COLLEGE ECOLOGIE
DOCTORAL GEOSCIENCES
BRETAGNE AGRONOMIE ALIMENTATION

L'INSTITUT agro Rennes Angers

THÈSE DE DOCTORAT DE

L'INSTITUT AGRO RENNES ANGERS

ÉCOLE DOCTORALE N° 600
Écologie, Géosciences, Agronomie, Alimentation
Spécialité : *Génétique, génomique et bio-informatique*

Par

Guillaume DORÉ

Production de ressources génomiques pour l'étude des mécanismes d'acclimatation au milieu terrestre de l'espèce aquatique invasive, *Ludwigia grandiflora* subsp. *hexapetala*

Thèse présentée et soutenue à L'institut Agro Rennes, le 02/12/2025
Unité de recherche : UMR DECOD
Thèse N° : 2025AGRORA20

Rapporteur·trice·s avant soutenance :

Patricia FAIVRE-RAMPANT	Chargée de recherche (HDR) - EPGV - CEA Institut de biologie François Jacob 2
Christophe HITTE	Ingénieur de recherche (HDR) - IGDR - INSERM

Composition du Jury :

Président·e :	Florence VAL	Professeure - IGEPP - Institut Agro Rennes-Angers
Examinateur·trice·s :	Pierre NOUHAUD	Chargé de recherche - CBGP - INRAE, Montpellier
	Mark COCK	Directeur de recherche - UMR 8227 - Station Biologique de Roscoff
Dir. de thèse :	Dominique BARLOY	Professeure - DECOD - Institut Agro Rennes-Angers

Invité·e·(s) :

Co-dir. de thèse :	Frédérique HUBLER	Chargée de recherche (HDR) - CreAAH - Université Rennes

Remerciements

Je tenais tout d'abord à remercier Dominique et Frédérique pour m'avoir encadré au cours de cette thèse sans qui tout ceci n'aurait pas été possible. Merci pour votre bienveillance, votre pédagogie ainsi que pour tout les échanges scientifiques riches que nous avons pu avoir. Même s'il y a eu des moments difficiles, je me suis toujours senti soutenu et cette thèse a été un plaisir pour moi.

Je voulais remercier les membres de mon CSI, Emmanuelle Becker et Stéphane Maury, de m'avoir accompagné et aidé au cours de ma thèse. Nos échanges ont amené beaucoup de réflexions intéressantes.

Je remercie les rapporteurs de ce manuscrit, Patricia Faivre-Rampant et Christophe Hitte, d'avoir accepté de relire ce travail. Merci également aux examinateurs, Florence Val, Pierre Nouhaud et Mark Cock, pour votre participation à mon jury de thèse.

Merci aussi à Liva pour son aide sur ma thèse lors de son stage de M1.

Merci à tous les collègues et amis doctorants et CDD, Ambre, Pauline, Jean, Marianne, Abel, Israel, Hugo, Hélène, Manu, Clément, Claire, Maëlle, Violetta, Baptiste, Paul et tous les autres avec qui on a passé de supers moments en espérant que ça va continuer.

Bien sûr, merci à tout le reste du couloir de l'UP ESP qui m'ont accueilli chaleureusement et avec qui ça a toujours été agréable de discuter. Un merci particulier à Ingrid pour son aide sur quelques galères (à Toulouse par exemple).

Enfin, un grand merci à ma famille et mes amis d'Alsace qui m'ont soutenu et qui ont essayé de comprendre tant bien que mal ce que je faisais. Merci à mes parents et à ma soeur de venir assister à ma thèse et d'avoir été là pour moi.

Pour finir, un merci tout particulier à Ségolène avec qui je partage ma vie depuis 6 ans et pour encore longtemps. Merci d'avoir été un soutien lors des moments difficiles et de m'avoir encouragé quand j'en avais besoin. Je n'y serais pas arrivé sans toi, je t'aime.

Table des matières

Introduction 1

Introduction générale 1

1 Ressources génomiques chez les plantes 2

1.1 Plastome et mitogénome 2

1.2 Génomes nucléaires 12

1.3 Gènes et transcrits chez les plantes 17

1.4 Méthylome 21

2 Stratégies de séquençages et bioinformatiques en génomiques végétales ... 23

2.1 Méthode de séquençage de l'ADN 23

2.2 Assemblage *de novo* 27

2.3 Spécificité de l'assemblage d'un génome d'organelles de plantes 29

2.4 Spécificité de l'assemblage d'un génome nucléaire de plantes 31

2.5 Décontamination 33

2.6 Annotation d'un génome 35

2.7 Transcriptome 37

3 La jussie, une plante aquatique invasive colonisant le milieu terrestre 43

3.1 Les invasions biologiques 43

3.2 La jussie 45

3.3 Colonisation du milieu terrestre de *Lgh* 49

3.4 Fonctions d'acclimatations au milieu terrestre des plantes amphibies . 50

4 Objectifs et stratégies de la thèse 53

4.1 Contexte de la thèse 53

4.2 Objectifs 53

4.3 Matériel végétal et stratégies de séquençage et d'assemblage 55

Chapitre 1 : Assemblage des mitogénomes de *Lgh* et *Lpm* 58

1 Introduction 59

1.1 Forme des mitogénomes de plantes 59

1.2 Mitogénomes dans l'ordre des Myrtales 60

1.3 Utilité de l'assemblage d'un mitogénome chez les plantes 60

2 Synthèse de l'article "De novo hybrid assembly unveils multi-chromosomal mitochondrial genomes in *Ludwigia* species highlighting genomic recombination, gene transfer, and RNA editing events" 61

2.1 Objectifs et stratégies 61

2.2 Principaux résultats 62

2.3 Conclusion 63

3 Article "De novo hybrid assembly unveils multi-chromosomal mitochondrial genomes in *Ludwigia* species highlighting genomic recombination, gene transfer, and RNA editing events" 63

Chapitre 2 : Assemblage du génome nucléaire de *Lgh* 93

1 Introduction 94

1.1 Génomique des invasions biologiques 94

1.2 Ressources génomiques dans l'ordre des Myrtales 96

2 Synthèse de l'article "From Water to land, introducing the first decaploid genome in Ludwigioideae : *Ludwigia grandiflora* subsp. *hexapetala*, an invasive amphibious plant" 98

2.1 Objectifs et stratégies 98

2.2 Principaux résultats 99

2.3 Conclusion 100

3 Article "From Water to land, introducing the first decaploid genome in Ludwigioideae : *Ludwigia grandiflora* subsp. *hexapetala*, an invasive amphibious plant " 100

Chapitre 3 : Assemblage du transcriptome de *Lgh* et analyses différentielles 112

1 Introduction 113

1.1 Plasticité phénotypique et invasions biologiques 113

1.2 Acclimatation de la jussie au milieu terrestre 114

1.3 Objectifs et hypothèses 114

2 Matériels et méthodes 115

2.1 Matériel végétal 115

2.2 Extraction de l'ARN 116

2.3 Assemblage *de novo* du transcriptome 117

2.4 Annotations du transcriptome *de novo* 119

2.5 Analyse différentielle en utilisant le génome de *Lgh* 119

2.6 Mapping et analyses différentielles en utilisant le transcriptome *de novo* de *Lgh* 120

3 Résultats et Discussion 121

3.1 Assemblage *de novo* du transcriptome de *Lgh* 121

3.2 Mapping des transcrits assemblés sur le génome 122

3.3 Annotations fonctionnelles du transcriptome de *Lgh* 130

3.4 Analyse globale de l'expression des gènes dans le génome 131

3.5 Identification des gènes différentiellement exprimés entre les tiges-feuilles et les racines de *Lgh* en utilisant le génome 132

3.6 Comparaison deux à deux des échantillons *via* le transcriptome de référence et analyse de l'effet de l'alternance sur *Lgh* lors d'une acclimatation au milieu terrestre 141

3.7 Analyses différentielles sur le transcriptome *de novo* 150

4 Conclusion 155

Discussion et perspectives 159

1 Bilan général de la thèse 160

2 Stratégies de séquençage 161

3 Erreurs dans les bases de données 163

3.1 Erreurs dans la prédiction des séquences protéiques 163

3.2 Erreurs dans les annotations fonctionnelles 164

3.3 Contaminations 165

4 Intérêt des données multi-omiques 166

5 Apports de la thèse 169

5.1 Apport pour la compréhension de l'acclimatation de *Lgh* au milieu terrestre 169

5.2 Apport des ressources génomiques de *Lgh* pour les études phylogénétiques chez les Myrtales et les Onagraceae 170

5.3 Apport dans la méthodologie d'assemblage de ressources génomiques complexes 171

6 Perspectives 174

6.1 Analyses différentielles du transcriptome plus poussées 174

6.2 Amélioration des ressources génomiques 175

6.3 Méthylome 177

6.4 Comparaison entre populations natives et invasives 178

Introduction

Introduction générale

Les espèces invasives font partie des principales menaces pour la biodiversité à l'échelle mondiale [Souza et Prevedello, 2021]. La jussie, *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*), est une des plantes aquatiques invasives les plus importantes d'Europe [Hussner, Stiers et al., 2017]. En France, un arrêté du 2 mai 2007 interdit sa commercialisation, son utilisation et son introduction dans le milieu naturel [legifrance, 2007].

L'étude des invasions biologiques s'est d'abord concentrée sur des questions écologiques [Hodgins, Battlay et Bock, 2025]. Aujourd'hui, des aspects génétique et d'évolution sont devenus fondamentaux dans les recherches sur les espèces exotiques envahissantes. Ainsi, dans le cadre des invasions biologiques, l'utilisation de ressources génomiques peut être une clé dans la compréhension de leur succès invasif [McGaughran et al., 2024]. Or, seul le génome chloroplastique de *Lgh* était publié au commencement de ma thèse [Barloy-Hubler et al., 2025]. Afin de disposer de ressources génomiques diversifiées et d'apporter des connaissances fonctionnelles à l'acclimatation de la jussie au milieu terrestre, ma thèse s'est, dans un premier temps, focalisée sur la production de ressources génomiques variées, i.e. le mitogènome, le génome et le transcriptome de *Lgh*. Puis une première analyse différentielle a été réalisée en comparant le comportement

de *Lgh* à des conditionnements en milieux terrestre et aquatique.

1 Ressources génomiques chez les plantes

Les ressources génomiques peuvent permettre l'amélioration variétale de plantes cultivées [Perez-de-Castro et al., 2012], l'identification d'espèces ou encore une meilleure compréhension des relations phylogénétiques [Wambugu, Ndjiondjop et Henry, 2018]. Dans les cellules de plantes, on trouve trois génomes distincts présents dans trois compartiments cellulaires : le génome nucléaire, le génome mitochondrial et le génome chloroplastique. Ils possèdent chacun des caractéristiques uniques. L'assemblage de ces génomes va permettre l'étude des fonctions et processus cellulaires en passant par l'annotation des gènes. L'ARN constitue l'ensemble du génome transcrit appelé transcriptome. Grâce à ce dernier, il est possible d'étudier l'expression des gènes et de la comparer entre différentes conditions. La régulation de l'expression des gènes peut se faire par des modifications épigénétiques.

1.1 Plastome et mitogénome

Structure et rôle du chloroplaste

Le chloroplaste est un organite cellulaire semi-autonome, présent dans les cellules photosynthétiques des plantes vertes (Viridiplantae) et des algues photosynthétiques, central pour la conversion de l'énergie lumineuse en énergie chimique et la régulation métabolique.

Les chloroplastes sont issus d'un événement endosymbiotique primaire, au cours duquel une cellule eucaryote ancestrale a intégré une cyanobactérie photosynthétique, il y a plus de 1,2 milliard d'années [Dobrogojski, Adamiec et Luciński, 2020]. Cette symbiose a permis à l'hôte de bénéficier de la capacité de photosynthèse oxygénique, conférant

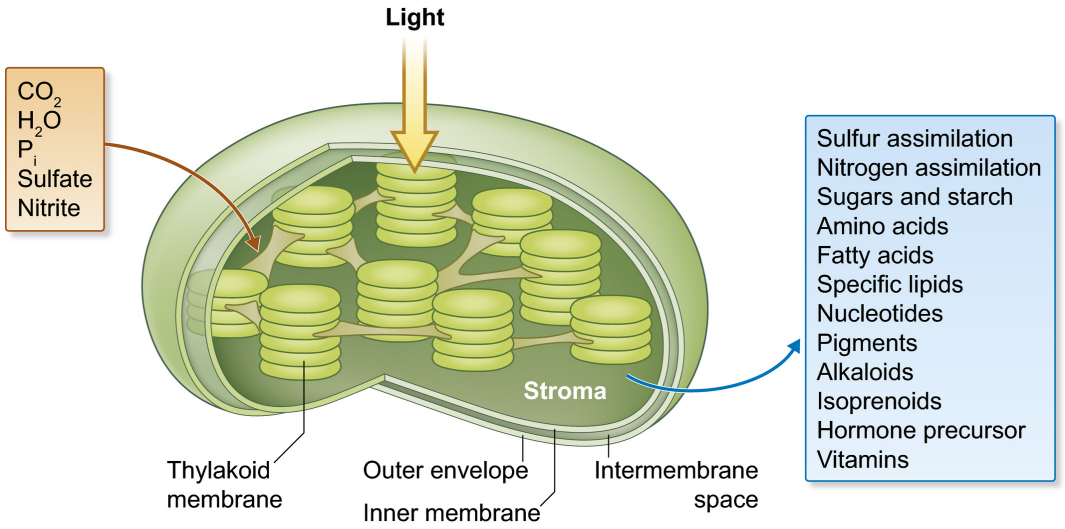


FIGURE 1 – Schéma d'un chloroplaste [Kirchhoff, 2019].

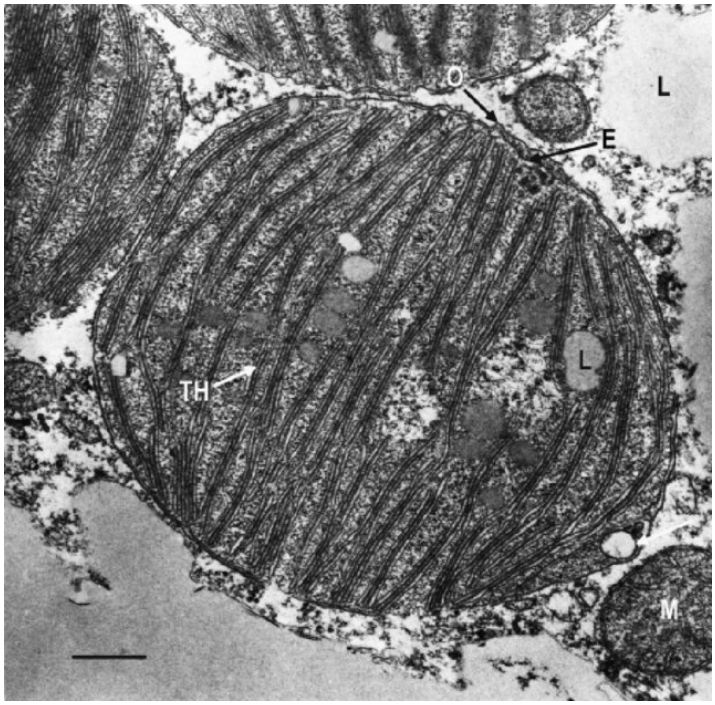


FIGURE 2 – Photographie au microscope électronique du chloroplaste de *Elysia clarki* [Curtis, Massey et Pierce, 2006]. TH : Thylakoïdes ; L : Lipides ; O : Outer membrane (membrane extérieure) ; E : Enveloppe ; M : Mitochondrie.

un avantage adaptatif majeur et ouvrant la voie à l'évolution des plantes et des algues [Raven et Allen, 2003]. Cette origine endosymbiotique a plusieurs conséquences fonctionnelles et évolutives, notamment une semi-autonomie puisque même s'ils possèdent leur propre génome, une grande partie des gènes nécessaires à leur fonctionnement a été transférée vers le noyau de la cellule hôte. Cela implique un trafic protéique complexe, où les protéines nucléaires doivent être synthétisées dans le cytosol puis importées dans le chloroplaste *via* des séquences signal et translocons [Raven et Allen, 2003]. Cette endosymbiose a favorisé la diversification évolutive des plantes et des algues, permettant le développement de multiples voies métaboliques dans le chloroplaste (photosynthèse, biosynthèse des acides aminés, lipides, hormones et métabolites secondaires). Chez les plantes terrestres, le chloroplaste mesure de 2 à 10 µm de diamètre et est entouré d'une double membrane qui délimite le stroma, un compartiment aqueux contenant les enzymes nécessaires au cycle de Calvin (fixation du carbone), à la biosynthèse d'acides aminés, de lipides et de certains pigments, ainsi qu'à la régulation de la réponse au stress oxydatif (Figures 1 et 2) [Cooper, 2000].

La principale fonction des chloroplastes est la photosynthèse qui se déroule en deux grandes étapes fonctionnelles : la phase claire et la phase sombre (ou cycle de Calvin). La phase claire, est localisée dans les thylakoïdes [Kirchhoff, 2019], membranes aplaties organisées en granums reliés par des lamelles stromales (Figures 1 et 2). Elle utilise directement l'énergie lumineuse pour exciter les électrons des pigments chlorophylliens dans les photosystèmes I et II. Cette énergie permet la photolyse de l'eau, libérant de l'oxygène, et la génération d'ATP et de NADPH *via* la chaîne de transport des électrons et l'ATP synthase [Arnon, 1971]. L'ATP et le NADPH produits sont ensuite utilisés dans la phase sombre, qui se déroule dans le stroma chloroplastique et ne nécessite pas directement la lumière. Cette étape, connue sous le nom de cycle de Calvin, fixe le dioxyde de carbone en composés organiques à trois carbones (triose phosphates), grâce aux en-

zymes clés comme la ribulose-1,5-bisphosphate carboxylase/oxygénase (RuBisCO). Les produits du cycle de Calvin servent de précurseurs pour la synthèse de sucres, d'amidons et d'autres métabolites essentiels à la croissance et au développement végétal [Sharkey, 2024].

Au-delà de la photosynthèse, le chloroplaste joue un rôle clé dans la biosynthèse d'hormones végétales (auxines, gibbérellines, jasmonates) ; la production de métabolites secondaires (pigments, caroténoïdes, flavonoïdes) ; l'assimilation de nitrate et sulfate et la synthèse d'acides aminés ; d'acides gras, de la chlorophylle et des caroténoïdes [Dobrogojski, Adamiec et Luciński, 2020] ; la régulation des réponses aux stress biotiques et abiotiques *via* la génération de ROS et la modulation de voies de signalisation intracellulaire [Bajguz et Piotrowska-Niczyporuk, 2023]. En plus des thylakoïdes et du stroma, le chloroplaste contient également des amyloplastes et des plastoglobules, qui jouent des rôles métaboliques et structuraux importants. Les amyloplastes sont des inclusions intracellulaires spécialisées dans le stockage temporaire de glucides sous forme d'amidon, constituant une réserve énergétique mobilisable par la plante lorsque la photosynthèse ne suffit pas, notamment pendant la nuit ou lors de périodes de stress [Bajguz et Piotrowska-Niczyporuk, 2023].

Les plastoglobules sont des gouttelettes lipidiques associées aux membranes thylakoïdiennes et au stroma. Elles contiennent des lipides neutres, des pigments caroténoïdes, des quinones et des protéines enzymatiques impliquées dans le métabolisme des lipides, la biosynthèse des hormones et la réponse au stress oxydatif [Austin et al., 2006]. Les plastoglobules servent également de sites de stockage et de régulation pour certains cofacteurs nécessaires aux complexes photosynthétiques et participent à la dégradation et au remodelage des membranes thylakoïdiennes pendant le vieillissement et le stress.

En conclusion, le chloroplaste est un organite central des cellules végétales, intégrant la photosynthèse, le stockage énergétique et la biosynthèse métabolique, tout en com-

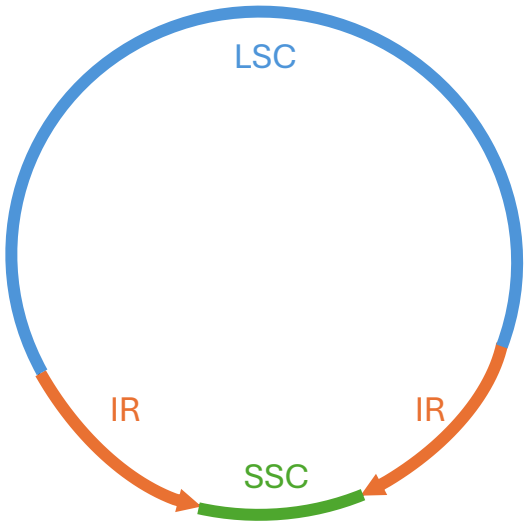


FIGURE 3 – Représentation schématique de l'organisation d'un plastome basée sur Wicke, Schneeweiss et al., 2011. LSC : Long Single Copy ; SSC : Short Single Copy ; IR : Inverted Repeat.

TABLE 1 – Liste des gènes codant des protéines dans les plastomes des angiospermes [Wicke, Schneeweiss et al., 2011].

Fonctions des gènes	**Gènes**
Protéine de la membrane interne	*cemA*
Biogenèse du cytochrome C	*ccsA*
Large sous-unité de la RuBisCO	*rbcL*
Synthèse des lipides	*accD*
ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*
NAD(P)H déshydrogénase	*ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
Facteur d'assemblage du photosystème I	*pafI* (*ycf3*), *pafII* (*ycf4*)
Complexe cytochrome b6f	*petA*, *petB*, *petD*, *petG*, *petL*, *petN*
Photosystème I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
Photosystème II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
Maturase	*matK*
ARN polymérase ADN-dépendante	*rpoA*, *rpoB*, *rpoC1*, *rpoC2*
Sous-unité protéolytique de la protéase Clp	*clpP*
Facteur A d'initiation de la traduction	*infA*
Larges sous-unités ribosomales	*rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23*, *rpl32*, *rpl33*, *rpl36*
Petites sous-unités ribosomales	*rps2*, *rps3*, *rps4*, *rps7*, *rps8*, *rps11*, *rps12*, *rps14*,*rps15*, *rps16*, *rps18*, *rps19*
Protéines avec une fonction inconnue	*ycf1*, *ycf2*

muniquant étroitement avec le noyau *via* des signaux rétrogrades. Sa structure complexe et son origine endosymbiotique lui confèrent une plasticité fonctionnelle essentielle à la croissance, au développement et à l'adaptation des plantes à leur environnement.

Les plastomes ou génomes chloroplastiques

Les génomes chloroplastiques des plantes ont été longtemps considérés comme ayant une forme circulaire, forme issue de la cyanobactérie originelle. D'après de récentes analyses, des structures linéaires des plastomes sont aussi présentes chez de nombreuses angiospermes [Dobrogojski, Adamiec et Luciński, 2020]. Les plastomes sont très bien conservés notamment concernant leur organisation quadripartite, leur contenu ainsi que l'ordre des gènes. En effet, l'organisation des plastomes se caractérise par la présence de deux répétitions inversées ("inverted repeats" = IR), deux fragments identiques mais inversés d'une taille de 20-25 kb (Figure 3). Ces séquences sont séparées d'un côté par une longue copie unique ("long single copy" = LSC) et de l'autre par une petite copie unique ("short single copy" = SSC) (Figure 3) [Wicke, Schneeweiss et al., 2011]. Il est possible dans de rare cas qu'une copie de l'IR soit perdue comme chez le pois chiche (*Cicer arietinum*) [Jansen et al., 2008].

La taille du plastome chez les plantes terrestres (Embryophytes) photosynthétiques ne varie pas beaucoup se trouvant entre 120 et 160 kb [Wicke, Schneeweiss et al., 2011]. Très peu de grandes séquences répétées, hormis les IR, sont présentes dans le plastome ce qui limite les variations de tailles et favorise la bonne conservation de la séquence. En effet, le génome chloroplastique contient des petites répétitions mais qui font en général moins de 50 pb limitant ainsi les recombinaisons [Wicke, Schneeweiss et al., 2011]. Les gènes sont aussi très bien conservés dans les plastomes. On trouve jusqu'à 81 gènes codant des protéines chez les angiospermes (Table 1) [Wicke, Schneeweiss et al., 2011]. Ces gènes codent notamment des protéines des photosystèmes I et II (5 et 15 gènes),

de la NAD(P)H déshydrogénase (11 gènes) ou encore des larges et petites sous-unités ribosomales (9 et 12 gènes). Les gènes des sous-unités ribosomales ou de la NAD(P)H peuvent être perdus ou pseudogénéisés dans certaines familles d'angiospermes [Wicke, Schneeweiss et al., 2011]. On trouve également des gènes codant des ARN ribosomaux (ARNr) et des ARN de transfert (ARNt). Il y a 4 gènes d'ARNr différents dans le plastome : *rrn4.5*, *rrn5*, *rrn16* et *rrn23*. Ils sont localisés dans les IR et sont donc présents en deux copies [Wicke, Schneeweiss et al., 2011]. Enfin, on trouve généralement 30 gènes d'ARNt différents dans les génomes chloroplastiques pouvant être présents en plusieurs copies. Exceptionnellement, 37 gènes ARNt ont été trouvés chez 3 espèces du genre *Spinacia* [She et al., 2022].

Ainsi, l'assemblage des plastomes des plantes vertes est généralement considéré comme relativement accessible en raison de leur taille modérée, de leur structure quadripartite et de leur contenu génique conservé [Freudenthal et al., 2020]. Toutefois, une certaine variation existe, avec notamment des inversions ou des changements d'orientation comme c'est le cas chez les fabacées [Doyle et al., 1996]. Les données sont une source d'information en phylogénétique végétale, permettant l'étude des relations entre les plantes, de créer de nouvelles classifications et une meilleure compréhension globale [Daniell et al., 2021]. Cependant, plusieurs cas de conflits dans les arbres phylogénétiques en focnion ses gènes ont été identifiés appelant à une certaine prudence [Walker et al., 2019 ; Tyszka et al., 2023]. Les plastomes sont donc des ressources génomiques pouvant jouer un rôle important dans la compréhension évolutive des plantes et relativement simple à obtenir.

Structure et rôle de la mitochondrie chez les plantes

La mitochondrie est un organite cellulaire présent dans toutes les cellules eucaryotes. Elle mesure de 1 à 3 µm de long et 0.5 µm de diamètre [Møller, Rasmusson et Van Aken,

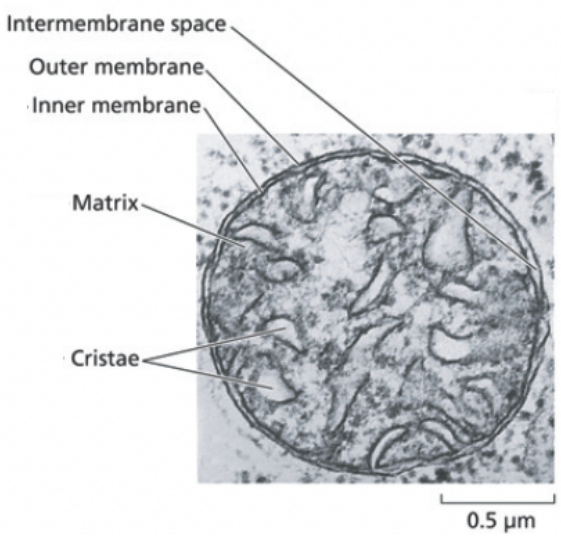


FIGURE 4 – Photographie d'une mitochondrie obtenue par microscopie électronique [Møller, Rasmusson et Van Aken, 2021].

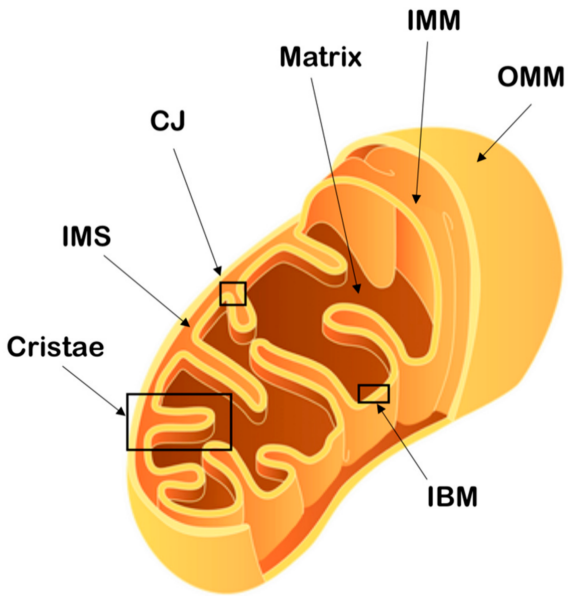


FIGURE 5 – Schéma de l'architecture d'une mitochondrie [Protasoni et Zeviani, 2021]. IMM : Membrane interne (Inner Mitochondrial Membrane) ; OMM : Membrane externe (Outer Mitochondrila Membrane) ; IBM : Limite de la membrane interne (Inner Boundary Membrane) ; CJ : Jonction des crêtes (Cristae Junction) ; IBM : Espace intermembranaire (Intermembrane Space).

2021]. La mitochondrie comporte une membrane externe avec des porines permettant le passage de petites molécules et une membrane interne plissée pour former des crêtes (Figures 4 et 5), augmentant la surface pour des réactions de phosphorylation oxydative [Pan et Hu, 2015]. La matrice est l'espace intérieur de la mitochondrie entourée par la membrane interne. Cette matrice est riche en protéines, notamment en enzymes pour le cycle de Krebs, en ADN mitochondrial et en ribosomes [Douce, 2012].

La mitochondrie provient à l'origine de l'endosymbiose d'une alphaprotéobactérie [Møller, Rasmusson et Van Aken, 2021] même si son placement dans l'arbre de la vie est encore débattu [Wang et Wu, 2015]. La mitochondrie est impliquée dans la respiration de la plante. Elle permet notamment la création d'ATP en passant par le cycle de Krebs et par le système de phosphorylation oxydative. Le système de phosphorylation oxydative est composé de 4 complexes de la chaîne de transport d'électrons (complexes I-IV) et un complexe de synthèse de l'ATP (complexe V) [Møller, Rasmusson et Van Aken, 2021]. Les 4 complexes de la chaîne de transport d'électrons sont le NADH déshydrogénase (35 sous-unités chez les plantes), le succinate déshydrogénase (8 sous-unités chez les plantes), le cytochrome c réductase (10 sous-unités chez les plantes) et le cytochrome c oxydase (14 sous-unités chez les plantes). Ils vont ainsi permettre le transport d'électrons du NADH vers l'oxygène et la création d'ATP à partir d'ADP [Jacoby et al., 2012]. Les sous-unités de ces complexes ne sont pas toutes codées dans la mitochondrie, elles proviennent du noyau et sont donc importées [Mackenzie et McIntosh, 1999].

La mitochondrie interagit également avec le chloroplaste, notamment lors du cycle de Krebs [Mackenzie et McIntosh, 1999]. Elle permet de maintenir le niveau d'ATP et l'équilibre redox dans le noir ainsi que d'oxyder les substrats produits par la photosynthèse comme le malate, le pyruvate ou le NAD(P)H [Hoefnagel, Atkin et Wiskich, 1998]. En plus de ça, la mitochondrie est impliquée dans les réponses aux stress des plantes notamment par la production de ROS (Reactive Oxygen Species = Dérivé réactif de

TABLE 2 – Liste des gènes codant des protéines dans les mitogénomes des angiospermes [Skippington et al., 2015].

Fonctions des gènes	**Gènes**
Complexe I (NADH déshydrogénase)	*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*
Complexe II (Succinate déshydrogénase)	*sdh3*, *sdh4*
Complexe III (Cytochrome c réductase)	*cob*
Complexe IV (Cytochrome c oxydase)	*cox1*, *cox2*, *cox3*
Complexe V (ATP synthase)	*atp1*, *atp4*, *atp6*, *atp8*, *atp9*
Cytochrome c maturation	*ccmB*, *ccmC*, *ccmFc*, *ccmFn*
Maturase	*matR*
Protéine translocase	*mttb*
Larges sous-unités ribosomales	*rpl2*, *rpl5*, *rpl10*, *rpl16*
Petites sous-unités ribosomales	*rps1*, *rps2*, *rps3*, *rps4*, *rps7*, *rps10*, *rps11*, *rps12*, *rps13*, *rps14*, *rps19*

l'oxygène) [Jacoby et al., 2012]. Les ROS vont ensuite induire de nombreux processus de stress oxydatif dans la mitochondrie comme la peroxydation des lipides, des changements de la teneur en métaux ou du contenu en protéines. Enfin, la mitochondrie joue un rôle crucial dans l'apoptose ou mort cellulaire programmée [Halestrap et al., 2000]. Ce processus est très important pour la réponse à de multiples stress ainsi que pour le bon développement de la plante [Kabbage et al., 2017].

Le nombre de mitochondries dans une cellule de plantes supérieures peut aller de plusieurs centaines à des milliers suivant le type de cellule [Douce, 2012]. Par exemple, chez *Arabidopsis thaliana*, 300 à 450 mitochondries sont présentes dans les cellules de feuilles, jusqu'à 600 dans les cellules du mésophylle et au moins 400 dans les cellules des racines [Preuten et al., 2010].

Les mitogénomes de plantes

On trouve jusqu'à 41 gènes codant des protéines dans les mitogénomes des angiospermes [Skippington et al., 2015]. Ils sont impliqués dans différentes fonctions (Table 2) : complexe I (9 gènes), complexe II (2 gènes), complexe III (1 gène), complexe IV (3 gènes), complexe V (5 gènes), cytochrome *c* maturation (4 gènes), maturase (1 gène), protéine translocase (1 gène) et larges et petites sous-unités ribosomales (4 et 11 gènes). Parmi ces gènes, ceux des complexes I, III, IV, V, de maturation et de protéine translocase sont presque toujours tous présents [Zhong et al., 2023]. Le contenu en gènes du complexe II, de maturation du cytochrome *c* et des sous-unités ribosomales est plus variable dans les mitogénomes des angiospermes [Zhong et al., 2023]. Dans les mitogénomes de plantes, on trouve aussi les gènes codant des ARNt et les gènes codant des ARNr. Les gènes d'ARNt sont toujours présents dans les mitogénomes des plantes mais le set n'est jamais complet [Mower, Sloan et Alverson, 2012]. Les trois gènes mitochondriaux d'ARNr (5S, 18S et 26S) sont conservés dans presque tous les mitogénomes de

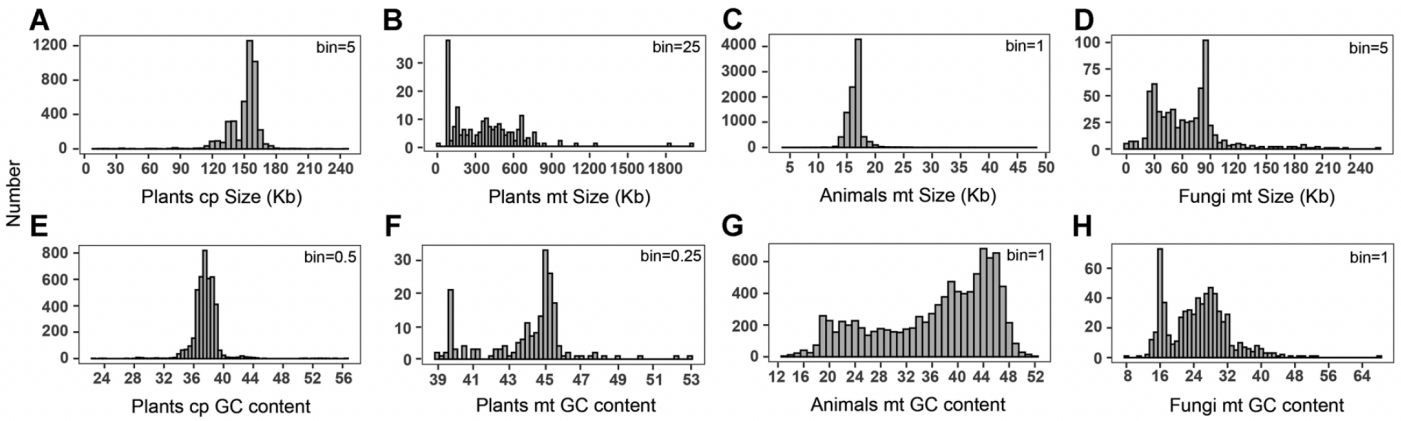


FIGURE 6 – Variations de la taille et du contenu en GC des organelles [Wu et al., 2020]. mt = mitogénome ; cp = plastome. Les graphes de A à D représentent les variations de tailles des plastomes de plantes terrestres, des mitogénomes de plantes terrestres, des mitogénomes des animaux et des mitogénomes des champignons. Les graphes de E à H montrent le contenu en GC pour les 4 mêmes organelles.

TABLE 3 – Distributions de la taille des répétitions au sein de différents groupes de plantes [Wynn et Christensen, 2019]. Les chiffres indiquent la fraction des espèces ayant au moins une répétition de la taille précisée au-dessus. Le gradient de couleur est bleu pour une fraction de 1 et jaune pour une fraction de 0.

Group	Subgroup	# of species	50-199	200-499	500-999	1000-2499	2500-4999	5000-7499	7500-9999	≥ 10000
Chlorophytes	Chlorophyta	26	0.88	0.19	0.08	0.00	0.00	0.00	0.00	0.00
Chlorophytes	Prasinophytes	8	0.63	0.00	0.00	0.00	0.00	0.00	0.50	0.25
Anthocerotophyta	Anthocerotophyta	4	1.00	0.50	1.00	0.25	0.00	0.00	0.00	0.00
Marchantiophyta	Marchantiophyta	6	1.00	1.00	0.67	0.00	0.00	0.00	0.00	0.00
Bryophytes	Bryophytes	26	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Streptophyta	Charophyta	8	0.88	0.25	0.00	0.00	0.00	0.00	0.00	0.00
Tracheophyta	Fern	2	1.00	1.00	1.00	1.00	1.00	0.00	1.00	0.00
Tracheophyta	Lycophyte	2	1.00	1.00	1.00	1.00	1.00	0.00	0.50	0.00
Tracheophyta	Gymnosperm	3	1.00	1.00	0.33	0.67	0.33	0.33	0.00	0.00
Tracheophyta	Angiosperm	72	1.00	0.96	0.47	0.49	0.46	0.36	0.25	0.43

plantes [Petersen et al., 2015]. Par exemple, chez trois espèces du genre *Viscum* les gènes d'ARNr 5S sont manquants et les 26S sont présents mais sous la forme de pseudogènes [Petersen et al., 2015].

La taille des mitogénomes de plantes est très variable, contrairement aux mitogénomes des animaux ou aux plastomes des plantes (Figure 6) [Wu et al., 2020]. Par exemple, chez les angiospermes, la taille des mitogénomes peut varier de 66 kb chez le gui (*Viscum scurruloideum*) [Skippington et al., 2015] à 11,7 Mb chez le mélèze de Sibérie (*Larix sibirica*) [A. Putintseva et al., 2020]. Par contre, le nombre de gènes ne variant pas avec la taille des mitogénomes, ces variations se font dans les séquences intergéniques [Wu et al., 2020]. Ceci peut être notamment attribué à la présence de régions répétées de tailles très variables dans les mitogénomes des plantes (Table 3) [Wynn et Christensen, 2019]. À travers l'étude de 72 mitogénomes d'angiospermes, Wynn *et al.* (2019) montrent la présence de régions répétées de plus de 600 pb dans 82% des espèces étudiées et de plus de 10 kb dans 43% des espèces. Cependant, la proportion de séquences répétées n'est pas forcément proportionnelle à la taille du mitogénomes. Par exemple, l'un des plus grands mitogénomes des angiospermes, celui du silène conique (*Silene conica* ; 11 Mb) [Sloan et al., 2012] et le plus petit mitogénome des angiospermes, celui de *Viscum scurruloideum* (66 kb) [Skippington et al., 2015] possèdent un taux de séquences répétées similaires (40,8 et 39%).

Une autre source de variation de la taille des mitogénomes de plantes est l'incorporation d'ADN par transferts horizontaux. En effet, des séquences du chloroplaste, du noyau ou même d'autres espèces peuvent être transférées vers le génome mitochondrial [Mower, Sloan et Alverson, 2012]. Chez les trachéophytes (plantes vasculaires), des insertions de séquences chloroplastiques sont présentes dans tous les mitogénomes ; la première datant de plus de 300 millions d'années au niveau de l'ancêtre commun des gymnospermes et angiospermes [Mower, Sloan et Alverson, 2012 ; Wang, Wu et al.,

2007]. En général, ces séquences représentent entre 1 et 12% du génome mitochondrial [Mower, Sloan et Alverson, 2012]. Par exemple, le mitogénome du melon (*Cucumis melo*) comprend 2,73% de séquences chloroplastiques [Cui et al., 2021], tandis que celui de l'orchidée pantoufle (*Paphiopedilum micranthum*), 10,34% [Yang, Dierckxsens et al., 2023]. Certains gènes chloroplastiques peuvent être transférés vers le mitogénome, notamment les gènes d'ARN de transfert qui peuvent être fonctionnels [Joyce et Gray, 1989]. Ils vont soit remplacer les tRNA mitochondriaux perdus au cours de l'évolution ou être présents sous la forme de pseudogènes ou fragments non fonctionnels [Marechal-Drouard et al., 1995]. Parfois, des gènes codant des protéines sont aussi transférés mais très rarement complets et/ou fonctionnels [Mower, Sloan et Alverson, 2012].

Des éléments mobiles comme les rétrotransposons, originaires du noyau, ont été trouvés dans certains mitogénomes de plantes [Marienfeld, Unseld et Brennicke, 1999]. Des transferts de séquences ou de gènes de la mitochondrie vers le noyau sont aussi possibles [Mackenzie et McIntosh, 1999]. Sachant cela, il peut être compliqué de déterminer si une séquence a été transférée de la mitochondrie vers le noyau ou inversement. Par ailleurs, les gènes transférés de la mitochondrie vers le noyau ont en général évolué et codent des protéines mitochondriales dans moins de 30% des cas [Møller, Rasmusson et Van Aken, 2021]. Ces transferts auraient lieu lors de contact entre la mitochondrie et le chloroplaste et entre la mitochondrie et le noyau durant le développement cellulaire [Mackenzie et McIntosh, 1999].

Des transferts horizontaux de gènes entre mitogénomes de plantes existent également [Richardson et Palmer, 2007]. Ces transferts sont souvent entre une espèce parasite et son hôte par contact direct. Cependant, tous les transferts ne peuvent pas être expliqués par une relation hôte-parasite ; même si une espèce parasite pourrait servir de vecteur entre deux espèces non liées par ce type de relation [Richardson et Palmer, 2007]. Les transferts d'ADN entre mitochondries chez les angiospermes sont plus fréquents que

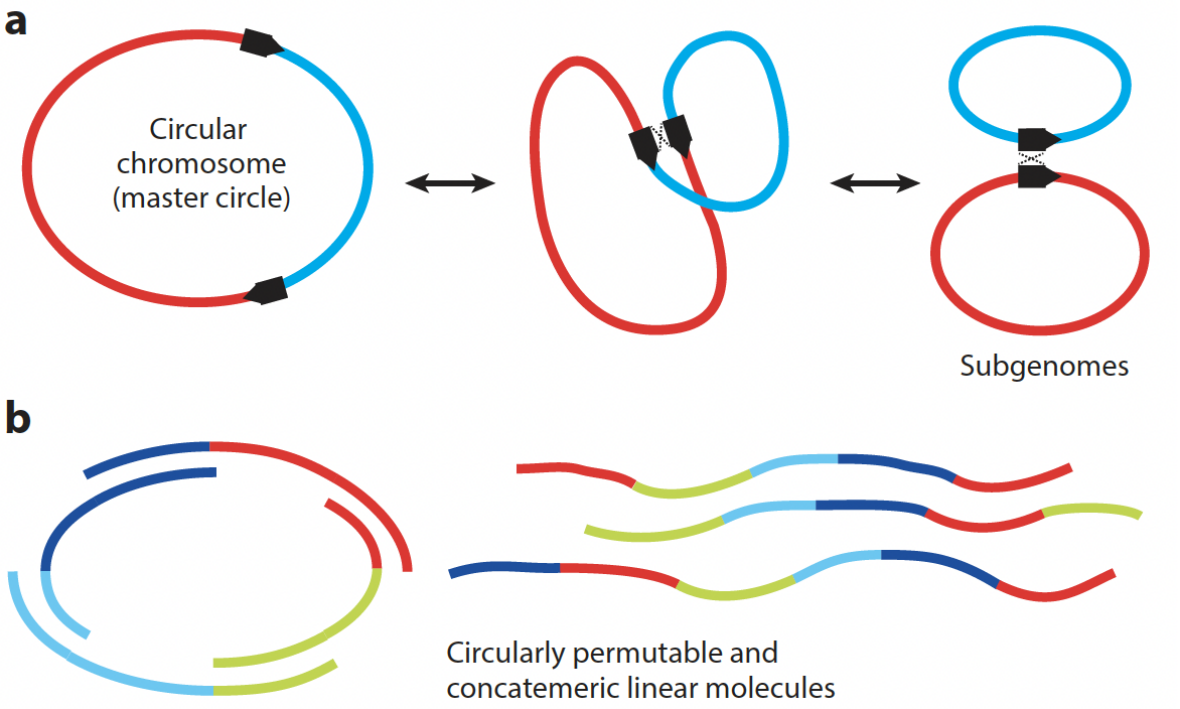


FIGURE 7 – Deux modèles de représentation des mitogénomes de plantes. (a) Le modèle du "Master circle" pouvant donner de plus petites molécules par recombinaison. (b) Un modèle alternatif qui serait composé de plusieurs molécules linéaires avec des régions communes pouvant ainsi formés un cercle [Gualberto et Newton, 2017].

les autres transferts entre organelles ou génome nucléaire [Sanchez-Puerta, 2014]. Ceci s'explique notamment par : (1) leur bonne capacité à incorporer de l'ADN ; (2) leur système de recombinaison permettant un réarrangement du génome ; (3) la présence de grandes régions intergéniques favorisant une incorporation sans endommager les gènes essentiels déjà présents [Sanchez-Puerta, 2014].

Traditionnellement, les génomes mitochondriaux de plantes étaient souvent représentés sous la forme d'une seule molécule circulaire, appelée "Master circle" (Figure 7) [Kozik et al., 2019]. Au vu de l'origine bactérienne des mitochondries et de la forme circulaire des génomes bactériens, cette forme semblait la plus cohérente. Cependant, ce modèle est actuellement discuté car 10% des mitogénomes assemblés possèdent une forme différente du "Master circle".

De plus, les séquences répétées (surtout celles > à 500 pb) peuvent être impliquées dans des recombinaisons des mitogénomes [Gualberto et Newton, 2017]. Ces recombinaisons vont créer des haplotypes ou formes alternatives. Elles consistent en la division d'une molécule circulaire en de plus petites molécules pouvant être circulaires ou linéaires [Gualberto et Newton, 2017]. En réalité, toutes ces formes existeraient dans différentes mitochondries d'une même cellule/espèce, et chaque mitochondrie ne contiendrait pas forcément le mitogénome complet [Preuten et al., 2010]. Ces mécanismes de recombinaisons basées sur des répétitions ont été mis en évidence chez le haricot mungo (*Vigna radiata*) [Alverson, Zhuo et al., 2011], l'épicéa commun (*Picea abies*) [Sullivan et al., 2020] ou encore la sauge rouge (*Salvia miltiorrhiza*) [Yang, Chen et al., 2022].

Une conséquence de toutes les raisons citées précédemment est qu'il y a très peu d'assemblages de mitogénomes dans les bases de données. Ceci génère un déficit des mitogénomes par rapport aux plastomes (200 vs 6000 chez les angiospermes en février 2024 dans Refseq [O'Leary et al., 2016]). Cependant, l'assemblage des mitogénomes

fournit notamment des clés pour l'identification d'espèces cryptiques, la reconstruction d'arbres phylogénétiques robustes et une compréhension de l'évolution des plantes [Zardoya, 2020].

1.2 Génomes nucléaires

Génomes de plantes dans les bases de données

En novembre 2025, 1529 espèces de plantes terrestres ont un génome assemblé (base de données RefSeq) soit $\approx 0,43\%$ des espèces connues (sur environ 350 000 espèces) [O'Leary et al., 2016]. Les trois familles ayant le plus de génomes assemblés sont les Poacées (blé, riz), les Brassicacées (colza, choux et moutarde) et les Fabacées (légumineuses) [Xie, Gong et al., 2024]. Il y a un biais "écologique" puisque les plantes d'intérêt pour l'Homme sont privilégiées [Schwacke, Bolger et Usadel, 2025]. En effet, ces plantes d'intérêt ont été majoritairement séquencées et assemblées alors que l'on recense entre 1200 et 2000 espèces domestiquées contre plus de 350 000 plantes sauvages [Marks et al., 2021]. De plus, la plupart des assemblages actuellement publiés sont incomplets (draft ou scaffold) [Kyriakidou et al., 2018]. Par ailleurs, environ 80% des espèces de plantes sont polyploïdes mais seulement 62 génomes de ces espèces ont été assemblés [Kyriakidou et al., 2018 ; Sun, Shang et al., 2022]. Tout ceci, révèle un manque dans les bases de données. Ce manque peut-être dû à la complexité des génomes de plantes, à la taille des génomes très variables, à la présence de nombreuses séquences répétées notamment d'éléments transposables ou encore à la polyploïdie [Kress et al., 2022]. L'assemblage d'un génome demande également un temps de calcul et de biocuration important. Tout ceci a un impact élevé sur la recherche en biologie végétale, limitant la compréhension de la diversité évolutive des plantes.

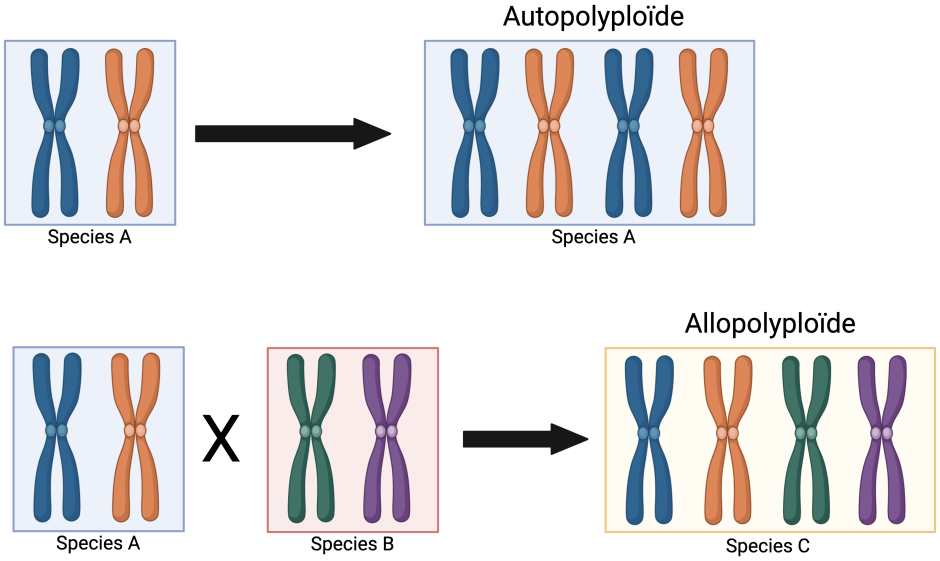


FIGURE 8 – Représentation schématique des deux types de polyploïdie : autopolyploïdie (en haut) et allopolyploïdie (en bas) basée sur Wendel et al., 2016.

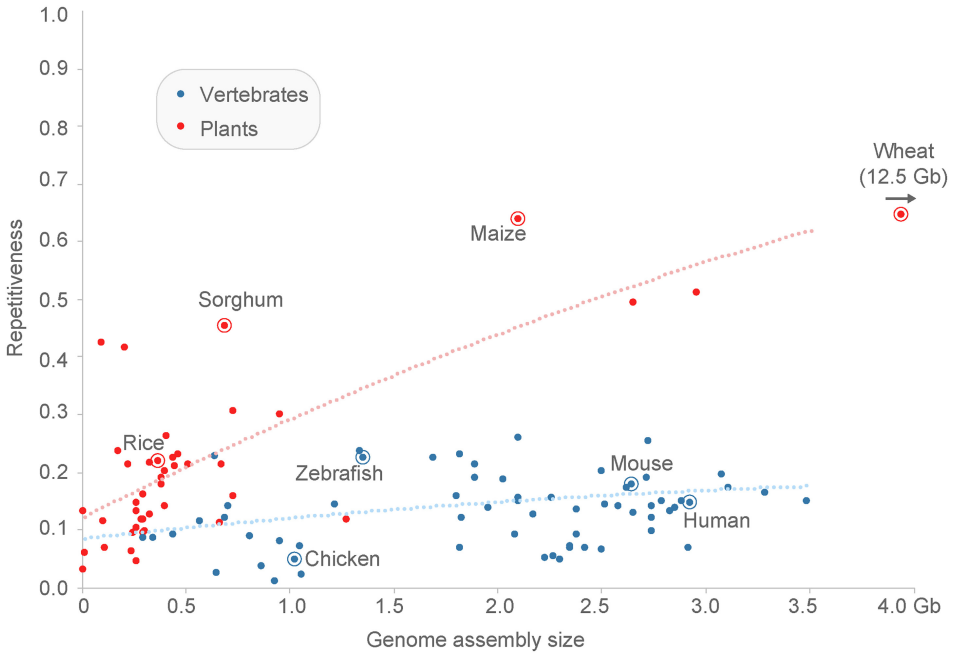


FIGURE 9 – Comparaison de la taille et du taux de répétitions dans les génomes de plantes et de vertébrés [Jiao et Schneeberger, 2017].

Polyploïdisation

De nombreux événements de polyploïdisation ont eu lieu dans la lignée de chaque angiosperme car c'est un phénomène cyclique [Wendel et al., 2016]. On peut diviser ces événements en deux catégories : l'autopolyploïdie provenant du doublement des chromosomes au sein d'une seule espèce et l'allopolyploïdie provenant d'un croisement entre deux espèces différentes (Figure 8). Suite à ça, il y a la création de sous-génome (= fraction du génome qui appartenait avant à une autre espèce). La polyploïdisation peut entraîner des échanges de gènes homoéologues (= gènes homologues provenant d'allopolyploïdie), une perte des gènes dupliqués, la diffusion d'éléments transposables, une augmentation de la taille du génome et une redondance génétique [Wendel et al., 2016 ; Kellogg et Bennetzen, 2004]. Sur le plus long terme, la polyploïdisation implique une perte de l'ADN redondant, le réarrangement des chromosomes et la réduction de la taille du génome [Wendel et al., 2016]. La présence de copie de gènes très similaire complexifie l'assemblage et l'annotation de génome [Kyriakidou et al., 2018].

Séquences répétées et éléments transposables

La taille des génomes chez les plantes est très variable pouvant aller de 63,6 Mb chez la plante carnivore, *Genlisea aurea* [Leushkin et al., 2013] à 160,45 Gb chez la fougère, *Tmesipteris oblanceolata* [Kozlov, 2024]. En plus de la polyploïdie déjà évoqué précédemment, cette variation est aussi due à la présence de nombreuses séquences répétées dont le nombre est proportionnel à la taille du génome (Figure 9) [Jiao et Schneeberger, 2017]. Le taux de répétitions peut atteindre plus de 60% chez le maïs ou le blé et peut être de 20% même dans des génomes de petites tailles comme celui d'*Arabidopsis thaliana* [Jiao et Schneeberger, 2017 ; Kellogg et Bennetzen, 2004]. Ces répétitions sont composés d'éléments transposables, de répétitions en tandem ou de microsatellites [Kellogg et Bennetzen, 2004]. La prolifération d'éléments transposables

a un rôle important dans l'augmentation de la taille des génomes [Wendel et al., 2016]. En effet, les séquences intergéniques sont surtout composées de rétrotransposons à LTR (long terminal repeat) ou de MITEs (Miniature Inverted-repat Transposable Elements) intact ou fragmenté [Kellogg et Bennetzen, 2004]. Ils sont très actifs chez les plantes pouvant diffuser des centaines ou milliers de copies à travers tout le génome lors d'un pic d'activité [Kellogg et Bennetzen, 2004]. C'est d'ailleurs la quantité de rétrotransposons à LTR avec la polyploïdie et la délétion d'ADN qui sont les causes principales de la variation de taille des génomes de plantes [Kellogg et Bennetzen, 2004]. Les éléments transposables comme d'autres séquences d'ADN non essentiels peuvent être rapidement retirés du génome par des recombinaisons de petites tailles ($<$ 100 pb) [Kellogg et Bennetzen, 2004].

Séquences provenant des organelles

Le dynamisme du génome est aussi montré par le nombre et la taille de séquences intégrées du mitogénome (NMUTs pour nuclear integrants of mitochondrial DNA) et du plastome (NPUTs pour nuclear integrants of plastid DNA) qui vont impactées la taille du génome [Zhang, Dong et al., 2020]. Ces insertions sont très fréquentes chez les génomes de plantes. Les NUMTs peuvent atteindre des tailles importantes comme par exemple 620 kb chez *Arabidopsis thaliana* [Stupar et al., 2001]. Les NUPTs sont généralement plus petits même si un fragment de 135 kb a été trouvé chez le coton (*Gossypium hirsutum*) [Zhang, Dong et al., 2020]. Ces insertions peuvent représenter plus de 1% et 2% du génome nucléaire pour les NMUTs et NUPTs, respectivement [Zhang, Dong et al., 2020]. Les transferts de séquences des organelles vers le génome ont lieu depuis l'apparition des organelles [Timmis et al., 2004]. Comme évoqué dans les paragraphes sur les mitogénomes et les plastomes, la plupart des gènes codant des protéines utilisées dans les processus des organelles ont été transférés dans le génome nucléaire

[Martin et Herrmann, 1998]. Les protéines sont ensuite importées dans les organelles. Les insertions ont plutôt lieu dans les régions centromériques et péricentromériques des chromosomes ; c'est-à-dire à fort niveau d'hétérochromatine et contenant en général peu de gènes. Une fois intégrés, les NMUTs et les NUPTs peuvent être éliminés, mutés, fragmentés ou réarrangés, mais peuvent également proliférer [Zhang, Dong et al., 2020]. Tous ces événements peuvent contribuer à la dynamique globale du génome nucléaire. Par exemple, lors de la délétion d'une insertion il est possible qu'une partie ne provenant pas du plastome ou du mitogénome soit aussi perdue [Zhang, Dong et al., 2020]. Ensuite, des insertions à l'intérieur de gènes sont assez fréquentes (25% des NMUTs et NUPTs chez le riz et *Arabidopsis*) [Zhang, Dong et al., 2020]. Elles peuvent modifier la séquence du transcrit en y ajoutant de nouveaux exons ou également influencer l'expression des gènes. Enfin, dans le chromosome sexuel Y des espèces dioïques, beaucoup de NUPTs et NMUTs sont retrouvés, pouvant ainsi contribuer à la différenciation des deux chromosomes sexuels [Zhang, Dong et al., 2020].

Organisation des chromosomes

Le nombre de chromosomes dans les génomes végétaux est extrêmement variable et ne reflète pas la taille du génome [Kellogg et Bennetzen, 2004]. Par exemple, chez les graminées (Poaceae), on observe une variation spectaculaire : *Zingeria biebersteiniana* possède seulement 2n = 4 chromosomes [Kim, Bol'sheva et al., 2009], tandis que *Poa litorosa* atteint 2n = 266 chromosomes [Edgar, 1986]. Malgré cette variabilité numérique, l'organisation générale des chromosomes est largement conservée. Chez la majorité des plantes à fleurs, les chromosomes sont monocentriques, présentant un centromère unique qui assure une ségrégation classique lors de la mitose et de la méiose. Certaines familles, comme les Cyperaceae ou Juncaceae, possèdent des chromosomes holocentriques, où le kinétochore s'étend sur toute la longueur du chromosome, entraînant une ségrégation

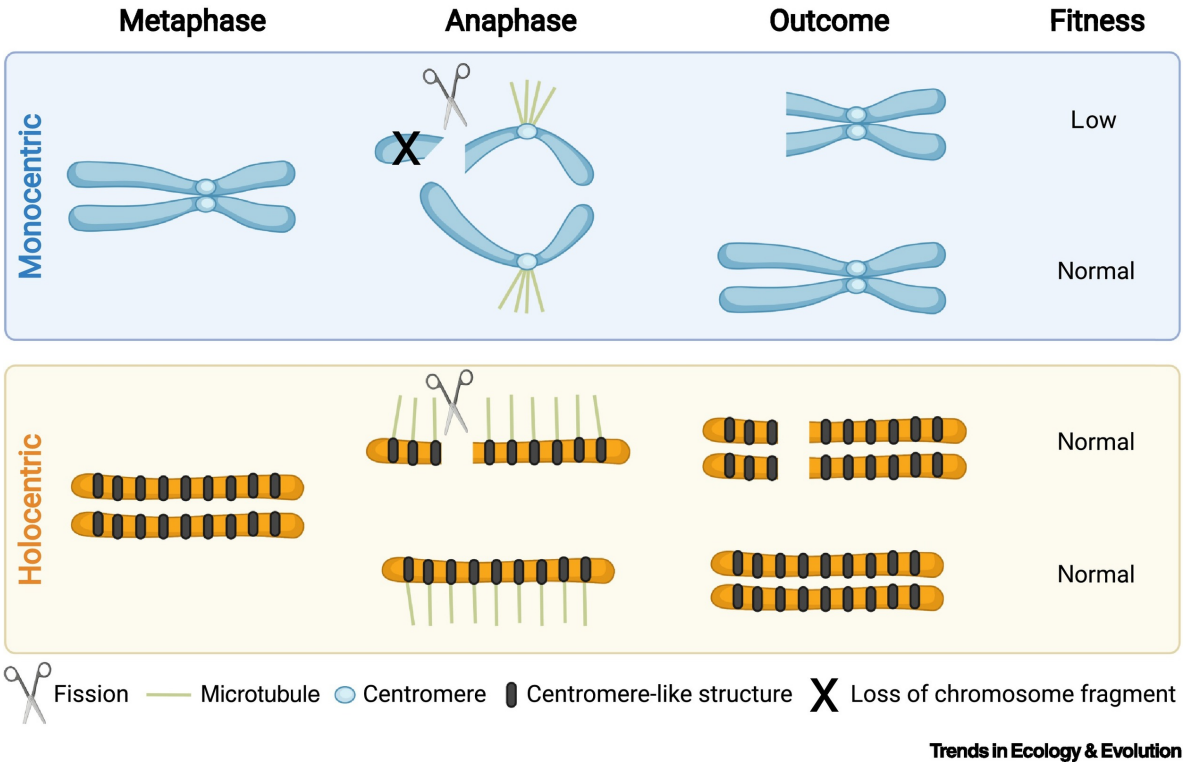


FIGURE 10 – Schéma simplifié de l'organisation des chromosomes de plantes [Lucek, Augustijnen et Escudero, 2022]. Si la fission a lieu durant l'anaphase, le fragment non attaché au centromère est perdu pour les espèces monocentriques. Les espèces holocentriques peuvent, quant à elles, conserver les deux fragments grâce aux structures centromériques le long du chromosome.

non conventionnelle (Figure 10).

Les chromosomes sont constitués d'ADN linéaire organisé autour d'histones, formant les chromatides visibles lors de la mitose. Chaque chromosome comporte des régions centromériques, riches en répétitions en tandem et en éléments transposables, dont la séquence est très variable entre espèces. Parmi ces séquences, on retrouve des répétitions satellites et des séquences spécifiques comme cenH3, qui jouent un rôle clé dans l'assemblage du kinétochore [Heslop-Harrison et Schwarzacher, 2011]. Les régions euchromatiques représentent la majorité des chromosomes et sont caractérisées par un ADN peu condensé, riche en gènes et sites de recombinaison pendant la méiose [Kellogg et Bennetzen, 2004]. À l'inverse, les régions hétérochromatiques sont fortement condensées, pauvres en gènes et concentrées en répétitions en tandem. Ces deux types de régions structurent la chromatine et influencent la dynamique de l'expression génétique et de la recombinaison.

Enfin, les télomères situés aux extrémités des chromosomes sont constitués de séquences répétitives conservées, généralement TTTAGGG, qui protègent les extrémités contre la dégradation et la fusion chromosomique, assurant ainsi la stabilité du génome sur les générations [Belyayev et al., 2023 ; Kumawat et Choi, 2023].

Les chromosomes peuvent également subir des réarrangements structurels, tels que des fusions, des fissions, des inversions ou des translocations. Ces modifications changent soit le nombre total de chromosomes, soit la structure interne d'un chromosome unique, contribuant à la diversité génomique observée chez les plantes [Kellogg et Bennetzen, 2004].

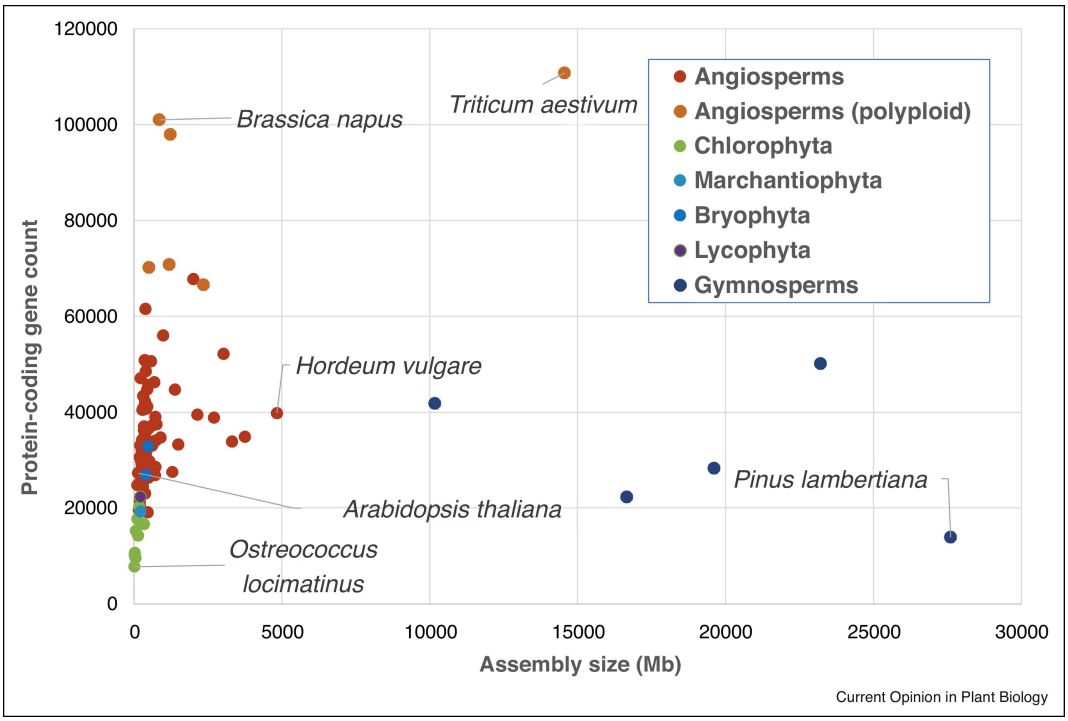


FIGURE 11 – Nombre de gènes codant des protéines par rapport à la taille des génomes chez les plantes [Kersey, 2019].

1.3 Gènes et transcrits chez les plantes

Les gènes codant des protéines

Il y a en général autour de 40,000 gènes codant des protéines dans les génomes d'espèces de plantes diploïdes [Kress et al., 2022]. Ce nombre peut varier avec par exemple chez le blé qui est héxaploïde, 107 891 gènes codant des protéines [THE INTERNATIONAL WHEAT GENOME SEQUENCING CONSORTIUM (IWGSC) et al., 2018] tandis que le génome de la plante parasite, *Sapria himalayana*, n'en contient que 13 670 [Guo, Hu et al., 2023]. En effet, le nombre de gène peut être supérieur pour les espèces ayant subies une polyploïdisation récente ou ayant un génome d'une taille importante (Figure 11) [Kersey, 2019].

Chez les plantes supérieures, une grande majorité des gènes codant des protéines contiennent des introns (∼80% chez *A. thaliana* et le riz) avec en moyenne 4 introns par gène [Morello et Breviario, 2008]. La longueur des introns chez les plantes est très variable et peut atteindre plus de 100 kb [Kress et al., 2022]. Un nombre non négligeable des gènes contenant des introns est soumis à un ou plusieurs épissages alternatifs (∼33%) [Zhou, Moshgabadi et Adams, 2011]. Ce mécanisme va réguler l'expression des gènes mais également permettre la production de plusieurs protéines différentes à partir d'un seul gène [Morello et Breviario, 2008]. Un gène et ses composants font en général entre 1 et 5 kb [Kellogg et Bennetzen, 2004]. Chez les espèces polyploïdes, il existe un effet de "silencing" des homoéologues qui va favoriser l'expression d'un gène dans un sous-génome [Voshall et Moriyama, 2020].

Les gènes vont coder des protéines ayant diverses fonctions au sein de la plante [Fang, Jiang et al., 2022]. Les gènes de résistance permettent à la plante de gérer différents stress externes comme la chaleur, le froid ou les maladies. Certains gènes sont impliqués dans la production de métabolites et donc dans les fonctions clés de la physiologie de

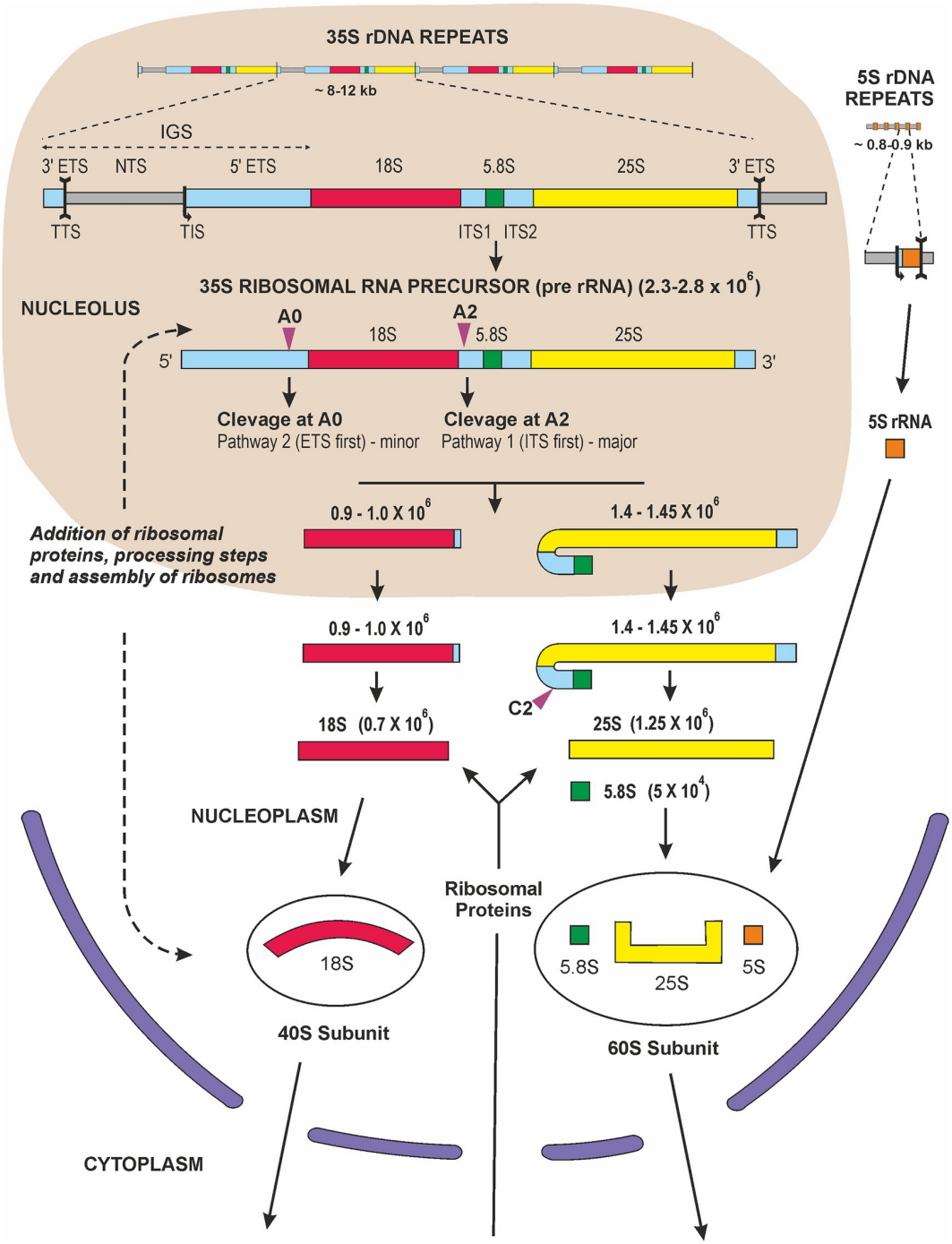


FIGURE 12 – Schéma de la structure et de la biogenèse des ribosomes dans une cellule eucaryote [Hemleben et al., 2021]. L'ARN 25S est l'équivalent chez les levures du 28S chez les autres eukaryotes.

la plante. D'autres gènes codent des protéines utilisées pour la structure de la cellule. Enfin, des gènes codent les facteurs de transcriptions qui permettent la régulation des processus évoqués précédemment. En plus de toutes ces fonctions connues, environ 1 gène sur 3 ne possède pas de fonction définie et est annoté comme "hypothetical" [Sivashankari et Shanmughavel, 2006]. Ces gènes peuvent participer à des fonctions importantes comme le développement ou la réponse aux stress [Wang, Wang et Yuan, 2023]. Malgré cela, leur caractérisation reste complexe mais essentielle pour comprendre les différents processus des plantes.

Les ARN non codants

Une grande majorité des ARN transcrits ne correspondent pas à des ARN messagers codant des protéines [Ariel et al., 2015]. Ces autres ARN sont appelés ARN non codants.

Il y a notamment les ARN ribosomaux (ARNr) et les ARN de transfert (ARNt) qui font des partis des gènes de ménages présents dans tous les génomes de plantes [Ariel et al., 2015]. En plus des ARNr des plastomes et mitogénomes, on trouve 4 ARNr dans les génomes nucléaires des plantes : 18S (petite sous-unité 40S du ribosome), 5.8S, 28S et 5S (large sous-unité 60S) (Figure 12) [Hemleben et al., 2021]. Ils permettent d'assurer la biosynthèse des protéines en vérifiant leur traduction (petite sous-unité 40S) et en formant les liaisons peptidiques (large sous-unité 60S). Chez les plantes, ces ARN ribosomaux sont organisés en deux unités répétées, l'ADNr 35S et l'ADNr 5S (Figure 12). L'ADNr 35S mesure entre 8 et 12 kb et contient les gènes codant les ARNr 18S, 5.8S et 28S tandis que l'ADNr 5S mesure environ 800 à 900 pb et contient le gène codant l'ARNr 5S (Figure 12). Ces deux unités sont organisées en clusters [Wicke, Schneeweiss et al., 2011]. Les clusters d'ADNr 35S sont localisés dans une régiondu chromosome appelée organisateur nucléaire (ou NOR pour nucleolus organizer regions) [Rosato et al., 2016]. Les régions d'ADNr ont été étudiées pour servir de marqueur

phylogénétique ou pour des analyses fonctionnelles et structurales [Hemleben et al., 2021].

Les ARNt, quant à eux, permettent de lier les codons à leur équivalent en acide aminés ; ils font entre 70 et 90 nt [Cech et Steitz, 2014]. Dans un génome de plantes, il est possible de trouver entre 18 et 59 gènes différents codant des ARNt [Mohanta et al., 2020]. Leur fréquence est très variable, les anticodons les plus abondants étant Met-CAU (5,033%), Asp-GUC (4,274%) et Asn-GUU (4,020%) tandis que les moins abondants sont Arg-GCG (0,004%), Leu-GAG (0,009%) et Sup-CUA (0,011%). La fréquence élevée de l'anticodon CAU de la méthionine est attribué au rôle d'initiateur de la traduction de cette acide aminé dans la plupart des protéines. Le nombre total de gènes ARNt peut être très variable allant de 41 chez l'algue verte unicellulaire, *Ostreococcus tauri* à 6475 chez l'ipomée, *Ipomoea nil* [Mohanta et al., 2020]. Cependant, ce nombre n'est pas proportionnel à la taille du génome.

Ils existent également d'autres ARN non codants exerçant une influence déterminante sur l'organisation, l'expression et l'évolution des génomes eucaryotes [Ariel et al., 2015]. Chez les plantes, la diversité et l'importance biologique de ces molécules sont particulièrement marquées, avec des implications majeures dans la régulation du développement, la plasticité phénotypique et les réponses aux stress biotiques et abiotiques [Cech et Steitz, 2014].

Les petits ARN représentent les familles les mieux caractérisées. Les microARN (miRNAs) sont des petits ARN régulateurs caractérisés par des précurseurs en épingle à cheveux et une maturation par DCL1. Généralement composés de 20 à 24 nucléotides et endogènes, ils assurent un contrôle post-transcriptionnel par clivage des ARNm cibles ou par inhibition de la traduction [Dugas et Bartel, 2004]. Par exemple, miR156 est un des microARN les plus étudiés chez les plantes. Il régule directement la famille des facteurs de transcription SPL (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE),

qui contrôlent de nombreux processus développementaux, notamment la transition de juvénile à adulte (stade végétatif), le timing de la floraison, la morphogenèse des feuilles et le développement de certaines structures reproductives [Guo, Xu et al., 2017]. Les petits ARN interférents (pARNi ou siRNAs) sont dérivés d'ARN double brin (endogènes ou exogènes), et sont souvent impliqués dans des mécanismes de silence transcriptionnel ou post-transcriptionnel. C'est le cas notamment des siRNA constitués de 24 nucléotides, produits en abondance dans les plantes et centraux dans la voie de méthylation de l'ADN dirigée par les ARN (RNA-directed DNA methylation, RdDM). Ces siRNA sont générés par Dicer-like 3 (DCL3) et associés à AGO4/6/9, guidant la méthylation de l'ADN sur les séquences répétées et les transposons, assurant ainsi la stabilité du génome et le contrôle épigénétique [He et al., 2009]. D'autres sous-catégories, telles que les tasiRNA et phasiRNA, dérivent de précurseurs spécifiques clivés par Dicer-like (DCL) et participent à la régulation développementale, en particulier dans les tissus reproducteurs [Liu, Teng et al., 2020].

Les longs ARN non codants (lncRNAs), définis par une taille supérieure à 200 nucléotides, forment un répertoire encore largement inexploré mais en expansion rapide. Plusieurs lncRNAs végétaux se sont révélés indispensables à la régulation transcriptionnelle et épigénétique. Un exemple emblématique est fourni par les transcrits COOLAIR et COLDAIR, impliqués dans la répression stable du gène FLOWERING LOCUS C (FLC) lors de la vernalisation chez Arabidopsis thaliana [Kim, Xi et Sung, 2017]. Ces travaux soulignent la capacité des lncRNAs à agir comme échafaudages (scaffolds) moléculaires ou comme guides pour les complexes de modification de la chromatine.

Enfin, des classes plus récentes d'ARN non codants suscitent un intérêt croissant en biologie végétale. Les ARN circulaires (circRNAs), générés par épissage rétrospectif, ont été identifiés dans plusieurs espèces végétales, avec des fonctions potentielles dans le contrôle de l'expression génique et l'adaptation aux stress [Litholdo et Fonseca, 2018].

TABLE 4 – Caractéristiques des différents ARN non codant des plantes.

Classe	**Taille (approx.)**	**Biogenèse courte**	**Fonction principale**	**Références clés (PMID)**
rRNA	100–5000 nt	Transcrit par RNA Pol I/III, maturé par endo/exonucléases	Constituants des ribosomes, traduction	31919210
tRNA	∼75–95 nt	Transcrit par RNA Pol III, modifications post-transcriptionnelles	Transport d'acides aminés	27698249
snRNA	∼100–300 nt	Transcrit par RNA Pol II/III, assemblage avec protéines Sm	Épissage (spliceosome)	27484477
snoRNA	∼60–300 nt	Maturés dans le noyau, dérivent souvent d'introns	Modification des rRNA et snRNA (méthylation, pseudouridylation)	24840859
miRNA	∼20–24 nt	Clivage précurseur par Dicer/DCL, incorporation dans RISC	Silencing post-transcriptionnel	29883606
siRNA	∼21–24 nt	Dérivent de dsRNA clivés par Dicer	Silencing génique et antiviral, RdDM	30626911
phasiRNA / tasiRNA	∼21–24 nt	Clivage de précurseurs spécifiques, DCL4/DCL5	Régulation développementale, reproduction	30787907
piRNA	∼24–31 nt	Biogenèse indépendante de Dicer, associée à PIWI	Silence des transposons (germline, animaux)	29719251
lncRNA	>200 nt	Transcrits Pol II, épissage, souvent polyadénylés	Régulation transcriptionnelle et épigénétique	35668249
circRNA	Variable	Formés par back-splicing	Régulation (miRNA sponge, signalisation)	32321745
NATs	Variable	Transcrits antisens d'autres gènes	Régulation transcriptionnelle et post-transcriptionnelle	32579934
eRNA	Variable (souvent courts)	Transcrits des enhancers	Activation transcriptionnelle locale	28439097
scaRNA	∼60–300 nt	Localisés aux corps de Cajal, maturation spécifique	Modification snRNA	15215318

D'autres transcrits, tels que les ARN antisens naturels (NATs) et les ARN d'enhancers (eRNAs), commencent également à être décrits dans les génomes de plantes [Chekanova, 2021].

L'ensemble de ces exemples témoignent de la richesse et de la complexité du paysage des ARNnc chez les plantes (Table 4). Leur étude contribue à révéler des couches de régulation jusque-là insoupçonnées, offrant une perspective nouvelle sur la manière dont les organismes végétaux coordonnent leur développement et s'adaptent à un environnement changeant.

Modification post-transcriptionelle

Après la transcription de l'ADN en ARN, un phénomène appelé "RNA editing" (édition de l'ARN) peut modifier la séquence du transcrit [Gray et Covello, 1993]. Ce phénomène semble ubiquitaire dans les mitogénomes et les plastomes des angiospermes et bien conservé entre les espèces. Il consiste surtout en la transformation d'une cytosine en un uracile et affecte, généralement, 1 à 2 % des séquences codantes [Mower, Sloan et Alverson, 2012]. Cette modification de la séquence va, dans une majorité des cas, créer un codon non-synonyme au précédent c'est-à-dire codant pour un acide aminé différent [Gray et Covello, 1993]. Le but de ce mécanisme serait de corriger les séquences d'ARN, par exemple en restaurant la séquence d'un codon pour un acide aminé conservé ou en créant un codon start ou stop [Mower, Sloan et Alverson, 2012].

1.4 Méthylome

Les modifications épigénétiques d'un génome sont l'ensemble des mécanismes transmissibles qui vont affecter l'expression des gènes sans changement dans la séquence d'ADN. La méthylation de l'ADN est une des modifications épigénétiques les plus importantes [Moore, Le et Fan, 2013]. Il s'agit de l'ajout d'un groupe méthyl (-CH3), en

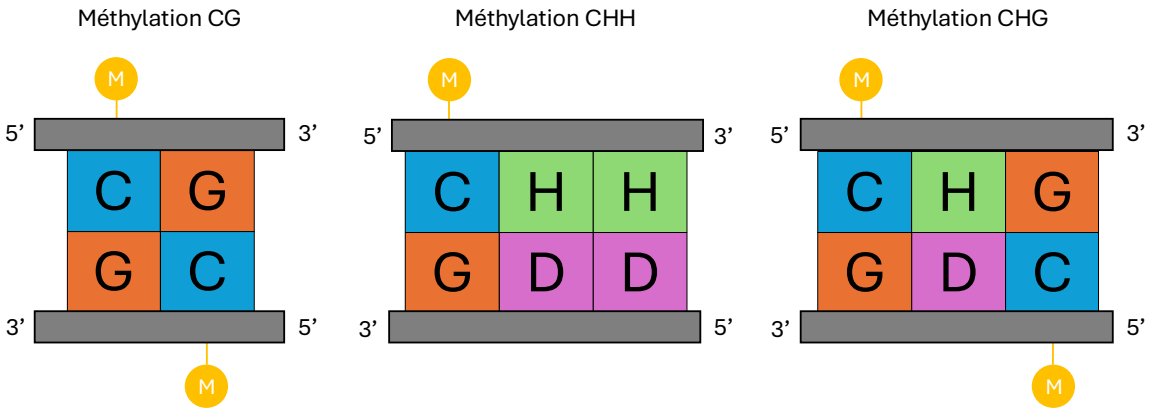


FIGURE 13 – Les différents contextes de la méthylation de la cytosine chez les plantes basés sur Bartels et al., 2018. H = A, T ou C ; D = A, T ou G.

général, au carbone n°5 de la cytosine formant ainsi une 5-méthylcytosine.

Ce processus est réversible de façon passive (durant la réplication de l'ADN) ou active (avec le mécanisme de réparation par excision de bases) [Bartels et al., 2018]. Chez les plantes, la méthylation peut avoir lieu dans trois contextes (Figure 13) : CG, CHH et CHG (où H peut désigner A, C ou T) [Bartels et al., 2018]. La méthylation CG est la plus commune chez tous les eucaryotes tandis que les deux autres se retrouvent surtout chez les plantes.

Cette modification épigénétique de l'ADN régule énormément de processus vitaux de la plante comme la morphologie des feuilles, la fertilité ou l'extinction (silencing) des éléments transposables et des répétitions [Bartels et al., 2018]. La méthylation est aussi impliquée dans les réponses aux stress biotiques et abiotiques et peut même permettre à une plante de développer une mémoire du stress conduisant à une réponse plus rapide lors de futures perturbations [Kumar et Mohapatra, 2021]. La régulation de ces nombreux mécanismes implique que la méthylation peut être spécifique à l'espèce, au tissu et même aux cellules d'un seul organisme [Vanyushin et Ashapkin, 2011].

L'expression des éléments transposables peut causer des dégâts à l'intégrité du génome, leur régulation par méthylation permet une stabilisation du génome [Law et Jacobsen, 2010]. Les variations interspécifiques du taux de méthylation sont liées à la taille du génome et plus précisément aux nombres d'éléments transposables et de répétitions qui sont fortement méthylés dans les trois contextes (CG, CHG et CHH) [Vidalis et al., 2016]. De plus, la méthylation des éléments transposables va aussi avoir un rôle dans la réponse aux stress biotiques et abiotiques [Ramakrishnan et al., 2021].

La régulation de l'expression des gènes par la méthylation des gènes se fait surtout en CG [Kumar et Mohapatra, 2021]. Chez *Arabidopsis thaliana*, environ 1/3 des gènes sont méthylés. Ces méthylations peuvent se trouver dans la région promotrice mais également dans le corps du gène [Bewick et Schmitz, 2017]. Une méthylation dans une

TABLE 5 – Caractéristiques des séquençages de deuxième et troisième génération basées sur Kchouk, Gibrat et Elloumi, 2017 ; Goodwin, McPherson et McCombie, 2016 ; Adewale, 2020.

Génération	**Société ou technologie**	**Avantages/Inconvénients**	**Taille**
Deuxième	Illumina	Très haut débit et coût faible	150-300 pb
	Roche/454 (Pyrosequencing)	Faible débit, arrêté en 2016	100-700 pb
	ABI/SOLiD	Très précis, quasiment plus utilisé	75 pb
	Ion Torrent	Peu précis, séquençage rapide	200-600 pb
Troisième	PacBio	Taux d'erreurs entre 0,1 et 1%	5-200 kb
	Oxford Nanopore	Taux d'erreurs de 5%	5-2000 kb

région promotrice est souvent associée à la répression de l'expression du gène associé mais dans certains cas elle va au contraire booster la transcription du gène [Kumar et Mohapatra, 2021]. Pour ce qui est des méthylations dans le corps du gène leurs fonctions restent pour le moment vagues.

La méthylation de l'ADN a des rôles importants dans la régulation de processus précis de la vie des plantes. Par exemple, elle est très associée à la maturation des fruits où elle va activer les gènes de maturation et également réduire l'expression des inhibiteurs de la maturation [Kumar et Mohapatra, 2021]. Chez le soja, la germination va également induire une augmentation de la méthylation CHH qui peut passer de 6 à 11% [Bartels et al., 2018].

2 Stratégies de séquençages et bioinformatiques en génomiques végétales

2.1 Méthode de séquençage de l'ADN

Le choix de la méthode de séquençage va avoir un impact important sur l'assemblage du génome qui suivra. Deux types de méthodes de séquençage de l'ADN sont utilisées aujourd'hui (Table 5) : les méthodes de deuxième génération (aussi appelée NGS pour Next Generation Sequencing ; 2005-2015) et de troisième génération (2010-aujourd'hui) [Kchouk, Gibrat et Elloumi, 2017].

Technologies de deuxième génération

À partir de 2005, la deuxième génération de séquençage a permis de réduire fortement le coût du séquençage et d'en augmenter radicalement la vitesse et le débit [Kchouk, Gibrat et Elloumi, 2017]. Les méthodes de deuxième génération produisent des millions

de shorts reads (SR) ; c'est-à-dire des séquences de tailles courtes (entre 75 et 700 pb) et avec un taux d'erreurs faible [Adewale, 2020]. La première technologie de deuxième génération ayant été développée est le pyroséquençage par Roche/454 en 2005. Cette technologie est basée sur le séquençage par synthèse qui va détecter l'incorporation d'un nucléotide par une polymérase lors de l'allongement d'un brin d'ADN [Goodwin, McPherson et McCombie, 2016]. Le séquençage de Roche/454 permet de générer des reads assez longs (jusqu'à 700 pb) et avec un taux d'erreurs de 1% [Kchouk, Gibrat et Elloumi, 2017]. Cette technologie a cependant été arrêté en 2016, suite à un manque de compétitivité en termes de coût et de rendement [Goodwin, McPherson et McCombie, 2016]. La technologie Ion Torrent, lancée en 2010, utilise aussi le séquençage par synthèse. Elle est très rapide et permet de produire des reads jusqu'à 600 pb avec un taux d'erreurs de $\approx 1\%$ [Kchouk, Gibrat et Elloumi, 2017].

Une autre technique de séquençage est le séquençage par ligation qui va se baser sur le spectre d'émission d'un fluorophore, préalablement lié à un oligonucléotide, pour identifier les bases [Goodwin, McPherson et McCombie, 2016]. Lancée en 2007, la technologie ABI/SOLiD utilise cette technique pour générer des reads avec un taux d'erreurs faible ($\approx 0,1\%$) mais avec une taille plus petite (75 bp) [Kchouk, Gibrat et Elloumi, 2017]. La taille trop courte de ces reads limitant son utilisation, cette technologie est très peu utilisée [Goodwin, McPherson et McCombie, 2016]. Enfin, Illumina est la technologie la plus utilisée aujourd'hui. Elle se base sur la technique de séquençage par synthèse et permet la génération de reads entre 150 et 300 pb avec un taux de précision élevé (99,5%) et un faible coût [Goodwin, McPherson et McCombie, 2016]. Les génomes de la tomate sauvage (*Solanum pimpinellifolium*) et de la pomme de terre (*Solanum tuberosum*) ont pu être assemblés en utilisant les technologies Illumina et/ou Roche/454 [Razali et al., 2018 ; Xu, Pan et al., 2011]. Cependant, les shorts reads rendent difficile la résolution de génomes complexes comportants des éléments transposables, de longues

répétitions ou des variants structuraux [Goodwin, McPherson et McCombie, 2016].

Technologies de troisième génération

Les méthodes de troisième génération produisent des longs reads (LR) qui permettent de résoudre les problèmes d'assemblage des technologies de deuxième génération. Contrairement aux SR, les LR permettent de bien gérer les zones répétées de grandes tailles et les différentes conformations [Goodwin, McPherson et McCombie, 2016]. En plus, il n'y pas besoin d'amplification par PCR, ce qui rend ces méthodes plus rapides et peu coûteuses [Kchouk, Gibrat et Elloumi, 2017]. Deux principales technologies pour produire des LR existent : la "Single Molecule Real-Time" (SMRT) de Pacific Biosciences et le séquençage nanopore par Oxford Nanopore [Jiao et Schneeberger, 2017]. Commercialisée en 2011, la technologie SMRT utilise un séquençage en temps réel avec une polymérase fixée au fond de puits qui permet le passage du brin d'ADN et la lecture des bases [Goodwin, McPherson et McCombie, 2016]. Généralement, les reads générés font entre 5 et 30 kb même s'ils peuvent parfois atteindre 200 kb avec un taux d'erreurs de 1% [Adewale, 2020 ; Dijk et al., 2023]. Les reads HiFi de PacBio peuvent même atteindre une précision similaire aux technologies de deuxième génération (99,9%) mais avec des longueurs comprises entre 10 et 25 kb.

La deuxième technologie développée en 2014 par Oxford Nanopore va faire passer le brin d'ADN à travers un nanopore où le courant ionique va être détecté ce qui permettra la lecture des bases [Kchouk, Gibrat et Elloumi, 2017]. Les reads ainsi générés peuvent atteindre des tailles jusqu'à 2 Mb mais avec un taux d'erreurs autour de 5% [Adewale, 2020 ; Dijk et al., 2023]. La technologie nanopore permet la génération de reads de très grandes tailles tandis que la technologie SMRT possède des avantages sur la grande précision de ces long reads. C'est cette dernière qui est la plus utilisée bien que plus coûteuse [Adewale, 2020]. Les génomes de *Houttuynia cordata* et *Cardamine*

chenopodiifolia ont notamment été assemblés avec des reads HiFi [Huang et al., 2024 ; Emonet et al., 2024]. Le génome de *Rhodomyrtus tomentosa* a été, quant à lui, assemblé avec des reads issus de la technologie d'Oxford Nanopore [Detcharoen, Bumrungsri et Voravuthikunchai, 2023]. Les technologies long reads permettent de bien gérer les régions répétées et de faciliter grandement les assemblages de génomes. Cependant, d'autres technologies permettent de les compléter en fournissant des informations sur la localisation des reads au sein du génome, notamment pour l'identification de variants structuraux [Jiang et al., 2024]

Autres technologies

Ces technologies alternatives utilisent surtout des séquençages SR. Les plus communes sont les méthodes Hi-C (High-throughput chromosome conformation capture) et Linked Reads (notamment celle de 10X Genomics) [Srikanth et al., 2020]. La technologie Hi-C marque la proximité physique des fragments d'ADN dans leur conformation naturelle [Jiao et Schneeberger, 2017]. Cette information, incluse dans les reads Hi-C, permet d'organiser les contigs en chromosome lors d'une étape de scaffolding [Jiang et al., 2024]. Le séquençage Hi-C a notamment été utilisé chez le blé [THE INTERNATIONAL WHEAT GENOME SEQUENCING CONSORTIUM (IWGSC) et al., 2018]. La technologie 10X permettant de créer des linked reads en coupant l'ADN en longs fragments de plus de 100 kb qui vont chacun être marqué à l'aide de différents barcodes [Jiang et al., 2024]. Ces longs fragments vont ensuite être coupés et séquencés avec une technologie short reads classique. Les reads appartenant au même long fragment d'ADN initial vont ainsi posséder le même barcode, ce qui permettra de faciliter ensuite l'assemblage. La technologie 10X peut à la fois faciliter le scaffolding mais également assembler *de novo* avec une grande précision même pour des gros génomes d'espèces polyploïdes [Jiang et al., 2024]. Cette technologie a notamment permis d'assembler

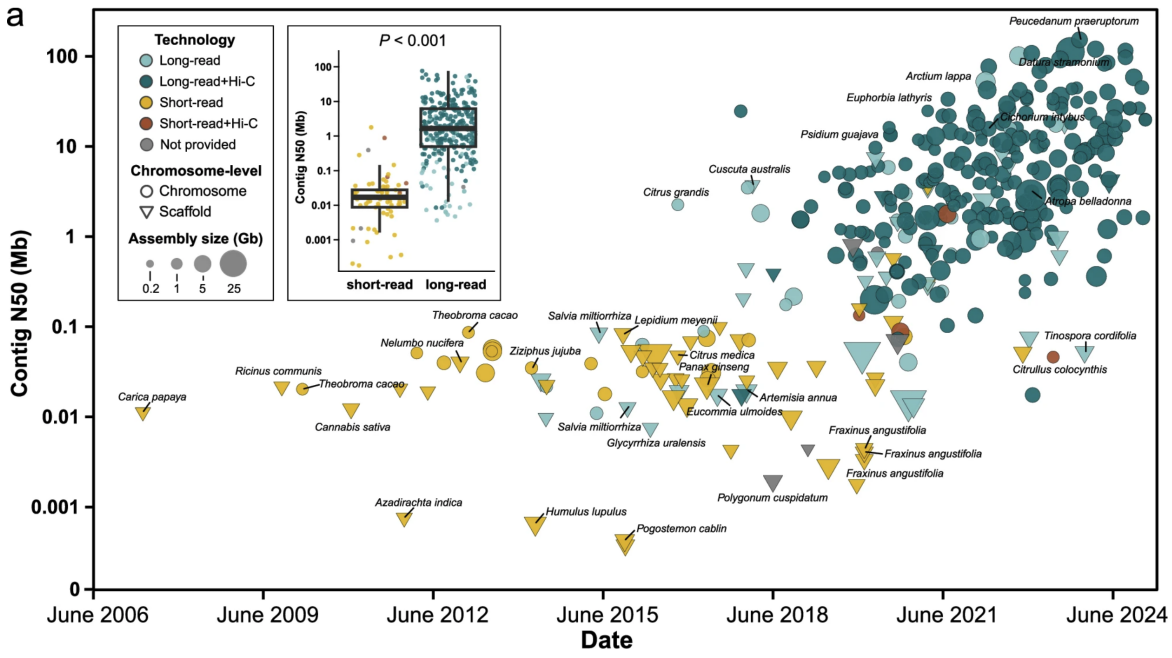


FIGURE 14 – Évolution des technologies de séquençages utilisées pour l'assemblage de génomes de plantes médicinales [Cheng, Wang et al., 2025].

TABLE 6 – Liste non exhaustive d'assembleurs de génome.

Type de reads	Algorithme	Nom de l'assembleur	Dernière mise à jour	Référence
SR	DBG	MEGAHIT	2019	[Li, Liu et al., 2015]
		ABySS	2024	[Simpson et al., 2009]
		SPAdes	2025	[Prjibelski et al., 2020]
		SOAPdenovo	2020	[Luo et al., 2012]
LR	OLC	Canu	2024	[Koren et al., 2017]
		FALCON	2016	[Chin et al., 2016]
	DBG-like	Flye	2025	[Kolmogorov et al., 2019]
		wtdg2	2019	[Ruan et Li, 2020]
Hybride	DBG	HYBRIDSPAdes	2020	[Antipov et al., 2016]
		HASLR	2024	[Haghshenas et al., 2020]

le génome du noisetier commun (*Coryllus avellana*) [Pavese et al., 2021]. Une autre technologie également disponible est la méthode d'optical mapping qui va utiliser des motifs enzymatiques fluorescents pour localiser les séquences [Dijk et al., 2023]. Cette méthode a été utilisée en combinaison d'un séquençage nanopore chez le sorgho (*Sorghum bicolor*) [Deschamps et al., 2018]. Ces technologies sont très intéressantes à utiliser en combinaison avec d'autres séquençages pour reconstruire le génome d'espèces complexes.

Évolution des technologies de séquençage utilisées dans les assemblages de génome de plantes

L'arrivée des technologies de deuxième génération a permis d'augmenter le nombre de génomes de plantes assemblés [Cheng, Wang et al., 2025]. Cependant, ce sont les technologies de troisième génération qui ont vraiment révolutionné la génomique végétale en augmentant drastiquement le nombre de génomes assemblés ainsi que leur complétude (Figure 14). Au cours des trois dernières années, 98% des génomes ont été assemblés avec ces technologies de troisième génération, souvent en combinaison avec la technologie Hi-C (Figure 14) [Cheng, Wang et al., 2025].

2.2 Assemblage *de novo*

Les assembleurs permettent d'aider à la reconstruction des génomes en fonction du type de séquençage (Table 6). Deux principaux algorithmes d'assemblage sont utilisés : le graph de De Bruijn (DBG) et l'overlap layout consensus (OLC).

Graphe de De Bruijn

Le DBG va transformer les séquences en une suite de k-mers (sous-séquence de longueur k) [Sohn et Nam, 2016]. Ces k-mers sont connectés par le chevauchement de

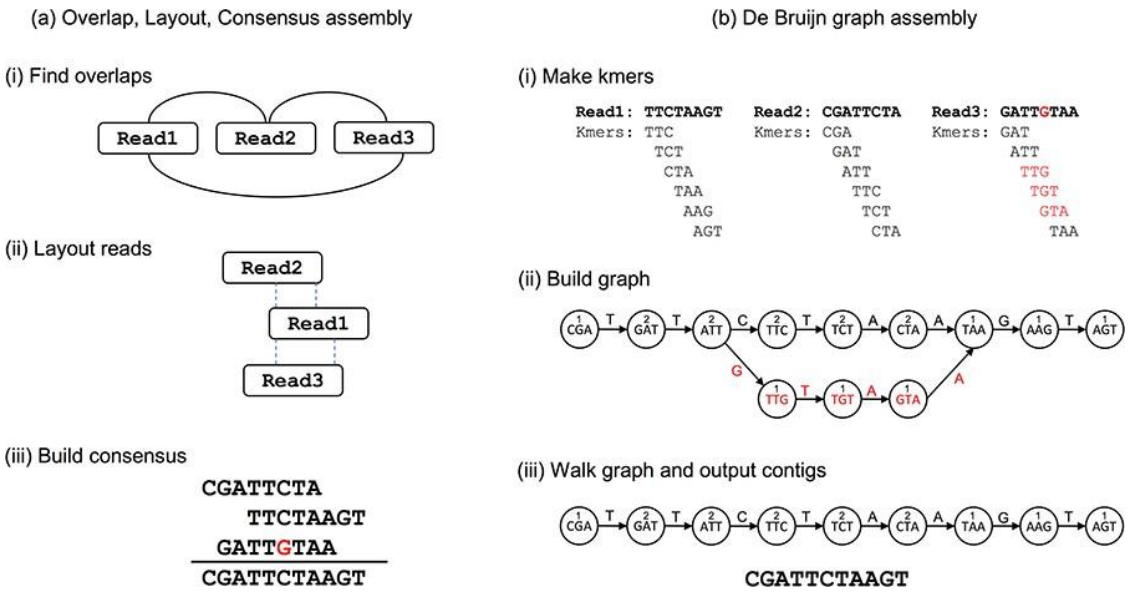


FIGURE 15 – Les deux algorithmes d'assemblage de génome : (a) Overlap Layout Consensus (OLC) et (b) Graph de De Bruijn (DBG) [Ayling, Clark et Leggett, 2020].

préfixe et suffixe aboutissant à la construction d'un graphe (Figure 15). Le chemin optimal du graphe est ensuite identifié et permettra de générer des contigs. Cet algorithme est utilisé pour l'assemblage de short reads car il ne permet pas de bien gérer les erreurs dans les séquences [Rice et Green, 2019]. Les assembleurs MEGAHIT [Li, Liu et al., 2015], SPAdes [Prjibelski et al., 2020], ABySS [Simpson et al., 2009] ou SOAPdenovo [Luo et al., 2012] utilisent le DBG. Cependant, l'algorithme ne permet pas une bonne gestion des séquences répétées [Sohn et Nam, 2016]. En effet, les répétitions créent des ambiguïtés et des gaps dans le graph impactant l'assemblage final.

Overlap layout consensus

Avec l'arrivée des technologies de séquençage de troisième génération, l'algorithme OLC a été favorisé car il permet une meilleure gestion de longs reads avec beaucoup d'erreurs [Sun, Shang et al., 2022]. Il va d'abord trouver des chevauchements entre les reads puis regrouper ces reads pour enfin créer un consensus avec la séquence la plus probable (Figure 15) [Ayling, Clark et Leggett, 2020]. Des assembleurs pour les LR comme Canu [Koren et al., 2017] ou FALCON [Chin et al., 2016] utilisent l'algorithme OLC. Cet algorithme demande beaucoup plus de temps de calcul et de mémoire que le DBG surtout pour la gestion de séquences répétées [Li, Chen et al., 2012]. C'est pourquoi certains assembleurs LR ont été développés en utilisant une variante de l'algorithme DBG gérant les erreurs dans les LR comme Flye [Kolmogorov et al., 2019] ou wtdbg2 [Ruan et Li, 2020]. Ces assembleurs sont plus rapides que les assembleurs OLC.

Assemblage hybride

Il est aussi possible de combiner les séquençages LR et SR qui sont complémentaires pour réaliser un assemblage hybride [Sun, Shang et al., 2022]. Un assemblage de ce type permet à la fois une bonne gestion des nombreuses séquences répétées tout en

ayant un nombre d'erreurs limitées. Des assembleurs hybrides comme HYBRIDSPAdes [Antipov et al., 2016] ou HASLR [Haghshenas et al., 2020] peuvent permettre une reconstruction fiable du génome [Štorchová et Krüger, 2024]. Il est aussi possible de corriger les LR avec les SR avant de les assembler en utilisant Ratatosk [Holley et al., 2021] ou FMLRC [Wang, Holt et al., 2018]. Par ailleurs, il est aussi recommandé, pour faciliter l'assemblage, de combiner plusieurs assembleurs [Štorchová et Krüger, 2024].

2.3 Spécificité de l'assemblage d'un génome d'organelle de plantes

Les génomes des organelles de plantes possèdent quelques particularités qui doivent être prises en compte lors de leur assemblage et demandent donc des méthodes adaptées.

Stratégies de séquençage

Comme présenté auparavant, lors d'un séquençage d'un génome de plante, l'ADN peut provenir de trois sources : mitochondrie, chloroplaste et noyau. Afin de récupérer uniquement les reads des organelles, il est possible de les isoler avant le séquençage. Cela a été réalisé pour les assemblages des mitogénomes du chanvre du Deccan (*Hibiscus cannabinus*) et de la moutarde blanche (*Brassica hirta*) [Liao et al., 2018 ; Palmer et Herbo, 1987]. Cependant, cela peut s'avérer long et compliqué et demande un protocole spécifique. L'autre option, plus utilisée aujourd'hui, est un séquençage de tout le génome (WGS = Whole Genome Sequencing) [Štorchová et Krüger, 2024]. Ici, des séquences du noyau, du chloroplaste et de la mitochondrie seront présentes. Une fois le séquençage effectué, une sélection des reads des organelles devra être effectuée pour permettre leur assemblage.

Assemblage sur référence

Cette sélection peut être faite en effectuant un assemblage sur référence en utilisant le génome déjà assemblé d'une espèce proche. Cela a été fait pour le plastome de trois espèces d'orchidées du genre *Paraphalaenopsis* [Chen, Wang et al., 2023]. Cependant, cela peut induire des erreurs si le plastome n'a pas une structure conventionnelle (par exemple, la perte d'une copie d'une IR) ou si la référence utilisée présente des erreurs [Wang, Schalamun et al., 2018]. Pour les plastomes, un assemblage *de novo* est préférable s'il n'y a pas de référence suffisamment proche [Twyford et Ness, 2017]. Pour les mitogénomes, l'assemblage *de novo* est très fortement recommandé à cause de leur variabilité trop importante [Zardoya, 2020].

Sélection *in silico* des reads des organelles

Les reads des organelles devront donc être sélectionnés sans assemblage sur référence. Pour les plastomes, la sélection des reads *in silico* peut se faire en mappant directement les reads sur un ou plusieurs plastome d'espèces assez proche phylogénétiquement. Cette méthode a été utilisé pour les assemblages des plastomes de *Picea rubens* [Parmar et al., 2022] et de cinq espèces du genre *Vaccinium* [Fahrenkrog et al., 2022]. La bonne conservation des plastomes permet au mapping de récupérer toutes les séquences voulues et d'ensuite les assembler *de novo*. Le manque de conservation des séquences intergéniques du mitogénome empêche une sélection des reads par mapping direct sur d'autres mitogénomes. La sélection se fera plutôt en se basant sur des séquences conservées comme les CDS (Coding DNA Sequence = séquences des gènes codant des protéines) dans les mitogénomes d'autres espèces, même éloignées phylogénétiquement [Yin, 2022 ; Fang, Li et al., 2021].

Assembleurs spécialisés

En plus des assembleurs déjà présentés précédemment, il existe des assembleurs spécialisés dans l'assemblage de génomes d'organelles comme GetOrganelle [Jin et al., 2020], MitoZ [Meng et al., 2019] ou NovoPlasty [Dierckxsens, Mardulyn et Smits, 2017]. GetOrganelle et NovoPlasty sont très efficaces pour la reconstruction de plastome grâce à la bonne conservation de ces derniers [Freudenthal et al., 2020]. Cependant, pour les mitogénomes, GetOrganelle et MitoZ, ne permettant que l'utilisation de SR, sont surtout efficaces pour les mitogénomes d'animaux qui sont moins complexes. Étant donné le nombre important de séquences répétées dans les génomes de plantes, l'utilisation uniquement de SR ne permettra pas un assemblage complet [Wynn et Christensen, 2019]. L'utilisation de méthode hybride ainsi que de plusieurs assembleurs est à privilégier [Štorchová et Krüger, 2024].

2.4 Spécificité de l'assemblage d'un génome nucléaire de plantes

L'assemblage de génome nucléaire de plantes représente un véritable challenge à cause de leur taille, leur taux de répétitions élevé ou encore la polyploïdie [Jiao et Schneeberger, 2017]. Il existe cependant des méthodes de séquençages et bioinformatiques permettant de limiter les erreurs et d'arriver, au moins, à un draft du génome. En effet, une majorité des génomes actuellement assemblés reste très fragmenté [Kress et al., 2022].

Séquençage et estimation de la taille et de la ploïdie

Le séquençage est un point important pour pouvoir réaliser un assemblage de qualité par la suite. Il est tout d'abord fortement recommandé d'utiliser plusieurs technologies de séquençage notamment de combiner LR et SR [Kersey, 2019 ; Jiao et Schneeberger, 2017]. Les séquençages 10X et Hi-C sont également très intéressants et faciliteront

l'assemblage comme expliqués précédemment [Jiang et al., 2024]. Il est aussi important d'estimer la taille ainsi que la ploïdie du génome étudié afin de pouvoir choisir la bonne stratégie selon le cas de figure [Jeon et Kim, 2024]. Cela peut se faire par cytométrie de flux. La difficulté majeure pour l'assemblage de génomes de plantes est la polyploïdie [Jeon et Kim, 2024]. Pour la gérer, il est possible de séquencer soit un haploïde naturel ou généré *in vitro* de l'espèce étudiée soit un de ces progéniteurs diploïdes (s'ils sont connus) [Kyriakidou et al., 2018]. Il pourra ensuite servir de référence pour l'assemblage.

Gestion de la polyploïdie *in silico*

Le principal problème de la polyploïdie va être de pouvoir attribuer chaque séquence au bon sous-génome pour les espèces polyploïdes [Jeon et Kim, 2024]. Des algorithmes d'haplotyping qui vont attribuer des variants/allèles à des chromosomes permettent cela. Cependant ils demandent beaucoup de profondeur de séquençage et de temps de calcul [Kyriakidou et al., 2018]. HiFiasm est un assembleur "haplotype-resolved" qui va permettre la construction de génome polyploïde avec des long reads HiFi, Oxford Nanopore ou Hi-C [Cheng, Concepcion et al., 2021]. Il a été utilisé pour assembler le génome de l'espèce tétraploïde, *Sauropus androgynus* [Xia et al., 2024].

L'utilisation de génomes diploïdes des progéniteurs est également pertinente ici, puisque chaque sous-génome doit être très similaire aux génomes de leur parent [Jeon et Kim, 2024]. Bien sûr, les génomes de ces espèces doivent déjà avoir été assemblés et être de suffisamment bonne qualité. Une autre méthode est basée sur les k-mers et les LTR qui sont caractéristiques des différents sous-génomes [Jeon et Kim, 2024]. Cependant, cette méthode fonctionne surtout pour les allopolyploïdes et moins pour les autopolyploïdes.

L'assemblage d'un pan-génome contenant toutes les variations de chaque séquence individuelle est aussi une solution puisque son assemblage permettra de représenter le

génome polyploïde [Kyriakidou et al., 2018]. Cependant, sa reconstruction pour un génome polyploïde est très complexe. Enfin, suivant les espèces, l'utilisation d'algorithmes non spécialisés pour les espèces polyploïdes peut suffire si les sous-génomes sont assez différents. Pour les autopolyploïdes, cela peut s'avérer plus compliqué et ils sont en général assemblés comme des diploïdes en annotant les sites polymorphiques [Jiao et Schneeberger, 2017].

L'assemblage de génome de plantes notamment polyploïde est très complexe et permet rarement d'obtenir un génome complet. C'est pourquoi il peut être important d'en évaluer la qualité et la complétude.

Évaluation de l'assemblage d'un génome

L'un des outils les plus utilisés est BUSCO (Benchmarking Universal Single-Copy Orthologs) [Manni et al., 2021]. Il utilise un set de gènes conservés et présents normalement en une seule copie pour évaluer la complétude du génome. Des statistiques comme le N50 (Nombre de contigs minimum nécessaire pour atteindre 50% de la longueur totale de l'assemblage) ou le L50 (Longueur du dernier contigs nécessaire pour atteindre 50% de la longueur totale de l'assemblage) peuvent permettre une évaluation qualitative du génome [Kress et al., 2022].

2.5 Décontamination

Une autre étape cruciale dans la reconstruction d'un génome est sa décontamination.

Origine des contaminants

Lors du séquençage, il est possible qu'il y ait l'introduction de séquences non voulues et contaminantes [Cornet et Baurain, 2022]. Ces contaminations peuvent avoir lieu à plusieurs étapes du séquençage. Ils peuvent être dû à des contaminations de micro-

organismes introduits involontairement ou endosymbiotes de l'organisme étudié. Les contaminations humaines lors de l'extraction ou du séquençage sont aussi fréquentes [Cornet et Baurain, 2022].

Stratégies de décontamination

Ces séquences contaminantes peuvent poser des problèmes au cours de l'assemblage et lors des analyses entrainant ainsi de fausses conclusions et la publication de séquences erronées dans les bases de données [Lupo et al., 2021]. Pour éviter cela, des outils utilisant différentes approches permettent la décontamination du génome.

La première approche est l'alignement direct des séquences avec celles présentes dans des bases de données de référence. Pour cela, il est possible d'utiliser Blast [Altschul et al., 1990], Diamond [Buchfink, Reuter et Drost, 2021] ou FCS GX [Astashyn et al., 2023]. Ce dernier a été développé spécifiquement dans ce but. Cette approche est efficace mais est soumise aux erreurs potentielles dans les bases de données [Lupo et al., 2021].

Une autre approche utilise les k-mers. Elle est notamment implémentée dans Kraken [Wood, Lu et Langmead, 2019], Kaiju [Menzel, Ng et Krogh, 2016] ou CLARK [Ounit et al., 2015]. Cela permet une classification très rapide des séquences mais toujours basé sur des données existantes.

Une dernière approche permet de quasiment s'affranchir de l'utilisation de base de données. Les programmes utilisant cette approche, comme BlobTools [Laetsch et Blaxter, 2017] ou Anvi'o [Eren et al., 2015], vont plutôt se baser sur des caractéristiques de l'ADN comme le contenu en GC ou la couverture de séquençage. Ces programmes nécessitent tout de même des informations taxonomiques pour aider la classification et ne permettent pas de travailler sur de gros jeu de données [Cornet et Baurain, 2022].

La détection de contaminant peut être compliqué, surtout pour les espèces n'ayant aucun représentant proche dans les bases de données [Cornet et Baurain, 2022]. Enfin,

certains transferts horizontaux de gènes peuvent être confondus avec des contaminations. Malgré cela, la décontamination reste une étape clé de l'assemblage d'un génome.

2.6 Annotation d'un génome

L'annotation est une partie importante de l'assemblage d'un génome permettant ensuite des analyses plus poussées. Cependant, ce processus est long et demande de combiner plusieurs analyses [Kress et al., 2022].

Prédiction des structures des gènes codant des protéines

La première étape pour annoter des gènes codant des protéines consistent souvent à prédire leur localisation dans le génome [Hoff et Stanke, 2015]. Il s'agit de trouver les structures qui pourraient correspondre à des gènes. Il est possible d'utiliser des programmes spécifiques comme AUGUSTUS [Stanke et al., 2008] ou SNAP [Korf, 2004]. Ils ont cependant besoin d'avoir des exemples de gènes de l'espèce analysée ou d'espèces proches afin de comprendre les structures particulières de gènes pouvant être présentes [Hoff et Stanke, 2015]. Il est aussi possible d'utiliser des données de séquençage de l'ARN pour prédire les structures et les séquences des gènes codant des protéines. Cependant, les transcrits ne sont pas toujours traduits et la séquence des transcrits ne permet pas de prédire la partie codante (ORF pour Open Reading Frame) [Hoff et Stanke, 2015].

Annotation fonctionnelle des gènes codant des protéines

L'annotation d'un gène codant une protéine peut se faire en se basant sur sa similarité avec d'autres séquences dans des bases de données en utilisant BLAST par exemple [Altschul et al., 1990]. C'est la méthode classiquement utilisée. Cependant elle peut poser des problèmes [Sivashankari et Shanmughavel, 2006]. Tout d'abord, il n'est pas toujours possible de trouver une séquence similaire et par ailleurs, il est possible d'avoir

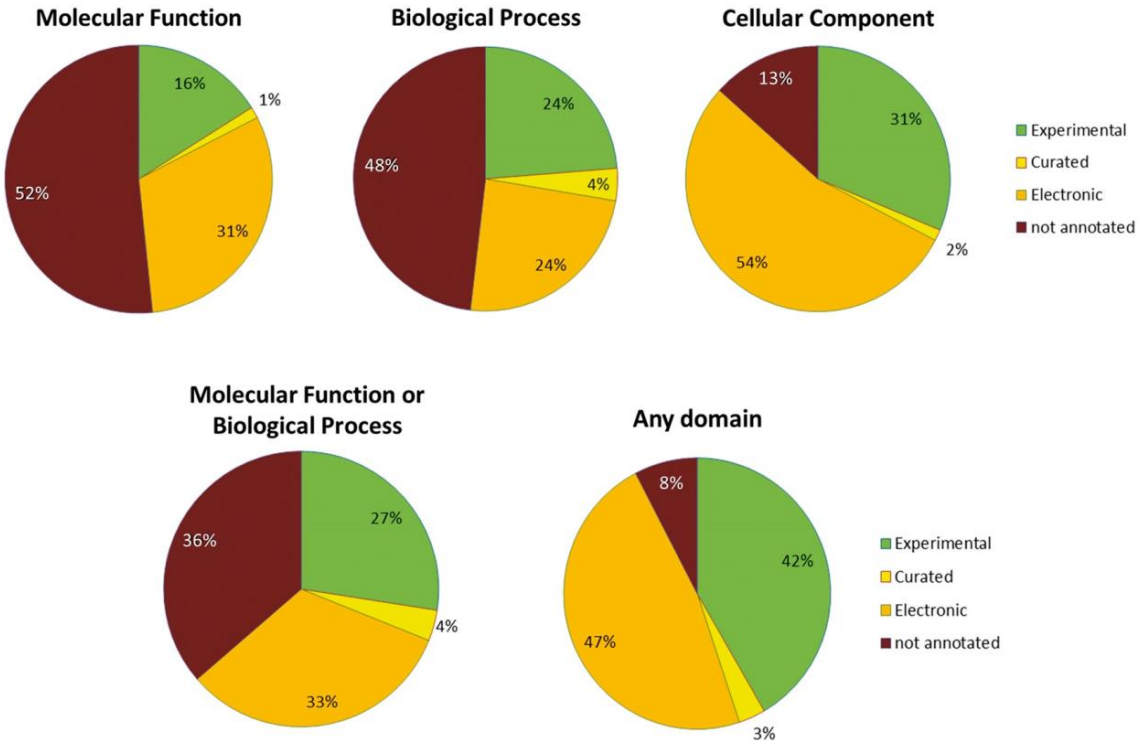


FIGURE 16 – Nombre de gènes annotés chez *Arabidopsis thaliana* basé sur une analyse GO [Bolger, Arsova et Usadel, 2018]. Les catégories ont été crées selon les "evidence code" c'est-à-dire les preuves des annotations. La catégorie "Experimental" contient les gènes annotés par les "evidence code" IDA (inferred from direct assay), IMP (inferred from mutant phenotype), IGI (inferred from genetic interaction), IPI (inferred from physical interaction) or IEP (inferred from expression profile). La catégorie "Curated" contient les gènes annotés par les "evidence codes" IC (Inferred by Curator), NAS (Non-traceable Author Statement) and TAS (Traceable Author Statement) mais ne contenant aucune annotation de la catégorie "Experimental". La catégorie "Electronic" contient les gènes annotés par les "evidence codes" ISS (Inferred from Sequence or Structural Similarity), ISO (Inferred from Sequence Orthology), ISM (Inferred from Sequence Model), IBA (Inferred from Biological Aspect of Ancestor), RCA (Inferred from Reviewed Computational Analysis) or IEA (Inferred from Electronic Annotation), mais ne contenant aucune annotation des catégories "Experimental" ou "Curated".

une annotation erronée. En effet, très peu d'annotations ont été validées expérimentalement (Figure 16) [Kersey, 2019]. De plus, le transfert de fonction basé uniquement sur une similarité de séquence peut créer de fausses annotations [Bolger, Arsova et Usadel, 2018].

