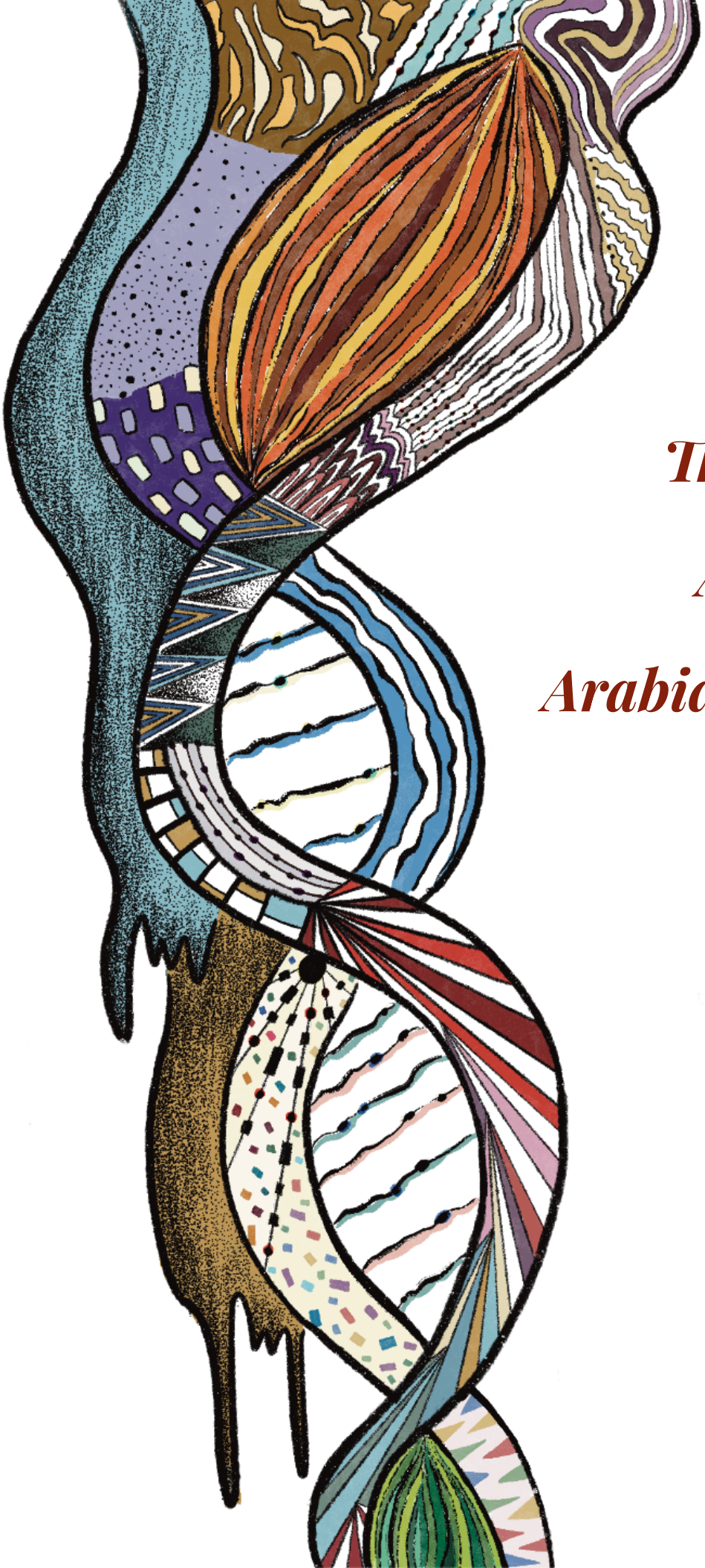


The Genetics of Adaptation in Arabidopsis Species

Loc Thuy Nhu Tran

The Genetics of Adaptation in *Arabidopsis* Species

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Loc Thuy Nhu Tran
aus Vietnam

Köln, Januar 2026

UNIVERSITÄT
ZU KÖLN

BERICHTERSTATTENDE:

Prof. Dr Juliette de Meaux
Prof. Dr. Markus Stetter

Diese Arbeit wurde von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln im Jahr 2026 als Dissertation angenommen.

IMPRINT

The Genetics of Adaptation in *Arabidopsis* Species

All illustrations, including the cover, chapter frontispieces, and internal figures, were created by the author using Procreate and Affinity Designer; unless stated otherwise.

COLOPHON

This thesis uses Zotero for bibliography management. Figures were edited with Affinity Designer. Body text is set in Lora with 1.5 line spacing.

A boat beneath a sunny sky
Lingering onwards dreamily
In an evening of July -

Children three that nestle near,
Eager eye and willing ear,
Pleased a simple tale to hear-

Long has paled that sunny sky;
Echoes fade and memories die;
Autumn frost has slain July.

[…]

In a Wonderland they lie,
Dreaming as the days go by,
Dreaming as the summers die;

Ever drifting down the stream -
Lingering in the golden gleam -
Life, what is it but a dream?

- Lewis Carroll, A Boat Beneath A Sunny Sky

Table of Contents

Abstract

As environmental change accelerates in the Anthropocene, a central challenge in evolutionary biology is understanding how populations respond to novel and rapidly changing conditions. Adaptation underpins whether species can persist and diverge under increasingly variable selective pressures. While adaptive potential is often inferred from phenotypic change or standing genetic variation, it remains unclear what determines the evolutionary "fuel" that enables sustained response. Using *Arabidopsis* species as model systems, this thesis examines the genetic basis of adaptation and how variation is generated and structured across biological scales, from life-history traits to gene expression and genomic interactions, with a particular focus on how genetic architecture shapes the pace and predictability of evolutionary change.

The first study shows that variation in heat-induced secondary dormancy in *Arabidopsis thaliana* - an ecologically critical life-history trait - reflects local adaptation to climatic gradients and mediates population persistence under environmental uncertainty. In the second study, analyses of gene expression in natural populations of *Arabidopsis lyrata* reveal that although heritable expression variation is widespread, the amount of non-additive variance is substantial and shaped by past selection and genomic features. Adaptive potential at the molecular scale is therefore limited not by a lack of variation, but by functional and selective constraints. Finally, analyses of *Arabidopsis lyrata* hybrid populations in the third study demonstrate that genetic background can reshape the components of genetic variance. As hybrid outcomes are often shaped by context-dependent epistasis, the role of admixture in adaptation depends on the type of epistasis and how genetic variance components are reorganised within the genetic architecture.

Together, this thesis conceptualises adaptive potential as an emergent population-level property arising from interactions among ecological traits, genetic architecture, molecular regulation, and environmental context. Adaptive potential depends not on the amount of variation present, but on its structure, heritability, and exposure to selection across evolutionary timescales.

Zusammenfassung

Mit der Beschleunigung von Umweltveränderungen im Anthropozän steht die Evolutionsbiologie vor der zentralen Herausforderung, zu verstehen, wie Populationen auf neuartige und sich rasch verändernde Bedingungen reagieren. Die adaptive Kapazität entscheidet darüber, ob Arten unter zunehmend variablen Selektionsdrücken bestehen oder divergieren können. Obwohl das adaptive Potenzial häufig aus phänotypischen Veränderungen oder stehender genetischer Variation abgeleitet wird, ist bislang unklar, welche evolutionären Faktoren die Maschinerie antreibt, die eine langfristige Anpassungsreaktion ermöglicht. In dieser Dissertation werden *Arabidopsis*-Arten als Modellsysteme verwendet, um die genetischen Grundlagen von Adaptation zu untersuchen. Im Fokus steht dabei, wie genetische Variation auf unterschiedlichen biologischen Ebenen – von Lebenszyklusmerkmalen über Genexpression bis hin zu genomischen Interaktionen – generiert und strukturiert wird. Besonderes Augenmerk gilt der Frage, wie die genetische Architektur das Tempo und die Vorhersagbarkeit evolutionärer Veränderungen beeinflusst.

Die erste Studie zeigt, dass die Variation in der durch Hitze induzierten, sekundären Dormanz in *Arabidopsis thaliana* - ein ökologisch bedeutsames lebensgeschichtliches Merkmal – lokale Anpassung entlang klimatischer Gradienten widerspiegelt und zur Persistenz von Populationen unter Umweltunsicherheit beiträgt. In der zweiten Studie wird anhand von Genexpressionsdaten natürlicher Populationen von *Arabidopsis lyrata* deutlich, dass die erbliche Variation in der Expression weit verbreitet ist. Zugleich zeigt sich, dass ein erheblicher Anteil der Varianz nicht-additiv ist und durch frühere Selektion sowie genomische Eigenschaften geprägt wird. Das adaptive Potenzial auf molekularer Ebene wird somit weniger durch fehlende Variation begrenzt als durch funktionelle und

selektive Einschränkungen. Die dritte Studie untersucht schließlich *Arabidopsis lyrata* Hybridpopulationen und zeigt, dass der genetische Hintergrund die Komponenten der genetischen Varianz tiefgreifend neu gestalten kann. Da hybride Ergebnisse häufig durch kontextabhängige Epistase bestimmt werden, hängt die Rolle von genetischer Admixtur für Anpassungsprozesse entscheidend davon ab, wie die Komponenten der genetischen Varianz innerhalb der genetischen Architektur reorganisiert werden.

Insgesamt konzeptualisiert meine Dissertation das adaptive Potenzial als eine emergente Eigenschaft auf Populationsebene, die aus dem Zusammenspiel ökologischer Merkmale, genetischer Architektur, molekularer Regulation und des Umweltkontexts hervorgeht. Adaptives Potenzial hängt dabei nicht von der bloßen Menge vorhandener Variation ab, sondern von deren Struktur, Erblichkeit und Exposition gegenüber Selektion über evolutionäre Zeitskalen hinweg.

Acknowledgement

Completing this PhD has been a long journey - one filled with moments of discovery and doubt, hope and distress, shared joy and solitude. This thesis is the sum of countless acts of generosity, guidance, and friendship that have carried me here. It would not have been possible without many people, to whom I am deeply grateful.

My thanks to my supervisor - **Juliette de Meaux**, and my thesis advisory committee - **Martin Lercher**, **Holger Schielzeth**, and **Josselin Clo**, for your time, expertise, and valuable feedback which have been essential in shaping my research and the synthesis of this work. I also thank **Markus Stetter** for serving as one of my examiners, **Thomas Wiehe** for serving as Chair of the Examination Committee, and **Tahir Ali** for taking notes during my defence. I want to express my sincere gratitude to **Nadine Rademacher** and **CEPLAS Graduate School** for your administrative help as well as generous financial and career support, to **Martina Reiter** and **Charalampos Mantziaris** for your help in the past years, and to **Katerina Vlantis** for your great kindness and support during the final stage of my Ph.D.

My thanks to **Cansu Aslan**, **Tahir Ali**, and **Gregor Schmitz** for your contributions to *Chapter 2*; to **Margarita Takou** for laying the foundation for *Chapter 3*; to **Kim Steige** for the great help in the population genetic analyses; to **Holger Schielzeth** for teaching me the animal model, and **Josselin Clo** for the significant contribution to *Chapter 4*; to **Yasar Özoglan** and **Thorben Kunisch** for helping with the *Zusammenfassung*; and to **Swan Portalier** for many great discussions. I extend my thanks to the **gardeners** who have cared for and maintained the large amount of plant materials used in this work, to all the **student helpers** who have assisted me throughout the years. I am also grateful to **Markus Stetter and his group** for their continued attention and feedback throughout this journey,

To **Leen Nanchira Abraham** – who taught me far more than just bioinformatics - thank you for teaching me how to think logically, to problem-solve in science, to persevere, and to be a kind person first and foremost. Through the many ups and downs of this journey, your presence has made the road far less lonely. To **Axel Julian Touw**, my first supervisor during my bachelor's study – who set the example for what a mentor could and should be to me. Your humility and scientific mindset have been a compass I carry with me until today.

To **all my colleagues**, to **Swan Portalier**, **Tahir Ali**, **Yasar Özoglan**, and **Renan Granado Chaves**, thank you for making our shared space a home - I could not have asked for better companions during these years. Especially to sweet **Lina Abdelwahed**, thank you for making the hard parts of this journey more bearable through your care and kindness. To **Thorben Kunisch**, thank you for brightening up many of my rainy days and going through everything with me in the last stretch of this journey.

To my best friend, **Do Hong Van** - your presence reminds me that, in sunshine or rain, I am never truly alone. To **all my dear friends** across the world who have always rooted for me, thank you for your quiet support over the past many years.

To **my parents**, without your support and sacrifices, I would not be living my dreams or standing where I am today; especially to **my mother** - like an anchor in rough seas, your presence has kept me grounded and hopeful. Thank you for believing and taking pride in me, for teaching me to hold my integrity and determination. **I dedicate every success that this thesis represents to you.**

Hy vọng con gái có thể làm bố mẹ được tự hào và hạnh phúc hơn một chút.

List of Figures

List of Tables

Publications

Tran, N. L. T., T. Ali, G. Schmitz, and J. de Meaux. 2025. Heat-Induced Secondary Dormancy Contributes to Local Adaptation in *Arabidopsis thaliana*. *Molecular Ecology*. https://doi.org/10.1111/mec.70086.

Touw, A.J., **Tran, N**., Schedl, A., Pajar, J.A., Doan, C.V., Uthe, H., van Dam, N.M. 2025. Root-knot nematode infection enhances the performance of a specialist root herbivore via plant-mediated interactions, *Plant Physiology*. https://doi.org/10.1093/plphys/kiaf109

Author Contributions

Chapter	Study Design	Data Collection	Data Analysis		Manuscript Preparation
2	**LTNT**, GS, JdM	**LTNT**	**LTNT**	Regression analyses, species distribution model, variant calling, GWAS, synthesis	**LTNT**, TA, GS, JdM
3	**LTNT**, MT, JdM, HS, LNA	**LTNT**	**LTNT**	RNAseq analysis from raw data to final data table, variant calling, image analysis, animal model, GOs, gene clustering, eQTL, machine learning, πN/πS,	**LTNT**, MT, LNA, JdM

				Ka/Ks, DFE, synthesis	
4	**LTNT**, MT, JdM, HS, LNA	**LTNT**	**LTNT**	RNAseq analysis from raw data to final data table, variant calling, image analysis, animal model, data integration, synthesis	-
			JC	Mathematical modelling and simulations	

Loc Thuy Nhu Tran (**LTNT**), *Leen Nanchira Abraham* (LNA); *Gregor Schmitz* (GS), *Tahir Ali* (TA), *Juliette de Meaux* (JdM), *Margarita Takou* (MT), *Josselin Clo* (JC), *Holger Schielzeth* (HS).

PART I – *General Introduction*

It may be said that natural selection is daily and hourly scrutinising, throughout the world, every variation, even the slightest; rejecting that which is bad, preserving and adding up all that is good.

- Charles Darwin, "On the Origin of Species" (1859)

1 Concepts, Motivation, and Thesis Aims

1.1. Adaptation and Adaptive Potential

Understanding why organisms vary in their potential to cope with environmental change has been one of the central questions in evolutionary biology. It is through this lens that we can ask one of the most imperative yet unresolved questions in biology: *to what extent can organisms persist in the face of incremental and unpredictable environmental change*? In an era of rapid anthropogenic change, predicting these dynamics has become both scientifically important and societally necessary, informing conservation management, ecological forecasting, and agricultural resilience. The capacity of populations to persist under such pressures depends on their *adaptive potential*, or, in other words, their ability to respond to selection and evolve.

Environmental change, such as fluctuating temperatures or drought, imposes selective pressures on populations (**Figure 1.1**). Yet exposure to selection alone does not guarantee adaptive response. Populations differ in their ability to translate selective pressures into evolutionary change, depending on whether heritable variation in fitness-related traits is available and expressed in ways that selection can act upon. When these conditions are met, selection can favour phenotypes with higher fitness under the new conditions, leading to shifts in phenotypic distributions across generations. In population-genetic terms, such adaptive change corresponds to shifts in allele frequencies driven by selection. Alleles that confer higher fitness tend to increase in frequency over time; however, selection acts on *phenotypic variation* - the observable differences among individuals - rather than directly on genotypes. Different theoretical frameworks have come to bridge

the gap by formalising the relationship between phenotypes and their underlying genetic variation and by predicting evolutionary responses to selection. The development of these concepts, from early evolutionary thought to modern theories, reflects long-standing efforts to understand the conditions under which adaptive evolution is possible.

Figure 1.1. Simplified example of adaptation. A population exhibiting phenotypic variation experiences environmental change, such as fluctuations in climate, that imposes selective pressures. The outcome of adaptation depends on the adaptive potential of the population, which determines its capacity to respond to selection.

1.2. Historical Foundations of Adaptation

The modern view of adaptation has deep historical roots. The first idea of adaptative potential in natural populations traces back to Charles Darwin in his *On the Origin of Species* (**Darwin et al., 1859**). Darwin conceptualised adaptation as the process by which individuals with traits better suited to their environment would have higher survival and reproduction, passing those traits to future generations. His theory was deeply ecological: he coined the term "*economy of nature*" to describe the web of interactions shaping adaptation and grounded it in empirical evidence from biogeography, morphology, and embryology. However, Darwin lacked a formal theory of heredity, therefore, his ideas faced competition from

alternative evolutionary theories. During what has been called the "*eclipse of Darwinism*" (1880s–1930s), mechanisms such as Lamarckian inheritance, orthogenesis, and mutationism were proposed as alternatives to Darwin's natural selection. While these frameworks sought to explain evolutionary change, none offered the same explanatory power.

The rediscovery of Mendelian genetics at the turn of the 20th century set the stage for reconciling heredity with evolution. This reconciliation culminated in the *Modern Synthesis* (also known as *neo-Darwinism*) of the 1930s and 1940s, when figures such as Ronald A. Fisher (**Fisher, 1930**), Sewall Wright (**Wright, 1931**) and John. B. S. Haldane (**Haldane, 1932**) united Mendelian genetics with evolutionary theory. Ronald A. Fisher, in particular, emphasised that adaptation is not natural selection itself, but rather the *movement of a population toward a phenotype that best fits the contemporary environment*. This distinction sharpened the concept of adaptation and laid the groundwork for mathematical frameworks linking genetics with evolutionary dynamics. Together with Wright and Haldane, Fisher established the foundations of population genetics and quantitative genetics, which became central pillars of the *Modern Synthesis*. Their contributions formalised the study of allele frequency change and firmly re-established adaptation as a central principle of biology. From the *Modern Synthesis* onward, two complementary traditions crystallised (**Figure 1.2**).

Population genetics focuses on changes in allele frequencies driven by mutation, migration, drift, recombination, and selection, describing the genetic composition of populations and formalising adaptation as directional allele frequency change. *Quantitative genetics* provides statistical tools for analysing continuous phenotypic variation and predicting evolutionary responses by linking selection gradients to genetic variances and covariances. *Molecular genetics* then adds mechanistic depth by revealing how genes, physiology, and developmental pathways generate

phenotypic variation. *Evolutionary ecology* complements these perspectives by embedding genetic and phenotypic variation within ecological contexts, examining how interactions with climate, resources, competitors, pathogens, and mutualists shape selection across space and time.

Each of these fields illuminates distinct aspects of the adaptive process, yet their historical separation has limited our understanding of how adaptation unfolds across molecular, phenotypic, and ecological scales. Today, adaptation is recognised as a unifying concept that spans genetics, ecology, and evolution. Researchers now integrate ecological insight with genomic and phenotypic data to understand how organisms adapt through changes in morphology, physiology, behaviour, and gene regulation. Advances in genomic technologies now enable unprecedented resolution, linking molecular variation to quantitative trait architecture and situating both within ecological and evolutionary contexts.

The intersecting frameworks of population genetics, quantitative genetics, genomics, and evolutionary ecology together provide a more complete view of adaptation than any single approach alone. This thesis is situated at this intersection, aiming to connect phenotypic variation, genetic architecture, ecological interactions, and evolutionary outcomes. In doing so, it highlights not only well-studied components of adaptation, but also neglected dimensions, complex interactions, and cryptic sources of variation that may become relevant under novel environmental conditions.

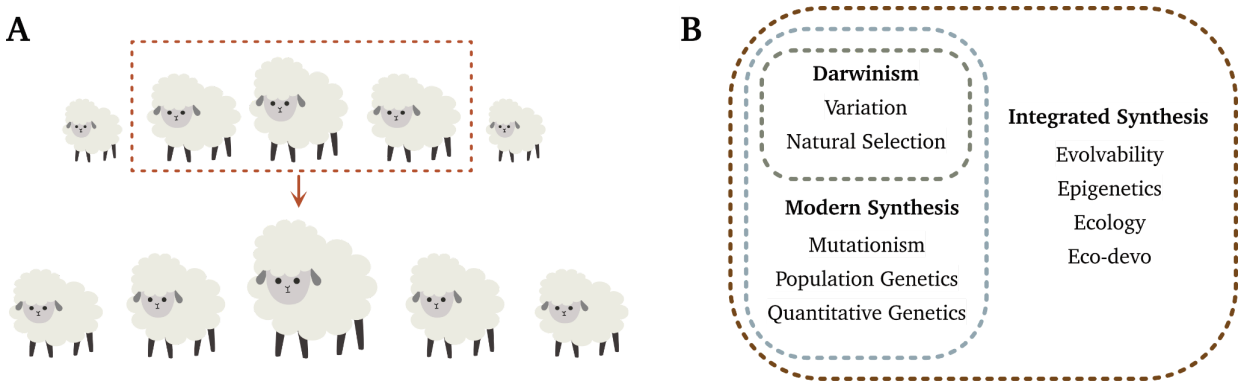


Figure 1.2. Foundations of adaptation. (A) Response to selection. A simplified illustration of directional selection acting on a quantitative trait. When individuals with the largest body size are preferentially selected as parents (highlighted by the dashed box), the mean body size of the population increases in the next generation—provided that the trait has sufficient heritable genetic variation. **(B) Integrated Synthesis.** Conceptual overview of the major intellectual frameworks that have shaped our understanding of adaptation. Darwinism introduced the principles of heritable variation and natural selection. The Modern Synthesis later unified Mendelian genetics with evolutionary theory through the development of population genetics, mutationism, and quantitative genetics. The emerging Integrated Synthesis builds on these foundations by incorporating new dimensions, including evolvability, epigenetics, ecological context, and eco-evolutionary and developmental processes, highlighting that adaptation arises from interactions across multiple biological scales.

1.3. Levels of Biological Variation

1.3.1. Genetic variants and genetic variation

Building on the integrative perspective outlined above, it is necessary to define more precisely the concepts that underpin the genetic basis of adaptation. *Genetic variants* are the most fundamental unit of variation, referring to specific alterations in the DNA sequence of a gene. These changes can be as small as a single nucleotide

polymorphism (SNP), or as large as insertions and deletions (INDELs) or duplications. SNPs are the most abundant genetic variation in eukaryotes and a key tool in plant genetics. Their high density and stability make them powerful markers for studying diversity, population structure, quantitative trait loci (QTL), and adaptation (**Brookes, 1999; Mammadov et al., 2012**). Variants may be inherited, arise *de novo* during gametogenesis, or occur somatically, and they represent the altered forms of DNA sequence that can influence traits or contribute to disease susceptibility. The collective presence of these variants across individuals constitutes *genetic variation*, a broader concept referring to the total diversity of DNA sequences within a population or species. Genetic variation provides the raw material for natural selection and evolutionary change, arising from processes such as mutation, recombination, fertilisation, and gene flow. Yet, genetic variation itself is not visible to selection. Its evolutionary relevance emerges when it is expressed as *phenotypic variation*.

1.3.2. From genotype to phenotype

Phenotypic variation arises through an organism's development and physiology, and natural selection "sees" this variation within its environmental context. When phenotypic differences are underpinned by genetic variation, selection can shape the genetic variation, which then contributes to heritable phenotypic variation, eventually leading to evolutionary change (**Figure 1.3**). Population-level processes such as mutation, genetic drift, and migration shape the reservoir of genetic diversity that provides the raw material for selection. In other words, the evolutionary trajectory is continually moulded by developmental processes, population dynamics, and ecological context. Together, these factors generate and shape the variation that natural selection acts upon, ultimately determining how populations can adapt over time.

Figure 1.3. Conceptual diagram of how evolutionary forces generate and shape variation. **(A) Variation-selection-outcome pathway.** Illustration of the basic logic of evolution: populations begin with phenotypic variation (differences in shape, size, or type), selection filters individuals according to their fitness in a given environment, and the resulting outcome reflects the subset of variants that persist and propagate. Only variation that is heritable contributes to evolutionary change. **(B) Interactions between evolutionary forces.** Genetic variation gives rise to phenotypic variation, which selection acts upon in an environment-dependent manner. Selection can reduce variation by removing deleterious alleles or shape it by favouring beneficial combinations, thereby influencing the evolutionary trajectory of populations. Mutation introduces new genetic variation, while genetic drift may reduce it through stochastic loss. Together, these forces determine how adaptive potential is maintained and transformed through time.

1.3.3. Natural variation and standing genetic variation

At a broader level, *natural variation* encompasses all observable differences among individuals and populations, spanning morphology, physiology, life history, and phenology (**Fusco & Minelli, 2010**). While much of this variation has a genetic basis, it also reflects environmental influences, developmental noise, and epigenetic factors, making it a more inclusive term than genetic variation. Most natural variation is *polygenic*, determined by polymorphisms across multiple loci (**Alonso-Blanco et al., 2009**), although single-gene variants can also play a role (**Salomon et al., 2009**). Importantly, natural variation is typically structured both within and among populations. This *population structure* arises through geographic isolation, barriers to gene flow, non-random mating, selection, or demographic history, producing systematic differences in allele frequencies across

subpopulations. As a result, populations are not panmictic but instead show varying degrees of relatedness and heterozygosity. Population structure is a central consideration in modern genomic analyses, since uncorrected allele frequency differences can confound association studies, producing spurious signals of adaptation or trait linkage. Within the broader pool of natural and genetic variation, *standing genetic variation* refers specifically to the allelic diversity already present in a population at a given time, prior to the onset of new selective pressures (**Barrett & Schluter, 2008**). Unlike novel mutations, which arise sporadically and may take time to contribute to adaptation, standing variation provides an immediate substrate for selection. Populations with greater levels of standing genetic variation can adapt more quickly and effectively, as they can draw upon alleles already segregating at low frequencies (**Chung et al., 2023; Hou et al., 2025; Lai et al., 2019**). This reservoir of diversity is therefore critical for evolutionary resilience, enabling rapid responses to environmental change or biotic challenges. Conversely, populations with reduced standing variation are more vulnerable to ecological stress and extinction (**Matuszewski et al., 2015**).

1.4. Fitness and Quantitative Traits

The study of genetic variation gains its evolutionary significance only when linked to *fitness*. *Fitness* is a central concept in evolutionary biology, in the simplest term, it describes an organism's success in surviving and reproducing relative to others in a population (**Orr, 2009**). In plants, key *fitness components* include the ability to withstand environmental stress and survive to reproductive maturity, fecundity as reflected in the number of seeds produced, and growth traits such as biomass and size, which affect both survival and reproductive output. *Phenology* is also a crucial determinant of fitness, as the timing of life events such as germination and flowering strongly shapes reproductive success under variable environmental conditions (**Anderson, 2016**). New approaches such as high-throughput

phenotyping, gene expression profiling, and machine learning, now provide unprecedented opportunities to link genetic variability to these fitness-related traits (**Henry & Stinchcombe, 2025**).

Many of the traits underlying fitness are *quantitative traits* - traits that vary continuously across a range of values in a population, and are typically influenced by multiple genes (*polygenic*) - each contributing a small additive effect to the overall phenotype - as well as environmental factors (**Caballero, 2020; Falconer & Mackay, 2009**). Because of this multilayered genetic and environmental influence, these traits usually follow a normal distribution and have complex inheritance. Unlike qualitative traits, which have distinct categories (e.g., flower color or pea seed shape), quantitative traits show continuous variation, such as flowering time, plant height, seed number in plants (**Figure 1.4**). Importantly, quantitative traits and fitness traits often show genetic correlations: for example, larger plant size may be genetically correlated with higher seed output, thereby enhancing reproductive success. Such correlations can arise through pleiotropy, genetic linkage, or shared physiological pathways, and they can themselves be targets of natural selection (**Dwivedi et al., 2021**). Genomic approaches, including QTL mapping and association studies, have been able to identify specific genomic regions contributing to quantitative variation and its fitness consequences (**Mauricio, 2001**). Because of the polygenic nature, quantitative traits often embody a vast pool of standing genetic variation that serves as raw material for adaptation, and form the bridge between molecular variation and evolutionary outcomes.

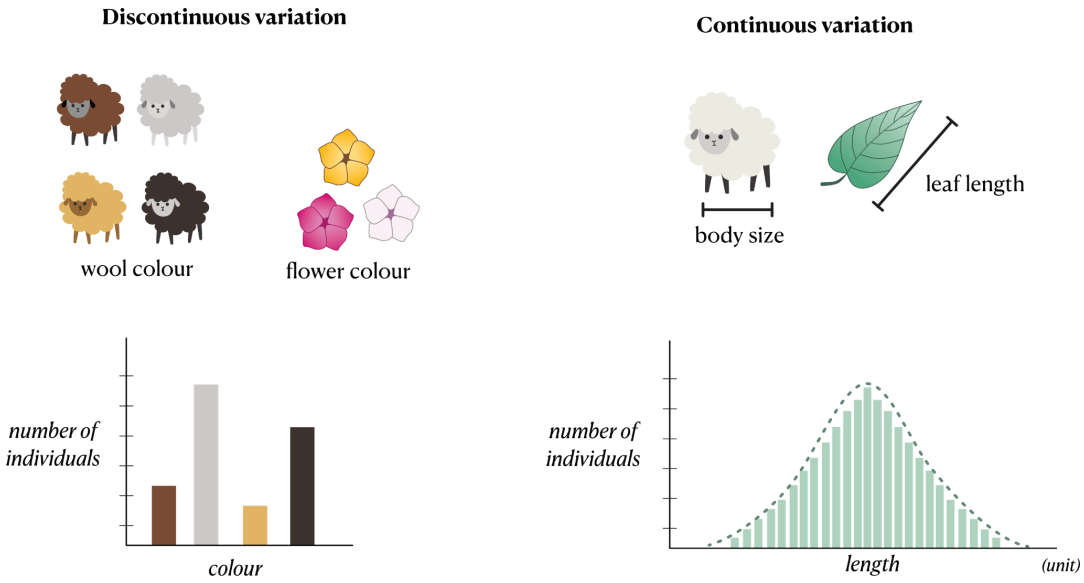


Figure 1.4. Examples of discontinuous and continuous variation. Discontinuous traits, such as wool colour in sheep or flower colour, fall into discrete categories with no intermediates, producing distributions with separate classes. Continuous traits, such as sheep body size or leaf length, vary gradually across a population, typically following a distribution.

In general, traits can be broadly categorized by how closely they are genetically correlated with fitness **(Figure 1.5)**. Traits related to growth rate, seed development, and stress tolerance often have very high correlations and are shaped by directional selection, since increases in these traits usually enhance reproductive success. Hypothetically, other traits such as flowering time and self-incompatibility might be under stabilising selection, where intermediate values are favoured and both extremes reduce fitness. Finally, traits that show little or no correlation with fitness do not usually undergo selection and instead provide information about demographic history and genetic drift.

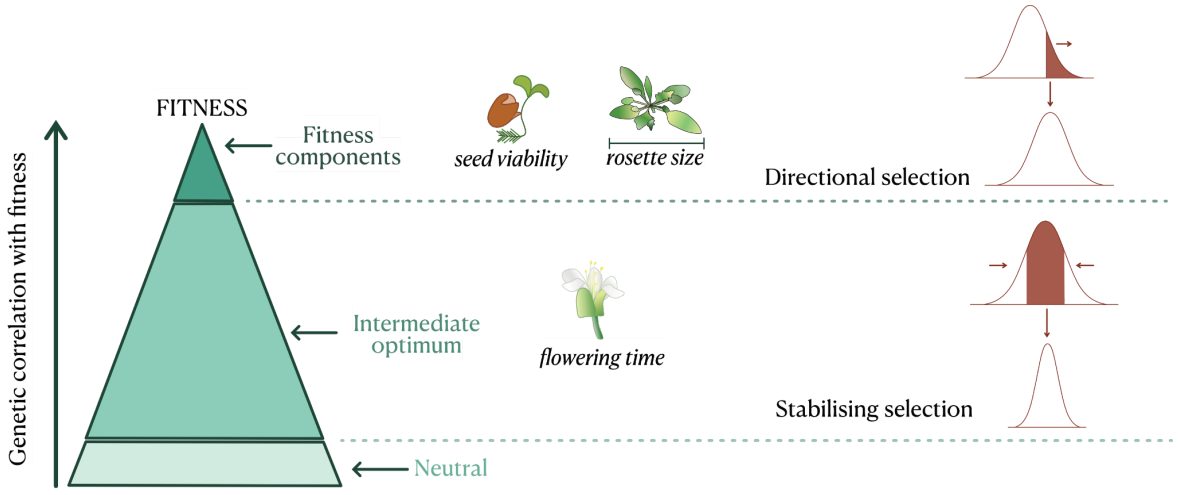


Figure 1.5. Genetic correlation with fitness and modes of selection. Traits closely associated with fitness components (e.g., seed viability, rosette size) are often subject to directional selection, shifting trait means across generations. Traits with intermediate optima (e.g., self-incompatibility, flowering time) are typically influenced by stabilizing selection, reducing phenotypic variance around the intermediate optimum. Traits with no detectable effect on fitness evolve neutrally. Adapted from **Caballero (2020)**, Figures 10.1 and 10.2.

1.5. The Model Systems: *Arabidopsis thaliana* and *Arabidopsis lyrata*

1.5.1. *Arabidopsis thaliana*

Arabidopsis thaliana is the most classic model species in plant biology due to its rapid life cycle, compact genome (~125 Mb), and predominantly selfing mating system (**Krämer, 2015; Meinke et al., 1998**). Extensive genomic resources, natural accessions spanning diverse environments, and well-characterised life-history traits make it ideal for studying natural variation and local adaptation (**Woodward & Bartel, 2018**). The substantial standing genetic variation in A. *thaliana* has enabled

fine-scale mapping of the genetic basis of phenotypes such as flowering time, growth rate, and seed dormancy (**Alonso-Blanco & Koornneef, 2000**).

1.5.2. *Arabidopsis lyrata*

A. *lyrata* (**Figure 1.6**) is a close relative of A. *thaliana*, but differs markedly in ecology, life history, and mating system. It is an obligate outcrosser due to a functional self-incompatibility system, resulting in high heterozygosity and pervasive segregating variation (**Jonsell et al., 1995; Kärkkäinen et al., 1999; Kusaba et al., 2001; Li et al., 2023; Schierup, 2004**). With a larger genome (~207 Mb) rich in structural diversity (transposable element content and structural variation), A. *lyrata* is particularly suited for quantitative genetic analyses that require controlled crosses and decomposing variance components (**de la Chaux et al., 2012; Hu et al., 2011**). Its perennial life history and adaptation to colder, circumpolar habitats also provide ecological contrast to A. *thaliana* (**Chang et al., 2016**).

These properties have made A.*lyrata* well suited for many population genetics studies (**Clauss & Mitchell-Olds, 2006; Muller et al., 2008; Van Treuren et al., 1997; Wright et al., 2003),** but also quantitative genetic analyses, as they require generating breeding designs and rely on segregating variation at multiple loci to dissect the genetic architecture of complex traits. Moreover, its close relationship with A. *thaliana* (ca. 5 MYA, (**Koch et al., 2000**)) allows the rich genomic and functional resources developed in A. *thaliana* to be leveraged in A. *lyrata*, further enhancing its utility as a study system.

This thesis leverages both systems: A. *thaliana* for studying natural phenotypic variation and climatic adaptation (**Chapter 2**), and A. *lyrata* for dissecting the structure of genetic variance, including additive, dominance, and epistatic components, in natural and hybrid populations (**Chapter 3-4**).

Figure 1.6. *Arabidopsis lyrata* in its natural habitat. Photographed at Pfaffenhofen, Plech (49.62°N, 11.51°E) on 22nd June 2023 by Dr. Tahir Ali.

1.6. Linking Genetic and Molecular Perspectives

Forecasting evolutionary responses to environmental change requires understanding how variation at different biological scales contributes to adaptation. Natural populations exhibit phenotypic divergence shaped by past environments, often forming adaptive clines across geography and climate. Classic studies showed concordance between environmental gradients and life-history traits such as flowering time or dormancy (**Gregor, 1938; Turesson, 1930**), and more recent analyses demonstrate that such clines can persist over millennia, reflecting responses to historical climate cycles and postglacial range expansion (**Perrier et al., 2025; Urquhart-Cronish et al., 2025**). These patterns reveal how past selection has shaped present-day phenotypic diversity, especially for traits governed by complex genetic architectures and strong environmental interactions.

Seed dormancy in *Arabidopsis thaliana* exemplifies such a life-history trait. Dormancy integrates ecological pressures, developmental pathways, and genotype-environment interactions, and its natural variation has been repeatedly linked to climatic adaptation (**Debieu et al., 2013; Fournier-Level et al., 2011**).

Because dormancy mediates the timing of germination – an important fitness-component - it provides insight into how populations historically responded to environmental variation, and how these responses may predict their capacity to cope with ongoing climatic shifts. Studying natural dormancy variation therefore connects past adaptive trajectories with the future evolutionary potential of populations.

However, understanding how organisms will respond to rapid and multivariate environmental change requires going beyond life-history traits. Phenotypes like dormancy are the outcome of many upstream molecular processes. Only few individual molecular changes have predictable impacts on adaptive responses, and it is often the amount and structure of standing genetic variation, not single loci, that determine the potential for evolution under novel climates (**Hancock et al., 2025**). Population genetic models show that adaptation may proceed via new mutations (**Smith & Haigh, 1974**), standing variation (**Hermisson & Pennings, 2005**), or polygenic shifts (**Barghi et al., 2020; Pritchard et al., 2010**). Yet these models typically treat traits as single axes, while real adaptive challenges, such as climate change, are multivariate, with selection acting on suites of interacting phenotypes (**Etterson & Shaw, 2001**).

Transcriptional traits offer a complementary perspective: they operate on shorter timescales, reflect immediate environmental responses, and constitute molecular quantitative traits with complex, polygenic bases. Regulatory variation can generate rapid phenotypic shifts and has been implicated in climatic adaptation across taxa, often through environmentally robust cis-regulatory changes **(Ballinger et al., 2023**). Moreover, gene expression traits capture the influence of pleiotropy and epistasis, which theoretical work and empirical data reveal as critical determinants of adaptive potential and evolutionary constraints (**Frachon et al., 2017; Østman et al., 2012**).

By quantifying the genetic variance components, namely, additive, dominance, and especially epistatic variance, underlying transcriptional traits in *Arabidopsis lyrata*, I addressed how molecular architecture shapes short-term evolutionary responsiveness. This is essential as the capacity of populations to adapt depends not only on standing genetic variation but also on how that variation is partitioned across genetic effects and how hybridisation, recombination, and gene flow restructures this architecture.

Broadly speaking, adaptive evolution links three levels: sequence-level variation (mutations, standing polymorphisms), molecular traits (gene expression, regulatory architecture), and phenotypic traits (dormancy, phenology, life-history strategies). By combining ecological, phenotypic, and molecular analyses across two *Arabidopsis* species, this thesis bridges these levels to reveal how genetic diversity translates into adaptive potential, across timescales ranging from historical climatic adaptation to rapid, contemporary evolutionary change.

1.7. Thesis Structure and Aims

The overarching goal of this thesis is to understand the role of natural variation and the evolutionary factors that determine the capacity of populations to adapt and evolve. To address this, I integrate modern genomic, transcriptomic, and phenotypic data with classical frameworks from quantitative genetics and population genetics. The entire work is structured into four main parts (**Figure 1.7**).

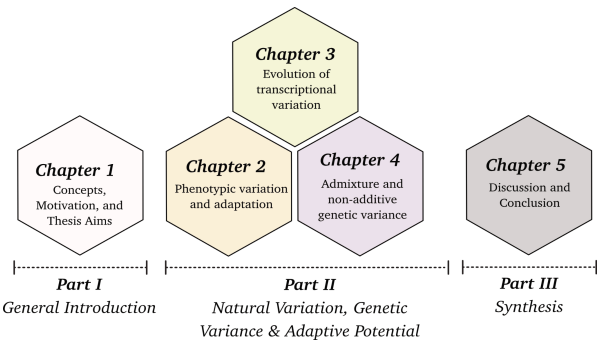


Figure 1.7. Summary of the thesis structure.

Part I (**Chapter 1**) established the conceptual foundations, introducing key definitions, historical context and the biological systems necessary to place the studies in context.

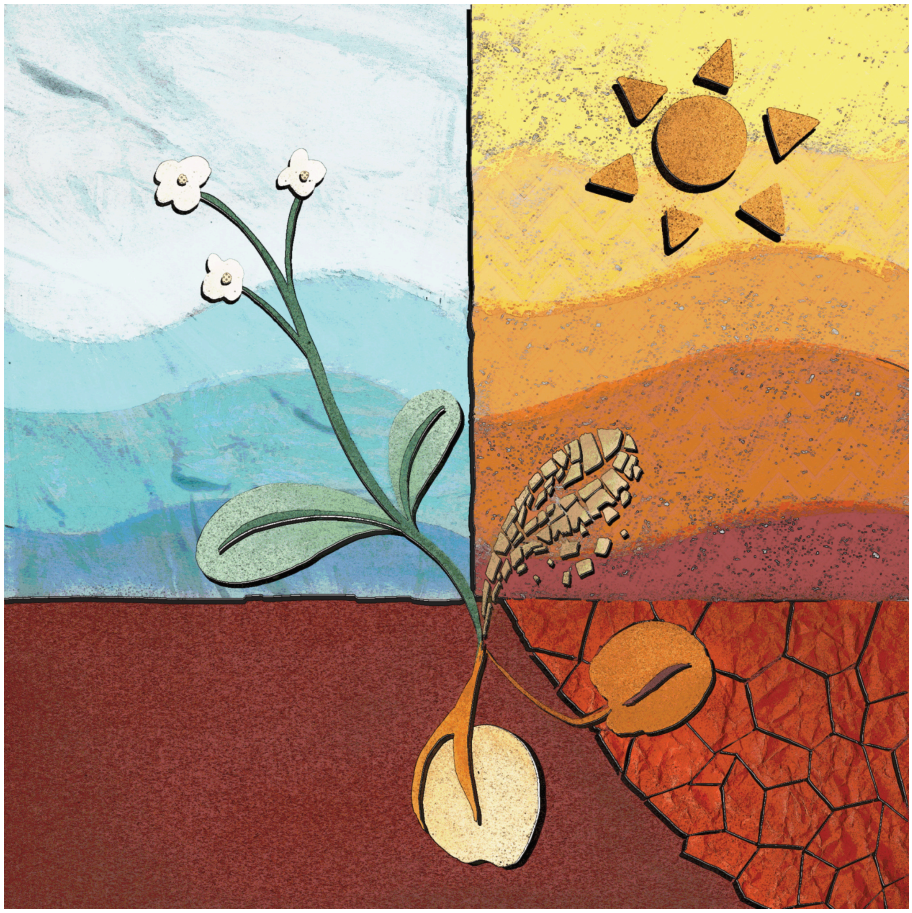
The main part begins in **Part II** (**Chapter 2-4**) which presents the empirical and theoretical studies: **Chapter 2**, at the most accessible level of adaptive potential: *phenotypic variation*. I examined phenotypic and genetic variation in seed dormancy across European A. *thaliana* accessions, testing whether heat-induced secondary dormancy contributes to local adaptation. This study demonstrates how phenotypic diversity, grounded in natural genetic variation, can contribute to adaptive potential. Large parts of this chapter are adapted from **Tran et al. (2025)**. Building on this, **Chapter 3** moves beyond phenotypes to the theoretical foundations of adaptation, where the focus shifts to the structure of genetic variance. Here I investigated the evolution and heritability of transcriptional traits in natural populations of A.*lyrata*, and connecting the past and future evolutionary trajectories. **Chapter 4** then extends this analysis to experimentally created hybrid populations of A.*lyrata* spanning two generations. Here I aimed to quantify how admixture reshapes genetic variance and to estimate the contribution of non-additive components, including dominance and epistasis, across generations.

Finally, **Part III** (**Chapter 5**) synthesises these findings, linking patterns of natural variation, molecular architecture, and genetic variance structure to a general framework of adaptive potential.

Together, these studies reveal how phenotypic diversity, standing genetic variation, and the often-overlooked components of genetic architecture interact to determine the evolutionary capacity of populations. By integrating perspectives from quantitative genetics, population genomics, and evolutionary ecology, the thesis provides new insights into the mechanisms that enable or constrain

adaptation in changing environments: from traits to theory, from patterns to mechanisms, and eventually from specific populations to general principles of evolutionary capacity.

PART II - *Natural Variation, Genetic Variance & Adaptive Potential*



Nature has introduced great variety into the landscape, but man has displayed a passion for simplifying it. Thus he undoes the built-in checks and balances by which nature holds the species within bounds.

- Rachel Carson, "Silent Spring" (1962)

2

Natural Variation in Seed Dormancy Contributes to Local Adaptation in *Arabidopsis thaliana*

2.1. Chapter Summary

This study characterises the natural variation of heat-induced secondary dormancy in A. *thaliana* and its role in local adaptation using 361 *Arabidopsis thaliana* accessions from across Europe. The results show how secondary dormancy varies across populations and co-varies with climatic variables such as temperature and precipitation, indicating its ecological relevance. Species distribution models revealed that genotypes with high secondary dormancy would show greater resilience to future climate changes. Moreover, secondary dormancy appears to be part of a broader dormancy trait syndrome, aligning germination timing with environmental conditions before and after dispersal. By linking natural variation in secondary dormancy to both local adaptation and potential future resilience, this work highlights its importance as a key axis of adaptive potential in A. *thaliana*.

2.2. Introduction

2.2.1. Germination timing and seed dormancy

Natural variation in life-history traits plays a central role in how plants persist and adapt across heterogeneous environments. Among these traits, the timing of seed germination is particularly critical, as it marks the first life-history transition and determines the conditions under which seedling establishment, growth, and

reproduction occur (**Baskin & Baskin, 1997**). Variation in germination timing is largely regulated by seed dormancy, a mechanism that prevents germination under unfavorable conditions (**Chahtane et al., 2016; Lamont & Pausas, 2023**). This natural variation in dormancy enables populations to synchronise life cycles with local environmental regimes, and thereby contributes to fitness differences within and among populations (**Burghardt et al., 2015; Donohue et al., 2010; Wagmann et al., 2012**). While environmental fluctuations such as shifts in rainfall or temperature can disrupt these finely tuned responses, they also provide the selective context in which variation in dormancy and germination timing evolves (**Blackman, 2017; Hamann et al., 2021**).

Dormancy can vary between populations of a species, as well as among different species (**Willis et al., 2014**). Two main types of dormancy exist: primary and secondary (**Figure 2.1**). Primary dormancy is established while the seed develops on the mother plant; secondary dormancy is established after seed dispersal, upon seed exposure to unfavorable conditions (**Baskin & Baskin, 1997; Soltani et al., 2019**). This capacity to respond dynamically to environmental cues provides additional flexibility beyond primary dormancy and can be particularly advantageous in unpredictable climates.

Variation in primary dormancy has been documented extensively in the model plant *A. thaliana* (**Bentsink et al., 2006; Debieu et al., 2013; Footitt et al., 2011, 2015; Martel et al., 2018**), but also within other species growing in a wide range of climatic zones, such as in Korean milkweeds, *Asclepias* (**Kaye et al., 2018**) or among species of the *Fabaceae* family (**Wyse & Dickie, 2018**). In *A. thaliana*, genetic variation in primary dormancy contributes to local adaptation (**Chiang et al., 2013; Kerdaffrec & Nordborg, 2017; Kronholm et al., 2012**). The cline of genetic variation for seed dormancy observed in *A. thaliana* follows the pattern observed across plant species: higher levels of primary dormancy are observed in regions where the

growing season is long, but summer can be very dry (**Klupczyńska & Pawłowski, 2021; Postma et al., 2015**).

2.2.2. Heat-induced secondary dormancy

Studies in *A. thaliana* and other model species, such as oilseed rape and oats, have shown that secondary dormancy can be induced after dispersal by even brief exposures to extreme temperatures under both experimental and natural conditions (**Buijs, 2020; Footitt et al., 2011, 2015; Gulden et al., 2004; Malavert et al., 2017; Martel et al., 2018; Ņečajeva et al., 2021; Pawłowski et al., 2020**). For example, late-summer heat delays germination in *A. thaliana* and modifies the expression of genetic variation in flowering time and fitness (**Schmitz et al., 2024**). Variation in secondary dormancy therefore represents a potentially adaptive trait that could sustain population persistence under locally variable and unpredictable environments.

Although cold-induced secondary dormancy has been better studied, for example, in *Butia odorata*, *Rosaceae*, poplars, and conifers such as *Pinus brutia* and *Alnus glutinosa*, (**Klupczyńska & Pawłowski, 2021; Schlindwein et al., 2019; Yamane et al., 2021**), heat-induced secondary dormancy remains less understood. Early studies (**Baskin & Baskin, 1997; Bouwmeester & Karssen, 1993**) established its occurrence in summer annuals, and subsequent work has documented it across diverse species, including tropical (*Comanthera bisulcata*, *Syngonanthus verticillatus*) and Mediterranean taxa (*Cistus* species) (**Duarte & Garcia, 2015; Zomer et al., 2022**). In fire-prone ecosystems, variation in heat-induced dormancy may also be highly relevant for biodiversity conservation (**Cuena Lombraña et al., 2024**). Taken together, these findings highlight heat-induced secondary dormancy as an important yet under-explored axis of natural variation, with potential significance for adaptation across environments.

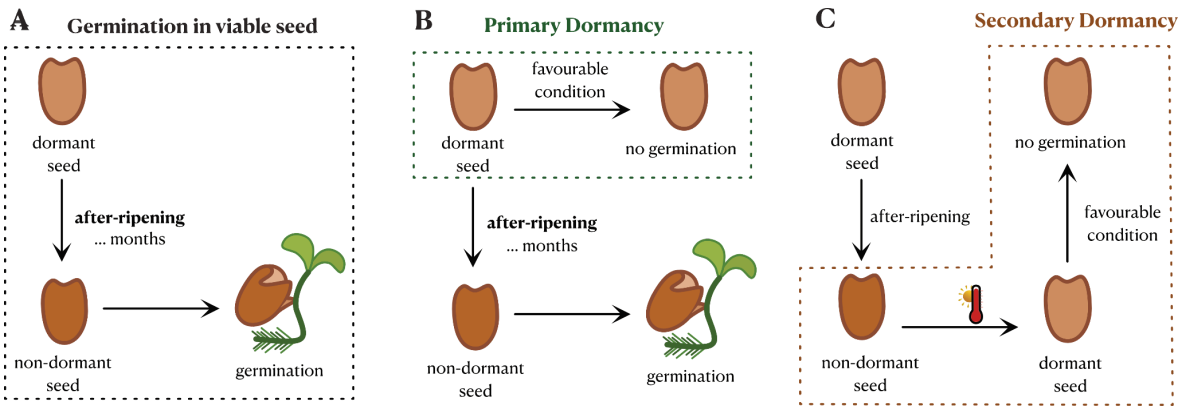


Figure 2.1. Illustrations of different types of dormancy. (**A**) Germination in a viable seed: after-ripening converts a dormant seed into a non-dormant state, allowing germination under favourable conditions. (**B**) Primary dormancy: a dormant seed fails to germinate even under favourable conditions until after-ripening occurs. (**C**) Secondary dormancy: a non-dormant seed re-enters dormancy when exposed to unfavourable conditions, preventing germination despite viability.

