
H. sapiens	Senescent cells	Hypomethylation	Increase	Colombo et al. (2018)
H. sapiens	Mammary luminal epithelial cells	Loss of methylation	Increase in breast cancer tissue	Senapati et al. (2023)

ND, not determined.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

2016; Jones et al. 2016; Wood et al. 2016; Brown et al. 2020a; Fabian et al. 2021; Giordani et al. 2022; Yang et al. 2022; Schneider et al. 2023), fish (Teefy et al. 2023; Xu et al. 2023), mice (De Cecco et al. 2013a; Van Meter et al. 2014), and humans (Colombo et al. 2018; Pehrsson et al. 2019; Senapati et al. 2023). Although it is generally agreed that this activation is due to an age-associated weakening of TE silencing processes, the precise molecular mechanisms underlying this phenomenon remain unclear.

In *Drosophila melanogaster*, differences in TE expression with age have been associated with changes in heterochromatin structure (Li et al. 2013; Chen et al. 2016; Wood et al. 2016; Brown et al. 2020a). For example, in older flies, a general loss of H3K9me2, a repressive histone modification typical of heterochromatin, has been observed at repetitive elements (Brown et al. 2020a). TE control in the soma of *D. melanogaster* is mediated by the siRNA pathway. Genetic manipulations of *Ago2* gene, a key component of this pathway, in young flies resulted in elevated *R2* and *mdg4* TE transcripts in heads, together with an age-dependent expression acceleration of these two TE families, which are both classified as Class I retroelements (Li et al. 2013). In *D. melanogaster* fat bodies, an overall increased expression of total retrotransposons was found in old flies compared to young ones (Chen et al. 2016). Specifically, around 16% of the annotated retrotransposons had a significant upregulation (Chen et al. 2016). Since ageing has been associated with a gradual reduction of lamin-B (LAM), a protein involved in heterochromatin maintenance, the authors used LAM mutant flies to demonstrate that its depletion in young fat bodies leads to a similar upregulation of retrotransposons as observed in old fat bodies (Chen et al. 2016). They found that upon LAM depletion, there was a decrease in H3K9me3 (repressive histone mark) and an increase in H3K4me3 (activating histone mark) in some de-repressed TE families. This suggested that the age-related reduction of LAM could contribute to the loss of heterochromatin and the activation of retrotransposons in the fat bodies (Chen et al. 2016).

Building on this finding, another study also observed increased TE expression with age in both *D. melanogaster* fat bodies and heads (Wood et al. 2016). The researchers then genetically enhanced the activity of genes known to influence heterochromatin structure, such as *Sir2*, *Su(var)3-9*, and *Dicer-2* (Wood et al. 2016). This intervention prevented the age-related loss of TE silencing observed in wild-type lines, further suggesting a common link between heterochromatin factors and the proper maintenance of TE silencing (Wood et al. 2016). Although genetic studies have demonstrated that components of the siRNA/piRNA pathway, such as Ago2 and Dicer-2, are crucial for TE control, their transcript levels have not been found significantly reduced in aged flies (Li et al. 2013; Chen et al. 2016). This suggests that factors beyond simple gene expression changes may be involved in age-associated TE reactivation and need further investigation. Additionally, a fully functional piRNA pathway was described in the adult fly fat body, whose suppression shortened lifespan (Jones et al. 2016). In this work, authors suggested that the presence of this piRNA pathway in normal somatic tissues may offer an additional cellular defence against TE reactivation and possible somatic genomic damage (Jones et al. 2016). However, further studies are needed to better understand the contribution of somatic small RNAs in the control of TE activity during ageing in wild-type populations of *Drosophila*.

In mammals, including humans, numerous studies have documented an age-associated increase in TE transcripts (De Cecco et al. 2013a, 2013b; Van Meter et al. 2014; Colombo et al. 2018; LaRocca et al. 2020; Stow et al. 2021; Liu et al. 2023; Senapati et al. 2023; Pabis et al. 2024; Tsai et al. 2024). The mechanisms proposed to explain the increase in TE expression are primarily related to changes in chromatin conformation (Jintaridth and Mutirangura 2010; De Cecco et al. 2013b; Van Meter et al. 2014; Colombo et al. 2018; Senapati et al. 2023). Early evidence of this phenomenon was observed in mice, where various families of retrotransposons showed elevated transcription levels during ageing, both *in vitro* and *in vivo* (De Cecco et al. 2013a). Indeed, this increase in transcription was associated with a relatively more open chromatin in senescent cells (De Cecco et al. 2013b). In the same species, a specific mechanism was described to be responsible for the expression of LINE1 retroelements with age and stress involving the protein Sirtuin 6 (SIRT6) (Fig. 2; Van Meter et al. 2014). This protein represses LINE1 activity by binding to the 5′UTR of LINE1 loci and facilitates heterochromatin formation at these sites through its interaction with the Heterochromatin Protein 1a (HP1a) via the KRAB-associated protein 1 (Fig. 2; Van Meter et al. 2014). However, during ageing, SIRT6 is depleted from LINE1 loci, impeding the formation of heterochromatin and thus allowing LINE1 activation (Fig. 2; Van Meter et al. 2014).

In humans, artificially induced senescent cell lines exhibit a higher expression of TEs compared to non-senescent cell lines (Colombo et al. 2018). While epigenetic modifications do not always directly result in changes in TE expression, they can be strong indicators of potential shifts in their expression. Supporting this, a comprehensive epigenetic analysis of human TEs across various normal tissues and developmental stages revealed age-dependent variations in TE epigenetic profiles (Pehrsson et al. 2019). Similarly, DNA methylation profiling using whole-genome bisulfite sequencing in human mammary luminal epithelial cells (key cells implicated in breast cancer) from younger and older women showed that several TE subfamilies consistently exhibited methylation loss with age (Senapati et al. 2023).

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

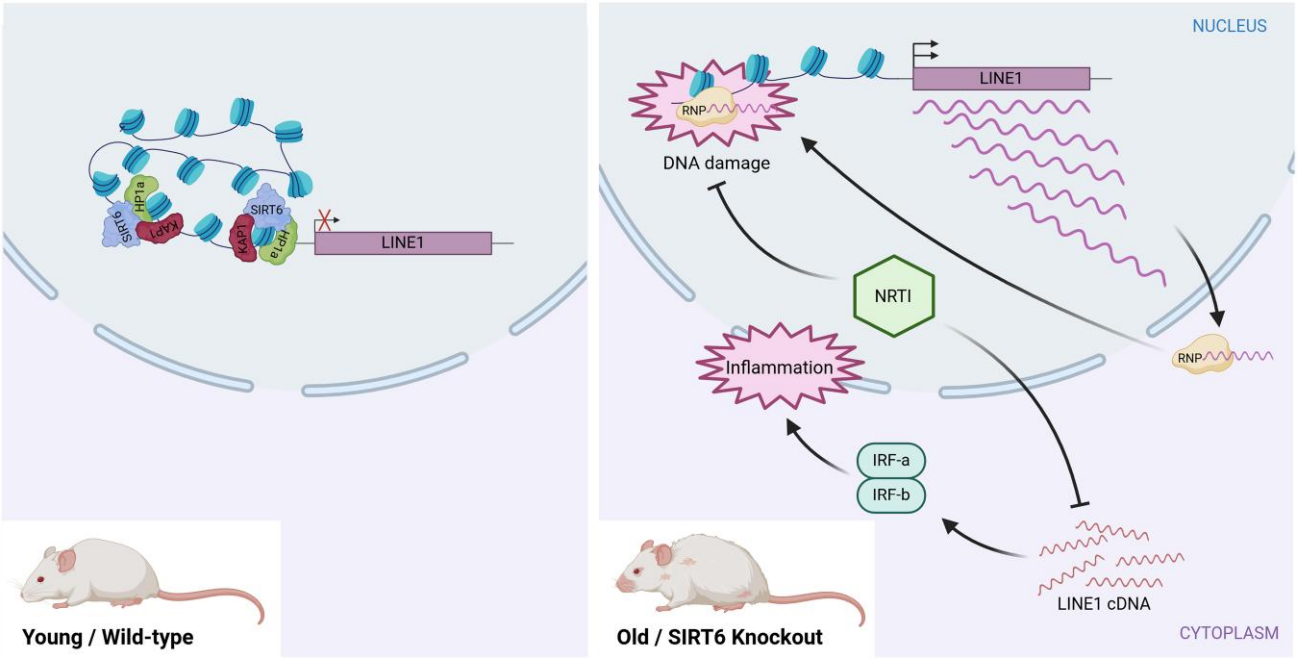


Fig. 2. Age-related activation of LINE1 elements due to SIRT6 depletion and its impact on genome stability and inflammation. In young/wild-type mouse cells, LINE1 elements are kept in a repressed state due to the regulatory role of SIRT6. SIRT6 ensures that the 5′ UTR region of LINE1 is packaged into transcriptionally inactive heterochromatin with the coordination of other factors such as KAP1 and HP1a. However, in old and senescent cells or in SIRT6 knockout mice, LINE1 elements become increasingly active. This activation is partly due to the inability to maintain the LINE1 5′ UTR in a stable heterochromatic state. As cells age, SIRT6 is depleted from the LINE1 5′ UTR, as it relocates to sites of DNA damage to facilitate DNA repair. The reduced presence of SIRT6 at the LINE1 5′ UTR weakens the silencing mechanisms, leading to the activation of LINE1 elements. The LINE1 transcripts can assemble into active ribonucleoprotein (RNP) complexes, which, upon transport to the nucleus, enhance LINE1 transposition and lead to DNA damage. Moreover, the cytoplasmic accumulation of LINE1 cDNA triggers a type I interferon (IRF) response, leading to pathological inflammation. Notably, inhibiting LINE1 replication with nucleoside reverse-transcriptase inhibitors (NRTIs) in aged mice significantly improved both mice health and lifespan.

Recent research suggests that transcriptional defects, such as increased intron retention and transcriptional readthrough may drive elevated TE expression during ageing and cellular senescence in mammals (Pabis et al. 2024). These defects become more frequent with age, potentially contributing to the rise in TE expression (Pabis et al. 2024). These findings indicate that transcriptional defects, rather than the sole expression of autonomous insertions, may be also responsible for age-related changes in transposon expression (Pabis et al. 2024).

Overall, the literature consistently reports an increased TE expression with ageing (Table 1), a trend that has been proposed as a reliable marker of biological age (LaRocca et al. 2021).

Impact of Ageing on TE Transposition

As previously mentioned, ageing can trigger the somatic expression of TEs, often as a result of disruptions in TE silencing mechanisms. However, evidence is limited regarding whether this age-induced TE expression leads to new transposition events. While there is some evidence of specific TE families increasing their copy number with age, this phenomenon has been observed in only a few species (Table 2).

Some of the earliest evidence of age-related transposition was documented in *D. melanogaster* (Driver and McKechnie 1992). In this study, the increase in transposition with ageing was examined for the *412* and *copia* Class I retroelements (Driver and McKechnie 1992). Southern blot analysis of genomic DNA from old flies showed an increase in band intensity compared to young flies, indicating that new insertions of these two TE types occur with age (Driver and McKechnie 1992). In the same species, the use of the *mdg4*-TRAP reporter system was successfully used to reveal *de novo* transposition during ageing (Li et al. 2013; Wood et al. 2016). This system is designed so that green fluorescent protein (GFP) is expressed when an endogenous *mdg4* retrotransposon inserts into an engineered site containing sequences from the *ovo* locus, a known hotspot for *mdg4* insertion (Dej et al. 1998). Additionally, the use of the *mdg4*-TRAP system in flies subjected to a dietary restriction regimen (known to extend lifespan) demonstrated that nutritional shortage suppressed the age-related increase in *mdg4* transposition events compared to flies on a high-calorie diet (Wood et al. 2016). Similarly, the development of a novel reporter system named

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Table 2 Summary table of the impact of ageing on TE transposition in different organisms

Organism	Tissue	Impact on TE transposition	TE family(ies)	Method	References
Drosophila melanogaster	Whole body	Increase	412, copia (Class I)	Southern blot	Driver and McKechnie (1992)
D. melanogaster	Heads	Increase	mdg4 (Class I)	mdg4-TRAP system	Li et al. (2013)
D. melanogaster	Whole body	Increase	mdg4 (Class I)	mdg4-TRAP system	Wood et al. (2016)
D. melanogaster	Glial and neuronal cells	Increase	mdg4 (Class I)	CLEVR	Chang and Dubnau (2019)
D. melanogaster	Whole body/brain	Increase	ND	TIDAL software	Yang et al. (2022)
D. melanogaster	Whole body/heads/abdomens	Increase	ND	DeviaTE software	Fabian et al. (2021)
D. melanogaster	Brains	No differences	ND	TEMP software	Treiber and Waddell (2017)
D. melanogaster	Thoraces	No differences	ND	Retrofind software	Schneider et al. (2023)
M. musculus	Liver and muscle	Increase	LINE1 and MusD (Class I)	qPCR assay of genomic DNA	De Cecco et al. (2013a)
H. sapiens	Brain	Increase	LINE1 and Alu (Class I)	TLDR software	Ramirez et al. (2025)

ND, not determined.

Cellular Labelling of Endogenous retroVirus Replication (CLEVR), which is more sensitive and more versatile, also revealed an age-dependent increase of *mdg4* retrotransposition in glial and neuronal cells (Chang and Dubnau 2019).

Subsequently, rather than focusing on specific transposition events within particular TE families, some studies expanded their analysis to nearly all described TE families in *Drosophila* using bioinformatic approaches on sequencing data (BOX 1) (Fabian et al. 2021; Yang et al. 2022). Using the software TIDAL to analyze whole-genome sequencing data from three different wild-type strains, researchers observed a higher number of uniquely detected TE insertions in aged flies compared to young fly genomes (Yang et al. 2022). Moreover, analysis of TE copy number using the software DeviaTE in populations selected for longevity revealed that TE families were generally more abundant in long-lived populations (Fabian et al. 2021). However, additional RNA-seq analysis showed a tendency toward a reduced TE expression in these populations, suggesting that controlling TE activity at the transcriptional level may be more critical for lifespan extension than regulating genomic insertions (Fabian et al. 2021).

Nevertheless, not all studies align with these findings. For example, the software TEMP did not detect a significant increase in new transposition events with age in genomic DNA from single adult fly neuronal cells (Treiber and Waddell 2017). Likewise, analysis of single-nucleus genome sequencing data using the software Retrofind revealed no significant increase in the number of TE insertions in nuclei from old flies compared with young ones (Schneider et al. 2023). Indeed, 76% of nuclei examined from indirect flight muscle and approximately 50% from thoraxes had no new insertions (Schneider et al. 2023).

In mice, age-related increase in TE transposition was reported for two specific Class I TE families, LINE1 and MusD. Quantitative PCR analysis of genomic DNA from skeletal muscle and liver tissues in 36-month-old mice revealed a significant increase in the copy number of these two TE families (De Cecco et al. 2013a). Moreover, authors also observed that LINE1 and MusD increase in naturally occurring age-associated cancers such as lymphoma and hepatocellular carcinoma (De Cecco et al. 2013a).

In humans, advances in long-read sequencing techniques have enabled the detection of novel transposition events in 18 frontal cortex brain samples from individuals between 67- and 92-year old (Ramirez et al. 2025). Using the software TLDR, which identifies retrotransposon insertions from long-read sequencing data, researchers found novel insertions of the Alu and LINE1 families in the ageing human brain (Ramirez et al. 2025). However, the authors noted that the number of Alu and LINE1 insertions identified per sample is likely underestimated due to the relatively low coverage of their long-read sequencing (Ramirez et al. 2025).

These findings highlight the complexity of TE regulation during ageing and indicate that further investigation is needed to fully understand the link between ageing and new transposition events, or whether such insertions are tissue-specific, as experimental and bioinformatic approaches can significantly influence the detection and interpretation of TE activity in ageing cells.

Impact of TE Activation in Longevity

TE activation (both expression and transposition) can impact cellular function in multiple ways. As already mentioned, TEs can modify the genome through their mobilization and influence host gene expression both in *cis* and *trans*, at transcriptional and post-transcriptional levels (Elbarbary et al. 2016; Chuong et al. 2017; Bourque et al. 2018; Bourgeois and Boissinot 2019). These alterations are a clear source of cellular damage, which can

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

subsequently impact the organism's lifespan. Hence, the transposon ageing model suggests that the exponential increase in TE copy number over time within the genome could ultimately lead to cell or organism death as a result of an accumulation of somatic mutations (Murray 1990).

The effect of increased TE transposition on age-related phenotypes was first shown in *Drosophila* species (Driver and McKechnie 1992; Woodruff and Nikitin 1995). In these studies, artificial somatic transposition of *P-element* Class II insertions in *D. melanogaster* flies induced a reduction of lifespan in both males and females (Driver and McKechnie 1992; Woodruff and Nikitin 1995). Similar findings were observed with the somatic mobilization of the *mariner* element in *D. simulans* (Class II) (Nikitin and Woodruff 1995). However, this was not the case in *D. melanogaster*, likely because the tested strains contained only two potentially mobile copies of the *mariner* element (Nikitin and Woodruff 1995). Consequently, the somatic genomic damage caused by *mariner* elements would be less than that of *P-elements*, which had 17 putatively active copies in the tested *D. melanogaster* strains (Nikitin and Woodruff 1995). More recently, in the same species, no increase in TE transposition with age was observed. However, reduced expression of *412* and *Roo* Class I elements led to an extension in lifespan, suggesting that TE expression, rather than an increase in copy number, plays a key role in longevity (Schneider et al. 2023).

TE expression has also been associated with age-related phenotypes in *Ago2* mutant flies. As expected, these mutants showed elevated TE expression in somatic tissues, along with progressive, age-related memory impairment and shorter lifespans compared to wild-type flies (Li et al. 2013). This finding aligned with previous works showing that mutations in *Dicer-2* and *loquacious* genes, similarly result in reduced lifespans (Lim et al. 2011; Liu et al. 2012; Wood et al. 2016).

To investigate the contribution of TE activity to age-related effects, pharmacological inhibition of TE transposition using the reverse transcriptase inhibitor lamivudine (3TC) resulted in a lifespan extension in *Dicer-2* mutants. These mutants exhibited increased double-strand breaks, likely caused by TE transposition, which was mitigated by 3TC treatment (Wood et al. 2016). Similarly, the loss of function of *Hinfp*, a zinc-finger transcription factor that interacts with Histone 1, a key chromatin assembly factor, leads to the derepression of numerous TEs (Nirala et al. 2021). This derepression accelerates ageing and promotes cancer-related phenotypes, such as abnormal cell proliferation and tissue transformation (Nirala et al. 2021). Rather than modifying genes involved in TE silencing, directly activating *mdg4* insertions with a UAS-Gal4 system increased mortality in middle-aged flies, despite the retrotransposon being active early in life (Rigal et al. 2022). Authors suggested that young flies might tolerate TE activation, but as they age, TE activity could become detrimental, ultimately affecting their survival (Rigal et al. 2022). This may be due to the combined effects of age-related metabolic and physiological changes with TE activity, which together exacerbate cellular damage. Notably, this period coincides with the natural expression of endogenous TEs during normal ageing (Rigal et al. 2022).

In mice, as mentioned in a previous section, SIRT6 deficiency led to the derepression of LINE1 elements, resulting in the cytoplasmic accumulation of LINE1 cDNA. This triggered a type I interferon (IFN-I) response, causing pathological inflammation, which is a hallmark of ageing (Fig. 2; Simon et al. 2019). As a consequence, mice exhibited an acute, degenerative ageing-like phenotype, characterized by reduced lifespan and growth retardation. Notably, inhibiting LINE1 replication with nucleoside reverse-transcriptase inhibitors significantly improved the health and extended the lifespan of SIRT6 knockout mice, further validating these findings (Fig. 2; Mostoslavsky et al. 2006; Simon et al. 2019).

Similarly, in humans, LINE1 derepression during cellular senescence in fibroblasts leads to its upregulation, which in turn activates the IFN-I response. Again, this activation appears to depend on the accumulation of LINE1 cDNA in the cytoplasm, ultimately causing inflammation (De Cecco et al. 2019). Furthermore, LINE1 elements, which contain lamin-associated domains, are anchored to the nuclear lamina through SIRT7-mediated deacetylation, a mechanism that contributes to the regulation of their expression (Vazquez et al. 2019). Lamin proteins (structural proteins of the nuclear lamina that help maintain nuclear integrity and organize chromatin) are often lost or reduced in ageing and cancer cells. This loss likely compromises the ability of SIRT7 to tether LINE1 elements to the nuclear periphery, leading to their activation. As a result, LINE1 activation may contribute to phenotypes such as increased DNA damage and changes in gene expression (Vazquez et al. 2019).

Also in humans, a recent study has shown that chronological age (the time an organism has been alive) is not directly correlated with the expression of LTR, LINE, or SINE TE families in blood samples from multiple cancer-free cohorts (Tsai et al. 2024). Instead, the expression of these retrotransposons was found to be correlated with biological age (the physiological and functional state of an organism), with LTR and LINE expression associated with inflammatory responses, and SINE expression linked with the upregulation of the DNA repair pathways, both key hallmarks of ageing (Tsai et al. 2024). Therefore, SINE might contribute to ageing through genome instability, whereas LTR and LINE expression could be associated with inflammation (Tsai et al. 2024).

In models of premature ageing such as Hutchinson–Gilford progeria syndrome and Werner syndrome, senescent human muscle precursor cells exhibited a derepression

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

of HERV-K insertions (Liu et al. 2023). This derepression led to the production of retrovirus-like particles, contributing to cellular senescence and accelerating ageing effects in younger cells, while also activating the innate immune response (Liu et al. 2023). In addition, neutralizing these HERV-K particles with antibodies has shown potential in mitigating these age-related processes (Liu et al. 2023). Likewise, early activation of LINE1 elements in progeroid cells, with signs of early-onset ageing, has been associated with heterochromatin loss and the emergence of senescent phenotypes (Della Valle et al. 2022). Targeting LINE1 RNA with antisense oligonucleotides helped restore heterochromatin epigenetic marks and lowered the expression of senescence-associated genes in both human cells and in a mouse model of Hutchinson–Gilford progeria syndrome, ultimately extending lifespan (Della Valle et al. 2022).

While TEs are primarily associated with promoting ageing or age-related phenotypes, there are also examples in the literature highlighting their beneficial effects on lifespan.

The evolution of telomeres is an example of antagonistic evolution of TEs. DNA polymerase is unable to fully replicate the ends of chromosomes, leading to their gradual shortening as the organism ages. Telomeres, which are chromosomal extensions, serve to protect against this loss of chromosomal length (Aubert and Lansdorp 2008). These structures are primarily composed of repeated sequences that, in *Drosophila*, are derived from TEs (Mason and Biessmann 1995). Notably, the elongation of telomeres in *Drosophila* is facilitated by specific retrotransposons: HeT-A, TART, and TAHRE (Pardue and DeBaryshe 2008). Consequently, the maintenance of telomeres and the control of TE activity within an organism must be carefully coordinated. Telomeric TEs are targeted by piRNAs in the same manner as other TEs, and disruptions in the piRNA pathway can lead to increased transcript levels of these TEs (Savitsky et al. 2006). Unlike other insertions, telomeric TEs are consistently transcribed in both sense and antisense orientations, which allows telomere maintenance despite the regulation exerted by the piRNA pathway (Danilevskaya et al. 1999).

Another example of the beneficial impact of TEs on ageing was observed in natural populations of *D. melanogaster*. The Hoppel insertion within the first intron of the longevity gene *Indy* enhances both fecundity and lifespan (Zhu et al. 2014). Moreover, the expression levels of *Indy* were found to be positively correlated with the presence of the Hoppel insertion (Zhu et al. 2014). However, homozygosity for the *Indy* insertion variant results in reduced fitness, thereby favouring the retention of the insertion allele in the heterozygous state (Zhu et al. 2014).

In summary, while there are rare instances where TE activity may positively influence the evolution of longevity, the overall evidence suggests that TE activity predominantly contributes to reduce lifespan. While the most evident consequence of TE expression with age is its potential to cause genomic instability through TE transposition, emerging evidence suggests that TE expression alone may also have an important role in reducing longevity, likely due to the activation of inflammatory responses. Although further research is necessary, targeting retrotransposon activity and the associated inflammation could potentially become a therapy for age-related diseases in the future (Simon et al. 2019).

The Role of TE Activity in Age-Related Diseases

Increased TE activity has been observed in various age-related diseases, often due to epigenetic control failures that occur naturally with ageing, as previously discussed. These disruptions can be exacerbated by genetic predispositions and environmental factors, influencing disease through different mechanisms (Gorbunova et al. 2021; Chénais 2022; Copley and Shorter 2023).

Elevated TE activity is known to cause DNA damage, primarily through double-strand breaks, which activate cell death pathways and contribute to neurodegeneration in diseases such as amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), ataxia, and tauopathies like Alzheimer's disease (Frost et al. 2014; Takahashi et al. 2022; Liu et al. 2023). Comprehensive genome analyses have also shown that increased retrotransposition contributes to genomic instability across multiple cancer types (Rodriguez-Martin et al. 2020). This instability arises through mechanisms such as deletions, translocations, duplications, ultimately promoting cancer development and driving metastasis (Rodriguez-Martin et al. 2020).

The inflammatory response triggered by TE expression is also a common feature in many diseases including cancer, ALS/FTD, Alzheimer, multiple sclerosis, and autoimmune disorders (Zhao et al. 2018; Tossberg et al. 2020; Mosaddeghi et al. 2023; Ochoa et al. 2023; Zhen et al. 2023). This occurs because double-stranded RNA generated during retrotransposition are typically recognized by cells as foreign, often triggering the innate immune response, which in some cases leads to apoptosis.

The potential role of insertions as cis-regulatory sequences has also been linked with cancer and neurodegenerative diseases. In cancer, several studies have demonstrated that epigenetic changes in TE insertions can confer promoter-like functions to TEs, potentially driving oncogene expression (Lock et al. 2014; Deniz et al. 2020; Senapati et al. 2023; Lanciano and Cristofari 2024). Supporting this, a large tumour dataset analysis identified 1,068 cases of cryptic promoter activation involving TEs across 33 tumour types (Shah et al. 2023). In neurodegenerative diseases, SINE-VNTR-Alu (SVA) elements have been shown to regulate the expression of their neighbouring genes in human neuronal cells. For example, the genetic

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

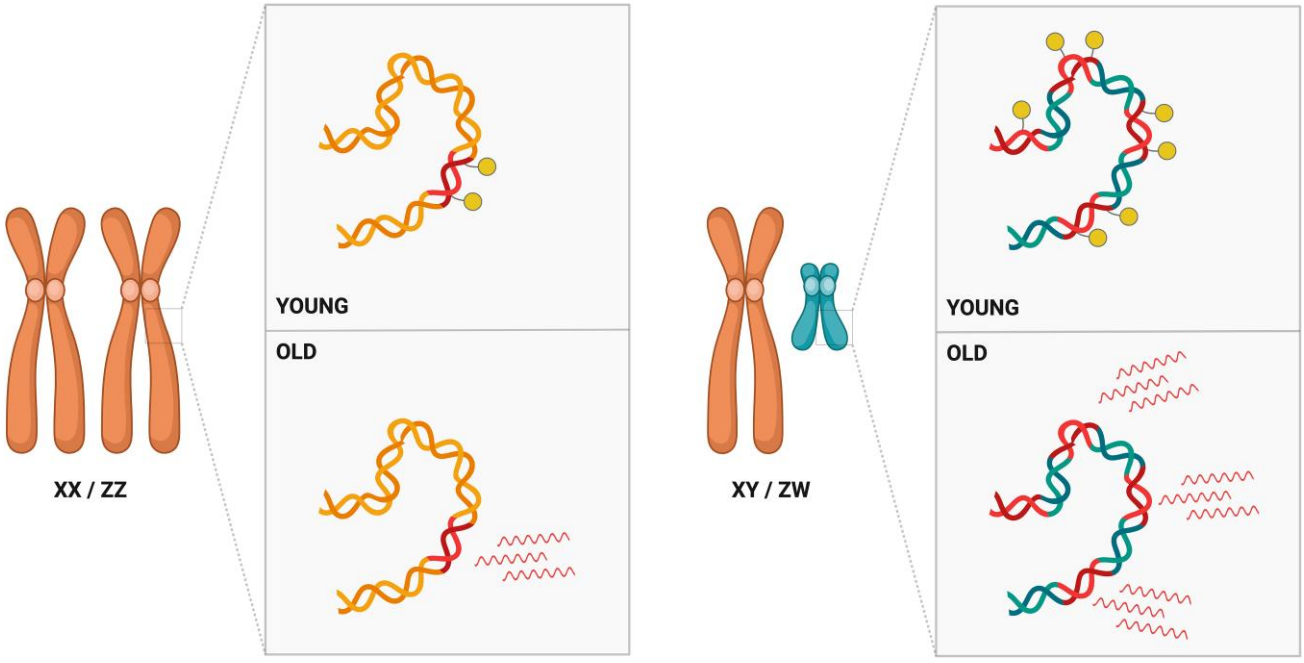


Fig. 3. Schematic representation of the toxic Y effect. The heterogametic sex (XY or ZW chromosomes, right) harbours a higher abundance of TE sequences (red) compared to the homogametic sex (XX or ZZ chromosomes, left). These TEs are typically silenced by epigenetic mechanisms (yellow lollipops). However, in aged individuals, these silencing mechanisms may weaken, leading to the activation of TEs. Because the Y or W chromosomes are TE-rich, more TEs may become active in one sex than the other, potentially contributing to sex differences in ageing and reduced lifespan.

deletion of SVA elements in Alzheimer and Parkinson disease-associated risk loci resulted in epigenomic changes in the surrounding sequences through a significant reduction of H3K4me3 histone mark (van Bree et al. 2022). Moreover, the deletion of the SVA element at the *BCKDK* locus was associated with increased expression of the gene encoding the RNA-binding protein FUS (van Bree et al. 2022), which its abnormal aggregation has been previously linked to neurodegenerative diseases such as ALS and FTD (Deng et al. 2014).

Finally, TEs have been shown to contribute to cancer development by facilitating the tissue-specific transfer of circulating cell-free tumour-derived DNA (ctDNA; Cinar et al. 2024). These molecules facilitate tumour progression by transferring oncogenes between tumour cells, altering the tumour microenvironment, and modifying cellular responses to treatment (García-Olmo et al. 2010; Trejo-Becerril et al. 2012; Dvořáková et al. 2013; Cinar et al. 2024).

Overall, increased TE activity is implicated in a variety of age-related diseases, including neurodegenerative disorders and cancer, with mechanisms such as DNA damage, inflammatory responses, and dysregulated gene expression playing central roles in disease progression.

Transposable Elements and Sex Gap in Longevity

The age-related activation of TEs has been linked to the phenomenon of the sex gap in longevity. This refers to the fact that, in many animal species, males and females exhibit distinct mortality patterns leading to variations in longevity (Marais et al. 2018; Lemaître et al. 2020). Between males and females, the amount of TE sequences can differ substantially due to the presence of a repeat-rich Y or W chromosome in the heterogametic sex. One plausible explanation for the sex gap in longevity is the "toxic Y effect", which suggests that the higher abundance of TEs on the Y (or W) chromosome compared to other chromosomes may have harmful consequences for the heterogametic sex (Fig. 3; Brown et al. 2020a; Nguyen and Bachtrog 2021; Peona et al. 2021; Warmuth et al. 2022; Teoli et al. 2024). This hypothesis proposes that, due to the TE-rich Y chromosome, more TEs may become active in older males than in females, potentially contributing to sex differences in ageing and the shorter lifespan observed in males (Fig. 3; Brown et al. 2020a; Nguyen and Bachtrog 2021; Peona et al. 2021; Warmuth et al. 2022; Teoli et al. 2024).

In *D. melanogaster*, it has been shown that ageing males experience a more rapid loss of H3K9me2 compared to females, leading to a greater TE activation during ageing in males (Brown et al. 2020b). This aligns with the observation that the Y chromosome, and TEs in general, impact chromatin structure genome-wide, supporting the "heterochromatin sink" model (Gatti and Pimpinelli 1992; Brown et al. 2020b). This model refers to the idea that the TE activation during ageing drains (hence "sink") proteins and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

other factors that are usually involved in the maintenance of the heterochromatin. When these factors are depleted or mislocalized, the heterochromatin weakens further, creating a feedback loop where more TEs become activated, further straining the heterochromatin (Gatti and Pimpinelli 1992). Furthermore, putatively Y-linked TEs seem to lose heterochromatin more readily as individuals age, indicating that male-specific TEs are particularly susceptible to derepression in older flies (Brown et al. 2020a). Longevity assays on flies with abnormal sex chromosome configurations (XO, XXY, and XYY) also suggest that additional Y chromosomes could decrease lifespan (Brown et al. 2020a).

Recent transcriptomic data from human individuals with different karyotypes (XX, XY, XXY, and XYY) suggested that the presence and number of Y chromosomes may be associated with increased expression of certain TE subfamilies. This elevated TE expression could potentially contribute to a reduced lifespan in individuals carrying one or more Y chromosomes (Teoli et al. 2024).

However, using CRISPR/Cas9 to generate *D. melanogaster* Y chromosomes of varying sizes indicated that the size of the Y chromosome, and thus the amount of heterochromatin and TEs, did not contribute to sex-specific longevity differences (Delanoue et al. 2023). Similarly, in *D. pseudoobscura*, while larger Y chromosomes were associated with slightly increased TE expression, particularly in older males, these differences did not accelerate ageing (Nguyen and Bachtrog 2021).

As a result, the role of the Y chromosome in reducing male longevity through increased TE activity remains controversial, highlighting the need to investigate the contribution of TEs to sex differences in longevity across other species.

TEs and Their Association With Longevity in Long-Lived Organisms

Studying long-lived species could provide valuable insights into the mechanisms of ageing and cellular senescence, potentially clarifying the role of TEs in longevity. The concept of "negligible senescence" was first introduced to describe organisms with exceptionally slow ageing rates (characterized by gradual changes in physical appearance, reproductive capacity, and susceptibility to disease) such as certain conifers and fish species (Finch 1998). The mechanisms that contribute to slow senescence are diverse and species-specific, though some examples involving TEs have been reported (Sulak et al. 2016; Zhao et al. 2021; Jové et al. 2023; Sahm et al. 2024).

As expected given their mutagenic capacity, most of the studies of long-lived species show a consistent pattern of overall reduction in both TE abundance and TE expression (Sturm et al. 2017; Elsner et al. 2018; Teefy et al. 2020; Ricci et al. 2023). In *Hydra*, a long-lived freshwater cnidarian, mortality does not increase with age, indicating an absence of senescence (Schaible et al. 2015). One proposed mechanism behind *Hydra*'s ability to evade ageing is the continued activity of the piRNA pathway in somatic cells (Sturm et al. 2017; Teefy et al. 2020). Indeed, the somatic activity of the piRNA pathway in *Hydra* represses TE expression in its somatic cell lineages, potentially contributing to the organism's exceptional longevity (Teefy et al. 2020).

The naked mole rat (*Heterocephalus glaber*) is a well-known organism studied for its resistance to cancer and its remarkable longevity compared to closely related species. A comparative genomic study between *H. glaber* and short-lived cancer-prone rodents (guinea pigs, rats, and mice) revealed a distinct TE landscape in *H. glaber*, reflecting a recent TE dynamic markedly different from the other rodents (Ricci et al. 2023). Notably, it has been observed that *H. glaber* did not accumulate non-LTR retrotransposons in recent evolutionary times, unlike the other species. Furthermore, an analysis of open reading frames with intact protein domains showed that naked mole rats possess only LINE elements lacking functional domains (Ricci et al. 2023).

Additionally, the comparison of the TE content and dynamic in long-lived bats from the genera *Myotis*, *Rhinolophus*, *Pteropus,* and *Rousettus* against the short-lived *Molossus molossus* (one of the bats with the shortest lifespan in the order Chiroptera) revealed a higher overall TE content in *M. molossus* (40.6%) compared to the long-lived bats (ranging from 28.84% to 34.29%; Ricci et al. 2023). Similar to the observations of the naked mole rat, long-lived bats showed a reduction in non-LTR retrotransposons. In contrast, the short-lived *M. molossus* had a high proportion of SINE elements, which rely on the transcriptional machinery of LINE elements for transposition (Ricci et al. 2023). This may impose a double burden of TE expression, potentially triggering inflammatory responses. The study suggests that both TE content and evolutionary TE dynamics, not just expression levels, may be linked to longevity (Ricci et al. 2023).

Social insects like honey bees and termites present valuable new models for ageing research due to the striking lifespan differences between colony members. The reproductive individuals, such as queens and kings, live for several decades, while sterile workers typically survive only a few weeks (Elsner et al. 2018). A transcriptomic analysis of old and young queens, kings, and termite workers (*Macrotermes bellicosus*) reported a 14.9% increase in TE expression only in old workers (Fig. 4; Elsner et al. 2018). This increase in TE expression appears to be linked to the downregulation of piRNA-pathway-related genes in workers compared to queens and kings, suggesting a reduction in the ping-pong cycle response in older workers (Fig. 4; Elsner et al. 2018).

Although TEs are typically associated with negative effects on ageing, they have, in some cases, been found to

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

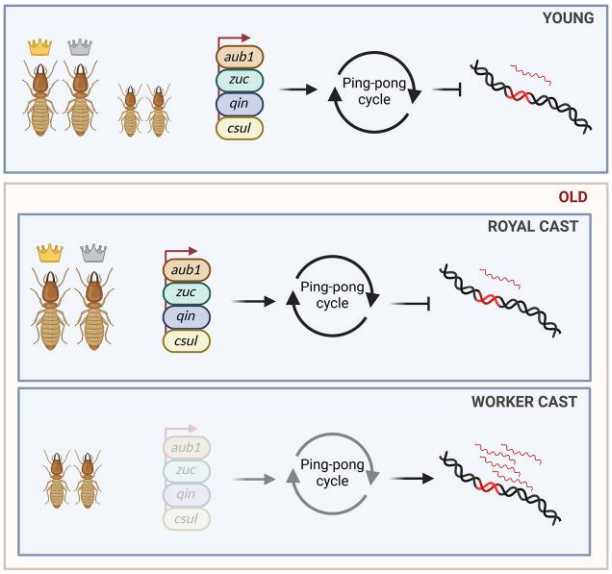


Fig. 4. Schematic representation of the mechanism underlying the difference in longevity between royal and worker castes in termites. In termites, the queen and king greatly outlive the workers. Analysis of TE expression in young and old samples of either cast revealed an increase in TE transcripts only in old workers, likely due to a piRNA pathway downregulation. Indeed, the downregulation of piRNA pathway effectors, such as *aub1*, *zuc*, *qin* and *csul*, in older workers results in decreased TE control, which subsequently leads to an increase in TE expression.

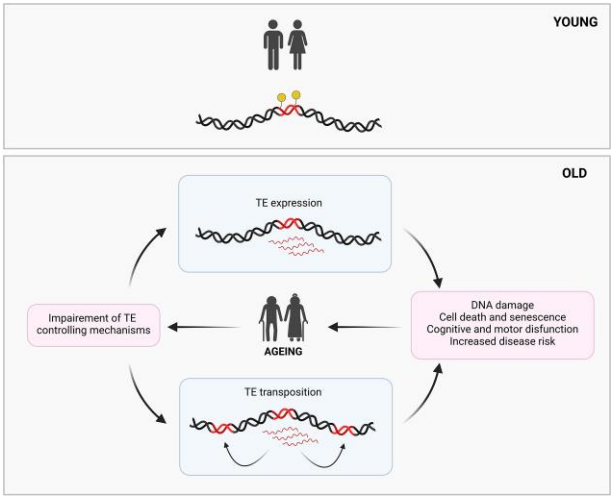


Fig. 5. TE activity as a cause and consequence of ageing. TEs are typically repressed by epigenetic mechanisms (represented as yellow lollipops). With ageing, the breakdown of these regulatory mechanisms leads to their deregulation, resulting in increased expression and occasional transposition. This elevated TE activity can contribute to DNA damage, cell death, and cellular senescence, as well as cognitive and motor dysfunction, ultimately raising disease risk and potentially accelerating age-related decline.

play a beneficial role in the ageing process (Sulak et al. 2016; Zhao et al. 2021; Sahm et al. 2024). For example, blind mole rats (*Spalax galili*), small rodents known for their remarkably long lifespan (over 21 years) and their resistance to both spontaneous and induced tumour formation, provide a compelling example of how TE activation can function as a tumour suppressor (Zhao et al. 2021). In this species, the activation of retrotransposons triggers the innate immune system, initiating a coordinated cell death process that helps prevent tumour formation. Specifically, retrotransposons contribute to the formation of cytoplasmic RNA–DNA hybrids, which activate the cyclic GMP–AMP synthase (stimulator of IFN genes) IFN pathway, leading to the elimination of premalignant cells (Zhao et al. 2021). This protective mechanism is associated with reduced DNA methylation levels, allowing for the increased retrotransposon expression and, in turn, enhance tumour suppression (Zhao et al. 2021).

Another relevant example can be found in African elephants (*Loxodonta africana*), the largest living land mammal, where it has been observed that their genome contains 20 copies of the master tumour suppressor gene *TP53,* known as *TP53* retrogenes (*TP53RTG*) (Sulak et al. 2016). *TP53* was retroduplicated and increased in copy number through repeated segmental duplications (Sulak et al. 2016). Indeed, each of the *TP53RTG* retrogenes was found to be flanked by nearly identical clusters of TEs (Sulak et al. 2016). This expansion of the *TP53RTG* family might have enhanced the DNA damage response, resulting in increased resistance to cancer and extended longevity (Sulak et al. 2016; Vazquez and Lynch 2021).

Recently, a chromosome-level genome assembly of the Greenland shark (*Somniosus microcephalus*), the longest-lived vertebrate known with an estimated lifespan of around 400 years, revealed a TE content of 70.6%. This represents the highest proportion of TE content reported among all known shark species (Sahm et al. 2024). The expansion of TEs might be related to the expansion of DNA repair genes. Particularly, retrotransposon activity may have facilitated the expansion of DNA repair-related retrogenes, which in turn enabled the organism to tolerate higher levels of TE activity (Sahm et al. 2024).

Overall, these studies likely represent a glimpse of the vast potential for discovery regarding the relationship between TEs and extreme longevity, with examples of both negative and positive associations.

Conclusions and Perspectives

Based on the analysis of existing literature, we conclude that TE activity plays a feedforward role in ageing, acting both as a cause and a consequence (Fig. 5). TEs interact with various cellular pathways and epigenetic mechanisms that influence ageing, highlighting their complex involvement within a broader network of ageing-related processes.

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

As ageing progresses, the breakdown of mechanisms that regulate TEs leads to their deregulation, resulting in increased expression and sometimes transposition. This activity contributes to genomic instability and can significantly impact cellular health, for example, through the production of TE-derived transcripts that may activate inflammatory pathways. Together, these effects drive age-related decline and disease (Fig. 5). Yet, intriguingly, some organisms with a high TE abundance do not experience reduced lifespans, suggesting that the relationship between TE abundance and ageing is far from straightforward.

Future research should aim to unravel the mechanisms that drive TE activation during ageing and investigate whether TE expression alone can have significant effects on the ageing process in different species. Advanced sequencing and bioinformatics techniques will be crucial for obtaining a more detailed assessment of TE activity at the copy-specific level, potentially revealing how individual elements contribute to age-related phenomena. Additionally, exploring the interactions between TEs and epigenetic regulators in the context of ageing holds great promise for developing novel therapeutic interventions, where targeted modulation of TE activity could help alleviate age-related dysfunctions.

BOX 1: Tools to Study TE Activity

Annotating TEs and analyzing their expression or transposition rates is challenging due to their repetitive nature, requiring specialized bioinformatic tools and pipelines (Teissandier et al. 2019; Lanciano and Cristofari 2020; Loreto et al. 2024). Here, we present a selection of some of the most recent and/or widely used tools that can aid in studying the dynamics of TEs.

A key question in the study of TEs and ageing is whether transposition increases with age, which can be addressed by comparing sequencing data from young and old samples. Nevertheless, inferring TE abundance or identifying low-frequency polymorphic TEs from sequencing reads, rather than from complete genome assemblies, presents unique challenges, especially when analyzing samples with nearly identical genome assemblies. This issue frequently arises in age-related comparisons of the same organism or tissue type. Indeed, a commonly employed experimental approach, pooled sequencing, involves combining tissues from multiple individuals and sequencing the DNA without tracking the origin of each read. In pooled sequencing data, methods for detecting structural variations in nearly identical genomes are generally limited to identifying variations shared by the majority within the pool. Detecting TE transpositions and estimating their frequencies in pooled sequencing datasets introduces additional computational challenges. Key obstacles include confidently identifying rare TE transposition events, recognizing reads that support the same transposition event, and mitigating biases due to non-uniform sequencing depth across the genome.

In studies of TE transposition and ageing, software such as TEMP (Zhuang et al. 2014) and TIDAL (Rahman et al. 2015) have been used to identify polymorphic TEs in short-read pool-sequencing data and estimate their frequencies, helping to determine whether ageing increases transposition rates (Treiber and Waddell 2017; Yang et al. 2022). Similarly, DeviaTE can be applied to both pool and individual sequencing data, and latest updates support the use of either long-read or short-read sequencing (Weilguny and Kofler 2019).

Recently, tools like TrEMOLO (Mohamed et al. 2023) and GraffiTE (Groza et al. 2024) have been developed to infer insertion frequencies of polymorphic TEs in pooled data from long-reads or genome assemblies, with the latter employing a pangenomic approach.

It is also essential in age-related studies to precisely measure TE expression, including copy-specific information (Lanciano and Cristofari 2020). TEs within the same family, by definition, share over 80% sequence identity, which complicates the investigation of TE regulation and expression in a copy-specific manner (Wicker et al. 2007; Lanciano and Cristofari 2020). Most genome-wide studies of TE expression analyze data at the family level. This process typically involves mapping short reads to either TE consensus sequences or specific TE copy sequences, followed by read counts for each family. Tools such as TEcount from the TEtools package (Lerat et al. 2017) and TEtranscripts (Jin et al. 2015) are commonly used for these analyses. Recently, numerous methods have emerged that leverage short-read sequencing datasets to address the multi-mapping issue and enable copy-level analyses (Lanciano and Cristofari 2020). These approaches utilize various algorithms to statistically reassign multi-mapped reads to specific locations, such as the expectation-maximization algorithm employed in TEtranscripts (Jin et al. 2015), SQuIRE (Yang et al. 2019), and Telescope (Bendall et al. 2019). TE expression pipelines that support the use of long-read sequencing data have been also recently developed (Kelsey et al. 2024; Rebollo et al. 2024).

A wide array of tools are available for analyzing TE transposition and expression, with new ones emerging as sequencing technologies continue to evolve (TE Hub Consortium et al. 2021). An extensive list of these tools, along with various other resources, can be found on the community-driven platform TEhub (https://tehub.org/).

Acknowledgments

We thank Tomás Carrasco-Valenzuela, Matthieu Boulesteix, Marie Fablet, and Rita Rebollo for their insightful comments on the manuscript. Figures were created with BioRender.com.

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Funding

This work was supported by the Agence Nationale de la Recherche (project LongevitY, grant Projet-ANR-20-CE02-0015), the HORIZON EUROPE Marie Sklodowska Curie Actions (project GeEpiAdaptation, grant 101065313), by the INSA Lyon and by the Université Claude Bernard Lyon 1.

Data Availability

No new data were generated or analyzed in support of this research.

Literature Cited

Arkhipova IR. Neutral theory, transposable elements, and eukaryotic genome evolution. Mol Biol Evol. 2018:35(6):1332–1337. https://doi.org/10.1093/molbev/msy083.

Aubert G, Lansdorp PM. Telomeres and aging. Physiol Rev. 2008:88(2):557–579. https://doi.org/10.1152/physrev.00026.2007.

Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res. 2013:23(1):169–180. https://doi.org/10.1101/gr.139618.112.

Bendall ML, de Mulder M, Iñiguez LP, Lecanda-Sánchez A, Pérez-Losada M, Ostrowski MA, Jones RB, Mulder LCF, Reyes-Terán G, Crandall KA, et al. Telescope: characterization of the retrotranscriptome by accurate estimation of transposable element expression. PLoS Comput Biol. 2019:15(9):e1006453. https://doi.org/10.1371/journal.pcbi.1006453.

Biémont C. A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics. 2010:186(4):1085–1093. https://doi.org/10.1534/genetics.110.124180.

Biémont C, Vieira C. Junk DNA as an evolutionary force. Nature. 2006:443(7111):521–524. https://doi.org/10.1038/443521a.

Bourgeois Y, Boissinot S. On the population dynamics of junk: a review on the population genomics of transposable elements. Genes (Basel). 2019:10(6):419. https://doi.org/10.3390/genes10060419.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. Ten things you should know about transposable elements. Genome Biol. 2018:19(1):199. https://doi.org/10.1186/s13059-018-1577-z.

Brown EJ, Nguyen AH, Bachtrog D. The Y chromosome may contribute to sex-specific ageing in Drosophila. Nat Ecol Evol. 2020a:4(6):853–862. https://doi.org/10.1038/s41559-020-1179-5.

Brown EJ, Nguyen AH, Bachtrog D. The Drosophila Y chromosome affects heterochromatin integrity genome-wide. Mol Biol Evol. 2020b:37(10):2808–2824. https://doi.org/10.1093/molbev/msaa082.

Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Mol Ecol. 2013:22(6):1503–1517. https://doi.org/10.1111/mec.12170.

Chang Y-H, Dubnau J. The gypsy endogenous retrovirus drives non-cell-autonomous propagation in a Drosophila TDP-43 model of neurodegeneration. Curr Biol. 2019:29(19):3135–3152.e4. https://doi.org/10.1016/j.cub.2019.07.071.

Chen H, Zheng X, Xiao D, Zheng Y. Age-associated de-repression of retrotransposons in the Drosophila fat body, its potential cause and consequence. Aging Cell. 2016:15(3):542–552. https://doi.org/10.1111/acel.12465.

Chénais B. Transposable elements and human diseases: mechanisms and implication in the response to environmental pollutants. Int J Mol Sci. 2022:23(5):2551. https://doi.org/10.3390/ijms23052551.

Chénais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene. 2012:509(1):7–15. https://doi.org/10.1016/j.gene.2012.07.042.

Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017:18(2):71–86. https://doi.org/10.1038/nrg.2016.139.

Cinar M, Martinez-Medina L, Puvvula PK, Arakelyan A, Vardarajan BN, Anthony N, Nagaraju GP, Park D, Feng L, Sheff F, et al. Transposon DNA sequences facilitate the tissue-specific gene transfer of circulating tumor DNA between human cells. Nucleic Acids Res. 2024:52(13):7539–7555. https://doi.org/10.1093/nar/gkae427.

Colombo AR, Triche T, Ramsingh G. Transposable element expression in acute myeloid leukemia transcriptome and prognosis. Sci Rep. 2018:8(1):16449. https://doi.org/10.1038/s41598-018-34189-x.

Copley KE, Shorter J. Repetitive elements in aging and neurodegeneration. Trends Genet. 2023:39(5):381–400. https://doi.org/10.1016/j.tig.2023.02.008.

Coronado-Zamora M, González J. Transposons contribute to the functional diversification of the head, gut, and ovary transcriptomes across Drosophila natural strains. Genome Res. 2023:33(9):1541–1553. https://doi.org/10.1101/gr.277565.122.

Danilevskaya ON, Traverse KL, Hogan NC, DeBaryshe PG, Pardue ML. The two Drosophila telomeric transposable elements have very different patterns of transcription. Mol Cell Biol. 1999:19(1):873–881. https://doi.org/10.1128/MCB.19.1.873.

Davidovic M, Sevo G, Svorcan P, Milosevic DP, Despotovic N, Erceg P. Old age as a privilege of the "selfish ones.". Aging Dis. 2010:1:139.

De Cecco M, Criscione SW, Peckham EJ, Hillenmeyer S, Hamm EA, Manivannan J, Peterson AL, Kreiling JA, Neretti N, Sedivy JM. Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. Aging Cell. 2013b:12(2):247–256. https://doi.org/10.1111/acel.12047.

