The polygenic basis of local adaptation in *Arabidopsis lyrata*

Inaugural-Dissertation

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

Vorgelegt von

Margarita Takou

aus Alexandroupoli, Grienchenland

Köln, 2020

Berichterstatter: Prof. Dr. Juliette de Meaux

(Gutacher) Prof. Dr. Michael Nothnangel

Tag der Disputation: 20.01.2021

Abstract

The global environment changes at an unprecedented speed, posing novel threats and challenges for many plant and animal species. Populations in the wild will have to adapt to these new environments. Adaptation requires phenotypic variation that affects fitness, much of which is controlled by complex polygenic backgrounds. The aim of the present thesis is to investigate the polygenic basis of adaptation, using as study model the outcrossing and perennial plant species *Arabidopsis lyrata* ssp *petraea.* I focused the analysis on two *A. lyrata* populations originating from constrasted environments in the core and edge of the species distribution range in Europe.

I first asked whether there is maintenance of adaptive dynamics and genetic diversity in the range edge population. During range expansion, edge populations are expected to face increased genetic drift, which in turn can alter and potentially compromise adaptive dynamics, preventing the removal of deleterious mutations and slowing down adaptation. I documented a sharp decline in effective population size in the range-edge population and observed that non-synonymous variants segregate at higher frequencies. A 4.9% excess of derived non-synonymous variants per individual was detected, suggesting an increase of the genomic burden of deleterious mutations in the range-edge population. We predicted, however, that most of these variants have a small effect on fitness. Consistent with this prediction, the range-edge population was not impaired in its growth and survival measured in a common garden experiment. Genomic footprints indicative of selective sweeps were broader in the Northern population but not less frequent, indicating that the smaller population had maintained its ability to adapt.

Adaptation at the level of transcript regulation is believed to make a significant contribution to complex trait adaptation in natural populations. From an evolutionary perspective, it is of great importance to identify the part of a phenotypes' genetic variance, that can be directly inherited to the next generation (additive variance) and thus directly contribute to adaptation. In fact, the global gene expression as documented by a complex set of crosses between populations, revealed substantial additive variance, distributed all over the genome. Yet, dominance variance, which results from allelic interactions within or between loci, forms the most significant part of the genetic variance in gene expression. I could show that dominance variance is related to gene structural properties, such as length, number of exons and number of transcription factor binding

sites, as well as to the degree of gene co-regulation. On the contrary, additive variance was independent of such signals. Additionally, I could identify that transcript with less genetic variance at the transcript level exhibit stronger constraints at the amino-acid level, indicating that purifying selection acts at both amino-acid and transcript levels. By contrast, transcripts with highest additive variance tended to evolve under relaxed selection.

The genetic basis of gene expression can be further investigated, by identifying the *trans*- and *cis*-regulatory effects. A small number of *trans* effect variants were identified controlling gene expression variation. I only detected significant associations for about 1% of the expressed transcripts, indicating that expression variation in most transcripts has a polygenic basis. Interestingly, these also showed higher levels of additive variance than transcripts whose variation showed a significant association. In addition, we found that transcript with a detectable *cis*-acting variant tended to show higher additive variance.

Altogether, these studies show that despite a dramatic bottleneck and a mild expansion load, adaptive mutations were present in sufficient number to maintain adaptive dynamics at the range-edge of the strictly outcrossing species *Arabidopsis lyrata* ssp. *petraea*. However, although we find evidence that gene expression variation contributes to the evolutionary potential of these populations, we also observe that a significant fraction of genetic variation is not directly available for selection.

Zusammenfassung

Die globale Umwelt verändert sich in einem beispiellosen Tempo und stellt viele Pflanzen- und Tierarten vor neue Bedrohungen und Herausforderungen. Die Populationen in der freien Natur müssen sich an diese neuen Umgebungen anpassen. Die Anpassung erfordert phänotypische Variationen, die sich auf die Fitness auswirken und zu einem großen Teil durch komplexe polygene Hintergründe gesteuert werden. Das Ziel der vorliegenden Arbeit ist es, die polygene Grundlage der Anpassung zu untersuchen, wobei als Studienmodell die auskreuzende und mehrjährige Pflanzenart Arabidopsis lyrata ssp petraea verwendet wird. Ich konzentrierte die Analyse auf zwei A. lyrata-Populationen, die aus kontrastreichen Umgebungen im Kern und Rand des Verbreitungsgebietes der Art in Europa stammen.

Ich fragte zunächst, ob die Anpassungsdynamik und genetische Vielfalt in der Population am Rande des Verbreitungsgebiets erhalten bleibt. Während der Ausdehnung des Verbreitungsgebiets ist zu erwarten, dass die Populationen am Rande des Verbreitungsgebiets einer verstärkten genetischen Drift ausgesetzt sind, die wiederum die Anpassungsdynamik verändern und möglicherweise beeinträchtigen kann, wodurch die Beseitigung schädlicher Mutationen verhindert und die Anpassung verlangsamt wird. Ich dokumentierte einen starken Rückgang der effektiven Populationsgröße in der Randpopulation und beobachtete, dass sich nicht-synonyme Varianten bei höheren Frequenzen segregieren. Es wurde ein Überschuss von 4,9% an abgeleiteten nicht-synonymen Varianten pro Individuum festgestellt, was auf eine Zunahme der genomischen Belastung durch schädliche Mutationen in der Range-Randpopulation hindeutet. Wir sagten jedoch voraus, dass die meisten dieser Varianten einen geringen Einfluss auf die Fitness haben. In Übereinstimmung mit dieser Vorhersage wurde die Population der Range-Redge-Population in ihrem Wachstum und Überleben, gemessen in einem gemeinsamen Gartenexperiment, nicht beeinträchtigt. Genomische Fußabdrücke, die auf selektive Sweeps hindeuteten, waren in der nördlichen Population breiter, aber nicht weniger häufig, was darauf hindeutet, dass die kleinere Population ihre Anpassungsfähigkeit erhalten hatte.

Man geht davon aus, dass die Anpassung auf der Ebene der Transkriptionsregulation einen bedeutenden Beitrag zur komplexen Anpassung von Merkmalen in natürlichen Populationen leistet. Aus evolutionärer Sicht ist es von großer Bedeutung, den Teil der genetischen Varianz

eines Phänotyps zu identifizieren, der direkt an die nächste Generation vererbt werden kann (additive Varianz) und somit direkt zur Anpassung beiträgt. Tatsächlich zeigt die globale Genexpression, wie sie durch einen komplexen Satz von Kreuzungen zwischen Populationen dokumentiert ist, eine beträchtliche additive Varianz, die über das gesamte Genom verteilt ist. Dennoch bildet die Dominanzvarianz, die aus allelischen Interaktionen innerhalb oder zwischen Loci resultiert, den bedeutendsten Teil der genetischen Varianz in der Genexpression. Ich konnte zeigen, dass die Dominanzvarianz mit genstrukturellen Eigenschaften, wie Länge, Anzahl der Exons und Anzahl der Bindungsstellen des Transkriptionsfaktors, sowie mit dem Grad der Gen-Koregulation zusammenhängt. Im Gegenteil, die additive Varianz war unabhängig von solchen Signalen. Zusätzlich konnte ich feststellen, dass Transkripte mit geringerer genetischer Varianz auf der Transkriptebene stärkere Einschränkungen auf der Aminosäureebene aufweisen, was darauf hindeutet, dass die reinigende Selektion sowohl auf der Aminosäure- als auch auf der Transkriptebene wirkt. Im Gegensatz dazu neigten Transkripte mit der höchsten additiven Varianz dazu, sich unter entspannter Selektion zu entwickeln.

Die genetische Basis der Genexpression kann weiter untersucht werden, indem die trans- und cis-regulierenden Effekte identifiziert werden. Es wurde eine kleine Anzahl von Varianten des trans-Effekts identifiziert, die die Genexpressionsvariation kontrollieren. Ich konnte nur bei etwa 1% der exprimierten Transkripte signifikante Assoziationen feststellen, was darauf hindeutet, dass die Expressionsvariation in den meisten Transkripten eine polygene Basis hat. Interessanterweise zeigten diese auch höhere Niveaus additiver Varianz als Transkripte, deren Variation eine signifikante Assoziation zeigte. Darüber hinaus fanden wir heraus, dass Transkripte mit einer nachweisbaren cis-wirkenden Variante tendenziell eine höhere additive Varianz aufwiesen.

Insgesamt zeigen diese Studien, dass trotz eines dramatischen Engpasses und einer milden Expansionsbelastung adaptive Mutationen in ausreichender Zahl vorhanden waren, um die adaptive Dynamik an der Verbreitungsgrenze der streng auskreuzenden Art Arabidopsis lyrata ssp. petraea aufrechtzuerhalten. Obwohl wir jedoch Hinweise darauf finden, dass die Genexpressionsvariation zum evolutionären Potenzial dieser Populationen beiträgt, stellen wir auch fest, dass ein signifikanter Anteil der genetischen Variation nicht direkt für die Selektion zur Verfügung steht.

Table of Contents

Publications

M Takou, B Wieters, S Kopriva, G Coupland, A Linstädter, J De Meaux (2019). Linking genes with ecological strategies in Arabidopsis thaliana.. Journal of Experimental Botany 70 (4), 1141-1151

M Takou, T Hämälä, KA Steige, E Koch, H Dittberner, L Yant, M Genete, S Sunyaev, V Castric, X Vekemans, O Savolainen, J de Meaux (2020). Maintenance of adaptive dynamics in a bottlenecked range-edge population that retained out-crossing. BioRXiv, doi: https://doi.org/10.1101/709873

Author's Contribution

Part of the thesis was a collaborative project between the labs of Prof Dr. Juliette de Meaux, Prof Dr. Outi Savolainen, prof Dr. Xavier Vekemans and prof Dr. Shamil Sunyaev. Bellow, the contribution of each author to the different aspect of the analysis is given.

Chapter	**Study Design**	**Data Collection**	**Data Analysis**		**Manuscript Preparation**
1	**MT**, JdM, OS, VC, XV, SY	**MT**	**MT**	Data preparation, genomic load, sweeps, genetic differentiations, GOs, phenotypic data analysis, synthesis	**MT**, JdM, SY, KS, VC, EK, XV
			TH	Demography	
			KS, EK	DFE	
			XV, VC	S locus analysis	
2	**MT**, JdM	**MT**	**MT**	Data preparation, phenotypic variance, GOs, gene co-expression clustering, statistical analysis, synthesis	**MT**, JdM
			KS	DFE	
			HD	Random forest	
3	**MT**, JdM	**MT**	**MT**	Data preparation, GWAS, GOs, statistical analysis, synthesis	**MT**, JdM
			FH	Mapping bias	

Kim A Steige (KA), Tuomas Hämälä (TH), Evan Koch (EK), Outi Savolainen (OS), Vincent Castric (VC), Xavier Vekemans (XV), Shamil Sunyaev (SY),, Hannes Dittberner (HD), Fei He (FH), Juliette de Meaux (JdM), **Margarita Takou (MT)**

List of Figures

List of Tables

1.Introduction

Global climate change is a reality that proceeds fast. Extended periods of high temperature and net declines in soil moisture are expected in many regions (IPCC, 2013). These changes will impact the distributions and survival of species, which will have to respond to the fast shifting environment, or they have to face extinction. Five percent of all species predicted to go extinct due to increasing temperatures (IPBES 2019). One possible scenario is that the species will have to shift their range distribution, in line (Waldvogel et al. 2020). The impact of such range shifts can be studied by investigating the impact of the postglacial colonization events.

1.1 Genetic load accumulated during range expansions interferes with adaptive dynamics

Range expansion events, such as the postglacial colonization of Northern Europe and Scandinavia from Southern refugia, have had wide influence on the distribution of genetic diversity within species (Hewitt 2000). Through its impact on multiple population genetic processes, range expansion has cascading effects on adaptive dynamics (Excoffier et al. 2009). Indeed, it increases drift (Hallatschek et al. 2007), and leads to both a progressive loss of genetic diversity and increased levels of population differentiation along the expansion route (Austerlitz et al. 1997; Corre and Kremer 1998; Muller et al. 2008; Excoffier et al. 2009; Slatkin and Excoffier 2012). As a consequence, fitness is expected to decrease at the front of the expanding range, causing what is known as the expansion load. The majority of those mutations remain at low frequencies or are lost, but some quickly fix, a phenomenon sometimes termed allele surfing (Klopfstein et al. 2006; Peischl et al. 2013). Although non-synonymous and potentially deleterious mutations are more likely to fix in bottlenecked populations, where the removal of new deleterious mutations is less efficient, it takes some evolutionary time until a significant load accumulates (Lohmueller 2014; Simons et al. 2014; Balick et al. 2015; Do et al. 2015).

Expansion load can interfere with adaptive dynamics. Locally adapted populations that move out of their core range are expected to evolve towards new adaptive peaks (Colautti and Barrett 2013; Savolainen et al. 2013; Wos and Willi 2018). In a population carrying an expansion load, larger adaptive steps might be required to establish a novel range edge, resulting in a slowdown of expansion, especially when dispersal is limiting (Henry et al. 2015). Theoretical studies report

complex interactions among parameters such as the strength and heterogeneity of selection, the rate of expansion, as well as the architecture of traits under selection. Expansion rate and adaptive requirements to the newly colonized environments can jointly modulate the fitness decrease observed at the range edge (Gilbert et al. 2017; Gilbert et al. 2018). However, to the best of our knowledge, these predictions remain practically untested in natural populations.

The speed of range expansion can also be limited by species interactions, if these are necessary for reproductive success and survival (Louthan et al. 2015). Many flowering plants rely on insects for pollination and thus fertility (Gibbs 2014). As species expand their range, efficient pollinators can become rare, and a shift towards selfing may help restore reproductive assurance and avoid Allee effects (Jain 1976; Morgan et al. 2005; Gascoigne et al. 2009). Transitions to selfing or mixed-mating systems have often been associated with range expansion (Baker 1995; Goodwillie et al. 2005; Levin 2010; Cheptou 2012; Laenen et al. 2018). However, mating system shifts can compromise adaptive processes by exposing populations to inbreeding depression and loss of genetic diversity as they face stress at the margin of their ecological niche (Baker 1995; Slatkin 1995; Ingvarsson 2002; Barrett 2003; Glémin and Ronfort 2013). Yet, increases in the selfing rate can also contribute to the purging of deleterious mutations (Pujol et al. 2009; Glémin and Ronfort 2013; Hadfield et al. 2017; Roessler et al. 2019) and promote the emergence of high fitness individuals at the front range of expansion (Klopfstein et al. 2006). In fact, selfing species generally show the greatest overall range size (Grossenbacher et al. 2015). In this context, plant species that have maintained a strictly outcrossing mating system across their expanded distribution range are particularly intriguing.

1.2 Complex adaptive dynamics required: example of coordinating flowering time and dormancy in plant populations

Adaptation is a complex process which occurs via selection on fitness traits and can lead to ecological specialization and speciation (Kawecki and Ebert 2004). It is complex because fitness results from the inter-dependent action of multiple phenotypes, each controlled by different genetic architecture, I illustrate this with the example of flowering time and dormancy, whose contribution to local adaptation cannot be understood in isolation. These two traits have been well studied in the model plant species *Arabidopsis thaliana.*

Local adaptation of *A. thaliana* populations has been documented throughout its range, despite a history of pervasive gene flow (Fournier-Level et al. 2011; Hancock et al. 2011; Ågren and Schemske 2012; Savolainen et al. 2013; Weigel and Nordborg 2015; Svardal et al. 2017). Field experiments and correlation analyses with climate parameters identified numerous genomic regions associated with local climatic conditions (Hancock et al. 2011; Lasky et al. 2014). Among them, SNPs associating with fitness differences in the field were also over-represented (Hancock et al. 2011), while alleles associating with fruit production are more frequent in populations closer to field sites where the selective advantage was expressed (Fournier-Level et al. 2011). Therefore, it is clear that much of the variation found in this species has played a role in optimizing plant performance to local environmental conditions.

Flowering time is one of the developmental traits underpinning adaptation in *A. thaliana*. Variation in flowering time follows climatic clines, both at the regional and species levels (Mendez-Vigo et al. 2011; Montesinos-Navarro et al. 2011; Debieu et al. 2013; Li et al. 2014; Sasaki et al. 2015). Warmer climates appear to favor earlier flowering time, a pattern that has been documented for a great number of species (Austen et al. 2017; Whittaker and Dean 2017). Strong selection for early flowering has been detected in Italy, where as selection for this trait is weaker in Sweden (Ågren et al. 2017). Population genetics studies also uncovered signatures of natural selection on genes controlling flowering time (Corre 2005; Toomajian et al. 2006). Finally, the analysis of co-variation between environmental and phenotypic variance consolidated evidence for the adaptive distribution of this trait (Heerwaarden et al. 2015).

The adaptive relevance of variants which modulate flowering time must be examined in the context of variation for the timing of germination. Much of the flowering time variation measured in the lab, does not manifest as variation in flowering phenology in the field (Wilczek et al. 2009; Brachi et al. 2010; Hu et al. 2017). Rather, flowering time in the field is tightly dependent on the prevailing environmental conditions during seedling establishment, and hence on another developmental trait: the timing of germination (Donohue 2002; Wilczek et al. 2009). Both field experiments and theoretical models integrating seed and flowering phenology have shown that seed dormancy is often decisive for controlling the life cycle across environments (Chiang et al. 2013; Burghardt et al. 2015).

There is indeed consistent support for the adaptive relevance of traits determining the timing of germination. Seed dormancy has a strong fitness advantage before the hot season but can impair fitness if it delays plant growth before winter (Donohue 2002; Donohue et al. 2005; Chiang et al. 2013). Population genetics analysis of seed dormancy and its major QTL *DOG1* supported the adaptive importance of strong dormancy in Southern regions, to escape dry and hot summers, whereas weaker dormancy was reported in Norway, where the season is shorter (Kronholm et al. 2012; Kerdaffrec et al. 2016; Postma and Ågren 2016).

Since flowering time determines the maternal environment the seeds experience during their maturation, it also impacts life history traits expressed by the next generation (Chiang et al. 2013; He et al. 2014; Morrison and Linder 2014; Kerdaffrec et al. 2016). Light, temperature, nutrient availability, and water status have all been identified as significant environmental factors influencing the maternal inheritance of seed dormancy (Footitt et al. 2013; He et al. 2014; Morrison and Linder 2014; Kerdaffrec et al. 2016). Germination can also be distributed over more than one seasonal window. For example, maintaining a spring germinating cohort is important for the maintenance of populations exposed to low winter temperature (Picó 2012; Akiyama and Ågren 2014). Furthermore, later flowering can lead to late seed dispersal, which can result in overwintering at the seed stage (Hu et al. 2017).

A first consequence of increased climatic unpredictability is that ecological shifts towards increasingly ruderal strategies will be promoted. The study of flowering time variation in *A. thaliana* has demonstrated that species are not limited in the number of mutations that can promote accelerated flowering (Mendez-Vigo et al. 2011; Sanchez-Bermejo et al. 2012; Hepworth and Dean 2015; Whittaker and Dean 2017). As a matter of fact, earlier flowering seems to be globally under selection (Munguía-Rosas et al. 2011). Flowering time and seed dormancy are therefore jointly subject to fluctuating seasonal selective forces. They can evolve as a syndrome, defining distinct life history strategies that have diversified across environments (Chiang et al. 2013; Vidigal et al. 2016). An analysis of seed dormancy and flowering time co-variation revealed that the optimization of the two traits probably depends on latitudinal differences in climate. Late flowering (i.e. vernalization requirement) and strong dormancy are more frequent in regions where summer drought is typically more severe, whereas late flowering in Northern latitudes co-varies negatively with dormancy (Debieu et al. 2013). Co-variance between flowering time and

dormancy is also detected at much smaller scale, along steep altitudinal gradients (Vidigal et al. 2016). Normally, diverse life-history trait combinations can allow comparable population growth rates in field conditions (Taylor et al. 2017). In some years, however, early winter frost can wipe out genotypes expressing inadequate life histories (Hu et al. 2017). Minimum winter temperature and precipitation, in fact, were also the main climatic factors that acted as selective pressures on flowering time and their underpinning genes in a set of Iberian *A.thaliana* genotypes (Mendez-Vigo et al. 2011). This suggests that extreme deviations from seasonal climatic averages may be important in determining allelic combinations of genetic variants which contribute to the adjustment of development to optimize growth in different environments throughout the species range.

Thus, as illustrated by the variation and co-variation of flowering time and dormancy, life history traits are intertwined in their effect on fitness and in their genetic basis. Understanding adaptation therefore cannot be advanced by focusing on single trait but would benefit from a more complete view of the relationship between phenotypes, environment and genetics.

1.3 Gene expression variation provides insight on the selection potential of traits with polygenic background

It is not possible to study all phenotypes but studying the expression level of all genes may provide a first approximation of the extent of population variation that is available for adaptation. Gene expression evolution has a critical role in adaptation, as expression evolution can shape the function of genetic mechanisms (Romero et al. 2012). To date, some studies have demonstrated that within population differences in gene expression is not random but the result of selection acting on natural quantitative variance (Oleksiak et al. 2002) and that gene expression does not evolve according to strictly neutral models (Rifkin et al. 2005).

In many of these studies, however, our understanding of the potential role of gene expression for adaptation is limited by the missing information on the heritability of this variation. Determining the evolutionary potential of gene expression variance requires establishing the relative importance of heredity versus the environment, which is directly described by the broad sense heritability. Furthermore, we can use the narrow sense heritability, which is the extent to which transmitted genes affect the phenotypes, in order to further understand the evolutionary impact of the genetic background on gene expression, without having to identify the genes in direct control of its

variance (Falconer and Mackay 1996; Lynch and Walsh 1998). Heritability of gene expression variance has been studied in a number of species, including in humans (Price et al. 2011), yeast (Liu et al. 2020), mice (Cui et al. 2006), fish (Leder et al. 2015; He et al. 2017) and plants (Stupar et al. 2007; Groen et al. 2020), all of which indicate that gene expression has substantial narrow sense and broad sense heritability.

Quantifying the narrow and broad sense heritability is crucial to understand how evolution and adaptation proceeds in the populations. Indeed, according to Fisher's fundamental theorem of natural selection, selection on a specific phenotype will act on the additive variance, and thus the transmitted alleles controlling it, resulting in a progressive decreasing of the narrow sense heritability present in the adapted population (Mousseau and Roff 1987; Orr 2009). This, however, is the case only for traits with simple genetic architecture, because selection will lead to the fixation of the few advantageous alleles within a population and thus the additive variance drops (Hill et al. 2008). On the contrary, when genetic drift is the primary force affecting the allele frequencies, additive variance can increase (Crnokrak and Roff 1995; Hill et al. 2008), creating a complicated balance between selection and stochastic demographic events. Moreover, the difference between the narrow sense and broad sense heritability, which represents all the genic and allelic interaction within and between loci (dominance variance), should not be neglected. For instance, in sticklebacks, the narrow sense heritability of gene expression is high, despite high levels of dominance. This could be attributed to signatures of directional selection within differentially expressed genes (Leder et al. 2015). In *A. thaliana*, non-heritable expression manifested as hybrid heterosis can have a strong effect on complex traits (Vasseur et al. 2019). Therefore, the type of genetic variance has important implication both for the adaptive potential of a population and for its history of adaptation. Yet, few studies have investigated the genomic and evolutionary factors that influences level of additive and dominance variance present in natural populations.

Studying gene expression variance can give insights on the amount of heritable genetic factors shaped by selection, but it can also provide insight on the molecular functions that are targeted by selection. The initial adaptation of complex traits towards a new fitness optimum will often happen by the quick fixation of a few loci (Orr 2009). However, on a molecular level, natural selection as adaptation to a slowly changing environment can rely on mutations of small effect (Collins and Meaux 2009; Yeaman 2015). Small effect mutations can have a cumulative effect on the

phenotype, even if each separately explains only a small part of the phenotypic variance (Falke et al. 2013). The omnigenic model proposes that almost all expressed genes affect a specific phenotype. Each phenotype is controlled by a set of core genes, which have a direct effect on the phenotypic variance, as well as by the peripheral genes that control the phenotype via direct effect on the core genes (Boyle et al. 2017; Liu et al. 2019). Thus, incorporating the mutations with small effect can significantly improve our ability to detect natural selection on phenotypic traits (Berg and Coop 2014).

Using gene expression, we can study how small effect mutations are distributed in the genome by analyzing *cis* regulatory mutations controlling gene expression. *Cis* regulatory variants are in linkage with the expressed transcript, so that they can be easily identified by genome wide allele-specific expression differences in hybrid heterozygotes (de Meaux 2006; He et al. 2012; He et al. 2016; Steige et al. 2017). Their large numbers allows investigating the genomic and evolutionary properties that associate with their distribution (Fay and Wittkopp 2008; He et al. 2016; de Meaux 2018). *Cis* acting regulatory changes tend to accumulate within species at adaptive traits (Erwin and Davidson 2009). They can contribute to reshape polygenic molecular functions even in closely related species. For instance, *cis* acting regulatory changes have contributed to adaptation to heavy metal substrate in *Arabidopsis halleri* in contrast to its close relative *Arabidopsis thaliana* (He et al. 2016). Also, floral adaptations and mating system shift between *Capsella grandiflora* and *Capsella rubella* are supported by *cis* acting regulatory changes (Steige et al. 2015). Therefore, measuring natural selection by *cis* acting changes can lead to better understanding of phenotypic evolution, especially when combined with population genetics (Emerson and Li 2010).

Gene expression variance is encoded both by cis- and trans-regulatory changes, but the role of this variation may be different. *Cis* regulatory changes could be influenced by adaptive polygenic selection (Fraser et al. 2011) but also preferentially fixed due to weaker pleiotropic effects compared to *trans*- regulatory changes (Prud'homme et al. 2007). Different mutation rates and dominance allelic interactions of *cis* and *trans* regulators might also affect the proportion that each of them contributes to intra- and inter-specific diversity, with, for example, *cis* acting alleles being more often under positive selection in *Drosophila melanogaster* (Lemos et al. 2008). Alternatively, their mutational variance over time is small, due to the fact that most mutations affecting gene expression are deleterious with negative fitness impact (Gilad et al. 2006). *Cis*-regulatory variants

can accelerate the response to negative selection through exposing deleterious mutations (Erwin and Davidson 2009). In *C. grandiflora,* the *cis* regulatory expression was mainly deleterious and in low frequencies, providing evidence for purifying selection acting on genetic variation within population (Josephs et al. 2015), even though genes with *trans* effect are also under stronger negative selection (Josephs et al. 2020). Thus, it is clear that gene expression can provide insight in the selection dynamics of polygenic traits and complex traits, but that it is also necessary to study the role of cis- and trans- acting changes, because their contribution to adaptation may differ.

1.4 *Arabidopsis lyrata* as a study system to investigate the polygenic basis of selection

Arabidopsis thaliana is the model system for plant research. Despite the wealth of genetic and phenotypic information that the community has gained via projects such as the 1001 genomes (1001 Genomes Consortium 2016), the species has a few limitations for studying in depth the impact of selection and adaptation. *A. thaliana,* is an annual selfing species, which has experienced strong bottlenecks (Durvasula et al. 2017; Svardal et al. 2017), increased genetic drift (Weigel and Nordborg 2015; Svardal et al. 2017) and exhibits population structure along clines (Zou et al. 2017). Those two factors are often cofounded with selection signals and can mislead studies of adaptation. For these reasons, I have used for this study, *A.thaliana's* outcrossing and perennial relative, *Arabidopsis lyrata* (Clauss and Koch 2006).

Populations in *A. lyrata* can be exposed to very different local conditions. The European sub-species *Arabidopsis lyrata ssp. lyrata* has expanded its range Northwards after the last glaciation (Clauss and Koch 2006; Schierup et al. 2006; Koch 2019). Its patchy populations are found from Central Europe to the North of Scandinavia (Hoffmann 2005). Northern populations in *A. l. ssp. petraea* show a strong reduction in diversity (Wright et al. 2003; Muller et al. 2008; Ross-Ibarra et al. 2008; Pyhäjärvi et al. 2012; Mattila et al. 2017). Yet, there is evidence that *A. l. ssp. petraea* populations at the Northern range edge are locally adapted. Reciprocal transplant studies between the Northern and Central European populations showed that Northern populations have the highest survival rate in their location of origin consistent with signals of local adaptation (Leinonen et al. 2009). Major developmental traits such as flowering time, as well as the response to abiotic stress factors, seem to have been targets of natural selection (Sandring et al. 2007; Toivainen et al. 2014; Mattila et al. 2016; Davey et al. 2018; Hämälä and Savolainen 2018). Reciprocal transplant experiments across four sites in Europe, as well as between populations of different altitude in

Norway, indicated that populations at the range margins were locally adapted (Hämälä and Savolainen 2018). Furthermore, allele specific expression differences have been detected in interspecies hybrids as well as between populations (He et al. 2012; He et al. 2016; Videvall et al. 2016).

The European subspecies of *A. lyrata* has maintained outcrossing, via well documented mechanism. *A. lyrata ssp. lyrata* enforces self-incompatibility (SI) via the multiallelic S-locus specific to the Brassicaceae family (Bateman 1955; Kusaba et al. 2001). Phylogenetic and genomic analyses of the S-locus have shown that strong negative frequency-dependent selection caused early diversification of the S-locus within the family and a high degree of sharing of S-allele lineages across species (Dwyer et al. 1991; Vekemans et al. 2014). The loss of SI, however, evolved repeatedly in the family (Tsuchimatsu et al. 2012; Vekemans et al. 2014; Durvasula et al. 2017). In fact, some populations of the closely related North American subspecies *A. l. ssp. lyrata*, lost obligate outcrossing at their range margin (Mable et al. 2005; Griffin and Willi 2014; Willi et al. 2018). This transition to selfing has been recently associated with a sharp decrease in average population fitness (Willi et al. 2018). In the sub-species *A. l. ssp. petraea*, instead, SI appears to have been maintained, presumably due to the inbreeding depression, which has been demonstrated using forced selfing (Kärkkäinen et al. 1999; Sletvold et al. 2013).

Therefore, *A. lyrata ssp. petraea* is a good model system to study the effects of range expansion and the subsequent adaptation on the genetic basis of complex traits between outcrossing plant populations.

1.5 Aims of the study

The main goal of the present thesis is to investigate the signatures of polygenic selection in the outcrossing plant species *A. lyrata* ssp *petraea*. Towards that end, I contrasted two populations representative of the species' European distribution, and specifically I compared a range -edge population contrasted to a range core population. In order to explore the main research question, the project was organized in the following three investigational chapters:

Chapter 1: Is there detectable genetic diversity and adaptive evolution in the range edge population?

Range expansions have been documented to impact both the level of genetic diversity and therefore the efficacy of selection, especially in range edge populations. To gain insight into the combined effects of demographic history and selection, I compared the two populations to ask i) can we document a decreased efficacy of negative selection and an increase of genetic load in the range edge population? Furthermore, ii) does the range – edge population show a decrease in S -alleles, which is related to mating shift? iii) Is there impaired growth in the range edge population indicative of genetic load? And finally, iv) can I detect a slowdown of adaptive dynamics in the range edge population?

Chapter 2: What is the adaptive potential of gene expression variation?

Gene expression variation can provide insight on the impact of small effect mutations, and thus polygenic selection, within the species, as well as the potential to respond to selection. I generated between population hybrids and obtained their transcriptome in order to investigate the following questions: i) how much of the gene expression variation is heritable? ii) Is the relative importance of the non-heritable gene expression variance random? and if it is not random what are the factors that influence the amount and genomic distribution of heritable and non-heritable variance?

Chapter 3: What is the genetic basis of gene expression variation?

Divergence in the *trans* and *cis* regulatory elements between populations originating from different environments, accumulates in functions important for local adaptation. I used the transcriptome dataset generated for Chapter 2, to further ask the questions: i) how complex is the genetic basis of gene expression? ii) is there connection of the genetic basis with signals of adaptation? iii) do they accumulate in specific functions? and iii) do they associate with heritability?

2. Material and Methods

2.1 Chapter 1: Genetic Diversity and adaptive evolution in a range-edge population

2.1.1 Plant Material, Sequencing and Data Preparation

Genomic sequences of *A. l. ssp. petraea* populations of 22 individuals originating from Spiterstulen in Norway (SP; 61.41N, 8.25E), 17 individuals originating from Plech in Germany (PL; 49.65N, 11.45E) and a scattered sample of 7 individuals from Austria (AUS; 47.54N, 15.58E; 47.55N, 15.59E; 47.58N, 16.9E) were used in the analysis. Lab generated seeds of SP plants and field collected seeds of PL were grown in greenhouse conditions. I sequenced 11 unrelated PL and 10 unrelated SP individuals. The rest of the sequences were obtained from previously published data. I obtained 6 PL sequences and 5 sequences of field collected SP individuals from Mattila et al. (2017) and 7 sequences of field collected SP from Hämälä and Savolainen (2018). The seven sequences of diploid AUS individuals were provided by Marburger et al. (2019). In total, 22 SP, 17 PL and 7 AUS whole genome sequences were included in the analysis. All the genome sequences are available online (see section Data Availability).

Read quality was checked with FastQC and the last two bases of the sequences were removed with cutadapt v1.14 (Martin 2011). Reads were mapped against the reference genome of *Arabidopsis lyrata* (Hu et al. 2011) Ensembl v1.0 with bwa mem, options -M (Li and Durbin 2009). Only the reads mapped against the main eight chromosomes were kept in the analysis. Samtools v1.5 (Li et al. 2009) was used for quality filtering (view -f 3 -q30 -F 264) and remove of PCR duplicates (rmdup). Indels were realigned with Genome Analysis Toolkit (McKenna et al. 2010) version nightly-2017-12-11-1.

Single nucleotide polymorphisms (SNPs) were called with samtools mpileup, with options -E -q 30, and bcftools call (Li et al. 2009) with options -p 0.01. Since repeat regions align poorly, they were flagged using the script generate_masked_ranges.py (https://gist.github.com/danielecook/cfaa5c359d99bcad3200) and subsequently were removed with bedtools subtract, v2.25.0 (Quinlan and Hall 2010). Sites where all individuals are heterozygous for one population likely result from excessive paralogous mapping and thus were removed with a custom python script. Indels and sites that had more than two alleles, coverage less than 10, genotype quality less than 20 or quality less than 30 were filtered out with VCFtools

v0.1.5 (Danecek et al. 2011). Also, the sites that had more than 80% of the individual data missing were removed. In the end, 1,878,003 SNPs were used for the downstream analysis, and the mean depth of the individuals ranged from 11.7 to 40.6 (Table 1).

Table 1: Mean depth of each sample in the filtered dataset.

Sample Name	Average Depth
Spiterstulen (SP)	
SP 70535	18.39
SP 70536	16.47
SP 70537	15.41
SP 70538	15.74
SP 70540	16.59
SP 70541	16.67
SP 70542	16.42
SP 70543	17.78
SP 70544	16.90
SP 70545	15.32
SP 154	18.34
SP 164	34.48
SP 1	24.03
SP 21	23.98
SP 2	21.36
SP 3	29.51
SP 4	32.75
SP 5	26.98
SP 6	12.43
SP 76	35.27
SP 7	22.62

Plech (PL)	
PL 1a	29.54
PL 2a	14.03
PL 3a	24.9
PL 4a	30.66
PL 5a	15.02
PL 6a	27.20
PL 7a	29.38
PL 8a	20.84
PL 9a	20.96
PL 10	22.95
PL 80936	27.22
PL S2	31.48
PL S4	30.59
PL S5	36.41
PL S6	38.32
PL S7	34.99
PL S8	40.61
Austria (AUS)	
AUS PEQ6	11.70
AUS PEQ9	12.17
AUS PER11	10.22
AUS PER8	18.76
AUS VLH2	14.67
AUS VLH5	22.88
AUS VLH6	21.58

2.1.2 Analysis of population structure

Genetic diversity and differentiation along the chromosomes were calculated with PopGenome package (Pfeifer et al. 2014) in the R environment v3.4.4 (R Core Team 2018). Specifically, I calculated pairwise nucleotide fixation index (F_{ST}), nucleotide diversity between (d_{xy}) and within population (π) in 10kb non-overlapping windows for each chromosome with functions F_ST.stats, diversity.stats.between and diversity.stats.within, respectively (Hudson and Wayne 1992; Wakeley 1996). In order to avoid biased F_{ST} estimates (Cruickshank and Hahn 2014), the windows that had F_{ST} values above 0.8, dxy and π values below 0.001 in at least one population dyad, were removed from the analysis. Tajima's D was calculated with the function neutrality.stats of PopGenome. The linkage disequilibrium for the field collected SP and PL individuals was calculated along the genome with the default functions of PopLDdecay (Zhang et al. 2018) and the values were plotted in R.

Principal component analysis (PCA) of the genomic data was conducted with adegenet package (Jombart 2008) using a dataset including only every 300th SNP to reduce computational load. This reduced dataset of 233,075 SNPs was also used to calculate SNP based F_{IS} for each population with Hierfstat (Goudet 2005) package function basic.stats (Alexander et al. 2009; Goudet and Jombart 2015). The F_{IS} value of each gene was estimated based on the median F_{IS} value of its SNPs, for SP and PL.

For the admixture analysis (Alexander et al. 2009) bed files were generated with PLINK (Purcell et al. 2007), which were then analyzed for clusters K=1 till K=5, with 10 iterations of cross-validation each. The clusters were normalized across runs using CLUMPAK (Kopelman et al. 2015) and subsequently they were plotted in R.

2.1.3 Demography simulations

To study the demographic history of these populations, my collaborator Dr. T. Hämälä and I conducted site frequency spectra (SFS) based coalescent simulations with fastsimcoal2 v2.6.0.3 (Excoffier et al. 2013). Folded 3D SFS, comprising of SP, PL and AUS individuals, was estimated from 4-fold sites with ANGSD v0.917 (Korneliussen et al. 2014), using the same filtering steps as with variant calls. First models with all possible divergence orders were considered, and subsequently compared models with five different migration scenarios, guided by previous work on the SP and PL populations (Mattila et al. 2017; Hämälä and Savolainen 2018): no migration, current migration between PL and AUS, historic migration between PL and AUS, and historic

migration between all three populations. Each model was repeated 50 times and the ones with the highest likelihoods were used for model selection was based on Akaike information criterion (AIC) scores. Confidence intervals were estimated by fitting the supported model to 100 nonparametric bootstrap SFS. We used these models to define effective populations sizes (N_e), divergence times (T) and migration rates (m). To evaluate how the estimated demography influences measures of positive selection, we used the N_e, T and m parameters in combination with recombination rate estimates derived from an *A. lyrata* linkage map (Hämälä et al. 2017) to generate 10,000 10 kb fragments with *ms* (Hudson 2002). These data were then used to define neutral expectations for analysis of positive selection.

2.1.4 Estimating the distribution of fitness effects

For analyzing the strength of selection, vcf files were first re-filtered for each population separately, as described in the section "data preparation". This was done to retain positions that only needed to be removed in one population. Sites with data for more than 80% of the individuals were randomly down sampled so that each position had the same number of alleles. Because the SP and PL populations differed in the number of individuals sampled, individuals in the SP population were further randomly down sampled at each position to give the same number of alleles in both populations.

With the help of my collaborators, Dr. K.A. Steige and Dr. E. Koch, we implemented a modified version of the program fit∂a∂i (Kim et al. 2017) to estimate the distribution of fitness effects. This extension to the ∂a∂i program (Gutenkunst et al. 2009), which infers demographic history as well as selection based on genomic data, allows us to specify the demographic model when inferring selection. Because we only fit the DFE to variation in PL and SP, we first fit a simplified demographic model for these populations only using ∂a∂i (Figure 1a-b). The simplified demographic model was inferred by maximizing the composite likelihood of the folded SFS at 4-fold degenerate sites in PL and SP using the "L-BFGS-B" method and basinhopping algorithm implemented in scipy. These models provided a good fit of the predicted neutral SFS to the data (Figure 1c-d). They were compatible with the complex 3-population model, but assumed a larger ancestral population size to account for migration from AUS. This model also indicated that the increase in population size following the last bottleneck may have been underestimated in SP. The estimated 4-fold population scaled mutation rate theta reached 24,000 for PL. It was multiplied by 2.76 to get the 0-fold mutation rate, i.e. the non-synonymous mutation rate for PL. In all instances,

the theta used for the SP population had to be constrained to $theta_{PL}$*0.74, to account for the difference in number of retained sites after all filters differed between the populations.

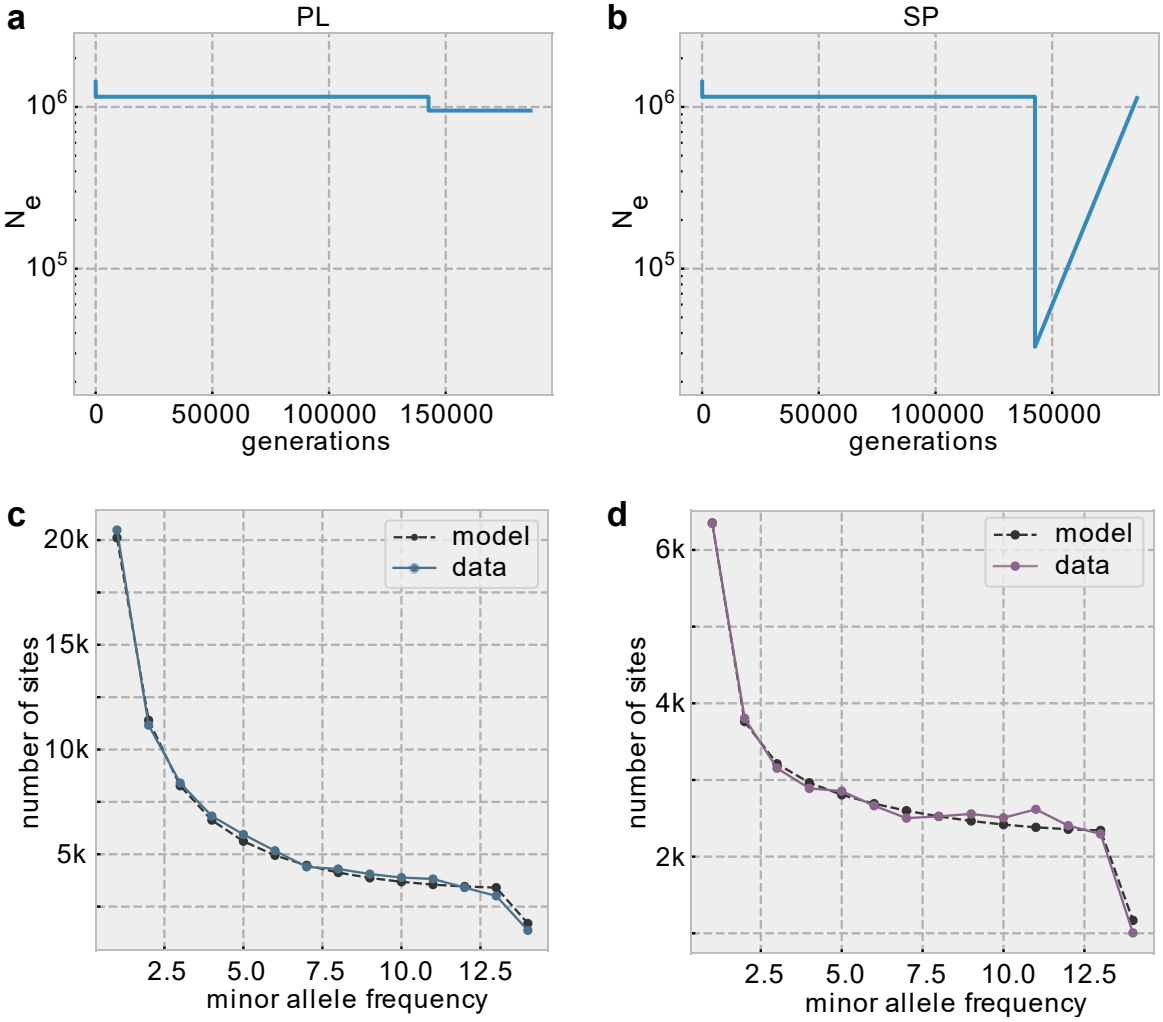


Figure 1: Simplified and re-fit demographic model used to assess the gamma distribution of the distribution of fitness effects fit∂a∂I for (a) PL and (b) SP. The simplified model in PL is a step wise population change and in PL shows a strong bottleneck following population expansion. (c) and (d) show the site frequency spectrum for synonymous sites. The solid line shows the data, the dashed line the estimate based on the model.

The 0-fold SFS and the demographic model were then used to fit the DFE by estimating the shape and scale parameter of a gamma distribution of selection coefficients. We used both a multinomial model (without using the population scaled mutation rate, theta) and a Poisson model (including theta) to estimate the DFE. The primary difference between these models is that the multinomial model only fits the DFE for variation sufficiently weakly selected to be observed in the sample.

The reason for this is that the multinomial model only fits the proportions of alleles observed at different frequencies (the "shape" of the SFS) and does not consider the decrease in per-site reduction in variation compared to 4-fold sites. Strongly deleterious variation will largely be absent from our moderate sample size and therefore does not affect the shape of the SFS. In practice we found that the gamma DFE fit using the multinomial method converged on a point mass at a single selection coefficient. After the DFE for observed variation was fit using the multinomial approach, we also estimated the fraction of strongly deleterious mutations by examining the ratio of the observed SFS to that under the multinomial DFE using the theta calculated for 0-fold sites. This ratio gives an estimate of the fraction of mutations that are sufficiently weakly selected to be observed in the sample.

Although fit∂a∂i includes a function for finding the maximum likelihood values for DFE parameters, we had to implement a different function because we were fitting the parameters to the composite likelihood of the SFS in both populations. We calculated the likelihood using corresponding ∂a∂i functions and determined maximum likelihood parameters using Sequential Least Squares Programming as implemented in scipy.

The DFE describes the distribution of fitness effects of new mutations arising in a population, and as such is independent of the demographic history. It was therefore assumed to be the same in both populations. To predict properties of genetic variation in the two populations, we calculated the distribution of selection coefficients for variants in each count of the SFS. For this, we first calculated the expected SFS for each selection coefficient using ∂a∂i functions. Then, we 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 under the Poisson model. Finally, we inferred the distribution of selection coefficients in each count of the SFS by applying Bayes' rule.

2.1.5 Genomic burden estimates

We estimated the difference in burden between the populations by first calculating the expected joint SFS for PL and SP under the selection coefficient fit by the multinomial model, using the theta value for PL as power for calling SNPs was higher in this population. For each entry in the SFS we then calculated the difference in the expected count between PL and SP, weighted by their frequency in the sample to account for their probability of being present in an individual genome. Crucially, we also counted alleles that were fixed in one population but not the other. The

cumulative difference over all frequencies gives the overall expected difference in the derived allele burden.

Additionally, I used the number of derived non-synonymous mutations per individual to quantify the population's genomic burden (Simons and Sella 2016). We used SNPeff (Cingolani et al. 2012) to annotate synonymous and non-synonymous sites, as well as sites with different level of high putative impact on the protein, whose ancestral state inference was done comparing to *A.thaliana* and *C.rubella*. For this, Dr. KA. Steige generated pairwise alignments with LASTZ v1.04 (Harris 2007), (astz 32 --ambiguous=n –notransition –step=25 –nogapped) and combined them in a multiple alignment file for the four species, which was generated using mugsy v1r2.3 (Angiuoli and Salzberg 2011) with default settings. To infer ancestral state genomic information of each SNP of the *A. lyrata* was combined with base information of *A. thaliana* and *C. rubella* using custom perl scripts. Ancestral state of the SNPs was inferred using the program est-sfs v2.03 (Keightley and Jackson 2018) using the Jukes-Cantor model. At probabilities between 1 – 0.9 the major allele was assumed to be ancestral and at probabilities between 0.1 and 0 the minor allele was assumed to be ancestral. 4-fold and 0-fold positions were identified using the NewAnnotateRef.py script (https://github.com/fabbyrob/science/blob/master/pileup_analyzers/NewAnnotateRef.py) and uSFS were extracted for these positions using custom R scripts.

Then I counted the respective numbers of synonymous and non-synonymous sites per individual, with weight of +1 for each instance of homozygous state of the derived allele and as +0.5 for the heterozygous sites (See Appendix). I divided the counts by the total number of genotyped sites, in order to correct for differences in genome mapping between the individuals. The genomic load of each population was calculated as the mean of the weighted number of non-synonymous sites of individual samples. The synonymous sites were used to confirm the robustness of the analysis, as they are expected to not differ among the populations. The confidence intervals for each population, were estimated by bootstrapping with replacement of 1Mbp windows to simulate each time a whole genome (207 1Mb regions). Significance of the mean load difference between SP and PL was estimated following Simon and Sella (2016). Briefly, I bootstrapped 16 1Mbp-windows of the genome with replacement and selected two random samples from the union of the two populations to create two groups of size equal to the original populations. This generated a random distribution of expected variance in the mean derived mutation counts. I used the quantile

of this distribution to determine the *p* value. Note that I verified that these estimates of per-individual burdens do not change if the regions carrying sweep signatures are removed from the analysis.

2.1.6 Scan for selective sweeps

Areas influenced by selective sweeps were inferred by estimating composite likelihood ratios with SweeD v4.0 (Pavlidis et al. 2013). The analysis was done in 2 kb grid sizes for the SP and PL samples. As a bottleneck can easily bias CLR estimates (Jensen et al. 2007), I used data simulated under the best supported demographic model to define limits to neutral variation among the observed estimates. Estimates exceeding the 99th percentile of neutral CLR values were considered putatively adaptive. I combined significant grid points within 10 kb regions to create outlier windows. Grid points that had no other outliers within 10 kb distance were removed from the analysis. Next, I examined the sweep regions in combination with regions showing elevated differentiation to find areas targeted by strong selection after the populations diverged. As with CLR, windows with F_{ST} values above the 99th percentile of their distribution were considered outliers. Genes from the regions showing higher than neutral differentiation with both CLR and F_{ST} were extracted. Gene Ontology enrichment analysis was performed in R with the topGO package (Subramanian et al. 2005; Alexa and Rahnenfuhrer 2016). Significance of the enrichment was evaluated with Fisher's exact test. Significance threshold was evaluated by permutating the sample's population identity 1,000 times.

2.1.7 Identification of S alleles

Our collaborators, X. Vekenmans, V. Castric and M. Genete, genotyped the individuals at the self-incompatibility locus (S-locus) with a genotyping pipeline (Genete et al. 2019) using raw Illumina reads from each individual and a database of all available sequences of *SRK* (the self-incompatibility gene expressed in the pistil) from *A. lyrata*, *A. halleri* and *Capsella grandiflora* (source: GenBank and unpublished sequences). Briefly, this pipeline uses Bowtie2 to align raw reads against each reference sequence from the database and produces summary statistics with Samtools (v1.4) allowing to identify alleles at the S-locus (S-alleles). Coverage statistics allow to reliably identify homozygote versus heterozygote individuals at the S-locus. Genotype data was used to compute population genetic statistics using Fstat (Goudet 2005): number of alleles per sample, sample allelic richness (a standardized estimate of the number of alleles taking into account differences in sample sizes among populations, computed after the rarefaction method

described in El Mousadik and Petit (1996), gene diversity (expected heterozygosity under Hardy-Weinberg hypotheses), and F_{ST} (unbiased estimate of the among population fixation index).

2.1.8 Identification of gene functional groups

F_{ST}, dxy and π were estimated for all genes according to the *A. lyrata* gene annotation v1.0.37 with PopGenome and as described above for the genomic windows. Genes that had functions involved in light, cold, flowering, plant development and dormancy were determined based on the gene ontology of the sister species *A. thaliana.* To explore whether the aforementioned groups of genes had genetic differentiation estimates that were significantly different from the genome-wide background, I performed a non-parametric, two sample Kolmogorov Smirnov test (Marsaglia et al. 2003) between the gene group of interest and the rest of the genomic genes identified in *A. thaliana* and belong in a GO group (ks.test function in R).

2.1.9 Comparative analysis of growth rate and biomass accumulation in a common garden experiment

I propagated clonally 10 genotypes from SP and 10 from PL to study growth in a common garden setting. The experiment was initiated in September 2017 and ended August 2018 and took place at the garden of the University of Cologne. During winter, the presence of vegetative plant material was scored during periods of prolonged frost, as well as the two subsequent months. Throughout the growing season (March to August), I scored monthly diameter size, in millimeters, as a proxy for vegetative growth. At the end of the growing season, I harvested the above ground material to estimate the dry to fresh weight ratio of the plants as a proxy for the plants' biomass. The phenotypic data are provided as Appendix Table 1. Differences between the two populations were tested in R with linear mixed models using the library lme4 (Bates et al. 2015). The model included population and month of the measurement taken as fixed effects. The genotype and replicate number were included as random effects in order to correct for pseudoreplication resulting from sampling the same individuals multiple times throughout the experiment. Significance levels were estimated with a type-II likelihood-ratio-test using the function Anova, from car library (Fox and Weisberg 2019).

