Evolutionary processes shaping natural variation in two *Brassicaceae* species

Inaugural-Dissertation

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

vorgelegt von

Hannes Dittberner

aus Schwerin

Köln, 2019

Berichterstatter:	Prof. Dr. Juliette de Meaux
(Gutacher)	Prof. Dr. Thomas Wiehe
	Prof. Dr. Michael Lenhard
Tag der Disputation:	15.07.2019

PUBLICATIONS

Dittberner, H., Korte, A., Mettler-Altmann, T., Weber, A. P. M., Monroe, G., & de Meaux, J. (2018). Natural variation in stomata size contributes to the local adaptation of water-use efficiency in Arabidopsis thaliana. *Molecular Ecology*, *27*(20), 4052–4065.

Dittberner, H., Becker, C., Jiao, W.-B., Schneeberger, K., Hölzel, N., Tellier, A., & Meaux, J. de. (2019). Strengths and potential pitfalls of hay transfer for ecological restoration revealed by RAD-seq analysis in floodplain Arabis species. Molecular Ecology, 28(17), 3887–3901.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In natural populations genetic variation is shaped by a complex interplay of evolutionary forces. For this thesis, I investigated patterns of natural variation in two selfing *Brassicaceae* species with contrasting demographic histories. I addressed the following questions: i) how do complex traits evolve in selfing populations when genetic drift is maximized and recombination strongly limited? ii) how can such populations be maintained when they are severely endangered? To answer the first question, I investigated natural variation in stomatal traits and water-use efficiency in 330 European accessions of the widely distributed human commensal *Arabidopsis thaliana*. A genome-wide association study (GWAS) revealed that natural variation in stomata density, stomata size and water-use efficiency has a complex largely polygenic genetic basis with few major effect loci at low frequency. Moreover, I found a significant correlation between stomata size and water-use efficiency, which has a genetic basis. All traits were significantly correlated with climatic variables and excessively differentiated among populations, suggesting a role of these traits in local adaptation.

To answer the second question, I investigated the distribution of genetic diversity in *Arabis nemorensis*, a strongly endangered floodplain species. *A. nemorensis* is a target species in an ecological restoration project at the Upper Rhine. To assess whether genetic diversity was maintained in the restoration process I genotyped and compared individuals from four pristine and six restored sites. Genetic analysis revealed that, in these sites, *A. nemorensis* co-occurs with its morphologically highly similar but ecologically divergent sibling species *Arabis sagittata* and that they naturally hybridize. In both species, there was no difference in the level of genetic diversity between pristine and restored sites. In *A. sagittata*, restoration resulted in admixture of previously isolated genotypes, suggesting that restoration can increase the adaptive potential of populations, depending on the initial structure of the donor populations.

Population genetic analysis of 15 additional pristine sites in Germany in Austria revealed that *A. nemorensis* is frequently confused with its sibling species, *A. sagittata* and *Arabis hirsuta*, in botanical surveys, indicating that the size of its total population might be overestimated. In three populations *A. nemorensis* co-occurs with *A. hirsuta*. However, the Rhine population is the only contact zone between *A. nemorensis* and *A. sagittata* I found. Intraspecific genetic diversity was low both in *A. nemorensis* and *A. sagittata*, likely due to habitat degradation. Thus, interspecific gene-flow through hybridization could be source of novel genetic variation for both species, which could be critical for their survival. Patterns of genomic ancestries of hybrids suggest that hybrids naturally back-cross with both parents, but preferentially with *A. sagittata*, which might have resulted in interspecific gene-flow. To test for interspecific gene-flow, I analyzed whole-genome sequences of 35 individuals from sympatric and allopatric populations of both species and an outgroup. In both sympatric and allopatric populations, I found signatures of substantial gene-flow among parental species, which was stronger from *A. nemorensis* into *A. sagittata* than vice versa and strongest into the sympatric

A. sagittata population. Haplotype network analyses suggest that gene-flow in this population was both recent and ancestral. To assess the adaptive potential of interspecific gene-flow, I investigated the phenotypic divergence of the species. I found that they significantly differ in several potentially adaptive traits: phenology, morphology, defense and flooding tolerance, highlighting the adaptive potential of interspecific gene-flow. Yet, additional studies will be needed to assess whether gene-flow is indeed adaptive.

ZUSAMMENFASSUNG

In natürlichen Populationen wird genetische Variation durch ein komplexes Zusammenspiel evolutionärer Kräfte beinflusst. In dieser Dissertation habe ich die natürliche Variation in zwei selbstenden *Brassicaceae* Spezies mit unterschiedlichen demografischen Vorgeschichten untersucht. Ich beschäftigte mich mit folgenden Fragestellungen: i) wie evolvieren komplexe Merkmale in selbstenden Populationen, wenn genetischer Drift maximal und Rekombination stark limitiert ist? ii) wie können solche Populationen erhalten bleiben, wenn sie stark gefährdet sind? Um die erste Frage zu beantworten habe ich natürliche Variation von Stomata-Merkmalen und Wassernutzungseffizienz in 330 europäischen Linien der weit-verbreiteten Spezies *Arabidopsis thaliana* untersucht. Eine genetische Assoziationsstudie zeigte, dass natürliche Variation von Stomatadichte, Stomatagröße und Wassernutzungseffizienz eine komplexe und größtenteils polygene genetische Grundlage hat, mit Ausnahme einiger Loci mit relativ großem Effekt und niedriger Frequenz. Außerdem fand ich eine signifikante Korrelation zwischen Stomatagröße und Wassernutzungseffizienz, die eine genetische Grundlage hat. Alle Merkmale waren signifikant mit klimatischen Variablen korreliert und übermäßig zwischen den Populationen differenziert, was eine Rolle dieser Merkmale in der Anpassung an lokale Bedingungen suggeriert.

Um die zweite Frage zu beantworten, habe ich die Verteilung der genetischen Diversität von *Arabis nemorensis*, einer stark gefährdeten Art der Stromtalwiesen, untersucht. *A. nemorensis* ist eine Zielart in einem Projekt zur ökologischen Restauration am Oberrhein. Um zu beurteilen ob genetische Diversität im Restaurationsprozess erhalten bleibt, habe ich Individuen aus vier ursprünglichen und sechs restaurierten Populationen genotypisiert und verglichen. Meine genetischen Untersuchungen haben gezeigt, dass *A. nemorensis* zusammen mit einer morphologisch ähnlichen aber ökologisch verschiedenen Schwesterart, *A. sagittata*, vorkommt und dass die Arten natürlich hybridisieren. In beiden Arten gab es keinen Unterschied bezüglich der genetischen Diversität zwischen ursprünglichen und restaurierten Populationen. Bei *A. sagittata* hat die Restauration zu Durchmischung von zuvor isolierten Genotypen geführt, was suggeriert, dass Restauration das Potential zur Anpassung von Populationen, in Abhängigkeit der Ausgangsbedingungen in den Spenderpopulationen, erhöhen kann.

Populationsgenetische Analysen von 15 zusätzlichen ursprünglichen Populationen in Deutschland und Österreich haben gezeigt, dass *A. nemorensis* in botanischen Untersuchungen häufig verwechselt wird, was bedeutet, dass die Größe der Gesamtpopulation möglichweise überschätzt wird. In drei Populationen kam *A. nemorensis* gemeinsam mit *Arabis hirsuta* vor. Jedoch ist die Rheinpopulation die einzige Kontaktzone zwischen *A. nemorensis* und *A. sagittata*, die ich finden konnte. Die intraspezifische Diversität in *A. nemorensis* und *A. sagittata* war niedrig, vermutlich aufgrund der Zerstörung ihres Lebensraums. Deshalb könnte interspezifischer Genfluss durch Hybridisierung eine wichtige Quelle für neue genetische Variation sein, die zum Überleben beider Arten beitragen könnte.

Genomische Abstammungsmuster in Hybriden suggeriert, dass Hybriden mit beiden Eltern rückkreuzen, aber vorzugsweise mit *A. sagittata*. Um den interspezifischen Genfluss zu charakterisieren, habe ich vollständige Genomsequenzen von 35 Individuen aus sympatrischen und allopatrischen Populationen beider Spezies und einer Außenseiterspezies analysiert. Sowohl in sympatrischen als auch allopatrischen Populationen habe ich Signaturen von erheblichem interspezifischem Genfluss gefunden, welcher stärker von *A. nemorensis* zu *A. sagittata* als andersrum und am stärksten in der sympatrischen *A. sagittata* Population war. Haplotypnetzwerkanalysen haben gezeigt, dass Genfluss in der sympatrischen Population sowohl vor kurzem als auch anzestral stattgefunden hat. Um das Anpassungspotential von interspezifischem Genfluss zu einzuschätzen, habe ich die phänotypische Divergenz der Spezies untersucht. Ich fand signifikante Unterschiede in mehreren potenziell adaptiven Phänotypen: Phänologie, Morphologie, Abwehr und Überflutungstoleranz, was das Anpassungspotential des interspezifischen Genflusses hervorhebt. Zusätzliche Studien sind nötig um zu testen, ob der Genfluss tatsächlich adaptiv ist.

1 INTRODUCTION

Life has evolved into a vast diversity of currently about 8.7 million species since its origin 3.5 billion years ago (Mora, Tittensor, Adl, Simpson, & Worm, 2011; Schopf, Kudryavtsev, Czaja, & Tripathi, 2007). Evolution requires natural genetic variation, i.e. variation in the DNA sequence among organisms. Genetic variation is found in all species, but the levels of variation differ among species (Leffler et al., 2012). As genetic variation can result in phenotypic variation, it is instrumental for the adaptation of populations to their environment. Yet, adaptation is not the only evolutionary force shaping variation in natural populations. The interplay of these forces is the basis for all evolutionary processes and thus understanding how these forces work is key to understand how life works.

1.1 EVOLUTIONARY FORCES SHAPING NATURAL VARIATION

Mutations are the major source for novel genetic variation. There are several types of mutations e.g., single nucleotide substitutions, insertions, deletions, inversions, transposable element insertions, chromosomal re-arrangements, transpositions and duplications (Houle & Kondrashov, 2006). The results of mutations, i.e. different versions of the same genetic locus, are called alleles. Mutations occur in every living organism and are caused by errors during DNA replication or external factors damaging DNA like chemicals and radiation. However, the mutation rate differs among taxa and is correlated, for example, with genome size and population size (Lynch, 2010). Mutations can have different effects depending on their type and location in the genome. Most mutations in intergenic regions and introns have no effect for the organism and are not directly targeted by natural selection (Palazzo & Gregory, 2014). In contrast, mutations in protein-coding or promotor regions can have various effects, e.g. amino acid changes, splicing variation, introduction of stop codon, reading frame shifts and changes in gene regulation. Additionally, large scale mutations may lead to gene duplications or loss of genes (Houle & Kondrashov, 2006). These functional mutations can translate into a phenotype and can have positive, neutral or negative effects on an organism's fitness (Eyre-Walker & Keightley, 2007). In sexually reproducing organisms, recombination is another source of novelty. It allows de-coupling of physically linked alleles with positive and negative effects and the creation of new advantageous allele combinations.

While mutation and recombination create novel genetic variation, other processes determine the fate genetic variants. These processes are genetic drift and natural selection. Genetic drift describes stochastic changes in allele frequencies due to random sampling of reproducing individuals (Masel, 2011). The strength of genetic drift depends on the size of the population and is strongest in small populations. However, the effective population size, an idealized measure of the strength genetic drift, can deviate strongly from the census size, e.g. due to non-random mating, selfing and differences in female and male abundance (Crow, 2010). In sufficiently large populations allele frequencies do not change from generation to generation in the absence of other evolutionary forces (Brian Charlesworth & Charlesworth, 2010). The effect of drift is especially strong on rare alleles, e.g. new mutations, as these can become lost easily and then cannot "recover" anymore. Thus, even in large populations most new mutations are quickly lost due to drift. Over long time periods, even common alleles can be lost or fixed by drift alone (Brian Charlesworth & Charlesworth, 2010). Thus, genetic drift can be responsible for significant divergence between populations, which can finally lead to speciation (N. H. Barton, 2010). The strength of genetic drift between populations depends not only on the size of the populations, but also increases with the degree of isolation between populations, e.g. due to geographic barriers or distance.

While genetic drift acts on all genetic variants, natural selection can only directly act on variants conferring a phenotype that affects fitness. However, due to physical linkage it can also affect neutral genetic variation. There are different modes of selection that depend on the fitness effect of the genotype. Negative or purifying selection purges deleterious genotypes from the population. Due to genetic linkage this affects not only loci with deleterious genotypes but also linked variant loci of the genetic background, an effect called background selection (B. Charlesworth, Morgan, & Charlesworth, 1993). For example, if an individual has a lethal genotype at one locus, all other alleles present in that individual also reduce in frequency in the population. Thus, negative and background selection reduce levels of genetic variation. Positive selection raises the frequency of beneficial alleles in the population. There are two types of positive selection: directional selection and balancing selection. Directional selection acts when there is a fitness optimum and the allele can shift the population distribution toward that optimum (Mitchell-Olds, Willis, & Goldstein, 2007). For example, if early flowering is advantageous in a population of plants, alleles conferring

early flowering will increase in frequency and eventually fix. Thus, directional selection also reduces genetic variation in the population. Again, this effect is magnified since sites linked to the beneficial allele are also fixed in the process, a phenomenon called selective sweep. In contrast, balancing selection actively maintains multiple alleles at a single locus, thus conserving genetic variation (Mitchell-Olds et al., 2007). This occurs, for example, when fitness of heterozygotes is higher than of homozygotes (overdominance) or when the effect of an allele depends on its frequency (frequency dependent selection). Divergent selection between populations, i.e. directional selection in opposite directions, can promote population isolation and differentiation, which can in turn result in speciation (Via, 2009).

Since genetic drift and natural selection act on the same loci, they are not independent of each other. Genetic drift is often disruptive for selective processes. For example, most beneficial mutations are lost due to drift while they are at low frequency (N. Barton & Partridge, 2000). Since drift is stronger on small populations, selection is more efficient in large populations. Thus, the probability of fixation of an allele depends not only on the strength of its selective advantage but also on the effective size of the population.

The interplay of evolutionary forces in natural populations can be complex depending on which evolutionary processes are at work. Adaptation to environmental conditions through natural selection is well documented for large effect loci in outcrossing species with large effective population sizes, e.g. *Drosophila* (Adrion, Hahn, & Cooper, 2015), humans (Fan, Hansen, Lo, & Tishkoff, 2016), sticklebacks (Shapiro et al., 2004) and trees (Savolainen, Pyhäjärvi, & Knürr, 2007). Yet, we know much less about how adaptation proceeds for more complex traits and in species with less idealized modes of reproduction, i.e. selfing or clonal reproduction, leading to different demographic effects. Moreover, in endangered species with strongly fragmented populations, genetic drift should quickly eradicate natural variation and thus the basis for genetic adaptation. However, not much is known to which degree this assumption reflects reality. Many plant species are strongly structured and selfing, yet local adaptation is often apparent (Ågren & Schemske, 2012; Böndel et al., 2015; Fournier-Level et al., 2011; Huang, Yan, Lascoux, & Ge, 2012) Understanding the processes by which selfing species adapt can help determine efficient strategies to maintain or restore degraded habitats.

Here, I investigate the population dynamics of natural variation in two selfing species displaying low levels of diversity at the local level. I address the following questions: i) how do complex traits evolve when drift is maximized and recombination impossible? ii) how can such populations be maintained when they are severely endangered?

1.2 EVOLUTIONARY PROCESSES IN NATURAL POPULATIONS

1.2.1 LOCAL ADAPTATION

If populations of the same species experience different environmental conditions, they can adapt to local conditions through natural selection. If local adaptation has occurred local genotypes outcompete non-local genotypes in a given environment (Kawecki & Ebert, 2004). Local adaptation is believed to be especially important in plants due to their sessile nature. However, in a meta-analysis of local adaptation in plants, it could only be detected in 45% of experiments (Leimu & Fischer, 2008). Thus, local adaptation might be less common than expected. The efficiency of local adaptation depends on several factors. The strength of the environmental gradient not only determines the selective pressure, but also contributes to genetic isolation among populations, both of which facilitate local adaptation. Strong environmental gradients can lead to local adaptation on very small scales. For example, in *Ranunculus reptans*, variation in natural flooding regime among lakeside microhabitats, lead to increased flooding resistance in plants experiencing more frequent flooding in their microhabitats, as demonstrated in a common garden experiment (Lenssen, Kleunen, Fischer, & Kroon, 2004). Nowadays, many plant populations are relatively small and isolated due to habitat fragmentation and degradation. Since, the effective population size determines the efficiency of selection, the potential for local adaptation might be diminished. Moreover, effective population size is often correlated with genetic diversity as strong genetic drift eliminates genetic variants. Genetic diversity is an important factor for local adaptation, as populations can adapt more quickly from standing variation than new mutations (Barrett & Schluter, 2008). This is facilitated by the fact that allelic fitness effects are dependent on the environment, i.e. an allele can be neutral in one environment, but beneficial in a different environment (Paaby & Rockman, 2014). Indeed, meta-analyses have shown that the frequency of local adaptation was positively correlated with the magnitude of environmental difference and population size (Hereford, 2009; Leimu & Fischer, 2008). Finally, the genetic

isolation of populations also influences the efficiency of local adaptation. If migration among populations is high, populations might get swamped by non-local alleles and adaptive alleles may never increase in frequency or fix. Further, if populations are not isolated, adaptation to one environment must have a cost in the other (Kawecki & Ebert, 2004). Otherwise, adaptive alleles will spread to all populations. However, low levels of migration can be beneficial due to the introduction of new genetic variation and the increase of the effective population size.

Selfing plants tend to have lower genetic diversity, smaller effective population size, lower recombination rates and stronger population structure than outcrossing plants (Hamrick J. L. & Godt M. J. W, 1996; M. Nordborg, 2000). These factors can influence a species ability to adapt to local conditions. On one hand, stronger isolation among populations should favor local adaptation (Linhart & Grant, 1996). Moreover, beneficial recessive alleles are exposed to selection even at low frequency (Glémin & Ronfort, 2013). On the other hand, the reduced effective population size in selfing species reduces the efficiency of selection. This is magnified by the fact that selfing populations undergo bottlenecks more frequently, because populations can be founded by single individuals (Schoen & Brown, 1991). Lower genetic diversity (standing variation) also means that local adaptation might be slower because it is limited by the occurrence of new mutations (Barrett & Schluter, 2008). Finally, reduced recombination rates lead to linkage of beneficial and deleterious alleles and inhibit the combination of independent beneficial mutations, thus also inhibiting adaptation. In meta-analyses comparing the frequency of local adaptation between selfing and outcrossing species, no significant influence of mating system/outcrossing rate was found, suggesting that benefits and drawbacks cancel each other out (Hereford, 2010; Leimu & Fischer, 2008). However, it is so far unclear if the mating system influences how local adaptation proceeds. Theoretical models predict that adaptation is more efficient in outcrossing species when mutations are dominant or codominant and when standing variation plays a significant role. Yet, adaptation, if it occurs, is expected to be more rapid in selfing species (Glémin & Ronfort, 2013).

1.2.1.1 HOW CAN WE DETECT LOCAL ADAPTATION?

There are several methods to detect local adaptation. Reciprocal transplant experiments are a direct way to test for local adaptation, as fitness advantage of native plants over non-native

is measured in common garden experiments, carried out at the native site of each tested population (Savolainen, Lascoux, & Merilä, 2013). Using this method, local adaptation was detected between populations of *Arabidopsis thaliana* from Sweden and from Italy (Ågren & Schemske, 2012). There are also several drawbacks of reciprocal transplants. Experiments are very work and time consuming especially if done for multiple populations or with organisms with a long generation time. Additionally, some fitness traits like male fitness or natural germination are difficult to assess in natural populations. Finally, selection strength might vary over time so that experiments should be repeated over multiple years, making them even more time-consuming (Savolainen et al., 2013).

Besides reciprocal transplant experiments there are indirect methods for testing local adaptation, which are less time intensive and feasible for a wider range of organisms. One method uses clinal variation along environmental gradients. Here, correlations between alleles or phenotypes, measured in common garden experiments, and environmental variables are tested (Savolainen et al., 2013). However, such correlations can also originate from neutral demographic processes, i.e. population structure, which will lead to false positive detection. Thus, it is important to take population structure into account (Coop, Witonsky, Rienzo, & Pritchard, 2010). Yet, correcting for population structure can easily lead to false negatives, if local adaptation contributed to shaping population structure. In *A. thaliana*, flowering time measured in common garden experiments has been shown to vary with latitude (Lempe et al., 2005; Stinchcombe et al., 2004). Moreover, patterns of co-variation between flowering time, seed dormancy and vegetative growth were also associated with latitude (Debieu et al., 2013). Climate-adaptive loci were identified by genotype-climate correlations, allowing fitness predictions of accessions grown in a common garden (Hancock et al., 2011).

Signatures of local adaptation can also be detected indirectly by analyzing differentiation between populations. Loci regulating local adaptation can be identified by F_{ST} outlier scans to detect regions with unusually strong differentiation (Savolainen et al., 2013). However, processes unrelated to local adaptation can also create strong F_{ST} outliers (Bierne, Roze, & Welch, 2013). Thus, false positive rates in F_{ST} scans can be very high (Fourcade, Chaput-Bardy, Secondi, Fleurant, & Lemaire, 2013) and results should be validated with additional tests or interpreted with caution. To detect locally adapted phenotypes, phenotypic differentiation

Q_{ST} can be estimated from genetic variation within and among populations. For neutral quantitative traits with an additive genetic basis, genetic drift acts similarly on phenotypes as on genotypes. Thus, the Q_{ST} distribution of neutral phenotypes is similar to the neutral F_{ST} distribution. Therefore, the distribution of F_{ST} estimates from neutral markers can serve as a null expectation for Q_{ST} (P. H. Leinonen, Remington, Leppälä, & Savolainen, 2013). For locally adapted phenotypes, Q_{ST} is expected to be in the tail of the F_{ST} distribution (Whitlock & Guillaume, 2009). Estimation of Q_{ST} can be difficult and often requires breeding experiments and common gardens to estimate the genetic variation. However, in selfing plants like *A. thaliana*, genotype replicates grown in common gardens can be used to achieve this relatively easily (Kronholm, Picó, Alonso-Blanco, Goudet, & de Meaux, 2012).

To identify the genetic basis of local adaptation, the methods outlined above can be combined with mapping approaches like quantitative trait locus (QTL) mapping or genome-wide association studies. For example, between Swedish and Italian *A. thaliana* populations, 15 quantitative trait loci (QTL) determining fitness in local environments were identified and a third of these showed evidence of trade-offs between populations (Ågren, Oakley, McKay, Lovell, & Schemske, 2013). So far, work on the genetic basis of local adaptation in plants has mostly focused on traits with major effect loci, e.g. flowering time (Hall & Willis, 2006; Méndez-Vigo, Picó, Ramiro, Martínez-Zapater, & Alonso-Blanco, 2011; Stinchcombe et al., 2004), seed dormancy (Kronholm et al., 2012; Postma, Lundemo, & Ågren, 2016) and nickel tolerance (Bratteler, Lexer, & Widmer, 2006). However, other traits with a more complex genetics basis likely also play an important role in local adaptation.

1.2.1.2 HOW CAN VARIATION IN STOMATAL TRAITS CONTRIBUTE TO LOCAL ADAPTATION?

In plants water-loss by transpiration, and gas-exchange for photosynthesis are tightly linked by a trade-off between growth and water conservation (Cowan, 1986; Cowan & Farquhar, 1977; Field, Merino, & Mooney, 1983). Stomata, the microscopic pores embedded in the epidermis of plant leaves, play a key role in the resolution of this trade-off. Their density, distribution and regulation control the rate of CO_2 and water exchange (Raven, 2002). As a result, they impact the ratio of photosynthetic carbon assimilation to water loss *via* transpiration. This ratio defines water-use efficiency (WUE), a physiological parameter that

directly determines plant productivity when the water supply is limited. The density of stomata on the leaf surface is expected to correlate positively with the rate of gas exchanges between the leaf and the atmosphere, also called “conductance”. Models based on gas diffusion theory predict that small stomata in high density can best maximize conductance (Franks & Beerling, 2009). A positive relationship between stomata density and conductance has been reported in a majority of studies looking at natural variation between species (Anderson & Briske, 1990; Pearce, Millard, Bray, & Rood, 2006) as well as within species (J. E. Carlson, Adams, & Holsinger, 2016; Muchow & Sinclair, 1989; Reich, 1984). Yet, higher stomata density does not always translate into higher rates of gas exchanges: in a diversity panel of rice (Ohsumi, Kanemura, Homma, Horie, & Shiraiwa, 2007) or within several vegetable crop species (Bakker, 1991), for example, the relationship was not observed. Nevertheless, molecular mutants in genes promoting stomata development show that reduced stomata density translates into decreased water loss and increased ability to survive after exposure to drought (Franks, W. Doheny-Adams, Britton-Harper, & Gray, 2015; Yoo et al., 2010).

Stomata density is not the only parameter modulating the balance between water loss and carbon uptake. Variation in stomata size also impacts the efficiency of stomata regulation (Raven, 2014). Stomata open and close in response to environmental and internal signals (Chater et al., 2011; Kinoshita et al., 2011). This ensures that plants do not desiccate when water evaporation is maximal and spares water when photosynthesis is not active (Daszkowska-Golec & Szarejko, 2013). The speed of stomata closure is higher in smaller stomata (Drake, Froend, & Franks, 2013; Raven, 2014). Stomatal responses are an order of magnitude slower than photosynthetic changes, so any increase in closure time lag may result in unnecessary water loss and reduce WUE (T. Lawson, Kramer, & Raines, 2012; Raven, 2014). However, it is often observed that decreases in stomata size occur at the expense of increased stomata density (reviewed in Hetherington & Woodward, 2003). This leads to a correlation that may at first be counter-intuitive: an increase in stomata density can result in improved WUE because of indirect effects on stomata size. The close link between stomatal traits and WUE suggest that stomatal traits may be an important factor in local adaptation to dry habitats.

Indeed, several examples of variation of stomatal traits along environmental clines, suggesting local adaptation, are known. In the *Mimulus guttatus* species complex, accessions from drier inland populations showed decreased stomatal density and increased WUE, compared to accessions collected in humid coastal populations (Wu, Lowry, Nutter, & Willis, 2010). By contrast, in 19 *Protea repens* populations measured in a common garden experiment, stomata density increased with decreasing summer rainfall at the source location (J. E. Carlson et al., 2016). In *Eucalyptus globulus*, plants from the drier sites had smaller stomata and higher WUE but no concomitant change in stomata density (Franks, Drake, & Beerling, 2009). This suggested that the developmental effect correlating stomata size and density may sometimes be alleviated. Altogether, these studies highlight interconnections between stomata size, stomata density and WUE that change across species or populations. How and whether variation in these traits and their connections support or constrain adaptive processes, however, is not clearly established.

The above-mentioned studies suggest a link between stomatal traits and environmental variation. However, they omit to take genetic population structure into account and don't compare genetic and phenotypic differentiation to provide further indications for local adaptation. Moreover, they do not attempt to link genotypic and phenotypic variation. While these analyses are crucial to further our understanding of the role of stomatal traits for local adaptation, they are difficult to perform in many organisms. However, the model plant is *A. thaliana* is a well-suited system to tackle these problems. In this species, genome-wide patterns of nucleotide variation can be contrasted to phenotypic variation and both the genetic architecture and the adaptive history of the traits can be reconstructed (Atwell et al., 2010; Fournier-Level et al., 2011; Alonso-Blanco et al., 2016). Environmental variation has a documented impact on local adaptation in this species (Debieu et al., 2013; Hamilton, Okada, Korves, & Schmitt, 2015; Hancock et al., 2011; Kronholm et al., 2012; Lasky et al., 2014; Postma & Ågren, 2016). In addition, natural variation in stomatal patterning is known to segregate among *A. thaliana* accessions (Delgado, Alonso-Blanco, Fenoll, & Mena, 2011). This species thus provides the ideal evolutionary context in which the adaptive contribution of variation in stomata patterning can be dissected.

1.2.2 HABITAT DEGRADATION AND ECOLOGICAL RESTORATION

High genetic diversity is considered an important feature of healthy populations as it increases potential of populations to adapt to changing environments (Reed & Frankham, 2003; Vrijenhoek, 1994) and has generally positive effects on ecosystems (A. R. Hughes, Inouye, Johnson, Underwood, & Vellend, 2008). For example, experimentally increasing the genetic diversity of *Solidago altissima* increased primary above-ground biomass productivity and arthropod diversity (Crutsinger et al., 2006). Genetic diversity also boosts resistance of populations to invasion and environmental fluctuations, presumably because it enhances the adaptive potential of populations (Reed & Frankham, 2003; Vrijenhoek, 1994). For example, high diversity experimental *Arabidopsis thaliana* populations showed higher resistance against invasion by *Senecio vulgaris* than low diversity populations (Scheepens, Rauschkolb, Ziegler, Schroth, & Bossdorf, 2017). Moreover, *Zoestra marina* populations with higher genetic diversity showed increased biomass production, plant density and faunal abundance during an extremely warm period (Reusch, Ehlers, Hämmerli, & Worm, 2005). Concordantly, restored populations of *Z. marina* with increased genetic diversity showed longer plant survival, increased faster in density and provided better ecosystem services. This effect was stable in a range of environmental conditions along a water-depth gradient (Reynolds, McGlathery, & Waycott, 2012).

