Aix Marseille Université

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 17 décembre 2024

Andrea MODICA

Genetic bases of microgeographic adaptation in European beech (*Fagus sylvatica* L.) and modelling of its evolutionary response to environmental gradient and climate change

Discipline
Sciences de l'Environnement

Spécialité
Ecologie

École doctorale
Sciences de l'Environnement

Laboratoire/Partenaires de recherche
Ecologie des forêts Méditerranéennes – UR629 INRA Avignon

Composition du jury

Elodie MAGNANOU PHAM IR, CNRS and Sorbonne Université	Rapporteuse
Katharina Birgit BUDDE CR, Northwest German Forest Research Institute	Rapporteuse
Oliver Brendel CR, INRAE UMR Silva	Examinateur
Myriam HEUERTZ DR, INRAE BIOGECO	Examinatrice
Olivier Panaud PR, Université de Perpignan	Président de jury
Ivan Scotti DR, INRAE URFM	Directeur de thése

Affidavit (FR)

Je soussigné, Andrea Modica, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Ivan Scotti et Sylvie Muratorio, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Avignon, le 23 October 2024



Affidavit (EN)

I, Andrea Modica, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific supervision of Ivan Scotti and Sylvie Muratorio, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Avignon, 23 October 2024



Liste de publications et/out brevets et participation aux conférences

Liste des publications réalisées dans le cadre du projet de thèse :

1. Modica Andrea, Lalagüe Hadrien, Muratorio Sylvie, Scotti Ivan (2024). **Rolling down that mountain: microgeographical adaptive divergence during a fast population expansion along a steep environmental gradient in European beech**, Heredity (2024) 133:99–112; https://doi.org/10.1038/s41437-024-00696-z

Participation aux conférences et écoles d'été au cours de la période de thèse :

2. 21-25 June 2021 - Introduction to Genome-Wide Association Stidies (GWAS) (**course**)
3. 4-7 September 2022 - 9th Congress Società Italiana Biologia Evoluzionistica, Ancona (**oral presentation**)
4. 13-14 October 2022 – Réunion du réseau de génétique du département ECODIV, Bordeaux (**oral presentation**)
5. 7 November 2022 – Colloque interne URFM (**oral presentation**)
6. 3-4 May 2023 - 30ème Congrès des Doctorants en "Sciences de l'Environnement", Marseille (**oral presentation**)
7. 19-23 September 2022 - Summer school on Software and Statistical Methods for Population, Assois (**poster**)
8. 20-22 April 2024 - Métaprogramme « Agriculture et Forêt face au changement climatique : adaptation et atténuation » - CLIMAE au traitment statistiques des données temporelles, Lyon (**poster**)
9. 18-19 September 2024 - Ecological Genetics Group Annual Meeting, online (**oral presentation**)

Summary (FR)

Long de gradients altitudinaux, les populations de hêtres européens (*Fagus sylvatica* L.) présentent une variation clinale de l'effeuillage. Dans le jardin commun, les arbres de haute altitude sortent leurs feuilles avant ceux de basse altitude, suggérant une différenciation génétique anticlinal.
Cette thèse teste l'hypothèse selon laquelle de fortes pressions sélectives à courte échelle spatiale conduisent à la différenciation génétique de l'effeuillage dans les peuplements de hêtres.
J'ai évalué la divergence génétique sur donnée d'exome-capture entre quatre peuplements différents par conditions environnementales, distance et passé démographique et contrôlé les effets démographiques. J'ai trouvé différents signaux (cinq loci et 17 haplotypes) entre tous les peuplements.
Une association génomique a étudié l'architecture de l'effeuillage –modélisé par une courbe logistique– au niveau du peuplement pendant deux années consécutives sur donnée de reséquençage du génome. 47 marqueurs uniques sont associés à l'effeuillage, placés principalement en régions intergéniques et distincts selon l'année.
En croisant les deux bases de données, j'ai estimé les valeurs génétiques de l'effeuillage le long de la montagne; ils ont suivi un contre-gradient. Des simulations ont évalué le potentiel évolutif de l'effeuillage dans des conditions plus chaudes.
Les résultats suggèrent que l'adaptation locale évolue rapidement, à de courtes échelles spatiales, et par quelques loci au sein d'une architecture polygénique, résultant une différenciation phénotypique clinale et une différenciation génotypique anticlinale.

Mots clés: adaptation locale, GWAS, phénologie végétative, gradient environnemental

Summary (EN)

Populations of European beech (*Fagus sylvatica* L.) along altitudinal gradients exhibit clinal variation in the timing of bud-break, a poorly understood polygenic trait. In common garden, trees from higher elevations break their buds before those from lower elevations, suggesting a pattern of anticlinal genetic differentiation.

This thesis tests the hypothesis that strong selective pressures at short spatial scales drive the genetic differentiation of spring phenology in beech stands.

I assessed the genetic divergence between four stands differing in environmental conditions, distance, and demographic past, on an exome-capture panel. Explicitly modelling of population history controlled demographic effects. Different signals of selections (five loci and 17 haplotypes) were found between all the stands.

I investigated the architecture of bud-break –modeled by a logistic curve– at the stand level in two consecutive years by genome-wide association on a whole genome resequencing panel. I found 47 unique polymorphisms associated with bud-break, mostly outside coding regions and distinct by year.

By mapping the signals onto the exome capture data, I estimated the breeding values of bud-break along the mountain slope and found a countergradient pattern. Forward-simulations assessed the evolutionary potential of bud-break under warmer conditions.

Results suggest that local adaptation evolves rapidly, at short spatial scales, and by changes at few loci within a polygenic architecture, resulting in clinal phenotypic differentiation and anticlinal genotipic differentiation.

Keywords: local adaptation, GWAS, vegetative phenology, environmental gradient

Genetic bases of microgeographic adaptation in European beech (*Fagus sylvatica* L.) and modelling of its evolutionary response to environmental gradient and climate change

"*There is always another secret."*
Brandon Sanderson, "The Last Empire"

"¿de dónde viene esta conspiración de invisibilidades?
ninguna palabra es visible"
Alejandra Pizarnik, "En esta noche, en este mundo"

To my parents, Angelo and Maria Teresa,
My grandmother Eva,
and Fripp, question mark-tailed cat

Acknowledgments

How am I supposed to start? It's a large crew, those people I feel I owe something during these three and a half years. No man is an island; nor is a dishwasher nor is a capybara -

I'm sorry. It started kind of fine but this capybara thing has driven everything off the rails. I'm not Elwood Blues enough to improvise the acknowledgements after a rushed entrance *and* improvise them with ease. (I don't even have some decent sunglasses).
But, believe me *bows* they are heartfelt and sincere.

Ivan, I can't write down how grateful I am for your supervision. Over these five years you have trusted me probably more than I trust myself, giving me room for my fair share of trials and errors and someone with whom to discuss them. I think having a very similar sense of humor has been a common ground for our mutual understanding - though, I admit, sometimes I swerved away on my own tangent - and I admire your enormous curiosity. You have gone well beyond being a mentor: you guided me through the civil perils of life in a country the language of which I barely spoke and which has a fame worth its depiction in *Asterix* comics. Since it has not found its place in Chapter 1 as we had planned during the pandemic, I feel obliged to write here the word ***Decameron***.

Thanks, Sylvie, for our exchanges and the insights on what to look for, especially when it came to understanding what was going on with these diverging stands - it was a thorough investigation into maps and papers. Though between us there was half of the France longitudinal extension, you helped me a lot in staying on the track and understanding how structure is important.

A great thanks to the members of my jury de thése, Olivier Panaud, Oliver Brendel, Katharina Budde, Elodie Magnanou, and Myriam Heuertz, for having accepted to review and evaluate my work.

I want to thank the members of my thesis committee, Katrin, Ludovic, and Thibaut, for your guidance during these meetings (I think I have improved my summarising skills) and the talks we had in person. Thibaut, I treasure the Summer School both for its content and the location - it's the first time I've not gotten scared by mountains (alpine lodge kind of scare me). Ludovic, the time you took to go through my project was immensely useful to pinpoint the pits of my reasoning, to rework which questions to ask and how to ask them. Katrin, I appreciated your tips on what to look for to get *true* signals in my data, and your encouragement was uplifting in that moment, at the start of June, when I had to make for a final leap forward. And thanks to you amazing lab members too!

Lauren, working with you is very enriching. We come from different backgrounds, have different methods, *different coding styles* (I'm going to catch up with dplyr, I'm definitely going to) and this is a great moment to enter into one's perspective - even more due to your kindness and

patience when I lagged behind. Sorry for how many times I ended up with a “well, it looks like I have been a fool so far” and you’ve been waiting with the results of the entire pipeline in your hands while I arrived saying “*there was a bug*”. I guess that’s kinda bugging. You can’t tell how many dead ends you have made me avoid. You have a panoply of experience on the field and pieces of knowledge I was unaware of. And your cat is a fine hunter!

A big thanks to my bureaumates, Jorge, Camille, and Lea for making me feel part of a group for how much I remained on my own in that corner of the office. I don’t want to come back again to you as the day of the soutenance; that big, hovering thank you sounded too formal to exchange with you, who shared that room for this long. I hope not to have come out as too weird or with a sense of humor too out of place. (I’m still sounding too formal? Too detached?)
(yes, I’m writing *sottovoce* to not jump-scare myself)

And thanks to Victor, Renaud, Alice, Bélen, and Marion. I wish I could have joined you more often in the Monday afterworks and - awrk, it’s sounding like a long “what if”, is it?
Victor, I owe you a lot for moving the first steps in Avignon, especially when it came to meeting the landlady. You can’t tell how scared I was!

To all the staff of the URFM for the on-and-off talks in the coffee room or in and out the bureaux. Laurent and Marie-Claude, thanks for guiding me through all the shallow waters of office documentations, both in my stay in 2019 and during my PhD - especially during that break. Fréderic, you have been a thorough guidance in explaining phenology to a complete foreigner -both in phenology *and* in language. You have been a treasure chest! Maurene, I want to be like you when I grow up.

To the swimmers of the Stade Nautique, Lucia, Suzie, Nathalie: it has been a pleasure to meet such a close-knitted group. I miss the pre-dawn chats in the cold and the swims under a colouring sky!

Dear cat of the Maison de Poesie who has been asked me to open her door each time we meet at 6h15 a.m.: unfortunately I don’t have the keys to your house.

To Clara, Marzia, Edoardo, Benedetta, and all the people of the tri-annual book club: you’ve been an amazing company and a source of new perspectives I may have never pondered before. Clara and Marzia, you did a great work in keeping up the club for this long and with such an amicable atmosphere, in spite of meeting only by video. Juggling it between your work and studies must have been quite a swim bounded upstream.

To Dr. Giammaria and Simona, thanks for having been at my side in this part of the path.
Emanuele, we’re both strangers in a strange land and maybe of another, more or strange, one: *fare forward, voyager!*

Fripp, demonic cat, stop eating my hairs.

To those who have been in touch but now we silently parted our ways because things got in the way; to those I'm sharing the bits and the dots in this interesting time we've happened to live (*Hard times, Hard times, very hard times, Mr. Lindsay / Hard times and winter so cruel*); to those I currently don't know who they are but we'll see: thanks.

Then, if this ship I've sailed on would ever happen to sink, I'm going to open a dancehall, a dancehall so big each night it will host multiple events and I will call it *The Whole*, so I will able to say that "*The Whole* is more than the sum of its parties."

Table of Contents

Glossary of Abbreviations Pag. 16

1. **Introduction** Pag. 17

1.1. The Word for World is Forest
Biology and Ecology of Trees Pag. 17

1.2. “Heart Is Where Home Is”
On (Local) Adaptation Pag. 19

1.2.1 Adaptation in the Time of Climate Change Pag. 24

1.3. How Much Spatially Fine Is a Fine Scale
Local Adaptation along an Environmental Gradient Pag. 26

1.4. Small, Average, Tall
Quantitative Traits and Their Genetic Control Pag. 29

1.4.1. Set Your Clock By the Season
The Time of Bud Break Pag. 33

1.5. The Mark of Selection
Detecting Adaptive Divergence in Natural Populations Pag. 37

1.5.1 Genome Wide Association Study (GWAS) Pag. 40

1.5.2 “There Is Demography in My Sample!”
Confounding Factors Pag. 40

1.6. Actor, Setting, Time, Events Plot Pag. 44

1.6.1. Who: *Fagus sylvatica* L. Pag. 44

1.6.2. Where: Mont Ventoux Pag. 46

1.6.3. When: Last Glacial Maximum – 1860 – Today Pag. 48

1.6.4. What: Anticlinal Spring Phenology and Molecular Divergence Pag. 49

1.6.5. How: Demographic Expansion, Selection Pag. 50

1.7 References Pag. 51

2. **Aims** Pag. 69

3. **Chapter 1**: *Rolling down that mountain: microgeographical adaptive divergence during a fast population expansion along a steep environmental gradient in European beech* Pag. 72

3.1. Summary Pag. 73

3.2. Introduction Pag. 74

3.3. Material and Methods Pag. 78

3.3.1. Species, study site, and sampling Pag. 78

3.3.2. Genotyping by Sequencing Pag. 79

3.3.3. Diversity and differentiation statistics Pag. 80

3.3.4. Demographic model and simulation of genotypic data Pag. 81

3.3.5. Detection of divergence outlier loci between stand pairs Pag. 82
3.3.5.1. “Standard” (non-demography controlled) Bamova outlier search Pag. 83
3.3.6. Annotation of SNPs and genes Pag. 87
3.4. Results Pag. 88
3.4.1. Diversity and differentiation statistics Pag. 88
3.4.2. Demographic model and simulation of genotypic data Pag. 91
3.4.1. Detection of divergence outlier loci between stand pairs Pag. 93
3.5. Discussion Pag. 99
3.6. Conclusions Pag. 102
3.7. Acknowledgements Pag. 102
3.8. Author Contribution Statement Pag. 103
3.9. References Pag. 104

4. **Chapter 2**: *Genotype-per-Phenotype Association reveals polygenic architecture for spring phenology traits of European beech* Pag. 112
4.1. Summary Pag. 113
4.2. Introduction Pag. 114
4.3. Material and Methods Pag. 116
4.3.1. Species, study site, and sampling Pag. 116
4.3.2 Genetic data Pag. 116
4.3.3. Traits Pag. 117
4.3.2.1. Modelling the relationship BBCH ~ Julian Day Pag. 118
4.3.2.2. Phenological Score Sum Pag. 120
4.3.2.3. Degree days Pag. 121
4.3.2.4. Inter-year differences Pag. 121
4.3.4. Genotype-Phenotype Association algorithms Pag. 123
4.3.4.1. GWAS: design Pag. 124
4.3.4.1.1. Training datasets analysis Pag. 124
4.3.4.1.2. Validation in the testing datasets Pag. 125
4.3.4.2. GWAS design: main datasets analysis Pag. 125
4.3.4 Annotation Pag. 125
4.4. Results Pag. 127
4.4.1. Genotyping and phenotyping Pag. 129
4.4.2. Phenotype analysis Pag. 129
4.4.2.1. Logistic models of Julian days Pag. 129
4.4.2.2. Inter-annual differences in the phenological score patterns Pag. 129
4.4.3. Phenotype-Genotype Associations Pag. 132
4.4.3.1. Main datasets Pag. 132
4.4.3.2. Training/testing replicates Pag. 132
4.4.3.3. Genomic patterns Pag. 134

4.5. Discussion Pag. 135
4.5.1. BBCH ~ date of passage Pag. 135
4.5.2. Training/Testing *in silva* Pag. 136
4.5.3. Genomic architecture of beech spring phenology Pag. 137
4.5.3.1. Inferred genomic architecture Pag. 137
4.5.3.2. Genomic distribution Pag. 138
4.5.4. Pattern of allelic effects Pag. 139
4.6. Perspectives Pag. 140
4.7. References Pag. 141
4.8. Supplementary Materials Pag. 148
4.8.1. Supplementary Tables Pag. 148
4.8.1. Supplementary Figures Pag. 163

5. **Chapter 3**: *Genomic bases of leaf budburst predict countergradient patterns of quantitative genetic variation in European beech* Pag. 195
5.1. Summary Pag. 196
5.2. Introduction Pag. 197
5.3. Material and Methods Pag. 199
5.3.1 Population sampling and genotype data Pag. 199
5.3.2. Estimation of individual breeding values Pag. 199
5.3.3. Prediction of evolutionary potential Pag. 199
5.4. Results Pag. 201
5.5. Discussion Pag. 207
5.6. References Pag. 209
5.7. Supplementary Materials Pag. 212
5.9.1. Supplementary Methods Pag. 212
5.9.2. Supplementary Tables Pag. 233
5.9.3. Supplementary Figures Pag. 236

6. **General Discussion** Pag. 263
6.1. “When Action Grows Unprofitable, Gather Information”
Explicitly Addressing Demography Using Interdisciplinary Sources Pag. 266
6.2. Not All Tangles Ask for a Razor
Parsimony in Complex Scenarios Pag. 270
6.3. This Year’s QTLs
Distinct Architectures Drive Bud-Break under Different Climates Pag. 271
6.4. Bud-Breaking against the Gradient
Clinal Phenotypic Differentiation, Anticlinal Genetic Differentiation Pag. 274
6.5. Dancing of Architecture Pag. 279
6.5.1. Genomic Signatures of Putatively Adaptive Microgeaographic Divergence Pag. 279
6.5.2. Genomic Architecture of Spring Phenology Pag. 280

6.6. As the Heat Goes on Pag. 282
6.7. Springing Forward Pag. 284
6.8. References Pag. 287

Glossary of Abbreviations

AFD: allele frequency difference
asl: above sea level
BBCH: **B**iologische **B**undesanstalt, Bundessortenamt and **CH**emical industry; scale to classify different stages of phonological development
DD: degree day
DOY: day of the year
FDR: False Discovery Rate
GLM: Generalised Linear Model
GWAS: Genome Wide Association Study
JDoY: Julian Day of the year (see ***DOY***)
kya: thousand years ago
PSS: phenological score sum
QTL: quantitative trait locus
SNP: single nucleotide polymorphism
TBB: time of bud-break

1. Introduction

1.1 The Word for World Is Forest

Biology and Ecology of Trees

Trees are habitat constructors. They define and dominate forest ecosystems, which are estimated to cover 31 percent of the land area (4.06 billion, **Figure 1**) (FAO, 2020). The tree growth habitus has evolved multiple times in different plant lineages, so species sharing this set of traits have similar ecological roles, in spite of their loose phyletic relationships, and fill the same ecological niches. Being the most abundant species and composing the bulk of the biomass, trees are considered foundation species of forest ecosystems (Dayton, 1972; Baiser *et al.*, 2013), that is, they have a huge impact on their community. They are primary producers and stand at the center of the food webs that characterize these ecosystems (Ellison *et al.*, 2005; Ellison, 2019). Not only do they support the other species: they directly regulate the cycles of energy and nutrients through the entire system by capturing and distributing the sun's energy (Pokorny *et al.*, 2010), affecting downstream species abundance and composition and the physical, chemical and biotic interactions of the forest ecosystem. Trees assimilate the atmospheric carbon into the biomass, acting both as a carbon sink (Sedjo and Sohngen, 2012; Pan *et al.*, 2024) and a gateway for carbon into the food web (Valladares *et al.*, 2012); they take part to the water cycle by supplying the atmosphere with moisture and by affecting rainfall regimes locally, by the means of evapotranspiration (Pokorny *et al.*, 2010), by enhancing soil infiltration and, in temperate ecosystems, by improving groundwater uptake (Neary *et al.*, 2009).

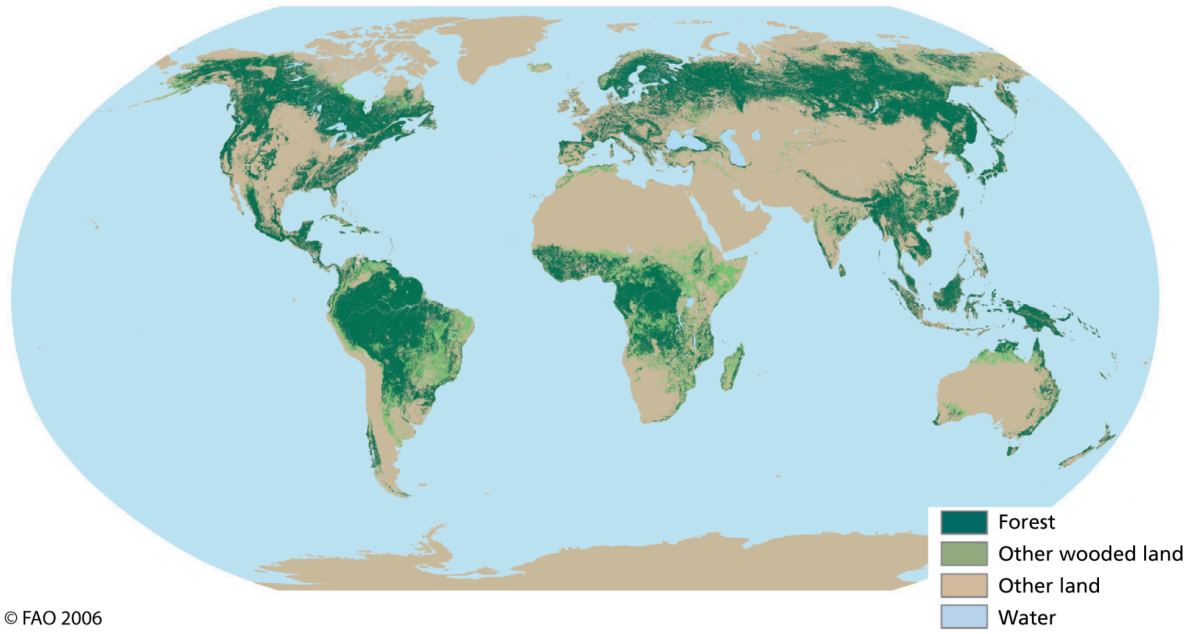


Figure 1: Distribution of forests and woodlands across landmasses (FAO, 2005).

Several features of the tree growth habitus - such as high lifetime fecundity, longevity, seed and pollen output, and the height from which seeds are released - contribute to maintaining large population sizes. Mortality peaks before maturity: for some species it has been estimated that one seed in one million sprouts successfully into a sapling that survives long enough to reach the tree growth habitus (Van Valen, 1975), but these individuals will be very likely to reproduce.
Trees are predominantly outcrossing and disperse seeds over a wide range of distances (e.g. 21 m in *Fagus sylvatica* L.*,* Bontemps *et al.*, 2013; 100 m - 3 km in *Fraxinus excelsior* L., Bacles *et al.*, 2005). In several tree species, populations produce seed and pollen in fluctuating patterns throughout the years, a phenomenon known as masting (Pearse *et al.*, 2016). Pollination contributes to the bulk of the gene flow: pollination from immigrant pollen has been reported in anemophilous species up to 10–30 km (*F. excelsior* and *Pinus sylvestris* L., respectively; Robledo-Anuncio and Gil, 2005; Bacles *et al.*, 2006). It follows that tree species have a wide geographical distribution where multiple, local populations are connected by gene flow both at short and large distances and constitute a larger metapopulation (Kremer *et al.*, 2012) with overall low levels of differentiation at neutral molecular loci (F_{ST}) even at large spatial scales (Petit and Hampe, 2006; Milesi *et al.*, 2024). Polymorphisms can be maintained for long periods of time, such as the case for trans-species polymorphisms shared by allopatric species of *Picea* spp that have diverged over 13 Mya (Bouillé and Bousquet, 2005).
Tree populations have large census population sizes (N_c): in the case of European beech (*F. sylvatica*) it is estimated that the 17 million ha covered by the species in central Europe corresponds to 1.5–2 billion adult individuals (Petit and Hampe, 2006). Such large distribution and the strong genetic connectivity may hinder defining whether a patch of trees within a forest constitutes a population in itself (Fady and Bozzano, 2021) and complicate the estimation of effective population size (N_e) to assess the evolutionary rate (Ryman *et al.*, 2019; Santos-del-Blanco *et al.*, 2022). Nonetheless, the random sampling of gametes (genetic drift) is buffered from eroding genetic variation in tree populations even across large time periods (Savolainen *et al.*, 2007); for example, Milesi *et al*. (2024) estimated stable or increasing N_e for seven tree species over the last glacial glacial cycles.
The high levels of within-population genetic diversity (Hamrick *et al.*, 1979; Nybom, 2004) and the preservation of tree populations' evolutionary potential over long time spans results in their ability to adapt to different environmental conditions in space (geographical range) and time (climatic fluctuations) (Petit and Hampe, 2006; Milesi *et al.*, 2024). On the one hand, the geographical range of many tree species spans over habitats with different climates and edaphic conditions (e. g. in Central Europe *F. sylvatica* populations are found in most undisturbed habitats; Magri, 2008); on the other hand, after the Last Glacial Maximum (11,700 yBP), several tree species have been able to persist or expand from their glacial refugia at a consistent pace (Aitken *et al.*, 2008), enabled by to both ecological and biological factors (Masek, 2001; Danby and Hik, 2007). Polymorphisms are spread across populations thanks to the intense gene flow. Neutral alleles are brought across the species range and supply local gene pools with new polymorphisms that may locally contribute to differences in reproductive performance and increase their occurrence in the following generation. Throughout the generations, the individual population becomes increasingly more able to thrive in its own environment - that is, it becomes adapted.

1.2 “Heart Is Where Home Is”
on (Local) Adaptation

In an evolutionary context, “adaptation” describes both the result of a process by which a population suits its environment and the process itself: it is “a feature [..] built by natural selection for the function it now performs” (Gould and Vrba, 1982) as well as “the acquisition within a population of such individual adaptations” (Simpson, 1953). There is no conscious nor individual effort to “improve” a trait toward an optimum: adaptation by natural selection occurs because, in a population, individuals carry different phenotypes for a given trait, each affecting differently the ability to survive and reproduce in different environments, and these phenotypes are heritable (Lewontin, 1970) (**Figure 2a**). The subject of selection is the population as a whole rather than the individual itself. Any habitat can sustain only a finite population, because its resources are finite: hence, the individuals that contribute more to the following generation are those whose phenotype(s), in a given environment, provides a higher rate of reproduction in comparison to those lacking it – either directly, by affecting their chances of reproduction; or indirectly, by improving the ability of these individuals to survive up to sexual maturity (Orr, 2010). Over generations, the differentially higher rate of reproduction a trait provides to its carrier within a given environment with respect to those lacking it leads to an increase of the frequency in the population of these phenotypes (**Figure 2b**) and the alleles that control its development (Monod, 1972) (**Figure 2c**).

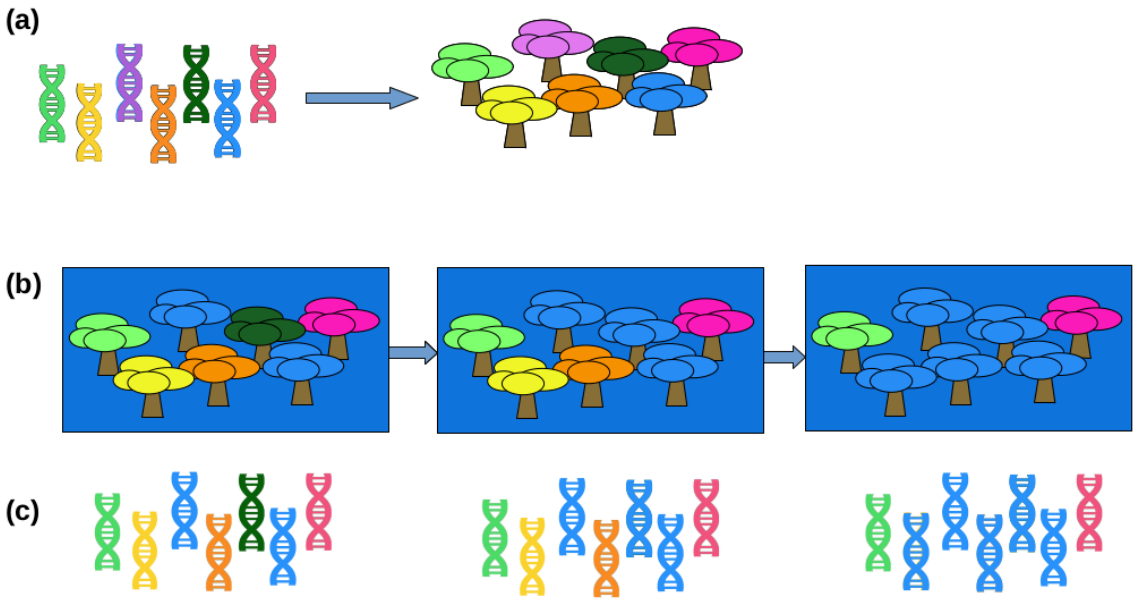


Figure 2: The process of natural selection. **(a)** Within a population, a genetically-controlled trait is expressed in multiple phenotypes, each driven by a different variant of the same gene. **(b)** In a given environmental condition, those individuals carrying a phenotype that result in a higher reproductive rate contribute more to the following generation; under a stable selective regime, the advantageous phenotype increases in frequency in the population as **(c)** well as the genetic variant that controls it.

The value of a phenotype, *P,* is conceptualised as $P = G + E + GxE$: the relationship between its genetic component *G*, the environment *E*, and the interaction between genotype and environment *GxE*. The same genotype can express multiple phenotypes for the trait it controls in response to a change in the environment during the lifetime of an individual (Scheiner, 1993) (**Figure 3a)**. This plastic change can either improve the individual's performance in that environment or be maladaptive for the interaction of evolutionary (*e.g.* gene drift, gene flow) or environmental factors (*e.g.* stress, nutrient limitation) (Merilä and Hendry, 2014). The plastic response of the phenotype arises by an "instantaneous" interaction between the environment and the developmental process of an individual without a change in genotype(s) controlling the trait (**Figure 3b)**. The environmental cues that drive the change of the phenotype do not cause change of allele frequencies in the following generation; hence, characteristics acquired during an organism's lifetime are not heritable (**Figure 3c**) (but see Pigliucci *et al*., 2006 for how phenotypic plasticity is being integrated into an evolutionary framework). Instead, the ability of undergoing a plastic change is a heritable trait in itself (Scheiner and Lyman, 1989); it has an adaptive value if changing phenotype to match the new conditions improves the reproductive rate (Gavrilets and Scheiner, 1993).

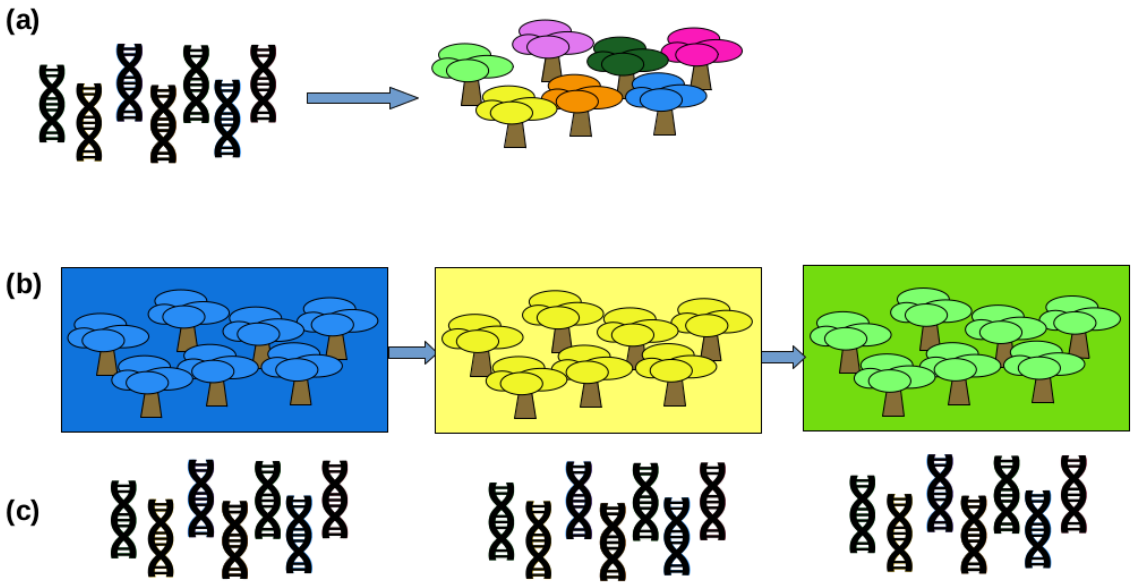


Figure 3: The mechanism of phenotypic plasticity. **(a)** The same genotype expresses multiple phenotypes depending on the environmental conditions in which the individuals carrying it live. **(b)** Changes in the environment over ecological times drive the phenotype of the individuals carrying that genotype to change as a response, **(c)** the genotype remaining the same in frequency.

Evolution is a small scale matter. The adaptive value of a trait is conditional to the very environment where a population has evolved (adaptiveness *sensu* Dobzhansky 1968). Environmental factors vary across space; a species will meet heterogeneous selective pressures across its spatial distribution; between two habitats with distinct environmental characteristics, a given trait may have different states which provide the highest fitness (optima). When two identical populations live in habitats that differ for phenotypic optima, over generations the spatial heterogeneity of the selective pressures drives each of them toward developing divergent phenotypes even if they exchange migrants, becoming each locally adapted to their own environment (Kawecki and Ebert, 2004; Savolainen *et al.*, 2007; Yeaman, 2022). In this way, local adaptation maintains high levels of genetic diversity (Hedrick *et al.*, 1976; Felsenstein, 1976; Hedrick, 1986) and, on a broader scale, contributes to the inception of the divergence of novel species (Turelli *et al.*, 2001; Via, 2001).

Kawecki and Ebert (2004) identified several demographic and ecological conditions under which local adaptation can evolve within the demes of a metapopulation. First, no single genotype must have the highest fitness in all the habitats. A genotype that drives multiple phenotypes for the same trait to local optima under different environmental conditions will lead to phenotypic differentiation without local adaptation (Pimm, 1979; Kingsolver *et al.*, 2002); instead, local adaptation comes with a local peak of fitness and a drop in fitness outside the home environment. Second, selection must be locally strong enough to cull migrants from neighbouring demes, but not so strong to disrupt genetic connectivity. A rapid drop in the fitness of the migrants' phenotype due to the mismatch with the local optimum promotes a scenario with genetically differentiated, locally specialised phenotypes (Ronce and Kirkpatrick, 2001; Kisdi, 2002) and a moderate selection against intermediate genotypes fuels local diversity (Spichtig and Kawecki, 2004). Third, there should be restricted gene flow between demes. On the one hand, gene flow opposes directional natural selection and thus homogenises allele frequencies between neighbouring demes; on the other hand, it brings new alleles and prevents genetic isolation (Lenormand, 2002). Fourth, small temporal heterogeneity in the local selective regime deters the immigration of phenotypes lacking specialisation for the deme they invade. In fact, generalist phenotypes successfully spread within a metapopulation whenever the local optima change over time (Kisdi, 2002; Dittmar and Schemske, 2023)

Due to the combination of their extended life cycle, late sexual maturity, and sessile growth habitus, trees also face environmental pressures that vary over time and have high levels of gene flow (Petit and Hampe, 2006). Furthermore, adaptation is hard to assess when plants grow in their own habitat because of the difficulty to tell how much of the phenotypic variation is controlled by genes and how much is driven by the environment (Silvertown, 1989). In spite of these characteristics, tree populations have been observed to have adapted to their own home environment. This is often observed by a decoupling in the genetic variances for loci controlling for quantitative traits (Q_{ST}) and neutral loci (F_{ST}), where the former can exceed the estimates the estimates of the latter (Whitlock, 1999; Mckay and Latta, 2002; Le Corre and Kremer, 2003, 2012a; Kremer and Le Corre, 2012a). Transplant experiments allow a direct assessment of the degree of adaptiveness of a population to its home habitat by growing multiple provenances together at experimental site(s) under controlled conditions to even out the effects of the environment. (Kawecki and Ebert, 2004; Savolainen *et al.*, 2007). The performance of a local

and a foreign provenance is compared based on traits that work as proxies for fitness. Transplant experiments can either test the performances when a provenance is grown in its own home habitat together with other provenances (reciprocal transplant: *in situ* or "local *vs.* immigrant", **Figure 4a**); or growing several provenances across multiple habitats (common garden: *ex situ* or "home *vs.* away", **Figure 4b**) (Savolainen *et al.*, 2013; De Villemereuil *et al.*, 2016; Johnson *et al.*, 2022). The two designs emphasise distinct facets of local adaptation: *in situ* transplants compare several provenances within the same habitat (deme: variable; habitat: constant), expecting that the local provenance is adapted to its home environment if it performs better than foreign provenances; *ex situ* transplants focus on the performance of a provenance across several habitats (deme: constant; habitat: variable), and the expectation for local adaptation is that a provenance has an higher fitness in its own habitat than when it's grown in any other. Whilst reciprocal transplants are the smoking gun for local adaptation, the replication of multiple common gardens allows to discriminate the observed pattern from other *deme x test* interactions (Kawecki and Ebert, 2004).

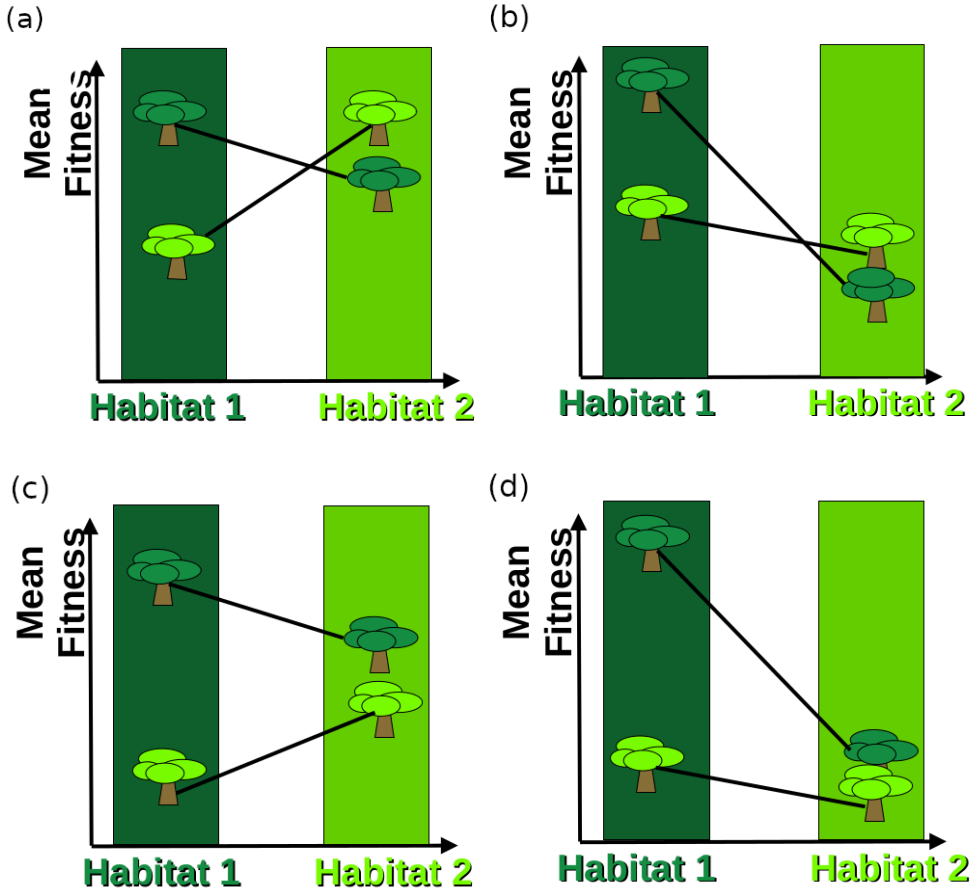


Figure 4: Transplant experiments. **(a-b)** Local *vs.* Immigrant. The **local** and **foreign** provenances are grown in the home environment of the former; the **local provenance** exhibits a higher fitness than the immigrants in its own environment, and its fitness drops when grown under different conditions. **(c-d)** Home *vs.* away. Several provenances are grown under multiple conditions. Even though a **provenance** performs better in its own habitat, it may still outcompete **other provenances** when grown in different habitats. Scenarios **a-c** suggest that the local provenance is locally adapted, the scenarios depicted in **c-d** points to a more complex deme x habitat interaction (adapted from Kawecki and Ebert, 2004).

Forestry has been resorting to common garden and transplant experiments for over 200 years to identify the most productive provenances for the inception of a new site or, more recently, for assisted migration (Langlet, 1971; Aitken and Whitlock, 2013; Aitken and Bemmels, 2016). The patterns assessed by the transplant vary across populations and species: from putatively genuine local adaptation (Sork *et al.*, 1993; Sáenz-Romero *et al.*, 2017), to "exaptive" life strategies in spite of the provenance (Vizcaíno-Palomar *et al.*, 2014), or differentiation without adaptation in the home or an environmentally similar habitat (Tíscar *et al.*, 2018; Ramírez-Valiente *et al.*, 2021). As a downside, adaptation may take time to be functional after the inception of the transplanted trees; so transplant experiments will reap what they sow about twenty years after their inception: "*agricola oleas non sibi plantat*" (Germino *et al.*, 2019). Genomic tools may be better suited to infer *putatively* adaptive patterns of genetic differentiation (Lepais and Bacles, 2014). Moreover, when the selective pressures - most notably, the climate - change over short timescales, transplant experiments established today may "arrive too late" when it comes to extrapolate predictions about the future of the forests.

1.2.1 Adaptation at the Time of Climate Change

Adaptation takes time. It's not necessarily *fast* in terms of absolute time, but it takes at least one *generation* to observe a change in the allele frequencies - and trees have a late sexual maturity. Pollen fossils, charcoal and genetic data show that at the end of the Würm glaciation (Last Glacial Maximum, 11,700 yBP), tree species were restricted to small refugial areas from which they progressively spread out and (re)colonised the now-iceless continents. The rate of advance has been estimated to be 100 m / y (Aitken *et al.*, 2008).

We are living in a progressively warmer world. Since the 1950s yearly temperatures have increased globally (**Figure 5**; Mann *et al.*, 1999). According to the Copernicus Climate Change Service, each individual year of the last decade has been the "hottest year on record [since 1850] globally"; in increasing order: 2015, 2019, 2020, 2016 and 2023 (Copernicus CCS, 2022, 2023). The year 2024 has been the warmest on record, with a global increase of 1.6 °C in comparison to the pre-industrial level (Copernicus CCS, 2024). Anthropogenic emissions of greenhouse gases in this time frame account for 40 % of the warming. Together with heatwaves and droughts, there comes also warmer winters and extreme climatic events like floods. Projections estimate a shift of near-surface temperature of 80-430 m / y, displacing northward the ecological niches (bioclimatic envelope) by 300-800 km by 2100 (Loarie *et al.*, 2009).

The climate-driven shift of local optima appears to jeopardise the permanence of tree populations adapted at the landscape scale, especially those at the edge of their species' distribution. Ecological and/or geographical marginal populations in fact have lower immigration rates from the edge to the core of the species distribution (Gaston, 2009) and, though they are not necessarily more prone to inbreeding, the harsher conditions at the edge makes them more susceptible to extinction (Birch, 1957; Theraroz *et al.*, 2024). Aitken *et al.*, (2008) reviewed three possible outcomes for the fate of tree populations under the current climate change scenario.

Trees may migrate. Paleobotanical surveys and phylogeographic reconstructions based on chloroplast DNA agree that tree species have expanded and contracted their range tracking the global climatic cycles (Davis and Shaw, 2001; Petit *et al.*, 2004; Magri, 2008) and long-distance dispersal events are far from rare (Kremer *et al.*, 2012). Alleles from populations adapted to harsher climates may migrate to the core area of their species distribution, which is expected to experience the conditions today found at the margin of their range, while a significant shift is expected in altitudinal treelines (Parmesan, 2006). However, the unprecedented velocity of the current warming coupled with habitat fragmentation appears to outpace the migration abilities of tree species (Anderson *et al.*, 2012; Zhu *et al.*, 2012).

Trees may adapt. Their high level of standing genetic diversity may allow for the evolution toward the new optima. The prolongation of the vegetative season that is being observed as a consequence of climate warming may provide a larger reproductive output - positive correlation has been found between canopy duration and rate of fruit production in *F. sylvatica* (Journé *et al.*, 2021). Multiple factors beside standing genetic variation contribute to a populations' ability to track changes in local optimum; even though trees have high fecundity rates, and large effective population sizes, environmental stochasticity (*ie*. extreme events) and strength of selection in the recruiting phase may hamper their adaptation to the new conditions (Lynch and Lande, 1993; Alberto, Aitken, *et al.*, 2013). More in general, failing to track the new optimum can lead locally adapted populations to find themselves in an habitat for which their current standing genetic diversity is detrimental (Rellstab, 2021). The concerns already expressed about tree

migration potential also stand for their adaptive potential over *short* times to *large* shifts (Jump and Peñuelas, 2005). In face of contrasting signals, forestry practices risk to either overestimate the adaptive potential of tree populations, leading to sharp decline; or to underestimate it, and rather implement unnecessary management decisions.

Trees may go locally extinct. It is still uncertain how the advancing of spring phenology will impact tree communities, because it anticipates the vegetative season and at the same time exposes the trees to late frost damage. Coupled with heatwaves, drought is expected to be a major hazard at the edge of the distribution of species that require a high water availability, resulting in increased mortality (Jump, Hunt, and Peñuelas, 2006). Extinction risks are predicted to increase as the warming goes on, both at a continental and at local scale (Cahill *et al.*, 2013; Urban, 2015) and their consequences will trickle down toward the entire forest ecosystem that trees define (García-Valdés *et al.*, 2018).

To understand whether trees will be able to adapt to a warmer climate, research must push to investigate the mechanisms at the lower limit of adaptation—spatial (in environments where the habitat conditions change over short distances) and temporal (over a few generations).

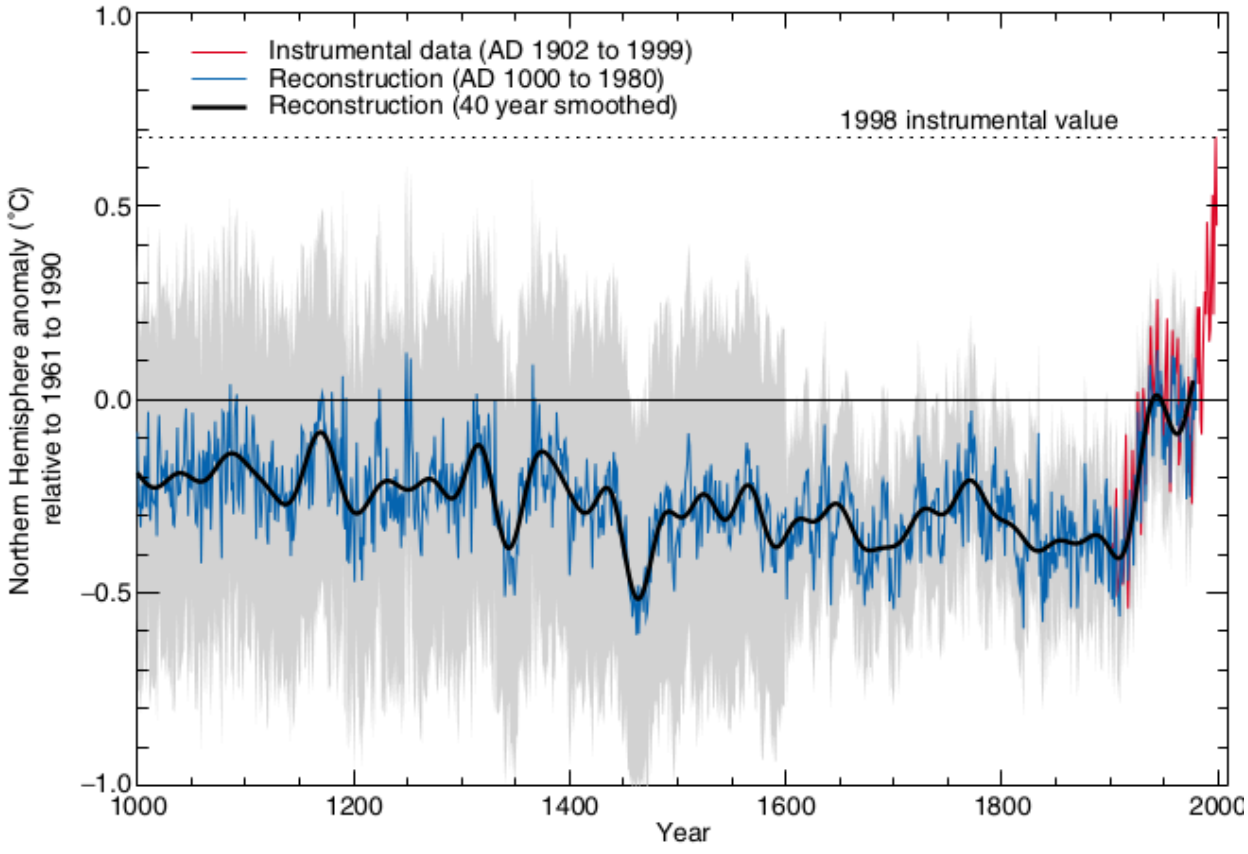


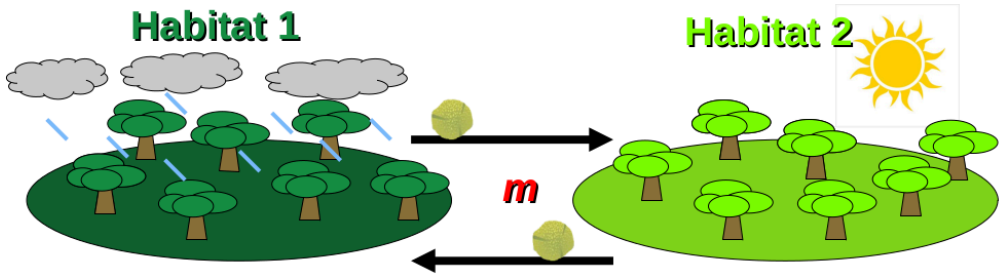
Figure 5: Trend of temperatures in the northern hemisphere in the last 2000 years: **observed time series** (1902 - 1999), **reconstructed trend** (1000 - 1980), **reconstructed trend smoothened on a 40-year average**, **confidence interval** (Watson *et al.*, 2001).

1.3 How Much Spatially Fine Is a Fine Scale

Local Adaptation along Environmental Gradients

A natural obstacle that disrupts the gene flow in a metapopulation facilitates each subpopulation to evolve adaptive genetic differentiation by tracking its local optimum. Whenever environmental conditions change across space as a continuum instead of discrete patches, that is a gradient. Along a spatial gradient, the change of a main variable drives a change in other correlated ones: climbing up a mountain (altitude), it becomes colder (temperature), water condenses (air humidity), and atmospheric pressure lessens (Körner, 2007); swimming downstream in a river, freshwater becomes first brackish then salty, current speed slows down, and so does the quantity of suspended organic matter.

A population that lives along a gradient will meet a continuous turnover of environmental pressures, which generates a series of phenotypic optima - each one an habitat distinct but within a continuum toward the neighbouring habitats. Spatial clines of phenotypic variation are often interpreted as the result of a selective process occurring along the chain of phenotypic optima along the gradient; in this case they reflect clinal patterns of allele frequencies at loci controlling the selected trait and evolve following the migration-selection equilibrium. The ability of a deme to undergo local adaptation depends on the relationship between migration and selection (**Figure 6**). While selection makes subpopulations diverge from each other toward their local optima, migration opposes the directional change in allele frequency. When allele A_1 has a selective advantage s over allele A_2 in *Habitat 1* and a disadvantage αs in *Habitat 2*, both are advantageous only in their own environment when the environmental effect (α) of A_1 is antagonistic ($\alpha > 0$). The polymorphism is maintained under a migration rate m when $m / s < \alpha / |1-\alpha|$, that is, selection strength exceeds the effect of migration either for low immigration rates (*e.g.* large spatial distance between demes) or under strong environmental pressures - when individuals with a trait state different from the optimum are selected against. Otherwise, gene flow leads to the fixation of the allele conferring the highest fitness across the demes (Bulmer, 1972; Lenormand, 2002; Tigano and Friesen, 2016).



A_1	$1 + s$	$1 - \alpha s$
A_2	1	1

Figure 6: Migration-selection balance. A_1 has a home advantage over A_2 when the ratio of migration between demes and selection strength does not exceed the environmental effect of A_1 outside its own environment; that is, the environmental effect on A_1 is synergic in **Habitat 1** and antagonistic in Habitat 2. (adapted from Bulmer, 1972; Lenormand, 2002).

Along an environmental gradient, there are as many values of α as the phenotypic optima; in this scenario, the environmental effect of A_1 (α) is better described as a function that describes the variation of selection intensity as a function of space. Clinal variation in the allele frequencies along a gradient will evolve when the spatial rate of change of the environmental conditions is less than a "characteristic length" l_c. Such distance depends on the level of dispersal of the species l and the selection strength s: $l_c = l / \sqrt{s}$ (Slatkin, 1973; Endler, 1977; Forester *et al.*, 2016). For adaptive differentiation to evolve, the gradient may span over much shorter distances than the continental scale: when the environment changes steeply enough within the "characteristic length" of a species, demes are able to receive migrants and the mean value of a trait to respond to the local selective conditions (Slatkin, 1973, 1978).

The relationship between spatial variance in traits and environmental variables is better understood by expanding the phenotype formula of quantitative genetics to include the covariance between genotypic and environmental sources of variances: $P = G + E + GxE + Cov(G,E)$ (Walsh and Lynch, 2018). The term $Cov(G,E)$ expresses to which degree the genotypes affecting the trait expression are non-randomly distributed among environmental gradients of a variable that influences that very trait (Conover and Schultz, 1995; **Figure 7**). A positive covariance (*co*-gradient variation) means that, in example, genotypes for frost avoidance occur in cold environments, suggesting that the environment and selection push the value of the trait in the same direction. Instead, a negative covariance (*counter*-gradient variation) means that *G* and *E* push the phenotype in opposite directions: in this case, genotypes for a vigorous growth are found in cold environments (Conover and Schultz, 1995). Countergradient patterns of phenotypic or genetic variation along environmental gradients are often interpreted as signatures of antagonistic selection on the trait and often as a long term compensation for maladaptive plasticity (Grether, 2005; Crispo, 2008). These patterns are not clear *in silva*, as the plastic response may mask the effect of the genotype (Conover and Schultz, 1995) and across heterogeneous environments gene flow is often non-random (Soularue and Kremer, 2012).

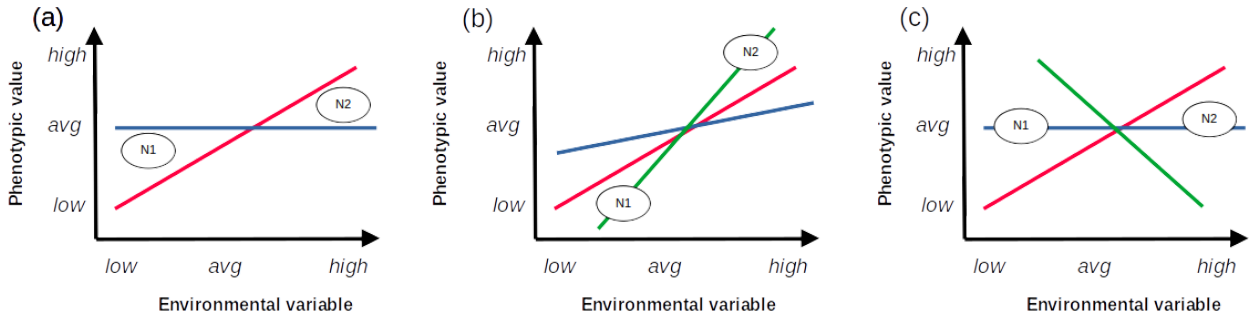


Figure 7: Different spatial **patterns of phenotype variation** depending on the **patterns of genetic variation** under the same **environmental gradient**. (**a**) Random distribution of the genotypes with respect to the environment variable, possibly due to strong plasticity; lack of covariance between environment and genotype(the pattern of phenotypes is here included with the collinear environmental gradient). (**b**) Co-gradient variation, $Cov(G,E) > 0$; the genetic and the environmental component of the phenotype push the trait in the same direction distribution. (**c**) Counter-gradient variation, $Cov(G,E) < 0$: along the gradient, the genetic and the environment push the phenotype in opposite directions (adapted from Conover and Schultz, 1995).

Populations along a gradient may differ in the state of a genetically-controlled trait in spite of being located well within their dispersal distance, where migration is expected to drive the equilibrium. Richardson *et al.*, (2014) defined this instance *micro*geographical adaptation, where demes exchange migrants yet retain a fitness advantage in their patch so that migrants are selected against and the demes diverge from each other. They expanded the definition of "characteristic length" to a radius of two standard deviations from the dispersal kernel of a species (**Figure 8**). Depending on the species, this distance can be as low as 100 m, as observed in transplant experiments of *Agrostis tenuis* Sibth. populations between an abandoned lead mine and a free pasture (Bradshaw, 1952, 1960). Two questions arise as the spatial scale as adaptive divergence is understood to be even narrower than previously thought: What is its lower spatial boundary? Which mechanisms contribute to its evolution?

Cases of microgeographic adaptive divergence have been reported in several species, such as salamanders (Richardson and Urban, 2013), riverine (Fitzpatrick *et al.*, 2015) and riverine-lacustrine fish (Aguirre, 2009; Weber *et al.*, 2017; Maciejewski *et al.*, 2020; Bal *et al.*, 2021), lizards (Rosenblum, 2006), annual plants (Antonovics, 1968; Turner *et al.*, 2010), and palms (Savolainen *et al.*, 2006; Babik *et al.*, 2009). In tree species, molecular signatures of adaptive divergence have been detected between subpopulations located at distances as short as less than the kilometre in distinct patchy habitats (Audigeos *et al.*, 2013; Brousseau *et al.*, 2015). In particular, altitudinal gradients display a marked environmental variation over short spatial scales, minimising factors such as biogeography and genetic isolation that confound the identification of adaptive signatures. Furthermore, the interplay of strong gene flow and the strong divergent local selection results in a discrepancy between molecular differentiation at neutral and quantitative loci (Savolainen *et al.*, 2013; Halbritter *et al.*, 2018).

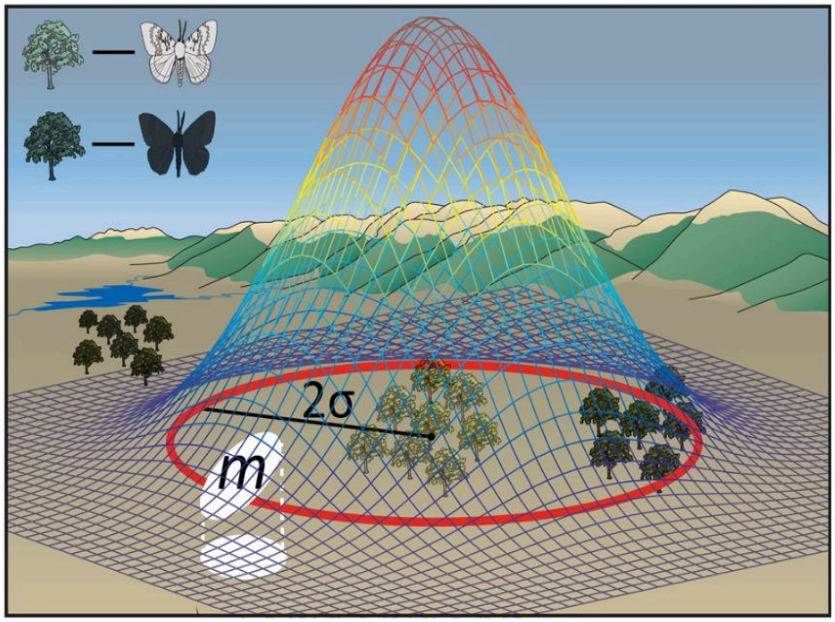


Figure 8: Microgeographic adaptation. When the environment changes within 2σ of the moth's dispersal distance and the deme within the same radius differentiates from other demes, it can be deemed to be microgeographic. Differentiation occurring outside this range is instead deemed to be "local". Migration can still occur at a rate *m* between the habitats. (Adapted from (Richardson *et al.*, 2014).

1.4 Small, Average, Tall
Quantitative Traits and Their Genetic Control

Quantitative traits are traits that display continuous variation among individuals in natural populations and their expression is influenced both by environmental and genetic factors. An extreme case of hyper-diffuse genetic control of a quantitative trait is height in humans (Boyle *et al.*, 2017; Conery and Grant, 2023), which, in addition to having a very complex genetic architecture, changes significantly between populations. Identifying the causal genes has been a conundrum because the trait appears to be associated with different loci in different populations. Examples of quantitative traits in trees are the timing of bud-break, the seed output, the diameter of the trunk, and adult height. The proportion of phenotypic variance of a quantitative trait in a population that is ascribed neither to environment or genotype x environment interaction is its heritability. In quantitative genetics it's useful to distinguish H^2, broad-sense heritability from h^2, narrow-sense heritability. H^2 represents all genetic contributions to phenotypic variance and encompasses additivity, dominance, epistasis, and parental effects (*ie.* influence of the parental phenotype on the offspring phenotype, such as seed size); it is the ratio between genetic variance and phenotypic variance, *var(G) / var(P)*. Instead, h^2 captures the contribution to phenotypic variance due solely to the additive effect of alleles driving the phenotype *var(A)* : *var(A) / var(P)*; it is the genetic component responsible for the parent-offspring resemblance.

Multiple genes contribute to driving the phenotype at quantitative traits, each of small effect size; the more loci underlie the trait, the larger the number of possible genotypes contributing to it. These loci are thus named Quantitative Trait Loci (QTL), regions of the DNA spanning over one or several genes and that are associated with a given phenotype; taken all together, the QTLs associated with a trait constitute its genetic architecture (**Fig 9a)**. The additive genetic effects, together with the environmental effects and the plastic responses (genotype x environment interactions) results in a large degree of phenotypic variation of quantitative traits (Fisher, 1919). The spectrum of allele effect sizes among the QTLs controlling a trait is roughly L-shaped, with multiple variants having small effect on the phenotype and few loci with large impact (**Fig 9b)**. Due to this continuum, the inheritance of loci with large effects is modeled in a Mendelian fashion, while small-effect loci segregate as quantitative genetic variation: their effect is scattered across multiple loci across the genome and the contribution of each locus to the phenotypic variance is understood to be additive (Lynch and Walsh, 1997).

QTLs are mapped across chromosomes using line-cross, pedigree analyses and inbreeding experiments (or genealogical analyses in humans) together with molecular markers (nowadays, typically Single Nucleotide Polymorphism, or SNP). An individual QTL can cover more than one causal variant: its location pinpoints a genomic region harbouring the linked causal polymorphism(s) that influences the phenotype. The physical location of QTLs along a chromosome can bias the estimate of the effect of the loci because regions of the genome that are inherited together may harbour several loci influencing the trait (Lynch and Walsh, 1997).

Identifying the QTLs responsible for trait determination and variation is a shared goal between quantitative and population genetics, as well as telling how much of the standing phenotypic variation in a population is driven by genetic and environmental factors. At the population level, the genetic component of phenotypic variance determines the rate at which a given trait responds to selection. The environmental component instead shifts the phenotypes of the individuals from their (theoretical) genetic value, thus reducing the efficacy of selection. Due to the effect of spatial environmental heterogeneity, the heritability for quantitative traits in tree natural populations has been estimated to be moderate (Howe *et al.*, 2003).

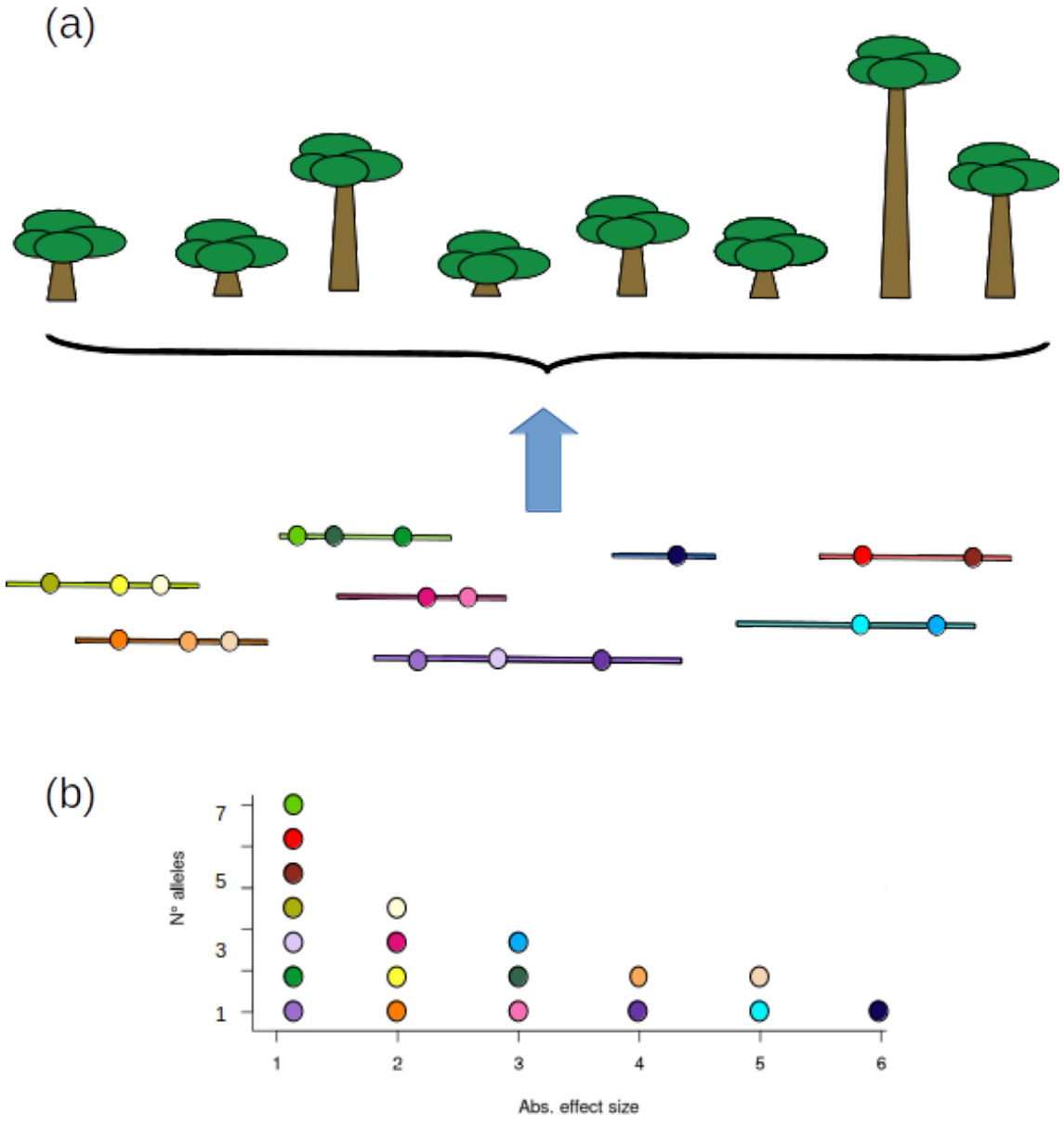


Figure 9: Polygenic bases of quantitative traits. **(a)** Quantitative traits are controlled by multiple Quantitative Trait Loci across the genome balance. **(b)** The distribution of allelic effects of the loci driving a quantitative trait are often enriched in the class of the small-effect size.

The diffuse architecture of quantitative traits also implies that a population can accomplish large changes in a trait's values by small shifts of allele frequencies at multiple loci. Two populations experiencing selection toward the same optimum but under different selective pressures may reach it with distinct combinations of QTLs (**Figure 10**), as observed in trait-agnostic studies in landscape genomics (Feulner *et al.*, 2015; Rellstab *et al.*, 2017). Studies have reported large heterogeneity in the effect size of QTLs driving adaptation at quantitative traits, from few, large effect QTLs (*e.g.* flower architecture in *Erythranthe guttata* (Fisch. DC.) G.L.Nesom, Bradshaw *et al.*, 1998), to many, small-effect QTLs (*e.g.* cold hardiness in conifers, Howe *et al.*, 2003). These scenarios represent the extremes, where adaptation has evolved either from big sweeps (Stephan, 2019) or many small shifts from standing genetic variation (Rockman, 2012).

The architecture of local adaptation will depend on the specific relationship between migration, selection, and drift in a population. The effect of selection on a phenotype on its architecture depends on the individual effect size of the individual QTL (Yeaman, 2022). In a scenario of adaptation with gene flow, selection may be strong enough to drive population toward adaptive phenotypic divergence, but an individual allele of small effect, contributing little to local adaptation, will be preserved only if migration happens below a critical threshold rate. Above this threshold, two non-mutually exclusive mechanisms can take place: the consecutive recruitment of beneficial alleles of small effect in tight linkage with each other, so that while their individual contribution to divergence is small, the global QTL behaves like a locus with large effect; or an immigrant allele with large effect may replace several small-effect alleles by approximately matching their combined contribution to the adaptation while gene flow erodes the otherwise sparse architecture (Yeaman and Otto, 2011; Yeaman and Whitlock, 2011). In the first phases of divergence, the local optima are tracked by combinations of alleles at different loci at different populations; only afterward the signals become detectable as changes in allelic frequencies (Le Corre and Kremer, 2012; Kremer and Le Corre, 2012). The result is that, over generations, a locally adapted polygenic trait that evolves under migration-selection balance will be controlled by few, tightly linked alleles of large effects, concentrated in small clusters across the genome (Yeaman, 2022).

Traits with a strong link to the survival and reproductive output are often quantitative (Le Corre and Kremer, 2012); their complex architecture allows for different adaptive phenotypes to evolve in different environmental contexts (but see Barson *et al.*, (2015) for a major effect locus that determines earlier sexual maturation in male individuals of *Salmo trutta* Linnaeus, 1758). The study of quantitative traits started from animal husbandry and crop breeding, fields where inbreeding experiments target a specific value of a phenotype in a controlled environment to select lineages that maximize that phenotype. The pedigrees of the selected individuals are often available and explicitly taken into account when designing experiments. Wild populations come with none of these advantages because little or no pedigree information is available. Whilst artificial selection explicitly targets a trait, in wild populations researchers must infer which trait and value of that trait is the target of selection by looking at changes in candidate traits within and/or across generations. Furthermore, environmental factors are neither controlled nor it’s clear which affects a trait directly or indirectly because of phenotypic correlation with a selected one (*e.g.* selection gradients; Lande, 1980), thus confounding the effects of genes on the trait (Walsh and Lynch, 2018). In trees this is complicated by long generation times, late sexual maturity, and the need for multiple vegetative and reproductive cycles over many years

for traits to be fully expressed, so that it can take decades to observe the effects of a round of selection on adult individuals (Abbott *et al.*, 2015; Germino *et al.*, 2019).
Nonetheless, it is possible to observe a great degree of phenotypic variation for fitness-related traits in natural populations, especially along gradients; a straightforward case is the time of bud-break (TBB) across a latitudinal gradient, such as in conifers (Mikkola, 1982; Alberto, Aitken, *et al.*, 2013). Along altitudinal gradients, gene flow is intense between stands at different elevations; nonetheless, traits such as spring phenology (Davi *et al.*, 2011; Gaüzere *et al.*, 2020) or strategies for drought avoidance (Bontemps *et al.*, 2017) are differentiated. In particular, the the time at which beech trees flush their leaves has a both a marked inter-annual variation due to the yearly climatic conditions (Badeck *et al.*, 2004; Arnold *et al.*, 2019) and, at the scale of an individual mountain, it can span over a time window as broad as that of populations along a spatially wider latitudinal gradient (Vitasse *et al.*, 2009).

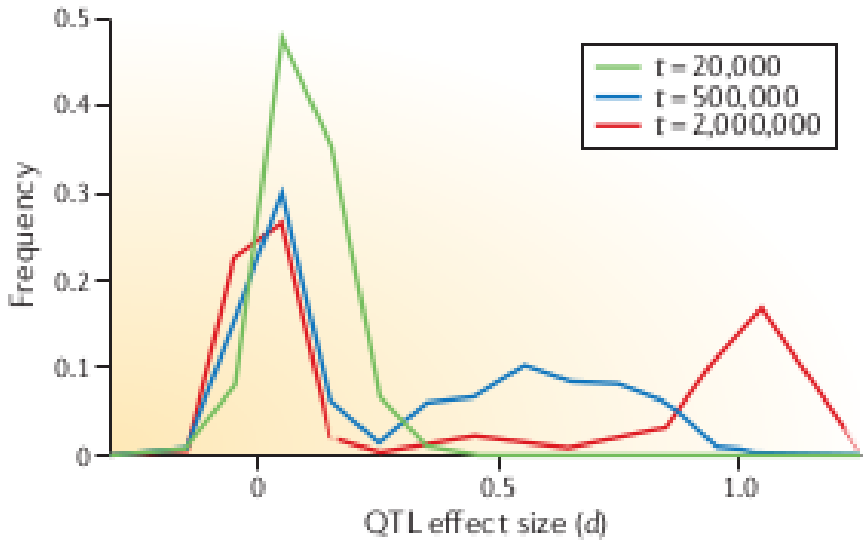


Figure 10: Evolution of the architecture of a quantitative trait in two populations under migration-selection balance. The different colours represent the distribution of allelic effects (*d*) after *t* generations. In the **early stages**, many loci scattered across the genome drive the trait. **Over time**, gene flow replaces much of this sparse architecture with fewer loci of larger but comparable effect; they are more likely to fix in linkage with other loci driving the trait. **After enough generations**, the architecture will be composed by few loci of large effect, the effect of which is comparable to the many other small-effect loci. (Adapted from Yeaman and Whitlock, 2011; Savolainen *et al.*, 2013)

1.4.1 Set Your Clock by the Season

The Timing of Bud-Break

Vegetative spring phenology represents the cyclical events that regulate leaf development and unfolding of leaves; since it tracks the local climate, it contributes to the length of the growing season, to fecundity, and to water use efficiency (Battey, 2000; Rowland *et al.*, 2020). In several boreal and temperate tree species (e. g. European beech) vegetative and reproductive phenology are also synchronous (Gaüzere *et al.*, 2013; Oddou-Muratorio *et al.*, 2024). The timing of vegetative spring phenology is a tradeoff between genetic and environmental control: while the genetic regulation has evolved through generations of trees undergoing historical climatic regimes, phenotypic plasticity is an “instantaneous” response of the phenotype expressed by a certain genotype to the current season. The time at which the buds open is pushed earlier or pulled later in the year by temperature cues toward a longer vegetative season and the avoidance of spring frost. The two factors weigh differently on the plastic and the genetic component of the trait, and the phenotype expressed each year is a tradeoff between these two pressures (Gömöry and Paule, 2011). In literature the process has many names: leaf flushing, leaf-out, bud break, bud-burst (TBB), or a general “spring phenology”. In literature, the progression of bud-break is described with the stages 1-10 of the BBCH scale (**B**iologische **B**undesanstalt, Bundessortenamt und **CH**emische Industrie), a decimal score divided into principal and secondary growth representing different phenological processes of mono- and dicotyledonous plant species (Meier, 2001; Finn *et al.*, 2007), from bud-set to bud-break, flower and fruit development, and leaf senescence (**Figure 11**). Representing TBB as a longitudinal trait better describes its dynamic.

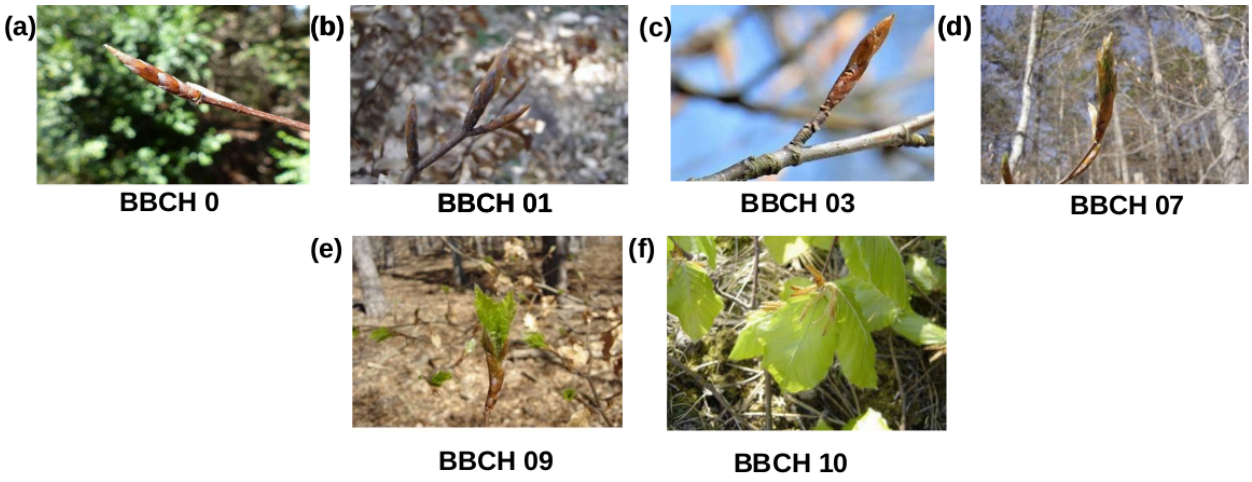


Figure 11: BBCH stages of bud-break European beech. **(a)** BBCH 0: all buds in winter dormancy; **(b)** BBCH 01: most leaf or mixed buds are swelling; **(c)** BBCH 03: most leaf or mixed buds have finished swelling and are ready to break and in some species the scales may have changed colour; **(d)** BBCH 07: espectively at least 10 % and 50 % of the leaf or mixed buds have break (the tips of the leaves should appear at the distal extremity of the bud, with the scales spread out; **(e)** BBCH 09: respectively at least 10 % and 50% of the buds have the first leaves begin to individualise: the tips of these first leaves are clearly visible and are taller than the tip of the scales, if these are still present; **(d)** BBCH 10: first leaves are spread out (Jean *et al.*, 2023); © Fréderic Jean and Oliver Gilg.

Plant phenology responds to cyclical environmental signals, chiefly changes in temperature and photoperiod (**Figure 12**). The combination of the different cues at a given time triggers different mechanisms, so that a tree will respond differently to the same temperature in fall (shorter photoperiod) and spring (longer photoperiod). In temperate perennial species, the leaf primordia differentiate in summer and by the end of the vegetative season the shortening of the day makes them enter in a state of arrested growth (dormancy) (Anderson *et al.*, 2010; Shim *et al.*, 2014). As the photoperiod further decreases and temperatures fall as well, internal signals stunt growth of the bud, which enters into *endo*dormancy. Genetic factors control the length of endodormancy; within this state, the growth of the bud is suppressed even when environmental conditions become again favourable to growth, like in the early-middle fall. Ecodormant buds are insensitive to positive changes of temperatures, preventing the leaf from unfolding in a time when frost is most likely to happen, depleting the resources of the plant (Anderson *et al.*, 2010). During endodormancy chill accumulation takes place (*ie.* time of exposure to temperatures between 0 and 5-7.5 C) which controls the release of winter arrest (Abbott *et al.*, 2015).
As the endogenous control ceases, the buds transition to ecodormancy, a suppression of growth controlled by environmental cues. Buds that haven been exposed to low temperature for long enough start accumulating warming. In late successional species, photoperiod works as an additional cue in avoiding early bud-break under inter-annual fluctuation in spring temperatures (Basler and Körner, 2012; Laube *et al.*, 2014). When the daylight stimulus is coupled with a time window of exposure to a certain amount of warmth (forcing temperature) ecodormancy is released and leaves start to unfold. For some species (*e.g. F. sylvatica*, *Betula pubescens* Ehrh.) it has been observed that a consistent lengthening of the photoperiod also compensate for an insufficient chilling during endodormancy, resulting in lower warming requirement and earlier bud-break (Caffarra and Donnelly, 2011; Vitasse and Basler, 2013; Basler, 2016).
Though TBB is an heritable trait (Frewen *et al.*, 2000; Hurme *et al.*, 2000; Jermstad *et al.*, 2001; Howe *et al.*, 2003), its inter-year variation within populations makes it difficult to tell apart the plastic from the genetic components. Since the 1950s, the warmer winters of the current climate scenario have pushed the time of bud-break earlier in the year for several boreal and temperate tree species (Körner and Basler, 2010; Vitasse *et al.*, 2011), and predicting the time of bud-break has become a priority to forecast the fate of tree populations. Sensitivity to photoperiod further muddles the modeling of the process, as photosensitive species are expected to respond differently to early warmths. Process-based models integrating empirical knowledge of the environmental drivers of bud-break have thus started to account both for the interplay between chilling and photoperiod as well as for the improvements in the understanding of endodormancy. Current models differ from each other in how they integrate chilling and photoperiod effects on the ability of buds to accumulate warming through (Gaüzere *et al.*, 2017).
Studies on fruit trees have identified several genetic networks that regulate the vegetative phenology by integrating internal signals (phytohormones such as auxin, gibberellin, and abscisic acid) and environmental cues: phytochromes and phytochromes-related transcription factors mediate and process Short-Day signal perception to *FLOWERING LOCUS T 2* to initiate bud set and regulate the entrance into dormancy at the end of the growing season (Howe *et al.*, 1995; Welling *et al.*, 2002); the auxin-receptor TIR1 mediates the response of the transcriptome to changes in auxin concentration (Baba *et al.*, 2011); different classes of *Dormancy-associated*

MADS-box (*DAM*) transcription factors are expressed throughout the year and regulate bud set and endodormancy (Ruttink *et al.*, 2007; Falavigna *et al.*, 2019; Yang *et al.*, 2021); as well as histone epigenetic modifications (Shim *et al.*, 2014). In particular, the product of *FLOWERING LOCUS T 1* and its paralogs play a crucial role in the release of winter dormancy and, indirectly, in bud-break (Horvath, 2009; Satake *et al.*, 2019; André *et al.*, 2022).

A significant part of the loci that control phenology is still unknown: only 5-10 QTLs of modest effect have been identified in species as *Pseudotsuga menziesii* (Mirb.) Franco (Douglas fir; Jermstad *et al.*, 2001); in *Quercus rubor* L (pedunculate oak) 32 QTLs were found to account for 3 - 11% of the phenotypic variance in the date of bud-break (Scotti-Saintagne *et al.*, 2004); in cases where bud-break is controlled by few QTLs of large effect, like in poplar and Scots pine (Frewen *et al.*, 2000; Hurme *et al.*, 2000), the heritability of the trait has been estimated to be even higher (Brachi, 2011). This hovering missing heritability in natural populations suggests that the control of spring phenology involves way more genes scattered across the genome. Genome-wide association studies in *Populus trichocarpa* (black cottonwood) Torr. & A.Gray ex. Hook. have identified QTLs in genes related to response to heat and daily thermocycles (several cytochromes P450, transporter interacting with histones), chilling (heat shock proteins), meristem regulation, cell growth and divisions, cell-to-cell signalling (target of *DAMs*) (Mckown *et al.*, 2014; McKown *et al.*, 2018); the same methodology identified found that in *Juglans regia* L (Persian walnut) the genetic control of bud-break regulatory genes for the phytochromes and genes for the cytoskeleton (Bükücü *et al.*, 2020).

A highly polygenic control of bud-break is to be expected in Fagaceae: the analysis of a full-sibling family of pedunculate oak detected at least one QTL for the date of bud-break in each of the 12 linkage groups, a clue that their number may be larger (Scotti-Saintagne *et al.*, 2004); a similar pattern has been found in *Castanea sativa* L (chestnut) (Casasoli *et al.*, 2004). The *CONSTANS-like* gene has been found associated with bud-break across multiple provenances of European beech, in *Quercus petrea* (Matt.) Liebl., and in two red oaks (Alberto, Derory, *et al.*, 2013; Lind-Riehl *et al.*, 2014; Müller *et al.*, 2017). An analysis of the bud and leaf transcriptome in *Q. rubor* and *Quercus pubescens* Willd (Lesur *et al. 2015*) that ecodormant buds over-express several genes related to protein synthesis and degradation (ribosomal proteins and proteins involved in the ubiquitin-dependent degradation of proteins, respectively), likely to maintain the protein turnover machinery under adverse climatic conditions. Other genes were involved in cold acclimation, responses to water stress, and light sensitivity. Dormancy control was found to be regulated by genes for gibberellin response and activation of the expression of *CONSTANS* (BAG proteins). Swelling buds were instead characterised by expression of genes for DNA replication, control of cell cycle, responses to phytohormones (auxin, gibberellin, brassinosteroids), and to sucrose as a re-initiator of mitotic activity.

A common trend across species is the finding of different QTLs driving phenology when analysing it in different years but in the same individuals (Jermstad *et al.*, 2001; Lazic *et al.*, 2024), suggesting that distinct sets of “general” QTLs driving TBB are expressed under different environmental conditions (Scotti-Saintagne *et al.*, 2004). Even though the short series of surveyed years may hamper detecting weak across-year signals, similar results have been reported also from *Arabidopsis thaliana* (L.) Heynh., where climatic adaptation drives genomic variation at fine scales involving distinct genetic bases across microhabitats (Frachon *et al.*, 2017, 2018).

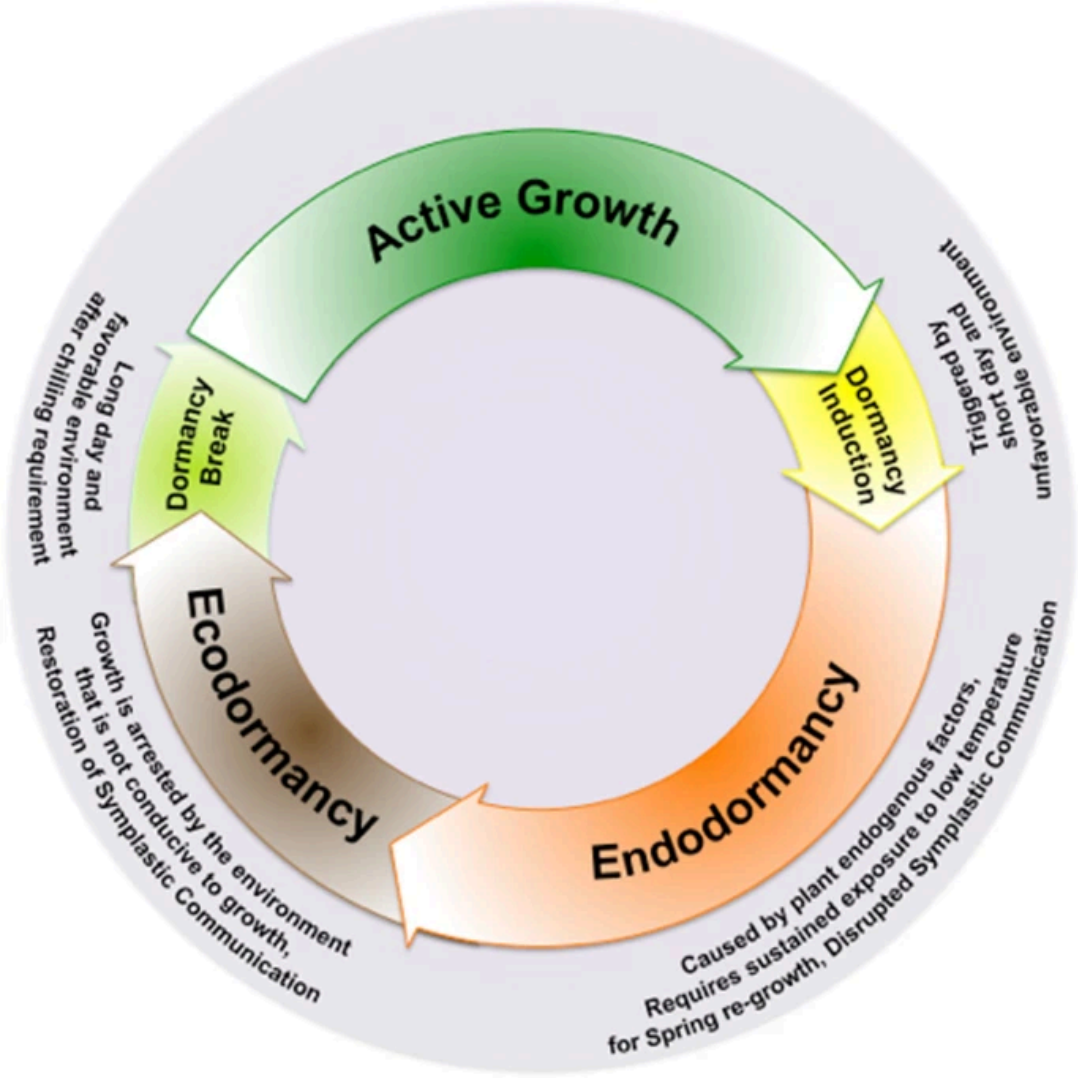


Figure 12: Growth-dormancy cycle of buds (Shim *et al.*, 2014). **Active growth** takes place from spring to end of the summer, when the signal of reduced photoperiod **induces dormancy**. **Endodormancy** is regulated by endogenous signals such as phytohormones; during this timeframe the plant accumulates cold. After a set time, environmental cues replace the endogenous control of dormancy (**endodormancy**) and the plant starts accumulating warming. The combination of signals of photoperiod elongation and successful accumulation of warming **breaks dormancy**.

1.5 The Mark of Selection

Detecting Adaptive Divergence in Natural Populations

While selection can be inferred at the phenotypic level by assessing the performances of individuals carrying different phenotypes for a trait, at the molecular level one has to evaluate the patterns of allele frequencies. Until the 1960s, selection was thought to be the major driver of differences between species. Based on comparison on the newly available information on protein sequence, Mooto Kimura (1968, 1969) proposed that the major part of the genetic variation in a single or a set of species does not affect fitness: most alleles at a locus have the same relative fitness and are selectively neutral (*ie*. evolve by drift); an allele may thus increase in frequency simply due to the random sampling of gametes between generations, reaching fixation in $4N_e$ generations (Nielsen, 2005). Tomoko Ohta (1973) further expanded the neutral theory of molecular evolution by introducing nearly neutral mutations, which have an effect on fitness between strictly neutral mutations and strongly deleterious mutations. Nearly neutral mutations are purged out by selection in large populations, while in small populations they behave neutrally and can reach fixation (**Figure 13**).

Nowaday, when looking for selection signatures across the genome, researchers use the hypothesis that the observed patterns of allele frequency across loci have evolved neutrally as the null model against which to test for selection: to deem a sequence to be subject of selection, one must first reject that the sequence is evolving by drift. The tenet is that without the action of selective forces the number of variant sites in the genome and the divergence between populations both depend solely on the (neutral) mutation rate and are thus proportional to each other. Since regions evolving deterministically tend to diverge more rapidly than expected under neutrality (Maynard Smith and Haigh, 1974), a signature of selection can be identified by comparing different loci to the genetic background of the population (Siol *et al.*, 2010). The direction of selection changes the distribution of genetic diversity at the genomic level (namely, the aforementioned excess and deficit of diversity). Tests for selection take advantage of theoretical predictions to evaluate selective signatures on molecular data.

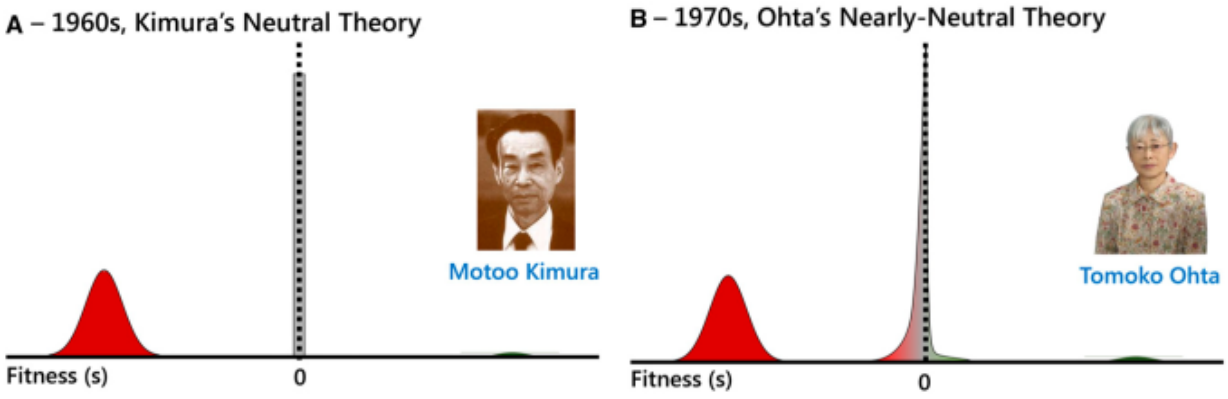


Figure 13: Distribution of the effect of mutations under (**a**) Kimura's neutral theory and (**b**) Ohta's nearly-neutral theory (Casillas and Barbadilla, 2017). In Ohta's theory, there is a gradient from **neutral** to **slightly neutral** mutation that evolve by drift; in both cases **strongly deleterious mutations** are culled and **beneficial mutations** fixed.

A first approach relies on the site frequency spectrum, the histogram of allele counts in a population (**Figure 14**). Under neutrality, allele frequencies change by chance alone and there is a progressive enrichment in the lower classes. When diversifying selection drives the allele dynamic, the genomic regions that control the selected phenotype diverge more than the genomic background: the allele with the stronger effect increases in frequency together with the neutral alleles at physically linked sites (hitchhiking) and while the frequency of this haplotype increases in the population, the genetic diversity in that genomic neighbourhood decreases. The number of alleles in the classes of higher counts increases slightly, mirrored by a slight decrease of alleles with low frequency (either selected against or generated by mutation). Instead, under a regime of balancing selection the site frequency spectrum is enriched at the intermediate classes (Fay and Wu, 2000; Nielsen, 2005; Siol *et al.*, 2010).

The regions linked to a locus under selection rises in frequency since recombinations happen too slowly to break up the association with the selected allele. The result is a local decrease of nucleotide diversity and high linkage disequilibrium (Maynard Smith and Haigh, 1974; Andolfatto, 2001). This high linkage disequilibrium can be used to detect new mutations that reach fixation very fast and change the local landscape of genetic diversity or follow the spread of an allele in a population (Sabeti *et al.*, 2002). However, the short fixation time of the beneficial allele is also the pitfall of these methods: once the variant has reached fixation, the haplotype is rapidly eroded (McVean, 2007).

Lewontin and Krakauer (1973) developed a test to detect recent signatures of adaptive divergence and speciation occurring among populations. The underlying idea is that there is some baseline metapopulation allele frequency. They reasoned that, in a deme, all loci should experience the same level of departure from that baseline because of their shared demographic history experienced at those loci: “While natural selection will operate differently for each locus and each allele at a locus, the effect of breeding structure is uniform over all loci and all alleles”. Instead, directional selection at individual loci dives their alleles to higher frequencies with respect to neutral loci. Signatures of selection can thus be detected by assessing the single-locus level of diversification between demes using Wright’s F_{ST} (1954) with respect to the expectation under neutrality. F_{ST} is the probability that two genes within a deme share a common ancestor within that deme and is commonly estimated with the Weir and Cockerham (1984)‘s method; it measures the degree of departure of the frequencies within a deme with respect to the metapopulation average due to drift within the deme. Since it relates to the breeding structure of a population, it’s often referred to as “inbreeding coefficient” (Beaumont, 2005).