L'annotation d'un gène codant une protéine peut aussi se faire en se basant sur les domaines protéiques conservés. Cela permet une annotation plus sensible et donc d'annoter plus de gènes [Bolger, Arsova et Usadel, 2018]. Les domaines protéiques ne renseignent pas toujours sur la fonction exacte de la protéine mais peuvent donner des indications sur celle-ci. InterProScan [Jones, Binns et al., 2014] permet ce genre d'analyse en se basant sur de multiple base de données comme Pfam, PANTHER ou NCBIFAM. InterProScan va aussi annoter des termes GO (Gene Ontology). Les termes GO sont les annotations fonctionnelles les plus utilisées et sont classées en trois catégories : les processus biologiques (BP pour Biological Process), les fonctions moléculaires (MF pour Molecular Function) et les composants cellulaires (CC pour Cellular Component) [The Gene Ontology Consortium et al., 2023]. Une autre ressource très utilisée est le KEGG (Kyoto Encyclopedia of Genes and Genomes) [Kanehisa, Furumichi et al., 2025] qui contient les réseaux moléculaires des fonctions importantes. Les annotations se font grâce à BlastKOALA et GhostKOALA [Kanehisa, Sato et Morishima, 2016] puis l'outil KEGG Mapper [Kanehisa et Sato, 2020] permet ensuite de reconstruire les voies métaboliques.

Malgré l'existence de tous ces outils, il est très improbable de pouvoir annoter plus de deux tiers des gènes du génome d'une plante [Sivashankari et Shanmughavel, 2006]. En effet, même pour le génome d'*Arabidopsis thaliana*, le mieux annoté chez les plantes, seulement 64% des gènes possèdent au moins un terme GO correspondant à une fonction moléculaire ou à un processus biologique (Figure 16) [Bolger, Arsova et Usadel, 2018].

Annotation des autres éléments d'un génome

L'annotation d'un génome ne concerne pas seulement les gènes codant des protéines. Les gènes codant des ARN sont aussi présents dans les génomes. La base de données Rfam [Kalvari et al., 2021] permet de répertorier un certain nombre d'ARN non codant comme les ARNr ou les miARN. Pour les ARNt, l'outil le plus communément utilisé est tRNAscan-SE [Chan et al., 2021]. L'annotation de ces éléments se fait par homologie de séquence mais également par comparaison de leur structure secondaire. Les éléments transposables, quant à eux, peuvent être annotés de différentes façons. Comme les gènes codant des protéines, il est possible de les annoter en se basant sur une simple similarité de séquence ou sur leur structure par rapport à une base de données [Loreto et al., 2024]. De nombreux outils permettent l'annotations des éléments transposables et autres éléments répétés comme RepeatMasker [Nishimura, 2000], RepeatModeler [Flynn et al., 2020], DANTE [Novák et al., 2024], REPuter [Kurtz et Schleiermacher, 1999] ou Tandel repeats finder [Benson, 1999]. Une annotation *de novo* est aussi possible et a l'avantage de ne pas avoir besoin de base de données mais les résultats peuvent perdre en fiabilité [Storer et al., 2022]. La meilleure méthode est de combiner ces différentes approches afin d'obtenir une annotation précise et plus juste [Loreto et al., 2024].

2.7 Transcriptome

Le transcriptome correspond à l'ensemble des gènes transcrits en ARN. Il permet notamment l'étude de l'expression des gènes et de l'épissage alternatif mais cela nécessite plusieurs étapes.

Conception expérimentale d'un transcriptome

La première étape consiste à déterminer l'objectif de l'étude. En effet, la génération de séquençage de l'ARN (ou RNA-Seq) peut avoir plusieurs buts [Han, Gao et al.,

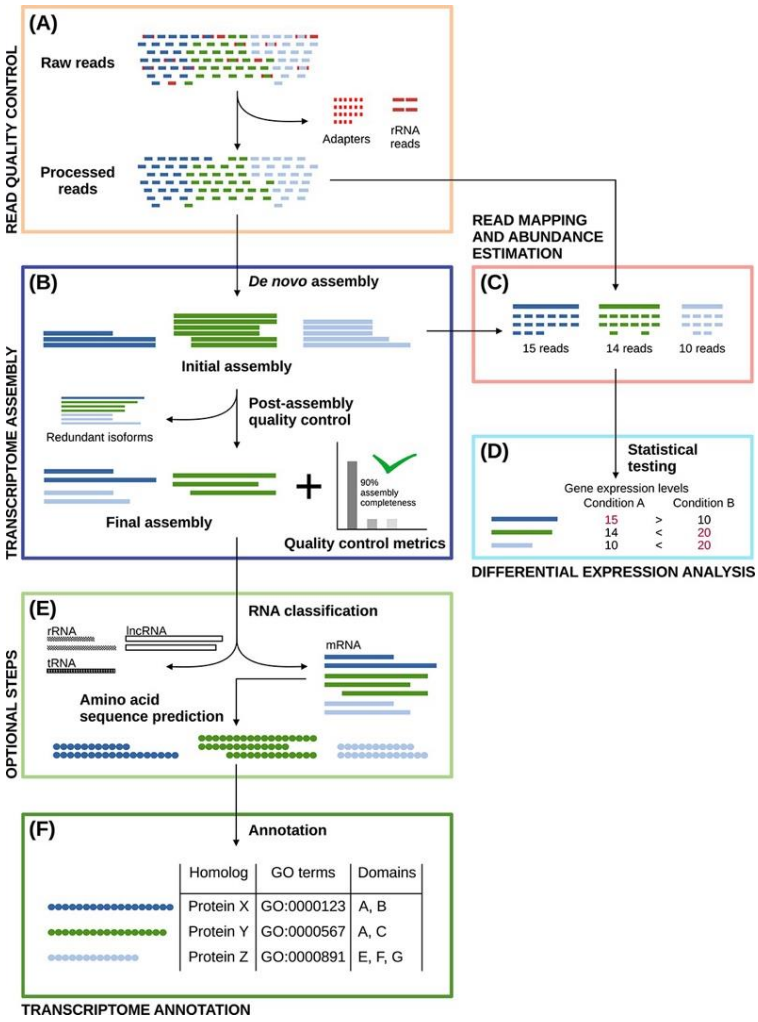


FIGURE 17 – Pipeline d'assemblage et d'annotation d'un transcriptome [Raghavan et al., 2022].

2015]. La comparaison de l'expression des gènes entre différentes conditions ou stade de développement est souvent l'objectif principal. Le design expérimental devra donc prendre en compte la question biologique, le budget ainsi que la variabilité des données [Todd, Black et Gemmell, 2016]. Pour avoir un jeu de données prenant suffisamment en compte la variabilité biologique, un nombre minimum de 3 réplicats biologiques est recommandé. Le RNA-Seq peut aussi servir à l'annotation d'un génome comme expliqué auparavant (Voir 2.6 Annotation d'un génome), notamment des parties non codantes difficilement annotables autrement [Hoff et Stanke, 2015]. Une fois l'expérimentation mise en place, l'ARN sera extrait et souvent enrichi en ARN messager à cause de la forte abondance d'ARN ribosomaux [Sultan et al., 2014]. Après ça, l'ARN sera séquencé ; ce qui se fait surtout en utilisant des technologies produisant des shorts reads (généralement entre 50 et 250 pb) [Raghavan et al., 2022]. En effet, le manque de précision et le faible débit des technologies de séquençage long reads les rendent assez peu efficaces pour la quantification d'expression de gènes même si elles peuvent trouver leur utilité dans l'identification de transcrits [Pardo-Palacios et al., 2024].

Pré-traitement des reads

Après le séquençage, la qualité des reads doit être contrôlée (Figure 17A). Cela peut se faire avec FastQC notamment. Ensuite, les adaptateurs devront être retirés par un logiciel comme cutadapt [Martin, 2011]. Puis les reads seront corrigés et filtrés selon leur qualité. Fastp [Chen, Zhou et al., 2018] est un outil permettant à la fois le "trimming" des adaptateurs, le filtrage et la correction des reads.

Assemblage d'un transcriptome

Les reads pourront ensuite être assemblés. Comme pour le génome, cela peut être fait sur référence ou *de novo* (Figure 17B). L'assemblage sur référence ne sera utilisé que s'il

TABLE 7 – Liste non exhaustive d'assembleurs de transcriptome.

Nom de l'assembleur	**Dernière mise à jour**	**Référence**
inGAP-CDG	2020	[Peng, Ji et Zhao, 2016]
Oases	2015	[Schulz et al., 2012]
SOAPdenovo-Trans	2022	[Xie, Wu et al., 2014]
rnaSPAdes	2025	[Bushmanova et al., 2019]
Trans-ABySS	2018	[Robertson et al., 2010]
TransLiG	2019	[Liu, Yu et al., 2019]
Trinity	2024	[Grabherr et al., 2011]

existe un génome de bonne qualité pour l'espèce étudiée [Behera, Voshall et Moriyama, 2021]. En effet, des gaps dans le génome peuvent empêcher une bonne reconstruction des transcrits [Raghavan et al., 2022]. L'assemblage *de novo* sera donc privilégié pour les espèces non-modèles. L'assemblage d'un transcriptome *de novo* a besoin d'assembleurs spécifiques. Il en existe un certain nombre (Table 7). Ces assembleurs ne fonctionnent pas toujours de la même façon mais la plupart utilisent des k-mers pour reconstruire les transcrits [Raghavan et al., 2022]. La taille de ces k-mers est un paramètre central pour les assembleurs de transcriptome. En effet, une petite taille de k-mers permet l'assemblage de transcrits peu exprimés mais va créer plus de faux positifs [Raghavan et al., 2022]. Tandis qu'une taille de k-mers plus grande va réduire le nombre de transcrits assemblés et ne pas reconstruire ceux peu exprimés. Pour obtenir un transcriptome complet et avec peu d'erreurs, l'utilisation de plusieurs tailles de k-mers et de plusieurs assembleurs est préconisée [Behera, Voshall et Moriyama, 2021].

Post-Assemblage

Après l'assemblage, il est possible d'obtenir un nombre important de transcrits qui peut notamment être dû à l'épissage alternatif qui va produire plusieurs transcrits par gènes, appelés isoformes [Raghavan et al., 2022]. Il est important de réduire la redondance de l'assemblage en utilisant certains outils de clustering comme Cd-hit [Li et Godzik, 2006] (Figure 17B). Ensuite, comme pour le génome, il sera important de décontaminer l'assemblage (Voir 2.5 Décontamination).

Annotations

La dernière étape de l'assemblage d'un transcriptome *de novo* est l'annotation (Figure 17E et F). Pour annoter les ARN non codant comme les ARNt et ARNr, les mêmes outils que pour le génome peuvent être utilisés (tRNAscan, base de données

Rfam). Pour les ARNm (ARN messagers = ARN qui vont ensuite être traduits en protéines), il sera nécessaire d'identifier les séquences codantes dans un premier temps avec des logiciels spécialisés comme ESTScan [Iseli, Jongeneel et Bucher, 1999] ou Transdecoder [Haas, 2024]. En effet, les ARNm sont composés de régions qui ne seront pas traduites (appelées UTRs pour untranslated regions) [Raghavan et al., 2022]. Une fois les séquences codantes identifiées, il sera possible de les annoter à l'aide des mêmes outils que pour le génome comme InterProScan ou BLAST. Il existe aussi Trinotate qui est un outil complet et spécialisé dans l'annotation du transcriptome permettant de combiner plusieurs méthodes et bases de données [Bryant et al., 2017].

Quantification de l'expression

Une fois le transcriptome assemblé et annoté, il va permettre de quantifier l'expression des gènes en fonction de différentes conditions. D'abord les reads peuvent être alignés sur le transcriptome ce qui permettra l'obtention d'une valeur d'abondance par transcrit pour chaque condition (Figure 17C). Pour l'alignement, il est possible d'utiliser un mappeur de short reads classique comme bowtie2 [Langmead et Salzberg, 2012] ou STAR [Dobin et al., 2013]. Il faudra ensuite se servir d'un logiciel de comptage de reads comme HTSeq [Putri et al., 2022]. Une autre option est l'utilisation d'un pseudo-aligneur comme Kallisto [Bray et al., 2016] ou Salmon [Patro et al., 2017] qui va à la fois mapper les reads et fournir un comptage. Cette alternative est plus rapide et plus simple tout en restant précise [Raghavan et al., 2022].

Normalisation des données

Pour pouvoir comparer les comptages de reads entre différents échantillons, il est nécessaire de passer par une étape de normalisation [Evans, Hardin et Stoebel, 2018]. En effet, la profondeur de séquençage (effet inter-échantillon), la longueur des gènes ou

le pourcentage de GC (effets intra-échantillon) vont créer des variations non voulues [Zyprych-Walczak et al., 2015]. Pour corriger ces biais, plusieurs méthodes existent comme la TMM (Trimmed Mean of M values), la RLE (Relative Log Expression) ou la TPM (Transcripts Per Million). La méthode TPM sera plutôt utilisée afin de filtrer les transcrits en amont des analyses différentielles si leur expression est globalement trop faible par exemple [Raghavan et al., 2022]. Les deux autres méthodes vont normaliser par la distribution en se basant sur l'hypothèse que le nombre de gènes différentiellement exprimés sera sensiblement le même entre les échantillons [Evans, Hardin et Stoebel, 2018]. Ces deux méthodes sont souvent comparées mais aucune des deux ne semblent clairement meilleure que l'autre [Corchete et al., 2020 ; Zyprych-Walczak et al., 2015].

Analyses différentielles

Le but de cette étape est d'identifier des gènes différentiellement exprimés entre plusieurs conditions suivant le dispositif expérimental mis en place au départ (Figure 17D). Les packages R edgeR [Robinson, McCarthy et Smyth, 2010] et DeSeq2 [Love, Huber et Anders, 2014] sont généralement utilisés pour identifier les gènes différentiellement exprimés. Ils vont comparer les expressions entre conditions en utilisant un test similaire au test de Fisher pour edgeR et un modèle linéaire généralisé pour DeSeq2 tout en prenant en compte la variabilité entre réplicats biologiques. Ainsi, des gènes différentiellement exprimés seront identifiés selon les conditions. Grâce à l'annotation de leur fonction, il sera possible de déterminer les processus biologiques impliqués dans l'adaptation de l'organisme à différentes conditions par exemple. Cela peut se faire grâce à des analyses d'enrichissement. Ces analyses vont tester statistiquement si une annotation est présente plus que ce qui pourrait être attendu lors d'un tirage aléatoire [Tipney et Hunter, 2010]. Le package R clusterProfiler permet d'effectuer des analyses d'enrichissement sur les annotations GO et KEGG [Yu, 2024].

Le transcriptome est une ressource permettant de mieux comprendre le fonctionnement biologique d'une espèce. Les choix faits lors de son assemblage et des analyses auront un impact important sur les conclusions de l'étude. C'est pourquoi les différentes étapes présentées ci-dessus doivent être réfléchies et adaptées aux objectifs de recherche.

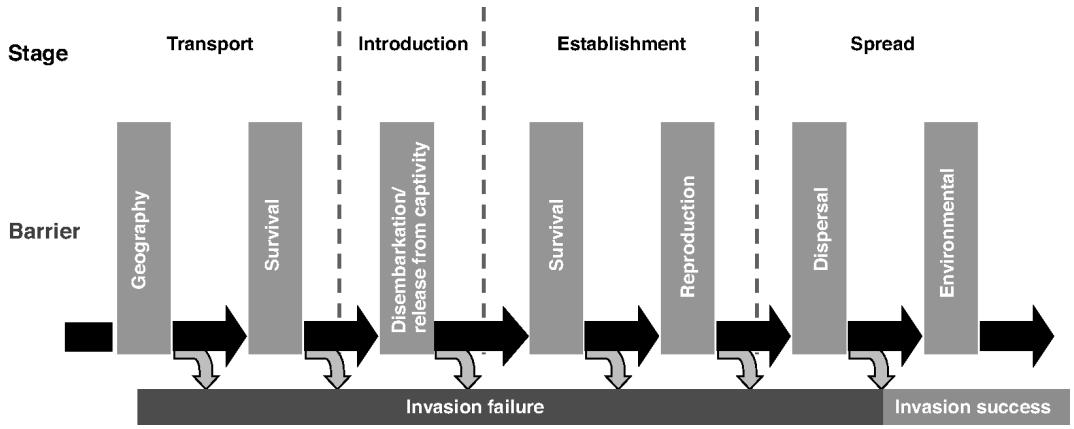


FIGURE 18 – Les différentes étapes aboutissant à une invasion biologique [Chapple et Wong, 2016].

3 La jussie, une plante aquatique invasive colonisant le milieu terrestre

3.1 Les invasions biologiques

Définition

Les invasions biologiques sont actuellement considérées comme l'une des principales causes de perte de la biodiversité [Souza et Prevedello, 2021]. Une espèce invasive, par opposition aux espèces indigènes, est une espèce exotique envahissante dont la présence dans une région a été causée par l'Homme de façon délibérée ou non et qui a réussi à surmonter les barrières biogéographiques [Pyšek, Hulme et al., 2020]. Elle va ensuite s'installer et se répandre largement dans son aire d'introduction [Williamson et Fitter, 1996]. Par extension, une invasion biologique peut être considérée comme la réussite d'une espèce à s'établir, se développer et à maintenir des populations en dehors de sa zone géographique d'origine [Theoharides et Dukes, 2007].

Le processus d'invasion peut être divisé en 4 étapes : le transport, l'introduction, l'installation et l'invasion (Figure 18) [Williamson et Fitter, 1996]. Tout d'abord, l'espèce doit être transportée de son aire de répartition vers une nouvelle aire. Cela peut être effectué de manière volontaire (plantes ornementales, animaux de compagnie) ou involontaire (contamination de marchandises, transport dans des conteneurs) [Hulme, 2021]. Après avoir été transporté, l'espèce doit s'introduire dans sa nouvelle aire ; c'est-à-dire s'échapper de captivité si elle y était puis y survivre. Vient ensuite l'installation qui consiste en l'établissement d'une petite population autosuffisante. Enfin, l'espèce deviendra invasive si elle se répand rapidement et largement [Pyšek, Hulme et al., 2020]. Le caractère invasif d'une espèce peut aussi se définir par les dégâts occasionnés par cette dernière [Williamson et Fitter, 1996].

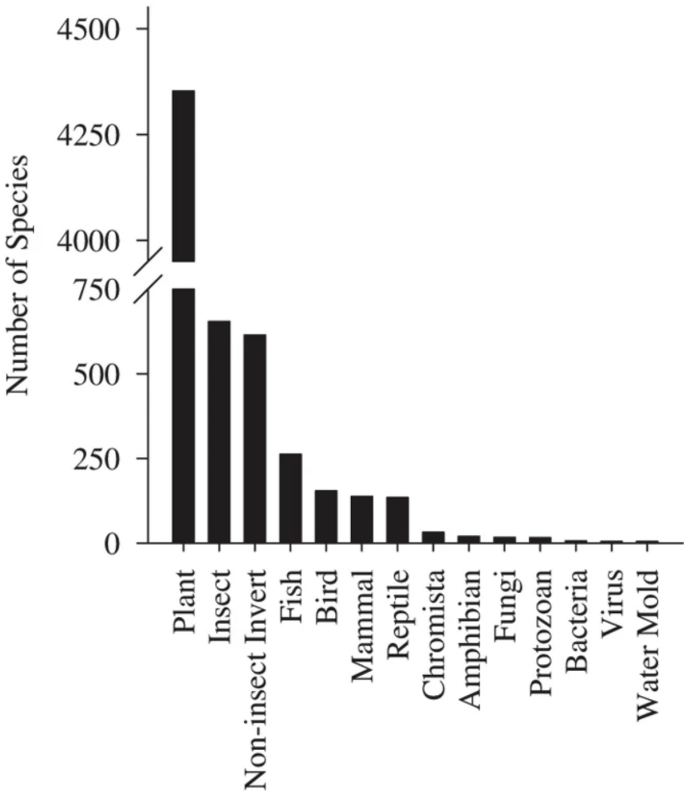


FIGURE 19 – Nombre d'espèces invasives enregistrées dans 20 pays selon leur groupe taxonomique [Pagad et al., 2018].

Problèmes et caractéristiques des espèces invasives

Si l'espèce arrive à s'installer et à se disperser, les dégâts peuvent être très importants sur l'environnement [Carneiro et al., 2024], l'économie [Rai, 2022], en particulier pour les plantes aquatiques invasives [Macêdo et al., 2024] ou même sur l'Homme directement [VilÀ et al., 2021]. Les espèces invasives sont notamment responsables de 25% des extinctions de plantes et 33% des extinctions d'animaux [Blackburn, Bellard et Ricciardi, 2019]. Elles ont un impact très important sur l'agriculture représentant, à l'échelle mondiale, plusieurs milliards de dollars de dégâts [Paini et al., 2016]. La santé de l'Homme et d'autres espèces animales ou végétales sont aussi menacées par la propagation de pathogènes invasifs [Morand, 2017 ; VilÀ et al., 2021] comme *Botryosphaeria dothidea*, un champignon ascomycète responsable de l'apparition du chancre et du dépérissement des arbres [Marsberg et al., 2017] ou d'espèces invasives venimeuses ou vénéneuses comme la Rascasse volante [Norton et Norton, 2021], le frelon asiatique [Franklin et al., 2017], l'ambroisie [Gladieux et al., 2011] ou la berce du Caucase [Renčo et al., 2021].

Les espèces invasives possèdent souvent des caractéristiques facilitant leur établissement et leur expansion comme une importante plasticité phénotypique, un taux de croissance élevé, une reproduction rapide ou une allélopathie [Gioria et al., 2023]. Une étude de 2018 réalisée sur 20 pays montre qu'environ 70% des espèces exotiques envahissantes sont des plantes (Figure 19) [Pagad et al., 2018]. La liste des espèces préoccupantes pour l'Union Européenne est établie régulièrement en application du règlement européen (n° 1143/2014) relatif à la prévention et à la gestion de l'introduction et de la propagation des espèces exotiques envahissantes, et de la réglementation nationale de 2018 (code de l'environnement L411-5) (OFB). En 2025, la dernière version de cette liste comporte 114 espèces (48 végétales et 66 animales) concernées par ce règlement (CDR-EEE) dont la jussie.



Ludwigia grandiflora subsp. *hexapetala*

Ludwigia peploides subsp. *montevidensis*

FIGURE 20 – Fleurs de *Ludwigia grandiflora* subsp. *hexapetala* et *Ludwigia peploides* subsp. *montevidensis*. Crédit photos : D. Barloy ; H. Tinguy, INPN.

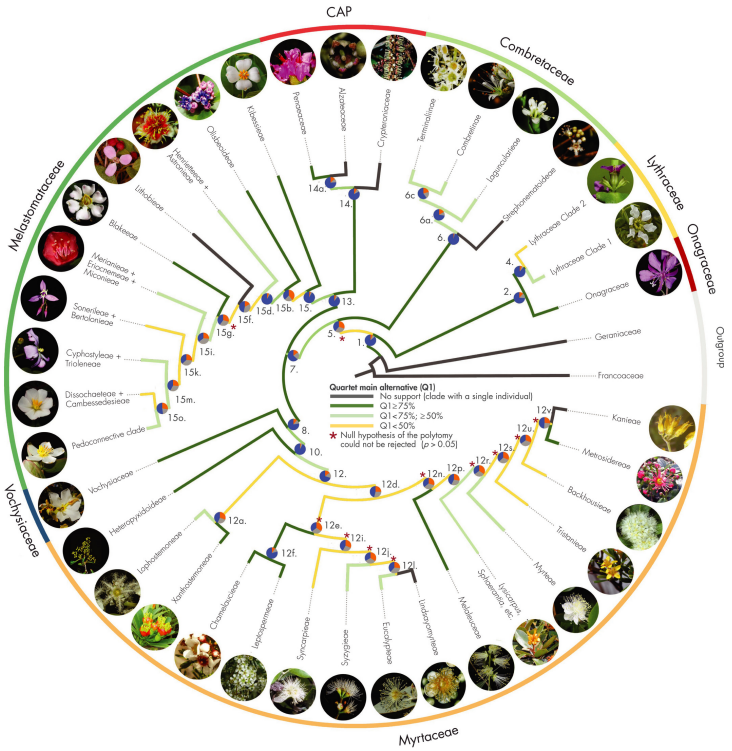


FIGURE 21 – Phylogénie multi-locus de l'ordre des Myrtales [Maurin et al., 2021].

3.2 La jussie

Taxonomie

La jussie est classée parmi les plantes aquatiques invasives les plus importantes d'Europe [Hussner, Stiers et al., 2017]. En réalité, ce terme désigne en France deux espèces (Figure 20), *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh* ; décaploïde : 2n = 10x = 80 chromosomes) et *Ludwigia peploides* subsp. *montevidensis* (*Lpm* ; diploïde : 2n = 2x = 16 chromosomes) [Dandelot, 2004]. Elles font partie de la sous-famille des Ludwigioideae (83 espèces), de la famille des Onagraceae (660 espèces) et de l'ordre des Myrtales (9 familles, 399 genres, environ 14 000 espèces) [Wagner, Hoch et Raven, 2007 ; Maurin et al., 2021] (Figure 21). Les Onagraceae regroupent des plantes herbacées, de petits arbustes, des plantes semi-aquatiques (par exemple, certaines espèces de *Ludwigia* L.) et quelques arbres [Maurin et al., 2021]. La famille des Onagraceae comprend actuellement deux sous-familles, qui se distinguent par la persistance (Ludwigioideae W.L. Wagner & Hoch) ou la chute (Onagroideae) des sépales au stade de la fructification, et par plusieurs caractéristiques anatomiques florales [Wagner, Hoch et Raven, 2007]. La sous-famille des Onagroideae est divisée en 6 tribus et la sous-famille des Ludwigioideae n'est constituée que par le genre *Ludwigia* L.. La famille des Lythraceae est la famille phylogénétiquement la plus proche des Onagraceae. Elle comprend environ 260 espèces, réparties en 28 genres ; correspondantes à des plantes allant de type herbacé à de grands arbres avec une majorité des espèces de type sous-arbustes [Maurin et al., 2021 ; POWO, 2025].

De récentes études cytologiques ont montré que l'espèce diploïde *Lpm* était une des espèces progénitrices de *Lgh* et participait à hauteur de 4x au génome 10x de *Lgh* (Figure 22) [Barloy et al., 2024]. *Ludwigia helminthorrhiza* (diploide, 2n = 16 chromosomes) constituait une des autres espèces progénitrices et participait au génome de *Lgh* à

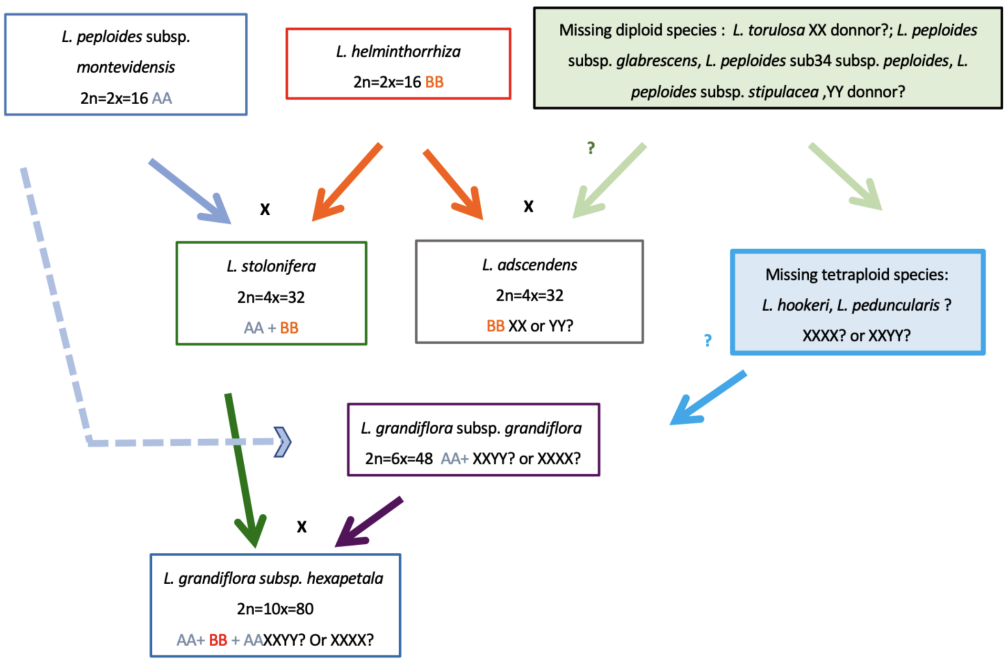


FIGURE 22 – Histoire phylogénétique hypothétique des espèces du genre *Ludwigia* [Barloy et al., 2024].

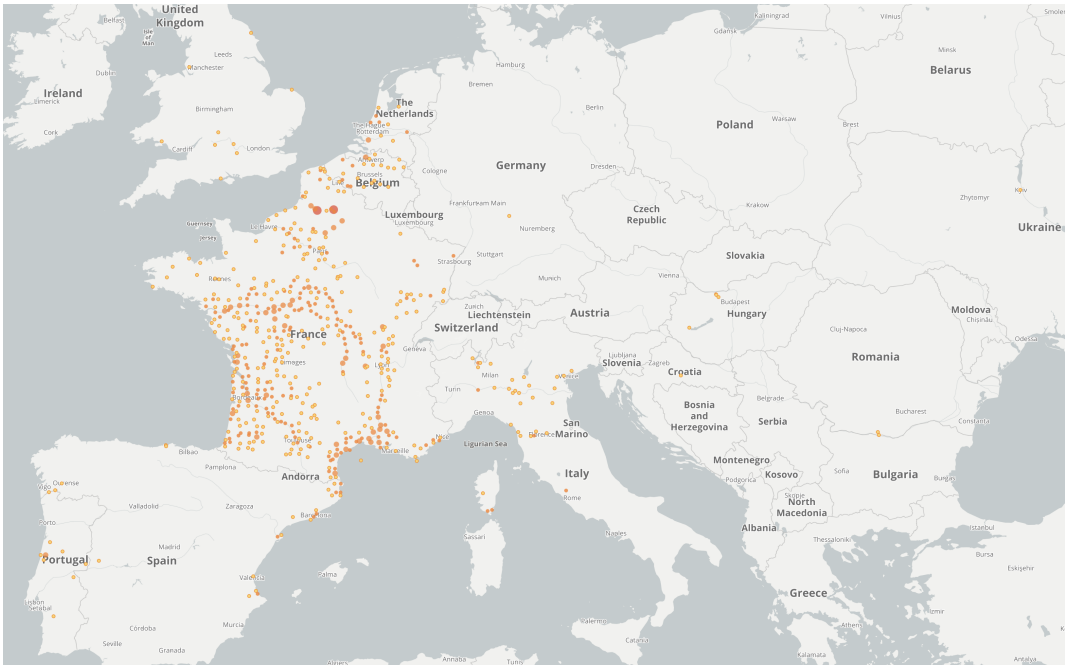


FIGURE 23 – Carte de répartition de *Ludwigia grandiflora* subsp. *hexapetala* et *Ludwigia peploides* subsp. *montevidensis* en Europe. Source : GBIF.

hauteur de 2x (Figure 22). Le(s) progéniteur(s) de la partie du génome (4x) restant est (sont) encore inconnu(s). Ces travaux ont également permis de démontrer la nature auto-allopolyploïde de *Lgh*. L'espèce *Ludwigia grandiflora* subsp. *grandiflora*, espèce reconnue comme invasive aux USA [Grewell, Netherland et Skaer Thomason, 2016], est sans doute une des deux espèces ayant participé à l'apparition de *Lgh* suite à une hybridation interspécifique avec *Ludwigia stolonifera* (Figure 22) [Barloy et al., 2024].

Répartition et biologie

Les jussies sont originaires d'Amérique du Sud et ont été importées pour la première fois en Europe en 1823 à Montpellier [Dandelot, 2004]. Depuis, elles se sont largement répandues sur le territoire français mais sont également présentes en Allemagne [Nehring et Kolthoff, 2011], en Italie [Celesti-Grapow et al., 2009], en Suisse [Lindemann-Matthies, 2016], en Grande-Bretagne [Gallardo et Aldridge, 2013] et dans beaucoup d'autres pays européens (Figure 23). Dans le reste du monde, des pays comme les États-Unis [Reddy et al., 2021], le Japon [Hieda et al., 2020] ou l'Australie [Grewell, Netherland et Skaer Thomason, 2016] sont également touchés. En France, *Lpm* a surtout colonisé le sud-ouest tandis que *Lgh* se retrouve dans la plupart des régions. Les jussies ont été importées comme plantes ornementales de bassin ou d'aquarium [Grewell, Netherland et Skaer Thomason, 2016]. Ce sont des plantes vivaces qui forment des tapis très denses à la surface de plan d'eau ou de rivière avec peu de courant [EPPO, 2011].

Plusieurs caractéristiques des jussies permettent leur succès invasif : un taux de croissance élevé, une résistance à l'hiver, une tolérance à divers environnements et une allélopathie [Dandelot, Robles et al., 2008 ; EPPO, 2011]. Leur reproduction est principalement végétative c'est-à-dire qu'une nouvelle plante peut se développer à partir d'un seul fragment [Dandelot, 2004]. Cependant, certaines populations en France sont

capables de produire des graines avec un taux de germination de quasiment 100% [Portillo Lemus et al., 2021]. L'estimation du taux de clonalité de populations de *Lgh* en France a révélé que ces populations se reproduisent principalement de manière clonale mais avec des taux estimés d'événements sexuels pouvant aller jusqu'à 40% [Stoeckel et al., 2024]. En France, les jussies ne possèdent que très peu de prédateurs naturels identifiés comme le ragondin, l'écrevisse de Louisiane ainsi que quelques espèces d'insectes [EPPO, 2011]. Cependant, l'écrevisse de Louisiane semble favoriser la dispersion des jussies en les fragmentant [Thouvenot, Haury, Pottier et al., 2017]. En Amérique du Sud, quelques insectes se nourrissent de jussie mais ne sont pas des prédateurs spécifiques de la plante [Reddy et al., 2021].

Dégâts et gestion

Les jussies ont des impacts écologiques importants sur la biodiversité locale en modifiant fortement les propriétés physico-chimiques du milieu aquatique, avec une diminution de l'oxygène, des sulfates et nitrates et une augmentation des phosphates et sulfures dans l'eau [Dandelot, Matheron et al., 2005]. Ces impacts sont dus à la grande capacité de production de biomasse de la jussie, couvrant la surface et bloquant, ainsi, l'accès à la lumière, limitant le courant et augmentant la sédimentation. Les jussies ont également un impact économique non négligeable en bloquant les voies navigables et les canaux et en perturbant l'accès à l'eau impactant ainsi la pêche, la chasse et l'agriculture [EPPO, 2011 ; VNF, 2024].

Leur gestion est très compliquée malgré l'interdiction de leur importation depuis 2007 [EPPO, 2011]. Des herbicides chimiques pourraient être efficace comme cela a été le cas aux États-Unis [Grewell, Netherland et Skaer Thomason, 2016] mais leur utilisation est interdite en France [Thouvenot, Haury et Thiebaut, 2013]. Des moyens mécaniques comme le faucardage ou l'arrachage manuel peuvent être utilisés mais il

faut éviter de fragmenter les plantes à cause de leur capacité de reproduction végétative. L'utilisation de l'arrachage manuel en Allemagne constitue un rare cas de gestion réussie de la jussie avec 99% de la biomasse retirée [Hussner, Windhaus et Starfinger, 2016]. Cependant, ceci fut uniquement possible car la population était de petite taille. Le contrôle biologique ne semble pas être une solution envisageable de par le manque d'ennemi naturel des jussies et de l'appétence limitée de ces dernières [Thouvenot, Haury et Thiebaut, 2013]. En effet, les chevaux et les vaches peuvent en consommer mais seulement si c'est leur unique option.

Ressources génomiques

Les jussies sont des espèces non-modèles pour lesquelles très peu de ressources génomiques sont actuellement disponibles. Seuls les génomes chloroplastiques de *Lgh* et *Lpm* ont été assemblés récemment [Barloy-Hubler et al., 2025]. Aucun génome mitochondrial dans la sous-famille des Ludwigioideae n'a été assemblé et seulement 4 dans la famille des Onagraceae sont disponibles (consultation de Genbank en avril 2024) [O'Leary et al., 2016]. Il s'agit des mitogénomes de l'épilobe, *Chamaenerion angustifolium* (OX328283) et des 3 onagres, *Oenothera elata*, *Oenothera biennis* et *Oenothera villaricae* [Fischer et al., 2022]. Le même constat peut être fait pour les génomes nucléaires avec seulement 4 génomes incomplets dans la famille des Onagraceae, aucun n'ayant été annoté (consultation de Genbank en juin 2025). Il s'agit des génomes de deux épilobes, *Chamaenerion angustifolium* (GCA_946814005.1) et *Epilobium hirsutum* (GCA_965153245.1) assemblés au niveau chromosomique et de ceux de l'onagre, *Oenothera biennis* (GCA_025802395.1) et d'une clarkia, *Clarkia xantiana* (GCA_036873135.1) assemblés en scaffolds seulement. La taille des génomes de *Lgh* et *Lpm* a été estimé par cytométrie de flux à respectivement, 1418,9 Mb (± 107,3 Mb) et 261,7 Mb (± 9,2 Mb) [Barloy et al., 2024].

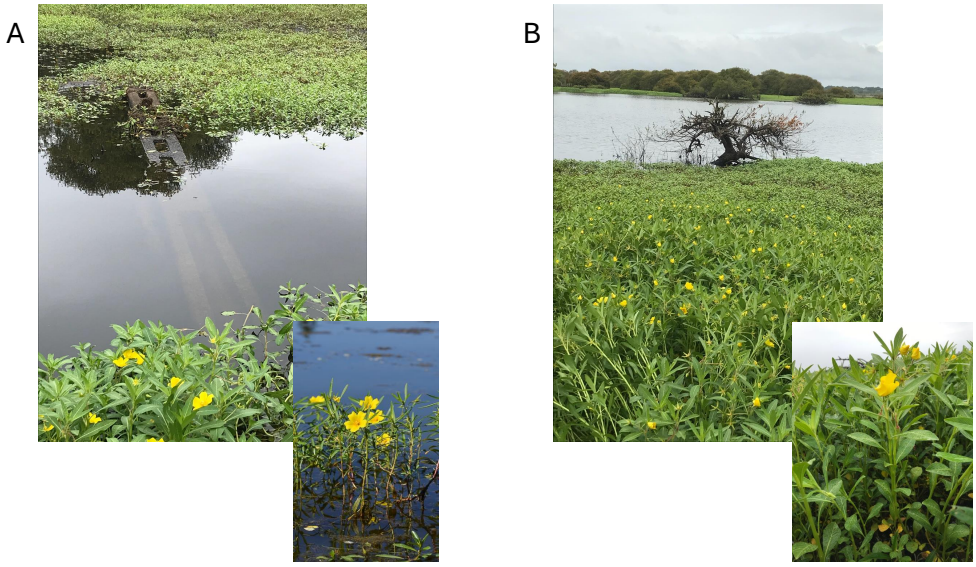


FIGURE 24 – *Ludwigia grandiflora* subsp. *hexapetala* en A) milieu aquatique et B) milieu terrestre. Crédits photos : D. Barloy ; Florealpes.

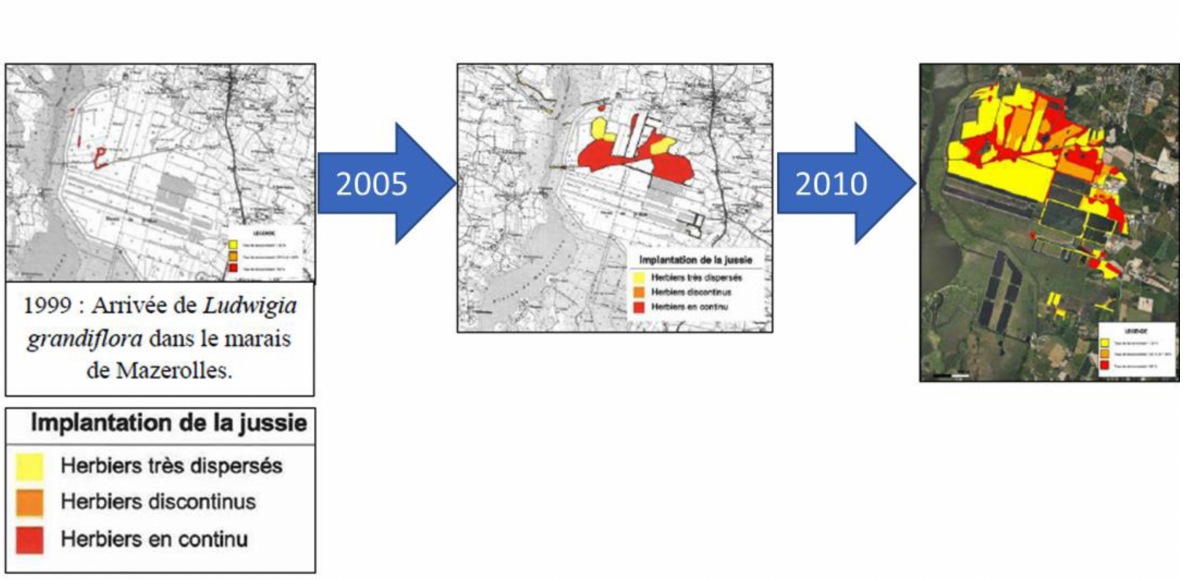


FIGURE 25 – Progression de 1999 à 2010 de *Ludwigia grandiflora* subsp. *hexapetala* en milieu terrestre au sein du marais de Mazerolles (Syndicat Mixte EDENN, 2011).

3.3 Colonisation du milieu terrestre de *Lgh*

Cas du marais de Mazerolles

Les jussies ont développé une capacité à coloniser le milieu terrestre dans certaines prairies humides (Figure 24) [Haury, Noël et al., 2011]. Les plantes de *Lgh* utilisées dans notre étude proviennent du marais endigué de Mazerolles, situé à Suce-sur-erdre (près de Nantes). Ce marais est une zone protégée Natura 2000 qui est constituée de prairies inondables sur près de 750 ha. Les niveaux d'eau y sont en partie régulés grâce à un système de pompage et les prairies restent sous l'eau tout l'hiver et jusqu'en juin. Le marais est envahi par *Ludwigia grandiflora* subsp. *hexapetala* depuis 1999 (Figure 25). On y trouve deux formes de jussie, une aquatique et la seconde terrestre. La première est apparue en 1999 et se développe aujourd'hui dans la douve qui traverse les prairies et la seconde a été observée 5 ans plus tard dans la prairie (Syndicat Mixte EDENN, 2011). La forme terrestre de la jussie a colonisé une grande partie des prairies du marais de Mazerolles (Figure 25). La présence de *Lgh* dans les prairies humides conduit à une dépréciation de leur qualité fourragère et à l'abandon de leurs exploitations par les agriculteurs (Haury, com pers.).

Existence de deux morphotypes en fonction du milieu

Chez *Lgh*, deux morphotypes ont été identifiés suite à ce changement de milieu, un morphotype terrestre et un morphotype aquatique. Le morphotype aquatique possède des tiges fines et allongées avec des feuilles courtes arrondies, alors que le morphotype terrestre est caractérisé par un port ramassé, trapu avec des feuilles plus allongées [Haury, Druel et al., 2014] (Figure 26). En comparant le comportement des deux morphotypes élevés en conditions contrôlés en milieu aquatique et en milieu terrestre, Billet et al (2018) et Genitoni et al (2025) ont montré que le morphotype terrestre présente

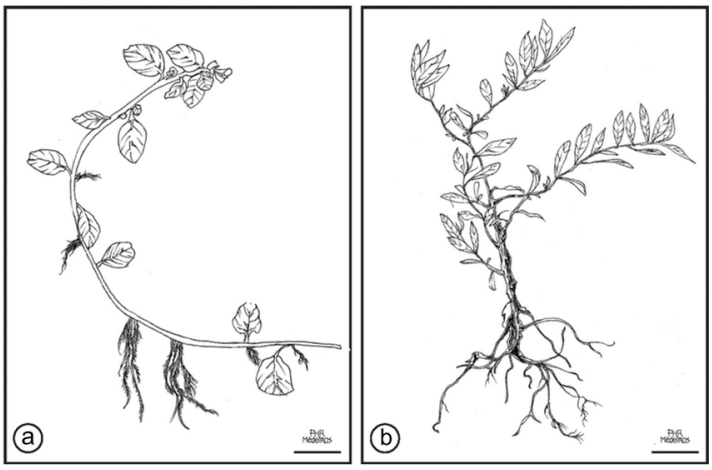


FIGURE 26 – Morphologie des morphotypes aquatique (a) et terrestre (b) chez *Ludwigia grandiflora* subsp. *grandiflora* [Scremin-Dias et al., 2023]. Cette illustration représente une sous-espèce différente de *Lgh* mais montre bien les différences morphologiques entre les deux morphotypes.

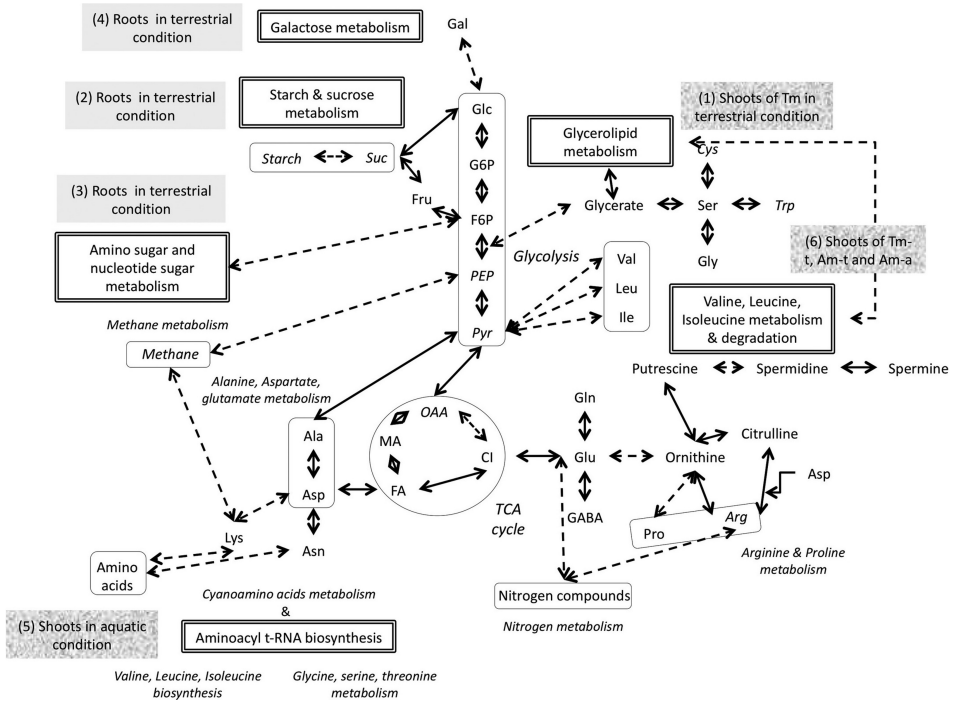


FIGURE 27 – Représentation des voies métaboliques du sucre, des acides aminés et des composants apparenté liées au changement de milieu de la jussie selon le morphotype [Billet et al., 2018].

des biomasses racinaire et aérienne plus importantes et un nombre de bourgeons plus élevé que le morphotype aquatique, quel que soit le milieu dans lequel il se développe. Billet et al (2018) ont également mis en évidence que *Lgh* activait des voies métaboliques différentes suivant la condition où elle se trouve (Figure 27). Par exemple, en condition terrestre, *Lgh* produit plus de sucres dans les racines mobilisant les voies métaboliques comme le métabolisme du galactose et le métabolisme de l'amidon et du sucre. En milieu aquatique, *Lgh* produit au niveau des tiges plus d'acides aminés impliquant la voie de la biosynthèse des aminoacyl-ARNt.

De plus, des réponses métaboliques différentes ont pu être observées entre les deux morphotypes mis dans les mêmes conditions pendant 14 ou 28 jours [Genitoni, Vassaux, Renault et al., 2025]. En condition terrestre, le morphotype aquatique va produire plus de sucres et de polyols que le morphotype terrestre. En condition aquatique, le morphotype terrestre va avoir des réponses physiologiques intermédiaires entre le morphotype aquatique dans la même condition et le morphotype terrestre en condition terrestre. Par ailleurs, l'implication de la méthylation dans le changement de milieu de *Lgh* a également pu être montré [Genitoni, Vassaux, Delaunay et al., 2020]. Les effets d'une hypométhylation de l'ADN affectent différemment, selon le morphotype, les traits morphologiques et la production de régulateurs de croissance.

3.4 Fonctions d'acclimatations au milieu terrestre des plantes amphibies

Les plantes amphibies, comme la jussie, vivent à l'interface entre milieux aquatique et terrestre et peuvent se développer et survivre dans ces deux environnements [Koga, Ikematsu et Kimura, 2024]. Cette particularité demande à ce type de plantes de s'acclimater au changement de milieu. En effet, les milieux aquatique et terrestre peuvent notamment varier par la disponibilité de l'eau, la luminosité ou l'accessibilité du CO_2

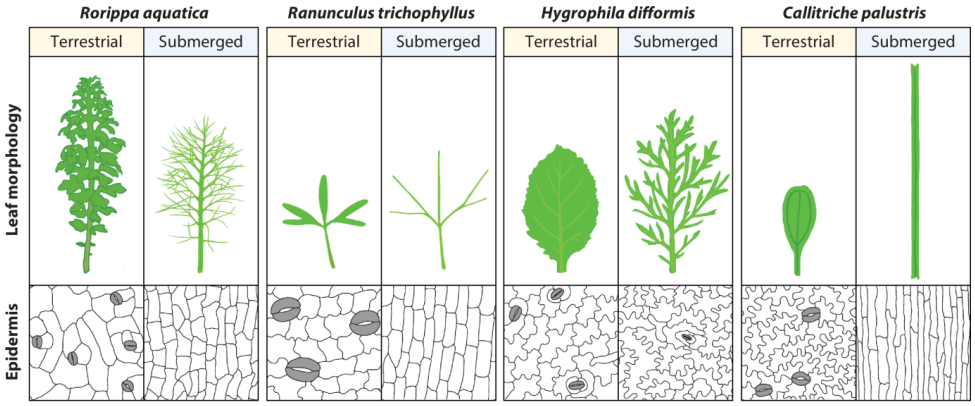


FIGURE 28 – Comparaison des traits entre des feuilles aériennes et immergées chez 4 espèces de plantes amphibies modèles illustrant l'hétérophyllie [Koga, Ikematsu et Kimura, 2024].

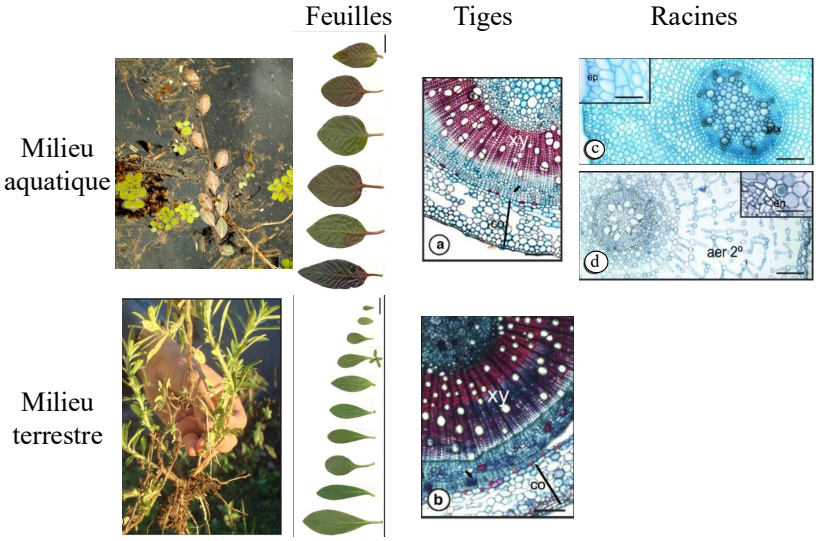


FIGURE 29 – Comparaison de la morphologie des plantes, des feuilles, des tiges et des racines de *Ludwigia grandiflora* subsp. *grandiflora* en milieu aquatique et en milieu terrestre (modifié à partir de Scremin-Dias et al., 2023). a : Tige présentant un cortex étroit ; b : Tige présentant un cortex large et lignifié ; c : Détail de la racine adventive mettant en évidence les cellules épidermiques ; d : Détail du pneumatophore mettant en évidence les aérenchymes secondaires. aer 2° : aérenchyme secondaire, ep : épiderme, co : cortex, plx : phloème, xy : xylème.

[Han, Xing et al., 2021]. L'hétérophyllie, c'est-à-dire la capacité à développer des feuilles de formes différentes pour les parties immergées et aériennes de la plante, est une capacité assez commune chez les plantes amphibies [Li, Hu et al., 2019]. Cela peut se caractériser par des feuilles immergées plus fines, allongées et sans cuticule ni stomate (Figure 28).

Le changement de milieux aquatique versus terrestre conduit, chez *Ludwigia grandiflora* subsp. *grandiflora*, à une modification de la forme des feuilles ainsi que de la densité et de la répartition des stomates entre la face supérieure et la face inférieure des feuilles [Scremin-Dias et al., 2023] (Figure 29). Quelle que soit la face de la feuille, les individus élevés en milieu terrestre présentent une densité de stomates plus élevée. Une forte densité de trichomes, des tissus lignifiés et compactés, ainsi que des parois cellulaires épaisses sont des caractéristiques fréquemment observées au niveau des tiges des plantes s'étant développées en milieu terrestre (Figure 29 a,b). La présence d'espaces intercellulaires plus larges et d'un périderme avec aérenchyme secondaire à la base des racines et des tiges était notable chez les individus du milieu aquatique. Par ailleurs, la présence de racines adventices et de pneumatophores n'a été observée que chez les plantes s'étant développées en milieu aquatique alors que les plantes cultivées en milieu terrestre développent des racines pivotantes (Figure 29 c,d). Ces auteurs concluent que les fluctuations de la disponibilité en eau au cours des cycles saisonniers induisent d'importantes réponses structurelles, qui contribuent au succès de la colonisation par *L. grandiflora* subsp. *grandiflora*.

Dans les feuilles immergées, l'expression de gènes impliqués dans le développement des stomates et de la cuticule sera réduite par rapport aux feuilles aériennes (Figure 30) [Han, Xing et al., 2021 ; Li, Zhao et al., 2024 ; Koga, Ikematsu et Kimura, 2024]. De plus, dans l'eau, l'accès au CO_2 étant réduit, la plante va mettre en œuvre des mécanismes permettant la concentration du CO_2 (CCMs) comme l'utilisation du HCO3- ou

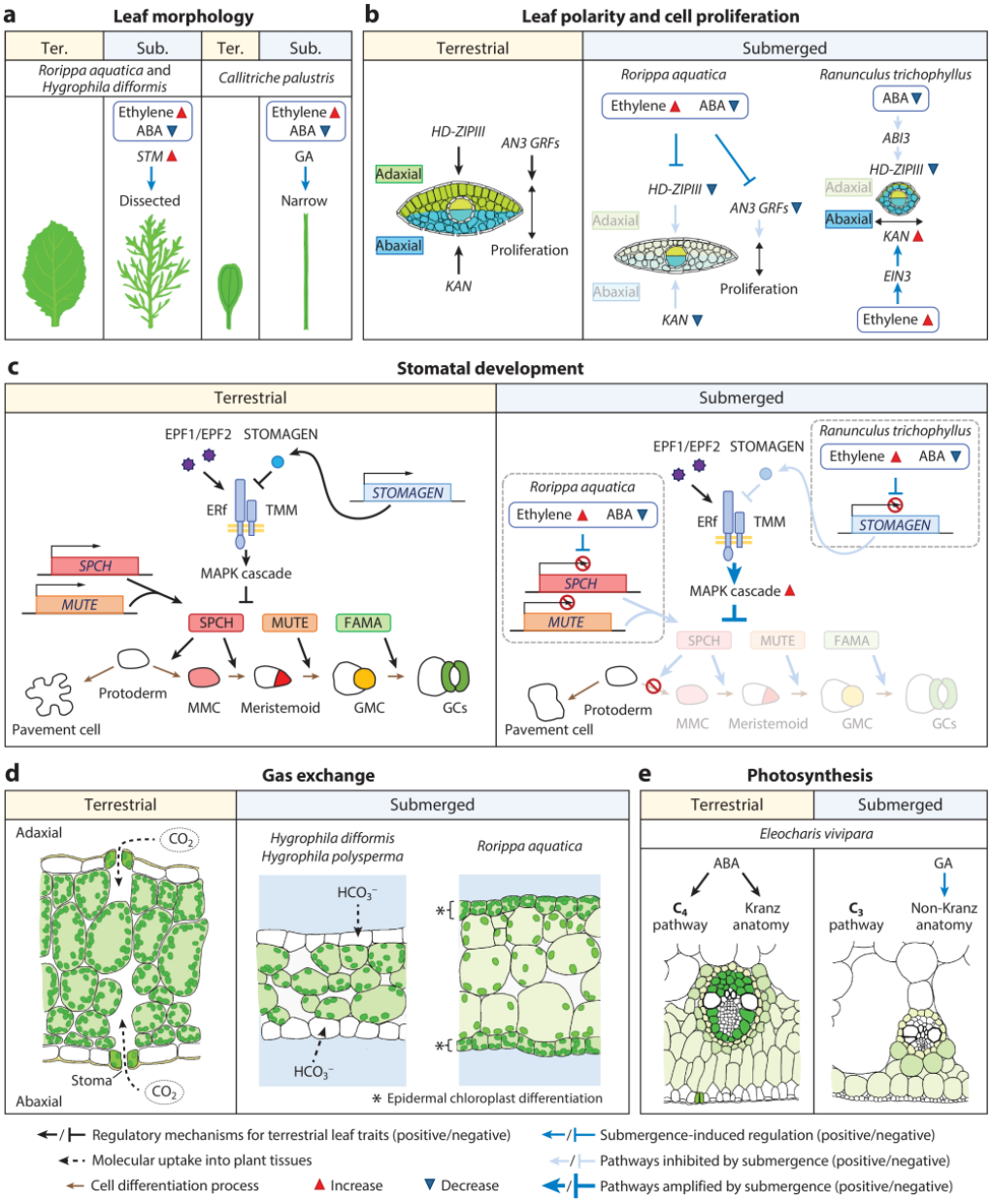


FIGURE 30 – Mécanismes moléculaires montrant la plasticité phénotypique des plantes amphibies [Koga, Ikematsu et Kimura, 2024].

la fixation du carbone en C4. Les stratégies permettant l'utilisation du HCO3- sont les plus fréquentes des CCMs chez les plantes d'eau douce (présentes chez 50% des espèces) [Han, Xing et al., 2021]. Afin de réduire la résistance à la diffusion du CO_2, les chloroplastes du mésophylle sont souvent proches de l'épiderme. Cela va aussi permettre une meilleure acclimatation dans un environnement plus sombre sous l'eau. De plus, les gènes liés au « light-harvesting complex », permettant de capter la lumière, sont surexprimés dans les feuilles immergées [Han, Xing et al., 2021]. Dans les feuilles aériennes, ce sont les gènes de la photosynthèse qui sont surexprimés. Une autre contrainte du changement de milieu est la différence de température. Des régulateurs de la thermomorphogénèse, i.e. les réponses morphologiques à l'élévation de la température, peuvent être différentiellement exprimés entre les milieux terrestre et aquatique comme *ELF3* chez *Hygrophila difformis* [Li, Zhao et al., 2024].

Par ailleurs, certaines phytohormones sont très importantes lors de réponse à un stress abiotique. Sous l'eau, une augmentation de l'éthylène et une diminution de la signalisation ABA sont observées, ces deux composés agissant de manière antagoniste (Figure 30) [Koga, Ikematsu et Kimura, 2024]. Par ailleurs, la biosynthèse de l'ABA est également fortement inhibée, ce qui entraîne une diminution de la signalisation ABA. Koga et al (2024) concluent que la plasticité phénotypique des plantes amphibies est fondamentalement déterminée par ces mécanismes de reconnaissance à la submersion.

4 Objectifs et stratégies de la thèse

4.1 Contexte de la thèse

Cette thèse s'inscrit dans le cadre des travaux de recherche de Dominique Barloy sur la jussie au sein de l'UMR DECOD, notamment sur la compréhension du passage du milieu aquatique au milieu terrestre de *Ludwigia grandiflora* subsp. *hexapetala*. Cela vient également à la suite de la thèse de Julien Genitoni (soutenue en 2019) qui a travaillé sur des approches physiologiques et épigénétiques de l'acclimatation de *Lgh* au milieu terrestre. Au cours de cette thèse, des séquençages du transcriptome et du méthylome ont été réalisés. Par ailleurs, des séquençages de l'ADN de *Lgh* et *Lpm* ont été réalisés en 2020. Ils ont permis l'assemblage de leurs génomes chloroplastiques par Anne-Laure Le Gac et Frédérique Barloy-Hubler produisant ainsi les premières ressources génomiques pour ces espèces orphelines [Barloy-Hubler et al., 2025].

4.2 Objectifs

Les études morphologiques et métabolomiques de *Ludwigia grandiflora* subsp. *hexapetala* ont permis de caractériser les différentes réponses des deux morphotypes dans les milieux terrestre et aquatique (cf partie 3 de l'introduction). Au vu de ces résultats, nous faisons l'hypothèse que l'acclimatation au milieu terrestre de *Lgh* implique des réponses génétiques et épigénétiques. Pour pouvoir étudier cela, développer des ressources génomiques chez cette espèce orpheline s'avère nécessaire. L'objectif principal de la thèse a donc été la production de ces ressources génomiques chez *Lgh*. Cet objectif se décline en trois sous-objectifs :

— (1) L'assemblage du génome mitochondrial de *Lgh*

— (2) L'assemblage du génome nucléaire de *Lgh*

— (3) L'assemblage et l'analyse différentielle du transcriptome de *Lgh* lors d'une

acclimatation au milieu terrestre.

Pour ce faire, nous avons été confrontés à différents challenges :

— (1) Construire des ressources génomiques chez une espèce peu étudiée dans ce domaine et pour laquelle les données génomiques existantes au sein du genre, de la famille et de l'ordre étaient limitées.

— (2) Pouvoir difficilement exploiter les données existantes de par l'éloignement phylogénétique de *Lgh* et des espèces appartenant à l'ordre des Myrtales et plus généralement aux Viridiplantae.

— (3) Travailler sur une espèce polyploïde sans qu'aucun génome des espèces progénitrices ne soit disponible et avec très peu d'assembleurs disponibles permettant la résolution d'haplotype.

— (4) Réaliser une analyse fonctionnelle chez une espèce polyploïde où la redondance d'information génétique dû à la ploïdie complique l'interprétation des données.

Le premier chapitre des résultats présente l'assemblage du génome mitochondrial de *Lgh* et *Lpm*. L'objectif a été d'assembler les génomes mitochondriaux complets de *Lgh* et *Lpm* est de comparer leur structure. Pour ce faire, nous avons été confrontés au manque de conservation de séquences et de structure entre les mitogénomes au sein de l'ordre des Myrtales [Stage M2 Guillaume Doré]. Nous avons donc fait l'hypothèse que l'exploitation en entier des mitogénomes d'espèces proches serait difficile. Ainsi, nous avons choisie la stratégie de sélectionner les séquences mitochondriales de notre set de séquençage en utilisant les CDS (Coding DNA Sequence = séquences des gènes codant des protéines) des mitogénomes des espèces du clade des Malvidées, clade contenant l'ordre des Myrtales auquel les *Ludwigia* appartiennent. Ce chapitre est composé de l'article intitulé "*De novo* hybrid assembly unveils multi-chromosomal mitochondrial genomes in *Ludwigia* species, highlighting genomic recombination, gene transfer, and RNA editing events" publié dans la revue International Journal of Molecular Science

(IJMS).

Le deuxième chapitre des résultats décrit l'assemblage du génome nucléaire de *Lgh*. Au vu des données disponibles et de la complexité attendue du génome en lien avec la polyploïdie, l'objectif est l'assemblage d'un draft de génome le plus complet possible et avec des annotations fonctionnelles pouvant servir pour de futures analyses. Ce travail a été valorisé au travers d'un article soumis à la revue GigaByte et intitulé "From Water to land, introducing the first decaploid genome in Ludwigioideae : *Ludwigia grandiflora* subsp. *hexapetala*, an invasive amphibious plant".

La troisième partie présente l'assemblage du transcriptome de *Lgh*, la comparaison des apports du génome et du transcriptome pour l'obtention de ressources les plus précises et complètes possibles ainsi que des premiers résultats d'analyses différentielles. La première hypothèse de ce chapitre est que le transcriptome pourra être utilisé pour compléter l'assemblage et les annotations du génome [Hoff et Stanke, 2015 ; Ghurye et Pop, 2019]. Ensuite, nous pensons que l'assemblage d'un transcriptome *de novo* et sur référence est complémentaire et permettra des analyses plus précises et complètes [Lee, Na et Park, 2021]. La troisième hypothèse est que l'expression d'un nombre important de gènes dépendra du tissu analysé [Yaschenko et al., 2022]. Enfin, nous testerons l'hypothèse que l'alternance de cycles "aquatique-terrestre" est un des facteurs de l'acclimatation de *Lgh* au milieu terrestre. L'attendu sera que le morphotype aquatique ayant subi ce traitement présentera des réponses fonctionnelles similaires au morphotype terrestre [Genitoni, 2019].

4.3 Matériel végétal et stratégies de séquençage et d'assemblage

La préparation du matériel végétal ainsi que les différents séquençages ont été réalisés avant ma thèse par Anne-Laure Le Gac et Luis Portillo-Lemus pour l'ADN (en 2020)

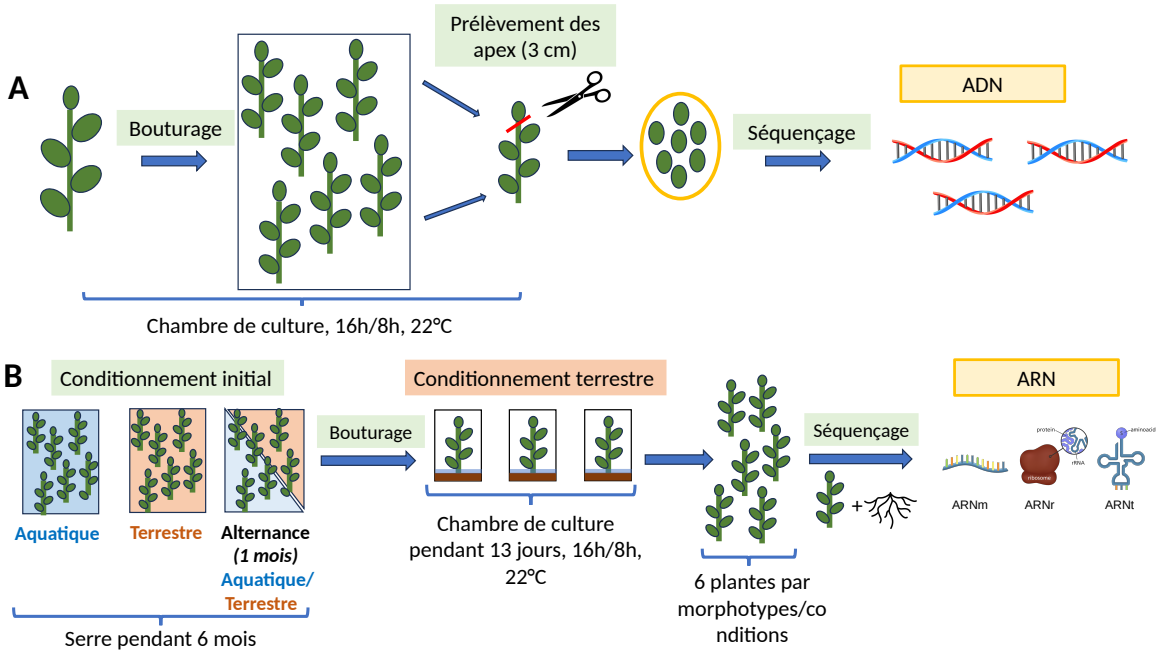


FIGURE 31 – Dispositif général simplifié du matériel végétal utilisé pour les séquençages de A : l'ADN de *Lgh* et *Lpm* et B : l'ARN de *Lgh*.

TABLE 8 – Les différents séquençages disponibles avant la thèse.

Espèce	**Type de séquences**	**Type de séquençage**	**Méthode de séquençage**	**Utilisation dans la thèse**
Lpm	ADN	Long reads	GridION (Oxford Nanopore)	Mitogénome
Lgh	ADN	Long reads	GridION (Oxford Nanopore)	Mitogénome et génome
Lgh	ADN	Short reads	MiSeq (Illumina)	Mitogénome et génome
Lgh	ARN	Short reads	NovaSeq (Illumina)	Transcriptome

et Julien Genitoni pour l'ARN (en 2019).

Matériel végétal

Pour le séquençage de l'ADN ayant servi aux assemblages du mitogénome et du génome, les plantes ont été prélevées au marais de Mazerolles pour *Lgh* et à La Musse pour *Lpm* [Billet et al., 2018]. Elles ont ensuite été placées en chambre de culture à 22°C et avec un cycle 16h/8h (lumière/obscurité). Une seule tige de 10 cm pour chaque espèce a ensuite été prélevée (Figure 31). Toutes les plantes ayant servies au séquençage de l'ADN sont issues de ces tiges afin d'éviter toute variabilité génétique. Les bourgeons (3 cm depuis l'apex) de plusieurs plantes pour chaque espèce ont été prélevés. L'ADN a ensuite été extrait suivant un protocole spécifique [Barloy-Hubler et al., 2025].

Pour le séquençage de l'ARN utilisé pour l'obtention du transcriptome et les analyses différentielles, des plantes de *Lgh* des deux morphotypes (aquatique et terrestre) ont été prélevées au marais de Mazerolles [Genitoni, Vassaux, Renault et al., 2025]. Durant 6 mois, elles ont été conditionnées en serre dans différentes conditions (aquatique, terrestre ou une alternance de 15 jours en aquatique et 15 jours en terrestre) (Figure 31). Au bout des 6 mois, des boutures de 10 cm en partant de l'apex ont été prélevées pour chacun des morphotypes/conditions. Ces boutures ont été conditionnées en condition terrestre en chambre de culture à 22°C et avec un cycle 16h/8h (lumière/obscurité), selon un dispositif détaillé au chapitre 3.

Stratégies de séquençage et d'assemblage

Plusieurs séquençages étaient disponibles avant la thèse et ont été utilisés en fonction des objectifs de la thèse (Table 8). Les séquençages de l'ADN ont été effectués par les technologies MiSeq d'Illumina pour les short reads (uniquement *Lgh*) et GridION d'Oxford Nanopore pour les long reads (*Lgh* et *Lpm*). Ces données de séquençage

ont ensuite servi pour les assemblages des mitogénome de *Lgh* et *Lpm* et du génome nucléaire de *Lgh*. Pour l'obtention du transcriptome et les analyses différentielles, les ARNs des racines et des tiges-feuilles ont été séquencés avec la technologie NovaSeq d'Illumina.

Le mitogénome de *Lgh* a été assemblé *de novo*. Les régions codantes des mitogénomes des espèces les plus proches phylogénétiquement ont permis de sélectionner les reads mitochondriaux. Le mitogénome de *Lpm* a ensuite été assemblé en s'appuyant sur celui de *Lgh*. Le génome de *Lgh* a été assemblé en utilisant les long reads et les short reads. L'utilisation de différents assembleurs ainsi que de plusieurs stratégies ont permis de réduire la fragmentation du génome. À cause du manque de profondeur de séquençage, l'utilisation de méthodes spécifiques aux génomes polyploïdes n'a pas été possible pour cet assemblage.

Le transcriptome de *Lgh* a été assemblé *de novo* avec les reads issus du séquençage de l'ARN. Un transcriptome sur référence a également été obtenu en mappant les reads issus du séquençage de l'ARN sur le génome. Les deux transcriptomes ont ensuite servi comme référence pour des analyses d'expression différentielle de gènes.

Chapitre 1

Assemblage des mitogénomes de *Lgh* et *Lpm*

1 Introduction

1.1 Forme des mitogénomes de plantes

Les mitogénomes de plantes ont souvent été représentés sous la forme d'une unique molécule circulaire appelé "Master circle". Ce modèle est basé sur les mitochondries des animaux et les bactéries. Les premiers séquençages et assemblages soutenaient cette structure malgré les contradictions avec les observations microscopiques et électrophorétiques qui notaient la présence de plusieurs molécules parfois linéaires [Sloan, 2013]. La découverte, dans un premier temps, de structure particulière des mitogénomes de plantes a questionné sur la validité du modèle du "Master circle". En 1984, une structure tripartie du mitogénome de la navette (*Brassica campestris*) est identifiée, comprenant un "Master circle" pouvant se recombiner en deux plus petites molécules [Palmer et Shields, 1984]. En 2011, le mitogénome du concombre (*Cucumis sativus*) est le premier composé de plusieurs molécules indépendantes à avoir été assemblé [Alverson, Rice et al., 2011]. Il est composé de 3 molécules circulaires de 1556, 84 et 45 kb qui ne peuvent pas toutes recombiner ensemble pour former un "Master circle". L'arrivée des technologies de séquençage de deuxième mais surtout de troisième génération ont permis de se détacher complétement de ce modèle [Wu et al., 2020]. Elles ont favorisé une meilleure compréhension de la structure des mitogénomes. Ainsi, en 2020, 10% des mitogénomes assemblés contenaient plusieurs molécules [Wu et al., 2020]. Par exemple les mitogénomes de l'orchidée, *Paphiopedilum micranthum* [Yang, Dierckxsens et al., 2023] et du peuplier (*Populus simonii*) [Yin, 2022] sont composés de 26 et 3 molécules circulaires, respectivement. Le mitogénomes de l'eucalyptus, *Eucalyptus urophylla* X *grandis* est composé de 3 molécules linéaires (Accession : OQ947725, OQ947726, OQ947727). Enfin, le mitogénome de la silène (*Silene conica*) possède quant à lui 128 chromosomes dont la majorité (86) n'ont même pas un seul gène complet [Sloan et al., 2012]. Les

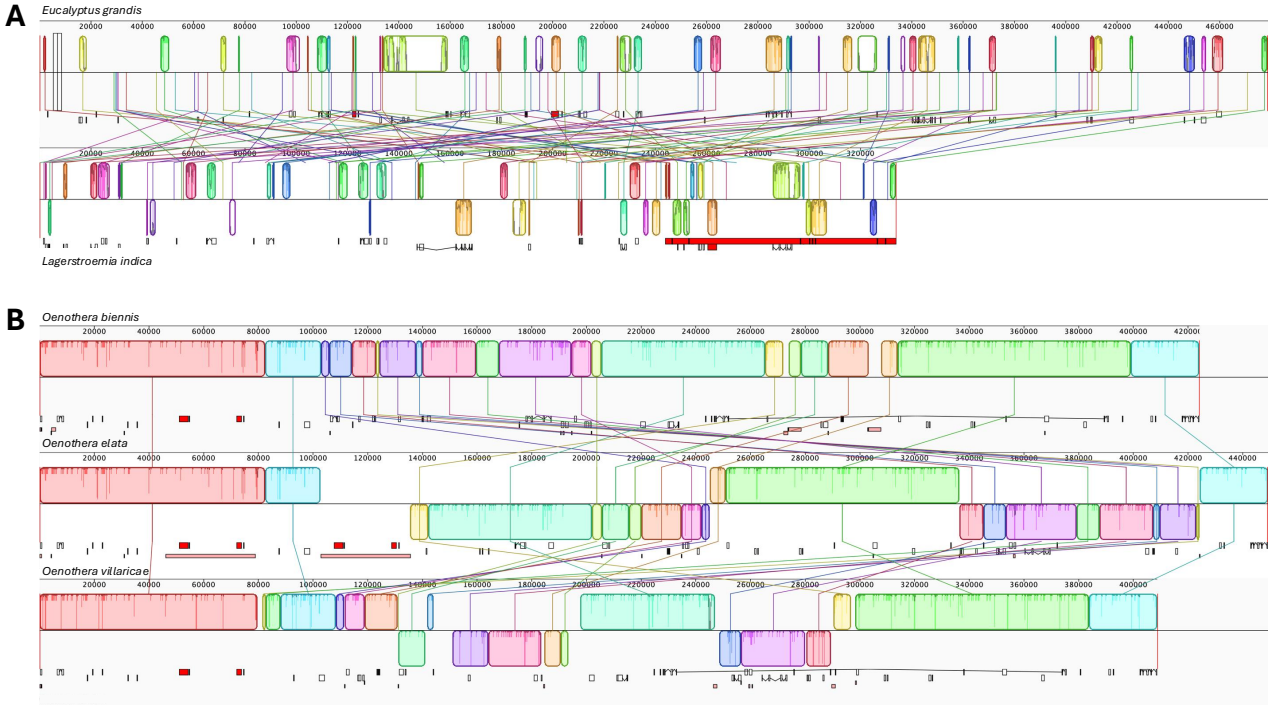


FIGURE 1.1 – Comparaison de la structure et de l'organisation des mitogénomes appartenant à l'ordre des Myrtales *via* l'utilisation de l'algorithme Mauve [Stage M2 Guillaume Doré]. A : Comparaison des mitogénomes d'*Eucalyptus grandis* (en haut) et de *Lagerstroemia indica* (en bas). B : Comparaison des mitogénomes d'*Oenothera biennis* (en haut), d'*Oenothera elata* (au milieu) et d'*Oenothera villaricae* (en bas). Les séquences conservées sont indiquées par une même couleur sur les mitogénomes et sont reliées par une ligne. L'entre-croisement des lignes révèle des différences dans l'organisation et la structure de ces 2 mitogénomes.

formes de mitogénomes chez les plantes sont très variées et ne sont pas toujours un unique cercle comme souvent présenté.

1.2 Mitogénomes dans l'ordre des Myrtales

Ludwigia grandiflora subsp. *hexapetala* (*Lgh*) et *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) font partie de la sous-famille des Ludwigioideae, de la famille des Onagraceae et de l'ordre des Myrtales. Au moment du début de cette étude (Janvier 2022, stage de M2), aucun mitogénome n'était disponible dans la famille des Onagraceae et seulement 2 l'étaient dans l'ordre des Myrtales. Il s'agit des mitogénomes de *Lagerstroemia indica* (famille des Lythraceae ; Accession : NC_035616) et d'*Eucalyptus grandis* (famille des Myrtaceae ; Accession : NC_040010). La comparaison de ces deux mitogénomes a mis en évidence le peu de conservation de séquences et de structure entre eux (Figure 1.1A) [Stage M2 Guillaume Doré]. Cela souligne le manque de conservation des mitogénomes au sein de l'ordre des Myrtales. En mars 2022, 3 mitogénomes d'espèces appartenant à la famille des Onagraceae et au genre *Oenothera* ont été assemblés [Fischer et al., 2022]. Malgré la proximité phylogénétique de ces trois espèces, leurs mitogénomes sont relativement variables entre eux (Figure 1.1B) [Stage M2 Guillaume Doré]. Cela montre bien l'impossibilité d'utiliser un génome d'une autre espèce comme référence pour l'assemblage d'un mitogénome de plantes. Dans le clade des Malvidées, englobant l'ordre des Myrtales et 7 autres ordres, 41 mitogénomes étaient assemblés en janvier 2022.

1.3 Utilité de l'assemblage d'un mitogénome chez les plantes

L'assemblage d'un mitogénome peut permettre des études de phylogénie [Duminil et Besnard, 2021]. En effet, grâce à leur taille importante, les mitogénomes de plantes contiennent plus de polymorphismes que les plastomes [Štorchová et Krüger, 2024]. En

plus, les gènes mitochondriaux, notamment le gène *cox1*, font partie des marqueurs génétiques les plus utilisés dans ces études [Smith, 2016]. La reconstruction d'arbre phylogénétique grâce aux mitogénomes permettra de résoudre les controverses toujours présentes et de mieux comprendre les mécanismes d'évolution de diversification des plantes [Zardoya, 2020]. Par ailleurs, il a été démontré que les réarrangements du mitogénomes avait un rôle dans la stérilité mâle cytoplasmique (CMS pour Cytoplasmic male sterility) [Horn, Gupta et Colombo, 2014]. Il s'agit d'une caractéristique héritée de la mère rendant la plante inapte à la production de pollen fonctionnel. Cette caractéristique est utilisée pour la production de variété cultivée de type hybride. L'assemblage de mitogénomes de plantes permettra de faciliter la compréhension de ce mécanisme [Štorchová et Krüger, 2024]. Enfin, l'ADN mitochondrial est très utilisé par les archéologues grâce à son héritabilité monoparentale, son nombre de copies élevé et son temps de dégradation plus faible que l'ADN nucléaire [Smith, 2016].

2 Synthèse de l'article "*De novo* hybrid assembly unveils multi-chromosomal mitochondrial genomes in *Ludwigia* species, highlighting genomic recombination, gene transfer, and RNA editing events"

2.1 Objectifs et stratégies

L'objectif principal de cette étude est l'assemblage et l'annotation des mitogénomes de *Lgh* et *Lpm*. Ils seront ensuite comparés au niveau de leur contenu en gène et de leur structure. Enfin, des comparaisons avec d'autres mitogénomes d'espèces plus éloignées seront aussi réalisées.

Les ADNs de *Lgh* et *Lpm* ont été séquencés en utilisant les technologies MiSeq d'Illumina (seulement pour *Lgh*) produisant des short reads (SR) et GridION d'Oxford Nanopore (pour les deux espèces) produisant des long reads (LR). Pour *Lgh*, les SR ont été corrigés avec les LR et inversement. Ensuite, différents assembleurs ont été utilisés avec les SR corrigés uniquement, les LR corrigés uniquement et en combinant les SR et LR (méthode hybride). Puis, les LR mitochondriaux de *Lgh* ont été sélectionnés en utilisant les CDS (Coding DNA Sequence = séquences des gènes codant des protéines) des mitogénomes des espèces du clade des Malvidées. Ces LR ont ensuite été allongés grâce aux différents assemblages. Une fois le mitogénome de *Lgh* assemblé, il a servi à l'assemblage de celui de *Lpm*. Pour finir, les gènes, les séquences répétées ainsi que les insertions provenant du chloroplaste ont été annotés dans les deux mitogénomes.

2.2 Principaux résultats

Le mitogénome de *Lgh* est composé de deux molécules circulaires, nommées M1 et M2, de respectivement 544 782 pb et 166 796 pb de long pour une taille totale de 711 578 pb (Accession : PP727126, PP727127). Le mitogénome de *Lpm* est lui aussi composé de deux molécules circulaires (également nommées M1 et M2), de 555 518 pb et 167 000 pb donnant une longueur totale de 722 518 pb (Accession : PP727128, PP727129). Les molécules M2 des deux mitogénomes sont entièrement colinéaires et similaires à 99,5%. Les molécules M1, quant à elles, ont été divisées en 8 blocs colinéaires représentant environ 90% de la longueur totale des mitogénomes. Ces blocs ne sont pas dans le même ordre chez *Lgh* et *Lpm* et certains sont même inversés. Les mitogénomes de *Lgh* et *Lpm* contiennent respectivement 42 et 46 gènes codant des protéines, 10 gènes d'ARNr et 34 gènes d'ARNt. Certains de ces gènes sont des copies complètes provenant d'insertion de séquences chloroplastiques. En effet, 7,8% (55 kb) de chacun des mitogénomes sont des fragments provenant de leur plastome respectif. De nombreuses séquences répétées

ont aussi été identifiées, représentant environ 7% des génomes mitochondriaux des deux *Ludwigia*. Trois de ces séquences répétées font plus de 1 kb. Elles sont impliquées dans la recombinaison des molécules M1 des deux espèces. Plusieurs conformations possibles ont été trouvées pour la molécule M1 chez *Lgh* et *Lpm*, conformations résultant de différents événements de recombinaisons. Dans l'ordre des Myrtales, les mitogénomes de *Lgh* et *Lpm* sont les plus grands assemblés à ce jour. Ils font également partie de ceux possédant le plus de gènes complets annotés.

2.3 Conclusion

Nous avons réalisé les assemblages des mitogénomes de *Ludwigia grandiflora* subsp. *hexapetala* et *Ludwigia peploides* subsp. *montevidensis*. Ce sont les premiers du genre *Ludwigia* et de la sous-famille des Ludwigioideae regroupant 83 espèces. Ces résultats permettent aussi de compléter les connaissances sur les génomes mitochondriaux de plantes, notamment sur leurs structures multi-moléculaires. Les mitogénomes de plantes sont très compliqués à reconstruire comme en témoigne le manque de mitogénomes assemblés dans les bases de données (342 mitogénomes vs plus de 6000 plastomes répertoriés dans RefSeq en avril 2024). Les mitogénomes de *Lgh* et *Lpm* pourront servir pour de futurs assemblages ou analyses chez d'autres plantes.

3 Article "*De novo* hybrid assembly unveils multi-chromosomal mitochondrial genomes in *Ludwigia* species, highlighting genomic recombination, gene transfer, and RNA editing events"

International Journal of
Molecular Sciences

MDPI

Article

De Novo Hybrid Assembly Unveils Multi-Chromosomal Mitochondrial Genomes in *Ludwigia* Species, Highlighting Genomic Recombination, Gene Transfer, and RNA Editing Events

Guillaume Doré [1], Dominique Barloy [1,*,†] and Frédérique Barloy-Hubler [2,*,†]

[1] DECOD (Ecosystem Dynamics and Sustainability), Institut Agro, INRAE, IFREMER, 35042 Rennes, France; guillaume.dore@agrocampus-ouest.fr
[2] UMR 6553 ECOBIO, CNRS, Université de Rennes 1, 35042 Rennes, France
* Correspondence: dominique.barloy@agrocampus-ouest.fr (D.B.); fhubler@univ-rennes1.fr (F.B.-H.)
† These authors contributed equally to this work.

Abstract: Biological invasions have been identified as the fifth cause of biodiversity loss, and their subsequent dispersal represents a major ecological challenge. The aquatic invasive species *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) are largely distributed in aquatic environments in North America and in Europe. However, they also present worrying terrestrial forms that are able to colonize wet meadows. To comprehend the mechanisms of the terrestrial adaptation of *Lgh* and *Lpm*, it is necessary to develop their genomic resources, which are currently poorly documented. We performed de novo assembly of the mitogenomes of *Lgh* and *Lpm* through hybrid assemblies, combining short reads (SR) and/or long reads (LR) before annotating both mitogenomes. We successfully assembled the mitogenomes of *Lgh* and *Lpm* into two circular molecules each, resulting in a combined total length of 711,578 bp and 722,518 bp, respectively. Notably, both the *Lgh* and *Lpm* molecules contained plastome-origin sequences, comprising 7.8% of the mitochondrial genome length. Additionally, we identified recombinations that were mediated by large repeats, suggesting the presence of multiple alternative conformations. In conclusion, our study presents the first high-quality mitogenomes of *Lpm* and *Lgh*, which are the only ones in the Myrtales order found as two circular molecules.

Keywords: water primrose; Myrtales; mitogenome; hybrid assemblies; DNA transfer; repeated sequences

check for updates

Citation: Doré, G.; Barloy, D.; Barloy-Hubler, F. De Novo Hybrid Assembly Unveils Multi-Chromosomal Mitochondrial Genomes in *Ludwigia* Species, Highlighting Genomic Recombination, Gene Transfer, and RNA Editing Events. *Int. J. Mol. Sci.* **2024**, *25*, 7283. https://doi.org/10.3390/ijms25137283

Academic Editor: Abir U. Igamberdiev

Received: 4 June 2024
Revised: 28 June 2024
Accepted: 29 June 2024
Published: 2 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Origin, Biological Traits, and Distribution of the Aquatic Invasive Plants, Ludwigia Species

Biological invasions are now recognized as one of the primary drivers behind global biodiversity loss. Invasive alien plant species (IAPs) lead to significant economic and ecological losses by causing disruption to ecosystems [1]. IAPs are characterized by their strong growth, excellent competitiveness, high adaptive value, an absence of predators, and also reproduce abundantly [2,3].

Of all these invasive aquatic plant species, *Ludwigia grandiflora* subsp. *hexapetala* (or *Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (or *Lpm*) stand out in particular. They're native to South America, but both species have colonized many countries around the world [4]. In North America, *Lgh* and *Lpm* are distributed across several states where they have both been found to degrade major watersheds as well as aquatic and riparian ecosystems [5]. *Lpm* and *Lgh* show some advantageous biological traits that might explain their success of colonization in news areas. *Lpm* and *Lgh* reproduce essentially by clonal propagation with high grow rate, which contribute to the dispersal and establishment of propagules [6]. Both

Ludwigia species are capable of thriving in a wide range of environments, revealing their broad ecological tolerance and great plasticity [7]. Sexual production is also effective in both *Ludwigia* species. In France, *Lpm* is self-compatible and produces many capsules and seeds [8]. *Lgh*, on the other hand, possesses a heteromorphic late-acting self-incompatible system and is also able to produce seeds [9]. Seeds can be dispersed by floating fruit, which contributes to long-distance colonization, as is the case for propagules [7]. *Lgh* has provoked similar damage in Europe, and has been observed in France [10], Germany [11], Italy [12], Spain [13], Switzerland [14], Great Britain/Ireland [15], and Turkey [16]. *Lgh* is also present in Japan [17]. *Lpm* is very common in the southeast and west of France, but its distribution is more limited in Belgium, Corsica, Greece, Italy, the Netherlands, Spain, Turkey, the Balkans, and the UK [18,19]. The oldest documented introductions for both species were in France in 1830 [20].

In its natural habitat, both *Ludwigia* ssp. exhibit an amphibious nature [21], and are commonly found in wetland areas or in transitional zones between aquatic and terrestrial environments [22,23]. However, it is now developing particularly worrying terrestrial forms and is able to colonize wet meadows at an alarming rate [24]. *Lgh* has two distinct morphological types (morphotypes) depending on its environment (aquatic or terrestrial). The two morphotypes present different metabolic processes: fermentation for the aquatic morphotype, and respiration for the terrestrial morphotype [25].

1.2. Genomic Resources in the Onagraceae Family

Genomics plays a central role in identifying genes and metabolic pathways involved in the adaptation process leading to plant invasion [3], with genomic information helping to predict and control invasiveness [26]. For instance, in species such as *Reynoutria japonica* and *Phalaris arundinacea*, genes that play a critical role in invasion are associated with stress response, reproduction, and growth [27,28]. Hence, the genomics of invasive plants represents a powerful tool for comprehending the molecular foundations of plant invasion, enabling the prediction and management of their proliferation [29,30].

Lgh and *Lpm* belong to the malvids clade, Myrtales order, Onagraceae family, and Ludwigioideae subfamily. To date, *Ludwigia* species have not undergone comprehensive genome sequencing, with the exception of the chloroplast genomes of *L. octovalvis* [31], *Lgh* and *Lpm* [32]. In the related subfamily Onagroideae, two incomplete draft genomes are available for *Chamaenerion angustifolium* and *Oenothera biennis* (unpublished GenBank records). *Lpm*, one of the progenitor species of *Lgh*, is a diploid species (2n = 16) with a relatively small nuclear genome (261.7 Mb), and *Lgh* is a decaploid (2n = 80) with a very large nuclear genome (1.4 Gb) [33]. Assembly of the *Lgh* genome is anticipated to be challenging and time-consuming, as has been acknowledged in previous studies dealing with such big and complex genomes [34]. Having successfully assembled and annotated their plastomes [32], we decided to subsequently assemble their mitogenomes due to the crucial role they play in the plants' biological functions [35,36]. In addition, no Ludwigioideae mitogenomes are currently available, and only four Onagraceae mitogenomes: *Chamaenerion angustifolium* (OX328283) and *Oenothera elata*, *O. biennis*, and *O. villaricae* [37].

1.3. Form and Size of Plant Mitogenomes

Commonly depicted as circular, mitogenomes of terrestrial plants exhibit intricate, diverse, and dynamic structures. Hence, certain plants possess multipartite mitochondrial genomes consisting of several chromosomes, predominantly in a circular form, although they can be linear or branched [38–41]. The most commonly reported structure consists of two circular mitochondrial chromosomes, as demonstrated by numerous studies for *Allium cepa* [42], *Saccharum officinarum* [43], *Fallopia multiflora* [44], *Salvia officinalis* [45], and *Zantedeschia odorata* [46]. Another common structure consists of three circular molecules like *Cynanchum wilfordii* [47], *Brassica campestris* [48], and cucumber [49]. Recent studies have revealed the existence of even more intricate mitochondrial genomes. For instance, *Amborella trichopoda* has five circular chromosomes [50], *P. micranthum* has 26 circular

subgenomes [51], *Lophophytum mirabile* is divided into 54 circular chromosomes [52], *Silene noctiflora* has 59 to 63 circular chromosomes [53], and *Silene conica* has at least 128 circular chromosomes, some of which are empty (i.e., contain no functional genes) [54].

In addition to these structural variations, angiosperm mitochondrial genomes are particularly variable in size. While the mitogenome of *Viscum scurruloideum* is frequently referenced in articles as being the smallest, measuring 66 kb [55], it is worth noting that smaller unpublished mitogenomes exist in the GenBank database including that of Asteraceae *Carthamus tinctorius* (63 kb, OQ621746) or Orchidaceae *Dendrobium amplum* (48 kb, MH591889). In contrast, the largest mitogenome currently described is that of *Larix sibirica* with a total of 11.7 Mb [56].

1.4. Repeat Sequences and Exogenous DNA in Plant Mitogenomes

Vascular plant mitogenomes also contain a large fraction of repeat sequences, making up between five and ten percent of the genome [57]. Repetitive elements, especially within the Angiosperm clade, can reach significant sizes, with repeats exceeding 10 kb being relatively common [58]. A portion of these repeated sequences can be traced back to transposable elements such as retrotransposons and mitoviruses [59,60], and numerous TEs fragments have been reported in plant mitogenomes [61–63]. The presence of these extensive repeated sequences within mitogenomes has made it difficult to assemble them accurately [57,64]. Repeats, especially those exceeding 1000 bp, play a significant role in the homologous recombination phenomena [65,66], resulting in different conformations coexisting within the same species [67,68].

The ability of the mitochondrial genome to incorporate exogenous DNA also partly explains these size variations [69]. The most frequent transfer concerns the insertion of chloroplast DNA sequences [70], although sequences of nuclear origin can also be found inserted [71]. Nucleotide sequences derived from plastids and nuclei represent 1–12% and 0.1–13.4% of the mitogenome, respectively [71,72]. Most of these insertions occur in the intergenic regions, even though substitutions of protein-coding gene fractions have been cited [73]. One of the roles of chloroplast-derived sequences is to provide tRNA genes for mitochondrial translation [74], although these sequences may also function as promoters for mitochondrial genes [75]. Most of these chloroplast foreign sequences are non-coding or pseudogenic, although they sometimes appear in coding regions [76] due in particular to alterations in the genetic code between the two organelles [77] and/or differences in the initiation and termination signals of transcription and translation [78]. Remarkably, the amounts of DNA attributed to external sources (plastids and nuclei) are insufficient to account for the origins of all the intergenic regions because more than 60% of these regions show no detectable homology with other known sequences, and their conservation is too weak to support strong functional constraints [79].

1.5. Limits and Objectives of This Study

Another striking feature of angiosperm mitochondrial genomes is their low gene density and a very low collinearity [80,81]. Thus, although the size of their plant mitogenomes is variable, their conserved genetic pool is generally small, with sequences encoding central mitochondrial functions such as subunits of oxidative phosphorylation chain complexes (OXPHOS), proteins involved in the biogenesis of these complexes, and several ribosomal proteins [70,82]. Mitochondrial protein-coding genes typically undergo RNA modification processes to fulfill their functional roles through RNA editing, a post-transcriptional regulation mechanism [83,84].

Due to this remarkable plasticity, it is difficult to assemble plant mitochondrial genomes [85]. However, the recent hybrid assembly approach, which combines long and short reads has proven to be suitable [38,86,87].

In order to better understand the terrestrial adaptation mechanisms of *Lgh* and *Lpm*, and to augment the genomic reservoir of both species, we carried out de novo assemblies of the mitogenomes of *Lgh* and *Lpm*. To achieve this, we first generated *Lpm* and *Lgh* Illumina

Mi-seq short-read (SR) sequences and/or Oxford Nanopore Technologies long reads (LR). Next, we used hybrid assemblies to combine the SR and LR sequences, facilitating the assembly of *Lgh* and *Lpm* mitogenomes. Afterward, we completed functional annotations and analyzed repeats and chloroplast insertions to determine the *Lpm* and *Lgh* mitogenome structures. Finally, we identified genes under selection and constructed a phylogenetic tree using mitogenomes from the Myrtales order.

2. Results

2.1. De Novo Assembly of Ludwigia Multichromosomal Mitogenomes

Sequencing of *Lgh* via Oxford Nanopore Technologies yielded a total of 482,619 reads (denoted as LRs for long reads), amounting to 2,386,362,939 bp. These reads exhibited an average length of 4944 bp, with the longest read extending to 72.9 kb. The sequencing conducted with Illumina Mi-seq technology resulted in 2 × 23,067,490 reads (abbreviated as SRs for short reads), totaling 9,606,411,104 bp. The reads had an average length of 208 base pairs, with the maximum read length reaching 301 base pairs. Error correction using Ratatosk for LRs and SPAdes for SRs retained 96% of the short reads and all long reads. Using these corrected reads, we conducted three distinct de novo assemblies using Megahit for SR-only assembly, Flye for LR-only assembly, and a hybrid assembly combining both read types using SPAdes.

Statistics for *Lgh* mitogenome assemblies were calculated using QUAST and are summarized in Table S1. We were able to observe that the SPAdes assembly stood out for generating the largest contig (192 kb), the greatest total length (806 kb), and achieving the highest genome fraction (94.9%). However, it also exhibited the highest error rate, with 147 mismatches per 100 kb and three misassemblies. Conversely, Megahit exhibited the lowest error rate, tallying just 46 mismatches per 100 kb and zero misassembled contigs. However, it showed weakness in terms of completion, and only reconstructed 89% of the genome, which stood out as the lowest completion rate among the three assemblers. Finally, Flye emerged as the optimal choice, striking a balance between the two other assemblers, with no misassemblies and less mismatches (71 per 100 kb) than SPAdes, and better LGA50 and total aligned length (664 kb) than Megahit. Simultaneously, with these assemblies, we selected 1118 mitochondrial LRs using coding sequences (CDS) from the malvids clade species mitogenomes. These reads were assembled into 12 contigs with the Geneious assembler, allowing us to identify mitochondrial contigs from the de novo assemblies. Subsequently, we extended all of these contigs to achieve the complete mitogenome. Ultimately, we successfully assembled the *Lgh* mitogenome into two circular molecules, referred to as M1 and M2, with respective sizes of 544,782 bp and 166,796 bp, resulting in a combined total of 711,578 bp (Figure 1). The *Lgh* mitogenome coverage depths were 38x for LRs and 229x for SRs, and these values were similar for both M1 and M2. The *Lgh* mitogenome GC content was 45.2%.

Sequencing *Lpm* using Oxford Nanopore Technologies generated a dataset comprising 68,907 reads, which was reduced to 67,384 following self-correction by RATATOSK. These cumulatively amounted to 510,468,382 bp. These reads exhibited an average length of 7434 bp, with some extending as long as 83.03 kilobases. Without the availability of short reads, only the Flye assembly could be carried out and generated 496 contigs. Using the *Lgh* mitogenome as a reference, we found one *Lpm* contig of 158,656 kb that corresponded to 95% of the *Lgh* M2 molecule, and eight *Lpm* contigs that covered 83.02% of the *Lgh* M1 molecule (453,279 kb in total). Using these contigs and incorporating *Lpm* LRs that were not assembled by Flye, we successfully reconstructed the *Lpm* mitogenome, forming two circular molecules (also designated as M1 and M2), mirroring the structure seen in *Lgh*. These two molecules had respective sizes of 555,518 bp and 167,000 bp, culminating in a total mitogenome size of 722,518 bp (Figure 1). The coverage depth of the *Lpm* mitogenome was 10.1 for M1 and 5.8 for M2, with a GC content of 45.2%.

The mitogenomes of *Lgh* and *Lpm* were submitted to GenBank under the accession numbers PP727126, PP72717, PP727128, and PP727129.

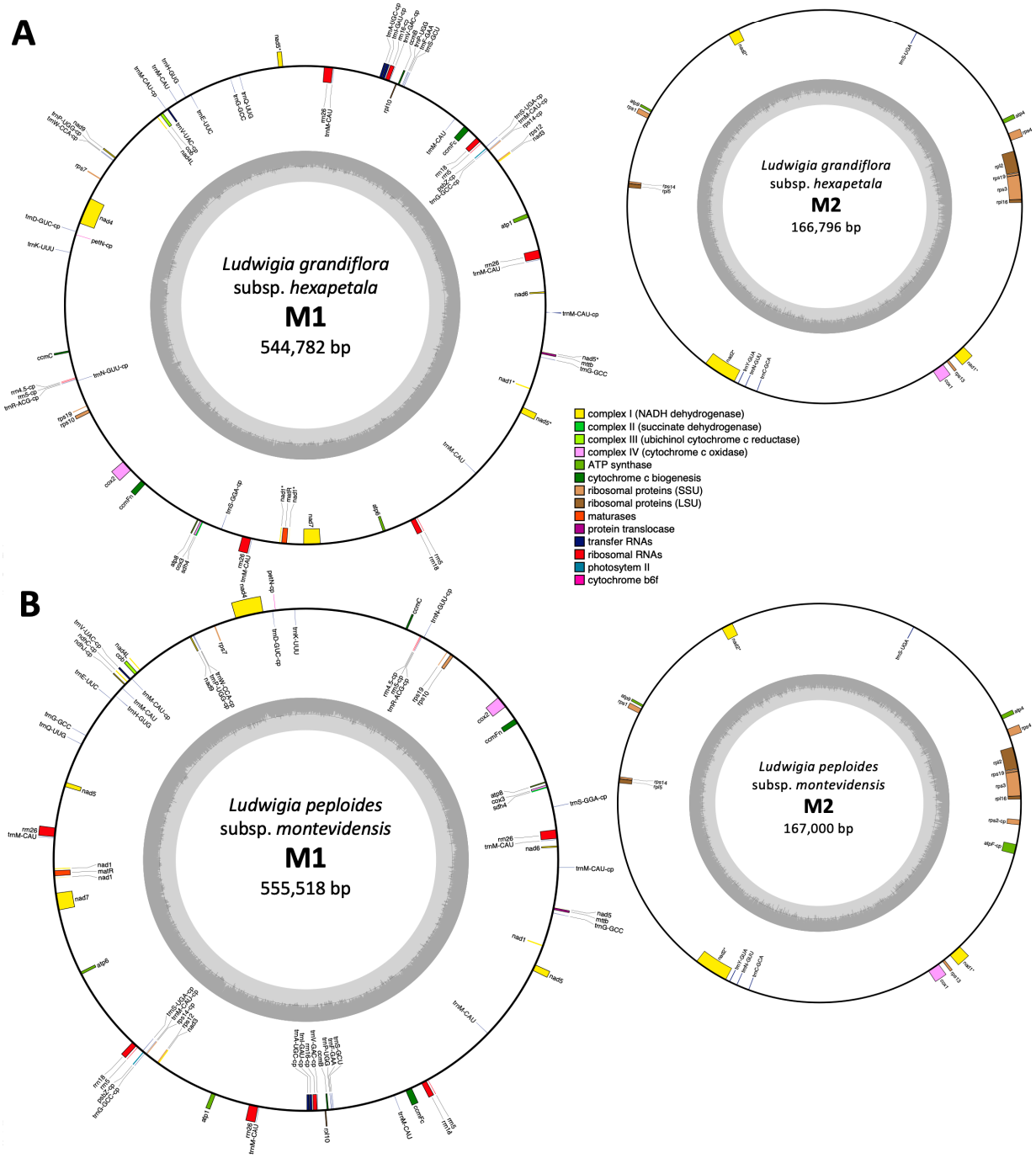


Figure 1. The two circular mitochondrial genomes of *Ludwigia grandiflora* subsp. *hexapetala* (**A**) and *Ludwigia peploides* subsp. *montevidensis* (**B**). Genes outside the circle are transcribed in a counter-clockwise direction, while inner genes are transcribed clockwise. Spliced genes are marked with an asterisk *. Genes labeled with -cp originate from the chloroplast. Inner gray circles represent the GC content. Genes are color-coded, as indicated in the legend.

2.2. Annotation and Comparisons of Ludwigia Mitogenome Contents

2.2.1. Mitogenome Organization Comparisons

To assess the rearrangement and collinearity of the *Lgh* and *Lpm* mitogenomes, their mtDNA sequences were compared using ProgressiveMauve and Lastz (Figure 2). While

the two M2 molecules were colinear and identical to 99.5%, the M1 counterparts were divided into eight locally collinear blocks (LCBs), collectively representing 88.43% and 90.27% of their total lengths. The sizes of these LCBs were well-preserved, with that of *Lpm* being consistently slightly larger than that of *Lgh* (Table S2). The LCB positions varied, with (as shown in Figure 2C) three inverted blocks following a central symmetry (LCB3, LCB4, LCB7), highlighting extensive mitogenome rearrangements between the two species of *Ludwigia*. In the five inter-LCB regions larger than 500 bp, three contained a repeated sequence of 5253 bp that was conserved between the two *Ludwigia* species. Two of these contained the genes *rrn26* and *trnM-CAU*, and the third contained a region of approximately 6156 kb, also conserved between the two species, carrying the genes *rrn18* and *rrn5*, of which a copy was present at the end of LCB3.

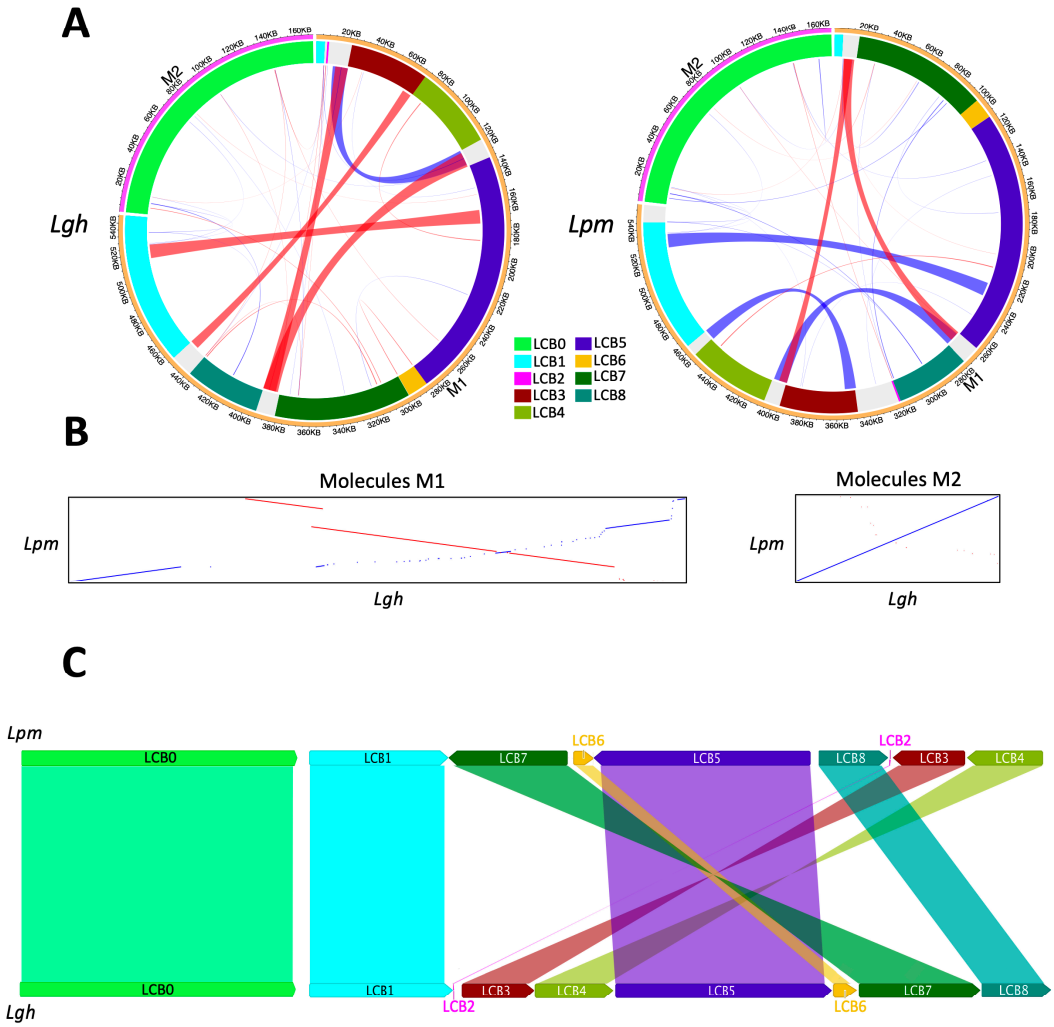


Figure 2. Structural comparisons between the *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) mitogenomes. (**A**) Repeats spanning at least 100 bp are illustrated with curved lines, with the width of the lines proportional to the size of the repeats. Inverted repeats are denoted in red, while direct repeats are indicated in blue. LCBs (locally collinear blocks) represent conserved regions between the *Lgh* and *Lpm* mitogenomes, as depicted in panel (**C**). (**B**) Comparisons of the M1 and M2 sequences of the two *Ludwigia* species were performed using LastZ: blue lines indicate common sequences in direct orientation, while red lines denote sequences that are common but in inverted orientation. (**C**) Locally collinear blocks (LCBs) (>1 kb) identified by ProgressiveMauve between the mitogenomes of *Lgh* and *Lpm*.

2.2.2. Protein Coding Gene Content

The mitogenomes of both *Ludwigia* species encompassed a total of 42–46 protein-coding genes, all located inside the LCBs, 68–70% of which were on the M1 molecule. Among these genes, 23 were involved in the oxidative phosphorylation (OXPHOS) machinery (Table 1), and of these, nine were associated with complex I (NADH dehydrogenase; *nad1*–*7*, *nad9*, and *nad4L*), with three trans-splicing genes (*nad1*, *nad2* and *nad5*) exhibiting distinct exon distributions. Exons of *nad2* were all localized on molecule M2, whereas exons of *nad5* were encoded on molecule M1, spread across LCB1 and LCB5, and exons of *nad1* were split between the two molecules. Specifically, exons 1, 4, and 5 resided in M1(LCB1 and LCB8), while exons 2 and 3 were located in M2. Significantly, the *matR* gene was encoded within exon 1 of *nad1*, a phenomenon consistently observed across all angiosperm mitogenomes (PMID: 24600456). The genes of the other complexes were as follows: *sdh4* for complex II (succinate dehydrogenase), *cob* for complex III (cytochrome c reductase; three genes in complex IV (cytochrome c oxidase; *cox1*–*3*), and five in complex V (ATP synthase; *atp1*, *4*, *6*, *8*, and *9*). Additionally, there were four subunits dedicated to the biogenesis of cytochrome c (*ccmB*, *ccmC*, *ccmFN*, and *ccmFC*), along with genes encoding maturase (*matR*) and transport membrane protein (*mttB*). These genes were distributed across the different LCBs. Furthermore, 14 non-core genes encoding small and large subunits of ribosome proteins were found to be present (Table 1). We identified seven cis-splicing protein-coding genes (*ccmFc*, *cox2*, *nad4*, *nad7*, *rpl2*, *rps3*, and *rps10*) and annotated seven protein-coding genes that originated from their respective plastids. The genes *petN* (encoding cytochrome b6f), *psbZ* (involved in photosystem II), and *rps14* (ribosomal protein S14) were transferred intact in both the *Lgh* and *Lpm* mitogenomes, whereas *ndhJ* and *ndhC* (NADH dehydrogenase), *atpF* (ATP synthase), and *rps2* (ribosomal protein S2) were intact in *Lpm* mtDNA but underwent pseudogenization in the *Lgh* mitogenome (Table S3).

Table 1. Mitochondrial protein coding genes annotated in the *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) mitogenomes. Numbers within parentheses indicate the number of exons for spliced genes. Trans-spliced genes are marked by "*". Locally collinear blocks (LCBs) correspond to conserved regions between *Lgh* and *Lpm* mitogenomes, identified by ProgressiveMauve. The molecular weight and isoelectric point of protein were calculated using the Sequence Manipulation Suite. Hypothetical localization was estimated by Plant-mPLoc.

Gene	LCB	Molecule	Size (in bp)	Molecular Weight (in kDa)	Protein Isoelectric Point	Hypothetical Localization
Complex I (NADH dehydrogenase)						
nad1 (5*)	LCB0 LCB1 LCB8	M2 M1 M1	1008	36.81	10.10	Mitochondrion
nad2 (5*)	LCB0	M2	1518	55.34	9.28	Mitochondrion
nad3	LCB3	M1	357	13.39	4.33	Mitochondrion
nad4 (4)	LCB5	M1	1473	54.27	8.97	Mitochondrion
nad4L	LCB5	M1	303	10.94	6.70	Mitochondrion
nad5 (5*)	LCB1 LCB5	M1 M1	1794	74.55	7.90	Mitochondrion
nad6	LCB1	M1	618	23.45	10.53	Mitochondrion
nad7 (5)	LCB8	M1	1182	44.32	7.87	Mitochondrion
nad9	LCB5	M1	573	22.63	8.20	Mitochondrion
Complex II (succinate dehydrogenase)						
sdh4	LCB7	M1	393	15.25	10.76	Mitochondrion

Table 1. *Cont.*

Gene	LCB	Molecule	Size (in bp)	Molecular Weight (in kDa)	Protein Isoelectric Point	Hypothetical Localization
Complex III (cytochrome c reductase)						
cob	LCB5	M1	1182	44.04	7.52	Cell membrane
Complex IV (cytochrome c oxidase)						
cox1	LCB0	M2	1623	58.75	9.25	Mitochondrion
cox2 (2)	LCB7	M1	780	29.09	5.53	Mitochondrion
cox3	LCB7	M1	798	29.54	7.53	Mitochondrion
Complex V (ATP synthase)						
atp1	LCB3	M1	1533	55.51	6.21	Mitochondrion
atp4	LCB0	M2	573	21.25	10.26	Mitochondrion
atp6	LCB8	M1	1044	38.30	4.80	Mitochondrion
atp8	LCB7	M1	480	18.19	9.50	Mitochondrion
atp9	LCB0	M2	330	11.61	9.09	Mitochondrion
Cytochrome c biogenesis						
ccmB	LCB4	M1	621	23.07	8.62	Mitochondrion
ccmC	LCB6	M1	723	26.92	10.98	Mitochondrion
ccmFc (2)	LCB4	M1	1500	56.40	10.74	Mitochondrion
ccmFn	LCB7	M1	1983	74.21	10.45	Mitochondrion
Maturase						
matR	LCB8	M1	1983	73.76	10.71	Chloroplast
Protein translocase						
mttb	LCB1	M1	768	29.07	7.75	Mitochondrion
Ribosomal large subunit						
rpl2 (3)	LCB0	M2	1389	50.69	11.28	Mitochondrion
rpl5	LCB0	M2	564	21.55	5.89	Mitochondrion
rpl10	LCB4	M1	486	18.53	7.45	Cell membrane, chloroplast, and nucleus
rpl16	LCB0	M2	435	15.98	11.81	Mitochondrion
Ribosomal small subunit						
rps1	LCB0	M2	693	25.87	11.89	Mitochondrion
rps3 (2)	LCB0	M2	1659	63.57	10.96	Mitochondrion
rps4	LCB0	M2	1047	41.18	11.63	Mitochondrion
rps7	LCB5	M1	447	17.04	11.21	Chloroplast and mitochondrion
rps10 (2)	LCB7	M1	408	16.40	11.25	Mitochondrion
rps12	LCB3	M1	378	14.30	11.80	Chloroplast and mitochondrion
rps13	LCB0	M2	351	13.27	11.21	Mitochondrion
rps14	LCB0	M2	303	11.96	12.04	Mitochondrion
rps19	LCB0 LCB7	M2 M1	285	11.20	12.32	Chloroplast

2.2.3. RNA Editing

In the mitogenomes of *Lgh* and *Lpm*, 738 and 788 post-transcriptional C to U editing events, with a probability of more than 50%, were identified within the coding DNA sequences (CDS) (Table S4; Figures S1 and S2). Among these, the *psab*, *nad4*, *ccmB*, and *mttB* genes exhibited the highest numbers of RNA editing sites, with 49, 46, 45, and 41 sites, respectively. Approximately 80% of the post-transcriptional modifications were non-synonymous (589–623 non-synonymous modifications compared to 149–165 synonymous modifications in *Lgh* and *Lpm*, respectively). Eight and twelve RNA editing events led to the creation of stop codons in the mitogenomes of *Lgh* and *Lpm*. These stop codon occurrences were predominantly observed in the chloroplast genes (*atpE*, *atpI*, and *psaB* in both mitogenomes, and *ndhK* in the *Lpm* mitogenome only), resulting in the annotation of these genes as pseudogenes. Additionally, two RNA editing sites generated stop codons in two mitochondrial genes (*atp9* and *ccmFc*), resulting in their truncation.

2.2.4. Content of RNA Genes

In both *Ludwigia* mitogenomes, we identified and annotated a comprehensive set of 34 tRNA genes (Tables S3 and S5). This included 19 native mitochondrial tRNAs and an additional 15 tRNAs derived from the plastome. Among these, sixteen tRNA genes were unique whereas *trnM-CAU* was found in nine copies, *trnG-GCC* in three copies, and *trnN-GUU*, *trnP-UGG*, and *trnS-UGA* in two copies each. Three of them were cis-splicing tRNAs (*trnA-UGC*, *trnI-GAU*, and *trnV-UAV*), all of which originated from the plastome. As shown in Table S5, all tRNAs were located in LCBs, except for the three *trnM-CAU* copies present in the previously described 5253 bp repeated region. As described previously, our analysis revealed the presence of multiple copies of mitochondrial ribosomal RNA genes (*rrn5*, *rrn18*, and *rrn26*; Table S5) along with a singular copy of the plastid ribosomal RNA genes (*rrn4.5*, *rrn5*, and *rrn16*) (Table S3).

2.2.5. Chloroplast-Derived Fragments

In addition to the previously mentioned three to seven protein-coding genes, fifteen tRNA genes, and three rRNA genes, our analysis revealed twenty-four homologous fragments between the *Lgh* and *Lpm* chloroplast genomes (Figure 3; Table S3). Cumulatively, these fragments had a size of approximately 55 kb, contributing to about 7.8% of the entire mitochondrial genome length in both *Ludwigia* species. Notably, these insertions of chloroplast DNA were dispersed throughout the entire M1 and M2 molecules (Figure 3), indicating potential independent transfers or regions modified through subsequent recombination. The largest of these fragments (8051 bp), located on the M2 molecule, was 98.4% identical to the original chloroplastic region present in both the *Lgh* and *Lpm* plastomes. This included the three pseudogenes *atpA*, *atpH*, and *atpI*, which harbors a stop codon due to post-transcriptional RNA editing, and the two genes *atpF* and *rps2*, which were complete in *Lpm* and pseudogenized in *Lgh* as well as a segment of the *rpoC2* gene. All pseudogenizations were associated with point mutations that led to frameshifts. The next largest fragment, spanning 7.6 kb and located on the M1 molecule, exhibited a lower identity to the plastome (93%) and featured an intact *trnM-CAU* gene at its extremity, accompanied by degraded forms of the *ycf2* and *ycf15* genes. In addition, we identified a 6.1 kb region with 99.6% identity to cpDNA, encompassing three tRNA genes, the *rrn16* gene, and a fragment of *rrn23*.

Another region, 5.4 kb in length, exhibited more degradation in *Lgh* compared to *Lpm* (89.3% and 91.9% identity with cpDNA, respectively), resulting in the pseudogenization of the *ndhJ* and *ndhC* genes in *Lgh* only, whereas *ndhK* and *atpE* were pseudogenized in both mitogenomes. Finally, the ultimate two extended regions of plastid origin, measuring 4.8 and 3.6 kb, respectively, were identical in both species and corresponded to the original cpDNA copy (Table S3), albeit with gene fragments (*psaA*, *ycf1*, and *rrn23*) at their termini. The remaining plastidic-originated fragments, ranging in size from 1215 to 147 bp contained either complete tRNA sequences or, for the majority, fragments of protein-coding genes.

Some of these corresponded to remains of *rps12*, *atpB*, or *psaA*, present in two copies (Table S3).