2.2.3. Molecular basis of seed dormancy

The molecular underpinnings of natural variation in seed dormancy are best understood for primary dormancy, but much less so for secondary dormancy (**Buijs, 2020; Gianinetti, 2023; Iwasaki et al., 2022**). While both forms of dormancy differ in the conditions under which they are induced, they may share overlapping regulatory pathways.

The gene *DELAY OF GERMINATION 1* (*DOG1*) is a well-established regulator of natural variation in primary dormancy (**Bentsink et al., 2006; Chiang et al., 2013**). *DOG1* conditions seed responsiveness to abscisic acid (ABA), a hormone central to the suppression of germination (**Iwasaki et al., 2022**). Its activity decreases during after-ripening, a process influenced by redox changes but still not fully understood

(**Née et al., 2017**). Allelic variation at *DOG1* contributes to population-level differences in dormancy: for instance, seeds that failed to re-enter dormancy after prolonged cold exposure often carried weak or non-functional *DOG1* haplotypes (**Martínez-Berdeja et al., 2020**). This also suggests that variation in secondary dormancy may partly depend on residual levels of primary dormancy and thus be influenced by the same pathway (**Buijs, 2020; Coughlan et al., 2017**). If so, genetic variation in primary dormancy genes like *DOG1* may indirectly shape responses to post-dispersal cues such as heat or cold. Clarifying the extent to which primary and secondary dormancy share molecular regulators remains an open challenge, but one that could reveal how natural variation in dormancy is partitioned across environmental contexts.

2.2.4. Research questions

In this chapter, I asked three main questions:

- What is the extent and pattern of natural variation in heat-induced secondary dormancy across the species range?
- How does this variation contribute to local adaptation to environmental change?
- What is the molecular basis underlying variation in secondary dormancy?

2.3. Data and Methodology

2.3.1. Seed material

The samples used in this study originated from a collection of 361 accessions across Europe (**Figure S2.1**). Genotype and accession information was obtained from the 1001 genome database (1001genomes.org; (**Alonso-Blanco et al., 2016**)) as well as from (**Wieters et al., 2021**) (**Table S2.1**). Seed material was amplified at University of

Cologne: plants were grown in growth chambers (Dixell, Germany) under long day condition 16:8 (hours) light: dark at 20^0C (day): 18^0C (night) (hereafter referred to as "standard condition"). Plants were vernalised for 4-6 weeks, depending on their genotype, to ensure the required vernalisation period was met. As a result, most plants produced seeds around the same time, with a maximum difference of one week. Mature dry seeds were harvested and packed in paper bags; these seeds were stored at room temperature for approximately six months post-harvest and subsequently transferred to 4°C for long-term storage. For each genotype, one seed batch of around 1000 seeds was harvested.

2.3.2. Germination test

For each genotype, I performed dormancy experiments 6 months, one year, and two years after harvesting, i.e. May 2022 - Trial 1, December 2022 - Trial 2, and December 2023 - Trial 3. Producing and phenotyping multiple independent seed batches for each genotype was beyond our logistical capacities. Each seed dormancy measurement was thus performed using 50-200 seeds of a single seed batch. This number of seeds allowed a robust estimation of germination rate (especially when germination rate was low). While this experimental design limits our ability to quantify the uncertainty around individual genotypic values, the size of the total number of genotypes allows us to assess patterns of genetic covariation in a statistically conservative manner. Due to practical limitations, the first Trial 1 used a set of 295 accessions; Trial 2 and the Trial 3, a set of 344 accessions. 50-200 seeds per genotype were sown on wet filter paper in 12-well plates in a randomized design. There were three incubation treatments, namely, "control," "primary dormancy," and "secondary dormancy" (**Figure 2.2**). As a viability control, seeds were pre-incubated at 4^0C for three days and moved to standard condition. To assess primary dormancy, I put seeds for germination without pre-treatment. To assess secondary dormancy, seeds were pre-incubated at 4^0C in the dark for three days to

release primary dormancy, then exposed to 37°C for four days, and finally moved to standard condition for germination. Germination rate was assessed after seven days, with radicle protrusion serving as the criterion for germination. The number of germinated and non-germinated seeds was recorded by carefully examining radicle emergence under a stereomicroscope. Approximately 50-200 seeds were placed in each well and carefully counted.

2.3.3. Bioclimatic data

Climatic variables were obtained from the WorldClim database (worldclim.org/version2) using the worldclim_global function (R/geoData package, (**Hijmans et al., 2022**)). Last Glacial Maximum data were obtained from Chelsa database (chelsa-climate.org). Latitude and longitude data were used for merging with the bioclimatic data. A resolution of 10-min was selected for this study. Bioclimatic variables (BIO1 – BIO19) were derived from monthly temperature and rainfall values representing annual trends, seasonality, and extreme environmental factors (**Fick & Hijmans, 2017**). The interpretation of bioclimatic variables is provided in **Table S2.8.** Future data were obtained similarly and will be clarified more in the "Species Distribution Modelling" section below.

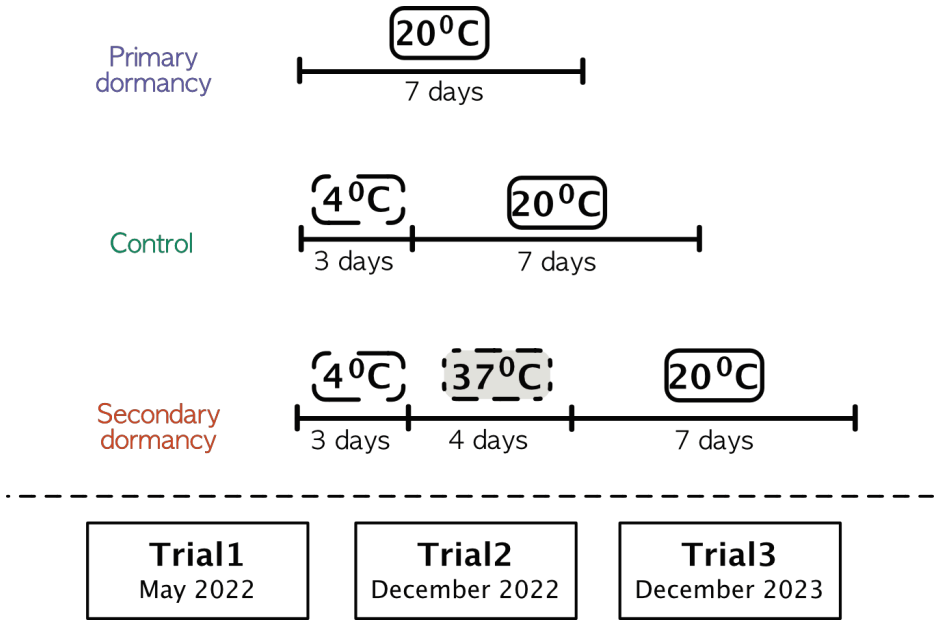


Figure 2.2. Overview of the experimental design for studying primary and heat-induced secondary dormancy. This schematic illustrates the three treatments used to assess germination behaviour and seed dormancy: primary dormancy, secondary dormancy, and control. For the control treatment, seeds were stratified at 4°C for three days to release any dormancy before being exposed to germination conditions. In the primary dormancy treatment, seeds were tested directly without any pre-treatment to assess the baseline dormancy established during seed maturation. For the secondary dormancy treatment, seeds underwent stratification at 4°C for three days to release dormancy, followed by a heat stress of 37°C for four days to induce secondary dormancy. Approximately 50–100 seeds were used per treatment in each genotype. Germinated and non-germinated seeds were counted after a seven-day germination period under long-day conditions (16 hours light, 20°C). Germination rates, calculated based on the proportion of germinated seeds, were used in subsequent analyses.

2.3.4. Regression analysis

To determine which bioclimatic variables influenced the strength of secondary dormancy, data were analysed using a generalised linear mixed model (GLMM) with a binomial distribution and individual relatedness as random effect. Secondary dormancy was modelled as a binomial response variable, defined by the number of germinated seeds (successes) relative to the total number of seeds per replicate. Thus, dormancy was analysed as a binary outcome (success/failure) at the seed level using a logistic regression framework, with climatic variables included as predictors.

In such a logistic regression model (binomial family with logit link function), the probability of success πi for the outcome yi is defined as: $logit(\pi_i) = ln[\pi_i/(1- \pi_i)]$

The linear predictor including fixed effects terms: $n_i = \beta_o + \beta_1 * (treatment) + \beta_2 * (bioclimatic\ variable) + \beta_3 *(treatment)* (bioclimatic\ variable) + u_i$

where n_i is germination rate; β_o is the intercept; β_1, β_2, β_3 are coefficients for treatments, bioclimatic variables, and their interactions, respectively; $u_i \sim N(0, \sigma^2 K)$ is the random effect, where K is the kinship matrix, accounting for population structure. The kinship-correlated random effect was thereby controlling for potential confounding effects on associations between phenotypes and explanatory variables due to shared ancestry.

The model was fitted with the relmatGlmer function of package lme4qtl (**Ziyatdinov et al., 2018**) using logit link function. The kinship matrix had to be converted to positive definite matrix, using the nearPD function of the Matrix package (**R Core Team, 2021**). The resulting model coefficients reflect the effect size of each predictor on dormancy variation.

2.3.5. Species distribution modelling

Species distribution model (SDM) was used to compare past, present and future habitat suitability of two ecotypic groups split around the median of corrected secondary dormancy measured in trial 1 (i.e., the residuals of secondary dormancy regressed on primary dormancy for the trial where secondary dormancy was the strongest). The aim of the SDM analysis was to evaluate whether these two groups had equal likelihood to find suitable habitats over time. Genotypes of the strong-secondary-dormancy group had values above the trial 1 median, whereas genotypes of the weak-secondary-dormancy group had values below the trial 1 median. I used the R/biomod2 package (**Thuiller et al., 2009**) with the parameters explained in the following four steps. This procedure and the results are reported according to the standard protocol proposed by (**Zurell et al., 2020**).

Step 1 - Overview: The spatial extent was limited from -10^0 to 50^0 for longitude and from 35^0 to 70^0 for latitude. The biodiversity data type was set on presence-only, with the locations of the samples marked as present. For predictor variables, I selected bioclimatic factors that are significantly associated with dormancy. The spatial resolution was set to 10-minute.

Step 2 - Data preparation: As only presence data were available, pseudo-absence data needed to be created, because presence/absence data are required for most SDM algorithms. To ensure these pseudo-absences were environmentally distinct from presence locations, I used the Surface Range Envelope (SRE) strategy within the biomod2 framework. This method excludes areas with environmental conditions similar to those of the presence points, reducing the risk of misclassifying potential presences. I generated three sets of pseudo-absences, with the number of pseudo-absences equal to the number of presence points in each replicate, following the recommendations of **Barbet-Massin et al. (2012).**

For future data, I chose the representative concentration pathway (RCP) of 4.5 from the Coupled Model Intercomparison Project Phase 5 (CMIP5). RCP 4.5 corresponds to an intermediate scenario in which emissions peak at around 2040 and then decline. RCP 4.5 - the most probable scenario given that no climate policies are applied - considers non-renewable fuel availability. All bioclimatic variables were tested for multi-collinearity. I tested all predictor variables for multicollinearity and found that none were highly correlated ($|r| < 0.7$); thus, no variable re-selection or dimensionality reduction (e.g., PCA) was required.

Step 3 - Model options: I tested different model algorithms, namely, Random Forest (RF), Gradient Boosting Machine (GBM), Generalized Linear Model (GLM), and Maxent. For calibration and evaluation, I implemented 3-fold cross-validation using a random splitting approach, whereby 80% of the data were used for training and 20% for testing in each replicate. This corresponds to random k-fold cross-validation, a standard method for evaluating model robustness when spatial autocorrelation is not the primary focus. I adopted this approach as the primary goal was to assess differences in habitat suitability between genotype groups. Assuming these two groups have similar dispersal abilities, the limitation of SDM applies equally to the two groups. I was aware of imperfect detection because the data had sampling bias issues, that is, not all *A. thaliana* in the given geographical range were sampled. The sampling of individuals in the Central European region was less dense than that in Sweden and Spain. Cross-validation, which represents the random effects of selecting data, can show the sensitivity of the models to the input data. The evaluation metrics included TSS, ROC, and KAPPA scores.

Step 4 - Assessment and Prediction: The biomod2 package examines the importance of each variable in the final model. Once the model is trained, a standard prediction is made, one variable is randomised, and a new prediction is made. The correlation score between the new prediction and standard prediction is given as an estimation

of the variable importance in the model. The influence of each variable was visualised by plotting response curves. The ensemble modelling option combines individual models to build a meta-model. Models with TSS, ROC, and KAPPA lower than 0.8 were excluded. I used binary projection using ROC and a so-called weighted mean, which was done for all chosen models for each run, to project current and future climate models.

2.3.6. Genome-wide association analysis

Variant calling was performed from raw sequence data for all 361 lines simultaneously. Genomic data of 309 genotypes from the 1001Genome data set are available on the European Nucleotide Archive (ENA), and data of 52 accessions were obtained from **Wieters et al. (2021)** (**Table S2.1** for accession information). The SRA files were downloaded from ENA using prefetch and converted into FASTQ using fastq-dump (from sra toolkit, (**SRA Toolkit Development Team, 2023**)). Low-quality signals were detected, and polyG and sequencing adapters were removed using fastp (**Chen et al., 2018**). For this, the minimum read length was set to 50bp, the minimum Phred quality score was 15, and the quality threshold was 40%. The sequences were mapped to the reference genome (TAIR10, arabidopsis.org) using bwa-mem (**Li, 2013**) and converted to bam files using samtools (**Danecek et al., 2021**). The mapping quality was evaluated using multiqc (**Ewels et al., 2016**). Variant calling was performed using bcftools mpileup and bcftools call (**Danecek et al., 2021**; **Li, 2011**). The minimum mapping quality was set to 20, and the minimum base quality was set to 30. SNP data were processed with bcftools v1.18, vcftools v0.1.16, and PLINK2 (**Danecek et al., 2011, 2021; Purcell et al., 2007**) for minor allele frequency filtering (0.05), maximum missingness (0.95), minimum depth and maximum depth (minDP = 10 and maxDP = 50), indels and low-quality calls (< 30), and prune linkage disequilibrium ($r > 0.1$, scanning 50kb window, 10bp step size). Genome-wide association analysis (GWAS) was conducted to identify genetic

variants associated with secondary dormancy. Association analysis was performed with GEMMA using PLINK files (**Zhou & Stephens, 2012**). Kinship matrix was calculated using the KING algorithm of PLINK, and GWAS was performed using the Mixed Linear Model (MLM) approach (-lmm 4). The –log10 p-values of the model were adjusted using Bonferroni correction at a significance level of 5%. Chromosomal position and candidate genes were checked using the available annotated genes in the TAIR database (arabidopsis.org). I controlled for population structure using kinship matrix and added primary dormancy as a covariate to control for seed age difference across the trials. I combined the p-values from the different GWAS trials using Fisher's combined probability test, which provides a robust approach for synthesizing results from multiple studies **(Walsh & Lynch, 2018a**). Fisher's method is based on the assumption that the p-values under the null hypothesis are uniformly distributed between 0 and 1. The method transforms each p-value using the natural logarithm and then sums these transformed values. It computes a combined test statistic χ^2 given by

$$\chi^2 = -2\sum ln(p_i)$$

where p_i represents the p-values from the individual trials. This test statistic follows a chi-square distribution with $2k$ degrees of freedom, where k is the number of studies combined. I used Python and libraries namely numpy (**Harris et al., 2020**), scipy.stats (**McKinney, 2010**), and pandas (**Virtanen et al., 2020**).

2.4. Results

2.4.1. Secondary dormancy variation correlates with primary dormancy

To quantify genetic variation in heat-induced dormancy with latitude, I used seeds collected from 361 European accessions of A. *thaliana* grown in a common garden. Seed dormancy was measured in three trials, with seeds sampled from the same

seed batch. The trials differed in the number of genotypes that were analyzed: 295, 361 and 344 genotypes, and seeds were six, twelve and twenty-four months after-ripened in trials 1, 2, and 3, respectively. I implemented three treatment groups for the germination tests: a germination test without any treatment to assess primary dormancy; a germination test following exposure to 4°C to evaluate seed viability, hereafter the control treatment; and a germination test after sequential exposure to 4°C and 37°C to assess secondary dormancy (**Figure 2.2**).

Both the number of lines exhibiting secondary dormancy and the strength of this dormancy decreased across the experiments (**Figure 2.3**). Seed viability remained high after 12 months of storage, because the control germination rate approached 100% after cold treatment. This rate, however, dropped to 50% for seeds after 24 months of seed storage. Primary dormancy was progressively reduced from Trial 1 (6-month-old seeds) to Trial 2 (12 months old seeds) and almost completely lost in Trial 3 (24-month-old seeds) (**Figure 2.3**).

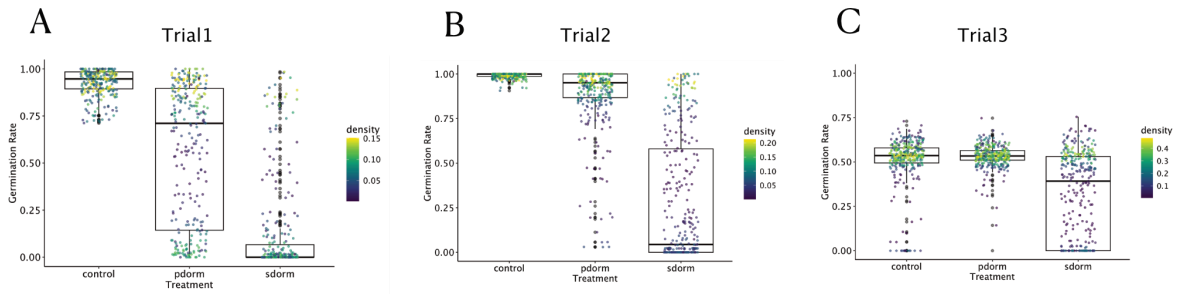


Figure 2.3. Germination rates across three treatments over three trials. Germination rates were measured for primary dormancy (pdorm), secondary dormancy (sdorm), and control treatments across three trials. Germination rates were measured for primary dormancy (pdorm), secondary dormancy (sdorm), and control treatments across three trials. The colour gradient represents the density of data points. (**A**) Trial 1: conducted in May 2021 (6 months post-harvest). (**B**) Trial 2: conducted in December 2021. (**C**) Trial 3: conducted in December 2022. The "sdorm" treatment represents germination rates following a 3-day stratification at 4°C to release dormancy, followed by a 4-day treatment at 37°C. The pdorm treatment tested germination rates without any pre-treatment, whereas the control treatment involved seeds stratified for 3 days at 4°C to release dormancy before testing germination. All germination tests were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days.

Levels of seed secondary dormancy were consistently correlated across trials ($p < 2.814e-09$, **Table S2.2**). While primary dormancy was released by stratification prior to heat exposure, secondary dormancy remained significantly correlated with primary dormancy in all three trials (maximum $p = 0.00135$, **Table S2.3**). However, the strength of this correlation declined over time (Trial 1: Spearman rho = 0.427; Trial 3: Spearman rho = 0.172). Residual variation also increased across trials, confirming the increasing divergence from primary dormancy signal (**Figure S2.2**).

2.4.2. Bioclimatic drivers of heat-induced secondary dormancy

Despite the variation in levels of dormancy observed across trials, secondary dormancy decreased with increasing latitude in all three trials (**Figure 2.4**, **Table S2.4**). This correlation may arise from the history of post-glaciation expansion in the species, from the indirect correlation to primary dormancy, or from the specific adaptation of secondary dormancy along climatic clines. To discriminate between these hypotheses, I tested whether the geographic distribution of secondary dormancy variation was significantly associated with climatic variables after accounting for individual relatedness and both variance in primary dormancy and variance in germination rate. Using publicly available global climate data with 10-minute resolution in longitudinal and latitudinal coordinates, I conducted a generalized linear mixed model (GLMM) analysis to investigate the association between 19 bioclimatic variables and germination outcome (**Figure 2.5**).

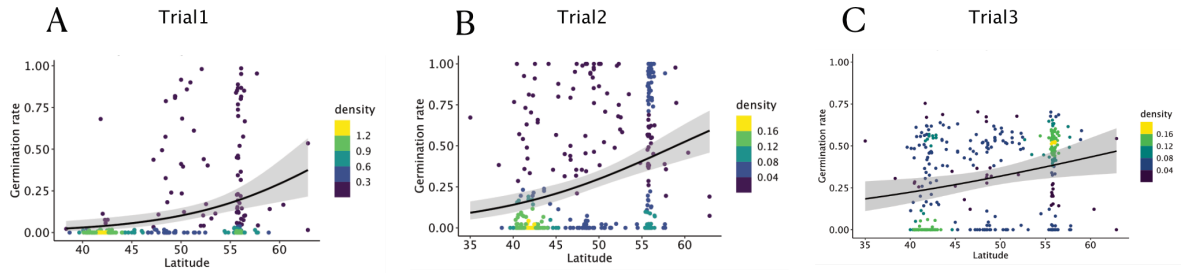


Figure 2.4. Germination rates of heat-induced secondary dormancy treatment as a function of latitude of origin. The secondary dormancy treatment represents germination rates following a 3-day stratification at 4°C to release dormancy, followed by a 4-day treatment at 37°C. Germination tests were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days. The correlation between germination rate and latitude of origin was assessed for three trials: (**A**) Trial 1, (**B**) Trial 2, and (**C**) Trial 3.

Original values were used without any corrections or adjustments. Spearman's rank correlation coefficients (rho) and corresponding p-values were as follows: (**A**) rho = 0.386, $p = 6.823 \times 10^{-12}$; (**B**) rho = 0.339, $p = 2.986 \times 10^{-11}$; and (**C**) rho = 0.278, $p = 1.144 \times 10^{-7}$. Colour density highlights regions with increasingly high concentration of values.

Genotypes exhibiting strong secondary dormancy tended to originate from locations associated with increasing mean temperatures and decreasing rainfall (mean temperature of driest quarter - BIO9 values). Summer precipitation (mean precipitation of warmest quarter - BIO18 values) was negatively correlated with secondary dormancy (minimum p < 2.2e-16, **Table S2.5**). Additionally, there was a positive correlation between winter precipitation (mean precipitation of coldest quarter - BIO19 values) and heat-induced secondary dormancy (maximum p = 3.55e-8, **Table S2.5**). Variation in secondary dormancy is further associated with climatic fluctuations. Higher isothermality (BIO3) - indicative of greater diurnal temperature variation relative to annual temperature range - was significantly associated with increased levels of secondary dormancy (maximum p = 1.77e-6; **Table S2.5**). Altogether, these results suggest that A. *thaliana* populations evolved stronger heat-induced secondary dormancy in environments characterized by low precipitation, high temperatures, and pronounced short-term temperature fluctuations, conditions likely serving as ecological cues signaling germination stress (**Figure 2.5**, **Figure S2.3**, **Table S2.5**). Since this correlation of phenotypic variation with environmental parameters was significant after accounting for population structure (see **2.3. Data and Methodology**), it can be concluded that the distribution of genetic variation for heat-induced secondary dormancy was shaped by natural selection.

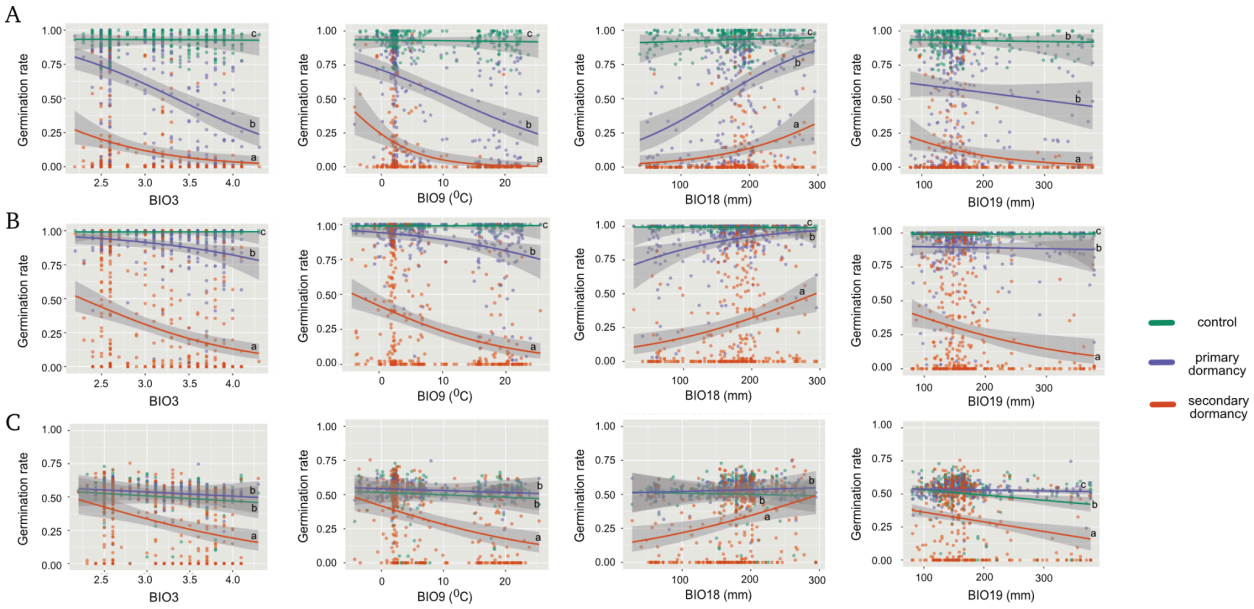


Figure 2.5. Germination rate across three treatments correlated with four bioclimatic variables (BIO3, BIO9, BIO18, and BIO19). The relationship between germination rates and four bioclimatic predictors—BIO3 (isothermality), BIO9 (mean temperature of the driest quarter), BIO18 (mean precipitation of the warmest quarter), and BIO19 (mean precipitation of the coldest quarter) - across three trials: (**A**) Trial 1, (**B**) Trial 2, and (**C**) Trial 3. The so-called corrected secondary dormancy treatment represents germination rates following a 3-day stratification at 4°C to release dormancy, followed by a 4-day treatment at 37°C, corrected for the residual effects of primary dormancy. The primary dormancy treatment tested germination rates without any pre-treatment, and the control treatment involved seeds stratified for 3 days at 4°C to release dormancy before testing germination. All germination tests were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days. The bioclimatic variables include BIO3, which is unitless; BIO9, measured in degrees Celsius; and BIO18 and BIO19, measured in millimeters.

Primary dormancy variation was also associated with isothermality (BIO3), mean temperature of driest quarter (BIO9), and summer precipitation (BIO18) in the first

two trials (maximum p < 2.2e-16, **Table S2.5**), but not with winter precipitations (BIO19) (maximum p = 0.16196, **Table S2.5**). In the third trial, only BIO19 was associated with remaining primary dormancy variation (p = 0.00143, **Table S2.5**). As primary dormancy decreased across trials, the slope of its correlation with bioclimatic variables also decreased (**Table S2.5**, **Figure 2.5**). Interestingly, the slope of the relationship between secondary dormancy and each of these four climatic variables was always significantly stronger than the relationship with primary dormancy, irrespective of the level of residual primary dormancy (**Figure 2.5**). The shared association between secondary dormancy and primary dormancy with temperature- and rainfall-related bioclimatic variables suggests that these traits may constitute an adaptive dormancy syndrome that coordinates dormancy responses to seasonal and climatic fluctuations. The sharper climatic clines of secondary dormancy compared to primary dormancy suggest that secondary dormancy is a better descriptor of local adaptation than primary dormancy.

2.4.3. Suitable habitats are predicted to increase for ecotypes with high secondary dormancy

Since secondary dormancy levels displayed a pattern of local adaptation along several bioclimatic gradients, we reasoned that the fitness of high vs. low secondary dormancy ecotypes would depend on the relative size of their suitable habitat (**Banta et al., 2012; Ikeda et al., 2017**). Using the four identified bioclimatic factors shown above to drive local adaptation of secondary seed dormancy (BIO3, BIO9, BIO18, BIO19), I quantified the amount of suitable habitat inferred by projecting species distribution models for both ecotypic groups (**Figure 2.6**). I used current data to determine the relationship of each ecotype with bio-climatic variables drawing their respective niches. I then used past climate data from the Last Glacial Maximum and future climate data based on RCP4.5 (see **2.3. Data and Methodology**).

Among the different models, the best model performance was obtained with Random Forest (ROC and TSS scores > 0.95; **Figure S2.4**). As anticipated, given its strong influence on dormancy ecotypes, BIO9 determined about 40% of the distribution of both strong- and weak-secondary dormancy ecotypes (**Figure S2.5**). BIO9 is particularly influential because seeds experience both heat and drought during this period, making the triggering of dormancy-related responses crucial for seedling survival. Furthermore, BIO9 likely interacts with other predictors, such as precipitation variables (e.g., BIO18 and BIO19), shaping the seasonal microclimatic conditions that define the ecological niches in which dormancy strategies confer an advantage.

Our model showed both the proportion of stable habitats for both ecotypic groups from LGM to current time, as well as the change in habitat boundaries. Past and predicted habitat range change depended on the secondary seed dormancy strategy (**Figure 2.6**, LGM-Current: X-squared = 108.03, df = 2, p-value < 2.2e-16; Current-2050: X-squared = 814.45, df = 2, p-value < 2.2e-16). Habitat suitability was quantified by the amount of habitat, measured as pixel in the grid, that fits the current distribution of the two secondary dormancy types. Habitat suitability increased for strong dormancy ecotypes in northwestern and southern areas of Europe such as in the Balkans and Spain, compared to weak secondary dormancy (measured in pixels). In total, weak secondary dormancy genotypes appeared to have benefitted from a larger habitat gain due to significant gains in central and northeastern Europe. However, in the close future, only strong dormancy ecotypes are predicted to maintain most of their current habitat, weak secondary dormancy ecotypes, instead, are predicted to lose around half of their current habitat (primarily in western and southern Europe) with only slight gains (in northeastern Europe) (**Figure 2.6**, LGM - current: X-squared = 200.6, df = 2, p-value < 2.2e-16, current - 2050: X-squared = 4484, df = 2, p-value < 2.2e-16). In summary, the climate

projection, which here is based on an intermediate scenario with emission peak around 2040, indicates that ecotypes with strong secondary dormancy should lose less habitat than those with weak secondary dormancy. Strong secondary dormancy ecotypes are therefore predicted to be more resilient to expected environmental changes. Based on the size of predicted suitable habitats in changing climatic conditions, high-secondary dormancy ecotypes are predicted to make up an ever increasing fraction of the A. *thaliana* population. Assuming that high- and low-dormancy ecotypes do not differ in their ability to migrate to new habitats, these findings suggest that high-dormancy genotypes have a fitness advantage.

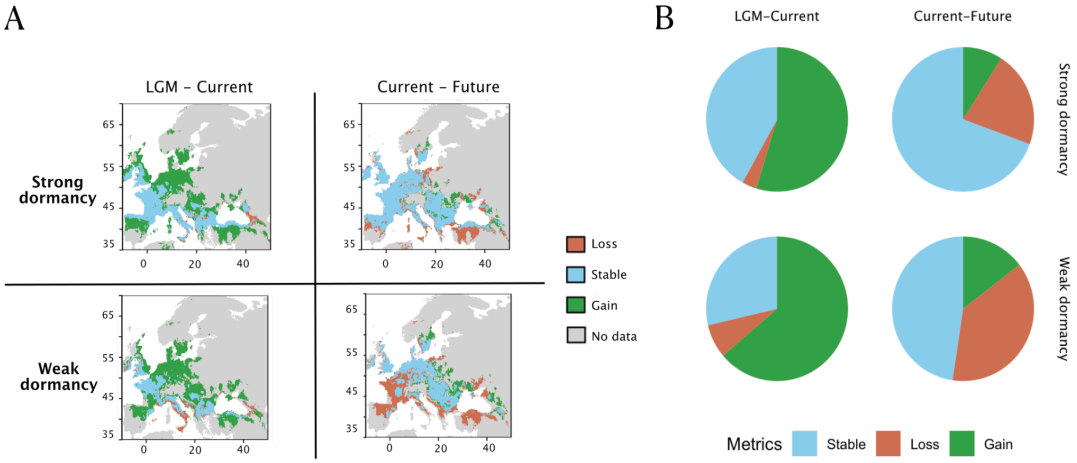


Figure 2.6. Species distribution model (SDM) for strong and weak heat-induced secondary dormancy genotypes. To construct the SDM, four bioclimatic variables were used: temperature seasonality (BIO3), mean temperature of the driest quarter (BIO9), precipitation of the warmest quarter (BIO18), and precipitation of the coldest quarter (BIO19). Data from Trial 2, the most complete dataset, was used for the analysis. Genotypes were divided into strong and weak secondary dormancy groups based the secondary dormancy corrected for primary dormancy, so-called corrected secondary dormancy, which represents germination rates following a 3-day stratification at 4°C to release

dormancy, followed by a 4-day treatment at 37°C, corrected for the residual effects of primary dormancy. The figure presents species range changes in three categories—loss (orange), stable (blue), and gain (green)—alongside areas (gray) that have no data due to insufficient number of data points in those regions (e.g., northeastern Europe). Results are derived from an ensemble model, which combines the best-performing model based on ROC metrics and calculates weighted means across three cross-validation runs. (**A**) Species range changes for strong secondary dormancy genotypes (upper) and weak secondary dormancy genotypes (lower) during two transitions: LGM to the present (left) and the present to 2050 (right). (**B**) Pie charts depicting the proportion of habitat loss, stability, and gain for each group during these transitions. Statistical analyses reveal significant differences in range change patterns between strong and weak dormancy genotypes (chi-squared test: LGM-current: $X^2 = 200.66$, $df = 2$, $p < 2.2 \times 10^{-16}$; current-future: $X^2 = 4484$, $df = 2$, $p < 2.2 \times 10^{-16}$).

2.4.4. Genetic variants associated with the environment and secondary dormancy

To investigate the genetic basis of the dormancy variation revealed in our experiment, I performed genome-wide association studies (**Figure 2.7**). For primary dormancy, there was a significant peak located only 14bp away from the region of the *DELAY OF GERMINATION 1* (*DOG1*) gene, which was not found to be associated with secondary dormancy (**Figure S2.6D**; **Table S2.6**). In order to identify the specific genetic basis of secondary dormancy, I accounted for variation in primary dormancy in the model of genetic association. Several SNPs with significant associations were identified, notably, on chromosomes 2, 3, and 4 (**Figure S2.7**, **Table S2.7**), as annotated in **Figure 2.7A**. The location of these three associated SNPs did not overlap with the genomic regions associated with primary dormancy (**Figure 2.7B**). SNPs on chromosomes 2 and 4 did not fall close to any obvious candidate gene for secondary dormancy, but they may represent novel loci, or the

regulatory elements involved in dormancy-related pathways that have yet to be characterised.

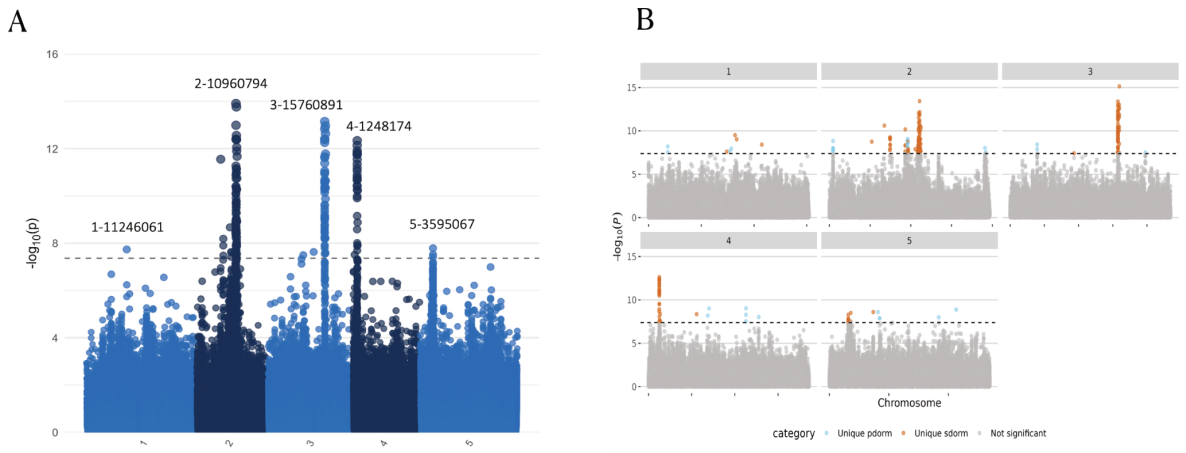


Figure 2.7. GWAS results for heat-induced secondary dormancy and primary dormancy SNP associations. **(A)** Manhattan plot illustrating the association of 1.2 million SNP markers with heat-induced secondary dormancy, using primary dormancy as a covariate. In the primary dormancy treatment, seeds were tested directly without any pre-treatment to assess the baseline dormancy established during seed maturation. For the secondary dormancy treatment, seeds underwent stratification at 4°C for three days to release dormancy, followed by a heat stress of 37°C for four days to induce secondary dormancy. The results are based on a combined p-value GWAS analysis of three trials, calculated using Fisher's combined probability formula. The dashed horizontal line represents the significance threshold after applying a 5% Bonferroni correction. Significant SNPs crossing this threshold highlight genomic regions associated with secondary dormancy independent of primary dormancy. **(B)** Overlay of two Manhattan plots to compare SNP associations for heat-induced secondary dormancy without covariates (orange) and primary dormancy (blue). SNPs uniquely associated with primary dormancy are highlighted in blue (unique pdorm), and SNPs uniquely associated with secondary dormancy are highlighted in orange (unique sdorm). Gray dots represent SNPs that do not meet the 5% Bonferroni correction significance threshold for either dormancy trait.

The SNPs on chromosome 3 were particularly intriguing, because they were located approximately 100kb upstream and 60kb downstream of the *DNAJ Homolog 3* gene, (AT3G44110, *AtJ3*, or *AtDjA3*). Although several genes are present in this region, *AtJ3* encodes for a heat-shock co-chaperone that is involved in the control of germination in stressful conditions, making it a likely candidate (**Barghetti et al., 2017; Salas-Muñoz et al., 2016; Wu et al., 2019**). Several elements indicate that the absence of associated SNPs within and or close to the gene may be due to mapping issues. The gene is surrounded by numerous repeats that perturb the quality of mapping. In addition, in the *Col-0* reference genome, substantial deletions in the gene further complicate SNP detection within the gene body.

2.5. Discussion

In this study, it was shown that heat-induced secondary dormancy represents a distinct but connected axis of natural variation in seed dormancy within A. *thaliana*. While primary dormancy is well known to contribute to adaptation to local climates (**Chiang et al., 2013; Debieu et al., 2013; Kronholm et al., 2012**), our results demonstrate that secondary dormancy, too, covaries with climate and likely contributes to the diversification of germination strategies across environments.

At the molecular level, our findings highlight both shared and unique pathways underlying secondary dormancy. On the one hand, its dependence on residual primary dormancy and its correlation with *DOG1* haplotypes confirm overlap with established dormancy regulators (**Coughlan et al., 2017; Martínez-Berdeja et al., 2020**). On the other hand, the identification of *AtJ3* - a heat-shock chaperone involved in stress responses - as a candidate locus suggests that secondary dormancy also recruits pathways beyond the canonical dormancy network. This combination of pleiotropic effects and independent regulators indicates that

natural variation in secondary dormancy is shaped by both constraints and novel opportunities for adaptation.

Ecologically, secondary dormancy variation tracks climatic variables such as summer heat and precipitation, pointing to its adaptive relevance. Genotypes from warmer, drought-prone regions exhibit stronger secondary dormancy, consistent with the hypothesis that delaying germination under harsh conditions increases survival (**Springthorpe & Penfield, 2015**). These associations persisted after controlling for population structure, confirming that the observed covariation reflects local adaptation rather than demographic history. Together with previous work, our findings suggest that secondary dormancy forms part of an adaptive trait syndrome with primary dormancy and life-history timing traits (**Debieu et al., 2013; Takou et al., 2019**).

Finally, our exploratory distribution models illustrate how the climatic envelope associated with secondary dormancy may have shifted in the past and may continue to shift under future climate scenarios. While such projections cannot capture the full evolutionary potential of the species, they emphasise that secondary dormancy is likely to become increasingly important under intensifying summer heat and drought (**Exposito-Alonso et al., 2019**). Future experimental tests of fitness differences among genotypes with contrasting secondary dormancy capacities will be essential to validate its adaptive role.

This study advances the understanding of natural variation in seed dormancy by showing that heat-induced secondary dormancy is both genetically distinct and ecologically significant. This dual role highlights secondary dormancy as a key but previously underappreciated component of the adaptive variation that shapes germination strategies in A. *thaliana*.

2.6. Conclusion

This chapter sets out to investigate how natural variation in seed dormancy contributes to adaptive potential in A. *thaliana*, focusing on three central questions. Both primary and secondary dormancy exhibit adaptive clines along climatic gradients, but secondary dormancy shows a stronger and more consistent signal, suggesting it is a key axis of natural variation shaping germination strategies. Next, by combining trait measurements with climatic associations and species distribution models, it is shown that genotypes with stronger secondary dormancy are favored in environments with harsh summers, and may be more resilient under future climate change scenarios. The analyses further reveal that while secondary dormancy shares regulatory pathways with primary dormancy, including connections to *DOG1*, it also involves distinct loci such as *AtJ3*. This highlights both pleiotropic constraints and novel pathways that generate natural variation in dormancy.

Building on these empirical insights, it is important to place such fitness-related traits within a broader theoretical framework of quantitative genetics. Understanding how these traits respond to selection requires evaluating how their genetic variance contributes to heritability and long-term adaptive potential. By connecting natural variation in phenotypes with the underlying variance components, we can better assess not only its ecological relevance but also its evolutionary capacity to fuel adaptation.

2.7. Supplementary Materials

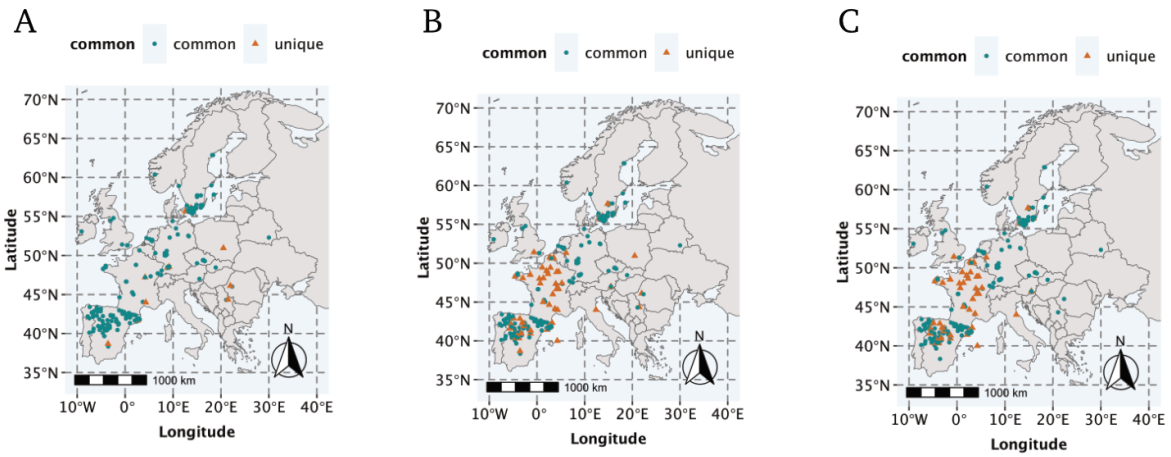


Figure S2.1. Geographic origin of European *Arabidopsis thaliana* accessions used in this study. The map displays the geographic locations of *Arabidopsis thaliana* accessions included in the study, shown separately for each trial. (**A**) Trial 1, (**B**) Trial 2, and (**C**) Trial 3. Green round markers represent common genotypes that are present across all three trials, while orange triangular markers indicate unique genotypes that are specific to a single trial.

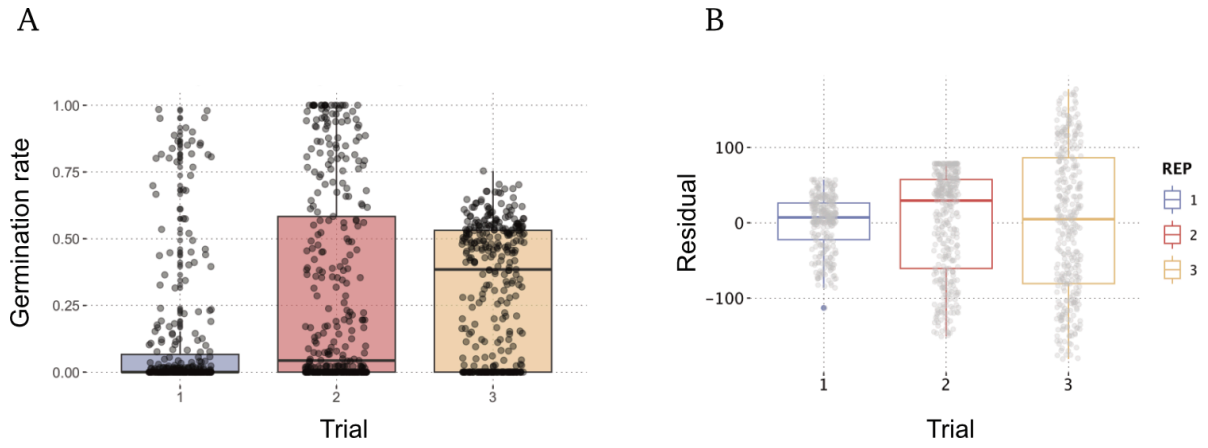


Figure S2.2. Temporal dynamics of heat-induced secondary dormancy. (**A**) Distribution of germination rates over time in response to the heat-induced secondary dormancy treatment across three trials. The x-axis represents time points corresponding to the three trials. Secondary dormancy treatment represents germination rates following a 3-day stratification at 4°C to release dormancy, followed by a 4-day treatment at 37°C. (**B**) Residuals of the linear regression of secondary dormancy rank on primary dormancy across the three trials. Primary dormancy tested germination rates without any pre-treatment. The residuals represent the component of secondary dormancy not explained by primary dormancy, highlighting the change in dormancy traits over time. All germination tests were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days.

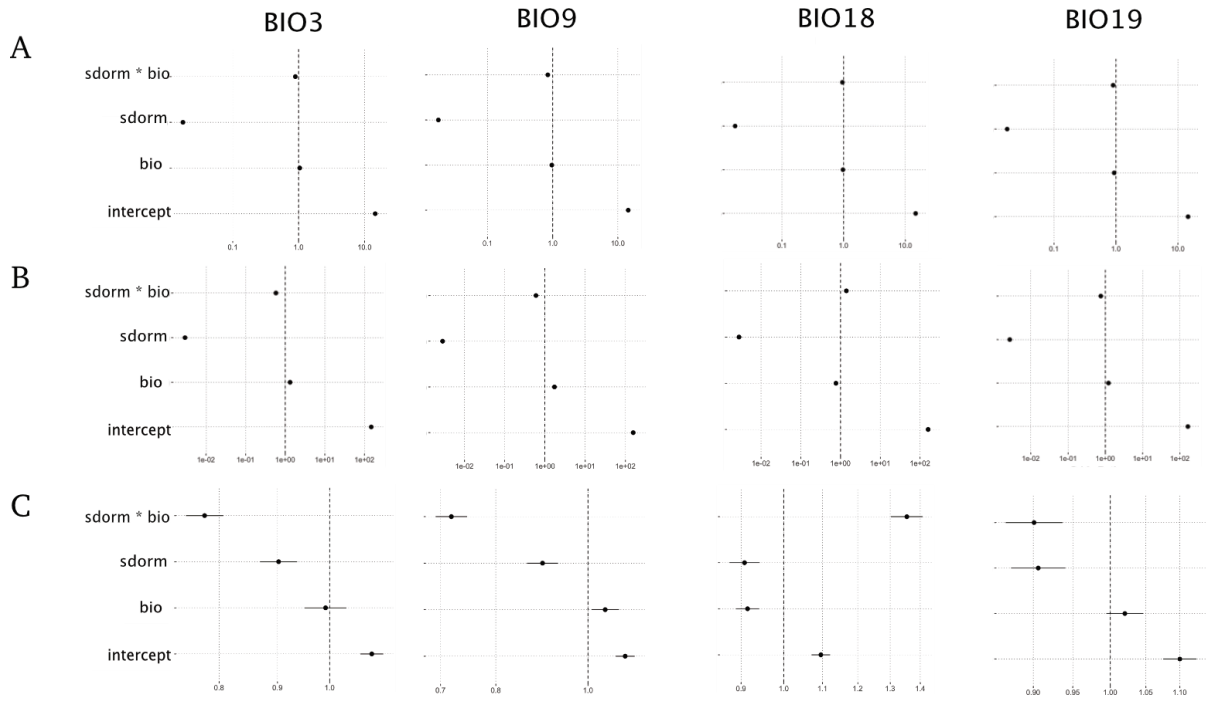


Figure S2.3. Bioclimatic variables as predictors of heat-induced secondary dormancy. Odds ratios are derived from binomial models with a logit link function for the effect of four bioclimatic variables - BIO3 (isothermality), BIO9 (mean temperature of the driest quarter), BIO18 (mean precipitation of the warmest quarter), and BIO19 (mean precipitation of the coldest quarter)—as predictors of secondary dormancy. Results are displayed separately for three replicates: (**A**) Trial 1, (**B**) Trial 2, and (**C**) Trial 3. Secondary dormancy (sdorm) was induced by a 4-day treatment at 37°C following a 3-day stratification at 4°C to release dormancy. Primary dormancy (pdorm) seeds were tested for germination without pre-treatment, and control seeds, used here as the reference level, were stratified for 3 days at 4°C and tested for germination. The sdorm values used in the models represent secondary dormancy corrected for primary dormancy (residuals of the regression of secondary dormancy on primary dormancy). The models assess the effect of each bioclimatic variable (bio) on germination outcomes under long-day conditions at 20°C, with germination tested after 7 days. BIO3 = isothermality (ratio; unitless); BIO9 = mean temperature of the driest quarter (°C); BIO18 = mean

precipitation of the warmest quarter (mm); BIO19 = mean precipitation of the coldest quarter (mm).