Almost all, but especially endangered species suffer from habitat fragmentation due to human activities (Haddad et al., 2015). Habitat fragmentation splits up populations and reduces effective population size, thus increasing genetic drift. Moreover, it reduces gene-flow among populations, which leads to a further decrease of genetic diversity (S. M. Carlson, Cunningham, & Westley, 2014; Lienert, 2004; Pavlova et al., 2017). Indeed, populations of endangered vertebrate species usually show lower genetic variation than non-endangered species (Willoughby et al., 2015). Additionally, increasing efforts are being made to restore degraded habitats, a process called ecological restoration. Depending on the methods used strong genetic drift or selection regimes might lead to a reduction of genetic diversity or shift population fitness away from the optimum. Thus, the preservation and recovery of genetic diversity is a major concern both in conservation and restoration biology.

Population genetics can contribute to our understanding of how genetic diversity is shaped in endangered species and (Allendorf, Hohenlohe, & Luikart, 2010; Supple & Shapiro, 2018) and ecological restoration (Mijangos, Pacioni, Spencer, & Craig, 2015; Williams, Nevill, & Krauss, 2014). Thus, the young research fields of conservation genetics and restoration genetics are becoming increasingly important.

1.2.2.1 HOW DOES HABITAT FRAGMENTATION AFFECT GENETIC DIVERSITY IN ENDANGERED SPECIES?

Population genetic parameters of genetic diversity within and among populations, including population isolation and inbreeding, form the basis of conservation genetic assessments for any species (Ottewell, Bickerton, Byrne, & Lowe, 2016). These parameters can be estimated from neutral markers which are straightforwardly obtainable even for non-model species (Allendorf, 2017). This allows the application of these analyses to a wide range of species. In *Lepidium subulatum*, analysis of microsatellite markers in 344 individuals from 20 sites revealed that genetic diversity was negatively correlated with the degree of isolation of populations, although the species was locally relatively abundant (Gómez-Fernández, Alcocer, & Matesanz, 2016). Similarly, in the strongly endangered grassland species *Dianthus seguieri ssp. glaber*, genetic diversity was higher in larger and more extended populations (Busch & Reisch, 2016). A genetic comparison of palm trees (*Oenocarpus bataua*) from pristine and recently fragmented (<2 generations ago) habitats revealed stronger genetic structure in fragmented sites than continuous forests. This effect was only observed in young trees (grown after fragmentation) but not in older trees, strengthening the link to habitat fragmentation. Additionally, no reduction in genetic diversity was observed, suggesting that effects on genetic diversity might take more time (Browne, Ottewell, & Karubian, 2015). Although the consequences of habitat fragmentation have been studied in many plant species, several groups are underrepresented in these studies, e.g. monocots, abiotically pollinated, clonal and self-compatible plants (Heinken & Weber, 2013).

1.2.2.2 CAN ECOLOGICAL RESTORATION MAINTAIN OR IMPROVE GENETIC DIVERSITY?

A major concern of restoration ecologists is the transfer genetic diversity from donor to restored populations and its subsequent maintenance. However, studies comparing the level of genetic diversity in pristine and restored populations frequently report limited success, with

a reduction of genetic diversity in restored populations. This decline in genetic diversity may be caused by genetic bottlenecks in plant nurseries, biases introduced by seed harvesting strategies, founder effects during recolonization and/or unreliable commercial seeds (Mijangos, Pacioni, Spencer, & Craig, 2015). By contrast, the transfer of seed-containing hay from pristine (donor) to restoration (donee) sites, termed hay-transfer, is expected to limit the loss of genetic diversity and maintain site-specific local adaption (Hufford & Mazer, 2003; Kiehl, Kirmer, & Shaw, 2014). In addition, this method has the unique feature that it can, theoretically, restore an entire community without altering the genetic composition of populations and thus is the best method available for restoring entire plant communities (Hölzel & Otte, 2003; Kiehl, Kirmer, Donath, Rasran, & Hölzel, 2010). So far, however, there is no empirical support for the efficiency of this practice (Bucharova et al., 2017), especially since many species maintain seed banks in the soil. Indeed, the genetic diversity specific to the seed bank will not be sampled with the hay, although it is known that it can contribute significantly to the maintenance of diversity (Tellier, Laurent, Lainer, Pavlidis, & Stephan, 2011).

1.2.2.3 HOW RELIABLE ARE GENOTYPING-BY-SEQUENCING METHODS FOR CONSERVATION AND RESTORATION GENETICS?

The fields of conservation and restoration genetics have witnessed a major technological shift over recent years. Studies have initially relied on allozymes, microsatellites and AFLP markers and thus provided a limited overview on patterns of genetic variation within and between restored or pristine populations (Allendorf, 2017; Mijangos et al., 2015). Now, genotyping-by-sequencing (GBS) methods are beginning to be more broadly adopted (Gruenthal et al., 2014; Massatti, Doherty, & Wood, 2018; O'Leary, Hollenbeck, Vega, Gold, & Portnoy, 2018; Torres-Martinez & Emery, 2016). These methods drastically reduce sequencing costs through strategies to sequence a reduced portion of the genome, e.g. Restriction site Associated DNA-sequencing (RAD-seq) (Elshire et al., 2011; Etter, Bassham, Hohenlohe, Johnson, & Cresko, 2011; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). In contrast to previous methods (AFLP, microsatellites), GBS approaches sample proportions of the genome that are sufficiently large to allow resolving patterns of genetic diversity and spatial structure even at very local scale where overall levels of genetic diversity are low (Bradbury et al., 2015; Jeffries et al., 2016; Reitzel, Herrera, Layden, Martindale, & Shank, 2013). In principle, GBS approaches have a third major advantage: they are well suited to unravel genetic diversity in non-model species

without prior genomic information. Yet, the accuracy of genotyping in the absence of a reliable reference genome has been questioned (Shafer et al., 2016). Since target species in conservation and restoration projects rarely coincide with species or genera with advanced prior genomic knowledge, it is important to assess whenever possible, whether conclusions from RAD-seq-based restoration genetics studies depend on the availability of a reference genome.

1.2.3 INTERSPECIFIC HYBRIDIZATION

Reproductively isolated populations diverge over time, both genetically and phenotypically through the evolutionary forces of genetic drift and natural selection. This process builds up reproductive barriers between individuals from these populations, eventually resulting in speciation (Loren H. Rieseberg & Willis, 2007). Since speciation is usually not instantaneous, individuals with incomplete reproductive barriers often come into secondary contact and hybridize (Abbott et al., 2013). In plants, hybridization is common but not universal, occurring in 40% of families and 16% of genera (Whitney, Ahern, Campbell, Albert, & King, 2010). Hybridization is an important evolutionary process with a variety of potential consequences for the hybrids as well as the parental taxa.

Hybrid individuals sometimes grow faster and to larger size than their parental taxa, an effect called hybrid vigor or heterosis (Goulet, Roda, & Hopkins, 2017). Hybrid vigor is usually observed in the F1 generation and passes away in following generations. It is commonly used in plant breeding to increase crop performance and yield (Tsaftaris & Kafka, 1997). Hybrid vigor is caused by a combination of genetic effects, dominance, overdominance and epistasis as well as epigenetic effects (Groszmann, Greaves, Fujimoto, James Peacock, & Dennis, 2013; Shang et al., 2016; Shen, Zhan, Chen, & Xing, 2014).

Transgressive segregation is another potential consequence of hybridization that like heterosis produces phenotypes outside of the parental range. However, in contrast to heterosis, it usually manifests in the F2 generation or later and is persistent afterwards (Goulet et al., 2017). Over 97% of studies reporting parent hybrid trait values in plants included at least one transgressive trait, suggesting transgressive segregation is common in plants (L. H. Rieseberg, Archer, & Wayne, 1999). Transgressive segregation is a key mechanism to create novel phenotypic variation through hybridization and allow more diverse ecological adaptation. This may be especially important in endangered species with small effective populations sizes and strong adaptive constraints.

Frequent hybridization can break down reproductive barriers and reverse the process of speciation (Taylor et al., 2006). In contrast, if hybrid fitness is reduced compared to parental taxa, e.g. due to maladaptation or genetic incompatibilities, selection against hybridization can increase reproductive isolation. This effect is called reinforcement (Hopkins, 2013).

Selection only favors reproductive isolation before costly hybridization, thus mostly pre-zygotic isolation is facilitated. In plants, reductions in hybridization rates can be achieved e.g. by changes in flowering time (Silvertown, Servaes, Biss, & Macleod, 2005), altered flower color (Hopkins & Rausher, 2012) and morphology (Whalen, 1978), increased self-fertilization (Fishman & Wyatt, 1999) and pollen-stigma incompatibilities (Kay & Schemske, 2008).

Hybridization can also lead to the formation of a third species, which is distinct from parental taxa. In plants, this often happens instantaneous through allopolyploidization, which strongly inhibits gene-flow between hybrids and parental taxa (Abbott et al., 2013; Mallet, 2007). Cases of homoploid speciation, i.e. without a change in ploidy, are rarer, take more time and require strong ecological differentiation among taxa to ensure reproductive isolation. However, transgressive segregation may allow hybrids to colonize new ecological niches, facilitating this process (Mallet, 2007). In summary, hybridization can not only slow down or reverse speciation but also but also facilitate speciation and even result in completely new species.

Fertile, homoploid hybrids are often not reproductively isolated from their parental taxa. Thus, they can backcross to one or both parents. Recurrent backcrossing can lead to gene-flow among parental taxa, resulting in introgressions, i.e. small genomic fragments from one fragment are inserted into the genome of the other parent (Suarez-Gonzalez, Lexer, & Cronk, 2018). In populations and species experiencing gene-flow, introgressions are a major source for genetic variation, besides *de novo* mutation and standing variation, increasing the potential for adaptation (Tigano & Friesen, 2016). Thus, interspecific gene-flow may be especially important in species with low intraspecific genetic variation and small effective population size as a means of recovering genetic diversity. Yet, while many examples of hybridization between endangered species and their relatives are known (Walter Bleeker, Schmitz, & Ristow, 2007), detailed descriptions of introgression patterns are missing in most cases.

1.2.3.1 HOW CAN INTROGRESSED GENOMIC FRAGMENTS BE DETECTED?

Introgressed fragments are expected to show less divergence than expected based on the overall phylogenetic distance of the two taxa, resulting in a locally discordant phylogenetic tree. However, detection of introgressions is aggravated by the fact that ancestral polymorphism persisting after the divergence of the two species, i.e. incomplete lineage

sorting, can also result in locally discordant phylogenetic trees (Goulet et al., 2017). Yet, it is possible to distinguish between these scenarios, using the so-called ABBA-BABA-test or D-statistic. The test compares allele counts of ancestral (A) and derived alleles (B) in a four-taxon phylogeny with the overall allele pattern AABA and BBAA. The D-statistic is the relative difference between the frequency of two allele patterns, ABBA and BABA, incongruent with the overall phylogeny. In a scenario of incomplete lineage sorting, ABBA and BABA should occur with the same frequency. Thus, a significant excess of one of the two patterns (D-statistic unequal to 0) is indicative of introgression (Green et al., 2010). While the test has been originally developed for single sequences, it was extended for usage with populations based on allele frequencies (Durand, Patterson, Reich, & Slatkin, 2011). Additionally, a derived statistic, f_D, can be used to locate candidate regions of introgression along the genome (S. H. Martin, Davey, & Jiggins, 2015). This statistic estimates the magnitude/frequency of introgression in genomic regions, by comparing observed frequencies of ABBA and BABA with expected frequencies under a complete introgression scenario. These methods in combination with advances in next-generation sequencing have made the analysis of introgression patterns in non-model species feasible.

1.2.3.2 CAN INTROGRESSION PLAY A ROLE IN ADAPTATION?

Adaptive introgressions, i.e. introgressions carrying alleles conferring a fitness advantage, are expected rise to high frequencies. The $q95_{(1,99)}$ statistic can be a first indicator of adaptive introgression (Racimo, Marnetto, & Huerta-Sánchez, 2017). This statistic is the 95% quantile of the distribution of alleles, which are close to fixation between the donor species and populations without introgression, in the population with introgression. Thus, it can be used to detect high frequency introgressions, which might be a result of selection. However, to demonstrate adaptive introgression has indeed occurred, additional evidence, e.g. genomic signatures of selection, phenotypic or fitness effects of introgressed fragments, are required, but often difficult to obtain (Suarez-Gonzalez et al., 2018). Despite these experimental challenges, there are several well documented examples of adaptive introgression in plants (Suarez-Gonzalez et al., 2018). For example, the allotetraploid *Senecio vulgaris* has lost its large attractive petals (radiate phenotype) due to a shift to high self-fertilization. Some populations in the United Kingdom, however, have regained the radiate phenotype after the introduction of a diploid sister species *Senecio squalidus*. Comparative analysis of haplotypes

of the single locus conferring the radiate phenotype among species has revealed that *S. vulgaris* has regained the phenotype through introgression from *S. squalidus,* via backcrossing of triploid hybrids (Kim et al., 2008). This demonstrates that introgression allows species to regain lost traits, which may be an evolutionary advantage. In this example, reversal of the selfing syndrome may lead to higher outcrossing rates, which allow for more efficient recombination. In sunflowers, gene-flow between *Helianthus annuus annuus* and its more herbivore resistant relative *Helianthus debilis* resulted in a hybrid subspecies, *Helianthus annuus texanus*, which showed increased herbivore resistance and seed production than *H. annuus annuus* in two common garden experiments (Whitney, Randell, & Rieseberg, 2006). In additional common garden experiments, phenotypes of *H. annuus texanus* for water-use efficiency, specific leaf area, seed maturation time, disk diameter, height of lowest branch, and relative branch diameter were all significantly different from *H. annuus annuus* and shifted towards *H. debilis.* These changes were again associated with increased fitness of introgressed lineages (Whitney, Randell, & Rieseberg, 2010). In *Arabidopsis arenosa*, a genomic analysis of serpentine populations detected strong selective sweeps which overlapped with introgressions from *Arabidopsis lyrata*, suggesting that these introgressions confer adaptation to the multi-hazardous (drought, phytotoxic metals, nutrient poor) serpentine habitat (B. J. Arnold et al., 2016). These studies demonstrate that introgression can affect not only one but multiple adaptive traits in the same species. Thus adaptive introgression has been suggested as a means to cope with changing environmental conditions (Hamilton & Miller, 2016). Yet, further understanding of the consequences of hybridization in endangered species is required to adjust management strategies of natural hybridization.

1.3 STUDY SYSTEMS

To understand how complex traits evolve in selfing plants with strong drift and low recombination rate, I analyzed natural variation of stomatal traits and water-use efficiency in *Arabidopsis thaliana*. The annual plant *A. thaliana* is native across Eurasia and Africa and is now also established in America (Krämer, 2015). Its selfing rate is about 97% resulting in very high homozygosity. Yet, outcrossing rates are high enough to have resulted in considerable haplotypic diversity (Platt et al., 2010), which might facilitate local adaptation. Its genome is one of the most well studied genomes (Berardini et al., 2015) and, with the 1001 genomes collection (Alonso-Blanco et al., 2016), more than 1100 fully sequenced accessions are available for phenotypic analysis. Thus, it is an ideal model to study the genetic basis of natural variation and local adaptation. Studies of local adaptation in *A. thaliana* have so far focused on relatively simple traits. (Ågren et al., 2013; Debieu et al., 2013; Exposito-Alonso et al., 2018; Fournier-Level et al., 2011; Hancock et al., 2011; Kronholm et al., 2012; Lasky et al., 2014; Mojica et al., 2016; Postma & Ågren, 2016). Stomatal traits have the potential to play a role in local adaptation due to their crucial function of regulating the trade-off between carbon-uptake and water-loss (see above). Yet, while natural variation of stomatal traits has been detected in a set of 80 accessions (Delgado et al., 2011), its role in local adaptation and its genetic basis remain unknown.

To study how populations of endangered species are affected by habitat degradation and how they can be maintained, I used a less known relative of *A. thaliana*, *Arabis nemorensis* as a model. In contrast to *A. thaliana*, which is a human commensal (Lee et al., 2017), *A. nemorensis* has become greatly endangered due to human activity. *A. nemorensis* is a short-lived, mostly biennial hemicryptophyte that is known to maintain a long-lived soil seed bank (Hölzel & Otte, 2004). It is part of the *Arabis hirsuta* species aggregate, which comprises several morphologically similar but ecologically diverse species, which diverged about 1.2 million years ago (Karl & Koch, 2014). *A. nemorensis* is a stenoecious species typically restricted to floodplain meadows, as it requires regular flooding to release competition pressure. However, floodplain meadows were degraded in large areas of Europe due to agriculture and river regulation. Thus, *A. nemorensis* has become strongly endangered throughout Europe (Schnittler & Günther, 1999). The effects of this habitat destruction on the genetic diversity and population structure of *A. nemorensis* have not been studied so far. *A.*

nemorensis is also a target species in an ongoing ecological restoration project of floodplain meadows in the Upper Rhine valley. In this project, seed containing hay was transferred from pristine donor sites to target sites to establish new plant communities (Donath, Bissels, Hölzel, & Otte, 2007; Hölzel & Otte, 2003). This project offers the opportunity to evaluate the influence of the hay-transfer method on genetic diversity in several restored sites and to analyze how ecological restoration shapes patterns of genetic diversity among populations. While there are other endangered species in floodplain meadows, which are also restoration targets, *A. nemorensis* is most well-suited for genetic studies, because it is a member of the *Brassicaceae* family and thus a close relative of the model plants *A. thaliana* and *Arabis alpina*. This allowed me to make use of the excellent genomic resources available for these species (Berardini et al., 2015; Jiao et al., 2017).

During the conservation and restoration genetics analyses, I discovered an active zone of hybridization between *A. nemorensis* and *Arabis sagittata*. *A. sagittata* is a sister species of A. nemorensis with a very similar morphology. However, it prefers drier habitats such as calcareous grasslands (Hand & Gregor, 2006) and its distribution is shifted towards Southern Europe compared to *A. nemorensis* (Jalas & Suominen, 1994). While it is known that the species can hybridize and sporadic hybrids have been discovered earlier (Novotná & Czapik, 1974; Titz, 1979), a hybrid zone of this extent was not previously described. Thus, I used genomic and phenotypic analyses to study the evolutionary potential of this hybridization, which might be critical for the survival of the species.

1.4 THESIS AIMS

1.4.1 CHAPTER 1 – THE ROLE OF NATURAL VARIATION OF STOMATAL TRAITS AND WATER-USE EFFICIENCY IN LOCAL ADAPTATION IN *ARABIDOPSIS THALIANA*

I investigated the role of stomata traits and water-use efficiency for local adaptation in *A. thaliana*. Using high-throughput methods, I characterized natural variation in stomatal traits and water-use efficiency in over 300 European accessions. This dataset allowed me to ask i) how variable are natural *A. thaliana* accessions in stomata patterning? ii) does variation in stomata patterning influence the carbon-water trade-off? iii) what is the genetic architecture of traits describing stomata patterning? iv) is stomata patterning optimized by natural selection?

1.4.2 CHAPTER 2 – THE EFFECT OF ECOLOGICAL RESTORATION ON THE GENETIC DIVERSITY OF TWO FLOODPLAIN ARABIS SPECIES

I analyzed how ecological restoration by hay-transfer shapes natural variation within and among pristine and restored populations of two *Arabis* species. Using RAD-seq, I genotyped over 130 individuals from ten sites, resulting in thousands of SNP markers for each species. I used this dataset to answer the following questions: i) what is the level of genetic diversity and structuration of the pristine sites that served as source populations for restoration? ii) do restored sites show a lower level of genetic diversity than the pristine sites? iii) how did restoration affect the distribution of diversity within and among restored sites? iv) is the use of a reference genome necessary to reliably characterize the impact of restoration on genetic diversity?

1.4.3 CHAPTER 3 – PATTERNS OF GENETIC DIVERSITY IN *A. NEMORENSIS*

I studied how habitat fragmentation and degradation have shaped the patterns of genetic diversity within and among populations of *A. nemorensis*, throughout Germany. I collected plants in 15 putative *A. nemorensis* populations. Using RAD-seq, I obtained genotype data for nearly 300 individuals. Genetic analysis that only nine populations were actually *A. nemorensis*. This dataset allowed me to ask the following questions: i) does *A. nemorensis* frequently co-occur with its sister species and hybridize? ii) what are the interspecific and

intraspecific level of genetic diversity iii) what is the level of genetic diversity within *A. nemorensis* populations? iv) how isolated are populations? v) which factors shape the population structure of *A. nemorensis*?

1.4.4 CHAPTER 4 – PHENOTYPIC DIVERGENCE BETWEEN *A. NEMORENSIS* AND *A. SAGITTATA*

I measured phenotypic variation within and among *A. nemorensis* and *A. sagittata* using two experimental approaches. This information is important to understand the evolutionary potential of hybridization between these species. I aimed to answer the following questions: i) do the two species differ in morphology, phenology and biotic stress resistance in semi-natural growth conditions? ii) is *A. nemorensis*, the floodplain native, more flooding resistant than *A. sagittata*, the floodplain invader?

1.4.5 CHAPTER 5 – PATTERNS OF GENE-FLOW BETWEEN *A. NEMORENSIS* AND *A. SAGITTATA*

In Chapter 5, I analyzed patterns of introgression between *A. nemorensis* and *A. sagittata*. I obtained whole-genome sequencing data for 34 accessions collected within and outside of the hybrid zone, as well as the outgroup species *Arabis androsacea*. Using this dataset, I asked the following questions: i) is there gene-flow between species i) can I detect regions of introgression in the genome? ii) is gene-flow bi-directional and in equal proportions? iv) are introgressions more frequent in the hybrid zone, i.e. are they recent or ancestral? v) what are the source populations of introgressions?

2 METHODS

2.1 CHAPTER 1 – THE ROLE OF NATURAL VARIATION OF STOMATAL TRAITS AND WATER-USE EFFICIENCY IN LOCAL ADAPTATION IN *ARABIDOPSIS THALIANA*

2.1.1 PLANT MATERIAL, PLANT GENOTYPES AND GROWTH CONDITIONS

I performed the experiments and phenotypic measurements for my master thesis at the University of Münster. I did most of the statistical analysis and wrote the publication of the results during my PhD.

In total, 330 accessions, spanning a wide geographical range were selected from the 1001 collection of fully sequenced genotypes (Table S 1). Accessions were assigned to five groups based on their geographic origin and genetic clustering (Alonso-Blanco et al., 2016): Spain, Western Europe, Central Europe, Southern Sweden and Northern Sweden (Figure 1). In 20 cases, for which genetic information contradicted geographic information, I prioritized geographic information since I am focusing on local adaptation and expect that geography is more relevant for local adaptation than demographic history. To avoid oversampling of some genotypes for the analysis of heritability, regional differentiation (Q_{ST}-F_{ST}) and climatic correlation, I randomly sampled one plant if multiple plants were collected at the same location, resulting in 287 accessions.

I downloaded the genome sequences of the 330 genotypes included in the analysis from the 1001 genome database (Alonso-Blanco et al., 2016) on May 12th, 2017. I extracted single nucleotide polymorphism (SNP) data using *vcftools* (Danecek et al., 2011). I randomly thinned genomic data to 1 SNP per 1000bp to reduce the computational load. In *A. thaliana*, linkage disequilibrium extends beyond 1kb (Magnus Nordborg et al., 2002). Thus, this data-size reduction should not impact statistics describing the geographical structure of genomic variation. Additionally, I set minimum minor allele frequency to 5% and removed sites exceeding 5% missing data, resulting in 70,410 SNPs among all genotypes. I loaded SNP information into R using the *vcfR* package (Knaus, Grunwald, Anderson, Winter, & Kamvar, 2017). For genome-wide association studies, I used the full, unthinned SNP dataset. Missing SNPs were imputed using BEAGLE version 3.0 (B. L. Browning & Browning, 2009).

I stratified seeds on wet paper for 6 days at 4°C in darkness. I grew plants on soil in 5x5 cm paper pots in 3 replicates with one plant per pot. I randomized genotypes within each of 3 blocks of 12 trays containing 8x4 pots. Plants grew for 7 weeks in growth chambers (one per block) under the following conditions: 16 h light; 95 µmol s^{-1} mm^{-2} light intensity; 20 °C day- and 18 °C night-temperature. Plants were watered twice a week and trays shuffled and rotated every two to three days to account for variable conditions within the chambers.

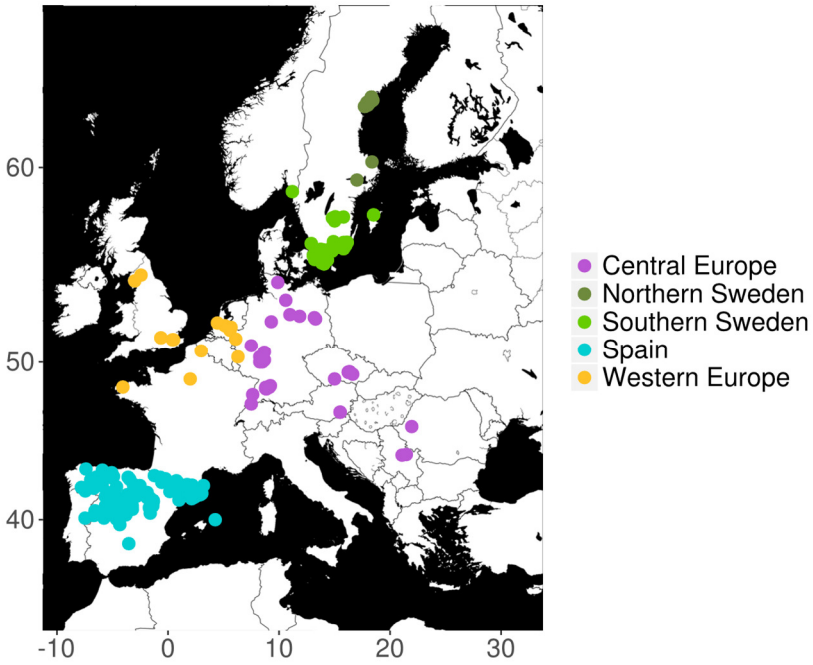

Figure 1: **Map of all accessions used in the study**
Each point represents one accession. Accessions are colored by assigned region of origin.

2.1.2 HIGH THROUGHPUT MICROSCOPY

After 7 weeks, one fully-expanded intact adult leaf (one of the largest leaves that developed after leaf 4) was selected on each plant for microscopic analysis. From each leaf, two discs were cut mid-way along the length of the leaf on both sides of the main vein, using a 6mm hole punch. If the leaf was too small to cut two discs, an additional leaf was collected. The discs were loaded onto an array of 80 spring mounted stamps with the abaxial side up and fixed on a thin layer of dental adhesive cream (blend-a-dent Super-Haftcreme). The leaf discs were stained using 25µl of a 100µg/ml propidium iodide solution for specific staining of stomata and cell walls (Fitzgibbon et al., 2013). A Zellkontakt 96-well glass-bottom plate was then put on top of the stamp array and fixed using four screws. To infiltrate the leaf discs with

the stain, the plate was put under vacuum three times for one minute. Microscopic images were taken using the Opera High Content Screening System from Perkin Elmer. The following settings were used: excitation wavelength 561nm; laser power 11000µW; magnification 20x; camera filter 600/40 nm; dichro filter 568 nm; exposure time 200ms; binning 1. Images were taken in 15 fields (0.15 mm^2) per well/sample. For each field 11 images were taken along the z-axis with 3µm distance to acquire image stacks. In total, we acquired 341,000 microscopic images of abaxial leaf epidermises from 31,000 image fields.

2.1.3 IMAGE PROCESSING AND ANALYSIS

2.1.3.1 MAXIMUM PROJECTION OF STACKS

The first step of image processing was performed using Acapella, the image analysis software designed specifically for the Opera by Perkin Elmer. The aim was to project the image stacks acquired for each field into single 2D images using maximum projection. First, a sliding parabola transformation was applied to each image layer in to reduce background noise. Second, the maximum projection function was applied to each stack. Third, the resulting image was saved to Bitmap format. The output folder was named after the plate number and the date of capturing of the images. The filename contained the following information separated by "_": Plate number, plant ID, plate well coordinates and image field within well. The Plant-ID contained additional information separated by '-': genotype, tray number, tray position and leaf ID (for corresponding leaf size measures). The bitmap images were further processed in MATLAB.

2.1.3.2 PRE-PROCESSING

In MATLAB, as in most other programming languages, images are read as 2D matrices with each pixel represented as one value in the matrix. For grayscale images the value is an integer between 0 (black) and 255 (white). On these matrices different mathematical operations can be performed to transform or analyze the images.

Images were enhanced using a histogram expansion function. This function stretches the pixel intensity values over the whole grayscale range, thus increasing brightness and contrast. Next, images were divided in 3x3 fields and for each field entropy and thresholding effectiveness of Otsu's method (Otsu, 1975) were calculated. Based on these values the quality of each image

part was determined. Only if at least 7 of 9 parts of the imaged matched the criteria the image was further analyzed. The critical values were determined on sets of manually selected high and low-quality images. This step is crucial as it not only decreases computing time by eliminating of low-quality images before complex analysis, but also because automated feature detection works more reliably if the images are relatively uniform. High quality images were then analyzed to detect stomata.

2.1.3.3 STOMATA DETECTION

Stomata appeared in images as bright, small and mostly elliptic objects with a gap in their center. These morphological features were used for stomata detection. Thresholds for all parameters were determined using training datasets and manual curation and would likely have to be adjusted for use with a different dataset.