De Cecco M, Criscione SW, Peterson AL, Neretti N, Sedivy JM, Kreiling JA. Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. Aging (Albany, NY). 2013a:5(12):867–883. https://doi.org/10.18632/aging.100621.

De Cecco M, Ito T, Petrashen AP, Elias AE, Skvir NJ, Criscione SW, Caligiana A, Brocculi G, Adney EM, Boeke JD, et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature. 2019:566(7742):73–78. https://doi.org/10.1038/s41586-018-0784-9.

Dej KJ, Gerasimova T, Corces VG, Boeke JD. A hotspot for the Drosophila gypsy retroelement in the ovo locus. Nucleic Acids Res. 1998:26(17):4019–4025. https://doi.org/10.1093/nar/26.17.4019.

Delanoue R, Clot C, Leray C, Pihl T, Hudry B. Y chromosome toxicity does not contribute to sex-specific differences in longevity. Nat Ecol Evol. 2023:7:1245–1256. https://doi.org/10.1038/s41559-023-02089-7.

Della Valle F, Reddy P, Yamamoto M, Liu P, Saera-Vila A, Bensaddek D, Zhang H, Prieto Martinez J, Abassi L, Celii M, et al. LINE-1 RNA causes heterochromatin erosion and is a target for amelioration of senescent phenotypes in progeroid syndromes. Sci Transl Med. 2022:14(657):eabl6057. https://doi.org/10.1126/scitranslmed.abl6057.

Deng H, Gao K, Jankovic J. The role of FUS gene variants in neurodegenerative diseases. Nat Rev Neurol. 2014:10(6):337–348. https://doi.org/10.1038/nrneurol.2014.78.

Deniz Ö, Ahmed M, Todd CD, Rio-Machin A, Dawson MA, Branco MR. Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. Nat Commun. 2020:11(1):3506. https://doi.org/10.1038/s41467-020-17206-4.

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. Nature. 1980:284(5757):601–603. https://doi.org/10.1038/284601a0.
Driver CJ, McKechnie SW. Transposable elements as a factor in the aging of *Drosophila melanogaster*. Ann N Y Acad Sci. 1992:673(1):83–91. https://doi.org/10.1111/j.1749-6632.1992.tb27439.x.
Dvořáková M, Karafiát V, Pajer P, Kluzáková E, Jarkovská K, Peková S, Krutílková L, Dvořák M. DNA released by leukemic cells contributes to the disruption of the bone marrow microenvironment. Oncogene. 2013:32(44):5201–5209. https://doi.org/10.1038/onc.2012.553.
Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. Science. 2016:351(6274):aac7247. https://doi.org/10.1126/science.aac7247.
TE Hub Consortium, Elliott TA, Heitkam T, Hubley R, Quesneville H, Suh A, Wheeler TJ. TE hub: a community-oriented space for sharing and connecting tools, data, resources, and methods for transposable element annotation. Mob DNA. 2021:12(1):16. https://doi.org/10.1186/s13100-021-00244-0.
Elsner D, Meusemann K, Korb J. Longevity and transposon defense, the case of termite reproductives. Proc Natl Acad Sci U S A. 2018:115(21):5504–5509. https://doi.org/10.1073/pnas.1804046115.
Fabian DK, Dönertaş HM, Fuentealba M, Partridge L, Thornton JM. Transposable element landscape in Drosophila populations selected for longevity. Genome Biol Evol. 2021:13(4): evab031. https://doi.org/10.1093/gbe/evab031.
Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, et al. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet. 2009:41(5):563–571. https://doi.org/10.1038/ng.368.
Finch CE. Variations in senescence and longevity include the possibility of negligible senescence. J Gerontol A Biol Sci Med Sci. 1998:53A(4):B235–B239. https://doi.org/10.1093/gerona/53A.4.B235.
Frost B, Hemberg M, Lewis J, Feany MB. Tau promotes neurodegeneration through global chromatin relaxation. Nat Neurosci. 2014:17(3):357–366. https://doi.org/10.1038/nn.3639.
García-Olmo DC, Domínguez C, García-Arranz M, Anker P, Stroun M, García-Verdugo JM, García-Olmo D. Cell-free nucleic acids circulating in the plasma of colorectal cancer patients induce the oncogenic transformation of susceptible cultured cells. Cancer Res. 2010:70(2):560–567. https://doi.org/10.1158/0008-5472.CAN-09-3513.
Gatti M, Pimpinelli S. Functional elements in *Drosophila melanogaster* heterochromatin. Annu Rev Genet. 1992:26(1):239–275. https://doi.org/10.1146/annurev.ge.26.120192.001323.
Gebert D, Rosenkranz D. RNA-based regulation of transposon expression. Wiley Interdiscip Rev RNA. 2015:6(6):687–708. https://doi.org/10.1002/wrna.1310.
Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler ELW, Zapp ML, Weng Z, et al. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science. 2008:320(5879):1077–1081. https://doi.org/10.1126/science.1157396.
Giordani G, Cavaliere V, Gargiulo G, Lattanzi G, Andrenacci D. Retrotransposons down- and up-regulation in aging somatic tissues. Cells. 2022:11(1):79. https://doi.org/10.3390/cells11010079.
González J, Karasov TL, Messer PW, Petrov DA. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila. PLoS Genet. 2010:6(4):e1000905. https://doi.org/10.1371/journal.pgen.1000905.
Gorbunova V, Seluanov A, Mita P, McKerrow W, Fenyö D, Boeke JD, Linker SB, Gage FH, Kreiling JA, Petrashen AP, et al. The role of retrotransposable elements in ageing and age-associated diseases. Nature. 2021:596(7870):43–53. https://doi.org/10.1038/s41586-021-03542-y.
Groza C, Chen X, Wheeler TJ, Bourque G, Goubert C. A unified framework to analyze transposable element insertion polymorphisms using graph genomes. Nat Commun. 2024:15(1):8915. https://doi.org/10.1038/s41467-024-53294-2.
Guio L, González J. New insights on the evolution of genome content: population dynamics of transposable elements in flies and humans. In: Anisimova M, editors. Evolutionary genomics: statistical and computational methods. New York, NY: Springer; 2019. p. 505–530. https://doi.org/10.1007/978-1-4939-9074-0_16.
Haldane JBS. New Paths in genetics. London, UK: Allen & Unwin; 1941.
Hickey DA. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics. 1982:101(3-4):519–531. https://doi.org/10.1093/genetics/101.3-4.519.
Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. The struggle for life of the genome's selfish architects. Biol Direct. 2011:6(1):19. https://doi.org/10.1186/1745-6150-6-19.
Hyun S. Small RNA pathways that protect the somatic genome. Int J Mol Sci. 2017:18(5):912. https://doi.org/10.3390/ijms18050912.
Jin K. Modern biological theories of aging. Aging Dis. 2010:1.
Jin Y, Tam OH, Paniagua E, Hammell M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics. 2015:31(22):3593–3599. https://doi.org/10.1093/bioinformatics/btv422.
Jintaridth P, Mutirangura A. Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. Physiol Genomics. 2010:41(2):194–200. https://doi.org/10.1152/physiolgenomics.00146.2009.
Jones BC, Wood JG, Chang C, Tam AD, Franklin MJ, Siegel ER, Helfand SL. A somatic piRNA pathway in the Drosophila fat body ensures metabolic homeostasis and normal lifespan. Nat Commun. 2016:7(1):13856. https://doi.org/10.1038/ncomms13856.
Jové M, Mota-Martorell N, Fernàndez-Bernal A, Portero-Otin M, Barja G, Pamplona R. Phenotypic molecular features of long-lived animal species. Free Radic Biol Med. 2023:208:728–747. https://doi.org/10.1016/j.freeradbiomed.2023.09.023.
Kelsey MMG, Kalekar RA, Sedivy JM. TE-Seq: a transposable element annotation and RNA-Seq pipeline. Preprint. bioRxiv. 2024. https://doi.org/10.1101/2024.10.11.617912.
Kirkwood TBL, Austad SN. Why do we age? Nature. 2000:408(6809):233–238. https://doi.org/10.1038/35041682.
Kissinger JC, DeBarry J. Genome cartography: charting the apicomplexan genome. Trends Parasitol. 2011:27(8):345–354. https://doi.org/10.1016/j.pt.2011.03.006.
Lanciano S, Cristofari G. Measuring and interpreting transposable element expression. Nat Rev Genet. 2020:21(12):721–736. https://doi.org/10.1038/s41576-020-0251-y.
Lanciano S, Cristofari G. Cancer immunotherapy: how to exploit transposable elements? Clin Chem. 2024:70(1):17–20. https://doi.org/10.1093/clinchem/hvad091.
LaRocca D, Lee J, West MD, Labat I, Sternberg H. No time to age: uncoupling aging from chronological time. Genes (Basel). 2021:12(5):611. https://doi.org/10.3390/genes12050611.
LaRocca TJ, Cavalier AN, Wahl D. Repetitive elements as a transcriptomic marker of aging: evidence in multiple datasets and models. Aging Cell. 2020:19(7):e13167. https://doi.org/10.1111/acel.13167.
Lee JC, Bruck DJ, Dreves AJ, Ioriatti C, Vogt H, Baufeld P. In focus: spotted wing Drosophila, *Drosophila suzukii*, across perspectives. Pest Manag Sci. 2011:67(11):1349–1351. https://doi.org/10.1002/ps.2271.
Lemaître J-F, Ronget V, Tidière M, Allainé D, Berger V, Cohas A, Colchero F, Conde DA, Garratt M, Liker A, et al. Sex differences in adult lifespan and aging rates of mortality across wild mammals.

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Proc Natl Acad Sci U S A. 2020:117(15):8546–8553. https://doi.org/10.1073/pnas.1911999117.
Lerat E, Fablet M, Modolo L, Lopez-Maestre H, Vieira C. TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. Nucleic Acids Res. 2017:45(4):e17. https://doi.org/10.1093/nar/gkw953.
Li W, Prazak L, Chatterjee N, Grüninger S, Krug L, Theodorou D, Dubnau J. Activation of transposable elements during aging and neuronal decline in Drosophila. Nat Neurosci. 2013:16(5):529–531. https://doi.org/10.1038/nn.3368.
Lim D-H, Oh C-T, Lee L, Hong J-S, Noh S-H, Hwang S, Kim S, Han S-J, Lee YS. The endogenous siRNA pathway in Drosophila impacts stress resistance and lifespan by regulating metabolic homeostasis. FEBS Lett. 2011:585(19):3079–3085. https://doi.org/10.1016/j.febslet.2011.08.034.
Liu N, Landreh M, Cao K, Abe M, Hendriks G-J, Kennerdell JR, Zhu Y, Wang L-S, Bonini NM. The microRNA miR-34 modulates ageing and neurodegeneration in Drosophila. Nature. 2012:482(7386):519–523. https://doi.org/10.1038/nature10810.
Liu X, Liu Z, Wu Z, Ren J, Fan Y, Sun L, Cao G, Niu Y, Zhang B, Ji Q, et al. Resurrection of endogenous retroviruses during aging reinforces senescence. Cell. 2023:186(2):287-304.e26. 10.1016/j.cell.2022.12.017.
Lock FE, Rebollo R, Miceli-Royer K, Gagnier L, Kuah S, Babaian A, Sistiaga-Poveda M, Lai CB, Nemirovsky O, Serrano I, et al. Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. Proc Natl Acad Sci U S A. 2014:111(34):E3534–E3543. https://doi.org/10.1073/pnas.1405507111.
López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013:153(6):1194–1217. https://doi.org/10.1016/j.cell.2013.05.039.
Loreto ELS, de Melo ES, Wallau GL, Gomes TMFF. The good, the bad and the ugly of transposable elements annotation tools. Genet Mol Biol. 2024:46(3 suppl 1):e20230138. https://doi.org/10.1590/1678-4685-GMB-2023-0138.
Mamillapalli A, Pathak RU, Garapati HS, Mishra RK. Transposable element 'roo' attaches to nuclear matrix of the Drosophila melanogaster. J Insect Sci. 2013:13(111):111. https://doi.org/10.1673/031.013.11101.
Marais GAB, Gaillard J-M, Vieira C, Plotton I, Sanlaville D, Gueyffier F, Lemaitre J-F. Sex gap in aging and longevity: can sex chromosomes play a role? Biol Sex Differ. 2018:9(1):33. https://doi.org/10.1186/s13293-018-0181-y.
Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. 2005:24(4):800–812. https://doi.org/10.1038/sj.emboj.7600545.
Mason JM, Biessmann H. The unusual telomeres of Drosophila. Trends Genet. 1995:11(2):58–62. https://doi.org/10.1016/S0168-9525(00)88998-2.
McClintock B. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A. 1950:36(6):344–355. https://doi.org/10.1073/pnas.36.6.344.
McClintock B. Induction of instability at selected loci in maize. Genetics. 1953:38(6):579–599. https://doi.org/10.1093/genetics/38.6.579.
Medawar PB. Old age and natural death. Mod. Q. 1946:1:30–56. https://doi.org/10.4324/9780429299759-2.
Medawar PB. An unsolved problem of biology. London, UK: H. K. Lewis; 1952.
Mérel V, Boulesteix M, Fablet M, Vieira C. Transposable elements in Drosophila. Mob DNA. 2020:11(1):23. https://doi.org/10.1186/s13100-020-00213-z.
Mohamed M, Sabot F, Varoqui M, Mugat B, Audouin K, Pélisson A, Fiston-Lavier A-S, Chambeyron S. TrEMOLO: accurate transposable element allele frequency estimation using long-read sequencing data combining assembly and mapping-based approaches. Genome Biol. 2023:24(1):63. https://doi.org/10.1186/s13059-023-02911-2.
Mosaddeghi P, Farahmandnejad M, Zarshenas MM. The role of transposable elements in aging and cancer. Biogerontology. 2023:24(4):479–491. https://doi.org/10.1007/s10522-023-10028-z.
Mostoslavsky R, Chua KF, Lombard DB, Pang WW, Fischer MR, Gellon L, Liu P, Mostoslavsky G, Franco S, Murphy MM, et al. Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. Cell. 2006:124(2):315–329. https://doi.org/10.1016/j.cell.2005.11.044.
Murray V. Are transposons a cause of ageing? Mutat Res. 1990:237(2):59–63. https://doi.org/10.1016/0921-8734(90)90011-F.
Nguyen AH, Bachtrog D. Toxic Y chromosome: increased repeat expression and age-associated heterochromatin loss in male Drosophila with a young Y chromosome. PLOS Genet. 2021:17(4):e1009438. https://doi.org/10.1371/journal.pgen.1009438.
Nikitin AG, Woodruff RC. Somatic movement of the manner transposable element and lifespan of Drosophila species. Mutat. Res. DNA Aging Gen. Instability Aging. 1995:338:43–49. https://doi.org/10.1016/0921-8734(95)00010-4.
Nirala NK, Li Q, Ghule PN, Chen H-J, Li R, Zhu LJ, Wang R, Rice NP, Mao J, Stein JL, et al. Hinfp is a guardian of the somatic genome by repressing transposable elements. Proc Natl Acad Sci U S A. 2021:118(41):e2100839118. https://doi.org/10.1073/pnas.2100839118.
Ochoa E, Ramirez P, Gonzalez E, De Mange J, Ray WJ, Bieniek KF, Frost B. Pathogenic tau-induced transposable element–derived dsRNA drives neuroinflammation. Sci Adv. 2023:9(1):eabq5423. https://doi.org/10.1126/sciadv.abq5423.
Oliveira DS, Fablet M, Larue A, Vallier A, Carareto CMA, Rebollo R, Vieira C. ChimeraTE: a pipeline to detect chimeric transcripts derived from genes and transposable elements. Nucleic Acids Res. 2023:51(18):9764–9784. https://doi.org/10.1093/nar/gkad671.
Orgel LE, Crick FHC. Selfish DNA: the ultimate parasite. Nature. 1980:284(5757):604–607. https://doi.org/10.1038/284604a0.
Pabis K, Barardo D, Sirbu O, Selvarajoo K, Gruber J, Kennedy BK. A concerted increase in readthrough and intron retention drives transposon expression during aging and senescence. eLife. 2024:12:RP87811. https://doi.org/10.7554/eLife.87811.
Pardue M-L, DeBaryshe PG. Drosophila telomeres: a variation on the telomerase theme. Fly (Austin). 2008:2(3):101–110. https://doi.org/10.4161/fly.6393.
Pehrsson EC, Choudhary MNK, Sundaram V, Wang T. The epigenomic landscape of transposable elements across normal human development and anatomy. Nat Commun. 2019:10(1):5640. https://doi.org/10.1038/s41467-019-13555-x.
Peona V, Palacios-Gimenez OM, Blommaert J, Liu J, Haryoko T, Jønsson KA, Irestedt M, Zhou Q, Jern P, Suh A. The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. Philos. Trans. R. Soc. B Biol. Sci. 2021:376(1833):20200186. https://doi.org/10.1098/rstb.2020.0186.
Perrat PN, DasGupta S, Wang J, Theurkauf W, Weng Z, Rosbash M, Waddell S. Transposition-driven genomic heterogeneity in the Drosophila brain. Science. 2013:340(6128):91–95. https://doi.org/10.1126/science.1231965.
Rahman R, Chirn G, Kanodia A, Sytnikova YA, Brembs B, Bergman CM, Lau NC. Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes. Nucleic Acids Res. 2015:43(22):10655–10672. https://doi.org/10.1093/nar/gkv1193.

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Ramirez P, Sun W, Dehkordi SK, Zare H, Fongang B, Bieniek KF, Frost B. Nanopore long-read sequencing unveils genomic disruptions in Alzheimer's disease. Preprint. bioRxiv. 2025:2024.02.01.578450, https://doi.org/10.1101/2024.02.01.578450 .

Rebollo R., Gerenton P., Cumunel E., Mary A., Sabot F., Burlet N., Gillet B., Hughes S., Oliveira S., Goubert D., et al. Identification and quantification of transposable element transcripts using long-read RNA-Seq in Drosophila germline tissues. Peer Community J. 2024:4:e89. https://doi.org/10.24072/pcjournal.457 .

Ricci M, Peona V, Boattini A, Taccioli C. Comparative analysis of bats and rodents' genomes suggests a relation between non-LTR retrotransposons, cancer incidence, and ageing. Sci Rep. 2023:13(1):9039. https://doi.org/10.1038/s41598-023-36006-6.

Rigal J, Martin Anduaga A, Bitman E, Rivellese E, Kadener S, Marr MT. Artificially stimulating retrotransposon activity increases mortality and accelerates a subset of aging phenotypes in Drosophila. eLife. 2022:11:e80169. https://doi.org/10.7554/eLife.80169.

Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, Santamarina M, Ju YS, Temes J, Garcia-Souto D, et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet. 2020:52(3):306–319. https://doi.org/10.1038/s41588-019-0562-0.

Sahm A, Cherkasov A, Liu H, Voronov D, Siniuk K, Schwarz R, Ohlenschläger O, Förste S, Bens M, Groth M, et al. The Greenland shark (*Somniosus microcephalus*) genome provides insights into extreme longevity. Preprint. BioRxiv. 2024:09.09.611499. https://doi.org/10.1101/2024.09.09.611499 .

Savitsky M, Kwon D, Georgiev P, Kalmykova A, Gvozdev V. Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline. Genes Dev. 2006:20(3):345–354. https://doi.org/10.1101/gad.370206.

Schaible R, Scheuerlein A, Dańko MJ, Gampe J, Martínez DE, Vaupel JW. Constant mortality and fertility over age in Hydra. Proc. Natl. Acad. Sci. 2015:112(51):15701–15706. https://doi.org/10.1073/pnas.1521002112.

Schmitz J, Brosius J. Exonization of transposed elements: a challenge and opportunity for evolution. Biochimie. 2011:93:1928–1934. https://doi.org/10.1016/j.biochi.2011.07.014.

Schneider BK, Sun S, Lee M, Li W, Skvir N, Neretti N, Vijg J, Secombe J. Expression of retrotransposons contributes to aging in Drosophila. Genetics. 2023:224(2):iyad073. https://doi.org/10.1093/genetics/iyad073.

Senapati P, Miyano M, Sayaman RW, Basam M, Leung A, LaBarge MA, Schones DE. Loss of epigenetic suppression of retrotransposons with oncogenic potential in aging mammary luminal epithelial cells. Genome Res. 2023:33(8):1229–1241. https://doi.org/10.1101/gr.277511.122.

Shah NM, Jang HJ, Liang Y, Maeng JH, Tzeng S-C, Wu A, Basri NL, Qu X, Fan C, Li A, et al. Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements. Nat Genet. 2023:55(4):631–639. https://doi.org/10.1038/s41588-023-01349-3.

Shmookler Reis RJ, Goldstein S. Mitochondrial DNA in mortal and immortal human cells. Genome number, integrity, and methylation. J Biol Chem. 1983:258(15):9078–9085. https://doi.org/10.1016/S0021-9258(17)44633-3.

Simon M, Meter MV, Ablaeva J, Ke Z, Gonzalez RS, Taguchi T, Cecco MD, Leonova KI, Kogan V, Helfand SL, et al. LINE1 derepression in aged wild-type and SIRT6-deficient mice drives inflammation. Cell Metab. 2019:29(4):871–885.e5. https://doi.org/10.1016/j.cmet.2019.02.014.

Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet. 2007:8(4):272–285. https://doi.org/10.1038/nrg2072.

Stow EC, Kaul T, deHaro DL, Dem MR, Beletsky AG, Morales ME, Du Q, LaRosa AJ, Yang H, Smither E, et al. Organ-, sex- and age-dependent patterns of endogenous L1 mRNA expression at a single locus resolution. Nucleic Acids Res. 2021:49(10):5813–5831. https://doi.org/10.1093/nar/gkab369.

Strobel E, Dunsmuir P, Rubin GM. Polymorphisms in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. Cell. 1979:17(2):429–439. https://doi.org/10.1016/0092-8674(79)90169-7.

Sturm Á, Perczel A, Ivics Z, Vellai T. The Piwi-piRNA pathway: road to immortality. Aging Cell. 2017:16(5):906–911. https://doi.org/10.1111/acel.12630.

Sulak M, Fong L, Mika K, Chigurupati S, Yon L, Mongan NP, Emes RD, Lynch VJ. TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. eLife. 2016:5:e11994. https://doi.org/10.7554/eLife.11994.

Takahashi F, Zhang C, Hohjoh H, Raveney B, Yamamura T, Hayashi N, Oki S. Immune-mediated neurodegenerative trait provoked by multimodal derepression of long-interspersed nuclear element-1. iScience. 2022:25(5):104278. https://doi.org/10.1016/j.isci.2022.104278.

Teefy BB, Adler A, Xu A, Hsu K, Singh PP, Benayoun BA. Dynamic regulation of gonadal transposon control across the lifespan of the naturally short-lived African turquoise killifish. Genome Res. 2023:33(1):141–153. https://doi.org/10.1101/gr.277301.122.

Teefy BB, Siebert S, Cazet JF, Lin H, Juliano CE. PIWI–piRNA pathway-mediated transposable element repression in Hydra somatic stem cells. RNA. 2020:26(5):550–563. https://doi.org/10.1261/rna.072835.119.

Teissandier A, Servant N, Barillot E, Bourc'his D. Tools and best practices for retrotransposon analysis using high-throughput sequencing data. Mob DNA. 2019:10(1):52. https://doi.org/10.1186/s13100-019-0192-1.

Teoli J, Merenciano M, Fablet M, Necsulea A, Siqueira-de-Oliveira D, Brandulas-Cammarata A, Labalme A, Lejeune H, Lemaitre J-F, Gueyffier F, et al. Transposable element expression with variation in sex chromosome number: insights into a toxic Y effect on human longevity. Preprint. BioRxiv. 2024:08.03.550779. https://doi.org/10.1101/2023.08.03.550779 .

Tossberg JT, Heinrich RM, Farley VM, Crooke PS, Aune TM. Adenosine-to-inosine RNA editing of Alu double-stranded (ds)RNAs is markedly decreased in multiple sclerosis and unedited Alu dsRNAs are potent activators of proinflammatory transcriptional responses. J. Immunol. 2020:205(10):2606–2617. https://doi.org/10.4049/jimmunol.2000384.

Treiber CD, Waddell S. Resolving the prevalence of somatic transposition in Drosophila. eLife. 2017:6:e28297. https://doi.org/10.7554/eLife.28297.

Trejo-Becerril C, Pérez-Cárdenas E, Taja-Chayeb L, Anker P, Herrera-Goepfert R, Medina-Velázquez LA, Hidalgo-Miranda A, Pérez-Montiel D, Chávez-Blanco A, Cruz-Velázquez J, et al. Cancer progression mediated by horizontal gene transfer in an in vivo model. PLoS One. 2012:7(12):e52754. https://doi.org/10.1371/journal.pone.0052754.

Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, Perry GH, Lynch VJ, Brown CD. Transposable elements are the primary source of novelty in primate gene regulation. Genome Res. 2017:27(10):1623–1633. https://doi.org/10.1101/gr.218149.116.

Tsai Y-T, Seymen N, Thompson IR, Zou X, Mumtaz W, Gerlevik S, Mufti GJ, Karimi MM. Expression of most retrotransposons in human blood correlates with biological aging. eLife. 2024:13:RP96575. https://doi.org/10.7554/eLife.96575.

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

van Bree EJ, Guimarães RLFP, Lundberg M, Blujdea ER, Rosenkrantz JL, White FTG, Poppinga J, Ferrer-Raventós P, Schneider A-FE, Clayton I, et al. A hidden layer of structural variation in transposable elements reveals potential genetic modifiers in human disease-risk loci. Genome Res. 2022:32(4):656–670. https://doi.org/10.1101/gr.275515.121.

Van Meter M, Kashyap M, Rezazadeh S, Geneva AJ, Morello TD, Seluanov A, Gorbunova V. SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. Nat Commun. 2014:5(1):5011. https://doi.org/10.1038/ncomms6011.

Vazquez BN, Thackray JK, Simonet NG, Chahar S, Kane-Goldsmith N, Newkirk SJ, Lee S, Xing J, Verzi MP, An W, et al. SIRT7 mediates L1 elements transcriptional repression and their association with the nuclear lamina. Nucleic Acids Res. 2019:47(15):7870–7885. https://doi.org/10.1093/nar/gkz519.

Vazquez JM, Lynch VJ. Pervasive duplication of tumor suppressors in Afrotherians during the evolution of large bodies and reduced cancer risk. eLife. 2021:10:e65041. https://doi.org/10.7554/eLife.65041.

Vijg J, Mullaart E, Lohman PHM, Knook DL. UV-induced unscheduled DNA synthesis in fibroblasts of aging inbred rats. Mutat. Res. Repair Rep. 1985:146:197–204. https://doi.org/10.1016/0167-8817(85)90011-2.

Villanueva-Cañas JL, Horvath V, Aguilera L, González J. Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in Drosophila melanogaster. Nucleic Acids Res. 2019:47(13):6842–6857. https://doi.org/10.1093/nar/gkz490.

Warmuth VM, Weissensteiner MH, Wolf JBW. Accumulation and ineffective silencing of transposable elements on an avian W chromosome. Genome Res. 2022:32(4):671–681. https://doi.org/10.1101/gr.275465.121.

Warren IA, Naville M, Chalopin D, Levin P, Berger CS, Galiana D, Volff J-N. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. Chromosome Res. 2015:23(3):505–531. https://doi.org/10.1007/s10577-015-9493-5.

Weilguny L, Kofler R. DeviaTE: assembly-free analysis and visualization of mobile genetic element composition. Mol Ecol Resour. 2019:19(5):1346–1354. https://doi.org/10.1111/1755-0998.13030.

Weismann A. Ueber die Dauer des Lebens: ein Vortrag. Jena, Germany: G. Fischer; 1882.

Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. Annu Rev Genet. 2020:54(1):539–561. https://doi.org/10.1146/annurev-genet-040620-022145.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007:8(12):973–982. https://doi.org/10.1038/nrg2165.

Williams GC. Pleiotropy, natural selection, and the evolution of senescence. Evolution. 1957:11(4):398–411. https://doi.org/10.1111/j.1558-5646.1957.tb02911.x.

Wood JG, Helfand SL. Chromatin structure and transposable elements in organismal aging. Front Genet. 2013:4:274. https://doi.org/10.3389/fgene.2013.00274.

Wood JG, Jones BC, Jiang N, Chang C, Hosier S, Wickremesinghe P, Garcia M, Hartnett DA, Burhenn L, Neretti N, et al. Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in Drosophila. Proc Natl Acad Sci U S A. 2016:113(40):11277–11282. https://doi.org/10.1073/pnas.1604621113.

Woodruff RC, Nikitin AG. P DNA element movement in somatic cells reduces lifespan in *Drosophila melanogaster*: evidence in support of the somatic mutation theory of aging. Mutat Res. 1995:338(1-6):35–42. https://doi.org/10.1016/0921-8734(95)00009-u.

Xia B, Zhang W, Zhao G, Zhang X, Bai J, Brosh R, Wudzinska A, Huang E, Ashe H, Ellis G, et al. On the genetic basis of tail-loss evolution in humans and apes. Nature. 2024:626(8001):1042–1048. https://doi.org/10.1038/s41586-024-07095-8.

Xu Y, Tang Y, Feng W, Yang Y, Cui Z. Comparative analysis of transposable elements reveals the diversity of transposable elements in decapoda and their effects on genomic evolution. Mar. Biotechnol. 2023:25(6):1136–1146. https://doi.org/10.1007/s10126-023-10265-w.

Yang N, Srivastav SP, Rahman R, Ma Q, Dayama G, Li S, Chinen M, Lei EP, Rosbash M, Lau NC. Transposable element landscapes in aging Drosophila. PLOS Genet. 2022:18(3):e1010024. https://doi.org/10.1371/journal.pgen.1010024.

Yang WR, Ardeljan D, Pacyna CN, Payer LM, Burns KH. SQuIRE reveals locus-specific regulation of interspersed repeat expression. Nucleic Acids Res. 2019:47(5):e27. https://doi.org/10.1093/nar/gky1301.

Yushkova E, Moskalev A. Transposable elements and their role in aging. Ageing Res Rev. 2023:86:101881. https://doi.org/10.1016/j.arr.2023.101881.

Zhao K, Du J, Peng Y, Li P, Wang S, Wang Y, Hou J, Kang J, Zheng W, Hua S, et al. LINE1 contributes to autoimmunity through both RIG-I- and MDA5-mediated RNA sensing pathways. J Autoimmun. 2018:90:105–115. https://doi.org/10.1016/j.jaut.2018.02.007.

Zhao Y, Oreskovic E, Zhang Q, Lu Q, Gilman A, Lin YS, He J, Zheng Z, Lu JY, Lee J, et al. Transposon-triggered innate immune response confers cancer resistance to the blind mole rat. Nat Immunol. 2021:22(10):1219–1230. https://doi.org/10.1038/s41590-021-01027-8.

Zhen Z, Chen Y, Wang H, Tang H, Zhang H, Liu H, Jiang Y, Mao Z. Nuclear cGAS restricts L1 retrotransposition by promoting TRIM41-mediated ORF2p ubiquitination and degradation. Nat Commun. 2023:14(1):8217. https://doi.org/10.1038/s41467-023-43001-y.

Zhu C-T, Chang C, Reenan RA, Helfand SL. Indy gene variation in natural populations confers fitness advantage and life span extension through transposon insertion. Aging (Albany NY). 2014:6(1):58–69. https://doi.org/10.18632/aging.100634.

Zhuang J, Wang J, Theurkauf W, Weng Z. TEMP: a computational method for analyzing transposable element polymorphism in populations. Nucleic Acids Res. 2014:42(11):6826–6838. https://doi.org/10.1093/nar/gku323.

Associate editor: Rachel Mueller

Downloaded from https://academic.oup.com/gbe/article/17/6/evaf088/8133184 by guest on 02 June 2025

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Annexe 4

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1

Intricate interactions between antiviral immunity and transposable element control in *Drosophila*

Camille A Mayeux[1], Anaïs Larue[1,2], Daniel S. Oliveira[1,3], Marion Varoqui[4], Hélène Henri[1], Rita Rebollo[2], Natacha Kremer[1], Séverine Chambeyron[4] and Marie Fablet[1,5,*]

[1] Laboratoire de Biométrie et Biologie Evolutive; Université Lyon 1; CNRS; VAS; UMR 5558, Villeurbanne, France

[2] Univ Lyon, INRAE, INSA-Lyon, BF2I, UMR 203, 69621 Villeurbanne, France

[3] Institute of Biosciences, Humanities and Exact Sciences, São Paulo State University (Unesp), São José do Rio Preto, São Paulo, 15054-000, Brazil

[4] Institute of Human Genetics, UMR9002, CNRS and Université de Montpellier, Montpellier, France

[5] Institut Universitaire de France

* Corresponding author: marie.fablet@univ-lyon1.fr

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Transposable elements (TEs) are parasite DNA sequences that are controlled by RNA interference pathways in many organisms. In insects, antiviral immunity is also achieved by the action of small RNAs. In the present study, we analyzed the impacts of an infection with Drosophila C Virus (DCV) and found that TEs are involved in a dual response: on the one hand TE control is released upon DCV infection, and on the other hand TE transcripts help the host reduce viral replication. This discovery highlights the intricate interactions in the arms race between host, genomic parasites, and viral pathogens.

Significance statement

Transposable elements (TEs) are widespread components of all genomes. They were long considered as mere DNA parasites but are now acknowledged as major sources of genetic diversity and phenotypic innovations. Using *Drosophila* C virus, here we show that TEs are at the center of defense and counter-attack between host and virus. On the one hand, TE control is released upon viral infection, and on the other hand, TE transcripts help the host reduce viral replication. To our knowledge, this is the first time such a complex host-pathogen interaction involving TEs is shown.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Transposable elements (TEs) are DNA sequences considered as genomic parasites because they have the ability to move and multiply along chromosomes at the expense of the host. TEs are very diverse in structure (number of genes, length, presence of repeats, *etc*.) and abundance, from a few percent in the honeybee *Apis mellifera* (1) to ~ 80% of the recently sequenced arctic krill genome (2). Tolerance to TEs is possible thanks to epigenetic mechanisms that inhibit their activity, and there is a strong selective pressure for these mechanisms to be particularly efficient in gonads, where the genomes of the next generation lie (3–5). In insects, this control is essentially achieved through RNA interference (RNAi) pathways, and in particular *via* the piRNA pathway. piRNAs are 23-30 nt-long single-stranded RNAs that target TE sequences through sequence complementarity. They form complexes with proteins displaying RNAse activity or triggering heterochromatinization (6–8).

RNAi is also the first line of defense against viruses in Arthropods (9–12). Antiviral immunity relies on small RNAs known as siRNAs, which are 21 nt-long single-stranded RNAs, produced by Dicer-2 from double-stranded RNA templates. Whereas siRNAs are known to play a role in the somatic control of TE activity (13, 14), the intricate interplay between TE regulation and antiviral immunity through RNAi remains largely unexplored. Upon viral infection, viral RNA genomes are converted to DNA by TE reverse transcriptases, and this process leads to the increased production of viral siRNAs, facilitating the establishment of persistent infection (15, 16). Moreover, the activation of the *mdg4* TE at the pupal stage was recently found to prime the host's antiviral immunity for adult stage (17). In addition, in a previous investigation in *Drosophila* using the Sindbis virus (SINV), which naturally infects mosquitoes, we demonstrated that the host antiviral response triggered an amplified production of TE-derived small RNAs. This resulted in the reduction of TE transcript amounts (18). In the present study, we investigated the more closely intertwined interaction between the *Drosophila* host and its natural pathogen Drosophila C virus (DCV).

DCV is a non-enveloped virus that belongs to the Dicistroviridae family. It has a 9.2 kb-long, single-stranded RNA genome of positive polarity (19). DCV is a natural pathogen of *Drosophila* (20, 21), and is horizontally transferred through oral infection. A few days after infection, and independently of the route of infection –either oral or systemic through experimental injections–, DCV viral particles can be detected in the fat body, trachea and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

visceral muscle of the crop, midgut, hindgut, and gonads (22). Systemic DCV infection causes intestinal obstruction that leads to fly death (23). DCV was first described in *D. melanogaster* but it can also naturally infect *Drosophila simulans* (20). The 99 most N-terminal residues of DCV ORF1 –encoding non-structural proteins– correspond to 1A, a viral suppressor of RNA interference (VSR) that binds long double-stranded RNAs (dsRNAs) and therefore prevents Dicer2-mediated production of siRNAs (24, 25).

Here we show that DCV infection disrupts the regulation of TEs in a more intricate way than what was observed with SINV (18). This is particularly pronounced in the *D. melanogaster* strain, as compared to the *D. simulans* strain. In addition, we show that hosts with a higher TE load in the transcriptome display a more efficient antiviral immunity. This study highlights the prominent contribution of TEs in the immune response of *Drosophila*.

Results and Discussion

Host response to DCV infection in *D. melanogaster* w^{1118} and *D. simulans* Makindu

We infected flies by intra-thoracic pricking (26), and followed fly mortality and DCV replication. The infection rapidly led to fly death, and significantly more rapidly in *D. simulans* Makindu (log-rank test, p-value = 0.008). On average, half of the flies were dead at 5.4 days post infection (dpi) in *D. melanogaster* w^{1118}, and 3.8 dpi in *D. simulans* Makindu (Fig. 1A). Accordingly, RT-qPCR revealed that DCV replication was 2.6-fold higher in *D. simulans* Makindu compared to *D. melanogaster* w^{1118} at 4 dpi, which corresponds to the peak of viral titer (Fig. 1B).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

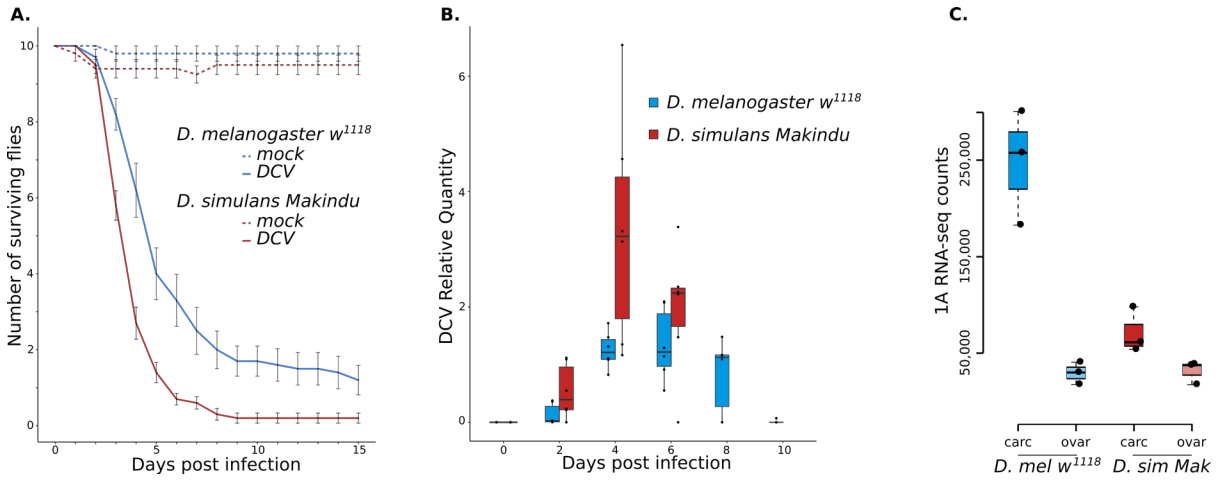


Figure 1. Response to DCV infection in *D. melanogaster* w^{1118} and *D. simulans* Makindu. **A.** Fly survival upon DCV infection. **B.** Kinetics of DCV titers followed using RT-qPCR. **C.** Raw RNA-seq read counts mapping against 1A sequence. RNA-seq read mapping along the sequence of DCV genome is shown in Fig S1.

We performed RNA-seq and small RNA-seq experiments at the peak of viral titer, *i.e.,* 4 dpi in both species, in dissected ovaries and the rest of the body –hereafter called “carcasses”. DCV RNAs were highly abundant among the sequenced molecules in carcasses. We could also detect DCV RNAs in ovaries, however in much lower amounts, potentially corresponding to the presence of DCV in the muscle cells of the ovarian peritoneal sheath, as described by Ferreira *et al.* (22).

As previously described (23), DCV infection by pricking leads to many down-regulated genes in *D. melanogaster* w^{1118} somatic tissues, as well as in *D. simulans* Makindu. In particular, down-regulated genes were enriched in genes related to metabolic functions, whereas the few activated genes were enriched in stress-response activity, as reported by Chtarbanova *et al.* (23), but also in immune processes (Fig S2).

It should be noticed that we cannot directly draw conclusions on a *D. melanogaster versus D. simulans* species comparison based on only one strain per species. Instead, it would require the inclusion of several strains for each species. Nevertheless, we have to acknowledge the differences in patterns between *D. melanogaster* w^{1118} and *D. simulans* Makindu, which could correspond to several hypotheses. i) The sequencing data were produced at 4 dpi for both strain, which corresponds to a higher number of dead flies and a stronger viral replication in *D. simulans* Makindu. Therefore, 4 dpi seems to be further along in the infection process in *D. simulans* Makindu. ii) It is possible that the infection protocol

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

leads to acute infection in *D. simulans* Makindu while persistence is achieved in *D. melanogaster* w^{1118}. iii) DCV was first described as a *D. melanogaster* pathogen. The differences between *D. melanogaster* and *D. simulans* outcomes may be due to *D. simulans* not being its preferential host, even though DCV has been found in natural samples of *D. simulans* (20).

DCV infection impacts the transcript amounts of transposable elements

We analyzed the amounts of transcripts for 237 TE families in the carcasses of the different samples, and we found a global significant increase upon DCV infection in *D. melanogaster* w^{1118} (median per TE family log2 Fold Change (log2FC) = 0.28, paired Wilcoxon test: p-value = 8e-12). This is in agreement with our previous reanalysis of data produced from *D. melanogaster* y^{1} fat bodies upon DCV infection, which also showed a clear increase in TE transcript amounts (18, 27). In the present data, forty three TE families are significantly upregulated upon DCV infection (DESeq2 adjusted p-value < 0.05), whereas seven are significantly down-regulated (Fig. 2A; See Fig S3 for TE family details). We could not detect significant differences across TE classes (Table S4). In *D. simulans* Makindu, there is virtually no modulation: we observed a significant increase for six TE families and a significant decrease for four TE families (Fig. 2B).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

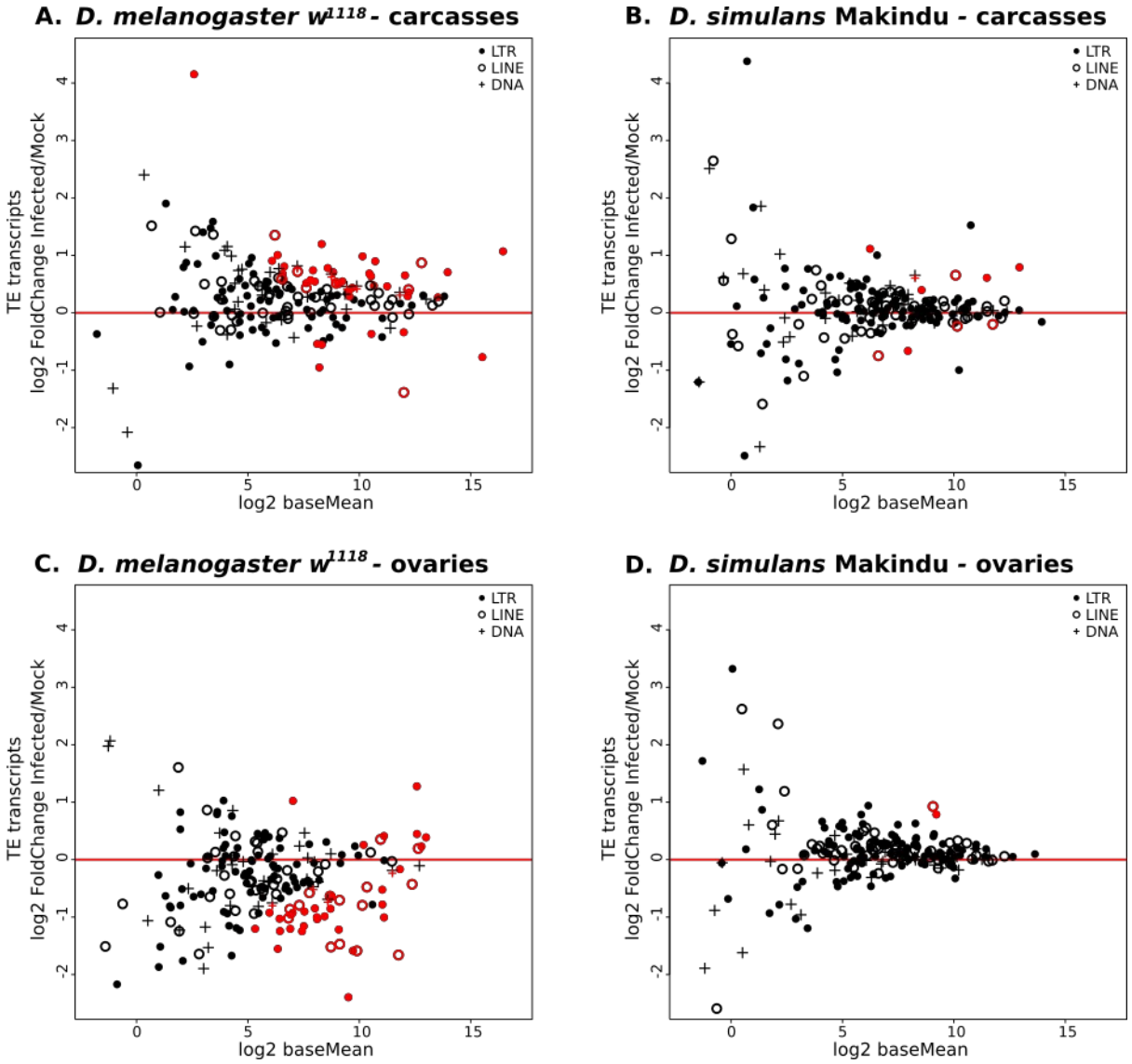


Figure 2. TE transcript modulation upon DCV infection. A. TE transcripts log2FC between DCV infected and mock conditions. Dot shapes indicate TE classes, red line is log2FC=0, *i.e.* no modulation. Red dots are TE families displaying significant differential expression at the 0.05 level for DESeq2 adjusted p-values. **A.** *D. melanogaster* w^{1118} carcasses, **B.** *D. simulans* Makindu carcasses, **C.** *D. melanogaster* w^{1118} ovaries, **D.** *D. simulans* Makindu ovaries.

In ovaries, we observed an opposite pattern for *D. melanogaster* w^{1118} samples, with a clear decrease of TE transcript amounts: median per TE family log2FC = -0.38 (paired Wilcoxon test : p-value = 1e-9). Transcript amounts increased significantly for nine TE families whereas they significantly decreased for 41 TE families (Fig. 2C). In *D. simulans* Makindu, DCV infection had very little impact on ovarian TE transcript amounts (median per TE family log2FC = 0.10, paired Wilcoxon test : p-value = 5e-9) (Fig. 2D).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In order to understand the mechanisms underlying changes in TE transcript amounts upon DCV infection, we analyzed small RNA-sequencing data from the same experimental conditions. The data showed the expected patterns in size distribution and nucleotide composition (Fig S5-S6). Ping-pong signatures were detected in all conditions, indicating that flies infected with DCV could still produce functional piRNAs (Fig S7). TE-derived siRNAs and piRNAs of both polarities clearly increased upon DCV infection in carcasses of *D. melanogaster* w^{1118} and *D. simulans* Makindu, whereas we detected very low size effects in ovaries (Fig S8-S10). The increase was stronger in piRNAs compared to siRNAs (all TE families together, median log2FC: piRNAs: [0.25; 1.28], siRNAs: [0.05, 0.75], Fig S11).

When TE control is achieved *via* small RNA interference, the expectation is that an increase in small RNAs leads to a decrease in RNA amounts, and reciprocally. However, here, the TE families displaying small RNA increase upon infection are globally not significantly enriched in downregulated families, and this was consistent across all conditions (Table S12). Such a result reveals a dissociation between TE transcript amounts and TE-derived small RNAs in the present experimental system.

It has been demonstrated that viral infections in *D. melanogaster* trigger the active uptake of dsRNA molecules of viral origin in haemocytes, which fuels the siRNA machinery and allows the systemic production of antiviral siRNAs (16, 28). Using Sindbis virus (SINV), we previously proposed that dsRNA molecules of TE sequences hitchhike the uptake pathway, and are opportunistically imported within haemocytes along with viral dsRNAs. This would then allow the enhancement of TE-derived small RNA production and explain the observed systemic increase of TE-derived small RNAs, which leads to a decrease in TE transcript amounts (18). Here, we found a global increase of both TE RNAs and TE-derived small RNAs in *D. melanogaster* w^{1118} carcasses. We propose that such a discrepancy is due to 1A, the VSR encoded by DCV (no VSR has been described for SINV). In the tissues where DCV is expressed, 1A is produced and binds long dsRNAs, however without any sequence specificity (24, 25, 29). Therefore, 1A mostly binds viral dsRNAs, but also other dsRNAs such as those produced from TE sequences, either due to antisense transcripts annealing to sense RNAs or to single-stranded transcripts folding back on repeated regions. Eventually, the production of TE-derived siRNAs is prevented, resulting in increased TE RNA amount. This is in agreement with the strong increase in TE transcript amounts that we previously observed in *dcr2* mutants upon SINV infection (18). Therefore, we propose that 1A is responsible for the decorrelation between TE RNA and siRNA amounts observed above. TE-derived small RNA amounts would thus result from these two mechanisms with opposed effects: the activation of the dsRNA uptake pathway leads to piRNA and siRNA increase

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

while the action of 1A leads to the reduction of siRNA production. Indeed, since 1A specifically binds dsRNA molecules, it is expected to interfere mainly with siRNA production, which explains the resulting increase in small RNAs is lower for siRNAs compared to piRNAs. In the case of the previously studied SINV infection, the tripartite interaction between host, virus and TEs is simpler and only results from TEs hitchhiking the dsRNA uptake pathway. Here, the production of a VSR adds another layer of interaction. This ends up in an intricate arms race between the virus, the host and its TEs (Fig. 3). Moreover, we propose that this explains the virtual absence of any TE phenotype in *D. simulans* Makindu. Indeed, the RNA-seq data reveal that the number of reads mapping against the sequence corresponding to 1A is the highest in *D. melanogaster* w^{1118} carcasses, and very low in ovarian samples. In carcasses, the number of reads mapping against 1A is significantly higher in *D. melanogaster* w^{1118} compared to *D. simulans* Makindu (Fig 1C). These species-specific differences may be due to contrasted efficiencies in the production of 1A.

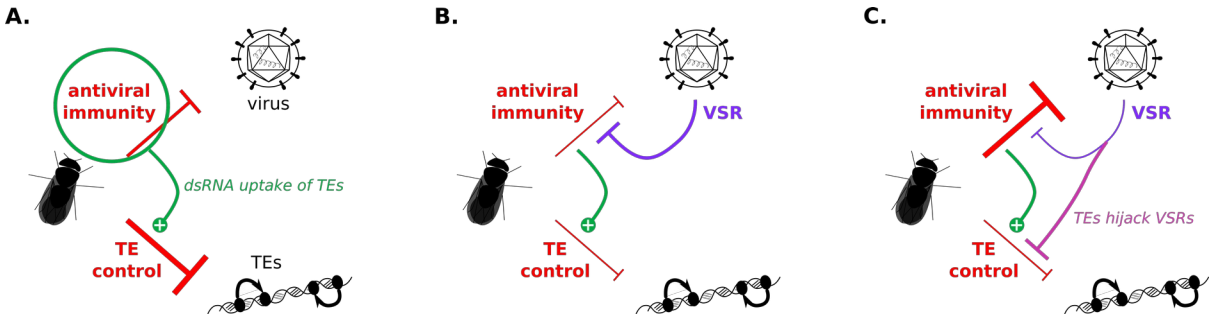


Figure 3. Reciprocal impacts of TE control and antiviral immunity. A. The fly host fights against viruses and TEs. Using SINV, we recently demonstrated that antiviral immunity enhanced TE control *via* the dsRNA uptake pathway. **B.** However, some viruses encode VSRs that inhibit antiviral immunity, such as the 1A protein produced by DCV. **C.** Here we propose that TEs hijack VSRs, which allows the release of TE control.