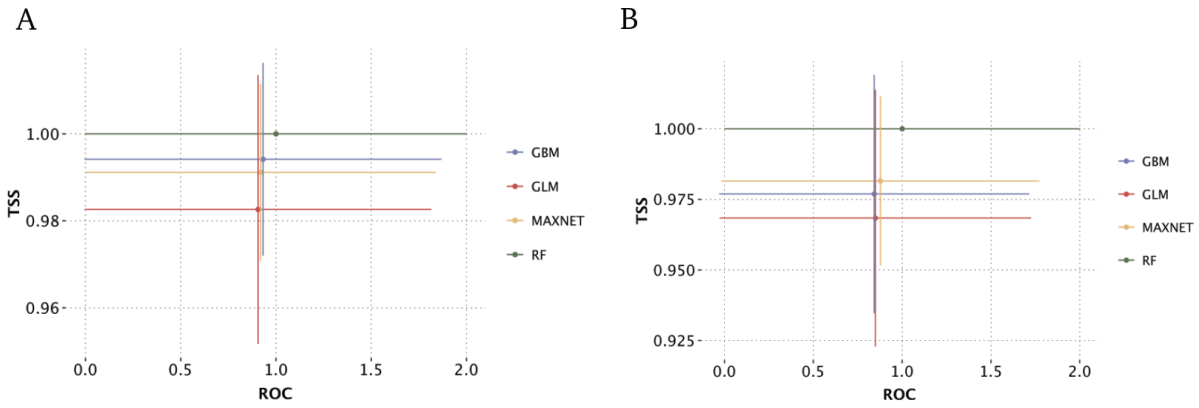


Figure S2.4. Performance comparison of modelling approaches using ROC and TSS metrics in predicting heat-induced secondary dormancy ecological niche. Comparisons of the performance of four modeling approaches: gradient boosting machine (GBM, blue), generalized linear model (GLM, red), maxnet (MAXNET, yellow), and random forest (RF, green). The comparison is based on two evaluation metrics: Receiver Operating Characteristic (ROC) and True Skill Statistic (TSS). The ROC metric assesses the model's ability to distinguish between classes, according to which a higher value indicates improved ability to discriminate. TSS evaluates the model's predictive skill, balancing sensitivity and specificity. These metrics provide a comprehensive evaluation of model performance across different approaches, facilitating the identification of the best method for predicting secondary dormancy. The color coding corresponds to the modeling approaches: blue for GBM, red for GLM, yellow for MAXNET, and green for RF. Each point on the plot represents the performance of a model under the respective metric.

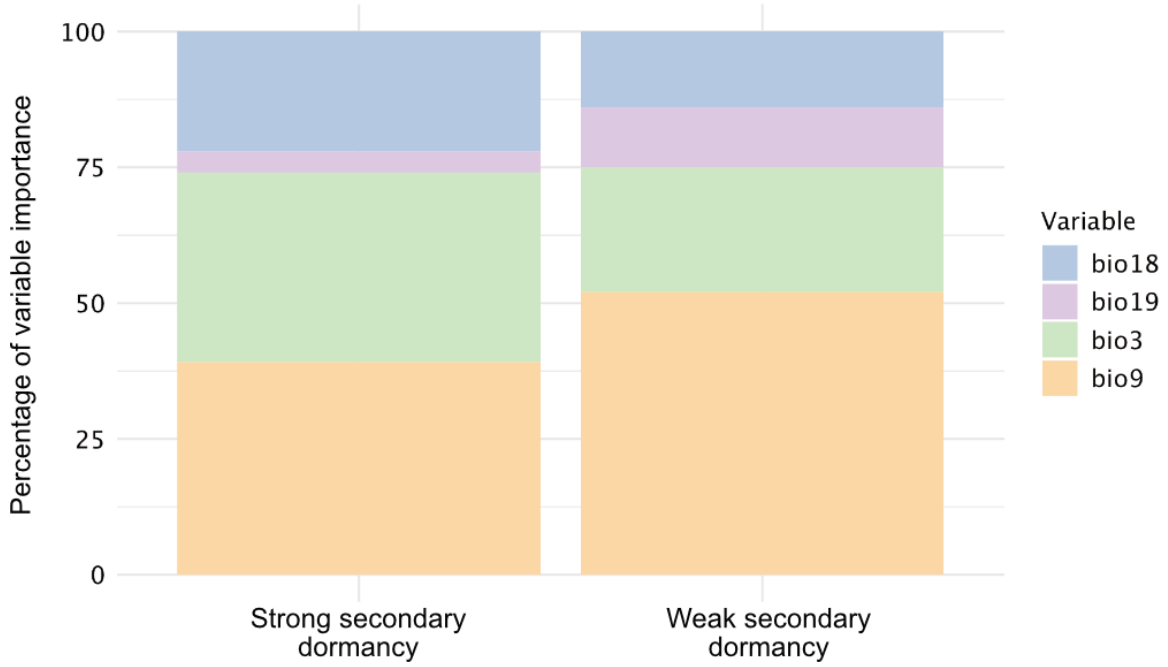


Figure S2.5. Importance of predictor variables in the species distribution model for strong and weak heat-induced secondary dormancy ecotypes. The relative importance of four bioclimatic variables in the species distribution model (SDM) used to predict the habitat suitability for strong and weak secondary dormancy ecotypes of Arabidopsis thaliana. The variables included in the model are isothermality (BIO3), mean temperature of the driest quarter (BIO9), mean precipitation of the warmest quarter (BIO18), and mean precipitation of the coldest quarter (BIO19). Trial 1 data were used for this analysis.

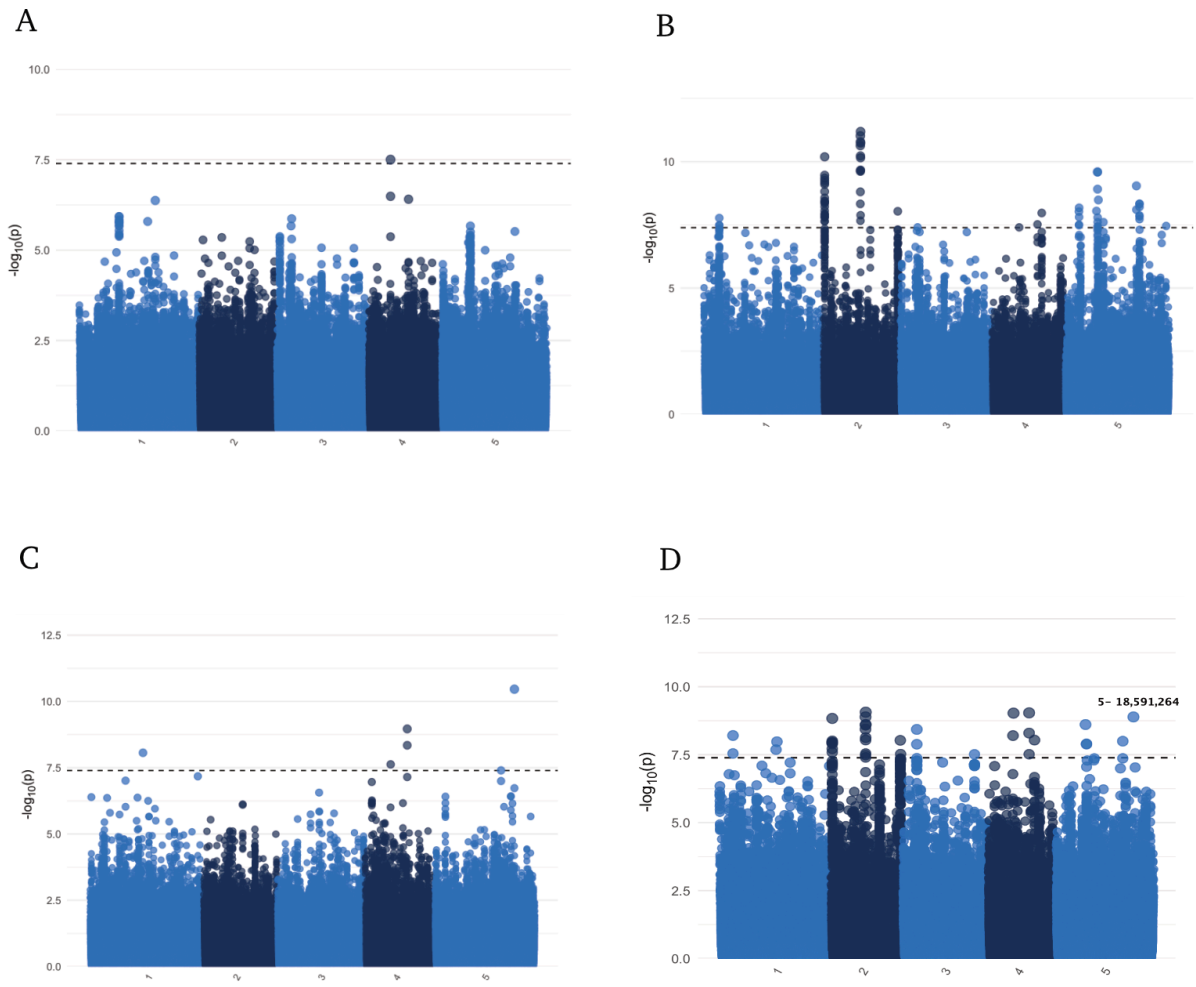


Figure S2.6. Manhattan plot displaying the association of 1.2M SNP markers with primary dormancy across trials. This plot presents the genome-wide association results for primary dormancy using 1.2 million SNP markers and three independent replicates of European *Arabidopsis thaliana* genotypes. (**A**) Results for Trial 1 of 295 lines, (**B**) Trial 2 of 361 lines, and (**C**) Trial 3 of 344 lines. Population structure was controlled for in the analysis using a kinship matrix between individuals. (**D**) The combined results from all three trials were calculated using Fisher's combined P-value method. A significant peak on chromosome 5 in panel (**D**) is annotated, which is located just 14 bp from the region of the *DELAY OF GERMINATION 1* (*DOG1*) gene, a well-known

regulator of seed dormancy. The dashed horizontal line indicates the significance threshold after a 5% Bonferroni correction was applied.

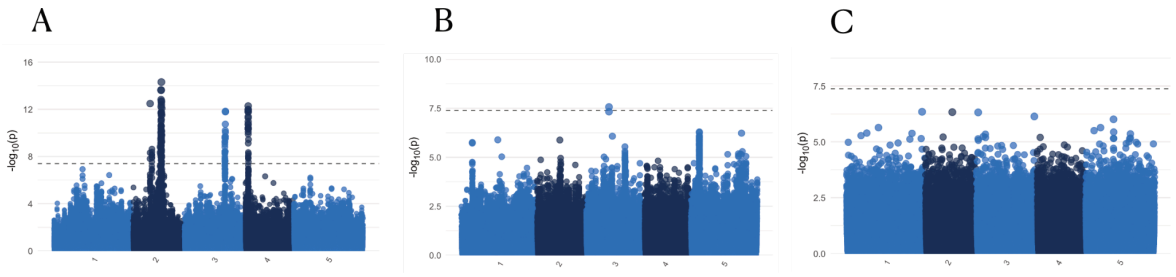


Figure S2.7. Manhattan plot displaying the association of 1.2M SNP markers with heat-induced secondary dormancy across trials. The genome-wide association results for heat-induced secondary dormancy using 1.2 million SNP markers across three independent replicates of European *Arabidopsis thaliana* genotypes. The analysis includes 295 lines in Trial 1 (**A**), 361 lines in Trial 2 (**B**), and 344 lines in Trial 3 (**C**). Primary dormancy was included as a covariate in the analysis, and population structure was accounted for using a kinship matrix between individuals. The dashed horizontal line represents the significance threshold after a 5% Bonferroni correction was applied. Significant loci are highlighted and indicate genomic regions associated with heat-induced secondary dormancy. So-called sdorm treatment represents germination rates following a 3-day stratification at 4°C to release dormancy, followed by a 4-day treatment at 37°C. The pdorm treatment tested germination rates without any pre-treatment, and the control treatment involved seeds stratified for 3 days at 4°C to release dormancy before testing germination. All germination tests were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days.

Table S2.1. Information of 361 studied *Arabidopsis thaliana* accessions. Table records Genotype ID (GenotypeID) according to ENA database and Genotype Name (GenotypeName) accordingly. Geographical origin (origin) and precise latitude and longitude of the genotype are recorded, together with extra information such as sequencer, collector, and CSS accession ID (AccessionID).

Available as “Table S1” at https://doi.org/10.6084/m9.figshare.28409279.v1

Table S2.2. Spearman correlation coefficients for heat-induced secondary dormancy across three trials. The Spearman correlation values quantifying the consistency of heat-induced secondary dormancy measurements across three independent trials. The secondary dormancy treatment represents germination rates following a 3-day stratification at 4°C to release dormancy, followed by a 4-day treatment at 37°C. The germination test was conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days. All germination tests were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days. Trial 1 included 295 genotypes and was conducted in May 2022; Trial 2 comprised the full set of 361 genotypes and was conducted in December 2022; Trial 3 included 344 genotypes and was conducted in December 2023.

	rho	p value
Trial 1 vs Trial 2	0.494	< 2.2e-16
Trial 2 vs Trial 3	0.409	3.661e-12
Trial 1 vs Trail 3	0.350	4.393e-09

Table S2.3. Spearman correlation of residual primary dormancy and heat-induced secondary dormancy of three trials. Secondary dormancy treatment represents germination rates following a 3-day stratification at 4°C to release dormancy, followed by a 4-day treatment at 37°C. The primary dormancy treatment tested germination rates without any pre-treatment. All germination tests were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days. Trial 1 is a set of 295 samples, Trial 2 has the complete set of 361 samples, and Trial 3 is a set of 344 samples. Trial 1 was performed in May 2022, Trial 2 in December 2022, and Trial 3 in December 2023.

	rho	**p value**
Trial 1	0.427	< 1.6e-14
Trial 2	0.204	< 8.98e-05
Trial 3	0.172	= 0.00135

Table S2.4. Descriptive statistics of germination rates under heat-induced secondary dormancy treatment across low- (<50°) and high-latitude (≥50°) regions in three trials. We used 50°N as a threshold to separate northern and southern populations based on our assumptions of the biogeographical transitions in Central and Northern Europe. Secondary dormancy treatment represents germination rates following a 3-day stratification at 4°C to release dormancy, followed by a 4-day treatment at 37°C. All germination tests were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days. Trial 1 is a set of 295 samples, Trial 2 has the complete set of 361 samples, and Trial 3 is a set of 344 samples. Trial 1 was performed in May 2022, Trial 2 in December 2022, and Trial 3 in December 2023.

	Low latitude	**High latitude**
Trial 1	mean = 0.0511, var = 0.0278	mean = 0.1808, var = 0.0826
Trial 2	mean = 0.1999, var = 0.1012	mean = 0.4193, var = 0.152
Trial 3	mean = 0.2546, var = 0.0611	mean = 0.3871, var = 0.0502

Table S2.5. Regression results of heat-induced secondary dormancy with four bioclimatic variables as predictors of genetic variation in germination after three treatments, across three trials. The results of regression models that assess the influence of four bioclimatic variables on secondary dormancy and its genetic variation. The models employed a binomial likelihood with a logit link function. The bioclimatic variables included BIO3 (isothermality), BIO9 (mean temperature of the driest quarter), BIO18 (mean precipitation of the warmest quarter), and BIO19 (mean precipitation of the coldest quarter). The analysis was performed across three treatments: primary dormancy (pdorm), secondary dormancy (sdorm), and control. Secondary dormancy was induced by a 4-day 37°C treatment following a 3-day stratification at 4°C to release primary dormancy. Primary dormancy was

tested without any pre-treatment, whereas control seeds were stratified at 4°C for 3 days before germination testing. The sdorm value used in this model represents secondary dormancy corrected for primary dormancy, calculated as the residual of the regression of secondary dormancy on primary dormancy. All germination tests were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days. Trial 1 is a set of 295 samples, Trial 2 has the complete set of 361 samples, and Trial 3 is a set of 344 samples. Trial 1 was performed in May 2022, Trial 2 in December 2022, and Trial 3 in December 2023. (A), (B), and (C) represent models with the control group as the baseline for Trial 1, Trial 2, and Trial 3, respectively; (D), (E), and (F) represent models in which primary dormancy is used as the baseline, for Trial 1, Trial 2, and Trial 3, respectively.

Available as "Table S5" at https://doi.org/10.6084/m9.figshare.28409279.v1

Table S2.6. Genome-wide association results of primary dormancy across three trials and shared genome-wide association peaks across three experimental trials. The primary dormancy treatment tested germination rates without any pre-treatment. All germination tests were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days. (A) Trial 1 is a set of 295 samples, (B) Trial 2 is the complete set of 361 samples, and (C) Trial 3 is a set of 344 samples. Trial 1 was performed in May 2022, Trial 2 in December 2022, and Trial 3 in December 2023. (D) Shared genome-wide association peaks of primary dormancy across all experimental trials, computed by Fisher's combined probability test.

Available as "Table S6" at https://doi.org/10.6084/m9.figshare.28409279.v1

Table S2.7. Genome-wide association results of heat-induced secondary dormancy across three trials and shared genome-wide association peaks across three experimental trials. The secondary dormancy represents germination rates following a 3-day stratification at 4°C to release dormancy, followed by a 4-day treatment at 37°C. All germination test s were conducted under long-day conditions at 20°C, and germination rates were recorded after 7 days. (A) Trial 1 is a set of 295 samples, (B) Trial 2 is the complete set of 361 samples, and (C) Trial 3 is a set of 344 samples. Trial 1 was performed in May 2022, Trial 2 in December 2022,

and Trial 3 in December 2023. (D) Shared genome-wide association peaks of heat-induced secondary dormancy across all experimental trials, computed by Fisher's combined probability test.

Available as "Table S7" at https://doi.org/10.6084/m9.figshare.28409279.v1

Table S2.8. Definitions of bioclimatic (BIO) variables used in the analysis and interpretation of their values. Descriptions of the nineteen bioclimatic (BIO) variables included in the generalized linear mixed model (GLMM). The ecological meaning of each BIO variable is provided, along with guidance on interpreting their values—particularly clarifying conditions represented by low and high values.

Bioclimatic variables	Definition	Interpretation
BIO1 Annual Mean Temperature	Mean of monthly temperature averages	High: warmer overall climate; Low: cooler year-round
BIO2 Mean Diurnal Range	Mean of monthly (max temp - min temp)	High: large day-night temperature shifts
BIO3 Isothermality	= (BIO2 / BIO7) × 100	High: more uniform temp across year (diurnal ≈ seasonal); Low: high seasonal contrast relative to daily variation
BIO4 Temperature Seasonality	Standard deviation of temperature × 100	High: large seasonal variability; Low: stable seasonal temps

BIO5 Max Temperature of Warmest Month	Highest monthly max temperature	High: hot extremes in summer
BIO6 Min Temperature of Coldest Month	Lowest monthly min temperature	Low: cold winters
BIO7 Temperature Annual Range	= BIO5 - BIO6	High: strong summer–winter contrast; Low: mild year-round
BIO8 Mean Temperature of Wettest Quarter	Average temp during 3-month wettest period	Ecologically varies; important for timing of growth or dormancy
BIO9 Mean Temperature of Driest Quarter	Average temp during 3-month driest period	High: warm/dry periods; Low: cold/dry stress
BIO10 Mean Temp of Warmest Quarter	Average temp during warmest 3 months	High: hot growing seasons
BIO11 Mean Temp of Coldest Quarter	Average temp during coldest 3 months	Low: cold stress potential
BIO12 Annual Precipitation	Total yearly precipitation	High: wet climates; Low: arid regions

BIO13 Precipitation of Wettest Month	Total precipitation in the wettest month	High: strong rainy season
BIO14 Precipitation of Driest Month	Total precipitation in driest month	Low: drought severity
BIO15 Precipitation Seasonality	Coefficient of variation of monthly precipitation	High: unpredictable rain patterns
BIO16 Precipitation of Wettest Quarter	Precipitation total in wettest 3-month period	High: monsoon-heavy systems
BIO17 Precipitation of Driest Quarter	Precipitation total in driest 3-month period	Low: drought-prone periods
BIO18 Precipitation of Warmest Quarter	Precipitation in the warmest 3 months	Low: hot/dry stress; High: wet growing season
BIO19 Precipitation of Coldest Quarter	Precipitation in coldest 3 months	Low: potential for overwintering stress

3

Non-additive Genetic Variance in Gene Expression in Natural Populations of *Arabidopsis lyrata*

3.1. Chapter Summary

Understanding the architecture of genetic variance is key to predicting how populations adapt and evolve. Since populations cannot respond effectively to natural selection if most of their genetic variance is non-additive, identifying the evolutionary and genomic factors driving non-additive variance in natural populations is critical. This chapter examines how genetic variance in gene expression shapes adaptive potential at the molecular level in natural populations of *Arabidopsis lyrata*. I quantified expression variance for over 26,000 genes and showed that genetic variance components were unevenly distributed across the genome and strongly structured by past selection and genomic context. Genes with high non-additive variance were found to exhibit lower synonymous diversity, reduced amino acid divergence, and stronger signatures of purifying selection compared to genes with high additive variance. By integrating classical quantitative genetics and modern molecular genetic data, this study provides a novel perspective on the role of non-additive genetic variance in shaping transcriptomic variation. Understanding these patterns will advance our knowledge of how genetic variance components contribute to evolutionary trajectories and adaptation in natural populations.

3.2. Fundamental Concepts

3.2.1. Decomposition of genetic variance components

The genetics of a quantitative trait is understood through its variation, which is measured as *variance*. The central idea in quantitative genetics is to partition the total variance of phenotypic values into components attributable to different causes. Classical quantitative genetics rests on the principle that, even without identifying specific loci, one can infer inheritance patterns and evolutionary potential by comparing phenotypic similarities among relatives and unrelated individuals. Greater resemblance among close relatives indicates a genetic contribution to trait variation. Most traits relevant to fitness are assumed to be influenced by many loci of small effect, a framework known as the *infinitesimal model* (**Fisher, 1919**), which has provided the foundation for predicting responses to selection for more than a century.

At the core of this framework lies the *decomposition of variance*. Phenotypic variance (V_P) can be partitioned into genetic and non-genetic components (**Figure 3.1**). Genetic variance (V_G) itself consists of *additive* (V_A) and *non-additive* (V_{NA}) components; the latter can be further partitioned into dominance (V_D), arising from interactions between alleles at a locus, and epistasis (V_I), arising from interactions across loci (**Figure 3.2; [3.1-3.3]**).

First and foremost, it is important to *distinguish between genetic effects and genetic variance*, a distinction often overlooked outside quantitative genetics. *Effects* describe how alleles influence phenotype; *variance* reflects how much of the phenotypic variation in a population is attributable to those effects, which depends not only on the genetic architecture but also on allele frequencies. The same genetic effect may contribute little or greatly to variance depending on how common the allele is in the population (**Huang & Mackay, 2016**).

Additive genetic variance is the component of genetic variance due to the additive effects of the individuals bearing them. Provided that one can assume random mating, additive variance is not difficult to measure directly on a given population. It is the component transmitted predictably from parents to offspring which makes it the key driver of heritable resemblance. By contrast, *non-additive variance* cannot be predicted from the combined additive effects of the genotype's collective alleles; it arises when allele effects depend on genetic background. *Dominance variance* reflects phenotypic outcomes that deviate from additive expectations at a locus; for example, two populations of identical genotypic composition will exhibit different genetic variances if they differ in the degree of dominance. *Epistasis variance* arises due to epistatic interactions among alleles at different loci, in which the expression of a genotype depends on the genotype at one or more other loci. Because recombination breaks up such interactions, they are less reliably transmitted across generations. For this reason, heritable variation in the context of evolutionary response is usually equated with additive genetic variance.

$$V_P = V_G + V_R = V_A + V_{NA} + V_R \quad \textbf{[3.1]}$$

$$H^2 = V_G / V_P \quad \textbf{[3.2]}$$

$$h^2 = V_A / V_P \quad \textbf{[3.3]}$$

With V_P being the total phenotypic variance, V_G being genetic variance component and V_R being environmental variance. V_G is further partitioned into additive (V_A) and non-additive variance (V_{NA}) – consisting of dominance and epistasis deviations.

Therefore, most empirical studies focus on additive variance, while the remainder is grouped as residual variance (V_R), largely attributed to environmental factors. The proportion of phenotypic variance explained by additive variance is termed *narrow-sense heritability* (h^2) (**Figure 3.2; [3.3])**, which captures the efficiency of

selection. A related measure, *broad-sense heritability* (H^2) (**Figure 3.2; [3.2])**, reflects the total genetic contribution, including dominance and epistasis. Broad-sense heritability is often of theoretical interest and in selfing species (in plants), but not for breeding regimes which require crossings, segregation, and combinations of genotypes. In general, narrow-sense heritability is more widely used in breeding programs, and equally relevant in natural populations where resemblance among relatives determines the potential for selection to drive evolutionary change (**Falconer & Mackay, 2009**).

Additive variance has thus remained the dominant focus in both evolutionary and applied breeding studies, as it provides the best proxy for a population's short-term adaptive potential. However, non-additive variance is not merely "noise." Theory and empirical evidence show that dominance and epistasis can shape trait architecture, buffer phenotypes, and store cryptic genetic variation that may be released under novel conditions (**Paaby & Rockman, 2014**). For example, during population bottlenecks, genetic drift can disrupt allele combinations, converting non-additive into additive variance. This process releases previously hidden additive variation from dominance and epistatic interactions, potentially facilitating evolutionary change despite reduced population size (**Neiman & Linksvayer, 2006; van Heerwaarden et al., 2008**).

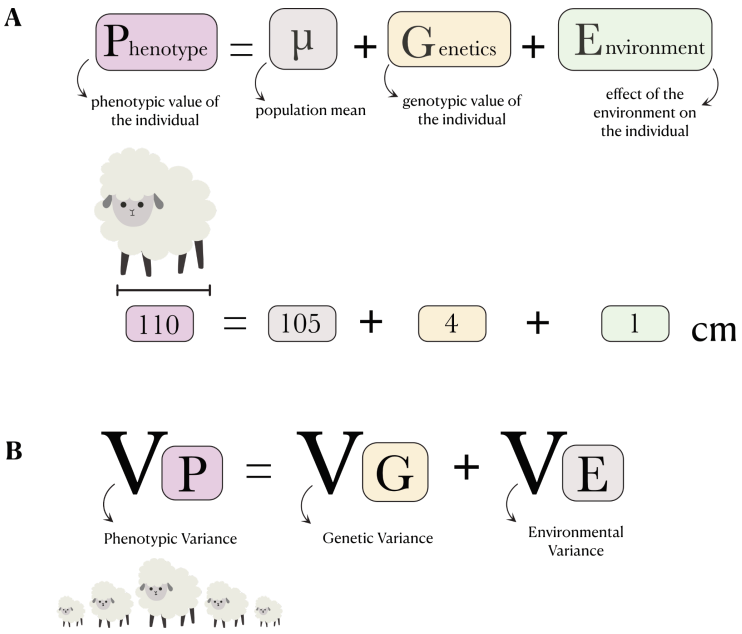


Figure 3.1. Decomposition of phenotype into genetic and environmental components at the individual and population level. (A) The phenotype of an individual is expressed as the sum of the population mean, the genotypic value, and the environmental effect. **(B)** At the population level, the same principle extends to variances: the total phenotypic variance (V_P) can be partitioned into genetic variance (V_G) and environmental variance (V_E). For simplicity, this equation assumes the covariance between genetics and environment is zero.

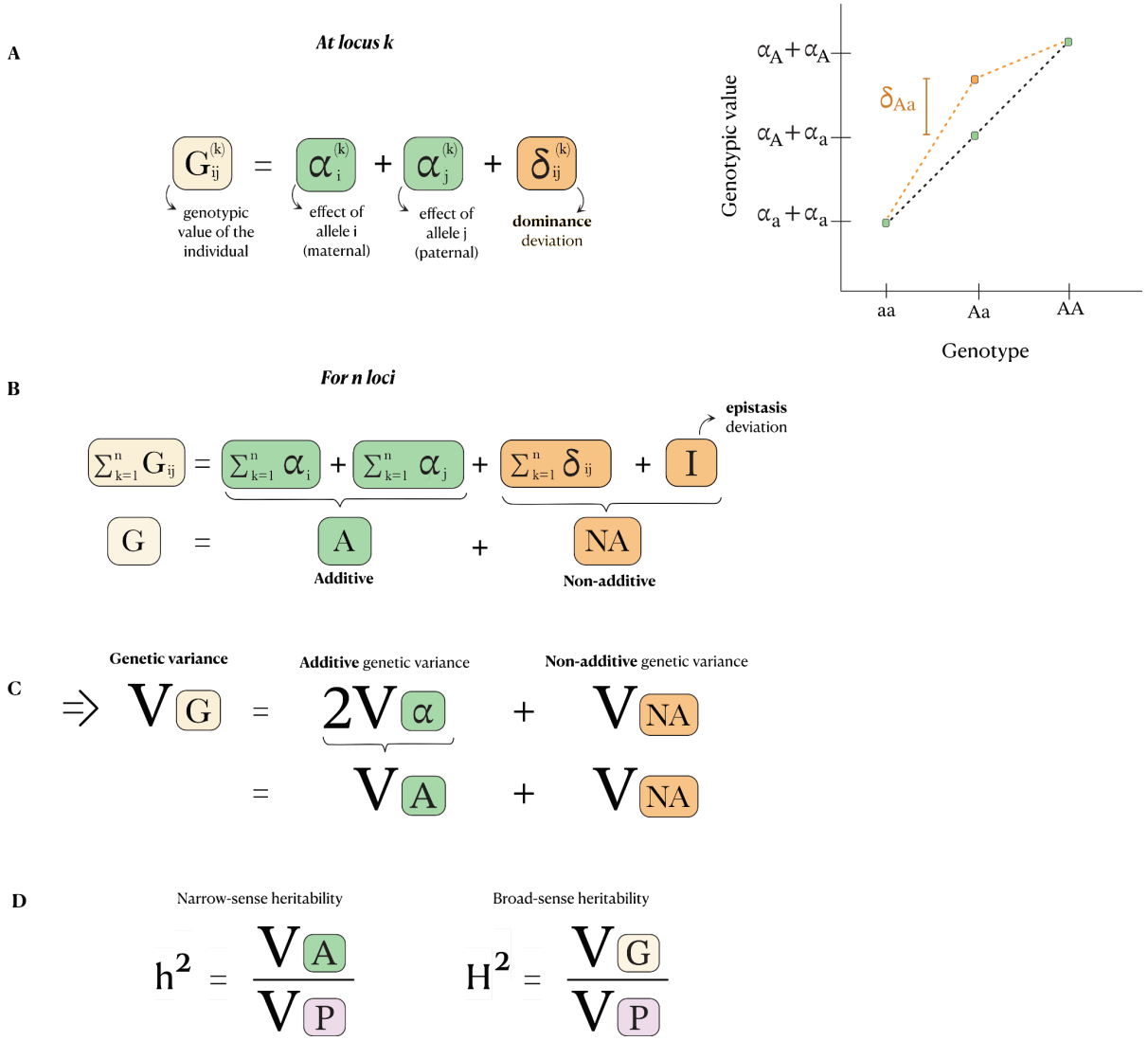


Figure 3.2. Decomposition of genotypic values into additive and non-additive components across loci. (A) At a single locus, the genotypic value of an individual can be partitioned into the average effect of alleles and a dominance deviation, illustrated on a genotype–phenotype map. **(B)** Across multiple loci, the total genotypic value is the sum of additive effects and non-additive contributions (dominance and epistasis). These variance components, therefore, are dependent on the distribution of allele frequencies. **(C)** This decomposition extends to genetic variance: the total genetic variance (V_G) can be partitioned into additive variance (V_A), and non-additive variance (V_{NA}), arising from dominance and epistatic interactions

3.2.2. Pedigree and animal model

Estimating quantitative genetic parameters in natural populations ultimately requires statistical approaches that relate phenotypic resemblance among individuals to their genetic relatedness. **Figure 3.3** gave a simplified example of how comparing the genetic makeup of the relatives can give information for the estimation of different variance components. From a parent-offspring perspective, they share half of additive variance and none of dominance variance. This can be extrapolated to other kinds of relatives. Full-sibs, for example, share additionally one fourth of dominance variance. Mathematically, combining different types of relatives in a pedigree allows us to estimate different components of genetic variance. For instance, excess similarity between full-sibs compared to half-sibs can be used to estimate dominance variance, since half-sibs share a quarter of additive variance but none of dominance variance. Similarly for excess similarity between full-sibs compared to parent-offspring, as in addition to half of additive variance, full-sibs also share one fourth of dominance variance. This picture will be fully presented in **Chapter 4** that takes into account also the epistatic components.

Early methods such as parent–offspring regressions and ANOVA-based designs using full-sib or half-sib families are not practical in wild populations due to unbalanced pedigrees and uncontrolled breeding structures (**Falconer & Mackay, 2009; Walsh & Lynch, 2018b**). The development of the animal model has transformed this field. Originating in animal breeding (**Henderson, 1953, 1976**), it has become the method of choice in evolutionary ecology because it efficiently uses all available information in complex, natural pedigrees. Unlike traditional methods, it can incorporate multiple types of relationships, tolerate missing data (e.g. uncertain paternity, unmeasured phenotypes), and account for non-genetic effects such as sex, age, or environment (**Wilson et al., 2010**).

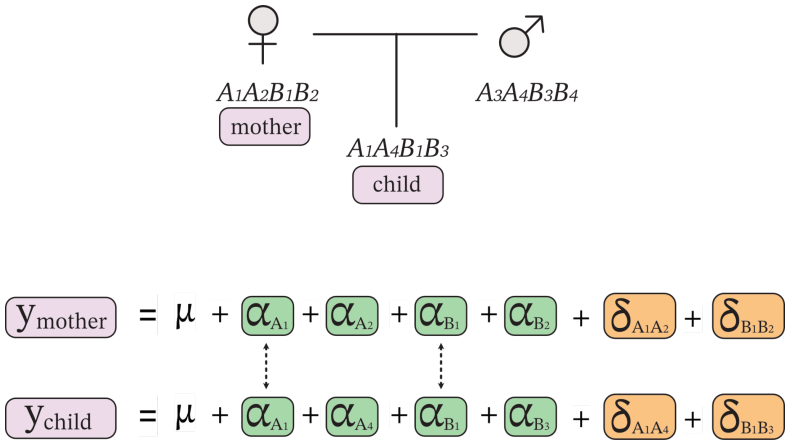


Figure 3.3. Inheritance of a two-locus trait, simplified to only additive and dominance effects. Two individuals mate to produce an offspring, which inherits alleles A1 and B1 from the mother and alleles A4 and B3 from the father. The genotypic values of the mother and of the offspring are presented. The dashed arrows indicate the shared effects. This shows that only half the additive effects are shared between a parent and the offspring, and none of the dominance effects. Also note that the additive effects of the offspring are inherited from the two parents, while none of the dominance effects are inherited.

In practice, the animal model is a *linear mixed-effects model*. The phenotypic value of an individual is expressed as the sum of the population mean, the additive genetic contribution (*breeding value*), any non-additive genetic contribution (e.g. dominance and/or epistasis), and a residual term (**Figure 3.4**). Each of these terms is treated as a random effect with its own variance component, such that the total phenotypic variance can be partitioned into additive, non-additive, and residual components. Importantly, additive genetic values are not independent across individuals but are correlated according to their relatedness, quantified by the *coefficient of coancestry*. Similarly, non-additive values may be correlated through shared genotype configurations, described by the *coefficient of fraternity*.

As mentioned earlier, the knowledge of the pedigree is required so that relatedness among individuals can be represented by the additive relationship matrix and any additional, such as dominance matrix. These matrices define the expected covariances in the respective values. This study made use of a pedigree of full-sibs and half-sibs: additive variance (V_A) can be estimated from half-sib families because half-sibs share, on average, 25% of their genes by descent through a common parent, and this covariance reflects only additive effects. By contrast, estimating dominance variance (V_D) requires full-sib families, since siblings that share both parents capture not only additive similarity but also additional covariance due to interactions between alleles at the same locus. Thus, while half-sibs provide information about V_A alone, full-sibs are needed to disentangle V_A and V_D.

It is noteworthy that in practice, partitioning non-additive variance into dominance (V_D) and epistasis (V_I) is particularly challenging. For this reason, **in this chapter, I only use the dominance matrix despite referring to this component as V_{NA}.** A more detailed background and attempts on partitioning epistasis variance component is provided in the **Chapter 4**.

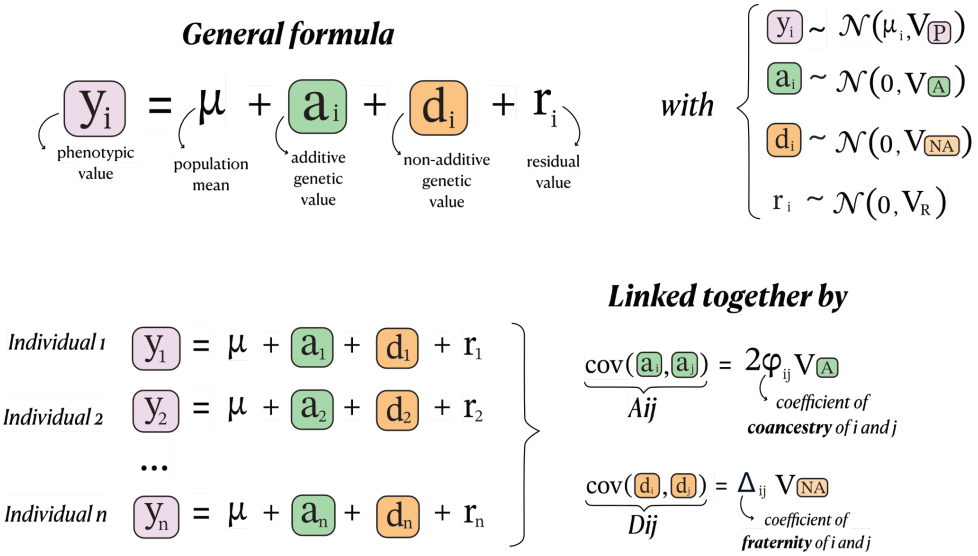


Figure 3.4. General formulation of the animal model. The phenotype of an individual (y_i) is expressed as the sum of the population mean (μ), the additive genetic value (a_i), non-additive genetic value (d_i), and a residual component (r_i). Each term is assumed to follow a normal distribution with mean 0 and variance corresponding to the relevant variance component (V_A, V_{NA}, V_R), such that total phenotypic variance is V_P. Across individuals, additive effects are correlated according to their relatedness, quantified by the coefficient of coancestry (ϕ_{ij}), while non-additive effects are linked by the coefficient of fraternity (Δ_{ij}). By modelling these covariance structures, the animal model partitions phenotypic variance into its genetic and environmental components.

3.2.3. On the πN/πS and Ka/Ks metrics

Quantitative and population genetics are deeply intertwined, as the analysis of quantitative traits is always carried out in a population context (**Caballero, 2020**). While quantitative genetics partitions phenotypic variance into its genetic and environmental components, population genetics provides the framework for understanding how evolutionary forces shape the underlying genetic variation. Mutation, selection, drift, and gene flow act on allele frequencies, determining how genetic diversity is structured across the genome and maintained through time. This perspective has been essential for explaining molecular diversity, the maintenance of polymorphism, and the probability of allele fixation or loss. Ultimately, allele frequencies and their distribution across loci influence the genetic variance in traits. Ronald A. Fisher's *Fundamental Theorem of Natural Selection* (**Fisher, 1930)** formalized this link by showing, in mathematical terms, that the rate of increase in mean fitness under natural selection is proportional to the additive genetic variance in fitness.

In population genetics, selection is often introduced with fundamental regimes - positive selection, purifying selection, and balancing selection: *Positive selection*

favors the spread of beneficial mutations, while *balancing selection* maintains genetic diversity through mechanisms such as heterozygote advantage or frequency dependence; by contrast, *purifying selection* - sometimes referred to as negative selection - is the most pervasive form of selection in genomes, acting consistently to eliminate deleterious mutations and preserve the integrity of functional sequences (**Wills, 2007**). Because most new mutations are harmful rather than advantageous, the evolutionary fate of much of the genome is shaped less by adaptive innovation than by this constant pruning.

To connect these evolutionary principles to observable data, population genetics makes use of measures of genetic diversity at the nucleotide level. A common measure is *nucleotide diversity* (π), which quantifies the average number of pairwise differences among DNA sequences in a population (**Kimura, 1968**). This can be calculated across all sites, or specifically on *synonymous* (πS) and *non-synonymous* (πN) mutations. *Synonymous* changes typically do not alter amino acids and are therefore often neutral, whereas *non-synonymous* changes modify protein sequences and are more likely to disrupt function (**Figure 3.5**). The ratio of non-synonymous to synonymous diversity ($\pi N/\pi S$) indicates the relative constraint on protein-coding sequences: values much less than one reflect the pervasive action of purifying selection, whereas values closer to or above one suggest neutrality or positive selection.

At a broader timescale, divergence between species can be assessed using the analogous comparison of substitution rates, with *Ka* representing non-synonymous divergence and Ks representing synonymous divergence. In other words, the *Ka/Ks ratio* compares the rate of nonsynonymous to synonymous substitutions per site between orthologous coding sequences (**Nei and Gojobori, 1986**). *Ka/Ks ratio* is a cornerstone measure in molecular evolution: a value much less than one again signals purifying selection, while values above one indicate adaptive change.

Because many mutations within populations remain polymorphic rather than fixed, closely related or within-species *Ka/Ks* can be difficult to interpret. To address this limitation, combining polymorphism (πN/πS) with divergence (*Ka/Ks*) can offer a more powerful framework for distinguishing neutral from selective processes across different evolutionary timescales.

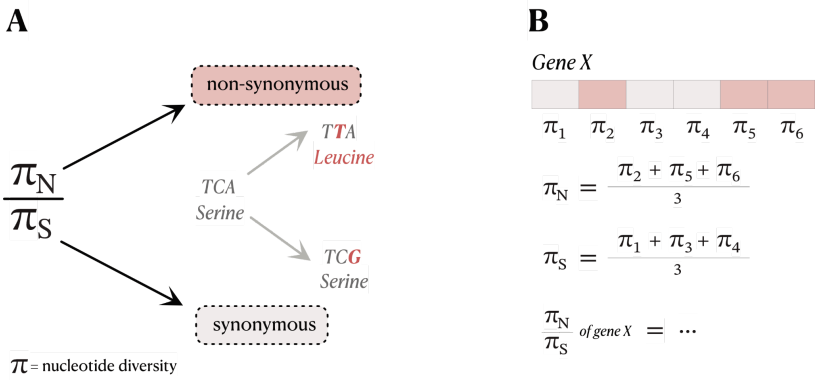


Figure 3.5. Illustration of the πN/πS ratio as a measure of selection on coding sequences. **(A)** Nucleotide diversity (π) can be partitioned into nonsynonymous (πN) and synonymous (πS) components depending on whether codon changes alter the encoded amino acid. For example, the codon TCA (Serine) can mutate to TCG (synonymous, still Serine) or TTA (nonsynonymous, Leucine). **(B)** Within a gene, π is estimated for each site, and averages across nonsynonymous and synonymous sites yield πN and πS, respectively. Their ratio (pN/pS) reflects the strength and direction of selection: values << 1 indicate purifying selection, ≈ 1 neutrality, and > 1 positive selection.

3.2.4. Allele frequencies

Because population genetics relies on statistical models that treat variation in terms of expectations and uncertainty, genetic data are typically summarised into population-level statistics rather than analysed site by site. One widely-used statistic is the *site frequency spectrum* (SFS) (**Fu, 1995**). The SFS is essentially a

histogram that records how many genetic variants occur at each possible allele frequency in a sample (**Figure 3.6**). In diploid organisms, the number of possible frequency categories equals twice the number of sampled individuals. For instance, sequencing five diploid individuals yields ten possible allele-frequency bins. By capturing the distribution of allele frequencies, the SFS provides a compact yet powerful summary of genetic variation across the genome, and is widely used to infer demographic history, mutation rates, and the action of natural selection.

The way an SFS is constructed depends on which allele is counted. If the evolutionary history of the species is unclear, the *minor allele frequency* (MAF) is typically used. At a biallelic SNP, the minor allele is the one occurring less often in the sample. Using this allele produces a *folded* SFS, which avoids making assumptions about ancestry. If ancestral states can be identified, one can distinguish between ancestral and derived alleles. In that case, the SFS is based on the derived allele frequency, giving an *unfolded* SFS that more directly represents the distribution of new mutations under coalescent theory.

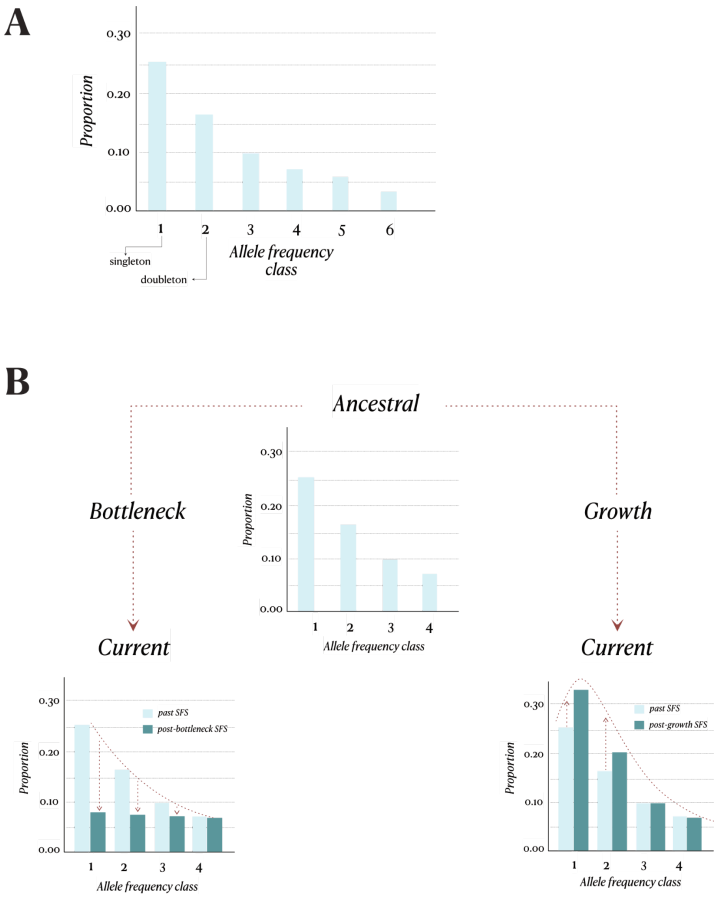


Figure 3.6. Illustration of site frequency spectrum (SFS) and the impact of demographic history. (A) The 'realistic' one-dimensional SFS (1D-SFS) for a population typically shows an exponential decline across allele frequency classes, with rare variants being the most common. The first frequency classes are often called singletons (alleles observed once) and doubletons (alleles observed twice). **(B)** Demographic processes distort this expected distribution. Before a bottleneck, variants follow the exponential SFS shape; after a bottleneck, the disproportionate loss of rare variants flattens the distribution. By contrast, population growth promotes the retention or appearance of new mutations, generating an excess of rare alleles relative to the ancestral expectation.

3.2.5. Mutations and distribution of fitness effects

The works of Fisher, Wright, and Haldane have also established that the path to evolution depends, to a large extent, on mutations; and the impact of new mutations on evolution is not determined solely by their frequencies, but by how they affect fitness once they arise (**Charlesworth et al., 2017; Chen et al., 2022**). The effects of mutations can be broadly divided into three categories: harmful – negative effect on survival and fecundity, neutral – no effect on fitness, and advantageous – positive effect on adaptation and fitness. However, in reality, mutations span a spectrum of selective effects from strongly deleterious, through weakly deleterious, to neutral and to mildly or highly beneficial. In nature, most mutations are either deleterious or effectively neutral, with strongly deleterious variants quickly removed by purifying selection and strongly beneficial ones extremely rare (**Eyre-Walker & Keightley, 2007**). Effectively neutral mutations occupy the portion of the distribution where selective effects are so weak that they are indistinguishable from strictly neutral variants in their evolutionary behaviour. Their fate is therefore shaped by drift rather than deterministic selection (**Ohta, 1992**). Capturing this picture is the role of methods that compute the *distribution of fitness effects* (*DFE*) of new mutations. DFE provides a framework for understanding the nature of quantitative genetic variation and predicting evolutionary consequences under different population dynamics.

Distribution of fitness effects (*DFE*) methods use polymorphism data within species, they model the *site frequency spectrum* (SFS) at neutral versus selected sites (e.g. synonymous vs. nonsynonymous) to infer the frequency distribution of selection coefficients, spanning strongly deleterious to advantageous mutations Although estimating the fraction of beneficial mutations remains difficult due to their rarity, DFE-based methods manage to reflect the mutations currently segregating in

populations and have been applied across plant and animal systems (**Castellano et al., 2019; Chen et al., 2017; Galtier, 2016; Moutinho et al., 2020**).

The fitness effect of a new mutation influences the frequency at which it is expected to segregate in a population, which is affected by both demography and selection. The fitness effect of a new mutation influences the frequency at which it segregates in a population, a pattern shaped jointly by demography and selection. Although synonymous mutations are often treated as approximately neutral, and thus their SFS primarily reflects demographic history, selection on synonymous sites can occur. In contrast, the SFS of non-synonymous mutations integrates both demographic processes and purifying selection. By comparing these two SFS, it is therefore possible to disentangle demographic effects from selection on protein-changing sites.

The efficacy of selection acting on a mutation depends on the scaled selection coefficient, $N_e s$. When $N_e s \gg 1$, selection is efficient, whereas when $N_e s \ll 1$, the fate of mutations is dominated by genetic drift. Thus, in small populations many weakly deleterious mutations behave as if they were neutral, a central idea of Ohta's nearly neutral theory of molecular evolution (**Ohta, 1992**). Although N_e sets the baseline for how effectively selection can act, namely, larger populations can purge deleterious mutations of smaller effect, differences in the distribution of fitness effects (DFE) also influence selection efficacy. Species with similar N_e can nevertheless differ in how efficiently they remove deleterious mutations or fix beneficial ones if their DFEs differ in shape or scale (**Boyko et al., 2008; Eyre-Walker & Keightley, 2007**).

The efficiency of selection acting on a mutation depends on the effective population size. This provides population-scale selection coefficient $N_e s$ – the product of effective population size N_e and selection coefficient s. If this product is

much greater than one, selection is effective. If this product is much less than one, the fate of mutations is largely subject to random genetic drift. Essentially, in small populations, mutations are largely affected by drift and behave like neutral ones, which is the premise of the nearly neutral theory of molecular evolution. *Ne* sets the baseline for selection efficacy, in a way that higher Ne allows for effective selection against deleterious mutations of smaller effect, compared to populations with lower Ne. However, DFE modifies how selection acts on mutations: species with the same *Ne* can still differ in how effectively they remove deleterious mutations or fix beneficial ones.