We estimated the per individual heterozygosity level (inbreeding coefficient F) for the derived sites, using vcftools. The phenotypes of the clonal plants were averaged per genotype and correlated to F and genomic load using spearman's rank correlation (ρ).

2.1.10 Seedling growth of between and within population crosses in controlled conditions.

Seeds from random within population (within SP and within PL) and between population crosses were produced in controlled environment (see for more details section 2.2.1). The seeds were stratified on wet filter paper in the dark and +4^{o}C, for 4 and 7 days the SP and 7 PL seeds, respectively. The hybrid seeds were stratified as long as their mother population would. Subsequently, the seeds were allowed to germinate in 20^{o}C, 16 hrs of daylight. Each seedling was transferred to pots as soon as the cotyledons were fully open and were transferred to a walk-in growth chamber (Dixel, Germany) set at +12^{o}C and 16 hrs of daylength. 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 (440nm), red (660nm) and white light with total measured intensity of 224 +/- 10 μmol * sec^{-1} *m^{-2} . During both experiments a light pulse of far red (750nm) was implemented for 10mins at the end of the day.

For each plant, the days till they have 6, 10 and 25 leaves were estimated. When they had 25 leaves, the rosette diameter was also measured. The data were analysed with the R library lme4, with population (SP, PL or hybrid) as the fixed effect and individuals identification number as random effect, to correct for pseudoreplication. Significance levels were estimated with a type-II likelihood-ratio-test using the function Anova, from car library (Fox and Weisberg 2019).

2.2 Chapter 2: The adaptive potential of gene expression variation

2.2.1 Preparation of Inter-population Crossings and plant material generation

I crossed six plants from SP with six plants from PL to obtain inter-population hybrids. The parental individuals were propagated clonally and vernalised for nine weeks in 4^{o}C and 12 hours daylength, before being transferred to greenhouse conditions till they flowered. Crosses were performed so that everyone from one population was crossed with two individuals from the other population. All crosses were reciprocal, which means that each individual plant was used as both pollen receiver and donor, to be able to control for maternal effects. Since the exact same plants were used for the reciprocal crosses, the resulting seeds are considered full siblings and they are termed "family". Each family's members are half-siblings with the members of at least another family with one parent common. Details about the resulting crossing scheme, families, and their kinship are given in Table 2.

For each reciprocal cross, I germinated between 4 and 10 plants as described in section 2.1.10. Once the cotyledons were fully open, the plants were transferred to pots and placed into a walk-in growth chamber (Dixel, Germany) set at 12^{o}C, 16 hrs of daylength. The light conditions were the same as the HL treatment described in section 2.1.10. We sampled the individuals when the 10th leaf was visible on the rosette. All samplings were performed at 4 hours Zeitgeber Time. Two families (annotated as fam11 and fam12 in Table 2) did not germinate well, and the remaining individuals died before producing the 10th leaf and thus these families had to be removed from the subsequent analysis.

Table 2: Information about the full sibling families used in the analysis of Chapter 2 and 3. For each full sibling family, the SP parent sequence ID, the PL parent sequence ID and the number of full-sibling samples sequenced are provided.

Family	**SP parent**	**PL parent**	**Number of family members**
Fam01	70539	10a	11
Fam02	70539	80936	18
Fam03	70535	80936	13
Fam04	70535	5a	12
Fam05	70537	6a	9
Fam06	70536	2a	8

Fam07	70544	6a	20
Fam08	70544	5a	14
Fam09	70545	2a	6
Fam10	70545	10a	6
Fam11	7056	11a	0
Fam12	70537	11a	0
Fam15	70539	2a	4
Fam16	70545	80936	12

2.2.2 RNA extractions and data preparation

We extracted RNA and DNA using the AllPrep DNA/RNA Mini Kit (QIAGEN, USA). We used custom primers to amplify DNA around regions that the restriction cut enzyme PvuII could not digest at a specific position within one of the two populations. The forward primer sequence was 5'-GCACAAGACTGCTGTAACGC-3' and the reverse primer sequence was 5'-AATGGCCTCCCGTATTTGCA-3'. Then the amplified area was digested and visualised with a 0.8% agarose gel. A sample from SP or PL would have unique sized bands, while a hybrid between the populations would show all the possible bands. Thus, we confirmed that all the samples were crosses between the two populations, by heterozygosity at the site. The amplification and digestion protocols can be found in the Appendix. RNA quality and quantity were checked with Qubit 4 Fluorometer (Thermofisher Scientific, Germany), 2100 Bioanalyzer (Agilent, USA) and a 0.8% agarose gel. Whole transcriptome sequencing for 133 plants was performed in 4 subsequent batches at the former Cologne Center for Genomics (now West Germany Center). We sequenced 75bp long RNA, paired end with average depth 80X. Sequence quality was checked with FastQC. The first 3 bp and unpaired reads were discarded with Trimmomatic v0.36 (Bolger et al. 2014). Transcriptome mapping against the *A. lyrata* v1 (Hu et al. 2011) reference genome was performed with STAR v2.5.3 (Dobin et al. 2013) using standards settings plus a cut off for maximum indel length of 10kbp. The ribosomal RNA was identified and removed based on the *A. lyrata* genome annotation v37 and use of bedtools subtract (Quinlan and Hall 2010).

All the reciprocal crosses had been performed in the greenhouse, which is not pollinator free. To make sure that all the parents were the ones we intended and not the result of natural pollination, we assigned parents to each hybrid sample. The SNPs of the hybrids' transcriptome as well as of

the parental genome sequences were called using samtools v1.5 and bcftools v1.5 (Li et al. 2009). The sites were filtered based on their quality (minimum 60), depth (minimum 10) and minor allele frequency (minimum 0.01), using vcftools v0.1.15 (Danecek et al. 2011). All indels were removed and no missing sites were allowed. I used custom python scripts (see Appendix) to calculate the genetic distance between the hybrids and the parental genomes. The genetic distance between two lines was estimated as:

((Number of pairwise differences / Number of comparisons) * Number of variant sites) / Number of non-variant sites

Furthermore, the relatedness between all individuals was estimated with vcftools. PCA of the samples was performed with R Statistical Language (R Core Team 2018) and specifically using the adegenet package (Jombart 2008). At the end of the analysis 22 samples were reassigned into new families.

I calculated the gene counts per individual using htseq-count (Anders et al. 2015) and estimated the number of fragments per kilobase per million reads (FPKM) in R. Based on the total FPKM and log2 distribution of the gene counts, I removed two individuals from the analysis as they showed indications of lower transcriptome quality than the rest.

2.2.3 Partitioning of gene expression variance to its components

I partitioned the expression variance of each gene to its genetic and environmental components. I filtered the resulting gene list based on their expression level. Genes with total reads less than 130 across all individuals, genes with median counts more than 180,000, as well as genes which 80% of the individuals had FPKM less than 0.31 were removed from the analysis. Then, for each of the remaining genes, the animal model (Lynch and Walsh 1998; Wilson et al. 2010) was implemented in R v3.6 statistical language (R Core Team 2018), via the package MCMCglmm (Hadfield 2010). MCMCglmm takes a Bayesian approach to fitting general linear model. The fitted multivariate models included the population of the mother as a fixed effect, to account for the different population origin of the parents. Random effects included the additive and dominance matrix as well as the identity of the mother plant to correct for maternal effects. The model, thus followed the following formula:

log10(transcript_counts + 1) ~ Population of Mother Plant + (1|Additive Matrix) + (1|Dominance Matrix) + (1|Mother Plant).

The prior distribution for both fixed and random effects was an inverse gamma (0.001; 0.001). The counts were transformed to fit a gaussian distribution. Each model was run twice, for 2.2 million iterations, from which the first 200,000 iterations were discarded and every 2,000th iteration was sampled (see Appendix). To correct for poor model fit, I removed from the dataset genes that at least one of random or fixed effects had effective size less than 500. Additionally, I removed genes from the analysis that the Markov chains showed low quality of convergence based on Gelman Rubin criterion values less than 1.1 (Gelman and Rubin 1992). In the end, we used the expression counts of 17,657 genes for the rest of the analysis.

For each gene, the additive (V_A), maternal (V_M), dominance (V_D) and residual (V_R) variance was extracted from the model. The sum of the four variances comprises the total phenotypic variance (V_P), while the sum of the V_A and V_D is the genetic variance (V_G). All variances are presented as proportion of V_P. Note that the narrow sense heritability corresponds to the estimated proportion of V_A out of total V_P and the broad sense heritability to the estimated V_G / V_P. The above analysis was implemented for the growth rate of each plant, which was defined as the number of days between potting and sampling.

Furthermore, for each gene we run an additional linear mixed model, excluding the dominance matrix from the random effects, and compared it to the original model. The best fitting model selection was done based on the Bayesian Information Criterion (BIC). We identified 26 genes that the inclusion of the dominance matrix did not improve the inference of the model.

I tested for potential clustering of the data and therefore non-independence of the phenotypes by firstly estimating the pairwise correlation of all the genes 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 (Müllner 2013). Potential clusters were identified by examining the within groups sum of squares for clusters between 1 and 300. The gene tree then was cut with cutree function in 23, 50, 100 and 200 clusters. The impact of clustering on the genes was tested by correlating the groups' median dominance, additive, and transcript length with the number of genes within each cluster.

Gene ontology enrichment analysis for decreasing values of V_A and V_D were performed as described above in section 2.1.6. For each enrichment performed, the gene universe was the list of orthologues genes with *A.thaliana* present in the *A. lyrata* genome. The *p* value thresholds were estimated by permuting the gene identity 1,000 times.

2.2.4 Correlation of genome architecture, selection and population genetic parameters with dominance and additive genetic variance

I aimed to assess whether there is a relationship between the genetic variance components and transcription factor binding sites, gene properties and population genetic features of *A. lyrata* genes. Dr. Hannes Dittberner and I collaborated to use a database created by Dr. U. Göbel, to extract the counts of each transcription factor binding site in all genes of *A. lyrata*. We treated the count of each TF binding site per gene as a separate feature and additionally determined the total number of TF binding sites per gene, the number of exons, the transcript length, and the gene length. Furthermore, we determined the mean value of the following population genetic features for each gene: F_{ST}, Tajima's D, SNP density per kb, dxy and π (per population of origin) as described in section 2.1.6. Additionally, we categorized each gene based on whether it is located within a selective sweep area defined as described in section 2.1.6, and if the gene is differentially expressed between at least one pair of families. The differentially expressed genes (DEG) were estimated using the R library DESeq2 (Love et al. 2014) and testing for a family effect. The significant change in gene expression level was tested based on corrected *p* values for multiple testing.

We trained a random forest model using the ranger library (Wright and Ziegler 2017) with V_A or V_D as the response variable and all other variables mentioned above as predictors. We set the number of trees of the forest to 500, the number of variables to possibly split at each node (mtry) to 200 and the impurity mode to impurity_corrected. From the trained random forest, we extracted the relative importance (i.e. explanatory power) of each predictor variable as well as the out-of-bag prediction error, which is informative about the predictive power of the model. Finally, we estimated a *p* value for each predictor variable using the "Altmann" method, using 60 permutations.

I explored further the impact of the genome architecture on V_A and V_D by, firstly correlating the V_A, and V_D values with the transcript and total length by using spearman's ρ via the R function cor.test. Then, the combined effect of transcript length and median gene expression, correcting for

the level of broad sense heritability, was estimated by a generalized linear model in R, with function glm and poisson distribution of the residual variance. Moreover, the difference between gene length for significantly enriched Gene Ontology categories was tested with lme4, having the variance type as fixed effect and the Gene Ontology category as random effect, to correct for pseudoreplication.

To investigate the level of constraint of the groups of genes with different levels of genetic variance in the PL and SP populations, I grouped the genes in four categories based on the different proportions of the additive, dominance, genetic and a combination of residual and maternal variance as a group with little genetic variance. Additionally, a random set of 3,500 genes was used as a control group. Dr Kim A. Steige and I collaborated on assessing the distribution of fitness effects (DFE) using a combination of general python functions, δaδi and fitδaδi, as described in section 2.1.4. The confidence intervals for each group were based on 200 bootstrap replicates for each group. Significant differences between the groups were calculated by pairwise comparisons within each bin of the DFE (0-1, 1-10 and 1-inf) by using the union of the bootstrap values to calculate what proportion of the mutations belong to the same group, multiplied by 2, to account for the fact that the test is one-sided (Keightley and Eyre-Walker 2007).

2.3.1 Genome wide association study for identifying *trans* effects on gene expression

I used the SNP calls of the 131 available hybrid transcriptomes to identify *trans* effects using the Genome Wide Association study (GWAS) method (Kang et al. 2010; Korte and Farlow 2013), with the corresponding R packages downloaded from https://github.com/arthurkorte/GWAS. I filtered the SNP calls with vcftools to maintain only biallelic sites, with no missing information in any sample, minimum depth of 10 reads per sample, minimum SNP calling quality of 30 and minor allele frequency of 0.05. The remaining 127,585 SNPs were used to estimate the kinship matrix with the function emma.kinship. The log transformed counts of each of the 17,657 genes were successively used as the phenotypic input for each GWAS. I retained the top hits for each GWAS using the Bonferonni cut off (p = 3.918956e-07) to correct for multiple testing. Moreover, we checked the coefficient of variance (cv) of gene expression between the groups of samples with specific genotype on the site. If the cv was smaller than the 25th percentile of their distribution, it was removed from the further analysis.

2.3.2 Estimation of *cis* regulatory variation within the hybrids by allele specific expression

Dr F. He and I used 105 out of 131 samples that were available at the time of the analysis to estimate the allele specific expression in the between SP and PL crosses. All transcriptome sequences and genome sequences were mapped as described in sections 2.2.1 and 2.1.1, respectively. We used the default pipeline of GATK variant caller v4.1.1.0 (McKenna et al. 2010) to call SNPs in the whole genomes of the SP and PL parents. Those SNPs were contrasted to the transcriptome SNPs called by freebayes v1.3.2-38-g71a3e1ce (Garrison and Marth 2012) to assign population origin in the hybrid samples. For each gene, the median depth of the SP and the median depth of the PL allele were estimated from the vcf files. The deviation from the 1:1 ratio for each gene per sample was tested with a chi square test. The *p* values were corrected for multiple testing using the Bonferroni correction.

We used the whole genome sequence of two extra sample from fam03 and fam08 to see for possible mapping errors. We mapped the genomes using bwa mem, called SNPs with freebayes and identified population origin of the alleles as described above. Subsequently, we tested for the deviance of the allelic depth from the 1:1 ratio and we removed from the transcriptome dataset of

fam03 and fam08 all the alleles that deviated significantly from the ratio as they showed signs of mapping bias towards the allele of one of the two populations.

2.3.3 Allele specific ratio analysis

I calculated the ratio of allelic depth for all genes using the formula :

$$\text{Ratio} = \log10 ((\text{SP depth} +1) / (\text{PL depth} + 1))$$

Defining the ratio that way allowed to know that all the genes with ratio above 0 had overexpressed the SP allele and the genes with ratio below 0 had overexpressed the PL allele. In the case that the ratio was equal to zero, then the two alleles are expressed in the same quantity. Also, adding 1 to both allelic depths allowed me to define the ratio for those genes that one of the two alleles is not expressed. Genes with total depth less than 20 were removed from the analysis.

Allele specific expression (ASE) is defined for the genes with significant deviance of the allelic depth from the 1:1 ratio. For each family, I defined genes as ASE for the family, if they are ASE in at least 30% of the individuals. An ASE gene was polymorphic for one population, if it fullfilled the following two criteria. Firstly, the allele from that population should have been overexpressed in the full sibling families with common parent the plant from that population. Secondly, the population specific allele should be overexpressed within the cross in more than two instances of half siblings comparison.

We used the list of ASE and polymorphic genes to perform gene ontology enrichment analysis, as described in Sections 2.1. Moreover, we used the population genetic parameters, as well as V_A and V_D values estimated previously, to explore for signals of genetic differentiation for the ASE genes. All the above analysis was performed with R statistical language.

3. Results

3.1 Chapter 1: Genetic Diversity and adaptive evolution in a range-edge population

3.1.1 Demographic history of three European *A. lyrata ssp. petraea* populations confirms a scenario of range expansion

I analyzed whole genome sequence data for 46 *Arabidopsis lyrata* individuals, of which, 22 were collected in a range edge population in Norway (Spiterstulen, SP), and 17 and 7 individuals from two core populations in Germany (Plech, PL) and Austria (a scattered sample, AUS; Figure 2a), respectively. A principal component analysis (PCA) confirmed that the sample was partitioned in three geographically and genetically distinct populations. The first principal component (PC) explained 24.95% of the variance, separating the Northern site from the two Central European sites (PL and AUS). The second PC (6.82%) differentiated the AUS and PL sites. AUS individuals were more scattered than SP and PL individuals, presumably because AUS individuals were collected over a broader area (Figure 2b). Admixture analysis showed that the samples formed three populations, without indication of admixture between the populations. The samples were well described with K=2 clusters (cross-validation error, cv = 0.397). The SP individuals formed a unique cluster, while PL-AUS individuals grouped together in one cluster. The second most probable scenario (cv = 0.419) was K=3, with each population forming its own cluster (Figure 2c). I further estimated the F_{ST} across 10 kb non-overlapping windows along the genome. Mean F_{ST} was 0.231 (median of 0.232) and 0.234 (median of 0.236) for SP vs. PL or AUS, respectively. Between PL and AUS, differentiation was much lower, with a mean F_{ST} value of 0.079 (median of 0.047). Thus, most of the genetic differentiation resides between Northern and Central European populations and not between PL and AUS. Average number of nucleotide differences between pairs of individuals from distinct sites (d_{xy}) confirmed the pattern of inter-population differentiation (Table 2). Within populations, nucleotide diversity was estimated as the average number of pairwise differences per sites (π) across the same non-overlapping 10 kb windows. Mean nucleotide diversity of the genomic windows was π=0.0081, π=0.0067 and π = 0.0055 for PL, AUS and SP, respectively (Table 3).

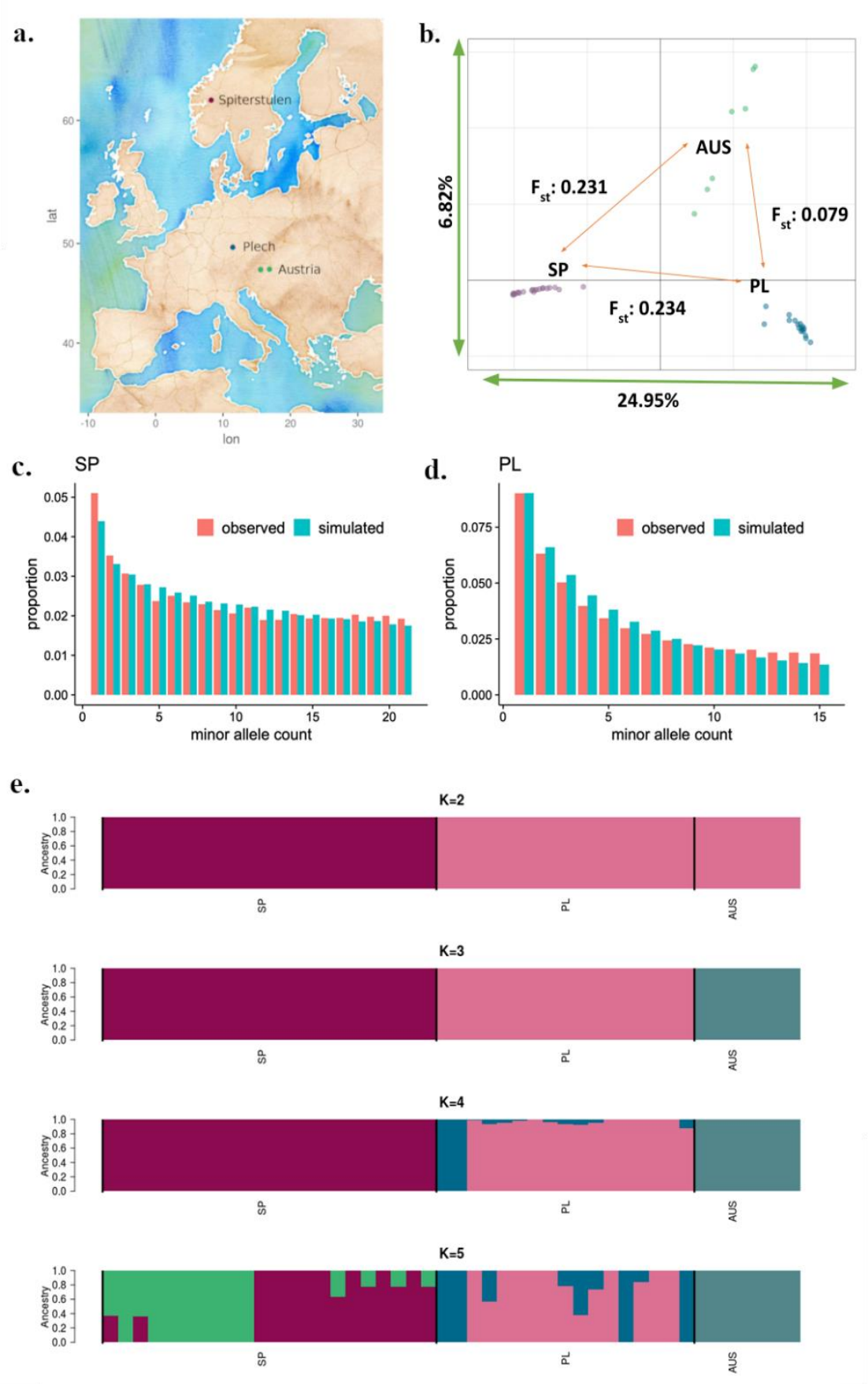

a.
Spiterstulen
Plech
Austria
lat
lon
b.
AUS
SP
PL
F_{st}: 0.231
F_{st}: 0.079
F_{st}: 0.234
6.82%
24.95%
c.
SP
observed
simulated
proportion
minor allele count
d.
PL
observed
simulated
proportion
minor allele count
e.
K=2
K=3
K=4
K=5
Ancestry
SP
PL
AUS

Figure 2: Population differentiation of 3 *A. lyrata* populations **a.** Geographical distribution of the Spiterstulen (SP), Plech (PL) and Austrian (AUS) populations. **b**. Principal Components analysis of SP, PL and AUS. The first Principal Component (PC) explains 24.95% of the sample variation and the second PC explains 6.82%. Within the PCA plot the F_{ST} values between all the population pairs are given. **c-d.** Observed, and estimated site frequency spectra of SP and PL used for the fastsimcoal analysis. **e.** Admixture analysis results for all samples. From top to bottom, the clustering in 2, 3, 4, or 5 clusters is shown. The samples are arranged in the same order in all five panels, with SP samples shown first, then PL and lastly AUS. According to the cross-validation error the most probable clustering is the K=2, and second best the K=3.

PCA, Fst and STRUCTURE provide measures genetic differentiation between individuals and populations. Genetic differentiation, in turn, is a result of the time since divergence, the intensity of gene flow, and the size of the population. Two populations could have split a long time ago, and nevertheless remain genetically similar if their population size is large and/or if there is gene flow. Conversely, populations could be genetically differentiated if they experienced a strong reduction in population size, even if they split recently. To identify the most likely demographic history explaining the observed pattern of genetic differentiation between populations, we used our dataset to model the demographic history of the three populations with *fastsimcoal2* (Excoffier et al. 2013).We tested models assuming different population split times. Model selection based on the Akaike information criterion (AIC) indicated that it was significantly more probable that the ancestral population of SP and PL (SP, PL) split from the AUS lineage first (Table 3; Figure 3b). Divergence between (SP, PL) and AUS (T) was estimated to have occurred approximately 292,210 generations ago (CI: 225,574 – 336,016). The split between SP and PL was estimated to have occurred more recently, approximately T = 74,042 generations ago (CI: 51,054 – 100,642). Demographic modelling further indicated that the most probable migration scenario entailed historical migration between all populations. The model indicated that gene-flow was higher between PL and AUS (PL to AUS, $4N_em$ =2.113, (CI: 1.668 – 6.771) and from AUS to PL $4N_em$= 0.039 (CI: 0.05 – 0.125)) than between SP and PL (SP to PL $4N_em$ = 0.038 (CI: 0.013 – 1.699), and PL to SP $4N_em$ = 0.162, (CI: 0.062 – 1.924).

Table 3: Summary of genome wide statistics calculated along 10kb windows in each of the three populations. The mean nucleotide diversity (π), Tajima's D, F_{IS}, and differentiation (pairwise F_{ST} above the diagonal and dxy values below the diagonal and in bold) between the 3 populations. Finally, the ratio of synonymous to non-synonymous derived alleles (Pn/Ps) are given. Whenever it is applicable, the 75th percentile of the distribution is given between parenthesis.

Population	**Watterson's θ**	**Π**	**Pn/Ps**	**Differentiation (**Fst , **dxy)**			**TajD**	**F_{IS}**
				SP	**PL**	**AUS**		
SP	0.00512 (0.00682)	0.0055 (0.007)	0.5028	-	0.231 (0.349)	0.234 (0.367)	1.23 (1.85)	-0.193 (0.130)
PL	0.0093 (0.0121)	0.0081 (0.011)	0.4826	**0.0094 (0.013)**	-	0.079 (0.129)	0.31 (0.70)	-0.04 (0.17)
AUS	0.00768 (0.0102)	0.0067 (0.0096)	0.4612	**0.0083 (0.012)**	**0.0079 (0.011)**	-	0.24 (0.59)	-

In addition, the estimated effective population sizes before and after divergence events indicated bottlenecks in all populations. The size estimate of the ancestral population reached N_e= 839,169 (CI: 823,959 – 877,924). The effective population size (N_e) of SP was reduced approximately 6-fold after it diverged from PL, from N_e= 206,610 (CI: 100,945 – 308,029) to N_e=35,479 (CI: 21,624 – 54,855), respectively before and after the split. In contrast, the PL population experienced a weaker initial bottleneck with N_e reduced by 40% after the split from SP: 127,100 (CI: 87,666 – 162,171). Both SP and PL also experienced more recent population size changes, with a slight increase in SP to a current N_e of 40,886 (23,081 – 47,713), approximately 4,421 (CI: 2,755 – 39,967) generations ago, and a very recent drop in PL to a current N_e of 11,190 (2,573 – 20,751), approximately 143 (CI: 4 – 361) generations ago. The population size in AUS decreased to 219,078 (CI: 148,664 – 249,105) after splitting from an ancestral population shared with PL. We note, however, that the AUS sample consists of individuals collected from three closely located sites, and thus might reflect diversity at a coarser grain than the SP and PL samples.

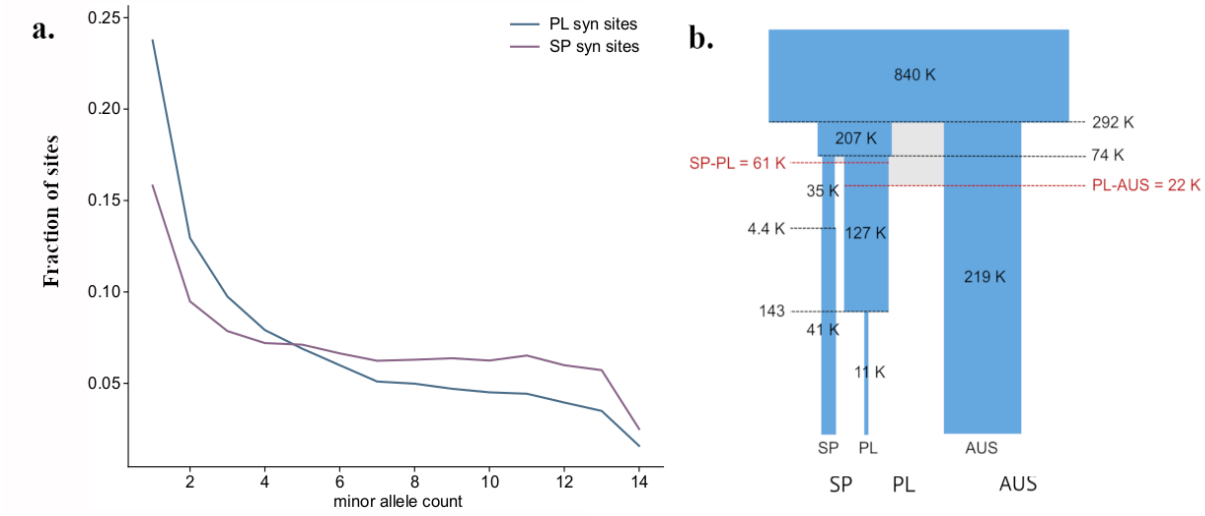


Figure 3: Demographic analysis of 3 Arabidopsis lyrata ssp. petraea populations. **a**. Folded site frequency spectrum of synonymous sites for PL and SP b. Schematic representation of the best-fit demography model. Shown within the boxes are the effective number of diploid individuals (Ne), divergence times (horizontal black lines) are indicated in thousands (k) of generations, with the exception of the final bottleneck in PL, which occurred only 143 years ago. The time since migration ended (horizontal red lines and numbers in red) is also indicated in thousands of individuals or generations. Width of the elements represents relative differences in Ne (in logarithmic scale), while time-differences in logarithmic-scale are represented by the height of the elements.

Table 4: Selection criteria for the population divergence models tested. The three possible trees were compared based on the Aikake Information Criterion (AIC). The best fitting model is shown in bold.

Split model	**Log Likelihood**	**AIC**	**Akaike weight**
(AUS, (PL, SP))	**-1438975.068**	**2877964**	**1**
(PL, (AUS, SP))	-1439477.708	2878969	~0
(SP, (AUS, PL))	-1439527.678	2878809	~0

Table 5: Selection criteria for the demographic model including migration. The three possible trees were compared based on the Aikake Information Criterion (AIC). The best fitting model is shown in bold.

Migration model	**Log likelihood**	**AIC**	**Akaike weight**
No migration	-1438975.068	2877964	~0
Current migration between PL and AUS	-1439112.311	2878247	~0
Historic migration between PL and AUS	-1438901.002	2877824	~0
Historic migration between all populations	**-1438455.012**	**2876936**	**1**

Table 6: Demographic parameters estimated for the complete data set are compared against two downsampled data sets. SP and PL datasets were reduced to 2/3 and 1/3 of their original sizes.

N = effective number of diploids (Ne)

T = time in number of generations

M = population migration rate 4Nem

NSP-H and NPL-H denote historic population sizes after population split, and TSP-H and TPL-H are times since population sizes changed to current estimates

TISOL are times since migration ended

Parameter	**3/3**	**2/3**	**1/3**
NSP	40,886	43,607	46,809
NSP-H	35,479	30,198	14,935
NPL	11,190	9,041	2,606
NPL-H	127,100	111,881	59,107
NAUS	219,078	187,331	181,429
NANC_SP-PL	206,610	233,398	197,094

NANC_PL-AUS	839,169	863,225	838,472
TSP-H	4,421	3,390	4,134
TPL-H	143	312	676
TISOL_PL-AUS	22,308	27,453	4,912
TISOL_SP-PL	61,148	59,642	20,622
TSP-PL	74,043	65,341	33,507
TPL-AUS	292,210	244,191	335,141

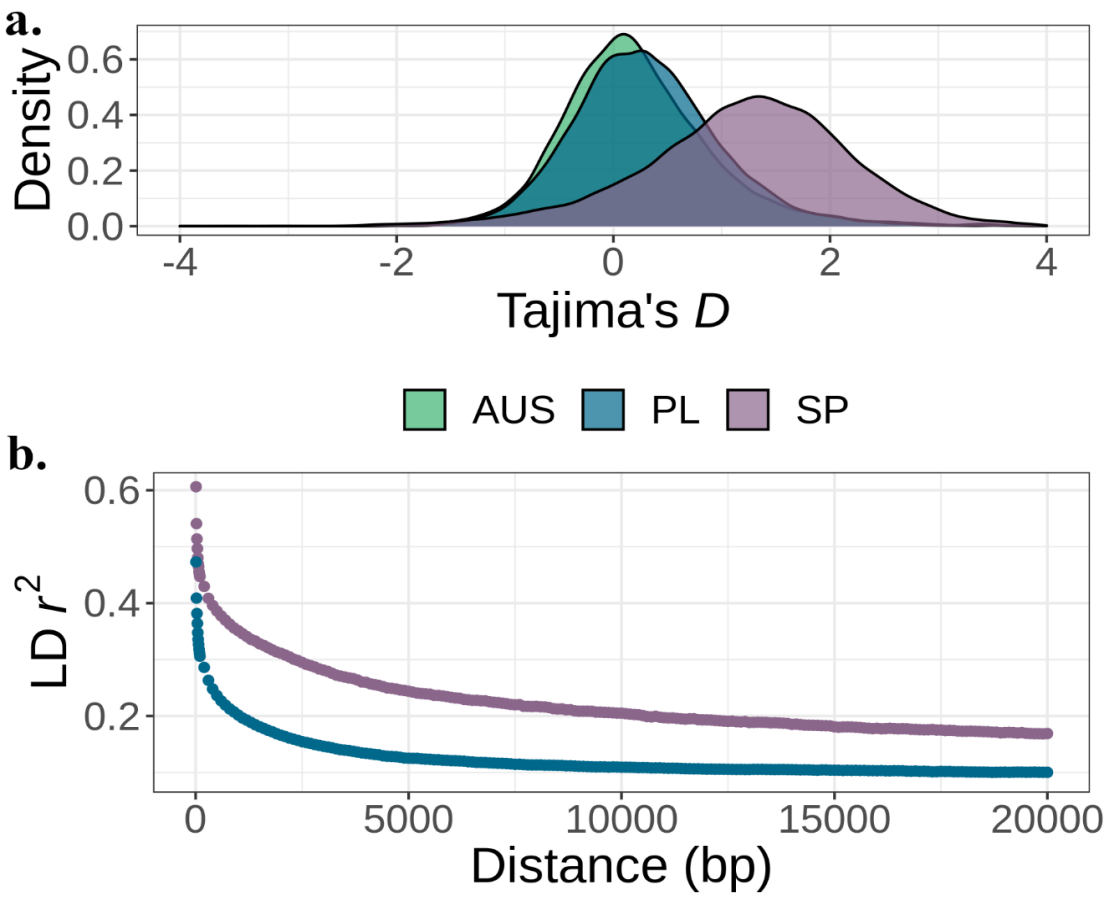


Figure 4: Evidence of a strong bottleneck along the SP genome. **a**. Tajima's D distribution for AUS, PL and SP calculated along the chromosomes in 10kb non-overlapping windows. **b**. Linkage disequilibrium decay in SP and PL given by SNP pairwise r^2 as a function of the distance between the SNPs. For comparison, both populations were down-sampled to 12 individuals each.

To test the robustness of bottleneck inference to sample size, we also conducted the demographic modeling with down-sampled data sets (1/3 and 2/3 of SP and PL sample sizes). Even though down-sampling changed the N_e estimates, the fold reductions in population size remained comparable and the bottleneck events are always supported (Table 6). Furthermore, we observed a good correspondence between the observed population-specific SFS (Figure 3a) and those simulated under the best-fit demography model, indicating that the model captures the evolutionary history of these populations reasonably well (Figure 3c-d).

I calculated Tajima's D values in 10kb windows for each population (Figure 4a). The distribution of Tajima's D values for SP was shifted towards positive values (mean = 1.230, median = 1.286), which was consistent with the inferred demographic history of a strong recent bottleneck in SP. Tajima's D values for PL and AUS were also mainly positive (mean = 0.313, median 0.265 for PL and mean =0.240, median =0.151 for AUS) but both were significantly lower than in SP (Kolmogorov-Smirnov, KS test p < 2.2e-16 in both cases). The two distributions also differed significantly (KS test p < 2.2e-16).

Additionally, analysis of linkage disequilibrium (LD) decay further confirmed the stronger bottleneck experienced by the SP population. LD decay was calculated on the subsample of 12 field collected SP individuals to ensure that native LD levels were analyzed (individuals obtained from crosses in the greenhouse were removed). LD was halved within 2.2 kb in SP, which is considerably slower than for an equally sized sample of PL individuals (LD halved within 0.5kb; Figure 4b).

Demographic modeling indicates that the large and fairly stable effective population sizes along with the persistence of gene flow for quite some time has resulted in a modest population differentiation between PL and AUS, despite their early split. By contrast, a more severe bottleneck and the lack of gene flow led to a stronger differentiation between SP and the other two populations. Demographic modeling therefore confirmed that SP is a range-edge population that can be contrasted to the more range-core populations PL and AUS (Muller et al. 2008; Pyhäjärvi et al. 2012; Mattila et al. 2017). This supports a scenario of colonization in Scandinavia with genetic material from Central European glacial refugia, a history that is common to several plant species (Clauss and Mitchell-Olds 2006; Pyhäjärvi et al. 2007; Ross-Ibarra et al. 2008; Ansell et al. 2010; Schmickl et al. 2010; Pyhäjärvi et al. 2012; Laenen et al. 2018).

3.1.2 The distribution of fitness effects

To infer the efficiency of negative selection, we estimated the distribution of fitness effects of new mutations in both range-edge (SP) and core (PL) populations, taking the demographic history into account (Williamson et al. 2005; Boyko et al. 2008). As the AUS population had a smaller sample size, as well as individuals taken from three different local sites, it was excluded. For SP and PL, we used a modified version of the software fit∂a∂i (Kim et al. 2017). We also fit a simplified demographic model that excluded AUS to the 4-fold SFS using ∂a∂i (Gutenkunst et al. 2009). This model was compatible with the complex model described above but assumed a larger population size in PL to account for migration from AUS. The demographic model showed a very good fit with (putatively neutral) SFS at 4-fold degenerate sites of both PL and SP (Figure 1a-d). Interestingly, the simplified inferred model in SP, which consisted of a bottleneck followed by exponential growth, corresponded very well to the scenario expected for range-core and -margin populations in an expanding species (Figure 1a-b). Distributions of fitness effects of new mutations (DFE) were estimated based on the non-synonymous (0-fold) folded SFS and taking the demographic history into account. We estimated the shape and scale parameters of a gamma-distributed DFE by fitting the demographic model and the 0-fold SFS of both populations (shape=0.213, scale=552.394, Figure 1c-d). Using the estimated gamma distribution of effects, we predicted for each frequency bin, the proportion of variants within four bins of selection coefficients (Figure 5a;c-d). The strength of s among segregating variants differed between the populations. Neutral and nearly-neutral mutations were found to contribute to a greater proportion of variation in the PL population compared to SP, whereas mutations with a stronger s were found to contribute more to variants segregating in SP (Figure 5a). Additionally, as a robustness check against our assumed non-synonymous mutation rate and gamma-distributed DFE, we used a multinomial model to predict the DFE fitting only the observed proportions in the folded 0-fold SFS. In this case, the best fit DFE was a point mass at $2*N_{anc}*s=1.2$, indicating that only slightly deleterious mutations were segregating in the two populations. Although the latter model ignores variation too deleterious to show up in the sample, we found that fixing the proportion of strongly deleterious new mutations to 44% provides a good fit to the observed 0-fold SFS in both populations, indicating that the Ne.s estimate of 1.2 was a reasonable approximation to the strength of selection against mildly deleterious non-synonymous variants (Figure 5b).

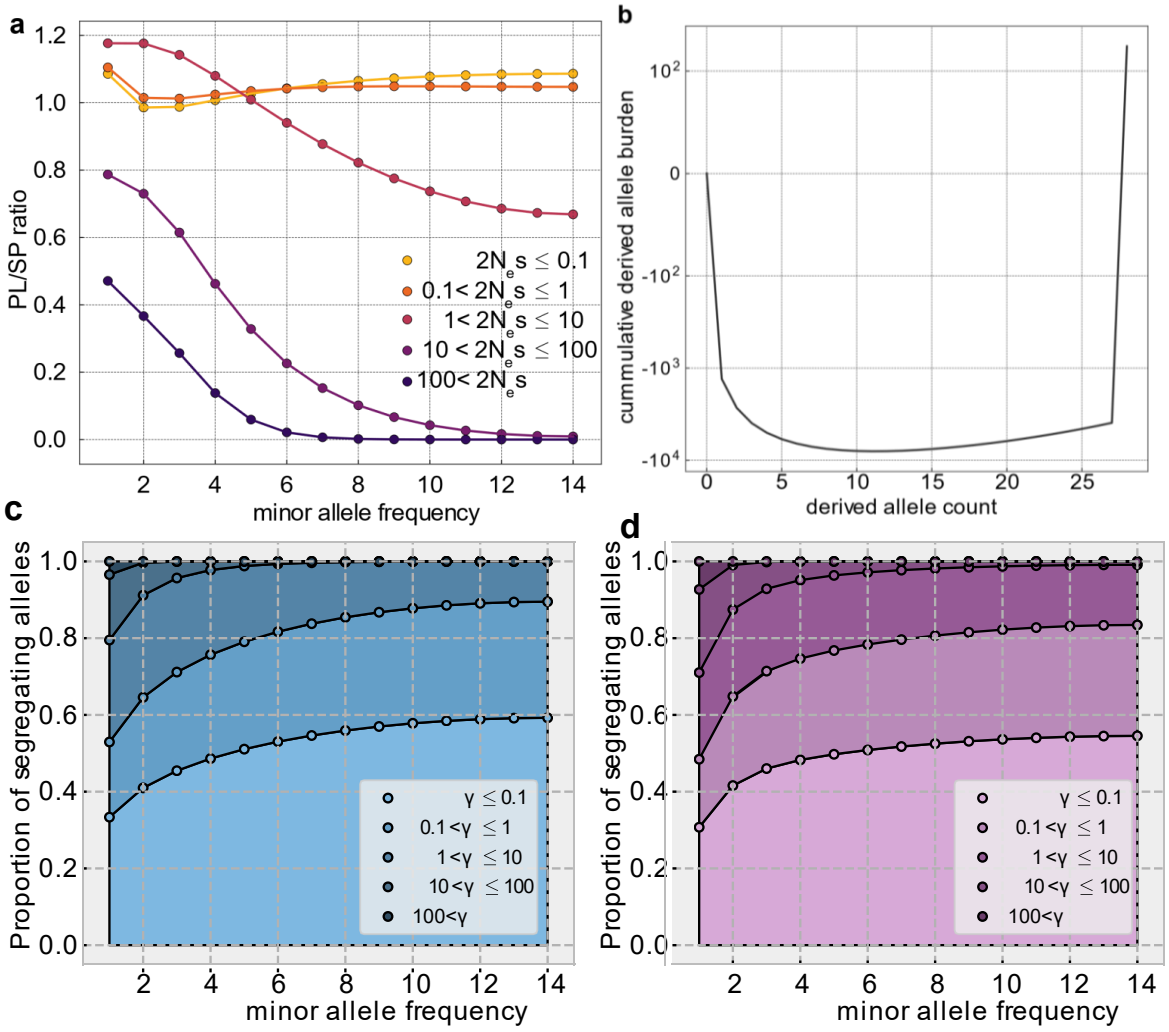


Figure 5: Comparative efficacy of selection and genomic burden in SP and PL. **a.** ratio of PL/SP of the proportion of variants for each s category and each allele frequency bin. Values below 1 indicate that mutations of a given size effect are less abundant in PL than in SP, within each frequency bin. This estimate is based on the gamma distribution of the DFE given by fit∂a∂I and the expected SFS in each category of s. As a proportion of the total number of variants at each count, PL has more slightly neutral and nearly neutral mutations (orange lines) at low frequency and considerably less strongly deleterious mutations (purple lines). **b**. Difference in per-individual cumulative derived allele burden between PL and SP given as the ratio of PL/SP. The cumulative derived allele burden is based on the contribution of deleterious variants depending on their count in the population assuming the point s estimate of deleterious mutations. Low frequency mutations contribute more to the burden in PL – negative values indicate that an excess of up to 10 000 deleterious mutations with count 10 or less in the population accumulate in each individual in PL-, whereas fixed mutations (count 28 in the population) play an important role in SP. The net difference, given by the end of the line, is 185. **c.** Distribution of selection coefficients for deleterious variants of each size category in each frequency bin of the SFS for PL **d.** Distribution of selection coefficients for deleterious variants of each size category in each frequency bin of the SFS for SP.

3.1.3 Estimates and Measures of accumulated burden in SP individuals

The severe bottleneck in the range edge population is expected to have facilitated the fixation of derived variants. Because some of them are expected to be deleterious, we investigated whether the per-individual burden in SP was higher than in PL. The number of derived non-synonymous mutations per Mbp of each lineage has been shown to be an appropriate proxy for the genomic load of a population, because its expectation is unaffected by demographic events (Simons and Sella 2016).

We used the inferred DFE to get an estimate of the expected burden in each of the two populations. SP and PL differed in the frequency of the variants contributing to the burden (Figure 5a). Low frequency mutations contribute more to the burden in PL, the core population. We inferred that an excess of about 10,000 slightly deleterious mutations of frequency below 30% were expected in PL, compared to SP. In the latter, instead, we inferred an almost equal excess of fixed derived mutations in the range-edge population (SP). Fixed mutations thus played a more important role in the estimated burden of SP individuals. The net difference, however, was comparatively small with an excess burden of 185 mutations per diploid genome in SP, compared to PL (Figure 5b). Although this number remains a crude estimation, it clearly indicates that the bottleneck in the range-margin population SP was not sufficiently long and severe to allow the accumulation of a much stronger burden.

I also directly measured the accumulated burden of deleterious mutations per individual haploid genome in the range edge and core population by calculating the mean count of derived mutations per haploid genome and corrected by the total number of genotyped sites. As expected, the mean per-individual count of derived synonymous mutations did not differ significantly between SP and PL (p = 0.121, Table 7; Figure 5c-d). There was a shift towards a smaller average number of synonymous mutations per genome in AUS (Figure 6a), which likely reflects a residual effect of the overall lower genomic coverage of AUS individuals. Thus, AUS individuals had to be excluded from this analysis. For each of the other two populations, I estimated the mean count of derived non-synonymous mutations (Figure 6b). The average burden accumulated by SP (range-edge) individuals reached a mean 0.0123 non-synonymous mutation per site (CI: 0.0118 – 0.0127). For the core population, PL, the mean burden was 0.0117 (CI: 0.0113 – 0.0121), which is 4.9% less than in SP. Permuting individuals among populations revealed that the mean difference between

the two populations is significantly different from zero (p <10e-4 for SP vs PL). These estimates remained identical when we removed regions with signatures of selective sweeps from the analysis (Figure 6bc, see also Methods as well as results below). Based on the approximate total of 2M non-synonymous sites per genome, I deduce that there are about 1,200 additional derived non-synonymous mutations per diploid genome in SP individuals, on average, compared to PL. Based on the estimated effect size of deleterious mutations above (point mass 2*Nanc*s estimated to be 1.2), this excess would result in a fitness loss of approximately 1,200 * 1.2.10-6 = 0.014%. Although this is higher than the theoretical prediction, it is much less than the approximately 3% fitness loss predicted in simulations (Gilbert et al. 2017).

I further used SNPeff (Cingolani et al. 2012) to identify mutations with a high deleterious impact and evaluate whether SP and PL could differ in the number of strongly deleterious mutations. Individuals in SP contained approximately 4.5% more such mutations (0.000164, CI: 0.000148-0.00018) than in PL (0.000156, CI:0.000142-0.000171, Figure 6b; Table 7). Individuals in SP contained approximately 4.5% more such mutations (0.000164, CI: 0.000148-0.00018) than in PL (0.000156, CI:0.000142-0.000171, Figure 6b; Table 7). Bootstrap across genomic regions, however, showed that this difference was not significant, with many regions in the genome showing no detectable difference in the number of mutations with high deleterious impact (p=0.183, Table 7). This indicates that the bottleneck was not severe enough to allow detecting a reduction of selection efficacy against strongly deleterious mutations.

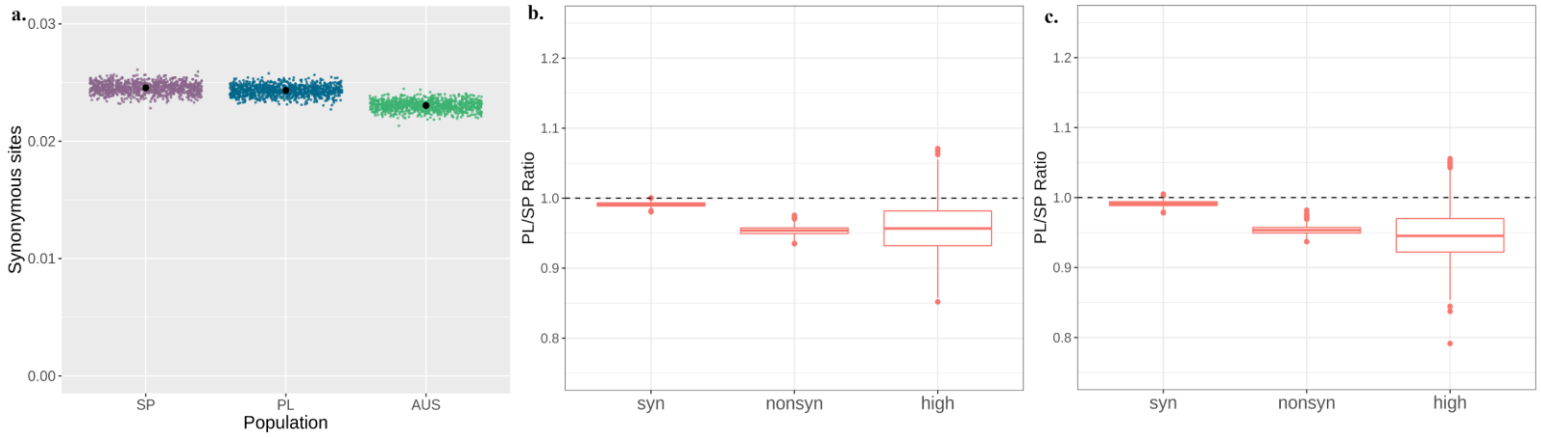


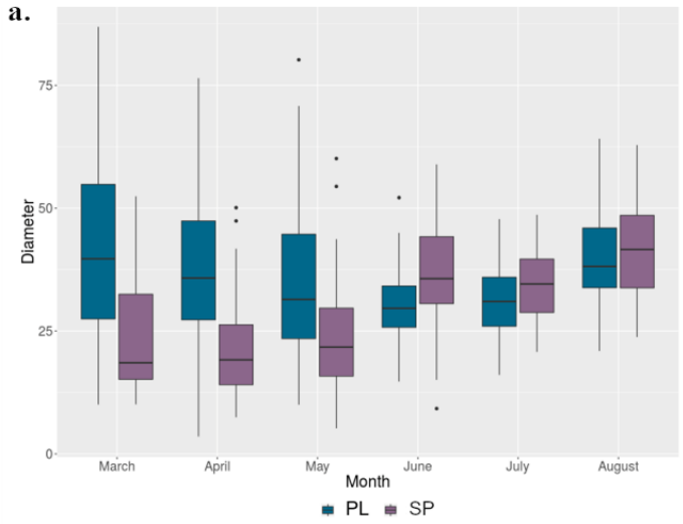
Figure 6: Accumulated burden per individual. **a.** The number of synonymous sites corrected by the total number of genotyped sites per sample for each population. For each population, the mean obtained during each bootstrap iteration is shown in color and the original mean is marked in black. The expectation is that all three populations should show no differences among the number of accumulated synonymous sites. The discrepancy noted between AUS and the other populations is the result of lower genome wide coverage, which lessened the power to detect derived mutations in this population. **b.** Comparison of genomic load in PL and SP, for synonymous, non-synonymous and high impact mutations. For each population, the genomic load was calculated as the mean number of non-synonymous corrected by the total number of genotyped sites for each sampled individual. The ratio of mean per individual genomic load of SP vs. PL is given. The distribution was established by bootstrap of the genome. **c.** Comparison of genomic load in PL and, for synonymous, non-synonymous and high impact mutations, when the areas with signatures of selective sweep have been removed. The values per category were not altered drastically compared to Fig 3c, which includes all the derived sites. For each population, the genomic load was calculated as the mean number of non-synonymous corrected by the total number of genotyped sites for each sampled individual. The ratio of mean per individual genomic load of PL vs SP is given, for 1000 permutations.