In the Lewontin-Krakauer test, loci evolving under directional selection, either directly or indirectly by linkage, should exhibit larger differences between populations than neutrally evolving loci—and vice versa for loci under balancing selection (Charlesworth *et al.*, 2003; Siol *et al.*, 2010). The neutral (null) distribution of F_{ST} at neutral sites required for the comparisons can be derived from coalescent simulations (Beaumont and Nichols, 1996; Vitalis *et al.*, 2001) or the genomic background (Gompert and Buerkle, 2011). Applying the test on population pairs rather than on the metapopulation should better highlight weak but genuine signals of divergence that a more noisy background would hush (Vitalis *et al.*, 2001).

Several F_{ST}-based tests have been developed to evaluate patterns of differentiation between populations and account for deme-effect, demographic history, or to evaluate other statistical properties of the allele frequencies. The Bayesian F_{ST} (Foll and Gaggiotti, 2008) is decomposed in a locus-component and a population-component to assess if the the level of differentiation at the locus can be assessed only by the latter (neutrality) or requires also the former (selection). The X^tX statistics (Coop *et al.*, 2010; Günther and Coop, 2013) is a derivative of F_{ST} standardised by the estimated frequencies in the ancestral gene pool of the investigated populations to control the correlation of allele frequencies between demes. The Φ-statistics evaluates the *variance* at a locus between populations in a hierarchical structure (Excoffier *et al.*, 1992; Gompert and Buerkle, 2011). Along environmental gradients, F_{ST}-based tests will report the highest divergence between the habitats with the strongest environmental difference.

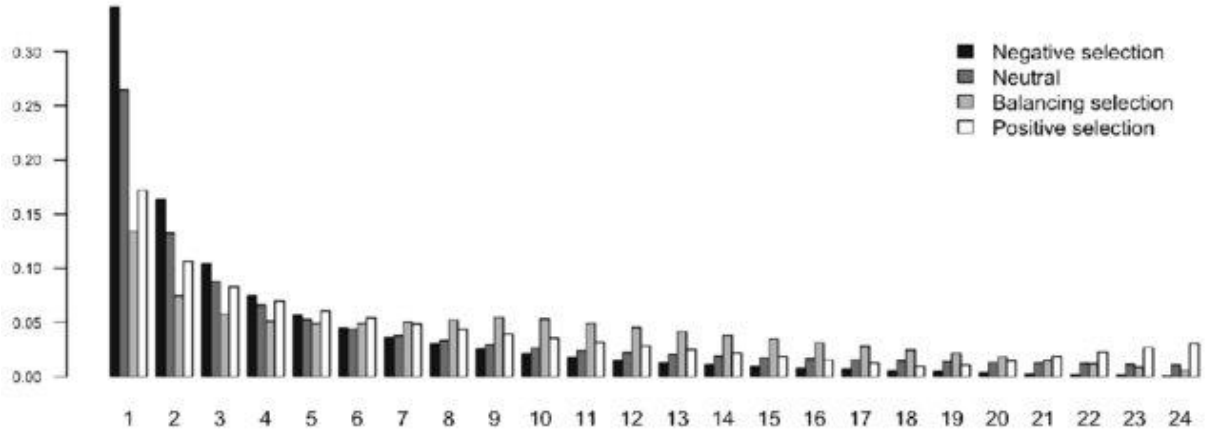


Figure 14: Unfolded site frequency spectra under neutrality (**dark gray**), negative (**black**), balancing (**light gray**), and positive selection (**white**) (Bernhardsson, 2012). The *x*-axis are the count classes of allele with the same frequence, each x_i class being the number of sites with a frequency *i* for the derived allele; the 0-class represents sites where individuals are homozygous for the ancestral allele, the (2*k* - 1)-class (in diploid organisms, *k* being the number of individuals in the sample) harbours sites where individuals are homozygous for the derived allele. A folded spectrum has half the classes of an unfolded one because the information on the ancestral state is not available (*ie*. lack of an outgroup) and it is not possible to polarise the alleles between ancestral and derived; hence the 0-class harbours the monomorphic sites.

1.5.1 Genome Wide Association Study (GWAS)

Also known as Genotype-Phenotype Association (GPA), GWAS methods investigate the association between the phenotype of an heritable trait and the genotypes at multiple loci. the studied loci are single nucleotide polymorphisms (SNP), individual sites in the DNA sequence where the population has been observed to have (most commonly) either one of two nucleotides (The 1000 Genomes Project Consortium, 2010). The relationship *Phenotype ~ Genotype* assesses the degree of change in the trait by changing an allele in the genotype; the proportion of phenotypic change is the allelic effect (**Figure 15**). GWAS can analyse both discrete (categorical) and continuous (quantitative) characters; in the former case, they implement a logistical model, in the latter a linear model. GWAS exploits the linkage disequilibrium between the locus responsible for the expression of the phenotype and the genetic marker the association of which is tested with the linear or logistic model. Markers that are tightly linked to the QTL will have higher frequencies among the individuals that express a given phenotype. Even if a QTL has not been sequenced, its genomic position will be revealed by the number of loci in linkage with it that will turn up to be associated with the phenotype.

A prerequisite of GWAS is to analyse unrelated individuals. The genomic distance at which linkage disequilibrium falls to zero depends on the level of consanguinity in the population. Two siblings have the same alleles at the same loci (identity by inheritance) for large stretches of a chromosome. Being separated by a single recombination event, in these individuals the distance at which a random locus and the QTL are associated by chance is way larger than for two unrelated individuals. If the individuals have the same trait value, many loci will turn up significant because they are in *cis* with the QTL; but the reason is the shared pedigree rather than a causal relationship.

GWA can involve from hundreds to (*mirabile dictu*) billion markers (Carré *et al.*, 2024). To avoid False Positives due to multiple hypothesis testing, markers are deemed significant either by strongly negative p-value (10^{-8}), after Bonferroni correction, or using a False Discovery Rate (Benjamini and Hochberg, 1995; François *et al.*, 2016).

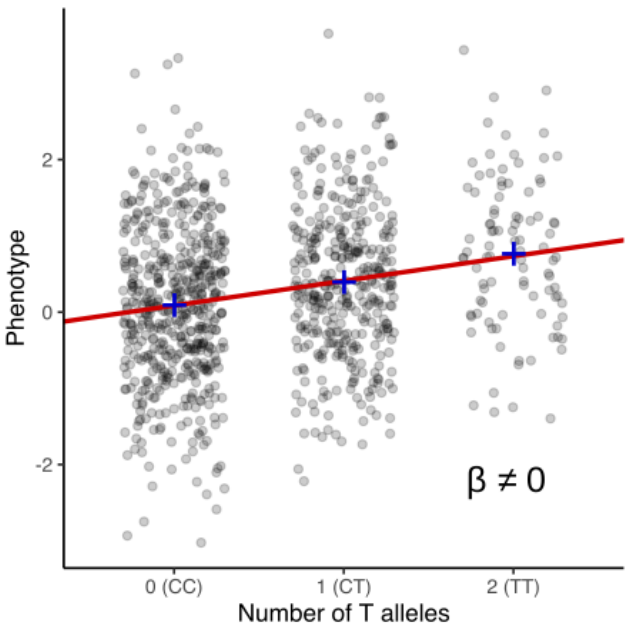


Figure 15: Genotype-Phenotype regression for a single SNP in a diploid population. Dots represents individual combinations of genotype x phenotype The association of the SNP with the trait is significant when the slope β of the regression is different from, meaning that changing an allele at the genotype also changes the phenotype value The regression is repeated for all SNPs in the panel.

More complex genotype ~ phenotype relationships can be used, depending on the aim of the researcher: Mixed Linear Models better control for relatedness between individuals and population structure by incorporating a matrix of coancestry (Zhang *et al.*, 2005) or kinship (Yu *et al.*, 2006) and principal components (Price *et al.*, 2006) respectively. Some GWAS models extend the test to multilocus phenotypes (Segura *et al.*, 2012; Wang *et al.*, 2016; Tamba and Zhang, 2018), decompose the SNP effect size in poly and oligogenic (Zhou *et al.*, 2013) and tackle linkage by testing each SNP with a model that incorporates the effect of its neighbouring markers (Liu *et al.*, 2016).

In the study of adaptation, GWAS aims at disentangling the architecture of the trait in the studied population. The precursors of GWAS were the linkage mapping studies, which were performed mostly on landraces and animal breeds with a known pedigree. Controlled crosses informed on which individual to include in a sample to study a given trait - and which breed to observe a certain phenotype. When performed in natural populations, GWAS lacks the advantage of knowing the pedigree of the individuals to analyse and, in many practical applications, the sample sizes are relatively small. A Kinship matrix is used to control for the probability that two individuals with the same phenotype have the same genotype at a locus because of shared inheritance (Goudet *et al.*, 2018). Environment also affects trait values through plasticity; while environmental factors may be included in the GWAS model, a panoply of them will escape the control of the researcher either because they are too expensive to measure at the individual level or simply because the phenotype x environment interaction is unknown. Furthermore, even in more controlled contexts the identified QTLs explain only a fraction of the phenotypic variance; this missing heritability is expected to be due the interaction among the SNPs in the panel (*ie*. epistasis), which is not easy to model (Manolio *et al.*, 2009; Chatelain *et al.*, 2018). Nonetheless, GWAS *in silva* allows for the broader expression of the studied trait and the detection of adaptive architectures (Santure and Garant, 2018).

1.5.2 “There Is Demography in My Sample!”

Confounding Factors

Methods based on F_{ST} (or equivalent) assume stable population sizes, a condition unlikely to be met *in natura*. Population expansions have several evolutionary consequences on the genetic diversity of the newly colonised areas, and non-adaptive processes can leave the same genomic signature expected by sweeps or directional selection. Demographic changes affect the patterns of linkage disequilibrium across the genome, as well as the shape of the site frequency spectrum (Wall *et al.*, 2002; Siol *et al.*, 2010). Tests that rely on hitchhiking will report higher rates of False Positives because of the random increases in the variance in levels of diversity across the genome (Wright and Gaut, 2005; Andolfatto, 2008) (**Figure 16**).

Plants can propagate vegetatively; populations are incepted by very small groups of colonists - down to a single individual - are thus particularly sensitive to bottlenecks (Foxe *et al.*, 2008). Due to their immobile growth habitus, plants have a tendency to mate with their neighbours and disperse pollen and seeds locally, and the resulting population structure. In a demographic wave, the random sampling of gametes causes some alleles to surf to high frequencies (and some others to be lost): the process is stochastic but can result in strong variation of allele frequencies in relatively short times. When a tree population expands, individuals on the edge of the home habitat contribute more than those in the core. The unequal contribution to the demographic wave causes a larger presence among the immigrants of alleles from the individuals on the edge of the home habitat. These alleles will appear at higher frequencies in the expanded area uniquely because of their parent’s position in the home habitat, but at the genomic level there will appear a low level of diversity around these loci, akin to the effect of a sweep (Klopfstein *et al.*, 2006; Excoffier and Ray, 2008) (**Figure 17**).

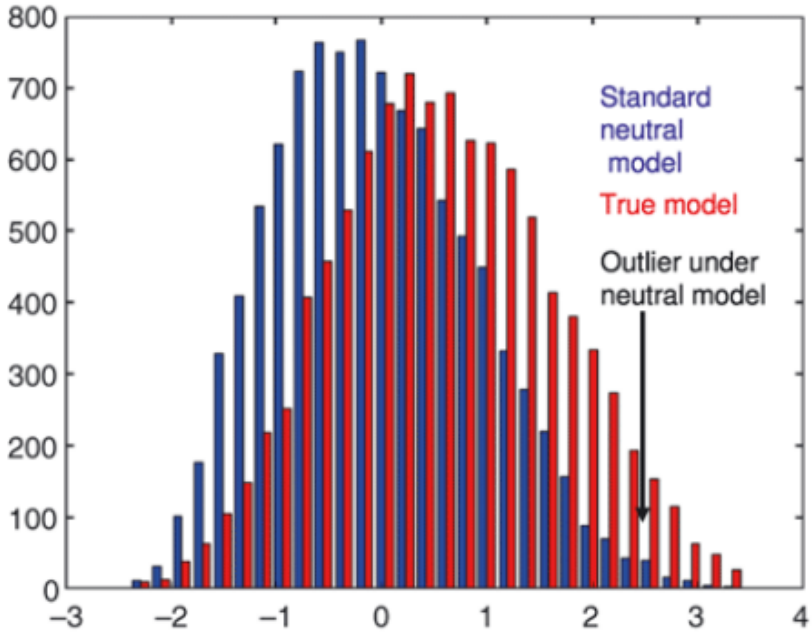


Figure 16: Distribution of the observed statistic in a population that experienced a size change under the **neutral** and the **true** model. The right-shift of the **true** distribution due to demography can lead to an higher proportion of False Positives.

When the surfing alleles have intermediate frequencies in the home population, multiple bouts of immigration - that is, serial founder events - can generate a cline of genetic differentiation (Edmonds *et al.*, 2004), even though there is no positive selection acting on these loci. This complicates the interpretation of clines of allele frequencies collinear with an environmental gradient, as they may be the signature of a demographic expansion, instead of selection (Slatkin and Excoffier, 2012). Another source of noise is the heterogeneous dispersal of propagules across the landscape due to geographical constraints, generating patches of different diversity (Barbujani *et al.*, 1995; Excoffier *et al.*, 2009). However, with a musical analogy, some noises can still sound nice: neutral alleles from the source population that have surfed in the expansion area may turn out to be advantageous in the new context, and, even if the trigger event is random chance, the following process is genuinely deterministic (Colautti and Barrett, 2013).

Controlling these factors is of capital importance to detect true positive signals of adaptive divergence. The problem when dealing with demography is thus generating a realistic null distribution of genetic diversity under the known demographic model and to use it to compute the chosen statistic of differentiation. Demography affects all loci in the genome in the same way while selection acts locally; hence the demography-corrected distribution will be shifted with respect to the one expected under constant population size but it should preserve the signals of selection while tuning down those that are due to the size change. One can incorporate pieces of information about the most plausible demographic scenario to generate such demography-corrected distribution by the use of (coalescent) simulations. The simulations estimate the parameters of the plausible demographic scenario that generated the observed genetic data; afterward, these parameters can be used to generate the expected distribution of genetic diversity under the aforementioned demographic history (Siol *et al.*, 2010).

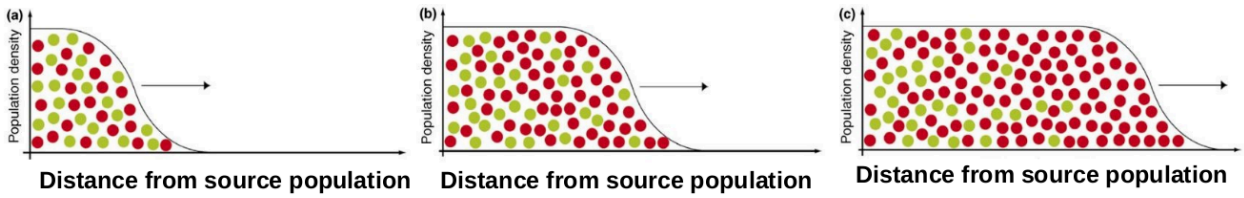


Figure 17: Allele surfing in an expanding population. In the home habitat (**a**), there is an equal proportion of individuals with **maroon** and **chartreuse** alleles. Within the demographic wave (**b**) more **maroon** alleles than **chartreuse** are random sampled for the next generation, resulting in (**c**) a cline of the **maroon** allele along the spatial gradient. If an environmental factor changes across the same spatial scale, it can be wrongfully interpreted as the selective driver for the observed pattern (Excoffier and Ray, 2008).

1.6 Actor, Setting, Time, Events, Plot

1.6.1 Who: *Fagus sylvatica* L.

Fagus sylvatica L. (European beech) is a late-successional forest tree that establishes in the later phases of forest succession (**Figure 18a**). European beech trees reach 30-35 m of height, with rare individuals that grow as high as 40-45 m (**Figure 18b**). Bud-break starts in March; the requirement of a vegetative season of at least ~14 days makes European beech sensitive to frost damage (Lenz *et al.*, 2013). Individuals reach sexual maturity at age 40-50; flowers develop between April and May (Oddou-Muratorio *et al.*, 2018) showing within-population synchronicity (Bogdziewicz *et al.*, 2023). The species is monoecious (individuals bear both male and female flowers) and predominantly outcrossed, with observed cases of autogamy (Piotti *et al.*, 2012). Pollen is dispersed by wind up to 46 m away, with rare long-distance events; seeds are produced heterogeneously between years (biannual masting cycles) and dispersed over 21 m away by gravity, with non-negligible zoochory.

The bark is smooth and dark gray, with small grey lenticels. Buds are 2.5 x 10 mm long and light brown. Leaf have a 10-15 mm long petiole, slightly asymmetric elliptic lamina (40-47 x 60 - 70 mm) with rounded base, obtuse apex and smooth or slightly crenate margin; the adaxial face is glabrous and glossy, the abaxial face presents tufts at the branching of the veins (**Figure 18c**). Catking are 1.5-2 cm long. Beech nuts (nuculae) are 8 x 20 mm large and completely enclosed in a 4-5-valved woody cupule bearing curve or patent (45°-90°-bent) needles (**Figure 18c**) (Pignatti and Guarino, 1982).

European beech is shade-tolerant growing on moderately deep soils and requires relatively high humidity rates in summers (Jump, Hunt, Martínez-Izquierdo, *et al.*, 2006; Bréda *et al.*, 2006; Gaüzere *et al.*, 2020). It grows on well drained, moderately deep and deep soils with different pH. In favorable conditions, European beech displays an efficient use of light that allows it to outcompete other tree species; in beech-dominated forest, the weak light of the undergrowth favours the growth of beech seedling with respect to other species. enough light once it. Humidity is a stronger environmental constraint than precipitation for European beech distribution range, and at high latitudes and on mountain slopes fog compensates for the lack of rainfall (Teissier Du Cros *et al.*, 1988).

European beech is classified within the family Fagaceae (subfamily Fagoideae; order Fagales) which includes the genera *Fagus* and *Engleriana*. Beeches are a dominant forest element at mid-latitudes at mid-elevations and have a disjunct distribution across of western Eurasia, East Asia and eastern North America. and found in habitats ranging from moist hot subtropical (*Fagus grandifolia* Ehrh. in Florida) to snow-rich cold temperate montane or oceanic settings Eurasian species) (Denk *et al.*, 2024). *Fagus* is monophyletic (Denk *et al.*, 2005) and likely diversified in the North Pacific region in the early Eocene. In an expansion wave it crossed the Bering channel and colonised Eurasia. The 12-15 species classified within the genus have emerged at different times through divergent and reticulate speciation processes; most Eurasian beeches diversified in the Middle and Upper Miocene (Denk and Grimm, 2009).

The few macrofossils of European beech (from now on, simply “beech”) allow to date its presence on the continent before the Quaternary (Denk et al. 2002). The species’ distribution range spans from the Cantabrian Mountains to the Carpathian, from Sicily to Sweden (**Figure**

18f), from 500 to 200 m of altitude. Beech is the main species of the undisturbed transnational UNESCO World Heritage SIte of the Ancient and Primeval Beech Forests of the Carpathians and Other Regions of Europe. Central Europe is the core of its distribution range nearly everywhere in anthropogenically undisturbed environments; in the northern part of its range it occurs in low elevation down to sea level while in the southern part it occurs at elevations of up to 2000 m a.s.l. (Magri *et al.*, 2006; Magri, 2008). The eastern edge of the distribution range overlaps with *Fagus orientalis* Lipsky, with which it can hybridise; its most proximate congeneric species appears to be *Fagus crenata* Blume (Denk and Grimm, 2009), from which the clade *F. sylvatica-F. orientalis* diversified during Miocene due to habitat fragmentation. Water content and photoperiod length determine its southern and northern edges, respectively.

Beech is diploid, with n = 12 and varying chromosome length (range between 28 and 73 Mb). The genome is 541 Mb long with a GC-content of 37.34% and a 2C value of 1.191±0.003 pg (Mishra *et al.*, 2018, 2022). In the Bhaga genome assembly 62,012 protein-coding have been predicted, with a density from 0 to 49.7 % per 100 kb window and a length range from 180 to 54,183 bp; the average number of exons per gene was 4.9. Insertions of organelles in the nuclear genome have been reported.

1.6.2 Where: Mont Ventoux

Mont Ventoux (44°21′N, 5°27′E) sits in southern France and with 1,910 m it towers over the Provencal prefecture of Vaucluse (**Figure 18d**). The mountaintop is predominantly limestone and numerous scree slopes, devoid of vegetation (**Figure 18e**); there, snow lasts until May. The intense rainfall during autumn and spring causes erosion and extensive karstification. Across the mountain, the soil is homogeneously heterogeneous - its composition changes over very short spatial scales without clear trend. Local climatic conditions change at different elevations from Alpine to temperate down to Mediterranean, with scorching summer temperatures. The northerly Mistral wind blows hard throughout the year (Barbero *et al.*, 1978). A mixed forest starts growing a little below the barren mountaintop and descends valley-ward. The fFloristic composition of the forest changes as one climbs up. From the bottom to the top, the forest is composed of deciduous oaks (*Quercus ilex* L., *Quercus pubescens* Willd.) maples, rowan, lime, boxwood *Buxus sempervirens* L., yew *Taxus baccata* L., European beech *Fagus sylvatica* L., Atlantic cedar *Cedrus atlantica* Manetti ex Endl., service tree *Sorbus domestica* Spach, and pines (Thinon, 2007). Above 1,500 metres dominates the cenosis of silver fir (*Abies alba* Mill.), Obier maple (*Acer opalus* Mill.), mountain pine (*Pinus uncinata* Turra), and Scots pine (*Pinus sylvestris* L.). Beech is distributed between 750 m to 1695 m. Human activities have been reported since the Neolithic, consistently with a progressive deforestation that peaked during the Middle Age (Ningre, 2007; Martin-Gousset *et al.*, 2019; Gamba *et al.*, 2023).

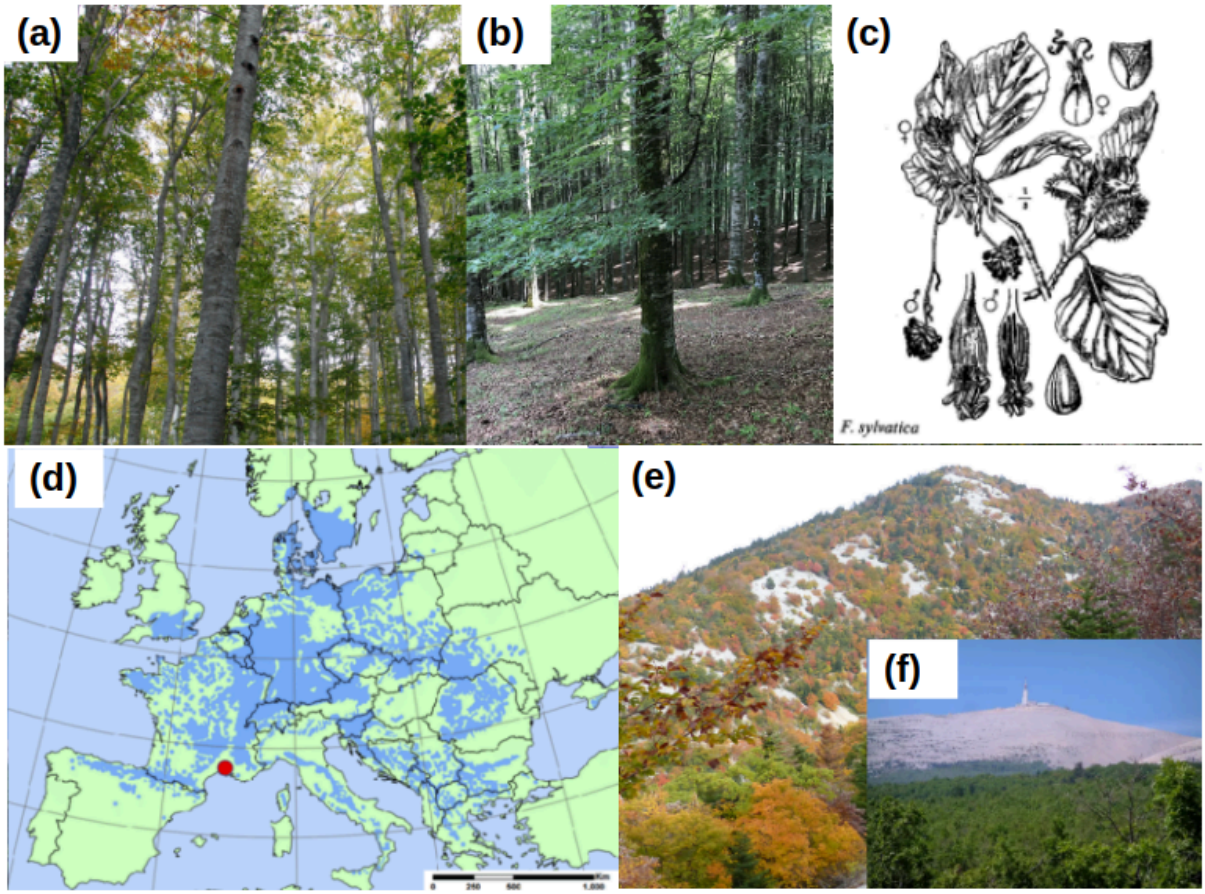


Figure 18: (**a**) Forest of European beech on Mont Ventoux (© Fréderic Jean); (**b**) adult tree; (**c**); leaves, flowers and fruit (Pignatti and Guarino, 1982)**;** (**e**) location of **Mont Ventoux** within the distribution area of European beech (Von Wühlisch, 2008); (**e**) forested slope of Mont Ventoux (© Fréderic Jean); (**f**) the barren mountaintop of Mont Ventoux as seen from the lowlands.

1.6.3 When: Last Glacial Maximum - 1860 - Today

Chloroplast markers distinguish three main groups of European beech: Italian Peninsula, southern Balkan Peninsula and the rest of Europe (Postolache *et al.*, 2021). During the Last Glacial Maximum (110 ka BP to 10 ka BP) the 52 identified beech refugia were located along the Mediterranean Basin, in the Pyrenees, the Western Alps and Balkans for an estimated coverage of 15,000 km^2 (Barbero *et al.*, 1978; Akman *et al.*, 1979). Population density increased slowly from the late glacial period and beech glacial refugia (13-10 kya) have been tracked to a few sites in Cantabrian mountains, Pyrenees, Croatia, Slovenia, southern Italy, and southern France. The species started expanding about 10-9 kya; it colonised central Europe about 5-4 kya from refugia in Slovenia, eastern Alps, possibly Moravia and southern Bohemia

Post-glacial colonisation was favoured by cooler and wetter climatic conditions; it was not simultaneous nor occurred with a spatially close front even in sites very close to each other, indicating that local and regional factors controlled the process. Signatures of strong founder effects are weak. Beech modern distribution corresponds to its maximum post-glacial extension (**Figure 19**; Magri, 2006; Magri *et al*., 2008).

Mont Ventoux is one of the 52 glacial refugia European beech. Since the Neolithic, human activities have pushed upward the distribution of the species, and several stands have been used for coppicing. In the Middle Age the forest was reduced to a relict ring at high elevations; patterns of charcoal findings date to this area (Gamba *et al.*, 2023) and even literary records confirm the barren state of Mont Ventoux slopes. By 1876 the estimated coverage of beech forest was approximately 2276 ha of beech, causing severe floods in the valley villages. With the social changes at the beginning of the XIX century, a rural flight from the towns at the foot of the mountain lowered human pressures on the forest. This factor, together with a reforestation program of the Office National de Forêts, was instrumental for a new valleyward expansion of the mixed forest of Mont Ventoux.

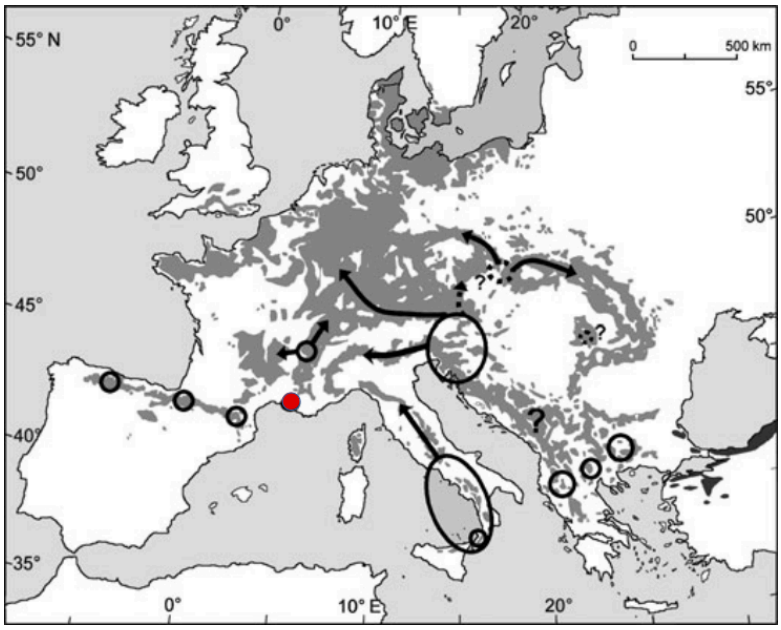


Figure 19: **Glacial refugia** (among which **Mont Ventoux**) and post-glacial colonisation patterns of beech overimposed to its **current distribution** (Magri *et al.* 2008)

1.6.4 What: Anticlinal Spring Phenology and Molecular Divergence

Local adaptation of beech stands on Mont Ventoux has been proven on a panel of candidate genes (Lalagüe *et al.*, 2013; Csilléry, Lalagüe *et al.*, 2014) between stands on the northerns and southern slope of the mountain (**Figure 20a-b**). Three provenances of different elevation were grown in a common garden experiment conducted by Gaüzere, Klein *et al.* (2020) **(Figure 20c-d)**. They observed that the provenances from higher elevation had an early spring, an anticline to what Davi *et al*. (2011) observed *in silva* (**Figure 20e**). The phenomenon has been observed since the 1980s (Von Wuehlisch, 1995) but its very local span had been assessed only in 2009 by Vitasse *et al.* Interestingly, Gaüzere *et al.* (2020) also estimated a higher contribution by immigrant local pollen in differentiation at quantitative traits that while plasticity has an adaptive role it drives a mismatch with the optimum for TBB at high elevations (Gaüzere, Teuf *et al.*, 2020). These results show that adaptation and environment push the timing of bud burst in opposite directions. However a finer understanding of the genomic control was required to successfully predict the anticline and understand its evolution.

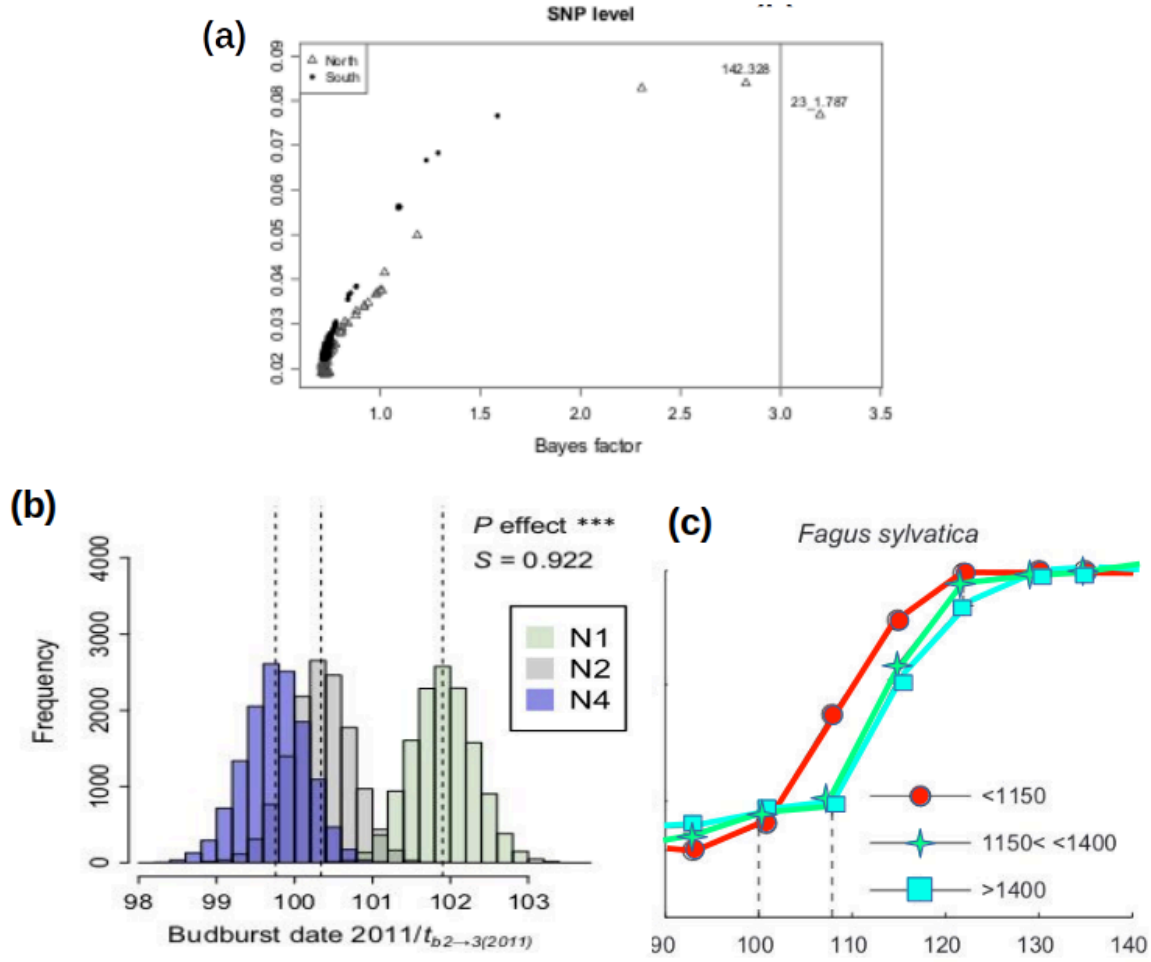


Figure 20: Signatures of selection have been identified on Mont Ventoux by (**a**) population genetics (Csilléry, Lalagüe *et al*., 2014) and (**b**) common garden experiments coupled with quantitative genetics (Gaüzere, Klein *et al*., 2020). The pattern for the time of bud-break observed in common garden is the opposite of the observed pattern *in silva* (**c**) (Daviiet al. 2011).

1.6.5 How: demographic expansion, selection

Mont Ventoux appears to be the stage where different, contrasting evolutionary forces unravel. At the shortest spatial scale, along the mountain slope, the changes in the environmental variables shape the genome of local beech trees in contrast to the extensive gene flow. At the same time, a downward demographic expansion facilitates non-adaptive evolution by bringing subsample of the gene pool of the refugial area in novel habitats. Nonetheless, in a common garden, seedlings from the stands encountered by climbing up the mountain will consistently show differences in budburst time—a clear phenotypic cline that looks like it is detached from the pattern of differentiation of its genetic bases.

Nowadays, the maps of the Office National de Fôrets make it possible to identify the old growth area and the area that has been colonised since 1860 (**Figure 21**; Lander *et al.*, 2011, 2020), explicitly addressing question about the differentiation of tree stands growing under sharply different conditions over few generations.

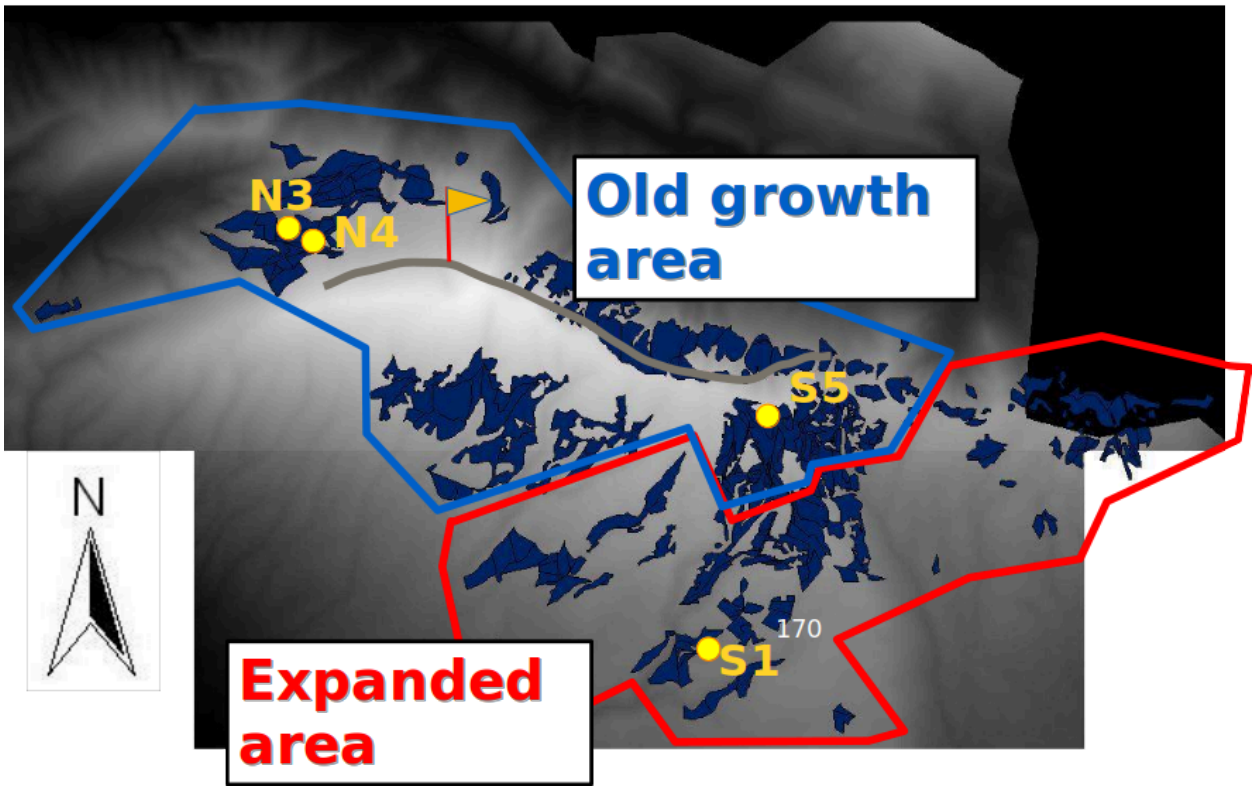


Figure 21: Map of modern day coverage of beech on Mont Ventoux, highlighting the **old growth area** and the **area colonised** since 1860 (Lander *et al*. 2011).

References

Abbott AG, Zhebentyayeva T, Barakat A, Liu Z (2015). The Genetic Control of Bud-Break in Trees. In: *Advances in Botanical Research*, Elsevier Vol 74, pp 201–228.

Aguirre WE (2009). Microgeographical diversification of threespine stickleback: Body shape-habitat correlations in a small, ecologically diverse Alaskan drainage. *Biol J Linn Soc* **98**: 139–151.

Aitken SN, Bemmels JB (2016). Time to get moving: Assisted gene flow of forest trees. *Evol Appl* **9**: 271–290.

Aitken SN, Whitlock MC (2013). Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. *Annu Rev Ecol Evol Syst* **44**: 367–388.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol Appl* **1**: 95–111.

Akman Y, Barbéro M, Quézel P (1979). Contribution à l'étude de la végétation forestière d'Anatolie méditerranéenne. *Phytocoenologia* **5**: 277–346.

Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, *et al.* (2013). Potential for evolutionary responses to climate change - evidence from tree populations. *Glob Change Biol* **19**: 1645–1661.

Alberto FJ, Derory J, Boury C, Frigerio J-M, Zimmermann NE, Kremer A (2013). Imprints of Natural Selection Along Environmental Gradients in Phenology-Related Genes of *Quercus petraea*. *Genetics* **195**: 495–512.

Anderson JV, Horvath DP, Chao WS, Foley ME (2010). Bud Dormancy in Perennial Plants: A Mechanism for Survival. In: Lubzens E, Cerda J, Clark M (eds) *Dormancy and Resistance in Harsh Environments*, Topics in Current Genetics. Springer Berlin Heidelberg: Berlin, Heidelberg Vol 21, pp 69–90.

Anderson JT, Panetta AM, Mitchell-Olds T (2012). Evolutionary and Ecological Responses to Anthropogenic Climate Change. *Plant Physiol* **160**: 1728–1740.

Andolfatto P (2001). Adaptive hitchhiking effects on genome variability. *Curr Opin Genet Dev* **11**: 635–641.

Andolfatto P (2008). Controlling Type-I Error of the McDonald–Kreitman Test in Genomewide Scans for Selection on Noncoding DNA. *Genetics* **180**: 1767–1771.

André D, Marcon A, Lee KC, Goretti D, Zhang B, Delhomme N, *et al.* (2022). FLOWERING LOCUS T paralogs control the annual growth cycle in Populus trees. *Curr Biol* **32**: 2988-2996.e4.

Antonovics J (1968). Evolution in closely adjacent plant populations VI. Manifold effects of gene flow. *Heredity* **23**: 508–524.

Arnold P, Kruuk, Loeske E. B., Nicotra, Adrienne B (2019). How to analyse plant phenotypic

plasticity in response to a changing climate. *New Phytol*: 1235–1241.

Audigeos D, Brousseau L, Traissac S, Scotti-Saintagne C, Scotti I (2013). Molecular divergence in tropical tree populations occupying environmental mosaics. *J Evol Biol* **26**: 529–544.

Baba K, Karlberg A, Schmidt J, Schrader J, Hvidsten TR, Bako L, *et al.* (2011). Activity-dormancy transition in the cambial meristem involves stage-specific modulation of auxin response in hybrid aspen. *Proc Natl Acad Sci U S A* **108**: 3418–3423.

Babik W, Butlin RK, Baker WJ, Papadopulos AST, Boulesteix M, Anstett M, *et al.* (2009). How sympatric is speciation in the *Howea* palms of Lord Howe Island? *Mol Ecol* **18**: 3629–3638.

Bacles CFE, Burczyk J, Lowe AJ, Ennos RA (2005). Historical and contemporary mating patterns in remnant populations of the forest tree *Fraxinus excelsior* L. *Evolution* **59**: 979.

Bacles CFE, Lowe AJ, Ennos RA (2006). Effective Seed Dispersal across a Fragmented Landscape. *Sci New Ser* **311**: 628.

Badeck F, Bondeau A, Böttcher K, Doktor D, Lucht W, Schaber J, *et al.* (2004). Responses of spring phenology to climate change. *New Phytol* **162**: 295–309.

Baiser B, Whitaker N, Ellison AM (2013). Modeling foundation species in food webs. *Ecosphere* **4**: 1–14.

Bal TMP, Llanos-garrido A, Chaturvedi A, Verdonck I, Hellemans B, Raeymaekers JAM (2021). Adaptive divergence under gene flow along an environmental gradient in two coexisting stickleback species. *Genes* **12**.

Barbero M, Du Merle P, Guende G, Quézel P (1978). La végétation du Mont Ventoux. *Terre Vie Supplément* **32**: 21–37.

Barbujani G, Sokal RR, Oden NL (1995). Indo-European origins: A computer-simulation test of five hypotheses. *Am J Phys Anthropol* **96**: 109–132.

Barson NJ, Aykanat T, Hindar K, Baranski M, Bolstad GH, Fiske P, *et al.* (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528**: 405–408.

Basler D (2016). Evaluating phenological models for the prediction of leaf-out dates in six temperate tree species across central Europe. *Agric For Meteorol* **217**: 10–21.

Basler D, Körner C (2012). Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agric For Meteorol* **165**: 73–81.

Battey NH (2000). Aspects of seasonality. *J Exp Bot* **51**: 1769–1780.

Beaumont MA (2005). Adaptation and speciation: What can Fst tell us? *Trends Ecol Evol* **20**: 435–440.

Beaumont MA, Nichols RA (1996). Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc B Biol Sci* **263**: 1619–1626.

Benjamini Y, Hochberg Y (1995). Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 (1995), Publi. *J R Stat Soc* **57**: 289–300.

Bernhardsson C (2012). *Molecular population genetics of inducible defense genes in* Populus tremula. Department of Ecology and Environmental Science, Umeå University: Umeå.

Birch LC (1957). The Role of Weather in Determining the Distribution and Abundance of Animals. *Cold Spring Harb Symp Quant Biol* **22**: 203–218.

Bogdziewicz M, Journé V, Hacket‐Pain A, Szymkowiak J (2023). Mechanisms driving interspecific variation in regional synchrony of trees reproduction. *Ecol Lett* **26**: 754–764.

Bontemps A, Davi H, Lefèvre F, Rozenberg P, Oddou‐Muratorio S (2017). How do functional traits syndromes covary with growth and reproductive performance in a water‐stressed population of *Fagus sylvatica* ? *Oikos* **126**: 1472–1483.

Bontemps A, Klein EK, Oddou-Muratorio S (2013). Shift of spatial patterns during early recruitment in Fagus sylvatica: Evidence from seed dispersal estimates based on genotypic data. *For Ecol Manag* **305**: 67–76.

Bouillé M, Bousquet J (2005). Trans‐species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long‐term maintenance of genetic diversity in trees. *Am J Bot* **92**: 63–73.

Boyle EA, Li YI, Pritchard JK (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**: 1177–1186.

Brachi B (2011). Genome-wide association studies in plants: the missing heritability is in the field.

Bradshaw AD (1952). Populations of Agrostis tenuis Resistant to Lead and Zinc Poisoning. *Nature* **169**: 1098–1098.

Bradshaw AD (1960). Population differentiation in *Agrostis tenuis* Sibth: III. Populations in varied environments. *New Phytol* **59**: 92–103.

Bradshaw HD, Otto KG, Frewen BE, McKay JK, Schemske DW (1998). Quantitative Trait Loci Affecting Differences in Floral Morphology Between Two Species of Monkeyflower (Mimulus). *Genetics* **149**: 367–382.

Bréda N, Huc R, Granier A, Dreyer E (2006). Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences. *Ann For Sci* **63**: 625–644.

Brousseau L, Fine PVA, Dreyer E, Vendramin GG, Scotti I (2021). Genomic and phenotypic divergence unveil microgeographic adaptation in the Amazonian hyperdominant tree Eperua falcata Aubl. (Fabaceae). *Mol Ecol* **30**: 1136–1154.

Brousseau L, Foll M, Scotti-Saintagne C, Scotti I (2015). Neutral and adaptive drivers of

microgeographic genetic divergence within continuous populations: The case of the neotropical tree Eperua falcata (aubl.). *PLoS ONE* **10**: 1–23.

Bükücü ŞB, Sütyemez M, Kefayati S, Paizila A, Jighly A, Kafkas S (2020). Major QTL with pleiotropic effects controlling time of leaf budburst and flowering-related traits in walnut (Juglans regia L.). *Sci Rep* **10**: 15207.

Bulmer MG (1972). Multiple Niche Polymorphism. *Am Nat* **106**: 254–257.

Caffarra A, Donnelly A (2011). The ecological significance of phenology in four different tree species: Effects of light and temperature on bud burst. *Int J Biometeorol* **55**: 711–721.

Cahill AE, Aiello-Lammens ME, Fisher-Reid MC, Hua X, Karanewsky CJ, Yeong Ryu H, *et al.* (2013). How does climate change cause extinction? *Proc R Soc B Biol Sci* **280**: 20121890.

Carré C, Carluer JB, Chaux C, Estoup-Streiff C, Roche N, Hosy E, *et al.* (2024). Next-Gen GWAS: full 2D epistatic interaction maps retrieve part of missing heritability and improve phenotypic prediction. *Genome Biol* **25**: 76.

Casasoli M, Pot D, Plomion C, Monteverdi MC, Barreneche T, Lauteri M, *et al.* (2004). Identification of QTLs affecting adaptive traits in *Castanea sativa* Mill. *Plant Cell Environ* **27**: 1088–1101.

Casillas S, Barbadilla A (2017). Molecular Population Genetics. *Genetics* **205**: 1003–1035.

Charlesworth B, Charlesworth D, Barton NH (2003). The Effects of Genetic and Geographic Structure on Neutral Variation. *Annu Rev Ecol Evol Syst* **34**: 99–125.

Chatelain C, Durand G, Thuillier V, Augé F (2018). Performance of epistasis detection methods in semi-simulated GWAS. *BMC Bioinformatics* **19**: 231.

Colautti RI, Barrett SCH (2013). Rapid Adaptation to Climate Facilitates Range Expansion of an Invasive Plant. *Science* **342**: 364–366.

Conery M, Grant SFA (2023). Human height: a model common complex trait. *Ann Hum Biol* **50**: 258–266.

Conover DO, Schultz ET (1995). Phenotypic similarity and the evolutionary significance of countergradient variation. *Trends Ecol Evol* **10**: 248–252.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**: 1411–1423.

Copernicus CCS (2022). Global climate highlights: 2022.

Copernicus CCS (2023). Global climate highlights: 2023.

Copernicus CCS (2024). Global climate highlights: 2024.

Crispo E (2008). Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. *J Evol Biol* **21**: 1460–1469.

Csilléry K, Lalagüe H, Vendramin GG, González-Martínez SC, Fady B, Oddou-Muratorio S (2014). Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (Fagus sylvatica) populations. *Mol Ecol* **23**: 4696–4708.

Danby RK, Hik DS (2007). Variability, contingency and rapid change in recent subarctic alpine tree line dynamics. *J Ecol* **95**: 352–363.

Davi H, Gillmann M, Ibanez T, Cailleret M, Bontemps A, Fady B, *et al.* (2011). Diversity of leaf unfolding dynamics among tree species: New insights from a study along an altitudinal gradient. *Agric For Meteorol* **151**: 1504–1513.

Davis MB, Shaw RG (2001). Range shifts and adaptive responses to quaternary climate change. *Science* **292**: 673–679.

Dayton PK (1972). Toward an understanding of community resilience and the potential effects of enrichments to the benthos at McMurdo Sound, Antarctica. In: Parker B (ed) *Proceedings of the Colloquium on Conservation Problems in Antarctica*, Allen Press, pp 81–96.

De Villemereuil P, Gaggiotti OE, Mouterde M, Till-Bottraud I (2016). Common garden experiments in the genomic era: new perspectives and opportunities. *Heredity* **116**: 249–254.

Denk T, Grimm GW (2009). The biogeographic history of beech trees. *Rev Palaeobot Palynol* **158**: 83–100.

Denk T, Grimm GW, Cardoni S, Csilléry K, Kurz M, Schulze E-D, *et al.* (2024). A subgeneric classification of Fagus (Fagaceae) and revised taxonomy of western Eurasian beeches. *Willdenowia* **54**.

Denk T, Grimm GW, Hemleben V (2005). Patterns of Molecular and Morphological Differentiation in Fagus (Fagaceae): Phylogenetic Implications Author (s): Thomas Denk , Guido W . Grimm and Vera Hemleben Published by : Wiley Stable URL : https://www.jstor.org/stable/4126078 REFERENCES Linked. **92**: 1006–1016.

Dittmar EL, Schemske DW (2023). Temporal Variation in Selection Influences Microgeographic Local Adaptation. *Am Nat* **202**: 471–485.

Dobzhansky T (1968). On Some Fundamental Concepts of Darwinian Biology. In: Dobzhansky T, Hecht MK, Steere WC (eds) *Evolutionary Biology*, Springer US: Boston, MA, pp 1–34.

Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004). Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci* **101**: 975–979.

Ellison AM (2019). Foundation Species, Non-trophic Interactions, and the Value of Being Common. *iScience* **13**: 254–268.

Ellison AM, Bank MS, Clinton BD, Colburn EA, Elliott K, Ford CR, *et al.* (2005). Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. *Front Ecol Environ* **3**: 479–486.

Endler JA (1977). *Geographic Variation, Speciation and Clines* (RM May, Ed.). Princenton University Press.

Excoffier L, Foll M, Petit RJ (2009). Genetic Consequences of Range Expansions. *Annu Rev Ecol Evol Syst* **40**: 481–501.

Excoffier L, Ray N (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol* **23**: 347–351.

Excoffier L, Smouse PE, Quattro JM (1992). Analysis of Molecular Variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.

Fady B, Bozzano M (2021). Effective population size does not make a practical indicator of genetic diversity in forest trees. *Biol Conserv* **253**: 108904.

Falavigna VDS, Guitton B, Costes E, Andrés F (2019). I Want to (Bud) Break Free: The Potential Role of DAM and SVP-Like Genes in Regulating Dormancy Cycle in Temperate Fruit Trees. *Front Plant Sci* **9**: 1990.

FAO (2005). Global Forest Resources Assessments 2005. *https://www.fao.org*.

FAO (2020). Global Forest Resources Assessment 2020. In: Rome, p .

Fay JC, Wu C-I (2000). Hitchhiking Under Positive Darwinian Selection. *Genetics* **155**: 1405–1413.

Felsenstein J (1976). Theoretical population genetics of variable selection and migration. *Annu Rev Genet* **10**: 253–280.

Feulner PGD, Chain FJJ, Panchal M, Huang Y, Eizaguirre C, Kalbe M, *et al.* (2015). Genomics of Divergence along a Continuum of Parapatric Population Differentiation (J Zhang, Ed.). *PLOS Genet* **11**: e1004966.

Finn GA, Straszewski AE, Peterson V (2007). A general growth stage key for describing trees and woody plants. *Ann Appl Biol* **151**: 127–131.

Fisher RA (1919). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans R Soc Edinb* **52**: 399–433.

Fitzpatrick SW, Gerberich JC, Kronenberger JA, Angeloni LM, Funk WC (2015). Locally adapted traits maintained in the face of high gene flow (G Grether, Ed.). *Ecol Lett* **18**: 37–47.

Foll M, Gaggiotti O (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics* **180**: 977–993.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Mol Ecol* **25**: 104–120.

Foxe JP, Dar V -u.-N, Zheng H, Nordborg M, Gaut BS, Wright SI (2008). Selection on Amino Acid Substitutions in Arabidopsis. *Mol Biol Evol* **25**: 1375–1383.

Frachon L, Bartoli C, Carrère S, Bouchez O, Chaubet A, Gautier M, *et al.* (2018). A genomic map of climate adaptation in arabidopsis thaliana at a micro-geographic scale. *Front Plant Sci* **9**: 1–15.

Frachon L, Libourel C, Villoutreix R, Carrère S, Glorieux C, Huard-Chauveau C, *et al.* (2017). Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. *Nat Ecol Evol* **1**: 1551–1561.

François O, Martins H, Caye K, Schoville SD (2016). Controlling false discoveries in genome scans for selection. *Mol Ecol* **25**: 454–469.

Frewen BE, Chen THH, Howe GT, Davis J, Rohde A, Boerjan W, *et al.* (2000). Quantitative Trait Loci and Candidate Gene Mapping of Bud Set and Bud Flush in Populus. *Genetics* **154**: 837–845.

Gamba E, Aurelle S, Vergnaud C, Talon B (2023). A Holocene history of forest vegetation on the northern slopes of a Mediterranean mountain : the Mont Ventoux, Southeast France. *Ecologia Mediterranea*, 48(2):53-68.

García‐Valdés R, Bugmann H, Morin X (2018). Climate change‐driven extinctions of tree species affect forest functioning more than random extinctions (D Zurell, Ed.). *Divers Distrib* **24**: 906–918.

Gaston KJ (2009). Geographic range limits: achieving synthesis. *Proc R Soc B Biol Sci* **276**: 1395–1406.

Gaüzere J, Delzon S, Davi H, Bonhomme M, Garcia de Cortazar-Atauri I, Chuine I (2017). Integrating interactive effects of chilling and photoperiod in phenological process-based models. A case study with two European tree species: Fagus sylvatica and Quercus petraea. *Agric For Meteorol* **244–245**: 9–20.

Gaüzere J, Klein EK, Brendel O, Davi H, Oddou-Muratorio S (2020). Microgeographic adaptation and the effect of pollen flow on the adaptive potential of a temperate tree species. *New Phytol* **227**: 641–653.

Gaüzere J, Klein EK, Oddou-Muratorio S (2013). Ecological determinants of mating system within and between three Fagus sylvatica populations along an elevational gradient. *Mol Ecol* **22**: 5001–15.

Gaüzere J, Teuf B, Davi H, Chevin L, Caignard T, Leys B, *et al.* (2020). Where is the optimum? Predicting the variation of selection along climatic gradients and the adaptive value of plasticity. A case study on tree phenology. *Evol Lett* **4**: 109–123.

Gavrilets S, Scheiner SM (1993). The genetics of phenotypic plasticity. V. Evolution of reaction norm shape. *J Evol Biol* **6**: 31–48.

Germino MJ, Moser AM, Sands AR (2019). Adaptive variation, including local adaptation, requires decades to become evident in common gardens. *Ecol Appl* **29**: e01842.

Gömöry D, Paule L (2011). Trade-off between height growth and spring flushing in common beech (Fagus sylvatica L.). *Ann For Sci* **68**: 975–984.

Gompert Z, Buerkle CA (2011). A hierarchical Bayesian model for next-generation population genomics. *Genetics* **187**: 903–17.

Goudet J, Kay T, Weir BS (2018). How to estimate kinship. *Mol Ecol* **27**: 4121–4135.

Gould SJ, Vrba ES (1982). Exaptation — a missing term in the science of form. *Paleobiology* **8**: 4–15.

Grether GF (2005). Environmental Change, Phenotypic Plasticity, and Genetic Compensation. *Am Nat* **166**: E115–E123.

Günther T, Coop G (2013). Robust identification of local adaptation from allele frequencies. *Genetics* **195**: 205–220.

Halbritter AH, Fior S, Keller I, Billeter R, Edwards PJ, Holderegger R, *et al.* (2018). Trait differentiation and adaptation of plants along elevation gradients. *J Evol Biol* **31**: 784–800.

Hamrick JL, Linhart YB, Mitton JB (1979). Relationships Between Life History Characteristics and Electrophoretically Detectable Genetic Variation in Plants. *Annu Rev Ecol Syst* **10**: 173–200.

Hedrick PW (1986). Genetic Polymorphism in Heterogeneous Environments: A Decade Later. *Ann Rev Ecol Syst* **17**: 535–456.

Hedrick PW, Ginevan ME, Ewing EP (1976). Genetic Polymorphism in Heterogeneous Environments. *Annu Rev Ecol Syst* **7**: 1–32.

Horvath D (2009). Common mechanisms regulate flowering and dormancy. *Plant Sci* **177**: 523–531.

Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen TH (2003). From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Can J Bot* **81**: 1247–1266.

Howe GT, Hackett WP, Furnier GR, Klevron RE (1995). Photoperiodic responses of a northern and southern eeotype of black cottonwood. *Physiol Plant* **93**: 695–708.

Hurme P, Sillanpää MJ, Arjas E, Repo T, Savolainen O (2000). Genetic Basis of Climatic Adaptation in Scots Pine by Bayesian Quantitative Trait Locus Analysis. *Genetics* **156**: 1309–1322.

Jean F, Davi H, Oddou-Muratorio S, Fady B, Scotti I, Scotti-Saintagne C, *et al.* (2023). A 14-year series of leaf phenological data collected for European beech (Fagus sylvatica L.) and silver fir (Abies alba Mill.) from their geographic range margins in south-eastern France. *Ann For Sci* **80**: 35.

Jermstad KD, Bassoni DL, Jech KS, Wheeler NC, Neale DB (2001). Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush: *Theor Appl Genet* **102**: 1142–1151.

Johnson LC, Galliart MB, Alsdurf JD, Maricle BR, Baer SG, Bello NM, *et al.* (2022). Reciprocal transplant gardens as gold standard to detect local adaptation in grassland species: New opportunities moving into the 21st century. *J Ecol* **110**: 1054–1071.

Journé V, Caignard T, Hacket-Pain A, Bogdziewicz M (2021). Leaf phenology correlates with fruit production in European beech (Fagus sylvatica) and in temperate oaks (Quercus robur and Quercus petraea). *Eur J For Res* **140**: 733–744.

Jump AS, Hunt JM, Martínez-Izquierdo JA, Peñuelas J (2006). Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Mol Ecol* **15**: 3469–3480.

Jump AS, Hunt JM, Peñuelas J (2006). Rapid climate change-related growth decline at the southern range edge of Fagus sylvatica. *Glob Change Biol* **12**: 2163–2174.

Jump AS, Peñuelas J (2005). Running to stand still: Adaptation and the response of plants to rapid climate change. *Ecol Lett* **8**: 1010–1020.

Kawecki TJ, Ebert D (2004). Conceptual issues in local adaptation. *Ecol Lett* **7**: 1225–1241.

Kimura M (1968). Evolutionary rate at the molecular level. *Nature* **217**: 624–626.

Kimura M (1969). The Rate of Molecular Evolution Considered from the Standpoint of Population Genetics. *Proc Natl Acad Sci* **63**: 1181–1188.

Kingsolver JG, Pfennig DW, Servedio MR (2002). Migration, local adaptation and the evolution of plasticity. *Trends Ecol Evol* **17**: 540–541.

Kisdi É (2002). Dispersal: Risk Spreading versus Local Adaptation. *Am Nat* **159**: 579–596.

Klopfstein S, Currat M, Excoffier L (2006). The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* **23**: 482–490.

Körner C (2007). The use of 'altitude' in ecological research. *Trends Ecol Evol* **22**: 569–574.

Körner C, Basler D (2010). Phenology Under Global Warming. *Science* **327**: 1461–1462.

Kremer A, Le Corre V (2012a). Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity* **108**: 375–385.

Kremer A, Le Corre V (2012b). Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity* **108**: 375–385.

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, *et al.* (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecol Lett* **15**: 378–392.

Lalagüe H, Csilléry K, Oddou-Muratorio S, Safrana J, de Quattro C, Fady B, *et al.* (2013). Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (Fagus sylvatica L.) population from southeastern France. *Tree Genet Genomes* **10**: 15–26.

Lande R (1980). The Genetic Covariance between Characters Maintained by Pleiotropic Mutations. *Genetics* **94**: 203–215.

Lander TA, Klein EK, Roig A, Oddou-Muratorio S (2020). Weak founder effects but significant spatial genetic imprint of recent contraction and expansion of European beech populations. *Heredity*.

Lander TA, Oddou-Muratorio S, Prouillet-Leplat H, Klein EK (2011). Reconstruction of a beech population bottleneck using archival demographic information and Bayesian analysis of genetic data. *Mol Ecol* **20**: 5182–5196.

Langlet O (1971). Two hundred Years genecology. *Taxon* **20**: 653–721.

Laube J, Sparks TH, Estrella N, Höfler J, Ankerst DP, Menzel A (2014). Chilling outweighs photoperiod in preventing precocious spring development. *Glob Change Biol* **20**: 170–182.

Lazic D, Geßner C, Liepe KJ, Lesur-Kupin I, Mader M, Blanc-Jolivet C, *et al.* (2024). Genomic variation of European beech reveals signals of local adaptation despite high levels of phenotypic plasticity. *Nat Commun* **15**: 8553.

Le Corre V, Kremer A (2003). Genetic Variability at Neutral Markers, Quantitative Trait Loci and Trait in a Subdivided Population Under Selection. *Genetics* **164**: 1205–1219.

Le Corre V, Kremer A (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Mol Ecol* **21**: The genetic differentiation at quantitative trait.

Lenormand T (2002). Gene flow and the limits to natural selection. *Trends Ecol Evol* **17**: 183–189.

Lenz A, Hoch G, Vitasse Y, Körner C (2013). European deciduous trees exhibit similar safety margins against damage by spring freeze events along elevational gradients. *New Phytol* **200**: 1166–1175.

Lepais O, Bacles CFE (2014). Two are better than one: combining landscape genomics and common gardens for detecting local adaptation in forest trees. *Mol Ecol* **23**: 4671–4673.

Lesur I, Bechade A, Lalanne C, Klopp C, Noirot C, Leplé J-C, *et al.* (2015). A unigene set for European beech (*Fagus sylvatica* L.) and its use to decipher the molecular mechanisms involved in dormancy regulation. *Mol Ecol Resour* **15**: 1192–1204.

Lewontin RC (1970). The Units of Selection. *Annu Rev Ecol Evol Syst* **1**: 1–18.

Lewontin RC, Krakauer J (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**: 175–195.

Lind-Riehl JF, Sullivan AR, Gailing O (2014). Evidence for selection on a CONSTANS-like gene between two red oak species. *Ann Bot* **113**: 967–975.

Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies (J Listgarten, Ed.). *PLOS Genet* **12**: e1005767.

Loarie SR, Duffy PB, Hamilton H, Asner GP, Field CB, Ackerly DD (2009). The velocity of climate change. *Nature* **462**: 1052–1055.

Lynch M, Lande R (1993). Evolution and extinction in response to environmental change. In: *Biotic interactions and global change.*, Sinauer Associates Inc.: Sunderland, p .

Lynch M, Walsh B (1997). *Genetics and analysis of quantitative traits*. Sinauer Associates Inc.: Sunderland.

Maciejewski MF, Jiang C, Stuart YE, Bolnick DI (2020). Microhabitat contributes to microgeographic divergence in threespine stickleback. *Evolution* **74**: 749–763.

Magri D (2008). Patterns of post-glacial spread and the extent of glacial refugia of European beech (Fagus sylvatica). *J Biogeogr* **35**: 450–463.

Magri D, Vendramin GG, Comps B, Dupanloup I, Geburek T, Gömöry D, *et al.* (2006). A new scenario for the Quaternary history of European beech populations: Palaeobotanical evidence and genetic consequences. *New Phytol* **171**: 199–221.

Mann ME, Bradley RS, Hughes MK (1999). Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophys Res Lett* **26**: 759–762.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, *et al.* (2009). Finding the missing heritability of complex diseases. *Nature* **461**: 747–753.

Martin-Gousset L, Terracol J, Roux J-P (2019). Habitats et dynamiques de végétation en ubac du Mont-Ventoux – Étude de la Réserve biologique intégrale. *Naturae* **2019**: 113–164.

Masek JG (2001). Stability of boreal forest stands during recent climate change: evidence from Landsat satellite imagery. *J Biogeogr* **28**: 967–976.

Maynard Smith JM, Haigh J (1974). The hitch-hiking effect of a favourable gene. *Genet Res* **23**: 23–35.

Mckay JK, Latta RG (2002). *Adaptive population divergence: markers, QTL and traits*.

McKown AD, Klápště J, Guy RD, El-Kassaby YA, Mansfield SD (2018). Ecological genomics of variation in bud-break phenology and mechanisms of response to climate warming in *Populus trichocarpa*. *New Phytol* **220**: 300–316.

Mckown AD, Klápště J, Guy RD, Geraldes A, Porth I, Hannemann J, *et al.* (2014). Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa. *New Phytol* **203**: 535–553.

McVean G (2007). The Structure of Linkage Disequilibrium Around a Selective Sweep. *Genetics* **175**: 1395–1406.

Meier U (2001). Growth stages of mono-and dicotyledonous plants: BBCH Monograph, 2nd ed. *Fed Biol Res Cent Agric For* **49(2)**: 66–70.

Merilä J, Hendry AP (2014). Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evol Appl* **7**: 1–14.

Mikkola J (1982). Bud-set phenology as an indicator of climatic adaptation of Scots pine in

Finland. *Silva Fenn* **16**: 178–184.

Milesi P, Kastally C, Dauphin B, Cervantes S, Bagnoli F, Budde KB, *et al.* (2024). Resilience of genetic diversity in forest trees over the Quaternary. *Nat Commun* **15**: 8538.

Mishra B, Gupta DK, Pfenninger M, Hickler T, Langer E, Nam B, *et al.* (2018). A reference genome of the European beech (Fagus sylvatica L.). *GigaScience* **7**.

Mishra B, Ulaszewski B, Meger J, Aury J-M, Bodénès C, Lesur-Kupin I, *et al.* (2022). A Chromosome-Level Genome Assembly of the European Beech (*Fagus sylvatica*) Reveals Anomalies for Organelle DNA Integration, Repeat Content and Distribution of SNPs. *Front Genet* **12**: 691058.

Monod J (1972). *Chance and Necessity: An Essay on the Natural Philosophy of Biology*. Random House Trade Paperbacks: New York City.

Mosca E, Gugerli F, Eckert AJ, Neale DB (2016). Signatures of natural selection on Pinus cembra and P. mugo along elevational gradients in the Alps. *Tree Genet Genomes* **12**: 9.

Müller M, Seifert S, Finkeldey R (2017). Comparison and confirmation of SNP-bud burst associations in European beech populations in Germany. *Tree Genet Genomes* **13**: 59.

Neary DG, Ice GG, Jackson CR (2009). Linkages between forest soils and water quality and quantity. *For Ecol Manag* **258**: 2269–2281.

Nielsen R (2005). Molecular Signatures of Natural Selection. *Annu Rev Genet* **39**: 197–218.

Ningre J (2007). Le reboisement du Ventoux au XIX siècle : techniques et travaux. *For Méditerranéenne* **XXVIII**: 319–326.

Nybom H (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol* **13**: 1143–1155.

Oddou-Muratorio S, Bontemps A, Gaüzere J, Klein EK (2024). Interplay between fecundity, sexual and growth selection on the spring phenology of European beech (Fagus sylvatica L.). *Peer Community J* **4**: e27.

Oddou-Muratorio S, Gaüzere J, Bontemps A, Rey J-F, Klein EK (2018). Tree, sex and size: Ecological determinants of male vs. female fecundity in three *Fagus sylvatica* stands. *Mol Ecol* **27**: 3131–3145.

Ohta T (1973). Slightly deleterious mutant substitutions in evolution. *Nat Phys Sci* **241**: 20.

Orr HA (2010). Fitness and its role in evolutionary genetics. *Nat Rev Genet* **10**: 531-539.

Pan Y, Birdsey RA, Phillips OL, Houghton RA, Fang J, Kauppi PE, *et al.* (2024). The enduring world forest carbon sink. *Nature* **631**: 563–569.

Parmesan C (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annu Rev Ecol Evol Syst* **37**: 637–669.

Pearse IS, Koenig WD, Kelly D (2016). Mechanisms of mast seeding: resources, weather, cues, and selection. *New Phytol* **212**: 546–562.

Petit RJ, Bialozyt R, Garnier-Géré P, Hampe A (2004). Ecology and genetics of tree invasions:

From recent introductions to Quaternary migrations. *For Ecol Manag* **197**: 117–137.

Petit RJ, Hampe A (2006). Some Evolutionary Consequences of Being a Tree. *Annu Rev Ecol Evol Syst* **37**: 187–214.

Pigliucci M, Murren CJ, Schlichting CD (2006). Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol* **209**: 2362–2367.

Pignatti S, Guarino R (1982). *Flora d'Italia, Vol I*. Ediagricole: Bologna.

Pimm SL (1979). Sympatric speciation: a simulation model. *Biol J Linn Soc* **11**: 131–139.

Piotti A, Leonardi S, Buiteveld J, Geburek T, Gerber S, Kramer K, *et al.* (2012). Comparison of pollen gene flow among four European beech (Fagus sylvatica L.) populations characterized by different management regimes. *Heredity* **108**: 322–331.

Pokorny J, Brom J, Cermak J, Hesslerova P, Huryna H, Nadezhdina N, *et al.* (2010). Solar energy dissipation and temperature control by water and plants. *Int J Water* **5**: 311.

Postolache D, Oddou‐Muratorio S, Vajana E, Bagnoli F, Guichoux E, Hampe A, *et al.* (2021). Genetic signatures of divergent selection in European beech (Fagus sylvatica L.) are associated with the variation in temperature and precipitation across its distribution range. *Mol Ecol*: 1–19.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**: 904–909.

Ramírez-Valiente JA, Santos del Blanco L, Alía R, Robledo-Arnuncio JJ, Climent J (2021). Adaptation of Mediterranean forest species to climate: Lessons from common garden experiments. *J Ecol*: 1–21.

Rellstab C (2021). Genomics helps to predict maladaptation to climate change. *Nat Clim Change* **11**: 85–86.

Rellstab C, Fischer MC, Zoller S, Graf R, Tedder A, Shimizu KK, *et al.* (2017). Local adaptation (mostly) remains local: Reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. *Heredity* **118**: 193–201.

Richardson JL, Urban MC (2013). Strong selection barriers explain microgeographic adaptation in wild salamander populations. *Evolution* **67**: 1729–1740.

Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends Ecol Evol* **29**: 165–176.

Robledo-Arnuncio JJ, Gil L (2005). Patterns of pollen dispersal in a small population of Pinus sylvestris L. revealed by total-exclusion paternity analysis. *Heredity* **94**: 13–22.

Rockman MV (2012). The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution* **66**: 1–17.

Ronce O, Kirkpatrick M (2001). When sources become sinks: migrational meltdown in

heterogeneous habitats. *Evolution* **55**: 1520–1531.

Rosenblum EB (2006). Convergent Evolution and Divergent Selection: Lizards at the White Sands Ecotone. *Am Nat* **167**: 1–15.

Rowland L, Oliveira RS, Bittencourt PRL, Giles AL, Coughlin I, Costa P de B, *et al.* (2020). Plant traits controlling growth change in response to a drier climate. *New Phytol*: 1363–1374.

Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, *et al.* (2007). A Molecular Timetable for Apical Bud Formation and Dormancy Induction in Poplar. *Plant Cell* **19**: 2370–2390.

Ryman N, Laikre L, Hössjer O (2019). Do estimates of contemporary effective population size tell us what we want to know? *Mol Ecol* **28**: 1904–1918.

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, *et al.* (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832–837.

Sáenz-Romero C, Lamy J, Ducousso A, Musch B, Ehrenmann F, Delzon S, *et al.* (2017). Adaptive and plastic responses of *Quercus petraea* populations to climate across Europe. *Glob Change Biol* **23**: 2831–2847.

Santos-del-Blanco L, Olsson S, Budde KB, Grivet D, González-Martínez SC, Alía R, *et al.* (2022). On the feasibility of estimating contemporary effective population size (Ne) for genetic conservation and monitoring of forest trees. *Biol Conserv* **273**: 109704.

Santure AW, Garant D (2018). Wild GWAS—association mapping in natural populations. *Mol Ecol Resour* **18**: 729–738.

Satake A, Kawatsu K, Chiba Y, Kitamura K, Han Q (2019). Synchronized expression of *FLOWERING LOCUS T* between branches underlies mass flowering in *Fagus crenata*. *Popul Ecol* **61**: 5–13.

Savolainen V, Anstett MC, Lexer C, Hutton I, Clarkson JJ, Norup MV, *et al.* (2006). Sympatric speciation in palms on an oceanic island. *Nature* **441**: 210–213.

Savolainen O, Lascoux M, Merilä J (2013). Ecological genomics of local adaptation. *Nat Rev Genet 2013 1411* **14**: 807–820.

Savolainen O, Pyhäjärvi T, Knürr T (2007). Gene Flow and Local Adaptation in Trees. *Annu Rev Ecol Evol Syst* **38**: 595–619.

Scheiner SM (1993). Genetics and evolution of phenotypic plasticity. *Annu Rev Ecol Syst* **24**: 35–68.

Scheiner SM, Lyman RF (1989). The genetics of phenotypic plasticity I. Heritability. *J Evol Biol* **2**: 95–107.

Scotti I, Lalagüe H, Saintagne CS-, Ruiz R, Delphine D, Lefevre F, *et al.* (2022). Common microgeographical selection patterns revealed in four European conifers. *Mol Ecol*: 1–19.

Scotti-Saintagne C, Bodénès C, Barreneche T, Bertocchi E, Plomion C, Kremer A (2004). Detection of quantitative trait loci controlling bud burst and height growth in Quercus robur L. *Theor Appl Genet* **109**: 1648–1659.

Sedjo R, Sohngen B (2012). Carbon Sequestration in Forests and Soils. *Annu Rev Resour Econ* **4**: 127–144.

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, *et al.* (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet 2012 447* **44**: 825–830.

Shim D, Ko J-H, Kim W-C, Wang Q, Keathley DE, Han K-H (2014). A molecular framework for seasonal growth-dormancy regulation in perennial plants. *Hortic Res* **1**: 14059.

Silvertown J (1989). The paradox of seed size andf adaptation. *Trends Ecol Evol* **4**: 24–26.

Simpson GG (1953). *The Major Features of Evolution*. Columbia Press: New York City.

Siol M, Wright SI, Barrett SCH (2010). The population genomics of plant adaptation. *New Phytol* **188**: 313–332.

Slatkin M (1973). Gene flow and selection in a cline. *Genetics* **75**: 733–756.

Slatkin M (1978). Spatial patterns in the distributions of polygenic characters. *J Theor Biol* **70**: 213–228.

Slatkin M, Excoffier L (2012). Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. *Genetics* **191**: 171–181.

Sork VL, Stowe KA, Hochwender C (1993). Evidence for Local Adaptation in Closely Adjacent Subpopulations of Northern Red Oak (Quercus rubra L.) Expressed as Resistance to Leaf Herbivores. *Am Nat* **142**: 928–936.

Soularue J-P, Kremer A (2012). Assortative mating and gene flow generate clinal phenological variation in trees. *BMC Evol Biol* **12**: 79.

Spichtig M, Kawecki TJ (2004). The Maintenance (or Not) of Polygenic Variation by Soft Selection in Heterogeneous Environments. *Am Nat* **164**: 70–84.

Stephan W (2019). Selective Sweeps. *Genetics* **211**: 5–13.

Tamba CL, Zhang YM (2018). A fast mrMLM algorithm for multi-locus genome-wide association studies. *bioRxiv*.

Teissier Du Cros E, Thiebaut B, Duval H (1988). Variability in beech : budding, height growth and tree form. *Ann Sci For* **45**: 383–398.

The 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. *Nature* **467**: 1061–1073.

Theraroz A, Guadaño‐Peyrot C, Archambeau J, Pinosio S, Bagnoli F, Piotti A, *et al.* (2024). The genetic consequences of population marginality: A case study in maritime pine. *Divers*

Distrib **30**: e13910.

Thinon M (2007). La végétation du mont Ventoux au cours des derniers millénaires. *For Méditerranéenne* **26**: 289-294.

Tigano A, Friesen VL (2016). Genomics of local adaptation with gene flow. *Mol Ecol* **25**: 2144–2164.

Tíscar PA, Lucas-Borja ME, Candel-Pérez D (2018). Lack of local adaptation to the establishment conditions limits assisted migration to adapt drought-prone Pinus nigra populations to climate change. *For Ecol Manag* **409**: 719–728.

Turelli M, Barton NH, Coyne JA (2001). Theory and speciation. *Trends Ecol Evol* **16**: 330–343.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010). Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nat Genet* **42**: 260–263.

Urban MC (2015). Accelerating extinction risk from climate change. *Science* **348**: 571–573.

Valladares G, Cagnolo L, Salvo A (2012). Forest fragmentation leads to food web contraction. *Oikos* **121**: 299–305.

Van Valen L (1975). Life, Death, and Energy of a Tree. *Biotropica* **7**: 259.

Via S (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol Evol* **16**: 381–390.

Vitalis R, Dawson K, Boursot P (2001). Interpretation of Variation Across Marker Loci as Evidence of Selection. *Genetics* **158**: 1811–1823.

Vitasse Y, Basler D (2013). What role for photoperiod in the bud burst phenology of European beech. *Eur J For Res* **132**: 1–8.

Vitasse Y, Delzon S, Bresson CC, Michalet R, Kremer A (2009). Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. *Can J For Res* **39**: 1259–1269.

Vitasse Y, François C, Delpierre N, Dufrêne E, Kremer A, Chuine I, *et al.* (2011). Assessing the effects of climate change on the phenology of European temperate trees. *Agric For Meteorol* **151**: 969–980.

Vizcaíno-Palomar N, Revuelta-Eugercios B, Zavala MA, Alía R, González-Martínez SC (2014). The Role of Population Origin and Microenvironment in Seedling Emergence and Early Survival in Mediterranean Maritime Pine (Pinus pinaster Aiton) (S Delzon, Ed.). *PLoS ONE* **9**: e109132.

Von Wühlisch G (2008). *EUFORGEN Technical Guidelines for genetic conservation and use for European beech (*Fagus sylvatica*)*. Biodiversity International: Rome.

Wall JD, Andolfatto P, Przeworski M (2002). Testing models of selection and demography in Drosophila simulans. *Drosoph Simulans Genet* **162**: 203.

Walsh B, Lynch M (2018). *Evolution and Selection of Quantitative Traits*. Oxford University Press: New York, Oxford.

Wang SB, Feng JY, Ren WL, Huang B, Zhou L, Wen YJ, *et al.* (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci Rep* **6**: 1–10.

Watson RT, Albritton DL, Intergovernmental Panel on Climate Change, Intergovernmental Panel on Climate Change, Intergovernmental Panel on Climate Change (Eds.) (2001). *Climate change 2001: synthesis report*. Cambridge University Press: Cambridge ; New York.

Weber JN, Bradburd GS, Stuart YE, Stutz WE, Bolnick DI (2017). Partitioning the effects of isolation by distance, environment, and physical barriers on genomic divergence between parapatric threespine stickleback. *Evolution* **71**: 342–356.

Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure.

Welling A, Moritz T, Palva ET, Junttila O (2002). Independent Activation of Cold Acclimation by Low Temperature and Short Photoperiod in Hybrid Aspen. *Plant Physiol* **129**: 1633–1641.

Whitlock MC (1999). Neutral additive genetic variance in a metapopulation. *Genet Res* **74**: 215–221.

Wright S (1954). The Genetical Structure of Populations. **1**: 1–8.

Wright SI, Gaut BS (2005). Molecular Population Genetics and the Search for Adaptive Evolution in Plants. *Mol Biol Evol* **22**: 506–519.

Yang Q, Gao Y, Wu X, Moriguchi T, Bai S, Teng Y (2021). Bud endodormancy in deciduous fruit trees: advances and prospects. *Hortic Res* **8**: 139.

Yeaman S (2022). Evolution of polygenic traits under global *vs* local adaptation. *Genetics* **220**: iyab134.

Yeaman S, Otto SP (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* **65**: 2123–2129.

Yeaman S, Whitlock MC (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**: 1897–1911.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, *et al.* (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* **38**: 203–208.

Zhang Y-M, Mao Y, Xie C, Smith H, Luo L, Xu S (2005). Mapping Quantitative Trait Loci Using Naturally Occurring Genetic Variance Among Commercial Inbred Lines of Maize (Zea mays L.). *Genetics* **169**: 2267–2275.

Zhou X, Carbonetto P, Stephens M (2013). Polygenic Modeling with Bayesian Sparse Linear Mixed Models. *PLoS Genet* **9**.

Zhu K, Woodall CW, Clark JS (2012). Failure to migrate: lack of tree range expansion in response to climate change. *Glob Change Biol* **18**: 1042–1052.

2. Aims

This thesis aims to understand the spatial, temporal, and genomic pattern of adaptive diversification in the European beech population growing on Mont Ventoux, with a focus on how the chain of different selective pressures along its slope shape the spatial variation of the genetic control of TBB. The work proceeds in three steps: first it search for signals of adaptive diversification at the landscape scale between four tree stands with differ in environment, reciprocal distances, and demographic history; it moves then to investigate the genetic control of spring phenology, a trait which has been proved to be adaptive and phenotypically diversified along the northern slope; and to predict the breeding values of stands along the gradient, a way to introduce the modelling of the evolutionary fate of these stands under climate change.

The studies cover different spatial scales across the slopes of Mont Ventoux (**Figure 1**): Chapter involves four stands in contrasting environmental conditions and with different demographic histories, two neighbouring stands of high elevation on the northern slope and two on the southern slope, one old-growth stand high elevation and one recently incepted one at low elevation; Chapter 2 focus on the a single stand on the northern slope, a little low than those studied in the previous Chapter; Chapter 3 covers the whole mountain, with three, previously studied stands and a new one on the northern slope and the two stands of the southern slope.

The first two Chapters take different, apparently opposite paths: whilst Chapter 1 investigates signatures of putatively adaptive divergence *between* stands remaining agnostic of the trait affected, Chapter 2 deals with the relationship between genotype and phenotype *within* a single stand. Chapter 3 joins the findings of the previous steps to predict the spatial pattern of genetic divergence for a phenotypic trait (time of bud-break). It also evaluates whether the genetic diversity at the identified QTLs suffice for populations of higher elevation to evolve toward a phenotypic optimum nowadays typical of lower elevations. Whilst the result of these simulations are mostly as a proof of concept for the evolutionary potential of TBB, they also provide insights (to handle *very* carefully) about the fate of beech populations under a warming climate.

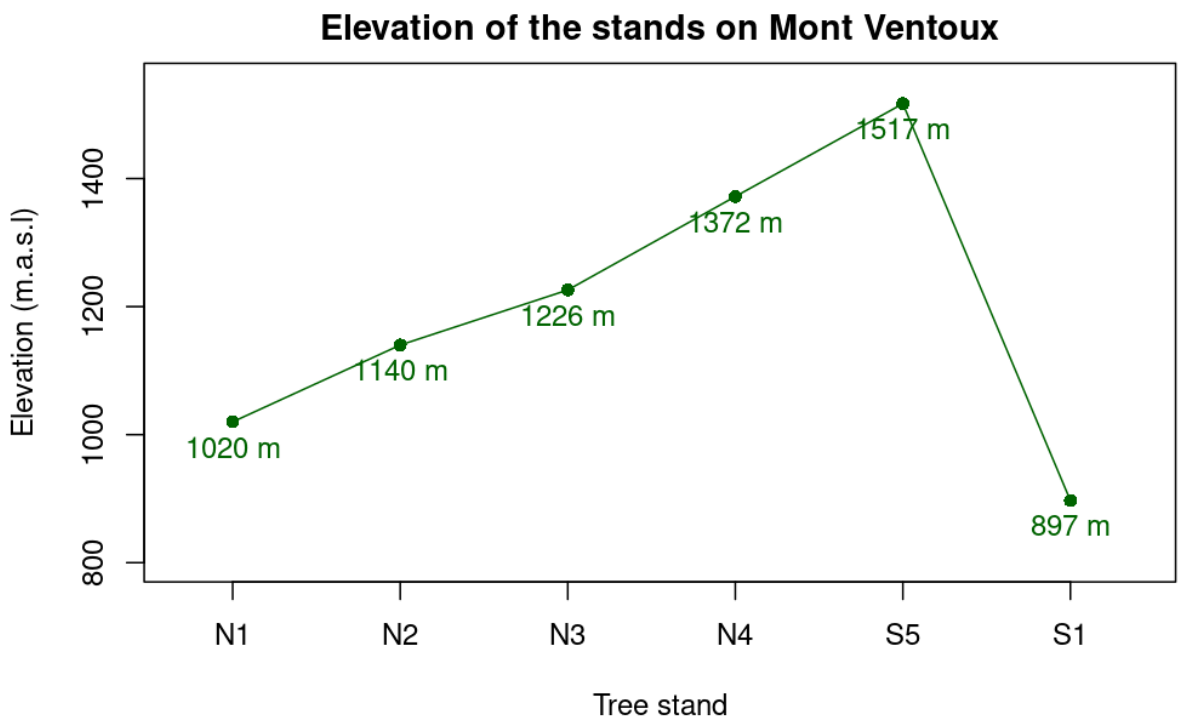


Figure 1: Stands of European beech on Mont Ventoux studied in this thesis.

(i) Investigating Microgeographic Adaptation

In the **Chapter 1** of the thesis I investigate molecular adaptive differentiation at the landscape scale between tree stands that differ for environmental conditions, demographic history and reciprocal distance between each other. Three stands belong to the old-growth forest of high elevation and two of them are very close to each other, while the fourth has been established during the last 150 years during a downward demographic expansion. On the basis of archival records and previous studies on the population's recent past, I am going to control for the signature left by the demographic expansion.

(ii) Genomic Architecture of the Timing of Bud Burst

Chapter 2 concerns the characterisation of the genomic bases of the bud-burst and trait related to spring phenology using a Genotype-Phenotype Association experiment on a stand on the northern slope (N1). Precision of the record of phenological observations is improved by modelling the relationship *PhenologicalScore ~ DayOfTheYear* with a logistic function, and power of the dataset to detect signals is assessed with a training/testing approach.

(iii) Modelisation of the Response to Climate Change

Based on common garden findings of an anticline of spring phenology along the northern slope, **Chapter 3** evaluates whether the genomic bases identified in **Chapter 2** can explain the observed pattern. The loci retrieved in the previous chapter are mapped on the SNP panel used in the **Chapter 1** and breeding values of the stand along the gradient are computed from the estimated allelic effect in N1. By changing the phenotypic optimum of a population, forward simulations predict the pattern of genetic diversity under changing climatic conditions.

Chapter 1:

Rolling down that mountain: microgeographical adaptive divergence during a fast population expansion along a steep environmental gradient in European beech

Andrea Modica[1], Hadrien Lalagüe[2], Sylvie Muratorio[3], Ivan Scotti[1](*)

This article has been published on Heredity as:

Andrea Modica, Hadrien Lalagüe, Sylvie Muratorio, Ican Scotti, (2024). **Rolling down that mountain: microgeographical adaptive divergence during a fast population expansion along a steep environmental gradient in European beech**. Heredity, 133:99–112; https://doi.org/10.1038/s41437-024-00696-z

Authors:

[1]INRAE-URFM, Avignon, France

[2]INRAE-EcoFoG, Kourou, French Guiana

[3]INRAE-EcoBioP, Saint-Pée-sur-Nivelle, France

* Corresponding author. ivan.scotti@inrae.fr

Summary

Forest tree populations harbour high genetic diversity thanks to large effective population sizes and strong gene flow, allowing them to diversify through adaptation to local environmental pressures within dispersal distance. Many tree populations also experienced historical demographic fluctuations, including spatial population contraction or expansions at various temporal scales, which may constrain their ability to adapt to environmental variations. Our aim is to investigate how recent contraction and expansion events interfere with local adaptation, by studying patterns of adaptive divergence between closely related stands undergoing environmentally contrasted conditions, and having or not recently expanded.

To investigate genome-wide signatures of local adaptation while accounting for demography, we analysed divergence in a European beech population by testing pairwise differentiation among four tree stands at ~35k Single Nucleotide Polymorphisms from ~9k genomic regions. We applied three divergence outlier search methods resting on different assumptions and targeting either single SNPs or contiguous genomic regions, while accounting for the effect of population size variations on genetic divergence.

We found 27 signals of selective signatures in 19 target regions. Putatively adaptive divergence involved all stand pairs. We retrieved signals both when comparing old-growth stands and recently colonised areas and when comparing stands within the old-growth area. Therefore, adaptive divergence processes have taken place both over short time spans, under strong environmental contrasts, and over short ecological gradients, in populations that have been stable in the long term. This suggests that standing genetic variation supports local, microgeographic divergence processes, which can maintain genetic diversity at the landscape level.

Introduction

The patterns of genetic diversity currently displayed by wild populations are the result of multiple demographic, ecological and evolutionary processes triggered by historical events, such as long-term climatic cycles, disturbances, and land-use change by humans. Understanding and/or predicting patterns of genome diversity requires unveiling the complex interactions between the demographic fluctuations caused by historical contingencies, on the one hand, and the intrinsic capacity of populations to evolve in response to these contingencies, on the other hand (e.g. selection is less effective, and migration more effective, in smaller populations, Hedrick, 2000). Conversely, evolutionary forces can shape population demography (e.g. inbreeding can depress fecundity, Charleswoth and Charlesworth, 1987; hard selection (i.e., the situation in which an individual's fitness depends only on its own phenotype) may reduce population size; Bell *et al.*, 2021). Such feedback loops are often hard to tease apart, and this can be particularly challenging for long-lived species playing key ecological roles, such as widespread forest trees. Indeed, their often large populations can persist over tens of millennia, involving complex demographic histories, marked by range shifts, long stretches of time in isolation from other populations, multiple disturbance regimes and the ever-present danger posed by anthropogenic deforestation (Petit and Hampe, 2006). Yet, deconvoluting the intricate roles of demographic and selective histories on tree population diversity and structure is all the more urgent in view of the ongoing biodiversity crisis, which calls for the implementation of adaptive management or/and conservation strategies for forest ecosystems benefitting from and/or caring for intraspecific diversity and evolutionary processes (Godineau *et al.*, 2023).

A wide range of methods have been developed in the last decades to infer population demographic (reviewed in Beichman *et al.*, 2018) or selective histories (reviewed in Nielsen, 2005; Bank *et al.*, 2014) from patterns of genomic diversity. Across non-model taxa, and specifically in trees, the predominant approaches used for uncovering sets of loci underlying local adaptation have neglected explicit demography (Lind *et al.*, 2018). These approaches generally rely on population divergence patterns and can be roughly divided into two categories: (1) those relying on the partition of between-population genetic variance (resting on the approach originally formulated in Lewontin and Krakauer, 1973) to identify loci showing unusually high or low between-population differentiation (typically measured by F_{ST}) and (2) those relying on the regression of population allele frequencies onto one or more environmental variables to identify loci showing strong correlations (i.e., gene-environment association, GEA). In the first category, some explicit variants of the island model (including hierarchical versions) were historically used to obtain neutral F_{ST} expectations against which to identify divergence outliers (e.g. Foll *et al.*, 2008). More recently, for both categories, more flexible, less model-based parameterization approaches have been used where the null model generally assumes that the allele frequencies in each population may deviate away from ancestral (or global) allele frequencies due to genetic drift and gene flow, and where this covariance structure is accounted for to estimate divergent selection (e.g. Coop *et al.*, 2010; Gautier, 2015). Although these different approaches to detect divergent selection intentionally avoid using any explicit model of the historical relationships between populations, they were shown to be remarkably robust to demographic history (De Villemereuil *et al.*, 2014; Lotterhos and Whitlock, 2014). At

the same time, many genomic studies confirmed that the current level and distribution of genetic diversity of tree populations show pervasive signatures of demographic fluctuations, be they ancient (such as the well-studied post-glacial expansion history, Petit *et al.*, 2003) or recent (Zhao *et al.*, 2020). A major risk when studying adaptation in recently expanded populations is that the marks left by demographic events on genomic patterns of diversity can mimic adaptive patterns. The classic case is allele surfing, when a neutral allele increases in frequency across the expansion range due to genetic drift and founder effect (Excoffier and Ray, 2008; Excoffier *et al.*, 2009). A typical example of the combined effect of range / demographic expansion and concurrent adaptation on population genetic structure is provided by Zhao *et al.* (2020), who manually removed geographically structured loci from the list of loci showing association with environmental variables. In addition, new adaptations arising during expansion can contribute to the success of the expansion itself, as shown in the case of the invasive plant, *Lythrium solicaria* L., by Colautti and Barrett (2013). Importantly, it is likely that the distribution of genetic diversity in recently expanded populations is far from any migration-selection or migration-selection-drift equilibrium. The incorporation of the demographic background in any population model describing such non-equilibrium situations is thus essential: any inference of the selective history should explicitly test selection against the backdrop of ongoing or recent demographic changes. Any conclusion drawn from the analysis of such “space with time” models is likely to provide more realistic predictions on how climate change will affect patterns of local adaptation than do the standard “space for time” approaches resting on stable, supposedly fully locally adapted populations (Hansen *et al.*, 2012).

In this study, we advocate for a generalised approach to the incorporation of detailed demographic models, while taking into account the uncertainties of the estimation of demographic model parameters, into the search for divergence outliers. The approach we propose involves the generation of simulated neutral distributions of the chosen divergence statistic, based on the demographic model that has been previously determined, followed by the comparison of the empirical values of the statistic with the simulated distribution. Myers et al. (2008) suggest that a given genomic data set is compatible with large numbers of alternative demographic scenarios, and Lapierre et al. (2017) suggest that small errors in genomic data can lead to major biases in demographic inference. By simulating data from a range of estimates of demographic parameters, our method accounts for the uncertainties associated with demographic inferences from allele frequency data; because such simulations can be performed for any demographic model, data set, and divergence statistic, the method is general.