The distribution of fitness effects (DFE) is most commonly modeled using a gamma distribution (**Figure 3.7**). This distribution is described by *a shape parameter* (*k*) and *a scale parameter* that together determine its mean. Importantly, the same mean can produce very different distributions depending on the value of k. When k is large, the gamma distribution becomes narrow and spike-like, reflecting a scenario where most mutations have similar selection coefficients. By contrast, smaller values of k yield broader, more skewed distributions, where most mutations have small effects but a few contribute disproportionately large effects. In such cases, the high kurtosis (peakedness) reflects that the variance is driven largely by these rare, large-effect mutations (**Eyre-Walker & Keightley, 2007**).

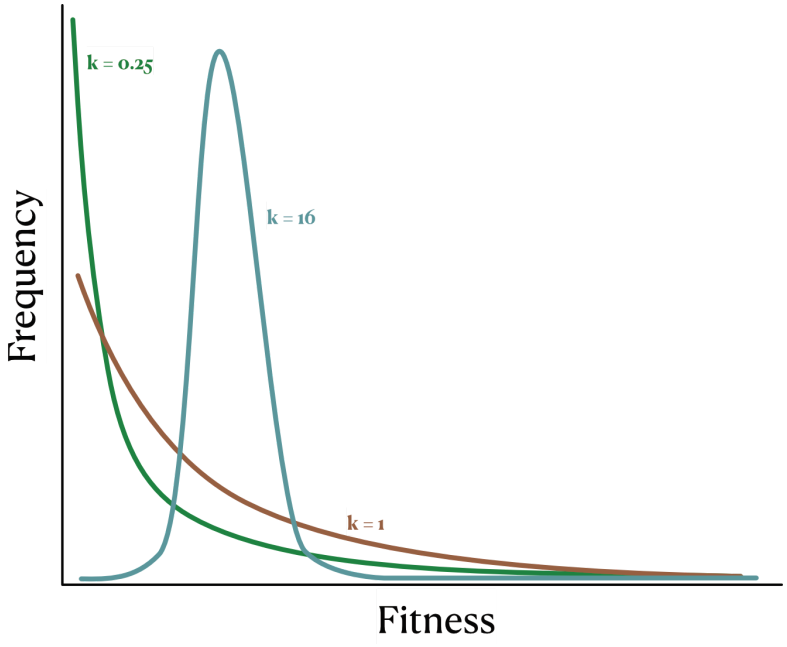


Figure 3.7. Gamma distribution is the most common distribution to model the distribution of fitness effects. The distribution has two parameters, shape parameter k and a parameter which determines the mean of the distribution. When k < 1, the distribution is L-shaped, when k = 1, the distribution is close to a skew-normal distribution, and when k > 1, it resembles a normal distribution. Varying shape parameters can shape the probability density of the distribution differently, given the same mean. The larger the k is, the more "spiky" the distribution is, in which all mutations have the same strength of selection. *Adapted from Eyre-Walker and Keightley* (2007).

3.3. Introduction

3.3.1. Genetic variance and its role in adaptive potential

The ability of natural populations to adapt to changing environments is governed by genetic and phenotypic variation. Genetic variance, the foundation of evolutionary processes, is shaped by complex allele interactions which determine the fraction of genetic variance accessible to selection. Additive genetic variance is directly

heritable and contributes to predictable evolutionary responses, whereas non-additive genetic variance, arising from dominance and epistasis, is less easily captured by selection (**Fisher, 1919**). The balance between these components is critical in determining a population's adaptive capacity, particularly under rapid environmental change (**Walsh & Lynch, 2018b**).

Theoretical studies suggest that natural selection depletes additive genetic variance over time, potentially increasing the proportion of non-additive variance (**Turelli, 1988**). However, empirical evidence remains inconsistent. Studies in maize (*Zea mays*) have shown that non-additive variance significantly contributes to yield traits (**Yang et al., 2019**), whereas in *Drosophila serrata*, non-additive variance was unrelated to fitness components in wing traits (**Sztepanacz & Blows, 2015**). Similarly, studies in plants have shown that additive and non-additive variance components vary depending on trait types (**Kumar et al., 2015; Seymour et al., 2016; Xu, 2013)**.

Processes such as population bottlenecks, genetic drift, and admixture also shape genetic variance by altering allele frequencies and inter-locus interactions. Bottlenecks tend to increase the proportion of additive variance, as genetic variation at interacting loci is lost (**Waldmann, 2001; Willis & Orr, 1993**). Admixture introduces novel allele combinations, potentially enhancing both additive and non-additive variance (**Rieseberg et al., 1999; Vasseur et al., 2019**). Additionally, the genetic architecture of a trait, such as the number and effect sizes of loci involved, can influence the amount of additive or non-additive variance (**Mäki-Tanila & Hill, 2014**).

The impact of dominance variance on heritability estimates in wild populations remains an area of debate. **Class and Brommer (2020)** investigated how ignoring dominance variance affects estimates of additive variance and thus heritability in blue tits (*Cyanistes caeruleus*). Their study found that dominance variance

constituted a minor fraction of total genetic variance, leading to only a slight overestimation of additive variance when dominance variance was excluded. However, this overestimation became more pronounced when maternal effects were also omitted, highlighting the importance of accounting for maternal contributions in quantitative genetic analyses.

3.3.2. The dynamics of genetic variance

Selection alters genetic variance components through both short-term linkage disequilibrium (LD) effects and long-term processes such as recombination and mutation-selection balance. In the short term, selection reduces additive variance due to LD, where favored allele combinations become non-randomly associated, restricting independent segregation (**Walsh & Lynch, 2018b**). Over generations, recombination breaks LD, allowing partial recovery of additive variance, while in the long term, mutation-selection balance stabilizes genetic variance. In contrast, dominance variance does not accumulate through LD in the same way, as it depends on heterozygote effects rather than additive allele contributions.

Epistasis plays a key role in shaping the persistence or depletion of genetic variance (**Hemani et al., 2013**). While additive-by-additive epistasis maintains genetic variation by preserving polymorphism at interacting loci, it does not necessarily maintain additive variance because the contributions of alleles depend on genetic background rather than independent additive effects (**Mackay, 2014**). In contrast, disruptive selection can amplify additive variance if epistatic interactions promote genetic correlations between loci, increasing overall trait variance. Hence, the *Bulmer Equation* predicts short-term declines in additive variance due to LD but highlights the importance of recombination and epistasis in long-term variance maintenance and evolutionary responses to selection (**Bulmer, 1971**). This is indeed central to *Wright's Shifting Balance Theory*, which proposes that populations can evolve adaptive gene complexes - sets of alleles that work well together in a given

genetic background but perform poorly when recombined with alleles from other populations (**Wright, 1948**). Because of this, inter-population fitness reduces when hybridization occurs, particularly when epistasis strongly influences allele effects - this will be further explored in **Chapter 4**.

3.3.3. The evolutionary paradox of additive genetic variance maintenance

An ongoing question in evolutionary and quantitative genetics is the persistence and loss of additive genetic variance over generations. The presence of non-additive effects like epistasis and dominance can maintain or even increase additive genetic variance under certain conditions. Epistasis, in particular, allows deleterious mutations to persist and can convert non-additive variance into additive variance, thereby sustaining genetic variation under selection (**Hallander & Waldmann, 2007**).

Although this hypothesis of genetic variance components conversion was supported by selection experiments in model organisms (e.g., *Drosophila and E. coli*) (**Carson & Wisotzkey, 1989; López-Fanjul & Villaverde, 1989; Wiser et al., 2013**), theoretical models often predict a rapid decline in additive variance and simulation studies could not support this mechanism (**Duenk et al., 2020; Esfandyari et al., 2017**). Empirical studies also support that population bottlenecks can increase additive genetic variance from non-additive sources, although this does not always translate into stronger responses to selection (**Jarvis et al., 2011; Neiman & Linksvayer, 2006; van Heerwaarden et al., 2008**). These dynamics also raise concerns about the long-term stability of genomic prediction, the methodology that uses genome-wide marker effects to predict phenotypes, because allele-frequency shifts alter average quantitative trait loci (QTL) effects and hence genomic relationships through time (**Meuwissen et al., 2001**). Despite efforts to model genetic variance dynamics, the persistence of additive variance in real

populations remains an unresolved evolutionary paradox, with important consequences for breeding strategies and the reliability of genomic prediction.

Such discrepancy between simulations and real populations remains a critical gap in our understanding, potentially indicating that existing models may have missed key biological or evolutionary contexts. A more complete understanding of adaptive potential, therefore, requires renewed attention to this neglected dimension of quantitative genetics, the gap that will be explored in this thesis.

3.3.4. The evolutionary potential of gene expression

Traditional quantitative genetics has primarily relied on phenotypic data to estimate genetic variance and predict evolutionary responses. While this approach has been instrumental in understanding heritability, selection, and trait evolution, it does not fully resolve many questions. Potential explanations may lie at the molecular level, where gene expression variation and regulatory networks may contribute to genetic variance in ways that traditional phenotype-based approaches cannot fully capture. By integrating molecular quantitative genetics, particularly through transcriptomics and regulatory variation, researchers can gain deeper insights into how genetic variance is maintained, how selection acts on gene expression, and how cryptic variation can be uncovered under novel environmental conditions. Given that many adaptive traits are polygenic and regulated through complex gene networks, studying transcriptional variation alongside traditional phenotype-based models may reveal previously hidden sources of heritable variation, ultimately refine evolutionary predictions and improve our understanding of adaptive potential in changing environments.

Over the past 20 years, many studies have highlighted the importance of gene regulatory networks and gene expression variation in understanding genetic architecture of complex traits or their evolutionary potential (**Ayroles et al., 2009;**

Ciliberti et al., 2007; Crombach & Hogeweg, 2008; Davidson & Erwin, 2010; Erwin & Davidson, 2009; Groen et al., 2020; Kittelmann et al., 2018). On a large scale, transcription is commonly used as a proxy for gene expression (**Lockhart & Winzeler, 2000; Schena et al., 1995; Leder et al, 2015**). Although transcription is only the first step in gene expression, even small variations in mRNA abundance can significantly impact protein levels (**Bar-Even et al., 2006; Ghazalpour et al., 2011; Kærn et al., 2005**); studies show that variation in protein abundance can be attributed to transcriptional differences, depending on taxa, cellular location, and protein function (**Greenbaum et al., 2003; Leder et al., 2015**). It is noteworthy that these estimates may be conservative because gene-regulatory processes often saturate, meaning that linear variance-partitioning approaches can underestimate the true influence of transcription on downstream protein levels (**Cai & Des Marais, 2023**). Therefore, transcription can nevertheless serve as a useful proxy for gene expression, and transcriptional variation itself is likely a key driver of phenotypic diversity and evolutionary change.

From a classical quantitative genetics perspective, transcriptional variation can be treated as a phenotype and, like any phenotype, is influenced by both genetic and environmental factors, with its standing variation shaped by natural selection. Gene expression is generally considered to be highly polygenic and largely additive, with many loci contributing small effects to transcript abundance (**Gilad et al., 2008; Hajheidari et al., 2025; Kim & Gibson, 2010**) (**Figure 3.8**). To date, transcriptomic data have been used predominantly in the context of expression quantitative trait locus (eQTL) mapping, which has proven powerful for identifying genomic regions associated with variation in gene expression in a wide range of systems, including mice (**Bao et al., 2006**), yeast (**Brem et al., 2002**), *Caenorhabditis elegans* (**Li et al., 2006**), and fish (**Pritchard et al., 2017**). However, beyond their use in eQTL mapping and studies of plasticity, transcriptomic data also provide a powerful opportunity to

quantify genetic variance components at the molecular level. By treating gene expression as a quantitative trait, it becomes possible to partition transcriptional variance into additive and non-additive components, thereby linking molecular variation directly to quantitative genetic theory. Intriguingly, little study to date has effectively bridged modern genetics with the classical fields to provide a comprehensive explanation of adaptation at the molecular level. To my knowledge, **Leder et al. (2015)** and **Tsouris et al. (2024)** were among the first to report the relevance of non-additive genetic components to population-wide gene expression variation. While quantitative genetics has long relied on phenotypic variance to estimate genetic parameters, integrating molecular-level variation, such as transcriptional regulation and gene expression networks, offers an unprecedented opportunity to refine our understanding of how genetic variation is maintained and how selection acts on regulatory mechanisms.

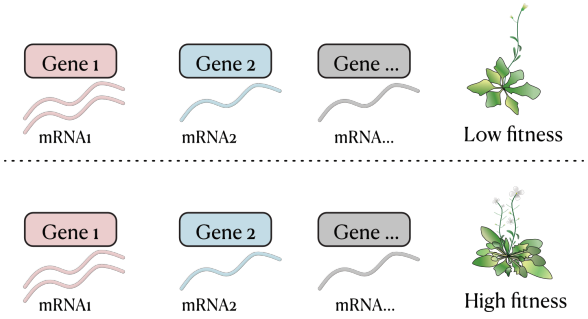


Figure 3.8. Variation in gene expression contributes to fitness. Differences in the expression levels of genes, reflected in the abundance of their corresponding mRNAs, influence phenotypic traits and ultimately affect individual fitness. Certain gene expression combinations can result in reduced fitness, whereas others enhance fitness.

Moreover, a major missing link remains between heritability at phenotypic level and at the molecular level. Traditionally, quantitative genetic studies have focused on partitioning phenotypic variation into additive, dominance, and environmental components; they often lack direct insight into how specific molecular processes

drive these variance patterns. On the other hand, molecular studies provide rich datasets on gene expression, regulatory elements, and sequence variation, yet they frequently do not quantify how these contribute to heritable variation at the population level. To obtain accurate estimates of genetic and environmental variance, careful experimental design is required, and outcrossing species provide a practical advantage because they permit the generation of controlled pedigrees or replicate genotypes while maintaining genetic diversity. Understanding how genetic variance at the transcriptional level translates into phenotypic variability is critical for predicting adaptive potential in natural populations.

3.3.5. Research questions

In this chapter, I asked three main questions:

- How heritable is transcriptional variation in natural populations of *Arabidopsis lyrata*?
- What is the link between genetic variance components at the molecular level co-vary and those observed at the phenotypic level?
- How has selection shaped transcriptional variation in *Arabidopsis lyrata*?

3.4. Data and Methodology

3.4.1. Plant material preparation

Nine plants from an *Arabidopsis lyrata* ssp. *petraea* population located in Germany (PL; 49.65N, 11.45E) were crossed to obtain intra-population F1 offspring progenies. For crossing, parental individuals were propagated clonally and vernalised for nine weeks at 4°C and 12-hour daylength and subsequently transferred to the greenhouse (16-hour daylight, 12°C) until they flowered. Each individual was then crossed with two or more individuals (**Figure 3.10**). Most crosses were reciprocal -

each individual plant was used as both pollen receiver and donor - to be able to control for maternal effects. Since the same plants were used for the reciprocal crosses, the resulting seeds were considered full-siblings, and they were termed "family". All family members were half-siblings with the members of at least another family that shared the same parent (**Table S3.1**).

Seeds obtained from each reciprocal cross were stratified on wet filter paper in the dark at +4°C and kept under these conditions for 7 days. Subsequently, the seeds were allowed to germinate at 20°C, 16 hrs of daylight. Once the cotyledons were fully open, the plants were transferred to pots and placed in a walk-in growth chamber (Dixel, Germany) set at 12°C, 16 hrs of day length. The light intensity was adjusted via the light ratio of the LEDs (LED Modul III DR-B-W-FR lights by dlicht®). The LEDs were set at 100% intensity of blue (440 nm), red (660 nm), and white light with total measured intensity of 224 +/- 10 µmol * sec-1 *m-2. During both experiments, a light pulse of far red (750 nm) was implemented for 10 mins at the end of the day. The above-ground leaf material of each individual was sampled when the 20th leaf was visible on the rosette. This time point was chosen because previous observations had shown that it is the stage at which plants grow exponentially. All sampling was performed at 3 hours Zeitgeber time.

3.4.2. Phenotyping fitness-related traits

Rosette images were captured on the day of harvest at the 20th-leaf stage and processed with PlantCV (**Gehan et al., 2017**) and OpenCV (**Bradski, 2000**) using a custom Python script (see **Code Availability**) to calculate plant rosette area (**Figure 3.9A**). Following imaging, plants were moved to the common garden, two weeks afterwards, they underwent a four-week vernalization at 4 °C before being returned to growth conditions for flowering. Flowering time was recorded as the number of days from the end of vernalization to the appearance of the first flower. Flower

number was scored on three primary branches per plant and averaged. Leaf thickness was measured on three mature, fully expanded young leaves per plant using digital calipers and averaged. For leaf serration, individual leaves were imaged, and serration was quantified as the ratio of leaf area to the area of the leaf's convex hull (calculated by PlantCV and OpenCV); a ratio near 1 indicates a simple (non-serrated) leaf, with lower values reflecting increased serration (**Figure 3.9B**).

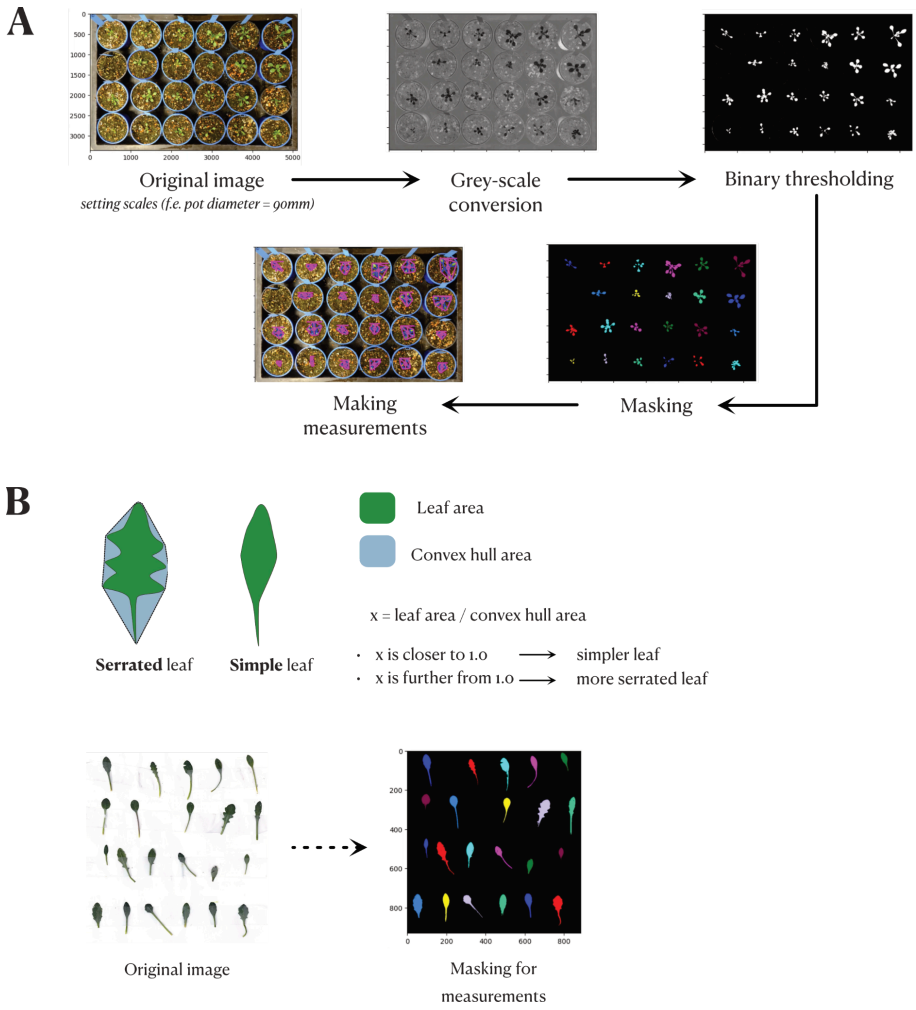


Figure 3.9. Image-based quantification of rosette size and leaf serration using PlantCV. **(A)** Workflow for measuring rosette size: original images were processed by greyscale conversion, binary thresholding, and masking, followed by extraction of rosette area measurements. **(B)** Workflow for quantifying leaf serration: leaf contours were extracted

from images, and serration was quantified as the ratio of leaf area to convex hull area (x = leaf area / convex hull area), where values closer to 1.0 indicate simpler leaves and lower values indicate more serrated leaves.

3.4.3. RNA extractions and transcriptome sequencing

To quantify the relative importance of additive and non-additive variance in gene expression, I extracted RNA using the ReliaPrep™ RNA Miniprep Systems (Promega). RNA quality and quantity was assessed with 4200 TapeStation Systems (Agilent). At this stage, any RNA extract that showed signs of degradation was discarded.

I obtained high-quality RNA samples for 186 individuals. RNA was sequenced on Illumina HiSeq4000 at Azenta Life Sciences (Leipzig, Germany) following the TruSeq protocol. Sequence reads were paired-end, 75bp-long, and stranded. They yielded on average 80X coverage of the transcriptome of each sample. Sequence quality was assessed with FastQC. Low-quality signals were detected and polyG and sequencing adapters were removed using fastp (**Chen et al., 2018**). Transcriptome mapping against the A. *lyrata* NT1 (**Kolesnikova et al., 2023**) reference genome was performed with STAR v2.5.3 (**Dobin et al., 2013**) using standard settings plus a cut-off for maximum INDEL length of 10 kbp. I calculated the gene counts per individual using STAR v2.5.3 followed by normalisation with the median-of-ratios method implemented in DESeq2. This approach accounts for differences in sequencing depth and RNA composition across samples without relying on transcript length.

3.4.4. Partitioning of gene expression variance into its components

The next step was to partition the expression variance of each gene into its genetic and environmental components. For each gene, an animal model (**Walsh & Lynch,**

2018b; Wilson et al., 2010) was fitted in R (**R Core Team, 2021**) using the brms package (**Bürkner, 2017**) which takes a Bayesian approach to fit generalised linear mixed effect models. Random effects included the additive and dominance matrix, which was calculated using the nadiv package in R (**Wolak, 2012**).

The additive relationship matrix (**A**) quantifies the probability that individuals share an allele identical-by-descent (IBD). I computed **A** using the Amat() function, which implements Henderson's tabular recursion to calculate pairwise additive coefficients. Each entry A_{ij} equals $2f_{ij}$, where f_{ij} is the coefficient of coancestry, and diagonal elements are $1+F_i$, with F_i the individual inbreeding coefficient. Mixed-model software typically computes the inverse of **A** internally, so no separate inversion step is required. The dominance relationship matrix (**D**) captures the expected probability that two individuals share both alleles IBD at a locus, that is, their similarity at the diploid genotype level. Whereas the additive matrix reflects the sharing of single alleles, the dominance matrix represents sharing of whole genotypes. I constructed **D** using the Dmat() function, which derives dominance coefficients from the pedigree following the classical rules for diploid identity relationships. Because animal-model fitting requires the inverse of **D** to assemble the mixed-model equations, I then obtained the sparse inverse matrix using Dinverse(). Thus, **A** was provided directly, whereas **D** was supplied in its inverse form, as required for REML estimation of both additive and dominance genetic variance components.

Additionally, the identity of the mother plant was included to control for effects of the plant that the seeds matured on (maternal effect). The model formula was as follows:

$$y_i = \mathcal{N}(\mu_i, \sigma^2)$$

$$\mu_i = \beta_0 + a_i + d_i + m_i$$

where:

$\boldsymbol{y_i}$ *is the observed gene expression (or phenotype) for individual i*

$\boldsymbol{\sigma^2}$ *is the residual variance*

$\boldsymbol{\beta_0}$ *is the overall intercept*

$\boldsymbol{a_i} \sim \mathcal{N}(0, \sigma_a^2 \mathrm{A})$ *is the additive genetic effect*

$\boldsymbol{d_i} \sim \mathcal{N}(0, \sigma_d^2 \mathrm{D})$ *is the non-additive genetic effect*

$\boldsymbol{m_i} \sim \mathcal{N}(0, \sigma_m^2 \mathrm{M})$ *is the maternal genetic effect*

The skew_normal() family was used to accommodate asymmetry in the phenotypic distribution and specified 4 chains of 20000 iterations each (5000 warmup), with threads=4, thin=10, and adapt_delta 0.99. For each gene, the additive (V_A), maternal (V_M), non-additive (V_{NA}), and residual (V_R) variance components were extracted from the model. The sum of the four variances comprises the total phenotypic variance (V_P), whereas the sum of the V_A and V_{NA} represents the total genetic variance (V_G). To compare the genetic composition of phenotypic variance across the genome, I presented V_A, V_{NA}, V_M, and V_R as fractions of V_P, unless stated otherwise. I also computed the **evolvability** ($\mathbf{e_\mu}$) of each gene as the absolute value of additive variance divided by the log10 of squared mean normalised transcript count (**Hansen & Houle, 2008**).

Randomisation analysis: To assess whether estimated variance components arose from true biological signal or noise, I repeated the variance partitioning after randomising individual identities in the gene expression table, while keeping the pedigree and model structure unchanged. Variance components (additive, non-additive, maternal, and residual) were then re-estimated for each gene using the same animal model, filtering criteria, and inference pipeline as in the main analysis.

3.4.5. Gene ontology enrichment analyses and clustering of gene expression

Gene ontology enrichment analyses were performed in R with the topGO package (**Alexa et al., 2006; Subramanian et al., 2005**), using as gene universe the list of annotated orthologue genes in the A. *thaliana* genome and applying the test to genes ranked by decreasing values of either V_A or V_{NA}. Enrichment was assessed in topGO using Fisher's exact test with the **'elim'** algorithm, against the universe of annotated A. *thaliana* orthologues and genes ranked by decreasing V_A or V_{NA}.

In order to test for potential clustering of the data and therefore non-independence of the expression phenotypes, I first computed the pairwise Spearman correlation coefficient between gene expressions across all the gene pairs in the dataset. To cluster the genes, the pairwise ρ values were transformed to pairwise Euclidean distance with the formula $D = (1-\rho)/2$. Hierarchical clustering was done in R with the function hclust (stats package v4.2.0; (**R Core Team, 2021**)). Potential clusters were identified by examining the within-group sum of squares for clusters 1 to 300. Clusters were pruned to 25, 50, 100, and 200 clusters by R function cuttree (stats package v4.2.0; (**R Core Team, 2021**)). The impact of clustering on the genetic variances and gene properties was tested by correlating the median V_A, V_{NA}, or transcript length of each group of clustering genes with the size of each cluster (i.e.. the number of genes within each cluster).

3.4.6. Linking genomic features to additive and non-additive variance

Another question was whether there are relationships between the components of genetic variance and transcription factor (TF) binding sites and gene properties in A.*lyrata*. To identify genomic features predictive of genes with high non-additive genetic variance (V_{NA}), I trained a gradient boosting classifier using genomic annotations as input features. Putative binding sites were predicted from A. *thaliana* using JASPAR2018 Bioconductor package (JASPAR 2020). I subsequently

extracted counts of each TF binding site in the 1 kb upstream of each gene of A. *lyrata*. I treated the count of each TF binding site per gene as a separate feature and determined the total number of TF binding sites per gene, the number of exons, the transcript length and gene length. Taken together, these descriptors of genomic variation provided a total of 443 features in which one could assess the predictive impact on the composition of variance **(Table S3.3).** A binary response variable was formed by labelling genes previously categorized as high V_{NA} (based on their variance decomposition profiles). The feature set was cleaned to remove missing values, and class imbalance was addressed by randomly subsampling 1,300 genes from the larger class (i.e., all genes that are not "high-V_{NA}") for 100 times, while retaining all high-V_{NA} genes. Similarly, for "high-V_A" genes, 2000 genes from the larger class (i.e., all genes that are not "high-V_A") were subsampled. The balanced datasets were used to train and evaluate the classifier across these 100 bootstrapped iterations. In each iteration, the data were split into training and test sets (80/20), and a gradient boosting model was trained with 100 estimators and a learning rate of 0.1.

3.4.7. Variant calling for population genetics analyses

Genetic variation data were analysed using 17 individuals from the Plech population of *Arabidopsis lyrata* and 23 from Spiterstulen (**Takou et al., 2021**). These populations are known for its high genetic diversity, making it a particularly informative system for studying patterns of polymorphism and evolutionary processes in natural populations. I mapped the sequencing reads to the *NT1* reference genome from **Kolesnikova et al. (2023)** using Bowtie2 (**Langmead & Salzberg, 2012**) and performed variant calling using GATK v4.6.1 (**McKenna et al., 2010**). After variant calling, SNP data were subjected to a series of filtering steps. First, I masked variants overlapping annotated repetitive elements, using the repeat annotation (EDTA-based transposable element prediction) generated from the A.

lyrata NT1 reference assembly as a mask file (**Kolesnikova et al., 2023**). This step was performed with *bedtools intersect* (bedtools v2.31) (**Quinlan & Hall, 2010**) ensuring that SNPs located in transposable elements or other low-complexity regions were excluded. Next, I removed sites showing fixed heterozygosity, i.e., where all individuals were called as heterozygotes and no homozygotes were present, as these are likely to reflect mapping artifacts or misalignments. To further refine the dataset, I extracted site-level depth (DP) and quality (QUAL) statistics using bcftools v1.18 (**Danecek et al., 2021)**, providing an overview of sequencing coverage and variant confidence. Finally, using VCFtools v0.1.16 (**Danecek et al., 2011**), I applied variant-level filters: INDELs were removed, only biallelic SNPs were retained, and thresholds were set to exclude sites with low quality (QUAL < 30), insufficient read depth (DP < 10), or excessive missing data (>20% missing genotypes).

This dataset then serves as the foundation for downstream analyses, such as site frequency spectrum (SFS) estimation and demographic inference.

3.4.7. Ratio of nonsynonymous to synonymous polymorphisms

To investigate patterns of nucleotide diversity, I calculated the ratio of nonsynonymous to synonymous diversity (π_N/π_S). The unfolded SFS was generated using easySFS (**Gutenkunst et al., 2009; Overcast, 2023**). I used the above-mentioned genetic variation of 17 individuals from the Plech (PL) population of *Arabidopsis lyrata*, together with that of 23 individuals from the Spiterstulen (SP) population of *Arabidopsis lyrata* (**Takou et al., 2021**). The variant calling result containing both variant and non-variant sites was used to compute nucleotide diversity (π) for all sites using pixy (**Korunes & Samuk, 2021**), a tool specifically designed to account for invariant sites in diversity calculations. Next, I annotated all the positions for 0-fold (nonsynonymous) and 4-fold (synonymous) sites using

degenotate (**Harvardinformatics, 2024; Mirchandani et al., 2024**). Subsequently, I calculated mean π_N and mean π_S for each gene, creating a final table of genes with their respective π_N and π_S values (**Table S3.8**). These genes were further categorized into groups based on genetic variance components. For each group, I performed 1000 bootstrap resampling iterations and calculated the mean π_N/π_S ratio for each bootstrap, generating the distributions of bootstrapped π_N/π_S ratios across groups.

3.4.8. Quantification of amino acid divergence

The *Arabidopsis thaliana* TAIR10 coding sequences (CDS) FASTA was retrieved from the Joint Genome Institute's Phytozome portal (released on 16 December 2023). For *Arabidopsis lyrata* NT1, I downloaded the reference genome FASTA and corresponding GTF annotation (**Kolesnikova et al., 2023**). The orthologs of *Arabidopsis lyrata* NT1 and *Arabidopsis thaliana* TAIR 10 were also obtained from Joint Genome Institute's Phytozome portal (*Arabidopsis lyrata* v2.1). I used a custom Python/GFFutils pipeline (see **Code Availability**) to extract and concatenate CDS exon features for each transcript directly from the GFF3 and genome FASTA, writing one CDS FASTA per transcript. I aligned each filtered CDS pair in codon space via a two-step pipeline using MAFFT v7.480 (**Katoh, 2002**) and PAL2NAL (**Suyama et al., 2006**). First, protein translations were aligned with MAFFT–auto. Second, PAL2NAL v14 was used to back-translate to a two-sequence codon alignment in FASTA. Alignments were then filtered to retain only those with 100% coverage i.e., CDS pairs whose codon alignments achieved 100% coverage (no missing codons in either sequence). All other pairs were discarded prior to Ka/Ks estimation. This step ensures that our Ka/Ks values are computed on fully homologous, gap-free alignments and are not inflated by misaligned regions. This choice is conservative and may exclude rapidly evolving or structurally complex genes. However, it prioritises reliability of codon alignment and avoids artefactual

inflation of Ka/Ks due to misalignment or incomplete CDS annotation. Because the analyses compare relative patterns across gene sets rather than absolute rates, this conservative filtering should not bias qualitative conclusions, but instead reduces noise.

Pairwise nonsynonymous (Ka) and synonymous (Ks) substitution rates were calculated on the codon alignments using KaKs_Calculator v2.0 (**Wang et al., 2010**) with the Yang–Nielsen (YN00) method (**Yang & Nielsen, 2000**). In the end, I obtained Ka/Ks estimates for 6,419 ortholog pairs. Of these, 3,453 showed Ka/Ks significantly ≠ 1 ($p < 0.05$, YN00 test) (**Table S3.9**). I then classified these 3453 into groups based on the amount of genetic variance components. For each group, I performed 1000 bootstrap replicates with replacement, recalculated the mean Ka and mean Ks. For each group, I performed 1,000 bootstrap replicates with replacement, recalculated the mean Ka and mean Ks, and computed the mean Ka/Ks ratio. I then summarised each group by its original mean Ka, mean Ks, and Ka/Ks ratio, along with the bootstrap distribution of these statistics (1,000 values).

3.4.9. Distribution of fitness effects across gene classes based on predominant genetic variance components

To investigate the level of constraint of the groups of genes with different levels of genetic variance in the PL and SP, the parental populations of our crosses, I grouped the genes in four categories based on whether more than half of V_P was due to additive, non-additive or residual variance, grouping all other genes in an intermediate group iVg, with intermediate fractions of all three variance components. A fifth set of 4,000 random genes was further added as a control group. The distribution of fitness effects in each group of genes (DFEs) was assessed using a combination of general python functions, δaδi and fitδaδi

(**Gutenkunst et al., 2009**), following the exact procedure that was previously reported in **Takou et al. (2021**), and for which a code is available.

In brief, synonymous variation was used to determine the best-fitting demographic parameters to the parental population PL (or SP). The distribution of fitness effects was then predicted given the simplified demographic model in each population and for each group of genes. The simplified demographic model was inferred by maximizing the composite likelihood of the folded SFS at 4-fold degenerate sites of each gene group in PL (or in SP) using the "L-BFGS-B" method and basin-hopping algorithm implemented in scipy (**Virtanen et al., 2020**). These models provided a good fit of the predicted neutral SFS to the data of each gene set. After estimating the demographic parameters, I proceeded to the second step of the analysis and used the 0-fold SFS of each gene set to fit their respective DFE by estimating the shape and scale parameter of a gamma distribution of selection coefficients. Analysis was performed assuming that deleterious variants were all co-dominant (h = 0.5). For this, the 4-fold population-scaled mutation rate theta was estimated, which reached 24,000 for PL. This rate was multiplied by 2.76 to get the 0-fold mutation rate, that is, the non-synonymous mutation rate, for PL. Note that the theta used for the SP population had to be constrained to theta PL × 0.90, to account for the difference in number of sites retained in each population after all filters for sequence quality. To estimate the DFE from the data collected for each gene group, I used a Poisson model including the population scaled mutation rate, theta, and compared the likelihood of the data along the parameter space. This model fits a DFE that is unbiased by mutations possibly too rare to be observed in the sample of genotypes used for computing the site frequency spectrum. Having determined the demographic parameters of the two populations and the DFE of each gene group, I proceeded to the third step of the analysis, which predicts the properties of genetic variation of each gene set and allows us to visualize the fit of

the DFE and demography to the data. These properties follow from the DFE and demographic histories under the standard diffusion model. For this step, I calculated the distribution of selection coefficients for variants in each count of the SFS of each gene set. First, the expected SFS for each selection coefficient under the demographic model using ∂a∂i functions was calculated. Then, I calculated the expected distribution of s using the python function gamma.cdf with the shape and scale parameter calculated for the joint estimate of the DFE of each gene set. Finally, I inferred the distribution of selection coefficients in each count of the SFS of each gene group by applying Bayes' rule (**Eyre-Walker & Keightley, 2007**). The confidence intervals for each group were based on values derived from 200 bootstraps obtained by resampling with replacement over genes. The statistical significance of differences between the groups in the DFE distribution was calculated by pairwise comparisons within each bin of the DFE (0-1, 1-10 and 1-inf). This method assumes that allelic effects are purely additive. If in fact variants had recessive effects on fitness, the method underestimates the magnitude of their effect. It is important to note that this assumption should lead to a pattern opposite to the one reported in this chapter, namely, variants in high-V_A genes experienced weaker selection than variants in high-V_{NA} genes. The results are therefore robust to the assumption of additive allelic effects on fitness.

3.4.10. Expression quantitative trait loci study for identification of trans and cis regulatory effects on gene expression

I performed expression quantitative trait locus (eQTL) mapping to identify *cis*- and *trans*-regulatory effects on gene expression variation using the single nucleotide polymorphism (SNPs) calls from 186 RNA-seq samples. SNPs calling was performed using GATK v4.6.1 (**McKenna et al., 2010**). The analysis was performed with QTLtools v1.3.1 (**Delaneau et al., 2017**), a software designed for efficient and robust eQTL detection. Using bcftools v1.18 (**Danecek et al., 2021**), variants were filtered

based on the following thresholds: information depth (INFO/DP) greater than 30 but less than 60, a maximum missingness rate of 10% (max-missing = 0.9), and a minor allele frequency (MAF) threshold of at least 0.05, all INDELs were removed and no missing sites were allowed. I retained the top hits for each eQTL by applying a false discovery rate (FDR) threshold of 0.05 for *cis*-eQTLs and a Bonferroni-corrected p-value threshold of 0.05 for *trans*-eQTLs to control for multiple testing. All necessary genome annotation information required for the input in the eQTL analysis are arranged in **Table S3.10**.

3.4.11. Using machine learning approach to link variation in gene expression with variation in phenotypes

To identify transcriptomic predictors of phenotypic variation, I implemented two complementary machine learning approaches: an *unsupervised* principal component regression (PCR) framework and a *supervised* gradient boosting classifier.

In the unsupervised analysis, I followed the procedure described by **Henry & Stinchcombe (2025)**. Firstly, gene expression counts (26,154 genes) were standardized using a z-score transformation so that each gene contributed equally to the analysis. A principal component analysis (PCA) was then performed on the standardized expression matrix to reduce dimensionality, retaining components that together explained 95% of the total variance. The principal component (PC) scores were subsequently used as predictors of relative fitness, defined as individual rosette size divided by the population mean. Each PC was first tested in a univariate linear regression against relative fitness to obtain approximate significance rankings, after which models were evaluated using repeated five-fold cross-validation to estimate performance metrics including R^2, RMSE, and MAE. R^2 (coefficient of determination) measures the proportion of variance in relative

fitness that is explained by the model; values closer to 1 indicate better explanatory power. RMSE (root mean square error) quantifies the average magnitude of prediction error on the original scale of relative fitness, with lower values indicating more accurate predictions. MAE (mean absolute error) measures the average absolute difference between predicted and observed values and is less sensitive to large errors than RMSE. Together, these metrics provide complementary assessments of how well the model captures variation in fitness and how accurately it predicts unseen individuals. The optimal subset of PCs was selected as the configuration yielding the highest cross-validated R^2. Regression coefficients from the final model were then back-transformed from PC space into the original gene expression space using the PCA rotation matrix, yielding per-gene selection gradients that represent the strength and direction of association between each gene's expression level and relative fitness. Genes were ranked according to the absolute value of these gradients, and the top 300 genes were retained as those most strongly associated with fitness variation.

For the supervised approach, I trained a Gradient Boosting Classifier (GBC) using gene expression data as predictors and the binary phenotype (above vs. below median rosette size) as the target variable. The data were split into 60% training and 40% test sets and standardized prior to model fitting. The model was trained using 100 estimators, a learning rate of 0.1, and a maximum tree depth of 3. Model performance was evaluated on the held-out test set using overall accuracy and class-specific metrics derived from the classification report. Gene-level feature importance scores were extracted from the trained model and ranked in descending order, with the top 300 genes selected as the most influential predictors.

3.5. Results

3.5.1. Predominance of non-additive genetic variance in gene expression

In total, 14 full- and half-sib families of the outcrossing plant species *Arabidopsis lyrata* ssp. *petraea* were generated by crossing individuals of a population from Plech, Germany (49.65N, 11.45E) (**Mattila et al., 2017; Takou et al., 2021**) (**Table S3.1**). Leaf material was collected during the exponential rosette growth phase of the plants and sequenced their transcriptome. Full- and half-sib relationships allow the partitioning of the total phenotypic variance (V_P) of gene expression into additive variance (V_A), non-additive (V_{NA}), maternal (V_M), and residual (V_R) components of variance in the expression for each transcript in the genome (**Figure 3.10A**, **Figure S3.1**). The main non-additive component is often termed dominance variance (V_{NA}) in quantitative genetics studies, and although strictly defined for intra-locus interactions, the use of the term typically also includes inter-locus allelic interactions that give rise to epistatic genetic variance. Indeed, the statistical variance due to intra- and inter-locus interactions, i.e. dominance and epistatic variance, cannot easily be disentangled.

In 26,154 genes analyzed, 38% of genetic variance in gene expression was explained by non-additive variance (**Figure 3.10B**). The maternal variance (V_M) was close to zero for most transcripts (median = 0.018) (**Figure S3.1**). Hence, even when gene expression variance is due largely to genetics, only a limited fraction of that genetic variance is inherited additively. Although amounts of non-additive genetic variance are rarely quantified in natural populations, the amount detected here is not unexpected. Similar fractions of non-additive variance have been reported in phenotypic traits (**Crnokrak & Roff, 1995**). In the current study, the amount of non-additive variance was substantial, despite not being higher than additive variance for most transcripts expressed in threespine stickleback (**Leder et al.,**

2015). The 95% credible intervals for both V_A and V_{NA} excluded zero in all transcripts, confirming that the levels of genetic variance reported here are systematically non-zero and not sampling artifacts. I performed a genome-wide *trans*-eQTL scan across all 26,154 transcripts and found no *trans*-acting hotspots, indicating that each transcript's expression behaved as an essentially independent phenotype (**Figure S3.2**).

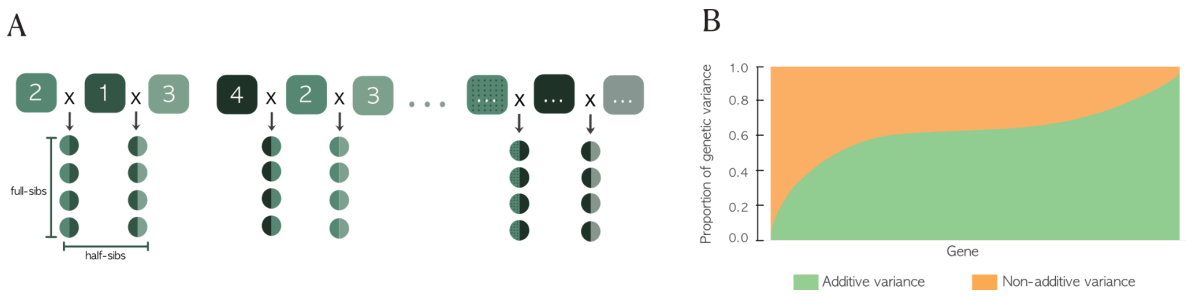


Figure 3.10. A predominant fraction of genetic variance of gene expression is composed of non-additive genetic variance. (A) Schematic representation of the full- and half-sib structure used in the North Carolina II design. For illustration clarity, only a subset of parental combinations is shown. To dissect the genetic architecture of gene expression, this design was applied to generate 186 F1 individuals derived from reciprocal crosses between nine natural *Arabidopsis lyrata* genotypes sampled in Plech (Germany). Each genotype was reciprocally crossed with two other genotypes or more, producing a series of full-sibs (two individuals share both parents) and half-sib families (two individuals only have one common parent). **(B)** The relative fractions of additive and non-additive genetic variance over total genetic variance (i.e., sum of additive and non-additive genetic variance) for each of the 26,154 genes analysed. Average broad-sense heritability is 11%.

To assess whether the estimated variance components could arise as artefacts of pedigree structure or sampling noise, I examined their behaviour under phenotype randomisation (**Figure S3.7**). When individual identities were permuted while keeping the pedigree constant, estimates of additive variance (V_A) collapsed to

values near zero, confirming that additive covariance in the real data reflects genuine heritable structure rather than model artefacts. In contrast, estimates of non-additive variance (V_{NA}) persisted under randomisation, indicating that dominance-related variance is not eliminated by phenotype shuffling in the same way as additive variance.

Unlike additive covariance, dominance covariance is strongly influenced by full-sib relationships and finite sampling structure. When phenotypes are randomised, individuals that are accidentally similar and also share dominance relationships can generate spurious dominance covariance, inflating V_{NA} estimates under the null. As a result, permutation provides a stringent test for V_A but has limited power to fully eliminate residual V_{NA} signal. Consequently, genome-wide levels of non-additive variance are difficult to separate cleanly from random expectations on a per-gene basis. However, this limitation does not imply that individual genes lack meaningful non-additive variance; rather, it reflects that pedigree-based estimates of dominance variance are intrinsically harder to distinguish from noise than additive variance.

Despite this limitation, relative differences in V_{NA} between genes remained informative. All variance components were estimated under the same model, using identical pedigree structure and sampling design, such that any residual inflation affects genes in a comparable manner. The downstream analyses focused on genes enriched for comparatively high V_{NA} to test whether transcripts with elevated non-additive genetic variation show consistent patterns in molecular evolution or regulatory architecture, without requiring that absolute V_{NA} magnitudes be estimated without error.

3.5.2. Strength of purifying selection associates with inheritance of regulatory variation

As the levels of non-additive genetic variance in gene expression were determined, the next question was whether the predominant mode of expression inheritance was associated with past selective pressures acting on the encoded proteins. The expressed transcripts were partitioned into three non-overlapping sets based on which genetic variance component (V_A, V_{NA}, V_R) was most strongly represented in their expression: high-V_A, high-V_{NA}, and high-V_R genes. For each category, I selected the 2,000 genes whose median value for the focal component exceeded the other two. Genes without a dominant component were assigned to an intermediate group (iVg), and an additional set of 4,000 transcripts was sampled at random to provide a genomic baseline. These gene sets were mutually exclusive. I then estimated amino-acid divergence between A. *lyrata* and A. *thaliana* for each group.

The five gene sets differed significantly in their rates of amino-acid evolution (**Fig. 3.11A**). High-V_A genes exhibited the highest Ka/Ks, indicating comparatively weak long-term purifying selection, whereas high-V_{NA} genes showed the lowest Ka/Ks, indicative of strong constraint. High-V_R genes also displayed low Ka/Ks values, while iVg and random genes were intermediate. Because Ka/Ks reflects long-term fixation patterns and can be influenced by heterogeneity in synonymous divergence, I complemented the analysis with polymorphism-based inference of the distribution of fitness effects (DFE) using demographic models previously inferred for these populations (**Takou et al., 2021**).

The DFE analyses supported a clear contrast between high-V_A and high-V_{NA} genes (**Fig. 3.11C**). High-V_A genes contained a significantly larger fraction of nearly neutral mutations ($0 < Nes < 1$; $p = 1.2 \times 10^{-11}$) and fewer strongly deleterious mutations ($Nes > 10$; $p = 3.15 \times 10^{-13}$) than high-V_{NA} genes, consistent with weaker purifying selection

and in agreement with their elevated πN/πS values (**Fig. 3.11B**). High-V_{NA} genes, by contrast, harboured the largest proportion of strongly deleterious sites, exhibited the lowest Ka/Ks and πN/πS ratios, and showed reduced synonymous diversity (πS; **Fig. 3.11D**). The iVg and random sets behaved similarly and served as baselines: iVg genes displayed slightly relaxed constraint in the DFE (higher fractions of nearly neutral variants) but otherwise resembled the genomic background in divergence and polymorphism measures. High-V_R genes, however, displayed a different pattern. Although their Ka/Ks values were low, indicating strong long-term constraint, these genes also exhibited the highest πN/πS ratios and a DFE enriched for nearly neutral mutations. This High-V_R gene set experienced elevated synonymous divergence (Ks) but at the same time, low synonymous diversity (πS) (**Figure 3.11D,E**). It is important to note that across High-V_A, High-V_{NA}, and High-V_R gene sets, the fitted DFE models showed comparable correspondence between observed and expected site frequency spectra (**Figure S3.8**); therefore, the different behaviour of High-V_R genes was unlikely due to the model adequacy. One hypothesis for the observed combination in High-V_R genes can be *background selection*, which reduces local effective population size and the efficacy of selection on mildly deleterious variants (**Charlesworth et al., 1993**).