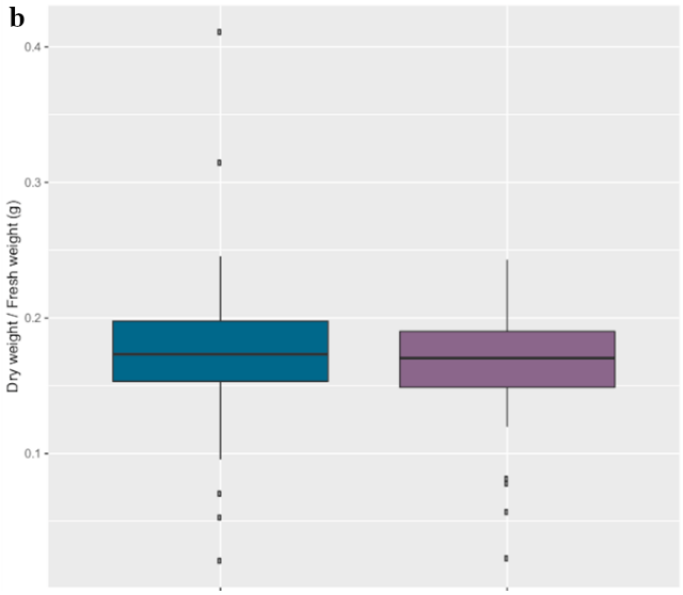
Table 7: Genomic load per population for synonymous, non-synonymous and high impact mutations as defined by SNPeff. Confidence intervals are provided within the parenthesis (see methods). The excess of mutations in SP was estimated by multiplying the genomic load difference by the number on non-synonymous sites in the A. lyrata genome as estimated by SNPeff.

Type of Mutations	**Total Number of Derived Positions**	**Genomic load in SP**	**Genomic load in PL**	**Genomic load Ratio of SP to PL**	**P-value**	**Excess of mutations in SP**
synonymous	125,228	0.0245 (0.0237-0.0252)	0.0243 (0.0235-0.0251)	1.008	0.121	400
non-synonymous	77,781	0.0123 (0.0118-0.0127)	0.0117 (0.0113-0.0121)	1.049	0.00099	1,200
high impact	1,323	0.000164 (0.000148-0.0001800)	0.000156 (0.000142-0.000171)	1.099	0.183	16

3.1.4 SP and PL show similar growth rate in a common garden of the species in the range core.

I further investigated whether a significant fitness erosion could be detected at the phenotypic level in the range edge population. I planted six replicate cuttings of 10 genotypes of each of the two populations in the common garden of University of Cologne, at a latitude that is comparable to that experienced in the species core range. The experiment was initiated early autumn and terminated a year later at the end of the growth season. During February, there was a week-long frost period, which resulted in 46% of SP plants to maintain their vegetative mass. In contrast, 86% of the PL plants maintained their vegetative mass, which was significantly higher than the SP proportion (p=0.000183). However, two months later most plants started producing new leaves and finally 86% of both SP and PL plants had vegetative mass. From this point onwards, none of the other replicates recovered and thus were considered dead. The effect of month (p = 2.362e-05) and of population (p=0.0001298) on the plant vegetative mass presence between January and April was significant. This indicates that the plants originating from PL are better equipped to handle frost than SP plants, which usually are covered by insulating snow during winter.





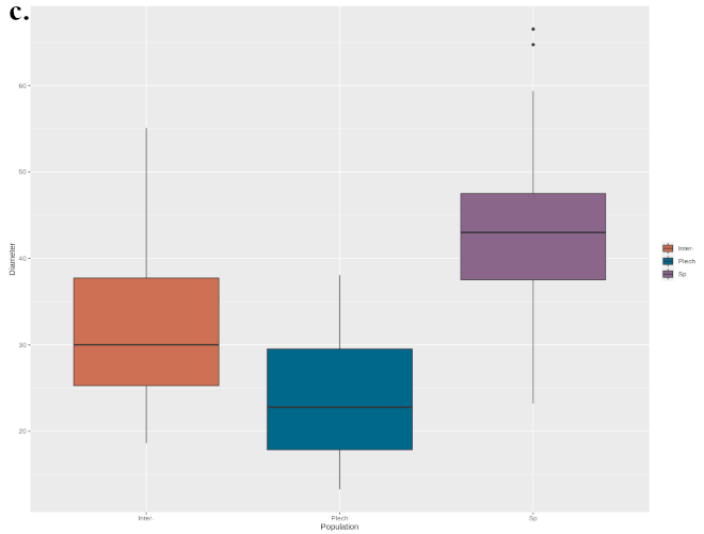


Figure 7: SP and PL show similar growth rate in a common garden experiment performed in the range core of the species' distribution. Six replicates of 10 genotypes per population were grown for one year in common garden setting in Cologne, which has climate representative of the core range of the species distribution. **a.** Diameter size (mm) per population for each month of the growing season in Cologne. Population and Month had a significant interaction ($p < 2.23\text{-}16$). The overall population effect was significant ($p = 0.01403$), even though SP and PL did not differ at the end of the growing season (August $p = 0.265$). **b.** Biomass of the plants at the end of the experiment as the dry to fresh weight ratio. The populations did not differ significantly ($p = 0.2873$). **c.** Diameter size (mm) per population when seedlings of *A. lyarata* had 25 leaves under controlled conditions. Seedlings from within population and between population crossings (orange color) were used for this experiment.

Although individuals of SP had a comparatively smaller rosette diameter after winter, the rosette diameter as well as their accumulated biomass did not differ from that of PL individuals at the end of the growth season (GLM model, p=0.26, and p=0.28, for the population effects of rosette diameter and accumulated biomass, respectively; Figure 7), due the comparatively higher growth rate of SP individuals during the growth season (Month and Population interaction $p < 2.2e-16$). Furthermore, none of these fitness measure correlated with the per-individual burden ($\rho = -0.111$, $p = 0.66$ for weight; $\rho = -0.149$, $p = 0.55$ for diameter at end of the season), nor with the level of heterozygosity ($\rho = 0.243$, $p = 0.34$ for diameter at end of the season; $\rho = 0.243$, $p = 0.29$ for biomass), which was estimated as the inbreeding coefficient F. Seedling growth was also not impaired in SP. I compared the growth of seedlings, which were produced by crossing the available individuals within and between populations (see Material and Methods). While growing under controlled conditions, the seedlings did not require significantly different time to produce the same number of leaves ($p = 0.4507$) It is notable that the between population hybrids did not produce leaves at different rates from the SP or PL seedlings, even though the diameter size ($p = 8.571e-14$***) was significantly different between the populations when they had 25 leaves. The SP had the largest rosette with median 43.01mm and PL the smaller diameters with median 22.79mm. The between population hybrids had intermediate values at 32.78mm diameter length

These analyses show that despite its increased per-individual burden and the potential impact of recessive deleterious variants, the cumulative effect of these mutations in the SP population did not result in a detectable decrease in complex fitness component traits such as growth. This observation is in agreement with previous reciprocal transplant experiments involving the same set of *A. lyrata ssp. petraea* populations, which concluded that the SP population is locally adapted (Leinonen et al. 2009). However, it stands in strong contrast with the clear effect of range expansion detected on plant survival and population growth rate in the relative *A. lyrata ssp. lyrata* (Willi et al. 2018).

3.1.5 Potential differences in recessive load in SP and PL

Recessive mutations with deleterious effects can segregate at higher frequency in a bottlenecked population and thus lead to a genomic load in the population that is higher than predicted by measures of per-individual burden (Balick et al. 2015). To evaluate whether recessive deleterious

mutations may contribute to the genomic load in SP and PL, I tested whether the F_{IS} distribution of non-synonymous mutations showed a departure from Hardy Weinberg expectations indicative of a selective removal of homozygotes. I found that both in range edge and core populations, the F_{IS} distribution of synonymous and non-synonymous mutations was significantly shifted towards lower values, revealing an excess of heterozygous non-synonymous mutations (Figure 8, KS test p < 2.2e-16). The shift towards negative F_{IS} values was more pronounced for the high impact variants, which were significantly different from the distribution of the synonymous sites (KS test p < 2.2e-16). This pattern suggests that, in both populations, offspring homozygous for deleterious alleles tend to be removed by selection. Compared to PL, the F_{IS} distribution of all sites in SP was shifted towards negative values, indicating a stronger excess of heterozygotes in this range edge population (Figure 8; p<2.2e-16).

While the F_{IS} statistic is generally useful to detect excess homozygosity, revealing cryptic population subdivision or partial selfing, excess levels of heterozygotes genome wide is more difficult to explain. I verified that this result was not influenced by unanticipated mapping biases by using the mean read coverage of each population and the distribution of genic coverage to set depth read filters. Then, for each filter, I correlated the new gene F_{IS} values with the median gene depth. Spearman's ρ was in the range of -0.125 to -0.145 for SP and -0.126 to -0.164 for PL. Filtering stringency did not modify the correlation, indicating that the F_{IS} bias that I specifically observe in SP is independent of read depth. In addition, I observed that SNPs with F_{IS} = -1 were not clustered in the genome, as would be expected from paralogous mapping. In contrast, the distribution of the physical distance between 2 consecutive such SNPs was significantly shifted towards higher values than two consecutive SNPs with any F_{IS} value (KS test for each population p < 2.2e-16). Therefore, I cannot fully rule out that this effect is not due to mapping inaccuracies, but it was independent from coverage thresholds or SNP density. It therefore suggests that the preferential removal of recessive homozygous might be more important in SP. Taken together, these data indicate that the per-individual burden we calculated may not fully recapitulate the deleterious load of the populations.

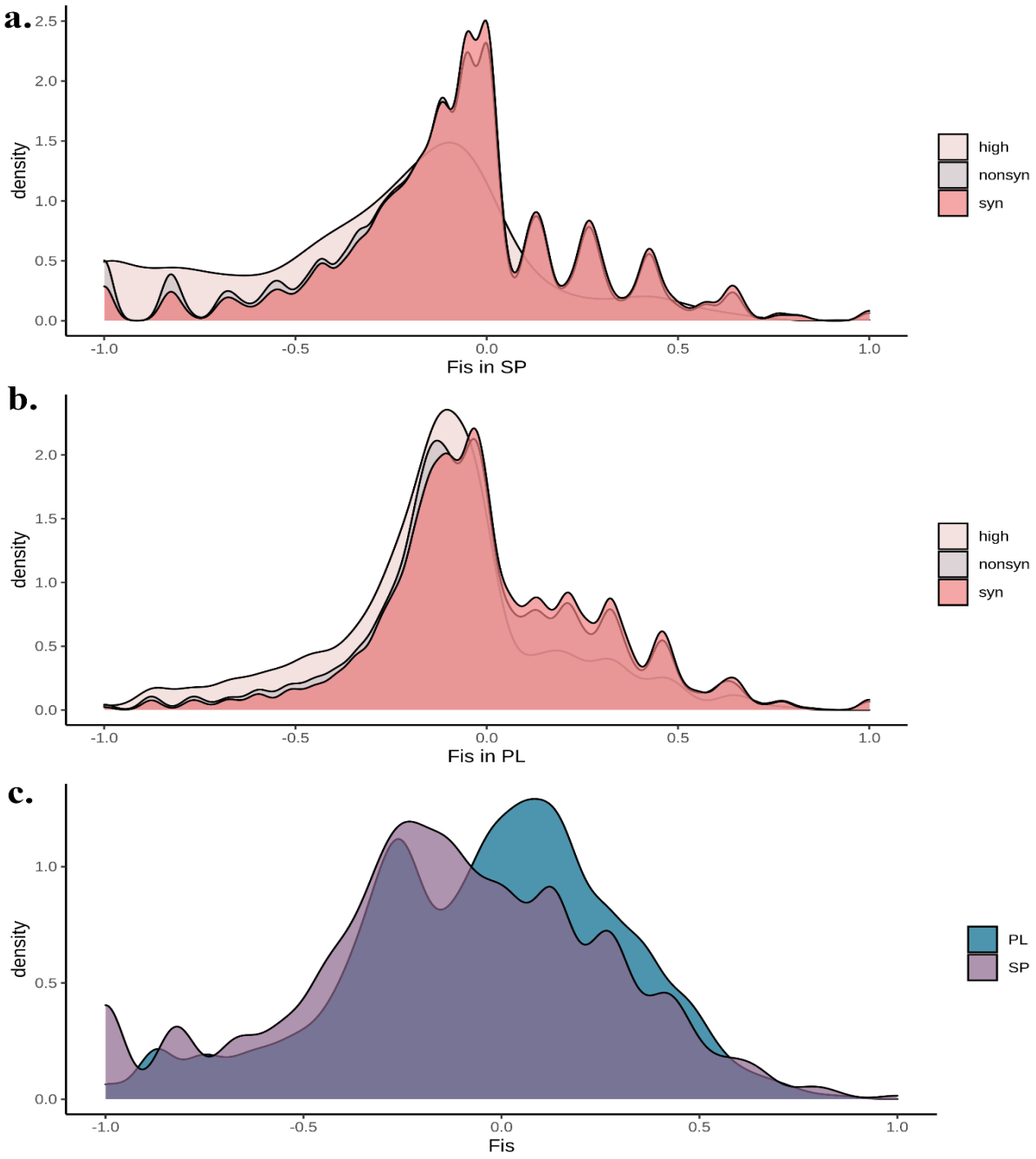


Figure 8: FIS distribution of SP and PL. **a.** The F_{IS} of high, non-synonymous and synonymous sites of SP is shown. **b.** The F_{IS} of high, non-synonymous and synonymous sites of PL is shown. **c.** In blue the F_{IS} distribution for the PL individuals is shown, and in pink the F_{IS} distribution for the field collected SP individuals is provided.

3.1.6 Selective sweeps in the range edge are broader than in the core but equally frequent

I searched for the footprints of selective sweeps within SP and PL – the two populations with the largest sample sizes using the Composite Likelihood Ratio (CLR) test. CLR estimates were computed in windows along the chromosomes with *SweeD* (Pavlidis et al. 2013). Significant deviations from neutral expectation were defined by comparing the observed diversity estimates to neutral diversity estimates simulated under the demographic model obtained above. I used the overlap of outlier CLR and F_{ST} to identify putative selective sweep regions specific to SP or PL (and thus indicative of local selection). I detected signatures of local sweeps within both populations despite their large differences in recent effective population size. In SP, I identified 1,620 local sweep windows, which grouped in 327 genomic regions of average size 7051bp and they cover 0.17% of the genome. Within PL, 745 windows, covering 104 genomic regions (average size 4,384bp; 0.87% of the genome), had PL specific signatures for sweep. In both populations, the sweeps were distributed along all the chromosomes (Appendix Table 2). Hence, the rate of adaptive evolution in the SP populations does not seem to have been compromised by the recent bottleneck.

Genes within the genomic regions carrying a population-specific signature of a selective sweep were extracted and tested for functional enrichment. In SP, fifteen Gene Ontology (GO) terms were enriched among genes showing signatures of positive selection (significance based on permutation derived *p* threshold of 0.0295). Interestingly, the top three GO terms were related to plant growth in response to environmental stimuli: “cellular response to iron ion”, “response to mechanical stimulus” and “response to hormone”. This observation agrees with the higher growth rate displayed by SP individuals in the common garden experiment. In PL, three GO enriched terms were significant (*p* threshold of 0.02137) and they were “intra-Golgi vesicle-mediated transport”, “regulation of anion transport” and “hexose metabolic process” (Table 8). Some of these functions have been associated with abiotic stress reactions in plants (Howell 2013) and may indicate adaptation in PL to the absence of snow cover protection during the cold season.

I further investigated whether specific groups of candidate genes carried signatures of adaptive evolution. Phenotypic differences in flowering time and especially selection related to the photoperiodic pathway, or to development have been shown to contribute to local adaptation in SP (Toivainen et al. 2014; Mattila et al. 2016; Hämälä and Savolainen 2019), as well as response to

abiotic factors such as cold and drought (Vergeer and Kunin 2013; Davey et al. 2018). I thus explored whether specific groups of genes associated with these traits carried signatures of adaptive evolution. I used the *A. thaliana* annotation to identify the *A. lyrata* orthologs of genes involved in these phenotypes. I then tested whether their F_{ST} estimates tended to be higher than the rest of the annotated genes (Table 9). An excess of high F_{ST} values was detected for genes involved in development and light ($p = 0.018$ and $p = 0.036$, respectively). Yet, genes related to dormancy, flowering, cold and water conditions did not exhibit significantly higher F_{ST} values than the control group (Table 9).

Table 8: Gene Ontology (GO) categories that are significantly enriched for genes located within areas with signature of selective sweep in SP or PL. The GO ID number, the description of the term and the p value of the Fisher's exact test (topGO) are provided. The p value threshold for determining the significance of the enrichment analysis was set by randomly picking genes to belong in area with signature of selective sweep, as described in the material and methods. Non significant GO.IDs are not shown.

GO.ID	**Term**	***P***	**Population**
GO:0071281	cellular response to iron ion	0.0043	SP
GO:0009612	response to mechanical stimulus	0.0066	SP
GO:0009725	response to hormone	0.0080	SP
GO:0009743	response to carbohydrate	0.0086	SP
GO:1901699	cellular response to nitrogen compound	0.0094	SP
GO:0042592	homeostatic process	0.0150	SP
GO:0019725	cellular homeostasis	0.0162	SP
GO:0090304	nucleic acid metabolic process	0.0165	SP
GO:0009581	detection of external stimulus	0.0200	SP
GO:0046483	heterocycle metabolic process	0.0203	SP
GO:0016070	RNA metabolic process	0.0246	SP

GO:0042991	transcription factor import into nucleus	0.024	SP
GO:0006338	chromatin remodelling	0.0247	SP
GO:1901360	organic cyclic compound metabolic processes	0.0260	SP
GO:0050801	ion homeostasis	0.0261	SP
GO:0006891	intra-Golgi vesicle-mediated transport	0.0024	PL
GO:0044070	regulation of anion transport	0.0063	PL
GO:0019318	hexose metabolic process	0.0141	PL

Table 9: Comparative analysis of F_{ST} distributions for gene groups. The genes were grouped based on candidate adaptive traits in the populations, or based on functions related to environmental differences between SP and PL, compared to the rest of the genome. The adjusted p values for multiple KS tests are provided.

Gene Group	**Control Group**	**Adjusted *p***
Flower	Athaliana GO annotated	0.235
Cold	Athaliana GO annotated	0.331
Water	Athaliana GO annotated	1
Light	Athaliana GO annotated	0.0364
Development	Athaliana GO annotated	0.0183
Dormancy	Athaliana GO annotated	0.899

3.1.7 Negative frequency-dependent selection maintained S-locus diversity in the range-edge population

Despite a smaller effective population size in SP, strong negative frequency-dependent selection acting on the self-incompatibility locus effectively maintained or restored S-allele diversity. In SP, 15 S-alleles (allelic richness was equal to 7.6) were detected across 22 individuals, with gene diversity at the S-locus equal to 0.828. These values were only slightly lower than to those observed within the 18 PL individuals (14 S-alleles; allelic richness = 8.1; gene diversity = 0.877) and the 7 AUS individuals (10 S-alleles; allelic richness = 10.0; gene diversity = 0.940) (Table 10). High S-allele diversity in SP (while a drastic reduction of the diversity at the S-locus would have been expected if a shift in the mating system had occurred), suggests that individuals are highly outcrossing and thus that the past bottleneck does not seem to have affected the mating system. The S-locus F_{ST} between SP and either PL or AUS was equal to 0.027 or 0.037, respectively, values much lower than the whole genome (0.231 or 0.234, respectively) as predicted by (Schierup et al. (2001).

Table 10: S –locus allelic diversity has been maintained within SP. The number of S-alleles for each population sample, as well as the number of individuals is provided. For each population the allelic richness has been calculated according to a rarefaction protocol with N=7.

Population	**S – alleles**	**Allelic Richness**	**Sample Size**
SP	15	7.6	22
PL	14	8.1	17
AUS	10	10.0	7

3.2 Chapter 2: The adaptive potential of gene expression variation

3.2.1 Gene expression variance is mostly not heritable

Whole transcriptome sequences of controlled inter-population crosses were used to analyse the gene expression variance of the species. The crossing scheme included the production of full-sib and half-sib families (see Methods), as this pedigree allows the accurate partitioning of the genetic variance into additive (V_A) and dominance (V_D) genetic variance (Falconer and Mackay, 1996; Lynch and Walsh, 1998). In total, I grew 131 individuals, from 10 full sibling families, each of which had between 4 and 20 members (Table 2). I implemented the animal model using the R library MCMCglmm (Hadfield 2010; see Material and Methods) to partition the total phenotypic variance (V_P) of 17,657 expressed transcripts to V_A, V_D, residual variance (V_R) and maternal variance (V_M). The sum of the V_A and V_D corresponds to the total genetic variance of the gene expression.

Most of the phenotypic variance could be attributed to genetics, as V_G composed more than 50% of the total phenotypic variance for 67.7% of the genes (Figure 9). However, the additive and dominance variance contributed to the phenotypic variance disproportionally, with 6.39% and 25.4% of the transcripts having more than 50% of V_P explained by V_A or V_D, respectively. I compared the generalized mixed model against a model that does not include the dominance matrix. Thus, it was possible to check whether V_D contributes significantly to the phenotypic variance. Based on the model's AIC values, all but 26 genes' expression was described better by the inclusion of the dominance matrix in the model. Most full-sib families included reciprocal crossings between the parents, and thus, I estimated the proportion of maternal variance (V_M) out of the V_P as being close to zero, with mean and median values of 0.088 and 0.075, respectively. I tested for effect of the population of origin of the mother plant on the gene expression level, by including the mother's population as fixed effect. There was no significant effect of the mother plant population on any genes' expression (Bonferroni corrected threshold of p = 2.486e-06). Thus, gene expression variance is genetic and mostly attributed to the fraction of genetic variance that is not heritable.

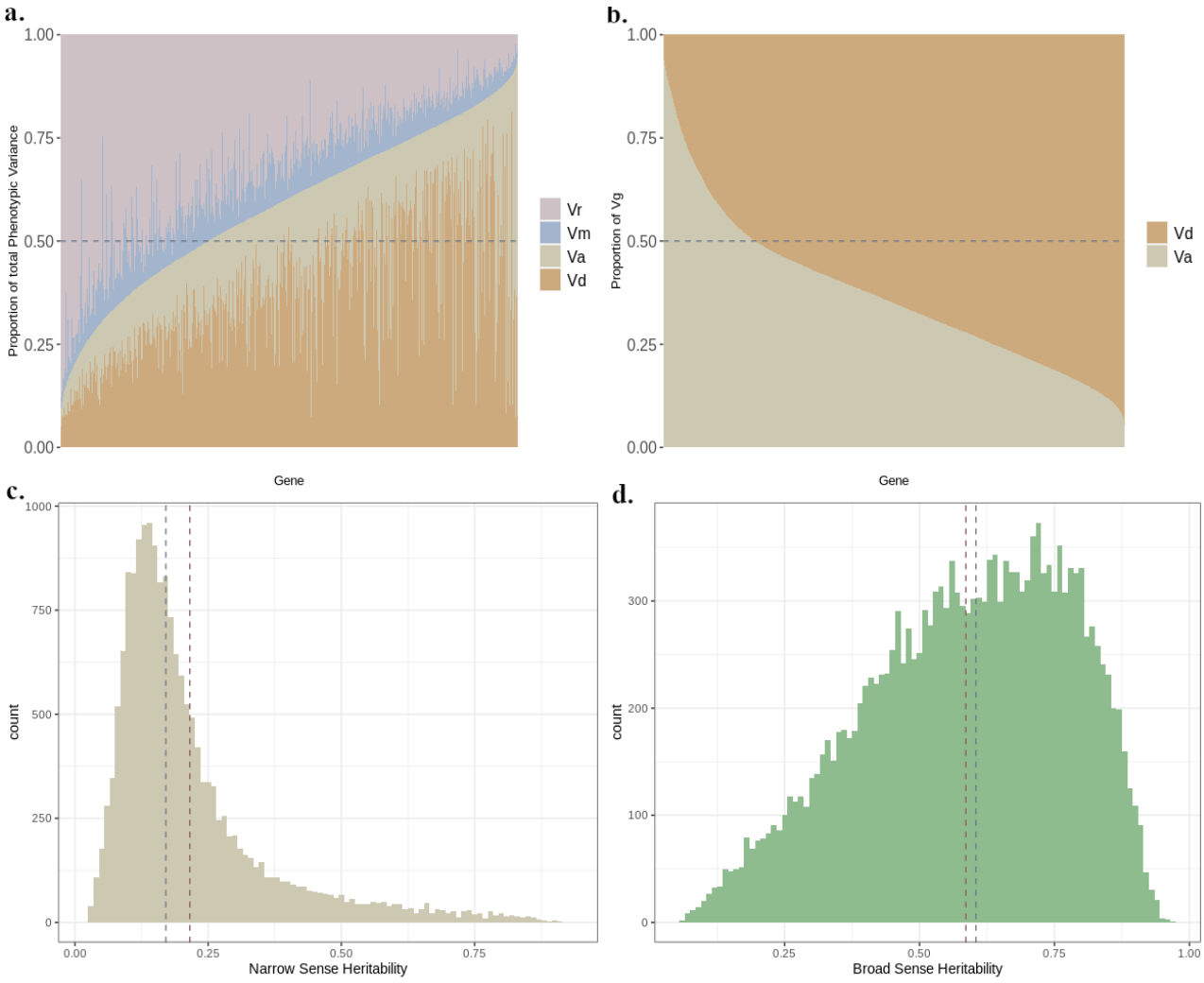


Figure 9: Gene expression variance is mostly described by genetic dominance variance. **a.** Proportion of the total phenotypic variance for each of the 17,657 genes' expression, which is composed of residual (Vr), maternal (Vm), additive (Va) and dominance (Vd) variance. The horizontal dashed line marks the 0.5. **b.** Proportion of the genetic variance that is additive or dominance variance, per gene. The horizontal dashed line marks the 0.5. **c.** Distribution of the Narrow Sense Heritability or else Va. Narrow sense heritability was estimated as the additive variance divided by the total phenotypic variance. **d.** Distribution of the Broad Sense Heritability or else Vg. Broad Sense heritability was estimated as the sum of additive and dominance variance divided by the total phenotypic variance. In both c. and d. the red and blue dashed lines mark the mean and median of the distributions, respectively.

The distributions of the additive and dominance variance along the genome are not random. I determined whether specific molecular and biological functions can be associated with either V_A or V_D based on their decreasing values. In total, 90 and 174 GOs were significantly enriched for high V_A and high V_D values at significance threshold of $p = 0.03813$ and $p=0.3810$, respectively.

The significantly enriched GOs for high values of V_A were accumulating in different functions than the high values of V_D. The five enriched GOs for the highest V_A values are "Glucose catabolic process" (p = 8.3-05), "Glycolytic process" (p = 1.6e-05), "gluconeogenesis" (p = 9.1e-05), "response to cadmium ion" (p = 0.00015) and "proteasome core complex assembly" (p = 0.00016; Appendix Table 3). Interestingly, the enriched categories for high V_A are reminiscent of the enriched GOs for genes within sweep regions identified in SP population (see section 3.1.6 and Table 8). In contrast, the five top enriched GOs for the highest V_D values are "chromatin silencing" (p = 4.5e-16), "cytokinesis by cell plate formation" (p = 2e-11), "DNA methylation" (p = 4.3e-11), "nuclear-transcribed mRNA catabolic processes" (p = 9.2e-11) and "methylation-dependent chromatin silencing" (p = 6.3e-10; Appendix Table 4). Based on the p-value of enrichments, functional enrichment appeared to be much stronger for genes displaying high V_D.

3.2.3 Gene structural properties correlate with the degree of dominance variance along

Dominance variance is defined as the non-heritable genetic variance derived by the allelic interactions within a locus (Fisher 1918; Falconer and Mackay 1996). Thus, we tested the impact of the gene structural properties that would create differences in the genic sequences and this way, they could contribute to the proportion of V_D out of the V_P. The nonrandom accumulation of the V_A and V_D values among the functions has been found to correlate with gene structural properties. We trained a random forest model (see Methods) to identify the genomic or population genetic parameters that best explain variation in dominance and additive variance in gene expression. The model's predicted value correlated poorly with the true values for V_A (R^2 = 0.009 and mean square error of 0.021). We therefore could not associate variation in V_A with any such parameters (Figure 10a-b).

For V_D, instead, the model's predicted value correlated well with the true values for (R^2 = 0.18 and mean square error of 0.026). The random forest analysis revealed that various components of gene architecture form the most important factor for explaining the level of the V_D observed in the transcriptome (Figure 10a). The number of exons per gene, the total gene length and the transcript length were the factors that most contributed to determine dominance variance in transcript expression. The level of variable importance of each parameter persists even when we correct all tested parameters for gene length. Transcript length has a significant positive correlation of ρ=0.258 (p=1.22e-266) with V_D, but it does not correlate with median gene expression (ρ=0.093,

p=9.54e-16; Table 11). We used a linear model to show that the significant effect of transcript length on V_D ($p < 2.2e-16$) remains significant even after correcting for median gene expression ($p = 0.0176$) and V_G ($p < 2.2e-16$). The tested model did not show significant interaction between transcript length and median gene expression (p=0.224). On the other hand, we found that V_A has a significant but negative correlation with transcript length (ρ= -0.177, $p < 2.2e-16$; Figure 10c-d).

The density of SNP sites within a gene, as well as the total number of transcription factors accumulating within a gene are also predicted to explain V_D well. The most important transcription factor was found to be the Dof zinc finger protein (DOF5.3; p=0.01639).

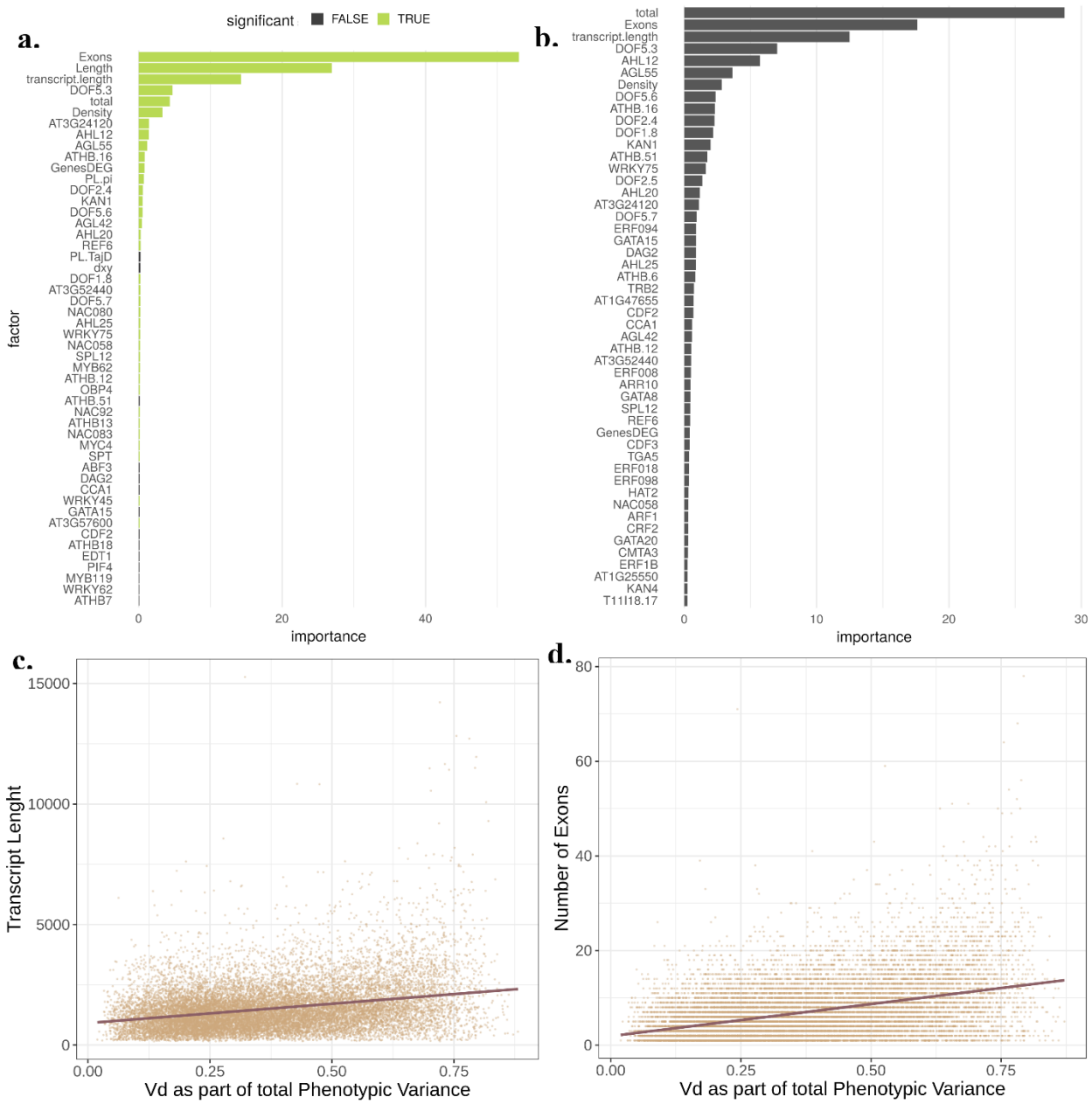


Figure 10: Gene architecture traits can predict the level of dominance variance. The level of variance that each factor included in the random forest model explains for **a.** dominance variance and **b.** additive variance. The bars are colored green and black when their effect is significant or not significant, respectively. **c.** Transcript length correlation with the dominance variance. **d.** Number of exons per gene correlate with the dominance variance.

3.2.4 Gene clustering highlights the impact of gene structural variation and population divergence on the components of the genetic variance

Gene expression is a highly regulated phenotype, as genes are inter-connected in co-expression and co-regulation pathways. To gain insight into how many effectively independent phenotypes could be observed in our dataset, I used a hierarchical clustering approach to group the genes based on their spearman correlation coefficient ρ. Based on the within groups sum of squares, the identifiable number of potential clusters is between 23 and at least 300 (Figure 11a). This indicates that variation in distinct gene expression is at least partially independent from each other, with at most 300 observations.

I investigated the impact of gene clustering. I grouped the data into 200 clusters. Each clusters' size was very variable, with number of genes per cluster between 5 and 2,829 genes (median=39.5 genes per cluster; Figure 11b). The number of genes per cluster showed a significant positive correlation with the median V_D of the cluster (ρ = 0.49 and p = 1.253e-13; Figure 11c). The opposite relationship was observed when we correlated the gene number per cluster with the V_A (ρ = -0.277, p = 7.101e-05; Figure 11d). Moreover, the median V_D per cluster was positively correlated with the median transcript length (ρ = 0.201 and p = 0.0042) and median gene exon number (ρ = 0.439 and p = 7.752e-11). This result does not differ from the pattern observed for each gene (see section 3.2.2 and Table 11).

Correlation between V_A and population genetic parameters of the SP and PL was detectable, indicating that the genetic divergence of the populations (described in Chapter 1) has contributed to the additive variance. The median V_A of each cluster was positively correlated with the median F_{ST} (ρ = -0.201 and p = 0.0041) and dxy (ρ = 0.32 and p = 2.59e-06) of SP vs PL. There was correlation of the median V_A with the median π values of SP (ρ = 0.34 and p = 8.643e-07) and PL (ρ = 0.33 and p = 1.627e-07). The V_A per gene was also significantly correlated with ρ= 0.115 (p<2.2e-16) and ρ= 0.1152 (p<2.2e-16) for π of PL and dxy of SP vs PL, respectively.

A subset of the dataset is well described by epigenetic non-heritable variance. I detected a core of approximately 2,829 genes that were almost always clustered in the same group during the previous analysis. This cluster was the largest one and had median V_D and V_A values of 0.501 and 0.141, respectively. Gene Ontology enrichment analysis highlights 294 GOs (significance threshold p=0.077; Appendix Table 5). The top processes are related to developmental functions,

such as “gravitropism” (p = 6.4e-28), “trichome morphogenesis” (p = 8.9e-15), “protein glycosylation” (p = 9.7e-15) and “protein N-linked glycosylation” (p = 2e-14).

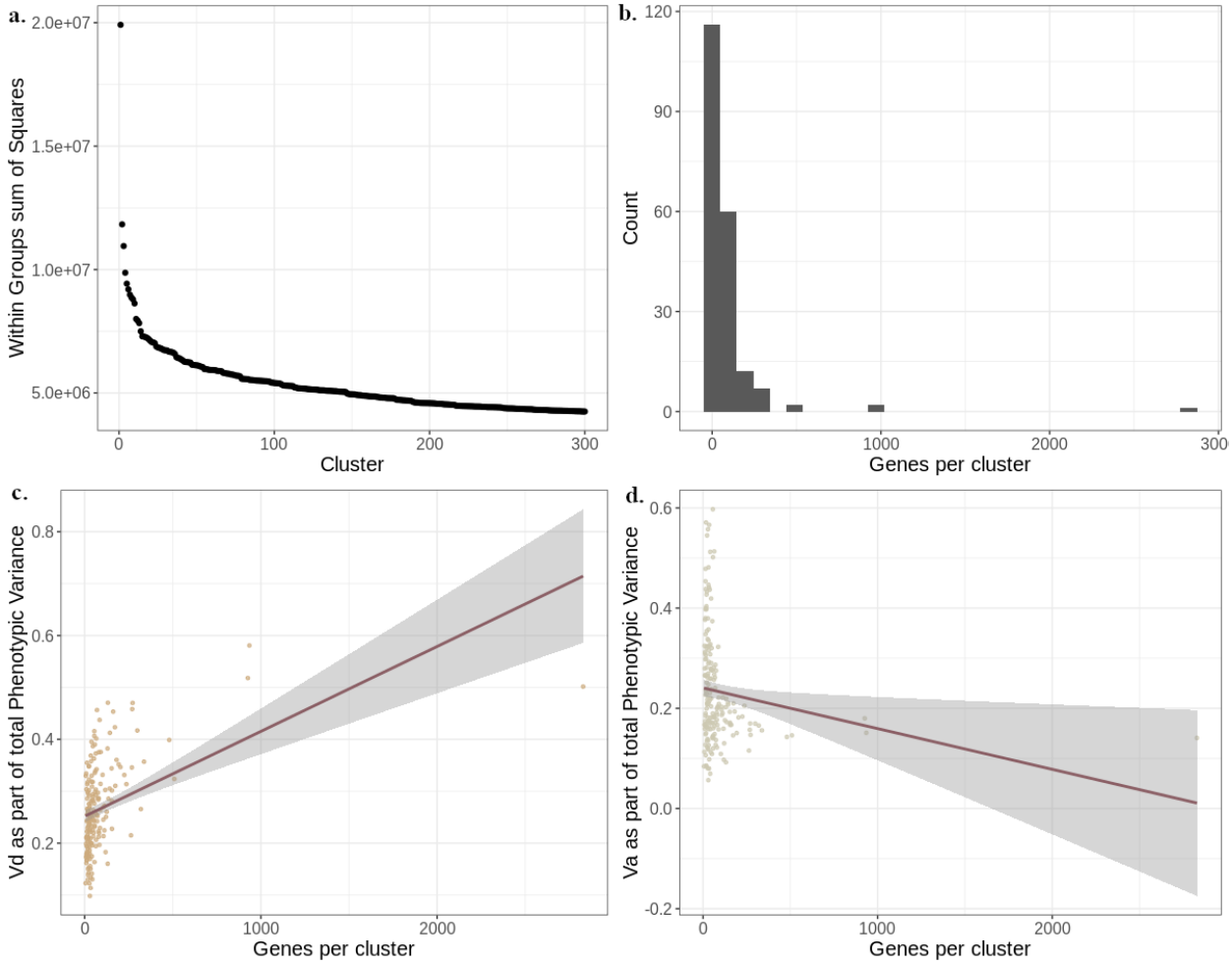


Figure 11: Clustering of genes based on pairwise gene expression correlation. **a.** The within sum of squares for 1 to 300 clusters. **b.** Genes per cluster when the genes are grouped in 200 clusters. **c.** Median dominance variance per cluster has a positive significant correlation with the number of genes within each cluster. **d.** Median additive variance per cluster has a negative significant correlation with the number of genes within each cluster.

Table 11: Correlation of dominance and additive variance with genome architecture traits and population genetics statistics. For each correlation of variance and feature, the p value, spearman's ρ as well as the number of clusters we had grouped the genes into are given. When the number of clusters is 1, the expression level of each gene was treated as independent observation.

Variance	**Feature**	**Number of clusters**	***P* value**	**Spearman's ρ**
Dominance	Transcript Length	1	1.22e-266	0.2588
	Median Gene Expression	1	9.54e-16	0.093
	Number of Exons	1	<2.2e-16	0.3852
	Gene Number	200	1.253e-13	0.49
	Transcript Length	200	0.0042	0.201
	Number of Exons	200	7.752e-11	0.439
Additive	Transcript Length	1	<2.2e-16	-0.177
	F_{ST}	1	0.0152	-0.023
	Dxy	1	<2.2e-16	0.1152
	π of SP	1	<2.2e-16	0.0916
	π of PL	1	<2.2e-16	0.115
	Gene Number	200	1.101e-05	-0.277
	F_{ST}	200	0.0041	-0.201
	Dxy	200	2.59e-06	0.32
	π of SP	200	8.643e-07	0.34
	π of PL	200	1.627e-07	0.33

3.2.5 Genes with high additive variance have signals of relaxed purifying selection

Genes with high dominance variance in the dataset show signals of pleiotropy and network connectivity, two properties that are often connected with strong constraints on the genome level (Hahn and Kern 2005; Josephs et al. 2015). Furthermore, to test for signals of directional selection, we estimated the DFE of non-synonymous mutations for the SP and PL genomes, which were described in Chapter 1. For each population, I extracted non-overlapping sets of genes based on the level of different variances they had (Figure 12a). In total, I have tested 4 different groups in addition to a random set of genes. Two group of them had either high V_D or V_A values, that are equal or more than 0.5 of the total phenotypic variance. Moreover, a similar gene group with

combined V_R and V_M (V_R+V_M) values of equal or more that 0.5 of the total phenotypic variance was tested. Finally, a group of genes with V_G more or equal to 0.5 of the total phenotypic variance, which is that high due to similar contribution of V_A and V_D (more than 0.25 but less than 0.5 of total phenotypic variance each) was tested.

The results show, that the groups of genes with different levels of genetic variance have evolved under different levels of purifying selection (Figure 12, Table 12). The group of genes with low genetic variance (V_R+V_M) has a significantly lower proportion of nearly neutral non-synonymous mutations ($0 < N_{anc}s < 1$) than the other groups, as well as a significantly higher proportion of mutations under strong purifying selection ($N_{anc}s > 10$). Genes with high genetic variance but intermediate dominance and additive variance (V_G) show stronger constraint than V_A, V_D and the control group, but less constraint than the genes with (V_R+V_M). The fraction of nearly neutral sites for the V_G group is significantly lower than the other groups as well as having a significantly higher fraction of mutations under strong purifying selection, except for (V_R+V_M). In contrast, genes with high V_A has signal of relaxed constraints. It has a significantly higher fraction of nearly neutral non-synonymous mutations ($0 < N_{anc}s < 1$) than the other groups, as well as a significantly lower fraction of mutations under strong purifying selection ($N_{anc}s > 10$). Genes with high level of V_D did not significantly differ from the control group consisted of random genes.

Table 12: The p values for all pairwise comparisons within each DFE bin. The gene groups were specified based on their level of additive, dominance, genetic as well as the sum of the residual and maternal variance, as seen in Figure 12a..

$0 < N_{anc}s < 1$				
	V_A	**V_D**	**V_G**	**V_R+V_M**
Rand	0.03	0.3	0.01	0.01
V_A	NA	0.01	0.01	0.01
V_D		NA	0.01	0.01
V_G			NA	0.01
$1 < N_{anc}s < 10$				
	V_A	**V_D**	**V_G**	**V_R+V_M**
Rand	0.23	0.19	0.01	0.01
Va	NA	0.52	0.01	0.01
Vd		NA	0.01	0.01

V_G			NA	0.43
10< N_{anc}s < Inf				
	V_A	**V_D**	**V_G**	**V_R+V_M**
Rand	0.01	0.39	0.01	0.01
V_A	NA	0.07	0.01	0.01
V_D		NA	0.01	0.01
V_G			NA	0.01

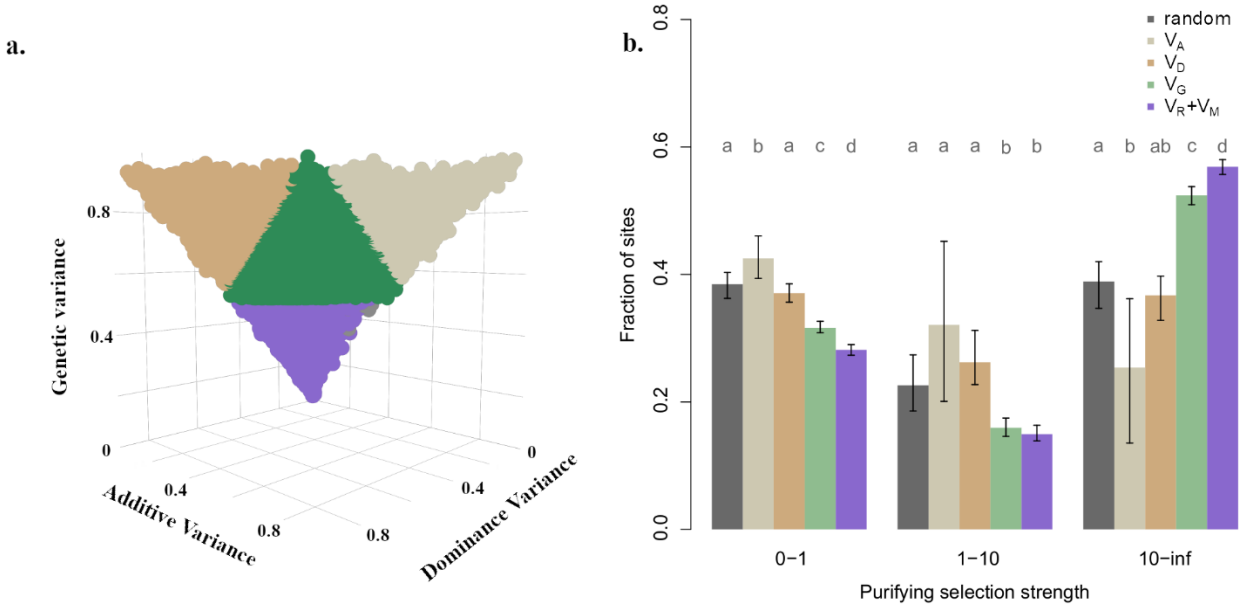


Figure 12: Different levels of constraints for genes with excess of different variance type. **a.** All the genes are plotted based on their additive, dominance, and genetic variance. The colors correspond to each of the 5 groups as indicated on the legend of figure b. The V_D and V_A correspond to genes with more than 0.5 additive and dominance variance, respectively. V_G corresponds to the group of genes with genetic variance more than 0.5 and additive and dominance variance between 0.25 and 0.5. The V_{R+M} group corresponds to the genes with a combined value of maternal and residual variance above or equal to 0.5. **b.** The distribution of fitness effects in bins of N_{anc}*s for different set of genes. It was estimated based on the site frequency spectrum of the non-synonymous sites in the genomic samples of SP and PL.

3.3 The polygenic basis of gene expression in *A. lyrata*

3.3.1 *Trans* effects indicate the polygenicity of gene expression variation

I next asked whether we could determine the genetic basis of gene expression variance, in order to confirm whether the genetic architecture explains dominance variance. Since genetic variation segregated in the 131 individuals, it was possible to perform GWAS, controlling for the structure of the population. The number of variants was restricted to the number of heterozygote sites in the 10 parents of the population. This approach was used to estimate the genetic architecture of gene expression variance. Evidence of the polygenic basis of gene expression variation was found via genome wide association study (GWAS). I used the transcript counts per gene to associate with the genotypic variance, which is observed in the transcriptome. In total, 127,585 SNPs were included in the GWAS analysis. Few genetic associations were identified, with only 174 transcripts having at least one significant association. This was revealing of the polygenic basis of gene expression variation (Figure 13a). The unique significant associations were 156 across all analysis were in total 156 and they corresponded to equal number of genes. On average, each of the 174 transcripts had only one significant association. Similarly, on average, each significant hit was identified successfully in only one association analyses, with a few exceptions (Figure 13b; Table 12). However, there are a few SNPSs that were identified as significant hits in more than 5 association analysis. It is notable, that none of the GWAS returned an association located within the genomic region of the transcript tested. Thus, all the associations have a potential *trans* regulatory effect on the transcripts and no *cis* regulatory effect. The V_D of the associating genes has a bimodal distribution (median = 0.33 vs control median = 0.37) and it is significantly higher (KS test p= 0.00598). The distribution of V_A values of the associating genes (median=0.22, median=0.15) is also significantly higher (KS test p=2.253e-05) than the rest of the genes. There was no difference for V_G (p=0.1).

I compared the V_A and V_D values of the genes within which a SNP with significant association is located (*trans* effect gene), to the values of genes with no significant hit located within them (control genes). *Trans* effect genes have more V_A (median = 0.3) but less V_D (median = 0.33), compared to the control genes (median of 0.20 and 0.39 for V_A and V_D respectively; Figure 13c). This indicates that genes with polygenic background have higher V_D than the genes under the control of a few genes. Moreover, less of the *trans* effect genes (24%) were located within sweep

regions than the control genes (36%). I used a generalized linear mixed model to test whether F_{ST} or positive selection signal can have an impact on the values of V_A, V_D and V_G. Also, I tested with the same model whether V_A, V_D and V_G distributions are different. The variances are all significantly different from each other (p < 2.2e-16) and the signal of positive selection has a significant impact on each component of genetic variance (p = 0.0001319). Overall, it was observed that *trans* effect genes located within a sweep area, have more V_A and less V_D than the ones with no positive selection (Figure 13d). Thus, genes with large effect on the expression of another gene have more additive variance, which could enable the action of positive selection.

Table 13: The genes that were identified in the majority of the GWAS analysis as significant associations. The gene name, chromosome, start and end position are given. Furthermore, the number of transcripts they associate with are given.