TE upregulation leads to reduced DCV titers

On the other side, we wondered whether TEs could modify DCV titers. Such a study is rather difficult because it requires to make TE transcript amounts vary within the same genetic background. We thus used a sophisticated transgenic *D. melanogaster* strain, which displays an inducible *piwi* knock-down in the somatic cells surrounding the ovary (30–32)

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[constitutive *piwi* knock-down leads to fly sterility and thus prevents from obtaining progeny (33, 34)]. The repetitive, punctual knock-down of *piwi* along 73 successive generations – hereafter named S73, S for shift in temperature causing *piwi* knock-down– led to the accumulation of TE copies from the ZAM and gtwin families (31, 32). It was accompanied by an increase in transcript amounts for these TE families, as compared to G0-F100 –no temperature shift corresponding to the original strain– whereas gene expression remained largely unaffected (Fig S13). It has to be noted that S73 is not an isogenic strain. Instead, flies are raised as large cohorts along generations, ensuring the lowest genetic drift, and resulting in pools of flies displaying distinct, low frequency TE insertions within the same genetic background. Upon intra-thoracic injection, we found that DCV replicated earlier but at strongly reduced levels in S73 compared to G0-F100 (DCV_relative_quantity ~ dpi + infection_status + strain; mean strain difference = 0.23; strain effect p-value = 0.004) (Fig. 4A). Fifteen days after injection, mean DCV RNA levels were 8.1-fold higher in G0-F100. Fly death rates were higher in S73 at the beginning of the infection, but the curves inverted around day 15 (log-rank test, p-value = 0.887) (Fig. 4B). Using RT-qPCR, we followed ZAM and gtwin somatic transcript amounts in both strains and both conditions (Fig. 4C and 4D). Similar to the global increase in TE transcript amounts observed in *D. melanogaster* w^{1118}, we could detect that ZAM and gtwin transcript amounts in S73 are significantly higher upon DCV infection compared to mock samples around the peak of DCV titer.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

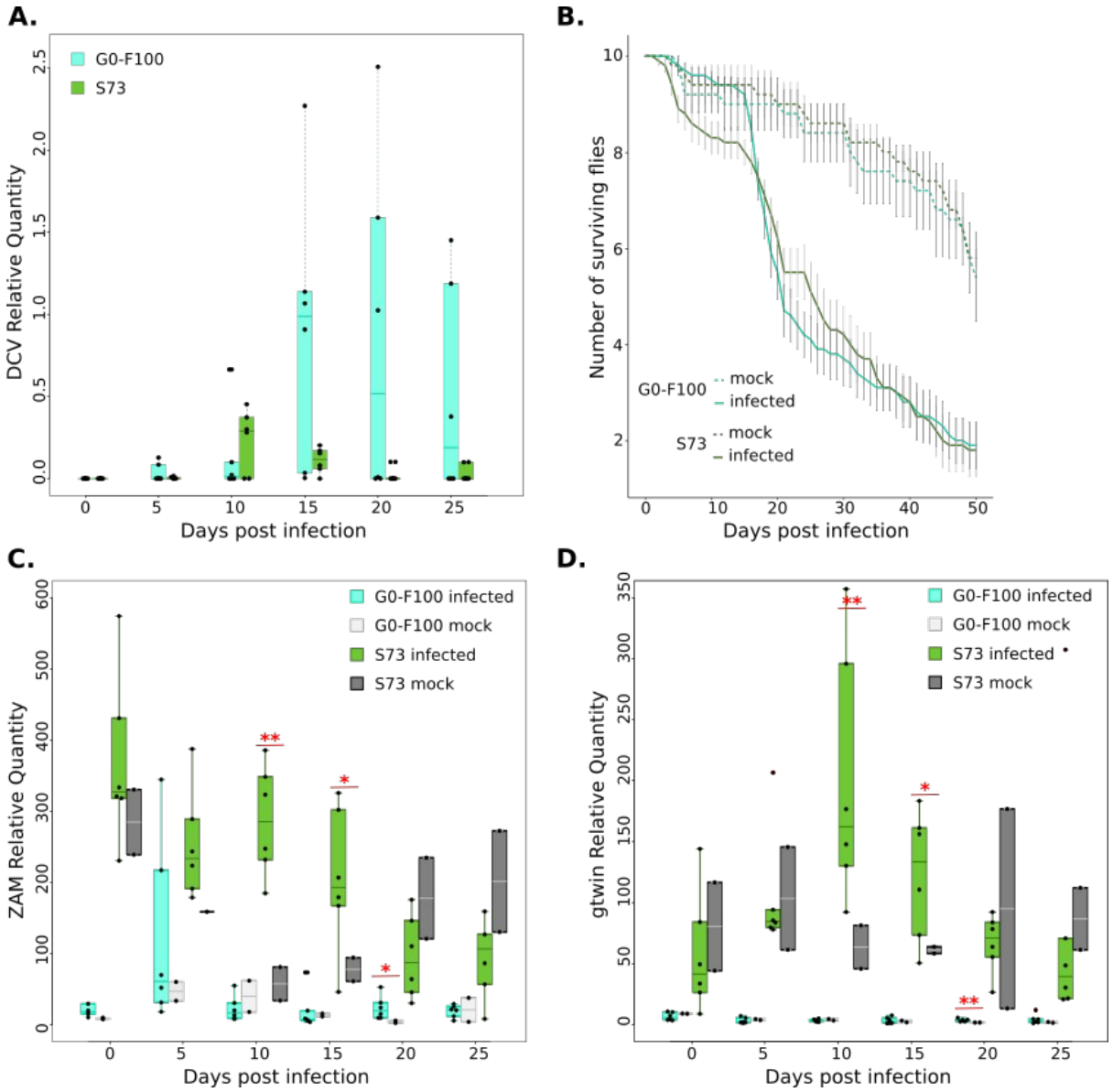


Figure 4. Response to DCV infection in S73 and G0-100. A. DCV titer in G0-F100 and S73 measured using RT-qPCR. **B.** Fly survival upon DCV and mock infections in G0-F100 and S73. **C.** ZAM transcript quantification upon DCV and mock infection in G0-F100 and S73. Stars indicate significant differences between infected and mock conditions using t-tests (p-value: 0.05 * 0.01 **). **D.** gtwin transcript quantification upon DCV and mock infection in G0-F100 and S73. Stars indicate significant differences between infected and mock conditions using t-tests (p-value: 0.05 * 0.01 **).

We propose that the higher TE transcript amounts in S73 produce increased amounts of dsRNA molecules, which can titrate 1A. The reduction in 1A availability removes DCV protection against the fly RNAi machinery; this allows the host to produce more siRNAs against the virus and therefore set up a more efficient RNAi response. This ends up in

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reduced viral loads (Fig. 3).

Altogether, our results suggest that a higher TE load could be beneficial in case of viral infection. At a larger evolutionary scale, such results may shed a new light on *Aedes* mosquitoes, which are known as major vectors of arboviruses –which cause no harm to them–, and are described to carry a large TE load (*ca* half of their genome). Finally, our results uncover TEs as major players of a complex host-pathogen interaction built along long-lasting coevolution.

Material and Methods

Drosophila strains and rearing conditions

D. simulans Makindu strain was previously described by Akkouche *et al.* (35) and Fablet *et al.* (36). G0-F100 and S73 transgenic *D. melanogaster* flies were previously described by Barckmann *et al.* (30) and Mohamed *et al.* (31). All experiments were performed using 3-6 day-old mated females.

Flies were reared on corn medium and maintained under standard laboratory conditions: 12/12 L/D cycle, 25°C (or 18°C for G0-F100 and S73, in order not to induce *piwi* knock-down) and 70% RH. Chronic viral infections were eliminated by bleaching the eggs, as described previously (37), except for G0-F100 and S73. Three to 6-hour-old eggs were incubated in 50% household bleach (2.6% active chlorine) for 10 minutes, washed three times for 5 minutes in deionised water and then transferred to fresh medium for adult emergence. As expected after the treatment, we could not detect amplification corresponding to the viruses using RT-PCR (SuperScript™ IV VILO™ Master Mix (without ezDNase enzyme treatment) kit and DreamTaq DNA polymerase) and the following primers: *Drosophila* A virus (DAV): 5′- AGGAGTTGGTGAGGACAGCCCA -3′ and 5′- AGACCTCAGTTGGCAGTTCGCC -3′, Nora Virus (NV): 5′- ATGGCGCCAGTTAGTGCAGACCT -3′ and 5′- CCTGTTGTTCCAGTTGGGTTCGA -3′, *Drosophila melanogaster* Sigma Virus (DmelSV): 5'- ATGTAACTCGGGTGTGACAG -3' and 5'- CCTTCGTTCATCCTCCTGAG -3) (37). *rp49* was used as control (5′- CGGATCGATATGCTAAGCTGT -3′ and 5′- GCGCTTGTTCGATCCGTA -3′) (38). To eliminate the *Wolbachia* endosymbiotic bacteria, 3- to 6-hour-old eggs were collected and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

placed on fresh standard medium containing 0.25 mg/mL tetracycline hypochloride (Sigma-Aldrich), for all *Drosophila* strains. The treatment was performed during two generations and then three generations recovered on standard medium without treatment. We validated the absence of amplification by PCR using *Wolbachia 16S* primers (5′-TTGTAGCCTGCTATGGTATAACT -3′ and 5′- GAATAGGTATGATTTTCATGT -3′) (39), *Wolbachia wsp* (5′- TGGTCCAATAAGTGATGAAGAAAC -3′ and 5′-AAAAATTAAACGCTACTCCA -3′) (40) and *FtsZ* (5'- CGAGATGGGCAAAGCGATGA -3' and 5'- ATTCCTTGCGCACCTTTCAT -3' (41)).

Virus production and quantification

DCV was produced and titrated in Schneider's *Drosophila* Line DL2 cells, both kindly provided by Luis Teixeira/Ewa Chrostek. DL2 cells were kept in Schneider's *Drosophila* Medium supplemented with 10% Fetal Bovine Serum, 1% penicillin-streptomycin 10,000 U/mL (all Gibco). Seven days after infection (MOI: 2) in Schneider's *Drosophila* Medium supplemented with 1% penicillin-streptomycin 10,000 U/mL, the cell culture was collected and frozen at -80°C for 40 min. The culture was thawed, frozen again at -80°C for 40 min, and thawed to disrupt cells, and centrifuged at 4000 g for 10 min to remove cell debris. The supernatant was aliquoted and stored at -80°C. DCV titer was calculated by the Reed and Muench end-point calculation method: DL2 confluent cells in 96-well plates were infected with a serial 10-fold dilutions of virus suspension. The presence of active DCV was scored by cell death or clear cytopathic effects, resulting in a DCV titer of $4.22x10^{9}$ $TCID_{50}$/mL. Similar dilutions with extracts of DL2 cells that were not inoculated with DCV did not cause any cytopathic effect in culture (42).

Virus inoculation and survival assays

CO_2-anesthetized flies were pricked in the left pleural suture on the thorax with a 0.15 mm diameter anodized steel needle dipped into the DCV solution, as described in Martinez *et al.* (26). Flies were pricked with the 'mock' solution –extract from uninfected DL2 cells– to control for the absence of mortality in absence of DCV infection. For the survival assays, 100 infected females and 50 'mock' females were kept in rich medium vials, 10 flies per vial. Flies were transferred to fresh media every 3-4 days. The number of dead flies was recorded each day after infection.

It has to be noted that infection experiments were performed at 25°C for *D. simulans* Makindu and *D. melanogaster* w^{1118} but at 18°C for *D. melanogaster* G0-F100 and S73 in

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

order not to induce *piwi* knock-down (30). This difference in temperature leads to differences in DCV replication kinetics and in lifespan across these experiments. Accordingly, each infected sample should be analyzed in comparison with the corresponding mock sample obtained in the exact same experimental conditions.

RNA extraction and RT-qPCR

To quantify DCV and TE expression, total RNAs were extracted individually for 6 infected and 2 mock flies at 0, 2, 4, 6 and 8 dpi using Qiazol and the RNeasy mini kit (Qiagen). Purified RNAs were processed using Turbo DNase (Ambion DNAfree kit). Reverse transcription was performed on 4 µL of total RNAs using SuperScript IV VILO Master Mix (Invitrogen). qPCR was performed using 2 µL of cDNA, 0.5 µL of each primer (DCV : 5'-GACACTGCCTTTGATTAG -3' and 5'- CCCTCTGGGAACTAAATG -3', ZAM : 5'-CTACGAAATGGCAAGATTAATTCCACTTCC -3' and 5'-CCCGTTTCCTTTATGTCGCAGTAGCT -3', gtwin : 5'- TTCGCACAAGCGATGATAAG -3' and 5'- GATTGTTGTACGGCGACCTT -3') and 5 µL of SsoADV SYBR® Green Supermix (Bio Rad) in a QuantStudio™ 6 Flex Real-Time PCR System (Applied Biosystems™) and following : 30 sec at 95 °C; 40 cycles of 95 °C for 15 sec and 58 °C for 30 sec. For standardization, we tested the *rp49*, *Adh*, and *EF1* genes, and kept *rp49* for further experiments because it displayed the highest stability across all conditions.

RNA-seq

Thirty pairs of ovaries were carefully and manually dissected and separated from the rest of the bodies ('carcasses') at 4 dpi. Total RNAs from 30 pairs of ovaries or 30 carcasses were extracted according to the protocol described above, using Qiazol, the RNeasy Mini Kit (Qiagen) and TurboDNAse (Ambion DNAfree kit) on the total amount of eluate to eliminate residual DNA. RNA quality was validated using a Bioanalyzer (Agilent). Library preparation was performed at the GenomEast platform at the Institute of Genetics and Molecular and Cellular Biology (IGBMC, France) using Illumina Stranded mRNA Prep Ligation - Reference Guide - PN 1000000124518. Libraries were sequenced as paired-end 100 base reads (HiSeq 4000: Illumina).

For S73 and G0-100 samples, total RNA was extracted from 3-16 h embryos using TRIzol. 4 mg of total RNA was subjected to Ribo-Zero ribosomal depletion. RNA was further purified using RNA Clean & Concentrator-5. Libraries were prepared using the Illumina Stranded mRNA Prep Ligation kit and 50 nt paired-end read sequencing were performed by MGX

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequencing services (Montpellier, France) on a SP flow cell using Novaseq 6000. The RNA-seq experiments were performed on three biological replicates.

Read alignment was performed using Hisat2 (43) on *D. melanogaster* and *D. simulans* reference genes available from FlyBase, versions r6.16 and r2.02, respectively, and previously masked using RepeatMasker. The reference sequence used for DCV was NC_001834.1 from GenBank. Alignments were processed using SAMtools (44). Gene count tables were generated using eXpress (45). TE count tables were obtained using the TEcount module of TEtools (46) and the same TE reference files as used in (18), and corresponding to 237 TE families. Gene and TE count tables were concatenated and then analyzed using the DESeq2 R package (version 4.3) (47). Four complete count tables were produced ({*D. melanogaster* carcasses}, {*D. melanogaster* ovaries}, {*D. simulans* carcasses}, and {*D. simulans* ovaries}). Each complete table was then analyzed by the DESeq2 procedure, ensuring that TE counts are normalized the same way as gene counts. The gene ontology (GO) analysis of RNA-seq data was performed on the differentially expressed genes with the gseGO function of the clusterProfiler package (48, 49). The GO terms were obtained using the R package org.Dm.eg.db v3.10.0 as database (50), corresponding to all *D. melanogaster* annotated genes. We converted *D. simulans* genes into their *D. melanogaster* orthologs using the orthology file from FlyBase ftp://ftp.flybase.net/releases/FB2020_04/precomputed_files/orthologs/dmel_orthologs_in_drosophila_species_fb_2020_04.tsv.gz. The GO overrepresentation analysis of biological process (BP) was performed with adjusted p-values by the FDR method, p-value cutoff at 0.05, and a minimum of 3 genes per term.

small RNA-seq

Ovaries of each female fly were carefully and manually dissected at 4 dpi. Small RNAs from 30 pairs of ovaries or 30 carcasses were extracted using the TraPR (Trans-kingdom, rapid, affordable Purification of RISCs) small RNA Isolation kit (Lexogen), as described previously (51). The TraPR method allows the specific isolation of fully functional and physiologically relevant interfering RNAs (microRNAs, piRNAs, siRNAs and scnRNAs), by anion exchange chromatography. Size selection (18-40 nt) was performed on gel at the sequencing platform. Purified small RNAs were used for library preparation, using TruSeq Small RNA Sample Preparation Guide - PN 15004197. Libraries were sequenced on an Illumina HiSeq 4000 sequencer as single-end 50 base reads.

Sequencing adapters were removed using cutadapt (52) -a

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TGGAATTCTCGGGTGCCAAGGAACTCCAGTCACTTA -m 1. Size filtering was performed using PRINTSEQ (53), in order to distinguish 21 nt-long reads (considered as siRNAs) from 23-30 nt-long reads (considered as piRNAs).

TE count tables for sense and antisense alignments were obtained using a modified version of the TEcount module (36) of TEtools (46) on the same TE reference sequences as above. Ping-pong signatures were checked using signature.py with min_size = 23 and max_size = 30 options (54). Read count normalization using microRNA counts as supposed invariants is a common strategy. However, it has been shown that some microRNA amounts could be affected by DCV infection (55). This is why we performed three different normalization strategies : 1) normalization using all microRNA sequences annotated from FlyBase (19-39 nt-long reads mapped using bowtie --best on FlyBase reference sequences dmel-all-miRNA-r6.16.fasta and dsim-all-miRNA-r2.02.fasta), 2) normalization using microRNA sequences but excluding those described to vary upon DCV infection in Monsanto-Hearnes *et al.* (55), 3) normalization using the endo-siRNAs described by Malone *et al.* (56), as we previously did in Roy *et al.* (18)). However, this last normalization procedure is only possible in *D. melanogaster* because endo-siRNA producing loci are not annotated in *D. simulans*. All normalization strategies provided similar patterns (Fig S8-S10).

Data availability

The RNA-seq and small RNA-seq data generated in this study have been submitted to the NCBI BioProject database under accession number PRJNA996035.

Acknowledgments

This work was performed using the computing facilities of Laboratoire de Biométrie Biologie Evolutive / Pôle Rhône-Alpes de bioinformatique (CC LBBE/PRABI) and the Symbiotron Pasteur platform, Laboratoire Biologie Fonctionnelle Insectes et Interactions (BF2i, Lyon). Sequencing was performed by the GenomEast platform, a member of the ‘France Génomique’ consortium (ANR-10-INBS-0009). We thank Matthieu Boulesteix, Miriam Merenciano, Cristina Vieira for helpful discussions. We thank Nelly Burlet, Sonia Janillon,

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Corinne Régis, Zaïnab Belgaidi, Carole Monégat, and Agnès Vallier for technical help. We thank also Clément Gilbert, Anna Zaidman-Rémy, and Feth el Zahar Haichar for helpful discussions. This work was funded by Agence Nationale de la Recherche (TEMIT grant) to M.F. and by the Agence Nationale de la Recherche (ANR-19-CE12-0012 -Top53 grant) to S.C.. M.V. was funded by CNRS-Japon PhD Joint Program.

References

1. Honeybee Genome Sequencing Consortium, Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* **443**, 931–949 (2006).

2. C. Shao, *et al.*, The enormous repetitive Antarctic krill genome reveals environmental adaptations and population insights. *Cell* **186**, 1279-1294.e19 (2023).

3. K. F. Tóth, D. Pezic, E. Stuwe, A. Webster, The piRNA Pathway Guards the Germline Genome Against Transposable Elements. *Adv. Exp. Med. Biol.* **886**, 51–77 (2016).

4. K.-A. Senti, J. Brennecke, The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet.* **26**, 499–509 (2010).

5. M. C. Siomi, K. Sato, D. Pezic, A. A. Aravin, PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12**, 246–258 (2011).

6. B. Czech, *et al.*, piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet* **52**, 131–157 (2018).

7. H. Yamashiro, M. C. Siomi, PIWI-Interacting RNA in Drosophila: Biogenesis, Transposon Regulation, and Beyond. *Chem Rev* **118**, 4404–4421 (2018).

8. X. Huang, K. Fejes Tóth, A. A. Aravin, piRNA Biogenesis in Drosophila melanogaster. *Trends Genet* **33**, 882–894 (2017).

9. A. Nayak, M. Tassetto, M. Kunitomi, R. Andino, RNA interference-mediated intrinsic antiviral immunity in invertebrates. *Curr Top Microbiol Immunol* **371**, 183–200 (2013).

10. M. Ghildiyal, P. D. Zamore, Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108 (2009).

11. D. Galiana-Arnoux, C. Dostert, A. Schneemann, J. A. Hoffmann, J.-L. Imler, Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. *Nat. Immunol.* **7**, 590–597 (2006).

12. S. H. Lewis, *et al.*, Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat Ecol Evol* **2**, 174–181 (2018).

13. W.-J. Chung, K. Okamura, R. Martin, E. C. Lai, Endogenous RNA interference provides a somatic defense against Drosophila transposons. *Curr. Biol.* **18**, 795–802 (2008).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

14. Y. Kawamura, *et al.*, Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**, 793–797 (2008).

15. B. Goic, *et al.*, RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model Drosophila. *Nat. Immunol.* **14**, 396–403 (2013).

16. M. Tassetto, M. Kunitomi, R. Andino, Circulating Immune Cells Mediate a Systemic RNAi-Based Adaptive Antiviral Response in Drosophila. *Cell* **169**, 314-325.e13 (2017).

17. L. Wang, *et al.*, Retrotransposon activation during Drosophila metamorphosis conditions adult antiviral responses. *Nat Genet* **54**, 1933–1945 (2022).

18. M. Roy, *et al.*, Viral infection impacts transposable element transcript amounts in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12249–12257 (2020).

19. K. N. Johnson, P. D. Christian, The novel genome organization of the insect picorna-like virus Drosophila C virus suggests this virus belongs to a previously undescribed virus family. *J Gen Virol* **79 (Pt 1)**, 191–203 (1998).

20. M. Kapun, V. Nolte, T. Flatt, C. Schlötterer, Host range and specificity of the Drosophila C virus. *PLoS One* **5**, e12421 (2010).

21. C. L. Webster, *et al.*, The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster. *PLoS Biol.* **13**, e1002210 (2015).

22. Á. G. Ferreira, *et al.*, The Toll-dorsal pathway is required for resistance to viral oral infection in Drosophila. *PLoS Pathog* **10**, e1004507 (2014).

23. S. Chtarbanova, *et al.*, Drosophila C virus systemic infection leads to intestinal obstruction. *J Virol* **88**, 14057–14069 (2014).

24. R. P. van Rij, *et al.*, The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. *Genes Dev.* **20**, 2985–2995 (2006).

25. A. Nayak, *et al.*, Cricket paralysis virus antagonizes Argonaute 2 to modulate antiviral defense in Drosophila. *Nat Struct Mol Biol* **17**, 547–554 (2010).

26. J. Martinez, *et al.*, Symbionts commonly provide broad spectrum resistance to viruses in insects: a comparative analysis of Wolbachia strains. *PLoS Pathog* **10**, e1004369 (2014).

27. S. H. Merkling, *et al.*, The epigenetic regulator G9a mediates tolerance to RNA virus infection in Drosophila. *PLoS Pathog.* **11**, e1004692 (2015).

28. M.-C. Saleh, *et al.*, Antiviral immunity in Drosophila requires systemic RNA interference spread. *Nature* **458**, 346–350 (2009).

29. B. Berry, S. Deddouche, D. Kirschner, J.-L. Imler, C. Antoniewski, Viral suppressors of RNA silencing hinder exogenous and endogenous small RNA pathways in Drosophila. *PLoS One* **4**, e5866 (2009).

30. B. Barckmann, *et al.*, The somatic piRNA pathway controls germline transposition over generations. *Nucleic Acids Res.* **46**, 9524–9536 (2018).

31. M. Mohamed, *et al.*, A Transposon Story: From TE Content to TE Dynamic Invasion of

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Drosophila Genomes Using the Single-Molecule Sequencing Technology from Oxford Nanopore. *Cells* **9** (2020).

32. M. Mohamed, *et al.*, TrEMOLO: accurate transposable element allele frequency estimation using long-read sequencing data combining assembly and mapping-based approaches. *Genome Biol* **24**, 63 (2023).

33. N. V. Rozhkov, M. Hammell, G. J. Hannon, Multiple roles for Piwi in silencing Drosophila transposons. *Genes Dev.* **27**, 400–412 (2013).

34. Z. Jin, A. S. Flynt, E. C. Lai, Drosophila piwi mutants exhibit germline stem cell tumors that are sustained by elevated Dpp signaling. *Curr Biol* **23**, 1442–1448 (2013).

35. A. Akkouche, *et al.*, Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells. *EMBO Rep.* **14**, 458–464 (2013).

36. M. Fablet, *et al.*, Dynamic Interactions Between the Genome and an Endogenous Retrovirus: Tirant in Drosophila simulans Wild-Type Strains. *G3 (Bethesda)* **9**, 855–865 (2019).

37. S. H. Merkling, R. P. van Rij, Analysis of resistance and tolerance to virus infection in Drosophila. *Nat Protoc* **10**, 1084–1097 (2015).

38. A. Akkouche, *et al.*, tirant, a newly discovered active endogenous retrovirus in Drosophila simulans. *J. Virol.* **86**, 3675–3681 (2012).

39. P. R. Holden, J. F. Brookfield, P. Jones, Cloning and characterization of an ftsZ homologue from a bacterial symbiont of Drosophila melanogaster. *Mol Gen Genet* **240**, 213–220 (1993).

40. M. Riegler, M. Sidhu, W. J. Miller, S. L. O'Neill, Evidence for a global Wolbachia replacement in Drosophila melanogaster. *Curr. Biol.* **15**, 1428–1433 (2005).

41. D. Monnin, *et al.*, Influence of oxidative homeostasis on bacterial density and cost of infection in Drosophila-Wolbachia symbioses. *J Evol Biol* **29**, 1211–1222 (2016).

42. L. J. Reed, H. Muench, A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS. *Am J Epidemiol* **27**, 493–497 (1938).

43. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907–915 (2019).

44. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

45. A. Roberts, C. Trapnell, J. Donaghey, J. L. Rinn, L. Pachter, Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* **12**, R22 (2011).

46. E. Lerat, M. Fablet, L. Modolo, H. Lopez-Maestre, C. Vieira, TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. *Nucleic Acids Res.* **45**, e17 (2017).

47. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.18.585529; this version posted March 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

48. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

49. T. Wu, *et al.*, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).

50. M. Carlson, *org.Dm.eg.db: Genome wide annotation for Fly. R package version 3.8.2.* (2019).

51. T. Grentzinger, *et al.*, A universal method for the rapid isolation of all known classes of functional silencing small RNAs. *Nucleic Acids Res* **48**, e79 (2020).

52. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

53. R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).

54. C. Antoniewski, Computing siRNA and piRNA overlap signatures. *Methods Mol. Biol.* **1173**, 135–146 (2014).

55. V. Monsanto-Hearne, A. L. Y. Tham, Z. S. Wong, S. Asgari, K. N. Johnson, Drosophila miR-956 suppression modulates Ectoderm-expressed 4 and inhibits viral replication. *Virology* **502**, 20–27 (2017).

56. C. D. Malone, *et al.*, Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137**, 522–535 (2009).

Figures

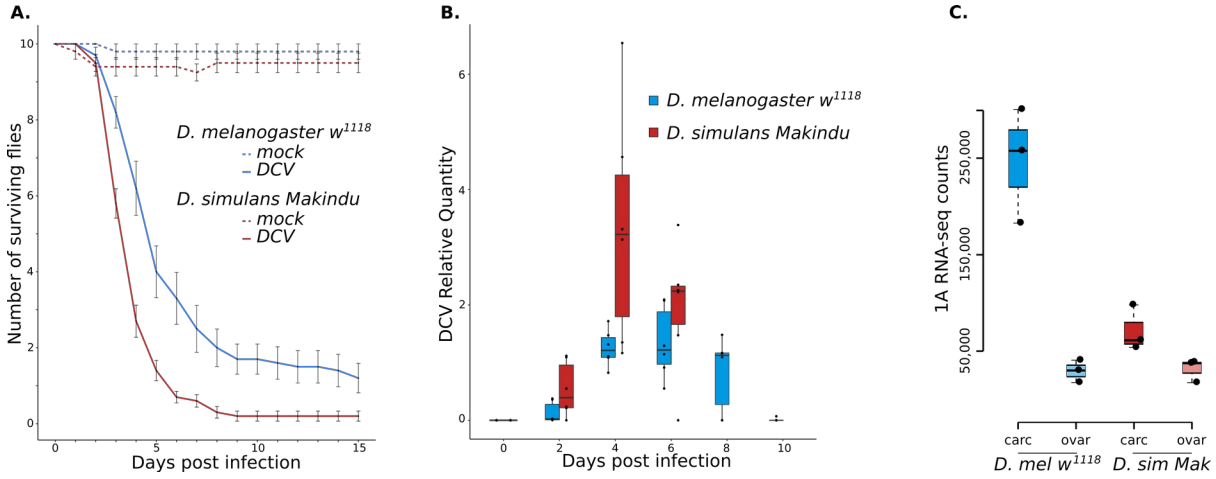


Figure 1. Response to DCV infection in *D. melanogaster w*^{1118} and *D. simulans* Makindu. **A.** Fly survival upon DCV infection. **B.** Kinetics of DCV titers followed using RT-qPCR. **C.** Raw RNA-seq read counts mapping against 1A sequence. RNA-seq read

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Annexe 5

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Transposable elements as evolutionary driving force to host shift in cactophilic *Drosophila* species

Oliveira D. S.[1,2], Larue A.[2], Nunes W. V. B.[2], Sabot F.[3], Bodelón A.[4], García Guerreiro M. P.[4], Vieira C.[2,*], Carareto C. M. A.[1,*]

[1] São Paulo State University (Unesp), Institute of Biosciences, Humanities and Exact Sciences, São José do Rio Preto, São Paulo, 15054-000, Brazil.

[2] Universite Claude Bernard Lyon 1, Laboratoire de Biométrie et Biologie Évolutive, CNRS, UMR5558, Villeurbanne 69100, France.

[3] DIADE, University of Montpellier, Cirad, IRD, 911 avenue Agropolis BP64501, 34394 Montpellier, France.

[4] Grup de Genòmica, Bioinformática i Biologia Evolutiva, Departament de Genètica i Microbiologia (Edifici C), Universitat Autònoma de Barcelona, Barcelona, Spain

* Corresponding authors

Daniel Siqueira de Oliveira: daniel.siqueira-de-oliveira@etu.univ-lyon1.fr

Anaïs Larue: anais.larue@univ-lyon1.fr

William Vilas Boas Nunes: william.vilas-boas-nunes@univ-lyon1.fr

François Sabot: francois.sabot@ird.fr

Alejandra Bodelón: Alejandra.Bodelon@uab.cat

María Pilar García Guerreiro: MariaPilar.Garcia.Guerreiro@uab.cat

Cristina Vieira: cristina.vieira@univ.lyon1.fr

Claudia Marcia Aparecida Carareto: claudia.carareto@unesp.br

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ABSTRACT

Background

The host shift in insects has been considered a key process with potential to collaborate with reproductive isolation and speciation. Both genomics and transcriptomics variation has been attributed to such process, in which gene families with functions for host location, acceptance and usage have been proposed to evolve. In this context, cactophilic *Drosophila* species are an excellent model to study host shift evolution, since they use a wide-range of cacti as hosts, and many species have different preferences. Transposable elements are engines of genetic novelty between populations and species, driving rapid adaptive evolution. However, the extent of TEs' contribution to host shift remains unexplored.

Results

We performed genomic and transcriptomic analysis in seven genomes of cactophilic species/subspecies to investigate how TEs interact with genes associated with host shift. Our results revealed enrichment of TEs at promoter regions of host shift-related genes, with *Helitrons* representing ~60% of the cases, demonstrating an unprecedented putative *cis*-regulatory role of *Helitrons* in *Drosophila*. Differential expression analysis between species with different preferred hosts demonstrated divergence on gene expression in head and larvae tissues. Although TEs' presence does not affect overall gene expression, we observed 1.31% of genes generating gene-TE chimeric transcripts, including those with function affecting host preference.

Conclusions

Our combined genomic and transcriptomic approaches provide evidence of TE-driven divergence between species, highlighting the evolutionary role of TEs in the context of host shift, a key adaptive process that can cause reproductive isolation.

Keywords: comparative transcriptomics, adaptation, chimeric transcripts.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Background

The molecular consequences of host shift evolution can be closely related to ecological divergence, specialization and ultimately ecological speciation (1). Due to the wide variety of hosts, successful plant-insect interactions require adaptations from the insect. In many species, reproductive isolation arises as a result of successive adaptations driven by host shift (HS), which results from divergent selection between different ecological environments (1–4). Thus, natural selection increases the frequency of alleles that confer fitness benefits when HS occurs. *Drosophila* species use a wide range of necrotic host tissues as feeding and breeding sites, providing a suitable model to study the HS process and its role in insect speciation. Among them, the *repleta* group comprises a clade with ~100 cactophilic *Drosophila* species (5). The specialization in cacti represents an important ecological challenge to the flies, due to the lack of nitrogen and phosphorus compared to other hosts, as well as the presence of toxic compounds (6). In the *repleta* species, the HS was initially to the *Opuntia sp*. cacti, and posteriorly several independent shifts happened to columnar cacti (6), considered a host with higher toxicity than *Opuntia sp.* (7). Such HS is observed between sibling species with recent divergence in two clusters of the subgroup *mulleri*: cluster *buzzatii* and cluster *mojavensis*. In the cluster *buzzatii*, *Drosophila buzzatii* and *D. koepferae* are sibling species (divergence 4-5 Mya (8)) with different primary hosts. The former species use preferentially *Opuntia sp.* cacti, whereas the latter is a columnar cacti dweller (9). In the cluster *mojavensis*, *D. arizonae* and the four allopatric subspecies of *D. mojavensis* have a high diversity regarding cacti hosts' preference. *D. arizonae* use both *Opuntia sp.* and columnar cacti, whereas *D. m. mojavensis* uses *Ferocactus cylindraceus* as preferential host, *D. m. wrigleyi* uses only *Opuntia sp*., *D. m. baja* uses *Stenocereus gummosus*, and *D. m. sonorensis* uses *S. thurberi* (10). Therefore, this set of species from *buzzatii* and *mojavensis* clusters provide an excellent model to investigate the molecular mechanisms underlying HS evolution.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The radiation of cactophilic *Drosophila* species across different hosts is an evolutionary outcome that gave rise to specific adaptations (11), which may be summarized into three steps: location, acceptation and host usage (12). In the location step, odorant-binding proteins (OBPs) and odorant receptors (ORs) are key proteins to distinguish specific host in the environment, through chemoreceptors that integrate the visual and olfactory systems. Subsequently, in the acceptance step occurs the evaluation of the nutritional compounds present on the host, as well as the presence of competitors, predators, pathogens and parasites (12). Several receptors are associated with this process, such as gustatory receptors (GRs) (13) and ionotropic receptors (IRs). (14–16). Finally, the flies must be able to complete the feeding and breeding processes, but it depends upon the overcoming of toxic substances that might be present on the host. Several gene families are associated with detoxification, such as P450s (CYPs), Glutathione S-Transferases (GSTs), UDP-Glicosiltransferases (UGTs), esterases (ESTs) and ATP binding-cassette transporters (ABCs). Altogether, these nine gene families associated with the Host Location, Acceptance and Usage will henceforth be referred to as HLAU.

A previous study with 14 *Drosophila* genomes, representing both *Sophophora* and *Drosophila* subgenera, has proposed that 82 genes, including GSTs, CYPs, and UDPs, are under positive selection (17). Signatures of positive selection were also found between *D. mojavensis* subspecies and *D. arizonae* in subsets of GRs and ORs (16). In addition to the selective pressure acting into HLAU gene families, modifications in the gene expression also play an important role in the HS adaptation in *Drosophila* species. It has been shown that CYPs, GSTs and OBPs are differentially expressed between *D. m. wrigleyi* and *D. m. sonorensis* (18), indicating that gene expression in the nervous system might play a role into detoxification and olfactory systems, hence contributing to HS adaptation in *D. mojavensis*.

Transposable elements (TEs) are mobile sequences with the ability to create structural variants in the genomes. In most cases, mutations caused by TEs are likely to be deleterious or neutral. Throughout evolutionary time, TEs that remain in the genome tend to be silenced

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

by epigenetic control, accumulating mutations, and losing their transposition ability (19). Despite they become degraded, the remaining TE sequences may still contain either regulatory motifs or protein domains (20). These sequences might be co-opted by the cell machinery modifying gene expression or protein sequences of the nearby genes (reviewed in (21)). Depending upon the positive outcome of the cooption into the individual fitness, these TEs can increase their frequency in the population, contributing to populational evolution. For instance, in *D. melanogaster*, the gene *CHKov1* has a truncated mRNA derived from a *DOC* TE located into the intron, resulting in resistance to viral infection and insecticides (22). Such exaptation and domestication events can be often identified by the occurrence of chimeric transcripts, which are mRNAs with both gene and TE-derived sequences (23). A recent transcriptome-wide study has identified that ~1.12% of *D. melanogaster* genes produce chimeric transcripts in different populations (24). Among all genes, 76 generate chimeric transcripts from TE insertions that were present in one strain, but absent in another, highlighting the potential of TEs as a source of genetic novelty between different *Drosophila* ecotypes.

Many aspects involving HS adaptation in *repleta* species have been shown previously, such as detoxification pathways (25), morphology (26), life history traits (27), behavior (28), and genomic and transcriptomic differences (18,29–32). However, the contribution of TEs to the evolution of HLAU genes has not yet been assessed. Here we aimed to uncover the extension of variability derived from TEs in cactophilic *Drosophila* species, using both genome- and transcriptome-wide analysis. We tested the hypothesis regarding the contribution of TEs to the HS in cactophilic species. The genome-wide analysis revealed that HLAU gene families are enriched with TEs in regulatory regions, particularly in ORs. Furthermore, the results show that approximately 35% of ORs with TEs in their promoter regions contain transcription factor binding sites (TFBSs) within the TE sequence, of which around 60% are Helitrons, highlighting their unprecedented potential regulatory role in *Drosophila*. The transcriptomic analysis revealed that HLAU genes are differentially expressed

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

between cactophilic species. Additionally, we also show that 1.31% of genes produce chimeric transcripts in head and larvae tissues, with HLAU genes are also being involved.

Results

Genome assemblies and gene annotation

To investigate the potential role of TEs into HS of cactophilic *Drosophila* species, we performed Nanopore long-reads sequencing on species that have different preferential cacti as hosts: *D. buzzatii* (*Opuntia sp.*), *D. koepferae* (columnar cacti), *D. arizonae* (*Opuntia sp.* or columnar cacti), *D. m. mojavensis* (*F. cylindraceus*), *D. m. wrigleyi* (*Opuntia sp.*), *D. m. sonorensis* (*S. thurberi*), and *D. m. baja* (*S. gummosus*). We obtained high resolution assemblies for all genomes, with an average of 567 scaffolds, 13.8 Mb of N50 (Sup. Table 1), and ~98% of BUSCO genes (Sup. Fig. 1). Furthermore, the gene annotation for these genomes covered 98.63% of all reference genes in the *D. mojavensis* subspecies, and 95.98% in *D. arizonae*. The *de novo* gene annotation in *D. buzzatii* and *D. koepferae* revealed a total of 18,050 and 17,848 coding genes, for which 63.77%, and 64.13% were successfully assigned as one-to-one orthologs with *D. mojavensis*, respectively. These results represent the higher gene repertoire identified for *D. buzzatii* and *D. koepferae* species, altogether with its ortholog relationship with *D. mojavensis* and *D. arizonae*. Since analysis on HLAU genes is the focus of this study, their efficient annotation is relevant. They are represented by nine gene families associated with the three steps on the use of host resources: (1) Location: OBPs, ORs; (2) Acceptance: GRs, IRs; and (3) Host usage: ABCs, ESTs, GSTs, UDPs, CYPs (Sup. Table 2). In *D. arizonae* and in the *D. mojavensis* subspecies, we recovered ~98.21% of the total HLAU genes from the reference genome. In *D. buzzatii* and *D. koepferae* the number of annotated HLAU genes covered ~95% in both species compared to *D. mojavensis* (Table 1).

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Species/subspecies	Total genes	Host location, acceptance and usage genes								
		Location		*Acceptance*		*Host use*				
		OBP	OR	GR	IR	ABC	EST	GST	UDP	CYP
Ref. *D. m. wrigleyi*	**15,045**	**25**	**59**	**45**	**22**	**27**	**17**	**28**	**21**	**59**
D. m. mojavensis	14,810	25	59	44	22	27	17	27	20	59
D. m. baja	14,797	25	57	44	22	27	17	27	19	59
D. m. wrigleyi	14,948	25	59	45	22	27	17	27	20	59
D. m. sonorensis	14,808	25	57	44	22	27	17	27	20	58
D. arizonae	14,441	25	56	41	21	27	17	27	20	59
D. buzzatii	18,050	25	52	34	21	27	16	27	20	68
D. koepferae	17,848	25	45	34	24	27	13	25	23	62

Table 1. Total number of genes annotated in seven *Drosophila* cactophilic species. The "Ref. *D. m. wrigleyi*" represents the annotation in the reference genome used to perform the gene annotation for our long-read Nanopore genomes. HLAU genes are separated according to their functions: Location: odorant-binding proteins (OBP), odorant receptor (OR); Acceptance: gustatory receptor (GR), ionotropic receptor (IR); Host usage: ATP-binding cassette (ABC), carboxylesterase (EST), glutathione S-Transferase (GST), glycosiltransferase (UDP), and cytochrome P450 (CYP).

Molecular evolution of HLAU genes between species with preference for Opuntia sp. and columnar cacti

The use of *Opuntia sp.* has been proposed to be an ancestral state in cactophilic *Drosophila* species, whereas columnar cacti is the derived state (6). To investigate whether species with preferential columnar cacti usage have HLAU genes that undergone positive selection, we used a branch-site approach to assess their evolutionary rates compared to species that use *Opuntia sp*. For this analysis, we removed *D. arizonae* due to its non-preferential usage between columnar and *Opuntia sp.* cacti (6). Moreover, we added *D. navojoa*, which is a basal species of the *mojavensis* cluster that use solely *Opuntia* sp. (33). Therefore, we carried out the branch-site selection test with *D. buzzatii*, *D. m. wrigleyi*, and *D. navojoa* as *Opuntia*-related branches (background), and *D. m. baja*, *D. m. mojavensis*, *D. sonorensis*, and *D. koepferae* as columnar-related branches (foreground). It is important to highlight that due to the need of one-to-one ortholog genes between the seven genomes,

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

our analysis have decreased the total number of HLAU genes to ~42% of the initial number depicted on Table 1. From 128 genes, we did not observe any gene under positive selection with FDR < 0.05. To further explore the evolutionary forces driving gene HLAU genes evolution, we also performed a test for relaxed selection. We observed the genes *Obp56d* (LOC6580688) and *Cyp304a1* (LOC6572829) with intensified selection strength (k > 1 and FDR < 0.05). Other six HLAU genes had significant relaxed selection, but with k < 1 (Sup. Table 4).

Evolutionary dynamics of TEs in cactophilic Drosophila species

We used a combination of automatic and manual methods to build a curated TE-library for each species. The method was designed to perform polishing steps in both TE consensus and copies to remove artefacts (see Methods). The total number of TE consensuses (LTRs separated of internal sequences) ranged from 243 in *D. m. baja* to 313 in *D. m. mojavensis*. The TE content in the genomes ranged from 14.07% in *D. buzzatii* to 20.4% in *D. m. mojavensis* (Fig. 1A). In addition, we observed a positive correlation between TE content and genome size estimated from the assemblies (*Pearson* correlation: p = 0.013; r = 0.86) (Sup. Fig. 3). Following the correlation, both species from the *buzzatii* cluster had the lower genome sizes and TE contents compared to species of the *mojavensis* cluster (Fig. 1A). Regarding the proportional load of TE orders in the genomes, LTRs contribute with 31.21% and 31.68% in the sister species *D. buzzatii* and *D. koepferae*, respectively. The main differences in the mobilomes are due to *Helitrons* and LINEs (Fig. 1B).

In the *mojavensis* cluster (*D. arizonae* and the four *D. mojavensis* subspecies), the abundance of TEs in the genome varied substantially (Fig. 2B), despite their recent divergence time of 1.5 Mya (34). *D. m. mojavensis* has the TE-richest genome (20.45%), with the abundance following from the highest to the lowest: LTR > LINE > Helitron > DNA > Unknown (Fig. 1B). The distribution of TEs in *D. arizonae* and the other three *D. mojavensis* does not follow such order. All of them have LTRs with the highest abundance, but followed

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

by Helitrons, LINEs, DNA, and Unknown (Fig. 1B). The observed differences are often associated with the divergence on the evolutionary patterns of the TE classes between the genomes. Our analysis revealed that *Helitrons* had a recent burst on *D. m. mojavensis* and *D. m. wrigleyi*, followed by a decline (Fig. 1C). In *D. arizonae* the patterns reveals high genome content derived from TEs with low Kimura 2-Parameters divergence, which indicates conserved and potentially active copies. The antagonistic patterns of TE dynamics observed in the seven analyzed genomes suggest that TEs contribute differently to genome evolution across these species.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

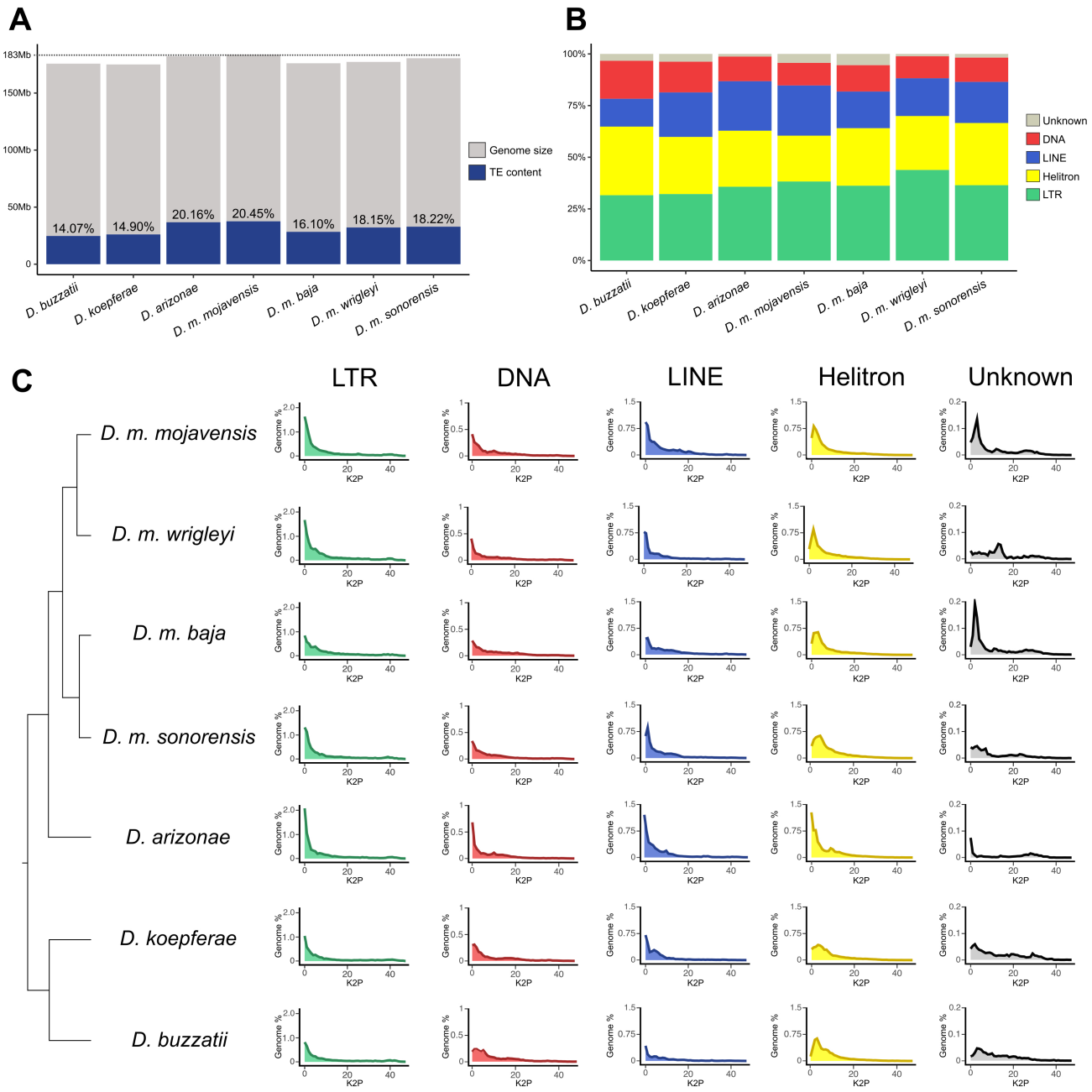


Figure 1. A) Genome size of the seven cactophilic *Drosophila* species sequenced in this work, and their respective TE content. **B**) Proportional distribution of the total TE content annotated in the genomes, divided by DNA transposon, LINE, Helitron, LTR, and Unknown. **C**) Phylogenetic relationship among species (adapted from (25,35)), alongside with TE families landscapes representing TE content in genome % (y-axis), and the divergence of TE insertions compared to the consensuses (x-axis), based on Kimura 2-parameters (K2P) corrected with CpG. The TE content located on the left of the graph is represented by TE insertions with low divergence with their respective consensus, hence inferring that they might be conserved/recent/active, whereas TE content on the right part represents copies with high divergence due to the accumulation of mutations, corresponding to old/inactive copies.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

HLAU genes are enriched in TEs at regulatory regions

To investigate whether TEs may have been maintained over time in a non-random distribution in the genomes relative to genes, we compared their positions and analyzed their frequencies in each genome. We analyzed whether specific TE orders are preferentially located within genes, which would suggest possible co-evolution. To do so, we compared the frequency of TE families that are preferentially inserted within gene regions (>50% of all insertions overlapping with genes), with a random distribution of TE insertions from the same order. We did not observe any TE order enriched within genes, which is expected with the hypothesis of purifying selection against TEs disrupting gene sequences (Sup. Fig. 4).

Despite TEs were not enriched on gene regions overall, a few insertions may be selected in the promoter of genes if they provide beneficial outcome to the individual fitness, through donation of histone marks enrichment, or *cis*-regulatory elements (36,37). We performed permutation tests to investigate whether TEs are enriched in the promoters (2 Kb upstream of the TSS) from the nine HLAU gene families in the seven genomes. To make a comparison of our findings with non-host shift related genes, we also included zinc-finger and serine/threonine kinases as two non-HLAU gene families. Regarding genes associated with location (OBPs and ORs), we observed that ORs have enrichment of TEs at the promoters in all genomes from species of the *mojavensis* cluster (p-value < 0.05), (Fig. 2A). For acceptance genes, *D. koepferae* had enrichment of TEs in the promoters of GRs (Fig. 2A). Finally, enrichment of TEs upstream to detoxification genes was observed in *D. arizonae* and *D. m. baja*. The latter had the enrichment for the CYP and Esterase families, whereas the latter had ABCs enriched by TEs (Fig. 2A). Overall, detoxification genes had z-score towards depletion of TEs on promoters, for which we observed significant depletion on *D. buzzatii* (Fig. 2A). In the non-HLAU gene families, we observed the expected result by chance: either z-score indicated the expected number of genes with TEs on promoters in a genome-wide comparison (z-score near 0), or it indicated depletion of TEs, as observed in *D. buzzatii* and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

D. koepferae. This result reinforces our findings indicating the significant and evolutionary unexpected enrichment of TEs, especially on the promoters of ORs.