First noise was removed by applying a Gaussian filter. Then the image contrast was strongly increased using histogram expansion and eliminating the darkest 60% of all pixels. The image was then converted from grayscale to logical by setting all pixels with value of 255 to one and all others to zero to mask the brightest objects in the images. The result was an image of foreground objects (connected 1-pixels) and background (all 0-pixels). Because stomata were among the detected objects, but not exclusively, the following filters were applied to the initial detection image:

- Objects located closely together were merged using image dilation
- Objects smaller than 700px or larger than 2000px were removed
- If afterwards the number of objects was still higher than 60, the image was discarded
- Holes in objects were filled

Upon detection of very large objects e.g. trichomes the image was discarded as this often led to inaccurate stomata recognition. Objects were also filtered based on their eccentricity as stomata are usually ellipsoid. Stomata with eccentricity below 0.4 were removed instantly. Major axis length of the ellipse had to be shorter than 80px. Moreover, the area of the ellipse with the same second moment as the presumed stomata was calculated. Out of this area and the actual object area a ratio was calculated (“area ratio”) to determine how well the object fit into an elliptic shape. Furthermore, the intensity profile through the minor axis of the object

and two offset parallels was analyzed to detect the characteristic stoma gap. The gap was showing up in the profile by two large peaks separated by a low intensity minimum. Based on whether the gap was detected in each of the three lines a gap score was calculated for each object. Because stomata were not necessarily uniformly shaped (e.g. open and closed stomata look different) various combinations of thresholds on these criteria were allowed for true stomata. For example, if an object is too round for a typical stoma it was still considered if the gap is very prominent. The following combinations were allowed for stomata:

- Area ratio > 0.9 & eccentricity > 0.8
- Profile score > 3 & eccentricity > 0.4 & area ratio > 0.8
- Profile score > 0 & eccentricity > 0.7 & area ratio > 0.85
- Profile score > 2 & eccentricity > 0.6 & area ratio > 0.8

Stomata detected by the different combinations of filters were then joined. The minimum convex area spanned by stomata an additional quality criterion. If this area was smaller than 50% of the image size, the result was discarded. Finally, the median of all images for each sample was calculated. If the median was smaller than 33 the image was discarded, as such a low value was never observed in manual controls and likely caused by low quality images which passed pre-filtering.

To evaluate detection accuracy, I manually determined stomata density on a random set of 14 individuals and on a set of 32 independently-grown individuals. Automatic and manual measurements were highly correlated (Pearson correlation coefficient r^2=0.88, p<<0.01and r^2=0.81, p<<0.01, for the 14 and 32 individuals, Figure 2). The algorithm was conservative and tended to slightly under-estimate stomata numbers, resulting in a low false-positive rate. This ensured that stomata area was generally quantified on objects that corresponded to real stomata.

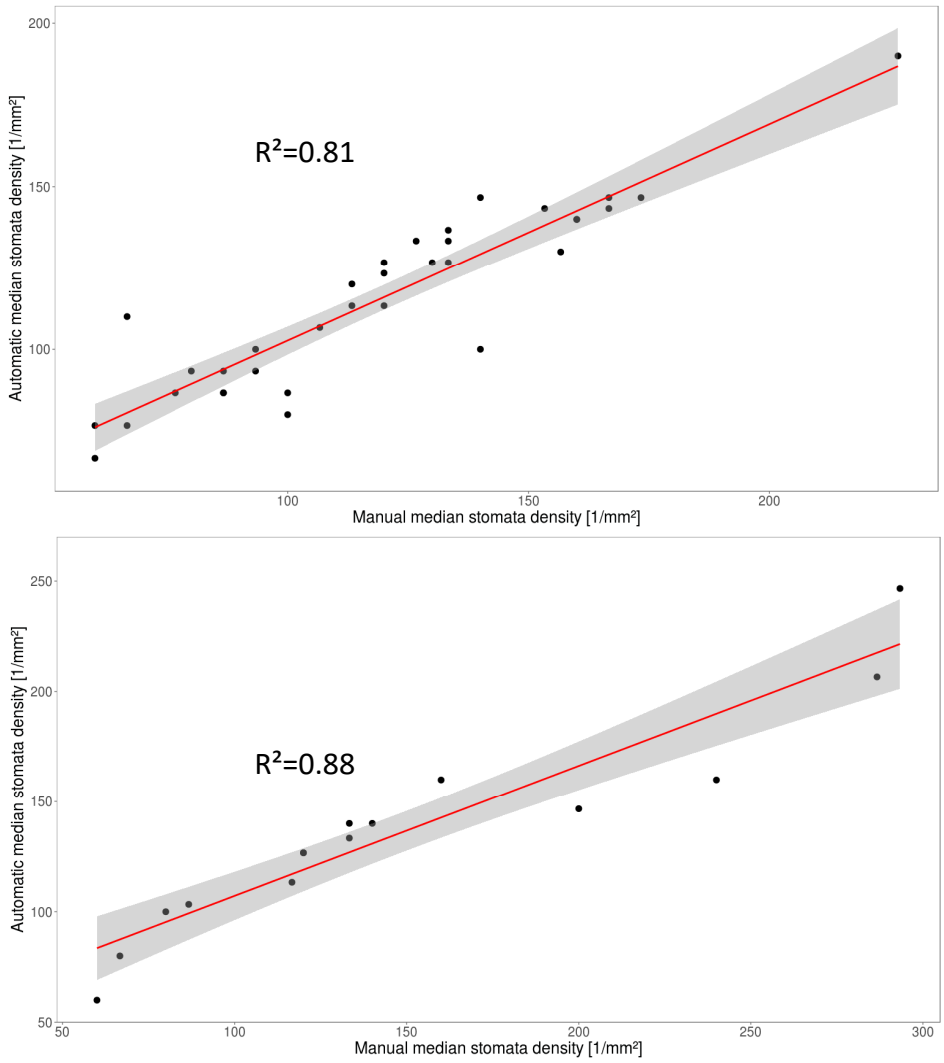

Figure 2: **Automatic and manual counts were highly correlated**
Correlation plot of manual and automatic stomata counts in a pre-experiment (top) and in a subset of accessions from the main experiment (bottom). Each point represents one leaf. The red line represents a linear model fit of the data and the gray shadows indicate the error of the fit. Correlation coefficients (R^2) are shown in each plot and both estimates are significant ($p<<0.001$).

Due to quality filters in the pipeline, the number of analyzed images differed between samples (Figure 3 left). I found a significant correlation between the number of images analyzed and stomata density ($r=0.21$, $p<<0.01$, Figure 3 right), but not stomata size ($r=0.02$, $p>0.05$). Thus, I included the number of images as a co-factor into all statistical models for stomata density (see below).

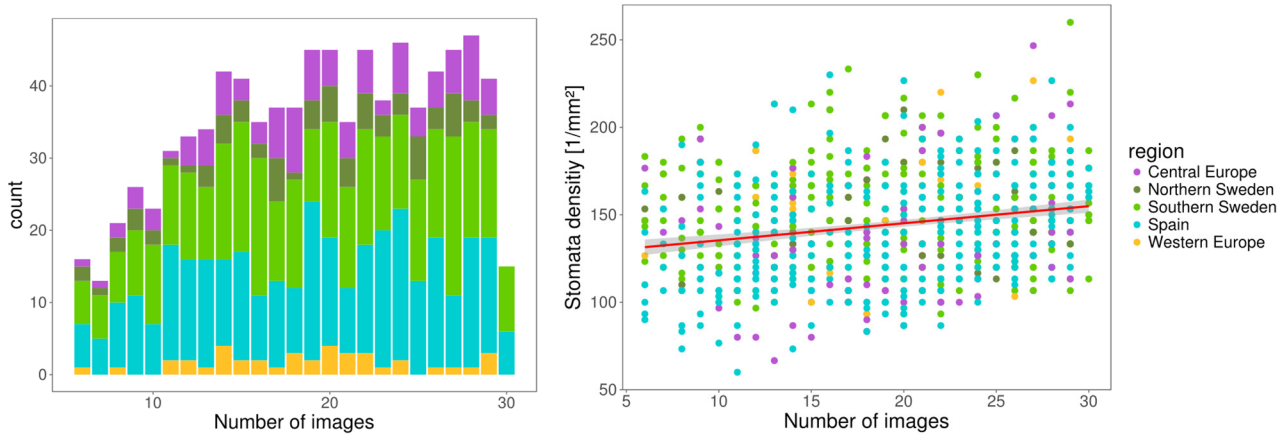

Figure 3: **Number of images differs among samples and is correlated with stomata density**
Left) Due to differences in image quality the number of images analyzed for stomata differs among samples. The plot shows the distribution of the number of images analyzed per sample. Right) Correlation plot of number of analyzed images and stomata density per sample. Points are colored by regions. The red line shows the linear fit and the grey shadows show the error of the fit. Pearson correlation coefficient was 0.21 ($p<0.01$). The region color scale on the right applies to both plots.

2.1.4 LEAF SIZE MEASUREMENTS

For leaf size measurements, each hole-punched leaf was fixed on a gridded A4 paper sheet (8 per page) using transparent tape. Possible gaps in the edge of leaves were closed using a pen. Paper sheets were then digitized using a common flatbed scanner.

The gridded A4 paper sheets with the fixed leaves were enumerated and scanned. Images were manually checked for closed leaf borders. This step was important as closed borders were necessary to fill the holes from disc cutting during image processing. The image analysis was also performed in MATLAB:

- Splitting of image into 8 leaf fields and the reference field (black 2cm² square) using relative coordinates
- For each sub-image:
    - Inversion of grayscale values
    - Conversion to logical image by intensity thresholding
    - Removal of small objects by area opening
    - Calculation of leaf area: number of white pixels in image/pixelarea in mm²

- Pixel-area was calculated using the reference field of known size

In case two leaves were used from one plant, these were detected by the algorithm as two objects and the mean was calculated.

2.1.5 CARBON ISOTOPE DISCRIMINATION MEASUREMENTS

Carbon isotope discrimination measurements ($\delta^{13}C$) of whole rosettes were performed by Andreas Weber and Tabea Mettler Altmann at the University of Düsseldorf.

The rosettes of block 1 were placed in individual paper bags after microscopic imaging was completed and dried at 70 °C for 3 weeks. Plant material was then ground to fine powder using a 25mm steel bead and a mixer mill (Retsch, MM 301). Isotope composition was determined using an ISOTOPE cube elemental analyzer coupled to an Isoprime 100 isotope ratio mass spectrometer (both from Elementar, Hanau, Germany) according to (Gowik, Bräutigam, Weber, Weber, & Westhoff, 2011). The carbon isotope ratio is expressed as ‰ against the Vienna Pee Dee Belemnite (VPDB) standard.

2.1.6 HERITABILITY ESTIMATES

I estimated broad-sense heritability H^2, the genetic proportion of the observed phenotypic variance, as:

$$H^2 = VarG/(VarG + VarE)$$

where VarG is the genetic variance and VarE is the environmental variance. Because I worked with inbred lines, VarE and VarG could be estimated as the variance between replicates of a genotype and the variance between genotypes, respectively, with a linear-mixed-model using block as fixed effect and genotype as random effect. I created the linear mixed model using the *lme* function from *nlme* package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2015) (Suppl. Document 2). For $\delta^{13}C$, no replicates were available but a pseudo-heritability estimate was extracted from the GWAS mixed model including the kinship matrix (Atwell et al., 2010).

2.1.7 GENOME-WIDE ASSOCIATION STUDY (GWAS)

I performed this analysis in collaboration with Arthur Korte from the University of Würzburg. For GWAS, SNPs with a minor allele count <5 were removed, leaving a dataset of 2.8-3M SNPs, depending on missing data for the phenotypes. Minor allele frequency spectra for all three datasets show that the subset of 261 genotypes, for which all three phenotypes were determined, has a lower proportion of rare SNPs (Figure 4). GWAS was performed using a mixed model correcting for population structure using a kinship matrix that was calculated under the assumption of the infinitesimal model. SNPs were first analyzed with a fast approximation (Kang et al., 2010) and the 1000 SNPs with the strongest association were reanalyzed with the complete model that estimates the respective variance components for each SNP separately (Kang et al., 2008).

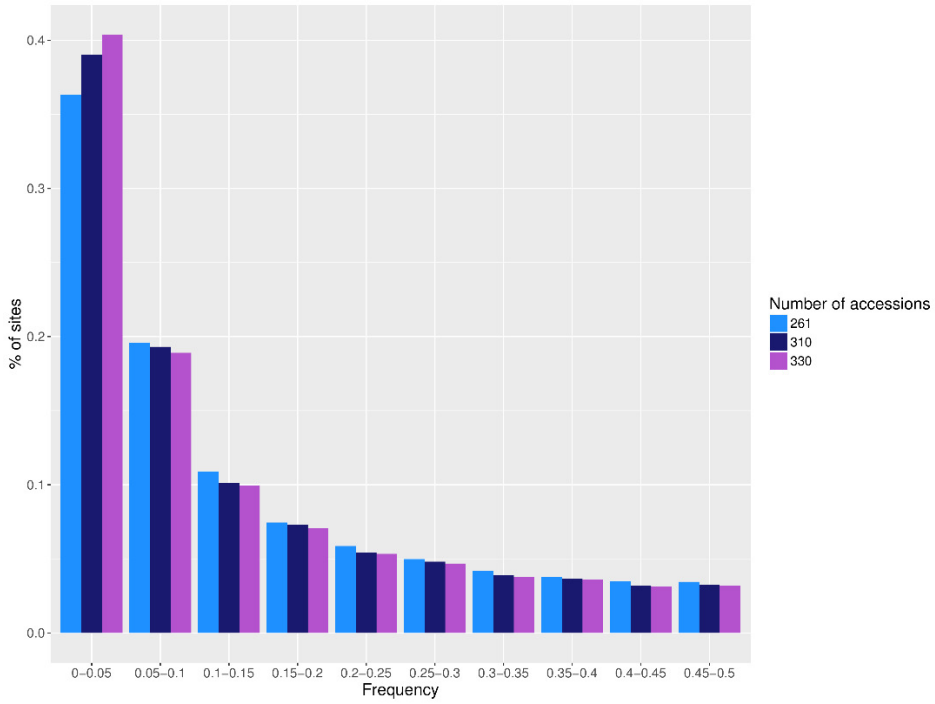

Figure 4: **Sites frequency spectra of GWAS datasets**
Each bar represents the number of sites with a specific allele frequency represented on the x-axis. Data is shown for all three GWAS datasets, with different numbers of accessions.

For trait pairs measured on the same plant, a Multi-Trait Mixed Model (MTMM) was applied to distinguish common and trait-specific SNP-phenotype association (Korte et al., 2012).

The MTMM performs three different statistical tests on a bivariate phenotype including each trait pair. The first model tests whether a given SNP has the same effect (direction and

magnitude) on both traits. This model has increased power to detect significant associations, which may fall under the significance threshold when traits are analyzed in isolation. The second model identifies SNPs having distinct effects (different direction) on the two traits. It is well suited to detect SNPs with antagonistic effects on both traits. The last model combines both trait-specific and common effects. This last model is particularly powerful for detecting markers affecting both traits with different intensity (same direction, different magnitude). The MTMM analysis also provides estimates of the genetic and environmental correlation for each pair of traits. The statistical details of the models are described in (Korte et al., 2012).

For all analyses (GWAS and MTMM), the significance threshold for QTL identification was determined as a 5 % Bonferroni threshold, i.e. 0.05 divided by the number of SNPs in the dataset.

2.1.8 CLIMATIC DATA

Climatic data included average precipitation, temperature, water vapor pressure (humidity), wind speed and solar radiation estimates with 2.5 min grid resolution (WorldClim2 database (Fick & Hijmans, 2017) on May 30th, 2017) and soil water content (Trabucco & Zomer, 2010). For each variable and accession, I extracted a mean over the putative growing season, i.e. the months in the year with average temperature greater than 5 °C and average soil water content over 25%.

Grey Monroe from the Colorado State University (Fort Collins, USA) computed historical drought frequencies at *A. thaliana* collection sites using 30+ years of the remotely-sensed Vegetative Health Index (VHI). The VHI is a drought detection method that combines the satellite measured Vegetative Health and Thermal Condition Indices to identify drought induced vegetative stress globally at weekly 4 km^2 resolution (Kogan, 1995). This is a validated method for detecting drought conditions in agriculture. Specifically, we used VHI records to calculate the historic frequency of observing drought conditions (VHI<40) during the spring (quarter surrounding spring equinox) and summer (quarter surrounding summer solstice). These are the typical reproductive seasons of *Arabidopsis* populations (reviewed in Burghardt, Metcalf, Wilczek, Schmitt, & Donohue, 2015). The drought regime in each location was

quantified as the log-transformed ratio of spring over summer drought frequency. Positive values of this drought regime measure reflect environments where the frequency of drought decreases over the typical reproductive growing season, and vice versa for negative values. This ratio quantifies the seasonality of water availability. It correlates with the ratio of soil water content of the first and third month of the reproductive season ($r=0.54$, $p<0.01$), which we defined as the first and third growing month in the year, giving similar estimates as Burghardt, Metcalf, Wilczek, Schmitt, & Donohue (2015).

Because the seven climate variables were correlated, I combined them in seven principal components (PCs) for 316 *A. thaliana* collection sites (Figure 5). I excluded fourteen genotypes with missing climate data. I estimated the climatic distance between each region pair as the F-statistic of a multivariate analysis of variance (MANOVA) with climatic PCs as response variables and region of origin as predictor.

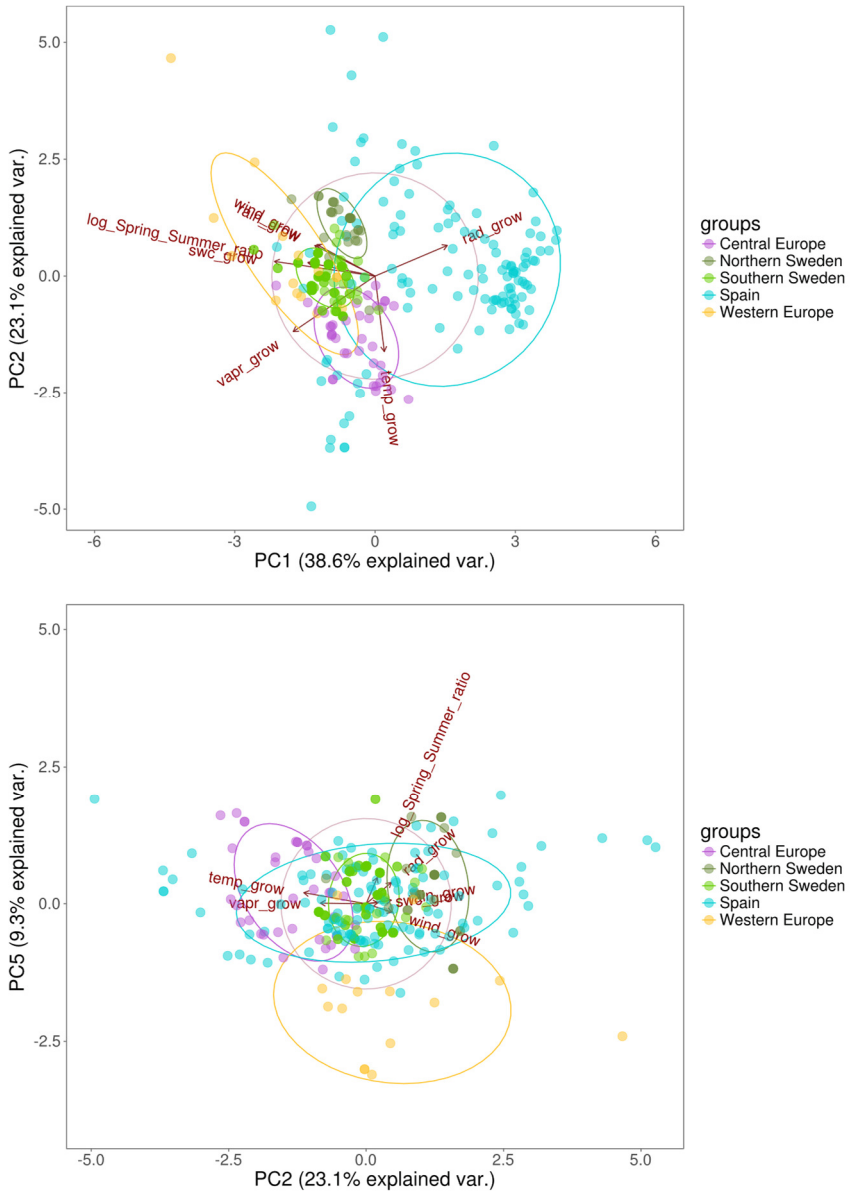

Figure 5: **PCA of climatic variation among accessions**
Plot of first two climatic principal components. Each point is one accession and accessions are colored by region of origin. Arrows indicate loadings of each climatic variable: rad_grow=solar radiation; temp_grow=temperature; vapr_grow=humidity; swc_grow=soil water content; wind_grow= wind speed; rain_grow= precipitation.

2.1.9 POPULATION GENOMIC ANALYSIS

I performed principal component analysis (PCA) of genomic data (thinned to 1kb) using the *adegenet* package (Jombart et al., 2016) with missing data converted to the mean (Figure 6).

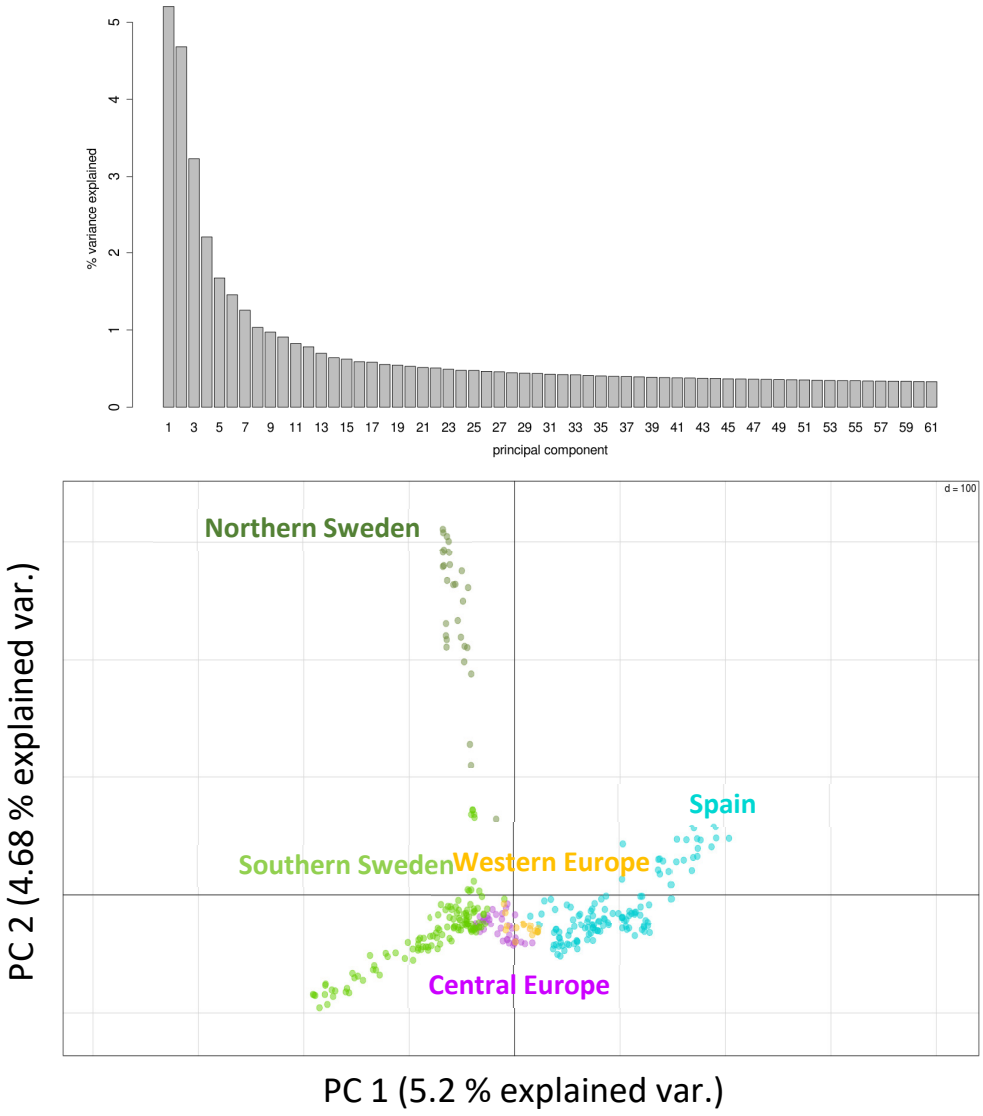

Figure 6: **PCA of genetic variation among accessions**
Top) Barplot of amount of variance explained by each genetic principal component; bottom) plot of first two genetic principal components. Each point is one accession and accessions are colored by region of origin.

Comparing phenotypic differentiation (Q_{ST}) to the distribution of F_{ST} is a useful method to reveal signatures of local adaptation (T. Leinonen, McCairns, O'hara, & Merilä, 2013; Whitlock & Guillaume, 2009). I calculated genome-wide, pairwise F_{ST} estimates between using the *hierfstat* package (function *basic.stats,* Nei's Fst) (Jérôme Goudet, 2005). I set negative F_{ST} values to zero before the 95th percentile was calculated.

For stomata density, stomata size and WUE, I estimated the respective phenotypic differentiation between regions, Q_{ST}, as:

$$Qst = VarB/(VarW + VarB)$$

where VarW is the genetic variance within regions and VarB the genetic variance between regions as described in Kronholm et al. (2012). Variance components were estimated with the lme function mentioned previously, including block as a fixed effect and genotype nested within region as a random effect. We used the genotype effect as an estimate for within-region variance and the region effect as an estimate for between-region variance. Since we did not measure replicates for δ13C, we could not estimate the environmental variance component. Thus, we adapted the model to include only region as a random effect, which was our estimate of between-region variance and used the residual variance as within-region variance. This approach underestimates Q_{ST} and its use for detecting signatures of local adaptation at the phenotypic level is conservative (Whitlock & Guillaume, 2009).

To test whether Q_{ST} estimates significantly exceed the 95th percentile of the F_{ST} distribution, I permuted the phenotypic data by randomizing genotype labels to keep heritability constant. For each permutation and phenotype, I calculated the difference between each Q_{ST} value and the 95th percentile of the F_{ST} distribution. I used the 95th percentile of the maximum Q_{ST}-F_{ST} distance distribution as a threshold for determining if phenotypic differentiation significantly exceeds neutral expectations. Since this test takes the maximum Q_{ST}-F_{ST} distance for all population combinations in each permutation, it does not require multiple testing correction.

2.1.10 STATISTICAL ANALYSIS

I conducted statistical analysis using R (R Development Core Team, 2008). I created plots using the following libraries: *ggplot2* (Wickham, 2009)*, ggthemes* (J. B. Arnold et al., 2017), *ggmap* (Kahle & Wickham, 2013)*, ggbiplot* (Vu, 2011) and *effects* (Fox et al., 2016).

I used Generalized Linear Models (GLM) to test the effect of block, origin, pot position in tray (edge or center) and leaf size on each phenotype (stomata density, stomata size and $\delta^{13}C$). For stomata density I also included the number of analyzed images as a co-factor. The error distribution was a quasi-Poisson distribution for stomata density and size and Gaussian for $\delta^{13}C$. I log transformed stomata density to avoid over-dispersion. I determined the significance of each predictor via a type-II likelihood-ratio test (*Anova* function of the *car* package).

Significant differences between regions were based on GLMs including only significant predictor variables and determined with Tukey's contrasts using the *glht* function of the *multcomp* package (Hothorn et al., 2017). I also used GLMs to test the impact of all climatic PCs on phenotypic traits, while accounting for population structure with the first 20 PCs for genetic variation, which explain 28% of genetic variation (see above). Additionally, for $\delta^{13}C$ I also tested a simpler model including climatic parameters but not population structure. From the resulting models, I created effect plots for significant environmental PCs using the *effects* package (Fox et al., 2016). Further, I used GLMs with binomial distribution to test whether any of the climatic PCs significantly predicts the allelic states of loci associated with WUE in GWAS.

2.2 CHAPTER 2 – THE EFFECT OF ECOLOGICAL RESTORATION ON THE GENETIC DIVERSITY OF TWO FLOODPLAIN *ARABIS* SPECIES

2.2.1 PLANT MATERIAL AND DNA EXTRACTION

The sampling area comprises the fossil, dyke-protected flood plain of the River Rhine near Riedstadt in Hessen, Germany. The area is dominated by arable fields, but also contains remnants of pristine flood meadow communities in low lying depressions that are submerged by ascending groundwater during high floods of the River Rhine. Since ca. 20 years, new flood meadow communities have been restored on ex-arable land using the transfer of green hay from the pristine sites as donors to overcome significant dispersal limitation (Hölzel & Otte, 2003). During this process, hay from different donor sites was placed in distinct yet adjacent patches, making admixture possible. Since restoration is still ongoing, restored sites differ in age (Table 1).

Table 1: **Overview of sampled populations**

Population	Type	Year sampled	Year restored
A-1	restored	2016	1997
A-2	restored	2015 & 2016	1997
B	pristine	2015 & 2016	-
C	pristine	2015	-
D	pristine	2015 & 2016	-
E	restored	2015 & 2016	unknown
F	restored	2016	2000
G	restored	2016	2014
H	restored	2016	2006
I	pristine	2016	-

I performed the sampling in collaboration with Norbert Hölzel from the University of Münster. From previous monitoring of the sites we knew that populations that *Arabis* plants were not present every year in all sites. Thus, we sampled in two consecutive years to maximize the number of study sites. In the few sites sampled in both years, the genetic composition of the samples was not markedly different across years, so samples from both years were bulked for

these sites (Table 1) We harvested seeds from a total of 134 plants of *Arabis nemorensis/sagittata* in nine sites named from A to I, in order of collection. The presence of *A. sagitata* was not previously reported in these sites and was thus unexpected. Although some individuals showed the reduced stem leaf density and shorter siliques typical of *A. sagittata*, these phenotypic criteria were not always clearly distinguishable in the field, especially at the end of the season when siliques matured, so that the presence of the two species has remained overlooked in previous studies (Burmeier, Eckstein, Donath, & Otte, 2011). Thus, I determined species identity by post-hoc analysis, after the RAD-seq analysis revealed the presence of two taxonomic units.