Gene Name	**Chr**	**Start**	**End**	**Transcript**
gene:fgenesh2_kg.4__134__AT2G22120.1	4	665407	668437	36
gene:scaffold_302751.1	3	10320792	10322007	11
gene:fgenesh2_kg.8__2670__AT5G67420.1	8	22300392	22301941	10
gene:scaffold_701797.1	7	7237139	7237787	8
gene:fgenesh1_pg.C_scaffold_8000314	8	2411704	2412746	7
gene:fgenesh2_kg.1__1019__AT1G09650.1	1	3676022	3677313	7
gene:fgenesh2_kg.6__1519__AT5G15390.1	6	6272766	6273086	6

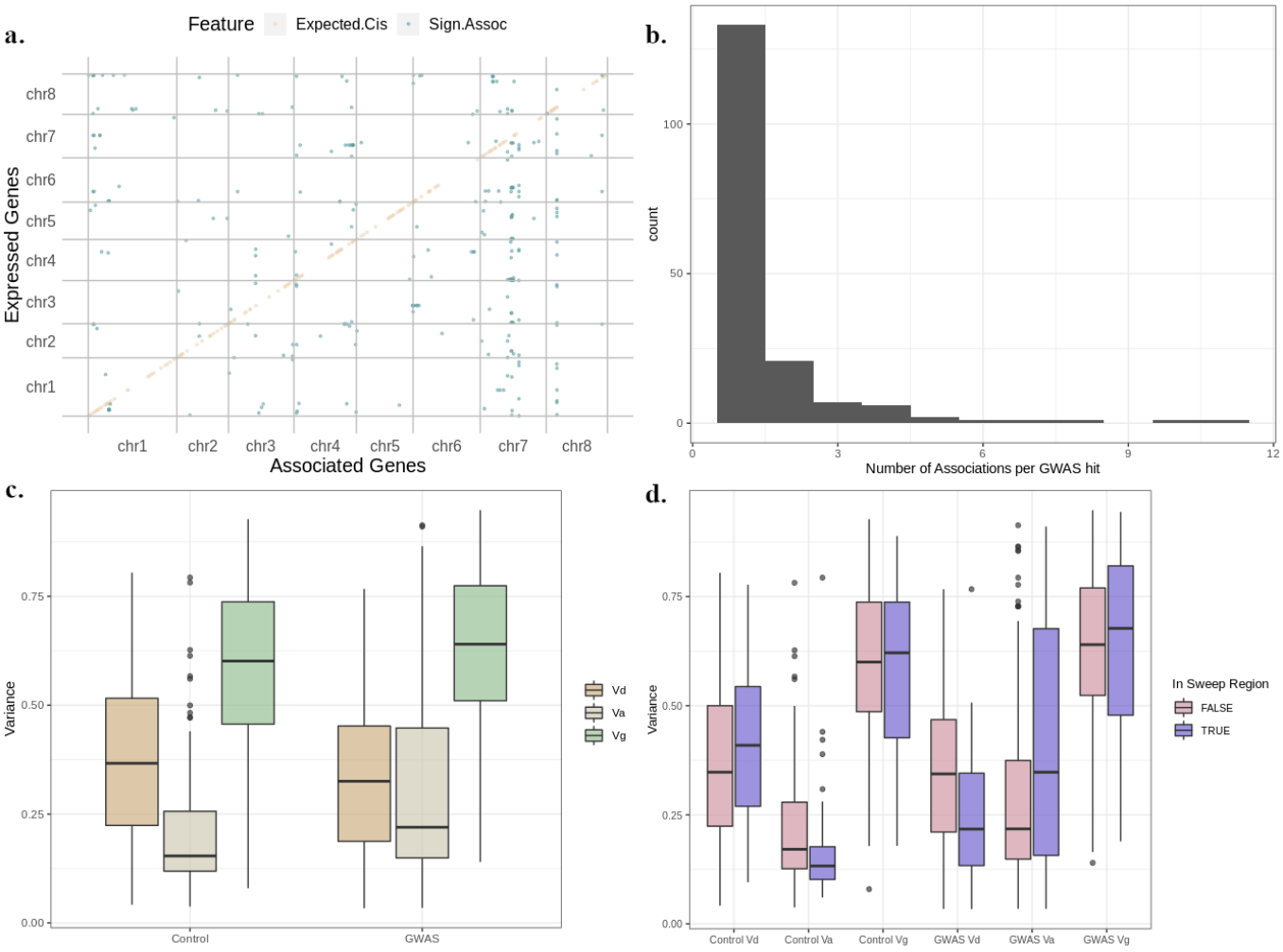


Figure 13: Genome wide association study reveals the polygenic basis of gene expression variation. **a**. The results of the GWAS run for each individual gene are summarized. On the y axis, the position of the expressed gene whose phenotype was tested is shown. On the x axis the position along the chromosomes of the *trans* effect gene, which associates significantly with the expressed gene is given. Brown dots show the position of the expected cis effect, blue dots show significant associations of genetic variant with variation in transcript level. Horizontal clusters of variant association reveal a set of transacting variants regulating many transcripts. **b**. Histogram showing the number of transcripts associating with each variant. **c**. Boxplot representation of the dominance (V_D), additive (V_A) and genetic (V_G) variance for the GWAS significant hits and a control set of genes. **d**. Comparison of the dominance (V_D), additive (V_A) and genetic (V_G) variance for the GWAS significant hits and a control set of genes, grouped by their presence or absence within a sweep region.

3.3.2 PL specific alleles drive most of the allele specific polymorphism in the dataset

The signals of selection identified within the two populations indicate the presence of polygenic selection along the whole genome. Furthermore, the substantial heritability of the gene expression variance, as well as the impact of transcription factors on the degree of dominance variance indicate that *cis* regulatory variance is important for the diversification and evolution of the two populations.

I used 105 out of 131 available transcriptomes at the time of the analysis, to explore the *cis* regulatory divergence between the families. I had crossed plants from SP to plants from PL, a design that enables the creation of heterozygous sites and the identification of allele specific expression (ASE). ASE can be used to identify the *cis* regulatory elements that vary between the parents. Thus, for each individual, we identified the parental origin of the expressed alleles, using the parental SNP calls as guide (see Methods). A gene was flagged as being ASE if the ratio of the expressed alleles was significantly different to the 1:1 proportion, based on Bonferroni corrected p value (threshold of $p = 0.001$), which were obtained by chi square test. In total, the median number of ASE genes per individual was 2,454, which is approximately 50% of all expressed genes per individual (Figure 14a). It is noted that for members of fam03 and fam08, the median number of genes with ASE is 16.1% and 14.6% respectively. Furthermore, I estimated the ASE per family, as the genes that are ASE in 30% of the family members (Figure 14b). The median gene number with ASE per family is 3,594 genes, which is the 72% of all genes within the family. I observe here the same discrepancy in the ASE numbers as before; fam03 and fam08 have 24.7% and 18.5% genes as ASE, respectively.

This difference between the fam03, fam08 and the rest of the families can be attributed to the way the gene expression datasets were filtered. Using genome sequences of family members as guide, areas, where the mapping shows a bias towards the mapping of one allele, have been removed (see methods). Thus, the ASE number per individual in the rest of the full sibling families, is inflated by possible mapping errors. However, this preliminary dataset allows to explore the *cis* regulatory elements divergence and the potential impact of selection on them.

A first aspect of the dataset to be explored, is the potential of preferential overexpression of the allele originating from one population. In order to do this, I categorized ASE genes as polymorphic

or not for one population, based on the comparison of half sibling families (see Methods). There is evidence for preferential expression of the allele from the PL parents over the SP parent.

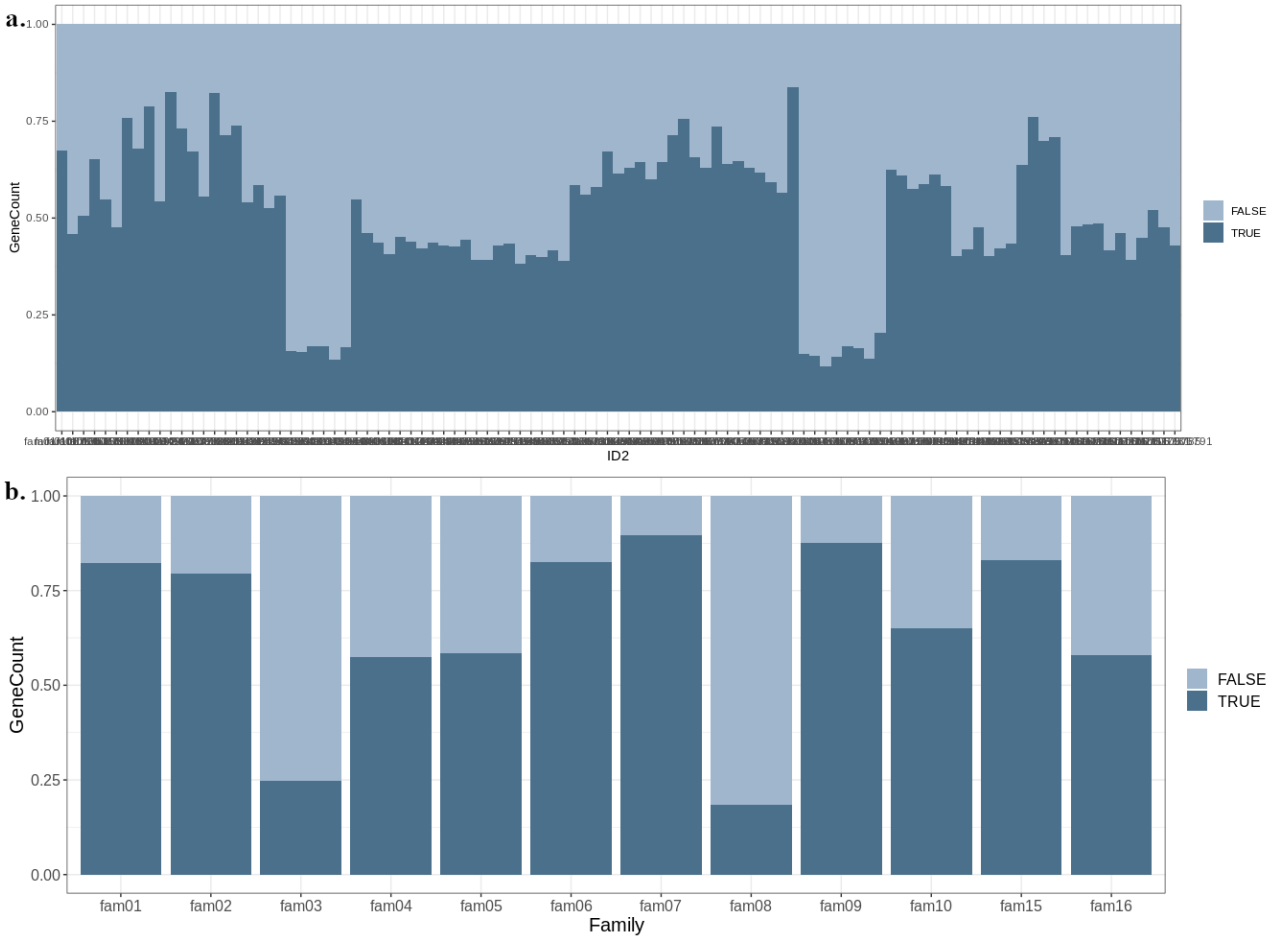


Figure 14: Proportion of ASE genes out of the total expressed genes per **a**. individual and **b.** per family. The individuals are grouped by family and their order corresponds the order presented in b.

I identified 184 and 506 polymorphic genes for SP and PL, respectively. The population specific polymorphic alleles are enriched for different set of functions within the SP and PL. When compared to the rest of the expressed genes, the SP specific polymorphic alleles were enriched in functions were in total 81 ($p < 0.05$). The top three functions were "anthocyanin-containing compound biosynthesis", "secondary metabolic process" and "proteasome-mediated ubiquitin-dependent processes" (Appendix Table 6). In contrast, the PL specific polymorphic alleles were enriched in 125 functions ($p < 0.05$) and a lot of them were related to response to stress and photosystem reactions (Appendix Table 7). The top three functions were "translation", "pentose-phosphate shunt" and "rRNA processing". It is notable though, that there were only 4 genes that

had overexpressed the allele from one population. This was an unexpected aspect, which requires further investigation.

3.3.3 Evidence for impact of positive selection and adaptive potential of the ASE genes

The next question I wanted to explore with this preliminary dataset, was the impact of positive selection on genes with ASE. In fact, when I investigate the genetic diversity and association with sweep regions in the genomic regions of the candidate ASE genes, there is evidence for genetic divergence and selection. I tested for signals of population divergence and selection in the SP and PL genomes for the ASE genes in contrast to the non ASE but expressed genes. The percentage of ASE genes (26.8%) accumulating within a sweep region is similar to the rate of the non ASE genes (27.1%). However, the population specific polymorphic sites are present in higher percentages within sweep regions than the non ASE, with 29.3% and 31.1% accumulating within the sweep regions, respectively. Furthermore, the distribution of F_{ST} values were significantly different (KS test $p = 0.0214$), with median values of 0.325 and 0.334, for ASE and non ASE genes respectively. Similarly, the distributions of the dxy values were significantly different (KS test $p = 6.684e\text{-}10$), with median values of 0.0101 and 0.009 for ASE and non ASE genes respectively. Thus, the ASE genes accumulate more genetic differences between the populations. Furthermore, they are located in areas with higher level of nucleotide diversity within populations. The ASE genes had in SP a median $\pi = 0.0044$, which was significantly shifted towards higher values than the non ASE genes ($\pi = 0.004$; KS test = 5.821e-08). Similarly, the distribution was significantly (KS test p=2.453e-13) shifted towards higher values in PL, with $\pi = 0.0079$ and $\pi = 0.007$ for ASE and non ASE genes (Figure 15).

The Tajima's D values for the ASE genes was higher in both populations, indicating that genetic drift have influenced the presence of alleles capable to cause ASE. Both within SP (KS test p = 0.035) and PL (KS test $p = 0.0003$), the ASE genes had higher values of Tajima's D than the non ASE. Specifically, the median values for ASE where 1.142 in SP and 0.1545 in PL. The distributions for non ASE had median values of 1.037 and 0.071, respectively.

Finally, I explored the adaptive potential of the genes accumulating *cis* acting changes. The ASE genes have more heritable genetic variance and less dominance variance than the non ASE genes. The distribution of V_A values for the ASE genes was significantly higher (KS test $p < 2.2e\text{-}16$) with median of 0.182 and 0.156 for ASE and non ASE genes, respectively. On the contrary, the

distribution of V_D was significantly lower (KS test p < 2.2e-16), with median values 0.334 and 0.359 for ASE and non ASE.

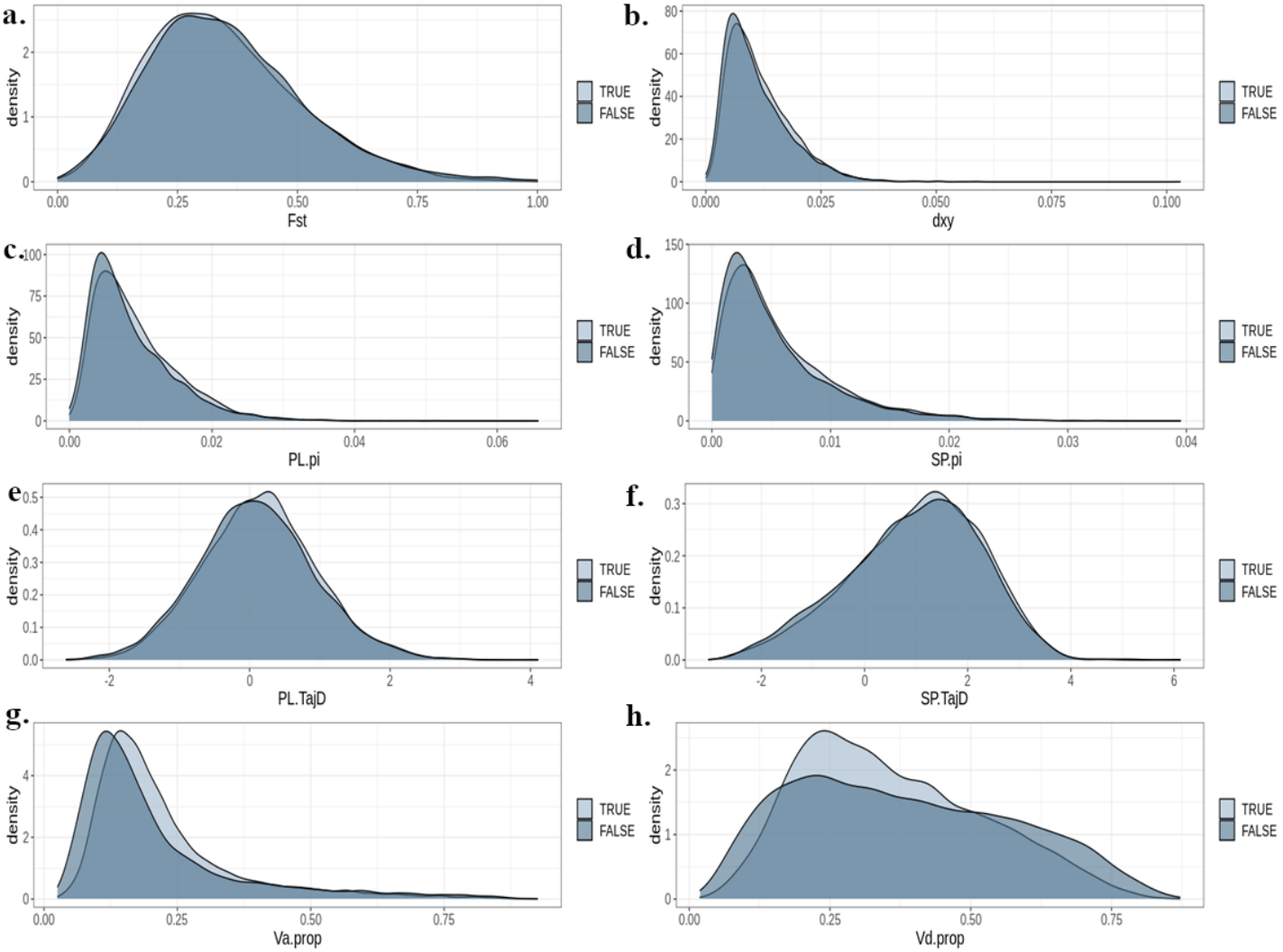


Figure 15: Distributions of population genetic parameters and genetic variances for ASE genes (TRUE) and non ASE genes (FALSE). **a.** Distribution of Fst values for SP vs PL **b.** Distributions of dxy values for SP vs PL. **c.** Distributions of π values in PL and **d.** SP. **e.** Tajima's D values for PL and f. SP. **g.** Distribution of additive and h. dominance variance between ASE and non ASE genes.

4. Discussion

4.1 Genomic burden detectable in range edge population, but no evidence of impaired fitness

The relationship between population size and selection is a centerpiece of population genetics theory. At equilibrium, smaller populations have a lower adaptive potential and increased burden of deleterious alleles (Kimura et al. 1963). These premises formed a viewpoint that population bottlenecks inhibit the removal of deleterious mutations (Kirkpatrick and Jarne 2000; Hamilton 2009; Glémin and Ronfort 2013; Balick et al. 2015). In reality however, it takes time until the equilibrium between gain and loss of mutations is restored in a bottlenecked population, so that population size reduction does not immediately associate with the presence of an increased mutation burden (Simons et al. 2014; Do et al. 2015).

The SP population provides a clear case of a range-edge population likely exposed to a severe bottleneck but with only a mild increase in average burden of deleterious mutations. Demographic modeling estimated that the population progressively decreased to about 4.8% of its initial size, despite the population growth estimated in recent generations. In agreement with previous reports (Mattila et al. 2017; Hämälä and Savolainen 2018), this decrease had pronounced population genetics consequences: a markedly lower level of diversity, a slower LD decay and non-synonymous variants segregating at higher frequency. The genome-wide elevation of Tajima's *D* further indicates that the population has not yet returned to equilibrium, since it is still depleted in rare alleles relative to common ones.

Significant mutation load has been associated to post-glacial expansion in several instances, where expansion occurred along with a mating system shift. Individuals of the sister sub-species *A. l. ssp. lyrata* showed a marked increase in phenotypic load at the range edge, particularly in populations that shifted to selfing (Willi et al. 2018). In *Arabis alpina,* individuals sampled in a selfing population of the species Northern European range also appeared to have accumulated a load of deleterious mutations greater than that of populations closer to the range-core (Laenen et al. 2018). Here, we investigated the footprint left by post-glacial range expansion in populations that did not experience a shift in mating system.

To measure the per individual genomic burden of deleterious variation we calculated the number of derived non-synonymous mutations in individual genomes. This metric has the considerable advantage that it is insensitive to variation in population size (Simons et al. 2014; Do et al. 2015) and we verified it is not influenced by the presence of selective sweep areas (data not shown). Other metrics, such as those which use the proportion of variation that is non-synonymous are confounded by demographic history (Do et al. 2015; Brandvain and Wright 2016; Simons and Sella 2016; Koch and Novembre 2017).

In the range-edge population of *A. lyrata*, prediction based on the estimated DFE indicated that the differences in the demographic histories of the two populations had a strong effect on the frequency of the mutations contributing to the per-individual burden. In SP, fixed mutations contributed comparatively more to the individual per-genome burden, whereas in PL, it was sustained by a greater number of low frequency mutations. Overall, our model predicted only an average excess of 185 non-synonymous mutations per diploid genome in SP. This prediction was within an order of magnitude of the excess non-synonymous burden of about 1,200 observed in the data. The predicted burden may be less than the observed for a number of reasons. It is possible that the SP population evolved a greater number of adaptive substitutions when expanding its range into a new environment. Since our predictions assumed free recombination, it is indeed possible that linkage with adaptive variants could have caused the faster accumulation of a burden (Marsden et al. 2016), along with adaptive non-synonymous variants themselves being potentially mischaracterized as deleterious. However, three elements suggest that linked selection will not have a strong impact on our predictions. First, the estimate of per-individual burden obtained after excluding regions carrying sweep signatures was similar (see methods). Second, the increased accumulation of deleterious mutation in the range-edge population is caused by nearly-neutral variants that become effectively neutral in the bottlenecked population, and the rate of fixation of neutral mutations is not expected to be affected by linked selection (Birky and Walsh 1988). Third, linked selection tends to distort allelic distribution in very large samples, because they mostly affect the low and high frequency ends of the spectrum, (Cvijović et al. 2018). The effect of linked purifying selection is therefore unlikely to be important with our limited sample sizes. We note, however, that the population bottleneck could have been underestimated, if we overcorrected for the reduced power to call variants due to the somewhat lower coverage of the range-edge population. This would indeed lead to an underestimation of the burden.

This number of deleterious mutations per individual genome, however, remains a crude estimator. First, it underestimates the impact of recessive deleterious mutations (Balick et al. 2015). The strong deficit of homozygous large effect mutations within SP and PL clearly shows that recessive deleterious variants do contribute to the load in these populations. We suspect, based on the very crude impression given by the F_{IS} analysis, that these variants may be more frequent in SP, in which mutations of strong deleterious effect tended to segregate at higher frequency. This would be in line with previous estimates, which show that more deleterious mutations are recessive in the species in comparison to the sister species *A.thaliana* (Huber et al. 2018). This data will be useful for theoretical studies investigating how bottlenecks impact the recessive loads. Second, indirect methods may be more powerful. For example, patterns of Neanderthal introgression in the modern human genome, revealed the increased deleterious load of the introgressing genome and its preferential removal in the larger *Homo sapiens* population (Juric et al. 2016). In maize, an outcrossing crop, which experienced two successive drastic bottlenecks during domestication, the variance in gene expression revealed a burden of deleterious regulatory mutations that significantly impaired fitness (Kremling et al. 2018).

The accumulated effect of deleterious mutations in the genome is expected to negatively impact any polygenic fitness trait, such as e.g. growth rate in plants (Leinonen et al. 2009; Debieu et al. 2013; Younginger et al. 2017). Our analysis indicated that the predicted effect of deleterious mutations is around $1.2.10^{-6}$ and therefore too small to lead to a detectable decrease in fitness. The lack of growth and survival difference observed in common gardens within the range-core area of the species both here and in a previous study, also support the notion that SP individuals do not suffer from a massive deleterious burden (Leinonen et al. 2009). Furthermore, seedlings of SP grow slower, but also larger than PL individuals, despite the presence of burden in the parental lines. The presence of the burden can be detected by the observed heterosis of the between populations hybrids; the crosses would have masked the effect of any strongly deleterious mutations of the parental lines, enabled them to perform better than their parents (Falconer and Mackay 1996). Even though, the hybrid plants grow faster than PL or SP, they do not grow bigger, indicating that the burden did not have a strong impact on SP. Our results therefore indicate that in this plant system, the severity and duration of the bottleneck experienced at the range-edge was not sufficient to allow the emergence of an impactful load of deleterious mutations. In this sense, the accumulated deleterious burden in SP is more similar to the consequences of the out-of-Africa

bottleneck in humans, which has had substantial effects on the SFS of deleterious variation, but no detectable effect on the genetic load (Simons et al. 2014; Do et al. 2015).

4.2 Absence of a bottleneck signature at the self-incompatibility locus

The S-locus diversity, both in terms of allelic richness and heterozygosity, was found to be only marginally lower in SP compared to PL and AUS. Similar levels of S-allele diversity were also reported for 12 Icelandic *A. lyrata* ssp *petraea* populations (Schierup et al. 2008), that share recent history with SP (Pyhäjärvi et al. 2012). This, together with the observation that homozygote genotypes are not more frequent throughout the genome, confirms that SP has maintained a functional self-incompatibility system, despite the historical genetic bottleneck. The persistence of obligate outcrossing in Scandinavian *A. l.* ssp. *petraea* populations has previously been discussed by Sletvold et al. (2013). Several North American populations of *A. lyrata* ssp. *Lyrata*, in contrast, have shifted to predominant selfing at the species distribution edges (Mable et al. 2005; Griffin and Willi 2014). Low inbreeding depression (Willi et al. 2013) along with a reduced diversity of S-alleles (Mable et al. 2017) may have contributed to parallel breakdowns of self-incompatibility in these bottlenecked populations, as predicted by theory (Brom et al. 2020). Accordingly, loss of self-incompatibility has been frequently reported after range expansion or strong genetic bottlenecks [e.g. in *Arabis alpina* (Laenen et al. 2018), *Leavenworthia alabamica* (Busch et al. 2011) or *Capsella rubella* (Guo et al. 2009)]. Our result illustrates the remarkable power of negative frequency-dependent selection acting on the S-locus at promoting a high level of resilience against the effect of a bottleneck on allelic diversity. Similar results were found in *L. alabamica*, were the authors did not find reduced S-allele diversity or mate limitation in outcrossing populations from small patches as compared to large populations (Busch 2005). Even if allelic diversity could have been reduced at the time of bottleneck in Scandinavian populations of *A. lyrata*, theory predicts that negative frequency-dependent selection promotes higher effective migration rates at the S-locus as compared to control loci (Schierup 1998), suggesting that high allelic diversity could have also been restored subsequently by gene flow.

4.3 Adaptive dynamics maintained in SP

Small size populations are also expected to require larger effect mutations to adapt, although these mutations are rare (Hamilton 2009). Whether a population size reduction immediately reduces adaptive evolution is, however, a complex question in the context of range expansion (Gilbert et al. 2017). If populations have to adapt locally at the range edge, the rate of geographical expansion

slows down, along with the severity of the expansion bottleneck (Gilbert et al. 2017). A decrease in population size, however, increases the range of beneficial alleles that behave effectively neutrally (Lynch 2007). Searching for signals of selective sweeps in SP, after accounting for its demography, we identified 327 regions that formed outlier for both CLR and F_{ST} statistics. In fact, the number of genomic regions displaying a signature of positive selection was greater in SP than in PL, a pattern that has been observed also in Northern Swedish populations of *A. thaliana* in contrast to Southern Swedish populations (Huber et al. 2014). However, we cannot exclude that some of the signal detected in SP could also result from the surfing of new alleles towards the range margin, which can mimic signatures of adaptive evolution and create false positive signatures of adaptation (Excoffier et al. 2009). In addition, some of the selective sweep signatures could be caused by background selection, although theoretical work indicates that genetic signatures of selective sweeps and adaptive divergence are unlikely to be mimicked by background selection (Lynch 2007; Matthey-Doret and Whitlock 2019; Schrider 2020). Adaptive dynamic therefore appear to be maintained despite evidence for a slight reduction in the ability of the SP population to purge deleterious alleles. This agrees with basic population genetics theory showing that the fixation probability of deleterious mutation is much more sensitive to changes in population size than that of deleterious alleles (Kimura 1964; Otto and Whitlock 1997).

Functional enrichments among regions displaying signatures of local positive selection, however, indicate the presence of true positive signals. Within those regions, functions involved in the response to stress were enriched, in agreement with a previous study investigating micro-geographical patterns of local adaptation in Norwegian populations connected by gene flow (Hämälä and Savolainen 2018; Hämälä and Savolainen 2019). We also found a significant enrichment in genes involved in light perception, a function enriched in loci differentiating the SP population from a close-by population of lower elevation (Hämälä and Savolainen 2019). Furthermore, the F_{ST} distribution of genes related to development was significantly shifted towards higher values, suggesting polygenic selection on alleles associated to this function (Foll et al. 2014; Daub et al. 2015; Stephan 2016). Previous work has documented that Scandinavian populations display differences in several traits related to growth and resource allocation, including plant size, inflorescence production and fruit production (Quilot-Turion et al. 2013; Hämälä et al. 2018). Both local and regional reciprocal transplant experiments have revealed local adaptation in this species via life history traits and growth-related phenotypes (Leinonen et al. 2009; Hämälä et al. 2018).

This shows that adaptive dynamics are ongoing also at smaller geographical scale in this system and is consistent with the broad genomic signals of positive selection we observed.

4.4 The importance of understanding the adaptive potential of gene expression variance

In plant and animal breeding research, it has been vital to estimate the amount of additive and non-additive genetic variance, to predict the response of selection. Especially, the additive variance is important, as it describes the heritable proportion of the genetic variance and it is therefore intertwined with selection (Falconer and Mackay 1996; Lynch and Walsh 1998). For evolutionary biology, it is also of great interest to identify the additive genetic variance, as it can be a great tool to understand the complex traits' response to selection in natural populations (Fisher 1918). Life history traits are closely associated with fitness and for that reason studies have tried to identify the relation of additive variance and selection. The life history traits are expected to have been under strong selection, and thus the amount of additive variance will be reduced. As expected, in a lot of wild populations, these traits do not have as high additive genetic variance as for instance traits related to physiology (Mousseau and Roff 1987; Shaw and Shaw 2014).

One of the greatest challenges of understanding the genetic variance of phenotypes in the wild, has been the limited number of phenotypes that we can sample (Shaw and Shaw 2014). I used the transcriptome of inter- population crossings to expand the range of studied phenotypes. Nowadays, it is possible to cost effectively sequence the whole transcriptome, increasing exponentially the number of phenotypes we can analyze. Moreover, gene expression is a phenotype of great importance as it directly influences other organismal phenotypes, even fitness. Despite the evolutionary importance of gene expression (Oleksiak et al. 2002; Fay and Wittkopp 2008; Romero et al. 2012; He et al. 2016), there has not been an extensive investigation of the genetic basis of its variance. Recently, an investigation of the relationship between selection strength and genetic variance of gene expression in rice was done (Groen et al. 2020). The authors estimated the strength and direction of selection by correlating the populations fecundity with gene expression under both stressed and control conditions. Stronger signals of selection were associated with the genetic variance of gene expression in stressed conditions than under control conditions. While the authors have identified an important connection between the two, the study confounds the dominance, additive and epistatic effects on gene expression variance. The importance of the ignoring the difference between additive and genetic variance could be derived by looking at another study of gene expression variance in natural populations of sticklebacks (Leder et al. 2015). Using

microarray data, the authors have successfully demonstrated that gene expression variance has substantial dominance variance, which is moreover linked to environmental fluctuations. Furthermore, they were able to associate signals of directional selection with differential gene expression, despite the limitations of microarray assays and the limitations imposed by microsatellite datasets.

I used a crossing design of full and half siblings to partition the genetic variance of gene expression in *A. lyrata* to its components. The relationship between the individual plants allowed the partition of genetic variance to both additive and dominance variance. Moreover, I aimed to have an extensive phenotypic dataset by sequencing the whole transcriptome of each of the 131 samples. As a result, I analyzed the gene expression variance of approximately 17,500 genes. Broad sense heritability, which is the proportion of genetic variance out of the total phenotypic variance, was more than half of the phenotypic variance for approximately 67% of the transcriptome. This indicates that, within *A. lyrata,* gene expression has a largely genetic basis. However, the amount of the actual heritable genetic variance, expressed as additive variance, is low. The median additive variance is approximately 20% of the phenotypic variance and only around 6% of the genes have additive variance more than half of the total phenotypic. It is noteworthy that the growth rate dominance and additive variance of the plants (V_D = 0.62 and V_A=0.005) is similar to the genes with the genes that have more than 50% of the phenotypic variance dominance variance. Nevertheless, the amount of additive variance observed in the transcriptome is comparable to what has been found in sticklebacks and human blood tissues with additive variance of approximately 20% of the total phenotypic variance (Price et al. 2011; Leder et al. 2015). even though a large part is controlled by within and between loci interactions, these results indicate that gene expression variance between *A. lyrata* populations have adaptive potential.

4.5 Evolution of gene expression variance has been shaped by directional selection

Directional selection has shaped the gene expression evolution within the species. Genes with high level of residual variance, and thus low levels of genetic variance, were shown to be under stronger constraints than a control group of randomly selected genes. On the contrary, genes with high level of additive variance, are under relaxed selection. Therefore, a substantial amount of genes' expression along the transcriptome has been under purifying selection. This result is in line with the expectation of Fisher's fundamental theorem of natural selection, according to which, selection on a specific phenotype will act on the additive variance by decreasing its proportion out of the total phenotypic variance. At the same time the genetic variance will also decrease (Fisher 1918;

Mousseau and Roff 1987; Orr 2009). Gene expression is a phenotype that influences fitness and its changes can even be connected to severe diseases manifesting in human populations (Emilsson et al. 2008; Cookson et al. 2009; Cooper-Knock et al. 2012). Thus, it is expected that deleterious mutations will be exposed to natural selection and they will be prevented from accumulating. While, this is especially true for mutations with large effect on the phenotype (Charlesworth 2012), natural selection can also act when the mutations are nearly neutral within a population (Bedford and Hartl 2009). Gene expression divergence is often under stabilizing selection, especially when species are compared (Whitehead and Crawford 2006; Bedford and Hartl 2009; Hodgins-Davis et al. 2015; Metzger et al. 2017). However, signals of purifying selection are not uncommon; regulatory elements have been found to be under different strengths of purifying selection, as for instance in *Capsella grandiflora* and *Theobroma cacao* (Fay and Wittkopp 2008; Steige et al. 2015; Hämälä et al. 2020).

The different types of genetic variance accumulate in different functions, and therefore so are the associating signals of selection. A similar observation was reported lately for gene co-expression modules in cacao, in which specific functional modules are under negative selection (Hämälä et al. 2020). Similarly, in *Capsella grandiflora*, not only the *cis* acting changes are under weaker purifying selection (Steige et al. 2017), but also the intensity of purifying selection depends on the level of interconnection between the genes (Josephs et al. 2017). Gene expression levels of wild populations are not always optimal for high fitness in certain environments (Keren et al. 2016), and, based on theoretical models, the selection strength will be dependent, firstly, on how far away from the optimum is the gene expression and also on its location in a gene network (Hurst and Randerson 2000; Huber et al. 2018). Therefore, it can be expected that the type and intensity of selection we observe within a species is not uniform across the transcriptome. In fact, in *A. lyrata*, I identified genes with higher level of additive variance, accumulate in functional groups that partially overlap with the enriched functions located within selective sweep areas. This observation is important as it shows that there is still potential for adaptation via modulating gene expression variance. However, this is a small number of genes that have the potential to contribute to selection. One consideration, to be taken into account is the impact of the bottleneck and the subsequent genetic drift in both SP and PL. However, in a similar crossing design between yeast populations, adaptive divergence was found to have played a role in promoting additive variance in admixed populations (Liu et al. 2020). Adaptive divergence between our populations has influenced part of

the gene expression evolution, as genes located within sweep areas have decreased additive variance.

The selection strength is affected by the number of loci that control gene expression. (Hodgins-Davis et al. 2015). Selection, especially if it is positive, is usually harder to identify when a trait has polygenic basis (Stephan 2016), with the extreme case of the infinitesimal model attributing all genetic variance to genetic drift (Barton et al. 2017). More plant and animal traits have a polygenic basis than previously thought. In plants, vegetative growth (Wieters et al. 2020), and stomata size and density (Dittberner et al. 2018) are some examples of traits with polygenic basis. Perhaps, one might expect gene expression to be mostly regulated by large effect mutations located within its own sequence. I observed the opposite; of the approximately 17,500 gene wide association studies, only 174 genes had significant associations. This is a clear indication that the gene expression's genetic basis is polygenic. Interestingly, is that genes regulated by *trans* genes have a higher additive variance than the rest of the genes and additionally they are more often under positive selection. Those genes have a large effect on the phenotype and therefore the selective sweeps are easier to identify than if they were mutations of small effect (Pritchard et al. 2010; Stephan 2016; Barghi et al. 2020). The large number of phenotypes with small effect mutations regulating gene expression, in combination with the evidence of selection for a large number of genes, indicates that a large proportion of the *A. lyrata* transcriptome is under polygenic selection. Note, however, that this outcome is not true for the small number of genes that have little genetic variance. This observation is in line with the overall signals of polygenic adaptation observed in genomes of the two populations.

4.6 Investigating the dominance variance can further enrich our understanding of the missing heritability

One aspect of polygenic traits that has arisen in the recent years, is the problem known as the mystery of the missing heritability, which refers to the phenomenon reported by genome wide association studies. Genome wide association studies can be used to infer the genetic variance of a trait in addition to its genetic architecture (Visscher et al. 2006). Often, the cumulative effect of the most important loci that control the tested phenotype explain only a small proportion of the predicted genetic variance (Manolio et al. 2009; Boyle et al. 2017). The missing heritability is attributed to rare alleles, which, due to their frequency, they are often not represented in the datasets (Simons et al. 2014; Marouli et al. 2017). However, often investigating the narrow sense

heritability alone does not provide the whole picture, due to the within and between locus allelic interactions. In my study, for example, there was a discrepancy between the values of genetic and additive variance; the difference between the two can be attributed to dominance variance. More than 25% of the transcripts had dominance variance greater than 50% of the total phenotypic variance. Moreover, in almost all cases the importance of including the dominance variance as a predictive variable in the model was significant. Dominance variance is often neglected from studies because of the assumption that it will be minimal under a demographic model of mutation balance and infinite population size. However, dominance variance has been shown to be a considerable part of phenotypic variance in wild (Waldmann 2001) and domesticated populations (Yang et al. 2019). Moreover, ignoring the contribution of dominance variance can lead to overestimation of narrow sense heritability as it is seen both by theoretical models (Ovaskainen et al. 2008) and field population analysis (Class and Brommer, 2020). Class and Brommer used a time series dataset of bird families to investigate the level of additive and non-additive variance within the species. Moreover, using model fits including and excluding the dominance matrix, as well as phenotypic modeling, they reached the conclusion that dominance variance is important to accurately explain the phenotypic variance observed in the wild. Therefore, partitioning genetic variance to both additive and dominance variance can provide insights into the part that is the result of the interactions. The interactions between loci and genes are not detectable via genome wide association study, due to the per SNP basis correlations (Korte and Farlow 2013). However, since the quantitative genetics approach is blind to the genetic architecture of the trait; it can provide insight into the mystery of missing heritability.

4.7 Dominance variance levels within the transcriptome support the omnigenic model due to pleiotropic effects and genic interactions

The accumulation of dominance variance observed in this study, align with the omnigenic model prediction. The omnigenic model predicts that gene regulatory networks are sufficiently interconnected, in such a way that they all can affect the phenotype (Boyle et al. 2017). I have identified three aspects of the genes with high dominance variance that support the model. First of all, genes with high dominance variance tend to accumulate in functions that are related to epigenetic functions such as chromatin silencing and DNA and protein methylation, which affect many genes (Robertson 2005; Choi et al. 2008). Secondly, most of the genes with high dominance variance

show evidence of network connectivity, as they are often clustered together based on their expression correlation. Network connectivity is a measure of pleiotropy (Langfelder and Horvath 2008), which is assumed to exist in the omnigenic model. Thirdly, the correlation of the high dominance variance and the number of transcription factors is further evidence of the pleiotropic and epigenic role of those genes. Hence, the polygenic basic of the transcriptome and the gene interconnection, as they are evident by analyzing the additive and dominance variance respectively, provide support for the omnigenic model in gene expression.

However, there is an interesting aspect that has arisen regarding dominance variance. In different species there has been evidence of strong constraints and negative selection related to pleiotropy in gene networks or co-expressed genes (see for example Hahn and Kern 2005; Orr 2009; Josephs et al. 2017; Masalia et al. 2017), including the species *A.lyrata ssp petraea* (Huber et al. 2018). The genes with high dominance variance within *A.lyrata* show the same level of constraints as a random set of genes, indicating no relation to negative selection. Furthermore, the genes do not show correlation with population genetic parameters indicative of divergence between the two populations in the study. This observation could have been due to the inevitable cofounding of within locus and between locus interactions. A proportion of the observed dominance is due to the within locus interactions. I have identified a significant association of specific genome architecture, such as gene length and exon number and dominance variance. This can be connected with the definition of dominance genetic variance (within locus interactions) and specifically, dominance deviation, as described in Falconer & MacCay (1996). Dominance deviation is the difference of the breeding value (or additive variance) from the genotypic value (or genetic variance), and it represents the interactions between the alleles within or across loci. Thus, increasing the number of potential alleles in a locus can have an impact on the amount of dominance variance.

One way, to disentangle the dominance variance from the epistatic variance is to partition the phenotypic variance in the next generation (Lynch and Walsh 1998) and then look for signals of selection in the two groups separately. An additional gene pathway analysis would be useful to investigate constraints between the genes that belong in the same pathway.

4.8 Considerations arising by the unique study design

One unique aspect of the study design has to be considered; I have crossed populations, while most of the studies are investigating the genetic variance within populations (see for example Leder et

al. 2015; Koch and Guillaume 2020). My design, even though it allows for investigation of the interaction of an increased number of alleles and gene structures, at the same time, it masks the effect of some alleles. The effect of the within population fixed alleles is lost, while the alleles that are in fixed heterozygous state within population, change in frequency in the next generation. Especially, the impact of fixed alleles is very important, as for traits with simple genetic architecture, selection acts on the allele frequencies, leading to fixation of advantageous alleles within a population and thus the reduction of additive variance (Hill et al. 2008). On the contrary, when genetic drift is the primary force affecting the allelic frequencies, as for example after a strong bottleneck, then additive variance increases (Crnokrak and Roff 1995; Hill et al. 2008). This indicates that the degree of additive variance is underestimated in the design.

The impact of the crossing design on dominance variance, on the other hand is more complicated to disentangle than the impact on additive variance. The proportion of dominance variance due to within locus effects should be represented fairly, except for the cases that a recessive allele is fixed within one population (Falconer and Mackay 1996; Lynch and Walsh 1998). When considering the impact of genetic load in the two populations, those cases would be rare, if they exist at all. It has also been shown that in yeast strongly deleterious mutations are additive (Agrawal and Whitlock 2011) and thus they would not contribute to dominance variance in a disproportional way.

Even though the type of allelic interactions and their frequencies within populations does not inflate the dominance variance within the design, I cannot readily exclude the inflation of dominance variance due to genome-wide incompatibilities. Genome wide incompatibilities can occur by merging two divergent genomes and lead to miss regulation of transcriptome (Lafon-Placette and Köhler 2015; Todesco et al. 2016). This does not seem to be the case here, as there was no large genomic area associating with the gene expression of most genes, which would have been the signal of genome incompatibility. On the contrary, the location of *trans* effects were dispersed along the genome, indicating that we do not have a misinterpretation of the dominance variance values in the design. One aspect of the design that would still need to be explored is a possible inflation of the dominance variance due to the private alleles. Each population have accumulated a number of unique alleles since their divergence, which could increase the number of interactions within and between loci, inflating dominance variance. Simulations of phenotypes based on interaction of private alleles could help solve this problem.

4.9 Indications of the important role that *cis* regulatory elements have in the evolution within species

It has been previously documented, that the *cis* regulatory elements have had a significant role in species divergent adaptation (He et al. 2012; He et al. 2016; Steige et al. 2017). The *cis* regulatory elements are used to understand the impact of small effect mutations on the phenotype (Fay and Wittkopp 2008; He et al. 2016; de Meaux 2018) and they have an additive nature of their allelic interactions, which generally contribute more to the additive genetic variance than alleles with non-additive interactions (Falconer and Mackay 1996). Using the GWAS of each gene's expression, I was able to identify mutations with large effect for approximately 1% of the transcriptome. Interestingly, almost all of the significant associations have a *trans* effect, leading to the conclusion that most of the gene expression variance has polygenic basis. Thus, it would be interesting to identify those elements with *cis* regulatory effect on gene expression, as they can add to the understanding of polygenic selection via accumulation of small effect mutations.

However, using the allele specific expression within the generated transcriptomes poses a problem that is yet to be overcome. There is evidence of mapping bias, with alleles from SP population often being identified as overexpressed more often than the alleles from PL. Mapping bias can distort the accurate detection of ASE and *cis* elements (Degner et al. 2009) and it could stem from mapping errors in areas with high divergence both between populations and the reference genome. Further investigation into the causes of it, can provide solutions for improved filtering of the ASE dataset.

Despite this fact, it is evident that even within species, the *cis* elements tend to accumulate in different functions between the populations. Also, it indicates that there is potential for adaptation via accumulation of small effect mutations, as genes with ASE seem to have higher additive variance than the genes with alleles expressed in equal proportion. Taken together, these observations illustrate the potential of studying the accumulation of *cis* regulatory variation within species.

4.10 Concluding remarks

The aim of this thesis was to investigate the polygenic basis of adaptation in the European subspecies of *A. lyrata*, using two populations from the edge and the core of its distribution. Firstly, I identified that, despite a strong bottleneck, the range-edge population has accumulated a moderate genomic load, which has neither compromised plant fitness nor the adaptive dynamics. This result

highlights the importance of studying populations with natural variation to better understand the impact of selection and genomic load.

Furthermore, I established that the gene expression variance within the species has substantial additive variance, which is related to signals of directional selection in the two populations. Gene expression in most of the cases is not a trait with simple genetic architecture, but with non-additive, polygenic basis. I identified evidence that the genetic variance of the trait is also highly dependent on pleiotropic effects and also the gene's structure. Taken together, these results support with evidence the novel omnigenic model, which aims to explain the complex gene interactions of complex phenotypes. Finally, the exploration of the *cis* regulatory variation indicates that small effects mutations have had an important role in adaptation of the two populations.

5. References

1001 Genomes Consortium. 2016. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell

Agrawal AF, Whitlock MC. 2011. Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* 187:553–566.

Ågren J, Oakley CG, Lundemo S, Schemske DW. 2017. Adaptive divergence in flowering time among natural populations of Arabidopsis thaliana: Estimates of selection and QTL mapping. *Evolution* 71:550–564.

Ågren J, Schemske DW. 2012. Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. *New Phytologist* 194:1112–1122.

Akiyama R, Ågren J. 2014. Conflicting selection on the timing of germination in a natural population of Arabidopsis thaliana. *Journal of Evolutionary Biology* 27:193–199.

Alexa A, Rahnenfuhrer J. 2016. topGO: Enrichment Analysis for Gene Ontology. R package version 2.30.1.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.

Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169.

Angiuoli SV, Salzberg SL. 2011. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* 27:334–342.

Ansell SW, Stenøien HK, Grundmann M, Schneider H, Hemp A, Bauer N, Russell SJ, Vogel JC. 2010. Population structure and historical biogeography of European Arabidopsis lyrata. *Heredity* 105:543–553.

Austen EJ, Rowe L, Stinchcombe F JR, J.R.K. 2017. Explaining the apparent paradox of persistent selection for early flowering. New Phytol

Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H. 1997. Evolution of Coalescence Times, Genetic Diversity and Structure during Colonization.

Baker HG. 1995. Self-Compatibility and Establishment After “Long-Distance” Dispersal. 9:347–349.

Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR. 2015. Dominance of Deleterious Alleles Controls the Response to a Population Bottleneck. *PLOS Genetics* 11:e1005436.

Barghi N, Hermisson J, Schlötterer C. 2020. Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics*:1–13.

Barrett SCH. 2003. Mating strategies in flowering plants: the outcrossing–selfing paradigm and beyond.Dickinson HG, Hiscock SJ, Crane PR, editors. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358:991–1004.

Barton NH, Etheridge AM, Véber A. 2017. The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology* 118:50–73.

Bateman AJ. 1955. Self-incompatibility systems in angiosperms: III. Cruciferae. *Heredity* 9:53–68.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67:1–48.

Bedford T, Hartl DL. 2009. Optimization of gene expression by natural selection. *PNAS* 106:1133–1138.

Berg JJ, Coop G. 2014. A Population Genetic Signal of Polygenic Adaptation. *PLOS Genetics* 10:e1004412.

Birky CW, Walsh JB. 1988. Effects of linkage on rates of molecular evolution. *PNAS* 85:6414–6418.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, et al. 2008. Assessing the Evolutionary Impact of Amino Acid Mutations in the Human Genome. *PLoS Genet* [Internet] 4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2377339/

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

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F. 2010. Linkage and association mapping of Arabidopsis thaliana flowering time in nature. *PLoS Genetics* 6:1000940.

Brandvain Y, Wright SI. 2016. The Limits of Natural Selection in a Nonequilibrium World. *Trends in Genetics* 32:201–210.

Brom T, Castric V, Billiard S. 2020. Breakdown of gametophytic self-incompatibility in subdivided populations. *Evolution* 74:270–282.

Burghardt LT, Metcalf C, Wilczek AM, Schmitt J. 2015. Modeling the Influence of Genetic and Environmental Variation on the Expression of Plant Life Cycles across Landscapes. The American Naturalist

Busch JW. 2005. Inbreeding depression in self-incompatible and self-compatible populations of Leavenworthia alabamica. *Heredity* 94:159–165.

Busch JW, Joly S, Schoen DJ. 2011. Demographic Signatures Accompanying the Evolution of Selfing in Leavenworthia alabamica. *Molecular Biology and Evolution* 28:1717–1729.

Charlesworth B. 2012. The Effects of Deleterious Mutations on Evolution at Linked Sites. *Genetics* 190:5–22.

Cheptou P-O. 2012. Clarifying Baker's Law. *Ann Bot* 109:633–641.

Chiang GCK, Barua D, Dittmar E, Kramer EM, Casas RR, Donohue K. 2013. Pleiotropy in the wild: the dormancy gene DOG1 exerts cascading control on life cycles. *Evolution* 67:883–893.

Choi JK, Hwang S, Kim Y-J. 2008. Stochastic and Regulatory Role of Chromatin Silencing in Genomic Response to Environmental Changes. *PLOS ONE* 3:e3002.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* 6:80–92.

Class B, Brommer JE. 2020. Can dominance genetic variance be ignored in evolutionary quantitative genetic analyses of wild populations? *Evolution* doi/abs/10.1111/evo.14034

Clauss MJ, Koch MA. 2006. Poorly known relatives of Arabidopsis thaliana. *Trends in Plant Science* 11:449–459.

Clauss MJ, Mitchell-Olds T. 2006. Population genetic structure of Arabidopsis lyrata in Europe. *Molecular Ecology* 15:2753–2766.

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

Collins S, Meaux JD. 2009. Adaptation to Different Rates of Environmental Change in Chlamydomonas. *Evolution* 63:2952–2965.

Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. 2009. Mapping complex disease traits with global gene expression. *Nature Reviews Genetics* 10:184–194.

Cooper-Knock J, Kirby J, Ferraiuolo L, Heath PR, Rattray M, Shaw PJ. 2012. Gene expression profiling in human neurodegenerative disease. *Nature Reviews Neurology* 8:518–530.

Corre VL. 2005. Variation at two flowering time genes within and among populations of Arabidopsis thaliana: comparison with markers and traits. *Molecular Ecology* 14:4181–4192.

Corre VL, Kremer A. 1998. Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. *Journal of Evolutionary Biology* 11:495–512.

Crnokrak P, Roff DA. 1995. Dominance variance: associations with selection and fitness. *Heredity* 75:530–540.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 23:3133–3157.

Cui X, Affourtit J, Shockley KR, Woo Y, Churchill GA. 2006. Inheritance Patterns of Transcript Levels in F1 Hybrid Mice. *Genetics* 174:627–637.

Cvijović I, Good BH, Desai MM. 2018. The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics* 209:1235–1278.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.

Daub JT, Dupanloup I, Robinson-Rechavi M, Excoffier L. 2015. Inference of Evolutionary Forces Acting on Human Biological Pathways. *Genome Biology and Evolution* 7:1546–1558.

Davey MP, Palmer BG, Armitage E, Vergeer P, Kunin WE, Woodward FI, Quick WP. 2018. Natural variation in tolerance to sub-zero temperatures among populations of Arabidopsis lyrata ssp. petraea. *BMC Plant Biology* 18:277.

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:e61075.

Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y, Pritchard JK. 2009. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* 25:3207–3212.

Dittberner H, Korte A, Mettler-Altmann T, Weber A, Monroe G, Meaux J. 2018. Natural variation in stomata size contributes to the local adaptation of water-use efficiency in Arabidopsis thaliana. *Molecular Ecology* 20:4052-4065.

Do R, Balick D, Li H, Adzhubei I, Sunyaev S, Reich D. 2015. No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nature Genetics* 47:126–131.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Donohue K. 2002. Germination timing influences natural selection on life-history characters in Arabidopsis thaliana. *Ecology* 83:1006–1016.

Donohue K, Dorn D, Griffith C, Kim E, Aguilera A, Polisetty CR, Schmitt J. 2005. Niche construction through germination cueing: Life-history responses to timing of germination in Arabidopsis thaliana. *Evolution* 59:771–785.

Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu T, Burbano HA, Picó FX, Alonso-Blanco C, et al. 2017. African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. *PNAS* 114:5213–5218.

Dwyer KG, Balent MA, Nasrallah JB, Nasrallah ME. 1991. DNA sequences of self-incompatibility genes from Brassica campestris and B. oleracea: polymorphism predating speciation. *Plant Mol Biol* 16:481–486.

El Mousadik A, Petit RJ. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [Argania spinosa (L.) Skeels] endemic to Morocco. *Theoret. Appl. Genetics* 92:832–839.

Emerson JJ, Li W-H. 2010. The genetic basis of evolutionary change in gene expression levels. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:2581–2590.

Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, et al. 2008. Genetics of gene expression and its effect on disease. *Nature* 452:423–428.

Erwin DH, Davidson EH. 2009. The evolution of hierarchical gene regulatory networks. *Nature Reviews Genetics* 10:141–148.