The results depicting the histograms with frequency of genes with TEs for each permutation test, and p-values are shown in the Sup. Fig. 5 and Sup. Table 5. To obtain further insights in the potential regulatory role of TEs, we also analyzed the enrichment at the 2 kb downstream region. In *D. buzzatii* and *D. koepferae*, we did not observe significant enrichment, indicating that the frequency of TEs downstream to HLAU genes follows the genomic distribution in these species. However, in the species of the *mojavensis* cluster, ORs had enrichment for TEs in the downstream region, as we have also observed for the 2 kb upstream. Finally, in the four gene families associated to detoxification, we observed that esterases are enriched by TEs in *D. arizonae*, while GSTs are depleted by TEs in *D. m. wrigleyi*. Our results suggest a possible ancient evolutionary path of TE cooption in species of the *mojavensis* cluster, since enrichment of TEs in ORs was observed in the five genomes.

The enrichment of TEs upstream and downstream of ORs indicates a possible coevolution of this gene family with TEs. Alternatively, it may also indicate that ORs are located in TE-rich regions of the genomes. To investigate this hypothesis, we identified TE-rich regions (Sup. Fig. 6) and analyzed the frequency of ORs. We observed on average 6 and 3 ORs in such regions in species of the *mojavensis* and *buzzatii* clusters, respectively (Sup. Table 6). Therefore, we rejected the hypothesis that OR genes are enriched by TEs in their vicinity due to their TE-rich genomic location. Other factors might be involved in the evolutionary maintenance of TEs nearby ORs, such as the cooption of regulatory motifs.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

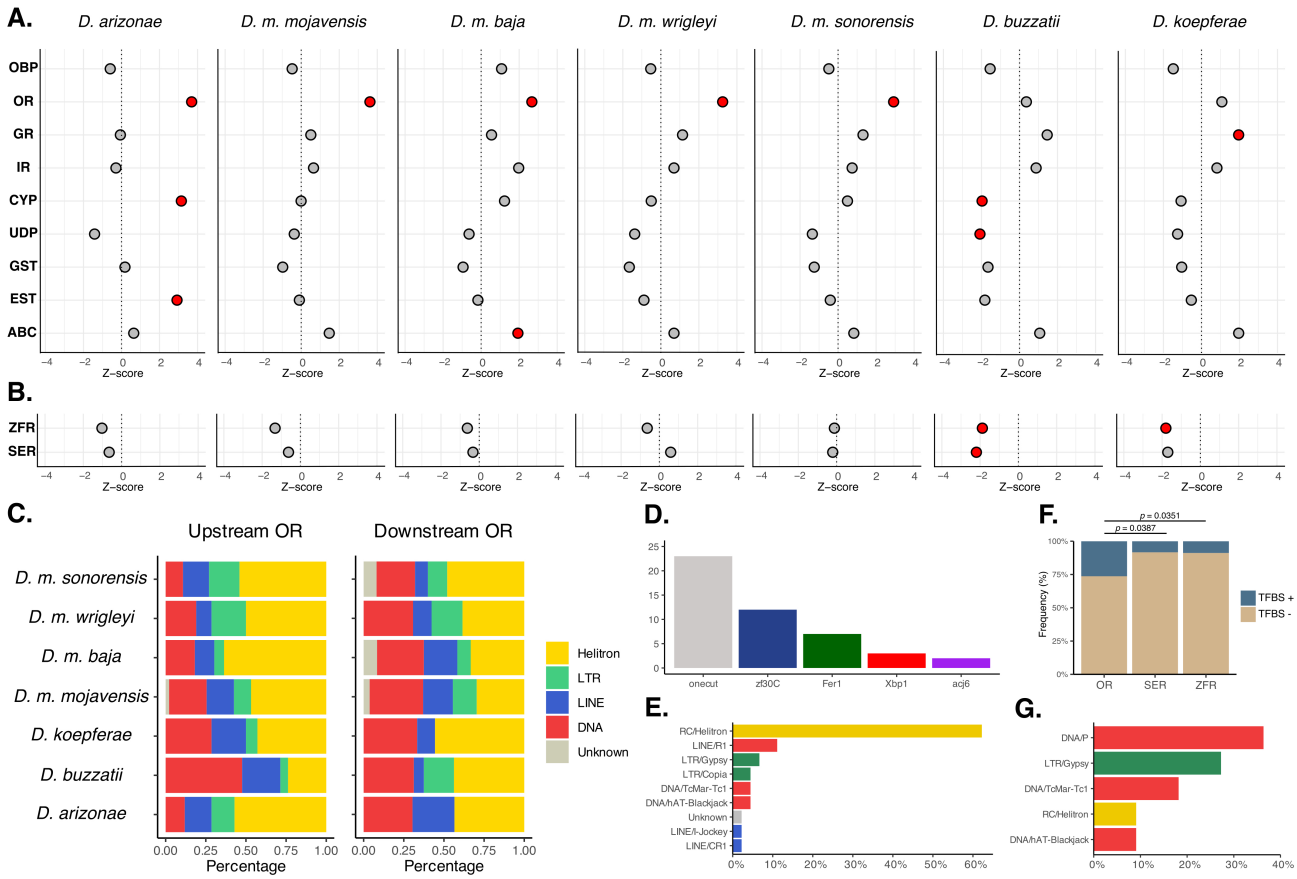


Figure 2. A) Enrichment of TEs in the promoter region (2kb upstream) of HLAU gene families. Z-score > 0 indicates enrichment, and Z-score < 0 depletion. Red dots: p-value < 0.05, grey dots: p-value > 0.05. **B**) The same enrichment analysis as depicted on A, but for non-HLAU gene families, represented by zinc-finger proteins (ZFR) and Serine/Threonine Kinases (SER). **C**) Proportion of TE orders located upstream and downstream ORs. **D**) Transcription factor binding sites (TFBSs) frequency on TEs located upstream OR genes in all species under study. **E**) Proportional frequency of TE families with TFBSs embedded on their sequences from TE copies located upstream OR genes. **F**) Significant Fisher's exact test between OR genes with OR-related TFBSs from TE insertions compared to the frequency observed in the non-HLAU gene families Serine/Threonine Kinases (SER) and zinc-finger proteins (ZFR). **G**) Proportion of TE orders carrying OR-related TFBSs in non-HLAU gene families. *Helitrons* represent only 10% of the copies.

A few TE families have been shown to be more prone to contribute with *cis*-elements than others, such as retroelements in *D. melanogaster* (38). Therefore, we investigated whether the observed enrichment of TEs in the HLAU gene families is TE-specific. Considering all species, 50.42% and 40% of the insertions located upstream and downstream of ORs

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

correspond to *Helitrons*, respectively (Fig. 3C). *Helitrons* have been documented to act on gene evolution of bats by carrying gene sequences (39), mainly 5′ UTRs and the first exons, within their own internal regions (40). Thus, we analyzed whether these *Helitron* insertions upstream to HLAU genes, mainly ORs, could be part of the first exons from genes. By intersecting exons from genes enriched with their respective upstream *Helitrons*, we observed that only *Or13a* and *Or22c* have their first exons overlapping with the TEs. The former was observed only in *D. m. wrigleyi*, suggesting a subspecies-specific insertion, whereas the latter is observed in the four *D. mojavensis* subspecies, likely be to inherited from their common ancestor (Sup. table 7). At the ORs' downstream region, we observed five genes with *Helitrons* overlapping exons: *Or71a*, *Or10a*, *Or13a*, *Or82a* and *Or30a*. Interestingly, the same *Helitron* in the *Or82a* was observed in the four *D. mojavensis* subspecies, highlighting the inheritance from the common ancestor and maintenance of the insertion overtime (Sup. table 7). Taken together, our results rejected the hypothesis regarding the possible role of *Helitrons* carrying HLAU genes' fragments. However, these elements could still act on gene regulation if they carry *cis*-elements as TFBSs.

TEs nearby gene regions have been shown to modulate gene expression, through donation of TFBSs or polyA sites to the ancestral regulatory motifs. To investigate the potential functional role of the observed enrichment of TEs in the upstream region of ORs, we carried out TFBSs identification for all TE insertions located in such position. We used the transcription factors (TFs) previously described to participate on OR's transcription in *D. melanogaster*: acj6, onecut, xbp1, fer1 and zf30C (41). Our analysis revealed that 35.71% of the ORs with TEs at the promoter region have the TFBS located within the insertion (Sup. table 8). Taking all species together, the most frequent TFBSs found was for the TF onecut (48.93%), followed by zf30C (25.53%), and Fer1 (14.89%) (Fig. 3D). *Helitrons* were the TEs with the highest frequency, representing 59.57% of all insertions carrying ORs' TFBSs (Fig. 3E). To test whether these results are specific to ORs, we also performed the same TFBSs' detection and frequency of TEs ZFR SER genes as non-HLAU families. Our results show that ZFR and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SER have on average 9.75% and 9.09% of the genes with OR-related TFBSs, respectively, which can be considered as noise from the analysis. Comparing these proportions with the ones found in ORs, we had significantly more TFBSs compared to ZFR (*Fisher's exact test: p-value = 0.0387*) and SER (*Fisher's exact test: p-value = 0.0351*) (Fig. 2F and Sup. Table 9). In addition, *Helitrons* represented only 10% of the insertions carrying TFBSs in ZFR and SER genes (Fig. 2G). These comparative results suggest that our findings regarding the presence of OR-related TFBSs from TE insertions, and the enrichment of *Helitrons*, are unlikely to be due to chance. A few cases of ancestral inheritance were observed, for instance, the *Or22c* and *Or82a* have a conserved *Helitron* insertion carrying the onecut TFBS in the five genomes from the *mojavensis* cluster (Sup. table 7). We also observed a case of loss, where the *Or85d* has the TFBS Fer1 in the three *D. mojavensis* subspecies that uses primarily columnar cacti as host, but absent in *D. m. wrigleyi*. These results highlight a putative functional role of TEs as donators of *cis*-elements to ORs in the cactophilic species *D. mojavensis* and *D. arizonae*.

HLAU genes are differentially expressed between Opuntia sp. and columnar cacti dwellers

Differential expression of genes has been associated with adaptation to new hosts in insects (31). Here, aiming to understand how gene expression divergence is associated with adaptation to different cactus hosts, we performed differential expression analysis between species with preference for *Opuntia sp.* and columnar cacti, in head and larvae tissues. Overall, the principal component analysis (PCA) from HLAU gene expression revealed remarkable differences between species in both tissues (Fig. 3C). The variance of gene expression resembles the phylogenetic divergence of the species, with the highest variance between the *buzzatii* and *mojavensis* clusters (PC1), followed by divergence between *D. buzzatii* and *D. koepferae* (PC2). The species of the *mojavensis* cluster were clustered together with a slight separation between *D. arizonae* and the subspecies of *D. mojavensis*.

We performed pairwise analysis of HLAU genes expression between species with preference for *Opuntia sp.* (*D. m. wrigleyi* and *D. buzzatii*) versus species with preference for

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

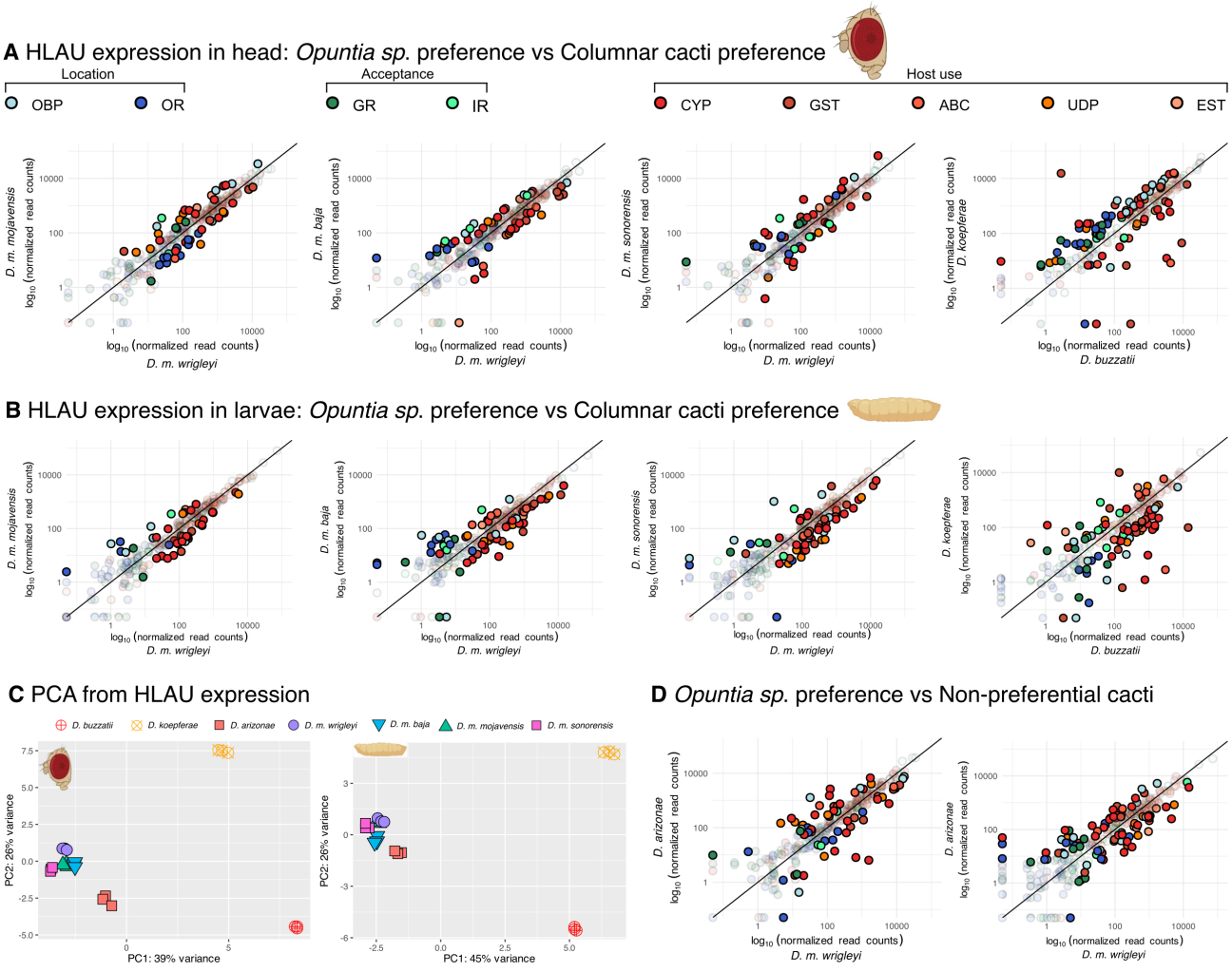
columnar cacti, and *D. arizonae* as non-preferential cacti. In head, *D. m. wrigleyi* had 27, 24, and 19 HLAU genes with higher expression compared to *D. m. mojavensis*, *D. m. baja*, and *D. m. sonorensis*, respectively (Fig. 3A). Four genes were found with higher expression in the head of *D. m. wrigleyi* in the three comparisons: *UDP-Ugt4*, *Gst-theta3*, *Abc-a3*, *Cyp6a21*. In the subspecies with preference for columnar cacti, we observed on average 26 HLAU genes with higher expression compared to *D. m. wrigleyi*. Among them, seven were recurrent in the three subspecies: *Ir21a, EstB1*, *UDP2*, *UDP-Ugt5*, *Cyp9b2*, *UDP-Ugt5-1*, *Cyp12a2*. Since they were observed in all subspecies with preference to columnar cacti, the higher expression of these genes can be related to the adaptation to columnar cacti. Finally, the same analysis on larvae revealed that *D. m. wrigleyi* has three genes with recurrent higher expression compared to the other subspecies: *Or82a*, *Obp59a*, and *Ir21a*; whereas in *D. mojavensis* subspecies we observed 17 common up-regulated genes compared to *D. m. wrigleyi*, which comprise seven CYP genes, three GSTs, four UDPs, one Esterase and one OBP (Fig. 3B). Regarding *D. buzzatii* and *D. koepferae*, we observed 27 HLAU genes in head and 54 in larvae with higher expression in *D. buzzatii*, and 63 in head and 29 with higher expression in *D. koepferae*. We also compared the expression of HLAU genes between *D. m. wrigleyi* and its sister species without preferential cacti host, *D. arizonae*. In head, the *Opuntia* sp. dweller *D. m. wrigleyi* had 31 HLAU genes with higher expression compared to its counterpart *D. arizonae*. In larvae, *D. m. wrigleyi* had 43 HLAU genes with higher expression compared to *D. arizonae* (Fig. 3D). The result demonstrates that HLAU genes have divergent differential expression in both head and larvae tissues.

To further investigate whether differential expression of HLAU genes have higher prevalence compared to non-HLAU genes, we compared the proportion of differentially expressed genes associated to location, acceptance, and usage with the proportion of differentially expressed genes from the non-HLAU gene families Ser/Thr Kinases and Zinc-finger. In head, our analysis revelaed that location and usage have higher differential expression than non-HLAU genes between *D. buzzatii* and *D. koepferae* (Sup. Figure 7). In *D. mojavensis* subspecies, we observed higher differential expression for location genes

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

between *D. m. wrigleyi* and *D. m. mojavensis*, and for usage between *D. m. wrigleyi* and *D. m. baja*. In larvae, we did not observe significant difference between non-HLAU and HLAU genes, but the pairwise differential expression between *D. m. wrigleyi* and three subspecies revealed higher prevalence of gene families associated with usage (detoxification) compared to non-HLAU (Sup. Figure 7). In *D. m. wrigleyi* and *D. m. baja*, we also observed higher differential expression of location genes than non-HLAU. Finally, the differential expression between *D. m. wrigleyi* and *D. arizonae* demonstrated significantly higher proportion of genes associated with location and usage compared to non-HLAU in larvae, and higher prevalence of differentially expressed usage genes in head (Sup. Figure 7). Taken together, the quantitative results suggest that adaptations that have allowed the shift to columnar cacti may arise from the evolution of gene expression in both head and larvae.



Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 3: **A**) HLAU expression in head: differential expression analysis of HLAU genes from head in the *Opuntia sp.* dweller *D. m. wrigleyi* versus the three columnar dwellers *D. mojavensis,* and *D. buzzatii* versus *D. koepferae*. In A and B: Filled dots represent HLAU genes with log2 fold-change > |1| and FDR < 0.05; transparent dots are non-significant differential expression. **B**) HLAU expression in larvae: The same pairwise species as in A with *Opuntia sp.* dweller vs columnar dwellers, but from larvae tissue. **C**) Principal component analysis of HLAU genes' expression in head and larvae tissues. Both tissues had HLAU genes variance representing the phylogenetic distance between species: *D. buzzatii* and *D. koepferae* compared to *mojavensis* cluster corresponds to the PC1 (45%), and 26% between them in PC2. Species from the *mojavensis* cluster had strong similarity, with *D. arizonae* slightly separated from *D. mojavensis*. **D**) HLAU genes' expression in the *Opuntia sp.* dweller *D. m. wrigleyi* versus its sister species *D. arizonae*, which does not have preference between *Opuntia sp.* and columnar cacti.

The enrichment of TEs on HLAU genes is not associated with changes on gene expression

Since we observed enrichment of TEs on the regulatory regions of a few HS families, mainly ORs, and we also found differentially expressed HLAU genes between species, we tested whether the presence of TE insertions is associated with differential expression. To do so, we selected expressed HLAU genes (see Methods) and performed X^2 independency test, splitting them into two variables with two categories each: 1) expression as columns: differentially expressed, and non-differentially; 2) TE presence as rows: with TEs at 2kb upstream, and without TE at 2kb upstream. Considering all pairwise analysis (Fig. 3A,B-D), the X^2 results demonstrated that the differential expression of HLAU genes is not associated with the presence of TEs on the promoter region (Sup. Table 10), except in *D. m. sonorensis* compared to *D. m. wrigleyi* ($X^2 = 8.18$, *p-value = 0.004*). Since we did not observe overall effect of TEs on the expression of all HLAU genes together, we investigated the association of TEs with differential expression of HLAU genes from families with TE enrichment (Fig. 2A). None of the enriched gene families had significant association between the presence of TEs and their expression level (Sup. Table 10). Thus, enrichment of TEs at the regulatory region is not likely to be associated with the differential expression of HLAU genes.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

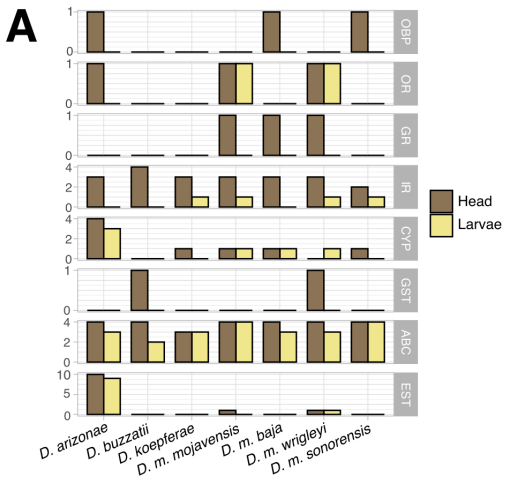
bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

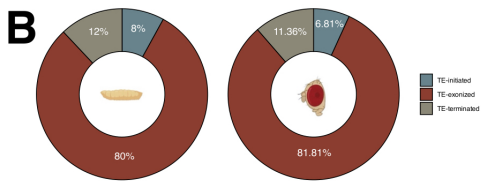
HLAU genes generate chimeric transcripts

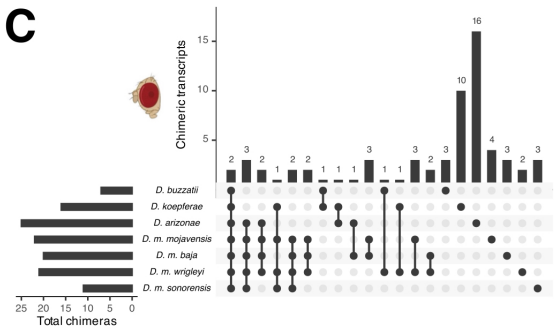
To investigate further cases where TEs produce chimeric transcripts with HLAU genes, and their possible role in the host preference between species, we focused our analysis on the nine HLAU gene families. In the head and larvae tissues, we observed an average of 5 and 3 HLAU genes producing chimeric transcripts per species/subspecies, respectively. We found more chimeric transcripts in head than in larvae for all HLAU gene families, except for UDPs (Fig. 4A). For instance, only head tissue had chimeras in the OBP, GR and GST gene families (Fig. 4A). TE-initiated transcripts constitute 11.36% of all chimeras found in head, and 12% in larvae (Fig. 4B). TE-exonized transcripts had a slightly higher proportion in larvae (80%) compared to head (81.81%). Finally, TE-terminated transcripts were the ones with the biggest difference between tissues, with 6.81% in head and 8% in larvae. Taken together, these results demonstrate unprecedented transcriptional interaction between TEs and HLAU genes in a tissue-specific manner.

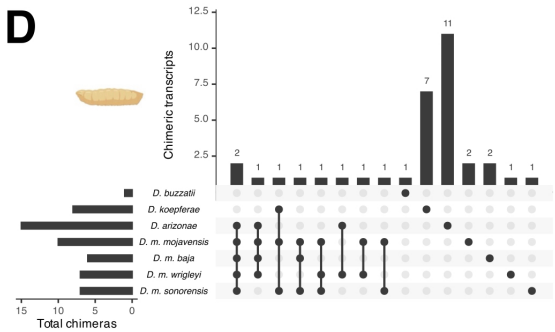
Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

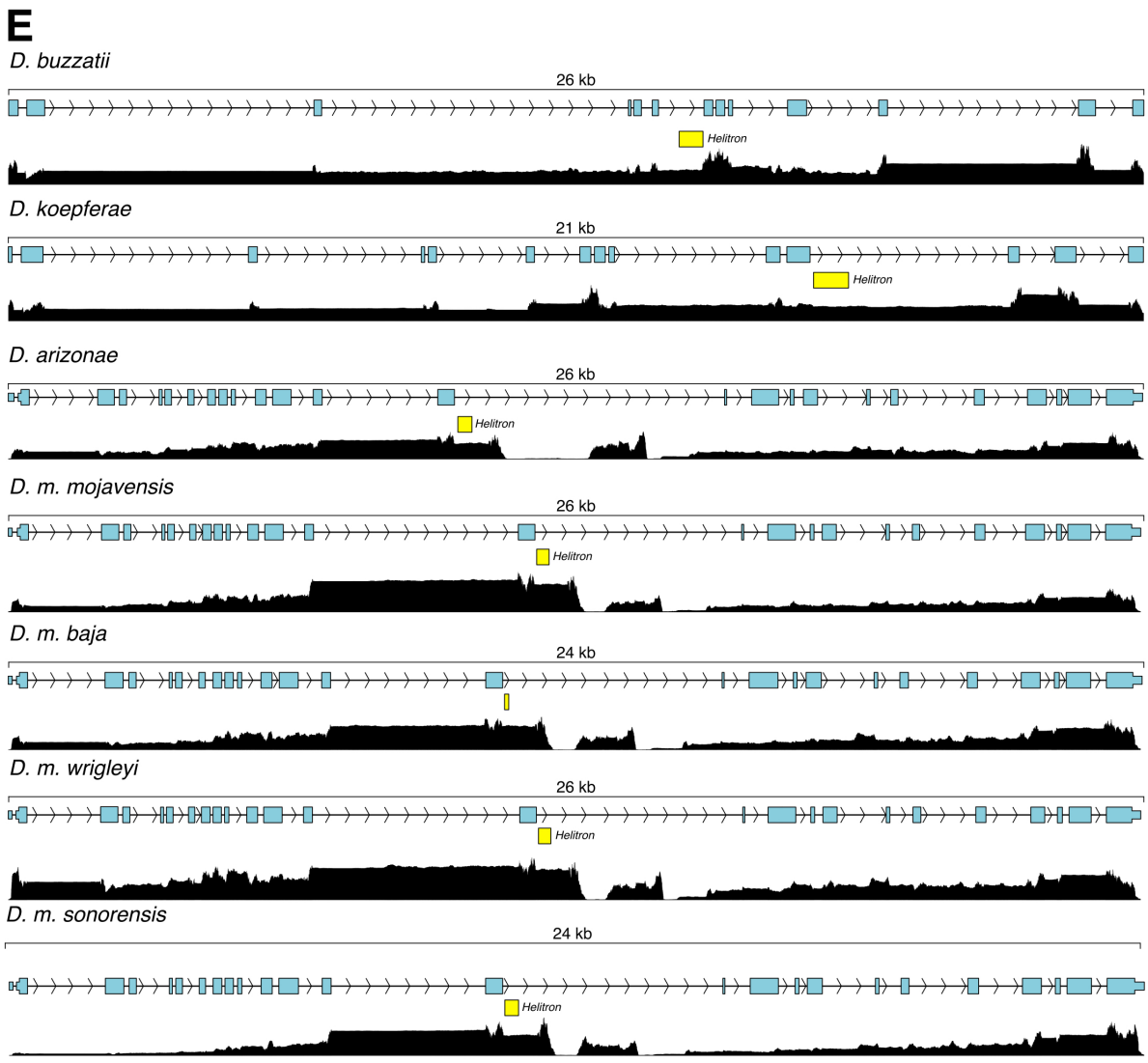
bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.











Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 4: **A**) Number of HLAU genes generating chimeric transcripts in each genome. **B**) Categories of chimeric transcripts in head and larvae tissues. **C-D**) Common chimeric transcripts between the seven genomes in head (D) and larvae (E) tissues. **E**) TE-exonized transcript from Glutamate IR, kainate 2 (LOC6585822) gene, with an *Helitron* insertion located in the intron. The same chimera was found in the seven genomes analyzed in this study.

The quantitative analysis on chimeric transcripts revealed a clear transcriptional polymorphism on HLAU genes due to TEs. We investigated common chimeras present in more than one species, highlighting their retention over evolutionary time, as well as species-specific chimeras, possibly associated with recent insertions. Our results revealed that 62.12%and 73.52% of the chimeric transcripts derived from HLAU genes are species-specific in head and larvae tissues, respectively (Fig. 4C-D). *D. arizonae* and *D. koepferae* were the ones with the highest number of HLAU genes with species-specific chimeric transcripts. In head, *D. arizonae* had 16 and *D. koepferae* 10; whereas in larvae we found 11 and 7, respectively. Regarding common chimeras, there are two expressed in head observed in all analyzed transcriptomes (Fig. 4E). The *AbcD (LOC6585769*) and *Glutamate IR, kainate 2* (*LOC6585822*) have TE-exonized transcripts expressed in the head with *Helitron* insertions. In *AbcD*, the copy is located in the intron, ~100nt away from the nearby exon, representing a case where the chimeras generated through an exon retention process. In the IR, the same process is observed since the TE is also located in the intron. Regarding the *buzzatii* cluster, we observed the gene *AbcA3* expressed only in *D. buzzatii* and *D. koepferae* with multiple insertions generating chimeras (Supp. Table 12). In the *mojavensis* cluster, two genes from the ABC gene family produced chimeras in the five transcriptomes. *Abc10* had a chimeric transcript with an *DNA/P* insertion, which is embedded in the 5' UTR. Although the insertion generates the same chimera in the seven genomes, the element is not conserved between species. In *D. m. sonorensis*, it is a 3.9kb insertion overlapping the 5' UTR of the *Abc10*, whereas in the other subspecies the TE is fragmented in 3 small insertions. The second gene is *AbcF2*, which has a TE-exonized transcript derived from LINE/R1 insertion in all species, except *D. m. sonorensis*, which has an *Helitron*. Although not the same TE in all species, the

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

copies are located partially in the intron and partially in the *AbcF2*'s exon. Both recurrent chimeras in the species of the *mojavensis* cluster were observed in larvae and head. When narrowing down our analysis for chimeras observed only among *D. mojavensis* subspecies, we did not find any.

Discussion

Insect host shift and its role in speciation have been the subject of research to understand the evolution of reproductive isolating barriers and adaptation to local environments. Indeed, many insects have finely tuned adaptations to locate, feed, reproduce, and develop on the host tissue. Adaptations to new hosts can lead to reproductive isolation, wherein ecological speciation arises as a consequence within divergent natural populations inhabiting distinct environments. In this context, we investigated whether TEs could play a role contributing to genetic novelty of HLAU genes facilitating rapid adaptive radiation to new hosts. To test this hypothesis, we selected seven *Drosophila* species/subspecies that belong to the *repleta* group and have different primary cacti hosts.

Relaxed selection on HLAU genes

The necrotic tissue of each cactus-host used by cactophilic *Drosophila* species differs in its composition in terms of yeasts and alkaloids (42,43). Therefore, species have adapted to feed and breed on different hosts, as demonstrated by behavioral and physiological response when species feed and breed in non-preferential hosts. Positive selection in the HS process acts as an evolutionary force, favoring mutations that lead to amino acid substitutions. These substitutions enhance the ability to recognize a broader spectrum of odors, metabolize nutrients, and activate detoxification pathways (44). Indeed, specialist species are likely to lose some of their genes related to sensory pathways, while the genes that are retained are subject to positive selection (45,46). A preliminary genome-wide analysis of signatures of positive selection in the four *D. mojavensis* subspecies revealed high ω rates on genes associated with chemosensory, perception, immunity, behavior, detoxification, and

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

reproduction (47). Although a previous genome-wide analysis has identified 1,294 genes under positive selection (48), here we focused the analysis to investigate specifically HLAU gene families. Our results with brach-site model did not have any HLAU gene under positive selection when considering the species' host preference. We argue that perhaps the branch-site approach was too stringent. The unprecedent method used in this work focused on the common differences of species adapted to columnar cacti, rather than independent evolutionary events. The lack of genes with statistical support for positive selection may be also associated with technical limitations, since we based our analysis on 1:1 orthologs, and a certain proportion of the genes evolve too fast to be correctly assigned during orthologs identification (49). Additionally, the ω values are higher than 1.5, for instance as *Gr59d* with ω = 3.49, which despite not significant still indicates ongoing divergence between species/subspecies.

Due to the lack of positive selection from our branch-site test, we decided to chose to apply the RELAX model to identify evidence of relaxed selection across HLAU genes. This model is particularly suited for detecting shifts from strong purifying selection to relaxed selection (50), which may not be identified by conventional methods focusing solely on positive selection, such as the one employed by CODEML. This is critical because evolutionary processes often involve not only episodes of adaptive evolution but also phases of selection relaxation, where selective pressures diminish but still influence gene evolution (51). The two HLAU genes (*Obp56d* and *Cyp304a1*) with significant relaxed selection suggest undergoing evolutionary transitions in response to changes in host preference. This result demonstrate that selective pressures on certain genes may diminish either during HS or after, leading to a period of relaxed selection as the species adapt to novel hosts. This relaxation of selection could indicate a functional change or reduced evolutionary constraint as these species exploit new ecological niches (52). Given that the species analyzed in this study have different primary hosts, the observed relaxed selection across these genes could reflect their

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

involvement in the physiological or behavioral adaptations required for successfully exploiting new hosts.

OR promoters are enriched with TE insertions

The literature has recurrently demonstrated that TEs can create new regulatory elements, including gene promoters (55–57), enhancers (58,59), and insulators (60,61). In mammals, TE-derived sequences contribute up to 40% of genome-wide binding sites for transcription factors located in TE-rich regions (62). It has been proposed that such co-option of TE-regulatory elements for gene regulatory networks can provide substantial modification of gene expression over short time scales, contributing to generate rapid genetic variability on the transcriptome level (63–65). Because the invasion of new habitats is accompanied by modifications in behavior, which has been associated with alterations in sensory cues and environment perception (66,67), our results demonstrate the potential of TEs to change gene regulation of specific HLAU gene families on several species/subspecies, and thus contributing to the HS from *Opuntia sp*. to columnar cacti.

We observed species-specific enrichment of TEs in the promoter region for CYP, GR, EST, and ABC genes, but OR gene family had the highest enrichment in five out of seven genomes from our cactophilic *Drosophila* species. In ants, ORs have also been found enriched in TEs, but in this case they were associated with tandem duplication by crossing over, allowing massive duplication events (68). In our case, since there is no expansion of ORs compared to other *Drosophila* species, the enrichment in TEs is likely to have another function, probably related with gene regulation due to its specific location in promoters. *Helitrons* have been extensively reported on promoter regions of other eukaryotic species (69–72), in which a few cases of embedded TFBSs were reported (73,74). Here, we demonstrated that ~60% of the enrichment in TEs observed at the promoter region of ORs might be due to the exaptation of TFBSs mostly derived from *Helitrons*. In addition, the recurrence of the enrichment in species of the *mojavensis* cluster might demonstrate

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

potential cases of TE co-option inherited by common ancestor. Our genomic analysis provides a robust list of potential candidate genes in which *Helitrons* might play a regulatory role, an interesting hypothesis that can be determined through cutting-edge technologies such CRISPR methods to delete insertions in site-specific manner (75).

Head transcriptome reveals subspecies specific patterns

The differential expression of HLAU genes in the head tissue revealed that location and host use (detoxification) gene families have the highest divergence across the subspecies, since we observed more differentially expressed genes for these families. Furthermore, in *D. m. wrigleyi*, which is the only subspecies with preference for *Opuntia sp*., we observed that CYP genes had on average 12 up-regulated genes compared to other species. Taken together, these results highlight that the divergence of the host preference between *D. mojavensis* subspecies might be explained by modifications in gene expression of HLAU genes in the head tissue. In addition, despite lower toxicity in *Opuntia sp*. compared to columnar cacti (76), we found a similar number of up-regulated host use genes between *D. m. wrigleyi* and the other subspecies, demonstrating that *D. m. wrigleyi* exhibits a constitutive transcriptional response for detoxification that is not driven by the host.

The extension of differentially expressed genes corresponds to 6.17% from all genes analyzed. These results demonstrate that the regulation of host location genes might be an important factor that evolved and allowed host shift. Previous studies with *D. melanogaster* have shown that the divergence in the amino acid sequences due to single polymorphic mutations in ORs and OBPs have been suggested as factor of the modulation in odor sensitivity in *Drosophila* (67,77), which could drive to divergence of host location. In addition, many ORs are involved in a large repertoire of ecological relationships, influencing not only behavior, but also oviposition and feeding (78). Another study has identified 18 ORs differentially expressed between heads of *D. m. wrigleyi* and *D. m. mojavensis* subspecies, demonstrating that modifications in the sensory transcriptome have evolved between these

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

species, as previously reported for the same tissue (79). Here, we extended the comprehension of such divergence with the pairwise comparison between *Opuntia sp.* and columnar cacti dwellers, and pointed out that each species has a large set of differentially expressed HLAU genes.

Gene expression in larvae is one of the factors shaping the adaptation to new hosts

In the larvae tissue, we found a common gene expression pattern between columnar cacti dwellers, which in turn diverge from species that use *Opuntia sp.* Location and acceptance genes had mild alterations, whereas gene families associated with host use (detoxification) had on average more up-regulated genes than the others, especially due to overexpression of CYP genes. As we have observed for the head tissue, there is also an increase of CYP genes expression in *D. m. wrigleyi* larvae compared to the other subspecies. Indeed, in *D. m. wrigleyi*, we found the highest number of CYP up-regulation, with 23 CYP genes overexpressed in comparison to *D. m. sonorensis*. Among the other subspecies that preferentially use different species of columnar cacti as hosts, we also found detoxification genes with the highest numbers of differentially expressed genes. *Opuntia sp.* produces less benzadelhyde volatiles than the columnar cacti *S. thurberi*, but both cacti are lethal to *D. melanogaster* after 48 hours (76). Therefore, the adaptation to use *Opuntia sp.* must have a constitutive gene expression of detoxification pathways in *D. m. wrigleyi*. Similarly with our observations in the head tissue, these results indicate that larvae tissue exhibits subspecies-specific expression, demonstrating part of the genetic background that allows these subspecies to be specialized in specific cacti species.

The variance in gene expression in larvae revealed that species of the *buzzatii* cluster have higher divergence compared to *mojavensis* cluster, and lower between *D. buzzatii* and *D. koepferae*, following the phylogenetic relationship between the species. Although we have species with preference for *Opuntia sp.* and for columnar cacti in both clusters, *D. koepferae* and *D. buzzatii* are adapted to different hosts for a longer time than those of the *mojavensis*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cluster. In addition, a preliminary study demonstrated that *D. buzzatii* have a high divergence on larvae gene expression compared to *D. koepferae* (80), reinforcing the variance that we have observed. Indeed, differences in larvae behavior and survival has been observed when *Opuntia*-related species are reared on columnar cacti (81,82). Our results show that divergence in the regulation of gene expression in larvae is one of the factors shaping the adaptation to new hosts.

The up-regulated HLAU genes were mostly represented by host use gene families, which are associated with detoxification. A previous study demonstrated that when *D. m. mojavensis*, which has *S. thurbery* as primary host, is reared on media with *S. gummosus*, the larval viability had a reduction of 60% (83). The authors also proposed that 21% of all genes would be differentially expressed between larvae reared in the primary and alternative cacti-host. Many genes were reported with detoxification, energy production, among others. Another similar study with larvae transcriptome, but with *D. buzzatii* and *D. koepferae*, proposed that *D. buzzatii* flies reared in columnar cacti media have a stronger detoxification response in comparison with *D. koepferae* flies, suggesting that *D. koepferae* has a canalized transcriptome response to the toxic alkaloids from columnar cacti (80). We expected to find a higher number of detoxification genes in the subspecies with preference for columnar cacti compared to *D. m. wrigleyi* and *D. buzzatii* that use primarily *Opuntia sp*. However, we observed a similar number of differentially expressed detoxification genes regardless of the host preference. Since CYP genes also have other functions than detoxification, and they were overrepresented in our results, we suggest that their differential expression might also be associated with developmental pathways in larvae (84,85). Therefore, further analysis in the biological role of up-regulated CYP genes in *D. m. wrigleyi* and *D. buzzatii* might be assessed. Alternatively, despite *Opuntia sp.* has lower toxic compounds compared to columnar cacti, it also has a basal level of toxicity (76). Thus, any cactophilic species is expected to have evolved a certain level of detoxification ability, which explains our observations. In addition, the high frequency of common HLAU genes associated with

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

detoxification on columnar cacti dweller species is expected and already demonstrated in previous works.

Transposable elements are a source of transcriptome variability between cactophilic species

TEs contribute to genome and transcriptome evolution by rewiring regulatory networks, and by incorporating protein domains into gene transcripts. In both cases, TE copies that are inserted within genes can integrate their sequences into the mRNA, generating chimeric transcripts. A recent study with different ecotypes of *D. melanogaster* has shown that these chimeras have the potential to generate transcriptome novelties (24). These chimeras have also been demonstrated to be active in a tissue-specific manner (86). Although the production of chimeric transcripts derived from polymorphic TE insertions is well-known, there is no study addressing the question on how the transcriptomic variability derived from TEs could impact the HS between species/subspecies. Here, we took advantage of our unprecedented genomic and transcriptomic data to analyze the extent of genetic variability derived from chimeric transcripts in cactophilic species, as well as HLAU genes producing chimeric transcripts.

A preliminary transcriptome-wide study with *D. melanogaster* revealed that ~1.12% of the genes produce chimeric transcripts (24). Here we observed ~1.31% on average between head and larvae tissues. The slightly higher number of genes in *repleta* species might reflect the difference between their mobilome composition compared to *D. melanogaster*, which is depleted of active *Helitrons* for example (87). Despite this difference, the proportion of TE-exonized, TE-terminated, and TE-initiated transcripts follows similar values between them, in which TE-exonized have the higher frequency with ~70% of all chimeras (24,86). In our analysis, TE-initiated transcripts represent chimeric transcripts in which the TSS of the mRNA is located within the TE insertion (24). *Helitrons* were the most frequent TEs falling in the category in four of the seven species. They have been extensively reported as a source of TFBSs when located at the promoter region of genes. In the bat *Myotis lucifugus*, 1.4% of the

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

transcripts from the salivary gland have *Helitron*-derived sequences, for which ~6% correspond to promoters (70). In the plant *Brassica napus*, an *Helitron* insertion disrupts the ancestral promoter of the gene *BnSP11-1*, altering the mating system of the populations carrying this insertion, which in turn contributes to the adaptation and speciation (88). In *Arabidopsis thaliana*, *Helitrons* have been responsible for the genome-wide distribution of the TFBS for the PHE1 TF, which is associated with crucial endosperm regulation (73). In *Caenorhabditis* species, TFBSs for the TF heat shock factor 1 have been widely spread by *Helitrons*, rewiring and diversifying heat shock response between species and strains (89). Combining our results from TFBSs from OR genes with our transcriptome-wide analysis of TE-initiated transcripts, the results reinforce that *Helitrons* have a *cis*-regulatory role in cactophilic *Drosophila* species. Regarding TE-terminated transcripts, the high prevalence of *Helitrons* in six transcriptomes also highlights their role on the termination of gene transcription. This phenomenon has been already described in bats, in which the genes *Otud3* and *Tor1b* have novel PASs from *Helitrons* (40). Another study with *A. thaliana* revealed by poly(A)-tag sequencing data that ~48% of all TE-derived PASs are from *Helitrons* (90). Although there is vast documentation of TEs acting as alternative PASs (91), here we provide the first transcriptome-wide report of *Helitrons* as a major source of putative PASs in *Drosophila*.

Chimeric transcripts reveals recent and old insertions interacting with HLAU genes

A comparative analysis between generalist and specialist *Drosophila* species revealed that niche amplitude is not likely to play a role in TE dynamics in the genomes (92). Nevertheless, several studies have proposed that the successful adaptation to new environments in invasive species might be associated with TE activity, despite reduced genetic diversity caused by founder effects (93,94). Through TE-initiated and TE-terminated transcripts, we provide evidence for TE insertions playing a regulatory role on HLAU genes, whereas TE-exonized transcripts suggest a putative source for TE-driven coding sequence evolution. Here, we demonstrate that TEs interact with genes associated with HS, providing

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

a transcriptomic variability in seven cactophilic species with different hosts preferences. We observed that species *D. arizonae,* which does not have preferential usage between *Opuntia sp.* and columnar cacti, are the ones with the highest production of unique HLAU chimeras in both head and larvae. This result demonstrates a possible relationship of broader niche and TE-driven adaptation.

The detection of chimeric transcripts is a reliable way to infer transcriptional interactions between genes and TEs. Nevertheless, even applying strict parameters as we did, the contribution of the chimeric isoform to the total gene transcript abundance is unknown. Thus, perhaps the integration of TE-derived sequences into the gene transcript is not beneficial, but still maintained in the transcriptome due to its low expression. Furthermore, chimeric transcripts might be degraded by surveillance pathways that control for aberrant mRNA production, such as no-go decay (95), non-stop decay (96), and non-sense-mediated RNA (97). Our study provides a set of HLAU genes producing chimeric transcripts, in which part of them might have phenotypic impact and ultimately association with host preference. However, it is fundamental to perform comparative phenotypic studies of host adaptation cues with mutant strains for the TEs interacting with HLAU genes identified in this study. The application of techniques of RNAi aiming chimeric transcripts, or deletion of TE insertions producing chimeras with CRISPR (75) will be able to confirm the predictions of our results.

Transposable elements driving host shift

Here, we demonstrated that HLAU genes are enriched by TEs in their regulatory regions, and produce different chimeric transcripts across species with different primary hosts. Host specificity is a key mechanism of reproductive isolation in plant pathogens, and is regulated by the repertoire of effector genes within each pathogen. Effector proteins modify host physiology, facilitating colonization by specific pathogens (98). In *Phytophthora sp.*, the host specificity is partly regulated by RXLR class effectors that facilitate host exploitation. Notably, synthetic chimeras of a short interspersed element (SINE) linked to an

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

effector gene in *P. infestans* induce silencing of effector genes (99). This silencing likely occurs naturally, as transcriptional inactivation of effectors is known to occur, and over half of RXLR effectors in the *P. infestans* genome are located in TE-rich regions (100). Consequently, TEs inserted near these genes may have influenced HS in *P. infestans*. Our findings demonstrate multiple interactions between genes and TEs, highlighting the contribution of TEs to the genome evolution of cactophilic species. Through the TE-cooption process, we postulate that TEs can play a role in the HS in these species. Although many TEs located on promoters of HLAU genes can have low frequency in natural populations, or chimeric transcripts identified in this work can be a result of ectopic expression instead of adaptive, our results demonstrate the first transcriptome-wide evidence of TE co-option in cactophilic species. The genes producing chimeras with TEs must be further studied, since the lack of gene-phenotype association between HLAU genes is not fully reported, preventing us to pinpoint specific relation between chimeras and possibly reproductive isolation.

The existence of reproductive barriers between populations that undergone a host shift is the key process to ecological speciation. In our model, with species from the *buzzatii* and *mojavensis* clusters, reproductive isolation has been reported between *D. m. sonorensis* and *D. m. baja* (26,101,102). The premating isolation between them occurs due to the difference on cuticular hydrocarbons, which participate in the signaling pathway of mate recognification (103). Importantly, artificially shifting the cactus-host was the main cause of changes on the hydrocarbon profiles. Thus, the source of reproductive barriers are dependent of the cacti species on which the flies feed and breed (104,105). The close relationship between HLAU genes to the HS process, and the observed contribution of TEs to the evolution of these genes, supports our proposal that TEs may play a role in the HS of cactophilic species. Hence, such process rises the hypothesis that TEs can play a role in HS. Hence, such process rises the hypothesis that TEs could drive reproductive isolation via adaptation to different environments, ultimately driving ecological speciation.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Conclusions

TEs have been shown to contribute to genome evolution of cactophilic *Drosophila* species. Their enrichment in promoter regions of HLAU gene families, along with potential presence of TFBSs within their sequences highlights their potential role in the evolution of adjacent genes. Additionally, their frequent exonization reveals a mechanism by which TEs increase genetic diversity between populations, potentially contributing to host shifts. Taken together, our combined genomic and transcriptomic approaches provide new insights regarding TE-related mechanisms that promote genome evolution between species and subspecies with different primary cactus hosts.

Material and methods

Fly stocks and genome sequencing

Most of the strains were obtained from the UC San Diego *Drosophila* Stock Center: *[D. m. mojavensis*: Anza (01), Anza Borrego Desert, California, USA, Stock Center n°: 15081-1352.01; *D. m. wrigleyi*: CI (22), Catalina Island, California, USA, Stock Center n°: 15081-1352.22; *D. m. baja*: BC (20), Cape Region, Santiago, Baja California Sur, México, Stock Center n°: 15081-1352.20 e *D. m. sonorensis*: AG (26), Agiabampo Bay, Sonora, México, Stock Center n°: 15081-1352.26] and *D. arizonae* [HI (17), collected in Metztitlan, Hidalgo, Mexico, Stock Center n°: 15081-1271.17]. *D. buzzatii* (Bu28) and *D. koepferae* (Ko2) stocks correspond to two laboratory lines used in previous works (106,107). Both stocks were maintained by brother-sister mating for more than a decade and then kept by mass culturing.

We used the same following protocols of DNA extraction and genome sequencing to all samples. DNA was extracted from 10 males and 10 females from each species using the Qiagen DNeasy Blood&Tissue kit. Then, we evaluated the genomic DNA amount and quality with NanoDropTM One UV-Vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit ® 1.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). Three micrograms of DNA were repaired using the NEBNext FFPE DNA Repair Mix (NEB M6630). We performed end

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

repair and dA-tailing using the NEBNext End repair/dA-tailing Module (E7546, NEB). Ligation was then assessed with the Ligation Sequencing Kit 1D. MinION sequencing was performed according to the manufacturer's guidelines using R9.4.1 flow cells (FLO-MIN106, ONT) and a Nanopore MinIon Mk1b sequencer (ONT) controlled by *ONT MinKNOW v.18.3.1*. Base calling was performed after sequencing using *Guppy v.4.0.5* in high accuracy mode for isogenic wild-type strains *v.3.1.5*.

Genome assemblies and gene annotation

The quality control for Nanopore reads was performed with *Nanoplot* v.1.10.2 (https://github.com/wdecoster/NanoPlot). Reads with quality lower than 7 were removed from downstream analysis. In order to assemble contigs, *Flye v.2.8* (108) was used with default parameters, except –plasmids. All contigs were aligned with *minimap2* v2.16 (109), with –x map-ont option. The alignment was used to perform the polishing of contigs with four rounds of *RACON* v1.3.2 (110), default parameters. The assembly quality metrics was assessed with *Assembly-Stats* v1.0.1 (https://github.com/sanger-pathogens/assembly-stats). Assembly incongruences were manually visualized with *D-genies* v1.2.0 (111) and corrected with *samtools* v1.9.0. (112), function *faidx*; and *Gepard* v1.4.0(113) for determination of breaking points. The super scaffolding of all corrected assemblies was performed with *RaGOO* v.1.1 (114), with –s and –t 4 parameters, using the respective reference genome assembly for each species. Subsequently, the chromosome-scale assemblies were submitted to a Benchmarking Universal Single-Copy Orhologs (BUSCO) v.5.4.3 analysis (115) with lineage *diptera_odb10* (*-l* parameter). The gene annotations for *D. mojavensis* subspecies and *D. arizonae* were performed with the software *Liftoff v.1.6.3* (116), with the reference genome of *D. mojavensis* (GCA_018153725.1) (117). We have used the parameters *–exclude_partial*; *-a 0.5* and *–s 0.75* to keep annotated genes with at least 75% of the reference gene length. We removed from our gene annotation HLAU genes flagged as "low quality protein" and "unestablished gene function" in the reference *D. mojavensis* genome.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

D. buzzatii and *D. koepferae* genomes were annotated with a de novo strategy with the software *BRAKER3 v.3.0.8* (118). To optimize the prediction of gene structures and repertoire of isoforms, we used a mix of RNA-seq data from different tissues publicly available and produced in this work (Sup. Table 11), in addition to protein sequences from *D. mojavensis* as reference. Ortholog genes between *D. buzzatii, D. koepferae*, *D. mojavensis* and *D. arizonae* were obtained with *Orthofinder v.2.5.4* (119) using *DIAMOND v.0.9.14* (120) as search engine and default parameters. One-to-one orthology were firstly obtained directly from orthofinder, and remaining proteins without an ortholog pair were analyzed with *blastp v.2.5.0+* (121), for which we considered as one-to-one proteins only those with a single homologous with parameters *-qcov_hsp_perc 95* and identity >= 90%. Since *Orthofinder* performs its search with protein sequences, any non-coding RNA gene has been identified. Then, we used *Liftoff v.1.6.3* (122) with all non-coding RNA genes from *D. mojavensis* as reference to maximize our homology-based gene annotation.