Four sites were sampled in pristine habitat (B, C, D and I) and five in restored habitat (A, E, F, G and H, Figure 7). In site A two distinct stands of plants separated by about 100 m were sampled and were treated as sub-sites A-1 and A-2 throughout this study.

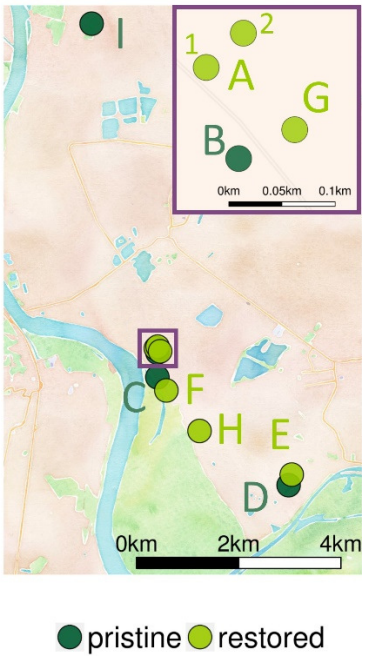

Figure 7: **Overview of study sites**
Each point on the map represents one site. The color of the points represents the type of the site, i.e. whether it is pristine or restored. The purple square represents the area shown in the zoomed inset. All sites, except site F are protected by a dyke.

To produce material for DNA extraction, I stratified seeds on wet filter paper for 6 days at 4°C in darkness. Afterwards I sowed seeds in soil (33% VM, 33% ED-73, 33% Seramis (clay granules)). After 4 to 6 weeks of growth in the greenhouse, I harvested about 200mg of leaf material from one offspring of each wild parent (genotype). I homogenized freshly harvested

leaf material using a Precellys Evolution homogenizer (Bertin technologies) for 2x20 seconds at 6800 rpm. I extracted DNA using the NucleoSpin Plant II Mini kit (Macherey-Nagel) following the manufacturer's instructions. I verified DNA quality using gel-electrophoresis with a 0.8% agarose gel. I measured DNA quantity using Qubit (broad-range kit) following manufacturer's instructions.

2.2.2 DRAFT GENOME ASSEMBLY

To facilitate genotype calling I assembled a draft genome for one *A. nemorensis* accession (ID 29). Library preparation and sequencing was done at the Cologne Center for Genomics. Three libraries were created: one paired-end library with approximately 280 bp insert size creating an overlap of 20 bp with 150 bp reads, and two mate-pair libraries with 3 kbp and 6 kbp inserts, respectively. The paired-end library was created using the Illumina TruSeq® Nano DNA Library Prep Kit, with 2µg of input DNA, without PCR. The mate-pair libraries were created using the Illumina Nextera® Mate Pair Library Prep Kit, with 4 µg of input DNA and 10 cycles of PCR for the 3 kbp library and 15 cycles of PCR for the 6 kbp library. All libraries were sequenced together as 150bp paired-end reads as part of an Illumina HiSeq 4000 lane for a total of 66 Gbp.

I used *FastQC* (Andrews, 2010) to quality-check the resulting reads. I filtered the resulting reads to remove reads shorter than 100bp and trimmed Illumina adapters using *Cutadapt* (M. Martin, 2011). I assembled reads using the *ALLPATHS-LG* assembler (Gnerre et al., 2011) with default settings, running it on the CHEOPS high-performance computing cluster of the University of Cologne. The resulting contig assembly had a size of 199 Mbp and an N50 of 47 kbp. The scaffold assembly had a size of 206 Mbp and an N50 102 kbp. To further scaffold the genome, I generated 2.9 Gbp of PacBio sequence data. Library preparation and sequencing was done at the Max-Planck-Institute for Plant Breeding Research (Cologne, Germany). I scaffolded the genome using *OPERA-LG* with default settings (Gao, Bertrand, Chia, & Nagarajan, 2016). This increased the N50 to 150kbp. To achieve chromosome-level assembly, I made use of the available reference genome of *Arabis alpina* (Jiao et al., 2017; Willing et al., 2015). To facilitate whole-genome alignment, repetitive regions in the genome were masked using RepeatMasker (Smit, Hubley, & Green, 2013). I used 'brassicaceae' as the search term

('-species' option) for the RepeatMasker repeat-database. I created the pseudo-chromosome assembly using the *CoGe* website (Lyons & Freeling, 2008). I aligned the *A. nemorensis* draft genome with the *Arabis alpina* genome using *SynMap2* (Haug-Baltzell, Stephens, Davey, Scheidegger, & Lyons, 2017) and afterwards performed synthenic path assembly (Lyons, Freeling, Kustu, & Inwood, 2011) to assemble the chromosomes. The pseudo-chromosomes had a total size of 192 Mbp and were used for all following analyses.

The size of the genome was estimated by flow-cytometry, which was performed commercially at Plant Cytometry Services. The estimated genome size was 274 Mbp. Thus, the assembly size was 70% of the genome size.

2.2.3 ANNOTATION

The genome annotation was performed by Wen-Biao Jao and Korbinian Schneeberger from the Max-Planck-Institute for Plant Breeding Research in Cologne. To detect and annotate Transposable elements (TE) specific to *A. nemorensis* (not in the previously used database) the softwares RepeatModeler (Smit & Hubley, 2008) and RepeatMasker (Smit et al., 2013) were used. The consensus repeat library was first constructed by RepeatModeler and then used by RepeatMasker to search TEs. Protein-coding genes were annotated by integrating predictions of *ab initio* gene annotation tools and alignments of homologous proteins. Three different tools including Augustus v3.2.3 (Stanke & Waack, 2003), GlimmerHMM v3.0 (Majoros, Pertea, & Salzberg, 2004) and SNAP v2013 (Korf, 2004) were used to predict the initial gene models. Protein sequences from *A. thaliana*, *A. lyrata* and *A. alpina* (Arabidopsis Genome Initiative, 2000; Hu et al., 2011; Willing et al., 2015) were aligned to the assembly by the tool Exonerate v2.2.0 (Slater & Birney, 2005). Then, the *ab initio* predictions and protein alignment hits were further combined to build the consensus gene models by the tool EVidenceModeler (EVM) v2012 (Haas et al., 2008). Finally, TE related genes in these models were annotated by checking the TE annotation, blastp (Altschul, Gish, Miller, Myers, & Lipman, 1990) alignments with Plant TE related proteins and blastp alignments with *A. thaliana* proteins. If a gene's protein sequence had blastp alignment identity and coverage both larger than 50% with a TE related protein, or at least 30% of its exon regions overlapped with TEs but had no good blastp hit (identity >50% and coverage >50%) from *A. thaliana* protein sequences, this gene would be marked as a TE related gene.

2.2.4 RAD-SEQUENCING AND SNP CALLING

I genotyped 134 samples using the original RAD-sequencing (RADseq) protocol (Etter et al., 2011), with the following modifications (see Appendix for complete protocol): i) I used the enzyme KpnI-HF (New England Biolabs) for DNA digestion, ii) I ligated digested DNA with complementary adapters containing one of ten different barcodes and a stretch of five random nucleotides, used for post-hoc removal of PCR duplicates (Table 2), iii) I created 14 pools of 10 barcoded samples each in equal amounts, iv) I used indexed reverse primers for amplification, described in (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012), to allow multiplexing of pools. Libraries were sequenced on two Illumina HiSeq 4000 lanes with 2x150bp.

Table 2: **Adapter sequences used for RAD-seq library construction**
For adapters, the name contains the enzyme specific overhang and the barcode sequence. P1f and P1r with the lame barcode are annealed to create the adapter. The second adapter is universal with only one base overhang. PCR 1 is the universal PCR forward primer and PCR2 are reverse primers with a specific barcode sequence.

first adapter	
KpnI_P1f_AACGTT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAACGTTGTA*C
KpnI_P1r_AACGTT	[PHO]AACGTTNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
KpnI_P1f_CGATAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCGATAAGTA*C
KpnI_P1r_CGATAA	[PHO]TTATCGNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
KpnI_P1f_GCTAGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGCTAGCGTA*C
KpnI_P1r_GCTAGC	[PHO]GCTAGCNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
KpnI_P1f_TTGGCC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTTGGCCGTA*C
KpnI_P1r_TTGGCC	[PHO]GGCCAANNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
KpnI_P1f_CCTCAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCCTCATGTA*C
KpnI_P1r_CCTCAT	[PHO]ATGAGGNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
KpnI_P1f_GGCTTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGGCTTAGTA*C
KpnI_P1r_GGCTTA	[PHO]TAAGCCNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
KpnI_P1f_TGTAAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTGTAATGTA*C
KpnI_P1r_TGTAAT	[PHO]ATTACANNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
KpnI_P1f_ACATTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNACATTAGTA*C

KpnI_P1r_ACATTA	[PHO]TAATGTNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
KpnI_P1f_CTACTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCTACTAGTA*C
KpnI_P1r_CTACTA	[PHO]TAGTAGNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
KpnI_P1f_GTTACC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGTTACCGTA*C
KpnI_P1r_GTTACC	[PHO]GGTAACNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
second adapter	
flex_P2f	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
flex_P2r	[PHO]AATTAGATCGGAAGAGCGAGAACAA
PCR primers	
PCR1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG
PCR2_Idx_1_ATCACG	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_2_CGATGT	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_3_TTAGGC	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_4_TGACCA	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_5_ACAGTG	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_6_GCCAAT	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_7_CAGATC	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_8_ACTTGA	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_9_GATCAG	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_10_TAGCTT	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_11_GGCTAC	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_12_CTTGTA	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGC

I used *FastQC* (Andrews, 2010) to quality-check the resulting reads. I trimmed adapters and removed reads shorter than 100bp using *Cutadapt* (M. Martin, 2011). I removed PCR duplicates based on a 5 bp stretch of random nucleotides at the end of the adapter, using the *clone_filter* module of *Stacks* version 1.37 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). I de-multiplexed samples using the *process_radtags* module from *Stacks*. I filtered

reads with ambiguous barcodes (allowed distance 2) and cut-sites, reads with uncalled bases and low-quality reads (default threshold).

For reference-based genotyping, I mapped reads using *BWA* (Li & Durbin, 2009) with default settings. I filtered mapped reads using *SAMtools* (Li et al., 2009) and custom python scripts using the following criteria to remove reads: mapping quality < 30, number soft-clipped bases > 30; reads were unpaired; mates mapped on different chromosomes; mate mapping distance > 700. I called genotypes using *SAMtools mpileup* and *VarScan2* (Koboldt et al., 2012) with the following options: base quality > 20; re-calculation of base quality on the fly (-E option); read depth > 14; strand filter de-activated; SNP calling p-value < 0.01. I filtered genotyped loci using *VCFtools* (Danecek et al., 2011) and custom python scripts removing loci with missing data in more than 5% of individuals and loci in masked (repetitive) regions. I clustered contiguous loci spaced less than 100 bp apart into RAD-regions. Regions with excessively low or high coverage are likely results of allele dropout or paralogous mapping, respectively. Thus, I removed regions fulfilling one of the following criteria: mean coverage of the region greater than twice the overall mean coverage; mean coverage of the region smaller than a third of the overall mean coverage; region maximum coverage greater than twice the mean maximal coverage over all regions; region shorter than 250bp; region longer than 1300bp.

After depth filtering I still found loci with a high frequency of heterozygotes (up to 100%), which is unlikely in highly selfing species. Notably, in over 90% of these loci only one of the two homozygous genotypes was observed. Moreover, these highly heterozygous loci clustered in high density on single RAD-fragments. Thus, these loci likely resulted from paralogous mapping, which passed the depth filter due to sequencing depth variation among RAD-regions. Therefore, I removed RAD-regions containing loci with a frequency of heterozygotes greater than 20%. This threshold was picked to limit the impact of noise from mapping artifacts while still allowing for reasonable levels of heterozygosity, since some low-level outcrossing likely occurs. From the resulting genotype dataset, I extracted single nucleotide polymorphisms (SNPs) using *VCFtools* (Danecek et al., 2011).

For *de-novo*-genotyping I used the *Stacks* 2.2 *denovo_map.pl* pipeline (Catchen et al., 2013). The aim of the de-novo analysis was to test whether one would reach similar results and conclusions without the use of a reference genome. Thus, I performed this analysis without

knowledge about the two species (since we gained this from the reference-based analysis) and ran *Stacks* for all samples combined. As recommended by the authors of the tool (Rochette & Catchen, 2017), I first used a subset of 15 representative genotypes to tune the parameters (-M and -n) of the algorithm, which control the number of mismatches between stacks within (M) and between (n) individuals. I varied M and n from 1 to 9. For each set of parameters, I analysed the number of loci shared between 80% of the samples. This measure peaked at the value six for M and n. Thus, I used this value for both parameters for the full analysis. We ran the *denovo_map.pl* pipeline using 0.01 as the p-value threshold for calling genotypes and SNPs, and otherwise default options. I used the *populations* program to create a VCF-file for further analysis using the following filters: 5% maximum missing data per locus; 20% maximum observed heterozygosity per locus; locus must be present in all sites.

2.2.5 POPULATION GENETICS STATISTICS

I did all statistical analysis using R version 3.4.4 (R Development Core Team, 2008). The following packages were used for plotting: *ggplot2* (Wickham, 2009), *ggmap* (Kahle & Wickham, 2013), *ggthemes* (J. B. Arnold et al., 2017), *ggsn* (Baquero, 2017) and *heatmap3* (Zhao, Guo, Sheng, & Shyr, 2015). I performed all analysis for the reference-based as well as the *de-novo*-based dataset and compared the results. I used the *vcfR* package (Knaus & Grünwald, 2017) to load VCF-files into R and make the SNP data available for processing with other libraries. Based on our annotation I determined whether SNPs are in coding regions and whether they are synonymous or non-synonymous using the *PopGenome* package (Pfeifer, Wittelsbuerger, Li, & Handsaker, 2018). I performed principal component analysis (PCA) of SNP data for all samples using the *adegenet* package (Jombart et al., 2016). Missing data was scaled to the mean for PCA.

Based on the first principal component most individuals (83%) could be assigned to two distinct taxonomic groups. I used molecular methods to assign species labels to the two taxonomic groups. I sequenced the internal transcribed spacer (ITS) sequence of nine individuals, three from the first (left) and five from the second (right) cluster and one located between the clusters (Table 8). Primers for amplification were taken from (Mummenhoff, Franzke, & Koch, 1997). I used the sequences as input to the taxonomy tool of the *Brassibase* website (Kiefer et al., 2014). Interpretation of the output required ploidy information.

Therefore, I collected leaf samples from twelve individuals (seven putative *A. sagittata*, two putative *A. nemorensis*, three putative hybrids; Table 9) and their ploidy determined by Plant Cytometry Services (Didam, Netherlands).

I conducted all population genetic analysis separately for the two species. Individuals, which were not assigned to any species, were likely interspecific hybrids and excluded from further analysis. I used the *pegas* library to calculate within-site genetic diversity (Nei's π; average pairwise nucleotide differences) for each site (Paradis, Jombart, Schliep, Potts, & Winter, 2016), excluding sites with less than two individuals per respective species. To scale the estimates of π, I divided the average number of pairwise nucleotide differences among samples by the total number of successfully genotyped bases, excluding all bases which failed any of our previously described genotype filters (missing data, region heterozygosity, region depth and length). For the de-novo pipeline, I extracted the total number of genotyped bases from *Stacks* output. I calculated correlation coefficients between reference-based and *de-novo*-based π estimates using Pearson's method. I calculated pairwise F_{ST} (Nei, 1987) and genetic distance (Cavalli-Sforza & Edwards, 1967) between all pairs of sites using the *hierfstat* package (Jerome Goudet & Jombart, 2015). Negative F_{ST} values were set to zero. Differences of genetic distance and F_{ST} among pristine and restored sites were tested using a Wilcoxon-rank-sum-test on pairwise distance matrices. I tested for correlation between the distance matrices of the reference and *de novo* datasets using a Mantel test with 10000 permutations.

2.2.6 ADMIXTURE ANALYSIS

For ADMIXTURE analysis (Alexander, Novembre, & Lange, 2009), I converted VCF-files to bed-files using PLINK (Purcell, 2009; Purcell et al., 2007). First, I conducted ADMIXTURE analysis for all samples combined for K=1 to K=10 (reference-pipeline only). Then for each of the species and pipeline (reference/*de novo*), I ran ADMIXTURE analysis for K=2 to K=10, with 10 iterations of cross-validation each. Before plotting, I normalized clusters across runs using CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). I created plots using a custom R-script.

2.3 CHAPTER 3 – PATTERNS OF GENETIC DIVERSITY IN *A. NEMORENSIS*

2.3.1 FIELD SAMPLING OF POTENTIAL *A. NEMORENSIS* POPULATIONS

I identified potential populations of *A. nemorensis* through the search of the Deutschlandflora database (https://karten.deutschlandflora.de/map.phtml; focused on sites with "Schwerpunkt nach 2000"), inspection of herbarium material and personal communication with government workers, scientists and hobby botanists. This resulted in a list of 30 target sites, which were geographically reasonably well defined. Target sites were located in Hungary, Germany, Austria and the Czech Republic. Several target sites were located in natural reserves (Nvw, Alp, Con-1, Con-2, GDB, Lob, Mam-1, Mam-2) I was able to acquire sampling permits for all sites in Germany and Austria, but not Hungary. Since all Hungarian sites were natural reserves, I did not visit any of them.

To visit the target sites for sampling I did a round-trip from 07.07.2016 to 27.07.2016 and visited 24 sites. I was able to find plants resembling *A. nemorensis* in 13 sites (Figure 8), which were all located in Germany, except one which was in Austria. I was not able to visit the two target sites in the Czech Republic due to either closed roads and or fenced (private) property. For all other sites, *A. nemorensis* plants were either not present or I could not locate them despite extensive search of the surrounding area (e.g. due to inaccurate coordinates/descriptions). For all *A. nemorensis* populations I found, I took photographs of the surrounding and representative individuals, sampled seeds from at least 10 plants, marking down the GPS-location of each plant. Sampled siliques/seeds were stored in paper bags. To comply with the Nagoya protocol, each sampled individual was given a unique name.

After the trip, seeds were allowed to after-ripen in siliques for two weeks to ensure complete drying of the material. Then I removed the seeds from siliques and separated them from the rest of the plant using several sieves (RETSCH-Analysensieb) of different sizes (450 µm - 1.7 mm).

In 2017, I collected an additional population in the Alps, located at a lakeshore and obtained seeds from 30 plants from an *ex situ* population of *A. nemorensis* from the Botanical Garden Halle, which was originally collected in the drainage area of the Elb river. Additionally, I

collected herbarium samples of *A. sagittata* and *A. hirsuta* from the University of Munich (Botanische Staatssammlung).

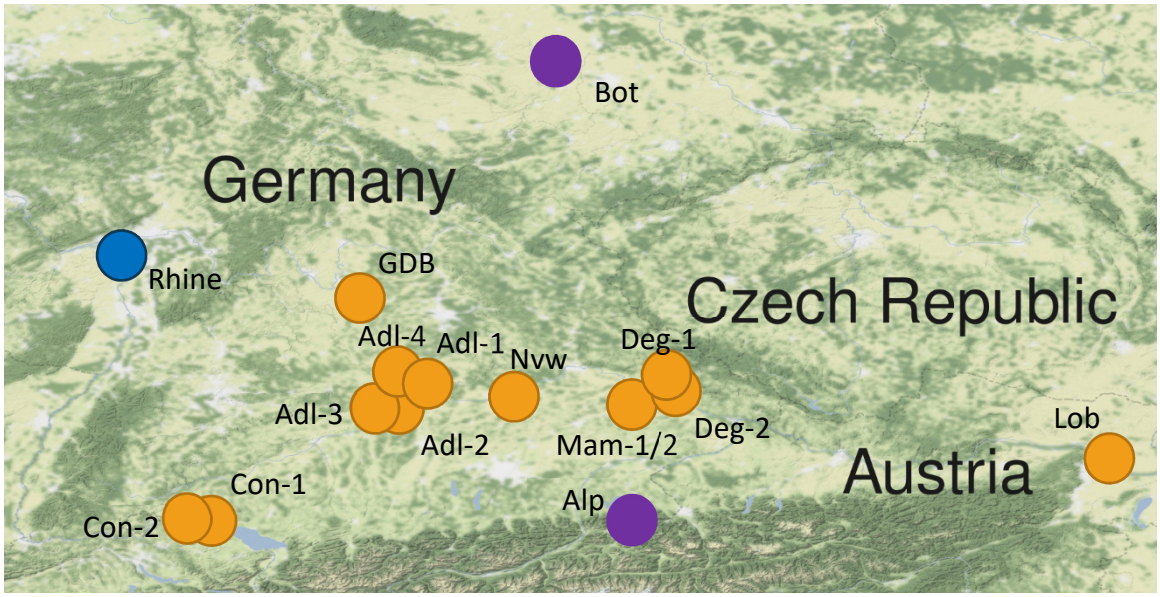

Figure 8: **Map of sampled populations**
Each population is represented by a dot. Populations not previously described in this work are marked by orange (sampled in 2016) and purple (sampled in 2017) dots. The blue dot is the previously described Rhine population. Population Mam-1/2 are two populations in very close proximity whose dots completely overlap.

2.3.2 SEED AMPLIFICATION, DNA EXTRACTION AND RAD-SEQ

I vernalized seeds from 10 plants (accessions) per population for 6 days at 4 °C in darkness. I sew several seeds per accession in pots filled with VM soil and put them in the greenhouse for germination. After germination was completed, I removed all but one plant per accession. After 4 weeks of growth I harvested about 100mg of young, small leaves from each plant into tubes containing ceramic beads for DNA extraction (see below). After 8 weeks of growth I moved the plants to the greenhouse garden for vernalization (November) to induce flowering. In spring, I moved the plants back to the greenhouse. When plants started to bolt, I attached the bolt to a wooden stick and covered each plant with a transparent plastic bag with small holes to prevent cross-pollination. I harvested seeds of each plants, when all siliques were brown, and the first siliques started to open by themselves. After several weeks of after-ripening, I cleaned the seeds with sieves as described above and assigned a new number to each seed bag (genotype information stored in a table). Note that, these seeds displayed low germination rates even after months of storage, indicating potential viability problems.

For DNA extraction, I added 100µl of TE Buffer to each freshly harvested leaf sample and homogenized the leaf samples with the Precellys Evolution homogenizer for 2 x 20s with 20s break at 6800rpm. I extracted DNA using the NucleoSpin Plant II Mini Kit (Macherey-Nagel, Düren, Germany), following manufacturer's instructions. Note, for dried herbarium samples I used twice the amount of the Buffers PL 1 and PC as well as RNAase A. I checked DNA quality using gel-electrophoresis on a 0.8% agarose gel, requiring a band > 10 kb and only little degradation. I quantified the amount of DNA using a Qubit device (Invitrogen, Carlsbad, USA) with High Sensitivity kits. I re-calibrated the Qubit device for each new working solution. To ensure adequate mixing for accurate measurements, I vortexed DNA samples briefly before taking the sample for measuring. Minimum concentration for the RAD-seq protocol was 12.5 ng/µl with at least 500ng of total DNA. I concentrated samples with lower concentrations but sufficient total DNA amount using a vacuum centrifuge (Eppendorf Concentrator Plus) and measured DNA quantity again.

RAD-libraries were prepared following protocol used in Chapter 2 (see Appendix for RAD-protocol) with 50 µl reaction volume for the digest. Samples were multiplexed together with all samples from the Upper Rhine populations and sequenced on five HiSeq4000 lanes, with 2 x 150 bp reads. The fifth lane was also loaded with samples for whole genome sequencing. The yield on the first four lanes was strongly reduced since they were pure RAD-seq material, which caused reads to fail one of the Illumina quality filters (library not complex enough in the first 20 bp). Thus, each of the four libraries was sequenced again on 25% of a HiSeq 4000 lane, together with other genomic libraries by the CCG.

I used *FastQC* (Andrews, 2010) to check the quality of the resulting reads. I trimmed Illumina adapters and removed reads shorter than 100bp using *Cutadapt* (M. Martin, 2011). I used the *clone_filter* module of *Stacks* version 1.37 (Catchen et al., 2013) to remove PCR duplicates based on a 5 bp stretch of random nucleotides at the end of the adapter. I de-multiplexed samples using the *process_radtags* module from *Stacks*, with the following filter settings: max. ambiguous barcode distance = 2; filter out reads with invalid cut-sites, uncalled bases and low-quality (default threshold).

I mapped reads using *BWA* (Li & Durbin, 2009) with default settings. I filtered mapped reads by combining *SAMtools* (Li et al., 2009) and custom python scripts applying the following

criteria to filter reads: mapping quality < 30, number soft-clipped bases > 30; reads were unpaired; mates mapped on different chromosomes; mate mapping distance > 700. For genotype calling, I used *SAMtools mpileup* and *VarScan2* (Koboldt et al., 2012) with the following options: read depth > 14; base quality > 20; strand filter de-activated; re-calculation of base quality on the fly (-E option); SNP calling p-value < 0.01. To maximize the number of called genotypes (e.g. by reducing the amount of missing information), the genotype filtering pipeline was applied to different sets of individuals (e.g. all samples, each separate species, two species together) depending on which individuals were needed for a specific analysis. I filtered genotyped loci using *VCFtools* (Danecek et al., 2011) to remove loci with missing information for more than 5% of individuals. I clustered contiguous loci spaced less than 100bp apart into RAD-regions. Since RAD- regions with excessively low or high coverage are likely results of allele dropout or paralogous mapping, respectively, regions fulfilling one of the following criteria were removed: mean coverage of the region greater than twice the overall mean coverage; mean coverage of the region smaller than a third of the overall mean coverage; region maximum coverage greater than twice the mean maximal coverage over all regions; region shorter than 250bp; region longer than 1300bp. In the predominantly selfing *Arabis* species, regions containing loci with high heterozygosity likely resulted from paralogous mapping errors. Thus, for the full dataset, I removed regions from the dataset with heterozygous frequency over 50% or 30% if either alleles' homozygous frequency was below 5%. For *A. nemorensis*- and *A. sagittata*-specific datasets these values were reduced to 20% and 10%, as they are diploid (in contrast to *A. hirstuta*) and the expected heterozygosity is thus lower. From the resulting genotype dataset, I extracted single nucleotide polymorphisms (SNPs) using *VCFtools* (Danecek et al., 2011). In total, 22 individuals were removed due to high proportion of missing data. Thus, 288 individuals were successfully genotyped.

Table 3: **Overview of population locations and the number of sampled and sequenced individuals per population**

Population	*Latitude*	*Longitude*	*Indiv. sampled*	*Indiv. sequenced*
Adl-1	48.70828	10.78728	16	10
Adl-2	48.594906	10.638422	12	9
Adl-3	48.62403	10.445017	15	10
Adl-4	48.84075	10.698425	10	10
Nvw	48.713775	11.58942	13	10

Deg-1	48.834098	12.917208	15	10
Deg-2	48.772525	12.952117	18	10
Mam-1	48.653949	12.589727	14	7
Mam-2	48.664051	12.591689	15	7
Lob	48.191911	16.482789	20	10
Alp	47.690921	12.581824	11	10
Con-1	47.696701	9.11086396	13	10
Con-2	47.737604	8.91845402	15	10
GDB	49.531369	10.349882	13	10
Bot	51.48848	11.961173	30	8

2.3.3 POPULATION GENETIC ANALYSIS

I performed most analyses, unless otherwise stated, using R (R Development Core Team, 2008) and the ggplot2 package (Wickham, 2009) for plotting, unless otherwise stated. For these analyses I included the previously collected data of both pristine and restored sites in the Upper Rhine population (Chapter 2) and merged the samples into a single “Rhine” population. I loaded the genotype information from vcf-files and converted it to other formats using the package vcfR (Knaus et al., 2017). I used the *adegenet* package (Jombart et al., 2016) to perform principal component analysis (PCA) of genetic variation and plot the results. For PCA, I converted missing genotype information the overall mean for each polymorphism.

Additionally, I applied the clustering software ADMIXTURE (Alexander et al., 2009), which clusters individuals into K (user specified) groups, based on genetic variation but allows individuals to share ancestry from multiple clusters. To convert vcf-format to bed-format (ADMIXTURE input), I used the software PLINK (Purcell, 2009; Purcell et al., 2007). I ran ADMIXTURE with varying K from 2 to 9 and 10 iterations of cross-validation for each run. I used the minimum cross-validation error or the point of stabilization as an orientation for the optimal value of K, but visualized and considered results for all values of K. Before plotting, I normalized clusters across runs using CLUMPAK (Kopelman et al., 2015). I created stacked barplots using a custom R script. Additionally, I created a map with pie charts depicting the mean ancestry components for each population using the packages *ggmap* (Kahle & Wickham, 2013) and *scatterpie* (G. Yu, 2018).

Based on the analysis described above and previous results (Chapter 2), I suspected the presence of several species in my sample. Species identity of two clusters could be derived from previous results (Chapter 2) due to overlap of the datasets. The third species was identified based on herbarium samples and the ITS phylogeny, as described in Chapter 2. Since this species is expected to be tetraploid, this was confirmed by cytometric analysis of a subset of six performed commercially by Plant Cytometry Services (Didam, Netherlands). I repeated PCA and ADMIXTURE analysis for samples identified as *A. nemorensis* and performed all following analysis for these samples only. I calculated diversity within and between species using the *dist.dna* function from the *ape* package (Paradis et al., 2019). Diversity estimates were scaled by the number of successfully genotyped bases. To calculate the F_{IS} for each species, I used the *hierfstat* package (Jerome Goudet & Jombart, 2015). I calculated the mean genome-wide F_{IS} for each population in each species and subsequently calculated the species-wide mean. For *A. nemorensis* I excluded low diversity populations with π<5e-05, due to a low number of informative SNPs.