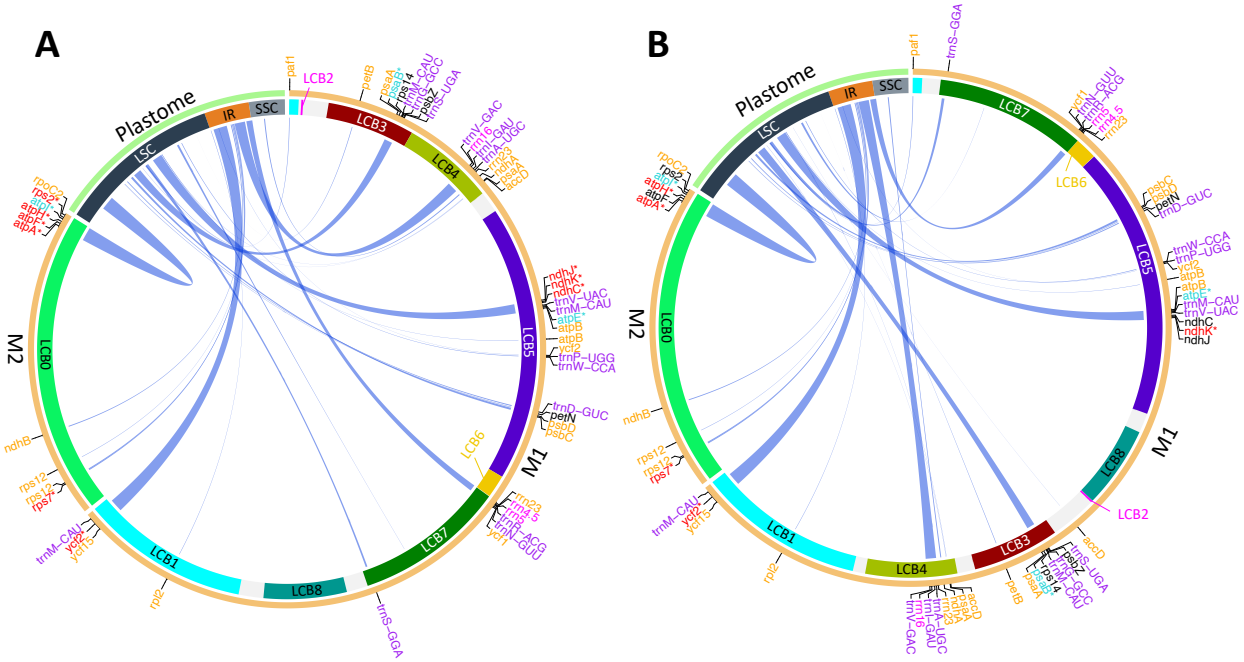


Figure 3. Maps illustrating the chloroplast sequence insertions (>80 bp) in the mitogenomes *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*; (**A**)) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*; (**B**)). Curved lines connect chloroplast insertions in the mitogenome to their respective origins in the plastome and the line width is proportional to the insertion size. Locally collinear blocks (LCBs) represent conserved regions between the *Lgh* and *Lpm* mitogenomes, identified by ProgressiveMauve. Plastome regions including large single copy (LSC), inverted repeat (IR; represented once for simplification), and small single copy (SSC), are indicated. Genes that originated from the plastome are color-coded: pseudogenes in red, pseudogenes due to RNA editing in blue, fragment genes in orange, complete protein coding genes in black, tRNA genes in purple, and rRNA genes in magenta. * Correspond to pseudogene.

2.2.6. Fragments of Mobile Genetic Elements

Four main types of endogenous fragments of mobile genetic elements were widespread in both the M1 and M2 *Ludwigia* mitogenomes: mitovirus NERVEs (nonretroviral endogenized RNA virus elements), all three classes of nuclear retrotransposons, the *Tyl/copia*, *Ty3/gypsy-*, and non-LTR/LINE-families, and plasmid derivatives (Table S6). For mitovirus NERVES, all fragments corresponded to partial or pseudogenized copies of the RNA dependent RNA polymerase (RdRP) of mitoviruses and were identical in *Lgh* and *Lpm*. The retrotransposon fragments integrated into the *Ludwigia* mitochondrial genomes were dispersed throughout the intergenic regions of the mitogenomes without showing a significant preference for integration hotspots. No complete copy was observed for any of the mobile elements identified.

2.2.7. Repeated Sequences and Recombination

Three non-tandem repeats exceeding 1 kb in size were detected within both the *Ludwigia* M1 mitogenomes, whereas none were detected in the M2 molecules (Figure 4). The first repeat encompassed a central region spanning 5641 bp (called "Core R1"; Figure 4A), was found in triplicate within both *Ludwigia* M1 molecules, and carried a gene for both 26S rRNA (*rrn26*) and the initiator methionine tRNA *trnM-CAU*. This core region exhibited a 99.9% identity between *Lgh* and *Lpm*, except for a divergent trinucleotide at position 386 (TCC in *Lgh*, GGA in *Lpm*) and a dinucleotide at position 5261 (TC in *Lgh*, GA in *Lpm*). In the

vicinity of this central "Core R1" area, we identified three scenarios: "R1-right" (Figure 4A), corresponding to an expansion in the 3′ region, which included both the "Extension" and "Right" regions present in *Lpm* and *Lgh* as well as an additional region called "Right_lgh" exclusive to *Lgh*; "R1-left" (Figure 4A), corresponding to an additional 5′ region (called "Left" for both and a small region unique to *Lgh* called "Left_lgh"), which was added to the "Core R1" part in *Lpm* and to the "Core_R1" and "Extension" regions in *Lgh*; and finally, "R1-left-right" (Figure 4A), which combined both the "R1-right" and "R1-left" regions. The second and third repeats (R2 and R3) were classic repetitive sequences, each appearing twice in the *Lpm* and *Lgh* M1 molecules. The R2 repeats, which harbored a solitary gene encoding trnG-GCC at their 3′ termini, spanned 9384 and 9376 base pairs in *Lgh* and *Lpm*, respectively. These repeats showcased a remarkable 99.9% identity across the two *Ludwigia* mitogenomes. Notably, copies found in *Lgh* possessed an additional eight bases and exhibited two single nucleotide polymorphisms (SNPs) compared to those present in *Lpm*. R2 repeats were in direct orientation in *Lpm* but were reversed in *Lgh*. *Lpm* repeats R3 were 1554 nt longer than *Lgh* repeats R3, but their common regions of 7817 nt were highly conserved (99.9% identity, 5 SNPs). Like R2, the R3 repeats were in direct orientation in *Lpm* and reversed in *Lgh.* Placing these repeats within the framework of the locally collinear blocks (LCBs) shared between *Lpm* and *Lgh*, it was apparent that all R1 repeats and one of the R3 repeats were positioned in LCBs within non-conserved regions, as depicted in Figure 4B. Meanwhile, the second iteration of R3 resided within LCB3, close to a fragment of chloroplastic origin. Additionally, the two R2 repeats were identified on LCB1 and LCB5.

By using these large repeated sequences, multiple alternative conformations were tested. In the *Lgh* mitogenome, an alternative form of the M1 molecule was found with a probability of 50%, mediated by R1-right (Figures 4B and 5). In this case, the M1 molecule was divided into two smaller molecules of 435 and 109 kb. Some other conformations mediated by R2, R1-core (The conserved part between R1-left and R1-right), and R3 repeats could not be completely validated but might also exist (Figures 4B and 5). Recombinations mediated by these three repeats led to all molecules of the same size as the original M1 mitogenome, each with a specific inversion (Figure 5). By combining all of these recombinations, three molecules of 109, 174, and 261 kb could be obtained. In the *Lpm* mitogenome, an alternative form of the M1 molecule was identified as mediated by R1-right like in *Lgh* (Figure 6). This form was composed of two molecules of 425 and 130 kb. Due to the lack of reads, we could not identify as many putative forms as in the *Lgh* M1 molecule, and only one was found with R1-core (Figure 6). This consisted of an M1 molecule of 555 kb with an inversion. A combination of the two rearrangements may also exist.

A total of 266–294 non-tandem repeats (Table S7; Figure S3) of at least 30 bp were identified in the *Ludwigia* mitogenomes and corresponded to 127–169 direct repeats and 125–139 reverse repeats. Non-tandem repeats represented 7–7.4% (50,752–52,800 bp) of each mitogenome. Most of them (161–185; 60.5–62.9%) were only found on M1, while 34—36.8% (98–100) of repeats were found between the M1 and M2 molecules and 2.6–3% (7–9) were only on the M2 molecule. The majority of repeats (192–220; 72.2–74.8%) were between 30 and 59 bp (Figure S3). Only 10.2–13.5% (30–36) of repeats were larger than 100 bp. A total of 25 and 22 tandem repeats were found in the *Lgh* and *Lpm* mitogenomes (20–24 on M1 and one or two on M2; Table S8). These repeats were in non-coding regions, except for the three of 18-period size present in each of the three copies of the rrn26 genes.

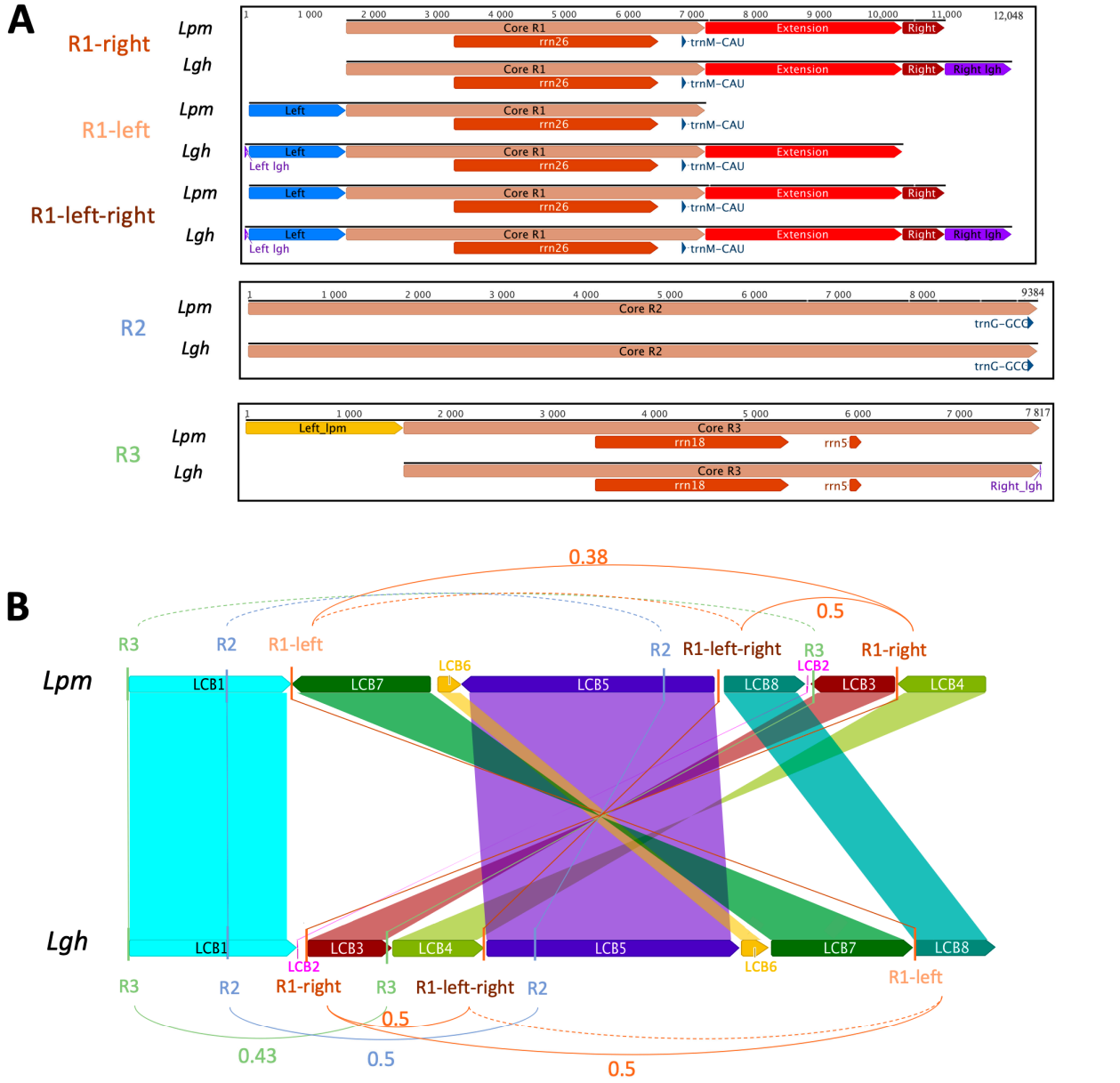


Figure 4. This representation of repeats exceeding 1 kb in the mitogenomes of *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*). "R" denotes repeat. R1-left-right combines R1-right and R1-left. (**A**) The five repeats exceeding 1 kb in the mitogenomes of *Lpm* and *Lgh*. Core sequences represent conserved sequences among repeats. "Extension" sequences were present in all repeats except R1-left in *Lpm*. Left and right sequences were unique to R1-left and R1-right, respectively, and both were present in R1-left-right. Purple and yellow sequences were unique to *Lgh* and *Lpm*, respectively. (**B**) Sequence comparison of M1 molecules between the two *Ludwigia* species and their repeats. Locally collinear blocks (LCBs) represent conserved regions between the *Lgh* and *Lpm* mitogenomes, identified by ProgressiveMauve. Recombination frequencies are specified when >0, while a dotted line indicates the absence of recombination.

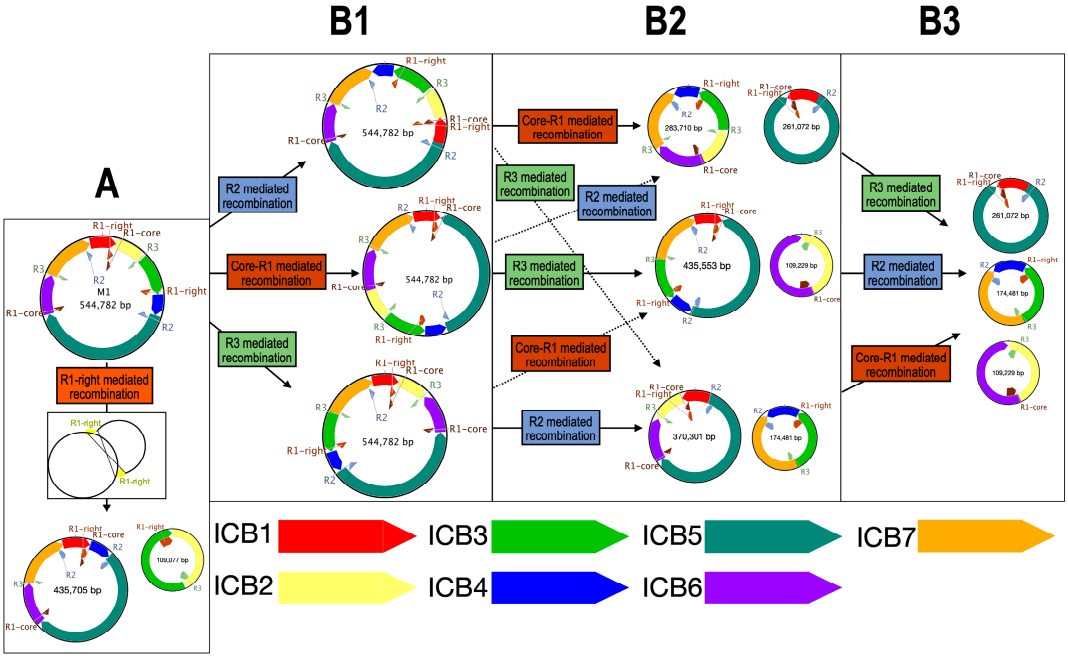


Figure 5. Schematic of mitogenome *Lgh* recombinations: (**A**) Validated conformations after M1 molecule rearrangement in *Ludwigia grandiflora* subsp. *hexapetala* mitogenome. (**B1**–**B3**) Different steps of the feasible successive recombinations. Only intramolecular rearrangements are depicted. "R" denotes repeat. ICB (intramolecular collinear block) represents regions conserved among all forms.

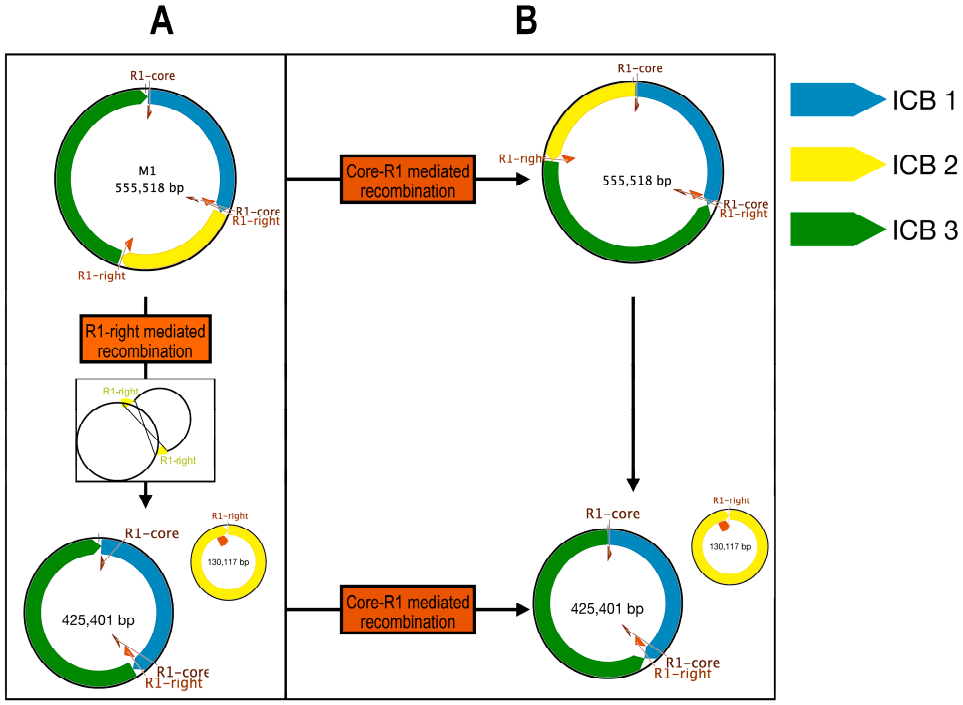


Figure 6. Schematic of mitogenome *Lpm* recombinations: (**A**) Validated conformations after M1 molecule rearrangements of *Ludwigia peploides* subsp. *montevidensis* mitogenome. (**B**) Putative conformations after M1 molecule rearrangements of *Ludwigia peploides* subsp. *montevidensis* mitogenome. Only intramolecular rearrangements are depicted. "R" denotes repeat. ICB (intramolecular collinear block) represents regions conserved among all forms.

In the *Ludwigia* mitogenomes, 183–155 SSRs were identified (137–117 and 46–38 found on the M1 and M2 molecules, respectively; Figure S4). The majority of SSRs were monomeric (66–44) and tetrameric (63–56). Monomeric repeats were almost entirely composed of A or T (30–20 A and 34–24 T). A total of 32–36 dimeric, 15–13 trimeric, and 7–6 pentameric SSRs were identified. Dimeric repeats of TA were the third most common SSR (9–10) after monomeric A and T repeats.

2.3. Gene Comparisons and Phylogeny

Relative synonymous codon usage (RSCU) was calculated using MEGA in the *Lgh* and *Lpm* mitogenome CDS on 11,595 and 12,141 codons, respectively (Figure 7; Table S9). A total of 30–29 codons had an RSCU > 1, which means that they were used more often than another synonymous codon. Among them, the most preferentially used codons were A-ended or U-ended, except for the UGG leucine codon, showing a strong bias in the third letter of the codon. In both mitogenomes, the UGA stop codon had the biggest RSCU value (1.6–1.64). In *Lgh*, the smallest RSCU value was for the UAG stop codon (0.29) while in *Lpm*, it was the GCG alanine codon (0.42).

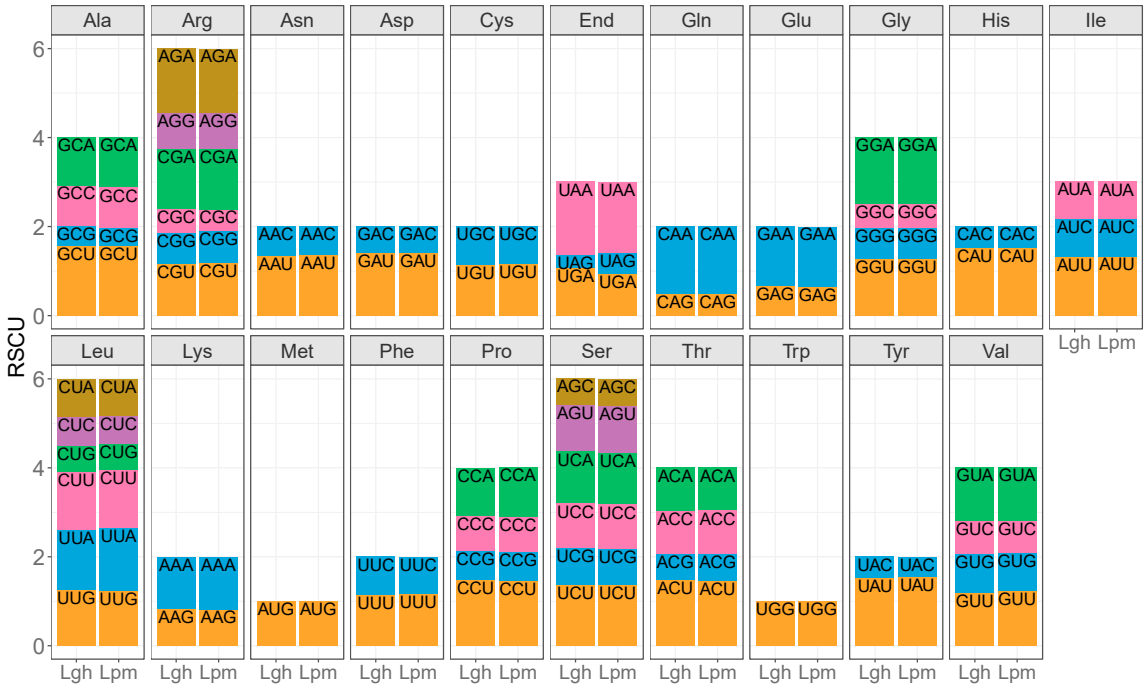


Figure 7. The relative synonymous codon usage (RSCU) bias in *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) mitogenome CDS (sequences of protein coding genes).

To identify genes under selection, Ka (non-synonymous substitution rate) and Ks (synonymous substitution rate) were calculated on 31 genes by comparing *Ludwigia* mitogenomes (these genes were identical in *Lgh* and *Lpm* mitogenomes) to each of the other 14 assembled Myrtales mitogenomes one by one, and to the *Geranium maderense* mitogenome as an outgroup. Genes under positive selection (not conserved) will have a Ka/Ks value above one, whereas genes under negative selection (conserved) will have a Ka/Ks value under one. A Ka/Ks ratio value of one points to a neutral selection. The *rps1* gene is the only gene showing a Ka/Ks >1 in the majority of species studied (Figure 8, Table S10). Moreover, the *ccmB* and *mttB* genes were also under negative selection in *Lgh* and *Lpm* in seven and five species, respectively. However, the majority of Ka/Ks ratios (92%) were under 1, showing the conservation of most genes.

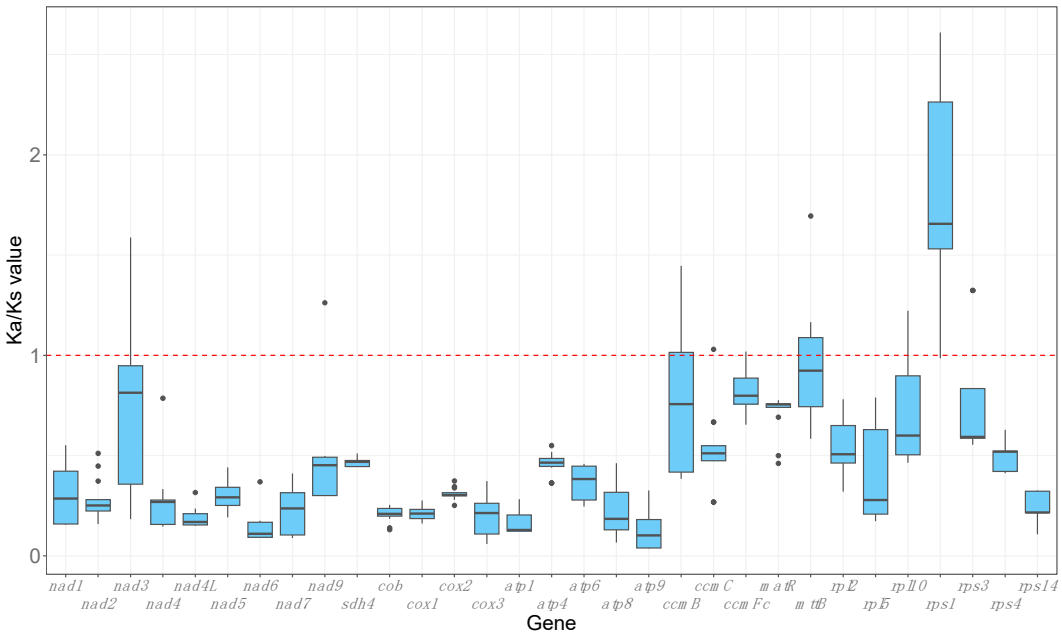


Figure 8. Graphical representation of pairwise comparisons of non-synonymous substitution rate (Ka) and synonymous substitution rate (Ks) ratios among species from the orders Myrtales and Geraniales, utilizing *Ludwigia grandiflora* subsp. *hexapetala* and *Ludwigia peploides* subsp. *montevidensis* as references, across 31 mitochondrial genes. Genes were included if they were present in at least 14 out of the 16 selected species.

Sixteen conserved protein-coding genes were concatenated and aligned to make a phylogenetic tree of the Myrtales order with *Geranium maderense* used as an outgroup (Figure 9). The tree fit with the phylogeny in the Myrtales order. The outgroup *G. maderense* was separated from the Myrtales order species. Species from the same family were grouped. The four Onagroideae subfamily species were clustered together within the Onagraceae family. Finally, the two *Ludwigia* species were on the same branch.

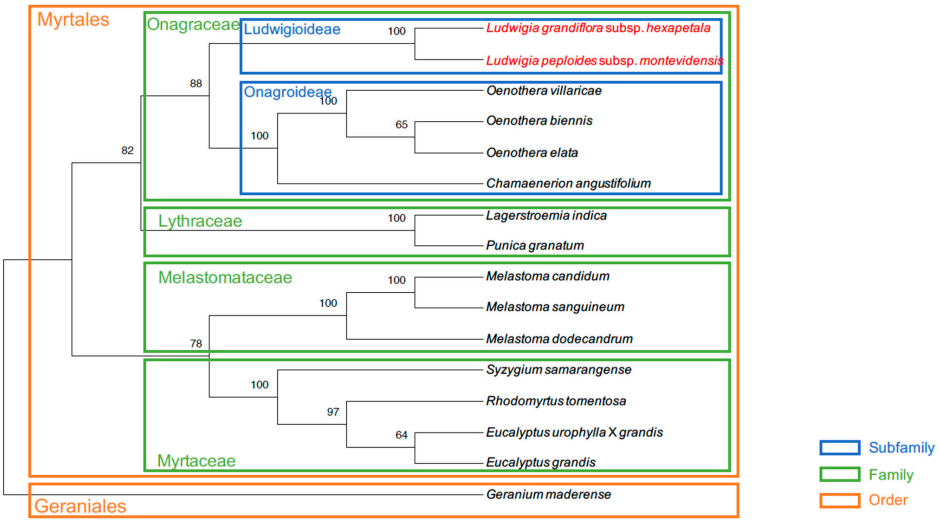


Figure 9. Phylogenetic tree of the Myrtales order constructed using 16 protein-coding genes via the maximum likelihood method with a bootstrap value of 1000. *Geranium maderense* was designated as the outgroup. Branch frequencies are denoted by numbers on branches. The *Ludwigia* species sequenced in this paper are highlighted in red.

Gene compositions were compared in the 15 Myrtales order species and *G. maderense* mitogenomes (Figure 10). Sixteen genes were common between all of the studied mitogenomes. Ten genes were only missing in one mitogenome. Among them, nine were missing in *R. tomentosa* and one in *E. urophylla* × *grandis*. Mitochondrial genes missing in more than one mitogenome were ribosomal subunit genes (*rpl16*, *rps7*, *rps10*, *rps12*, *rps13*, and *rps19*), *ccmFn* and *sdh3*. Out of all species, 83% of the identified chloroplast protein coding genes were incomplete. Only five species contained one to nine complete chloroplast genes: *Lgh*, *Lpm*, *M. candidum*, *M. sanguineum*, and *E. grandis*. These genes coded for different functions: PSI assembly factor (*ycf3*), photosystem II (*psbM* and *psbZ*), ATP synthase (*atpF*, *atpH*, and *atpI*), cytochrome b6f (*petN*), NADH dehydrogenase (*ndhB*, *ndhC*, and *ndhJ*), transcription (*rpoC1*), ribosomal large subunit (*rpl14*, *rpl22*, and *rpl23*), and ribosomal small subunit (*rps2* and *rps14*). Each of these genes were present in only one or two species, always of the same genus.

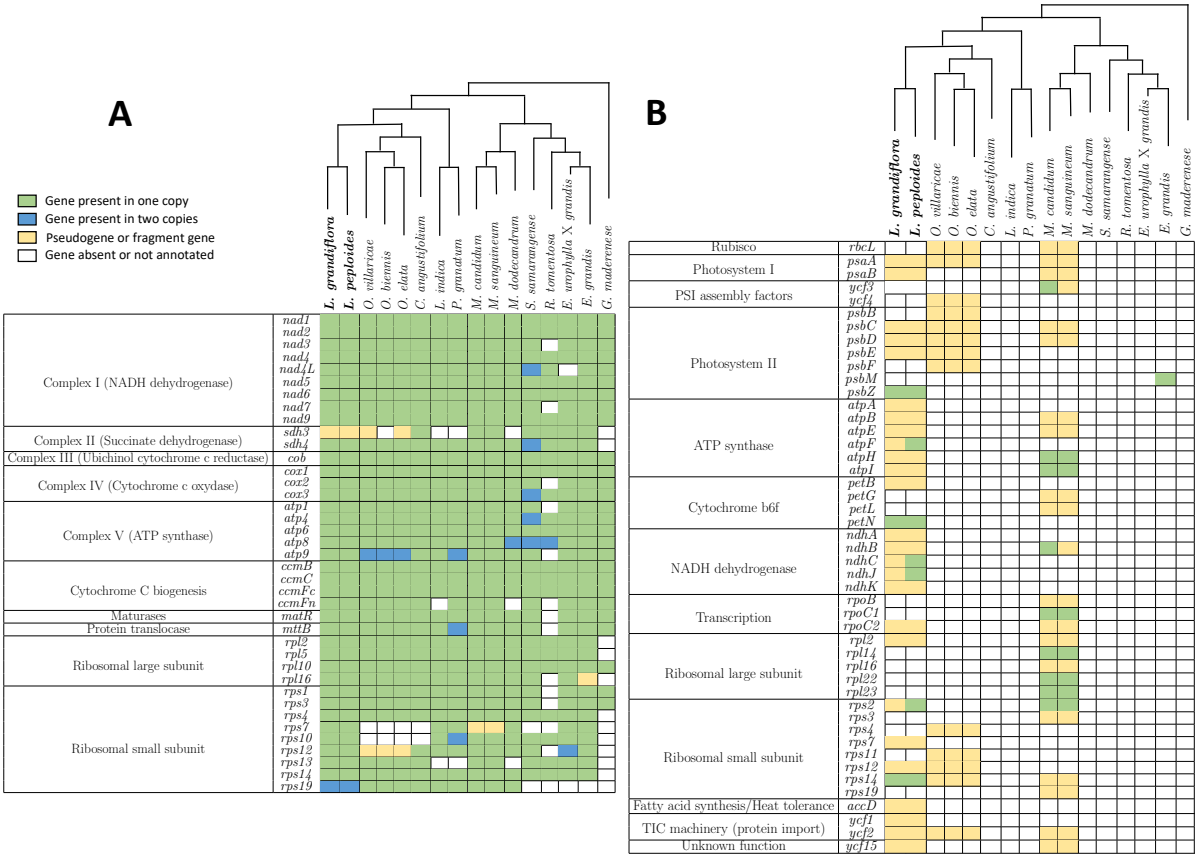


Figure 10. Gene compositions of mitogenomes from 15 Myrtales and one Geraniales order species. Panel (**A**) displays the mitochondrial genes, while Panel (**B**) shows the chloroplast genes. The phylogenetic trees presented were based on 16 protein-coding genes and constructed using the maximum likelihood method (refer to Figure 9). The *Ludwigia* species sequenced in this paper are highlighted in bold. Color boxes correspond to presence/absence panel (see legend in figure).

3. Discussion

In this study, *Lgh* and *Lpm* mitogenomes were assembled de novo from the SR or/and LR sequences. For *Lgh*, the best assembling combination was found to be SPAdes and Flye. Two circular molecules called M1 and M2 were obtained for both mitogenomes and had similar sizes (544,782 bp and 555,518 bp for M1 and 166,796 bp and 158,656 bp for *Lgh* and *Lpm*, respectively). For these two closely related species, their mtDNAs showed similarities in their structures and sequences for both the M1 and M2 molecules. While the

M2 molecules were highly similar (99.5%) and colinear, the M1 molecules from *Lpm* and *Lgh* mitogenomes contained eight conserved blocks including three inverted blocks.

3.1. Comparison of Plant Mitogenomes in Myrtales Order

Plant mitogenomes vary greatly in size, ranging from 66 kb to 12 Mb. In the Myrtales order, which comprises 9 families, 399 genera, and about 13,000 species [88], only the mitochondria genomes of 17 species have been sequenced. Most of these sequenced mitogenomes are around 400 to 450 kb in size, as follows for the Lythraceae family (*Punica granatum* (404,807 bp) [89], *Lagerstroemia indica* (333,948 bp; NC 035616), *Trapa incisa* (381,774 bp; NC 086691), *T. bicornis* (383,262 bp; NC 086690)); Myrtaceae family (*Rhodomyrtus tomentosa* (400,482 bp) [90], *Eucalyptus grandis* (478,813 bp) [91], *E. camaldulensis* (463,134 bp) [92], *E. urophylla* × *grandis* (481,982 bp; OQ947725-OQ947727), *Syzygium samarangense* (530,242 bp; NC 079700)); Melastomataceae family (*Medinilla magnifica* (377,864 bp) [93], *Melastoma candidum* (391,595 bp; NC 071383), *M. sanguineum* (395,542 bp; NC 071384), and *M. dodecandrum* (411,954 bp) [94]); and the Onagraceae family (*Oenothera villaricae*, *O. biennis* and *O. elata*, with respective lengths of 408,260, 424,132, and 418,451 bp [37]) and *Chamaenerion angustifolium* (495,176 bp; OX328283)). For *Lgh* and *Lpm*, however, the total length of mitogenomes was notably higher, standing at approximately 700 kb (711,578 bp and 722,518 bp, respectively), and even the largest M1 molecules alone exhibited a greater size than that of the mitogenomes of the previous 17 species.

The origin of this variation in size between different plant mitochondrial genomes is multifaceted, encompassing the proliferation of repeat elements, the integration of foreign DNA, and the acquisition or loss of large intragenic segments [95]. Homologous recombination has led to rearrangements that play a significant role in plant mitogenome evolution [96]. In their study, Zhou et al. (2023) conducted pairwise comparisons of the mitogenomes of five Myrtales species and observed sequence similarities ranging from 17% to 36% [94]. Notably, the highest percentage of shared sequences was observed between *M. magnifica* and *M. dodecandrum*, which belong to the same family. Similarly, the comparison of the mitogenomes of *E. camaldulensis* and *E. grandis* revealed a distinct alteration in repeat sequences, shedding light on evolutionary changes between the two genomes [92]. In the mitogenomes of *Lgh* and *Lpm*, the smaller M2 molecules exhibited colinearity and shared 99.5% identity, whereas the larger M1 molecules displayed rearrangements in five inter-LCB regions (size > 1 kb) between the two species, although no alterations were observed in the repeat sequences. One explanation for this result could be the genomic proximity of these closely related species, a hypothesis supported by the fact that *Lpm* is one of the maternal progenitor species of *Lgh* [33]. Given that most plants inherit mitochondrial genomes from the maternal parent, the minimal difference observed between the two mitogenomes might signify the evolutionary divergence between *Lpm* and *Lgh*.

3.2. Existence of Multi-Chromosomal Mitogenomes

Initially, most plant mitogenomes were depicted as single circular molecules. However, the increased availability of plant mitogenomes has unveiled a greater complexity, with multi-chromosomal mitogenomes identified in many species. However, research by Wu et al. (2020) indicates that multiple chromosome structures are present in only about 10% of sequenced plant mitogenomes [95]. Significantly, the mitochondrial genomes of *Lgh* and *Lpm* stand out as inaugural examples of multiple mitogenome structures documented within the Myrtales order. In this context, Bi et al. (2020) posed an intriguing question regarding the mechanisms that regulate and control the replication and segregation of multi-chromosomal mitogenomes during cell division [97]. In our study, we observed that for *Ludwigia*, sequence coverage was consistent for both the M1 and M2 molecules, indicating an equal number of copies of both molecules in the cell and therefore the likelihood of equi-replication.

Recently, Yang et al. (2022) reconstructed the mitogenome of *Salvia miltiorrhiza*, which was initially assembled as a single circular molecule through short-read sequence analy-

sis [98]. By employing both long-read and short-read technologies, the researchers identified two mitochondrial chromosomes, highlighting the value of combining both sequencing technologies. This finding reinforces our research, demonstrating that the integration of short-read (SR) and long-read (LR) sequences leads to improved mitogenome assembly. Notably, SR sequences, which are shorter than the repeat size, may not fully span the repeats independently, highlighting the necessity of combining both technologies for comprehensive assembly.

The presence of large repeat sequences (longer than 1 kb) has been shown to induce structural changes or isomerization in plant mitogenomes [58,99]. Hence, recent studies suggest the presence of one to two dominant forms of sub-genomic mitochondrial chromosomes, with repeat sequences capable of mediating recombination, leading to a diverse array of minor conformations. For instance, in *Gelsemium elegans*, four pairs of repeat sequences have been shown to facilitate the formation of one major and five minor conformations of the mitochondrial genome [100]. In *Salvia miltiorrhiza*, Yang et al. (2022) discovered two mitochondrial genomes and identified multiple minor conformations resulting from homologous recombination mediated by nine repeat sequences [98]. In our study, we identified two dominant forms, M1 and M2, and validated, through long-read (LR) sequence mapping, the existence of two minor conformations for the M1 molecule in both *Lgh* and *Lpm*, mediated by three repeat sequences. Other configurations are theoretically possible, but could not be confirmed with the sequences available.

3.3. Plastid DNA Transfer in Lpm and Lgh Mitogenomes

Inter-organelle DNA transfer can occur between the mitochondria and plastids, although it is extremely rare in the direction of mitogenomes to plastids. The plastome sequences represent 0% to 10% of the mitochondrial genome. The presence of chloroplast sequences and the number of genes in mitochondrial genome are very fluctuant. Recently, Wang et al. (2018) demonstrated that plastid-derived DNA fragments comprised 10.34% of the entire mitogenome length of *P. micranthus*, encompassing twelve intact protein-coding genes, three tRNAs, and twenty-nine pseudogenes [101]. Similarly, Tang et al. (2024) observed for *Acorus tatarinowii* and Yang et al. (2024) for *Haloxylon Ammodendron* that homologous sequences between the mitochondrial genome and the chloroplast genome accounted for 3.78% and 2.49%, respectively [102,103]. In our study, both the *Lgh* and *Lpm* M1 and M2 molecules contained plastome-origin sequences, encompassing 21–25 intact genes and constituting 7.8% of the mitochondrial genome length, which is quite a high percentage compared to the studies cited above, particularly given the large size of *Ludwigia* mitogenomes.

RNA editing is a post-transcriptional mechanism in higher plant organelles, and contributes to improved protein folding [104]. The conversion of C-to-U is the main form of RNA editing, as we observed in *Lgh* and *Lpm* [105]. RNA editing plays a crucial role in various plant developmental processes and could serve as a vital mechanism in adapting to a changing environment, allowing plants to acclimatize and survive [106]. Given the remarkable plasticity of both *Ludwigia* species to thrive in aquatic to dry environments [23], the role of RNA editing in facilitating this acclimatization warrants investigation. We identified the highest number of RNA editing sites among all mitochondrial genes in the *psab*, *nad4*, *ccmB*, and *mttB* genes. Similarly, Lu et al. (2024) reported elevated RNA editing sites in the *ccmB* and *nad4* genes in *Syzygium samarangense* [107]. Additionally, Xiong et al. (2017) investigated the responses of six mitochondrial genes including *ccmB* to oxidative stress, commonly associated with abiotic stresses like drought [108]. They demonstrated that differences in the RNA editing of *ccmB* could modulate plant performance responses under oxidative stress.

To summarize, the *Lpm* and *Lgh* mitogenomes consisted of two circular molecules with similar structures and sequences. The presence of large repeat sequences could result in multiple minor conformations. Plastome-origin sequences accounted for 7.8% of both mitochondrial genomes. A high number of RNA editing was observed.

4. Materials and Methods

4.1. Plant Material and Sequencing Data

To avoid genetic variability, *Lgh* and *Lpm* plants used for the sequencing were obtained using vegetative reproduction from a single fragment of stem with buds for each species. DNA extractions were carried out from new buds according to a specific protocol for extracting DNA long fragments [32]. DNA fragments were sequenced using Illumina MiSeq for short reads (SRs) and Oxford Nanopore GridION for long reads (LRs) at the Genome Transcriptome Facility of Bordeaux (INRAE-UMR 1202 BIOGECO, Bordeaux, France), as described in Barloy-Hubler et al. (2023). Both LRs and SRs were available for *Lgh* but only LRs for *Lpm*. Quality controls and filtering were performed using Guppy v4.0.14 for the nanopore reads and fastp v0.20.0 for the MiSeq reads [109]. Short reads were then paired and merged using BBMerge [110].

4.2. Assembly Strategies and Annotations

For the assembly of the *Lgh* mitogenome, SRs were corrected with LRs using SPAdes and LRs with SRs using Ratatosk [111,112]. Subsequently, Flye was used to assemble the corrected LRs, Megahit to assemble the corrected SRs, and SPAdes for the corrected SR-LR hybrid assemblies [113,114].

To recruit mitochondrial reads, we developed a strategy centered on the coding DNA sequences (CDSs) due to their high conservation across species [57]. Therefore, all mitogenomes from the malvids clade were retrieved from the NCBI RefSeq database, and each CDS was identified, aligned, and a consensus sequence was constructed to serve as references to map corrected LRs. LRs selected in this manner were then assembled using the 'De Novo Assemble' tool within Geneious Prime 2023 (http://www.geneious.com/, accessed on 28 June 2024) and the resulting contigs were elongated by successive mapping of the LRs, SRs, and SPAdes (only for *Lgh*) and Flye assemblies. The *Lpm* mitogenome was assembled both de novo and using the *Lgh* mitogenome as a reference. For *Lgh*, as we had both LRs and SRs, we compared the hybrid (SPAdes), LR-only (Flye), and SR-only (Megahit) assemblies using the QUality ASsessment Tool (QUAST) [115].

CDS and tRNA annotations were performed using Geseq and BLASTx [116,117]. Rfam was used to annotate the rRNA [118]. All annotations were manually verified and biocurated. OGDraw was used for the graphic representation of the mitogenomes [119]. For the RNA editing predictions, CDSs were extracted and Deepred-MT was used to find the C-to-U post-transcriptional modifications [120]. Five genes were only analyzed in *Lpm* (*atpF*, *ndhC*, *ndhJ*, *ndhK*, and *rps2*) as they were annotated as pseudogenes in the *Lgh* mitogenome. Mobile elements were annotated using DANTE via the RepeatExplorer Galaxy server and each region identified was checked for both gene designation and boundaries using BlastN and BlastP [121].

4.3. Sequence Analysis

REPuter on BiBiServ2 (https://bibiserv.cebitec.uni-bielefeld.de, accessed on 28 June 2024) identified direct and reverse repeats (maximum computed repeats = 500; minimal repeat size = 8) [122]. Simple sequence repeats (SSR) were annotated by MISA with the parameters set to 10, 5, 4, 3, 3, and 3 for the mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, respectively [123]. Tandem Repeats Finder was used to identify tandem repeats [124]. Chloroplast sequence insertions in the mitochondrial genome were found using ProgressiveMauve with the plastome [125]. Circular repeats maps were made using shinyCircos [126]. Each repeat longer than 1 kb was extracted, along with 1 kb regions on each side, and recombination events were constructed. Then, the LRs were mapped to all of these sequences using the 'Map to Reference' function in Geneious, and sequences that overlapped the repeats and covered at least 100 bp of the two flanking 1 kb regions were tallied for all conformations and utilized to identify potential recombinations. If the repeats were in the same direction, the molecule was divided into two molecules at the repeat locations. If the repeats were in the opposite direction, an inversion occurred between the

two repeats. A general sequences comparison of the *Ludwigia* mitogenomes was conducted by using LastZ [127]. ProgressiveMauve identified LCBs (locally collinear blocks) between the *Lgh* and *Lpm* mitogenomes.

4.4. Codon Usage Bias and Ka/Ks Calculation

All CDSs were extracted from each *Ludwigia* mitogenome and concatenate. MEGA was used to count the presence of each codon and calculate their RSCU (relative synonymous codon usage) [126].

The complete mitogenomes of 13 species belonging to the Myrtales order were collected from the NCBI RefSeq database: *Oenothera biennis* (MZ934756), *Oenothera elata* (MZ934757), *Oenothera villaricae* (MZ934755) [37], *Chamaenerion angustifolium* (OX328283), *Lagerstroemia indica* (NC 035616), *Eucalyptus grandis* (NC 040010) [91], *Eucalyptus urophylla* × *grandis* (OQ947727), *Melastoma dodecandrum* (OR148386) [94], *Melastoma candidum* (NC 071383), *Melastoma sanguineum* (NC 071384), *Punica granatum* (NC 071229) [89], *Rhodomyrtus tomentosa* (NC 071968) [90], and *Syzygium samarangense* (NC 079700) [107]. In addition to these 13 mitogenomes, the mitogenome of *Geranium maderense* (NC 027000) from the Geraniales order was also collected and used as the outgroup species. *C. angustifolium* was annotated with Geseq [116]. Thirty-one genes present in at least 14 of the 16 studied mitogenomes (14 collected in database and *Lpm* and *Lgh*) were selected (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *sdh4*, *cob*, *cox1*, *cox2*, *cox3*, *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *matR*, *mttb*, *rpl2*, *rpl5*, *rpl10*, *rps1*, *rps3*, *rps4*, and *rps14*). Protein coding genes were aligned using MAFFT, and MEGA was used to calculate Ka (non-synonymous substitution rate) and Ks (synonymous substitution rate) with the Nei–Gojobori (Jukes–Cantor) method, of these 31 genes between *Lgh* as the reference, and the other 15 species [127,128].

4.5. Molecular Mass, Isoelectric Point and Subcellular Localization Predictions

Sequences of mitochondrial protein coding genes were translated by Geneious. Protein sequences were used to calculate the molecular mass and isoelectric point with Sequence Manipulation Suite [129]. These were also used for subcellular localization prediction with Plant-mPLoc [130].

4.6. Phylogenetic and Comparative Analysis

Mitogenomes used in the Ka/Ks calculation were also used for phylogenetic analysis. Sixteen genes (*nad1*, *nad2*, *nad4*, *nad5*, *nad6*, *nad9*, *cob*, *cox1*, *cox3*, *atp4*, *atp6*, *atp8*, *ccmB*, *ccmC*, *ccmFc*, and *rpl10*), common between the 16 mitogenomes selected, were concatenated and aligned with MAFFT. MEGA was used to make a tree with the maximum likelihood method using a bootstrap of 1000 r.

5. Conclusions

To conclude, we obtained the first high-quality mitogenomes of two *Ludwigia* species, *Ludwigia peploides* subsp. *montevidensis* and *Ludwigia grandiflora* subsp. *hexapetala*. The *Lpm* and *Lgh* mitogenomes were the first of the Ludwigioideae subfamily to be sequenced, which regroups 83 species, and the fifth and sixth in the Onagraceae family, which includes approximately 660 species, and correspond to the eighteenth and nineteenth to be sequenced in the Myrtales order. The fact that there are so few mitogenomes in the Myrtales order compared to plastomes (342 in the RefSeq database in April 2024) highlights how challenging it is to assemble mitogenomes. In addition, *Lgh* and *Lpm* mitogenomes were the first to be found as two circular molecules in the Myrtales order. Our results can be used as a reference for the enrichment of mitochondrial resources in the Myrtales order, which is considered to be one of the largest in the Angiosperms with around 12,000 species [131].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms25137283/s1.

Author Contributions: D.B. and F.B.-H. designed the project and defined the methodological approaches. G.D. carried out the data analysis and tested all of the data assembly and biocuration strategies. All authors wrote different parts of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funding by INRAE through the doctoral grant of Guillaume Doré.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contribution presented in the study are included in this article (see Supplementary Materials). Accession numbers of the *Lpm* and *Lgh* mitogenomes are indicated in the main manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rai, P.K. Environmental Degradation by Invasive Alien Plants in the Anthropocene: Challenges and Prospects for Sustainable Restoration. *Anthr. Sci.* **2022**, *1*, 5–28. [CrossRef]
2. Pyšek, P.; Hulme, P.E.; Simberloff, D.; Bacher, S.; Blackburn, T.M.; Carlton, J.T.; Dawson, W.; Essl, F.; Foxcroft, L.C.; Genovesi, P.; et al. Scientists' Warning on Invasive Alien Species. *Biol. Rev.* **2020**, *95*, 1511–1534. [CrossRef]
3. Gioria, M.; Hulme, P.E.; Richardson, D.M.; Pyšek, P. Why Are Invasive Plants Successful? *Annu. Rev. Plant Biol.* **2023**, *74*, 635–670. [CrossRef] [PubMed]
4. Mikulyuk, A. *Ludwigia grandiflora* (Water Primrose). *CABI Compend.* **2009**, 109148. [CrossRef]
5. Grewell, B.; Netherland, M.; Skaer Thomason, M. Establishing Research and Management Priorities for Invasive Water Primroses (*Ludwigia* spp.). 2016. Available online: https://www.researchgate.net/publication/295856718_Establishing_Research_and_Management_Priorities_for_Invasive_Water_Primroses_Ludwigia_spp (accessed on 28 June 2024).
6. Okada, M.; Grewell, B.J.; Jasieniuk, M. Clonal Spread of Invasive *Ludwigia hexapetala* and *L. Grandiflora* in Freshwater Wetlands of California. *Aquat. Bot.* **2009**, *91*, 123–129. [CrossRef]
7. Thouvenot, L.; Haury, J.; Thiebaut, G. A Success Story: Water Primroses, Aquatic Plant Pests. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **2013**, *23*, 790–803. [CrossRef]
8. Dandelot, S.; Matheron, R.; Le Petit, J.; Verlaque, R.; Cazaubon, A. Variations Temporelles des Paramètres Physicochimiques et Microbiologiques de Trois Écosystèmes Aquatiques (Sud-est de la France) Envahis par des *Ludwigia*. *Comptes Rendus Biol.* **2005**, *328*, 991–999. [CrossRef]
9. Portillo Lemus, L.O.; Bozec, M.; Harang, M.; Coudreuse, J.; Haury, J.; Stoeckel, S.; Barloy, D. Self-Incompatibility Limits Sexual Reproduction Rather than Environmental Conditions in an Invasive Water Primrose. *Plant-Environ. Interact.* **2021**, *2*, 74–86. [CrossRef]
10. Thiébaut, G.; Dutartre, A. Management of invasive aquatic plants in France. In *Aquatic Ecosystem Research Trends*; Nova Science Pub Inc.: New York, NY, USA, 2009.
11. Nehring, S.; Kolthoff, D. The Invasive Water Primrose *Ludwigia grandiflora* (Michaux) Greuter & Burdet (Spermatophyta: Onagraceae) in Germany: First Record and Ecological Risk Assessment. *Aquat. Invasions* **2011**, *6*, 83–89. [CrossRef]
12. Celesti-Grapow, L.; Alessandrini, A.; Arrigoni, P.V.; Banfi, E.; Bernardo, L.; Bovio, M.; Brundu, G.; Cagiotti, M.R.; Camarda, I.; Carli, E.; et al. Inventory of the Non-native Flora of Italy. *Plant Biosyst.-Int. J. Deal. All Asp. Plant Biol.* **2009**, *143*, 386–430. [CrossRef]
13. Bou Manobens, J.; Portillo, L.; Curcó i Massip, A. New Contributions to Allochthonous Ludwigia Species (Onagraceae) on Catalonia. *Butlletí Inst. Catalana D'història Nat.* **2019**, 45–48. [CrossRef]
14. Lindemann-Matthies, P. Beasts or Beauties? Laypersons' Perception of Invasive Alien Plant Species in Switzerland and Attitudes towards Their Management. *NeoBiota* **2016**, *29*, 15–33. [CrossRef]
15. Gallardo, B.; Aldridge, D.C. The 'Dirty Dozen': Socio-Economic Factors Amplify the Invasion Potential of 12 High-Risk Aquatic Invasive Species in Great Britain and Ireland. *J. Appl. Ecol.* **2013**, *50*, 757–766. [CrossRef]
16. Arslan, Z.F.; Uludag, A.; Uremis, I. Status of Invasive Alien Plants Included in EPPO Lists in Turkey. *EPPO Bull.* **2015**, *45*, 66–72. [CrossRef]
17. Hieda, S.; Kaneko, Y.; Nakagawa, M.; Noma, N. *Ludwigia grandiflora* (Michx.) Greuter & Burdet subsp. *Hexapetala* (Hook. & Arn.) GL Nesom & Kartesz, an Invasive Aquatic Plant in Lake Biwa, the Largest Lake in Japan 2020. *Acta Phytotaxon. Geobot.* **2020**, *71*, 65–71.
18. EPPO. *Ludwigia grandiflora* and *L. Peploides* onagraceae—Water Primroses. *EPPO Bull.* **2011**, *41*, 414–418. [CrossRef]
19. Sarika, M. *Ludwigia peploides* subsp. Montevidensis, a New Allien Taxon for the Flora of Greece and the Balkans. *J. Biol. Res.* **2006**, *5*, 71–78.
20. Dandelot, S. Les *Ludwigia* spp. invasives du Sud de la France: Historique, Biosystématique, Biologie et Ecologie. Ph.D. Thesis, Université Paul Cézanne: Aix-Marseille III, Marseille, France, 2004.

21. Muller, S. *Plantes invasives en France*; Muséum National d'Histoire Naturelle: Paris, France, 2004.
22. Rolon, A.S.; Lacerda, T.; Maltchik, L.; Guadagnin, D.L. Influence of Area, Habitat and Water Chemistry on Richness and Composition of Macrophyte Assemblages in Southern Brazilian Wetlands. *J. Veg. Sci.* **2008**, *19*, 221–228. [CrossRef]
23. Hernández-R, J.; Rangel-Ch, J.O. Vegetation of the Wetland Jaboque (Bogotá, DC). *Caldasia* **2009**, *31*, 355–379.
24. Haury, J.; Druel, A.; Cabral, T.; Paulet, Y.; Bozec, M.; Coudreuse, J. Which Adaptations of Some Invasive *Ludwigia* Spp. (Rosidae, Onagraceae) Populations Occur in Contrasting Hydrological Conditions in Western France? *Hydrobiologia* **2014**, *737*, 45–56. [CrossRef]
25. Billet, K.; Genitoni, J.; Bozec, M.; Renault, D.; Barloy, D. Aquatic and Terrestrial Morphotypes of the Aquatic Invasive Plant, *Ludwigia grandiflora*, Show Distinct Morphological and Metabolomic Responses. *Ecol. Evol.* **2018**, *8*, 2568–2579. [CrossRef] [PubMed]
26. Moravcová, L.; Pyšek, P.; Jarošík, V.; Pergl, J. Getting the Right Traits: Reproductive and Dispersal Characteristics Predict the Invasiveness of Herbaceous Plant Species. *PLoS ONE* **2015**, *10*, e0123634. [CrossRef] [PubMed]
27. Mounger, J.; Ainouche, M.L.; Bossdorf, O.; Cavé-Radet, A.; Li, B.; Parepa, M.; Salmon, A.; Yang, J.; Richards, C.L. Epigenetics and the Success of Invasive Plants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2021**, *376*, 20200117. [CrossRef] [PubMed]
28. Hodkinson, T.R.; Perdereau, A.; Klaas, M.; Cormican, P.; Barth, S. Genotyping by Sequencing and Plastome Analysis Finds High Genetic Variability and Geographical Structure in *Dactylis glomerata* L. in Northwest Europe Despite Lack of Ploidy Variation. *Agronomy* **2019**, *9*, 342. [CrossRef]
29. Chown, S.L.; Hodgins, K.A.; Griffin, P.C.; Oakeshott, J.G.; Byrne, M.; Hoffmann, A.A. Biological Invasions, Climate Change and Genomics. *Evol. Appl.* **2015**, *8*, 23–46. [CrossRef]
30. Hamelin, R.; Roe, A. Genomic Biosurveillance of Forest Invasive Alien Enemies: A Story Written in Code. *Evol. Appl.* **2019**, *13*, 12853. [CrossRef] [PubMed]
31. Liu, S.-H.; Edwards, C.; Hoch, P.C.; Raven, P.H.; Barber, J.C. Complete Plastome Sequence of *Ludwigia octovalvis* (Onagraceae), a Globally Distributed Wetland Plant. *Genome Announc.* **2016**, *4*, e01274-16. [CrossRef]
32. Barloy-Hubler, F.; Gac, A.-L.L.; Boury, C.; Guichoux, E.; Barloy, D. Sequencing, de Novo Assembly of *Ludwigia* Plastomes, and Comparative Analysis within the Onagraceae Family. *bioRxiv* **2023**. [CrossRef]
33. Barloy, D.; Portillo-Lemus, L.; Krueger-Hadfield, S.; Huteau, V.; Coriton, O. Genomic relationships among diploid and polyploid species of the genus *Ludwigia* L. section *Jussiaea* using a combination of molecular cytogenetic, morphological, and crossing investigations. *Peer Community J.* **2024**, *4*, e8. [CrossRef]
34. Wang, Y.; Yu, J.; Jiang, M.; Lei, W.; Zhang, X.; Tang, H. Sequencing and Assembly of Polyploid Genomes. *Methods Mol. Biol.* **2023**, *2545*, 429–458. [CrossRef]
35. Jacoby, R.P.; Li, L.; Huang, S.; Pong Lee, C.; Millar, A.H.; Taylor, N.L. Mitochondrial Composition, Function and Stress Response in PlantsF. *J. Integr. Plant Biol.* **2012**, *54*, 887–906. [CrossRef]
36. Mackenzie, S.; McIntosh, L. Higher Plant Mitochondria. *Plant Cell* **1999**, *11*, 571–585. [CrossRef] [PubMed]
37. Fischer, A.; Dotzek, J.; Walther, D.; Greiner, S. Graph-Based Models of the Oenothera Mitochondrial Genome Capture the Enormous Complexity of Higher Plant Mitochondrial DNA Organization. *NAR Genom. Bioinform.* **2022**, *4*, lqac027. [CrossRef]
38. Kozik, A.; Rowan, B.A.; Lavelle, D.; Berke, L.; Schranz, M.E.; Michelmore, R.W.; Christensen, A.C. The Alternative Reality of Plant Mitochondrial DNA: One Ring Does Not Rule Them All. *PLoS Genet.* **2019**, *15*, e1008373. [CrossRef]
39. Liu, D.; Zeng, M.; Wu, Y.; Du, Y.; Liu, J.; Luo, S.; Zeng, Y. Comparative Transcriptomic Analysis Provides Insights into the Molecular Basis Underlying Pre-Harvest Sprouting in Rice. *BMC Genom.* **2022**, *23*, 771. [CrossRef] [PubMed]
40. Varré, J.-S.; D'Agostino, N.; Touzet, P.; Gallina, S.; Tamburino, R.; Cantarella, C.; Ubrig, E.; Cardi, T.; Drouard, L.; Gualberto, J.M.; et al. Complete Sequence, Multichromosomal Architecture and Transcriptome Analysis of the Solanum Tuberosum Mitochondrial Genome. *Int. J. Mol. Sci.* **2019**, *20*, 4788. [CrossRef]
41. Li, J.; Xu, Y.; Shan, Y.; Pei, X.; Yong, S.; Liu, C.; Yu, J. Assembly of the Complete Mitochondrial Genome of an Endemic Plant, Scutellaria Tsinyunensis, Revealed the Existence of Two Conformations Generated by a Repeat-Mediated Recombination. *Planta* **2021**, *254*, 36. [CrossRef] [PubMed]
42. Tsujimura, M.; Kaneko, T.; Sakamoto, T.; Kimura, S.; Shigyo, M.; Yamagishi, H.; Terachi, T. Multichromosomal Structure of the Onion Mitochondrial Genome and a Transcript Analysis. *Mitochondrion* **2019**, *46*, 179–186. [CrossRef] [PubMed]
43. Lloyd Evans, D.; Hlongwane, T.T.; Joshi, S.V.; Riaño Pachón, D.M. The Sugarcane Mitochondrial Genome: Assembly, Phylogenetics and Transcriptomics. *PeerJ* **2019**, *7*, e7558. [CrossRef]
44. Kim, C.-K.; Kim, Y.-K. The Multipartite Mitochondrial Genome of *Fallopia multiflora* (Caryophyllales: Polygonaceae). *Mitochondrial DNA B Resour.* **2018**, *3*, 155–156. [CrossRef]
45. Yang, H.; Chen, H.; Ni, Y.; Li, J.; Cai, Y.; Wang, J.; Liu, C. Mitochondrial Genome Sequence of *Salvia officinalis* (Lamiales: Lamiaceae) Suggests Diverse Genome Structures in Cogeneric Species and Finds the Stop Gain of Genes through RNA Editing Events. *Int. J. Mol. Sci.* **2023**, *24*, 5372. [CrossRef]
46. Guo, Y.; Li, Z.; Jin, S.; Chen, S.; Li, F.; Wu, H. Assembly and Comparative Analysis of the Complete Mitochondrial Genome of Two Species of Calla Lilies (Zantedeschia, Araceae). *Int. J. Mol. Sci.* **2023**, *24*, 9566. [CrossRef]
47. Kang, S.-H.; Oh, J.-H.; Kim, H.-J.; Kim, C.-K. The Multipartite Mitochondrial Genome of *Cynanchum wilfordii* (Gentianales: Apocynaceae). *Mitochondrial DNA B Resour.* **2017**, *2*, 720–721. [CrossRef]

48. Chang, S.; Yang, T.; Du, T.; Huang, Y.; Chen, J.; Yan, J.; He, J.; Guan, R. Mitochondrial Genome Sequencing Helps Show the Evolutionary Mechanism of Mitochondrial Genome Formation in Brassica. *BMC Genom.* **2011**, *12*, 497. [CrossRef]
49. Alverson, A.J.; Rice, D.W.; Dickinson, S.; Barry, K.; Palmer, J.D. Origins and Recombination of the Bacterial-Sized Multichromosomal Mitochondrial Genome of Cucumber. *Plant Cell* **2011**, *23*, 2499–2513. [CrossRef]
50. Rice, D.W.; Alverson, A.J.; Richardson, A.O.; Young, G.J.; Sanchez-Puerta, M.V.; Munzinger, J.; Barry, K.; Boore, J.L.; Zhang, Y.; dePamphilis, C.W.; et al. Horizontal Transfer of Entire Genomes via Mitochondrial Fusion in the Angiosperm Amborella. *Science* **2013**, *342*, 1468–1473. [CrossRef]
51. Yang, J.-X.; Dierckxsens, N.; Bai, M.-Z.; Guo, Y.-Y. Multichromosomal Mitochondrial Genome of Paphiopedilum Micranthum: Compact and Fragmented Genome, and Rampant Intracellular Gene Transfer. *Int. J. Mol. Sci.* **2023**, *24*, 3976. [CrossRef]
52. Sanchez-Puerta, M.V.; García, L.E.; Wohlfeiler, J.; Ceriotti, L.F. Unparalleled Replacement of Native Mitochondrial Genes by Foreign Homologs in a Holoparasitic Plant. *New Phytol.* **2017**, *214*, 376–387. [CrossRef]
53. Wu, Z.; Cuthbert, J.M.; Taylor, D.R.; Sloan, D.B. The Massive Mitochondrial Genome of the Angiosperm Silene Noctiflora Is Evolving by Gain or Loss of Entire Chromosomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10185–10191. [CrossRef]
54. Sloan, D.B.; Alverson, A.J.; Wu, M.; Palmer, J.D.; Taylor, D.R. Recent Acceleration of Plastid Sequence and Structural Evolution Coincides with Extreme Mitochondrial Divergence in the Angiosperm Genus Silene. *Genome Biol. Evol.* **2012**, *4*, 294–306. [CrossRef]
55. Skippington, E.; Barkman, T.J.; Rice, D.W.; Palmer, J.D. Miniaturized Mitogenome of the Parasitic Plant Viscum Scurruloideum Is Extremely Divergent and Dynamic and Has Lost All Nad Genes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3515–E3524. [CrossRef] [PubMed]
56. Putintseva, Y.A.; Bondar, E.I.; Simonov, E.P.; Sharov, V.V.; Oreshkova, N.V.; Kuzmin, D.A.; Konstantinov, Y.M.; Shmakov, V.N.; Belkov, V.I.; Sadovsky, M.G.; et al. *Siberian larch* (Larix sibirica Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. *BMC Genom.* **2020**, *21*, 654. [CrossRef] [PubMed]
57. Mower, J.P.; Sloan, D.B.; Alverson, A.J. Plant Mitochondrial Genome Diversity: The Genomics Revolution. In *Plant Genome*; Wendel, J.F., Greilhuber, J., Dolezel, J., Leitch, I.J., Eds.; Springer: Vienna, Austria, 2012; Volume 1, pp. 123–144, ISBN 978-3-7091-1129-1.
58. Wynn, E.L.; Christensen, A.C. Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution. *G3 Genes Genomes Genet.* **2019**, *9*, 549–559. [CrossRef] [PubMed]
59. Bruenn, J.A.; Warner, B.E.; Yerramsetty, P. Widespread Mitovirus Sequences in Plant Genomes. *PeerJ* **2015**, *3*, e876. [CrossRef] [PubMed]
60. Choi, I.-S.; Wojciechowski, M.F.; Ruhlman, T.A.; Jansen, R.K. In and out: Evolution of Viral Sequences in the Mitochondrial Genomes of Legumes (Fabaceae). *Mol. Phylogenet. Evol.* **2021**, *163*, 107236. [CrossRef]
61. Knoop, V.; Unseld, M.; Marienfeld, J.; Brandt, P.; Sünkel, S.; Ullrich, H.; Brennicke, A. Copia-, Gypsy- and LINE-like Retrotransposon Fragments in the Mitochondrial Genome of Arabidopsis Thaliana. *Genetics* **1996**, *142*, 579–585. [CrossRef] [PubMed]
62. Yu, X.; Jiang, W.; Tan, W.; Zhang, X.; Tian, X. Deciphering the Organelle Genomes and Transcriptomes of a Common Ornamental Plant Ligustrum Quihoui Reveals Multiple Fragments of Transposable Elements in the Mitogenome. *Int. J. Biol. Macromol.* **2020**, *165*, 1988–1999. [CrossRef]
63. de Abreu, V.A.C.; Moysés Alves, R.; Silva, S.R.; Ferro, J.A.; Domingues, D.S.; Miranda, V.F.O.; Varani, A.M. Comparative Analyses of Theobroma Cacao and T. Grandiflorum Mitogenomes Reveal Conserved Gene Content Embedded within Complex and Plastic Structures. *Gene* **2023**, *849*, 146904. [CrossRef]
64. Sloan, D.B. One Ring to Rule Them All? Genome Sequencing Provides New Insights into the 'Master Circle' Model of Plant Mitochondrial DNA Structure. *New Phytol.* **2013**, *200*, 978–985. [CrossRef]
65. Arrieta-Montiel, M.P.; Mackenzie, S.A. Plant Mitochondrial Genomes and Recombination. In *Plant Mitochondria*; Kempken, F., Ed.; Springer: New York, NY, USA, 2011; pp. 65–82. ISBN 978-0-387-89781-3.
66. Zhang, X.; Shan, Y.; Li, J.; Qin, Q.; Yu, J.; Deng, H. Assembly of the Complete Mitochondrial Genome of Pereskia Aculeata Revealed That Two Pairs of Repetitive Elements Mediated the Recombination of the Genome. *Int. J. Mol. Sci.* **2023**, *24*, 8366. [CrossRef]
67. Yang, Z.; Ni, Y.; Lin, Z.; Yang, L.; Chen, G.; Nijiati, N.; Hu, Y.; Chen, X. De Novo Assembly of the Complete Mitochondrial Genome of Sweet Potato (*Ipomoea batatas* [L.] Lam) Revealed the Existence of Homologous Conformations Generated by the Repeat-Mediated Recombination. *BMC Plant Biol.* **2022**, *22*, 285. [CrossRef]
68. Sullivan, A.R.; Eldfjell, Y.; Schiffthaler, B.; Delhomme, N.; Asp, T.; Hebelstrup, K.H.; Keech, O.; Öberg, L.; Møller, I.M.; Arvestad, L.; et al. The Mitogenome of Norway Spruce and a Reappraisal of Mitochondrial Recombination in Plants. *Genome Biol. Evol.* **2020**, *12*, 3586–3598. [CrossRef]
69. Lonsdale, D.M.; Brears, T.; Hodge, T.P.; Melville, S.E.; Rottmann, W.H.; Leaver, C.J.; Lonsdale, D.M. The Plant Mitochondrial Genome: Homologous Recombination as a Mechanism for Generating Heterogeneity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1997**, *319*, 149–163. [CrossRef]
70. Gualberto, J.M.; Newton, K.J. Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. *Annu. Rev. Plant Biol.* **2017**, *68*, 225–252. [CrossRef]
71. Kubo, T.; Newton, K.J. Angiosperm Mitochondrial Genomes and Mutations. *Mitochondrion* **2008**, *8*, 5–14. [CrossRef]

72. Marienfeld, J.; Unseld, M.; Brennicke, A. The Mitochondrial Genome of Arabidopsis Is Composed of Both Native and Immigrant Information. *Trends Plant Sci.* **1999**, *4*, 495–502. [CrossRef]
73. Hao, W.; Palmer, J.D. Fine-Scale Mergers of Chloroplast and Mitochondrial Genes Create Functional, Transcompartmentally Chimeric Mitochondrial Genes. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 16728–16733. [CrossRef]
74. Marechal-Drouard, L.; Weil, J.H.; Dietrich, A. Transfer RNAs and Transfer RNA Genes in Plants. *Annu. Rev. Plant Biol.* **1993**, *44*, 13–32. [CrossRef]
75. Nakazono, M.; Nishiwaki, S.; Tsutsumi, N.; Hirai, A. A Chloroplast-Derived Sequence Is Utilized as a Source of Promoter Sequences for the Gene for Subunit 9 of NADH Dehydrogenase (Nad9) in Rice Mitochondria. *Mol. Gen. Genet.* **1996**, *252*, 371–378. [CrossRef]
76. Wang, D.; Rousseau-Gueutin, M.; Timmis, J.N. Plastid Sequences Contribute to Some Plant Mitochondrial Genes. *Mol. Biol. Evol.* **2012**, *29*, 1707–1711. [CrossRef] [PubMed]
77. Fox, T.D.; Leaver, C.J. The Zea Mays Mitochondrial Gene Coding Cytochrome Oxidase Subunit II Has an Intervening Sequence and Does Not Contain TGA Codons. *Cell* **1981**, *26*, 315–323. [CrossRef] [PubMed]
78. Chao, S.; Sederoff, R.R.; Levings, C.S. Partial Sequence Analysis of the 5S to 18S rRNA Gene Region of the Maize Mitochondrial Genome. *Plant Physiol.* **1983**, *71*, 190–193. [CrossRef]
79. Sloan, D.B.; Alverson, A.J.; Chuckalovcak, J.P.; Wu, M.; McCauley, D.E.; Palmer, J.D.; Taylor, D.R. Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. *PLoS Biol.* **2012**, *10*, e1001241. [CrossRef]
80. Knoop, V. The Mitochondrial DNA of Land Plants: Peculiarities in Phylogenetic Perspective. *Curr. Genet.* **2004**, *46*, 123–139. [CrossRef] [PubMed]
81. Unseld, M.; Marienfeld, J.R.; Brandt, P.; Brennicke, A. The Mitochondrial Genome of Arabidopsis Thaliana Contains 57 Genes in 366,924 Nucleotides. *Nat. Genet.* **1997**, *15*, 57–61. [CrossRef]
82. Adams, K.L.; Palmer, J.D. Evolution of Mitochondrial Gene Content: Gene Loss and Transfer to the Nucleus. *Mol. Phylogenet. Evol.* **2003**, *29*, 380–395. [CrossRef] [PubMed]
83. Ichinose, M.; Sugita, M. RNA Editing and Its Molecular Mechanism in Plant Organelles. *Genes* **2016**, *8*, 5. [CrossRef] [PubMed]
84. Small, I.D.; Schallenberg-Rüdinger, M.; Takenaka, M.; Mireau, H.; Ostersetzer-Biran, O. Plant Organellar RNA Editing: What 30 Years of Research Has Revealed. *Plant J.* **2020**, *101*, 1040–1056. [CrossRef]
85. Gualberto, J.M.; Mileshina, D.; Wallet, C.; Niazi, A.K.; Weber-Lotfi, F.; Dietrich, A. The Plant Mitochondrial Genome: Dynamics and Maintenance. *Biochimie* **2014**, *100*, 107–120. [CrossRef]
86. Wang, S.; Song, Q.; Li, S.; Hu, Z.; Dong, G.; Song, C.; Huang, H.; Liu, Y. Assembly of a Complete Mitogenome of Chrysanthemum Nankingense Using Oxford Nanopore Long Reads and the Diversity and Evolution of *Asteraceae mitogenomes*. *Genes* **2018**, *9*, 547. [CrossRef]
87. Kovar, L.; Nageswara-Rao, M.; Ortega-Rodriguez, S.; Dugas, D.V.; Straub, S.; Cronn, R.; Strickler, S.R.; Hughes, C.E.; Hanley, K.A.; Rodriguez, D.N.; et al. PacBio-Based Mitochondrial Genome Assembly of *Leucaena trichandra* (Leguminosae) and an Intrageneric Assessment of Mitochondrial RNA Editing. *Genome Biol. Evol.* **2018**, *10*, 2501–2517. [CrossRef] [PubMed]
88. The Angiosperm Phylogeny Group; Chase, M.W.; Christenhusz, M.J.M.; Fay, M.F.; Byng, J.W.; Judd, W.S.; Soltis, D.E.; Mabberley, D.J.; Sennikov, A.N.; Soltis, P.S.; et al. An Update of the Angiosperm Phylogeny Group Classification for the Orders and Families of Flowering Plants: APG IV. *Bot. J. Linn. Soc.* **2016**, *181*, 12385. [CrossRef]
89. Lu, G.; Zhang, K.; Que, Y.; Li, Y. Assembly and Analysis of the First Complete Mitochondrial Genome of Punica Granatum and the Gene Transfer from Chloroplast Genome. *Front. Plant Sci.* **2023**, *14*, 1132551. [CrossRef] [PubMed]
90. Yu, X.; Ma, Z.; Liu, S.; Duan, Z. Analysis of the *Rhodomyrtus tomentosa* Mitochondrial Genome: Insights into Repeat-Mediated Recombination and Intra-Cellular DNA Transfer. *Gene* **2024**, *909*, 148288. [CrossRef] [PubMed]
91. Pinard, D.; Myburg, A.A.; Mizrachi, E. The Plastid and Mitochondrial Genomes of Eucalyptus Grandis. *BMC Genom.* **2019**, *20*, 132. [CrossRef] [PubMed]
92. Fukasawa, Y.; Driguez, P.; Bougouffa, S.; Carty, K.; Putra, A.; Cheung, M.-S.; Ermini, L. Plasticity of Repetitive Sequences Demonstrated by the Complete Mitochondrial Genome of Eucalyptus Camaldulensis. *Front. Plant Sci.* **2024**, *15*, 1339594. [CrossRef]
93. Wu, H.-H.; Zhao, X.-H.; Zong, X.-Y.; Ding, R.; Chen, X.-H. Complete Mitochondrial Genome of *Medinilla magnifica* (Myrtales, Melastomataceae). *Mitochondrial DNA Part B* **2020**, *5*, 1716–1717. [CrossRef]
94. Zhou, Y.; Zheng, R.; Peng, Y.; Chen, J.; Zhu, X.; Xie, K.; Ahmad, S.; Chen, J.; Wang, F.; Shen, M.; et al. The First Mitochondrial Genome of *Melastoma dodecandrum* Resolved Structure Evolution in Melastomataceae and Micro Inversions from Inner Horizontal Gene Transfer. *Ind. Crops Prod.* **2023**, *205*, 117390. [CrossRef]
95. Wu, Z.-Q.; Liao, X.-Z.; Zhang, X.-N.; Tembrock, L.R.; Broz, A. Genomic Architectural Variation of Plant Mitochondria—A Review of Multichromosomal Structuring. *J. Syst. Evol.* **2020**, *60*, 160–168. [CrossRef]
96. Cole, L.W.; Guo, W.; Mower, J.P.; Palmer, J.D. High and Variable Rates of Repeat-Mediated Mitochondrial Genome Rearrangement in a Genus of Plants. *Mol. Biol. Evol.* **2018**, *35*, 2773–2785. [CrossRef]
97. Bi, C.; Qu, Y.; Hou, J.; Wu, K.; Ye, N.; Yin, T. Deciphering the Multi-Chromosomal Mitochondrial Genome of *Populus simonii*. *Front. Plant Sci.* **2022**, *13*, 914635. [CrossRef] [PubMed]

98. Yang, H.; Chen, H.; Ni, Y.; Li, J.; Cai, Y.; Ma, B.; Yu, J.; Wang, J.; Liu, C. De Novo Hybrid Assembly of the *Salvia miltiorrhiza* Mitochondrial Genome Provides the First Evidence of the Multi-Chromosomal Mitochondrial DNA Structure of Salvia Species. *Int. J. Mol. Sci.* **2022**, *23*, 14267. [CrossRef] [PubMed]
99. Sloan, D.B.; Alverson, A.J.; Štorchová, H.; Palmer, J.D.; Taylor, D.R. Extensive Loss of Translational Genes in the Structurally Dynamic Mitochondrial Genome of the Angiosperm Silene Latifolia. *BMC Evol. Biol.* **2010**, *10*, 274. [CrossRef] [PubMed]
100. You, C.; Cui, T.; Zhang, C.; Zang, S.; Su, Y.; Que, Y. Assembly of the Complete Mitochondrial Genome of Gelsemium Elegans Revealed the Existence of Homologous Conformations Generated by a Repeat Mediated Recombination. *Int. J. Mol. Sci.* **2023**, *24*, 527. [CrossRef] [PubMed]
101. Wang, X.-C.; Chen, H.; Yang, D.; Liu, C. Diversity of Mitochondrial Plastid DNAs (MTPTs) in Seed Plants. *Mitochondrial DNA Part A* **2018**, *29*, 635–642. [CrossRef] [PubMed]
102. Tang, J.; Luo, Z.; Zhang, J.; Chen, L.; Li, L. Multi-Chromosomal Mitochondrial Genome of Medicinal Plant *Acorus Tatarinowii* (Acoraceae): Firstly Reported from Acorales Order. *Gene* **2024**, *892*, 147847. [CrossRef]
103. Yang, L.; Liu, J.; Guo, W.; Zheng, Z.; Xu, Y.; Xia, H.; Xiao, T. Insights into the Multi-Chromosomal Mitochondrial Genome Structure of the Xero-Halophytic Plant *Haloxylon ammodendron* (C.A.Mey.) Bunge Ex Fenzl. *BMC Genom.* **2024**, *25*, 123. [CrossRef]
104. Hao, W.; Liu, G.; Wang, W.; Shen, W.; Zhao, Y.; Sun, J.; Yang, Q.; Zhang, Y.; Fan, W.; Pei, S.; et al. RNA Editing and Its Roles in Plant Organelles. *Front. Genet.* **2021**, *12*, 757109. [CrossRef]
105. Yan, J.; Zhang, Q.; Yin, P. RNA Editing Machinery in Plant Organelles. *Sci. China Life Sci.* **2018**, *61*, 162–169. [CrossRef]
106. Gommans, W.M.; Mullen, S.P.; Maas, S. RNA Editing: A Driving Force for Adaptive Evolution? *BioEssays* **2009**, *31*, 1137–1145. [CrossRef]
107. Lu, G.; Li, Q. Complete Mitochondrial Genome of Syzygium Samarangense Reveals Genomic Recombination, Gene Transfer, and RNA Editing Events. *Front. Plant Sci.* **2024**, *14*, 1301164. [CrossRef]
108. Xiong, J.; Tao, T.; Luo, Z.; Yan, S.; Liu, Y.; Yu, X.; Liu, G.; Xia, H.; Luo, L. RNA Editing Responses to Oxidative Stress between a Wild Abortive Type Male-Sterile Line and Its Maintainer Line. *Front. Plant Sci.* **2017**, *8*, 2023. [CrossRef]
109. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
110. Bushnell, B.; Rood, J.; Singer, E. BBMerge—Accurate Paired Shotgun Read Merging via Overlap. *PLoS ONE* **2017**, *12*, e0185056. [CrossRef]
111. Prjibelski, A.; Antipov, D.; Meleshko, D.; Lapidus, A.; Korobeynikov, A. Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinform.* **2020**, *70*, e102. [CrossRef]
112. Holley, G.; Beyter, D.; Ingimundardottir, H.; Møller, P.L.; Kristmundsdottir, S.; Eggertsson, H.P.; Halldorsson, B.V. Ratatosk: Hybrid Error Correction of Long Reads Enables Accurate Variant Calling and Assembly. *Genome Biol.* **2021**, *22*, 28. [CrossRef]
113. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of Long, Error-Prone Reads Using Repeat Graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [CrossRef]
114. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph. *Bioinformatics* **2015**, *31*, 1674–1676. [CrossRef]
115. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]
116. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq—Versatile and Accurate Annotation of Organelle Genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [CrossRef]
117. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef] [PubMed]
118. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) Version 1.3.1: Expanded Toolkit for the Graphical Visualization of Organellar Genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [CrossRef] [PubMed]
119. Edera, A.A.; Small, I.; Milone, D.H.; Sanchez-Puerta, M.V. Deepred-Mt: Deep Representation Learning for Predicting C-to-U RNA Editing in Plant Mitochondria. *Comput. Biol. Med.* **2021**, *136*, 104682. [CrossRef] [PubMed]
120. Novák, P.; Neumann, P.; Pech, J.; Steinhaisl, J.; Macas, J. RepeatExplorer: A Galaxy-Based Web Server for Genome-Wide Characterization of Eukaryotic Repetitive Elements from next-Generation Sequence Reads. *Bioinformatics* **2013**, *29*, 792–793. [CrossRef]
121. Kurtz, S.; Schleiermacher, C. REPuter: Fast Computation of Maximal Repeats in Complete Genomes. *Bioinformatics* **1999**, *15*, 426–427. [CrossRef] [PubMed]
122. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-Web: A Web Server for Microsatellite Prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [CrossRef]
123. Benson, G. Tandem Repeats Finder: A Program to Analyze DNA Sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [CrossRef]
124. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **2010**, *5*, e11147. [CrossRef]
125. Yu, Y.; Ouyang, Y.; Yao, W. shinyCircos: An R/Shiny Application for Interactive Creation of Circos Plot. *Bioinformatics* **2018**, *34*, 1229–1231. [CrossRef]
126. Harris, R.S. *Improved Pairwise Alignment of Genomic DNA*; The Pennsylvania State University: University Park, PA, USA, 2007.

127. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]
128. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef]
129. Stothard, P. The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *BioTechniques* **2000**, *28*, 1102–1104. [CrossRef] [PubMed]
130. Chou, K.-C.; Shen, H.-B. Plant-mPLoc: A Top-Down Strategy to Augment the Power for Predicting Plant Protein Subcellular Localization. *PLoS ONE* **2010**, *5*, e11335. [CrossRef]
131. Kriebel, R.; Khabbazian, M.; Sytsma, K.J. A Continuous Morphological Approach to Study the Evolution of Pollen in a Phylogenetic Context: An Example with the Order Myrtales. *PLoS ONE* **2017**, *12*, e0187228. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

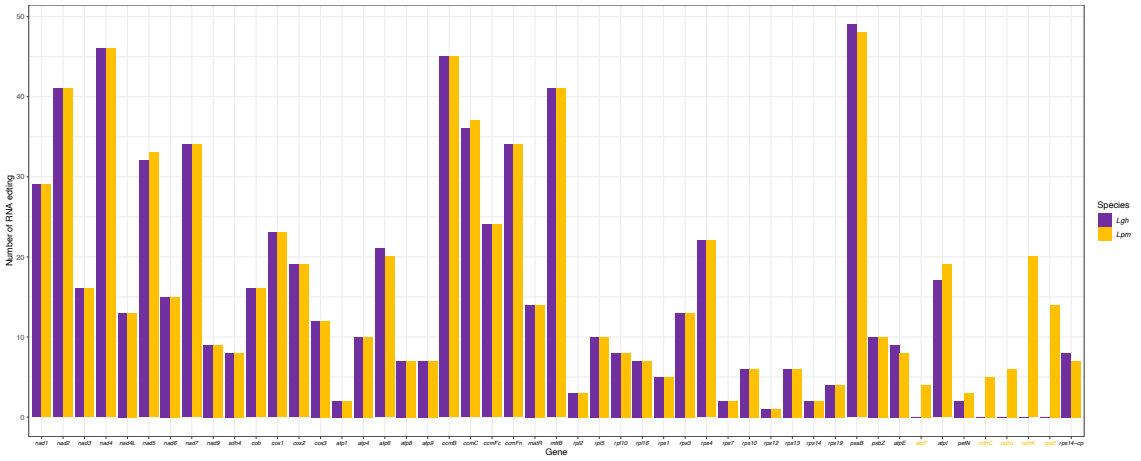
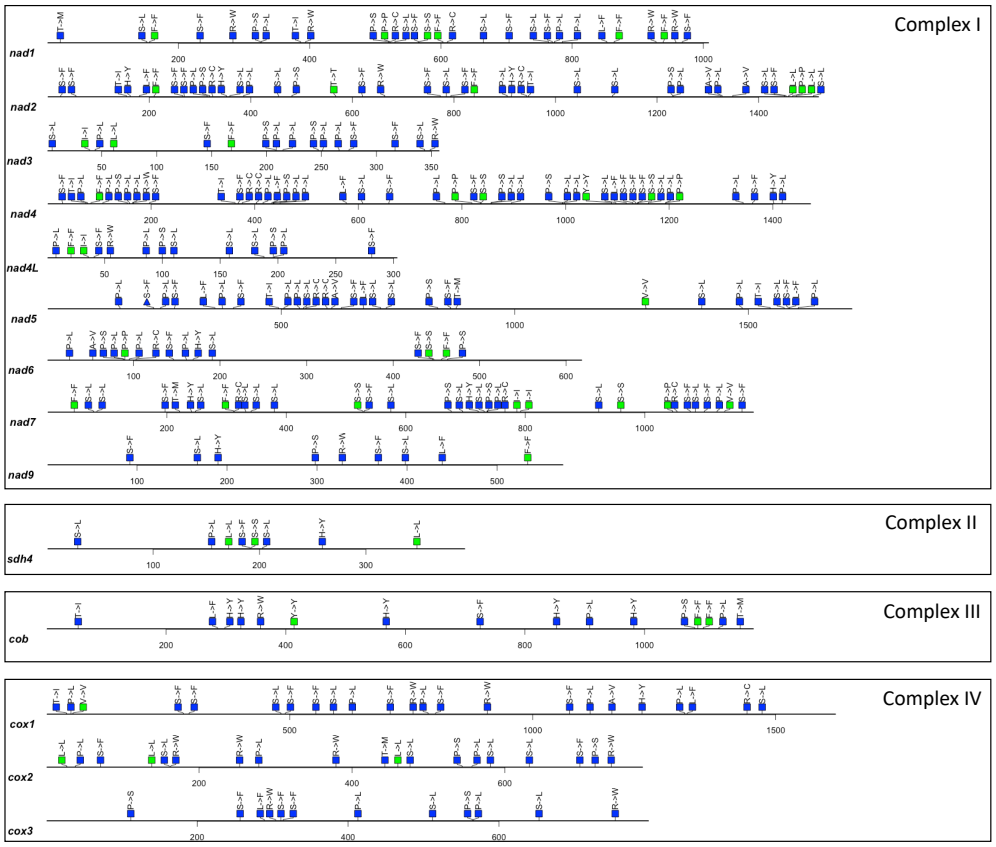
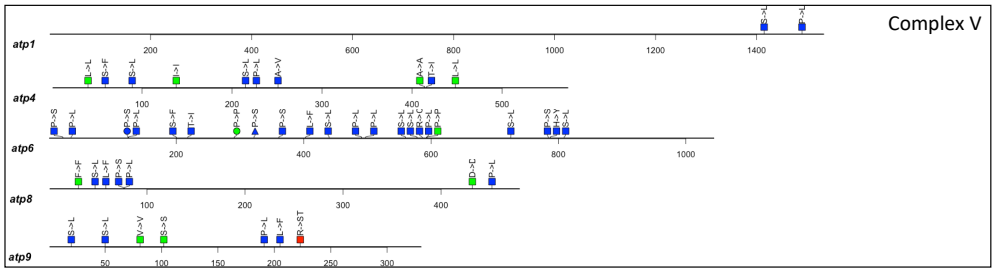
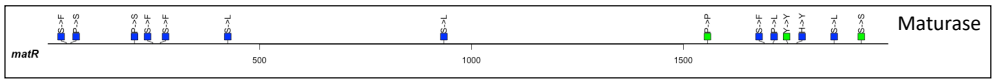
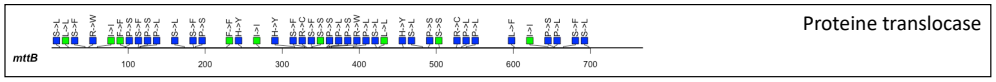


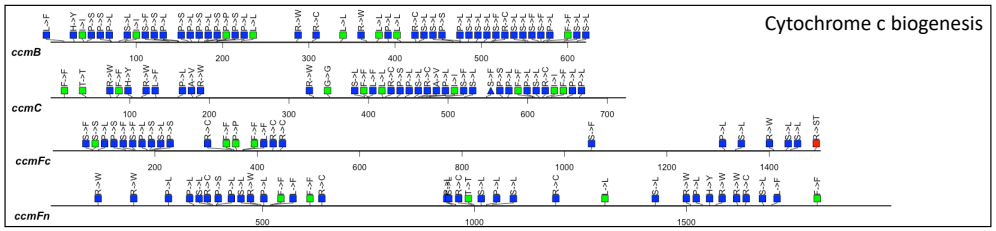
Figure S1. Number of RNA editing per gene, identified by Deepred-Mt, in *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh* in purple) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm* in yellow) mitogenomes.

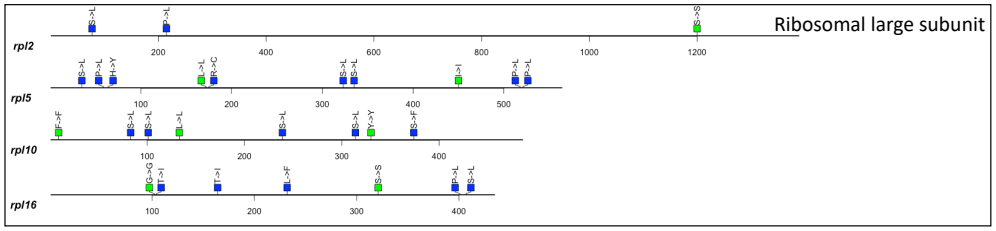


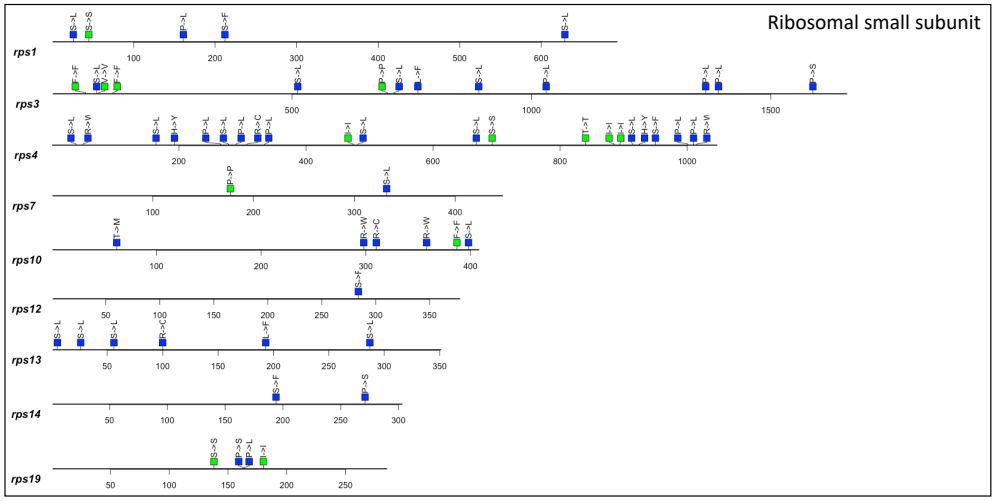




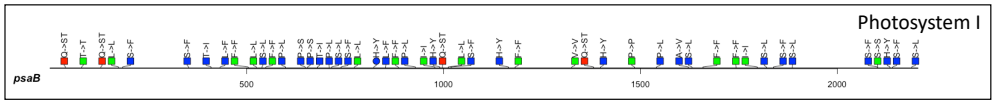




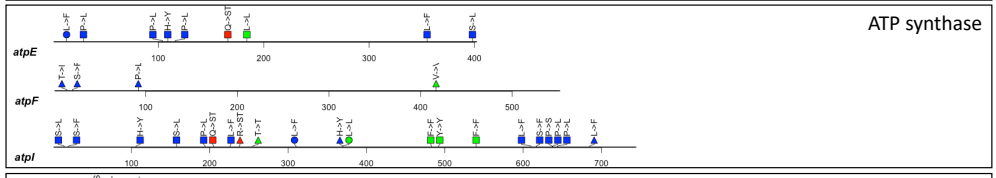


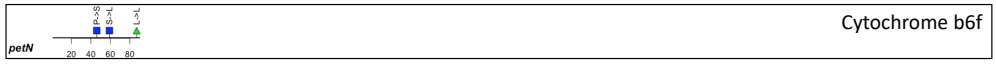


Chloroplast originated genes









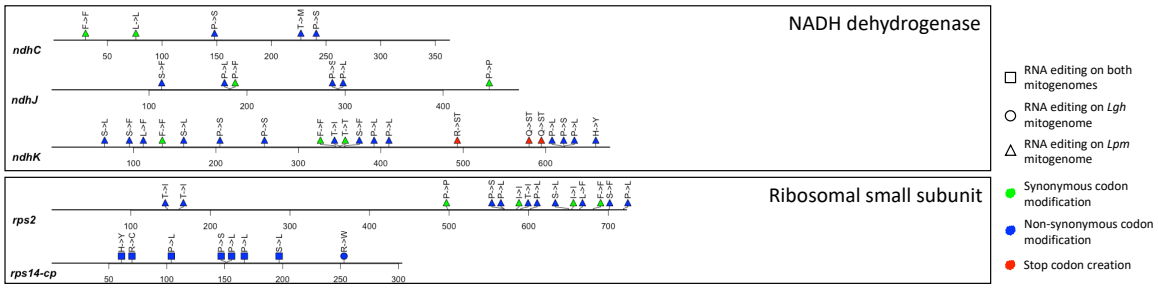


Figure S2. Representation of RNA editing sites in *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) mitogenomes protein-coding genes.

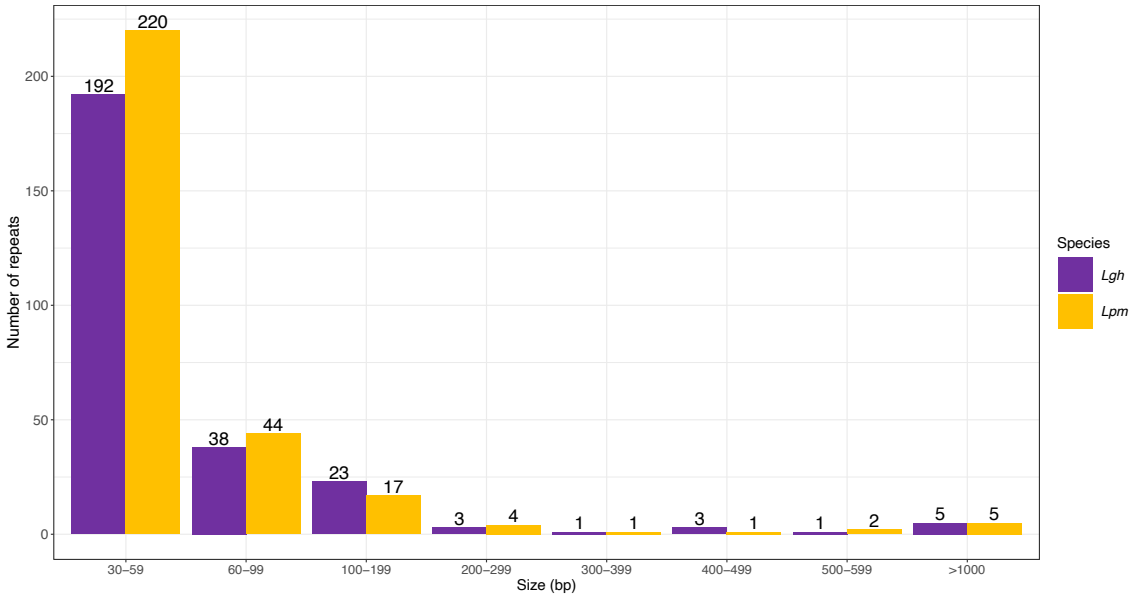


Figure S3. Number of non-tandem repeats based on their size in *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) mitogenomes.

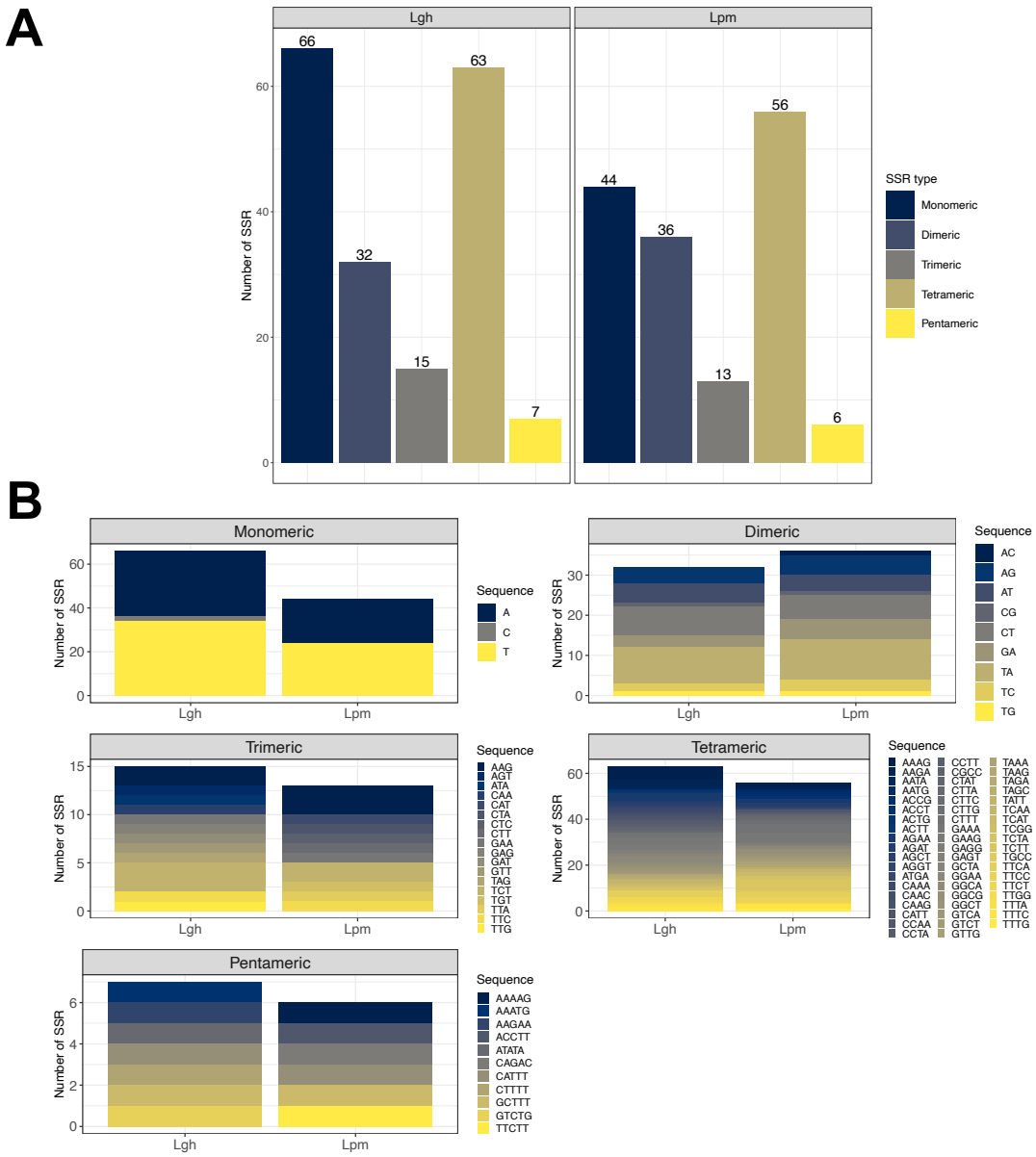


Figure S4. Simple Sequence Repeats (SSR) in *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) mitogenomes. A: Number of each SSR type. B: Detailed compositions of each SSR type.

Table S1: Comparison of Flye, Megahit and SPAdes assembly performances using *Ludwigia grandiflora* subsp. *hexapetala* mitogenomes M1 and M2 as the reference, evaluated with QUAST. All statistics are derived from contigs with a size ≥ 500 bp, unless otherwise specified (e.g., "# contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" encompass all contigs).

Alignment-based statistics	**Flye**	**Megahit**	**SPAdes**
Genome fraction (%)	92.42	89.066	94.922
Duplication ratio	1.003	1.007	1.038
Largest alignment	105 008	38 614	191 601
Total aligned length	663 802	642 161	704 311
NGA50	53 304	15 088	80 045
LGA50	5	15	3
Misassemblies			
# misassemblies	0	0	3
Misassembled contigs length	0	0	439 030
Per base quality			
# mismatches per 100 kbp	70.76	46.22	146.93
# mismatches	473	296	1041
# indels per 100 kbp	12.91	5.44	34.29
# indels	87	35	244
# indels (≤ 5 bp)	78	32	208
# indels (> 5 bp)	9	3	36
Indels length	226	91	1012
Statistics without reference			
# contigs (≥ 0 bp)	39	179	35
# contigs (≥ 500 bp)	37	70	20
# contigs (≥ 1000 bp)	22	67	13
# contigs (≥ 5000 bp)	16	35	8
# contigs (≥ 10000 bp)	13	18	7
# contigs (≥ 25000 bp)	11	7	5
# contigs (≥ 50000 bp)	6	0	5
Largest contig	105 008	43 938	191 601
Total length (≥ 0 bp)	661 029	681 941	812 440
Total length (≥ 500 bp)	660 613	648 165	806 438
Total length (≥ 1000 bp)	650 062	646 177	801 733
Total length (≥ 5000 bp)	641 311	550 392	789 051
Total length (≥ 10000 bp)	619 455	414 136	780 610
Total length (≥ 25000 bp)	596 263	214 902	752 738
Total length (≥ 50000 bp)	420 228	0	752 738
GC (%)	45.1	45.24	43.79

Table S2: Comparisons of locally collinear blocks (LCBs) identified by Progressive-Mauve between *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) mitogenomes.

LCB name	**Length in *Lpm* (bp)**	**Length in *Lgh* (bp)**	**Identity %**
	Molecule M2		
LCB0	167,000	166,796	99.5
	Molecule M1		
LCB1	97,795	97,852	99.7
LCB2	1,046	1,046	99.1
LCB3	48,239	48,292	99
LCB4	50,794	50,805	99.6
LCB5	152,180	152,309	99.7
LCB6	12,325	12,514	86.6
LCB7	83,140	83,214	99.8
LCB8	45,749	45,778	99.9

Table S3: Chloroplast insertions in *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) mitogenomes. Genes are color-coded as follows: red for pseudogenes, cyan for pseudogenes due to RNA editing, orange for fragment genes, black for complete protein coding genes, purple for tRNA genes, and magenta for rRNA genes. Locally Collinear Blocks (LCBs) correspond to conserved regions between *Lgh* and *Lpm* mitogenomes, identified by Progressive-Mauve. The term "interLCB" indicates that the sequence is situated between two LCBs.

LCB	Coordinates	Length (bp)	Identity %	Genes
		Ludwigia grandiflora* subsp. *hexapetala		
		Molecule M2		
LCB0	17,289-17,785	497	100	*rps7**
LCB0	17,776-18,317	542	95.8	Δ*rps12*
LCB0	26,101-26,343	243	88.5	Δ*rps12*
LCB0	45,455-45,982	528	92.1	Δ*ndhB*
LCB0	158,507-166,557	8051	98.4	*atpA**, *atpF**, *atpH**, *atpI**, *rps2**, Δ*rpoC2*
		Molecule M1		
LCB1	15-379	365	97.1	Δ*paf1*
LCB3	37,376-37,737	362	67.8	Δ*petB*
LCB3	57,951-62,745	4795	99.9	Δ*psaA*, *psaB**, ***rps14***, *trnM-CAU*, *trnG-GCC*, ***psbZ***, *trnS-UGA*
LCB4	103,876-110,040	6165	99.6	*trnV-GAC*, ***rrn16***, *trnI-GAU*, *trnA-UGC*, Δ*rrn23*
LCB4	112,517-113,116	600	100	Δ*ndhA*
LCB4	115,795-115,940	146	94.6	Δ*psaA*
LCB4	116,140-116,292	153	81	Δ*accD*
LCB5	185,591-191,050	5460	89.3	*ndhJ**, *ndhK**, *ndhC**, *trnV-UAC*, *trnM-CAU*, *atpE**, Δ*atpB*
LCB5	206,734-206,941	208	86.5	Δ*atpB*
LCB5	215,323-215,452	130	77.9	Δ*ycf2*
LCB5	215,984-216,324	341	83.8	*trnP-UGG*, *trnW-CCA*
LCB5	245,287-245,855	569	74.7	*trnD-GUC*
LCB5	246,325-246,633	309	93.5	***petN***
LCB5	246,921-248,135	1215	89.7	Δ*psbD*, Δ*psbC*
LCB7	297,606-301,261	3656	100	Δ*rrn23*, ***rrn4.5***, ***rrn5***, *trnR-ACG*, *trnN-GUU*, Δ*ycf1*
LCB7	353,346-353,464	119	84.9	
LCB7	377,423-378,440	1018	73.7	*trnS-GGA*
LCB1	491,174-491,425	252	100	Δ*rpl2*
LCB1	534,557-542,192	7636	92.8	Δ*ycf15*, *ycf2**, *trnM-CAU*
		Ludwigia peploides* subsp. *montevidensis		
		Molecule M2		
LCB0	17,286-17,783	498	97.4	*rps7**
LCB0	17,774-18,315	542	93.9	Δ*rps12*
LCB0	26,097-26,339	243	88.9	Δ*rps12*
LCB0	45,440-45,967	528	100	Δ*ndhB*
LCB0	158,507-166,557	8051	98.4	*atpA**, ***atpF***, *atpH**, *atpI**, ***rps2***, Δ*rpoC2*
		Molecule M1		
LCB1	1-382	382	88.7	Δ*paf1*
LCB3	379,206-379,537	332	68	Δ*petB*
LCB3	354,205-358,879	4675	99.9	Δ*psaA*, *psaB**, ***rps14***, *trnM-CAU*, *trnG-GCC*, ***psbZ***, *trnS-UGA*
LCB4	415,927-422,111	6185	99.6	*trnV-GAC*, ***rrn16***, *trnI-GAU*, *trnA-UGC*, Δ*rrn23*
LCB4	412,854-413,449	596	96	Δ*ndhA*
LCB4	410,033-410,170	138	95.5	Δ*psaA*
LCB4	409,710-409,854	145	79.2	Δ*accD*
LCB5	204,896-210,306	5411	91.9	***ndhJ***, *ndhK**, ***ndhC***, *trnV-UAC*, *trnM-CAU*, *atpE**, Δ*atpB*
LCB5	189,010-189,217	208	86.1	Δ*atpB*
LCB5	180,503-180,632	130	78.7	Δ*ycf2*
LCB5	179,632-179,994	363	82.8	*trnP-UGG*, *trnW-CCA*
LCB5	150,130-150,687	558	79.4	*trnD-GUC*
LCB5	149,349-149,657	309	94.8	***petN***
LCB5	147,831-149,052	1222	93.5	Δ*psbD*, Δ*psbC*
LCB7	95,009-98,538	3530	99.8	Δ*rrn23*, ***rrn4.5***, ***rrn5***, *trnR-ACG*, *trnN-GUU*, Δ*ycf1*
LCB7	42,831-42,965	135	86.7	
LCB7	17,297-18,884	1588	65.1	*trnS-GGA*
LCB1	501,949-502,200	252	96	Δ*rpl2*
LCB1	545,340-553,004	7665	93	Δ*ycf15*, *ycf2**, *trnM-CAU*
interLCB	335,484-335,571	88	85.2	Δ*accD*

Table S4: RNA editing sites in *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) mitogenomes protein-coding genes identified by Deepred-Mt. Yellow and purple lines represent RNA editing unique to *Lpm* and *Lgh*, respectively. Amino acid changes in blue and green indicate non-synonymous and synonymous modifications, respectively. Changes in red signify the appearance of a stop codon.