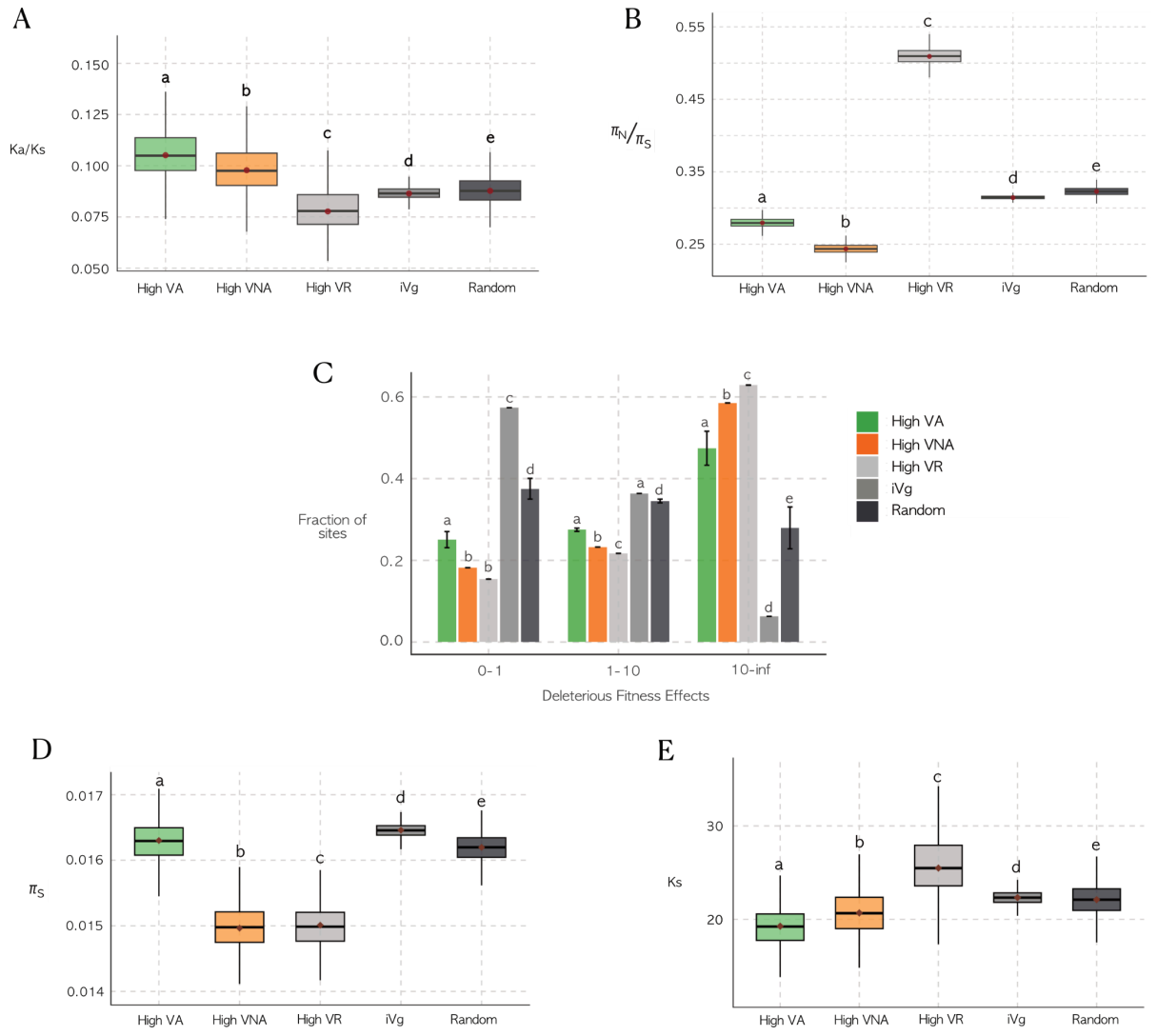


Figure 3.11. The strength of purifying selection at the amino-acid level is associated with the predominant component of the variance in expression. Expressed transcripts were partitioned into three non-overlapping sets called **high-V_A**, **high-V_{NA}**, or **high-V_R** genes. Genes not strongly dominated by a single variance component were assigned to an "intermediate" category (**iVg**). As a control a fourth set of 4000 genes sampled at random (**random**) was included. **(A)** The rate of amino acid divergence Ka/Ks that evolved between A. *lyrata* and A. *thaliana* across these groups of genes was estimated, and the mean and variance of average Ka/Ks values based on 1000 bootstrap iterations were computed. **(B)**

Estimates of the π_N/π_S ratio for the five gene groups based on population genomics datasets for *Arabidopsis lyrata* natural population and the mean and variance of average π_N/π_S values based on 1000 bootstrap iterations. **(C)** Inferred distribution of fitness effects of new mutations in the gene groups for each of three classes of deleterious fitness effect N_es (nearly neutral, $N_es<1$, weakly deleterious $1<N_es<10$, strongly deleterious $N_es>10$). **(D)** Estimates of π_S values for the five gene groups based on population genomics datasets for *Arabidopsis lyrata* natural population and the mean and variance of average π_S values based on 1000 bootstrap iterations. **(E)** Ks values from the analysis of Ka/Ks in **(A)**, and the mean and variance of average Ks values based on 1000 bootstrap iterations were computed.

3.5.3. Gene architecture associates with the level of non-additive variance in transcripts

In order to better understand whether genomic factors modulate the relative fractions of non-additive and additive variance, I compiled a total of 443 descriptors of genetic variation at each gene; these included the number of known transcription-factor binding sites in the 1kb upstream the respective gene, the length of the genes and of the transcripts, and the number of exons per gene. I trained a gradient boosting classifier model to identify the genomic architecture that best explains variation in the fraction of V_A or V_{NA} in gene expression variance. Because the gene groups were unbalanced (e.g., ca. 2,000 high-V_{NA} genes *versus* ca. 17,000 genes in “the rest” group), I randomly sampled 2,000 genes from “the rest" group, repeated this process 200 times, and ran the model for each replicate. The error bars represent variability across these bootstrapped replicates, providing a robust estimate of model performance and feature importance despite the group size disparity.

For variation in V_{NA}, the model provided a good performance: 5-fold cross-validation yielded a mean accuracy of 0.590 ± 0.035 (range 0.540–0.636). On

an independent test set, the model achieved 62.6 % accuracy, an AUC of 0.665, and a log-loss of 0.657 (the confusion matrix is shown in **Table S3.4).** Specifically, the total gene length (importance: 0.28); transcript length (importance: 0.11), and the number of exons (importance: 0.025) were the three factors that best predicted levels of non-additive variance. Since these significant descriptors tend to increase the mutational target size of the transcript, it can be hypothesised that non-additive variance is likely to be observed as the mutational target size increases.

In addition, the transcription factor binding sites associated with high V_{NA} were putative targets of AHL12; this protein (AT1G63480), is a nuclear transcriptional regulator that binds AT-rich DNA regions and likely modulates gene expression through chromatin remodeling: its interaction with other transcription factors potentially regulate plant growth and development (**Zhao et al., 2013**). The second most significantly associated with high-V_{NA} values is a putative target of a DOF zinc finger protein expressed in young rosettes (DOF 5.3); this protein (AT5G60200 TMO6) was recently shown to drive the stem cell division that is necessary for plants to thicken (**Miyashima et al., 2019**). This regulator was likely active in the collected tissue samples, because plant material was sampled during the exponential phase of rosette growth.

3.5.4. Non-additive variance increases with the size of the co-regulatory network

As the results indicated that larger mutational target sizes (longer transcripts and more exon–intron boundaries) elevated non-additive variance, the next question is how regulatory network complexity may contribute to the opportunities for the inter-allelic interactions that give rise to non-additive variance. I defined co-regulatory modules by cutting the transcript-coexpression tree into 25, 50, 100, or 200 clusters, then asked how strongly non-additive variance (V_{NA}) is associated

with gene length, exon count, and transcript length. At every clustering resolution, genes in the high-V_{NA} group showed a markedly stronger Spearman ρ and consistently significant p-values than those in high-V_A. (**Table S3.5**, **Figure S3.4**). These results suggested that transcripts embedded in larger, more complex regulatory modules, and thus exposed to more inter-allelic interactions, tended to harbor greater non-additive variance. Taken together, these findings showed that both mutational target size and network complexity amplify opportunities for non-additive genetic variance.

3.5.5. Trait- and transcriptome-level non-additive variance reflect fitness-linked selection

Similar to transcriptomic level, in all five fitness-related traits, the 95% credible intervals for both V_A and V_{NA} also excluded zero. These five traits spanned a wider range of non-additive variance (V_{NA}/V_G) than expected from the transcriptome-wide background, from mostly additive control (leaf serration, flowering time) through intermediate (leaf thickness, V_{NA}/V_G = 0.301) to strongly non-additive control (rosette area, number of flowers) (**Figure 3.12**). While *mean* V_{NA}/V_G did not significantly deviate from the that of transcriptome-wide (V_{NA}/V_G = 0.403) (Monte Carlo test, 10,000 draws, p = 0.78, **Figure S3.5**), the *variance* among the five traits (0.087) was an order of magnitude greater than the null variance calculated from the transcriptome analysis (0.007), indicating that these traits sampled a much broader distribution of the V_{NA}/V_G landscape than expected by chance. Thus, although their average non-additive variance mirrored the transcriptome background, fitness-related traits occupied a more dispersed distribution of the genetic-variance components than gene transcript levels.

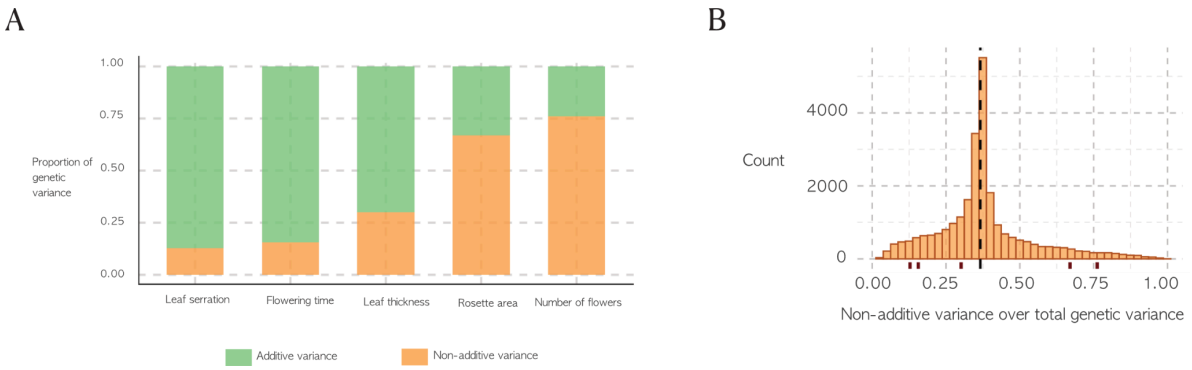


Figure 3.12. Non-additive variance in fitness-related traits is consistent with the transcriptome-wide background. (A) The proportion of total genetic variance (V_G) attributable to additive and non-additive components for five fitness-related phenotypes: leaf serration, flowering time, leaf thickness, rosette area, and number of flowers. **(B)** Histogram of the fraction of non-additive variance (V_{NA}/V_G) across all 26,154 transcripts. The vertical dashed line marks the mean V_{NA} across genes, and the red tick marks indicate each of the five phenotypes' V_{NA} values, in the same order as in (**A**).

3.5.6. Gene ontology enrichments in genes with high non-additive variance

I further quantified the distribution of genes with high V_{NA} across gene ontology (GO) categories of expressed genes. Among the top enriched GO terms for the highest estimates of relative V_{NA}, I found several functions of interest, such as "post-transcriptional gene silencing", "epigenetic regulation of gene expression", "regulation of developmental process", "vegetative to reproductive phase transition" (**Table S3.6**). In addition, among the top GO categories, the most enriched among genes with high V_A were several functions linked to central metabolism such as "quinone biosynthetic process", "peroxisome organization", but I also detected enrichments among genes involved in "epigenetic regulation of gene expression", and specifically to RNA-related functions, such as "RNA splicing" and "mRNA processing" (**Table S3.7**).

3.5.7. Predictive modeling to associate gene expression with rosette size

To evaluate whether transcriptomic variation reflects phenotypic differences in plant performance, I applied both unsupervised and supervised machine-learning approaches to identify genes whose expression covaries with rosette size during exponential growth. The unsupervised principal component regression (PCR) approach captured a moderate proportion of the variation in relative fitness. Cross-validated evaluation of the optimal model yielded an R^2 of approximately 0.59, with an RMSE of 0.29 and MAE of 0.23, indicating that multivariate patterns in gene expression explain a measurable portion of the observed variation in rosette size. In contrast, the supervised gradient boosting classifier, trained to distinguish individuals with large versus small rosettes, achieved an overall accuracy of 0.49 on the held-out test set (macro-precision = 0.50, recall = 0.50, F1 = 0.49), indicating performance close to chance for binary classification.

Given the limited predictive performance of the supervised classification approach, the top 300 genes identified in the unsupervised principal component regression were used for further downstream analysis. These genes exhibited significantly higher additive genetic variance (V_A) than non-additive variance (V_{NA}) (paired t-test, $p < 2.2e\text{-}16$, **Figure 3.13**), suggesting that genes contributing most strongly to expression–phenotype covariance exhibit a higher proportion of additive genetic variance relative to non-additive variance.

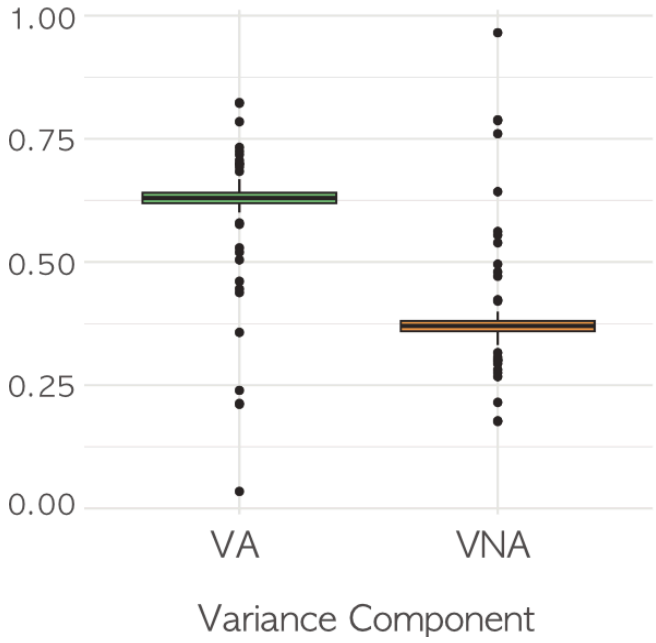


Figure 3.13. Comparison of additive and dominance variance components among genes associated with rosette size identified by the unsupervised principal component regression model. Boxplots show the distributions of V_A/V_G and V_{NA}/V_G, where total genetic variance is defined as $V_G = V_A + V_{NA}$. Points represent individual genes. The difference between two components is statistically significant (t-test, p-value < 2.2e-16)

3.6. Discussion

Delineating the genomic and evolutionary factors that increase the non-additive components of genetic variance is crucial for understanding the adaptive potential of a population. Non-additive variance, a statistical component of genetic variance, contributes to similarity between siblings but is not inherited in their offspring and thus cannot be selected for (**Fisher, 1919**). As such, the only way to quantify non-additive variance is by applying quantitative genetics methods. Plants are especially suited for dissecting genetic variance because both genetic crosses and environmental variance can be completely controlled; reliable estimates of

non-additive components of variance depend on such control (**Walsh & Lynch, 2018b; Yang et al., 2019**). Here, the estimates of non-additive variance in the 26,154 gene transcripts revealed that non-additive variance forms a considerable fraction of regulatory variation.

Gene regulation, which is central to the formation of phenotypes, is subject to natural selection (**Groen et al., 2020; Kremling et al., 2018**) and thus ideal to test whether the strength of selection shapes the amount of non-additive variance. The current data showed that purifying selection that has operated over long evolutionary time scales in the past, impact levels of non-additive variance today, and thus the potential of a gene to adapt in the future. Genes whose genetic variance at the transcript level is predominantly non-additive were significantly more constrained than genes showing a large fraction of additive variance. Past studies have been also exploring expression variation under different selection regimes and evolutionary constraints (**Fay & Wittkopp, 2008; Hämälä et al., 2020; Romero et al., 2012**), but only few recent work has begun to link the subject to components of genetic variance (**Leder et al., 2015**; **Tsouris et al., 2024**). This chapter's results provide a complementary perspective by explicitly contrasting genes enriched for additive versus non-additive variance. It was shown that genes with high additive variance were under weaker purifying selection, whereas those with high non-additive variance showed the strongest constraint. In other words, as purifying selection intensified, the genetic architecture of expression shifted from primarily additive variance toward non-additive variance. This pattern suggests that inter-allelic interactions (dominance and epistasis) become increasingly important for buffering and maintaining essential gene functions under strong selective pressure - an insight not captured by studies that focus solely on total genetic variance. The data further revealed an important additional feature: the rate of synonymous divergence and polymorphism is significantly lower

in genes with high non-additive variance compared to genes with high additive variance, suggesting that variation in the mutation rate may also play a role on the composition of genetic variance for gene expression.

Furthermore, I identified genomic features associated with the emergence of non-additive variance in gene expression. Non-additive variance increased with transcript length, consistent with a larger mutational target size facilitating the accumulation of interaction-based effects. It was also observed that transcripts embedded in larger co-expression networks tended to show higher non-additive variance. This pattern does not necessarily imply stronger selective constraint on these genes; rather, a larger regulatory neighbourhood provides more opportunities for interactions among loci to generate non-additive expression variance. Whether increased network connectivity also contributes to reduced additive variance, for example, due to stronger constraints by natural selection on central genes (**Hahn & Kern, 2005**), cannot be determined from the current data, and both hypotheses remain plausible. More broadly, dominance variance may emerge from the functional significance of genes and the optimal levels of gene expression (**Huber et al., 2018**), and regulatory network complexity and functional connectivity do not necessarily predict the mode of inheritance at the phenotypic level.

To my knowledge, this study was the first to investigate how the proportions of genetic variance observed at molecular level are translating into those measured at phenotypic level and vice versa. It was found that, on average, fitness-related traits did not exhibit higher non-additive variance than random genes; however, the dispersion of non-additive variance across these traits was markedly greater than expected by chance. This suggests that while not all fitness traits are enriched for non-additive variance, a subset of them is, such as rosette size and flower number, which are close proxies for fitness in our study. Studies on classical phenotypic and

gene-expression traits alike have been exploring the maintenance of additive variance in complex fitness- or adaptation-related traits (**Hill et al., 2008**; **Merilä & Sheldon, 1999, 2000; Tsouris et al., 2024**), and a recent work (**Bonnet et al., 2022**) has shown that additive variance for fitness-related traits can persist even under long-term selection.

This study's results add nuance to this debate: although the mean additive variance in fitness traits is not uniformly eroded, the elevated non-additive variance in key traits could suggest a partial depletion of additive variance in both molecular and higher-level phenotypes that are highly linked to fitness. Importantly, the transcripts that predict variation in rosette size show a higher distribution of V_A than V_{NA}. This disconnection indicates that much of the transcriptional variation underlying the phenotype does not translate directly into heritable (additive) variance, highlighting a key feature of the genotype–phenotype map: non-additive expression variance can buffer or mask molecular variation from contributing to adaptive evolution. Consistent with this, genes with high V_{NA} also tend to show reduced genetic diversity, suggesting that strong purifying selection may constrain the accumulation of additive genetic variance in these loci.

Together, these findings emphasise that the architecture of genetic variance at the transcript level is shaped jointly by regulatory interactions and past selection. Looking forward, epistatic processes, such as those modelled in **Chapter 4**, may transform additive genetic inputs into predominantly non-additive transcript-level variation.

3.7. Conclusion

Populations can only respond to selection in proportion to their additive genetic variance, yet little is known about the evolutionary forces that shape the balance between additive and non-additive components in natural systems. This study uses

a quantitative genetic breeding design to disentangle additive and non-additive components of expression variance in 26,154 gene transcripts of the outcrossing plant *Arabidopsis lyrata*. First, the findings reveal the widespread role of dominance and epistasis in shaping gene expression. Second, genes with high non-additive variance were found to exhibit lower synonymous diversity, reduced amino acid divergence, and stronger signatures of purifying selection compared to genes with high additive variance. These results suggest that the balance between additive and non-additive variance is itself influenced by natural selection.

Therefore, non-additive genetic variance is a pervasive and evolutionarily relevant feature of transcriptional variation. Its association with specific regulatory mechanisms and stronger purifying selection suggests that the balance between additive and non-additive variance is not merely a statistical property of traits, but itself an evolutionary outcome. Because only additive variance contributes directly to the response to selection, understanding why some molecular traits accumulate non-additive rather than additive variance is essential for determining whether underlying molecular variation can translate into heritable phenotypic change. In this sense, the distribution of variance components forms a critical part of the genotype–phenotype map, shaping the extent to which regulatory variation can generate or limit future evolutionary responses.

3.8. Supplementary Materials

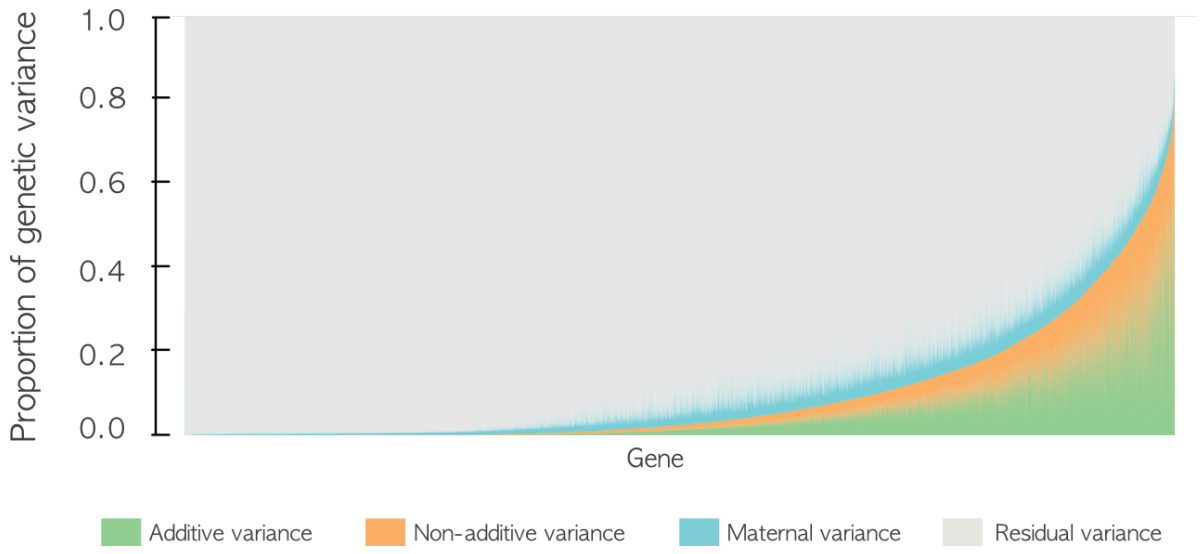


Figure S3.1. The total phenotypic variance of each one of the 26,154 genes included in the analysis was partitioned into non-additive, additive, maternal, and residual components of variance. The genes were ordered by increasing values of the fraction of total phenotypic variance in transcript level that was attributed to genetic variance (V_G).

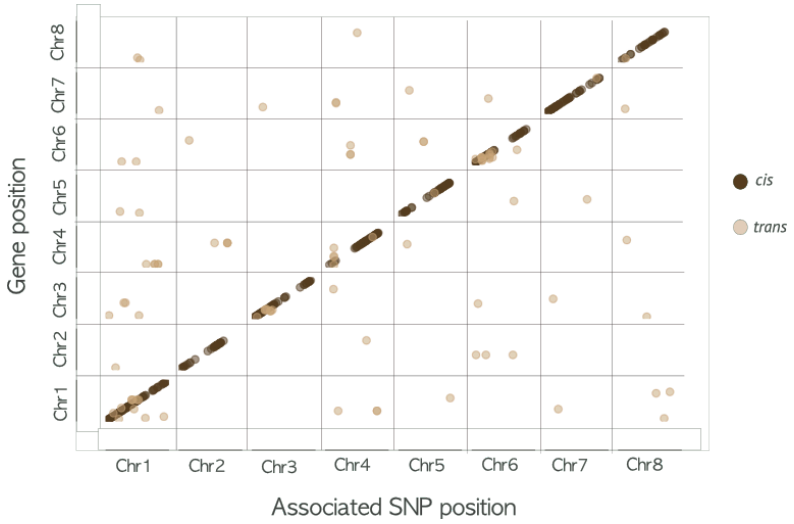


Figure S3.2. Scatterplot of *cis*- and *trans*-eQTL positions versus their associated gene locations across the eight chromosomes of A. *lyrata*. Each point represents one significant SNP–gene association, colored by regulatory mode (dark brown = cis, light brown = trans). Diagonal clustering along the identity line indicates local (cis) effects, whereas off-diagonal points reveal distant (trans) associations. The sparse distribution of high-affinity hits outside the diagonal demonstrates that few large-effect loci account for the observed variance in transcript expression.

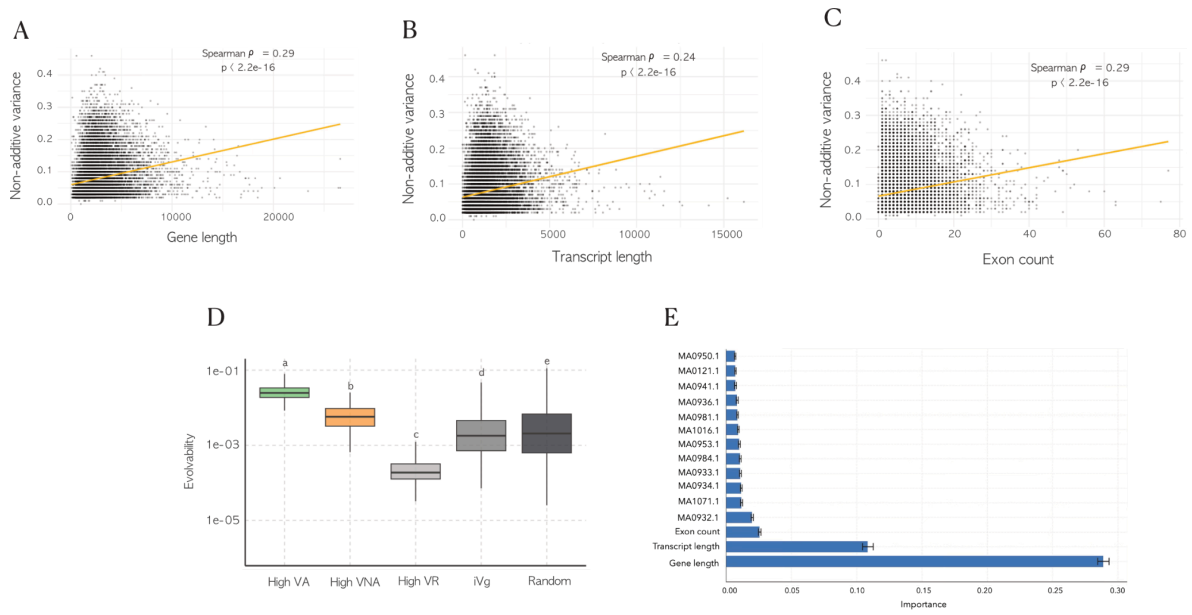


Figure S3.3. Architectural correlates of non-additive variance and evolvability. (A) Gene length versus the proportion of expression variance attributable to non-additive genetic effects (V_{NA}), with Spearman's $\rho = 0.29$ ($p < 2.2 \times 10^{-16}$). **(B)** Transcript length versus the proportion of expression variance attributable to non-additive genetic effects (V_{NA}), with Spearman's $\rho = 0.24$ ($p < 2.2 \times 10^{-16}$). **(C)** Exon count per gene versus the proportion of expression variance attributable to non-additive genetic effects (V_{NA}), with Spearman's $\rho = 0.29$ ($p < 2.2 \times 10^{-16}$). **(D)** Evolvability—defined as unstandardized additive genetic variance (V_A) divided by (log(mean expression))²—for four gene sets (high-V_A, high-V_{NA}, iVg, high-V_R, and a set of 4000 genes randomly sampled from the four gene groups. All pairwise differences among these distributions are significant by Tukey's HSD ($p < 0.05$). **(E)** Gradient Boosting Machine classifier model shows that the level of non-additive variance depends on gene architecture. Importance of genomic parameters and number of transcription factor binding sites 1kb upstream each gene on the level of non-additive variance in the respective gene. The twenty factors that explain the most variance are shown. 5-fold cross-validation yielded a mean accuracy of 0.590 ± 0.035 (range

0.540–0.636). On an independent test set, the model achieved 62.6 % accuracy, an AUC of 0.665, and a log-loss of 0.657.

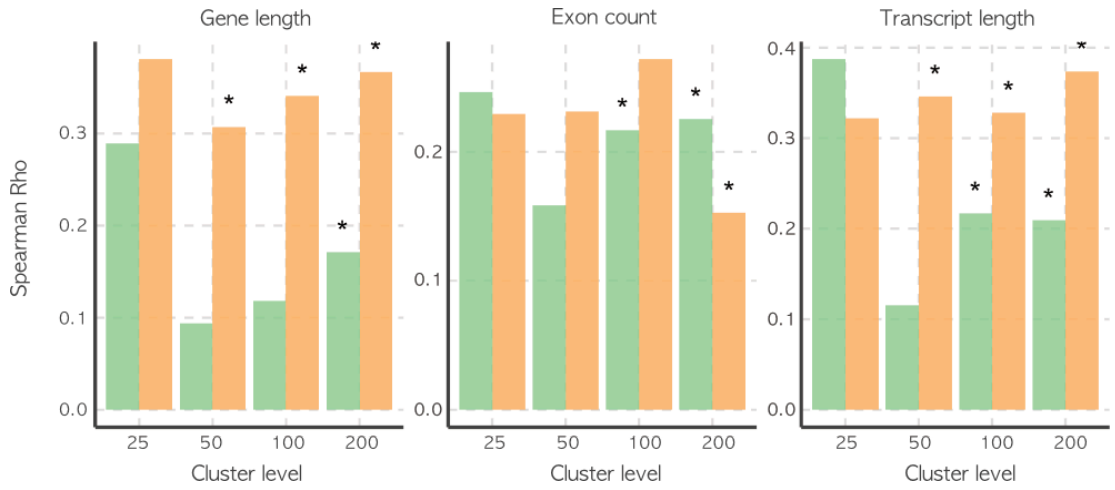


Figure S3.4. Strength of association between genetic architecture non-additive variance across co-regulatory module resolutions. I partitioned transcripts into co-expression clusters at four resolutions (25, 50, 100, and 200 modules) and computed Spearman's rank correlation (ρ) between V_{NA} and three metrics of genetic architecture: gene length, exon count, and transcript length. Within each panel, green bars show ρ for the high-V_A gene set and orange bars for the high-V_{NA} gene set; asterisks denote correlations significant at $p < 0.05$.

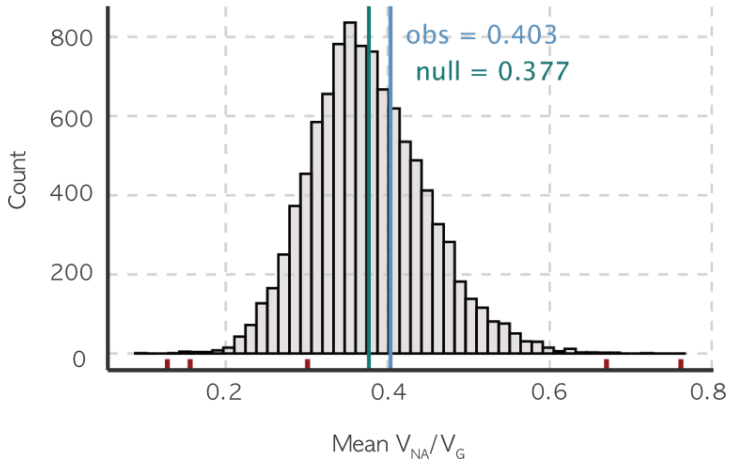


Figure S3.5. Null distribution of mean non-additive variance proportion (V_{NA}/V_G) based on 10,000 random samples of 5 transcripts each, drawn from the transcriptome-wide data. The histogram shows the distribution of mean V_{NA}/V_G under the null hypothesis that trait values reflect random gene behavior. The vertical cyan line indicates the null mean (0.38), while the blue line marks the observed mean for five fitness-related traits (0.403). Red rug ticks along the x-axis indicate individual trait V_{NA}/V_G values. The observed mean is not significantly different from the null expectation (two-sided Monte Carlo p = 0.7766). The variance among the five traits is 0.087, and of the null distribution is 0.007. The skewness of the distribution of five traits is 0.3067 and of the null distribution is 0.2187.

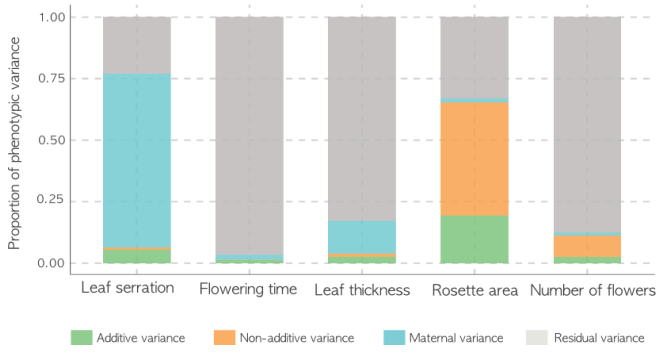


Figure S3.6. The total phenotypic variance of five fitness-related phenotypes included in the analysis was partitioned into non-additive, additive, maternal, and residual components of variance.

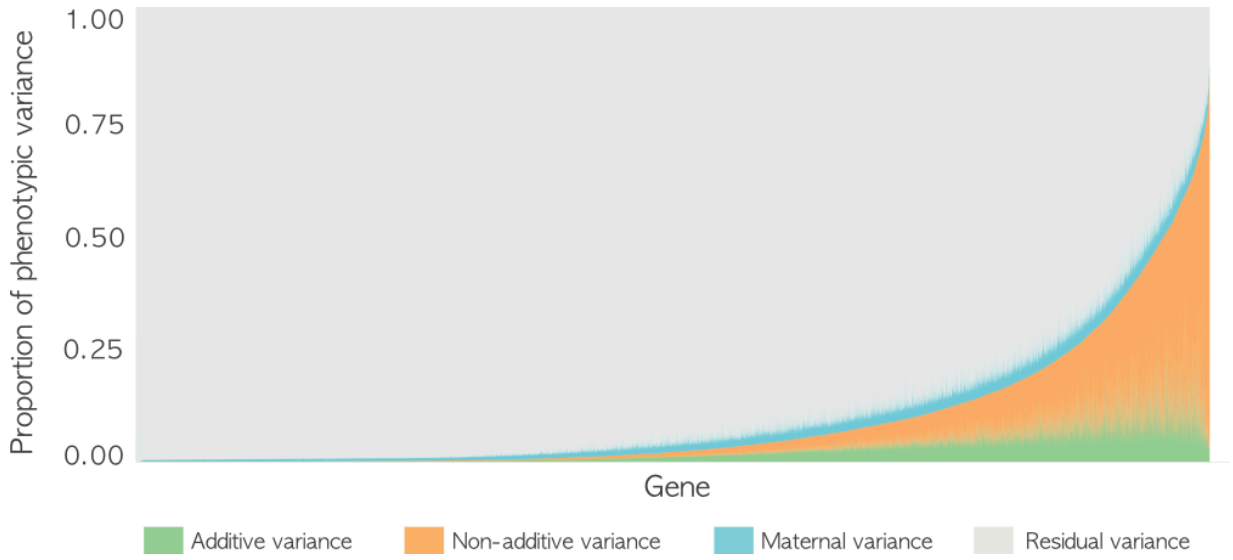


Figure S3.7. Partitioning of phenotypic variance for 26,154 genes under phenotype randomisation. For each gene, total variance in transcript abundance was decomposed into additive, non-additive, maternal, and residual components using the same animal model as in the main analysis, but with phenotypes randomised across individuals while preserving the pedigree structure. Genes are ordered by increasing proportion of total phenotypic variance attributed to genetic variance (V_G). Additive genetic variance is strongly reduced under randomisation.

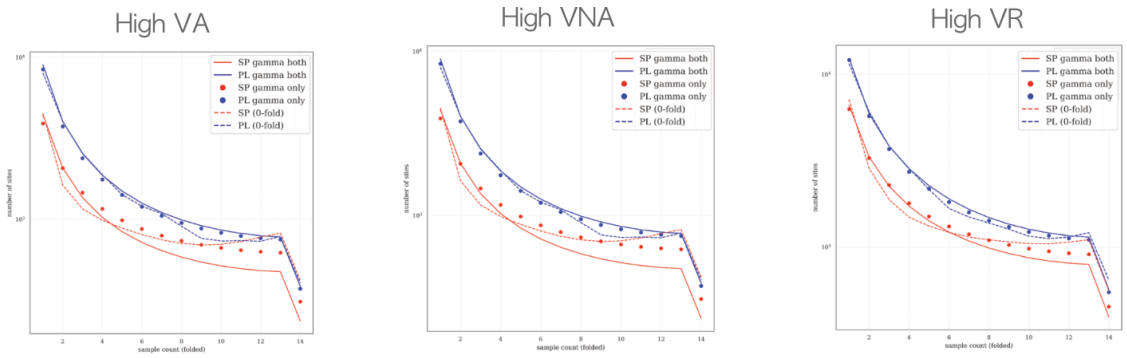


Figure S3.8. Model fit of the folded site frequency spectrum (SFS) for genes partitioned by variance component category (High-V_A, High-V_{NA}, and High-V_R) in the distribution of fitness effects (DFE) analysis, corresponding to Figure 3.11. Dashed lines show the observed SFS; lines and points represent the expected SFS under alternative DFE models, including gamma-distributed models inferred either jointly using both populations (SP and PL; *gamma both*) or separately for each population (*gamma only*). The x-axis shows the folded sample count, and the y-axis shows the number of sites per frequency class (log scale).

Table S3.1. Information about the full sibling families used in the analysis. Each full sibling family was generated based on the scheme presented in Figure 3.10. For each full sibling family, the Plech (PL) sequence ID numbers, as described in Takou (2021) and annotated in ENA archive repository, are provided (PRJEB49153). For each family, the number of offspring used in the final analysis is shown.

Family	Parent 1	Parent 2	Number of family members
1	80936	11a	12
2	10a	11a	1

2	11a	10a	3
3	2a	11a	10
4	11a	9a	3
5	6a	80936	16
5	80936	6a	4
6	80936	1a	3
7	4a	80936	2
7	80936	4a	27
8	10a	3a	1
9	3a	10a	19
9	1a	3a	3
9	3a	1a	4
10	2a	3a	3
11	10a	6a	27
11	6a	10a	3
12	1a	6a	1
12	6a	1a	27

13	1a	11a	2
14	10a	6a	13

Table S3.2. Tukey's HSD pairwise comparisons of Plech population across gene groups within each fitness-effect class. For each fitness-effect category ("Nearly Neutral", "Weakly Deleterious", "Strongly Deleterious"), I performed a one-way ANOVA on the bootstrap-derived fractions and applied Tukey's honestly-significant-difference test to all pairwise group contrasts. Columns report the mean difference in fractions between the two groups, the lower and upper bounds of the 95 % family-wise confidence interval ("Conf.low" and "Conf.high"), and the Tukey-adjusted p-value.

Fitness Effect Class	Group1	Group2	Estimate	Conf.low	Conf.high	p.adj
Nearly Neutral	High V_A	High V_{NA}	-0.0687	-0.0947	-0.0426	1.2E-11
	High V_A	High V_R	-0.0968	-0.1229	-0.0708	2.89E-13
	High V_A	iVg	0.3228	0.2968	0.3489	2.42E-13
	High V_A	Random	0.1245	0.0985	0.1506	2.42E-13
	High V_{NA}	High V_R	-0.0281	-0.0541	-0.0022	0.0259
	High V_{NA}	iVg	0.3915	0.3655	0.4174	2.42E-13
	High V_{NA}	Random	0.1932	0.1673	0.2192	2.42E-13
	High V_R	iVg	0.4196	0.3937	0.4456	2.42E-13
	High V_R	Random	0.2214	0.1954	0.2473	2.42E-13
	iVg	Random	-0.1983	-0.2242	-0.1723	2.42E-13
Weakly Deleterious	High V_A	High V_{NA}	-0.0422	-0.0537	-0.0308	2.9E-13

	High V_A	High V_R	-0.0576	-0.0691	-0.0462	2.42E-13
	High V_A	iVg	0.0890	0.0775	0.1004	2.42E-13
	High V_A	Random	0.0704	0.0590	0.0819	2.42E-13
	High V_{NA}	High V_R	-0.0154	-0.0268	-0.0040	0.00223
	High V_{NA}	iVg	0.1312	0.1198	0.1426	2.42E-13
	High V_{NA}	Random	0.1127	0.1013	0.1241	2.42E-13
	High V_R	iVg	0.1466	0.1352	0.1580	2.42E-13
	High V_R	Random	0.1280	0.1166	0.1395	2.42E-13
	iVg	Random	-0.0186	-0.0300	-0.0071	9.54E-05
Strongly Deleterious	High V_A	High V_{NA}	0.1109	0.0735	0.1483	3.15E-13
	High V_A	High V_R	0.1544	0.1170	0.1918	2.54E-13
	High V_A	iVg	-0.4118	-0.4492	-0.3744	2.42E-13
	High V_A	Random	-0.1950	-0.2324	-0.1576	2.42E-13
	High V_{NA}	High V_R	0.0435	0.0063	0.0808	0.0126
	High V_{NA}	iVg	-0.5227	-0.5599	-0.4854	2.42E-13
	High V_{NA}	Random	-0.3059	-0.3431	-0.2686	2.42E-13
	High V_R	iVg	-0.5662	-0.6035	-0.5290	2.42E-13

	High V_R	Random	-0.3494	-0.3867	-0.3121	2.42E-13
	iVg	Random	0.2168	0.1796	0.2541	2.42E-13

Table S3.3. Feature Importance for Predicting High V_{NA} from Gradient-Boosted Classifier. Summary of the mean relative importance and its standard error for each genomic feature, as estimated over 200 bootstrapped replicates (2,000 high-V_{NA} vs. 2,000 randomly sampled "rest" genes per run). Importance scores reflect each feature's contribution to the model's accuracy in classifying genes with high non-additive variance (V_{NA}). Higher values indicate greater predictive power; standard errors quantify the variability of each importance score across bootstrap iterations.

Available at 10.6084/m9.figshare.29986699

Table S3.4. Confusion matrix for the held-out test set (n = 530) in Gradient-Boosting Classifier model. The table shows counts of true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP) when classifying genes as high-VNA gene versus "other" genes. Row totals indicate the number of actual negatives and positives; column totals indicate the number of predicted negatives and positives. Class-level metrics for the high-VNA class were precision = 0.620, recall = 0.684, and F_1 = 0.650. These results demonstrate that our features carry genuine predictive signal and that the model is not merely over-fitting to bootstrapped class sizes.

	Predicted Negative	**Predicted Positive**
Actual Negative	TN = 148	FP = 113
Actual Positive	FN = 85	TP = 184

Table S3.5. Correlation of non-additive and additive variance with genome architecture traits and summary statistics of population genetics. For each correlation between the fraction of variance and the genomic feature, the p value, Spearman's ρ.

Cluster size	Feature	Spearman's ρ	p-value	Group	Significant
25	Transcript Length	0.3872	0.0558	High V_A	Not Significant
	Number of Exons	0.2463	0.2353	High V_A	Not Significant
	Gene Length	0.2895	0.1605	High V_A	Not Significant
	Transcript Length	0.3215	0.1172	High V_{NA}	Not Significant
	Number of Exons	0.2296	0.2696	High V_{NA}	Not Significant
	Gene Length	0.3808	0.0613	High V_{NA}	Not Significant
50	Transcript Length	0.1155	0.4342	High V_A	Not Significant
	Number of Exons	0.1584	0.2823	High V_A	Not Significant
	Gene Length	0.0939	0.5256	High V_A	Not Significant
	Transcript Length	0.3463	0.0159	High V_{NA}	Significant
	Number of Exons	0.2312	0.1138	High V_{NA}	Not Significant

	Gene Length	0.3073	0.0336	High V_{NA}	Significant
100	Transcript Length	0.2169	0.0358	High V_A	Significant
	Number of Exons	0.2170	0.0357	High V_A	Significant
	Gene Length	0.1183	0.2561	High V_A	Not Significant
	Transcript Length	0.3278	0.0015	High V_{NA}	Significant
	Number of Exons	0.2719	0.0091	High V_{NA}	Significant
	Gene Length	0.3409	0.0009	High V_{NA}	Significant
200	Transcript Length	0.2095	0.0044	High V_A	Significant
	Number of Exons	0.2256	0.0021	High V_A	Significant
	Gene Length	0.1709	0.0207	High V_A	Significant
	Transcript Length	0.3737	0.0000	High V_{NA}	Significant
	Number of Exons	0.1529	0.0446	High V_{NA}	Significant
	Gene Length	0.3664	0.0000	High V_{NA}	Significant

Table S3.6. GO enrichment among genes with high non-additive variance (V_{NA}). For each significant GO term (elim algorithm, Fisher's exact test $p < 0.05$), the table lists the GO ID,

term name, number of genes in the V_{NA} set annotated to the term, number expected by chance, fold enrichment, and p-value.

Available at 10.6084/m9.figshare.29986699

Table S3.7. GO enrichment among genes with high additive variance (V_A). For each significant GO term (elim algorithm, Fisher's exact test $p < 0.05$), the table lists the GO ID, term name, number of genes in the V_A set annotated to the term, number expected by chance, fold enrichment, and p-value.

Available at 10.6084/m9.figshare.29986699

Table S3.8. Nucleotide diversity of synonymous (πS) and non-synonymous (πN) variants in 17 Arabidopsis lyrata individuals from the Plech population. Sequencing reads were mapped to the NT1 reference genome with Bowtie2, and variants were called with GATK v4.6.1. Nucleotide diversity (π) was estimated using pixy from variant and invariant sites, with 0-fold (nonsynonymous) and 4-fold (synonymous) sites annotated by degenotate. Mean πN and πS were calculated per gene, and πN/ratios were summarized across gene groups defined by genetic variance components using 1000 bootstrap replicates.

Available at 10.6084/m9.figshare.29458916

Table S3.9. Pairwise Ka/Ks analysis of A. *thaliana* (TAIR10) and A. *lyrata* (NT1) coding sequences. CDS were extracted from reference annotations and aligned in codon space using MAFFT + PAL2NAL. Ka and Ks were estimated with KaKs_Calculator v2.0 (YN00 method) after quality filtering, yielding 6419 ortholog pairs. Genes were grouped by functional or expression categories, and 1000 bootstrap replicates were performed per group to calculate mean Ka/Ks ratios with confidence estimates.

Available at 10.6084/m9.figshare.29459012

Table S3.10. Gene annotation table for the *Arabidopsis lyrata* NT1 reference genome. The table lists all annotated genes together with their chromosome assignment, genomic start and end coordinates, and strand orientation (+/−).

Available at 10.6084/m9.figshare.30870323

Table S3.11. Normalised gene expression matrix for F_1 intra-population individuals. Rows correspond to annotated genes and columns to individual samples; values represent

normalised expression levels used as input for downstream statistical and variance-component analyses. For more details, refer to section **3.4.3.**

Available at 10.6084/m9.figshare.29986699

Table S3.12. Phenotypic trait data for F1 intra-population individuals. List of sample identifiers together with measurements of rosette size (millimeter), leaf serration (no unit), leaf thickness (millimeter), flowering time (days), and number of flowers (count), and was used for downstream quantitative genetic analyses. For more details, refer to section **3.4.2.**

Available at 10.6084/m9.figshare.29986699

4

Dominance and Epistasis Variance Components in Hybrid Populations of *Arabidopsis lyrata*

4.1. Chapter Summary

This chapter extends the analysis of genetic variance in **Chapter 3** to a hybrid population derived from crosses between genetically diverged lineages of *Arabidopsis lyrata*. Using experimental hybrids generated by crossing previously characterised German and Norwegian lines, I quantified genetic variance components and assessed how admixture reshapes the genetic architecture of gene expression traits. Empirical variance-component estimates showed that additive genetic variance increases from F_1 to F_2, consistent with recombination restoring homozygous and recombinant genotypes absent from the F_1 generation. Theoretical models demonstrated that epistasis does not generate this effect but modulates its magnitude, amplifying or attenuating the F_1–F_2 contrast depending on its direction. Together, these results show that the evolutionary consequences of hybridisation depend on how recombination interacts with underlying genetic architecture.

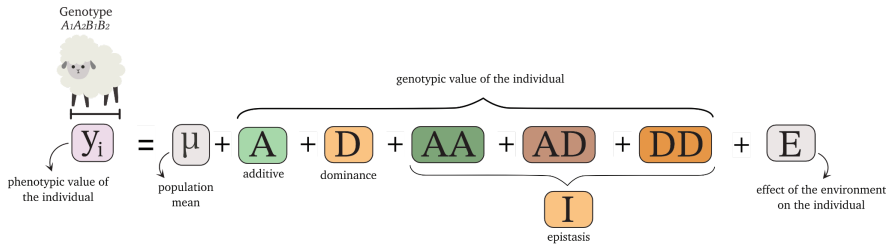
All mathematical modelling and the methodology thereof were performed in collaboration with Josselin Clo, as stated in the Author Contributions.

4.2. Fundamental Concepts

4.2.1. Understanding epistatic variance

To sum up, earlier in **Chapter 3** (**Figure 3.1 - 3.3**), we learned that the phenotypic value of an individual can be represented as the population mean plus the sum of additive effects contributed by all alleles in that individual's genome. However, there is no biological guarantee that allelic effects at the same locus combine purely additively. To account for deviations from additivity within a locus, a *dominance deviation term* is introduced, which captures non-additive interactions between alleles at the same locus.

A similar logic applies to interactions among alleles at different loci. In this case, we include an *epistatic term*, representing pairwise (or higher-order) deviations from the additive prediction due to interactions across loci. For simplicity, **Figure 4.1** illustrates a trait influenced by two bi-allelic loci. In such a two-locus system, we can partition the epistatic variance into three categories: additive-by-additive (AA) effects, additive-by-dominance (AD) effects - interactions between a single allele at one locus and the pair of alleles at another - and dominance-by-dominance (DD) effects, which involve all four alleles across the two loci. This logic can be extended to additional loci, giving rise to higher-order terms such as AAA effects for three-way interactions, and so forth.



$$y_i = \mu + \alpha_{A_1} + \alpha_{A_2} + \alpha_{B_1} + \alpha_{B_2} + \delta_{A_1A_2} + \delta_{B_1B_2} + \alpha\alpha_{A_1B_1} + \alpha\alpha_{A_1B_2} + \alpha\alpha_{A_2B_1} + \alpha\alpha_{A_2B_2}$$

Figure 4.1. Partitioning of genetic effects including epistasis. Statistical genetics model for a trait determined by two loci: the phenotypic value of an individual can be expressed as the sum of the population mean, additive, dominance, plus epistatic deviations and environmental effects. Epistasis captures deviations from the additive prediction that arise from interactions among alleles at different loci. In this two-locus example, three types of epistatic effects are represented: additive-by-additive (AA), additive-by-dominance (AD), and dominance-by-dominance (DD) interactions. The expanded expression below illustrates how each component can be decomposed into the contributions of specific alleles (α), and dominance deviations (δ) at locus A and locus B.

(*Adapted from Hansen et al, 2023. under CC BY-NC license. Legend text was reworded*)

4.2.2. The inheritance of epistatic variance

Statistical components of genetic variance differ in how they are transmitted across generations. In an offspring, all additive deviations are inherited from its parents, but only half of the additive-by-additive (AA) epistatic deviations and none of the dominance deviations are passed on. This is because an individual can never inherit both alleles at a locus from the same parent; therefore, its dominance deviations are unique unless alleles are identical by descent. Such non-inherited deviations arise through recombination, continuously reshuffling genotypes and generating new combinations of alleles each generation. Although selection acts on all components of genetic variance, only selection acting on additive variance leads to a predictable, heritable response. Some epistatic deviations can also contribute to selection response when they generate additive effects in the next generation, but these contributions are transient and unpredictable, as recombination can readily break down the interacting allelic combinations that underlie them.

Earlier it was explained how pedigrees can give information for the estimation of different variance components. Here, the picture is fully presented in **Figure 4.2**, in which we can see that in case of parent-offspring, they share half of additive variance, zero of dominance variance, and a quarter of additive-by-additive variance. This can be extrapolated to other kinds of relatives. Full-sibs, for example, share additionally one fourth of dominance variance. Combining different types of relatives in a pedigree allows us to disentangle different components of genetic variance, as applied in **Chapter 3**. For instance, excess similarity between full-sibs compared to half-sibs can be used to estimate dominance variance, since half-sibs share a quarter of additive variance but none of dominance variance, and one sixteenth of additive-by-additive variance.