The following analyses I performed only for *A. nemorensis* samples. I calculated pairwise genetic distances (d_{XY}) between all *A. nemorensis* samples using a custom python script and plotted the result in R using the *heatmap3* package (Zhao et al., 2015). D_{XY} values were scaled by the number of genotyped bases. To calculate genetic differentiation (F_{ST}) and genetic distance (Cavalli-Sforza & Edwards, 1967) between all pairs of populations, I used the *hierfstat* package (Jerome Goudet & Jombart, 2015). I visualized the results as a heatmap using the *heatmap3* package. To calculate within population diversity (π) (Nei, 1987) I used the *pegas* package (Paradis et al., 2016) and normalized the results per base-pair.

To analyse the variation in climatic conditions among populations, I obtain publicly available climate data for all sites. Specifically, I obtained altitude data and average monthly data for temperature, precipitation, wind speed, solar radiation, water vapor pressure and all bioclimatic variables from the *WorldClim* database (Fick & Hijmans, 2017). I downloaded all data in 2.5-arcminute resolution. Further, I extracted average monthly soil water-content data from the *Global High-Resolution Soil-Water Balance dataset* (Trabucco & Zomer, 2010) which I obtained from the CGIAR-CSI geoportal, which offers datasets related to climate and agriculture. For each population I calculated the annual average of all monthly climatic variable. In order to reduce the number of variables, I calculated pairwise correlation

coefficients for all variables and removed bioclimatic variables, which with r>0.8 with any other variable. “Raw” variables, e.g. temperature and precipitation, were not removed. To characterize climatic differences among populations while accounting for correlations among variables, I conducted PCA. PCA was plotted using *ggbiplot* (Vu, 2011). Climatic distance among populations was calculated as the Euclidian distance of populations on the first three PCs. I used pairwise Mantel-tests to test for correlation between climatic distance and F_{ST} and genetic distance, respectively. To compare the amount of relative variation of each climatic variable, I normalized each variable by dividing all values by the maximum of that variable in the whole sampling region and scaled the results to percent. The sampling region was defined by a rectangle encompassing all sampled sites.

2.4 CHAPTER 4 – PHENOTYPIC DIVERGENCE BETWEEN *A. NEMORENSIS* AND *A. SAGITTATA*

2.4.1 COMMON GARDEN EXPERIMENT IN SEMI-NATURAL CONDITIONS

I performed a common garden experiment in semi-natural conditions to quantify of phenotypic diversity among *A. nemorensis* and *A. sagittata* genotypes collected in the Rhine populations.

I stratified field-collected seeds of 50 *Arabis* accessions from five Rhine populations (10 each) on wet filter paper for 6 days at 4 °C in the dark. I sowed seeds in pots on the 16. and 17.09.2015, sowing four seeds per pot. Seeds sown on the second day were kept outside on wet filter paper until sown. I used two-liter round pots filled with 'Topferde' (Einheitserde, Sinntal-Altengronau, Germany). I randomized pots in 12 blocks with one replicate per accession per block, for a total of 600 pots. After germination, seedlings were thinned to one plant per pot. Pots without germination were marked, to exclude plants with germination after winter. Due to seed contamination from nearby fields, pots were weeded regularly. In spring, the gardeners started watering the plants, when necessary to avoid death due to drought.

Phenotyping started in spring 2016. On March 18th I measured the diameter of all plants. Additionally, I scored herbivore damage on the rosette. Plants were assigned to 6 categories (0 to 5), based on the approximate area of the rosette, damaged by herbivores: 0: 0% damage; 1: 1 - 20%; 2: 21-40%; and so on (for examples see Figure 9). I monitored bolting and flowering of plants several times a week. Scoring was done every day after first plant bolted. On April 12th I measured the height of each plant. Additionally, I determined the number of stem leaves. When seed set was completed, I measured the final height of each plant and harvested all siliques. I collaborated with Marina Sell, a student from the University of Münster, who used material collected in the common garden to measure several fitness related phenotypes in the first 6 blocks of the experiment: Number of siliques, seeds per silique, seed weight and silique length.

Figure 9: **Overview of degrees of herbivory damage**
Example images of rosettes for each level of herbivory damage measured in the experiment.

For statistical analysis I used R (R Development Core Team, 2008), with the following libraries: ggplot2 (Wickham, 2009), nlme (Pinheiro et al., 2015), car (Fox et al., 2018), psych (Revelle, 2019), MASS (Ripley et al., 2018), corrplot (Wei et al., 2017), multcomp (Hothorn et al., 2017) and plyr (Wickham, 2016). Flowering and bolting times are expressed in number of days since sowing. Flowering time data was incomplete (22% of bolting plants did not have a flowering date). This is likely because I failed to transfer data of late flowering plants to the digital table. Unfortunately, the original (written) data sheets were not available anymore when I realized the mistake. Thus, I excluded flowering time from the analysis. Flowering time and bolting time were correlated ($r=0.60$, $p<<0.0001$).

I calculated pairwise correlations between all phenotypes for each species separately (p-values adjusted with Holm method) and plotted results using the *corrplot* function (with clustering activated (hclust)). To test for genetic differences within and between species, I

created a regression model for each phenotype, with the phenotype as the response variable and the following predictors: block, the taxonomic group (*A. nemorensis, A. sagittata*, hybrid), and the accession nested within the taxonomic group ("genotype effect"). Whenever possible, I assumed a Gaussian error distribution for all models. However, if over- or underdispersion of residuals was observed (residual deviance differing by more than an order of magnitude from the degrees of freedom) a dispersion parameter for the negative binomial error distribution was estimated and then used for the model. If this failed, data was additionally square-root transformed. Predictor significance was assessed using likelihood ratio tests (type-II test; *car* package *Anova* function). To visualize phenotypic variance among taxonomic groups, I calculated the phenotypic mean per accession and plotted stripcharts comparing taxonomic groups. To assess pairwise significance among taxonomic groups I performed generalized linear hypothesis tests (GLHT) with Tukey contrasts. I calculated broad sense heritability for each phenotype in the same manner as in Chapter 1.

Table 4: **Overview of accessions used in common garden experiment**

Exp.ID	Accession	Site	Type	Species
1	2	A-2	*restored*	*A. sagittata*
2	3	A-2	*restored*	*A. nemorensis*
3	5	A-2	*restored*	*A. sagittata*
4	6	A-2	*restored*	*Hybrid*
5	7	A-2	*restored*	*A. nemorensis*
6	8	A-2	*restored*	*A. nemorensis*
7	9	A-2	*restored*	*A. nemorensis*
8	10	A-2	*restored*	*A. nemorensis*
9	13	B	*pristine*	*A. sagittata*
10	15	B	*pristine*	*A. sagittata*
11	16	B	*pristine*	*A. sagittata*
12	17	B	*pristine*	*A. sagittata*
13	19	B	*pristine*	*A. sagittata*
14	20	B	*pristine*	*A. sagittata*
15	24	C	*pristine*	*Hybrid*
16	25	C	*pristine*	*A. nemorensis*
17	26	C	*pristine*	*A. nemorensis*
18	27	C	*pristine*	*A. nemorensis*

19	29	C	*pristine*	*A. nemorensis*
20	30	C	*pristine*	*A. nemorensis*
21	31	D	*pristine*	*Hybrid*
22	32	D	*pristine*	*Hybrid*
23	33	D	*pristine*	*A. sagittata*
24	34	D	*pristine*	*A. sagittata*
25	35	D	*pristine*	*Hybrid*
26	36	D	*pristine*	*A. nemorensis*
27	37	D	*pristine*	*Hybrid*
28	38	D	*pristine*	*A. nemorensis*
29	39	D	*pristine*	*Hybrid*
30	40	D	*pristine*	*Hybrid*
31	63	E	*restored*	*A. sagittata*
32	64	E	*restored*	*A. sagittata*
33	73	E	*restored*	*A. sagittata*
34	66	E	*restored*	*A. sagittata*
35	67	E	*restored*	*A. sagittata*
36	68	E	*restored*	*A. sagittata*
37	69	E	*restored*	*A. sagittata*
38	70	E	*restored*	*A. sagittata*
39	71	E	*restored*	*A. sagittata*
40	59	A-2	*restored*	*A. nemorensis*
41	60	A-2	*restored*	*A. nemorensis*
42	61	B	*pristine*	*A. sagittata*
43	62	B	*pristine*	*A. sagittata*
44	72	B	*pristine*	*A. sagittata*
45	51	C	*pristine*	*A. nemorensis*
46	52	C	*pristine*	*A. nemorensis*
47	53	C	*pristine*	*Hybrid*
48	55	A-2	*restored*	*A. nemorensis*
49	56	D	*pristine*	*Hybrid*
50	74	A-2	*restored*	*A. sagittata*

2.4.2 SUBMERGENCE EXPERIMENTS

I conducted two independent experiments to quantify tolerance of plants to prolonged submergence underwater (flooding tolerance). The first started June 24th, 2016. I put seeds for 8 *A. nemorensis* accessions, 6 *A. sagittata* accessions, 3 *Arabidopsis thaliana* accessions (non-resistant control) and one *Rorippa palustris* accession (resistant control) on wet VM soil (Einheitserde, Sinntal-Altengronau, Germany) and stratified them in darkness at 5 °C for 3 days (two additional Arabis accessions were used but removed from analysis since they were not successfully genotyped). Then I moved plants to the greenhouse for germination. On July 6th plants pricked out to pots (8x8x8.5 cm) containing a soil mixture of VM, sand and coarse volcanic soil in equal parts. Each pot contained a single plant. Pots were randomly assigned to 21 identical blocks with 20 plants per block for a total of 420 plants in the experiment. Treatments (1 to 6 weeks of flooding + one control without flooding) were randomly assigned to each block with three blocks per treatment. On August 3rd at approximately 3 PM I submerged all non-control plants in water-filled transparent plastic boxes. All boxes were filled the day before with 15 l of water and 750 µl AlGo Universal (algal growth inhibitor). The algal growth inhibitor does not affect growth of water-plants (manufacturer's description). After each week, I removed the respective treatment group from the water for recovery. I scored plant survival two weeks after the end of each treatment, by checking whether new leaves had formed. Note, *R. palustris* started flowering during the experiment and was not completely submerged.

Starting December 5th, 2016, I conducted the second experiment, mostly following the protocol of the first, but with several changes: I used the same accessions but replaced the two not genotyped accessions with *A. sagittata*. Further I replaced the *A. thaliana* accessions with three *Arabis* hybrid accessions. I reduced the number of treatments to four (2-4 weeks submergence + control) but in turn increased the replicates per treatment to five. Immediately before submergence, I determined the rosette diameter, the number of leaves and the length of one marked leaf (nail polish) for each plant. The plan was to track these measures during submergence. However, this proved to be difficult and measures are likely not very reliable, indicated e.g. by negative growth measures (possibly because growth just stopped). On January 10th, 2017 at approximately 3 PM I submerged all treatment plants. For the first two weeks, plants were phenotyped weekly while submerged. Afterwards, only de-submerged

plants were phenotyped. Two weeks after de-submergence, I scored survival of all treatment plants, as in the first trial. Then I removed plants from pots and washed off the soil to free the roots. I measured the length from base to tip of each root. Finally, I oven dried all plants at 65 °C for one week and determined the dry weight.

For statistical analysis I used R (R Development Core Team, 2008), with the following libraries: ggplot2 (Wickham, 2009), multcomp (Hothorn et al., 2017), car (Fox et al., 2018) and MASS (Ripley et al., 2018). To plot the time-course of flooding survival, I calculated the mean and standard error per species. To test statistical differences in flooding resistance among species I used a generalized linear model, with mean survival rate per genotype as the response variable and treatment (flooding duration as numeric variable) and species nested within treatment as predictors. For the second experiment I also included the rosette diameter before flooding as a predictor. I used a negative binomial error distribution with an automatically estimated theta. To assess significance of the predictors I used a type-II likelihood ratio test implemented in the *Anova* function of the *car* package. To contrast all species, I used the model summary output and generalized linear hypothesis tests.

For the second experiment, I also tested for differences between *A. nemorensis* and *A. sagittata* in the following post-submergence phenotypes: rosette diameter and number of leaves 2 weeks after submergence, mean diameter growth rate and leaf creation rate 2 to 4 weeks after submergence, dry weight 4 weeks after submergence, root length 4 weeks after submergence. Diameter growth rate had a few negative values, likely due to measurement error. These were removed. For all phenotypes I build generalized linear models with treatment (numeric), species, their interaction term and block within treatment as predictors. For growth and leaf production rates I also included the starting values of the traits (2 weeks post-submergence) as covariate. For rosette diameter and number of leaves at two weeks post-submergence, I also included the pre-submergence values as covariate. If models with gaussian error distribution showed signs of over-/under-dispersion, I used a negative binomial error distribution with an automatically determined theta.

Table 5: **Overview of accessions used in the submergence experiments.**

Accession	Species	Experiment
29	*A. nemorensis*	both
36	*A. nemorensis*	both

71	*A. sagittata*	both
51	*A. nemorensis*	both
61	*A. sagittata*	both
34	*A. sagittata*	both
26	*A. nemorensis*	both
74	*A. sagittata*	both
55	*A. nemorensis*	both
53	*Hybrid*	2
68	*A. sagittata*	2
60	*A. nemorensis*	both
30	*A. nemorensis*	both
35	*A. sagittata*	both
73	*A. sagittata*	both
38	*A. nemorensis*	both
32	*Hybrid*	2
37	*Hybrid*	2
5	*A. sagittata*	2
Rorippa	*R. palustris*	both
Eden-7	*A. thaliana*	1
Got-7	*A. thaliana*	1
Pu-2-8	*A. thaliana*	1

2.5 CHAPTER 5 – PATTERNS OF GENE-FLOW BETWEEN *A. NEMORENSIS* AND *A. SAGITTATA*

2.5.1 DETECTION OF INTROGRESSION BETWEEN *A. NEMORENSIS* AND *A. SAGITTATA*

To accurately detect introgressions between *A. nemorensis* and *A. sagittata* a higher density of genetic markers than I had previously obtained from RAD-seq was required. Thus, I performed whole-genome re-sequencing for 35 accessions of *A. nemorensis* and *A. sagittata* (Table 6). Additionally, I sequenced one accession of *Arabis androsacea*, which was provided by Jean-Gabriel Valay (Jardin Alpin du Lautaret, France). For all samples, except *A. androsacea*, I extracted DNA for sequencing from 100 mg of freshly harvested leaf material. For each sample leaf material was mixed with 100µl TE-Buffer and ground in Precellys Evolution tissuelyser at 6800 rpm for 2x20 s with 20 s break. DNA was then extracted using the Macherey-Nagel NucleoPlant II Mini Kit, following the manufacturer's instruction. For *A. androsacea*, I was provided with a dried sample. About 40 mg of dried leaf material were ground using Precellys Evolution tissuelyser with setting 'hard'. DNA was extracted using Macherey-Nagel NucleoPlant II Mini Kit, following the manufacturer's instruction, but with twice the amount of buffers PL1 and PC and of RNAase. I checked DNA integrity using gel-electrophoresis with a 0.8% agarose gel. I measured DNA quantity using Qubit with the broad-range kit and 3µl of input DNA. Sequencing and library preparation were done at the Cologne Center for Genomics. Library preparation was PCR-free for all samples. Six samples were sequenced on Illumina HiSeq 4000 with 2x75 bp reads for 110 to 120 million reads. The remaining samples were sequenced on Illumina NovaSeq 6000 with 2x150 bp reads for 40 to 50 million reads.

Table 6: **Overview of accessions used for introgression analysis**.
For accession 174 the population was re-assigned because genetic analysis revealed a sample mix-up.

Accession	Species	Population
271	*A. sagittata*	Lob
267	*A. sagittata*	Lob
277	*A. sagittata*	Lob
272	*A. sagittata*	Lob
261	*A. sagittata*	Lob

174	*A. sagittata*	Adl-1 (re-assigned to Lob based on PCA)
173	*A. sagittata*	Adl-1
175	*A. sagittata*	Adl-1
176	*A. sagittata*	Adl-1
177	*A. sagittata*	Adl-1
19	*A. sagittata*	Rhine
325	*A. sagittata*	Rhine
380	*A. sagittata*	Rhine
344	*A. sagittata*	Rhine
94	*A. sagittata*	Rhine
370	*A. sagittata*	Rhine
116	*A. sagittata*	Rhine
323	*A. sagittata*	Rhine
359	*A. sagittata*	Rhine
114	*A. sagittata*	Rhine
360	*A. sagittata*	Rhine
34	*A. sagittata*	Rhine
326	*A. sagittata*	Rhine
72	*A. sagittata*	Rhine
25	*A. nemorensis*	Rhine
29	*A. nemorensis*	Rhine
30	*A. nemorensis*	Rhine
358	*A. nemorensis*	Rhine
337	*A. nemorensis*	Rhine
163	*A. nemorensis*	Adl-4
233	*A. nemorensis*	Deg-1
293	*A. nemorensis*	Con-1
309	*A. nemorensis*	Con-2
393	*A. nemorensis*	GDB
435	*A. nemorensis*	Alp
androsacea	*A. androsacea*	-

I cleaned the raw reads using STACKS process_shortreads version 2.2 (Catchen et al., 2013) with default quality filters and additionally removing reads shorter than 100 bp and trimming

reads to 150 bp. I mapped reads from all samples against the *A. nemorensis* reference genome using the mem algorithm of BWA (Li & Durbin, 2009) with default settings. I filtered mapped reads by combining *SAMtools* (Li et al., 2009) and custom python scripts applying the following criteria to remove reads: mapping quality < 30, number soft-clipped bases > 50% of the read length; reads were unpaired; mates mapped on different chromosomes; mate mapping distance > 700.

For genotype calling, I used *SAMtools mpileup* and *VarScan2* (Koboldt et al., 2012) with the following options: read depth per sample > 5; base quality > 20; strand filter de-activated; re-calculation of base quality on the fly (-E option); SNP calling p-value < 0.01. I filtered genotyped loci using *VCFtools* (Danecek et al., 2011) and custom python scripts to remove loci with missing information for more than 20% of individuals, loci with ambiguous bases or Ns in the reference and loci not passing filters of the VarScan2. In the predominantly selfing *Arabis* species, regions containing loci with high heterozygosity likely resulted from paralogous mapping errors. Thus, for the full dataset, I removed regions from the dataset with heterozygous frequency over 20%. For further analysis I retained only loci variable among or within *A. sagittata* and *A. nemorensis*, resulting in over 2,000,000 SNPs. To test for sample mix-ups, I conducted PCA of thinned dataset (1 SNP per 1000 kb). This analysis revealed that accession 174 clustered with the Lob population instead of Adl-1, likely due to a sample mix-up in during wet-lab work. Thus, I re-assigned this sample to Lob.

I created custom python scripts to implement the calculation of the following statistics based on population allele frequencies: D (Durand et al., 2011), f_D (Martin, Davey, & Jiggins, 2015) and q95(1,100) (Racimo, Sankararaman, Nielsen, & Huerta-Sánchez, 2015). D gives an estimate of relative amounts of gene-flow between a donor population and two potential donee populations. To obtain an unbiased estimate of gene-flow, one of the donee populations (control) should be isolated. Since all populations are located within the range overlap of the two species, this might not be the case, so estimates of gene-flow might be underestimated. D gives a good estimate of the genome-wide magnitude of gene-flow. However, it can be biased in small windows. Thus, to locate introgressed regions, f_D is more appropriate (S. H. Martin et al., 2015). f_D estimates the magnitude of introgression by comparing observed allele frequencies with expected frequencies under complete introgression. To detect candidate regions for adaptive introgressions (which are expected to

rise to high frequencies) the $q95_{(1,100)}$ statistic calculates the 95th percentile of frequencies of derived alleles which are fixed in the donor population and below 1% in the isolated (control) population. Thus, large values of $q95_{(1,100)}$ suggest adaptive introgression.

Additionally, I calculated d_{XY} (average pairwise genetic distance) for all pairs of populations. All statistics were calculated in 50 kbp sliding windows with 25 kbp overlap. ABBA-BABA statistics (D, f_D, q95) assume a specific phylogeny of tested populations (((P1, P2), P3), Outgroup). The outgroup is used to determine the ancestral allelic state and here was *A. androsacea* for all tests. Thus, for all analysis requiring the outgroup, I removed all sites which were missing or heterozygous in *A. androsacea,* retaining over 900,000 SNPs among *A. nemorensis* and *A. sagittata*. To test different scenarios of introgression, I calculated ABBA-BABA statistics for several different taxonomic combinations: P1 = *A. sagittata* non-Rhine populations, P2 = *A. sagittata* Rhine population, P3 = all A. nemorensis populations; P1 = A. sagittata population Lob (Austria), P2 = *A. sagittata* population Adl-1, P3 = all *A. nemorensis* populations; P1 = *A. sagittata* population Adl-1 (Germany), P2 = *A. sagittata* population Lob (Austria), P3 = all *A. nemorensis* populations; P1 = *A. nemorensis* non-Rhine populations, P2 = *A. nemorensis* Rhine population, P3 = *A. sagittata* population Lob; P1 = *A. nemorensis* Rhine population, P2= *A. nemorensis* non-Rhine populations; P3 = *A. sagittata* population Lob.

For the genome-wide estimate of D, I used a block-jack-knife procedure to assess significance with 5 Mb block size. D is a relative statistic depending on the difference between P1 and P2. Thus, to detect introgression in all taxa, I analysed the distribution of f_D across the genome for all windows with D>0 for all taxon combinations. Since f_D is only defined is D>0, I set all windows with D==NA to f_D=NA and all windows with D<=0 to f_D=0. To select candidate introgression windows, I defined three outlier levels: 1) f_D > 95% quantile of f_D; 2) condition 1 ∩ d_{XY}(P1, P2)>d_{XY}(P2, P3); and 3) condition 1 ∩ condition 2 ∩ q95(1, 100) > 0.9. Outlier level 2 encompasses loci with likely introgression since f_D is high and individuals from P2 are closer two P3 than P1, contradicting the whole-genome phylogeny. Outlier levels 3 encompasses loci with high frequency of alleles that are fixed between P1 and P3 in P2, which is a first indicator of potentially adaptive introgression (Racimo et al., 2015). To define introgressed blocks, I manually checked all windows with outlier level >= 2. I used Integrative Genomics Viewer (Robinson et al., 2011) and custom R plots to visualize genotype patterns in candidate regions. Based on these visualizations, I assessed whether introgressed blocks could be clearly defined,

based on the genotype distribution, i.e. whether stretches in the genome deviated from the overall pattern of genotypes (see Figure 10 for example). Candidate regions with high f_D not matching this criterion (Figure 10, middle) might still be the result of complex (ancestral) gene-flow. However, I decided to exclude these regions because further analysis required clearly defined introgressed blocks. For each introgressed region I defined break points of introgressed blocks and assessed which accessions have the introgression.

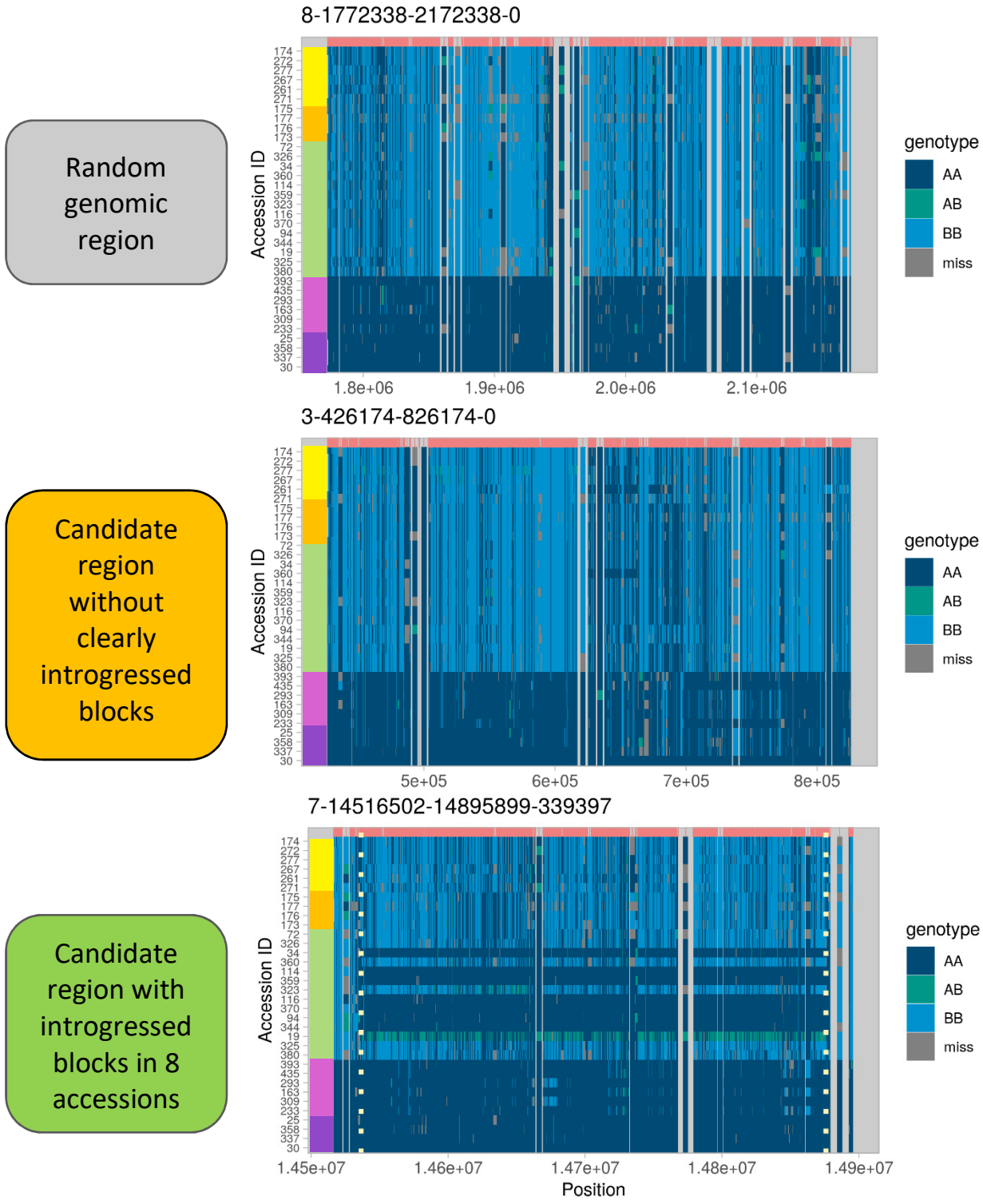

Figure 10: **Examples of genotype composition in genomic regions with/without introgression**
Each plot depicts the genotype composition of all lines in a genomic region. Each SNP is colored according to its genotype. Genotypes are polarized so that accession 30 (*A. nemorensis* Rhine) is always genotype AA. For better visualization of isolated SNPs, SNP width was extended to up to 100th of the region width. Marker positions are indicated by red bars on top of each plot. Colored bars on the left side of the plot indicate the population membership of the accessions: (from top to bottom) yellow – *A. sagittata* Lob, orange – *A. sagittata* Adl-1, green – *A. sagittata* Rhine, violet – *A. nemorensis*

scattered, purple – *A. nemorensis* Rhine. Dashed vertical lines show the border of the introgressed block if it was defined. Otherwise lines were centered. Numbers on top of each plot give the region coordinates in format: chromosome – start – end – width of introgression.

2.5.2 CHARACTERIZATION OF INTROGRESSED BLOCKS

To characterize the genetic relationship between introgressed blocks and orthologous regions in donor accessions, I created haplotype networks of accessions with introgressions and donor accessions, for each introgressed block. I excluded accessions without the introgression to focus on the phylogenetic history of the introgressed haplotype and its donor species. To convert genotypes to haplotypes, I phased all genotypes in the SNP *vcf*-file using Beagle version 5.0 (B. L. Browning, Zhou, & Browning, 2018; S. R. Browning & Browning, 2007) with default options. Beagle automatically imputed missing genotypes, but to avoid mis-inference I reverted the states to missing using a custom python script. I extracted the introgressed block from the phased SNP *vcf*-file using VCFftools (Danecek et al., 2011). I converted the data to *fasta*-format using a custom python script. I created haplotype networks in R using functions from the *pegas* package (Paradis et al., 2016).

3 RESULTS

3.1 CHAPTER 1 – THE ROLE OF NATURAL VARIATION OF STOMATAL TRAITS AND WATER-USE EFFICIENCY IN LOCAL ADAPTATION IN *ARABIDOPSIS THALIANA*

3.1.1 SUBSTANTIAL GENETIC VARIATION IN STOMATA DENSITY AND SIZE

I analyzed over 31,000 images collected in leaves of 330 *A. thaliana* genotypes and observed high levels of genetic variation in stomata patterning. Genotypic means ranged from 87 to 204 stomata/mm² for stomata density and from 95.0 µm² to 135.1 µm² for stomata size. Leaf size was not significantly correlated with stomata density (r=-0.02, p=0.7) and stomata size (r=-0.08, p=0.15), as expected in fully developed leaves. Broad-sense heritability reached 0.41 and 0.30 for stomata size and density, respectively. Mean stomata density and stomata size were negatively correlated (r=-0.51, p<<0.001; Figure 11). Due to the strong correlation between stomata size and density, I focus primarily on stomata size in the following report, but results for stomata density are in the supplemental material.