Gene	Position	Base codon	New codon	Aminoacid change	Probability
nad1	20	ACG	ATG	T->M	0.963
nad1	152	TCG	TTG	S->L	0.76
nad1	156	TTC	TTT	F->F	0.685
nad1	233	TCC	TTC	S->F	0.999
nad1	283	CGG	TGG	R->W	0.999
nad1	325	CCG	TCG	P->S	0.995
nad1	326	CCG	CTG	P->L	1
nad1	386	ACA	ATA	T->I	0.998
nad1	394	CGG	TGG	R->W	0.994
nad1	520	CCC	TCC	P->S	0.997
nad1	522	CCC	CCT	P->P	0.989
nad1	523	CGT	TGT	R->C	0.995
nad1	530	TCG	TTG	S->L	1
nad1	566	TCC	TTC	S->F	0.997
nad1	567	TCC	TCT	S->S	0.996
nad1	603	TTC	TTT	F->F	1
nad1	610	CGT	TGT	R->C	1
nad1	665	TCA	TTA	S->L	0.999
nad1	704	TCT	TTT	S->F	1
nad1	764	TCG	TTG	S->L	0.826
nad1	770	TCT	TTT	S->F	0.937
nad1	773	CCA	CTA	P->L	0.993
nad1	785	CCG	CTG	P->L	0.995
nad1	853	CTC	TTC	L->F	0.99
nad1	864	TTC	TTT	F->F	0.671
nad1	928	CGG	TGG	R->W	0.999
nad1	939	TTC	TTT	F->F	0.558
nad1	958	CGG	TGG	R->W	0.999
nad1	968	TCT	TTT	S->F	0.871
nad2	26	TCC	TTC	S->F	0.996
nad2	56	TCC	TTC	S->F	0.999
nad2	155	ACT	ATT	T->I	0.687
nad2	163	CAT	TAT	H->Y	0.532
nad2	181	CTT	TTT	L->F	0.997
nad2	210	TTC	TTT	F->F	0.653
nad2	266	TCT	TTT	S->F	0.997
nad2	269	TCC	TTC	S->F	0.999
nad2	314	CCA	CTA	P->L	0.995
nad2	319	CCT	TCT	P->S	0.999
nad2	325	CGC	TGC	R->C	0.989
nad2	352	CAT	TAT	H->Y	0.997
nad2	359	TCA	TTA	S->L	0.995
nad2	386	CCT	CTT	P->L	0.999
nad2	455	TCG	TTG	S->L	0.999
nad2	481	CCC	TCC	P->S	0.997
nad2	567	ACC	ACT	T->T	0.697
nad2	620	TCT	TTT	S->F	0.996
nad2	664	CGG	TGG	R->W	0.656
nad2	688	CTA	TTA	L->L	0.511
nad2	773	TCT	TTT	S->F	0.999
nad2	785	TCA	TTA	S->L	0.966
nad2	794	TCT	TTT	S->F	0.998
nad2	840	TTC	TTT	F->F	0.538
nad2	905	CCT	CTT	P->L	0.999
nad2	905	CCT	CTT	P->L	0.997
nad2	913	CAT	TAT	H->Y	0.995
nad2	913	CAT	TAT	H->Y	0.993
nad2	943	CGT	TGT	R->C	1
nad2	947	ACT	ATT	T->I	0.997
nad2	1043	TCA	TTA	S->L	0.996
nad2	1112	TCG	TTG	S->L	0.999

nad2	1231	CCA	TCA	P->S	0.995
nad2	1232	CCA	CTA	P->L	0.683
nad2	1325	GCA	GTA	A->V	0.757
nad2	1328	CCG	CTG	P->L	0.846
nad2	1349	GCG	GTG	A->V	0.999
nad2	1451	TCA	TTA	S->L	0.986
nad2	1454	TCC	TTC	S->F	0.97
nad2	1459	CTA	TTA	L->L	0.962
nad2	1467	CCC	CCT	P->P	0.966
nad2	1474	CTG	TTG	L->L	0.76
nad2	1508	TCA	TTA	S->L	0.996
nad3	5	TCA	TTA	S->L	0.966
nad3	39	ATC	ATT	I->I	0.855
nad3	44	CCG	CTG	P->L	1
nad3	61	CTA	TTA	L->L	0.514
nad3	146	TCC	TTC	S->F	0.999
nad3	168	TTC	TTT	F->F	0.692
nad3	208	CCT	TCT	P->S	0.998
nad3	209	CCT	CTT	P->L	0.996
nad3	215	CCG	CTG	P->L	0.997
nad3	247	CCT	TCT	P->S	0.996
nad3	251	CCC	CTC	P->L	0.981
nad3	266	CCC	CTC	P->L	0.999
nad3	275	TCT	TTT	S->F	0.992
nad3	317	TCT	TTT	S->F	1
nad3	344	TCG	TTG	S->L	0.999
nad3	349	CGG	TGG	R->W	0.999
nad4	29	TCT	TTT	S->F	1
nad4	74	ACT	ATT	T->I	0.997
nad4	77	CCT	CTT	P->L	0.999
nad4	84	TTC	TTT	F->F	0.678
nad4	107	CCG	CTG	P->L	0.999
nad4	154	CCC	TCC	P->S	0.999
nad4	158	CCT	CTT	P->L	0.997
nad4	164	CCT	CTT	P->L	0.994
nad4	166	CGG	TGG	R->W	1
nad4	197	TCT	TTT	S->F	0.998
nad4	362	ACA	ATA	T->I	0.999
nad4	368	TCT	TTT	S->F	0.996
nad4	376	CGT	TGT	R->C	0.995
nad4	403	CGC	TGC	R->C	0.999
nad4	416	CCT	CTT	P->L	0.992
nad4	433	CTT	TTT	L->F	0.999
nad4	436	CCC	TCC	P->S	0.999
nad4	437	CCC	CTC	P->L	0.992
nad4	449	CCA	CTA	P->L	0.993
nad4	577	CTC	TTC	L->F	0.55
nad4	608	TCA	TTA	S->L	0.999
nad4	659	TCT	TTT	S->F	1
nad4	767	CCT	CTT	P->L	0.999
nad4	819	CCC	CCT	P->P	0.601
nad4	836	TCC	TTC	S->F	0.995
nad4	837	TCC	TCT	S->S	0.613
nad4	856	CCA	TCA	P->S	0.979
nad4	857	CCA	CTA	P->L	1
nad4	896	TCA	TTA	S->L	1
nad4	994	CCA	TCA	P->S	0.606
nad4	995	CCA	CTA	P->L	0.999
nad4	998	CCG	CTG	P->L	0.999
nad4	1089	TAC	TAT	Y->Y	0.906
nad4	1097	TCA	TTA	S->L	0.999
nad4	1117	CTC	TTC	L->F	0.997
nad4	1130	TCC	TTC	S->F	0.989
nad4	1136	TCT	TTT	S->F	0.998
nad4	1139	TCC	TTC	S->F	0.995
nad4	1140	TCC	TCT	S->S	0.923
nad4	1160	TCA	TTA	S->L	1
nad4	1193	CCC	CTC	P->L	0.999
nad4	1194	CCC	CCT	P->P	0.634
nad4	1343	CCA	CTA	P->L	0.995
nad4	1361	TCC	TTC	S->F	0.999
nad4	1402	CAC	TAC	H->Y	0.997
nad4	1418	CCG	CTG	P->L	0.998
nad4L	8	CCT	CTT	P->L	0.847
nad4L	21	TTC	TTT	F->F	0.573
nad4L	36	ATC	ATT	I->I	0.557
nad4L	41	TCT	TTT	S->F	0.994
nad4L	55	CGG	TGG	R->W	0.996

nad4L	86	CCT	CTT	P->L	0.996
nad4L	100	CCA	TCA	P->S	0.999
nad4L	110	TCA	TTA	S->L	0.998
nad4L	158	TCG	TTG	S->L	0.996
nad4L	188	TCA	TTA	S->L	1
nad4L	196	CCA	TCA	P->S	0.734
nad4L	197	CCA	CTA	P->L	1
nad4L	281	TCT	TTT	S->F	0.999
nad5	155	CCG	CTG	P->L	1
nad5	227	TCC	TTC	S->F	1
nad5	242	CCG	CTG	P->L	0.999
nad5	242	CCG	CTG	P->L	1
nad5	272	TCC	TTC	S->F	0.999
nad5	358	CTT	TTT	L->F	0.994
nad5	374	CCA	CTA	P->L	0.997
nad5	398	TCT	TTT	S->F	1
nad5	494	ACA	ATA	T->I	1
nad5	506	CCT	CTT	P->L	0.995
nad5	539	CCT	CTT	P->L	1
nad5	548	TCG	TTG	S->L	0.998
nad5	553	CGT	TGT	R->C	0.999
nad5	598	CGT	TGT	R->C	0.999
nad5	608	GCC	GTC	A->V	0.993
nad5	629	TCT	TTT	S->F	0.999
nad5	676	CTT	TTT	L->F	1
nad5	713	TCG	TTG	S->L	0.999
nad5	725	TCA	TTA	S->L	1
nad5	835	CCA	TCA	P->S	1
nad5	863	TCT	TTT	S->F	0.999
nad5	875	ACG	ATG	T->M	1
nad5	1281	GTC	GTT	V->V	0.577
nad5	1400	TCA	TTA	S->L	1
nad5	1487	CCC	CTC	P->L	0.999
nad5	1547	ACC	ATC	T->I	0.998
nad5	1577	TCA	TTA	S->L	0.996
nad5	1586	TCT	TTT	S->F	0.99
nad5	1588	CTT	TTT	L->F	0.651
nad5	1607	CCC	CTC	P->L	0.984
nad5	1925	TCT	TTT	S->F	0.999
nad5	1927	CGT	TGT	R->C	0.999
nad5	1967	TCG	TTG	S->L	0.981
nad5	2009	TCA	TTA	S->L	0.804
nad6	26	CCA	CTA	P->L	0.991
nad6	53	GCA	GTA	A->V	0.999
nad6	88	CCC	TCC	P->S	0.991
nad6	89	CCC	CTC	P->L	0.998
nad6	90	CCC	CCT	P->P	0.557
nad6	95	CCA	CTA	P->L	0.994
nad6	103	CGC	TGC	R->C	0.979
nad6	146	TCC	TTC	S->F	1
nad6	161	CCA	CTA	P->L	0.999
nad6	169	CAT	TAT	H->Y	1
nad6	191	TCA	TTA	S->L	0.999
nad6	446	TCC	TTC	S->F	0.999
nad6	447	TCC	TCT	S->S	0.588
nad6	456	TTC	TTT	F->F	0.999
nad6	463	CCT	TCT	P->S	0.996
nad7	45	TTC	TTT	F->F	0.994
nad7	77	TCA	TTA	S->L	0.991
nad7	83	TCA	TTA	S->L	1
nad7	197	TCT	TTT	S->F	0.994
nad7	221	ACG	ATG	T->M	0.999
nad7	241	CAT	TAT	H->Y	0.998
nad7	248	TCA	TTA	S->L	0.993
nad7	312	TTC	TTT	F->F	0.61
nad7	313	CGT	TGT	R->C	0.997
nad7	332	TCA	TTA	S->L	0.999
nad7	341	TCA	TTA	S->L	0.999
nad7	380	TCA	TTA	S->L	1
nad7	528	TCC	TCT	S->S	0.522
nad7	530	TCC	TTC	S->F	0.998
nad7	575	TCA	TTA	S->L	1
nad7	676	CCA	TCA	P->S	0.987
nad7	695	TCG	TTG	S->L	0.995
nad7	721	CAT	TAT	H->Y	1
nad7	731	TCG	TTG	S->L	1
nad7	736	CCT	TCT	P->S	0.999
nad7	737	CCT	CTT	P->L	0.999

nad7	766	CGC	TGC	R->C	0.998
nad7	786	ATC	ATT	I->I	0.922
nad7	792	ATC	ATT	I->I	0.634
nad7	923	TCA	TTA	S->L	0.998
nad7	960	TCC	TCT	S->S	0.806
nad7	1047	CCC	CCT	P->P	0.966
nad7	1054	CGT	TGT	R->C	0.998
nad7	1076	TCT	TTT	S->F	0.999
nad7	1085	TCA	TTA	S->L	0.996
nad7	1100	TCT	TTT	S->F	0.999
nad7	1121	CCA	CTA	P->L	1
nad7	1134	GTC	GTT	V->V	0.979
nad7	1163	TCT	TTT	S->F	1
nad9	92	TCT	TTT	S->F	0.996
nad9	167	TCG	TTG	S->L	0.998
nad9	190	CAT	TAT	H->Y	1
nad9	298	CCG	TCG	P->S	0.996
nad9	328	CGG	TGG	R->W	1
nad9	368	TCC	TTC	S->F	1
nad9	398	TCA	TTA	S->L	0.998
nad9	439	CTT	TTT	L->F	0.998
nad9	534	TTC	TTT	F->F	0.524
sdh4	29	TCG	TTG	S->L	1
sdh4	155	CCA	CTA	P->L	0.983
sdh4	171	CTC	CTT	L->L	0.912
sdh4	191	TCC	TTC	S->F	0.608
sdh4	192	TCC	TCT	S->S	0.596
sdh4	203	TCA	TTA	S->L	0.935
sdh4	259	CAT	TAT	H->Y	0.9
sdh4	348	CTC	CTT	L->L	0.679
cob	53	ACA	ATA	T->I	0.685
cob	286	CTC	TTC	L->F	0.999
cob	298	CAC	TAC	H->Y	0.999
cob	325	CAT	TAT	H->Y	1
cob	358	CGG	TGG	R->W	1
cob	414	TAC	TAT	Y->Y	0.699
cob	568	CAT	TAT	H->Y	0.997
cob	725	TCT	TTT	S->F	0.999
cob	853	CAT	TAT	H->Y	0.999
cob	908	CCA	CTA	P->L	0.998
cob	982	CAC	TAC	H->Y	0.992
cob	1084	CCT	TCT	P->S	0.998
cob	1089	TTC	TTT	F->F	0.601
cob	1098	TTC	TTT	F->F	0.579
cob	1124	CCG	CTG	P->L	0.998
cob	1160	ACG	ATG	T->M	0.673
cox1	41	ACA	ATA	T->I	0.99
cox1	50	CCG	CTG	P->L	0.992
cox1	54	GTC	GTT	V->V	0.957
cox1	281	TCT	TTT	S->F	0.998
cox1	293	TCT	TTT	S->F	1
cox1	482	TCA	TTA	S->L	0.996
cox1	491	TCT	TTT	S->F	1
cox1	554	TCC	TTC	S->F	1
cox1	590	TCA	TTA	S->L	1
cox1	629	CCA	CTA	P->L	0.999
cox1	707	TCT	TTT	S->F	0.999
cox1	754	CGG	TGG	R->W	0.997
cox1	785	CCC	CTC	P->L	0.998
cox1	800	TCC	TTC	S->F	0.996
cox1	907	CGG	TGG	R->W	0.999
cox1	1076	TCC	TTC	S->F	0.999
cox1	1118	CCG	CTG	P->L	0.996
cox1	1163	GCG	GTG	A->V	0.999
cox1	1225	CAC	TAC	H->Y	1
cox1	1313	CCG	CTG	P->L	0.994
cox1	1318	CTC	TTC	L->F	0.987
cox1	1441	CGT	TGT	R->C	0.999
cox1	1472	TCA	TTA	S->L	0.999
cox2	27	CTC	CTT	L->L	0.618
cox2	38	CCT	CTT	P->L	0.69
cox2	71	TCT	TTT	S->F	1
cox2	138	CTC	CTT	L->L	0.934
cox2	138	CTC	CTT	L->L	0.923
cox2	161	TCA	TTA	S->L	0.993
cox2	161	TCA	TTA	S->L	0.999
cox2	163	CGG	TGG	R->W	1
cox2	253	CGG	TGG	R->W	0.998

cox2	278	CCG	CTG	P->L	0.998
cox2	379	CGG	TGG	R->W	0.998
cox2	443	ACG	ATG	T->M	0.999
cox2	460	CTA	TTA	L->L	0.658
cox2	476	TCA	TTA	S->L	1
cox2	544	CCT	TCT	P->S	0.999
cox2	557	CCT	CTT	P->L	0.996
cox2	581	TCA	TTA	S->L	1
cox2	632	TCG	TTG	S->L	0.999
cox2	698	TCT	TTT	S->F	0.939
cox2	718	CCT	TCT	P->S	0.984
cox2	739	CGG	TGG	R->W	0.952
cox3	112	CCA	TCA	P->S	0.998
cox3	257	TCT	TTT	S->F	0.995
cox3	289	CTT	TTT	L->F	1
cox3	304	CGG	TGG	R->W	0.998
cox3	311	TCT	TTT	S->F	0.947
cox3	314	TCT	TTT	S->F	0.968
cox3	413	CCT	CTT	P->L	0.995
cox3	512	TCA	TTA	S->L	1
cox3	565	CCC	TCC	P->S	0.999
cox3	566	CCC	CTC	P->L	0.995
cox3	653	TCG	TTG	S->L	1
cox3	754	CGG	TGG	R->W	1
atp1	1415	TCA	TTA	S->L	0.994
atp1	1490	CCA	CTA	P->L	0.999
atp4	40	CTA	TTA	L->L	0.913
atp4	59	TCT	TTT	S->F	0.995
atp4	89	TCA	TTA	S->L	0.998
atp4	138	ATC	ATT	I->I	0.755
atp4	215	TCG	TTG	S->L	0.997
atp4	227	CCA	CTA	P->L	0.989
atp4	251	GCG	GTG	A->V	0.562
atp4	414	GCC	GCT	A->A	0.716
atp4	416	ACT	ATT	T->I	0.967
atp4	448	CTA	TTA	L->L	0.576
atp6	22	CCG	TCG	P->S	0.635
atp6	35	CCG	CTG	P->L	0.751
atp6	124	CCA	TCA	P->S	0.544
atp6	125	CCA	CTA	P->L	0.85
atp6	125	CCA	CTA	P->L	0.949
atp6	200	TCT	TTT	S->F	0.609
atp6	218	ACT	ATT	T->I	0.715
atp6	291	CCC	CCT	P->P	0.537
atp6	322	CCT	TCT	P->S	0.637
atp6	361	CCA	TCA	P->S	1
atp6	403	CTT	TTT	L->F	0.623
atp6	440	TCA	TTA	S->L	1
atp6	491	CCG	CTG	P->L	0.999
atp6	497	CCG	CTG	P->L	0.996
atp6	560	TCG	TTG	S->L	0.998
atp6	578	TCG	TTG	S->L	0.998
atp6	586	CGT	TGT	R->C	0.999
atp6	593	CCC	CTC	P->L	0.997
atp6	594	CCC	CCT	P->P	0.968
atp6	725	TCA	TTA	S->L	0.999
atp6	784	CCT	TCT	P->S	0.999
atp6	787	CAT	TAT	H->Y	1
atp6	809	TCA	TTA	S->L	0.999
atp8	30	TTC	TTT	F->F	0.965
atp8	47	TCA	TTA	S->L	0.996
atp8	58	CTC	TTC	L->F	0.998
atp8	76	CCC	TCC	P->S	0.997
atp8	77	CCC	CTC	P->L	0.662
atp8	432	GAC	GAT	D->D	0.583
atp8	452	CCA	CTA	P->L	1
atp9	20	TCA	TTA	S->L	0.999
atp9	50	TCA	TTA	S->L	1
atp9	81	GTC	GTT	V->V	0.993
atp9	102	TCC	TCT	S->S	0.601
atp9	191	CCA	CTA	P->L	0.998
atp9	205	CTT	TTT	L->F	0.987
atp9	223	CGA	TGA	R->STOP	0.997
matR	44	TCC	TTC	S->F	0.886
matR	55	CCC	TCC	P->S	0.689
matR	205	CCA	TCA	P->S	0.994
matR	248	TCC	TTC	S->F	0.89
matR	266	TCT	TTT	S->F	0.77

matR	425	TCG	TTG	S->L	0.989
matR	935	TCA	TTA	S->L	1
matR	1557	CCC	CCT	P->P	0.876
matR	1691	TCC	TTC	S->F	0.996
matR	1712	CCT	CTT	P->L	1
matR	1746	TAC	TAT	Y->Y	0.998
matR	1768	CAC	TAC	H->Y	0.995
matR	1856	TCA	TTA	S->L	0.965
matR	1920	TCC	TCT	S->S	0.881
mttb	11	TCG	TTG	S->L	0.985
mttb	22	CTA	TTA	L->L	0.814
mttb	44	TCC	TTC	S->F	0.609
mttb	49	CGG	TGG	R->W	0.983
mttb	60	ATC	ATT	I->I	0.731
mttb	96	TTC	TTT	F->F	0.74
mttb	97	CCG	TCG	P->S	1
mttb	113	TCT	TTT	S->F	0.982
mttb	115	CCA	TCA	P->S	0.77
mttb	116	CCA	CTA	P->L	0.969
mttb	173	TCA	TTA	S->L	1
mttb	185	TCC	TTC	S->F	0.998
mttb	187	CCG	TCG	P->S	0.999
mttb	228	TTC	TTT	F->F	0.779
mttb	247	CAT	TAT	H->Y	0.999
mttb	270	ATC	ATT	I->I	0.729
mttb	316	CAT	TAT	H->Y	0.994
mttb	329	TCT	TTT	S->F	0.971
mttb	331	CGC	TGC	R->C	0.997
mttb	338	TCC	TTC	S->F	0.896
mttb	339	TCC	TCT	S->S	0.808
mttb	358	CCT	TCT	P->S	0.948
mttb	359	CCT	CTT	P->L	0.998
mttb	361	CCC	TCC	P->S	0.99
mttb	364	CGG	TGG	R->W	0.995
mttb	392	CCA	CTA	P->L	0.999
mttb	422	TCG	TTG	S->L	0.884
mttb	426	CTC	CTT	L->L	0.775
mttb	457	CAT	TAT	H->Y	0.997
mttb	482	TCG	TTG	S->L	0.994
mttb	490	CCA	TCA	P->S	0.981
mttb	504	TCC	TCT	S->S	0.626
mttb	526	CGT	TGT	R->C	1
mttb	533	CCA	CTA	P->L	0.996
mttb	539	CCG	CTG	P->L	0.988
mttb	601	CTC	TTC	L->F	0.886
mttb	645	ATC	ATT	I->I	0.988
mttb	652	CCT	TCT	P->S	0.98
mttb	653	CCT	CTT	P->L	0.687
mttb	665	TCT	TTT	S->F	0.767
mttb	698	TCG	TTG	S->L	0.999
ccmB	16	CTT	TTT	L->F	0.874
ccmB	28	CAT	TAT	H->Y	0.945
ccmB	39	ATC	ATT	I->I	0.665
ccmB	43	CCC	TCC	P->S	0.995
ccmB	70	CCG	TCG	P->S	0.905
ccmB	71	CCG	CTG	P->L	0.922
ccmB	80	TCG	TTG	S->L	0.975
ccmB	87	ATC	ATT	I->I	0.962
ccmB	137	TCC	TTC	S->F	0.997
ccmB	148	CCG	TCG	P->S	0.99
ccmB	149	CCG	CTG	P->L	0.992
ccmB	160	CCT	TCT	P->S	0.997
ccmB	164	CCG	CTG	P->L	0.995
ccmB	172	CCT	TCT	P->S	0.989
ccmB	179	CCT	CTT	P->L	0.988
ccmB	181	CCC	TCC	P->S	0.996
ccmB	183	CCC	CCT	P->P	0.542
ccmB	193	CCT	TCT	P->S	0.954
ccmB	194	CCT	CTT	P->L	0.997
ccmB	222	CTC	CTT	L->L	0.54
ccmB	286	CGG	TGG	R->W	0.999
ccmB	304	CGT	TGT	R->C	0.996
ccmB	337	CTG	TTG	L->L	0.569
ccmB	367	CGG	TGG	R->W	0.982
ccmB	379	CTA	TTA	L->L	0.938
ccmB	392	CCG	CTG	P->L	1
ccmB	406	CTG	TTG	L->L	0.749
ccmB	424	CGT	TGT	R->C	0.994

ccmB	428	TCG	TTG	S->L	0.996
ccmB	467	TCG	TTG	S->L	0.975
ccmB	475	CCA	TCA	P->S	0.997
ccmB	476	CCA	CTA	P->L	0.996
ccmB	485	TCA	TTA	S->L	0.985
ccmB	494	TCA	TTA	S->L	0.993
ccmB	503	TCA	TTA	S->L	0.973
ccmB	512	TCT	TTT	S->F	0.972
ccmB	514	CGT	TGT	R->C	0.997
ccmB	551	TCA	TTA	S->L	0.997
ccmB	554	TCG	TTG	S->L	0.994
ccmB	566	TCC	TTC	S->F	0.997
ccmB	569	TCT	TTT	S->F	0.979
ccmB	572	CCG	CTG	P->L	0.999
ccmB	576	TTC	TTT	F->F	0.566
ccmB	596	TCG	TTG	S->L	0.995
ccmB	611	TCA	TTA	S->L	0.835
ccmC	18	TTC	TTT	F->F	0.857
ccmC	45	ACC	ACT	T->T	0.535
ccmC	76	CGG	TGG	R->W	1
ccmC	84	TTC	TTT	F->F	0.532
ccmC	103	CAT	TAT	H->Y	0.999
ccmC	115	CGG	TGG	R->W	0.991
ccmC	133	CTT	TTT	L->F	0.987
ccmC	161	CCG	CTG	P->L	1
ccmC	179	GCG	GTG	A->V	0.999
ccmC	184	CGG	TGG	R->W	0.994
ccmC	331	CGG	TGG	R->W	0.999
ccmC	351	GGC	GGT	G->G	0.557
ccmC	395	TCG	TTG	S->L	0.995
ccmC	399	TTC	TTT	F->F	0.997
ccmC	400	CTT	TTT	L->F	0.997
ccmC	418	CTG	TTG	L->L	0.611
ccmC	421	CGT	TGT	R->C	0.987
ccmC	436	CCT	TCT	P->S	0.999
ccmC	446	CCG	CTG	P->L	0.996
ccmC	458	TCA	TTA	S->L	0.979
ccmC	463	CGT	TGT	R->C	0.999
ccmC	467	GCT	GTT	A->V	0.997
ccmC	473	CCG	CTG	P->L	0.995
ccmC	477	ATC	ATT	I->I	0.612
ccmC	497	TCT	TTT	S->F	0.999
ccmC	521	TCG	TTG	S->L	0.999
ccmC	548	TCT	TTT	S->F	0.96
ccmC	568	CCT	TCT	P->S	0.819
ccmC	568	CCT	TCT	P->S	0.686
ccmC	575	CCC	CTC	P->L	0.998
ccmC	606	TTC	TTT	F->F	0.514
ccmC	608	CCC	CTC	P->L	0.954
ccmC	614	TCA	TTA	S->L	0.982
ccmC	619	CGT	TGT	R->C	0.695
ccmC	624	ATC	ATT	I->I	0.888
ccmC	630	TTC	TTT	F->F	0.866
ccmC	650	CCT	CTT	P->L	0.993
ccmC	665	CCC	CTC	P->L	0.998
ccmFc	77	TCC	TTC	S->F	0.999
ccmFc	78	TCC	TCT	S->S	0.899
ccmFc	89	CCT	CTT	P->L	1
ccmFc	142	CCC	TCC	P->S	0.999
ccmFc	158	TCT	TTT	S->F	0.924
ccmFc	161	TCC	TTC	S->F	0.994
ccmFc	185	CCT	CTT	P->L	0.998
ccmFc	190	CCT	TCT	P->S	0.993
ccmFc	194	TCA	TTA	S->L	0.941
ccmFc	199	CCT	TCT	P->S	0.776
ccmFc	349	CGT	TGT	R->C	0.82
ccmFc	354	TTC	TTT	F->F	0.761
ccmFc	360	CCC	CCT	P->P	0.776
ccmFc	372	TTC	TTT	F->F	0.692
ccmFc	373	CTT	TTT	L->F	0.938
ccmFc	430	CGT	TGT	R->C	0.999
ccmFc	445	CGT	TGT	R->C	0.998
ccmFc	1052	TCT	TTT	S->F	0.99
ccmFc	1313	CCA	CTA	P->L	0.985
ccmFc	1334	TCG	TTG	S->L	0.999
ccmFc	1408	CGG	TGG	R->W	0.989
ccmFc	1433	TCG	TTG	S->L	0.989
ccmFc	1442	TCG	TTG	S->L	0.983

ccmFc	1489	CGA	TGA	R->STOP	0.994
ccmFn	112	CGG	TGG	R->W	0.585
ccmFn	196	CGG	TGG	R->W	0.91
ccmFn	278	CCG	CTG	P->L	0.999
ccmFn	338	CCT	CTT	P->L	1
ccmFn	377	TCG	TTG	S->L	0.762
ccmFn	382	CGT	TGT	R->C	0.939
ccmFn	391	CCG	TCG	P->S	0.773
ccmFn	392	CCG	CTG	P->L	0.815
ccmFn	488	TCA	TTA	S->L	0.999
ccmFn	496	CGG	TGG	R->W	0.999
ccmFn	503	CCA	CTA	P->L	1
ccmFn	519	TTC	TTT	F->F	0.672
ccmFn	523	CTT	TTT	L->F	0.993
ccmFn	612	TTC	TTT	F->F	0.917
ccmFn	640	CGT	TGT	R->C	0.52
ccmFn	947	CCT	CTT	P->L	0.999
ccmFn	956	TCA	TTA	S->L	0.986
ccmFn	994	CGT	TGT	R->C	0.998
ccmFn	999	ACC	ACT	T->T	0.852
ccmFn	1016	TCA	TTA	S->L	0.981
ccmFn	1028	CCA	CTA	P->L	0.999
ccmFn	1043	TCA	TTA	S->L	0.991
ccmFn	1192	CGC	TGC	R->C	0.957
ccmFn	1308	CTC	CTT	L->L	0.536
ccmFn	1427	TCG	TTG	S->L	0.886
ccmFn	1510	CGG	TGG	R->W	0.996
ccmFn	1538	CCA	CTA	P->L	1
ccmFn	1555	CAT	TAT	H->Y	1
ccmFn	1570	CGG	TGG	R->W	0.999
ccmFn	1621	CGG	TGG	R->W	0.999
ccmFn	1639	CGT	TGT	R->C	0.996
ccmFn	1682	TCG	TTG	S->L	0.957
ccmFn	1702	CTT	TTT	L->F	1
ccmFn	1809	TTC	TTT	F->F	0.602
rpl2	77	TCA	TTA	S->L	0.529
rpl2	215	CCA	CTA	P->L	0.995
rpl2	1200	TCC	TCT	S->S	0.597
rpl5	35	TCA	TTA	S->L	0.996
rpl5	59	CCG	CTG	P->L	0.969
rpl5	64	CAC	TAC	H->Y	0.912
rpl5	172	CTG	TTG	L->L	0.975
rpl5	175	CGC	TGC	R->C	1
rpl5	323	TCG	TTG	S->L	0.999
rpl5	335	TCG	TTG	S->L	0.946
rpl5	450	ATC	ATT	I->I	0.532
rpl5	518	CCA	CTA	P->L	0.97
rpl5	521	CCG	CTG	P->L	0.999
rpl10	9	TTC	TTT	F->F	0.576
rpl10	83	TCA	TTA	S->L	0.894
rpl10	101	TCG	TTG	S->L	0.995
rpl10	133	CTA	TTA	L->L	0.588
rpl10	239	TCG	TTG	S->L	0.996
rpl10	314	TCA	TTA	S->L	0.979
rpl10	330	TAC	TAT	Y->Y	0.804
rpl10	374	TCT	TTT	S->F	0.965
rpl16	102	GGC	GGT	G->G	0.814
rpl16	104	ACT	ATT	T->I	1
rpl16	164	ACA	ATA	T->I	1
rpl16	232	CTC	TTC	L->F	0.957
rpl16	321	TCC	TCT	S->S	0.913
rpl16	401	CCA	CTA	P->L	0.997
rpl16	407	TCG	TTG	S->L	0.992
rps1	26	TCA	TTA	S->L	0.678
rps1	45	TCC	TCT	S->S	0.86
rps1	161	CCT	CTT	P->L	0.96
rps1	212	TCC	TTC	S->F	0.867
rps1	629	TCG	TTG	S->L	0.878
rps3	69	TTC	TTT	F->F	0.974
rps3	69	TTC	TTT	F->F	0.974
rps3	92	TCA	TTA	S->L	0.997
rps3	92	TCA	TTA	S->L	0.997
rps3	96	GTC	GTT	V->V	0.578
rps3	96	GTC	GTT	V->V	0.578
rps3	126	TTC	TTT	F->F	0.574
rps3	126	TTC	TTT	F->F	0.574
rps3	512	TCA	TTA	S->L	0.992
rps3	512	TCA	TTA	S->L	0.992

rps3	699	CCC	CCT	P->P	0.802
rps3	699	CCC	CCT	P->P	0.802
rps3	713	TCG	TTG	S->L	0.993
rps3	713	TCG	TTG	S->L	0.993
rps3	763	CTC	TTC	L->F	0.619
rps3	763	CTC	TTC	L->F	0.619
rps3	890	TCG	TTG	S->L	0.905
rps3	890	TCG	TTG	S->L	0.905
rps3	1031	CCC	CTC	P->L	0.889
rps3	1031	CCC	CTC	P->L	0.889
rps3	1364	CCG	CTG	P->L	1
rps3	1364	CCG	CTG	P->L	1
rps3	1391	CCG	CTG	P->L	0.999
rps3	1391	CCG	CTG	P->L	0.999
rps3	1588	CCT	TCT	P->S	0.992
rps3	1588	CCT	TCT	P->S	0.992
rps4	38	TCA	TTA	S->L	0.932
rps4	38	TCA	TTA	S->L	0.932
rps4	49	CGG	TGG	R->W	1
rps4	49	CGG	TGG	R->W	1
rps4	164	TCA	TTA	S->L	0.999
rps4	164	TCA	TTA	S->L	0.999
rps4	193	CAT	TAT	H->Y	0.988
rps4	193	CAT	TAT	H->Y	0.988
rps4	266	CCA	CTA	P->L	0.99
rps4	266	CCA	CTA	P->L	0.99
rps4	278	TCG	TTG	S->L	0.997
rps4	278	TCG	TTG	S->L	0.997
rps4	290	CCG	CTG	P->L	0.999
rps4	290	CCG	CTG	P->L	0.999
rps4	307	CGT	TGT	R->C	0.984
rps4	307	CGT	TGT	R->C	0.984
rps4	335	CCG	CTG	P->L	0.996
rps4	335	CCG	CTG	P->L	0.996
rps4	474	ATC	ATT	I->I	0.673
rps4	474	ATC	ATT	I->I	0.673
rps4	482	TCA	TTA	S->L	0.983
rps4	482	TCA	TTA	S->L	0.983
rps4	668	TCA	TTA	S->L	0.662
rps4	668	TCA	TTA	S->L	0.662
rps4	693	TCC	TCT	S->S	0.685
rps4	693	TCC	TCT	S->S	0.685
rps4	840	ACC	ACT	T->T	0.652
rps4	840	ACC	ACT	T->T	0.652
rps4	885	ATC	ATT	I->I	0.944
rps4	885	ATC	ATT	I->I	0.944
rps4	894	ATC	ATT	I->I	0.786
rps4	894	ATC	ATT	I->I	0.786
rps4	914	TCG	TTG	S->L	0.995
rps4	914	TCG	TTG	S->L	0.995
rps4	925	CAT	TAT	H->Y	0.985
rps4	925	CAT	TAT	H->Y	0.985
rps4	950	TCT	TTT	S->F	0.997
rps4	950	TCT	TTT	S->F	0.997
rps4	1001	CCA	CTA	P->L	0.991
rps4	1001	CCA	CTA	P->L	0.991
rps4	1010	CCT	CTT	P->L	0.996
rps4	1010	CCT	CTT	P->L	0.996
rps4	1015	CGG	TGG	R->W	0.995
rps4	1015	CGG	TGG	R->W	0.995
rps7	177	CCC	CCT	P->P	0.619
rps7	177	CCC	CCT	P->P	0.619
rps7	332	TCA	TTA	S->L	0.986
rps7	332	TCA	TTA	S->L	0.986
rps10	62	ACG	ATG	T->M	0.998
rps10	298	CGG	TGG	R->W	0.986
rps10	310	CGT	TGT	R->C	0.633
rps10	358	CGG	TGG	R->W	0.995
rps10	387	TTC	TTT	F->F	0.962
rps10	398	TCA	TTA	S->L	0.643
rps12	284	TCC	TTC	S->F	0.999
rps13	5	TCA	TTA	S->L	0.518
rps13	26	TCA	TTA	S->L	0.999
rps13	56	TCA	TTA	S->L	0.995
rps13	100	CGT	TGT	R->C	0.964
rps13	193	CTT	TTT	L->F	0.767
rps13	287	TCG	TTG	S->L	0.999
rps14	194	TCC	TTC	S->F	0.932

rps14	271	CCT	TCT	P->S	0.744
rps19	138	TCC	TCT	S->S	0.977
rps19	163	CCT	TCT	P->S	0.958
rps19	164	CCT	CTT	P->L	0.947
rps19	180	ATC	ATT	I->I	0.777
psaB	40	CAG	TAG	Q->STOP	0.87
psaB	81	ACC	ACT	T->T	0.569
psaB	130	CAG	TAG	Q->STOP	0.98
psaB	175	CTG	TTG	L->L	0.784
psaB	185	TCC	TTC	S->F	0.625
psaB	353	TCT	TTT	S->F	0.535
psaB	410	ACA	ATA	T->I	0.591
psaB	433	CTT	TTT	L->F	0.953
psaB	504	TTC	TTT	F->F	0.596
psaB	525	CTC	CTT	L->L	0.67
psaB	539	TCG	TTG	S->L	0.741
psaB	549	TTC	TTT	F->F	0.843
psaB	599	CCT	CTT	P->L	0.504
psaB	649	CCG	TCG	P->S	0.795
psaB	655	CCC	TCC	P->S	0.69
psaB	680	ACA	ATA	T->I	0.586
psaB	710	CCC	CTC	P->L	0.602
psaB	716	TCA	TTA	S->L	0.9
psaB	740	TCC	TTC	S->F	0.592
psaB	774	CTC	CTT	L->L	0.72
psaB	823	CAT	TAT	H->Y	0.681
psaB	853	CTT	TTT	L->F	0.73
psaB	853	CTT	TTT	L->F	0.846
psaB	885	TTC	TTT	F->F	0.517
psaB	929	CCT	CTT	P->L	0.726
psaB	978	ATC	ATT	I->I	0.727
psaB	991	CAC	TAC	H->Y	0.959
psaB	997	CAA	TAA	Q->STOP	0.681
psaB	1012	CTA	TTA	L->L	0.839
psaB	1019	TCT	TTT	S->F	0.766
psaB	1165	CAT	TAT	H->Y	0.547
psaB	1182	TTC	TTT	F->F	0.628
psaB	1326	GTC	GTT	V->V	0.961
psaB	1381	CAA	TAA	Q->STOP	0.777
psaB	1399	CAT	TAT	H->Y	0.849
psaB	1485	CCC	CCT	P->P	0.672
psaB	1541	CCT	CTT	P->L	0.615
psaB	1622	GCA	GTA	A->V	0.633
psaB	1631	TCA	TTA	S->L	0.915
psaB	1659	TTC	TTT	F->F	0.858
psaB	1743	TTC	TTT	F->F	0.826
psaB	1797	ATC	ATT	I->I	0.673
psaB	1820	TCA	TTA	S->L	0.876
psaB	1838	TCC	TTC	S->F	0.778
psaB	1886	TCA	TTA	S->L	0.771
psaB	2102	TCC	TTC	S->F	0.853
psaB	2103	TCC	TCT	S->S	0.92
psaB	2134	CAC	TAC	H->Y	0.994
psaB	2141	TCT	TTT	S->F	0.882
psaB	2189	TCA	TTA	S->L	0.978
psbZ	11	TCC	TTC	S->F	0.625
psbZ	13	CTC	TTC	L->F	0.776
psbZ	158	TCA	TTA	S->L	0.764
psbZ	161	TCC	TTC	S->F	0.775
psbZ	182	CCC	CTC	P->L	0.931
psbZ	183	CCC	CCT	P->P	0.636
psbZ	200	CCT	CTT	P->L	0.588
psbZ	237	TCC	TCT	S->S	0.712
psbZ	280	CTT	TTT	L->F	0.694
psbZ	287	TCT	TTT	S->F	0.55
atpE	13	CTT	TTT	L->F	0.665
atpE	29	CCT	CTT	P->L	0.967
atpE	29	CCT	CTT	P->L	0.816
atpE	104	CCA	CTA	P->L	0.604
atpE	109	CAC	TAC	H->Y	0.706
atpE	116	CCT	CTT	P->L	0.571
atpE	166	CAA	TAA	Q->STOP	0.981
atpE	184	CTG	TTG	L->L	0.986
atpE	355	CTC	TTC	L->F	0.57
atpE	398	TCA	TTA	S->L	0.64
atpF	14	ACC	ATC	T->I	0.597
atpF	20	TCT	TTT	S->F	0.841
atpF	92	CCA	CTA	P->L	0.993

atpF	417	GTC	GTT	V->V	0.755
atpI	14	TCA	TTA	S->L	0.838
atpI	20	TCC	TTC	S->F	0.781
atpI	106	CAC	TAC	H->Y	0.689
atpI	158	TCA	TTA	S->L	0.677
atpI	194	CCA	CTA	P->L	0.603
atpI	205	CAG	TAG	Q->STOP	0.87
atpI	226	CTT	TTT	L->F	0.596
atpI	238	CGA	TGA	R->STOP	0.508
atpI	255	ACC	ACT	T->T	0.635
atpI	310	CTT	TTT	L->F	0.799
atpI	370	CAT	TAT	H->Y	0.856
atpI	370	CTA	TTA	L->L	0.944
atpI	483	TTC	TTT	F->F	0.594
atpI	492	TAC	TAT	Y->Y	0.617
atpI	540	TTC	TTT	F->F	0.589
atpI	610	CTT	TTT	L->F	0.758
atpI	617	TCT	TTT	S->F	0.808
atpI	637	CCT	TCT	P->S	0.878
atpI	637	CCT	TCT	P->S	0.709
atpI	638	CCT	CTT	P->L	0.585
atpI	638	CCT	CTT	P->L	0.886
atpI	644	CCT	CTT	P->L	0.609
atpI	644	CCT	CTT	P->L	0.945
atpI	685	CTT	TTT	L->F	0.62
petN	46	CCA	TCA	P->S	0.578
petN	59	TCA	TTA	S->L	0.533
petN	87	CTC	CTT	L->L	0.567
ndhC	30	TTC	TTT	F->F	0.827
ndhC	76	CTA	TTA	L->L	0.831
ndhC	148	CCA	TCA	P->S	0.553
ndhC	227	ACG	ATG	T->M	0.553
ndhC	241	CCA	TCA	P->S	0.505
ndhJ	113	TCC	TTC	S->F	0.829
ndhJ	182	CCC	CTC	P->L	0.812
ndhJ	183	CCC	CCT	P->P	0.682
ndhJ	292	CCG	TCG	P->S	0.583
ndhJ	293	CCG	CTG	P->L	0.788
ndhJ	447	CCC	CCT	P->P	0.61
ndhK	65	TCA	TTA	S->L	0.589
ndhK	95	TCT	TTT	S->F	0.778
ndhK	112	CTT	TTT	L->F	0.948
ndhK	135	TTC	TTT	F->F	0.765
ndhK	161	TCA	TTA	S->L	0.718
ndhK	205	CCT	TCT	P->S	0.605
ndhK	259	CCC	TCC	P->S	0.955
ndhK	345	TTC	TTT	F->F	0.62
ndhK	350	ACC	ATC	T->I	0.53
ndhK	351	ACC	ACT	T->T	0.692
ndhK	356	TCT	TTT	S->F	0.582
ndhK	392	CCT	CTT	P->L	0.503
ndhK	410	CCG	CTG	P->L	0.941
ndhK	493	CGA	TGA	R->STOP	0.6
ndhK	580	CAA	TAA	Q->STOP	0.634
ndhK	595	CAA	TAA	Q->STOP	0.871
ndhK	620	CCT	CTT	P->L	0.555
ndhK	622	CCT	TCT	P->S	0.68
ndhK	623	CCT	CTT	P->L	0.701
ndhK	661	CAC	TAC	H->Y	0.678
rps2	146	ACA	ATA	T->I	0,768
rps2	161	ACT	ATT	T->I	0,55
rps2	498	CCC	CCT	P->P	0,942
rps2	568	CCA	TCA	P->S	0,64
rps2	569	CCA	CTA	P->L	0,576
rps2	585	ATC	ATT	I->I	0,572
rps2	590	ACA	ATA	T->I	0,712
rps2	602	CCG	CTG	P->L	0,918
rps2	650	TCA	TTA	S->L	0,514
rps2	654	ATC	ATT	I->I	0,571
rps2	664	CTT	TTT	L->F	0,688
rps2	681	TTC	TTT	F->F	0,612
rps2	701	TCT	TTT	S->F	0,693
rps2	719	CCT	CTT	P->L	0,835
rps14-cp	61	CAT	TAT	H->Y	0.677
rps14-cp	70	CGT	TGT	R->C	0.821
rps14-cp	104	CCG	CTG	P->L	0.837
rps14-cp	151	CCG	TCG	P->S	0.842
rps14-cp	152	CCG	CTG	P->L	0.736

rps14-cp	167	CCT	CTT	P->L	0.691
rps14-cp	197	TCG	TTG	S->L	0.758
rps14-cp	250	CGG	TGG	R->W	0.761

Table S5: Mitochondrial RNA genes annotated in *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) mitogenomes. Locally Collinear Blocks (LCBs) correspond to conserved regions between *Lgh* and *Lpm* mitogenomes, identified by ProgressiveMauve.

Gene	LCB	Molecule
tRNA		
trnC-GCA	LCB0	M2
trnE-UUC	LCB5	M1
trnF-GAA	LCB4	M1
trnG-GCC	LCB1	M1
	LCB5	M1
trnH-GUG	LCB5	M1
trnK-UUU	LCB5	M1
trnM-CAU	LCB1	M1
	LCB4	M1
	LCB5	M1
	interLCB	M1
	interLCB	M1
	interLCB	M1
trnN-GUU	LCB0	M2
trnP-UGG	LCB4	M1
trnQ-UUG	LCB5	M1
trnS-GCU	LCB4	M1
trnS-UGA	LCB0	M2
trnY-GUA	LCB0	M2
rRNA		
rrn26	interLCB	M1
	interLCB	M1
	interLCB	M1
rrn5	LCB3	M1
	interLCB	M1
rrn18	LCB3	M1
	interLCB	M1

Table S6: Mobile elements in *Ludwigia grandiflora* subsp. *hexapetala* (A) and *Ludwigia peploides* subsp. *montevidensis* (B) mitogenomes. LINE stands for Long Interspersed Nuclear Elements, LTR refers to Long Terminal Repeat, and NERVE denotes Nonretroviral Endogenized RNA Virus Elements.

A

Ludwigia grandiflora* subsp. *hexapetala

Class	Coordinates	Length (bp)	Genes
	Molecule M1		
Class I / LINE	10,751-11,546	796	
	20,872-21,757	886	RH
	129,949-130,834	886	RH
	132,307-132,546	240	RH-like superfamily
	148,898-149,038	141	ENDO
	213,723-214,059	337	
	261,737-261,977	241	ENDO
	279,762-280,109	348	RH
	329,152-329,604	453	RT
	376,573-376,726	154	RT
	382,465-382,704	240	RH-like superfamily
	384,177-385,062	886	RH
	456,181-456,373	193	RT
	488,076-488,257	182	RT
	542,947-543,737	791	ENDO
Class I / LTR / Ty1-copia	3,928-4,132	205	PROT
	25,510-25,752	243	INT
	75,637-75,872	236	GAG
	148,595-148,792	198	RH
	176,581-176,676	96	RT
	353,782-353,973	192	RT
	404,483-405,581	1099	RT
Class I / LTR / Ty3-gypsy / chromovirus	42,455-42,728	274	INT
	166,953-168,071	1119	RH, RT
	183,238-183,359	122	GAG
	243,942-244,270	329	CHD, INT
	248,614-250,320	1707	PROT, INT, RT, RH
	265,861-266,763	903	RT, RH
	285,481-285,699	219	PROT
	429,875-430,905	1031	Integrase_H2C2 pfam17921 Integrase zinc binding domain
	479,719-479,991	273	INT
	492,676-496,668	3993	ENDO, RT, RH
	521,573-522,691	1119	RH, RT
Class I / LTR / Ty3-gypsy / non-chromovirus	53,135-54,241	1107	INT, RT
	88,483-89,405	923	RH, aRH
	166,168-166,592	425	RH
	320,597-320,828	232	PROT
	325,676-325,993	318	RNA-directed DNA polymerase
	397,214-397,401	188	RT
	457,374-457,550	177	RT
	457,707-457,853	147	RH
	523,052-523,476	425	RH
NERVE mitovirus	25,797-26,171	375	RNA-dependent RNA polymerase
	62,933-64,192	1260	RNA-directed DNA polymerase
	68,212-68,934	723	RNA-directed DNA polymerase
	150,031-150,348	318	RH domain
	168,815-169,924	1110	RNA-dependent RNA polymerase
	204,099-205,205	1107	RNA-dependent RNA polymerase, pfam05919
	426,804-427,376	573	RNA-dependent RNA polymerase
	446,229-446,951	723	RNA-directed DNA polymerase
	450,971-453,182	2212	RNA-directed DNA polymerase
	473,210-474,128	919	
	519,720-520,829	1110	RNA-dependent RNA polymerase
Plasmid-derived	291,293-294,250	2958	DNA polymerase type B
	Molecule M2		
Class I / LINE	83,895-84,162	268	RT
	101,506-101,677	172	RT
	103,198-103,410	213	RH
Class I / LTR / Ty1-copia	16,980-17,060	81	RT
	21,095-21,382	288	GAG
	21,896-22,099	204	PROT
	22,112-22,696	585	INT
	23,502-24,269	768	RT
	24,312-24,407	96	RH
	24,538-24,919	382	RH
	64,264-64,356	93	RT
	88,990-89,664	675	RT
	157,481-157,821	341	RT
Class I / LTR / Ty3-gypsy / chromovirus	122,504-122,781	278	INT
Class I / LTR / Ty3-gypsy / non-chromovirus	91,156-91,488	333	RH
	133,044-133,419	376	INT
	135,340-135,889	550	INT
	150,771-150,887	117	aRH
	166,472-166,643	172	GAG
NERVE mitovirus	75,427-77,316	1890	RNA-dependent RNA polymerase

B

Ludwigia peploides* subsp. *montevidensis

Class	Coordinates	Length (bp)	Genes
	Molecule M1		
Class I / LINE	11,250-12,135	886	RH
	13,608-13,847	240	RH-like superfamily
	19,581-19,734	154	RT
	66,699-67,139	441	ENDO
	115,874-116,220	347	RH
	133,999-134,239	241	ENDO
	181,894-182,230	337	RT
	246,839-246,979	141	ENDO
	263,310-263,549	240	RH-like superfamily
	265,022-265,907	886	RH
	322,717-323,538	822	RT
	395,139-396,024	886	RH
	466,969-467,161	193	RT
	498,852-499,033	182	RT
	553,680-554,329	650	ENDO
Class I / LTR / Ty1-copia	3,927-4,131	205	PROT
	10,273-10,365	93	GAG
	42,318-42,509	192	RT
	219,222-219,317	96	RT
	247,085-247,281	197	RH
	266,792-266,884	93	GAG
	286,067-287,165	1099	RT
	341,637-341,739	103	INT
	391,146-391,387	242	INT
	396,909-397,001	93	GAG
	450,093-450,328	236	GAG
Class I / LTR / Ty3-gypsy / chromovirus	99,166-99,384	219	PROT
	129,206-130,114	909	RH, RT
	145,613-147,356	1744	INT, RH, RT, PROT
	151,705-152,032	328	INT, CHD
	212,545-212,666	122	GAG
	227,810-228,943	1134	RH, RT
	311,444-312,349	906	INT
	328,680-328,832	153	INT
	331,981-332,085	105	RH
	374,181-374,454	274	INT
	490,497-490,769	273	INT
	503,451-507,445	3995	ENDO, RT, RH
	532,321-533,454	1134	RH, RT
Class I / LTR / Ty3-gypsy / non-chromovirus	70,282-70,602	321	RT
	75,428-75,659	232	PROT
	229,303-229,726	424	RH
	278,803-278,990	188	RT
	362,702-363,794	1093	RT, INT
	436,549-437,504	956	aRH, RH
	468,162-468,338	177	RT
	468,495-468,641	147	RH
	533,814-534,237	424	RH
NERVE mitovirus	190,740-191,842	1103	RNA-dependent RNA polymerase, pfam05919
	225,971-227,080	1110	RNA-dependent RNA polymerase
	245,529-245,846	318	RH domain
	308,373-308,945	573	RNA-dependent RNA polymerase
	348,014-348,736	723	RNA-directed DNA polymerase
	352,755-354,014	1260	RNA-directed DNA polymerase
	390,728-391,101	374	RNA-dependent RNA polymerase
	457,022-457,744	723	RNA-directed DNA polymerase
	461,763-463,045	1283	RNA-directed DNA polymerase
	463,335-464,077	743	
	483,993-484,911	919	
	530,482-531,591	1110	RNA-dependent RNA polymerase
Plasmid-derived	104,764-107,721	2958	DNA polymerase type B
	Molecule M2		
Class I / LINE	83,859-84,126	268	RT
	101,461-101,628	168	RT
	103,151-103,363	213	RH
Class I / LTR / Ty1-copia	16,976-17,056	81	RT
	21,091-21,378	288	GAG
	21,892-22,095	204	PROT
	23,498-24,265	768	RT
	24,308-24,403	96	RH
	24,534-24,915	382	RH
	64,245-64,337	93	RT
	88,953-89,627	675	RT
	157,381-157,721	341	RT
Class I / LTR / Ty3-gypsy / chromovirus	122,430-122,707	278	INT
Class I / LTR / Ty3-gypsy / non-chromovirus	91,117-91,448	332	RH
	132,966-133,341	376	INT
	135,262-135,811	550	INT
	150,675-150,791	117	aRH
	166,677-166,848	172	GAG
NERVE mitovirus	75,404-77,293	1890	RNA-dependent RNA polymerase

Table S7: Non-tandem repeated sequences in *Ludwigia grandiflora* subsp. *hexapetala* and *Ludwigia peploides* subsp. *montevidensis* mitogenomes.

Ludwigia grandiflora subsp. *hexapetala*							
Molecule 1	**Start 1**	**End 1**	**Molecule 2**	**Start 2**	**End 2**	**Direction**	**Length (bp)**
Repeats M1-M1							
M1	11549	21999	M1	120626	131076	forward	10451
M1	122334	132673	M1	382338	392677	reverse	10340
M1	163643	173026	M1	516618	526001	reverse	9384
M1	13257	21999	M1	383935	392677	reverse	8743
M1	62933	69195	M1	445968	452230	reverse	6263
M1	312774	313287	M1	437121	437634	reverse	514
M1	284951	285361	M1	297198	297608	reverse	411
M1	77591	77990	M1	184532	184931	reverse	400
M1	5195	5432	M1	17278	17515	forward	238
M1	5195	5432	M1	126355	126592	forward	238
M1	5195	5432	M1	388419	388656	reverse	238
M1	223090	223278	M1	316968	317156	forward	189
M1	415088	415275	M1	534368	534555	forward	188
M1	6538	6718	M1	367876	368056	reverse	181
M1	78043	78172	M1	184348	184477	reverse	130
M1	6577	6694	M1	438458	438575	forward	118
M1	367900	368017	M1	438458	438575	reverse	118
M1	367794	367910	M1	438585	438701	reverse	117
M1	7920	8032	M1	373826	373938	forward	113
M1	32962	33070	M1	148031	148139	forward	109
M1	145877	145972	M1	437035	437130	reverse	96
M1	194531	194625	M1	214479	214573	reverse	95
M1	108334	108418	M1	376774	376858	forward	85
M1	208207	208291	M1	368236	368320	forward	85
M1	34355	34432	M1	132192	132269	forward	78
M1	34355	34432	M1	382742	382819	reverse	78
M1	288585	288662	M1	333946	334023	forward	78
M1	120301	120374	M1	176144	176217	reverse	74
M1	78216	78288	M1	184232	184304	reverse	73
M1	19798	19860	M1	355714	355776	reverse	63
M1	69520	69582	M1	276915	276977	forward	63
M1	128875	128937	M1	355714	355776	reverse	63
M1	355714	355776	M1	386074	386136	forward	63
M1	143419	143480	M1	309849	309910	forward	62
M1	16901	16961	M1	338343	338403	forward	61
M1	125978	126038	M1	338343	338403	forward	61
M1	167874	167934	M1	249770	249830	forward	61
M1	249770	249830	M1	521710	521770	reverse	61
M1	338343	338403	M1	388973	389033	reverse	61
M1	31789	31848	M1	526931	526990	reverse	60
M1	44814	44873	M1	276543	276602	reverse	60
M1	528899	528957	M1	538059	538117	forward	59
M1	2423	2479	M1	38260	38316	forward	57
M1	31850	31906	M1	526873	526929	reverse	57
M1	194413	194469	M1	214612	214668	reverse	57
M1	6513	6568	M1	438394	438449	forward	56
M1	2373	2427	M1	38205	38259	forward	55
M1	49272	49325	M1	263886	263939	reverse	54
M1	77995	78047	M1	184475	184527	reverse	53
M1	82334	82385	M1	164604	164655	reverse	52
M1	82334	82385	M1	524989	525040	forward	52
M1	34434	34484	M1	132271	132321	forward	51
M1	34434	34484	M1	382690	382740	reverse	51
M1	288229	288279	M1	291066	291116	forward	51
M1	74955	75004	M1	236571	236620	reverse	50
M1	17913	17960	M1	255010	255057	forward	48
M1	126990	127037	M1	255010	255057	forward	48
M1	255010	255057	M1	387974	388021	reverse	48
M1	16817	16863	M1	337854	337900	forward	47
M1	54633	54679	M1	369262	369308	forward	47
M1	125894	125940	M1	337854	337900	forward	47
M1	321812	321858	M1	381274	381320	forward	47
M1	321936	321982	M1	381401	381447	forward	47
M1	337854	337900	M1	389071	389117	reverse	47

M1	288591	288636	M1	368106	368151	forward	46
M1	333952	333997	M1	368106	368151	forward	46
M1	225383	225426	M1	456557	456600	forward	44
M1	337253	337296	M1	490941	490984	reverse	44
M1	45486	45528	M1	433404	433446	reverse	43
M1	145210	145252	M1	289836	289878	forward	43
M1	151915	151957	M1	317898	317940	reverse	43
M1	370264	370306	M1	431518	431560	forward	43
M1	475310	475352	M1	516536	516578	reverse	43
M1	221	262	M1	378372	378413	reverse	42
M1	20065	20106	M1	142258	142299	reverse	42
M1	129142	129183	M1	142258	142299	reverse	42
M1	142258	142299	M1	385828	385869	forward	42
M1	34265	34305	M1	67335	67375	forward	41
M1	34265	34305	M1	447788	447828	reverse	41
M1	20002	20041	M1	509995	510034	reverse	40
M1	129079	129118	M1	509995	510034	reverse	40
M1	133896	133935	M1	284645	284684	reverse	40
M1	288599	288638	M1	438373	438412	reverse	40
M1	333960	333999	M1	438373	438412	reverse	40
M1	385893	385932	M1	509995	510034	forward	40
M1	34666	34704	M1	459040	459078	reverse	39
M1	45452	45490	M1	433449	433487	reverse	39
M1	78783	78821	M1	509998	510036	forward	39
M1	90452	90490	M1	236609	236647	forward	39
M1	149182	149220	M1	152789	152827	forward	39
M1	394376	394414	M1	470990	471028	reverse	39
M1	20388	20425	M1	315480	315517	forward	38
M1	91063	91100	M1	208160	208197	forward	38
M1	99975	100012	M1	198279	198316	forward	38
M1	108498	108535	M1	376938	376975	forward	38
M1	129465	129502	M1	315480	315517	forward	38
M1	228460	228497	M1	321182	321219	reverse	38
M1	315480	315517	M1	385509	385546	reverse	38
M1	368114	368151	M1	438375	438412	reverse	38
M1	20002	20038	M1	78783	78819	reverse	37
M1	55420	55456	M1	526995	527031	forward	37
M1	78783	78819	M1	129079	129115	reverse	37
M1	78783	78819	M1	385896	385932	forward	37
M1	178967	179003	M1	299580	299616	forward	37
M1	6413	6448	M1	192430	192465	forward	36
M1	144236	144271	M1	230033	230068	reverse	36
M1	378795	378830	M1	398734	398769	reverse	36
M1	6684	6718	M1	438585	438619	forward	35
M1	18481	18515	M1	66057	66091	forward	35
M1	18481	18515	M1	449072	449106	reverse	35
M1	66057	66091	M1	127558	127592	forward	35
M1	66057	66091	M1	387419	387453	reverse	35
M1	66977	67011	M1	105581	105615	reverse	35
M1	77684	77718	M1	189539	189573	reverse	35
M1	105581	105615	M1	448152	448186	forward	35
M1	127558	127592	M1	449072	449106	reverse	35
M1	184804	184838	M1	189539	189573	forward	35
M1	225465	225499	M1	500436	500470	forward	35
M1	387419	387453	M1	449072	449106	forward	35
M1	12568	12601	M1	392678	392711	forward	34
M1	121645	121678	M1	392678	392711	forward	34
M1	150385	150418	M1	194972	195005	reverse	34
M1	179010	179043	M1	299623	299656	forward	34
M1	316387	316420	M1	346497	346530	reverse	34
M1	321724	321757	M1	381193	381226	forward	34
M1	54744	54776	M1	369383	369415	forward	33
M1	66377	66409	M1	106191	106223	reverse	33
M1	79477	79509	M1	137937	137969	forward	33
M1	106191	106223	M1	448754	448786	forward	33
M1	142723	142755	M1	229919	229951	reverse	33
M1	171606	171638	M1	476145	476177	forward	33
M1	280769	280801	M1	372920	372952	forward	33
M1	321601	321633	M1	381081	381113	forward	33
M1	476145	476177	M1	518006	518038	reverse	33
M1	476146	476178	M1	497217	497249	forward	33
M1	45530	45561	M1	433371	433402	reverse	32
M1	79431	79462	M1	137892	137923	forward	32
M1	100222	100253	M1	203624	203655	forward	32
M1	170512	170543	M1	476102	476133	forward	32
M1	171607	171638	M1	497217	497248	forward	32
M1	312383	312414	M1	345838	345869	forward	32
M1	476102	476133	M1	519101	519132	reverse	32

M1	497217	497248	M1	518006	518037	reverse	32
M1	64993	65023	M1	183790	183820	forward	31
M1	170930	170960	M1	289925	289955	forward	31
M1	183790	183820	M1	450140	450170	reverse	31
M1	289925	289955	M1	518684	518714	reverse	31
M1	368026	368056	M1	438419	438449	reverse	31
M1	438208	438238	M1	438217	438247	forward	31
M1	20512	20541	M1	315558	315587	forward	30
M1	22928	22957	M1	170028	170057	forward	30
M1	22928	22957	M1	519587	519616	reverse	30
M1	75027	75056	M1	215550	215579	forward	30
M1	129589	129618	M1	315558	315587	forward	30
M1	171929	171958	M1	171970	171999	forward	30
M1	171929	171958	M1	517645	517674	reverse	30
M1	171970	171999	M1	517686	517715	reverse	30
M1	309773	309802	M1	415231	415260	reverse	30
M1	309773	309802	M1	534511	534540	reverse	30
M1	315558	315587	M1	385393	385422	reverse	30
M1	517645	517674	M1	517686	517715	forward	30

Repeats M1-M2

M1	414973	415421	M2	7384	7832	forward	449
M1	309847	310212	M2	4205	4570	reverse	366
M1	438375	438569	M2	140233	140427	reverse	195
M1	534368	534555	M2	7499	7686	forward	188
M1	260642	260825	M2	139747	139930	reverse	184
M1	6684	6849	M2	140064	140229	reverse	166
M1	131199	131351	M2	98976	99128	forward	153
M1	383660	383812	M2	98976	99128	reverse	153
M1	526048	526180	M2	70417	70549	forward	133
M1	6577	6688	M2	140233	140344	reverse	112
M1	367906	368017	M2	140233	140344	forward	112
M1	336061	336170	M2	75597	75706	forward	110
M1	526822	526929	M2	12068	12175	forward	108
M1	152756	152860	M2	77917	78021	reverse	105
M1	368114	368214	M2	140390	140490	forward	101
M1	31807	31906	M2	12119	12218	reverse	100
M1	101172	101269	M2	70284	70381	forward	98
M1	219940	220034	M2	95784	95878	reverse	95
M1	90954	91045	M2	80915	81006	reverse	92
M1	260827	260909	M2	139663	139745	reverse	83
M1	310262	310340	M2	2423	2501	reverse	79
M1	208120	208197	M2	80866	80943	reverse	78
M1	13643	13718	M2	98040	98115	forward	76
M1	122720	122795	M2	98040	98115	forward	76
M1	392216	392291	M2	98040	98115	reverse	76
M1	337531	337605	M2	143363	143437	reverse	75
M1	260569	260640	M2	139932	140003	reverse	72
M1	3552	3616	M2	104366	104430	forward	65
M1	193868	193932	M2	10275	10339	reverse	65
M1	143419	143480	M2	4507	4568	reverse	62
M1	300311	300372	M2	114078	114139	forward	62
M1	328849	328909	M2	11051	11111	forward	61
M1	193349	193407	M2	10545	10603	reverse	59
M1	277474	277532	M2	134536	134594	reverse	59
M1	183734	183791	M2	3186	3243	forward	58
M1	483474	483531	M2	38032	38089	forward	58
M1	6513	6568	M2	140353	140408	reverse	56
M1	114221	114276	M2	82507	82562	reverse	56
M1	250912	250966	M2	15374	15428	forward	55
M1	193470	193521	M2	10426	10477	reverse	52
M1	6851	6901	M2	140012	140062	reverse	51
M1	91063	91112	M2	80854	80903	reverse	50
M1	193813	193862	M2	10345	10394	reverse	50
M1	237530	237579	M2	6030	6079	forward	50
M1	237581	237629	M2	6081	6129	forward	49
M1	527845	527892	M2	28630	28677	reverse	48
M1	31836	31882	M2	11360	11406	reverse	47
M1	284529	284575	M2	74380	74426	forward	47
M1	10207	10252	M2	67926	67971	forward	46
M1	54422	54467	M2	41295	41340	reverse	46
M1	219380	219425	M2	87566	87611	reverse	46
M1	371703	371747	M2	156725	156769	reverse	45
M1	250867	250910	M2	15329	15372	forward	44
M1	277930	277973	M2	138175	138218	forward	44
M1	3668	3709	M2	103545	103586	reverse	42
M1	11767	11808	M2	67095	67136	reverse	42

M1	120844	120885	M2	67095	67136	reverse	42
M1	526931	526972	M2	12177	12218	forward	42
M1	149182	149220	M2	77950	77988	reverse	39
M1	334028	334066	M2	56171	56209	reverse	39
M1	495720	495758	M2	338	376	reverse	39
M1	11609	11646	M2	67265	67302	reverse	38
M1	120686	120723	M2	67265	67302	reverse	38
M1	288599	288636	M2	140390	140427	forward	38
M1	333960	333997	M2	140390	140427	forward	38
M1	525996	526033	M2	70366	70403	forward	38
M1	438583	438619	M2	140195	140231	reverse	37
M1	452906	452942	M2	75691	75727	reverse	37
M1	11655	11690	M2	67221	67256	reverse	36
M1	120732	120767	M2	67221	67256	reverse	36
M1	172165	172200	M2	153336	153371	forward	36
M1	398737	398772	M2	114244	114279	reverse	36
M1	517444	517479	M2	153336	153371	reverse	36
M1	312739	312773	M2	140538	140572	forward	35
M1	337489	337523	M2	143447	143481	reverse	35
M1	367876	367910	M2	140195	140229	forward	35
M1	505035	505068	M2	3638	3671	reverse	34
M1	527812	527845	M2	28676	28709	reverse	34
M1	17539	17571	M2	26713	26745	reverse	33
M1	126616	126648	M2	26713	26745	reverse	33
M1	180834	180866	M2	114667	114699	reverse	33
M1	228460	228492	M2	52856	52888	reverse	33
M1	321187	321219	M2	52856	52888	forward	33
M1	336028	336060	M2	75566	75598	forward	33
M1	378795	378827	M2	114247	114279	forward	33
M1	388363	388395	M2	26713	26745	forward	33
M1	526897	526929	M2	11360	11392	forward	33
M1	278160	278191	M2	10875	10906	reverse	32
M1	321189	321220	M2	33211	33242	reverse	32
M1	465213	465244	M2	61620	61651	forward	32
M1	542203	542234	M2	154385	154416	reverse	32
M1	228460	228490	M2	33212	33242	forward	31
M1	368026	368056	M2	140353	140383	forward	31
M1	378575	378605	M2	116632	116662	reverse	31
M1	35780	35809	M2	5418	5447	forward	30
M1	99593	99622	M2	84402	84431	reverse	30
M1	309773	309802	M2	7642	7671	reverse	30
M1	411878	411907	M2	111492	111521	forward	30

Repeats M2-M2

M2	70688	70753	M2	70817	70882	forward	66
M2	11411	11458	M2	12194	12241	forward	48
M2	54051	54098	M2	145529	145576	reverse	48
M2	11360	11406	M2	12143	12189	forward	47
M2	107670	107710	M2	142149	142189	reverse	41
M2	138121	138159	M2	151685	151723	reverse	39
M2	33212	33242	M2	52858	52888	reverse	31

Ludwigia peploides subsp. *montevidensis*

Repeats M1-M1

M1	264780	274180	M1	394897	404297	forward	9401
M1	222869	232244	M1	527380	536755	forward	9376
M1	346199	354003	M1	455207	463011	forward	7805
M1	6737	13908	M1	263249	270420	reverse	7172
M1	6737	12377	M1	394897	400537	reverse	5641
M1	210976	211556	M1	447793	448373	reverse	581
M1	82969	83482	M1	318686	319199	forward	514
M1	296661	296887	M1	545119	545345	forward	227
M1	6131	6324	M1	61904	62097	forward	194
M1	5902	6061	M1	61669	61828	forward	160
M1	79136	79292	M1	172716	172872	forward	157
M1	28282	28423	M1	320020	320161	forward	142
M1	146305	146413	M1	331979	332087	forward	109
M1	247735	247843	M1	383837	383945	forward	109
M1	6636	6738	M1	201799	201901	reverse	103
M1	249899	249994	M1	318600	318695	forward	96
M1	181382	181476	M1	201305	201399	reverse	95
M1	5799	5887	M1	61566	61654	forward	89
M1	19449	19533	M1	417565	417649	forward	85

M1	27979	28063	M1	187661	187745	forward	85
M1	13493	13570	M1	382475	382552	reverse	78
M1	263587	263664	M1	382475	382552	forward	78
M1	62258	62334	M1	102065	102141	reverse	77
M1	96580	96656	M1	216858	216934	forward	77
M1	219682	219755	M1	405594	405667	reverse	74
M1	13901	13973	M1	263185	263257	reverse	73
M1	211600	211672	M1	447677	447749	reverse	73
M1	329221	329291	M1	512329	512399	forward	71
M1	58980	59043	M1	501711	501774	forward	64
M1	10176	10238	M1	40517	40579	forward	63
M1	40517	40579	M1	266919	266981	reverse	63
M1	40517	40579	M1	397036	397098	reverse	63
M1	119004	119066	M1	347367	347429	forward	63
M1	119004	119066	M1	456375	456437	forward	63
M1	7279	7340	M1	57878	57939	reverse	62
M1	57878	57939	M1	269817	269878	forward	62
M1	57878	57939	M1	399934	399995	forward	62
M1	86336	86397	M1	252390	252451	forward	62
M1	146137	146197	M1	227961	228021	forward	61
M1	146137	146197	M1	532472	532532	forward	61
M1	385059	385118	M1	537681	537740	forward	60
M1	539648	539706	M1	548799	548857	forward	59
M1	2425	2481	M1	378590	378646	reverse	57
M1	181287	181343	M1	201461	201517	reverse	57
M1	385001	385057	M1	537624	537680	forward	57
M1	2375	2429	M1	378647	378701	reverse	55
M1	132038	132091	M1	367616	367669	reverse	54
M1	231236	231287	M1	443581	443632	reverse	52
M1	443581	443632	M1	535747	535798	reverse	52
M1	13572	13622	M1	382423	382473	reverse	51
M1	263535	263585	M1	382423	382473	forward	51
M1	146412	146461	M1	332087	332136	forward	50
M1	159351	159400	M1	450956	451005	reverse	50
M1	297661	297710	M1	359632	359681	forward	50
M1	14864	14910	M1	74281	74327	forward	47
M1	14991	15037	M1	74403	74449	forward	47
M1	17478	17524	M1	358924	358970	forward	47
M1	26993	27039	M1	362252	362298	forward	47
M1	339162	339208	M1	526061	526107	reverse	47
M1	28148	28193	M1	62283	62328	forward	46
M1	28148	28193	M1	102071	102116	reverse	46
M1	8291	8335	M1	140921	140965	reverse	45
M1	140921	140965	M1	268822	268866	forward	45
M1	140921	140965	M1	398939	398983	forward	45
M1	170538	170581	M1	467345	467388	reverse	44
M1	25995	26037	M1	313087	313129	reverse	43
M1	26920	26962	M1	362189	362231	forward	43
M1	78322	78364	M1	243920	243962	reverse	43
M1	103307	103349	M1	250619	250661	reverse	43
M1	314972	315014	M1	371393	371435	forward	43
M1	486092	486134	M1	527298	527340	reverse	43
M1	223	264	M1	17896	17937	forward	42
M1	7195	7236	M1	58387	58428	reverse	42
M1	10443	10484	M1	253567	253608	forward	42
M1	58387	58428	M1	269921	269962	forward	42
M1	58387	58428	M1	400038	400079	forward	42
M1	253567	253608	M1	266673	266714	reverse	42
M1	253567	253608	M1	396790	396831	reverse	42
M1	539708	539749	M1	548859	548900	forward	42
M1	28359	28399	M1	329751	329791	forward	41
M1	98556	98596	M1	111075	111115	reverse	41
M1	274262	274302	M1	404290	404330	reverse	41
M1	320097	320137	M1	329751	329791	forward	41
M1	333098	333138	M1	443444	443484	forward	41
M1	349573	349613	M1	382602	382642	forward	41
M1	382602	382642	M1	458581	458621	forward	41
M1	10380	10419	M1	520763	520802	reverse	40
M1	62281	62320	M1	319935	319974	forward	40
M1	101976	102015	M1	203291	203330	forward	40
M1	102079	102118	M1	319935	319974	reverse	40
M1	111302	111341	M1	261925	261964	reverse	40
M1	126852	126891	M1	169408	169447	forward	40
M1	266738	266777	M1	520763	520802	forward	40
M1	396855	396894	M1	520763	520802	forward	40
M1	104536	104574	M1	223833	223871	forward	39
M1	104536	104574	M1	528344	528382	forward	39
M1	159324	159362	M1	435477	435515	forward	39

M1	225829	225867	M1	393931	393969	forward	39
M1	243052	243090	M1	246657	246695	forward	39
M1	275967	276005	M1	481773	481811	reverse	39
M1	315017	315055	M1	371431	371469	forward	39
M1	334685	334723	M1	435476	435514	reverse	39
M1	382203	382241	M1	469828	469866	forward	39
M1	393931	393969	M1	530340	530378	forward	39
M1	447144	447182	M1	520766	520804	reverse	39
M1	10379	10416	M1	72778	72815	reverse	38
M1	10766	10803	M1	80740	80777	reverse	38
M1	19332	19369	M1	417448	417485	forward	38
M1	28148	28185	M1	319937	319974	forward	38
M1	72778	72815	M1	266741	266778	forward	38
M1	72778	72815	M1	396858	396895	forward	38
M1	75042	75079	M1	167467	167504	reverse	38
M1	80740	80777	M1	266354	266391	forward	38
M1	80740	80777	M1	396471	396508	forward	38
M1	159324	159361	M1	334685	334722	reverse	38
M1	180552	180589	M1	549889	549926	reverse	38
M1	187755	187792	M1	434869	434906	forward	38
M1	197619	197656	M1	425970	426007	forward	38
M1	10380	10416	M1	447146	447182	forward	37
M1	50463	50499	M1	351009	351045	reverse	37
M1	50463	50499	M1	460017	460053	reverse	37
M1	72778	72814	M1	447146	447182	reverse	37
M1	72778	72814	M1	520766	520802	forward	37
M1	266741	266777	M1	447146	447182	reverse	37
M1	361477	361513	M1	537745	537781	reverse	37
M1	396858	396894	M1	447146	447182	reverse	37
M1	17480	17515	M1	280322	280357	forward	36
M1	72776	72811	M1	358897	358932	forward	36
M1	165897	165932	M1	251600	251635	reverse	36
M1	280322	280357	M1	358926	358961	forward	36
M1	8859	8893	M1	350857	350891	reverse	35
M1	8859	8893	M1	459865	459899	reverse	35
M1	170465	170499	M1	511209	511243	reverse	35
M1	206354	206388	M1	211069	211103	forward	35
M1	206354	206388	M1	448246	448280	reverse	35
M1	268264	268298	M1	350857	350891	forward	35
M1	268264	268298	M1	459865	459899	forward	35
M1	319773	319807	M1	319782	319816	forward	35
M1	349937	349971	M1	420362	420396	reverse	35
M1	350857	350891	M1	398381	398415	forward	35
M1	398381	398415	M1	459865	459899	forward	35
M1	420362	420396	M1	458945	458979	reverse	35
M1	10383	10416	M1	358899	358932	reverse	34
M1	15085	15118	M1	74504	74537	forward	34
M1	49757	49790	M1	79839	79872	reverse	34
M1	200925	200958	M1	245459	245492	reverse	34
M1	266741	266774	M1	358899	358932	forward	34
M1	358899	358932	M1	396858	396891	forward	34
M1	358899	358932	M1	447149	447182	reverse	34
M1	358899	358932	M1	520766	520799	forward	34
M1	15198	15230	M1	74628	74660	forward	33
M1	23350	23382	M1	115183	115215	forward	33
M1	79105	79137	M1	172684	172716	forward	33
M1	110449	110481	M1	110506	110538	forward	33
M1	166014	166046	M1	253114	253146	reverse	33
M1	224257	224289	M1	486927	486959	reverse	33
M1	257894	257926	M1	446456	446488	forward	33
M1	350539	350571	M1	419754	419786	reverse	33
M1	419754	419786	M1	459547	459579	reverse	33
M1	486927	486959	M1	528768	528800	reverse	33
M1	486928	486960	M1	507990	508022	forward	33
M1	26887	26918	M1	362156	362187	forward	32
M1	50418	50449	M1	83840	83871	forward	32
M1	180376	180407	M1	450904	450935	forward	32
M1	192285	192316	M1	425729	425760	forward	32
M1	224257	224288	M1	507990	508021	reverse	32
M1	225352	225383	M1	486884	486915	reverse	32
M1	257940	257971	M1	446503	446534	forward	32
M1	486884	486915	M1	529863	529894	reverse	32
M1	507990	508021	M1	528768	528799	reverse	32
M1	5767	5797	M1	61534	61564	forward	31
M1	28243	28273	M1	319981	320011	forward	31
M1	103396	103426	M1	224935	224965	reverse	31
M1	103396	103426	M1	529446	529476	reverse	31
M1	212084	212114	M1	351924	351954	forward	31

M1	212084	212114	M1	460932	460962	forward	31
M1	10890	10919	M1	80670	80699	reverse	30
M1	80670	80699	M1	266238	266267	forward	30
M1	80670	80699	M1	396355	396384	forward	30
M1	86444	86473	M1	296804	296833	forward	30
M1	86444	86473	M1	545262	545291	forward	30
M1	223896	223925	M1	223937	223966	forward	30
M1	223896	223925	M1	528448	528477	forward	30
M1	223937	223966	M1	528407	528436	forward	30
M1	528407	528436	M1	528448	528477	forward	30

Repeats M1-M2

M1	296546	296994	M2	7381	7829	forward	449
M1	86034	86399	M2	4204	4569	forward	366
M1	319937	320186	M2	140087	140336	reverse	250
M1	545119	545345	M2	7496	7722	forward	227
M1	135150	135333	M2	139664	139847	forward	184
M1	12500	12652	M2	98932	99084	forward	153
M1	264505	264657	M2	98932	99084	reverse	153
M1	28282	28423	M2	140112	140253	reverse	142
M1	6473	6611	M2	11398	11536	reverse	139
M1	536802	536934	M2	70395	70527	forward	133
M1	60111	60220	M2	75574	75683	reverse	110
M1	537573	537680	M2	12065	12172	forward	108
M1	28085	28185	M2	140299	140399	reverse	101
M1	385001	385100	M2	12116	12215	forward	100
M1	424717	424814	M2	70262	70359	reverse	98
M1	320185	320281	M2	139993	140089	reverse	97
M1	175927	176021	M2	95742	95836	forward	95
M1	434924	435015	M2	80885	80976	forward	92
M1	135066	135148	M2	139580	139662	forward	83
M1	85906	85984	M2	2423	2501	forward	79
M1	187755	187832	M2	80836	80913	forward	78
M1	273060	273135	M2	97996	98071	reverse	76
M1	403177	403252	M2	97996	98071	reverse	76
M1	58676	58750	M2	143270	143344	forward	75
M1	135335	135406	M2	139849	139920	forward	72
M1	243052	243123	M2	77923	77994	forward	72
M1	3552	3616	M2	104319	104383	forward	65
M1	201996	202060	M2	10271	10335	forward	65
M1	95864	95925	M2	114015	114076	reverse	62
M1	252390	252451	M2	4506	4567	forward	62
M1	67367	67427	M2	11047	11107	reverse	61
M1	118449	118507	M2	134458	134516	forward	59
M1	202521	202579	M2	10541	10599	forward	59
M1	212113	212170	M2	3185	3242	reverse	58
M1	494253	494310	M2	38023	38080	forward	58
M1	328938	328994	M2	73650	73706	reverse	57
M1	411694	411749	M2	82473	82528	forward	56
M1	145001	145055	M2	15371	15425	reverse	55
M1	320294	320346	M2	139928	139980	reverse	53
M1	202407	202458	M2	10422	10473	forward	52
M1	158392	158441	M2	6029	6078	reverse	50
M1	202066	202115	M2	10341	10390	forward	50
M1	434857	434906	M2	80824	80873	forward	50
M1	6554	6601	M2	12191	12238	reverse	48
M1	538595	538642	M2	28625	28672	reverse	48
M1	111411	111457	M2	74357	74403	reverse	47
M1	385025	385071	M2	11357	11403	forward	47
M1	176536	176581	M2	87529	87574	forward	46
M1	362466	362511	M2	41283	41328	forward	46
M1	24554	24598	M2	156626	156670	forward	45
M1	118008	118051	M2	138095	138138	reverse	44
M1	145057	145100	M2	15326	15369	reverse	44
M1	463144	463187	M2	76217	76260	reverse	44
M1	3668	3709	M2	103498	103539	reverse	42
M1	6426	6467	M2	141272	141313	reverse	42
M1	405083	405124	M2	67074	67115	forward	42
M1	537681	537722	M2	12174	12215	forward	42
M1	246655	246695	M2	77921	77961	forward	41
M1	329751	329791	M2	140136	140176	reverse	41
M1	62214	62252	M2	56158	56196	forward	39
M1	506495	506533	M2	338	376	reverse	39
M1	62283	62320	M2	140299	140336	reverse	38
M1	102079	102116	M2	140299	140336	forward	38
M1	405245	405282	M2	67244	67281	forward	38
M1	536750	536787	M2	70344	70381	forward	38

M1	463697	463733	M2	75668	75704	reverse	37
M1	158355	158390	M2	6080	6115	reverse	36
M1	223695	223730	M2	153239	153274	reverse	36
M1	280325	280360	M2	114181	114216	reverse	36
M1	405201	405236	M2	67200	67235	forward	36
M1	528206	528241	M2	153239	153274	reverse	36
M1	58758	58792	M2	143354	143388	forward	35
M1	83483	83517	M2	140447	140481	reverse	35
M1	186364	186398	M2	82536	82570	reverse	35
M1	193827	193861	M2	7835	7869	forward	35
M1	426356	426390	M2	84363	84397	forward	35
M1	59259	59292	M2	76501	76534	reverse	34
M1	515807	515840	M2	3637	3670	reverse	34
M1	538562	538595	M2	28671	28704	reverse	34
M1	7917	7949	M2	26709	26741	reverse	33
M1	17483	17515	M2	114184	114216	reverse	33
M1	60221	60253	M2	75543	75575	reverse	33
M1	75042	75074	M2	52842	52874	reverse	33
M1	167472	167504	M2	52842	52874	forward	33
M1	215038	215070	M2	114602	114634	forward	33
M1	269208	269240	M2	26709	26741	forward	33
M1	358929	358961	M2	114184	114216	reverse	33
M1	399325	399357	M2	26709	26741	forward	33
M1	537648	537680	M2	11357	11389	forward	33
M1	75041	75072	M2	33203	33234	forward	32
M1	117790	117821	M2	10871	10902	forward	32
M1	243019	243050	M2	77890	77921	forward	32
M1	475999	476030	M2	61604	61635	forward	32
M1	552936	552967	M2	154288	154319	reverse	32
M1	17705	17735	M2	116566	116596	forward	31
M1	28243	28273	M2	140262	140292	reverse	31
M1	167474	167504	M2	33204	33234	reverse	31
M1	86444	86473	M2	7639	7668	forward	30
M1	293455	293484	M2	111434	111463	forward	30
M1	381099	381128	M2	5417	5446	reverse	30

Repeats M2-M2

M2	157915	158155	M2	158212	158452	forward	241
M2	70666	70731	M2	70795	70860	forward	66
M2	11408	11455	M2	12191	12238	forward	48
M2	54037	54084	M2	145435	145482	reverse	48
M2	11357	11403	M2	12140	12186	forward	47
M2	107621	107661	M2	142058	142098	reverse	41
M2	138041	138079	M2	151589	151627	reverse	39
M2	33204	33234	M2	52844	52874	reverse	31
M2	158155	158185	M2	158450	158480	forward	31

Table S8: Tandem repeats in *Ludwigia grandiflora* subsp. *hexapetala* and *Ludwigia peploides* subsp. *montevidensis* mitogenomes.

Coordinates	Period size	Copy number	Percent matches
Ludwigia grandiflora* subsp. *hexapetala			
Molecule M2			
146,440-146,485	22	2.1	100
Molecule M1			
18,305-18,347	18	2.4	81
44,204-44,240	17	2.2	90
44,754-44,787	17	2.0	88
62,384-62,413	15	2.0	100
96,281-96,313	17	1.9	93
127,742-127,424	18	2.4	81
154,823-154,847	12	2.1	100
170,080-170,113	15	2.3	94
171,929-172,009	41	2.0	90
225,692-225,734	11	3.9	93
251,276-251,316	21	1.9	90
287,158-287,206	21	3.9	70
287,208-287,249	18	2.0	100
287,208-287,249	21	2.0	95
296,977-297,053	24	3.5	74
296,993-297,033	18	2.3	95
314,519-314,556	16	2.4	91
332,357-332,406	25	2.0	100
382,053-382,089	19	1.9	88
387,587-387,629	18	2.4	81
438,208-438,268	9	6.8	94
482,654-482,687	16	2.1	94
514,004-514,082	39	2.0	95
519,531-519,564	15	2.3	94
Ludwigia peploides* subsp. *montevidensis			
Molecule M2			
146,342-146,387	22	2.1	100
157,911-158,505	297	2.0	98
Molecule M1			
8,683-8,725	18	2.4	81
14,223-14,259	19	1.9	88
63,874-63,923	25	2.0	100
81,700-81,737	16	2.4	91
100,869-100,932	18	3.6	91
144,651-144,691	21	1.9	90
170,230-170,272	11	3.9	93
225,782-225,815	15	2.3	94
241,032-241,056	12	2.1	100
268,432-268,474	18	2.4	81
319,773-319,833	9	6.8	98
372,671-372,707	17	2.2	90
398,549-398,591	18	2.4	81
493,434-493,467	16	2.1	94
501,801-501,830	15	2.0	100
524,767-524,845	39	2.0	95
530,293-530,326	15	2.3	94

Table S9: The relative synonymous codon usage (RSCU) bias in *Ludwigia grandiflora* subsp. *hexapetala* and *Ludwigia peploides* subsp. *montevidensis* mitogenomes CDS (Sequences of protein coding genes).

Ludwigia grandiflora* subsp. *hexapetala			
Codon	**Aminoacid**	**Count**	**RSCU**
GCA	Ala	200	1.09
GCC	Ala	169	0.92
GCG	Ala	79	0.43
GCU	Ala	288	1.57
AGA	Arg	204	1.47
AGG	Arg	111	0.8
CGA	Arg	189	1.36
CGC	Arg	72	0.52
CGG	Arg	99	0.71
CGU	Arg	160	1.15
AAC	Asn	122	0.66
AAU	Asn	247	1.34
GAC	Asp	110	0.6
GAU	Asp	256	1.4
UGC	Cys	76	0.86
UGU	Cys	100	1.14
UAA	End	23	1.64
UAG	End	4	0.29
UGA	End	15	1.07
CAA	Gln	235	1.52
CAG	Gln	75	0.48
GAA	Glu	332	1.35
GAG	Glu	160	0.65
GGA	Gly	307	1.51
GGC	Gly	108	0.53
GGG	Gly	144	0.71
GGU	Gly	256	1.26
CAC	His	68	0.48
CAU	His	216	1.52
AUA	Ile	246	0.84
AUC	Ile	252	0.86
AUU	Ile	385	1.31
CUA	Leu	176	0.87
CUC	Leu	131	0.64
CUG	Leu	120	0.59
CUU	Leu	266	1.31
UUA	Leu	275	1.35
UUG	Leu	251	1.24
AAA	Lys	325	1.18
AAG	Lys	225	0.82
AUG	Met	302	1
UUC	Phe	311	0.86
UUU	Phe	410	1.14
CCA	Pro	167	1.09
CCC	Pro	121	0.79
CCG	Pro	99	0.65
CCU	Pro	224	1.47
AGC	Ser	110	0.61
AGU	Ser	186	1.04
UCA	Ser	209	1.17
UCC	Ser	178	0.99
UCG	Ser	148	0.83
UCU	Ser	245	1.37
ACA	Thr	140	0.97
ACC	Thr	141	0.98
ACG	Thr	84	0.58
ACU	Thr	213	1.47
UGG	Trp	174	1
UAC	Tyr	82	0.48
UAU	Tyr	261	1.52
GUA	Val	215	1.21
GUC	Val	133	0.75
GUG	Val	155	0.87
GUU	Val	210	1.18

Ludwigia peploides* subsp. *montevidensis			
Codon	**Aminoacid**	**Count**	**RSCU**
GCA	Ala	219	1.13
GCC	Ala	176	0.9
GCG	Ala	81	0.42
GCU	Ala	302	1.55
AGA	Arg	212	1.46
AGG	Arg	117	0.81
CGA	Arg	196	1.35
CGC	Arg	71	0.49
CGG	Arg	105	0.72
CGU	Arg	169	1.17
AAC	Asn	127	0.64
AAU	Asn	272	1.36
GAC	Asp	117	0.6
GAU	Asp	273	1.4
UGC	Cys	77	0.84
UGU	Cys	107	1.16
UAA	End	24	1.6
UAG	End	7	0.47
UGA	End	14	0.93
CAA	Gln	248	1.51
CAG	Gln	81	0.49
GAA	Glu	360	1.37
GAG	Glu	166	0.63
GGA	Gly	321	1.51
GGC	Gly	110	0.52
GGG	Gly	150	0.71
GGU	Gly	268	1.26
CAC	His	70	0.48
CAU	His	223	1.52
AUA	Ile	257	0.83
AUC	Ile	263	0.85
AUU	Ile	410	1.32
CUA	Leu	181	0.85
CUC	Leu	133	0.62
CUG	Leu	123	0.58
CUU	Leu	280	1.31
UUA	Leu	301	1.41
UUG	Leu	263	1.23
AAA	Lys	343	1.19
AAG	Lys	234	0.81
AUG	Met	316	1
UUC	Phe	318	0.85
UUU	Phe	428	1.15
CCA	Pro	175	1.12
CCC	Pro	122	0.78
CCG	Pro	101	0.65
CCU	Pro	228	1.46
AGC	Ser	112	0.61
AGU	Ser	193	1.05
UCA	Ser	214	1.16
UCC	Ser	183	0.99
UCG	Ser	151	0.82
UCU	Ser	251	1.36
ACA	Thr	148	0.97
ACC	Thr	147	0.97
ACG	Thr	91	0.6
ACU	Thr	222	1.46
UGG	Trp	187	1
UAC	Tyr	83	0.46
UAU	Tyr	277	1.54
GUA	Val	224	1.21
GUC	Val	133	0.72
GUG	Val	160	0.86
GUU	Val	226	1.22

Table S10: Pairwise comparisons of non-synonymous substitution rate (Ka) and synonymous substitution rate (Ks) ratios of Myrtales and Geraniales order species with *Ludwigia grandiflora* subsp. *hexapetala* and *Ludwigia peploides* subsp. *montevidensis* for 31 mitochondrial genes. Genes present in at least 14 of the 16 selected species were studied. Values in blue are >1.

Genes	*O. villaricae*	*O. biennis*	*O. elata*	*C. angustifolium*	*L. indica*	*P. granatum*	*M. candidum*	*M. sanguineum*	*M. dodecandrum*	*S. samarangense*	*R. tomentosa*	*E. urophylla X grandis*	*E. grandis*	*G. maderense*
nad1	0.15	0.15	0.15	0.15	0.19	0.17	0.29	0.29	0.29	0.45	0.55	0.45	0.45	0.33
nad2	0.22	0.22	0.22	0.16	0.45	0.37	0.26	0.26	0.27	0.28	0.24	0.19	0.25	0.51
nad3	1.59	1.59	1.59	0.95	0.63	0.79	0.87	0.87	0.81	0.18	NA	0.24	0.18	0.36
nad4	0.28	0.28	0.28	0.33	0.15	0.15	0.15	0.15	0.17	0.27	0.27	0.3	0.26	0.79
nad4L	0.21	0.21	0.21	0.32	0.24	0.24	0.15	0.15	0.15	0.15	0.15	NA	0.15	0.17
nad5	0.29	0.29	0.29	0.35	0.3	0.24	0.19	0.19	0.23	0.36	0.32	0.32	0.35	0.44
nad6	0.09	0.11	0.09	0.11	0.17	0.17	0.18	0.18	0.17	0.09	0.09	0.09	0.09	0.37
nad7	0.09	0.09	0.09	0.13	0.24	0.1	0.31	0.31	0.31	0.16	NA	0.27	0.37	0.41
nad9	0.3	0.3	0.3	0.45	0.45	0.45	0.5	0.5	0.5	0.45	0.3	0.3	1.26	0.48
sdh4	0.48	0.48	0.48	0.51	0.45	0.46	0.44	0.44	0.44	0.5	0.5	0.47	0.47	NA
cob	0.24	0.24	0.24	0.25	0.14	0.13	0.2	0.2	0.2	0.22	0.18	0.2	0.22	0.23
cox1	0.17	0.18	0.18	0.21	0.19	0.2	0.28	0.28	0.28	0.23	0.21	0.22	0.23	0.16
cox2	0.28	0.31	0.31	0.34	0.3	0.25	0.3	0.3	0.3	0.34	NA	0.37	0.3	0.31
cox3	0.06	0.06	0.06	0.08	0.37	0.23	0.2	0.2	0.2	0.26	0.26	0.29	0.23	0.27
atp1	0.12	0.12	0.12	0.12	0.16	0.13	0.28	0.28	0.28	0.13	NA	0.13	0.12	0.2
atp4	0.46	0.46	0.46	0.49	0.44	0.55	0.36	0.36	0.36	0.47	0.48	0.52	0.48	0.51
atp6	0.26	0.28	0.28	0.28	0.28	0.35	0.45	0.45	0.45	0.42	0.45	0.46	0.46	0.25
atp8	0.13	0.07	0.07	0.28	0.46	0.46	0.13	0.13	0.15	0.15	0.33	0.22	0.22	0.45
atp9	0.04	0.04	0.04	0.04	0.33	0.33	0.1	0.1	0.1	0.18	NA	0.18	0.18	0.05
ccmB	0.38	0.38	0.38	1.01	1.45	1.2	0.42	0.42	0.42	0.5	1.01	1.01	1.01	1.04
ccmC	0.55	0.55	0.55	0.55	0.67	0.67	0.27	0.27	0.27	0.47	0.47	0.47	0.47	1.03
ccmFc	0.69	0.79	0.79	0.65	0.89	0.67	0.8	0.8	0.75	1.02	0.86	1	1	0.82
matR	0.75	0.76	0.76	0.76	0.5	0.46	0.74	0.74	0.76	0.69	NA	0.78	0.78	0.76
mttB	1.01	0.92	0.92	1.7	0.62	0.58	0.74	0.74	0.74	1.09	NA	1.16	1.16	0.68
rpl2	0.47	0.51	0.51	0.32	0.63	0.64	0.44	0.44	0.46	0.68	0.78	0.65	0.72	NA
rpl5	0.21	0.24	0.24	0.79	0.38	0.57	0.75	0.75	0.63	0.17	0.28	0.17	0.2	NA
rpl10	0.49	0.64	0.64	0.64	0.49	0.56	1.22	1.22	1.07	0.48	0.46	0.55	0.55	0.98
rps1	2.26	2.26	2.26	1.09	0.98	0.98	1.66	1.66	1.66	2.59	NA	2.61	2.61	1.53
rps3	1.32	1.32	1.32	0.83	0.75	0.57	0.59	0.59	0.59	0.59	NA	0.55	0.57	0.65
rps4	0.42	0.41	0.41	0.41	0.58	0.63	0.52	0.52	0.5	0.49	0.62	0.52	0.52	NA
rps14	0.22	0.22	0.22	0.11	0.33	0.33	0.32	0.32	0.32	0.22	0.22	0.22	0.22	NA

Chapitre 2

Assemblage du génome nucléaire de *Lgh*

1 Introduction

1.1 Génomique des invasions biologiques

Le rythme des invasions biologiques s'accélère sous l'effet des activités humaines et les aires de répartition s'accroissent sous l'effet du changement climatique. Après des décennies de recherche, il est toujours difficile de comprendre comment les espèces exotiques envahissantes parviennent à coloniser de nouveaux habitats malgré leur exposition à des défis environnementaux et écologiques potentiellement nouveaux [Novoa et al., 2020]. La plupart des travaux de ces trente dernières années se sont intéressés presque exclusivement à des questions d'écologie en focalisant sur l'importance des caractéristiques phénotypiques (capacité à se disperser, reproduction sexuée, plasticité phénotypique...) [Hodgins, Battlay et Bock, 2025]. De nombreuses hypothèses sur les mécanismes du potentiel invasif des plantes envahissantes ont été avancées dont la duplication du génome. Ces changements génomiques à grande échelle permettent aux organismes de s'adapter à des environnements nouveaux et pourraient donc favoriser les invasions [Beest et al., 2012]. L'utilisation des approches génomiques montre un lien étroit entre la variation génomique et le succès des invasions. Elle devrait éclairer notre compréhension des types et des rythmes d'adaptation lors des invasions biologiques [McGaughran et al., 2024]. Au cours du processus d'invasion, une multitude de facteurs entre en jeu dont le mode de reproduction et la polyploïdie.

La clonalité est une caractéristique surreprésentée chez les plantes invasives puisque 81% des plantes invasives figurant sur la liste des 100 espèces envahissantes les plus nuisibles (30/37) sont capables de se reproduire par clonage [Mounger et al., 2021]. La reproduction clonale est avantageuse lors de la colonisation en raison de la garantie de reproduction qu'elle offre [Baker, 1955]. Pourtant, même chez ces espèces, il existe des preuves d'adaptation. Ainsi des signatures génomiques d'adaptation ont été observées

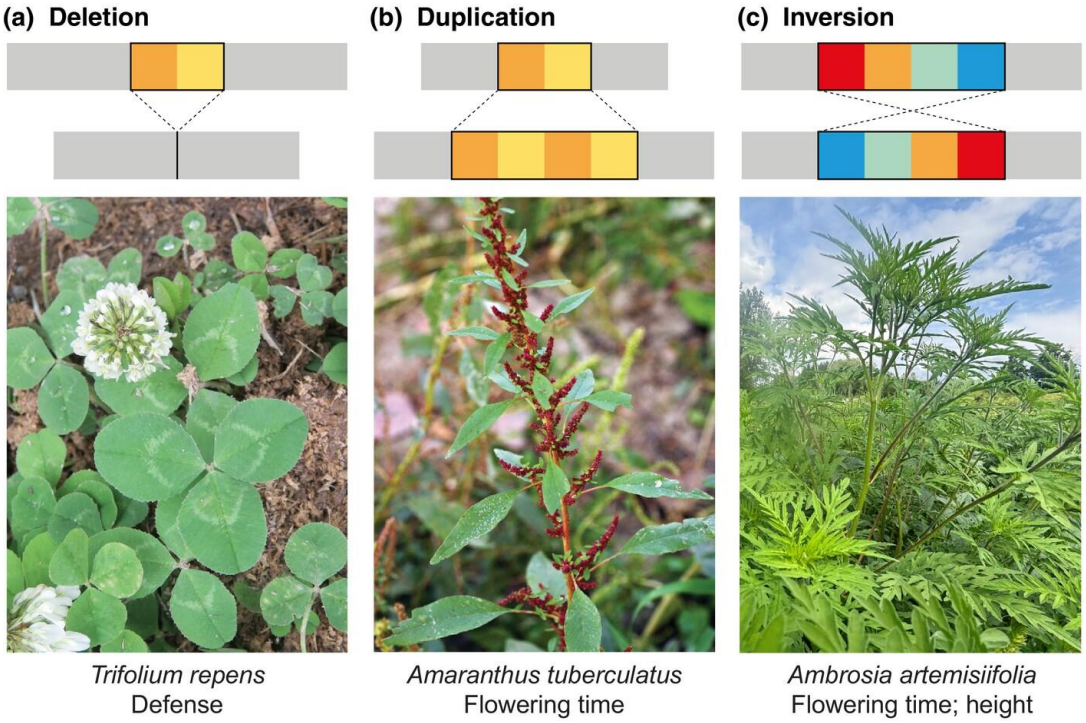


FIGURE 2.1 – Formes courantes de variation structurelle et exemples de leur implication dans les traits adaptatifs chez les espèces végétales envahissantes [Hodgins, Battlay et Bock, 2025]. (a) La délétion de gènes impliqués dans la voie de défense par cyanogénèse est fréquente dans les populations de *Trifolium repens* [Santangelo et al., 2022]. (b) Chez *Amaranthus tuberculatus*, les duplications d'un gène de l'ATP synthase, qui ont donné lieu à pas moins de 14 copies, expliquent environ 20% de la variation du temps de floraison [Kreiner et al., 2023]. (c) Les inversions chez *Ambrosia artemisiifolia* sont porteuses de variants génétiques affectant des traits clés d'adaptation locale tels que le moment de la floraison [Battlay, Wilson et al., 2023].

non seulement entre les aires de répartition natives et envahies à l'échelle locale, mais aussi entre différentes invasions à l'échelle mondiale [Battlay, Craig et al., 2025]. Chez l'ambroisie, *Ambrosia artemisiifolia*, une variabilité du nombre de copies (changements structuraux incluant la duplication et la délétion de gènes ainsi que la variation de l'abondance d'ET) présentant des signatures d'adaptation locale dans les aires de répartition natives et introduites a été identifié à l'aide d'une approche de génomique des populations [Wilson et al., 2025]. Ces résultats soulignent l'importance d'une forme souvent négligée de variation génomique pour l'adaptation locale et l'adaptation rapide des espèces envahissantes.

La polyploïdie est souvent évoquée comme un facteur influençant le potentiel d'invasion des espèces invasives. En effet, elle est courante chez les plantes exotiques envahissantes [Beest et al., 2012]. La polyploïdisation augmente souvent la variabilité génétique, de sorte que certaines espèces polyploïdes se sont avérées plus aptes à coloniser de nouveaux habitats que leurs parents diploïdes [DeWalt et Hamrick, 2004 ; Cadotte, McMahon et Fukami, 2006]. En 2023, Pyšek et al. ont réalisé une méta-analyse chez les plantes, compilant des informations sur la taille du génome et la ploïdie, ainsi que sur la naturalisation et le succès d'invasion de 11 049 espèces [Pyšek, Lučanová et al., 2023]. Leurs analyses ont révélé que les plantes ayant un génome plus grand et un niveau de ploïdie plus élevé sont envahissantes dans un plus grand nombre de régions géographiques. Le développement de la génomique a permis d'identifier des mécanismes par lesquels la polyploïdisation peut favoriser les invasions (Figure 2.1). Grâce à l'identification chez la centaurée, *Centaurea solstitialis*, de régions du génome ayant une forte influence sur les traits qui diffèrent entre les populations indigènes et envahissantes, il a été possible de montrer que des variants structurels du génome et des duplications génomiques passées pourraient jouer un rôle dans l'adaptation rapide des populations envahissantes [Reatini et al., 2024]. Par ailleurs, chez les espèces du genre *Cochlearia*, les

génomes polyploïdes possèdent une proportion de variants structurels plus importants que les génomes diploïdes [Hämälä et al., 2024]. La plupart de ces variants semblaient être délétères et masqués par la sélection en raison des copies alléliques supplémentaires présentes chez les polyploïdes. Lors de l'invasion, ces variants structurels délétères qui s'accumulent dans les génomes polyploïdes pourraient être recrutés par sélection positive en réponse au changement environnemental [Hodgins, Battlay et Bock, 2025]. Les ressources génomiques dans l'étude d'espèces invasives peuvent donc avoir un rôle central.

1.2 Ressources génomiques dans l'ordre des Myrtales

Pour *Ludwigia grandiflora* subsp. *hexapetala*, l'assemblage des génomes chloroplastique [Barloy-Hubler et al., 2025] et mitochondrial [Doré, Barloy et Barloy-Hubler, 2024] ont permis de compléter les ressources génomiques pour cette espèce. Cependant, le génome nucléaire est toujours manquant. L'assemblage de ce dernier risque d'être complexe à cause de sa taille importante (1,419 Gb) ainsi que de sa décaploïdie et son auto-allopolyploïdie [Barloy et al., 2024]. En effet, la redondance génétique des génomes polyploïdes ainsi que la présence de chromosomes homéologues rendent difficile un assemblage complet et sans erreur [Sun, Shang et al., 2022]. De plus, il y a un manque de ressources génomiques chez les espèces proches de *Lgh*.

L'ordre des Myrtales contient 9 familles comprenant plus de 13 000 espèces [The Angiosperm Phylogeny Group et al., 2016]. Il comprend des espèces très variées pouvant aller d'arbres à plantes herbacées annuelles, souvent terrestres mais parfois aquatiques [Dahlgren et Thorne, 1984]. L'aire de répartition des espèces de l'ordre des Myrtales s'étend dans une grande partie du monde, principalement dans les régions tropicales et subtropicales [Maurin et al., 2021]. Cinq familles contiennent un nombre plus important d'espèces : les Onagraceae, les Lythraceae, les Melastomataceae, les Combretaceae et

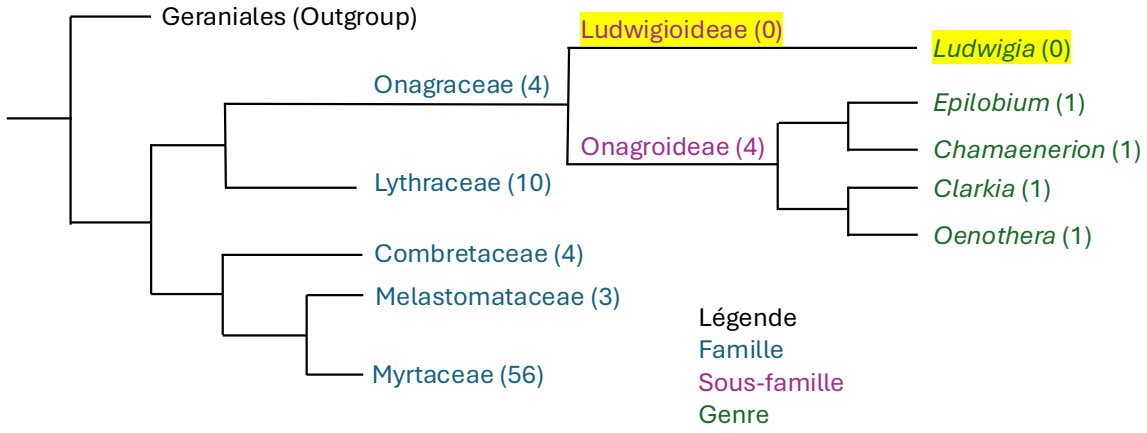


FIGURE 2.2 – Nombre de génomes assemblés dans l'ordre des Myrtales dans la base de données GenBank en juin 2025. La phylogénie est basée sur Wagner, Hoch et Raven, 2007 et Maurin et al., 2021. L'arbre ne contient que les familles, sous-familles ou genre avec au moins 1 génome assemblé sauf ceux surlignés qui correspondent aux classifications de *Lgh*.

TABLE 2.1 – Statistiques des assemblages des génomes des espèces de la famille des Onagraceae dans la base de données Genbank en juin 2025.

Espèce	**Taille du génome (en Mb)**	**Niveau de l'assemblage**	**Nombre de chromosomes ou scaffolds assemblés**
Chamaenerion angustifolium	655,9	Chromosomes	18 (+ 103 scaffolds non placés)
Clarkia xantiana subsp. *parviflora*	695	Scaffolds	58
Epilobium hirsutum	336	Chromosomes	18 (+ 257 scaffolds non placés)
Oenothera biennis	312,4	Scaffolds	276 660

les Myrtaceae [Maurin et al., 2021]. La phylogénie au sein de l'ordre des Myrtales n'est pas encore complétement résolue. Le placement de la famille des Combretaceae est encore débattu tandis que des liens de proximité sont bien établis entre les Onagraceae et Lythraceae d'un côté et les Melastomataceae et Myrtaceae d'un autre [Wang, Li, Yang et al., 2025].

Un des freins dans la résolution de la phylogénie dans l'ordre des Myrtales est le manque de ressources génomiques. En effet, on trouve seulement 77 génomes assemblés, dont 56 au niveau des chromosomes et 21 au niveau de scaffolds, dans l'ordre des Myrtales dans la base de données Genbank en juin 2025 (Figure 2.2) [O'Leary et al., 2024]. Ces 77 génomes assemblés sont répartis inégalement entre les familles. On en retrouve une grande majorité chez les Myrtaceae (72,7% ; Figure 2.2) surtout dans le genre *Eucalyptus* concentrant, à lui seul, 44% (34) des génomes assemblés dans l'ordre des Myrtales. Dans les 4 autres plus grandes familles, on trouve entre 3 et 10 espèces avec un génome assemblé (Figure 2.2).

La famille des Onagraceae (famille de *Lgh*) compte seulement 4 espèces avec des génomes nucléaires assemblés (Figure 2.2). Ces espèces appartiennent à la sous-famille des Onagroideae qui contient toutes les espèces de la famille des Onagraceae à l'exception de celles appartenant au genre *Ludwigia* [Wagner, Hoch et Raven, 2007]. Les 4 espèces dont le génome a été assemblé sont ceux de deux épilobes, *Chamaenerion angustifolium* (GCA_946814005.1) et *Epilobium hirsutum* (GCA_965153245.1), d'une onagre, *Oenothera biennis* (GCA_025802395.1) et d'une clarkia *Clarka xantiana* subsp. *parviflora* (GCA_036873135.1) (Table 2.1). Elles sont toutes diploïdes et les tailles de leur génome varient entre 312,4 Mb pour *Oenothera biennis* et 695 Mb pour *Clarkia xantiana* subsp. *parviflora*. Cependant, seuls les génomes de *Chamaenerion angustifolium* et de *Epilobium hirsutum* ont été assemblés en chromosomes tandis que les deux autres assemblages ne contiennent que des scaffolds. L'assemblage du génome d'*Oenothera*

biennis est d'ailleurs très fragmenté avec 276 660 scaffolds, ce qui le rend difficilement exploitable pour des comparaisons structurelles ou du scaffolding. Les génomes des autres espèces de la famille des Onagraceae sont plus complets mais les espèces restent très éloignées de *Lgh* au vu de la situation d'isolement phylogénétique de la sous-famille des Ludwigioideae. Le déficit de ressources génomiques dans la famille des Onagraceae (seulement 4 espèces avec un génome assemblé sur 660 espèces au total) est donc très marqué.

2 Synthèse de l'article "From Water to land, introducing the first decaploid genome in Ludwigioideae : *Ludwigia grandiflora* subsp. *hexapetala*, an invasive amphibious plant"

2.1 Objectifs et stratégies

L'objectif de cette étude est l'assemblage du génome nucléaire de *Lgh*. Les gènes et les éléments répétés ont été annotés. Enfin, des analyses d'orthologie et de synténie avec des espèces proches ont été réalisées.

L'ADN de *Lgh* a été extrait puis séquencé avec les technologies MiSeq d'Illumina (short reads) et GridION d'Oxford Nanopore (long reads). Les longs reads ont été corrigés grâce aux short reads. Trois assembleurs long reads différents (Flye, Canu et wtdbg2) ont été utilisés ainsi qu'un assembleur hybride (SPAdes) utilisant les long et short reads. Les reads et assemblages correspondants au plastome et au mitogénome ont été retirés dans un premier temps. Ensuite, les résultats des différents assembleurs ont été combinés. Après ça, deux stratégies de biocuration ont été utilisées pour essayer

d'assembler entre eux les contigs restants :

(1) L'annotation puis l'utilisation des CDS

(2) La réintroduction des long reads et des assemblages appartenant au plastome et au mitogénome

Les contigs ainsi assemblés ont été décontaminés en plusieurs étapes successives. Tout d'abord en utilisant BLASTp sur les CDS, puis différents outils directement sur les contigs (kraken2, BLASTn et FCS GX, notammment). Après la décontamination, les génomes des deux espèces phylogénétiquement les plus proches (Sous-famille des Onagroideae), *Chamaenerion angustifolium*, *Epilobium hirsutum* et celui du grenadier, *Punica granatum*, espèces appartenant à la famille des Lythraceae, ont été utilisés comme référence pour le scaffolding des contigs. Les ARNr, ARNt, éléments transposables et répétitions ont été annotés.

2.2 Principaux résultats

L'assemblage final est composé de 111 219 contigs pour une taille totale d'environ 1,487 Gb. Sa complétude a été évalué avec BUSCO montrant un score élevé de 96,3%. Plus de 246 000 gènes codant des protéines ont été annotés dans le génome de *Lgh*. Un peu plus de la moitié des gènes codant des protéines n'ont pas reçu d'annotation fonctionnelle. Le génome de *Lgh* contient également 15 834 gènes d'ARNr et 8417 gènes d'ARNt. La présence de nombreux ARNr suggère la présence de beaucoup de régions répétées ce qui a compliqué l'assemblage. Les éléments transposables composent 8.5% du génome. Ce taux est assez bas, ce qui pourrait être dû à des problèmes d'assemblage. La comparaison des gènes orthologues entre *Lgh* et des espèces du clade des Malvidées a révélé la présence de beaucoup de gènes spécifiques à *Lgh*. Ces gènes pourraient être impliqués dans le caractère invasif de la jussie.

2.3 Conclusion

L'assemblage du génome de *Lgh* est le premier dans la sous-famille des Ludwigioideae et le cinquième dans la famille des Onagraceae. Il vient également compléter les ressources génomiques au sein de l'ordre des Myrtales ce qui permettra des analyses évolutives et phylogénétiques.

3 Article "From Water to land, introducing the first decaploid genome in Ludwigioideae : *Ludwigia grandiflora* subsp. *hexapetala*, an invasive amphibious plant"

GigaByte, 2020, 1–13

doi: xx.xxxx/xxxx
Manuscript in Preparation
Data release

DATA RELEASE

From Water to land, introducing the first decaploid genome in Ludwigioideae: *Ludwigia grandiflora* subsp. *hexapetala*, an invasive amphibious plant

Guillaume Doré[1], Dominique Barloy[1,*,†] and Frédérique Barloy-Hubler[2,*,†]

[1]DECOD (Ecosystem Dynamics and Sustainability), Institut Agro, INRAE, IFREMER, 35042 Rennes, France and
[2]UMR 6553 ECOBIO, CNRS, Université de Rennes 1, 35042 Rennes, France

[*]dominique.barloy@agrocampus-ouest.fr; fhubler@univ-rennes1.fr
[†]Contributed equally.

Abstract

Invasive species are one of the biggest drivers of species extinction. *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) is widely invasive in aquatic ecosystems of Europe, North America, and Japan, and can also establish on emergent freshwater soils. The lack of genomic data for this species, and its relatives, limits our ability to explore the genetic basis of its invasiveness. Here, we present a draft genome assembly of the decaploid species, *Lgh* (2n = 10x = 80), based on Illumina Miseq, and GridION Oxford nanopore sequencing. The assembled genome is 1.487 Gb long and contains 111,219 contigs. A total of 246,242 protein-coding genes were annotated. The presence of numerous large repetitive sequences, particularly rDNA clusters and transposable elements, made the assembly process complicated, and contributed to the observed genome fragmentation. This first genome of a Ludwigioideae provides a valuable reference for functional genomics, phylogenetic reconstruction, and evolutionary analyses across the Onagraceae family.

Classifications: Genetics and Genomics; Plants Genetics; Botany

Introduction

Biological invasions are one of the largest threats to biodiversity [1]. Invasive aquatic plants have particularly severe ecological and economic impacts [2]. *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*), commonly known as the water primrose, is one of the most invasive plants in the world (Figure 1A). Native to South America, *Lgh* was first introduced into France around 1830, likely for ornamental purposes [3]. Since then, *Lgh* has spread rapidly across Europe (Figure 1D) and is now also invading regions in Japan and North America [4, 5, 6]. In France, dense mats of *Lgh* (Figure 1C) severely disrupt freshwater ecosystems, causing damage to local biodiversity by obstructing waterways, and interfering with fishing, navigation, and agriculture [7]. Moreover, this amphibious invasive plant has developed the ability to colonize terrestrial environments within its introduction range (Figure 1C) [8]. *Lgh* belongs to the Onagraceae family within the Myrtales order [10]. The Onagraceae family comprises 664 species, and is divided into two subfamilies: Ludwigioideae, represented exclusively by the genus Ludwigia (82 species), and Onagroideae, which is further subdivided into six tribes: Epilobieae, Onagreae, Fuchsiaeeae, Clarkieae, Circaeeae, and Oenothereae [10]. *Lgh* is decaploid (2n = 10x = 80 chromosomes), having arisen through multiple allopolyploidization events (Figure 1B) [9]. Cytogenetic studies revealed that *Lgh* is an auto-allopolyploid species. Its autotetraploid donor (4x) is the diploid species, *Ludwigia peploides* subsp. *montevidensis* (2n = 2x = 16 chromosomes). *Ludwigia helminthorrhiza* (2n = 2x = 16 chromosomes) was also

Compiled on: August 14, 2025.
Draft manuscript prepared by the author.

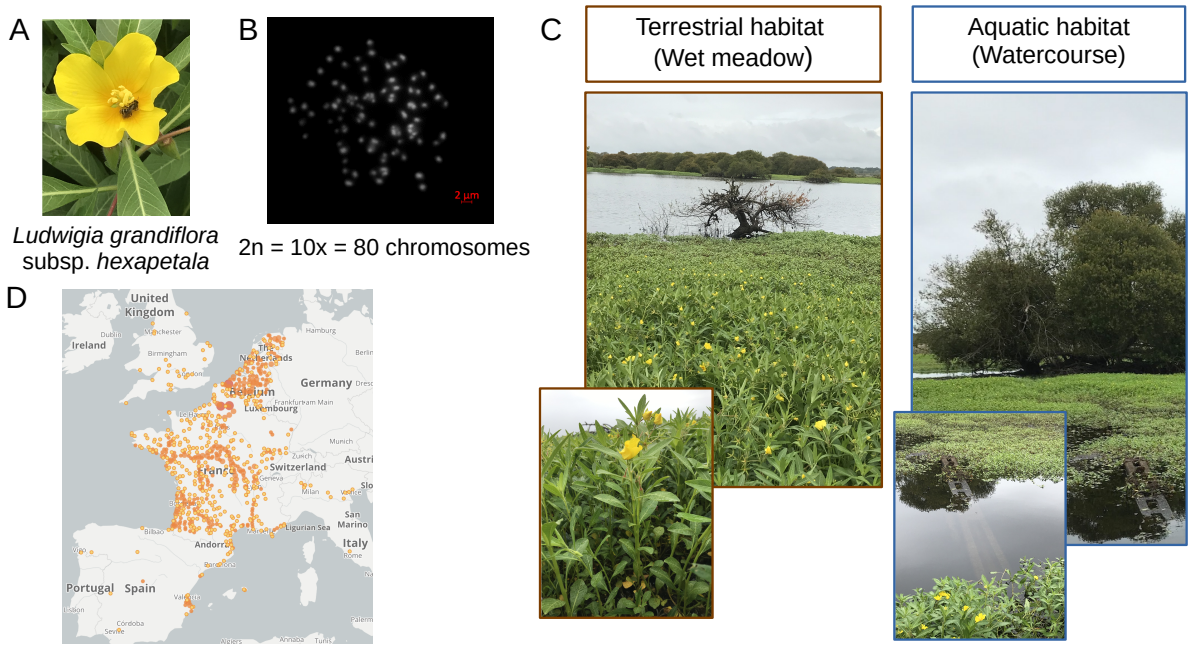


Figure 1. Overview of the amphibious invasive plant, *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*). A: Photography of *Lgh* (D. Barloy). B: Metaphase chromosomes of *Lgh* [9]. C: Growing ability of *Lgh* in various habitats (D. Barloy). D: Map of the distribution of *Lgh* in Europe (GBIF), size of points correspond to number recorded.

identified as a contributor of one genome part (2x), while the progenitor(s) corresponding to the last tetraploid (4x) genome part remains unknown. The genome size of *Lgh* was estimated to be 1.419 Gb (± 0.107 Gb) using flow cytometry [9]. Genomic analyses are essential for deepening our understanding of invasive species' adaptative abilities [11]. However, genomic data for *Ludwigia grandiflora* subsp. *hexapetala* remain scarce, with only the mitochondrial genome and plastome having been recently sequenced and assembled [12, 13]. Furthermore, as of June 2025, only four genomes from the Onagraceae family are available on GenBank, none of which are fully annotated [14]. These include *Chamaenerion angustifolium* and *Epilobium hirsutum*, which are both assembled at the chromosome level, and *Oenothera biennis* and *Clarka xantiana*, which are only available at the scaffold level. Within the Myrtales order, which is composed of about 13,000 species, there are 84 complete or draft genomes, but only four annotated in the RefSeq database.

Context

Here, we report the first annotated draft nuclear genome assembly of *Ludwigia grandiflora* subsp. *hexapetala*, to complement the organelle genomes already available. This genome helps to fill the substantial gap in genomic resources for the Onagraceae family, and more broadly, for the Myrtales order. It also constitutes the first genome reported for the Ludwigioideae sub-family. These genomic resources offer one of the most comprehensive references to date for decoding the acclimation mechanisms of *Ludwigia grandiflora* subsp. *hexapetala* to terrestrial habitats. Given the scarcity of genomic resources for the Myrtales order, especially within the Onagraceae, this new genome paves the way for comprehensive studies on genetic diversity, phylogenetics, and evolutionary dynamics within this complex family.

Material and methods

Plant material

Ludwigia grandiflora subsp. *hexapetala* (*Lgh*) plants were collected from the Mazerolles swamps (N47 23.260, W1 28.206) near Nantes (France). Several plants were grown in a growth chamber from a single 10 cm stem, using vegetative propagation in order to avoid potential genetic variation, as described in Barloy-Hubler et al., 2025 [13].

DNA extraction and Sequencing

High-molecular-weight (HMW) DNA was extracted from new buds following the protocol described in Barloy-Hubler et al., 2025 [13]. DNA libraries were then sequenced using Illumina MiSeq (RRID:SCR_016379) and GridION Oxford Nanopore technologies (RRID:SCR_016379) to obtain short (SR) and long (LR) reads, respectively (Table 1). The sequencing was carried out at the Genome Transcriptome Facility of Bordeaux (INRAE-UMR 1202 BIOGECO, Bordeaux, France).

Table 1. Statistics of sequencing data used for assembling *Ludwigia grandiflora* subsp. *hexapetala* genome.

Sequencing technology	Sequencing type	Number of reads	Mean size	Total size	Depth
Illumina MiSeq	Short reads	2 x 23,067,490	207.95	9,606,411,104	6.5x
Oxford Nanopore	Long reads	482,619	4944.6	2,386,362,939	1.6x

De novo genome assembly

Sequence quality controls have already been described in Barloy-Hubler et al., 2025 [13]. In short, LR and SR were filtered with Guppy v4.0.14 (RRID:SCR_023196) and fastp v0.20.0 (RRID:SCR_016962), respectively [15]. We performed a hybrid *de novo* genome assembly by integrating both long-read (Nanopore) and short-read (Illumina) sequencing data. Long reads were first error-corrected using Ratatosk (RRID:SCR_021824), which leverages short-read Illumina data to reduce long-read error rates sixfold compared to raw reads [16]. Reads corresponding to chloroplast and mitochondrial genomes were initially depleted by mapping against reference sequences using minimap2 (RRID:SCR_018550) [17]. Next, three long-reads assemblers were employed in parallel to maximize contiguity and assembly accuracy: Flye (RRID:SCR_017016), Canu (RRID:SCR_015880) and wtdbg2 (RRID:SCR_017225) [18, 19, 20]. A complementary hybrid assembly was realized using SPAdes (RRID:SCR_000131) [21]. Subsequently, we scaffolded and merged individual assemblies in order to resolve conflicts and select the highest-quality contigs and scaffolds from each tool. This multi-assembler strategy allowed us to exploit the strengths of each algorithm, resulting in a more complete and accurate nuclear genome assembly. Finally, Cd-hit v4.8.1 (RRID:SCR_007105) (parameters: -c 0.90) was used to reduce redundancy among contigs [22].

CDS annotation

To aid assembly of the remaining contigs, CDS were annotated by MEGANTE (release 2018-02), using *Arabidopsis thaliana* as the gene prediction reference [23]. Genes with the exact same sequence were considered as duplicated. Amino acid sequences predicted by MEGANTE, but lacking functional annotation were subsequently aligned against the NCBI non-redundant (NR) database for functional inference (28-07-2023) using diamond BLASTp v2.1.11 (RRID:SCR_016071) (parameters: –max-targets-seqs 1) on the Europe Galaxy server (RRID:SCR_006281) [24, 25]. All ORF were also annotated by InterProScan (v5.59-91.0) (RRID:SCR_005829) [26]. All matching GO terms were submitted to ReViGo (RRID:SCR_005825) with their occurrence frequency in *Lgh* genes (Parameters: Resulting list size = "Medium (0.7)", "values associated with GO terms represent" = "Higher value is better", Database = "*Arabidopsis thaliana*") [27]. ReViGo results were visualized as representative clusters via the TreeMap method, where the size of each rectangle is based on the occurrence frequency of each term in the *Lgh* genome. Proteins were classified as having uncharacterized proteins if all associated annotations contained at least one of the terms: "hypothetical", "uncharacterized", "detected", "unknown", "predicted protein" or "unnamed". WoLF PSORT (RRID:SCR_002472) was used to predict subcellular protein localization, keeping the best match for each protein [28]. GhostKOALA assigned KEGG annotations to each ORF [29]. KEGG Mapper was used to represent the metabolic pathway maps, and chloroplast and mitochondria coded proteins were added [30].

Biocuration

All CDS were extracted and *de novo* assembled using Geneious Prime "*de novo* assembly" option (version 2025.0.3, RRID:SCR_010519, https://www.geneious.com) (parameters: Minimum Overlap: 100, Minimum overlap identity: 95%, other parameters were the same as the "Low Sensitivity/Fastest"). These assemblies were used to align all genome contigs to identify potential overlaps and enable their further scaffolding using the "Map to Reference(s)" function (parameters: Minimum Overlap: 100, other parameters were the same as the "Low Sensitivity / Fastest") from Geneious. Contigs mapped to the same CDS were then assembled using the "*de novo* assembly" function (same parameters as before) from Geneious. As plant nuclear genomes contain mitochondrial and chloroplast sequences [31], long reads and assemblies belonging to the mitogenome and plastome were subsequently reintegrated into the assembly process to evaluate whether they could aid in scaffolding, and improving contig connectivity. Contigs that were elongated by the first step were *de novo* assembled with the reads mapped to them (same parameters as before). Elongated contigs after this second step were all *de novo* assembled (same parameters as before). Finally, all contigs were mapped to the new elongated contigs with "Map to Reference" function (same parameters as before).

Decontamination

Decontaminating the genome assembly is essential for removing foreign DNA sequences, such as those from bacteria, fungi, or environmental contaminants, which can lead to misannotations and erroneous conclusions. To remove contaminant contigs, all amino acid sequences predicted by MEGANTE were mapped back to the final contigs using Miniprot v0.13 (parameters: –outc 1 –outs 1) [32]. Each mapped predicted protein was assigned to a phylum via diamond BLASTp. Those classified within Viridiplantae (Streptophyta and Chlorophyta Phylums) were retained as non-contaminants, whereas proteins matching other phyla were flagged as contaminants. Accordingly, any contig harboring at least one contaminant predicted protein, or predominately comprising unannotated proteins (at least 17 out of 20) were then classified by kraken2 v2.1.3 (RRID:SCR_005484), as well as contigs that didn't contain any protein [33]. Contigs classified as non-viridiplantae were compared to the nt database, and BLASTn v2.16.0 (RRID:SCR_004870) [34]. NCBI FCS GX v0.5.5 (RRID:SCR_026367) (parameters: –tax-id 33090) was used on all contigs to check for contamination [35]. Contigs that were still not classified were then compared to other Myrtales species genomes from EnsemblPlants with BLAT (v.39x1) (RRID:SCR_011919) [36, 37]. Those matching with at least two species were conserved, while the others were sent to Kraken2 classification using the UniVec_Core and four Onagraceae home-made database (*Chamaenerion angustifolium* (Accession: GCA_946814005.1), *Oenothera biennis* (Accession: GCA_025802395.1), *Epilobium hirsutum* (Accession: GCA_965153245.1), and *Clarkia xantiana* (Accession: GCA_036873135.1)). Contigs classified as Onagraceae were kept.

Non-coding genomic features annotation

Ribosomal and transfer RNA were respectively annotated by pybarrnap v0.5.1 (Parameters : –kingdom euk –reject 0.8) and tRNAscan-SE v2.0.12 (RRID:SCR_008637) [38, 39]. DANTE v0.1.9 predicted transposable elements in the *Lgh* genome using the database "Viridiplantae_v3.0" [40]. Other repeats were annotated by RepeatMasker v4.1.8 (RRID:SCR_012954) [41] and Tandem Repeats Finder (RRID:SCR_022065) [42].