To investigate the joint demographic and selective histories in recently expanded tree populations, we focused on relatively common situations encountered in European areas that were densely populated by humans in the (pre)historic past. A general trend exists in Europe, with deforestation steadily reducing forest cover from the Holocene until the industrial revolution, followed by a partial reversal of this trend starting mid-nineteenth century (Kaplan *et al.*, 2009; Zanon *et al.*, 2018). The slopes of Mont Ventoux, in southern France, are a typical example, where anthropogenic pressure has been increasing since the Neolithic, throughout historical times, until it tapered off between the end of the nineteenth century and mid-twentieth century. From 1860, a massive rural flight, locally accompanied by measures aiming at reforestation and at supporting natural forest regeneration, favoured the expansion of forest remnants into newly

thriving forests (Ningre, 2007; Martin-Gousset *et al.*, 2019; Gamba *et al.*, 2023). The remnants of beech stands having persisted in historical times were restricted to high-elevation forest patches, and they gradually colonised land at lower elevations, thus meeting conditions that are both warmer and wetter. Our hypothesis is that such population expansions involved both demographic changes and adaptive processes, even on these restricted spatial scales, causing a reshuffling or filtering of genetic diversity within the population. In agreement with these differential selection pressures, signatures of local adaptation were identified for growth and phenological characters (Gaüzere *et al.*, 2020) based on a quantitative genetic approach. Moreover, a candidate-gene study suggested adaptive divergence at the molecular level (Csilléry *et al.*, 2014). Such characteristics make it likely that the recolonisation process has left genomic signatures of microgeographic adaptation, and it should therefore be possible to identify genetic loci showing divergence both between sub-sections of the historical relict areas, historically exposed to different environmental conditions (e.g., north- vs south-facing slopes), and between historical relict and recently colonised areas.

The present study takes advantage of this well-studied population to investigate genome-wide signature of local adaptation while accounting for demography. We build on the extant knowledge of the Mont Ventoux beech population history, which was previously investigated based on paleoecological and historical records as well as neutral genetic diversity (Lander *et al.*, 2011, 2020; Gamba *et al.*, 2023). This exceptional knowledge of the recent demography of the study population made it possible to finely calibrate a selection inference method accounting for population demography. Another major asset on which this study relies is the recent establishment of a high-quality reference genome sequence for *F. sylvatica* (Mishra *et al.*, 2022). Studies relying on annotated forest tree genomes are so far restricted to few species (*Pinus taeda* L., Lu *et al.*, 2019; De La Torre *et al.*, 2019; *Quercus robur* L., Leroy *et al.*, 2020; *Populus* spp., Liu *et al.*, 2022).

To carry out our study, we focused on single-copy gene regions of the *F. sylvatica* reference genome sequence (Mishra *et al.*, 2022), and on four stands belonging to the Mont Ventoux beech population (three in historical relict areas and one in a recently colonised area). These stands show varying degrees of divergence in environmental conditions (Table 1), and lie at varying spatial distances. The combination of these three variables leads to different expectations on adaptive divergence patterns, depending on the strength of selective pressures in time and space (**Figure 1**). Our results show (adaptive) molecular divergence at all hierarchical levels of population subdivision.

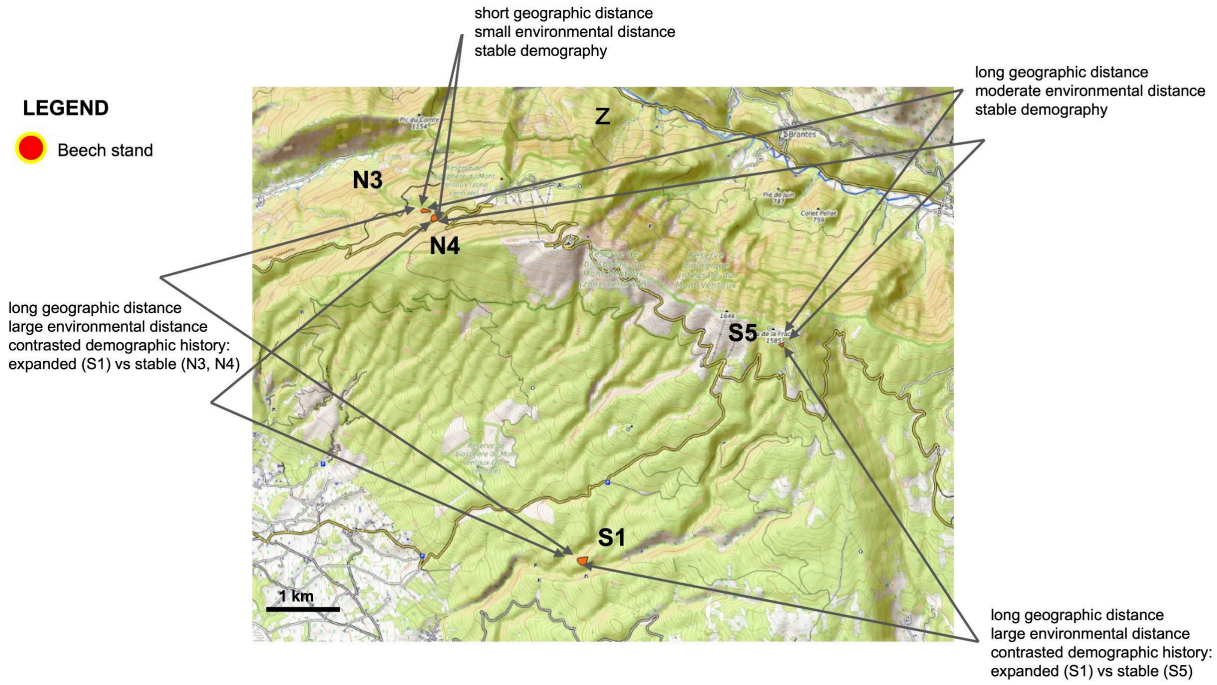


Figure 1 Qualitative description of combinations of levels of temporal, spatial, and environmental divergence for population pairs. Notice that “long” and “short” should be interpreted in a landscape-level context, not to be confused with processes occurring at larger divergence scales (regional, rangewide, phylogenetic). See Materials and Methods and Table 1 for the detailed description of the sampled stands.

Material and Methods

Species, Study Site, and Sampling

Fagus sylvatica L., the European common beech, is a dominant broadleaved tree species of many lowland and mountain forests across Europe, extending from Spain to the Carpathians and from Sicily to southern Sweden. Its genome is diploid with 2n = 24 chromosomes. It has a monoecious, anemophilous, and mainly allogamous mating system (Vornam *et al.*, 2004). Sexual maturity is reached at 40-50 years of age. Seeds are mostly produced in mast years and are mainly dispersed at short distances (mean primary dispersal distance of 25 m, Bontemps *et al.*, 2013), but non-negligible long-distance dispersal can be achieved (typically with birds such as the European jay, *Garrrulus glandarius* (Nilsson, 1985).

The population we have studied on Mont Ventoux is one of the 52 glacial refugia identified by Médail and Diadema (2009) in the Mediterranean basin. Mont Ventoux is a 1,920 m-high mountain located in southern France, with a steeper, colder northern slope and a flatter, warmer southern slope. Both harbour several floristic altitudinal zones encompassing a wide range of ecosystems, from mediterranean to alpine, with beech occurring in most of them (Barbero *et al.*, 1978), on both slopes.

In historical times, beech populations have been restricted to a relict area on the top of both the north and the south slope, with an extent that shrinked progressively until reaching its minimum in the mid-19th century. The paleoanthracological survey of Gamba et al. (2023) reports the retrieval of remnants of beech charcoal at 1,130 m of altitude on the northern slope and dated them to 346-460 years before present (1490-1604 CE). Since 1850, beech stands expanded continuously until almost doubling the surface they occupied, reaching in the process elevations of 300-600 m below the lower margin of the historical relict areas. A previous investigation of demographic scenarios using coalescent-based simulations and microsatellite genetic data revealed the genetic signature of population expansion, but not that of bottlenecks (Lander et al. 2011), with the total “ancestral” (i.e., refugial) effective population size estimated to approximately 1,000 to 13,000 individuals (Lander *et al.*, 2011). Weak but significant divergence among different portions of the historical relict population was detected, showing that, in spite of the close proximity of the forest patches, gene flow was somewhat restricted, allowing sub-populations to differentiate (mean divergence of each subpopulation from the global population β_{WT} = 0.03; Lander *et al.*, 2011). On the contrary, no sign of differentiation was detected between recently expanded forests and the relicts they expanded from (Lander *et al.*, 2021), suggesting that the time elapsed since the start of the expansion is insufficient for neutral divergence processes to leave genome-wide marks. Therefore, the beech stands of Mont Ventoux show signs of recent, ecologically-driven divergence (Csilléry *et al.*, 2014; Gaüzere *et al.*, 2020), in the absence of neutral divergence (Lander *et al.*, 2021), which enables the detection of even relatively weak signals of adaptive genomic divergence.

To carry out this study, we sampled 20 to 26 adult trees from four stands identified in Lander et al. (2020): N3 and N4 on the northern slope and S1 and S5 on the southern slope (see Table 1

for details on elevation and environmental parameters). N3 and N4 belong to a continuous mixed forest of beech and silver fir (*Abies alba* Mill.) which extends over a steep altitudinal gradient, and are close to each other, even though they are separated by 150 m in elevation, and thus have a difference in mean temperature of about 2 °C. The S5 stand grows at the highest elevation, and is a part of a continuous beech forest encompassing the mountain ridge. Its mean annual temperature is intermediate between N3 and N4, but it experiences colder winter temperatures. Summer vapour pressure deficit is similar among N3, N4, and S5, but the latter is globally the least exposed to drought (Table 1, stress May to September). S1 is a mixed stand of beech and oak (*Quercus* spp. L.) located on a small gorge within the lowest part of the south-eastern mountainside, and it is the population at the lowest elevation, experiencing the highest mean temperatures and the most severe drought stress (Table 1, stress May to September and self-calibrated Palmer Drought Severity Index, scPDSI). According to Lander et al. (2011; 2020), stands N3, N4, and S5 belong to the historical relict area, while stand S1 is the result of the demographic expansion process that occurred between the XIX and the XX centuries, likely originating in the upper south slope to which S5 belongs. In the Supplementary Materials of Lander *et al.* (2021), N3, N4, S1, and S5 correspond respectively to W_N3, W_384, S_1, and S_5.

Genotyping by sequencing

The genotypes were obtained based on data from a sequence capture experiment following Peñalba et al. (2014). The sequence capture was carried out by RapidGenomics (Gainesville (FL), US) with the probes reported in Supplementary Table ST1. The probes were based on a beech unigene set (Lesur *et al.*, 2015), and therefore the targets were exclusively coding sequences. Probe preparation, fragment capture, sequencing, *de novo* assembly, read mapping, as well as variant calling and filtering and the preparation of genotype tables, were executed as described in Scotti et al. (2023; the whole process was actually carried out concurrently with the treatment reported there). In summary, reads were *de novo* assembled using the velvet algorithm (Zerbino 2010) with K-mers (that is, the substrings used to identify alignments; see Compeau et al. 2011) of length 85 nucleotides, coverage cut-off = 8, minimum contig length = 100 and expected coverage = 20. Because assembly algorithms are very sensitive to polymorphism that prevents contigs from being correctly assembled, we carried out *de* novo assembly only from one individual (Sample Fs_T_S1_005), chosen because it had the largest number of high-quality reads.

Notice that the data were obtained and curated before the beech genome reference (Mishra *et al.*, 2022) was obtained and published, and therefore, at the time of bioinformatics data treatment, *de novo* assembly was the best available strategy. The *de novo* contigs were mapped to the reference genome (Mishra *et al.*, 2022) for further curation, insight and annotation, using BLAST 2.9.0+ (Altschul *et al.*, 1990; Zhang *et al.*, 2000). To obtain the positions of target regions and SNPs on the reference genome, we proceeded as follows. We first extracted the best hit for each target region and retrieved the positions of each nucleotide in the target regions relative to the reference scaffolds (including gaps in the alignment). Pairs or small groups of contigs mapping to the same genomic regions, or with no hits, were removed

from the dataset. We then built a correspondence table providing the position of each nucleotide of each target region on reference scaffolds. Using this table, target region positions of SNPs were converted into scaffold positions, which were used in all subsequent analyses. For analyses at the target region level, the position of each target region on the reference genome was defined as the start of its alignment on the scaffolds. The random distribution of target regions along chromosomes was also tested using a test for conformity to a Poisson distribution (see Supplementary Methods SM1 for details).

Based on the reference contigs obtained above, we performed read mapping with BOWTIE2 v 2.0.6 (Langmead & Salzberg, 2012) using “very- fast-local” parameters; we used SAMTOOLS MPILEUP (Li et al., 2009) to build a mpileup file with all the individual alignments to the reference The parameters applied for variant calling (using VARSCAN v. 2.3.9 , Koboldt et al., 2009) were the following: minimum coverage = 20, minimum average quality = 30, minimum variant frequency = 0.3, minimum frequency for homozygote call = 0.7, maximum missing data = 20%. We further filtered the dataset according to the distribution and the biological properties of the variant were used to further filter the dataset with the ad hoc R (R Development Core Team, 2008) *sieve()* script. The *sieve()* script works on a VCF file and filters out contigs based on SNP density (variants per base) and heterozygosity thresholds; it also returns genotype tables. We excluded the contigs with variant density > 0.05 (i.e., having more than one variant every 20 bases; *varDensThreshold* argument) and carrying only heterozygous genotypes to remove contigs potentially containing paralogues (*hetThreshold* argument = 1). The filtering process was refined by removing monomorphic sites with the ad-hoc *monoRemove()* R script.

Diversity and differentiation statistics

We carried out diversity and differentiation analyses in R v. 4.1.2 (R Core Team, 2023), unless stated otherwise. The values of π per base were computed for each population as the average π over all nucleotides of that sequence, including the invariant ones. Global F_{ST}, following (Weir and Cockerham, 1984), and its variance components relative to stand and slope, were computed with the function *varcomp.glob()* from the package Hierfstat 0.5-7 (Goudet, 2005). We assessed the 95% confidence intervals for the global statistics by bootstrapping 1,000 times the loci with *boot.vc()*. We also estimated F_{ST} values for each stand pair with the function *pairwise.WCfst()*. The hypothesis of non-null genetic differentiation was tested by bootstrapping over loci 1,000 times with *boot.ppfst()*, and checking significance both at the 95th and 99th quantile. We obtained an UPGMA clustering tree from pairwise F_{ST} distances using Phangorn 2.5.5 (Schliep, 2011).

Because F_{ST} estimates of divergence are sensitive to within-population diversity (Jost, 2007; Lefèvre and Gallais, 2020) we also computed pairwise estimates of Jost’s D (Jost, 2008), which is unaffected by such biases (see DataVerse repository for the script). We tested the hypothesis of non-null genetic differentiation by bootstrapping over loci 1,000 times and assessed the significance at the 0.95 confidence interval. See DataVerse repository for the scripts (n°4-10).

Demographic model and simulation of genotypic data

We built on the previous work of Lander *et al*. (2011), which allowed the major demographic events to be identified and the parameter range to be set. We adapted their demographic model to our sampling design with four stands as follows and as illustrated in Figure 4. We used the demographic parameters they estimated to inform the starting parameter ranges for maximum-likelihood estimation. Lander et al. (2011, 2021) also report that the Mont Ventoux beech stands reached their minimum extent in the 1850s, while Gamba et al. (2023) report beech charcoal in pits of the northern slope of Mont Ventoux from 925 to 1,340 m a.s.l. and dated most of them not earlier than the Middle Ages. They also dated beech charcoal found at 1,130 m a.s.l. to the Modern Era, confirming the existence of beech trees on the northern slope before the reforestation program started in the 1850s..A continental-scale palynological survey by Magri (2008) observed that, after the Wûrm glaciation, beech naturally expanded from its glacial refugia. However she noted that climate and anthtopogenic pressures may have locally affected its colonisation patterns. Martin-Gousset *et al.* (2019) provided further information on the dynamic of the post-Neolitic deforestation dynamic on the Mont Ventoux. These dates allowed us to define starting parameter ranges for historical expansion and contraction dates. We relied on these pieces of information and the divergence estimates obtained from our genomic data (see above and Results section) to build our scenario, that expands from the event times and population sizes, estimated by Lander et al (2011, 2021), further back in the past until the Holocene expansion: we started with an expansion (simulating the Holocene colonisation following climate warming; TIMEE to TIMED, GROWTHRATE2), followed by a time of stability (TIMED to TIME C) and contraction (matching the progressive deforestation since the Neolithic; TIMEC to TIMEB, GROWTHRATE1), until reaching the minimum-extent size, follwoed again by stability (TIMEB to TIMEA); and finally, we allowed the four stands to expand again and diverge starting from the minimum-extent population, while exchanging migrants (TIMEA to TIME3, GROWTHRATE0), until present tim. The demographic model included 21 parameters in total (Supplementary Table ST3).

To identify the most-likely values for demographic parameters, and to simulate genetic data matching the estimated demographic history of the populations, while taking into account uncertainties in model parameter estimation (Myers *et al.*, 2008; Lapierre *et al.*, 2017), we used the following three-steps strategy (see Supplementary Methods SM2 for details):

(a) the first step consisted in estimating the 21 parameters from empirical data. To do this, we used the Maximum-Composite-Likelihood estimation tool implemented in FastSimCoal2 v2.7.0.5 (Excoffier and Foll, 2011), based on the folded pairwise site-frequency spectra (2D-fSFS) for each population pair (see Supplementary Methods SM2 for details on how the 2D-fSFS were obtained from genotype data). Note that the FastSimCoal2 optimisation algorithm is allowed to sample parameter values outside the boundaries set by the user, so the maximum likelihood estimations of the parameters (**Table 2**) do not necessarily reflect their starting ranges (see Supplementary Methods SM2). We performed 100 runs to obtain a distribution of values for each parameter;

(b) the second step consisted of sampling meaningful combinations of correlated parameter values obtained in (a). Since several parameters were correlated partly by construction, because of their boundaries were defined by other parameters, we sampled them together to avoid) ending up with biologically implausible combinations of parameter values as input for simulations. Hence, we analysed the correlation among parameters and identified “correlation blocks”. We then selected a “core” parameter in each block (i.e., the one with the strongest correlation with all others in the block) and drew 100 values for this core parameter from a uniform distribution, with boundaries set to the minimum and maximum of the distribution obtained in (a). The values of the remaining parameters were computed by performing linear regressions between them and the value of the “core” parameters values (including error). This allowed us to generate 100 new sets of values for the 21 parameters.

(c) In the third step, these 100 sets of parameter values were used as input for FastSimCoal2 to simulate genotypes for four populations, which we could use as the demographic-neutral reference in the following outlier detection analyses (see below for details). In each simulation (each starting from one set of parameter values), we simulated 750 1000-bp DNA independent fragments, with per-base recombination rates obtained from the comparison of the linkage map distances and physical genome distances reported in Mishra et al. (2022); we used the same mutation rate of 9.5×10^{-10} reported in Milesi et al. (2023). This way we obtained 100 genome-level 4-population SNP data sets, which were compiled together to obtain a single data set (“*simulated data set*” from here on) with composite diversity properties, representing the properties of data obtained from all sets of parameter values.

Detection of divergence outlier loci between stand pairs

We applied three different methods (see below) to identify genomic loci that show unusual allele frequency differentiation among stands (divergence outlier loci in the following). For each method, we followed the same approach to account explicitly for population demography,, as described below.

We first applied the method on the *simulated data set*, to generate the neutral, demography-driven *simulated distribution* of the method-specific divergence statistic for each population pair. Then, we applied the method on the empirical loci, thus obtaining single-locus values of the divergence statistic and an *empirical distribution* of the divergence statistic.

To identify divergence outliers, we compared the value of the divergence statistic obtained for each empirical locus and stand pair to the corresponding *simulated distribution*. This allowed us to compute a P-value for each empirical value of the statistic as the normalised rank $r_{i,p}$ of the value of the statistic for locus *i* for stand pair *p* relative to the simulated distribution for stand pair *p*. The P-value $P_{i,p}$ was computed as $P_{i,p} = r_{i,p} / N$, with N = total number of loci in the simulated distribution. P-values were converted into q-values using the R package “qvalue” (Storey *et al.*, 2023). An FDR = 0.01 was used as a threshold to identify outliers.

The comparison of the *empirical* and *simulated* distributions of the divergence statistic allowed us to assess whether the outlier tests of our methods were liberal, unbiased, or conservative

(Figure 2): when the *simulated* distribution turned out to be shifted to the *left* relative to the *empirical* distribution, we considered the test as liberal (i.e., too many empirical values exceeded the threshold built on the simulated distribution); when the *simulated* distribution turned out to be shifted to the *right* relative to the *empirical* distribution, we considered the test as conservative (i.e., too few empirical values exceeded the threshold built on the simulated distribution); when there was no difference between the distributions, we considered the tests as unbiased. A two-sided Wilcoxon test was used to assess the shift between distributions.

We tested signatures of selection with three methods resting on different assumptions, with null expectations obtained, for each of them, based on the aforementioned demographic simulations:

(1) The G2D statistic (Nielsen *et al.*, 2009) is based on the calculation of the multinomial composite likelihood ratio (CLR) of observing a given combination of allele frequencies, at the SNPs carried by a given stretch of DNA sequence, in a pair of stands, relative to the expectation obtained by genome-wide allele frequencies. It therefore applies to DNA sequences carrying at least two SNPs and uses 2D-fSFS as a starting point. G2D was computed using the equations for the G statistic as in Nielsen et al. (2005), as modified to apply to a 2D-fSFS. The analyses were carried out at the contig level (see the script in°18 on the online repository for details of the calculation of the G2D statistic).

(2) For each stand pair and for each locus, the absolute difference in allele frequency was computed between stands, and then the distribution of allele frequency differences was obtained for all SNPs for a given stand pair, both for empirical and simulated data.

(3) Bamova (Gompert and Buerkle, 2011) estimates components of molecular variance, genome-wide and for individual loci. To run our analyses, we extracted the modes of locus-specific posterior distributions of Φ_{ST} as the divergence statistic, and proceeded as described above to identify outliers. See Supplementary Methods SM3 for details.

Notice that the first method (G2D) assesses divergence at the level of a whole sequence (here, a target region), while the other methods assess divergence at the single SNP level.

"Standard" (non-demography controlled) Bamova outlier search

For comparison, we also applied the "standard" Bamova (Gompert and Buerkle, 2011; Scotti et al. 2023) method to the raw data (that is, without accounting for population demography). Note that G2D and allele frequency difference tests can only be run with a demographic model, and that allele frequency difference (AFD) has no theoretical expectation and therefore no parametric test could be applied to it. Instead, Bamova relies on a theoretical island population model as a null model, so we were able to test the methods under their theoretical assumptions, following Scotti et al. (2023). Outlier loci were identified based on the Bayes Factor statistics (Makowski *et al.*, 2019), the difference between the value of the prior and the posterior densities at x – that is, the ratio of the data under the null hypotesis and the alternative hypothesis. Hence, results are drawn directly from the observed data.

Table 1

Characteristics of the stands; all measurements refer to years 1959-1990 (Muratorio et al. in prep.). **Tmin**, **Tmax** = average minimum, maximum temperature; **sum GDD** = sum of the number of days with temperature > 5.5 °C; **NFD** = Number of frost days, number of days with temperature < 4 °C; **VPD** = vapour pressure deficit (the difference between the actual amount of air moisture and the amount of air moisture at saturation); **stress May to September** = difference between evapotranspiration and precipitations; **scPDSI_** self-calibrated Palmer Drought severity index (Water stress between May and September (stress) was calculated as the difference between ETP and precipitation, and the self-calibrated Palmer Drought Severity Index (scPDSI) was calculated using the scPDSI R package (Wells, 2004)). See Davi and Cailleret (2017) for raw data.

Stand	origin	Slope	Elevation (m a.s.l.)	Tmax (°C)	Tmin (°C)	sum of GDD (d)	NFD (d)	VPD May to September (mbar)	stress May to September (mm)	scPDSI May to September (adimensional)
N3	Refugial	N	1226	20.4	-2.0	2496	27	414.1343	102.5	0.55
N4	Refugial	N	1363	18.5	-2.8	2187	35	391.9002	69.2	0.44
S1	Expanded	S	895	25.2	-0.5	3531	16	620.3911	167.3	-0.26
S5	Refugial	S	1517	19.4	-3.8	2162	52	421.0679	18.7	0.54

Table 2.

Summary statistics of the 100 independent estimates for each of the 21 demographic simulation parameter. See text and Supplementary Table ST3 for the description of simulation parameters. The mode corresponds to the peak to the density distribution of the parameter (see Supplementary Material).

	Description	**lower CI boundary**	**median**	**mode**	**upper CI boundary**
POPSIZE-N3	Current size of population N3	601	23383	18521	37882
POPSIZE-N4	Current size of population N4	42482	44528	46345	46573
POPSIZE-S1	Current size of population S1	652	23729	30362	38128
POPSIZE-S5	Current size of population S5	304	20242	29609	33290
MIG N3-N4	Symmetrical migration rate for all population pairs (independent draws for each population pair)	0.0013	0.0022	0.0021	0.0057
MIG N3-S1		0.0014	0.0026	0.0043	0.0070
MIG N3-S5		0.0013	0.0024	0.0028	0.0067
MIG N4-S1		0.0013	0.0022	0.0015	0.0056
MIG N4-S5		0.0013	0.0022	0.0035	0.0058
MIG S1-S5		0.0012	0.0023	0.0062	0.0061
TIME1	Time of N3-N4 merger	1	1	1	3
TIME2	Time of S1-(N3-N4) merger	1	1	1	3
TIME3	Time of final population merger (S5-(S1-(N3-N4)))	1	1	1	5

TIMEA	End of population contraction	2	2	2	12
GROWTHRATE0	Growth rate until TIMEA	0.00443	0.00313	0.00366	-0.00010
TIMEB	End of population stability at minimum size	2	4	3	16
TIMEC	End of population expansion	90	98	99	99
GROWTHRATE1	Growth rate until TIMEC	-0.00010	-0.00127	-0.00125	-0.00544
TIMED	Start of population contraction	101	159	103	268
GROWTHRATE2	Growth rate until TIMEE	0.00001	-0.00001	-0.00004	-0.00004
TIMEE	End of population contraction	176	937	987	989

Annotation of SNPs and genes

We used SnpEff 4.3 (Cingolani *et al.*, 2012) to functionally annotate each outlier SNP and target region relative to the reference genome annotation (Mishra *et al.*, 2022) (see Supplementary Methods SM4). We retrieved the gene-level functional annotation from the GFF3 annotation reported in Mishra et al. (2022) and matched with orthologs with the same accession on GeneBank.

To search for correspondences with published studies, the genomic regions containing outliers were compared, using Blast (or manually, when locus positions relative to Mishra et al. (2022) were reported), to outlier sequences reported in Müller (2017; drought response expression outliers), Pfenninger et al. (2021; association to drought damage), Csilléry et al. (2014; altitudinal divergence in the same populations as in the present study), Lesur et al. (2015; differential expression during budburst), Postolache et al. (2021; rangewide signals of adaptive divergence), Meger et al. (2021; adaptive divergence for phenology).

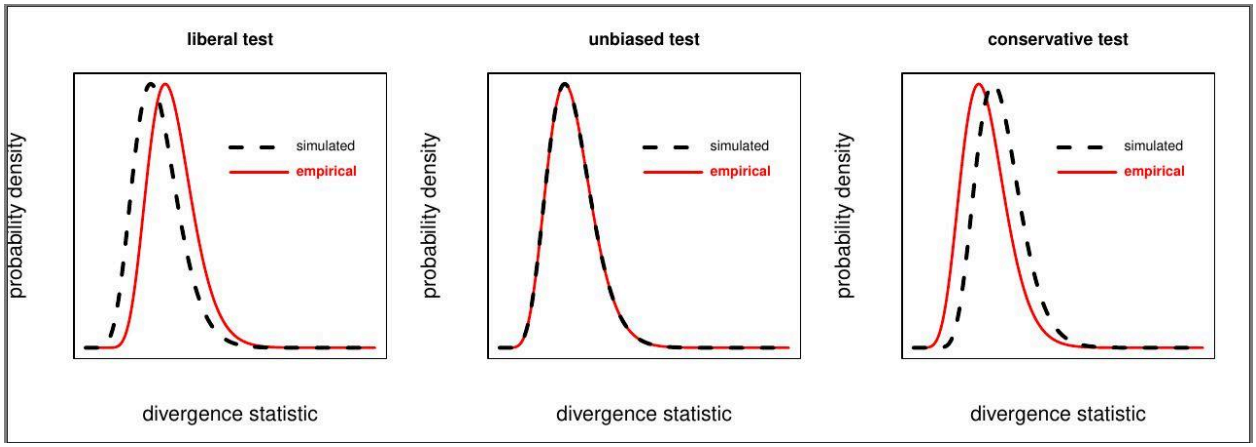


Figure 2: Illustration of the biases in tests for the search of divergence outliers. Filled line: distribution of *empirical* values of the divergence statistic; dashed line: distribution of the *simulated* values of the divergence statistic. Liberal test: the simulated distribution is shifted to the left relative to the empirical distribution; it is easier for empirical values to have extreme values relative to the simulated values they are compared to, and this generates false positives; unbiased test: the simulated distribution is centred on the empirical distribution; this should generate little false negatives or positives; conservative test: the simulated distribution is shifted to the right relative to the empirical distribution; it is more complicated for empirical values to have extreme values relative to the simulated values they are compared to, and this generates false negatives.

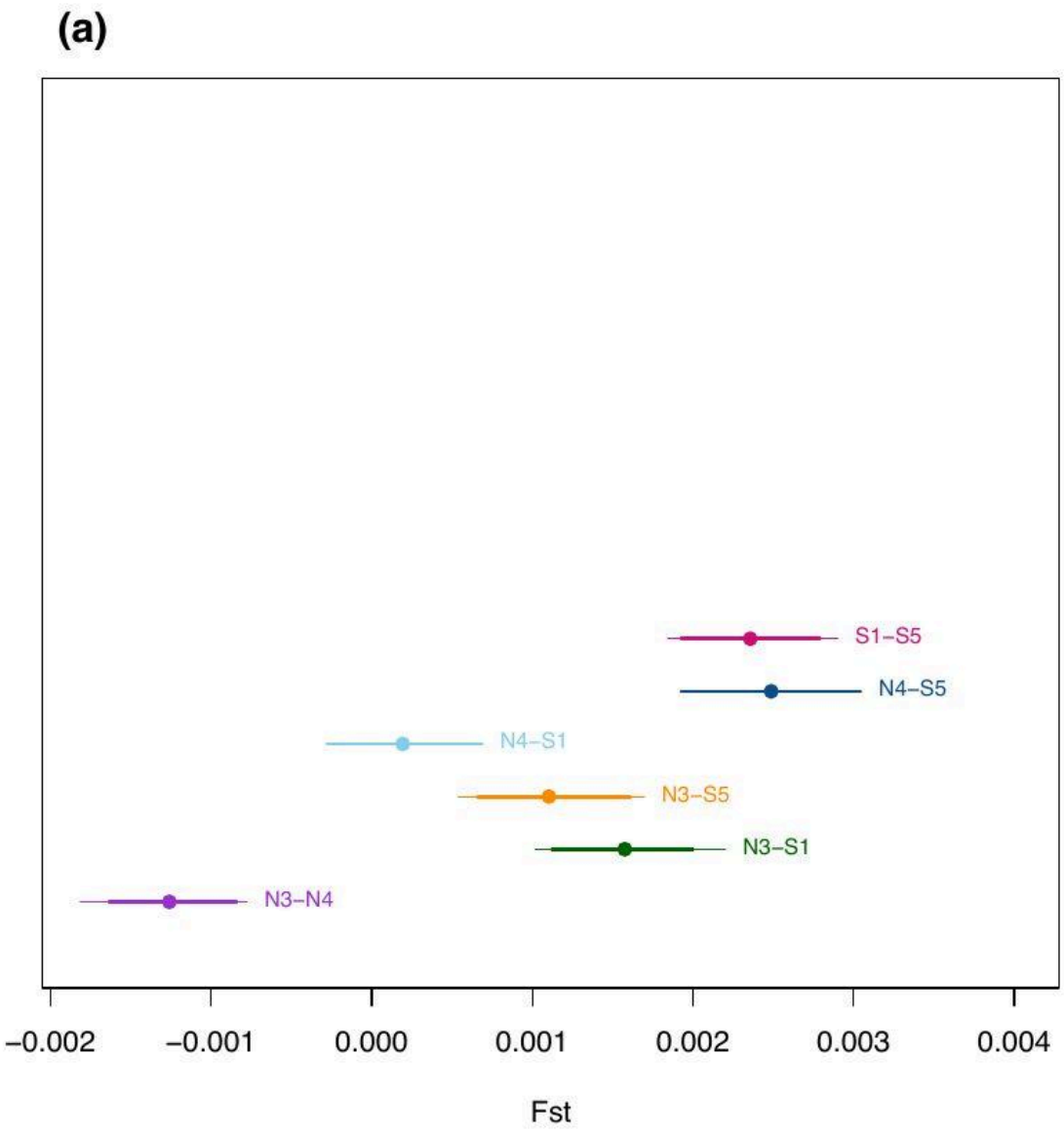
Results

The *de novo* assembly resulted in 11,527 contigs, amounting to 4,407,713 bp. Variant calling produced 40,594 variants in 9,195 genomic regions. After filtering, the dataset included 35,997 SNP in 8,970 genomic regions (3,487,133 bp), with 2.9 % missing genotypes. After mapping the contig sequences onto the beech reference genome, only contigs with a unique position on the reference were retained, resulting in a VCF file containing 34,889 SNPs from 8,791 target regions mapped unequivocally to as many scaffold positions (amounting to 3,429,878 bp, or 0.6 % of total genome length), and 2.8 % missing data. Target regions were not randomly distributed along chromosomes (see Supplementary Results SR1).

After subsetting the dataset into stand pairs, the number of polymorphisms for each stand pair ranged between 30,966 (N3 - S1, 45 individuals) and 32,033 (N4 - S5, 51 individuals) (see Supplementary Table ST2). This reduction of the number of polymorphisms is, of course, caused by the removal of the variants only carried by the samples excluded from each subset.

Diversity and differentiation statistics

We computed diversity statistics for each stand as well as global and pairwise values of differentiation statistics. The values of normalised diversity π, averaged over genomic regions, were 2.089×10^{-3} for N3, 2.121×10^{-3} for N4, 2.122×10^{-3} for S1, and 2.215×10^{-3} for S5. Global F_{ST} was 0.0010 (95% confidence interval (CI): 0.0008-0.0013, median 0.0011). Pairwise F_{ST}-values ranged between -0.0013 and 0.0025, with a mean value of 0.00130 (setting negative values as 0) (median = 0.0013) (**Figure 3a**). The N4-S5 plot pair displayed the largest differentiation (median = 0.0025), and the N3-N4 the smallest (medians -0.0013; the 99% CI for N3-N4 was entirely negative). Pairwise Jost's D values ranged between 0.0060 (N4-S1) and 0.0070 (N3-S1) (mean = 0.0065, median = 0.0066). No confidence interval on Jost's D-values encompassed zero (**Figure 3b**). Most confidence intervals for pairwise divergence overlapped at least with the confidence interval of another population pair, making it difficult to establish a clear dendrogram of population split, although globally, N3 and N4 seem to form a pair, and this pair seems to be more tightly related to S1 than to S5. We used the topology; (S5-(S1-(N3-N4))) as a starting point for the demographic analyses, but left the algorithm free to establish the order of population separation.


(a)
S1–S5
N4–S5
N4–S1
N3–S5
N3–S1
N3–N4
−0.002
−0.001
0.000
0.001
0.002
0.003
0.004
Fst

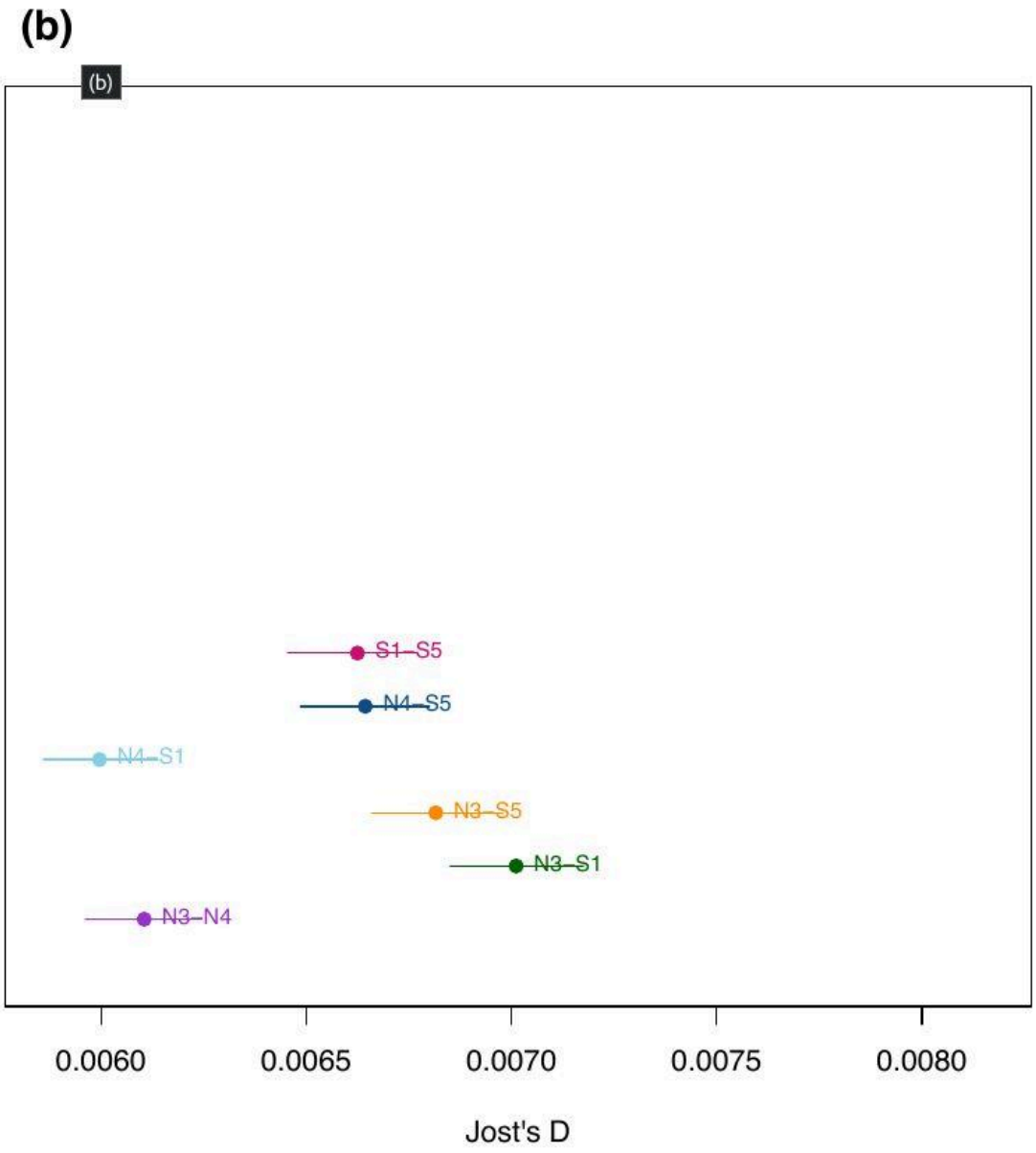


Figure 3: Levels of pairwise molecular divergence for all population pairs, computed as an average of divergence over all SNP. Filled dots indicate the empirical (observed) values, and horizontal bars indicate 95% confidence intervals, as obtained by bootstrap. (a) F_{ST}; (b) Jost's *D*.

Demographic model and simulation of genotypic data

After having computed diversity statistics, we moved to computing the maximum-likelihood estimates of the demographic parameters. The mode and median of estimated current population size (Table 2) were an order of magnitude larger than the upper boundary of the parameter starting range (Supplementary Table ST3). Population sizes were in the order of few tens of thousands for all stands, with the confidence interval on N4 stand size much narrower than the other stands' (Table 2). Splitting of populations was estimated to have occurred not earlier than five generations ago (a timeframe that corresponds to the upper boundaries of the estimates of TIME1, TIME2, and TIME3; Table 2, **Figure 4**~~),~~ . This suggests a very recent expansion from historical relict areas, which started at the earliest twelve generations ago (upper boundary for TIMEA; Table 2, **Figure 4**), with relatively small growth rates of effective (not census) population sizes. In fact, the estimates of GROWTHRATE0 (Table 2, **Figure 4**) were between 0.00010 and 0.00443 , with a mode (0.00366) closer to the upper boundary, thus supporting *expansion*). However, the exact order of population separation remained unresolved, as the intervals for all population split times overlapped. Contraction (possibly due to deforestation in historical times) ended at the earliest 16 generations ago (upper boundary for TIMEB; Table 2, **Figure 4**), and started in the order of one hundred generations ago (TIMEC; Table 2, **Figure 4**); the model correctly returned negative values (i.e. *contraction*) for GROWTHRATE1 (-0.00010/0.00544, Table 2, **Figure 4**). Deep-time population expansion is estimated to have occurred between 100-300 and 1,000 generations ago (TIMED, TIMEE; Table 2, **Figure 4**), but these estimates are not very precise. Growth rates would have been very small, with estimates around zero, and mostly with negative values, suggesting *contraction* instead of expansion of effective population sizes. The distribution of the maximul-likelihood estimates of the demographic parameters are reported in Supplementary Methods 2, Supplementary Figure SM2-2.

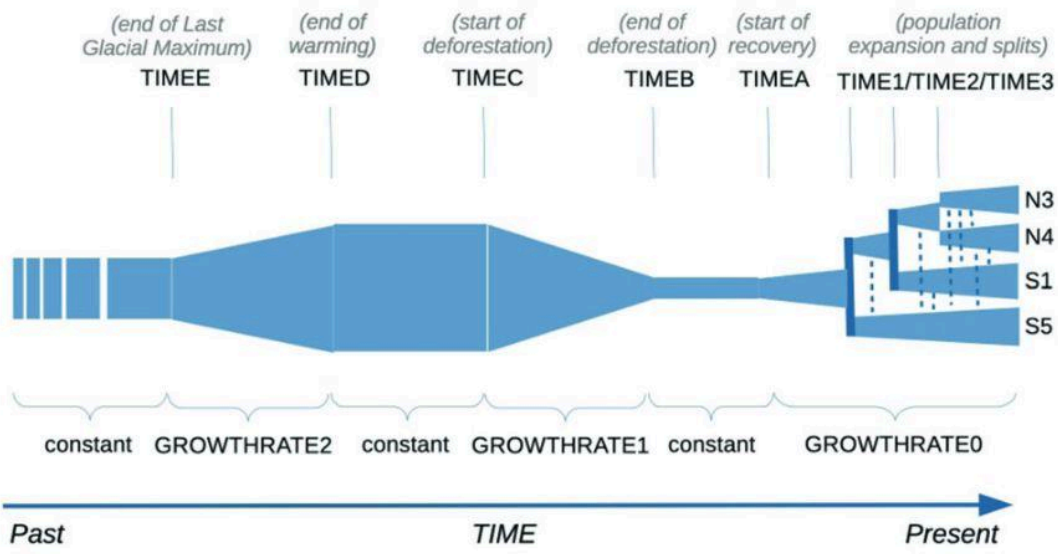


Figure 4 Graphical illustration of the events that were modelled in the background demographic model. The time arrow is depicted horizontally, with present time to the right (arrow at the bottom of the plot). The names of parameters estimated in the demographic inference are shown in bold capitals: TIME parameters are shown above the diagram (italic texts in parentheses describe the geological and historical events corresponding to each time point), and GROWTHRATE (population growth) parameters are shown below the diagram. The filled polygons indicate population sizes (POPSIZE) through time (with thicker vertical width indicating larger population size). Towards the present, the ancestral population splits progressively into the four current (sub)populations that have been studied in the paper (N3, N4, S1, S5). Dotted vertical lines indicate continued gene flow among all population pairs. The broken box at the left end indicates that the population was considered as constant over a long and undefined time in the past.

Polygon sizes, lengths, and shapes are entirely arbitrary and do not to scale at all with either the starting range or the maximum likelihood estimations of the demographic parameters; (b) the topology of population splits is shown only for illustration, as the exact timing of each event was left for the inference algorithm to estimate.

Detection of divergence outlier loci between stand pairs

Based on the demographic background established through the empirical simulation we could move to the identification of the outlier loci in the empirical data. Globally, we detected 27 signatures of selection from 19 non-overlapping genomic regions (Table 3). All the six tests of the G2D method were unbiased (Supplementary Table ST4). Seventeen target regions showed significant divergece in at least one stand pair at the 0.01 FDR; one of them (Bhaga_8.g1884) showed divergence for two stand pairs, and one (Bhaga_7.g1751) for three stand pairs. Population pairs showed variable numbers of outlier target regions (N3-N4: three; N3-S1: two; N3-S5: four; N4-S1: four; N4-S5: five; and S1-S5: one).
Allele frequency difference tests were either unbiased (for the N4-S5 and S1-S5 pairs) or conservative (for the four remaining pairs) (Supplementary Table ST4). Allele frequency differences exceeded the neutral expectation at two loci (Bhaga_1_66575559 and Bhaga_4_14168662) for the S1-S5 pair (See Table 3).

All the Bamova tests were conservative (Supplementary Table ST4). We retrieved five outliers: one in N4-S1 (Bhaga_5_40781880), two in N4-S5 (Bhaga_7_3608194 and Bhaga_7_3608089), and two in S1-S5 (Bhaga_1_66575559 and Bhaga_4_14168662).

The two AFD outliers were retrieved also by Bamova, and several Bamova outliers were inside an outlier target region identified by G2D–(Table 3 and Figure 4), while fifteen out of the seventeen G2D outliers were unique. With the “standard” approach, many more outliers were found with BAMOVA: 65 between populations N3 and N4 ; 120 between S1 and S5; 88 between N3 and S5; 109 for N4 and S5; 79 for N3-S1; and 74 for N4-S1 (Supplementary Result SR1, Supplementary Table 5).

Outlier gene and SNP annotations (gene functions and substitution effects) are reported in Table 3. All outlier-containing genomic regions matched at least one sequence in the transcriptomic or genomic regions analysed in Müller et al. (2017), Pfenninger et al. (2021), and Lesur et al. (2015), but none of our outliers appeared in the outlier lists reported in those studies. Our outliers did not match any of the candidate genes analysed in Postolache et al. (2021), Meger et al. (2021), or Csilléry et al. (2014). Some of the outliers found with the “standard” approach overlapped with those found in the cited studies (Supplementary Result SR1, Supplementary Table 5).

Table 3.

List of outliers. The position on the reference genome, the annotation of the corresponding genomic region, and the stand pairs for which the outlier was identified are reported. When sequence-level (G2D) and/or nucleotide-level (Allele Frequency Difference (AFD), Bamova) outliers belong to the same region, they are displayed on the same row and their stand pair and position are highlighted in bold. All genes were annotated using the gff3 provided by Mishra *et al.* (2002) and GeneBank, unless indicated otherwise.
Gene number = gene number reported in the Mishra et al. (2022) annotation. **chr** = chromosome number (Mishra et al. 2022). **Region** = genomic coordinates of the target sequence in bases pairs. **Gene description** = GenBank accession number of the most similar DNA sequence and summary description of function. **pair** = stand pair for which the locus was an outlier. POS **=** position on the chromosome. **Ann** = functional annotation of SNP : **I** = intron variant; **M** = missense; **S**= synonymous; **mo** = modifier; **md** = moderate; **lo** = low.

Gene number	chr	Region	Gene description	Divergence tests						Ann	Non-demographic outlier
				G2D		AFD		Bamova			
				pair	POS	pair	POS	pair	POS		
Bhaga_1.g6628	1	66574103 - 66580231 (+)	• **XP_023915805.1** • glycerol-3-phosphate dehydrogenase [NAD(+)] GPDHC1, cytosolic • Energy production and conversion			**S1-S5**	**66575559**	**S1-S5**	**66575559**	I,mo	S1-S5
Bhaga_2.g4572	2	41650033 - 41655512 (-)	• **KAB1208090.1** • Glucan endo-1,3-beta-glucosidase 2 • Possibly involved in carbohydrate binding	N4-S1	41651123-41651459						

Bhaga_2.g4634	2	42363076 - 42376634 (-)	• **XP_023922067.1** • tRNA threonylcarbamoyladenosine dehydratase • Activating enzymes (E1) of ubiquitin-like proteins	N3-S5	42370573-42370965						
Bhaga_4.g1667	4	14159632 - 14169373 (+)	• **XP_030964976.1** • (unknown gene name) • probable disease resistance protein RF9			**S1-S5**	**14168662**	**S1-S5**	**14168662**	M,md	S1-S5, N4-S1,
Bhaga_4.g1705	4	14494635 - 14499931 (-)	• **XP_023889455.1** • isoflavone reductase homolog • oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptors	N3-S5	14494838-14495102						
Bhaga_5.g1593	5	14195661 - 14197196 (+)	• **PON33942.1** • (unknown gene name) • NTF2-like domain containing protein	N3-S5	14195973-14196472						
Bhaga_5.g4748	5	40781598 - 40785294 (+)	• **XP_023913931.1** • adenylyltransferase and sulfurtransferase MOCS3	**N4-S1**	**40781565-40781959**			**N4-S1**	**40781880**	S,lo	N4-S1, N3-S1

			• Cofactor biosynthesis: molybdopterin biosynthesis. tRNA modification: 5-methoxycarbonylmethyl-2-thiouridine-tRNA biosynthesis								
Bhaga_6.g2513	6	22741089 - 22752585 (-)	• **RVW32758.1** • Trasposon TX1 uncharacterised 149 kDa protein • Non-LTR retrotransposon and non-LTR retrovirus reverse transcriptase (RT) which catalyzes the conversion of single-stranded RNA into a single-stranded DNA	N4-S5	22743486-22743973						
Bhaga_7	7	-	• **KU700965.1** • putative LOV domain-containing protein [(1)] • sensors for light and oxygen in signal transduction	**N4-S5**	**3608009-3608280**			**N4-S5** **N4-S5**	**3608194** **3608089**		N4-S5 N4-S5
Bhaga_7.g1751	7	14815926 - 14822900 (+)	• **XP_023905227.1** • TIP41-like protein • Regulation of phosphoprotein phosphatase activity and TOR signaling; could play a role in cytoskeleton functions.	N3-N4 N4-S1 N4-S5	14822493-14822895						

Bhaga_7.g4066	7	35377853 - 35383008 (+)	• **XP_030971157.1** • putative PAP-specific phosphatase, mitochondria • catalyses the hydrolysis of 3'-phosphoadenosine-5'-phosphate (PAP) to AMP	N3-S5	35385076-35385440						
Bhaga_8.g97	8	774878 - 784334 (-)	• **XP_030974634.1** • exosome complex exonuclease RRP44 homolog A • rRNA processing and catabolic process; RNA and metal ion binding; 3'-5' RNA exonuclease activity into RNase complex (nucleus and cytosol)	N3-S1	773129-773569						
Bhaga_8.g1884	8	15669815 - 15678133 (-)	• **XP_024017681.1** • CMP-sialic acid transporter • pyrimidine nucleotide-sugar transmembrane and sialic acid transporter activity (Golgi)	N3-N4 N4-S5	15673749-15674000						
Bhaga_9.g1595	9	12879937 - 12884370 (-)	• **KAB1210082.1** • (unknown gene name) • Dynein assembly factor 1, axonemal	N4-S5	12881196-12881598						

Bhaga_9.g2261	9	18401049 - 18401988 (+)	• **KAB1227406.1** • putative vacuolar protein sorting-associated protein 13C	N3-S1	18401752-18402280						
Bhaga_10	10	-	• **XM_024042142.1** • (unknown gene name) • mitochondrial transcription termination factor family isoform 1 [(1)]	N4-S1	10719018-10719314						
Bhaga_10.g2539	10	20777894 - 20784927 (-)	• **VVA22558.1** • (unknown gene name) • uncharacterised B456_004G223300	S1-S5	20781693-20782152						
Bhaga_10.g3505	10	28867540 - 28869330 (-)	• **XP_023906757.1** • (unknown gene name) • NAC domain-containing protein 87-like	N4-S5	28868621-28868946						
Bhaga_11.g539	11	4738139 4743880 (+)	• **KAA3460747.1** • BRO1 domain-containing protein BROX • Metallo-hydrolase/oxidoreductase	N3-N4	4738984-4739355						

[(1)] annotation derived from orthologous genes identified by a BLAST search, followed by a match for the accession on GeneBank.

Discussion

Using a combination of high-throughput genomic data and demo-genetic modelling tools, we found molecular signatures of microgeographic adaptation in a European beech population encompassing the two slopes of a Mediterranean mountain. Although the studied stands have been separated only by a few generations, this genomic signal of divergent selection confirms that population recolonisation and the occupation of novel environments was followed by selective processes strong enough to provoke population divergence. The demographically explicit approach used here, based on a combination of historical records with the results previously obtained by Lander et al. (2011), opens new perspectives for understanding the genomic architecture of microgeographic selection. After controlling for demographic effects, 27 signatures of divergent selection were detected (in 19 out of 8,791 sequenced genomic regions): 20 in G2D, two as allele frequency differences (AFD), and five in BAMOVA (Table 3). A subset of divergence outliers showed a consistent signal over stand pairs and/or methods: out of 19 outlier genomic regions, four appeared with two methods, and two for at least two stand pairs (Table 3, Figure 4). This suggests that signals of divergent selection are strong enough to be picked up by multiple methods and that, for at least some loci, similar selective processes occur within multiple stand pairs, because of shared environmental contrasts (Table 1).

We estimate 0.19 % of the genome (sequence level; G2D method) and 0.01 % of the sequenced genomic polymorphisms (Bamova, AFD) to be under selection. This reflects the findings of Kelly (2022), who found a comparable proportion of genomic loci responsive to temporal selection in a longitudinal study spanning 23 generations in monkeyflower (0.1 %, 994 SNPs the frequency of which fluctuated significantly from positive to negative, harboured in 1,000 genes), but also the findings on several other forest tree populations : Mosca *et al*. (2016) (0.16 % among 768 SNPs from *Pinus cembra* and 1152 SNPs from *Pinus mugo*); Brousseau *et al*., (2020) (0.19 % among 97,062 SNPs in *Eperua falcata* Aubl.); and Scotti et al. (2023) (0.1 - 1.0 % on panels of 7114 - 8462 SNPs in four conifer species: *Abies alba*, *Cedrus atlantica*, *Pinus halepensis*, and *Pinus pinaster*). Microgeographic adaptation apparently involves a small fraction of the genome because, even though Scotti et al. (2023) hypothesised that a fraction of truly adaptive loci went undetected because of the lack of power. Because our tests were often conservative, it is possible that we also missed a fraction of the true divergence outliers.

Population-genetic theory predicts that fewer, larger-effect loci should underlie *local* adaptation relative to *global* adaptation (Yeaman, 2022). Therefore, it is not surprising that we identified few outliers, seemingly undergoing strong selection, even though a compact genetic control should limit adaptation to very divergent local conditions (Cubry *et al.*, 2022). It is worth noticing that large-effect mutations that confer advantage in new environments can actually favour population expansion (Gilbert and Whitlock, 2017); such loci could therefore actually have facilitated recolonisation.

A clear (methodological) message to be received from the small numbers of outliers found here comes from the comparison with the analyses carried out without demographic correction: taking into account demography greatly reduces the number of positive signals in the Bamova

analyses, probably by removing large numbers of false positives. The comparison, however, is imperfect, because the method not controlling for demography (Scotti *et al.*, 2023) uses Bayes Factors (computed from empirical values only) instead of q-values (computed by comparison with simulated values) to identify divergence outliers, and because our tests were often conservative (that is: divergence levels obtained based on the demographic model were on average higher than the empirical divergence levels), producing some amount of false negatives (that is, even high empirical divergence levels fell within the middle range of the simulated values). These limitations notwithstanding, this comparison suggests that incorporating historical information in demographic models, and then using them as a background for the analysis of adaptation, should become standard practice whenever possible. This will allow tests to be performed without relying on the equilibrium hypotheses that often underlie the neutral expectations used to identify outliers. In general, it's always better to rely on diverse methods to detect divergence outliers according to the available knowledge on the studied population: methods resting on different assumptions can capture different aspects of the adaptive process. Whenever linkage disequilibrium is strong between neighboring SNPs, a multilocus method such as G2D should be more fitting to capture the sequence signal. On the other hand, single-locus methods are better suited to capture individual signals scattered across the genome (*ie*. recombination has broken physical association, or they are located far enough from each other). The single-locus methods we used are likely to respond to different facets of the selective process: while raw allele frequency differences (AFD) may become apparent early in the divergence process, F-statistics-based methods rest on ratios of between populations / within-populations variances, which may increase only after the selective process has pushed each population near fixation. We can safely suggest using more than one method when an hypothesis on selection dynamics is lacking.

Our historical demography results are mostly consistent with Lander et al. (2011), both for population split times and for the upper boundaries of population size confidence intervals. Nevertheless, our medians and modes for population sizes are an order of magnitude larger, and the lower boundaries an order of magnitude smaller, so that globally, we tend to obtain larger estimates than Lander et al. (2011). The duration of the period over which population size would have stayed at a minimum is in the order of 90-99 generations, placing it approximately within historical times. A slight difference with Lander et al. (2011) is that the lower boundary of our growth rate estimates also includes contraction. However, the timing of the most ancient events (post-glacial expansion) is poorly estimated. Interestingly though, Milesi et al. (2023) detect very stable effective population sizes for beech over the last hundreds of thousands of years, at the continental level, somehow pushing expansion events far back in the past, supporting our results. While our approach is likely to carefully control divergence in allele frequencies stemming from purely neutral, demographic processes (i.e., allele surfing), we cannot rule out that the demographic parameters estimated here can be biased by the presence of undetected background selection (Charlesworth, 1993), since strongly deleterious, low-frequency alleles could have been interpreted as signatures of expansion (Ewing and Jensen, 2016; Johri et al., 2020, 2022). This happens because intermediately deleterious mutations accumulate in a non-neutral fashion in the "rare alleles" classes of an SFS in the context of non-equilibrium processes. During demographic expansions novel mutations are

generated, with the risk of inflating the "rare alleles" classes of the SFS. *F sylvatica* genome size (542 Mb, Mishra *et al.,* 2018) and mutation rate (9.5×10^{-10}, Milesi et al, 2023) allow us to infer that each gamete carries ≈ 0.5 new mutations. We can expect ≈ 5000 novel mutation having occurred in the population of Mont Ventoux over the last 150 years given that its estimated effective population size N_e is the number of gametes that constitute the next generation (5000), the number of elapsed generation since the reforestation process (2), and the number of novel mutations for each gamete (0.5). In addition to this, the use of SFS-derived statistics to infer demography has been criticised because of its high sensitivity to fluctuations in SFS (Myers *et al.*, 2008; Lapierre *et al.*, 2017). Another limitation of our work is that we mostly focused on allele frequency shifts rather than changes in allele frequency covariances, which are known to be more involved in the early stages of adaptation (Le Corre and Kremer, 2003, 2012; Cubry *et al.*, 2022). This may explain the low number of outliers detected here.

The second (evolutionary) take-home message can be derived from the combination of the detection of divergence outliers and the estimation of population divergence times: the different stands have become separated only a few generations ago, yet signatures of the effect of divergent selection appear in the data. This suggests that population recolonisation and the occupation of novel environments was followed by selective processes strong enough to provoke population divergence. First, we showed that separate parts of the historical relict population (pairs N3-S5 and N4-S5) show signatures of adaptive molecular divergence. This suggests that geographical separation and environmental divergence can establish adaptation-driven population divergence even if the populations are quite close geographically (here, S5 suffers less summer drought stress than N3-N4, but undergoes more intense frost; Table 1). Second, newly colonised areas (stand S1) showed signatures of adaptive divergence. As S1 was generated by recolonisation over the last one-and-a-half century at most, this suggests that adaptive processes can be very rapid, likely driven by the very different environmental conditions experienced by S1 relative to the three historical relict populations. Third, and quite surprisingly, sub-populations within the same part of the historical relict zone (N3 and N4), but experiencing slightly different environmental conditions, also showed adaptive divergence, despite the conservativeness of the tests. So, stable sub-populations, undergoing environmental contrasts, putatively undergo divergence even when they are geographically contiguous.

We expected that the loci that diverge between stand par among an environmental gradient are associated with phenotypic traits that usually vary in these landscape contexts, like budburst (Gaüzere et al., 2020) or response to drought (Bontemps et al., 2017). The only gene with a possible relationship to phenology is a putative LOV domain-containing protein (Ito et al., 2012; Glantz et al., 2016), flavoproteins that modulate the photosensitivity. We can also hypothesise that the identified isoflavone reductase homolog, (Cheng et al., 2015) may be involved in this pathway. We also noticed that four genes have been studied in the context of plant defense from different kind of pathogens (endo-1,3-beta-glucosidase, Sperisen et al., (1991) and Kebede and Kebede (2021); probable disease resistance protein RF9, Milc et al., (2019); isoflavone reductase homolog, Cheng et al., (2015)) and two are invovled in the response to oxidative (GPDHC1; Shen et al., (2006)) and osmotic stress (NAC domain-containing protein; Duval et al., (2002).

A third (evolutionary) message is that novel adaptation was fuelled by the genetic variation already present in the historical relict zone, through changes in allele frequencies between the source and the derived stands. This is an indication that divergent selection can be detected over both short *distances* (as already reported in Audigeos *et al.*, 2013; Lind *et al.*, 2017; Brousseau *et al.*, 2020; Scotti *et al.*, 2023; Budde *et al.*, 2023) and short *times*, and that populations can diverge quickly from standing variation. The same stands were shown to also be genetically differentiated at adaptive traits (Gaüzere *et al.*, 2020), which reasonably can only have diverged from the same genetic background and over the same time span as genomic loci.

Conclusions

The adaptive processes ongoing in the European beech population of Mont Ventoux have been the object of several studies (Lalagüe *et al.*, 2013; Csilléry *et al.*, 2014; Gaüzere *et al.*, 2020) relying on candidate gene and phenotype data. Here we take a genome-wide perspective (even though restricted to exons) to study adaptive processes. We integrated genomic data with a reconstruction of the forest demographic history building on the work of (Lander *et al.*, 2011, 2021) and exploited the recently assembled genome of *Fagus sylvatica* from (Mishra *et al.*, 2022) to map our polymorphisms and retrieve functional annotations.

We were able to retrieve divergence signals, both at the single-SNP and the gene level, possibly caused by local environmental pressures, which can be detected even on very small spatial scales (~150 m) but also between newly expanded and old-growth forests. This indicates that, while overall genome-wide divergence is as small as expected for such short geographical distances, selection can target portions of the genome and cause divergence, as already shown for phenological characters. Such observations suggest that within-population genetic diversity has been a powerful reservoir of adaptive potential, which allowed expanding cohorts to adapt to new conditions over very few generations. Such genetic properties may contribute to help forests adjust to climate upheavals, and are a potential ally for establishing forest management practices favouring adaptation.

Acknowledgements

The Authors wish to thank Rasmus Nielsen and Josephine Paris for stimulating discussions and suggestions while at the SIBE 2022 conference in Ancona, Italy, Élodie Vercken for commenting our results, Katharina Budde for a very deep and critical reading of the manuscript, as well as three anonymous reviewers for very useful criticisms. This work was funded by the INRAE-funded CLIMAE MODEGRAD research program, as well as by the ANR FLAG grant (ANR-12-ADAP-0007-01). A. Modica's Ph.D. program is supported by a grant co-funded by the INRAE ECODIV Department and the European Commission's Horizon 2020-funded project FORGENIUS (EC grant No. 862221).

Author contribution statement

IS, SM conceived the experiment ; IS, HL produced the data ; AM, SM, HL, IS analysed the data ; AM, SM, IS interpreted the results ; AM, SM, HL, IS wrote and edited the manuscript.

Conflict of interest

The Authors declare no conflict of interest.

Data archiving

Data and scripts used to analyse them are available at

https://doi.org/10.57745/BE4HIG

Research ethic statement

All permissions required at the time of sampling were obtained by the Authors.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Audigeos D, Brousseau L, Traissac S, Scotti-Saintagne C, Scotti I (2013). Molecular divergence in tropical tree populations occupying environmental mosaics. *J Evol Biol* 26: 529–544.

Bank C, Ewing GB, Ferrer-Admettla A, Foll M, Jensen JD (2014). Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends Genet* **30**: 540–546.

Barbero M, Du Merle P, Guende G, Quézel P (1978). La végétation du Mont Ventoux. *Terre Vie Supplément* **32**: 21–37.

Beichman AC, Huerta-Sanchez E, Lohmueller KE (2018). Using Genomic Data to Infer Historic Population Dynamics of Nonmodel Organisms. *Annu Rev Ecol Evol Syst* **49**: annurev-ecolsys-110617-062431.

Bontemps A, Klein EK, Oddou-Muratorio S (2013). Shift of spatial patterns during early recruitment in Fagus sylvatica: Evidence from seed dispersal estimates based on genotypic data. *For Ecol Manag* **305**: 67–76.

Bontemps A, Davi H, Lefèvre F, Rozenberg P, Oddou-Muratorio S (2017). How do functional traits syndromes covary with growth and reproductive performance in a water-stressed population of Fagus sylvatica? *Oikos* **126**: 1472–1483.

Brousseau L, Fine PVA, Dreyer E, Vendramin GG, Scotti I (2020). Genomic and phenotypic divergence unveil microgeographic adaptation in the Amazonian hyperdominant tree Eperua falcata Aubl. (Fabaceae). *Mol Ecol*: 1–19.

Budde KB, Rellstab C, Heuertz M, Gugerli F, Hanika T, Verdú M, *et al.* (2024). Divergent selection in a Mediterranean pine on local spatial scales. *J Ecol* **112**: 278–290.

Charlesworth, D, Charlesworth, B (1987) Inbreeding depression and its evolutionary consequences. *Annual review of ecology and systematics.* **18**: 237-268.

Charlesworth B, Morgan MT, Charlesworth D (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.

Cheng Q, Li N, Dong L, Zhang D, Fan S, Jiang L, *et al.* (2015). Overexpression of Soybean Isoflavone Reductase (GmIFR) Enhances Resistance to Phytophthora sojae in Soybean. *Front Plant Sci* **6**.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, *et al.* (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**: 80–92.

Colautti RI, Barrett SCH (2013). Rapid Adaptation to Climate Facilitates Range Expansion of an Invasive Plant. *Science* **342**: 364–366.

Compeau PEC, Pevzner PA, Tesler G (2011). How to apply de Bruijn graphs to genome assembly. *Nature Biotech* **29**: 987-991.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**: 1411–1423.

Csilléry K, Lalagüe H, Vendramin GG, González-Martínez SC, Fady B, Oddou-Muratorio S (2014). Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (Fagus sylvatica) populations. *Mol Ecol* **23**: 4696–4708.

Cubry P, Oddou-Muratorio S, Scotti I, Lefèvre F (2022). Interactions between microenvironment, selection and genetic architecture drive multiscale adaptation in a simulation experiment. *J Evol Biol* **35**: 451–466.

Davi, H, Cailleret M (2017) Assessing drought-driven mortality trees with physiological process-based models. *Agricultural and Forest Meteorology* **232**: 279-290.

De La Torre AR, Wilhite B, Neale DB, Slotte T (2019). Environmental Genome-Wide Association Reveals Climate Adaptation Is Shaped by Subtle to Moderate Allele Frequency Shifts in Loblolly Pine. *Genome Biol Evol* **11**: 2976–2989.

De Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE (2014). Genome scan methods against more complex models: When and how much should we trust them? *Mol Ecol* **23**: 2006–2019.Duval M, Hsieh T-F, Kim SY, Thomas TL (2002). Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. *Plant Mol Biol*: 237–248.

Ewing GB, Jensen JD (2016). The consequences of not accounting for background selection in demographic inference. *Mol Ecol* **25**: 135–141.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013). Robust Demographic Inference from Genomic and SNP Data (JM Akey, Ed.). *PLoS Genet* **9**: e1003905.

Excoffier L, Foll M (2011). fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**: 1332–1334.

Excoffier L, Foll M, Petit RJ (2009). Genetic consequences of range expansions. *Annu Rev Ecol Evol Syst* **40**: 481–501.

Excoffier L, Ray N (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol* **23**: 347–351.

Foll M, Beaumont MA, Gaggiotti O (2008). An approximate Bayesian computation approach to overcome biases that arise when using Amplified Fragment Length Polymorphism

markers to study population structure. *Genetics* **179**: 927–939.

Gamba E, Aurelle S, Vergnaud C, Talon B (2023). A Holocene history of forest vegetation on the northern slopes of a Mediterranean mountain : the Mont Ventoux, Southeast France. *Ecologia Mediterranea*, 48(2):53-68.

Gautier M (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* **201**: 1555–1579.

Gaüzere J, Klein EK, Brendel O, Davi H, Oddou-Muratorio S (2020). Microgeographic adaptation and the effect of pollen flow on the adaptive potential of a temperate tree species. *New Phytol* **227**: 641–653.

Gilbert KJ, Whitlock MC (2017). The genetics of adaptation to discrete heterogeneous environments: frequent mutation or large-effect alleles can allow range expansion. *J Evol Biol* **30**: 591–602.

Glantz ST, Carpenter EJ, Melkonian M, Gardner KH, Boyden ES, Wong GK-S, *et al.* (2016). Functional and topological diversity of LOV domain photoreceptors. *Proc Natl Acad Sci* **113**.

Godineau C, Fririon V, Beudez N, De Coligny F, Courbet F, Ligot G, *et al.* (2023). A demo-genetic model shows how silviculture reduces natural density-dependent selection in tree populations. *Evol Appl* **16**: 1830–1844.

Gompert Z, Buerkle CA (2011). A hierarchical Bayesian model for next-generation population genomics. *Genetics* **187**: 903–17.

Goudet J (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes* **5**: 184–186.

Hansen MM, Olivieri I, Waller DM, Nielsen EE, The GeM Working Group (2012). Monitoring adaptive genetic responses to environmental change. *Mol Ecol* **21**: 1311–1329.

Hedrick, P W. *Genetics of populations*. Sudbury, MS: Jones and Bartlett, 2000.

Ito S, Song YH, Imaizumi T (2012). LOV Domain-Containing F-Box Proteins: Light-Dependent Protein Degradation Modules in Arabidopsis. *Mol Plant* **5**: 573–582.

Johri P, Aquadro CF, Beaumont M, Charlesworth B, Excoffier L, Eyre-Walker A, *et al.* (2022). Recommendations for improving statistical inference in population genomics. *PLOS Biol* **20**: e3001669.

Johri P, Charlesworth B, Jensen JD (2020). Toward an Evolutionarily Appropriate Null Model: Jointly Inferring Demography and Purifying Selection. *Genetics* **215**: 173–192.

Jost L (2007). Partitioning diversity into independent alpha and beta components. *Ecology* **88**: 2427–2439.

Jost L (2008). G ST and its relatives do not measure differentiation. *Mol Ecol* **17**: 4015–4026.

Kaplan JO, Krumhardt KM, Zimmermann N (2009). The prehistoric and preindustrial deforestation of Europe. *Quat Sci Rev* **28**: 3016–3034.

Kebede A, Kebede M (2021). In silico analysis of promoter region and regulatory elements of glucan endo-1,3-beta-glucosidase encoding genes in Solanum tuberosum: cultivar DM 1-3 516 R44. *J Genet Eng Biotechnol* **19**: 145.

Kelly JK (2022). The genomic scale of fluctuating selection in a natural plant population. *Evol Lett* **6**: 506–521.

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, *et al.* (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**: 2283–5.

Lalagüe H, Csilléry K, Oddou-Muratorio S, Safrana J, de Quattro C, Fady B, *et al.* (2013). Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (Fagus sylvatica L.) population from southeastern France. *Tree Genet Genomes* **10**: 15–26.

Lander TA, Klein EK, Roig A, Oddou-Muratorio S (2021). Weak founder effects but significant spatial genetic imprint of recent contraction and expansion of European beech populations. *Heredity* **126**: 491–504.

Lander TA, Oddou-Muratorio S, Prouillet-Leplat H, Klein EK (2011). Reconstruction of a beech population bottleneck using archival demographic information and Bayesian analysis of genetic data. *Mol Ecol* **20**: 5182–5196.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Lapierre M, Lambert A, Achaz G (2017). Accuracy of Demographic Inferences from the Site Frequency Spectrum: The Case of the Yoruba Population. *Genetics* **206**: 439–449.

Le Corre V, Kremer A (2003). Genetic Variability at Neutral Markers, Quantitative Trait Loci and Trait in a Subdivided Population Under Selection. *Genetics* **164**: 1205–1219.

Le Corre V, Kremer A (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Mol Ecol* **21**: 1548–1566.

Lefèvre F, Gallais A (2020). Partitioning heterozygosity in subdivided populations: Some misuses of Nei's decomposition and an alternative probabilistic approach. *Mol Ecol* **29**: 2957–2962.

Leroy T, Louvet JM, Lalanne C, Le Provost G, Labadie K, Aury JM, *et al.* (2020). Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New*

Phytol **226**: 1171–1182.

Lesur I, Bechade A, Lalanne C, Klopp C, Noirot C, Leplé J-C, *et al.* (2015). A unigene set for European beech (*Fagus sylvatica* L.) and its use to decipher the molecular mechanisms involved in dormancy regulation. *Mol Ecol Resour* **15**: 1192–1204.

Lewontin RC, Krakauer J (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**: 175–195.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Lind BM, Friedline CJ, Wegrzyn JL, Maloney PE, Vogler DR, Neale DB, *et al.* (2017). Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Englm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Mol Ecol* **26**: 3168–3185.

Lind BM, Menon M, Bolte CE, Faske TM, Eckert AJ (2018). The genomics of local adaptation in trees: are we out of the woods yet? *Tree Genet Genomes* **14**: 29.

Liu S, Zhang L, Sang Y, Lai Q, Zhang X, Jia C, *et al.* (2022). Demographic History and Natural Selection Shape Patterns of Deleterious Mutation Load and Barriers to Introgression across *Populus* Genome (Y Kim, Ed.). *Mol Biol Evol* **39**: msac008.

Lotterhos KE, Whitlock MC (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol* **23**: 2178–92.

Lu M, Loopstra CA, Krutovsky KV (2019). Detecting the genetic basis of local adaptation in loblolly pine (*Pinus taeda* L.) using whole exome‐wide genotyping and an integrative landscape genomics analysis approach. *Ecol Evol* **9**: 6798–6809.

Martin-Gousset L, Terracol J, Roux J-P (2019). Habitats et dynamiques de végétation en ubac du Mont-Ventoux – Étude de la Réserve biologique intégrale. *Naturae* **2019**: 113–164.

Médail F, Diadema K (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *J Biogeogr* **36**: 1333–1345.

Meger J, Ulaszewski B, Burczyk J (2021) Genomic signatures of natural selection at phenology-related genes in a widely distributed tree species *Fagus sylvatica* L. *BMC Genomics* **22**: 1-20.

Milc J, Bagnaresi P, Aragona M, Valente MT, Biselli C, Infantino A, *et al.* (2019). Comparative transcriptome profiling of the response to Pyrenochaeta lycopersici in resistant tomato cultivar Mogeor and its background genotype—susceptible Moneymaker. *Funct Integr Genomics* **19**: 811–826.

Milesi P, Kastally C, Dauphin B, Cervantes S, Bagnoli F, Budde KB, *et al.* (2023). Synchronous effective population size changes and genetic stability of forest trees through glacial

cycles. *bioRxiv*: 2023.01.05.522822.

Mishra B, Gupta DK, Pfenninger M, Hickler T, Langer E, Nam B, *et al.* (2018). A reference genome of the European beech (Fagus sylvatica L.). *GigaScience* **7**.

Mishra B, Ulaszewski B, Meger J, Aury J-M, Bodénès C, Lesur-Kupin I, *et al.* (2022). A Chromosome-Level Genome Assembly of the European Beech (Fagus sylvatica) Reveals Anomalies for Organelle DNA Integration, Repeat Content and Distribution of SNPs. *Front Genet* **12**: 691058.

Mosca E, Gugerli F, Eckert AJ, Neale DB (2016). Signatures of natural selection on *Pinus cembra* and *P. mugo* along elevational gradients in the Alps. *Tree Genet Genomes* **12**: 9.

Müller M, Seifert S, Lübbe T, Leuschner C, Finkeldey R (2017). De novo transcriptome assembly and analysis of differential gene expression in response to drought in European beech (Z-H Chen, Ed.). *PLoS ONE* **12**: e0184167.

Myers S, Fefferman C, Patterson N (2008). Can one learn history from the allelic spectrum? *Theor Popul Biol* **73**: 342–348.

Nielsen R (2005). Molecular Signatures of Natural Selection. *Annu Rev Genet* **39**: 197–218.

Nielsen R, Hubisz MJ, Clark AG (2004). Reconstituting the Frequency Spectrum of Ascertained Single-Nucleotide Polymorphism Data. *Genetics* **168**: 2373–2382.

Nielsen R, Hubisz MJ, Hellmann I, Torgerson D, Andrés AM, Albrechtsen A, *et al.* (2009). Darwinian and demographic forces affecting human protein coding genes. *Genome Res* **19**: 838–849.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005). Genomic scans for selective sweeps using SNP data. *Genome Res* **15**: 1566–1575.

Nilsson SG (1985). Ecological and Evolutionary Interactions between Reproduction of Beech Fagus silvatica and Seed Eating Animals. *Oikos* **44**: 157.

Ningre J (2007). Le reboisement du Ventoux au XIX siècle : techniques et travaux. *For Méditerranéenne* **XXVIII**: 319–326.

Peñalba JV, Smith LL, Tonione M a, Sass C, Hykin SM, Skipwith PL, *et al.* (2014). Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Mol Ecol Resour* **14**: 1000–1010.

Petit RJ, Aguinagalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, *et al.* (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.

Petit RJ, Hampe A (2006). Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics* **37**: 187-214.

Pfenninger M, Reuss F, Klebler A, Schönnenbeck P, Caliendo C, Gerber S, *et al.* (2021).

Genomic basis for drought resistance in European beech forests threatened by climate change. *eLife* **10**.

Postolache D, Oddou-Muratorio S, Vajana E, Bagnoli F, Guichoux E, Hampe A, *et al.* (2021). Genetic signatures of divergent selection in European beech (Fagus sylvatica L.) are associated with the variation in temperature and precipitation across its distribution range. *Mol Ecol* **30**: 5029–5047.

R Core Team (2023). R: A Language and Environment for Statistical Computing.

Schliep KP (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**: 592–593.

Scotti I, Lalagüe H, Oddou-Muratorio S, Scotti-Saintagne C, Ruiz Daniels R, Grivet D, *et al.* (2023). Common microgeographical selection patterns revealed in four European conifers. *Mol Ecol* **32**: 393–411.

Shen W, Wei Y, Dauk M, Tan Y, Taylor DC, Selvaraj G, *et al.* (2006). Involvement of a Glycerol-3-Phosphate Dehydrogenase in Modulating the NADH/NAD+ Ratio Provides Evidence of a Mitochondrial Glycerol-3-Phosphate Shuttle in *Arabidopsis*. *Plant Cell* **18**: 422–441.

Sperisen C, Ryals J, Meins F (1991). Comparison of cloned genes provides evidence for intergenomic exchange of DNA in the evolution of a tobacco glucan endo-1,3-beta-glucosidase gene family. *Proc Natl Acad Sci* **88**: 1820–1824.

Storey JD, Bass A, Dabney A, Robinson D (2023). qvalue: Q-value estimation for false discovery rate control. R package version 2.32.0.

Vornam B, Decarli N, Gailing O (2004). Spatial Distribution of Genetic Variation in a Natural Beech Stand (Fagus sylvaticaL.) Based on Microsatellite Markers. *Conserv Genet* **5**: 561–570.

Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Wells N, Goddard S, Hayes MJ (2004). A Self-Calibrating Palmer Drought Severity Index. *J Clim* **17**: 2335–2351.

Yeaman S (2022). Evolution of polygenic traits under global vs local adaptation. *Genetics* **220**: iyab134.

Zanon M, Davis BAS, Marquer L, Brewer S, Kaplan JO (2018). European Forest Cover During the Past 12,000 Years: A Palynological Reconstruction Based on Modern Analogs and Remote Sensing. *Front Plant Sci* **9**: 253.

Zerbino DR (2010) Using the Velvet de novo assembler for short-read sequencing technologies. *Current protocols in bioinformatics* Chapter 11, Unit 11.5.

Zhang Z, Schwartz S, Wagner L, Miller W (2000). A greedy algorithm for aligning DNA

sequences. *J Comput Biol* **7**: 203–214.

Zhao W, Sun Y-Q, Pan J, Sullivan AR, Arnold ML, Mao J-F, *et al.* (2020). Effects of landscapes and range expansion on population structure and local adaptation. *New Phytol* **228**: 330–343.

Chapter 2:

Genotype-per-Phenotype Association reveals polygenic architecture for spring phenology traits of European beech

Andrea Modica[1], Sylvie Oddou-Muratorio[2], Katrin Heer[3], Ivan Scotti[1]

[1]INRAE-URFM, Avignon, France,

[2]EcoBioP, Saint-Pée-sur-Nivelle, France

[3]Forest Genetics, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany

Summary

In forest trees, budburst dynamic is determined by the interplay of the individual genotype and its plastic response to environmental factors. Within the same population, individual trees can break the buds with different timings. European beech seedlings from higher elevation grown in common gardens have been observed to flush leaves earlier than those from lower altitudes, suggesting that genetic and environmental factors contrast each other. Our goal is to identify the Quantitative Trait Loci controlling beech spring phenology. A whole genome resequencing experiment on 162 individual trees provided a panel of ~7 million SNPs. To estimate the dates of passage in incomplete time series of phenological scores recorded in 2009 and 2010, we modelled individual time series as a logistic function of the day of the year and estimated the dates of the phenological transition. We also computed the degree days and the phenological score sum, and estimated synthetic parameters of budburst dynamics from the equation. The traits were analysed by a Genotype-Phenotype Association with a replicate training/testing approach to account for the sample size. The traits measured in the two years were associated with different sets of QTLs, whilst several traits shared elements of their architecture. More than half of the QTLs were mapped outside coding regions and another fraction within introns. The distribution of allelic effects for several traits conformed to the expectations of a complex architecture.

Introduction

The timing of bud-break (TBB) is a highly heritable trait (Howe *et al.*, 2000), the adaptive value of which results from an optimal tradeoff between a longer growing season and late frost avoidance (Gömöry and Paule, 2011). By determining for how long trees will be able to fix carbon, consequently TBB affects the functioning and productivity of an ecosystem and its carbon balance (Rathcke and Lacey, 1985; Kramer *et al.*, 2000; Vitasse *et al.*, 2009; Davi *et al.*, 2011) and individual fitness - because of the synchronicity between leaf and flower phenology (Gömöry and Paule, 2011; Gaüzere *et al.*, 2013). Spring phenology is tuned to local climate, keeps pace with the ongoing season, and it is regulated by multiple genes (Nicotra *et al.*, 2010; De Villemereuil *et al.*, 2016; Rellstab *et al.*, 2017); The onset of bud burst depends on the interplay of chilling and forcing temperatures, that enables an individual tree to respond to a longer photoperiod (Heide, 1993; Laube *et al.*, 2014) by triggering the genes controlling bud opening (Caffarra and Donnelly, 2011; Gaüzere *et al.*, 2017). Environment hence plays a strong role in determining TBB of a population in years with different climatic conditions, both in terms of phenotypic plasticity (Badeck *et al.*, 2004; Arnold *et al.*, 2019), and as a long-term evolutionary response (Kawecki and Ebert, 2004; Richardson *et al.*, 2014). This makes it hard to tell apart the roles of plastic environmental contribution and the individual breeding values in determining TBB; this is best done in common-garden conditions (De Villemereuil *et al.*, 2016; Gaüzere *et al.*, 2020) and highlights the genetic factors underpinning divergence for TBB. Each individual thus expresses the trait in the same conditions, easing the investigations of genetic bases of a highly pleiotropic trait such as phenology (Gienapp *et al.*, 2017; Johnson *et al.*, 2022). The shared environment of common gardens evens out the bulk of random environmental noise, but it may introduce a systematic error on the trait expression by imposing constrained conditions that may be not met *in silva*. These very homogeneous conditions constrain the spectrum of phenotypic diversity, unlike what happens in a natural setting. In example, Lazic *et al*. (2024) identified an allelic variant associated with winter temperature that retards TBB in *Fagus sylvatica* L. in natural populations, but does not show a fitness effect neither in common garden conditions nor in reciprocal transplant.

While phenology in natural populations is affected by environmental noise, it also comes in a wider range of variation because traits that share an architecture with the phenology may not be expressed at full in a constrained environment. In addition to this, local conditions, in combination with (random or deterministic) variations in allele frequencies and the polygenic bases of trait control cause selective pressures to act on different loci in different populations (Pritchard and Di Rienzo, 2010; Savolainen *et al.*, 2013), leading to population-specific genetic architectures of traits. Thus, studying the genetic bases of complex, adaptive traits directly in focal wild populations is a reasonable option, notwithstanding the presence of uncontrolled variables affecting the phenotype. *Fagus sylvatica* L. (European beech) dominates forest landscapes in central and northern Europe over a wide range of climatic and edaphic conditions (Magri, 2008; Hickler *et al.*, 2012; Hanewinkel *et al.*, 2013). Water availability constraints its distribution (Pluess and Weber, 2012; Choat *et al.*, 2018; Pfenninger *et al.*, 2021) and global-change induced droughts are predicted to shift its range northward and toward higher altitudes (Capblancq *et al.*, 2020). Over the last decade, European beech has been subjected to a thorough investigation of its adaptive potential at different spatial scales. Phenology-related genes have been been identified with genomic approaches at a continental (Müller *et al.*, 2015, 2017; Pluess *et al.*, 2016; Krajmerová *et al.*, 2017; Frank *et al.*, 2017; Meger *et al.*, 2021; Postolache *et al.*, 2021; Lazic *et al.*, 2024),

regional (Müller *et al.*, 2015, 2017; Pluess *et al.*, 2016; Frank *et al.*, 2017; Meger *et al.*, 2021) and local scale (Csilléry and Lalagüe *et al.*, 2014). The publication of a European beech genome reference (Mishra *et al.*, 2018, 2022) has paved the path to a genome-wide investigation of the loci underpinning spring phenology. The elective tool for the identification of the genomic architecture of a trait are Genotype-to-Phenotype Association studies (more commonly named GWAS, genome-wide association studies, and so called from now on). They require the measurement of the investigated trait on a large number of individuals, that must be as unrelated as possible, so that genomic blocks of linkage are as small as it is possible, and be genotyped at enough loci to represent as many linkage blocks as possible (González-Martínez *et al.*, 2006). Whilst whole-genome resequencing enables a complete coverage of the beech genome, the cost of resequencing is roughly proportional to sample size, which is therefore the limiting factor when studying wild populations: too few individuals can result in either lower phenotypic variation, hence less predictive power, or in an overestimation of the effect of the alleles on the trait, if the phenotypic variation is too large relative to genetic variation in the sample size (Xu, 2003).

Here, we tap into a whole-genome resequencing panel and in a two-year time series of phenology observations to

(i) improve the accuracy and completeness of field-recorded BBCH scores during the vegetative phenological transition and better capture and quantify processes linked both to the timing and the pace of TBB;

(ii) and infer which genes participate in the architecture of TBB and associated phenological traits, which is their effect on the phenotype, and whether the trait(s) exhibit a polygenic architecture.

Material and Methods

Species, study site, and sampling

European beech (*Fagus sylvatica* L.) is a monoecious, shade tolerant forest tree species (Gaüzere *et al.*, 2013). It is predominantly outcrossed; pollen dispersal occurs by wind. Seed are produced in mast years and are spread either by gravity or small animals (Jensen, 1985; Piotti *et al.*, 2012; Oddou-Muratorio *et al.*, 2018), with a higher degree of long-distance dispersal events for pollen than seeds (Bontemps *et al.*, 2013; Gaüzere *et al.*, 2013). Leaf development occurs from March to May and is paired with flowering (April to May) (Bontemps *et al.*, 2017; Oddou-Muratorio *et al.*, 2018). We studied the stand N1 (1.3 ha, 1,020 m a.s.l.) on the northern slope of Mont Ventoux, in southern France, at the dry edge of the species distribution range. There, European beech grows within a continuous mixed forest with *Abies alba* Mill., and *Quercus*. spp. On the northern slope this forest spans uninterrupted from 900 to 1500 m.s.l.m. (Oddou-Muratorio *et al.*, 2024).

Genetic data

The BEECHGENOMEs consortium dataset had been obtained from Illumina short-read sequences of 150 bpwith an expected coverage of 20x per individual, and then treated with the BeechMarkET pipeline as reported in Mishra et al. (2022). The BCF contained 7,793,706 variants genotyped in a panel of 362 individuals, of which 166 from Mont Ventoux, France (44°21′N, 5°27′E). We first converted the BCF file using BCFtools v. 1.10.2 (Li, 2011; Danecek *et al.*, 2021) in VCF.bgz (command `convert -O z`); following that, and after each subsequent manipulation of the VCF file, we counted the number of loci and names of the individual trees and checked their names. Afterward, we skimmed out the triallelic sites using (`bcftools view --max-alleles 2`) and re-gzipped the uncompressed output. We manually mapped genotyped individuals with the list of phenotyped trees; then we removed an individual that had been genotyped twice to avoid possible biases in either of the samples (`bcftools view -O z -s [`*`IndividualName`*`]`). The individuals without a phenotype were then filtered out with VCFtools (Danecek *et al.*, 2011) (v`cftools --keep [`*`File of IndividualNames`*`] --recode`). We inspected the percent of missing data per site and removed sites with more than 20% of missing calls with VCFtools (`vcftools --max-missing 0.8 --recode`). Each locus was assigned a name in the format "Chromosome_Position" using BCFtools (`bcftools annotate --set-id '%CHROM\_%POS'`). We imputed the resulting VCF with Beagle v. 5.4 (Browning *et al.*, 2018), allowing an upper bound of the memory pool at 100 GB. A reference panel of *F. sylvatica* is currently lacking, so we ran the software package with the default parameters of burnin, iterations, `em` (to allow the update of the N_e and *err* parameter with the values obtained by an expectation maximisation algorithm), overlap, and window size. We updated the header of the resulting VCF with the fasta file of the Bhaga assembly (Mishra *et al.*, 2022) (`bcftools reheader --fai [`*`fasta file`*`]`) and sorted the chromosome in numerical order. Eventually we renamed each SNP as `CHR_POS`. This was the Main Raw Dataset 2009. An individual had been phenotyped only in 2009, so it was removed to obtain a second main Main Raw Dataset 2010. A MAF filter at 0.1 was applied to the two with BCFtools (`bcftools view -q 0.1:minor`), obtaining the two Main Datasets: MD.2009 and MD.2010.

Traits

The mean temperature in 2009 was 4.3°C, and 10.10 °C in 2010. Restricting to the window 1 - 130 DOY (1st January to 10th May, the end of observations) the means are 4.50 and 3.1 °C respectively (**Figure 1**). Phenological scores have been recorded from 23 March to 4 May on 162 trees (2009) and from 8 April to 10 May on 161 trees (2010), using the 1-5 scale published by Davi *et al.*, (2011), which map biunivocally to as many BBCH scores. The resulting time series consisted of a pair of daily scores assigned to the top and bottom of the canopy at each of the 15 and eight observation days (Jean *et al.,* 2023). We relied on the instructions provided by Jean *et al.*, (2023) to first convert the partial scores into the BBCH scale following Table 3, and thus to assign an individual unique score based on the mean between top and bottom canopy (Table 2, “Selected BBCH score”; Jean *et al.*, 2023). We then converted the dates of the observations in Julian days of the year (JDoY from now on) using a set of functions from the R package Lubridate (Grolemund and Wickham, 2011). They are, sequentially: *dmy()* to parse the date in a readable format; *ad.Date()* to change the object class; and *yday()* to do the conversion proper.

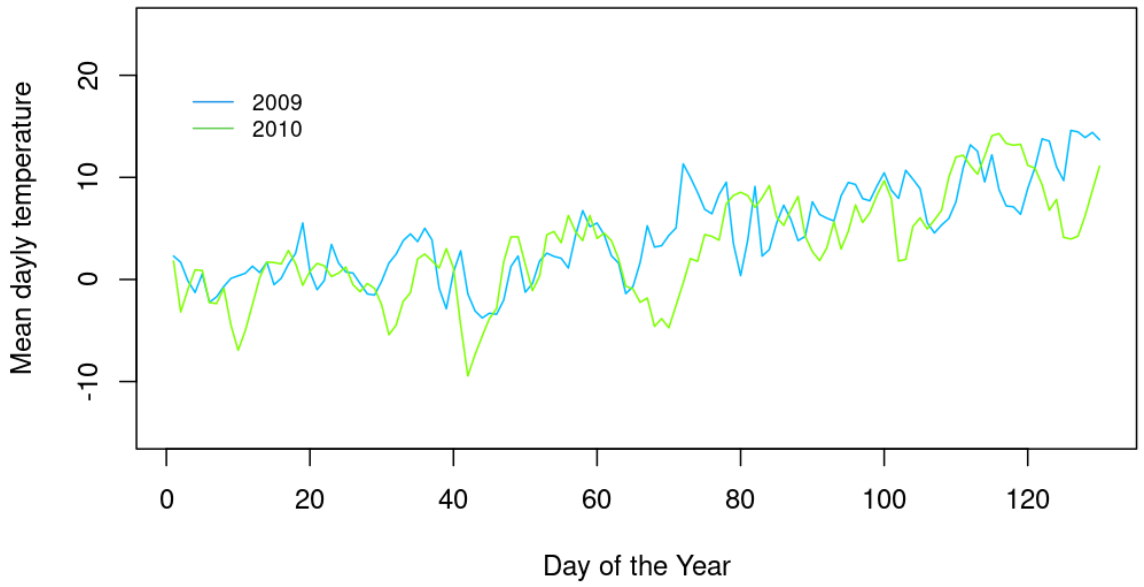


Figure 1: Temperature trend from 2st January to 10th May (1 - 130 DOY) in **2009** and **2010**.

Modelling the relationship BBCH ~ Julian Day

To better compare the phenological stages among different plant species, Hack (1992) developed the BBCH scale (Meier, 2001; Finn *et al.*, 2007), which codes similar phenological growth stages between different taxa of mono- and dicotyledonous plant species. The BBCH scale has since become a valuable tool to describe TBB as a longitudinal trait. However, collecting phenological data during the budding season can possibly result in incomplete time series of the transition for all the followed trees because it's not always feasible to be on the field. Any researcher that relies on these data for a follow-up study will have to adjust their experimental design to the patterns from the field data, thus requiring to remove those individuals whose phenology has not been followed fully because of the tree physiological dynamic. Furthermore, the way the BBCH scale represents a tree depends on many aspects of the canopy (*e.g.* percent of unfolded upper to lower canopy). A tree can be reported with the same score for multiple days, even though the parameters determining the score have slightly changed - until they change enough that the score has to be changed accordingly. The most relevant BBCH stages are the 7th (sensitive to late frost) and 9th (complete individualisation of the leaves; TBB proper).

The sampling on Mont Ventoux had been conducted in a time window and with a frequency in which trees are broadly expected to be attributed the BBCH scores 1, 3, 7, 9, and 10. Trees have been recorded repeatedly in the same stage across multiple observation days. Most of them (132 in 2009 and 129 in 2010 respectively) have been recorded only in four or less stages; in particular, out of 161 individuals in 2009 137 and 119 trees had been reported with the scores 7 and 9, respectively, while in 2010 they were 124 and 151.