Molecular evolution analysis

To identify HLAU genes with signatures of positive selection in species that have preference for columnar cacti compared to species with preference for *Opuntia sp.*, we selected the four *D. mojavensis* subspecies, *D. navojoa*, *D. koepferae* and *D. buzzatii*. *D. arizonae* species was removed from this analysis due to its generalist preference for both columnar and *Opuntia* cacti. The *D. navojoa* genome was retrieved from NCBI (GCF_001654015.2), which was produced with Illumina reads (123). For the other Nanopore genomes sequenced in this work, we performed a polishing with Illumina paired-end reads. We used the software *NextPolish* (124) to fix base errors in the assembled genomes. The four *D. mojavensis* subspecies were polished with genomic Illumina reads sequenced previously (125), whereas for *D. koepferae* and *D. buzzatii* we used publicly available genomic Illumina reads (126) and RNA-seq reads (48), respectively. In the latter, to increase the read coverage and optimize the polishing to the higher number of genes and exons as possible, we merged RNA-seq reads from adult male and female (48). All genomes were corrected using two

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

rounds of polishing, as recommended by the authors (124). Then, we extracted the longest CDS sequences for all genes combining AGAT (127) *agat_extract_sequences.pl* parameters *-t mrna*, *samtools faidx* (128), and *seqtk* (https://github.com/lh3/seqtk). Posteriorly, we intersected the occurrence of orthologous genes between the seven genomes, recovering a list of genes that were present in all of them. The sequences were aligned using *MAFFT v.7.487* [67] and the non-aligned codons were removed using *pal2nal* (129). To obtain a strongly supported tree for this analysis, we inferred the species phylogeny tree with 2,043 shared BUSCO genes. The phylogeny was estimated using the concatenated alignment (3,730,914 sites) with *IQ-TREE v. 2.2.2.6* (130), using the best model for each gene, 10,000 replicates of ultrafast bootstrap and 1000 replicates for Approximate Likelihood-Ratio Test (ALRT) (parameters *-m MFP -bb 10000 -alrt 1000*). Finally, we carried out analysis to detect signatures of selection analysis for each gene individually with *CODEML* from *PAML v.4.10.7* (131), using *branch-site* approach. In order to find signatures of positive selection in species that use columnar cacti, we marked the respective branches of *D. koepferae*, *D. m. baja*, *D. m. mojavensis* and *D. m. sonorensis*, with the same label (#1), to compare them with the background species that use *Opuntia sp.* (*D. buzzatii*, *D. m. wrigleyi*). For this, we tested two hypotheses: (i) H0 – Null hypothesis, in which all branches evolved under negative selection (model = 2, Nssites = 2, fix_omega = 1, ômega = 1); (ii) H1 – Alternative hypothesis, in which the labeled branches (foreground) have sites with signatures of positive selection (model = 2, NSsites = 2, fix_omega = 0, omega = 1.5). The hypotheses were tested through chi-square test, calculated by the log-likelihood (lnL) ratio between H1 and H0 (2ΔlnL) and degrees of freedom obtained from the difference between the number of estimated parameters for H1 and H0 (Δnp). Sites under positive selection were identified by Bayes Empirical Bayes and posterior probability, implemented in *CODEML*.

We used the software *HYPHY* (132) to test relaxed selection with the same phylogeny framework used for branch-site analysis, using the RELAX model (50). The phylogenetic tree had its branches corresponding to columnar cacti dwellers as foreground, and *Opuntia sp.* as

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

background. We used the parameters *--test Foreground* and *--model Minimal*. Multiple testing correction was performed with False discovery rate (FDR). Genes with FDR < 0.05 and k > 1 were considered under intensified selection.

RNA extraction and sequencing

We performed RNA-seq for larvae and head tissues from *D. buzzatii*, *D. koepferae*, *D. arizonae* and the four *D. mojavensis* subspecies. The RNA extraction was assessed with RNeasy Plus Kit for 10 female heads (10 days old) and 15 larvae with three technical replicates. The library preparation was done with stranded mRNA-seq/standard quantity (ligation), and 100 bp paired-end reads were produced in an Illumina Hiseq 4000. Illumina adaptors and low quality reads were removed for downstream analysis with *Trimommatic v.0.39* (133), parameters*: LEADING:3*; *TRAILING:3*; *SLIDINGWINDOW:4:15*; *HEADCROP:10*; *CROP:90*.

TE annotation

For each sequenced genome, we developed a novel pipeline combining automatic and manual approaches to construct TE consensuses. First, we used *Earl Grey v.2.2* (134) to build de novo TE libraries with *RepeatModeler2 v.2.0.4* (135), which uses RECON *v.1.08* (136) and *RepeatScout v.1.0.6* (137). Then, in order to retrieve full-length TE consensus, the pipeline performs an extension of each predicted consensus sequence using a method named "Blast, Extract, Extend" (BEE) (138). Such method extracts the sequences of the 20 full-length (or longest) TE insertions for each consensus, and then it retrieves the 1 kb flanking regions from them. Posteriorly, insertions altogether with flanking regions are aligned with *MAFFT v.7.487* (139), and the alignment is trimmed with *trimAl v.1.4* (140), parameters *–gt 0.6* and *–cons 60*. Subsequently, the alignment is analyzed to check whether the flanking regions from the full-length insertions have high quality, meaning that they are part of the TE rather than genomic regions. Finally, the consensus sequence for each TE family is updated with *EMBOSS v.6.6.0.0* (141). The BEE method is repeated up to five times for each consensus. Once the extended

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

TE libraries were obtained for the seven genomes, they were submitted to a CDS filtering step, in which we removed all consensus with high similarity with *D. mojavensis wrigleyi* CDSs using blastn, parameter *-qcov_hsp_perc 80*. A last filter step was applied to remove TE consensus with less than 3 *RepeatMasker* hits across the genome, with at least 80% of its length. A similar cutoff has been applied to *D. suzukii* TE annotation (93).

TE consensus that had passed in our filtering thresholds were then submitted to the classification analysis. We developed a pipeline to perform similarity analysis of non-reference consensus with a database provided by the user (https://github.com/OliveiraDS-hub/Pipelines-Cactophilic-Drosophila-Species). The pipeline identifies similarity analysis between the non-ref. consensus and a database provided. All consensuses were masked with *RepeatMasker v.4.1.4* (142), parameters *-lib*, *-cutoff 250*, *-s*, and *-norna*. Then, the pipeline performs different steps depending on the number and variety of hits that the non-ref. consensus has: (1) only one hit is considered unique and the total length of the masked hit is compared with the total length of the non-ref insertion; (2) multiple hits, but all from the same class and family: the length of the hits is summed up, and the total length is compared with the length from the non-ref consensus; (3) multiple hit, but different classes and/or families: in this case each class is treated separately. Due to the possibility of overlapped hits from TE families of the same class in the non-ref consensus, the pipeline removes the extent of overlapped regions from two insertions, hence decreasing the length of the insertions. Once all hits from a given class have no overlap, the pipeline sum up the total extent of the hits and compare it with the non-ref consensus. The process is repeated for all classes masked. Finally, the pipeline classifies only TEs with a corresponding match in >= 50% over their total length. In this work, we have used Dfam v.3.7 (143) as reference database. Due to the "curated" library contains only *D. melanogaster* consensuses, we used the "uncurated" library from Dfam. We extracted *Drosophila* TEs with *famdb.py v.0.4.3*, parameters *-i Dfam.h5 families --include-class-in-name -f fasta_name -ad 'Diptera'*. Then, we filtered out non-Drosophila sequences, and "Unknown" TEs with the unix commandline: *grep -i 'drosophila'*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

file.fa | *grep -v -i 'unknown'*, obtaining 49,916 out of 344,562 consensuses. Finally, we used all obtained IDs with *seqtk v.1.3* (https://github.com/lh3/seqtk) to subset the complete fasta file generated by *famdb*.

We used the polished TE libraries constructed to the seven cactophilic *Drosophila* genomes to perform TE annotation with *RepeatMasker v.4.1.4* (142), parameters *-cutoff 250*, *-s*, and *-norna*. Then, we have used *RepeatCraft* (144) with *LTR Finder v.1.2* results to merge LTRs and their respective internal regions. Subsequently, we removed all TE insertions shorter than 80 nucleotides, following the classification rule 80-80-80 (145). Finally, we performed a last filtering step in the TE annotation to remove insertions comprised by simple sequence repeat (SSR). We used *TRF v.4.09* (146) in all TE insertions from each genome, with parameters *2*, *5*, *6*, *75*, *20*, *50*, *500*, *-m*, *-h*, and removed TEs that had SSRs over more than 50% of their sequences. Such method has been previously used to remove TE copies that are indistinguishable from simple repeats (24,147).

TE enrichment at regulatory regions

To identify TE insertions within and nearby genes for each genome, we created *bed* files with the genomic positions of all insertions. Then, we created *bed* files for the gene regions corresponding to the 2 Kb upstream to the transcription start-site (TSS). In order to test whether the prevalence of TEs in each gene region was significantly higher than expected (enriched) compared to random genomic distributions of TEs in the same region of other genes, we performed permutation tests with the R package *regioneR* (148), parameter *randomization = resampleRegions*, with 2,000 random samples. Therefore, for each gene family with *n* genes, the number of genes with TEs was compared to other random *n* genes 1,000 times. Such analysis allowed us to infer whether there are HLAU gene families with the observed frequency of genes with TEs at the promoter region higher than expected compared to the rest of the genome. We also selected zinc-finger proteins and serine/threonine kinases as two non-HLAU genes families to compare our findings. Taking

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

into account that TEs might harbor regulatory and/or coding motifs in both strands (65), our analysis considered TEs within the investigated gene regions regardless the strand.

Identification of TE-rich genomic regions

We split the genomes in bins of 200 kb, generating a bed file with such intervals across the genomes. Then, we used the function *intersect* from *bedtools v.2.30.0* (149) to identify the number of TE insertions located in each genomic bin. Finally, we calculated the 3rd quartile with the frequency of TEs in each region and defined as TE-rich genomic regions those with more TE insertions than the 3rd quartile. We used the same function from *bedtools* to identify whether genes were present in the TE-rich regions, or in the TE-poor (less TEs than the 3rd quartile).

Transcription factor binding sites on TE insertions

We assessed the identification of transcriptions factor binding sites (TFBSs) with *FIMO* (150), from the *MEME v. 5.0.5* suite tools (151). We extracted the sequences of all TE insertions located 2kb upstream of ORs with *bedtools v. 2.30.0*, function *getfasta*, using parameter *-s* (149). We obtained the TFBS motifs described for *D. melanogaster* from Jaspar database (152), for the transcription factors (TFs) acj6, zf30c, fer1, onecut, and xbp1, which have been demonstrated to activate ORs (41). To control our findings with OR genes, we used IR as a HLAU gene family member, and Serine/Threonine Kinases as a non-HLAU gene family. We performed The same process of TFBSs identification performed for ORs on these two gene families. We only considered TEs with TFBSs those with p-value < 0.01 and score > 250.

Differential expression of genes

We aligned the RNA-seq reads with *STAR v.2.7.3* (153), using the assembled genomes sequences and annotations of their respective species. The alignments were used to create the count table reporting the expression of each gene, with *htseq v.0.13.5* (154), parameters *–order pos*; *-s reverse*; *–i gene*; *-t exon*; the latter indicates that the count for each gene

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

corresponds to the sum of all reads aligned into exons (CDSs and UTRs). We have merged the counts from species of the *buzzatii* and *mojavensis* clusters separately, maintaining only ortholog genes that have at least 90% of the length of the annotated gene in the reference genome. Such threshold allowed us to avoid misinterpretation of expression level by comparing gene with different lengths. Then, we performed independent statistical analysis for species of the *buzzatii* and *mojavensis* clusters with *DEseq2* v.1.28.1 (155), using single factor experimental design, which considers only the factor that different species may have differentially expressed genes. Due to the phylogenetic distance between species of the *buzzatii* and *mojavensis* clusters, we performed the differential expression analysis separately for each group. We only considered genes with adjusted *p-value* < 0.05 and *log2 fold-change* > |1| as differentially expressed. To test for association between TE enrichment and differential expression, we selected HLAU genes with the mean of normalized counts between replicates >= 100. Then, we build the contingency table to apply X^2 independency test splitting them into four categories: differentially expressed, non-differentially expressed, with TEs at promoter region, and without TEs at the promoter region. We carried out the test for all pairwise analysis between cactophilic species with preference for *Opuntia sp* and columnar cacti. The proportional differences between location, acceptance and usage genes compared to non-HLAU genes (Ser/Thr Kinases and zinc-finger gene families) were assessed with Fisher's exact test, with contingency table containing average number of genes per each group: 1) non-HLAU differentially expressed genes; 2) non-HLAU non-differentially expressed genes; 3) HLAU differentially expressed genes; and 4) HLAU non-differentially expressed genes. *Fisher's* Exact test < 0.05 were considered as significant.

Identification of gene-TE chimeric transcripts

We detected gene-TE chimeric transcripts with *ChimeraTE v.1.1.1* Mode 1 (24). Each species was analyzed using its respective genome, gene annotation, and TE annotation. Shortly, *ChimeraTE* identifies chimeras based on paired-end reads spanning from exons to TE sequences. Based on the TE position compared to the gene (upstream, exon, intron,

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

downstream), chimeric transcripts are classified in categories. We have used default parameters, except *--replicate 3*, which consider true chimeras only those identified in the three RNA-seq replicates.

Declarations

Availability of data and materials

The quality control data from both ONT and RNA-seq data, the genome assemblies and their respective gene and TE annotation (libraries and gtf files), and the data used to produce phylogenetic are available in the Zenodo repository, doi: https://zenodo.org/records/13117512. The code to reproduce the analyses are available on Github repository: https://github.com/OliveiraDS-hub/Pipelines-Cactophilic-Drosophila-Species.

Competing interests

The authors declare that they have no competing interests.

Funding

The present study has been funded by Brazilian agencies FAPESP- The São Paulo Research Foundation [grant number 2020/06238-2] and from the National Council for Scientific and Technological Development (308020/2021-9) to C. M. A. C; the Eiffel Fellowship Campus France [P769649C] and CAPES – PrInt [88887.716810/2022-00] to D. S. O.; by Ministerio de Ciencia e Innovación (Spain) [grant number PID2021-127107NB-I00], and Generalitat de Catalunya (Spain) [grant number 2021 SGR 00526] to M. P. G. G.; and by UdI-FAPESP ANR-IDEX-0005 to C. V.

Authors' contributions

A. L., A. B, M. P. G. G, performed experiments; F.S., Performed genome assemblies; W. V. B. N. performed phylogenetic analysis; D. S. O performed computational analysis for genomics and transcriptomics analysis; D. S. O. wrote and edited the paper; C. V. and C. M. A. C.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

designed the study, and reviewed the manuscript. All authors read, corrected, and approved the final manuscript.

Acknowledgements

We thank Séverine Chambeyron for useful discussions and advice. This work was performed using the computing facilities of the CC LBBE/PRABI.

References

1. Whiteman NK, Pierce NE. Delicious poison: genetics of Drosophila host plant preference. Trends in Ecology & Evolution. 2008 Sep 1;23(9):473–8.
2. Forbes AA, Devine SN, Hippee AC, Tvedte ES, Ward AKG, Widmayer HA, et al. Revisiting the particular role of host shifts in initiating insect speciation. Evolution. 2017 May;71(5):1126–37.
3. Rundle HD, Nosil P. Ecological speciation. Ecology Letters. 2005;8(3):336–52.
4. Howard DJ, Berlocher SH. Endless Forms: Species and Speciation. Oxford University Press; 1998. 486 p.
5. O'Grady PM, Markow TA. Rapid morphological, behavioral, and ecological evolution in Drosophila: comparisons between the endemic Hawaiian Drosophila and the cactophilic repleta species group. In: Singh RS, Xu J, Kulathinal RJ, editors. Rapidly Evolving Genes and Genetic Systems [Internet]. Oxford University Press; 2012 [cited 2024 Jul 1]. p. 0. Available from: https://doi.org/10.1093/acprof:oso/9780199642274.003.0018
6. Oliveira DCSG, Almeida FC, O'Grady PM, Armella MA, DeSalle R, Etges WJ. Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the Drosophila repleta species group. Molecular Phylogenetics and Evolution. 2012 Sep;64(3):533–44.
7. Stintzing FC, Carle R. Cactus stems (Opuntia spp.): A review on their chemistry, technology, and uses. Molecular Nutrition & Food Research. 2005;49(2):175–94.
8. Gomez GA. Transpecific Polymorphisms in an Inversion Linked Esterase Locus in Drosophila buzzatii. Molecular Biology and Evolution. 2003 Mar 1;20(3):410–23.
9. Hasson E, Naveira H, Fontdevila A. The breeding sites of the Argentinian species of the Drosophila mulleri complex (subgenus Drosophila-repleta group). Rev Chil Hist Nat. 1992 Jan 1;65:319–26.
10. Newby BD, Etges WJ. Host Preference Among Populations of Drosophila mojavensis (Diptera: Drosophilidae) that Use Different Host Cacti. Journal of Insect Behavior. 1998 Sep 1;11(5):691–712.
11. Rane RV, Clarke DF, Pearce SL, Zhang G, Hoffmann AA, Oakeshott JG. Detoxification Genes Differ Between Cactus-, Fruit-, and Flower-Feeding *Drosophila*. Journal of Heredity. 2019 Jan 7;110(1):80–91.
12. Markow TA. Host use and host shifts in Drosophila. Current Opinion in Insect Science. 2019 Feb;31:139–45.
13. Montell C. A taste of the Drosophila gustatory receptors. Current Opinion in Neurobiology. 2009 Aug;19(4):345–53.
14. Ni L. The Structure and Function of Ionotropic Receptors in Drosophila. Front Mol Neurosci. 2021 Feb 1;13:638839.
15. Chen Y, Amrein H. Ionotropic Receptors Mediate Drosophila Oviposition Preference through Sour Gustatory Receptor Neurons. Current Biology. 2017 Sep;27(18):2741-2750.e4.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

16. Gomez-Diaz C, Martin F, Garcia-Fernandez JM, Alcorta E. The Two Main Olfactory Receptor Families in Drosophila, ORs and IRs: A Comparative Approach. Front Cell Neurosci. 2018 Aug 30;12:253.
17. Rane RV, Clarke DF, Pearce SL, Zhang G, Hoffmann AA, Oakeshott JG. Detoxification Genes Differ Between Cactus-, Fruit-, and Flower-Feeding *Drosophila*. Journal of Heredity. 2019 Jan 7;110(1):80–91.
18. Benowitz KM, Coleman JM, Allan CW, Matzkin LM. Contributions of cis- and trans-Regulatory Evolution to Transcriptomic Divergence across Populations in the Drosophila mojavensis Larval Brain. Betran E, editor. Genome Biology and Evolution. 2020 Aug 1;12(8):1407–18.
19. Fedoroff NV. Transposable Elements, Epigenetics, and Genome Evolution. Science. 2012 Nov 9;338(6108):758–67.
20. Capy P. Taming, Domestication and Exaptation: Trajectories of Transposable Elements in Genomes. Cells. 2021 Dec 20;10(12):3590.
21. Drongitis D, Aniello F, Fucci L, Donizetti A. Roles of Transposable Elements in the Different Layers of Gene Expression Regulation. IJMS. 2019 Nov 15;20(22):5755.
22. Magwire MM, Bayer F, Webster CL, Cao C, Jiggins FM. Successive Increases in the Resistance of Drosophila to Viral Infection through a Transposon Insertion Followed by a Duplication. Begun DJ, editor. PLoS Genet. 2011 Oct 20;7(10):e1002337.
23. Lipatov M, Lenkov K, Petrov DA, Bergman CM. Paucity of chimeric gene-transposable element transcripts in the Drosophila melanogaster genome. BMC Biol. 2005 Dec;3(1):24.
24. Oliveira DS, Fablet M, Larue A, Vallier A, Carareto CMA, Rebollo R, et al. ChimeraTE: a pipeline to detect chimeric transcripts derived from genes and transposable elements. Nucleic Acids Research. 2023 Oct 13;51(18):9764–84.
25. Matzkin LM. The Molecular Basis of Host Adaptation in Cactophilic Drosophila: Molecular Evolution of a Glutathione *S* -Transferase Gene (*GstD1*) in *Drosophila mojavensis*. Genetics. 2008 Feb 1;178(2):1073–83.
26. Pfeiler E, Castrezana S, Reed LK, Markow TA. Genetic, ecological and morphological differences among populations of the cactophilic Drosophila mojavensis from southwestern USA and northwestern Mexico, with descriptions of two new subspecies. Journal of Natural History. 2009 Apr 1;43(15–16):923–38.
27. Etges WJ. Genetics of Host-Cactus Response and Life-History Evolution Among Ancestral and Derived Populations of Cactophilic Drosophila Mojavensis. Evolution. 1993;47(3):750–67.
28. Crowley-Gall A, Date P, Han C, Rhodes N, Andolfatto P, Layne JE, et al. Population differences in olfaction accompany host shift in *Drosophila mojavensis*. Proc R Soc B. 2016 Aug 31;283(1837):20161562.
29. Rajpurohit S, Oliveira CC, Etges WJ, Gibbs AG. Functional genomic and phenotypic responses to desiccation in natural populations of a desert drosophilid. Molecular Ecology. 2013;22(10):2698–715.
30. De Panis D, Dopazo H, Bongcam-Rudloff E, Conesa A, Hasson E. Transcriptional responses are oriented towards different components of the rearing environment in two Drosophila sibling species. BMC Genomics. 2022 Dec;23(1):515.
31. Matzkin L, Markow TA. Transcriptional Differentiation Across the Four Subspecies of Drosophila mojavensis. In: Speciation: Natural Processes, Genetics and Biodiversity. New York: Nova Scientific Publishers; 2013. p. 119–35.
32. Benowitz KM, Allan CW, Jaworski CC, Sanderson MJ, Diaz F, Chen X, et al. Chromosome-length genome assemblies of cactophilic Drosophila illuminate links between structural and sequence evolution [Internet]. bioRxiv; 2022 [cited 2023 Nov 2]. p. 2022.10.16.512445. Available from: https://www.biorxiv.org/content/10.1101/2022.10.16.512445v1

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

33. Oliveira DCSG, Almeida FC, O'Grady PM, Armella MA, DeSalle R, Etges WJ. Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the Drosophila repleta species group. Molecular Phylogenetics and Evolution. 2012 Sep;64(3):533–44.
34. Sanchez-Flores A, Peñaloza F, Carpinteyro-Ponce J, Nazario-Yepiz N, Abreu-Goodger C, Machado CA, et al. Genome Evolution in Three Species of Cactophilic *Drosophila*. G3 Genes|Genomes|Genetics. 2016 Oct 1;6(10):3097–105.
35. Smith G, Lohse K, Etges WJ, Ritchie MG. Model-based comparisons of phylogeographic scenarios resolve the intraspecific divergence of cactophilic *Drosophila mojavensis*: PHYLOGEOGRAPHY OF *DROSOPHILA MOJAVENSIS*. Molecular Ecology. 2012 Jul;21(13):3293–307.
36. Guio L, Vieira C, González J. Stress affects the epigenetic marks added by natural transposable element insertions in Drosophila melanogaster. Sci Rep. 2018 Aug 15;8(1):12197.
37. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, et al. A Single P450 Allele Associated with Insecticide Resistance in *Drosophila*. Science. 2002 Sep 27;297(5590):2253–6.
38. Chang NC, Rovira Q, Wells J, Feschotte C, Vaquerizas JM. Zebrafish transposable elements show extensive diversification in age, genomic distribution, and developmental expression. Genome Res. 2022 Jul;32(7):1408–23.
39. Pritham EJ, Feschotte C. Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. Proceedings of the National Academy of Sciences. 2007 Feb 6;104(6):1895–900.
40. Thomas J, Phillips CD, Baker RJ, Pritham EJ. Rolling-Circle Transposons Catalyze Genomic Innovation in a Mammalian Lineage. Genome Biology and Evolution. 2014 Oct;6(10):2595–610.
41. Jafari S, Alkhori L, Schleiffer A, Brochtrup A, Hummel T, Alenius M. Combinatorial activation and repression by seven transcription factors specify Drosophila odorant receptor expression. PLoS Biol. 2012;10(3):e1001280.
42. Fogleman JC, Danielson PB. Chemical Interactions in the Cactus-Microorganism-Drosophila Model System of the Sonoran Desert1. American Zoologist. 2001 Aug 1;41(4):877–89.
43. Fogleman JC, Heed WB, Kircher HW. Drosophila mettleri and senita cactus alkaloids: Fitness measurements and their ecological significance. Comparative Biochemistry and Physiology Part A: Physiology. 1982 Jan;71(3):413–7.
44. Markow TA. Host use and host shifts in Drosophila. Current Opinion in Insect Science. 2019 Feb;31:139–45.
45. McBride CS. Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. Proceedings of the National Academy of Sciences. 2007 Mar 20;104(12):4996–5001.
46. McBride CS, Arguello JR. Five Drosophila Genomes Reveal Nonneutral Evolution and the Signature of Host Specialization in the Chemoreceptor Superfamily. Genetics. 2007 Nov 1;177(3):1395–416.
47. Allan CW, Matzkin LM. Genomic analysis of the four ecologically distinct cactus host populations of Drosophila mojavensis. BMC Genomics. 2019 Oct 12;20(1):732.
48. Guillén Y, Rius N, Delprat A, Williford A, Muyas F, Puig M, et al. Genomics of Ecological Adaptation in Cactophilic Drosophila. Genome Biology and Evolution. 2015 Jan;7(1):349–66.
49. Kristensen DM, Wolf YI, Mushegian AR, Koonin EV. Computational methods for Gene Orthology inference. Briefings in Bioinformatics. 2011 Sep 1;12(5):379–91.
50. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. Molecular Biology and Evolution. 2015 Mar 1;32(3):820–32.
51. Lahti DC, Johnson NA, Ajie BC, Otto SP, Hendry AP, Blumstein DT, et al. Relaxed selection

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in the wild. Trends in Ecology & Evolution. 2009 Sep 1;24(9):487–96.
52. Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biology and Evolution. 1994 Sep 1;11(5):725–36.
53. Arguello JR, Cardoso-Moreira M, Grenier JK, Gottipati S, Clark AG, Benton R. Extensive local adaptation within the chemosensory system following Drosophila melanogaster's global expansion. Nat Commun. 2016 Jun 13;7(1):ncomms11855.
54. Rane RV, Pearce SL, Li F, Coppin C, Schiffer M, Shirriffs J, et al. Genomic changes associated with adaptation to arid environments in cactophilic Drosophila species. BMC Genomics. 2019 Jan 16;20(1):52.
55. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet. 2009 May;41(5):563–71.
56. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, et al. Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. Hoekstra HE, editor. PLoS Genet. 2013 Apr 25;9(4):e1003470.
57. Emera D, Casola C, Lynch VJ, Wildman DE, Agnew D, Wagner GP. Convergent Evolution of Endometrial Prolactin Expression in Primates, Mice, and Elephants Through the Independent Recruitment of Transposable Elements. Molecular Biology and Evolution. 2012 Jan;29(1):239–47.
58. Ye M, Goudot C, Hoyler T, Lemoine B, Amigorena S, Zueva E. Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. Proc Natl Acad Sci USA. 2020 Apr 7;117(14):7905–16.
59. Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, et al. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nat Genet. 2013 Jul;45(7):836–41.
60. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves Â, Kutter C, et al. Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages. Cell. 2012 Jan;148(1–2):335–48.
61. Wang J, Lunyak VV, Jordan IK. Genome-wide prediction and analysis of human chromatin boundary elements. Nucleic Acids Research. 2012 Jan;40(2):511–29.
62. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, et al. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res. 2014 Dec;24(12):1963–76.
63. Gill RA, Scossa F, King GJ, Golicz AA, Tong C, Snowdon RJ, et al. On the Role of Transposable Elements in the Regulation of Gene Expression and Subgenomic Interactions in Crop Genomes. Critical Reviews in Plant Sciences. 2021 Mar 4;40(2):157–89.
64. Hof AE van't, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, et al. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016 Jun;534(7605):102–5.
65. Mateo L, Ullastres A, González J. A Transposable Element Insertion Confers Xenobiotic Resistance in Drosophila. Feschotte C, editor. PLoS Genet. 2014 Aug 14;10(8):e1004560.
66. Dekker T, Ibba I, Siju KP, Stensmyr MC, Hansson BS. Olfactory Shifts Parallel Superspecialism for Toxic Fruit in Drosophila melanogaster Sibling, D. sechellia. Current Biology. 2006 Jan;16(1):101–9.
67. Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y. Odorant-Binding Proteins OBP57d and OBP57e Affect Taste Perception and Host-Plant Preference in Drosophila sechellia. Noor MAF, editor. PLoS Biol. 2007 Apr 24;5(5):e118.
68. McKenzie SK, Kronauer DJC. The genomic architecture and molecular evolution of ant odorant receptors. Genome Res. 2018 Jan 11;28(11):1757–65.
69. Cultrone A, Domínguez YR, Drevet C, Scazzocchio C, Fernández-Martín R. The tightly

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

regulated promoter of the xanA gene of Aspergillus nidulans is included in a helitron: A fungal helitron includes a regulated promoter. Molecular Microbiology. 2007 Jan 15;63(6):1577–87.
70. Thomas J, Phillips CD, Baker RJ, Pritham EJ. Rolling-Circle Transposons Catalyze Genomic Innovation in a Mammalian Lineage. Genome Biology and Evolution. 2014 Oct;6(10):2595–610.
71. Gupta S, Gallavotti A, Stryker GA, Schmidt RJ, Lal SK. A novel class of Helitron- related transposable elements in maize contain portions of multiple pseudogenes. Plant Mol Biol. 2005 Jan;57(1):115–27.
72. Huang K, Li CF, Wu J, Wei JH, Zou Y, Han MJ, et al. Enhancer activity of *Helitron* in *sericin-1* gene promoter from *Bombyx mori*: Enhancer activity of Helitrons in sericin-1. Insect Science. 2016 Jun;23(3):396–405.
73. Batista RA, Moreno-Romero J, Qiu Y, van Boven J, Santos-González J, Figueiredo DD, et al. The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. Zilberman D, Hardtke CS, editors. eLife. 2019 Dec 2;8:e50541.
74. Carareto CMA, Hernandez EH, Vieira C. Genomic regions harboring insecticide resistance-associated Cyp genes are enriched by transposable element fragments carrying putative transcription factor binding sites in two sibling Drosophila species. Gene. 2014 Mar;537(1):93–9.
75. Merenciano M, Aguilera L, González J. Two-step CRISPR-Cas9 protocol for transposable element deletion in *D. melanogaster* natural populations. STAR Protocols. 2023 Sep 15;4(3):102501.
76. Wright CR, William S. Characterization of Volatiles of Necrotic Stenocereus thurberi and Opuntia littoralis and Toxicity and Olfactory Preference of Drosophila melanogster, D. mojavensis wrigleyi, and D. mojavensis sonorensis to Necrotic Cactus Volatiles. Natural Product Communications. 2014;9(6):1934578X1400900833.
77. Hickner PV, Rivaldi CL, Johnson CM, Siddappaji M, Raster GJ, Syed Z. The making of a pest: Insights from the evolution of chemosensory receptor families in a pestiferous and invasive fly, Drosophila suzukii. BMC Genomics. 2016 Dec;17(1):648.
78. Mansourian S, Stensmyr MC. The chemical ecology of the fly. Current Opinion in Neurobiology. 2015 Oct;34:95–102.
79. Crowley-Gall A, Date P, Han C, Rhodes N, Andolfatto P, Layne JE, et al. Population differences in olfaction accompany host shift in *Drosophila mojavensis*. Proc R Soc B. 2016 Aug 31;283(1837):20161562.
80. De Panis D, Dopazo H, Bongcam-Rudloff E, Conesa A, Hasson E. Transcriptional responses are oriented towards different components of the rearing environment in two Drosophila sibling species. BMC Genomics. 2022 Dec;23(1):515.
81. Ruiz A, Heed WB. Host-Plant Specificity in the Cactophilic Drosophila mulleri Species Complex. The Journal of Animal Ecology. 1988 Feb;57(1):237.
82. Mateus RP, Nazario-Yepiz NO, Ibarra-Laclette E, Ramirez Loustalot-Laclette M, Markow TA. Developmental and Transcriptomal Responses to Seasonal Dietary Shifts in the Cactophilic Drosophila mojavensis of North America. Journal of Heredity. 2019 Jan 7;110(1):58–67.
83. Matzkin LM. Population transcriptomics of cactus host shifts in *Drosophila mojavensis*: Population Transcriptomics of *D. mojavensis*. Molecular Ecology. 2012 May;21(10):2428–39.
84. Feyereisen R. Insect CYP Genes and P450 Enzymes. In: Insect Molecular Biology and Biochemistry [Internet]. Elsevier; 2012 [cited 2022 Sep 28]. p. 236–316. Available from: https://linkinghub.elsevier.com/retrieve/pii/B978012384747810008X
85. Tijet N, Helvig C, Feyereisen R. The cytochrome P450 gene superfamily in Drosophila melanogaster : Annotation, intron-exon organization and phylogeny. Gene. 2001 Jan;262(1–2):189–98.
86. Coronado-Zamora M, González J. Transposons contribute to the functional diversification of the head, gut, and ovary transcriptomes across Drosophila natural strains. Genome Res. 2023 Jan

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

9;33(9):1541–53.
87. Bartolomé C, Maside X, Charlesworth B. On the Abundance and Distribution of Transposable Elements in the Genome of Drosophila melanogaster. Molecular Biology and Evolution. 2002 Jun 1;19(6):926–37.
88. Gao C, Zhou G, Ma C, Zhai W, Zhang T, Liu Z, et al. Helitron-like transposons contributed to the mating system transition from out-crossing to self-fertilizing in polyploid Brassica napus L. Sci Rep. 2016 Sep 21;6(1):33785.
89. Garrigues JM, Tsu BV, Daugherty MD, Pasquinelli AE. Diversification of the Caenorhabditis heat shock response by Helitron transposable elements. Landry CR, Wittkopp PJ, editors. eLife. 2019 Dec 11;8:e51139.
90. Guo C, Spinelli M, Liu M, Li QQ, Liang C. A Genome-wide Study of "Non-3UTR" Polyadenylation Sites in Arabidopsis thaliana. Sci Rep. 2016 Jun 15;6(1):28060.
91. Lee JY, Ji Z, Tian B. Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3′-end of genes. Nucleic Acids Research. 2008 Oct 1;36(17):5581–90.
92. Fonseca PM, Moura RD, Wallau GL, Loreto ELS. The mobilome of Drosophila incompta, a flower-breeding species: comparison of transposable element landscapes among generalist and specialist flies. Chromosome Res. 2019 Sep 1;27(3):203–19.
93. Mérel V, Gibert P, Buch I, Rodriguez Rada V, Estoup A, Gautier M, et al. The Worldwide Invasion of *Drosophila suzukii* Is Accompanied by a Large Increase of Transposable Element Load and a Small Number of Putatively Adaptive Insertions. Arkhipova I, editor. Molecular Biology and Evolution. 2021 Sep 27;38(10):4252–67.
94. Marin P, Jaquet A, Picarle J, Fablet M, Merel V, Delignette-Muller ML, et al. Phenotypic and Transcriptomic Responses to Stress Differ According to Population Geography in an Invasive Species. Genome Biology and Evolution. 2021 Sep 1;13(9):evab208.
95. Harigaya Y, Parker R. No-go decay: a quality control mechanism for RNA in translation. WIREs RNA. 2010 Jul;1(1):132–41.
96. Vasudevan S, Peltz SW, Wilusz CJ. Non-stop decay—a new mRNA surveillance pathway. BioEssays. 2002;24(9):785–8.
97. Hug N, Longman D, Cáceres JF. Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Res. 2016 Feb 29;44(4):1483–95.
98. Beckerson WC, Rodríguez de la Vega RC, Hartmann FE, Duhamel M, Giraud T, Perlin MH. Cause and Effectors: Whole-Genome Comparisons Reveal Shared but Rapidly Evolving Effector Sets among Host-Specific Plant-Castrating Fungi. mBio. 2019 Nov 5;10(6):10.1128/mbio.02391-19.
99. Whisson SC, Vetukuri RR, Avrova AO, Dixelius C. Can silencing of transposons contribute to variation in effector gene expression in Phytophthora infestans? Mobile Genetic Elements. 2012 Mar 1;2(2):110–4.
100. Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, et al. Genome Evolution Following Host Jumps in the Irish Potato Famine Pathogen Lineage. Science. 2010 Dec 10;330(6010):1540–3.
101. Zouros E, d'Entremont CJ. Sexual Isolation Among Populations of Drosophila mojavensis: Response to Pressure from a Related Species. Evolution. 1980;34(3):421–30.
102. Markow TA. Sexual Isolation among Populations of Drosophila mojavensis. Evolution. 1991;45(6):1525–9.
103. Howard RW, Blomquist GJ. ECOLOGICAL, BEHAVIORAL, AND BIOCHEMICAL ASPECTS OF INSECT HYDROCARBONS*. Annual Review of Entomology. 2005 Jan 1;50(Volume 50, 2005):371–93.
104. Stennett MD, Etges WJ. Premating Isolation Is Determined by Larval Rearing Substrates in

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

CactophilicDrosophila mojavensis. III. Epicuticular Hydrocarbon Variation Is Determined by Use of Different Host Plants inDrosophila mojavensis andDrosophila arizonae. J Chem Ecol. 1997 Dec 1;23(12):2803–24.
105. Havens JA, Etges WJ. Premating isolation is determined by larval rearing substrates in cactophilic Drosophila mojavensis. IX. Host plant and population specific epicuticular hydrocarbon expression influences mate choice and sexual selection. Journal of Evolutionary Biology. 2013 Mar 1;26(3):562–76.
106. Romero-Soriano V, Burlet N, Vela D, Fontdevila A, Vieira C, García Guerreiro MP. *Drosophila* Females Undergo Genome Expansion after Interspecific Hybridization. Genome Biol Evol. 2016 Mar;8(3):556–61.
107. Bodelón A, Fablet M, Veber P, Vieira C, García Guerreiro MP. High Stability of the Epigenome in *Drosophila* Interspecific Hybrids. Betancourt A, editor. Genome Biology and Evolution. 2022 Feb 4;14(2):evac024.
108. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019 May;37(5):540–6.
109. Li H. Minimap2: pairwise alignment for nucleotide sequences. Birol I, editor. Bioinformatics. 2018 Sep 15;34(18):3094–100.
110. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bio] [Internet]. 2013 May 26 [cited 2021 Nov 5]; Available from: http://arxiv.org/abs/1303.3997
111. Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 2018 Jun 4;6:e4958.
112. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.
113. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics. 2007 Apr 15;23(8):1026–8.
114. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019 Dec;20(1):224.
115. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015 Oct 1;31(19):3210–2.
116. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Valencia A, editor. Bioinformatics. 2021 Jul 19;37(12):1639–43.
117. Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, et al. Highly contiguous assemblies of 101 drosophilid genomes. eLife. 2021;10(e66405):33.
118. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, et al. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res [Internet]. 2024 Jun 12 [cited 2024 Jul 15]; Available from: https://genome.cshlp.org/content/early/2024/05/28/gr.278090.123
119. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019 Dec;20(1):238.
120. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015 Jan;12(1):59–60.
121. Altschul S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 1997 Sep 1;25(17):3389–402.
122. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Valencia A, editor. Bioinformatics. 2021 Jul 19;37(12):1639–43.
123. Vanderlinde T, Dupim EG, Nazario-Yepiz NO, Carvalho AB. An Improved Genome

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Assembly for Drosophila navojoa, the Basal Species in the mojavensis Cluster. Journal of Heredity. 2019 Jan 7;110(1):118–23.
124. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Berger B, editor. Bioinformatics. 2020 Apr 1;36(7):2253–5.
125. Banho CA, Mérel V, Oliveira TYK, Carareto CMA, Vieira C. Comparative transcriptomics between Drosophila mojavensis and D. arizonae reveals transgressive gene expression and underexpression of spermatogenesis-related genes in hybrid testes. Sci Rep. 2021 May 10;11(1):9844.
126. Moreyra NN, Almeida FC, Allan C, Frankel N, Matzkin LM, Hasson E. Phylogenomics provides insights into the evolution of cactophily and host plant shifts in Drosophila. Molecular Phylogenetics and Evolution. 2023 Jan 1;178:107653.
127. Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. Zenodo [Internet]. 2023 [cited 2023 Nov 9]; Available from: https://zenodo.org/records/8178877
128. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.
129. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research. 2006 Jul 1;34(suppl_2):W609–12.
130. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution. 2020 May 1;37(5):1530–4.
131. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution. 2007 Apr 18;24(8):1586–91.
132. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, et al. HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Molecular Biology and Evolution. 2020 Jan 1;37(1):295–9.
133. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.
134. Baril TJ, Galbraith JD, Hayward A. Earl Grey: a fully automated user-friendly transposable element annotation and analysis pipeline [Internet]. bioRxiv; 2024 [cited 2024 Feb 20]. p. 2022.06.30.498289. Available from: https://www.biorxiv.org/content/10.1101/2022.06.30.498289v3
135. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020 Apr 28;117(17):9451–7.
136. Bao Z, Eddy SR. Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. Genome Res. 2002 Aug 1;12(8):1269–76.
137. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005 Jun 1;21(Suppl 1):i351–8.
138. Platt RN, Blanco-Berdugo L, Ray DA. Accurate Transposable Element Annotation Is Vital When Analyzing New Genome Assemblies. Genome Biol Evol. 2016 Feb;8(2):403–10.
139. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution. 2013 Apr 1;30(4):772–80.
140. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009 Aug 1;25(15):1972–3.
141. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics. 16(6):276–7.
142. SMIT, Arian FA. RepeatMasker Open 3.0. 2004;
143. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, et al. The Dfam database of repetitive DNA families. Nucleic Acids Res. 2016 Jan 4;44(D1):D81–9.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.27.587021; this version posted October 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

144. Wong WY, Simakov O. RepeatCraft: a meta-pipeline for repetitive element de-fragmentation and annotation. Hancock J, editor. Bioinformatics. 2019 Mar 15;35(6):1051–2.
145. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007 Dec;8(12):973–82.
146. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research. 1999 Jan 1;27(2):573–80.
147. Flutre T, Permal E, Quesneville H. Transposable Element Annotation in Completely Sequenced Eukaryote Genomes. In: Grandbastien MA, Casacuberta JM, editors. Plant Transposable Elements [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012 [cited 2023 Feb 27]. p. 17–39. (Topics in Current Genetics; vol. 24). Available from: https://link.springer.com/10.1007/978-3-642-31842-9_2
148. Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics. 2015 Sep 30;btv562.
149. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.
150. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011 Apr 1;27(7):1017–8.
151. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Research. 2015 Jul 1;43(W1):W39–49.
152. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Research. 2019 Nov 8;gkz1001.
153. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan;29(1):15–21.
154. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015 Jan 15;31(2):166–9.
155. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

References

Akkouche, Abdou, Rita Rebollo, Nelly Burlet, et al. 2012. 'Tirant, a Newly Discovered Active Endogenous Retrovirus in Drosophila Simulans'. *Journal of Virology* 86 (7): 3675–81. https://doi.org/10.1128/jvi.07146-11.

Alhosin, Mahmoud. 2023. 'Epigenetics Mechanisms of Honeybees: Secrets of Royal Jelly'. *Epigenetics Insights* 16 (January): 25168657231213717. https://doi.org/10.1177/25168657231213717.

Aliaga, Benoît, Ingo Bulla, Gabriel Mouahid, David Duval, and Christoph Grunau. 2019. 'Universality of the DNA Methylation Codes in Eucaryotes'. *Scientific Reports* 9 (1): 173. https://doi.org/10.1038/s41598-018-37407-8.

Alonge, Michael, Ludivine Lebeigle, Melanie Kirsche, et al. 2022. 'Automated Assembly Scaffolding Using RagTag Elevates a New Tomato System for High-Throughput Genome Editing'. *Genome Biology* 23 (1): 258. https://doi.org/10.1186/s13059-022-02823-7.

Alruiz, José M., Ignacio Peralta-Maraver, Francisco Bozinovic, Mauro Santos, and Enrico L. Rezende. 2023. 'Temperature Adaptation and Its Impact on the Shape of Performance Curves in Drosophila Populations'. *Proceedings of the Royal Society B: Biological Sciences* 290 (1998): 20230507. https://doi.org/10.1098/rspb.2023.0507.

Ameri, Abigail J, and Zachary A Lewis. 2021. 'Shannon Entropy as a Metric for Conditional Gene Expression in Neurospora Crassa'. *G3 Genes|Genomes|Genetics* 11 (4): jkab055. https://doi.org/10.1093/g3journal/jkab055.

Andrade, Pedro. 2025. 'Highlight: "Jumping Genes" Caught in an Aging Merry-Go-Round'. *Genome Biology and Evolution* 17 (6): evaf094. https://doi.org/10.1093/gbe/evaf094.

Aranda, Ana, and Angel Pascual. 2001. 'Nuclear Hormone Receptors and Gene Expression'. *Physiological Reviews* 81 (3): 1269–304. https://doi.org/10.1152/physrev.2001.81.3.1269.

Aravamudan, Bharathi, Tim Fergestad, Warren S. Davis, Chris K. Rodesch, and Kendal Broadie. 1999. 'Drosophila Unc-13 Is Essential for Synaptic Transmission'. *Nature Neuroscience* 2 (11): 965–71. https://doi.org/10.1038/14764.

Arkhipova, Irina R. 2006. 'Distribution and Phylogeny of Penelope-Like Elements in Eukaryotes'. *Systematic Biology* 55 (6): 875–85. https://doi.org/10.1080/10635150601077683.

Arkhipova, Irina R, and Irina A Yushenova. 2019. 'Giant Transposons in Eukaryotes: Is Bigger Better?' *Genome Biology and Evolution* 11 (3): 906–18. https://doi.org/10.1093/gbe/evz041.

Arking, Robert, Allan G. Force, Steven P. Dudas, Steven Buck, and George T. Baker. 1996. 'Factors Contributing to the Plasticity of the Extended Longevity Phenotypes of Drosophila'. *Experimental Gerontology* 31 (6): 623–43. https://doi.org/10.1016/S0531-5565(96)00096-4.

Arnone, Maria I., and Eric H. Davidson. 1997. 'The Hardwiring of Development: Organization and Function of Genomic Regulatory Systems'. *Development* 124 (10): 1851–64. https://doi.org/10.1242/dev.124.10.1851.

Asahina, Kenta. 2018. 'Sex Differences in *Drosophila* Behavior: Qualitative and Quantitative Dimorphism'. *Current Opinion in Physiology*, Sex Differences, vol. 6 (December): 35–45. https://doi.org/10.1016/j.cophys.2018.04.004.

Asselin, Angelique, and Karyn Johnson. 2025. 'The Infectivity of Virus Particles from Wolbachia-Infected Drosophila'. *BMC Microbiology* 25 (1): 25. https://doi.org/10.1186/s12866-024-03722-6.

Azad, Priti, Julie Ryu, and Gabriel G. Haddad. 2011. 'Distinct Role of Hsp70 in *Drosophila* Hemocytes during Severe Hypoxia'. *Free Radical Biology and Medicine* 51 (2): 530–38. https://doi.org/10.1016/j.freeradbiomed.2011.05.005.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Bakker, K. 1959. 'Feeding Period, Growth, and Pupation in Larvae Ofdrosophila Melanogaster'. *Entomologia Experimentalis et Applicata* 2 (3): 171–86. https://doi.org/10.1111/j.1570-7458.1959.tb00432.x.

Ballard, J. W. O., and R. G. Melvin. 2007. 'Tetracycline Treatment Influences Mitochondrial Metabolism and mtDNA Density Two Generations after Treatment in Drosophila'. *Insect Molecular Biology* 16 (6): 799–802. https://doi.org/10.1111/j.1365-2583.2007.00760.x.

Barckmann, Bridlin, Marianne El-Barouk, Alain Pélisson, et al. 2018. 'The Somatic piRNA Pathway Controls Germline Transposition over Generations'. *Nucleic Acids Research* 46 (18): 9524–36. https://doi.org/10.1093/nar/gky761.

Barrón, Maite G., Anna-Sophie Fiston-Lavier, Dmitri A. Petrov, and Josefa González. 2014. 'Population Genomics of Transposable Elements in Drosophila'. *Annual Review of Genetics* 48 (Volume 48, 2014): 561–81. https://doi.org/10.1146/annurev-genet-120213-092359.

Barro-Trastoy, Daniela, and Claudia Köhler. 2024. 'Helitrons: Genomic Parasites That Generate Developmental Novelties'. *Trends in Genetics*, February, S0168952524000295. https://doi.org/10.1016/j.tig.2024.02.002.

Bartel, David P. 2018. 'Metazoan MicroRNAs'. *Cell* 173 (1): 20–51. https://doi.org/10.1016/j.cell.2018.03.006.

Barter, Basil, and Jungnam Cho. 2025. 'RNA Methylation in Retrotransposon Control'. *Trends in Genetics* 41 (7): 556–58. https://doi.org/10.1016/j.tig.2025.04.013.

Bassett, J. H. Duncan, Clare B. Harvey, and Graham R. Williams. 2003. 'Mechanisms of Thyroid Hormone Receptor-Specific Nuclear and Extra Nuclear Actions'. *Molecular and Cellular Endocrinology*, Proceedings of the Conference entitled Thyroid and Gonadal Hormones in Health and Disease., vol. 213 (1): 1–11. https://doi.org/10.1016/j.mce.2003.10.033.

Bates, Douglas, Martin Maechler, Ben Bolker [aut, et al. 2025. *Lme4: Linear Mixed-Effects Models Using 'Eigen' and S4*. V. 1.1-36. Released January 11. https://cran.r-project.org/web/packages/lme4/index.html.

Bauerfeind, Stephanie S., Vanessa Kellermann, Neda N. Moghadam, Volker Loeschcke, and Klaus Fischer. 2014. 'Temperature and Photoperiod Affect Stress Resistance Traits in Drosophila Melanogaster'. *Physiological Entomology* 39 (3): 237–46. https://doi.org/10.1111/phen.12068.

Baumann, Paul, Nancy A. Moran, and Linda C. Baumann. 2013. 'Bacteriocyte-Associated Endosymbionts of Insects'. In *The Prokaryotes*, edited by Eugene Rosenberg, Edward F. DeLong, Stephen Lory, Erko Stackebrandt, and Fabiano Thompson. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30194-0_19.

Baythavong, Brooke S. 2011. 'Linking the Spatial Scale of Environmental Variation and the Evolution of Phenotypic Plasticity: Selection Favors Adaptive Plasticity in Fine-Grained Environments.' *The American Naturalist* 178 (1): 75–87. https://doi.org/10.1086/660281.

Behrman, E. L., S. S. Watson, K. R. O'Brien, M. S. Heschel, and P. S. Schmidt. 2015. 'Seasonal Variation in Life History Traits in Two Drosophila Species'. *Journal of Evolutionary Biology* 28 (9): 1691–704. https://doi.org/10.1111/jeb.12690.