Scaffolding

RagTag (v2.1.0) was used to scaffold the *Lgh* genome (Parameters: –mm2-params "-x asm20") (Alonge et al. 2019). Three nuclear reference genomes were used iteratively: *Chamaenerion angustifolium*, *Epilobium hirsutum* and Punica granatum (See Supplementary Table 1 for accessions and details). We initially scaffolded the genome using *Chamaenerion angustifolium*, as it yielded the highest number of mapped *Lgh* contigs. We used *Epilobium hirsutum* as a second scaffold, due to it being phylogenetically closer to *Lgh*. This improved assembly coherence compared to *Punica granatum* [10, 43]. We applied this multi-reference scaffolding strategy because no fully assembled genome existed for any species that was sufficiently close to *Lgh* phylogenetically [44].

Estimation of chromosome size

A study compiling 419 angiosperm species showed that the average length of a chromosome (L, in μm) follows a power law with respect to the amount of chromosomal DNA (C, in picograms): $L = a \cdot C\alpha$ with $a \approx 4.64$ and $\alpha \approx 0.506$ [45]. This relationship can be inverted to estimate C from L: $C = (L/4.64)^{1/0.506}$ then converted to bases with the following reference conversion: $1\,pg \approx 0.978 \times 10^9 bp$ [46]. Using these formulae, *Lgh*'s chromosome size range was estimated to be approximately 30 to 134 Mb per chromosome, based on chromosome length measured by cytological observations between 0.8 and 1.7 μm [3]. This is greater than the estimated average chromosome size of 17.7 Mb, corresponding to the size of the decaploid genome ($\approx 1.42 Gb$) divided by the number of chromosomes (2n = 10× = 80). Estimates of genome size by flow cytometry are approximate and subject to variations estimated at ±29% or more, depending on the internal standard and specific experimental conditions (extraction buffer, fluorochrome, plant tissue/metabolites or tissue type) [47]. This correction would take us from a minimum size of 21.3 Mb to a maximum size of 172.9 Mb.

Qualitative and quantitative genome evaluation

BUSCO v5.8.0 (RRID:SCR_015008) was used to evaluate the quality of the assembly [48]. Long reads and short reads were mapped to the genome using Minimap2 v2.28 (RRID:SCR_018550) (Parameters: -x map-ont for LR and -x sr for SR) [17]. The statswrapper script from BBMap (RRID:SCR_016965) calculated genome statistics (N50, L50, N90 and L90) [49].

Orthologous genes analysis

Proteomes from 12 species, corresponding to all nine Myrtales order species with annotated genomes, and three Malvid clade species used as outgroups, were downloaded from NCBI Genbank (See Supplementary Table 1 for accessions and details). *Chamaenerion angustifolium* and *Epilobium hirsutum* genomes were annotated by the Helixer Web Service v0.3.4 using only the chromosomes [50]. To find common orthogroups and orthologous genes between *Lgh* and other species, these 15 proteomes (the 12 downloaded from NCBI Genbank + *C. angustifolium*, *E. hirsutum* and *Lgh*) were analyzed by Orthofinder v3.0.1 (RRID:SCR_017118) [51]. The R package UpSetR (RRID:SCR_026112) was used for the upset graph [52].

Genome synteny

LASTZ v1.04.15 (RRID:SCR_018556) in Geneious was used to compare synteny between *Lgh* and *Chamaenerion angustifolium* genomes [53]. *Lgh* scaffolds bigger than 1 Mb, and *C. angustifolium* chromosomes were used. Sequence alignments of at least 1 Mb in total were conserved.

Results and discussion

Sequencing, assembling and scaffolding

Following *de novo* assembly, we obtained 175,249 initial contigs. To reduce redundancy, we applied CD-HIT-est at a 90% identity threshold, yielding 158,330 non-redundant contigs. Next, contigs sharing overlapping coding sequence (CDS) annotations were merged, resulting in a refined set of 149,835 contigs. To improve scaffold continuity, we incorporated long reads along with plastome and mitogenome assemblies, which slightly increased contig contiguity, producing 149,534 contigs. Finally, after comprehensive decontamination removing bacterial, fungal, and other non-target sequences, we retained 117,972 high-confidence nuclear contigs for downstream analysis. These contigs were iteratively scaffolded against *C. angustifolium*, *E. hirsutum* and *P. granatum*. Initially, 3088 contigs aligned to the *C. angustifolium* genome (Onagraceae), producing 49 scaffolds. These scaffolds, along with any remaining contigs, were then mapped to the *E. hirsutum* genome, resulting in 2,260 alignments and the formation of 32 additional scaffolds. In the final step, 1,550 scaffolds and unscaffolded contigs were aligned to the *P. granatum* genome (Punicaceae), generating 50 new scaffolds. Overall, this multi-tiered scaffolding strategy reduced the total number of nuclear contigs from 117,972 to 111,219 (Table 2), creating a final assembly of 111,219 scaffolds. This elevated number of scaffolds likely results from the low lengths of ONT long reads, which are not long enough to span repetitive regions, leading to unresolved repeats during assembly, and consequently, highly fragmented scaffolds (Figure 2).

Table 2. *Ludwigia grandiflora* subsp. *hexapetala* genome statistics.

Statistics	Data
Number of contigs and scaffolds	111,219
Number of scaffolds	50
Number of contigs	111,169
Total size (bp)	1,486,583,937
Maximum scaffold length (bp)	53,046,275
Minimum contig length (bp)	4,970
N50 (bp)	13,498
L50	26,764
N90 (bp)	6,387
L90	91,755
GC content (%)	39.7
Long reads mapped (%)	99.65
Short reads mapped (%)	98.62
BUSCO score (%)	C:96.3 [S:15.5, D:80.8], F:1.6, M:2.1

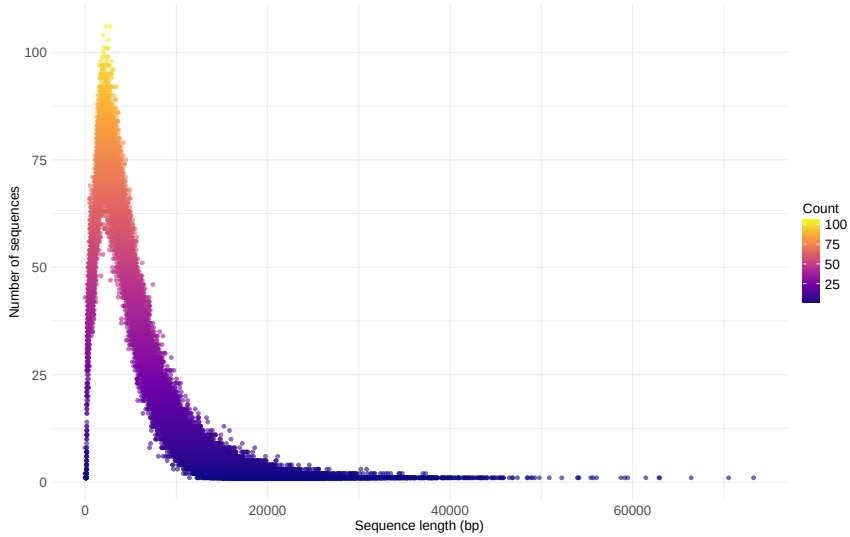


Figure 2. Distribution of Oxford nanopore long read lengths in *Ludwigia grandiflora* subsp. *hexapetala* sequencing.

Indeed, obtaining high-quality, high molecular weight (HMW) DNA in plants is critical for successful long-read sequencing. However, plant genomes such as *Lgh* often pose significant challenges due to the presence of secondary metabolites such as polysaccharides and polyphenols, which can co-purify with DNA, and cause fragmentation, or inhibit downstream enzymatic reactions [54]. Given that our efforts to extract high molecular weight (HMW) DNA from *Lgh* have been unsuccessful, and the DNA quality remains suboptimal, alternative approaches such as stLFR or TELL-Seq could represent robust solutions for obtaining chromosome-scale genome assemblies. Rather than depending on ultra-long intact DNA fragments, these synthetic long-read technologies barcode fragmented DNA molecules within a single tube, enabling the in silico reconstruction of long-range genomic information ($\simeq$ 10–100 kb) from standard short-read sequencing data. These methods have been successfully applied in plants; for example, TELL-Seq has demonstrated effective de novo genome assembly, and accurate structural variant detection in *Arabidopsis thaliana* [55]. Moreover, given that *Lgh* exhibits decaploidy (10x), employing capture sequencing (Hi-C, Pore-C, Omni-C) can be considered, as this technique has been successfully applied to complex polyploid species such as *Brassica napus* and *Eutrema japonicum* [56, 57].

Genome qualitative and quantitative assessment

Different statistics were used to assess *Lgh* genome assembly (Table 2). The assembled genome has a total length of 1.487 Gb, which falls within the range of the genome size estimated by flow cytometry (1.419 $\pm$ 0.107 Gb) [9]. Scaffold lengths range from 5 kb to 53 Mb. Based on calculated chromosome size estimates, ranging from 21.3 Mb to 172.9 Mb, four scaffolds (from 35.2 to 53 Mb) fall within the expected range for complete chromosomes (out of the 40 expected). The biggest scaffold size is bigger (53 Mb) than *C. angustifolium* and *E. hirsutum*'s biggest chromosomes (46 Mb and 22 Mb, respectively). Furthermore, five other scaffold sizes are between 15 and 46 Mb long, and thereby within the range of those two species chromosome lengths, suggesting these scaffolds might be complete or almost complete (Supplementary Figure 1). The N50 value is 13,498 bp which shows that there are still a lot of very small contigs in the assembly. Almost all LR and SR (99.65% and 98.62%, respectively) are mapped on *Lgh* genome. The BUSCO score is 96.3% showing a good completeness of the genome. The presence of 80.8% of duplicated BUSCOs in the *Lgh* genome is not surprising for this decaploid species, as the octoploid species *Cardamine chenopodiifolia* presented 80.9% duplicated BUSCOs [58].

Protein-coding genes annotations

MEGANTE found 246,242 protein-coding genes (PCG) in the *Lgh* genome (Supplementary Table 2). This total is much bigger than other related species such as *Arabidopsis thaliana*, which had 48,265 PCG and *Punica granatum* which had 36,608 PCG (See Supplementary Table 1 for accession). However, *Lgh* is a decaploid species and these two species are diploid, which might explain this difference, as polyploid species tend to have more PCG [59]. The decaploid genome of *Houttuynia cordata* (2n = 10x = 90), with a genome size of 2.63 Gb, contains only 139,087 PCG, which represents a little more than half of the *Lgh* PCG number [60]. *H. cordata*, like *Lgh*, is reported to be an autoallopolyploid species. However, *H. cordata* only has two ancestral parents, with one parent contributing 9x parts of the 10x genome, whereas the two *Lgh* donors identified contribute to 4x and 2x of 10x genome [9]. Another example is the *Prunus domestica* genome, which is a hexaploid species, possessing 138,361 PCG with a genome size of 1,3 Gb [61]. All these data suggest that although the number of genes can be linked to genome size and/or polyploidy, it is not really consistent between species and can vary a lot. Among the 246,242 PCG annotated in the *Lgh* genome, 93,6% (208,914) of them are in single copy while 6,4% (14,247) are duplicated (Supplementary Table 2). This gave a total of 223,161 unique genes in the *Lgh* genome. Among them, almost half (106,292) had at least one of diamond BLASTp, GO, Interpro, or KEGG annotation (Figure 3). The other 116,869 proteins are considered as hypothetical proteins since no homology was found with any sequences in the databases. Among the annotated proteins, almost 75% (77,391) were annotated as unknown function proteins. The observation that fewer than 7% of genes are identified as duplicated in *Lgh* decaploid genomes may be attributed to several factors. Firstly, many duplicated genes undergo rapid silencing through mechanisms such as reciprocal silencing, where redundant gene copies are epigenetically silenced without affecting the organism's phenotype. This silencing can occur shortly after polyploidization, leading to the apparent absence of duplicated genes in functional analyses [62]. Secondly, duplicated genes often experience functional divergence through processes like subfunctionalization and neofunctionalization. In subfunctionalization, each gene copy retains a subset of the original gene's functions, while in neofunctionalization, one copy acquires a new function. These processes can make it challenging to identify duplicated genes using standard annotation methods, as the functional roles of the duplicates may differ from the original gene [63].

Protein-coding genes functional characterization

GO terms were assigned to 55,954 proteins. Analysis of these GO terms according to the number of occurrences show that key molecular functions such as oxidoreductase activity, nucleic acid binding or protein binding were overrepresented (Supplementary Figure 2). Biological process GO annotation such as protein phosphorylation, transmembrane transport or proteolysis were very present in *Lgh* proteins (Supplementary Figure 3). GO terms for cellular components such as the membrane, nucleus and ribosome particularly stood out (Supplementary Figure 4). Some of the most frequent keywords in all GO terms are process, exact or activity (Figure 4). Protein subcellular localizations were predicted by WoLF PSORT (Figure 5, Supplementary Table 3). Chloroplasts are the most common localization with 84,283 proteins (34%), followed by the nucleus with 78,267 (31%) proteins, then the cytosol with 48,179 proteins. As chloroplasts host various important processes such as photosynthesis, it is not surprising to find a high proportion of proteins here [64]. Plastids are morphologically and functionally diverse organelles that are dependent on nuclear-encoded, plastid-targeted proteins for all biochemical and regulatory functions. In plant

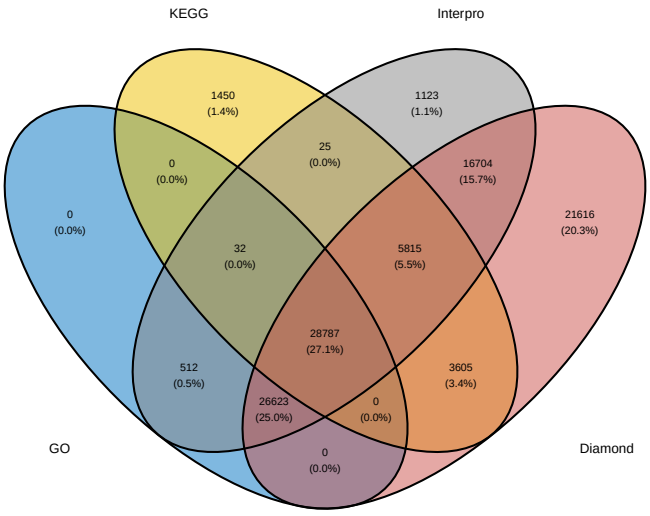


Figure 3. Venn diagram of *Ludwigia grandiflora* subsp. *hexapetala* genome annotations : KEGG (yellow), Interpro (grey), GO (blue) and diamond (red).

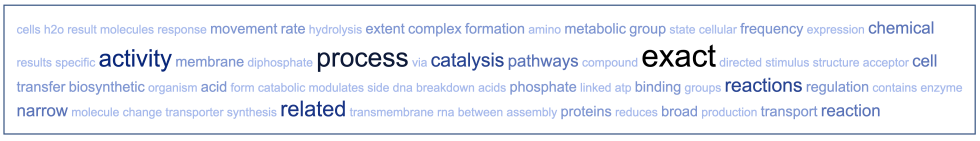


Figure 4. REVIGO word cloud of frequent GO term keywords in *Ludwigia grandiflora* subsp. *hexapetala* genome, based on occurrence frequency.

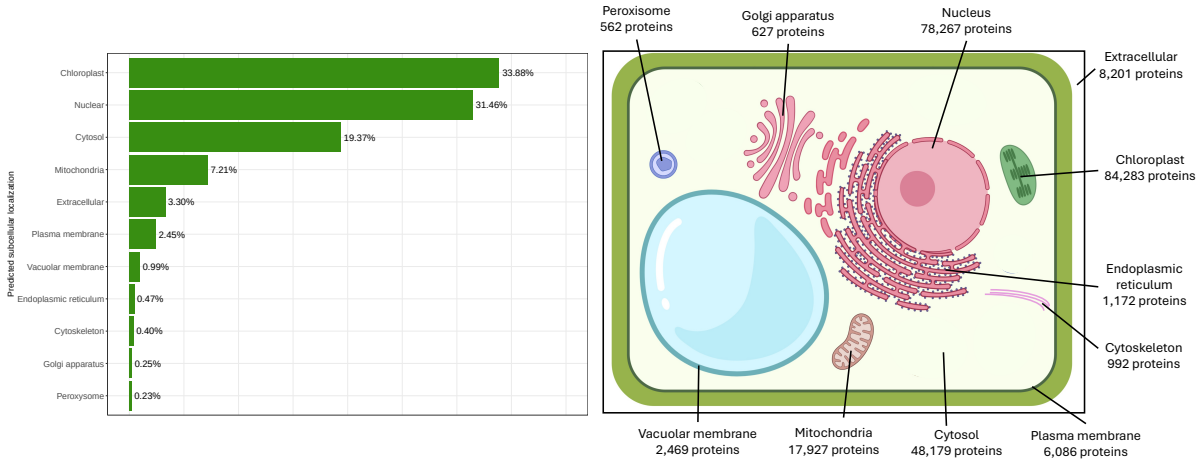


Figure 5. *Ludwigia grandiflora* subsp. *hexapetala* nuclear-encoded proteins subcellular localization prediction by WoLF PSORT.

genomes, the proportion of proteins predicted to be localized in the plastids typically ranges from 15–25%, with notable variation among species: in tomatoes (*Solanum lycopersicum*), a comprehensive plastid protein atlas predicted approximately 7,473 nuclear-encoded plastid proteins, which represents about 22% of the proteome [65]. In *Arabidopsis thaliana* and other angiosperms, genome-wide analyses predict 3,000–5,000 plastid-targeted proteins, accounting for roughly 15–20% of total proteins [66]. A 34% prediction for *Lgh* likely reflects a blend of biological expansions in plastid functions linked to its water to land invasive abilities. What is more surprising is the difference between Cellular Component GO terms and the predicted localization. The GO term "membrane" (GO:0016020) is a broad parent category that encompasses various membrane types like thylakoid, mitochondrial, nuclear, endoplasmic reticulum membranes. Thus, a protein localized in the chloroplast, mitochondrion, or nucleus membrane can still carry the general annotation "membrane" in its GO terms. KEGG annotations were assigned to 39,714 genes and important metabolic pathways such as the glycolysis and TCA cycle appear to be complete (Supplementary Figure 5). Annotating $\simeq$ 16% of proteins in the *Lgh* genome with KEGG seems surprisingly low, as using KEGG-based tools, the prediction reaches 26% for *Brassica napus* [67]. This low rate of KEGG annotation in the *Lgh* genome could be linked to its invasive ability. It may be due to the presence of invasion-specific genes not yet represented in existing databases.

Non-coding RNA genes annotation

The *Lgh* genome comprises a total of 15,834 rRNA genes, including 2,018 5S, 3,939 5.8S, 6,237 18S and 3,640 28S rRNA genes (Supplementary Table 2). The genome of *Ludwigia grandiflora* subsp. *hexapetala* harbors a number of rRNA genes that is notably high compared to other plant species, even though this number can be very variable in plants (generally from 500 to 40,000) [68], and for the 5S rRNA gene, it can even reach 75,000 copies [69]. This abundance may be attributed to the species' polyploid nature and the presence of multiple ribosomal DNA (rDNA) loci. Usually, rRNA genes are organized in two repeated units : the 35S rDNA, which contains the 18S, 5.8S and 28S rRNA genes, and is 8 to 12 kb long, and the 5S rDNA, which contains only the 5S rRNA gene, and is about 800-900 pb long [70]. Both rDNA units are often grouped in clusters. The 35S rDNA clusters are long tandem repeats localized in nucleolar organizer regions (NORs) [69, 71]. In the *Lgh* genome, 17 clusters of 35S rDNA of more than 50 kb were found. However, NORs can span millions of base pairs, suggesting that these high repetitive regions might be incomplete in the *Lgh* genome [72]. Even though they are incomplete, we can expect the presence of multiple NORs, which are often associated with polyploidy and genomic redundancy [73]. Such a high copy number of rRNA could have implications on ribosome biogenesis and cellular metabolism, potentially enhancing the plant's adaptability to various environmental conditions. A total of 8,417 tRNA genes were annotated in the *Lgh* genome, surpassing the 6,475 tRNA genes identified in *Ipomea nil*, the plant species with the highest number of tRNA genes out of 128 others in a study [74]. In plant genomes, the most abundant tRNA isotypes are tRNA-Ala, tRNA-Pro, tRNA-Ser, tRNA-Arg and tRNA-Leu [75]. Three of them were the most abundant tRNA isotypes in the *Lgh* genome (tRNA-Arg, tRNA-Ser and tRNA-Leu), with frequency comprised between 8.5 to 9.2%, representing 26.6% of total tRNA number (Supplementary Figure 6). All anticodons were present in *Lgh* tRNA genes, although some of them very underrepresented, such as Pro-GGG (0.07%), Gly-ACC (0.07%), Leu-GAG (0.04%) and Tyr-AUA (0.04%). This was unsurprising, as these anticodons are among the least abundant in the Plant Kingdom (from 0.009 to 0.038%) [74]. On the other hand, Met-CAU (7.6%), Asn-GUU (7.5%), Arg-ACG (4.8%) and Asp-GUC (4.3%) were overrepresented in *Lgh* tRNA genes. The three anticodons Met-CAU, Asn-GUU and Asp-GUC are the three most abundant in plants (from 4.02 to 5.04%) [74].

Repeated elements

The *Lgh* genome is made up of 8.5% (126 Mb) of repeated elements (Figure 6, Supplementary Table 4). A total of 625,820 simple sequence repeats, 388,071 tandem repeats and 140,494 transposable elements were identified. Among the transposable elements, almost 70% (98,220) were LTR Ty3-gypsy. The proportion of repeated elements in the *Lgh* genome was very small compared to other polyploid plant genomes, such as the decaploid species *H. cordata* (55.52%), or the tetraploid species *Camelia granthamiana* (68.48%) genomes [60, 76]. This low rate of *Lgh* repeated elements (8.5%) could simply be due to the fact that our analysis did not assemble them and that they are therefore missing. LTR elements, frequently span 4–31 kb and are often nested within each other, as observed in maize or wheat, with clusters of up to 155 kb [77, 78]. This produces contig breaks where reads cannot span unique flanking regions. This phenomenon results in

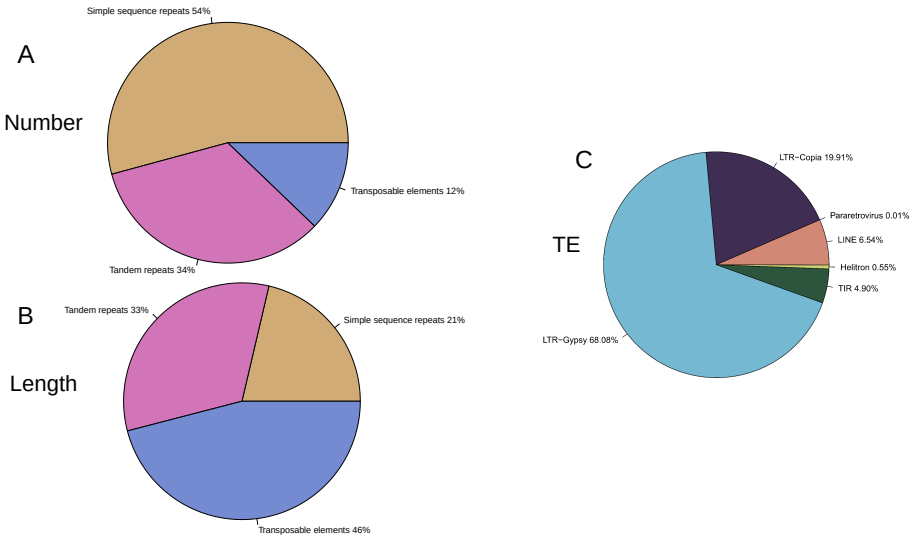


Figure 6. Repeated elements in *Ludwigia grandiflora* subsp. *hexapetala* genome. A: Distribution of number of repeated element types. B: Distribution of sizes of repeated element types. C: Distribution of number of transposable elements major families. Detailed number can be found in Supplementary Table 4.

collapsed or missing copies, and as repetitive regions are not properly assembled, they are underrepresented or completely missing in the final genome. Another hypothesis for the repeat content in the *Lgh* genome is the low percentage of repeat content in some polyploid species compared to diploid species [79]. This observation can be explained by a purging mechanism of repeated elements, and in particular LTRs, following whole genome duplication [80].

Orthologs comparison in the Malvids clade

The *Lgh* proteome was compared with those of 14 Malvid species using OrthoFinder to identify orthologous gene sets. This analysis revealed that 9,691 orthogroups are shared across all species, illustrating a substantial core proteome conserved within the clade (Figure 7A). However, 8,605 orthogroups were unique to *Lgh*, which is a lot more than in any other analyzed species. Furthermore, *Lgh* is the only species which does not contain a majority of genes belonging to orthogroups common to all species (Figure 7B, Supplementary Table 5). Almost 42% (77,448/185,876) of genes assigned to orthogroups in *Lgh* were species-specific, which is a much bigger proportion than in any other species (from 0.5% in *R. argentea* to 12.5% in *A. thaliana*). This result is in agreement with the specificity of the Ludwigia genus which is the only genus available for the Ludwigioideae subfamily, and is phylogenetically distinct from all other species of Onagraceae family [43]. This prevalence of orphan genes in the *Lgh* polyploid invasive plant may contribute to its adaptability and invasive success. In polyploid plants, whole-genome duplications (WGDs) can lead to the formation of orphan genes through mechanisms such as neofunctionalization and subfunctionalization, processes that allow duplicated genes to acquire novel functions or divide ancestral functions between them [81]. Studies have identified orphan genes in various plant species, including *Mikania micrantha*, which shows a significant number of unique gene families related to oxidative phosphorylation, biosynthesis of unsaturated fatty acids, photosynthesis, and ether lipid metabolism. These pathways are crucial for growth, metabolism, and defense responses, highlighting the functional importance of orphan genes in adaptation to new environments [82]. The functional annotation of orphan genes poses challenges due to their lack of sequence similarity to known genes. Integrating transcriptomic data can provide insights into their roles as they can play a pivotal role in *Lgh*'s adaptive capacity for land habitats. Their unique functions, arising from genomic duplications, enable these plants to exploit new ecological niches and outcompete native species. Understanding the origin and function of orphan genes is crucial for unraveling the mechanisms underlying plant invasiveness, and for developing strategies to manage invasive species. Among these orthologs, 83% (64,350) were annotated as hypothetical proteins. These genes could either be unique *Lgh* genes, or an 'overannotation' resulting from our annotation methods. If we remove them from the analysis, the proportion of *Lgh*-specific genes assigned to orthogroups drops to 7%, which is within a similar range to other species. Very few genes in each genome belonged to family or order specific orthogroups.

Synteny conservation in the Onagraceae family

The *Lgh* genome was aligned to the *C. angustifolium* genome with LASTZ to evaluate the synteny conservation (Figure 8). Due to *C. angustifolium* being used in the scaffold process, the synteny was often conserved between the two species. Multiple chromosomes of *C. angustifolium* were aligned to each *Lgh* scaffold, excluding S33 and S36, which didn't have enough similarities to the *C. angustifolium* genome. The chromosomes aligned almost entirely to *Lgh* scaffolds, showing an important sequence conservation between the two species, apart from for chromosomes 15 and 18. Eleven *C. angustifolium* chromosomes (out of the 16 that aligned to the *Lgh* genome) were aligned discontinuously with the *Lgh* genome with a big gap around the middle. These regions may correspond to centromeric regions which are known to contains a lot of repetitive elements, and which may be missing in the *Lgh* assembly [83]. These results are preliminary, however, as the analysis of synteny in incomplete plant genomes presents significant challenges that can compromise the accuracy and reliability of comparative genomic studies. In polyploid plants, where multiple subgenomes may exist, incomplete assemblies can lead to misinterpretation of syntenic regions, as gaps or misassembled contigs can disrupt the continuity of gene clusters. In addition, the

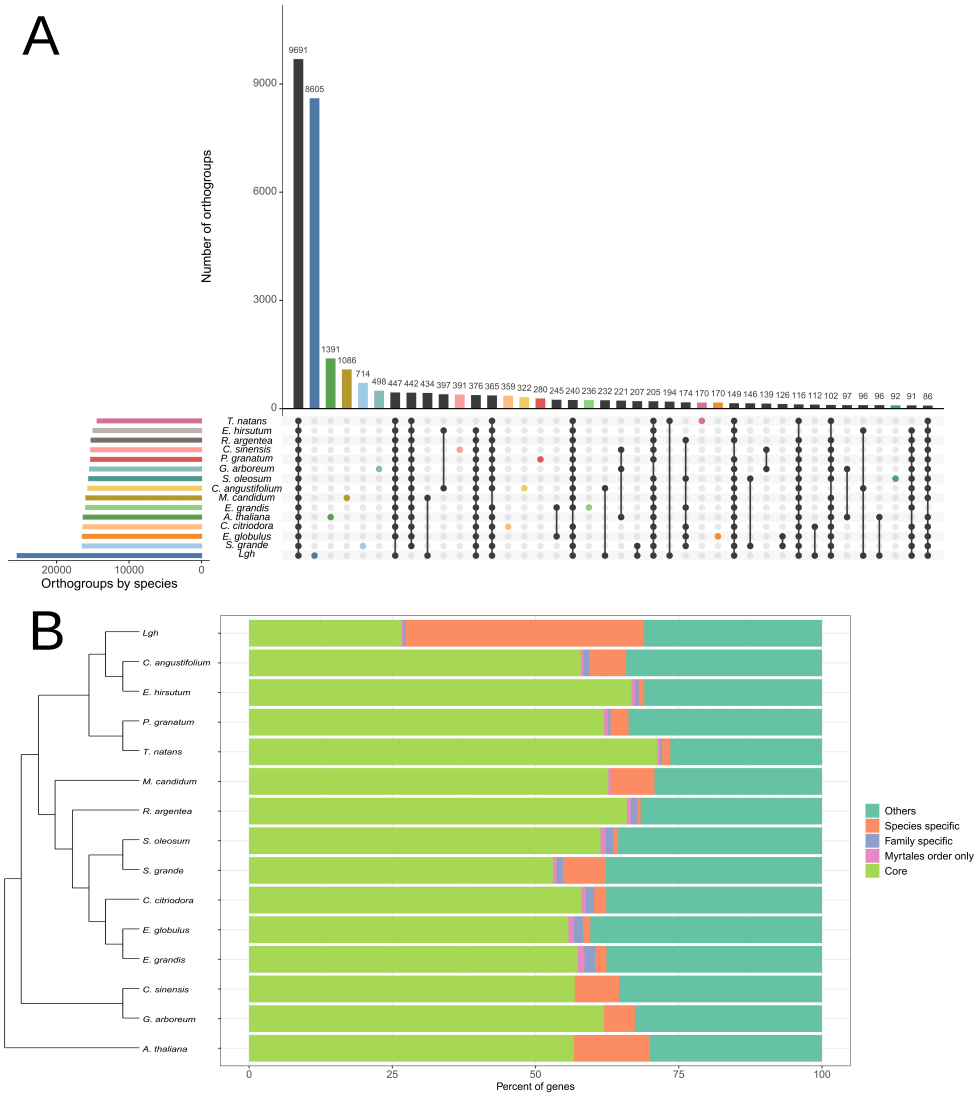


Figure 7. Orthologs comparison of Malvid species proteomes. A: UpSet plot of orthogroups shared among different species, sorted by frequency. B: Phylogenetic tree based on all genes and percent of genes classified in orthogroups shared or unique among species. Core refers to genes in orthogroups present in all species. References for genomes used in this analysis can be found in Supplementary Table 1. Detailed number are in Supplementary Table 5.

presence of repetitive sequences and structural variations, common in plant genomes, further complicates the identification of syntenic blocks. These problems can lead to erroneous conclusions about chromosome evolution and gene function. To alleviate these difficulties, researchers are developing advanced computational tools and methodologies that integrate long-term sequencing data and use algorithms capable of handling incomplete assemblies. These approaches aim to improve the accuracy of synteny analysis and provide deeper insights into the genomic architecture of plants.

Conclusion

This first draft genome assembly of *Ludwigia grandiflora* subsp. *hexapetala* marks a pivotal milestone as the inaugural reference for the Ludwigioideae subfamily. By addressing the considerable gap in genomic data within the Onagraceae family, and more broadly within the Myrtales order, this genomic resource significantly enhances our capacity to study genetic diversity, infer evolutionary relationships, and trace phylogenetic history across related taxa. Although the genome remains fragmented and incomplete, its availability provides a valuable foundation for future efforts in scaffolding, annotation refinement, and downstream comparative and functional genomic analyses.

Availability of supporting data and materials

The genome assembly generated in this study has been deposited at NCBI GenBank under the accession (not yet available). All sequencing reads can be accessed through the NCBI SRA (BioProject number: PRJNA1102104). The data sets supporting the results of this article are available in the GigaDB repository, [cite unique persistent identifier].

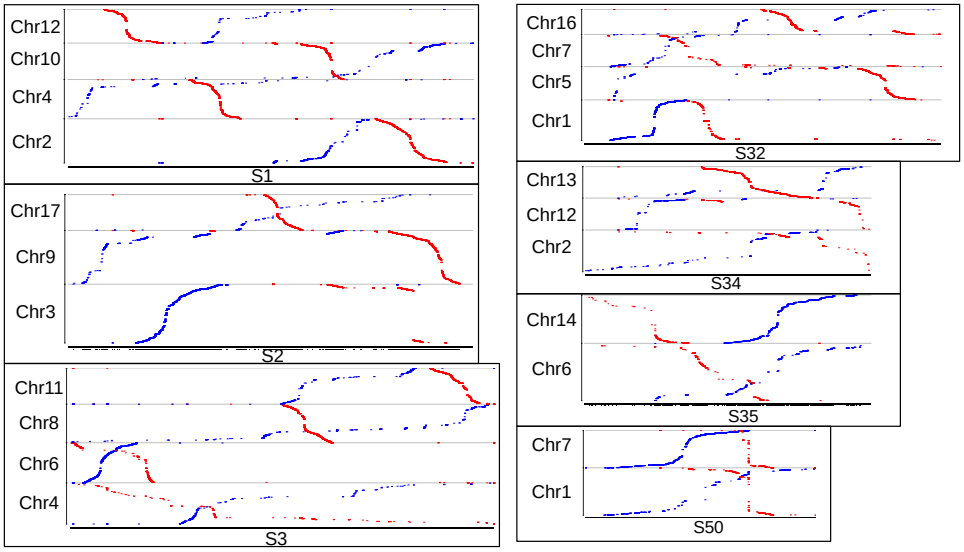


Figure 8. Dot plots of genome alignment between *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Chamaenerion angustifolium*. Only scaffolds > 1 Mb in *Lgh* were used. "S" correspond to *Lgh* scaffolds while "Chr" correspond to *C. angustifolium* chromosomes.

Declarations

List of abbreviations

BUSCO, Benchmarking Universal Single-Copy Orthologs; CDS, Coding-DNA sequence; GO, gene ontology; HMW, High-molecular-weight; *Lgh*, *Ludwigia grandiflora* subsp. *hexapetala*; LR, long reads; LTR, Long terminal repeat; NR, non-redundant; PCG, protein-coding gene; rRNA, ribosomal RNA; SMRT, single-molecule real-time; SR, short reads; tRNA, transfer RNA.

Consent for publication

Not applicable

Competing Interests

The author(s) declare that they have no competing interests.

Funding

This research was funding by INRAE through the doctoral grant of Guillaume Doré.

Author's Contributions

D.B. and F.B.-H. designed the project and defined the methodological approaches for which D.B. acquired funding. G.D. conducted the bioinformatics analyses, supervised by F.B.-H and DB. All authors wrote different parts of the manuscript and, together reviewed the manuscript.

Acknowledgements

Not applicable

References

1. Pyšek P, Hulme PE, Simberloff D, Bacher S, Blackburn TM, Carlton JT, et al. Scientists' warning on invasive alien species;95(6):1511–1534. https://onlinelibrary.wiley.com/doi/abs/10.1111/brv.12627, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/brv.12627.
2. Hussner A. Alien aquatic plant species in European countries;52(4):297–306. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3180.2012.00926.x, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-3180.2012.00926.x.
3. Dandelot S, Les Ludwigia spp. invasives du Sud de la France : Historique, Biosystématique, Biologie et Ecologie;.
4. Thiebaut G, Dutartre A. Management of invasive aquatic plants in France;p. 25–46.

5. Okada M, Grewell BJ, Jasieniuk M. Clonal spread of invasive *Ludwigia hexapetala* and *L. grandiflora* in freshwater wetlands of California;91(3):123–129. https://www.sciencedirect.com/science/article/pii/S0304377009000515.
6. Hieda S, Kaneko Y, Nakagawa M, Noma N. Ludwigia grandiflora (Michx.) Greuter & Burdet subsp. hexapetala (Hook. & Arn.) G. L. Nesom & Kartesz, an Invasive Aquatic Plant in Lake Biwa, the Largest Lake in Japan;71(1):65–71.
7. Thouvenot L, Haury J, Thiebaut G. A success story: water primroses, aquatic plant pests;23(5):790–803. https://onlinelibrary.wiley.com/doi/abs/10.1002/aqc.2387, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aqc.2387.
8. Haury JJ, Noël F, Bozec MM, Coudreuse J, Guil J, Marrel G, et al. Importance of Ludwigia grandiflora as invasive weed on meadows and pastures in Western France;. https://institut-agro-rennes-angers.hal.science/hal-00842412.
9. Barloy D, Portillo-Lemus L, Krueger-Hadfield S, Huteau V, Coriton O. Genomic relationships among diploid and polyploid species of the genus *Ludwigia* L. section *Jussiaea* using a combination of molecular cytogenetic, morphological, and crossing investigations;4. https://peercommunityjournal.org/articles/10.24072/pcjournal.364/.
10. Wagner WL, Hoch PC, Raven PH. Revised classification of the Onagraceae;http://hdl.handle.net/10088/7611.
11. McGaughran A, Dhami MK, Parvizi E, Vaughan AL, Gleeson DM, Hodgins KA, et al. Genomic Tools in Biological Invasions: Current State and Future Frontiers;16(1):evad230. https://doi.org/10.1093/gbe/evad230.
12. Doré G, Barloy D, Barloy-Hubler F. De Novo Hybrid Assembly Unveils Multi-Chromosomal Mitochondrial Genomes in Ludwigia Species, Highlighting Genomic Recombination, Gene Transfer, and RNA Editing Events;25(13):7283. https://www.mdpi.com/1422-0067/25/13/7283, number: 13 Publisher: Multidisciplinary Digital Publishing Institute.
13. Barloy-Hubler F, Le Gac AL, Boury C, Guichoux E, Barloy D. Sequencing, de novo assembly of *Ludwigia* plastomes, and comparative analysis within the Onagraceae family;5. https://peercommunityjournal.org/articles/10.24072/pcjournal.536/.
14. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation;44:D733–D745. https://doi.org/10.1093/nar/gkv1189.
15. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor;34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
16. Holley G, Beyter D, Ingimundardottir H, Møller PL, Kristmundsdottir S, Eggertsson HP, et al. Ratatosk: hybrid error correction of long reads enables accurate variant calling and assembly;22(1):28. https://doi.org/10.1186/s13059-020-02244-4.
17. Li H. Minimap2: pairwise alignment for nucleotide sequences;34(18):3094–3100. http://arxiv.org/abs/1708.01492.
18. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs;37(5):540–546. https://www.nature.com/articles/s41587-019-0072-8, number: 5 Publisher: Nature Publishing Group.
19. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation;27(5):722–736. http://genome.cshlp.org/content/27/5/722, company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.
20. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2;17(2):155–158. https://www.nature.com/articles/s41592-019-0669-3, publisher: Nature Publishing Group.
21. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler;70(1):e102. https://onlinelibrary.wiley.com/doi/abs/10.1002/cpbi.102, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpbi.102.
22. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences;22(13):1658–1659.
23. Numa H, Itoh T. MEGANTE: A Web-Based System for Integrated Plant Genome Annotation;55(1):e2. https://doi.org/10.1093/pcp/pct157.
24. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND;18(4):366–368. https://www.nature.com/articles/s41592-021-01101-x, publisher: Nature Publishing Group.
25. The Galaxy Community. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update;52:W83–W94. https://doi.org/10.1093/nar/gkae410.
26. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification;30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
27. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms;6(7):e21800. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021800, publisher: Public Library of Science.
28. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor;35:W585–W587. https://doi.org/10.1093/nar/gkm259.
29. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences;428(4):726–731.
30. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences;29(1):28–35.
31. Zhang GJ, Dong R, Lan LN, Li SF, Gao WJ, Niu HX. Nuclear Integrants of Organellar DNA Contribute to Genome Structure and Evolution in Plants;21(3):707. https://www.mdpi.com/1422-0067/21/3/707, number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
32. Li H. Protein-to-genome alignment with miniprot;39(1):btad014.
33. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2;20(1):257. https://doi.org/10.1186/s13059-019-1891-0.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool;215(3):403–410. https://www.sciencedirect.com/science/article/pii/S0022283605803602.
35. Astashyn A, Tvedte ES, Sweeney D, Sapojnikov V, Bouk N, Joukov V, et al., Rapid and sensitive detection of genome contamination at scale with FCS-GX. bioRxiv;. https://www.biorxiv.org/content/10.1101/2023.06.02.543519v1, pages: 2023.06.02.543519 Section: New Results.
36. Yates A, Allen J, Amode RM, Azov AG, Barba M, Becerra A, et al. Ensembl Genomes 2022: an expanding genome resource for non-vertebrates;50:D996–D1003. https://doi.org/10.1093/nar/gkab1007.
37. Kent WJ. BLAT—The BLAST-Like Alignment Tool;12(4):656–664. http://genome.cshlp.org/content/12/4/656, company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.
38. Shimoyama Y, pybarrnap: Python implementation of barrnap;. https://github.com/moshi4/pybarrnap.
39. Chan P, Lin B, Mak A, Lowe T. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes;49(16):9077–

9096. https://doi.org/10.1093/nar/gkab688.
40. Novák P, Hoštáková N, Neumann P, Macas J. DANTE and DANTE_LTR: lineage-centric annotation pipelines for long terminal repeat retrotransposons in plant genomes;6(3):lqae113. https://doi.org/10.1093/nargab/lqae113.
41. Smit A, Hubley R, Green P, RepeatMasker Open-4.0;. http://www.repeatmasker.org.
42. Benson G. Tandem repeats finder: a program to analyze DNA sequences;27(2):573–580. https://doi.org/10.1093/nar/27.2.573.
43. Overson RP, Johnson MG, Bechen LL, Kinosian SP, Douglas NA, Fant JB, et al. A phylogeny of the evening primrose family (Onagraceae) using a target enrichment approach with 303 nuclear loci;23(1):66. https://doi.org/10.1186/s12862-023-02151-9.
44. Rice ES, Green RE. New Approaches for Genome Assembly and Scaffolding;7:17–40. https://www.annualreviews.org/content/journals/10.1146/annurev-animal-020518-115344, publisher: Annual Reviews.
45. Kramer EM, Tayjasanant PA, Cordone B. Scaling Laws for Mitotic Chromosomes;9:684278. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8262490/.
46. Doležel J, Bartoš J, Voglmayr H, Greilhuber J. Nuclear DNA Content and Genome Size of Trout and Human;51A(2):127–128. https://onlinelibrary.wiley.com/doi/abs/10.1002/cyto.a.10013, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cyto.a.10013.
47. Nix J, Ranney TG, Lynch NP, Chen H. Flow Cytometry for Estimating Plant Genome Size: Revisiting Assumptions, Sources of Variation, Reference Standards, and Best Practices;149(3):131–141. https://journals.ashs.org/view/journals/jashs/149/3/article-p131.xml#T1, publisher: American Society for Horticultural Science Section: Journal of the American Society for Horticultural Science.
48. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes;38(10):4647–4654. https://doi.org/10.1093/molbev/msab199.
49. Bushnell B, BBMap;. https://sourceforge.net/projects/bbmap/.
50. Stiehler F, Steinborn M, Scholz S, Dey D, Weber APM, Denton AK. Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning;36(22):5291–5298. https://doi.org/10.1093/bioinformatics/btaa1044.
51. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics;20(1):238. https://doi.org/10.1186/s13059-019-1832-y.
52. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties;33(18):2938–2940. https://doi.org/10.1093/bioinformatics/btx364.
53. Harris RS, IMPROVED PAIRWISE ALIGNMENT OF GENOMIC DNA;.
54. Sahu SK, Thangaraj M, Kathiresan K. DNA Extraction Protocol for Plants with High Levels of Secondary Metabolites and Polysaccharides without Using Liquid Nitrogen and Phenol;2012:205049. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4890884/.
55. Chen T. Simple and Scalable Genome Analysis with Transposase Enzyme Linked Long-Read Sequencing (TELL-Seq): From Haplotype Phasing to De Novo Assembly in a Tube;30:S37. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6938038/.
56. Kyriakidou M, Tai HH, Anglin NL, Ellis D, Strömvik MV. Current Strategies of Polyploid Plant Genome Sequence Assembly;9. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2018.01660.
57. Jeon D, Kim C. Polyploids of Brassicaceae: Genomic Insights and Assembly Strategies;13(15):2087. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11314605/.
58. Emonet A, Awad M, Tikhomirov N, Vasilarou M, Pérez-Antón M, Gan X, et al. Polyploid genome assembly of Cardamine chenopodiifolia;2024:0–0. https://gigabytejournal.com/articles/145, publisher: GigaScience Press.
59. Kersey PJ. Plant genome sequences: past, present, future;48:1–8. https://www.sciencedirect.com/science/article/pii/S1369526618300918.
60. Huang P, Li Z, Wang H, Huang J, Tan G, Fu Y, et al. A genome assembly of decaploid Houttuynia cordata provides insights into the evolution of Houttuynia and the biosynthesis of alkaloids;11(9):uhae203. https://doi.org/10.1093/hr/uhae203.
61. Abbas Q, Wilhelm M, Kuster B, Poppenberger B, Frishman D. Exploring crop genomes: assembly features, gene prediction accuracy, and implications for proteomics studies;25(1):619. https://doi.org/10.1186/s12864-024-10521-w.
62. Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, et al. Reciprocal Silencing, Transcriptional Bias and Functional Divergence of Homeologs in Polyploid Cotton (Gossypium);182(2):503–517. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2691759/.
63. Adams KL, Wendel JF. Polyploidy and genome evolution in plants;8(2):135–141. https://www.sciencedirect.com/science/article/pii/S1369526605000051.
64. Millar AH. Location, location, location: surveying the intracellular real estate through proteomics in plants;31(6):563–572. https://www.publish.csiro.au/fp/FP04034, publisher: CSIRO PUBLISHING.
65. Bhattacharya O, Ortiz I, Hendricks N, Walling LL. The tomato chloroplast stromal proteome compendium elucidated by leveraging a plastid protein-localization prediction Atlas;14:1020275.
66. Christian R, Labbancz J, Usadel B, Dhingra A. Understanding protein import in diverse non-green plastids;14:969931.
67. Bolger ME, Arsova B, Usadel B. Plant genome and transcriptome annotations: from misconceptions to simple solutions;19(3):437–449. https://doi.org/10.1093/bib/bbw135.
68. Rogers SO, Bendich AJ. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer;9(5):509–520. https://doi.org/10.1007/BF00015882.
69. Wicke S, Costa A, Muñoz J, Quandt D. Restless 5S: The re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants;61(2):321–332. https://www.sciencedirect.com/science/article/pii/S1055790311003125.
70. Hemleben V, Grierson D, Borisjuk N, Volkov RA, Kovarik A. Personal Perspectives on Plant Ribosomal RNA Genes Research: From Precursor-rRNA to Molecular Evolution;12. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2021.797348/full, publisher: Frontiers.
71. Mohannath G, Pontvianne F, Pikaard CS. Selective nucleolus organizer inactivation in Arabidopsis is a chromosome position-effect phenomenon;113(47):13426–13431. https://www.pnas.org/doi/full/10.1073/pnas.1608140113, publisher: Proceedings of the National Academy of Sciences.
72. Fultz D, McKinlay A, Enganti R, Pikaard CS. Sequence and epigenetic landscapes of active and silent nucleolus organizer regions in Arabidopsis;9(44):eadj4509. https://www.science.org/doi/full/10.1126/sciadv.adj4509, publisher: American Association for the Advancement of Science.
73. Pikaard CS. Nucleolar dominance: uniparental gene silencing on a multi-megabase scale in genetic hybrids;43(2):163–177. https:

//doi.org/10.1023/A:1006471009225.
74. Mohanta TK, Mishra AK, Hashem A, Abd_Allah EF, Khan AL, Al-Harrasi A. Construction of anti-codon table of the plant kingdom and evolution of tRNA selenocysteine (tRNASec);21(1):804. https://doi.org/10.1186/s12864-020-07216-3.
75. Monloy KC, Planta J. tRNA gene content, structure, and organization in the flowering plant lineage;15. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1486612/full, publisher: Frontiers.
76. Hui JHL, Chan TF, Chan LL, Cheung SG, Cheang CC, Fang JKH, et al. Genome assembly of the rare and endangered Grantham's camellia, Camellia granthamiana;2024:1–12. https://gigabytejournal.com/articles/124, publisher: GigaScience Press.
77. SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, et al. Nested retrotransposons in the intergenic regions of the maize genome;274(5288):765–768.
78. Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B. Analysis of a contiguous 211 kb sequence in diploid wheat (Triticum monococcum L.) reveals multiple mechanisms of genome evolution;26(3):307–316.
79. Moreno-Aguilar MF, Inda LA, Sánchez-Rodríguez A, Arnelas I, Catalán P. Evolutionary Dynamics of the Repeatome Explains Contrasting Differences in Genome Sizes and Hybrid and Polyploid Origins of Grass Loliinae Lineages;13:901733. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9284676/.
80. Michael TP. Plant genome size variation: bloating and purging DNA;13(4):308–317.
81. Casola C, Lawing AM. The nonrandom evolution of gene families;106(1):14–17. https://onlinelibrary.wiley.com/doi/abs/10.1002/ajb2.1211, _eprint: https://bsapubs.onlinelibrary.wiley.com/doi/pdf/10.1002/ajb2.1211.
82. Ruan X, Wang Z, Su Y, Wang T. Population Genomics Reveals Gene Flow and Adaptive Signature in Invasive Weed Mikania micrantha;12(8):1279.
83. Jiang J, Birchler JA, Parrott WA, Kelly Dawe R. A molecular view of plant centromeres;8(12):570–575. https://www.sciencedirect.com/science/article/pii/S1360138503002747.

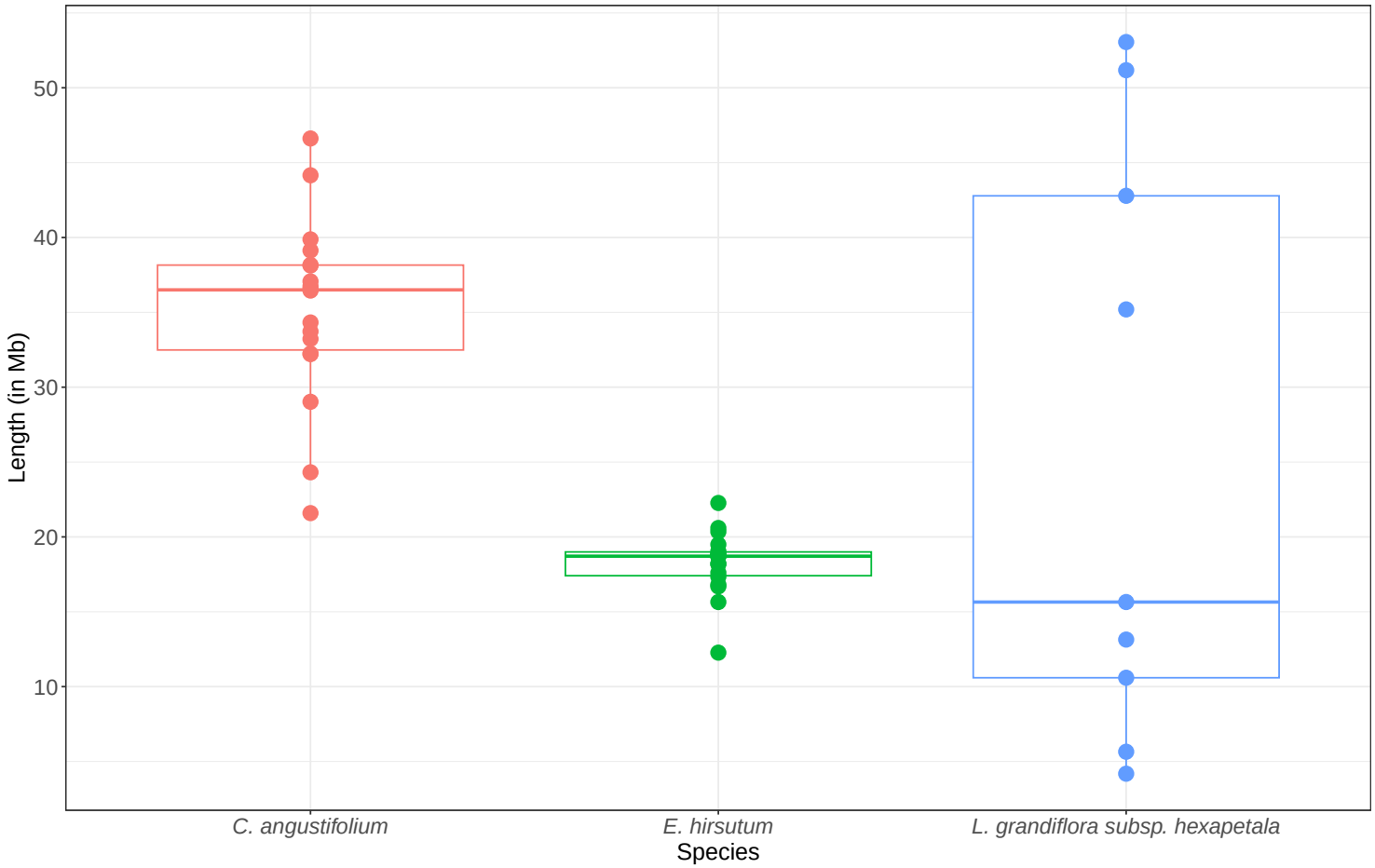


Figure S1: Chromosome/scaffold length comparison between *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*), *Chamaenerion angustifolium* and *Epilobium hirsutum*. Only scaffolds bigger than 4 Mb were represented for *Lgh*.

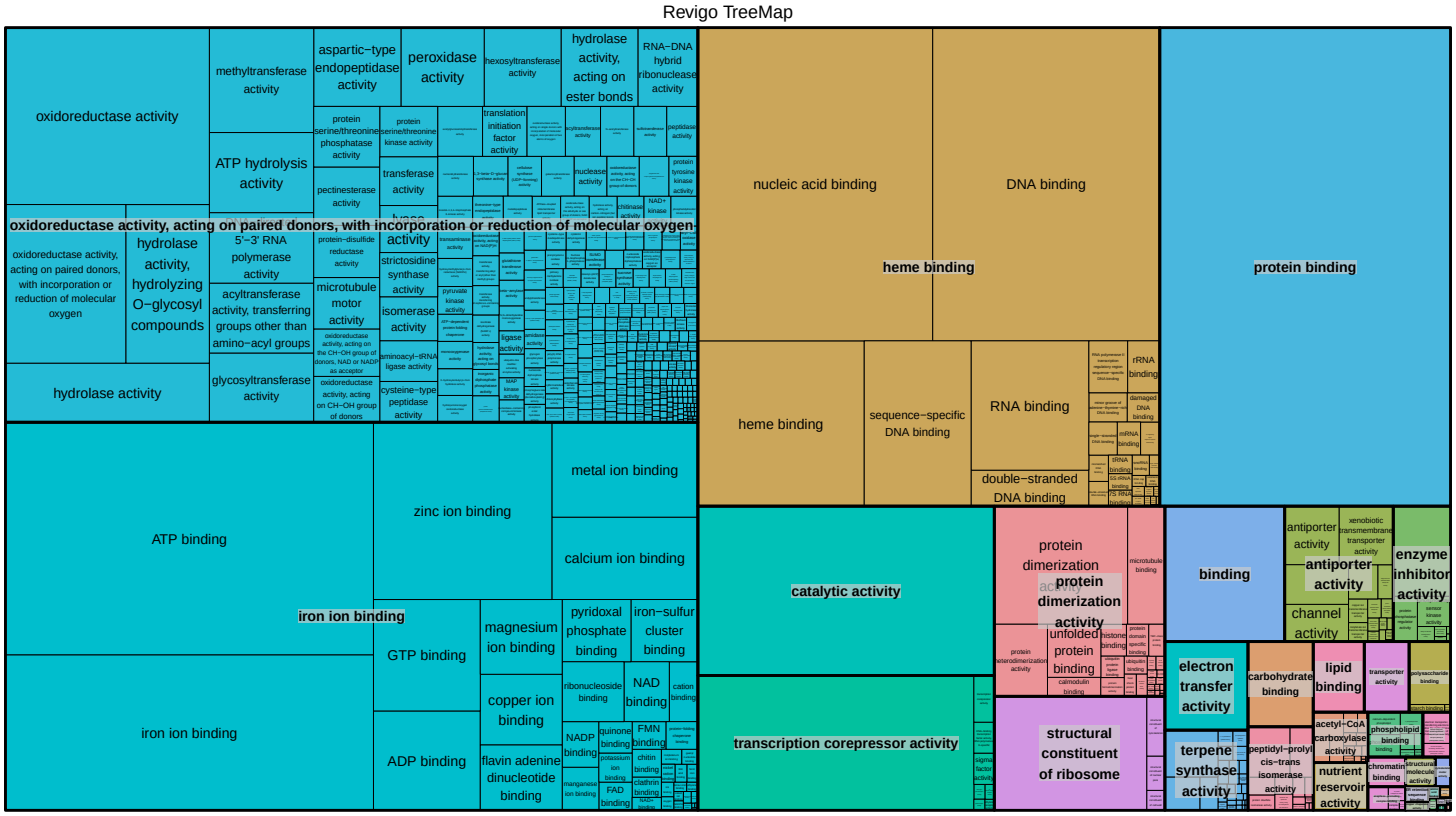


Figure S2: REVIGO treemap of GO Molecular function terms based on occurrence frequency in *Ludwigia grandiflora* subsp. *hexapetala*.

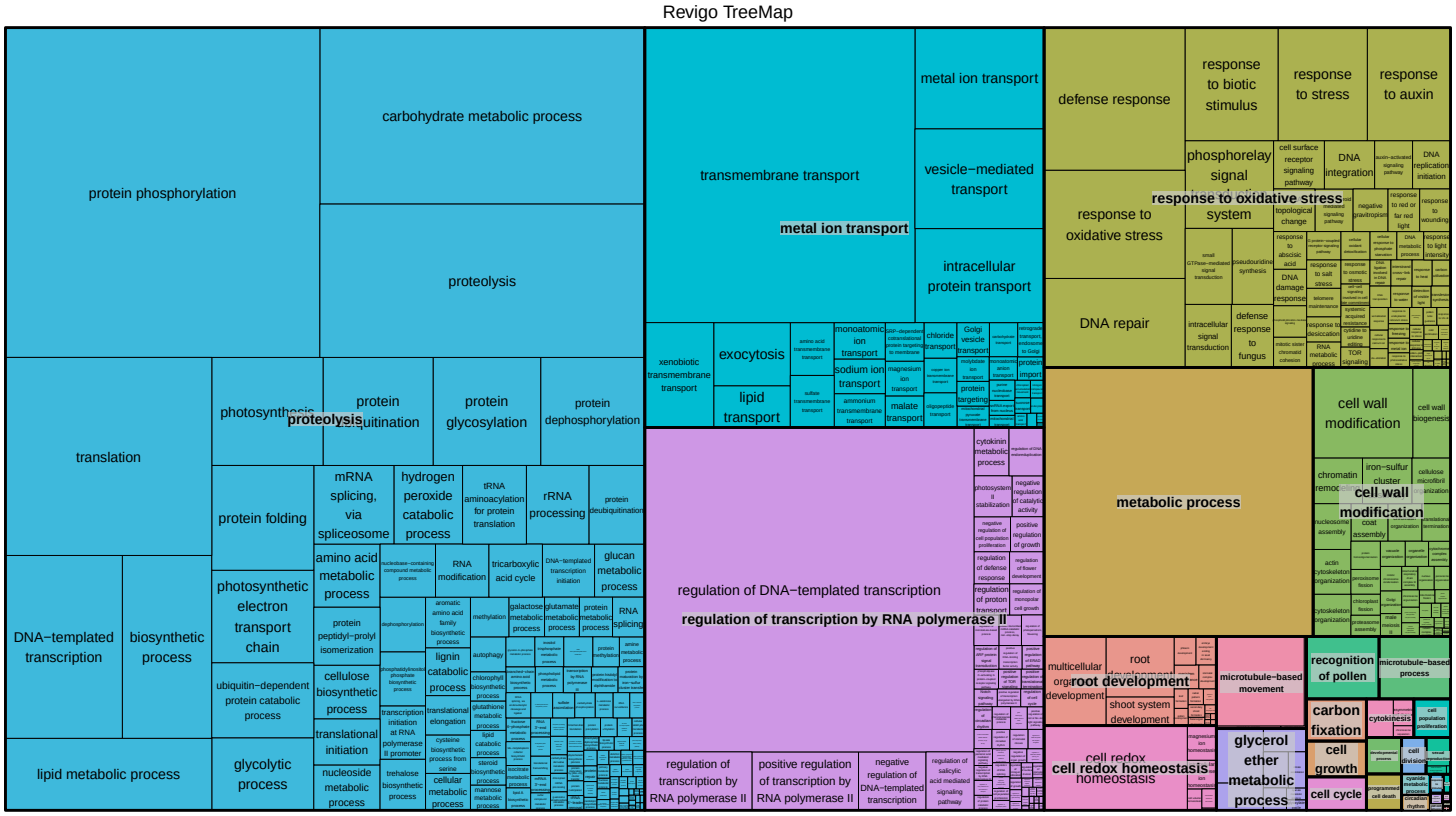


Figure S3: REVIGO treemap of GO Biological process terms based on occurrence frequency in *Ludwigia grandiflora* subsp. *hexapetala*.

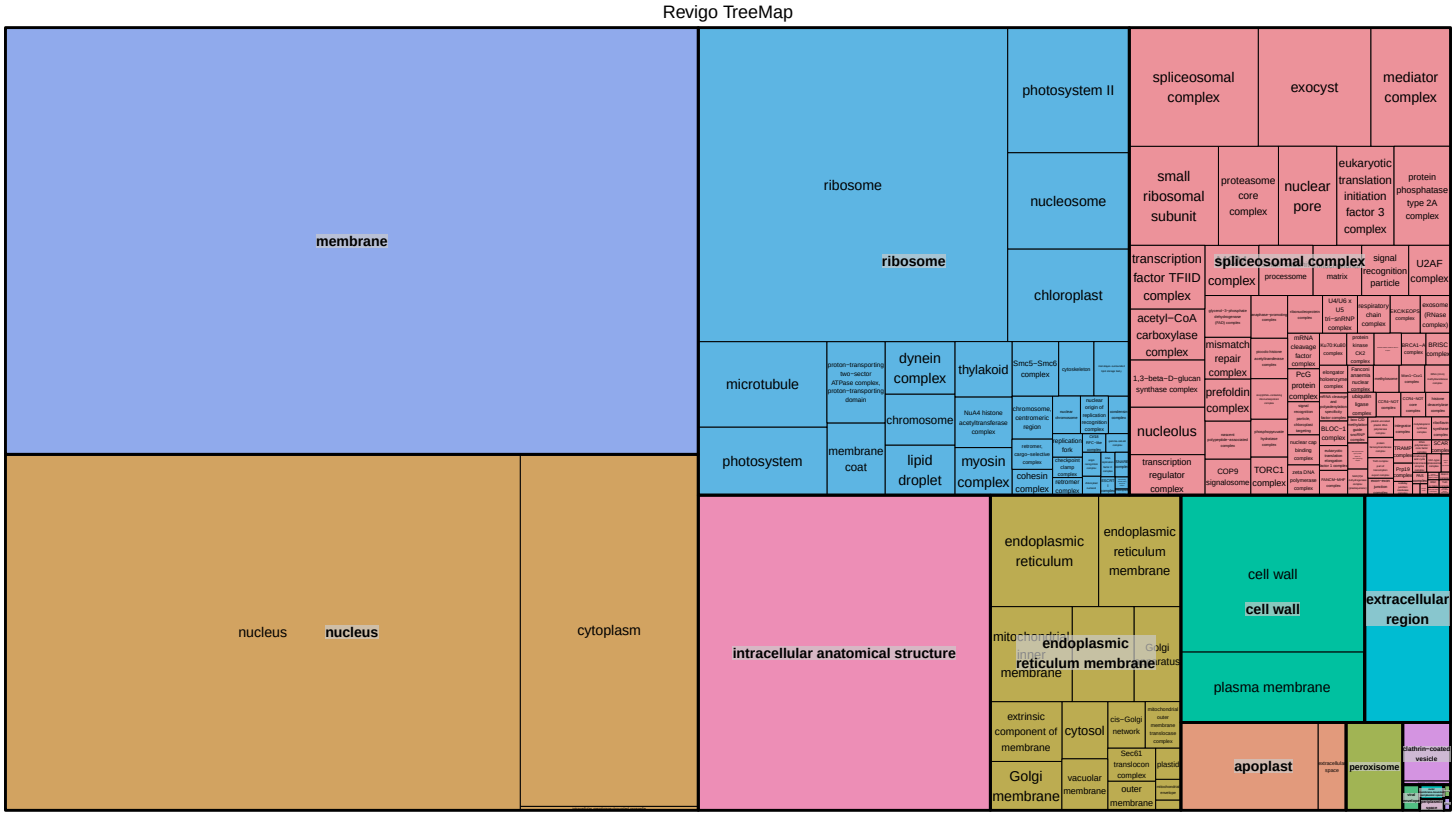


Figure S4: REVIGO treemap of GO Cellular component terms based on occurrence frequency in *Ludwigia grandiflora* subsp. *hexapetala*.

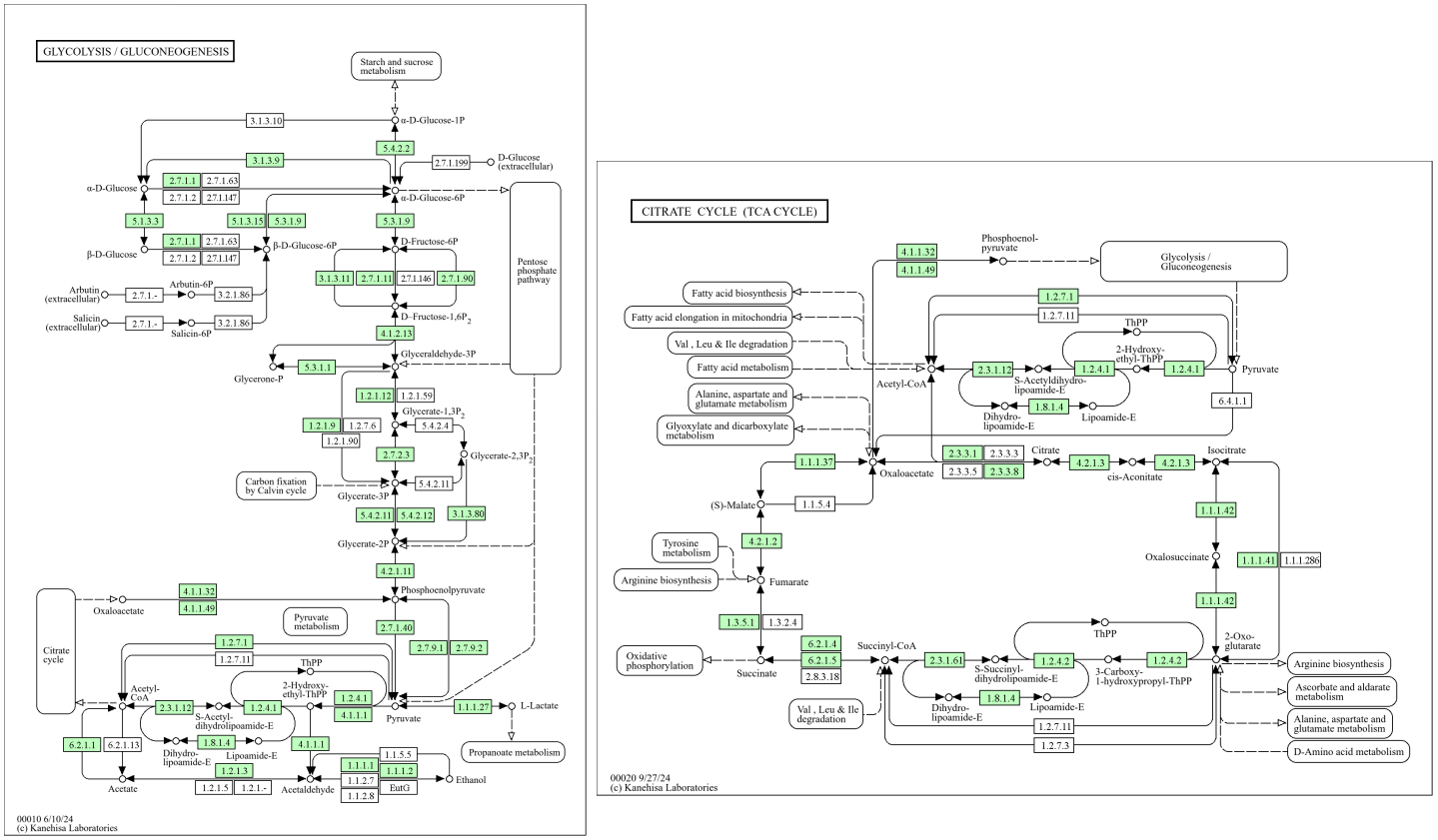


Figure S5: Glycolysis and citrate cycle genes in *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) genome by KEGG annotations. Genes in green are present in *Lgh* genome.

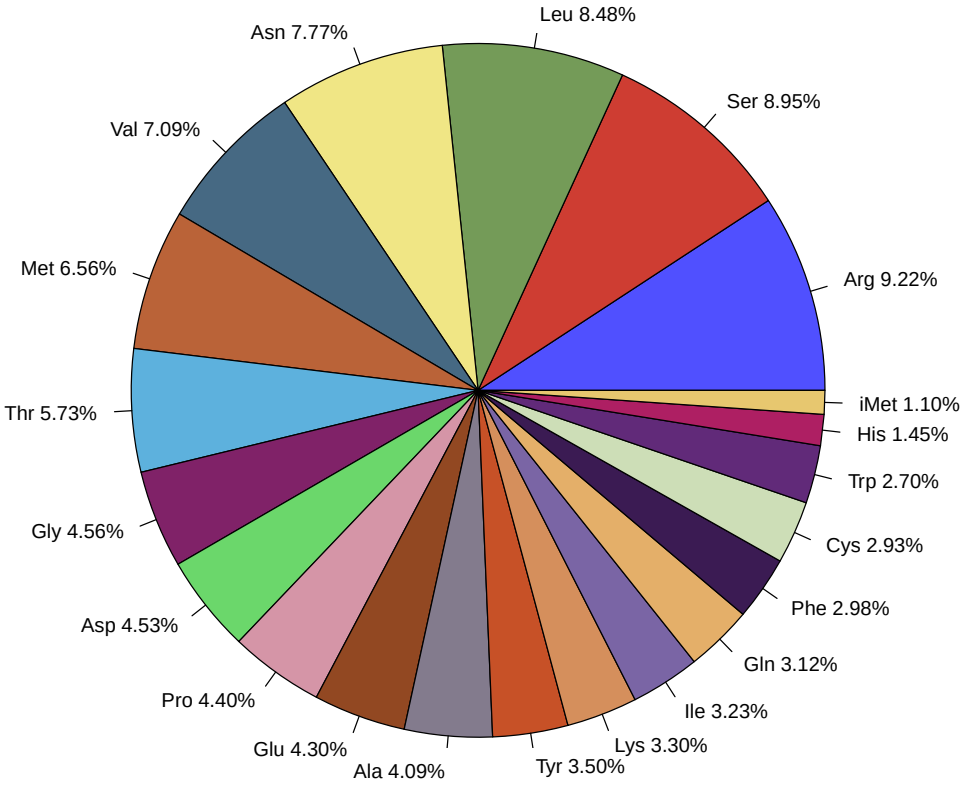


Figure S6: Transfer RNA gene repartition by isotype in *Ludwigia grandiflora* subsp. *hexapetala* genome.

Table S1: Information about genomes used in this study.

Species	NCBI Accession	Annotation source	Taxonomy (Order/Family)
Epilobium hirsutum	GCA_965153245.1	Annotated in this paper with Helixer [ref]	Myrtales/Onagraceae
Chamaenerion angustifolium	GCA_946814005.1	Annotated in this paper with Helixer [ref]	Myrtales/Onagraceae
Punica granatum	GCA_007655135.2	NCBI RefSeq	Myrtales/Lythraceae
Trapa natans	GCA_035582545.1	Submitter	Myrtales/Lythraceae
Eucalyptus grandis	GCA_016545825.1	NCBI RefSeq	Myrtales/Myrtaceae
Eucalyptus globulus	GCA_016545825.1	Submitter	Myrtales/Myrtaceae
Corymbia citriodora	GCA_016545825.1	Submitter	Myrtales/Myrtaceae
Rhodamnia argentea	GCA_016545825.1	NCBI RefSeq	Myrtales/Myrtaceae
Syzygium oleosum	GCA_016545825.1	NCBI RefSeq	Myrtales/Myrtaceae
Syzygium grande	GCA_016545825.1	Submitter	Myrtales/Myrtaceae
Melastoma candidum	GCA_016545825.1	Submitter	Myrtales/Melasomataceae
Gossypium arboreum	GCA_025698485.2	NCBI RefSeq	Malvales/Malvaceae
Citrus sinensis	GCA_022201045.1	NCBI RefSeq	Sapindales/Rutaceae
Arabidopsis thaliana	GCA_000001735.2	NCBI RefSeq	Brassicales/Brassicaceae

Table S2: Annotation statistics of genes in *Ludwigia grandiflora* subsp. *hexapetala* genome.

Annotation type	**Number of annotation**
Total genes	270,493
rRNA genes	15,834
5S rRNA	2,018
5.8S rRNA	3,939
18S rRNA	6,237
28S rRNA	3,640
tRNA genes	8,417
Protein coding genes	246,242
Unique protein coding genes	223,161
Duplicated protein coding genes	14,247
Single copy protein coding genes	208,914
Blast annotations	103,150
GO annotations	55,954
Interpro annotations	79,621
KEGG annotations	39,714
Uncharacterized proteins*	77,391
Hypothetical proteins**	116,869

*"Uncharacterized proteins" here correspond to proteins that had a Blast annotations but without a defined name.

**"Hypothetical proteins" here correspond to proteins that didn't get any of Blast, GO, Interpro or KEGG annotations.

Table S3: Predicted subcellular localizations of proteins in *Ludwigia grandiflora* subsp. *hexapetala* genome by WoLF PSORT.

Subcellular localization	**Number of proteins**
Chloroplast	84,283
Nucleus	78,267
Cytosol	48,179
Mitochondria	17,927
Extracellular	8,201
Plasma membrane	6,086
Vacuolar membrane	2,469
Endoplasmic reticulum	1,172
Cytoskeleton	992
Golgi apparatus	627
Peroxisome	562

Table S4: Repeated elements in *Ludwigia grandiflora* subsp. *hexapetala* genome. Some transposable elements have multiple family annotations due to ambiguous domain, which explains why the total sum of TE is different from the addition of all TE families.

Classification	**Number**	**Length**
Simple Sequence Repeats (SSR)	625,820	26,674,375
Tandem repeats	388,071	40,880,736
All transposable elements	140,494	57,451,184
Class I/LINE	9,434	4,377,968
Class I/pararetrovirus	16	3,231
Class I/LTR/Ty1-copia	28,730	10,609,375
Class I/LTR/Ty3-gypsy	98,220	40,151,004
Class II/Subclass 1/TIR	7,070	2,904,260
Class II/Subclass 2/Helitron	797	267,502

Table S5: Number of genes in orthogroups by species. Core refers to genes in orthogroups present in all species.

Species	**Core**	**Myrtales order only**	**Family specific**	**Species specific**	**Others**	**Total**
Arabidopsis thaliana	25844	0	0	6047	13682	45573
Chamaenerion angustifolium	21308	181	370	2360	12589	36808
Corymbia citriodora	25001	327	615	916	16232	43091
Citrus sinensis	22534	0	0	3085	14021	39640
Eucalyptus globulus	22533	378	663	505	16355	40434
Eucalyptus grandis	25243	528	836	847	16553	44007
Epilobium hirsutum	18482	161	180	261	8593	27677
Gossypium arboreum	29580	0	0	2611	15570	47761
Ludwigia grandiflora subsp. *hexapetala*	49488	517	670	77448	57753	185876
Melastoma candidum	26044	204	0	3140	12145	41533
Punica granatum	22257	247	152	1131	12124	35911
Rhodamnia argentea	25413	267	450	206	12196	38532
Syzygium grande	17089	209	352	2404	12161	32215
Syzygium oleosum	24662	362	545	313	14327	40209
Trapa natans	20705	142	102	409	7696	29054

Chapitre 3

Assemblage du transcriptome de *Lgh* et analyses différentielles

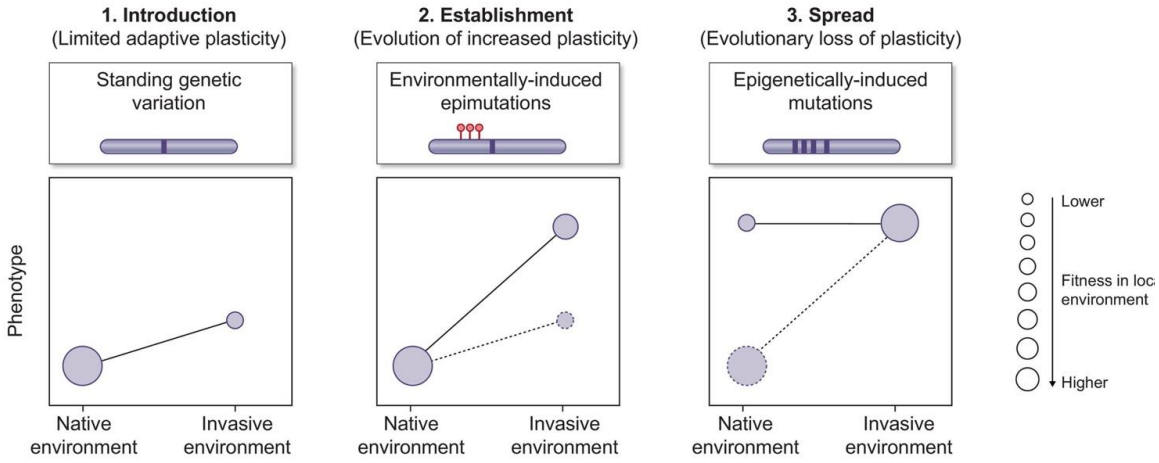


FIGURE 3.1 – Évolution de la plasticité phénotypique au cours des invasions [Hodgins, Battlay et Bock, 2025]. Aux différentes phases du processus d'invasion sont associées des épi(génotypes) hypothétiques à un locus de plasticité phénotypique le long d'un chromosome (en haut) et les changements de plasticité attendus (en bas) c'est-à-dire les variations des normes de réaction d'un phénotype dans des environnements contrastés. Les normes de réaction caractéristiques de la phase précédente sont illustrées par des lignes pointillées.

1 Introduction

1.1 Plasticité phénotypique et invasions biologiques

La plasticité phénotypique est la capacité d'un génotype donné à produire des phénotypes alternatifs en réponse à son environnement de développement [Via et Lande, 1985]. Les mécanismes génétiques et épigénétiques peuvent modifier l'expression des gènes et donc, à le fois, la capacité d'adaptation locale et la plasticité phénotypique [Marin et al., 2020]. Dans le cadre d'une invasion biologique, la plasticité phénotypique va évoluer et peut être diviser en trois phases (Figure 3.1) [Hodgins, Battlay et Bock, 2025]. La phase 1 a lieu au moment de l'introduction, le niveau de plasticité est modéré et provient de mutation génétique. En phase 2, des variations épigénétiques héritables induites par l'environnement vont apparaître montrant une plasticité accrue pendant l'établissement de l'espèce invasive. Enfin, en phase 3, si le maintien de cette réponse plastique a un coût, il y aura alors une perte des mutations induites épigénétiquement menant à une diminution de la plasticité à mesure que l'invasion progresse. Par exemple, après une période de présence de 10 ans dans leur zone d'invasion, des populations d'*Ambrosia trifida* ont vu leur plasticité phénotypique se réduire, permettant une croissance végétative plus vigoureuse [Savić et al., 2024]. Les auteurs de cette étude ont suggéré que l'héritabilité de traits favorables permettait l'adaptation locale.

La génomique peut permettre de comprendre les mécanismes impliqués dans l'évolution de la plasticité au cours des invasions. De plus, les analyses combinées du transcriptome et du métabolome peuvent être utilisées pour interpréter les différences moléculaires entre des plantes présentant des phénotypes différents résultant de l'expression de la plasticité. L'exploitation des données transcriptomiques pour comprendre l'invasion est un domaine émergent, et de plus en plus d'études génèrent des données transcriptomiques afin de constituer les ressources nécessaires à leur réutilisation future

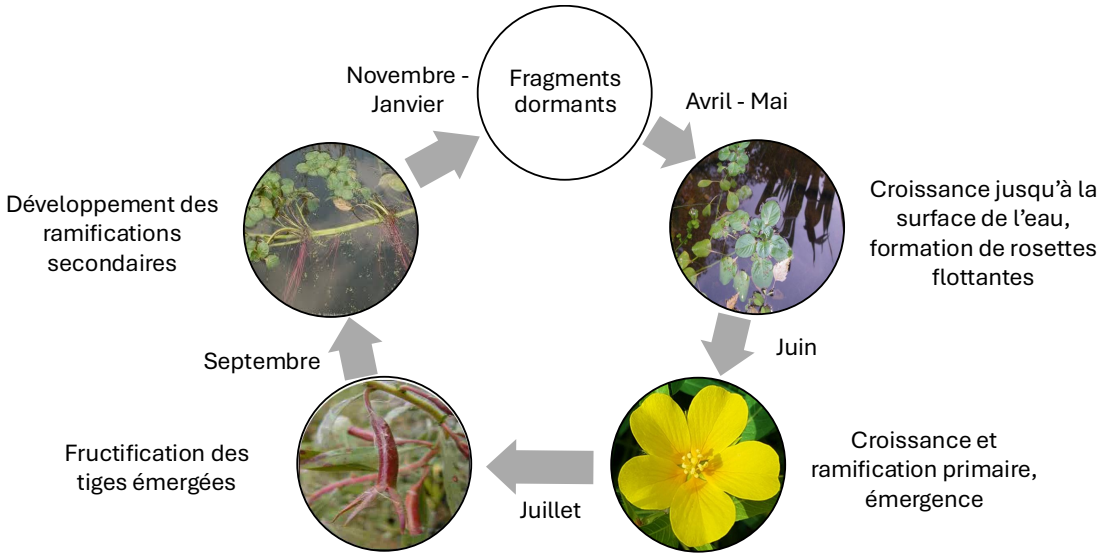


FIGURE 3.2 – Cycle de développement annuel de la jussie (d'après Ruaux, 2008 et Dandelot, Robles et al., 2008). La jussie est en état de dormance durant la période hivernale et à partir des tiges souterraines, graines et fragments. Elle reprend sa croissance dès avril. Durant le printemps, elle croit jusqu'à produire des rosettes à la surface de l'eau. Puis à partir de juin, elle produit des tiges érigées qui permettront de juillet à septembre d'effectuer la floraison et la fructification. Enfin, de septembre à novembre, elle continue sa croissance avant de redémarrer un nouveau cycle. Ce cycle est modifié pour les formes terrestres qui subissent une alternance des niveaux d'eau durant l'année. En effet, de novembre à juin la jussie terrestre sera sous l'eau alors que de juin à octobre, elle sera complètement terrestre.

[Maldonado et al., 2020].

1.2 Acclimatation de la jussie au milieu terrestre

Cette approche fonctionnelle (transcriptomique) pourrait permettre de mieux comprendre le passage de *Lgh* du milieu aquatique au milieu terrestre et ainsi son acclimatation. En effet, il a été montré que le morphotype aquatique de *Lgh* présentait une plasticité phénotypique plus importante que celle présentée par le morphotype terrestre alors que le morphotype terrestre avait des valeurs de traits supérieures à celles du morphotype aquatique, et cela aussi bien dans des habitats terrestres qu'aquatiques [Genitoni, Vassaux, Renault et al., 2025]. L'acclimatation de *Lgh* au milieu terrestre semble assez rapide car au marais de Mazerolles, la forme terrestre est apparue 5 ans après l'arrivée de la jussie en milieu aquatique [Haury, Noël et al., 2011]. *Lgh* présente des cycles de développement différents entre les deux morphotypes. En effet, la jussie dans sa forme aquatique a une partie de son appareil végétatif continuellement immergé tout au long de l'année et seule une partie de l'appareil aérien est émergée à partir du printemps lors d'apparition de tiges érigées (Figure 3.2). Dans sa forme terrestre, la jussie passe 6 mois complétement immergée et 6 mois totalement émergée, lorsque l'eau se retire des prairies inondées ce qui pourrait avoir favorisé l'apparition du morphotype terrestre [Genitoni, 2019].

1.3 Objectifs et hypothèses

L'objectif principal du chapitre est l'assemblage d'un transcriptome de *Lgh* et d'identifier des gènes différentiellement exprimés suivant des conditionnement préalable et le morphotype lors d'une acclimatation au milieu terrestre. Ce transcriptome va permettre de remplir plusieurs buts :

— (1) Compléter l'assemblage et l'annotation du génome de *Lgh*. Nous faisons l'hy-

pothèse que le génome de *Lgh* n'est pas complet et ne contient pas tous les gènes. Ainsi l'obtention d'un transcriptome assemblé de *Lgh* permettra de le compléter [Voir Chapitre 2]. De plus, le transcriptome pourrait permettre d'identifier des nouvelles régions transcrites non annotées dans le génome comme des nouveaux gènes codants ou non codant (ARN non codants) ainsi que les régions régulatrices (5' et 3' UTR ou ilôts CpG) appartenant à des gènes déjà annotés [Hoff et Stanke, 2015].

— (2) Comparer l'assemblage d'un transcriptome *de novo* et sur référence en utilisant le génome de *Lgh*. Nous formulons l'hypothèse que les deux ressources seront différentes et complémentaires [Lee, Na et Park, 2021].

— (3) Identifier des gènes ayant une expression spécifique en fonction du tissu. Nous pensons que l'expression d'un grand nombre de gènes dépendra du tissu étudié [Yaschenko et al., 2022]. Dans notre étude, il s'agira des gènes exprimés dans les racines et ceux exprimés dans les tiges/feuilles.

— (4) Tester si l'alternance de conditions immergées et terrestres est un des facteurs environnementaux ayant favorisé l'apparition du morphotype terrestre de *Lgh*. L'hypothèse est qu'en condition terrestre, le morphotype aquatique soumis à plusieurs cycles d'alternances aura des réponses fonctionnelles similaires à celle du morphotype terrestre [Genitoni, 2019].

2 Matériels et méthodes

2.1 Matériel végétal

Des plantes de *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) des morphotypes aquatique et terrestre ont été collectés au marais de Mazerolles, près de Nantes en septembre 2017 ; dans 5 quadrats de 0,5 m2 distancés de 10 mètres les uns par rapport aux autres.

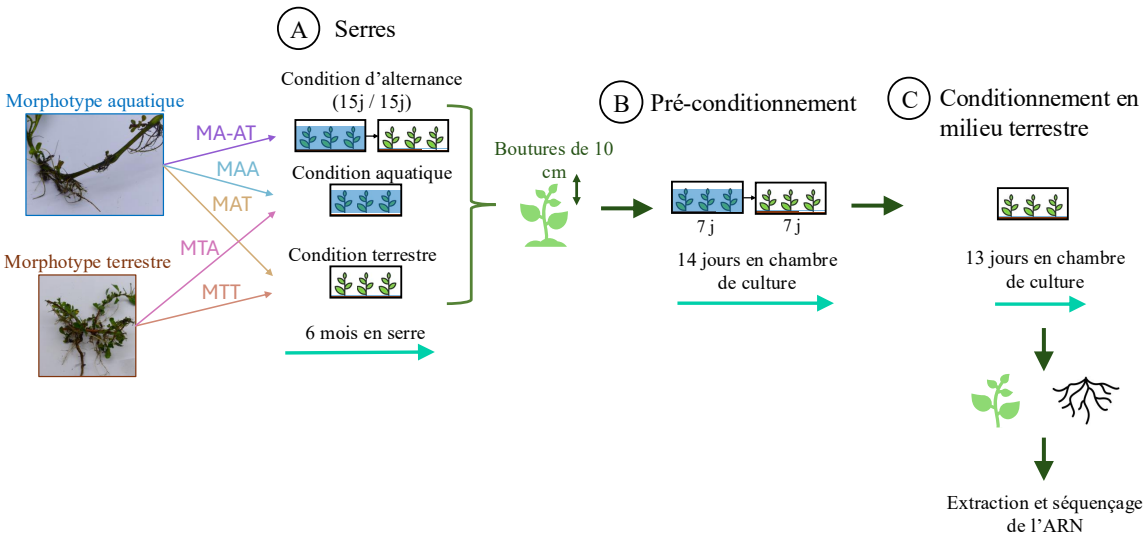


FIGURE 3.3 – Dispositif expérimental pour le séquençage de l'ARN de *Ludwigia grandiflora* subsp. *hexapetala*.

Ces plantes ont été élevées en serre froide et conditionnées soit dans leur condition d'origine terrestre ou aquatique (MTT = Morphotype Terrestre en condition Terrestre et MAA = Morphotype Aquatique en condition Aquatique), soit en conditions croisées (MTA = Morphotype Terrestre en condition Aquatique et MAT = Morphotype Aquatique en condition Terrestre) (Figure 3.3). De plus, des plantes du morphotype aquatique ont été élevées en conditions d'alternance ; à savoir 15 jours en condition aquatique suivi de 15 jours en conditions terrestre ce qui correspond à 1 cycle d'alternance condition aquatique/condition terrestre (Figure 3.3). Ce morphotype aquatique ayant subi des cycles d'alternance sera appelé MA-AT.

Au bout de 6 mois, des boutures ont été prélevées et normalisées à 10 cm en partant de l'apex pour chacun des morphotypes x conditions décrits ci-dessus (Figure 3.3). Une étape de pré-conditionnement des fragments a été réalisée par un passage de 7 jours en condition aquatique (immersion des plantes) suivi de 7 jours en condition terrestre de 7 jours (émersion des plantes, avec un niveau d'eau affleurant afin de simuler un sol hydromorphe [Genitoni, 2019] en chambre de culture (22°C, photopériode 16h/8h, jour/nuit)). Cette étape de pré-conditionnement permet aux plantes de jussie de former des racines et de s'acclimater à la condition terrestre. A l'issue de ce pré-traitement, les plantes ont été placées en condition terrestre pendant encore 13 jours dans la même chambre de culture. La condition terrestre a été simulée par un apport initial de 120 mL d'eau et un ajout de 30 mL tous les 3 jours. Pour le prétraitement et le conditionnement en milieu terrestre, les plantes ont été élevées individuellement dans des conteneurs en plastique contenant un substrat.

2.2 Extraction de l'ARN

Les racines et les parties aériennes (tiges et feuilles) de 6 plantes ont été prélevées séparément et plongées directement dans l'azote liquide. Trois réplicats biologiques ont

TABLE 3.1 – Explication des dénominations des échantillons séquencés.

Morphotype	Condition	Tissu	Nom abrégé	Nombre de banque
Aquatique	**A**quatique	**T**iges-**F**euilles	M**AA_TF**	1
Aquatique	**A**quatique	**R**acines	M**AA_R**	3
Aquatique	Alternance (**A**quatique-**T**errestre)	**T**iges-**F**euilles	M**A-AT_TF**	3
Aquatique	Alternance (**A**quatique-**T**errestre)	**R**acines	M**A-AT_R**	3
Aquatique	**T**errestre	**T**iges-**F**euilles	M**AT_TF**	3
Aquatique	**T**errestre	**R**acines	M**AT_R**	3
Terrestre	**A**quatique	**T**iges-**F**euilles	M**TA_TF**	3
Terrestre	**A**quatique	**R**acines	M**TA_R**	3
Terrestre	**T**errestre	**T**iges-**F**euilles	M**TA_TF**	3
Terrestre	**T**errestre	**R**acines	M**TT_R**	3

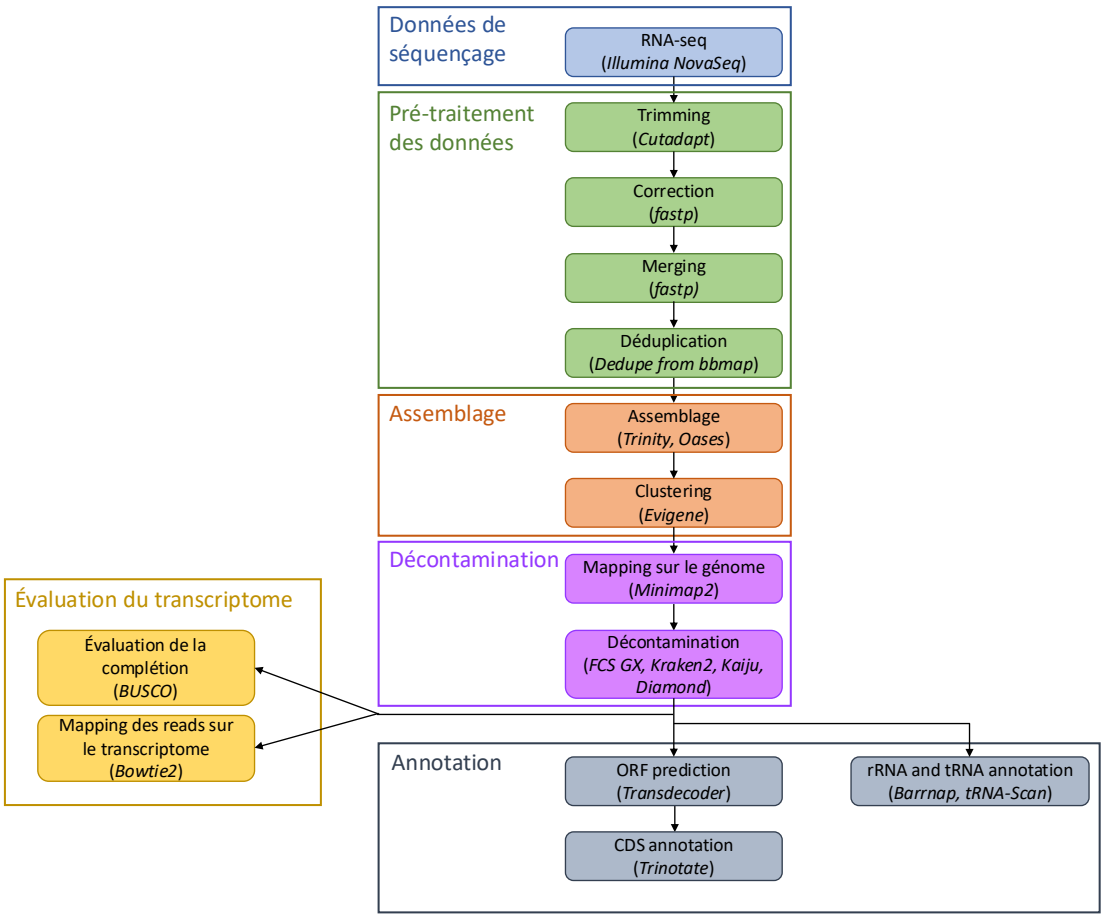


FIGURE 3.4 – Pipeline de l'assemblage *de novo* du transcriptome de *Lgh*.

été réalisés pour chaque condition/morphotype (Table 3.1). La présence importante de polyphénols et de produits phénoliques comme des tanins chez la jussie a conduit à une mise au point méthodologique d'extraction des ARNs qui a demandé 18 mois. Le protocole proposé par Chang et al. (1993) qui a travaillé sur le pin, espèce récalcitrante à l'extraction des ARNs de par sa richesse en polyphénols et produits phénoliques a été utilisé pour extraire les ARNs. Puis, une étape de purification des ARNs a été réalisée en utilisant le kit Nucleospin RNA de Macherey Nagel. Ces extractions ont permis l'obtention de 15 banques pour les racines et 13 banques pour les "tiges/feuilles", un seul échantillon (pas de réplicat biologique) ayant pu être séquencé pour le morphotype aquatique en condition aquatique (MAA) (Table 3.1). Ces banques ont été séquencées sur la plateforme Montpellier GenomiX (MGX) *via* la réalisation de deux runs NovaSeq d'Illumina (2*150 pb). Pour la réalisation du transcriptome de référence de *Lgh*, l'ensemble des données de séquençage ont été poolées pour l'assemblage *de novo* et pour l'analyse différentielle, chaque banque séquencée a été analysée séparément.

2.3 Assemblage *de novo* du transcriptome

Le pipeline montrant les différentes étapes du trimming des séquences à l'annotations du transcriptome est présenté en figure 3.4. Les adaptateurs présents dans les reads d'ARN ont été trimmés par Cutadapt v1.15 (paramètre : -m 2 et séquences adaptateurs : R1 : "AGATCGGAAGAGCACACGTCTGAACTCCAGTCA" ; R2 : "AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT") [Martin, 2011]. Fastp v0.23.4 a été utilisé pour corriger et merger les reads (paramètres : –disable_adapter_trimming –dedup -l 6 –trim_poly_g –trim_poly_x –average_qual 20 –cut_right –cut_right_mean_quality 20 -c –overlap_len_require 10 -m) [Chen, 2023]. Les reads mergés et non mergés ont été gardés pour l'assemblage. Ils ont tous été regroupés en un seul fichier puis dédupliqués avec dedupe de BBMap [Bushnell, 2024].

Les reads ont été assemblé par Trinity v2.15.1 (paramètre : –run_as_paired) et Velvet-Oases v0.2.09 (avec des tailles de k-mer de 49 et 73 et -min_trans_lgth 200) [Grabherr et al., 2011 ; Schulz et al., 2012]. Le script tr2aacds4.pl de EvidentialGene (dernière mise à jour : 15-07-23) a été utilisé pour réduire la redondance et identifier les isoformes des différents assemblages [Gilbert, 2016].

Pour les décontaminer, les transcrits ont été d'abord mappés sur le génome de *Lgh* avec minimap2 v2.15 (option : –ax splice) [Li, 2018]. Les transcrits qui ont mappés ont été gardés comme non contaminants. Les autres ont été classifiés d'abord par FCS GX v0.5.5 (taxonomic identifier : 1450660) sur Galaxy org [Astashyn et al., 2023 ; The Galaxy Community, 2024]. Les transcrits classés comme « primary-div » sont gardés pour *Lgh* et les transcrits classés comme « contaminant » sont retirés. Les transcrits non classifiés par FCS GX sont analysés par Kraken2 v2.14 dans deux bases de données (nt_core et plusPFP) [Wood, Lu et Langmead, 2019]. Les transcrits classifiés en Viridiplantae (taxid 33090) sont gardés pour *Lgh* et les transcrits classifiés comme non-Viridiplantae sont retirés. Les transcrits non classifiés par Kraken2 sont analysés par Kaiju v1.10.1 dans deux bases de données (nr_euk et refseq_nr) [Menzel, Ng et Krogh, 2016]. Les transcrits classifiés sont retirés et considérés comme contaminants. Enfin, les transcrits toujours sans classification sont analysés par Diamond Blastx v2.1.11 (paramètres : Base de données nr, maximum de séquences par transcrits dont l'alignement sera reporté = 10, format d'output = 6 (tableau tsv) avec skingdoms) [Buchfink, Reuter et Drost, 2021]. Les transcrits sont gardés quand ils ont soit uniquement une classification Streptophyte (skingdoms = Streptophytina) ou au moins 2 classifications plantes comptant pour au moins 50% de toutes les classifications pour la séquence.

La complétude de l'assemblage a été évalué par BUSCO (base de données : eudicotylédones) [Manni et al., 2021]. Tous les reads ont été mappés sur l'assemblage final par Bowtie2 [Langmead et Salzberg, 2012]. Le N50 a été calculé avec le script statswrapper

de BBMap [Bushnell, 2024].

2.4 Annotations du transcriptome *de novo*

TransDecoder v5.7.1 a identifié les ORFs dans les transcrits [Haas, 2024]. Trinotate v4.0.2 (paramètres : –run "swissprot_blastp swissprot_blastx pfam signalp6 tmhmmv2 infernal EggnogMapper") a ensuite annoté le transcriptome de *Lgh*. GhostKoala a donné des annotations KEGG aux transcrits [Kanehisa, Sato et Morishima, 2016]. Enfin, les ARNr et ARNt ont été annoté par pybarrnap v0.5.1 (paramètre : –kingdom euk) [Shimoyama, 2024] et tRNAscan-SE v2.0.12 [Chan et al., 2021].

2.5 Analyse différentielle en utilisant le génome de *Lgh*

Pour cette expérience, les RNA-seq pairés et mergés des triplicats (sauf pour MAA_TF qui est unique) ont été poolés par conditions puis mappés sur le génome concatené, chaque contig ont été séparé artificiellement par un gap de 100 N, avec le mappeur minimap2 en mode spliced (-ax splice alignement aware de jonctions d'intron). Le comptage a été réalisé en TPM pour chaque condition avec Geneious.

Dans un premier temps, les données d'expression génique ont été préalablement transformées par log2 puis analysées à l'aide du modèle linéaire général (MLG) implémenté dans le package limma. Ce modèle linéaire a été ajusté à chaque gène en fonction du plan expérimental à l'aide de la fonction lmFit. Pour comparer les conditions d'intérêt, un contraste spécifique (tiges feuilles) a été défini à l'aide de la fonction makeContrasts. Les coefficients correspondants ont ensuite été estimés en appliquant le contraste au modèle ajusté *via* contrasts.fit. Enfin, la fonction eBayes a été utilisée pour modérer les variances des coefficients estimés, permettant ainsi une identification plus robuste des gènes différentiellement exprimés entre les conditions TF et R. Les différences d'expression entre les conditions ont été quantifiées à l'aide du logarithme du

fold change (logFC), qui représente le logarithme en base 2 du ratio d'expression d'un gène entre deux conditions. Les échantillons ont ensuite été comparés 2 à 2 en séparant les tissus. Pour identifier les gènes différentiellement exprimés de manière significative, un seuil de logFC ≥ 5 a été appliqué afin de sélectionner les gènes présentant une variation d'expression importante entre les conditions. Parallèlement, la significativité statistique a été évaluée à l'aide de la valeur p ajustée (adj.P.Val) $< 0{,}05$, obtenue après correction pour les tests multiples selon la méthode de Benjamini-Hochberg. Seuls les gènes respectant simultanément ces deux critères ont été considérés comme significativement différentiellement exprimés, garantissant à la fois l'importance biologique et la robustesse statistique des résultats.

2.6 Mapping et analyses différentielles en utilisant le transcriptome *de novo* de *Lgh*

Salmon v1.10.3 (paramètre : -l IU) [Patro et al., 2017] a été utilisé pour mapper les reads R1 et R2 corrigés de chaque échantillon sur le transcriptome "decoy" de *Lgh*. Il a été généré à l'aide du génome et du transcriptome de *Lgh* avec le script generateDecoyTranscriptome.sh de SalmonTools. Les résultats de Salmon ont été importé dans RStudio avec le package tximport [Soneson, Love et Robinson, 2016] afin d'obtenir des comptages par gène en utilisant les gènes ayant été identifié par EvidentialGene. Les gènes ARNr et ARNt ont été retirés de l'analyse. Ensuite, les analyses différentielles ont été réalisés à l'aide du package edgeR [Robinson, McCarthy et Smyth, 2010]. D'abord une normalisation avec la méthode TMM puis une comparaison globale de tous les échantillons a été effectué. Les enrichissement GO des gènes différentiellement exprimés dans les tissus ont été effectué en utilisant le package R clusterProfiler [Yu, 2024]. Ensuite, ils ont été séparés par type de tissus et normalisés pour effectuer les analyses différentielles.

3 Résultats et Discussion

3.1 Assemblage *de novo* du transcriptome de *Lgh*

Séquençage, correction et assemblage

Le séquençage NovaSeq a été réalisé sur les 28 banques produites et a permis d'obtenir 1 976 483 628 reads soit une moyenne de 70 588 701 reads par échantillon. Ces reads ont été trimmés par Cutadapt puis corrigés et dédupliqués par fastp aboutissant à 1 058 600 564 reads soit 37 807 163 reads par échantillon. Ensuite, 122 992 877 reads ont été obtenus après avoir été mergés, poolés puis dédupliqués par Dedupe. Ces reads ont été assemblés par Trinity et Oases, produisant, respectivement, 446 772 et 242 070 transcrits. Evigene a été utilisé pour réduire la redondance des assemblages et a sélectionné 249 124 transcrits.

Classification des transcrits

Un total de 174 487 transcrits ($\approx 70\%$) a été mappé sur le génome de *Lgh*. Les 74,637 reads non mappés sur le génome sont passé par plusieurs étapes de décontamination qui ont permis d'identifier 7604 transcrits de plantes et 57 561 transcrits contaminants en plus de 9472 transcrits restés sans identification. Parmi les séquences contaminantes identifiées par FCS GX, on trouve beaucoup de microorganismes comme des bactéries (6358) ou des protistes (2194). Les espèces de microorganismes les plus présentes sont des espèces de la flore humaine comme *Lactobacillus crispatus* ou *Escherichia coli* ainsi que des pathogènes des plantes comme des espèces du genre *Cladosporiaceae* ou *Phytopythium helicoides*. On trouve aussi des insectes comme *Aphis gossypii*, le puceron du cotonnier ou *Rhagoletis zephyria*, une mouche des fruits. Les transcrits correspondant à des plantes ont été conservés en plus des transcrits mappant sur le génome de *Lgh* donnant ainsi 182 091 transcrits pour l'assemblage final (Table 3.2).

TABLE 3.2 – Statistiques de l'assemblage du transcriptome de *Lgh*.

Statistiques	**Valeurs**
Nombre de transcrits assemblés	182 091
Nombre de bases totales (pb)	298 730 243
N50 (pb)	2304
Contenu en GC (%)	44,2
Longueur moyenne	1641
BUSCO score (%)	C : 98 [S : 6,9 ; D : 91,1] F : 0,5 ; M : 1,5

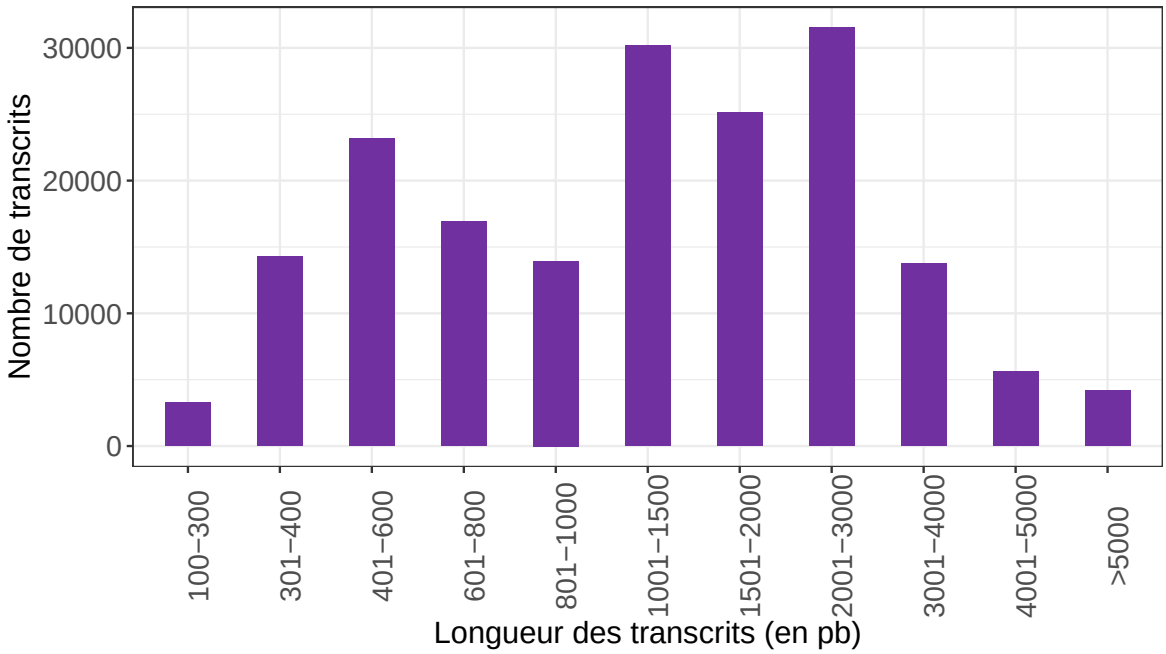


FIGURE 3.5 – Longueur des transcrits assemblés dans le transcriptome de *Lgh*.

Evaluation de l'assemblage final

Le transcriptome de *Lgh* possède un N50 de 2304 pb et une longueur moyenne des transcrits de 1641 pb (Table 3.2). Beaucoup de transcrits font entre 1000 et 3000 pb de long (Figure 3.5). L'analyse par BUSCO a permis d'évaluer la complétude du transcriptome à 98% (Table 3.2). Une grande majorité des BUSCO retrouvés sont dupliqués (91,1%) tandis que 6,9% sont présents en une seule copie. Le nombre de BUSCO dupliqués n'est pas surprenant compte tenu de la polyploïdie de *Lgh*, un taux élevé de BUSCO dupliqué avait également été retrouvé dans le génome de *Lgh* (80,8%) [Chapitre 2, article génome soumis].

3.2 Mapping des transcrits assemblés sur le génome

La comparaison entre un transcriptome de *Lgh*, assemblé *de novo* et le génome de référence constitue une étape essentielle pour évaluer à la fois la qualité de notre assemblage transcriptomique et l'exhaustivité de la référence génomique.

Transcrits non mappés

Sur les 182091 transcrits étudiés, 7604, soit 4,2% ne se mappent pas sur le génome de référence malgré l'application de paramètres adaptés pour le mappeur sélectionnée précédemment (Minimap2, long reads, -ax spliced) et après décontamination préalable. Différentes causes peuvent être responsables de ces transcrits ne mappant pas sur le génome.

Le génome de référence est incomplet : ces transcrits détectés par RNA-seq mais non alignés sur le génome de référence peuvent refléter des lacunes ou des limitations de la version actuelle. En effet, de nombreux exemples montrent, notamment chez les plantes polyploïdes, que certains transcrits apparemment "orphelins" ne sont en réalité que des séquences exprimées issues de régions non assemblées initialement. Par exemple, même

si la première version du génome de la vigne (*Vitis vinifera*) était un draft de haute qualité [Jaillon et al., 2007], les premières analyses transcriptomiques ont montré des lacunes avec notamment 3368 transcrits liés au développement des bourgeons, absents du premier assemblage [Ramos et al., 2017]. Le problème évoqué est notamment lié à des phénomènes biologiques tels que la polyploïdie, la reproduction clonale, la présence de régions répétées ou une diversité génétique non représentée. Notre génome de référence est actuellement constitué d'une seule version de chaque locus, les variations génétiques présentes dans d'autres individus ou populations, telles que les polymorphismes nucléotidiques (SNPs), les insertions, les délétions ou les duplications, issu d'un individu ou d'une lignée spécifique, peuvent empêcher l'alignement des transcrits correspondants.

Ces variations sont particulièrement fréquentes dans les régions non codantes ou répétitives, qui sont souvent sous-représentées ou absentes des génomes de référence classiques [Sun, Shang et al., 2022]. La polyploïdie, ajoute une complexité supplémentaire avec plusieurs copies de chaque gène qui peuvent être homoéologues (provenant de génomes parents différents) ou paralogues (résultant de duplications génétiques) [Glover, Redestig et Dessimoz, 2016]. Ces copies peuvent diverger au fil du temps, accumulant des mutations spécifiques à chaque nouvelle copie. Actuellement, notre génome de référence, comme la grande majorité des génomes de plantes, ne capturent qu'une seule version de ces gènes, négligeant ainsi les autres copies et leurs variantes associées. Cela peut entraîner l'absence d'alignement des transcrits provenant de copies non représentées dans la référence. Enfin, les régions répétitives du génome, telles que les transposons, les satellites et les séquences intergéniques, sont souvent difficiles à assembler et à annoter dans les génomes de référence [Loreto et al., 2024]. Ces régions peuvent contenir des transcrits fonctionnels ou régulateurs, mais leur absence ou leur mauvaise représentation dans la référence peut empêcher l'alignement des transcrits correspondants. Ces observations soulignent la nécessité d'améliorer notre génome de

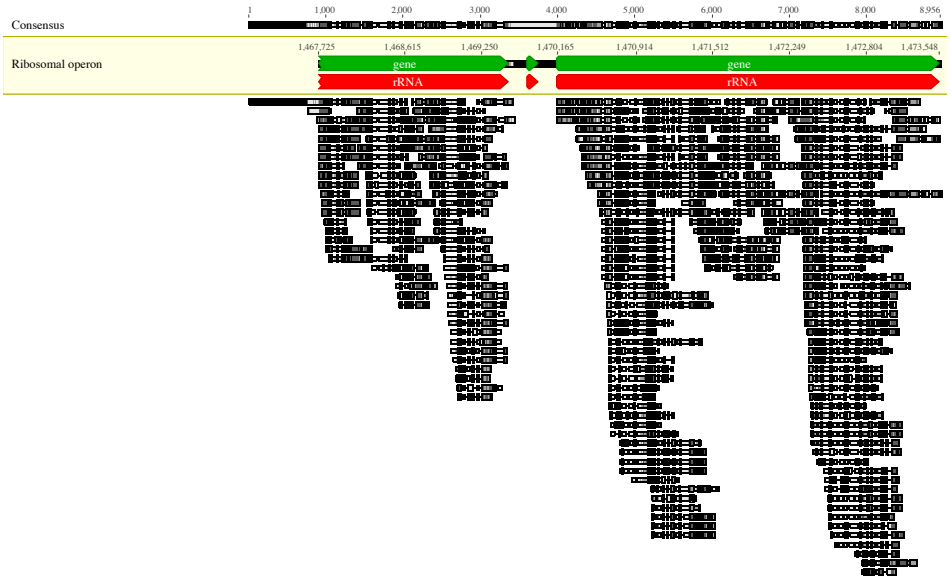


FIGURE 3.6 – Observations d'environ 300 variants des gènes ribosomaux identifiés dans le transcriptome *de novo* de *Lgh*, en relâchant les paramètres de spécificité/sensibilité et non mappés avec les paramètres "classiques".

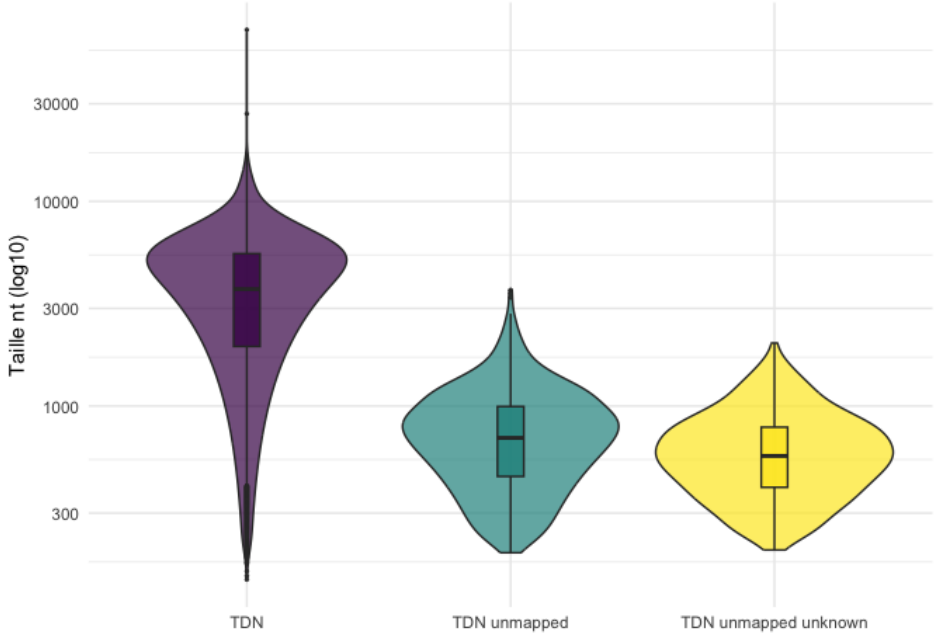


FIGURE 3.7 – Répartition des tailles des transcrits de *Lgh* assemblés *de novo* (TDN), non mappés (TDN unmapped) et non mappés sans homologue (TDN unmapped unknown), en log10 (nt).

référence en les rendant plus représentatifs de la diversité génétique réelle, notamment en intégrant des données provenant de multiples individus, populations ou espèces (cf. perspectives).

Une autre explication pour les transcrits non mappés serait la présence d'isoformes, de chimères ou d'assemblage erronés. Les assembleurs *de novo* de transcriptome, tels que Trinity, ABySS ou SPAdes, rencontrent des difficultés notables lorsqu'il s'agit de traiter les régions répétitives du génome comme les gènes ribosomiques [Lima et al., 2017]. Ces gènes, qui codent les ARN ribosomiques (ARNr), sont organisés en clusters de répétitions en tandem [Wicke, Costa et al., 2011], ce qui complique leur assemblage fiable, surtout en utilisant un séquençage short read. Ces assembleurs *de novo* reposent sur des graphes de De Bruijn, qui segmentent les lectures en k-mers pour assembler les transcrits, méthodes mal adaptées pour gérer les répétitions présente en de nombreuses copies, car elles ne disposent pas de modèles explicites pour les répétitions [Lima et al., 2017]. Cela peut entraîner la formation de sous-graphes combinatoires complexes, rendant difficile la résolution correcte des transcrits associés aux régions répétitives. Nous constatons par exemple qu'environ 10% des séquences non mappées sont à des variantes de l'ilôt ribosomique 35S correspondant aux gènes 18S, 5.8S et 28S (Figure 3.6).

La plupart des transcrits non mappés sur le génome sont plutôt de petites tailles, en moyenne 781 nt et 67,8% de séquences inférieurs à 500 nt comparé au transcrits mappés (Figure 3.7). Le mappeur Minimap2 a été retenu pour l'étude suite à des tests effectués au préalable. Cependant, selon le manuel de Minimap2, l'outil est plutôt désigné pour l'alignement de longues séquences d'ARN [Li, 2018]. Certains de ces transcrits peuvent correspondre à des exons isolés ou des UTR très courts, difficile à mapper s'ils chevauchent une jonction exon-intron ou une fin de contigs. Enfin, il est difficile de savoir si ces petits transcrits représentent des réalités biologiques ou des artefacts ou

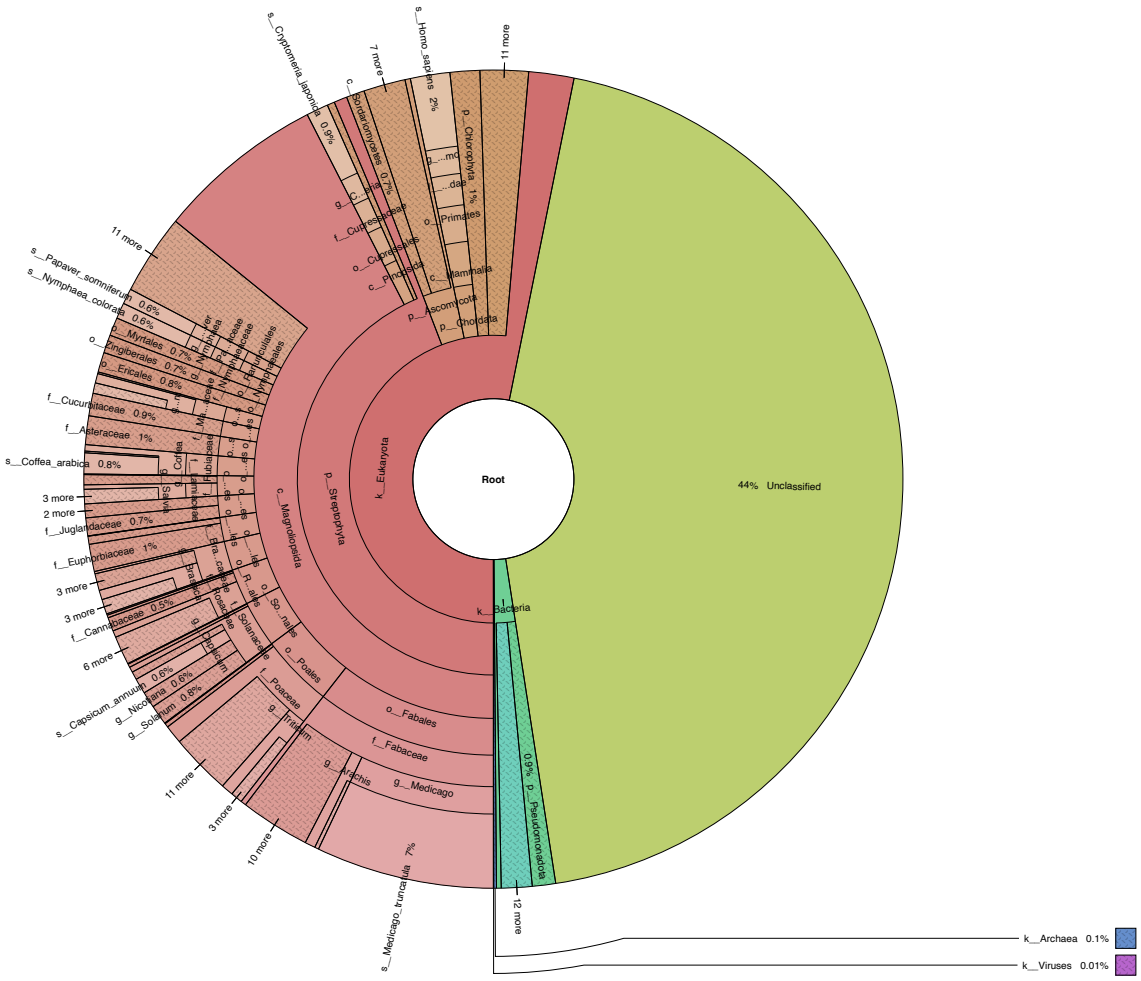


FIGURE 3.8 – Répartition taxonomique des homologues des transcrits de *Lgh* analysés avec Kraken2 et visualisés *via* Krona. Les séquences sont majoritairement assignées mais une fraction significative (44% ; partie verte) reste non classée.

fragments non informatifs.