To avoid losing statistical power by excluding these individuals from the analysis, we opted to estimate the individual dates at which each tree had been expected to be reported into a given stage based on the available record of scores. These modelled dates of transition would be then used as phenotypes for the GWAS.

We defined the budburst dynamic as $\frac{d(BBCH)}{d(JDoY)}$ to bypass the discrete nature of the BBCH scale. Since the curve of BBCH observation is a function of time, the distribution of these data points approximates to a sigmoid resulting from a logistic function. The model will have a twofold scope: interpolate the not-recorded dates of the BBCH 07 and 09 and provide a summary of the budburst dynamics at the individual level. Logistic regression is a type of Generalised Linear Model, an extension of linear regression that models a categorical outcome to predict the class of the observations based on multiple predictors. Given a Julian day of the year (JDoY) (X), a logistic model returns a Y, the BBCH score (*ie*. a class) a tree is expected to be classified on that day. The user must set the threshold at which a category switches into another. The standard logistic function has equation

$$y = \frac{1}{1 + e^{-x}} \qquad [1]$$

The independent variable *x* is defined in]−∞;∞[, and its $f(x)$ is bound (0,1). The point of inflection is at *x* = 0 and it reaches an asymptotic level at 1. In this case, the explanatory variable *x* is the JDoY and the response the BBCH. To interpolate the JDoY of a given BBCH score, we are seeking the JDoY at which a tree is expected to be in a given BBCH score, so now *y* (BBCH) becomes the explanatory variable and *x* (JDoY) the response. The process consists in fitting a straight line in the interval of *x* in which the curve changes shape and making it pass through the flex point of the logistic. Analytically, it corresponds to approximating the set of points of logistic as a straight line with its own slope and intercept. We can thus solve equation [1] for *x* and linearise it as

$$y^{-1} = 1 + e^{-x}$$
$$y^{-1} - 1 = e^{-x}$$
$$log(y^{-1} - 1) = -x$$
$$-log(y^{-1} - 1) = x \quad [2]$$

Equation [2] is the inverse function of the logistic, the logit. Getting back to [1], we generalise it to better accommodate the properties of the relationship between BBCH and JDoY; it is obtained by adding the parameters *a, b*, and *c*. The *a* is the slope of the curve at the inflection point; smaller the *a*, steeper the curve; the, *b* is the ordinate at which the curve passes through *x = 0* and determines where the midpoint occurs: smaller the *b*, the smoother the curve; and the *c* is the value of *y* at which the function becomes asymptotic. To retain the mathematical properties of the curve, we normalised the response variable BBCH by its maximum value so that the scores are all < 1 and $lim_{x->max(BBCH)} f(x) = 1$. Consequently, the point of inflection too has a BBCH > 1 and we could leave *c* apart. Whilst this was straightforward on our studied population, since the maximum reported score was 10, the assumption holds also for higher values of BBCH. Therefore, the generalised logistic function with two parameters is

$$y = \frac{1}{1 + e^{-a(x-b)}}$$

The R function *glm()* takes a vector of two variables (*x, y)* - JDoY and BBCH - and identifies the parameters of a linear equation corresponding to the linearised logistic function. We used the custom function *LoopingGLMgeneral()* to compute the GLM for each tree and obtain the parameters *s* and *b*. With these values the, equation [3] was linearised and solved for *x* as

$$y^{-1} = 1 + e^{-a(x-b)}$$
$$y^{-1} - 1 = e^{-a(x-b)}$$
$$log(y^{-1} - 1) = -ax + ab$$
$$-log(y^{-1} - 1) = ax - ab \quad [3]$$

From equation [3b] we linearised the response variable (BBCH observations)

$$Y = -log(y^{-1} - 1)$$

Since this is a linear equation, its parameters are

$$Slope = a \quad [5]$$
$$Intercept = -ab \quad [6]$$

Putting into a system the equation [5-6] we describe the linearised equation which parameters *slope* and *Intercept* the Generalised Linear Model estimates as

$$Y = -Slope \cdot x + Intercept$$

From these linear parameters and equations [5-6], we derive those of the logistic curve:

$$a = Slope$$
$$b = -\frac{Intercept}{Slope}$$

We can thus write the response (BBCH) as the sum of the intercept and the predictor times its coefficient:

$$BBCH = a(JDoY) - ab$$
$$BBCH = Slope(JDoY) + Intercept$$

With the parameters *a* and *b*, we come back to the equation of the logit [2] and interpolate the date in which a given BBCH stage has been attained by inverting equation [3].

$$\frac{-log(y^{-1}-1)}{-a} = \frac{ax}{-a} - \frac{ab}{-a}$$

$$\frac{log(y^{-1}-1)}{a} = -x + b$$

$$\frac{log(y^{-1}-1)}{a} - b = -x$$

$$x = -\frac{log(y^{-1}-1)}{a} + b \qquad [10]$$

In this way we obtained a series of expected dates of passage for the five stages recorded in the field. We used the expected date of passage for BBCH 07 and 09 in the two years as phenotypes in the GWAS (JDoY.7_2009, JDoY.9_2009 and JDoY.7_2010, JDoY.9_2010, respectively). From now on, when we refer to pair JDoY.7 and 9 we will use the denomination “spring phenology”. In addition to the expected dates of passage, we also investigated the individual bases of the parameters *a* and *b* considering them proxies for the rate of change of the phenological transition and the overall timing of spring phenology, respectively (Logistic.a_2009, Logistic.a_2010; Logistic.b_2009, Logistic.b_2010), akin to other GWAS studies that used as phenotypes the values of the parameters of the model that describes the dynamic of biological growth processes (Duan *et al*. 2021). **Figure 2a** represents a logistic fitted to the timeseries of phenological scores of a tree.

To assess the robustness of the model we assessed the effect of the scarcity of points on the estimation of the date of passage. We degraded the time series by removing one even-numbered observation each two (simulating a less accurate sampling) and the observations of the three days when transition from BBCH 07 to BBC 09 occurs more often (simulating a less precise sampling). We estimated the individual parameters of the logistic fitted to the degraded time series and, using the logit function, interpolated the dates of passage for the BBCH scores under the degraded time series (**Figure 2b)**. We compared the three models (complete time series, inaccurate time series, imprecise time series) computing the offset between the observed and interpolated dates, the McFadden’s R^2 (Veall and Zimmermann, 1996; Smith and McKenna, 2013), and the residual sum of squares.

Phenological Score Sum

The Phenology Score Sum (PSS) is the sum of all BBCH stages a tree has been assigned during the observation window. A higher PSS corresponds to having spread the leaves for a longer time during the year and a quicker pace of the bud burst dynamic. PSS values were computed after the scores had been converted to the BBCH scale and averaged between the upper and lower parts of the tree crown. We computed the PSS_2009 and PSS_2010 by summing the individual BBCH scores across the observation window.

Degree Days

Degree days for a BBCH score are the sum of the mean temperatures of the days in which the mean temperature was equal or above 5 °C (forcing temperature, Osada *et al.*, 2018) before a tree has attained BBCH 07 and 09. We used the hourly records of temperatures in the stand N1 in the two investigated years that have been made available by Jean *et al*. (2023). We rounded up to the next integer the expected dates of passage obtained from the logistic interpolation and summed up the mean daily temperatures above 5 °C for each day until the end of the highest dates of passage. We then matched the individual temperature sum of the (rounded up) Julian date in which a tree was expected to be assigned a given score to the individual date itself, obtaining the sum of temperatures that the tree must have experienced when it was expected to be assigned a given score (DD.7_2009, DD.9_2009, DD.7_2010, DD.9_2010).

Inter-annuals differences

We checked whether data collected in 2009 and 2010 were homogeneous in terms of data collection methods, so that they could be compared in a meaningful way (notice that, because variations in the rank of budburst dates were observed between the two years (see below) the data sets from the two years were treated independently). In 2010 we found that the dates of transition of BBCH 07 and BBCH 09 were homogeneous among trees: many more trees reached a given phase in the same day than in 2009. This suggests that the data series in 2009 and 2010 may have intrinsic differences caused e.g. by different observers or an inadvertent change in the interpretation of stage observations between the two campaigns. At the same time, 2010 had about half the observations of 2009; so it may just be that we observed fewer different types of dynamics in 2010 relative to 2009 just because of the simplified data set. To test whether those differences could only be caused by the resolution in the data sets (thus excluding a methodological bias) or there were deeper reasons to the observed differences, we proceed as follows. We reduced the number of observations per time series of 2009 to those of 2010 to make the two data sets comparable. Then we counted how many time series have been observed in a given number of trees (*e.g.* five time series were observed only in one tree, three time series were observed in two trees, and so on) using the custom function *spectre()*. This way we obtained two multinomial observed distributions, one for each year. We then ran a multinomial G-test, using the distribution of 2009 as the expectation, and asking whether the distribution of 2010 conformed to that expectation, and vice versa. When the expected frequencies are those observed in 2010:

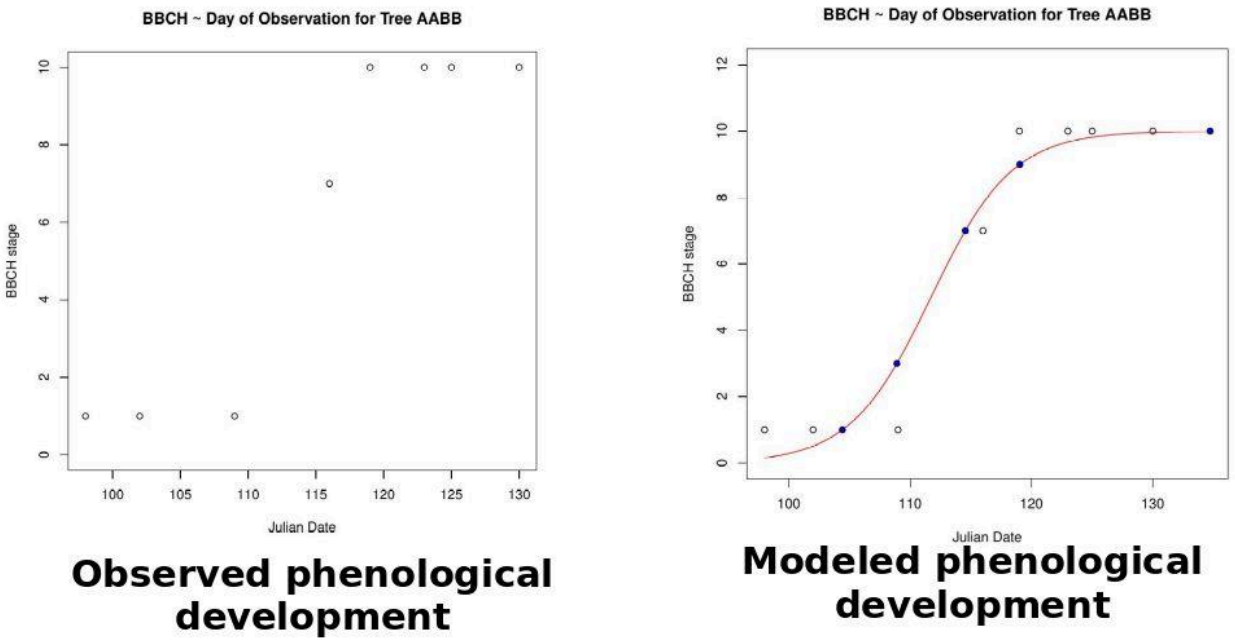
$$Likelihood\ (Patterns\ 2009 |\ Patterns\ 2010) \ = \ 2 \cdot \ log_{10}\left(\frac{L(O(frequencies)_{2009} \mid E(frequencies)_{2010}}{L(O(frequencies)_{2010} \mid E(frequencies)_{2010}}\right)$$

(tested against a chi-square distribution with 11 d.f.), and when the expected frequencies are those observed in 2009:

$$Likelihood\ (Patterns\ 2010 |\ Patterns\ 2009) \ = 2 \ \cdot \ log_{10}\left(\frac{L(O(frequencies)_{2010} \mid E(frequencies)_{2009}}{L(O(frequencies)_{2009} \mid Efrequencies)_{2009}}\right)$$

(tested against a chi-square distribution with 12 d.f.)

(a)



f(Julian date) = BBCH score

(b)

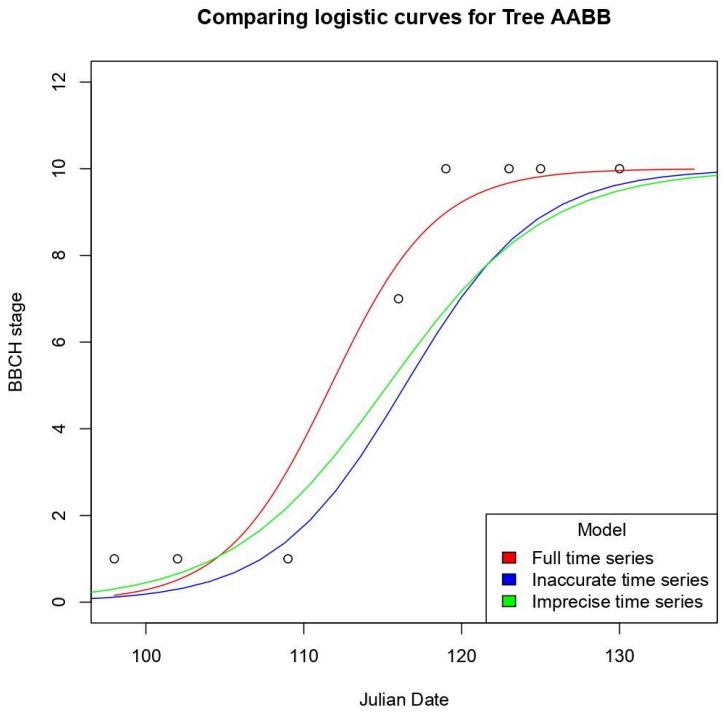


Figure 1: (**a**) Interpolation of bud-break scores by the means of a logistic function. (**b**) Application of the logistic model to an **imprecise time series** (only the even-numbered observations of the **full time series**) and an **inaccurate time series** (removing the third, fourth, and fifth observation of the **full time series**).

Genotype-Phenotype Association algorithms

We tested the association between SNPs and phenotypes with the FarmCPU version (Fixed-and-random-model Circulating Probability Unification; Liu *et al.*, 2012) implemented in the R package rMVP (Yn *et al.*, 2021). The input genotypes, phenotypes, genomic maps and kinship matrix files were obtained with the function *MVP.Data()* with the option *fileVCF* for the genotypic input. FarmCPU builds on the Multi-Locus Linear Mixed Model (Segura *et al.*, 2012) and decomposes it into two sub-algorithms which work separately to update each other. The algorithm of the Fixed Effect Model (FEM) tests one marker at time to remove the confounding effect between the testing marker, kinship, and population structure. Each marker is modeled independently from the others and any SNP which can be associated with it is included as covariate to tackle the issues of False Positives. The Random Effect Model (REM) estimates the associated markers included into the FEM to avoid overfitting. It uses a Maximum-Likelihood approach which relies on the associated markers linkage map and their test statistics derived from the FEM to compute the kinship among individuals (SUPER, Wang et al., 2014). The overfitting in the FEM is avoided because the REM uses few known parameters of the markers to include as covariates, while their genetic and residual variance remains unknown. The FEM model is

$$y_1 = M_{i1}b_1 + \ldots + M_{it}b_t + S_{ij}d_j + e_i$$

where M_{i1} to M_{it} are the genotypes of the individual *i* at *t* associated markers; b_1 to b_t are the corresponding associated marker effects; S_{ij} is the genotype of the *i*th individual at the *j*th tested marker; d_j is the effect of the *j*th thested marker; and e_i is the distribution of the residuals with mean 0 and variance σ_i^2. At the start of the algorithm, the set from M_{i1} to M_{it} is empty. At each iteration for a given testing marker S_i, FarmCPU examines the association between S_i and the associated marker M_i and assigns to the latter the most significant p-value which has been computed for it. After S_i and each M has been assigned a p-value, the REM algorithm kicks in. The REM model is

$$y_1 = u_i + e_i$$

where u_i is the total genetic effect of the individual *i*, hence a random term. In a population of *n* individuals, $E(\Sigma_{1\ to\ n}^{n}\ u_i) = 0$, whose variance-covariance matrix is $G = 2K\sigma_a^2$, where K is the kinship estimated from the M_{it} and σ_a^2 an unknown genetic variance. The set of M_{it} who maximise the likelihood of the REM model replace those used in the FEM. The iterations go on until REM provides to FEM the same set of M_{it} it had been received as input. FarmCPU performs the GP tests on all S_i markers using the M as covariates, hence significant p-values are not expected for M not included as covariates.

GWAS design

A GWAS requires a large sample of genotyped individuals to reach the statistical power needed to identify marker-trait associations (Ekblom and Galindo, 2011; Santure and Garant, 2018). Furthermore, whenever the number of markers exceeds that of the phenotyped individuals, there is a high risk of False Positives association tests: this is also known as the "large *p*, small *n*" problem. Tandemly, a small sample size ($n < 500$) also causes the algorithm to be affected by the Baevis effect (Otto and Jones, 2000; Xu, 2003). This pitfall consists in missing the small-effect loci because of low power, whilst overestimating the effects of the large-effect loci. Since our sample size was relatively small, we risked lacking power, to get inflated estimates of the allelic effects, and to identify False Positives. Whilst reducing the sample size indeed exacerbates the "small *n*, large *p"* problem and the Baevis effect, we reasoned that the entire dataset can be deemed to provide enough statistical power to identify true QTLs if those identified on a subsample of it can explain the observed trait variation in another (non-overlapping) subsample. The number of false positives used in the testing step can be reduced by running the GWAS on a batch of multiple subsets, instead of a single one, and then choosing the loci that resulted significant more than twice (consensus loci (**Figure 3**).

We sampled 100 individuals from the two main raw datasets as the training dataset and the remaining individuals composed the testing dataset. In this way we generated from each main raw datasets eight randomly shuffled replicated training datasets and as many testing datasets to make the global results more robust to the random composition of the training datasets. We conducted the random sampling on R v. 4.1.2 (R Core Team, 2022) using the base function *sample(x = individualNames, n = 100)* and outputted a text file. This file enabled us to subset the VCF file with BCFtools (`view -S [file of IndividualNames])` to keep only the individuals of each training dataset. We applied a MAF filter of 0.1 with BCFtools (`view -q 0.1:minor`).

GWAS: training datasets analysis

The GWAS experiments were run using the function *MVP()* with a maximum of 1,000,000,000 iterations of the FEM/REM cycle, 1,000 permutations the (`maxLoop` parameter), EMMA as variance decomposition method (Kang *et al.*, 2008), and a static bin. We converted p-values into q-values with the function *qvalue()* from the R package *qvalue* (Storey and Tibshirani, 2003) and deemed as significant those SNPs which passed the FDR 5 %. Among all the SNPs which resulted significant across the eight training datasets, we identified those SNPs which have been retrieved twice or more times as consensus loci. Different breeding values may have been estimated for a given SNPs retrieved in multiple training datasets; we selected the maximum allelic effect estimates in absolute value among them. The FEM/REM loop disentangles clustered signals, yet, due to differences in sample subsets, FarmCPU identifies slightly different groups of signals for a given genomic region in different runs, with variations in effect size, significance, and actual sets of significant variants. It was likely that the algorithm spreaded the signal of a central QTL over a genomic window, resulting in multiple False Positives, each with its own estimated allelic effect value. We first identified the clusters by inspecting the distribution of the pattern of distances between significant variants along each chromosome. We deemed two variants to be part of the same cluster if they were closer to each other than a threshold - the average intra-chromosomal distance between significant variants multiplied by a scaling factor. The

variants farther than this threshold from the previous one along the chromosome would mark the end of the cluster. If the distance from the following variant were less than the threshold, a new cluster would start; otherwise, this very variant would be deemed an isolated peak. We empirically assessed the trend of identified clusters, cluster length, and isolated peaks identified by a range of scaling factors values from 0.1 to 1 with a pace of 0.05. We chose the value at which the rate of the statistics change was the steepest. Among the significant variants close enough to be part of the same cluster, we selected as the peaks the one with the largest effect size in absolute value.

GWAS: validation in the testing datasets

The eight testing datasets had the same number of loci as the training datasets before the application of the MAF filters, because they both descended from the same main raw datasets. We extracted the genotypes of the testing datasets individuals at the consensus loci identified as peak using BCFtools (Danececk *et al*., 2011, Li *et al*., 2011; Danececk *et al*., 2021) (`filter -R [ContigPOS.file.txt]`) and converted the genotypes into the BIMBAM format using QCtool. We estimated the breeding values in the testing datasets as $ALT.Allele_EffectSize_{locusL} \cdot ALT.Allele_Dosage_{locusL}$. The degree to which these loci explained the variation in the trait within the testing datasets was assessed with the R^2 of the linear regression of the computed (expected) breeding values against the observed phenotypes. The Beavis effect was likely to affect the allelic effects of the loci retrieved in the training datasets because of these subsets' small sample size. However, the eight training/test replicates made sure that the effect sizes of the identified loci were able to explain enough the phenotypes in any of these independent (to a degree) subsets. We thus scaled phenotypes and breeding values and centred them to 0 to highlight the regression itself rather than the magnitude of its factors. Since we were using a smaller subset of the significant loci retrieved in the replicate GWAS, we also tested whether the estimated breeding values of the peak loci alone were able to estimate the phenotypes within the training datasets

GWAS design: main datasets analysis

The main datasets were analysed with the same parameters as the training datasets.P-values were converted into q-values. SNPs which passed the FDR threshold of 5 % were deemed as significant. Clustered and isolated peaks were determined based on the scaled chromosomal distance threshold applied in the training datasets. We checked the distribution of the allelic effects in absolute value. Patterns of shared architecture between traits have been investigated graphically with UpSetR (Conway *et al.*, 2017).

Annotations

We retrieved the annotation of the significant SNPs using the beech genome browser created by Mishra (2022) on the platform JBrowse (Buels *et al.*, 2016). Accessions with the same gene name were subsequently searched on UniProt (The UniProt Consortium, 2023) to retrieve the GO terms (https://www.uniprot.org/).

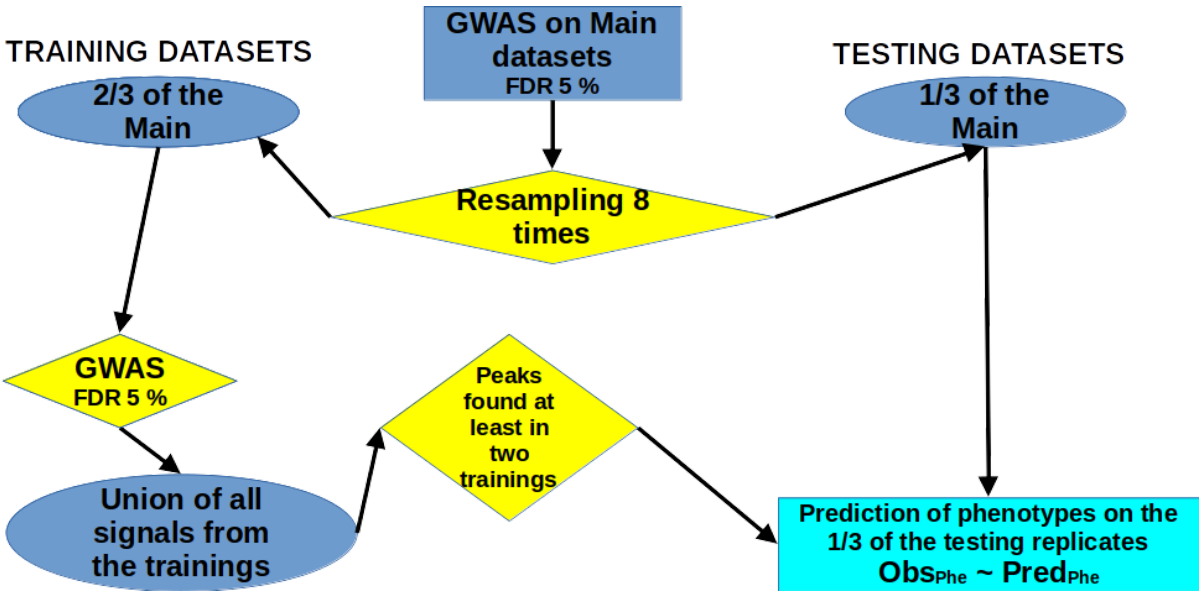


Figure 2 Pipeline for the training/testing approach. To obtain the Training datasets 100 individuals have been sampled eight times from the Main dataset, and those that were not included composed the Testing datasets. A MAF filter was then applied. GWAS was run in the Main datasets, then replicated in the eight Training datasets. From the union of all the SNPs associated to a trait across the eight Training datasets, all those that were found twice or more times were used to estimate the phenotypes of the Test dataset.

Results

Genotyping and phenotyping

We matched the 162 phenological time series recorded in 2009 and the 161 recorded in 2010 phenotypes from Muratorio et al. (2018) with the 166 genotypes provided by the BEECHENOMEs consortium. We retrieved an individual tree sequenced twice and retained the genome assembly with more polymorphisms. Hence, we counted 162 unique individuals for which we had both the phenotype measurements for at least one year and the genotypes. The percent of missing data was 0.097 %. After the removal of the sites with more than 20% of missing data, the VCF contained 6,717,158 polymorphisms, with 0.0822 % missing data. The VCF obtained by the BCF through BCFtools (Danececk *et al*., 2011, Li *et al*., 2011; Danececk *et al*., 2021) included 382 individuals genotyped at 7,793,706 loci. After the removal of the triallelic sites, the dataset included 7,424,658 polymorphisms. After having removed a copy of the individual genotyped twice and retained the individuals phenotyped on the Mont Ventoux population, we were left with 162 individual trees genotyped at 7,424,658 loci; the number of polymorphisms dropped to 6,703,614 after removing the sites with more than 20% of missing data. In **Table 1** is reported the number of loci after having selected the individuals and filtered with MAF < 0.1, both for the VCF files of the two years and the training datasets obtained from them, which brought the numbers of variants down to under four millions.

Year	Dataset composition	Code	N° polymorphisms
2009	162 individuals	Main 2009	3,891,047
	Training Dataset: 100 individuals *Testing Dataset*: 62 individuals	Rep 1 2009	3,842,004
		Rep 2 2009	3,849,916
		Rep 3 2009	3,844,969
		Rep 4 2009	3,858,985
		Rep 5 2009	3,835,664
		Rep 6 2009	3,844,969
		Rep 7 2009	3,844,852
		Rep 8 2009	3,838,403
2010	161 individuals	Main 2010	3,878,975
	Training Dataset: 100 individuals *Testing Dataset*: 61 individuals	Rep 1 2010	3,844,493
		Rep 2 2010	3,837,059
		Rep 3 2010	3,849,121
		Rep 4 2010	3,841,005
		Rep 5 2010	3,839,069
		Rep 6 2010	3,849,121
		Rep 7 2010	3,853,446
		Rep 8 2010	3,843,690

Table 1. Table reports: **Year**: year in which the phenological scores have been recorded; **Dataset composition**: number of samples phenotyped by year and repartition between training and testing datasets; **Code**: identification code of Main and Training datasets; **N° polymorphisms**: number of SNPs in each dataset.

Phenotype analysis

Logistic models of Julian days

The logistically interpolated (expected) dates of attaining BBCH 07 had a median positive offset of 0.01 and 0.06 Julian days with respect to the day in which a tree has been recorded with the same score (ranges: -4.479 to 10.254 DOY in 2009; -1.452 to 4.376 DOY in 2010); the median difference between the field and interpolated dates for BBCH 09 was 4.63 and 3.04 DOY (ranges: 1.117 to 25.799 DOY and -0.0482 to 7.311 d). Notably, BBCH 10 displayed the largest difference between the fields and observed dates, 36.02 and 17.35 in 2009 and 2010, respectively. The median of Logistic.a were 0.12 and 0.23 (2009, 2010), while those of Logistic.b 98.75 and 114.52 and 114 (2009, 2010). The R^2 of the models over the full time series was in between the models built on incomplete time series but with a broader range of values. **Table 2** reports range and mean of the R^2 of the models built over the full, inaccurate (less intense sampling), and imprecise (sampling missed the days when usually occurs a BBCH stage), the mean offset for date of BBCH 09 under the three models, and the percent difference between the observed and interpolated dates of BBCH 09.

Inter-annual differences in dates of passage and in phenological score patterns

The dates of passage recorded in 2010 were overall late with respect to 2009, with mean offsets of 11.1, 10.4, and 10.60 DOY (for BBCH 07, 09, and 10 respectively). The conversion of the 1-5 score scale published by Davi *et al.* (2011) to BBCH scores resulted in 88 classes of time series in 2009 and 20 in 2010, respectively. These numbers remained unchanged after removing the individual not followed in both years. In 2009 the four pattern classes that were observed in at least five trees were shared by 106 individuals, while in 2010 the patterns classes observed in at least five trees were 11 and they were common to 139 individuals. The distribution of trees with a given pattern of BBCH scores had 22 and 17 classes (2009 and 2010, respectively). The final data sets for the multinomial tests (after removal of empty classes in the data sets forming the expectation) comprised 13 and 12 classes, respectively. The two tests turned out to be non-significant (test of 2010 against the 2009 expectation: $G = 0.8997488$, 12 d.f., $P = 1$; test of 2009 against the 2010 expectation: $G = 13.3063$, 11 d.f., $P = 0.29$). We therefore concluded that the two groups of score patterns did not differ due to differences in method. Nonetheless, the ranks of individual trees in JDoY for BBCH 07 and BBCH 09 differed greatly between years, suggesting the presence of Genotype x Environment effects (**Figure 3**) Therefore we concluded that the data of 2009 and 2010 may not be treated as a merged data set.

Year	2009	2010
R^2 full	0.76 (0.37 – 0.98)	0.88 (0.81 – 0.92)
R^2 inaccurate	0.75 (0.28 – 0.98)	0.920 (0.78 – 0.98)
R^2 imprecise	0.79 (0.78 – 98)	0.85 (0.77 –0.91)
BBCH 09 full	+6.0 (+1.1 – +25.8)	+3.3 (-0.05 – +7.3)
BBCH 09 inaccurate	+6.4 (+0.9 – +35.7)	+3.8 (-1.1 – 9.4)
BBCH 09 imprecise	+4.7 (+0.01 – +21.70)	+4.1 (+0.48 – +8.80)
Mean % offset Pred(BBCH09) – O(BBCH 09)	+5.24	+2.70
Mean % offset P.Inaccurate(BBCH09) – O(BBCH 09)	+5.21	+3.60
Mean % offset P.Imprecise(BBCH09) – O(BBCH 09)	+4.10	+3.35

Table 2: Range and mean McFadden R^2; range and average offset for BBCH 09 (in DOY) with respect to the observed for the models obtained from the full and degraded time series; mean percentage offset for BBCH 09 (in DOY) with respect to the observed for the models obtained from the full and degraded time series.

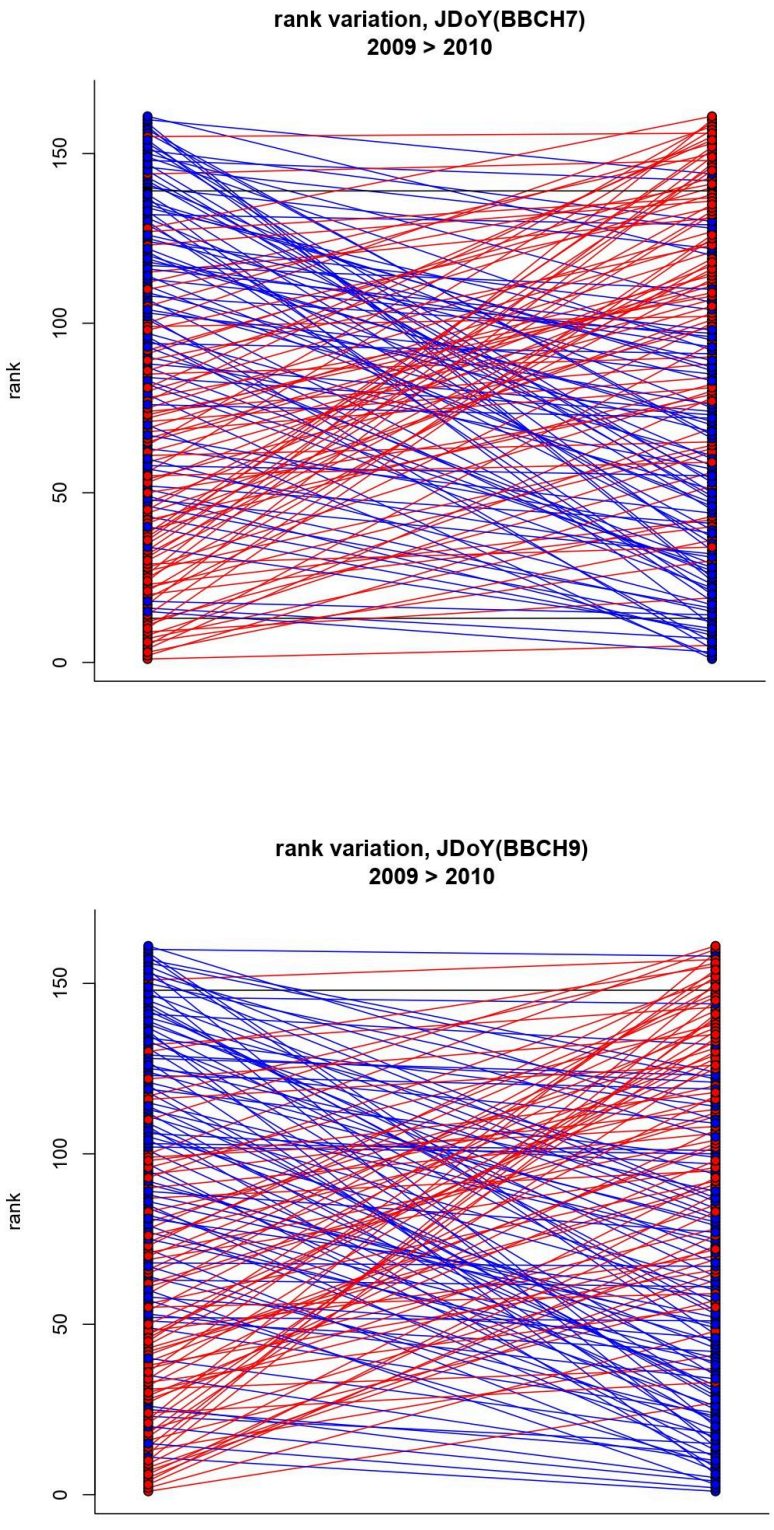


Figure 3: Change in rank for the date of passage of BBCH 07 and 09 between 2009 and 2010. **Red** lines connect trees the 2010 rank of which was higher than 2009, **blue** lines connect trees the 2010 rank of which was lower than 2009, and **black** lines connect trees the rank of which remained the same across the two years.

Phenotype-Genotype Associations

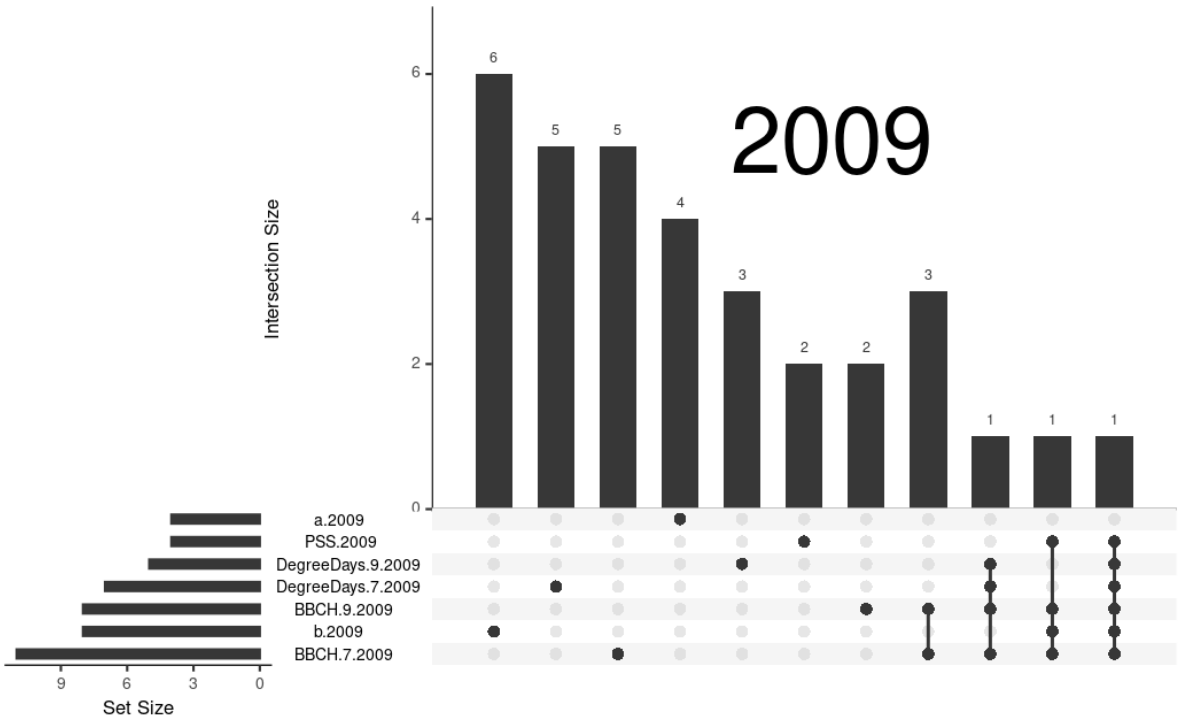
Main Datasets

Across the fourteen traits, 73 tests were significant at an FDR of 5%, for a total of 47 unique SNPs. The traits measured in the two years did not share any SNP. The trait associated with most QTLs was JDoY.7_2009 (11) and Logistic.a_2010 the one associated with fewer QTLs (1). In spite of the small number of QTL and the inflation of the allelic effects, traits associated with more QTLs displayed a distribution of allelic effects skewed toward the classes with the smallest effect, with only one QTL havin129g a large effect (JDoY.7_2009, JDoY.9_2009, DD.7_2009, Logistic.b_2009, DD.9_2010). We observed that the allelic effects estimated in the main datasets were smaller by a factor of 2 to 10 than those estimated in the training datasets0. The estimated values abundantly exceeded the standard errors. In 2009, the dates of passage of BBCH 07 and 09 shared seven QTLs out of 13; two of them were common to the two DegreeDay and two with Logistic.b and PSS. In both years, the two DegreeDays shared two out of 10 (2009) and one out of 9 (2010) QTLs. aThe QTL of Logistic.a_2009 were private. In 2009, only two QTLs were in common between JDoY.7 and 9 the dates of passage of BBCH 07 and 09, while JDoY.7_2010 and Logistic.a_2010 shared two QTL. A summary of the intersection between traits is available in **Figure 4**.

Training/Testing Replicates

The R^2 of the *ObservedPhenotypes ~ BreedingValues* across replicates spanned from 0.06 in training dataset 3 of Logistic.b_2009 to 0.74 in training dataset 16 of JDoY.7_2009). We observed a wide variation in the number of significant SNPs across the replicated training datasets at the FDR of 5 %. The replicates of PSS, JDoY.7 and Logistic.b (2010) were consistently associated with 10 variants or fewer but in training dataset 16, while Logistic.a_2010 was found to be associated with between 3 and 32,879 variants. The number of SNPs retrieved at least twice over the eight replicates spanned from 18 to 38,379. For the removal of duplicated signals, the inspection of the scaling factor revealed that the steepest drop in the cluster statistics occurred when the intra-chromosomal distances were scaled by 0.015 (**Supplementary Figure 1**). Most of the SNPs associated with the traits were mapped to clusters (max: 96.3 %, min: 0 %); on the whole, isolated SNPs tended to be more numerous than clusters (proportion of isolated SNPs, max: 100 %, min: 46.2 %). Depending on the trait, the estimated breeding values explained from 6.6-56 % (Logistic.b_2009) to 53-74 % (JDoY.7_2010) of the phenotypic variation in the testing datasets. The R^2 of the GWAS models used by FarmCPU in the training datasets were larger than those from *ObservedPhenotypes ~ BreedingValues* (**Supplementary Table 1**).

(a)



(b)

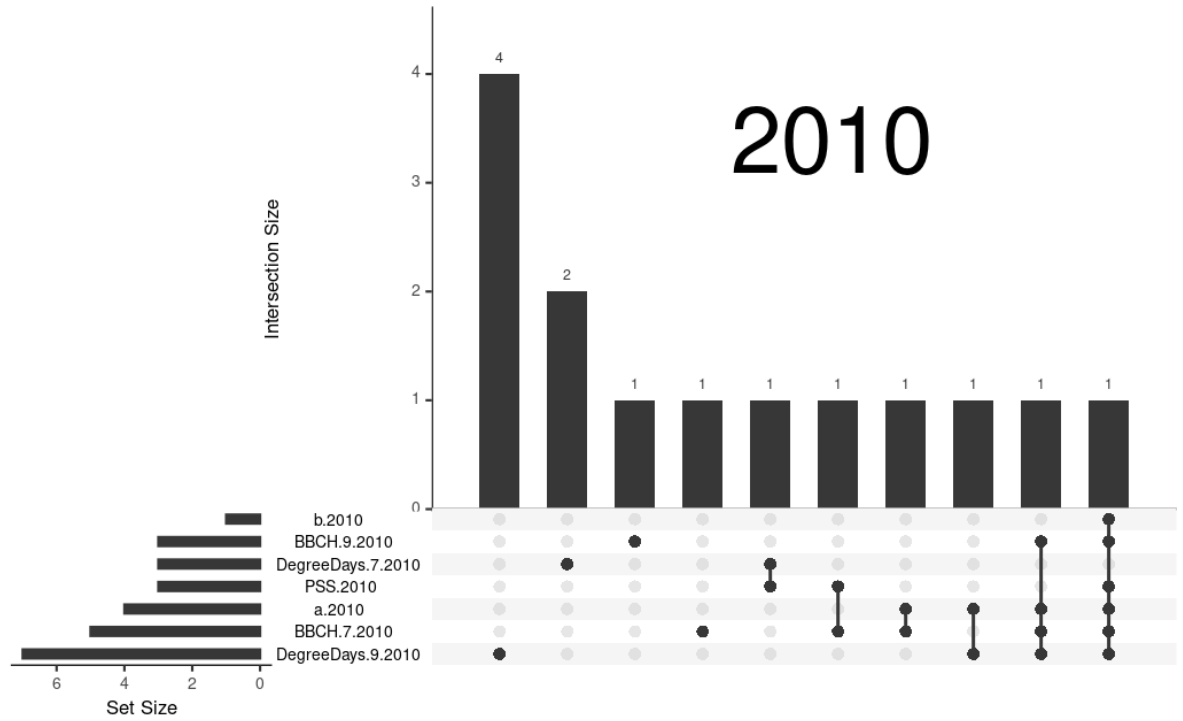


Figure 4 Repartition of the identified QTLs associated with the phenological traits in (**a**) 2009 and (**b**) 2010.

Genomic patterns

The chromosomes 1, 5, 6 and 7 harboured 57.4 % of the QTL. Chromosome 7 harboured most of the QTL (seven) whilst chromosome 9 and 11 only one. The R package karyoploteR was used to obtain a depiction of the distribution of the QTLs (Gel and Serra, 2017) across and along the chromosomes (**Figure 5**). QTL associated with traits measured in 2009 were mapped mostly on chromosome 1 (nine), seldom on 2 and 8 (two) and never on 9, while those from 2010 were concentrated on chromosome 7 and 11 (six QTLs each); no QTL for a trait of 2010 was found on chromosomes 2, 4, 10, and 12. Overall, 22 QTLs were found in genes, especially within introns (11), while the remaining 25 were located in non-coding regions within 10 kbp from at least one gene; seven of them were within 1000 bp from a 5''- or 3'' UTR. Three QTLs mapped at ~ 11 kbp from each other on chromosome 6 and two were ~39 kbp apart on chromosome 1 on opposite strands. A complete table of annotation is available in **Table 3**. Distribution of the allelic effects are reported in **Supplementary Figure 2** (Main datasets) and **Supplementary Figure 3** (training datasets).

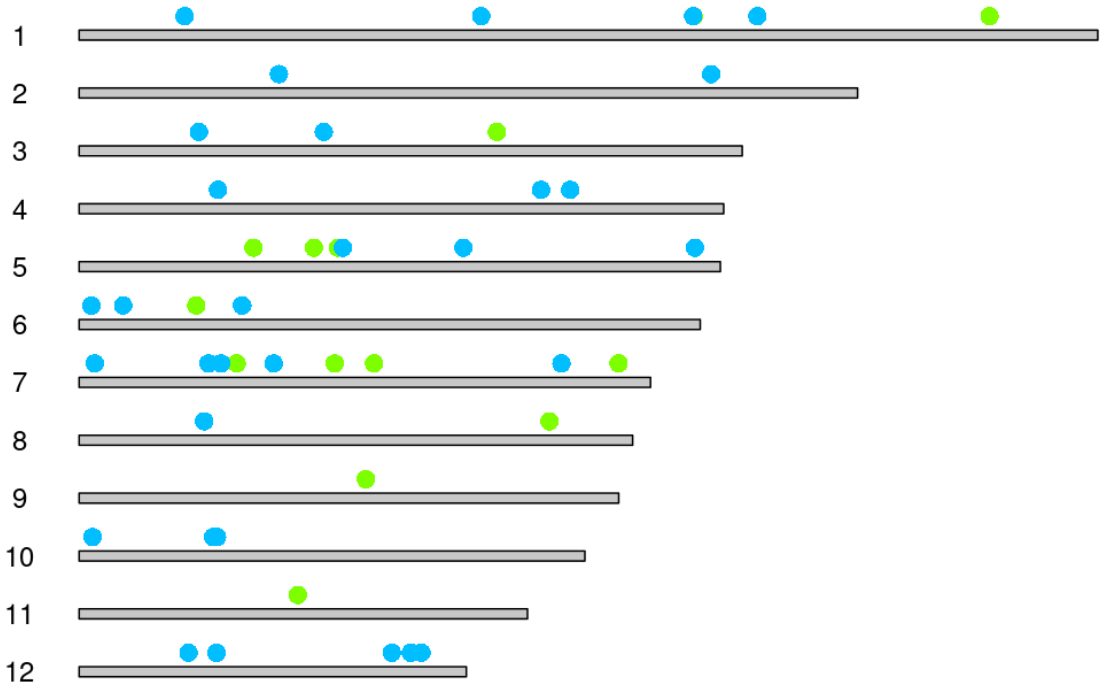


Figure 5 Distribution of the QTL associated with TBB in **2009** and **2010**.

Discussion

In this work we have identified the genomic bases of European beech spring phenology traits, as they are expressed in wild conditions. To do this, we have modelled the phenological development through time, resting on incomplete observed time series of BBCH scores, thus interpolating a new set of expected, but unobserved, dates of passage.
We first discuss the advantages of the logistic modelisation of the BBCH as a function of time to describe the bud burst as a dynamic process. Next, we show how the training/testing approach can assess whether the sample composition of our dataset harboured both enough power to detect significant associations and was robust to the Beavis effect. Though our sample size caused overestimation of allelic effects (Otto and Jones, 2000; Xu, 2003), we found that, for some traits, they were distributed according to the expectation of the infinitesimal model - one locus with a larger effect and more loci with smaller ones - both in the training datasets and in the main datasets. Finally, we put the identified QTLs into a wider context of plant genomics.

BBCH ~ date of passage

The BBCH scale has been developed to compare the progression of the phenology among species that differ in timing, environmental triggers, and morphological characteristics (Finn *et al.*, 2007). When used to follow the phenological transition in a single species, the patterns it describes are strictly dependent on the frequency of observations. Because the transition through stages does not proceed at a uniform rate, it is likely that field observations miss the date of passage for the stages that transition more quickly, which also happen to be the most crucial ones in leaf unfolding. Therefore, dates for such stages are not always observed directly in the field and need to be interpolated. We tackled this constraint by representing the phenological development as a logistic curve, which fits the observed pattern of scores as a function of time, relying on and complementing the information provided by field observations.
A logistic curve provides an individual-based description of the phenological development based on its record of BBCH scores. No time series is actually “complete” and, indeed, the model can be reasonably used when there are missing scores in the record. Equation [11] allows a straightforward estimation of the day in which a tree is expected to attain a given phenological score using the parameters of the curve itself and the BBCH score of interest.
The parameters of the logistic curves are mathematical representations of individual tree physiology. The larger the Logistic.a, the faster the pace of a phenological transition; the smaller the logistic.b, the earlier a tree will start the bud-break.
Opposite to the BBCH scale, the findings from a logistic phenological model should be used to compare different populations only after careful manipulation. In fact their values depend on the number of observations on which the individual curves have been built on; a different observation frequency can thus lead to different estimates.
The average estimate for date of BBCH 09 on the complete time series were late by 6 and -3.3 DOY (2009 and 2010); the delay decreased for the imprecise (6.4 DOY in 2009 and 3.3 DOY in 2010) and the inaccurate time series (4.7 DOY in 2009 and 4.1 DOY in 2010). Even at first sight it may appear that a less accurate time series record may even improve the description of the phenomenon, it must be noted that this happens because the BBCH “jumps” scores in place of progressing continuously, so that a tree may be described by the

same score in multiple days. Even though increasing the days of observation adds noise to the model, it is not possible to know *a priori* on which day the tree will be deemed to have significantly progressed the bud-break. In spite of these offsets, the interpolated dates of BBCH 09 were larger than the field dates by 2.7 (2009) and 5.4 % (**Table 2**). We tested it also on the window 1-15 (1,9,11,15) using field observation recorded by another researcher between 2022 and 2023 in the beech forest of La Massane (Clément and Modica, in preparation). The median difference between modelled and field dates were -0.6 and 0.12 DOY (BBCH 09), -1.99 and 2.68 DOY (BBCH 11), and 32.26 DOY and 28.91 DOY (BBCH 15) for 2022 and 2023, respectively.

The model was robust for estimating the expected date of passage, but the estimate of the highest score of the chosen window will have a large positive offset with respect to the observed. In fact, a process with a logistic dynamic reaches saturation at the limit of the greatest value of *Y* - the asymptote - whereas leaf development actually continues beyond it. The logistic model therefore predicted the dates around the inflection point with high accuracy and precision. The higher the score, the nearer the asymptote and the larger will be the offset between predicted and observed date. This pitfall can be avoided by fitting to the time series sigmoidal models that better capture the dynamic of the bud-break, such as the Gompertz curve (Winsor, 1932; Tjørve and Tjørve, 2017). This curve is asymmetric for the point of inflection and slows down as it approaches the asymptote; it is commonly used to describe growth processes, from animals to plants and tumors (Tjørve and Tjørve, 2017).

Following our model, trees followed significantly different patterns of spring phenology between years, though individual timing was consistent. The short inter-annual span of the available time series hampens the identification of the reason for this discrepancy, though we propose this could be a consequence of the different climatic conditions, resulting in a distinct number of degree days. Such a hypothesis is better tested with an explicit phenological model that accounts for chilling and warming requirements (*e.g.* PHENOFIT; Chuine and Beaubien, 2001; Gaüzere, Teuf, *et al.*, 2020) and a set of time series spanning across multiple years.

Training/Testing *in silva*

GWAS takes advantage of a combination of whole-genome resequencing and small size of linkage blocks to minimise both error type I and error type II: with many, genetically independent markers, one makes sure to cover the genome exhaustively (thus maximising power of detection of signals), while avoiding spurious signals (because only markers tightly associated to a QTN turn out as being associated to trait values). Using large sample sizes further improves detection and reduces false positives, as the probability that a marker is detected as associated by chance to trait value diminishes with increasing sample size. Our dataset included 162 (2009) and 161 (2010) individual trees, sample sizes in which an overestimation of the allelic effects is to be expected. To partially cope with the problems associated with small sample size, we used a resampling approach in combination with the standard testing/training data set strategy, to assess the robustness of the analysis to limited sample size. A dataset that has enough statistical power to detect small-effect loci is also more likely to display a lower inflation of the allelic effect estimates for the larger-effect loci. It is possible that the algorithm used by FarmCPU is less efficient in removing correlated genetic effects in smaller than in larger samples, due to intrinsically lower power, and as a consequence a subsample of the original dataset will detect more signals than the larger one, though several of them will be false positives. To minimise the signals resulting from the

random sampling we analysed multiple training datasets and tagged as “consensus loci” only the SNPs that appear more than in two of them. In this way, if a locus has been retrieved only in a single training datasets because of its idiosyncratic sample composition, it will lack the consensus to be included in the testing set. More training datasets results in more loci being retrieved, while a higher threshold for the consensus enables a better discrimination between true and false positives. It is to be noted that applying a high consensus threshold over many training datasets is likely to lead to a steep decrease in the number of consensus loci: because of the different sampling composition, the training datasets are likely to converge around very few or no loci at all. It’s up to the investigators to select a combination of the number of training datasets and threshold of consensus.

Genomic architecture of beech spring phenology

Inferred genomic architecture

The complete segregation between the QTLs detected in the two years suggests that genotype x environment interactions are widespread (as suggested as well by the changes in the ranks in event dates). Hence, we will focus more on the trends in the architecture of a set of traits rather than on the actual sets of underlying loci. Completely distinct sets of loci controlling spring phenology in 2022 and 2023 have been retrieved in a GWAS experiment by Lazic *et al*. (2024) on a larger sample size with common garden-grown provenances from the whole species’s range. Since in Lazic et al. (2024) phenotypes were correlated between years, it is possible that different suites of genes regulate TBB depending on the climatic conditions of each year.

A suite of seven QTLs underlie spring phenology in 2009 with few others associated with the individual date of passage: Bhaga_1_7659492, Bhaga_3_8690855, Bhaga_4_10083900, Bhaga_5_44718895, Bhaga_6_11825186, Bhaga_6_11847868, Bhaga_12_24859600. The central phase of the spring phenological transition appears to be coordinated by a core set of genes which determine the date of transition of these phases together with phase-specific genes. We did retrieve few phase-specific genes (BBCH 07: Bhaga_6_885681, Bhaga_10_9729709, Bhaga_12_7940839, Bhaga_12_24085234; BBCH 09: Bhaga_1_49253423, Bhaga_2_45895074).

A subset of three loci from this suite is also shared with DegreeDay.7_2009 and Logistic.b_2009. Bhaga_6_11847868 is one of the seven genes underlying the dates of passage in 2009; the QTL is upstream of ANTI-SILENCING 1-like isoform, a histone chaperonin that participates in the induction of transcription under thermal stress. It’s possible that its heat-modulated activation also works to induce transcription in other temperature-mediated processes (Liu *et al.*, 2015; Jin *et al.*, 2024). Another QTL of these seven (Bhaga_4_10083900) is upstream of the gene coding for a calcineurin B-like protein, which mediates Ca^{2+} signalling pathways (Luan *et al.*, 2002) and possibly participates in the orchestration of the molecular switches activated during leaf unfolding. Bhaga_1_49253423 (associated with JDoY.9_2009) falls within an intron of the cytochrome P450, possibly modulating its expression during the final phase of the bud burst. This protein is a monooxygenase that, by forming complexes, takes part into a number of developmental processes, like synthesis of the cell wall, cell division and expansion (Minerdi *et al.*, 2023). Annotations are reported in **Supplementary Table 2**.

While in 2009 Logistic.a stands apart from the other traits, in 2010 it shares a QTL with JDoY.7, JDoY.9, and DegreeDay.9. The temperature requirement for the investigated

phenological phases have distinct architectures with a degree of overlap - these QTLs are also shared by the date of passage. Globally, the rate of change of the phenological transition appears genomically disjointed from the other phenological traits, and is likely tied to other physiological processes that underlie the spring transition in the tree. On the other side, many QTLs of the suite of seven are pleiotropic to the timing of spring phenology; in 2010, the sole QTL associated with Logistic.b is in fact pleiotropic to most of the other traits.
We found no overlap with the loci identified by Lazic *et al.* (2024) nor with loci identified in another population phenotyped in 2022 and 2023 (La Massane; Lauren Clément, personal communication) or across multiple years on larger sample size (Niels Müller, personal communication). In these cases, different biogeographical origins may have had a role in the distinct signals identified by this group of studies.

Genomic distribution

More than half of the QTLs we identified were mapped in non-coding regions, and most of them at less than 5 kbp from a gene; in particular, 18 QTLs were upstream of a gene and, in three cases, of two genes, one on each side of the QTL along the chromosome. Eleven QTLs fell within introns.
With its 541 Mb (Mishra *et al.*, 2022), European beech has a genome size in line with the mode of the angiosperms (0.588 Gb/1C; Pellicer *et al.*, 2018; Soto Gomez *et al.*, 2024). Due to differences in sheer total DNA length between coding and non-coding regions, the majority of mutations are found in non-coding regions and, whilst most of them have a low probability to affect the phenotype (Mei *et al.*, 2018), a small proportion of them has been reported to be functional (Lloyd *et al.*, 2019). Intergenic regions also harbour a variety of binding sites for transcription factors or methylation targets that regulate either the quantity (promoters) or the timing of transcription (enhancers/suppressors) (Weber *et al.*, 2016; Marand *et al.*, 2023); even if understudied, there is evidence of their activity (Lozano *et al.*, 2021) though the affected genes have not been yet identified. Intron can play an even wider role on the phenotypes thanks to alternative splicing, which can either generate tissue-specific proteins (Martín *et al.*, 2021) or regulate transcription (Zhang *et al.*, 2015; Wright *et al.*, 2022). Already in 2010, Derory *et al.* hypothesised that intergenic regions could account for a sizable fraction of the genetic variants that affect the TBB: following that, both McKown (2014) and Evans *et al.* (2014) found that most of the QTLs underlying the architecture of bud burst and leaf fall in *Populus trichocarpa* Torr. & Gray was harboured within intronic or intergenic regions. Introns and noncoding regions can interact with each other as well, in example by miRNA interference, where a RNA transcript produced by an intergenic region binds an untranslated (intronic) mRNA region, allowing the latter transcript quantity to be regulate in specific tissues (Meng *et al.*, 2013; Stepien *et al.*, 2017; Michlewski and Cáceres, 2019).The findings we report in this paper suggests that variation at intergenic sites in *F. sylvatica* is likely to have strong phenotypic effects.

Pattern of allelic effects

The allelic effects estimated in the training datasets are inflated by a factor that ranges between 2 and 10 relative to the full sample size (**Supplementary Figure 2-3**). Effect size varied among loci associated with a trait, with a smallest-to-largest effect ratio smaller than 5; for some traits we observed many loci with relatively smaller effects and one with a way larger effect. We tested whether the effects of loci associated with a trait both in the training datasets and in the main datasets were affected by the Beavis effect by the linear regression $Effects_{training\ datasets} \sim Effects_{main\ datasets}$. Among the five traits in which it was possible to assess a relationship, it was possible to observe a lack of effect overestimation in two cases (JDoY.7_2009 and Logistic.b_2009, with intercept of 0.97 and 1.24 respectively). Numbers are too small to assume we have identified a major part of the trait architectures if we are to expect them to conform to the infinitesimal model (Neale and Kremer, 2011; Rockman, 2012). We stand on the fact that the main datasets contained enough phenotypic diversity to correctly identify true phenotype-genotype associations.

Perspectives

The accuracy of the GWAS can be improved by including environmental covariables that can cause environmental noise on the trait, as well as cryptic population structure that may evolve even within a given stand. More importantly, an extensive sampling can both identify more loci of minor effect and tackle the overestimation of their effect; while a GO enrichment analysis would pinpoint which biological functions the QTLs are mostly known for. We do not discard the possibility that many loci affecting phenology may be rather more known for being involved in other processes, especially physiological - as suggested by the general small overlap between the QTL for the rate of change and the others. An ideal cross validation would involve a second population of European beech on which to estimate the breeding values with the effects estimated in the one presented in this study.

These results provide a starting point to understand the pattern of selection across the slope of Mont Ventoux by assessing the level of between-population differentiation for the identified QTLs. Further modeling can test the evolvability of TBB under different scenarios of climate warming, migration rate, and, possibly, ascertaining the respective role of phenotypic plasticity and current standing genetic variation in these contexts.

References

Arnold P, Kruuk, Loeske E. B., Nicotra, Adrienne B (2019). How to analyse plant phenotypic plasticity in response to a changing climate. *New Phytol*: 1235–1241.

Badeck F, Bondeau A, Böttcher K, Doktor D, Lucht W, Schaber J, *et al.* (2004). Responses of spring phenology to climate change. *New Phytol* **162**: 295–309.

Bontemps A, Davi H, Lefèvre F, Rozenberg P, Oddou-Muratorio S (2017). How do functional traits syndromes covary with growth and reproductive performance in a water-stressed population of Fagus sylvatica? *Oikos* **126**: 1472–1483.

Bontemps A, Klein EK, Oddou-Muratorio S (2013). Shift of spatial patterns during early recruitment in Fagus sylvatica: Evidence from seed dispersal estimates based on genotypic data. *For Ecol Manag* **305**: 67–76.

Browning BL, Zhou Y, Browning SR (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet* **103**: 338–348.

Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, *et al.* (2016). JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol* **17**: 66.

Caffarra A, Donnelly A (2011). The ecological significance of phenology in four different tree species: Effects of light and temperature on bud burst. *Int J Biometeorol* **55**: 711–721.

Capblancq T, Morin X, Gueguen M, Renaud J, Lobreaux S, Bazin E (2020). Climate-associated genetic variation in Fagus sylvatica and potential responses to climate change in the French Alps. *J Evol Biol* **33**: 783–796.

Choat B, Brodribb TJ, Brodersen CR, Duursma RA, López R, Medlyn BE (2018). Triggers of tree mortality under drought. *Nature* **558**: 531–539.

Chuine I, Beaubien EG (2001). Phenology is a major determinant of tree species range. *Ecol Lett* **4**: 500–510.

Conway JR, Lex A, Gehlenborg N (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**: 2938–2940.

Csilléry K, Lalagüe H, Vendramin GG, González-Martínez SC, Fady B, Oddou-Muratorio S (2014). Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (Fagus sylvatica) populations. *Mol Ecol* **23**: 4696–4708.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, *et al.* (2011). The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, *et al.* (2021). Twelve years of SAMtools and BCFtools. *GigaScience* **10**: giab008.

Davi H, Gillmann M, Ibanez T, Cailleret M, Bontemps A, Fady B, *et al.* (2011). Diversity of leaf unfolding dynamics among tree species: New insights from a study along an altitudinal gradient. *Agric For Meteorol* **151**: 1504–1513.

De Villemereuil P, Gaggiotti OE, Mouterde M, Till-Bottraud I (2016). Common garden experiments in the genomic era: new perspectives and opportunities. *Heredity* **116**: 249–254.

Derory J, Scotti-Saintagne C, Bertocchi E, Le Dantec L, Graignic N, Jauffres A, *et al.* (2010). Contrasting relations between diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. *Heredity* **105**: 401–411.

Duan X, An B, Du L, Chang T, Liang M, Yang B-G, *et al.* (2021). Genome-Wide Association Analysis of Growth Curve Parameters in Chinese Simmental Beef Cattle. *Animals* **11**: 192.

Ekblom R, Galindo J (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**: 1–15.

Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, *et al.* (2014). Population genomics of Populus trichocarpa identifies signatures of selection and adaptive trait associations. *Nat Genet* **46**: 1089–1096.

Finn GA, Straszewski AE, Peterson V (2007). A general growth stage key for describing trees and woody plants. *Ann Appl Biol* **151**: 127–131.

Frank A, Pluess AR, Howe GT, Sperisen C, Heiri C (2017). Quantitative genetic differentiation and phenotypic plasticity of European beech in a heterogeneous landscape: Indications for past climate adaptation. *Perspect Plant Ecol Evol Syst* **26**: 1–13.

Gaüzere J, Delzon S, Davi H, Bonhomme M, Garcia de Cortazar-Atauri I, Chuine I (2017). Integrating interactive effects of chilling and photoperiod in phenological process-based models. A case study with two European tree species: Fagus sylvatica and Quercus petraea. *Agric For Meteorol* **244–245**: 9–20.

Gaüzere J, Klein EK, Brendel O, Davi H, Oddou-Muratorio S (2020). Microgeographic adaptation and the effect of pollen flow on the adaptive potential of a temperate tree species. *New Phytol* **227**: 641–653.

Gaüzere J, Klein EK, Oddou-Muratorio S (2013). Ecological determinants of mating system within and between three Fagus sylvatica populations along an elevational gradient. *Mol Ecol* **22**: 5001–15.

Gaüzere J, Teuf B, Davi H, Chevin L, Caignard T, Leys B, *et al.* (2020). Where is the optimum? Predicting the variation of selection along climatic gradients and the adaptive value of plasticity. A case study on tree phenology. *Evol Lett* **4**: 109–123.

Gel B, Serra E. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics, 31–33. doi:10.1093/bioinformatics/btx346

Gienapp P, Fior S, Guillaume F, Lasky JR, Sork VL, Csilléry K (2017). Genomic Quantitative Genetics to Study Evolution in the Wild. *Trends Ecol Evol* **32**: 897–908.

Gömöry D, Paule L (2011). Trade-off between height growth and spring flushing in common beech (Fagus sylvatica L.). *Ann For Sci* **68**: 975–984.

González-Martínez SC, Krutovsky KV, Neale DB (2006). Forest-tree population genomics and adaptive evolution. *New Phytol* **170**: 227–238.

Grolemund G, Wickham H (2011). Dates and Times Made Easy with **lubridate**. *J Stat Softw* **40**.

Hanewinkel M, Cullmann DA, Schelhaas M-J, Nabuurs G-J, Zimmermann NE (2013). Climate change may cause severe loss in the economic value of European forest land. *Nat Clim Change* **3**: 203–207.

Heide OM (1993). Daylength and thermal time responses of budburst during dormancy release in some northern deciduous trees. *Physiol Plant* **88**: 531–540.

Hickler T, Vohland K, Feehan J, Miller PA, Smith B, Costa L, *et al.* (2012). Projecting the future distribution of European potential natural vegetation zones with a generalized, tree species‐based dynamic vegetation model. *Glob Ecol Biogeogr* **21**: 50–63.

Howe GT, Saruul P, Davis J, Chen THH (2000). Quantitative genetics of bud phenology, frost damage, and winter survival in an F2 family of hybrid poplars. *Theor Appl Genet* **101**: 632–642.

Jean F, Davi H, Oddou-Muratorio S, Fady B, Scotti I, Scotti-Saintagne C, *et al.* (2023). A 14-year series of leaf phenological data collected for European beech (Fagus sylvatica L.) and silver fir (Abies alba Mill.) from their geographic range margins in south-eastern France. *Ann For Sci* **80**: 35.

Jensen TS (1985). Seed-Seed Predator Interactions of European Beech, Fagus silvatica and Forest Rodents, Clethrionomys glareolus and Apodemus flavicollis. *Oikos* **44**: 149.

Jin Q, Chachar M, Ali A, Chachar Z, Zhang P, Riaz A, *et al.* (2024). Epigenetic Regulation for Heat Stress Adaptation in Plants: New Horizons for Crop Improvement under Climate Change. *Agronomy* **14**: 2105.

Johnson LC, Galliart MB, Alsdurf JD, Maricle BR, Baer SG, Bello NM, *et al.* (2022). Reciprocal transplant gardens as gold standard to detect local adaptation in grassland species: New opportunities moving into the 21st century. *J Ecol* **110**: 1054–1071.

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, *et al.* (2008). Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics* **178**: 1709–1723.

Kawecki TJ, Ebert D (2004). Conceptual issues in local adaptation. *Ecol Lett* **7**: 1225–1241.

Krajmerová D, Hrivnák M, Ditmarová Ľ, Jamnická G, Kmeť J, Kurjak D, *et al.* (2017). Nucleotide polymorphisms associated with climate, phenology and physiological traits in European beech (Fagus sylvatica L.). *New For*: 1–15.

Kramer K, Leinonen I, Loustau D (2000). The importance of phenology for the evaluation of impact of climate change on growth of boreal, temperate and Mediterranean forests ecosystems: an overview. *Int J Biometeorol* **44**: 67–75.

Laube J, Sparks TH, Estrella N, Höfler J, Ankerst DP, Menzel A (2014). Chilling outweighs photoperiod in preventing precocious spring development. *Glob Change Biol* **20**: 170–182.

Lazic D, Geßner C, Liepe KJ, Lesur-Kupin I, Mader M, Blanc-Jolivet C, *et al.* (2024a). Genomic variation of European beech reveals signals of local adaptation despite high levels of phenotypic plasticity. *Nat Commun* **15**: 8553.

Lazic D, Geßner C, Liepe KJ, Lesur-Kupin I, Mader M, Blanc-Jolivet C, *et al.* (2024b).

Genomic variation of European beech reveals signals of local adaptation despite high levels of phenotypic plasticity. *Nat Commun* **15**: 8553.

Lazic D, Geßner C, Liepe KJ, Lesur-Kupin I, Mader M, Blanc-Jolivet C, *et al.* (2024c). Genomic variation of European beech reveals signals of local adaptation despite high levels of phenotypic plasticity. *Nat Commun* **15**: 8553.

Li H (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.

Liu J, Feng L, Li J, He Z (2015). Genetic and epigenetic control of plant heat responses. *Front Plant Sci* **06**.

Liu X, Huang M, Fan B, Buckler E, Zhang Z (2012). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS Genet* **1**: e1005767.

Lloyd JP, Bowman MJ, Azodi CB, Sowers RP, Moghe GD, Childs KL, *et al.* (2019). Evolutionary characteristics of intergenic transcribed regions indicate rare novel genes and widespread noisy transcription in the Poaceae. *Sci Rep* **9**: 12122.

Lozano R, Booth GT, Omar BY, Li B, Buckler ES, Lis JT, *et al.* (2021). RNA polymerase mapping in plants identifies intergenic regulatory elements enriched in causal variants (D-J De Koning, Ed.). *G3 GenesGenomesGenetics* **11**: jkab273.

Luan S, Kudla J, Rodriguez-Concepcion M, Yalovsky S, Gruissem W (2002). Calmodulins and Calcineurin B–like Proteins: Calcium Sensors for Specific Signal Response Coupling in Plants. *Plant Cell* **14**: S389–S400.

Magri D (2008). Patterns of post-glacial spread and the extent of glacial refugia of European beech (Fagus sylvatica). *J Biogeogr* **35**: 450–463.

Marand AP, Eveland AL, Kaufmann K, Springer NM (2023). *cis* -Regulatory Elements in Plant Development, Adaptation, and Evolution. *Annu Rev Plant Biol* **74**: 111–137.

Martín G, Márquez Y, Mantica F, Duque P, Irimia M (2021). Alternative splicing landscapes in Arabidopsis thaliana across tissues and stress conditions highlight major functional differences with animals. *Genome Biol* **22**: 35.

Mckown AD, Klápště J, Guy RD, Geraldes A, Porth I, Hannemann J, *et al.* (2014). Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa. *New Phytol* **203**: 535–553.

Meger J, Ulaszewski B, Burczyk J (2021). Genomic signatures of natural selection at phenology-related genes in a widely distributed tree species Fagus sylvatica L. *BMC Genomics* **22**: 583.

Mei W, Stetter MG, Gates DJG, Stitzer MC, Ross-Ibarra J (2018). Adaptation in plant genomes: Bigger is different. *Am J Bot*.

Meier U (2001). Growth stages of mono-and dicotyledonous plants: BBCH Monograph, 2nd ed. *Fed Biol Res Cent Agric For* **49(2)**: 66–70.

Meng Y, Shao C, Ma X, Wang H (2013). Introns targeted by plant microRNAs: a possible

novel mechanism of gene regulation. *Rice* **6**: 8.

Michlewski G, Cáceres JF (2019). Post-transcriptional control of miRNA biogenesis. *RNA* **25**: 1–16.

Milesi P, Kastally C, Dauphin B (2023). Synchronous effective population size changes and genetic stability of forest trees through glacial cycles Synchronous effective population size changes and genetic stability of forest trees through glacial cycles.

Minerdi D, Savoi S, Sabbatini P (2023). Role of Cytochrome P450 Enzyme in Plant Microorganisms' Communication: A Focus on Grapevine. *Int J Mol Sci* **24**: 4695.

Mishra B, Gupta DK, Pfenninger M, Hickler T, Langer E, Nam B, *et al.* (2018). A reference genome of the European beech (Fagus sylvatica L.). *GigaScience* **7**.

Mishra B, Ulaszewski B, Meger J, Aury J-M, Bodénès C, Lesur-Kupin I, *et al.* (2022). A Chromosome-Level Genome Assembly of the European Beech (Fagus sylvatica) Reveals Anomalies for Organelle DNA Integration, Repeat Content and Distribution of SNPs. *Front Genet* **12**.

Müller M, Seifert S, Finkeldey R (2015). A candidate gene-based association study reveals SNPs significantly associated with bud burst in European beech (Fagus sylvatica L.). *Tree Genet Genomes* **11**.

Müller M, Seifert S, Finkeldey R (2017). Comparison and confirmation of SNP-bud burst associations in European beech populations in Germany. *Tree Genet Genomes* **13**: 59.

Neale DB, Kremer A (2011). Forest tree genomics: growing resources and applications. *Nat Rev Genet* **12**: 111–122.

Nicotra AB, Atkin OK, Bonser SP, Davidson AM, Finnegan EJ, Mathesius U, *et al.* (2010). Plant phenotypic plasticity in a changing climate. *Trends Plant Sci* **15**: 684–692.

Oddou-Muratorio S, Bontemps A, Gaüzere J, Klein EK (2024). Interplay between fecundity, sexual and growth selection on the spring phenology of European beech (Fagus sylvatica L.). *Peer Community J* **4**: e27.

Oddou-Muratorio S, Gaüzere J, Bontemps A, Rey J-F, Klein EK (2018). Tree, sex and size: Ecological determinants of male vs. female fecundity in three *Fagus sylvatica* stands. *Mol Ecol* **27**: 3131–3145.

Osada N, Murase K, Tsuji K, Sawada H, Nunokawa K, Tsukahara M, *et al.* (2018). Genetic differentiation in the timing of budburst in Fagus crenata in relation to temperature and photoperiod. *Int J Biometeorol* **62**: 1763–1776.

Otto SP, Jones CD (2000). Detecting the Undetected: Estimating the Total Number of Loci Underlying a Quantitative Trait. *Genetics* **156**: 2093–2107.

Pellicer J, Hidalgo O, Dodsworth S, Leitch I (2018). Genome Size Diversity and Its Impact on the Evolution of Land Plants. *Genes* **9**: 88.

Pfenninger M, Reuss F, Kiebler A, Schönnenbeck P, Caliendo C, Gerber S, *et al.* (2021). Genomic basis for drought resistance in european beech forests threatened by climate change. *eLife* **10**.

Piotti A, Leonardi S, Buiteveld J, Geburek T, Gerber S, Kramer K, *et al.* (2012). Comparison of pollen gene flow among four European beech (Fagus sylvatica L.) populations characterized by different management regimes. *Heredity* **108**: 322–331.

Pluess AR, Frank A, Heiri C, Lalagüe H, Vendramin GG, Oddou-Muratorio S (2016). Genome-environment association study suggests local adaptation to climate at the regional scale in Fagus sylvatica. *New Phytol* **210**: 589–601.

Pluess AR, Weber P (2012). Drought-adaptation potential in *Fagus sylvatica*: linking moisture availability with genetic diversity and dendrochronology. *PloS One* **7**: e33636.

Postolache D, Oddou-Muratorio S, Vajana, Bagnoli F, Guichoux E, Hampe A, *et al.* (2021). Genetic signatures of divergent selection in European beech (Fagus sylvatica L.) are associated with the variation in temperature and precipitation across its distribution range. *Mol Ecol*: mec.16115.

Pritchard JK, Di Rienzo A (2010). Adaptation - Not by sweeps alone. *Nat Rev Genet* **11**: 665–667.

Rathcke B, Lacey EP (1985). Phenological patterns of terrestrial plants. *Annu Rev Ecol Evol Syst* **16**: 179–214.

Rellstab C, Fischer MC, Zoller S, Graf R, Tedder A, Shimizu KK, *et al.* (2017). Local adaptation (mostly) remains local: Reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. *Heredity* **118**: 193–201.

Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends Ecol Evol* **29**: 165–176.

Rockman MV (2012). The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution* **66**: 1–17.

Santure AW, Garant D (2018). Wild GWAS—association mapping in natural populations. *Mol Ecol Resour* **18**: 729–738.

Savolainen O, Lascoux M, Merilä J (2013). Ecological genomics of local adaptation. *Nat Rev Genet 2013 1411* **14**: 807–820.

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, *et al.* (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet 2012 447* **44**: 825–830.

Smith TJ, McKenna CM (2013). A Comparison of Logistic Regression Pseudo R2 Indices. **39**.

Soto Gomez M, Brown MJM, Pironon S, Bureš P, Verde Arregoitia LD, Veselý P, *et al.* (2024). Genome size is positively correlated with extinction risk in herbaceous angiosperms. *New Phytol* **243**: 2470–2485.

Stepien A, Knop K, Dolata J, Taube M, Bajczyk M, Barciszewska-Pacak M, *et al.* (2017). Posttranscriptional coordination of splicing and MIRNA biogenesis in plants. *WIREs RNA* **8**: e1403.

Storey JD, Tibshirani R (2003). Statistical significance for genomewide studies. *Proc Natl*

Acad Sci **100**: 9440–9445.

Tjørve KMC, Tjørve E (2017). The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family (RMH Merks, Ed.). *PLOS ONE* **12**: e0178691.

Veall MR, Zimmermann KF (1996). Pseudo-R2 Measures for Some Common Limited Dependent Variable Models. *J Econ Surv* **10**: 241–259.

Vitasse Y, Porté AJ, Kremer A, Michalet R, Delzon S (2009). Responses of canopy duration to temperature changes in four temperate tree species: relative contributions of spring and autumn leaf phenology. *Oecologia* **161**: 187–198.

Wang Q, Tian F, Pan Y, Buckler ES, Zhang Z (2014). A SUPER Powerful Method for Genome Wide Association Study (Y Li, Ed.). *PLoS ONE* **9**: e107684.

Weber B, Zicola J, Oka R, Stam M (2016). Plant Enhancers: A Call for Discovery. *Trends Plant Sci* **21**: 974–987.

Winsor CP (1932). The Gompertz Curve as a Growth Curve. *Proc Natl Acad Sci* **18**: 1–8.

Wright CJ, Smith CWJ, Jiggins CD (2022). Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet* **23**: 697–710.

Xu S (2003). Theoretical Basis of the Beavis Effect. *Genetics* **165**: 2259–2268.

Yin L (2021). rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. *Genomics Proteomics Bioinformatics*.

Zhang C, Yang H, Yang H (2015). Evolutionary Character of Alternative Splicing in Plants. *Bioinforma Biol Insights* **9s1**: BBI.S33716.

Supplementary Materials

Supplementary Table 1: GWAS results

Table 2 Table reports: **FDR 5%**: significant SNPs found in the dataset at FDR 5 %; **Found once**: SNPs retrieved only once among all significant SNPs retrieved across the training datasets; **Found 2+**: SNPs retrieved at least twice among all significant SNPs retrieved across the training datasets; **SNPs in clusters**: SNPs retrieved at least twice that mapped into a cluster (distance from next SNP < mean scaled chromosomal distance) ; **Clustered peaks**: n° of association signals tied to a cluster of SNPs; **Isolated peaks**: SNPs retrieved at least twice that mapped at a distance from next SNP > mean scaled chromosomal distance; **Final**: Signals from clusters + signals from isolated SNPs; **R^2 validation**: prediction of the phenotype of the training dataset using the SNPs retrieved in the training datasets.

BBCH 07

Dataset	Year	FDR 5%	Found once	Found 2+	SNPs in clusters	Clustered peaks	Isolated peaks	Final	R^2 validation
Training 1	2009	2,469	2,863	2,648	2,514	122	134	256	0.39
Training 2	2009	3							0.64
Training 3	2009	2,558							0.17
Training 4	2009	51							0.530
Training 5	2009	1009							0.48
Training 6	2009	2,558							0.17
Training 7	2009	0							0.56
Training 8	2009	147							0.55
Main 2009	2009	12			2	1	11	12	
Training 9	2010	9	18,245	26	2	1	24	25	0.65
Training 10	2010	6							0.53
Training 11	2010	7							0.61
Training 12	2010	5							0.63
Training 13	2010	7							0.71
Training 14	2010	7							0.61
Training 15	2010	10							0.570
Training 16	2010	18,250							0.740
Main 2010	2010	5			0	0	5	5	

BBCH 09

Dataset	Year	FDR 5%	Found once	Found 2+	SNPs in clusters	Clustered peaks	Isolated peaks	Final	R^2 validation
Training 1	2009	5,844	5,331	5,500	2,206	245	294	539	0.38
Training 2	2009	3							0.65
Training 3	2009	5,277							0.18
Training 4	2009	7							0.500
Training 5	2009	2,007							0.490
Training 6	2009	5,277							0.18
Training 7	2009	0							0.62
Training 8	2009	5							0.51
Main 2009	2009	8			0	0	8	8	
Training 9	2010	7	30,802	7,908	7,529	366	379	25	0.67
Training 10	2010	2							0.62
Training 11	2010	2							0.68
Training 12	2010	9							0.65
Training 13	2010	29,297							0.60
Training 14	2010	2							0.68
Training 15	2010	2							0.65
Training 16	2010	17,308							0.53
Main 2010	2010	3			0	0	3	.3	

Logistic a

Dataset	Year	FDR 5%	Found once	Found 2+	SNPs in clusters	Clustered peaks	Isolated peaks	Final	R^2 validation
Training 1	2009	5,753	5,884	214	182	21	32	53	0.57
Training 2	2009	0							0.62
Training 3	2009	6							0.56
Training 4	2009	7							0.57
Training 5	2009	5							0.58
Training 6	2009	6							0.56
Training 7	2009	540							0.43
Training 8	2009	2							0.54
Main 2009	2009	4			0	0	4	4	
Training 9	2010	10,599	37,256	38,379	36,126	1,862	2,253	4,115	0.54
Training 10	2010	3							0.59
Training 11	2010	28,536							0.45
Training 12	2010	4							0.67
Training 13	2010	21,170							0.63
Training 14	2010	28,536							0.45
Training 15	2010	32,879							0.52
Training 16	2010	19,667							0.55
Main 2010	2010	4			0	0	4	4	

Logistic b

Dataset	Year	FDR 5%	Found once	Found 2+	SNPs in clusters	Clustered peaks	Isolated peaks	Final	R^2 validation
Training 1	2009	265	285	604	556	37	48	85	0.42
Training 2	2009	0							0.56
Training 3	2009	604							0.066
Training 4	2009	0							0.46
Training 5	2009	0							0.41
Training 6	2009	604							0.067
Training 7	2009	30							0.39
Training 8	2009	3							0.48
Main 2009	2009	8			0	0	8	8	
Training 9	2010	10	16,684	21	3	2	18	20	0.55
Training 10	2010	7							0.59
Training 11	2010	7							0.45
Training 12	2010	6							0.67
Training 13	2010	6							0.63
Training 14	2010	7							0.45
Training 15	2010	4							0.53
Training 16	2010	16,684							0.55
Main 2010	2010	1			0	0	1	1	

PSS

Dataset	Year	FDR 5%	Found once	Found 2+	SNPs in clusters	Clustered peaks	Isolated peaks	Final	R^2 validation
Training 1	2009	882	1,492	1,129	1,060	67	69	136	0.46
Training 2	2009	0							0.56
Training 3	2009	1,116							0.15
Training 4	2009	0							0.53
Training 5	2009	746							0.47
Training 6	2009	1,116							0.15
Training 7	2009	37							0.46
Training 8	2009	3							0.54
Main 2009	2009	4			0	0	4	4	
Training 9	2010	25	17,949	18	0	0	18	18	0.59
Training 10	2010	6							0.45
Training 11	2010	2							0.64
Training 12	2010	6							0.6
Training 13	2010	7							0.67
Training 14	2010	2							0.64
Training 15	2010	2							0.56
Training 16	2010	17,936							0.72
Main 2010	2010	3			0	0	3	3	

DegreeDays 7

Dataset	Year	FDR 5%	Found once	Found 2+	SNPs in clusters	Clustered peaks	Isolated peaks	Final	R^2 validation
Training 1	2009	1,916	1808	662	599	42	63	105	0.35
Training 2	2009	0							0.66
Training 3	2009	659							0.14
Training 4	2009	5							0.500
Training 5	2009	3							0.46
Training 6	2009	659							0.14
Training 7	2009	9							0.56
Training 8	2009	6							0.66
Main 2009	2009	7			0	0	7	7	
Training 9	2010	7	32,653	7,036	6,723	304	313	617	0.67
Training 10	2010	6							0.64
Training 11	2010	6							0.62
Training 12	2010	32,076							0.57
Training 13	2010	14,632							0.55
Training 14	2010	6							0.62
Training 15	2010	7							0.64
Training 16	2010	0							0.63
MD.2010	2010	3			0	0	3	3	

DegreeDays 9

Dataset	Year	FDR 5%	Found once	Found 2+	SNPs in clusters	Clustered peaks	Isolated peaks	Final	R^2 validation
Training 1	2009	3,855	2,968	2,665	2,492	126	173	299	0.3
Training 2	2009	3							0.700
Training 3	2009	2,659							0.12
Training 4	2009	6							0.47
Training 5	2009	26							0.47
Training 6	2009	2,659							0.12
Training 7	2009	0							0.64
Training 8	2009	5							0.51
Main 2009	2009	5			0	0	5	5	
Training 9	2010	5	21,540	4,701	4,494	241	207	448	0.67
Training 10	2010	8							0.63
Training 11	2010	3							0.67
Training 12	2010	7							0.64
Training 13	2010	22,000							0.59
Training 14	2010	3							0.67
Training 15	2010	4							0.66
Training 16	2010	8,890							0.51
Main 2010	2010	7			0	0	7	7	

Supplementary Table 2 Table reports: **SNP name** according to the Bhaga assembly (Mishra *et al*. 2022); **Associated trait(s)**: traits that have been found associated with the SN; **Distance from genes** located less than 10 kbp away upstream, downstream, or genic region the SNP map into (whenever two genes were within 10 kbp, both are reported); **Gene/nearest gene(s)**: name of the most proximate genes; **Gene(s) description in *F. sylvatica***: name of the gene in European beech assigned by Mishra *et al*. 2022 based on homology; **Reference gene(s)**: homologous gene that allowed the characterisation of the sequence in European beech; **Ontology**: molecular function, cellular component, biological process (species the homologous gene was characterised in).