Ben Shachar, Mattan, Daniel Lüdecke, and Dominique Makowski. 2020. 'Effectsize: Estimation of Effect Size Indices and Standardized Parameters Aims of the Package'. *The Journal of Open Source Software* 5 (December): 2815. https://doi.org/10.21105/joss.02815.

Bennett, E Andrew, Laura E Coleman, Circe Tsui, W Stephen Pittard, and Scott E Devine. 2004. 'Natural Genetic Variation Caused by Transposable Elements in Humans'. *Genetics* 168 (2): 933–51. https://doi.org/10.1534/genetics.104.031757.

Bergman, Aviv, and Mark L. Siegal. 2003. 'Evolutionary Capacitance as a General Feature of Complex Gene Networks'. *Nature* 424 (6948): 549–52. https://doi.org/10.1038/nature01765.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Berrigan, D., and L. Partridge. 1997. 'Influence of Temperature and Activity on the Metabolic Rate of Adult *Drosophila Melanogaster*'. *Comparative Biochemistry and Physiology Part A: Physiology* 118 (4): 1301–7. https://doi.org/10.1016/S0300-9629(97)00030-3.

Berrigan, David. 1997. 'Acclimation of Metabolic Rate in Response to Developmental Temperature in Drosophila Melanogaster'. *Journal of Thermal Biology* 22 (3): 213–18. https://doi.org/10.1016/S0306-4565(97)00015-6.

Berthelier, Jérémy, Leonardo Furci, Shuta Asai, et al. 2023. 'Long-Read Direct RNA Sequencing Reveals Epigenetic Regulation of Chimeric Gene-Transposon Transcripts in Arabidopsis Thaliana'. *Nature Communications* 14 (1): 3248. https://doi.org/10.1038/s41467-023-38954-z.

Bewick, Adam J., Kevin J. Vogel, Allen J. Moore, and Robert J. Schmitz. 2017. 'Evolution of DNA Methylation across Insects'. *Molecular Biology and Evolution* 34 (3): 654–65. https://doi.org/10.1093/molbev/msw264.

Bi, Jie, and Yu-Feng Wang. 2020. 'The Effect of the Endosymbiont Wolbachia on the Behavior of Insect Hosts'. *Insect Science* 27 (5): 846–58. https://doi.org/10.1111/1744-7917.12731.

Biémont, C., F. Lemeunier, M. P. Garcia Guerreiro, et al. 1994. 'Population Dynamics of the Copia, Mdg1, Mdg3, Gypsy, and P Transposable Elements in a Natural Population of Drosophila Melanogaster'. *Genetics Research* 63 (3): 197–212. https://doi.org/10.1017/S0016672300032353.

Biémont, Christian. 2010a. 'A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution'. *Genetics* 186 (4): 1085–93. https://doi.org/10.1534/genetics.110.124180.

Biémont, Christian. 2010b. 'A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution'. *Genetics* 186 (4): 1085–93. https://doi.org/10.1534/genetics.110.124180.

Biémont, Christian, and Cristina Vieira. 2006. 'Junk DNA as an Evolutionary Force'. *Nature* 443 (7111): 7111. https://doi.org/10.1038/443521a.

Bobrovskikh, Margarita A., and Nataly E. Gruntenko. 2023. 'Mechanisms of Neuroendocrine Stress Response in Drosophila and Its Effect on Carbohydrate and Lipid Metabolism'. *Insects* 14 (5): 5. https://doi.org/10.3390/insects14050474.

Böhme, Mathias A., Christina Beis, Suneel Reddy-Alla, et al. 2016. 'Active Zone Scaffolds Differentially Accumulate Unc13 Isoforms to Tune Ca2+ Channel–Vesicle Coupling'. *Nature Neuroscience* 19 (10): 1311–20. https://doi.org/10.1038/nn.4364.

Bolar, Kartikeya. 2019. *STAT: Interactive Document for Working with Basic Statistical Analysis*. V. 0.1.0. Released April 1. https://cran.r-project.org/web/packages/STAT/index.html.

Bordenstein, Seth R., and Kevin R. Theis. 2015. 'Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes'. *PLOS Biology* 13 (8): e1002226. https://doi.org/10.1371/journal.pbio.1002226.

Boros, Imre M. 2012. 'Histone Modification in Drosophila'. *Briefings in Functional Genomics* 11 (4): 319–31. https://doi.org/10.1093/bfgp/els029.

Bourque, Guillaume, Kathleen H. Burns, Mary Gehring, et al. 2018. 'Ten Things You Should Know about Transposable Elements'. *Genome Biology* 19 (1): 199. https://doi.org/10.1186/s13059-018-1577-z.

Bowman, Jc. 1972. 'Genotype × Environment Interactions'. *Genetics Selection Evolution* 4 (1): 117. https://doi.org/10.1186/1297-9686-4-1-117.

Bradshaw, A. D. 1965. 'Evolutionary Significance of Phenotypic Plasticity in Plants'. In *Advances in Genetics*, edited by E. W. Caspari and J. M. Thoday, vol. 13. Academic Press. https://doi.org/10.1016/S0065-2660(08)60048-6.

Brasset, E., A. R. Taddei, F. Arnaud, et al. 2006. 'Viral Particles of the Endogenous Retrovirus ZAM from Drosophila Melanogaster Use a Pre-Existing Endosome/Exosome Pathway for Transfer to the

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Oocyte'. *Retrovirology* 3 (May): 25. https://doi.org/10.1186/1742-4690-3-25.

Brennecke, Julius, Alexei A. Aravin, Alexander Stark, et al. 2007. 'Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila'. *Cell* 128 (6): 1089–103. https://doi.org/10.1016/j.cell.2007.01.043.

Brennecke, Julius, Colin D. Malone, Alexei A. Aravin, Ravi Sachidanandam, Alexander Stark, and Gregory J. Hannon. 2008. 'An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing'. *Science* 322 (5906): 1387–92. https://doi.org/10.1126/science.1165171.

Bright, Monika, and Silvia Bulgheresi. 2010. 'A Complex Journey: Transmission of Microbial Symbionts'. *Nature Reviews Microbiology* 8 (3): 218–30. https://doi.org/10.1038/nrmicro2262.

Brockhurst, Michael A., Ellie Harrison, James P. J. Hall, Thomas Richards, Alan McNally, and Craig MacLean. 2019. 'The Ecology and Evolution of Pangenomes'. *Current Biology* 29 (20): R1094–103. https://doi.org/10.1016/j.cub.2019.08.012.

Broecker, Felix, and Karin Moelling. 2019. 'Evolution of Immune Systems From Viruses and Transposable Elements'. *Frontiers in Microbiology* 10 (January). https://doi.org/10.3389/fmicb.2019.00051.

Bronstein, Judith L. 2015. *Mutualism*. Oxford University Press.

Brookfield, John F. Y. 2004. 'Evolutionary Genetics: Mobile DNAs as Sources of Adaptive Change?' *Current Biology* 14 (9): R344–45. https://doi.org/10.1016/j.cub.2004.04.021.

Bruner-Montero, Gaspar, and Francis M. Jiggins. 2023. 'Wolbachia Protects Drosophila Melanogaster against Two Naturally Occurring and Virulent Viral Pathogens'. *Scientific Reports* 13 (1): 1. https://doi.org/10.1038/s41598-023-35726-z.

Bryant, Kaeli N., and Irene L. G. Newton. 2020. 'The Intracellular Symbiont Wolbachia Pipientis Enhances Recombination in a Dose-Dependent Manner'. *Insects* 11 (5): 5. https://doi.org/10.3390/insects11050284.

Burgess, Scott C., and Dustin J. Marshall. 2014. 'Adaptive Parental Effects: The Importance of Estimating Environmental Predictability and Offspring Fitness Appropriately'. *Oikos* 123 (7): 769–76. https://doi.org/10.1111/oik.01235.

Burns, Kathleen H. 2020. 'Our Conflict with Transposable Elements and Its Implications for Human Disease'. *Annual Review of Pathology: Mechanisms of Disease* 15 (Volume 15, 2020): 51–70. https://doi.org/10.1146/annurev-pathmechdis-012419-032633.

Cabral-de-Mello, Diogo C., and Octavio M. Palacios-Gimenez. 2024. 'Repetitive DNAs: The "Invisible" Regulators of Insect Adaptation and Speciation'. *Current Opinion in Insect Science*, November, 101295. https://doi.org/10.1016/j.cois.2024.101295.

Cai, Yimei, Xiaomin Yu, Songnian Hu, and Jun Yu. 2009. 'A Brief Review on the Mechanisms of miRNA Regulation'. *Genomics, Proteomics & Bioinformatics* 7 (4): 147–54. https://doi.org/10.1016/S1672-0229(08)60044-3.

Cappello, J. 1985. 'Sequence of Dictyostelium DIRS-1: An Apparent Retrotransposon with Inverted Terminal Repeats and an Internal Circle Junction Sequence'. *Cell* 43 (1): 105–15. https://doi.org/10.1016/0092-8674(85)90016-9.

Cappucci, Ugo, Fabrizia Noro, Assunta Maria Casale, et al. 2019. 'The Hsp70 Chaperone Is a Major Player in Stress-Induced Transposable Element Activation'. *Proceedings of the National Academy of Sciences* 116 (36): 17943–50. https://doi.org/10.1073/pnas.1903936116.

Capy, Pierre. 2021. 'Taming, Domestication and Exaptation: Trajectories of Transposable Elements in Genomes'. *Cells* 10 (12): 3590. https://doi.org/10.3390/cells10123590.

Capy, Pierre, Giuliano Gasperi, Christian Biémont, and Claude Bazin. 2000. 'Stress and Transposable Elements: Co-Evolution or Useful Parasites?' *Heredity* 85 (2): 101–6. https://doi.org/10.1046/j.1365-2540.2000.00751.x.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Carbone, Mary Anna, Katherine W. Jordan, Richard F. Lyman, et al. 2006. 'Phenotypic Variation and Natural Selection at Catsup, a Pleiotropic Quantitative Trait Gene in Drosophila'. *Current Biology* 16 (9): 912–19. https://doi.org/10.1016/j.cub.2006.03.051.

Carey, Matthew R., C. Ruth Archer, James Rapkin, et al. 2022. 'Mapping Sex Differences in the Effects of Protein and Carbohydrates on Lifespan and Reproduction in Drosophila Melanogaster: Is Measuring Nutrient Intake Essential?' *Biogerontology* 23 (1): 129–44. https://doi.org/10.1007/s10522-022-09953-2.

Carroll, Sean B., Jennifer K. Grenier, and Scott D. Weatherbee. 2005. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. 2nd ed. Blackwell publ.

Casacuberta, Elena, and Josefa González. 2013. 'The Impact of Transposable Elements in Environmental Adaptation'. *Molecular Ecology* 22 (6): 1503–17. https://doi.org/10.1111/mec.12170.

Casadesús, Josep, and María A. Sánchez-Romero. 2022. 'DNA Methylation in Prokaryotes'. In *DNA Methyltransferases - Role and Function*, edited by Albert Jeltsch and Renata Z. Jurkowska. Springer International Publishing. https://doi.org/10.1007/978-3-031-11454-0_2.

Catlin, Nathan S., and Emily B. Josephs. 2022a. 'The Important Contribution of Transposable Elements to Phenotypic Variation and Evolution'. *Current Opinion in Plant Biology* 65 (February): 102140. https://doi.org/10.1016/j.pbi.2021.102140.

Catlin, Nathan S., and Emily B. Josephs. 2022b. 'The Important Contribution of Transposable Elements to Phenotypic Variation and Evolution'. *Current Opinion in Plant Biology* 65 (February): 102140. https://doi.org/10.1016/j.pbi.2021.102140.

Cech, T. R. 2012. 'The RNA Worlds in Context'. *Cold Spring Harbor Perspectives in Biology* 4 (7): a006742–a006742. https://doi.org/10.1101/cshperspect.a006742.

Chacón, Jeremy M., Sarah P. Hammarlund, Jonathan N. V. Martinson, Leno B. Smith, and William R. Harcombe. 2021. 'The Ecology and Evolution of Model Microbial Mutualisms'. *Annual Review of Ecology, Evolution, and Systematics* 52 (Volume 52, 2021): 363–84. https://doi.org/10.1146/annurev-ecolsys-012121-091753.

Chalvet, Fabienne, Laure Teysset, Christophe Terzian, et al. 1999. 'Proviral Amplification of the Gypsy Endogenous Retrovirus of Drosophila Melanogaster Involves Env-independent Invasion of the Female Germline'. *The EMBO Journal* 18 (9): 2659–69. https://doi.org/10.1093/emboj/18.9.2659.

Chandler, Michael, Martin Gellert, Alan M. Lambowitz, Phoebe A. Rice, and Suzanne B. Sandmeyer. 2020. *Mobile DNA III*. John Wiley & Sons.

Chatterjee, Rabindra Nath. 2024. 'The Role of Epigenetics in Phenotypic Evolution: Lessons from the Genomic Response to Stresses, Genetic Conflicts and Developmental Plasticity in Drosophila'. *The Nucleus* 67 (3): 675–88. https://doi.org/10.1007/s13237-024-00520-3.

Cheeseman, Kevin, Jeanne Ropars, Pierre Renault, et al. 2014. 'Multiple Recent Horizontal Transfers of a Large Genomic Region in Cheese Making Fungi'. *Nature Communications* 5 (1): 2876. https://doi.org/10.1038/ncomms3876.

Chen, Bing, Martin E. Feder, and Le Kang. 2018a. 'Evolution of Heat-Shock Protein Expression Underlying Adaptive Responses to Environmental Stress'. *Molecular Ecology* 27 (15): 3040–54. https://doi.org/10.1111/mec.14769.

Chen, Bing, Martin E. Feder, and Le Kang. 2018b. 'Evolution of Heat-Shock Protein Expression Underlying Adaptive Responses to Environmental Stress'. *Molecular Ecology* 27 (15): 3040–54. https://doi.org/10.1111/mec.14769.

Chen, Bing, and Andreas Wagner. 2012. 'Hsp90 Is Important for Fecundity, Longevity, and Buffering of Cryptic Deleterious Variation in Wild Fly Populations'. *BMC Evolutionary Biology* 12 (1): 25. https://doi.org/10.1186/1471-2148-12-25.

Chen, Jie, Yunjie Ma, Ziying Guan, et al. 2023. 'Labor Division of Worker Ants Can Be Controlled by Insulin

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Synthesis Targeted through miR-279c-5p in Solenopsis Invicta (Hymenoptera: Formicidae)'. *Pest Management Science* 79 (12): 5029–43. https://doi.org/10.1002/ps.7704.

Chen, Wenqiong J., and Tong Zhu. 2004. 'Networks of Transcription Factors with Roles in Environmental Stress Response'. *Trends in Plant Science* 9 (12): 591–96. https://doi.org/10.1016/j.tplants.2004.10.007.

Choi, Jae Young, and Yuh Chwen G. Lee. 2020. 'Double-Edged Sword: The Evolutionary Consequences of the Epigenetic Silencing of Transposable Elements'. *PLOS Genetics* 16 (7): e1008872. https://doi.org/10.1371/journal.pgen.1008872.

Choi, Kwang-Wook. 2024. 'Piwi and piRNA in Germline and Epigenetic Regulation'. In *Genetic Studies in Model Organisms: From Classical to Modern Genetics*, edited by Kwang-Wook Choi. Springer Nature. https://doi.org/10.1007/978-981-97-0830-7_13.

Christensen, Rune Haubo B. (2018) 2025. *Runehaubo/lmerTestR*. HTML. January 17, released January 14. https://github.com/runehaubo/lmerTestR.

Chrostek, Ewa, Marta S. P. Marialva, Sara S. Esteves, et al. 2013. 'Wolbachia Variants Induce Differential Protection to Viruses in Drosophila Melanogaster: A Phenotypic and Phylogenomic Analysis'. *PLOS Genetics* 9 (12): e1003896. https://doi.org/10.1371/journal.pgen.1003896.

Chung, Henry, Michael R Bogwitz, Caroline McCart, et al. 2007. 'Cis-Regulatory Elements in the Accord Retrotransposon Result in Tissue-Specific Expression of the Drosophila Melanogaster Insecticide Resistance Gene Cyp6g1'. *Genetics* 175 (3): 1071–77. https://doi.org/10.1534/genetics.106.066597.

Chuong, Edward B., Nels C. Elde, and Cédric Feschotte. 2016. 'Regulatory Evolution of Innate Immunity through Co-Option of Endogenous Retroviruses'. *Science* 351 (6277): 1083–87. https://doi.org/10.1126/science.aad5497.

Chuong, Edward B., Nels C. Elde, and Cédric Feschotte. 2017. 'Regulatory Activities of Transposable Elements: From Conflicts to Benefits'. *Nature Reviews Genetics* 18 (2): 2. https://doi.org/10.1038/nrg.2016.139.

Claycomb, Julie M. 2014. 'Ancient Endo-siRNA Pathways Reveal New Tricks'. *Current Biology* 24 (15): R703–15. https://doi.org/10.1016/j.cub.2014.06.009.

Coen, Dario, Bruno Lemaitre, Marion Delattre, et al. 1994. 'DrosophilaP Element: Transposition, Regulation and Evolution'. *Genetica* 93 (1–3): 61–78. https://doi.org/10.1007/BF01435240.

Colinet, Hervé, Siu Fai Lee, and Ary Hoffmann. 2010. 'Temporal Expression of Heat Shock Genes during Cold Stress and Recovery from Chill Coma in Adult Drosophila Melanogaster'. *The FEBS Journal* 277 (1): 174–85. https://doi.org/10.1111/j.1742-4658.2009.07470.x.

Colonna Romano, Nunzia, and Laura Fanti. 2022. 'Transposable Elements: Major Players in Shaping Genomic and Evolutionary Patterns'. *Cells* 11 (6): 1048. https://doi.org/10.3390/cells11061048.

Colwell, Robert K. 1974. 'Predictability, Constancy, and Contingency of Periodic Phenomena'. *Ecology* 55 (5): 1148–53. https://doi.org/10.2307/1940366.

Copley, Katie E., and James Shorter. 2023. 'Repetitive Elements in Aging and Neurodegeneration'. *Trends in Genetics* 39 (5): 381–400. https://doi.org/10.1016/j.tig.2023.02.008.

Cornette, Richard, Hiroki Gotoh, Shigeyuki Koshikawa, and Toru Miura. 2008. 'Juvenile Hormone Titers and Caste Differentiation in the Damp-Wood Termite *Hodotermopsis Sjostedti* (Isoptera, Termopsidae)'. *Journal of Insect Physiology* 54 (6): 922–30. https://doi.org/10.1016/j.jinsphys.2008.04.017.

Corona, Miguel, Romain Libbrecht, and Diana E Wheeler. 2016. 'Molecular Mechanisms of Phenotypic Plasticity in Social Insects'. *Current Opinion in Insect Science*, Insect genomics *

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Development and regulation, vol. 13 (February): 55–60. https://doi.org/10.1016/j.cois.2015.12.003.

Correa, Claudia C., and J. William O. Ballard. 2012. '*Wolbachia* Gonadal Density in Female and Male *Drosophila* Vary with Laboratory Adaptation and Respond Differently to Physiological and Environmental Challenges'. *Journal of Invertebrate Pathology* 111 (3): 197–204. https://doi.org/10.1016/j.jip.2012.08.003.

Costantini, David. 2024. 'For Better or Worse: How Early Life Oxidative Stress Moulds the Phenotype'. In *The Role of Organismal Oxidative Stress in the Ecology and Life-History Evolution of Animals*, edited by David Costantini. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65183-0_7.

Cotman, Carl W., and James L. McGaugh. 1980. '5 - Synaptic Transmission'. In *Behavioral Neuroscience*, edited by Carl W. Cotman and James L. McGaugh. Academic Press. https://doi.org/10.1016/B978-0-12-191650-3.50010-6.

Cox, Daniel N., Anna Chao, Jeff Baker, Lisa Chang, Dan Qiao, and Haifan Lin. 1998. 'A Novel Class of Evolutionarily Conserved Genes Defined by *Piwi* Are Essential for Stem Cell Self-Renewal'. *Genes & Development* 12 (23): 3715–27. https://doi.org/10.1101/gad.12.23.3715.

Craig, Elizabeth A., and Milton J. Schlesinger. 1985. 'The Heat Shock Respons'. *Critical Reviews in Biochemistry*, ahead of print, January 1. world. https://doi.org/10.3109/10409238509085135.

Cranz-Mileva, Susanne, Eve Reilly, Noor Chalhoub, et al. 2024. 'Transposon Removal Reveals Their Adaptive Fitness Contribution'. *Genome Biology and Evolution* 16 (2): evae010. https://doi.org/10.1093/gbe/evae010.

Curcio, M. Joan, and Keith M. Derbyshire. 2003. 'The Outs and Ins of Transposition: From Mu to Kangaroo'. *Nature Reviews Molecular Cell Biology* 4 (11): 865–77. https://doi.org/10.1038/nrm1241.

Czech, Benjamin, and Gregory J. Hannon. 2016. 'One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing'. *Trends in Biochemical Sciences* 41 (4): 324–37. https://doi.org/10.1016/j.tibs.2015.12.008.

Czech, Benjamin, Colin D. Malone, Rui Zhou, et al. 2008. 'An Endogenous Small Interfering RNA Pathway in Drosophila'. *Nature* 453 (7196): 798–802. https://doi.org/10.1038/nature07007.

Czech, Benjamin, Marzia Munafò, Filippo Ciabrelli, et al. 2018. 'piRNA-Guided Genome Defense: From Biogenesis to Silencing'. *Annual Review of Genetics* 52 (1): 131–57. https://doi.org/10.1146/annurev-genet-120417-031441.

Dana, Hassan, Ghanbar Mahmoodi Chalbatani, Habibollah Mahmoodzadeh, et al. 2017. 'Molecular Mechanisms and Biological Functions of siRNA'. *International Journal of Biomedical Science* 13 (2): 48–57. https://doi.org/10.59566/IJBS.2017.13048.

Danecek, Petr, James K Bonfield, Jennifer Liddle, et al. 2021. 'Twelve Years of SAMtools and BCFtools'. *GigaScience* 10 (2): giab008. https://doi.org/10.1093/gigascience/giab008.

Darwin, Charles. 1859. *On the Origin of Species*. John Murray. https://archive.org/details/onoriginofspec00darw.

David, J., Y. Cohet, and P. Fouillet. 1975. 'The Variability between Individuals as a Measure of Senescence: A Study of the Number of Eggs Laid and the Percentage of Hatched Eggs in the Case of *Drosophila Melanogaster*'. *Experimental Gerontology* 10 (1): 17–25. https://doi.org/10.1016/0531-5565(75)90011-X.

David, Jean R., and Pierre Capy. 1988. 'Genetic Variation of Drosophila Melanogaster Natural Populations'. *Trends in Genetics* 4 (4): 106–11. https://doi.org/10.1016/0168-9525(88)90098-4.

David, Jean R., Patricia Gibert, Emmanuelle Gravot, et al. 1997. 'Phenotypic Plasticity and Developmental Temperature in Drosophila: Analysis and Significance of Reaction Norms of Morphometrical Traits'. *Journal of Thermal Biology* 22 (6): 441–51. https://doi.org/10.1016/S0306-4565(97)00063-6.

Dawkins, Richard. 1976. *The Selfish Gene*. Oxford University Press.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Dawkins, Richard. 1982. *Richard Dawkins The Extended Phenotype The Long Reach Of The Gene (1999, Oxford University Press)*. http://archive.org/details/RichardDawkinsTheExtendedPhenotypeTheLongReachOfTheGene1999OxfordUniversityPress.

De Coster, Wouter, and Rosa Rademakers. 2023. 'NanoPack2: Population-Scale Evaluation of Long-Read Sequencing Data'. *Bioinformatics* 39 (5): btad311. https://doi.org/10.1093/bioinformatics/btad311.

Deininger, Prescott. 2011. 'Alu Elements: Know the SINEs'. *Genome Biology* 12 (12): 236. https://doi.org/10.1186/gb-2011-12-12-236.

Dej, Kimberley J., Tatiana Gerasimova, Victor G. Corces, and Jef D. Boeke. 1998. 'A Hotspot for the Drosophila Gypsy Retroelement in the Ovo Locus'. *Nucleic Acids Research* 26 (17): 4019–24. https://doi.org/10.1093/nar/26.17.4019.

Diamond, Sarah, and Ryan Martin. 2016. 'The Interplay between Plasticity and Evolution in Response to Human-Induced Environmental Change'. *F1000Research* 5 (December): 2835. https://doi.org/10.12688/f1000research.9731.1.

Diamond, Sarah, and Ryan Martin. 2021. *Buying Time: Plasticity and Population Persistence*. https://doi.org/10.1201/9780429343001-11.

Díaz-Castillo, Carlos. 2017. 'Junk DNA Contribution to Evolutionary Capacitance Can Drive Species Dynamics'. *Evolutionary Biology* 44 (2): 190–205. https://doi.org/10.1007/s11692-016-9404-5.

Ding, Wenyong, Lin Lin, Bing Chen, and Jianwu Dai. 2006. 'L1 Elements, Processed Pseudogenes and Retrogenes in Mammalian Genomes'. *IUBMB Life* 58 (12): 677–85. https://doi.org/10.1080/15216540601034856.

Dittman, Jeremy S. 2019. 'Unc13: A Multifunctional Synaptic Marvel'. *Current Opinion in Neurobiology*, Molecular Neuroscience, vol. 57 (August): 17–25. https://doi.org/10.1016/j.conb.2018.12.011.

Doak, T G, F P Doerder, C L Jahn, and G Herrick. 1994. 'A Proposed Superfamily of Transposase Genes: Transposon-like Elements in Ciliated Protozoa and a Common "D35E" Motif.' *Proceedings of the National Academy of Sciences* 91 (3): 942–46. https://doi.org/10.1073/pnas.91.3.942.

Dobson, Andy, Kevin D. Lafferty, Armand M. Kuris, Ryan F. Hechinger, and Walter Jetz. 2008. 'Homage to Linnaeus: How Many Parasites? How Many Hosts?' *Proceedings of the National Academy of Sciences* 105 (supplement_1): 11482–89. https://doi.org/10.1073/pnas.0803232105.

Dobzhansky, Theodosius. 1973. 'Nothing in Biology Makes Sense except in the Light of Evolution'. *The American Biology Teacher* 35 (3): 125–29. https://doi.org/10.2307/4444260.

Dominguez-Salas, Paula, Sophie E. Moore, Maria S. Baker, et al. 2014. 'Maternal Nutrition at Conception Modulates DNA Methylation of Human Metastable Epialleles'. *Nature Communications* 5 (1): 3746. https://doi.org/10.1038/ncomms4746.

Drew, Georgia C., Emily J. Stevens, and Kayla C. King. 2021. 'Microbial Evolution and Transitions along the Parasite–Mutualist Continuum'. *Nature Reviews Microbiology* 19 (10): 623–38. https://doi.org/10.1038/s41579-021-00550-7.

Drongitis, Denise, Francesco Aniello, Laura Fucci, and Aldo Donizetti. 2019. 'Roles of Transposable Elements in the Different Layers of Gene Expression Regulation'. *International Journal of Molecular Sciences* 20 (22): 22. https://doi.org/10.3390/ijms20225755.

Du, Juan, and Michael A. D. Goodisman. 2024. 'The Role of Epigenetics in Insects in Changing Environments'. *Insect Molecular Biology* 33 (5): 429–31. https://doi.org/10.1111/imb.12947.

Dubin, Manu J, Ortrun Mittelsten Scheid, and Claude Becker. 2018. 'Transposons: A Blessing Curse'. *Current Opinion in Plant Biology*, 42 Genome studies and molecular genetics 2018, vol. 42 (April): 23–29. https://doi.org/10.1016/j.pbi.2018.01.003.

Dufty, A. 2002. 'Hormones, Developmental Plasticity and Adaptation'. *Trends in Ecology & Evolution* 17 (4):

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

190–96. https://doi.org/10.1016/S0169-5347(02)02498-9.

Duncan, Elizabeth J., Christopher B. Cunningham, and Peter K. Dearden. 2022. 'Phenotypic Plasticity: What Has DNA Methylation Got to Do with It?' *Insects* 13 (2): 110. https://doi.org/10.3390/insects13020110.

Duncan, Elizabeth J., Peter D. Gluckman, and Peter K. Dearden. 2014. 'Epigenetics, Plasticity, and Evolution: How Do We Link Epigenetic Change to Phenotype?' *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 322 (4): 208–20. https://doi.org/10.1002/jez.b.22571.

Dunkler, Daniela, Maria Haller, Rainer Oberbauer, and Georg Heinze. 2020. 'To Test or to Estimate? P-Values versus Effect Sizes'. *Transplant International* 33 (1): 50–55. https://doi.org/10.1111/tri.13535.

Duron, Olivier, and Yuval Gottlieb. 2020. 'Convergence of Nutritional Symbioses in Obligate Blood Feeders'. *Trends in Parasitology* 36 (10): 816–25. https://doi.org/10.1016/j.pt.2020.07.007.

Ecker, Simone, Vera Pancaldi, Alfonso Valencia, Stephan Beck, and Dirk S. Paul. 2018. 'Epigenetic and Transcriptional Variability Shape Phenotypic Plasticity'. *BioEssays* 40 (2): 1700148. https://doi.org/10.1002/bies.201700148.

Eddy, Sean R. 2001. 'Non–Coding RNA Genes and the Modern RNA World'. *Nature Reviews Genetics* 2 (12): 919–29. https://doi.org/10.1038/35103511.

Eickbush, T. H. 1992. 'Transposing without Ends: The Non-LTR Retrotransposable Elements'. *The New Biologist* 4 (5): 430–40.

Eickbush, Thomas H., and Harmit S. Malik. 2007. 'Origins and Evolution of Retrotransposons'. In *Mobile DNA II*. John Wiley & Sons, Ltd. https://doi.org/10.1128/9781555817954.ch49.

Elgin, Sarah C. R., and Gunter Reuter. 2013. 'Position-Effect Variegation, Heterochromatin Formation, and Gene Silencing in Drosophila'. *Cold Spring Harbor Perspectives in Biology* 5 (8): a017780. https://doi.org/10.1101/cshperspect.a017780.

Eugénio, Ana T, Marta S P Marialva, and Patrícia Beldade. 2023. 'Effects of Wolbachia on Transposable Element Expression Vary Between Drosophila Melanogaster Host Genotypes'. *Genome Biology and Evolution* 15 (3): evad036. https://doi.org/10.1093/gbe/evad036.

Evans, Abigail DiVito, Regina A. Fairbanks, Paul Schmidt, and Mia T. Levine. 2023. 'An Epigenetic Basis of Adaptive Plasticity in Drosophila Melanogaster'. Preprint, bioRxiv, January 12. https://doi.org/10.1101/2022.10.11.511590.

Evans, Oliver, Eric P. Caragata, Conor J. McMeniman, et al. 2009. 'Increased Locomotor Activity and Metabolism of Aedes Aegyptiinfected with a Life-Shortening Strain of Wolbachia Pipientis'. *Journal of Experimental Biology* 212 (10): 1436–41. https://doi.org/10.1242/jeb.028951.

Evgen'ev, M.B., and I.R. Arkhipova. 2005. 'Penelope-like Elements – a New Class of Retroelements: Distribution, Function and Possible Evolutionary Significance'. *Cytogenetic and Genome Research* 110 (1–4): 510–21. https://doi.org/10.1159/000084984.

Evgen'ev, Michael B., Helena Zelentsova, Natalia Shostak, et al. 1997. 'Penelope, a New Family of Transposable Elements and Its Possible Role in Hybrid Dysgenesis in Drosophila Virilis'. *Proceedings of the National Academy of Sciences* 94 (1): 196–201. https://doi.org/10.1073/pnas.94.1.196.

Fablet, Marie, and Cristina Vieira. 2011. *Evolvability, Epigenetics and Transposable Elements*. Biomolecular Concepts. 2 (5): 333–41. https://doi.org/10.1515/BMC.2011.035.

Fabry, Martin H, Federica A Falconio, Fadwa Joud, Emily K Lythgoe, Benjamin Czech, and Gregory J Hannon. 2021. 'Maternally Inherited piRNAs Direct Transient Heterochromatin Formation at Active Transposons during Early Drosophila Embryogenesis'. *eLife* 10 (July): e68573. https://doi.org/10.7554/eLife.68573.

Falconer, D S. 1960. *QUANTITATIVE GENETICS*.

Falconer, Douglas S. 1996. *Introduction To Quantitative Genetics 4th Edition*. http://archive.org/details/IntroductionToQuantitativeGenetics.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Fattash, Isam, Rebecca Rooke, Amy Wong, et al. 2013. 'Miniature Inverted-Repeat Transposable Elements: Discovery, Distribution, and Activity'. *Genome* 56 (9): 475–86. https://doi.org/10.1139/gen-2012-0174.

Feder, Martin E., and Gretchen E. Hofmann. 1999. 'HEAT-SHOCK PROTEINS, MOLECULAR CHAPERONES, AND THE STRESS RESPONSE: Evolutionary and Ecological Physiology'. *Annual Review of Physiology* 61 (Volume 61, 1999): 243–82. https://doi.org/10.1146/annurev.physiol.61.1.243.

Feige, U., and B. S. Polla. 1994. 'Heat Shock Proteins: The Hsp70 Family'. *Experientia* 50 (11): 979–86. https://doi.org/10.1007/BF01923452.

Feil, Robert, and Mario F. Fraga. 2012. 'Epigenetics and the Environment: Emerging Patterns and Implications'. *Nature Reviews Genetics* 13 (2): 2. https://doi.org/10.1038/nrg3142.

Ferrari, Julia, and Fabrice Vavre. 2011. 'Bacterial Symbionts in Insects or the Story of Communities Affecting Communities'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366 (1569): 1389–400. https://doi.org/10.1098/rstb.2010.0226.

Feschotte, Cédric. 2008. 'Transposable Elements and the Evolution of Regulatory Networks'. *Nature Reviews Genetics* 9 (5): 397–405. https://doi.org/10.1038/nrg2337.

Feschotte, Cédric. 2023. 'Transposable Elements: McClintock's Legacy Revisited'. *Nature Reviews Genetics* 24 (11): 797–800. https://doi.org/10.1038/s41576-023-00652-3.

Feschotte, Cédric, and Ellen J. Pritham. 2007. 'DNA Transposons and the Evolution of Eukaryotic Genomes'. *Annual Review of Genetics* 41 (Volume 41, 2007): 331–68. https://doi.org/10.1146/annurev.genet.40.110405.090448.

Field, L. M., F. Lyko, M. Mandrioli, and G. Prantera. 2004. 'DNA Methylation in Insects'. *Insect Molecular Biology* 13 (2): 109–15. https://doi.org/10.1111/j.0962-1075.2004.00470.x.

Finnegan, David J. 1992. 'Transposable Elements'. *Current Opinion in Genetics & Development* 2 (6): 861–67. https://doi.org/10.1016/S0959-437X(05)80108-X.

Fitz-James, Maximilian H., and Giacomo Cavalli. 2022. 'Molecular Mechanisms of Transgenerational Epigenetic Inheritance'. *Nature Reviews Genetics* 23 (6): 325–41. https://doi.org/10.1038/s41576-021-00438-5.

Flatt, Thomas. 2020. 'Life-History Evolution and the Genetics of Fitness Components in Drosophila Melanogaster'. *Genetics* 214 (1): 3–48. https://doi.org/10.1534/genetics.119.300160.

Fleury, F., F. Vavre, N. Ris, P. Fouillet, and M. Boulétreau. 2000. 'Physiological Cost Induced by the Maternally-Transmitted Endosymbiont Wolbachia in the Drosophila Parasitoid Leptopilina Heterotoma'. *Parasitology* 121 (5): 493–500. https://doi.org/10.1017/S0031182099006599.

Foret, Sylvain, Robert Kucharski, Yvonne Pittelkow, Gabrielle A. Lockett, and Ryszard Maleszka. 2009. 'Epigenetic Regulation of the Honey Bee Transcriptome: Unravelling the Nature of Methylated Genes'. *BMC Genomics* 10 (1): 472. https://doi.org/10.1186/1471-2164-10-472.

Fox, Rebecca J., Jennifer M. Donelson, Celia Schunter, Timothy Ravasi, and Juan D. Gaitán-Espitia. 2019. 'Beyond Buying Time: The Role of Plasticity in Phenotypic Adaptation to Rapid Environmental Change'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374 (1768): 20180174. https://doi.org/10.1098/rstb.2018.0174.

Frías-Lasserre, Daniel, and Cristian A. Villagra. 2017a. 'The Importance of ncRNAs as Epigenetic Mechanisms in Phenotypic Variation and Organic Evolution'. *Frontiers in Microbiology* 8 (December): 2483. https://doi.org/10.3389/fmicb.2017.02483.

Frías-Lasserre, Daniel, and Cristian A. Villagra. 2017b. 'The Importance of ncRNAs as Epigenetic Mechanisms in Phenotypic Variation and Organic Evolution'. *Frontiers in Microbiology* 8. https://www.frontiersin.org/articles/10.3389/fmicb.2017.02483.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Fry, A. J., M. R. Palmer, and D. M. Rand. 2004. 'Variable Fitness Effects of Wolbachia Infection in Drosophila Melanogaster'. *Heredity* 93 (4): 4. https://doi.org/10.1038/sj.hdy.6800514.

Fry, Adam J., and David M. Rand. 2002. 'WOLBACHIA INTERACTIONS THAT DETERMINE DROSOPHILA MELANOGASTER SURVIVAL'. *Evolution* 56 (10): 1976–81. https://doi.org/10.1111/j.0014-3820.2002.tb00123.x.

Fu, Xiang-Dong. 2014. 'Non-Coding RNA: A New Frontier in Regulatory Biology'. *National Science Review* 1 (2): 190–204. https://doi.org/10.1093/nsr/nwu008.

Fueyo, Raquel, Julius Judd, Cedric Feschotte, and Joanna Wysocka. 2022. 'Roles of Transposable Elements in the Regulation of Mammalian Transcription'. *Nature Reviews Molecular Cell Biology* 23 (7): 481–97. https://doi.org/10.1038/s41580-022-00457-y.

Galagan, James E., and Eric U. Selker. 2004. 'RIP: The Evolutionary Cost of Genome Defense'. *Trends in Genetics* 20 (9): 417–23. https://doi.org/10.1016/j.tig.2004.07.007.

Galbraith, James D., and Alexander Hayward. 2023. 'The Influence of Transposable Elements on Animal Colouration'. *Trends in Genetics* 0 (0). https://doi.org/10.1016/j.tig.2023.04.005.

García-Nieto, Pablo E., Ban Wang, and Hunter B. Fraser. 2022. 'Transcriptome Diversity Is a Systematic Source of Variation in RNA-Sequencing Data'. *PLOS Computational Biology* 18 (3): e1009939. https://doi.org/10.1371/journal.pcbi.1009939.

Gautier, Mathieu. 2015. 'Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates'. *Genetics* 201 (4): 1555–79. https://doi.org/10.1534/genetics.115.181453.

Gebert, Daniel, and David Rosenkranz. 2015. 'RNA - based Regulation of Transposon Expression'. *WIREs RNA* 6 (6): 687–708. https://doi.org/10.1002/wrna.1310.

Gebhardt, M. D., and S. C. Stearns. 1993. 'Phenotypic Plasticity for Life History Traits in Drosophila Melanogaster. I. Effect on Phenotypic and Environmental Correlations'. *Journal of Evolutionary Biology* 6 (1): 1–16. https://doi.org/10.1046/j.1420-9101.1993.6010001.x.

Gebrie, Alemu. 2023. 'Transposable Elements as Essential Elements in the Control of Gene Expression'. *Mobile DNA* 14 (August): 9. https://doi.org/10.1186/s13100-023-00297-3.

Ghalambor, C. K., J. K. McKAY, S. P. Carroll, and D. N. Reznick. 2007. 'Adaptive versus Non-Adaptive Phenotypic Plasticity and the Potential for Contemporary Adaptation in New Environments'. *Functional Ecology* 21 (3): 394–407. https://doi.org/10.1111/j.1365-2435.2007.01283.x.

Ghanim, George E., Donald C. Rio, and Felipe Karam Teixeira. 2020. 'Mechanism and Regulation of P Element Transposition'. *Open Biology* 10 (12): 200244. https://doi.org/10.1098/rsob.200244.

Gibert, Jean-Michel, Emmanuèle Mouchel-Vielh, Sandra De Castro, and Frédérique Peronnet. 2016. 'Phenotypic Plasticity through Transcriptional Regulation of the Evolutionary Hotspot Gene Tan in Drosophila Melanogaster'. *PLoS Genetics* 12 (8): e1006218. https://doi.org/10.1371/journal.pgen.1006218.

Gibert, Jean-Michel, Emmanuèle Mouchel-Vielh, and Frédérique Peronnet. 2017. 'Modulation of Yellow Expression Contributes to Thermal Plasticity of Female Abdominal Pigmentation in Drosophila Melanogaster'. *Scientific Reports* 7 (1): 43370. https://doi.org/10.1038/srep43370.

Gibert, Patricia, Brigitte Moreteau, Ashok Munjal, and Jean R David. 1999. *Phenotypic Plasticity of Abdominal Pigmentation in Drosophila Kikkawai: Multiple Interactions between a Major Gene, Sex, Abdomen Segment and Growth Temperature*.

Gibert, Patricia, Samuel M Scheiner, and Jean R David. 1998. *Phenotypic Plasticity of Body Pigmentation in Drosophila: Correlated Variations between Segments*.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Gibson, Greg, and Ian Dworkin. 2004. 'Uncovering Cryptic Genetic Variation'. *Nature Reviews Genetics* 5 (9): 681–90. https://doi.org/10.1038/nrg1426.

Gibson, Greg, and Laura K. Reed. 2008. 'Cryptic Genetic Variation'. *Current Biology* 18 (21): R989–90. https://doi.org/10.1016/j.cub.2008.08.011.

Gilbert, Clément, Jean Peccoud, and Richard Cordaux. 2021a. 'Transposable Elements and the Evolution of Insects'. *Annual Review of Entomology* 66 (Volume 66, 2021): 355–72. https://doi.org/10.1146/annurev-ento-070720-074650.

Gilbert, Clément, Jean Peccoud, and Richard Cordaux. 2021b. 'Transposable Elements and the Evolution of Insects'. *Annual Review of Entomology* 66 (Volume 66, 2021): 355–72. https://doi.org/10.1146/annurev-ento-070720-074650.

Gilbert, Scott F. 2012. 'Ecological Developmental Biology: Environmental Signals for Normal Animal Development'. *Evolution & Development* 14 (1): 20–28. https://doi.org/10.1111/j.1525-142X.2011.00519.x.

Glastad, K. M., B. G. Hunt, S. V. Yi, and M. a. D. Goodisman. 2011. 'DNA Methylation in Insects: On the Brink of the Epigenomic Era'. *Insect Molecular Biology* 20 (5): 553–65. https://doi.org/10.1111/j.1365-2583.2011.01092.x.

Gluck-Thaler, Emile, Timothy Ralston, Zachary Konkel, et al. 2022. 'Giant Starship Elements Mobilize Accessory Genes in Fungal Genomes'. *Molecular Biology and Evolution* 39 (5): msac109. https://doi.org/10.1093/molbev/msac109.

Goldstein, Ilan, and Ian M. Ehrenreich. 2021. 'Genetic Variation in Phenotypic Plasticity'. In *Phenotypic Plasticity & Evolution*. CRC Press.

González, Josefa, Kapa Lenkov, Mikhail Lipatov, J. Michael Macpherson, and Dmitri A. Petrov. 2008. 'High Rate of Recent Transposable Element–Induced Adaptation in Drosophila Melanogaster'. *PLOS Biology* 6 (10): e251. https://doi.org/10.1371/journal.pbio.0060251.

González, Josefa, and Dmitri A. Petrov. 2009. 'The Adaptive Role of Transposable Elements in the *Drosophila* Genome'. *Gene*, Genomic Impact of Eukaryotic Transposable Elements, vol. 448 (2): 124–33. https://doi.org/10.1016/j.gene.2009.06.008.

González-Tokman, Daniel, Alex Córdoba-Aguilar, Wesley Dáttilo, Andrés Lira-Noriega, Rosa A. Sánchez-Guillén, and Fabricio Villalobos. 2020. 'Insect Responses to Heat: Physiological Mechanisms, Evolution and Ecological Implications in a Warming World'. *Biological Reviews* 95 (3): 802–21. https://doi.org/10.1111/brv.12588.

Goodacre, Sara L., Oliver Y. Martin, Dries Bonte, et al. 2009. 'Microbial Modification of Host Long-Distance Dispersal Capacity'. *BMC Biology* 7 (1): 32. https://doi.org/10.1186/1741-7007-7-32.

Goodier, John L., and Haig H. Kazazian. 2008. 'Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites'. *Cell* 135 (1): 23–35. https://doi.org/10.1016/j.cell.2008.09.022.

Goodwin, Timothy J. D., Margaret I. Butler, and Russell T. M. Poulter. 2003. 'Cryptons: A Group of Tyrosine-Recombinase-Encoding DNA Transposons from Pathogenic Fungi'. *Microbiology* 149 (11): 3099–109. https://doi.org/10.1099/mic.0.26529-0.

Goodwin, Timothy J. D., and Russell T. M. Poulter. 2004. 'A New Group of Tyrosine Recombinase-Encoding Retrotransposons'. *Molecular Biology and Evolution* 21 (4): 746–59. https://doi.org/10.1093/molbev/msh072.

Goriaux, Coline, Sophie Desset, Yoan Renaud, Chantal Vaury, and Emilie Brasset. 2014. 'Transcriptional Properties and Splicing of the Flamenco piRNA Cluster'. *EMBO Reports* 15 (4): 411–18. https://doi.org/10.1002/embr.201337898.

Graba, Yacine, Denise Aragnol, and Jacques Pradel. 1997. 'Drosophila Hox Complex Downstream Targets and the Function of Homeotic Genes'. *BioEssays* 19 (5): 379–88. https://doi.org/10.1002/bies.950190505.

Grabundzija, Ivana, Simon A. Messing, Jainy Thomas, et al. 2016. 'A Helitron Transposon Reconstructed from Bats Reveals a Novel Mechanism of Genome

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Shuffling in Eukaryotes'. *Nature Communications* 7 (1): 10716. https://doi.org/10.1038/ncomms10716.

Grandi, Fiorella C., and Wenfeng An. 2013. 'Non-LTR Retrotransposons and Microsatellites: Partners in Genomic Variation'. *Mobile Genetic Elements* 3 (4): e25674. https://doi.org/10.4161/mge.25674.

Grangeteau, Cédric, Fairouz Yahou, Claude Everaerts, et al. 2018. 'Yeast Quality in Juvenile Diet Affects Drosophila Melanogaster Adult Life Traits'. *Scientific Reports* 8 (1): 13070. https://doi.org/10.1038/s41598-018-31561-9.

Green, Delbert A., and Cassandra G. Extavour. 2014. 'Insulin Signalling Underlies Both Plasticity and Divergence of a Reproductive Trait in Drosophila'. *Proceedings of the Royal Society B: Biological Sciences* 281 (1779): 20132673. https://doi.org/10.1098/rspb.2013.2673.

Gregory, M., L. Alphey, N. I. Morrison, and S. M. Shimeld. 2016. 'Insect Transformation with piggyBac: Getting the Number of Injections Just Right'. *Insect Molecular Biology* 25 (3): 259–71. https://doi.org/10.1111/imb.12220.

Groza, Cristian, Xun Chen, Travis J. Wheeler, Guillaume Bourque, and Clément Goubert. 2024. 'A Unified Framework to Analyze Transposable Element Insertion Polymorphisms Using Graph Genomes'. *Nature Communications* 15 (1): 8915. https://doi.org/10.1038/s41467-024-53294-2.

Gruntenko, Nataly E., Yury Yu. Ilinsky, Natalya V. Adonyeva, et al. 2017. 'Various Wolbachia Genotypes Differently Influence Host Drosophila Dopamine Metabolism and Survival under Heat Stress Conditions'. *BMC Evolutionary Biology* 17 (2): 252. https://doi.org/10.1186/s12862-017-1104-y.

Grzebelus, Dariusz. 2006. *TRANSPOSON INSERTION POLYMORPHISM AS A NEW SOURCE OF MOLECULAR MARKERS*. 14.

Guio, Lain, and Josefa González. 2019. 'New Insights on the Evolution of Genome Content: Population Dynamics of Transposable Elements in Flies and Humans'. In *Evolutionary Genomics: Statistical and Computational Methods*, edited by Maria Anisimova. Springer. https://doi.org/10.1007/978-1-4939-9074-0_16.

Guio, Lain, Cristina Vieira, and Josefa González. 2018. 'Stress Affects the Epigenetic Marks Added by Natural Transposable Element Insertions in Drosophila Melanogaster'. *Scientific Reports* 8 (1): 1. https://doi.org/10.1038/s41598-018-30491-w.

Guo, Ya-Long, Jesper S. Bechsgaard, Tanja Slotte, et al. 2009. 'Recent Speciation of Capsella Rubella from Capsella Grandiflora, Associated with Loss of Self-Incompatibility and an Extreme Bottleneck'. *Proceedings of the National Academy of Sciences* 106 (13): 5246–51. https://doi.org/10.1073/pnas.0808012106.

Gupta, Ayushi, and Suresh Nair. 2025. 'Epigenetic Processes in Insect Adaptation to Environmental Stress'. *Current Opinion in Insect Science* 67 (February): 101294. https://doi.org/10.1016/j.cois.2024.101294.

Gusev, Oleg, Richard Cornette, Takahiro Kikawada, and Takashi Okuda. 2011. 'Expression of Heat Shock Protein-Coding Genes Associated with Anhydrobiosis in an African Chironomid *Polypedilum Vanderplanki*'. *Cell Stress and Chaperones* 16 (1): 81–90. https://doi.org/10.1007/s12192-010-0223-9.

Gutzeit, Herwig O., and Roswitha Koppa. 1982. 'Time-Lapse Film Analysis of Cytoplasmic Streaming during Late Oogenesis of Drosophila'. *Development* 67 (1): 101–11. https://doi.org/10.1242/dev.67.1.101.

Haag, Karen Luisa. 2018. 'Holobionts and Their Hologenomes: Evolution with Mixed Modes of Inheritance'. *Genetics and Molecular Biology* 41: 189–97. https://doi.org/10.1590/1678-4685-GMB-2017-0070.

Haapa-Paananen, Saija, Niklas Wahlberg, and Harri Savilahti. 2014. 'Phylogenetic Analysis of Maverick/Polinton Giant Transposons across Organisms'. *Molecular Phylogenetics and Evolution* 78 (September): 271–74. https://doi.org/10.1016/j.ympev.2014.05.024.

Hall, Michael B. 2022. 'Rasusa: Randomly Subsample Sequencing Reads to a Specified Coverage'. *Journal of*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Open Source Software 7 (69): 3941. https://doi.org/10.21105/joss.03941.

Hallgrímsson, Benedikt, and Brian K Hall. 2005. *Variation and Variability: Central Concepts in Biology*.

Hamilton, W. D. 1964. 'The Genetical Evolution of Social Behaviour. II'. *Journal of Theoretical Biology* 7 (1): 17–52. https://doi.org/10.1016/0022-5193(64)90039-6.

Hammes, Stephen R., and Ellis R. Levin. 2007. 'Extranuclear Steroid Receptors: Nature and Actions'. *Endocrine Reviews* 28 (7): 726–41. https://doi.org/10.1210/er.2007-0022.

Havula, Essi, Mari Teesalu, Tuulia Hyötyläinen, et al. 2013. 'Mondo/ChREBP-Mlx-Regulated Transcriptional Network Is Essential for Dietary Sugar Tolerance in Drosophila'. *PLOS Genetics* 9 (4): e1003438. https://doi.org/10.1371/journal.pgen.1003438.

Haws, Spencer A., Zoltan Simandi, R. Jordan Barnett, and Jennifer E. Phillips-Cremins. 2022. '3D Genome, on Repeat: Higher-Order Folding Principles of the Heterochromatinized Repetitive Genome'. *Cell* 185 (15): 2690–707. https://doi.org/10.1016/j.cell.2022.06.052.