Il est aussi possible que notre étape de décontamination contienne des erreurs. Pour en savoir plus, nous avons cherchez des homologues des transcrits non mappés par megablast et par kraken2 et avons trouvé environ 44% des séquences sans homologues (Figure 3.8). Le fait d'avoir autant de séquences sans homologue peut avoir plusieurs explications. Soit ces transcrits sont liés à un assemblage incomplet ou chimérique, soit, le fait de travailler sur une espèce non modèle font que ces séquences sont novatrices, spécifiques à l'espèce *Ludwigia* et n'ont pas d'homologues dans les bases de données, contenant principalement des espèces modèles. L'analyse taxonomique des séquences avec homologues révèle la persistance de contaminations au sein de l'assemblage *de novo* avec des homologues d'origine microbien (bactéries, archées, virus) et d'autres contaminants comme des séquences de levures et d'homme également observés, soulignant la présence de séquences exogènes dans l'assemblage qu'il convient d'éliminer. La présence de ces séquences peut provenir de notre méthode de décontamination itératives qui ne permet pas de comparaison entre les outils utilisés contrairement à ce qui est recommandé dans la littérature [Chorlton, 2024 ; Lupo et al., 2021]. Par contre, des homologues végétaux ont été identifiés pour environ 45% des séquences non mappées, ce qui suggère que ces transcrits sont probablement issus de l'espèce étudiée (Figure 3.8). Ces transcrits non mappés constituent donc une ressource précieuse pour améliorer la complétude et l'exactitude de l'assemblage génomique lors de futures versions. Leur intégration permettra non seulement de combler les lacunes existantes, mais aussi de mieux représenter les isoformes, les gènes spécifiques et les régions peu conservées, renforçant ainsi la qualité globale de la référence génomique de cette plante non-modèle.

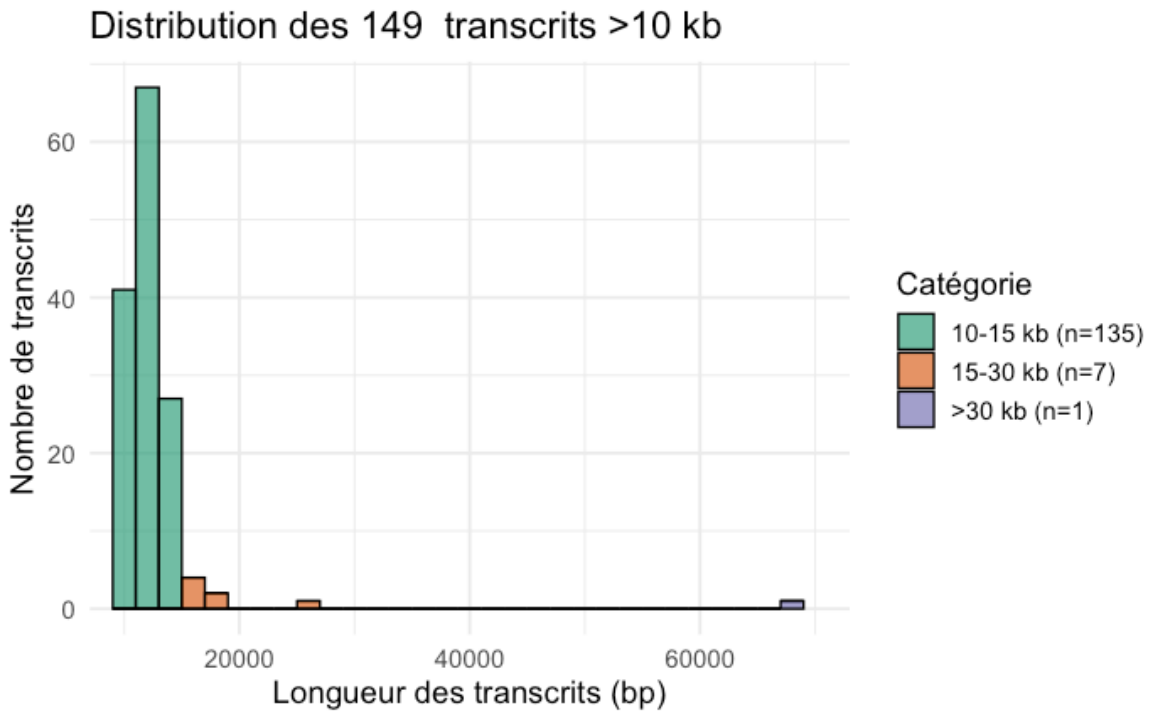


FIGURE 3.9 – Distribution des tailles des 149 transcrits de *Lgh* supérieures à 10 kb.

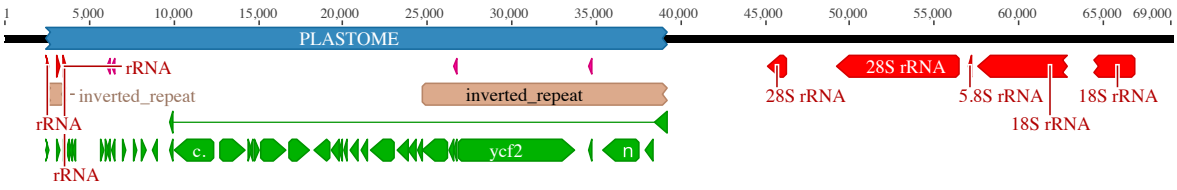


FIGURE 3.10 – Transcrit de *Lgh* chimérique de 69 kb issu de notre assemblage contenant de l'ADN contaminant et des ARN ribosomaux.

Transcrits mappés

Les transcrits *de novo*, dont la longueur dépasse 10 kb, doivent être considérés comme suspects puisque la plupart des ARNm matures de plantes ont des longueurs de l'ordre du kb (généralement entre 1 et 5 kb pour la plupart des gènes codants) [Kellogg et Bennetzen, 2004]. Le transcriptome *de novo* de *Lgh* contient 149 transcrits de taille > 10 kb, correspondant à 64 unigènes, dont la grande majorité ayant une taille comprise entre 10 et 15 kb (Figure 3.9). Un unigène est définit comme un groupe non redondant d'ARN associés à un même locus regroupant ainsi tous les isoformes exprimés d'un gène [Pontius, Wagner et Schuler, 2003]. Ces contigs très longs peuvent résulter d'artefacts (contamination par de l'ADN génomique, chimères d'assemblage, ou collapse de répétitions) mais peuvent également être réellement des ARNm. Voici les trois cas de figure rencontrés pour ces 64 unigènes.

Le premier cas est un transcrit chimérique composé d'ADN contaminant et de séquences répétées (Figure 3.10). L'annotation de ce contig par BlastN montre une structure chimérique en 2 régions : une première région de 36,8 kb d'origine chloroplastique et une seconde région contenant plusieurs fragments d'ARNs ribosomaux nucléaires. La région chloroplastique correspond clairement à une contamination par de l'ADN plutôt qu'à de l'ARN chloroplastique car la couverture est uniforme sur les régions intergéniques attendues comme non transcrites et l'identité est supérieure à 98% avec le plastome. En effet, il est possible que la DNase ne digère pas complétement l'ADN, causant ainsi une contamination [Li, Zhang et al., 2022].

De plus, le nombre de copie des organelles important (mitochondrie, chloroplaste) conduit à une quantité d'ADN d'organelles résiduels pouvant représenter une part non négligeable des reads contaminés par le génome [Morley, Ahmad et Nielsen, 2019]. La seconde région montre les difficultés d'assemblage des régions répétées, telles que les gènes d'ARN ribosomaux (ARNr) comme évoqué précédemment. Dans un graphe

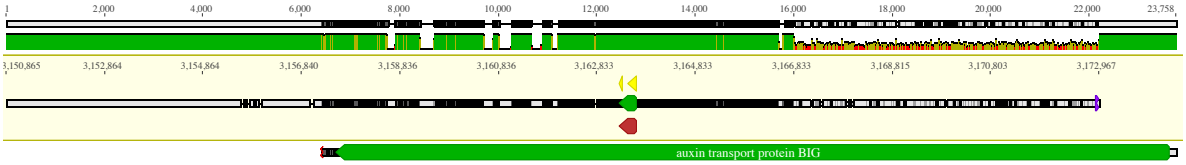


FIGURE 3.11 – Transcrit codant pour la protéine BIG qui mappe en bout de contig du génome de *Lgh*.

de Bruijn, des k-mers communs relient artificiellement des fragments qui ne sont pas biologiquement contigus, ce qui conduit à la production de chimères fragmentaires : des contigs qui associent des segments provenant de différents loci, voire de différents compartiments comme c'est le cas de ce contig, chimère entre un ADN chloroplastique et un long missassemblage de rRNA nucléaires, probablement lié *via* un ARN ou un k-mer commun [Sacomoto et al., 2014].

Cet exemple illustre l'importance d'avoir un génome de référence, même sous forme de draft pour pouvoir identifier ces transcrits artéfactuels. Sur les 64 unigènes de taille $>$ 10 kb, 3 correspondent à de l'ADN chloroplastique et un à de l'ADN mitochondrial M1. Afin d'être exhaustif et d'éliminer tous les fragments d'ADN des organites, nous avons mappé les transcripts sans autoriser d'épissage et de manière stringente ce qui nous à permis d'identifier et d'éliminer 131 transcrits correspondant à de l'ADN chloroplastique, 135 correspondant à de l'ADN mitochondrial M1 et 54 à de l'ADN mitochondrial M2. Si l'on réalise le même mapping sur le génome nucléaire, on retrouve 41581 transcrits correspondant à des régions avec plus de 96 % d'identité. Une étude complémentaire devra être réalisé afin de filtrer les contaminants ADN, susceptibles de biaiser les estimations d'expression, l'annotation des isoformes et la complétude transcriptomique, leur retrait systématique ou un filtrage trop « agressif » peuvent entrainer une perte de signaux biologiques comme de nouveaux isoformes ou de longs ARN non-épissées. Il faut donc documenter et conserver les données filtrées séparément.

Le deuxième cas correspond à un ARN-messager exceptionnellement long, non épissé, présentant une forte similarité avec le gène BIG (alias DOC1 / TIR3), codant une protéine de très grande taille nécessaire au transport polaire de l'auxine (Figure 3.11) [Gil et al., 2001]. Dans ce cas de figure, l'absence d'épissage et la taille inhabituelle de la protéine produite génère ce transcrit de grande taille, qui n'est pas un artefact. Compte tenu de l'importance fonctionnelle d'un gène de transport de l'auxine, il est

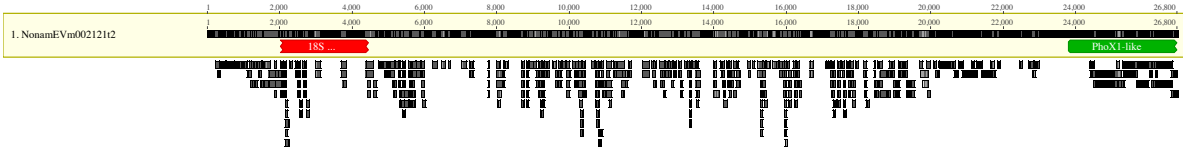


FIGURE 3.12 – Transcrit artéfactuel composé de plusieurs petits fragments du génome de *Lgh*.

important de conserver ce transcrit dans le jeu de données. De plus, l'alignement partiel en fin de contig et l'absence d'annotation montrent la complémentarité des approches « omiques » (*de novo* versus sur référence) et illustre ici comment le transcriptome *de novo* peut corriger notamment les « effet de bord des contigs », limite de l'assemblage par manque de k-mers et identifier des zones non ou mal annotées. L'analyse rapide des grands unigènes montre qu'au moins quinze grands ARNm (> 10 kb) existent réellement dans le transcriptome de *Lgh*. On trouve notamment un ARNm codant une protéine de type BEACH domain-containing protein C2. Le domaine BEACH est un module pouvant aller de 1000 à 4000 acides aminés présent dans une grande famille de protéines eucaryotes impliquées dans le trafic vésiculaire et les processus membranaires [Cullinane, Schäffer et Huizing, 2013].

Au total, 15 transcrits de plus de 10 kb s'avèrent correspondre à des ARNm codant de grandes protéines. Il faut également noter que ces gènes ne contiennent pas d'introns et sont apparus lors de l'évolution des plantes terrestres, possiblement en lien avec des duplications du génome (WGD) [Liu, Lyu et al., 2021]. En conclusion, l'observation de transcrits végétaux de très grande taille, notamment mono-exoniques, ne doit pas être systématiquement interprétée comme un artefact, car ils peuvent coder des protéines structurales ou régulatrices majeures dans l'adaptation physiologique de *Lgh*. Dans un contexte où nos assemblages *de novo* et sur référence restent incomplets, il semblera important de porter une attention particulière à ces transcrits afin de détecter de nouvelles unités fonctionnelles, mais aussi de guider l'amélioration des annotations futures et la complétion des ressources génomiques.

Enfin, certains grands transcrits semblent correspondre à des séquences nucléiques courtes, provenant de plusieurs loci et artificiellement concaténées entre elles lors de l'assemblage *de novo*. Par exemple, un transcrit était composé de plusieurs petits fragments génomiques non contigus (Figure 3.12). Ce type de concaténation artéfactuelle

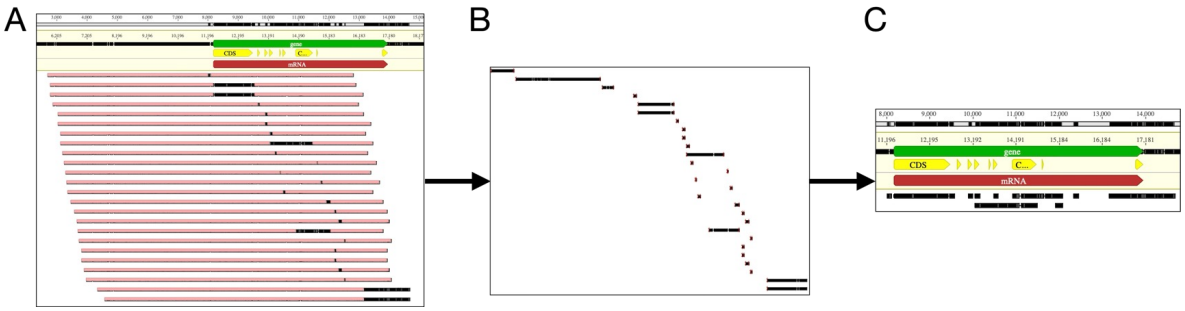


FIGURE 3.13 – Exemple de redondance lors du mapping des transcrits sur le génome de *Lgh*. A : Mapping avec les régions trimmées par Minimap2 en rose. B : Transcrits sans les régions trimmées. C : Mapping des transcrits après réduction de la redondance.

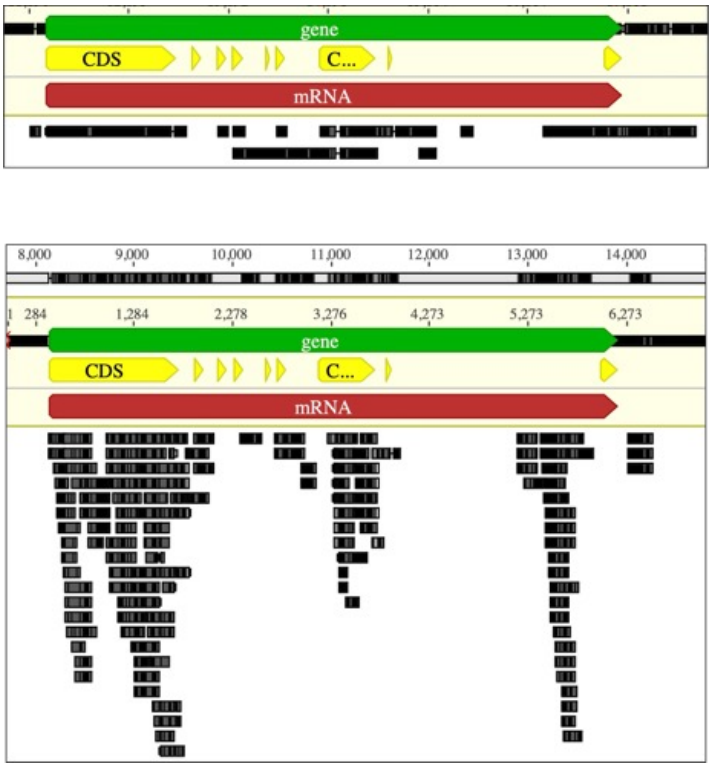


FIGURE 3.14 – Comparaison du mapping d'un transcrit *de novo* après réduction de la redondance (en haut) et du mapping des reads issus du RNA-Seq pairés et mergés (en bas) sur un gène du génome de *Lgh*.

est provoquée par la présence de k-mers partagés entre les séquences, notamment très abondantes comme les rRNA ou les répétitions, provoquant des erreurs d'assemblage. La polyploïdie importante de *Lgh* complique encore l'assemblage *de novo* de son transcriptome avec la multiplication des copies homéologues, divergentes par seulement quelques SNPs ou petits indels, que les assembleurs basés sur des k-mers ont du mal à distinguer. L'assembleur peut mélanger les isoformes d'un gène et les homoéologues d'un paralogue, produisant des chimères ou une redondance massive de contigs [Behera, Voshall et Moriyama, 2021]. Il convient donc d'identifier ces assemblages artefactuels provenant de collapsus (plusieurs copies fusionnées en un seul transcrit) à l'aide du génome de référence et de les éliminer dans un premier temps.

Un autre problème rencontré avec le transcriptome *de novo* est la présence de transcrits redondants et fragmentés. En effet, lors du mapping des transcrits assemblés *de novo* sur la référence génomique, nous avons observé que de nombreux transcrits s'alignaient sur un même locus, souvent avec des régions homologues très courtes conservées après soft-clipping successif par Minimap2. Ce phénomène reflète principalement une combinaison de facteurs évoqués précédemment : fragmentation, et redondance de l'assemblage (présence d'isoformes partielles et de contigs chevauchants), artefacts d'assemblage (chimères), extrémités résiduelles (polyA, adaptateurs) ou complexité locus-spécifique (répétitions, paralogies). Après un nettoyage avec trimming des régions non homologues (en rose Figure 3.13A) et réduction de la redondance (Figure 3.13B), le mapping est beaucoup plus cohérent et informatif (Figure 3.13C), ce qui confirme l'intérêt du génome pour produire une annotation fiable et pour sélectionner les transcrits pertinents. C'est ce nettoyage que nous allons réaliser pour l'ensemble des transcrits mappés. La comparaison des transcrits assemblés *de novo* nettoyés versus RNA-seq mergés mappés pour ce locus demontrent l'intérêt du nettoyage à l'aide du génome (Figure 3.14).

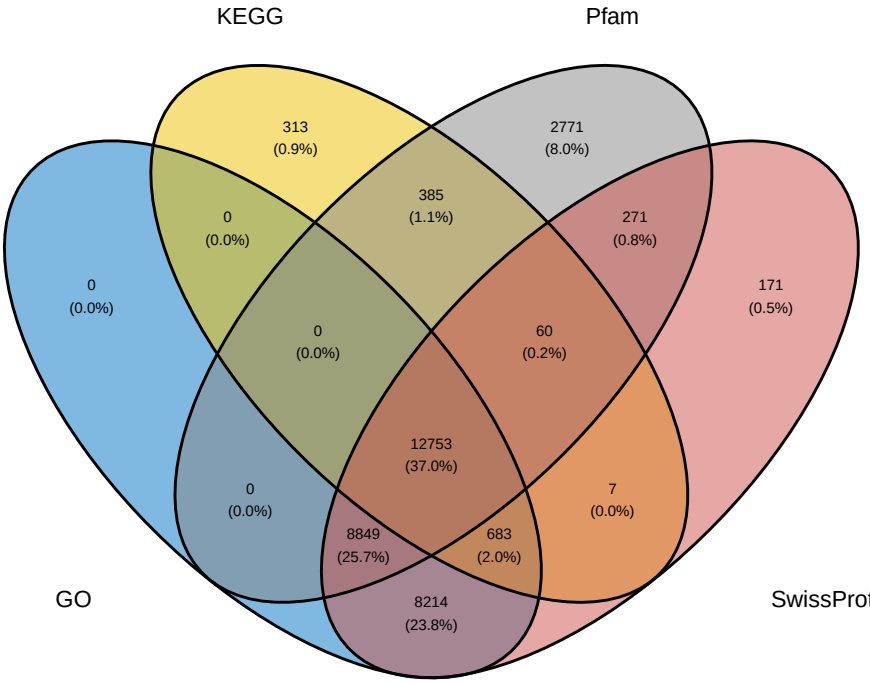


FIGURE 3.15 – Diagramme de Venn du nombre d'unigènes de *Lgh* annotés par Trinotate et GhostKoala en utilisant GO, Pfam, SwissProt et KEGG comme référence.

L'assemblage *de novo* des transcriptomes végétaux comme celui de *Lgh* reste un exercice complexe, marqué par des défis spécifiques : redondances, fragments chimériques, isoformes difficiles à distinguer, ou encore séquences issues de contaminants. Ces limites peuvent être en grande partie évaluées et corrigées grâce au génome de référence, qui offre un cadre structurel robuste pour ancrer et interpréter les transcrits. À l'inverse, le génome de référence dans sa première version est incomplet, fragmenté et imparfaitement annoté, en particulier parce que *Lgh* est une espèce non modèle. Dans ce contexte, les transcrits *de novo* constituent une ressource précieuse pour détecter des régions génomiques manquantes, corriger des erreurs de structure, enrichir l'annotation et mettre en évidence de nouvelles unités de transcription. Dans notre cas, le mapping des transcrits retenus après assemblage et décontamination sur le génome complet de *Lgh* (nucléaire, plastidique et mitochondrial) met clairement en lumière ces apports.

3.3 Annotations fonctionnelles du transcriptome de *Lgh*

Un total de 57 446 unigènes ont été identifiés par Evigene dans le transcriptome *de novo* de *Lgh*. Parmi les unigènes, 359 gènes ARNr et 169 gènes ARNt ont été identifiés et retirés des analyses. Le nombre de gènes identifié ici est faible comparé au nombre de gènes présent dans le génome ($\approx$ 246 000 gènes codant des protéines) [Chapitre 2, article génome soumis]. Ceci peut s'expliquer par la présence de beaucoup de gènes dupliqués dans le génome. En effet, plus de 14 000 gènes sont présents en copie identique et un nombre important doit aussi être présent en copie quasi-identique ne donnant pas des gènes ou transcrits distincts dans le transcriptome et augmentant ainsi le nombre de gènes total dans le génome de *Lgh*. Une autre explication est la possibilité que des gènes ne soient pas détectés à cause d'une expression trop faible ou nulle [García-Ortega et Martínez, 2015]. En effet, plusieurs milliers de gènes peuvent manquer dans les transcriptomes, ce qui peut représenter jusqu'à 50% du nombre total de gènes. Les

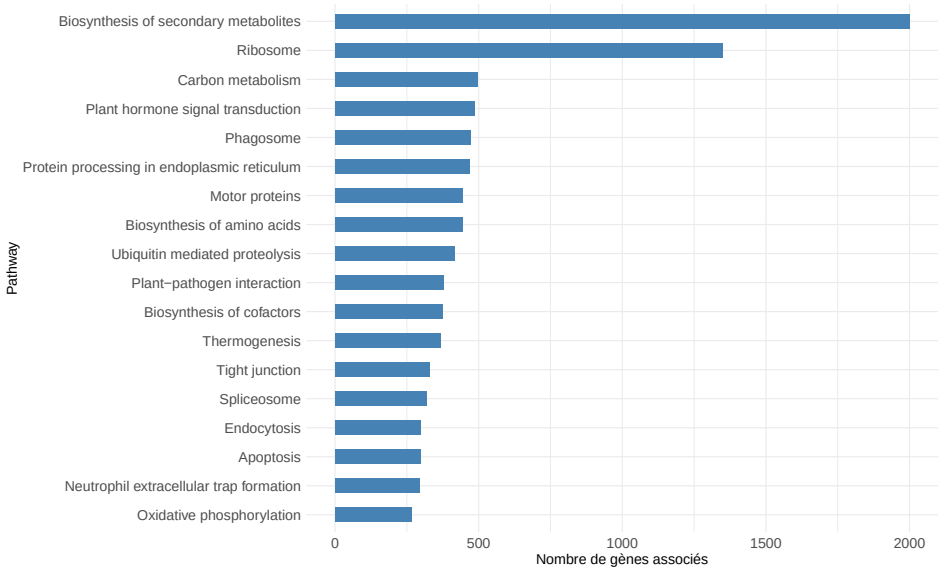


FIGURE 3.16 – Les 20 voies KEGG les plus représentés dans le transcriptome de *Lgh* annotés par GhostKoala.

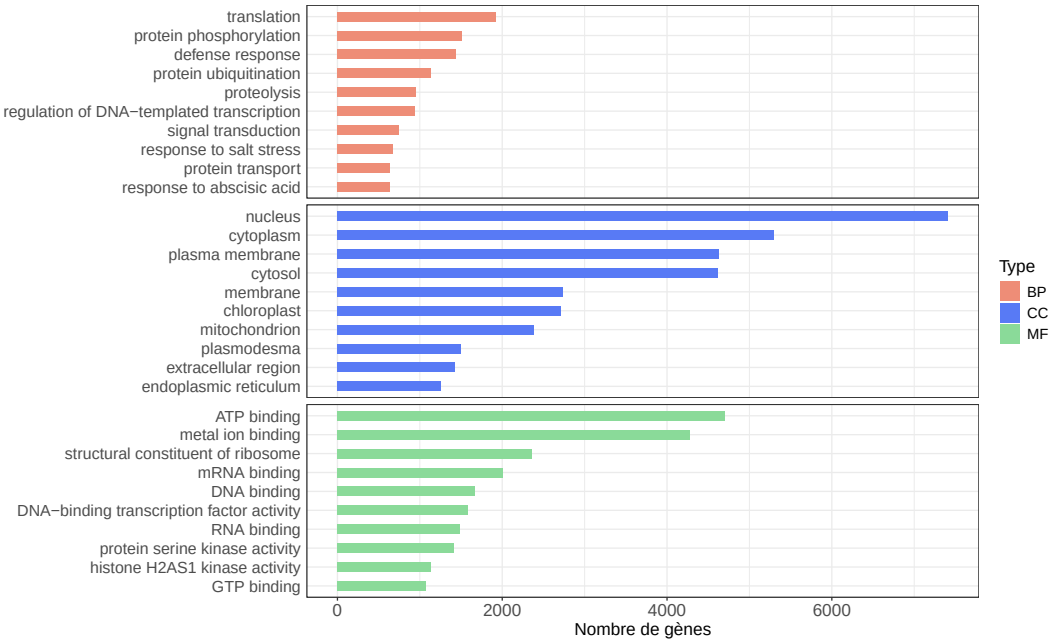


FIGURE 3.17 – Les 10 termes GO les plus représentés dans le transcriptome de *Lgh* par catégorie. BP correspond à Biological Process (Processus biologique) ; CC correspond à Cellular Component (Composant cellulaire) ; MF correspond à Molecular Function (Fonction moléculaire).

gènes non exprimés dans le transcriptome peuvent correspondre à des gènes avec des fonctions spécifiques exprimées seulement dans des tissus ou conditions particulières comme cela a été montré chez le pois chiche et le maïs [Garg et al., 2011 ; Chang et al., 2012]. Des gènes n'ont donc pas pu être retrouvés dans le transcriptome de *Lgh* qui est un transcriptome spécifique aux conditions et aux tissus étudiés.

Trinotate a annoté 30 499 (53,6%), 25 089 (44,1%) et 31 008 (54,5%) gènes avec, respectivement, GO, Pfam et SwissProt (Figure 3.15). GhostKoala a identifié 14 201 (24,9%) gènes avec des annotations KEGG (Figure 3.15). Plus de 60% (34 477) des gènes ont ainsi obtenu une annotation fonctionnelle. Parmi les 40% de gènes sans annotations fonctionnelles, 448 et 1987 ont obtenu des annotations de peptide signal et de domaine transmembranaire avec SignalP et TMHMM. L'annotation KEGG a permis de révéler dans le transcriptome de *Lgh* la présence de nombreux gènes impliqués notamment dans la biosynthèse de métabolites secondaires, dans le ribosome et dans le métabolisme du carbone (Figure 3.16). Les annotations GO mettent en avant les processus biologiques de traduction, de phosphorylation des protéines ou de réponse de défense (Figure 3.17). Le noyau, le cytoplasme et la membrane plasmique sont les compartiments cellulaires les plus présents dans les annotations GO du transcriptome de *Lgh* (Figure 3.17). Pour les fonctions moléculaires, ce sont les liaisons de l'ATP et des ions métalliques et la structure du ribosome qui ressortent (Figure 3.17). Les annotations GO les plus présentes dans le transcriptome semblent être aussi les plus présentes dans le génome de *Lgh* [Chapitre 2, article génome soumis] montrant une bonne adéquation entre les deux ressources.

3.4 Analyse globale de l'expression des gènes dans le génome

Sur les 270 758 gènes présents sur le génome actuel de *Lgh*, nous avons d'abord chercher à identifier le taux d'expression par morphotype/condition. Sur la totalité des expériences que nous avons mené sur les expressions dans les tiges-feuilles et les racines,

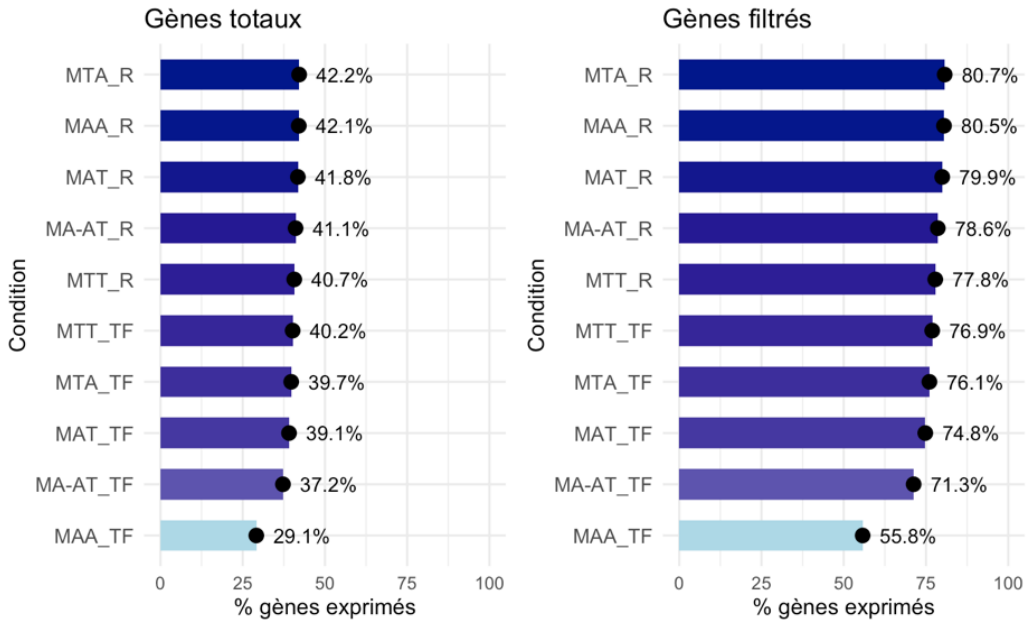


FIGURE 3.18 – Expression des gènes du génome de *Lgh* par échantillon. Les gènes filtrés correspondent aux gènes exprimés dans au moins un échantillon.

un total de 47,7% des gènes (129 271) ne sont jamais exprimés, soit un taux d'expression de 52,3%, très proche de celui de 54% décrit chez *Brassica napus* par exemple. Cette observation est cohérente avec les études précédentes chez les plantes, où une fraction importante du génome est exprimée uniquement dans certains organes, types cellulaires ou stades de développement spécifiques. Ainsi, de nombreux gènes non exprimés dans notre dataset correspondent probablement à une expression spécifique dans un tissu ou à un stade de développement, et leur absence n'indique pas nécessairement une erreur d'annotation, même si la présence de pseudogènes ou de mauvaises annotations n'est pas exclue dans la première version du génome de *Lgh*. Ces conclusions sont similaires à nos observations faites pour le transcriptome *de novo*. Il restera cependant à identifier les gènes exprimés trouvés uniquement dans le génome qui n'ont pas leur transcrit dans le transcriptome *de novo*.

Le filtrage des gènes jamais exprimés améliore la puissance statistique des analyses ultérieures, puisque l'on observe un meilleur taux d'expression, passant de 39,2 à 75,2% en moyenne par échantillon, retenant les signaux biologiquement pertinents (Figure 3.18). Un total de 141 487 gènes du génome sont exprimés au moins une fois dans les échantillons. L'échantillon MAA_TF contient un pourcentage de gènes exprimés bien inférieur aux autres échantillons. Ceci s'explique par le manque de réplicats biologiques pour cet échantillon, diminuant ainsi grandement la représentativité des gènes. Ceci souligne l'importance des triplicats rendant cet échantillon difficilement exploitable par la suite.

3.5 Identification des gènes différentiellement exprimés entre les tiges-feuilles et les racines de *Lgh* en utilisant le génome

Nous présentons ici une sélection de résultats encore en cours d'analyse concernant les gènes différentiellement exprimés (DEGs) entre les tiges-feuilles et les racines en

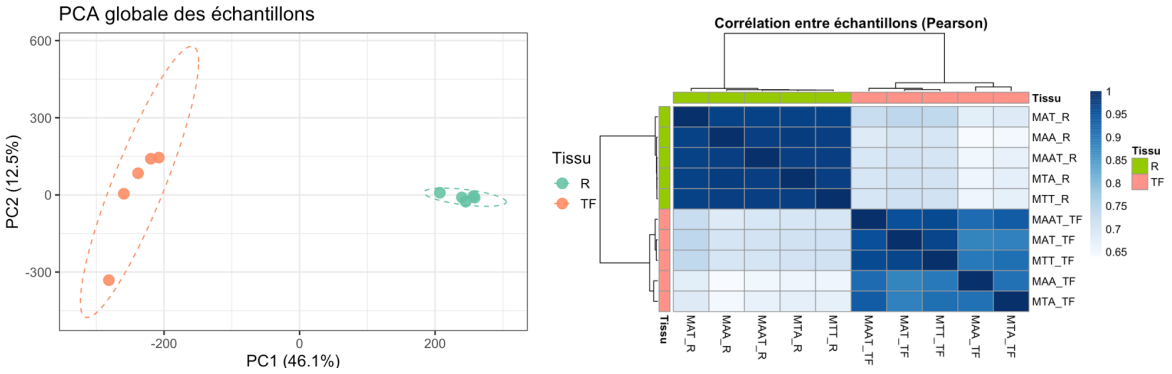


FIGURE 3.19 – Comparaisons entre les échantillons de *Lgh* provenant des tiges-feuilles et des racines. À gauche : ACP ; à droite : Clustering hiéarchique avec la corrélation de Pearson.

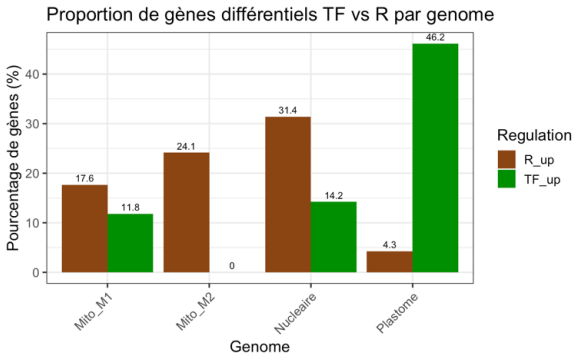


FIGURE 3.20 – Proportion des gènes différentiellement entre les différents génomes de *Lgh* (mitochondrial, chloroplastiques et nucléaire). Le génome mitochondrial de *Lgh* est composé de deux molécules appelées ici Mito_M1 et Mito_M2.

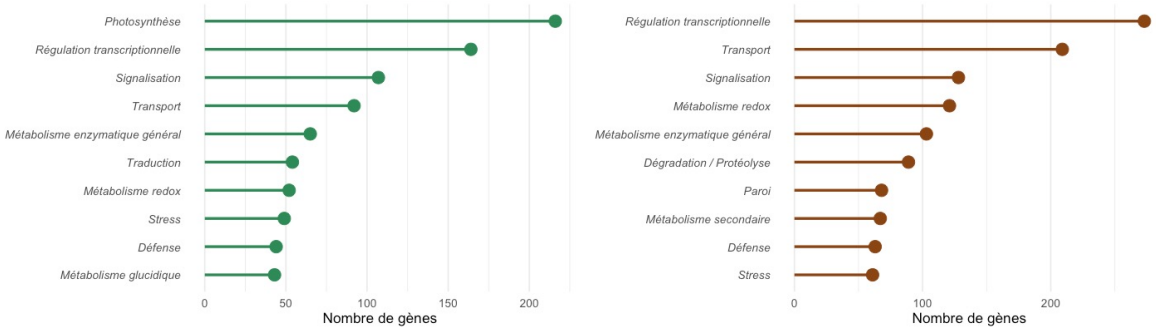


FIGURE 3.21 – Fonctions avec le nombre de gènes surexprimés le plus importants dans les tiges-feuilles (à gauche) et dans les racines (à droite) de *Lgh*.

utilisant le génome comme référence pour les analyses.

Analyses globales

L'ACP permet de mettre en évidence la séparation nette des échantillons en fonction des tissus étudiés (Figure 3.19). En effet, l'axe 1 de l'ACP, participant à 46,1% des variations, sépare les échantillons issus des tiges-feuilles et des racines, révélant ainsi une expression tissu dépendant. Le même constat peut être fait pour la classification hiérarchique séparant bien les échantillons issus des tiges-feuilles et des racines (Figure 3.19). Avant d'identifier les fonctions précises des DEGs entre les tiges-feuilles et les racines, la proportion de ces gènes a été mesuré par génome (Figure 3.20). Près de la moitié (46,2%) des gènes du plastome sont surexprimés dans les tiges feuilles. Ce résultat n'est pas étonnant au vu de l'implication centrale du chloroplaste dans la photosynthèse qui a lieu dans les feuilles de la plante. Par contre, les gènes du plastome surexprimés dans les racines (4,3%) devront être analysés pour comprendre leur rôle. Pour les autres génomes (mitogénome M1 et M2 et génome nucléaire), les racines ont un plus grand pourcentage de DEGs que dans les tiges-feuilles. Ces résultats révèlent que les gènes nucléaires et mitochondriaux sont globalement plus exprimés dans les racines que dans les tiges-feuilles. Ainsi, les variations dues à l'acclimatation de *Lgh* au milieu terrestre pourraient avoir principalement lieu dans les racines.

Pour aller plus loin, les 10 fonctions les plus présentes dans les DEGs dans chaque tissu ont été identifiées (Figure 3.21). On note la présence de la photosynthèse comme étant la fonction la plus représentées dans les DEGs des tiges-feuilles. Côté racines, les fonctions liées à la paroi cellulaire sont très importantes et spécifiques à ce tissu dans les DEGs. On retrouve aussi 7 fonctions communes aux deux tissus que sont la régulation transcriptionnelle, la signalisation, le transport, le métabolisme enzymatique général, le métabolisme redox, le stress et la défense. Dans la suite des analyses nous

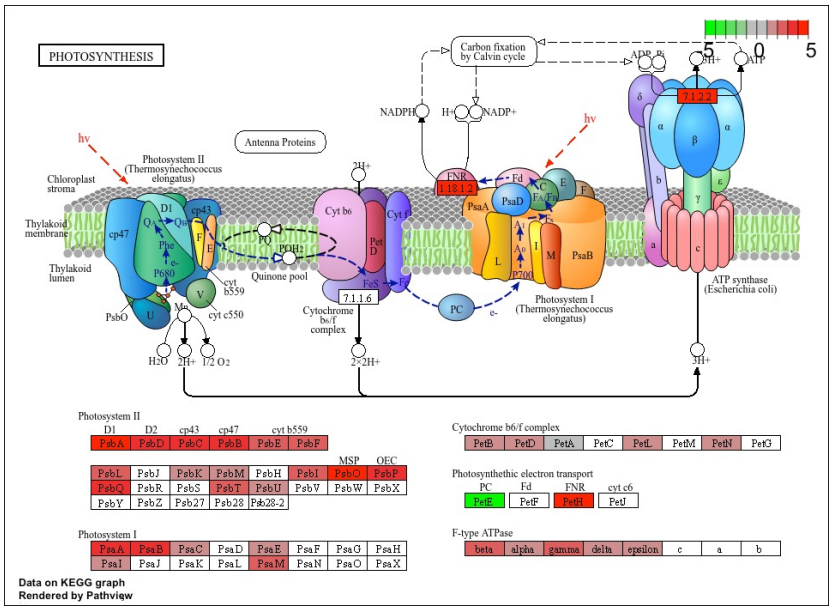


FIGURE 3.22 – Représentation de l'expression différentielle des gènes dans les voies de la photosynthèse entre les tiges-feuilles et les racines réalisée de *Lgh* avec KEGG Mapper. La couleur rouge représente une surexpression dans les tiges-feuilles, le gris une expression équivalente entre tiges-feuilles et racines et le vert une répression dans les tiges-feuilles. Les gènes blancs sont des gènes absents ou non exprimés.

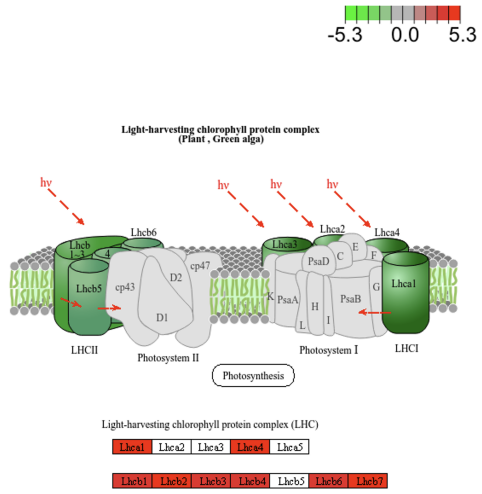


FIGURE 3.23 – Représentation de l'expression différentielle des gènes dans les voies du light-harvesting complex entre les tiges-feuilles et les racines de *Lgh* réalisée avec KEGG Mapper. La couleur rouge représente une surexpression dans les tiges-feuilles, le gris une expression équivalente entre tiges-feuilles et racines et le vert une répression dans les tiges-feuilles. Les gènes blancs sont des gènes absents ou non exprimés.

commencerons par commenter deux fonctions clés spécifiques aux tissus étudiés : la photosynthèse associée aux feuilles et le métabolisme de la paroi cellulaire spécifique aux racines. Nous décrirons ensuite six voies fonctionnelles communes à ces tissus : la régulation transcriptionnelle, la signalisation, le transport et la réponse aux stress, la défense et le métabolisme redox.

Gènes liés à la photosynthèse dans les tiges-feuilles

L'analyse de l'expression différentielle des gènes associés à la photosynthèse chez *Lgh* révèle une forte induction dans les tissus foliaires, ce qui est cohérent avec le rôle central des feuilles comme principal site de capture de l'énergie lumineuse. La majorité des gènes codant les sous-unités structurales des photosystèmes I et II, les composants de la chaîne de transport d'électrons, ainsi que les protéines du complexe de l'ATP synthase, présentent une surexpression marquée, suggérant une activité photosynthétique intense (Figure 3.22).

Dans ce contexte de forte induction globale des gènes photosynthétiques, la répression de *petE*, codant la plastocyanine, est remarquable (en vert sur la figure 3.22) et suggère une modulation fine de l'étape de transfert d'électrons entre le cytochrome b6f et le photosystème I. Cette protéine joue un rôle crucial dans la liaison fonctionnelle des deux photosystèmes, et une réduction de son expression peut indiquer une limitation volontaire du flux linéaire d'électrons, possiblement pour maintenir l'équilibre entre production d'ATP et pouvoir réducteur (NADPH). L'association d'une surexpression globale des autres composants avec une répression ciblée de *petE* suggère donc une régulation post-transcriptionnelle ou transcriptionnelle stratégique pour ajuster la performance photosynthétique aux conditions physiologiques spécifiques des feuilles étudiées [Zhou, Wang et al., 2018].

L'analyse de l'expression génique dans les feuilles a également révélé une surexpres-

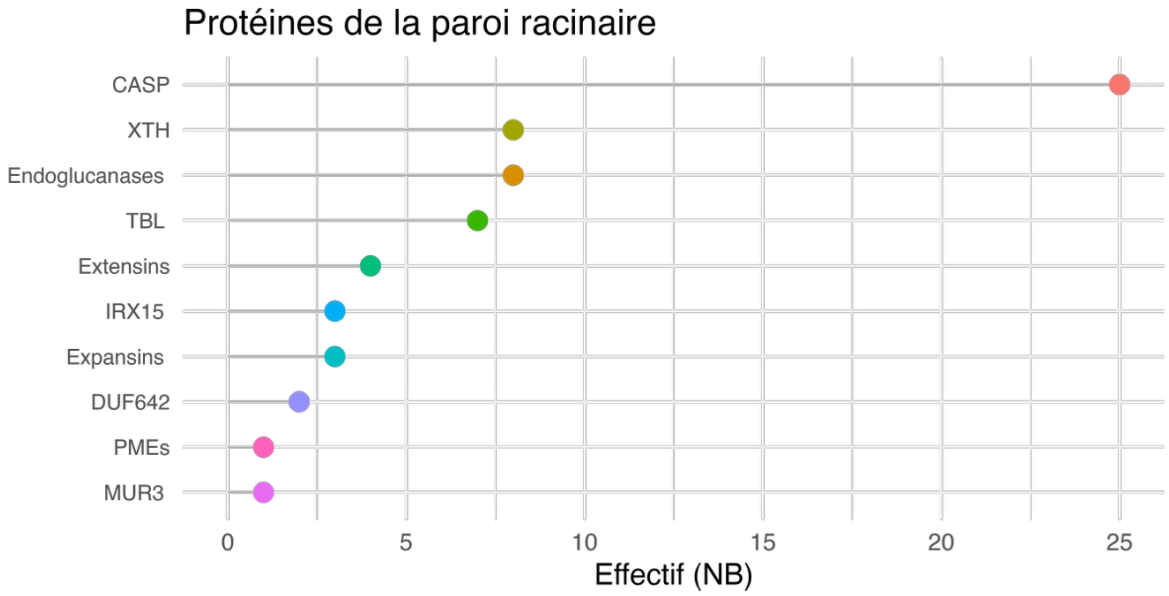


FIGURE 3.24 – Nombre de différentes protéines de la paroi racinaires surexprimés dans les racines de *Lgh*.

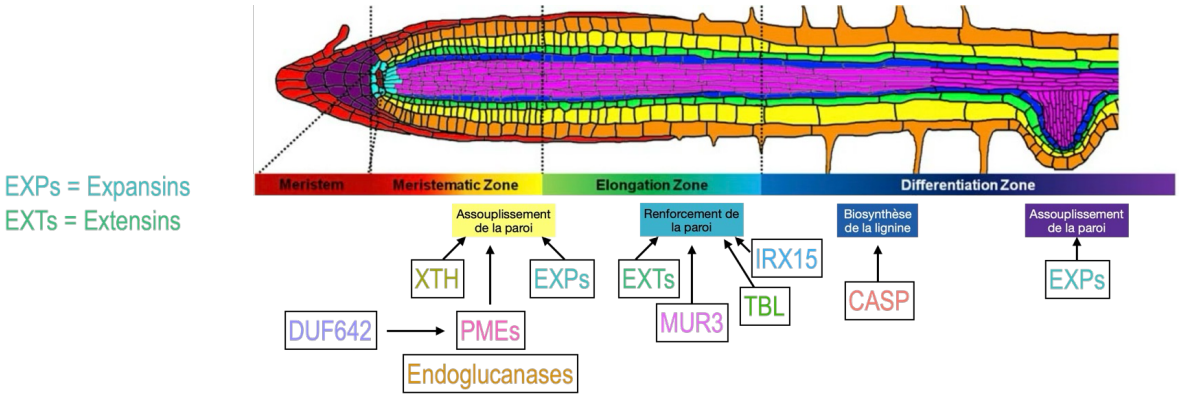


FIGURE 3.25 – Répartition dans la racine et rôle de différentes protéines de la paroi surexprimés dans les racines de *Lgh*.

sion marquée des gènes associés au light-harvesting complex (Figure 3.23). Cependant, les gènes Lhca2, Lhca3, Lhca5 et Lhcb5 n'ont pas été détectés dans le génome nucléaire de *Lgh*. Cette absence peut être attribuée à plusieurs facteurs, tels que des variations spécifiques à l'espèce et/ou le besoin d'une biocuration pour vérifier leur absence ou leur manque d'annotation. Néanmoins, l'absence de ces gènes ne remet pas en cause la capacité photosynthétique de la plante, car d'autres protéines antennes peuvent assumer leurs fonctions, assurant ainsi une capture efficace de l'énergie lumineuse et son transfert vers les centres réactionnels des photosystèmes [Arshad et al., 2022]. Dans l'ensemble, les données indiquent que la plante déploie une activation coordonnée de la machinerie photosynthétique, avec une surexpression des gènes structuraux et enzymatiques essentiels dans les feuilles, résultat attendu qui valide notre transcriptome et explique la spécificité de tissus.

Gènes liés à la paroi cellulaire dans les racines

Le transcriptome racinaire de *Lgh* révèle un profil centré sur le remodelage des hémicelluloses et pectines ainsi que sur le renforcement de la paroi cellulaire (Figure 3.24, 3.25). Cette signature reflète les besoins racinaires typiques : croissance cellulaire (allongement radiculaire, poils absorbants), formation de barrières sélectives (endoderme et cadre de Caspary), et adaptation aux contraintes du sol (osmose, ions, pathogènes). Les modules de structuration et d'acétylation semblent particulièrement mobilisés dans les racines.

Le métabolisme de la paroi racinaire chez *Lgh* combine deux stratégies complémentaires : la formation de barrières spécialisées (CASP, lignification/localisation) pour contrôler la sélectivité d'entrée [Roppolo et al., 2014 ; Brown et al., 2011] et une plasticité des parois primaires (XTH, expansines, PMEs, PGs, extensines) pour l'allongement et la réponse rapide aux contraintes du sol [De Caroli et al., 2021 ; Marowa,

Ding et Kong, 2016 ; Anderson et Pelloux, 2025]. La modulation chimique de la paroi (acétylation par TBL, galactosylation par MUR3, méthylation/déméthylation pectique par PMEs/DUF642) régule simultanément porosité, résistance enzymatique et sensibilité ionique, illustrant un équilibre fin entre perméabilité sélective et capacité d'expansion/renforcement requis à la fonction racinaire [Gao, He et al., 2017 ; Cruz-Valderrama et al., 2019 ; Madson et al., 2003].

La comparaison du comportement de *L. grandiflora* subsp. *grandiflora* (*Lgg*) en sols non inondés (terrestre) et inondés (aquatique) a révélé que la racine est l'organe présentant la plus grande plasticité morphologique en réponse aux variations locales de l'eau [Scremin-Dias et al., 2023]. Ces auteurs ont montré que pour *Lgg*, cultivée sur un sol non inondé, le système vasculaire secondaire des racines présentait un épaississement marqué des cellules du xylème et un diamètre plus petit des éléments vasculaires et des cellules du système axial, avec une lumière nettement réduite. Nos observations sont en accord avec ces résultats.

Facteurs de transcription

Dans l'analyse fonctionnelle des gènes les plus différentiellement exprimés chez *Lgh* (Figure 3.21), les régulateurs transcriptionnels apparaissent en première position pour les racines et en seconde pour les tiges et feuilles. Cette surreprésentation traduit leur rôle central dans la coordination des programmes développementaux et des réponses adaptatives. Les facteurs de transcription assurent le contrôle de l'expression de multiples gènes cibles impliqués dans la différenciation cellulaire, la croissance des organes et les réponses aux signaux hormonaux et environnementaux. L'analyse de la distribution tissulaire des régulateurs met en lumière une double logique avec, d'une part, l'existence de facteurs transversaux qui coordonnent des programmes fondamentaux de croissance et d'adaptation dans les deux compartiments, tiges-feuilles et racines et d'autre part, la

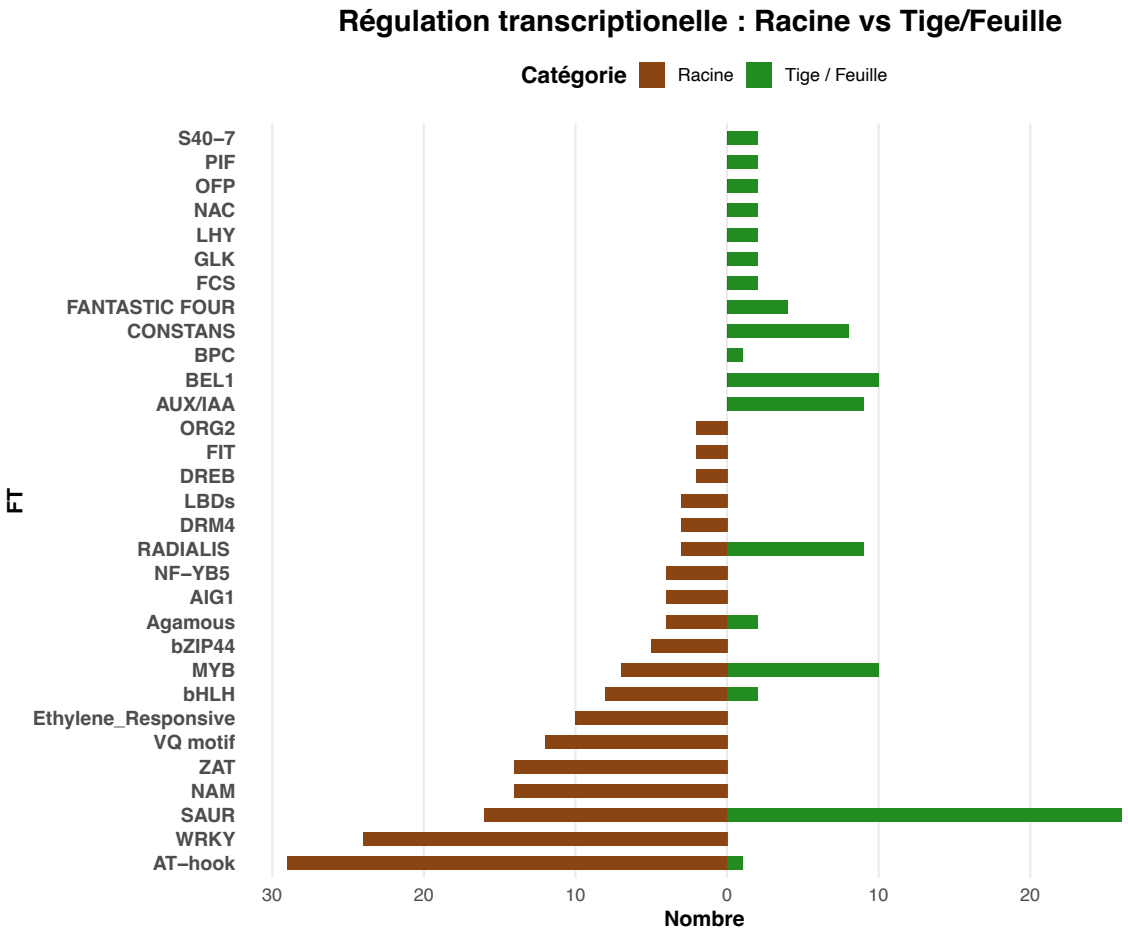


FIGURE 3.26 – Les différents facteurs de transcriptions surexprimés dans les tiges-feuilles et les racines de *Lgh*.

présence de familles spécialisées qui répondent aux contraintes physiologiques propres aux tissus aériens ou souterrains (Figure 3.26). Certaines familles se retrouvent dans les deux compartiments. Par exemple, les SAUR (Small Auxin Up-Regulated RNAs) sont des effecteurs rapides de la voie de l'auxine, promouvant l'expansion cellulaire et la plasticité de croissance. Ils sont très abondants dans les tissus aériens en raison de l'intense activité d'élongation et des gradients d'auxine dans les feuilles et pétioles. Cependant, on les retrouve aussi dans la racine, souvent exprimés de façon plus localisée (zone de transition, coiffe) et donc moins détectés globalement [Stortenbeker et Bemer, 2019 ; Ren et Gray, 2015 ; Spartz et al., 2014].

Les tissus aériens affichent une surreprésentation de familles particulières de facteur de transcription. Le facteur RADIALIS intervient dans l'organisation de la symétrie des organes, expliquant son enrichissement dans les organes aériens [Putterill et al., 1995]. Les régulateurs de la différenciation chloroplastique et de la réponse à la lumière, GLK, PIF, GATA et LHY, reflètent la fonction photosynthétique des feuilles et la nécessité d'une régulation circadienne/photopériodique fine [Waters et al., 2009 ; Leivar et Quail, 2011 ; Reyes, Muro-Pastor et Florencio, 2004].

Les racines montrent une forte représentation de familles impliquées dans la spécification cellulaire, l'acquisition des nutriments et la réponse au stress souterrain. Les AT-hook (AHL), régulateurs de la structure chromatinienne, sont massivement exprimés dans les racines et contrôlent des traits tels que la longueur de la zone de prolifération et l'initiation des racines latérales [Machaj et Grzebelus, 2021]. Les WRKY s'expriment dans les primordia des racines latérales et modulent les réponses au phosphate, à la salinité et aux métaux lourds *via* les voies ABA/auxine et de régulation de l'éthylène [Wang, Wei et al., 2024 ; Devaiah, Karthikeyan et Raghothama, 2007].

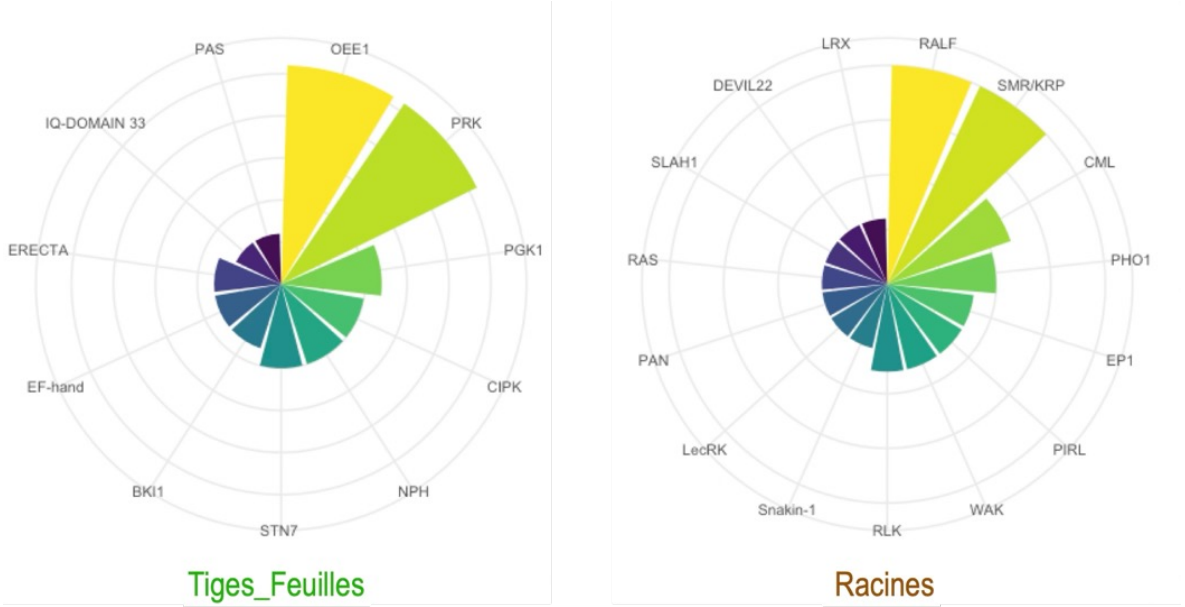


FIGURE 3.27 – Gènes impliqués dans la signalisation surexprimés dans les tiges-feuilles (à gauche) et dans les racines (à droite) de *Lgh*.

Signalisation

L'expression des gènes impliqués dans la signalisation a ensuite été comparé entre les tiges-feuilles et les racines (Figure 3.27). Ici, le terme signalisation est utilisé au sens très large et englobe la perception d'un signal (récepteur, capteur), sa transduction (kinases, messagers) et la réponse de modulation métabolique initial. La phosphoribulokinase (PRK), enzyme clé du cycle de Calvin–Benson (cf. paragraphe Photosynthèse), occupe une position stratégique à l'interface entre métabolisme et signalisation, ce qui peut expliquer sa forte expression dans les tiges et les feuilles. Localisée dans le chloroplaste, PRK catalyse la phosphorylation du ribulose-5-phosphate en ribulose-1,5-bisphosphate (RuBP), substrat essentiel pour la fixation du CO_2 par la Rubisco, et commande ainsi le flux de carbone au sein du cycle photosynthétique. L'Oxygen-evolving enhancer 1 (OEE1) est un constituant essentiel du complexe d'évolution de l'oxygène (OEC) du photosystème II (PSII). En stabilisant l'architecture du site catalytique et en contribuant à l'efficacité de la photolyse de l'eau, OEE1 influence indirectement les signaux oxydatifs et lumineux, modulant la photoprotection et l'équilibre redox du chloroplaste [Deng et al., 2025].

Les racines déploient des réseaux de signalisation adaptés à leur fonction souterraine : nutrition, ancrage et perception du milieu. Les peptides RALF (rapid alkalinization factor) régulent l'expansion cellulaire et la morphogenèse des poils racinaires en modulant l'apoplaste et en activant des récepteurs membranaires, influençant la croissance locale et les réponses de défenses [Schoenaers et al., 2024]. Les inhibiteurs des kinases dépendantes des cyclines (familles SMR, KRP) contrôlent le cycle cellulaire aux méristèmes, assurant l'équilibre prolifération/différenciation racinaire [Kumar et Mohapatra, 2021].

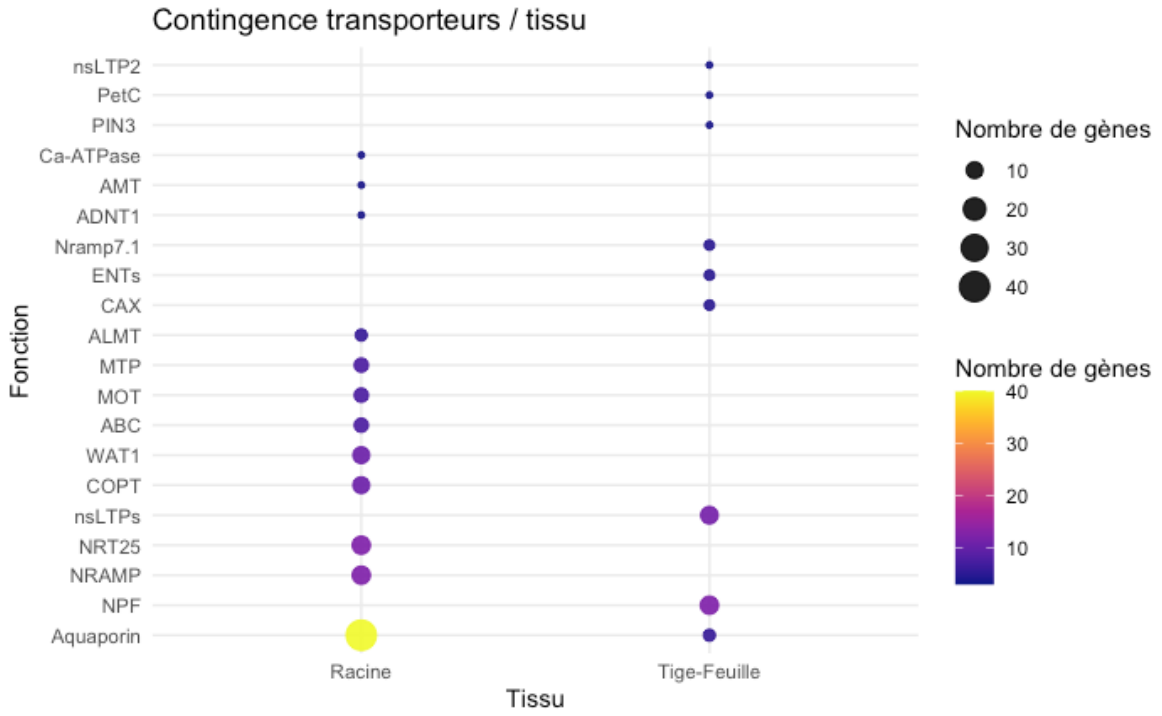


FIGURE 3.28 – Gènes impliqués dans le transport surexprimés dans les tiges-feuilles et dans les racines de *Lgh*.

Transporteurs

D'après nos analyses, les gènes associés au transport font partie des gènes les plus différentiellement exprimés entre les tiges-feuilles et les racines chez *Lgh* (Figure 3.28). Les transporteurs NRT1/PTR (NPF) sont les plus abondants dans les tiges-feuilles et sont impliqués dans le transport de nitrates, de peptides et d'hormones comme l'auxine, modulant ainsi la nutrition azotée, la signalisation et la différenciation tissulaire [Nedelyaeva et al., 2024]. Les nsLTPs facilitent le transfert de lipides entre membranes, contribuant à la formation de la cuticule et à la défense contre les pathogènes ; nsLTP2 joue un rôle central dans la protection contre les stress et le maintien de l'intégrité membranaire [Molina, Segura et García-Olmedo, 1993 ; Edqvist et al., 2018]. Les aquaporines, bien que moins abondantes dans les feuilles que dans les racines, régulent l'hydratation cellulaire et soutiennent l'efficacité photosynthétique et la tolérance au stress hydrique [Byrt et al., 2023].

Dans les racines, les aquaporines, notamment les PIP, dominent ce réseau et contrôlent la conductance hydraulique ainsi que l'homéostasie osmotique, jouant un rôle majeur dans l'entrée d'eau et l'adaptation hydrique [Chaumont et Tyerman, 2014]. Les transporteurs NRT2.5 assurent l'acquisition de l'azote et sont spécifiquement induit en conditions de carence pour optimiser la captation des nitrates [Lezhneva et al., 2014]. Pour les micronutriments, la famille NRAMP contrôle l'import intracellulaire des cations métalliques (Fe, Mn, Zn, Cd), contribuant à la nutrition minérale et à la tolérance aux métaux toxiques [Bozzi et Gaudet, 2021].

Réponse au stress, défense et métabolisme redox

La réponse au stress fait partie des mécanismes clés identifiés dans les tiges-feuilles et dans les racines chez *Lgh*(Figure 3.29). Dans les tiges et les feuilles, les Monothiol glutaredoxins assurent la régulation redox intracellulaire et le transfert de groupes fer-

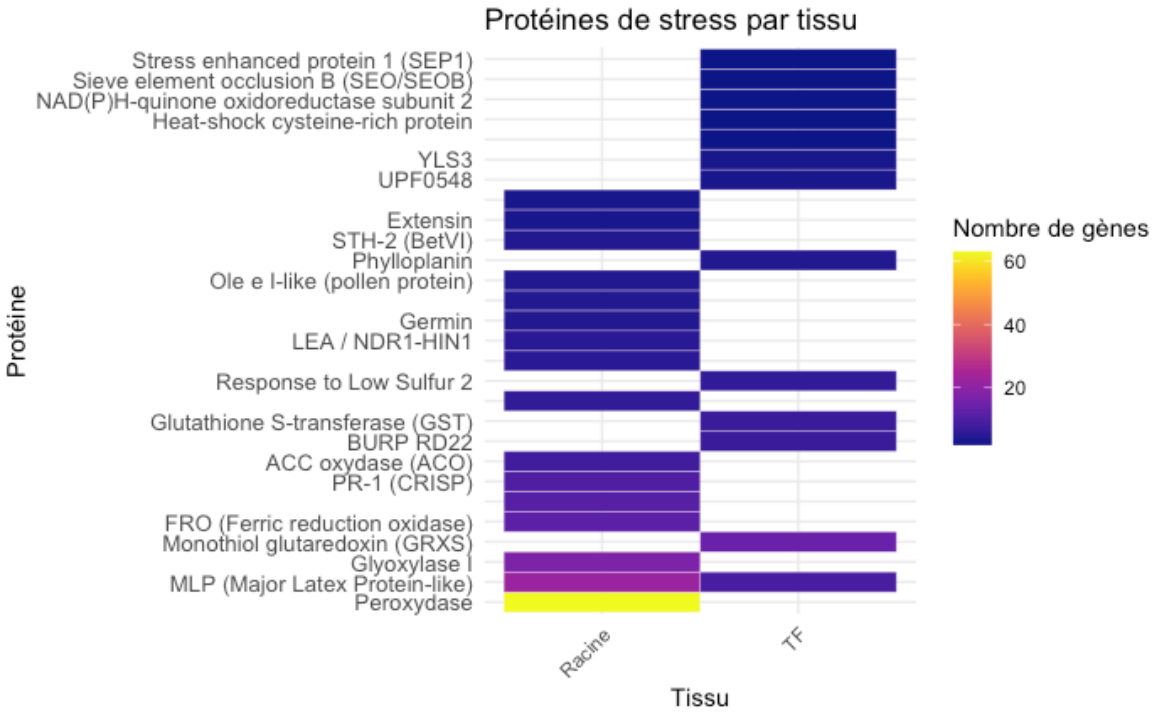


FIGURE 3.29 – Gènes impliqués dans le stress, la défense ou le métabolisme redox surexprimés dans les tiges-feuilles et dans les racines de *Lgh*.

soufre, essentiels au fonctionnement enzymatique et à la protection contre le stress oxydatif [Herrero et Torre-Ruiz, 2007]. Les MLP, petites protéines de signalisation, contribuent à la réponse aux stress biotiques et abiotiques ainsi qu'à la régulation de la croissance [Kang et al., 2023]. Les peroxydases, majoritaires dans les racines, interviennent à la fois dans la génération et la détoxification de ROS (Reactiv Oxygen Species), la lignification et le remodelage de la paroi [Almagro et al., 2009]. Ces fonctions contribuent simultanément à la signalisation d'alerte et à la barrière physique contre les envahisseurs. La multiplicité d'isoformes observée chez les peroxydases s'explique par des duplications de gènes, des ciblages subcellulaires et des régulations tissulaires spécifiques, permettant des réponses rapides et localisées. Des systèmes de détoxification métabolique comme la voie glyoxalase sont fortement représentés dans les racines car ils éliminent le méthylglyoxal, un composé très réactif qui s'accumule sous stress abiotique, et participent ainsi à la tolérance générale aux conditions défavorables [Garai et al., 2021].

Toutes ces analyses démontrent bien les différences d'expression de gènes chez *Lgh* entres les tiges-feuilles et les racines dans plusieurs familles clés de la plante comme les facteurs de transcription, les transporteurs ou les gènes impliqués dans les réponses au stress. De plus, ces fonctions démontrent la fiabilité de notre génome sur ces annotations en plus d'y apporter des éléments de validations expérimentales. Au vu de leur grande différence, les jeux de données tiges-feuilles et racines ont été séparés pour les autres analyses sur le génome.

TABLE 3.3 – Bilan des gènes différentiellement exprimés par *Lgh* lors de la comparaison des morphotypes x condition pris deux à deux. Sont indiqués en jaune les comparaisons qui seront ensuite présentées ; en rouge les comparaisons avec MAA dans les tiges-feuilles (sans réplicat biologique) ; en bleu les comparaisons avec MTA dans les tiges-feuilles qui sont les plus variables. Les gènes surexprimés et réprimés le sont dans le premier échantillon de la comparaison.

Organe	**Comparaison 2 à 2**	**Nb de gènes surexprimés**	**Nb de gènes réprimés**	**Nb total de gènes différentiellement exprimés**
Racine	MAA vs MAT	815	541	1356
	MAA vs MA-AT	744	634	1378
	MAA vs MTA	593	638	1231
	MAA vs MTT	893	593	1486
	MA-AT vs MAT	797	705	1502
	MAT vs MTA	554	838	1392
	MAT vs MTT	791	722	1513
	MA-AT vs MTA	639	830	1469
	MA-AT vs MTT	959	778	1737
	MTA vs MTT	940	610	1550
Tiges-feuilles	MAA vs MAT	4954	5967	10921
	MAA vs MA-AT	2350	6014	8364
	MAA vs MTA	2017	6497	8514
	MAA vs MTT	2438	6211	8649
	MA-AT vs MAT	1283	1559	2842
	MAT vs MTA	871	7417	8288
	MAT vs MTT	696	997	1693
	MA-AT vs MTA	1048	2055	3103
	MA-AT vs MTT	1017	1468	2485
	MTA vs MTT	4875	925	5800

3.6 Comparaison deux à deux des échantillons *via* le transcriptome de référence et analyse de l'effet de l'alternance sur *Lgh* lors d'une acclimatation au milieu terrestre

Le nombre de gènes différentiellement exprimés dans chaque comparaison a d'abord été mesuré (Table 3.3). On trouve un nombre total de gènes différentiellement exprimés variant entre 1231 et 1550 dans les racines. Pour les tiges-feuilles, une quantité importante de gènes différentiellement exprimés est observée pour les comparaisons avec MAA (de 8364 à 10921). Il faut garder en mémoire qu'un seul échantillon de MAA a été séquencé, ce qui nécessitera de regarder ces données avec prudence. Pour les autres comparaisons, le nombre de gènes différentiellement exprimés dans les tiges-feuilles varie entre 1693 et 8288 avec des proportions les plus élevées quand l'un des échantillons comparés est le morphotype terrestre conditionné en aquatique (MTA).

Afin d'évaluer si l'alternance de condition terrestre-aquatique a eu un effet sur le morphotype aquatique, nous avons décidé de comparer deux à deux les trois échantillons suivants : MA-AT, le morphotype aquatique conditionné en alternance, MTT, le morphotype terrestre conditionné en terrestre et MAT, le morphotype aquatique conditionné en terrestre. Les comparaisons seront effectuées uniquement dans les racines.

Comparaison MA-AT vs MTT

En milieu terrestre, le morphotype aquatique ayant subi l'alternance (MA-AT) montre une réorganisation transcriptionnelle majeure de la racine par rapport au morphotype terrestre conditionné en terrestre (MTT) (Figure 3.30).

Les programmes liés à la croissance rapide et à la plasticité cellulaire sont globalement réprimés, ce qui n'est, à contrario, pas le cas pour les racines du morphotype terrestre (MTT). On observe une baisse d'expression de gènes centraux de la division

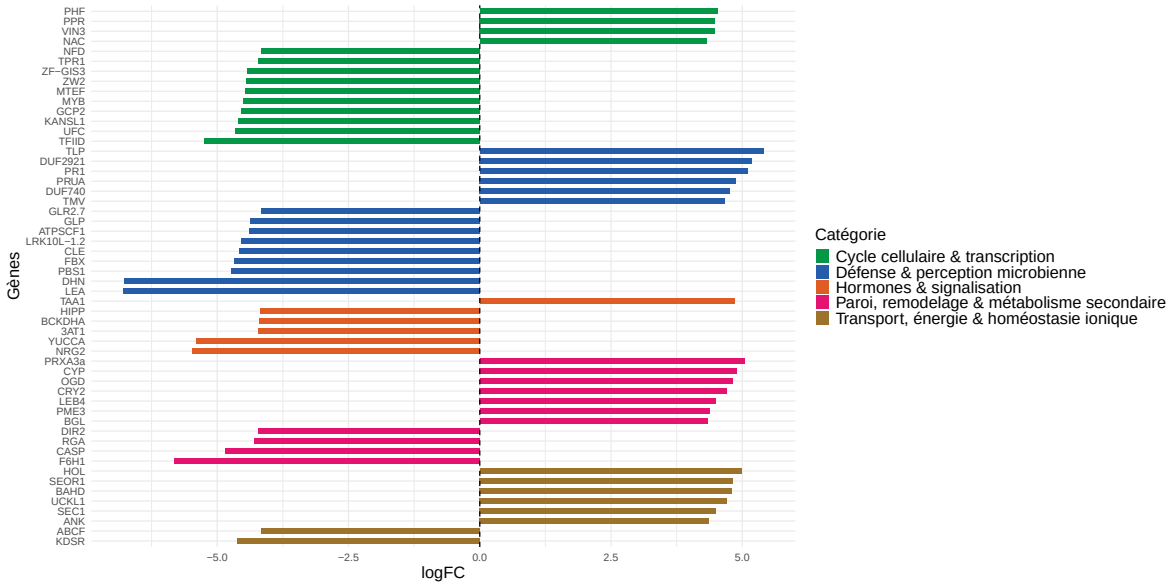


FIGURE 3.30 – Gènes différentiellement exprimés entre MA-AT et MTT chez *Lgh*. Les logFC positif correspondent à une surexpression dans MA-AT tandis que les négatif à une surexpression dans MTT. Les correspondances entre les noms de gènes et leur fonction sont en Table S1.

cellulaire et de la machinerie transcriptionnelle, (NUCLEAR FUSION DEFECTIVE, TPR1, TFIID, MTEF18, MYB, GIS3, GCP2p, KAT8, NSL et UPSTREAM OF FLC). Cette régulation traduit un ralentissement des divisions cellulaires. En parallèle, des gènes favorisant la différenciation et le renforcement tissulaire sont surexprimés. Les familles NAC (stimulation de l'épaississement des parois et différenciation des cellules conductrices), PPR (amélioration de l'efficacité mitochondriale), VIN3-like (état d'expression durables) et PHD finger (lecture et modulation des marques d'histones) sont activées, soutenant la lignification, la subérisation et l'optimisation énergétique pour la construction de structures racinaires résistantes.

La réorganisation entre croissance, énergie et défense apoplastique est également marquée. La baisse de GLR2.7 suggère une réduction des signaux calciques rapides [Toyota et al., 2018]. La diminution de germines et d'ATP synthase CF1 modifie le métabolisme énergétique et la capacité à répondre aux ROS [Dunwell et al., 2008]. Par ailleurs, la répression de récepteurs et kinases de défense (PBS1, LR10-type) et de peptides méristématiques (CLE) indique un ajustement des capteurs et du turnover protéique pour réduire certains programmes de croissance et de surveillance constitutive [Swiderski et Innes, 2001]. La baisse de Dehydrin/LEA signale un retard ou une absence d'induction des protections anti-dessiccation [Battaglia et al., 2008].

À l'inverse, des marqueurs de défense apoplastique sont fortement activés comme PR1, thaumatin, allergènes de type Pru et protéines DUF, indiquant une protection renforcée contre la flore microbienne et les agressions du sol [Han, Xiong et al., 2023 ; Feuillet, Schachermayr et Keller, 1997]. Concrètement hez MA-AT, la racine réalloue ses ressources : moins d'énergie et de flux pour la croissance rapide, plus de défense et de barrière apoplastique, favorisant une racine plus épaisse et plus protégée plutôt que ramifiée.

Les ajustements hormonaux et métaboliques confirment cette stratégie. On note

une baisse de l'expression de HIPP, 2-oxoisovalerate dehydrogenase, C3GT (coumaroyl-CoA :anthocyanidin 3-O-glucoside-6"-O-coumaroyltransferase), YUCCA et NRG2, et une augmentation de CYP734A1 et TAA/TAR. La réduction de YUCCA et l'augmentation de TAA/TAR réorientent le flux auxinique, perturbant la formation d'auxine active et limitant l'élongation et la ramification des radicelles [Stepanova et al., 2011 ; Zhao, 2014]. La modulation de NRG2 ajuste la signalisation nitrate [Xu, Wang et al., 2016], tandis que la baisse de la dégradation des acides aminés branchés et d'acyltransférases anthocyaniques limite l'investissement dans certaines voies phénoliques et énergétiques [Binder, 2010]. L'augmentation de CYP734A1, enzyme d'inactivation des brassinostéroïdes, contribue à limiter l'élongation cellulaire.

Le remodelage des parois et du métabolisme secondaire reflète également cette stratégie. La répression des gènes, codant les Dirigent protein 2, récepteurs de résistance (NLR/RLP), CYP82D47, metacaspase 2a1 et certaines 2-ODD (oxoglutarate/ Fe-dépendantes dioxygénases), réduit la mort cellulaire programmée et certains programmes de défense [Gao, Huang et al., 2023 ; Contreras et al., 2023 ; Tsiatsiani et al., 2011 ; Wang, Shi et al., 2021]. En parallèle, l'augmentation des bêta-galactosidase, pectinesterase 3, legumin type B, Cryptochrome-2, CYP94C115, lignin-forming anionic peroxidase indique un remodelage actif de la paroi, une redirection métabolique et une modulation hormonale. Ces changements favorisent l'épaississement et la résistance mécanique de la racine.

Enfin, la régulation du transport membranaire et de la sécrétion est ajustée. La répression de l'ABC transporter F [Hellsberg, Montanari et Ecker, 2015] et de la 3-dehydrosphinganine reductase [Mandon et al., 1992] limite certains flux membranaires et la synthèse de sphingolipides. L'induction de SEC1 renforce le trafic vésiculaire [Karnik et al., 2013], tandis que l'expression des uridine kinase-like, BAHD acyltransferase et holotricin-3 soutient la biosynthèse et les défenses apoplastiques. La présence de SEO

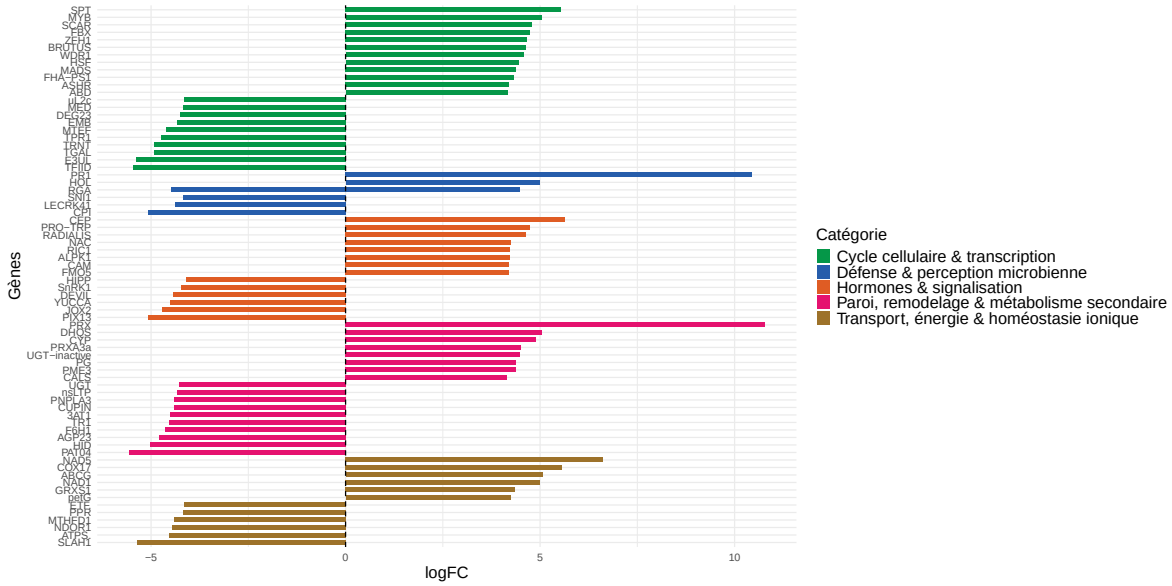


FIGURE 3.31 – Gènes différentiellement exprimés entre MA-AT et MAT chez *Lgh*. Les logFC positif correspondent à une surexpression dans MA-AT tandis que les négatif à une surexpression dans MAT. Les correspondances entre les noms de gènes et leur fonction sont en Table S1.

(sieve element occlusion) assure la protection du phloème [Ernst et al., 2012].

Dans l'ensemble, la racine du morphotype aquatique ayant subi l'alternance transféré en condition terrestre passe d'un mode « croissance rapide et plasticité » à un mode « consolidation et maintien ». Le développement accru de tissus lignifiés dans les racines, suite à un stress hydrique (sécheresse) est souvent rapporté et aboutit à un épaississement des parois cellulaires [De Micco et Aronne, 2012]. La racine limite également l'élongation et l'exploration, réduit la mort cellulaire programmée (PCD), tout en renforçant la paroi, la lignification, la défense apoplastique et l'ajustement énergétique et hormonal. Ces changements reflètent une stratégie adaptative de MA-AT vers un environnement terrestre, contrastant avec les réponses du morphotype terrestre déjà optimisées pour ces conditions. En effet, le morphotype terrestre va, quant à lui ; favoriser une plasticité cellulaire et une croissance racinaire, traduisant son adaptation aux contraintes environnementales du milieu terrestre.

Comparaison MA-AT vs MAT

En milieu terrestre, le morphotype aquatique ayant subi l'alternance (MA-AT) montre une réorganisation transcriptionnelle nette des racines, comparé à ce qui est observé dans les racines produites par le morphotype aquatique ayant été directement conditionné en terrestre (MAT) (Figure 3.31). On observe une réorganisation des priorités cellulaires, avec l'activation coordonnée de régulateurs et d'effecteurs de protection et la répression de gènes associés à la croissance. Cette réorganisation suggère que MA-AT ne « déclenche » pas uniquement des réponses de défense, mais redéfinit l'allocation des ressources pour privilégier des fonctions liées à la structure cellulaire et à la tolérance en contexte terrestre.

Les régulateurs transcriptionnels constituent le point central de cette réorganisation. MA-AT montre une induction marquée de familles classiques (NAC, MADS, MYB,

HSF, TGAL) et de kinases de signalisation qui contrôlent des réseaux cibles larges. Les HSF (Heat Shock Factors) sont induits et, comme attendu, préparent aux stress abiotiques en activant des chaperonnes, des systèmes antioxydants et des protéines stabilisatrices de membrane [Guo, Liu et al., 2016]. Les NAC interviennent dans la réorganisation de la paroi et la maturation vasculaire et s'articulent avec les voies hormonales (notamment l'ABA) pour ajuster la tolérance hydrique [Xiong et al., 2025]. Les MYB orientent le métabolisme phénylpropanoïde et la lignification, ce qui est cohérent avec une stratégie de renforcement des parois [Liu, Osbourn et Ma, 2015]. Les MADS-box traduisent une capacité de reprogrammation développementale pouvant faciliter la reconfiguration morphologique nécessaire à la vie terrestre [Castelán-Muñoz et al., 2019]. Les TGAL / trihelix et les kinases agissent comme nœuds d'intégration des signaux hormonaux et de stress (ABA, jasmonate, éthylène), permettant une modulation rapide des cascades transcriptionnelles [Xu, Shi et al., 2018].

Parallèlement à ces régulateurs, MA-AT induit un ensemble d'effecteurs de surface et apoplastiques et d'enzymes de remodelage de la paroi. On note l'activation des synthases de callose, des protéines PR et de peroxydases impliquées dans la polymérisation de la lignine, ainsi que l'induction de transporteurs et canaux (ABC, SLAH) et de peptides antimicrobiens (holotricine). Ces changements soutiennent la mise en place d'une double barrière — physique (callose, lignine) et chimique (PR, peptides) — tout en maintenant la signalisation ionique et calcique nécessaire pour coordonner la réponse locale [Almagro et al., 2009 ; Wang, Li, Fan et al., 2021 ; Dos Santos et Franco, 2023 ; Ma, 2024 ; Kurusu et al., 2018 ; Weerden, Bleackley et Anderson, 2013 ; Virdi, Singh et Singh, 2015]. La combinaison de ces effecteurs traduit une stratégie d'interface sol–racine : renforcer la zone d'échange et réduire la vulnérabilité face à la flore du sol et à la dessiccation potentielle.

Dans le même temps, des composantes cellulaires soutenant le trafic vésiculaire, la

signalisation et l'homéostasie énergétique sont accrues (actin-binding/SCAR, RAB6A-GEF, alpha-protein kinase, calmoduline, glutaredoxine, nad1/nad5, petG), ce qui favorise l'acheminement ciblé des effecteurs apoplastiques et fournit l'énergie/redox nécessaire aux processus coûteux de lignification et de défense [Pollitt et Insall, 2009 ; Feathers, Vignogna et Fromme, 2024 ; Delay, Imin et Djordjevic, 2013]. Ces modules garantissent que la réorganisation structurelle soit accompagnée d'un trafic cellulaire et d'un métabolisme capable de la soutenir.

Les gènes réprimés chez MA-AT sont surexprimés chez MAT. Ainsi, chez MAT, ces gènes sont associés à la plasticité de croissance et à la production de nouveaux organes : on observe l'augmentation des facteurs de transcription (complexe Médiateur, initiation de la transcription et TFIID, un facteur général de la transcription), ainsi que de régulateurs clefs de la SAR (NPR1 et SNI1). Cette configuration augmente le niveau basal d'activité transcriptionnelle de certains gènes de défense constitutive [Chen, Yang et al., 2022 ; Durrant, Wang et Dong, 2007]. De même, la surexpression de composantes ribosomales (uL2c) et de protéines PPR impliquées dans la maturation des ARNm signale une accélération de la traduction et de la machinerie post-transcriptionnelle, augmentant la plasticité et la prolifération cellulaires. MAT augmente aussi l'expression d'enzymes de biosynthèse (YUCCA pour l'auxine et voies flavonoïdes/anthocyaniques) et de protéines de remodelage pectique (pectinestérases, polygalacturonases, arabinogalactanes) ; autant d'éléments qui augmentent l'investissement dans l'expansion et la ramification racinaire [Stepanova et al., 2011 ; Ma, 2024]. Enfin, la surexpression ciblée de gènes de polarité et développement (SOSEKI, « embryo defective ») confirme l'activation de programmes morphogénétiques [Yoshida et al., 2019 ; Matthew et al., 2022].

Le morphotype aquatique ayant subi l'alternance (MA-AT) semble capable d'anticiper et de résister aux contraintes terrestres (variation hydrique, stress oxydant, pres-

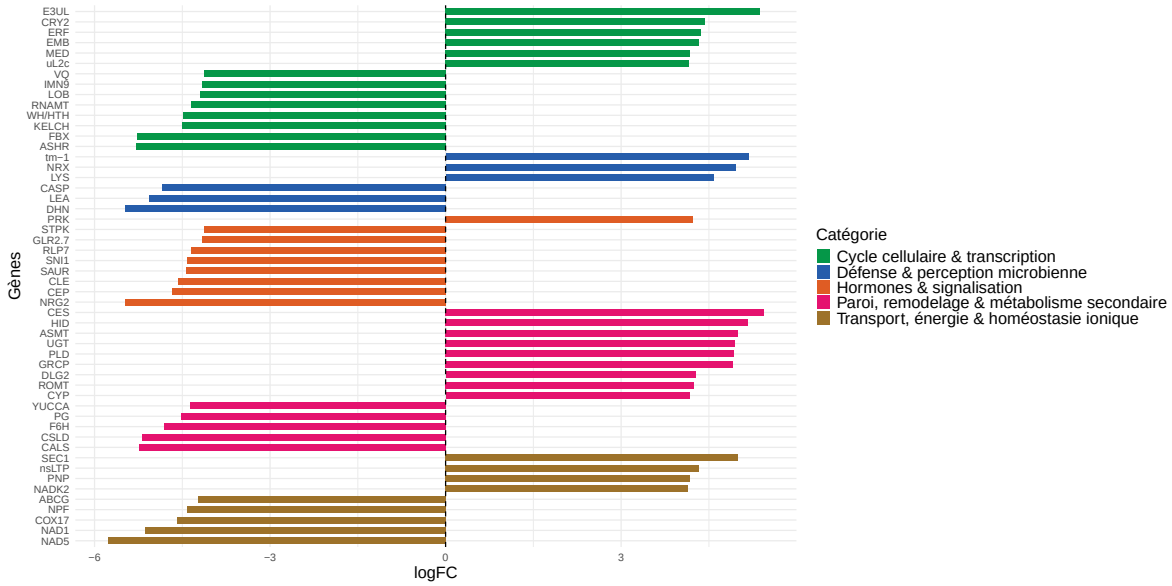


FIGURE 3.32 – Gènes différentiellement exprimés entre MAT et MTT chez *Lgh*. Les logFC positif correspondent à une surexpression dans MAT tandis que les négatif à une surexpression dans MTT. Les correspondances entre les noms de gènes et leur fonction sont en Table S1.

sion microbienne) en réorganisant l'activité transcriptionnelle et métabolique, ce que ne montre pas le morphotype aquatique conditionné en terrestre (MAT). Cependant, ceci comporte chez MA-AT, une contrepartie avec une capacité réduite d'expansion racinaire et d'exploration du sol si des conditions d'acquisition rapide d'eau/nutriments l'exigent. MAT, quant à lui, présente un comportement favorisant la division et l'expansion racinaire.

Comparaison MAT vs MTT

En milieu terrestre, le morphotype aquatique conditionné 6 mois en terrestre (MAT) montre une réorganisation profonde de l'expression des gènes liés au cycle cellulaire et à la transcription par rapport au morphotype terrestre (MTT), révélant un basculement stratégique de son programme physiologique (Figure 3.32).

L'induction de gènes chez MAT tels que uL2c, MED, EMB, ERF, CRY2 et les ligases E3 met en évidence un recentrage des ressources vers un maintien de la transcription de base, un renforcement de la signalisation hormonale (notamment *via* l'éthylène), une adaptation photophysiologique et un remodelage post-traductionnel rapide du protéome. MED notamment relie des facteurs de transcription activateurs/répresseurs, impliquées dans la signalisation hormonale (ABA/éthylène), les stress (sécheresse ou stress oxydatif). Son activation transcriptionnelle massive montre que le passage en milieu terrestre nécessite l'activation rapide de programmes de protection (ABA, LEA, antioxydants) [Chong, Guo et Zhu, 2020]. Les ERFs régulent également les signaux éthylène, acide abscissique, jasmonate et redox dans la réponse des plantes à un certain nombre de stress abiotiques. Le passage du milieu aquatique au milieu terrestre peut modifier la diffusion d'éthylène et la dynamique de l'ABA, expliquant l'induction d'ERF et un remodelage tissulaire [Müller et Munné-Bosch, 2015 ; Kong et al., 2023].

La protéine NRX (nucleoredoxin), une protéine redox, protège les enzymes antioxy-

dantes comme la catalase en maintenant leurs résidus cystéine dans un état réduit, évitant ainsi leur inactivation par oxydation [Hazra et al., 2024]. Cette activité est très importante car en conditions terrestres (sol, sécheresse, fluctuations hydriques, exposition aux rayonnements, etc.), les plantes sont soumises à des stress oxydants car la production de radicaux libres (ROS, par exemple H2O2) est accrue et peut endommager les protéines, les lipides, l'ADN et les membranes. Pour survivre, les plantes disposent de systèmes antioxydants (catalases, peroxydases, enzymes de la famille thioredoxine/glutaredoxine, etc.) qui éliminent ou neutralisent les ROS. En protégeant ces enzymes, NRX renforce la capacité de MTA à contrôler l'équilibre redox et à limiter les dommages oxydatifs. Les autres gènes surexprimés dans MAT incluent des kinases de paroi cellulaire (PERK), des enzymes de biosynthèse secondaire telles que CYP et ROMT, ainsi que des protéines structurales de la paroi comme GRCP et PLD qui améliorent la signalisation lipidique. Des enzymes métaboliques comme UGT, ASMT, HID, CES, NADK2 et PNP sont également surexprimées, reflétant une réorientation métabolique pour la production de composés protecteurs, la régulation énergétique et la tolérance aux stress oxydatifs.

Concernant les mécanismes de défense chez MAT, certains gènes sont réprimés comme par exemple, le gène codant la protéine CASP. Cette protéine est impliquée dans la formation de la bande de Caspary, structure riche en lignine, localisée dans les cellules de l'endoderme. Comme cette bande agit comme une barrière de diffusion apoplastique, elle bloque le passage non sélectif de l'eau et des solutés et les oblige à passer par des transporteurs membranaires "sélectifs" vis-à-vis de ce qui entre dans le cylindre vasculaire. Sa réduction conduit à une moindre lignification de la bande de Caspary, donc une barrière endodermique moins rigide, ce qui peut favoriser le passage de l'eau même si elle est moins sélective.

La répression de gènes tels que ceux codant les protéines VQ, les importines, les pro-

téines LOB, les méthyltransférases d'ARN et d'histones, les protéines Kelch et F-box traduit une réduction coordonnée des mécanismes de régulation associés à la croissance et au développement ce qui ne semble pas être le cas de MTT. VQ est impliqué dans la régulation du développement et dans les réponses aux stress abiotiques et biotiques (jasmonate, lumière, stress) [Jing et Lin, 2015]. Sa répression dans MAT (par rapport à MTT) montre une réorganisation hormonale et un basculement prioritaire vers des réponses de survie. La répression de l'importine, qui contrôle le transport nucléaire de facteurs de transcription de division et de croissance, montre que MAT économise ses ressources pour faire face à la dessiccation et au stress mécanique en surface [Wang, Selinski et al., 2020 ; Zhao et al., 2018]. La répression des WH/HTH qui contrôlent des programmes de division cellulaire s'accorde avec une réduction générale de la prolifération/croissance quand la plante fait face à stress de dessiccation afin de rediriger l'énergie vers la protection (réparation, cuticule) [Teichmann, Dumay-Odelot et Fribourg, 2012].

Plusieurs autres gènes liés à la croissance, à la signalisation et à la respiration mitochondriale, tels que STPK, SNF1, RLP7, GLR2.7, CLE, NRG2, NPF, ABCG, PG, CALS, NAD1 et NAD5, sont réprimés chez MAT. Ceci traduit un ralentissement des voies énergétiques et de croissance et une optimisation de l'absorption de nutriments et de l'énergie, pour concentrer les ressources sur la survie et la protection cellulaire dans un milieu plus exigeant. Cette réorientation expressionnelle de MAT confère à la plante un avantage immédiat pour tolérer le stress lumineux, hydrique et oxydatif propre au milieu terrestre, mais elle se fait au détriment d'une croissance rapide et d'une morphogenèse active. Elle illustre une stratégie adaptative d'une plante soumise à une transition environnementale brutale, qui choisit de suspendre la croissance, de stabiliser l'appareil transcriptionnel et d'activer des voies de survie permettant de traverser une phase critique. Ce type de réponse, efficace à court terme pour résister aux

TABLE 3.4 – Bilan des principaux mécanismes mis en œuvre dans les racines par les différents morphotypes /conditions de *Lgh* lors d'une acclimatation au milieu terrestre.

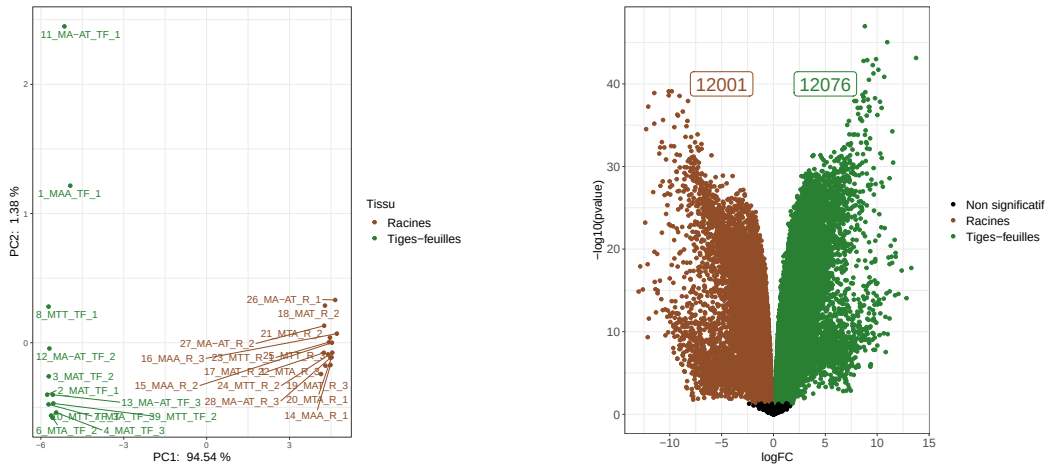
	Comparaison					
Caractéristiques	MA-AT vs MTT		MA-AT vs MAT		MAT vs MTT	
Réponse au stress	+	-	+	-	+	-
Croissance	-	+	-	+	-	+
Renforcement des parois	+	-	+	-	-	+
Plasticité cellulaire	-	+	-	+	-	+

nouvelles contraintes, peut toutefois limiter la capacité de développement à long terme si la situation stressante persiste.

En conclusion, lors de l'acclimatation au milieu terrestre, le morphotype aquatique ayant subi l'alternance (MA-AT) va mobiliser rapidement les voies de réponses aux stress et favoriser le renforcement de ses parois racinaires. Cependant, il va réduire sa croissance racinaire et sa plasticité cellulaire comparé à MAT et MTT (Table 3.4). MAT va avoir une réponse au stress plus importante que MTT tout en diminuant à la fois sa croissance, sa plasticité cellulaire et surtout le renforcement de ces parois. Le morphotype aquatique, quel que soit son conditionnement, perçoit le milieu terrestre comme un milieu stressant. À ce stade de notre étude, il est difficile de conclure sur l'impact de l'alternance sur l'acclimatation de *Lgh* au milieu terrestre. Par contre, nous avons observés que MA-AT mobilise des gènes de défense précoce comparé à MAT qui mobilise des gènes de défense tardive. Ceci traduit une réponse différente du morphotype aquatique en fonction du conditionnement.

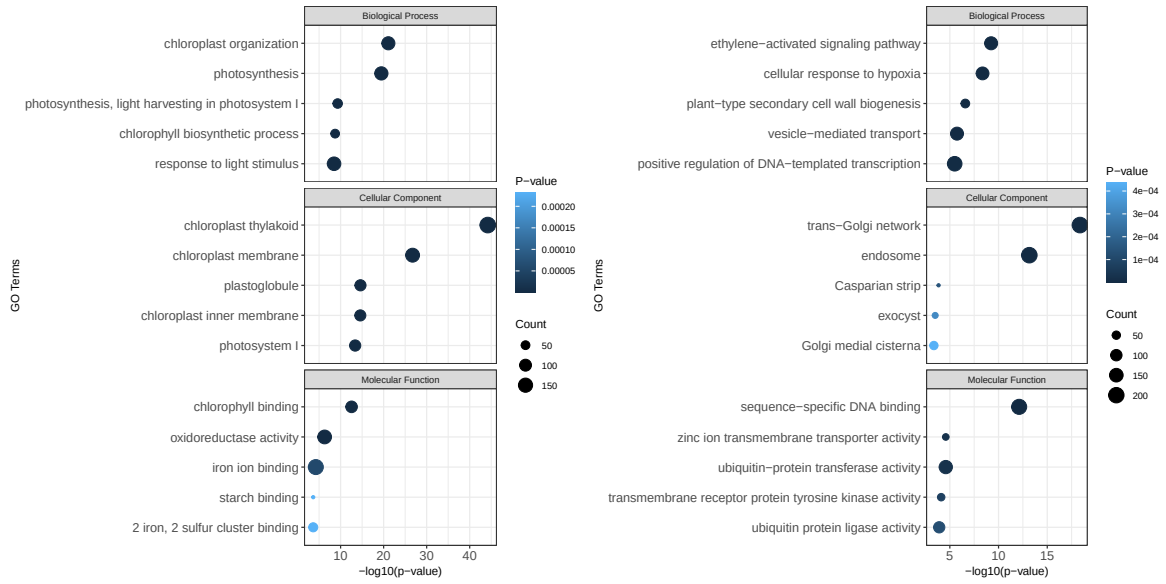
3.7 Analyses différentielles sur le transcriptome *de novo*

Les analyses présentées ici concernent une autre stratégie d'analyse de nos données transcriptomiques. Nous avons comparé l'expression différentielle de nos morphotype/ condition en utilisant le transcriptome *de novo* de *Lgh* comme référence. Ces comparaisons sont moins poussées que celles réalisées sur le génome et devront être approfondies. Ici, les triplicats ont été étudiés séparément permettant d'évaluer la variabilité entre eux.



A **B**

FIGURE 3.33 – Comparaison entre les échantillons provenant des tiges-feuilles et racines de *Lgh*. A : ACP montrant les différences d'expression des gènes entre les tiges-feuilles et les racines. B : Volcano plot des gènes différentiellement exprimés entre les tiges-feuilles et les racines.



A **B**

FIGURE 3.34 – Enrichissement GO des gènes différentiellement exprimés entre les tiges-feuilles et les racines *Lgh*. A : Gènes surexprimés dans les tiges-feuilles. B : Gènes surexprimés dans les racines.

Identification des gènes différentiellement exprimés entre les tiges-feuilles et les racines de *Lgh*

Dans le but d'effectuer des analyses différentielles, les reads de chacune des 28 banques séquencées ont été mappés séparément sur le transcriptome *de novo* de *Lgh* en utilisant Salmon. Dans un premier temps, les 28 échantillons ont été normalisés ensemble pour pouvoir comparer l'expression suivant le type de tissus (tiges-feuilles et racines). Une première Analyse en Composante Principale (ACP) a permis de mettre en évidence un échantillon trop éloigné des autres (5_MTA_TF_1 : Tiges et feuilles du premier réplicat du morphotype terrestre en condition aquatique). Cet échantillon sera par la suite retiré de l'analyse. Suite à ça, une deuxième ACP a été réalisée sans cet échantillon. La dimension 1 de cette ACP, représentant 94,54% des variations observés, sépare les tiges-feuilles des racines montrant ainsi une grande différence dans les gènes exprimés entres les deux tissus (Figure 3.33A) comme cela été également montré avec l'approche précédente. En effet, sur un total de 24 077 gènes différentiellement exprimés entre les tissus, 12076 et 12001 sont respectivement surexprimés dans les tiges-feuilles et les racines (Figure 3.33B).

Les termes GO de processus biologique significativement enrichis dans les tiges-feuillles sont "chloroplast organization", "photosynthesis" et "response to light stimulus". Ceux pour les termes GO de compartiment cellulaire et de fonction moléculaire sont, respectivement, "chloroplast thylakoid", "chloroplast membrane", "plastoglobule" et "chlorophyll binding" et "oxidoreductase activity" (Figure 3.34A). Les termes enrichis, que nous avons identifiés, sont des termes relatifs à la photosynthèse ou au chloroplaste ce qui n'est pas étonnant car les feuilles sont les organes centraux de la photosynthèse contrairement aux racines non exposées à la lumière. Pour les racines, la voie de l'éthylène est significativement enrichie avec le terme de processus biologique "ethylen-activated signaling pathway" (Figure 3.34B). L'éthylène est une hormone des plantes

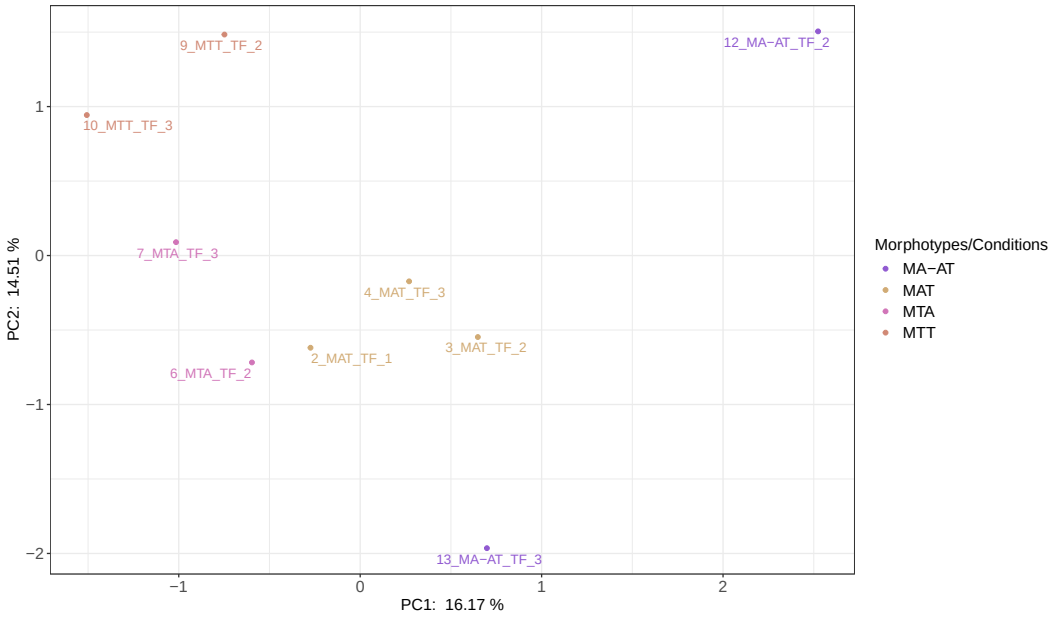


FIGURE 3.35 – ACP montrant les différences d'expression des gènes entre les échantillons de *Lgh* provenant des tiges-feuilles sans les échantillons 1_MAA_TF_1, 5_MTA_TF_1 et 8_MTT_TF_1, 11_MA-AT_TF_1.

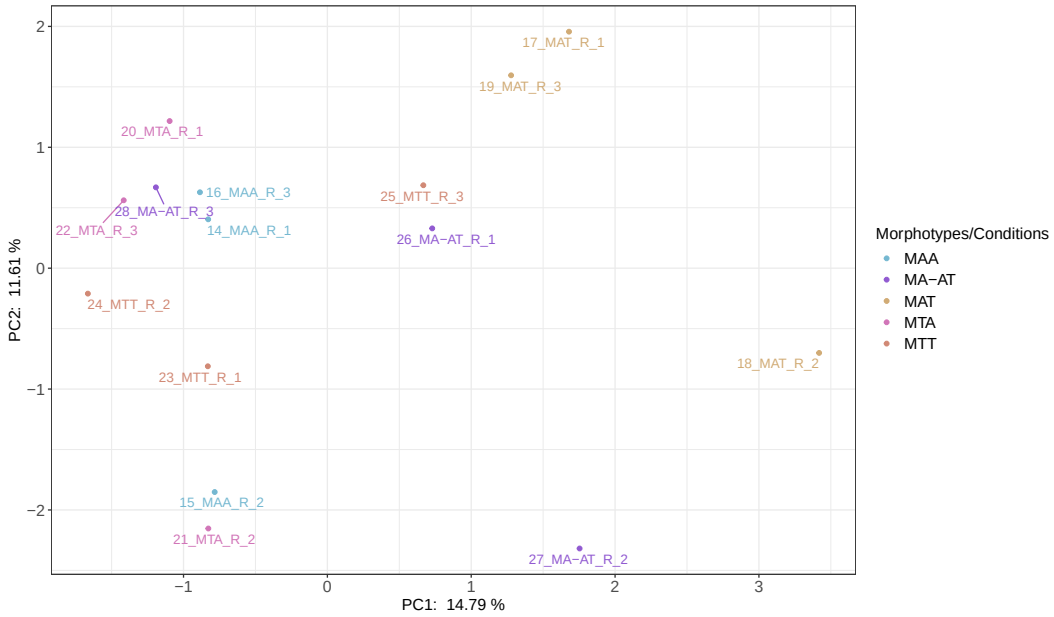


FIGURE 3.36 – ACP montrant les différences d'expression des gènes entre les échantillons de *Lgh* provenant des racines.

impliquée dans le développement de la plante notamment celui des racines [Casson et Lindsey, 2003]. Les termes "plant-type secondary cell wall biogenesis" ou "trans-Golgi network" se rapportent au développement cellulaire (Figure 3.34B). En effet, l'appareil de Golgi est impliqué dans le développement de la paroi cellulaire par la synthèse des polysaccharides [Dupree et Sherrier, 1998]. L'appareil de Golgi est très actif dans les racines des plantes permettant ainsi leur croissance. Chez le riz, les termes GO lié à la paroi cellulaire sont associés avec le développement de racines, ce qui corrobore nos résultats [Xu, Zou et al., 2019]. En plus, le terme GO "vesicle-mediated transport" est aussi lié à l'appareil de Golgi puisque l'appareil de Golgi produit des vésicules transportant les polysaccharides jusqu'à la membrane [Dupree et Sherrier, 1998].

Première exploration des réponses fonctionnelles de *Lgh* à son acclimatation au milieu terrestre

Étant donné la spécificité des gènes exprimés en fonction du type de tissu que nous avons observé, les échantillons provenant des tiges-feuilles et des racines ont été normalisés et comparés séparément. Pour les tiges-feuilles, en plus de l'échantillon 5_MTA_TF_1 déjà retiré de l'analyse précédente, le seul échantillon du morphotype aquatique en condition aquatique (1_MAA_TF_1) a aussi été retiré de l'analyse car sans réplicat biologique. Une première ACP a montré que les variations des axes 1 et 2 (36% et 10% des variations) était quasiment uniquement dû à deux échantillons issus du morphotype terrestre conditionné en terrestre (8_MTT_TF_1) pour l'axe 2 et du morphotype aquatique en condition d'alternance (11_MA-AT_TF_1) pour l'axe 1 (ACP non montré). Ceci a conduit à les retirer également pour la suite de l'analyse. Au final, pour les "morphotype-condition", MTT, MTA et MA-AT, seuls 2 réplicats biologiques ont été conservés. La seconde ACP des tiges-feuilles montre un bon regroupement des réplicats biologiques entre eux (Figure 3.35). L'axe 1, expliquant 16,2% des variations,

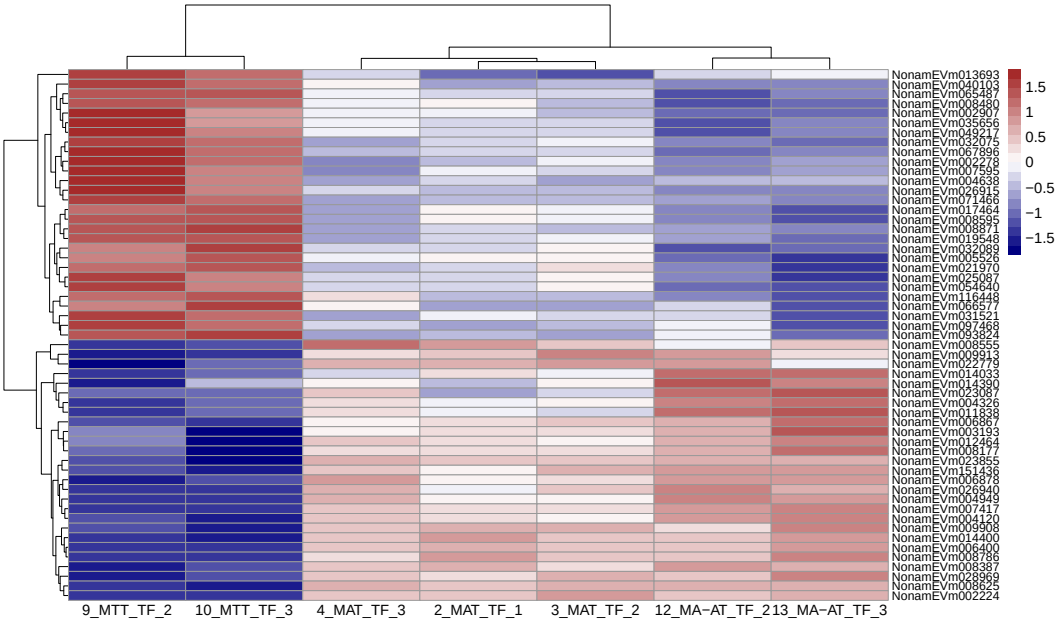


FIGURE 3.37 – Heatmap des gènes différentiellement exprimés entre MAAT, MAT et MTT chez *Lgh*. Les couleurs correspondent aux z-scores, une coloration rouge indique une surexpression du gène alors qu'une coloration bleue une sous-expression par rapport à la moyenne.

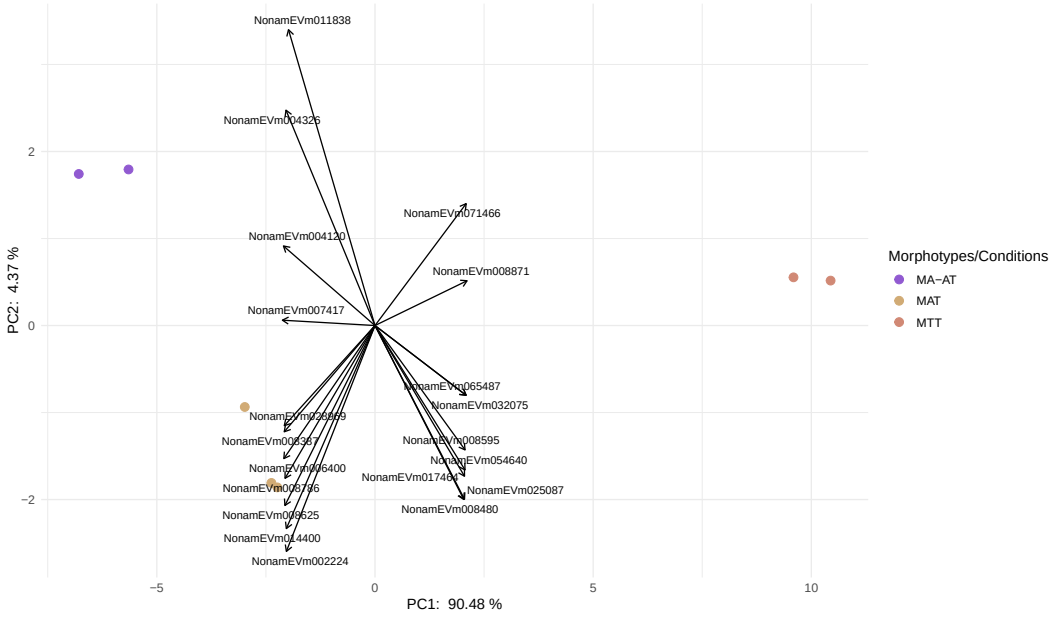


FIGURE 3.38 – ACP sur les gènes différentiellement exprimés entre MAAT, MAT et MTT chez *Lgh*. Les 20 gènes participant le plus aux deux axes sont représentés.

sépare les morphotypes terrestre et aquatique. L'axe 2 (14,5% de variation) sépare le morphotype terrestre en condition terrestre (MTT) des autres "morphotype-condition". Pour les racines, les réplicats biologiques ne se regroupent pas très bien ensemble (Figure 3.36) montrant ainsi une grande variabilité entre réplicats ou une faible variabilité entre les différentes conditions et/ou morphotypes.

Afin d'évaluer si l'alternance des conditions immersion et d'émersion subies par *Lgh in natura* a favorisé son acclimatation au milieu terrestre, nous avons comparé les expressions différentielles des gènes du morphotype aquatique ayant été directement conditionné en milieu terrestre (MAT), du morphotype aquatique ayant subi 6 cycles d'alternance "condition aquatique-condition terrestre" (MA-AT) et du morphotype terrestre élevé en condition terrestre (MTT). Il s'agit ici de comparer les réponses de ces 3 morphotypes x conditions à un développement en condition terrestre.

Pour les tiges-feuilles, l'expression différentielle des gènes du morphotype aquatique conditionné en terrestre (MAT), conditionné en alternance (MA-AT) et le morphotype terrestre conditionné en terrestre (MTT) a été comparée deux à deux. La comparaison de l'expression des transcrits de MA-AT vs MTT a mis en évidence 50 gènes différentiellement exprimés dont 24 sur-exprimés chez MA-AT et 26 chez MTT. La comparaison de l'expression des transcrits de MA-AT vs MAT n'a pas permis d'identifier de gène différentiellement exprimé. Enfin, la comparaison de l'expression des transcrits de MAT vs MTT a mis en évidence 28 gènes différentiellement exprimés dont 14 sur-exprimés chez MAT et 14 chez MTT. Dans les racines, aucun gène différentiellement exprimé n'a pu être identifié pour ces trois comparaisons.