Supplementary Table 2: Annotations

SNP name	Associated trait(s)	Distance from genes	Gene/nearest gene(s)	Gene description(s) in *F. sylvatica*	Reference gene(s)	Ontology	GO terms	Enzyme
Bhaga_1_7659492	JDoY.7_2009, JDoY.9_2009, Logistic.b_2009, DegreeDays.7_2009, DegreeDays.9_2009, PSS_2009	4,294 bp upstream (-)	Bhaga_1.g805	XP_030957484.1 CST complex subunit TEN1 (-)	CJ030_MR8G020236	CST complex single-stranded DNA binding telomerase inhibitor activity telomeric DNA binding negative regulation of telomere maintenance via telomerase negative regulation of telomerase activity	GO:1990879 GO:0003697 GO:0010521 GO:0042162 GO:0032211 GO:0051974	NA
		1,510 bp downstream (-)	Bhaga_1.g806	XP_030958239.1 phospholipase D zeta 2-like isoform X1 (-)	RchiOBHm_Chr3g0456481	plasma membrane, phospholipase D activity N-acylphosphatidylethanolamine-specific phospholipase D activity phosphatidic acid biosynthetic process phospholipid catabolic process inositol lipid-mediated signaling	GO:0005886 GO:0004630 GO:0070290 GO:0006654 GO:0009395 GO:0048017	Phospholipase D
Bhaga_1_29194858	PSS_2009	7,808, bp upstream (-)	Bhaga_1.g2811	LTR/Copia XP_030957059.1 chitinase-like protein 2 (-)	NA	NA	NA	NA
Bhaga_1_44593854	Logistic.b_2009	Intron	Bhaga_1.g4336	uncharacterised RNA-coding gene (-)	NA	NA	NA	NA
Bhaga_1_44633170	DegreeDays.9_2010	9,520 bp upstream (+)	Bhaga_1.g4339	uncharacterised RNA-coding gene (+)	NA	NA	NA	NA
Bhaga_1_49253423	JDoY.9_2009	Intron	Bhaga_1.g4809	XP_030926572.1 cytochrome P450 81Q32-like	FH972_026945	binding,oxidoreductase activity	GO:0005488 GO:0016491	NA

				isoform X1 (+)				
Bhaga_1_66110114	PSS_2010, DegreeDays.7_2010	Intron	Bhaga_1.g6576	XP_030957883.1 DNA repair protein recA homolog 2, mitochondrial-like (+)	ES332_D11G386900v1	nucleotide binding DNA binding DNA metabolic process cellular response to DNA damage stimulus	GO:0000166 GO:0003677 GO:0006259 GO:0006974	NA
Bhaga_2_14512420	DegreeDays.7_2009	3,504 bp downstream (+)	Bhaga_2.g2089	XP_030966553.1 protein encore-like (+)	LOC109009862	nucleic acid binding	GO:0003676	NA
		5,612 bp upstream (+)	Bhaga_2.g2090	XP_023897521.1 probable aspartyl aminopeptidase isoform X2 (+)	CR513_30966	aminopeptidase activity, metallopeptidase activity, zinc ion binding, proteolysis	GO:0004177 GO:0008237 GO:0008270 GO:0006508	Aminopeptidase B, Prolyl aminopeptidase, Xaa-Pro aminopeptidase
Bhaga_2_45895074	JDoY.9_2009	Intron	Bhaga_2.g5032	RVW60234.1 Transposon Tf2-12 polyprotein	VITISV_026680	NA	NA	NA
Bhaga_3_8690855	JDoY.7_2009, JDoY.9_2009	2,033 bp downstream (-)	Bhaga_3.g1038	XP_018841506.1 V-type proton ATPase subunit B 1-like (-)	ES332_A11G004800v1	proton-transporting V-type ATPase, V1 domain, ATP binding, ATP metabolic process, proton transmembrane transport	GO:0033180 GO:0005524 GO:0046034 GO:1902600	NA
Bhaga_3_17765409	Logistic.a_2009	270 bp downstream (+)	Bhaga_3.g2091	PRQ55093.1 putative nucleotidyltransferase, Ribonuclease H (+)	pol_1575	catalytic activity	GO:0003824	NA
Bhaga_3_30336349	DegreeDays.9_2010	2,004 bp downstream (-)	Bhaga_3.g3468	OMO75792.1 Reverse transcriptase, RNA-dependent DNA polymerase (-)	F0562_018326	NA	NA	NA
Bhaga_4_10083900	JDoY.7_2009, JDoY.9_2009	2,147 bp upstream (+)	Bhaga_4.g1232	XP_030947970.1 calcineurin B-like protein 9 (+)	PanWU01x14_323560	calcium ion binding	GO:0005509	NA
Bhaga_4_33547731	DegreeDays9.2009	408 bp upstream (+)	Bhaga_4.g3911	KAB1228035.1 Mediator of RNA polymerase II transcription subunit 13 (+)	FH972_019012	cytoplasm, translation elongation factor activity, protein kinase activity, calcium ion binding, ATP binding, translational elongation, protein phosphorylation	GO:0005737 GO:0003746 GO:0004672 GO:0005509 GO:0005524 GO:0006414 GO:0006468	NA
Bhaga_4_356	DegreeDays.7_2	Intron	Bhaga_4.g4163	XP_023906581.1	NA	NA	NA	NA

54050	009			enoyl-CoA hydratase 2, peroxisomal (-)				
Bhaga_5_12660371	DegreeDays.9_2010	Intron	Bhaga_5.g1428	(-)	NA	NA	NA	NA
Bhaga_5_17049977	DegreeDays.9_2010	3,355 bp upstream (-)	Bhaga_5.g1889	OIT32365.1 putative ribonuclease h protein (-)	A4A49_59360	NA	NA	NA
		4,350 upstream (+)	Bhaga_5.g1890	XP_030942234.1 polygalacturonase-like (+)	LOC109005045	extracellular region, polygalacturonase activity, carbohydrate metabolic process, cell wall organization	GO:0005576 GO:0004650 GO:0005975 GO:0071555	Endo-polygalacturonase
Bhaga_5_18794174	DegreeDays.7_2010	Intron	Bhaga_5.g2106	EOX99142.1 Immunoglobulin E-set (+)	LOC108989724	Golgi membrane, cytosol, RIC1-RGP1 guanyl-nucleotide exchange factor complex, Rab guanyl-nucleotide exchange factor activity, retrograde transport, endosome to Golgi	GO:0000139 GO:0005829 GO:0034066 GO:0017112 GO:0042147	NA
Bhaga_5_19150120	DegreeDays.7_2009	5''-UTR	Bhaga_5.g2152	5''-UTR of VVA22314.1 calcium sensing receptor (-)	NA	NA	NA	NA
Bhaga_5_27898737	DegreeDays.9_2009	380 bp downstream (+)	Bhaga_5.g3119	KAB1221678.1COP9 signalosome complex subunit 8 (+)	CFOL_v3_10871	plasma membrane, integral component of membrane	GO:0005886 GO:0016021	NA
Bhaga_5_44718895	JDoY.7_2009, JDoY.9_2009, Logistic.b_2009, PSS_2009	Intron	Bhaga_5.g5235	XP_023879717.1 protein transport protein SFT2-like	CJ030_MR4G018816	integral component of membrane, protein transport, vesicle-mediated transport	GO:0016021 GO:0015031 GO:0016192	NA
Bhaga_6_885681	JDoY.7_2009	Intron	Bhaga_6_105	PON88643.1 Aminotransferase-like mobile domain containing protein (+)	TorRG33x02_155480	NA	NA	NA
Bhaga_6_3193982	Logistic.b_2009	Exon	Bhaga_6.g364	RVX09900.1 Retrovirus-related Pol polyprotein from transposon 17.6	pol_1575	NA	NA	NA
Bhaga_6_8499851	DegreeDays.7_2010	mRNA	Bhaga_6.g989	RWR75228.1 TBC1 domain-containing protein (+)	VITISV_003620	NA	NA	NA

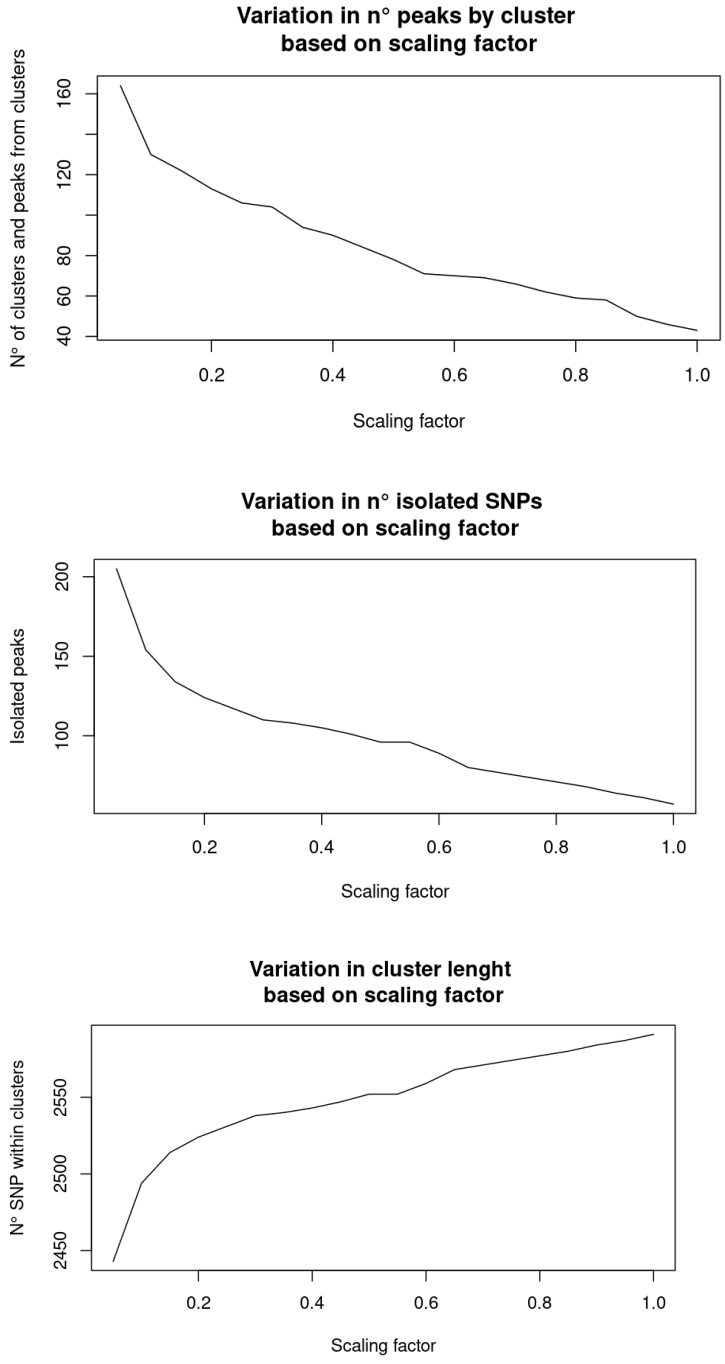
Bhaga_6_11825186	JDoY.7_2009, JDoY.9_2009, Logistic.a_2009, Logistic.b_2009	Intron	Bhaga_6.g1353	VVA36562.1 alpha/beta (+)	LOC108996369	integral component of membrane, catalytic activity	GO:0016021 GO:0003824	NA
Bhaga_6_11836592	Logistic.b_2009	Exon	Bhaga_6.g.1354	XP_028091572.1 protein ANTI-SILENCING 1-like isoform X6 (+)	TCM_028427	binding, regulation of cellular metabolic process, regulation of nitrogen compound metabolic process, regulation of macromolecule metabolic process, regulation of primary metabolic process	GO:0005488 GO:0031323 GO:0051171 GO:0060255 GO:0080090	NA
Bhaga_6_11847868	JDoY.7_2009, JDoY.9_2009	4,460 bp downstream (+)	Bhaga_6.g.1354,	XP_028091572.1 protein ANTI-SILENCING 1-like isoform X6 (+)	TCM_028427	binding, regulation of cellular metabolic process, regulation of nitrogen compound metabolic process, regulation of macromolecule metabolic process, regulation of primary metabolic process	GO:0005488 GO:0031323 GO:0051171 GO:0060255 GO:0080090	NA
		5,966 bp downstream (-)	Bhaga_6.g.1355	XP_030971839.1 serine-threonine kinase receptor-associated protein-like (-)	D5086_0000218220	SMN complex, RNA binding, kinase activity, spliceosomal snRNP assembly, phosphorylation	GO:0032797 GO:0003723 GO:0016301 GO:0000387 GO:0016310	NA
Bhaga_7_1137423	Logistic.a_2009	mRNA	Bhaga_7.g124	XP_031279372.1 alanine--tRNA ligase (+)	EZV62_023592	Mitochondrion, tRNA binding, alanine-tRNA ligase activity, ATP binding, zinc ion binding, mitochondrial alanyl-tRNA aminoacylation	GO:0005739 GO:0000049 GO:0004813 GO:0005524 GO:0008270 GO:0070143	Alanine-tRNA ligase
Bhaga_7_9394885	Logistic.b_2009	mRNA	Bhaga_7.g1080	KAB1206725.1 Transcription factor GTE3, chloroplastic (-)	NA	NA	NA	NA
Bhaga_7_10306779	PSS_2009	mRNA	Bhaga_7.g1194	XP_027163848.1 histidinol-phosphate aminotransferase, chloroplastic-like (+)	BC332_12801	NA	NA	NA
Bhaga_7_11442207	JDoY.9_2010	6,457 bp downstream (+)	Bhaga_7.g1338	KYP48096.1 Transposon Ty3-G Gag-Pol polyprotein	KK1_029954	nucleic acid binding RNA-DNA hybrid ribonuclease activity DNA integration RNA phosphodiester bond	GO:0003676 GO:0004523GO:0015074 GO:0090502	Ribonuclease H

						hydrolysis endonucleolytic		
		4,391 bp upstream (+)	Bhaga_7.1337	RVW53829.1 Transposon TX1 uncharacterized 149 kDa protein (+)	YTX2_137	intracellular anatomical structure, membrane, nucleic acid binding, transferase activity, cellular macromolecule metabolic process, regulation of response to stimulus, nucleic acid metabolic process	GO:0005622 GO:0016020 GO:0003676 GO:0016740 GO:0044260 GO:0048583 GO:0090304	NA
Bhaga_7_14128893	JDoY.7_2009	2,218 bp upstream (-)	Bhaga_7.1665	Uncharacterised RNA-coding gene (-)	NA	NA		NA
Bhaga_7_18561401	Logistic.a_2010, DegreeDays.9_2010	Intron	Bhaga_7.g2208	TKY47971.1 Glycerophosphodiester phosphodiesterase gde1 (+)	LOC109021288	integral component of membrane	GO:0016021	NA
Bhaga_7_21418349	JDoY.7_2010	mRNA	Bhaga_7.g2491	PON78534.1 Ribosomal protein (-)	PanWU01x14_017920	Ribosome	GO:0005840	NA
Bhaga_7_35021952	DegreeDays.9_2009	mRNA	Bhaga_7.g4019	XP_030972515.1 ADP-ribosylation factor GTPase-activating protein AGD5-like isoform X2 (+)	CICLE_v10015133mg	NA	NA	NA
Bhaga_7_39176789	JDoY.7_2010, PSS_2010	1,528 bp upstream (-)	Bhaga_7.g4539	KAB1213363.1 Retrovirus-related Pol polyprotein from transposon TNT 1-94 (-)	CJ030_MR5G003567	catalytic activity, cellular process, organic substance metabolic process	GO:0003824 GO:0009987 GO:0071704	NA
Bhaga_8_9082384	DegreeDays.7_2009	3,572 bp upstream (-)	Bhaga_8.g1149	PSS30657.1 Deoxyribodipyrimidine photo-lyase (-)	CEY00_Acc05942			NA
		3,371 bp upstream (-)	Bhaga_8.g1150	VVA23316.1 CISIN_1g039139mg (+)	FH972_006803	cytoplasm, microtubule, GTPase activity, structural constituent of cytoskeleton, GTP binding, cytoskeleton organization, microtubule-based process	GO:0005737 GO:0005874 GO:0003924 GO:0005200 GO:0005525 GO:0007010 GO:0007017	Nucleoside-triphosphate phosphatase

Bhaga_8_34150623	Logistic.a_2010, JDoY.7_2010, JDoY.9_2010, DegreeDays.9_2010	275 bp upstream (-)	Bhaga_8.g4081	XP_023884882.1 protein KINESIN LIGHT CHAIN-RELATED 2 (-)	TorRG33x02_297100	plasma membrane, transferase activity, pollen tube growth, regulation of defense response	GO:0005886 GO:0016740 GO:0009860 GO:0031347	NA
Bhaga_9_20825271	Logistic.a_2010, JDoY.7_2010	2,056 bp upstream (-)	Bhaga_9.g2536	KAB1219016.1 putative transporter MCH1 (+)	LOC108984543	integral component of membrane, transmembrane transporter activity, transmembrane transport	GO:0016021 GO:0022857 GO:0055085	NA
Bhaga_10_974374	Logistic.a_2009	3,246 bp upstream (-)	Bhaga_10.g122	LINE/RTE-BovB of 243 bp XP_018854097.1 E3 ubiquitin-protein ligase ATL41-like (-)	NA	NA	NA	NA
		779 bp upstream (-)	Bhaga_10.g123	XP_030955181.1 mitochondrial dicarboxylate/tricarboxylate transporter DTC (+)	NA	NA	NA	NA
Bhaga_10_9729709	JDoY.7_2009	265 bp downstream (+)	Bhaga_10.g1214	XP_030941086.1 transcription factor PIF7 (-)	RCOM_1470600	NA	NA	NA
Bhaga_10_10012648	Logistic.b_2009	956 bp upstream (-)	Bhaga_10.g1256	XP_030940478.1 SNAP25 homologous protein SNAP33 (-)	LOC109011467	plasma membrane	GO:0005886	NA
		2,421 bp downstream (-)	Bhaga_10.g1257	XP_030940807.1 intracellular protein transport protein USO1 isoform X1	NA	NA	NA	NA
Bhaga_11_15878198	PSS_2010, Logistic.a_2010, Logistic.b_2010, JDOY.7_2010, JDOY.9_2010, DegreeDays.9_2010	1,642 bp upstream (+)	Bhaga_11.g1815	LTR/Gypsy XP_018824015.1 thiamine thiazole synthase, chloroplastic-like (+)	LOC108993518	cytosol, chloroplast stroma, iron ion binding, kinase activity, transferase activity, transferring pentosyl groups, thiamine biosynthetic process, phosphorylation, thiazole biosynthetic process	GO:0005829 GO:0009570 GO:0005506GO:0016301 GO:0016763GO:0009228GO:0016310 GO:0052837	NA
Bhaga_12_7940839	JDoY.7_2009	Exon	Bhaga_12.g1015	XP_008244435.1 sec-independent protein translocase protein TatB-like isoform X1 (-)	FH972_018256	integral component of membrane	GO:0016021	NA
Bhaga_12_9983927	Logistic.b_2009	mRNA	Bhaga_12.g1251	RVW85322.1 Transposon TX1	VITISV_031269	NA	NA	NA

				uncharacterized 149 kDa protein (+)				
Bhaga_12_22708997	DegreeDays.7_2009	2,311 bp upstream (-)	Bhaga_12.g2696	KAE8676151.1 putative CCCH-type zinc finger family protein (-)	F3Y22_tig00112540pilonHSYRG00059	binding, transferase activity, metabolic process	GO:0005488 GO:0016740 GO:0008152	NA
		3,198 bp downstream (-)	Bhaga_12.g2697	XP_030961260.1 protein DMR6-LIKE OXYGENASE 1-like (-)	TorRG33x02_165340	metal ion binding, dioxygenase activity, oxidation-reduction process	GO:0046872 GO:0051213 GO:0055114	NA
Bhaga_12_24085234	JDoY.7_2009	967 bp upstream (+)	Bhaga_12.g2861	XP_023918358.1 potassium transporter 6-like (+)	TCM_010271	plasma membrane, integral component of membrane, potassium ion transmembrane transporter activity, potassium ion transmembrane transport	GO:0005886 GO:0016021 GO:0015079 GO:0071805	NA
Bhaga_12_24859600	JDoY.7_2009, JDoY.9_2009, DegreeDays.7_2009, DegreeDays.9_2009	3,314 bp downstream (+)	Bhaga_12.g2946	XP_030956337.1 transcription factor BHLH089 isoform X1 (+)	LOC108988637	nucleus, DNA-binding transcription factor activity, protein dimerization activity, regulation of transcription, DNA-templated	GO:0005634 GO:0003700 GO:0046983 GO:0006355	NA
		4,251 bp upstream (+)	Bhaga_12.g2947	XP_030957026.1 leucine zipper protein 1 isoform X1 (+)	NA	NA	NA	NA

Supplementary Figure 1: Parameters ~ Scaling factor



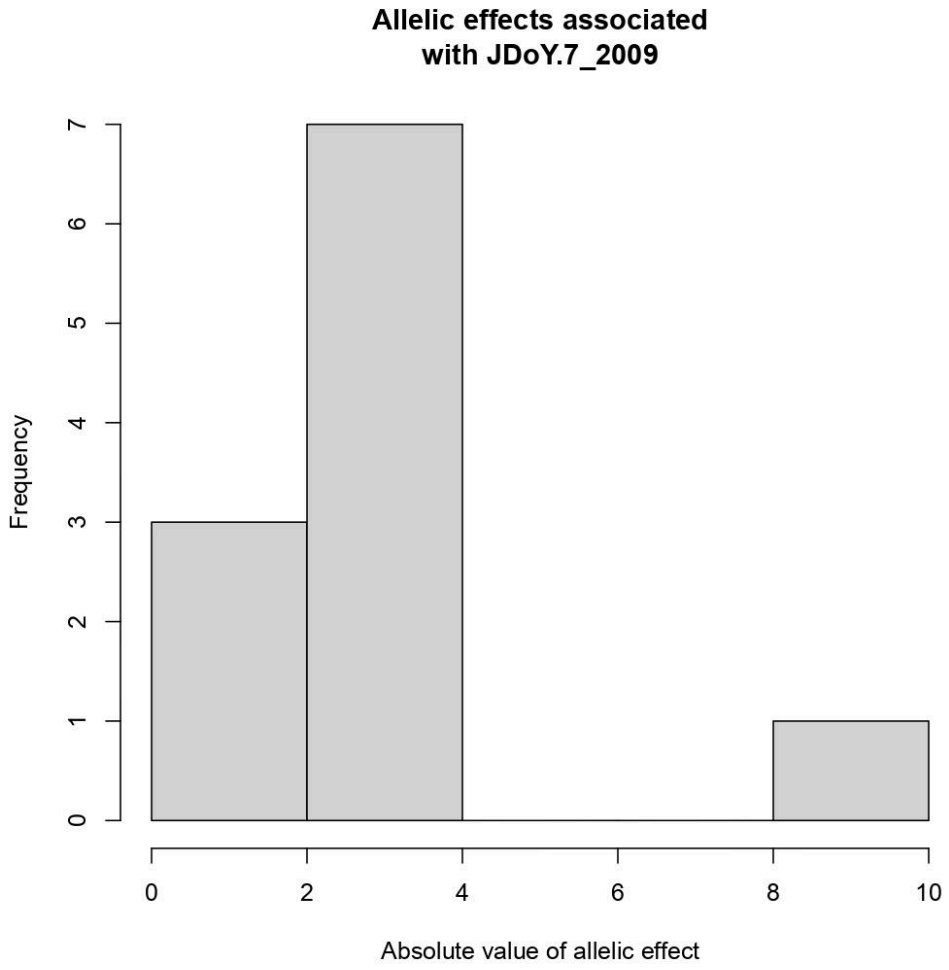
Supplementary Figure 3 Variation in the statistics of clustering of significant SNPs with increasing scaling factors. The trait represented is JDoY.7_2009.

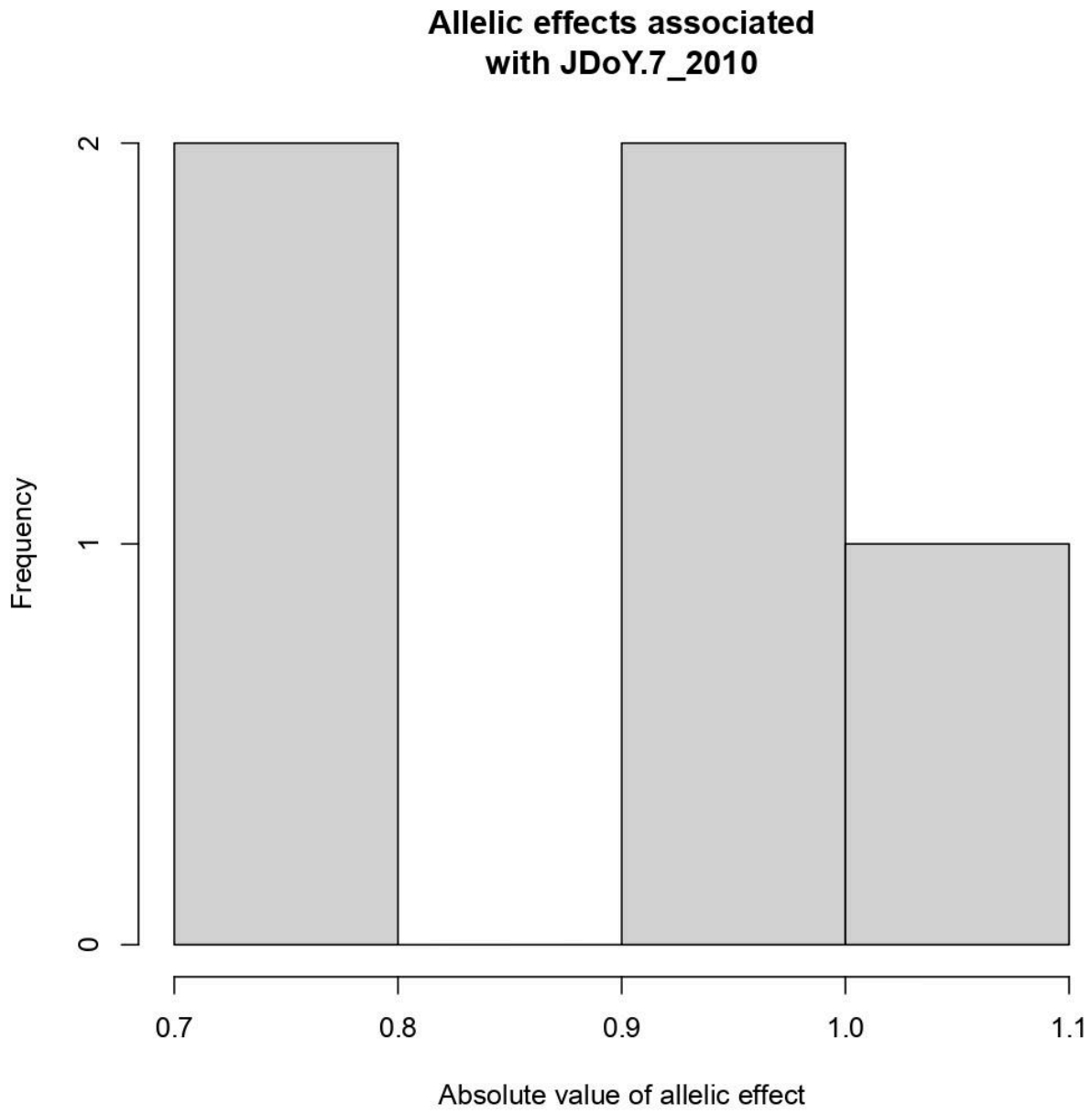
Supplementary Figure 2 : Allelic Effects Main Datasets

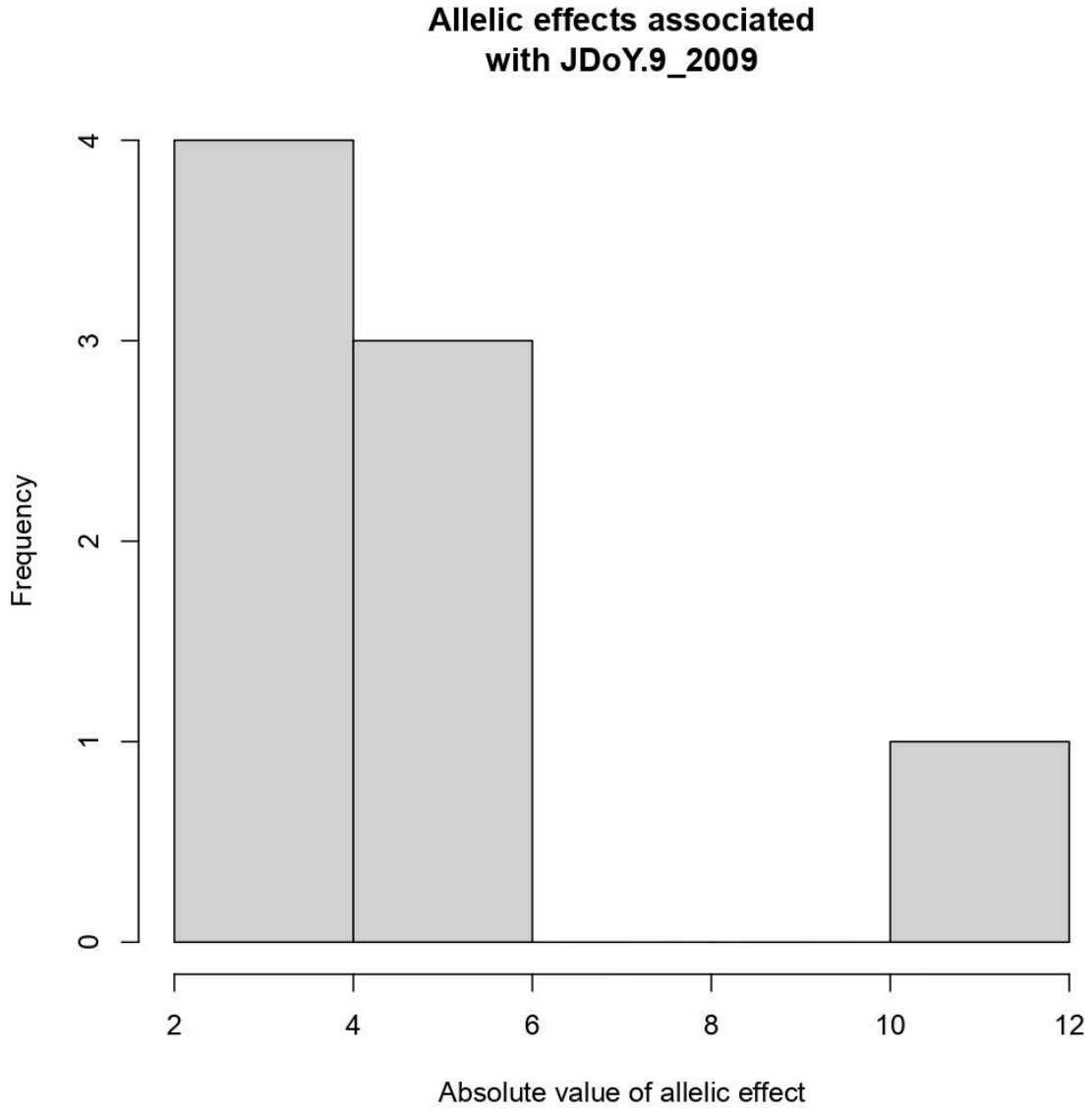
Traits are ordered as:

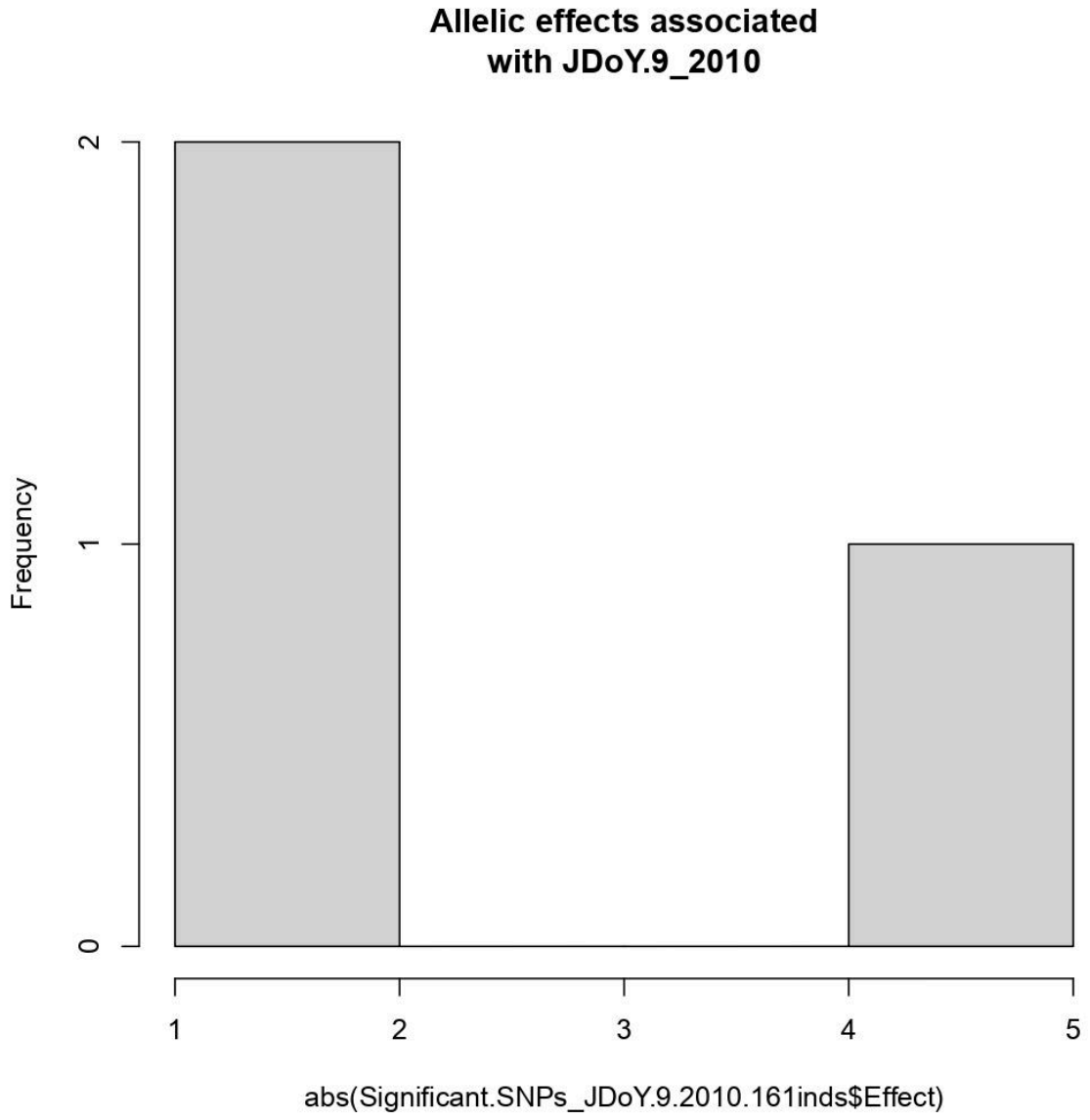
DOY BBCH 07 2009
DOY BBCH 07 2010
DOY BBCH 09 2009
DOY BBCH 09 2010
Logistic slope *a* 2009
Logistic slope *a* 2010
Logistic mid-point *b* 2009
Logistic mid-point *b* 2010
PSS 2009
PSS 2010
Degree Days at BBCH 07 2009
Degree Days at BBCH 07 2010
Degree Days at BBCH 09 2009
Degree Days at BBCH 09 2010

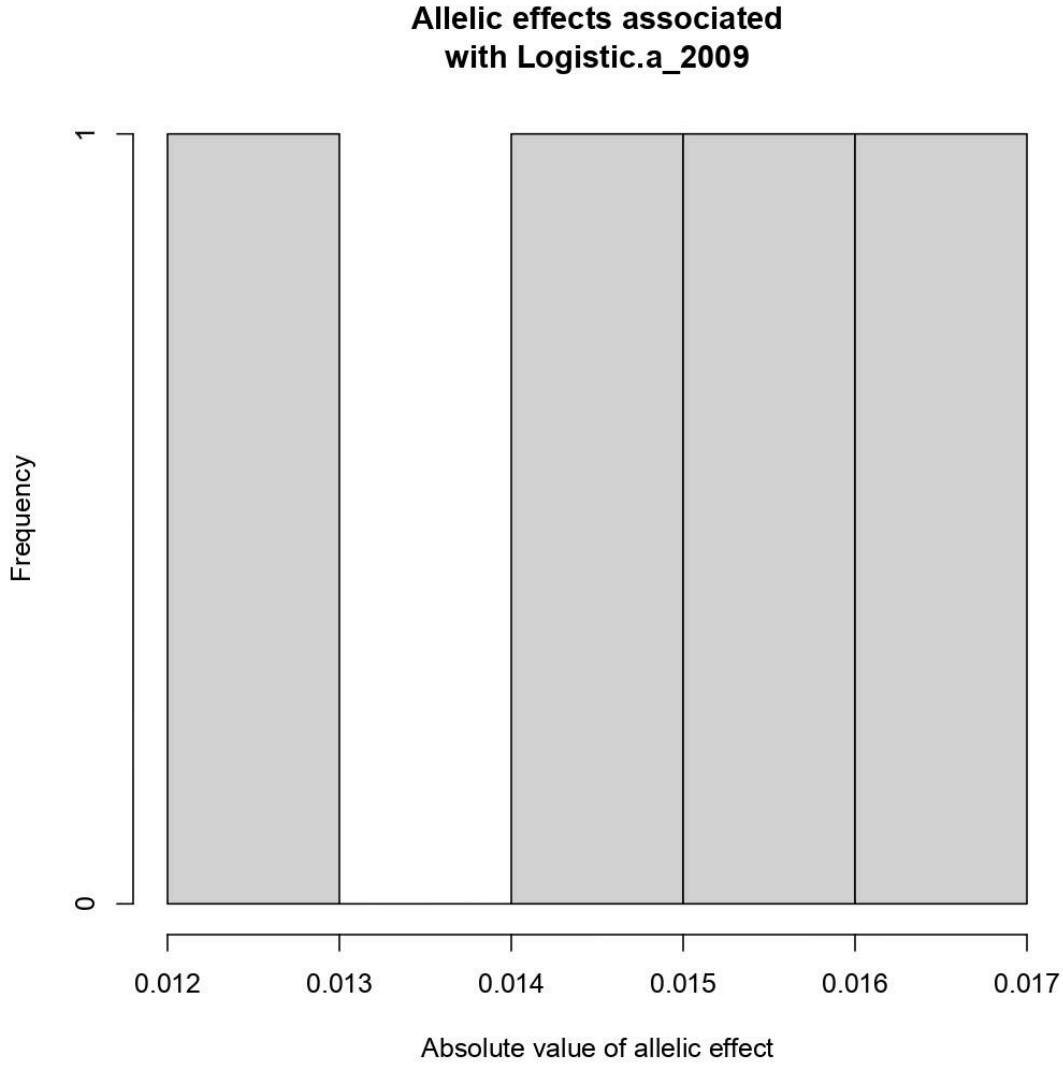
DOY BBCH 07 2009
DOY BBCH 07 2010
DOY BBCH 09 2009
DOY BBCH 09 2010
Logistic slope a 2009
Logistic slope a 2010
Logistic mid-point b 2009
Logistic mid-point b 2010
PSS 2009
PSS 2010
Degree Days at BBCH 07 2009
Degree Days at BBCH 07 2010
Degree Days at BBCH 09 2009
Degree Days at BBCH 09 2010

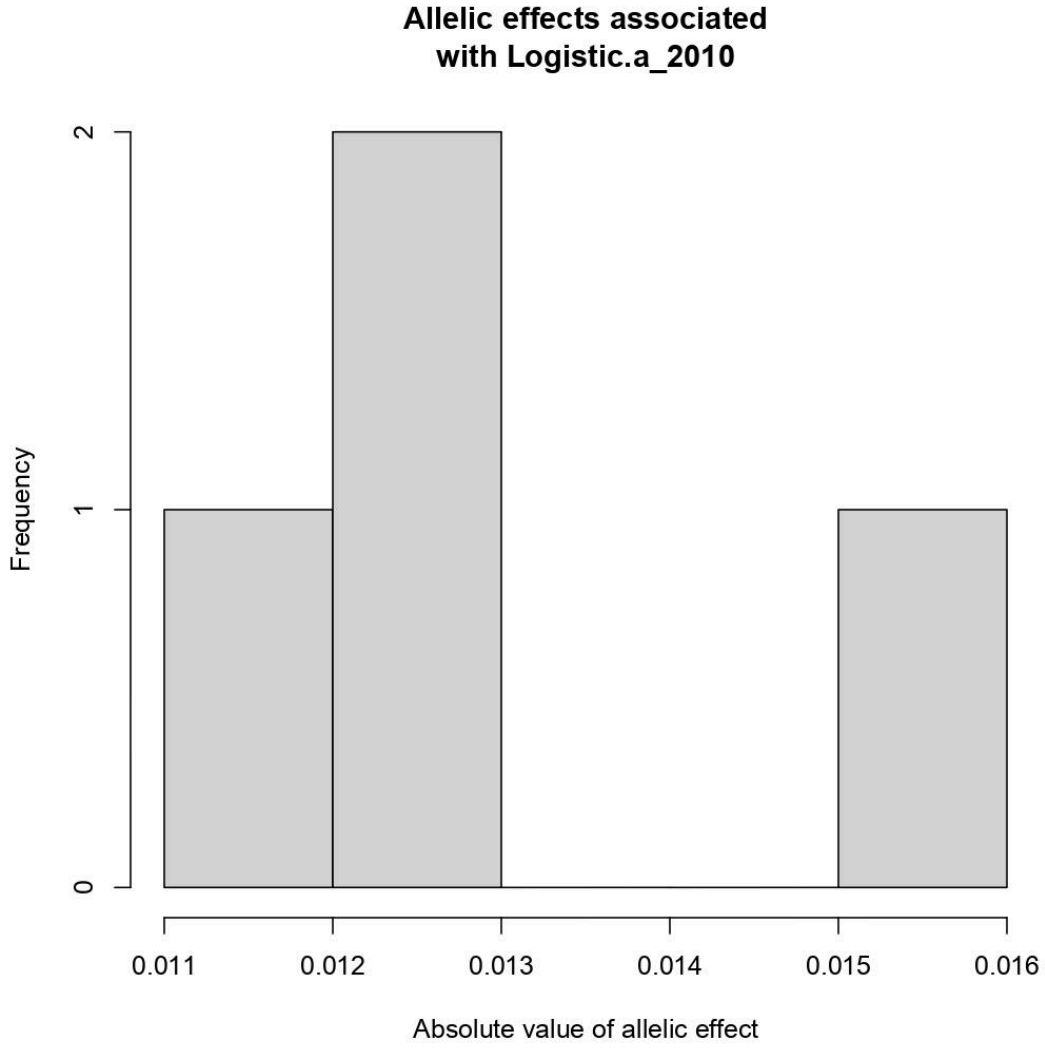

Allelic effects associated
with JDoY.7_2009
Frequency
0
1
2
3
4
5
6
7
0
2
4
6
8
10
Absolute value of allelic effect

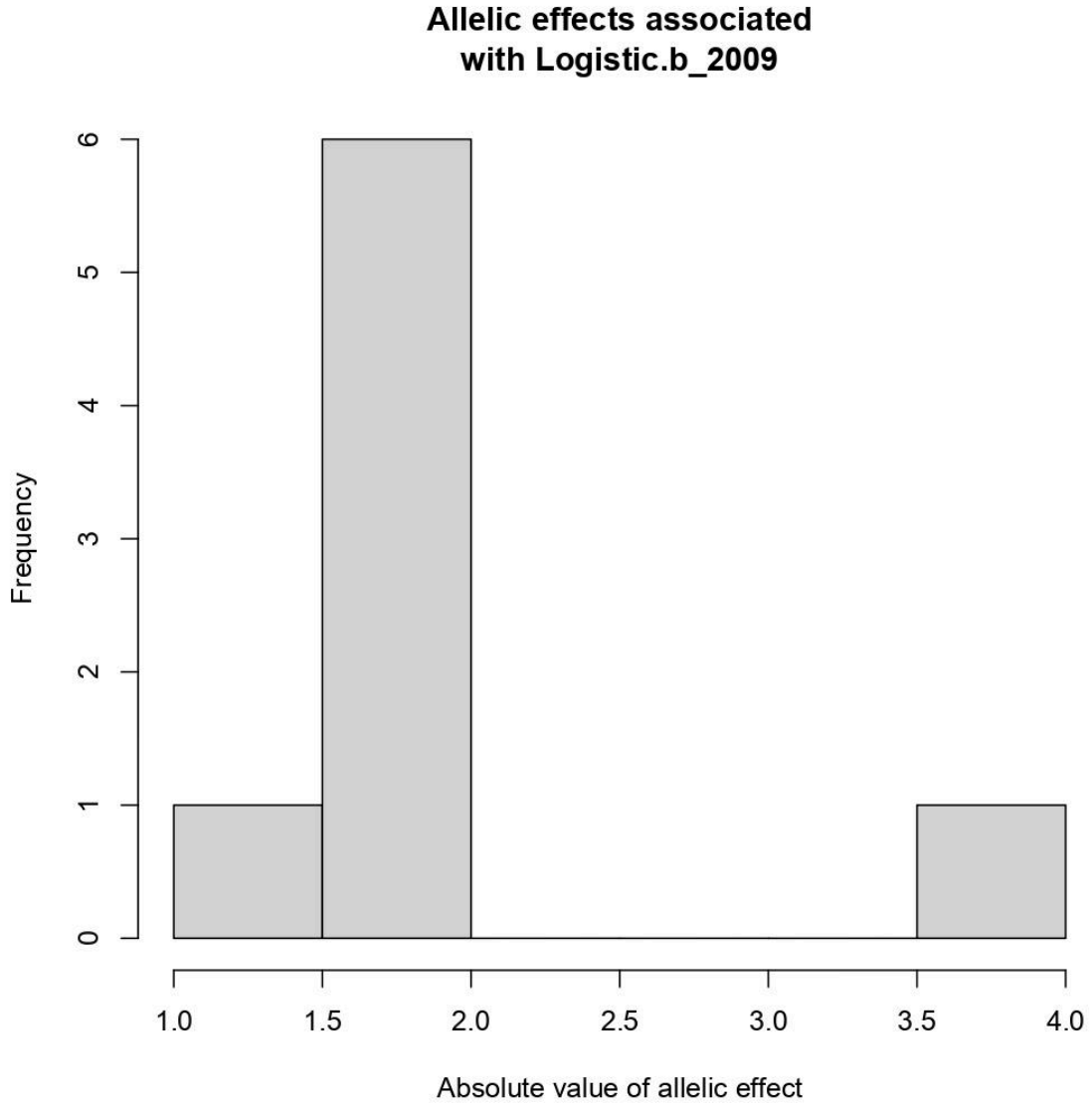

Allelic effects associated
with JDoY.7_2010
Frequency
0
1
2
0.7
0.8
0.9
1.0
1.1
Absolute value of allelic effect

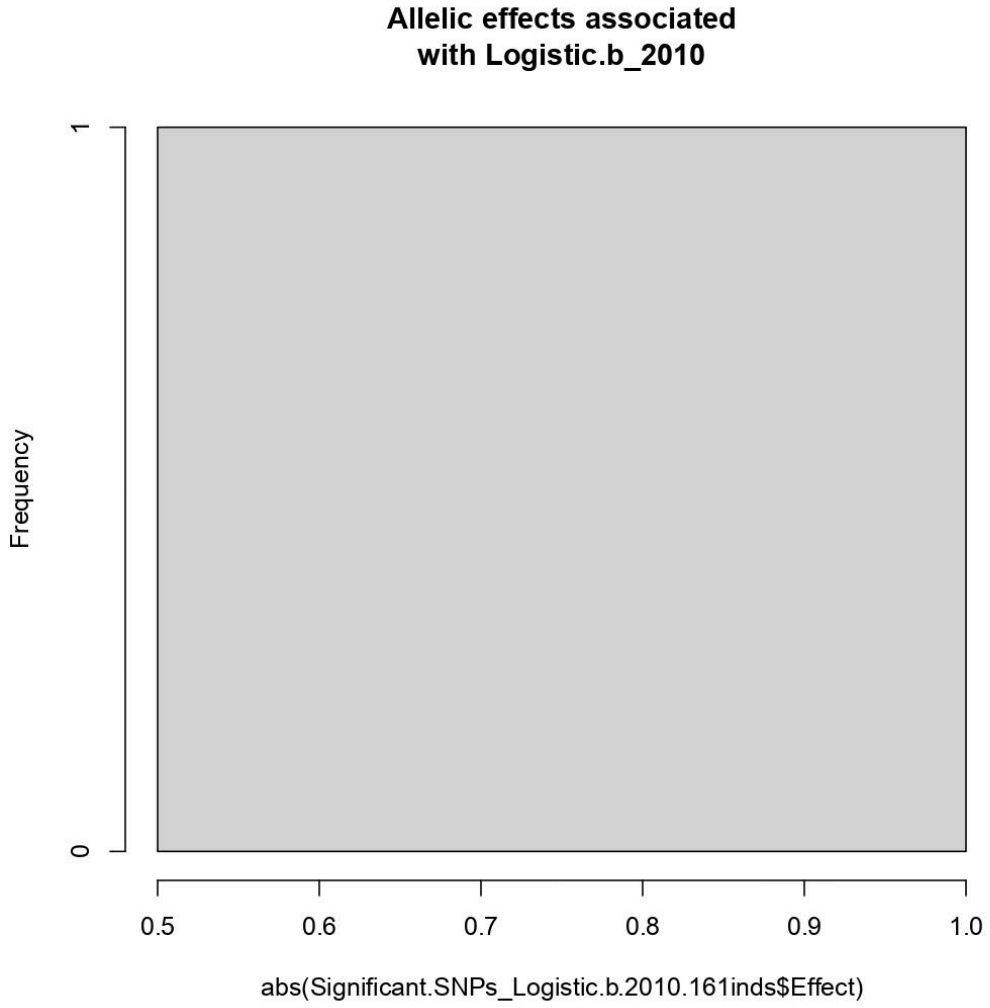

Allelic effects associated
with JDoY.9_2009
Frequency
0
1
2
3
4
2
4
6
8
10
12
Absolute value of allelic effect

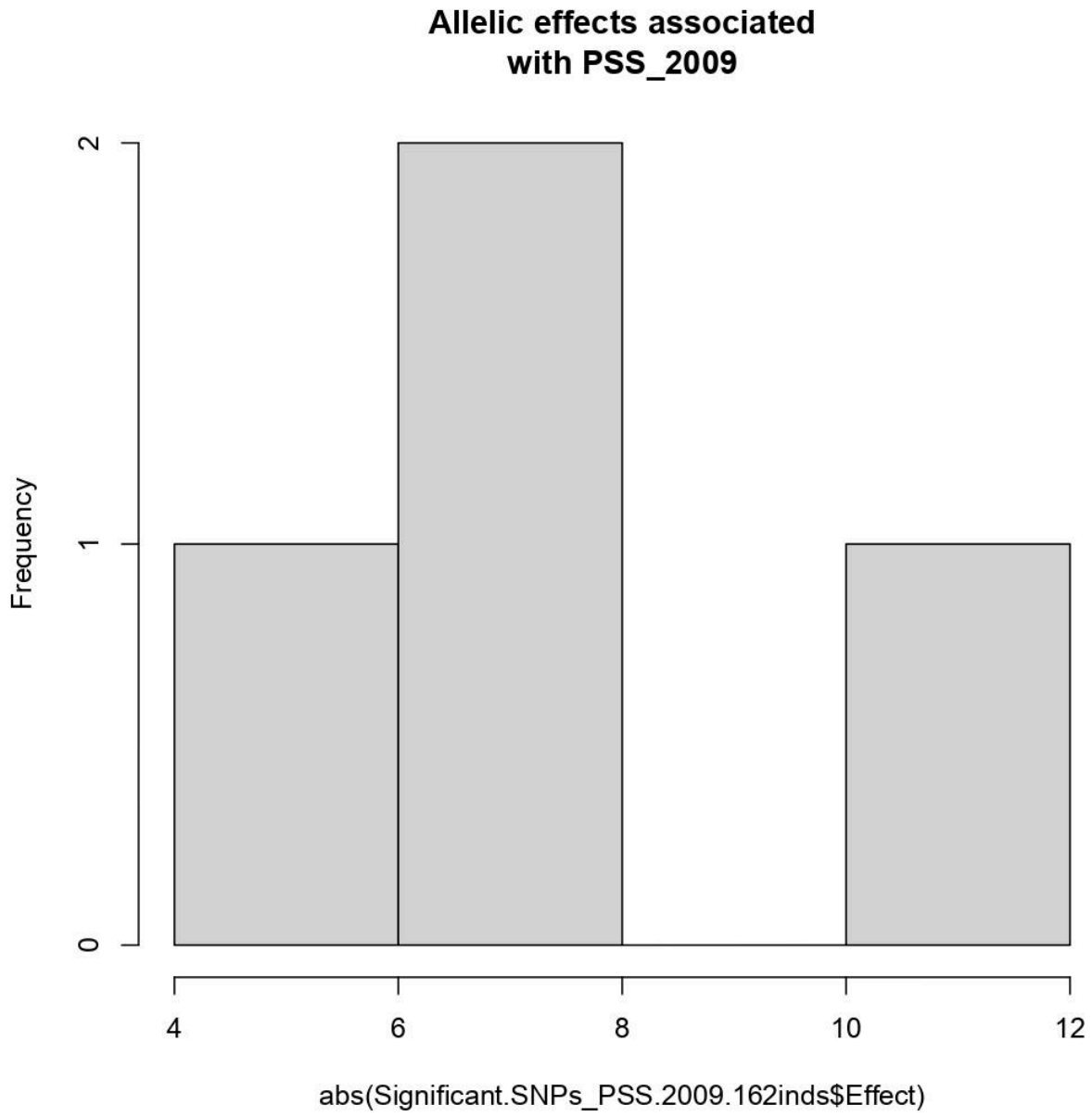

Allelic effects associated
with JDoY.9_2010
Frequency
0
1
2
1
2
3
4
5
abs(Significant.SNPs_JDoY.9.2010.161inds$Effect)

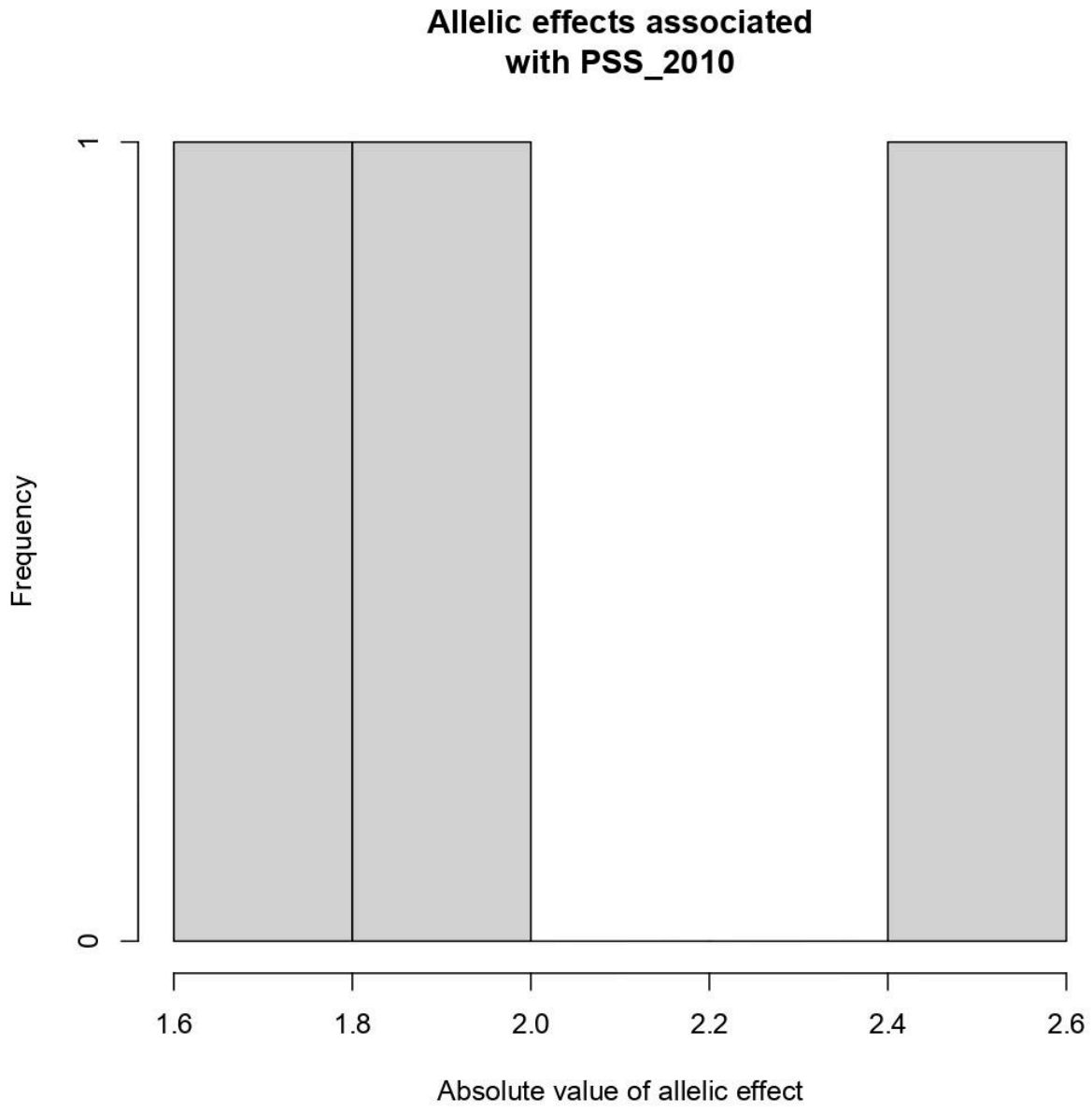

Allelic effects associated
with Logistic.a_2009
Frequency
1
0
0.012
0.013
0.014
0.015
0.016
0.017
Absolute value of allelic effect

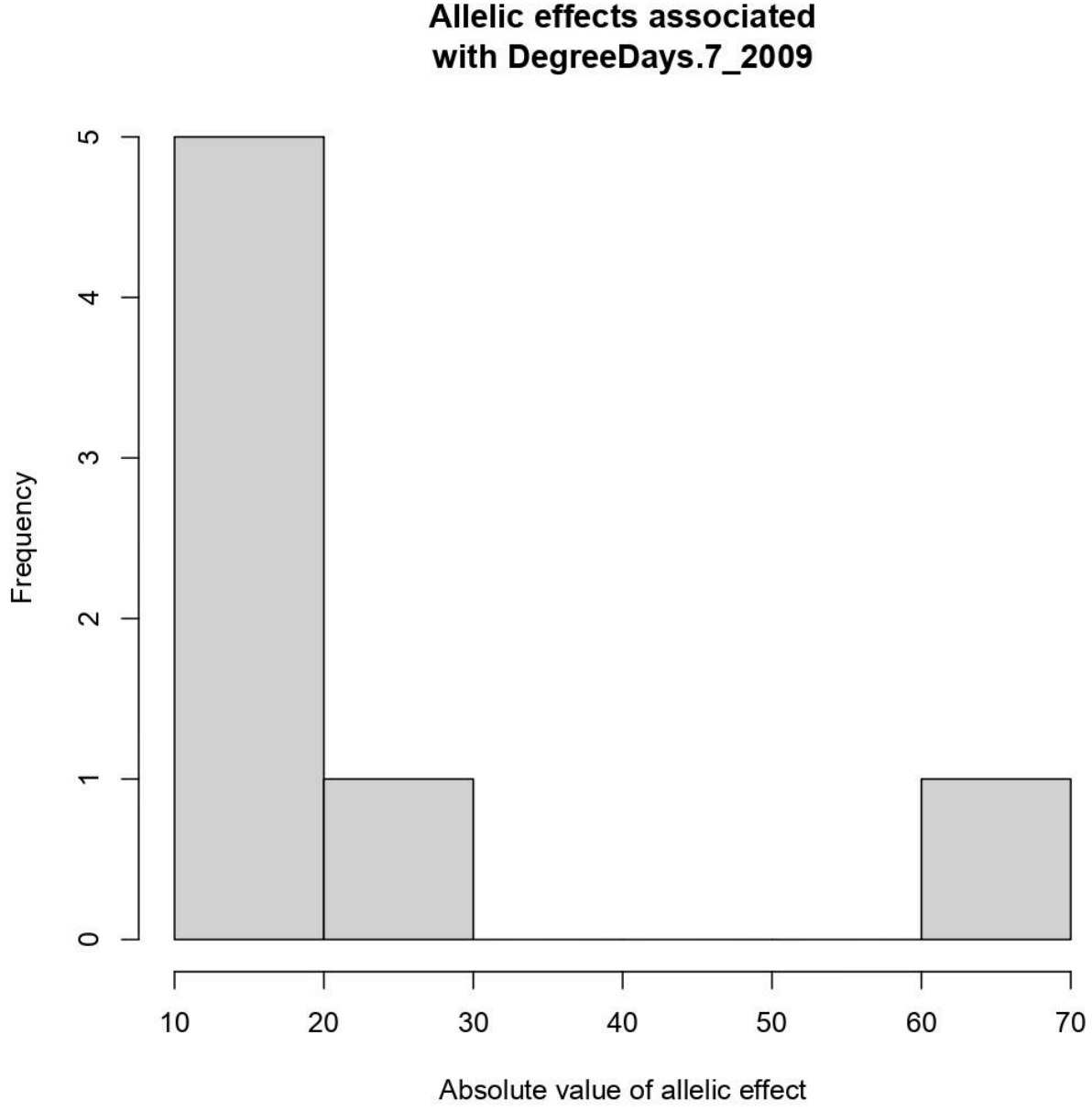

Allelic effects associated
with Logistic.a_2010
Frequency
0
1
2
0.011
0.012
0.013
0.014
0.015
0.016
Absolute value of allelic effect

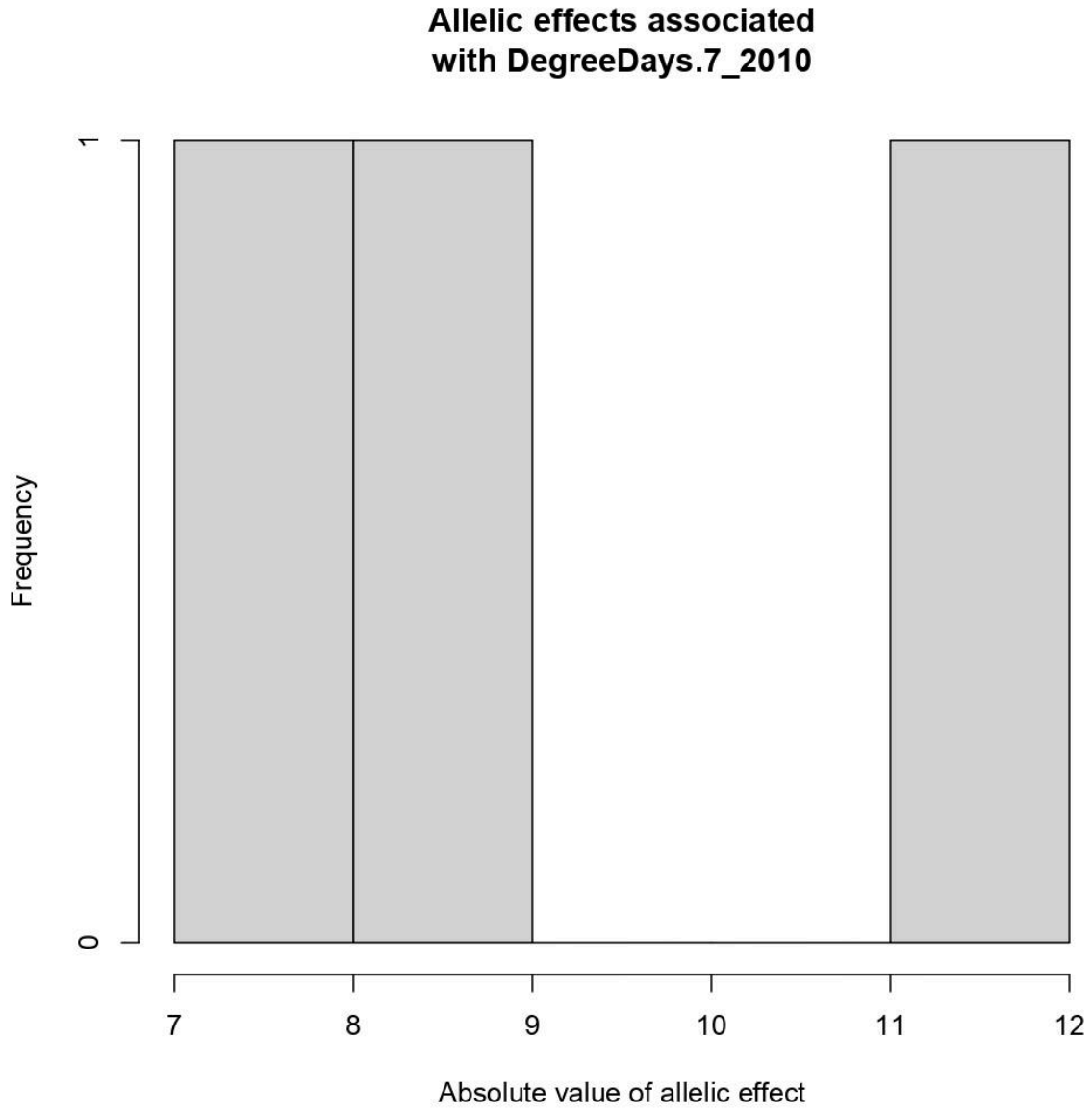

Allelic effects associated
with Logistic.b_2009
Frequency
0
1
2
3
4
5
6
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Absolute value of allelic effect

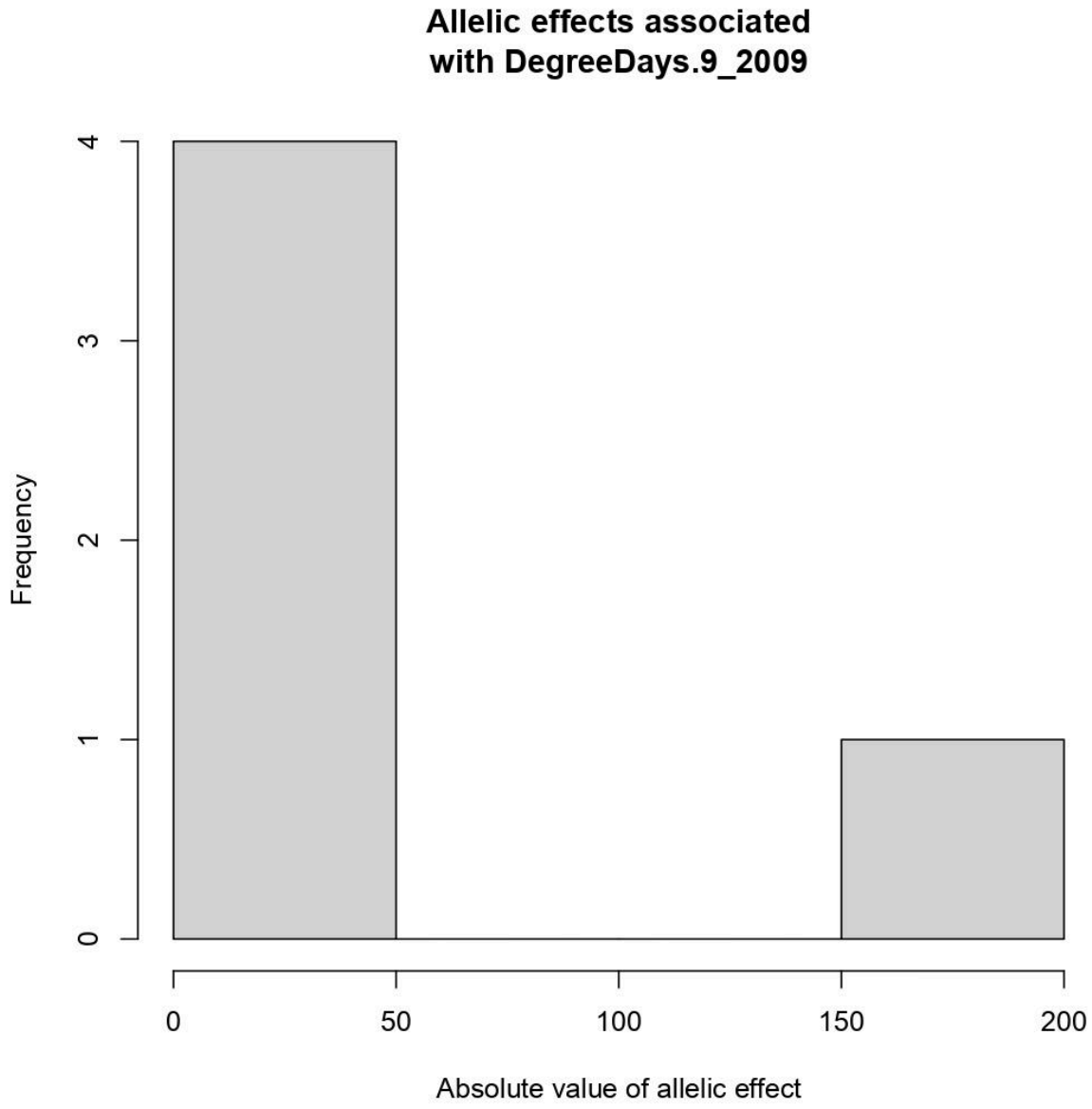

Allelic effects associated
with Logistic.b_2010
Frequency
1
0
0.5
0.6
0.7
0.8
0.9
1.0
abs(Significant.SNPs_Logistic.b.2010.161inds$Effect)

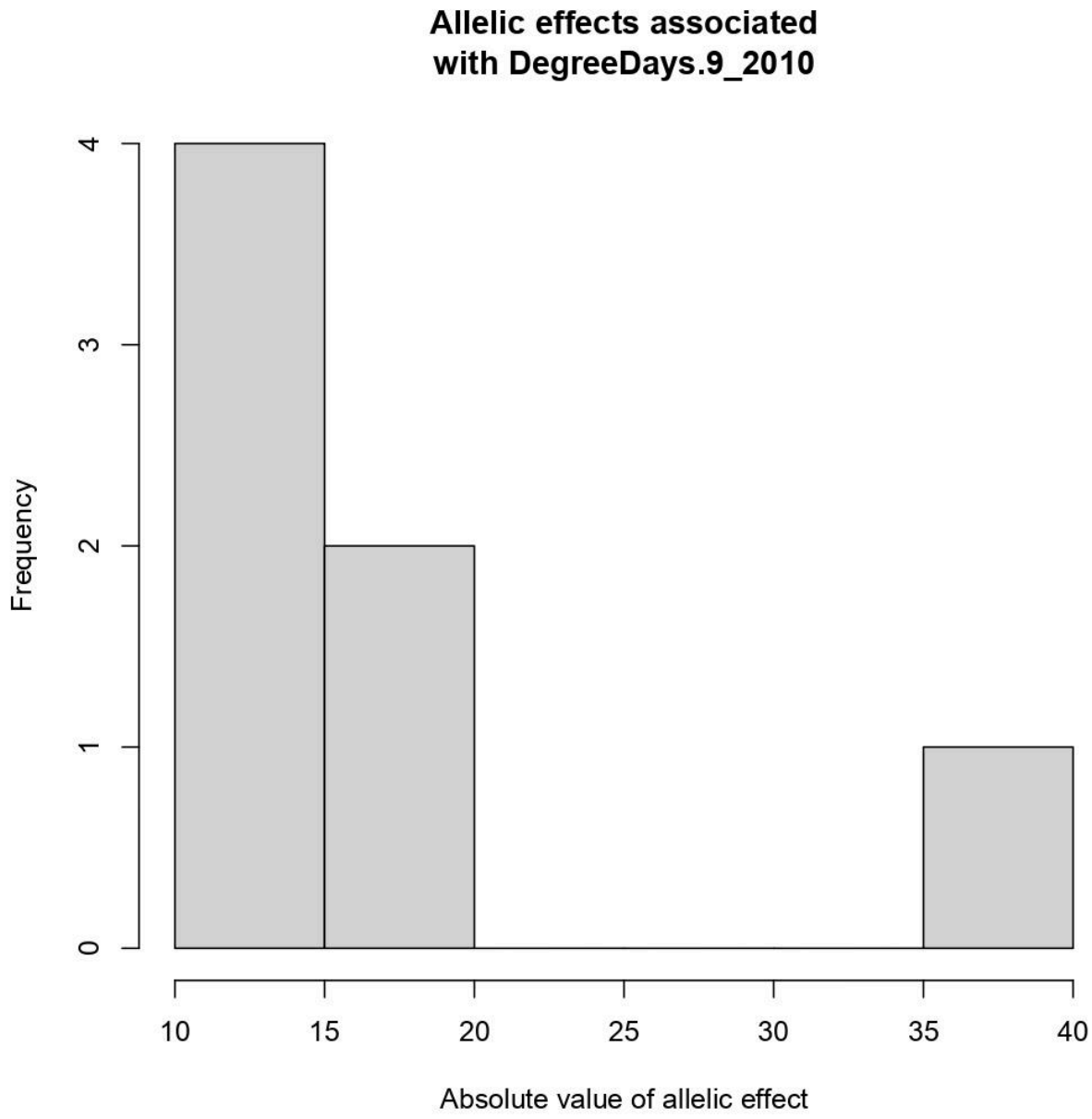

Allelic effects associated
with PSS_2009
Frequency
0
1
2
4
6
8
10
12
abs(Significant.SNPs_PSS.2009.162inds$Effect)


Allelic effects associated
with PSS_2010
Frequency
0
1
1.6
1.8
2.0
2.2
2.4
2.6
Absolute value of allelic effect


Allelic effects associated
with DegreeDays.7_2009
Frequency
0
1
2
3
4
5
10
20
30
40
50
60
70
Absolute value of allelic effect


Allelic effects associated
with DegreeDays.7_2010
Frequency
0
1
7
8
9
10
11
12
Absolute value of allelic effect

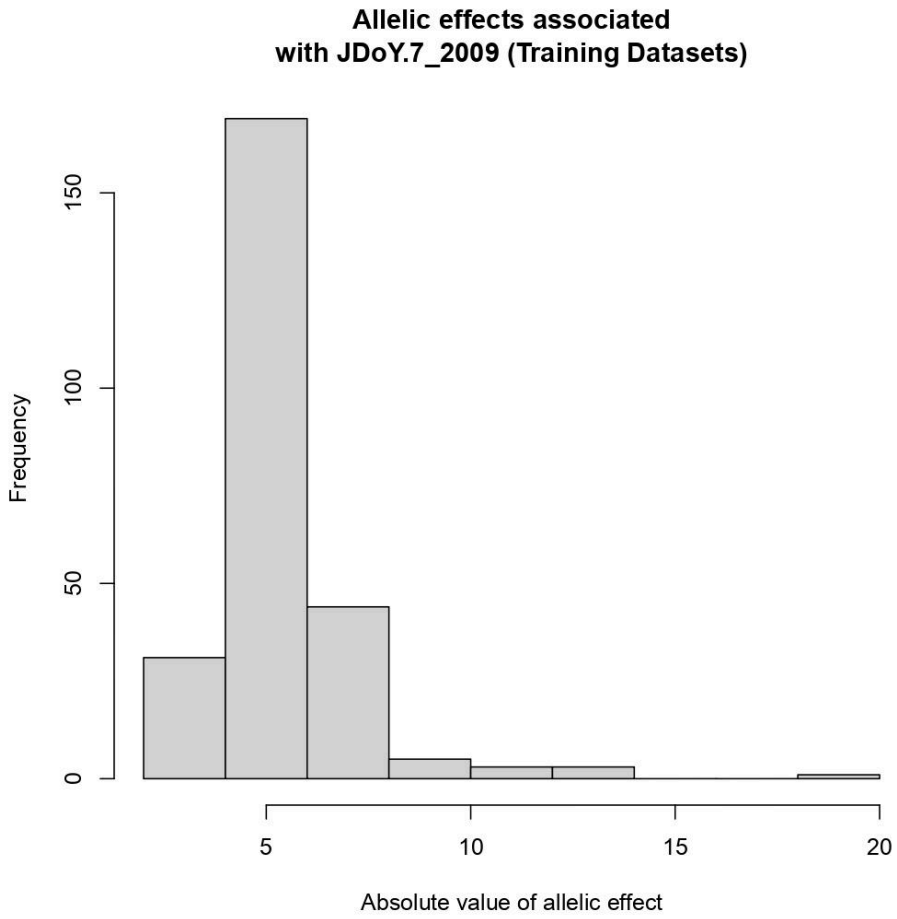

Allelic effects associated
with DegreeDays.9_2009
Frequency
0
1
2
3
4
0
50
100
150
200
Absolute value of allelic effect

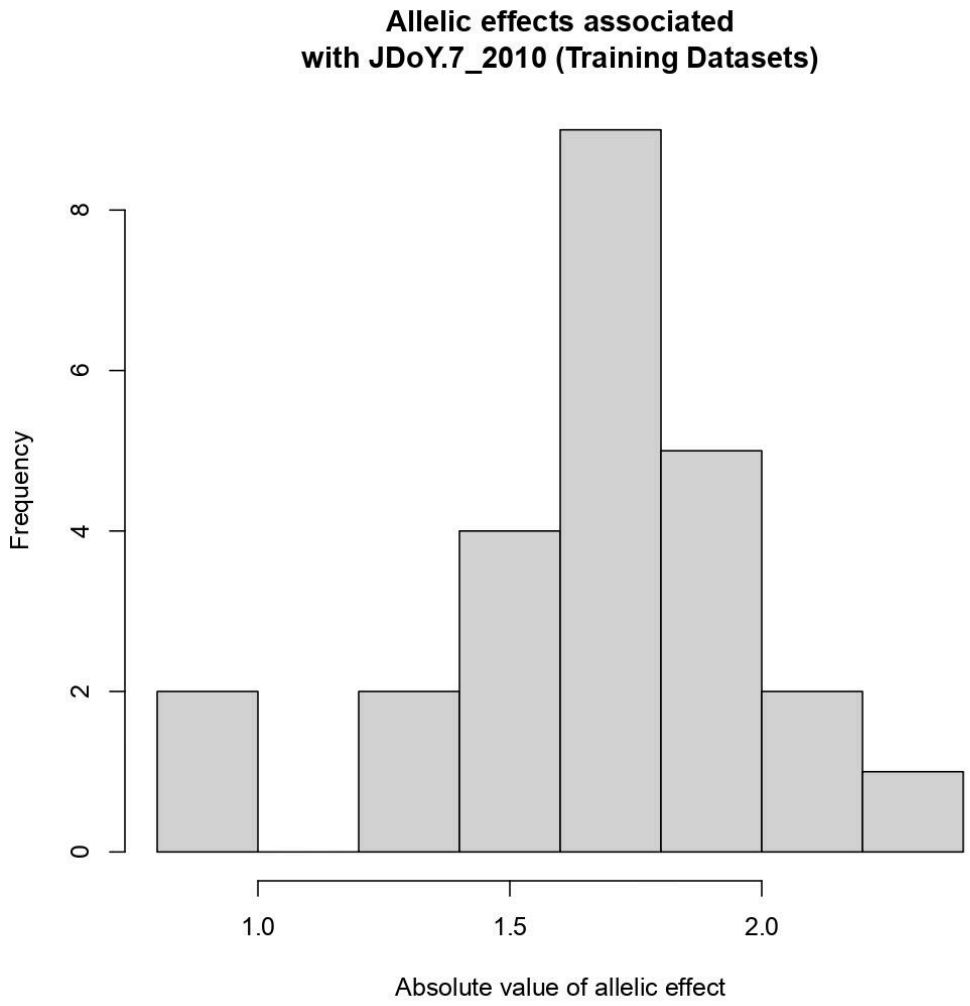

Allelic effects associated
with DegreeDays.9_2010
Frequency
0
1
2
3
4
10
15
20
25
30
35
40
Absolute value of allelic effect

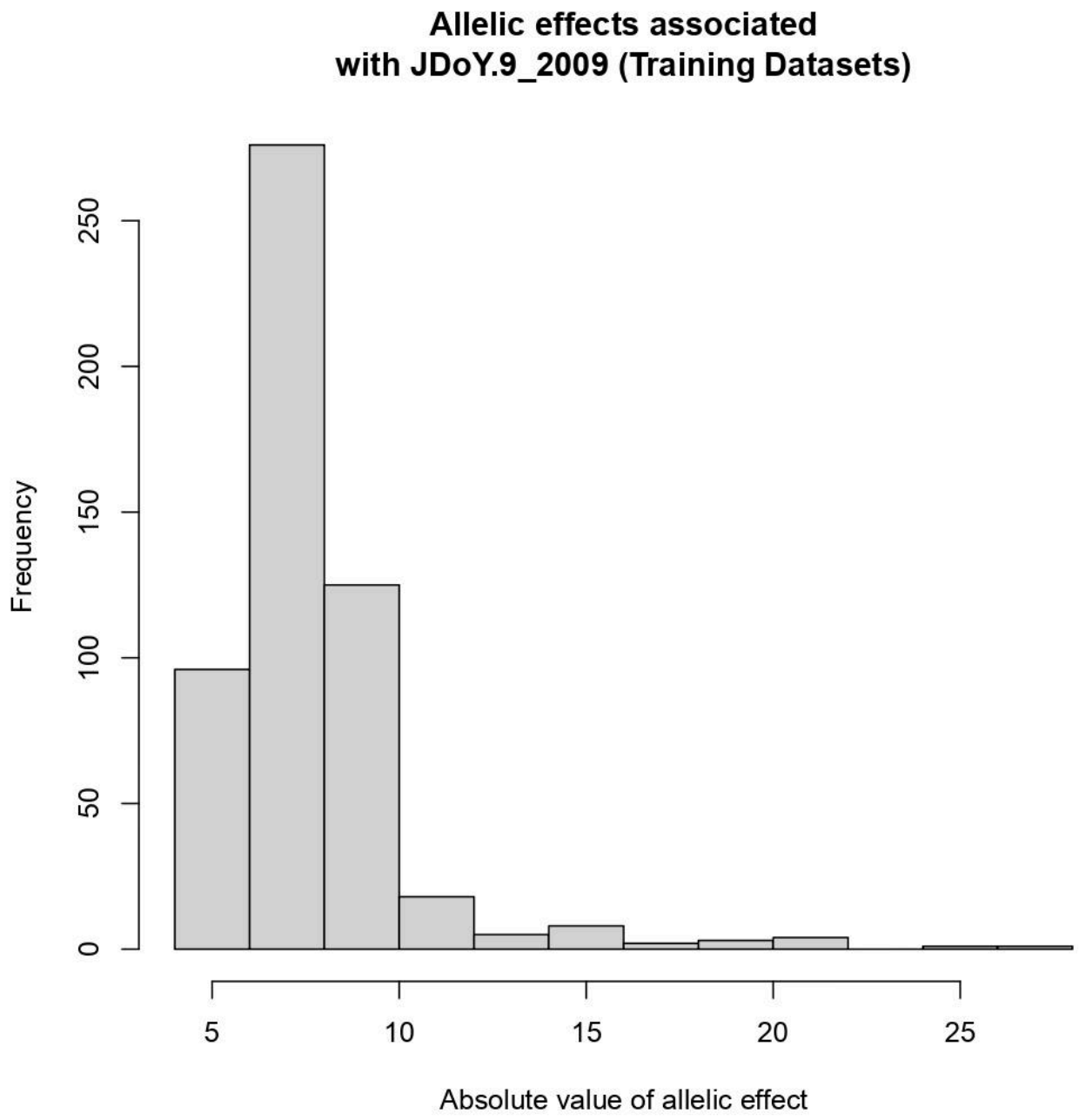
SupplementaryFigure 3 : Allelic Effects Training Datasets

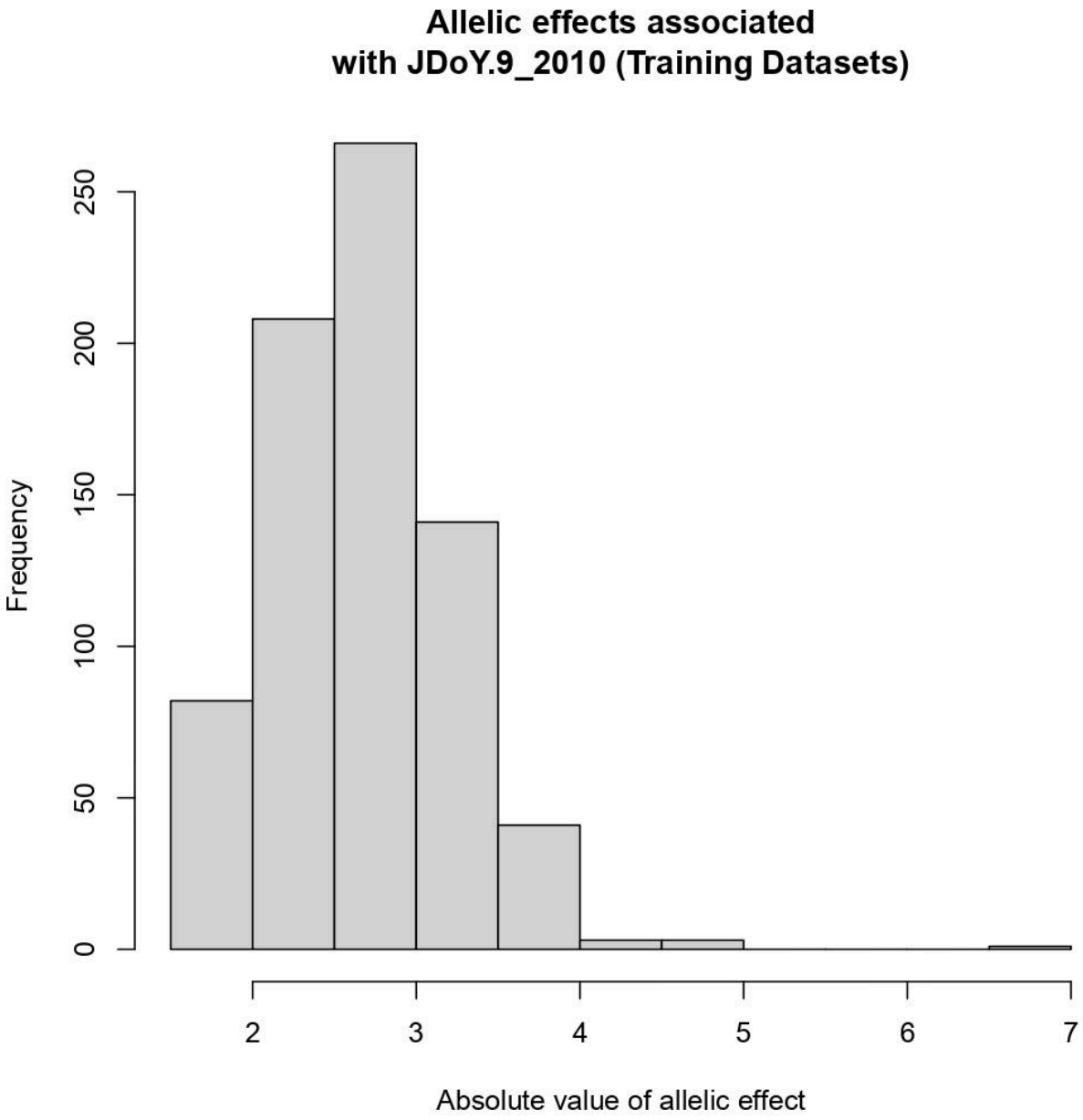
Traits are ordered as

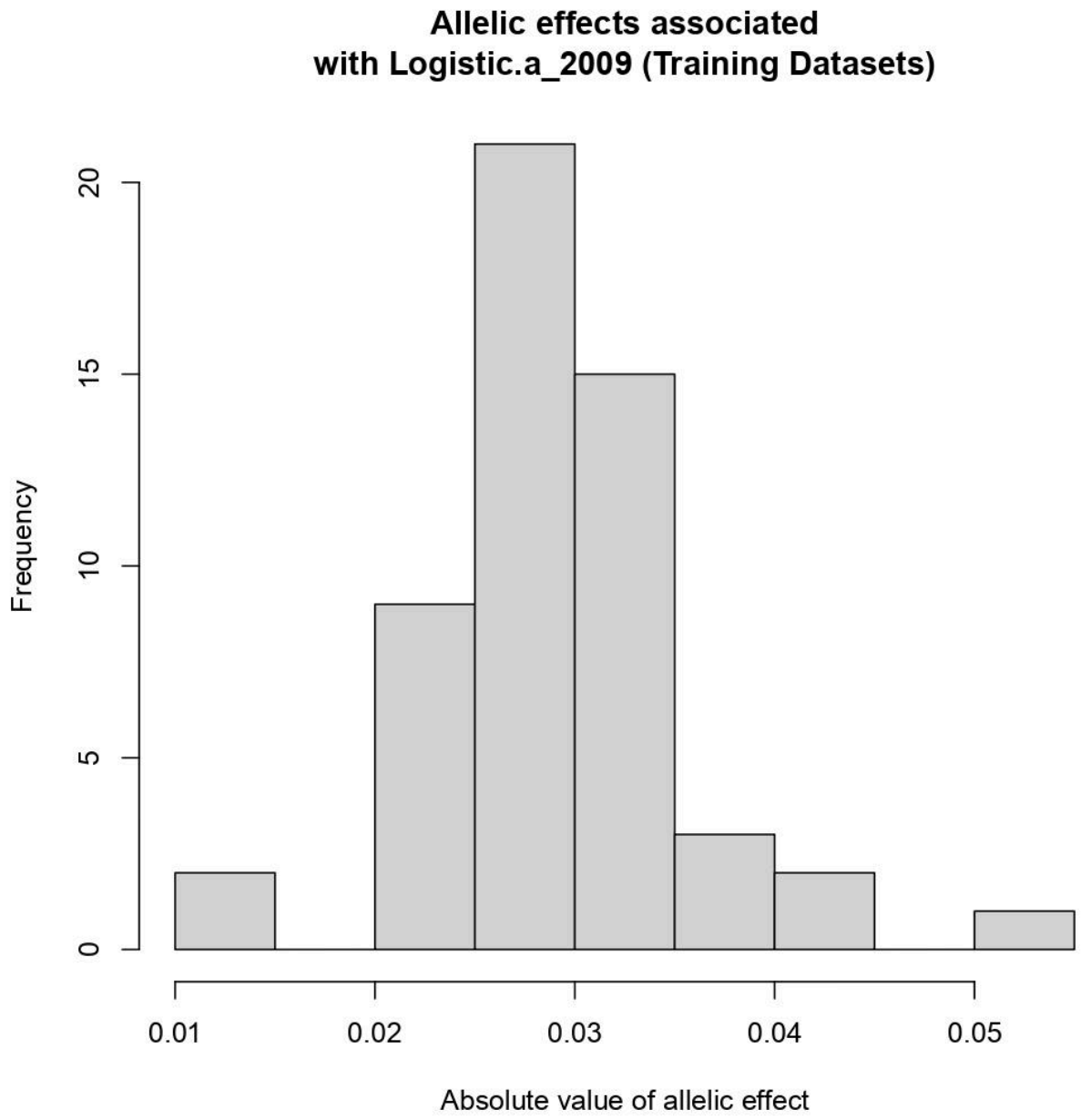
DOY BBCH 07 2009
DOY BBCH 07 2010
DOY BBCH 09 2009
DOY BBCH 09 2010
Logistic slope *a* 2009
Logistic slope *a* 2010
Logistic mid-point *b* 2009
Logistic mid-point *b* 2010
PSS 2009
PSS 2010
Degree Days at BBCH 07 2009
Degree Days at BBCH 07 2010
Degree Days at BBCH 09 2009
Degree Days at BBCH 09 2010

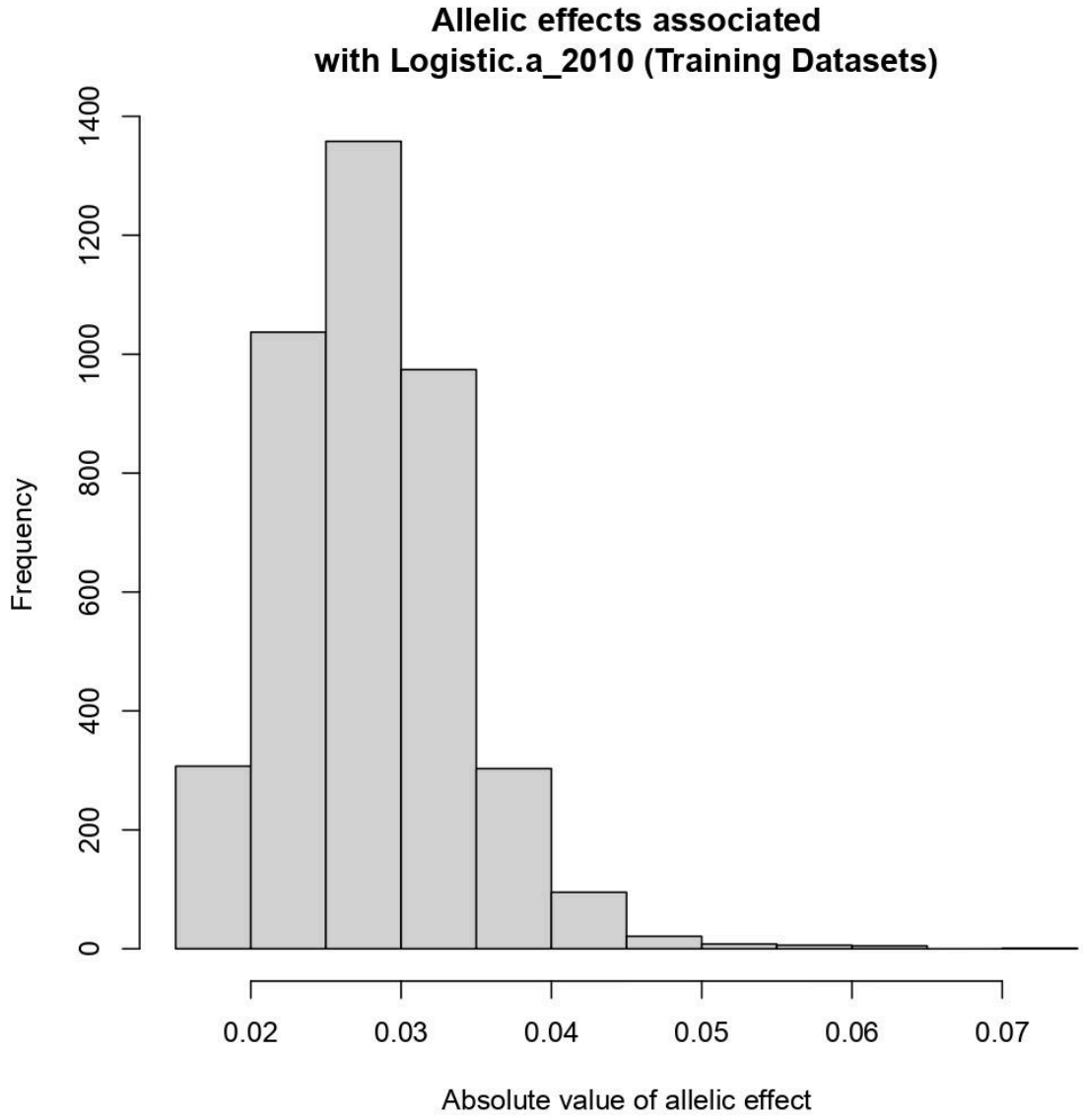

Allelic effects associated
with JDoY.7_2009 (Training Datasets)
Frequency
0
50
100
150
5
10
15
20
Absolute value of allelic effect

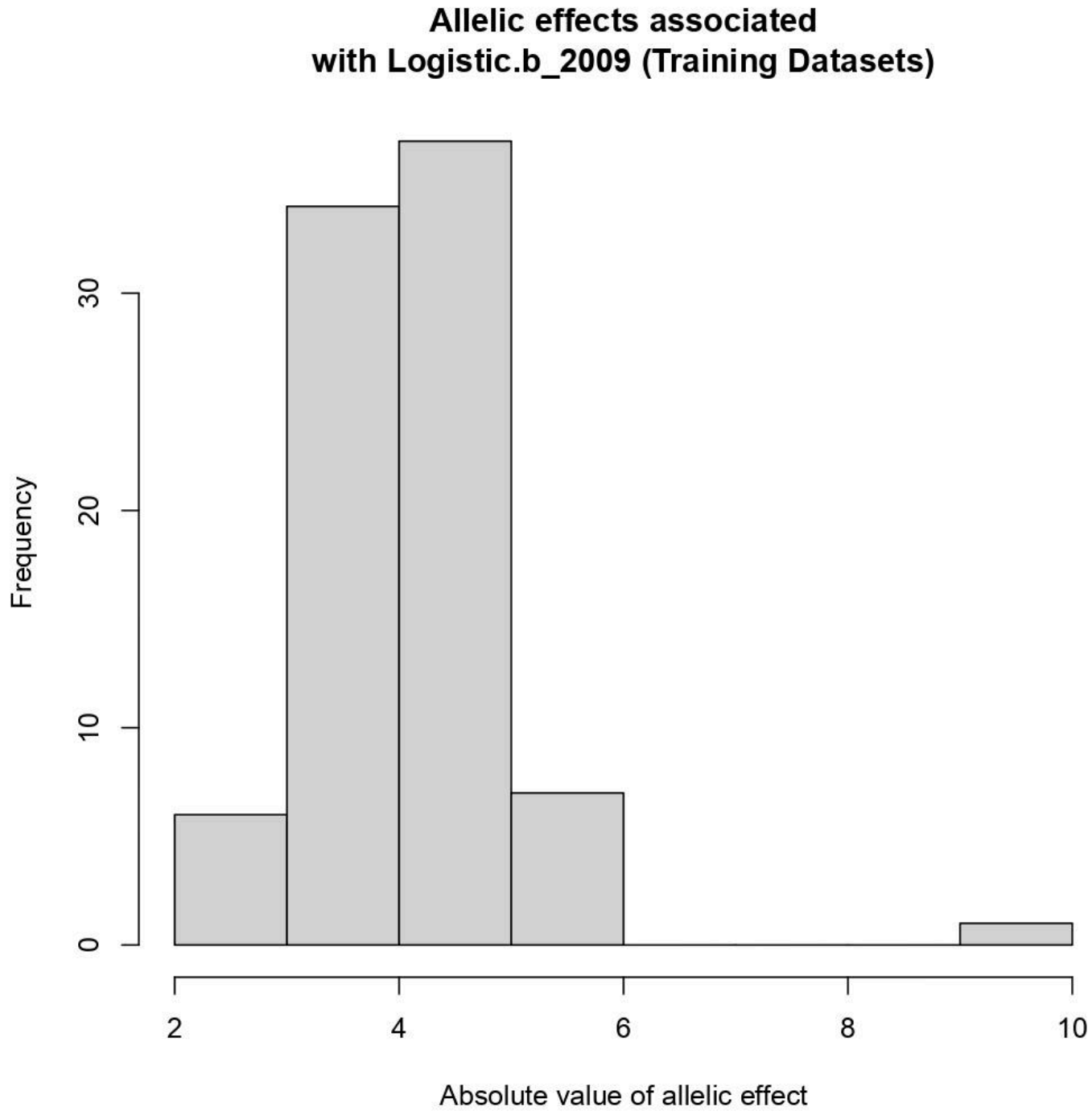

Allelic effects associated
with JDoY.7_2010 (Training Datasets)
Frequency
0
2
4
6
8
1.0
1.5
2.0
Absolute value of allelic effect

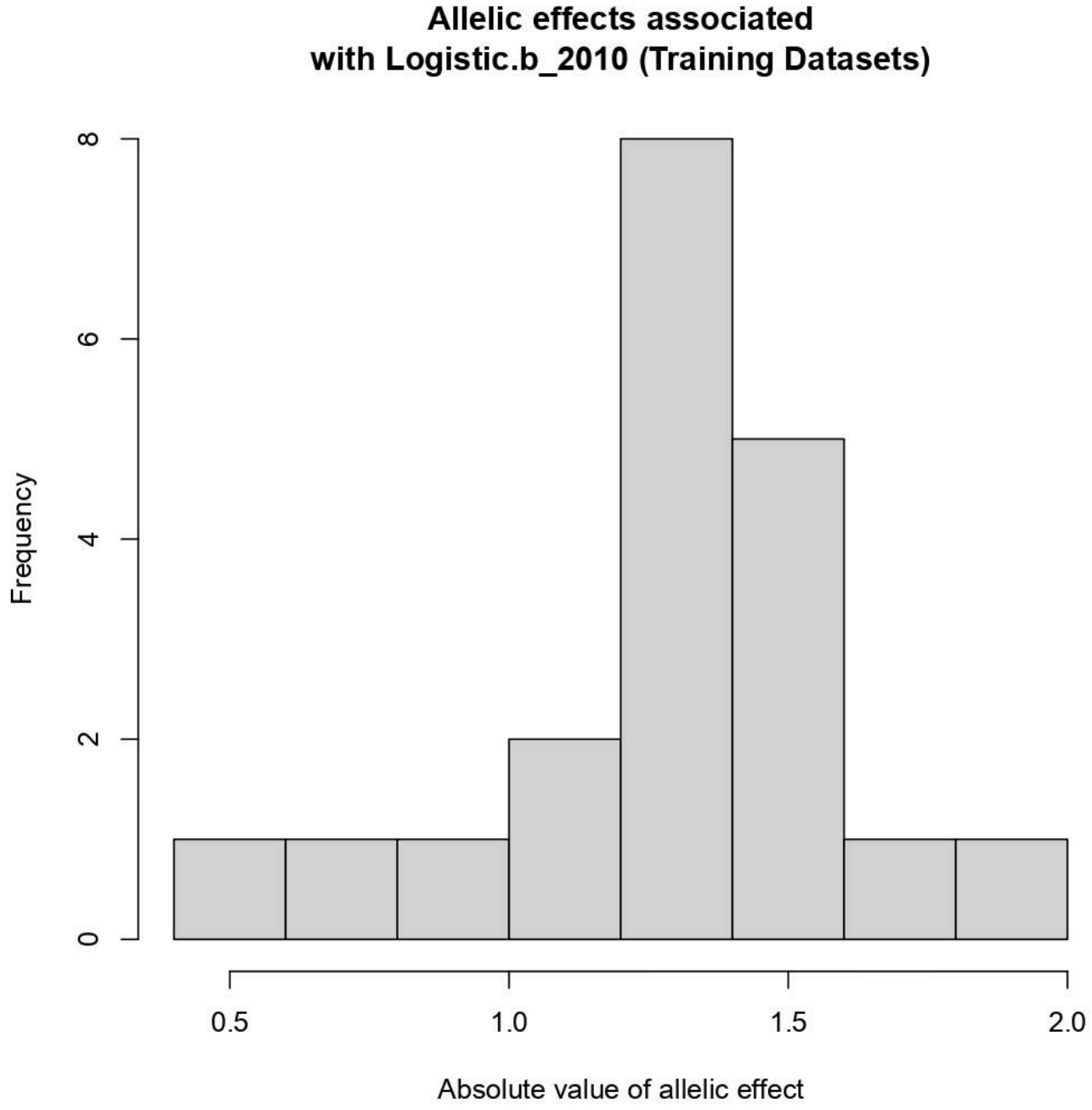

Allelic effects associated
with JDoY.9_2009 (Training Datasets)
Frequency
0
50
100
150
200
250
5
10
15
20
25
Absolute value of allelic effect