Hayward, Alexander, and Clément Gilbert. 2022. 'Transposable Elements'. *Current Biology* 32 (17): R904–9. https://doi.org/10.1016/j.cub.2022.07.044.

Hayward, April, and James F. Gillooly. 2011. 'The Cost of Sex: Quantifying Energetic Investment in Gamete Production by Males and Females'. *PLOS ONE* 6 (1): e16557. https://doi.org/10.1371/journal.pone.0016557.

He, Xin-Jian, Taiping Chen, and Jian-Kang Zhu. 2011. 'Regulation and Function of DNA Methylation in Plants and Animals'. *Cell Research* 21 (3): 442–65. https://doi.org/10.1038/cr.2011.23.

Heard, Edith, and Robert A. Martienssen. 2014. 'Transgenerational Epigenetic Inheritance: Myths and Mechanisms'. *Cell* 157 (1): 95–109. https://doi.org/10.1016/j.cell.2014.02.045.

Hedges, D. J., and P. L. Deininger. 2007. 'Inviting Instability: Transposable Elements, Double-Strand Breaks, and the Maintenance of Genome Integrity'. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, Dedicated in memory of Dr. Tony Carrano, vol. 616 (1): 46–59. https://doi.org/10.1016/j.mrfmmm.2006.11.021.

Henikoff, Steven. 2023. 'The Epigenetic Landscape: An Evolving Concept'. *Frontiers in Epigenetics and Epigenomics* 1 (March). https://doi.org/10.3389/freae.2023.1176449.

Heyland, Andreas, Jason Hodin, and Adam M. Reitzel. 2005. 'Hormone Signaling in Evolution and Development: A Non-Model System Approachs'. *BioEssays* 27 (1): 64–75. https://doi.org/10.1002/bies.20136.

Hilgenboecker, Kirsten, Peter Hammerstein, Peter Schlattmann, Arndt Telschow, and John H. Werren. 2008. 'How Many Species Are Infected with Wolbachia? – A Statistical Analysis of Current Data'. *FEMS Microbiology Letters* 281 (2): 215–20. https://doi.org/10.1111/j.1574-6968.2008.01110.x.

Hill, William G, and Trudy F C Mackay. 2004. 'D. S. Falconer and Introduction to Quantitative Genetics'. *Genetics* 167 (4): 1529–36. https://doi.org/10.1093/genetics/167.4.1529.

Hill, William, and Trudy Mackay. 2004. 'D. S. Falconer and Introduction to Quantitative Genetics'. *Genetics* 167 (September): 1529–36. https://doi.org/10.1093/genetics/167.4.1529.

Hirsch, Cory D., and Nathan M. Springer. 2017. 'Transposable Element Influences on Gene Expression in Plants'. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, Plant Gene Regulatory Mechanisms and Networks, vol. 1860 (1): 157–65. https://doi.org/10.1016/j.bbagrm.2016.05.010.

Hoffmann, Jules A. 2003. 'The Immune Response of Drosophila'. *Nature* 426 (6962): 33–38. https://doi.org/10.1038/nature02021.

Hofmann, Gretchen E. 2005. 'Patterns of Hsp Gene Expression in Ectothermic Marine Organisms on Small to Large Biogeographic Scales1'. *Integrative and*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Comparative Biology 45 (2): 247–55. https://doi.org/10.1093/icb/45.2.247.

Holliday, Robin. 2006. 'Epigenetics: A Historical Overview'. *Epigenetics* 1 (2): 76–80. https://doi.org/10.4161/epi.1.2.2762.

Hopper, Keith R. 1999. 'RISK-SPREADING AND BET-HEDGING IN INSECT POPULATION BIOLOGY1'. *Annual Review of Entomology* 44 (Volume 44, 1999): 535–60. https://doi.org/10.1146/annurev.ento.44.1.535.

Horváth, Barbara, and Alex T. Kalinka. 2016. 'Effects of Larval Crowding on Quantitative Variation for Development Time and Viability in Drosophila Melanogaster'. *Ecology and Evolution* 6 (23): 8460–73. https://doi.org/10.1002/ece3.2552.

Horváth, Vivien, Miriam Merenciano, and Josefa González. 2017. 'Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response'. *Trends in Genetics* 33 (11): 832–41. https://doi.org/10.1016/j.tig.2017.08.007.

Hosokawa, Takahiro, Ryuichi Koga, Yoshitomo Kikuchi, Xian-Ying Meng, and Takema Fukatsu. 2010. 'Wolbachia as a Bacteriocyte-Associated Nutritional Mutualist'. *Proceedings of the National Academy of Sciences* 107 (2): 769–74. https://doi.org/10.1073/pnas.0911476107.

Huang, Hsin-Ya, Siqi Zhang, Fadi Abou Choucha, et al. 2024. 'Harbinger Transposon Insertion in Ethylene Signaling Gene Leads to Emergence of New Sexual Forms in Cucurbits'. *Nature Communications* 15 (1): 4877. https://doi.org/10.1038/s41467-024-49250-9.

Huang, Wen, Mary Anna Carbone, Richard F. Lyman, Robert R. H. Anholt, and Trudy F. C. Mackay. 2020. 'Genotype by Environment Interaction for Gene Expression in Drosophila Melanogaster'. *Nature Communications* 11 (1): 5451. https://doi.org/10.1038/s41467-020-19131-y.

Huang, Wen, Mary Anna Carbone, Michael M. Magwire, et al. 2015. 'Genetic Basis of Transcriptome Diversity in Drosophila Melanogaster'. *Proceedings of the National Academy of Sciences* 112 (44): E6010–19. https://doi.org/10.1073/pnas.1519159112.

Huang, Yuheng, and Yuh Chwen G. Lee. 2024. 'Blessing or Curse: How the Epigenetic Resolution of Host-Transposable Element Conflicts Shapes Their Evolutionary Dynamics'. *Proceedings of the Royal Society B: Biological Sciences* 291 (2020): 20232775. https://doi.org/10.1098/rspb.2023.2775.

Hunter, Richard G. 2024. 'Transposons as Environmental Stress Detectors, Are Eukaryotic Genomes Evolved to Evolve?' In *Epigenetics in Biological Communication*, edited by Guenther Witzany. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-59286-7_13.

Huntzinger, Eric, and Elisa Izaurralde. 2011. 'Gene Silencing by microRNAs: Contributions of Translational Repression and mRNA Decay'. *Nature Reviews Genetics* 12 (2): 99–110. https://doi.org/10.1038/nrg2936.

Husby, Arild. 2022. 'Wild Epigenetics: Insights from Epigenetic Studies on Natural Populations'. *Proceedings of the Royal Society B: Biological Sciences* 289 (1968): 20211633. https://doi.org/10.1098/rspb.2021.1633.

Hwang, Sung-Yeon, Hyunchul Jung, Seyoung Mun, et al. 2021. 'L1 Retrotransposons Exploit RNA m6A Modification as an Evolutionary Driving Force'. *Nature Communications* 12 (1): 880. https://doi.org/10.1038/s41467-021-21197-1.

Hyun, Seogang. 2017. 'Small RNA Pathways That Protect the Somatic Genome'. *International Journal of Molecular Sciences* 18 (5): 912. https://doi.org/10.3390/ijms18050912.

Iida, Shigeru, Yasumasa Morita, Jeong-Doo Choi, Kyeung-Il Park, and Atsushi Hoshino. 2004. 'Genetics and Epigenetics in Flower Pigmentation Associated with Transposable Elements in Morning Glories'. *Advances in Biophysics* 38 (January): 141–59. https://doi.org/10.1016/S0065-227X(04)80136-9.

Ishikawa, Asano, Hiroki Gotoh, Taisuke Abe, and Toru Miura. 2013. 'Juvenile Hormone Titer and Wing-Morph Differentiation in the Vetch Aphid *Megoura Crassicauda*'. *Journal of Insect Physiology* 59 (4): 444–49. https://doi.org/10.1016/j.jinsphys.2013.02.004.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Ito, Hidetaka. 2022. 'Environmental Stress and Transposons in Plants'. *Genes & Genetic Systems* 97 (4): 169–75. https://doi.org/10.1266/ggs.22-00045.

Ivanov, Dobril K., Valentina Escott-Price, Matthias Ziehm, et al. 2015. 'Longevity GWAS Using the Drosophila Genetic Reference Panel'. *The Journals of Gerontology: Series A* 70 (12): 1470–78. https://doi.org/10.1093/gerona/glv047.

Jaenisch, Rudolf, and Adrian Bird. 2003. 'Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals'. *Nature Genetics* 33 (3): 3. https://doi.org/10.1038/ng1089.

Jarosz, Daniel F., and Susan Lindquist. 2010. 'Hsp90 and Environmental Stress Transform the Adaptive Value of Natural Genetic Variation'. *Science* 330 (6012): 1820–24. https://doi.org/10.1126/science.1195487.

Jedlicka, Paul, Mark A. Mortin, and Carl Wu. 1997. 'Multiple Functions of Drosophila Heat Shock Transcription Factor in Vivo'. *The EMBO Journal* 16 (9): 2452–62. https://doi.org/10.1093/emboj/16.9.2452.

Jeffreys, Harold. 1948. *Theory Of Probability*. http://archive.org/details/in.ernet.dli.2015.2608.

Jiang, Jingchen, Liuya Zhao, Lan Yan, et al. 2014. 'Structural Features and Mechanism of Translocation of Non-LTR Retrotransposons in Candida Albicans'. *Virulence* 5 (2): 245–52. https://doi.org/10.4161/viru.27278.

Joly-Lopez, Zoé, and Thomas E Bureau. 2018. 'Exaptation of Transposable Element Coding Sequences'. *Current Opinion in Genetics & Development* 49 (April): 34–42. https://doi.org/10.1016/j.gde.2018.02.011.

Joly-Lopez, Zoé, Ewa Forczek, Emilio Vello, Douglas R. Hoen, Akiko Tomita, and Thomas E. Bureau. 2017. 'Abiotic Stress Phenotypes Are Associated with Conserved Genes Derived from Transposable Elements'. *Frontiers in Plant Science* 8 (November). https://doi.org/10.3389/fpls.2017.02027.

Jones, Brian C., Jason G. Wood, Chengyi Chang, et al. 2016. 'A Somatic piRNA Pathway in the Drosophila Fat Body Ensures Metabolic Homeostasis and Normal Lifespan'. *Nature Communications* 7 (1): 1. https://doi.org/10.1038/ncomms13856.

Jones, Emma F, Anisha Haldar, Vishal H Oza, and Brittany N Lasseigne. 2024. 'Quantifying Transcriptome Diversity: A Review'. *Briefings in Functional Genomics* 23 (2): 83–94. https://doi.org/10.1093/bfgp/elad019.

Jones, Peter A. 2012. 'Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and Beyond'. *Nature Reviews Genetics* 13 (7): 484–92. https://doi.org/10.1038/nrg3230.

Jordan, Katherine W, Theodore J Morgan, and Trudy F C Mackay. 2006. 'Quantitative Trait Loci for Locomotor Behavior in Drosophila Melanogaster'. *Genetics* 174 (1): 271–84. https://doi.org/10.1534/genetics.106.058099.

José Gosalbes, María, Amparo Latorre, Araceli Lamelas, and Andrés Moya. 2010. 'Genomics of Intracellular Symbionts in Insects'. *International Journal of Medical Microbiology* 300 (5): 271–78. https://doi.org/10.1016/j.ijmm.2009.12.001.

Julian Huxley. 1942. *Evolution The Modern Synthesis*. http://archive.org/details/in.ernet.dli.2015.280031.

Kapitonov, Vladimir V., and Jerzy Jurka. 2006. 'Self-Synthesizing DNA Transposons in Eukaryotes'. *Proceedings of the National Academy of Sciences* 103 (12): 4540–45. https://doi.org/10.1073/pnas.0600833103.

Karam, Joseph A., Rasesh Y. Parikh, Dhananjaya Nayak, David Rosenkranz, and Vamsi K. Gangaraju. 2017. 'Co-Chaperone Hsp70/Hsp90-Organizing Protein (Hop) Is Required for Transposon Silencing and Piwi-Interacting RNA (piRNA) Biogenesis'. *Journal of Biological Chemistry* 292 (15): 6039–46. https://doi.org/10.1074/jbc.C117.777730.

Karavioti, Kyza. 2024. 'Wolbachia Endosymbionts in Drosophila Regulate the Resistance to Zika Virus Infection in a Sex Dependent Manner'. *Frontiers in Microbiology*.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Kassambara, Alboukadel, Marcin Kosinski, Przemyslaw Biecek, and Scheipl Fabian. 2021. *Survminer: Drawing Survival Curves Using 'Ggplot2'*. V. 0.4.9. Released March 9. https://cloud.r-project.org/web/packages/survminer/index.html.

Kelly, Scott A., Tami M. Panhuis, and Andrew M. Stoehr. 2012. 'Phenotypic Plasticity: Molecular Mechanisms and Adaptive Significance'. In *Comprehensive Physiology*, 1st edn, edited by Ronald Terjung. Wiley. https://doi.org/10.1002/cphy.c110008.

Kerkut, G. A., and B. J. R. Taylor. 1958. 'The Effect of Temperature Changes on the Activity of Poikilotherms'. *Behaviour* 13 (3/4): 259–79. https://www.jstor.org/stable/4532903.

Khan, Aziz, and Anthony Mathelier. 2017. 'Intervene: A Tool for Intersection and Visualization of Multiple Gene or Genomic Region Sets'. *BMC Bioinformatics* 18 (1): 287. https://doi.org/10.1186/s12859-017-1708-7.

Khazaeli, Aziz A., Wayne Van Voorhies, and James W. Curtsinger. 2005. 'The Relationship between Life Span and Adult Body Size Is Highly Strain-Specific in Drosophila Melanogaster'. *Experimental Gerontology* 40 (5): 377–85. https://doi.org/10.1016/j.exger.2005.02.004.

Kidwell, Margaret G. 2002. 'Transposable Elements and the Evolution of Genome Size in Eukaryotes'. *Genetica* 115 (1): 49–63. https://doi.org/10.1023/A:1016072014259.

Kikuchi, Yoshitomo, Masahito Hayatsu, Takahiro Hosokawa, Atsushi Nagayama, Kanako Tago, and Takema Fukatsu. 2012. 'Symbiont-Mediated Insecticide Resistance'. *Proceedings of the National Academy of Sciences* 109 (22): 8618–22. https://doi.org/10.1073/pnas.1200231109.

Kim, Adele, and Ilmari Pyykko. 2011. 'Size Matters: Versatile Use of PiggyBac Transposons as a Genetic Manipulation Tool'. *Molecular and Cellular Biochemistry* 354 (1): 301–9. https://doi.org/10.1007/s11010-011-0832-3.

Kim, Keonhee E., Taehwan Jang, and Kwang Pum Lee. 2020. 'Combined Effects of Temperature and Macronutrient Balance on Life-History Traits in Drosophila Melanogaster: Implications for Life-History Trade-Offs and Fundamental Niche'. *Oecologia* 193 (2): 299–309. https://doi.org/10.1007/s00442-020-04666-0.

Kin Chan, Sunny Sun. 2013. 'What Is a Master Regulator?' *Journal of Stem Cell Research & Therapy* 03 (02). https://doi.org/10.4172/2157-7633.1000e114.

King, Allison M., and Thomas H. MacRae. 2015. 'Insect Heat Shock Proteins During Stress and Diapause'. *Annual Review of Entomology* 60 (1): 59–75. https://doi.org/10.1146/annurev-ento-011613-162107.

Kingsolver, Joel G., and H. Arthur Woods. 2016. 'Beyond Thermal Performance Curves: Modeling Time-Dependent Effects of Thermal Stress on Ectotherm Growth Rates'. *The American Naturalist* 187 (3): 283–94. https://doi.org/10.1086/684786.

Klepsatel, Peter, Diana Knoblochová, Thirnahalli Nagaraj Girish, Heinrich Dircksen, and Martina Gáliková. 2020. 'The Influence of Developmental Diet on Reproduction and Metabolism in Drosophila'. *BMC Evolutionary Biology* 20 (July): 93. https://doi.org/10.1186/s12862-020-01663-y.

Klepsatel, Peter, Emanuel Procházka, and Martina Gáliková. 2018. 'Crowding of *Drosophila* Larvae Affects Lifespan and Other Life-History Traits via Reduced Availability of Dietary Yeast'. *Experimental Gerontology* 110 (September): 298–308. https://doi.org/10.1016/j.exger.2018.06.016.

Klepzig, K. D., A. S. Adams, J. Handelsman, and K. F. Raffa. 2009. 'Symbioses: A Key Driver of Insect Physiological Processes, Ecological Interactions, Evolutionary Diversification, and Impacts on Humans*'. *Environmental Entomology* 38 (1): 67–77. https://doi.org/10.1603/022.038.0109.

Kojima, Kenji K., and Jerzy Jurka. 2011. 'Crypton Transposons: Identification of New Diverse Families and Ancient Domestication Events'. *Mobile DNA* 2 (1): 12. https://doi.org/10.1186/1759-8753-2-12.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Kolmogorov, Mikhail, Derek M. Bickhart, Bahar Behsaz, et al. 2020. 'metaFlye: Scalable Long-Read Metagenome Assembly Using Repeat Graphs'. *Nature Methods* 17 (11): 1103–10. https://doi.org/10.1038/s41592-020-00971-x.

Koštál, Vladimír, and Michaela Tollarová-Borovanská. 2009. 'The 70 kDa Heat Shock Protein Assists during the Repair of Chilling Injury in the Insect, Pyrrhocoris Apterus'. *PLOS ONE* 4 (2): e4546. https://doi.org/10.1371/journal.pone.0004546.

Krupovic, Mart, Dennis H. Bamford, and Eugene V. Koonin. 2014. 'Conservation of Major and Minor Jelly-Roll Capsid Proteins in Polinton (Maverick) Transposons Suggests That They Are Bona Fide Viruses'. *Biology Direct* 9 (1): 6. https://doi.org/10.1186/1745-6150-9-6.

Kubiak, Magdalena Regina, and Izabela Makałowska. 2017. 'Protein-Coding Genes' Retrocopies and Their Functions'. *Viruses* 9 (4): 80. https://doi.org/10.3390/v9040080.

Kucharski, R., J. Maleszka, S. Foret, and R. Maleszka. 2008. 'Nutritional Control of Reproductive Status in Honeybees via DNA Methylation'. *Science* 319 (5871): 1827–30. https://doi.org/10.1126/science.1153069.

Lafuente, E., D. Duneau, and P. Beldade. 2024. 'Genetic Basis of Variation in Thermal Developmental Plasticity for Drosophila Melanogaster Body Pigmentation'. *Molecular Ecology* 33 (6): e17294. https://doi.org/10.1111/mec.17294.

Lafuente, Elvira, and Patrícia Beldade. 2019. 'Genomics of Developmental Plasticity in Animals'. *Frontiers in Genetics* 10 (August). https://doi.org/10.3389/fgene.2019.00720.

Lahondère, Chloé. 2023. 'Recent Advances in Insect Thermoregulation'. *Journal of Experimental Biology* 226 (18): jeb245751. https://doi.org/10.1242/jeb.245751.

Lahti, David C., Norman A. Johnson, Beverly C. Ajie, et al. 2009. 'Relaxed Selection in the Wild'. *Trends in Ecology & Evolution* 24 (9): 487–96. https://doi.org/10.1016/j.tree.2009.03.010.

Lanciano, Sophie, and Marie Mirouze. 2018. 'Transposable Elements: All Mobile, All Different, Some Stress Responsive, Some Adaptive?' *Current Opinion in Genetics & Development*, Genome Architecture and Expression, vol. 49 (April): 106–14. https://doi.org/10.1016/j.gde.2018.04.002.

Lander, Eric S., Lauren M. Linton, Bruce Birren, et al. 2001. 'Initial Sequencing and Analysis of the Human Genome'. *Nature* 409 (6822): 860–921. https://doi.org/10.1038/35057062.

Landmann, Frédéric. 2019. 'The Wolbachia Endosymbionts'. *Microbiology Spectrum* 7 (2): 10.1128/microbiolspec.bai-0018–2019. https://doi.org/10.1128/microbiolspec.bai-0018-2019.

Larue, Anaïs, Alexander Mauro, Miriam Merenciano, et al. in prep. 'On the Trail of Transposable Elements: Exploring Their Impact on Phenotypes, Phenotypic Variability, and Plasticity in Drosophila Melanogaster'. Preprint.

Latchman, David S. 1997. 'Transcription Factors: An Overview'. *The International Journal of Biochemistry & Cell Biology* 29 (12): 1305–12. https://doi.org/10.1016/S1357-2725(97)00085-X.

Latzel, Vít, Javier Puy, Michael Thieme, Etienne Bucher, Lars Götzenberger, and Francesco de Bello. 2023. 'Phenotypic Diversity Influenced by a Transposable Element Increases Productivity and Resistance to Competitors in Plant Populations'. *Journal of Ecology* 111 (11): 2376–87. https://doi.org/10.1111/1365-2745.14185.

Lavialle, Christian, Guillaume Cornelis, Anne Dupressoir, et al. 2013. 'Paleovirology of "Syncytins", Retroviral Env Genes Exapted for a Role in Placentation'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1626): 20120507. https://doi.org/10.1098/rstb.2012.0507.

Leblanc, P., S. Desset, B. Dastugue, and C. Vaury. 1997. 'Invertebrate Retroviruses: ZAM a New Candidate in D.Melanogaster'. *The EMBO Journal* 16 (24): 7521–31. https://doi.org/10.1093/emboj/16.24.7521.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Leclerc, Vincent, and Jean-Marc Reichhart. 2004. 'The Immune Response of Drosophila Melanogaster'. *Immunological Reviews* 198 (1): 59–71. https://doi.org/10.1111/j.0105-2896.2004.0130.x.

Ledón-Rettig, Cris C., David W. Pfennig, Amanda J. Chunco, and Ian Dworkin. 2014. 'Cryptic Genetic Variation in Natural Populations: A Predictive Framework'. *Integrative and Comparative Biology* 54 (5): 783–93. https://doi.org/10.1093/icb/icu077.

Lee, Tong Ihn, and Richard A. Young. 2000. 'TRANSCRIPTION OF EUKARYOTIC PROTEIN-CODING GENES'. *Annual Review of Genetics* 34 (Volume 34, 2000): 77–137. https://doi.org/10.1146/annurev.genet.34.1.77.

Lee, Yuh Chwen G, and Gary H Karpen. 2017. 'Pervasive Epigenetic Effects of Drosophila Euchromatic Transposable Elements Impact Their Evolution'. *eLife* 6 (July): e25762. https://doi.org/10.7554/eLife.25762.

Lefranc, Agnès, and Jørgen Bundgaard. 2000. 'The Influence of Male and Female Body Size on Copulation Duration and Fecundity in Drosophila Melanogaster'. *Hereditas* 132 (3): 243–47. https://doi.org/10.1111/j.1601-5223.2000.00243.x.

Lehmann, Fritz-Olaf, and Peter Schützner. 2010. 'The Respiratory Basis of Locomotion in Drosophila'. *Journal of Insect Physiology* 56 (5): 543–50. https://doi.org/10.1016/j.jinsphys.2009.04.019.

Lema, Sean C. 2014. 'Hormones and Phenotypic Plasticity in an Ecological Context: Linking Physiological Mechanisms to Evolutionary Processes'. *Integrative and Comparative Biology* 54 (5): 850–63. https://doi.org/10.1093/icb/icu019.

Lema, Sean C. 2020. 'The Adaptive Value of Hormones: Endocrine Systems as Outcomes and Initiators of Evolution'. *Molecular and Cellular Endocrinology* 517 (November): 110983. https://doi.org/10.1016/j.mce.2020.110983.

Lema, Sean C., and Jun Kitano. 2013. 'Hormones and Phenotypic Plasticity: Implications for the Evolution of Integrated Adaptive Phenotypes'. *Current Zoology* 59 (4): 506–25. https://doi.org/10.1093/czoolo/59.4.506.

Leung, T. L. F., and R. Poulin. 2008. 'PARASITISM, COMMENSALISM, AND MUTUALISM: EXPLORING THE MANY SHADES OF SYMBIOSES'. *Vie et Milieu / Life & Environment*, 107. https://hal.sorbonne-universite.fr/hal-03246057.

Levis, Nicholas A., and David W. Pfennig. 2019a. 'Plasticity-Led Evolution: Evaluating the Key Prediction of Frequency-Dependent Adaptation'. *Proceedings of the Royal Society B: Biological Sciences* 286 (1897): 20182754. https://doi.org/10.1098/rspb.2018.2754.

Levis, Nicholas A., and David W. Pfennig. 2019b. 'Plasticity-Led Evolution: Evaluating the Key Prediction of Frequency-Dependent Adaptation'. *Proceedings of the Royal Society B: Biological Sciences* 286 (1897): 20182754. https://doi.org/10.1098/rspb.2018.2754.

Lewontin, Richard C. 1974. *The Genetic Basis of Evolutionary Change*. With Internet Archive. New York, Columbia University Press. http://archive.org/details/geneticbasisofev00lewo.

Lewontin, Richard C. 2000. 'The Triple Helix: Gene, Organism, and Environment'. https://philpapers.org/rec/LEWTTH.

Li, Heng. 2021. 'New Strategies to Improve Minimap2 Alignment Accuracy'. *Bioinformatics* 37 (23): 4572–74. https://doi.org/10.1093/bioinformatics/btab705.

Lima, Fernando P., Filipa Gomes, Rui Seabra, et al. 2016. 'Loss of Thermal Refugia near Equatorial Range Limits'. *Global Change Biology* 22 (1): 254–63. https://doi.org/10.1111/gcb.13115.

Lin, Yanzhu, Zhen-Xia Chen, Brian Oliver, and Susan T Harbison. 2016. 'Microenvironmental Gene Expression Plasticity Among Individual Drosophila Melanogaster'. *G3 Genes|Genomes|Genetics* 6 (12): 4197–210. https://doi.org/10.1534/g3.116.035444.

Lindquist, Susan. 1986. *Department of Molecular Genetics and Cell Biology, The University of Chicago,*.

Lippman, Zachary, Anne-Valérie Gendrel, Michael Black, et al. 2004. 'Role of Transposable Elements in

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Heterochromatin and Epigenetic Control'. *Nature* 430 (6998): 6998. https://doi.org/10.1038/nature02651.

Little, Alexander G., and Frank Seebacher. 2024. 'Endocrine Responses to Environmental Variation'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 379 (1898): 20220515. https://doi.org/10.1098/rstb.2022.0515.

Liu, Ruiqi, Jiajun Wu, Haiwei Guo, et al. 2023. 'Post-Translational Modifications of Histones: Mechanisms, Biological Functions, and Therapeutic Targets'. *MedComm* 4 (3): e292. https://doi.org/10.1002/mco2.292.

Long, Jacob A. 2024. 'Jtools: Analysis and Presentation of Social Scientific Data'. *Journal of Open Source Software* 9 (101): 6610. https://doi.org/10.21105/joss.06610.

López-Madrigal, Sergio, and Elves H Duarte. 2019. 'Titer Regulation in Arthropod-Wolbachia Symbioses'. *FEMS Microbiology Letters* 366 (23): fnz232. https://doi.org/10.1093/femsle/fnz232.

Lu, Thomas X., and Marc E. Rothenberg. 2018. 'MicroRNA'. *Journal of Allergy and Clinical Immunology* 141 (4): 1202–7. https://doi.org/10.1016/j.jaci.2017.08.034.

Luan, Dongmei D., Malka H. Korman, John L. Jakubczak, and Thomas H. Eickbush. 1993. 'Reverse Transcription of R2Bm RNA Is Primed by a Nick at the Chromosomal Target Site: A Mechanism for Non-LTR Retrotransposition'. *Cell* 72 (4): 595–605. https://doi.org/10.1016/0092-8674(93)90078-5.

Lushchak, Oleh, Olha Strilbytska, and Kenneth B. Storey. 2023. 'Gender-Specific Effects of pro-Longevity Interventions in *Drosophila*'. *Mechanisms of Ageing and Development* 209 (January): 111754. https://doi.org/10.1016/j.mad.2022.111754.

Lyko, F., C. Beisel, J. Marhold, and R. Paro. 2006. 'Epigenetic Regulation in Drosophila'. In *DNA Methylation: Development, Genetic Disease and Cancer*, edited by Walter Doerfler and Petra Böhm, vol. 310. Current Topics in Microbiology and Immunology. Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-31181-5_3.

Lyko, Frank, Sylvain Foret, Robert Kucharski, Stephan Wolf, Cassandra Falckenhayn, and Ryszard Maleszka. 2010. 'The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers'. *PLoS Biology* 8 (11): e1000506. https://doi.org/10.1371/journal.pbio.1000506.

Lyons, Pierre J., Daneck Lang-Ouellette, and Pier Jr Morin. 2013. 'CryomiRs: Towards the Identification of a Cold-Associated Family of microRNAs'. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 8 (4): 358–64. https://doi.org/10.1016/j.cbd.2013.10.001.

Mackay, Trudy F. C. 1986. 'Transposable Element-Induced Fitness Mutations in Drosophila Melanogaster'. *Genetics Research* 48 (2): 77–87. https://doi.org/10.1017/S0016672300024794.

Mackay, Trudy F. C., and Robert R. H. Anholt. 2024. 'Pleiotropy, Epistasis and the Genetic Architecture of Quantitative Traits'. *Nature Reviews Genetics* 25 (9): 639–57. https://doi.org/10.1038/s41576-024-00711-3.

Mackay, Trudy F. C., Stephen Richards, Eric A. Stone, et al. 2012. 'The Drosophila Melanogaster Genetic Reference Panel'. *Nature* 482 (7384): 173–78. https://doi.org/10.1038/nature10811.

Mackay, Trudy F. C., Eric A. Stone, and Julien F. Ayroles. 2009. 'The Genetics of Quantitative Traits: Challenges and Prospects'. *Nature Reviews Genetics* 10 (8): 565–77. https://doi.org/10.1038/nrg2612.

Mackay, Trudy F.C, and Richard F Lyman. 2005. 'Drosophila Bristles and the Nature of Quantitative Genetic Variation'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1459): 1513–27. https://doi.org/10.1098/rstb.2005.1672.

Mahat, Dig B., H. Hans Salamanca, Fabiana M. Duarte, Charles G. Danko, and John T. Lis. 2016. 'Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-Wide Transcriptional Regulation'. *Molecular Cell* 62 (1): 63–78. https://doi.org/10.1016/j.molcel.2016.02.025.

Majumdar, Sharmistha, and Donald C. Rio. 2015. 'P Transposable Elements in Drosophila and Other

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Eukaryotic Organisms'. *Microbiology Spectrum* 3 (2): 10.1128/microbiolspec.mdna3-0004–2014. https://doi.org/10.1128/microbiolspec.mdna3-0004-2014.

Maleszka, Ryszard. 2008. 'Epigenetic Integration of Environmental and Genomic Signals in Honey Bees: The Critical Interplay of Nutritional, Brain and Reproductive Networks'. *Epigenetics* 3 (4): 188–92. https://doi.org/10.4161/epi.3.4.6697.

Maleszka, Ryszard. 2024. 'Reminiscences on the Honeybee Genome Project and the Rise of Epigenetic Concepts in Insect Science'. *Insect Molecular Biology* 33 (5): 444–56. https://doi.org/10.1111/imb.12888.

Manni, Mosè, Matthew R Berkeley, Mathieu Seppey, Felipe A Simão, and Evgeny M Zdobnov. 2021. 'BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes'. *Molecular Biology and Evolution* 38 (10): 4647–54. https://doi.org/10.1093/molbev/msab199.

Margulis. 1938. 'Symbiosis in Cell Evolution : Life and Its Environment on the Early Earth : Margulis, Lynn, 1938- : Free Download, Borrow, and Streaming : Internet Archive'. https://archive.org/details/symbiosisincelle00marg/page/206/mode/2up.

Margulis, Lynn, and René Fester. 1991. *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis*. MIT Press.

Marie, Léa, and Lorraine S. Symington. 2022. 'Mechanism for Inverted-Repeat Recombination Induced by a Replication Fork Barrier'. *Nature Communications* 13 (1): 32. https://doi.org/10.1038/s41467-021-27443-w.

Marin, Pierre, Julien Genitoni, Dominique Barloy, et al. 2020. 'Biological Invasion: The Influence of the Hidden Side of the (Epi)Genome'. *Functional Ecology* 34 (2): 385–400. https://doi.org/10.1111/1365-2435.13317.

Marmorstein, Ronen. 2001. 'Protein Modules That Manipulate Histone Tails for Chromatin Regulation'. *Nature Reviews Molecular Cell Biology* 2 (6): 422–32. https://doi.org/10.1038/35073047.

Martienssen, Robert A, and Eric J Richards. 1995. 'DNA Methylation in Eukaryotes'. *Current Opinion in Genetics & Development* 5 (2): 234–42. https://doi.org/10.1016/0959-437X(95)80014-X.

Martinez, Julien, Ignacio Tolosana, Suzan Ok, et al. 2017. 'Symbiont Strain Is the Main Determinant of Variation in Wolbachia-Mediated Protection against Viruses across Drosophila Species'. *Molecular Ecology* 26 (15): 4072–84. https://doi.org/10.1111/mec.14164.

Masel, Joanna. 2006. 'Cryptic Genetic Variation Is Enriched for Potential Adaptations'. *Genetics* 172 (3): 1985–91. https://doi.org/10.1534/genetics.105.051649.

Masel, Joanna. 2013. 'Q&A: Evolutionary Capacitance'. *BMC Biology* 11 (1): 103. https://doi.org/10.1186/1741-7007-11-103.

Masel, Joanna, and Meredith V. Trotter. 2010. 'Robustness and Evolvability'. *Trends in Genetics* 26 (9): 406–14. https://doi.org/10.1016/j.tig.2010.06.002.

Mateo, Lidia, Anna Ullastres, and Josefa González. 2014. 'A Transposable Element Insertion Confers Xenobiotic Resistance in Drosophila'. *PLOS Genetics* 10 (8): e1004560. https://doi.org/10.1371/journal.pgen.1004560.

Mathis, Kaitlyn A., and Judith L. Bronstein. 2020. 'Our Current Understanding of Commensalism'. *Annual Review of Ecology, Evolution, and Systematics* 51 (Volume 51, 2020): 167–89. https://doi.org/10.1146/annurev-ecolsys-011720-040844.

Mattei, Alexandra L., Nina Bailly, and Alexander Meissner. 2022. 'DNA Methylation: A Historical Perspective'. *Trends in Genetics* 38 (7): 676–707. https://doi.org/10.1016/j.tig.2022.03.010.

Mayeux, Camille A., Anaïs Larue, Daniel S. Oliveira, et al. 2024. 'Intricate Interactions between Antiviral Immunity and Transposable Element Control in

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Drosophila'. Preprint, bioRxiv, March 20. https://doi.org/10.1101/2024.03.18.585529.

Mayoral, Jaime G., Kayvan Etebari, Mazhar Hussain, Alexander A. Khromykh, and Sassan Asgari. 2014. 'Wolbachia Infection Modifies the Profile, Shuttling and Structure of MicroRNAs in a Mosquito Cell Line'. *PLoS ONE* 9 (4): e96107. https://doi.org/10.1371/journal.pone.0096107.

McClintock, Barbara. 1950. 'The Origin and Behavior of Mutable Loci in Maize'. *Proceedings of the National Academy of Sciences* 36 (6): 344–55. https://doi.org/10.1073/pnas.36.6.344.

McClintock, Barbara. 1953. 'Induction of Instability at Selected Loci in Maize'. *Genetics* 38 (6): 579. https://doi.org/10.1093/genetics/38.6.579.

McClintock, Barbara. 1956. 'Controlling Elements and the Gene'. *Cold Spring Harbor Symposia on Quantitative Biology* 21 (January): 197–216. https://doi.org/10.1101/SQB.1956.021.01.017.

McDonald, Megan C., Adam P. Taranto, Erin Hill, et al. 2019. 'Transposon-Mediated Horizontal Transfer of the Host-Specific Virulence Protein ToxA between Three Fungal Wheat Pathogens'. *mBio* 10 (5): 10.1128/mbio.01515-19. https://doi.org/10.1128/mbio.01515-19.

McLean, Ailsa H. C., Benjamin J. Parker, Jan Hrček, James C. Kavanagh, Peter A. D. Wellham, and H. Charles J. Godfray. 2018. 'Consequences of Symbiont Co-Infections for Insect Host Phenotypes'. *Journal of Animal Ecology* 87 (2): 478–88. https://doi.org/10.1111/1365-2656.12705.

Merçot, Herve, and Sylvain Charlat. 2004. *Wolbachia Infections in Drosophila Melanogaster and D. Simulans: Polymorphism and Levels of Cytoplasmic Incompatibility*.

Mérel, Vincent, Matthieu Boulesteix, Marie Fablet, and Cristina Vieira. 2020. 'Transposable Elements in Drosophila'. *Mobile DNA* 11 (1): 23. https://doi.org/10.1186/s13100-020-00213-z.

Mérel, Vincent, Patricia Gibert, Inessa Buch, et al. 2021. 'The Worldwide Invasion of Drosophila Suzukii Is Accompanied by a Large Increase of Transposable Element Load and a Small Number of Putatively Adaptive Insertions'. *Molecular Biology and Evolution* 38 (10): 4252–67. https://doi.org/10.1093/molbev/msab155.

Merenciano, Miriam, and Josefa González. 2023. 'The Interplay Between Developmental Stage and Environment Underlies the Adaptive Effect of a Natural Transposable Element Insertion'. *Molecular Biology and Evolution* 40 (3): msad044. https://doi.org/10.1093/molbev/msad044.

Merenciano, Miriam, Anaïs Larue, Chloé Garambois, William Vilas Boas Nunes, and Cristina Vieira. 2025. 'Exploring the Relationship of Transposable Elements and Ageing: Causes and Consequences'. *Genome Biology and Evolution* 17 (6): evaf088. https://doi.org/10.1093/gbe/evaf088.

Merenciano, Miriam, Anaïs Larue, Cristian Groza, Cristina Vieira, Rita Rebollo, and Clément Goubert. 2024. 'Chapter 6 - Epigenetics and Genotypic Variation: A Transposable Elements' Perspective'. In *On Epigenetics and Evolution*, edited by Carlos M. Guerrero-Bosagna. Translational Epigenetics. Academic Press. https://doi.org/10.1016/B978-0-443-19051-3.00006-1.

Michaud, M. Robert, Nicholas M. Teets, Justin T. Peyton, Brandon M. Blobner, and David L. Denlinger. 2011. 'Heat Shock Response to Hypoxia and Its Attenuation during Recovery in the Flesh Fly, *Sarcophaga Crassipalpis*'. *Journal of Insect Physiology* 57 (1): 203–10. https://doi.org/10.1016/j.jinsphys.2010.11.007.

Mikheenko, Alla, Andrey Prjibelski, Vladislav Saveliev, Dmitry Antipov, and Alexey Gurevich. 2018. 'Versatile Genome Assembly Evaluation with QUAST-LG'. *Bioinformatics* 34 (13): i142–50. https://doi.org/10.1093/bioinformatics/bty266.

Miller, Richard S., and J. L. Thomas. 1958. 'The Effects of Larval Crowding and Body Size on the Longevity of Adult Drosophila Melanogaster'. *Ecology* 39 (1): 118–25. https://doi.org/10.2307/1929973.

Minchiotti, Gabriella, Cristina Contursi, Franco Graziani, Giuseppe Gargiulo, and Pier Paolo Di Nocera. 1994. 'Expression of Drosophila

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Melanogaster F Elements in Vivo'. *Molecular and General Genetics MGG* 245 (2): 152–59. https://doi.org/10.1007/BF00283262.

Miura, Toru. 2019. 'Juvenile Hormone as a Physiological Regulator Mediating Phenotypic Plasticity in Pancrustaceans'. *Development, Growth & Differentiation* 61 (1): 85–96. https://doi.org/10.1111/dgd.12572.

Miyoshi, Keita, Tomohiro Miyoshi, Julia Verena Hartig, Haruhiko Siomi, and Mikiko C. Siomi. 2010. 'Molecular Mechanisms That Funnel RNA Precursors into Endogenous Small-Interfering RNA and microRNA Biogenesis Pathways in Drosophila'. *RNA* 16 (3): 506–15. https://doi.org/10.1261/rna.1952110.

Modenini, Giorgia, Paolo Abondio, and Alessio Boattini. 2022. 'The Coevolution between APOBEC3 and Retrotransposons in Primates'. *Mobile DNA* 13 (1): 27. https://doi.org/10.1186/s13100-022-00283-1.

Mohamed, Mourdas, Nguyet Thi-Minh Dang, Yuki Ogyama, et al. 2020. 'A Transposon Story: From TE Content to TE Dynamic Invasion of Drosophila Genomes Using the Single-Molecule Sequencing Technology from Oxford Nanopore'. *Cells* 9 (8): 8. https://doi.org/10.3390/cells9081776.

Mohamed, Mourdas, François Sabot, Marion Varoqui, et al. 2023. 'TrEMOLO: Accurate Transposable Element Allele Frequency Estimation Using Long-Read Sequencing Data Combining Assembly and Mapping-Based Approaches'. *Genome Biology* 24 (1): 63. https://doi.org/10.1186/s13059-023-02911-2.

Monnin, D., N. Kremer, C. Berny, et al. 2016. 'Influence of Oxidative Homeostasis on Bacterial Density and Cost of Infection in Drosophila–Wolbachia Symbioses'. *Journal of Evolutionary Biology* 29 (6): 1211–22. https://doi.org/10.1111/jeb.12863.

Monnin, David, Natacha Kremer, Caroline Michaud, et al. 2021. 'Experimental evolution of virulence and associated traits in a *Drosophila melanogaster*–*Wolbachia* symbiosis'. *Peer Community Journal* 1. https://doi.org/10.24072/pcjournal.9.

Montgomery, Sara L., Daria Vorojeikina, Wen Huang, Trudy F. C. Mackay, Robert R. H. Anholt, and Matthew D. Rand. 2014. 'Genome-Wide Association Analysis of Tolerance to Methylmercury Toxicity in Drosophila Implicates Myogenic and Neuromuscular Developmental Pathways'. *PLOS ONE* 9 (10): e110375. https://doi.org/10.1371/journal.pone.0110375.

Moore, Lisa D., Thuc Le, and Guoping Fan. 2013. 'DNA Methylation and Its Basic Function'. *Neuropsychopharmacology* 38 (1): 23–38. https://doi.org/10.1038/npp.2012.112.

Moran, Nancy A. 2006. 'Symbiosis'. *Current Biology* 16 (20): R866–71. https://doi.org/10.1016/j.cub.2006.09.019.

Morgante, Fabio, Peter Sørensen, Daniel A. Sorensen, Christian Maltecca, and Trudy F. C. Mackay. 2015. 'Genetic Architecture of Micro-Environmental Plasticity in Drosophila Melanogaster'. *Scientific Reports* 5 (May): 9785. https://doi.org/10.1038/srep09785.

Morris, J. Jeffrey. 2018. 'What Is the Hologenome Concept of Evolution?' *F1000Research* 7 (October): 1664. https://doi.org/10.12688/f1000research.14385.1.

Moschetti, Roberta, Antonio Palazzo, Patrizio Lorusso, Luigi Viggiano, and René Massimiliano Marsano. 2020. '"What You Need, Baby, I Got It": Transposable Elements as Suppliers of Cis-Operating Sequences in Drosophila'. *Biology* 9 (2): 2. https://doi.org/10.3390/biology9020025.

Mostoufi, Sabrina L, and Nadia D Singh. 2022. 'Diet-Induced Changes in Titer Support a Discrete Response of Wolbachia-Associated Plastic Recombination in Drosophila Melanogaster'. *G3 Genes|Genomes|Genetics* 12 (1): jkab375. https://doi.org/10.1093/g3journal/jkab375.

Mostoufi, Sabrina L, and Nadia D Singh. 2024. 'Pathogen Infection Alters the Gene Expression Landscape of Transposable Elements in Drosophila

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Melanogaster'. *G3 Genes|Genomes|Genetics*, August 12, jkae171. https://doi.org/10.1093/g3journal/jkae171.

Muller, H. J. 1930. 'Types of Visible Variations Induced by X-Rays inDrosophila'. *Journal of Genetics* 22 (3): 299–334. https://doi.org/10.1007/BF02984195.

Murphy, Patrick J., and Frédéric Berger. 2023. 'The Chromatin Source-Sink Hypothesis: A Shared Mode of Chromatin-Mediated Regulations'. *Development* 150 (21): dev201989. https://doi.org/10.1242/dev.201989.

Murray, Kevin O., Thomas L. Clanton, and Michal Horowitz. 2022. 'Epigenetic Responses to Heat: From Adaptation to Maladaptation'. *Experimental Physiology* 107 (10): 1144–58. https://doi.org/10.1113/EP090143.

Mustafin, R. N. 2018. 'Hypothesis on the Origin of Viruses from Transposons'. *Molecular Genetics, Microbiology and Virology* 33 (4): 223–32. https://doi.org/10.3103/S0891416818040067.

Nässel, Dick R. 2012. 'Insulin-Producing Cells and Their Regulation in Physiology and Behavior of Drosophila1This Review Is Part of a Virtual Symposium on Recent Advances in Understanding a Variety of Complex Regulatory Processes in Insect Physiology and Endocrinology, Including Development, Metabolism, Cold Hardiness, Food Intake and Digestion, and Diuresis, through the Use of Omics Technologies in the Postgenomic Era.' *Canadian Journal of Zoology* 90 (4): 476–88. https://doi.org/10.1139/z2012-009.

Negi, Pooja, Archana N. Rai, and Penna Suprasanna. 2016. 'Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response'. *Frontiers in Plant Science* 7 (October). https://doi.org/10.3389/fpls.2016.01448.

Nelson, R. J. 2010. 'Hormones and Behavior: Basic Concepts'. In *Encyclopedia of Animal Behavior*, edited by Michael D. Breed and Janice Moore. Academic Press. https://doi.org/10.1016/B978-0-08-045337-8.00236-9.

Nesmelova, Irina V., and Perry B. Hackett. 2010. 'DDE Transposases: Structural Similarity and Diversity'. *Advanced Drug Delivery Reviews*, Stem Cell Gene Manipulation and Delivery as Systemic Therapeutics, vol. 62 (12): 1187–95. https://doi.org/10.1016/j.addr.2010.06.006.

Nijhout, H. Frederik. 2003. 'Development and Evolution of Adaptive Polyphenisms'. *Evolution & Development* 5 (1): 9–18. https://doi.org/10.1046/j.1525-142X.2003.03003.x.

Ninova, Maria, Yung-Chia Ariel Chen, Baira Godneeva, et al. 2020. 'Su(Var)2-10 and the SUMO Pathway Link piRNA-Guided Target Recognition to Chromatin Silencing'. *Molecular Cell* 77 (3): 556-570.e6. https://doi.org/10.1016/j.molcel.2019.11.012.

Nishihara, Hidenori. 2019. 'Transposable Elements as Genetic Accelerators of Evolution: Contribution to Genome Size, Gene Regulatory Network Rewiring and Morphological Innovation'. *Genes & Genetic Systems* 94 (6): 269–81. https://doi.org/10.1266/ggs.19-00029.

Nishimasu, Hiroshi, Hirotsugu Ishizu, Kuniaki Saito, et al. 2012. 'Structure and Function of Zucchini Endoribonuclease in piRNA Biogenesis'. *Nature* 491 (7423): 284–87. https://doi.org/10.1038/nature11509.

Niu, Xiao-Min, Yong-Chao Xu, Zi-Wen Li, et al. 2019. 'Transposable Elements Drive Rapid Phenotypic Variation in Capsella Rubella'. *Proceedings of the National Academy of Sciences* 116 (14): 6908–13. https://doi.org/10.1073/pnas.1811498116.

Noble, Denis. 2015. 'Conrad Waddington and the Origin of Epigenetics'. *Journal of Experimental Biology* 218 (6): 816–18. https://doi.org/10.1242/jeb.120071.

Okamura, Katsutomo, Wei-Jen Chung, J. Graham Ruby, Huili Guo, David P. Bartel, and Eric C. Lai. 2008. 'The Drosophila Hairpin RNA Pathway Generates Endogenous Short Interfering RNAs'. *Nature* 453 (7196): 803–6. https://doi.org/10.1038/nature07015.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Okonechnikov, Konstantin, Ana Conesa, and Fernando García-Alcalde. 2016. 'Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data'. *Bioinformatics* 32 (2): 292–94. https://doi.org/10.1093/bioinformatics/btv566.

Oleson, Bryndon J., Daphne Bazopoulou, and Ursula Jakob. 2021. 'Shaping Longevity Early in Life: Developmental ROS and H3K4me3 Set the Clock'. *Cell Cycle* 20 (22): 2337–47. https://doi.org/10.1080/15384101.2021.1986317.

Oliveira, Daniel S, Marie Fablet, Anaïs Larue, et al. 2023. 'ChimeraTE: A Pipeline to Detect Chimeric Transcripts Derived from Genes and Transposable Elements'. *Nucleic Acids Research* 51 (18): 9764–84. https://doi.org/10.1093/nar/gkad671.

Oliver, Kerry M, and Clesson HV Higashi. 2019. 'Variations on a Protective Theme: *Hamiltonella Defensa* Infections in Aphids Variably Impact Parasitoid Success'. *Current Opinion in Insect Science*, Ecology • Parasites/Parasitoids/Biological control, vol. 32 (April): 1–7. https://doi.org/10.1016/j.cois.2018.08.009.

O'Malley, Bert W. 2007. 'Coregulators: From Whence Came These "Master Genes"'. *Molecular Endocrinology* 21 (5): 1009–13. https://doi.org/10.1210/me.2007-0012.

Onishi, Ryo, Soichiro Yamanaka, and Mikiko C Siomi. 2021. 'piRNA- and siRNA-mediated Transcriptional Repression in *Drosophila* , Mice, and Yeast: New Insights and Biodiversity'. *EMBO Reports* 22 (10): e53062. https://doi.org/10.15252/embr.202153062.

Paaby, Annalise B., and Matthew V. Rockman. 2014. 'Cryptic Genetic Variation: Evolution's Hidden Substrate'. *Nature Reviews Genetics* 15 (4): 247–58. https://doi.org/10.1038/nrg3688.

Palacio-López, Kattia, Brian Beckage, Samuel Scheiner, and Jane Molofsky. 2015. 'The Ubiquity of Phenotypic Plasticity in Plants: A Synthesis'. *Ecology and Evolution* 5 (16): 3389–400. https://doi.org/10.1002/ece3.1603.

Palazzo, Alexander F., Yi Qiu, and Yoon Mo Kang. 2024. 'mRNA Nuclear Export: How mRNA Identity Features Distinguish Functional RNAs from Junk Transcripts'. *RNA Biology* 21 (1): 1–12. https://doi.org/10.1080/15476286.2023.2293339.

Paracer, Surindar, and Vernon Ahmadjian. 2000. *Symbiosis: An Introduction to Biological Associations*. Oxford University Press. https://books.google.no/books?hl=en&lr=&id=Z3Y8DwAAQBAJ&oi=fnd&pg=PR9&ots=FO1MllJkEa&sig=hw9f6ZnuWUgJ7IryAMQgMhphLbo&redir_esc=y#v=onepage&q&f=false.

Pardue, Mary-Lou, and P.G. DeBaryshe. 2008. 'Drosophila Telomeres: A Variation on the Telomerase Theme'. *Fly* 2 (3): 101–10. https://doi.org/10.4161/fly.6393.

Park, Ah Rume, Na Liu, Nils Neuenkirchen, Qiaozhi Guo, and Haifan Lin. 2019. 'The Role of Maternal HP1a in Early Drosophila Embryogenesis via Regulation of Maternal Transcript Production'. *Genetics* 211 (1): 201–17. https://doi.org/10.1534/genetics.118.301704.

Pasyukova, E. G., S. V. Nuzhdin, T. V. Morozova, and T. F. C. Mackay. 2004. 'Accumulation of Transposable Elements in the Genome of Drosophila Melanogaster Is Associated with a Decrease in Fitness'. *Journal of Heredity* 95 (4): 284–90. https://doi.org/10.1093/jhered/esh050.