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

Excoffier L, Foll M, Petit RJ. 2009. Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics* 40:481–501.

Falconer DS, Mackay TFC. 1996. Introduction to Quantitative Genetics. 4th ed. Essex: Prentice Hall

Falke KC, Glander S, He F, Hu J, de Meaux J, Schmitz G. 2013. The spectrum of mutations controlling complex traits and the genetics of fitness in plants. *Current Opinion in Genetics & Development* 23:665–671.

Fay JC, Wittkopp PJ. 2008. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* 100:191–199.

Fisher RA. 1918. The correlation between relatives on the supposition of mendelian inheritance. *Trans. R. Soc. Edinb* 53:399–433.

Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L. 2014. Widespread Signals of Convergent Adaptation to High Altitude in Asia and America. *The American Journal of Human Genetics* 95:394–407.

Footitt S, Huang Z, Clay HA, Mead A, Finch-Savage WE. 2013. Temperature, light and ni-trate sensing coordinate Arabidopsis seed dormancy cycling, resulting in winter and summer annual phenotypes. *The Plant Journal* 74:1003–1015.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in Arabidopsis thaliana. *Science* 334:86–89.

Fox J, Weisberg S. 2019. An R Companion to Applied Regression. Third. Sage Available from: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fraser HB, Babak T, Tsang J, Zhou Y, Zhang B, Mehrabian M, Schadt EE. 2011. Systematic Detection of Polygenic cis-Regulatory Evolution. *PLOS Genetics* 7:e1002023.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv:1207.3907 [q-bio]* [Internet]. Available from: http://arxiv.org/abs/1207.3907

Gascoigne J, Berec L, Gregory S, Courchamp F. 2009. Dangerously few liaisons: a review of mate-finding Allee effects. *Population Ecology* 51:355–372.

Gelman A, Rubin DB. 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statist. Sci.* 7:457–472.

Genete M, Castric V, Vekemans X. 2019. Genotyping and de novo discovery of allelic variants at the Brassicaceae self-incompatibility locus from short read sequencing data. *Mol Biol Evol* [37 (4): 1193-1201.

Gibbs PE. 2014. Late-acting self-incompatibility – the pariah breeding system in flowering plants. *New Phytologist* 203:717–734.

Gilad Y, Oshlack A, Rifkin SA. 2006. Natural selection on gene expression. *Trends in Genetics* 22:456–461.

Gilbert KJ, Peischl S, Excoffier L. 2018. Mutation load dynamics during environmentally-driven range shifts.Bomblies K, editor. *PLOS Genetics* 14:e1007450.

Gilbert KJ, Sharp NP, Angert AL, Conte GL, Draghi JA, Guillaume F, Hargreaves AL, Matthey-Doret R, Whitlock MC. 2017. Local Adaptation Interacts with Expansion Load during Range Expansion: Maladaptation Reduces Expansion Load. *The American Naturalist* 189:368–380.

Glémin S, Ronfort J. 2013. Adaptation and Maladaptation in Selfing and Outcrossing Species: New Mutations Versus Standing Variation. *Evolution* 67:225–240.

Goodwillie C, Kalisz S, Eckert CG. 2005. The Evolutionary Enigma of Mixed Mating Systems in Plants: Occurrence, Theoretical Explanations, and Empirical Evidence. *Annual Review of Ecology, Evolution, and Systematics* 36:47–79.

Goudet J. 2005. hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 5:184–186.

Goudet J, Jombart T. 2015. hierfstat: Estimation and Tests of Hierarchical F-statistics. R package version 0.04-22. Available from: https://CRAN.R-project.org/package=hierfstat

Griffin PC, Willi Y. 2014. Evolutionary shifts to self-fertilisation restricted to geographic range margins in North American Arabidopsis lyrata. *Ecology Letters* 17:484–490.

Groen SC, Ćalić I, Joly-Lopez Z, Platts AE, Choi JY, Natividad M, Dorph K, Mauck WM, Bracken B, Cabral CLU, et al. 2020. The strength and pattern of natural selection on gene expression in rice. *Nature*:1–5.

Grossenbacher D, Runquist RB, Goldberg EE, Brandvain Y. 2015. Geographic range size is predicted by plant mating system. *Ecology Letters* 18:706–713.

Guo Y-L, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, Weigel D, Schierup MH. 2009. Recent speciation of Capsella rubella from Capsella grandiflora, associated with loss of self-incompatibility and an extreme bottleneck. *Proceedings of the National Academy of Sciences* 106:5246–5251.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genetics* 5:e1000695.

Hadfield J, Harris SR, Seth-Smith HMB, Parmar S, Andersson P, Giffard PM, Schachter J, Moncada J, Ellison L, Vaulet MLG, et al. 2017. Comprehensive global genome dynamics of Chlamydia trachomatis show ancient diversification followed by contemporary mixing and recent lineage expansion. *Genome Res* 27:1220–1229.

Hadfield JD. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* 33:1–22.

Hahn MW, Kern AD. 2005. Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks. *Mol Biol Evol* 22:803–806.

Hallatschek O, Hersen P, Ramanathan S, Nelson DR. 2007. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences* 104:19926–19930.

Hämälä T, Guiltinan MJ, Marden JH, Maximova SN, dePamphilis CW, Tiffin P. 2020. Gene Expression Modularity Reveals Footprints of Polygenic Adaptation in Theobroma cacao. *Mol Biol Evol* 37:110–123.

Hämälä T, Mattila TM, Leinonen PH, Kuittinen H, Savolainen O. 2017. Role of seed germination in adaptation and reproductive isolation in Arabidopsis lyrata. *Molecular Ecology* 26:3484–3496.

Hämälä T, Mattila TM, Savolainen O. 2018. Local adaptation and ecological differentiation under selection, migration, and drift in Arabidopsis lyrata. *Evolution* 72:1373–1386.

Hämälä T, Savolainen O. 2018. Local adaptation under gene flow: Recombination, conditional neutrality and genetic trade-offs shape genomic patterns in. Available from: http://biorxiv.org/lookup/doi/10.1101/374900

Hämälä T, Savolainen O. 2019. Genomic patterns of local adaptation under gene flow in Arabidopsis lyrata. *Mol Biol Evol*.

Hamilton MB. 2009. Population Genetics. Chichester, UK; Hoboken, NJ: Wiley-Blackwell

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to climate across the Arabidopsis thaliana genome. *Science* 334:83–86.

Harris RS. 2007. Improved Pairwise Alignmnet of Genomic DNA. PhD thesis, Pennsylvania State University.

He F, Arce AL, Schmitz G, Koornneef M, Novikova P, Beyer A, de Meaux J. 2016. The Footprint of Polygenic Adaptation on Stress-Responsive *Cis* -Regulatory Divergence in the *Arabidopsis Genus*. *Mol Biol Evol* 33:2088–2101.

He F, Zhang X, Hu J, Turck F, Dong X, Goebel U, Borevitz J, de Meaux J. 2012. Genome-wide Analysis of Cis-regulatory Divergence between Species in the Arabidopsis Genus. *Mol Biol Evol* 29:3385–3395.

He H, Souza Vidigal D, Snoek LB, Schnabel S, Nijveen H, Hilhorst H, Bentsink L. 2014. Interaction between parental environment and genotype affects plant and seed performance in Arabidopsis. *Journal of Experimental Botany* 65:6603–6615.

He X, Houde ALS, Pitcher TE, Heath DD. 2017. Genetic architecture of gene transcription in two Atlantic salmon (Salmo salar) populations. *Heredity* 119:117–124.

Heerwaarden J van, Zanten M van, Kruijer W. 2015. Genome-Wide Association Analysis of Adaptation Using Environmentally Predicted Traits. *PLOS Genetics* 11:e1005594.

Henry RC, Barto KA, Travis JMJ. 2015. Mutation accumulation and the formation of range limits. *Biology Letters* 11:20140871–20140871.

Hepworth J, Dean C. 2015. Flowering Locus C's Lessons: Conserved Chromatin Switches Underpinning Developmental Timing and Adaptation. *Plant Physiology* 168:1237.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907–913.

Hill WG, Goddard ME, Visscher PM. 2008. Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits.Mackay TFC, editor. *PLoS Genet* 4:e1000008.

Hodgins-Davis A, Rice DP, Townsend JP. 2015. Gene Expression Evolves under a House-of-Cards Model of Stabilizing Selection. *Mol Biol Evol* 32:2130–2140.

Hoffmann MH. 2005. Evolution of the Realized Climatic Niche in the Genus: Arabidopsis (brassicaceae). *Evolution* 59:1425-1436.

Howell SH. 2013. Endoplasmic reticulum stress responses in plants. *Annual review of plant biology* 64:477–499.

Hu J, Lei L, Meaux J. 2017. Temporal fitness fluctuations in experimental Arabidopsis tha-liana populations. In: K Ingvarsson P, Ed), editors. PLoS ONE 12. p. 0178990.

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, et al. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics* 43:476–481.

Huber CD, Durvasula A, Hancock AM, Lohmueller KE. 2018. Gene expression drives the evolution of dominance. *Nature Communications* 9:2750.

Huber CD, Nordborg M, Hermisson J, Hellmann I. 2014. Keeping It Local: Evidence for Positive Selection in Swedish Arabidopsis thaliana. *Mol Biol Evol* 31:3026–3039.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.

Hudson RR, Wayne MS. 1992. Estimation of Levels of Gene FlowFrom DNA Sequence Data. 132:583–589.

Hurst LD, Randerson JP. 2000. Dosage, Deletions and Dominance: Simple Models of the Evolution of Gene Expression. *Journal of Theoretical Biology* 205:641–647.

IPBES. 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In: IPBES secretariat. D´ıaz, S., J. Settele, E. S. Brondizio, H. T. Ngo, M. Gu`eze, J. Agard, A. Arneth, et al. (eds.). IPBES, Bonn, Germany, pp. 45.

IPCC. 2013. Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Intergovernmental Panel on Climate Change Fifth Assessment Report. Cambridge, UK: Cambridge University Press.

Ingvarsson P. 2002. A Metapopulation Perspective on Genetic Diversity and Differentiation in Partially Self-Fertilizing Plants. *Evolution* 56:2368–2373.

Jain SK. 1976. The Evolution of Inbreeding in Plants. *Annual Review of Ecology and Systematics* 7:469–495.

Jensen JD, Thornton KR, Bustamante CD, Aquadro CF. 2007. On the Utility of Linkage Disequilibrium as a Statistic for Identifying Targets of Positive Selection in Nonequilibrium Populations. *Genetics* 176:2371–2379.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405.

Josephs EB, Lee YW, Stinchcombe JR, Wright SI. 2015. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *PNAS* 112:15390–15395.

Josephs EB, Lee YW, Wood CW, Schoen DJ, Wright SI, Stinchcombe JR. 2020. The Evolutionary Forces Shaping Cis- and Trans-Regulation of Gene Expression within a Population of Outcrossing Plants. *Mol Biol Evol* 37:2386–2393.

Josephs EB, Wright SI, Stinchcombe JR, Schoen DJ. 2017. The Relationship between Selection, Network Connectivity, and Regulatory Variation within a Population of Capsella grandiflora. *Genome Biol Evol* 9:1099–1109.

Juric I, Aeschbacher S, Coop G. 2016. The Strength of Selection against Neanderthal Introgression. *PLOS Genetics* 12:e1006340.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* 42:348–354.

Kärkkäinen K, Kuittinen H, Treuren R van, Vogl C, Oikarinen S, Savolainen O. 1999. Genetic Basis of Inbreeding Depression in Arabis Petraea. *Evolution* 53:1354–1365.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecology Letters* 7:1225–1241.

Keightley PD, Eyre-Walker A. 2007. Joint Inference of the Distribution of Fitness Effects of Deleterious Mutations and Population Demography Based on Nucleotide Polymorphism Frequencies. *Genetics* 177:2251–2261.

Keightley PD, Jackson BC. 2018. Inferring the Probability of the Derived vs. the Ancestral Allelic State at a Polymorphic Site. 209:897–906.

Kerdaffrec E, Filiault DL, Korte A, Sasaki E, Nizhynska V, Seren Ü, Nordborg M. 2016. Multiple alleles at a single locus control seed dormancy in Swedish Arabidopsis.Hardtke CS, editor. *eLife* 5:e22502.

Keren L, Hausser J, Lotan-Pompan M, Vainberg Slutskin I, Alisar H, Kaminski S, Weinberger A, Alon U, Milo R, Segal E. 2016. Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* 166:1282-1294.e18.

Kim BY, Huber CD, Lohmueller KE. 2017. Inference of the Distribution of Selection Coefficients for New Nonsynonymous Mutations Using Large Samples. *Genetics* 206:345–361.

Kimura M. 1964. Diffusion models in population genetics. *Journal of Applied Probability* 1:177–232.

Kimura M, Maruyama T, Crow JF. 1963. The Mutation Load in Small Populations. *Genetics* 48:1303–1312.

Kirkpatrick M, Jarne P. 2000. The Effects of a Bottleneck on Inbreeding Depression and the Genetic Load. *The American Naturalist* 155:154–167.

Klopfstein S, Currat M, Excoffier L. 2006. The Fate of Mutations Surfing on the Wave of a Range Expansion. *Molecular Biology and Evolution* 23:482–490.

Koch E, Novembre J. 2017. A Temporal Perspective on the Interplay of Demography and Selection on Deleterious Variation in Humans. *G3: Genes, Genomes, Genetics* 7:1027–1037.

Koch EL, Guillaume F. 2020. Additive and mostly adaptive plastic responses of gene expression to multiple stress in Tribolium castaneum.Mauricio R, editor. *PLoS Genet* 16:e1008768.

Koch MA. 2019. The plant model system Arabidopsis set in an evolutionary, systematic, and spatio-temporal context. *J Exp Bot* 70:55–67.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 15:1179–1191.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356.

Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29.

Kremling KAG, Chen S-Y, Su M-H, Lepak NK, Romay MC, Swarts KL, Lu F, Lorant A, Bradbury PJ, Buckler ES. 2018. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* 555:520–523.

Kronholm I, Picó FX, Alonso-Blanco C, Goudet J, Meaux J de. 2012. GENETIC BASIS OF ADAPTATION IN ARABIDOPSIS THALIANA: LOCAL ADAPTATION AT THE SEED DORMANCY QTL DOG1: LOCAL ADAPTATION FOR SEED DORMANCY QTL DOG1. *Evolution* 66:2287–2302.

Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, Nasrallah ME. 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* 13:627–643.

Laenen B, Tedder A, Nowak MD, Toräng P, Wunder J, Wötzel S, Steige KA, Kourmpetis Y, Odong T, Drouzas AD, et al. 2018. Demography and mating system shape the genome-wide impact of purifying selection in *Arabis alpina*. *Proceedings of the National Academy of Sciences* 115:816–821.

Lafon-Placette C, Köhler C. 2015. Epigenetic mechanisms of postzygotic reproductive isolation in plants. *Current Opinion in Plant Biology* 23:39–44.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.

Lasky JR, Des Marais DL, Lowry DB, Povolotskaya I, McKay JK, Richards JH, Keitt TH, Juenger TE. 2014. Natural Variation in Abiotic Stress Responsive Gene Expression and Local Adaptation to Climate in Arabidopsis thaliana. *Mol Biol Evol* 31:2283–2296.

Leder EH, McCairns RJS, Leinonen T, Cano JM, Viitaniemi HM, Nikinmaa M, Primmer CR, Merilä J. 2015. The Evolution and Adaptive Potential of Transcriptional Variation in Sticklebacks—Signatures of Selection and Widespread Heritability. *Mol Biol Evol* 32:674–689.

Leinonen PH, Sandring S, Quilot B, Clauss MJ, Mitchell-Olds T, Agren J, Savolainen O. 2009. Local adaptation in European populations of Arabidopsis lyrata (Brassicaceae). *American Journal of Botany* 96:1129–1137.

Lemos B, Araripe LO, Fontanillas P, Hartl DL. 2008. Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *PNAS* 105:14471–14476.

Levin DA. 2010. Environment-enhanced self-fertilization: implications for niche shifts in adjacent populations. *Journal of Ecology* 98:1276–1283.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Li PJ, Filiault D, Box MS, Kerdaffrec E, Oosterhout C, Wilczek AM, Schmitt J, McMullan M, Bergelson J, Nordborg M. 2014. Multiple FLC haplotypes defined by independent cis-regulatory variation underpin life history diversity in Arabidopsis thaliana. *Genes & Development* 28:1635–1640.

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.Zhang J, editor. *Molecular Biology and Evolution*:msaa085.

Liu X, Li YI, Pritchard JK. 2019. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* 177:1022-1034.e6.

Lohmueller KE. 2014. The Impact of Population Demography and Selection on the Genetic Architecture of Complex Traits. *PLOS Genetics* 10:1–16.

Louthan AM, Doak DF, Angert AL. 2015. Where and When do Species Interactions Set Range Limits? *Trends in Ecology & Evolution* 30:780–792.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.

Lynch M. 2007. The Origins Of Genome Architecture. 1 edition. Sinauer Associates Inc

Lynch M, Walsh B. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, Massachusetts

Mable BK, Dart AVR, Berardo CD, Witham L. 2005. Breakdown of Self-Incompatibility in the Perennial Arabidopsis Lyrata (brassicaceae) and Its Genetic Consequences. *Evolution* 59:1437–1448.

Mable BK, Hagmann J, Kim S-T, Adam A, Kilbride E, Weigel D, Stift M. 2017. What causes mating system shifts in plants? Arabidopsis lyrata as a case study: Updated online 7 December 2016: This article was originally published under a standard licence, but has now been made available under a CC BY 4.0 licence. The PDF and HTML versions of the paper have been modified accordingly. A corrigendum has also been published. *Heredity* 118:52–63.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–753.

Marburger S, Monnahan P, Seear PJ, Martin SH, Koch J, Paajanen P, Bohutínská M, Higgins JD, Schmickl R, Yant L. 2019. Interspecific introgression mediates adaptation to whole genome duplication. *bioRxiv*:636019.

Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, Fine RS, Lu Y, Schurmann C, Highland HM, et al. 2017. Rare and low-frequency coding variants alter human adult height. *Nature* 542:186–190.

Marsaglia G, Tsang WW, Wang J. 2003. Evaluating Kolmogorov's Distribution. *Journal of Statistical Software* 8:1–4.

Marsden CD, Vecchyo DO-D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, Marques-Bonet T, Schnabel RD, Wayne RK, Lohmueller KE. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *PNAS* 113:152–157.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10–12.

Masalia RR, Bewick AJ, Burke JM. 2017. Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. *PLOS ONE* 12:e0182289.

Matthey-Doret R, Whitlock MC. 2019. Background selection and FST: Consequences for detecting local adaptation. *Molecular Ecology* 28:3902–3914.

Mattila TM, Aalto EA, Toivainen T, Niittyvuopio A, Piltonen S, Kuittinen H, Savolainen O. 2016. Selection for population-specific adaptation shaped patterns of variation in the photoperiod pathway genes in Arabidopsis lyrata during post-glacial colonization. *Molecular Ecology* 25:581–597.

Mattila TM, Tyrmi J, Pyhäjärvi T, Savolainen O. 2017. Genome-Wide Analysis of Colonization History and Concomitant Selection in Arabidopsis lyrata. *Mol Biol Evol* 34:2665–2677.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.

de Meaux J. 2006. An adaptive path through jungle DNA. *Nature Genetics* 38:506–507.

de Meaux J. 2018. Cis-regulatory variation in plant genomes and the impact of natural selection. *American Journal of Botany* 105:1788–1791.

Mendez-Vigo B, Pico FX, Ramiro M, Martinez-Zapater JM, Alonso-Blanco C. 2011. Altitudinal and climatic adaptation is mediated by flowering traits and FRI, FLC, and PHYC genes in Arabidopsis. *Plant Physiol* 157:1942–1955.

Metzger BPH, Wittkopp PJ, Coolon JD. 2017. Evolutionary Dynamics of Regulatory Changes Underlying Gene Expression Divergence among Saccharomyces Species. *Genome Biol Evol* 9:843–854.

Montesinos-Navarro A, Wig J, Picó FX, Tonsor SJ. 2011. Arabidopsis thaliana populations show clinal variation in a climatic gradient associated with altitude. *New Phytologist* 189:282–294.

Morgan MT, Wilson WG, Knight TM. 2005. Plant Population Dynamics, Pollinator Foraging, and the Selection of Self-Fertilization. *The American Naturalist* 166:169–183.

Morrison GD, Linder CR. 2014. Association Mapping of Germination Traits in Arabidopsis thaliana Under Light and Nutrient Treatments: Searching for G x E Effects. *G3: Genes, Ge-nomes, Genetics* 3.

Mousseau TA, Roff DA. 1987. Natural selection and the heritability of fitness components. *Heredity* 59:181–197.

Muller M-H, Leppälä J, Savolainen O. 2008. Genome-wide effects of postglacial colonization in Arabidopsis lyrata. *Heredity* 100:47–58.

Müllner D. 2013. fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. *Journal of Statistical Software* 53:1–18.

Munguía-Rosas MA, Ollerton J, Parra-Tabla V, De-Nova JA. 2011. Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured. *Ecol-ogy Letters* 14:511–521.

Oleksiak MF, Churchill GA, Crawford DL. 2002. Variation in gene expression within and among natural populations. *Nature Genetics* 32:261–266.

Orr HA. 2009. Fitness and its role in evolutionary genetics. *Nat Rev Genet* 10:531–539.

Otto SP, Whitlock MC. 1997. The Probability of Fixation in Populations of Changing Size. *Genetics* 146:723–733.

Ovaskainen O, Cano JM, Merila J. 2008. A Bayesian framework for comparative quantitative genetics. *Proceedings of the Royal Society B: Biological Sciences* 275:669–678.

Pavlidis P, Živković D, Stamatakis A, Alachiotis N. 2013. SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Molecular Biology and Evolution* 30:2224–2234.

Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L. 2013. On the accumulation of deleterious mutations during range expansions. *Molecular Ecology* 22:5972–5982.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Molecular Biology and Evolution* 31:1929–1936.

Picó FX. 2012. Demographic fate of Arabidopsis thaliana cohorts of autumn- and spring-germinated plants along an altitudinal gradient. *Journal of Ecology* 100:1009–1018.

Postma FM, Ågren J. 2016. Early life stages contribute strongly to local adaptation in Arabidopsis thaliana. In: Proceedings of the National Academy of Sciences. Vol. 113. p. 7590–7595.

Price AL, Helgason A, Thorleifsson G, McCarroll SA, Kong A, Stefansson K. 2011. Single-Tissue and Cross-Tissue Heritability of Gene Expression Via Identity-by-Descent in Related or Unrelated Individuals. *PLOS Genetics* 7:e1001317.

Pritchard JK, Pickrell JK, Coop G. 2010. The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology* 20:R208–R215.

Prud'homme B, Gompel N, Carroll SB. 2007. Emerging principles of regulatory evolution. *PNAS* 104:8605–8612.

Pujol B, Zhou S-R, Sanchez Vilas J, Pannell JR. 2009. Reduced inbreeding depression after species range expansion. *Proceedings of the National Academy of Sciences* 106:15379–15383.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* 81:559–575.

Pyhäjärvi T, Aalto E, Savolainen O. 2012. Time scales of divergence and speciation among natural populations and subspecies of *Arabidopsis lyrata* (Brassicaceae). *American Journal of Botany* 99:1314–1322.

Pyhäjärvi T, García-Gil MR, Knürr T, Mikkonen M, Wachowiak W, Savolainen O. 2007. Demographic History Has Influenced Nucleotide Diversity in European Pinus sylvestris Populations. *Genetics* 177:1713–1724.

Quilot-Turion B, Leppälä J, Leinonen PH, Waldmann P, Savolainen O, Kuittinen H. 2013. Genetic changes in flowering and morphology in response to adaptation to a high-latitude environment in Arabidopsis lyrata. *Annals of Botany* 111:957–968.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/

Rifkin SA, Houle D, Kim J, White KP. 2005. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* 438:220–223.

Robertson KD. 2005. DNA methylation and human disease. *Nature Reviews Genetics* 6:597–610.

Roessler K, Muyle A, Diez CM, Gaut GRJ, Bousios A, Stitzer MC, Seymour DK, Doebley JF, Liu Q, Gaut BS. 2019. The genome-wide dynamics of purging during selfing in maize. *Nature Plants* 5:980–990.

Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* 13:505–516.

Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut BS. 2008. Patterns of Polymorphism and Demographic History in Natural Populations of Arabidopsis lyrata.Fay JC, editor. *PLoS ONE* 3:e2411.

Sanchez-Bermejo E, Mendez-Vigo B, Pico FX, Martinez-Zapater JM, Alonso-Blanco C. 2012. Novel natural alleles at FLC and LVR loci account for enhanced vernalization responses in Arabidopsis thaliana. *Plant, Cell Env* 35:1672–1684.

Sandring S, Riihimäki M-A, Savolainen O, Ågren J. 2007. Selection on flowering time and floral display in an alpine and a lowland population of Arabidopsis lyrata. *Journal of Evolutionary Biology* 20:558–567.

Sasaki E, Zhang P, Atwell S, Meng D, Nordborg M. 2015. Missing' G x E Variation Con-trols Flowering Time in Arabidopsis.thaliana, editor. *PLoS Genetics* 11:1005597.

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* 14:807–820.

Schierup MH. 1998. The Number of Self-Incompatibility Alleles in a Finite, Subdivided Population. *Genetics* 149:1153–1162.

Schierup MH, Bechsgaard JS, Christiansen FB. 2008. Selection at Work in Self-Incompatible Arabidopsis lyrata. II. Spatial Distribution of S Haplotypes in Iceland. *Genetics* 180:1051–1059.

Schierup MH, Bechsgaard JS, Nielsen LH, Christiansen FB. 2006. Selection at Work in Self-Incompatible Arabidopsis lyrata: Mating Patterns in a Natural Population. *Genetics* 172:477–484.

Schierup MH, Mable BK, Awadalla P, Charlesworth D. 2001. Identification and Characterization of a Polymorphic Receptor Kinase Gene Linked to the Self-Incompatibility Locus of Arabidopsis lyrata. *Genetics* 158:387-399.

Schmickl R, Jørgensen MH, Brysting AK, Koch MA. 2010. The evolutionary history of the Arabidopsis lyrata complex: a hybrid in the amphi-Beringian area closes a large distribution gap and builds up a genetic barrier. *BMC Evolutionary Biology* 10:98.

Schrider DR. 2020. Background selection does not mimic the patterns of genetic diversity produced by selective sweeps. *bioRxiv*:2019.12.13.876136.

Shaw RG, Shaw FH. 2014. Quantitative genetic study of the adaptive process. *Heredity* 112:13–20.

Simons YB, Sella G. 2016. The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Current Opinion in Genetics & Development* 41:150–158.

Simons YB, Turchin MC, Pritchard JK, Sella G. 2014. The deleterious mutation load is insensitive to recent population history. *Nature Genetics* 46:220–224.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462.

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

Sletvold N, Mousset M, Hagenblad J, Hansson B, Ågren J. 2013. Strong Inbreeding Depression in Two Scandinavian Populations of the Self-Incompatible Perennial Herb Arabidopsis Lyrata. *Evolution* 67:2876–2888.

Steige KA, Laenen B, Reimegård J, Scofield DG, Slotte T. 2017. Genomic analysis reveals major determinants of cis-regulatory variation in Capsella grandiflora. *PNAS* 114:1087–1092.

Steige KA, Reimegård J, Koenig D, Scofield DG, Slotte T. 2015. Cis-Regulatory Changes Associated with a Recent Mating System Shift and Floral Adaptation in Capsella. *Mol Biol Evol* 32:2501–2514.

Stephan W. 2016. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Mol Ecol* 25:79–88.

Stupar RM, Hermanson PJ, Springer NM. 2007. Nonadditive Expression and Parent-of-Origin Effects Identified by Microarray and Allele-Specific Expression Profiling of Maize Endosperm. *Plant Physiology* 145:411–425.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 102:15545–15550.

Svardal H, Farlow A, Exposito-Alonso M, Ding W, Novikova P, Alonso-Blanco C, Weigel D, LEE C-R, Nordborg M. 2017. On the post-glacial spread of human commensal Arabidopsis thaliana. *Nature Communications* 8:1–12.

Taylor MA, Cooper MD, Sellamuthu R, Braun P, Migneault A, Browning A, Perry E, Schmitt J. 2017. Interacting effects of genetic variation for seed dormancy and flowering time on phenology, life history, and fitness of experimental Arabidopsis thalianapopulations over multiple generations in the field. *New Phytologist* 216:291–302.

Todesco M, Pascual MA, Owens GL, Ostevik KL, Moyers BT, Hübner S, Heredia SM, Hahn MA, Caseys C, Bock DG, et al. 2016. Hybridization and extinction. *Evolutionary Applications* 9:892–908.

Toivainen T, Pyhäjärvi T, Niittyvuopio A, Savolainen O. 2014. A recent local sweep at the PHYA locus in the Northern European Spiterstulen population of Arabidopsis lyrata. *Molecular Ecology* 23:1040–1052.

Toomajian C, Hu TT, Aranzana MJ, Lister C, Tang C, Zheng H, Zhao K, Calabrese P, Dean C, Nordborg M. 2006. A nonparametric test reveals selection for rapid flowering in the Arabidopsis genome. *PLoS Biology* 4.

Tsuchimatsu T, Kaiser P, Yew C-L, Bachelier JB, Shimizu KK. 2012. Recent Loss of Self-Incompatibility by Degradation of the Male Component in Allotetraploid Arabidopsis kamchatica.Mauricio R, editor. *PLoS Genetics* 8:e1002838.

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.Moyle LC, editor. *PLoS Biol* 17:e3000214.

Vekemans X, Poux C, Goubet PM, Castric V. 2014. The evolution of selfing from outcrossing ancestors in Brassicaceae: what have we learned from variation at the *S*- locus? *Journal of Evolutionary Biology* 27:1372–1385.

Vergeer P, Kunin WE. 2013. Adaptation at range margins: common garden trials and the performance of Arabidopsis lyrata across its northwestern European range. *New Phytologist* 197:989–1001.

Videvall E, Sletvold N, Hagenblad J, Ågren J, Hansson B. 2016. Strong Maternal Effects on Gene Expression in Arabidopsis lyrata Hybrids. *Mol Biol Evol* 33:984–994.

Vidigal DS, Marques ACSS, Willems LAJ, Buijs G, Méndez-Vigo B, Hilhorst HWM, Ben-tsink L, Picó FX, Alonso-Blanco C. 2016. Altitudinal and climatic associations of seed dormancy and flowering traits evidence adaptation of annual life cycle timing in Arabidopsis tha-liana. *Plant, Cell & Environment* 39:1737–1748.

Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG. 2006. Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. *PLOS Genetics* 2:e41.

Wakeley J. 1996. The Variance of Pairwise Nucleotide Differences in Two Populations with Migration. *Theoretical Population Biology* 49:39–57.

Waldmann P. 2001. Additive and non-additive genetic architecture of two different-sized populations of Scabiosa canescens. *Heredity* 86:648–657.

Waldvogel A-M, Feldmeyer B, Rolshausen G, Exposito-Alonso M, Rellstab C, Kofler R, Mock T, Schmid K, Schmitt I, Bataillon T, et al. 2020. Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters* 4:4–18.

Weigel D, Nordborg M. 2015. Population Genomics for Understanding Adaptation in Wild Plant Species. *Annual Review of Genetics* 49:315–338.

Whitehead A, Crawford DL. 2006. Neutral and adaptive variation in gene expression. *PNAS* 103:5425–5430.

Whittaker C, Dean C. 2017. The FLC Locus: A Platform for Discoveries in Epigenetics and Adaptation. *Annual Review of Cell and Developmental Biology* 33:555–575.

Wieters B, Steige KA, He F, Koch EM, Ramos-Onsins SE, Gu H, Guo Y-L, Sunyaev S, de Meaux J. 2020. Polygenic adaptation of rosette growth in Arabidopsis thaliana. *bioRxiv* [Internet]. Available from: https://www.biorxiv.org/content/early/2020/09/09/2020.03.31.018341

Wilczek AM, Roe JL, Knapp MC. 2009. Effects of genetic perturbation on seasonal life history plasticity. *Science* 323:930–934.

Willi Y, Fracassetti M, Zoller S, Van Buskirk J. 2018. Accumulation of Mutational Load at the Edges of a Species Range. *Molecular Biology and Evolution* 35:781–791.

Willi Y, Griffin P, Van Buskirk J. 2013. Drift load in populations of small size and low density. *Heredity* 110:296–302.

Williamson SH, Hernandez R, Fledel-Alon A, Zhu L, Nielsen R, Bustamante CD. 2005. Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc Natl Acad Sci U S A* 102:7882–7887.

Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LEB, Nussey DH. 2010. An ecologist's guide to the animal model. *Journal of Animal Ecology* 79:13–26.

Wos G, Willi Y. 2018. Genetic differentiation in life history traits and thermal stress performance across a heterogeneous dune landscape in Arabidopsis lyrata. *Ann Bot* 122:473–484.

Wright MN, Ziegler A. 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* 77:1–17.

Wright SI, Lauga B, Charlesworth D. 2003. Subdivision and haplotype structure in natural populations of Arabidopsis lyrata. *Molecular Ecology* 12:1247–1263.

Yang CJ, Samayoa LF, Bradbury PJ, Olukolu BA, Xue W, York AM, Tuholski MR, Wang W, Daskalska LL, Neumeyer MA, et al. 2019. The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proc Natl Acad Sci USA* 116:5643–5652.

Yeaman S. 2015. Local Adaptation by Alleles of Small Effect. *Am Nat* 186 Suppl 1:S74-89.

Younginger BS, Sirová D, Cruzan MB, Ballhorn DJ. 2017. Is biomass a reliable estimate of plant fitness? *Applications in Plant Sciences* 5:1600094.

Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. 2018. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files.Schwartz R, editor. *Bioinformatics* 35:1786–1788.

Zou Y-P, Hou X-H, Wu Q. 2017. Adaptation of Arabidopsis thaliana to the Yangtze River basin. *Genome Biology* 18:239.

6.Appendix

6.1 Custom scripts

6.1.1 Python script for calculation of genetic distance

```
#!/usr/bin/env python


import numpy as np

from numpy import array

import sys

import csv


#genetic pairwise diversit per sample

def genetPairDiver(vcfIO):

	lines = 0

	filename = input("Give file name to save the matrix as 'name.txt':")

	for line in vcfIO:

		if line[1] == '#':

			pass

		else:

			if line[1] == 'C':

				line2 = line.split('\t')

				line2 = line2[9::]

				indexes = [x for x in range(0,len(line2))]

				num_diff = np.zeros((len(indexes), len(indexes))) #create the table that I can add if different the two indiv
```

```
                num_comp = np.zeros((len(indexes), len(indexes))) #create the table that I can add number of comparisons
                #print indexes, line2
            else:
                lines = lines + 1 #count the number of lines in the variant file
                line2 = line.split('\t')
                line2 = line2[9::]
                for indv in indexes:
                    i = indv + 1
                    gt1 = line2[indv].split(':') #get the correct index for each indv
                    gt1 = gt1[0]
                    while i < len(indexes): #to compare with the next indv in the line, The comparison with the ones before has already be done, and no need to compare with itself
                        gt2 = line2[i].split(':')
                        gt2 = gt2[0]
                        #print lines,indv, i
                        if gt1 == './.' or gt2 =='./.': #skip if missing info
                        #       print 1
                            pass
                        else:
                            if gt1 != gt2:
                                num_diff[indv, i] = num_diff[indv, i] + 1
```

```
                                        num_comp[indv, i] = num_comp[indv, i] + 1
                                        num_diff[i, indv] = num_diff[i, indv] + 1 #add in both columns and rows, the matrix for heatmap3 must be symmetrical
                                        num_comp[i, indv] = num_comp[i, indv] + 1
                                        #print num_comp
                                #       print 2
                                elif gt1 == gt2:
                                        num_comp[indv, i] = num_comp[indv, i] + 1
                                        num_comp[i, indv] = num_comp[i, indv] + 1 #count it only in the number of comparisons.
                                #       print 3
                                i = i + 1
        diff_1 = num_diff / num_comp
        diff_1 = diff_1 * float(lines)
        diff_1 = diff_1 / float(nonvar)  #normalise by multiple total number of sites in variant sites file and divide by total number of lines in nonvariant file
        np.savetxt('{}'.format(filename), diff_1, delimiter='\t')


file = sys.argv[1] #vcf file containing all the variant sites for the gene
nonvar = sys.argv[2] #number of nonvariant sites, pre counted
vcfIO = open(file, 'r')
```

```
genetPairDiver(vcfIO, nonvar)

sys.exit("Live long and prosper!")
```

6.1.2 Python script for the estimation of the genomic load per individual

```
#!/usr/bin/env python


import os

import sys

import csv

import numpy as np


###get the number of derived alleles, based on the polyDFE dervided state file.

#

def counts2(vcf, non, outTable, r, name):

	dictPos = {'1':[], '2':[], '3':[], '4':[], '5':[], '6':[], '7':[], '8':[]}

	print("Creating dictionary with positions of input variants...")

	for line in non:

		if line[0] == '#':

			pass

		else:

			line2 = line.split("\t")

			chrom = line2[0]

			pos = line2[1][0:-1]

			dictPos[chrom].append(pos)

	#print(dictPos['1'])
```

```
outFile = open("{}.derived.pos.alls.txt".format(name), 'w')
print("Going through the csv file...")
for line in vcf:
        if line[0] == 'C':
                outFile.write(line)
        else:
                line2 = line.split(",")
                chrom = line2[0]
                pos = line2[1]
                #print pos
                #print line2
                if pos in dictPos[chrom]:
                        line3 = line2[4::]
                        #print line3
                        l = len(line3)
                        outFile.write(line)
                        #print l
                        i = 0
                        while i < l:
                                all1 = line3[i]
                                all2 = line3[(i + 1)][0]
                                #print all1, all2
```

```
                    if all1 == 'NA' or all2 == 'NA': #missing info for the individual
                        pass
                    elif all1 == '0' and all2 == '0': #it is the ancestral state, then no change
                        pass
                    elif all1 == '0' and all2 == '1': #it is a change in the individual
                        outTable[i/2, int(r)] = outTable[i/2, int(r)] + 1 #this is like before, the general count of het or hom state
                        outTable[i/2, (int(r) + 1)] = outTable[i/2, (int(r) + 1)] + 0.5 #for the haploid genome the heter is +0.5
                    elif all1 == '1' and all2 == '0': #it is a change in the individual
                        outTable[i/2, int(r)] = outTable[i/2, int(r)] + 1 #this is like before, the general count of het or hom state
                        outTable[i/2, (int(r) + 1)] = outTable[i/2, (int(r) + 1)] + 0.5 #for the haploid genome the heter is +0.5
                    elif all1 == '1' and all2 == '1':
                        outTable[i/2, int(r)] = outTable[i/2, int(r)] + 1 #this is like before, the general count of het or hom state
                        outTable[i/2, (int(r) + 1)] = outTable[i/2, (int(r) + 1)] + 1 #for the haploid genome the hom is +1
                        #print 1
                    i = i + 2
                else:
                    pass
```

```
        print("Counted the variants per individual.")
        #print(outTable)
        outFile.close
        return(outTable)

############Main Part of the Code######
vcf1 = open(sys.argv[1], 'r')
output = np.zeros((int(sys.argv[4]), 2)) #create the table that I can add the data if the individual has the syn and nonsyn variance. Non/Syn/ratio, the order of the things

non = open(sys.argv[2], 'r')
output = counts2(vcf1, non, output, 0, sys.argv[3])
np.savetxt('derived.{}.counts.txt'.format(sys.argv[3]), output, delimiter='\t')
sys.exit("Live long and prosper!")
```

6.1.3 R Script for running the MCMCglmm analysis on the cluster or locally

```
#!/usr/bin/env Rscript

args = commandArgs(trailingOnly=TRUE)

start_time <- Sys.time()


####Arguements are the array number, and number of genes to be run in each file

library(MCMCglmm)

library(pedigreemm)

library(nadiv)

library(coda)


load("counts.RData")

#load("genes100.RData")


#give the table with the counts, the first gene index, last gene index

#return a table with the VA, Vd, Vr, h2, diagnostics

variancesCalc <- function(tre, st, end){

  out1 <- NULL

  flags <- NULL

  gelman <- NULL

  prior4 <- list(R=list(V=1, nu=0.02), G = list(G1 = list(V=1, nu=0.02), G2 = list(V=1, nu=0.02),
G3 = list(V=1, nu=0.02)))

  #counts <- tre[st:end,] #pick a subset to iterate over
```

```
counts <- tre[,st:end] #already transposed dataset
pdf(paste(st, end, "gelman", "pdf", sep = "."))
par(mfrow=c(2,2), mar=c(2, 1, 1, 1))
for (iter in 1:length(colnames(counts))){
  temp <- as.data.frame(counts[,iter])
  gene <- colnames(counts)[iter]
        print(gene)
  colnames(temp) <- "counts"
  temp$animal <- rownames(temp)
  temp$dom <- rownames(temp)
  #print(temp$animal)
  temp <- merge(temp,  toMergeFam, by="animal")
  #print(summary(temp))
        temp$counts <- as.numeric(as.character(temp$counts))
        print(summary(temp))
  print(hist(log2(temp$counts)))
  f <- "o"
  tryCatch({
    gene.model <- MCMCglmm(log2(counts) ~ pop, random= ~ animal + dom + Dam, ginverse = list(dom=Dinv), start=list(QUASI=FALSE), singular.ok = FALSE, family="gaussian", prior=prior4,  pedigree=fam_matrix2, data=temp, nitt = 2200000, burnin = 200000, thin=2000, verbose = FALSE)
    print(summary(gene.model))
```

```
EfR <- effectiveSize(gene.model$Sol)[1] #effective size of fixed effects. More than 1000, close or above 10000 is good

EfA <- effectiveSize(gene.model$VCV)[1] #effective size of random effects

EfD <- effectiveSize(gene.model$VCV)[2]

EfM <- effectiveSize(gene.model$VCV)[3]

EfU <- effectiveSize(gene.model$VCV)[4]

Va <- median(gene.model$VCV[,"animal"])

Vd <- median(gene.model$VCV[,"dom"])

Vr <- median(gene.model$VCV[,"units"])

Vm <- median(gene.model$VCV[,"Dam"])

h2 <- Va / (Va + Vm + Vr + Vd)

a.cor <- autocorr.diag(gene.model$VCV)[2]

d.cor <- autocorr.diag(gene.model$VCV)[7]

m.cor <- autocorr.diag(gene.model$VCV)[12]

u.cor <- autocorr.diag(gene.model$VCV)[17]

sdA <- sd(gene.model$VCV[,"animal"])

sdD <- sd(gene.model$VCV[,"dom"])

sdU <- sd(gene.model$VCV[,"units"])

sdM <- sd(gene.model$VCV[,"Dam"])

outp <- c(as.character(gene), Va, Vd, Vm, Vr, h2, sdA, sdD, sdM, sdU, a.cor, d.cor,m.cor, u.cor,EfR, EfA, EfD, EfM, EfU)

out1 <- rbind(out1, outp)

f <- "e"
```

```
}, error = function(cond){print("Error in first model.")})

###run the other chain

tryCatch({

gene.model1 <- MCMCglmm(log2(counts) ~ pop, random= ~ animal + dom + Dam, ginverse = list(dom=Dinv), start=list(QUASI=FALSE), singular.ok = TRUE, family="gaussian", prior=prior4, pedigree=fam_matrix2, data=temp, nitt = 2200000, burnin = 200000, thin=2000, verbose = FALSE)

chains <- mcmc.list(gene.model$Sol, gene.model1$Sol) #collect the chain

gelman <- rbind(gelman, c(as.character(gene), gelman.diag(chains)$psrf[1], gelman.diag(chains)$psrf[2]))#the actual diagnostic. Has to be close to 1

print(plot(chains, ask=F, auto.layout=F)) #plot them on top of each other

f <- "y"

}, error = function(cond){print("Error in second model.")})

if (f == "y"){flags <- rbind(flags, c(as.character(gene), "pass"))} #both chains run properly

else if (f == "e"){flags <- rbind(flags, c(as.character(gene), "fail2nd"))} #only the first passed

else if (f == "o"){flags <- rbind(flags, c(as.character(gene), "fail"))} #both chains failed

}

colnames(out1) <- c("Gene", "Va", "Vd","Vm", "Vr", "h2", "Va.sd", "Vd.sd","Vm.sd", "Vr.sd", "Va.cor", "Vd.cor","Vm.cor", "Vr.cor", "Ef.R","Ef.A", "Ef.D","Ef.M", "Ef.U")

colnames(gelman) <- c("Gene", "Inter", "PopSp")

dev.off()

assign("gelmanValues", gelman, envir =.GlobalEnv)

assign("flags", flags, envir =.GlobalEnv)

return(out1)
```

```
}

csvprint <- function(x, nameP, row.names=FALSE, col.names=TRUE){
 write.table(x, file=nameP, append=FALSE, eol='\n', sep="\t", na = "NA", dec='.', row.names=FALSE, col.names=TRUE)
}

#pick the index based on the array iteration on the cluster
#args[1] is the array being run.
#args[2] is the number of genes I want to run in each array
end = as.numeric(as.character(args[1])) * as.numeric(as.character(args[2]))
start = end - (as.numeric(as.character(args[2])) - 1)
print(start)
print(end)
variances1.2k <- variancesCalc(counts, start, end)
#variances1.2k <- variancesCalc(genes100, start, end)
csvprint(variances1.2k, paste("variances", start, end, "csv", sep="."))
csvprint(gelmanValues, paste("gelmanValues", start, end, "csv", sep="."))
csvprint(flags, paste("flags", start, end, "csv", sep="."))
end_time <- Sys.time()
end_time - start_time
```

6.2 Protocols for PCR and digestion of DNA

PCR with the custom primers

The annealing Temperature of the primers is 56°C

Add per sample:

H_2O	15.9 µl
10x Buffer	2 µl
dNTPs 10mM	0.3 µl
Forward Primer	0.5 µl
Reverse Primer	0.5 µl
Dream Taq	0.1 µl

Run the following program in the PCR machine:

1. 95^{o}C for 2 mins
2. 95^{o}C for 25secs
3. 60^{o}C for 30 secs
4. 72^{o}C for 30 secs
5. 72^{o}C for 5 mins

Repeat 35 times the steps 2 to 3.

Digestion with the enzyme Pvu II (Promega)

Add per sample:

H_2O	16.3 µl
RE 10x Buffer	2 µl
Acetylated BSA	0.2 µl
Restrivtion Enzyme	0.5 µl
PCR Product	10 U/µl

Incubate at 37^{o}C for2 hours.

6.3 SupplementaryTables

Appendix Table 1: Phenotypic data of PL and SP individuals, collected during common garden experiment in Cologne. The experiment started in September 2017 and lasted for a year. The Diameter per month is given in mm and the fresh weight, dry weight and dry to fresh weight for each individual in gramms. Each of our original individuals was replicated by vegetative clones. ID signifies the clones and Family the original plant.