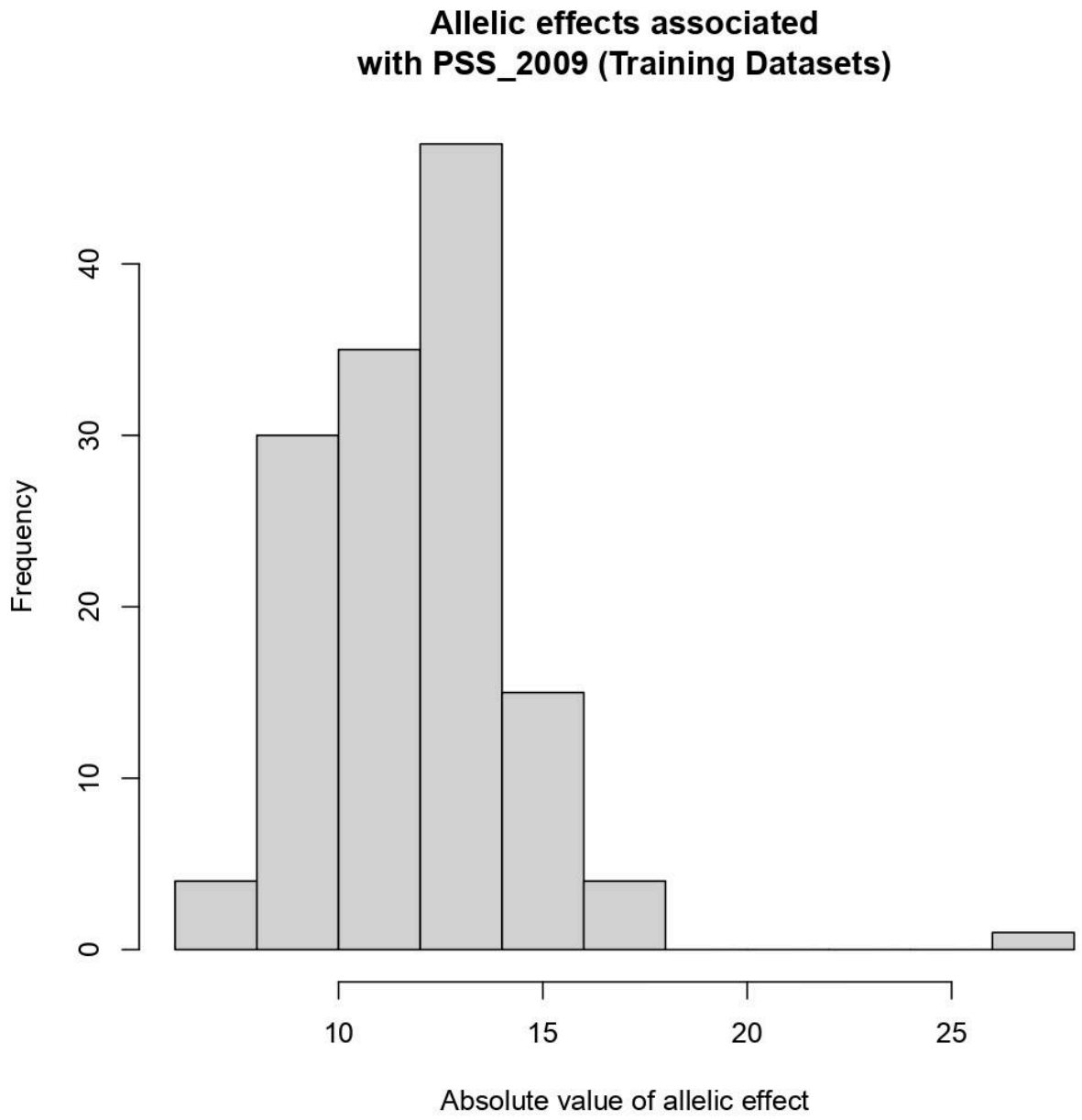

Allelic effects associated
with JDoY.9_2010 (Training Datasets)
Frequency
0
50
100
150
200
250
2
3
4
5
6
7
Absolute value of allelic effect

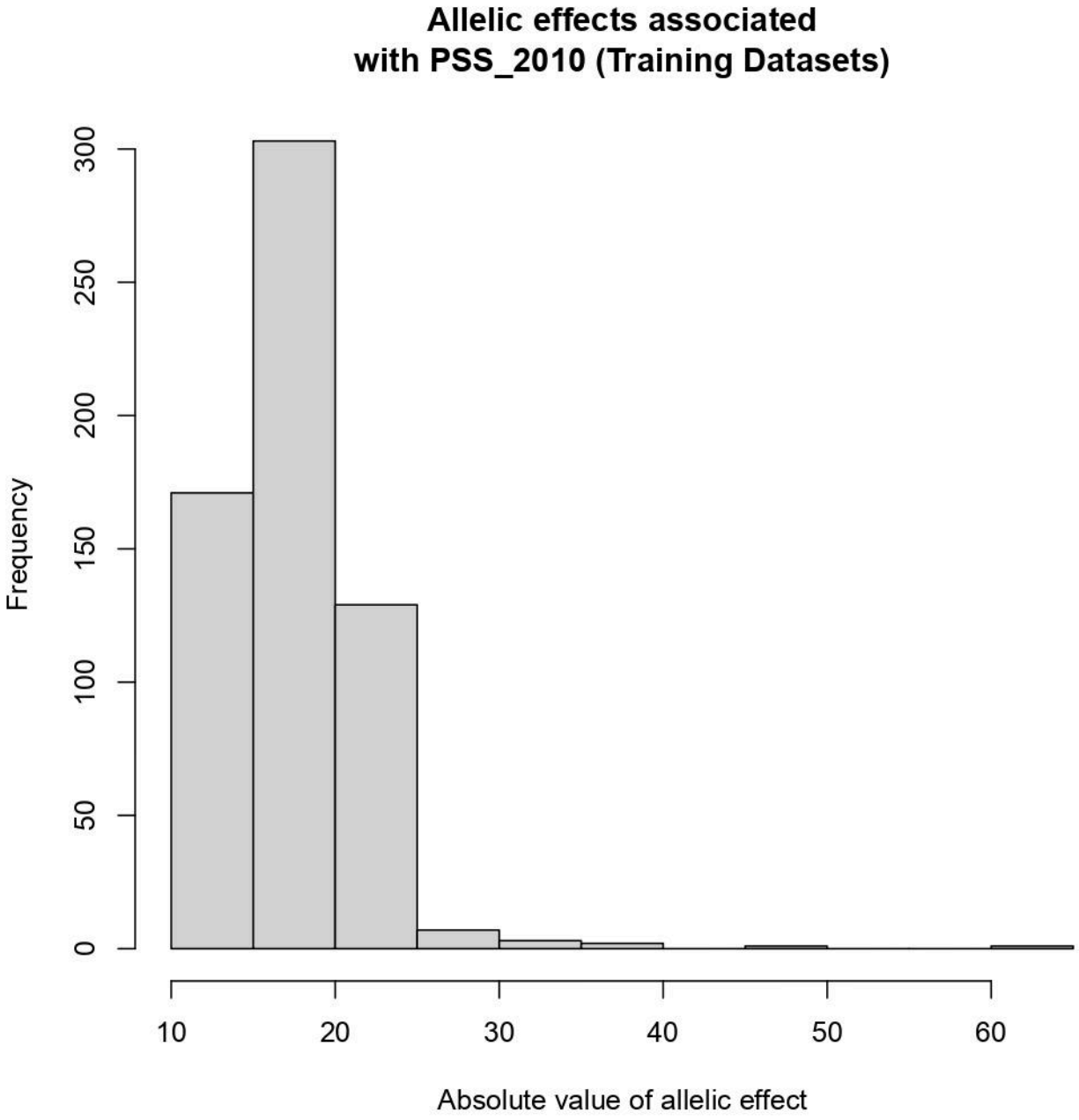

Allelic effects associated
with Logistic.a_2009 (Training Datasets)
Frequency
0
5
10
15
20
0.01
0.02
0.03
0.04
0.05
Absolute value of allelic effect

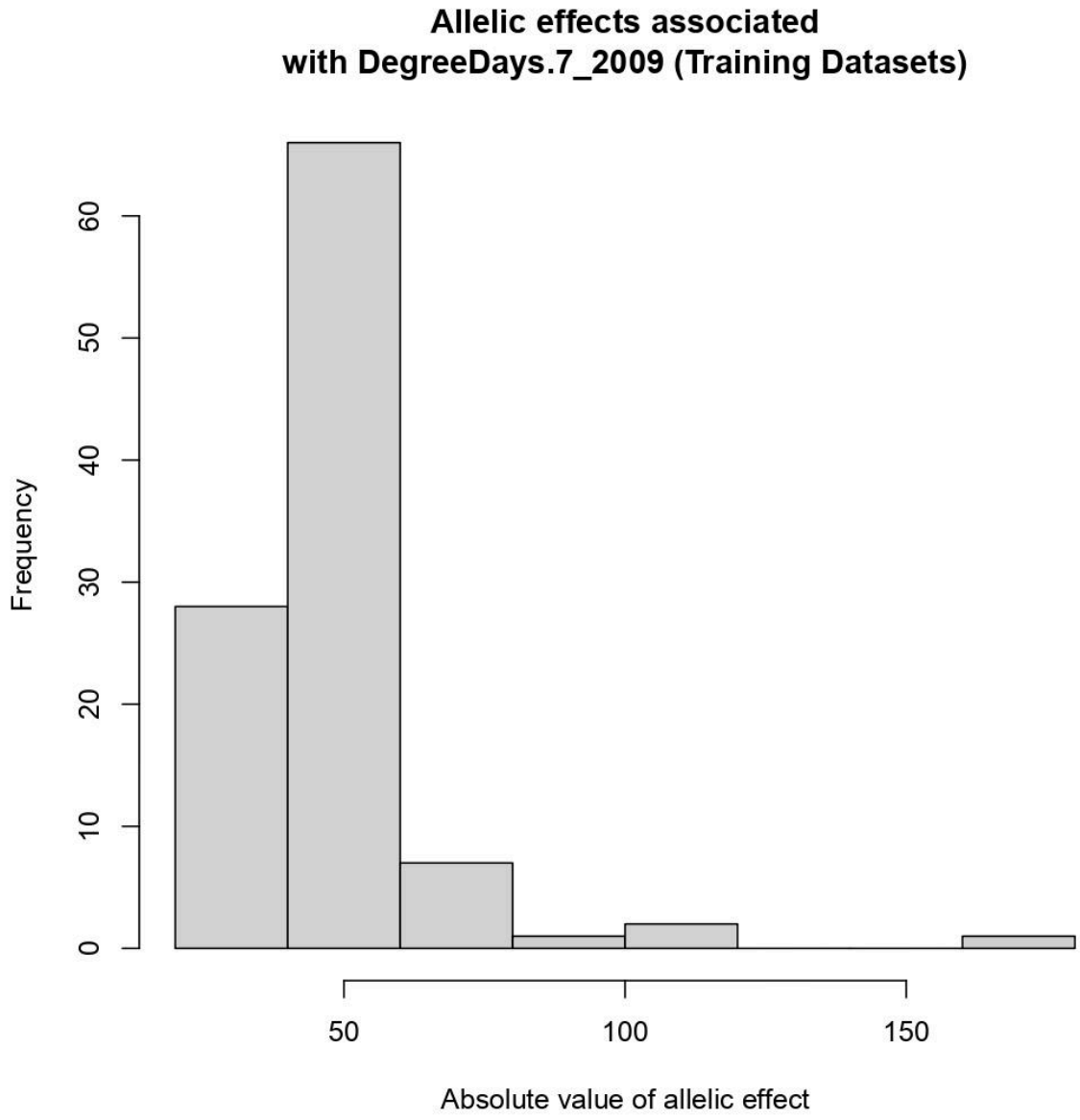

Allelic effects associated
with Logistic.a_2010 (Training Datasets)
Frequency
0
200
400
600
800
1000
1200
1400
0.02
0.03
0.04
0.05
0.06
0.07
Absolute value of allelic effect

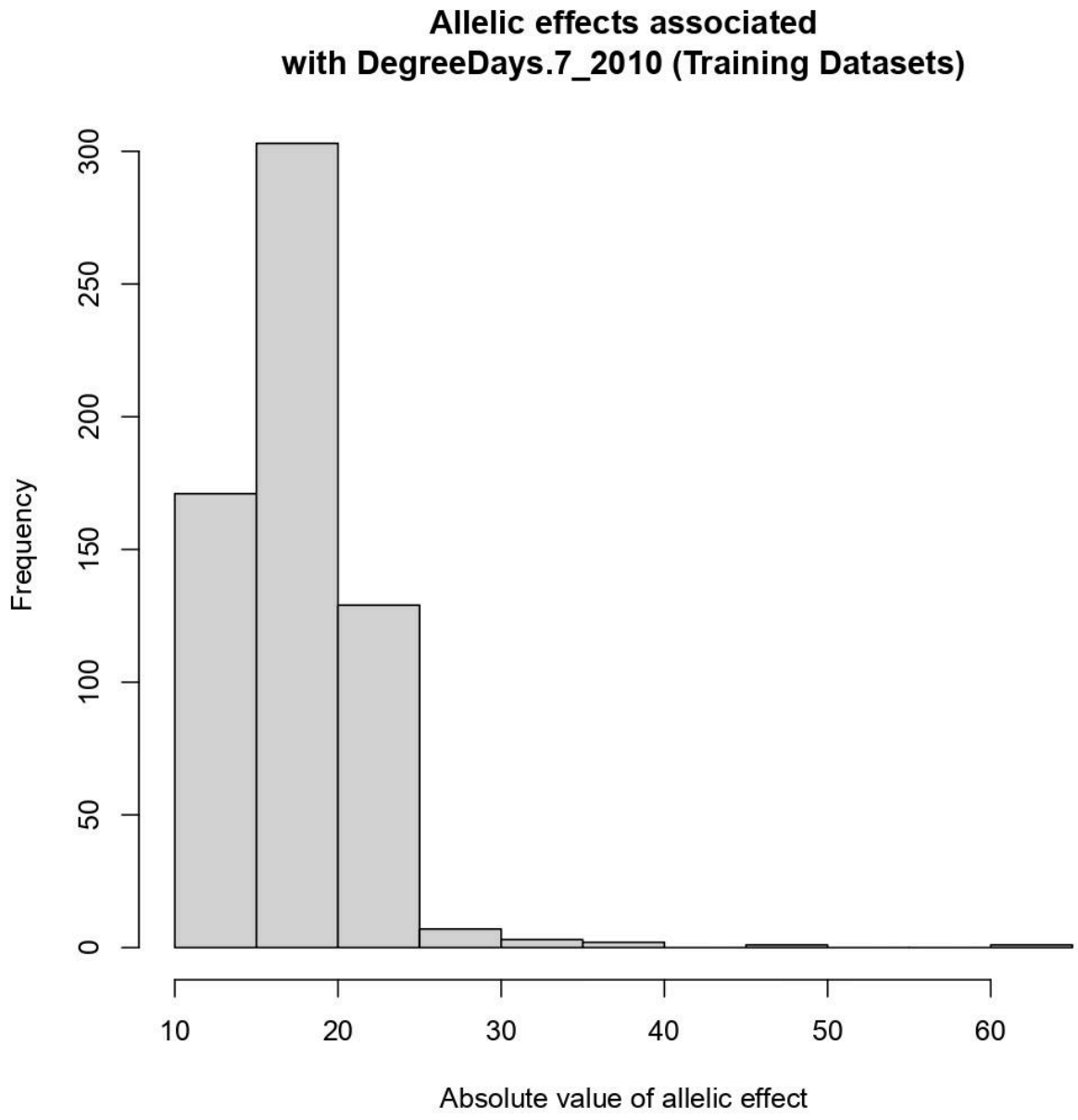

Allelic effects associated
with Logistic.b_2009 (Training Datasets)
Frequency
0
10
20
30
2
4
6
8
10
Absolute value of allelic effect

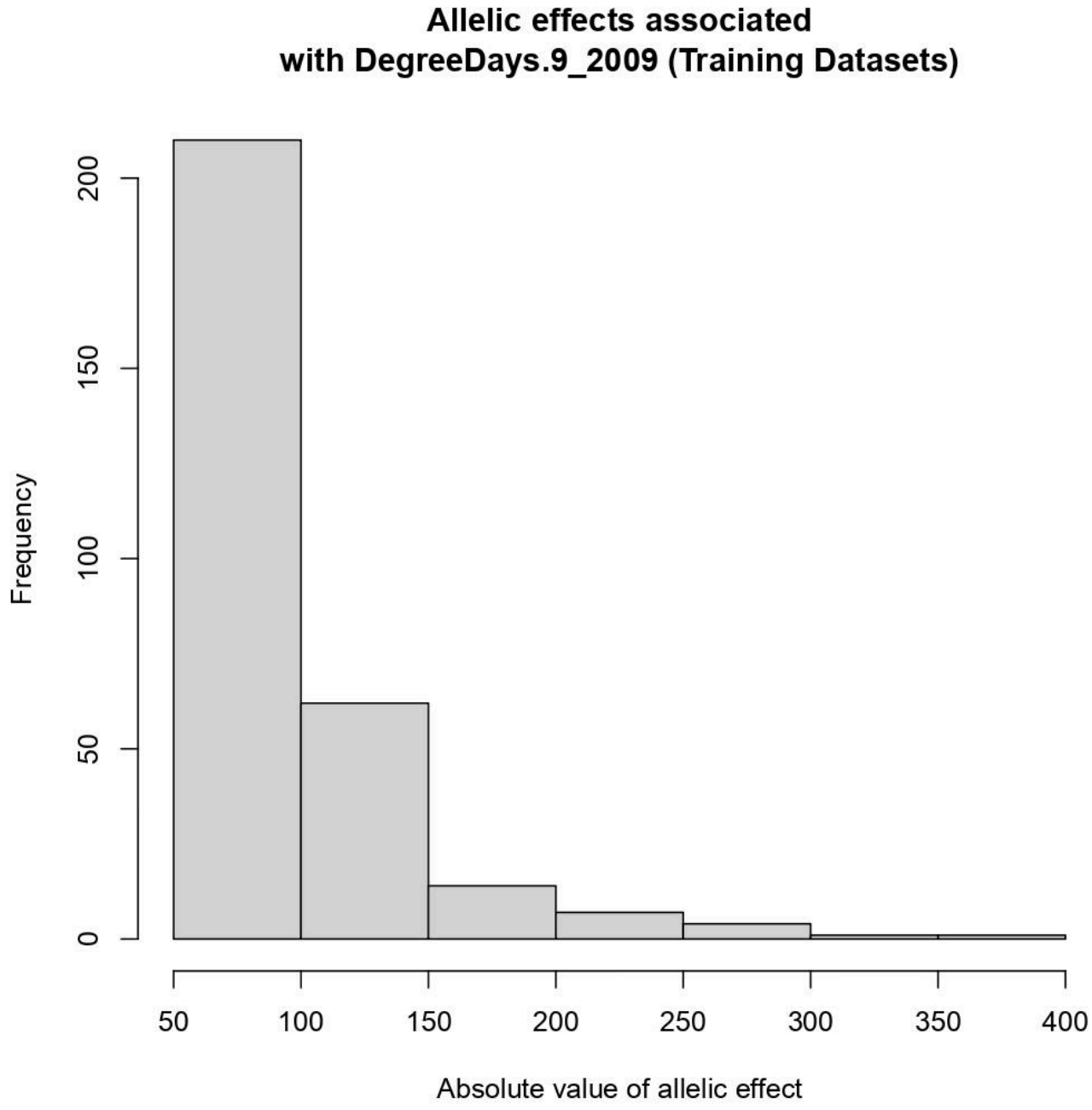

Allelic effects associated
with Logistic.b_2010 (Training Datasets)
Frequency
0
2
4
6
8
0.5
1.0
1.5
2.0
Absolute value of allelic effect

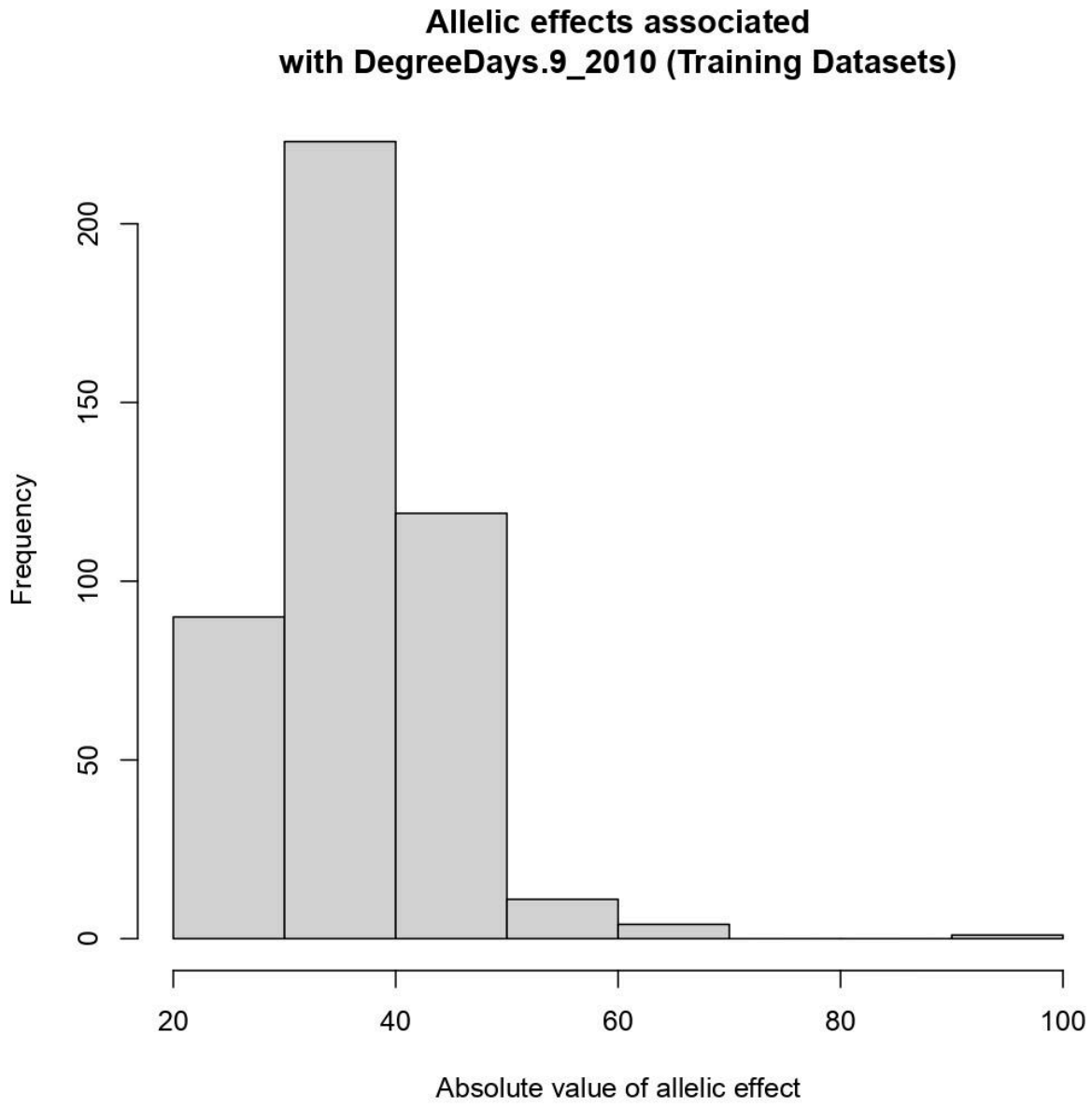

Allelic effects associated
with PSS_2009 (Training Datasets)
Frequency
0
10
20
30
40
10
15
20
25
Absolute value of allelic effect


Allelic effects associated
with PSS_2010 (Training Datasets)
Frequency
0
50
100
150
200
250
300
10
20
30
40
50
60
Absolute value of allelic effect


Allelic effects associated
with DegreeDays.7_2009 (Training Datasets)
Frequency
0
10
20
30
40
50
60
50
100
150
Absolute value of allelic effect


Allelic effects associated
with DegreeDays.7_2010 (Training Datasets)
Frequency
0
50
100
150
200
250
300
10
20
30
40
50
60
Absolute value of allelic effect


Allelic effects associated
with DegreeDays.9_2009 (Training Datasets)
Frequency
0
50
100
150
200
50
100
150
200
250
300
350
400
Absolute value of allelic effect


Allelic effects associated
with DegreeDays.9_2010 (Training Datasets)
Frequency
0
50
100
150
200
20
40
60
80
100
Absolute value of allelic effect

Chapter 3:

Genomic bases of leaf budburst predict countergradient patterns of quantitative genetic variation in European beech

Ivan Scotti , Andrea Modica

INRAE, URFM, Avignon, France

Clarification

This chapter's main focus is the genomic prediction of phenotypic values for populations along the Mont Ventoux northern slope's gradient. A second part, on the prediction of the evolutionary trajectory that populations may follow under climate change, is very preliminary, and used here only to illustrate the potential for evolution provided by within-population genetic variation. The simulation approach used here is very rudimentary, as discussed later in the chapter, and in no way should these simulations be considered as a realistic representation of field populations or of the processes they may undergo in the future.

Summary

Local adaptation proceeds through the adjustment of population-level phenotypic values to local optima. For phenotypes with a genetic control, the selective process involves the selection of the corresponding genotypes, and theory predicts that the underlying genetic loci are selected (in the case of polygenic traits, first in terms of covariation of genotypes at multiple loci, and then in terms of individual allele frequencies) in a way that the resulting phenotypes tend to match phenotypic optima.

European beech (*Fagus sylvatica* L.) shows signs of local adaptation not only over wide geographical distances, but also over steep and short microgeographical gradients. In particular, it displays weak but sizable molecular divergence, as well as divergence for bud-burst time in common gardens, over altitudinal gradients within the same landscape on Mont Ventoux, France (European Union). This suggests that strong selective forces operate in such populations, leading to divergence among populations otherwise homogenised by continued gene flow.

In such conditions, beech stands show *counter-gradient* genetic variation: timing of bud-burst increases with altitude in natural stands (thus reflecting shorter vegetation seasons at higher altitude), bud trees from lower altitudes have a later bud break in common gardens. Based on the results of a previous GWAS experiment and on the genotypes of loci underlying phenotypic traits, we estimate here the breeding values of trees along the Mont Ventoux (France) altitudinal gradient, predicting the countergradient pattern, expected from common garden experiments, for some traits (and in particular, for bud-burst date), but not for all traits related to vegetative phenology. These results show that it is possible to infer local adaptation from genotypic information at quantitative trait loci.

Introduction

The study of the distribution of phenotypic values in wild populations—and of their association with environmental variables—is the staple of the analysis of local adaptation since the foundation of evolutionary biology (Endler, 1977; Ford, 1964). The underlying theory is that genetically controlled characters that have an impact on fitness can be selected for matching the local environmental conditions; through this process, individuals carrying phenotypes that are closer to the optimum have a higher chance to survive and reproduce, thus resulting in stabilising selection favouring such phenotypes. Local adaptation and population divergence are the observable consequence of such differential fitness, and are the empirical patterns that drive the identification of both the process of local adaptation and the characters that undergo local selection.

Plants in general, and forest trees in particular, typically show signs of local adaptation over large latitudinal gradients, particularly in their phenological characters (timing of leaf budburst, flowering, leaf senescence), that are tightly linked to the duration of the vegetation season and have a major impact on growth and allocation of resources to reproduction (Howe *et al.*, 2003; Savolainen, 2011; Alberto *et al.*, 2013; Müller *et al.*, 2015). Selection on phenological traits responds to a double optimisation requirement: extending the growth season as far as possible, both by developing leaves as early, and shedding them as late, as possible—thus pushing for early bud burst and late senescence—and avoiding leaf-damaging (late) frosts, on the one hand, and summer droughts and early autumn chills, on the other hand (which would damage the leaves while still functional, preventing their organised recovery of useful nutrients by the (perennial) plant before leaf shedding)—which would push for late budbreak and early senescence (Battey, 2000; Gömöry and Paule, 2011). Natural selection and plastic responses will thus tend to favour patterns of budbreak and leaf senescence that allow the plant to make the most of their growing season, without unnecessarily investing resources in organs that will be lost to environmental accidents. The results of these trade-offs are complex and fascinating; they involve a large number of characters related to leaf development, function, and protection; they call into action multiple signalling networks that collect environmental cues, particularly in terms of accumulation of winter chill and spring heat and of photoperiod (Heide, 1993; Vitasse *et al.*, 2009; Vitasse and Basler, 2013; Laube *et al.*, 2014), helping the plant to adjust to year-to-year meteorological variation; and finally, because of the large array of processes involved, they have a wide genetic base (Howe *et al.*, 2003), but also a strong plastic component (Vitasse *et al.*, 2010).

In European beech (*Fagus sylvatica* L.), multiple studies have focused on the description of patterns of variation of phenological traits (Gaüzere, Klein, et al., 2020; Gaüzere, Teuf, et al., 2020; Jean et al., 2023; Oddou-Muratorio et al., 2018; Westergren et al., 2023), providing a picture of large phenotypic and genetic variation in adaptive phenological traits, suggesting that beech stands may have relevant adaptive potential. In particular, one salient feature of the adaptation of beech to climatic gradients is that it shows *countergradient* adaptation, that is, populations show patterns in the genetic component of phenological traits, that go the opposite way relative to phenotypic variation as observed in natural settings. These characteristics may have a major effect on the capacity of beech to adapt to climate change, and it has been shown that accounting for genetic variation can greatly improve predictions

on the future distribution of species (Benito Garzón et al., 2019).

In this context, obtaining data on the genetic architecture of phenological traits, and using them to predict the adaptive fate of extant populations, is paramount. The study of how the *genetic component* of trait variation is distributed across locally adapted populations can corroborate the observations made in common gardens on the distribution of *phenotypic* values and link them to selective processes; focusing on the heritable part of variation, and on how it is controlled by genetic loci, modelling resting on knowledge about the genetic architecture of traits can contribute to refine evolutionary predictions. The availability of genomic resources for beech has greatly improved over recent years (Mishra et al., 2022), and this has led to a flourishing of analyses of genomic patterns of adaptation (Lazic et al., 2024; Meger et al., 2021; Modica et al., 2024; Pfenninger et al., 2021), showing that both large-scale and microgeographic environmental variation adaptively shape the distribution of genetic diversity in the species. In parallel, studies of the variation of phenological traits (and their quantitative-genetic bases) haves shown that they can be finely-tuned to micro-geographic conditions (Gaüzere, Klein, et al., 2020), and that in particular the genetic bases of the traits follow countergradient variation, due to the complex interplay of genetic adaptation and plasticity.

Genomic predictions of traits and fitness are increasingly used as tools for the investigation of evolutionary forces in simulations (Cubry et al., 2022), where QTL can be pre-defined based on empirical information on their genomic distribution and effects, or left to arise through mutations. Such simulation studies are useful to explore the effect of variable genetic architectures of traits on the capacity of populations to adapt. In other applications, (Meger et al., 2021; Pfenninger et al., 2021) GWAS results are used to assess the effect of individual loci on the trait, to compute polygenic scores (Khera et al. 2018), and to predict phenotypes. Thanks to the available annotation of the European beech genome, the variants underlying trait variation could, at least in part, be interpreted functionally and in terms of mutational effects. In particular, Meger et al. (2021) focused on the genomic bases of phenological traits and their variability patterns at the range scale (though excluding samples from the northern- and southernmost parts of the distribution) using a candidate gene approach, and found that hundreds of SNPs were associated with phenological characters and with environmental variables. SImilar results were obtained by Postolache et al. (2021), who studied the covariation of candidate gene polymorphisms with environmental variables. Lazic et al. (2024) took a true genomic approach to run a GWAS in a rangewide common garden, but could identify QTL with a consistent effect over successive years, nor could they explain large amounts of phenotypic variability using the significantly associated loci.

In this study, building on the previously identified loci underlying phenological traits (Modica et al. in prep.) and on genomic data collected on natural populations along an altitudinal gradient, we wish to answer the following questions: (a) can we estimate breeding values for individual trees in populations, based on their genotypes at QTL? (b) do estimated breeding values correctly predict the countergradient variation observed at phenotypic traits? (c) can populations evolve and adapt to climate change, resting on their standing genetic variation at QTL?

Materials and methods

Population sampling and genotype data

The population samples used here are the same as in Modica et al. (2024), with the addition of 25 samples from each of two further plots, N1 and N2 (Gaüzere, Klein, et al., 2020). Molecular-biology and bioinformatics methods are the same as reported in Modica et al. (2024).

Estimation of individual breeding values

To obtain individual breeding values, the list of significant loci (SNP) obtained in replicate subsets of a GWAS analysis of multiple phenology-related traits (Modica et al. in prep.) was matched to the list of SNP loci genotyped in stands N3, N4, S1, S5 (Modica et al., 2024) as well as N1 and N2 (see above). N1 through N4 stands are located on the north face of the mountain (N1 the lowest, N4 the highest), and S1 and S5 on the south face (S5 being the highest stand, and S1 the lowest). A variable number of loci, depending on the trait, was found to be shared between the GWAS and the population-genetic analyses (Table 1); genotypes at these loci were combined with allelic effects obtained in the GWAS study (Modica et al. in prep.) to obtain breeding values as follows:

$G_i = \sum g_{i,l} a_l$

with G_i the breeding value of individual i, $g_{i,l} = \{0,1,2\}$, the number of alternate alleles at locus l for individual i, and a_l the allelic effect of the alternate allele on the character; that is, breeding values were computed under the assumption of pure additivity both within locus (*ie.*, co-dominance) and over loci (no epistatic interactions).

Differences in the distribution of breeding values between populations were tested, separately for populations on each face of the mountain, using a non-parametric Wilcoxon test.

Prediction of evolutionary potential

To assess whether the genetic component of phenotype could evolve under climate change, and to assess whether such adaptive process may erode evolutionary potential, we simulated a population with an optimum phenotypic value corresponding to the mean breeding value of N4 trees (stabilising selection phase); once the simulated population's mean phenotype had stabilised around the optimum value, we instantaneously changed the optimum phenotypic value to the mean breeding value of N2 trees (directional selection phase), thus simulating a sudden warming of climate and consequently the need to adjust to a new optimum (assuming that current climate at lower altitudes will become the future climate at higher altitudes).

The simulations were carried out using SLiM 3 (Haller & Messer, 2019), and were repeated

100 times. At the end of both the stabilising selection phase and the directional selection phase, allele frequencies at the QTL and mean phenotypes were recorded. Single-locus and mean expected heterozygosity were computed and compared between the two temporal phases, as well as with the empirical data, and a principal component analysis was performed to compare the genetic composition of the two simulated temporal data sets to each other and to the empirical population (stand N4). The simulation approach was applied only to the character “date of BBCH 07” for year 2009 (see Modica et al., in prep., for details on trait definition). See Supplementary methods SM1 for the code.

For comparing levels of diversity and computing genetic distances between simulations and empirical populations, a sample of 25 individuals was taken from the simulated populations and used to compute the statistics. This way, the sampling variance that affects the estimation of allele frequencies is the same in the *in natura* and in the *in silico* populations. Based on the allele frequencies in the 25-individual samples, we computed expected heterozygosity at each QTN and the mean over all the loci, and carried out a Principal component analysis including all the simulations and the N4 population, using the R package ade4 (Dray & Dufour, 2007). Comparisons of expected heterozygosity between *in silico* populations after the stabilising selection phase and after the directional selection phase was carried out on the entire populations, using a t-test after logit transformation of the expected heterozygosity values.

Results

Number of shared SNP between the GWAS (Modica et al., in prep.) and the population-genetic study (Modica et al., 2024 and this study) varied between 1 and 449 depending on the character, the year of observation, and the FDR level chosen for the GWAS (**Table 1**).

Individual breeding values were computed for all combinations of trait, year, and FDR. The distribution for trait BBCH 7, year 2009, FDR = 5% is represented in **Figure 1**, and those of the remaining combinations are shown in Supplementary Figure SF1. **Table 2** shows the *P*-values of pairwise Wilcoxon tests (not corrected for multiple testing) for the same data, with the remaining tests reported in Supplementary Table ST1.

With the exception of population N1, populations at higher elevations tend to have a lower breeding value for BBCH 07 date in 2009: at a nominal threshold $\alpha = 0.05$, four comparisons out of six for the north-slope stands lend support to the *countergradient* hypothesis, while one out of six lends support to the *gradient* hypothesis (the south slope comparison did not reject identity of medians) (**Table 1**). This trend is visible in **Figure 1**. Similar trends are apparent for date of BBCH 09, which is correlated with date for BBCH 07, and for Degree Days for BBCH 07 and BBCH 09, with the combination of Degree Days for BBCH 09, year 2009, FDR = 5% showing four tests in favour of countergradient variation, and one in favour of co-gradient variation. PSS shows an opposed trend (thus covarying with the gradient) and parameters of the logistic curve show no clear pattern.

Globally, combining all traits, years, and FDR levels, twenty tests lean toward a countergradient variation, and five favour co-gradient variation.

Simulations of a single population of 5000 individuals reached a mean phenotype value close to the optimum (i.e. the mean breeding value predicted for N4) in fewer than 300 generations and remained stable thereafter. However, allele frequencies in simulated populations differed from the empirical population's frequencies, as shown by a PCA based on allele frequencies at the 23 QTL (**Figure 2**); the level of diversity shown by the empirical sample of the N4 population also differs from simulations, lying at the lower end of the distribution of values of mean expected heterozygosity shown by samples of simulated populations (**Figure 3**).

After change in selection regime, instantly moving the optimum from one corresponding to the mean estimated breeding value of population N4 to mean breeding value of population N2, the phenotypic mean of the simulated population responded to selection and moved to the new optimum in less than 300 generations, then stabilising again.

The bout of directional selection was associated with a major loss of genetic diversity, as estimated by expected heterozygosity (Student t-test, $t = 21.909$, $df = 116.09$, p-value < 2.2e-16; **Figure 4**).

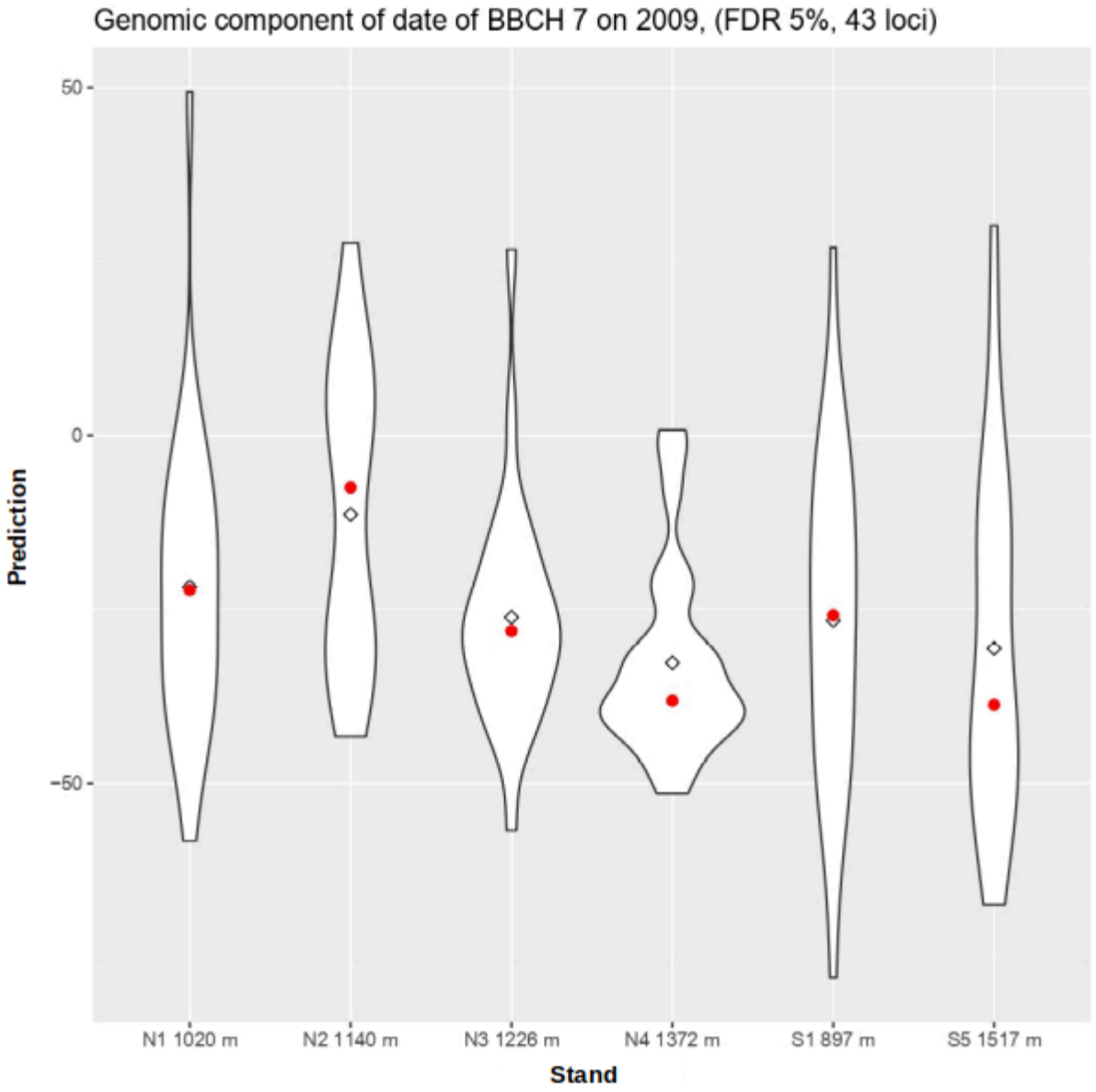


Figure 1. Distribution of breeding values by population for character BBCH 7 (Day of appearance of stage BBCH 7), year 2009, FDR 5%. The absolute breeding values are arbitrary, as the allelic effects may have biases and inflation, due to the small sample size use in GWAS (Modica et al., in prep.)

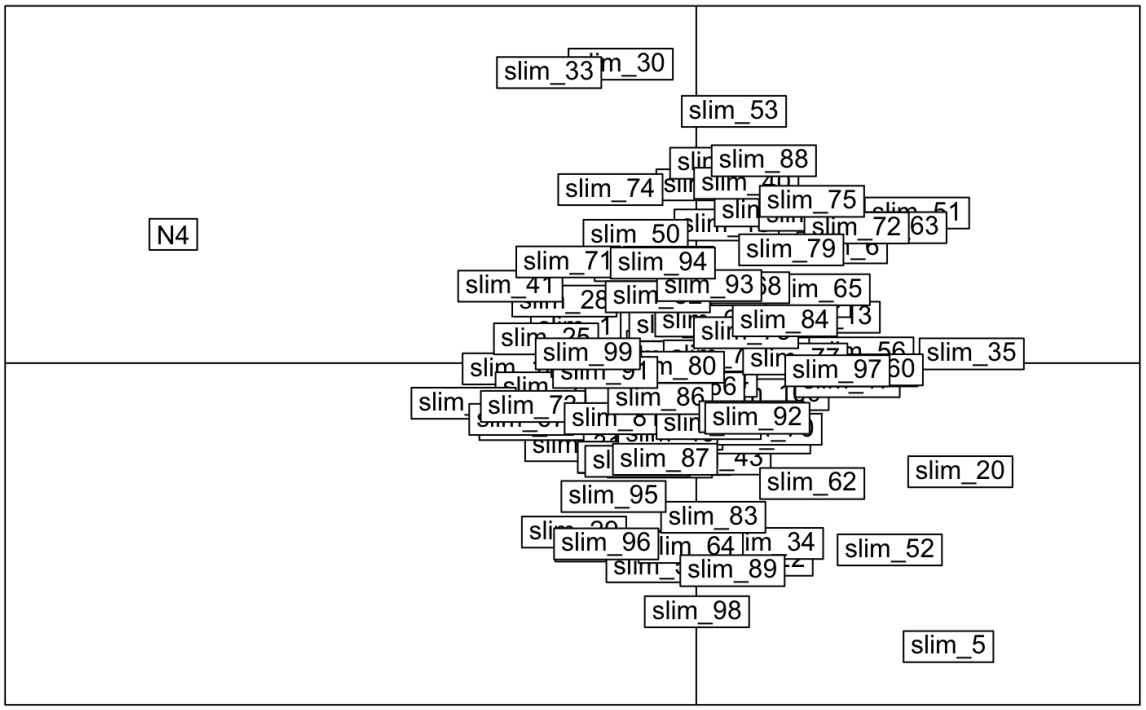


Figure 2. Principal Component Analysis (PCA) of the 100 populations simulated using SLiM (slim_1 through slim_100) and of empirical population N4, based on allele frequencies at QTL.

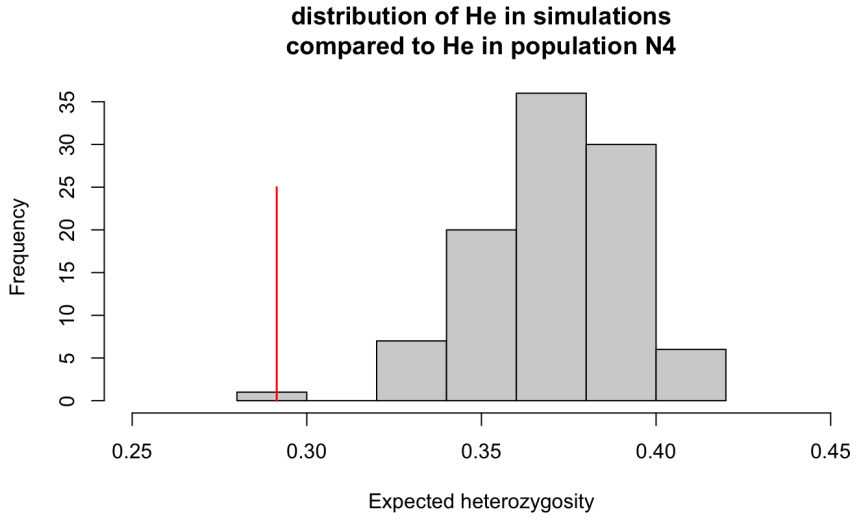


Figure 3. Distribution of values of mean expected heterozygosity (computed over the 23 QTN) in samples of simulated data sets (histogram) and in the empirical sample from the N4 population (red bar).

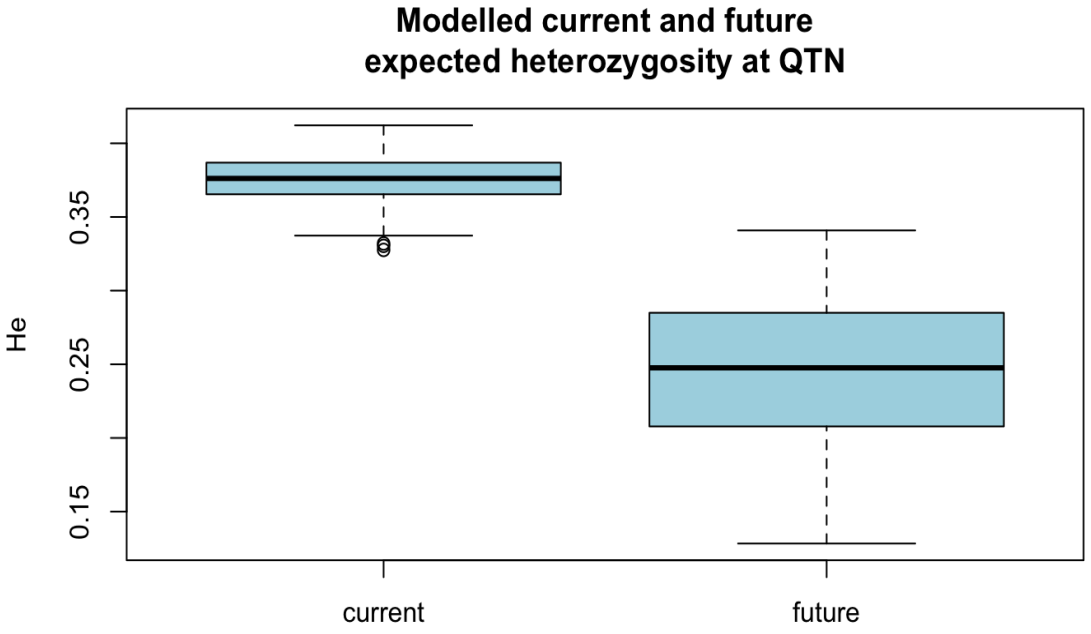


Figure 4: Comparison of expected heterozygosity levels in simulated populations after the first round of selection (stabilising selection; “current” population) versus after the second round (directional selection; “future” population).

Trait	Year	FDR 5%	FDR 10%
BBCH 7	2009	23	102
BBCH 7	2010	37	213
BBCH 9	2009	43	189
BBCH 9	2010	101	443
a (slope logistic)	2009	15	81
a (slope logistic)	2010	195	752
b (inflection logistic)	2009	1	28
b (inflection logistic)	2010	41	222
PSS	2009	4	62
PSS	2010	43	222
Degree Days 7	2009	17	58
Degree Days 7	2010	120	449
Degree Days 9	2009	32	98
Degree Days 9	2010	71	322

Table 1: Number of SNP shared between the GWAS in Modica et al. (in prep.) and the population-genetic data (Modica et al., 2024 and this study). BBCH 07, BBCH 09: dates of observation of developmental stages; a, b: parameters of the logistic curve fitted to the observed data (developmental stage as a function of time; see Modica et al. (in prep.) for details); PSS: Phenotypic Sum Score (Bontemps et al., 2017); Degree Days 7, Degree Days 9: sum of Degree Days until developmental stage 7, 9.

(a)

	N2	N3	N4	S5
N1	0.96	0.27	**0.014**	
N2		**0.015**	**0.00004**	
N3			**0.026**	
S1				0.22

(b)

	N2	N3	N4	S5
N1	**0.034**	0.73	0.99	
N2		0.99	1	
N3			0.97	
S1				0.78

Table 2: P-values of Wilcoxon tests for character BBCH 07, year 2009, predictions at FDR = 5% for (a) *countergradient* variation (i.e., earlier budburst for populations at higher altitudes) and for (b) *gradient* variation (i.e., earlier budburst for populations at lower altitudes)

Discussion

The breeding values, predicted based on the allelic effects estimated by GWAS and the allele frequencies detected in samples of populations, show altitudinal trends consistent with the countergradient selection pattern observed by Gaüzere et al. (2020). Interestingly, the phenotypes used here were measured the same year that the seeds, used by that study, were collected in the very same stands. It is worth noticing that, in the Gaüzere et al. (2020) study, the phenotypic mean of BBCH 07 in population N2 was indistinguishable from N4's, and differed from N1, while here N2 differs from N4 and N1 is similar to N4. While these differences cannot be overlooked, the general trend observed here over populations and characters confirms countergradient variation of growth phenology-associated characters, despite the very partial representation of the characters' genetic control: indeed, the starting point of this analysis is an in situ GWAS with relatively low power and potentially numerous sources of noise, which blur the signal; on top of this, the dataset used here for the genomic prediction has been "repurposed" from a previous study (Scotti et al. 2023) and only consists of transcriptome-based variants—thus excluding all intergenic variants; and finally, the population sample is small (only twenty to twenty-five trees), limiting the precision of the estimated per-population phenotypic mean. Despite all these downsides, our analysis still managed to capture a biological signal that agrees with previous, more precise estimations of the phenotypic gradient. This suggests that the signal is strong enough, and the method robust enough, to produce meaningful results even in very limiting conditions.

The study carried out by Gaüzère et al. (2020) proved, using maternal half-sib families and controlled conditions, that budbreak in seedlings follows a countergradient and is heritable. This very clear result is now corroborated by the present study, which provides an *in situ*-only chain of proof going from phenological observations to identification of the genetic control of traits to prediction of trait values in unobserved trees. The result is inevitably noisy, and this analysis should benefit from field observations to confirm the prediction. Unfortunately, this is inaccessible, because the "predictions" are only valid for the years the phenotypes were observed (2009 and 2010), preventing field validation (the GWAS performed on the field data showed that QTL are year-specific, thus making entirely unsafe to use contemporary observations to validate predictions of past phenotypes).

This is a major step towards understanding how selection operates in natural beech populations and paves the way to a more precise assessment of the relative role of the genetic bases of traits and plastic responses to the environment. Indeed, the fact that countergradient variation is hard-wired in the genetic setup of populations suggests that selection acts by compensating the effect of plastic responses to environmental cues: selection may favour advancement of flushing dates at higher altitudes because the plastic response delays it so much that it excessively reduces the duration of the growing season and, consequently, growth. Plastic response would therefore be partially maladaptive, and selection would favour genotypes that balance this maladaptive response. Building on the availability of genomic data, including a full genome sequence (Mishra et al., 2022) and resequencing data, and on phenotype data that can be collected in long-term common gardens (Robson et al., 2018), it will be possible in the near future to test this hypothesis. It is worth noticing that the breeding value predictions reported here are quite rudimentary, as they rest on a GWAS experiment with relatively low power (Modica et al. in prep.) and on population-genomic data that only partially overlap the loci controlling the traits and have themselves low power due to sample size. Notwithstanding these limitations, we succeeded

in highlighting the countergradient patterns of variation expected based on previous literature (Gaüzere, Klein, et al., 2020; Vitasse et al., 2010).

The relatively incomplete array of loci and allelic effects that we considered here also allowed us to model the effect of selection on the genetic setup of populations, in a very preliminary way. Assuming that the currently observed phenotypic mean of population matches their optimum phenotype, and that climate change will shift such optimum, we simulated one population at equilibrium with the current environmental conditions, corresponding to population N4, and then applied a sudden change of conditions, mimicking climate change moving the N4 optimum to the N3 optimum. In our simulation, we observed that the population can and does adapt to the new conditions, yet loses significant amounts of diversity in doing so. This modelling exercise therefore sends mixed signals relative to the fate of populations under climate change.

Notice, though, that our simulations are extremely crude, and should only be taken as a proof-of-concept of the adaptive potential carried by the variability of loci underlying the traits. Indeed, several limitations affect the current simulations: (i) they were carried out with a single population, ignoring the presence of neighbouring populations that are possibly preadapted to the future climate; these neighbouring populations may contribute alleles facilitating the selection of favourable genotypes, thus making adaptation much faster than what we observe here (and in agreement with the observation that past adaptation has occurred over just a few generations, Modica et al., 2024), and sustained gene flow would counteract the loss of diversity observed here; (ii) the initialisation of gene frequencies was very crude: we randomly drew alleles to build individual genotypes, based on allele frequencies observed in the empirical sample. This is obviously very imprecise, as it can generate multi-locus genotypes from which totally implausible phenotype values can arise; in addition, this approach entirely ignores covariation of allele frequencies across loci. Possible fixes would be to constrain genotypes based on the observed phenotypic mean and variance; to draw alleles taking (linkage) disequilibrium into account; to initialise the population starting from the empirical sample of genotypes; (iii) considering that some amount of genetic lag is expected in more complex modelling settings (Cubry et al., 2022), the currently observed phenotypic means may not be considered as the true optima. For all these reasons, the phenotypic and genetic diversity dynamics observed in the simulations cannot at all be trusted; this preliminary simulation study is only an empirical test that the phenotypic mean, as obtained through genomic prediction, can respond to selection. This preliminary study shows that it is possible to model the evolution of important adaptive traits in wild populations, and possibly to provide guidelines to foster a forest management approach that favours adaptability to new conditions (Lefèvre et al., 2014).

In conclusion, we show here that “A to Z *in situ*” genomic predictions (resting only on *in situ* phenotypes, genotypes, and analyses) can effectively represent biologically sound differentiation patterns and adaptive processes, and that populations predicted based on these results show relevant adaptive potential. Further development (and validation) of predictions and modelling approaches will make this promising strategy more precise, and will allow us to inform management decisions oriented towards facilitating the adaptation of forests to future climatic challenges.

References

Benito Garzón, M., Robson, T. M., & Hampe, A. (2019). ΔTraitSDMs: Species distribution models that account for local adaptation and phenotypic plasticity. *New Phytologist*, *222*(4), 1757–1765. https://doi.org/10.1111/nph.15716

Bontemps, A., Davi, H., Lefèvre, F., Rozenberg, P., & Oddou-Muratorio, S. (2017). How do functional traits syndromes covary with growth and reproductive performance in a water-stressed population of *Fagus sylvatica*? *Oikos*, *126*(10), 1472–1483. https://doi.org/10.1111/oik.04156

Cubry, P., Oddou‐Muratorio, S., Scotti, I., & Lefèvre, F. (2022). Interactions between microenvironment, selection and genetic architecture drive multiscale adaptation in a simulation experiment. *Journal of Evolutionary Biology*, *35*(3), 451–466. https://doi.org/10.1111/jeb.13988

Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*, 1–20. https://doi.org/10.18637/jss.v022.i04

Ford EB (1964). Ecological genetics. Chapman & Hall: London

Gaüzere, J., Klein, E. K., Brendel, O., Davi, H., & Oddou-Muratorio, S. (2020). Microgeographic adaptation and the effect of pollen flow on the adaptive potential of a temperate tree species. *New Phytologist*, *227*(2), 641–653. https://doi.org/10.1111/nph.16537

Gaüzere, J., Teuf, B., Davi, H., Chevin, L.-M., Caignard, T., Leys, B., Delzon, S., Ronce, O., & Chuine, I. (2020). Where is the optimum? Predicting the variation of selection along climatic gradients and the adaptive value of plasticity. A case study on tree phenology. *Evolution Letters*, *4*(2), 109–123. https://doi.org/10.1002/evl3.160

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Molecular Biology and Evolution*, *36*(3), 632–637. https://doi.org/10.1093/molbev/msy228

Heide OM (1993). Daylength and thermal time responses of budburst during dormancy release in some northern deciduous trees. *Physiol Plant* **88**: 531–540.

Jean, F., Davi, H., Oddou-Muratorio, S., Fady, B., Scotti, I., Scotti-Saintagne, C., Ruffault, J., Journe, V., Clastre, P., Marloie, O., Brunetto, W., Correard, M., Gilg, O., Pringarbe, M., Rei, F., Thevenet, J., Turion, N., & Pichot, C. (2023). A 14-year series of leaf phenological data collected for European beech (Fagus sylvatica L.) and silver fir (Abies alba Mill.) from their geographic range margins in south-eastern France. *Annals of Forest Science*, *80*(1), 35. https://doi.org/10.1186/s13595-023-01193-9

Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**: 1219–1224.

Lazic, D., Geßner, C., Liepe, K. J., Lesur-Kupin, I., Mader, M., Blanc-Jolivet, C., Gömöry, D.,

Liesebach, M., González-Martínez, S. C., Fladung, M., Degen, B., & Müller, N. A. (2024). Genomic variation of European beech reveals signals of local adaptation despite high levels of phenotypic plasticity. *Nature Communications*, *15*(1), 8553. https://doi.org/10.1038/s41467-024-52933-y

Laube J, Sparks TH, Estrella N, Höfler J, Ankerst DP, Menzel A (2014). Chilling outweighs photoperiod in preventing precocious spring development. *Glob Change Biol* **20**: 170–182.

Lefèvre, F., Boivin, T., Bontemps, A., Courbet, F., Davi, H., Durand-Gillmann, M., Fady, B., Gaüzere, J., Gidoin, C., Karam, M. J., Lalagüe, H., Oddou-Muratorio, S., & Pichot, C. (2014). Considering evolutionary processes in adaptive forestry. *Annals of Forest Science*, *71*(7), 723–739. https://doi.org/10.1007/s13595-013-0272-1

Meger, J., Ulaszewski, B., & Burczyk, J. (2021). Genomic signatures of natural selection at phenology-related genes in a widely distributed tree species *Fagus sylvatica* L. *BMC Genomics 2021 22:1*, *22*(1), 1–20. https://doi.org/10.1186/S12864-021-07907-5

Mishra, B., Ulaszewski, B., Meger, J., Aury, J.-M., Bodénès, C., Lesur-Kupin, I., Pfenninger, M., Da Silva, C., Gupta, D. K., Guichoux, E., Heer, K., Lalanne, C., Labadie, K., Opgenoorth, L., Ploch, S., Le Provost, G., Salse, J., Scotti, I., Wötzel, S., … Thines, M. (2022). A Chromosome-Level Genome Assembly of the European Beech (Fagus sylvatica) Reveals Anomalies for Organelle DNA Integration, Repeat Content and Distribution of SNPs. *Frontiers in Genetics*, *12*, 691058. https://doi.org/10.3389/fgene.2021.691058

Modica, A., Lalagüe, H., Muratorio, S., & Scotti, I. (2024). Rolling down that mountain: Microgeographical adaptive divergence during a fast population expansion along a steep environmental gradient in European beech. *Heredity*, 1–14. https://doi.org/10.1038/s41437-024-00696-z

Oddou-Muratorio, S., Gaüzere, J., Bontemps, A., Rey, J. F., & Klein, E. K. (2018). Tree, sex and size: Ecological determinants of male vs. Female fecundity in three Fagus sylvatica stands. *Molecular Ecology*, *27*(15), 3131–3145. https://doi.org/10.1111/mec.14770

Postolache D, Oddou-Muratorio S, Vajana E, Bagnoli F, Guichoux E, Hampe A, et al. (2021). Genetic signatures of divergent selection in European beech (Fagus sylvatica L.) are associated with the variation in temperature and precipitation across its distribution range. *Molecular Ecology* **30**: 5029–5047.

Pfenninger, M., Reuss, F., Klebler, A., Schönnenbeck, P., Caliendo, C., Gerber, S., Cocchiararo, B., Reuter, S., Blüthgen, N., Mody, K., Mishra, B., Bálint, M., Thines, M., & Feldmeyer, B. (2021). Genomic basis for drought resistance in European beech forests threatened by climate change. *eLife*, *10*, e65532. https://doi.org/10.7554/elife.65532

Robson, T. M., Garzón, M. B., Alia Miranda, R., Barba Egido, D., Bogdan, S., Borovics, A., Božič, G., Brendel, O., Clark, J., De Vries, S. M. G., Delehan, I. I., Ducousso, A., Fady, B., Fennessy, J., Forstreuter, M., Frýdl, J., Geburek, T., Gömöry, D.,

Hauke-Kowalska, M., … Wesoły, W. (2018). Phenotypic trait variation measured on European genetic trials of Fagus sylvatica L. *Scientific Data*, *5*, 180149. ttps://doi.org/10.1038/sdata.2018.149

Vitasse Y, Basler D (2013). What role for photoperiod in the bud burst phenology of European beech. *Eur J For Res* **132**: 1–8.

Vitasse, Y., Bresson, C. C., Kremer, A., Michalet, R., & Delzon, S. (2010). Quantifying phenological plasticity to temperature in two temperate tree species. *Functional Ecology*, *24*(6), 1211–1218. https://doi.org/10.1111/j.1365-2435.2010.01748.x

Westergren, M., Archambeau, J., Bajc, M., Damjanić, R., Theraroz, A., Kraigher, H., Oddou-Muratorio, S., & González-Martínez, S. C. (2023). Low but significant evolutionary potential for growth, phenology and reproduction traits in European beech. *Molecular Ecology*, *n/a*(n/a). https://doi.org/10.1111/mec.17196

Supplementary Material

Supplementary Methods SM1: SLiM script

```
// Keywords: migration, dispersal, QTL, quantitative trait loci

initialize() {
      initializeMutationRate(1e-11);  // low mut rate, to minimise the
burden of neutral variants
      // rec rates & chrom breaks
      rates=c(1e-8, 0.5, 1e-8, 0.5, 1e-8, 0.5, 1e-8, 0.5, 1e-8, 0.5, 1e-8,
0.5, 1e-8, 0.5, 1e-8);

ends=c(4999999,5000000,36999999,37000000,43999999,44000000,97999999,98000000,
122999999,123000000,
      166499999,166500000,204999999,205000000,220500000);                 //
not "true" chromosomes, see RMD file
      initializeRecombinationRate(rates,ends);
      // neutral mutations in non-coding regions
      initializeMutationType("m1", 0.5, "f", 0.0);     // neutral
//
//
      // mutations representing alleles in QTLs
                  // not used:
                  //scriptForQTLs = "if (runif(1) < 0.5) -1; else 1;";
      // changed to accommodate my mutations:
      // one mutation type per QTN:
      initializeMutationType("m2", 0.5, "f", 4.15263736509172); // this is
QTN 1
      m2.convertToSubstitution = F;
      m2.mutationStackPolicy = "l";
      m2.color = "red";  // QTN mutations are marked in red
      initializeMutationType("m3", 0.5, "f", -13.2893793189314); // this is
QTN 2
      m3.convertToSubstitution = F;
      m3.mutationStackPolicy = "l";
      m3.color = "red";  // QTN mutations are marked in red
      initializeMutationType("m4", 0.5, "f", -13.2893793189314); // this is
QTN 3
      m4.convertToSubstitution = F;
      m4.mutationStackPolicy = "l";
      m4.color = "red";  // QTN mutations are marked in red
      initializeMutationType("m5", 0.5, "f", 4.54276804881715); // this is
QTN 4
      m5.convertToSubstitution = F;
      m5.mutationStackPolicy = "l";
      m5.color = "red";  // QTN mutations are marked in red
      initializeMutationType("m6", 0.5, "f", 5.05507627554576); // this is
QTN 5
      m6.convertToSubstitution = F;
      m6.mutationStackPolicy = "l";
      m6.color = "red";  // QTN mutations are marked in red
      initializeMutationType("m7", 0.5, "f", 4.96247269325974); // this is
QTN 6
      m7.convertToSubstitution = F;
      m7.mutationStackPolicy = "l";
      m7.color = "red";  // QTN mutations are marked in red
      initializeMutationType("m8", 0.5, "f", 4.96247269325974); // this is
QTN 7
      m8.convertToSubstitution = F;
      m8.mutationStackPolicy = "l";
```

```
      m8.color = "red";  // QTN mutations are marked in red
      initializeMutationType("m9", 0.5, "f", 4.58830862571148); // this is
QTN 8
      m9.convertToSubstitution = F;
      m9.mutationStackPolicy = "l";
      m9.color = "red";  // QTN mutations are marked in red
      initializeMutationType("m10", 0.5, "f", 5.02684381507689); // this is
QTN 9
      m10.convertToSubstitution = F;
      m10.mutationStackPolicy = "l";
      m10.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m11", 0.5, "f", -4.32484133512907); // this is
QTN 10
      m11.convertToSubstitution = F;
      m11.mutationStackPolicy = "l";
      m11.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m12", 0.5, "f", 4.32484133512912); // this is
QTN 11
      m12.convertToSubstitution = F;
      m12.mutationStackPolicy = "l";
      m12.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m13", 0.5, "f", 5.66703293686413); // this is
QTN 12
      m13.convertToSubstitution = F;
      m13.mutationStackPolicy = "l";
      m13.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m14", 0.5, "f", 5.66703293686413); // this is
QTN 13
      m14.convertToSubstitution = F;
      m14.mutationStackPolicy = "l";
      m14.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m15", 0.5, "f", 5.49605751740467); // this is
QTN 14
      m15.convertToSubstitution = F;
      m15.mutationStackPolicy = "l";
      m15.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m16", 0.5, "f", 5.49605751740467); // this is
QTN 15
      m16.convertToSubstitution = F;
      m16.mutationStackPolicy = "l";
      m16.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m17", 0.5, "f", 5.01905185415728); // this is
QTN 16
      m17.convertToSubstitution = F;
      m17.mutationStackPolicy = "l";
      m17.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m18", 0.5, "f", 4.20007009177104); // this is
QTN 17
      m18.convertToSubstitution = F;
      m18.mutationStackPolicy = "l";
      m18.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m19", 0.5, "f", -4.83860362183782); // this is
QTN 18
      m19.convertToSubstitution = F;
      m19.mutationStackPolicy = "l";
      m19.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m20", 0.5, "f", -3.98590533697532); // this is
QTN 19
      m20.convertToSubstitution = F;
      m20.mutationStackPolicy = "l";
      m20.color = "red"; // QTN mutations are marked in red
```

```
      initializeMutationType("m21", 0.5, "f", -3.95651186144391); // this is
QTN 20
      m21.convertToSubstitution = F;
      m21.mutationStackPolicy = "l";
      m21.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m22", 0.5, "f", -4.11515055783445); // this is
QTN 21
      m22.convertToSubstitution = F;
      m22.mutationStackPolicy = "l";
      m22.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m23", 0.5, "f", -4.04716401435689); // this is
QTN 22
      m23.convertToSubstitution = F;
      m23.mutationStackPolicy = "l";
      m23.color = "red"; // QTN mutations are marked in red
      initializeMutationType("m24", 0.5, "f", -4.43554686481838); // this is
QTN 23
      m24.convertToSubstitution = F;
      m24.mutationStackPolicy = "l";
      m24.color = "red"; // QTN mutations are marked in red
// Genomic element types
      initializeGenomicElementType("g1", m1, 1.0);
      initializeGenomicElementType("g2", m2, 1.0);
      initializeGenomicElementType("g3", m3, 1.0);
      initializeGenomicElementType("g4", m4, 1.0);
      initializeGenomicElementType("g5", m5, 1.0);
      initializeGenomicElementType("g6", m6, 1.0);
      initializeGenomicElementType("g7", m7, 1.0);
      initializeGenomicElementType("g8", m8, 1.0);
      initializeGenomicElementType("g9", m9, 1.0);
      initializeGenomicElementType("g10", m10, 1.0);
      initializeGenomicElementType("g11", m11, 1.0);
      initializeGenomicElementType("g12", m12, 1.0);
      initializeGenomicElementType("g13", m13, 1.0);
      initializeGenomicElementType("g14", m14, 1.0);
      initializeGenomicElementType("g15", m15, 1.0);
      initializeGenomicElementType("g16", m16, 1.0);
      initializeGenomicElementType("g17", m17, 1.0);
      initializeGenomicElementType("g18", m18, 1.0);
      initializeGenomicElementType("g19", m19, 1.0);
      initializeGenomicElementType("g20", m20, 1.0);
      initializeGenomicElementType("g21", m21, 1.0);
      initializeGenomicElementType("g22", m22, 1.0);
      initializeGenomicElementType("g23", m23, 1.0);
      initializeGenomicElementType("g24", m24, 1.0);
// genomic elements
      initializeGenomicElement(g1, 0, 4264619);
      initializeGenomicElement(g2, 4264620, 4264620);
      initializeGenomicElement(g1, 4264621, 36800944);
      initializeGenomicElement(g3, 36800945, 36800945);
      initializeGenomicElement(g1, 36800946, 36801016);
      initializeGenomicElement(g4, 36801017, 36801017);
      initializeGenomicElement(g1, 36801018, 43973961);
      initializeGenomicElement(g5, 43973962, 43973962);
      initializeGenomicElement(g1, 43973963, 64325363);
      initializeGenomicElement(g6, 64325364, 64325364);
      initializeGenomicElement(g1, 64325365, 64326667);
      initializeGenomicElement(g7, 64326668, 64326668);
      initializeGenomicElement(g1, 64326669, 64326698);
      initializeGenomicElement(g8, 64326699, 64326699);
      initializeGenomicElement(g1, 64326700, 64326827);
```

```
      initializeGenomicElement(g9, 64326828, 64326828);
      initializeGenomicElement(g1, 64326829, 64329733);
      initializeGenomicElement(g10, 64329734, 64329734);
      initializeGenomicElement(g1, 64329735, 74068481);
      initializeGenomicElement(g11, 74068482, 74068482);
      initializeGenomicElement(g1, 74068483, 74068541);
      initializeGenomicElement(g12, 74068542, 74068542);
      initializeGenomicElement(g1, 74068543, 97665867);
      initializeGenomicElement(g13, 97665868, 97665868);
      initializeGenomicElement(g1, 97665869, 97666005);
      initializeGenomicElement(g14, 97666006, 97666006);
      initializeGenomicElement(g1, 97666007, 97688907);
      initializeGenomicElement(g15, 97688908, 97688908);
      initializeGenomicElement(g1, 97688909, 97688934);
      initializeGenomicElement(g16, 97688935, 97688935);
      initializeGenomicElement(g1, 97688936, 122593004);
      initializeGenomicElement(g17, 122593005, 122593005);
      initializeGenomicElement(g1, 122593006, 134696337);
      initializeGenomicElement(g18, 134696338, 134696338);
      initializeGenomicElement(g1, 134696339, 166071343);
      initializeGenomicElement(g19, 166071344, 166071344);
      initializeGenomicElement(g1, 166071345, 200272537);
      initializeGenomicElement(g20, 200272538, 200272538);
      initializeGenomicElement(g1, 200272539, 204394772);
      initializeGenomicElement(g21, 204394773, 204394773);
      initializeGenomicElement(g1, 204394774, 204751201);
      initializeGenomicElement(g22, 204751202, 204751202);
      initializeGenomicElement(g1, 204751203, 220347999);
      initializeGenomicElement(g23, 220348000, 220348000);
      initializeGenomicElement(g1, 220348001, 220348177);
      initializeGenomicElement(g24, 220348178, 220348178);
      initializeGenomicElement(g1, 220348179, 220500000);
                  // not used:
                  // set up our chromosome: 10 QTLs, surrounded by neutral
regions
                  //    defineConstant("C", 10);    // number of QTLs
                  //    defineConstant("W", 1000);  // size of neutral
buffer on each side
                  //    pos = 0;
                  //    q = NULL;
                  //
                  //    for (i in 1:C)
                  //    {
                  //          initializeGenomicElement(g1, pos, pos + W-1);
                  //          pos = pos + W;
                  //
                  //          initializeGenomicElement(g2, pos, pos);
                  //          q = c(q, pos);
                  //          pos = pos + 1;
                  //
                  //          initializeGenomicElement(g1, pos, pos + W-1);
                  //          pos = pos + W;
                  //    }
                  //
      defineConstant("Q", c(4264620,36800945,36801017,43973962,64325364,

64326668,64326699,64326828,64329734,74068482,

74068542,97665868,97666006,97688908,97688935,

122593005,134696338,166071344,200272538,204394773,
```

```
204751202,220348000,220348178));     // remember our QTL positions
                    //
                          // we want the QTLs to be unlinked; build a
recombination map for that
                    //    rates = c(rep(c(1e-8, 0.5), C-1), 1e-8);
                    //    ends = (repEach(Q + W, 2) + rep(c(0,1), C))[0:(C*2 -
2)];
                    //    initializeRecombinationRate(rates, ends);
}
// modelling one (large) population
1 late() { sim.addSubpop("p1", 5000);
//
                    // not used: second population & migration parameters
                    //    sim.addSubpop("p2", 500);

                    // set up migration; comment these out for zero gene flow
                    //    p1.setMigrationRates(p2, 0.01);
                    //    p2.setMigrationRates(p1, 0.01);

// replicate the s1 event to output in tick 2 also
      community.registerEarlyEvent("s2", s1.source, 2, 2);
//
// optional: give m2 mutations to everyone, as standing variation
g = sim.subpopulations.genomes;
// changed assigned values from +1, -1 to (allelic effect), 0 (unsure this is
right, though)
for (q in Q[0])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m2, +4.15263736509172, q);
g[!isPlus].addNewMutation(m2, 0.0, q);
}
for (q in Q[1])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m3, -13.2893793189314, q);
g[!isPlus].addNewMutation(m3, 0.0, q);
}
for (q in Q[2])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m4, -13.2893793189314, q);
g[!isPlus].addNewMutation(m4, 0.0, q);
}
for (q in Q[3])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m5, 4.54276804881715, q);
g[!isPlus].addNewMutation(m5, 0.0, q);
}
for (q in Q[4])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m6, 5.05507627554576, q);
g[!isPlus].addNewMutation(m6, 0.0, q);
}
for (q in Q[5])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m7, 4.96247269325974, q);
```

```
g[!isPlus].addNewMutation(m7, 0.0, q);
}
for (q in Q[6])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m8, 4.96247269325974, q);
g[!isPlus].addNewMutation(m8, 0.0, q);
}
for (q in Q[7])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m9, 4.58830862571148, q);
g[!isPlus].addNewMutation(m9, 0.0, q);
}
for (q in Q[8])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m10, 5.02684381507689, q);
g[!isPlus].addNewMutation(m10, 0.0, q);
}
for (q in Q[9])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m11, -4.32484133512907, q);
g[!isPlus].addNewMutation(m11, 0.0, q);
}
for (q in Q[10])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m12, 4.32484133512912, q);
g[!isPlus].addNewMutation(m12, 0.0, q);
}
for (q in Q[11])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m13, 5.66703293686413, q);
g[!isPlus].addNewMutation(m13, 0.0, q);
}
for (q in Q[12])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m14, 5.66703293686413, q);
g[!isPlus].addNewMutation(m14, 0.0, q);
}
for (q in Q[13])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m15, 5.49605751740467, q);
g[!isPlus].addNewMutation(m15, 0.0, q);
}
for (q in Q[14])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m16, 5.49605751740467, q);
g[!isPlus].addNewMutation(m16, 0.0, q);
}
for (q in Q[15])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m17, 5.01905185415728, q);
g[!isPlus].addNewMutation(m17, 0.0, q);
```

```
}
for (q in Q[16])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m18, 4.20007009177104, q);
g[!isPlus].addNewMutation(m18, 0.0, q);
}
for (q in Q[17])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m19, -4.83860362183782, q);
g[!isPlus].addNewMutation(m19, 0.0, q);
}
for (q in Q[18])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m20, -3.98590533697532, q);
g[!isPlus].addNewMutation(m20, 0.0, q);
}
for (q in Q[19])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m21, -3.95651186144391, q);
g[!isPlus].addNewMutation(m21, 0.0, q);
}
for (q in Q[20])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m22, -4.11515055783445, q);
g[!isPlus].addNewMutation(m22, 0.0, q);
}
for (q in Q[21])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m23, -4.04716401435689, q);
g[!isPlus].addNewMutation(m23, 0.0, q);
}
for (q in Q[22])
{
isPlus = asLogical(rbinom(size(g), 1, 0.5));
g[isPlus].addNewMutation(m24, -4.43554686481838, q);
g[!isPlus].addNewMutation(m24, 0.0, q);
}
}
//
mutationEffect(m2) { return 1.0; }
mutationEffect(m3) { return 1.0; }
mutationEffect(m4) { return 1.0; }
mutationEffect(m5) { return 1.0; }
mutationEffect(m6) { return 1.0; }
mutationEffect(m7) { return 1.0; }
mutationEffect(m8) { return 1.0; }
mutationEffect(m9) { return 1.0; }
mutationEffect(m10) { return 1.0; }
mutationEffect(m11) { return 1.0; }
mutationEffect(m12) { return 1.0; }
mutationEffect(m13) { return 1.0; }
mutationEffect(m14) { return 1.0; }
mutationEffect(m15) { return 1.0; }
mutationEffect(m16) { return 1.0; }
mutationEffect(m17) { return 1.0; }
```

```
mutationEffect(m18) { return 1.0; }
mutationEffect(m19) { return 1.0; }
mutationEffect(m20) { return 1.0; }
mutationEffect(m21) { return 1.0; }
mutationEffect(m22) { return 1.0; }
mutationEffect(m23) { return 1.0; }
mutationEffect(m24) { return 1.0; }
 1: late() {
      // evaluate and save the additive effects of QTLs
// for (subpop in c(p1)) // only one subpop, but format kept for further
generalisation
// for (subpop in c(p1,p2))
// {
            subpop = p1 ;   // static definition of subpop, so that all the
variables are also stored statically
            inds = subpop.individuals;
            phenotype = inds.sumOfMutationsOfType(m2) +
inds.sumOfMutationsOfType(m3) +
                                    inds.sumOfMutationsOfType(m4) +
inds.sumOfMutationsOfType(m5) +
                                    inds.sumOfMutationsOfType(m6) +
inds.sumOfMutationsOfType(m7) +
                                    inds.sumOfMutationsOfType(m8) +
inds.sumOfMutationsOfType(m9) +
                                    inds.sumOfMutationsOfType(m10) +
inds.sumOfMutationsOfType(m11) +
                                    inds.sumOfMutationsOfType(m12) +
inds.sumOfMutationsOfType(m13) +
                                    inds.sumOfMutationsOfType(m14) +
inds.sumOfMutationsOfType(m15) +
                                    inds.sumOfMutationsOfType(m16) +
inds.sumOfMutationsOfType(m17) +
                                    inds.sumOfMutationsOfType(m18) +
inds.sumOfMutationsOfType(m19) +
                                    inds.sumOfMutationsOfType(m20) +
inds.sumOfMutationsOfType(m21) +
                                    inds.sumOfMutationsOfType(m22) +
inds.sumOfMutationsOfType(m23) +
                                    inds.sumOfMutationsOfType(m24);
//    optimum = (subpop == p1 ? -32.66114 else -10.0); // only one subpop,
but format kept for further

                          // generalisation. Here, optimum is equal

                          // to the mean of the prediction, -32.66114
            optimum = -32.66114 ;          // fixed optimum for the time
being

            inds.fitnessScaling = 1.0 + dnorm(optimum - phenotype, 0.0,
5.0); //

                                                                        //
            inds.tagF = phenotype;
//    }
            if (sim.cycle % 100 == 0) catn(sim.cycle + ": Mean phenotype ==
" + mean(phenotype));
 }
                          // mateChoice() {
                          //    phenotype = individual.tagF;
                          //    others = sourceSubpop.individuals.tagF;
                          //    return weights * dnorm(others, phenotype,
```

```
5.0);
                        // }
s1 1001 early() {
            // sim extended to 10000 cycles
      cat("-------------------------------\n");
      cat("Output for end of cycle " + (sim.cycle - 1) + ":\n\n");


      // Output population fitness values
      cat("p1 mean fitness = " + mean(p1.cachedFitness(NULL)) + "\n");
//    cat("p2 mean fitness = " + mean(p2.cachedFitness(NULL)) + "\n");

      // Output population additive QTL-based phenotypes
      cat("p1 mean phenotype = " + mean(p1.individuals.tagF) + "\n");
//    cat("p2 mean phenotype = " + mean(p2.individuals.tagF) + "\n");

      // Output frequencies of +1/-1 alleles at the QTLs

      cat("\nOverall frequencies:\n\n");  // modified to correctly track
allele freqs
for (q in Q[0])
{
muts = sim.mutationsOfType(m2);
plus = muts[muts.selectionCoeff == +4.15263736509172];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[1])
{
muts = sim.mutationsOfType(m3);
plus = muts[muts.selectionCoeff == -13.2893793189314];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[2])
{
muts = sim.mutationsOfType(m4);
plus = muts[muts.selectionCoeff == -13.2893793189314];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
```

```
}
for (q in Q[3])
{
muts = sim.mutationsOfType(m5);
plus = muts[muts.selectionCoeff == 4.54276804881715];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[4])
{
muts = sim.mutationsOfType(m6);
plus = muts[muts.selectionCoeff == 5.05507627554576];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[5])
{
muts = sim.mutationsOfType(m7);
plus = muts[muts.selectionCoeff == 4.96247269325974];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[6])
{
muts = sim.mutationsOfType(m8);
plus = muts[muts.selectionCoeff == 4.96247269325974];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[7])
{
muts = sim.mutationsOfType(m9);
plus = muts[muts.selectionCoeff == 4.58830862571148];
```

```
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[8])
{
muts = sim.mutationsOfType(m10);
plus = muts[muts.selectionCoeff == 5.02684381507689];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[9])
{
muts = sim.mutationsOfType(m11);
plus = muts[muts.selectionCoeff == -4.32484133512907];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[10])
{
muts = sim.mutationsOfType(m12);
plus = muts[muts.selectionCoeff == 4.32484133512912];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[11])
{
muts = sim.mutationsOfType(m13);
plus = muts[muts.selectionCoeff == 5.66703293686413];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
```

```
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[12])
{
muts = sim.mutationsOfType(m14);
plus = muts[muts.selectionCoeff == 5.66703293686413];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[13])
{
muts = sim.mutationsOfType(m15);
plus = muts[muts.selectionCoeff == 5.49605751740467];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[14])
{
muts = sim.mutationsOfType(m16);
plus = muts[muts.selectionCoeff == 5.49605751740467];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[15])
{
muts = sim.mutationsOfType(m17);
plus = muts[muts.selectionCoeff == 5.01905185415728];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
```

```
for (q in Q[16])
{
muts = sim.mutationsOfType(m18);
plus = muts[muts.selectionCoeff == 4.20007009177104];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[17])
{
muts = sim.mutationsOfType(m19);
plus = muts[muts.selectionCoeff == -4.83860362183782];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[18])
{
muts = sim.mutationsOfType(m20);
plus = muts[muts.selectionCoeff == -3.98590533697532];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[19])
{
muts = sim.mutationsOfType(m21);
plus = muts[muts.selectionCoeff == -3.95651186144391];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[20])
{
muts = sim.mutationsOfType(m22);
plus = muts[muts.selectionCoeff == -4.11515055783445];
minus = muts[muts.selectionCoeff == 0.0];
```

```
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[21])
{
muts = sim.mutationsOfType(m23);
plus = muts[muts.selectionCoeff == -4.04716401435689];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[22])
{
muts = sim.mutationsOfType(m24);
plus = muts[muts.selectionCoeff == -4.43554686481838];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
// sim.outputFull();
}
// 1001 late() {
//                               p1.outputVCFSample(25);
//                               }
//1000 late() { print(sim.mutationFrequencies(
//                                     NULL,
//                                     sim.mutationsOfType(m2),
//                                     sim.mutationsOfType(m3))); }

//
// Extension: change in optimum
// optimum becomes -11.33837 (that is, N2 mean)
//
//
 1002: late() {
      // evaluate and save the additive effects of QTLs
// for (subpop in c(p1)) // only one subpop, but format kept for further
generalisation
// for (subpop in c(p1,p2))
// {
            subpop = p1 ;   // static definition of subpop, so that all the
variables are also stored statically
            inds = subpop.individuals;
```

```
            phenotype = inds.sumOfMutationsOfType(m2) +
inds.sumOfMutationsOfType(m3) +
                                    inds.sumOfMutationsOfType(m4) +
inds.sumOfMutationsOfType(m5) +
                                    inds.sumOfMutationsOfType(m6) +
inds.sumOfMutationsOfType(m7) +
                                    inds.sumOfMutationsOfType(m8) +
inds.sumOfMutationsOfType(m9) +
                                    inds.sumOfMutationsOfType(m10) +
inds.sumOfMutationsOfType(m11) +
                                    inds.sumOfMutationsOfType(m12) +
inds.sumOfMutationsOfType(m13) +
                                    inds.sumOfMutationsOfType(m14) +
inds.sumOfMutationsOfType(m15) +
                                    inds.sumOfMutationsOfType(m16) +
inds.sumOfMutationsOfType(m17) +
                                    inds.sumOfMutationsOfType(m18) +
inds.sumOfMutationsOfType(m19) +
                                    inds.sumOfMutationsOfType(m20) +
inds.sumOfMutationsOfType(m21) +
                                    inds.sumOfMutationsOfType(m22) +
inds.sumOfMutationsOfType(m23) +
                                    inds.sumOfMutationsOfType(m24);
//    optimum = (subpop == p1 ? -11.33837 else -10.0); // only one subpop,
but format kept for further

                           // generalisation. Here, optimum is equal

                           // to the mean of the prediction, -32.66114
            optimum = -11.33837 ;          // fixed optimum for the time
being

            inds.fitnessScaling = 1.0 + dnorm(optimum - phenotype, 0.0,
5.0); //

                                                                          //
            inds.tagF = phenotype;
//    }
            if (sim.cycle % 100 == 0) catn(sim.cycle + ": Mean phenotype ==
" + mean(phenotype));
 }
                          // mateChoice() {
                          //    phenotype = individual.tagF;
                          //    others = sourceSubpop.individuals.tagF;
                          //    return weights * dnorm(others, phenotype,
5.0);
                          // }
2001 early() {
            // sim extended to 10000 cycles
      cat("-----------------------------\n");
      cat("After change in selection regime:\n");
      cat("Output for end of cycle " + (sim.cycle - 1) + ":\n\n");


      // Output population fitness values
      cat("p1 mean fitness = " + mean(p1.cachedFitness(NULL)) + "\n");
//    cat("p2 mean fitness = " + mean(p2.cachedFitness(NULL)) + "\n");

      // Output population additive QTL-based phenotypes
      cat("p1 mean phenotype = " + mean(p1.individuals.tagF) + "\n");
//    cat("p2 mean phenotype = " + mean(p2.individuals.tagF) + "\n");
```

```
      // Output frequencies of +1/-1 alleles at the QTLs

      cat("\nOverall frequencies:\n\n");  // modified to correctly track
allele freqs
for (q in Q[0])
{
muts = sim.mutationsOfType(m2);
plus = muts[muts.selectionCoeff == +4.15263736509172];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[1])
{
muts = sim.mutationsOfType(m3);
plus = muts[muts.selectionCoeff == -13.2893793189314];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[2])
{
muts = sim.mutationsOfType(m4);
plus = muts[muts.selectionCoeff == -13.2893793189314];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[3])
{
muts = sim.mutationsOfType(m5);
plus = muts[muts.selectionCoeff == 4.54276804881715];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
```

```
for (q in Q[4])
{
muts = sim.mutationsOfType(m6);
plus = muts[muts.selectionCoeff == 5.05507627554576];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[5])
{
muts = sim.mutationsOfType(m7);
plus = muts[muts.selectionCoeff == 4.96247269325974];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[6])
{
muts = sim.mutationsOfType(m8);
plus = muts[muts.selectionCoeff == 4.96247269325974];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[7])
{
muts = sim.mutationsOfType(m9);
plus = muts[muts.selectionCoeff == 4.58830862571148];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[8])
{
muts = sim.mutationsOfType(m10);
plus = muts[muts.selectionCoeff == 5.02684381507689];
minus = muts[muts.selectionCoeff == 0.0];
```

```
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[9])
{
muts = sim.mutationsOfType(m11);
plus = muts[muts.selectionCoeff == -4.32484133512907];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[10])
{
muts = sim.mutationsOfType(m12);
plus = muts[muts.selectionCoeff == 4.32484133512912];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[11])
{
muts = sim.mutationsOfType(m13);
plus = muts[muts.selectionCoeff == 5.66703293686413];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[12])
{
muts = sim.mutationsOfType(m14);
plus = muts[muts.selectionCoeff == 5.66703293686413];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
```

```
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[13])
{
muts = sim.mutationsOfType(m15);
plus = muts[muts.selectionCoeff == 5.49605751740467];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[14])
{
muts = sim.mutationsOfType(m16);
plus = muts[muts.selectionCoeff == 5.49605751740467];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[15])
{
muts = sim.mutationsOfType(m17);
plus = muts[muts.selectionCoeff == 5.01905185415728];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[16])
{
muts = sim.mutationsOfType(m18);
plus = muts[muts.selectionCoeff == 4.20007009177104];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[17])
```

```
{
muts = sim.mutationsOfType(m19);
plus = muts[muts.selectionCoeff == -4.83860362183782];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[18])
{
muts = sim.mutationsOfType(m20);
plus = muts[muts.selectionCoeff == -3.98590533697532];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[19])
{
muts = sim.mutationsOfType(m21);
plus = muts[muts.selectionCoeff == -3.95651186144391];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[20])
{
muts = sim.mutationsOfType(m22);
plus = muts[muts.selectionCoeff == -4.11515055783445];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[21])
{
muts = sim.mutationsOfType(m23);
plus = muts[muts.selectionCoeff == -4.04716401435689];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
```

```
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
for (q in Q[22])
{
muts = sim.mutationsOfType(m24);
plus = muts[muts.selectionCoeff == -4.43554686481838];
minus = muts[muts.selectionCoeff == 0.0];
qPlus = plus[plus.position == q];
qMinus = minus[minus.position == q];
pf = sum(sim.mutationFrequencies(NULL, qPlus));
mf = sum(sim.mutationFrequencies(NULL, qMinus));
pf1 = sum(sim.mutationFrequencies(p1, qPlus));
mf1 = sum(sim.mutationFrequencies(p1, qMinus));
// cat("   QTL " + q + ": f(+) == " + pf + ", f(-) == " + mf + "\n");
cat("         in p1: f(+) == " + pf1 + ", f(-) == " + mf1 + "\n");
}
// sim.outputFull();
}
```

Supplementary Table ST1

P-values for all Wilcoxon tests for pairwise comparisons of character value distributions.