La heatmap de tous les gènes différentiellement exprimés dans les feuilles-tiges montrent une séparation nette entre MTT d'un côté et MA-AT et MAT de l'autre (Figure 3.37). L'ACP réalisée avec les gènes différentiellement exprimés appuie également cette séparation avec l'axe 1 (90,5% des variations) séparant MTT et MA-AT et MAT (Figure

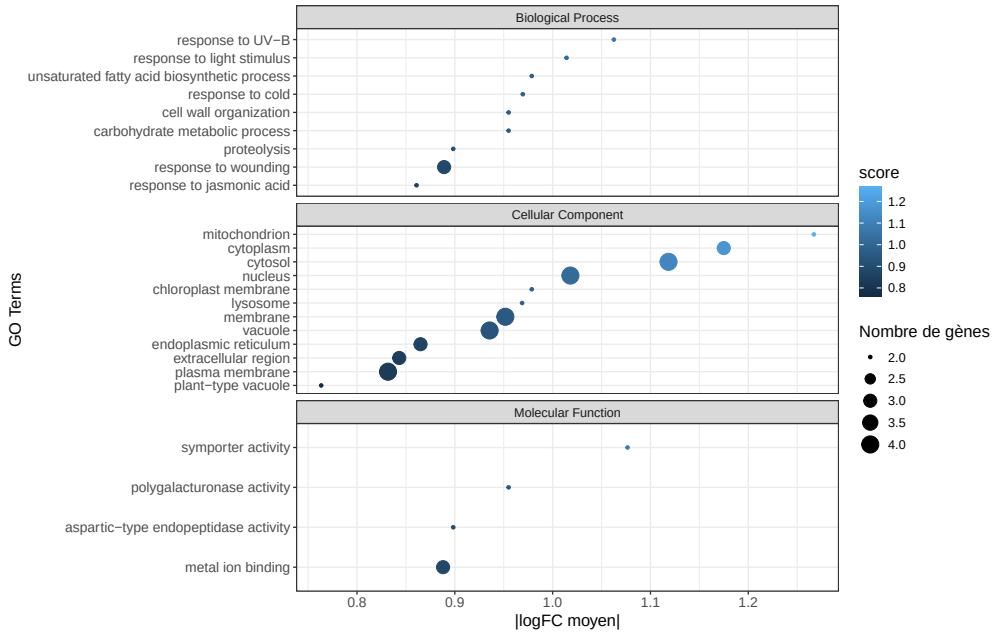


FIGURE 3.39 – Termes GO des gènes surexprimés dans MA-AT et/ou MAT par rapport à MTT dans les tiges-feuilles de *Lgh*.

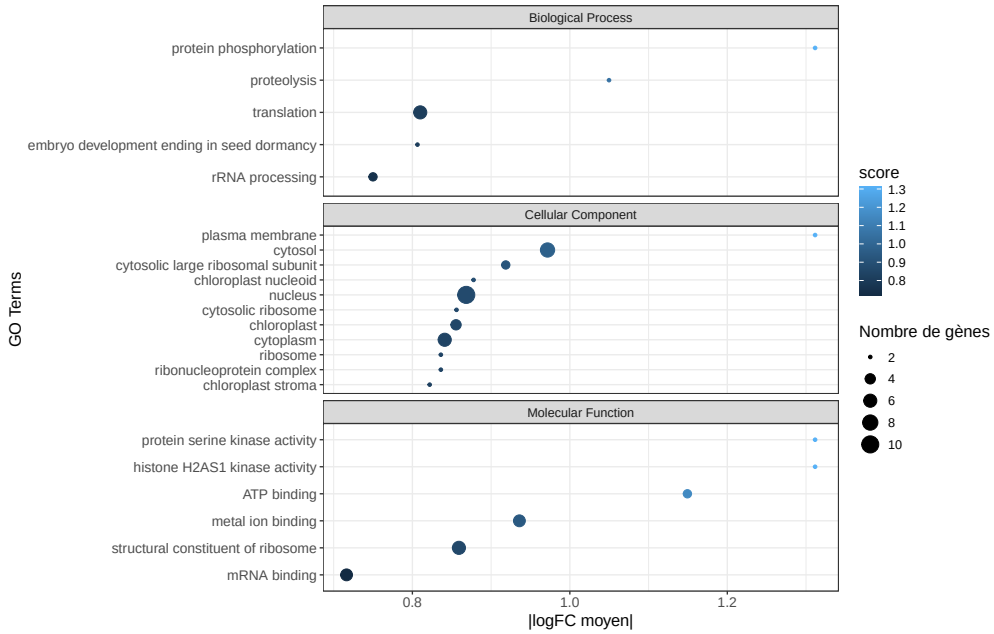


FIGURE 3.40 – Termes GO des gènes différentiellement surexprimés dans MTT par rapport à MA-AT et/ou MAT dans les tiges-feuilles de *Lgh*.

3.38). Ces résultats révèlent que des différences d'expression génique existent entre le morphotype aquatique élevé 6 mois en condition terrestre et le morphotype terrestre lorsque ces 2 morphotypes se développent en condition terrestre. De même, nos résultats montrent que suite au conditionnement en alternance "aquatique-terrestre", les réponses fonctionnelles du morphotype aquatique diffèrent de celles observées chez le morphotype terrestre lorsqu'ils sont élevés en condition terrestre. Cependant, l'application des cycles d'alternance sur le morphotype aquatique a conduit à des réponses différentes par rapport à son conditionnement direct en milieu terrestre (Figure 3.37).

L'étude des gènes différentiellement exprimés grâce à leurs annotations permettra de mieux comprendre les mécanismes sous-jacents à ces observations. Des voies de réponses au stress ont été identifiées dans les gènes sur-exprimés dans MA-AT et/ou MAT comme "response to UV-B", "response to cold" ou "response to wounding" (Figure 3.39). Tandis que pour MTT des voies plutôt relative à l'expression des gènes ("translation", "rRNA processing") ou au protéines ("protein phosphorylation", "proteolysis") sont sur-exprimés (Figure 3.40). Cependant, ces analyses restent préliminaires et devront être approfondies.

Comparaison des résultats obtenus *via* l'utilisation du transcriptome *de novo* et du transcriptome sur référence

Nous avons identifié beaucoup plus de gènes différentiellement exprimés dans nos comparaisons deux à deux lors de l'utilisation du transcriptome sur référence par rapport au transcriptome *de novo*. Les différences dans les méthodologies utilisées peuvent en partie expliquer ces observations comme par exemple, le fait de pooler les échantillons du transcriptome de référence. De plus, le génome et le transcriptome *de novo* sont des ressources pour le moment incomplètes. Il sera nécessaire de les améliorer pour pouvoir ensuite comparer les résultats obtenus et conclure pour les analyses sur l'acclimatation

de *Lgh* au milieu terrestre.

4 Conclusion

Le transcriptome *de novo* de *Lgh* vient compléter les ressources génomiques de cette espèce comprenant les génomes nucléaire, chloroplastique et mitochondrial. Il va permettre de réaliser des analyses d'expression de gènes et ainsi mieux comprendre les mécanismes de changement de milieu de la jussie. Il a encore besoin d'être améliorer mais il permettra d'obtenir une meilleure version du génome de *Lgh*. Les analyses réalisées pour le moment sur le transcriptome *de novo* et sur référence ont permis de mettre en évidence un nombre important de gènes avec une expression spécifiques aux racines et aux tiges-feuilles. La stratégie mise en place sur le transcriptome de référence semble plus efficace pour identifier les gènes différentiellement exprimés et sera à privilégier pour la suite de nos analyses.

Les premières comparaisons de l'expression des gènes lors d'une acclimatation en milieu terrestre montre les différences de réponses entre MA-AT, MAT et MTT. MA-AT et MAT semblent percevoir le milieu terrestre comme un milieu stressant et adoptent des stratégies différentes. MA-AT réduit sa croissance racinaire et renforce ces tissus racinaires alors que MAT favorise son développement racinaire en diminuant les barrières de transport de l'eau. Il faudra cependant des analyses plus approfondies pour pouvoir tirer des conclusions quant aux rôles de l'alternance dans l'apparition du morphotype terrestre.

Par ailleurs, l'épigénétique est souvent évoquée comme un "driver" de l'adaptation rapide des espèces invasives dans leur nouvel environnement [Marin et al., 2020]. Ainsi, l'activité des éléments transposables (ET) et les modifications épigénétiques peuvent contribuer au succès des invasions en facilitant à la fois l'évolution adaptative et la

plasticité phénotypique [Marin et al., 2020]. Les mécanismes épigénétiques peuvent participer à la variation de l'expression génétique et agir comme régulateurs du développement et de la réponse au stress [Mounger et al., 2021]. Ainsi, lors de l'invasion et de l'adaptation à un nouvel environnement, le rôle de l'activité des éléments transposables (ET) et des modifications épigénétiques peut être crucial. On sait en effet que l'activité des ET et les modifications épigénétiques sont stimulées par des environnements nouveaux ou stressants, et ces deux mécanismes peuvent agir indépendamment ou conjointement, car ils sont tous deux sensibles aux changements environnementaux [Nota et al., 2024]. Cette piste est à explorer pour comprendre l'acclimatation de *Lgh* au milieu terrestre.

Ce travail sera valorisé au travers de deux publications lors d'un post-doctorat en 2026 : (1) Une amélioration du génome de *Lgh* ainsi qu'une comparaison avec le transcriptome et (2) Une étude fonctionnelle sur l'acclimatation de *Lgh* au milieu terrestre en utilisant les données transcriptomiques.

Table S1: Correspondance des noms de gènes avec leur fonction.

Nom du gène	**Fonction**
F6'H1	oxoglutarate (2og) and fe(ii)-dependent oxygenase (flavonoid 6'-hydroxylase 1)
HID	2-hydroxyisoflavanone dehydratase
BCKDHA	2-oxoisovalerate dehydrogenase
DHQS	3-dehydroquinate synthase
KDSR	3-dehydrosphinganine reductase
ABCG	abc transporter g
ABCF	abc transporter f
PAT04	s-acyltransferase 4
ASMT	acetylserotonin o-methyltransferase
ABD	actin binding protein
ALPK1	alpha-protein kinase 1
ANK	ankyrin repeat-containing protein
AGP23	arabinogalactan protein 23
ATPSCF1	atp synthase cf1
ATPSϵ	atp synthase epsilon chain
BAHD	bahd acyltransferase
BGL	beta-galactosidase
CALS	callose synthase
CAM	calmodulin cam
CES	carboxylesterase
CASP	casp-like protein 2a1
CSLD	cellulose synthase-like protein d
CLE	clavata3/esr (peptide hormone receptor)
3AT1	coumaroyl-coa:anthocyanidin 3-o-glucoside-6"-o-coumaroyltransferase
DEG23	cttnbp 2 amino-terminal-like protein
CUPIN	cupin
CRY2	cryptochrome-2 (blue light receptor)
CPI	cysteine proteinase inhibitor
COX17	cytochrome c oxidase copper chaperone
petG	cytochrome b6-f complex subunit 5
CYP	cytochrome p450 (monooxygenase)
DHN	dehydrin (late embryogenesis abundant protein)
DIR2	dirigent protein 2-like
RGA	disease resistance protein
EMB	embryo defective
DUF2921	duf2921 (domain of unknown function 2921)
DLG2	duf642 l-galactono-1,4-lactone-responsive gene 2
DUF740	duf740 (domain of unknown function 740)
E3UL	e3 ubiquitin-protein ligase

ETFβ	electron transfer flavoprotein subunit beta (beta-etf)
ERF	ethylene-responsive transcription factor erf
FBX	f-box/kelch protein
FHA-PS1	fha domain-containing protein ps1
F6H	flavonoid-6-hydroxylase
FMO5	flavin-containing monooxygenase 5
GCP2	gamma-tubulin complex protein 2
GLP	germin-like protein
GRCP	glycine-rich cell wall structural protein
GLR2.7	glutamate receptor 2.7-like
HSF	heat shock factor protein hsf
HIPP	heavy metal-associated isoprenylated plant protein
ASHR	histone-lysine n-methyltransferase ashr
HOL	holotricin
UGT-inactive	inactive udp-glycosyltransferase
YUCCA	indole-3-pyruvate monooxygenase yucca
IMN9	importin-9 isoform
JOX2	jasmonate-induced oxygenase
KANSL1	kat8 regulatory nsl (histone acetylation complex subunit)
KELCH	kelch domain-containing protein
PIX13	kinase pix13
TAA1	l-tryptophan–pyruvate aminotransferase
LECRK41	l-type lectin-domain containing receptor kinase
uL2c	large ribosomal subunit protein ul2c
LRK10L-1.2	leaf rust 10 disease-resistance locus receptor-like protein kinase
LEA	late embryogenesis abundant protein-like
LEB4	legumin type b
PRXA3a	lignin-forming anionic peroxidase-like
LOB	lob domain-containing protein
LYS	lysozyme-like protein
MADS	mads-box protein (transcription factor)
PRUA	major allergen pru
MED	mediator of rna polymerase ii
GRXS1	monothiol glutaredoxin-s1
MYB	myb protein
NAC	nac domain-containing protein
NAD1	nadh dehydrogenase subunit 1
NAD5	nadh dehydrogenase subunit 5
NADK2	nadh kinase isoform x4
NDOR1	nadph-dependent diflavin oxidoreductase
nsLTP	non-specific lipid-transfer protein
NPF	nrt1/ptr (nitrate/peptide transporter family)
NRG2	nitrate regulatory gene2

NRX	nucleoredoxin
NFD	nuclear fusion defective
PNP	nucleoside phosphorylase
OGD	oxoglutarate-dependent dioxygenase
PNPLA3	patatin-like protein
PR1	pathogenesis-related protein 1-like
PME3	pectinesterase 3
PPR	pentatricopeptide repeat-containing protein
PRX	peroxidase
PHF	phd finger (zinc finger domain)
PLD	phospholipase d/transphosphatidylase
PG	polygalacturonase
CEP	precursor of cep
PRK	proline-rich receptor-like protein kinase perk
ZW2	protein zw2
PRO-TRP	transporteur de dipeptides (pro-trp)
RIC1	rab6a-gef complex partner (ric1)
RADIALIS	radialis
RLP7	receptor-like protein 7
RNAMT	rna n6-adenosine-methyltransferase
SAUR	saur family protein
SCAR	scar (actin regulatory protein)
SEC1	sec1 (vesicle trafficking protein)
SLAH1	s-type anion channel slah1
STPK	serine/threonine-protein kinase
PBS1	serine/threonine-protein kinase pbs1
SEOR1	sieve element occlusion amino-terminus protein
DEVIL	small polypeptide devil
SNI1	snf1-related protein kinase
SnRK1	soseki (polarity protein)
SNI1	suppressor of npr1-1
MTHFD1	tetrahydrofolate dehydrogenase/cyclohydrolase
TMV	tmv (tobacco mosaic virus protein)
tm-1	tomv susceptible protein tm-1
TPR1	tetratricopeptide repeat protein 1
TLP	thaumatin-like protein
ROMT	trans-resveratrol di-o-methyltransferase
MYB	transcription factor myb1
SPT	transcription factor spt20
TGAL	transcription factor tgal
TFIID	transcription initiation factor tfiid
MTEF	transcription termination factor mtef
TRNT	trna n6-adenosine threonylcarbamoyltransferase

TR1	tropinone reductase
UGT	udp-glycosyltransferase
UCKL1	uridine kinase-like
UFC	upstream of flc (regulator of flowering locus c)
VQ	vq motif-containing protein
VIN3	vin3-like protein
WDR1	wd repeat-containing protein
BRUTUS	zinc finger protein brutus
ZFH1	zinc-finger homeodomain protein
ZF-GIS3	zinc finger protein gis3
WH/HTH	winged helix-turn-helix dna-binding

Chapitre 4

Discussion et perspectives

1 Bilan général de la thèse

Les travaux de cette thèse viennent en complément des études de Dominique Barloy sur la jussie, *Ludwigia grandiflora* subsp. *hexapetala*, espèce aquatique envahissante. Cela s'ajoute également aux travaux de thèse de Julien Genitoni visant à mieux comprendre l'acclimatation de cette espèce au milieu terrestre. La thèse de Julien Genitoni s'est concentré sur la caractérisation des morphotypes aquatique et terrestre aux niveaux morphologique et métabolique mais également sur l'implication de l'épigénétique dans l'acclimatation au milieu terrestre.

L'objectif de mon travail a été la génération de ressource génomique pour *Ludwigia grandiflora* subsp. *hexapetala* dans le but de mieux comprendre son acclimatation au milieu terrestre. Dans le premier chapitre, les assemblages des génomes mitochondriaux de *Lgh* et *Lpm* ont été décrits. Le deuxième chapitre a présenté l'assemblage du génome nucléaire de *Lgh*. Le troisième chapitre a décrit l'assemblage du transcriptome de *Lgh* ainsi que des premières analyses relatives à son acclimatation au milieu terrestre. Les conclusions majeures de ces travaux sont :

— (1) Les génomes mitochondriaux de *Lgh* et *Lpm* assemblés sont composés deux molécules circulaires (M1 et M2). Ils ont une taille totale de, respectivement, 711 578 et 722 518 pb. Des recombinaisons basées sur des séquences répétées de plus de 1 kb ont été identifiées dans les deux génomes mitochondriaux montrant le dynamisme des mitogénomes de plantes.

— (2) Le draft du génome nucléaire de *Lgh* est composé de 111 219 contigs pour une taille totale de 1,487 Gb. Plus de 246 000 gènes codant des protéines ont été annotés mais moins de la moitié ont reçu une annotation fonctionnelle. Parmi ces gènes, un grand nombre est spécifique à *Lgh* lors d'analyses de comparaison d'orthologues avec des espèces proches phylogénétiquement.

— (3) Le transcriptome *de novo* de *Lgh* est composé de 182 091 transcrits. Une

grande partie des transcrits assemblés (95,8%) ont été mappé sur le génome montrant de *Lgh* sa bonne complétude.

— (4) L'analyse des transcriptomes *de novo* et sur référence ont permis d'identifier un grand nombre de gènes différentiellement exprimés entre les tiges-feuilles et les racines.

— (5) En se basant sur les analyses d'expression des gènes, l'alternance de cycle "aquatique-terrestre" n'aurait pas participé seul à l'acclimatation du morphotype aquatique au milieu terrestre.

Les travaux de ma thèse seront discutés sur trois points importants : (1) la stratégie de séquençage, (2) les erreurs dans les bases de données et (3) l'intérêt de données multi-omiques. Puis, trois apports principaux de ces travaux seront détaillés : (1) l'apport pour la compréhension de l'acclimatation de *Lgh* au milieu terrestre, (2) l'apport des ressources génomiques en elle-même dans la famille des Onagraceae, pour l'étude d'espèces polyploïdes et non-modèle d'intérêt écologique et enfin (3) l'apport pour la méthodologie d'assemblage de ressources génomiques complexes.

2 Stratégies de séquençage

Les séquençages long reads de l'ADN de *Lgh* et *Lpm* n'ont pas permis d'obtenir une profondeur de séquençage élevée. En effet, seulement 2386 Mb (taille du génome estimé à 1419 Mb) et 512 Mb (taille du génome estimé à 262 Mb) ont pu être séquencées pour *Lgh* et *Lpm*, respectivement. Ainsi, une profondeur de séquençage de 1,7x et 1,9x a été obtenue pour respectivement *Lgh* et *Lpm*. Bien que suffisantes pour l'assemblage des génomes des organelles, ces profondeurs de séquençage ont rendu l'assemblage du génome nucléaire compliqué et abouti à la production d'un génome assemblé de *Lgh* fragmenté (111 219 contigs). De plus, la faible profondeur du séquençage ne nous a

pas permis de prendre en compte la polyploïdie de *Lgh*. En effet, la résolution des haplotypes demande une profondeur de séquençage entre 5 et 20x par chromosome homologue [Kyriakidou et al., 2018]. Cependant, réaliser un autre run de séquençage aurait pu améliorer notre jeu de données mais cela n'a pas été envisagé durant ma thèse afin de valoriser les données déjà produites et à cause de la complexité du protocole d'extraction d'ADN longs brins. La richesse en polyphénols et polysaccharides dans les tissus de la jussie a conduit à la mise au point d'un protocole d'extraction d'ADNs longs fragments spécifique, n'utilisant que partiellement les kits standards. En effet, ces composés peuvent réduire le débit et la qualité d'un séquençage nanopore [Sawicki et al., 2024]. De plus, l'obtention de longs brins d'ADN a été complexe et les tailles des séquences obtenues suite au séquençage nanopore, étaient faible avec un N50 de 6840 et une taille maximum de $\approx$ 73 kb pour *Lgh*, en deçà de l'attendu de ce qui est rapporté pour un séquençage de ce type. Le séquenceur GridION, utilisé pour *Lgh*, devrait produire des reads avec un N50 entre 10 et 60 kb et une taille maximum pouvant atteindre plus de 1000 kb [Logsdon, Vollger et Eichler, 2020]. En plus, le débit attendu se situe entre 2 et 20 Gb par run de séquençage bien au-dessus de ce que nous avons obtenu pour *Lgh* (795 Mb en moyenne par run, $\approx$ 2 Gb en tout).

En outre, le séquençage a été effectué en 2020 avec les technologies présentes sur les plateformes, à savoir Illumina MiSeq et GridION d'Oxford Nanopore. Or, les techniques de séquençage, surtout de troisième génération, ont évolué depuis. Le séquenceur Sequel II de PacBio, produisant les premiers reads HiFi (99% de précision), a été mis sur le marché en 2019. Cette technologie a divisé le coût du séquençage par 8 pour les séquençages PacBio, les rendant beaucoup plus accessibles [Adewale, 2020]. La même année, la technologie PromethION d'Oxford Nanopore venait aussi d'être lancé, permettant un très haut débit de séquençage. L'utilisation de ces technologies aurait peut-être pu permettre d'obtenir une meilleure profondeur de séquençage pour *Lgh*. En effet, la tech-

TABLE 4.1 – Stratégies de séquençage d'assemblages de génomes de plantes récents (2024-2025).

Espèces	**Ploïdie**	**Séquençage SR**	**Séquençage LR**	**Autre séquençage**	**Référence**
Vitis vinifera	2x	Illumina	ONT PromethION	/	[Ritter et al., 2025]
Cardamine chenopodiifolia	8x	/	PacBio HiFi	Omni-C	[Emonet et al., 2024]
Ulmus minor	2x	Illumina	PacBio HiFi	Hi-C	[Pallares-Zazo et al., 2025]
Camellia rubituberculata	2x	Illumina	PacBio HiFi	Hi-C	[Yan et al., 2025]
Chamaerops humilis	2x	/	PacBio HiFi	Hi-C	[Beltran-Sanz et al., 2025]
Actinidia deliciosa	6x	Illumina	PacBio HiFi	/	[Parra et al., 2025]
Sarcandra glabra	/	/	PacBio HiFi	Hi-C	[Hu et al., 2025]
Hippophae salicifolia	2x	Illumina	PacBio HiFi	Hi-C	[Zeng et al., 2025]
Wurfbainia villosa	/	Illumina	ONT PromethION	Hi-C	[Zhang, Yang et al., 2025]
Solanum acaule	4x	Illumina	PacBio HiFi et ONT PromethION	Hi-C	[Achakkagari et al., 2025]
Camellia granthamiana	4x	/	PacBio HiFi et ONT PromethION	Omni-C	[Hui et al., 2024]
Lolium perenne	2x	/	ONT PromethION	Hi-C	[Chen, Kölliker et al., 2024]
Dysphania ambrosioides	4x	/	PacBio HiFi	Hi-C	[Frandsen et al., 2025]
Mitragyna parvifolia	4x	Illumina	ONT PromethION	Hi-C	[Laforest et al., 2025]
Ficus carica	2x	Illumina	PacBio HiFi	Hi-C	[Ren, Gu et al., 2025]
Zoysia macrostachya	4x	Illumina	ONT MinION	Omni-C	[Park et al., 2025]

nologie Sequel II a un débit moyen entre 15 et 30 Gb et la technologie PromethION utilisant 48 "flow cell" entre 2,4 et 4,8 Tb (entre 50 et 100 Gb par "flow cell") [Logsdon, Vollger et Eichler, 2020]. Aujourd'hui, le taux d'erreurs des reads HiFi est même descendu à 0,01%, équivalent aux taux d'erreurs de short reads [Dijk et al., 2023]. Dans les assemblages récents, le séquençage de reads HiFi combiné à un séquençage Hi-C ou Omni-C est la méthode la plus utilisée (Table 4.1). Les séquençages SR sont moins utilisés au vu de la précision des reads HiFi mais ils restent utiles pour la correction des LR et l'estimation des caractéristiques du génome (taille, proportion d'hétérozygote ou taux de répétitions) [Ren, Gu et al., 2025 ; Zeng et al., 2025].

3 Erreurs dans les bases de données

En plus de la difficulté rencontrée lors de l'assemblage, il est possible que d'autres erreurs soient présentes dans les ressources génomiques de *Lgh*. En effet, des problèmes au cours des étapes de décontamination et d'annotation peuvent être dues à des erreurs dans les bases de données utilisées.

3.1 Erreurs dans la prédiction des séquences protéiques

La prédiction de gènes des eucaryotes est complexe à cause du peu de gènes présents dans un génome comparé aux séquences intergéniques, de la présence d'introns et de l'épissage alternatif [Prosdocimi et al., 2012 ; Salzberg, 2019]. Les structures introns-exons seraient correctement prédites dans seulement 50 à 60% des gènes annotés [Prosdocimi et al., 2012]. En effet, surtout pour les espèces non-modèles, beaucoup de gènes n'ont pas de preuves expérimentales validant leur structure [Meyer et al., 2020]. Ces erreurs risquent ensuite de se propager dans les bases de données, entraîner de mauvaises interprétations et peuvent même s'amplifier. Par exemple, dans une étude réalisée chez

13 vertébrés, un taux d'erreurs de 41% dans les séquences protéiques a ensuite conduit à 57% d'erreurs dans la prédiction d'événements de duplication puis à 75% d'erreurs d'interprétations dans les analyses fonctionnelles correspondantes [Prosdocimi et al., 2012].

L'annotation du génome de *Lgh* contient donc très probablement un nombre d'erreurs important dans les séquences des gènes codant des protéines. En effet, comme très peu de ressources génomiques dans des espèces proches étaient disponibles, *Arabidopsis thaliana* a été utilisé comme modèle de prédiction de gènes. Bien qu'appartenant au clade des Malvidées comme *Lgh*, ces deux espèces restent très éloignées phylogénétiquement. L'utilisation du séquençage de l'ARN permettra d'améliorer l'annotation du génome de *Lgh* rendant les prédictions plus précises [Salzberg, 2019].

3.2 Erreurs dans les annotations fonctionnelles

L'annotation fonctionnelle des gènes est aussi sujet à des erreurs notamment de par le manque de validations expérimentales [Schnoes et al., 2009]. En effet, ces annotations se font surtout par des méthodes informatiques. L'augmentation du nombre de séquences dans les bases de données a conduit à une augmentation du nombre d'erreurs d'annotations [Schnoes et al., 2009]. La proportion d'erreurs a ainsi augmenté passant de presque aucune en 1993 à 40% en 2005. Pour les annotations GO, un taux d'erreurs de 30% a été estimé [Jones, Brown et Baumann, 2007]. Ce taux est surtout dû à des annotations par similarité de séquences (ISS pour Inferred by Sequence Similarity) qui contiennent 49% d'erreurs contre seulement $\approx$ 15% pour les non-ISS. Des logiciels comme InterProScan n'utilise que des termes GO provenant d'une base de données réduites contenant seulement des annotations vérifiées, ce qui permet de réduire ces erreurs [Jones, Binns et al., 2014]. Les bases de données UniProt et SwissProt contiendraient quant à elles entre 33 et 43 % d'erreurs [Jones, Brown et Baumann, 2007]. Les

propagations d'erreurs contribuent beaucoup aux mauvaises annotations de séquences [Schnoes et al., 2009]. En effet, une fois l'erreur rentrée dans une base de données si un autre génome se base sur celle-ci pour une annotation, il sera difficile de corriger la deuxième annotation même si l'erreur initiale est ensuite corrigée.

Le génome et le transcriptome de *Lgh* ne sont sûrement pas épargnés par les erreurs d'annotations fonctionnelles. À cause du nombre important de gènes/transcrits, il n'a pas été possible de biocurer manuellement les annotations. De plus, des outils utilisés comme Diamond [Buchfink, Reuter et Drost, 2021] ou BLAST [Altschul et al., 1990] se basent sur des similarités de séquences pour annoter, augmentant ainsi le risque d'erreurs. Certaines annotations pourront être vérifiées grâce aux données de RNA-Seq [Hoff et Stanke, 2015]. Malgré tout, les annotations des gènes devront être traitées avec précaution tout comme les interprétations qui en découlent [Kress et al., 2022].

3.3 Contaminations

Comme présenté en introduction, l'étape de décontamination est une étape clé lors de l'assemblage d'un génome ou d'un transcriptome. Cependant, tout comme l'annotation, elle nécessite l'utilisation de bases de données qui peuvent contenir des erreurs. Tout d'abord, la mauvaise assignation taxonomique est une erreur importante [Chorlton, 2024]. Par exemple, dans la base de données GenBank, 3,6% des génomes de procaryotes ne correspondent pas à la bonne espèce [Ciufo et al., 2018]. Les bases de données peuvent aussi avoir des problèmes de sur ou sous-représentation taxonomique avec certaines espèces modèles surreprésentées [Chorlton, 2024]. C'est le cas dans la base de données RefSeq, où chez les bactéries, plus de 50% des génomes correspondent à 20 espèces pathogènes tandis que de nombreuses espèces manquent toujours de ressources génomiques. Enfin, la présence de séquences contaminantes dans les génomes dans les bases de données est aussi un facteur d'erreur important [Lupo et al., 2021]. Dans la

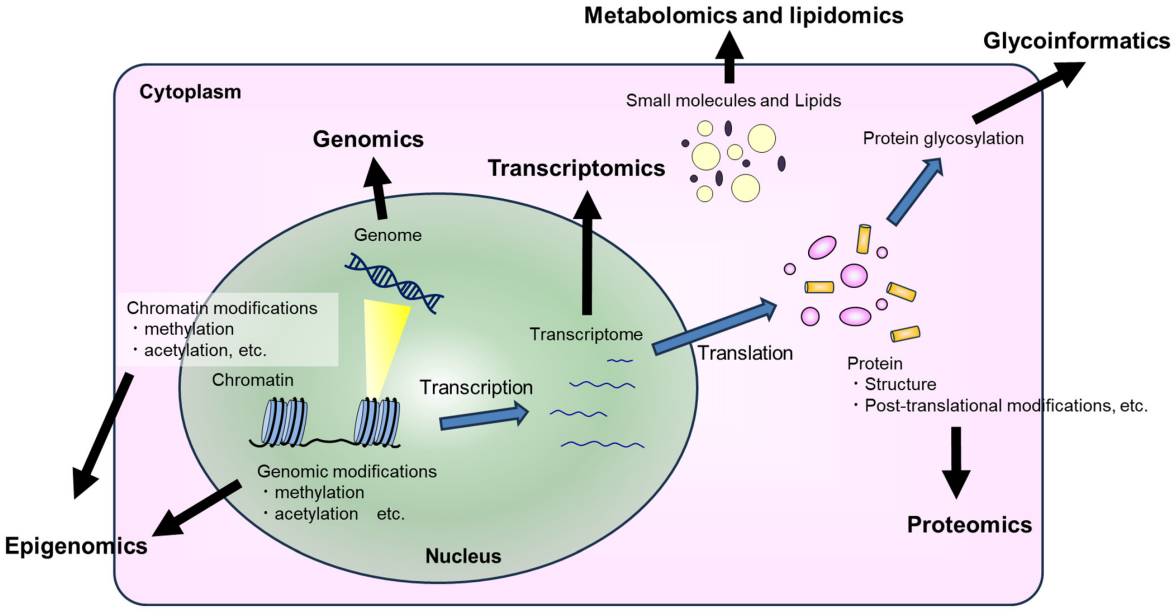


FIGURE 4.1 – Représentation de différentes méthodes "omiques" [Hayes et al., 2024].

base de données RefSeq, environ 1% des génomes de procaryotes contiennent plus de 5% de séquences contaminantes. Pour pallier les problèmes évoqués, il est recommandé d'utiliser plusieurs outils de décontamination ainsi que des bases de données adaptées [Chorlton, 2024 ; Lupo et al., 2021].

Pour la décontamination du génome et du transcriptome de *Lgh*, nous avons utilisés plusieurs outils différents (Kraken2, BLAST, Diamond, Kaiju, FCS GX). Cependant, nous avons fait le choix de les utiliser de façon itérative et nous n'avons donc pas pu comparer les résultats de ces outils entre eux. Par ailleurs, les bases de données, que nous avons choisies d'utiliser, sont les plus exhaustives (nr et nt notamment) mais elles contiennent plus d'erreurs que la base de données RefSeq par exemple. Enfin, à cause du manque de ressources génomiques pour *Lgh* dans les bases de données, nous avons fait le choix de garder tous les contigs correspondant à des plantes vertes (Viridiplantae), empêchant ainsi l'identification de contaminations provenant d'autres plantes. Certaines séquences du transcriptome de *Lgh* pourraient en fait être des séquences contaminantes, confortant l'idée d'améliorer l'étape de décontamination (Voir chapitre 3).

4 Intérêt des données multi-omiques

Les données "omiques" contribuent à la compréhension des processus cellulaires. Pour cela, différents types de données existent permettant des études à différentes échelles (Figure 4.1) [Hayes et al., 2024]. Dans ma thèse, nous avons à la fois généré et analysé des données génomiques et transcriptomiques. Au départ, l'assemblage du génome nucléaire de *Lgh* n'était pas forcément prévu car sa taille, sa polyploïdie ainsi qu'une profondeur de séquençage faible laissaient présager des difficultés pour son obtention. L'idée était d'assembler uniquement un transcriptome *de novo*. Un travail conséquent d'assemblage et de biocuration en utilisant différentes stratégies (Voir chapitre 2) a cependant permis

son assemblage. Par ailleurs, un transcriptome *de novo* a également été assemblé pour *Lgh*.

L'assemblage du génome de *Lgh* a permis l'obtention d'un transcriptome sur référence. Contrairement au transcriptome *de novo*, son obtention est basé sur le génome et sur les annotations de ce dernier. Les deux transcriptomes peuvent être complémentaires présentant chacun des avantages et inconvénients. Dans une étude menée sur des jeux de données de RNA-Seq chez l'Homme, il a été montré qu'un transcriptome sur référence était plus complets qu'un transcriptome *de novo* [Lee, Na et Park, 2021]. En effet, bien que 80% des gènes exprimés étaient retrouvés dans les deux méthodes, 17% des gènes exprimés étaient uniques au transcriptome sur référence. À l'inverse, une étude chez le chou (*Brassica oleracea*) a montré que l'assemblage *de novo* avait été plus efficace pour l'identification de gènes exprimés [Privitera et al., 2024]. Ce résultat s'explique par le manque d'annotations fonctionnelles robustes pour le génome de *Brassica oleracea*. Le génome de *Lgh* n'étant peut-être pas complet, l'utilisation des deux méthodes (*de novo* et sur référence) a permis d'obtenir à la fois un répertoire des gènes exprimés le plus complet en plus d'une quantification précise de l'expression des gènes [Lee, Na et Park, 2021].

L'utilisation de données transcriptomiques peut aussi permettre d'améliorer l'annotation du génome. Par exemple, chez le sorgo (*Sorghum bicolor*), la génération d'un atlas du transcriptome contenant l'expression de gènes dans les tissus majeurs ainsi que durant les phases juvénile, végétative et reproductive a contribué à l'amélioration du génome. Ainsi, le transcriptome a permis la découverte de nouveaux gènes (augmentation du nombre de gènes de 24%), l'amélioration des annotations existantes (meilleure identification des séquences des CDS) et de nouvelles caractérisations fonctionnelles [McCormick et al., 2018]. Par ailleurs, dans une autre étude, l'utilisation de données transcriptomiques chez 18 plantes ont fourni une description biologique supplémentaire

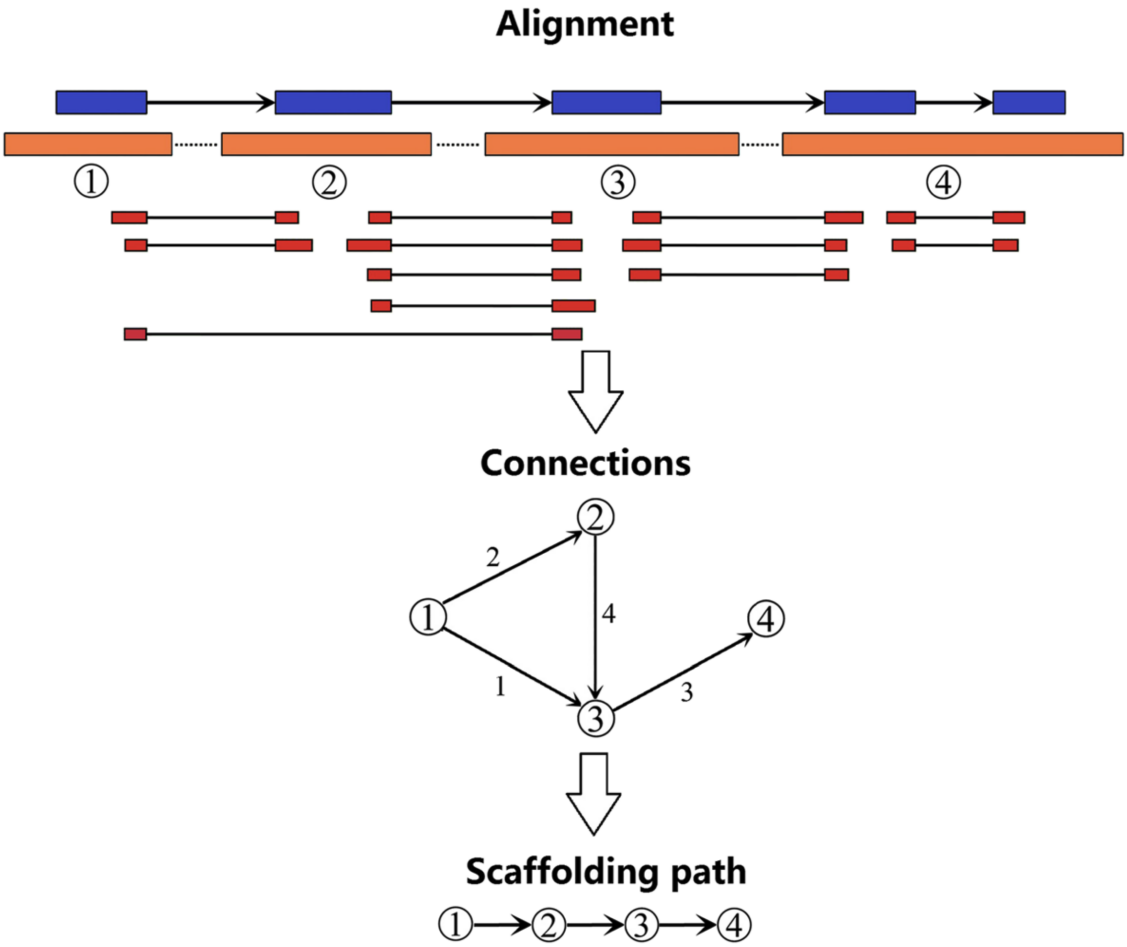


FIGURE 4.2 – Exemple de méthode de scaffolding avec des reads de RNA-seq avec le logiciel P_RNA_scaffolder [Zhu et al., 2018]

à 40% des gènes annotés dans leur génome dont 9,5% qui n'avaient pas de fonction connue [Sreedasyam et al., 2023].

Les données de RNA-seq peuvent également permettre de réduire la fragmentation du génome grâce à des méthodes de scaffolding particulière. En effet, les introns chez les plantes peuvent atteindre plus de 100 kb [Kress et al., 2022], il sera donc possible avec des reads de RNA-seq d'établir des liens entre des contigs (Figure 4.2) [Ghurye et Pop, 2019]. Cette méthodologie a permis l'amélioration du génome très fragmenté du panda géant (*Ailuropoda melanoleuca*) en réduisant le nombre de contigs de 200 593 à 197 567, en améliorant le N50 passant de 39 886 à 41 190 bp et en diminuant le nombre gaps de 119 126 à 116 170. Le scaffolding avec des données transcriptomiques a aussi été utilisé chez des plantes comme les deux espèces de ficus, *Ficus benghalensis* et *Ficus religiosa* [Chakraborty et al., 2022] et chez le groseillier de Ceylan (*Phyllanthus emblica*) [Mahajan et al., 2023].

L'assemblage du génome de *Lgh* pourra aussi servir comme base pour des analyses épigénétiques. En effet, des données sur le méthylome de *Lgh* sont disponibles (Voir perspectives). Un génome de référence est souvent nécessaire pour des analyses du méthylome [Nunn et al., 2021]. En effet, il servira de référence pour mapper les reads du méthylome et ainsi identifier les régions méthylés. Cependant, certaines méthodes existent permettant de détecter des variations de méthylations sans utilisation de génome de référence [Richards et al., 2017]. Mais dans ce cas, les variations détectées ne seront pas exhaustives et l'utilisation d'un génome de référence simplifiera les analyses et permettra une étude plus fine sur la régulation des gènes grâce aux annotations. Par ailleurs, il est possible de combiner des analyses transcriptomiques et du métylome ; comme cela a été fait chez deux espèces de coton, *Gossypium barbadense* et *Gossypium hirsutum* [You et al., 2022 ; Zhang, Guo et al., 2019]. Ces études ont permis de mettre en évidence l'effet de la méthylation sur l'expression des gènes et son implication dans

la stérilité mâle chez le coton.

5 Apports de la thèse

5.1 Apport pour la compréhension de l'acclimatation de *Lgh* au milieu terrestre

Les ressources génomiques produites au cours de ma thèse vont permettre des analyses à une échelle génomique sur la compréhension de l'acclimatation de *Lgh* au milieu terrestre. En effet, les morphotypes aquatique et terrestre et leurs réponses dans différents milieux avaient été caractérisées au niveau morphologique et métabolique [Haury, Druel et al., 2014 ; Billet et al., 2018 ; Genitoni, Vassaux, Renault et al., 2025]. L'étude du transcriptome a déjà permis de mettre en évidence des différences d'expression de gènes entre les morphotypes aquatique et terrestre de *Lgh*. Ainsi, les ressources produites durant ma thèse vont permettre des analyses nouvelles à des échelles génomiques, transcriptomiques et épigénomiques qui compléteront les connaissances déjà établies sur l'acclimatation de *Lgh* au milieu terrestre. En outre, l'expression différentielle de gènes entre les milieux terrestre et aquatique a été étudié chez de nombreuses plantes amphibies comme *Potamogeton wrightii*, *Rorippa aquatica* ou *Hygrophila difformis* [Koga, Ikematsu et Kimura, 2024 ; Han, Gao et al., 2015]. Nos données participeront à une meilleure compréhension du comportement des plantes amphibies et de leur capacité à changer de milieu.

5.2 Apport des ressources génomiques de *Lgh* pour les études phylogénétiques chez les Myrtales et les Onagraceae

Malgré l'importance écologique et économique des Myrtales et de nombreuses études phénotypiques (morphologiques, palynologiques, cytologiques et anatomiques), les relations de parenté entre les familles de cet ordre restent difficiles à cerner [Dahlgren et Thorne, 1984]. En 2021, un travail conséquent établissant les relations phylogénétiques chez les Myrtales a révélé que la lignée menant aux Myrtales se serait séparée des autres plantes il y a environ 104,90 millions d'années pendant le Crétacé inférieur, leur ancêtre commun avec les Geraniales étant daté de 111,59 Ma [Zhang, Landis et al., 2021]. Dans l'ordre des Myrtales, le groupe comprenant les Combretaceae aurait divergé avec les Onagraceae et les Lythraceae, il y a 96,22 Ma au Crétacé supérieur. Les Lythraceae et les Onagraceae se seraient, quant à eux, séparés il y a 89,59 Ma.

Les premières études moléculaires sur les relations phylogénétiques chez les Myrtales sont basées sur l'utilisation de marqueurs de type barcode chloroplastique ou de type nucléaire ayant pu conduire à des résultats incohérents. À titre d'exemple, des arbres phylogénétiques chez les *Ludwigia* ont été construits à partir de marqueurs nucléaires (ITS, waxy) et chloroplastiques (rps16, rpl16, trnL-trnF, trnL-CD, trnG) [Liu, Hoch et al., 2017]. Les types d'arbres révèlent des différences de structuration, soulignant l'intérêt de diversifier les types de marqueurs utilisés. Par ailleurs, une étude de phylogénie basée sur 33 plastomes d'Onagraceae a révélé que les Onagroideae ont connu une augmentation de la taille des plastomes liée à l'expansion de l'IR (Inverted repeat) alors que les Ludwigioideae, dont les *Ludwigia* représente le seul genre, présente un schéma évolutif distinct [Ou et al., 2025]. Les données génomiques, en particulier les gènes nucléaires à copie unique ou quasi-unique, fournissent également des informations précieuses pour résoudre les relations phylogénétiques. Le séquençage et l'assemblage des génomes de deux espèces représentatives de la tribu des Epilobieae, *Chamaenerion*

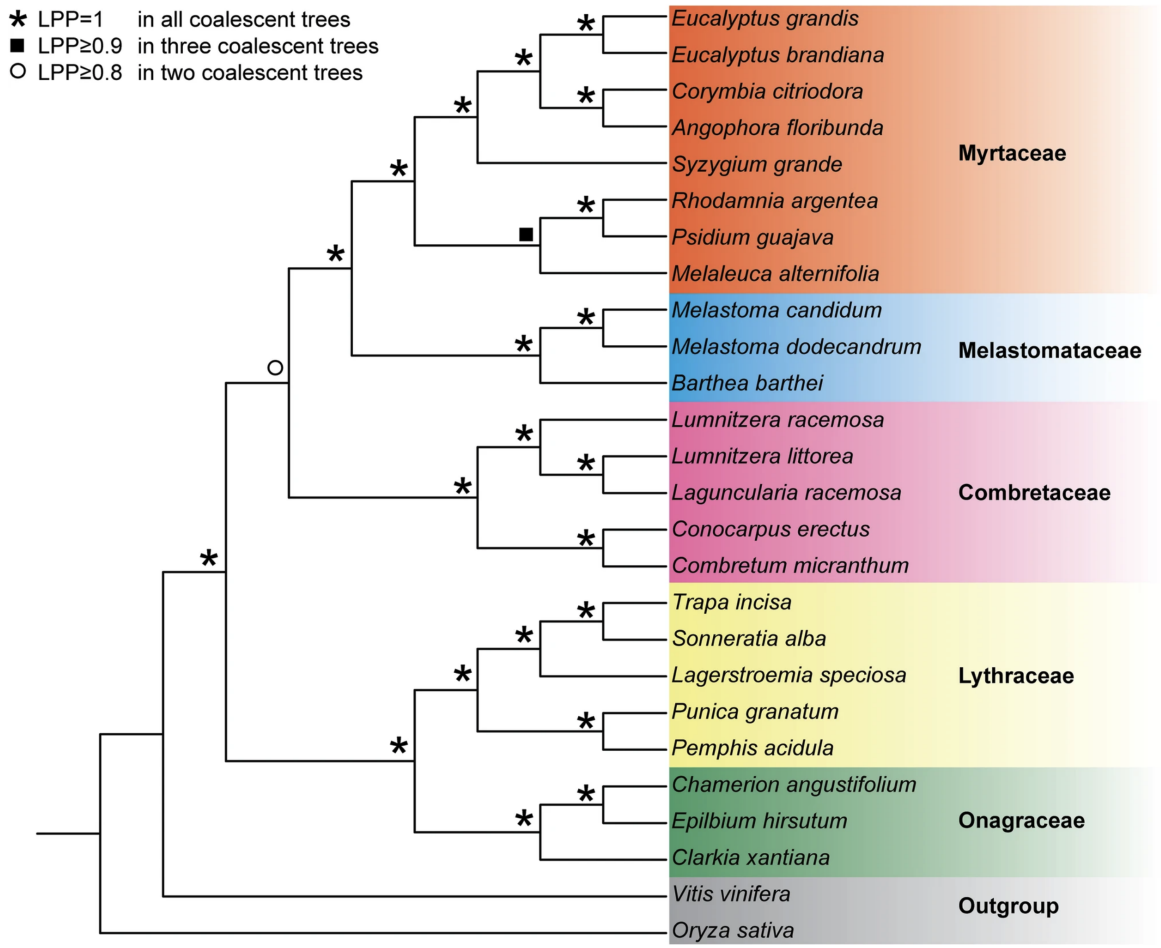


FIGURE 4.3 – Arbre phylogénétique de l'ordre des Myrtales issu de 4 arbres coalescents utilisant des gènes provenant de 994, 939, 706 et 339 orthogroupes [Wang, Li, Yang et al., 2025].

angustifolium et *Epilobium hirsutum*, ont permis de révéler deux événements de duplication génomique, au cours desquels le nombre de chromosomes est passé de 9 à 18, suivis d'une réduction aneuploïde, conduisant à la diversité du nombre de chromosomes observée au sein de cette tribu [Wang, Li, Yang et al., 2025]. Ainsi, la polyploïdie a joué un rôle important dans l'évolution et la diversification des Epilobieae. Ces auteurs ont également confirmé que les Combretaceae, dont le positionnement phylogénétique questionne toujours, sont groupes frères des Myrtaceae et des Melastomataceae, apportant ainsi de nouvelles informations sur les relations évolutives au sein des Myrtales (Figure 4.3). Ces auteurs soulignent l'intérêt d'intégrer les données génomiques pour ce type d'étude, fournissant ainsi des ressources supplémentaires riches pour la recherche phylogénétique, qui joueront un rôle important dans la poursuite des études sur les Myrtales.

Ces résultats soulignent l'importance de développer des ressources génomiques diversifiées afin de mieux comprendre l'histoire évolutive des espèces, le rôle de la polyploïdie dans l'évolution et la diversification des espèces. Ainsi, les ressources génomiques (nucléaire, plastome, mitochondrie) et transcriptomique générées durant cette thèse s'avèrent constituer un set de ressources unique qui pourra servir à améliorer la connaissance sur l'évolution des espèces au sein des Myrtales et plus particulièrement au sein des Onagraceae.

5.3 Apport dans la méthodologie d'assemblage de ressources génomiques complexes

La production de données génomiques pour *Lgh* a été difficile. Cependant, les méthodologies utilisées durant ma thèse pourraient servir pour d'autres espèces non-modèles, complexes (taille du génome importante, polyploïde, etc...) et avec des tissus riches en polyphénols et tanins rendant difficile les extractions nucléotidiques.

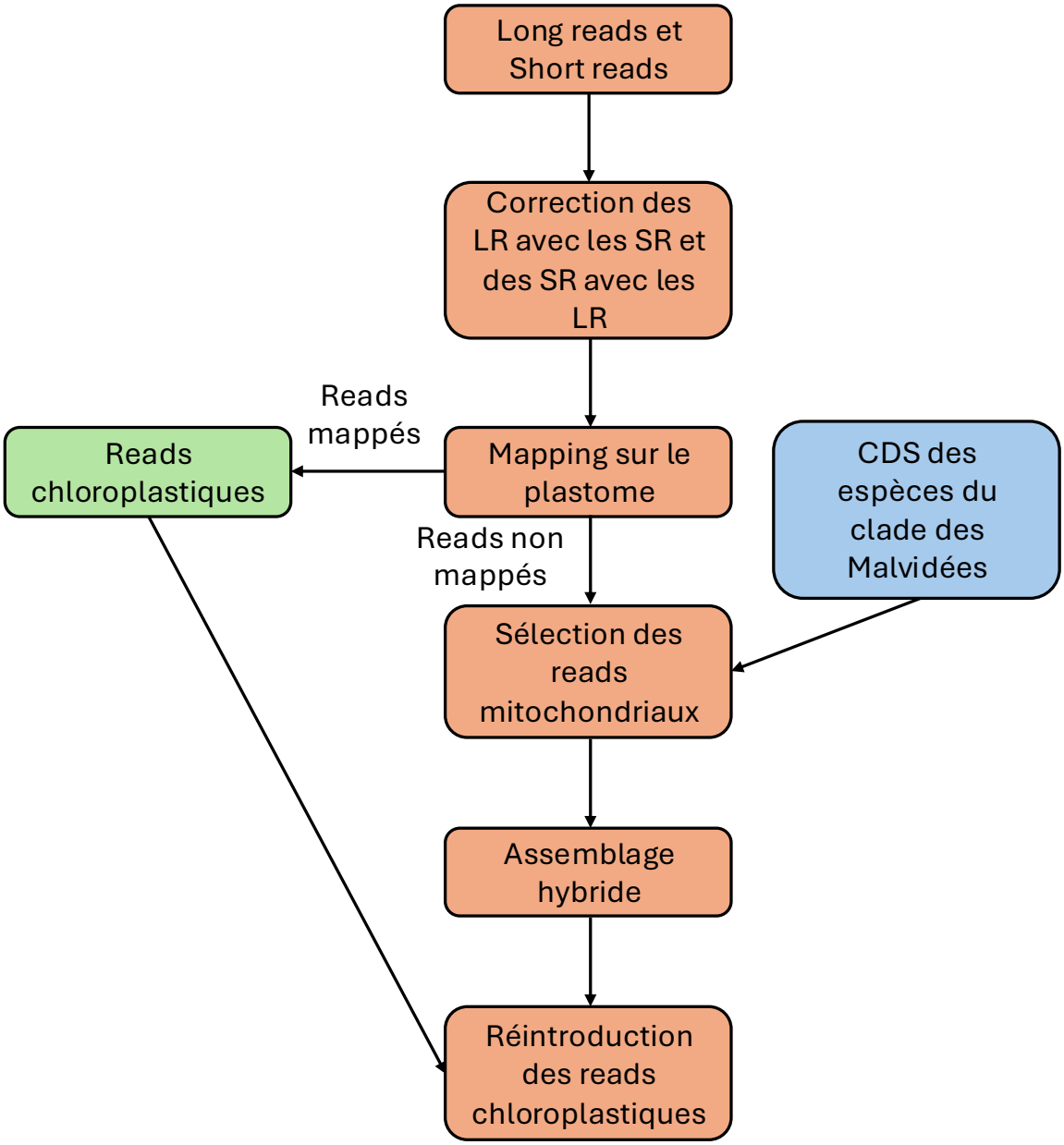


FIGURE 4.4 – Pipeline simplifié de l'assemblage du mitogénome de *Lgh*.

Méthodologie pour l'assemblage d'un mitogénome de plante

La stratégie que nous recommandons pour l'assemblage de mitogénome de plante est présentée en Figure 4.4. Comme présenté précédemment, les génomes mitochondriaux de plantes sont complexes ; ceci est notamment dû à leur taille assez importante, à leur grand nombre de répétitions et au peu de conservation entre espèces même proches phylogénétiquement [Wu et al., 2020 ; Wynn, 2019]. Dans notre cas, la première étape a été la correction des reads. Nous avons utilisé des corrections hybrides utilisant les LR pour corriger les SR et inversement. Ensuite, nous avons aussi décidé de retirer les reads appartenant au plastome. Dans notre cas, cette stratégie s'est avérée utile au départ pour simplifier le jeu de données. Cependant, étant donné la présence d'insertions du chloroplaste dans les mitogénomes de plantes [Wang, Wu et al., 2007], la réintroduction des séquences du plastome a été indispensable pour compléter l'assemblage. Le fait de les introduire tard dans le processus permet de réduire la présence d'erreurs qui pourraient faussement associer des séquences du plastome et du mitogénome. La stratégie choisie pour sélectionner les reads mitochondriaux a été l'utilisation comme référence des séquences des gènes codant des protéines (CDS) des espèces les plus proches phylogénétiquement de *Lgh* (espèces du clade des Malvidées). Cette approche a été utilisée chez deux espèces du genre *Zantedeschia* [Guo, Li et al., 2023] et chez le groseillier des Barbades (*Pereskia aculeata*) [Zhang, Shan et al., 2023]. Enfin, la stratégie d'assemblage combinant long reads et short reads avec plusieurs assembleurs a été gagnante. Cette stratégie est également recommandée dans la littérature [Štorchová et Krüger, 2024]. Notre pipeline d'assemblage a été efficace pour l'assemblage du mitogénome de *Lgh* et pourrait servir pour de futurs assemblages d'autres mitogénomes de plantes.

Méthodologie pour l'assemblage d'un génome de plante

L'assemblage du génome de *Lgh* a été très compliqué notamment à cause de sa taille importante (1,487 Gb) et de sa polyploïdie. De plus, les séquençages réalisés avaient peu de profondeur. Ainsi, notre stratégie s'adapte à ce cas particulier même s'il serait plutôt recommandé d'avoir plus de profondeur et éventuellement d'effectuer d'autres séquençages plus adaptés (Voir discussion sur les stratégies de séquençage). Un pipeline ne sera pas proposé ici mais les points clés de l'assemblage du génome seront rappelés et discutés. Comme pour le mitogénome, les long reads et les short reads ont été corrigés en utilisant une méthode hybride, comme conseillé dans la littérature [Kersey, 2019]. Dans un premier temps, les reads mappés sur les génomes des organelles ont été retirés de l'assemblage. Ensuite, la stratégie a surtout été axée sur les longs reads avec l'utilisation de trois assembleurs LR différents et un assembleur hybride. Nous avons fait ce choix car les LR permettent une meilleure gestion des régions répétées [Li, Chen et al., 2012] qui sont très présentes dans les génomes de plantes [Jiao et Schneeberger, 2017]. Un gros travail de biocuration manuelle a été nécessaire pour l'assemblage, notamment pour enlever les ADNs contaminants (micro-organismes, Homme, ..). Il est surprenant qu'aucun ou peu de papiers ne souligne l'intérêt de la décontamination des génomes ce qui peut conduire à l'accumulation d'erreurs dans les bases de données. Les séquences des génomes des organelles ont été réintroduites plus tard comme pour le mitogénome. Cette réintroduction est nécessaire à cause de la grande proportion de ces séquences dans les génomes de plantes [Zhang, Dong et al., 2020]. Les différentes stratégies utilisées pour l'assemblage du génome de *Lgh* pourront servir dans le cadre d'assemblage de génome polyploïde de plantes.

Méthodologie pour l'assemblage d'un transcriptome

Le pipeline de l'assemblage du transcriptome de *Lgh* a déjà été présenté (Voir Matériels et méthodes du chapitre 3). Nous avons utilisé plusieurs assembleurs et plusieurs tailles de kmers ce qui est conseillé dans la littérature [Behera, Voshall et Moriyama, 2021]. Le mapping des transcrits assemblé *de novo* sur le génome est une stratégie de décontamination qui n'est pas utilisé en général. Pourtant, elle s'est avérée pertinente même si aucun mappeur spécifique à cette tâche ne semble exister pour le moment. L'utilisation de minimap2 a pu montrer ses limites sur certains transcrits notamment courts et il n'est pas exclu que certains transcrits contaminants puissent avoir été mappés. Malgré tout, il s'agit, d'après notre benchmarking, du meilleur outil pour ce type de mapping. L'étape de décontamination suivante devra cependant être retravaillé comme discuté précédemment.

Ainsi, les assemblages du mitogénome, du génome et du transcriptome de *Lgh* peuvent avoir de multiples apports pour de futures recherches sur la jussie, sur la phylogénie au sein de l'ordre des Myrtales et pour les méthodologies d'assemblage de ressources génomiques des espèces complexes. La production de ces ressources génomique au cours de ma thèse ouvre également des perspectives sur l'étude de *Lgh*.

6 Perspectives

6.1 Analyses différentielles du transcriptome plus poussées

Les analyses différentielles sur l'expression des gènes impliqués dans l'acclimatation de *Lgh* au milieu terrestre entre les différents morphotypes conditionnés dans différents milieux n'ont pas pu être finalisées par manque de temps. Grâce aux ressources génomiques maintenant disponibles, il sera possible de continuer ces analyses. La compa-

raison entre le morphotype terrestre conditionné en terrestre (MTT) et le morphotype aquatique conditionné en terrestre (MAT) et en alternance (MA-AT) devra être approfondie. Ensuite, une comparaison pourra être faite entre les morphotypes aquatique et terrestre conditionnés dans leur milieu d'origine (MAA et MTT). Enfin, il sera aussi possible d'évaluer l'effet d'un conditionnement prolongé en milieu aquatique du morphotype terrestre (MTA) en le comparant avec le morphotype terrestre conditionné en terrestre (MTT) et le morphotype aquatique conditionné en aquatique (MAA). Ces comparaisons pourront être effectuées avec le transcriptome *de novo* et sur référence mais pour ça il sera d'abord nécessaire d'améliorer ces ressources génomiques.

6.2 Amélioration des ressources génomiques

Les ressources assemblées actuellement ne sont pas complètes et contiennent encore des erreurs, ce qui est fréquent pour toute version initiale d'un génome/transcriptome de plante. Il est possible d'améliorer la qualité des assemblages du transcriptome et du génome avec les données déjà à disposition ou en, en produisant des complémentaires. Le transcriptome *de novo* de *Lgh* contient quelques erreurs d'assemblage (Voir chapitre 3). Grâce au génome de *Lgh*, il sera possible de réduire la redondance du transcriptome et de corriger ou retirer certains assemblages chimériques. Des séquences d'ADN présentes dans les transcrits devront notamment être retirées avec l'aide du génome. L'étape de décontamination pourra aussi être amélioré pour éviter l'inclusion de séquences contaminantes dans le transcriptome de *Lgh*.

À l'inverse, le transcriptome ou les reads RNA-Seq pourront améliorer le génome de *Lgh*. Comme discuté précédemment (Partie 4 données multi-omiques), l'utilisation d'un jeu de données RNA-Seq peut réduire la fragmentation du génome grâce à des outils spécialisés [Ghurye et Pop, 2019]. Il sera aussi possible d'annoter de nouveaux gènes ainsi que d'améliorer les annotations existantes en identifiant mieux les CDS et en ajoutant

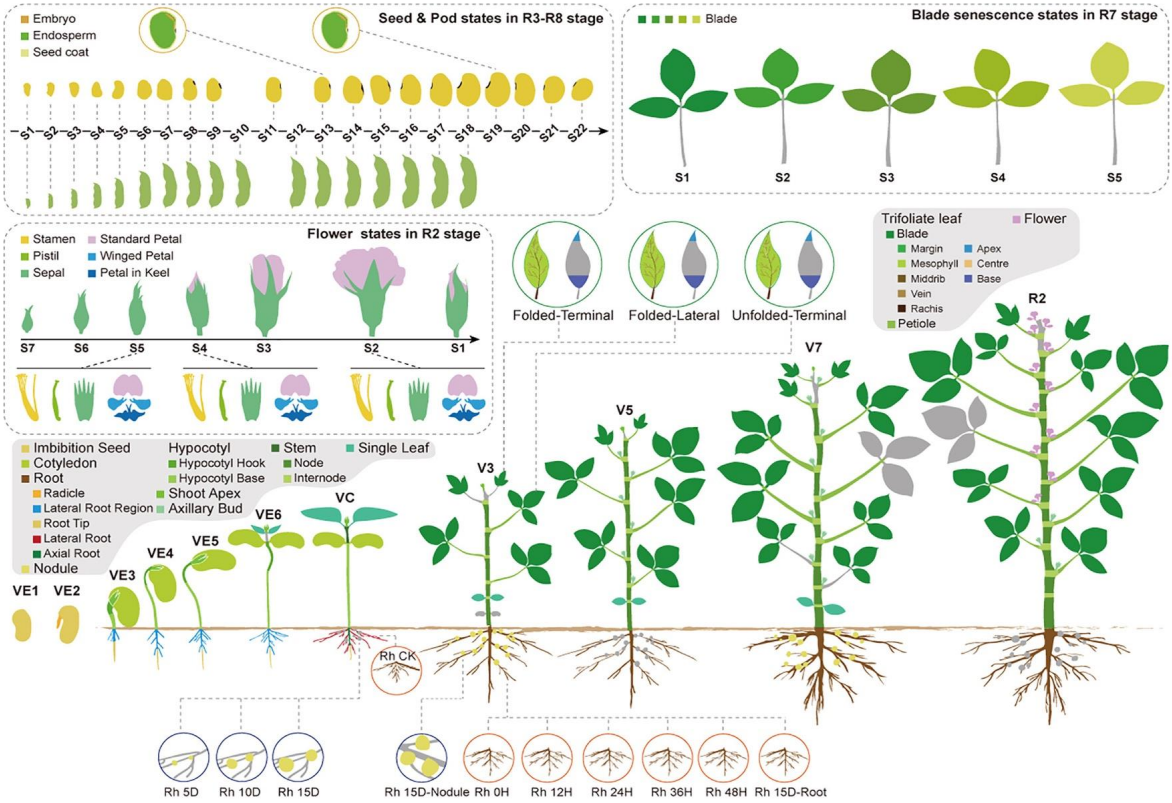


FIGURE 4.5 – Représentation graphique de 314 échantillons de RNA-Seq repartis sur les phases de développement du soja [Fan et al., 2025]. VE : Stade d'émergence végétative ; VC = Stade de cotylédon végétatif ; V : Stade végétatif ; R : Stade reproductif ; Rh : Inoculation de *Rhizobium* ; H : Heure ; D : Jour.

des validations expérimentales à certaines annotations. Cependant, le transcriptome que nous avons produit dans notre étude concerne les transcrits exprimés dans les racines, les tiges et les feuilles sous contrainte hydrique. C'est le cas de la majorité des études produisant un transcriptome *de novo* chez les plantes, et qui porte sur des transcrits produits dans des organes spécifiques (racine, tige, feuille) et en réponse par exemple à des contraintes environnementales particulières comme un stress hydrique [Ribeyre et al., 2025] ou un stress thermique [Sun, Feng et al., 2022]. Une version 2 du génome de *Lgh* sera réalisée.

L'obtention d'un transcriptome le plus exhaustif possible chez *Lgh* nécessiterait de conduire une expérimentation impliquant des prélèvements d'organes variés. Ainsi chez le soja, un atlas transcriptomique intégrant cinq organes (racine, nodule, apex de la pousse, feuille et tige) et 314 échantillons répartis sur tout le cycle de vie du soja a été produit afin de faciliter la caractérisation des profils d'expression spatio-temporelle de la plupart des gènes liés au développement des organes (Figure 4.5) [Fan et al., 2025]. Bien que la construction d'atlas transcriptomiques représente un coût élevé et un challenge technologique important, plusieurs auteurs soulignent l'intérêt de développer des transcriptomes intégrés complets chez différentes espèces afin de fournir des ressources génomiques précieuses pour faciliter les études fonctionnelles [Rich-Griffin et al., 2020 ; Seyfferth et al., 2021 ; Ribeyre et al., 2025]. La génération d'un tel jeu de données pourrait donc être envisagé pour *Lgh* afin d'obtenir un set de transcrits le plus exhaustif possible qui permettrait l'annotation de la plupart des gènes du génome.

Pour améliorer l'assemblage du génome de *Lgh*, de nouveaux séquençages pourraient être envisagés. Si le séquençage de *Lgh* devait être refait aujourd'hui, la combinaison de stratégies de séquençage LR (HiFi), SR et Hi-C serait la stratégie la plus efficace [Kyriakidou et al., 2018 ; Jeon et Kim, 2024]. De plus, le séquençage d'un ou plusieurs de ces progéniteurs pourraient également aider [Kyriakidou et al., 2018]. *Lpm*, qui est

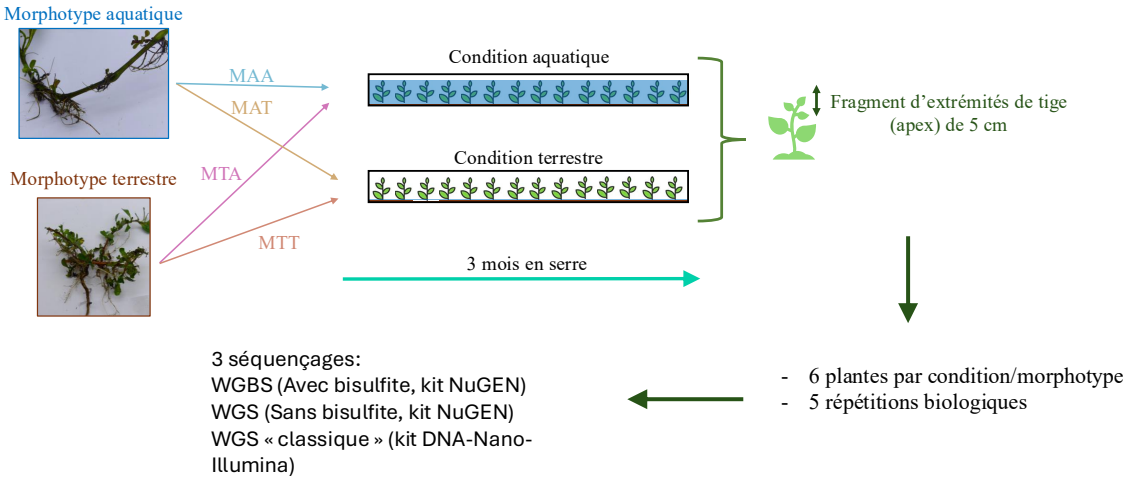


FIGURE 4.6 – Dispositif expérimental pour le séquençage du méthylome de *Lgh* .

diploïde et participe à hauteur de 4x au génome de *Lgh*, avait été séquencé dans ce but. Cependant, seuls des LR avaient été obtenus et avec très peu de profondeur (1,9x). Des séquençages LR, SR et Hi-C avec plus de profondeur de *Lpm* auraient permis d'aider pour l'assemblage du génome de *Lgh*. Par ailleurs, la description récente de la phylogénie au sein du genre *Ludwigia* a permis d'identifier certains autres progéniteurs de *Lgh* comme *L. helminthorrhiza* (diploïde, participe à hauteur de 2x), *L. stolonifera* (tétraploïde, participe à hauteur de 4x = *Lpm* + *L. helminthorrhiza*) ou *L. grandiflora* subsp. *grandiflora* (hexaploïde, participe à hauteur de 6x = *Lpm* + 4x d'espèce(s) inconnue(s)) [Barloy et al., 2024]. Le séquençage et l'assemblage des génomes de ces espèces permettraient d'améliorer l'assemblage du génome de *Lgh*, de disposer des sous-génomes de *Lgh* et de comprendre l'évolution des génomes suite à la polyploïdisation comme cela a été réalisée par exemple avec *Camelina neglecta*, espèce progénitrice diploïde de *Camelina sativa* (espèce hexaploide) [Chaudhary et al., 2023 ; Wang, Blume et al., 2024]. Une étude pour identifier la ou les espèce(s) progénitrices de *Lgh* manquantes pourrait être envisagée afin de séquencer chacun de ses progéniteurs et permettre l'assemblage de chacun des sous-génomes de *Lgh*.

6.3 Méthylome

Des données de séquençage du méthylome ont été obtenus au cours de la thèse de Julien Genitoni en 2019 (Figure 4.6). Les deux morphotypes aquatique et terrestre ont été placés en serre pendant trois mois dans leurs conditions d'origine ou dans des conditions croisées. Les 5 premiers centimètres à partir de l'apex de 6 plantes par morphotype/condition ont été prélevés 5 fois afin d'obtenir des réplicats biologiques. La chromatine a été extraite de ces échantillons et digérée par la DNAse I. Puis, l'euchromatine a été extraite car elle est potentiellement riche en séquences géniques. Enfin, trois séquençages ont été effectués pour chaque échantillon : un WGBS (séquençage

bisulfite, kit NuGEN), un WGS (sans bisulfite, kit NuGEN) et un WGS "classique" (kit DNA-Nano-Illumina). Le séquençage avec bisulfite comparé au séquençage sans bisulfite va permettre d'identifier les régions méthylées. Le génome de *Lgh* sera utilisé comme référence pour le mapping de ces séquences. Des différences de méthylations entre les morphotypes et peut-être même entre les conditionnements sont attendues. En effet, une hypométhylation artificielle de l'ADN de *Lgh* a affecté différemment les morphotypes aquatique et terrestre suggérant un rôle important de l'épigénétique dans la transition de *Lgh* du milieu aquatique au milieu terrestre [Genitoni, Vassaux, Delaunay et al., 2020].

6.4 Comparaison entre populations natives et invasives

Une étude plus globale de l'invasion de *Lgh* pourrait aussi être envisagée. L'analyse de différentes populations natives et invasives à travers le monde permettrait une meilleure compréhension des mécanismes de l'invasion et également d'évaluer sa capacité à s'adapter localement. Le génome de *Lgh* aiderait à réaliser ce genre d'analyse. En effet, l'assemblage du génome de la spartine à feuilles alternes (*Spartina alterniflora*) a permis d'identifier des trajectoires évolutives distinctes entre différentes populations natives et invasives [Hao et al., 2024]. Des preuves génomiques d'adaptation rapide de cette plante ont été découvertes, démontrant un dynamisme du génome important pour les populations introduites.

Chez l'ambroisie à feuilles d'armoise (*Ambrosia artemisiifolia*), des variations structurelles du génome ont été identifiées entre populations natives et invasives [Battlay, Craig et al., 2025]. De plus, des adaptations parallèles ont pu être mises en évidence entre différentes populations invasives grâce à l'étude de leur génome. Ainsi, le séquençage de plusieurs populations de *Lgh* collectées dans deux zones envahies comme en France et aux États-Unis et dans son aire native comme au Brésil pourrait améliorer

les connaissances des processus évolutif et d'adaptation lors des multiples invasions de *Lgh*.

Bibliographie

A. Putintseva, Yuliya et al. [2020]. « Siberian larch (Larix sibirica Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome ». In : Accepted : 2021-02-15T12 :29 :13Z Publisher : BMC. ISSN : 14712164. DOI : 10.1186/s12864-020-07061-4. URL : https://elib.sfu-kras.ru/handle/2311/137952 [visité le 08/03/2022].

Achakkagari, S. et al. [2025]. « Better together : Subgenomes for allotetraploid potato wild relative Solanum acaule Bitt. reveal origins in Petota Clade 3 and 4 ». In : *The Plant Genome* 18.3. _eprint : https ://acsess.onlinelibrary.wiley.com/doi/pdf/10.1002/tpg2.70095, e70095. ISSN : 1940-3372. DOI : 10.1002/tpg2.70095. URL : https://onlinelibrary.wiley.com/doi/abs/10.1002/tpg2.70095 [visité le 24/09/2025].

Adewale, Boluwatife A. [26 nov. 2020]. « Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years ? » In : *African Journal of Laboratory Medicine* 9.1, p. 1340. ISSN : 2225-2002. DOI : 10.4102/ajlm.v9i1.1340. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7736650/ [visité le 19/05/2025].

Almagro, L. et al. [2009]. « Class III peroxidases in plant defence reactions ». In : *Journal of Experimental Botany* 60.2, p. 377-390. ISSN : 1460-2431. DOI : 10.1093/jxb/ern277.

Altschul, Stephen F. et al. [5 oct. 1990]. « Basic local alignment search tool ». In : *Journal of Molecular Biology* 215.3, p. 403-410. ISSN : 0022-2836. DOI : 10.1016/S0022-2836(05)80360-2. URL : https://www.sciencedirect.com/science/article/pii/S0022283605803602 [visité le 23/02/2023].

Alverson, Andrew J., Danny W. Rice et al. [juill. 2011]. « Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber ». In : *The Plant Cell* 23.7, p. 2499-2513. ISSN : 1532-298X. DOI : 10.1105/tpc.111.087189.

Alverson, Andrew J., Shi Zhuo et al. [20 jan. 2011]. « The Mitochondrial Genome of the Legume Vigna radiata and the Analysis of Recombination across Short Mitochondrial Repeats ». In : *PLoS ONE* 6.1. Sous la dir. d'Orian S. Shirihai, e16404. ISSN : 1932-6203. DOI : 10.1371/journal.pone.0016404. URL : https://dx.plos.org/10.1371/journal.pone.0016404 [visité le 16/01/2023].

Anderson, Charles T. et Jérôme Pelloux [20 mai 2025]. « The Dynamics, Degradation, and Afterlives of Pectins : Influences on Cell Wall Assembly and Structure, Plant Development and Physiology, Agronomy, and Biotechnology ». In : *Annual Review of Plant Biology* 76 [Volume 76, 2025]. Publisher : Annual Reviews, p. 85-113. ISSN : 1543-5008, 1545-2123. DOI : 10.1146/annurev-arplant-083023-034055. URL : https://www.annualreviews.org/content/journals/10.1146/annurev-arplant-083023-034055 [visité le 09/10/2025].

Antipov, Dmitry et al. [1er avr. 2016]. « hybridSPAdes : an algorithm for hybrid assembly of short and long reads ». In : *Bioinformatics* 32.7, p. 1009-1015. ISSN : 1367-4803. DOI : 10.1093/bioinformatics/btv688. URL : https://doi.org/10.1093/bioinformatics/btv688 [visité le 20/09/2025].

Ariel, Federico et al. [1er juin 2015]. « Battles and hijacks : noncoding transcription in plants ». In : *Trends in Plant Science* 20.6. Publisher : Elsevier, p. 362-371. ISSN :

1360-1385. DOI : 10.1016/j.tplants.2015.03.003. URL : https://www.cell.com/trends/plant-science/abstract/S1360-1385(15)00055-2 [visité le 04/06/2025].

Arnon, Daniel I. [nov. 1971]. « The Light Reactions of Photosynthesis ». In : *Proceedings of the National Academy of Sciences of the United States of America* 68.11, p. 2883-2892. ISSN : 0027-8424. DOI : 10.1073/pnas.68.11.2883. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC389550/ [visité le 17/09/2025].

Arshad, Rameez et al. [27 juin 2022]. « A kaleidoscope of photosynthetic antenna proteins and their emerging roles ». In : *Plant Physiology* 189.3, p. 1204-1219. ISSN : 1532-2548. DOI : 10.1093/plphys/kiac175.

Astashyn, Alexander et al. [6 juin 2023]. *Rapid and sensitive detection of genome contamination at scale with FCS-GX*. Pages : 2023.06.02.543519 Section : New Results. DOI : 10.1101/2023.06.02.543519. URL : https://www.biorxiv.org/content/10.1101/2023.06.02.543519v1 [visité le 29/04/2025].

Austin, Jotham R. et al. [juill. 2006]. « Plastoglobules are lipoprotein subcompartments of the chloroplast that are permanently coupled to thylakoid membranes and contain biosynthetic enzymes ». In : *The Plant Cell* 18.7, p. 1693-1703. ISSN : 1040-4651. DOI : 10.1105/tpc.105.039859.

Ayling, Martin, Matthew D Clark et Richard M Leggett [23 mars 2020]. « New approaches for metagenome assembly with short reads ». In : *Briefings in Bioinformatics* 21.2, p. 584-594. ISSN : 1477-4054. DOI : 10.1093/bib/bbz020. URL : https://doi.org/10.1093/bib/bbz020 [visité le 20/09/2025].

Bajguz, Andrzej et Alicja Piotrowska-Niczyporuk [25 juill. 2023]. « Biosynthetic Pathways of Hormones in Plants ». In : *Metabolites* 13.8, p. 884. ISSN : 2218-1989. DOI : 10.3390/metabo13080884. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10456939/ [visité le 17/09/2025].

Baker, H. G. [1955]. « Self-Compatibility and Establishment After 'Long-Distance' Dispersal ». In : *Evolution* 9.3. Publisher : [Society for the Study of Evolution, Wiley], p. 347-349. ISSN : 0014-3820. DOI : 10.2307/2405656. URL : https://www.jstor.org/stable/2405656 [visité le 26/09/2025].

Barloy, Dominique et al. [2024]. « Genomic relationships among diploid and polyploid species of the genus *Ludwigia* L. section *Jussiaea* using a combination of molecular cytogenetic, morphological, and crossing investigations ». In : *Peer Community Journal* 4. ISSN : 2804-3871. DOI : 10.24072/pcjournal.364. URL : https://peercommunityjournal.org/articles/10.24072/pcjournal.364/ [visité le 23/02/2024].

Barloy-Hubler, Frédérique et al. [2025]. « Sequencing, de novo assembly of *Ludwigia* plastomes, and comparative analysis within the Onagraceae family ». In : *Peer Community Journal* 5. ISSN : 2804-3871. DOI : 10.24072/pcjournal.536. URL : https://peercommunityjournal.org/articles/10.24072/pcjournal.536/ [visité le 28/04/2025].

Bartels, Arthur et al. [juill. 2018]. « Dynamic DNA Methylation in Plant Growth and Development ». In : *International Journal of Molecular Sciences* 19.7. Number : 7 Publisher : Multidisciplinary Digital Publishing Institute, p. 2144. ISSN : 1422-0067. DOI : 10.3390/ijms19072144. URL : https://www.mdpi.com/1422-0067/19/7/2144 [visité le 26/02/2024].

Battaglia, Marina et al. [sept. 2008]. « The enigmatic LEA proteins and other hydrophilins ». In : *Plant Physiology* 148.1, p. 6-24. ISSN : 0032-0889. DOI : 10.1104/pp.108.120725.

Battlay, Paul, Samuel Craig et al. [1er jan. 2025]. « Rapid Parallel Adaptation in Distinct Invasions of Ambrosia Artemisiifolia Is Driven by Large-Effect Structural Variants ». In : *Molecular Biology and Evolution* 42.1, msae270. ISSN : 1537-1719. DOI : 10.

1093/molbev/msae270. URL : https://doi.org/10.1093/molbev/msae270 [visité le 26/09/2025].

Battlay, Paul, Jonathan Wilson et al. [27 mars 2023]. « Large haploblocks underlie rapid adaptation in the invasive weed Ambrosia artemisiifolia ». In : *Nature Communications* 14.1. Publisher : Nature Publishing Group, p. 1717. ISSN : 2041-1723. DOI : 10.1038/s41467-023-37303-4. URL : https://www.nature.com/articles/s41467-023-37303-4 [visité le 26/09/2025].

Beest, Mariska te et al. [jan. 2012]. « The more the better ? The role of polyploidy in facilitating plant invasions ». In : *Annals of Botany* 109.1, p. 19-45. ISSN : 0305-7364. DOI : 10.1093/aob/mcr277. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC3241594/ [visité le 26/09/2025].

Behera, Sairam, Adam Voshall et Etsuko Moriyama [20 mars 2021]. « Plant Transcriptome Assembly : Review and Benchmarking ». In : *Exon Publications*, p. 109-130. DOI : 10.36255/exonpublications.bioinformatics.2021.ch7. URL : https://exonpublications.com/index.php/exon/article/view/280 [visité le 26/02/2024].

Beltran-Sanz, Núria et al. [10 sept. 2025]. « Chromosome-level assembly of the 400-year-old Goethe's Palm (Chamaerops humilis L.) » In : *Scientific Data* 12.1. Publisher : Nature Publishing Group, p. 1542. ISSN : 2052-4463. DOI : 10.1038/s41597-025-05673-7. URL : https://www.nature.com/articles/s41597-025-05673-7 [visité le 24/09/2025].

Belyayev, Alexander et al. [26 sept. 2023]. « Telomere sequence variability in genotypes from natural plant populations : unusual block-organized double-monomer terminal telomeric arrays ». In : *BMC Genomics* 24.1, p. 572. ISSN : 1471-2164. DOI : 10.1186/s12864-023-09657-y. URL : https://doi.org/10.1186/s12864-023-09657-y [visité le 18/09/2025].

Benson, Gary [1er jan. 1999]. « Tandem repeats finder : a program to analyze DNA sequences ». In : *Nucleic Acids Research* 27.2, p. 573-580. ISSN : 0305-1048. DOI : 10.1093/nar/27.2.573. URL : https://doi.org/10.1093/nar/27.2.573 [visité le 26/02/2024].

Bewick, Adam J et Robert J Schmitz [1er avr. 2017]. « Gene body DNA methylation in plants ». In : *Current Opinion in Plant Biology*. 36 Genome studies and molecular genetics 36, p. 103-110. ISSN : 1369-5266. DOI : 10.1016/j.pbi.2016.12.007. URL : https://www.sciencedirect.com/science/article/pii/S1369526616301297 [visité le 29/07/2025].

Billet, Kevin et al. [2018]. « Aquatic and terrestrial morphotypes of the aquatic invasive plant, Ludwigia grandiflora, show distinct morphological and metabolomic responses ». In : *Ecology and Evolution* 8.5. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/ec p. 2568-2579. ISSN : 2045-7758. DOI : 10.1002/ece3.3848. URL : https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.3848 [visité le 23/02/2024].

Binder, Stefan [2010]. « Branched-Chain Amino Acid Metabolism in Arabidopsis thaliana ». In : *The Arabidopsis Book* 8, e0137. ISSN : 1543-8120. DOI : 10.1199/tab.0137.

Blackburn, Tim M, Céline Bellard et Anthony Ricciardi [2019]. « Alien versus native species as drivers of recent extinctions ». In : *Frontiers in Ecology and the Environment* 17.4. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/fee.2020, p. 203-207. ISSN : 1540-9309. DOI : 10.1002/fee.2020. URL : https://onlinelibrary.wiley.com/doi/abs/10.1002/fee.2020 [visité le 23/02/2024].

Bolger, Marie E, Borjana Arsova et Björn Usadel [1er mai 2018]. « Plant genome and transcriptome annotations : from misconceptions to simple solutions ». In : *Briefings in Bioinformatics* 19.3, p. 437-449. ISSN : 1477-4054. DOI : 10.1093/bib/bbw135. URL : https://doi.org/10.1093/bib/bbw135 [visité le 29/07/2025].

Bozzi, Aaron T. et Rachelle Gaudet [6 août 2021]. « Molecular Mechanism of Nramp-Family Transition Metal Transport ». In : *Journal of Molecular Biology* 433.16, p. 166991. ISSN : 1089-8638. DOI : 10.1016/j.jmb.2021.166991.

Bray, Nicolas L. et al. [mai 2016]. « Near-optimal probabilistic RNA-seq quantification ». In : *Nature Biotechnology* 34.5. Publisher : Nature Publishing Group, p. 525-527. ISSN : 1546-1696. DOI : 10.1038/nbt.3519. URL : https://www.nature.com/articles/nbt.3519 [visité le 05/08/2025].

Brown, David et al. [mai 2011]. « Arabidopsis genes IRREGULAR XYLEM (IRX15) and IRX15L encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall ». In : *The Plant Journal : For Cell and Molecular Biology* 66.3, p. 401-413. ISSN : 1365-313X. DOI : 10.1111/j.1365-313X.2011.04501.x.

Bryant, Donald M. et al. [17 jan. 2017]. « A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors ». In : *Cell Reports* 18.3. Publisher : Elsevier, p. 762-776. ISSN : 2211-1247. DOI : 10.1016/j.celrep.2016.12.063. URL : https://www.cell.com/cell-reports/abstract/S2211-1247(16)31770-3 [visité le 03/12/2024].

Buchfink, Benjamin, Klaus Reuter et Hajk-Georg Drost [avr. 2021]. « Sensitive protein alignments at tree-of-life scale using DIAMOND ». In : *Nature Methods* 18.4. Publisher : Nature Publishing Group, p. 366-368. ISSN : 1548-7105. DOI : 10.1038/s41592-021-01101-x. URL : https://www.nature.com/articles/s41592-021-01101-x [visité le 29/04/2025].

Bushmanova, Elena et al. [1er sept. 2019]. « rnaSPAdes : a de novo transcriptome assembler and its application to RNA-Seq data ». In : *GigaScience* 8.9, giz100. ISSN : 2047-217X. DOI : 10.1093/gigascience/giz100. URL : https://doi.org/10.1093/gigascience/giz100 [visité le 12/06/2024].

Bushnell, Brian [2024]. *BBMap*. SourceForge. URL : https://sourceforge.net/projects/bbmap/ [visité le 02/12/2024].

Byrt, Caitlin S. et al. [2023]. « Exploring aquaporin functions during changes in leaf water potential ». In : *Frontiers in Plant Science* 14, p. 1213454. ISSN : 1664-462X. DOI : 10.3389/fpls.2023.1213454.

Cadotte, Marc W., Sean M. McMahon et Tadashi Fukami [19 juill. 2006]. *Conceptual Ecology and Invasion Biology : Reciprocal Approaches to Nature*. Google-Books-ID : mG4XvlZdIOAC. Springer Science & Business Media. 506 p. ISBN : 978-1-4020-4925-5.

Carneiro, Laís et al. [20 oct. 2024]. « Biological invasions negatively impact global protected areas ». In : *Science of The Total Environment* 948, p. 174823. ISSN : 0048-9697. DOI : 10.1016/j.scitotenv.2024.174823. URL : https://www.sciencedirect.com/science/article/pii/S0048969724049726 [visité le 30/09/2025].

Casson, Stuart A. et Keith Lindsey [2003]. « Genes and signalling in root development ». In : *New Phytologist* 158.1. _eprint : https://nph.onlinelibrary.wiley.com/doi/pdf/10.1046/j.1469-8137.2003.00705.x, p. 11-38. ISSN : 1469-8137. DOI : 10.1046/j.1469-8137.2003.00705.x. URL : https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1469-8137.2003.00705.x [visité le 12/09/2025].

Castelán-Muñoz, Natalia et al. [2019]. « MADS-Box Genes Are Key Components of Genetic Regulatory Networks Involved in Abiotic Stress and Plastic Developmental Responses in Plants ». In : *Frontiers in Plant Science* 10, p. 853. ISSN : 1664-462X. DOI : 10.3389/fpls.2019.00853.

Cech, Thomas R. et Joan A. Steitz [27 mars 2014]. « The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones ». In : *Cell* 157.1. Publisher : Elsevier, p. 77-94. ISSN : 0092-8674, 1097-4172. DOI : 10.1016/j.cell.2014.03.008. URL :

https://www.cell.com/cell/abstract/S0092-8674(14)00338-9 [visité le 04/06/2025].

Celesti-Grapow, L. et al. [1er juill. 2009]. « Inventory of the non-native flora of Italy ». In : *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* 143.2. Publisher : Taylor & Francis _eprint : https ://doi.org/10.1080/11263500902722824, p. 386-430. ISSN : 1126-3504. DOI : 10.1080/11263500902722824. URL : https://doi.org/10.1080/11263500902722824 [visité le 03/05/2024].

Chakraborty, Abhisek et al. [21 oct. 2022]. « Genome sequencing and comparative analysis of Ficus benghalensis and Ficus religiosa species reveal evolutionary mechanisms of longevity ». In : *iScience* 25.10. Publisher : Elsevier. ISSN : 2589-0042. DOI : 10.1016/j.isci.2022.105100. URL : https://www.cell.com/iscience/abstract/S2589-0042(22)01372-4 [visité le 05/10/2025].

Chan, Patricia P et al. [20 sept. 2021]. « tRNAscan-SE 2.0 : improved detection and functional classification of transfer RNA genes ». In : *Nucleic Acids Research* 49.16, p. 9077-9096. ISSN : 0305-1048. DOI : 10.1093/nar/gkab688. URL : https://doi.org/10.1093/nar/gkab688 [visité le 03/12/2024].

Chang, Yao-Ming et al. [1er sept. 2012]. « Characterizing Regulatory and Functional Differentiation between Maize Mesophyll and Bundle Sheath Cells by Transcriptomic Analysis ». In : *Plant Physiology* 160.1, p. 165-177. ISSN : 0032-0889. DOI : 10.1104/pp.112.203810. URL : https://doi.org/10.1104/pp.112.203810 [visité le 11/09/2025].

Chapple, David G. et Bob B.M. Wong [2016]. « The Role of Behavioural Variation across Different Stages of the Introduction Process ». In : *Biological Invasions and Animal Behaviour*. Sous la dir. de Daniel Sol et Judith S. Weis. Cambridge : Cambridge University Press, p. 7-25. ISBN : 978-1-139-93949-2. DOI : 10.1017/CBO9781139939492.003. URL : https://www.cambridge.org/core/books/biological-invasions-

and-animal-behaviour/role-of-behavioural-variation-across-different-stages-of-the-introduction-process/A6D38D82C31193B88FE2A88AF53177E3 [visité le 20/05/2025].

Chaudhary, Raju et al. [2023]. « Sequencing of Camelina neglecta, a diploid progenitor of the hexaploid oilseed Camelina sativa ». In : *Plant Biotechnology Journal* 21.3. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/pbi.13968, p. 521-535. ISSN : 1467-7652. DOI : 10.1111/pbi.13968. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/pbi.13968 [visité le 08/10/2025].

Chaumont, François et Stephen D. Tyerman [avr. 2014]. « Aquaporins : highly regulated channels controlling plant water relations ». In : *Plant Physiology* 164.4, p. 1600-1618. ISSN : 1532-2548. DOI : 10.1104/pp.113.233791.

Chekanova, Julia A. [27 oct. 2021]. « Plant long non-coding RNAs in the regulation of transcription ». In : *Essays in Biochemistry* 65.4, p. 751-760. ISSN : 0071-1365. DOI : 10.1042/EBC20200090. URL : https://doi.org/10.1042/EBC20200090 [visité le 18/09/2025].

Chen, Jialuo, Su Yang et al. [31 mai 2022]. « The Mediator Complex : A Central Coordinator of Plant Adaptive Responses to Environmental Stresses ». In : *International Journal of Molecular Sciences* 23.11, p. 6170. ISSN : 1422-0067. DOI : 10.3390/ijms23116170.

Chen, Jinliao, Fei Wang et al. [jan. 2023]. « Complete Chloroplast Genomes and Comparative Analyses of Three Paraphalaenopsis (Aeridinae, Orchidaceae) Species ». In : *International Journal of Molecular Sciences* 24.13. Number : 13 Publisher : Multidisciplinary Digital Publishing Institute, p. 11167. ISSN : 1422-0067. DOI : 10.3390/ijms241311167. URL : https://www.mdpi.com/1422-0067/24/13/11167 [visité le 30/07/2025].

Chen, Shifu [2023]. « Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp ». In : *iMeta* 2.2. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/ e107. ISSN : 2770-596X. DOI : 10.1002/imt2.107. URL : https://onlinelibrary.wiley.com/doi/abs/10.1002/imt2.107 [visité le 02/12/2024].

Chen, Shifu, Yanqing Zhou et al. [1er sept. 2018]. « fastp : an ultra-fast all-in-one FASTQ preprocessor ». In : *Bioinformatics* 34.17, p. i884-i890. ISSN : 1367-4803. DOI : 10.1093/bioinformatics/bty560. URL : https://doi.org/10.1093/bioinformatics/bty560 [visité le 26/02/2024].

Chen, Yutang, Roland Kölliker et al. [6 mars 2024]. « An improved chromosome-level genome assembly of perennial ryegrass (Lolium perenne L.) » In : *Gigabyte* 2024. Publisher : GigaScience Press, p. 1-11. ISSN : 2709-4715. DOI : 10.46471/gigabyte.112. URL : https://gigabytejournal.com/articles/112 [visité le 24/09/2025].

Cheng, Haoyu, Gregory T. Concepcion et al. [fév. 2021]. « Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm ». In : *Nature Methods* 18.2. Publisher : Nature Publishing Group, p. 170-175. ISSN : 1548-7105. DOI : 10.1038/s41592-020-01056-5. URL : https://www.nature.com/articles/s41592-020-01056-5 [visité le 21/09/2025].

Cheng, Ling-Tong, Zi-Long Wang et al. [4 mars 2025]. « A long road ahead to reliable and complete medicinal plant genomes ». In : *Nature Communications* 16.1. Publisher : Nature Publishing Group, p. 2150. ISSN : 2041-1723. DOI : 10.1038/s41467-025-57448-8. URL : https://www.nature.com/articles/s41467-025-57448-8 [visité le 25/11/2025].

Chin, Chen-Shan et al. [déc. 2016]. « Phased diploid genome assembly with single-molecule real-time sequencing ». In : *Nature Methods* 13.12. Publisher : Nature Publishing Group, p. 1050-1054. ISSN : 1548-7105. DOI : 10.1038/nmeth.4035. URL : https://www.nature.com/articles/nmeth.4035 [visité le 20/09/2025].

Chong, Leelyn, Pengcheng Guo et Yingfang Zhu [20 oct. 2020]. « Mediator Complex : A Pivotal Regulator of ABA Signaling Pathway and Abiotic Stress Response in Plants ». In : *International Journal of Molecular Sciences* 21.20, p. 7755. ISSN : 1422-0067. DOI : 10.3390/ijms21207755.

Chorlton, Samuel D. [15 mars 2024]. « Ten common issues with reference sequence databases and how to mitigate them ». In : *Frontiers in Bioinformatics* 4. Publisher : Frontiers. ISSN : 2673-7647. DOI : 10.3389/fbinf.2024.1278228. URL : https://www.frontiersin.org/journals/bioinformatics/articles/10.3389/fbinf.2024.1278228/full [visité le 30/09/2025].

Ciufo, Stacy et al. [2018]. « Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI ». In : *International Journal of Systematic and Evolutionary Microbiology* 68.7. Publisher : Microbiology Society, p. 2386-2392. ISSN : 1466-5034. DOI : 10.1099/ijsem.0.002809. URL : https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.002809 [visité le 03/10/2025].

Contreras, Mauricio P. et al. [9 oct. 2023]. « NLR receptors in plant immunity : making sense of the alphabet soup ». In : *EMBO reports* 24.10, e57495. ISSN : 1469-3178. DOI : 10.15252/embr.202357495.

Cooper, Geoffrey M. [2000]. « Chloroplasts and Other Plastids ». In : *The Cell : A Molecular Approach. 2nd edition*. Sinauer Associates. URL : https://www.ncbi.nlm.nih.gov/books/NBK9905/ [visité le 17/09/2025].

Corchete, Luis A. et al. [12 nov. 2020]. « Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis ». In : *Scientific Reports* 10.1. Publisher : Nature Publishing Group, p. 19737. ISSN : 2045-2322. DOI : 10.1038/s41598-020-76881-x. URL : https://www.nature.com/articles/s41598-020-76881-x [visité le 04/08/2025].

Cornet, Luc et Denis Baurain [21 fév. 2022]. « Contamination detection in genomic data : more is not enough ». In : *Genome Biology* 23.1, p. 60. ISSN : 1474-760X. DOI : 10.1186/s13059-022-02619-9. URL : https://doi.org/10.1186/s13059-022-02619-9 [visité le 30/09/2025].

Cruz-Valderrama, José Erik et al. [6 juill. 2019]. « Overview of the Role of Cell Wall DUF642 Proteins in Plant Development ». In : *International Journal of Molecular Sciences* 20.13, p. 3333. ISSN : 1422-0067. DOI : 10.3390/ijms20133333.

Cui, Haonan et al. [déc. 2021]. « Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer ». In : *Scientific Reports* 11.1, p. 1595. ISSN : 2045-2322. DOI : 10.1038/s41598-020-80149-9. URL : http://www.nature.com/articles/s41598-020-80149-9 [visité le 24/05/2022].

Cullinane, Andrew R., Alejandro A. Schäffer et Marjan Huizing [juill. 2013]. « The BEACH is hot : A LYST of emerging roles for BEACH-domain containing proteins in human disease ». In : *Traffic (Copenhagen, Denmark)* 14.7, 10.1111/tra.12069. ISSN : 1398-9219. DOI : 10.1111/tra.12069. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC3761935/ [visité le 09/10/2025].

Curtis, Nicholas, Steven Massey et Sidney Pierce [1er nov. 2006]. « The symbiotic chloroplasts in the sacoglossan Elysia clarki are from several algal species ». In : *Invertebrate Biology* 125, p. 336-345. DOI : 10.1111/j.1744-7410.2006.00065.x.

Dahlgren, Rolf et Robert F. Thorne [1984]. « The Order Myrtales : Circumscription, Variation, and Relationships ». In : *Annals of the Missouri Botanical Garden* 71.3. Publisher : Missouri Botanical Garden Press, p. 633-699. ISSN : 0026-6493. DOI : 10.2307/2399158. URL : https://www.jstor.org/stable/2399158 [visité le 28/09/2025].

Dandelot, Sophie [2004]. « Les Ludwigia spp. invasives du Sud de la France : Historique, Biosystématique, Biologie et Ecologie ». Thèse de doct. Aix-Marseille III : Université Paul Cézanne. 207 p.

Dandelot, Sophie, Robert Matheron et al. [1er oct. 2005]. « Variations temporelles des paramètres physicochimiques et microbiologiques de trois écosystèmes aquatiques (Sud-Est de la France) envahis par des *Ludwigia* ». In : *Comptes Rendus Biologies* 328.10, p. 991-999. ISSN : 1631-0691. DOI : 10.1016/j.crvi.2005.09.007. URL : https://www.sciencedirect.com/science/article/pii/S1631069105001721 [visité le 23/02/2024].

Dandelot, Sophie, Christine Robles et al. [1er mai 2008]. « Allelopathic potential of two invasive alien *Ludwigia* spp. » In : *Aquatic Botany* 88.4, p. 311-316. ISSN : 0304-3770. DOI : 10.1016/j.aquabot.2007.12.004. URL : https://www.sciencedirect.com/science/article/pii/S0304377007001805 [visité le 23/02/2024].

Daniell, Henry et al. [2021]. « Green giant—a tiny chloroplast genome with mighty power to produce high-value proteins : history and phylogeny ». In : *Plant Biotechnology Journal* 19.3. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/pbi.13556, p. 430-447. ISSN : 1467-7652. DOI : 10.1111/pbi.13556. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/pbi.13556 [visité le 12/08/2025].

De Caroli, Monica et al. [juill. 2021]. « Ride to cell wall : Arabidopsis XTH11, XTH29 and XTH33 exhibit different secretion pathways and responses to heat and drought stress ». In : *The Plant Journal* 107.2, p. 448-466. ISSN : 0960-7412. DOI : 10.1111/tpj.15301. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC8453972/ [visité le 09/10/2025].

De Micco, Veronica et Giovanna Aronne [2012]. « Morpho-Anatomical Traits for Plant Adaptation to Drought ». In : *Plant Responses to Drought Stress : From Morphological to Molecular Features*. Sous la dir. de Ricardo Aroca. Berlin, Heidelberg :

Springer, p. 37-61. ISBN : 978-3-642-32653-0. DOI : 10.1007/978-3-642-32653-0_2. URL : https://doi.org/10.1007/978-3-642-32653-0_2 [visité le 12/10/2025].

Delay, Christina, Nijat Imin et Michael A. Djordjevic [déc. 2013]. « CEP genes regulate root and shoot development in response to environmental cues and are specific to seed plants ». In : *Journal of Experimental Botany* 64.17, p. 5383-5394. ISSN : 1460-2431. DOI : 10.1093/jxb/ert332.

Deng, Lele et al. [1er sept. 2025]. « *OEE1* affects the tobacco photosynthesis through the hormone regulation ». In : *Industrial Crops and Products* 231, p. 121171. ISSN : 0926-6690. DOI : 10.1016/j.indcrop.2025.121171. URL : https://www.sciencedirect.com/science/article/pii/S0926669025007174 [visité le 09/10/2025].

Deschamps, Stéphane et al. [19 nov. 2018]. « A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping ». In : *Nature Communications* 9.1. Publisher : Nature Publishing Group, p. 4844. ISSN : 2041-1723. DOI : 10.1038/s41467-018-07271-1. URL : https://www.nature.com/articles/s41467-018-07271-1 [visité le 19/09/2025].

Detcharoen, Matsapume, Sara Bumrungsri et Supayang Piyawan Voravuthikunchai [jan. 2023]. « Complete Genome of Rose Myrtle, Rhodomyrtus tomentosa, and Its Population Genetics in Thai Peninsula ». In : *Plants* 12.8. Number : 8 Publisher : Multidisciplinary Digital Publishing Institute, p. 1582. ISSN : 2223-7747. DOI : 10.3390/plants12081582. URL : https://www.mdpi.com/2223-7747/12/8/1582 [visité le 03/05/2025].

Devaiah, Ballachanda N., Athikkattuvalasu S. Karthikeyan et Kashchandra G. Raghothama [avr. 2007]. « WRKY75 transcription factor is a modulator of phosphate acquisition and root development in Arabidopsis ». In : *Plant Physiology* 143.4, p. 1789-1801. ISSN : 0032-0889. DOI : 10.1104/pp.106.093971.

DeWalt, Saara J. et J. L. Hamrick [2004]. « Genetic variation of introduced Hawaiian and native Costa Rican populations of an invasive tropical shrub, Clidemia hirta (Melastomataceae) ». In : *American Journal of Botany* 91.8. _eprint : https ://bsa-pubs.onlinelibrary.wiley.com/doi/pdf/10.3732/ajb.91.8.1155, p. 1155-1163. ISSN : 1537-2197. DOI : 10.3732/ajb.91.8.1155. URL : https://onlinelibrary.wiley.com/doi/abs/10.3732/ajb.91.8.1155 [visité le 26/09/2025].

Dierckxsens, Nicolas, Patrick Mardulyn et Guillaume Smits [28 fév. 2017]. « NOVOPlasty : de novo assembly of organelle genomes from whole genome data ». In : *Nucleic Acids Research* 45.4, e18. ISSN : 0305-1048. DOI : 10.1093/nar/gkw955. URL : https://doi.org/10.1093/nar/gkw955 [visité le 30/07/2025].

Dijk, Erwin L. van et al. [1er sept. 2023]. « Genomics in the long-read sequencing era ». In : *Trends in Genetics* 39.9. Publisher : Elsevier, p. 649-671. ISSN : 0168-9525. DOI : 10.1016/j.tig.2023.04.006. URL : https://www.cell.com/trends/genetics/abstract/S0168-9525(23)00119-1 [visité le 19/09/2025].

Dobin, Alexander et al. [1er jan. 2013]. « STAR : ultrafast universal RNA-seq aligner ». In : *Bioinformatics (Oxford, England)* 29.1, p. 15-21. ISSN : 1367-4811. DOI : 10.1093/bioinformatics/bts635.

Dobrogojski, Jędrzej, Małgorzata Adamiec et Robert Luciński [18 mai 2020]. « The chloroplast genome : a review ». In : *Acta Physiologiae Plantarum* 42.6, p. 98. ISSN : 1861-1664. DOI : 10.1007/s11738-020-03089-x. URL : https://doi.org/10.1007/s11738-020-03089-x [visité le 16/06/2025].

Doré, Guillaume, Dominique Barloy et Frédérique Barloy-Hubler [2 juill. 2024]. « De Novo Hybrid Assembly Unveils Multi-Chromosomal Mitochondrial Genomes in Ludwigia Species, Highlighting Genomic Recombination, Gene Transfer, and RNA Editing Events ». In : *International Journal of Molecular Sciences* 25.13. Number : 13 Publisher : Multidisciplinary Digital Publishing Institute, p. 7283. ISSN : 1422-0067.

DOI : 10.3390/ijms25137283. URL : https://www.mdpi.com/1422-0067/25/13/7283 [visité le 12/03/2025].

Dos Santos, Cristiane et Octávio Luiz Franco [5 juin 2023]. « Pathogenesis-Related Proteins (PRs) with Enzyme Activity Activating Plant Defense Responses ». In : *Plants (Basel, Switzerland)* 12.11, p. 2226. ISSN : 2223-7747. DOI : 10.3390/plants12112226.

Douce, Roland [2 déc. 2012]. *Mitochondria in Higher Plants : Structure, Function, and Biogenesis*. Google-Books-ID : bPUX8UckVvsC. Elsevier. 345 p. ISBN : 978-0-323-15505-2.

Doyle, Jeff J. et al. [1er avr. 1996]. « The Distribution and Phylogenetic Significance of a 50-kb Chloroplast DNA Inversion in the Flowering Plant Family Leguminosae ». In : *Molecular Phylogenetics and Evolution* 5.2, p. 429-438. ISSN : 1055-7903. DOI : 10.1006/mpev.1996.0038. URL : https://www.sciencedirect.com/science/article/pii/S105579039690038X [visité le 17/09/2025].

Dugas, Diana V et Bonnie Bartel [1er oct. 2004]. « MicroRNA regulation of gene expression in plants ». In : *Current Opinion in Plant Biology* 7.5, p. 512-520. ISSN : 1369-5266. DOI : 10.1016/j.pbi.2004.07.011. URL : https://www.sciencedirect.com/science/article/pii/S1369526604001037 [visité le 03/06/2025].

Duminil, Jérôme et Guillaume Besnard [2021]. « Utility of the Mitochondrial Genome in Plant Taxonomic Studies ». In : *Molecular Plant Taxonomy : Methods and Protocols*. Sous la dir. de Pascale Besse. New York, NY : Springer US, p. 107-118. ISBN : 978-1-0716-0997-2. DOI : 10.1007/978-1-0716-0997-2_6. URL : https://doi.org/10.1007/978-1-0716-0997-2_6 [visité le 23/09/2025].

Dunwell, Jim M. et al. [25 sept. 2008]. « Germin and Germin-like Proteins : Evolution, Structure, and Function ». In : *Critical Reviews in Plant Sciences* 27.5. Publisher : Taylor & Francis _eprint : https ://doi.org/10.1080/07352680802333938, p. 342-375.

ISSN : 0735-2689. DOI : 10.1080/07352680802333938. URL : https://doi.org/10.1080/07352680802333938 [visité le 12/10/2025].

Dupree, Paul et D. Janine Sherrier [14 août 1998]. « The plant Golgi apparatus ». In : *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1404.1, p. 259-270. ISSN : 0167-4889. DOI : 10.1016/S0167-4889(98)00061-5. URL : https://www.sciencedirect.com/science/article/pii/S0167488998000615 [visité le 12/09/2025].

Durrant, Wendy E., Shui Wang et Xinnian Dong [6 mars 2007]. « Arabidopsis SNI1 and RAD51D regulate both gene transcription and DNA recombination during the defense response ». In : *Proceedings of the National Academy of Sciences of the United States of America* 104.10, p. 4223-4227. ISSN : 0027-8424. DOI : 10.1073/pnas.0609357104.

Edgar, E. [1er juill. 1986]. « Poa L. in New Zealand ». In : *New Zealand Journal of Botany* 24.3. Publisher : Taylor & Francis _eprint : https ://doi.org/10.1080/0028825X.1986.10409820, p. 425-503. ISSN : 0028-825X. DOI : 10.1080/0028825X.1986.10409820. URL : https://doi.org/10.1080/0028825X.1986.10409820 [visité le 18/09/2025].

Edqvist, Johan et al. [août 2018]. « Plant lipid transfer proteins : are we finally closing in on the roles of these enigmatic proteins ? » In : *Journal of Lipid Research* 59.8, p. 1374-1382. ISSN : 1539-7262. DOI : 10.1194/jlr.R083139.

Emonet, Aurélia et al. [23 déc. 2024]. « Polyploid genome assembly of Cardamine chenopodiifolia ». In : *Gigabyte* 2024. Publisher : GigaScience Press, p. 0–0. ISSN : 2709-4715. DOI : 10.46471/gigabyte.145. URL : https://gigabytejournal.com/articles/145 [visité le 03/05/2025].

EPPO [2011]. « Ludwigia grandiflora and L. peploides Onagraceae – Water primroses ». In : *EPPO Bulletin* 41.3. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2338.2011.02511.x, p. 414-418. ISSN : 1365-2338. DOI : 10.1111/j.1365-2338.2011.

02511.x. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2338.2011.02511.x [visité le 06/08/2025].

Eren, A. Murat et al. [8 oct. 2015]. « Anvi'o : an advanced analysis and visualization platform for 'omics data ». In : *PeerJ* 3. Publisher : PeerJ Inc., e1319. ISSN : 2167-8359. DOI : 10.7717/peerj.1319. URL : https://peerj.com/articles/1319 [visité le 21/09/2025].

Ernst, Antonia M. et al. [10 juill. 2012]. « Sieve element occlusion (SEO) genes encode structural phloem proteins involved in wound sealing of the phloem ». In : *Proceedings of the National Academy of Sciences of the United States of America* 109.28, E1980-E1989. ISSN : 0027-8424. DOI : 10.1073/pnas.1202999109. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC3396537/ [visité le 12/10/2025].

Evans, Ciaran, Johanna Hardin et Daniel M Stoebel [28 sept. 2018]. « Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions ». In : *Briefings in Bioinformatics* 19.5, p. 776-792. ISSN : 1477-4054. DOI : 10.1093/bib/bbx008. URL : https://doi.org/10.1093/bib/bbx008 [visité le 28/07/2025].

Fahrenkrog, Annette M. et al. [14 déc. 2022]. « Chloroplast genome assemblies and comparative analyses of commercially important Vaccinium berry crops ». In : *Scientific Reports* 12.1. Publisher : Nature Publishing Group, p. 21600. ISSN : 2045-2322. DOI : 10.1038/s41598-022-25434-5. URL : https://www.nature.com/articles/s41598-022-25434-5 [visité le 30/07/2025].

Fan, Jingwei et al. [7 avr. 2025]. « A large-scale integrated transcriptomic atlas for soybean organ development ». In : *Molecular Plant* 18.4. Publisher : Elsevier, p. 669-689. ISSN : 1674-2052. DOI : 10.1016/j.molp.2025.02.003. URL : https://www.cell.com/molecular-plant/abstract/S1674-2052(25)00069-3 [visité le 10/10/2025].

Fang, Bo, Jingling Li et al. [déc. 2021]. « Assembly of the Complete Mitochondrial Genome of Chinese Plum (Prunus salicina) : Characterization of Genome Recombination and RNA Editing Sites ». In : *Genes* 12.12. Number : 12 Publisher : Multidisciplinary Digital Publishing Institute, p. 1970. ISSN : 2073-4425. DOI : 10.3390/genes12121970. URL : https://www.mdpi.com/2073-4425/12/12/1970 [visité le 18/12/2023].

Fang, Yuanpeng, Junmei Jiang et al. [7 sept. 2022]. « Plant protein-coding gene families : Their origin and evolution ». In : *Frontiers in Plant Science* 13. Publisher : Frontiers. ISSN : 1664-462X. DOI : 10.3389/fpls.2022.995746. URL : https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.995746/full [visité le 04/06/2025].

Feathers, J. Ryan, Ryan C. Vignogna et J. Christopher Fromme [4 déc. 2024]. « Structural basis for Rab6 activation by the Ric1-Rgp1 complex ». In : *Nature Communications* 15.1, p. 10561. ISSN : 2041-1723. DOI : 10.1038/s41467-024-54869-9.

Feuillet, C., G. Schachermayr et B. Keller [jan. 1997]. « Molecular cloning of a new receptor-like kinase gene encoded at the Lr10 disease resistance locus of wheat ». In : *The Plant Journal : For Cell and Molecular Biology* 11.1, p. 45-52. ISSN : 0960-7412. DOI : 10.1046/j.1365-313x.1997.11010045.x.

Fischer, Axel et al. [1er juin 2022]. « Graph-based models of the Oenothera mitochondrial genome capture the enormous complexity of higher plant mitochondrial DNA organization ». In : *NAR Genomics and Bioinformatics* 4.2, lqac027. ISSN : 2631-9268. DOI : 10.1093/nargab/lqac027. URL : https://doi.org/10.1093/nargab/lqac027 [visité le 20/01/2023].

Flynn, Jullien M. et al. [28 avr. 2020]. « RepeatModeler2 for automated genomic discovery of transposable element families ». In : *Proceedings of the National Academy of Sciences* 117.17. Publisher : Proceedings of the National Academy of Sciences,

p. 9451-9457. DOI : 10.1073/pnas.1921046117. URL : https://www.pnas.org/doi/abs/10.1073/pnas.1921046117 [visité le 21/09/2025].

Frandsen, Paul B et al. [19 août 2025]. « A chromosome-scale genome assembly and annotation of the tetraploid herb "epazote" (Dysphania ambrosioides) ». In : *G3 Genes|Genomes|Genetics*, jkaf191. ISSN : 2160-1836. DOI : 10.1093/g3journal/jkaf191. URL : https://doi.org/10.1093/g3journal/jkaf191 [visité le 24/09/2025].

Franklin, Daniel N. et al. [1er mai 2017]. « Invasion dynamics of Asian hornet, Vespa velutina (Hymenoptera : Vespidae) : a case study of a commune in south-west France ». In : *Applied Entomology and Zoology* 52.2, p. 221-229. ISSN : 1347-605X. DOI : 10.1007/s13355-016-0470-z. URL : https://doi.org/10.1007/s13355-016-0470-z [visité le 23/02/2024].

Freudenthal, Jan A. et al. [28 sept. 2020]. « A systematic comparison of chloroplast genome assembly tools ». In : *Genome Biology* 21.1, p. 254. ISSN : 1474-760X. DOI : 10.1186/s13059-020-02153-6. URL : https://doi.org/10.1186/s13059-020-02153-6 [visité le 30/07/2025].

Gallardo, Belinda et David C. Aldridge [2013]. « The 'dirty dozen' : socio-economic factors amplify the invasion potential of 12 high-risk aquatic invasive species in Great Britain and Ireland ». In : *Journal of Applied Ecology* 50.3. _eprint : https ://onlinelibrary.wiley.com/d 2664.12079, p. 757-766. ISSN : 1365-2664. DOI : 10.1111/1365-2664.12079. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2664.12079 [visité le 03/05/2024].

Gao, Yi-Qun, Jin-Quan Huang et al. [27 oct. 2023]. « A dirigent protein complex directs lignin polymerization and assembly of the root diffusion barrier ». In : *Science (New York, N.Y.)* 382.6669, p. 464-471. ISSN : 1095-9203. DOI : 10.1126/science.adi5032.

Gao, Yaping, Congwu He et al. [jan. 2017]. « Two Trichome Birefringence-Like Proteins Mediate Xylan Acetylation, Which Is Essential for Leaf Blight Resistance in Rice ». In : *Plant Physiology* 173.1, p. 470-481. ISSN : 1532-2548. DOI : 10.1104/pp.16.01618.

Garai, Sampurna et al. [oct. 2021]. « What signals the glyoxalase pathway in plants ? » In : *Physiology and Molecular Biology of Plants : An International Journal of Functional Plant Biology* 27.10, p. 2407-2420. ISSN : 0971-5894. DOI : 10.1007/s12298-021-00991-7.

García-Ortega, Luis Fernando et Octavio Martínez [24 juin 2015]. « How Many Genes Are Expressed in a Transcriptome ? Estimation and Results for RNA-Seq ». In : *PLOS ONE* 10.6. Publisher : Public Library of Science, e0130262. ISSN : 1932-6203. DOI : 10.1371/journal.pone.0130262. URL : https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0130262 [visité le 11/09/2025].

Garg, Rohini et al. [1er août 2011]. « Gene Discovery and Tissue-Specific Transcriptome Analysis in Chickpea with Massively Parallel Pyrosequencing and Web Resource Development ». In : *Plant Physiology* 156.4, p. 1661-1678. ISSN : 0032-0889. DOI : 10.1104/pp.111.178616. URL : https://doi.org/10.1104/pp.111.178616 [visité le 11/09/2025].

Genitoni, Julien [2019]. « Acclimatation de l'espèce aquatique invasive, Ludwigia grandiflora, au milieu terrestre : Approches physiologique et épigénétique ». Thèse de doct. Rennes : Agrocampus Ouest. 411 p.

Genitoni, Julien, Danièle Vassaux, Alain Delaunay et al. [2020]. « Hypomethylation of the aquatic invasive plant, Ludwigia grandiflora subsp. hexapetala mimics the adaptive transition into the terrestrial morphotype ». In : *Physiologia Plantarum* 170.2. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/ppl.13162, p. 280-

298. ISSN : 1399-3054. DOI : 10.1111/ppl.13162. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/ppl.13162 [visité le 23/02/2024].

Genitoni, Julien, Danièle Vassaux, David Renault et al. [1er sept. 2025]. *Adaptive Strategies of the invasive aquatic plant, Ludwigia grandiflora subps. hexapetala : Contrasting Plasticity Between Aquatic and Terrestrial Morphotypes*. ISSN : 2692-8205 Pages : 2025.08.27.672630 Section : New Results. DOI : 10.1101/2025.08.27.672630. URL : https://www.biorxiv.org/content/10.1101/2025.08.27.672630v1 [visité le 26/09/2025].

Ghurye, Jay et Mihai Pop [5 juin 2019]. « Modern technologies and algorithms for scaffolding assembled genomes ». In : *PLOS Computational Biology* 15.6. Publisher : Public Library of Science, e1006994. ISSN : 1553-7358. DOI : 10.1371/journal.pcbi.1006994. URL : https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006994 [visité le 05/10/2025].

Gil, Pedro et al. [1er août 2001]. « BIG : a calossin-like protein required for polar auxin transport in Arabidopsis ». In : *Genes & Development* 15.15, p. 1985-1997. ISSN : 0890-9369. DOI : 10.1101/gad.905201. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC312751/ [visité le 09/10/2025].

Gilbert, Don [13 juill. 2016]. « Gene-omes built from mRNA seq not genome DNA ». In : *F1000Research* 5. DOI : 10.7490/f1000research.1112594.1. URL : https://f1000research.com/posters/5-1695 [visité le 02/09/2025].