Pearl, Raymond. 1928. *The Rate Of Living*. With Osmania University and Digital Library Of India. By Alfred A.Knopf Inc. http://archive.org/details/rateofliving031726mbp.

Peng, Yu, John E. Nielsen, J. Paul Cunningham, and Elizabeth A. McGraw. 2008. 'Wolbachia Infection Alters Olfactory-Cued Locomotion in Drosophila Spp'. *Applied and Environmental Microbiology* 74 (13): 3943–48. https://doi.org/10.1128/AEM.02607-07.

Peng, Yu, and YuFeng Wang. 2009. 'Infection of Wolbachia May Improve the Olfactory Response of Drosophila'. *Chinese Science Bulletin* 54 (8): 1369–75. https://doi.org/10.1007/s11434-009-0183-6.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Perez, Rémy, and Serge Aron. 2020. 'Adaptations to Thermal Stress in Social Insects: Recent Advances and Future Directions'. *Biological Reviews* 95 (6): 1535–53. https://doi.org/10.1111/brv.12628.

Peric-Mataruga, Vesna, Vera Nenadovic, and Jelisaveta Ivanovic. 2006. 'Neurohormones in Insect Stress: A Review'. *Archives of Biological Sciences* 58 (1): 1–12. https://doi.org/10.2298/ABS0601006P.

Peterson, Craig L., and Marc-André Laniel. 2004. 'Histones and Histone Modifications'. *Current Biology* 14 (14): R546–51. https://doi.org/10.1016/j.cub.2004.07.007.

Pfaffl, Michael W. 2001. 'A New Mathematical Model for Relative Quantification in Real-Time RT–PCR'. *Nucleic Acids Research* 29 (9): e45. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC55695/.

Pfennig, David W. 2021a. 'Key Questions about Phenotypic Plasticity'. In *Phenotypic Plasticity & Evolution*, 1st edn, by David W. Pfennig. CRC Press. https://doi.org/10.1201/9780429343001-4.

Pfennig, David W. 2021b. *Phenotypic Plasticity & Evolution: Causes, Consequences, Controversies*. 1st edn. CRC Press. https://doi.org/10.1201/9780429343001.

Pfennig, David W. 2021c. *Phenotypic Plasticity & Evolution: Causes, Consequences, Controversies*. 1st edn. CRC Press. https://doi.org/10.1201/9780429343001.

Piacentini, Lucia, Laura Fanti, Valeria Specchia, et al. 2014. 'Transposons, Environmental Changes, and Heritable Induced Phenotypic Variability'. *Chromosoma* 123 (4): 345–54. https://doi.org/10.1007/s00412-014-0464-y.

Pigliucci, Massimo. 2001. *Phenotypic Plasticity: Beyond Nature and Nurture*. JHU Press.

Pigliucci, Massimo, and Karen Hayden. 2001. 'Phenotypic Plasticity Is the Major Determinant of Changes in Phenotypic Integration in Arabidopsis'. *New Phytologist* 152 (3): 419–30. https://doi.org/10.1046/j.0028-646X.2001.00275.x.

Pimentel, André C., Cássia S. Cesar, Marcos Martins, and Rodrigo Cogni. 2021. 'The Antiviral Effects of the Symbiont Bacteria Wolbachia in Insects'. *Frontiers in Immunology* 11 (January): 626329. https://doi.org/10.3389/fimmu.2020.626329.

Pimpinelli, Sergio, and Lucia Piacentini. 2020. 'Environmental Change and the Evolution of Genomes: Transposable Elements as Translators of Phenotypic Plasticity into Genotypic Variability'. *Functional Ecology* 34 (2): 428–41. https://doi.org/10.1111/1365-2435.13497.

Piper, Matthew D. W., Danielle Skorupa, and Linda Partridge. 2005. 'Diet, Metabolism and Lifespan in Drosophila'. *Experimental Gerontology*, Metabolism, Aging and Longevity, vol. 40 (11): 857–62. https://doi.org/10.1016/j.exger.2005.06.013.

Pippadpally, Srikanth, and Thejaswini Venkatesh. 2020. 'Deciphering piRNA Biogenesis through Cytoplasmic Granules, Mitochondria and Exosomes'. *Archives of Biochemistry and Biophysics* 695 (November): 108597. https://doi.org/10.1016/j.abb.2020.108597.

Planelló, R., J. L. Martínez-Guitarte, and G. Morcillo. 2008. 'The Endocrine Disruptor Bisphenol A Increases the Expression of HSP70 and Ecdysone Receptor Genes in the Aquatic Larvae of *Chironomus Riparius*'. *Chemosphere* 71 (10): 1870–76. https://doi.org/10.1016/j.chemosphere.2008.01.033.

Planelló, R., J. L. Martínez-Guitarte, and G. Morcillo. 2010. 'Effect of Acute Exposure to Cadmium on the Expression of Heat-Shock and Hormone-Nuclear Receptor Genes in the Aquatic Midge *Chironomus Riparius*'. *Science of The Total Environment* 408 (7): 1598–603. https://doi.org/10.1016/j.scitotenv.2010.01.004.

Platt, Roy N., Michael W. Vandewege, and David A. Ray. 2018. 'Mammalian Transposable Elements and Their Impacts on Genome Evolution'. *Chromosome Research* 26 (1): 25–43. https://doi.org/10.1007/s10577-017-9570-z.

Poças, Gonçalo M., Alexander E. Crosbie, and Christen K. Mirth. 2022. 'When Does Diet Matter? The Roles of

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Larval and Adult Nutrition in Regulating Adult Size Traits in *Drosophila Melanogaster*'. *Journal of Insect Physiology* 139 (May): 104051. https://doi.org/10.1016/j.jinsphys.2020.104051.

Polsinelli, Gregory A., and Hongwei D. Yu. 2018. 'Regulation of Histone Deacetylase 3 by Metal Cations and 10-Hydroxy-2E-Decenoic Acid: Possible Epigenetic Mechanisms of Queen-Worker Bee Differentiation'. *PLOS ONE* 13 (12): e0204538. https://doi.org/10.1371/journal.pone.0204538.

Poulain, Marius, Elodie Rosinski, Hélène Henri, et al. 2024. 'Development, Feeding, and Sex Shape the Relative Quantity of the Nutritional Obligatory Symbiont Wolbachia in Bed Bugs'. *Frontiers in Microbiology* 15 (May). https://doi.org/10.3389/fmicb.2024.1386458.

Poulin, Robert, and Haseeb S. Randhawa. 2015. 'Evolution of Parasitism along Convergent Lines: From Ecology to Genomics'. *Parasitology* 142 (S1): S6–15. https://doi.org/10.1017/S0031182013001674.

Poulter, Russell T. M., and Margi I Butler. 2015. 'Tyrosine Recombinase Retrotransposons and Transposons'. In *Mobile DNA III*. John Wiley & Sons, Ltd. https://doi.org/10.1128/9781555819217.ch55.

Pozzi, Carlo M., Angelo Gaiti, and Alberto Spada. 2025. 'Climate Change and Plant Genomic Plasticity'. *Theoretical and Applied Genetics* 138 (9): 231. https://doi.org/10.1007/s00122-025-05010-x.

Pressman, Abe, Celia Blanco, and Irene A. Chen. 2015. 'The RNA World as a Model System to Study the Origin of Life'. *Current Biology* 25 (19): R953–63. https://doi.org/10.1016/j.cub.2015.06.016.

Pulver, Stefan, and Jimena Berni. 2012. 'The Fundamentals of Flying: Simple and Inexpensive Strategies for Employing Drosophila Genetics in Neuroscience Teaching Laboratories'. *Journal of Undergraduate Neuroscience Education : JUNE : A Publication of FUN, Faculty for Undergraduate Neuroscience* 11 (October): A139–48.

Pyatkov, K. I., H. S. Zelentsova, and M. B. Evgen'ev. 2004. 'Endonuclease Activity Determined by the Penelope Retrotransposon'. *Molecular Biology* 38 (4): 509–15. https://doi.org/10.1023/B:MBIL.0000037002.82913.6d.

Quinlan, Aaron R., and Ira M. Hall. 2010. 'BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features'. *Bioinformatics* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* V. 4.4.0. R Foundation for Statistical Computing, Vienna, Austria., Released. https://www.R-project.org/.

Ramesh, Niraja, Marc J. F. Escher, Malou M. Mampell, et al. 2021. 'Antagonistic Interactions between Two Neuroligins Coordinate Pre- and Postsynaptic Assembly'. *Current Biology* 31 (8): 1711-1725.e5. https://doi.org/10.1016/j.cub.2021.01.093.

Rayment, Ivan. 1996. 'The Structural Basis of the Myosin ATPase Activity*'. *Journal of Biological Chemistry* 271 (27): 15850–53. https://doi.org/10.1074/jbc.271.27.15850.

Razin, A, and H Cedar. 1991. 'DNA Methylation and Gene Expression'. *Microbiological Reviews* 55 (3): 451–58. https://doi.org/10.1128/mr.55.3.451-458.1991.

Rebollo, Rita, Mark T. Romanish, and Dixie L. Mager. 2012. 'Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes'. *Annual Review of Genetics* 46 (1): 21–42. https://doi.org/10.1146/annurev-genet-110711-155621.

Rech, Gabriel E., Santiago Radío, Sara Guirao-Rico, et al. 2022a. 'Population-Scale Long-Read Sequencing Uncovers Transposable Elements Associated with Gene Expression Variation and Adaptive Signatures in Drosophila'. *Nature Communications* 13 (1): 1948. https://doi.org/10.1038/s41467-022-29518-8.

Rech, Gabriel E., Santiago Radío, Sara Guirao-Rico, et al. 2022b. 'Population-Scale Long-Read Sequencing Uncovers Transposable Elements Associated with Gene Expression Variation and Adaptive Signatures in Drosophila'. *Nature Communications* 13 (1): 1948. https://doi.org/10.1038/s41467-022-29518-8.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Rey, Olivier, Etienne Danchin, Marie Mirouze, Céline Loot, and Simon Blanchet. 2016a. 'Adaptation to Global Change: A Transposable Element–Epigenetics Perspective'. *Trends in Ecology & Evolution* 31 (7): 514–26. https://doi.org/10.1016/j.tree.2016.03.013.

Rey, Olivier, Etienne Danchin, Marie Mirouze, Céline Loot, and Simon Blanchet. 2016b. 'Adaptation to Global Change: A Transposable Element–Epigenetics Perspective'. *Trends in Ecology & Evolution* 31 (7): 514–26. https://doi.org/10.1016/j.tree.2016.03.013.

Riddiford, Lynn M., Peter Cherbas, and James W. Truman. 2000. 'Ecdysone Receptors and Their Biological Actions'. In *Vitamins & Hormones*, vol. 60. Elsevier. https://doi.org/10.1016/S0083-6729(00)60016-X.

Riegler, Markus, Manpreet Sidhu, Wolfgang J. Miller, and Scott L. O'Neill. 2005. 'Evidence for a Global Wolbachia Replacement in Drosophila Melanogaster'. *Current Biology* 15 (15): 1428–33. https://doi.org/10.1016/j.cub.2005.06.069.

Rij, Ronald P. van, and Eugene Berezikov. 2009. 'Small RNAs and the Control of Transposons and Viruses in Drosophila'. *Trends in Microbiology* 17 (4): 163–71. https://doi.org/10.1016/j.tim.2009.01.003.

Roach, Michael J., Simon A. Schmidt, and Anthony R. Borneman. 2018. 'Purge Haplotigs: Allelic Contig Reassignment for Third-Gen Diploid Genome Assemblies'. *BMC Bioinformatics* 19 (1): 460. https://doi.org/10.1186/s12859-018-2485-7.

Roe, Anne, and George Gaylord Simpson. 1958. *Behavior and Evolution*. With Internet Archive. New Haven, Yale University Press. http://archive.org/details/behaviorevolutio00roearich.

Rohner, Nicolas, Dan F. Jarosz, Johanna E. Kowalko, et al. 2013. 'Cryptic Variation in Morphological Evolution: HSP90 as a Capacitor for Loss of Eyes in Cavefish'. *Science* 342 (6164): 1372–75. https://doi.org/10.1126/science.1240276.

Rolandi, Carmen, John R. B. Lighton, Gerardo J. De La Vega, Pablo E. Schilman, and Julián Mensch. 2018. 'Genetic Variation for Tolerance to High Temperatures in a Population of *Drosophila Melanogaster*'. *Ecology and Evolution* 8 (21): 10374–83. https://doi.org/10.1002/ece3.4409.

Romero-Mujalli, Daniel, Laura I R Fuchs, Martin Haase, Jan-Peter Hildebrandt, Franz J Weissing, and Tomás A Revilla. 2024. 'Emergence of Phenotypic Plasticity through Epigenetic Mechanisms'. *Evolution Letters*, March 27, qrae012. https://doi.org/10.1093/evlett/qrae012.

Rosenberg, Eugene, and Ilana Zilber-Rosenberg. 2011. 'Symbiosis and Development: The Hologenome Concept'. *Birth Defects Research Part C: Embryo Today: Reviews* 93 (1): 56–66. https://doi.org/10.1002/bdrc.20196.

Rosenberg, Eugene, and Ilana Zilber-Rosenberg. 2018. 'The Hologenome Concept of Evolution after 10 Years'. *Microbiome* 6 (1): 78. https://doi.org/10.1186/s40168-018-0457-9.

Rothacher, Lukas, Mar Ferrer-Suay, and Christoph Vorburger. 2016. 'Bacterial Endosymbionts Protect Aphids in the Field and Alter Parasitoid Community Composition'. *Ecology* 97 (7): 1712–23. https://doi.org/10.1890/15-2022.1.

Roughgarden, Joan. 2023. 'Holobiont Evolution: Population Theory for the Hologenome'. *The American Naturalist* 201 (6): 763–78. https://doi.org/10.1086/723782.

Roy, Marlène, Barbara Viginier, Édouard Saint-Michel, Frédérick Arnaud, Maxime Ratinier, and Marie Fablet. 2020. 'Viral Infection Impacts Transposable Element Transcript Amounts in *Drosophila*'. *Proceedings of the National Academy of Sciences* 117 (22): 12249–57. world. https://doi.org/10.1073/pnas.2006106117.

Rust, Joshua. 2022. 'Phenotype-First Hypotheses, Spandrels and Early Metazoan Evolution'. *History and Philosophy of the Life Sciences* 44 (4): 48. https://doi.org/10.1007/s40656-022-00531-w.

Rutherford, Suzanne L., and Susan Lindquist. 1998. 'Hsp90 as a Capacitor for Morphological Evolution'. *Nature* 396 (6709): 336–42. https://doi.org/10.1038/24550.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Ryder, Edward, and Steven Russell. 2003. 'Transposable Elements as Tools for Genomics and Genetics in Drosophila'. *Briefings in Functional Genomics & Proteomics* 2 (1): 57–71. https://doi.org/10.1093/bfgp/2.1.57.

Ryu, Taewoo, Heather D. Veilleux, Jennifer M. Donelson, Philip L. Munday, and Timothy Ravasi. 2018. 'The Epigenetic Landscape of Transgenerational Acclimation to Ocean Warming'. *Nature Climate Change* 8 (6): 504–9. https://doi.org/10.1038/s41558-018-0159-0.

Sabarís, Gonzalo, Bernd Schuettengruber, Giorgio L. Papadopoulos, et al. 2025. 'A Mechanistic Basis for Genetic Assimilation in Natural Fly Populations'. *Proceedings of the National Academy of Sciences* 122 (11): e2415982122. https://doi.org/10.1073/pnas.2415982122.

Sangster, Todd A., Neeraj Salathia, Soledad Undurraga, et al. 2008. 'HSP90 Affects the Expression of Genetic Variation and Developmental Stability in Quantitative Traits'. *Proceedings of the National Academy of Sciences* 105 (8): 2963–68. https://doi.org/10.1073/pnas.0712200105.

Sawarkar, Ritwick, Cem Sievers, and Renato Paro. 2012. 'Hsp90 Globally Targets Paused RNA Polymerase to Regulate Gene Expression in Response to Environmental Stimuli'. *Cell* 149 (4): 807–18. https://doi.org/10.1016/j.cell.2012.02.061.

Scheiner, Samuel M. 1993. 'Genetics and Evolution of Phenotypic Plasticity'. *Annual Review of Ecology, Evolution, and Systematics* 24 (Volume 24, 1993): 35–68. https://doi.org/10.1146/annurev.es.24.110193.000343.

Scheiner, Samuel M., Michael Barfield, and Robert D. Holt. 2017. 'The Genetics of Phenotypic Plasticity. XV. Genetic Assimilation, the Baldwin Effect, and Evolutionary Rescue'. *Ecology and Evolution* 7 (21): 8788–803. https://doi.org/10.1002/ece3.3429.

Scheiner, Samuel M., Roberta L. Caplan, and Richard F. Lyman. 1989. 'A Search for Trade-Offs among Life History Traits inDrosophila Melanogaster'. *Evolutionary Ecology* 3 (1): 51–63. https://doi.org/10.1007/BF02147931.

Schlichting, Carl D. 2008. 'Hidden Reaction Norms, Cryptic Genetic Variation, and Evolvability'. *Annals of the New York Academy of Sciences* 1133 (1): 187–203. https://doi.org/10.1196/annals.1438.010.

Schlichting, Carl D., and Harry Smith. 2002. 'Phenotypic Plasticity: Linking Molecular Mechanisms with Evolutionary Outcomes'. *Evolutionary Ecology* 16 (3): 189–211. https://doi.org/10.1023/A:1019624425971.

Schmidleithner, Lisa, Philipp Stüve, and Markus Feuerer. 2025. 'Transposable Elements as Instructors of the Immune System'. *Nature Reviews Immunology*, April 29, 1–11. https://doi.org/10.1038/s41577-025-01172-3.

Schmitz, Jürgen, and Jürgen Brosius. 2011. 'Exonization of Transposed Elements: A Challenge and Opportunity for Evolution'. *Biochimie*, Coding or Non-coding: need they be exclusive?, vol. 93 (11): 1928–34. https://doi.org/10.1016/j.biochi.2011.07.014.

Schmitz, Robert J., Zachary A. Lewis, and Mary G. Goll. 2019. 'DNA Methylation: Shared and Divergent Features across Eukaryotes'. *Trends in Genetics* 35 (11): 818–27. https://doi.org/10.1016/j.tig.2019.07.007.

Schneider, Ralf F., and Axel Meyer. 2017. 'How Plasticity, Genetic Assimilation and Cryptic Genetic Variation May Contribute to Adaptive Radiations'. *Molecular Ecology* 26 (1): 330–50. https://doi.org/10.1111/mec.13880.

Schneuwly, Stephan, Atsushi Kuroiwa, and Walter J. Gehring. 1987. 'Molecular Analysis of the Dominant Homeotic Antennapedia Phenotype'. *The EMBO Journal* 6 (1): 201–6. https://doi.org/10.1002/j.1460-2075.1987.tb04739.x.

Schostak, Nataliya, Konstantin Pyatkov, Elena Zelentsova, et al. 2008. 'Molecular Dissection of Penelope Transposable Element Regulatory Machinery'. *Nucleic Acids Research* 36 (8): 2522–29. https://doi.org/10.1093/nar/gkm1166.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Schrader, Lukas, and Jürgen Schmitz. 2019. 'The Impact of Transposable Elements in Adaptive Evolution'. *Molecular Ecology* 28 (6): 1537–49. https://doi.org/10.1111/mec.14794.

Schrey, Aaron W., Travis R. Robbins, Jacob Lee, et al. 2016. 'Epigenetic Response to Environmental Change: DNA Methylation Varies with Invasion Status'. *Environmental Epigenetics* 2 (2): dvw008. https://doi.org/10.1093/eep/dvw008.

Sciamanna, Ilaria, Patrizia Vitullo, Angela Curatolo, and Corrado Spadafora. 2009. 'Retrotransposons, Reverse Transcriptase and the Genesis of New Genetic Information'. *Gene*, Genomic Impact of Eukaryotic Transposable Elements, vol. 448 (2): 180–86. https://doi.org/10.1016/j.gene.2009.07.011.

Sela, Noa, Eddo Kim, and Gil Ast. 2010. 'The Role of Transposable Elements in the Evolution of Non-Mammalian Vertebrates and Invertebrates'. *Genome Biology* 11 (6): R59. https://doi.org/10.1186/gb-2010-11-6-r59.

Seong, Hoon Je, Sang-Wook Han, and Woo Jun Sul. 2021. 'Prokaryotic DNA Methylation and Its Functional Roles'. *Journal of Microbiology* 59 (3): 242–48. https://doi.org/10.1007/s12275-021-0674-y.

Serga, Svitlana V., Oleksandr M. Maistrenko, Nataliya P. Matiytsiv, Alexander M. Vaiserman, and Iryna A. Kozeretska. 2021. 'Effects of Wolbachia Infection on Fitness-Related Traits in Drosophila Melanogaster'. *Symbiosis* 83 (2): 163–72. https://doi.org/10.1007/s13199-020-00743-3.

Shapiro, James A. 2000. 'Transposable Elements as the Key to a 21st Century View of Evolution'. In *Transposable Elements and Genome Evolution*, edited by John F. McDonald. Springer Netherlands. https://doi.org/10.1007/978-94-011-4156-7_19.

Sharma, Vineeta, Surbhi Kohli, and Vani Brahmachari. 2017. 'Correlation between Desiccation Stress Response and Epigenetic Modifications of Genes in Drosophila Melanogaster: An Example of Environment-Epigenome Interaction'. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1860 (10): 1058–68. https://doi.org/10.1016/j.bbagrm.2017.08.001.

Shields, Hazel J., Annika Traa, and Jeremy M. Van Raamsdonk. 2021. 'Beneficial and Detrimental Effects of Reactive Oxygen Species on Lifespan: A Comprehensive Review of Comparative and Experimental Studies'. *Frontiers in Cell and Developmental Biology* 9 (February): 628157. https://doi.org/10.3389/fcell.2021.628157.

Shingleton, Alexander W. 2010. 'The Regulation of Organ Size in Drosophila: Physiology, Plasticity, Patterning and Physical Force'. *Organogenesis* 6 (2): 76–87. https://doi.org/10.4161/org.6.2.10375.

Shingleton, Alexander W., Josephine R. Masandika, Lily S. Thorsen, Yuqing Zhu, and Christen K. Mirth. 2017. 'The Sex-Specific Effects of Diet Quality versus Quantity on Morphology in Drosophila Melanogaster'. *Royal Society Open Science* 4 (9): 170375. https://doi.org/10.1098/rsos.170375.

Shingleton, Alexander W., and Hui Yuan Tang. 2012. 'Plastic Flies'. *Fly* 6 (3): 147–52. https://doi.org/10.4161/fly.20323.

Shirasu-Hiza, Mimi M., Marc S. Dionne, Linh N. Pham, Janelle S. Ayres, and David S. Schneider. 2007. 'Interactions between Circadian Rhythm and Immunity in Drosophila Melanogaster'. *Current Biology* 17 (10): R353–55. https://doi.org/10.1016/j.cub.2007.03.049.

Shumate, Alaina, and Steven L Salzberg. 2021. 'Liftoff: Accurate Mapping of Gene Annotations'. *Bioinformatics* 37 (12): 1639–43. https://doi.org/10.1093/bioinformatics/btaa1016.

Sienski, Grzegorz, Julia Batki, Kirsten-André Senti, et al. 2015. 'Silencio/CG9754 Connects the Piwi–piRNA Complex to the Cellular Heterochromatin Machinery'. *Genes & Development* 29 (21): 2258–71. https://doi.org/10.1101/gad.271908.115.

Singal, Rakesh, and Gordon D. Ginder. 1999. 'DNA Methylation'. *Blood* 93 (12): 4059–70. https://doi.org/10.1182/blood.V93.12.4059.

Singh, Nadia D. 2019. 'Wolbachia Infection Associated with Increased Recombination in Drosophila'. *G3*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Genes|*Genomes*|*Genetics* 9 (1): 229–37. https://doi.org/10.1534/g3.118.200827.

Slotkin, R. Keith, and Robert Martienssen. 2007. 'Transposable Elements and the Epigenetic Regulation of the Genome'. *Nature Reviews Genetics* 8 (4): 272–85. https://doi.org/10.1038/nrg2072.

Smit, A. F. A., R. Hubley, and P. Green. 2021. '2013–2015. RepeatMasker Open-4.0'.

Song, Liang, Shao-shan Carol Huang, Aaron Wise, et al. 2016. 'A Transcription Factor Hierarchy Defines an Environmental Stress Response Network'. *Science* 354 (6312): aag1550. https://doi.org/10.1126/science.aag1550.

Sørensen, Jesper G., Mads F. Schou, and Volker Loeschcke. 2017. 'Evolutionary Adaptation to Environmental Stressors: A Common Response at the Proteomic Level'. *Evolution* 71 (6): 1627–42. https://doi.org/10.1111/evo.13243.

Spannhoff, Astrid, Yong Kee Kim, Noel J -M Raynal, et al. 2011. 'Histone Deacetylase Inhibitor Activity in Royal Jelly Might Facilitate Caste Switching in Bees'. *EMBO Reports* 12 (3): 238–43. https://doi.org/10.1038/embor.2011.9.

Stapley, Jessica, Anna W. Santure, and Stuart R. Dennis. 2015. 'Transposable Elements as Agents of Rapid Adaptation May Explain the Genetic Paradox of Invasive Species'. *Molecular Ecology* 24 (9): 2241–52. https://doi.org/10.1111/mec.13089.

Stearns, Stephen C. 1989. 'The Evolutionary Significance of Phenotypic Plasticity: Phenotypic Sources of Variation among Organisms Can Be Described by Developmental Switches and Reaction Norms'. *BioScience* 39 (7): 436–45. https://doi.org/10.2307/1311135.

Stinchcombe, John R., and Mark Kirkpatrick. 2012. 'Genetics and Evolution of Function-Valued Traits: Understanding Environmentally Responsive Phenotypes'. *Trends in Ecology & Evolution* 27 (11): 637–47. https://doi.org/10.1016/j.tree.2012.07.002.

Strunov, Anton, Sina Lerch, Wolf U. Blanckenhorn, Wolfgang J. Miller, and Martin Kapun. 2022. 'Complex Effects of Environment and Wolbachia Infections on the Life History of Drosophila Melanogaster Hosts'. *Journal of Evolutionary Biology* 35 (6): 788–802. https://doi.org/10.1111/jeb.14016.

Su, Qi, Xiaomao Zhou, and Youjun Zhang. 2013. 'Symbiont-Mediated Functions in Insect Hosts'. *Communicative & Integrative Biology* 6 (3): e23804. https://doi.org/10.4161/cib.23804.

Sudakaran, Sailendharan, Christian Kost, and Martin Kaltenpoth. 2017. 'Symbiont Acquisition and Replacement as a Source of Ecological Innovation'. *Trends in Microbiology* 25 (5): 375–90. https://doi.org/10.1016/j.tim.2017.02.014.

Sullivan, Gail M., and Richard Feinn. 2012. 'Using Effect Size—or Why the P Value Is Not Enough'. *Journal of Graduate Medical Education* 4 (3): 279–82. https://doi.org/10.4300/JGME-D-12-00156.1.

Sundaram, Vasavi, and Joanna Wysocka. 2020. 'Transposable Elements as a Potent Source of Diverse Cis-Regulatory Sequences in Mammalian Genomes'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1795): 20190347. https://doi.org/10.1098/rstb.2019.0347.

Sykiotis, Gerasimos P., and Dirk Bohmann. 2008. 'Keap1/Nrf2 Signaling Regulates Oxidative Stress Tolerance and Lifespan in *Drosophila*'. *Developmental Cell* 14 (1): 76–85. https://doi.org/10.1016/j.devcel.2007.12.002.

Tang, Hui Yuan, Martha S. B. Smith-Caldas, Michael V. Driscoll, Samy Salhadar, and Alexander W. Shingleton. 2011. 'FOXO Regulates Organ-Specific Phenotypic Plasticity In Drosophila'. *PLOS Genetics* 7 (11): e1002373. https://doi.org/10.1371/journal.pgen.1002373.

Taniguchi, Cullen M., Brice Emanuelli, and C. Ronald Kahn. 2006. 'Critical Nodes in Signalling Pathways: Insights into Insulin Action'. *Nature Reviews Molecular Cell Biology* 7 (2): 85–96. https://doi.org/10.1038/nrm1837.

Tata, Jamshed R. 2005. 'One Hundred Years of Hormones'. *EMBO Reports* 6 (6): 490–96. https://doi.org/10.1038/sj.embor.7400444.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Therneau, Terry M., Thomas Lumley (original S.->R port and R. maintainer until 2009), Atkinson Elizabeth, and Crowson Cynthia. 2024. *Survival: Survival Analysis*. V. 3.8-3. Released December 17. https://cran.r-project.org/web/packages/survival/index.html.

Thomas, Jainy, and Ellen J. Pritham. 2015. 'Helitrons, the Eukaryotic Rolling-Circle Transposable Elements'. *Microbiology Spectrum* 3 (4): 10.1128/microbiolspec.mdna3-0049–2014. https://doi.org/10.1128/microbiolspec.mdna3-0049-2014.

Tipton, Laura, John L. Darcy, and Nicole A. Hynson. 2019. 'A Developing Symbiosis: Enabling Cross-Talk Between Ecologists and Microbiome Scientists'. *Frontiers in Microbiology* 10 (February). https://doi.org/10.3389/fmicb.2019.00292.

Touret, Franck, François Guiguen, and Christophe Terzian. 2014. 'Wolbachia Influences the Maternal Transmission of the Gypsy Endogenous Retrovirus in Drosophila Melanogaster'. *mBio* 5 (5): e01529-01514. https://doi.org/10.1128/mBio.01529-14.

Truitt, Amy M., Martin Kapun, Rupinder Kaur, and Wolfgang J. Miller. 2019. 'Wolbachia Modifies Thermal Preference in Drosophila Melanogaster'. *Environmental Microbiology* 21 (9): 3259–68. https://doi.org/10.1111/1462-2920.14347.

Tu, Meng-Ping, and Marc Tatar. 2003. 'Juvenile Diet Restriction and the Aging and Reproduction of Adult Drosophila Melanogaster'. *Aging Cell* 2 (6): 327–33. https://doi.org/10.1046/j.1474-9728.2003.00064.x.

Ullastres, Anna, Miriam Merenciano, and Josefa González. 2021. 'Regulatory Regions in Natural Transposable Element Insertions Drive Interindividual Differences in Response to Immune Challenges in Drosophila'. *Genome Biology* 22 (1): 265. https://doi.org/10.1186/s13059-021-02471-3.

Urquhart, Andrew S., Nicholas F. Chong, Yongqing Yang, and Alexander Idnurm. 2022. 'A Large Transposable Element Mediates Metal Resistance in the Fungus *Paecilomyces Variotii*'. *Current Biology* 32 (5): 937-950.e5. https://doi.org/10.1016/j.cub.2021.12.048.

Van Voorhies, Wayne A. 2009. 'Metabolic Function in Drosophila Melanogaster in Response to Hypoxia and Pure Oxygen'. *Journal of Experimental Biology* 212 (19): 3132–41. https://doi.org/10.1242/jeb.031179.

Van't Hof, Arjen E., Pascal Campagne, Daniel J. Rigden, et al. 2016. 'The Industrial Melanism Mutation in British Peppered Moths Is a Transposable Element'. *Nature* 534 (7605): 102–5. https://doi.org/10.1038/nature17951.

Varoqui, Marion, Mourdas Mohamed, Bruno Mugat, et al. 2025. 'Temporal and Spatial Niche Partitioning in a Retrotransposon Community of the Drosophila Melanogaster Genome'. *Nucleic Acids Research* 53 (11): gkaf516. https://doi.org/10.1093/nar/gkaf516.

Vaser, Robert, Ivan Sović, Niranjan Nagarajan, and Mile Šikić. 2017. 'Fast and Accurate de Novo Genome Assembly from Long Uncorrected Reads'. *Genome Research* 27 (5): 737–46. https://doi.org/10.1101/gr.214270.116.

Vaury, C., M. C. Chaboissier, M. E. Drake, O. Lajoinie, B. Dastugue, and A. Pélisson. 1994. 'The Doc Transposable Element in Drosophila Melanogaster and Drosophila Simulans: Genomic Distribution and Transcription'. *Genetica* 93 (1–3): 117–24. https://doi.org/10.1007/BF01435244.

Vautrin, Emilie, and Fabrice Vavre. 2009. 'Interactions between Vertically Transmitted Symbionts: Cooperation or Conflict?' *Trends in Microbiology* 17 (3): 95–99. https://doi.org/10.1016/j.tim.2008.12.002.

Verma, K. K. 2008. 'Polyphenism in Insects and Juvenile Hormone (JH)'. In *Encyclopedia of Entomology*. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-6359-6_3055.

Vermeulen, C. J., and R. Bijlsma. 2004. 'Changes in Mortality Patterns and Temperature Dependence of Lifespan in Drosophila Melanogaster Caused by Inbreeding'. *Heredity* 92 (4): 275–81. https://doi.org/10.1038/sj.hdy.6800412.

Verspoor, Rudi L., and Penelope R. Haddrill. 2011. 'Genetic Diversity, Population Structure and Wolbachia Infection Status in a Worldwide Sample of

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Drosophila Melanogaster and D. Simulans Populations'. *PLOS ONE* 6 (10): e26318. https://doi.org/10.1371/journal.pone.0026318.

Via, Sara, and Russell Lande. 1985. 'Genotype-Environment Interaction and the Evolution of Phenotypic Plasticity'. *Evolution* 39 (3): 505–22. https://doi.org/10.2307/2408649.

Videlier, Mathieu, Howard D. Rundle, and Vincent Careau. 2019. 'Sex-Specific Among-Individual Covariation in Locomotor Activity and Resting Metabolic Rate in Drosophila Melanogaster'. *The American Naturalist* 194 (6): E164–76. https://doi.org/10.1086/705678.

Videlier, Mathieu, Howard D. Rundle, and Vincent Careau. 2021. 'Sex-specific Genetic (Co)Variances of Standard Metabolic Rate, Body Mass and Locomotor Activity in Drosophila Melanogaster'. *Journal of Evolutionary Biology* 34 (8): 1279–89. https://doi.org/10.1111/jeb.13887.

Vieira, Cristina, Marie Fablet, Emmanuelle Lerat, et al. 2012. 'A Comparative Analysis of the Amounts and Dynamics of Transposable Elements in Natural Populations of *Drosophila Melanogaster* and *Drosophila Simulans*'. *Journal of Environmental Radioactivity* 113 (November): 83–86. https://doi.org/10.1016/j.jenvrad.2012.04.001.

Vieira, Cristina, Elena G Pasyukova, Zhao-Bang Zeng, J Brant Hackett, Richard F Lyman, and Trudy F C Mackay. 2000. 'Genotype-Environment Interaction for Quantitative Trait Loci Affecting Life Span in Drosophila Melanogaster'. *Genetics* 154 (1): 213–27. https://doi.org/10.1093/genetics/154.1.213.

Villagra, C., and D. Frías-Lasserre. 2020. 'Epigenetic Molecular Mechanisms in Insects'. *Neotropical Entomology* 49 (5): 615–42. https://doi.org/10.1007/s13744-020-00777-8.

Vitte, C., and O. Panaud. 2003. 'Formation of Solo-LTRs Through Unequal Homologous Recombination Counterbalances Amplifications of LTR Retrotransposons in Rice Oryza Sativa L.' *Molecular Biology and Evolution* 20 (4): 528–40. https://doi.org/10.1093/molbev/msg055.

Vonesch, Sibylle Chantal, David Lamparter, Trudy F. C. Mackay, Sven Bergmann, and Ernst Hafen. 2016. 'Genome-Wide Analysis Reveals Novel Regulators of Growth in Drosophila Melanogaster'. *PLOS Genetics* 12 (1): e1005616. https://doi.org/10.1371/journal.pgen.1005616.

Vorburger, Christoph, and Romain Rouchet. 2016. 'Are Aphid Parasitoids Locally Adapted to the Prevalence of Defensive Symbionts in Their Hosts?' *BMC Evolutionary Biology* 16 (1): 271. https://doi.org/10.1186/s12862-016-0811-0.

Waddington, C. H. 1942. 'Canalization of Development and the Inheritance of Acquired Characters'. *Nature* 150 (3811): 563–65. https://doi.org/10.1038/150563a0.

Waddington, C. H. 1953. 'Genetic Assimilation of an Acquired Character'. *Evolution* 7 (2): 118–26. https://doi.org/10.1111/j.1558-5646.1953.tb00070.x.

Wang, Jie, Michael W. Itgen, Huiju Wang, et al. 2021. 'Gigantic Genomes Provide Empirical Tests of Transposable Element Dynamics Models'. *Genomics, Proteomics & Bioinformatics* 19 (1): 123–39. https://doi.org/10.1016/j.gpb.2020.11.005.

Wang, Kai, Tao Wang, Xiang-Qian Gao, Xin-Zhe Chen, Fei Wang, and Lu-Yu Zhou. 2021. 'Emerging Functions of Piwi-Interacting RNAs in Diseases'. *Journal of Cellular and Molecular Medicine* 25 (11): 4893–901. https://doi.org/10.1111/jcmm.16466.

Wang, Xin, Anne Ramat, Martine Simonelig, and Mo-Fang Liu. 2023. 'Emerging Roles and Functional Mechanisms of PIWI-Interacting RNAs'. *Nature Reviews Molecular Cell Biology* 24 (2): 123–41. https://doi.org/10.1038/s41580-022-00528-0.

Wang, Yiguan, Paul McNeil, Rashidatu Abdulazeez, et al. 2023. 'Variation in Mutation, Recombination, and Transposition Rates in Drosophila Melanogaster and Drosophila Simulans'. *Genome Research* 33 (4): 587–98. https://doi.org/10.1101/gr.277383.122.

Weger, Anastasia A., and Clare C. Rittschof. 2024. 'The Diverse Roles of Insulin Signaling in Insect

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Behavior'. *Frontiers in Insect Science* 4 (April). https://doi.org/10.3389/finsc.2024.1360320.

Wei, Liya, and Xiaofeng Cao. 2016. 'The Effect of Transposable Elements on Phenotypic Variation: Insights from Plants to Humans'. *Science China Life Sciences* 59 (1): 24–37. https://doi.org/10.1007/s11427-015-4993-2.

Wei, Xiaocui, Junxia Li, and Yang Li. 2023. 'RNA Interfering Machinery: New Insight into the Dicer-2 Complexes in the dsRNA-Processing Cycle'. *Signal Transduction and Targeted Therapy* 8 (1): 2. https://doi.org/10.1038/s41392-022-01272-9.

Weilguny, Lukas, and Robert Kofler. 2019. 'DeviaTE: Assembly-Free Analysis and Visualization of Mobile Genetic Element Composition'. *Molecular Ecology Resources* 19 (5): 1346–54. https://doi.org/10.1111/1755-0998.13030.

Wells, Jonathan N., and Cédric Feschotte. 2020. 'A Field Guide to Eukaryotic Transposable Elements'. *Annual Review of Genetics* 54 (November): 539–61. https://doi.org/10.1146/annurev-genet-040620-022145.

Wen, Kejia, Lijuan Yang, Tuanlin Xiong, et al. 2016. 'Critical Roles of Long Noncoding RNAs in Drosophila Spermatogenesis'. *Genome Research* 26 (9): 1233–44. https://doi.org/10.1101/gr.199547.115.

Werren, John H., Laura Baldo, and Michael E. Clark. 2008. 'Wolbachia: Master Manipulators of Invertebrate Biology'. *Nature Reviews Microbiology* 6 (10): 741–51. https://doi.org/10.1038/nrmicro1969.

West-Eberhard, Mary Jane. 2003. *Developmental Plasticity and Evolution*. Oxford University Press. https://doi.org/10.1093/oso/9780195122343.001.0001.

Wheeler, Matthew W, Robert M. Park, and A. John Bailer. 2006. 'Comparing Median Lethal Concentration Values Using Confidence Interval Overlap or Ratio Tests'. *Environmental Toxicology and Chemistry* 25 (5): 1441–44. https://doi.org/10.1897/05-320r.1.

Whitelaw, Emma, and David I. K. Martin. 2001. 'Retrotransposons as Epigenetic Mediators of Phenotypic Variation in Mammals'. *Nature Genetics* 27 (4): 4. https://doi.org/10.1038/86850.

Wicker, Thomas, François Sabot, Aurélie Hua-Van, et al. 2007. 'A Unified Classification System for Eukaryotic Transposable Elements'. *Nature Reviews Genetics* 8 (12): 973–82. https://doi.org/10.1038/nrg2165.

Williams, George C. (George Christopher). 1966. *Adaptation and Natural Selection; a Critique of Some Current Evolutionary Thought*. With Internet Archive. Princeton, N.J., Princeton University Press. http://archive.org/details/adaptationnatura00will.

Wingfield, John C. 2008. 'Comparative Endocrinology, Environment and Global Change'. *General and Comparative Endocrinology* 157 (3): 207–16. https://doi.org/10.1016/j.ygcen.2008.04.017.

Winter, David J., Austen R. D. Ganley, Carolyn A. Young, et al. 2018. 'Repeat Elements Organise 3D Genome Structure and Mediate Transcription in the Filamentous Fungus Epichloë Festucae'. *PLoS Genetics* 14 (10): e1007467. https://doi.org/10.1371/journal.pgen.1007467.

Wojciechowski, Marek, Robert Lowe, Joanna Maleszka, Danyal Conn, Ryszard Maleszka, and Paul J. Hurd. 2018. 'Phenotypically Distinct Female Castes in Honey Bees Are Defined by Alternative Chromatin States during Larval Development'. *Genome Research* 28 (10): 1532–42. https://doi.org/10.1101/gr.236497.118.

Wojda, Iwona. 2017. 'Temperature Stress and Insect Immunity'. *Journal of Thermal Biology*, SI:Ectotherms: Performance, vol. 68 (August): 96–103. https://doi.org/10.1016/j.jtherbio.2016.12.002.

Woltereck, R. 1909. 'Weitere experimentelle untersuchungen über Artänderung, speziell über das Wesen quantitativer Artunterschiede bei Daphniden'. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* 9 (1): 146–146. https://doi.org/10.1007/BF01876686.

Wong, Keith SK, and Walid A. Houry. 2006. 'Hsp90 at the Crossroads of Genetics and Epigenetics'. *Cell*

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Research 16 (9): 742–49. https://doi.org/10.1038/sj.cr.7310090.

Wood, Jason G., and Stephen L. Helfand. 2013. 'Chromatin Structure and Transposable Elements in Organismal Aging'. *Frontiers in Genetics* 4 (December). https://doi.org/10.3389/fgene.2013.00274.

Woodruff, R. C., J. N. Thompson, J. S. F. Barker, and H. Huai. 2000. 'Transposable DNA Elements and Life History Traits: II. Transposition of P DNA Elements in Somatic Cells Reduces Fitness, Mating Activity, and Locomotion of Drosophila Melanogaster'. In *Transposable Elements and Genome Evolution*, edited by John F. McDonald. Springer Netherlands. https://doi.org/10.1007/978-94-011-4156-7_26.

Wu, Carl. 1995. 'Heat Shock Transcription Factors: Structure and Regulation'. *Annual Review of Cell and Developmental Biology* 11 (Volume 11, 1995): 441–69. https://doi.org/10.1146/annurev.cb.11.110195.002301.

Wu, Qi, Guixiang Yu, Xingyi Cheng, et al. 2020. 'Sexual Dimorphism in the Nutritional Requirement for Adult Lifespan in Drosophila Melanogaster'. *Aging Cell* 19 (3): e13120. https://doi.org/10.1111/acel.13120.

Xia, Bo, Weimin Zhang, Guisheng Zhao, et al. 2024. 'On the Genetic Basis of Tail-Loss Evolution in Humans and Apes'. *Nature* 626 (8001): 1042–48. https://doi.org/10.1038/s41586-024-07095-8.

Xiong, Y., and T. H. Eickbush. 1990. 'Origin and Evolution of Retroelements Based upon Their Reverse Transcriptase Sequences.' *The EMBO Journal* 9 (10): 3353–62. https://doi.org/10.1002/j.1460-2075.1990.tb07536.x.

Yang, Meiling, Yuanyuan Wei, Feng Jiang, et al. 2014. 'MicroRNA-133 Inhibits Behavioral Aggregation by Controlling Dopamine Synthesis in Locusts'. *PLOS Genetics* 10 (2): e1004206. https://doi.org/10.1371/journal.pgen.1004206.

Ye, Yixin H., Andrei Seleznev, Heather A. Flores, Megan Woolfit, and Elizabeth A. McGraw. 2017. 'Gut Microbiota in *Drosophila Melanogaster* Interacts with *Wolbachia* but Does Not Contribute to *Wolbachia*-Mediated Antiviral Protection'. *Journal of Invertebrate Pathology* 143 (February): 18–25. https://doi.org/10.1016/j.jip.2016.11.011.

Yin, Hang, Sarah Sweeney, Debasish Raha, Michael Snyder, and Haifan Lin. 2011. 'A High-Resolution Whole-Genome Map of Key Chromatin Modifications in the Adult Drosophila Melanogaster'. *PLOS Genetics* 7 (12): e1002380. https://doi.org/10.1371/journal.pgen.1002380.

Yu, Yi, and Igor Iatsenko. 2025. '*Drosophila* Symbionts in Infection: When a Friend Becomes an Enemy'. *Infection and Immunity*, April 2, e00511-24. https://doi.org/10.1128/iai.00511-24.

Yun, Hyun Myoung, Bom Hyun, Xinge Song, and Seogang Hyun. 2024. 'Piwi Expressed in Drosophila Adipose Tissues Regulates Systemic IGF Signaling and Growth via IGF-Binding Protein'. *Biochemical and Biophysical Research Communications* 695 (February): 149495. https://doi.org/10.1016/j.bbrc.2024.149495.

Yushkova, Elena, and Alexey Moskalev. 2023. 'Transposable Elements and Their Role in Aging'. *Ageing Research Reviews* 86 (April): 101881. https://doi.org/10.1016/j.arr.2023.101881.

Zamore, Phillip D. 2010. 'Somatic piRNA Biogenesis'. *The EMBO Journal* 29 (19): 3219–21. https://doi.org/10.1038/emboj.2010.232.

Zhang, Huiming, Zhaobo Lang, and Jian-Kang Zhu. 2018. 'Dynamics and Function of DNA Methylation in Plants'. *Nature Reviews Molecular Cell Biology* 19 (8): 489–506. https://doi.org/10.1038/s41580-018-0016-z.

Zhang, Jianming, Antony M. Dean, Frédéric Brunet, and Manyuan Long. 2004. 'Evolving Protein Functional Diversity in New Genes of Drosophila'. *Proceedings of the National Academy of Sciences* 101 (46): 16246–50. https://doi.org/10.1073/pnas.0407066101.

Zhang, Yi, and Danny Reinberg. 2001. 'Transcription Regulation by Histone Methylation: Interplay between Different Covalent Modifications of the Core

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

Histone Tails'. *Genes & Development* 15 (18): 2343–60. https://doi.org/10.1101/gad.927301.

Zhao, L., and W. A. Jones. 2012. 'Expression of Heat Shock Protein Genes in Insect Stress Responses'. *Invertebrate Survival Journal* 9 (1): 1. https://www.isj.unimore.it/index.php/ISJ/article/view/264.

Zheng, Chunyang, and Jeffrey J. Hayes. 2003. 'Structures and Interactions of the Core Histone Tail Domains'. *Biopolymers* 68 (4): 539–46. https://doi.org/10.1002/bip.10303.

Zheng, Ya, Pan-Pan Ren, Jia-Lin Wang, and Yu-Feng Wang. 2011. 'Wolbachia-Induced Cytoplasmic Incompatibility Is Associated with Decreased Hira Expression in Male Drosophila'. *PLOS ONE* 6 (4): e19512. https://doi.org/10.1371/journal.pone.0019512.

Zhou, Shanshan, Terry G. Campbell, Eric A. Stone, Trudy F. C. Mackay, and Robert R. H. Anholt. 2012. 'Phenotypic Plasticity of the Drosophila Transcriptome'. *PLoS Genetics* 8 (3): e1002593. https://doi.org/10.1371/journal.pgen.1002593.

Zhu, Jinsong, Li Chen, and Alexander S. Raikhel. 2003. 'Posttranscriptional Control of the Competence Factor βFTZ-F1 by Juvenile Hormone in the Mosquito Aedes Aegypti'. *Proceedings of the National Academy of Sciences* 100 (23): 13338–43. https://doi.org/10.1073/pnas.2234416100.

Zwaan, B. J., R. Bijlsma, and R. F. Hoekstra. 1992. 'On the Developmental Theory of Ageing. II. The Effect of Developmental Temperature on Longevity in Relation to Adult Body Size in D. Melanogaster'. *Heredity* 68 (2): 123–30. https://doi.org/10.1038/hdy.1992.19.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés

FOLIO ADMINISTRATIF

THESE DE L'INSA LYON, MEMBRE DE L'UNIVERSITE DE LYON

NOM : LARUE
(avec précision du nom de jeune fille, le cas échéant)

DATE de SOUTENANCE : 28/11/2025

Prénoms : Anaïs, Espérance, Alice

TITRE : La contribution des éléments transposables à la plasticité phénotypique chez *Drosophila melanogaster*

NATURE : Doctorat

Numéro d'ordre : 2025ISAL0105

Ecole doctorale : Ecologie, Ecosystème, Microbiologie, Modélisation (E2M2)

Spécialité : Biologie des populations et écologie

RESUME :
La plasticité phénotypique désigne la capacité d'un organisme doté d'un génotype donné à produire des phénotypes distincts en réponse aux changements environnementaux. Dans un contexte adaptatif, cette plasticité peut améliorer la potentialité d'un organisme à faire face aux fluctuations environnementales. Les éléments transposables (ET), bien que souvent négligés dans ce contexte, peuvent jouer un rôle important dans la plasticité phénotypique grâce à leur capacité à moduler l'expression génétique en fonction de l'environnement. Les ET sont des séquences d'ADN mobiles et répétitives. Leur activité est régulée par des mécanismes épigénétiques, mais aussi par leurs propres séquences régulatrices, tous deux sensibles aux variations environnementales. Ces fluctuations peuvent entraîner une expression ou une mobilisation différentielle des ET, influençant l'expression génique et induisant une plasticité épigénomique, génomique et transcriptomique. Une telle plasticité au niveau moléculaire pourrait conduire à une diversité interindividuelle et promouvoir une réponse plastique aux changements environnementaux. Afin d'étudier cette hypothèse dans des conditions contrôlées, nous avons étudié des mouches *Drosophila melanogaster* génétiquement modifiées ayant un contenu en ET différent, mais partageant un patrimoine génétique commun. Le criblage phénotypique a révélé des différences significatives entre les populations. Une étude des phénotypes dans divers environnements a révélé des variations dans les réponses plastiques et une augmentation de la variance des traits pour la population la plus chargée en ET. Ces résultats suggèrent que les ET contribuent à l'hétérogénéité de la réponse environnementale et peuvent donc constituer un mécanisme sous-jacent à la plasticité phénotypique.

MOTS-CLÉS : éléments transposables, drosophile, plasticité phénotypique

Laboratoire (s) de recherche : Biologie Fonctionnelle, Insectes et Interactions (BF2I), Laboratoire de Biométrie et Biologie Evolutive (LBBE)

Directeur de thèse : Rita REBOLLO

Président de jury : Abdelaziz HEDDI

Composition du jury :

Rapporteur.trice.s : Cameron Ghalambor, Christoph Grunau, Pascale Lesage
Examinateur.trice.s Patricia Gibert, François Sabot, Abdelaziz Heddi,
Direction : Rita Rebollo, Cristina Vieira,
Invitée : Séverine Chambeyron.

Thèse accessible à l'adresse : https://theses.insa-lyon.fr/publication/2025ISAL0105/these.pdf © [A. Larue], [2025], INSA Lyon, tous droits réservés