Trait	BBCH 7	BBCH 7	BBCH 9	BBCH 9	PSS	PSS
Year	2009	2010	2009	2010	2009	2010
SNPs FDR 5%	*23*	*37*	*43*	*101*	*4*	*43*
SNPs FDR 10%	102	213	189	443	62	222
P-val support Countergradient (N1 > N2) FDR 5%	*0,9*	*0,91*	*0,97*	*0,62*	*0,2*	0,04
P-val support Countergradient (N1 > N2) FDR 10%	0,89	0,72	0,99	0,87	0,17	0,37
P-val support Gradient (N1 < N2) FDR 5%	*0,034*	*0,091*	*0,02*	*0,39*	*0,8*	*0,96*
P-val support Gradient (N1 < N2) FDR 10%	0,11	0,29	0,004	0,14	0,83	0,64
P-val support Countergradient (N1 > N3) FDR 5%	*0,2*	*0,51*	*0,37*	*0,66*	*0,943*	*0,48*
P-val support Countergradient (N1 > N3) FDR 10%	0,52	0,32	0,98	0,88	0,51	0,71
P-val support Gradient (N1 < N3) FDR 5%	*0,8*	*0,5*	*0,64*	*0,94*	*0,0598*	*0,52*
P-val support Gradient (N1 < N3) FDR 10%	0,48	0,68	0,015	0,12	0,49	0,3
P-val support Countergradient (N1 > N4) FDR 5%	0,014	*0,5*	*0,08*	*0,6*	*0,9*	*0,45*
P-val support Countergradient (N1 > N4) FDR 10%	0,27	0,69	0,91	0,52	0,96	0,3
P-val support Gradient (N1 < N4) FDR 5%	*0,99*	*0,51*	*0,92*	*0,94*	*0,1*	*0,55*
P-val support Gradient (N1 < N4) FDR 10%	0,74	0,32	0,09	0,48	0,044	0,7
P-val support Countergradient (N2 > N3) FDR 5%	0,015	*0,12*	0,012	*0,66*	*0,96*	*0,93*
P-val support Countergradient (N2 > N3) FDR 10%	0,23	0,88	0,3	0,43	0,83	0,83
P-val support Gradient (N2 < N3) FDR 5%	*0,99*	*0,14*	*0,99*	*0,94*	*0,041*	*0,068*
P-val support Gradient (N2 < N3) FDR 10%	0,77	0,88	0,69	0,58	0,18	0,177
P-val support Countergradient (N2 > N4) FDR 5%	4.47*10^5	*0,86*	0,002	*0,08*	*0,96*	*0,89*
P-val support Countergradient (N2 > N4) FDR 10%	0,05	0,51	0,28	0,21	0,998	0,37
P-val support Gradient (N2 < N4) FDR 5%	*1*	*0,51*	*0,998*	*0,92*	*0,045*	*0,12*
P-val support Gradient (N2 < N4) FDR 10%	0,95	0,5	0,72	0,79	0,0026	0,64
P-val support Countergradient (N3 > N4) FDR 5%	*0,26*	*0,49*	*0,11*	*0,54*	*0,45*	*0,42*
P-val support Countergradient (N3 > N4) FDR 10%	0,31	0,84	0,35	0,24	0,97	0,12
P-val support Gradient (N3 < N4) FDR 5%	*0,98*		*0,9*	*0,47*	*0,56*	*0,59*
P-val support Gradient (N3 < N4) FDR 10%	0,69	0,17	0,65	0,76	0,0297	0,88

Trait	a (slope logistic)	a (slope logistic)	b (inflection logistic)	b (inflection logistic)	Degree Days 7	Degree Days 7	Degree Days 9	Degree Days 9
Year	2009	2010	2009	2010	2009	2010	2009	2010
SNPs FDR 5%	*15*	*195*	*1*	*41*	*17*	*120*	*32*	*71*
SNPs FDR 10%	81	752	28	222	58	449	98	322
P-val support Countergradient (N1 > N2) FDR 5%	*0,8*	*0,09*	*0,5*	*0,89*	0,98	0,55	0,97	0,77
P-val support Countergradient (N1 > N2) FDR 10%	0,8	0,12	0,94	0,53	0,99	0,38	0,98	0,88
P-val support Gradient (N1 < N2) FDR 5%	*0,2*	*0,91*	*0,6*	*0,11*	*0,0247*	*0,46*	*0,031*	*0,24*
P-val support Gradient (N1 < N2) FDR 10%	0,2	0,88	0,06	0,48	0,01	0,62	0,015	0,12
P-val support Countergradient (N1 > N3) FDR 5%	*0,28*	*0,91*	*0,84*	*0,31*	*0,45*	*0,31*	*0,49*	*0,073*
P-val support Countergradient (N1 > N3) FDR 10%	0,6	0,37	0,45	0,21	0,94	0,23	0,86	0,45
P-val support Gradient (N1 < N3) FDR 5%	*0,72*	*0,09*	*0,17*	*0,7*	*0,56*	*0,7*	*0,52*	*0,93*
P-val support Gradient (N1 < N3) FDR 10%	0,4	0,64	0,55	0,8	0,07	0,77	0,14	0,55
P-val support Countergradient (N1 > N4) FDR 5%	*0,31*	*0,89*	*0,93*	*0,55*	*0,2*	*0,18*	0,0168	0,03547
P-val support Countergradient (N1 > N4) FDR 10%	0,44	0,63	0,07	0,59	0,86	0,1	0,65	0,26
P-val support Gradient (N1 < N4) FDR 5%	*0,69*	*0,11*	*0,077*	*0,46*	*0,8*	*0,84*	*0,984*	*0,97*
P-val support Gradient (N1 < N4) FDR 10%	0,57	0,38	0,93	0,42	0,14	0,9	0,36	0,74
P-val support Countergradient (N2 > N3) FDR 5%	*0,089*	*0,992*	*0,84*	*0,063*	0,008	*0,33*	0,027	*0,07*
P-val support Countergradient (N2 > N3) FDR 10%	0,26	0,77	0,025	0,14	0,046	0,38	0,088	0,085
P-val support Gradient (N2 < N3) FDR 5%	*0,91*	*0,00897*	*0,17*	*0,94*	*0,992*	*0,68*	*0,9741*	*0,93*
P-val support Gradient (N2 < N3) FDR 10%	0,74	0,23	0,98	0,86	0,96	0,63	0,92	0,92
P-val support Countergradient (N2 > N4) FDR 5%	*0,11*	*0,995*	*0,93*	*0,23*	0,0006	0,2	0,00124	0,03867
P-val support Countergradient (N2 > N4) FDR 10%	0,22	0,9	0,000557	0,58	0,0356	0,28	0,0215	0,058
P-val support Gradient (N2 < N4) FDR 5%	*0,89*	*0,00528*	*0,077*	*0,78*	*0,997*	*0,8*	*0,999*	*0,96*
P-val support Gradient (N2 < N4) FDR 10%	0,79	0,1	0,9995	0,42	0,97	0,72	0,98	0,94
P-val support Countergradient (N3 > N4) FDR 5%	*0,55*	*0,4*	*0,67*	*0,61*	*0,09*	*0,38*	0,0137	*0,34*
P-val support Countergradient (N3 > N4) FDR 10%	0,42	0,65	0,063	0,86	0,35	0,3	0,34	0,27
P-val support Gradient (N3 < N4) FDR 5%	*0,46*	*0,61*	*0,34*	*0,39*	*0,9*	*0,63*	*0,99*	*0,67*
P-val support Gradient (N3 < N4) FDR 10%	0,59	0,35	0,94	0,14	0,66	0,7	0,67	0,73

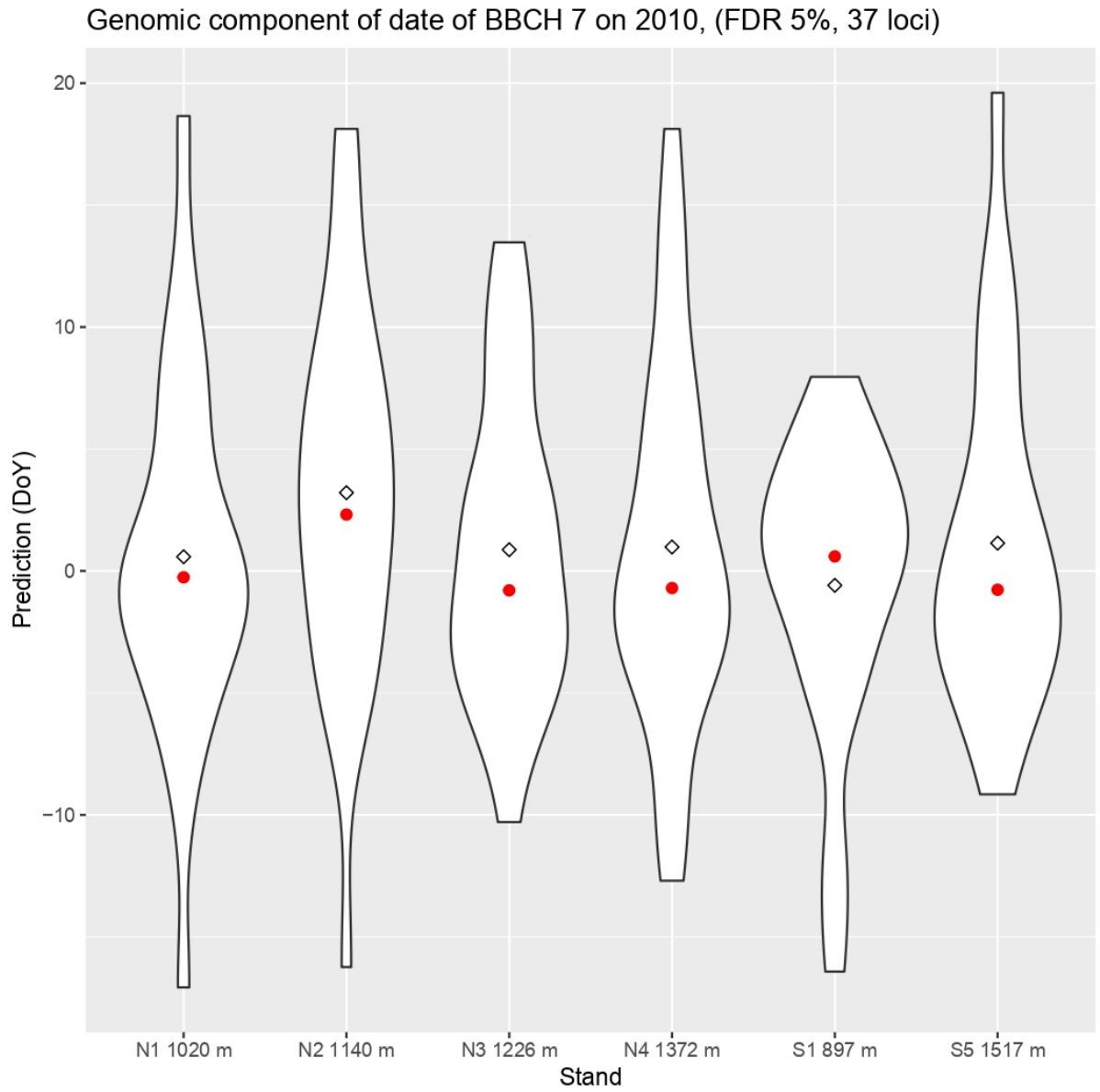
Supplementary Figure SF1

Distribution of breeding values by population for all characters, for years 2009 and 20120, and for FDR de 5% ond 10%. The absolute breeding values are arbitrary, as the allelic effects may have biases and inflation, due to the small sample size use in GWAS.

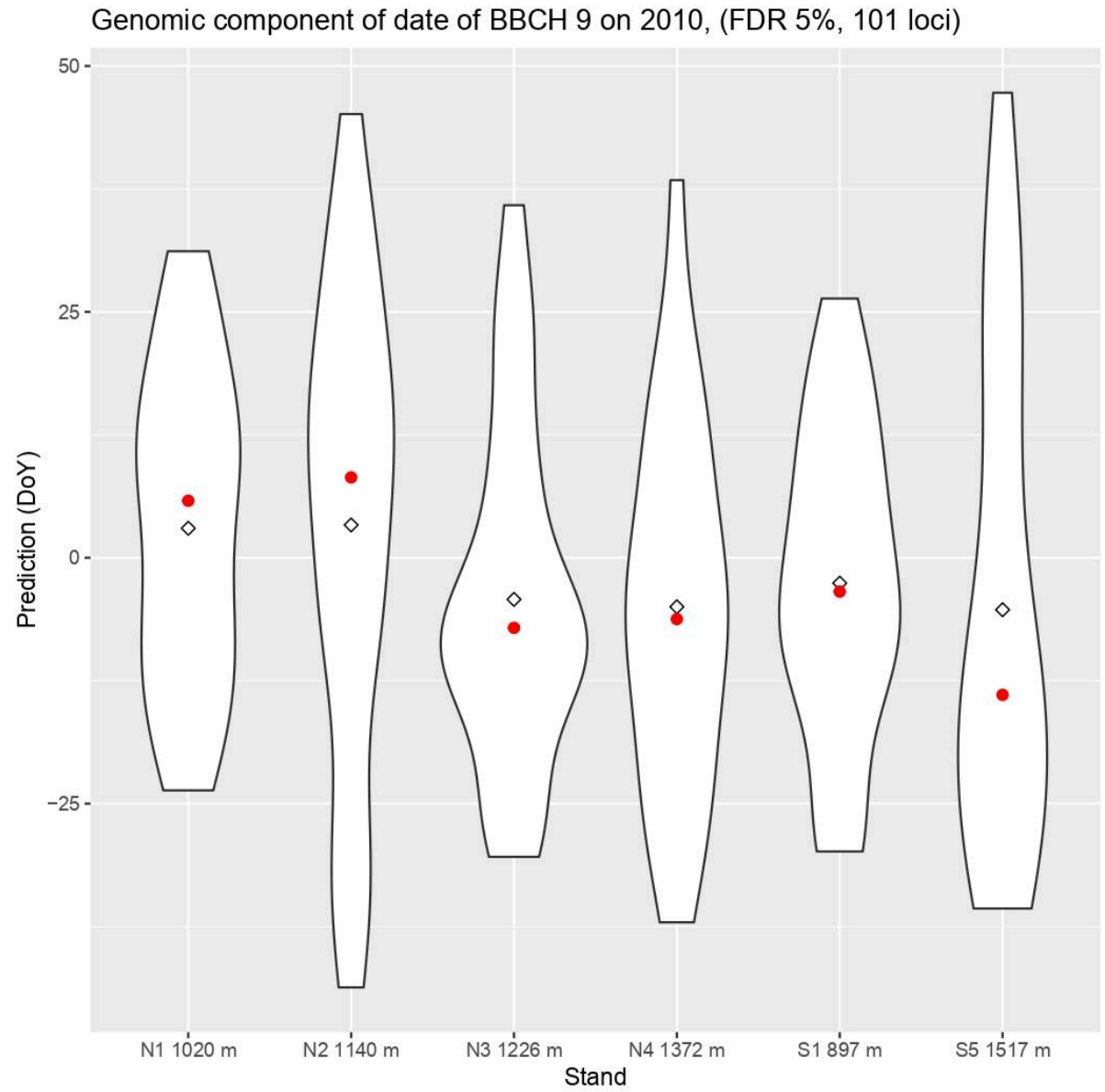
Traits are ordered as:

FDR 5%
DOY BBCH 07 2010
DOY BBCH 09 2009
DOY BBCH 09 2010
Logistic slope *a* 2009
Logistic slope *a* 2010
Logistic mid-point *b* 2009
Logistic mid-point *b* 2010
PSS 2009
PSS 2010
Degree Days at BBCH 07 2009
Degree Days at BBCH 07 2010
Degree Days at BBCH 09 2009
Degree Days at BBCH 09 2010

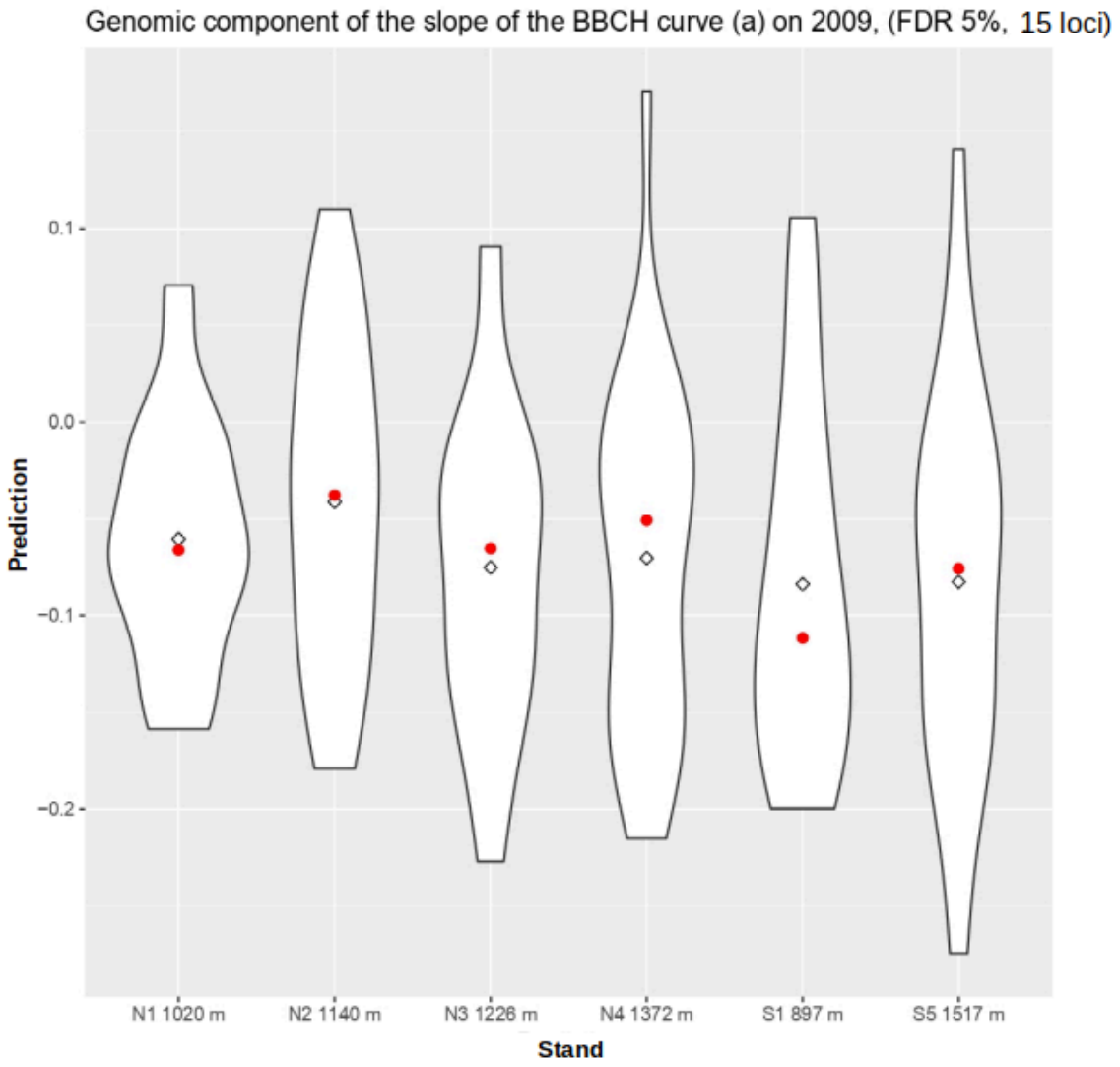
FDR 10%
DOY BBCH 07 2009
DOY BBCH 07 2010
DOY BBCH 09 2009
DOY BBCH 09 2010
Logistic slope *a* 2009
Logistic slope *a* 2010
Logistic mid-point *b* 2009
Logistic mid-point *b* 2010
PSS 2009
PSS 2010
Degree Days at BBCH 07 2009
Degree Days at BBCH 07 2010
Degree Days at BBCH 09 2009
Degree Days at BBCH 09 2010

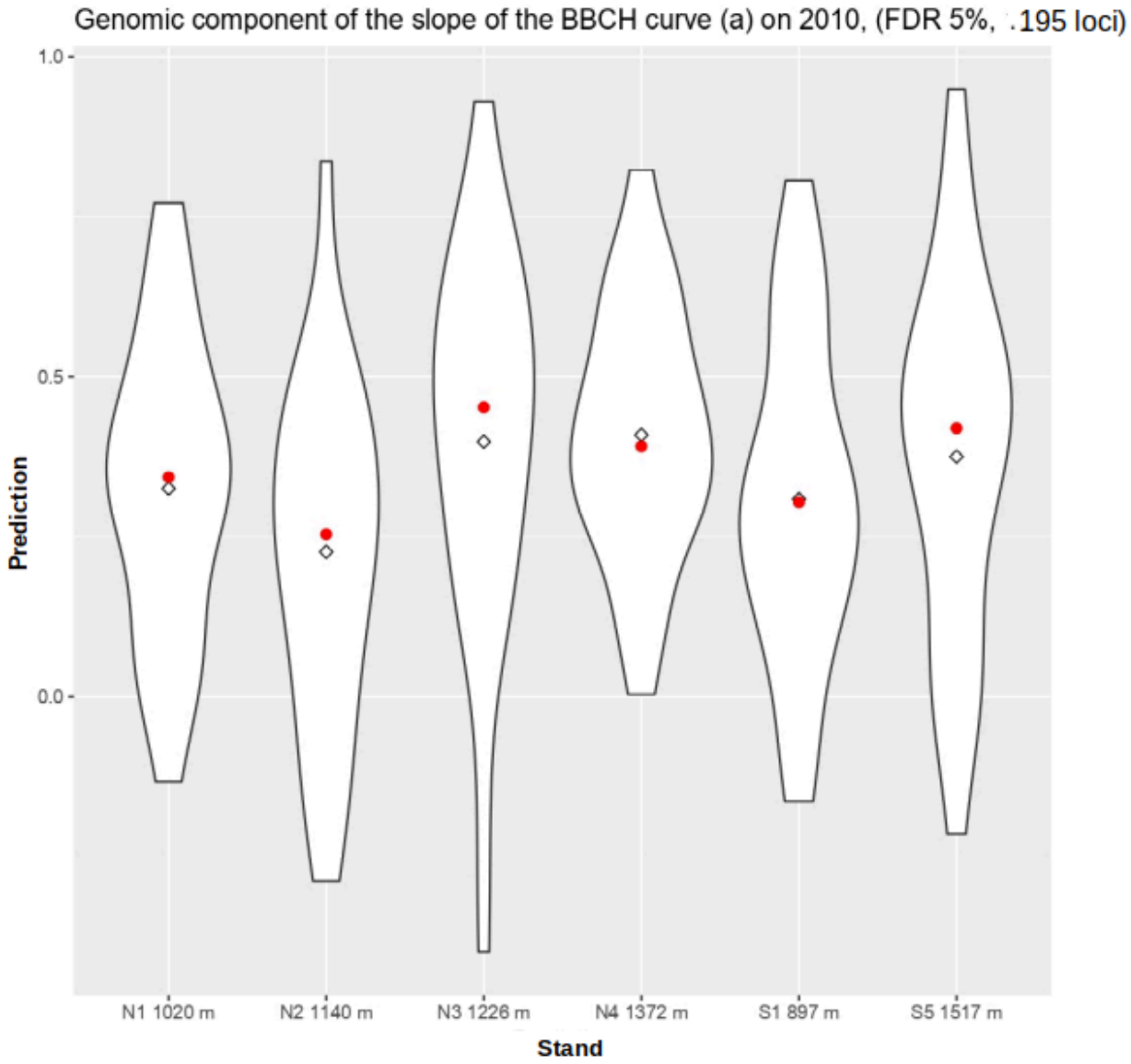

Genomic component of date of BBCH 7 on 2010, (FDR 5%, 37 loci)
Prediction (DoY)
20
10
0
−10
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

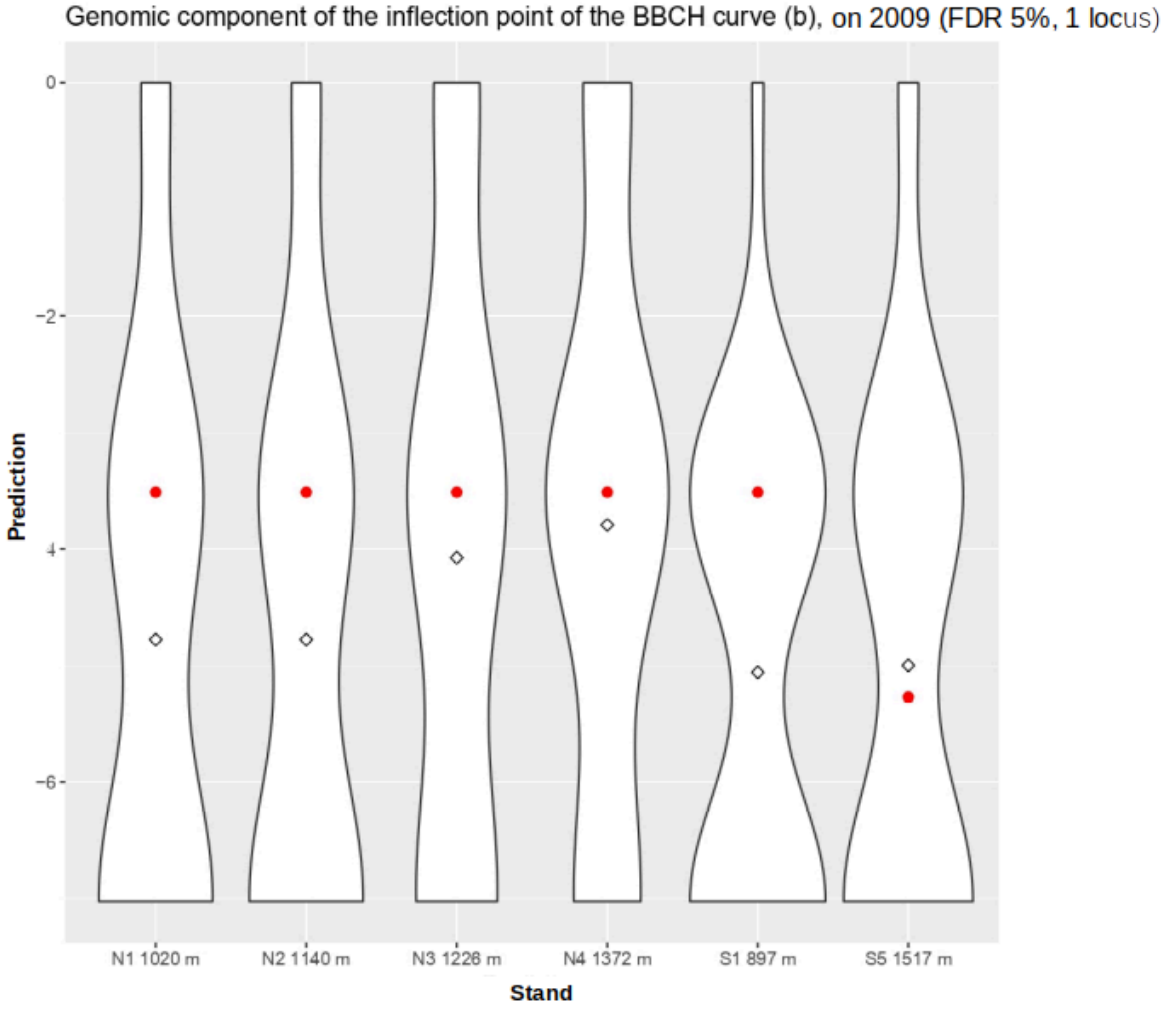
OBJ

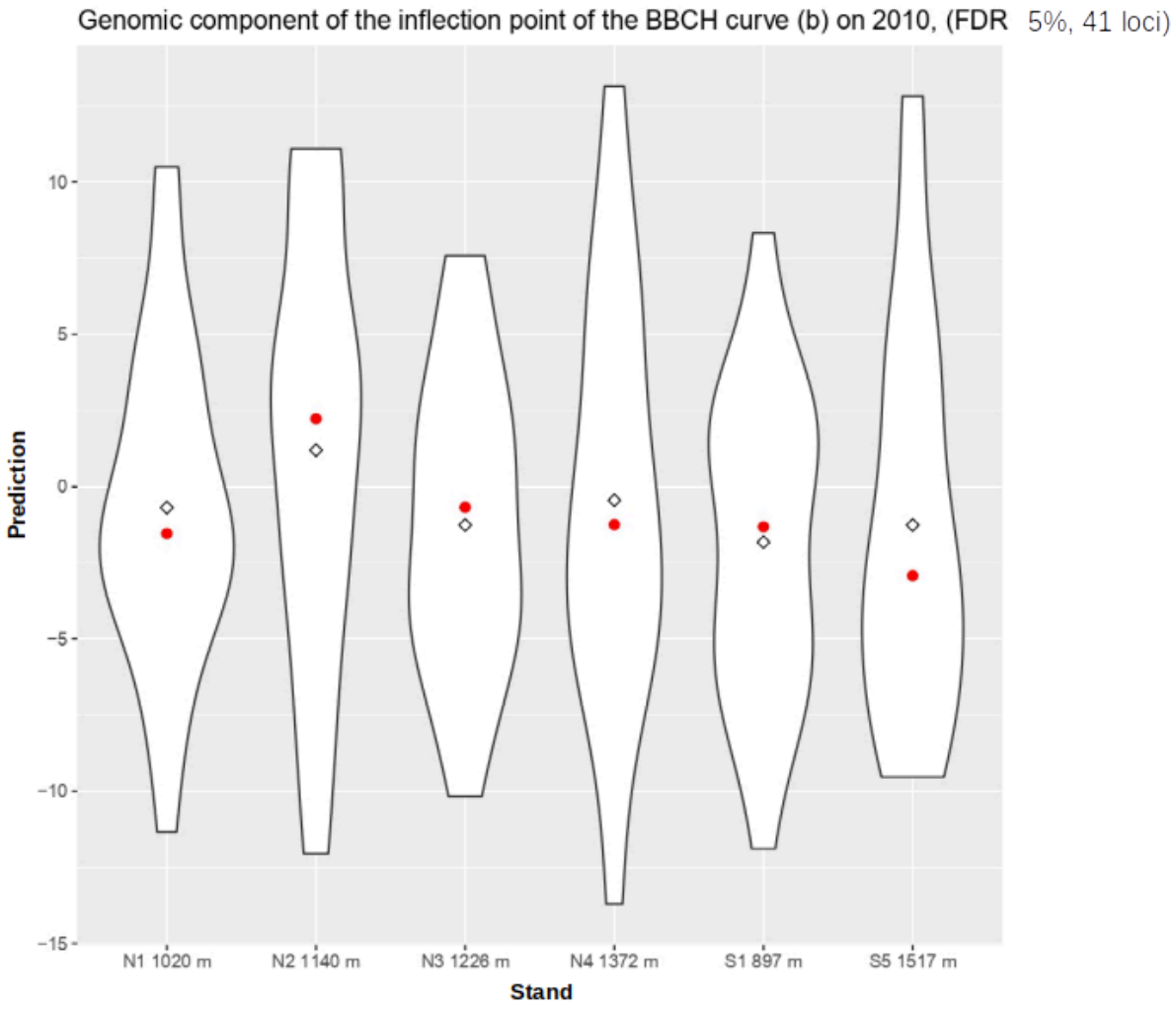

Genomic component of date of BBCH 9 on 2010, (FDR 5%, 101 loci)
Prediction (DoY)
50
25
0
−25
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

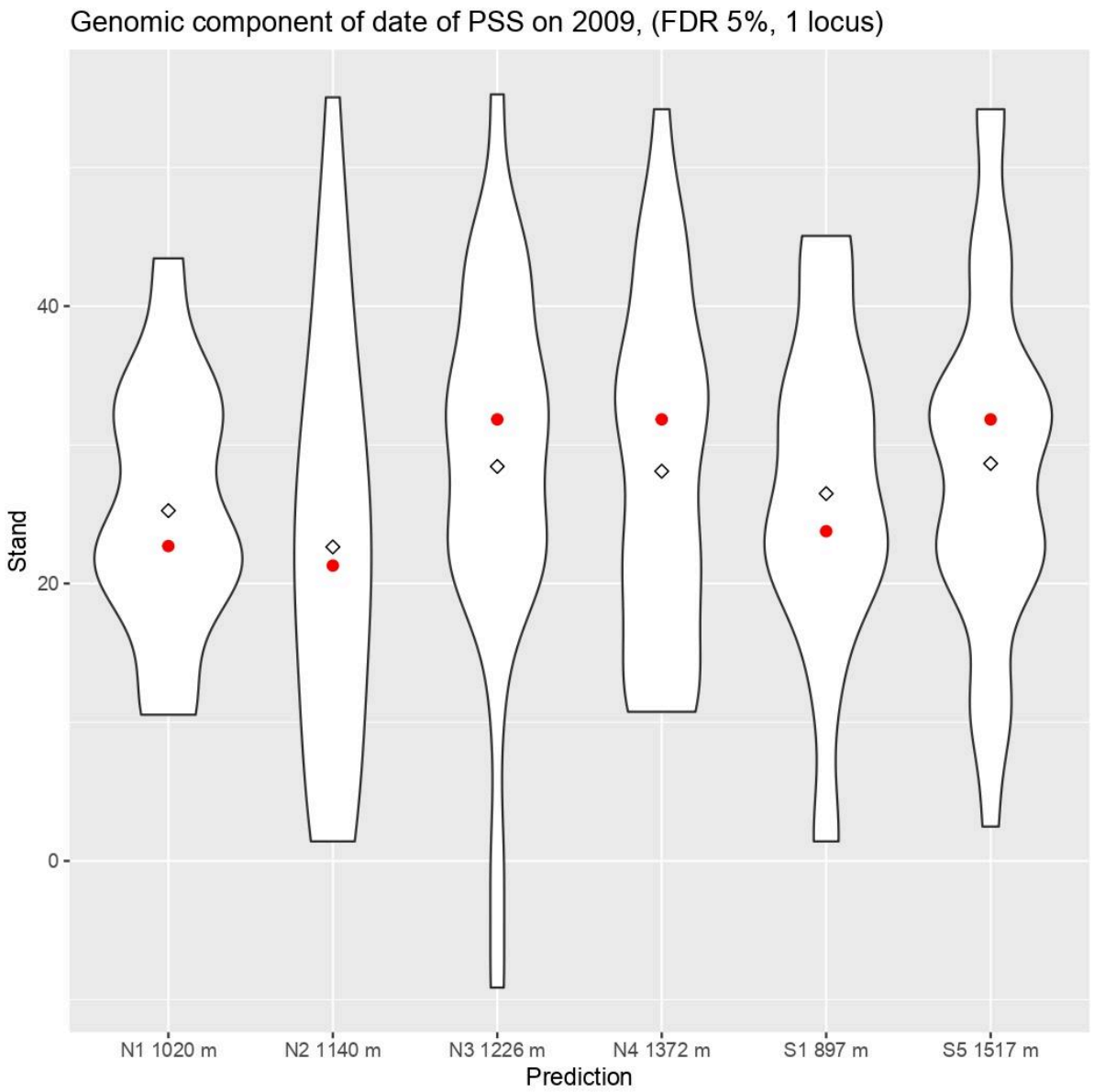
OBJ

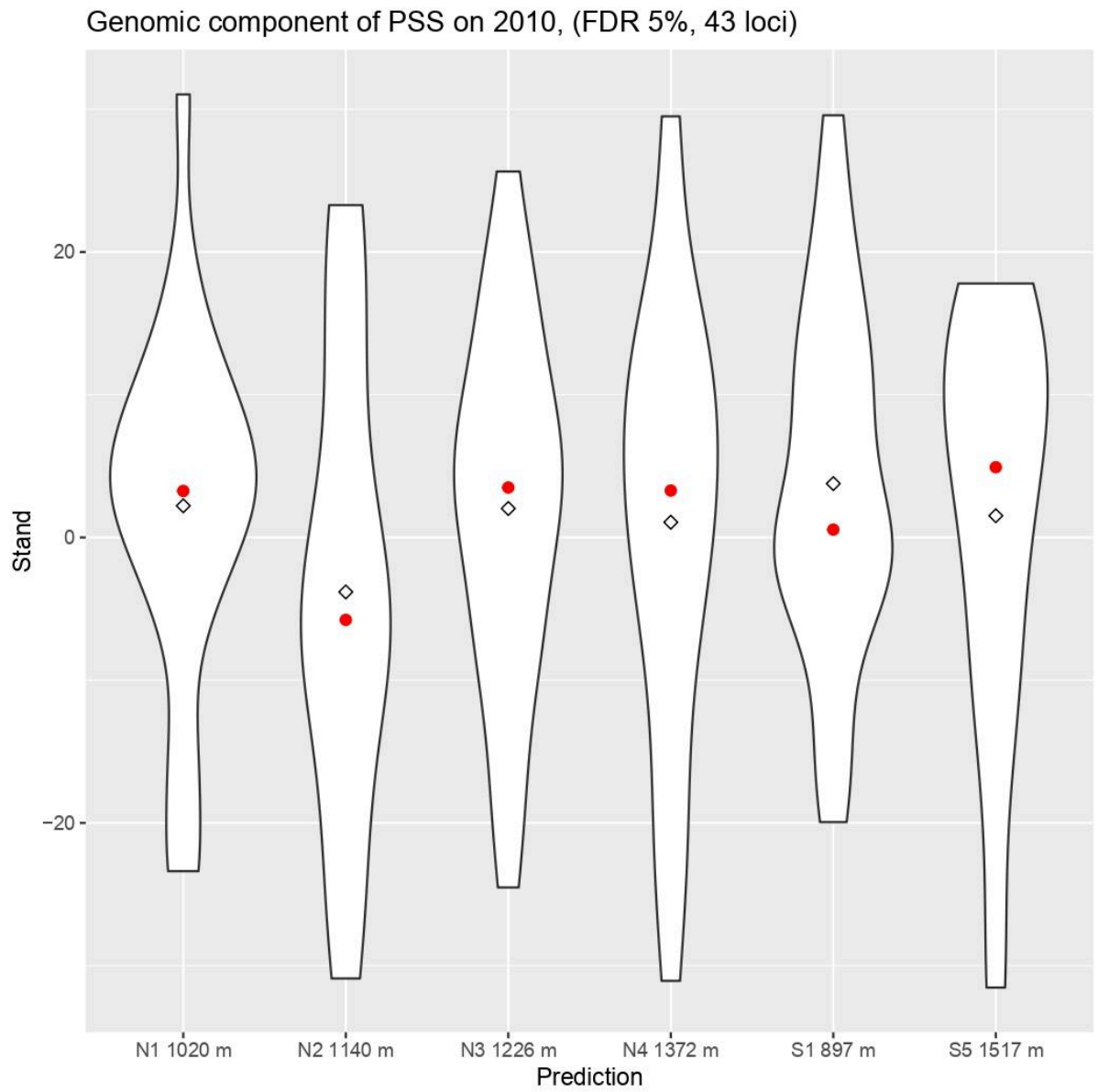

Genomic component of the slope of the BBCH curve (a) on 2009, (FDR 5%, 15 loci)
Prediction
0.1
0.0
−0.1
−0.2
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand


Genomic component of the slope of the BBCH curve (a) on 2010, (FDR 5%, .195 loci)
1.0
0.5
0.0
Prediction
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

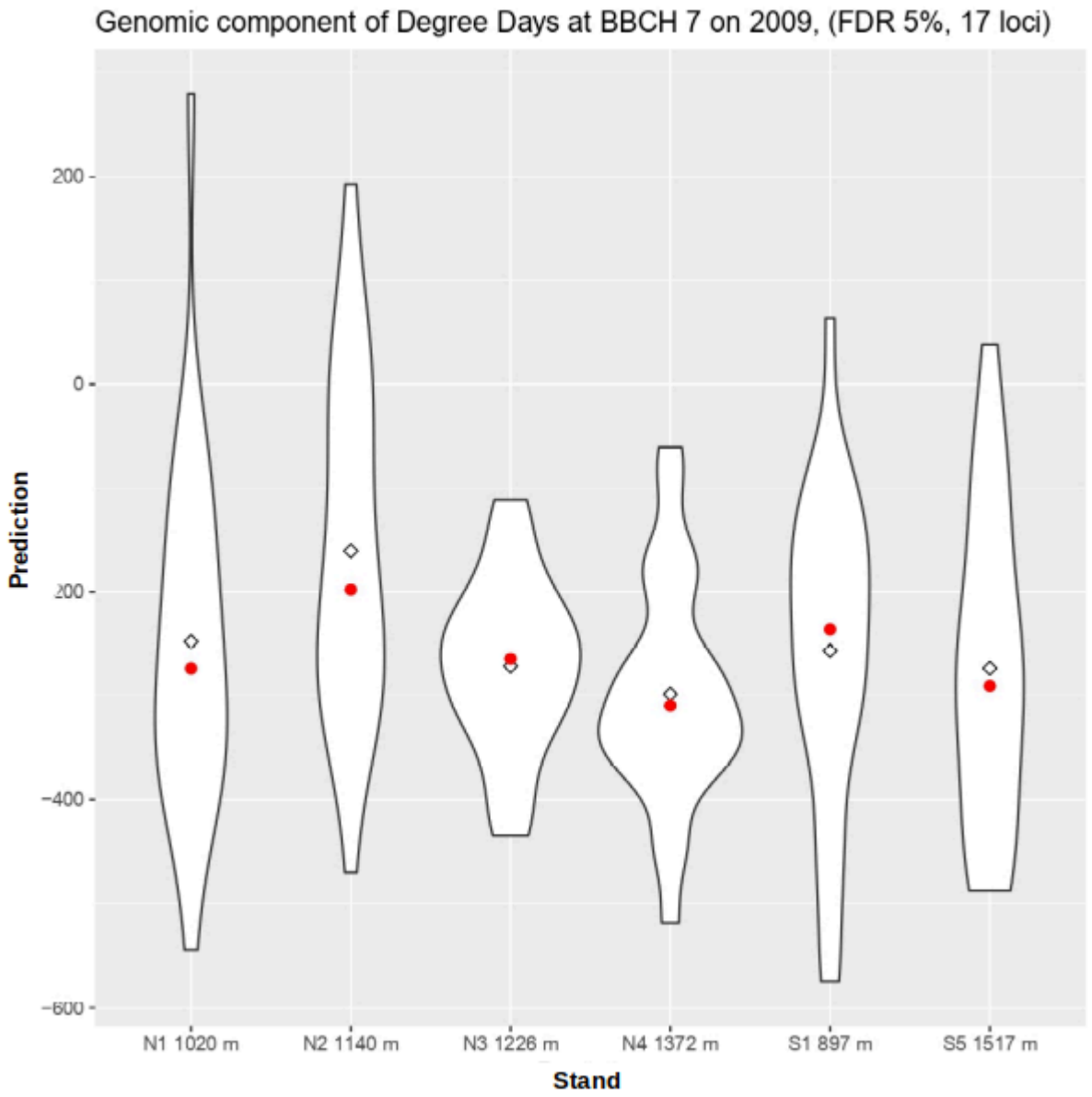

Genomic component of the inflection point of the BBCH curve (b), on 2009 (FDR 5%, 1 locus)
Prediction
0
−2
−4
−6
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

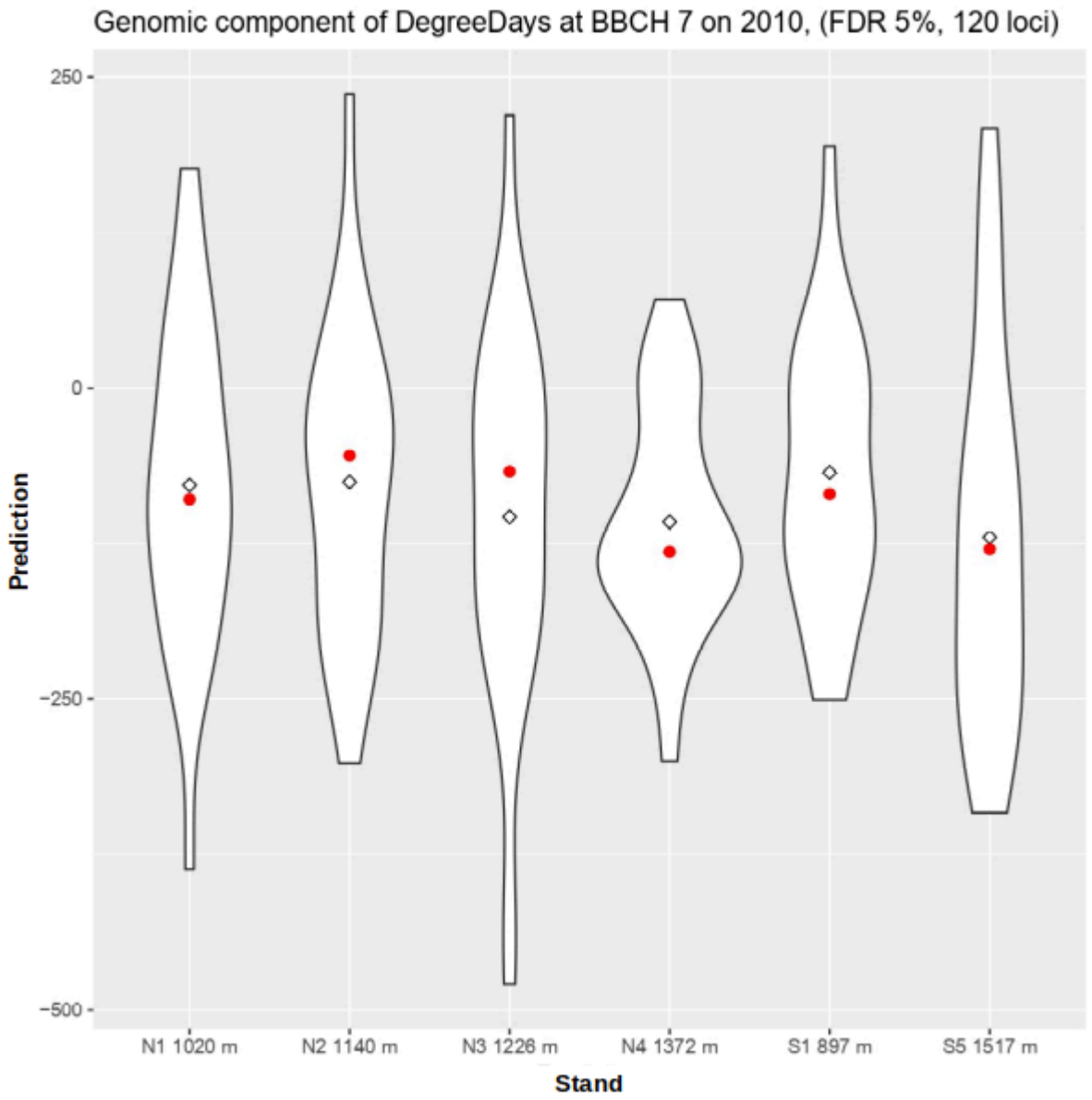

Genomic component of the inflection point of the BBCH curve (b) on 2010, (FDR 5%, 41 loci)
Prediction
10
5
0
−5
−10
−15
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

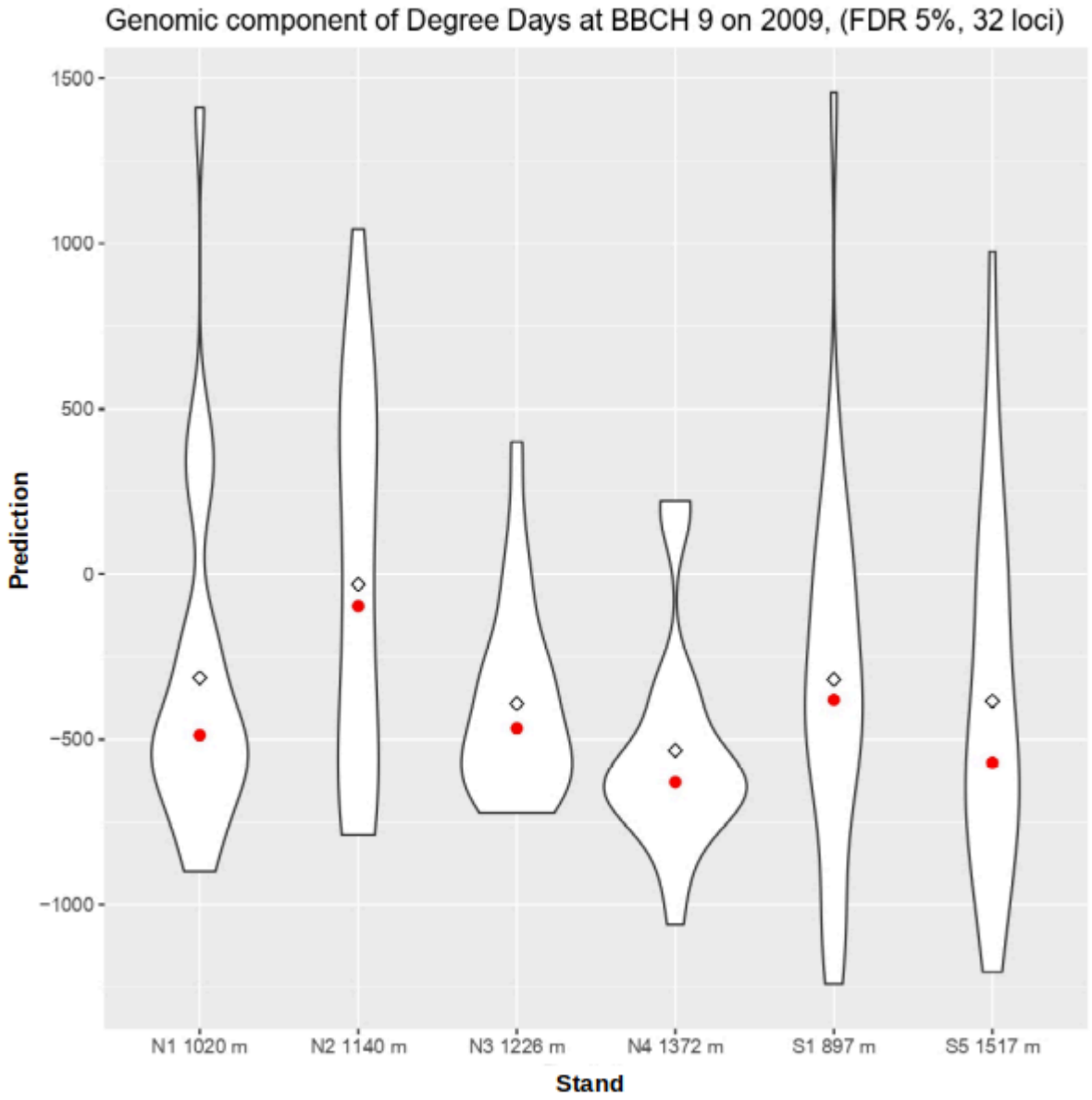

Genomic component of date of PSS on 2009, (FDR 5%, 1 locus)
Stand
40
20
0
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Prediction

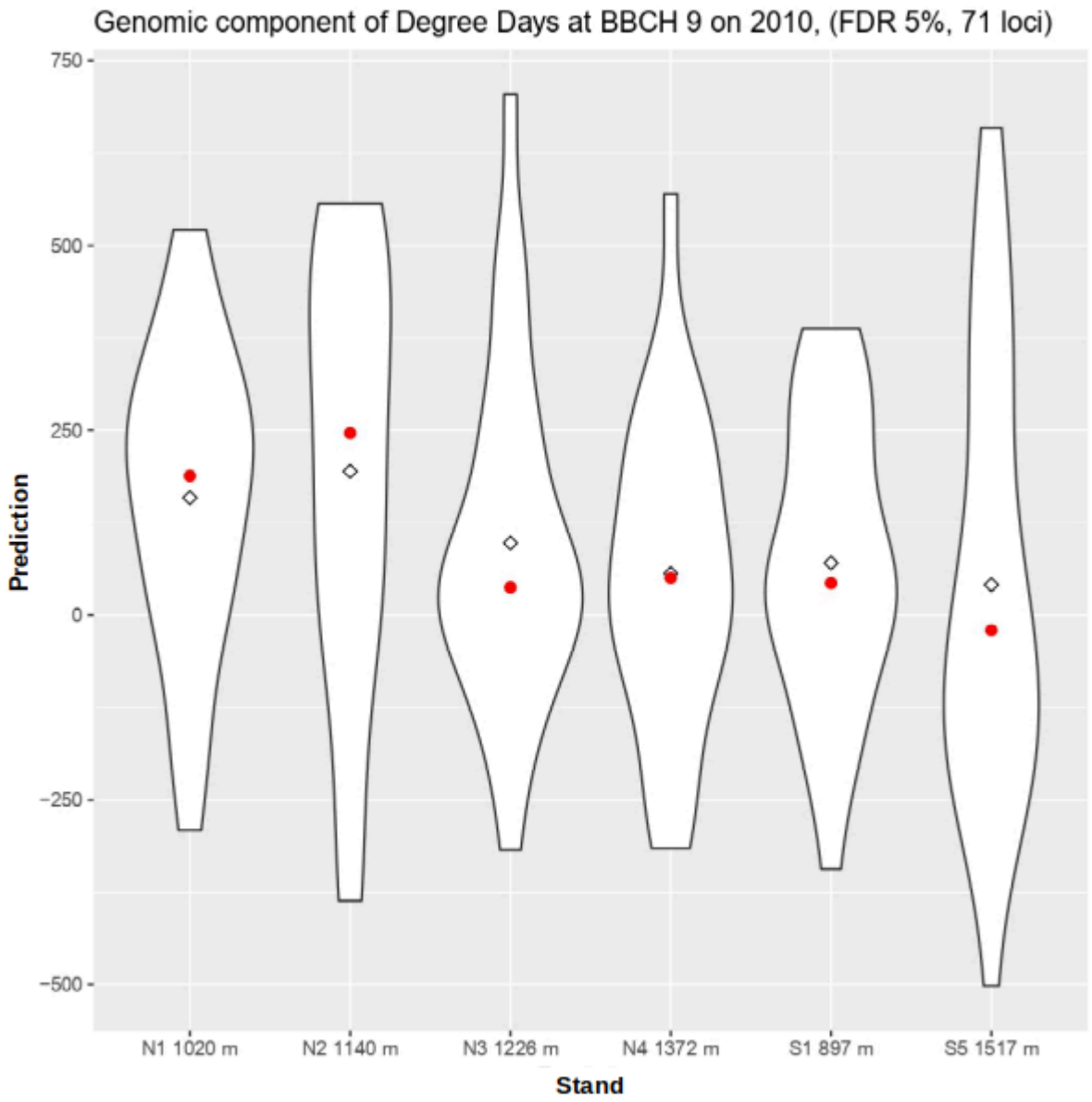

Genomic component of PSS on 2010, (FDR 5%, 43 loci)
Stand
20
0
−20
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Prediction

OBJ OBJ OBJ


Genomic component of Degree Days at BBCH 7 on 2009, (FDR 5%, 17 loci)
Prediction
200
0
−200
−400
−600
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

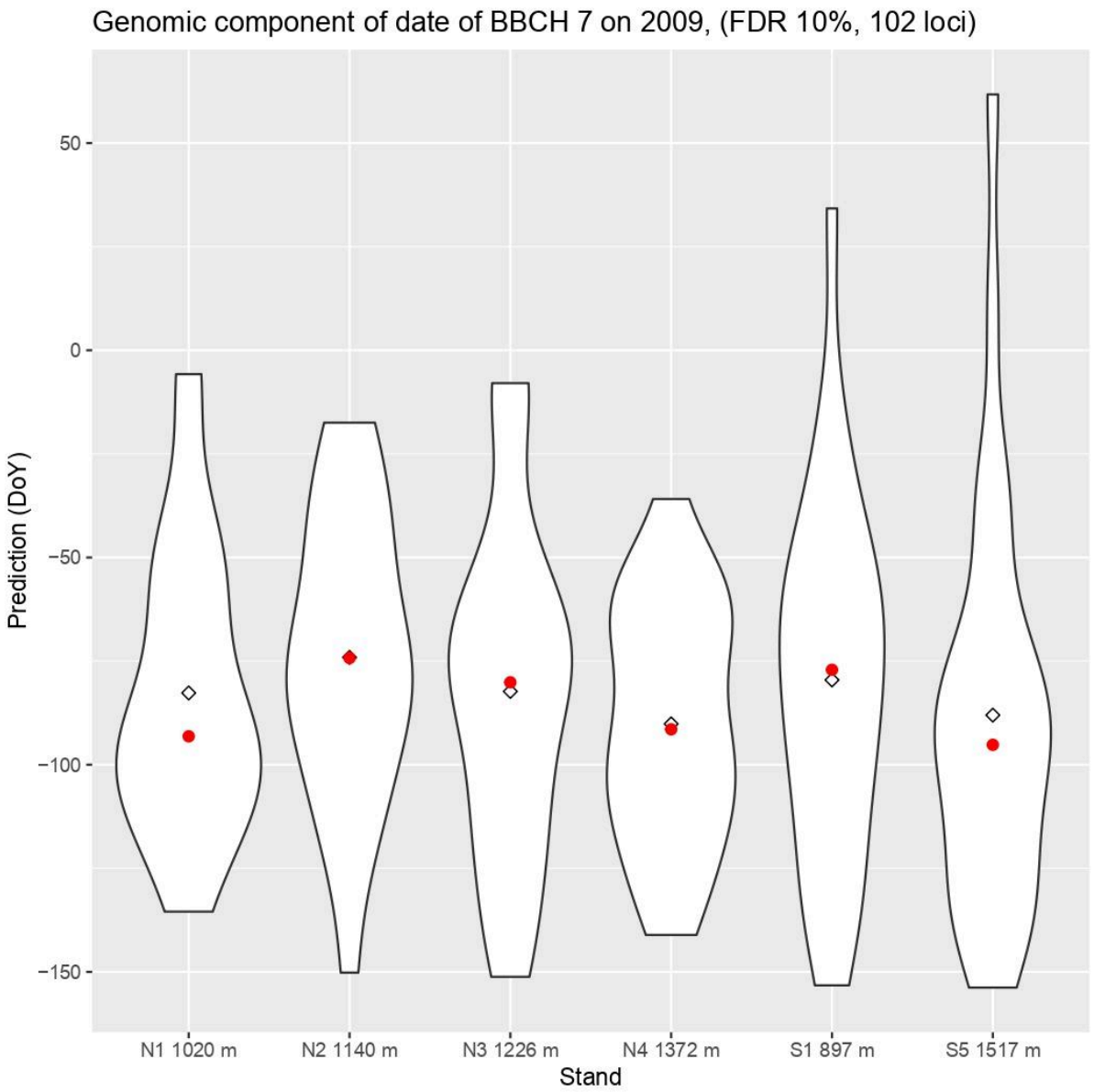

Genomic component of DegreeDays at BBCH 7 on 2010, (FDR 5%, 120 loci)
Prediction
250
0
−250
−500
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

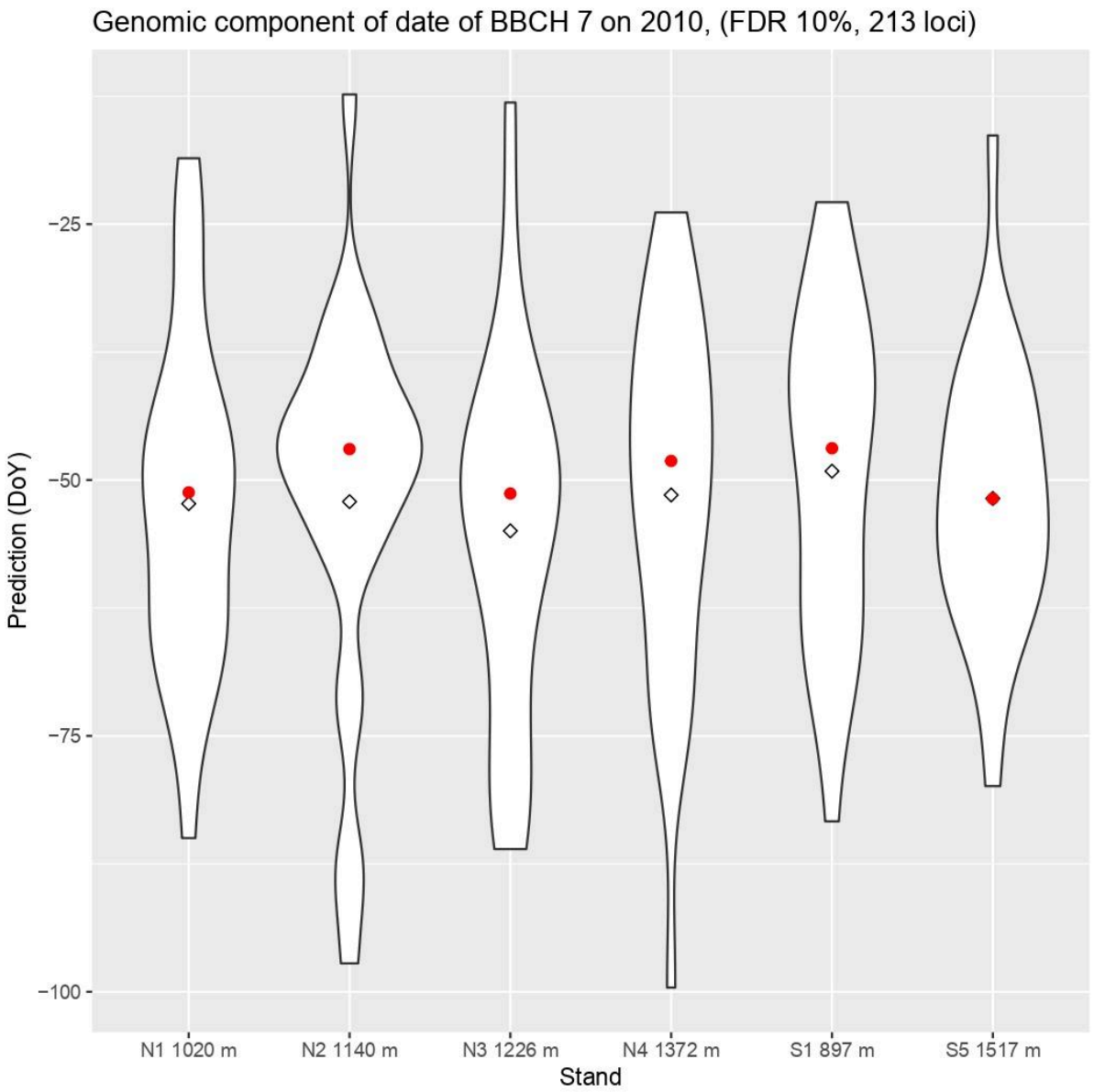

Genomic component of Degree Days at BBCH 9 on 2009, (FDR 5%, 32 loci)
Prediction
1500
1000
500
0
−500
−1000
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

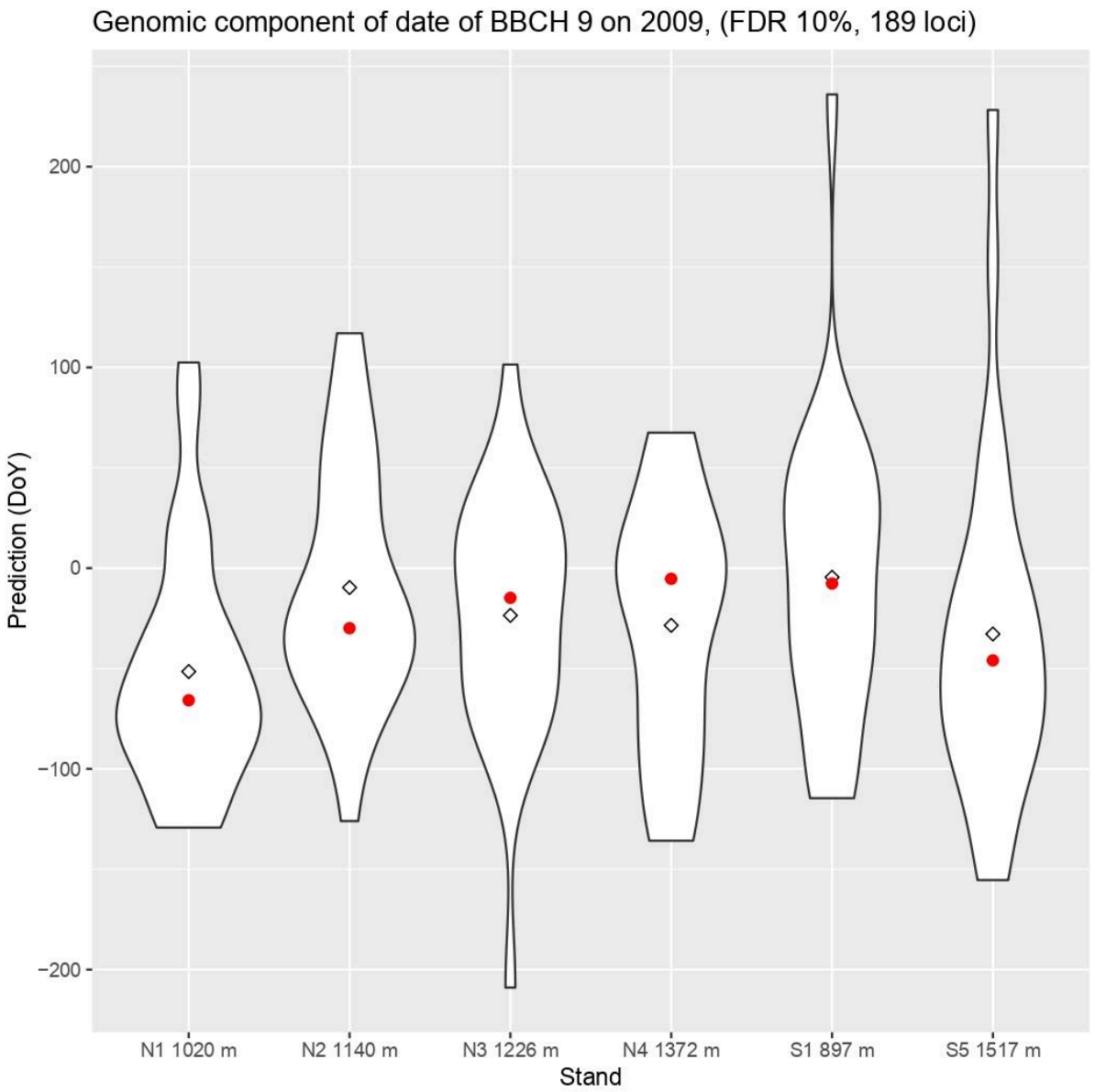


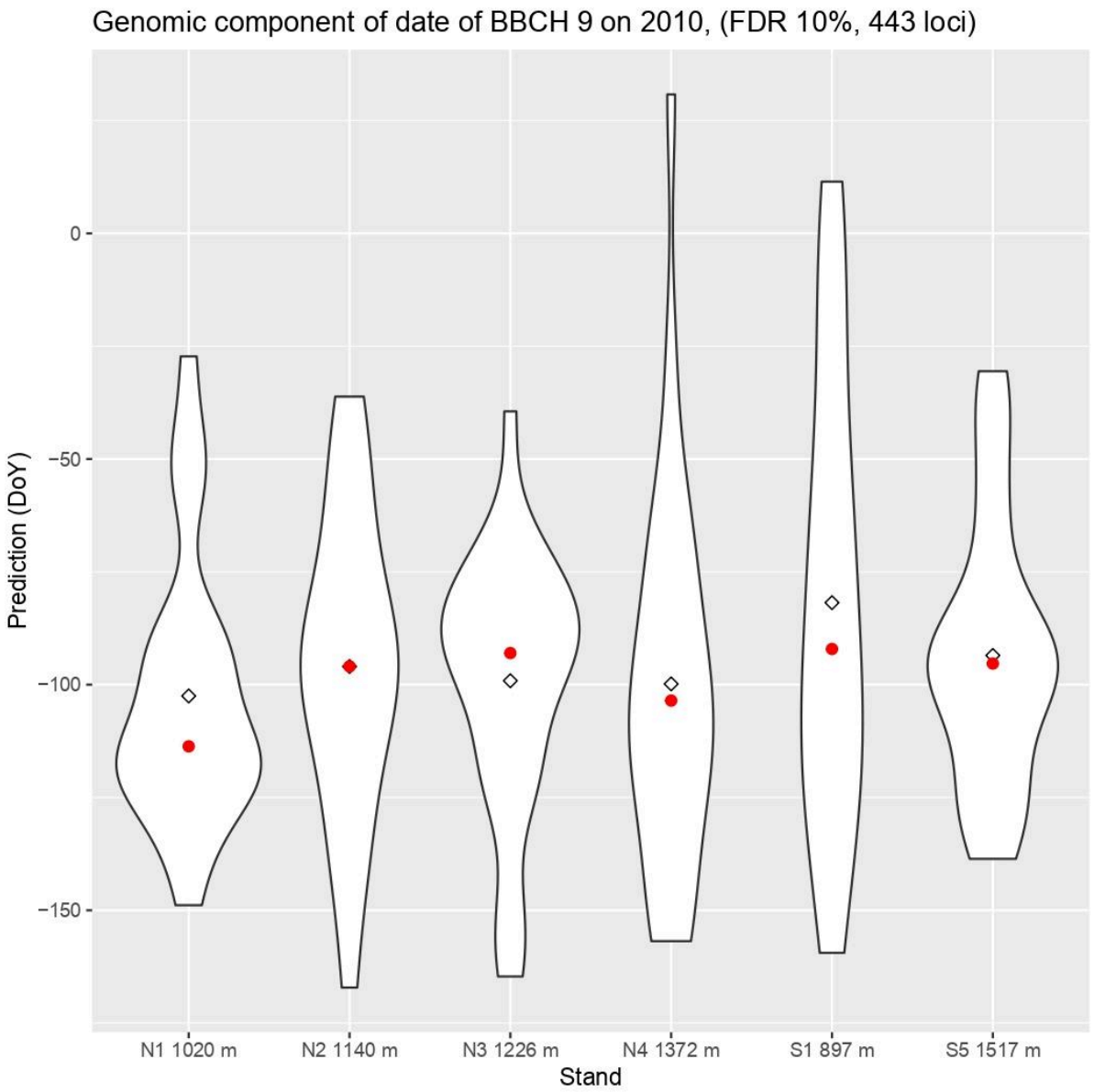
OBJ

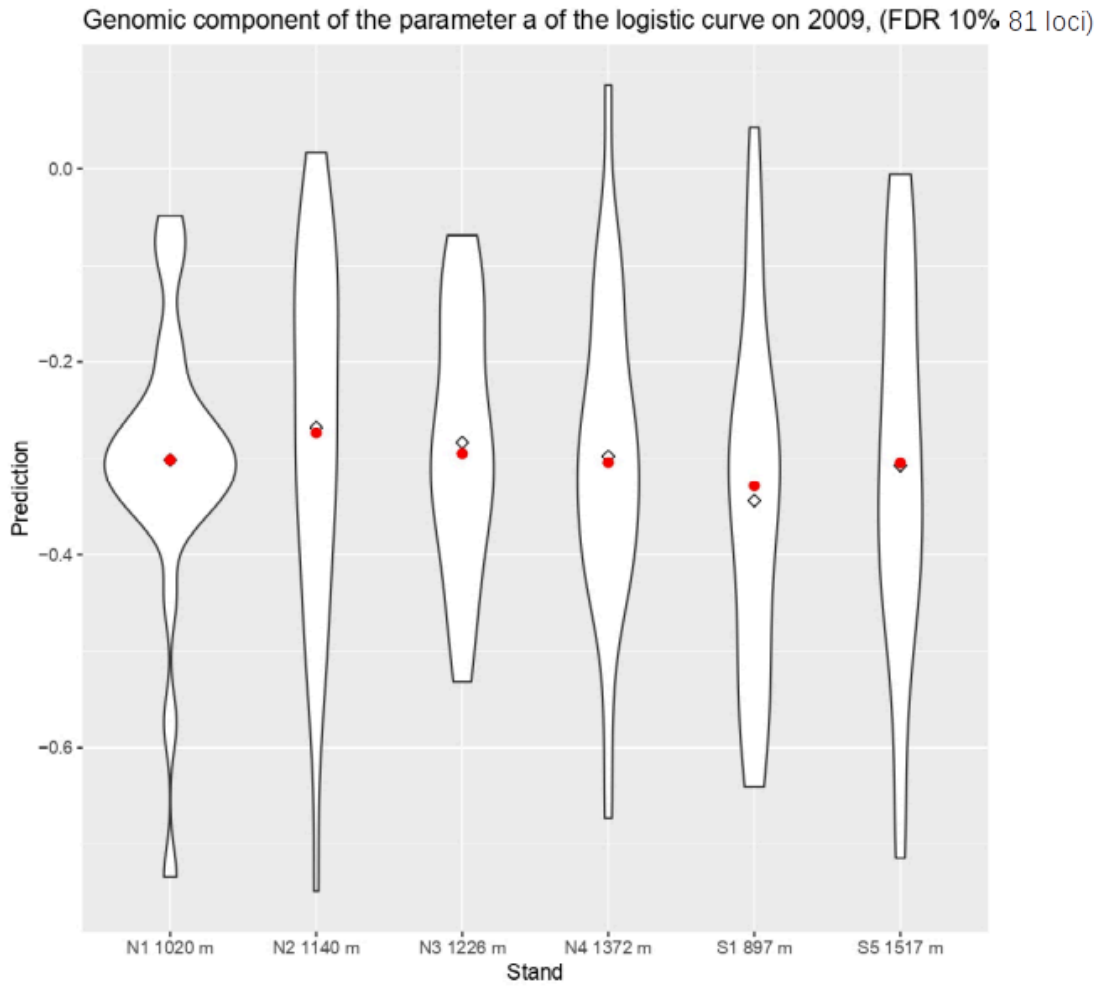
OBJ OBJ OBJ OBJ OBJ OBJ OBJ OBJ


Genomic component of date of BBCH 7 on 2009, (FDR 10%, 102 loci)
Prediction (DoY)
50
0
−50
−100
−150
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

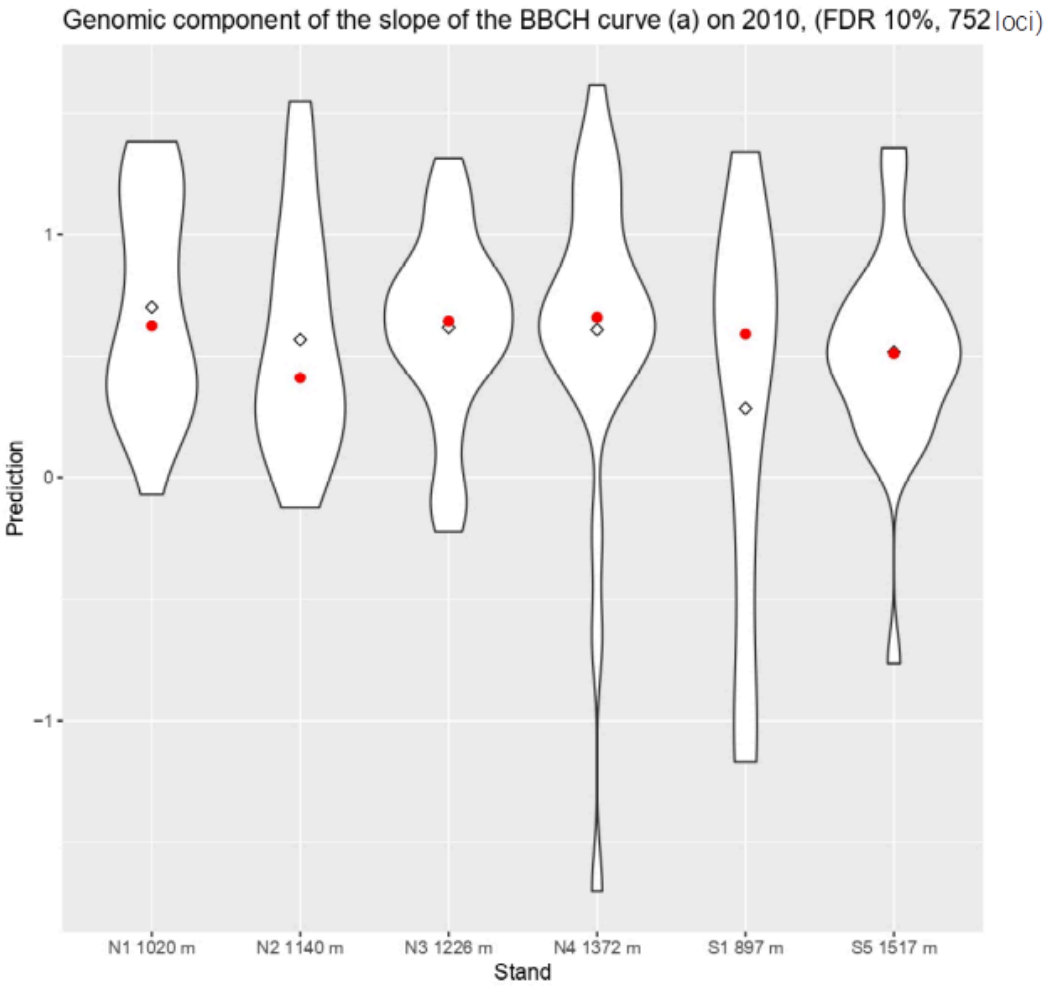

Genomic component of date of BBCH 7 on 2010, (FDR 10%, 213 loci)
Prediction (DoY)
−25
−50
−75
−100
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

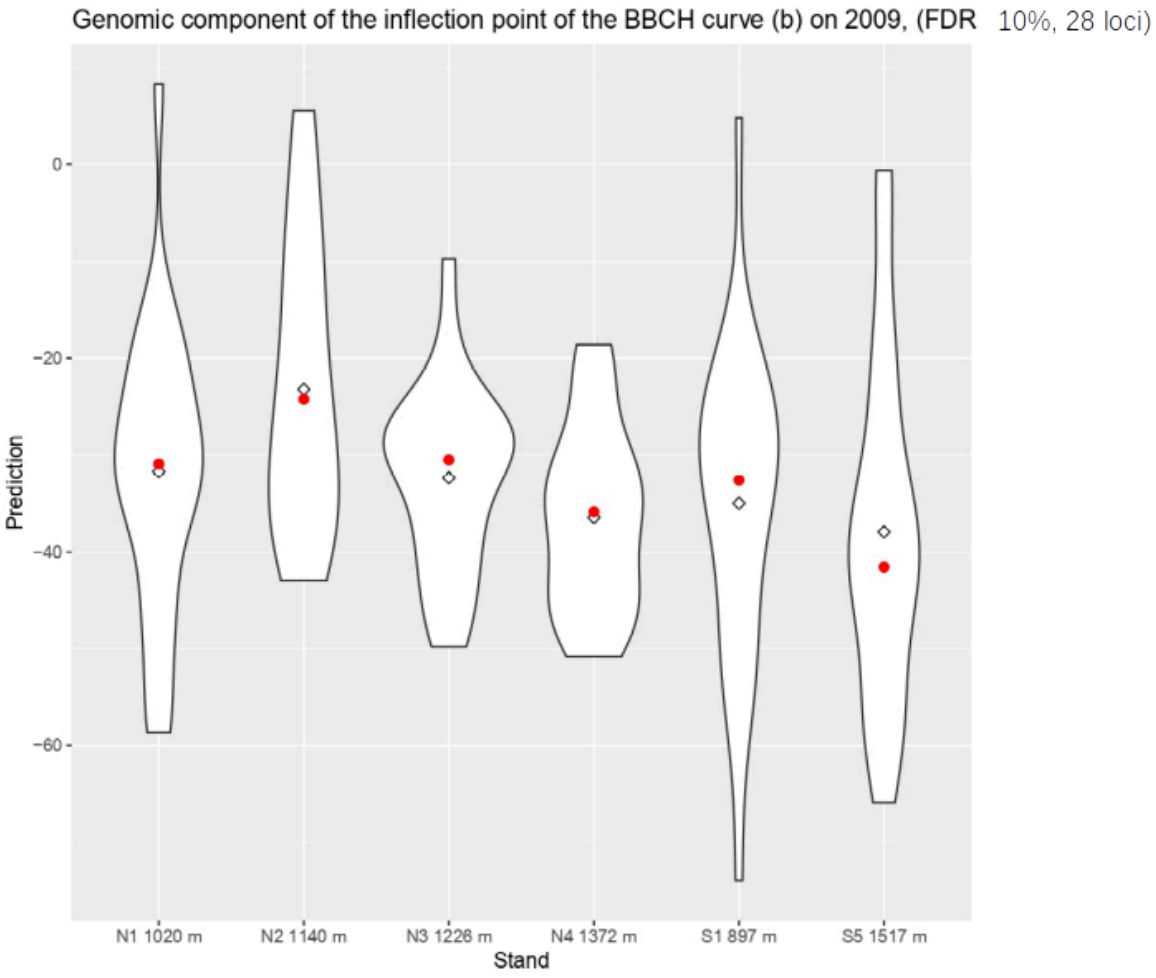

Genomic component of date of BBCH 9 on 2009, (FDR 10%, 189 loci)
Prediction (DoY)
200
100
0
−100
−200
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

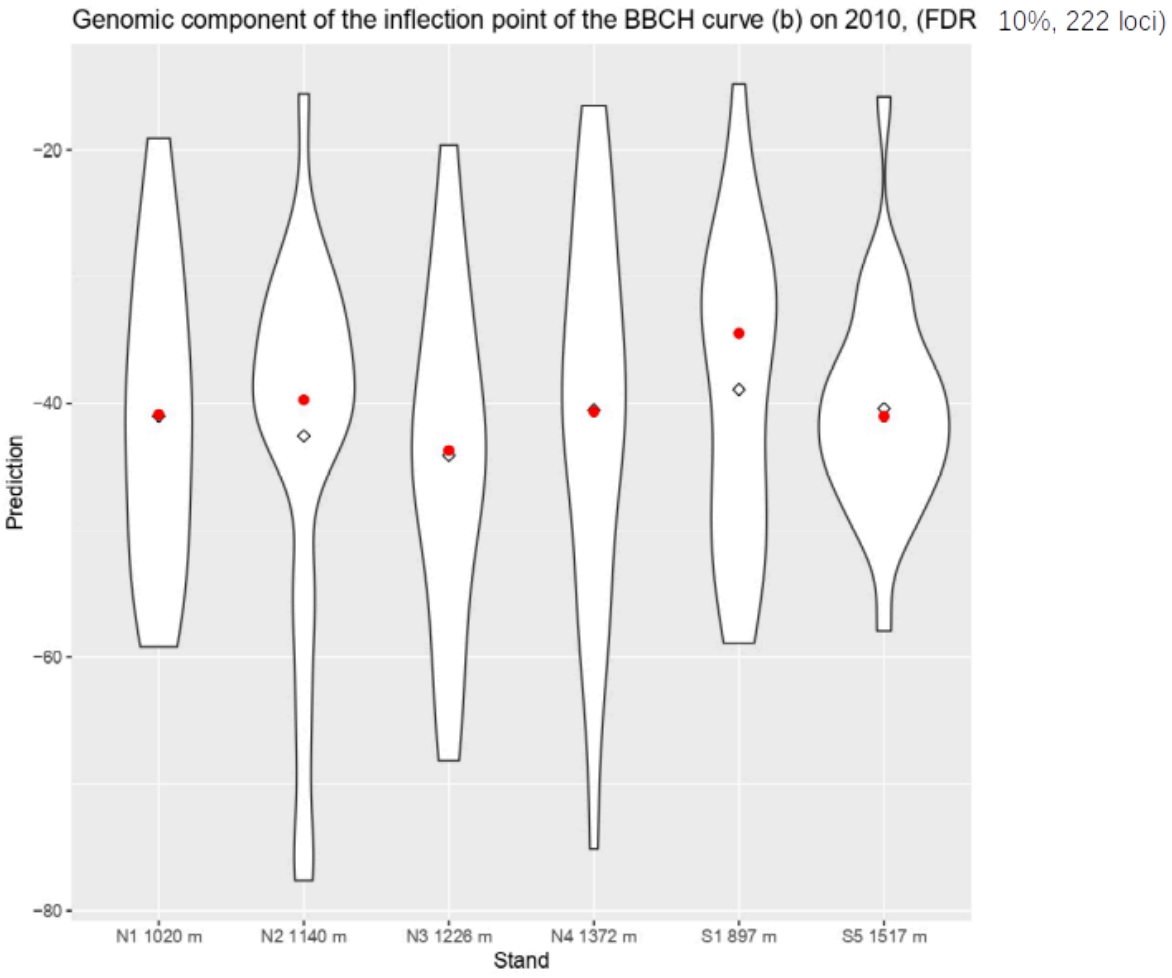

Genomic component of date of BBCH 9 on 2010, (FDR 10%, 443 loci)
Prediction (DoY)
0
−50
−100
−150
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

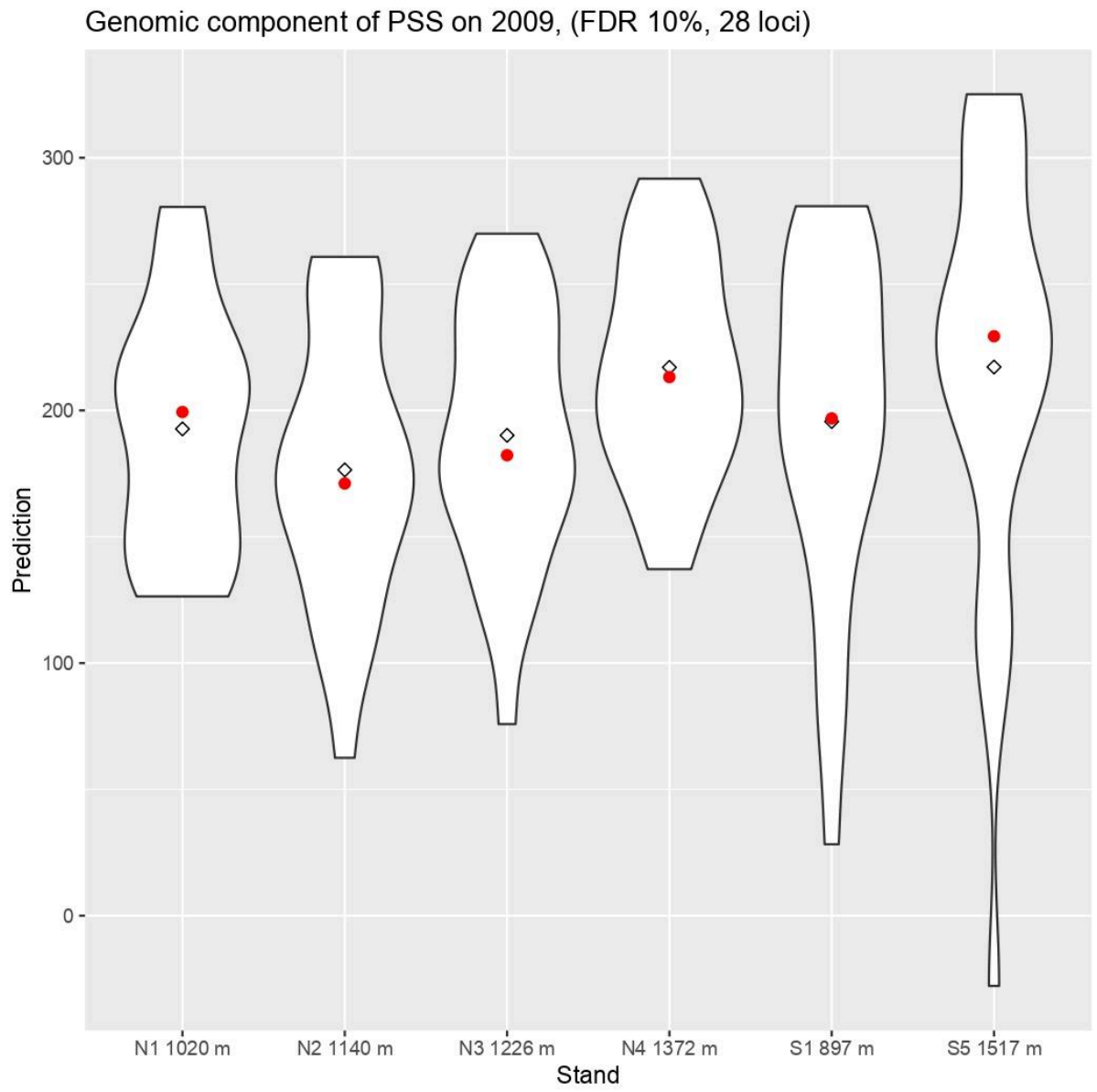

Genomic component of the parameter a of the logistic curve on 2009, (FDR 10% 81 loci)
0.0
−0.2
−0.4
−0.6
Prediction
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

OBJ

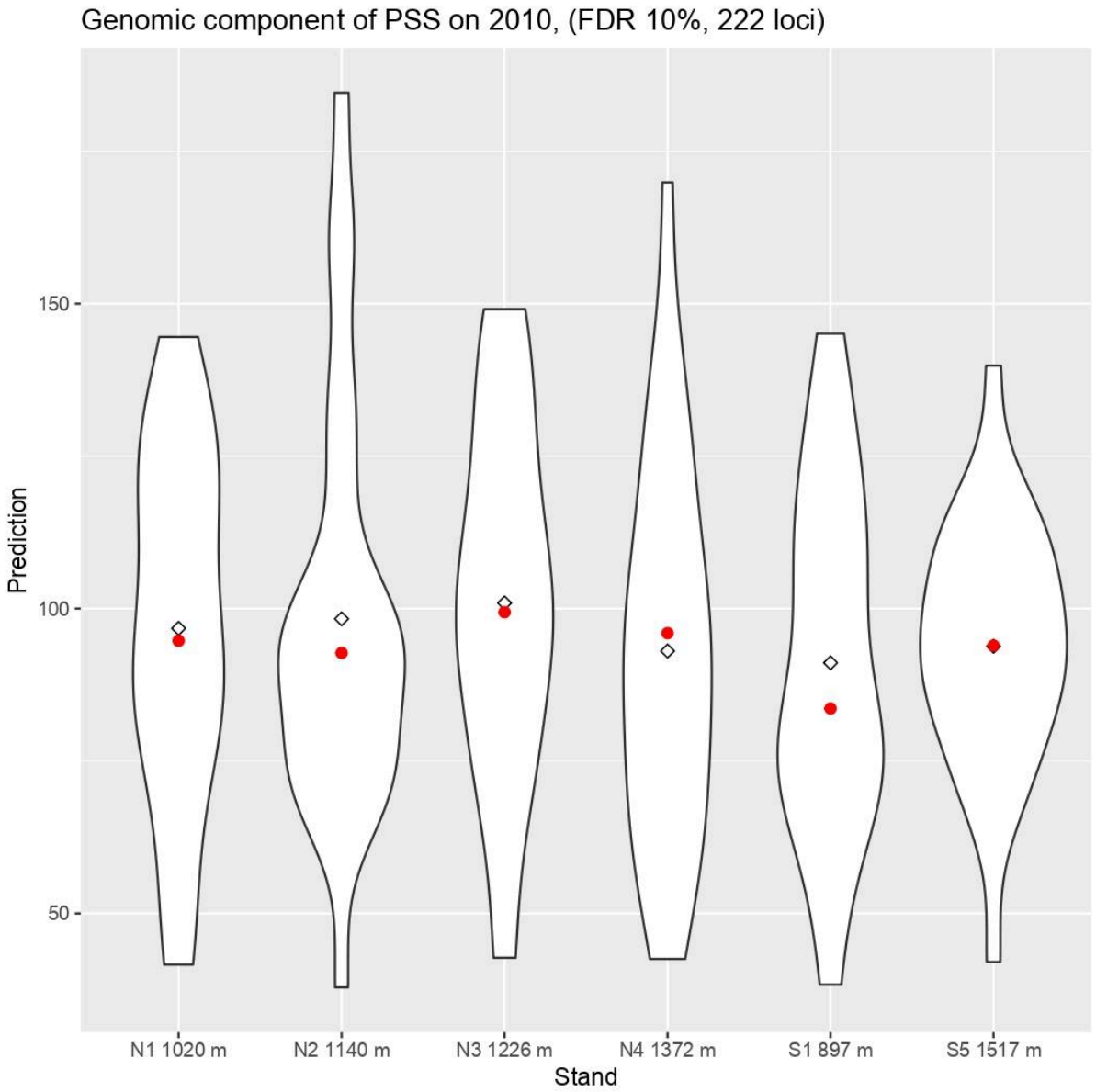

Genomic component of the slope of the BBCH curve (a) on 2010, (FDR 10%, 752 loci)
Prediction
1
0
−1
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

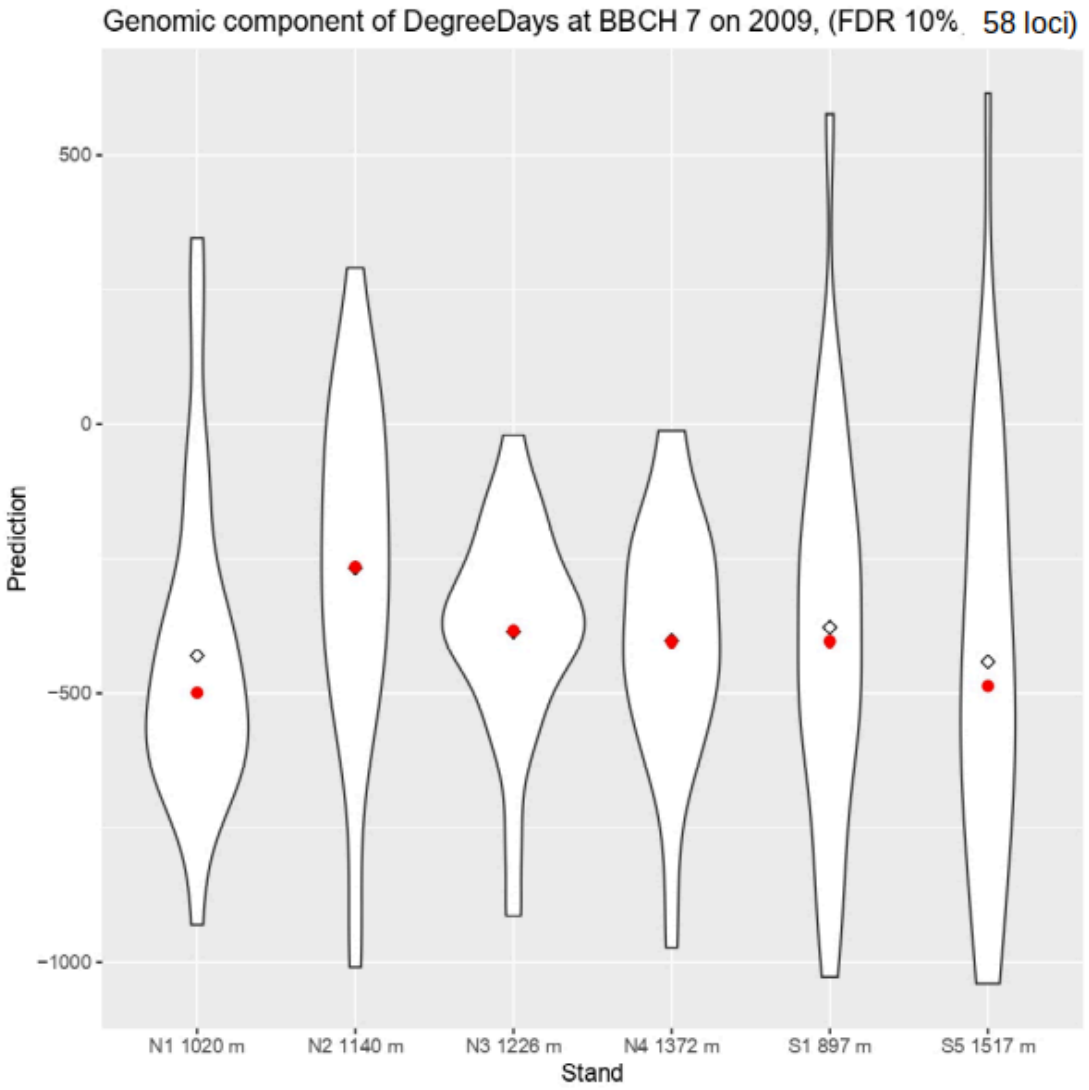

Genomic component of the inflection point of the BBCH curve (b) on 2009, (FDR 10%, 28 loci)
Prediction
0
−20
−40
−60
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

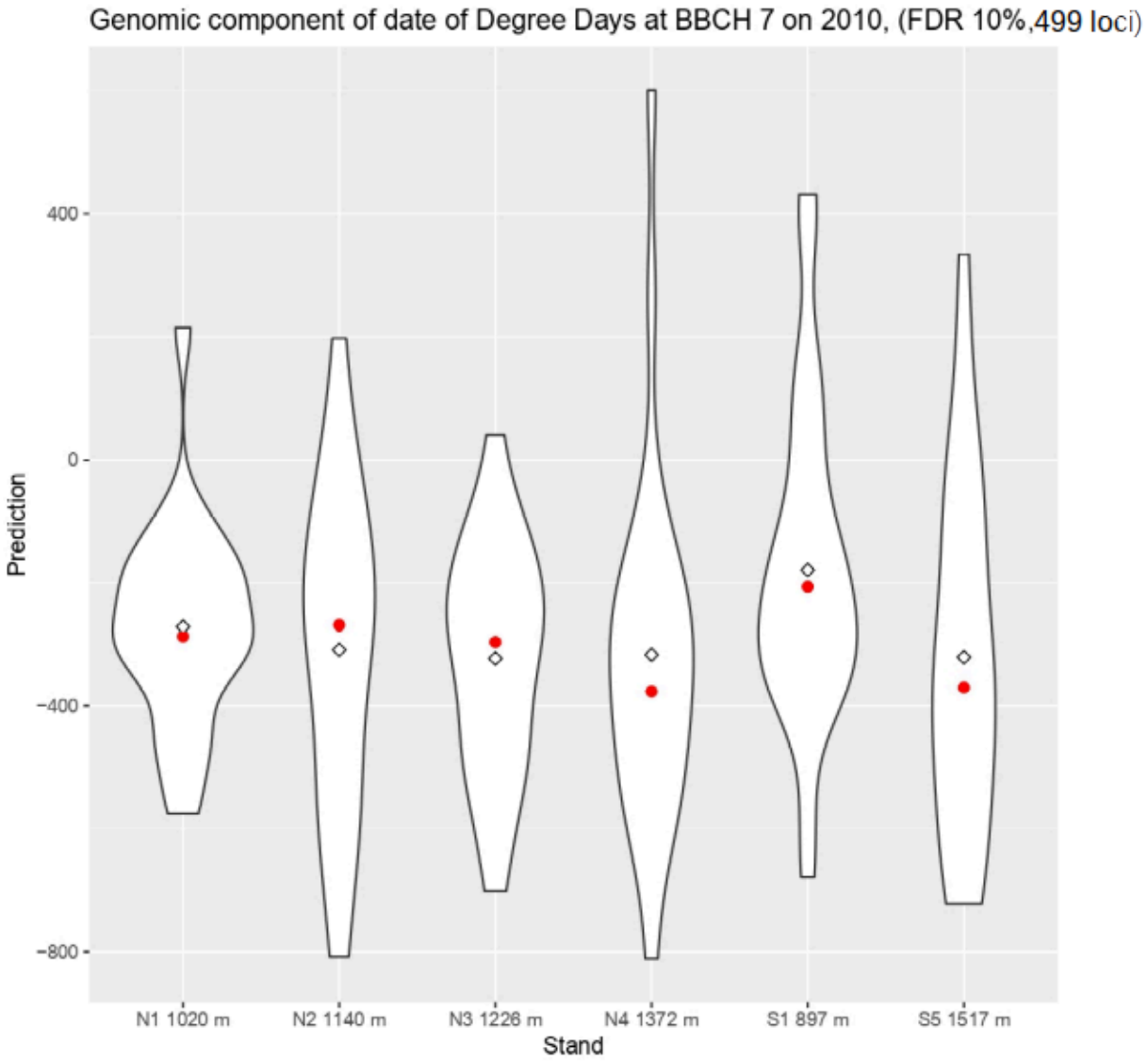

Genomic component of the inflection point of the BBCH curve (b) on 2010, (FDR 10%, 222 loci)
Prediction
−20
−40
−60
−80
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

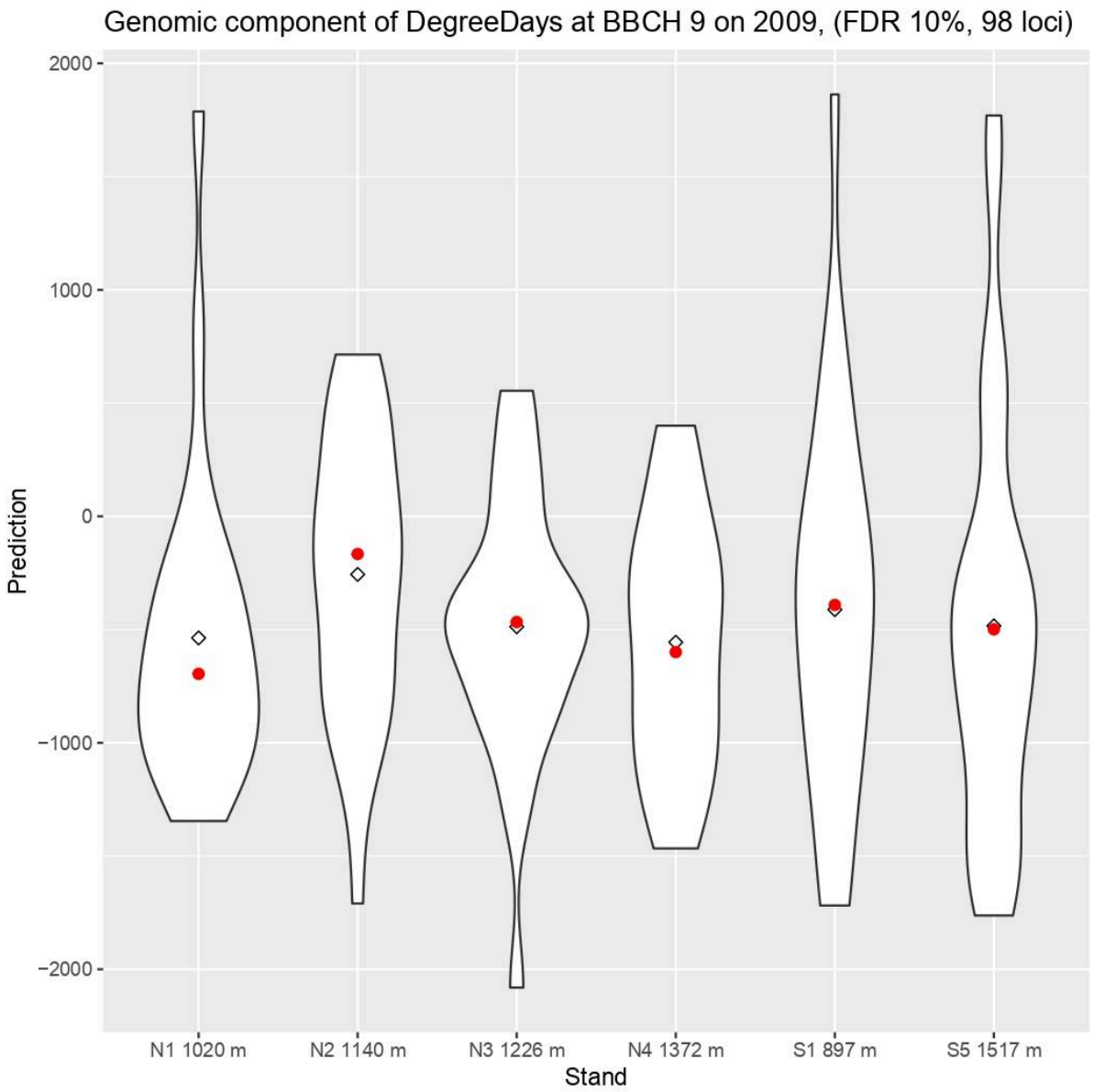


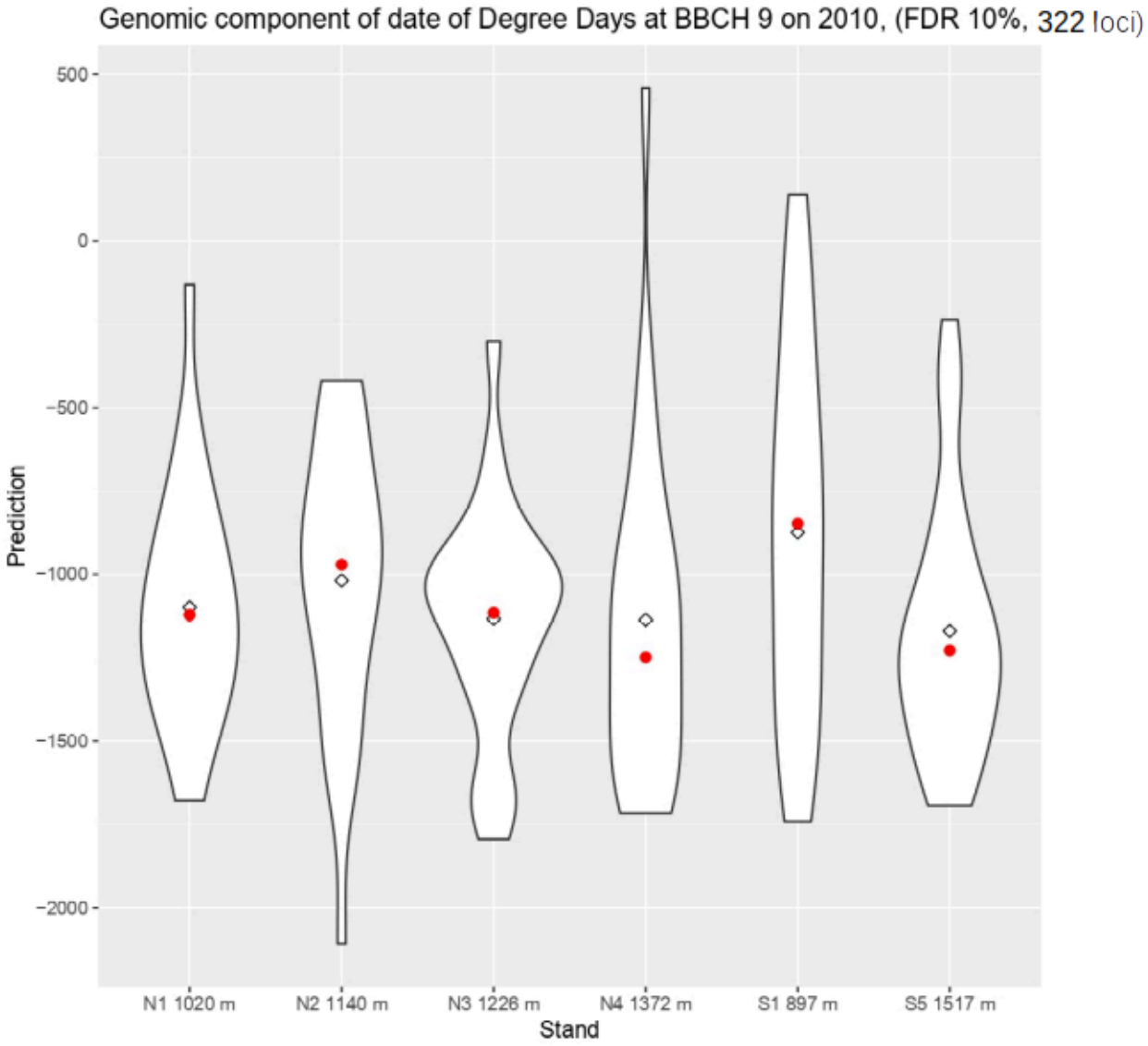
OBJ OBJ OBJ OBJ


Genomic component of PSS on 2010, (FDR 10%, 222 loci)
Prediction
150
100
50
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand


Genomic component of DegreeDays at BBCH 7 on 2009, (FDR 10%, 58 loci)
Prediction
500
0
−500
−1000
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand


Genomic component of date of Degree Days at BBCH 7 on 2010, (FDR 10%,499 loci)
Prediction
400
0
−400
−800
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand


Genomic component of DegreeDays at BBCH 9 on 2009, (FDR 10%, 98 loci)
Prediction
2000
1000
0
−1000
−2000
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand


Genomic component of date of Degree Days at BBCH 9 on 2010, (FDR 10%, 322 loci)
Prediction
500
0
−500
−1000
−1500
−2000
N1 1020 m
N2 1140 m
N3 1226 m
N4 1372 m
S1 897 m
S5 1517 m
Stand

6. General Discussion

My thesis provides new pieces of evidence to our knowledge on the adaptive response of forest trees to environmental gradients from the standpoint of spring phenology.