In the short term, the rate of evolutionary change is primarily determined by the amount of additive genetic variance available for selection. In the long term, however, evolution depends on the generation of new variation through mutation, the reshuffling of alleles by recombination and segregation, and processes such as genetic drift, effective population size, gene flow, and the structure of epistatic and dominance interactions that constrain or facilitate adaptive change.

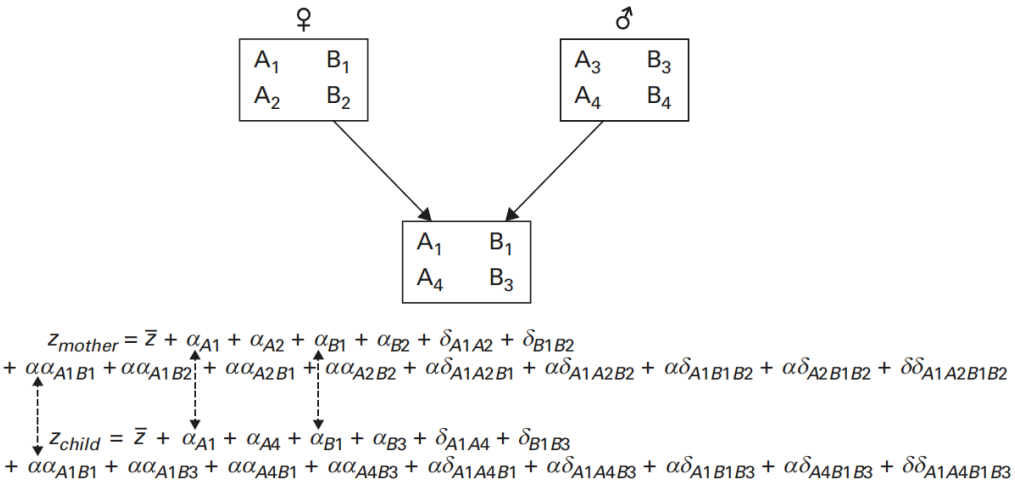


Figure 4.2. Inheritance of a two-locus trait. Two individuals mate to produce an offspring, which inherits the alleles A1 and B1 from the mother,and the alleles A4 and B3 from the father. The full genotypic values of the mother and the offspring are spelled out. The dashed arrows indicate the shared effects. These are two additive effects, αA1 and αB1, and one epistatic effect, ααA1B1. This shows that half the additive effects and one quarter of the AA epistatic effects are shared between a parent and the offspring. Also note that all the additive effects of the offspring are inherited from the two parents, while none of the dominance effects and only fractions of the epistatic effects are inherited.

(*Original figure and legend text from the source. Reproduced from Hansen et al* (2023) *under CC BY-NC license.*)

4.2.3. Disentangling epistasis

Epistasis is challenging to disentangle statistically because interactions among loci modify the apparent additive and dominance effects estimated at each locus. Even in a simple two-locus model, as shown by Table 3.6 **in Caballero (2020)**, the expression for the genotypic value of each genotype involves products of allele frequencies and interaction terms. This means that the additive, dominance, and epistasis variances are not independent: changes in allele frequencies or linkage disequilibrium continually redistribute variance among them. As a result, what is expressed as an epistatic effect in one population may appear as additive or dominance variance in another, depending on genetic background.

An *epistatic factor* (*k*) is often used to quantify the direction and strength of these interactions. When $k > 1$, the combined effect of alleles at two loci exceeds the additive expectation—referred to as *synergistic* or *positive epistasis*. Conversely, $k < 1$ indicates *antagonistic* or *negative epistasis*, where combined effects are weaker

than expected. This simple formulation highlights how epistasis can either amplify or buffer genetic effects, altering both the magnitude and the sign of variance components. Because recombination and allele frequency changes can transform these relationships over time, partitioning variance into additive, dominance, and epistatic components is inherently population-specific and dynamic.

Hence, while this chapter outlined the theoretical basis of epistasis and its contribution to genetic variance, the manifestation of these effects strongly depends on the population context in which alleles interact. Hybrid populations provide a natural testbed for this theory. When previously isolated lineages interbreed, recombination reshuffles alleles into new combinations, often disrupting co-adapted complexes and revealing hidden epistatic effects. At the same time, hybridisation alters the balance between additive and dominance variance through changes in heterozygosity. Therefore, understanding how variance components behave in natural and hybrid populations allows us to link the statistical framework of genetic variance to its evolutionary consequences under hybridisation.

4.3. Introduction

4.3.1. The evolutionary role of hybridisation

Hybrid populations play a central role in evolutionary genetics because admixture introduces novel allele combinations and increases heterozygosity, generating genetic interactions absent in parental lineages (**Vasseur et al., 2019**). This process can reveal cryptic genetic variation - phenotypically silent in parental populations but expressed in hybrid backgrounds (**Gibson & Dworkin, 2004; Rieseberg et al., 1999**). Through enhanced heterozygosity and novel epistatic interactions, hybridisation contributes to both additive and non-additive variance, thereby

expanding phenotypic variability and providing raw material for selection (**Barton, 2001**). Increased heterozygosity can also buffer populations against inbreeding and drift, enhancing fitness under fluctuating environments (**Grueber et al., 2008; Zhang et al., 2020**). Hybridisation, however, can be a double-edged sword. While it can generate heterosis and expand adaptive potential, it can also expose *Dobzhansky-Muller incompatibilities*, which are themselves manifestations of *negative epistasis* between divergent alleles (**Dobzhansky, 1936; Johnson, 2002; Muller, 1942; Wright, 1982**). Such incompatibilities contribute to segregational load and may reduce fitness in hybrid populations (**Abbott et al., 2013).** Thus, the evolutionary consequences of hybridization depend on how epistatic interactions among loci alter the partitioning of genetic variance into additive and dominance components. Understanding this balance between variance enhancement via novel epistatic combinations and fitness costs arising from incompatible interactions is crucial for predicting the long-term trajectories of hybrid populations in adaptation and speciation.

4.3.2. Genetic variance components in hybrid populations

Genetic divergence between parental populations alters the balance between additive and dominance variance in hybrids. Moderate divergence often maximizes heterosis, as both *general combining ability* (GCA, additive effects) and *specific combining ability* (SCA, dominance interactions) contribute meaningfully to hybrid performance (**Larièpe et al., 2017; Sprague & Tatum, 1942**). However, as divergence increases further, dominance variance declines because beneficial dominance interactions become rare (**Reif et al., 2007**). This shifts the genetic basis of hybrid performance toward additive variance, making GCA more predictive but weakening heterosis over successive generations. Excessive divergence therefore constrains long-term hybrid performance by depleting dominance variance and limiting the exploitation of hybrid vigour (**Utami et al., 2025**).

Inter-population crosses typically increase heterozygosity relative to intra-population crosses, but the fitness consequences are highly context-dependent. In some cases, heterozygosity restores variation and enhances fitness (**Luijten et al., 2002**). In other cases, increased heterozygosity may reduce fitness, for instance, *Physa acuta* exhibited after inter-population crosses (**Escobar et al., 2008**). These findings demonstrate that heterozygosity alone does not guarantee heterosis; outcomes depend on the evolutionary history and genetic architecture of the populations. Hybrid vigour is most likely at intermediate divergence, where heterozygosity enhances adaptive potential without triggering incompatibilities (**Wei & Zhang, 2018**).

Non-additive genetic interactions further shape divergence outcomes by altering genetic correlations across populations. Because dominance and, especially, epistasis cause allele effects to depend on the genetic background, their average effects change as allele frequencies diverge among isolated populations. As a result, allele effects become less predictable and less portable across populations, leading to reduced genetic correlations. Theory predicts that the magnitude of this decay depends on both the extent of allele-frequency divergence and the ratio of epistatic to additive variance in the ancestral population, with higher epistatic variance generating greater differences in allele effects across populations. In the absence of epistasis, genetic correlations between populations are expected to remain close to one, whereas epistatic interactions cause these correlations to fall below one as genetic backgrounds diverge (**Duenk et al., 2020**). Consistent with this prediction, stronger non-additive effects have been shown to correspond to weaker genetic correlations between populations, limiting the accuracy of genomic prediction - a method that relies on stable marker–trait associations across genetic backgrounds (**Duenk et al., 2020**). Empirical studies further support these expectations: hybrid crosses in *Tribolium castaneum* exhibit reduced fitness due to additive-by-additive

epistasis (**Drury & Wade, 2011**), while incompatibilities observed in *Drosophila* and microbial endosymbioses reveal that alleles beneficial in one genetic background can become deleterious in another (**Brucker & Bordenstein, 2013; Mack & Nachman, 2017; Walter et al., 2020**). Together, these findings illustrate how epistasis simultaneously drives the decay of genetic correlations, limits the transferability of trait predictions across populations, and contributes to reproductive isolation via Dobzhansky–Muller incompatibilities.

4.3.3. Quantifying epistasis variance from hybrid generations

Having established how hybridization reshapes additive and dominance variance, I next examine the contribution of epistatic interactions. Decomposing phenotypic variance remains the classical approach for quantifying epistasis within and among populations. In the context of our F_1–F_2 crosses, this decomposition allows estimation of how much of the total variance arises from pairwise or higher-order genetic interactions. Beyond estimating the magnitude of epistatic variance

Positive (or synergistic) epistasis occurs when the combined effect of alleles exceeds the additive expectation (captured by the *epistatic factor* $k > 1$, see "*epistatic factor*" in **Section 4.2.3**). Such interactions can increase the overall phenotypic variance and temporarily accelerate the response to selection. In hybrid populations, synergistic epistasis may underlie transgressive segregation - offspring that outperform both parents by expressing extreme phenotypes not present in either lineage (**Lotsy, 1916; Rieseberg et al., 1999**). Negative (or antagonistic) epistasis (*epistatic factor* $k < 1$) arises when allelic combinations diminish each other's effects. This interaction reduces phenotypic variance and can buffer deleterious mutations, maintaining stability of traits across environments. In hybrids, antagonistic epistasis can also generate *Dobzhansky–Muller* incompatibilities, where otherwise benign alleles reduce fitness when combined.

Because F_1 hybrids carry new combinations of divergent alleles, heterozygosity often increases dominance variance and can expose hidden epistatic effects. Recombination in F_2 progeny then reshuffles these interactions, sometimes leading to segregation variance or heterozygosity excess. Quantifying how heterozygosity contributes to non-additive variance helps separate dominance from true epistatic components.

It is important to note that apparent increases in additive variance across hybrid generations are sometimes interpreted as evidence for enhanced evolvability. However, as shown in the two-locus model (**Section 4.2.2**), epistasis can feed into the additive component: recombination converts part of epistasis (V_I) into additive variance (V_A). Thus, what seems to be an increase in heritable additive variance may simply reflect the re-expression of underlying epistatic interactions rather than a genuine, long-term gain in evolutionary potential. In other words, while epistasis can transiently increase phenotypic variance and accelerate early responses to selection, these effects do not necessarily translate into sustained additive variance across generations.

4.3.4. Research questions

Although these principles outline how dominance, epistasis, and heterozygosity can shape hybrid phenotypes, their relative contributions in natural systems remain difficult to quantify. In particular, we still lack empirical assessments of how genetic variance components - additive, dominance, and epistatic variance - change when previously isolated lineages are brought together, and whether such changes persist across generations. Hybridisation in *Arabidopsis lyrata* provides an ideal framework to address these gaps: crosses between genetically differentiated populations generate diverse combinations of alleles and reveal the extent to which non-additive interactions reconfigure the genetic architecture of traits. By

comparing natural and hybrid populations, and by following hybrids across successive generations, we can directly test how the balance of genetic variance components is reshaped and whether observed increases in non-additive variance arise from true epistatic interactions or from transient effects of heterozygosity.

In this chapter, I address three questions:

- How do genetic variance components differ between natural populations and hybrid populations of *Arabidopsis lyrata*?
- How do genetic variance components change across two successive hybrid generations?
- To what extent are changes in the non-additive genetic variance component across generations attributable to epistasis?

4.4. Data and Methodology

4.4.1. Plant material preparation

Nine plants from an *Arabidopsis lyrata* ssp. *petraea* population located in Germany (PL; 49.65N, 11.45E), were crossed with six plants from an *Arabidopsis lyrata* ssp. *petraea* located in Norway (SP; 61.41N, 8.25E) to obtain inter-population F1 offspring progenies. These F1 individuals were chosen at random to generate inter-population F2 offspring progenies. For crossing, parental individuals were propagated clonally and vernalized for nine weeks at 4°C and 12-hour daylength and subsequently transferred to the greenhouse (16-hour daylight, 12°C) until they flowered. Each individual was then crossed with two or more individuals from the other population (**Figure 3.10A**). Most crosses were reciprocal - each individual plant was used as both pollen receiver and donor - to be able to control for maternal effects. Since the same plants were used for the reciprocal crosses, the

resulting seeds are considered full-siblings, and they are termed “family”. All family members are half-siblings with the members of at least another family that share the same parent (**Table S4.1**, **Table S4.2**).

4.4.2. Phenotyping and gene expression variance analysis

Phenotyping of five fitness-related traits, namely, rosette area, leaf serration, leaf thickness, flowering time, and the number of flowers, were done exactly as described in **3.4.2.** Phenotyping fitness-related traits.

RNA extractions and quality control were performed as described in **3.4.3.** RNA extractions and transcriptome sequencing. High-quality RNA samples for **125** F1 and **225** F2 individuals were obtained. RNA was sequenced on Illumina HiSeq4000 at Novogene (Munich, Germany) following the TruSeq protocol. The specifics of the sequencing and the downstream analysis of the transcriptome was as described in **3.4.3.** RNA extractions and transcriptome sequencing.

The partitioning of gene expression variance was then performed using animal models as described in **3.4.4.** Partitioning of gene expression variance into its components. The two generations F1 and F2 were run in separate models.

4.4.3. Downstream analyses

Unless specified otherwise, all downstream analyses, such as variant calling and eQTL, were performed as described in **3.4**.

4.4.4. Modelling genetic variance using F_1 and F_2 inter-population crosses

4.4.4.1. Demographic scenario and crossing design

To examine how hybridisation and recombination change additive genetic variance (V_A) under alternative genetic architectures, two parental populations (Pop 1 and

Pop 2) were modelled. These two populations diverged independently but were otherwise identical (same demography and genetic parameters). The populations differed only in allele frequencies at one locus (single-locus model) or two loci (two-locus models).

At each bi-allelic locus, both populations retain a shared ancestral allele associated with the local adaptive background (denoted A at locus 1 and B at locus 2). Each population also carries a derived allele unique to that population (denoted A_1 in Pop 1 and A_2 in Pop 2 at locus 1; similarly, B_1 and B_2 at locus 2).

For locus 1, the ancestral allele frequencies are

p1–1 = Pr(A *in Pop 1*)
p1–2 = Pr(A *in Pop 2*),

and derived allele frequencies are

q1–1 = 1 − p1–1
q1–2 = 1 − p1–2

(Analogous notation is used for locus 2)

Hybrid generations were produced by a controlled crossing design: F_1 individuals were generated exclusively by between-population crosses (Pop 1 × Pop 2), implying an excess of heterozygotes relative to panmixia. F_2 individuals were generated by random mating among F_1 individuals, restoring Hardy–Weinberg proportions (and, in the two-locus models, restoring linkage equilibrium through recombination).

All calculations assumed neutrality during the formation of F_1 and F_2 (no mutation, selection, or drift), so allele frequencies are conserved across generations, while genotype frequencies differ due to the crossing design (**Gallais, 2003**).

4.4.4.2. Single-locus model without epistasis (baseline)

Genotype frequencies in F_1

At locus 1, four genotypes occur in F_1 because each parent contributes alleles from different populations:

G_{AA} with $F(AA) = p_{1\text{-}1}.p_{1\text{-}2}$
G_{AA1} with $F(AA_1) = p_{1\text{-}2}.q_{1\text{-}1}$
G_{AA2} with $F(AA_2) = p_{1\text{-}1}q_{1\text{-}2}$
G_{A1A2} with $F(A_1A_2) = q_{1\text{-}1}q_{1\text{-}2}$

From these, the allele frequencies in F_1 are:

$f_A^{F1} = F(AA) + 0.5*F(AA_1) + 0.5*F(AA_2) = p_{1\text{-}1}.p_{1\text{-}2} + 0.5(p_{1\text{-}2}.q_{1\text{-}1} + p_{1\text{-}1}q_{1\text{-}2}) = 1 - (q_{1\text{-}1} + q_{1\text{-}2})/2$
$f_{A1}^{F1} = 0.5*F(AA_1) + 0.5*F(A_1A_2) = 0.5(p_{1\text{-}2}q_{1\text{-}1} + q_{1\text{-}1}q_{1\text{-}2}) = 0.5(q_{1\text{-}1})$
$f_{A2}^{F1} = 0.5*F(AA_2) + 0.5*F(A_1A_2) = 0.5(p_{1\text{-}1}q_{1\text{-}2} + q_{1\text{-}1}q_{1\text{-}2}) = 0.5(q_{1\text{-}2})$

This design yields excess heterozygosity in F_1 because within-population matings are absent, so genotype frequencies deviate from panmictic expectations even when allele frequencies are fixed.

Genotype frequencies in F_2 (panmixia restored)

F_2 is formed by random mating among F_1 individuals, so genotype frequencies follow Hardy–Weinberg proportions using the F_1 allele frequencies:

G_{AA} with $F(AA) = (f_A^{F1})^2 = (2 - q_{1\text{-}1} - q_{1\text{-}2})^2/4$
G_{AA1} with $F(AA_1) = 2f_A^{F1}f_{A1}^{F1} = -0.5\,q_{1\text{-}1}^2 + q_{1\text{-}1}(1 - 0.5q_{1\text{-}2})$
G_{AA2} with $F(AA_2) = 2f_A^{F1}f_{A2}^{F1} = (1 - 0.5q_{1\text{-}1})\,q_{1\text{-}2} - 0.5\,q_{1\text{-}1}^2$
G_{A1A2} with $F(A_1A_2) = 2f_{A1}^{F1}f_{A2}^{F1} = 0.5\,q_{1\text{-}1}\,.\,q_{1\text{-}2}$

G_{A1A1} with $F(A_1A_1) = (f_{A1}^{F1})^2 = 0.25\ q_{1\text{-}1}^2$

G_{A2A2} with $F(A_2A_2) = (f_{A2}^{F1})^2 = 0.25\ q_{1\text{-}2}^2$

Under neutrality, allele frequencies are conserved:

$f_A^{F2} = f_A^{F1}$

$f_{A1}^{F2} = f_{A1}^{F1}$

$f_{A2}^{F2} = f_{A2}^{F1}$

Genotypic values, average allele effects, and V_A

Genotypic values at locus 1 were assigned as:

$G_{AA} = a$

$G_{AA1} = d_1$

$G_{AA2} = d_2$

$G_{A1A2} = d_3$

$G_{A1A1} = a_1$

$G_{A2A2} = a_2$

Average allele effects (**Cheverud & Routman, 1995; Falconer & Mackay, 2009**) were computed using the F_1 allele frequencies:

$\alpha_A = f_A^{F1}.\ a + f_{A1}^{F1}\ .\ d_1 + f_{A2}^{F1}.\ d_2$

$\alpha_{A1} = f_A^{F1}.\ d_1 + f_{A1}^{F1}\ .\ a_1 + f_{A2}^{F1}.\ d_3$

$\alpha_{A2} = f_A^{F1}.\ d_2 + f_{A1}^{F1}\ .\ d_3 + f_{A2}^{F1}.\ a_2$

Additive (breeding) values for each genotype were:

$\Lambda_{AA} = 2\alpha_A$

$\Lambda_{AA1} = \alpha_A + \alpha_{A1}$

$\Lambda_{AA2} = \alpha_A + \alpha_{A2}$

$\Lambda_{A1A2} = \alpha_{A1} + \alpha_{A2}$

$\Lambda_{A1A1} = 2\alpha_{A1}$

$\Lambda_{A2A2} = 2\alpha_{A2}$

Additive variance in each generation was calculated as the variance of Λ weighted by genotype frequencies:

$$V_{A\text{-}F1} = (p_{1\text{-}1}.p_{1\text{-}2}.\Lambda_{AA}^2 + p_{1\text{-}2}.q_{1\text{-}1}.\Lambda_{AA1}^2 + p_{1\text{-}1}.q_{1\text{-}2}.\Lambda_{AA2}^2 + q_{1\text{-}1}.q_{1\text{-}2}.\Lambda_{A1A2}^2) - (p_{1\text{-}1}.p_{1\text{-}2}.\Lambda_{AA} + p_{1\text{-}2}.q_{1\text{-}1}.\Lambda_{AA1} + p_{1\text{-}1}.q_{1\text{-}2}.\Lambda_{AA2} + q_{1\text{-}1}.q_{1\text{-}2}.\Lambda_{A1A2})^2$$

$$V_{A\text{-}F2} = (((2 - q_{1\text{-}1} - q_{1\text{-}2})^2/4).\Lambda_{AA}^2 + (- 0.5\, q_{1\text{-}1}^2 + q_{1\text{-}1}\,(1 - 0.5q_{1\text{-}2})).\Lambda_{AA1}^2 + ((1 - 0.5q_{1\text{-}1})\, q_{1\text{-}2} - 0.5\, q_{1\text{-}1}^2).\Lambda_{AA2}^2 + (0.5\, q_{1\text{-}1}\, .\, q_{1\text{-}2}).\Lambda_{A1A2}^2 + (0.25\, q_{1\text{-}1}^2).\Lambda_{A1A1}^2 + (0.25\, q_{1\text{-}2}^2).\Lambda_{A2A2}^2) - (((2 - q_{1\text{-}1} - q_{1\text{-}2})^2/4).\Lambda_{AA} + (- 0.5\, q_{1\text{-}1}^2 + q_{1\text{-}1}\,(1 - 0.5q_{1\text{-}2})).\Lambda_{AA1} + ((1 - 0.5q_{1\text{-}1})\, q_{1\text{-}2} - 0.5\, q_{1\text{-}1}^2).\Lambda_{AA2} + (0.5\, q_{1\text{-}1}\, .\, q_{1\text{-}2}).\Lambda_{A1A2} + (0.25\, q_{1\text{-}1}^2).\Lambda_{A1A1} + (0.25\, q_{1\text{-}2}^2).\Lambda_{A2A2})^2$$

Numerical exploration of parameter space

Because compact closed-form comparisons across all allele-frequency combinations are unwieldy, we evaluated $V_A(F_1)$ and $V_A(F_2)$ numerically on grids of ancestral allele frequencies (p1–1,p1–2). The generational change is summarised as

$\Delta V_A = V_A(F_2) - V_A(F_1)$

and visualised ΔV_A as heatmaps. Two dominance regimes were considered: pure additivity (h=0.5), and partial recessivity (h=0.25). As the full (p1–1,p1–2) space includes extreme, limiting cases, the heatmaps are interpreted as showing the range of possible outcomes, not their expected frequency in nature (**Clo et al., 2020**).

This single-locus model provides the baseline expectation that V_A can differ between F_1 and F_2 purely because genotype frequencies differ under the crossing design, even when allele frequencies are conserved.

4.4.4.3. Two-locus model with directional epistasis

To extend the baseline to a multilocus context, two loci were considered (locus 1: $A/A_1/A_2$; locus 2: $B/B_1/B_2$). Without epistasis, the two-locus genotypic value equals the sum of single-locus contributions:

$$G_{(A\text{-}genotype,\ B\text{-}genotype)} = G_{A\text{-}genotype} + G_{B\text{-}genotype}$$

Hybrid generations were produced with the same $F_1 \rightarrow F_2$ crossing design. In the parental populations, loci were assumed independent (no linkage disequilibrium), so two-locus genotype frequencies were obtained by products of the corresponding single-locus genotype frequencies. In F_2, random mating among F_1 restores Hardy–Weinberg proportions at each locus and generates new multilocus combinations through recombination. Additive values and V_A were computed using the same average-effects decomposition as in the single-locus model, now applied to the two-locus genotype set. This additive two-locus model isolates the effect of recombination across loci on V_A, providing the reference against which epistatic scenarios are compared (**Lande & Porcher, 2015; Abu Awad & Roze, 2018; Clo et al., 2020**).

We then introduced epistasis following the variance-decomposition framework of **Cheverud & Routman (1995)**. Each locus carried three alleles (shared ancestral allele plus two population-specific derived alleles), yielding nine possible allele combinations across loci. Epistatic deviations were assigned only to novel allele combinations created by hybridisation, specifically interactions between A_1–B_2 and A_2–B_1, which do not occur within either parental population.

Two-locus genotypic values were decomposed into (i) single-locus components, (ii) non-epistatic predicted values, and (iii) epistatic deviations. Epistatic effects scaled with allele copy number: each copy of an interacting allele contributed a

proportional ±5% modification to the relevant genotypic value (e.g., genotypes carrying two copies at both loci could experience up to a 20% modification under the same sign of interaction). Following **Cheverud & Routman (1995)**, marginal single-locus values were defined as the sum of the single-locus genotypic value and its mean epistatic effect; these marginal values were then used to compute additive values and V_A analogously to the non-epistatic model.

Two directional epistasis scenarios were evaluated: *Positive directional epistasis*, where epistatic interactions increase genotypic values on average (**Carter et al., 2005; Hansen, 2015; Le Rouzic et al., 2024**). *Negative directional epistasis*, where epistatic interactions decrease genotypic values on average (**Carter et al., 2005**). For each scenario, $\Delta V_A = V_A(F_2) - V_A(F_1)$ was computed across the same (p1–1,p1–2) grids. For comparability with the additive two-locus heatmaps, we assumed equal allele frequencies at the second locus when generating the visualisations.

4.5. Results

4.5.1. Strong correlation of non-additive variance between natural and hybrid populations

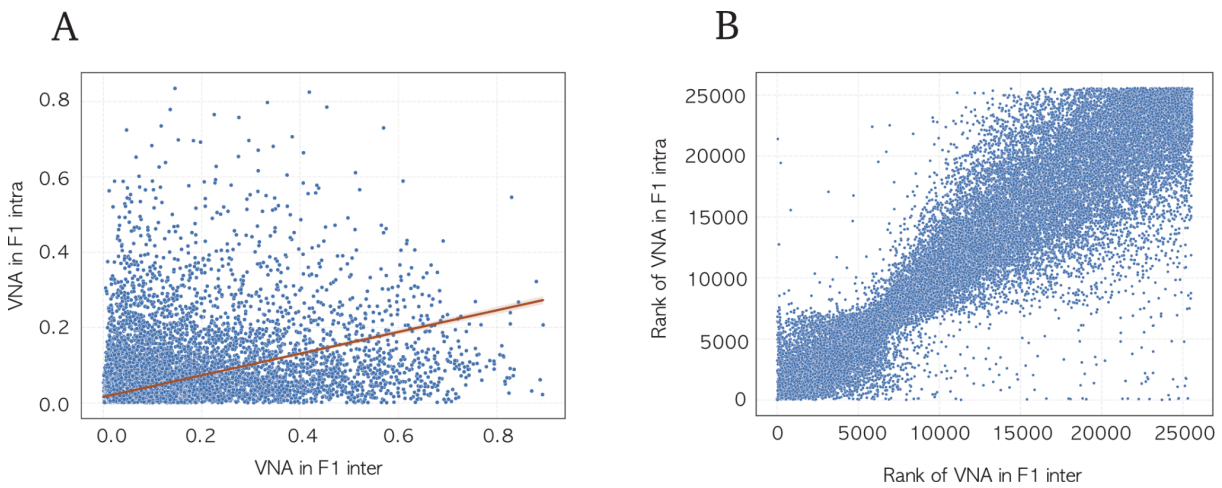


Figure 4.3. Correlation of non-additive genetic variance (V_{NA}) over total phenotypic variance between intra- and inter-population crosses in the F_1 generation. Each point represents a gene for which V_{NA} was estimated independently within (F_1 intra) and between (F_1 inter) natural populations. (**A**) Comparison of V_{NA} estimates (Kendall's $\tau = 0.74$, $p < 2.2 \times 10^{-16}$). The solid line represents the fitted regression line. As the data are highly skewed and show substantial dispersion and heteroscedasticity, (**B**) rank-rank comparison is also presented.

To assess whether non-additive variance reflects intrinsic gene properties or arises primarily from specific genetic combinations, I first compared gene-wise estimates of non-additive variance (V_{NA}) between intra-population and inter-population F_1 crosses. Gene-wise rankings were highly conserved between inter- and intra-population F1 crosses (Kendall's $\tau = 0.74$, $p < 2.2 \times 10^{-16}$; **Fig. 4.3A**). Since the estimates of non-additive variance were highly skewed, resulting in a diffuse raw scatter, rank-based comparisons were also presented (**Fig. 4.3B**). The high correspondence indicates that the magnitude of non-additive effects is not highly sensitive to the particular genetic backgrounds being combined. In this sense, the stability of V_{NA} across crosses is consistent with evolutionary constraint, insofar as the same genes repeatedly express strong non-additive effects despite changes in allele combinations and heterozygosity. In other words, genes showing strong intra-population non-additivity tend to maintain similar patterns when combined across populations, although part of the correspondence could also arise from similar allele-frequency patterns across populations, which influence the statistical expression of dominance and epistasis.

4.5.2. An increase of additive variance from F_1 to F_2 – empirical motivation for exploring epistasis

Across the transcriptome, a tendency for additive genetic variance to increase from the F_1 to the F_2 generation was observed (**Figure 4.4A**; histogram of $\Delta\log(V_A/V_G)$). Although individual genes vary in magnitude and direction, about 75% of all genes are centred below zero, indicating that recombination and segregation in the F_2 consistently expose additional additive variance relative to the structured heterozygous genotypes present in the F_1. This pattern matches theoretical expectations for hybrid populations under simple additive models, where new homozygous combinations and recombinant haplotypes reappear only in the F_2.

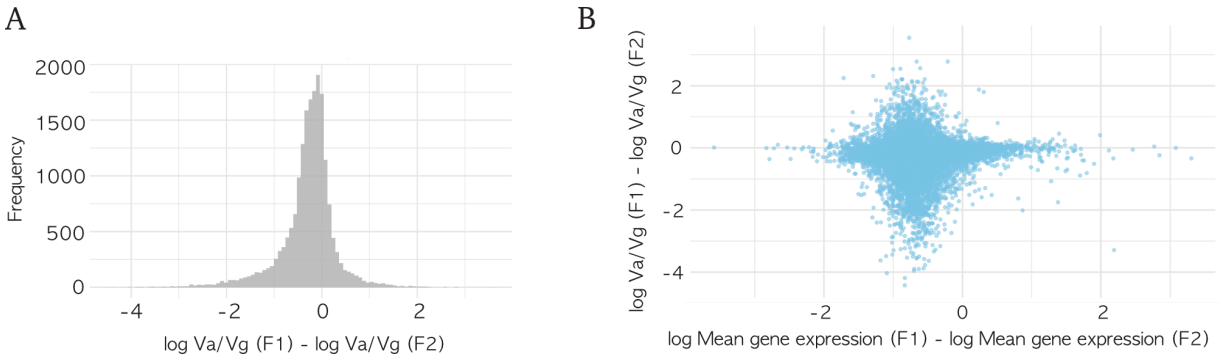


Figure 4.4. Changes in additive genetic variance and gene expression across F_1 and F_2 generations of inter-population crosses. (**A**) Histogram of genome-wide differences in additive-variance proportion between the F_1 and F_2 generations, expressed as $\log(V_A/V_G)F_1 - \log(V_A/V_G)F_2$. Each bar represents a gene for which additive (V_A) and total genetic variance (V_G) were estimated from transcriptomic data. The distribution is centred below zero, indicating that for the majority of genes, additive variance increases after recombination and segregation in the F_2. (**B**) Relationship between difference in gene expression and difference in V_A of each gene from F_1 to F_2. Each point represents a

gene, difference in V_A (log scale) is plotted against mean expression in F_1 and mean expression in F_2 (log scale). The differences in additive-variance proportion are statistically not associated with shifts in mean expression between F_1 and F_2.

However, the magnitude of the increase, together with the presence of genes showing substantial $F_2 > F_1$ shifts, raised the question of whether epistatic interactions could contribute to or amplify this pattern. Because F_1 individuals only express dominance-by-dominance interactions, whereas the F_2 generation reveals the full spectrum of epistasis (additive x additive, additive x dominance, dominance x dominance), the transition from F_1 to F_2 provides an ideal context in which directional epistasis could shape the amount of additive variance recovered.

This empirical motivation led us to develop the two complementary models presented here: first, a baseline additive model without epistasis, and second, an extended model incorporating positive or negative directional epistatic effects following **Cheverud & Routman (1995)**. These models allow us to test whether the observed genome-wide increase in V_A from F_1 to F_2 is consistent with purely additive expectations, or whether epistasis provides a mechanistic explanation for cases where the increase is unexpectedly strong, or, conversely, for genes where V_A decreases in the F_2. The theoretical predictions match these possibilities: positive epistasis amplifies the F_2 increase in V_A, whereas negative epistasis can reverse the pattern and make the F_1 more variable than the F_2.

To assess whether genes showing a decrease in additive variance from F_1 to F_2 also exhibit signatures consistent with negative epistasis at the level of mean expression, I examined the relationship between mean gene expression in the two generations (**Figure 4.4B**). When mean expression in F_2 was regressed against mean

expression in F_1, mean expression levels of the F_1–F_2 relationship did not differ significantly whether additive variance increased or decreased from F_1 to F_2.

4.5.3. Additive model as baseline theoretical expectation

To determine whether the genome-wide increase in additive variance from F_1 to F_2 necessarily requires epistatic interactions, we first examined a baseline theoretical expectation under a single-locus model without epistasis. This model isolates the effects of allele-frequency divergence and hybrid crossing design on additive variance, independent of multi-locus recombination or gene–gene interactions.

Across nearly the entire allele-frequency space, the model predicts higher additive variance in F_2 than in F_1 (**Figure 4.5**). This pattern arises because obligate between-population mating in F_1 eliminates homozygous derived genotypes, whereas recombination and segregation in F_2 restore these genotypes and increase additive diversity. Increasing dominance reduces the magnitude of the difference but does not alter its direction. These results demonstrate that the empirical $F_2 > F_1$ increase in additive variance can arise under purely additive genetic architectures and therefore does not, by itself, imply epistatic interactions.

While this model explains the general tendency for additive variance to increase from F_1 to F_2, it does not account for cases in which the increase is unusually large or, conversely, where additive variance decreases in F_2. We therefore next examined how directional epistasis modifies this baseline expectation.

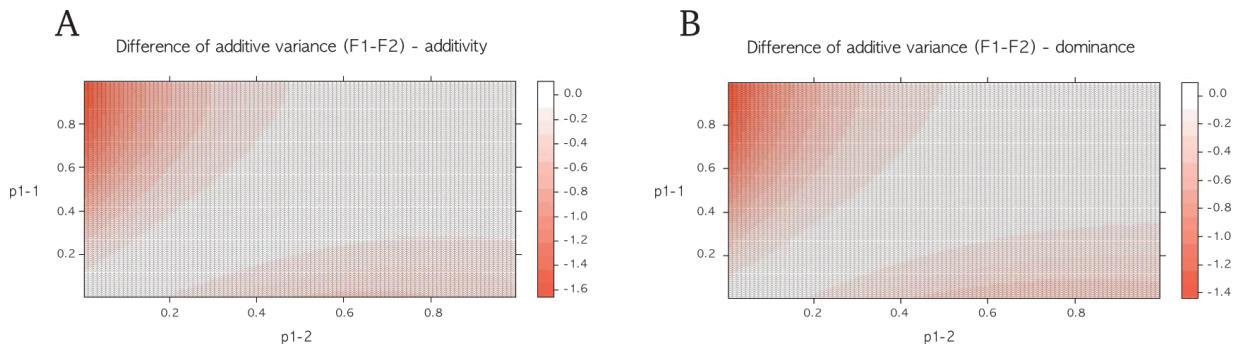


Figure 4.5. Heatmap of the ratio of additive genetic variance between F_1 and F_2 populations under a single-locus model without epistasis, as a function of ancestral allele frequencies in the two parental populations. Axes show the frequencies of the ancestral allele A at locus 1 in the two parental populations (p1-1 in population 1 and p1-2 in population 2), with derived alleles (A_1 and A_2) segregating at complementary frequencies. **(A)** Purely additive genetic effects. **(B)** Partial recessivity (h=0.25). Across most of the parameter space, additive variance is higher in F_2 than in F_1, reflecting the re-emergence of homozygous derived genotypes and the restoration of panmixia following recombination.

4.5.4. Additive variance in the presence of epistasis

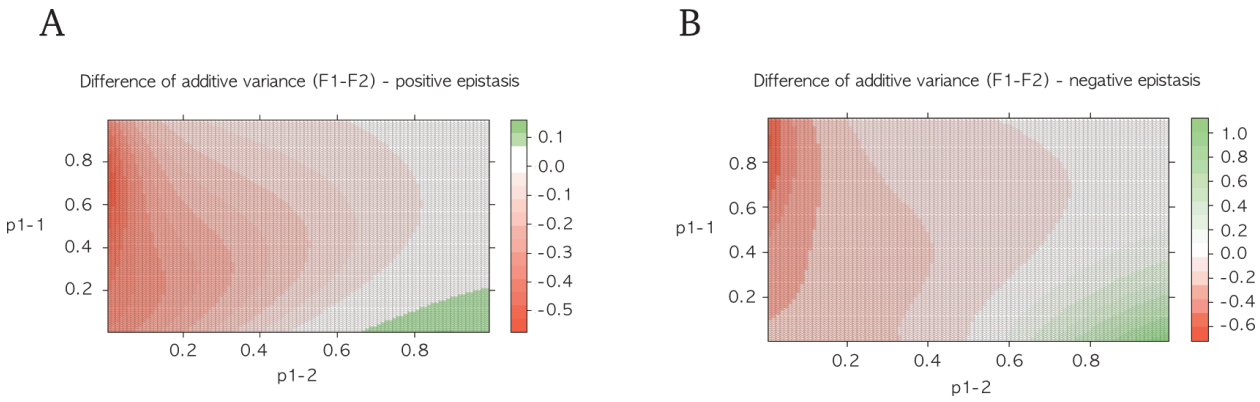


Figure 4.6. Heatmaps of the difference in additive genetic variance between F_1 and F_2 hybrid populations ($\Delta V_A = V_A(F_1) - V_A(F_2)$) under two-locus models with epistasis. Colour intensity indicates both the magnitude and direction of the difference, with negative values

(red) corresponding to higher additive variance in F_2 and positive values (green) corresponding to higher additive variance in F_1. Axes show the frequencies of the ancestral allele A at locus 1 in the two parental populations (p1-1 in population 1 and p1-2 in population 2), with derived alleles (A_1 and A_2) segregating at complementary frequencies. **(A)** Positive directional epistasis. **(B)** Negative directional epistasis.

As illustrated in **Figure 4.6**, the simulation results are consistent with the theoretical expectations outlined above. Positive directional epistasis - defined as epistatic interactions that increase genotypic values on average - tends to increase the additive variance of quantitative traits (**Carter et al., 2005; Hansen, 2015; Le Rouzic et al., 2024**). In the F_1 population, only dominance-by-dominance interactions are expressed, whereas the F_2 population allows the full spectrum of epistatic interactions to be realised, including additive-by-additive, additive-by-dominance, and dominance-by-dominance effects.

Under positive directional epistasis, this expanded set of interactions in F_2 leads to an amplification of the additive-variance difference between F_1 and F_2, while preserving the qualitative pattern observed under purely additive models (**Figure 4.6A**). In contrast, negative directional epistasis, where epistatic interactions reduce genotypic values on average, tends to decrease additive variance (**Carter et al., 2005**). Because epistatic interactions are more frequently expressed in the F_2 population, negative epistasis disproportionately reduces genetic diversity in F_2 relative to F_1. Under these conditions, the F_1 population can exhibit higher additive variance than the F_2, a reversal of the baseline expectation that is unlikely to occur in the absence of epistasis (**Figure 4.6B**). These simulation results confirm that the direction of epistatic interactions critically determines whether recombination amplifies or attenuates additive genetic variance in hybrid populations.

4.6. Discussion

Understanding how hybridisation reshapes the genetic architecture of traits requires disentangling the relative contributions of additive, dominance, and epistatic interactions. The results presented here demonstrate three main insights. First, non-additive variance is conserved across genetic backgrounds. Second, while additive expectations explain the general tendency for additive variance to increase from F_1 to F_2, they do not account for the magnitude and heterogeneity of this increase across genes. Third, the sign of directional epistasis fundamentally determines whether recombination elevates or depresses additive variance in hybrid populations. Together, these findings elucidate when hybridisation enhances evolutionary potential and when it uncovers hidden constraints.

First, the strong correlation in non-additive variance between natural populations and F_1 hybrids suggests that dominance and epistatic components often reflect inherent properties of gene-regulatory systems, rather than interactions that arise solely from allele combinations unique to hybrid backgrounds. This interpretation builds on earlier theoretical work showing that while inter-population crosses increase the frequency of heterozygous genotypes and thereby enhance the detectability of non-additive variance, they nevertheless reveal patterns that are strongly correlated with those expected from intra-population crosses (**Clo & Opedal, 2021; Takou et al., 2022**). Here, by directly quantifying non-additive variance in both intra- and inter-population contexts, we provide empirical support for this prediction. The persistence of gene-specific non-additive variance across these backgrounds indicates that regulatory dominance and epistasis are not merely artefacts of crossing design, but reflect stable, system-level properties of gene-regulatory networks. Regulatory robustness and molecular buffering have been reported across a wide range of systems (**Félix & Barkoulas, 2015; Garfield et al., 2013; Gibson & Dworkin, 2004; Macneil & Walhout, 2011; Payne & Wagner,**

2014), and such robustness implies that non-additive variance can shape evolutionary trajectories even in the absence of demographic perturbations such as bottlenecks, which are traditionally assumed to convert non-additive variance into additive variance via changes in allele frequencies and genetic background. Instead, the persistence of non-additive effects across genetic backgrounds suggests that gene-regulatory networks encode predictable interaction structures that recombination or selection may amplify, suppress, or redistribute.

Second, the genome-wide increase in additive variance from F_1 to F_2 mirrors the classical expectation that recombination restores homozygous genotypes absent from the F_1 (**Abu Awad & Roze, 2018; Clo et al., 2020; Lande & Porcher, 2015**). However, the increase for many genes may point to additional mechanisms. Recombination not only reshuffles additive effects but can also unmask cryptic epistatic interactions that were buffered in the F_1. In doing so, recombination can redistribute genetic variance into additive components without necessarily changing mean expression. This process aligns with long-standing theory showing that epistatic variance can be transformed into additive variance when genetic backgrounds change (**Cheverud & Routman, 1995**). Hybridisation provides a natural setting for such unmasking. In F_1 individuals, many translational and regulatory perturbations are masked due to heterozygosity and network robustness mechanisms, such as compensatory regulation or chaperone activity, that conceal genetic perturbations (**Gibson & Dworkin, 2004; Rieseberg et al., 1999**). As recombination disrupts combinations of alleles and regulatory interactions in the F_2, the buffering capacity diminishes, revealing interactions that manifest as increased additive variance.

Third, the theoretical models underscore that the sign of epistasis fundamentally changes the expected relationship between F_1 and F_2 variance. Under no epistasis, F_2 populations nearly always display increased additive variance due to the return of

recombinant and homozygous genotypes. Under negative epistasis, recombination introduces allele combinations that reduce trait values, making F_2 less variable than F_1 in parts of parameter space. This phenomenon is consistent with the role of antagonistic epistasis in generating Dobzhansky-Muller incompatibilities (**Bateson, 1909**; **Dobzhansky, 1937; Muller, 1942**) and hybrid breakdown observed in both plants and animals (**Brucker & Bordenstein, 2013; Mack & Nachman, 2017; Walter et al., 2020**). Furthermore, in the present data, eQTL mapping (**Figure S4.2**) indicates that expression variation is shaped by a highly distributed regulatory architecture, with abundant *trans*-acting associations but no evidence for a small number of major-effect regulatory loci. Rather than being driven by discrete *trans* hotspots, regulatory divergence appears polygenic, involving many loci of modest effect. This pattern may be less consistent with a model dominated by a small number of large-effect Dobzhansky–Muller incompatibilities, but is compatible with the hypothesis that recombination in the F_2 generation exposes numerous weak interaction effects across the genome. Studies have shown that when regulatory interactions are synergistic or aligned with selection, recombination can generate novel expression profiles, transgressive phenotypes, and enhanced adaptive potential, as documented in diverse hybrid systems (**Lippman & Zamir, 2007**; **Grant & Grant, 2019; Patton et al., 2022; Rieseberg et al., 1999**). Conversely, when regulatory interactions are antagonistic or misaligned with the fitness landscape, the same polygenic architecture can reduce additive variance, constrain evolvability, and contribute to hybrid breakdown or outbreeding depression without producing discrete, catastrophic failure points (**Abbott et al., 2013; Drury & Wade, 2011; Rahnamae et al., 2025**). Thus, whether hybridisation can either enhance or restrict evolutionary potential depends on the direction and context of gene-gene interactions.

Importantly, the empirical allele-frequency structure of the system constrains which theoretical regimes are most relevant. In the empirical data, ancestral and derived alleles segregate at broadly similar frequencies across parental populations, placing the empirical system close to the diagonal of the theoretical parameter space explored in Figures 4.5 and 4.6, where ancestral allele frequencies in the two parental populations are similar. In this region, additive-only models already predict a consistent excess of additive variance in F_2 relative to F_1, while strong reversals driven by negative epistasis are expected to be rare and localised.

Lastly, the interplay between epistasis and robustness can provide an additional interpretive framework. Genetic networks can maintain stable outputs despite underlying genetic variation, through redundancy, compensatory regulation, or protein-folding chaperones, which enables mutational robustness and accumulation of cryptic variation (**Gibson & Dworkin, 2004**). Hybridisation interacts with this robustness in two ways: (1) F_1 often benefits from heterozygosity-based buffering, as deleterious alleles are masked by alternative alleles and regulatory compensation stabilises expression outputs (**Schneemann et al., 2022**); (2) recombination in the F_2 disrupts these buffering mechanisms, exposing hidden interactions. Whether this exposure increases or decreases additive variance depends on the sign of epistasis, explaining why some F_2 lineages exhibit inflated additive variance while others show contraction. Thus, the empirical and modelling results converge on a common interpretation: the F_1 can represent a buffered genetic background, whereas the F_2 can reflect a de-buffered state, in which cryptic epistasis becomes phenotypically visible.

4.7. Conclusion

This chapter examines how hybridisation reshapes the structure of genetic variance by integrating empirical analyses of gene expression with theoretical models of

increasing genetic complexity. By first comparing variance components within and between natural populations, it is shown that non-additive variance is a stable, gene-specific feature of the transcriptome that persists across genetic backgrounds. This conservation indicates that dominance and epistatic effects largely reflect intrinsic properties of gene-regulatory networks, rather than interactions that arise solely as artefacts of hybrid crossing designs. Extending this framework to hybrid generations, a genome-wide increase in additive genetic variance from F_1 to F_2 is observed. This pattern is consistent with classical expectations that recombination and segregation restore homozygous and recombinant genotypes absent from the obligately heterozygous F_1 generation.

The magnitude and heterogeneity of the F_1–F_2 contrast across genes motivated the development of explicit theoretical models to assess whether additive architectures alone are sufficient to account for the observed variation. The modelling results show that recombination establishes a robust baseline expectation of higher additive variance in F_2 populations, even in the absence of epistasis. Directional epistatic interactions do not generate this baseline effect, but instead modulate its magnitude, either amplifying or attenuating the F_1–F_2 difference depending on their sign. Integrating these theoretical predictions with the empirical allele-frequency structure of the system indicates that the observed genome-wide patterns primarily reflect the reorganisation of standing genetic variation, with epistatic interactions shaping variance at specific loci rather than driving uniform shifts across the transcriptome.

4.8. Supplementary Materials

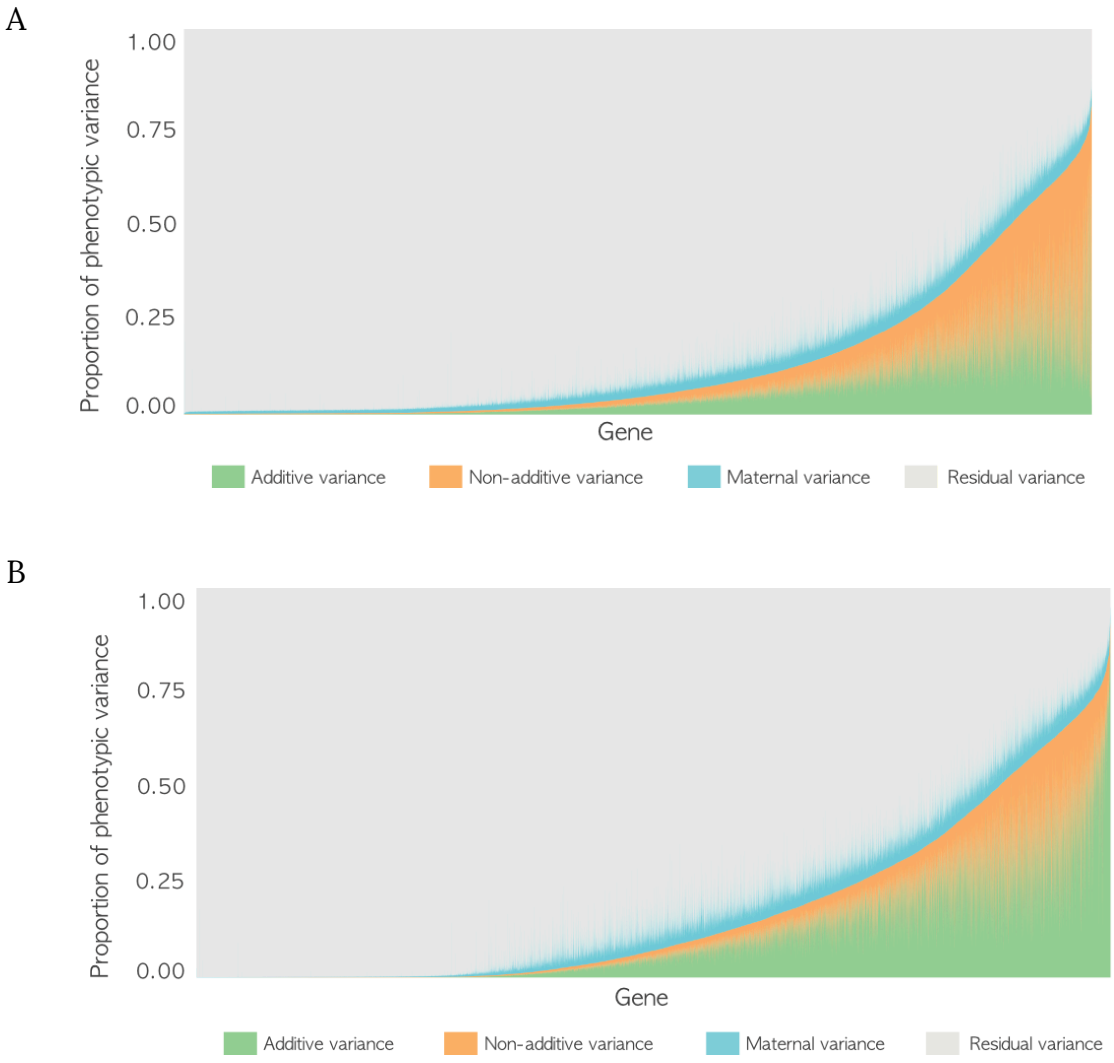


Figure S4.1. The total phenotypic variance of F_1 and F_2 inter-population of each gene was partitioned into non-additive, additive, maternal, and residual components of variance. (A) The total phenotypic variance of each of the 25,769 genes in F_1 included in the analysis. **(B)** The total phenotypic variance of each of the 25,781 genes in F_2 included in the analysis. The genes were ordered by increasing values of the fraction of total phenotypic variance in transcript level that was attributed to genetic variance (V_G).