ID	Fam-ily	Popu-lation	Spe-cies	Diam-March (mm)	Di-amApril(mm)	Di-amMay(mm)	Diam-June(mm)	Dia-mAug (mm)	Diam-July(mm)	Fresh Weight(gr)	Dry-Weight(gr)	Dryto Fresh WeightRatio
1897	10a	PL	A.ly-rata	NA	NA	NA	44,98	45,78	40,17	3,552	0,5635	0,158643018
1905	1a	PL	A.ly-rata	85,13	76,46	80,2	29,62	38,15	35,59	4,078	0,8181	0,200613046
1908	10a	PL	A.ly-rata	37,43	35,47	26,85	27,78	36,32	29,17	1,612	0,3099	0,192245658
1903	9a	PL	A.ly-rata	23,98	27,71	32,29	32,1	33,24	34,85	1,438	0,2855	0,198539638
1913	80936	PL	A.ly-rata	17,05	23,03	20,85	33,02	30,12	29,39	2,546	0,4361	0,171288295
1896	5a	PL	A.ly-rata	31,62	33,97	46,53	31,35	35,09	33,03	2,18	0,2434	0,111651376
1900	3a	PL	A.ly-rata	41,83	32,08	24,1	26,17	25,47	27,81	1,144	0,1829	0,159877622
1906	2a	PL	A.ly-rata	NA	NA	NA	NA	25,01	22,27	1,092	0,1043	0,095512821
1904	4a	PL	A.ly-rata	48,49	41,18	17,63	43,43	46,38	41,98	4,315	0,7226	0,167462341
1898	6a	PL	A.ly-rata	NA	NA	NA	NA	26,48	23,76	0,488	0,0879	0,180122951
1902	11a	PL	A.ly-rata	76,29	74,02	NA	52,15	39,95	43,36	3,124	0,696	0,222791293
1901	80936	PL	A.ly-rata	NA	5,1	23,56	NA	NA	NA	NA	NA	NA
1899	7a	PL	A.ly-rata	66,3	59,12	13,35	44,36	45,75	27,81	4,388	0,8178	0,186371923
1943	9a	PL	A.ly-rata	41,26	28,2	46,57	41,69	48,93	35,66	3,353	0,5808	0,173218014
1938	6a	PL	A.ly-rata	NA	4,85	NA	NA	NA	NA	NA	NA	NA
1939	7a	PL	A.ly-rata	55,24	38,16	18,9	41,29	48,22	34,3	2,77	0,4243	0,153176895
1942	11a	PL	A.ly-rata	35,34	38,44	40,4	28,95	40,75	30,88	2,07	0,266	0,128502415
1940	5a	PL	A.ly-rata	64,8	53,88	38,68	24,81	47,05	30,37	2,893	0,4672	0,16149326
1946	2a	PL	A.ly-rata	68,2	56,46	34,99	17,46	35,54	19,67	2,24	0,3544	0,158214286

1945	1a	PL	A.ly-rata	36,3	34,6	32,08	26,77	32,79	30,9	1,48	0,2401	0,16222973
1944	4a	PL	A.ly-rata	24,72	22,54	23,16	25,36	46,19	25,17	1,46	0,2439	0,16705479 5
1954	2a	PL	A.ly-rata	15,11	18,64	19,33	24,78	47,48	26,47	2,229	0,9158	0,41085686
1953	10a	PL	A.ly-rata	68,45	66,49	56,11	34,14	64,13	45,99	2,44	0,1286	0,05270491 8
1936	3a	PL	A.ly-rata	47,25	44,38	36,73	30,22	37,1	35,9	1,569	0,1102	0,07023581 9
1937	10a	PL	A.ly-rata	33,13	40,2	NA	27,41	43,58	33,54	2,045	0,3901	0,19075794 6
2001	11a	PL	A.ly-rata	54,92	53,93	17,29	40,8	49,84	26,05	3,396	0,6729	0,19814487 6
1983	9a	PL	A.ly-rata	28,33	29,11	23,76	28,61	37,9	30,84	1,935	0,0402	0,02077519 4
1981	80936	PL	A.ly-rata	69,64	74,47	64,78	38,98	52,1	40,6	3,457	0,4149	0,12001735 6
1984	4a	PL	A.ly-rata	36,44	34,22	33,57	27,71	36,03	32,55	2,002	0,2224	0,11108891 1
1986	2a	PL	A.ly-rata	26,52	27,94	26,7	22,58	42,27	27,18	1,974	0,2426	0,12289767
1977	10a	PL	A.ly-rata	37,33	40,62	40,58	32,32	47,81	40,56	1,059	0,1289	0,12171860 2
1976	3a	PL	A.ly-rata	15,46	28,63	25,2	37,85	45,78	31,57	1,709	0,3991	0,23352837 9
1978	6a	PL	A.ly-rata	48,04	64,3	60,7	24,79	49,86	39,51	3,681	0,5559	0,15101874 5
1980	5a	PL	A.ly-rata	NA	NA	NA	NA	29,71	20,45	0,718	0,1347	0,18760445 7
1985	1a	PL	A.ly-rata	17,78	32,07	30,8	26,39	49,47	22,78	0,954	0,2999	0,31436058 7
2023	9a	PL	A.ly-rata	51,43	62,22	70,8	22,19	43,79	33,24	4,207	0,8667	0,20601378 7
2019	7a	PL	A.ly-rata	31,88	3,56	NA	NA	NA	NA	NA	NA	NA
2020	5a	PL	A.ly-rata	10,05	NA	NA	NA	NA	NA	NA	NA	NA
2022	11a	PL	A.ly-rata	19,5	30,12	28,9	30,92	37,01	32,71	2,173	0,4297	0,19774505 3
2018	6a	PL	A.ly-rata	32,24	42,86	44,1	40,85	43,05	40,27	3,244	0,5627	0,17345869 3
2024	4a	PL	A.ly-rata	11,84	21,3	20,8	34,83	48,03	30,79	2,072	0,2943	0,14203668
2026	2a	PL	A.ly-rata	15,98	14,08	14,2	16,1	24,05	20,41	0,697	0,1194	0,17130559 5
2017	10a	PL	A.ly-rata	29,46	53,62	60,01	32,62	32,05	35,89	3,273	0,6081	0,18579285 1
2021	80936	PL	A.ly-rata	NA	16,42	20,1	30,74	28,43	16,07	1,745	0,2904	0,16641833 8
2025	1a	PL	A.ly-rata	20,39	36,11	38,9	37,03	37,55	NA	2,798	0,4869	0,17401715 5
2016	3a	PL	A.ly-rata	NA	3,54	22,9	32,08	37,15	31,4	1,415	0,2144	0,15151943 5

2074	2a	PL	A.ly-rata	48,76	58,01	41,12	23,76	35,46	32,07	1,358	0,2874	0,211634757
2067	7a	PL	A.ly-rata	14,34	27,14	23,94	33,04	32,17	24,16	1,683	0,2751	0,163458111
2061	80936	PL	A.ly-rata	86,91	41,92	43,45	22,83	35,06	21,71	1,497	0,3676	0,245557782
2056	3a	PL	A.ly-rata	78,58	29,53	29,65	21,78	20,95	16,74	NA	NA	NA
2066	2a	PL	A.ly-rata	NA	13,44	14,4	24,27	21,42	23,25	0,592	0,1221	0,20625
2062	11a	PL	A.ly-rata	64,29	60,96	60,7	39,04	33,51	38,18	3,004	0,6166	0,205259654
2072	11a	PL	A.ly-rata	43,06	38,95	40,6	25,25	35,08	NA	1,984	0,3064	0,154435484
2059	7a	PL	A.ly-rata	65,44	38,89	50,8	28,52	37,53	23,79	2,423	0,4399	0,181551795
2063	9a	PL	A.ly-rata	34,16	37,1	40	34,22	33,53	36,49	1,86	0,2899	0,155860215
2064	4a	PL	A.ly-rata	41,31	39,11	50,2	32,47	44,47	36,13	2,054	0,4212	0,205063291
2058	6a	PL	A.ly-rata	27,78	29,81	17,37	23,8	33,85	26,01	1,655	0,324	0,195770393
2060	5a	PL	A.ly-rata	53,72	40,89	58,9	44,9	49,15	47,78	2,599	0,4914	0,18907272
2065	1a	PL	A.ly-rata	54	23,97	34,48	26,5	43,88	25,82	3,612	0,6328	0,175193798
2136	3a	PL	A.ly-rata	25,62	25,76	20,1	34,08	35,78	31,76	0,649	0,1439	0,221725732
2143	9a	PL	A.ly-rata	38,15	31,15	28,08	28,77	45,78	40,01	2,598	0,4605	0,177251732
2139	7a	PL	A.ly-rata	48,81	NA	NA	NA	23,84	NA	NA	NA	NA
2144	4a	PL	A.ly-rata	NA	5,9	10,01	32,92	51,67	39,88	3,086	0,4836	0,156707712
2146	2a	PL	A.ly-rata	NA	47,39	26,59	26,67	41,59	27,44	2,42	0,4616	0,190743802
2138	6a	PL	A.ly-rata	32,83	29,28	30,1	21,22	33,78	20,01	1,397	0,2689	0,192483894
2141	80936	PL	A.ly-rata	50,13	44,76	50,81	14,73	37,16	21,2	2,509	0,5361	0,213670785
2154	6a	PL	A.ly-rata	60,08	55,25	60,1	26,15	51,27	30,35	3,639	0,5674	0,155921957
2137	10a	PL	A.ly-rata	NA	NA	NA	32,56	49,83	45,56	1,632	0,2294	0,140563725
2155	3a	PL	A.ly-rata	42,22	47,42	23,81	26,37	39,88	31,15	1,31	0,1763	0,134580153
2142	11a	PL	A.ly-rata	63,44	61,15	70,41	41,04	39,67	38,42	3,089	0,6325	0,204758822
2145	1a	PL	A.ly-rata	54,82	50,34	25,46	27,24	39,24	29,81	2,724	0,4114	0,1510279
2140	5a	PL	A.ly-rata	NA	18,13	NA	NA	NA	NA	NA	NA	NA
1915	70542	SP	A.ly-rata	10,67	26,81	23,89	35,66	32,41	30,92	1,631	0,2788	0,170938075

1916	70545	SP	A.ly-rata	NA	NA	NA	NA	NA	NA	0,102	0,0023	0,02254902
1914	70541	SP	A.ly-rata	NA	7,46	14,23	29,2	35,41	25,84	2,076	0,2624	0,126396917
1910	70538	SP	A.ly-rata	NA	15,68	NA	30,31	32,48	28,77	1,783	0,3141	0,176163769
1912	70539	SP	A.ly-rata	29,9	26,1	NA	45,52	NA	NA	NA	NA	NA
1909	70537	SP	A.ly-rata	NA	NA	NA	40,15	60,01	34,58	3,058	0,4491	0,146860693
1911	SP5	SP	A.ly-rata	NA	NA	54,43	NA	NA	NA	NA	NA	NA
1957	SPH7	SP	A.ly-rata	10,09	8,57	NA	NA	NA	NA	NA	NA	NA
1955	70542	SP	A.ly-rata	NA	NA	15,16	25,74	33,61	30,88	1,3	0,2221	0,170846154
1950	70538	SP	A.ly-rata	16,33	23,74	25,6	31,79	48,61	40,96	1,837	0,3376	0,183777899
1948	70537	SP	A.ly-rata	NA	14,47	37,51	NA	37,62	33,88	0,956	0,1625	0,169979079
1997	70538	SP	A.ly-rata	25,83	20,71	33,01	30,87	39,32	26,9	2,159	0,3115	0,144279759
2009	70539	SP	A.ly-rata	NA	NA	25,3	52,28	57,75	40,44	3,456	0,6985	0,202112269
1990	70536	SP	A.ly-rata	13,01	20,16	NA	31,69	48,35	45,74	1,706	0,2743	0,160785463
1979	70536	SP	A.ly-rata	NA	NA	14,97	16,95	26,39	24,9	0,791	0,0615	0,077749684
1995	70542	SP	A.ly-rata	NA	10,55	28,39	33,53	36,31	27,27	2,251	0,1826	0,081119502
1991	SP5	SP	A.ly-rata	15,17	17,29	18,2	31,57	26,19	28,17	0,726	0,1534	0,211294766
1996	70545	SP	A.ly-rata	18,55	37,13	43,7	45,75	52,15	40,41	2,653	0,502	0,189219751
1987	70536	SP	A.ly-rata	NA	7,67	8,5	15,07	23,78	20,76	1,06	0,0602	0,056792453
1994	70541	SP	A.ly-rata	NA	12,86	NA	NA	NA	NA	NA	NA	NA
1993	70540	SP	A.ly-rata	NA	12,92	19,85	40,19	39,9	38,79	1,855	0,3352	0,180700809
1992	70539	SP	A.ly-rata	17,21	24,16	21,81	39,34	44,9	34,88	1,766	0,4291	0,242978482
2031	SP5	SP	A.ly-rata	28,33	28,99	28,3	43,44	44,46	40,09	2,676	0,4365	0,163116592
2027	70536	SP	A.ly-rata	NA	16,17	17,8	18,48	23,79	26,54	0,741	0,0887	0,119703104
2033	70540	SP	A.ly-rata	NA	16,76	10,02	43,1	38,3	39,66	2,184	0,2911	0,133287546
2036	70545	SP	A.ly-rata	NA	NA	NA	9,23	24,01	NA	0,612	0,1003	0,163888889
2030	70538	SP	A.ly-rata	19,38	16,75	18,8	44,07	43,3	NA	2,763	0,4199	0,151972494
2035	70542	SP	A.ly-rata	14,72	22,38	24,2	50,22	37,4	34,47	1,188	0,2376	0,2

2037	SPH7	SP	A.ly-rata	17,97	31,14	30,1	25,94	NA	36,8	0,489	0,0892	0,182413088
2029	70537	SP	A.ly-rata	18,55	25,53	21,65	32,29	34,3	NA	1,099	0,2183	0,198635123
2071	SPH7	SP	A.ly-rata	NA	10,97	NA	NA	NA	NA	NA	NA	NA
2077	SPH7	SP	A.ly-rata	32,5	18,11	20,1	33,41	29,55	32,34	1,006	0,1734	0,172365805
2075	70542	SP	A.ly-rata	NA	23,35	25,5	47,91	58,95	43,97	4,315	0,7882	0,182665122
2069	70537	SP	A.ly-rata	48,11	41,76	60,1	37,75	48,19	30,44	2,136	0,4836	0,226404494
2070	70538	SP	A.ly-rata	NA	7,64	10,6	58,91	57,87	38,52	2,999	0,4884	0,162854285
2157	SPH7	SP	A.ly-rata	49,24	25,4	10,2	34,23	49,68	35,2	2,049	0,3063	0,149487555
2150	70538	SP	A.ly-rata	NA	17,8	19,61	44,34	46,48	35,71	2,99	0,5056	0,16909699
2152	70539	SP	A.ly-rata	47,17	35,98	21,41	23,4	46,75	25,58	1,573	0,3131	0,199046408
2149	70537	SP	A.ly-rata	51,04	47,41	36,2	38,01	44,38	40,64	2,681	0,5219	0,194666169
2153	70540	SP	A.ly-rata	13,05	28,45	38,1	44,62	49,39	48,65	2,82	0,4341	0,15393617
2156	70545	SP	A.ly-rata	NA	16,64	10,9	NA		NA	NA	NA	NA
2147	70536	SP	A.ly-rata	52,42	50,11	32,98	46,9	62,85	37,66	3,318	0,6391	0,192616034

Appendix Table 2: Candidate genome regions for selective sweep in SP and PL. For each candidate region, the chromosome, start and end position, as well as the length in kb is given. Also, the average CLR and Fst value for the whole region is provided.

Chr	Start	End	kb	CLR	Fst	win-dows	Population
1	1020482	1028479	7997	10,30	0,38	2	SP
1	1134441	1156433	21992	13,70	0,53	9	SP
1	1710234	1724228	13995	12,98	0,57	5	SP
1	1742222	1746221	3999	10,81	0,52	2	SP
1	1808198	1810198	1999	11,36	0,64	2	SP
1	1860180	1864178	3999	6,54	0,33	2	SP
1	2280028	2286026	5998	15,06	0,28	4	SP
1	2415979	2433973	17994	10,38	0,45	6	SP
1	2673887	2675886	1999	7,26	0,35	2	SP
1	2693879	2701876	7997	9,52	0,28	4	SP
1	2763854	2791844	27990	43,02	0,52	14	SP
1	2981776	2989773	7997	19,26	0,50	5	SP
1	3001769	3003768	1999	9,12	0,63	2	SP
1	3099733	3123725	23991	8,72	0,54	6	SP
1	4357280	4359280	1999	7,03	0,48	2	SP
1	4569204	4577201	7997	8,30	0,30	2	SP
1	4589197	4599193	9996	11,78	0,46	2	SP
1	4989053	4991052	1999	12,42	0,54	2	SP
1	5087018	5097014	9996	10,41	0,55	2	SP
1	5382911	5384910	1999	20,45	0,48	2	SP
1	5404903	5408902	3999	18,60	0,27	3	SP
1	5420897	5432893	11996	10,74	0,30	3	SP
1	6110649	6116647	5998	16,88	0,84	2	SP
1	6164630	6166629	1999	10,56	0,35	2	SP
1	7174266	7176265	1999	9,17	0,54	2	SP
1	7708074	7710073	1999	10,13	0,43	2	SP
1	8187901	8219890	31988	19,62	0,55	10	SP
1	8231885	8233885	1999	79,10	0,32	2	SP
1	8603751	8613748	9996	7,87	0,63	3	SP
1	8645736	8653733	7997	13,57	0,45	4	SP
1	8819674	8841666	21992	12,41	0,46	5	SP
1	8889649	8893647	3999	10,05	0,28	3	SP
1	8907642	8917638	9996	8,18	0,32	3	SP
1	8937631	8943629	5998	9,99	0,40	4	SP
1	9483435	9491432	7997	12,53	0,45	4	SP
1	10303140	10305139	1999	10,05	0,42	2	SP
1	10527059	10535056	7997	14,12	0,39	5	SP
1	10603032	10605031	1999	8,58	0,36	2	SP

1	11622665	11628662	5998	14,40	0,42	3	SP
1	14097773	14113768	15994	9,01	0,44	5	SP
1	14231725	14237723	5998	15,27	0,32	3	SP
1	14967460	14969459	1999	8,66	0,52	2	SP
1	15781167	15789164	7997	14,58	0,43	3	SP
1	16161030	16167028	5998	12,25	0,48	3	SP
1	17468560	17478556	9996	9,29	0,36	2	SP
1	17772450	17776449	3999	13,50	0,44	2	SP
1	18020361	18024360	3999	32,40	0,27	3	SP
1	18130321	18146316	15994	14,87	0,46	8	SP
1	21701036	21703035	1999	7,98	0,65	2	SP
1	22328810	22330809	1999	11,81	0,40	2	SP
1	22670687	22682682	11996	6,61	0,31	4	SP
1	22866616	22868615	1999	11,79	0,52	2	SP
1	23680323	23682322	1999	16,44	0,37	2	SP
1	23696317	23698317	1999	9,95	0,35	2	SP
1	24633980	24637978	3999	7,01	0,35	2	SP
1	25949506	25969499	19993	15,91	0,51	7	SP
1	26149434	26155432	5998	24,63	0,38	4	SP
1	26285385	26295382	9996	11,29	0,49	4	SP
1	26419337	26429333	9996	13,17	0,46	5	SP
1	26477316	26485313	7997	8,81	0,48	4	SP
1	26541293	26547291	5998	10,54	0,47	2	SP
1	26563285	26565284	1999	13,15	0,60	2	SP
1	26671246	26675245	3999	10,96	0,45	3	SP
1	27099092	27107089	7997	10,92	0,39	5	SP
1	28484593	28490591	5998	18,11	0,33	2	SP
1	29352281	29364277	11996	12,50	0,41	3	SP
1	29686161	29700156	13995	15,10	0,49	3	SP
1	29766132	29776128	9996	15,73	0,59	5	SP
1	30675804	30677804	1999	25,75	0,35	2	SP
1	31237602	31239601	1999	9,96	0,39	2	SP
1	31251597	31265592	13995	39,66	0,36	8	SP
2	249464	253464	4000	15,20	0,41	3	SP
2	361465	367465	6000	8,06	0,35	2	SP
2	549467	553467	4000	18,33	0,34	2	SP
2	3595500	3599500	4000	80,17	0,33	3	SP
2	3633501	3639501	6000	15,73	0,34	4	SP
2	3905504	3909504	4000	18,80	0,47	2	SP
2	4123506	4129506	6000	29,97	0,34	3	SP
2	7913547	7919548	6000	8,55	0,32	4	SP
2	9603566	9605566	2000	15,06	0,28	2	SP
2	9671567	9675567	4000	9,95	0,44	3	SP
2	11311585	11313585	2000	7,54	0,56	2	SP
2	11881591	11883591	2000	14,45	0,36	2	SP
2	11981592	11983592	2000	16,35	0,55	2	SP
2	12179594	12213594	34000	21,40	0,59	12	SP

2	12859601	12865602	6000	9,66	0,44	3	SP
2	12955603	12957603	2000	13,32	0,82	2	SP
2	13053604	13057604	4000	9,37	0,37	2	SP
2	13073604	13087604	14000	7,90	0,43	3	SP
2	13127604	13155605	28000	28,52	0,59	13	SP
2	13189605	13233606	44000	14,62	0,48	15	SP
2	13577609	13579609	2000	8,62	0,29	2	SP
2	13943613	13967614	24000	40,58	0,50	10	SP
2	14033614	14039614	6000	11,11	0,74	4	SP
2	15543631	15545631	2000	7,37	0,44	2	SP
2	16189638	16195638	6000	6,40	0,35	2	SP
2	16795644	16797645	2000	7,93	0,46	2	SP
2	17431651	17439652	8000	9,11	0,37	2	SP
2	17743655	17751655	8000	6,98	0,36	2	SP
2	18389662	18391662	2000	60,60	0,58	2	SP
3	28278	30278	2000	6,01	0,57	2	SP
3	180285	186285	6000	22,08	0,59	4	SP
3	322291	328291	6000	7,18	0,51	2	SP
3	346292	350292	4000	10,18	0,51	2	SP
3	970319	974320	4000	10,32	0,45	3	SP
3	1352336	1354336	2000	6,23	0,63	2	SP
3	2346379	2360380	14001	22,44	0,34	5	SP
3	2830401	2832401	2000	9,81	0,44	2	SP
3	2846401	2860402	14001	28,09	0,40	5	SP
3	3348423	3360424	12001	21,48	0,49	4	SP
3	3388425	3390425	2000	17,88	0,67	2	SP
3	3512430	3514430	2000	6,24	0,39	2	SP
3	3756441	3758441	2000	14,43	0,42	2	SP
3	4770485	4780486	10000	20,62	0,46	6	SP
3	4906491	4914491	8000	17,89	0,49	3	SP
3	4930492	4932492	2000	17,30	0,42	2	SP
3	5570520	5574520	4000	17,17	0,30	3	SP
3	6932580	6936580	4000	8,56	0,58	3	SP
3	7208592	7212592	4000	7,38	0,39	3	SP
3	7304596	7306596	2000	9,34	0,51	2	SP
3	7654611	7658611	4000	17,22	0,91	3	SP
3	7692613	7696613	4000	7,05	0,44	2	SP
3	7714614	7716614	2000	18,35	0,33	2	SP
3	7740615	7752615	12001	50,27	0,54	6	SP
3	8336641	8342641	6000	13,96	0,35	2	SP
3	8410644	8412644	2000	7,64	0,34	2	SP
3	8498648	8500648	2000	6,42	0,62	2	SP
3	8972669	8974669	2000	13,33	0,48	2	SP
3	9130675	9132676	2000	9,14	0,78	2	SP
3	9230680	9232680	2000	13,71	0,43	2	SP
3	10536737	10538737	2000	19,15	0,35	2	SP
3	10796748	10798748	2000	9,13	0,33	2	SP

3	11618784	11620784	2000	11,50	0,51	2	SP
3	13518867	13522867	4000	5,75	0,34	2	SP
3	14042890	14048890	6000	11,90	0,39	4	SP
3	14412906	14416906	4000	21,73	0,35	3	SP
3	17435038	17439038	4000	31,80	0,48	3	SP
3	18435081	18437082	2000	45,00	0,43	2	SP
3	18831099	18833099	2000	47,50	0,36	2	SP
3	20781184	20783184	2000	101,35	0,26	2	SP
3	20909189	20911190	2000	22,55	0,38	2	SP
3	21033195	21039195	6000	13,09	0,38	3	SP
3	21099198	21103198	4000	8,23	0,41	2	SP
3	21359209	21365209	6000	8,29	0,34	2	SP
3	22137243	22139243	2000	18,00	0,42	2	SP
4	900093	906093	6000	34,23	0,40	3	SP
4	2106089	2114089	8000	12,38	0,47	3	SP
4	2420088	2426088	6000	13,88	0,26	2	SP
4	2964087	2976087	12000	14,76	0,40	5	SP
4	3052087	3054087	2000	24,32	0,46	2	SP
4	3142086	3148086	6000	10,36	0,51	2	SP
4	3400085	3408085	8000	9,09	0,48	2	SP
4	4842081	4848081	6000	22,25	0,43	4	SP
4	5028080	5030080	2000	30,45	0,59	2	SP
4	7020074	7024074	4000	8,09	0,32	2	SP
4	10254064	10272064	18000	8,91	0,41	4	SP
4	10310064	10328064	18000	82,94	0,36	7	SP
4	10622063	10628063	6000	8,16	0,35	2	SP
4	11550060	11554060	4000	8,81	0,52	3	SP
4	13200055	13216055	16000	195,59	0,29	5	SP
4	14290052	14292052	2000	8,55	0,48	2	SP
4	14328051	14334051	6000	11,76	0,64	2	SP
4	15006049	15008049	2000	32,75	0,71	2	SP
4	16148046	16154046	6000	17,86	0,35	3	SP
4	16524045	16534045	10000	20,27	0,59	6	SP
4	18822037	18836037	14000	16,00	0,34	5	SP
4	19674035	19676035	2000	14,30	0,33	2	SP
4	19864034	19872034	8000	13,43	0,42	5	SP
4	20028034	20036034	8000	11,04	0,41	2	SP
4	20104033	20108033	4000	10,47	0,48	2	SP
4	20122033	20132033	10000	8,14	0,50	3	SP
4	20644032	20648032	4000	39,55	0,68	2	SP
4	21374029	21404029	30000	41,44	0,43	14	SP
4	21796028	21798028	2000	9,66	0,45	2	SP
4	22252027	22254027	2000	30,80	0,43	2	SP
4	22304027	22314027	10000	11,70	0,51	4	SP
4	23028024	23032024	4000	8,00	0,77	3	SP
5	694302	698302	3999	10,01	0,31	2	SP
5	2189984	2191983	2000	9,89	0,43	2	SP

5	2283964	2291962	7998	8,10	0,30	2	SP
5	2315957	2321956	5999	11,78	0,44	4	SP
5	2355949	2367946	11997	15,62	0,34	3	SP
5	2387942	2389941	2000	12,63	0,33	2	SP
5	2421935	2423934	2000	18,45	0,44	2	SP
5	2657884	2659884	2000	10,44	0,28	2	SP
5	2897833	2901832	3999	10,80	0,47	3	SP
5	2923828	2927827	3999	9,47	0,45	3	SP
5	3805640	3809639	3999	5,91	0,40	2	SP
5	4225550	4229549	3999	15,49	0,44	3	SP
5	4241547	4247546	5999	21,70	0,40	4	SP
5	4595472	4597471	2000	11,12	0,38	2	SP
5	4661457	4667456	5999	5,94	0,36	2	SP
5	4861415	4869413	7998	17,40	0,43	4	SP
5	5389302	5391302	2000	26,69	0,31	2	SP
5	5865201	5867201	2000	6,58	0,30	2	SP
5	5893195	5909192	15997	9,68	0,42	3	SP
5	9528421	9530420	2000	8,49	0,46	2	SP
5	10068306	10072305	3999	15,40	0,37	2	SP
5	12291832	12297831	5999	16,48	0,49	4	SP
5	12431802	12439801	7998	21,09	0,75	5	SP
5	13045672	13057669	11997	20,83	0,52	3	SP
5	13511572	13513572	2000	9,75	0,30	2	SP
5	13525569	13529569	3999	7,88	0,32	3	SP
5	13595555	13599554	3999	10,28	0,61	3	SP
5	13763519	13765518	2000	15,60	0,34	2	SP
5	13985472	13989471	3999	21,90	0,37	2	SP
5	14345395	14355393	9998	16,83	0,27	2	SP
5	16412954	16422952	9998	22,86	0,34	5	SP
5	16444948	16454946	9998	10,17	0,49	2	SP
5	17350755	17354754	3999	11,37	0,40	2	SP
5	17382748	17392746	9998	7,55	0,45	2	SP
5	17476728	17478727	2000	17,25	0,35	2	SP
5	17694681	17696681	2000	8,70	0,52	2	SP
5	17826653	17832652	5999	26,58	0,48	4	SP
5	17890640	17898638	7998	7,42	0,58	2	SP
5	18460518	18464517	3999	9,10	0,46	2	SP
5	19130376	19136374	5999	19,95	0,38	4	SP
5	19274345	19306338	31993	25,03	0,51	12	SP
5	19320335	19332333	11997	19,63	0,40	5	SP
5	19462305	19464305	2000	9,83	0,53	2	SP
5	19684258	19694256	9998	9,73	0,54	4	SP
5	20558072	20560071	2000	9,20	0,59	2	SP
6	372407	380407	8000	12,83	0,53	3	SP
6	530413	532413	2000	5,82	0,56	2	SP
6	1080432	1082432	2000	13,05	0,33	2	SP
6	1254438	1266439	12000	14,34	0,41	3	SP

6	1280439	1288440	8000	11,61	0,39	2	SP
6	2002465	2004465	2000	25,20	0,67	2	SP
6	3310511	3312511	2000	34,45	0,31	2	SP
6	3552520	3560520	8000	8,01	0,31	2	SP
6	5098574	5102575	4000	7,12	0,33	3	SP
6	5316582	5318582	2000	21,00	0,52	2	SP
6	6630629	6632629	2000	24,05	0,32	2	SP
6	7178648	7180648	2000	6,65	0,59	2	SP
6	7200649	7202649	2000	7,13	0,37	2	SP
6	7236650	7238650	2000	11,59	0,40	2	SP
6	7744668	7746668	2000	10,85	0,40	2	SP
6	8290687	8300688	10000	16,55	0,36	4	SP
6	8842707	8854707	12000	30,95	0,37	3	SP
6	8886709	8888709	2000	10,25	0,69	2	SP
6	9512731	9530731	18001	17,26	0,48	6	SP
6	9830742	9852743	22001	29,45	0,46	11	SP
6	9922745	9924745	2000	9,39	0,38	2	SP
6	10194755	10198755	4000	30,70	0,43	3	SP
6	11150789	11154789	4000	8,75	0,60	3	SP
6	17203003	17205003	2000	22,10	0,30	2	SP
6	17227004	17231004	4000	10,35	0,51	2	SP
6	18211039	18217039	6000	28,08	0,26	4	SP
6	18501049	18505049	4000	22,92	0,38	2	SP
6	18855061	18859062	4000	16,03	0,43	3	SP
6	18943065	18949065	6000	7,99	0,48	2	SP
6	19687091	19693091	6000	17,63	0,31	4	SP
6	20211109	20219110	8000	8,45	0,52	5	SP
6	20411117	20413117	2000	6,88	0,56	2	SP
6	20491119	20495120	4000	29,00	0,38	3	SP
6	20755129	20767129	12000	18,09	0,46	4	SP
6	20865133	20869133	4000	18,71	0,55	3	SP
6	21053139	21059139	6000	15,03	0,34	3	SP
6	21869168	21873168	4000	11,32	0,46	2	SP
6	21941171	21951171	10000	15,38	0,50	4	SP
6	23133213	23141213	8000	27,50	0,30	2	SP
6	23445224	23447224	2000	8,41	0,33	2	SP
6	23513226	23529227	16001	19,96	0,55	6	SP
6	23545227	23565228	20001	23,11	0,63	7	SP
6	23683232	23695233	12000	22,27	0,52	3	SP
6	23883239	23885240	2000	10,51	0,34	2	SP
6	24071246	24073246	2000	12,70	0,46	2	SP
6	24375257	24387257	12000	20,37	0,43	4	SP
7	1290947	1298947	8000	13,92	0,45	5	SP
7	1330946	1334946	4000	75,33	0,33	3	SP
7	1802944	1804944	2000	12,35	0,46	2	SP
7	2728939	2742938	14000	7,89	0,44	3	SP
7	5272924	5274924	2000	33,25	0,57	2	SP

7	5754921	5764921	10000	23,03	0,36	3	SP
7	6338918	6340918	2000	12,11	0,34	2	SP
7	6992914	6998914	6000	27,13	0,42	3	SP
7	7498911	7504911	6000	6,67	0,47	2	SP
7	8608905	8610905	2000	20,36	0,35	2	SP
7	8986903	8990903	4000	13,22	0,38	3	SP
7	9738899	9744899	6000	11,53	0,38	2	SP
7	9862898	9868898	6000	14,64	0,47	3	SP
7	10228896	10234896	6000	21,07	0,38	3	SP
7	10366895	10370895	4000	72,80	0,41	3	SP
7	12120885	12140885	20000	19,83	0,40	6	SP
7	12196885	12222885	26000	18,97	0,45	10	SP
7	12770882	12778881	8000	9,39	0,32	3	SP
7	14650871	14698871	48000	134,87	0,44	23	SP
7	14746870	14750870	4000	12,10	0,33	3	SP
7	17896852	17898852	2000	13,16	0,31	2	SP
7	18364850	18380850	16000	22,00	0,38	4	SP
7	18650848	18654848	4000	8,06	0,32	3	SP
7	20542837	20544837	2000	8,39	0,55	2	SP
7	20590837	20592837	2000	13,05	0,35	2	SP
7	21120834	21122834	2000	45,75	0,35	2	SP
7	23480821	23502821	22000	31,13	0,47	6	SP
7	23606820	23616820	10000	8,00	0,41	2	SP
7	23736819	23738819	2000	29,54	0,29	2	SP
7	24620814	24622814	2000	7,52	0,29	2	SP
8	490985	498984	7999	124,20	0,42	5	SP
8	1160926	1176925	15999	9,96	0,43	7	SP
8	1450900	1452900	2000	10,51	0,29	2	SP
8	1732876	1738875	5999	9,79	0,35	2	SP
8	1832867	1838866	5999	15,21	0,46	4	SP
8	2070846	2072846	2000	12,97	0,34	2	SP
8	2144839	2146839	2000	12,10	0,29	2	SP
8	2166837	2168837	2000	8,08	0,50	2	SP
8	2230832	2238831	7999	19,28	0,30	4	SP
8	3576713	3584712	7999	7,95	0,36	2	SP
8	3608710	3614710	5999	8,54	0,48	2	SP
8	3854689	3888686	33997	22,76	0,35	15	SP
8	3928682	3932682	4000	11,70	0,41	2	SP
8	4070669	4078669	7999	11,44	0,34	3	SP
8	4230655	4242654	11999	25,43	0,37	3	SP
8	4288650	4296650	7999	13,35	0,32	3	SP
8	4420639	4422638	2000	9,96	0,61	2	SP
8	4690615	4696614	5999	26,24	0,48	4	SP
8	5966502	5974502	7999	10,84	0,42	4	SP
8	6008499	6012498	4000	14,00	0,30	3	SP
8	11366026	11368026	2000	13,61	0,28	2	SP
8	11991971	11993971	2000	25,45	0,48	2	SP

8	12231950	12243949	11999	33,80	0,41	3	SP
8	14141782	14159780	17998	6,83	0,46	4	SP
8	14335765	14337765	2000	5,92	0,29	2	SP
8	14715731	14719731	4000	12,90	0,61	3	SP
8	14735729	14739729	4000	14,57	0,53	3	SP
8	14973708	14975708	2000	11,09	0,39	2	SP
8	15399671	15401671	2000	6,25	0,49	2	SP
8	17007529	17009529	2000	8,04	0,52	2	SP
8	17063524	17073523	9999	13,09	0,41	3	SP
8	17731465	17765462	33997	22,34	0,49	11	SP
8	17813458	17827457	13999	11,78	0,63	6	SP
8	17923448	17927448	4000	19,23	0,58	3	SP
8	20487222	20489222	2000	13,80	0,32	2	SP
8	21075171	21077170	2000	10,85	0,48	2	SP
8	22223069	22231069	7999	16,94	0,44	5	SP
1	188032	190032	2000	4,11	0,50	2	PL
1	1427936	1429936	2000	16,05	0,31	2	PL
1	1795907	1805906	9999	5,17	0,57	2	PL
1	6101572	6105572	4000	9,35	0,69	3	PL
1	6123570	6125570	2000	17,60	0,66	2	PL
1	8805362	8813361	7999	23,32	0,65	5	PL
1	9443312	9447312	4000	21,52	0,48	2	PL
1	11339165	11349164	9999	7,50	0,35	2	PL
1	11433157	11435157	2000	15,25	0,26	2	PL
1	11681138	11683138	2000	4,75	0,56	2	PL
1	17576679	17584678	7999	15,83	0,34	3	PL
1	22152323	22154323	2000	26,90	0,56	2	PL
1	22834270	22836270	2000	5,63	0,72	2	PL
1	24754120	24756120	2000	22,80	0,31	2	PL
1	25024099	25030099	6000	7,74	0,49	3	PL
1	25744043	25748043	4000	20,54	0,55	2	PL
1	26072018	26074018	2000	12,20	0,73	2	PL
1	26841958	26845958	4000	4,24	0,57	3	PL
1	29589744	29593744	4000	6,10	0,51	2	PL
1	29647740	29659739	11999	11,14	0,72	4	PL
1	29673738	29679737	6000	5,87	0,73	3	PL
1	30903642	30915641	11999	6,98	0,47	4	PL
2	3370791	3372791	2000	13,92	0,27	2	PL
2	11167248	11171248	4000	31,32	0,55	3	PL
2	15455499	15461499	6000	16,75	0,33	4	PL
2	15579506	15581506	2000	10,86	0,52	2	PL
2	16155540	16159540	4000	30,23	0,78	3	PL
2	17495618	17501618	6000	11,59	0,73	3	PL
2	18711689	18717690	6000	24,25	0,54	2	PL
2	18831696	18833696	2000	18,65	0,48	2	PL
3	2120309	2128309	8000	12,82	0,75	3	PL
3	2344319	2348319	4000	5,79	0,42	2	PL

3	2974348	2986349	12001	18,20	0,51	4	PL
3	4790433	4794433	4000	5,06	0,66	2	PL
3	6764525	6784526	20001	10,10	0,63	6	PL
3	7126541	7128542	2000	8,72	0,31	2	PL
3	7386554	7394554	8000	33,90	0,41	2	PL
3	11086726	11088726	2000	31,10	0,49	2	PL
3	13984860	13992861	8000	6,72	0,35	3	PL
3	18831086	18833086	2000	20,70	0,36	2	PL
3	23635309	23639309	4000	61,30	0,71	3	PL
3	23797317	23801317	4000	12,05	0,59	2	PL
4	1348079	1352079	4000	9,30	0,57	2	PL
4	1556079	1562079	6000	9,31	0,48	4	PL
4	3646080	3648080	2000	7,62	0,60	2	PL
4	4786081	4788081	2000	6,50	0,31	2	PL
4	5604082	5606082	2000	6,92	0,32	2	PL
4	11038085	11044085	6000	7,48	0,70	3	PL
4	12934087	12936087	2000	13,14	0,46	2	PL
4	14114087	14116087	2000	11,50	0,44	2	PL
4	16388089	16390089	2000	10,60	0,48	2	PL
4	16606089	16608089	2000	10,04	0,60	2	PL
4	16830089	16834089	4000	6,41	0,39	2	PL
4	17470090	17472090	2000	11,87	0,45	2	PL
4	19950091	19952091	2000	25,00	0,39	2	PL
4	22864093	22866093	2000	14,50	0,69	2	PL
5	3748665	3754662	5996	13,95	0,46	2	PL
5	5157830	5163826	5996	23,28	0,32	4	PL
5	13049150	13051148	1999	17,65	0,54	2	PL
5	13336979	13338978	1999	4,52	0,53	2	PL
5	13376955	13378954	1999	4,95	0,40	2	PL
5	16766945	16768944	1999	6,77	0,49	2	PL
5	16842900	16854893	11993	7,66	0,34	3	PL
5	18290041	18298037	7995	4,68	0,44	2	PL
5	19293446	19295445	1999	6,68	0,69	2	PL
5	19721193	19729188	7995	5,78	0,57	2	PL
5	20070985	20072984	1999	11,35	0,64	2	PL
5	20090973	20096970	5996	44,81	0,27	4	PL
6	1474430	1476430	2000	5,45	0,62	2	PL
6	1606433	1608434	2000	17,41	0,66	2	PL
6	5092521	5100522	8000	19,45	0,41	5	PL
6	6342553	6344553	2000	20,35	0,40	2	PL
6	7712587	7716588	4000	5,79	0,44	2	PL
6	13656737	13658737	2000	18,80	0,42	2	PL
6	18760866	18762866	2000	13,38	0,91	2	PL
6	21534936	21538936	4000	70,53	0,27	3	PL
6	22032948	22038949	6000	10,16	0,61	2	PL
6	22222953	22226953	4000	8,88	0,57	2	PL
6	22466959	22468959	2000	11,01	0,33	2	PL

6	23550987	23554987	4000	13,55	0,65	2	PL
6	23818993	23820993	2000	7,37	0,49	2	PL
6	23914996	23916996	2000	21,10	0,72	2	PL
7	148731	156732	8000	23,70	0,40	2	PL
7	1310751	1316751	6000	5,00	0,43	2	PL
7	1890761	1892761	2000	7,94	0,33	2	PL
7	2564772	2566772	2000	4,76	0,53	2	PL
7	3054780	3058780	4000	21,20	0,36	2	PL
7	11136916	11140916	4000	10,59	0,40	2	PL
7	17897029	17899029	2000	18,15	0,31	2	PL
7	18391038	18393038	2000	13,16	0,31	2	PL
7	20695076	20699076	4000	5,63	0,36	2	PL
8	694447	698448	4000	8,56	0,51	2	PL
8	1730479	1732479	2000	26,83	0,35	2	PL
8	2394500	2402500	8000	8,84	0,58	3	PL
8	2902515	2906516	4000	16,53	0,70	3	PL
8	3610537	3614537	4000	14,96	0,45	2	PL
8	6310620	6316621	6000	8,24	0,31	2	PL
8	15480903	15488903	8000	6,59	0,34	2	PL
8	16024920	16026920	2000	41,00	0,30	2	PL
8	16370931	16372931	2000	5,93	0,43	2	PL
8	16620938	16624938	4000	8,63	0,28	3	PL
8	17470964	17476965	6000	20,95	0,38	2	PL
8	19777036	19781036	4000	11,38	0,63	2	PL
8	19803036	19813037	10000	108,60	0,56	6	PL
8	20601061	20603061	2000	10,30	0,29	2	PL
8	21649093	21651093	2000	6,30	0,65	2	PL
8	22473119	22475119	2000	16,38	0,37	2	PL

Appendix Table 3: Gene ontology enrichment analysis for decreasing values of additive variance. The GOs ID, the term and the p value of the significant GOs ($p < 0.03813$) are given.

GO.ID	**Term**	***P* value**
GO:0008150	biological_process	5.4e-12
GO:0006007	glucose catabolic process	8.3e-06
GO:0006096	glycolytic process	1.6e-05
GO:0006094	gluconeogenesis	9.1e-05
GO:0046686	response to cadmium ion	0.00015
GO:0080129	proteasome core complex assembly	0.00016
GO:0006511	ubiquitin-dependent protein catabolic pr...	0.00016
GO:0051788	response to misfolded protein	0.00077
GO:0009853	photorespiration	0.00099
GO:0006412	translation	0.00231
GO:0006626	protein targeting to mitochondrion	0.00243
GO:0010498	proteasomal protein catabolic process	0.00382
GO:0000162	tryptophan biosynthetic process	0.00545
GO:0009805	coumarin biosynthetic process	0.00626
GO:0006575	cellular modified amino acid metabolic p...	0.00627
GO:0019760	glucosinolate metabolic process	0.00788
GO:0042542	response to hydrogen peroxide	0.00846
GO:0009651	response to salt stress	0.00865
GO:0009610	response to symbiotic fungus	0.00903
GO:0055114	oxidation-reduction process	0.00921
GO:0019395	fatty acid oxidation	0.01062
GO:1901566	organonitrogen compound biosynthetic pro...	0.01064
GO:0034440	lipid oxidation	0.01098
GO:0042180	cellular ketone metabolic process	0.01139
GO:0050801	ion homeostasis	0.01232

GO:0006635	fatty acid beta-oxidation	0.01251
GO:0009812	flavonoid metabolic process	0.01263
GO:0007030	Golgi organization	0.01349
GO:0010256	endomembrane system organization	0.01410
GO:0000041	transition metal ion transport	0.01428
GO:0030258	lipid modification	0.01529
GO:0009408	response to heat	0.01625
GO:0031365	N-terminal protein amino acid modificati...	0.01710
GO:0006498	N-terminal protein lipidation	0.01733
GO:0006499	N-terminal protein myristoylation	0.01733
GO:0006520	cellular amino acid metabolic process	0.01776
GO:0044272	sulfur compound biosynthetic process	0.01900
GO:0072329	monocarboxylic acid catabolic process	0.01904
GO:0048878	chemical homeostasis	0.01905
GO:0006873	cellular ion homeostasis	0.01940
GO:1901605	alpha-amino acid metabolic process	0.01967
GO:1901420	negative regulation of response to alcoh...	0.01988
GO:0009788	negative regulation of abscisic acid-act...	0.01988
GO:1905958	negative regulation of cellular response...	0.01988
GO:0016144	S-glycoside biosynthetic process	0.02042
GO:0019761	glucosinolate biosynthetic process	0.02042
GO:0019758	glycosinolate biosynthetic process	0.02042
GO:0046482	para-aminobenzoic acid metabolic process	0.02197
GO:0022904	respiratory electron transport chain	0.02218

GO:0009062	fatty acid catabolic process	0.02316
GO:0030003	cellular cation homeostasis	0.02317
GO:0000096	sulfur amino acid metabolic process	0.02332
GO:0006497	protein lipidation	0.02407
GO:0042157	lipoprotein metabolic process	0.02407
GO:0042158	lipoprotein biosynthetic process	0.02407
GO:0062012	regulation of small molecule metabolic p...	0.02437
GO:0044281	small molecule metabolic process	0.02465
GO:0044282	small molecule catabolic process	0.02651
GO:0009813	flavonoid biosynthetic process	0.02669
GO:0000097	sulfur amino acid biosynthetic process	0.02689
GO:0006820	anion transport	0.02731
GO:0055082	cellular chemical homeostasis	0.02735
GO:0019752	carboxylic acid metabolic process	0.02824
GO:0009060	aerobic respiration	0.02833
GO:0006082	organic acid metabolic process	0.02871
GO:0018377	protein myristoylation	0.02877
GO:0046351	disaccharide biosynthetic process	0.02878
GO:0044273	sulfur compound catabolic process	0.02948
GO:0022613	ribonucleoprotein complex biogenesis	0.02958
GO:0006833	water transport	0.03039
GO:0042044	fluid transport	0.03039
GO:0042773	ATP synthesis coupled electron transport	0.03137
GO:0042775	mitochondrial ATP synthesis coupled elec...	0.03137
GO:0043436	oxoacid metabolic process	0.03177

GO:0006576	cellular biogenic amine metabolic proces...	0.03278
GO:0042254	ribosome biogenesis	0.03338
GO:0046395	carboxylic acid catabolic process	0.03374
GO:0016054	organic acid catabolic process	0.03374
GO:0055080	cation homeostasis	0.03403
GO:1901659	glycosyl compound biosynthetic process	0.03423
GO:0019321	pentose metabolic process	0.03462
GO:0042732	D-xylose metabolic process	0.03462
GO:0046149	pigment catabolic process	0.03562
GO:0051596	methylglyoxal catabolic process	0.03602
GO:0009438	methylglyoxal metabolic process	0.03602
GO:0046185	aldehyde catabolic process	0.03602
GO:0042182	ketone catabolic process	0.03602
GO:0006661	phosphatidylinositol biosynthetic proces...	0.03637
GO:0009963	positive regulation of flavonoid biosynt...	0.03637
GO:0009269	response to desiccation	0.03800

Appendix Table 4: Gene ontology enrichment analysis for decreasing values of dominance variance. The GOs ID, the term and the p value of the significant GOs ($p < 0.03810$) are given.

GO.ID	**Term**	***P* value**
GO:0031048	chromatin silencing by small RNA	4.5e-12
GO:0000911	cytokinesis by cell plate formation	2.0e-11
GO:0006306	DNA methylation	4.3e-11
GO:0000956	nuclear-transcribed mRNA catabolic proce...	9.2e-11
GO:0006346	methylation-dependent chromatin silencin...	6.3e-10
GO:0008150	biological_process	2.1e-09
GO:0045010	actin nucleation	3.5e-09
GO:0010090	trichome morphogenesis	5.6e-09
GO:0051567	histone H3-K9 methylation	1.6e-08
GO:0007062	sister chromatid cohesion	1.8e-08
GO:0009630	gravitropism	3.1e-08
GO:0035196	production of miRNAs involved in gene si...	1.3e-07
GO:0007131	reciprocal meiotic recombination	7.9e-07
GO:0051225	spindle assembly	8.8e-07
GO:0006275	regulation of DNA replication	1.1e-06
GO:0010267	production of ta-siRNAs involved in RNA ...	1.2e-06
GO:0009616	virus induced gene silencing	2.0e-06
GO:0016572	histone phosphorylation	3.4e-06
GO:0016558	protein import into peroxisome matrix	3.7e-06
GO:0045132	meiotic chromosome segregation	4.8e-06

GO:0009909	regulation of flower development	5.4e-06
GO:0000278	mitotic cell cycle	5.5e-06
GO:0007267	cell-cell signaling	8.5e-06
GO:0000724	double-strand break repair via homologou...	9.0e-06
GO:0007129	synapsis	1.1e-05
GO:0015996	chlorophyll catabolic process	1.9e-05
GO:0007155	cell adhesion	2.1e-05
GO:0000226	microtubule cytoskeleton organization	2.6e-05
GO:0006635	fatty acid beta-oxidation	2.7e-05
GO:0006270	DNA replication initiation	4.0e-05
GO:0032204	regulation of telomere maintenance	6.2e-05
GO:0046855	inositol phosphate dephosphorylation	6.4e-05
GO:0043247	telomere maintenance in response to DNA ...	7.0e-05
GO:0010228	vegetative to reproductive phase transit...	7.8e-05
GO:0010389	regulation of G2/M transition of mitotic...	8.5e-05
GO:0042138	meiotic DNA double-strand break formatio...	8.9e-05
GO:0032957	inositol trisphosphate metabolic process	9.1e-05
GO:0010332	response to gamma radiation	9.1e-05
GO:0008283	cell proliferation	0.00018
GO:0009855	determination of bilateral symmetry	0.00025
GO:0043687	post-translational protein modification	0.00026

GO:0016926	protein desumoy-lation	0.00044
GO:0006342	chromatin silencing	0.00046
GO:0008380	RNA splicing	0.00050
GO:0006623	protein targeting to vacuole	0.00057
GO:0010264	myo-inositol hex-akisphosphate bio-synthet...	0.00062
GO:0006312	mitotic recombina-tion	0.00065
GO:0006486	protein glycosyla-tion	0.00067
GO:0016444	somatic cell DNA recombination	0.00089
GO:0010638	positive regulation of organelle or-ganiz...	0.00099
GO:0051276	chromosome organ-ization	0.00113
GO:0009887	animal organ mor-phogenesis	0.00120
GO:0016246	RNA interference	0.00126
GO:0006406	mRNA export from nucleus	0.00135
GO:0043414	macromolecule methylation	0.00146
GO:0008284	positive regulation of cell proliferatio...	0.00156
GO:0006606	protein import into nucleus	0.00158
GO:0010014	meristem initiation	0.00171
GO:0000280	nuclear division	0.00201
GO:0032875	regulation of DNA endoreduplication	0.00202
GO:0006665	sphingolipid meta-bolic process	0.00214
GO:0010050	vegetative phase change	0.00229
GO:0043543	protein acylation	0.00245
GO:0050665	hydrogen peroxide biosynthetic process	0.00270
GO:0016192	vesicle-mediated transport	0.00288

GO:0006487	protein N-linked glycosylation	0.00296
GO:0009640	photomorphogene-sis	0.00338
GO:0048573	photoperiodism, flowering	0.00357
GO:0030029	actin filament-based process	0.00362
GO:0006397	mRNA processing	0.00364
GO:0007346	regulation of mitotic cell cycle	0.00383
GO:0010564	regulation of cell cycle process	0.00400
GO:0031047	gene silencing by RNA	0.00400
GO:0048449	floral organ for-mation	0.00419
GO:0016570	histone modification	0.00441
GO:0009560	embryo sac egg cell differentiation	0.00441
GO:0010182	sugar mediated sig-naling pathway	0.00442
GO:0051726	regulation of cell cycle	0.00446
GO:1902275	regulation of chro-matin organization	0.00504
GO:0048440	carpel development	0.00623
GO:0010072	primary shoot apical meristem specifi-cat...	0.00652
GO:0010162	seed dormancy pro-cess	0.00703
GO:0016567	protein ubiquitina-tion	0.00725
GO:0051604	protein maturation	0.00728
GO:0051301	cell division	0.00770
GO:0006281	DNA repair	0.00791
GO:0010051	xylem and phloem pattern formation	0.00830
GO:0048451	petal formation	0.00949
GO:0048453	sepal formation	0.00949
GO:0042127	regulation of cell proliferation	0.00961
GO:0006497	protein lipidation	0.00973

GO:0072528	pyrimidine-containing compound biosynthe...	0.00996
GO:0000375	RNA splicing, via transesterification re...	0.01002
GO:0000377	RNA splicing, via transesterification re...	0.01002
GO:0006325	chromatin organization	0.01033
GO:0006261	DNA-dependent DNA replication	0.01090
GO:0031056	regulation of histone modification	0.01139
GO:0051168	nuclear export	0.01186
GO:0048229	gametophyte development	0.01219
GO:0018377	protein myristoylation	0.01237
GO:0031365	N-terminal protein amino acid modificati...	0.01237
GO:0043603	cellular amide metabolic process	0.01280
GO:2000242	negative regulation of reproductive proc...	0.01421
GO:0000398	mRNA splicing, via spliceosome	0.01444
GO:0019915	lipid storage	0.01513
GO:0070646	protein modification by small protein re...	0.01522
GO:0006498	N-terminal protein lipidation	0.01567
GO:0006499	N-terminal protein myristoylation	0.01567
GO:0006891	intra-Golgi vesicle-mediated transport	0.01834
GO:0010224	response to UV-B	0.01854
GO:0046519	sphingoid metabolic process	0.01864
GO:0030048	actin filament-based movement	0.01866

GO:0006928	movement of cell or subcellular compo-nen...	0.01866
GO:0010212	response to ionizing radiation	0.01882
GO:0043604	amide biosynthetic process	0.01902
GO:0070918	production of small RNA involved in gene...	0.01916
GO:0031050	dsRNA processing	0.01916
GO:0097435	supramolecular fi-ber organization	0.01928
GO:0051302	regulation of cell di-vision	0.01968
GO:0006220	pyrimidine nucleo-tide metabolic pro-cess	0.01980
GO:0006221	pyrimidine nucleo-tide biosynthetic proce...	0.01980
GO:0009220	pyrimidine ribonu-cleotide biosyn-thetic p...	0.01980
GO:0009218	pyrimidine ribonu-cleotide metabolic proc...	0.01980
GO:0016579	protein deubiquiti-nation	0.01985
GO:0009943	adaxial/abaxial axis specification	0.02030
GO:0009944	polarity specifica-tion of adaxial/ab-axia...	0.02030
GO:0065001	specification of axis polarity	0.02030
GO:0003002	regionalization	0.02159
GO:0009553	embryo sac devel-opment	0.02206
GO:0009955	adaxial/abaxial pat-tern specification	0.02224
GO:0001510	RNA methylation	0.02231
GO:0044275	cellular carbohy-drate catabolic pro-cess	0.02232

GO:0006518	peptide metabolic process	0.02252
GO:0010048	vernalization response	0.02275
GO:0000303	response to superoxide	0.02288
GO:0000305	response to oxygen radical	0.02288
GO:0045595	regulation of cell differentiation	0.02413
GO:0009910	negative regulation of flower developmen...	0.02466
GO:0032504	multicellular organism reproduction	0.02581
GO:0000394	RNA splicing, via endonucleolytic cleava...	0.02586
GO:0009057	macromolecule catabolic process	0.02587
GO:0009311	oligosaccharide metabolic process	0.02601
GO:0046520	sphingoid biosynthetic process	0.02609
GO:0048439	flower morphogenesis	0.02634
GO:0051338	regulation of transferase activity	0.02657
GO:0033044	regulation of chromosome organization	0.02670
GO:0043161	proteasome-mediated ubiquitin-dependent ...	0.02694
GO:0010876	lipid localization	0.02704
GO:0010351	lithium ion transport	0.02797
GO:0051093	negative regulation of developmental pro...	0.02803
GO:0006611	protein export from nucleus	0.02809
GO:0048481	plant ovule development	0.02810
GO:0035670	plant-type ovary development	0.02810

GO:0001676	long-chain fatty acid metabolic process	0.02849
GO:0005984	disaccharide metabolic process	0.02861
GO:0009798	axis specification	0.02964
GO:0043043	peptide biosynthetic process	0.02966
GO:0006259	DNA metabolic process	0.02975
GO:0006897	endocytosis	0.03171
GO:0098657	import into cell	0.03171
GO:0048193	Golgi vesicle transport	0.03182
GO:0006412	translation	0.03248
GO:0050826	response to freezing	0.03377
GO:0009410	response to xenobiotic stimulus	0.03426
GO:0007031	peroxisome organization	0.03434
GO:0016571	histone methylation	0.03488
GO:0006508	proteolysis	0.03595
GO:0050657	nucleic acid transport	0.03597
GO:0050658	RNA transport	0.03597
GO:0006403	RNA localization	0.03597
GO:0006405	RNA export from nucleus	0.03597
GO:0051236	establishment of RNA localization	0.03597
GO:0065009	regulation of molecular function	0.03611
GO:0000272	polysaccharide catabolic process	0.03805

Appendix Table 5: Gene ontology enrichment analysis for the genes included in the cluster with high coregulation. The GOs ID, the term and the p value of the significant GOs ($p < 0.077$) are given.