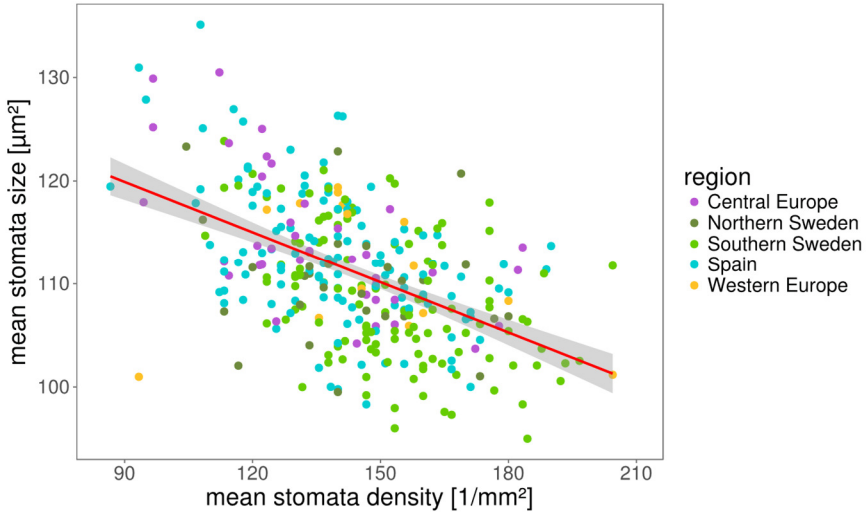

Figure 11: **Stomata density and stomata size were significantly correlated**
Stomata density and size were measured for 330 natural genotypes of *A. thaliana*. The plot shows genotypic means of stomata density and stomata size. Dots are colored based on the geographical origin of each accession. The red line shows a linear fit and gray shadows indicate the error of the fit. Pearson's product-moment correlation r=-0.5, p<0.001.

3.1.2 STOMATA SIZE CORRELATES WITH WATER-USE EFFICIENCY

I expected variation in stomatal traits to influence the trade-off between carbon uptake and transpiration. Thus isotopic carbon discrimination, $\delta^{13}C$, an estimator that increases with water-use efficiency (WUE), was measured (Farquhar, Hubick, Condon, & Richards, 1989; McKay et al., 2008). $\delta^{13}C$ ranged from -38.7‰ to -30.8‰ and was significantly correlated with stomata size (r=-0.18, p=0.004; Figure 12), indicating that accessions with smaller stomata have higher WUE. About ~4% of the total phenotypic variation (i.e. the sum of phenotypic and genetic variance) in $\delta^{13}C$ is explained by genetic variance in stomata size. There was no significant correlation between stomatal density and $\delta^{13}C$ (r=-0.007, p=0.9).

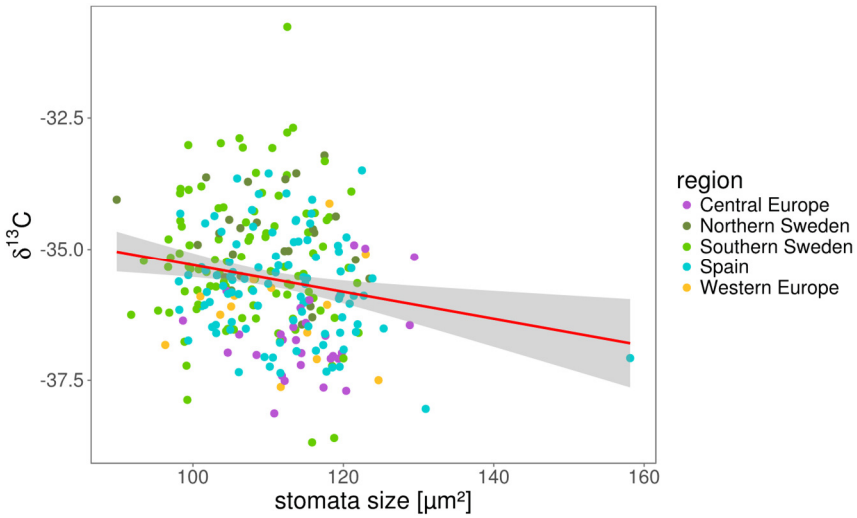

Figure 12: **Stomata size correlated with $\delta^{13}C$**
$\delta^{13}C$ was measured for all plants in block 1. Plots show correlation of stomata size (block 1 only) with $\delta^{13}C$. $\delta^{13}C$ is expressed as ‰ against the Vienna Pee Dee Belemnite (VPDB) standard. The red line shows a linear fit and gray shadows indicate the error of the fit. Pearson's product-moment correlation: r=-0.18, p=0.004. Correlation of $\delta^{13}C$ and stomata size is not only driven by the Spanish outlier (correlation without outlier: r=-0.16, p=0.009). Genetic correlation was calculated using the MTMM approach: r=-0.58, p<0.05.

3.1.3 COMMON GENETIC BASIS OF STOMATA SIZE AND $\delta^{13}C$

To identify the genetic basis of the phenotypic variance I observed, I conducted a genome-wide association study (GWAS) for each phenotype. I calculated for each phenotype a pseudo-heritability, which is the fraction of phenotypic variance explained by the empirically estimated relatedness matrix (e.g. kinship matrix computed on genome-wide SNP typing).

Pseudo-heritability estimates were 0.59 for stomata density, 0.56 for stomata size and 0.69 for $\delta^{13}C$, indicating that differences in stomata patterning and carbon physiology decreased with increasing relatedness. Despite considerable levels of heritability, I did not detect any variant associating with stomata density at a significance above the Bonferroni-corrected p-value of 0.05 (log10(p)=7.78). For stomata size, I detected one QTL with two SNPs significantly associating at positions 8567936 and 8568437 (Figure 13, left) These SNPs had an allele frequency of 1.5% (5 counts) and 2.1% (7 counts), respectively and mapped to gene *AT4G14990.1*, which encodes for a protein annotated with a function in cell differentiation. The former SNP is a synonymous coding mutation while the latter is in an intron.

For $\delta^{13}C$, one genomic region on chromosome 2 position 15094310 exceeded the Bonferroni significance threshold ($\log_{10}(p)=7.97$, Figure 13, right). Allele frequency at this SNP was 9.7% (30 counts) and all accessions carrying this allele, except four, were from Southern Sweden (3 Northern Sweden, 1 Central Europe). Southern Swedish lines carrying the allele showed significantly increased $\delta^{13}C$ compared to the remaining Southern Swedish lines (W=1868, p-value=6.569e-05). A candidate causal mutation is a non-synonymous SNP at position 15109013 in gene *AT2G35970.1*, which codes for a protein belonging to the Late Embryogenesis Abundant (LEA) Hydroxyproline-Rich Glycoprotein family. This SNP also shows elevated association with the phenotype. However, its significance was below the Bonferroni-threshold (log(p)=7). Since this SNP is not in linkage disequilibrium with the highest associating SNP in the region, it is possible that another, independent SNP in this region is causing the association.

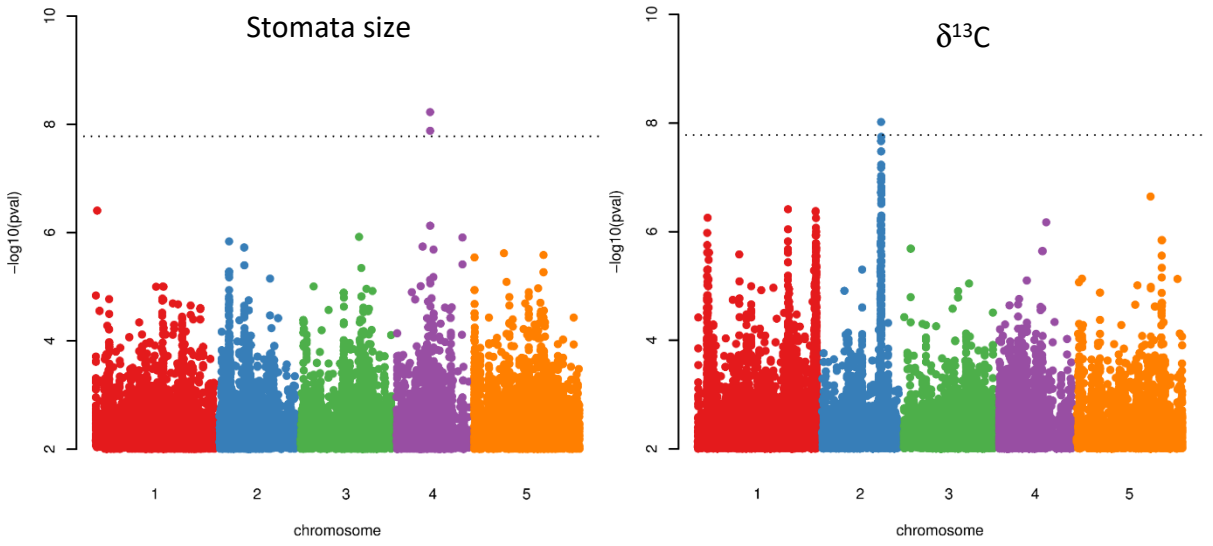

Figure 13: **Significant GWAS associations for stomata size and $\delta^{13}C$**
Manhattan plot of GWAS for mean stomata size with 330 accessions (left) and $\delta^{13}C$ with 310 accessions (right). Each point represents the association p-value of one polymorphism. The dotted horizontal line indicates the Bonferroni significance threshold at type I error rate alpha=0.05. Minor allele count is five.

Multi-Trait Mixed-Model (MTMM) analysis was used to disentangle genetic and environmental determinants of the phenotypic correlations. The significant correlation between stomata density and stomata size (r=-0.5) had no genetic basis, but had a significant (r=-0.9, p<0.05) residual correlation. This suggests that the correlation was not determined by common loci controlling the two traits, but by other, perhaps physical, constraints or by epistatic alleles at distinct loci. By contrast, the correlation between stomata size and $\delta^{13}C$ (r= -0.18) had a significant genetic basis (kinship-based correlation, r=-0.58, p<0.05). Thus, in contrast to the phenotypic variation, genetic variation in stomata size roughly explains over 33% of the genetic variation in $\delta^{13}C$.

To further investigate the genetic basis for the correlation between stomata size and $\delta^{13}C$, MTMM GWAS was performed, which tests three models: the first model tests whether a SNP has the same effect on both traits; the second model tests whether a SNP has differing effects on both traits and the third model is a combination of the first two to identify SNPs which have effects of different magnitude on the traits (Korte et al., 2012). No variants with same or differing effects on $\delta^{13}C$ and stomata size were observed. However, with the combined model, a marginally significant association on chromosome 4, which affected $\delta^{13}C$ but not stomata

size, was observed. GWAS of $\delta^{13}C$ restricted to the 261 individuals used for the MTMM analysis confirmed the QTL on chromosome 4. GWAS applied to different but overlapping sets of accessions yield similar results but can sometimes differ in the set of significant associations, since marginal changes in SNP frequency can affect significance levels (Figure 4). Indeed, the p-values of associations with $\delta^{13}C$ for the two datasets (310 and 261 accessions) were highly correlated (r=0.87, p<<0.0001). In this set of genotypes, two SNPS, at position 7083610 and 7083612, exceeded the Bonferroni-corrected significance threshold (α=0.05) (both p=4.8e-09, Figure 14) although they were under the significance threshold in the larger dataset. Allele frequency is 14% (37 counts) at these two loci and explains 11% of the phenotypic variation. The association is probably due to complex haplotype differences since it coincides with a polymorphic deletion and contains several imputed SNPs. Thirty-five of the 37 accessions carrying the minor allele originated from Southern Sweden and showed significantly higher $\delta^{13}C$ compared to other Southern Swedish accessions (mean difference=1.34; W=1707, p=1.15e-06). In summary, two genetic variants significantly associating with $\delta^{13}C$ independent of stomata size were detected despite the common genetic basis of the two traits.

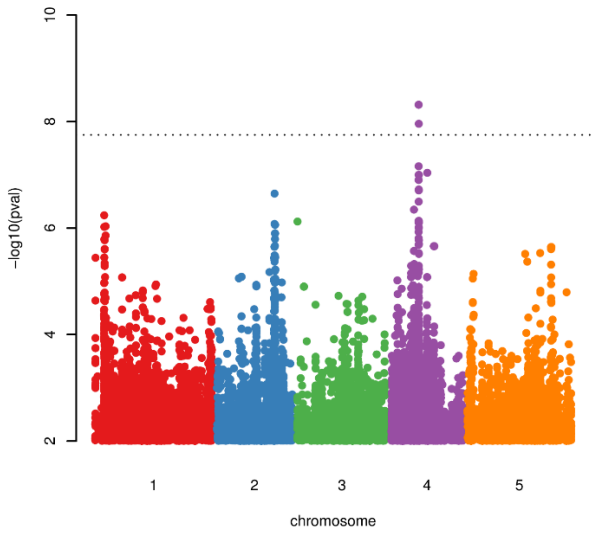

Figure 14: **Significant MTMM GWAS association for $\delta^{13}C$ independent of stomata size**
Manhattan plot of GWAS for $\delta^{13}C$ with the 261 accessions used in MTMM analysis. Each point represents the association p-value of one polymorphism. The dotted horizontal line indicates the Bonferroni significance threshold at type I error rate alpha=0.05. Minor allele count is five.

3.1.4 STOMATA SIZE AND STOMATA DENSITY CORRELATE WITH GEOGRAPHICAL PATTERNS OF CLIMATIC VARIATION

I used PCA to describe multivariate variation in climatic conditions reported for the locations of origins of the genotypes. I tested the correlation of each measured phenotype with climatic principal components (PCs) using a GLM which accounted for genetic population structure (see methods). I found a significant, negative relationship between genetic variation in stomata size and climatic PC2 (Likelihood ratio test Chi-Square (LRT X^2) =9.2784, degrees of freedom (df)=1, p=0.005) and PC5 (LRT X^2=5.7335, df=1, p=0.02, Figure 15). Climatic PC 2 explained 23.8% of climatic variation and had the strongest loadings (both negative) from temperature and water vapor pressure (humidity). Climatic PC 5 explained 9% of the climatic variation and mostly increased with increasing spring-summer drought probability ratio and increasing solar radiation.

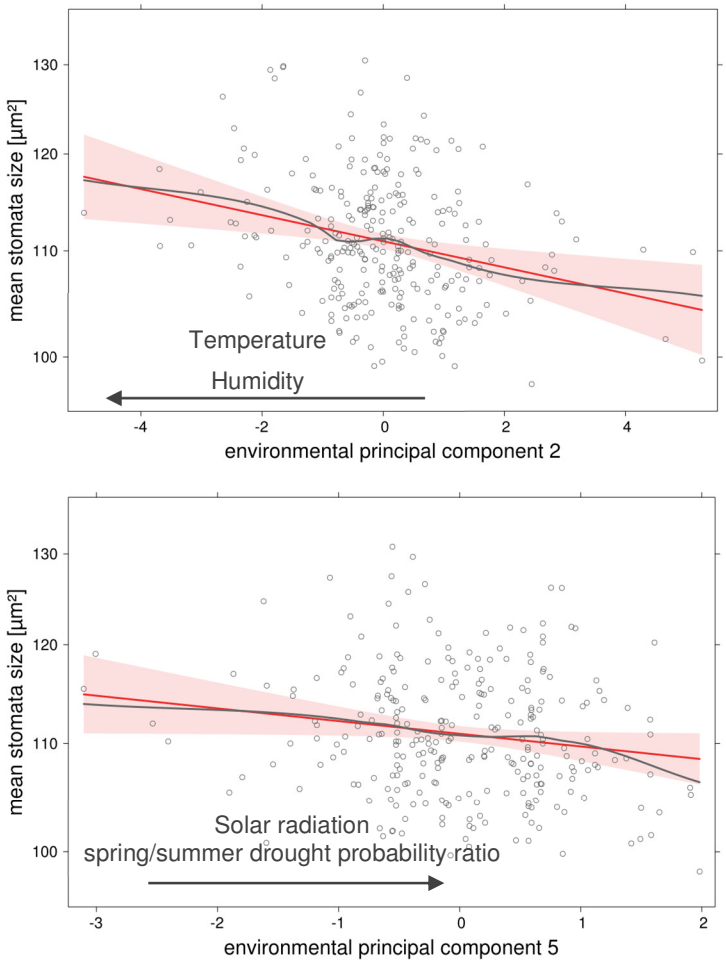

Figure 15: **Significant correlation between stomata size and climatic variables**

Correlation between stomata patterns and seven climatic principal components (PCs) was tested for each phenotype using a Generalized Linear Model (GLM) including genetic population structure as described by the 20 first genetic PCs. Plots are effect plots based on the GLM (see methods), showing the correlation between stomata size two climatic PCs. Black arrows indicate correlation with the climatic variables showing the strongest loadings for the respective PC. Plots show the linear fit (red solid line) and the smoothed fit of partial residuals (gray) of the specific predictor. Gray dots are partial residuals. The red shade shows the error of the linear fit. Both PCs shown here are significant predictors of the respective response variable ($p<0.05$).

I also found significant climatic predictors for the distribution of genetic variation in stomata density (PC 2: LRT X^2=8.6612, df=1 p=0.003; PC 5: LRT X^2=7.3773, df=1, p=0.007; PC 7: LRT X^2=6.6033, df=1, p=0.01). $\delta^{13}C$ did not correlate with any of the climatic PCs. However, removing population structure covariates from the model revealed significant correlations of $\delta^{13}C$ with climatic PC2 (+, LRT X^2=7.3564, df=1, p=0.006), PC3 (-, LRT X^2=3.8889, df=1, p=0.048) and PC4 (+, LRT X^2=6.6885, df=1, p= 0.009) (Figure 16). PC3 explained 13.7% of climatic variation and principally increased with rainfall and decreased with spring-summer drought probability ratio. PC4 explained 11.4% of the total variation and mostly increased with wind speed. Therefore, the covariation of $\delta^{13}C$ with climatic parameters describing variation in water availability and evaporation in *A. thaliana* is strong but confounded with the demographic history of the species. To test whether alleles associating with increased $\delta^{13}C$ in GWAS are involved in adaptation to local climate, I checked whether any climatic PC is a significant predictor of the allelic state of Southern Swedish accessions. However, none of the climatic PCs was a significant predictor for one of the two loci.

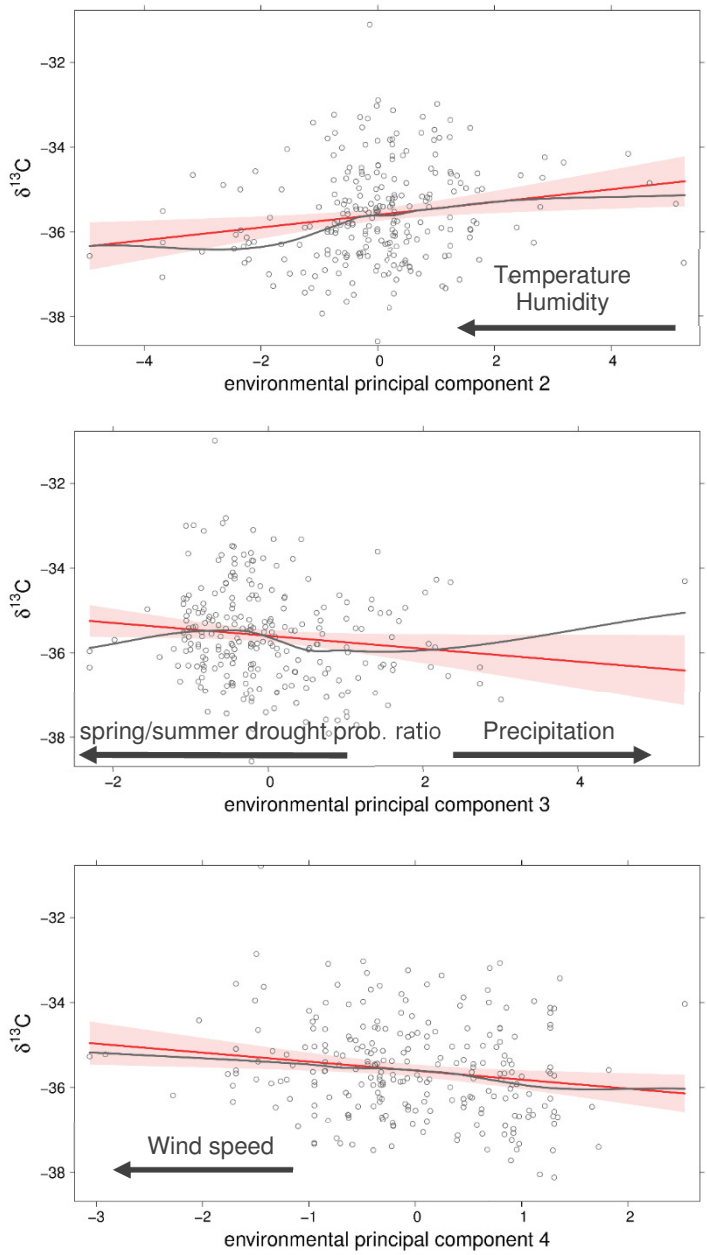

Figure 16: **Significant correlation between $\delta^{13}C$ and climatic variables without population structure correction**

Correlation between $\delta^{13}C$ and climatic principal components (PCs) was tested using a generalized linear model excluding genetic population structure. Plots are effect plots based on the model (see methods), showing the correlation between $\delta^{13}C$ and three PCs. Black arrows indicate correlation with the climatic variables showing the strongest loadings for the respective principal component Plots show the linear fit (red solid line) and the smoothed fit of partial residuals (grey) of the specific predictor. Grey dots are partial residuals. The red shade shows the error of the linear fit. All three principal components shown here are significant predictors of the respective response variable ($p<0.05$).

3.1.5 PATTERNS OF REGIONAL DIFFERENTIATION DEPART FROM NEUTRAL EXPECTATIONS

I divided genotypes into five regions based on genetic clustering (Alonso-Blanco et al., 2016) and their geographic origin (Figure 1, see Methods). I detected significant phenotypic differentiation among these regions for stomata size (LRT X^2=52.852, df=4, p=9.151e-11, Figure 17). Stomata size was significantly lower in Southern Sweden (mean=108 µm²) compared to Central Europe (mean=114µm², Generalized Linear Hypothesis Test (GLHT) z=-6.24, p<0.001), Western Europe (mean=111 µm², GLHT z=2.769, p=0.04) and Spain (mean=113 µm², GLHT z=6.709, p<0.001), which did not significantly differ from each other. Northern Sweden showed an intermediate phenotype and did not differ significantly from any region (mean=110 µm²). Variation for stomata density, showed a similar but inverted pattern.

Furthermore, we found significant regional differentiation in $\delta^{13}C$ measurements (LR X^2 =58.029, df=4 p=7.525e-12, Figure 17). Highest $\delta^{13}C$ levels (highest WUE) were found in accessions from Northern Sweden (mean=-34.8) and Southern Sweden (mean=-35.2), which were significantly higher than in accessions from Spain (mean=-35.7; GLHT Southern Sweden z=-3.472, p=0.008; GLHT North Sweden z=-3.49, p=0.001) and Western Europe (mean=-36.06; GLHT Southern Sweden z= -2.8, p=0.03; GLHT Northern Sweden z=-3.28, p=0.008). Lowest $\delta^{13}C$ levels were found in lines from Central Europe (mean=-36.6), which were significantly lower than in lines from Northern Sweden (GLHT z=5.676, p<0.001), Southern Sweden (GLHT z=6.992, p<0.001) and Spain (GLHT z=3.714, p=0.002).

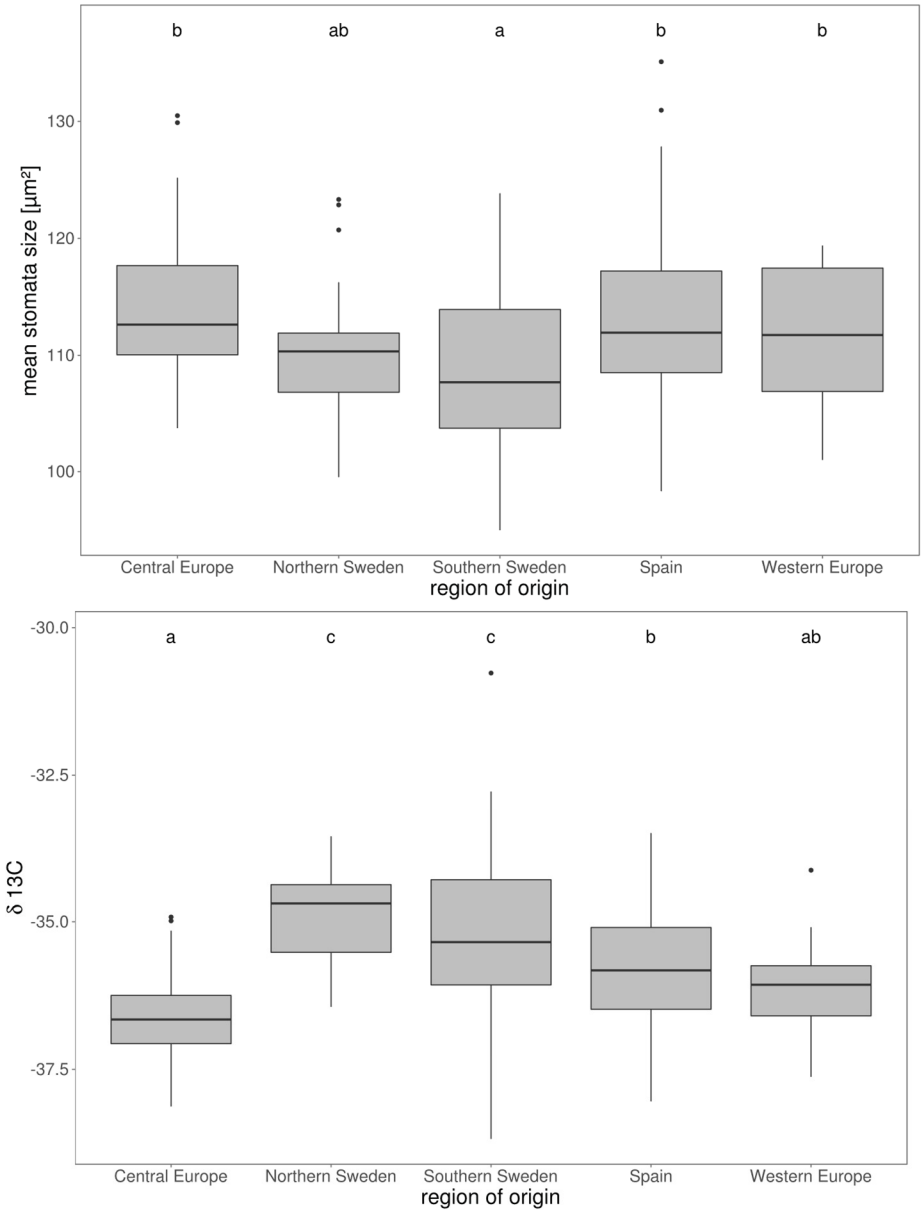

Figure 17: **Significant regional differentiation for stomata size and $\delta^{13}C$**
A. thaliana accessions were grouped based on their geographical origin. Boxplots show regional differentiation of stomata size (top) and $\delta^{13}C$ (bottom). Significance of differentiation was tested using Generalized Linear Models followed by a post-hoc test. Statistical significance is indicated by letters on top: Groups that do not share a common letter are significantly different. Significance levels: top) a-c, a-bc: p<0.001; ab-c: p<0.05; bottom) a-c, a-b: p<0.001, b-c: p<0.01, ab-c: p<0.05.

The observed regional differences result either from the demographic history of the regions or from the action of local selective forces. To tease these possibilities apart, the phenotypic differentiation (Q_{ST}) can be compared to nucleotide differentiation (F_{ST}) (Kronholm et al., 2012; T. Leinonen et al., 2013). I examined each pair of regions separately, since they are not equidistant from each other and calculated F_{ST} distributions for over 70,000 SNP markers (spaced at least 1kb apart, see methods). For each trait, Q_{ST} exceeded the 95th percentile of

the F_{ST} distribution in at least two pairs of regions (Table 7 A-C). I used permutations to calculate a significance threshold for the Q_{ST}/F_{ST} difference (see methods). Significant regional differentiation was pervasive in the sample, with Central Europe and Southern Sweden being significantly differentiated for all three phenotypes. This analysis suggests that natural selection has contributed to shape the phenotypic differentiation between regions.

Regional differences in climate may have imposed divergence in stomatal patterning. Thus, I estimated climatic distances between regions using estimates of regional effects extracted from a MANOVA. I did not observe significant correlations between adaptive phenotypic divergence (Q_{ST}-F_{ST}) and the climatic distance of the respective regions (Mantel test $p>0.05$ for each of the three traits). Regional divergence in δC^{13}, stomata density and stomata size was therefore not proportional to climatic divergence.

Table 7: **Patterns of regional differentiation depart from neutral expectations**
Pairwise Q_{ST} estimates were derived from linear mixed models for all regions. Genome-wide, pairwise F_{ST} distribution was calculated based on 70,000 SNPs for all regions. In the top half of each table, the difference Q_{ST}-F_{ST} for each pair of regions is shown. In the bottom half of each table the Q_{ST} estimate for each pair of regions is shown. Each table represents one phenotype as indicated by table headlines. Significant Q_{ST}-F_{ST} differences are written in bold. The significance threshold is based on the 95th percentile of a distribution of maximum Q_{ST}-F_{ST} values from 1000 random permutations of phenotypic data.