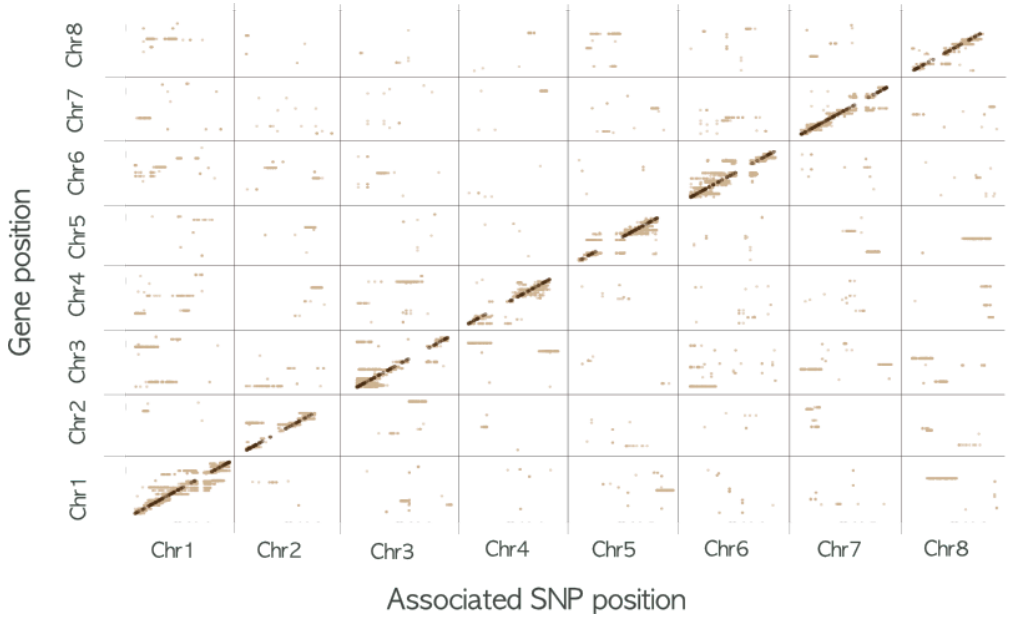


Figure S4.2. Genome-wide distribution of *cis*- and *trans*-eQTL in the hybrid population. Each point represents a significant association between a genetic variant (x-axis; SNP position) and a gene expression trait (y-axis; gene position), plotted across the eight major scaffolds of the *Arabidopsis lyrata* genome. Points along the diagonal correspond to local (*cis*) associations, defined here as variants located within ±5 kb of the target gene. Off-diagonal points represent *trans*-acting associations, where variants map to genomic regions distant from the regulated gene. *Trans*-acting associations were detected, these were broadly dispersed across the genome rather than concentrated at a small number of loci. No clear *trans*-eQTL hotspots or individual variants influencing large numbers of genes were observed. Instead, expression variation appears to be shaped by many loci of modest effect, consistent with a polygenic regulatory architecture.

Table S4.1. Information about the inter-population F1 families used in the analysis. Each full-sibling family was generated based on the scheme presented in Figure 3.10. For each full sibling family, the Plech (PL) and Spiterstulen (SP) sequence ID numbers, as described in **Takou et al. (2021)** and annotated in ENA archive repository, are provided (PRJEB49153). The number of offspring used in the final analysis is shown.

Family	Parent 1	Parent 2	Number of family members
2	11a	70536	4
3	70536	80936	32
4	10a	70536	19
5	6a	70536	15
6	1a	70538	22
7	2a	70539	1
10	1a	70537	3
12	70539	80936	14
13	10a	70539	14
14	1a	70539	1

Table S4.2. Information about the inter-population F2 families used in the analysis. Each full-sibling family was generated based on the scheme presented in Figure 3.10.

A. For each full sibling family, the parents ID from inter-population F1 individuals were provided. The origin of these parents can be found in **Table S4.2.B.** The number of offspring used in the final analysis is shown.

Family	Parent 1	Parent 2	Number of family members

15	531	538	60
16	447	490	6
17	471	519	14
18	482	540	10
19	490	528	24
20	406	449	14
21	520	544	7
22	482	541	11
23	531	547	11
24	403	453	16
25	495	522	12
26	442	529	10
27	403	515	4
28	529	536	5
29	484	515	3
30	470	505	2
31	425	515	7

33	442	462	4
34	449	497	1

B. Origin of the F1 individuals used to generate F2 generation as shown in **Table S4.2 A**. For each F1 parent, the Plech (PL) and Spiterstulen (SP) sequence ID numbers, as described in **Takou et al. (2021)** and annotated in ENA archive repository, are provided (PRJEB49153).

F2 parentID	F1 parent 1	F1 parent 2
531	1a	70537
447	10a	70536
471	70539	80936
482	70536	80936
490	1a	70538
406	70536	80936
520	1a	70538
403	11a	70536
495	1a	70537
442	10a	70536
529	2a	70539

484	6a	70536
470	70539	80936
425	70536	80936
449	10a	70536
538	6a	70536
519	1a	70538
540	1a	70538
528	70536	80936
544	10a	70539
541	1a	70538
547	1a	70539
453	6a	70536
522	6a	70536
515	10a	70539
536	10a	70536
505	10a	70539
462	1a	70537

497	70539	80936

Table S4.3. Normalised gene expression matrix for F_1 inter-population individuals. Rows correspond to annotated genes and columns to individual samples; values represent normalised expression levels used as input for downstream statistical and variance-component analyses. For more details, refer to section **3.4.3.**

Available at 10.6084/m9.figshare.29986699

Table S4.4. Normalised gene expression matrix for F_2 inter-population individuals. Rows correspond to annotated genes and columns to individual samples; values represent normalised expression levels used as input for downstream statistical and variance-component analyses. For more details, refer to section **3.4.3.**

Available at 10.6084/m9.figshare.29986699

Table S4.5. Phenotypic trait data for F1 inter-population individuals. List of sample identifiers together with measurements of rosette size (millimeter), leaf serration (no unit), leaf thickness (millimeter), flowering time (days), and number of flowers (count), and was used for downstream quantitative genetic analyses. For more details, refer to section **3.4.2.**

Available at 10.6084/m9.figshare.29986699

Table S4.6. Phenotypic trait data for F2 inter-population individuals. List of sample identifiers together with measurements of rosette size (millimeter), leaf serration (no unit), leaf thickness (millimeter), flowering time (days), and number of flowers (count), and was used for downstream quantitative genetic analyses. For more details, refer to section **3.4.2.**

Available at 10.6084/m9.figshare.29986699

PART III -

Synthesis and General Discussion

For Darwinian biology the organism is the nexus of the internal and external forces. It is only through natural selection of internally produced variations, which happen to match by chance the externally generated environmental demands, that what is outside and what is inside confront each other.

- Richard C. Lewontin (2009), Biology Under the Influence: Dialectical Essays on Ecology, Agriculture, and Health.

5 Understanding Adaptation and Adaptive Potential

5.1. Adaptive Potential

Accelerating environmental change in the Anthropocene is imposing unprecedented selective pressures on natural populations. As interest grows in interventions designed to promote adaptive responses, empirical data and evolutionary models provide essential tools for predicting population responses to future environments, identifying drivers of rapid adaptation, and evaluating the consequences of management strategies. At the beginning of this thesis, I therefore asked a fundamental question: *what determines the adaptive potential of natural populations*? This question lies at the core of evolutionary biology, because understanding whether populations can persist and diversify under environmental change requires identifying both the sources of variation and the constraints that shape evolutionary responses (**Olson-Manning et al., 2012)**.

Adaptive potential has traditionally been inferred from phenotypic responses to selection. Within the quantitative genetics framework, adaptation can be described as directional change in trait means driven by additive genetic variance and covariance in traits related to fitness (**Lai et al., 2024; Shaw & Shaw, 2014**). This framework has been highly successful in explaining short-term evolutionary responses and population tracking of environmental gradients. However, phenotypic evolution is ultimately underpinned by molecular architecture. Genes operate within complex regulatory networks and gene expression patterns, interactions, and inheritance determine how genetic variation is translated into phenotypic variation. With the advent of transcriptomic data and genome-wide quantitative approaches, it has become possible to partition molecular traits into

additive, dominance, and epistatic components, revealing constraints and capacities that are not apparent from phenotypes alone.

Populations do not adapt simply because they possess genes or traits in isolation, but because interacting layers of genetic architecture, gene expression, ecological traits, and environmental context jointly shape the flow of heritable variation across generations. At molecular level, adaptation is shaped by how genomic and regulatory variation modulates gene expression, facilitating population persistence and adaptive divergence (**Hämälä et al., 2022; Pavey et al., 2010**). Moreover, environmental exposure can generate heritable phenotypic and gene expression variation across generations, indicating that the environment can actively contribute to the generation of heritable variation rather than solely filtering existing genetic diversity (**Lin et al., 2024**). Through these interactions, ecological and evolutionary processes become tightly coupled: trait changes influence demography and environments which in turn reshape selection on genetic and regulatory architectures (**Yamamichi, 2022**) (**Figure 5.1**).

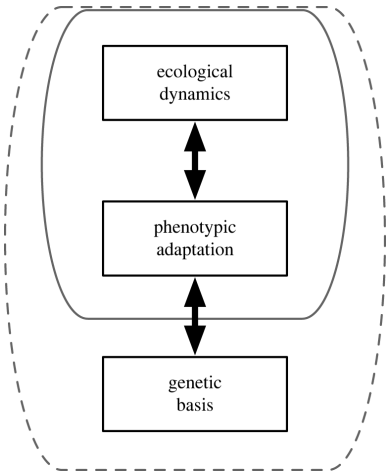


Figure 5.1. Conceptual framework of eco-evolutionary feedback. While eco-evolutionary dynamics have traditionally focused on feedback between ecology and phenotype (solid line), explicitly incorporating the genetic basis of traits (dashed line) can show how ecological change can restructure the targets of selection. *Original figure from Yamamichi, 2022. Legend text was reworded.*

Adaptive potential can thus be viewed as an emergent property of populations that can be inferred from phenotypic variation, genetic diversity, or molecular polymorphism, but is not fully captured by any single level alone. Addressing this multi-layered concept requires integrating complementary perspectives, which together structure this thesis. First, I investigated the genetic basis and adaptive significance of secondary dormancy in *Arabidopsis thaliana* and its interaction with primary dormancy across climatic gradients. Second, I examined how additive and non-additive components of genetic variance shape gene expression in natural populations of *Arabidopsis lyrata*, and what these patterns imply for evolutionary potential. Third, I explored how hybridisation redistributes additive and non-additive genetic variance and how these changes influence adaptive capacity in recombinant populations.

Although these chapters span different levels of biological organisation, from ecological traits to transcriptional regulation and genetic interaction architecture, they converge on a single question: *how does variation become available, expressed, and evolutionarily relevant*? The central message of this thesis is to consolidate that *adaptive potential resides not simply in the amount of genetic variance present, but in how biological systems generate, filter, buffer, and expose that variation across environments and evolutionary timescales.*

5.1.1. Persistence and adaptive divergence

The first component of this thesis addresses how trait variation can predict adaptive potential from an *ecological perspective*. Natural populations of *Arabidopsis thaliana* show structured variation along the latitudinal cline in both primary and secondary dormancy, reflecting adaptation to the local climates. Populations from harsher or more unpredictable environments tended to exhibit stronger or more secondary dormancy responses, illustrating *ecological filtering* imposed by

temperature regimes, precipitation, and seasonal predictability. Because dormancy directly governs germination timing, survival, and reproductive success, its variation provides a *forward-looking measure of adaptive potential*. Populations with more responsive dormancy strategies were better positioned to persist under shifting climatic regimes. Importantly, the genetic basis of dormancy can interact with environmental cues and hormonal pathways, embedding adaptive potential within a dynamic interplay between plasticity, local adaptation, and environmental responsiveness.

It is crucial to distinguish *persistence from adaptive divergence*. When populations encounter novel or stressful environments, phenotypic plasticity can rapidly shift physiology, behaviour, or life-history timing to maintain fitness, allowing populations to persist long enough for selection on genetic variants to occur (e.g. Baldwin effect or plasticity-mediated persistence) **(Hendry, 2016; Morris, 2014; Westneat et al., 2019**). However, such plastic responses do not, in themselves, generate lasting evolutionary change. When environmental differences are removed, for example, plastic phenotypic differences can disappear, revealing little or no underlying genetic divergence (**Hodgins-Davis & Townsend, 2009**). Dormancy illustrates this distinction particularly clearly. As a large-effect, switch-like life-history trait, dormancy can generate rapid population-level shifts by altering the timing of life-history transitions rather than spatial distributions. By altering exposure to selection, effective generation time, and demographic dynamics, dormancy can strongly influence population persistence. As a classic bet-hedging strategy, it maximises long-term fitness by reducing temporal variance in reproductive success, even if mean fitness within a given season is reduced (**Bruijning et al., 2020**). Because dormancy can operate through phenotypic plasticity, these ecologically consequential responses may occur with little immediate genetic change. Whether such responses translate into long-term

adaptive divergence depends on the presence of heritable variation in dormancy traits, or in the parameters governing their reaction norms (**Joschinski & Bonte, 2020; Simons, 2011**). Plasticity may therefore either buffer populations from selection, slowing divergence, or facilitate evolution when plastic responses align with underlying additive genetic variance (**Fitzpatrick, 2012; Radersma et al., 2020**).

Together, these insights underscore a broader principle: ecologically important variation may drive rapid responses and enhance persistence, but long-term adaptive divergence requires heritable variation that selection can act upon consistently, and coupling to other traits relevant for local fitness. Understanding such distinction has been a central aim of genetics; and over the past century, quantitative genetic approaches have provided the primary framework for disentangling heritable and environmental contributions to trait variation, both in applied contexts such as breeding and in theoretical studies of evolutionary potential.

5.1.2. Constraints and potentials

The second component of this thesis moves to the molecular level, examining how gene expression varies and is inherited across natural populations of *Arabidopsis lyrata*. Here, the focus shifts from the mere existence of variation to its *evolutionary usability*. Gene expression represents a key intermediate phenotype linking genetic variation to organismal function, and treating expression levels as quantitative traits extends the classical Fisherian framework into the molecular domain. Rather than focusing on gene expression *per se*, this perspective uses expression as a system for dissecting the sources, constraints, and evolutionary accessibility of additive and non-additive genetic variance.

Empirical evidence from gene expression showed that the balance between additive and non-additive variance reflected both the functional importance of genes and

their evolutionary history. Genes with elevated non-heritable expression variance were highly interconnected, long, or involved in core cellular and regulatory processes; and they were subject to strong purifying selection acting to constrain changes in protein sequence. Supporting this view, it has been shown that essential genes evolve more slowly at the protein level than non-essential genes in bacteria models (**Jordan et al., 2002**) and in *Arabidopsis* system (**Lloyd et al., 2015**). It is interesting to note that such reduced evolutionary rate of essential plant genes can be largely due to stronger constraint (subject to purifying selection) rather than differences in the underlying mutation rates (**Monroe et al., 2022**). These observations together indicate that functional constraint and regulatory architecture shape how variation is expressed, transmitted, and exposed to selection. Hence, having an abundant amount of variation does not necessarily translate into evolutionary accessibility.

The next question is: *how do additive and non-additive components of genetic variance arise, persist, and change across evolutionary time*? From a theoretical perspective, adaptive divergence provides one mechanism by which genetic variance can be restructured. As populations diverge under different selective regimes, genetic backgrounds become more homogeneous within populations but differentiated between them. During this process, epistatic and dominance effects can be transiently converted into additive variance, temporarily increasing heritability and facilitating short-term adaptive responses, consistent with Fisher's infinitesimal model. In this sense, adaptive divergence can be viewed as a process that *both reveals and consumes additive genetic variance*. Because selection acts most efficiently on additive effects, alleles contributing additively to fitness are more likely to fix during adaptation, whereas non-additive effects remain context-dependent and transient; however, after population admixed, continued selection can erode additive variance as favourable alleles approach fixation (**Liu et**

al., 2020) (Figure 5.2). This cycle of asymmetry can serve as one mechanistic explanation for a long-standing puzzle: why additive genetic variance in polygenic traits does not decrease despite the ubiquity of epistatic interactions observed in functional studies (**Merilä & Sheldon, 1999; Liu et al., 2020; Sackton & Hartl, 2016**).

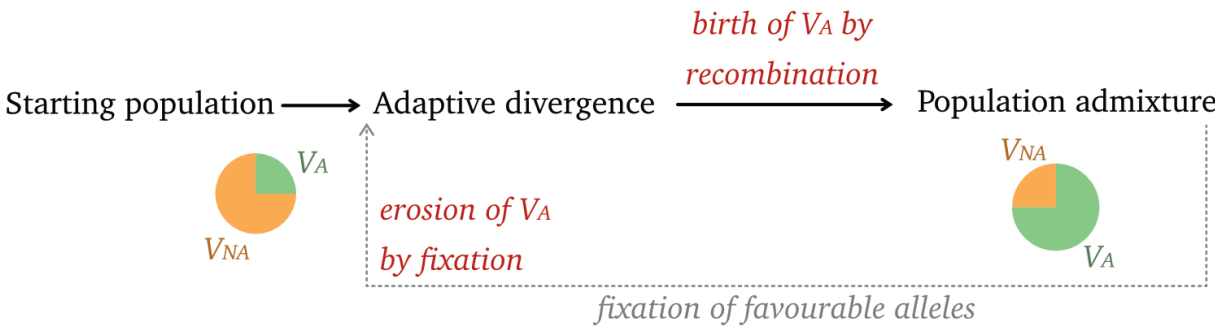


Figure 5.2. Conceptual schematic illustrating a birth–death cycle of additive genetic variance across evolutionary time. Adaptive divergence in isolated populations can lead to positive selection acting on different loci in different environments. During this phase, alleles with additive effects are more likely to fix, whereas non-additive effects remain largely hidden. When diverged populations subsequently admix, recombination can expose these previously fixed differences as additive genetic variance in the hybrid population, resulting in an increase in additive variance (V_A). Over time, continued selection erodes V_A as favourable alleles approach fixation, consistent with Fisher's fundamental theorem. *Redrawn from* ***Liu et al.* (2020),** *legend text was reworded.*

Finally, the relative contributions of additive and non-additive genetic variance to adaptive potential have been debated for more than a century. Although non-additive variation can contribute to adaptation under different scenarios (**Carroll et al., 2003**), it is often context-dependent and less predictably transmitted across generations, and adaptive evolution preferentially fixes the variation that responds predictably to selection. However, continued selection may deplete

additive variance at times, leaving behind constrained, buffered architectures dominated by non-additive effects. Hence, to this end, it is safe to say that non-additive variance should not be dismissed as mere statistical noise. To my understanding, loci that strongly influence core traits are subject to higher functional constraints. As a result, mutations affecting these loci are efficiently removed by purifying selection, limiting the accumulation of additive genetic variance. Thus, systems under strong functional constraint evolve robustness through purifying selection, whereas peripheral components retain greater evolvability. *Therefore, the evolutionary significance of non-additive variance lies primarily in how it reflects underlying constraints and latent potential rather than in its direct contribution to adaptive trajectories.*

5.1.3. The landscape under novel genomic contexts

While the dynamics described above can be inferred from theory and patterns of molecular constraint, they can become more evident when genetic backgrounds are actively reshuffled. The third component of this thesis therefore examines gene expression variance under hybridisation, a context in which novel combinations of alleles can expose previously hidden components of genetic architecture.

Transcriptomic analyses of F_1 and F_2 hybrids revealed substantial dominance and epistatic variance across genes. The empirical increase in additive variance from F_1 in the current data supported a long-standing theoretical view that recombination converts epistatic variance into additive variance (**Goodnight, 1995; Neiman & Linksvayer, 2006**). At the same time, the strong correspondence in non-additive expression variance between parental populations and hybrids indicated that such effects are intrinsic features of regulatory architecture rather than transient by-products of admixture.

More broadly, hybridisation can reshape genetic variance by combining divergent alleles from distinct parental backgrounds, generating phenotypic novelty and exposing cryptic genetic effects that were previously buffered within parental populations (**Paaby & Rockman, 2014**). For example, in *Triticum dicoccoides*, F_1 hybrids showed no fitness change relative to parents when both parents originated from core populations, and consistently higher fitness than the peripheral parent in core–periphery crosses, with fitness remaining stable across F_2 and F_3 generations (**Volis et al., 2016**). However, such restructuring does not always increase fitness or evolvability. Much of the resulting variation remains context-dependent, and epistatic interactions lacking directional effects contribute little to long-term adaptive change. Whether hybridisation enhances or constrains adaptive potential therefore depends critically on the underlying regulatory networks and on the nature of epistasis: positive epistasis can expand phenotypic space and generate *transgressive* phenotypes, whereas negative epistasis can constrain variation and contribute to *canalisation* (**Carter et al., 2005; Hansen et al., 2023**).

With respect to gene expression variation, hybridisation highlights the extent to which gene expression is shaped by non-additive regulatory interactions. When regulatory elements that have co-evolved within populations are combined across divergent genetic backgrounds, expression outcomes often depend on specific *cis–trans* combinations rather than allele dosage, producing discrete or highly variable expression states rather than intermediate phenotypes (**Moran et al., 2021**). While such restructuring can generate striking phenotypic diversity, it may not reliably increase additive genetic variance or evolvability, because expression outcomes remain context-dependent and poorly heritable. It therefore remains a question whether hybridisation simply reshuffle genetic variance or actually restore the conditions required for sustained adaptive divergence. In this sense, the study revolves back to a broader synthesis message: *while recombination can reveal*

hidden variation and generate short-term phenotypic novelty, long-term adaptive divergence depends on the directionally responsive variation, not on the mere reshuffling of genetic variation through hybrid recombination.

5.2. Going Forward

5.2.1. Adaptive potential across biological levels

In this thesis, we have seen how different classes of biological variation contribute to adaptation: *Life-history trait variation* is shaped primarily by environmental filtering, *expression variation* by regulatory architecture, and *hybrid variation* by epistatic interactions and recombination. Across systems, adaptive potential emerges not from the quantity of variation alone, but from how that variation is structured, filtered, and transmitted across generations. Several questions remain unresolved and motivate future research; they will be essential for understanding not only how populations adapt, but also when and why adaptation fails.

First, under what ecological and genetic conditions plastic responses align with additive genetic variance, enabling short-term flexibility to translate into long-term adaptive change; and how epistasis and regulatory architecture interact to modulate adaptive potential across different environmental and demographic contexts. From my point of view, it is useful to link genetic variation and genetic architecture explicitly to adaptive dynamics in changing environments over time, moving beyond static theoretical expectations toward context-dependent and mechanistic views of evolutionary potential (**Bomblies & Peichel, 2022; Shaw & Shaw, 2014**). As it has long been established, Mendelian and Fisherian models do not represent competing explanations, but sequential phases along a shared adaptive trajectory: rapid initial responses driven by large-effect or plastic mechanisms give way to slower or constrained evolution governed by polygenic architecture. As populations approach fitness optima, the contribution of large-effect alleles to

adaptation diminishes and evolutionary change becomes increasingly polygenic, dominated by many small-effect variants (**Höllinger et al., 2019**). This transition constrains the role of large-effect variants in core traits, favouring infinitesimal dynamics in which adaptation proceeds through subtle, distributed allele-frequency changes (**Parsons & Ralph, 2024; Thornton, 2019**).

Secondly, when genomic admixture promotes sustained evolutionary change, and when it merely exposes transient or context-dependent variation without long-term adaptive consequence. Genomic features such as reduced recombination near selected loci or dominance effects described by Haldane's sieve can further limit the fixation of beneficial alleles, constraining adaptive responses even when genetic variation is abundant (**Haldane, 1927; Hvala et al., 2018**).

Finally, an unresolved question is under what conditions *robustness* facilitates adaptive evolution, and when it instead constrains evolutionary trajectories by buffering variation from selection. Robustness operates across biological levels: dormancy buffers populations against environmental uncertainty, gene regulatory networks buffer transcriptional noise, and hybrid genotypes buffer incompatibilities between divergent genomes. Such buffering permits the accumulation of cryptic genetic variation, which can be released when buffering breaks down under strong environmental stress, in later hybrid generations, or in different regulatory contexts. Interestingly, **Deshpande and Fronhofer (2022)** showed that the genetic architecture of key traits shapes both the speed and predictability of adaptation during range expansion: gene-regulatory networks with reduced robustness and increased sensitivity to mutation accelerate local adaptation, but at the cost of less predictable evolutionary outcomes.

5.2.3. Predictability and uncertainty

A long-standing question in evolutionary biology is whether limits to prediction arise primarily from methodological constraints or from intrinsic unpredictability in natural systems (**Matsuno, 1992; Nosil et al., 2020**). The results of this thesis suggest that evolutionary change is only partially predictable, not simply because of stochasticity, but because adaptive outcomes emerge from interacting sources of constraint, plasticity, and historical contingency. At the ecological level, dormancy variation shows *relatively predictable* associations with climatic gradients. At the molecular level, inheritance of gene expression is *partly predictable* due to functional constraint and stabilising selection, yet remains influenced by stochastic and context-dependent effects. In contrast, hybridisation outcomes are often *highly unpredictable*, dominated by complex epistatic interactions that resist reliable forecasting.

These patterns point to a deeper limitation on predictability that goes beyond noise or incomplete information. They echo a biological analogue of the *Heisenberg uncertainty principle* in physics, which formalises how measurement and system dynamics are intrinsically coupled (**Heisenberg, 1927**): *the more precisely one aspect of change is characterised, the less predictable others become.* In evolutionary systems, *variation and selection* are likewise intertwined, and variation must not be treated as a pre-existing input on which selection subsequently acts, because the generation of variation and the action of selection are inseparable and ongoing. Variability is continuously generated within organisms and populations whose genetic and regulatory architectures have themselves been shaped by prior selection. In other words, variability is a common denominator of both evolutionary change and the selective processes acting upon it. Furthermore, as traits and regulatory systems become increasingly canalised through selection, their responses to perturbation become more robust but less flexible, enhancing

short-term stability at the cost of long-term predictability; contexts that expose hidden or cryptic variation, such as environmental stress or hybridisation, may greatly increase observable variability while simultaneously reducing heritability and evolutionary predictability. In this sense, increasing precision in one dimension of evolutionary description, for example, understanding functional constraint or short-term robustness, often comes at the cost of predictability in another, such as long-term adaptive trajectories.

From this perspective, unpredictability in evolution does not necessarily represent a failure of evolutionary theory or methodology, but a property of the system itself. Natural selection operates as an opportunistic, historically contingent process carried by organisms as organised, dynamic units. In doing so, selection not only filters existing variation but feeds back on the evolutionary "engine" that produces future variation, altering the space of future evolutionary possibilities. Evolution therefore cannot be fully anticipated by extrapolating from current patterns alone, because the mechanisms generating variation are themselves evolving. This is why evolutionary outcomes are often well-defined in retrospect but fundamentally open-ended in prospect.

Looking forward, key challenges lie not in eliminating uncertainty, but in understanding it. This includes identifying when and under which conditions variation becomes evolutionarily effective; and elucidating the balance between predictability and contingency in genetic architecture and regulatory networks across evolutionary timescales. Addressing such questions will be helpful for refining theoretical models and guiding empirical investigations of when and how variation translates into evolutionary potential.

5.3. Conclusion and Take-home Message

This thesis shows that adaptive potential is not a fixed property of genomes or populations, but an emergent outcome of how variation is generated, structured, and filtered across biological scales. Across ecological, molecular, and genomic levels, adaptation reflects the interaction between environmental context, genetic architecture, and evolutionary constraint.

First, ecological traits such as dormancy illustrate how selection acts on life-history strategies that cope with environmental uncertainty. These traits can mediate persistence under variable conditions and also contribute to local adaptation and species range change in the future. Second, analyses of gene expression variation suggest that much of the molecular variation observed in natural populations is not directly available for adaptation; and show how genetic variance components were constrained by past selection and associated with genomic features. Third, the hybridisation study reveals how genetic background reshapes variance. Recombination can expose epistatic interactions and generate novel phenotypes, but whether hybridisation enhances or constrains adaptation depends on the type of epistasis and how genetic variance components are reorganised within the genetic architecture.

Together, these results support the viewpoint that adaptation is a non-linear, scale-dependent process, and that adaptive potential is not determined by the amount of variation present, but by how variation is structured, transmitted, and exposed to selection.

In the end, the central message of this thesis echoes back to Lewontin's insight:
"'*the question was never really, How much genetic variation is there...? but rather, What is the nature of genetic variation for fitness in a population?*'" (**Lewontin, 1974**)

Data Availability

Chapter 2:

- The raw data of germination tests are available on Figshare at https://doi.org/10.6084/m9.figshare.28218308.v1.
- Genomic data of 309 genotypes from the 1001Genome data set are available on the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena), under the project number PRJNA273563 (https://www.ebi.ac.uk/ena/browser/view/PRJNA273563) , and data of 52 accessions were obtained from Wieters et al. (2021). note that the accession ID we see in 1001genome database is called "sample_alias" in ENA, and the SRR number is “run_accession”.

Chapter 3 and Chapter 4:

- All RNAseq and whole-genome data will be shortly published in ENA.
- All phenotypic data and gene count tables are available on Figshare at 10.6084/m9.figshare.29986699
- All parental Plech and Spiterstulen sequence data are available at either NCBI Short Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) or at the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena) with accession codes: SAMN06141173-SAMN06141198 (SRA; (**Mattila et al., 2017**)), SRP144592 (SRA; (**Hämälä et al., 2018**)), PRJEB34247 (ENA; (**Marburger et al., 2019**)), and PRJEB33206 (ENA).

Code Availability

Chapter 2: https://github.com/tntranloc/Athaliana_seed_secondary_dormancy

Chapter 3 and Chapter 4:

https://github.com/tntranloc/Alyrata_RNA_QuantitativeGenetics

Bibliography

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abu Awad, D., & Roze, D. (2018). Effects of partial selfing on the equilibrium genetic variance, mutation load, and inbreeding depression under stabilizing selection: SELFING AND STABILIZING SELECTION. *Evolution*, *72*(4), 751–769. https://doi.org/10.1111/evo.13449

Alexa, A., Rahnenführer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, *22*(13), 1600–1607. https://doi.org/10.1093/bioinformatics/btl140

Alonso-Blanco, C., Aarts, M. G. M., Bentsink, L., Keurentjes, J. J. B., Reymond, M., Vreugdenhil, D., & Koornneef, M. (2009). What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? *The Plant Cell*, *21*(7), 1877–1896. https://doi.org/10.1105/tpc.109.068114

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., Cao, J., Chae, E., Dezwaan, T. M., Ding, W., Ecker, J. R., Exposito-Alonso, M., Farlow, A., Fitz, J., Gan, X., Grimm, D. G., Hancock, A. M., Henz, S. R., Holm, S., … Zhou, X. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*, *166*(2), 481–491. https://doi.org/10.1016/j.cell.2016.05.063

Alonso-Blanco, C., & Koornneef, M. (2000). Naturally occurring variation in Arabidopsis: An underexploited resource for plant genetics. *Trends in Plant Science*, *5*(1), 22–29. https://doi.org/10.1016/s1360-1385(99)01510-1

Anderson, J. T. (2016). Plant fitness in a rapidly changing world. *New Phytologist*, *210*(1), 81–87. https://doi.org/10.1111/nph.13693

Ayroles, J. F., Carbone, M. A., Stone, E. A., Jordan, K. W., Lyman, R. F., Magwire, M. M., Rollmann, S. M., Duncan, L. H., Lawrence, F., Anholt, R. R. H., & Mackay, T. F. C. (2009). Systems genetics of complex traits in Drosophila melanogaster. *Nature Genetics*, *41*(3), 299–307. https://doi.org/10.1038/ng.332

Ballinger, M. A., Mack, K. L., Durkin, S. M., Riddell, E. A., & Nachman, M. W. (2023). Environmentally robust *cis* -regulatory changes underlie rapid climatic adaptation. *Proceedings of the National Academy of Sciences*, *120*(39), e2214614120. https://doi.org/10.1073/pnas.2214614120

Banta, J. A., Ehrenreich, I. M., Gerard, S., Chou, L., Wilczek, A., Schmitt, J., Kover, P. X., & Purugganan, M. D. (2012). Climate envelope modelling reveals intraspecific relationships among flowering phenology, niche breadth and potential range size in ***Arabidopsis thaliana*** . *Ecology Letters*, 15(8), 769–777. https://doi.org/10.1111/j.1461-0248.2012.01796.x

Bao, L., Wei, L., Peirce, J. L., Homayouni, R., Li, H., Zhou, M., Chen, H., Lu, L., Williams, R. W., Pfeffer, L. M., Goldowitz, D., & Cui, Y. (2006). Combining gene expression QTL mapping and phenotypic spectrum analysis to uncover gene regulatory relationships. *Mammalian Genome*, 17(6), 575–583. https://doi.org/10.1007/s00335-005-0172-2

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3(2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., & Barkai, N. (2006). Noise in protein expression scales with natural protein abundance. *Nature Genetics*, 38(6), 636–643. https://doi.org/10.1038/ng1807

Barghetti, A., Sjögren, L., Floris, M., Paredes, E. B., Wenkel, S., & Brodersen, P. (2017). Heat-shock protein 40 is the key farnesylation target in meristem size control, abscisic acid signaling, and drought resistance. *Genes & Development*, 31(22), 2282–2295. https://doi.org/10.1101/gad.301242.117

Barghi, N., Hermisson, J., & Schlötterer, C. (2020). Polygenic adaptation: A unifying framework to understand positive selection. *Nature Reviews Genetics*, 21(12), 769–781. https://doi.org/10.1038/s41576-020-0250-z

Barrett, R., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Barton, N. H. (2001). The role of hybridization in evolution. *Molecular Ecology*, 10(3), 551–568. https://doi.org/10.1046/j.1365-294x.2001.01216.x

Baskin, J., & Baskin, C. (1997). Methods of breaking seed dormancy in the endangered species Iliamna corei (Sherff) Sherff (Malvaceae), with special attention to heating. *Natural Areas Journal*, 17(4), 313–323.

Bateson, W. (1909). Heredity and variation in modern lights. In *Darwin and modern science* (A.C.Seward (Ed.), pp. 85–101). Cambridge University Press.

Bentsink, L., Jowett, J., Hanhart, C. J., & Koornneef, M. (2006). Cloning of *DOG1*, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 103(45), 17042–17047. https://doi.org/10.1073/pnas.0607877103

Blackman, B. K. (2017). Changing Responses to Changing Seasons: Natural Variation in the Plasticity of Flowering Time. *Plant Physiology*, 173(1), 16–26. https://doi.org/10.1104/pp.16.01683

Bomblies, K., & Peichel, C. L. (2022). Genetics of adaptation. *Proceedings of the National Academy of Sciences*, 119(30), e2122152119. https://doi.org/10.1073/pnas.2122152119

Bonnet, T., Morrissey, M. B., de Villemereuil, P., de Villemereuil P, Alberts, S. C., Arcese, P., Bailey, L. D., Boutin, S., Brekke, P., Brent, L. J. N., Glauco Camenisch, Charmantier, A., Clutton-Brock, T. H., Andrew Cockburn, Coltman, D. W., Courtiol, A., Davidian, E., Evans, S., Evans, S. R., … Loeske E. B. Kruuk. (2022). Genetic variance in fitness indicates rapid

contemporary adaptive evolution in wild animals. *Science*, 376(6596), 1012–1016. https://doi.org/10.1126/science.abk0853

Bouwmeester, H. J., & Karssen, C. M. (1993). Annual changes in dormancy and germination in seeds of *Sisymbrium officinale* (L.) Scop. *New Phytologist*, 124(1), 179–191. https://doi.org/10.1111/j.1469-8137.1993.tb03808.x

Boyko, A. R., Williamson, S. H., Indap, A. R., Degenhardt, J. D., Hernandez, R. D., Lohmueller, K. E., Adams, M. D., Schmidt, S., Sninsky, J. J., Sunyaev, S. R., White, T. J., Nielsen, R., Clark, A. G., & Bustamante, C. D. (2008). Assessing the Evolutionary Impact of Amino Acid Mutations in the Human Genome. *PLoS Genetics*, 4(5), e1000083. https://doi.org/10.1371/journal.pgen.1000083

Bradski, G. (2000). *The OpenCV Library*. [Computer software]. Dr Dobbs J Software Tools.

Brem, R. B., Yvert, G., Clinton, R., & Kruglyak, L. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. *Science*, 296(5568), 752–755. https://doi.org/10.1126/science.1069516

Brookes, A. J. (1999). The essence of SNPs. *Gene*, 234(2), 177–186. https://doi.org/10.1016/S0378-1119(99)00219-X

Brucker, R. M., & Bordenstein, S. R. (2013). The Hologenomic Basis of Speciation: Gut Bacteria Cause Hybrid Lethality in the Genus *Nasonia*. *Science*, 341(6146), 667–669. https://doi.org/10.1126/science.1240659

Bruijning, M., Metcalf, C. J. E., Jongejans, E., & Ayroles, J. F. (2020). The Evolution of Variance Control. *Trends in Ecology & Evolution*, 35(1), 22–33. https://doi.org/10.1016/j.tree.2019.08.005

Buijs, G. (2020). A Perspective on Secondary Seed Dormancy in Arabidopsis thaliana. *Plants*, 9(6), 749. https://doi.org/10.3390/plants9060749

Bulmer, M. G. (1971). The Effect of Selection on Genetic Variability. *The American Naturalist*, 105(943), 201–211. https://doi.org/10.1086/282718

Burghardt, L. T., Metcalf, C. J. E., Wilczek, A. M., Schmitt, J., & Donohue, K. (2015). Modeling the Influence of Genetic and Environmental Variation on the Expression of Plant Life Cycles across Landscapes. *The American Naturalist*, 185(2), 212–227. https://doi.org/10.1086/679439

Bürkner, P.-C. (2017). **brms**: An R Package for Bayesian Multilevel Models Using *Stan*. *Journal of Statistical Software*, 80(1). https://doi.org/10.18637/jss.v080.i01

Caballero, A. (2020). *Quantitative Genetics* (1st edn). Cambridge University Press. https://doi.org/10.1017/9781108630542

Cai, H., & Des Marais, D. L. (2023). Revisiting regulatory coherence: Accounting for temporal bias in plant gene co-expression analyses. *The New Phytologist*, 238(1), 16–24. https://doi.org/10.1111/nph.18720

Carroll, S. P., Dingle, H., & Famula, T. R. (2003). Rapid appearance of epistasis during adaptive divergence following colonization. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(suppl_1). https://doi.org/10.1098/rsbl.2003.0019

Carson, H. L., & Wisotzkey, R. G. (1989). Increase in Genetic Variance Following a Population Bottleneck. *The American Naturalist*, 134(4), 668–673. https://doi.org/10.1086/285004

Carter, A. J. R., Hermisson, J., & Hansen, T. F. (2005). The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theoretical Population Biology*, 68(3), 179–196. https://doi.org/10.1016/j.tpb.2005.05.002

Castellano, D., Macià, M. C., Tataru, P., Bataillon, T., & Munch, K. (2019). Comparison of the Full Distribution of Fitness Effects of New Amino Acid Mutations Across Great Apes. *Genetics*, 213(3), 953–966. https://doi.org/10.1534/genetics.119.302494

Chahtane, H., Kim, W., & Lopez-Molina, L. (2016). Primary seed dormancy: A temporally multilayered riddle waiting to be unlocked. *Journal of Experimental Botany*, erw377. https://doi.org/10.1093/jxb/erw377

Chang, C., Bowman, J. L., & Meyerowitz, E. M. (2016). Field Guide to Plant Model Systems. *Cell*, 167(2), 325–339. https://doi.org/10.1016/j.cell.2016.08.031

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134(4), 1289–1303. https://doi.org/10.1093/genetics/134.4.1289

Charlesworth, D., Barton, N. H., & Charlesworth, B. (2017). The sources of adaptive variation. *Proceedings of the Royal Society B: Biological Sciences*, 284(1855), 20162864. https://doi.org/10.1098/rspb.2016.2864

Chen, J., Bataillon, T., Glémin, S., & Lascoux, M. (2022). What does the distribution of fitness effects of new mutations reflect? Insights from plants. *New Phytologist*, 233(4), 1613–1619. https://doi.org/10.1111/nph.17826

Chen, J., Glémin, S., & Lascoux, M. (2017). Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. *Molecular Biology and Evolution*, 34(6), 1417–1428. https://doi.org/10.1093/molbev/msx088

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cheverud, J. M., & Routman, E. J. (1995). Epistasis and its contribution to genetic variance components. *Genetics*, 139(3), 1455–1461. https://doi.org/10.1093/genetics/139.3.1455

Chiang, G. C. K., Barua, D., Dittmar, E., Kramer, E. M., De Casas, R. R., & Donohue, K. (2013). PLEIOTROPY IN THE WILD: THE DORMANCY GENE *DOG1* EXERTS CASCADING CONTROL ON LIFE CYCLES. *Evolution*, 67(3), 883–893. https://doi.org/10.1111/j.1558-5646.2012.01828.x

Chung, M. Y., Merilä, J., Li, J., Mao, K., López-Pujol, J., Tsumura, Y., & Chung, M. G. (2023). Neutral and adaptive genetic diversity in plants: An overview. *Frontiers in Ecology and Evolution*, 11, 1116814. https://doi.org/10.3389/fevo.2023.1116814

Ciliberti, S., Martin, O. C., & Wagner, A. (2007). Innovation and robustness in complex regulatory gene networks. *Proceedings of the National Academy of Sciences*, 104(34), 13591–13596. https://doi.org/10.1073/pnas.0705396104

Class, B., & Brommer, J. E. (2020). Can dominance genetic variance be ignored in evolutionary quantitative genetic analyses of wild populations? *Evolution*, 74(7), 1540–1550. https://doi.org/10.1111/evo.14034

Clauss, M. J., & Mitchell-Olds, T. (2006). Population genetic structure of *Arabidopsis lyrata* in Europe. *Molecular Ecology*, 15(10), 2753–2766. https://doi.org/10.1111/j.1365-294X.2006.02973.x

Clo, J., & Opedal, Ø. H. (2021). Genetics of quantitative traits with dominance under stabilizing and directional selection in partially selfing species. *Evolution*, 75(8), 1920–1935. https://doi.org/10.1111/evo.14304

Clo, J., Ronfort, J., & Abu Awad, D. (2020). Hidden genetic variance contributes to increase the short-term adaptive potential of selfing populations. *Journal of Evolutionary Biology*, 33(9), 1203–1215. https://doi.org/10.1111/jeb.13660

Coughlan, J. M., Saha, A., & Donohue, K. (2017). Effects of pre- and post-dispersal temperature on primary and secondary dormancy dynamics in contrasting genotypes of A *rabidopsis thaliana* (B rassicaceae). *Plant Species Biology*, 32(3), 210–222. https://doi.org/10.1111/1442-1984.12145

Crnokrak, P., & Roff, D. A. (1995). Dominance variance: Associations with selection and fitness. *Heredity (Edinb.)*, 75(5), 530–540.

Crombach, A., & Hogeweg, P. (2008). Evolution of Evolvability in Gene Regulatory Networks. *PLoS Computational Biology*, 4(7), e1000112. https://doi.org/10.1371/journal.pcbi.1000112

Cuena Lombraña, A., Dessì, L., Podda, L., Fois, M., Luna, B., Porceddu, M., & Bacchetta, G. (2024). The Effect of Heat Shock on Seed Dormancy Release and Germination in Two Rare and Endangered Astragalus L. Species (Fabaceae). *Plants*, 13(4), 484. https://doi.org/10.3390/plants13040484

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. https://doi.org/10.1093/gigascience/giab008

Darwin, C., Murray, J., William Clowes and Sons, & Bradbury & Evans. (1859). *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life*. John Murray, Albemarle Street. https://doi.org/10.5962/bhl.title.82303

Davidson, E. H., & Erwin, D. H. (2010). Evolutionary innovation and stability in animal gene networks. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 314B(3), 182–186. https://doi.org/10.1002/jez.b.21329

de la Chaux, N., Tsuchimatsu, T., Shimizu, K. K., & Wagner, A. (2012). The predominantly selfing plant Arabidopsis thaliana experienced a recent reduction in transposable element abundance compared to its outcrossing relative Arabidopsis lyrata. *Mobile DNA*, 3(1), 2. https://doi.org/10.1186/1759-8753-3-2

Debieu, M., Tang, C., Stich, B., Sikosek, T., Effgen, S., Josephs, E., Schmitt, J., Nordborg, M., Koornneef, M., & De Meaux, J. (2013). Co-Variation between Seed Dormancy, Growth Rate and Flowering Time Changes with Latitude in Arabidopsis thaliana. *PLoS ONE*, 8(5), e61075. https://doi.org/10.1371/journal.pone.0061075

Delaneau, O., Ongen, H., Brown, A. A., Fort, A., Panousis, N. I., & Dermitzakis, E. T. (2017). A complete tool set for molecular QTL discovery and analysis. *Nature Communications*, *8*(1), 15452. https://doi.org/10.1038/ncomms15452

Deshpande, J. N., & Fronhofer, E. A. (2022). Genetic architecture of dispersal and local adaptation drives accelerating range expansions. *Proceedings of the National Academy of Sciences*, *119*(31), e2121858119. https://doi.org/10.1073/pnas.2121858119

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dobzhansky, T. (1937). *Genetics and the Origin of Species*. Columbia University Press.

Donohue, K., Rubio De Casas, R., Burghardt, L., Kovach, K., & Willis, C. G. (2010). Germination, Postgermination Adaptation, and Species Ecological Ranges. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 293–319. https://doi.org/10.1146/annurev-ecolsys-102209-144715

Drury, D. W., & Wade, M. J. (2011). Genetic variation and co-variation for fitness between intra-population and inter-population backgrounds in the red flour beetle, Tribolium castaneum: Genetic variation for hybrid fitness. *Journal of Evolutionary Biology*, *24*(1), 168–176. https://doi.org/10.1111/j.1420-9101.2010.02151.x

Duarte, D. M., & Garcia, Q. S. (2015). Interactions between substrate temperature and humidity in signalling cyclical dormancy in seeds of two perennial tropical species. *Seed Science Research*, *25*(2), 170–178. https://doi.org/10.1017/S0960258515000045

Duenk, P., Bijma, P., Calus, M. P. L., Wientjes, Y. C. J., & Van Der Werf, J. H. J. (2020). The Impact of Non-additive Effects on the Genetic Correlation Between Populations. *G3 Genes|Genomes|Genetics*, *10*(2), 783–795. https://doi.org/10.1534/g3.119.400663

Dwivedi, S. L., Reynolds, M. P., & Ortiz, R. (2021). Mitigating tradeoffs in plant breeding. *iScience*, *24*(9), 102965. https://doi.org/10.1016/j.isci.2021.102965

Erwin, D. H., & Davidson, E. H. (2009). The evolution of hierarchical gene regulatory networks. *Nature Reviews Genetics*, 10(2), 141–148. https://doi.org/10.1038/nrg2499

Escobar, J. S., Nicot, A., & David, P. (2008). The different sources of variation in inbreeding depression, heterosis and outbreeding depression in a metapopulation of Physa acuta. *Genetics*, 180(3), 1593–1608. https://doi.org/10.1534/genetics.108.092718

Esfandyari, H., Henryon, M., Berg, P., Thomasen, J. R., Bijma, P., & Sørensen, A. C. (2017). Response to selection in finite locus models with non-additive effects. *Journal of Heredity*, esw123. https://doi.org/10.1093/jhered/esw123

Etterson, J. R., & Shaw, R. G. (2001). Constraint to Adaptive Evolution in Response to Global Warming. *Science*, 294(5540), 151–154. https://doi.org/10.1126/science.1063656

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Exposito-Alonso, M., 500 Genomes Field Experiment Team, Burbano, H. A., Bossdorf, O., Nielsen, R., & Weigel, D. (2019). Natural selection on the Arabidopsis thaliana genome in present and future climates. *Nature*, 573(7772), 126–129. https://doi.org/10.1038/s41586-019-1520-9

Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8), 610–618. https://doi.org/10.1038/nrg2146

Falconer, D. S., & Mackay, T. (2009). *Introduction to quantitative genetics* (4. ed., [16. print.]). Pearson, Prentice Hall.

Fay, J. C., & Wittkopp, P. J. (2008). Evaluating the role of natural selection in the evolution of gene regulation. *Heredity*, 100(2), 191–199. https://doi.org/10.1038/sj.hdy.6801000

Félix, M.-A., & Barkoulas, M. (2015). Pervasive robustness in biological systems. *Nature Reviews Genetics*, 16(8), 483–496. https://doi.org/10.1038/nrg3949

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fisher, R. A. (1919). XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.*, 52(2), 399–433.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press. https://doi.org/10.5962/bhl.title.27468

Fitzpatrick, B. M. (2012). Underappreciated Consequences of Phenotypic Plasticity for Ecological Speciation. *International Journal of Ecology*, 2012, 1–12. https://doi.org/10.1155/2012/256017

Footitt, S., Douterelo-Soler, I., Clay, H., & Finch-Savage, W. E. (2011). Dormancy cycling in *Arabidopsis* seeds is controlled by seasonally distinct hormone-signaling pathways. *Proceedings of the National Academy of Sciences*, 108(50), 20236–20241. https://doi.org/10.1073/pnas.1116325108

Footitt, S., Müller, K., Kermode, A. R., & Finch-Savage, W. E. (2015). Seed dormancy cycling in A rabidopsis: Chromatin remodelling and regulation of DOG 1 in response to seasonal environmental signals. *The Plant Journal*, 81(3), 413–425. https://doi.org/10.1111/tpj.12735

Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J., & Wilczek, A. M. (2011). A Map of Local Adaptation in *Arabidopsis thaliana*. *Science*, 334(6052), 86–89. https://doi.org/10.1126/science.1209271

Frachon, L., Libourel, C., Villoutreix, R., Carrère, S., Glorieux, C., Huard-Chauveau, C., Navascués, M., Gay, L., Vitalis, R., Baron, E., Amsellem, L., Bouchez, O., Vidal, M., Le Corre, V., Roby, D., Bergelson, J., & Roux, F. (2017). Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. *Nature Ecology & Evolution*, 1(10), 1551–1561. https://doi.org/10.1038/s41559-017-0297-1

Fu, Y. X. (1995). Statistical Properties of Segregating Sites. *Theoretical Population Biology*, 48(2), 172–197. https://doi.org/10.1006/tpbi.1995.1025

Fusco, G., & Minelli, A. (2010). Phenotypic plasticity in development and evolution: Facts and concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1540), 547–556. https://doi.org/10.1098/rstb.2009.0267

Gallais, A. (2003). *Quantitative genetics and breeding methods in autopolyploid plants*. Institut national de la recherche agronomique.