GO.ID	**Term**	***P* value**
GO:0009630	gravitropism	6.4e-28
GO:0010090	trichome morpho-genesis	8.9e-15
GO:0006486	protein glycosyla-tion	9.7e-15
GO:0006487	protein N-linked glycosylation	2.0e-14
GO:0000956	nuclear-transcribed mRNA catabolic proce...	5.8e-14
GO:0010228	vegetative to repro-ductive phase transit...	4.0e-12
GO:0007155	cell adhesion	1.4e-11
GO:0007062	sister chromatid co-hesion	4.8e-11
GO:0045010	actin nucleation	1.1e-10
GO:0016926	protein desumoy-lation	2.4e-09
GO:0016567	protein ubiquitina-tion	2.5e-09
GO:0016579	protein deubiquiti-nation	2.5e-08
GO:0050665	hydrogen peroxide biosynthetic process	2.8e-08
GO:0048193	Golgi vesicle transport	3.8e-08
GO:0010182	sugar mediated sig-naling pathway	4.7e-08
GO:0010072	primary shoot api-cal meristem spec-ificat...	1.6e-07
GO:0033044	regulation of chro-mosome organiza-tion	5.2e-07
GO:0035196	production of miR-NAs involved in gene si...	5.6e-07

GO:0031048	chromatin silencing by small RNA	6.1e-07
GO:0006897	endocytosis	6.3e-07
GO:0000911	cytokinesis by cell plate formation	7.6e-07
GO:0007010	cytoskeleton organization	8.6e-07
GO:0009909	regulation of flower development	9.3e-07
GO:0009887	animal organ morphogenesis	1.0e-06
GO:0045132	meiotic chromosome segregation	1.0e-06
GO:0045595	regulation of cell differentiation	1.1e-06
GO:0009640	photomorphogenesis	1.5e-06
GO:0009845	seed germination	1.8e-06
GO:0016558	protein import into peroxisome matrix	2.0e-06
GO:0000278	mitotic cell cycle	2.3e-06
GO:0000226	microtubule cytoskeleton organization	2.6e-06
GO:0043687	post-translational protein modification	3.0e-06
GO:0050826	response to freezing	6.3e-06
GO:0008284	positive regulation of cell proliferatio...	6.7e-06
GO:0010332	response to gamma radiation	6.8e-06
GO:0010048	vernalization response	8.5e-06
GO:0030244	cellulose biosynthetic process	1.9e-05
GO:0009880	embryonic pattern specification	2.6e-05
GO:0043247	telomere maintenance in response to DNA ...	2.9e-05
GO:0006397	mRNA processing	2.9e-05
GO:0019915	lipid storage	4.7e-05

GO:0042138	meiotic DNA double-strand break formatio...	6.2e-05
GO:0009616	virus induced gene silencing	6.4e-05
GO:0008150	biological_process	6.5e-05
GO:0032204	regulation of telomere maintenance	6.6e-05
GO:0048765	root hair cell differentiation	6.7e-05
GO:0010073	meristem maintenance	7.9e-05
GO:0010267	production of ta-siRNAs involved in RNA ...	9.5e-05
GO:0010162	seed dormancy process	0.00010
GO:0007131	reciprocal meiotic recombination	0.00011
GO:0010498	proteasomal protein catabolic process	0.00012
GO:0051301	cell division	0.00012
GO:0000398	mRNA splicing, via spliceosome	0.00013
GO:0010050	vegetative phase change	0.00014
GO:0009855	determination of bilateral symmetry	0.00018
GO:0045893	positive regulation of transcription, DN...	0.00027
GO:0000281	mitotic cytokinesis	0.00032
GO:0009832	plant-type cell wall biogenesis	0.00038
GO:0010638	positive regulation of organelle organiz...	0.00038
GO:0006094	gluconeogenesis	0.00046
GO:0006346	methylation-dependent chromatin silencin...	0.00048
GO:0010051	xylem and phloem pattern formation	0.00048
GO:0046855	inositol phosphate dephosphorylation	0.00053

GO:0016571	histone methylation	0.00054
GO:0048573	photoperiodism, flowering	0.00058
GO:0006366	transcription by RNA polymerase II	0.00061
GO:0051302	regulation of cell division	0.00061
GO:0007033	vacuole organization	0.00061
GO:0010014	meristem initiation	0.00066
GO:0032957	inositol trisphosphate metabolic process	0.00068
GO:0006499	N-terminal protein myristoylation	0.00074
GO:0010315	auxin efflux	0.00080
GO:0048825	cotyledon development	0.00081
GO:0030422	production of siRNA involved in RNA inte...	0.00083
GO:0051726	regulation of cell cycle	0.00098
GO:0006468	protein phosphorylation	0.00100
GO:0046777	protein autophosphorylation	0.00102
GO:0006306	DNA methylation	0.00122
GO:0009894	regulation of catabolic process	0.00139
GO:0006342	chromatin silencing	0.00154
GO:0035194	posttranscriptional gene silencing by RN...	0.00173
GO:0006635	fatty acid beta-oxidation	0.00177
GO:0051273	beta-glucan metabolic process	0.00215
GO:0007034	vacuolar transport	0.00224
GO:0048364	root development	0.00260
GO:0009793	embryo development ending in seed dorman...	0.00272

GO:0090056	regulation of chlorophyll metabolic proc...	0.00275
GO:0022904	respiratory electron transport chain	0.00275
GO:0048437	floral organ development	0.00276
GO:0009933	meristem structural organization	0.00302
GO:0016192	vesicle-mediated transport	0.00310
GO:0048229	gametophyte development	0.00342
GO:0009910	negative regulation of flower developmen...	0.00354
GO:0048580	regulation of post-embryonic development	0.00364
GO:0090698	post-embryonic plant morphogenesis	0.00381
GO:0016036	cellular response to phosphate starvatio...	0.00402
GO:0006310	DNA recombination	0.00407
GO:0007267	cell-cell signaling	0.00440
GO:0006418	tRNA aminoacylation for protein translat...	0.00468
GO:0016570	histone modification	0.00488
GO:0010629	negative regulation of gene expression	0.00491
GO:0042732	D-xylose metabolic process	0.00608
GO:0009908	flower development	0.00635
GO:0030865	cortical cytoskeleton organization	0.00637
GO:0007129	synapsis	0.00663
GO:0000394	RNA splicing, via endonucleolytic cleava...	0.00818
GO:0048767	root hair elongation	0.00847

GO:0016049	cell growth	0.00859
GO:0040007	growth	0.00947
GO:0032504	multicellular organism reproduction	0.00949
GO:0010558	negative regulation of macromolecule bio...	0.01012
GO:2000113	negative regulation of cellular macromol...	0.01012
GO:0065007	biological regulation	0.01022
GO:0042773	ATP synthesis coupled electron transport	0.01039
GO:0042775	mitochondrial ATP synthesis coupled elec...	0.01039
GO:0008356	asymmetric cell division	0.01039
GO:0031329	regulation of cellular catabolic process	0.01039
GO:0051641	cellular localization	0.01068
GO:0010564	regulation of cell cycle process	0.01076
GO:0031324	negative regulation of cellular metaboli...	0.01163
GO:0006891	intra-Golgi vesicle-mediated transport	0.01205
GO:0042546	cell wall biogenesis	0.01223
GO:0009926	auxin polar transport	0.01313
GO:0032879	regulation of localization	0.01334
GO:0050789	regulation of biological process	0.01394
GO:0070592	cell wall polysaccharide biosynthetic pr...	0.01400
GO:0070589	cellular component macromolecule biosynt...	0.01400

GO:0044038	cell wall macromol-ecule biosynthetic pro...	0.01400
GO:0048827	phyllome develop-ment	0.01434
GO:0006816	calcium ion transport	0.01438
GO:0031327	negative regulation of cellular bio-synth...	0.01442
GO:0040008	regulation of growth	0.01452
GO:0048439	flower morphogen-esis	0.01493
GO:0006312	mitotic recombina-tion	0.01517
GO:0098727	maintenance of cell number	0.01517
GO:0019827	stem cell population maintenance	0.01517
GO:0010413	glucuronoxylan metabolic process	0.01533
GO:2000280	regulation of root development	0.01585
GO:0009553	embryo sac devel-opment	0.01615
GO:0045492	xylan biosynthetic process	0.01712
GO:0031537	regulation of antho-cyanin metabolic proc...	0.01719
GO:0040029	regulation of gene expression, epige-neti...	0.01731
GO:0044265	cellular macromole-cule catabolic pro-cess	0.01746
GO:0009890	negative regulation of biosynthetic proc...	0.01801
GO:0016441	posttranscriptional gene silencing	0.01912
GO:0097435	supramolecular fi-ber organization	0.01942
GO:0051274	beta-glucan biosyn-thetic process	0.01965

GO:0032875	regulation of DNA endoreduplication	0.02018
GO:0090329	regulation of DNA-dependent DNA replicat...	0.02018
GO:0006623	protein targeting to vacuole	0.02037
GO:0072665	protein localization to vacuole	0.02037
GO:0072666	establishment of protein localization to...	0.02037
GO:0045491	xylan metabolic process	0.02119
GO:0090567	reproductive shoot system develop-ment	0.02127
GO:0051168	nuclear export	0.02146
GO:0003002	regionalization	0.02164
GO:0030243	cellulose metabolic process	0.02194
GO:0010359	regulation of anion channel activity	0.02233
GO:0022898	regulation of trans-membrane trans-porter ...	0.02233
GO:0032412	regulation of ion transmembrane transpor...	0.02233
GO:0032409	regulation of trans-porter activity	0.02233
GO:0006119	oxidative phosphor-ylation	0.02293
GO:0010074	maintenance of me-ristem identity	0.02300
GO:0032880	regulation of pro-tein localization	0.02300
GO:0030029	actin filament-based process	0.02303
GO:0010410	hemicellulose meta-bolic process	0.02360
GO:0051172	negative regulation of nitrogen com-pound...	0.02522
GO:1905393	plant organ for-mation	0.02551

GO:0048449	floral organ formation	0.02583
GO:0051640	organelle localization	0.02583
GO:0006302	double-strand break repair	0.02587
GO:0006007	glucose catabolic process	0.02587
GO:0048509	regulation of meristem development	0.02598
GO:0016246	RNA interference	0.02661
GO:0061458	reproductive system development	0.02683
GO:0048608	reproductive structure development	0.02683
GO:0051240	positive regulation of multicellular org...	0.02794
GO:0048522	positive regulation of cellular process	0.02990
GO:0010540	basipetal auxin transport	0.03177
GO:0045787	positive regulation of cell cycle	0.03177
GO:1902532	negative regulation of intracellular sig...	0.03177
GO:1902275	regulation of chromatin organization	0.03177
GO:0090697	post-embryonic plant organ morphogenesis	0.03186
GO:0048589	developmental growth	0.03326
GO:0010383	cell wall polysaccharide metabolic proce...	0.03374
GO:0019320	hexose catabolic process	0.03417
GO:0046365	monosaccharide catabolic process	0.03417
GO:0009749	response to glucose	0.03417
GO:0090696	post-embryonic plant organ development	0.03424

GO:0010304	PSII associated light-harvesting complex...	0.03490
GO:0048582	positive regulation of post-embryonic de...	0.03495
GO:0009314	response to radiation	0.03541
GO:0052543	callose deposition in cell wall	0.03559
GO:2000243	positive regulation of reproductive proc...	0.03559
GO:0031123	RNA 3'-end processing	0.03603
GO:0009960	endosperm development	0.03603
GO:0051253	negative regulation of RNA metabolic pro...	0.03779
GO:0051049	regulation of transport	0.03794
GO:0048638	regulation of developmental growth	0.03861
GO:0006281	DNA repair	0.03991
GO:0051094	positive regulation of developmental pro...	0.03994
GO:0090693	plant organ senescence	0.04038
GO:0010150	leaf senescence	0.04038
GO:0045934	negative regulation of nucleobase-contai...	0.04114
GO:0015931	nucleobase-containing compound transport	0.04220
GO:0046467	membrane lipid biosynthetic process	0.04220
GO:0006643	membrane lipid metabolic process	0.04233
GO:0006874	cellular calcium ion homeostasis	0.04247
GO:0048574	long-day photoperiodism, flowering	0.04247
GO:0048440	carpel development	0.04253

GO:0050657	nucleic acid transport	0.04370
GO:0050658	RNA transport	0.04370
GO:0006611	protein export from nucleus	0.04370
GO:0006403	RNA localization	0.04370
GO:0006405	RNA export from nucleus	0.04370
GO:0051236	establishment of RNA localization	0.04370
GO:0009416	response to light stimulus	0.04371
GO:0052386	cell wall thickening	0.04384
GO:0009225	nucleotide-sugar metabolic process	0.04384
GO:0009057	macromolecule catabolic process	0.04399
GO:0006338	chromatin remodeling	0.04431
GO:1902531	regulation of intracellular signal trans...	0.04431
GO:0010017	red or far-red light signaling pathway	0.04431
GO:0071489	cellular response to red or far red ligh...	0.04431
GO:0018205	peptidyl-lysine modification	0.04468
GO:0050793	regulation of developmental process	0.04476
GO:0007389	pattern specification process	0.04487
GO:0048467	gynoecium development	0.04507
GO:0009555	pollen development	0.04515
GO:0071554	cell wall organization or biogenesis	0.04529
GO:0010118	stomatal movement	0.04537
GO:1903507	negative regulation of nucleic acid-temp...	0.04665
GO:0045892	negative regulation of transcription, DN...	0.04665

GO:1902679	negative regulation of RNA biosynthetic ...	0.04665
GO:0071555	cell wall organization	0.04707
GO:0006974	cellular response to DNA damage stimulus	0.04809
GO:0051604	protein maturation	0.05065
GO:0043269	regulation of ion transport	0.05111
GO:0034762	regulation of transmembrane transport	0.05145
GO:0034765	regulation of ion transmembrane transpor...	0.05145
GO:0007292	female gamete generation	0.05168
GO:0050794	regulation of cellular process	0.05258
GO:2000241	regulation of reproductive process	0.05267
GO:0006511	ubiquitin-dependent protein catabolic pr...	0.05340
GO:0009150	purine ribonucleotide metabolic process	0.05375
GO:0055074	calcium ion homeostasis	0.05508
GO:0002253	activation of immune response	0.05508
GO:0048571	long-day photoperiodism	0.05508
GO:0002218	activation of innate immune response	0.05508
GO:0048518	positive regulation of biological proces...	0.05845
GO:0042398	cellular modified amino acid biosynthetic...	0.05945
GO:0007051	spindle organization	0.05945
GO:0010608	posttranscriptional regulation of gene e...	0.05993

GO:0021700	developmental maturation	0.06017
GO:0010154	fruit development	0.06018
GO:0044036	cell wall macromolecule metabolic proces...	0.06127
GO:0019375	galactolipid biosynthetic process	0.06203
GO:0043480	pigment accumulation in tissues	0.06203
GO:0043481	anthocyanin accumulation in tissues in r...	0.06203
GO:0043473	pigmentation	0.06203
GO:0043476	pigment accumulation	0.06203
GO:0043478	pigment accumulation in response to UV l...	0.06203
GO:0043479	pigment accumulation in tissues in respo...	0.06203
GO:0009560	embryo sac egg cell differentiation	0.06306
GO:0000271	polysaccharide biosynthetic process	0.06317
GO:0044070	regulation of anion transport	0.06375
GO:1903959	regulation of anion transmembrane transp...	0.06375
GO:0006406	mRNA export from nucleus	0.06376
GO:0051028	mRNA transport	0.06376
GO:0071427	mRNA-containing ribonucleoprotein comple...	0.06376
GO:0043632	modification-dependent macromolecule cat...	0.06417
GO:0019941	modification-dependent protein catabolic...	0.06417
GO:0061647	histone H3-K9 modification	0.06542

GO:0051567	histone H3-K9 methylation	0.06542
GO:0009955	adaxial/abaxial pattern specification	0.06667
GO:2001057	reactive nitrogen species metabolic proc...	0.06810
GO:0031056	regulation of histone modification	0.06810
GO:0010252	auxin homeostasis	0.06810
GO:0016032	viral process	0.06810
GO:0009954	proximal/distal pattern formation	0.06810
GO:0010015	root morphogenesis	0.06830
GO:0006275	regulation of DNA replication	0.06933
GO:0072507	divalent inorganic cation homeostasis	0.06954
GO:0016444	somatic cell DNA recombination	0.06963
GO:0006473	protein acetylation	0.06963
GO:0051645	Golgi localization	0.06963
GO:0051646	mitochondrion localization	0.06963
GO:0060151	peroxisome localization	0.06963
GO:0019374	galactolipid metabolic process	0.06964
GO:0009411	response to UV	0.06982
GO:0043161	proteasome-mediated ubiquitin-dependent ...	0.06990
GO:0051649	establishment of localization in cell	0.07344
GO:0010646	regulation of cell communication	0.07411

Appendix Table 6: Gene ontology enrichment analysis for the genes with a polymorphic *cis* effect for SP. The GOs ID, the term and the p value of the significant GOs ($p < 0.05$) are given.

GO.ID	Term	resultKS
GO:0009718	anthocyanin-containing compound biosynth...	0.00057
GO:0019748	secondary metabolic process	0.00064
GO:0043161	proteasome-mediated ubiquitin-dependent ...	0.00165
GO:0051302	regulation of cell division	0.00179
GO:0017144	drug metabolic process	0.00228
GO:0046686	response to cadmium ion	0.00350
GO:1901617	organic hydroxy compound biosynthetic pr...	0.00355
GO:0022603	regulation of anatomical structure morph...	0.00531
GO:0009962	regulation of flavonoid biosynthetic pro...	0.00730
GO:0009651	response to salt stress	0.00757
GO:0008284	positive regulation of cell proliferatio...	0.00795
GO:0018958	phenol-containing compound metabolic pro...	0.00876
GO:0043085	positive regulation of catalytic activit...	0.00923
GO:0005982	starch metabolic process	0.01024
GO:0042023	DNA endoreduplication	0.01031
GO:0044786	cell cycle DNA replication	0.01088
GO:0043933	protein-containing complex subunit organ...	0.01127

GO:0044281	small molecule metabolic process	0.01146
GO:1901135	carbohydrate derivative metabolic proces...	0.01268
GO:0042274	ribosomal small subunit biogenesis	0.01271
GO:0031540	regulation of anthocyanin biosynthetic p...	0.01271
GO:0006595	polyamine metabolic process	0.01281
GO:0044085	cellular component biogenesis	0.01299
GO:0043248	proteasome assembly	0.01364
GO:0051788	response to misfolded protein	0.01364
GO:0006796	phosphate-containing compound metabolic ...	0.01387
GO:0006766	vitamin metabolic process	0.01494
GO:0019684	photosynthesis, light reaction	0.01522
GO:0019252	starch biosynthetic process	0.01539
GO:0016999	antibiotic metabolic process	0.01594
GO:0006793	phosphorus metabolic process	0.01665
GO:0044042	glucan metabolic process	0.01674
GO:0006073	cellular glucan metabolic process	0.01674
GO:0034622	cellular protein-containing complex asse...	0.01681
GO:0090407	organophosphate biosynthetic process	0.01716
GO:0009636	response to toxic substance	0.01764
GO:0046189	phenol-containing compound biosynthetic ...	0.01866

GO:0009404	toxin metabolic process	0.01936
GO:0017000	antibiotic biosynthetic process	0.01981
GO:0065003	protein-containing complex assembly	0.02076
GO:0051186	cofactor metabolic process	0.02129
GO:0009411	response to UV	0.02236
GO:0019637	organophosphate metabolic process	0.02365
GO:0009642	response to light intensity	0.02382
GO:0055086	nucleobase-containing small molecule met...	0.02386
GO:0009611	response to wounding	0.02394
GO:0000023	maltose metabolic process	0.02517
GO:0006364	rRNA processing	0.02679
GO:0019932	second-messenger-mediated signaling	0.02709
GO:0016072	rRNA metabolic process	0.02836
GO:0006260	DNA replication	0.03170
GO:0022607	cellular component assembly	0.03251
GO:0048528	post-embryonic root development	0.03338
GO:0051604	protein maturation	0.03756
GO:0006598	polyamine catabolic process	0.03942
GO:0006807	nitrogen compound metabolic process	0.03970
GO:0006091	generation of precursor metabolites and ...	0.04025
GO:0046677	response to antibiotic	0.04096
GO:0006261	DNA-dependent DNA replication	0.04326
GO:0034470	ncRNA processing	0.04326

GO:1901566	organonitrogen compound biosynthetic pro...	0.04340
GO:0006139	nucleobase-containing compound metabolic...	0.04362
GO:0009735	response to cytokinin	0.04448
GO:0009409	response to cold	0.04468
GO:0001666	response to hypoxia	0.04470
GO:0007020	microtubule nucleation	0.04470
GO:0009407	toxin catabolic process	0.04499
GO:0046785	microtubule polymerization	0.04722
GO:0009117	nucleotide metabolic process	0.04738
GO:1902531	regulation of intracellular signal trans...	0.04784
GO:0006301	postreplication repair	0.04784
GO:0005984	disaccharide metabolic process	0.04798
GO:0006753	nucleoside phosphate metabolic process	0.04806
GO:0080129	proteasome core complex assembly	0.04886
GO:0009697	salicylic acid biosynthetic process	0.04951
GO:0006979	response to oxidative stress	0.04960
GO:0051510	regulation of unidimensional cell growth	0.04981
GO:0036293	response to decreased oxygen levels	0.04981
GO:0031109	microtubule polymerization or depolymeri...	0.04981
GO:0070482	response to oxygen levels	0.04981

Appendix Table 7: Gene ontology enrichment analysis for the genes with a polymorphic *cis* effect for PL. The GOs ID, the term and the p value of the significant GOs ($p < 0.05$) are given.

GO.ID	Term	resultKS
GO:0006412	translation	8.8e-23
GO:0006098	pentose-phosphate shunt	6.8e-22
GO:0006364	rRNA processing	1.3e-18
GO:0046686	response to cadmium ion	2.5e-17
GO:0006096	glycolytic process	5.0e-16
GO:0010207	photosystem II assembly	4.2e-15
GO:0019344	cysteine biosynthetic process	7.2e-12
GO:0009651	response to salt stress	1.2e-11
GO:0006833	water transport	1.6e-11
GO:0007030	Golgi organization	2.1e-11
GO:0019288	isopentenyl diphosphate biosynthetic pro...	3.3e-11
GO:0010114	response to red light	3.5e-11
GO:0042744	hydrogen peroxide catabolic process	5.3e-11
GO:0009902	chloroplast relocation	9.9e-11
GO:0010027	thylakoid membrane organization	1.2e-10
GO:0010155	regulation of proton transport	1.5e-10
GO:0010218	response to far red light	4.6e-10
GO:0015995	chlorophyll biosynthetic process	9.7e-10
GO:0009637	response to blue light	5.3e-09
GO:0019252	starch biosynthetic process	9.3e-09
GO:0000023	maltose metabolic process	1.6e-08
GO:0043085	positive regulation of catalytic activit...	1.9e-08
GO:0009735	response to cytokinin	2.3e-08

GO:0009657	plastid organization	5.4e-08
GO:0001510	RNA methylation	1.6e-07
GO:0019761	glucosinolate biosynthetic process	1.3e-06
GO:0009773	photosynthetic electron transport in pho...	3.0e-06
GO:0009965	leaf morphogenesis	4.4e-06
GO:0006972	hyperosmotic response	4.6e-06
GO:0009409	response to cold	8.8e-06
GO:0009744	response to sucrose	1.2e-05
GO:0019684	photosynthesis, light reaction	1.3e-05
GO:0009266	response to temperature stimulus	1.3e-05
GO:0006094	gluconeogenesis	2.3e-05
GO:0042793	plastid transcription	3.7e-05
GO:0035304	regulation of protein dephosphorylation	6.1e-05
GO:0009644	response to high light intensity	8.6e-05
GO:0009269	response to desiccation	9.7e-05
GO:0019760	glucosinolate metabolic process	0.00012
GO:0006816	calcium ion transport	0.00013
GO:0045893	positive regulation of transcription, DN...	0.00014
GO:0042254	ribosome biogenesis	0.00015
GO:0009658	chloroplast organization	0.00017
GO:0016117	carotenoid biosynthetic process	0.00019
GO:0010043	response to zinc ion	0.00020
GO:0010304	PSII associated light-harvesting complex...	0.00036
GO:0010264	myo-inositol hexakisphosphate biosynthet...	0.00051

GO:0030003	cellular cation homeostasis	0.00051
GO:0070838	divalent metal ion transport	0.00058
GO:0009853	photorespiration	0.00067
GO:0042364	water-soluble vitamin biosynthetic proce...	0.00115
GO:0009073	aromatic amino acid family biosynthetic ...	0.00120
GO:0009750	response to fructose	0.00152
GO:0045036	protein targeting to chloroplast	0.00229
GO:0006007	glucose catabolic process	0.00235
GO:0042742	defense response to bacterium	0.00241
GO:0009805	coumarin biosynthetic process	0.00327
GO:0030154	cell differentiation	0.00440
GO:1901566	organonitrogen compound biosynthetic pro...	0.00499
GO:0043255	regulation of carbohydrate biosynthetic ...	0.00541
GO:0045037	protein import into chloroplast stroma	0.00541
GO:0051186	cofactor metabolic process	0.00600
GO:0015979	photosynthesis	0.00676
GO:0017004	cytochrome complex assembly	0.00704
GO:0090056	regulation of chlorophyll metabolic proc...	0.00704
GO:0006790	sulfur compound metabolic process	0.00995
GO:1901565	organonitrogen compound catabolic proces...	0.01011
GO:0009825	multidimensional cell growth	0.01016
GO:0009746	response to hexose	0.01038

GO:0009765	photosynthesis, light harvesting	0.01106
GO:0009749	response to glucose	0.01108
GO:0000272	polysaccharide catabolic process	0.01128
GO:0006979	response to oxidative stress	0.01177
GO:0016052	carbohydrate catabolic process	0.01191
GO:0009108	coenzyme biosynthetic process	0.01198
GO:0009664	plant-type cell wall organization	0.01202
GO:0048767	root hair elongation	0.01203
GO:0033559	unsaturated fatty acid metabolic process	0.01265
GO:0006636	unsaturated fatty acid biosynthetic proc...	0.01265
GO:0034284	response to monosaccharide	0.01266
GO:0051188	cofactor biosynthetic process	0.01369
GO:0009310	amine catabolic process	0.01421
GO:0042402	cellular biogenic amine catabolic proces...	0.01421
GO:0042168	heme metabolic process	0.01618
GO:0032880	regulation of protein localization	0.01654
GO:0031163	metallo-sulfur cluster assembly	0.01661
GO:0016226	iron-sulfur cluster assembly	0.01661
GO:0080129	proteasome core complex assembly	0.01745
GO:1901658	glycosyl compound catabolic process	0.01916
GO:0006518	peptide metabolic process	0.02165
GO:0071555	cell wall organization	0.02167

GO:0010118	stomatal movement	0.02221
GO:0006576	cellular biogenic amine metabolic proces...	0.02238
GO:0006598	polyamine catabolic process	0.02241
GO:0080147	root hair cell development	0.02275
GO:0009411	response to UV	0.02692
GO:0006301	postreplication repair	0.02977
GO:0010015	root morphogenesis	0.03114
GO:0019725	cellular homeostasis	0.03140
GO:0045229	external encapsulating structure organiz...	0.03206
GO:0044106	cellular amine metabolic process	0.03319
GO:0005982	starch metabolic process	0.03344
GO:1901136	carbohydrate derivative catabolic proces...	0.03388
GO:0043480	pigment accumulation in tissues	0.03480
GO:0043481	anthocyanin accumulation in tissues in r...	0.03480
GO:0043473	pigmentation	0.03480
GO:0043476	pigment accumulation	0.03480
GO:0043478	pigment accumulation in response to UV l...	0.03480
GO:0043479	pigment accumulation in tissues in respo...	0.03480
GO:0048481	plant ovule development	0.03498
GO:0035266	meristem growth	0.03498
GO:0035670	plant-type ovary development	0.03498

GO:0006109	regulation of carbo-hydrate metabolic pro...	0.03559
GO:0008361	regulation of cell size	0.03717
GO:0010119	regulation of sto-matal movement	0.03717
GO:0072525	pyridine-containing compound biosyn-theti...	0.03767
GO:0009809	lignin biosynthetic process	0.04291
GO:0010054	trichoblast differen-tiation	0.04419
GO:0051181	cofactor transport	0.04422
GO:0048364	root development	0.04667
GO:0048232	male gamete gener-ation	0.04783
GO:0009625	response to insect	0.04783
GO:0022622	root system devel-opment	0.04787
GO:0010053	root epidermal cell differentiation	0.04915

6.4 Review: Linking genes with ecological strategies in *Arabidopsis thaliana*

Authors: Margarita Takou[1,2], Benedict Wieters[1,2], Stanislav Kopriva[2], George Coupland[3], Anja Linstädter[2,4,] , Juliette de Meaux[2]

[1] Contributed equally

[2] Institute of Botany, University of Cologne, Germany

[3] Max Planck Institute of Plant Breeding Research, Cologne, Germany

[4] Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Germany

Abstract

Arabidopsis thaliana is the most prominent model system in plant molecular biology and genetics. Although its ecology was initially neglected, collections of various genotypes revealed a complex population structure, with high levels of genetic diversity and substantial levels of phenotypic variation. This helped identify the genes and gene pathways mediating phenotypic change. Population genetics studies further demonstrated that this variation generally contributes to local adaptation. Here, we review evidence showing that traits affecting plant life history, growth rate and stress reactions are not only locally adapted, they also often co-vary. Co-variation between these traits indicate that they evolve as trait syndromes, and reveals the ecological diversification that took place within *A. thaliana*. We argue that examining traits and the gene that control them within the context of global summary schemes that describe major ecological strategies will contribute to resolve important questions both in molecular biology and ecology.

Keywords: *Arabidopsis thaliana*, natural variation, local adaptation, trait syndrome, CSR strategy

Local adaptation suggests a diversity of ecological specializations within *A. thaliana*

Arabidopsis thaliana (L.) Heyhn is exceptional within its genus. It is the only annual species, it has adapted to open, dry habitats prone to seasonal drought (Ruppert *et al.*, 2015; Kiefer *et al.*, 2017), and its reproductive success is directly dependent on interannual variation in environmental

conditions (Segrestin *et al.*, 2018). It also has the widest geographic range in the *Arabidopsis* genus (Clauss and Koch, 2006; Novikova *et al.*, 2016). Natural populations have been found throughout Europe, from the North of Scandinavia to the South of Spain, in the Balkans, in Central Asia, China and parts of Africa (Hoffmann, 2005; He *et al.*, 2007; 1001 Genomes Consortium. 2016; Durvasula *et al.*, 2017). It is also naturalized in North America and Argentina (Stock *et al.*, 2015; Kasulin *et al.*, 2017; Exposito-Alonso *et al.*, 2018*a*). This exceptionally wide distribution range is only limited by very low spring or autumn temperatures or by high temperature in regions of low precipitation (Hoffmann, 2002).

The unrivaled genomic resources available for these populations have helped demonstrate that the last glacial period determined the current distribution of genetic variation (reviewed in this issue, Koch 2018). After the last glacial maximum, populations have spread towards Northern latitudes, experiencing successive bottlenecks (Svardal *et al.*, 2017; Durvasula *et al.*, 2017). As a result, regional diversity is highest in Africa and lowest in Scandinavia. Genetic variation in Eurasia also follows a longitudinal gradient (1001 Genomes Consortium, 2016; Zou *et al.*, 2017).

The local adaptation of *A. thaliana* populations has been documented throughout its range, despite a history of pervasive gene flow (Fournier-Level *et al.*, 2011; Hancock *et al.*, 2011; Ågren and Schemske, 2012; Savolainen *et al.*, 2013; Weigel and Nordborg, 2015; Svardal *et al.*, 2017). Field experiments and correlation analyses with climate parameters identified numerous genomic regions associated with local climatic conditions. Association studies correlating SNP variants with climatic variation showed that non-synonymous SNPs were enriched among SNPs associating with environmental variance (Hancock *et al.*, 2011; Lasky *et al.*, 2014). Among them, SNPs associating with fitness differences in the field were also over-represented (Hancock *et al.*, 2011). Furthermore, alleles associating with fruit production are more frequent in populations closer to field sites where the selective advantage was expressed (Fournier-Level *et al.*, 2011). Therefore, it is now clear that much of the variation found in this species has played a role in optimizing plant performance to local environmental conditions.

Combination of development traits underpin local adaptation in *A. thaliana*

Flowering time is one of the development traits underpinning adaptation in *A. thaliana*. It has been extensively studied and elevated levels of variation have been observed in the lab (Koornneef *et al.*, 2004). The adaptive relevance of its genetic variation is supported by multiple independent

studies. Variation in flowering time follows climatic clines, both at the regional and species levels (Mendez-Vigo *et al.*, 2011; Montesinos-Navarro *et al.*, 2011; Debieu *et al.*, 2013; Li *et al.*, 2014; Sasaki *et al.*, 2015). Warmer climates appear to favor earlier flowering time, a pattern that has been documented for a great number of species (Austen *et al.*, 2017; Whittaker and Dean, 2017). Strong selection for early flowering was detected in Italy but was weaker in Sweden (Ågren *et al.*, 2017). Population genetics studies also uncovered signatures of natural selection on genes controlling flowering time (Le Corre, 2005; Toomajian *et al.*, 2006). Finally, the analysis of co-variation between environmental and phenotypic variance consolidated evidence for the adaptive distribution of this trait (van Heerwaarden *et al.*, 2015).

Much of the flowering time variation measured in the lab, however, does not manifest as variation in flowering phenology in the field (Wilczek *et al.*, 2009; Brachi *et al.*, 2010; Hu *et al.*, 2017). It is indeed tightly dependent on the environmental conditions prevailing during seedling establishment, and hence on another developmental trait: the timing of germination (Donohue, 2002; Wilczek *et al.*, 2009). Both field experiments and theoretical models integrating seed and flowering phenology have shown that seed dormancy is often decisive for controlling the life cycle across environments (Chiang *et al.*, 2013; Burghardt *et al.*, 2015). Therefore, the adaptive relevance of variants modulating flowering time control must be examined in the context of variation for the timing of germination.

There is indeed consistent support for the adaptive relevance of traits determining the timing of germination. Seed dormancy has a strong fitness advantage before the hot season, but can impair fitness if it delays plant growth before winter (Donohue, 2002; Donohue *et al.*, 2005; Chiang *et al.*, 2013). Population genetics analysis of seed dormancy and its major QTL *DOG1* supported the adaptive importance of strong dormancy in Southern regions, to escape dry and hot summers, whereas weaker dormancy was reported in Norway, where the season is shorter (Kronholm *et al.*, 2012; Postma and Ågren, 2016; Kerdaffrec *et al.*, 2016).

Since flowering time determines the maternal environment the seeds experience during their maturation, it also impacts life history traits expressed by the next generation (Chiang *et al.*, 2013; He *et al.*, 2014; Postma and Ågren, 2015). Light, temperature, nutrient availability and water status have all been identified as significant environmental factors influencing the maternal inheritance of seed dormancy (Footitt *et al.*, 2013; Morrison and Linder, 2014; He *et al.*, 2014; Kerdaffrec *et*

al., 2016). Germination can also be distributed over more than one seasonal window. For example, maintaining a spring germinating cohort is important for the maintenance of populations exposed to low winter temperature (Akiyama and Ågren 2014; Picó, 2012). Furthermore, later flowering can lead to late seed dispersal, which can result in overwintering at the seed stage (Hu *et al.* 2017).

Flowering time and seed dormancy are therefore jointly subject to fluctuating seasonal selective forces. They can evolve as a syndrome, defining distinct life history strategies that have diversified across environments (Marcer *et al.*; Chiang *et al.*, 2013; Vidigal *et al.*, 2016). An analysis of seed dormancy and flowering time co-variation revealed that the optimization of the two traits probably depends on latitudinal differences in climate. Late flowering (i.e. vernalization requirement) and strong dormancy are more frequent in regions where summer drought is typically more severe, whereas late flowering in Northern latitudes co-varies negatively with dormancy (Debieu *et al.* 2013). Co-variance between flowering time and dormancy is also detected at much smaller scale, along steep altitudinal gradients (Vidigal *et al.* 2016). Normally, diverse life-history trait combinations can allow comparable population growth rates in field conditions (Taylor *et al.*, 2017). In some years, however, early winter frost can wipe out genotypes expressing inadequate life histories (Hu *et al.* 2017). Minimum winter temperature and precipitation, in fact, were also the main climatic factors that acted as selective pressures on flowering time and their underpinning genes in a set of Iberian *A.thaliana* genotypes (Méndez-Vigo *et al.*, 2011). This suggests that extreme deviation from seasonal averages may be important drivers of the allelic combination of life history variants adjusting development to the optimal growth season throughout the species range.

Patterns of co-variation between growth rate and developmental traits suggest the existence of trait syndromes

Beyond the combination of life history traits to target the best season for growth, *A. thaliana* also displays considerable genetic variation in its growth rate (Debieu *et al.*, 2013; Marchadier *et al.*, 2018). The pattern of co-variation linking growth rate with flowering time and seed dormancy is independent of population structure and changes from Southern to Northern latitude (Debieu *et al.*, 2013). This suggests that trade-offs between growth rate and life history change across the distribution range of the species (Debieu *et al.*, 2013). It further implies that complex multi-trait combinations, i.e. trait syndromes, are necessary to adjust to the changing trade-offs imposed by regional differences in climatic conditions. Co-variation between flowering time, final biomass

and average rate of biomass accumulation before flowering also suggests that genetic adaptation to local climate conditions is mediated by a trait syndrome (Vasseur *et al.*, 2018a).

Genetic variation for tolerance to drought stress, just like that for life history, displays signatures of local adaptation. Many genetic variants have been identified that also affect either root or rosette growth in the face of severe drought stress (El Soda *et al.*; Clauw *et al.*, 2016; Davila Olivas *et al.*, 2017). Several studies highlighted the adaptive relevance of variation in the ability to maintain growth and photosynthesis when water is limited. After accounting for the demographic history of the species, stomata size variation was found to correlate with water-use efficiency (i.e. rate of carbon fixation to water loss) and both air humidity and the local probability of spring drought severity (Dittberner *et al.* 2018), in agreement with field measurements (Mojica *et al.*, 2016). The molecular evolution of the gene P5CS, which contributes to the synthesis of proline, a potent osmoprotectant in *A. thaliana,* suggests it contributed to local adaptation (Kesari *et al.*, 2012). Nucleotide variants within genes displaying stress-dependent expression was also shown to be over-represented among variants correlating with climatic parameters (Lasky *et al.*, 2014; Exposito-Alonso *et al.*, 2018*b*).

In fact, genetic variation for stress tolerance is not only involved in local adaptation, it also appears to be part of a trait syndrome, because it is often reported to coincide with variation in life history. Early flowering individuals, which sometimes complete their life cycle within a few weeks, can escape conditions causing high pre-reproductive mortality (Franks et al., 2011; Fournier-Level *et al.*, 2013; Riboni *et al.*, 2013). In addition, the major flowering time QTL *FRIGIDA* controls not only the timing of flowering but also water-use efficiency (Johanson *et al.*, 2000; Lovell *et al.*, 2013). Improved performance in plants exposed to moderate drought stress is correlated with the ability to flower early (Bac Molenaar *et al.*, 2016), but genotypes with a strong vernalization requirement rather tend to avoid the effect of drought by maintaining their internal water level (McKay *et al.*, 2003; Des Marais *et al.*, 2012; Lovell *et al.*, 2013; Easlon *et al.*, 2014; Davila Olivas *et al.*, 2017). The most stress tolerant individuals actually appear to be either early flowering or slow growing (Davila Olivas *et al.*, 2017; Vasseur *et al.*, 2018a).

Because of its co-variation with life-history, adaptation to drought stress can show counter-intuitive patterns. *In A. thaliana,* local adaptation for increased tolerance to drought stress is not found in the driest regions, because, in these areas, natural selection rather promoted genotypes with the

ability to escape the stress (Tabas-Madrid *et al.*; Kronholm *et al.*, 2012; Vidigal *et al.*, 2016, Mojica *et al.* 2016). Genotypes showing smaller stomata, higher water-use efficiency and longer photosynthetic activity in the face of terminal drought have in fact evolved in Southern Scandinavia, where the growth season is too short to allow escaping the drier season but dry enough to require improved drought tolerance (Dittberner *et al.*, 2018; Exposito-Alonso *et al.*, 2018*b*, Mojica *et al.* 2016). Genotypes with a strong vernalization requirement are in fact more frequent in this region, limiting the possibility to escape drought during the growing season (Li *et al.*, 2014). Non-monotonic patterns of co-variation between flowering time and temperature have also been reported in Spain (Tabas-Madrid *et al.*, 2018), suggesting that the selective advantage of early flowering for persisting in dry regions depends on a broader ecological context, and thus presumably on the possibility to rely on earlier flowering to escape stressful conditions. The evolution of the response to abiotic stress in *A. thaliana* is therefore not independent of the evolution of the timing of life history transitions.

The ability of plants to face biotic stresses is also dependent on life history variation. Alleles accelerating flowering were shown to be often combined with alleles decreasing the expression of plant defense genes throughout natural *A. thaliana* populations (Glander *et al.*, 2018). Indication that this assortment coincides with differential fitness suggests that it has been driven by natural selection. In addition, plant growth in response to the specialist herbivore *Pieris rapae* was enhanced in fast flowering individuals but showed a trade-off with the drought response (Davila Olivas *et al.*, 2017). Here again, this hints to the evolution of a trait syndrome, where early flowering genotypes may have been selected for their ability to allocate fewer resources into defense, in order to maximize their growth rate or to reshuffle energetic priorities and ensure survival.

A. thaliana thus displays significant levels of genetic variation in traits controlling life-history, growth rate or tolerance to diverse stresses, all of which co-vary with each other and with climatic conditions at the location of origin. This suggests that adaptation to novel environments after the last glaciation has allowed the evolution of trait syndromes, i.e. combination of multiple traits. These combinations probably reflect both adaptive synergies and global trade-offs imposed by resource limitations. Identifying the exact composition of trait syndromes and their variation requires a careful monitoring of life-history transitions, growth rates, stress tolerance and plant fitness in natural conditions. We argue below that interpreting trait variation and co-variation in the

global context of plant ecological strategies, i.e. within summary schemes developed by ecologists to describe the major dimensions determining variation in form and function within and across habitats, may help resolve the ecological significance of traits and their combinations.

Interpreting *A. thaliana* trait syndromes in the context of major ecological strategies

Not all possible trait combinations are viable in natural environments: Natural selection indeed limits the diversity of forms and functions in plants as a result of trade-offs among the diverse options for resource allocation (Reich, 2014; Díaz *et al.*, 2016). This major constraint has motivated several attempts to classify plants with respect to their ecological strategies (reviewed in Westoby *et al.*, 2002; Díaz *et al.*, 2016). Among them, Grime's CSR theory (Grime, 1974; Grime, 1977) is a prominent strategy scheme (Pierce *et al.*, 2017). It distinguishes three primary strategies, i.e. competitive (C), stress-tolerant (S) and ruderal (R). The two first strategies evolve in rather constant environments, which differ in the severity of resource shortage (light, water, and/or nutrients). The third one prevails in disturbed environments, and involves investment of a large proportion of resources in propagules from which the population can regenerate in the face of repeated biomass destruction events. Distinct strategies may also co-occur within a given environment enhancing local niche separation between species (Kraft *et al.* 2008). The multivariate and complex traits that form the basis of ecological strategies are often difficult to measure. Yet, a small number of plant functional traits related to growth, survival and reproduction has been shown to efficiently summarize the overall diversity of plant life form and functions (Díaz *et al.*, 2016). Among them, as many as three leaf traits – leaf area, leaf dry matter content, and specific leaf area – can be used as surrogate to describe a species' CSR strategy (Pierce *et al.*, 2017).

Like many other annuals in the Brassicaceae family, these leaf traits position *A. thaliana* as a typical R-strategist (Pierce *et al.*, 2017). It is typically encountered in regularly disturbed habitat patches, such as urban, ruderal or mountainous habitats, and its seedlings are directly exposed to seasonal climatic fluctuations (Pico *et al.*, 2008; Bomblies *et al.*, 2010; Svardal *et al.*, 2017). However, the past years have shown that plant species are far from having a fixed CSR strategy. On the contrary, strategy classifications at the species level have been more and more challenged by the large spectrum of intraspecific variation (Des Roches *et al.*, 2017; Volaire, 2018). For this reason, it is now increasingly acknowledged that trade-offs in life-history and growth strategies can also occur at the level of genotypes and populations, and that more attention should be devoted

to inter-individual and inter-population variation of the CSR strategy (Astuti *et al.*, 2018). Intraspecific trade-offs have been found along the S-R axis of Grime's CSR strategy scheme (Bilton *et al.*, 2010; May *et al.*, 2017), but also along the C-S axis (Ravenscroft *et al.*, 2014; Astuti *et al.*, 2018). These trade-offs are commonly explained as a mechanism of local adaptation, for example in response to different levels of resource stress. It is thus not surprising that considerable intraspecific variation has also been found for *A. thaliana.* A study with 16 individual accessions sampled along a steep altitudinal gradient revealed that populations from hotter climates clustered towards the stress-tolerant end of the observed strategy spectrum, implying pronounced intraspecific variation along the S-R axis (May *et al.*, 2017). The extent of variation along the S-R axis was recently confirmed by a comprehensive analysis of variation in CSR positioning in some 300 genotypes in *A. thaliana* (Vasseur *et al.* 2018b). As for other annual plants, we could thus assume that *A. thaliana* populations growing in water- or temperature-limited habitats may be well-adapted to high levels of stress and thus characterized as stress tolerant (Volaire, 2018). By contrast, genotypes that grow fast and complete their life cycle within a few weeks may be described as extreme ruderals (Fig. 1).In *A. thaliana,* genotypic variation covers the whole S-R strategy spectrum. The geographical distribution of this variation contrasts with that of genome-wide patterns of variation, suggesting a role in local adaptation (Vasseur *et al.* 2018b).

Intraspecific variation along the S-C or R-C axes involves traits that have not been intensively investigated in *A. thaliana* (Fig. 1). The ability to compete with other species plays a presumably minor role for a pioneer species that only subsists in disturbed environments. Yet, the few studies that investigated this aspect (e.g. Masclaux *et al.* 2010; Baron *et al.*, 2015; Frachon *et al.*, 2017) suggest that intraspecific variation in C-strategic features will be significant as well (Fig.1). The disturbed environments in which *A. thaliana* can be found are sometimes densely populated (G. Schmitz, pers. Com, see also the population studied in Frachon *et al.* 2018). Interspecific competition has been shown to modify the pattern of natural selection operating on flowering traits in a collection of recombinant inbred lines (Brachi *et al.*, 2012). Some *A. thaliana* genotypes, initially collected in a densely populated habitat patch, displayed the ability to decrease the biomass of some of their interspecific competitors (Baron *et al.*, 2015; Frachon *et al.*, 2017).

Intraspecific competition is also expected to stand under strong selection in the species. The population census size is small in a newly colonized habitat patch but will increase with the age of the

habitat patch. Under increasing density of intraspecific competitors, plants differ in their ability to reach the fruiting phase and produce seeds (Masclaux *et al.* 2010; Muñoz-Parra *et al.* 2017). Root growth is also negatively impacted by intra-specific competition (Muñoz-Parra *et al.* 2017). Competitive ability may also modulate the intensity of selection on water-use efficiency variants (Campitelli *et al.*, 2016). Intraspecific variation along the S-C or

R-C axes might be less pronounced than along the S-R axis, but is probably not negligible, as indicated by the recent discovery of a gene locus promoting positive interactions between genotypes (Wuest and Niklaus, 2018; see also Vasseur *et al.* 2018b).

The CSR-strategy scheme is one of the conceptual frameworks that can help understand the role specific trait- (or gene-) combinations have played in the ecological diversification observed within *A. thaliana*. To date, their contribution remains mostly hypothetical (Fig. 1). Late flowering in controlled conditions was reported to associate with increased stress-tolerance in the CSR scheme, yet whether this trait mediates an increase in stress tolerance or associates with traits which control stress tolerance has not been elucidated (Vasseur *et al.* 2018b). Identifying causal links between traits, their underpinning genes and shifts in CSR strategy could considerably ameliorate knowledge transfer between model and non-model species, because this scheme was designed to facilitate interspecific comparisons (Pierce *et al.* 2018).

Towards linking molecular biological functions with their role in ecological strategies

Exploring how traits are combined in natural populations to tune the ecological strategy of local genotypes to their local environmental conditions can indeed cast new light on gene function at the molecular level. We illustrate this idea with two points: first we describe how natural variation can help identify genes controlling ecologically important traits and focus on plant nutrition as an exemplary trait. Second, we show that the function of the well-known flowering time regulator FLC can be revised in the perspective of ecological strategies.

Studies of natural variation have greatly assisted the discovery of the genes controlling functions that cannot be easily dissected in mutant screen approaches (reviewed in Alonso-Blanco *et al.*, 2009). For example, QTL analyses of nutrient unraveled the function of the anion channel AtCLC-c in nitrate transport to vacuoles (Loudet *et al.* 2003; Harada *et al.* 2004) or showed that the ATPase

subunit G and the multicopper oxidase LPR1 have a major impact on the accumulation of phosphate and phytate (Bentsink *et al.* 2003; Reymond *et al.* 2006). They further showed that foliar sulfate accumulation is dependent on sulfate reduction rates (Loudet *et al.* 2007; Koprivova *et al.* 2013). The fact that one of the two major QTLs controlling sulfate reduction, the gene *APR2*, has evolved loss of function alleles three times independently, in Central Asia, Czech Republic and Sweden, also came out as a striking result (Loudet *et al.* 2007, Chao *et al.* 2014).

Studies of natural variation can also inform about the genetic architecture and evolutionary potential of specific traits. For example, the major discoveries driven by differences in ionomes between *A. thaliana* populations were not achieved through genome-wide association mapping (Atwell *et al.* 2010), but through the use of ionomics to screen for genotypes with contrasted nutrient content for analysis (Lahner *et al.* 2003; Salt *et al.* 2008). Approximately 20-fold differences in molybdenum concentration was measured in leaves of 98 *A. thaliana* genotypes and the genetic analysis of the progeny of two of the most contrasted accessions Col-0 and Ler revealed the role of *Molybdenum Transporter-1 MOT1* (Tomatsu *et al.* 2007; Baxter *et al.* 2008). Similarly, tetraploidy was shown to increase potassium content in the progeny of Col-0 and the autotetraploid line Wa-1 (Chao *et al.* 2013). Such studies demonstrate that heritable variation in nutrient content often results from variants that are i) large effect mutations since they can be easily dissected in bi-parental progenies, and ii) rare because they do not give a detectable signal in GWAS. This indicates that plant ability to preempt resources for improved nutrition can be easily manipulated at the genetic level. Such genetic variants in nutrient uptake can be used to understand population maintenance and plant community formation in a context of nutrient depletion or plant-plant competition. In other words, they provide a valuable resource to understand how molecular mechanisms can contribute to ecological diversification.

For most ionomic traits, however, a contribution to plant growth rate, competitive ability or stress tolerance has not been established. Accumulation of sodium is one of the rare examples where the ecological relevance of genetic variation for mineral uptake could be documented. An allele of the sodium transporter *AtHKT1* was shown to mediate increased Na^+ concentration in *A. thaliana* genotypes originating from two coastal habitats in Spain and Japan and was found to co-segregate with salt tolerance (Rus *et al.* 2006, Baxter *et al.* 2010). Using distance to sea or to a known salinity soil as a quantitative measure, a strong relationship between high leaf Na^+ and origin in potentially

saline impacted soils was confirmed (Baxter *et al.* 2010). Recently, the mechanism by which the weak allele of *AtHKT1* confers Na^+ tolerance has been elucidated (An *et al.* 2017). High expression of *AtHKT1* in stems strongly limits the allocation of Na^+ to reproductive tissues and confers thus higher fertility specifically under salt stress (An *et al.* 2017).