Gioria, Margherita et al. [22 mai 2023]. « Why Are Invasive Plants Successful ? » In : *Annual Review of Plant Biology* 74 [Volume 74, 2023]. Publisher : Annual Reviews, p. 635-670. ISSN : 1543-5008, 1545-2123. DOI : 10.1146/annurev-arplant-070522-071021. URL : https://www.annualreviews.org/content/journals/10.1146/annurev-arplant-070522-071021 [visité le 03/05/2024].

Gladieux, Pierre et al. [1er avr. 2011]. « Distinct invasion sources of common ragweed (Ambrosia artemisiifolia) in Eastern and Western Europe ». In : *Biological Invasions* 13.4, p. 933-944. ISSN : 1573-1464. DOI : 10.1007/s10530-010-9880-y. URL : https://doi.org/10.1007/s10530-010-9880-y [visité le 21/05/2025].

Glover, Natasha M., Henning Redestig et Christophe Dessimoz [juill. 2016]. « Homoeologs : What Are They and How Do We Infer Them ? » In : *Trends in Plant Science* 21.7, p. 609-621. ISSN : 1878-4372. DOI : 10.1016/j.tplants.2016.02.005.

Goodwin, Sara, John D. McPherson et W. Richard McCombie [17 mai 2016]. « Coming of age : ten years of next-generation sequencing technologies ». In : *Nature reviews. Genetics* 17.6, p. 333-351. ISSN : 1471-0056. DOI : 10.1038/nrg.2016.49. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10373632/ [visité le 16/05/2025].

Grabherr, Manfred G. et al. [15 mai 2011]. « Trinity : reconstructing a full-length transcriptome without a genome from RNA-Seq data ». In : *Nature biotechnology* 29.7, p. 644-652. ISSN : 1087-0156. DOI : 10.1038/nbt.1883. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3571712/ [visité le 26/02/2024].

Gray, Michael W. et Patrick S. Covello [1993]. « RNA editing in plant mitochondria and chloroplasts ». In : *The FASEB Journal* 7.1. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1096/fasebj.7.1.8422976, p. 64-71. ISSN : 1530-6860. DOI : 10.1096/fasebj.7.1.8422976. URL : https://onlinelibrary.wiley.com/doi/abs/10.1096/fasebj.7.1.8422976 [visité le 02/02/2024].

Grewell, Brenda, Michael Netherland et Meghan Skaer Thomason [16 fév. 2016]. *Establishing Research and Management Priorities for Invasive Water Primroses (Ludwigia spp.)* DOI : 10.13140/RG.2.1.5020.6482.

Gualberto, José M. et Kathleen J. Newton [28 avr. 2017]. « Plant Mitochondrial Genomes : Dynamics and Mechanisms of Mutation ». In : *Annual Review of Plant Bio-*

logy 68.1, p. 225-252. ISSN : 1543-5008, 1545-2123. DOI : 10.1146/annurev-arplant-043015-112232. URL : https://www.annualreviews.org/doi/10.1146/annurev-arplant-043015-112232 [visité le 16/01/2023].

Guo, Changkui, Yunmin Xu et al. [juin 2017]. « Repression of miR156 by miR159 Regulates the Timing of the Juvenile-to-Adult Transition in Arabidopsis ». In : *The Plant Cell* 29.6, p. 1293-1304. ISSN : 1040-4651. DOI : 10.1105/tpc.16.00975. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC5502449/ [visité le 18/09/2025].

Guo, Meng, Jin-Hong Liu et al. [2016]. « The Plant Heat Stress Transcription Factors (HSFs) : Structure, Regulation, and Function in Response to Abiotic Stresses ». In : *Frontiers in Plant Science* 7, p. 114. ISSN : 1664-462X. DOI : 10.3389/fpls.2016.00114.

Guo, Xuelian, Xiaodi Hu et al. [6 juin 2023]. « The Sapria himalayana genome provides new insights into the lifestyle of endoparasitic plants ». In : *BMC Biology* 21, p. 134. ISSN : 1741-7007. DOI : 10.1186/s12915-023-01620-3. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC10246380/ [visité le 18/09/2025].

Guo, Yanbing, Ziwei Li et al. [31 mai 2023]. « Assembly and Comparative Analysis of the Complete Mitochondrial Genome of Two Species of Calla Lilies (Zantedeschia, Araceae) ». In : *International Journal of Molecular Sciences* 24.11, p. 9566. ISSN : 1422-0067. DOI : 10.3390/ijms24119566.

Haas, Brian John [2024]. *GitHub - TransDecoder/TransDecoder : TransDecoder source*. URL : https://github.com/TransDecoder/TransDecoder [visité le 03/12/2024].

Haghshenas, Ehsan et al. [21 août 2020]. « HASLR : Fast Hybrid Assembly of Long Reads ». In : *iScience* 23.8, p. 101389. ISSN : 2589-0042. DOI : 10.1016/j.isci.2020.101389. URL : https://www.sciencedirect.com/science/article/pii/S2589004220305770 [visité le 16/05/2025].

Halestrap, A. P. et al. [1er fév. 2000]. « Mitochondria and cell death ». In : *Biochemical Society Transactions* 28.2, p. 170-177. ISSN : 0300-5127. DOI : 10.1042/bst0280170. URL : https://doi.org/10.1042/bst0280170 [visité le 17/09/2025].

Hämälä, Tuomas et al. [25 juin 2024]. « Impact of whole-genome duplications on structural variant evolution in Cochlearia ». In : *Nature Communications* 15, p. 5377. ISSN : 2041-1723. DOI : 10.1038/s41467-024-49679-y. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC11199601/ [visité le 26/09/2025].

Han, Shijuan, Zhenfei Xing et al. [1er nov. 2021]. « Biological adaptive mechanisms displayed by a freshwater plant to live in aquatic and terrestrial environments ». In : *Environmental and Experimental Botany* 191, p. 104623. ISSN : 0098-8472. DOI : 10.1016/j.envexpbot.2021.104623. URL : https://www.sciencedirect.com/science/article/pii/S0098847221002537 [visité le 31/05/2024].

Han, Yixing, Shouguo Gao et al. [15 nov. 2015]. « Advanced Applications of RNA Sequencing and Challenges ». In : *Bioinformatics and Biology Insights* 9 [Suppl 1], p. 29-46. ISSN : 1177-9322. DOI : 10.4137/BBI.S28991. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC4648566/ [visité le 22/09/2025].

Han, Zhu, Dianguang Xiong et al. [juin 2023]. « The function of plant PR1 and other members of the CAP protein superfamily in plant-pathogen interactions ». In : *Molecular Plant Pathology* 24.6, p. 651-668. ISSN : 1364-3703. DOI : 10.1111/mpp.13320.

Hao, Yan et al. [8 avr. 2024]. « Genomic and phenotypic signatures provide insights into the wide adaptation of a global plant invader ». In : *Plant Communications*. Topical Review Focus Issue 5.4, p. 100820. ISSN : 2590-3462. DOI : 10.1016/j.xplc.2024.100820. URL : https://www.sciencedirect.com/science/article/pii/S2590346224000403 [visité le 10/10/2025].

Haury, Jacques, Arsène Druel et al. [1er oct. 2014]. « Which adaptations of some invasive Ludwigia spp. (Rosidae, Onagraceae) populations occur in contrasting hydrological

conditions in Western France ? » In : *Hydrobiologia* 737.1, p. 45-56. ISSN : 1573-5117. DOI : 10.1007/s10750-014-1815-7. URL : https://doi.org/10.1007/s10750-014-1815-7 [visité le 03/05/2024].

Haury, Jacques, Florian Noël et al. [3 oct. 2011]. « Importance of Ludwigia grandiflora as invasive weed on meadows and pastures in Western France ». In : 3rd International Symposium on Weeds and Invasive Plants. URL : https://institut-agro-rennes-angers.hal.science/hal-00842412 [visité le 23/02/2024].

Hayes, C. Nelson et al. [déc. 2024]. « From Omics to Multi-Omics : A Review of Advantages and Tradeoffs ». In : *Genes* 15.12. Publisher : Multidisciplinary Digital Publishing Institute, p. 1551. ISSN : 2073-4425. DOI : 10.3390/genes15121551. URL : https://www.mdpi.com/2073-4425/15/12/1551 [visité le 04/10/2025].

Hazra, Soham et al. [1er sept. 2024]. « Nucleoredoxin Vis-à-Vis a Novel Thioredoxin in Regulating Oxidative Stress in Plants : A Review ». In : *Agricultural Research* 13.3, p. 400-418. ISSN : 2249-7218. DOI : 10.1007/s40003-024-00737-3. URL : https://doi.org/10.1007/s40003-024-00737-3 [visité le 13/10/2025].

He, Xin-Jian et al. [1er mai 2009]. « An effector of RNA-directed DNA methylation in Arabidopsis is an ARGONAUTE 4- and RNA-binding protein ». In : *Cell* 137.3, p. 498-508. ISSN : 0092-8674. DOI : 10.1016/j.cell.2009.04.028. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC2700824/ [visité le 18/09/2025].

Hellsberg, Eva, Floriane Montanari et Gerhard F. Ecker [avr. 2015]. « The ABC of Phytohormone Translocation ». In : *Planta Medica* 81.6, p. 474-487. ISSN : 1439-0221. DOI : 10.1055/s-0035-1545880.

Hemleben, Vera et al. [21 déc. 2021]. « Personal Perspectives on Plant Ribosomal RNA Genes Research : From Precursor-rRNA to Molecular Evolution ». In : *Frontiers in Plant Science* 12. Publisher : Frontiers. ISSN : 1664-462X. DOI : 10.3389/fpls.

2021.797348. URL : https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2021.797348/full [visité le 04/06/2025].

Herrero, E. et M. A. de la Torre-Ruiz [juin 2007]. « Monothiol glutaredoxins : a common domain for multiple functions ». In : *Cellular and molecular life sciences : CMLS* 64.12, p. 1518-1530. ISSN : 1420-682X. DOI : 10.1007/s00018-007-6554-8.

Heslop-Harrison, J. S. (Pat) et Trude Schwarzacher [2011]. « Organisation of the plant genome in chromosomes ». In : *The Plant Journal* 66.1. _eprint : https ://onlinelibrary.wiley.com/doi/313X.2011.04544.x, p. 18-33. ISSN : 1365-313X. DOI : 10.1111/j.1365-313X.2011.04544.x. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-313X.2011.04544.x [visité le 23/07/2025].

Hieda, Shinya et al. [2020]. « Ludwigia grandiflora (Michx.) Greuter & Burdet subsp. hexapetala (Hook. & Arn.) G. L. Nesom & Kartesz, an Invasive Aquatic Plant in Lake Biwa, the Largest Lake in Japan ». In : *Acta Phytotaxonomica et Geobotanica* 71.1, p. 65-71. DOI : 10.18942/apg.201911.

Hodgins, Kathryn A., Paul Battlay et Dan G. Bock [2025]. « The genomic secrets of invasive plants ». In : *New Phytologist* 245.5. _eprint : https ://nph.onlinelibrary.wiley.com/doi/pdf/10.1 p. 1846-1863. ISSN : 1469-8137. DOI : 10.1111/nph.20368. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.20368 [visité le 26/09/2025].

Hoefnagel, Marcel H. N, Owen K Atkin et Joseph T Wiskich [7 sept. 1998]. « Interdependence between chloroplasts and mitochondria in the light and the dark ». In : *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1366.3, p. 235-255. ISSN : 0005-2728. DOI : 10.1016/S0005-2728(98)00126-1. URL : https://www.sciencedirect.com/science/article/pii/S0005272898001261 [visité le 17/09/2025].

Hoff, KJ et M Stanke [1er fév. 2015]. « Current methods for automated annotation of protein-coding genes ». In : *Current Opinion in Insect Science*. Insect genomics * Development and regulation 7, p. 8-14. ISSN : 2214-5745. DOI : 10.1016/j.cois.

2015.02.008. URL : https://www.sciencedirect.com/science/article/pii/S2214574515000267 [visité le 21/09/2025].

Holley, Guillaume et al. [8 jan. 2021]. « Ratatosk : hybrid error correction of long reads enables accurate variant calling and assembly ». In : *Genome Biology* 22.1, p. 28. ISSN : 1474-760X. DOI : 10.1186/s13059-020-02244-4. URL : https://doi.org/10.1186/s13059-020-02244-4 [visité le 26/02/2024].

Horn, Renate, Kapuganti J. Gupta et Noemi Colombo [1er nov. 2014]. « Mitochondrion role in molecular basis of cytoplasmic male sterility ». In : *Mitochondrion*. Plant Mitochondria in Mitochondrion 19, p. 198-205. ISSN : 1567-7249. DOI : 10.1016/j.mito.2014.04.004. URL : https://www.sciencedirect.com/science/article/pii/S1567724914000427 [visité le 23/09/2025].

Hu, Ying et al. [28 août 2025]. « Highly contiguous genome of the medicinal plant Sarcandra glabra (Thunb.) Nakai ». In : *Scientific Data* 12.1. Publisher : Nature Publishing Group, p. 1508. ISSN : 2052-4463. DOI : 10.1038/s41597-025-05796-x. URL : https://www.nature.com/articles/s41597-025-05796-x [visité le 24/09/2025].

Huang, Peng et al. [4 sept. 2024]. « A genome assembly of decaploid Houttuynia cordata provides insights into the evolution of Houttuynia and the biosynthesis of alkaloids ». In : *Horticulture Research* 11.9, uhae203. ISSN : 2662-6810. DOI : 10.1093/hr/uhae203. URL : https://doi.org/10.1093/hr/uhae203 [visité le 09/05/2025].

Hui, Jerome H. L. et al. [20 mai 2024]. « Genome assembly of the rare and endangered Grantham's camellia, Camellia granthamiana ». In : *Gigabyte* 2024. Publisher : GigaScience Press, p. 1-12. ISSN : 2709-4715. DOI : 10.46471/gigabyte.124. URL : https://gigabytejournal.com/articles/124 [visité le 03/05/2025].

Hulme, Philip E. [21 mai 2021]. « Unwelcome exchange : International trade as a direct and indirect driver of biological invasions worldwide ». In : *One Earth* 4.5, p. 666-

679. ISSN : 2590-3322. DOI : 10.1016/j.oneear.2021.04.015. URL : https://www.sciencedirect.com/science/article/pii/S2590332221002347 [visité le 19/05/2025].

Hussner, Andreas, I. Stiers et al. [1er jan. 2017]. « Management and control methods of invasive alien freshwater aquatic plants : A review ». In : *Aquatic Botany* 136, p. 112-137. ISSN : 0304-3770. DOI : 10.1016/j.aquabot.2016.08.002. URL : https://www.sciencedirect.com/science/article/pii/S0304377016300924 [visité le 23/02/2024].

Hussner, Andreas, M Windhaus et U Starfinger [2016]. « From weed biology to successful control : an example of successful management of Ludwigia grandiflora in Germany ». In : *Weed Research* 56.6. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/wre.1222 p. 434-441. ISSN : 1365-3180. DOI : 10.1111/wre.12224. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/wre.12224 [visité le 23/02/2024].

Iseli, C., C. V. Jongeneel et P. Bucher [1999]. « ESTScan : a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences ». In : *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, p. 138-148. ISSN : 1553-0833.

Jacoby, Richard P. et al. [2012]. « Mitochondrial Composition, Function and Stress Response in Plants ». In : *Journal of Integrative Plant Biology* 54.11. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1744-7909.2012.01177.x, p. 887-906. ISSN : 1744-7909. DOI : 10.1111/j.1744-7909.2012.01177.x. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1744-7909.2012.01177.x [visité le 17/09/2025].

Jaillon, Olivier et al. [sept. 2007]. « The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla ». In : *Nature* 449.7161. Publisher :

Nature Publishing Group, p. 463-467. ISSN : 1476-4687. DOI : 10.1038/nature06148. URL : https://www.nature.com/articles/nature06148 [visité le 09/10/2025].

Jansen, Robert K. et al. [1er sept. 2008]. « Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae) ». In : *Molecular Phylogenetics and Evolution* 48.3, p. 1204-1217. ISSN : 1055-7903. DOI : 10.1016/j.ympev.2008.06.013. URL : https://www.sciencedirect.com/science/article/pii/S105579030800328X [visité le 30/07/2025].

Jeon, Donghyun et Changsoo Kim [jan. 2024]. « Polyploids of Brassicaceae : Genomic Insights and Assembly Strategies ». In : *Plants* 13.15. Number : 15 Publisher : Multidisciplinary Digital Publishing Institute, p. 2087. ISSN : 2223-7747. DOI : 10.3390/plants13152087. URL : https://www.mdpi.com/2223-7747/13/15/2087 [visité le 29/07/2025].

Jiang, Libo et al. [1er août 2024]. « The Bioinformatic Applications of Hi-C and Linked Reads ». In : *Genomics, Proteomics & Bioinformatics* 22.4, qzae048. ISSN : 1672-0229. DOI : 10.1093/gpbjnl/qzae048. URL : https://doi.org/10.1093/gpbjnl/qzae048 [visité le 12/08/2025].

Jiao, Wen-Biao et Korbinian Schneeberger [1er avr. 2017]. « The impact of third generation genomic technologies on plant genome assembly ». In : *Current Opinion in Plant Biology*. 36 Genome studies and molecular genetics 36, p. 64-70. ISSN : 1369-5266. DOI : 10.1016/j.pbi.2017.02.002. URL : https://www.sciencedirect.com/science/article/pii/S1369526616301315 [visité le 29/07/2025].

Jin, Jian-Jun et al. [10 sept. 2020]. « GetOrganelle : a fast and versatile toolkit for accurate de novo assembly of organelle genomes ». In : *Genome Biology* 21.1, p. 241. ISSN : 1474-760X. DOI : 10.1186/s13059-020-02154-5. URL : https://doi.org/10.1186/s13059-020-02154-5 [visité le 26/02/2024].

Jing, Yanjun et Rongcheng Lin [sept. 2015]. « The VQ Motif-Containing Protein Family of Plant-Specific Transcriptional Regulators ». In : *Plant Physiology* 169.1, p. 371-378. ISSN : 1532-2548. DOI : 10.1104/pp.15.00788.

Jones, Craig E., Alfred L. Brown et Ute Baumann [22 mai 2007]. « Estimating the annotation error rate of curated GO database sequence annotations ». In : *BMC Bioinformatics* 8.1, p. 170. ISSN : 1471-2105. DOI : 10.1186/1471-2105-8-170. URL : https://doi.org/10.1186/1471-2105-8-170 [visité le 30/09/2025].

Jones, Philip, David Binns et al. [1er mai 2014]. « InterProScan 5 : genome-scale protein function classification ». In : *Bioinformatics* 30.9, p. 1236-1240. ISSN : 1367-4803. DOI : 10.1093/bioinformatics/btu031. URL : https://doi.org/10.1093/bioinformatics/btu031 [visité le 05/05/2025].

Joyce, Paul B.M. et Michael W. Gray [28 juill. 1989]. « Chloroplast-like transfer RNA genes expressed in wheat mitochondria ». In : *Nucleic Acids Research* 17.14, p. 5461-5476. ISSN : 0305-1048. DOI : 10.1093/nar/17.14.5461. URL : https://doi.org/10.1093/nar/17.14.5461 [visité le 17/09/2025].

Kabbage, Mehdi et al. [28 avr. 2017]. « The Life and Death of a Plant Cell ». In : *Annual Review of Plant Biology* 68 [Volume 68, 2017]. Publisher : Annual Reviews, p. 375-404. ISSN : 1543-5008, 1545-2123. DOI : 10.1146/annurev-arplant-043015-111655. URL : https://www.annualreviews.org/content/journals/10.1146/annurev-arplant-043015-111655 [visité le 17/09/2025].

Kalvari, Ioanna et al. [8 jan. 2021]. « Rfam 14 : expanded coverage of metagenomic, viral and microRNA families ». In : *Nucleic Acids Research* 49 [D1], p. D192-D200. ISSN : 0305-1048. DOI : 10.1093/nar/gkaa1047. URL : https://doi.org/10.1093/nar/gkaa1047 [visité le 26/02/2024].

Kanehisa, Minoru, Miho Furumichi et al. [6 jan. 2025]. « KEGG : biological systems database as a model of the real world ». In : *Nucleic Acids Research* 53 [D1], p. D672-

D677. ISSN : 1362-4962. DOI : 10.1093/nar/gkae909. URL : https://doi.org/10.1093/nar/gkae909 [visité le 01/08/2025].

Kanehisa, Minoru et Yoko Sato [jan. 2020]. « KEGG Mapper for inferring cellular functions from protein sequences ». In : *Protein Science : A Publication of the Protein Society* 29.1, p. 28-35. ISSN : 1469-896X. DOI : 10.1002/pro.3711.

Kanehisa, Minoru, Yoko Sato et Kanae Morishima [22 fév. 2016]. « BlastKOALA and GhostKOALA : KEGG Tools for Functional Characterization of Genome and Metagenome Sequences ». In : *Journal of Molecular Biology* 428.4, p. 726-731. ISSN : 1089-8638. DOI : 10.1016/j.jmb.2015.11.006.

Kang, Yunyan et al. [2 jan. 2023]. « Comprehensive Analysis of Major Latex-Like Protein Family Genes in Cucumber (Cucumis sativus L.) and Their Potential Roles in Phytophthora Blight Resistance ». In : *International Journal of Molecular Sciences* 24.1, p. 784. ISSN : 1422-0067. DOI : 10.3390/ijms24010784.

Karnik, Rucha et al. [avr. 2013]. « Arabidopsis Sec1/Munc18 protein SEC11 is a competitive and dynamic modulator of SNARE binding and SYP121-dependent vesicle traffic ». In : *The Plant Cell* 25.4, p. 1368-1382. ISSN : 1532-298X. DOI : 10.1105/tpc.112.108506.

Kchouk, Mehdi, Jean-Francois Gibrat et Mourad Elloumi [1er jan. 2017]. « Generations of Sequencing Technologies : From First to Next Generation ». In : *Biology and Medicine* 09. DOI : 10.4172/0974-8369.1000395.

Kellogg, Elizabeth A. et Jeffrey L. Bennetzen [2004]. « The evolution of nuclear genome structure in seed plants ». In : *American Journal of Botany* 91.10. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.3732/ajb.91.10.1709, p. 1709-1725. ISSN : 1537-2197. DOI : 10.3732/ajb.91.10.1709. URL : https://onlinelibrary.wiley.com/doi/abs/10.3732/ajb.91.10.1709 [visité le 08/04/2024].

Kersey, Paul Julian [1er avr. 2019]. « Plant genome sequences : past, present, future ». In : *Current Opinion in Plant Biology*. Genome studies and molecular genetics 48, p. 1-8. ISSN : 1369-5266. DOI : 10.1016/j.pbi.2018.11.001. URL : https://www.sciencedirect.com/science/article/pii/S1369526618300918 [visité le 29/07/2025].

Kim, Dong-Hwan, Yanpeng Xi et Sibum Sung [31 juill. 2017]. « Modular function of long noncoding RNA, COLDAIR, in the vernalization response ». In : *PLoS Genetics* 13.7, e1006939. ISSN : 1553-7390. DOI : 10.1371/journal.pgen.1006939. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC5552341/ [visité le 18/09/2025].

Kim, E. S., N. L. Bol'sheva et al. [nov. 2009]. « [The unique genome of two-chromosome grasses Zingeria and Colpodium, its origin, and evolution] ». In : *Genetika* 45.11, p. 1506-1515. ISSN : 0016-6758.

Kirchhoff, Helmut [2019]. « Chloroplast ultrastructure in plants ». In : *New Phytologist* 223.2. _eprint : https ://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/nph.15730, p. 565-574. ISSN : 1469-8137. DOI : 10.1111/nph.15730. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.15730 [visité le 16/06/2025].

Koga, Hiroyuki, Shuka Ikematsu et Seisuke Kimura [22 juill. 2024]. « Diving into the Water : Amphibious Plants as a Model for Investigating Plant Adaptations to Aquatic Environments ». In : *Annual Review of Plant Biology* 75 [Volume 75, 2024]. Publisher : Annual Reviews, p. 579-604. ISSN : 1543-5008, 1545-2123. DOI : 10.1146/annurev-arplant-062923-024919. URL : https://www.annualreviews.org/content/journals/10.1146/annurev-arplant-062923-024919 [visité le 30/09/2025].

Kolmogorov, Mikhail et al. [mai 2019]. « Assembly of long, error-prone reads using repeat graphs ». In : *Nature Biotechnology* 37.5. Number : 5 Publisher : Nature Publishing Group, p. 540-546. ISSN : 1546-1696. DOI : 10.1038/s41587-019-0072-

8. URL : https://www.nature.com/articles/s41587-019-0072-8 [visité le 26/02/2024].

Kong, Lingfei et al. [déc. 2023]. « The AP2/ERF transcription factor PtoERF15 confers drought tolerance via JA-mediated signaling in Populus ». In : *The New Phytologist* 240.5, p. 1848-1867. ISSN : 1469-8137. DOI : 10.1111/nph.19251.

Koren, Sergey et al. [5 jan. 2017]. « Canu : scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation ». In : *Genome Research* 27.5. Company : Cold Spring Harbor Laboratory Press Distributor : Cold Spring Harbor Laboratory Press Institution : Cold Spring Harbor Laboratory Press Label : Cold Spring Harbor Laboratory Press Publisher : Cold Spring Harbor Lab, p. 722-736. ISSN : 1088-9051, 1549-5469. DOI : 10.1101/gr.215087.116. URL : http://genome.cshlp.org/content/27/5/722 [visité le 10/07/2025].

Korf, Ian [14 mai 2004]. « Gene finding in novel genomes ». In : *BMC Bioinformatics* 5.1, p. 59. ISSN : 1471-2105. DOI : 10.1186/1471-2105-5-59. URL : https://doi.org/10.1186/1471-2105-5-59 [visité le 21/09/2025].

Kozik, Alexander et al. [30 août 2019]. « The alternative reality of plant mitochondrial DNA : One ring does not rule them all ». In : *PLOS Genetics* 15.8. Sous la dir. de Nathan M. Springer, e1008373. ISSN : 1553-7404. DOI : 10.1371/journal.pgen.1008373. URL : https://dx.plos.org/10.1371/journal.pgen.1008373 [visité le 18/01/2022].

Kozlov, Max [31 mai 2024]. « Biggest genome ever found belongs to this odd little plant ». In : *Nature*. Bandiera_abtest : a Cg_type : News Publisher : Nature Publishing Group Subject_term : Genomics, Plant sciences, Evolution. DOI : 10.1038/d41586-024-01567-7. URL : https://www.nature.com/articles/d41586-024-01567-7 [visité le 18/09/2025].

Kreiner, Julia M. et al. [27 déc. 2023]. « Quantifying the role of genome size and repeat content in adaptive variation and the architecture of flowering time in Amaranthus tuberculatus ». In : *PLOS Genetics* 19.12. Publisher : Public Library of Science, e1010865. ISSN : 1553-7404. DOI : 10.1371/journal.pgen.1010865. URL : https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010865 [visité le 26/09/2025].

Kress, W. John et al. [25 jan. 2022]. « Green plant genomes : What we know in an era of rapidly expanding opportunities ». In : *Proceedings of the National Academy of Sciences* 119.4. Publisher : Proceedings of the National Academy of Sciences, e2115640118. DOI : 10.1073/pnas.2115640118. URL : https://www.pnas.org/doi/abs/10.1073/pnas.2115640118 [visité le 29/07/2025].

Kumar, Suresh et Trilochan Mohapatra [21 mai 2021]. « Dynamics of DNA Methylation and Its Functions in Plant Growth and Development ». In : *Frontiers in Plant Science* 12. Publisher : Frontiers. ISSN : 1664-462X. DOI : 10.3389/fpls.2021.596236. URL : https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2021.596236/full [visité le 28/07/2025].

Kumawat, Surbhi et Jae Young Choi [2023]. « No end in sight : Mysteries of the telomeric variation in plants ». In : *American Journal of Botany* 110.11. _eprint : https ://bsapubs.onlinelibrary.wiley.com/doi/pdf/10.1002/ajb2.16244, e16244. ISSN : 1537-2197. DOI : 10.1002/ajb2.16244. URL : https://onlinelibrary.wiley.com/doi/abs/10.1002/ajb2.16244 [visité le 18/09/2025].

Kurtz, S et C Schleiermacher [1er mai 1999]. « REPuter : fast computation of maximal repeats in complete genomes. » In : *Bioinformatics* 15.5, p. 426-427. ISSN : 1367-4803. DOI : 10.1093/bioinformatics/15.5.426. URL : https://doi.org/10.1093/bioinformatics/15.5.426 [visité le 26/02/2024].

Kurusu, Takamitsu et al. [2018]. « Involvement of S-type anion channels in disease resistance against an oomycete pathogen in Arabidopsis seedlings ». In : *Communicative & Integrative Biology* 11.3, p. 1-6. ISSN : 1942-0889. DOI : 10.1080/19420889.2018.1495007.

Kyriakidou, Maria et al. [21 nov. 2018]. « Current Strategies of Polyploid Plant Genome Sequence Assembly ». In : *Frontiers in Plant Science* 9. Publisher : Frontiers. ISSN : 1664-462X. DOI : 10.3389/fpls.2018.01660. URL : https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2018.01660/full [visité le 29/07/2025].

Laetsch, Dominik R. et Mark L. Blaxter [31 juill. 2017]. *BlobTools : Interrogation of genome assemblies*. DOI : 10.12688/f1000research.12232.1. URL : https://f1000research.com/articles/6-1287 [visité le 21/09/2025].

Laforest, Larissa C et al. [1er sept. 2025]. « A chromosome-level Mitragyna parvifolia genome unveils spirooxindole alkaloid diversification and mitraphylline biosynthesis ». In : *The Plant Cell* 37.9, koaf207. ISSN : 1040-4651. DOI : 10.1093/plcell/koaf207. URL : https://doi.org/10.1093/plcell/koaf207 [visité le 24/09/2025].

Langmead, Ben et Steven L. Salzberg [avr. 2012]. « Fast gapped-read alignment with Bowtie 2 ». In : *Nature Methods* 9.4. Publisher : Nature Publishing Group, p. 357-359. ISSN : 1548-7105. DOI : 10.1038/nmeth.1923. URL : https://www.nature.com/articles/nmeth.1923 [visité le 03/12/2024].

Law, Julie A. et Steven E. Jacobsen [mars 2010]. « Establishing, maintaining and modifying DNA methylation patterns in plants and animals ». In : *Nature reviews. Genetics* 11.3, p. 204-220. ISSN : 1471-0056. DOI : 10.1038/nrg2719. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC3034103/ [visité le 18/09/2025].

Lee, Sung-Gwon, Dokyun Na et Chungoo Park [21 oct. 2021]. « Comparability of reference-based and reference-free transcriptome analysis approaches at the gene

expression level ». In : *BMC Bioinformatics* 22.11, p. 310. ISSN : 1471-2105. DOI : 10.1186/s12859-021-04226-0. URL : https://doi.org/10.1186/s12859-021-04226-0 [visité le 29/09/2025].

legifrance [2 mai 2007]. *Arrêté du 2 mai 2007 interdisant la commercialisation, l'utilisation et l'introduction dans le milieu naturel de Ludwigia grandiflora et Ludwigia peploides*. [Visité le 30/09/2025].

Leivar, Pablo et Peter H. Quail [jan. 2011]. « PIFs : pivotal components in a cellular signaling hub ». In : *Trends in Plant Science* 16.1, p. 19-28. ISSN : 1878-4372. DOI : 10.1016/j.tplants.2010.08.003.

Leushkin, Evgeny V et al. [15 juill. 2013]. « The miniature genome of a carnivorous plant Genlisea aurea contains a low number of genes and short non-coding sequences ». In : *BMC Genomics* 14, p. 476. ISSN : 1471-2164. DOI : 10.1186/1471-2164-14-476. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC3728226/ [visité le 18/09/2025].

Lezhneva, Lina et al. [oct. 2014]. « The Arabidopsis nitrate transporter NRT2.5 plays a role in nitrate acquisition and remobilization in nitrogen-starved plants ». In : *The Plant Journal : For Cell and Molecular Biology* 80.2, p. 230-241. ISSN : 1365-313X. DOI : 10.1111/tpj.12626.

Li, Dinghua, Chi-Man Liu et al. [15 mai 2015]. « MEGAHIT : an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph ». In : *Bioinformatics* 31.10, p. 1674-1676. ISSN : 1367-4803. DOI : 10.1093/bioinformatics/btv033. URL : https://doi.org/10.1093/bioinformatics/btv033 [visité le 19/05/2025].

Li, Gaojie, Shiqi Hu et al. [oct. 2019]. « Heterophylly : Phenotypic Plasticity of Leaf Shape in Aquatic and Amphibious Plants ». In : *Plants* 8.10. Number : 10 Publisher : Multidisciplinary Digital Publishing Institute, p. 420. ISSN : 2223-7747. DOI : 10.

3390/plants8100420. URL : https://www.mdpi.com/2223-7747/8/10/420 [visité le 22/05/2025].

Li, Gaojie, Xuyao Zhao et al. [2024]. « Water wisteria genome reveals environmental adaptation and heterophylly regulation in amphibious plants ». In : *Plant, Cell & Environment* 47.12. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/pce.15050, p. 4720-4740. ISSN : 1365-3040. DOI : 10.1111/pce.15050. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/pce.15050 [visité le 21/05/2025].

Li, Heng [15 sept. 2018]. « Minimap2 : pairwise alignment for nucleotide sequences ». In : *Bioinformatics* 34.18, p. 3094-3100. ISSN : 1367-4803, 1367-4811. DOI : 10.1093/bioinformatics/bty191. arXiv : 1708.01492[q-bio]. URL : http://arxiv.org/abs/1708.01492 [visité le 05/05/2025].

Li, Weizhong et Adam Godzik [1er juill. 2006]. « Cd-hit : a fast program for clustering and comparing large sets of protein or nucleotide sequences ». In : *Bioinformatics (Oxford, England)* 22.13, p. 1658-1659. ISSN : 1367-4803. DOI : 10.1093/bioinformatics/btl158.

Li, Xiangnan, Peipei Zhang et al. [3 août 2022]. « Genes expressed at low levels raise false discovery rates in RNA samples contaminated with genomic DNA ». In : *BMC Genomics* 23.1, p. 554. ISSN : 1471-2164. DOI : 10.1186/s12864-022-08785-1. URL : https://doi.org/10.1186/s12864-022-08785-1 [visité le 09/10/2025].

Li, Zhenyu, Yanxiang Chen et al. [1er jan. 2012]. « Comparison of the two major classes of assembly algorithms : overlap–layout–consensus and de-bruijn-graph ». In : *Briefings in Functional Genomics* 11.1, p. 25-37. ISSN : 2041-2649. DOI : 10.1093/bfgp/elr035. URL : https://doi.org/10.1093/bfgp/elr035 [visité le 20/09/2025].

Liao, Xiaofang et al. [24 août 2018]. « Complete sequence of kenaf (Hibiscus cannabinus) mitochondrial genome and comparative analysis with the mitochondrial genomes of

other plants ». In : *Scientific Reports* 8.1, p. 12714. ISSN : 2045-2322. DOI : 10.1038/s41598-018-30297-w.

Lima, Leandro et al. [22 fév. 2017]. « Playing hide and seek with repeats in local and global de novo transcriptome assembly of short RNA-seq reads ». In : *Algorithms for Molecular Biology* 12.1, p. 2. ISSN : 1748-7188. DOI : 10.1186/s13015-017-0091-2. URL : https://doi.org/10.1186/s13015-017-0091-2 [visité le 09/10/2025].

Lindemann-Matthies, Petra [16 mars 2016]. « Beasts or beauties ? Laypersons' perception of invasive alien plant species in Switzerland and attitudes towards their management ». In : *NeoBiota* 29. Publisher : Pensoft Publishers, p. 15-33. ISSN : 1314-2488. DOI : 10.3897/neobiota.29.5786. URL : https://neobiota.pensoft.net/article/5786/ [visité le 03/05/2024].

Litholdo, Celso Gaspar et Guilherme Cordenonsi da Fonseca [2018]. « Circular RNAs and Plant Stress Responses ». In : *Circular RNAs : Biogenesis and Functions*. Sous la dir. de Junjie Xiao. Singapore : Springer, p. 345-353. ISBN : 978-981-13-1426-1. DOI : 10.1007/978-981-13-1426-1_27. URL : https://doi.org/10.1007/978-981-13-1426-1_27 [visité le 18/09/2025].

Liu, Hui, Hai-Meng Lyu et al. [fév. 2021]. « The emergence and evolution of intron-poor and intronless genes in intron-rich plant gene families ». In : *The Plant Journal : For Cell and Molecular Biology* 105.4, p. 1072-1082. ISSN : 1365-313X. DOI : 10.1111/tpj.15088.

Liu, Jingying, Anne Osbourn et Pengda Ma [mai 2015]. « MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants ». In : *Molecular Plant* 8.5, p. 689-708. ISSN : 1752-9867. DOI : 10.1016/j.molp.2015.03.012.

Liu, Juntao, Ting Yu et al. [23 avr. 2019]. « TransLiG : a de novo transcriptome assembler that uses line graph iteration ». In : *Genome Biology* 20.1, p. 81. ISSN :

1474-760X. DOI : 10.1186/s13059-019-1690-7. URL : https://doi.org/10.1186/s13059-019-1690-7 [visité le 04/08/2025].

Liu, Shih-Hui, Peter C. Hoch et al. [2017]. « Multi-locus phylogeny of Ludwigia (Onagraceae) : Insights on infra- generic relationships and the current classification of the genus ». In : *TAXON* 66.5. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.12705/665.7, p. 1112-1127. ISSN : 1996-8175. DOI : 10.12705/665.7. URL : https://onlinelibrary.wiley.com/doi/abs/10.12705/665.7 [visité le 09/10/2025].

Liu, Yuanlong, Chong Teng et al. [oct. 2020]. « PhasiRNAs in Plants : Their Biogenesis, Genic Sources, and Roles in Stress Responses, Development, and Reproduction ». In : *The Plant Cell* 32.10, p. 3059-3080. ISSN : 1040-4651. DOI : 10.1105/tpc.20.00335. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC7534485/ [visité le 18/09/2025].

Logsdon, Glennis A., Mitchell R. Vollger et Evan E. Eichler [oct. 2020]. « Long-read human genome sequencing and its applications ». In : *Nature reviews. Genetics* 21.10, p. 597-614. ISSN : 1471-0056. DOI : 10.1038/s41576-020-0236-x. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC7877196/ [visité le 07/10/2025].

Loreto, Elgion L. S. et al. [19 fév. 2024]. « The good, the bad and the ugly of transposable elements annotation tools ». In : *Genetics and Molecular Biology* 46. Publisher : Sociedade Brasileira de Genética, e20230138. ISSN : 1415-4757, 1678-4685. DOI : https://doi.org/10.1590/1678-4685-GMB-2023-0138. URL : https://www.scielo.br/j/gmb/a/yv8YfNbKqwLm8CYcpjtFmWK/?lang=en [visité le 01/08/2025].

Love, Michael I., Wolfgang Huber et Simon Anders [5 déc. 2014]. « Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 ». In : *Genome Biology* 15.12, p. 550. ISSN : 1474-760X. DOI : 10.1186/s13059-014-0550-8. URL : https://doi.org/10.1186/s13059-014-0550-8 [visité le 05/08/2025].

Lucek, Kay, Hannah Augustijnen et Marcial Escudero [1er août 2022]. « A holocentric twist to chromosomal speciation ? » In : *Trends in Ecology & Evolution* 37.8, p. 655-662. ISSN : 0169-5347. DOI : 10.1016/j.tree.2022.04.002. URL : https://www.sciencedirect.com/science/article/pii/S0169534722000854 [visité le 18/09/2025].

Luo, Ruibang et al. [27 déc. 2012]. « SOAPdenovo2 : an empirically improved memory-efficient short-read de novo assembler ». In : *GigaScience* 1, p. 18. ISSN : 2047-217X. DOI : 10.1186/2047-217X-1-18. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC3626529/ [visité le 20/09/2025].

Lupo, Valérian et al. [22 oct. 2021]. « Contamination in Reference Sequence Databases : Time for Divide-and-Rule Tactics ». In : *Frontiers in Microbiology* 12. Publisher : Frontiers. ISSN : 1664-302X. DOI : 10.3389/fmicb.2021.755101. URL : https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2021.755101/full [visité le 30/09/2025].

Ma, Qing-Hu [25 fév. 2024]. « Lignin Biosynthesis and Its Diversified Roles in Disease Resistance ». In : *Genes* 15.3, p. 295. ISSN : 2073-4425. DOI : 10.3390/genes15030295.

Macêdo, Rafael L. et al. [15 jan. 2024]. « The economic costs of invasive aquatic plants : A global perspective on ecology and management gaps ». In : *Science of The Total Environment* 908, p. 168217. ISSN : 0048-9697. DOI : 10.1016/j.scitotenv.2023.168217. URL : https://www.sciencedirect.com/science/article/pii/S0048969723068444 [visité le 30/09/2025].

Machaj, Gabriela et Dariusz Grzebelus [18 mai 2021]. « Characteristics of the AT-Hook Motif Containing Nuclear Localized (AHL) Genes in Carrot Provides Insight into Their Role in Plant Growth and Storage Root Development ». In : *Genes* 12.5, p. 764. ISSN : 2073-4425. DOI : 10.3390/genes12050764.

Mackenzie, S et L McIntosh [avr. 1999]. « Higher plant mitochondria ». In : *The Plant Cell* 11.4, p. 571-586. ISSN : 1040-4651. DOI : 10.1105/tpc.11.4.571. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC144202/ [visité le 17/09/2025].

Madson, Michael et al. [juill. 2003]. « The MUR3 gene of Arabidopsis encodes a xyloglucan galactosyltransferase that is evolutionarily related to animal exostosins ». In : *The Plant Cell* 15.7, p. 1662-1670. ISSN : 1040-4651. DOI : 10.1105/tpc.009837.

Mahajan, Shruti et al. [1er sept. 2023]. « Genome of Phyllanthus emblica : the medicinal plant Amla with super antioxidant properties ». In : *Frontiers in Plant Science* 14. Publisher : Frontiers. ISSN : 1664-462X. DOI : 10.3389/fpls.2023.1210078. URL : https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1210078/full [visité le 05/10/2025].

Maldonado, Jose A. et al. [15 juin 2020]. « Transcriptome sequencing reveals signatures of positive selection in the Spot-Tailed Earless Lizard ». In : *PLOS ONE* 15.6. Publisher : Public Library of Science, e0234504. ISSN : 1932-6203. DOI : 10.1371/journal.pone.0234504. URL : https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0234504 [visité le 26/09/2025].

Mandon, E. C. et al. [5 juin 1992]. « Subcellular localization and membrane topology of serine palmitoyltransferase, 3-dehydrosphinganine reductase, and sphinganine N-acyltransferase in mouse liver ». In : *The Journal of Biological Chemistry* 267.16, p. 11144-11148. ISSN : 0021-9258.

Manni, Mosè et al. [1er oct. 2021]. « BUSCO Update : Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes ». In : *Molecular Biology and Evolution* 38.10, p. 4647-4654. ISSN : 1537-1719. DOI : 10.1093/molbev/msab199. URL : https://doi.org/10.1093/molbev/msab199 [visité le 03/12/2024].

Marechal-Drouard, Laurence et al. [1er jan. 1995]. « [23] Transfer RNA import into plant mitochondria ». In : *Methods in Enzymology*. T. 260. Mitochondrial Biogenesis and Genetics Part A. Academic Press, p. 310-327. DOI : 10.1016/0076-6879(95)60148-1. URL : https://www.sciencedirect.com/science/article/pii/0076687995601481 [visité le 17/09/2025].

Marienfeld, Joachim, Michael Unseld et Axel Brennicke [1er déc. 1999]. « The mitochondrial genome of Arabidopsis is composed of both native and immigrant information ». In : *Trends in Plant Science* 4.12. Publisher : Elsevier, p. 495-502. ISSN : 1360-1385. DOI : 10.1016/S1360-1385(99)01502-2. URL : https://www.cell.com/trends/plant-science/abstract/S1360-1385(99)01502-2 [visité le 01/02/2024].

Marin, Pierre et al. [2020]. « Biological invasion : The influence of the hidden side of the (epi)genome ». In : *Functional Ecology* 34.2. _eprint : https://onlinelibrary.wiley.com/doi/pdf/10.1112435.13317, p. 385-400. ISSN : 1365-2435. DOI : 10.1111/1365-2435.13317. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2435.13317 [visité le 23/02/2024].

Marks, Rose A. et al. [déc. 2021]. « Representation and participation across 20 years of plant genome sequencing ». In : *Nature Plants* 7.12. Publisher : Nature Publishing Group, p. 1571-1578. ISSN : 2055-0278. DOI : 10.1038/s41477-021-01031-8. URL : https://www.nature.com/articles/s41477-021-01031-8 [visité le 17/09/2025].

Marowa, Prince, Anming Ding et Yingzhen Kong [mai 2016]. « Expansins : roles in plant growth and potential applications in crop improvement ». In : *Plant Cell Reports* 35.5, p. 949-965. ISSN : 1432-203X. DOI : 10.1007/s00299-016-1948-4.

Marsberg, Angelica et al. [2017]. « Botryosphaeria dothidea : a latent pathogen of global importance to woody plant health ». In : *Molecular Plant Pathology* 18.4. _eprint : https://onlinelibrary.wiley.com/doi/pdf/10.1111/mpp.12495, p. 477-488.

ISSN : 1364-3703. DOI : 10.1111/mpp.12495. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/mpp.12495 [visité le 23/02/2024].

Martin, Marcel [2 mai 2011]. « Cutadapt removes adapter sequences from high-throughput sequencing reads ». In : *EMBnet.journal* 17.1. Number : 1, p. 10-12. ISSN : 2226-6089. DOI : 10.14806/ej.17.1.200. URL : https://journal.embnet.org/index.php/embnetjournal/article/view/200 [visité le 12/06/2024].

Martin, W. et R. G. Herrmann [sept. 1998]. « Gene Transfer from Organelles to the Nucleus : How Much, What Happens, and Why ? » In : *Plant Physiology* 118.1, p. 9-17. ISSN : 0032-0889. DOI : 10.1104/pp.118.1.9. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC1539188/ [visité le 18/09/2025].

Matthew, Louisa et al. [28 sept. 2022]. « DEFECTIVE EMBRYO AND MERISTEMS1 (DEM1) Is Essential for Cell Proliferation and Cell Differentiation in Tomato ». In : *Plants (Basel, Switzerland)* 11.19, p. 2545. ISSN : 2223-7747. DOI : 10.3390/plants11192545.

Maurin, Olivier et al. [2021]. « A nuclear phylogenomic study of the angiosperm order Myrtales, exploring the potential and limitations of the universal Angiosperms353 probe set ». In : *American Journal of Botany* 108.7. _eprint : https ://bsapubs.onlinelibrary.wiley.com/ p. 1087-1111. ISSN : 1537-2197. DOI : 10.1002/ajb2.1699. URL : https://onlinelibrary.wiley.com/doi/abs/10.1002/ajb2.1699 [visité le 28/09/2025].

McCormick, Ryan F. et al. [2018]. « The Sorghum bicolor reference genome : improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization ». In : *The Plant Journal* 93.2. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/tpj.1 p. 338-354. ISSN : 1365-313X. DOI : 10.1111/tpj.13781. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.13781 [visité le 04/10/2025].

McGaughran, Angela et al. [1er jan. 2024]. « Genomic Tools in Biological Invasions : Current State and Future Frontiers ». In : *Genome Biology and Evolution* 16.1,

evad230. ISSN : 1759-6653. DOI : 10.1093/gbe/evad230. URL : https://doi.org/10.1093/gbe/evad230 [visité le 13/08/2025].

Meng, Guanliang et al. [20 juin 2019]. « MitoZ : a toolkit for animal mitochondrial genome assembly, annotation and visualization ». In : *Nucleic Acids Research* 47.11, e63. ISSN : 0305-1048. DOI : 10.1093/nar/gkz173. URL : https://doi.org/10.1093/nar/gkz173 [visité le 26/02/2024].

Menzel, Peter, Kim Lee Ng et Anders Krogh [13 avr. 2016]. « Fast and sensitive taxonomic classification for metagenomics with Kaiju ». In : *Nature Communications* 7.1. Publisher : Nature Publishing Group, p. 11257. ISSN : 2041-1723. DOI : 10.1038/ncomms11257. URL : https://www.nature.com/articles/ncomms11257 [visité le 29/07/2025].

Meyer, Corentin et al. [10 nov. 2020]. « Understanding the causes of errors in eukaryotic protein-coding gene prediction : a case study of primate proteomes ». In : *BMC Bioinformatics* 21.1, p. 513. ISSN : 1471-2105. DOI : 10.1186/s12859-020-03855-1. URL : https://doi.org/10.1186/s12859-020-03855-1 [visité le 30/09/2025].

Mohanta, Tapan Kumar et al. [19 nov. 2020]. « Construction of anti-codon table of the plant kingdom and evolution of tRNA selenocysteine (tRNASec) ». In : *BMC Genomics* 21.1, p. 804. ISSN : 1471-2164. DOI : 10.1186/s12864-020-07216-3. URL : https://doi.org/10.1186/s12864-020-07216-3 [visité le 09/07/2025].

Molina, A., A. Segura et F. García-Olmedo [25 jan. 1993]. « Lipid transfer proteins (nsLTPs) from barley and maize leaves are potent inhibitors of bacterial and fungal plant pathogens ». In : *FEBS letters* 316.2, p. 119-122. ISSN : 0014-5793. DOI : 10.1016/0014-5793(93)81198-9.

Møller, Ian Max, Allan G. Rasmusson et Olivier Van Aken [2021]. « Plant mitochondria – past, present and future ». In : *The Plant Journal* 108.4. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/tpj.15495, p. 912-959. ISSN : 1365-313X. DOI :

10.1111/tpj.15495. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.15495 [visité le 23/09/2025].

Moore, Lisa D., Thuc Le et Guoping Fan [jan. 2013]. « DNA Methylation and Its Basic Function ». In : *Neuropsychopharmacology* 38.1. Publisher : Nature Publishing Group, p. 23-38. ISSN : 1740-634X. DOI : 10.1038/npp.2012.112. URL : https://www.nature.com/articles/npp2012112 [visité le 28/07/2025].

Morand, Serge [2017]. « Infections and Diseases in Wildlife by Non-native Organisms ». In : *Impact of Biological Invasions on Ecosystem Services*. Sous la dir. de Montserrat Vilà et Philip E. Hulme. Invading Nature - Springer Series in Invasion Ecology. Cham : Springer International Publishing, p. 177-190. ISBN : 978-3-319-45121-3. DOI : 10.1007/978-3-319-45121-3_11. URL : https://doi.org/10.1007/978-3-319-45121-3_11 [visité le 23/02/2024].

Morello, L. et D. Breviario [1er juin 2008]. « Plant Spliceosomal Introns : Not Only Cut and Paste ». In : *Current Genomics* 9.4, p. 227-238. ISSN : 13892029. DOI : 10.2174/138920208784533629. URL : http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1389-2029&volume=9&issue=4&spage=227 [visité le 16/05/2025].

Morley, Stewart A., Niaz Ahmad et Brent L. Nielsen [oct. 2019]. « Plant Organelle Genome Replication ». In : *Plants* 8.10. Publisher : Multidisciplinary Digital Publishing Institute, p. 358. ISSN : 2223-7747. DOI : 10.3390/plants8100358. URL : https://www.mdpi.com/2223-7747/8/10/358 [visité le 09/10/2025].

Mounger, Jeannie et al. [19 avr. 2021]. « Epigenetics and the success of invasive plants ». In : *Philosophical Transactions of the Royal Society B : Biological Sciences* 376.1826. Publisher : Royal Society, p. 20200117. DOI : 10.1098/rstb.2020.0117. URL : https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0117 [visité le 26/09/2025].

Mower, Jeffrey P., Daniel B. Sloan et Andrew J. Alverson [2012]. « Plant Mitochondrial Genome Diversity : The Genomics Revolution ». In : *Plant Genome Diversity Volume 1*. Sous la dir. de Jonathan F. Wendel et al. Vienna : Springer Vienna, p. 123-144. ISBN : 978-3-7091-1129-1 978-3-7091-1130-7. DOI : 10.1007/978-3-7091-1130-7_9. URL : http://link.springer.com/10.1007/978-3-7091-1130-7_9 [visité le 18/01/2022].

Müller, Maren et Sergi Munné-Bosch [sept. 2015]. « Ethylene Response Factors : A Key Regulatory Hub in Hormone and Stress Signaling ». In : *Plant Physiology* 169.1, p. 32-41. ISSN : 1532-2548. DOI : 10.1104/pp.15.00677.

Nedelyaeva, Olga I. et al. [20 déc. 2024]. « Functional and Molecular Characterization of Plant Nitrate Transporters Belonging to NPF (NRT1/PTR) 6 Subfamily ». In : *International Journal of Molecular Sciences* 25.24, p. 13648. ISSN : 1422-0067. DOI : 10.3390/ijms252413648.

Nehring, Stefan et Detlef Kolthoff [mars 2011]. « The invasive water primrose Ludwigia grandiflora (Michaux) Greuter & Burdet (Spermatophyta : Onagraceae) in Germany : First record and ecological risk assessment ». In : *Aquatic Invasions* 6.1, p. 83-89. ISSN : 18185487. DOI : 10.3391/ai.2011.6.1.10. URL : http://www.aquaticinvasions.net/2011/issue1.html [visité le 17/05/2024].

Nishimura, Darryl [avr. 2000]. « RepeatMasker ». In : *Biotech Software & Internet Report* 1.1. Publisher : Mary Ann Liebert, Inc., publishers, p. 36-39. ISSN : 1527-9162. DOI : 10.1089/152791600319259. URL : https://www.liebertpub.com/doi/abs/10.1089/152791600319259 [visité le 21/09/2025].

Norton, Benjamin B. et Scott A. Norton [1er jan. 2021]. « Lionfish envenomation in Caribbean and Atlantic waters : Climate change and invasive species ». In : *International Journal of Women's Dermatology*. Special Issue on Climate Change & Dermatology 7.1, p. 120-123. ISSN : 2352-6475. DOI : 10.1016/j.ijwd.2020.05.016. URL :

https://www.sciencedirect.com/science/article/pii/S2352647520301106 [visité le 23/02/2024].

Nota, Alessandro et al. [2024]. « Adaptation to bioinvasions : When does it occur ? » In : *Global Change Biology* 30.6. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.17362, e17362. ISSN : 1365-2486. DOI : 10.1111/gcb.17362. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.17362 [visité le 21/09/2025].

Novák, Petr et al. [1er sept. 2024]. « DANTE and DANTE_LTR : lineage-centric annotation pipelines for long terminal repeat retrotransposons in plant genomes ». In : *NAR Genomics and Bioinformatics* 6.3, lqae113. ISSN : 2631-9268. DOI : 10.1093/nargab/lqae113. URL : https://doi.org/10.1093/nargab/lqae113 [visité le 07/05/2025].

Novoa, Ana et al. [1er mai 2020]. « Invasion syndromes : a systematic approach for predicting biological invasions and facilitating effective management ». In : *Biological Invasions* 22.5, p. 1801-1820. ISSN : 1573-1464. DOI : 10.1007/s10530-020-02220-w. URL : https://doi.org/10.1007/s10530-020-02220-w [visité le 26/09/2025].

Nunn, Adam et al. [1er sept. 2021]. « Comprehensive benchmarking of software for mapping whole genome bisulfite data : from read alignment to DNA methylation analysis ». In : *Briefings in Bioinformatics* 22.5, bbab021. ISSN : 1477-4054. DOI : 10.1093/bib/bbab021. URL : https://doi.org/10.1093/bib/bbab021 [visité le 04/10/2025].

O'Leary, Nuala A. et al. [4 jan. 2016]. « Reference sequence (RefSeq) database at NCBI : current status, taxonomic expansion, and functional annotation ». In : *Nucleic Acids Research* 44 [D1], p. D733-D745. ISSN : 0305-1048. DOI : 10.1093/nar/gkv1189. URL : https://doi.org/10.1093/nar/gkv1189 [visité le 26/02/2024].

O'Leary, Nuala A. et al. [5 juill. 2024]. « Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets ». In : *Scientific Data*

11.1. Publisher : Nature Publishing Group, p. 732. ISSN : 2052-4463. DOI : 10.1038/s41597-024-03571-y. URL : https://www.nature.com/articles/s41597-024-03571-y [visité le 28/09/2025].

Ou, Chia-Ying et al. [1er juin 2025]. « Novel perspectives on plastome evolution in Onagraceae ». In : *AoB PLANTS* 17.3, plaf025. ISSN : 2041-2851. DOI : 10.1093/aobpla/plaf025. URL : https://doi.org/10.1093/aobpla/plaf025 [visité le 09/10/2025].

Ounit, Rachid et al. [25 mars 2015]. « CLARK : fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers ». In : *BMC Genomics* 16.1, p. 236. ISSN : 1471-2164. DOI : 10.1186/s12864-015-1419-2. URL : https://doi.org/10.1186/s12864-015-1419-2 [visité le 21/09/2025].

Pagad, Shyama et al. [23 jan. 2018]. « Introducing the Global Register of Introduced and Invasive Species ». In : *Scientific Data* 5.1. Publisher : Nature Publishing Group, p. 170202. ISSN : 2052-4463. DOI : 10.1038/sdata.2017.202. URL : https://www.nature.com/articles/sdata2017202 [visité le 19/05/2025].

Paini, Dean R. et al. [5 juill. 2016]. « Global threat to agriculture from invasive species ». In : *Proceedings of the National Academy of Sciences* 113.27. Publisher : Proceedings of the National Academy of Sciences, p. 7575-7579. DOI : 10.1073/pnas.1602205113. URL : https://www.pnas.org/doi/full/10.1073/pnas.1602205113 [visité le 19/05/2025].

Pallares-Zazo, Jorge et al. [2025]. « Chromosome-level genome assembly and annotation of Ulmus minor reveal dynamic intrageneric clusters of resistance genes ». In : *The Plant Genome* 18.3. _eprint : https ://acsess.onlinelibrary.wiley.com/doi/pdf/10.1002/tpg2.70114, e70114. ISSN : 1940-3372. DOI : 10.1002/tpg2.70114. URL : https://onlinelibrary.wiley.com/doi/abs/10.1002/tpg2.70114 [visité le 24/09/2025].

Palmer, Jeffrey D. et Laura A. Herbo [1er mars 1987]. « Unicircular structure of the Brassica hirta mitochondrial genome ». In : *Current Genetics* 11.6, p. 565-570. ISSN : 1432-0983. DOI : 10.1007/BF00384620. URL : https://doi.org/10.1007/BF00384620 [visité le 25/02/2022].

Palmer, Jeffrey D. et Clark R. Shields [fév. 1984]. « Tripartite structure of the Brassica campestris mitochondrial genome ». In : *Nature* 307.5950. Publisher : Nature Publishing Group, p. 437-440. ISSN : 1476-4687. DOI : 10.1038/307437a0. URL : https://www.nature.com/articles/307437a0 [visité le 22/09/2025].

Pan, Ronghui et Jianping Hu [28 août 2015]. « Plant mitochondrial dynamics and the role of membrane lipids ». In : *Plant Signaling & Behavior* 10.10, e1050573. ISSN : 1559-2316. DOI : 10.1080/15592324.2015.1050573. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4883875/ [visité le 17/09/2025].

Pardo-Palacios, Francisco J. et al. [juill. 2024]. « Systematic assessment of long-read RNA-seq methods for transcript identification and quantification ». In : *Nature Methods* 21.7. Publisher : Nature Publishing Group, p. 1349-1363. ISSN : 1548-7105. DOI : 10.1038/s41592-024-02298-3. URL : https://www.nature.com/articles/s41592-024-02298-3 [visité le 22/09/2025].

Park, Hyeonseon et al. [31 juill. 2025]. « Chromosome-level Genome Assembly of the Halophytic Turfgrass Zoysia macrostachya ». In : *Scientific Data* 12.1. Publisher : Nature Publishing Group, p. 1335. ISSN : 2052-4463. DOI : 10.1038/s41597-025-05702-5. URL : https://www.nature.com/articles/s41597-025-05702-5 [visité le 24/09/2025].

Parmar, Rajni et al. [jan. 2022]. « Assembly and Annotation of Red Spruce (Picea rubens) Chloroplast Genome, Identification of Simple Sequence Repeats, and Phylogenetic Analysis in Picea ». In : *International Journal of Molecular Sciences* 23.23. Number : 23 Publisher : Multidisciplinary Digital Publishing Institute, p. 15243.

ISSN : 1422-0067. DOI : 10.3390/ijms232315243. URL : https://www.mdpi.com/1422-0067/23/23/15243 [visité le 30/07/2025].

Parra, Samuel et al. [23 sept. 2025]. « Decoding the Hayward kiwi (Actinidia deliciosa var Hayward) genome : transcriptomic responses to drought and salinity and AdhSAP4's role in salinity stress responses ». In : *Frontiers in Plant Science* 16. Publisher : Frontiers. ISSN : 1664-462X. DOI : 10.3389/fpls.2025.1637092. URL : https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2025.1637092/full [visité le 24/09/2025].

Patro, Rob et al. [avr. 2017]. « Salmon provides fast and bias-aware quantification of transcript expression ». In : *Nature Methods* 14.4. Publisher : Nature Publishing Group, p. 417-419. ISSN : 1548-7105. DOI : 10.1038/nmeth.4197. URL : https://www.nature.com/articles/nmeth.4197 [visité le 05/08/2025].

Pavese, Vera et al. [1er juill. 2021]. « Whole-genome assembly of Corylus avellana cv "Tonda Gentile delle Langhe" using linked-reads (10X Genomics) ». In : *G3 Genes|Genomes|Genetics* 11.7, jkab152. ISSN : 2160-1836. DOI : 10.1093/g3journal/jkab152. URL : https://doi.org/10.1093/g3journal/jkab152 [visité le 12/08/2025].

Peng, Gongxin, Peifeng Ji et Fangqing Zhao [17 nov. 2016]. « A novel codon-based de Bruijn graph algorithm for gene construction from unassembled transcriptomes ». In : *Genome Biology* 17.1, p. 232. ISSN : 1474-760X. DOI : 10.1186/s13059-016-1094-x. URL : https://doi.org/10.1186/s13059-016-1094-x [visité le 04/08/2025].

Perez-de-Castro, A. M. et al. [1er mai 2012]. « Application of Genomic Tools in Plant Breeding ». In : *Current Genomics* 13.3. Publisher : Bentham Science Publishers, p. 179-195. ISSN : 1389-2029, 1875-5488. DOI : 10.2174/138920212800543084.

URL : https://www.benthamdirect.com/content/journals/cg/10.2174/138920212800543084 [visité le 16/09/2025].

Petersen, G. et al. [2 déc. 2015]. « Massive gene loss in mistletoe (Viscum, Viscaceae) mitochondria ». In : *Scientific Reports* 5.1. Number : 1 Publisher : Nature Publishing Group, p. 17588. ISSN : 2045-2322. DOI : 10.1038/srep17588. URL : https://www.nature.com/articles/srep17588 [visité le 02/02/2024].

Pollitt, Alice Y. et Robert H. Insall [1er août 2009]. « WASP and SCAR/WAVE proteins : the drivers of actin assembly ». In : *Journal of Cell Science* 122 [Pt 15], p. 2575-2578. ISSN : 0021-9533. DOI : 10.1242/jcs.023879.

Pontius, Joan, Lukas Wagner et Gregory Schuler [1er jan. 2003]. « 21. UniGene : A Unified View of the Transcriptome ». In : *The NCBI Handbook* 1.

Portillo Lemus, Luis O. et al. [avr. 2021]. « Self-incompatibility limits sexual reproduction rather than environmental conditions in an invasive water primrose ». In : *Plant-Environment Interactions* 2.2. Publisher : John Wiley & Sons, Ltd, p. 74-86. DOI : 10.1002/pei3.10042. URL : https://onlinelibrary.wiley.com/doi/full/10.1002/pei3.10042 [visité le 06/08/2025].

POWO [2025]. *Plants of the World Online*. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. URL : https://powo.science.kew.org/ [visité le 30/09/2025].

Preuten, Tobias et al. [2010]. « Fewer genes than organelles : extremely low and variable gene copy numbers in mitochondria of somatic plant cells ». In : *The Plant Journal* 64.6. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-313X.2010.04389.x, p. 948-959. ISSN : 1365-313X. DOI : 10.1111/j.1365-313X.2010.04389.x. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-313X.2010.04389.x [visité le 02/02/2024].

Privitera, Grete Francesca et al. [1er déc. 2024]. « Comparative transcriptome analysis of *B. oleracea* L. var. italica and *B. macrocarpa* Guss. genotypes under drought stress : *de novo* vs reference genome assembly ». In : *Plant Stress* 14, p. 100657. ISSN : 2667-064X. DOI : 10.1016/j.stress.2024.100657. URL : https://www.sciencedirect.com/science/article/pii/S2667064X24003105 [visité le 06/10/2025].

Prjibelski, Andrey et al. [2020]. « Using SPAdes De Novo Assembler ». In : *Current Protocols in Bioinformatics* 70.1. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/cpbi.102, e102. ISSN : 1934-340X. DOI : 10.1002/cpbi.102. URL : https://onlinelibrary.wiley.com/doi/abs/10.1002/cpbi.102 [visité le 26/02/2024].

Prosdocimi, Francisco et al. [4 jan. 2012]. « Controversies in modern evolutionary biology : the imperative for error detection and quality control ». In : *BMC Genomics* 13.1, p. 5. ISSN : 1471-2164. DOI : 10.1186/1471-2164-13-5. URL : https://doi.org/10.1186/1471-2164-13-5 [visité le 30/09/2025].

Protasoni, Margherita et Massimo Zeviani [jan. 2021]. « Mitochondrial Structure and Bioenergetics in Normal and Disease Conditions ». In : *International Journal of Molecular Sciences* 22.2. Publisher : Multidisciplinary Digital Publishing Institute, p. 586. ISSN : 1422-0067. DOI : 10.3390/ijms22020586. URL : https://www.mdpi.com/1422-0067/22/2/586 [visité le 17/09/2025].

Putri, Givanna H et al. [13 mai 2022]. « Analysing high-throughput sequencing data in Python with HTSeq 2.0 ». In : *Bioinformatics* 38.10, p. 2943-2945. ISSN : 1367-4803. DOI : 10.1093/bioinformatics/btac166. URL : https://doi.org/10.1093/bioinformatics/btac166 [visité le 26/02/2024].

Putterill, J. et al. [24 mars 1995]. « The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors ». In : *Cell* 80.6, p. 847-857. ISSN : 0092-8674. DOI : 10.1016/0092-8674(95)90288-0.

Pyšek, Petr, Philip E. Hulme et al. [2020]. « Scientists' warning on invasive alien species ». In : *Biological Reviews* 95.6. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/brv.126 p. 1511-1534. ISSN : 1469-185X. DOI : 10.1111/brv.12627. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/brv.12627 [visité le 19/05/2025].

Pyšek, Petr, Magdalena Lučanová et al. [2023]. « Small genome size and variation in ploidy levels support the naturalization of vascular plants but constrain their invasive spread ». In : *New Phytologist* 239.6. _eprint : https ://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/n p. 2389-2403. ISSN : 1469-8137. DOI : 10.1111/nph.19135. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.19135 [visité le 26/09/2025].

Raghavan, Venket et al. [1er mars 2022]. « A simple guide to de novo transcriptome assembly and annotation ». In : *Briefings in Bioinformatics* 23.2, bbab563. ISSN : 1477-4054. DOI : 10.1093/bib/bbab563. URL : https://doi.org/10.1093/bib/bbab563 [visité le 04/08/2025].

Rai, Prabhat Kumar [1er mars 2022]. « Environmental Degradation by Invasive Alien Plants in the Anthropocene : Challenges and Prospects for Sustainable Restoration ». In : *Anthropocene Science* 1.1, p. 5-28. ISSN : 2731-3980. DOI : 10.1007/s44177-021-00004-y. URL : https://doi.org/10.1007/s44177-021-00004-y [visité le 06/05/2024].

Ramakrishnan, Muthusamy et al. [jan. 2021]. « The Dynamism of Transposon Methylation for Plant Development and Stress Adaptation ». In : *International Journal of Molecular Sciences* 22.21. Publisher : Multidisciplinary Digital Publishing Institute, p. 11387. ISSN : 1422-0067. DOI : 10.3390/ijms222111387. URL : https://www.mdpi.com/1422-0067/22/21/11387 [visité le 18/09/2025].

Ramos, Miguel Jesus Nunes et al. [jan. 2017]. « Deep analysis of wild Vitis flower transcriptome reveals unexplored genome regions associated with sex specification ».

In : *Plant Molecular Biology* 93.1, p. 151-170. ISSN : 1573-5028. DOI : 10.1007/s11103-016-0553-9.

Raven, John A et John F Allen [2003]. « Genomics and chloroplast evolution : what did cyanobacteria do for plants ? » In : *Genome Biology* 4.3, p. 209. ISSN : 1474-7596. DOI : 10.1186/gb-2003-4-3-209. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC153454/ [visité le 17/09/2025].

Razali, Rozaimi et al. [4 oct. 2018]. « The Genome Sequence of the Wild Tomato Solanum pimpinellifolium Provides Insights Into Salinity Tolerance ». In : *Frontiers in Plant Science* 9, p. 1402. ISSN : 1664-462X. DOI : 10.3389/fpls.2018.01402. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC6186997/ [visité le 19/09/2025].

Reatini, Bryan et al. [27 nov. 2024]. « Chromosome-scale Reference Genome and RAD-based Genetic Map of Yellow Starthistle (Centaurea solstitialis) Reveal Putative Structural Variation and QTL Associated With Invader Traits ». In : *Genome Biology and Evolution* 16.12, evae243. ISSN : 1759-6653. DOI : 10.1093/gbe/evae243. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC11632367/ [visité le 26/09/2025].

Reddy, Angelica M. et al. [juill. 2021]. « Biological control of invasive water primroses, Ludwigia spp., in the United States : A feasibility assessment ». In : Accepted : 2024-02-28T13 :04 :33Z Publisher : Aquatic Plant Management Soc, Inc. ISSN : 0146-6623. URL : https://ri.conicet.gov.ar/handle/11336/228776 [visité le 06/08/2025].

Ren, Hong et William M. Gray [août 2015]. « SAUR Proteins as Effectors of Hormonal and Environmental Signals in Plant Growth ». In : *Molecular Plant* 8.8, p. 1153-1164. ISSN : 1752-9867. DOI : 10.1016/j.molp.2015.05.003.

Ren, Shixiong, Xun Gu et al. [5 août 2025]. « A chromosome-level genome assembly for Ficus carica provides genetic insights into flowerless fig fruit development, psoralen biosynthesis, and drought tolerance ». In : *Plant Communications* 0.0. Publisher :

Elsevier. ISSN : 2590-3462, 2590-3462. DOI : 10.1016/j.xplc.2025.101470. URL : https://www.cell.com/plant-communications/abstract/S2590-3462(25)00232-9 [visité le 24/09/2025].

Renčo, Marek et al. [oct. 2021]. « Long-Term Giant Hogweed Invasion Contributes to the Structural Changes of Soil Nematofauna ». In : *Plants* 10.10. Number : 10 Publisher : Multidisciplinary Digital Publishing Institute, p. 2103. ISSN : 2223-7747. DOI : 10.3390/plants10102103. URL : https://www.mdpi.com/2223-7747/10/10/2103 [visité le 23/02/2024].

Reyes, José C., M. Isabel Muro-Pastor et Francisco J. Florencio [avr. 2004]. « The GATA family of transcription factors in Arabidopsis and rice ». In : *Plant Physiology* 134.4, p. 1718-1732. ISSN : 0032-0889. DOI : 10.1104/pp.103.037788.

Ribeyre, Zoé et al. [3 jan. 2025]. « De novo transcriptome assembly and discovery of drought-responsive genes in white spruce (Picea glauca) ». In : *PLOS ONE* 20.1. Publisher : Public Library of Science, e0316661. ISSN : 1932-6203. DOI : 10.1371/journal.pone.0316661. URL : https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0316661 [visité le 10/10/2025].

Rice, Edward S. et Richard E. Green [15 fév. 2019]. « New Approaches for Genome Assembly and Scaffolding ». In : *Annual Review of Animal Biosciences* 7 [Volume 7, 2019]. Publisher : Annual Reviews, p. 17-40. ISSN : 2165-8102, 2165-8110. DOI : 10.1146/annurev-animal-020518-115344. URL : https://www.annualreviews.org/content/journals/10.1146/annurev-animal-020518-115344 [visité le 29/07/2025].

Rich-Griffin, Charlotte et al. [1er fév. 2020]. « Single-Cell Transcriptomics : A High-Resolution Avenue for Plant Functional Genomics ». In : *Trends in Plant Science* 25.2, p. 186-197. ISSN : 1360-1385. DOI : 10.1016/j.tplants.2019.10.008. URL :

https://www.sciencedirect.com/science/article/pii/S1360138519302729 [visité le 10/10/2025].

Richards, Christina L. et al. [2017]. « Ecological plant epigenetics : Evidence from model and non-model species, and the way forward ». In : *Ecology Letters* 20.12. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/ele.12858, p. 1576-1590. ISSN : 1461-0248. DOI : 10.1111/ele.12858. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.12858 [visité le 04/10/2025].

Richardson, Aaron O. et Jeffrey D. Palmer [1er jan. 2007]. « Horizontal gene transfer in plants ». In : *Journal of Experimental Botany* 58.1, p. 1-9. ISSN : 0022-0957. DOI : 10.1093/jxb/erl148. URL : https://doi.org/10.1093/jxb/erl148 [visité le 07/02/2024].

Ritter, Eleanore J. et al. [27 fév. 2025]. « The assembly and annotation of two teinturier grapevine varieties, Dakapo and Rubired ». In : *Gigabyte* 2025. Publisher : GigaScience Press, gigabyte149-. ISSN : 2709-4715. DOI : 10.46471/gigabyte.149. URL : https://gigabytejournal.com/articles/149 [visité le 24/09/2025].

Robertson, Gordon et al. [nov. 2010]. « De novo assembly and analysis of RNA-seq data ». In : *Nature Methods* 7.11. Publisher : Nature Publishing Group, p. 909-912. ISSN : 1548-7105. DOI : 10.1038/nmeth.1517. URL : https://www.nature.com/articles/nmeth.1517 [visité le 04/08/2025].

Robinson, Mark D., Davis J. McCarthy et Gordon K. Smyth [1er jan. 2010]. « edgeR : a Bioconductor package for differential expression analysis of digital gene expression data ». In : *Bioinformatics* 26.1, p. 139-140. ISSN : 1367-4803. DOI : 10.1093/bioinformatics/btp616. URL : https://doi.org/10.1093/bioinformatics/btp616 [visité le 26/02/2024].

Roppolo, Daniele et al. [août 2014]. « Functional and Evolutionary Analysis of the CASPARIAN STRIP MEMBRANE DOMAIN PROTEIN Family1[C][W] ». In :

Plant Physiology 165.4, p. 1709-1722. ISSN : 0032-0889. DOI : 10.1104/pp.114.239137. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC4119050/ [visité le 09/10/2025].

Rosato, Marcela et al. [13 sept. 2016]. « Conserved Organisation of 45S rDNA Sites and rDNA Gene Copy Number among Major Clades of Early Land Plants ». In : *PLOS ONE* 11.9. Publisher : Public Library of Science, e0162544. ISSN : 1932-6203. DOI : 10.1371/journal.pone.0162544. URL : https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0162544 [visité le 01/07/2025].

Ruan, Jue et Heng Li [fév. 2020]. « Fast and accurate long-read assembly with wtdbg2 ». In : *Nature Methods* 17.2. Publisher : Nature Publishing Group, p. 155-158. ISSN : 1548-7105. DOI : 10.1038/s41592-019-0669-3. URL : https://www.nature.com/articles/s41592-019-0669-3 [visité le 19/05/2025].

Ruaux, Brigitte [4 déc. 2008]. « Les plantes envahissantes des corridors fluviaux : traits biologiques, impacts de Ludwigia peploides et L. grandiflora en Loire moyenne et implications pour la gestion ». Thèse de doct. Université François Rabelais - Tours. URL : https://theses.hal.science/tel-00380073 [visité le 29/09/2025].

Sacomoto, Gustavo et al. [2014]. « Navigating in a Sea of Repeats in RNA-seq without Drowning ». In : *Algorithms in Bioinformatics*. Sous la dir. de Dan Brown et Burkhard Morgenstern. Berlin, Heidelberg : Springer, p. 82-96. ISBN : 978-3-662-44753-6. DOI : 10.1007/978-3-662-44753-6_7.

Salzberg, Steven L. [16 mai 2019]. « Next-generation genome annotation : we still struggle to get it right ». In : *Genome Biology* 20.1, p. 92. ISSN : 1474-760X. DOI : 10.1186/s13059-019-1715-2. URL : https://doi.org/10.1186/s13059-019-1715-2 [visité le 30/09/2025].

Sanchez-Puerta, Maria Virginia [31 déc. 2014]. « Involvement of plastid, mitochondrial and nuclear genomes in plant-to-plant horizontal gene transfer ». In : *Acta Societatis*

Botanicorum Poloniae 83.4. Number : 4, p. 317-323. ISSN : 2083-9480. DOI : 10.5586/asbp.2014.041. URL : https://pbsociety.org.pl/journals/index.php/asbp/article/view/asbp.2014.041 [visité le 08/02/2024].

Santangelo, James S. et al. [18 mars 2022]. « Global urban environmental change drives adaptation in white clover ». In : *Science* 375.6586. Publisher : American Association for the Advancement of Science, p. 1275-1281. DOI : 10.1126/science.abk0989. URL : https://www.science.org/doi/10.1126/science.abk0989 [visité le 26/09/2025].

Savić, Aleksandra et al. [août 2024]. « Is Ambrosia trifida L. Preparing for a Wider Invasion ? Changes in the Plant Morpho-Functional Traits over a Decade ». In : *Agronomy* 14.8. Publisher : Multidisciplinary Digital Publishing Institute, p. 1601. ISSN : 2073-4395. DOI : 10.3390/agronomy14081601. URL : https://www.mdpi.com/2073-4395/14/8/1601 [visité le 26/09/2025].

Sawicki, Jakub et al. [mars 2024]. « Nanopore Sequencing Technology as an Emerging Tool for Diversity Studies of Plant Organellar Genomes ». In : *Diversity* 16.3. Publisher : Multidisciplinary Digital Publishing Institute, p. 173. ISSN : 1424-2818. DOI : 10.3390/d16030173. URL : https://www.mdpi.com/1424-2818/16/3/173 [visité le 07/10/2025].

Schnoes, Alexandra M. et al. [11 déc. 2009]. « Annotation Error in Public Databases : Misannotation of Molecular Function in Enzyme Superfamilies ». In : *PLOS Computational Biology* 5.12. Publisher : Public Library of Science, e1000605. ISSN : 1553-7358. DOI : 10.1371/journal.pcbi.1000605. URL : https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000605 [visité le 30/09/2025].

Schoenaers, Sébastjen et al. [mars 2024]. « Rapid alkalinization factor 22 has a structural and signalling role in root hair cell wall assembly ». In : *Nature Plants* 10.3, p. 494-511. ISSN : 2055-0278. DOI : 10.1038/s41477-024-01637-8.

Schulz, Marcel H. et al. [24 fév. 2012]. « Oases : robust de novo RNA-seq assembly across the dynamic range of expression levels ». In : *Bioinformatics* 28.8, p. 1086. DOI : 10.1093/bioinformatics/bts094. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC3324515/ [visité le 02/12/2024].

Schwacke, Rainer, Marie E. Bolger et Björn Usadel [24 juin 2025]. « PubPlant – a continuously updated online resource for sequenced and published plant genomes ». In : *Frontiers in Plant Science* 16. Publisher : Frontiers. ISSN : 1664-462X. DOI : 10.3389/fpls.2025.1603547. URL : https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2025.1603547/full [visité le 25/11/2025].

Scremin-Dias, Edna et al. [11 avr. 2023]. « Vegetative organs morphological plasticity of Ludwigia grandiflora in flooded and flood-free habitats ». In : *Plant Systematics and Evolution* 309.3, p. 14. ISSN : 1615-6110. DOI : 10.1007/s00606-023-01844-4. URL : https://doi.org/10.1007/s00606-023-01844-4 [visité le 30/09/2025].

Seyfferth, Carolin et al. [17 juin 2021]. « Advances and Opportunities in Single-Cell Transcriptomics for Plant Research ». In : *Annual Review of Plant Biology* 72 [Volume 72, 2021]. Publisher : Annual Reviews, p. 847-866. ISSN : 1543-5008, 1545-2123. DOI : 10.1146/annurev-arplant-081720-010120. URL : https://www.annualreviews.org/content/journals/10.1146/annurev-arplant-081720-010120 [visité le 10/10/2025].

Sharkey, Thomas D [2 jan. 2024]. « The end game(s) of photosynthetic carbon metabolism ». In : *Plant Physiology* 195.1, p. 67-78. ISSN : 0032-0889. DOI : 10.1093/plphys/kiad601. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11060661/ [visité le 17/09/2025].

She, Hongbing et al. [17 jan. 2022]. « Comparative chloroplast genome analyses of cultivated spinach and two wild progenitors shed light on the phylogenetic relationships and variation ». In : *Scientific Reports* 12, p. 856. ISSN : 2045-2322. DOI : 10.1038/s41598-022-04918-4. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8763918/ [visité le 17/09/2025].

Shimoyama, Yuki [24 jan. 2024]. *pybarrnap : Python implementation of barrnap*. URL : https://github.com/moshi4/pybarrnap.

Simpson, Jared T. et al. [6 jan. 2009]. « ABySS : A parallel assembler for short read sequence data ». In : *Genome Research* 19.6. Company : Cold Spring Harbor Laboratory Press Distributor : Cold Spring Harbor Laboratory Press Institution : Cold Spring Harbor Laboratory Press Label : Cold Spring Harbor Laboratory Press Publisher : Cold Spring Harbor Lab, p. 1117-1123. ISSN : 1088-9051, 1549-5469. DOI : 10.1101/gr.089532.108. URL : http://genome.cshlp.org/content/19/6/1117 [visité le 20/09/2025].

Sivashankari, Selvarajan et Piramanayagam Shanmughavel [29 déc. 2006]. « Functional annotation of hypothetical proteins – A review ». In : *Bioinformation* 1.8, p. 335-338. ISSN : 0973-2063. DOI : 10.6026/97320630001335. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1891709/ [visité le 23/07/2025].

Skippington, Elizabeth et al. [7 juill. 2015]. « Miniaturized mitogenome of the parasitic plant Viscum scurruloideum is extremely divergent and dynamic and has lost all nad genes ». In : *Proceedings of the National Academy of Sciences* 112.27. Publisher : Proceedings of the National Academy of Sciences, E3515-E3524. DOI : 10.1073/pnas.1504491112. URL : https://www.pnas.org/doi/abs/10.1073/pnas.1504491112 [visité le 08/03/2022].

Sloan, Daniel B. [2013]. « One ring to rule them all ? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure ». In :

New Phytologist 200.4. _eprint : https ://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/nph.12395, p. 978-985. ISSN : 1469-8137. DOI : 10.1111/nph.12395. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.12395 [visité le 22/09/2025].

Sloan, Daniel B. et al. [2012]. « Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus Silene ». In : *Genome Biology and Evolution* 4.3, p. 294-306. ISSN : 1759-6653. DOI : 10.1093/gbe/evs006.

Smith, David Roy [1er jan. 2016]. « The past, present and future of mitochondrial genomics : have we sequenced enough mtDNAs ? » In : *Briefings in Functional Genomics* 15.1, p. 47-54. ISSN : 2041-2649. DOI : 10.1093/bfgp/elv027. URL : https://doi.org/10.1093/bfgp/elv027 [visité le 23/09/2025].

Sohn, Jang-Il et Jin-Wu Nam [14 oct. 2016]. « The present and future of de novo whole-genome assembly ». In : *Briefings in Bioinformatics* 19, bbw096. DOI : 10.1093/bib/bbw096.

Soneson, Charlotte, Michael I. Love et Mark D. Robinson [30 déc. 2016]. *Differential analyses for RNA-seq : transcript-level estimates improve gene-level inferences*. DOI : 10.12688/f1000research.7563.1. URL : https://f1000research.com/articles/4-1521 [visité le 03/09/2025].

Souza, Aline Cavalcante de et Jayme Augusto Prevedello [1er avr. 2021]. « Climate change and biological invasion as additional threats to an imperiled palm ». In : *Perspectives in Ecology and Conservation* 19.2, p. 216-224. ISSN : 2530-0644. DOI : 10.1016/j.pecon.2021.02.003. URL : https://www.sciencedirect.com/science/article/pii/S2530064421000183 [visité le 30/09/2025].

Spartz, Angela K. et al. [mai 2014]. « SAUR Inhibition of PP2C-D Phosphatases Activates Plasma Membrane H+-ATPases to Promote Cell Expansion in Arabidopsis ».

In : *The Plant Cell* 26.5, p. 2129-2142. ISSN : 1532-298X. DOI : 10.1105/tpc.114.126037.

Sreedasyam, Avinash et al. [8 sept. 2023]. « JGI Plant Gene Atlas : an updateable transcriptome resource to improve functional gene descriptions across the plant kingdom ». In : *Nucleic Acids Research* 51.16, p. 8383-8401. ISSN : 0305-1048. DOI : 10.1093/nar/gkad616. URL : https://doi.org/10.1093/nar/gkad616 [visité le 05/10/2025].

Srikanth, Krishnamoorthy et al. [20 mars 2020]. « A Comparison between Hi-C and 10X Genomics Linked Read Sequencing for Whole Genome Phasing in Hanwoo Cattle ». In : *Genes* 11.3, p. 332. ISSN : 2073-4425. DOI : 10.3390/genes11030332. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7140831/ [visité le 16/05/2025].

Stanke, Mario et al. [1er mars 2008]. « Using native and syntenically mapped cDNA alignments to improve de novo gene finding ». In : *Bioinformatics* 24.5, p. 637-644. ISSN : 1367-4803. DOI : 10.1093/bioinformatics/btn013. URL : https://doi.org/10.1093/bioinformatics/btn013 [visité le 21/09/2025].

Stepanova, Anna N. et al. [nov. 2011]. « The Arabidopsis YUCCA1 flavin monooxygenase functions in the indole-3-pyruvic acid branch of auxin biosynthesis ». In : *The Plant Cell* 23.11, p. 3961-3973. ISSN : 1532-298X. DOI : 10.1105/tpc.111.088047.

Stoeckel, Solenn et al. [2024]. « Reproductive modes in populations of late-acting self-incompatible and self-compatible polyploid *Ludwigia grandiflora* subsp. *hexapetala* in western Europe ». In : *Peer Community Journal* 4. ISSN : 2804-3871. DOI : 10.24072/pcjournal.458. URL : https://peercommunityjournal.org/articles/10.24072/pcjournal.458/ [visité le 06/08/2025].

Štorchová, Helena et Manuela Krüger [1er fév. 2024]. « Methods for assembling complex mitochondrial genomes in land plants ». In : *Journal of Experimental Botany*,

erae034. ISSN : 0022-0957. DOI : 10.1093/jxb/erae034. URL : https://doi.org/10.1093/jxb/erae034 [visité le 26/02/2024].

Storer, Jessica M. et al. [avr. 2022]. « Methodologies for the De novo Discovery of Transposable Element Families ». In : *Genes* 13.4. Number : 4 Publisher : Multidisciplinary Digital Publishing Institute, p. 709. ISSN : 2073-4425. DOI : 10.3390/genes13040709. URL : https://www.mdpi.com/2073-4425/13/4/709 [visité le 01/08/2025].

Stortenbeker, Niek et Marian Bemer [1er jan. 2019]. « The SAUR gene family : the plant's toolbox for adaptation of growth and development ». In : *Journal of Experimental Botany* 70.1, p. 17-27. ISSN : 1460-2431. DOI : 10.1093/jxb/ery332.

Stupar, Robert M. et al. [24 avr. 2001]. « Complex mtDNA constitutes an approximate 620-kb insertion on Arabidopsis thaliana chromosome 2 : Implication of potential sequencing errors caused by large-unit repeats ». In : *Proceedings of the National Academy of Sciences* 98.9. Publisher : Proceedings of the National Academy of Sciences, p. 5099-5103. DOI : 10.1073/pnas.091110398. URL : https://www.pnas.org/doi/full/10.1073/pnas.091110398 [visité le 18/09/2025].

Sullivan, Alexis R et al. [1er jan. 2020]. « The Mitogenome of Norway Spruce and a Reappraisal of Mitochondrial Recombination in Plants ». In : *Genome Biology and Evolution* 12.1, p. 3586-3598. ISSN : 1759-6653. DOI : 10.1093/gbe/evz263. URL : https://doi.org/10.1093/gbe/evz263 [visité le 19/07/2023].

Sultan, Marc et al. [11 août 2014]. « Influence of RNA extraction methods and library selection schemes on RNA-seq data ». In : *BMC Genomics* 15.1, p. 675. ISSN : 1471-2164. DOI : 10.1186/1471-2164-15-675. URL : https://doi.org/10.1186/1471-2164-15-675 [visité le 22/09/2025].

Sun, Xiaoxue, Daling Feng et al. [19 déc. 2022]. « Single-cell transcriptome reveals dominant subgenome expression and transcriptional response to heat stress in Chinese cabbage ». In : *Genome Biology* 23.1, p. 262. ISSN : 1474-760X. DOI : 10.1186/

s13059-022-02834-4. URL : https://doi.org/10.1186/s13059-022-02834-4 [visité le 18/01/2023].

Sun, Yanqing, Lianguang Shang et al. [1er avr. 2022]. « Twenty years of plant genome sequencing : achievements and challenges ». In : *Trends in Plant Science* 27.4. Publisher : Elsevier, p. 391-401. ISSN : 1360-1385. DOI : 10.1016/j.tplants.2021.10.006. URL : https://www.cell.com/trends/plant-science/abstract/S1360-1385(21)00281-8 [visité le 29/07/2025].

Swiderski, M. R. et R. W. Innes [avr. 2001]. « The Arabidopsis PBS1 resistance gene encodes a member of a novel protein kinase subfamily ». In : *The Plant Journal : For Cell and Molecular Biology* 26.1, p. 101-112. ISSN : 0960-7412. DOI : 10.1046/j.1365-313x.2001.01014.x.

Teichmann, Martin, Hélène Dumay-Odelot et Sébastien Fribourg [2012]. « Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes ». In : *Transcription* 3.1, p. 2-7. ISSN : 2154-1272. DOI : 10.4161/trns.3.1.18917.

The Angiosperm Phylogeny Group et al. [1er mai 2016]. « An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants : APG IV ». In : *Botanical Journal of the Linnean Society* 181.1, p. 1-20. ISSN : 0024-4074. DOI : 10.1111/boj.12385. URL : https://doi.org/10.1111/boj.12385 [visité le 03/05/2024].

The Galaxy Community [5 juill. 2024]. « The Galaxy platform for accessible, reproducible, and collaborative data analyses : 2024 update ». In : *Nucleic Acids Research* 52 [W1], W83-W94. ISSN : 0305-1048. DOI : 10.1093/nar/gkae410. URL : https://doi.org/10.1093/nar/gkae410 [visité le 17/06/2025].

The Gene Ontology Consortium et al. [2 mai 2023]. « The Gene Ontology knowledgebase in 2023 ». In : *Genetics* 224.1, iyad031. ISSN : 1943-2631. DOI : 10.1093/

genetics/iyad031. URL : https://doi.org/10.1093/genetics/iyad031 [visité le 01/08/2025].

THE INTERNATIONAL WHEAT GENOME SEQUENCING CONSORTIUM (IWGSC) et al. [17 août 2018]. « Shifting the limits in wheat research and breeding using a fully annotated reference genome ». In : *Science* 361.6403. Publisher : American Association for the Advancement of Science, eaar7191. DOI : 10.1126/science.aar7191. URL : https://www.science.org/doi/10.1126/science.aar7191 [visité le 18/09/2025].

Theoharides, Kathleen A. et Jeffrey S. Dukes [2007]. « Plant invasion across space and time : factors affecting nonindigenous species success during four stages of invasion ». In : *New Phytologist* 176.2. _eprint : https://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8137.2007.02207.x, p. 256-273. ISSN : 1469-8137. DOI : 10.1111/j.1469-8137.2007.02207.x. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.2007.02207.x [visité le 30/09/2025].

Thouvenot, Lise, Jacques Haury, Gaétan Pottier et al. [1er mars 2017]. « Reciprocal indirect facilitation between an invasive macrophyte and an invasive crayfish ». In : *Aquatic Botany* 139, p. 1-7. ISSN : 0304-3770. DOI : 10.1016/j.aquabot.2017.02.002. URL : https://www.sciencedirect.com/science/article/pii/S0304377016301747 [visité le 23/02/2024].

Thouvenot, Lise, Jacques Haury et Gabrielle Thiebaut [2013]. « A success story : water primroses, aquatic plant pests ». In : *Aquatic Conservation : Marine and Freshwater Ecosystems* 23.5. _eprint : https://onlinelibrary.wiley.com/doi/pdf/10.1002/aqc.2387, p. 790-803. ISSN : 1099-0755. DOI : 10.1002/aqc.2387. URL : https://onlinelibrary.wiley.com/doi/abs/10.1002/aqc.2387 [visité le 17/06/2025].

Timmis, Jeremy N. et al. [fév. 2004]. « Endosymbiotic gene transfer : organelle genomes forge eukaryotic chromosomes ». In : *Nature Reviews Genetics* 5.2. Publisher : Nature

Publishing Group, p. 123-135. ISSN : 1471-0064. DOI : 10.1038/nrg1271. URL : https://www.nature.com/articles/nrg1271 [visité le 18/09/2025].

Tipney, Hannah et Lawrence Hunter [1er fév. 2010]. « An introduction to effective use of enrichment analysis software ». In : *Human Genomics* 4.3, p. 202. ISSN : 1479-7364. DOI : 10.1186/1479-7364-4-3-202. URL : https://doi.org/10.1186/1479-7364-4-3-202 [visité le 22/09/2025].

Todd, Erica V., Michael A. Black et Neil J. Gemmell [2016]. « The power and promise of RNA-seq in ecology and evolution ». In : *Molecular Ecology* 25.6. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/mec.13526, p. 1224-1241. ISSN : 1365-294X. DOI : 10.1111/mec.13526. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13526 [visité le 30/07/2025].

Toyota, Masatsugu et al. [14 sept. 2018]. « Glutamate triggers long-distance, calcium-based plant defense signaling ». In : *Science (New York, N.Y.)* 361.6407, p. 1112-1115. ISSN : 1095-9203. DOI : 10.1126/science.aat7744.

Tsiatsiani, L. et al. [août 2011]. « Metacaspases ». In : *Cell Death and Differentiation* 18.8, p. 1279-1288. ISSN : 1476-5403. DOI : 10.1038/cdd.2011.66.

Twyford, Alex D. et Rob W. Ness [2017]. « Strategies for complete plastid genome sequencing ». In : *Molecular Ecology Resources* 17.5. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/1 0998.12626, p. 858-868. ISSN : 1755-0998. DOI : 10.1111/1755-0998.12626. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.12626 [visité le 30/07/2025].

Tyszka, Alexa S. et al. [30 mars 2023]. « Characterizing conflict and congruence of molecular evolution across organellar genome sequences for phylogenetics in land plants ». In : *Frontiers in Plant Science* 14, p. 1125107. ISSN : 1664-462X. DOI : 10.3389/fpls.2023.1125107. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10098128/ [visité le 17/09/2025].

Vanyushin, Boris F. et Vasili V. Ashapkin [1er août 2011]. « DNA methylation in higher plants : Past, present and future ». In : *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. Epigenetic control of cellular and developmental processes in plants 1809.8, p. 360-368. ISSN : 1874-9399. DOI : 10.1016/j.bbagrm.2011.04.006. URL : https://www.sciencedirect.com/science/article/pii/S1874939911000630 [visité le 28/07/2025].

Via, Sara et Russell Lande [1985]. « Genotype-Environment Interaction and the Evolution of Phenotypic Plasticity ». In : *Evolution* 39.3. Publisher : Oxford University Press, p. 505-522. ISSN : 0014-3820. DOI : 10.2307/2408649. URL : https://www.jstor.org/stable/2408649 [visité le 26/09/2025].

Vidalis, Amaryllis et al. [20 déc. 2016]. « Methylome evolution in plants ». In : *Genome Biology* 17.1, p. 264. ISSN : 1474-760X. DOI : 10.1186/s13059-016-1127-5. URL : https://doi.org/10.1186/s13059-016-1127-5 [visité le 28/07/2025].

VilÀ, Montserrat et al. [1er juill. 2021]. « Viewing Emerging Human Infectious Epidemics through the Lens of Invasion Biology ». In : *BioScience* 71.7, p. 722-740. ISSN : 0006-3568. DOI : 10.1093/biosci/biab047. URL : https://doi.org/10.1093/biosci/biab047 [visité le 30/09/2025].

Virdi, Amardeep S., Supreet Singh et Prabhjeet Singh [2015]. « Abiotic stress responses in plants : roles of calmodulin-regulated proteins ». In : *Frontiers in Plant Science* 6, p. 809. ISSN : 1664-462X. DOI : 10.3389/fpls.2015.00809.

VNF [13 fév. 2024]. *Assises nationales "Plantes Exotiques Envahissantes" 2024 : des solutions à long terme*. VNF. URL : https://www.vnf.fr/vnf/dossiers-actualitess/assises-nationales-plantes-exotiques-envahissantes-2024-des-solutions-a-long-terme/ [visité le 30/09/2025].

Voshall, Adam et Etsuko N. Moriyama [1er avr. 2020]. « Next-generation transcriptome assembly and analysis : Impact of ploidy ». In : *Methods*. RNA-Seq : Me-

thods and Applications 176, p. 14-24. ISSN : 1046-2023. DOI : 10.1016/j.ymeth.2019.06.001. URL : https://www.sciencedirect.com/science/article/pii/S1046202318303803 [visité le 26/02/2024].

Wagner, Warren L., P. C. Hoch et Peter H. Raven [2007]. « Revised classification of the Onagraceae ». In : URL : http://hdl.handle.net/10088/7611 [visité le 21/07/2025].

Walker, Joseph F. et al. [24 sept. 2019]. « Characterizing gene tree conflict in plastome-inferred phylogenies ». In : *PeerJ* 7, e7747. ISSN : 2167-8359. DOI : 10.7717/peerj.7747. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6764362/ [visité le 17/09/2025].

Wambugu, Peterson W, Marie-Noelle Ndjiondjop et Robert J Henry [1er mai 2018]. « Role of genomics in promoting the utilization of plant genetic resources in genebanks ». In : *Briefings in Functional Genomics* 17.3, p. 198-206. ISSN : 2041-2657. DOI : 10.1093/bfgp/ely014. URL : https://doi.org/10.1093/bfgp/ely014 [visité le 16/09/2025].

Wang, Dan, Wenkai Li, Bing Yang et al. [23 juill. 2025]. « Epilobieae genomes and the evolution of Myrtales ». In : *BMC Plant Biology* 25.1, p. 950. ISSN : 1471-2229. DOI : 10.1186/s12870-025-07003-w. URL : https://doi.org/10.1186/s12870-025-07003-w [visité le 28/09/2025].

Wang, Daryi, Yu-Wei Wu et al. [sept. 2007]. « Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA ». In : *Molecular Biology and Evolution* 24.9, p. 2040-2048. ISSN : 0737-4038. DOI : 10.1093/molbev/msm133.

Wang, Jeremy R., James Holt et al. [9 fév. 2018]. « FMLRC : Hybrid long read error correction using an FM-index ». In : *BMC Bioinformatics* 19.1, p. 50. ISSN : 1471-2105. DOI : 10.1186/s12859-018-2051-3. URL : https://doi.org/10.1186/s12859-018-2051-3 [visité le 29/07/2025].

Wang, Shuo, Rostislav Y. Blume et al. [31 jan. 2024]. « Chromosome-level assembly and analysis of Camelina neglecta : a novel diploid model for Camelina biotechnology research ». In : *Biotechnology for Biofuels and Bioproducts* 17.1, p. 17. ISSN : 2731-3654. DOI : 10.1186/s13068-024-02466-9. URL : https://doi.org/10.1186/s13068-024-02466-9 [visité le 08/10/2025].

Wang, Weiwen, Miriam Schalamun et al. [29 déc. 2018]. « Assembly of chloroplast genomes with long- and short-read data : a comparison of approaches using Eucalyptus pauciflora as a test case ». In : *BMC Genomics* 19.1, p. 977. ISSN : 1471-2164. DOI : 10.1186/s12864-018-5348-8. URL : https://doi.org/10.1186/s12864-018-5348-8 [visité le 30/07/2025].

Wang, Xi, Baoshan Wang et Fang Yuan [23 nov. 2023]. « Deciphering the roles of unknown/uncharacterized genes in plant development and stress responses ». In : *Frontiers in Plant Science* 14, p. 1276559. ISSN : 1664-462X. DOI : 10.3389/fpls.2023.1276559. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC10701545/ [visité le 18/09/2025].

Wang, Yan, Jennifer Selinski et al. [22 juin 2020]. « Linking mitochondrial and chloroplast retrograde signalling in plants ». In : *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 375.1801, p. 20190410. ISSN : 1471-2970. DOI : 10.1098/rstb.2019.0410.

Wang, Ying, Xifeng Li, Baofang Fan et al. [27 fév. 2021]. « Regulation and Function of Defense-Related Callose Deposition in Plants ». In : *International Journal of Molecular Sciences* 22.5, p. 2393. ISSN : 1422-0067. DOI : 10.3390/ijms22052393.

Wang, Yueyue, Yufeng Shi et al. [8 nov. 2021]. « Roles of the 2-Oxoglutarate-Dependent Dioxygenase Superfamily in the Flavonoid Pathway : A Review of the Functional Diversity of F3H, FNS I, FLS, and LDOX/ANS ». In : *Molecules (Basel, Switzerland)* 26.21, p. 6745. ISSN : 1420-3049. DOI : 10.3390/molecules26216745.

Wang, Zhang et Martin Wu [22 jan. 2015]. « An integrated phylogenomic approach toward pinpointing the origin of mitochondria ». In : *Scientific Reports* 5, p. 7949. ISSN : 2045-2322. DOI : 10.1038/srep07949. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4302308/ [visité le 17/09/2025].

Wang, Zhaoqiang, Xiangyan Wei et al. [2024]. « The transcription factor WRKY22 modulates ethylene biosynthesis and root development through transactivating the transcription of ACS5 and ACO5 in Arabidopsis ». In : *Physiologia Plantarum* 176.3, e14371. ISSN : 1399-3054. DOI : 10.1111/ppl.14371.

Waters, Mark T. et al. [avr. 2009]. « GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis ». In : *The Plant Cell* 21.4, p. 1109-1128. ISSN : 1040-4651. DOI : 10.1105/tpc.108.065250.

Weerden, Nicole L. van der, Mark R. Bleackley et Marilyn A. Anderson [oct. 2013]. « Properties and mechanisms of action of naturally occurring antifungal peptides ». In : *Cellular and molecular life sciences : CMLS* 70.19, p. 3545-3570. ISSN : 1420-9071. DOI : 10.1007/s00018-013-1260-1.

Wendel, Jonathan F. et al. [1er mars 2016]. « Evolution of plant genome architecture ». In : *Genome Biology* 17.1, p. 37. ISSN : 1474-760X. DOI : 10.1186/s13059-016-0908-1. URL : https://doi.org/10.1186/s13059-016-0908-1 [visité le 16/05/2025].

Wicke, Susann, Andrea Costa et al. [1er nov. 2011]. « Restless 5S : The re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants ». In : *Molecular Phylogenetics and Evolution* 61.2, p. 321-332. ISSN : 1055-7903. DOI : 10.1016/j.ympev.2011.06.023. URL : https://www.sciencedirect.com/science/article/pii/S1055790311003125 [visité le 01/07/2025].

Wicke, Susann, Gerald M. Schneeweiss et al. [2011]. « The evolution of the plastid chromosome in land plants : gene content, gene order, gene function ». In : *Plant*

Molecular Biology 76.3, p. 273-297. ISSN : 0167-4412. DOI : 10.1007/s11103-011-9762-4. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3104136/ [visité le 21/07/2025].

Williamson, Mark et Alastair Fitter [1996]. « The Varying Success of Invaders ». In : *Ecology* 77.6. Publisher : Ecological Society of America, p. 1661-1666. ISSN : 0012-9658. DOI : 10.2307/2265769. URL : https://www.jstor.org/stable/2265769 [visité le 23/02/2024].

Wilson, Jonathan et al. [11 mars 2025]. « Copy number variation contributes to parallel local adaptation in an invasive plant ». In : *Proceedings of the National Academy of Sciences* 122.10. Publisher : Proceedings of the National Academy of Sciences, e2413587122. DOI : 10.1073/pnas.2413587122. URL : https://www.pnas.org/doi/10.1073/pnas.2413587122 [visité le 26/09/2025].

Wood, Derrick E., Jennifer Lu et Ben Langmead [28 nov. 2019]. « Improved metagenomic analysis with Kraken 2 ». In : *Genome Biology* 20.1, p. 257. ISSN : 1474-760X. DOI : 10.1186/s13059-019-1891-0. URL : https://doi.org/10.1186/s13059-019-1891-0 [visité le 29/04/2025].

Wu, Zhi-Qiang et al. [8 juill. 2020]. « Genomic architectural variation of plant mitochondria—A review of multichromosomal structuring ». In : *Journal of Systematics and Evolution* 60.1. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/jse.12655, p. 160-168. ISSN : 1759-6831. DOI : 10.1111/jse.12655. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/jse.12655 [visité le 20/02/2023].

Wynn, Emily [avr. 2019]. « Plant mitochondrial genome evolution and structure has been shaped by double-strand break repair and recombination ». In.

Wynn, Emily L et Alan C Christensen [1er fév. 2019]. « Repeats of Unusual Size in Plant Mitochondrial Genomes : Identification, Incidence and Evolution ». In : *G3 Genes|Genomes|Genetics* 9.2, p. 549-559. ISSN : 2160-1836. DOI : 10.1534/g3.118.

200948. URL : https://academic.oup.com/g3journal/article/9/2/549/6026745 [visité le 18/01/2022].

Xia, Fagang et al. [jan. 2024]. « Polyploid Genome Assembly Provides Insights into Morphological Development and Ascorbic Acid Accumulation of Sauropus androgynus ». In : *International Journal of Molecular Sciences* 25.1. Publisher : Multidisciplinary Digital Publishing Institute, p. 300. ISSN : 1422-0067. DOI : 10.3390/ijms25010300. URL : https://www.mdpi.com/1422-0067/25/1/300 [visité le 21/09/2025].

Xie, Lingjuan, Xiaojiao Gong et al. [avr. 2024]. « Technology-enabled great leap in deciphering plant genomes ». In : *Nature Plants* 10.4. Publisher : Nature Publishing Group, p. 551-566. ISSN : 2055-0278. DOI : 10.1038/s41477-024-01655-6. URL : https://www.nature.com/articles/s41477-024-01655-6 [visité le 12/11/2025].

Xie, Yinlong, Gengxiong Wu et al. [15 juin 2014]. « SOAPdenovo-Trans : de novo transcriptome assembly with short RNA-Seq reads ». In : *Bioinformatics* 30.12, p. 1660-1666. ISSN : 1367-4803. DOI : 10.1093/bioinformatics/btu077. URL : https://doi.org/10.1093/bioinformatics/btu077 [visité le 04/08/2025].

Xiong, Haiyan et al. [mars 2025]. « Multiple roles of NAC transcription factors in plant development and stress responses ». In : *Journal of Integrative Plant Biology* 67.3, p. 510-538. ISSN : 1744-7909. DOI : 10.1111/jipb.13854.

Xu, Hongyun, Xinxin Shi et al. [1er mai 2018]. « Arabidopsis thaliana Trihelix Transcription Factor AST1 Mediates Salt and Osmotic Stress Tolerance by Binding to a Novel AGAG-Box and Some GT Motifs ». In : *Plant & Cell Physiology* 59.5, p. 946-965. ISSN : 1471-9053. DOI : 10.1093/pcp/pcy032.

Xu, Na, Rongchen Wang et al. [fév. 2016]. « The Arabidopsis NRG2 Protein Mediates Nitrate Signaling and Interacts with and Regulates Key Nitrate Regulators ». In : *The Plant Cell* 28.2, p. 485-504. ISSN : 1532-298X. DOI : 10.1105/tpc.15.00567.

Xu, Xun, Shengkai Pan et al. [juill. 2011]. « Genome sequence and analysis of the tuber crop potato ». In : *Nature* 475.7355. Publisher : Nature Publishing Group, p. 189-195. ISSN : 1476-4687. DOI : 10.1038/nature10158. URL : https://www.nature.com/articles/nature10158 [visité le 19/09/2025].

Xu, Yu, Junjie Zou et al. [mars 2019]. « RNA-Seq Transcriptome Analysis of Rice Primary Roots Reveals the Role of Flavonoids in Regulating the Rice Primary Root Growth ». In : *Genes* 10.3. Publisher : Multidisciplinary Digital Publishing Institute, p. 213. ISSN : 2073-4425. DOI : 10.3390/genes10030213. URL : https://www.mdpi.com/2073-4425/10/3/213 [visité le 12/09/2025].

Yan, Chao et al. [11 sept. 2025]. « Chromosome-level genome assembly and population genomics analysis of *Camellia rubituberculata* provide insights into adaptation to karst habitats ». In : *Plant Diversity*. ISSN : 2468-2659. DOI : 10.1016/j.pld.2025.09.004. URL : https://www.sciencedirect.com/science/article/pii/S2468265925001659 [visité le 24/09/2025].

Yang, Heyu, Haimei Chen et al. [jan. 2022]. « De Novo Hybrid Assembly of the Salvia miltiorrhiza Mitochondrial Genome Provides the First Evidence of the Multi-Chromosomal Mitochondrial DNA Structure of Salvia Species ». In : *International Journal of Molecular Sciences* 23.22. Number : 22 Publisher : Multidisciplinary Digital Publishing Institute, p. 14267. ISSN : 1422-0067. DOI : 10.3390/ijms232214267. URL : https://www.mdpi.com/1422-0067/23/22/14267 [visité le 03/05/2024].

Yang, Jia-Xing, Nicolas Dierckxsens et al. [jan. 2023]. « Multichromosomal Mitochondrial Genome of Paphiopedilum micranthum : Compact and Fragmented Genome, and Rampant Intracellular Gene Transfer ». In : *International Journal of Molecular Sciences* 24.4. Number : 4 Publisher : Multidisciplinary Digital Publishing Institute, p. 3976. ISSN : 1422-0067. DOI : 10.3390/ijms24043976. URL : https://www.mdpi.com/1422-0067/24/4/3976 [visité le 19/07/2023].

Yaschenko, Anna E. et al. [1er août 2022]. « Deciphering the molecular basis of tissue-specific gene expression in plants : Can synthetic biology help ? » In : *Current Opinion in Plant Biology* 68, p. 102241. ISSN : 1369-5266. DOI : 10.1016/j.pbi.2022.102241. URL : https://www.sciencedirect.com/science/article/pii/S136952662200070X [visité le 29/09/2025].

Yin, Tongming [2022]. « Deciphering the Multi-Chromosomal Mitochondrial Genome of Populus simonii ». In : *Frontiers in Plant Science* 13.

Yoshida, Saiko et al. [fév. 2019]. « A SOSEKI-based coordinate system interprets global polarity cues in Arabidopsis ». In : *Nature Plants* 5.2, p. 160-166. ISSN : 2055-0278. DOI : 10.1038/s41477-019-0363-6.

You, Jingyi et al. [27 avr. 2022]. « Integrated Methylome and Transcriptome Analysis Widen the Knowledge of Cytoplasmic Male Sterility in Cotton (Gossypium barbadense L.) » In : *Frontiers in Plant Science* 13, p. 770098. ISSN : 1664-462X. DOI : 10.3389/fpls.2022.770098. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9093596/ [visité le 20/01/2023].

Yu, Guangchuang [4 nov. 2024]. « Thirteen years of clusterProfiler ». In : *The Innovation* 5.6, p. 100722. ISSN : 2666-6758. DOI : 10.1016/j.xinn.2024.100722. URL : https://www.sciencedirect.com/science/article/pii/S2666675824001607 [visité le 21/09/2025].

Zardoya, Rafael [17 avr. 2020]. « Recent advances in understanding mitochondrial genome diversity ». In : Type : article. DOI : 10.12688/f1000research.21490.1. URL : https://f1000research.com/articles/9-270 [visité le 25/03/2024].

Zeng, Zhefei et al. [28 août 2025]. « Chromosome-level genome assembly of Hippophae salicifolia ». In : *Scientific Data* 12.1. Publisher : Nature Publishing Group, p. 1503. ISSN : 2052-4463. DOI : 10.1038/s41597-025-05844-6. URL : https://www.nature.com/articles/s41597-025-05844-6 [visité le 24/09/2025].

Zhang, Guo-Jun, Ran Dong et al. [jan. 2020]. « Nuclear Integrants of Organellar DNA Contribute to Genome Structure and Evolution in Plants ». In : *International Journal of Molecular Sciences* 21.3. Number : 3 Publisher : Multidisciplinary Digital Publishing Institute, p. 707. ISSN : 1422-0067. DOI : 10.3390/ijms21030707. URL : https://www.mdpi.com/1422-0067/21/3/707 [visité le 15/04/2024].

Zhang, Meng, Liping Guo et al. [jan. 2019]. « Integrated Methylome and Transcriptome Analysis between the CMS-D2 Line ZBA and Its Maintainer Line ZB in Upland Cotton ». In : *International Journal of Molecular Sciences* 20.23. Number : 23 Publisher : Multidisciplinary Digital Publishing Institute, p. 6070. ISSN : 1422-0067. DOI : 10.3390/ijms20236070. URL : https://www.mdpi.com/1422-0067/20/23/6070 [visité le 20/01/2023].

Zhang, Xiao-Feng, Jacob B. Landis et al. [15 mai 2021]. « Comparative analysis of chloroplast genome structure and molecular dating in Myrtales ». In : *BMC Plant Biology* 21.1, p. 219. ISSN : 1471-2229. DOI : 10.1186/s12870-021-02985-9. URL : https://doi.org/10.1186/s12870-021-02985-9 [visité le 09/10/2025].

Zhang, Xue, Yuanyu Shan et al. [6 mai 2023]. « Assembly of the Complete Mitochondrial Genome of Pereskia aculeata Revealed That Two Pairs of Repetitive Elements Mediated the Recombination of the Genome ». In : *International Journal of Molecular Sciences* 24.9, p. 8366. ISSN : 1422-0067. DOI : 10.3390/ijms24098366.

Zhang, Yingmin, Congwei Yang et al. [26 août 2025]. « A chromosome-level genome assembly of Wurfbainia villosa (Zingiberaceae) ». In : *Scientific Data* 12.1. Publisher : Nature Publishing Group, p. 1490. ISSN : 2052-4463. DOI : 10.1038/s41597-025-05874-0. URL : https://www.nature.com/articles/s41597-025-05874-0 [visité le 24/09/2025].

Zhao, Chenchen et al. [23 mars 2018]. « Roles of Chloroplast Retrograde Signals and Ion Transport in Plant Drought Tolerance ». In : *International Journal of Molecular Sciences* 19.4, p. 963. ISSN : 1422-0067. DOI : 10.3390/ijms19040963.

Zhao, Yunde [2014]. « Auxin biosynthesis ». In : *The Arabidopsis Book* 12, e0173. ISSN : 1543-8120. DOI : 10.1199/tab.0173.

Zhong, Furong et al. [2023]. « Comprehensive analysis of the complete mitochondrial genomes of three Coptis species (C. chinensis, C. deltoidea and C. omeiensis) : the important medicinal plants in China ». In : *Frontiers in Plant Science* 14. ISSN : 1664-462X. URL : https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1166420 [visité le 23/02/2024].

Zhou, Renchao, Noushin Moshgabadi et Keith L. Adams [20 sept. 2011]. « Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids ». In : *Proceedings of the National Academy of Sciences* 108.38. Publisher : Proceedings of the National Academy of Sciences, p. 16122-16127. DOI : 10.1073/pnas.1109551108. URL : https://www.pnas.org/doi/full/10.1073/pnas.1109551108 [visité le 16/05/2025].

Zhou, Xin-Tong, Fang Wang et al. [mars 2018]. « Ectopic expression of SsPETE2, a plastocyanin from Suaeda salsa, improves plant tolerance to oxidative stress ». In : *Plant Science : An International Journal of Experimental Plant Biology* 268, p. 1-10. ISSN : 1873-2259. DOI : 10.1016/j.plantsci.2017.12.006.

Zhu, Bai-Han et al. [2 mars 2018]. « P_RNA_scaffolder : a fast and accurate genome scaffolder using paired-end RNA-sequencing reads ». In : *BMC Genomics* 19.1, p. 175. ISSN : 1471-2164. DOI : 10.1186/s12864-018-4567-3. URL : https://doi.org/10.1186/s12864-018-4567-3 [visité le 06/10/2025].

Zyprych-Walczak, J. et al. [2015]. « The Impact of Normalization Methods on RNA-Seq Data Analysis ». In : *BioMed Research International* 2015.1. _eprint : https ://onli-

nelibrary.wiley.com/doi/pdf/10.1155/2015/621690, p. 621690. ISSN : 2314-6141. DOI : 10.1155/2015/621690. URL : https://onlinelibrary.wiley.com/doi/abs/10.1155/2015/621690 [visité le 05/08/2025].

COLLEGE ECOLOGIE
DOCTORAL GEOSCIENCES
BRETAGNE AGRONOMIE ALIMENTATION

L'INSTITUT agro Rennes Angers

Titre : Production de ressources génomiques pour l'étude des mécanismes d'acclimatation au milieu terrestre de l'espèce aquatique invasive, *Ludwigia grandiflora* subsp. *hexapetala*

Mot clés : jussie, mitogénome, génome, transcriptome, invasion biologique, adaptation

Résumé : Dans un contexte de dispersion des espèces invasives, leur survie est conditionnée par leur capacité à s'adapter. En France, la jussie, *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*), est une espèce aquatique invasive colonisant de nombreux milieux aquatiques et aussi des prairies humides. Pour comprendre cette acclimatation qui a conduit à l'apparition de deux morphotypes aquatique et terrestre, il est indispensable de produire des ressources génomiques pour cette espèce polyploïde non-modèle. L'objectif majeur de cette thèse a été l'assemblage *de novo* des génomes mitochondrial et nucléaire et d'un transcriptome de *Lgh*. Des séquences d'ADN long reads et short reads, et des transcrits issus de conditionnements différents des deux morphotypes ont été utilisées. L'assemblage du mitogénome a montré une structure composée de deux molécules. Un draft du génome et des transcriptomes *de novo* et sur référence de *Lgh* ont été assemblés et annotés. Ils ont montré une complémentarité soulignant leur intérêt pour une amélioration mutuelle. L'étude de l'expression différentielle des gènes en condition terrestre révèle des stratégies différentes du morphotype aquatique, fonction de son conditionnement, et pouvant contribuer à son acclimatation au milieu terrestre. Différentes ressources génomiques de qualité ont été produites permettant une première analyse des réponses fonctionnelles de la jussie. Des améliorations des génome et transcriptome de *Lgh* sont envisagées afin de poursuivre cette étude et de fournir des ressources uniques pour étudier l'histoire évolutive et le rôle de la polyploïdie chez les Onagraceae et dans l'ordre des Myrtales.

Title: Genomic resources production for the study of terrestrial environment acclimation of the invasive aquatic species, *Ludwigia grandiflora* subsp. *hexapetala*

Keywords: water primrose, mitogenome, genome, transcriptome, biological invasion, adaptation

Abstract: In a context of invasive species dispersion, their survival depends on their ability to adapt. In France, the water primrose, *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*), is an invasive aquatic species that colonizes many aquatic environments as well as wet meadows. To understand this acclimation, which led to the emergence of two aquatic and terrestrial morphotypes, it is essential to produce genomic resources for this non-model polyploid species. The main objective of this thesis was to *de novo* assemble the mitochondrial and nuclear genomes and a transcriptome of *Lgh*. Long read and short read DNA sequences and transcripts from different conditions of the two morphotypes were used. The assembly of the mitogenome showed a structure composed of two molecules. A draft of the genome and also *de novo* and reference transcriptomes of *Lgh* were assembled and annotated. They showed complementarity, highlighting their potential for mutual improvement. The study of differential gene expression in terrestrial conditions reveals strategies that differ from those of the aquatic morphotype, depending on its conditioning, and which may contribute to its acclimation to the terrestrial environment. Various high-quality genomic resources have been produced, enabling an initial analysis of the functional responses of water primrose. Improvements to the *Lgh* genome and transcriptome are planned in order to continue this study and provide unique resources for studying the evolutionary history and role of polyploidy in the Onagraceae family and in the Myrtales order.