Galtier, N. (2016). Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis. *PLOS Genetics*, 12(1), e1005774. https://doi.org/10.1371/journal.pgen.1005774

Garfield, D. A., Runcie, D. E., Babbitt, C. C., Haygood, R., Nielsen, W. J., & Wray, G. A. (2013). The impact of gene expression variation on the robustness and evolvability of a developmental gene regulatory network. *PLoS Biology*, *11*(10), e1001696. https://doi.org/10.1371/journal.pbio.1001696

Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., Doust, A. N., Feldman, M. J., Gilbert, K. B., Hodge, J. G., Hoyer, J. S., Lin, A., Liu, S., Lizárraga, C., Lorence, A., Miller, M., Platon, E., Tessman, M., & Sax, T. (2017). PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ*, 5, e4088. https://doi.org/10.7717/peerj.4088

Ghazalpour, A., Bennett, B., Petyuk, V. A., Orozco, L., Hagopian, R., Mungrue, I. N., Farber, C. R., Sinsheimer, J., Kang, H. M., Furlotte, N., Park, C. C., Wen, P.-Z., Brewer, H., Weitz, K., Camp, D. G., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., … Lusis, A. J. (2011). Comparative Analysis of Proteome and Transcriptome Variation in Mouse. *PLoS Genetics*, 7(6), e1001393. https://doi.org/10.1371/journal.pgen.1001393

Gianinetti, A. (2023). A Travel through Landscapes of Seed Dormancy. *Plants*, 12(23), 3963. https://doi.org/10.3390/plants12233963

Gibson, G., & Dworkin, I. (2004). Uncovering cryptic genetic variation. *Nature Reviews Genetics*, 5(9), 681–690. https://doi.org/10.1038/nrg1426

Gilad, Y., Rifkin, S. A., & Pritchard, J. K. (2008). Revealing the architecture of gene regulation: The promise of eQTL studies. *Trends in Genetics*, *24*(8), 408–415. https://doi.org/10.1016/j.tig.2008.06.001

Goodnight, C. J. (1995). EPISTASIS AND THE INCREASE IN ADDITIVE GENETIC VARIANCE: IMPLICATIONS FOR PHASE 1 OF WRIGHT'S SHIFTING-BALANCE PROCESS. *Evolution*, *49*(3), 502–511. https://doi.org/10.1111/j.1558-5646.1995.tb02282.x

Grant, P. R., & Grant, B. R. (2019). Hybridization increases population variation during adaptive radiation. *Proceedings of the National Academy of Sciences*, *116*(46), 23216–23224. https://doi.org/10.1073/pnas.1913534116

Greenbaum, D., Colangelo, C., Williams, K., & Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*, *4*(9), 117. https://doi.org/10.1186/gb-2003-4-9-117

Gregor, J. W. (1938). EXPERIMENTAL TAXONOMY: II. INITIAL POPULATION DIFFERENTIATION IN PLANTAGO MARITIMA L. OF BRITAIN. *New Phytologist*, *37*(1), 15–49. https://doi.org/10.1111/j.1469-8137.1938.tb06925.x

Groen, S. C., Ćalić, I., Joly-Lopez, Z., Platts, A. E., Choi, J. Y., Natividad, M., Dorph, K., Mauck, W. M., Bracken, B., Cabral, C. L. U., Kumar, A., Torres, R. O., Satija, R., Vergara, G., Henry, A., Franks, S. J., & Purugganan, M. D. (2020). The strength and pattern of natural selection on gene expression in rice. *Nature*, *578*(7796), 572–576. https://doi.org/10.1038/s41586-020-1997-2

Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2008). Heterozygosity–fitness correlations and their relevance to studies on inbreeding depression in threatened species. *Molecular Ecology*, *17*(18), 3978–3984. https://doi.org/10.1111/j.1365-294X.2008.03910.x

Gulden, R. H., Thomas, A. G., & Shirtliffe, S. J. (2004). Secondary dormancy, temperature, and burial depth regulate seedbank dynamics in canola. *Weed Science*, *52*(3), 382–388. https://doi.org/10.1614/ws-03-123r1

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genetics*, 5(10), e1000695. https://doi.org/10.1371/journal.pgen.1000695

Hahn, M. W., & Kern, A. D. (2005). Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks. *Molecular Biology and Evolution*, 22(4), 803–806. https://doi.org/10.1093/molbev/msi072

Hajheidari, M., Sunyaev, S., & De Meaux, J. (2025). Are complex traits underpinned by polygenic molecular traits? A reflection on the complexity of gene expression. *Plant And Cell Physiology*, 66(4), 444–460. https://doi.org/10.1093/pcp/pcae140

Haldane, J. B. S. (1927). A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23(7), 838–844. https://doi.org/10.1017/S0305004100015644

Haldane, J. B. S. (1932). *The Causes of Evolution* (1st edition). Longmans, Green and Co.

Hallander, J., & Waldmann, P. (2007). The effect of non-additive genetic interactions on selection in multi-locus genetic models. *Heredity*, 98(6), 349–359. https://doi.org/10.1038/sj.hdy.6800946

Hämälä, T., Guiltinan, M. J., Marden, J. H., Maximova, S. N., dePamphilis, C. W., & Tiffin, P. (2020). Gene Expression Modularity Reveals Footprints of Polygenic Adaptation in Theobroma cacao. *Molecular Biology and Evolution*, 37(1), 110–123. https://doi.org/10.1093/molbev/msz206

Hämälä, T., Mattila, T. M., & Savolainen, O. (2018). Local adaptation and ecological differentiation under selection, migration, and drift in *Arabidopsis lyrata* *: LOCAL ADAPTATION UNDER GENE FLOW AND DRIFT. *Evolution*, 72(7), 1373–1386. https://doi.org/10.1111/evo.13502

Hämälä, T., Weixuan Ning, Helmi Kuittinen, Nader Aryamanesh, & Outi Savolainen. (2022). Environmental response in gene expression and DNA methylation reveals factors influencing the adaptive potential of Arabidopsis lyrata. *eLife*, *11*. https://doi.org/10.7554/elife.83115

Hamann, E., Denney, D., Day, S., Lombardi, E., Jameel, M. I., MacTavish, R., & Anderson, J. T. (2021). Review: Plant eco-evolutionary responses to climate change: Emerging directions. *Plant Science*, 304, 110737. https://doi.org/10.1016/j.plantsci.2020.110737

Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., Toomajian, C., Roux, F., & Bergelson, J. (2011). Adaptation to Climate Across the *Arabidopsis thaliana* Genome. *Science*, 334(6052), 83–86. https://doi.org/10.1126/science.1209244

Hancock, A. M., Portalier, S., Fulgione, A., Stetter, M. G., & De Meaux, J. (2025). The Molecular Basis of Adaptation to Climatic Factors and Range Change in Plants. *Annual Review of Ecology, Evolution, and Systematics*, 56(1), 597–621. https://doi.org/10.1146/annurev-ecolsys-102723-053810

Hansen, T. F. (2015). Measuring Gene Interactions. In J. H. Moore & S. M. Williams (Eds), *Epistasis* (Vol. 1253, pp. 115–143). Springer New York. https://doi.org/10.1007/978-1-4939-2155-3_7

Hansen, T. F., Houle, D., Pavličev, M., & Pélabon, C. (Eds). (2023). *Evolvability: A unifying concept in evolutionary biology*? The MIT Press.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Harvardinformatics. (2024). *Degenotate* [Computer software]. https://github.com/harvardinformatics

Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3–4), 172–198. https://doi.org/10.1007/BF01397280

Hemani, G., Knott, S., & Haley, C. (2013). An Evolutionary Perspective on Epistasis and the Missing Heritability. *PLoS Genetics*, 9(2), e1003295. https://doi.org/10.1371/journal.pgen.1003295

Henderson, C. R. (1953). Estimation of Variance and Covariance Components. *Biometrics*, 9(2), 226. https://doi.org/10.2307/3001853

Henderson, C. R. (1976). A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics*, 32(1), 69. https://doi.org/10.2307/2529339

Hendry, A. P. (2016). Key Questions on the Role of Phenotypic Plasticity in Eco-Evolutionary Dynamics. *Journal of Heredity*, 107(1), 25–41. https://doi.org/10.1093/jhered/esv060

Henry, G. A., & Stinchcombe, J. R. (2025). Predicting Fitness-Related Traits Using Gene Expression and Machine Learning. *Genome Biology and Evolution*, 17(2), evae275. https://doi.org/10.1093/gbe/evae275

Hermisson, J., & Pennings, P. S. (2005). Soft Sweeps. *Genetics*, 169(4), 2335–2352. https://doi.org/10.1534/genetics.104.036947

Hijmans, R., Ghosh, A., & Mandel, A. (2022). *geodata: Download Geographic Data _ R package version* 0.5-3. https://CRAN.R-project.org/package=geodata

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genetics*, 4(2), e1000008. https://doi.org/10.1371/journal.pgen.1000008

Hodgins-Davis, A., & Townsend, J. P. (2009). Evolving gene expression: From G to E to G×E. *Trends in Ecology & Evolution*, 24(12), 649–658. https://doi.org/10.1016/j.tree.2009.06.011

Höllinger, I., Pennings, P. S., & Hermisson, J. (2019). Polygenic adaptation: From sweeps to subtle frequency shifts. *PLOS Genetics*, *15*(3), e1008035. https://doi.org/10.1371/journal.pgen.1008035

Hou, J., Liu, M., Yang, K., Liu, B., Liu, H., & Liu, J. (2025). Genetic variation for adaptive evolution in response to changed environments in plants. *Journal of Integrative Plant Biology*, jipb.13961. https://doi.org/10.1111/jipb.13961

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., Fahlgren, N., Fawcett, J. A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J. D., Ossowski, S., Ottilar, R. P., Salamov, A. A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., … Guo, Y.-L. (2011). The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nature Genetics*, *43*(5), 476–481. https://doi.org/10.1038/ng.807

Huang, W., & Mackay, T. F. C. (2016). The Genetic Architecture of Quantitative Traits Cannot Be Inferred from Variance Component Analysis. *PLOS Genetics*, *12*(11), e1006421. https://doi.org/10.1371/journal.pgen.1006421

Huber, C. D., Durvasula, A., Hancock, A. M., & Lohmueller, K. E. (2018). Gene expression drives the evolution of dominance. *Nature Communications*, *9*(1), 2750. https://doi.org/10.1038/s41467-018-05281-7

Hvala, J. A., Frayer, M. E., & Payseur, B. A. (2018). Signatures of hybridization and speciation in genomic patterns of ancestry*. *Evolution*, *72*(8), 1540–1552. https://doi.org/10.1111/evo.13509

Ikeda, D. H., Max, T. L., Allan, G. J., Lau, M. K., Shuster, S. M., & Whitham, T. G. (2017). Genetically informed ecological niche models improve climate change predictions. *Global Change Biology*, *23*(1), 164–176. https://doi.org/10.1111/gcb.13470

Iwasaki, M., Penfield, S., & Lopez-Molina, L. (2022). Parental and Environmental Control of Seed Dormancy in *Arabidopsis thaliana*. *Annual Review of Plant Biology*, *73*(1), 355–378. https://doi.org/10.1146/annurev-arplant-102820-090750

Jarvis, J. P., Cropp, S. N., Vaughn, T. T., Pletscher, L. S., King-Ellison, K., Adams-Hunt, E., Erickson, C., & Cheverud, J. M. (2011). The effect of a population bottleneck on the evolution of genetic variance/covariance structure: Evolution of genetic V/CV structure. *Journal of Evolutionary Biology*, *24*(10), 2139–2152. https://doi.org/10.1111/j.1420-9101.2011.02347.x

Jonsell, B., Kustås, K., & Nordal, I. (1995). Genetic variation in Arabis petraea, a disjunct species in northern Europe. *Ecography*, *18*(4), 321–332. https://doi.org/10.1111/j.1600-0587.1995.tb00135.x

Jordan, I. K., Rogozin, I. B., Wolf, Y. I., & Koonin, E. V. (2002). Essential Genes Are More Evolutionarily Conserved Than Are Nonessential Genes in Bacteria. *Genome Research*, *12*(6), 962–968. https://doi.org/10.1101/gr.87702

Joschinski, J., & Bonte, D. (2020). Transgenerational Plasticity and Bet-Hedging: A Framework for Reaction Norm Evolution. *Frontiers in Ecology and Evolution*, *8*, 517183. https://doi.org/10.3389/fevo.2020.517183

Kærn, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, *6*(6), 451–464. https://doi.org/10.1038/nrg1615

Kärkkäinen, K., Kuittinen, H., Van Treuren, R., Vogl, C., Oikarinen, S., & Savolainen, O. (1999). GENETIC BASIS OF INBREEDING DEPRESSION IN *ARABIS PETRAEA*. *Evolution*, *53*(5), 1354–1365. https://doi.org/10.1111/j.1558-5646.1999.tb05400.x

Katoh, K. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kaye, T. N., Sandlin, I. J., & Bahm, M. A. (2018). Seed dormancy and germination vary within and among species of milkweeds. *AoB PLANTS*, *10*(2). https://doi.org/10.1093/aobpla/ply018

Kerdaffrec, E., & Nordborg, M. (2017). The maternal environment interacts with genetic variation in regulating seed dormancy in Swedish Arabidopsis thaliana. *PLOS ONE*, *12*(12), e0190242. https://doi.org/10.1371/journal.pone.0190242

Kim, J., & Gibson, G. (2010). Insights from GWAS into the quantitative genetics of transcription in humans. *Genetics Research*, *92*(5–6), 361–369. https://doi.org/10.1017/S001667231000056X

Kimura, M. (1968). Evolutionary Rate at the Molecular Level. *Nature*, *217*(5129), 624–626. https://doi.org/10.1038/217624a0

Kittelmann, S., Buffry, A. D., Franke, F. A., Almudi, I., Yoth, M., Sabaris, G., Couso, J. P., Nunes, M. D. S., Frankel, N., Gómez-Skarmeta, J. L., Pueyo-Marques, J., Arif, S., & McGregor, A. P. (2018). Gene regulatory network architecture in different developmental contexts influences the genetic basis of morphological evolution. *PLoS Genetics*, *14*(5), e1007375. https://doi.org/10.1371/journal.pgen.1007375

Klupczyńska, E. A., & Pawłowski, T. A. (2021). Regulation of Seed Dormancy and Germination Mechanisms in a Changing Environment. *International Journal of Molecular Sciences*, *22*(3), 1357. https://doi.org/10.3390/ijms22031357

Koch, M. A., Haubold, B., & Mitchell-Olds, T. (2000). Comparative Evolutionary Analysis of Chalcone Synthase and Alcohol Dehydrogenase Loci in Arabidopsis, Arabis, and Related Genera (Brassicaceae). *Molecular Biology and Evolution*, *17*(10), 1483–1498. https://doi.org/10.1093/oxfordjournals.molbev.a026248

Kolesnikova, U. K., Scott, A. D., Van De Velde, J. D., Burns, R., Tikhomirov, N. P., Pfordt, U., Clarke, A. C., Yant, L., Seregin, A. P., Vekemans, X., Laurent, S., & Novikova, P. Y. (2023). Transition to Self-compatibility Associated With Dominant S -allele in a Diploid Siberian Progenitor of Allotetraploid *Arabidopsis kamchatica* Revealed by *Arabidopsis lyrata* Genomes. *Molecular Biology and Evolution*, *40*(7), msad122. https://doi.org/10.1093/molbev/msad122

Korunes, K. L., & Samuk, K. (2021). PIXY: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources*, *21*(4), 1359–1368. https://doi.org/10.1111/1755-0998.13326

Krämer, U. (2015). Planting molecular functions in an ecological context with Arabidopsis thaliana. *eLife*, 4, e06100. https://doi.org/10.7554/eLife.06100

Kremling, K. A. G., Chen, S.-Y., Su, M.-H., Lepak, N. K., Romay, M. C., Swarts, K. L., Lu, F., Lorant, A., Bradbury, P. J., & Buckler, E. S. (2018). Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature*, 555(7697), 520–523. https://doi.org/10.1038/nature25966

Kronholm, I., Picó, F. X., Alonso-Blanco, C., Goudet, J., & Meaux, J. D. (2012). GENETIC BASIS OF ADAPTATION IN ARABIDOPSIS THALIANA: LOCAL ADAPTATION AT THE SEED DORMANCY QTL DOG1: LOCAL ADAPTATION FOR SEED DORMANCY QTL DOG1. *Evolution*, 66(7), 2287–2302. https://doi.org/10.1111/j.1558-5646.2012.01590.x

Kumar, S., Molloy, C., Muñoz, P., Daetwyler, H., Chagné, D., & Volz, R. (2015). Genome-Enabled Estimates of Additive and Nonadditive Genetic Variances and Prediction of Apple Phenotypes Across Environments. *G3 Genes|Genomes|Genetics*, 5(12), 2711–2718. https://doi.org/10.1534/g3.115.021105

Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J. B., & Nasrallah, M. E. (2001). Self-incompatibility in the genus Arabidopsis: Characterization of the S locus in the outcrossing A. lyrata and its autogamous relative A. thaliana. *The Plant Cell*, 13(3), 627–643.

Lai, W.-Y., Nolte, V., Jakšić, A. M., & Schlötterer, C. (2024). Evolution of Phenotypic Variance Provides Insights into the Genetic Basis of Adaptation. *Genome Biology and Evolution*, 16(4), evae077. https://doi.org/10.1093/gbe/evae077

Lai, Y.-T., Yeung, C. K. L., Omland, K. E., Pang, E.-L., Hao, Y., Liao, B.-Y., Cao, H.-F., Zhang, B.-W., Yeh, C.-F., Hung, C.-M., Hung, H.-Y., Yang, M.-Y., Liang, W., Hsu, Y.-C., Yao, C.-T., Dong, L., Lin, K., & Li, S.-H. (2019). Standing genetic variation as the predominant source for

adaptation of a songbird. *Proceedings of the National Academy of Sciences*, 116(6), 2152–2157. https://doi.org/10.1073/pnas.1813597116

Lamont, B. B., & Pausas, J. G. (2023). Seed dormancy revisited: Dormancy-release pathways and environmental interactions. *Functional Ecology*, 37(4), 1106–1125. https://doi.org/10.1111/1365-2435.14269

Lande, R., & Porcher, E. (2015). Maintenance of Quantitative Genetic Variance Under Partial Self-Fertilization, with Implications for Evolution of Selfing. *Genetics*, 200(3), 891–906. https://doi.org/10.1534/genetics.115.176693

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. https://doi.org/10.1038/nmeth.1923

Larièpe, A., Moreau, L., Laborde, J., Bauland, C., Mezmouk, S., Décousset, L., Mary-Huard, T., Fiévet, J. B., Gallais, A., Dubreuil, P., & Charcosset, A. (2017). General and specific combining abilities in a maize (Zea mays L.) test-cross hybrid panel: Relative importance of population structure and genetic divergence between parents. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 130(2), 403–417. https://doi.org/10.1007/s00122-016-2822-z

Le Rouzic, A., Roumet, M., Widmer, A., & Clo, J. (2024). Detecting directional epistasis and dominance from cross-line analyses in alpine populations of *Arabidopsis thaliana*. *Journal of Evolutionary Biology*, 37(7), 839–847. https://doi.org/10.1093/jeb/voae056

Leder, E. H., McCairns, R. J. S., Leinonen, T., Cano, J. M., Viitaniemi, H. M., Nikinmaa, M., Primmer, C. R., & Merilä, J. (2015). The evolution and adaptive potential of transcriptional variation in sticklebacks – signatures of selection and widespread heritability. *Molecular Biology and Evolution*, 32(3), 674–689. https://doi.org/10.1093/molbev/msu328

Lewontin, R. C. (1974). *The genetic basis of evolutionary change*. Columbia Univ. Pr.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with* BWA-MEM (Version 2). arXiv. https://doi.org/10.48550/ARXIV.1303.3997

Li, Y., Álvarez, O. A., Gutteling, E. W., Tijsterman, M., Fu, J., Riksen, J. A. G., Hazendonk, E., Prins, P., Plasterk, R. H. A., Jansen, R. C., Breitling, R., & Kammenga, J. E. (2006). Mapping Determinants of Gene Expression Plasticity by Genetical Genomics in C. elegans. *PLoS Genetics*, 2(12), e222. https://doi.org/10.1371/journal.pgen.0020222

Li, Y., Mamonova, E., Köhler, N., Van Kleunen, M., & Stift, M. (2023). Breakdown of self-incompatibility due to genetic interaction between a specific S-allele and an unlinked modifier. *Nature Communications*, 14(1), 3420. https://doi.org/10.1038/s41467-023-38802-0

Lin, X., Yin, J., Wang, Y., Yao, J., Li, Q. Q., Latzel, V., Bossdorf, O., & Zhang, Y.-Y. (2024). Environment-induced heritable variations are common in Arabidopsis thaliana. *Nature Communications*, 15(1), 4615. https://doi.org/10.1038/s41467-024-49024-3

Lippman, Z. B., & Zamir, D. (2007). Heterosis: Revisiting the magic. *Trends in Genetics*, 23(2), 60–66. https://doi.org/10.1016/j.tig.2006.12.006

Liu, L., Wang, Y., Zhang, D., Chen, Z., Chen, X., Su, Z., & He, X. (2020). The Origin of Additive Genetic Variance Driven by Positive Selection. *Molecular Biology and Evolution*, 37(8), 2300–2308. https://doi.org/10.1093/molbev/msaa085

Lloyd, J. P., Seddon, A. E., Moghe, G. D., Simenc, M. C., & Shiu, S.-H. (2015). Characteristics of Plant Essential Genes Allow for within- and between-Species Prediction of Lethal Mutant Phenotypes. *The Plant Cell*, 27(8), 2133–2147. https://doi.org/10.1105/tpc.15.00051

Lockhart, D. J., & Winzeler, E. A. (2000). Genomics, gene expression and DNA arrays. *Nature*, 405(6788), 827–836. https://doi.org/10.1038/35015701

López-Fanjul, C., & Villaverde, A. (1989). INBREEDING INCREASES GENETIC VARIANCE FOR VIABILITY IN *DROSOPHILA MELANOGASTER*. *Evolution*, *43*(8), 1800–1804. https://doi.org/10.1111/j.1558-5646.1989.tb02628.x

Lotsy, J. P. (1916). *Evolution by means of hybridization*. M. Nijhoff. https://doi.org/10.5962/bhl.title.55656

Luijten, S. H., Kéry, M., Oostermeijer, J. G. B., & Den Nijs, H. (J.) C. M. (2002). Demographic consequences of inbreeding and outbreeding in *Arnica montana*: A field experiment. *Journal of Ecology*, *90*(4), 593–603. https://doi.org/10.1046/j.1365-2745.2002.00703.x

Mack, K. L., & Nachman, M. W. (2017). Gene Regulation and Speciation. *Trends in Genetics*, *33*(1), 68–80. https://doi.org/10.1016/j.tig.2016.11.003

Mackay, T. F. C. (2014). Epistasis and quantitative traits: Using model organisms to study gene–gene interactions. *Nature Reviews Genetics*, *15*(1), 22–33. https://doi.org/10.1038/nrg3627

Macneil, L. T., & Walhout, A. J. M. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research*, *21*(5), 645–657. https://doi.org/10.1101/gr.097378.109

Mäki-Tanila, A., & Hill, W. G. (2014). Influence of Gene Interaction on Complex Trait Variation with Multilocus Models. *Genetics*, *198*(1), 355–367. https://doi.org/10.1534/genetics.114.165282

Malavert, C., Batlla, D., & Benech-Arnold, R. L. (2017). Temperature-dependent regulation of induction into secondary dormancy of Polygonum aviculare L. seeds: A quantitative analysis. *Ecological Modelling*, *352*, 128–138. https://doi.org/10.1016/j.ecolmodel.2017.03.008

Mammadov, J., Aggarwal, R., Buyyarapu, R., & Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *International Journal of Plant Genomics*, *2012*, 728398. https://doi.org/10.1155/2012/728398

Marburger, S., Monnahan, P., Seear, P. J., Martin, S. H., Koch, J., Paajanen, P., Bohutínská, M., Higgins, J. D., Schmickl, R., & Yant, L. (2019). Interspecific introgression mediates adaptation to whole genome duplication. *Nature Communications*, 10(1), 5218. https://doi.org/10.1038/s41467-019-13159-5

Martel, C., Blair, L. K., & Donohue, K. (2018). PHYD prevents secondary dormancy establishment of seeds exposed to high temperature and is associated with lower PIL5 accumulation. *Journal of Experimental Botany*, 69(12), 3157–3169. https://doi.org/10.1093/jxb/ery140

Martínez-Berdeja, A., Stitzer, M. C., Taylor, M. A., Okada, M., Ezcurra, E., Runcie, D. E., & Schmitt, J. (2020). Functional variants of *DOG1* control seed chilling responses and variation in seasonal life-history strategies in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 117(5), 2526–2534. https://doi.org/10.1073/pnas.1912451117

Matsuno, K. (1992). The uncertainty principle as an evolutionary engine. *Biosystems*, 27(2), 63–76. https://doi.org/10.1016/0303-2647(92)90047-3

Mattila, T. M., Tyrmi, J., Pyhäjärvi, T., & Savolainen, O. (2017). Genome-Wide Analysis of Colonization History and Concomitant Selection in Arabidopsis lyrata. *Molecular Biology and Evolution*, 34(10), 2665–2677. https://doi.org/10.1093/molbev/msx193

Matuszewski, S., Hermisson, J., & Kopp, M. (2015). Catch Me if You Can: Adaptation from Standing Genetic Variation to a Moving Phenotypic Optimum. *Genetics*, 200(4), 1255–1274. https://doi.org/10.1534/genetics.115.178574

Mauricio, R. (2001). Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. *Nature Reviews Genetics*, 2(5), 370–381. https://doi.org/10.1038/35072085

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., & Koornneef, M. (1998). *Arabidopsis thaliana*: A Model Plant for Genome Analysis. *Science*, 282(5389), 662–682. https://doi.org/10.1126/science.282.5389.662

Merilä, J., & Sheldon, B. C. (1999). Genetic architecture of fitness and nonfitness traits: Empirical patterns and development of ideas. *Heredity*, 83(2), 103–109. https://doi.org/10.1046/j.1365-2540.1999.00585.x

Merilä, J., & Sheldon, B. C. (2000). Lifetime Reproductive Success and Heritability in Nature. *The American Naturalist*, 155(3), 301–310. https://doi.org/10.1086/303330

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157(4), 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Mirchandani, C. D., Shultz, A. J., Thomas, G. W. C., Smith, S. J., Baylis, M., Arnold, B., Corbett-Detig, R., Enbody, E., & Sackton, T. B. (2024). A Fast, Reproducible, High-throughput Variant Calling Workflow for Population Genomics. *Molecular Biology and Evolution*, 41(1), msad270. https://doi.org/10.1093/molbev/msad270

Miyashima, S., Roszak, P., Sevilem, I., Toyokura, K., Blob, B., Heo, J., Mellor, N., Help-Rinta-Rahko, H., Otero, S., Smet, W., Boekschoten, M., Hooiveld, G., Hashimoto, K., Smetana, O., Siligato, R., Wallner, E.-S., Mähönen, A. P., Kondo, Y., Melnyk, C. W., … Helariutta, Y. (2019). Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature*, 565(7740), 490–494. https://doi.org/10.1038/s41586-018-0839-y

Monroe, J. G., Srikant, T., Carbonell-Bejerano, P., Becker, C., Lensink, M., Exposito-Alonso, M., Klein, M., Hildebrandt, J., Neumann, M., Kliebenstein, D., Weng, M.-L., Imbert, E., Ågren, J., Rutter, M. T., Fenster, C. B., & Weigel, D. (2022). Mutation bias reflects natural selection in Arabidopsis thaliana. *Nature*, 602(7895), 101–105. https://doi.org/10.1038/s41586-021-04269-6

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., & Schumer, M. (2021). The genomic consequences of hybridization. *eLife*, 10, e69016. https://doi.org/10.7554/eLife.69016

Morris, M. R. J. (2014). Plasticity-Mediated Persistence in New and Changing Environments. *International Journal of Evolutionary Biology*, 2014, 1–18. https://doi.org/10.1155/2014/416497

Moutinho, A. F., Bataillon, T., & Dutheil, J. Y. (2020). Variation of the adaptive substitution rate between species and within genomes. *Evolutionary Ecology*, 34(3), 315–338. https://doi.org/10.1007/s10682-019-10026-z

Muller, H. J. (1942). Isolating mechanisms, evolution, and temperature. *Biology Symposium*, 6, 71–125.

Muller, M.-H., Leppälä, J., & Savolainen, O. (2008). Genome-wide effects of postglacial colonization in Arabidopsis lyrata. *Heredity*, 100(1), 47–58. https://doi.org/10.1038/sj.hdy.6801057

Ņečajeva, J., Bleidere, M., Jansone, Z., Gailīte, A., & Ruņģis, D. (2021). Variability of Seed Germination and Dormancy Characteristics and Genetic Analysis of Latvian Avena fatua Populations. *Plants*, 10(2), 235. https://doi.org/10.3390/plants10020235

Née, G., Kramer, K., Nakabayashi, K., Yuan, B., Xiang, Y., Miatton, E., Finkemeier, I., & Soppe, W. J. J. (2017). DELAY OF GERMINATION1 requires PP2C phosphatases of the ABA signalling pathway to control seed dormancy. *Nature Communications*, 8(1). https://doi.org/10.1038/s41467-017-00113-6

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*. https://doi.org/10.1093/oxfordjournals.molbev.a040410

Neiman, M., & Linksvayer, T. A. (2006). The conversion of variance and the evolutionary potential of restricted recombination. *Heredity*, *96*(2), 111–121. https://doi.org/10.1038/sj.hdy.6800772

Nosil, P., Flaxman, S. M., Feder, J. L., & Gompert, Z. (2020). Increasing our ability to predict contemporary evolution. *Nature Communications*, *11*(1), 5592. https://doi.org/10.1038/s41467-020-19437-x

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, *23*, 263–286.

Olson-Manning, C. F., Wagner, M. R., & Mitchell-Olds, T. (2012). Adaptive evolution: Evaluating empirical support for theoretical predictions. *Nature Reviews. Genetics*, *13*(12), 867–877. https://doi.org/10.1038/nrg3322

Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, *10*(8), 531–539. https://doi.org/10.1038/nrg2603

Østman, B., Hintze, A., & Adami, C. (2012). Impact of epistasis and pleiotropy on evolutionary adaptation. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1727), 247–256. https://doi.org/10.1098/rspb.2011.0870

Overcast, I. (2023). *easy*SFS [Computer software]. https://github.com/isaacovercast/easySFS

Paaby, A. B., & Rockman, M. V. (2014). Cryptic genetic variation: Evolution's hidden substrate. *Nature Reviews Genetics*, *15*(4), 247–258. https://doi.org/10.1038/nrg3688

Parsons, T. L., & Ralph, P. L. (2024). Large effects and the infinitesimal model. *Theoretical Population Biology*, *156*, 117–129. https://doi.org/10.1016/j.tpb.2024.02.009

Patton, A. H., Richards, E. J., Gould, K. J., Buie, L. K., & Martin, C. H. (2022). Hybridization alters the shape of the genotypic fitness landscape, increasing access to novel fitness peaks during adaptive radiation. *eLife*, *11*, e72905. https://doi.org/10.7554/eLife.72905

Pavey, S. A., Collin, H., Nosil, P., & Rogers, S. M. (2010). The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences*, 1206(1), 110–129. https://doi.org/10.1111/j.1749-6632.2010.05765.x

Pawłowski, T. A., Bujarska-Borkowska, B., Suszka, J., Tylkowski, T., Chmielarz, P., Klupczyńska, E. A., & Staszak, A. M. (2020). Temperature Regulation of Primary and Secondary Seed Dormancy in Rosa canina L.: Findings from Proteomic Analysis. *International Journal of Molecular Sciences*, 21(19), 7008. https://doi.org/10.3390/ijms21197008

Payne, J. L., & Wagner, A. (2014). The Robustness and Evolvability of Transcription Factor Binding Sites. *Science*, 343(6173), 875–877. https://doi.org/10.1126/science.1249046

Perrier, A., Turner, M. C., & Galloway, L. F. (2025). Shifts in vernalization and phenology at the rear edge hold insight into the adaptation of temperate plants to future milder winters. *New Phytologist*, 246(3), 1377–1389. https://doi.org/10.1111/nph.70005

Postma, F. M., Lundemo, S., & Ågren, J. (2015). Seed dormancy cycling and mortality differ between two locally adapted populations of *Arabidopsis thaliana*. *Annals of Botany*, mcv171. https://doi.org/10.1093/aob/mcv171

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology*, 20(4), R208–R215. https://doi.org/10.1016/j.cub.2009.11.055

Pritchard, V. L., Viitaniemi, H. M., McCairns, R. J. S., Merilä, J., Nikinmaa, M., Primmer, C. R., & Leder, E. H. (2017). Regulatory Architecture of Gene Expression Variation in the Threespine Stickleback *Gasterosteus aculeatus*. *G3 Genes|Genomes|Genetics*, 7(1), 165–178. https://doi.org/10.1534/g3.116.033241

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3), 559–575. https://doi.org/10.1086/519795

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Radersma, R., Noble, D. W. A., & Uller, T. (2020). Plasticity leaves a phenotypic signature during local adaptation. *Evolution Letters*, 4(4), 360–370. https://doi.org/10.1002/evl3.185

Rahnamae, N., Metzger, L., Hördemann, L., Korfmann, K., Khan, A. S., Özoglan, Y., Dent, C. I., Amar, S., Wijfjes, R. Y., Ali, T., Schmitz, G., Stich, B., Tellier, A., & De Meaux, J. (2025). Contemporary hybridization among *Arabis*floodplain species creates opportunities for adaptation. *New Phytologist*, nph.70779. https://doi.org/10.1111/nph.70779

Reif, J. C., Gumpert, F.-M., Fischer, S., & Melchinger, A. E. (2007). Impact of interpopulation divergence on additive and dominance variance in hybrid populations. *Genetics*, *176*(3), 1931–1934. https://doi.org/10.1534/genetics.107.074146

Richards, C. L., Rosas, U., Banta, J., Bhambhra, N., & Purugganan, M. D. (2012). Genome-Wide Patterns of Arabidopsis Gene Expression in Nature. *PLoS Genetics*, *8*(4), e1002662. https://doi.org/10.1371/journal.pgen.1002662

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*, *83*(4), 363–372. https://doi.org/10.1038/sj.hdy.6886170

Romero, I. G., Ruvinsky, I., & Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, *13*(7), 505–516. https://doi.org/10.1038/nrg3229

Sackton, T. B., & Hartl, D. L. (2016). Genotypic Context and Epistasis in Individuals and Populations. *Cell*, *166*(2), 279–287. https://doi.org/10.1016/j.cell.2016.06.047

Salas-Muñoz, S., Rodríguez-Hernández, A. A., Ortega-Amaro, M. A., Salazar-Badillo, F. B., & Jiménez-Bremont, J. F. (2016). Arabidopsis AtDjA3 Null Mutant Shows Increased Sensitivity to Abscisic Acid, Salt, and Osmotic Stress in Germination and Post-germination Stages. *Frontiers in Plant Science*, 7. https://doi.org/10.3389/fpls.2016.00220

Salomon, M. P., Ostrow, D., Phillips, N., Blanton, D., Bour, W., Keller, T. E., Levy, L., Sylvestre, T., Upadhyay, A., & Baer, C. F. (2009). Comparing Mutational and Standing Genetic Variability for Fitness and Size in *Caenorhabditis briggsae* and *C. elegans*. *Genetics*, 183(2), 685–692. https://doi.org/10.1534/genetics.109.107383

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*, 270(5235), 467–470. https://doi.org/10.1126/science.270.5235.467

Schierup, M. H. (2004). The Effect of Enzyme Heterozygosity on Growth in a Strictly Outcrossing Species, the Self-Incompatible Arabis Petraea (Brassicaceae). *Hereditas*, 128(1), 21–31. https://doi.org/10.1111/j.1601-5223.1998.00021.x

Schlindwein, G., Schlindwein, C. C. D., & Dillenburg, L. R. (2019). Seasonal cycle of seed dormancy controls the recruitment of *Butia odorata* (ARECACEAE) seedlings in savanna-like palm tree formations in southern Brazil. *Austral Ecology*, 44(8), 1398–1409. https://doi.org/10.1111/aec.12813

Schmitz, G., Linstädter, A., Frank, A. S. K., Dittberner, H., Thome, J., Schrader, A., Linne von Berg, K., Fulgione, A., Coupland, G., & De Meaux, J. (2024). Environmental filtering of life-history trait diversity in urban populations of *Arabidopsis thaliana*. *Journal of Ecology*, 112(1), 14–27. https://doi.org/10.1111/1365-2745.14211

Schneemann, H., Munzur, A. D., Thompson, K. A., & Welch, J. J. (2022). The diverse effects of phenotypic dominance on hybrid fitness. *Evolution*, evo.14645. https://doi.org/10.1111/evo.14645

Seymour, D. K., Chae, E., Grimm, D. G., Martín Pizarro, C., Habring-Müller, A., Vasseur, F., Rakitsch, B., Borgwardt, K. M., Koenig, D., & Weigel, D. (2016). Genetic architecture of

nonadditive inheritance in *Arabidopsis thaliana*hybrids. *Proceedings of the National Academy of Sciences*, *113*(46). https://doi.org/10.1073/pnas.1615268113

Shaw, R. G., & Shaw, F. H. (2014). Quantitative genetic study of the adaptive process. *Heredity*, *112*(1), 13–20. https://doi.org/10.1038/hdy.2013.42

Simons, A. M. (2011). Modes of response to environmental change and the elusive empirical evidence for bet hedging. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1712), 1601–1609. https://doi.org/10.1098/rspb.2011.0176

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, *23*(1), 23–35.

Soltani, E., Baskin, J. M., & Baskin, C. C. (2019). A review of the relationship between primary and secondary dormancy, with reference to the volunteer crop weed oilseed rape (*Brassica napus*). *Weed Research*, *59*(1), 5–14. https://doi.org/10.1111/wre.12342

Sprague, G. F., & Tatum, L. A. (1942). General vs. Specific Combining Ability in Single Crosses of Corn[1]. *Agronomy Journal*, *34*(10), 923–932. https://doi.org/10.2134/agronj1942.00021962003400100008x

Springthorpe, V., & Penfield, S. (2015). Flowering time and seed dormancy control use external coincidence to generate life history strategy. *eLife*, *4*, e05557. https://doi.org/10.7554/eLife.05557

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*(Web Server), W609–W612. https://doi.org/10.1093/nar/gkl315

Sztepanacz, J. L., & Blows, M. W. (2015). Dominance Genetic Variance for Traits Under Directional Selection in *Drosophila serrata*. *Genetics*, *200*(1), 371–384. https://doi.org/10.1534/genetics.115.175489

Takou, M., Balick, D. J., Steige, K. A., Clo, J., Dittberner, H., Goebel, U., Schielzeth, H., & De Meaux, J. (2022). *Strength of purifying selection on the amino-acid sequence is associated with the amount of non-additive variance in gene expression.* Evolutionary Biology. https://doi.org/10.1101/2022.02.11.480164

Takou, M., Hämälä, T., Koch, E. M., Steige, K. A., Dittberner, H., Yant, L., Genete, M., Sunyaev, S., Castric, V., Vekemans, X., Savolainen, O., & Meaux, J. D. (2021). Maintenance of Adaptive Dynamics and No Detectable Load in a Range-Edge Outcrossing Plant Population. *Molecular Biology and Evolution*, *38*(5), 1820–1836. https://doi.org/10.1093/molbev/msaa322

Takou, M., Wieters, B., Kopriva, S., Coupland, G., Linstädter, A., & De Meaux, J. (2019). Linking genes with ecological strategies in *Arabidopsis thaliana*. *Journal of Experimental Botany*, *70*(4), 1141–1151. https://doi.org/10.1093/jxb/ery447

Thornton, K. R. (2019). Polygenic Adaptation to an Environmental Shift: Temporal Dynamics of Variation Under Gaussian Stabilizing Selection and Additive Effects on a Single Trait. *Genetics*, *213*(4), 1513–1530. https://doi.org/10.1534/genetics.119.302662

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, *32*(3), 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Tran, N. L. T., Ali, T., Schmitz, G., & De Meaux, J. (2025). Heat-Induced Secondary Dormancy Contributes to Local Adaptation in *Arabidopsis thaliana*. *Molecular Ecology*, e70086. https://doi.org/10.1111/mec.70086

Tsouris, A., Brach, G., Schacherer, J., & Hou, J. (2024). Non-additive genetic components contribute significantly to population-wide gene expression variation. *Cell Genomics*, *4*(1), 100459. https://doi.org/10.1016/j.xgen.2023.100459

Turelli, M. (1988). Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution*, *42*(6), 1342–1347.

Turesson, G. (1930). THE SELECTIVE EFFECT OF CLIMATE UPON THE PLANT SPECIES. *Hereditas*, *14*(2), 99–152. https://doi.org/10.1111/j.1601-5223.1930.tb02531.x

Urquhart-Cronish, M., Thompson, K. A., & Angert, A. L. (2025). Joint test of historical vs. Contemporary biogeography supports abundant center hypothesis shaping spatial patterns of self-fertilization. *Evolution Letters*, qraf034. https://doi.org/10.1093/evlett/qraf034

Utami, C. Y., Violle, C., Vile, D., Perrier, L., & Vasseur, F. (2025). From Individual to Stand Performance in Hybrids: Challenging the Optimal Parental Genetic Distance. *Evolutionary Applications*, *18*(10), e70165. https://doi.org/10.1111/eva.70165

van Heerwaarden, B., Willi, Y., Kristensen, T. N., & Hoffmann, A. A. (2008). Population bottlenecks increase additive genetic variance but do not break a selection limit in rain forest Drosophila. *Genetics*, *179*(4), 2135–2146. https://doi.org/10.1534/genetics.107.082768

Van Treuren, R., Kuittinen, H., Karkkainen, K., Baena-Gonzalez, E., & Savolainen, O. (1997). Evolution of microsatellites in Arabis petraea and Arabis lyrata, outcrossing relatives of Arabidopsis thaliana. *Molecular Biology and Evolution*, *14*(3), 220–229. https://doi.org/10.1093/oxfordjournals.molbev.a025758

Vasseur, F., Fouqueau, L., De Vienne, D., Nidelet, T., Violle, C., & Weigel, D. (2019). Nonlinear phenotypic variation uncovers the emergence of heterosis in Arabidopsis thaliana. *PLOS Biology*, *17*(4), e3000214. https://doi.org/10.1371/journal.pbio.3000214

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Volis, S., Ormanbekova, D., Yermekbayev, K., Abugalieva, S., Turuspekov, Y., & Shulgina, I. (2016). Genetic architecture of adaptation to novel environmental conditions in a predominantly selfing allopolyploid plant. *Heredity*, *116*(6), 485–490. https://doi.org/10.1038/hdy.2016.2

Wagmann, K., Hautekèete, N.-C., Piquot, Y., Meunier, C., Schmitt, S. E., & Van Dijk, H. (2012). Seed dormancy distribution: Explanatory ecological factors. *Annals of Botany*, *110*(6), 1205–1219. https://doi.org/10.1093/aob/mcs194

Waldmann, P. (2001). Additive and non-additive genetic architecture of two different-sized populations of Scabiosa canescens. *Heredity*, *86*(Pt 6), 648–657. https://doi.org/10.1046/j.1365-2540.2001.00873.x

Walsh, B., & Lynch, M. (2018a). Appendix 4: Multiple comparisions: Bonferroni corrections, false-discovery rates, and meta-analysis, Combining p Values Over Independent Tests. In *Genetics and Analysis of Quantitative Traits* (pp. 1260–1262). Sinauer Associates, Inc.

Walsh, B., & Lynch, M. (2018b). *Evolution and Selection of Quantitative Traits*.

Walter, G. M., Richards, T. J., Wilkinson, M. J., Blows, M. W., Aguirre, J. D., & Ortiz-Barrientos, D. (2020). Loss of ecologically important genetic variation in late generation hybrids reveals links between adaptation and speciation. *Evolution Letters*, *4*(4), 302–316. https://doi.org/10.1002/evl3.187

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., & Yu, J. (2010). KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genomics, Proteomics & Bioinformatics*, *8*(1), 77–80. https://doi.org/10.1016/S1672-0229(10)60008-3

Wei, X., & Zhang, J. (2018). The optimal mating distance resulting from heterosis and genetic incompatibility. *Science Advances*, *4*(11), eaau5518. https://doi.org/10.1126/sciadv.aau5518

Westneat, David. F., Potts, L. J., Sasser, K. L., & Shaffer, J. D. (2019). Causes and Consequences of Phenotypic Plasticity in Complex Environments. *Trends in Ecology & Evolution*, *34*(6), 555–568. https://doi.org/10.1016/j.tree.2019.02.010

Wieters, B., Steige, K. A., He, F., Koch, E. M., Ramos-Onsins, S. E., Gu, H., Guo, Y.-L., Sunyaev, S., & De Meaux, J. (2021). Polygenic adaptation of rosette growth in Arabidopsis thaliana. *PLOS Genetics*, 17(1), e1008748. https://doi.org/10.1371/journal.pgen.1008748

Willis, C. G., Baskin, C. C., Baskin, J. M., Auld, J. R., Venable, D. L., Cavender-Bares, J., Donohue, K., Rubio De Casas, R., & The NESCent Germination Working Group. (2014). The evolution of seed dormancy: Environmental cues, evolutionary hubs, and diversification of the seed plants. *New Phytologist*, 203(1), 300–309. https://doi.org/10.1111/nph.12782

Willis, J. H., & Orr, H. A. (1993). Increased Heritable Variation Following Population Bottlenecks: The Role of Dominance. *Evolution*, 47(3), 949. https://doi.org/10.2307/2410199

Wills, C. (2007). Principles of Population Genetics, 4th edition. *Journal of Heredity*, 98(4), 382–382. https://doi.org/10.1093/jhered/esm035

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B., & Nussey, D. H. (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology*, 79(1), 13–26. https://doi.org/10.1111/j.1365-2656.2009.01639.x

Wiser, M. J., Ribeck, N., & Lenski, R. E. (2013). Long-Term Dynamics of Adaptation in Asexual Populations. *Science*, 342(6164), 1364–1367. https://doi.org/10.1126/science.1243357

Wolak, M. E. (2012). **nadiv**: An R package to create relatedness matrices for estimating non-additive genetic variances in animal models. *Methods in Ecology and Evolution*, 3(5), 792–796. https://doi.org/10.1111/j.2041-210X.2012.00213.x

Woodward, A. W., & Bartel, B. (2018). Biology in Bloom: A Primer on the *Arabidopsis thaliana* Model System. *Genetics*, 208(4), 1337–1349. https://doi.org/10.1534/genetics.118.300755

Wright, S. (1931). EVOLUTION IN MENDELIAN POPULATIONS. *Genetics*, 16(2), 97–159. https://doi.org/10.1093/genetics/16.2.97

Wright, S. (1948). On the Roles of Directed and Random Changes in Gene Frequency in the Genetics of Populations. *Evolution*, 2(4), 279. https://doi.org/10.2307/2405519

Wright, S. I., Lauga, B., & Charlesworth, D. (2003). Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Molecular Ecology*, *12*(5), 1247–1263. https://doi.org/10.1046/j.1365-294X.2003.01743.x

Wu, J.-R., Wang, T.-Y., Weng, C.-P., Duong, N. K. T., & Wu, S.-J. (2019). AtJ3, a specific HSP40 protein, mediates protein farnesylation-dependent response to heat stress in Arabidopsis. *Planta*, *250*(5), 1449–1460. https://doi.org/10.1007/s00425-019-03239-7

Wyse, S. V., & Dickie, J. B. (2018). Ecological correlates of seed dormancy differ among dormancy types: A case study in the legumes. *New Phytologist*, *217*(2), 477–479. https://doi.org/10.1111/nph.14777

Xu, S. (2013). Mapping Quantitative Trait Loci by Controlling Polygenic Background Effects. *Genetics*, *195*(4), 1209–1222. https://doi.org/10.1534/genetics.113.157032

Yamamichi, M. (2022). How does genetic architecture affect eco-evolutionary dynamics? A theoretical perspective. *Philosophical Transactions of the Royal Society B*, *377*(1855), 20200504. https://doi.org/10.1098/rstb.2020.0504

Yamane, H., Singh, A. K., & Cooke, J. E. K. (2021). Plant dormancy research: From environmental control to molecular regulatory networks. *Tree Physiology*, *41*(4), 523–528. https://doi.org/10.1093/treephys/tpab035

Yang, C. J., Samayoa, L. F., Bradbury, P. J., Olukolu, B. A., Xue, W., York, A. M., Tuholski, M. R., Wang, W., Daskalska, L. L., Neumeyer, M. A., Sanchez-Gonzalez, J. D. J., Romay, M. C., Glaubitz, J. C., Sun, Q., Buckler, E. S., Holland, J. B., & Doebley, J. F. (2019). The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proceedings of the National Academy of Sciences*, *116*(12), 5643–5652. https://doi.org/10.1073/pnas.1820997116

Yang, Z., & Nielsen, R. (2000). Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Molecular Biology and Evolution*, *17*(1), 32–43. https://doi.org/10.1093/oxfordjournals.molbev.a026236

Zhang, Z., Bendixsen, D. P., Janzen, T., Nolte, A. W., Greig, D., & Stelkens, R. (2020). Recombining Your Way Out of Trouble: The Genetic Architecture of Hybrid Fitness under Environmental Stress. *Molecular Biology and Evolution*, 37(1), 167–182. https://doi.org/10.1093/molbev/msz211

Zhao, J., Favero, D. S., Peng, H., & Neff, M. M. (2013). *Arabidopsis thaliana* AHL family modulates hypocotyl growth redundantly by interacting with each other via the PPC/DUF296 domain. *Proceedings of the National Academy of Sciences*, 110(48). https://doi.org/10.1073/pnas.1219277110

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7), 821–824. https://doi.org/10.1038/ng.2310

Ziyatdinov, A., Vázquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., & Soria, J. M. (2018). lme4qtl: Linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics*, 19(1), 68. https://doi.org/10.1186/s12859-018-2057-x

Zomer, M., Moreira, B., & Pausas, J. G. (2022). Fire and summer temperatures interact to shape seed dormancy thresholds. *Annals of Botany*, 129(7), 809–816. https://doi.org/10.1093/aob/mcac047

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., … Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261–1277. https://doi.org/10.1111/ecog.04960

UNIVERSITÄT
ZU KÖLN

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen

Fakultät der Universität zu Köln

vorgelegt von

Loc Thuy Nhu Tran

Köln, 18. Januar 2026