In the wild, European beech populations growing at low elevations are observed to start their spring phenology earlier than those at higher altitudes (Davi *et al.*, 2011). Since the 1980s, provenances from low-elevation populations were observed to flush leaves slightly later in the year than the seedlings sampled at higher altitudes when grown in controlled conditions (Teissier Du Cros *et al.*, 1988). Whilst the controlled conditions suggested that the trend was genetically determined and possibly adaptive, the study covered a regional scale and further hypotheses on what was the scale of differentiation were hampered by the lack of observations of samples from different altitudes within the same populations. A similar trend of "high altitude, early spring" was observed at a continental scale (von Wuehlisch, 1995) and replicated with Polish (Chmura and Rożwowski, 2002) and German provenances (Gömöry and Paule, 2011). A local *counter*gradient was observed investigating genetic differentiation for TBB in individuals sampled at the same site from different elevations along several Pyrenean mountain sides (Vitasse *et al.*, 2009). In common garden conditions, beech provenances of the same site growing in a colder temperature had a lower (earlier) TBB, that the authors linked to a lower temperature requirement to initiate spring phenology (Worrall, 1983). However, there still existed the chance that the phenomenon was an artefact due to the very common garden system - that is: higher elevation provenances flushed earlier in common garden because that time suited *the common garden* conditions rather than because the cues affecting them in the wild were evened out. Field quantifications of the TBB genetic component needed to be put together with common garden observation and the effect of phenotypic plasticity.

Meanwhile, historical records of phenological observations found that spring was happening earlier each year consistently with the warming of climate (Menzel *et al*. 2006), but spring advancing was asymmetric between latitudes: the late-spring northern provenances were becoming earlier *faster* than the southern ones, following Hopkins' Law (1920). The inferred reason was that the asymmetrical spring advance was the effect of a shorter exposition to cold. Studies were now highlighting a continental scale north-south phenotypic anticline in wild conditions. The same trend was found also at a local scale: Gaüzere et al., (2017) predicted an altitudinal anticline of TBB in a Pyrenean forest for the period 1994 - 2098 on the basis of phenological time series recorded between 1997 and 2009. They modelled the interactive effect of chilling and photoperiod on TBB under increasingly warmer conditions and found out that, by 2098, the expected TBBs for the forest was globally lower (*ie*. earlier), and the trees of higher altitude displayed a faster progression toward early dates with respect to those growing 700 m below. It was the *rate* of the advance ($\Delta days/decade$) that was distributed anticlinally rather than the advance itself. Gaüzere's model stated that trees that were experiencing less time at colder temperatures became more photosensitive; and that increasing exposure to light made them more sensitive to changes of temperature. Even without long winters the combination of longer days in early winter and the lower exposure to cold still cascaded in a slightly earlier TBB in trees in warmer conditions - from southern or lower provenances: since photosensitivity is *never*

0, a bout of warming in a climate that is *still* cold enough (northern, higher up) will cascade in a faster anticipation of TBB year by year: trees of higher altitudes were anticipating the TBB even in colder conditions, because these conditions were becoming less cold (Gaüzere et al., 2019). Vitasse et al. (2018) bakced these findings and confirmed that the advance of spring is already asymmetrical between altitudes.

Building on these findings, Gaüzere, Teuf *et al.* (2020) estimated the phenotypic optimum for TBB along a mountain slope and compared it with field and estimated dates of TBB. They found a congruence between the three kind of data, suggesting that plasticity is adaptive; but at higher elevations beech exhibited an hyperplastic response to the colder environment that shifted the observed and estimated TBB later than the optimum; hence, these stands experienced a strong selection for earlier TBB. From these data, Gaüzere, Klein *et al*., (2020) showed with a quantitative genetic approach that the saplings grown in common garden followed the anticline of genetic differentiation in the determination of TBB: on Mont Ventoux, low-altitude provenances stands were genetically *late* than high-altitude ones, but in the wild phenotypic plasticity generates a cline e of TBB (**Figure 1a-b**). Gene flow has been quantified to be consistent along the gradient (Gaüzere *et al*. 2013), so the genetic differentiation is a signature of adaptation at a spatial scale (232 m) within the dispersal distance of European beech (~100 m with rare long-distance events; Gaüzere *et al.*, 2013; Oddou-Muratorio *et al.*, 2018). The counter-gradient appears to be the result of different effects of genotype and environment on TBB (Conover and Schultz, 1995).

This thesis takes off from this cliff to investigate how contrasting selective pressures among altitudes have driven the differentiation between stands separated by a few hundred meters, highlighting the genetic patterns divergence in the genomic regions that control TBB, a relevant adaptive trait. The approach I used took two complementing paths.

Chapter 1 investigates selection by addressing the confounding effects of demofgraphy and remaining phenotype-agnostic. It reports the detection of signatures of molecular divergence at a fine geographical scale (120 to 620 m) between beech stands growing in contrasted environmental conditions. This finding is even more relevant because *and* in spite of the population having recently expanded (two generations): "*because*", as these results are evidence that beech can adapt to novel environment at a rapid pace; "*in spite of*", because range expansions complicate the detection of selection signatures in populations that have recently expanded their range.

Chapter 2 does the opposite: it puts the phenotype in the limelight by looking at its genetic architecture and frames its question within a single stand, evening out demography. It unravels a portion of the complex genomic basis of vegetative spring phenology by estimating new synthetic phenotypes that capture several processes concerning timing and pace of the phenomenon.

Chapter 3 comes full circle. It moves from the framework of genomic differentiation between stands assessed in Chapter 1 and, relying on breeding values of the QTL for TBB identified Chapter 2, it successfully predicts the genetic counter-gradient observed in the common garden. Then, it models the fate of a stand of high elevation under a different optimum and assess that the genetic diversity at these QTLs is enough for the stand to evolve toward a new optimum. This finding confirms the hypothesis that the genes controlling TBB are differentiated along the northern slope (352 m) and proves the evolutionary potential of TBB in a marginal population.

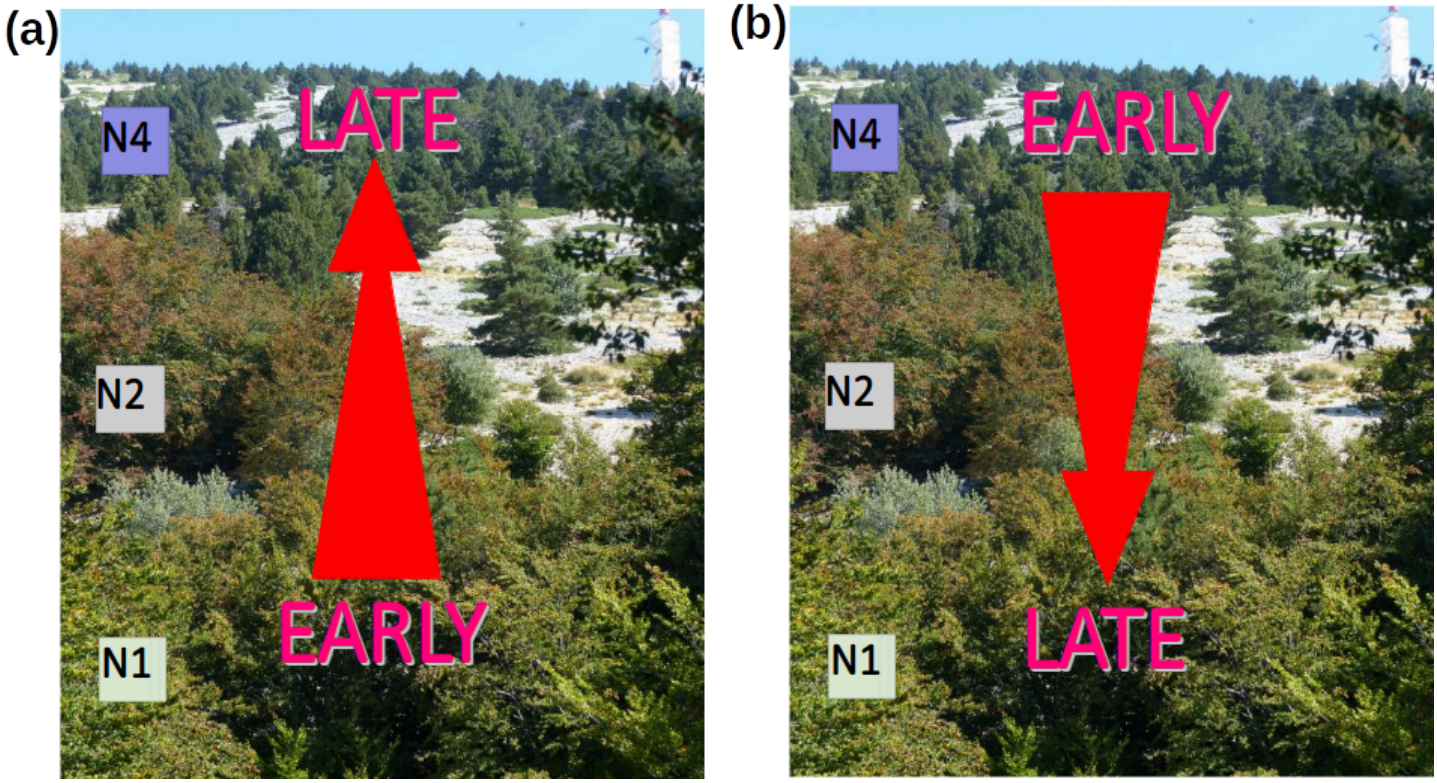


Figure 1: Clinal patterns of TBB. In the wild (**a**) environmental cues cause trees of higher elevations to flush their leaves later in the year than those from lower elevations (co-gradient), while in controlled conditions (**b**) trees from higher elevations flush their leaves earlier than those from from lower elevations (counter-gradient).

6.1 “When Action Grows Unprofitable, Gather Information”
Explicitly Addressing Demography Using Interdisciplinary Sources

The identification of signatures of putatively adaptive divergence along Mont Ventoux slopes confirms that there exists molecular differentiation at the scale of the landscape. During the second part of the XIX century the population has expanded downward from a high elevation historical relict area. To reduce the signals caused by the non-adaptive range shift, I explicitly modelled the demographic process and included it in the test for divergence as the null model; then, I used a set of methods resting on different statistics and targeting different genomic scales and found molecular signals of divergence that partially overlapped, with some single-locus outliers mapping within outlier genomic regions.

Recorded population expansions along environmental gradients represent a “natural” experiment to understand how spatially heterogeneous selection regimes shape the genetic diversity within the newly expanded demes and between the newly colonised area and the home habitat. External, environmental pressures are not the sole forces driving the genomic diversity of a (meta)population along a gradient; the way selection shapes genomic diversity depends on historical, internal and biological factors: the number of generations elapsed since the establishment of the population; the standing genetic diversity of the parental population; the rate of immigrants between demes; local genomic recombination rate.

In tree species, large population sizes preserve polymorphisms for long times and the extensive gene flow is expected to smoothen differentiation. However, in common garden experiments trees display very large differentiation at genomic regions controlling traits with strong link to fitness (*e.g.* TBB, cold tolerance) coupled with low differentiation at neutral markers (Savolainen *et al.*, 2007; Le Corre and Kremer, 2012). Due to their long generation times, today’s adult trees are the yesterday’s sapling that survived to yesterday’s environment, so they display patterns of climate-driven selection due to spatially *and* temporally heterogeneous selection regimes.

Methods that rely on different approaches to dissect genetic diversity can bring to light the various facets of adaptations, allowing the discrimination of more recent signatures (*ie*. single-locus sweeps) from signals that become noticeable only after more generations (*ie*. variance decomposition) or by looking at frequency covariances at multiple linked loci, either at distinct (*ie.* monogenic shift of allele frequency at a large effect locus) or small effect loci tattered across the genome (*ie*. polygenic adaptation). A first, simple, take-home message when investigating selection is to use multiple approaches that look for distinct facets of the process, because it unrolls in many convoluted ways.

On a molecular level, populations undergoing range expansion can exhibit genetic patterns of differentiation that can mimic signatures of selective processes, while they have instead resulted from the random sampling of gametes in the expansion area with no deterministic driver. Though deleterious variants are purged out in a few generations, neutral alleles may persist at high frequencies as if they have a selective advantage (Klopfstein *et al.*, 2006; Excoffier and Ray, 2008). Subsequent bouts of migration refuel the genetic diversity in the newly colonised habitat, but also reiterate the drift-like sampling of gametes of the previous expansion. Multiple founder events can lead common variants to structure a cline of frequency along the expansion axis that mimics a gradient of adaptive genetic differentiation (Barbujani *et al.*, 1997; Currat and

Excoffier, 2005; Slatkin and Excoffier, 2012) or result in population structure across the landscape (Barbujani *et al.*, 1995). This bundle of events swamps the signal from the loci that have undergone directional selection, as some loci across will diversify more between the parental and the newly colonised area only due to random sampling during the inception of the latter. Over generations, the genetic diversity and large effective population size of tree populations can buffer these effects, though it is difficult to state at which degree. Classic models to detect divergence test the distribution of their differentiation statistics against a theoretical null model where migration/selection processes are assumed at equilibrium. Tautologically, a population that has recently expanded or contracted is *not* at equilibrium: relying on a model that does not account for the (non-selective) diversification due to inception risks to overestimate the between-demes divergence at several loci (Charlesworth, 1994; De Mita *et al.*, 2013; De Villemereuil *et al.*, 2014; Lotterhos and Whitlock, 2015).

Being a discipline rooted in processes that take place over time, researching in the field of population genetics entails framing its questions into a multidisciplinary context. Historical archives are extensively used in human population genetics (*e.g.* Jobling, 2001; Robledo *et al.*, 2013; Luconi *et al.*, 2023) and findings from other fields often underpin reconstructions of past range change at a continental (Magri *et al.*, 2006; Aitken *et al.*, 2008) or regional (botanic surveys in Conte *et al.*, 2020; herbaria in Buldrini *et al.*, 2023). In the same way, any researcher that aims to work on adaptation at the landscape scale has to control the demographic sources of noise whenever possible by fetching information on the past of the population of study and use them to model these events. Being a discipline rooted in processes that take place over time, researching in the field of population genetics entails framing its questions into a multidisciplinary context. Historical archives are extensively used in human population genetics (*e.g.* Jobling, 2001; Robledo *et al.*, 2013; Luconi *et al.*, 2023) and findings from other fields often underpin reconstructions of past range change at a continental (Magri *et al.*, 2006; Aitken *et al.*, 2008) or regional (botanic surveys in Conte *et al.*, 2020; herbaria in Buldrini *et al.*, 2023).

Even though molecular data are often used to control for demography when investigating divergence (Brousseau *et al.*, 2021; Scotti *et al.*, 2022), to my knowledge most studies addressing the landscape scale have neglected to explicitly integrate interdisciplinary sources in their hypotheses. The first chapter of the thesis shows that past events may leave confounding information in the genetic data to rely only on the latter ones without a strong hypothesis. Historical maps, documents and experts from other fields - such as historians, palynologists, and forest rangers - can help to collate and interpret maps and documents, as they have expertises that the background of a population geneticist lacks. Historical information helps infer and date past events, or to constrain parameters to estimate within credible boundaries whenever the information available is scarce. Indeed, this thesis would not exist without the bibliographical work of Tonya Lander (2011, 2020) and Hélène Prouillet Leplat in the archives of the Office National de Forêts, and the charcoal residues retrieved by Gamba *et al*. (2023) gave further strength to Lander's reconstruction.

My approach follows the guideline proposed by Siol *et al*. (2010) for includeing demographic events within the null hypothesis. In the first step, I reconstructed and dated events to a time window thanks to historical maps, archeological, and biogeographical information. In a second phase I inputted the model to a coalescent-based simulation algorithm to estimate the most likely values of the demographic parameters of such a scenario assuming that the observed

(field) allele frequencies are the result of the dated events. In the final step, the algorithm uses the estimated parameters to simulate the genetic diversity expected from the processes that have historically involved the studied population(s) - it generates an “empirical” simulated distribution of allele frequencies, rather than theoretical. I computed the test statistics from the simulated frequencies to generate an “empirical” simulated *null* hypothesis of neutral differentiation that includes the fluctuations in allele frequencies expected from the modelled (*known*) events. The approach is non-parametric, so I used single-tail Wilcoxon tests to see whether the demographically-correct genetic diversity estimates were larger or lower than the field estimates.

Including the effect of demography in the null model resulted in a large reduction of the number of positive tests in the pairwise Bamova analyses (Gompert and Buerkle, 2011), which were very likely false positives. The magnitude of the difference ranged from 40-fold (S1 - S5: 120 SNPs without accounting for demography, three when accounting for it) to not detecting any signal when the demographic background was tested against (N3 - S5, N3 - S1). It also allowed me to test differentiation on crude differences in allele frequencies, that do not have a theoretical distribution, and multilocus shifts from a demography-explicit expectation.

This thesis provides new evidence of rapid adaptation to novel microenvironments within the species’ dispersal distance. I found that patches of trees within the larger beech population display genomic signatures of selection at a few (five) SNP loci mapped within coding regions. Differences in frequencies were identified also in 17 genomic regions (haplotypes). The proportion between single- and multilocus signals suggests that local adaptation has evolved through strong selection at few individual loci and on tightly linked loci in which individual signals are too weak to be detected by single-locus method (Le Corre and Kremer, 2003, 2012; Kremer and Le Corre, 2012; Cubry *et al.*, 2022; Yeaman, 2022). The percent of the genome identified to be under divergent selection (0.01 - 0.19%) is consistent with the estimated extent of the genome underlying local adaptation (0.16 % in *Pinus cembra* L. and *Pinus mugo* Turra, Mosca *et al.*, 2016; *Eperua falcata* Aubl., Brousseau *et al.*, 2021; *Abies alba* Mill., *Cedrus atlantica* (Manetti ex Endl.) Carrière, *Pinus halepensis* Mill., and *Pinus pinaster* Aiton, Scotti *et al.*, 2022).

Given the rigorous control of the effect of demography in the search for these loci, the beech stands are likely to be locally adapted. This is backed up by Gaüzere, Teuf *et al.* (2020), who reported genetic differentiation for TBB in sapling coming from stands at different altitudes when grown in a common garden experiment. Loci under divergent selection were identified between the expansion and the refugial area (three SNPs and seven genomic regions), and within the refugial area itself: between the two slopes (two SNPs and seven genomic regions) and along the same slope (three genomic regions). The findings suggest that in the population of Mont Ventoux adaptation has occurred very fast (about two generations) and within very short spatial scales (down to 150 m) (**Figure 3**). Differentiation occurring over approximately the same time scale has been reported between stands of *C. atlantica* growing in contrasted habitats (Scotti *et al.*, 2022). These findings add to a body of evidences that adaptive divergence is a widespread phenomenon within tree populations that inhabit granularly heterogeneous habitats; that the scale of its genomic signature is consistent between taxa; that it can occur over few generations following a range expansion; and demography must be explicitly accounted for when testing for selection, gathering as many sources as it is possible.

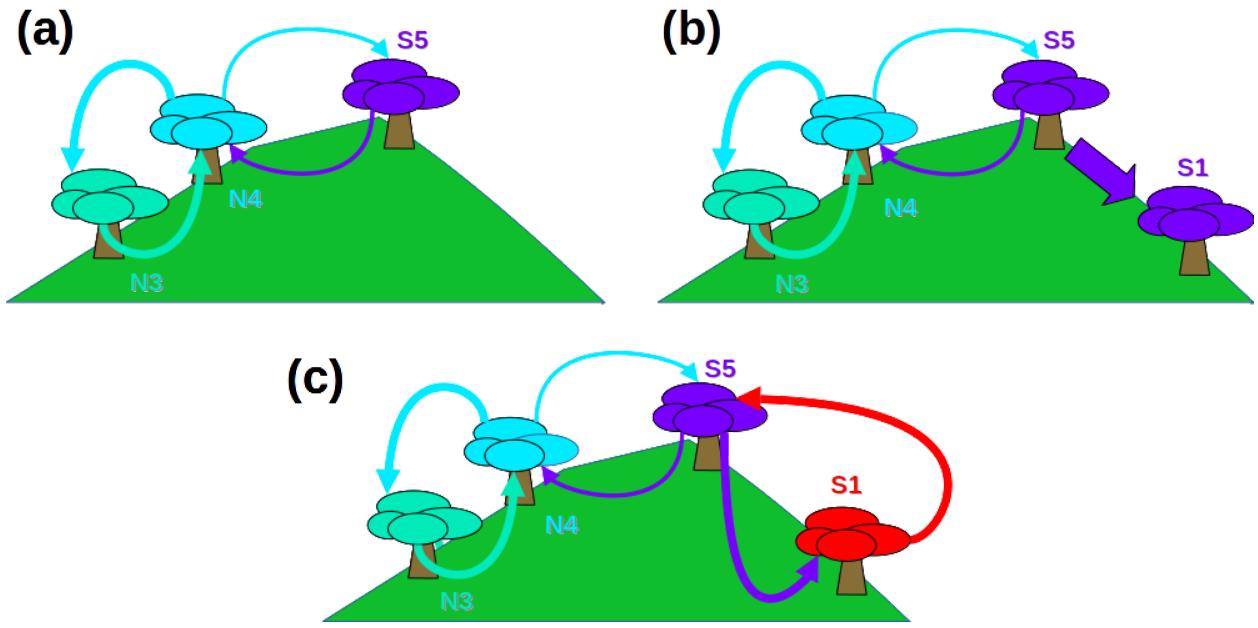


Figure 2: Stand diversification during the demographic expansion (1850s-200s). In a first (**a**) phase, the cold-adapted stands N3, N4, and **S5** in the relict area exchange migrants. Following decreased anthropic pressures and the ONF reforestation plan (**b**), migrant seeds and pollen from **S5** establish a new stand down the mountain slope. After a few generations (**c**), **S1** has diverged from **S5** and displays signatures of local adaptation to warmer conditions.

6.2 Not All Tangles Ask for a Razor

Parsimony in complex scenarios

Having a lot of documentation about the past of the studied population is a lucky, happy, and unusual situation. One would be tempted to write a very complex demographic model to cover all the known events expected to affect the modern patterns of genetic diversity. The more the events, the more the parameters; and overparametrising often leads to only describing the observed data instead of representing a phenomenon in a way that allows it to predict its outcome. Ockham's razor is a heuristic that argues for shaving away too complex explanations because that *pluralitas non est ponenda sine necessitate* - “entities should not be multiplied beyond necessity”. While the quote was an exhortation to not make stuff up, it is nowadays used as a justification to value the simplest explanation as the most likely one by virtue of that very simplicity. But it is evidence that should drive the explanation, not parsimony: what we may want to explain can *actually* be complex (Oreskes, 2024). A response that increases slowly for a large range of values of its predictor, followed by a steep increase by small increases of the predictors, that finally reaches a plateau will be better represented by a sigmoid rather than a straight line (**Figure 4**). To pick a biological example, Lynn Margulis’ theory of the endosymbiotic origin of eukaryotes (Sagan, 1967) looked very intricate *and* outrageous, even though it was backed by significant evidence. Quoting J.B.S. Haldane, “The universe is not only queerer than we suppose, but queerer than we *can* suppose” and to better understand it we can only examine what we see and make good use of it. If the information available on a population history say that it is efficiently described by 21 parameters, it will be described by 21 parameters - a reasonable number to model the evolutionary trajectories of 44,000 loci.

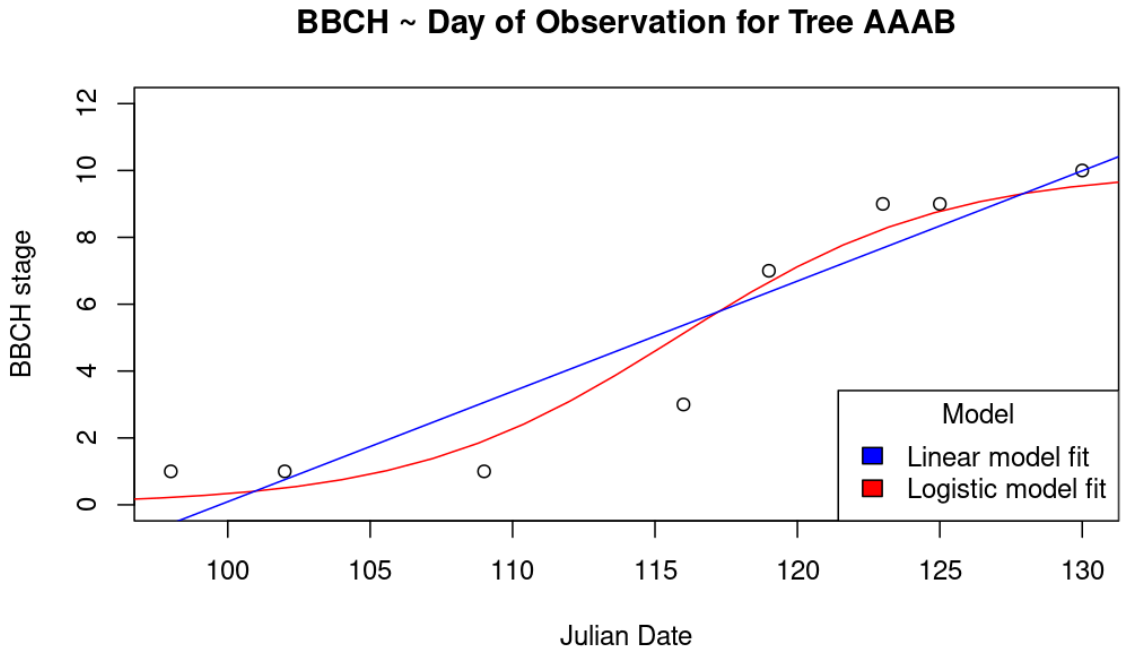


Figure 4: Though “simpler”, a linear model is a worse fit than a sigmoidal model to time series data of vegetative spring phenology.

6.3 This Year's QTLs

Distinct Architectures Drive Bud-Break Under Different Climates

Within Chapter 2 I described a Genotype-Phenotype Association (or Genome Wide Association Study, GWAS) experiment to identify the loci that underlie the phenotypic variation of spring phenology in a beech stand growing below those investigated in the first chapter (1,020 m). Starting from a time series of phenological scores recorded in 2009 and 2010 in 162 individual trees, I modelled the relationship *PhenologicalScore ~ Day* using a logistic model. From these phenotypes I computed for each tree the dates of passage for the BBCH stage 7 and 9, the respective degree days, the sum of phenological scores, and two parameters of the logistic curve that worked as proxies for the rate of change of the phenological transition (Logistic.a) and the timing of spring phenology (Logistic.b). The analyses were run with the algorithm of FarmCPU using a kinship matrix to control for the relatedness.

Traits with broad continuous phenotypic variation are likely to have a polygenic architecture, where few loci have a large effect on the trait and many others contribute to it by a small amount each. Vegetative spring phenology - in particular the time of bud burst - are among the most responsive traits to environmental cues. Along an environmental gradient, the rapid turnover of selective pressures generates a cline of phenotypic optima that subpopulations reach through the shift of various combinations of allele frequencies at the small-effect loci. Such a complex architecture results in the trait being able to respond to very different selective regimes. Which loci are involved in the process depends on the ancestral pool of allele frequencies, so the power to identify the genomic bases of the trait and the loci that get identified are strictly dependent on the studied population and on its environment. Performing GWAS on natural populations is a double edged sword. On one side the environmental effects on the trait hampers the identification of the causal loci. On the other hand, the observed phenotypic variation is expressed in the actual environment of the species, while in common garden conditions it may be constrained by conditions that trees do not meet *in silva*.

Several trees have not been recorded in the studied BBCH stages. Rather than reducing an already small sample size for each trait, I modelled the phenological score as a function of the day using a logistic function. The sigmoidal curve of the relationship maps a discrete variable (the BBCH) into a continuous one (the day). Its more straightforward advantage is to fill the gap of a time series, as it estimates the expected day in which a tree must have been recorded at a given score on the basis of the distribution of the observed scores. The logistic model of the day of passage also provides two synthetic traits in the form of the curve parameter: the slope *a*, that mathematically represents the rate of change of the phenological transition; and the inflection point *b*, the timing of spring phenology.

Most importantly, I detected loci significantly associated with variation in all the investigated traits. This is far from a foregone conclusion: phenotypes collected in natural populations are the result of several genotype x environment interactions that blur the association between trait value and genotype; this diminishes the power of the algorithm to detect signals Furthermore, were related individuals with a similar phenotype together within the same sample, they may generate spurious association because of the higher likelihood to have loci in the same allelic phase to a linked QTL due to inheritance, because recombination has not yet broken down the

association between locus and QTL. The Training/Testing approach assessed the robustness of the results of the Main datasets to its sampling composition, which could affect the power to detect significant associations. Were subsamples of different individuals from the same population able to replicate the findings of the main datasets? If the loci retrieved by at least two training datasets could predict the phenotypes of a congruent subsample of individuals, they could be deemed to have enough power - and by extension the Main datasets as well.

The results reveal that spring phenology has been regulated by distinct sets of loci each year. This was expected, because the pattern of scores within the time series were significantly similar. However individuals that were early one year were not so the following one, suggesting genotype x environment interactions have affected the phenological patterns in spite of their similarity (**Figure 5**). A similar non overlap of the genes associated with phenology in consecutive years was observed in beech provenances from all the species' range grown in common garden (Lazic *et al.*, 2024). On a broader taxonomic scale, different populations of *Arabidopsis thaliana* have been found to adapt to climate at a fine spatial scale relying on different suites of loci between microhabitats (Frachon *et al.*, 2017, 2018).

The main constraints to detect inter-annual QTLs appear to be the sample size and the extent of the series of years of observation, as they both improve the statistical power (Niels Müller, personal communication). Testing this hypothesis requires to record an analyse time series of spring phenology over more than two years and compare several climatic variables over the same time period. Spring phenology is a highly polygenic trait, so, on smaller samples, different climatic conditions between years may reveal distinct suites of loci underlying it (mean temperature from 1st January to 10th May: 4.5 °C in 2009, 3.1 °C in 2010).

In spite of annual heterogeneity in trait architecture, it is possible to detect intra-annual patterns. In 2009, a core suite of seven loci coordinates spring phenology, as they are associated with all the traits but the rate of change of the phenology transition (Logistic.a). These seven loci were always associated with the dates of passage, and fewer loci involved in determining the two individual stages. Whilst five of them are in different chromosomes, two are ~22 kbp apart on chromosome 6 (Bhaga_6_11825186, Bhaga_6_11847868); Bhaga_6_11836592 (associated with the timing of spring phenology) is mid-road between the two, at ~11 kbp. Bhaga_6_11847868 and Bhaga_6_11836592 are in an exon and downstream, respectively, to a gene coding for a chaperonine that regulate transcription under thermal stress; Bhaga_6_11825186 instead is upstream of such gene, and maps within the intron of a membrane protein with catalytic activity. The discovered QTLs mapped mostly outside coding regions or introns. Polymorphisms within intron or noncoding regions are believed to account for a great part of the QTLs that drive TBB (Derory *et al.*, 2010), an expectation that has been confirmed by other genomic studies on tree species (Mckown *et al.*, 2014; Evans *et al.*, 2014), and in particular in beech (Lauren Clément, personal communication). Intergenic regions can harbour a cornucopia of functions, from binding sites for transcription factors (enhancers/suppressors; Weber et al., 2016; Marand et al., 2023) to intronic sites for alternative splicing (Zhang *et al.*, 2015; Martín *et al.*, 2021; Wright *et al.*, 2022) and loci coding for miRNA that control the quantity of transcript of gene farther away. In this chapter I successfully identified loci associated with spring phenology in a wild population, a context where the environment introduces noise and the small sample sizes hinder the discovery of QTLs. They appear to have a small but significant effect on the trait.

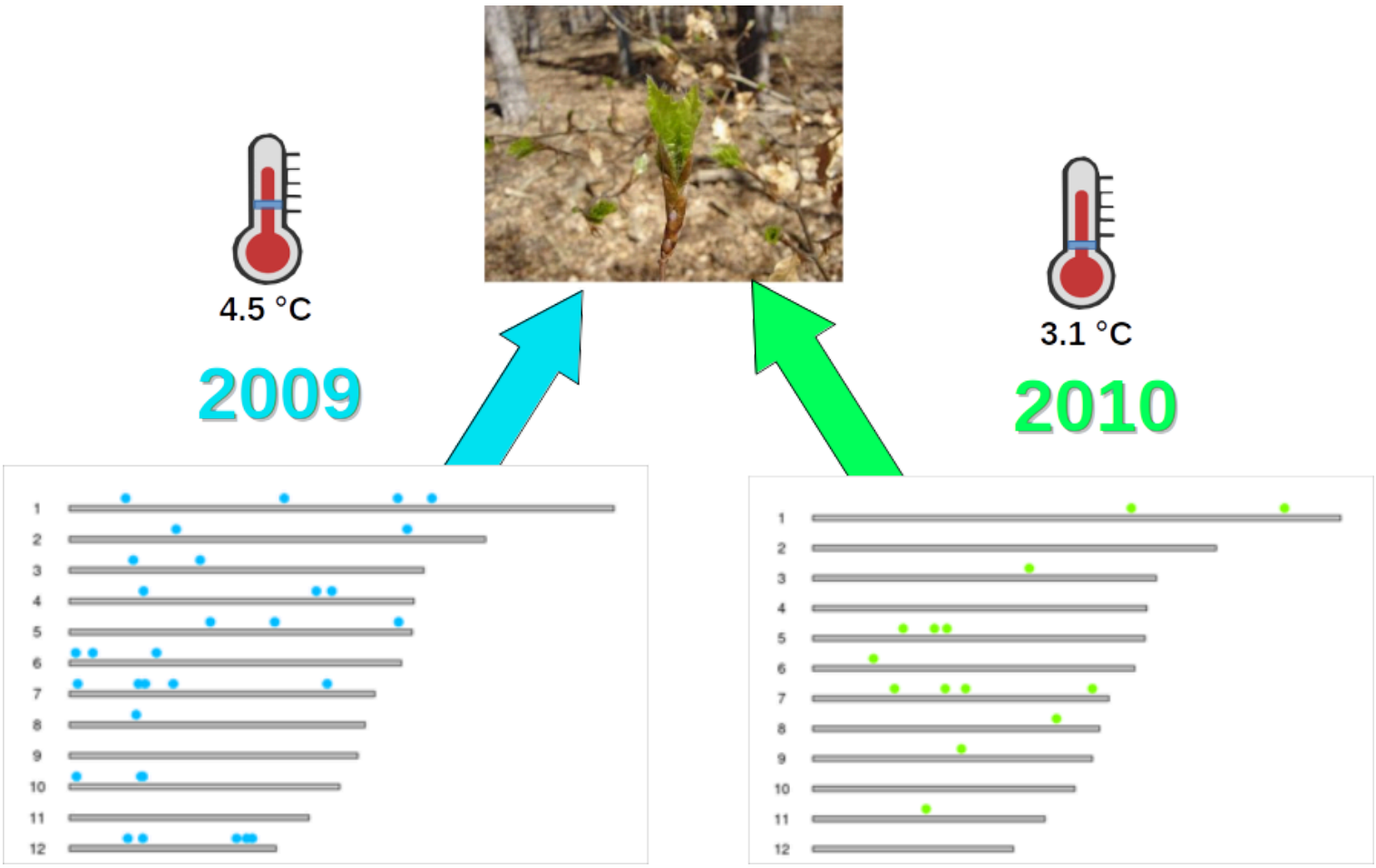


Figure 5: Vegetative spring phenology appears to be controlled by different QTLs in different climatic conditions (the first quarter of 2009 was warmer than the first quarter of 2010 by 1.4 °C), suggesting that different components of its architecture can track different environmental conditions.

6.4 Bud-breaking Against the Gradient: Clinal Phenotypic Differentiation, Anticlinal Genetic Differentiation

The quantification of the common garden observation by Gaüzere, Klein *et al.* (2020) found evidence of a countergradient of genetic differentiation for spring phenology within the historical relict area: trees from stands at lower elevation break the buds later relative to those at higher altitudes. The shared environment had evened out the cues that caused the plastic reaction norm that results in the observed pattern *in silva* (**Figure 6**). The observed counter-gradient is to be framed within the trend of genomic divergence within the dispersal scale between stands in different microhabitats, as highlighted in Chapter 1.

If beech trees follow a countergradient of genetic differentiation at adaptive traits on the spatial scale of the historical relict area (497 m of altitudinal difference), it should be predictable with the breeding values obtained from the allelic effects of genes underlying the architecture of spring phenology. With the genetically determined component of the trait it is possible to estimate what is the complementary effect of the plastic reaction norm to the environmental turnover along the gradient and to infer the adaptive roles of the two components. The expectation was that selection has genetically pushed TBB forward in the year in stands from lower altitudes, that would thus have greater (*ie*. less negative) breeding values, whilst the product of allelic effects and genotypes will be smaller (*ie.* more negative) at higher altitudes. This is what the third chapter accomplished: having identified the QTLs that control spring phenology on the N1 stand, we predicted the differentiation countergradient for several traits associated with spring phenology. Then we simulated a shift in climate that causes the phenotypic optimum to move towards what is currently expected for lower elevations (N2 in this particular example). We found that genetic diversity at the QTLs within the stand was enough for it to adapt to the novel conditions; however, it happened at the cost of a sizable loss of genetic diversity.

Several factors may have biased downward the extent of the unveiled architecture: small sample size, paucity of years of observation, and the investigation of the trait in a single deme (*ie*. the trait has been studied under a subset of the environmental cues that unfolds along the mountain slope) may have constrained the statistical power. Also, a sigmoidal model that better depict the TBB dynamic (*e.g.* a Gompertz curve, which is asymmetric to the midpoint; Winsor, 1932; Tjørve and Tjørve, 2017) could improve the precision of the associations; while cryptic population structure may have generated spurious associations.

Globally, the breeding values of the investigated spring phenology traits are anticlinal: in 20 stand pairings seven traits (chiefly, the dates of BBCH 07 and 09 in 2009) have larger genetic components at lower elevations which decline as one climbs upward. On the opposite, four traits (for a total of five stand pairings) exhibited a co-gradient trend. We counted the phenological scores sum, the breeding values of which increase with elevation, among the traits that support the counter-gradient because the trait value has an inverse relationship with the date of passage - the time series of trees that flush their leaves earlier include higher scores than later trees. The parameters of the logistic curves lack a clear message, possibly because they condense too many physiological processes that may follow opposite differentiation clines at different optima.

Along the altitudinal gradient of this study, temperature is the main environmental variable known to affect phenology. The stand where its genomic bases have been investigated sits at

the bottom of the mountain slope, which warms earlier during spring. As one climbs up the mountain slope in early April, they would pass from trees at a more advanced stage of bud-break to trees in earlier stages, or even closed buds.

Spring phenology is modulated by its genetic component and, through its plastic component, the environment; an individual tree breaks its buds at a time that is a “compromise” between the long term selective regime and how the current season is going. Interpreting the pattern of TBB *in silva* depends on which of the two forces (short-term plasticity or long-term evolutionary change) drive the trade off. This thesis shows that the (few) genomic regions associated with spring phenology are polymorphic along the mountain slope, and produce differential phenotypic values at different elevations. It thus provides evidence that it is the plastic response to drive the trade off on timing of TBB, since the plastic responses of trees at different elevations go in the same directions - “cold, later; warm, earlier”. We hypothesise that at higher elevations selection drives the genetic component of the trait toward an early spring to balance the delaying caused by the response to seasonal cold, that reduces the length of the vegetative season and the onset of the synchronous reproductive phenology (**Figure 7a**). The colder climate pushes the beech trees to flush the leaves later in the year to avoid frost (Dantec *et al.*, 2015). However, a late TBB results in a shorter growing season: though the predicted optimum for TBB is delayed in the year as the elevation increases, the maximum fitness decreases and becomes narrower (**Figure 8**). A tree that starts photosynthesising earlier can fix more carbon and better allocate resources to the co-occurring reproductive phenology (Gaüuzere, Klein *et al.* 2020). The advantage of an early spring is predicted to be large enough in terms of reproductive output to outcompete the damage of late spring frosts, which are more frequent than lower elevations.

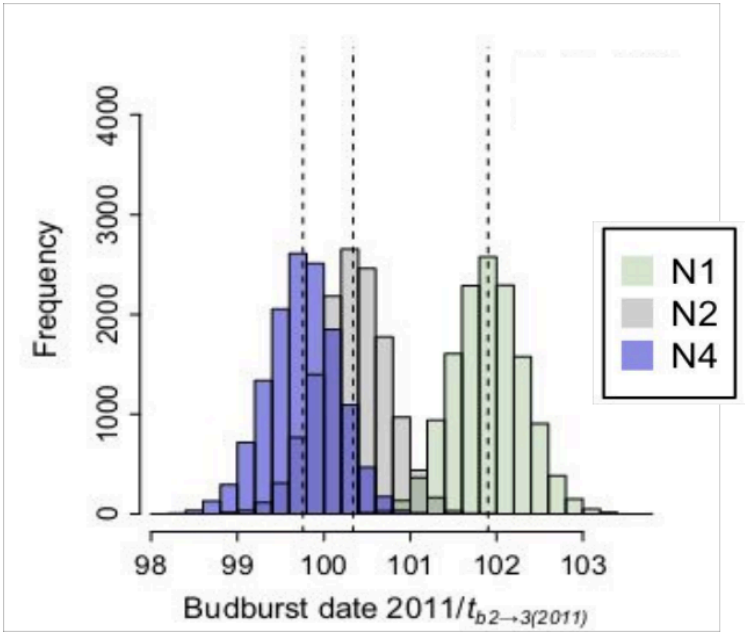


Figure 6: Counter-gradient of TBB observed in common garden: individuals from **N4** (1,363 m a.s.l.) flush their leaves earlier in the year than those from lower altitudes like **N2** (1,1040 m a.s.l.) and **N1** (1,020 m a.s.l.) (Gaüzere, Klein *et al.* 2020).

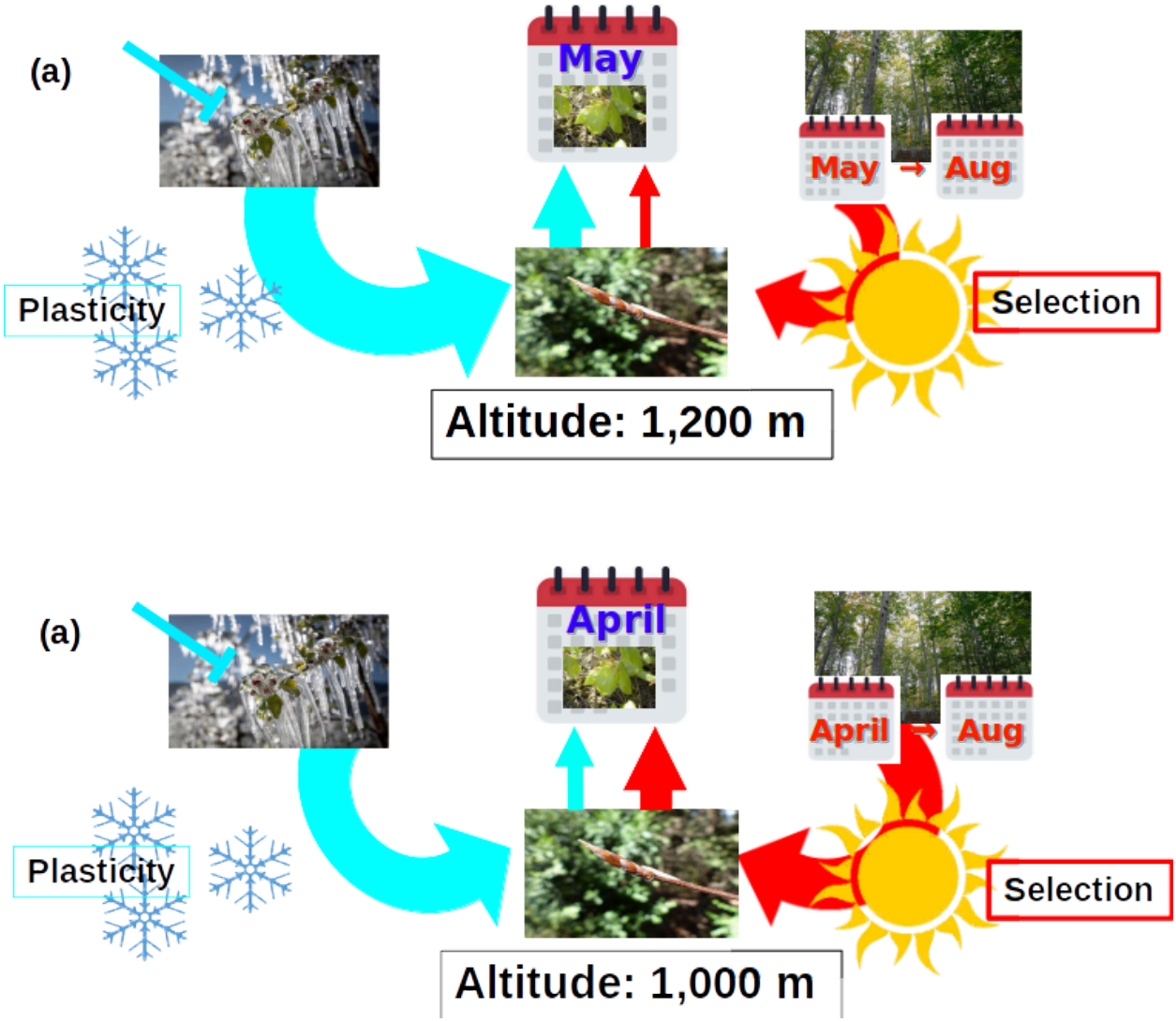


Figure 7: Tradeoff between environmental drivers of TBB at different elevations. At the scale of the mountain, the peak of the fitness optima toward early TBB and selection pushes the trait in that direction. At high elevations (**a**) the plastic response to low temperatures (frost avoidance) pushes the TBB late in the year, and selection is stronger to compensate for the hyperplasticity. At lower elevations (**b**) the warmer conditions result in a less intense pressure toward early dates that is still strong enough to overcome the retarding effect of plasticity on TBB.

Under the pressure to compensate for the shorter growing season, trees evolved toward a higher sensitivity to a positive change in temperature (*ie.* lower heat requirement) so whenever the signal from the photoperiod is coupled with warm enough temperatures, trees initiate the physiological processes required for the bud-break. The tension between a low threshold to initiate bud-break and the consistent reaction to the persistent cold results in a defensively shorter growing season (hyperplasticity, Gaüzere, Teuf *et al*., 2020): only a very warm spring would "unbind" the effect of the genotype. Selective regimes on other traits contribute to the selection for a precocious TBB and stand along the gradient experience different regimes of sexual selection pressures: at low altitude early TBB was maintained by assortative mating, while at high altitude, selection through male fecundity (*ie*. synchronicity between flowers of the two sexes) reinforces the directional selection for earlier TBB (Oddou-Muratorio *et al*. 2024).

At lower elevations the tradeoff that drives TBB is reversed (**Figure 7b**). The pressure to start assimilating carbon earlier relaxes as one climbs down, because the (relatively) milder climate warms earlier and trees quickly fulfill their heat requirements. The genetic component pushes toward a later TBB with a higher threshold of heat accumulation to counterbalance the higher rate of warming that would trigger a precocious spring phenology: these trees will reach fast even such a higher threshold, ensuring the avoidance to late spring frosts to a certain degree.

In estimating the phenotypic optima for TBB along the gradient, Gaüzere, Teuf *et al*. (2020) used the sole spatial variation of temperature as driver of variation in the trait and assumed a consistent plastic response between elevations - that is, stands along the gradient were not genetically differentiated in their plastic reaction and they all responded the same way to the same temperature. At high elevations, the optimum for TBB is earlier than the observed and estimated TBB and directional selection is expected to be strong for an early TBB.

The "anti-plastic", "elongating growing season" selective pressure allows high-elevation stands a weakly positive fitness tradeoff by taking advantage from the shorter growing season and appears to be the key for plasticity to remain adaptive. Other factors outside the length of vegetative season should be considered, as a retarded bud-break enables a water saving-strategy that maximises growth and seed quantity (Bontemps *et al.*, 2017; Oddou-Muratorio *et al.*, 2024). The assumption of a constant plastic response (*ie*. no diversity in its genetic bases at this spatial scale) can be tested by evaluating the rate of change of the phenotype from tree growing at different altitudes by accounting for the genotype.

The forward simulations we performed in Chapter 3 are, first and foremost, a proof of concept for the evolutionary potential of TBB. The detected loss of genetic diversity must thus be interpreted carefully, as the conditions (no neighbouring demes, no migration) differ from those *in silva*. In first place they confirm the ability of trees to adapt to novel environmental pressures; the instantaneous shift of the phenotypic optimum mimics the hockey stick graph of temperature increase by Mann *et al.*, (1999). Under a warmer climate, heat requirement are fulfilled early in the year and prolonging the growing season ceases to be the pressure to advanceTBB; rather, frost avoidance appears to be the driver of adaptation: trees that start photosynthesising (and developing flowers) later are more successful in propagating seeds and pollen. The projected and ongoing higher frequency of extreme events will play an important role in determining the selective regimes and the selected genotypes. Several, non mutually-exclusive factors can explain the drop in genetic diversity.

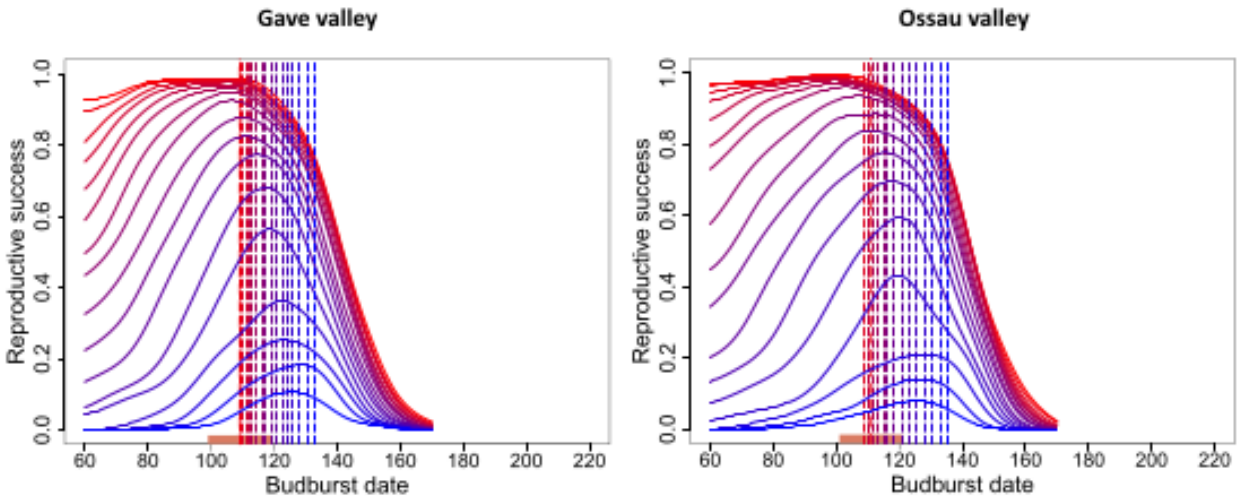


Figure 8: Predicted fitness landscape for TBB at different altitudes in two Pyrenean valleys. Increasing elevation is shown as a spectrum from **blue** to **red**. Peaks are more pronounced, broader and for earlier TBB at lower elevations than higher elevations (Gaüzere, Teuf *et al* 2020).

6.5 Dancing of Architecture

The thesis improves the understanding of adaptation of tree populations (long lived, with large populations and strong gene flow) along environmental gradients. It highlights processes that happen at the fine scales on distinct axes - temporal, spatial, and of genome structure and patterns. It also contributes to the understanding of the architecture of a suite of complex traits that represent the phenological transition from dormancy to bud-break.

65.5.1 Genomic Signatures of Putatively Adaptive Microgeaographic Divergence

After two generations from the start of downward expansion, trees growing in the newly colonised site of low altitude S1 showed signature of selection whenever compared to the stands in the historical relict area. More outliers were detected at the sequence level than to single loci (seven outlier regions *vs* three outlier SNPs). A similar pace has been reported by reciprocal transplants in grasses (Colautti and Barrett, 2013; Hancock, 2016; Magnoli, 2020) and with genetic data in conifers (Scotti *et al.*, 2023). The tests that detected these signals were conservative respect to the simulated distribution of genetic diversity hence the extent of rapid adaptation is likely to be underestimated. However, my findings follow the expectation that few individual loci of large effect contribute to local adaptation (Yeaman, 2022), as S1 was involved in the pairings that found three out of four single-locus outliers; this suggests that the novel environmental pressures were strong enough to impose a regime of strong selection at a few loci. The paucity of the numbers prevent me from stating what the short-term genomic consequences of a novel regime of environmental pressures are. In fact theory predicts that, during the early generations of population divergence, selection should capture advantageous combinations of alleles at multiple loci across the genome between different populations rather than driving differentiation at the level of allele frequencies (Le Corre and Kremer, 2012; Kremer and Le Corre, 2012; Cubry *et al.*, 2022). I did not pursue this path, because evaluating covariances on a panel of 44,000 SNPs is too time consuming and computationally expensive.
Selection can affect traits with a complex architecture only by acting on a small portion of the QTL, resulting in phenotypic divergence over a few generations (Messer *et al.*, 2016). It is not to be excluded that the range expansion has brought home-neutral alleles of intermediate frequency into an ecological context where they turned out to provide an advantage (Burton and Travis, 2008; Gilbert and Whitlock, 2017), so, while the process started as non-adaptive, it eventually became genuinely so. The consistent gene flow from the source population can also refuel the genetic diversity in the recently colonised stands and improve their response to selection (Gaüzere, Klein *et al.*, 2020). These findings indicate that European beech can adapt to novel environmental conditions over short time scales. However, since the high intra-population genetic diversity is the main source of such ability, it must be preserved to ensure the species' long term survival. Even though the preliminary results of the third chapter concern an individual population and rest on the theoretical assumption of no immigration from (possibly preadapted) stands, they show that adapting to a new optimum is paired with a substantial loss of genetic diversity at the QTLs.

Outliers were retrieved among all the stand pairs, indicating that adaptation can be extensive all along the landscape scale. Strong within-population divergence between environmental gradients within the dispersal distance is well reported in literature (European beech: Lalagüe *et al.*, 2013; Csilléry, Lalagüe *et al.*, 2014; Gaüzere, Klein *et al.*, 2020; Pfenninger *et al.*, 2021; *Eperua falcata*: Audigeos *et al.*, 2013; Brousseau *et al.*, 2020; sugar pine: Eckert *et al.*, 2015; stone and mountain pine: Mosca *et al.*, 2016; Alepp and maritime pine, silver fir, Atlantic cedar: Scotti *et al.*, 2022; maritime pine: Budde *et al.*, 2024). On Mont Ventoux, the 150 metres between N3 and N4 suffice for environmental conditions to generate different optima for beech and result in tree stands adapted to a very granular spatial scale. Outside of the three outlier genomic regions, also the QTLs that determine budburst appear to be differentiated at this very scale. Gene flow has been found to sustain divergence rather than constraining it, as Gaüzere, Klein *et al*. (2020) found that pollen from neighbouring stands contributes more to the variance in quantitative traits than local pollen, which is consistent with the preservation of the linkage between loci of small effect, since as the number of recombination events increases concentrated architectures are more favoured than sparse loci (Bürger and Gimelfarb, 2002; Yeaman and Otto, 2011; Yeaman and Whitlock, 2011).

6.5.2 Genomic Architecture of Spring Phenology

Vegetative spring phenology appears to be controlled by a complex network of genes with small effects that are triggered in suites by year- and site-specific conditions. On Mont Ventoux, European beech has become locally adapted to a short-scale turnover of environmental pressures through changes in allele frequencies at relatively few genes (0.01 - 0.19% of the genome from the patterns of divergence; 8×10^{-4} % for traits measured in 2009 and 3×10^{-4} for 2010; 2×10^{-3} and 0.01 % of the genomes determines the countergradient of breeding values on at FDR 5 and 10, respectively, and the 0.015 and 0.06 % in 2010). In the case of the genome-wide pattern of diversification, the process involved subtle shifts at groups of loci with minor individual impact that experience selection together thanks to their tight linkage and fewer individual variants (Barghi *et al.*, 2020; Yeaman, 2022). In the case of spring phenology even such small numbers suffice to match the local optima and produce phenotypic differentiation along the gradient - in this case, constituting an anticline. This thesis strengthens the body of evidence that, for highly polygenic traits with strong link to fitness, just a fraction of their architecture is affected by selection at a local scale, yet generating a marked response to selection. In spite of the small sample size, the power of the GWAS experiment was enough to highlight that some phenological traits follow a distribution of allelic effects coherent with the expectations under the infinitesimal model - few loci with larger impact on the trait and a constellation of small-effect loci. Due to such a large network, it appears that distinct elements of the architecture underlying spring phenology respond to different, year-specific climatic cues (Lazic *et al.*, 2024). My findings suggest that the fraction of causal variants involved in adaptation acts in a coordinated way, with core suites of loci regulating the spring transition and subsets or isolated ones responsible for the individual phases. Interestingly, the bulk of the identified QTLs driving trait variation maps outside coding sequences, suggesting that regulatory variants may be as impactful as protein-changing ones in modulating local adaptive

phenotypic diversity for fitness-related traits (Eckert *et al.*, 2015), supporting the observed correlation between pace of structural protein evolution and levels of genes expression (Gout *et al.*, 2010). Even though the small fraction of the genome involved in adaptation is small, it appears that phenotypic variation in TBB is mainly determined by regulatory variants that modulate the expression of other genes or variants that indirectly influence protein structure by producing protein isoforms.

6.6 As the Heat Goes On

The current trend of increasing temperatures is already affecting spring phenology of many tree species, European beech included. Climate warming affects spring phenology well before spring, changing how ecodormancy triggers. While the onset of endodormancy occurs in response to the shortening of the photoperiod and has a genetically determined length (Chuine, 2000; Battey, 2000; Caffarra and Donnelly, 2011; Lesur, Le Provost, *et al.*, 2015; Lesur, Bechade, *et al.*, 2015), the release of ecodormancy depends on the exposition to different temperatures: a tree needs to spend time in a window of low temperatures (chilling) followed by a continuous exposition to forcing temperatures for enough days (Chuine *et al.*, 1999; Chuine, 2000). When the bud has spent enough time at forcing temperatures and the photoperiod has extended (Gaüzere *et al.*, 2017), it opens and unfolds into mature leaves. The chilling requirement is playing a major role in determining how climate change affects spring phenology. While the onset of spring phenology has hastened at the European scale since 1960 (Menzel and Fabian, 1999; Menzel *et al.*, 2006), its pace is slackening because the reduced chill exposure lowers tree sensitivity to temperature (*ie*. for each degree of warming, the advance of leaf flush is lowering) (Körner and Basler, 2010; Fu *et al.*, 2015), likely for the increased temperature requirements to start the bud-break (Fu *et al.*, 2015). Considering that the timespan of these measurements is within the life expectancy of an individual tree, these phenomena are likely to have affected the plastic component of phenology.

If we assume that plastic responses are consistent in its direction (their genomic bases are not differentiated at a fine spatial scale), we see a congruence between the direction of the phenotypic optimum for TBB, the predicted and the observed TBB, but at high elevations the latter ones stride late in the year from the local optimum (Gaüzere, Klein *et al.* 2020). The clear counter-gradient pattern of the genomic bases of spring phenology both reinforce the assumption of no differentiation for plasticity and evidence a strong selection to constraining environments (different warming requirements between altitudes) and a buffering of the maladaptive reaction norm to microclimatic conditions.

The risk of climate warming is that both these evolutionary pressures (plasticity compensation, longer growing season) can cease to exist: spring advances at a syncopated pace in large part because temperatures are evening along altitudinal gradients; the plastic response pushes trees of high elevation to flush the leaves earlier, because they receive an adequate exposure to chilling. At the same time, trees at lower elevation experience less time in cold temperatures, and the lack of chilling leads to a delayed TBB. What follows is a synchronisation of vegetative *and* reproductive phenology among altitudes (Vitasse *et al.*, 2018), possibly also changing the regimes of directional sexual selection across the gradient (Oddou-Muratorio *et al.* 2024).

As selection appears to become relaxed, the adaptive value of phenology is going to depend on ecological as well as evolutionary factors. While the decoupled flowering between altitudes resulted in trees of higher elevation to disperse seeds mostly in their immediate surroundings, its synchronisation increases the gene flow between stands. The ability of a population to adapt is conditional to its genetic diversity, so the smoothening in the turnover of environmental pressures is expected to reduce the steepness of the phenotypic optima so that smaller and smaller fractions of a trait architecture will diversify, possibly toward a less pronounced countergradient. Indeed, the simulated N4 population described in Chapter 3 of the thesis did

reach the mean breeding value that today is observed at a lower altitude. The positive relationship between canopy duration and fruit production even points toward a larger reproductive output following an earlier bud burst (Journé *et al.*, 2021).
However, even if N4 managed to track the new optimum for TNN, it may not be demographically viable in the new environment (Chevin *et al.*, 2010). Whilst immigrant pollen increases diversity at quantitative traits (Barton, 1999; Le Corre and Kremer, 2003; Gaüzere, Klein *et al.*, 2020), under warmer conditions stands but the highest and lowest one will be under less different conditions. Most of the exchange pollen will come from stands that have been evolving under increasingly more similar environments; the immigration of adaptive alleles at lower elevations will homogenise genetic diversity at QTLs under a more relaxed selection regime, whilst the stands will still be able to exhibit an adaptive response to different inter-annual conditions. Should the phenotypic optima suddenly and abruptly shift because of extreme climatic events (*e.g.Le Corre V, Kremer A (2012b). The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol 21: The genetic differentiation at quantitative trait.Le Corre V, Kremer A (2012b). The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol 21: The genetic differentiation at quantitative trait.* heatwaves, drought, extreme late spring frost) the fewer QTLs that have diversified under the laxer regime risk not being able to match them.
In Chapter 3 the estimates of the breeding values along the northern slope followed a counter-gradient trend only for several phenological traits, and several others did not display differentiation at all. The genomic scale of local adaptation observed throughout this thesis shows that the degree of differentiation is strong but little and likely not global. Coming full circle: genetic differentiation at quantitative traits is constrained by the environmental gradient and the plastic response to the gradient. A reduced variance of the first is matched by the second and lessens the required differentiation; as the environment abruptly changes, the plastic response still moves in the direction of “cold, later; warm, earlier”, but the QTLs risks harbouring too little diversity to match them both. The yearly reproductive output will be affected by sudden frost waves for the trees with a genetic component pushing toward an early spring together with the plastic response - that is, the genetic component has become maladaptive, with possibly snowballing consequences. The ability of trees to match suddenly steeper optima will be conditional on how large is the diversity at the QTLs that have experienced neutral evolution under the less steep environmental gradient.

6.7 Springing Forward

The inferred pattern of adaptations that this thesis explores provides suggestions for evolution-driven forest management practices and paves the road for new research on the adaptation at the scale of the landscape.

Gaüzere, Klein *et al*. (2020) proposed to carry on assisted gene flow campaigns with local sources of seeds rather than relying on geographically distant populations, since the shared ancestral gene pool of the former can boost differentiation at quantitative loci while the latter risk to incur into maladaptive context. By identifying a clear countergradient for the spring phenology, this thesis confirms the advantages of using nearby seed sources, as the introduction of “early spring” alleles from higher elevation can buffer the effects of changing selective regimes.

Several paths can be taken from these results in different research directions.

To better disentangle the genomic bases of local adaptation, the most direct paths goes toward (i) perform multiple simulations to investigate the level of genetic diversity at the QTLs at different optima; (ii) replicating the GWAS in other natural populations to improve the understanding on the architecture of spring phenology, possibly looking for other short-scale gradients; (iii) increase the sample size at the same site I studied, strengthening the statistical power; and (iv) gather information on the differentiation at other traits like reproductive phenology, which is strictly linked to vegetative phenology

The forward simulations introduced in the third chapter will improve the prediction of how stands will respond to novel environmental conditions when moving from an optimum to another: will the stands harbour enough genetic diversity at the QTLs to match the shift? And what will be the outcome in terms of resulting diversity? A more extensive research will require simulating populations that move from distinct optima, so that the entire gradient will be covered; and including neighbouring demes to assess the role of immigration.

I am currently working together with Laurent Clément (PhD student at the University of Perpignan) on a joint GWAS project to identify causal variants associated with spring phenology traits in the two beech populations of Mont Ventoux and La Massane. This study aims at discovering the differences in the trait architecture between the sites: investigating the same trait in two populations with different biogeographic origin (Western Alps glacial refuge for Mont Ventoux, Pyreneean glacial refuge for La Massane; Magri *et al.*, 2006; Magri, 2008) will highlight different facets of the adaptive processes from distinct ancestral pool and more recent signature due to the two local environments (Rellstab *et al.*, 2017). Nonetheless, environmental heterogeneity between studied populations is the first factor that makes the results of Genotype to Phenotype Association analyses in the wild hard to replicate, because the expression of the trait may be affected by distinct sets and intensities of environmental factors. In light of these issues I have modelled the individual-level environmental covariates on the “crude” traits (PSS, JDoY.7 and 9, DegreeDays.7 and 9) and I am going to perform more other GWAS on the residuals of these models. Other variables are still expected to play a role in determining the trait value; gathering them at the individual level may appear a Sysiphean effort, as in example temperature and air moisture are inherent to the stand rather than the tree. However, variables like soil quality and water content can be measured tree by tree, or within a certain radius.

The results of the first two chapters of the thesis are the trampoline for PANGENOMEBEECH, a project that broadens these aims to other adaptive characters in more populations. Exploiting

two common garden experiments in France and Germany, 24 beech trees from different provenances will be sequenced to assemble a pangenome representative of the whole spectrum of gene diversity of the species and thus identify variants that have not been found in the two individuals on which the reference genome Bhaga has been assembled (Mishra *et al.*, 2022). From this reference pangenome, variants differentiated in response to environmental variables will be investigated with pangenome Genotype-per-Environment Analysis (panGEA) on a large sample of provenances from Italy, Germany (natural populations) and France (common garden). Building on the GEA interactions, they will move to identify genes that take part in the determination of different adaptive traits with GWAS experiments, as well as broaden the understanding on how a polygenic architecture and genomic redundancy may sustain GEA signatures, in spite of selection eroding them away by allele frequency changes.

FAGRESCUE is a second project that will take advantage of the second chapter of my thesis and Laurent Clément's one. It will rely on remote-sensing approaches to characterise phenology and perform surveys of drought stress resistance in the forests of Mont Ventoux and La Massane. The genes responsible for the control of these traits will be investigated through GWAS experiments. Based on these results and the finding of Laurent Clément and me, they will simulate physio-genetic processes at the individual scale using PHENOPHYT (phenology) and SUREAU (drought resistance) and *in silico-generated* genotypes to make projections of yearly performances of growth, fecundity, and survival. These results will inform about the genotypes that perform better under different climatic scenarios at different spatial scales, and which, instead will be maladapted.

The interaction of vegetative and reproductive phenology is the subject of REGEMAST. Mastings are non-annual bursts of seed production by anemophilous trees; beech and *Quercus petrea* (Matt.) Liebl. are among the species with such reproductive strategy. The project aims at investigating the causes of the different temporal pattern of masting of these two species (cyclical and synchronised within and sometime between populations in beech; stochastic and population-specific in oak) to provide forest management suggestions. The core hypothesis of REGEMAST is that different tempo of masting among these species stems from distinct prey-predator interactions - an individual species for beech, five for oak. Starting from the study of fruiting dynamics and its impact on insect communities, the project will move toward the physiological, ecological and genetic determinants of matsing patterns between species and the investigation of the mechanisms determining individual and population inter-annual variability of flowering time, and will culminate with the modelisation of masting. My thesis will kick in providing clues on the genetic bases of floral initiation, which is strictly linked and synchronous to the vegetative phenology.

The surprisingly rapid adaptation of S1 and the genetic cline on the old-growth stands on the northern slope led to a further investigation of the pace of adaptation of beech populations in the context of climate change. The genetic variance of quantitative traits (growth, mortality, phenology) will be first investigated within and between 217 provenance of gradients; the project will investigate maladaptation by analysing whether several fitness-related traits can compensate for environmental differences (mainly temperature) between populations. The fate of populations under changing conditions will be then assessed by estimating the necessary evolutionary change to match the new optima. The principle of this approach is similar to the

genomic offset, but framed within quantitative genetics: akin to the third chapter of my thesis, here the focus is the population's mean breeding values of fitness related traits.
Methods and results of this thesis can be applied or tested in other non-model species, with a clear emphasis on other Fagaceae and ecologically similar species. *Quercus petrea* L. appears to display a similar anticline of the genetic component of spring phenology (Vitasse *et al.*, 2009); the phylogenetic proximity may ease the detection of hortologous or gene that underlie the trait.

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol Appl* **1**: 95–111.

Audigeos D, Brousseau L, Traissac S, Scotti-Saintagne C, Scotti I (2013). Molecular divergence in tropical tree populations occupying environmental mosaics. *J Evol Biol* **26**: 529–544.

Barbujani G, Magagni A, Minch E, Cavalli-Sforza L (1997). An apportionment of human DNA diversity. *Proc Natl Acad Sci U S A* **94**: 4516–4519.

Barbujani G, Sokal RR, Oden NL (1995). Indo-European origins: A computer-simulation test of five hypotheses. *Am J Phys Anthropol* **96**: 109–132.

Barghi N, Hermisson J, Schlötterer C (2020). Polygenic adaptation: a unifying framework to understand positive selection. *Nat Rev Genet* **21**: 769–781.

Barton NH (1999). Clines in polygenic traits. *Genet Res* **74**: 223–236.

Battey NH (2000). Aspects of seasonality. *J Exp Bot* **51**: 1769–1780.

Bontemps A, Davi H, Lefèvre F, Rozenberg P, Oddou-Muratorio S (2017). How do functional traits syndromes covary with growth and reproductive performance in a water-stressed population of *Fagus sylvatica* ? *Oikos* **126**: 1472–1483.

Brousseau L, Fine PVA, Dreyer E, Vendramin GG, Scotti I (2021). Genomic and phenotypic divergence unveil microgeographic adaptation in the Amazonian hyperdominant tree *Eperua falcata* Aubl. (Fabaceae). *Mol Ecol* **30**: 1136–1154.

Budde KB, Rellstab C, Heuertz M, Gugerli F, Hanika T, Verdú M, *et al.* (2024). Divergent selection in a Mediterranean pine on local spatial scales. *J Ecol* **112**: 278–290.

Buldrini F, Alessandrini A, Mossetti U, Muzzi E, Pezzi G, Soldano A, *et al.* (2023). Botanical memory: five centuries of floristic changes revealed by a Renaissance herbarium (Ulisse Aldrovandi, 1551–1586). *R Soc Open Sci* **10**: 230866.

Bürger R, Gimelfarb A (2002). Fluctuating environments and the role of mutation in maintaining quantitative genetic variation. *Genet Res* **80**: 31–46.

Burton OJ, Travis JMJ (2008). The Frequency of Fitness Peak Shifts Is Increased at Expanding Range Margins Due to Mutation Surfing. *Genetics* **179**: 941–950.

Caffarra A, Donnelly A (2011). The ecological significance of phenology in four different tree species: Effects of light and temperature on bud burst. *Int J Biometeorol* **55**: 711–721.

Charlesworth B (1994). The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet Res* **63**: 213–227.

Chevin L-M, Lande R, Mace GM (2010). Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory (JG Kingsolver, Ed.). *PLoS Biol* **8**: e1000357.

Chmura DJ, Rożwowski R (2002). Variability of Beech Provenances in Spring and Autumn Phenology. *Silvae Genet* **51**: 2–3.

Chuine I (2000). A unified model for budburst of trees. *J Theor Biol* **207**: 337–347.

Chuine I, Cour P, Rousseau DD (1999). Selecting models to predict the timing of flowering of temperate trees: Implications for tree phenology modelling. *Plant Cell Environ* **22**: 1–13.

Colautti RI, Barrett SCH (2013). Rapid Adaptation to Climate Facilitates Range Expansion of an Invasive Plant. *Science* **342**: 364–366.

Conte L, Dallai D, Modica A, Pezzi G, Rebecchi C, Buldrini F (2020). A littoral grass growing inland: genetic diversity of Puccinellia fasciculata around mud volcanoes in Italy. *Plant Biosyst*.

Csilléry K, Lalagüe H, Vendramin GG, González-Martínez SC, Fady B, Oddou-Muratorio S (2014). Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (Fagus sylvatica) populations. *Mol Ecol* **23**: 4696–4708.

Cubry P, Oddou‐Muratorio S, Scotti I, Lefèvre F (2022). Interactions between microenvironment, selection and genetic architecture drive multiscale adaptation in a simulation experiment. *J Evol Biol*: 1–16.

Currat M, Excoffier L (2005). The effect of the Neolithic expansion on European molecular diversity. *Proc R Soc B Biol Sci* **272**: 679–688.

Dantec CF, Ducasse H, Capdevielle X, Fabreguettes O, Delzon S, Desprez‐Loustau M (2015). Escape of spring frost and disease through phenological variations in oak populations along elevation gradients (P Thrall, Ed.). *J Ecol* **103**: 1044–1056.

Davi H, Gillmann M, Ibanez T, Cailleret M, Bontemps A, Fady B, *et al.* (2011). Diversity of leaf unfolding dynamics among tree species: New insights from a study along an altitudinal gradient. *Agric For Meteorol* **151**: 1504–1513.

De Mita S, Thuillet AC, Gay L, Ahmadi N, Manel S, Ronfort J, *et al.* (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol Ecol* **22**: 1383–1399.

De Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE (2014). Genome scan methods against more complex models: When and how much should we trust them? *Mol Ecol* **23**: 2006–2019.

Derory J, Scotti-Saintagne C, Bertocchi E, Le Dantec L, Graignic N, Jauffres A, *et al.* (2010). Contrasting relations between diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. *Heredity* **105**: 401–411.

Eckert AJ, Maloney PE, Vogler DR, Jensen CE, Mix AD, Neale DB (2015). Local adaptation at fine spatial scales: an example from sugar pine (Pinus lambertiana, Pinaceae). *Tree Genet Genomes* **11**.

Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, *et al.* (2014). Population genomics of Populus trichocarpa identifies signatures of selection and adaptive trait associations. *Nat Genet* **46**: 1089–1096.

Excoffier L, Ray N (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol* **23**: 347–351.

Frachon L, Bartoli C, Carrère S, Bouchez O, Chaubet A, Gautier M, *et al.* (2018). A genomic map of climate adaptation in arabidopsis thaliana at a micro-geographic scale. *Front Plant Sci* **9**: 1–15.

Frachon L, Libourel C, Villoutreix R, Carrère S, Glorieux C, Huard-Chauveau C, *et al.* (2017). Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. *Nat Ecol Evol* **1**: 1551–1561.

Fu YH, Zhao H, Piao S, Peaucelle M, Peng S, Zhou G, *et al.* (2015). Declining global warming effects on the phenology of spring leaf unfolding. *Nature* **526**: 104–107.

Gamba E, Aurelle S, Vergnaud C, Talon B (2023). A Holocene history of forest vegetation on the northern slopes of a Mediterranean mountain : the Mont Ventoux, Southeast France. *Ecologia Mediterranea*, 48(2):53-68.

Gaüzere J, Delzon S, Davi H, Bonhomme M, Garcia de Cortazar-Atauri I, Chuine I (2017). Integrating interactive effects of chilling and photoperiod in phenological process-based models. A case study with two European tree species: Fagus sylvatica and Quercus petraea. *Agric For Meteorol* **244–245**: 9–20.

Gaüzere J, Klein EK, Brendel O, Davi H, Oddou-Muratorio S (2020). Microgeographic adaptation and the effect of pollen flow on the adaptive potential of a temperate tree species. *New Phytol* **227**: 641–653.

Gaüzere J, Klein EK, Oddou-Muratorio S (2013). Ecological determinants of mating system within and between three Fagus sylvatica populations along an elevational gradient. *Mol Ecol* **22**: 5001–15.

Gaüzere J, Lucas C, Ronce O, Davi H, Chiuine I (2019). Sensitivity analysis of tree phenology models reveals increasing sensitivity of their predictions to winter chilling temperature and photoperiod with warming climate.

Gaüzere J, Teuf B, Davi H, Chevin L, Caignard T, Leys B, *et al.* (2020). Where is the optimum? Predicting the variation of selection along climatic gradients and the adaptive value of plasticity. A case study on tree phenology. *Evol Lett* **4**: 109–123.

Gilbert KJ, Whitlock MC (2017). The genetics of adaptation to discrete heterogeneous environments: frequent mutation or large-effect alleles can allow range expansion. *J Evol Biol* **30**: 591–602.

Gömöry D, Paule L (2011). Trade-off between height growth and spring flushing in common beech (Fagus sylvatica L.). *Ann For Sci* **68**: 975–984.

Gompert Z, Buerkle CA (2011). A hierarchical Bayesian model for next-generation population

genomics. *Genetics* **187**: 903–17.

Gout J-F, Kahn D, Duret L, Paramecium Post-Genomics Consortium (2010). The Relationship among Gene Expression, the Evolution of Gene Dosage, and the Rate of Protein Evolution (JK Pritchard, Ed.). *PLoS Genet* **6**: e1000944.

Hancock AM (2016). Rapid adaptation to climate change. *Mol Ecol* **25**: 3525–3526.

Hopkins AD (1920). The Bioclimatic Law. *Mon Weather Rev* **48**: 355–355.

Jobling MA (2001). In the name of the father: surnames and genetics. *Trends Genet* **17**: 353–357.

Journé V, Caignard T, Hacket-Pain A, Bogdziewicz M (2021). Leaf phenology correlates with fruit production in European beech (Fagus sylvatica) and in temperate oaks (Quercus robur and Quercus petraea). *Eur J For Res* **140**: 733–744.

Klopfstein S, Currat M, Excoffier L (2006). The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* **23**: 482–490.

Körner C, Basler D (2010). Phenology Under Global Warming. *Science* **327**: 1461–1462.

Kremer A, Le Corre V (2012). Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity* **108**: 375–385.

Lalagüe H, Csilléry K, Oddou-Muratorio S, Safrana J, de Quattro C, Fady B, *et al.* (2013). Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (Fagus sylvatica L.) population from southeastern France. *Tree Genet Genomes* **10**: 15–26.

Lander TA, Klein EK, Roig A, Oddou-Muratorio S (2020). Weak founder effects but significant spatial genetic imprint of recent contraction and expansion of European beech populations. *Heredity*.

Lander TA, Oddou-Muratorio S, Prouillet-Leplat H, Klein EK (2011). Reconstruction of a beech population bottleneck using archival demographic information and Bayesian analysis of genetic data. *Mol Ecol* **20**: 5182–5196.

Lazic D, Geßner C, Liepe KJ, Lesur-Kupin I, Mader M, Blanc-Jolivet C, *et al.* (2024). Genomic variation of European beech reveals signals of local adaptation despite high levels of phenotypic plasticity. *Nat Commun* **15**: 8553.

Le Corre V, Kremer A (2003). Genetic Variability at Neutral Markers, Quantitative Trait Loci and Trait in a Subdivided Population Under Selection. *Genetics* **164**: 1205–1219.

Le Corre V, Kremer A (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Mol Ecol* **21**: The genetic differentiation at quantitative trait.

Lesur I, Bechade A, Lalanne C, Klopp C, Noirot C, Leplé J-C, *et al.* (2015). A unigene set for European beech (*Fagus sylvatica* L.) and its use to decipher the molecular mechanisms involved in dormancy regulation. *Mol Ecol Resour* **15**: 1192–1204.

Lesur I, Le Provost G, Bento P, Da Silva C, Leplé JC, Murat F, *et al.* (2015). The oak gene expression atlas: Insights into Fagaceae genome evolution and the discovery of genes regulated during bud dormancy release. *BMC Genomics* **16**.

Lotterhos KE, Whitlock MC (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol* **24**.

Luconi E, Boracchi P, Nodari R, Comandatore F, Marano G, Vaglienti F, *et al.* (2023). Spatial and Temporal Analyses of the Event of Death for 1480 in Milan Using the Data Contained in the Sforza's Registers of the Dead. *Int J Environ Res Public Health* **20**: 2783.

Magnoli SM (2020). Rapid adaptation (or not) in restored plant populations. *Evol Appl* **13**: 2030–2037.

Magri D (2008). Patterns of post-glacial spread and the extent of glacial refugia of European beech (Fagus sylvatica). *J Biogeogr* **35**: 450–463.

Magri D, Vendramin GG, Comps B, Dupanloup I, Geburek T, Gömöry D, *et al.* (2006). A new scenario for the Quaternary history of European beech populations: Palaeobotanical evidence and genetic consequences. *New Phytol* **171**: 199–221.

Mann ME, Bradley RS, Hughes MK (1999). Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophys Res Lett* **26**: 759–762.

Martín G, Márquez Y, Mantica F, Duque P, Irimia M (2021). Alternative splicing landscapes in Arabidopsis thaliana across tissues and stress conditions highlight major functional differences with animals. *Genome Biol* **22**: 35.

Mckown AD, Klápště J, Guy RD, Geraldes A, Porth I, Hannemann J, *et al.* (2014). Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa. *New Phytol* **203**: 535–553.

Menzel A, Fabian P (1999). Growing season extended in Europe. *Nature* **397**: 659–659.

Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, *et al.* (2006). European phenological response to climate change matches the warming pattern. *Glob Change Biol* **12**: 1969–1976.

Messer PW, Ellner SP, Hairston NG (2016). Can Population Genetics Adapt to Rapid Evolution? *Trends Genet* **32**: 408–418.

Mishra B, Ulaszewski B, Meger J, Aury J-M, Bodénès C, Lesur-Kupin I, *et al.* (2022). A Chromosome-Level Genome Assembly of the European Beech (Fagus sylvatica) Reveals Anomalies for Organelle DNA Integration, Repeat Content and Distribution of SNPs. *Front Genet* **12**: 691058.

Mosca E, Gugerli F, Eckert AJ, Neale DB (2016). Signatures of natural selection on Pinus cembra and P. mugo along elevational gradients in the Alps. *Tree Genet Genomes* **12**: 9.

Oddou-Muratorio S, Bontemps A, Gaüzere J, Klein EK (2024). Interplay between fecundity,

sexual and growth selection on the spring phenology of European beech (Fagus sylvatica L.). *Peer Community J* **4**: e27.

Oddou-Muratorio S, Gaüzere J, Bontemps A, Rey J-F, Klein EK (2018). Tree, sex and size: Ecological determinants of male vs. female fecundity in three *Fagus sylvatica* stands. *Mol Ecol* **27**: 3131–3145.

Oreskes N (2024). The Dull Edge of Occam's Razor. *Sci Am* **331**: 70–71.

Pfenninger M, Reuss F, Kiebler A, Schönnenbeck P, Caliendo C, Gerber S, *et al.* (2021). Genomic basis for drought resistance in european beech forests threatened by climate change. *eLife* **10**.

Rellstab C, Fischer MC, Zoller S, Graf R, Tedder A, Shimizu KK, *et al.* (2017). Local adaptation (mostly) remains local: Reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. *Heredity* **118**: 193–201.

Robledo R, Corrias L, Bachis V, Puddu N, Mameli A, Vona G, *et al.* (2013). Analysis of a Genetic Isolate: The Case of Carloforte (Italy).

Sagan L (1967). On the Origin of Mitosing Cdls. *J Theor Biol* **14**: 225–274.

Savolainen O, Pyhäjärvi T, Knürr T (2007). Gene Flow and Local Adaptation in Trees. *Annu Rev Ecol Evol Syst* **38**: 595–619.

Scotti I, Lalagüe H, Oddou‐Muratorio S, Scotti‐Saintagne C, Ruiz Daniels R, Grivet D, *et al.* (2023). Common microgeographical selection patterns revealed in four European conifers. *Mol Ecol* **32**: 393–411.

Scotti I, Lalagüe H, Saintagne CS-, Ruiz R, Delphine D, Lefevre F, *et al.* (2022). Common microgeographical selection patterns revealed in four European conifers. *Mol Ecol*: 1–19.

Siol M, Wright SI, Barrett SCH (2010). The population genomics of plant adaptation. *New Phytol* **188**: 313–332.

Slatkin M, Excoffier L (2012). Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. *Genetics* **191**: 171–181.

Teissier Du Cros E, Thiebaut B, Duval H (1988). Variability in beech : budding, height growth and tree form. *Ann Sci For* **45**: 383–398.

Tjørve KMC, Tjørve E (2017). The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family (RMH Merks, Ed.). *PLOS ONE* **12**: e0178691.

Vitasse Y, Delzon S, Bresson CC, Michalet R, Kremer A (2009). Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. *Can J For Res* **39**: 1259–1269.

Vitasse Y, Signarbieux C, Fu YH (2018). Global warming leads to more uniform spring phenology across elevations. *Proc Natl Acad Sci U S A* **115**: 1004–1008.

von Wuehlisch (1995). Variation in temperature sum requirement for flushing of beech provenances. *Silvae Genet* **44**: 343–346.

Winsor CP (1932). The Gompertz Curve as a Growth Curve. *Proc Natl Acad Sci* **18**: 1–8.

Worrall J (1983). Temperature — bud-burst relationships in amabilis and subalpine fir provenance tests replicated at different elevations. *Silvae Genet* **32**: 203–209.

Wright CJ, Smith CWJ, Jiggins CD (2022). Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet* **23**: 697–710.

Yeaman S (2022). Evolution of polygenic traits under global *vs* local adaptation. *Genetics* **220**: iyab134.

Yeaman S, Otto SP (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* **65**: 2123–2129.

Yeaman S, Whitlock MC (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**: 1897–1911.

Zhang C, Yang H, Yang H (2015). Evolutionary Character of Alternative Splicing in Plants. *Bioinforma Biol Insights* **9s1**: BBI.S33716.

Worrall, J. 1983. Temperature — bud-burst relationships in amabilis and subalpine fir provenance tests replicated at different elevations. Silvae Genet. 32(5/6): 203–209.