An ecological perspective on functional variation can also allow a more comprehensive description of gene function. For example, the gene *FLOWERING LOCUS C* was named after its typical effect on flowering time: *flc* mutants are considerably earlier flowering in long-day controlled conditions (Michaels & Amasino, 1999). The dissection of natural variation present at this locus in *A. thaliana* uncovered an allelic series conferring a wide range of flowering times and responses to vernalization (Lempe *et al.*, 2005; Coustham *et al.*, 2012; Shindo *et al.*, 2005; Li *et al.*, 2014). Allelic variation at *FLC* orthologues is also responsible for variation in flowering time or duration of flowering in other crucifer species (Albani *et al.*, 2012; Kemi *et al.*, 2013; Baduel *et al.*, 2016). Progressively, however, the specificity of FLC action on flowering time has been questioned. *FLC* variation was associated with the timing of germination (Chiang *et al.* 2009), raising the possibility that *FLC* acts pleiotropically on multiple phenotypes. Indeed, the genome-wide analysis of FLC binding sites uncovered several hundred genes, a large proportion of which were involved in response to cold stress (Deng *et al.*, 2011; Mateos *et al.*, 2015; Mateos *et al.*, 2017). Several genes involved in cold stress were strongly deregulated in *flc* mutants compared to *FLC* wild-type when plants were exposed to cold, but not at normal growth temperatures, suggesting that FLC has a role in modulating expression of genes conferring tolerance to cold (Mateos *et al.*, 2017). Pleiotropic genes such as *FLC* may respond to the fundamental requirement for ecological pleiotropy in natural environments that are marked by inevitable fluctuations. Indeed, the monitoring of frequency changes in alleles associated with various reproductive and phenological traits in the field within a single natural *A. thaliana* population showed that variants with intermediate levels of pleiotropy contributed the largest adaptive steps (Frachon *et al.* 2018). This is because selective forces fluctuate across years and seasons and they are more likely to act consistently on variants controlling more than one phenotype. Natural selection at this particular site thus appeared to have favored variants contributing to both increased tolerance to local warming and increased competivive ability (Frachon *et al.* 2018). Although it stands beyond the scope of this review to enumerate all molecular functions whose ecological role remains to be fully determined, the examples given by plant nutrition or the pleiotropic effects of *FLC* illustrate how placing molecular functions

within the context of ecological strategies helps identify genes and their ultimate role in natural conditions.

Towards resolving ecological questions with a genetically tractable plant system

Conversely, the diverse ecological strategies co-segregating in *A. thaliana* provides a unique system to address at the genetic and molecular level those questions that are key to ecologists. We illustrate this idea below with a pressing question in current ecological research: the impact of climate change.

In today's ecological research, discerning the mechanisms behind ecosystem responses to climate change is a central theme (Reed *et al.*, 2012; Ruppert *et al.*, 2015). Extended periods of high temperature and net declines in soil moisture are expected in many regions (IPCC, 2013). Intraspecific variation in functional traits associated with resource-use efficiency and stress tolerance may help understand the determinants of species growth and survival under climate change (Aspinwall *et al.*, 2013).

A first consequence of increased climatic unpredictability is that ecological shifts towards increasingly ruderal strategies will be promoted. The study of flowering time variation in *A. thaliana* has demonstrated that species are not limited in the number of mutations that can promote accelerated flowering (Mendez-Vigo *et al*., 2011; Sanchez-Bermejo *et al.*, 2012; Hepworth and Dean, 2015; Whittaker and Dean, 2017;): species will therefore have ample opportunities to adapt to drought by advancing their transition to flowering (Franks *et al.* 2009). As a matter of fact, earlier flowering seems to be globally under selection (Munguía-Rosas *et al.*, 2011).

However, adapting the timing of major life history transitions will probably not suffice. As water is a paramount factor in determining both the distribution and the productivity of plant species, drought stress responses will be increasingly critical for species assemblages in most environments (Volaire, 2018). Through manipulative experiments and data fusion approaches, ecologists have learned what they may basically expect for ecosystem dynamics: Individual-level responses are followed by species reordering within communities, and finally by species losses and immigration (Smith *et al.*, 2009). These observations lack a generic understanding of individual-level responses, which are typically initiated at the molecular level, and then cascade upwards to affect plant individuals' physiology and growth (Chaves *et al.*, 2003; Avolio *et al.*, 2018). Unfortunately,

detecting and linking cascading stress responses across levels of biological organization is highly challenging (Meyer *et al.*, 2014; Lovell *et al.*, 2016), partly due to the use of different conceptual frameworks and terminologies across the different disciplines and scales (Volaire, 2018). Although ecologists have grown increasingly interested in linking molecular drought responses with physiological data from plant individuals (Lovell *et al.*, 2016; Hoffman & Smith, 2018), very few studies have up to now examined the link between different levels of biological organization in plant water stress responses (Avolio *et al.*, 2018). Besides the research challenges described above, this is partly due to a reluctance of ecologists to include an ecological outlier such as the ruderal *A. thaliana*.

Ecologists nevertheless increasingly acknowledge that an understanding of gene expression is a critical hurdle for dissecting stress response mechanisms (Hoffman & Smith, 2018). Many studies focusing on drought ecology have been conducted in perennial grasses such as *Andropogon gerardii, Sorghastrum nutans* or *Panicum virgatum* (Hoover *et al.* 2014*)*. Agronomic studies have been mostly conducted on domesticated, annual grasses such as durum wheat (*Triticum turgidum*) or barley (*Hordeum vulgare*). In light of the comparatively high ecological diversification reviewed above, one can argue that the annual forb *A. thaliana* could efficiently complement these studies. Some recent, interdisciplinary attempts have exemplified how such a diverse species could help us understand the mechanisms and ecological trade-offs of stress responses. Combining a characterization of genetic variation in drought stress resistance with current and future climate envelopes revealed the enormous adaptive potential of *A. thaliana* in the face of climate change (Exposito-Alonso *et al.* 2018a, Fournier-Level *et al.* 2015). It also documented the genetics of this potential (Fournier-Level *et al.* 2015, Exposito-Alonso *et al.* 2018b, c). Among European genotypes, it is predicted that those originating from Northern and Southern latitudes will be able to adapt in the new climate, due to their higher drought resistance as well as the genetic variability of the populations (Exposito-Alonso *et al.,* 2018b). In fact, in *A. thaliana,* it is possible to perform experiments quantifying the impact of selection in populations faced with increasingly realistic scenarios of global climate change, where exposure to drought stress, average temperature, or increased frequency of major disturbance set new limits on plant plasticity (Exposito-Alonso *et al.,* 2018c). To enhance the comparability of drought studies across model species and disciplines, drought regimes (e.g. frequency and intensity) should also be characterized with standard methods in the species (Vicca *et al.*, 2012; Ruppert *et al.*, 2015), and diagnostic experimental procedures

should be adopted to identify the ecological mechanisms promoting drought resistance (Gilbert & Medina, 2016).

As an undomesticated species *A. thaliana* has been subject to a complex suite of environmental challenges over the course of its evolutionary history (see Koch, 2018 in this issue), which promoted a diversification in ecological strategy. Its amenability to genetic approaches (e.g. seed stocks, mapping populations, mutant collections, GWAS panels) can greatly facilitate trait analysis and reveal which functional trait or trait combinations are sufficient to promote shifts in ecological strategies. For example, dissecting how the plant combines tolerance to multiple stresses, whilst at the same time fine-tuning the balance between defence, growth, and productivity is of great importance for interpreting the dynamics of plant communities (Bechtold *et al.*, 2018). Knowing which genes contribute to unobservable traits underpinning key aspect of ecological strategy can also improve the ecological classification of species. Indeed, they provide measurable proxies for traits that are difficult to measure but make important contributions to dimenstions of the ecological strategy that leaf traits and the CSR strategy scheme do not fully recapitulate.

Conclusion and Outlook

The high natural variation and the unrivaled genomic resources of *Arabidopsis thaliana* are great assets in understanding pressing questions in contemporary plant ecology but also to comprehensively dissect gene function, from the molecular to the community level. This review has assembled recent conceptual and methodological developments that show how this field is advancing. The advent of new sequencing technologies has increasingly digitalized observations both in the lab and in the field. To enhance our interpretation of these data, links between specific genes and the evolution of novel ecological preferences must be established. Evidence that variation in CSR positioning contributes to local adaptation in *A. thaliana* already suggests that variation in global ecological strategy is both heritable and relevant for understanding plant performance in diverse part of the species range (Vasseur *et al.* 2018b). Yet, variation in CSR positioning also depends on the environment in which traits are measured (Vasseur *et al.* 2018b). Several key challenges remain to be addressed such as *e.g.* i) to what extent do intraspecific changes in CSR positioning translate in changes in competitive ability, stress-resistance or tolerance to disturbance?, ii) how many traits in a trait syndrome are sufficient to initiate significant ecological shifts ?, iii) what is the importance of plasticity in shifting ecological strategies? iv) which gene or gene activity can

be used as proxy to quantify ecological dimensions that are not correctly summarized in global strategy schemes? Answering these questions in *A. thaliana* will pave the way for bridging ecology and molecular biology in Plant Sciences.

References

1001 Genomes Consortium. 2016. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell.

Ågren J, Oakley CG, Lundemo S, Schemske DW. 2017. Adaptive divergence in flowering time among natural populations of Arabidopsis thaliana: Estimates of selection and QTL mapping. Evolution **71**, 550–564.

Ågren J, Schemske DW. 2012. Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. New Phytologist **194**, 1112–1122.

Akiyama R, Ågren J. 2014 Conflicting selection on the timing of germination in a natural population of Arabidopsis thaliana. Journal of Evolutionary Biology **27**, 193–199.

Albani MC, Castaings L, Wotzel S, Mateos JL, Wunder J, Wang R, Reymond M, Coupland G. 2012. PEP1 of Arabis alpina is encoded by two overlapping genes that contribute to natural genetic variation in perennial flowering. *PLoS Genet* **8**(12): e1003130.

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M. 2009. What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? The Plant cell **21**, 1877–1896.

An D, Chen JG, Gao YQ, Li X, Chao ZF, Chen ZR, Li QQ, Han ML, Wang YL, Wang YF, Chao DY. 2017. AtHKT1 drives adaptation of Arabidopsis thaliana to salinity by reducing floral sodium content. PLoS Genet.13(10):e1007086.

Aspinwall MJ, Lowry DB, Taylor SH, Juenger TE, Hawkes CV, Johnson M-VV, Kiniry JR, Fay PA. 2013**.** Genotypic variation in traits linked to climate and aboveground productivity in a widespread C4 grass: evidence for a functional trait syndrome. *New Phytologist* **199**(4): 966-980.

Astuti G, Ciccarelli D, Roma-Marzio F, Trinco A, Peruzzi L. 2018**.** Narrow endemic species *Bellevalia webbiana* shows significant intraspecific variation in tertiary CSR strategy. *Plant Biosystems*: 1-7.

Atwell S, *et al*. 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature. 3;465(7298):627-31

Austen EJ, Rowe L, Stinchcombe JR, Forrest JRK. 2017**.** Explaining the apparent paradox of persistent selection for early flowering. *New Phytol*.

Avolio ML, Hoffman AM, Smith MD. 2018**.** Linking gene regulation, physiology, and plant biomass allocation in Andropogon gerardii in response to drought. *Plant Ecology* **219**(1): 1-15.

Bac Molenaar JA, Granier C, Keurentjes JJB, Vreugdenhil D. 2016. Genome-wide association mapping of time-dependent growth responses to moderate drought stress in Arabidopsis. Plant, Cell & Environment **39**, 88–102.

Baduel P, Arnold B, Weisman CM, Hunter B, Bomblies K. 2016**.** Habitat-Associated Life History and Stress-Tolerance Variation in Arabidopsis arenosa. *Plant Physiol* **171**(1): 437-451.

Baron E, Richirt J, Villoutreix R, Amsellem L. 2015. The genetics of intra-and interspecific competitive response and effect in a local population of an annual plant species. Functional Ecology. 29:1361-1370.

Bechtold U, Ferguson JN, Mullineaux PM. 2018. To defend or to grow: lessons from Arabidopsis C24. *Journal of Experimental Botany* **69**(11): 2809-2821.

Baxter I, Muthukumar B, Park HC, Buchner P, Lahner B, Danku J, Zhao K, Lee J, Hawkesford MJ, Guerinot ML, Salt DE. 2008. Variation in molybdenum content across broadly distributed populations of Arabidopsis thaliana is controlled by a mitochondrial molybdenum transporter (MOT1). PLoS Genet.;4(2):e1000004.

Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M, Vitek O, Salt DE. 2010. A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. PLoS Genet. 11;6(11):e1001193

Bentsink L, Yuan K, Koornneef M, Vreugdenhil D. 2003. The genetics of phytate and phosphate accumulation in seeds and leaves of Arabidopsis thaliana, using natural variation. Theor Appl Genet. May;106(7):1234-43.

Bilton MC, Whitlock R, Grime JP, Marion G, Pakeman RJ. 2010**.** Intraspecific trait variation in grassland plant species reveals fine-scale strategy trade-offs and size differentiation that underpins performance in ecological communities. *Botany* **88**(11): 939-952.

Bomblies K, Yant L, Laitinen RA, Kim S-T, Hollister JD, Warthmann N, Fitz J, Weigel D. 2010. Local-Scale Patterns of Genetic Variability, Outcrossing, and Spatial Structure in Natural Stands of Arabidopsis thaliana (R Mauricio, Ed.). PLoS Genetics **6**, e1000890.

Brachi B, Aimé C, Glorieux C, Cuguen J, Roux F. 2012. Adaptive Value of Phenological Traits in Stressful Environments: Predictions Based on Seed Production and Laboratory Natural Selection (V Laudet, Ed.). PLoS ONE **7**, e32069.

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F. 2010. Linkage and association mapping of Arabidopsis thaliana flowering time in nature. PLoS Genetics **6**, e1000940.

Burghardt LT, Metcalf C, Wilczek AM, Schmitt J. 2015. Modeling the Influence of Genetic and Environmental Variation on the Expression of Plant Life Cycles across Landscapes. The American Naturalist.

Campitelli BE, Des Marais DL, Juenger TE. 2016. Ecological interactions and the fitness effect of water-use efficiency: Competition and drought alter the impact of natural MPK12 alleles in Arabidopsis (B Enquist, Ed.). Ecology Letters **19**, 424–434.

Chao DY, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, Salt DE. 2013 Polyploids exhibit higher potassium uptake and salinity tolerance in Arabidopsis. Science. 9;341(6146):658-9.

Chao DY, Baraniecka P, Danku J, Koprivova A, Lahner B, Luo H, Yakubova E, Dilkes B, Kopriva S, Salt DE. 2014. Variation in sulfur and selenium accumulation is controlled by naturally occurring isoforms of the key sulfur assimilation enzyme ADENOSINE 5'-PHOSPHOSULFATE REDUCTASE2 across the Arabidopsis species range. Plant Physiol. ;166(3):1593-608.

Chaves MM, Maroco JP, Pereira JS. 2003**.** Understanding plant responses to drought - from genes to the whole plant. *Functional Plant Biology* **30**: 239-264.

Chiang GCK, Barua D, Dittmar E, Kramer EM, de Casas RR, Donohue K. 2013. Pleiotropy in the wild: the dormancy gene DOG1 exerts cascading control on life cycles. Evolution **67**, 883–893.

Clauss MJ, Koch MA. 2006. Poorly known relatives of Arabidopsis thaliana. Trends in Plant Science **11**, 449–459.

Clauw P, Coppens F, Korte A, *et al.* 2016. Leaf Growth Response to Mild Drought: Natural Variation in Arabidopsis Sheds Light on Trait Architecture. The Plant Cell **28**, 2417.

Coustham V, Li PJ, Strange A, Lister C, Song J, Dean C. 2012**.** Quantitative Modulation of Polycomb Silencing Underlies Natural Variation in Vernalization. *Science* **337**(6094): 584-587.

Davila Olivas NH, Kruijer W, Gort G, Wijnen CL, van Loon JJA, Dicke M. 2017. Genome-wide association analysis reveals distinct genetic architectures for single and combined stress responses in Arabidopsis thaliana. The New phytologist **213**, 838–851.

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 (JO Borevitz, Ed.). PLoS ONE **8**, e61075.

Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, Dennis ES. 2011**.** FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis. *Proc Natl Acad Sci U S A* **108**(16): 6680-6685.

Des Marais DL, McKay JK, Richards JH, Sen S, Wayne T, Juenger TE. 2012. Physiological Genomics of Response to Soil Drying in Diverse Arabidopsis Accessions. THE PLANT CELL ONLINE **24**, 893–914.

Des Roches S, Post DM, Turley NE, Bailey JK, Hendry AP, Kinnison MT, Schweitzer JA, Palkovacs EP. 2017**.** The ecological importance of intraspecific variation. *Nature Ecology & Evolution.*

Díaz S, Kattge J, Cornelissen JHC, Wright IJ, Lavorel S, Dray S, Reu B, Kleyer M, Wirth C, Colin Prentice I, et al. 2016**.** The global spectrum of plant form and function. *Nature* **529**(7585): 167-171.

Dittberner H, Korte A, Mettler-Altmann T, Weber A, Monroe G, de Meaux J. 2018. Natural variation in stomata size contributes to the local adaptation of water-use efficiency in Arabidopsis thaliana. bioRxiv, 253021.

Donohue K. 2002. Germination timing influences natural selection on life-history characters in Arabidopsis thaliana. Ecology **83**, 1006–1016.

Donohue K, Dorn D, Griffith C, Kim E, Aguilera A, Polisetty CR, Schmitt J. 2005. Niche construction through germination cueing: Life-history responses to timing of germination in Arabidopsis thaliana. Evolution **59**, 771–785.

Durvasula A, Fulgione A, Gutaker RM, *et al.* 2017. African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **209**, 201616736.

Easlon HM, Nemali KS, Richards JH, Hanson DT, Juenger TE, McKay JK. 2014. The physiological basis for genetic variation in water use efficiency and carbon isotope composition in Arabidopsis thaliana. Photosynthesis Research **119**, 119–129.

El Soda M, Kruijer W, Malosetti M, Koornneef M, Aarts MGM. Quantitative trait loci and candidate genes underlying genotype by environment interaction in the response of Arabidopsis thaliana to drought. Plant, Cell & Environment **38**, 585–599.

Exposito-Alonso M, Becker C, Schuenemann VJ, *et al.* 2018*a*. The rate and potential relevance of new mutations in a colonizing plant lineage. (J Ross-Ibarra, Ed.). PLoS Genetics **14**, e1007155.

Exposito-Alonso M, Vasseur F, Ding W, Wang G, Burbano HA, Weigel D. 2018*b*. Genomic basis and evolutionary potential for extreme drought adaptation in Arabidopsis thaliana. Nature ecology & evolution **2**, 352–358.

Exposito-Alonso M, **Brennan A, Alonso-Blanco C, Picó F.X.** 2018c. Spatio-temporal variation in fitness responses to contrasting environments in *Arabidopsis thaliana*. Evolution, **72**, 1570-1586.

Footitt S, Huang Z, Clay HA, Mead A, Finch-Savage WE. 2013. Temperature, light and nitrate sensing coordinate Arabidopsis seed dormancy cycling, resulting in winter and summer annual phenotypes. The Plant Journal **74**, 1003–1015.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in Arabidopsis thaliana. Science **334**, 86–89.

Fournier-Level A, Wilczek AM, Cooper MD, Roe JL, Anderson J, Eaton D, Moyers BT, Petipas RH, Schaeffer RN, Pieper B, et al. 2013**.** Paths to selection on life history loci in different natural environments across the native range of Arabidopsis thaliana. *Mol Ecol* **22**(13): 3552-3566.

Franks SJ. 2011. Plasticity and evolution in drought avoidance and escape in the annual plant Brassica rapa. *New Phytologist* **190**(1): 249-257.

Frachon L, Libourel C, Villoutreix R, *et al.* 2017. Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. Nature Ecology & Evolution **1**, 1551–1561.

Gilbert ME, Medina V. 2016. Drought adaptation mechanisms should guide experimental design. *Trends in Plant Science* **21**(8): 639-647.

Glander S, He F, Schmitz G, Witten A, Telschow A, de Meaux J. 2018. Assortment of flowering time and immunity alleles in natural Arabidopsis thaliana populations suggests immunity and vegetative lifespan strategies coevolve. Genome Biology and Evolution.

Grime JP. 1974**.** Vegetation classification by reference to strategies. *Nature* **250**(5461): 26.

Grime JP. 1977**.** Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *American Naturalist* **111**(982): 1169-1194.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to climate across the Arabidopsis thaliana genome. Science **334**, 83–86.

Harada H, Kuromori T, Hirayama T, Shinozaki K, Leigh RA. 2004. Quantitative trait loci analysis of nitrate storage in Arabidopsis leading to an investigation of the contribution of the anion channel gene, AtCLC-c, to variation in nitrate levels. J Exp Bot. Sep;55(405):2005-14

He F, Kang D, Ren Y, Qu L-J, Zhen Y, Gu H. 2007. Genetic diversity of the natural populations of Arabidopsis thaliana in China. Heredity **99**, 423.

He H, de Souza Vidigal D, Snoek LB, Schnabel S, Nijveen H, Hilhorst H, Bentsink L. 2014. Interaction between parental environment and genotype affects plant and seed performance in Arabidopsis. Journal of Experimental Botany **65**, 6603–6615.

van Heerwaarden J, van Zanten M, Kruijer W. 2015. Genome-Wide Association Analysis of Adaptation Using Environmentally Predicted Traits. PLOS Genetics **11**, e1005594.

Hepworth J, Dean C. 2015. Flowering Locus C's Lessons: Conserved Chromatin Switches Underpinning Developmental Timing and Adaptation. Plant Physiology **168**, 1237.

Hoffmann MH. 2005. Evolution of the realized climatic niche in the genus Arabidopsis (Brassicaceae). Evolution **59**, 1425–1436.

Hoffman AM, Smith MD. 2018**.** Gene expression differs in codominant prairie grasses under drought. *Molecular ecology resources* **18**(2): 334-346.

Hoffmann MH. 2002. Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). Journal of Biogeography **29**, 125–134.

Hoover DL, Knapp AK, Smith MD. 2014. Resistance and resilience of a grassland ecosystem to climate extremes Ecology 95:2646-2656

Hu J, Lei L, de Meaux J. 2017. Temporal fitness fluctuations in experimental Arabidopsis thaliana populations (P r K Ingvarsson, Ed.). PLoS ONE **12**, e0178990.

IPCC. 2013**.** *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Intergovernmental Panel on Climate Change Fifth Assessment Report*. Cambridge, UK: Cambridge University Press.

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C. 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science **290**, 344–347.

Kasulin L, Rowan BA, León RJC, Schuenemann VJ, Weigel D, Botto JF. 2017. A single haplotype hyposensitive to light and requiring strong vernalization dominates Arabidopsis thaliana populations in Patagonia, Argentina. Molecular Ecology **19**, 1655.

Kemi U, Niittyvuopio A, Toivainen T, Pasanen A, Quilot-Turion B, Holm K, Lagercrantz U, Savolainen O, Kuittinen H. 2013**.** Role of vernalization and of duplicated FLOWERING LOCUS C in the perennial Arabidopsis lyrata. *New Phytol* **197**(1): 323-335.

Kerdaffrec E, Filiault DL, Korte A, Sasaki E, Nizhynska V, Seren Ü, Nordborg M. 2016. Multiple alleles at a single locus control seed dormancy in Swedish Arabidopsis. eLife **5**.

Kesari R, Lasky JR, Villamor JG, Des Marais DL, Chen Y-JC, Liu T-W, Lin W, JUENGER TE, Verslues PE. 2012. Intron-mediated alternative splicing of Arabidopsis P5CS1 and its association with natural variation in proline and climate adaptation. Proceedings of the National Academy of Sciences **109**, 9197–9202.

Kiefer C, Severing E, Karl R, Bergonzi S, Koch M, Tresch A, Coupland G. 2017. Divergence of annual and perennial species in the Brassicaceae and the contribution of cis-acting variation at FLC orthologues. *Molecular Ecology* **26**(13): 3437-3457.

Koornneef M, Alonso-Blanco C, Vreugdenhil D. 2004. Naturally occurring genetic variation in Arabidopsis thaliana. Annual Review of Plant Biology **55**, 141–172.

Koprivova A, Giovannetti M, Baraniecka P, Lee BR, Grondin C, Loudet O, Kopriva S. 2013. Natural variation in the ATPS1 isoform of ATP sulfurylase contributes to the control of sulfate levels in Arabidopsis. Plant Physiol.;163(3):1133-41

Kowalski SP, Lan TH, Feldmann KA, Paterson AH. 1994. QTL mapping of naturally-occurring variation in flowering time of Arabidopsis thaliana. Mol Gen Genet. Dec 1;245(5):548-55.

Kraft NJB, Valencia R, Ackerly DD. 2008. Functional Traits and Niche-Based Tree Community Assembly in an Amazonian Forest. Science **322**, 580:582.

Kronholm I, Picó FX, Alonso-Blanco C, Goudet J, Meaux J de. 2012. Genetic Basis of Adaptation in Arabidopsis thaliana: Local Adaptation at the Seed Dormancy QTL DOG1. Evolution **66**, 2287–2302.

Lahner B, Gong J, Mahmoudian M, Smith EL, Abid KB, Rogers EE, Guerinot ML, Harper JF, Ward JM, McIntyre L, Schroeder JI, Salt DE. 2003. Genomic scale profiling of nutrient and trace elements in Arabidopsis thaliana. Nat Biotechnol.;21(10):1215-21.

Lasky JR, Des Marais DL, Lowry DB, Povolotskaya I, Mckay JK, Richards JH, Keitt TH, Juenger TE. 2014. Natural variation in abiotic stress responsive gene expression and local adaptation to climate in Arabidopsis thaliana. Molecular Biology and Evolution **31**, 2283–2296.

Le Corre V. 2005. Variation at two flowering time genes within and among populations of Arabidopsis thaliana: comparison with markers and traits. Molecular Ecology **14**, 4181–4192.

Lempe J, Balasubramanian S, Sureshkumar S, Singh A, Schmid M, Weigel D. 2005**.** Diversity of flowering responses in wild Arabidopsis thaliana strains. *PLoS Genet* **1**(1): 109-118.

Li PJ, Filiault D, Box MS, Kerdaffrec E, van Oosterhout C, Wilczek AM, Schmitt J, McMullan M, Bergelson J, Nordborg M, et al. 2014**.** Multiple FLC haplotypes defined by independent cis-regulatory variation underpin life history diversity in Arabidopsis thaliana. *Genes & Development* **28**(15): 1635-1640.

Lovell JT, Shakirov EV, Schwartz S, Lowry DB, Aspinwall MJ, Taylor SH, Bonnette J, Palacio-Mejia JD, Hawkes CV, Fay PA. 2016**.** Promises and challenges of eco-physiological genomics in the field: tests of drought responses in switchgrass. *Plant physiology*: pp. 00545.02016.

Lovell JT, Juenger TE, Michaels SD, Lasky JR, Platt A, Richards JH, Yu X, Easlon HM, Sen S, McKay JK. 2013. Pleiotropy of *FRIGIDA* enhances the potential for multivariate adaptation. Proceedings of the Royal Society B: Biological Sciences **280**.

Loudet O, Chaillou S, Merigout P, Talbotec J, Daniel-Vedele F. 2003 . Quantitative trait loci analysis of nitrogen use efficiency in Arabidopsis. Plant Physiol. Jan;131(1):345-58.

Loudet O, Saliba-Colombani V, Camilleri C, Calenge F, Gaudon V, Koprivova A, North KA, Kopriva S, Daniel-Vedele F. 2007. Natural variation for sulfate content in Arabidopsis thaliana is highly controlled by *APR2*. Nat Genet.;39(7):896-900.

Marcer A, Vidigal DS, James PMA, Fortin M-J, Méndez-Vigo B, Hilhorst HWM, Bentsink L, Alonso-Blanco C, Picó FX. 2017 Temperature fine-tunes Mediterranean Arabidopsis thaliana life-cycle phenology geographically. Plant Biology **20**, 148–156.

Marchadier E, Hanemian M, Tisne S, Bach L, Bazakos C, Gilbault E, Haddadi P, Virlouvet L, Loudet O. 2018. The complex genetic architecture of shoot growth natural variation in Arabidopsis thaliana. bioRxiv.

Masclaux F, Hammond RL, Meunier J, Gouhier-Darimont C, Keller L, Reymond P. 2010. Competitive ability not kinship affects growth of Arabidopsis thaliana accessions. New Phytologist **185**, 322–331.

Mateos JL, Madrigal P, Tsuda K, Rawat V, Richter R, Romera-Branchat M, Fornara F, Schneeberger K, Krajewski P, Coupland G. 2015**.** Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in Arabidopsis. *Genome Biol* **16**: 31.

Mateos JL, Tilmes V, Madrigal P, Severing E, Richter R, Rijkenberg CWM, Krajewski P, Coupland G. 2017**.** Divergence of regulatory networks governed by the orthologous transcription factors FLC and PEP1 in Brassicaceae species. *Proceedings of the National Academy of Sciences of the United States of America* **114**(51): E11037-E11046.

May R-L, Warner S, Wingler A. 2017**.** Classification of intra-specific variation in plant functional strategies reveals adaptation to climate. *Annals of Botany* **119**(8): 1343-1352.

McKay JK, Richards JH, Mitchell-Olds T. 2003. Genetics of drought adaptation in Arabidopsis thaliana: I. Pleiotropy contributes to genetic correlations among ecological traits. Molecular Ecology **12**, 1137–1151.

Mendez-Vigo B, Pico FX, Ramiro M, Martinez-Zapater JM, Alonso-Blanco C. 2011**.** Altitudinal and climatic adaptation is mediated by flowering traits and FRI, FLC, and PHYC genes in Arabidopsis. *Plant Physiol* **157**(4): 1942-1955.

Meyer E, Aspinwall MJ, Lowry DB, Palacio-Mejía JD, Logan TL, Fay PA, Juenger TE. 2014**.** Integrating transcriptional, metabolomic, and physiological responses to drought stress and recovery in switchgrass (*Panicum virgatum* L.). *BMC Genomics* **15**(1): 527.

Michaels SD, Amasino RM. 1999**.** FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**(5): 949-956.

Mojica JP, Mullen J, Lovell JT, Monroe JG, Paul JR, Oakley CG, Mckay JK. 2016. Genetics of water use physiology in locally adapted Arabidopsis thaliana. Plant Science **251**, 12–22.

Montesinos-Navarro A, Wig J, Picó FX, Tonsor SJ. 2011. Arabidopsis thaliana populations show clinal variation in a climatic gradient associated with altitude. New Phytologist **189**, 282–294.

Morrison GD, Linder CR. 2014. Association Mapping of Germination Traits in Arabidopsis thaliana Under Light and Nutrient Treatments: Searching for G x E Effects. G3: Genes, Genomes, Genetics, g3.114.012427.

Munguía-Rosas MA, Ollerton J, Parra-Tabla V, De-Nova JA. 2011. Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured. Ecology Letters **14**, 511–521.

Muñoz-Parra E, Pelagio-Flores R, Raya-González J, Salmerón-Barrera G, Ruiz-Herrera LF, Valencia-Cantero E, López-Bucio J. 2017. Plant–plant interactions influence developmental phase transitions, grain productivity and root system architecture in Arabidopsis via auxin and PFT1/MED25 signalling. Plant, Cell & Environment **40**, 1887–1899.

Novikova PY, Hohmann N, Nizhynska V, *et al.* 2016. Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. Nature Genetics **48**, 1077–1082.

Picó FX. 2012. Demographic fate of Arabidopsis thaliana cohorts of autumn- and spring-germinated plants along an altitudinal gradient. Journal of Ecology **100**, 1009–1018.

Pico FX, Mendez-Vigo B, Martinez-Zapater JM, Alonso-Blanco C. 2008. Natural Genetic Variation of Arabidopsis thaliana Is Geographically Structured in the Iberian Peninsula. Genetics **180**, 1009–1021.

Pierce S, Negreiros D, Cerabolini BEL, Kattge J, Díaz S, Kleyer M, Shipley B, Wright SJ, Soudzilovskaia NA, Onipchenko VG, et al. 2017**.** A global method for calculating plant CSR ecological strategies applied across biomes world-wide. *Functional Ecology* **31**(2): 444-457.

Postma FM, Ågren J. 2015. Maternal environment affects the genetic basis of seed dormancy in Arabidopsis thaliana. Molecular Ecology **24**, 785–797.

Postma FM, Ågren J. 2016. Early life stages contribute strongly to local adaptation in Arabidopsis thaliana. Proceedings of the National Academy of Sciences **113**, 7590–7595.

Ravenscroft CH, Fridley JD, Grime JP. 2014**.** Intraspecific functional differentiation suggests local adaptation to long-term climate change in a calcareous grassland. *Journal of Ecology* **102**(1): 65-73.

Reed SC, Coe KK, Sparks JP, Housman DC, Zelikova TJ, Belnap J. 2012**.** Changes to dryland rainfall result in rapid moss mortality and altered soil fertility. *Nature Clim. Change* **2**(10): 752-755.

Reich PB. 2014. The world-wide 'fast–slow' plant economics spectrum: a traits manifesto. *Journal of Ecology* **102**(2): 275-301.

Reymond M, Svistoonoff S, Loudet O, Nussaume L, Desnos T. 2006**.** Identification of QTL controlling root growth response to phosphate starvation in Arabidopsis thaliana. Plant Cell Environ. ;29(1):115-25.

Riboni M, Galbiati M, Tonelli C, Conti L. 2013. GIGANTEA enables drought escape response via abscisic acid-dependent activation of the florigens and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS. *Plant Physiol* **162**(3): 1706-1719.

Ruppert JC, Harmoney K, Henkin Z, Snyman HA, Sternberg M, Willms W, Linstädter A. 2015**.** Quantifying drylands' drought resistance and recovery: The importance of drought intensity, dominant life history and grazing regime. *Global Change Biology* **21**: 258–1270.

Rus A, Baxter I, Muthukumar B, Gustin J, Lahner B, Yakubova E, Salt DE. 2006. Natural variants of AtHKT1 enhance Na+ accumulation in two wild populations of Arabidopsis. PLoS Genet. ;2(12):e210.

Sanchez-Bermejo E, Mendez-Vigo B, Pico FX, Martinez-Zapater JM, Alonso-Blanco C. 2012. Novel natural alleles at FLC and LVR loci account for enhanced vernalization responses in Arabidopsis thaliana. Plant, Cell Env. **35**:1672-1684

Salt DE, Baxter I, Lahner B.2008 Ionomics and the study of the plant ionome.. Annu Rev Plant Biol. 2008;59:709-33.

Sasaki E, Zhang P, Atwell S, Meng D, Nordborg M. 2015. 'Missing' G x E Variation Controls Flowering Time in Arabidopsis thaliana (G Gibson, Ed.). PLoS Genetics **11**, e1005597.

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. Nature Reviews Genetics **14**, 807–820.

Segrestin J, Bernard-Verdier M, Violle C, Richarte J, Navas ML, Garnier E. 2018**.** When is the best time to flower and disperse? A comparative analysis of plant reproductive phenology in the Mediterranean. *Functional Ecology*.

Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005**.** Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. *Plant Physiol* **138**(2): 1163-1173.

Smith MD, Knapp AK, Collins SL. 2009**.** A framework for assessing ecosystem dynamics in response to chronic resource alterations induced by global change. *Ecology* **90**(12): 3279-3289.

Stock AJ, McGoey BV, Stinchcombe JR. 2015. Water availability as an agent of selection in introduced populations of Arabidopsis thaliana: impacts on flowering time evolution (J Ross-Ibarra, Ed.). PeerJ **3**, e898.

Svardal H, Farlow A, Exposito-Alonso M, Ding W, Novikova P, Alonso-Blanco C, Weigel D, LEE C-R, Nordborg M. 2017. On the post-glacial spread of human commensal Arabidopsis thaliana. Nature Communications **8**, 1–12.

Tabas-Madrid D, Méndez-Vigo B, Arteaga N, Marcer A, Pascual-Montano A, Weigel D, Picó FX, Alonso-Blanco C. 2018. Genome-wide signatures of flowering adaptation to climate temperature: Regional analyses in a highly diverse native range of Arabidopsis thaliana. Plant, Cell & Environment **41**, 1806–1820.

Taylor MA, Cooper MD, Sellamuthu R, Braun P, Migneault A, Browning A, Perry E, Schmitt J. 2017. Interacting effects of genetic variation for seed dormancy and flowering time on phenology, life history, and fitness of experimental Arabidopsis thalianapopulations over multiple generations in the field. New Phytologist **216**, 291–302.

Tomatsu H, Takano J, Takahashi H, Watanabe-Takahashi A, Shibagaki N, Fujiwara T. 2007. An Arabidopsis thaliana high-affinity molybdate transporter required for efficient uptake of molybdate from soil. Proc Natl Acad Sci U S A.;104(47):18807-12

Toomajian C, Hu TT, Aranzana MJ, Lister C, Tang C, Zheng H, Zhao K, Calabrese P, Dean C, Nordborg M. 2006. A nonparametric test reveals selection for rapid flowering in the Arabidopsis genome. PLoS Biology **4**.

Vasseur F, Exposito-Alonso M, Ayala-Garay OJ, Wang G, Enquist BJ, Vile D, Violle C, Weigel D. 2018a. Adaptive diversification of growth allometry in the plant *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences **115**, 3416.

Vasseur F, Sartori K, Baron E, Fort F, Kazalou E, Segrestin J, Garnier E, Vile D, Violle C. 2018b. Climate as a driver of adaptive variations in ecological strategies in *Arabidopsis thaliana*. Ann. Bot. https://doi.org/10.1093/aob/mcy165

Vicca S, Gilgen AK, Camino Serrano M, Dreesen FE, Dukes JS, Estiarte M, Gray SB, Guidolotti G, Hoeppner SS, Leakey ADB, et al. 2012**.** Urgent need for a common metric to make precipitation manipulation experiments comparable. *New Phytologist* **195**(3): 518-522.

Vidigal DS, Marques ACSS, Willems LAJ, Buijs G, Méndez-Vigo B, Hilhorst HWM, Bentsink L, Picó FX, Alonso-Blanco C. 2016. Altitudinal and climatic associations of seed dormancy and flowering traits evidence adaptation of annual life cycle timing in Arabidopsis thaliana. Plant, Cell & Environment **39**, 1737–1748.

Volaire F. 2018**.** A unified framework of plant adaptive strategies to drought: Crossing scales and disciplines. *Global Change Biology* **24**(7): 2929-2938.

Weigel D, Nordborg M. 2015. Population Genomics for Understanding Adaptation in Wild Plant Species. Annual Review of Genetics **49**, 315–338.

Westoby M, Falster SD, Moles AT, Vesk PA, Wright IJ. 2002. Plant Ecological Strategies: Some leading dimensions of Variation Between Species. Annu. Rev. Ecol. Syst. **33**:125-159.

Whittaker C, Dean C. 2017. The FLC Locus: A Platform for Discoveries in Epigenetics and Adaptation. Annual Review of Cell and Developmental Biology **33**, 555–575.

Wilczek AM, Roe JL, Knapp MC, *et al.* 2009. Effects of genetic perturbation on seasonal life history plasticity. Science **323**, 930–934.

Wuest SE, Niklaus PA. 2018. A plant biodiversity effect resolved to a single locus. bioRxiv.

Zou Y-P, Hou X-H, Wu Q, *et al.* 2017. Adaptation of Arabidopsis thaliana to the Yangtze River basin. Genome Biology **18**, 239.

Data Availability

All sequence data used in Chapter 1 are available in either NCBI Short Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) or in 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; whole genome sequences generated for this project and the rest of PL sequences).All sequence data used for the analysis in Chapter 2 and Chapter 3 will be shortly available in ENA. Morphological data used in the analysis are either available as a Supplementary file, or will be available in the database dryad.

Acknowledgments

First and foremost, I would like to thank my supervisor Prof. Dr. Juliette de Meaux, not only for the opportunity to work on this exciting project but also for her excellent mentoring. Juliette's advice has helped me develop my scientific skills and she has been invaluable for my personal growth over the last four years.

I would like to thank Dr. Holger Schielzeth, Dr Helmi Kuittinen and Dr Markus Stetter for participating in my thesis committee and providing guidance whenever needed. My co-authors Dr Tuomas Hämälä, Dr Kim A Steige, Dr Evan Koch, Dr. Hannes Dittberner, Dr. Shamil Sunyaev, Dr. Xavier Vekemans, Dr. Vincent Castric and Dr. Outi Savolainen have provided me with excellent cooperation and have contributed significantly to the increase of my knowledge and I am greatful for the cooperation. The expertise and the invigorating scientific discussions with the various members of the AG de Meaux have been of great help to me. Thus, I would like to thank Dr. Fei He, Benedict Wieters, Dr. Gregor Schmitz, Lea Hoerdemann, Dr. Ulrike Goebel and Dr. Maroua Bouzid Elkhesairi. I greatly appreciate Kirsten Bell for her support during plant experiments and laboratory work.

Furthermore, this thesis would not have been completed without Hannes, Kim, Alex, Tuuli, Veri, Xuan, Leif, Laura and Aggelos, whose friendship has always been there for me, even if some of them were a couple of thousand kilometers away.

Finally, I have to thank my family, Konstantinos, Chrysoula and Alexandra, who always believe in me. Ευχαριστώ πολύ!

Margarita Takou

Soemmeringstr 45, 50823, Cologne. Germany

Mobile: +491728990997 email: mtakou1@uni-koeln.de

Current Position

I am currently working as a PhD student at the group of prof Dr Juliette de Meaux. The group is part of the Botanical Institute of the University of Cologne.

My PhD project "The polygenic basis of local adaptation in *Arabidopsis lyrata*" has as primary focus to study the adaptive potential of gene expression and *cis* regulatory evolution within species. As a first step, population differentiation and adaptation between a range edge and range core population of the species is studied on the genomic and phenotypic level. Then, the polygenic selection is quantified and studied in the transcriptomes of interpopulation crossings, produced via a nested crossing scheme. Specifically, the evolution and future adaptive potential is assessed both by allele specific expression and partitioning of gene expression variance to its heritable and non-heritable components.

Supervisor: prof Dr Juliette de Meaux

Education

June 2016-present

PhD Student in Evolutionary Biology, (AG de Meaux, Botanical Institute, University of Cologne, Germany)

During the PhD study skills related to genomics, transcriptomics, quantitative and evolutionary genetics, statistical analysis, and plant ecology are being acquired through working on the main project, as well as by interacting with senior scientists and participating in workshops.

PhD thesis title: "The polygenic basis of local adaptation in *Arabidopsis lyrata*"

September 2014-June 2016

Master Degree in Ecology and Population Genetics, (Department of Biology, University of Oulu, Finland)

Emphasis of the studies was given on population genetics, including: Basics of population genetics, DNA analysis in population genetics and evolutionary genomics. I also attended computer science and bioinformatics courses. Detailed record of courses can be provided upon request.

Master thesis title: "Altitudinal and latitudinal variation in cessation of flowering and growth and its molecular basis in *Arabidopsis lyrata*".

September 2009-November 2013

Bachelor Degree in Biology, (School of Biology, Faculty of Science, Aristotele University of Thessaloniki, Greece)

During my bachelor degree I studied genetics, zoology, botany and ecology. Emphasis was given to genetic topics and as well as plant physiology. The first semester of 2012 I participated in the LLP/Erasmus Exchange Program, at Universidad Complutense de Madrid. Detailed record of courses can be provided upon request.

Bachelor thesis title: "Photoprotective energy dissipation in the presence/absence of PsbS protein in *Arabidopsis thaliana* under drought stress"

Past Projects

Master Thesis Project: "Altitudinal and latitudinal variation in cessation of flowering and growth and its molecular basis in *Arabidopsis lyrata*"

Primary focus to detect morphological differences of growth and flowering traits of *Arabidopsis lyrata* from different altitudes and latitudes. Secondary focus to determine expression differences of flowering time network genes, evaluated and compared between populations and flowering states. The results indicated significant differences of the cessation of flowering and timing of growth cessation between Northern and Southern populations, as it was expected due to different growing seasons the populations have to adapt to. A role of *FT* and *GI* in those phenotypes was indicated, but the results were inconclusive due to small sample sizes.

During my thesis I mainly worked with *Arabidopsis lyrata* adult plants and gained knowledge of life history traits, flowering time gene network, identifying key plant morphological and developmental traits and qPCR, as well as statistical analysis.

Supervisors: Dr Helmi Kuittinen & Dr Outi Savolainen.

Bachelor Thesis Project: "Photoprotective energy dissipation in the presence/absence of PsbS protein in *Arabidopsis thaliana* under drought stress."

Study of dissipation of excess absorbed light energy in photosystem II (PSII) antenna as heat, through a process known as non-photochemical quenching (NPQ) under drought stress conditions. It was found that under drought stress the mutant line *npq4*, which lacks functional PsbS protein, allocated energy to electron transport to contribute to photoprotection resulting in lower excitation pressure of the PSII than in wild type plants. The existence of this photoprotective mechanism of energy dissipation in *npq4* mutants is believed to occur through cyclic electron flow around photosystem I.

During my thesis I mainly worked with imaging PAM fluometer machine, *Arabidopsis thaliana* and gained deeper knowledge of the photoprotective system of plants.

Supervisor: Dr Michael Moustakas.

Work experience

July 2015- August 2015

Research Assistant at the Department of Biology, University of Oulu, Finland.

Assisting within a field study, primarily focused on collecting phenotypic data related to flowering cessation, seed production and vegetative growth. Sample collections for RNA extractions and evaluation of related to the topic information. Work with the UNIX systems and the programming language Perl was also included.

April 2013-May 2013

Laboratory Technician at ELGO-"DEMETER" Cereal Institute, Thessaloniki, Greece

Preparation of cereal samples and execution of biochemical methodologies to identify quantity of tocopherols, tocotrienols and carotenoids. Analysis of the results with HPLC machine. Chlorophyll extraction from *Arabidopsis thaliana* and statistical analysis of the quantity.

Published Paper

M Takou, B Wieters, S Kopriva, G Coupland, A Linstädter, J De Meaux (2019). Linking genes with ecological strategies in *Arabidopsis thaliana*..Journal of experimental botany 70 (4), 1141-1151

M Takou, T Hämälä, KA Steige, E Koch, H Dittberner, L Yant, M Genete, S Sunyaev, V Castric, X Vekemans, O Savolainen, J de Meaux (2020). Maintenance of adaptive dynamics in a bottlenecked range-edge population that retained out-crossing. BioRXiv, doi: https://doi.org/10.1101/709873

Conference Presentations

Margarita Takou, Tuomas Hämälä, Kim Steige, Hannes Dittberner, Vincent Castric, Levi Yant, Xavier Vekemans, Outi Savolainen, Juliette de Meaux. How do Arabidopsis lyrata ssp. petreae populations at the edge of the distribution adapt, while coping with decreased genetic variation? SMBE Satellite Meeting: Towards an integrated concept of adaptation: Uniting molecular population genetics and quantitative genetics, 2019, Vienna.

Margarita Takou & Juliette de Meaux. "Local adaptation and contribution of small effect mutations to polygenic adaptation in *Arabidopsis lyrata*". 23. European Meeting of PhD Students in Evolutionary Biology, 2017, Krasiczyn Castle, Poland.

Margarita Takou, Ilektra Sperdouli and Michael Moustakas. "Photoprotective energy dissipation in the presence or absence of PsbS protein in Arabidopsis thaliana under drought stress". Proceedings 35th Annual Conference of Hellenic Society for Biological Sciences, Nafplio 2013, Greece, pp 340-341.

Computer Skills

Knowledge of Python and R programming languages for processing and statistical analysis of genomic, transcriptomic, and phenotypic datasets.

Knowledge of Linux based software and command line to analyse population genomic and transcriptomic data.

Knowledge of Markov Chains Monte Carlo algorithms and generalized linear mixed models for partitioning phenotypic variance into its components.

I have gained knowledge of using bioinformatics tools such as *fastsimcoal2*, BEAST2, STRUCTURE, Admixture, DNAsp, Arlequin, bwa, samtools.

Intermediate knowledge of Microsoft office and Adobe Photoshop.

Basic knowledge of topics of algorithms and data structures.

Language Skills

My mother language is Greek. In addition to this I have developed language skills that have been evaluated according to the Common European Reference for Languages: English (level C2), Spanish (level C1), German (level B1) and Finnish (level A1).

Additional Skills

Soft skills were acquired while volunteering work as event coordinator for Network of International Students in Oulu, NISO ry. The duties included coordination of events to help new students to integrate in the university society and network through promoting cultural exchange in a relaxed atmosphere. Through my experience I gained experience in organizing events, coordinating team work, communicating through social media and communicating with people with multicultural backgrounds, skills that can be applied for building relationships and manage team collaboration in any professional level.

Teaching and communication skills were acquired by teaching bioinformatic skills, such as statistical analysis in R and python, to various bachelor students working in the de Meaux lab. During the process, I gained insight in judging the abilities and information gaining abilities of persons collaborating with.

Clean driver's license, class B.

Personal Interests

Personal hobbies include participating in creative writing groups, pen and paper tabletop game, as well as in the activities of book clubs. Joined the training of archery club based in Cologne.

Future plans

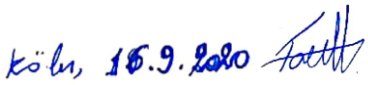
I am interested in exploring the evolution of small effect mutations and the potential for polygenic adaption in wild sexual populations. Specifically, what is the adaptive potential of small effect populations in the wild? What is the

impact of the selection and population parameter interaction on polygenic adaptation? Can we trace their impact on phenotypic differentiation between environments on a large scale?

References

Available upon request.

Date, Signature



Erklärung zum Promotionsverfahren

(gemäß der Promotionsordnung vom 02. Februar 2006 mit den Änderungsordnungen vom 10. Mai 2012, 16. Januar 2013 und 21. Februar 2014)

(Nichtzutreffendes bitte streichen; ggf. näher erläutern und entsprechende Unterlagen beifügen!)

1. Frühere Promotionsverfahren

1.1 Ich habe bereits einen Doktortitel erworben oder ehrenhalber verliehen bekommen.
(nähere Angaben)

Nein.

1.2 Für mich ist an einer anderen Fakultät / Universität ein Promotionsverfahren eröffnet worden, aber noch nicht abgeschlossen.
(nähere Angaben)

Nein

1.3 Ich bin in einem Promotionsverfahren gescheitert.
(nähere Angaben)

Nein

2. Ich bin wegen einer vorsätzlichen **Straftat** zu einer Freiheitsstrafe von mindestens einem Jahr ohne Bewährung rechtskräftig verurteilt worden bzw. es läuft ein Strafverfahren gegen mich.
(nähere Angaben)

Nein.

Ich versichere, alle Angaben wahrheitsgemäß gemacht zu haben.

Margarita Takou

Nachname in Druckbuchstaben

2.11.2020

Datum

Unterschrift

Erklärung zur Dissertation

gemäß der Promotionsordnung vom 02. Februar 2006 mit den Änderungsordnungen vom 10. Mai 2012, 16. Januar 2013 und 21. Februar 2014

Diese Erklärung muss in der Dissertation enthalten sein

"Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit − einschließlich Tabellen, Karten und Abbildungen −, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie − abgesehen von unten angegebenen Teilpublikationen − noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.
Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von (*Name des anleitenden Dozenten oder der anleitenden Dozentin*) betreut worden."

Teilpublikationen:

M Takou, B Wieters, S Kopriva, G Coupland, A Linstädter, J De Meaux (2019). Linking genes with ecological strategies in Arabidopsis thaliana.. Journal of Experimental Botany 70 (4), 1141-1151

M Takou, T Hämälä, KA Steige, E Koch, H Dittberner, L Yant, M Genete, S Sunyaev, V Castric, X Vekemans, O Savolainen, J de Meaux (2020). Maintenance of adaptive dynamics in a bottlenecked range-edge population that retained out-crossing. BioRXiv, doi: https://doi.org/10.1101/709873

Datum / Unterschrift

2.11.2020 , Frau

Non-official English translation **(The German version must be included in the doctoral thesis)**

"I declare that I have independently completed the dissertation I submitted, that the sources and tools used are completely cited and that parts of the dissertation - including tables, maps and figures -, taken from other sources (in the wording or the sense) in each individual case has been referred to as such. Further, I declare that this dissertation has not been submitted to any other Faculty or University and that - apart from the following partial publications - has not yet been published, and that I will not publish the dissertation before the end of the doctoral examination. I am aware of the requirements of the doctoral regulations. The doctoral project and Dissertation has been supervised by (*name of the supervisor*)."

Partial publications of the thesis:

Date / Signature