A) Stomata size

Q_{ST} \ Q_{ST}-F_{ST}	Cent. Europe	North. Sweden	South. Sweden	Spain	West. Europe
Central Europe		-0.32	**0.29**	-0.17	-0.13
Northern Sweden	0.15		-0.31	-0.38	-0.51
Southern Sweden	0.41	0.09		**0.12**	-0.03
Spain	0.01	0.06	0.32		-0.18
West. Europe	0.02	<0.01	0.21	<0.01	

B) Stomata density

Q_{ST} \ Q_{ST}-F_{ST}	Cent. Europe	North. Sweden	South. Sweden	Spain	West. Europe
Central Europe		-0.37	**0.31**	-0.16	**0.17**
Northern Sweden	0.09		-0.24	-0.44	-0.49
Southern Sweden	0.44	0.16		**0.17**	-0.19
Spain	0.01	0.01	0.36		0.07
West. Europe	0.32	0.02	<0.01	0.26	

c) δC^{13}

Q_{ST} \ Q_{ST}-F_{ST}	Cent. Europe	North. Sweden	South. Sweden	Spain	West. Europe
Central Europe		**0.21**	**0.28**	**0.13**	-0.07
Northern Sweden	0.7		-0.40	-0.16	-0.01
Southern Sweden	0.4	0.01		-0.08	-0.01
Spain	0.30	0.28	0.11		-0.12
West. Europe	0.07	0.40	0.17	0.05	

3.2 CHAPTER 2 – THE EFFECT OF ECOLOGICAL RESTORATION ON THE GENETIC DIVERSITY OF TWO FLOODPLAIN *ARABIS* SPECIES

3.2.1 RAD-SEQUENCING UNCOVERS TWO HYBRIDIZING SPECIES

Unless otherwise stated, all described results were obtained using the reference-based pipeline. I genotyped 134 individuals from 10 sites – 4 pristine and 6 restored (Figure 7) – yielding 3.6 Mb of sequence of which 32,880 single nucleotide positions were polymorphic (SNPs). Only 20% of SNPs were in coding regions, 40% and 56% of which were synonymous and non-synonymous, respectively (4% unassigned). To visualize patterns of genetic diversity across sites, I conducted principal component analysis (PCA) for all individuals (Figure 18, top). Almost 80% of the total genetic variation was explained by the first principal component, which separated most individuals into two clearly defined clusters, likely representing taxonomic units.

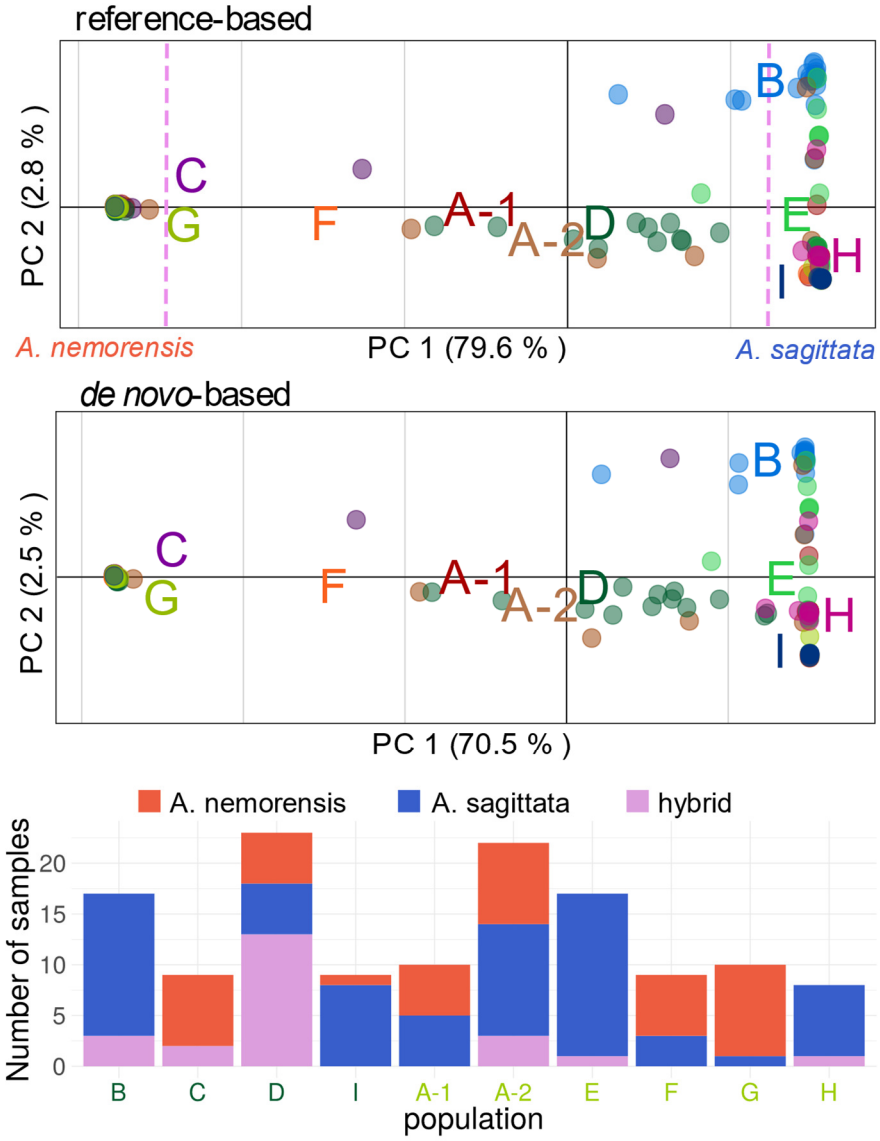

Figure 18: **Presence of two hybridizing *Arabis* species in study sites**
Top) Plot of the first two principal components (PC) of genetic variation based on the reference-based pipeline. Text labels and colors distinguish the different sites. Text labels are approximately in the centroid of the respective site (with offset to avoid overlaps). Values in brackets in the axis labels show the amount of variance explained by each PC. Note that the first PC explains about 80% of the total variation and splits most individuals into two clearly defined clusters. The clusters on the left and on the right end of PC1 correspond to two distinct species, identified here as *Arabis nemorensis* and *Arabis sagittata*. Species identity is indicated by text labels below the plot. Points in between the two clusters are likely natural interspecific hybrids. The dashed lines show the thresholds used for distinction between species and hybrids throughout this study. Middle) Plot showing the same analysis as above but based on the de-novo-pipeline. Bottom) Stacked barplot showing the distribution of species among the study sites. Population labels are colored by type: darkgreen = pristine, lightgreen = restored. Species are distributed heterogeneously among sites.

Based on phylogenetic analysis of the ITS region of 10 individuals (Table 8), I identified one taxonomic unit as *Arabis nemorensis.* The ITS region of individuals from the other taxonomic unit was highly similar to *A. sagittata* and *Arabis hirsuta*. Both are sibling species from the *A.*

hirsuta complex, but *A. hirsuta* is tetraploid and *A. sagittata* diploid. Since all twelve tested samples were diploid (Table 9), we conclude that the second cluster most likely corresponds to *A. sagittata*. Average genetic distance (d_{XY}) between *A. nemorensis* and *A. sagittata* was 8.1e-03, i.e. 8 fixed SNP differences for 1000 bp.

Table 8: **Results of the phylogenetic analysis of the ITS sequences**
For each genotype (ID) the closest species match of the ITS sequence is shown. Additionally, the position on PC1 of the PCA of genetic variation is shown to identify the corresponding cluster for each species. The two clusters match uniquely to the two species.

ID	ITS match	PC 1
30	*A. nemorensis*	319.89
63	*A. sagittata*	-178.23
2	*A. sagittata*	-181.67
5	*A. sagittata*	-180.34
15	*A. sagittata*	-177.64
19	*A. sagittata*	-176.14
29	*A. nemorensis*	332.1
36	*A. nemorensis*	325.64
32	*A. sagittata*	-55.95

Table 9: **Results of ploidy analysis**
For each genotype (ID) the level of ploidy is shown. Additionally, the putative species, as determined by ITS analysis and PCA is shown.

ID	put. species	ploidy
29	*A. nemorensis*	diploid
94	*A. sagittata*	diploid
104	*A. nemorensis*	diploid
114	*A. sagittata*	diploid
328	*A. sagittata*	diploid
336	hybrid	diploid

350	*A. sagittata*	diploid
364	*A. sagittata*	diploid
376	*A. sagittata*	diploid
382	*A. sagittata*	diploid
333	hybrid	diploid
375	hybrid	diploid

Twenty-three individuals showed a positioning along the first PC that was intermediate between the two clusters, suggesting they were interspecific hybrids. Most of these hybrids were closer to *A. sagittata* on the first PC. Since sterile F1-hybrids should be located exactly in the middle between the two species, this suggests that they are somewhat fertile and preferentially back-cross with *A. sagittata.* Laboratory observations have shown that hybrids are indeed fertile. Additionally, in an ADMIXTURE analysis of all samples with K=2, most hybrids showed ancestry over 50% from *A. sagittata*, as expected from preferential back-crossing to this parent (Figure 19).

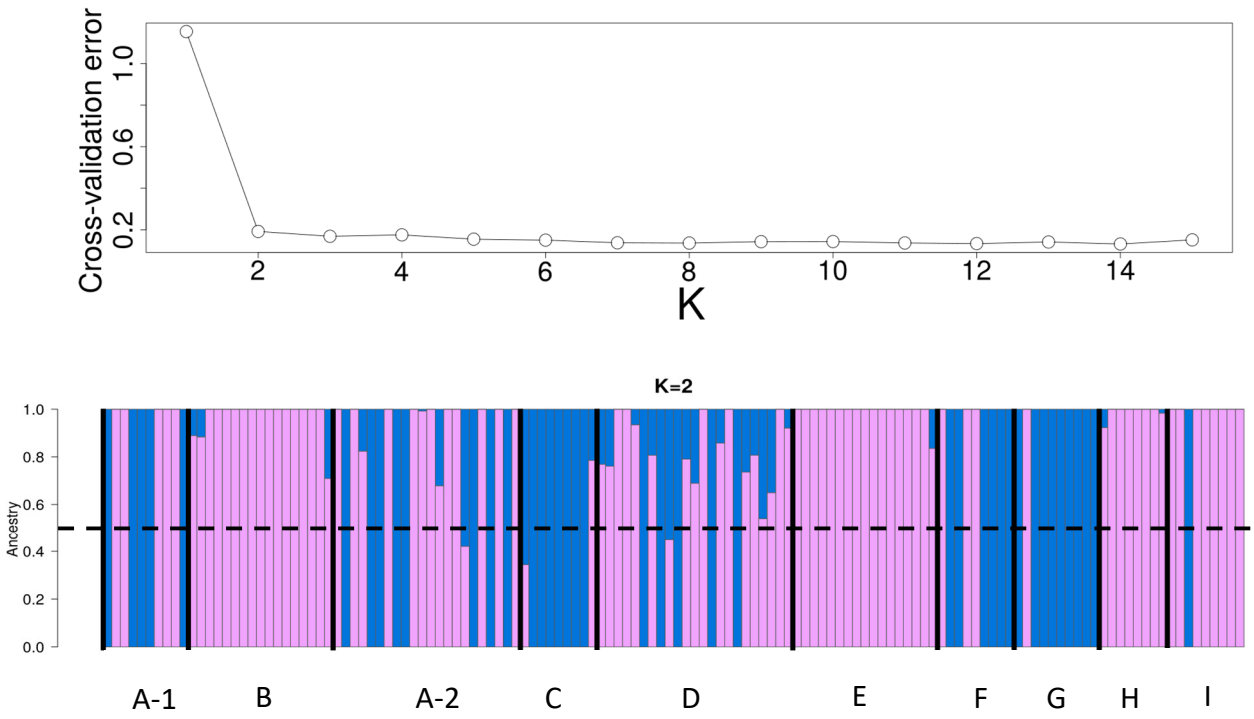

Figure 19: **ADMIXTURE analysis confirms presence of two taxonomic groups**
Top) Cross-validation error of ADMIXTURE analysis for different values of K. Bottom) ADMIXTURE plot for all individuals including hybrids for K=2. Blue represents the *A. nemorensis* cluster and pink the *A. sagittata* cluster. Individuals are ordered by site, sites are separated by black lines and site labels are below the plot. The dashed black line is approximately at ancestry level 0.5. This shows that most hybrids have *A. sagittata* ancestry greater than 0.5, indicating preferential back-crossing to this parent.

Overall the sample was composed of 31% *A. nemorensis*, 52% *A. sagittata* and 17% hybrids. The species composition differed among sites (Figure 18, bottom). *A. nemorensis* was present in 7 sites (3 pristine, 4 restored), *A. sagittata* in 9 sites (3 pristine, 6 restored) and hybrids in 6 sites (3 pristine, 3 restored). Notably, the pristine site D was dominated by hybrids with over 56%.

3.2.2 NO REDUCTION OF GENETIC DIVERSITY IN RESTORED SITES

I computed species-specific estimates of genetic diversity within each site, excluding hybrid genotypes. The *A. nemorensis* dataset consisted of 2746 SNPs. Levels of genetic diversity (π) varied up to two-fold among sites, ranging from 6.6e-05 in A-2 to 1.4e-04 in A-1 (Figure 20, left). Total diversity was 1.5e-04. However, pristine and restored sites did not differ significantly in their level of diversity (mean difference= +10% in restored; W = 4, p = 1).

The *A. sagittata* dataset consisted of 6,366 SNPs. Total genetic diversity in *A. sagittata* was about three times as high as in *A. nemorensis*. Yet, in contrast to *A. nemorensis*, genetic diversity differed strongly among sites, ranging from 1.03e-06 in site I to 4.26e-04 in site E (Figure 20, right)Notably, genetic diversity was low in two of three pristine sites. In contrast, levels of diversity were high in all restored sites, except site F. However, the overall difference between pristine and restored sites was not significant (mean difference= +163% in restored sites; W=13, p=0.14).

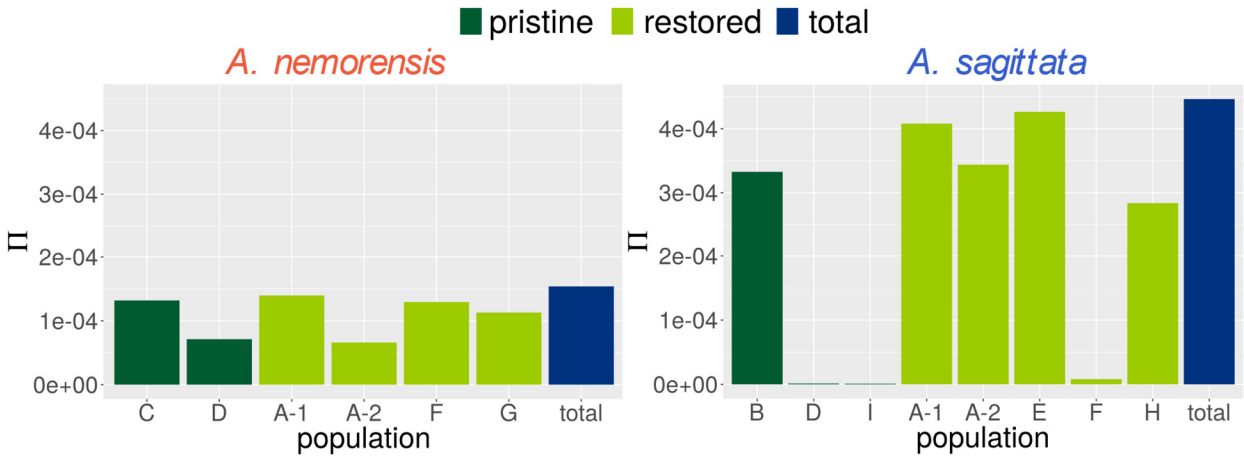

Figure 20: **No reduction in genetic diversity in restored sites**
Barplot of average pairwise genetic diversity (π) within each site of each species. Bar-color indicates the type of site.

Since hybridization potentially enables gene flow between the two species we also compared levels of genetic diversity for both species combined, including the hybrids. Overall genetic diversity increased by an order of magnitude and mixed sites were more diverse than mostly pure sites, as would be expected (Figure 21). Again, restored sites did not show significantly different levels of genetic diversity from pristine sites (mean difference= +22%; W=13, p=0.91).

Figure 21: **No reduction in genetic diversity in restored sites in the combined dataset**
Barplot of average pairwise genetic diversity (π) within each site of both species combined including hybrids. Bar-color indicates the type of site.

3.2.3 RESTORATION REDUCES POPULATION STRUCTURE AND FACILITATES RECOMBINATION

To quantify the degree of population structure, I estimated genetic distance and differentiation (F_{ST}) among all pairs of sites. Genetic distance among *A. nemorensis* sites ranged from 0.03 to 0.31 and F_{ST} estimates from 0 to 0.5 (Figure 22, A+C). Population structure was slightly more pronounced among pristine sites than restored sites: Mean genetic distance was 0.26 among pristine sites and 0.13 among restored sites (W=17, p=0.047); mean F_{ST} was 0.37 among pristine sites and 0.25 among restored sites (W=14, p=0.26).

In *A. sagittata*, genetic distance ranged from 0.32 to 0.34 and F_{ST} estimates from 0 to 0.91 (Figure 22, B+D). The difference in population structure between pristine and restored sites was stronger than in *A. nemorensis*: mean genetic distance was 0.33 among pristine sites and 0.12 among restored sites (W=45, p=0.002); mean F_{ST} was 0.81 among pristine sites and 0.2 among restored sites (W=43, p=0.015).

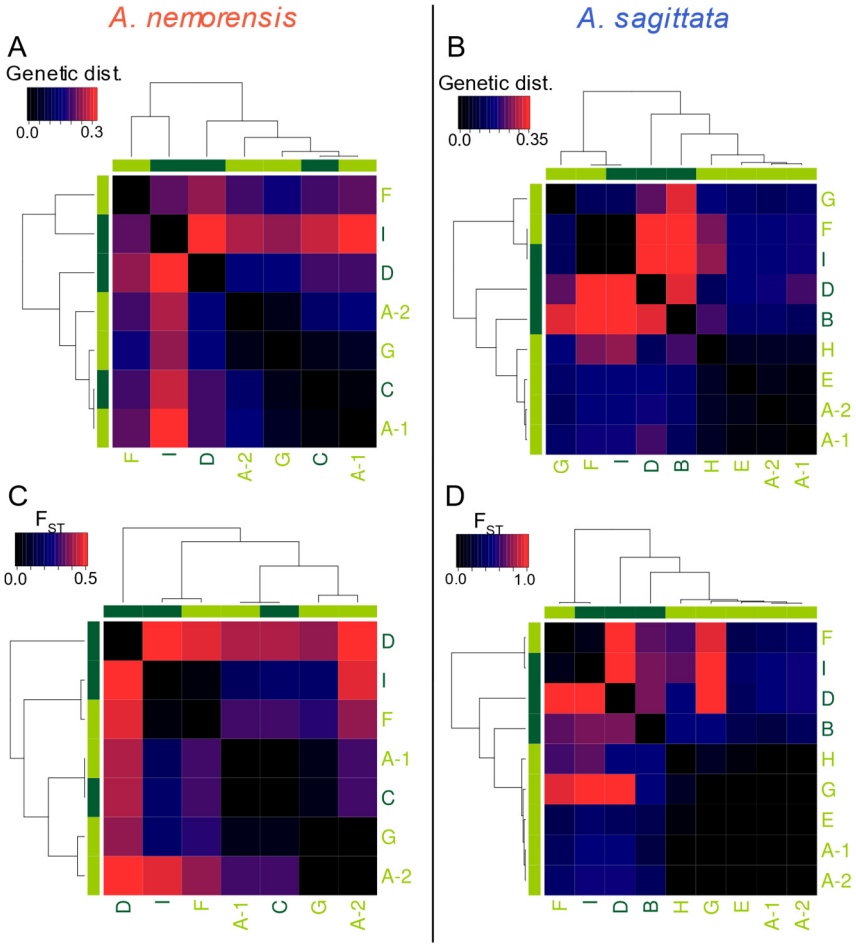

Figure 22: **Overall strong population structure with signatures of admixture in restored sites**
This panel shows different measurements of population structure of each species. A-D) Heatmaps showing pairwise comparisons between all sites of genetic distance (Cavalli-Sforza chord distance) and FST, respectively. Values of the respective variable are indicated by tile color, corresponding to the color scale in the top-left corner of each heatmap. Sites were clustered based on the respective variable, as indicated by the dendrograms. Colors of text labels and bars next to the grid indicate the type of the site: pristine – dark green, restored – light green.

Reduced differentiation among restored sites suggests that genetic material of pristine sites was mixed in restored sites. To visualize this in more detail, I conducted ADMIXTURE analysis for each species. This analysis assumes a given number (K) of genetic clusters (ancestral populations) and assigns cluster ancestry proportions to each individual, allowing for mixed ancestry. For *A. sagittata*, K=6 was estimated as the optimal value (Figure 23). Up to K=4, the three pristine sites consisted of three distinct genetic clusters, which were mixed in all but one of the restored sites. For higher values of K, pristine population B consisted in two clusters, with pure and admixed individuals in equal proportions, in agreement with the higher genetic diversity observed at this site. All individuals from other pristine sites had pure ancestry from a single cluster per population.

For *A. nemorensis*, K=7 was estimated as the optimal value for clustering (Figure 23). In contrast to *A. sagittata*, we found individuals with different or mixed ancestry within populations of *A. nemorensis*, even at low values of K, indicating more genetic mixture than in *A. sagittata*. Since genetic structure was less pronounced for *A. nemorensis*, the restoration procedure did not impact the genetic distribution of this species, by contrast with *A. sagittata*.

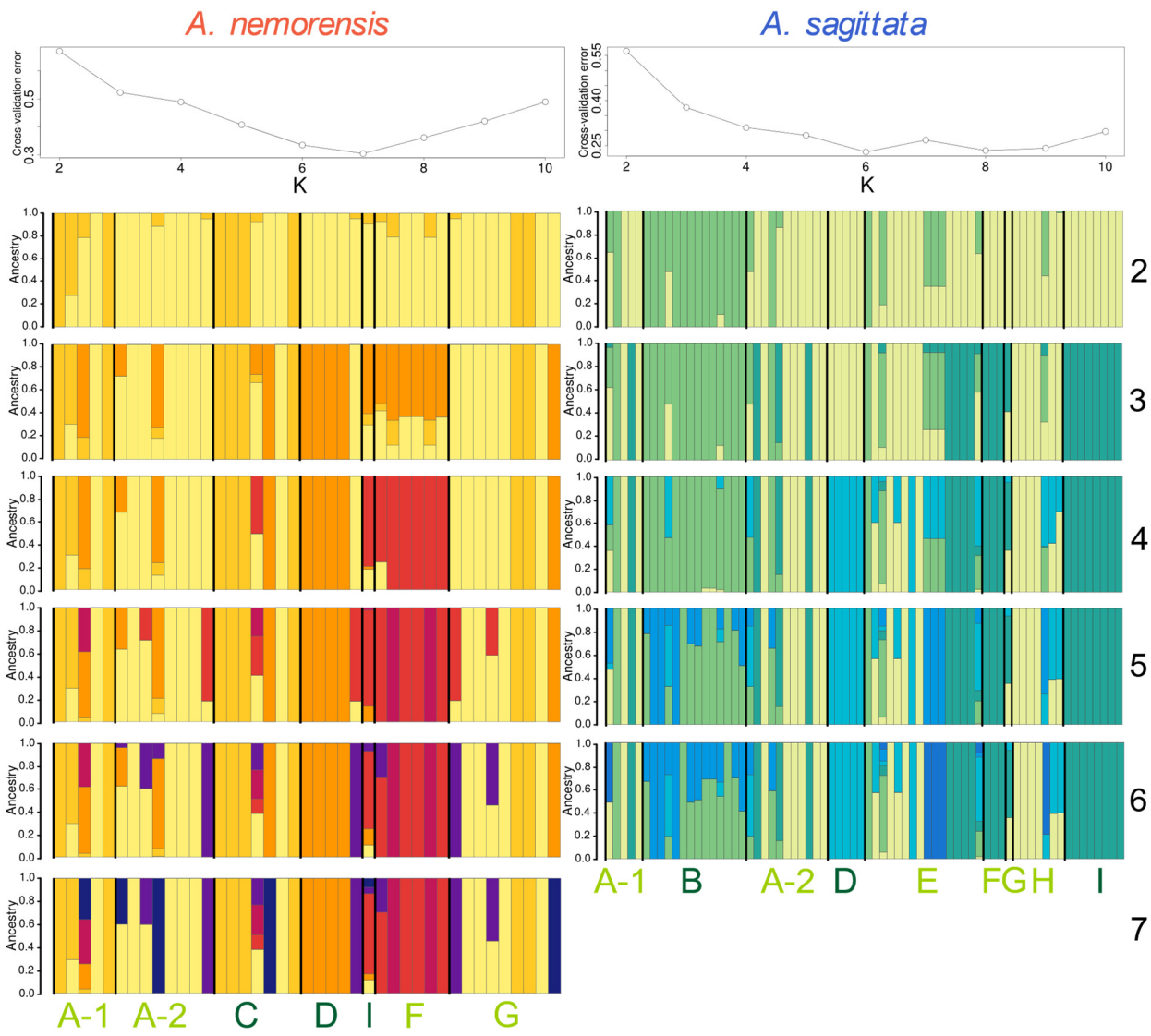

Figure 23: **Increased impact of genetic admixture of pristine sites in *A. sagittata***
The figures shows a comparison of ADMIXTURE results for both species. The top panels show the cross-validation (CV) error for all tested values of K. The following panels show the ancestry proportions of each cluster for each individual for all values of K until the first minimum of the CV-error curve. The numbers on the right represent the value of K. Site labels are colored according to the type of the site: light-green = restored, dark-green = pristine.

3.2.4 *DE-NOVO*- AND REFERENCE-BASED SUMMARY STATISTICS ARE HIGHLY CORRELATED

Finally, I tested whether a reference genome is required to determine the impact of restoration on genetic diversity by comparing results from a reference-based and a *de novo* pipeline. I found that estimates of genetic diversity, genetic distance or F_{ST} yielded by the two methods were highly correlated, with all correlation coefficients being greater than 0.95 (maximum $p < 0.001$, Figure 24). However, estimates of genetic diversity generated without a reference genome were deflated, especially for high diversity sites (Figure 24), by a median factor of 0.83 for *A. nemorensis* and 0.36 for *A. sagittata*. Estimates of genetic distance were deflated by a median factor of 0.61 in both species. Interestingly, however, both pipelines yielded almost identical estimates of F_{ST} (median factor of 0.93). Moreover, the presence of the two species and their hybrids was detected with both methods (Figure 18, middle). These results indicate that, in this study, the use of a reference genome was not required to determine the impact of restoration on genetic diversity.

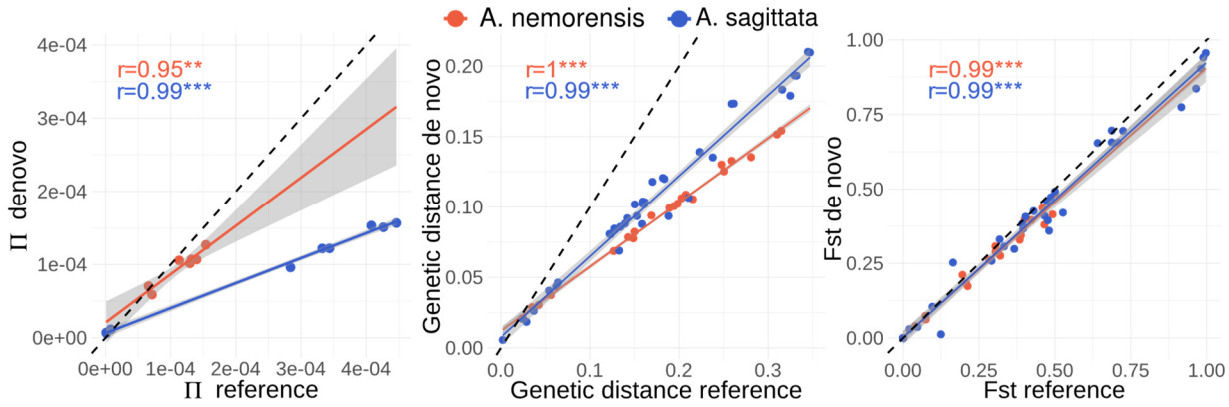

Figure 24: **De-novo- and reference-based summary statistics are highly correlated.**
Plots show the correlation of summary statistics based on the reference-based and the de-novo-based pipeline. Line and dot colors represent the species. Each dot represents one individual. Lines represent a linear fit through the dots and the gray shadows indicate the error of the fit. The dashed line represents a hypothetical 1:1 relationship of the variables. Correlation-coefficients (Pearson's product-moment correlation) are indicated as colored text corresponding to the species. Stars indicate the level of the coefficient: **p<0.01, ***p<0.001.

3.3 CHAPTER 3 – PATTERNS OF GENETIC DIVERSITY IN *A. NEMORENSIS*

3.3.1 PRESENCE THREE DIFFERENT *ARABIS* SPECIES IN THE COLLECTION

I genotyped 1.45 Mb (0.8 % of the genome) for each of 292 individuals. Overall, I identified 58,306 single nucleotide polymorphisms (SNPs). To obtain an overview of the genetic variation I performed PCA. The first four principal components explained 50 %, 15.4%, 10.8% and 2.7 % of the genetic variation, respectively. Most individuals clustered into three distinct groups (Figure 25). I previously identified two of these groups as *A. nemorensis* and *A. sagittata* (Chapter 2). Using the ITS phylogeny and cytometric information, I identified the third group as the tetraploid *Arabis hirsuta*. *A. nemorensis* x *A. sagitatta* hybrids from the Rhine populations were again located between the corresponding clusters on the PCA, as expected. Interestingly, two individuals of the Nvw population were located *between A. hirsuta* and *A. nemorensis*, suggesting another type of interspecific hybrid. However, from these individuals I could only collect leaf material, as the siliques appeared to be empty, suggesting sterility. Additionally, the species description of the herbarium samples did not match the genetic clustering, suggesting misidentification by the collector/curator. All putative *A. sagittata* samples clustered with *A. hirsuta* (one sample, sag-6, was slightly separated from the *A. hirsuta* cluster). One herbarium sample annotated as *A. hirsuta* (hir-1) in fact clustered with the *A. sagittata* genotypes. Interestingly, sag-6 and hir-1 are strong outliers on the third and fourth PC, respectively. Both samples were collected in France.

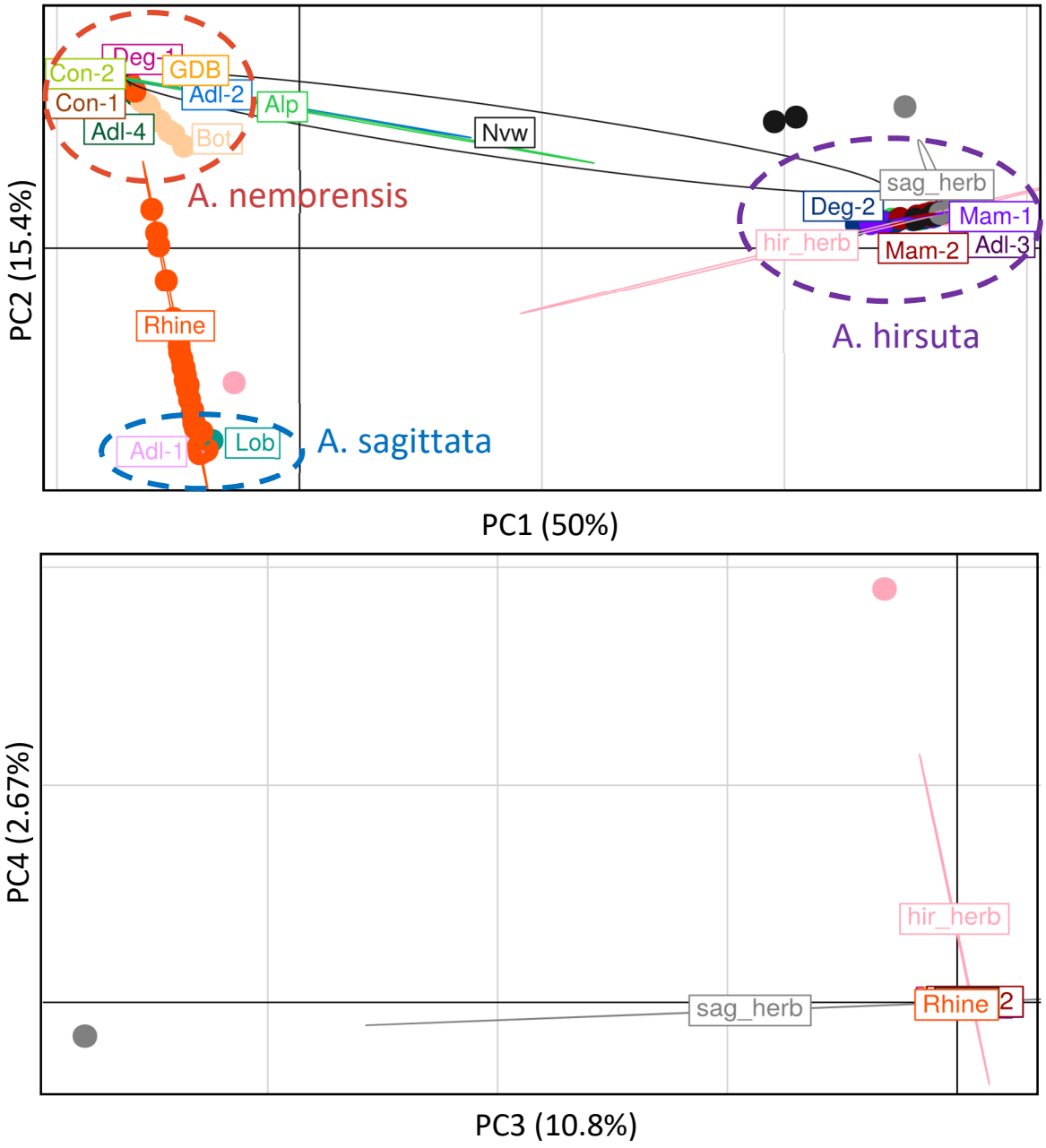

Figure 25: **PCA identifies three highly distinct genetic clusters corresponding to different species** Plots showing the first four principle components. Each individual is represented by a dot, colored by its population of origin. Population labels are located near the centroid of each population, but manual offsets were added to avoid overlaps. The values in brackets represent the % variance explained by each principle component. Phylogenetic analysis of the ITS sequence identified each cluster as a separate species. Species are marked by dashed circles.

Overall, the sample (excluding herbaria) consisted of 121 individuals of *A. nemorensis*, 95 of *A. sagittata*, 42 of *A. hirsuta*, 24 of *A. sagittata* x *A. nemorensis* hybrids and two *A. hirsuta* x *A. nemorensis* hybrids. To confirm the PCA results and visualize the geographic distribution of species I performed ADMIXTURE analysis. Cross-validation confirmed K=3 as the 'optimal'

number of clusters and the three clusters correspond to the three identified species. Most populations consisted of only one species (Figure 26). Four populations were pure *A. hirsuta*, two were pure *A. sagittata* and six were pure *A. nemorensis*. In three populations *A. nemorensis* co-occurred with *A. hirsuta*. *A. sagittata* co-occurred with *A. nemorensis* only in the previously analyzed Upper Rhine population.

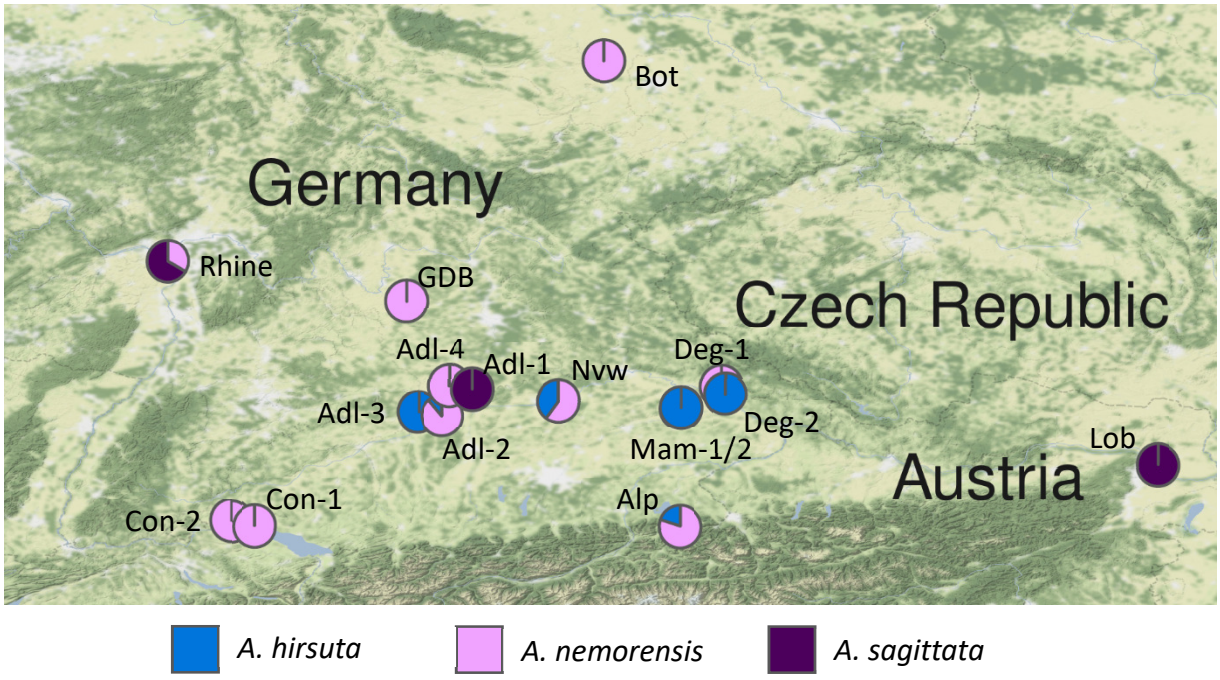

Figure 26: **Geographic distribution of the three *Arabis* species**
Each pie-chart represents a single population and shows the mean values of genetic ancestry from ADMIXTURE analysis. Populations Mam-1 and Mam-2 completely overlapped, but the pie charts are identical.

The presence of three species in my sample allowed me to compare genetic variation within and among *A. nemorensis* and its sister species (Table 10). *A. nemorensis* had the lowest level of intraspecific genetic diversity with π =0.0002. The tetraploid *A. hirsuta* had the highest diversity with π =0.012, almost two orders of magnitude higher than *A. nemorensis*. *A. sagittata* had three times the genetic diversity of *A. nemorensis*. Interspecific genetic diversity was highest between *A. hirsuta* and *A. nemorensis* with d_{XY}=0.017, followed by *A. hirsuta* and *A. sagittata* with d_{XY}=0.013. Genetic divergence was lowest between *A. nemorensis* and *A. sagittata* with d_{XY}=0.008. In *A. nemorensis* and *A. sagittata* F_{IS} was strongly 0.85 and 0.75, respectively, indicating an excess of homozygosity in these species. In contrast, F_{IS} in *A. hirsuta* was -0.85, indicating an excess of heterozygosity.

Table 10: **Interspecific comparison of genetic diversity**
In the upper triangle are the average pairwise genetic distances between species (d_{XY}). On the diagonal are the average pairwise genetic distances within species (π). Both values are normalized by the total number of genotyped bases.

d_{XY} (π)	***A. hirsuta***	***A. nemorensis***	***A. sagittata***
A. hirsuta	0.012	0.017	0.013
A. nemorensis		0.0002	0.0080
A. sagittata			0.0006

3.3.2 STRONG POPULATION STRUCTURE AND LOW DIVERSITY IN *ARABIS NEMORENSIS*

I genotyped 121 individuals of *A. nemorensis* at 4 Mb (2% of the genome) and identified 4713 SNPs. To obtain an overview of intraspecific genetic diversity I calculated pairwise genetic distance between all individuals and plotted them as a clustered heatmap (Figure 27). Most populations, except Rhine and Adl-2 consisted of genetically almost identical individuals. Adl-2 consisted of two genetic groups, while the Rhine population consisted of several groups. In Chapter 2, I have shown that these groups are relatively evenly distributed over the subpopulations.

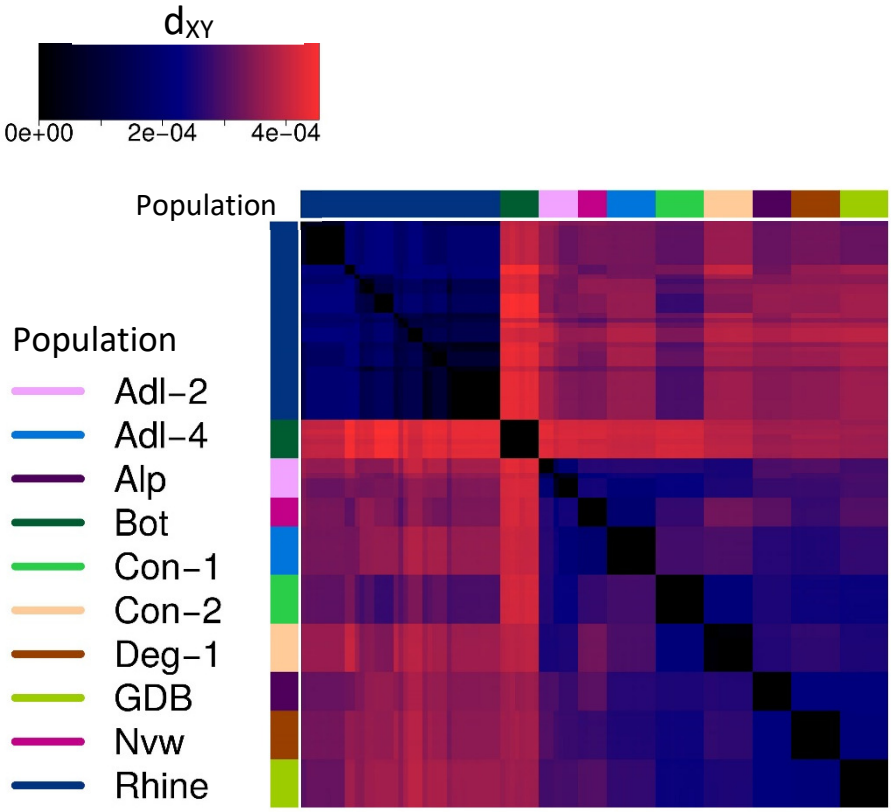

Figure 27: **Genetic distances among *A. nemorensis* individuals indicate strong structure**
Heatmap of all pairwise genetic distances between *A. nemorensis* individuals. Individuals were clustered based on genetic distance. Panel color represents genetic distance according to the scale on the top-left. Populations are indicated by color bars next to/above rows/columns.

Population were very highly differentiated, with a mean F_{ST} of 0.9. The minimum F_{ST} of 0.62 was observed between the two diverse populations Adl-2 and Rhine. Thus, F_{ST} is strongly driven by low within populations diversity and less by diversity between populations. To better resolve genetic relationships between populations, I calculated pairwise genetic distances (Cavalli-Sforza and Edwards chord distance) for all populations (Figure 28). The *ex situ* population from the Botanical Garden Halle, which was originally collected from a now extinct population in the drainage area of the Elb river, formed a clear outlier with minimum genetic distance of 0.39 to all other populations. In comparison the maximum difference between all other populations was 0.35 between Con-2 and Nvw. The Rhine population also shows relatively high divergence from other populations, with an average of 0.3. It was most similar to Adl-2 and Con-1 with approximately 0.26 for both populations. The other populations formed three clusters of increased genetic similarity, partly corresponding to their geography. One cluster consisted of Adl-2, Adl-4 and Nvw, which are also geographically close, with an

average genetic distance of 0.22. Another cluster was formed by the two populations at Lake Constance, Con-1 and Con-2, with an average genetic distance of 0.23. The third cluster was formed by Alp, Deg-1 and GDB, which are geographically distant, with an average genetic distance of 0.24. Overall, I found significant isolation by distance, but at least three populations deviated from this pattern, as described above (Mantel-test: r=0.66, p=0.003; Figure 28).

Total genetic diversity (π) in the sample was 2.91e-04, but within-population genetic diversity varied strongly among populations (Figure 28). Populations (excluding individuals identified as hybrids; see Chapter 2) Rhine and Adl-2 were most diverse with π=1.49e-04 and π=1.06e-04, respectively. Genetic diversity in all other populations was very low with an average of 3.91e-06, which is almost two orders of magnitude lower than the other two populations.

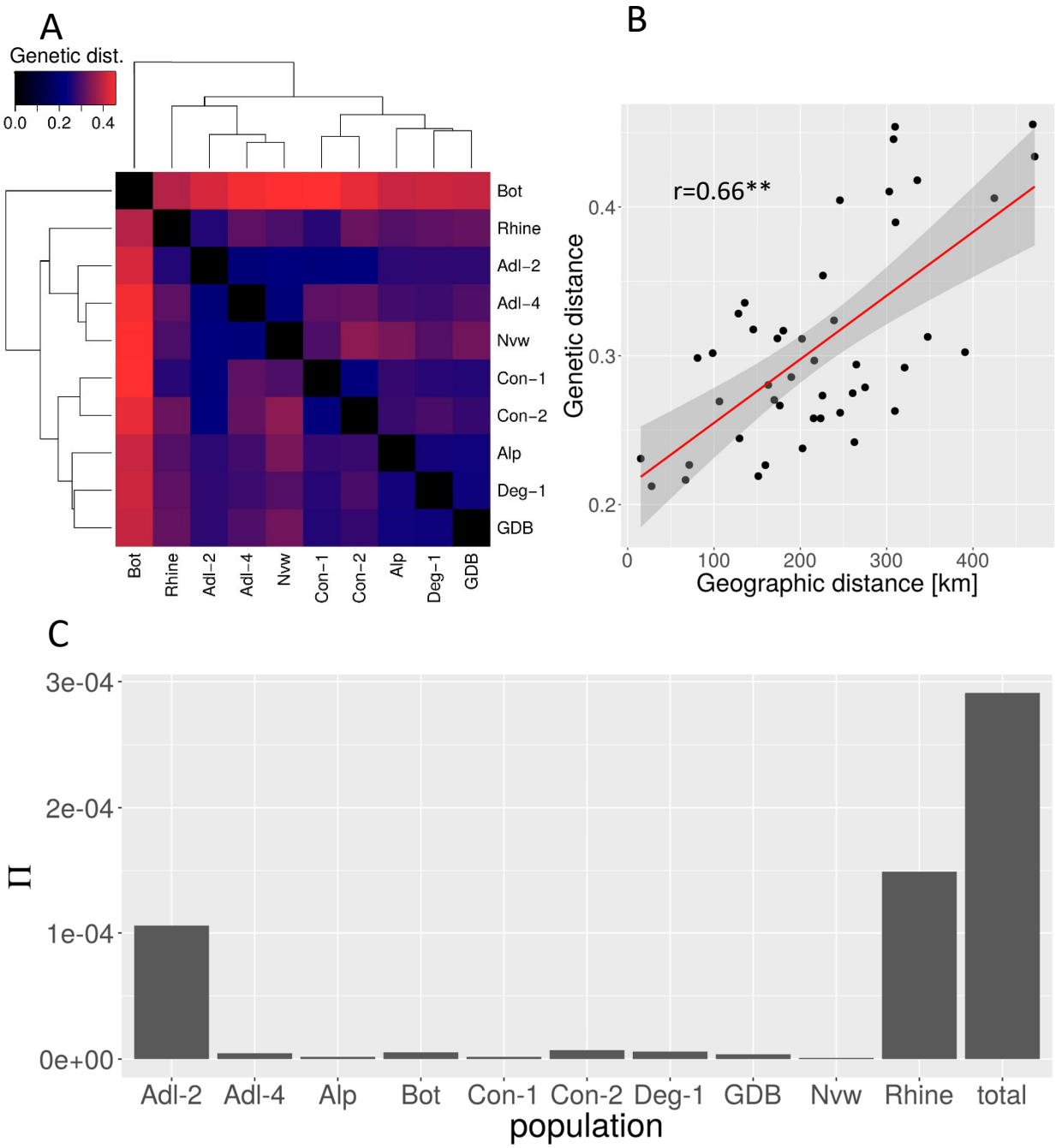

Figure 28: **Strong population structure and low diversity in *A. nemorensis***
A) Heatmap showing genetic distance (Cavalli-Sforza & Edwards chord distance) among populations. Panels are colored according to the genetic distance of the population pair and the color scale in the top-left corner. Dendrograms indicate distance-based clustering of populations. B) Correlation plot of geographic distance vs. genetic distance of all population pairs. The red line is the linear fit and the grey shadows are the errors of the fit. R was estimated by a mantel-test, which was significant at $p<0.01$. C) Barchart showing genetic diversity (average pairwise distance) within each *A. nemorensis* population.

3.3.3 POPULATION STRUCTURE OF *ARABIS NEMORENSIS* IS INDEPENDENT OF CLIMATIC VARIATION

To analyze climatic variation among confirmed *Arabis nemorensis* populations, I used publicly available climate data. To compare the degree of variance among climatic factors, I normalized each variable by its maximum value in the whole sampling region and scaled to percent (see methods). The degree of variance differed among climatic variables (Figure 29 A). Sites showed in the highest relative variance in mean temperature of the driest quarter with 1785.3, followed by altitude with 612.4 and precipitation with 291.6. In contrast, sites showed the lowest relative variance in iso-thermality with 4.6, followed by temperature seasonality with 8.9 and solar radiation with 12.3. I also compared the climatic variation among sites with the climatic variation in the sampling region. For precipitation, wind speed and altitude most samples were below the mean of the region. For humidity, temperature seasonality and max. temperature of the warmest month, most samples were above the mean of the region. For soil water-content populations were distributed over the whole range of the region. Interestingly, the soil water-content of the Rhine population was below the 5% quantile of the region. For solar radiation and iso-thermality the variation in the regions was very small and samples were distributed over the whole range.

To analyze climatic distances among populations I conducted PCA (Figure 29 B). The first PC explained 56% of the variance and had strongest loading from altitude (+), precipitation (+), temperature (-) and humidity (-). The Alp site was a strong outlier on PC 1, indicating that this high-elevation site is cold and wet. On the other extreme are the Rhine and Bot populations, which are relatively warm and dry. The other sites clustered in the center of PC 1. PC 2 explained 21% of the variance and had strongest loadings from wind speed (-), radiation (+), soil water content (+). The strong positive outliers on this axis were the Lake Constance sites, Con-1 and Con-2, featuring relatively low wind speed, but high radiation, high water content and high temperature in the driest season. On the opposite end of PC 2 is the site Bot. PC 3 (not shown) explained 12% of the variance and had the strongest loading from precipitation seasonality (+) and iso-thermality (-), thus higher values on this PC equate greater environmental variability. Most populations showed relatively high environmental variability except Rhine, GDB and Deg-1. To summarize, the sites Rhine, Bot, Alp, Deg-1, Con-1 and C on-

2 were climatic outliers with unique climatic conditions, while the other sites are relatively similar.

Environmental gradients can influence genetic differentiation among populations (Wang & Bradburd, 2014). Under this scenario genetic distance/F_{ST} between populations is expected to increase with increased climatic distance. Thus, I calculated the distance between sites on the first three PCs (which explain 92.8% of the variance) as a measure for climatic distance. I found no significant correlation between climatic distance and genetic distance (Mantel-test: r= 0.29, p=0.09; Figure 29 C) or F_{ST} (Mantel-test: r=0.11, p=0.44).

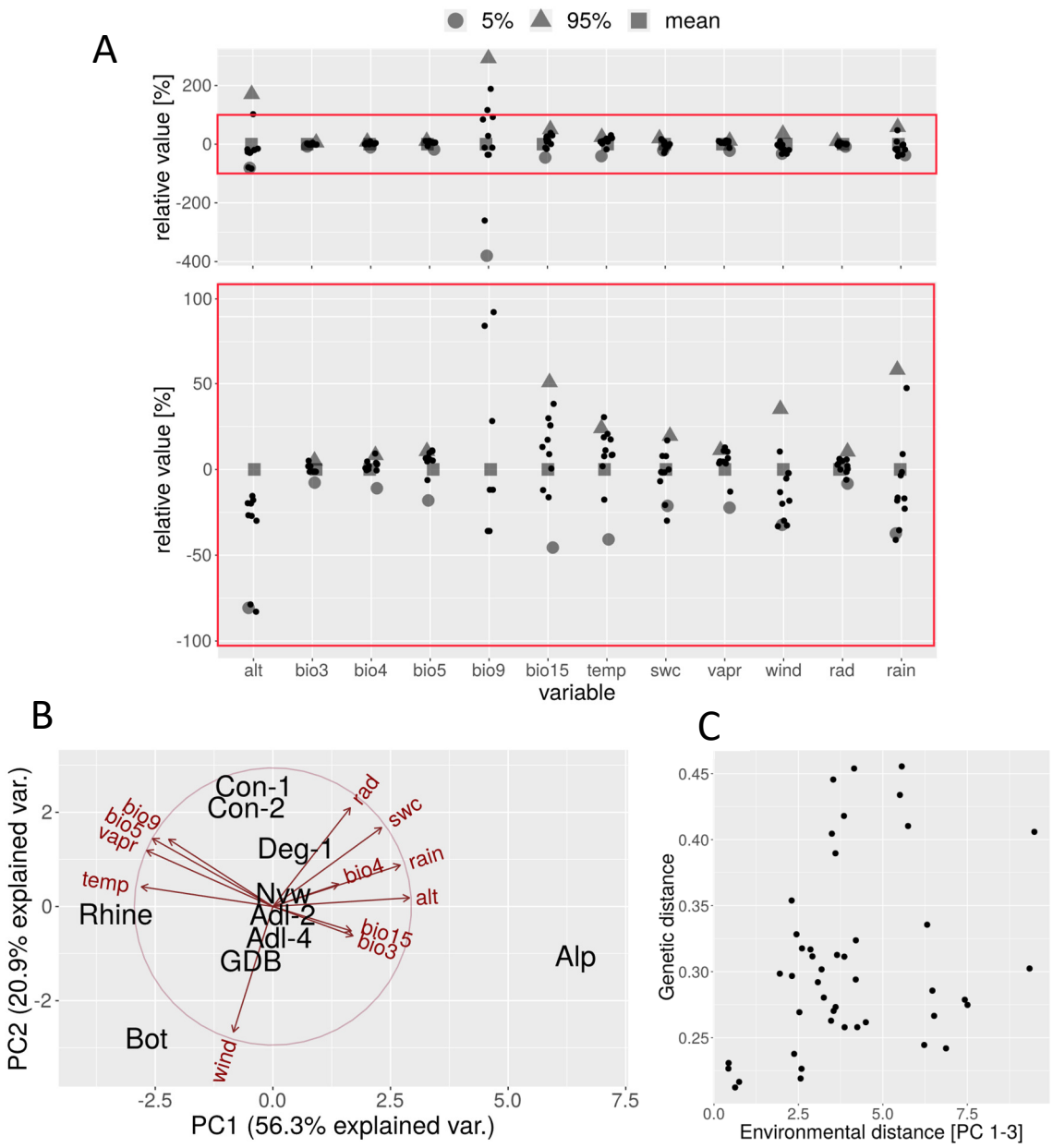

Figure 29: **Climatic variation among *A. nemorensis* populations**
A) Representation of the degree of variation of each climatic variable. Each point represents one site. The upper plot shows the whole range of y-values and the lower plot shows the window -100% < y < 100%. Values are the normalized difference to the mean of that variable and are scaled in %; Large symbols represent climatic variation of the sampling region showing the 5% quantile, 95% quantile or the mean according to the legend; B) PCA of 11 climatic variables represented by the first two PCs. Sites are represented by their respective label. Arrows indicate the loadings of climatic variables on the PCs: arrow direction represents the relative loading on the PCs and arrow length represents the strength of loading.; C) Correlation plot of genetic distance and the climatic distance taking into account the first three PCs explaining 92% of the variance.; Abbreviations for climatic variables: alt – altitude, bio4 – temperature seasonality, bio5 – max. temperature of the warmest month, bio3 – isothermality, bio9 – mean temperature of the driest quarter, bio15 – precipitation seasonality, temp – mean annual temperature, swc – mean annual soil water-content, vapr – mean annual water vapor pressure, wind – mean annual wind speed, rad – mean annual solar radiation, rain – mean annual precipitation

3.4 CHAPTER 4 – PHENOTYPIC DIVERGENCE BETWEEN *A. NEMORENSIS* AND *A. SAGITTATA*

3.4.1 COMMON GARDEN EXPERIMENT IN SEMI-NATURAL CONDITIONS

The goal of this experiment was to identify phenotypes which show genetic variation between species in a semi-natural environment. I calculated broad-sense heritability and tested the significance of species and accession (within species) effects for all phenotypes. For rosette diameter (Figure 30 A) both species label (Likelihood-Ratio-Test (LRT) $X^2 = 4.3$, df=1, p=0.04) and accession (LRT $X^2 = 174.8$, df=42, p<0.00001) were significant predictors. Broad sense heritability was 0.26. Per-accession mean rosette diameter ranged from 5.6 cm to 7.8 cm. Mean rosette diameter was 6.6 cm (+/- 0.5 cm) for *A. nemorensis*, 6.3 (+/- 0.8 cm) for *A. sagittata* and 6.7 (+/- 0.8 cm) for hybrids. Due to multiple comparison correction the species did not differ significantly from each other, despite an overall significant species effect.

For herbivory (Figure 30 B), species label (LRT $X^2 = 16.7$, df=1, p<0.0001) was a significant predictor, but not accession (LRT $X^2 = 37.8$, df=42, p=0.65). Broad sense heritability was indeed very low with 0.03. Per accession mean herbivory ranged from 0.71 to 2.0. Mean herbivory was 1.15 (+/- 0.21) for *A. nemorensis*, 1.51 (+/- 0.29) for *A. sagittata* and 1.52 (+/- 0.22) for hybrids. *A. nemorensis* was significantly different from the two other groups.

For bolting time (Figure 30 C), both species label (LRT $X^2 = 161.2$, df=1, p<0.00001) and accession (LRT $X^2 = 194.1$, df=42, p<0.00001) were significant predictors. Broad-sense heritability was 0.45. Per-accession mean bolting time ranged from 201 to 214 days. Mean bolting time was 204.1 d (+/- 1.6 d) for *A. nemorensis*, 208.3 d (+/- 2.6) for *A. sagittata* and 206.5 d (+/- 2.7) for hybrids. All groups were significantly different from each other.

For stem-leaf density (Figure 30 D), both species label (LRT $X^2 = 359.6$, df=1, p<0.00001) and accession (LRT $X^2 = 95.9$, df=42, p<0.00001) were significant predictors. Broad-sense heritability was 0.54. Per-accession means for stem-leaf density ranged from 0.64 to 1.52 leaves/cm. Mean stem leaf density was 1.36 (+/- 0.11) for *A. nemorensis*, 0.82 (+/- 0.11) for *A. sagittata* and 1.12 (+/- 0.14) for hybrids and all groups differed significantly from each other.

For final height (Figure 30 E), accession (LRT X^2 = 37.8, df=42, p=0.65) was a significant predictor, but not species label (LRT X^2 = 0.19, df=1, p=0.65). Broad-sense heritability was 0.22. Per-accession means for final height ranged from 48.23 to 76.57 cm. Mean final height was 61.3 cm (+/- 6.1 cm) for *A. nemorensis*, 61.4 cm (+/- 7.8 cm) for *A. sagittata* and 64.1 cm (+/- 8.7 cm) for hybrids.

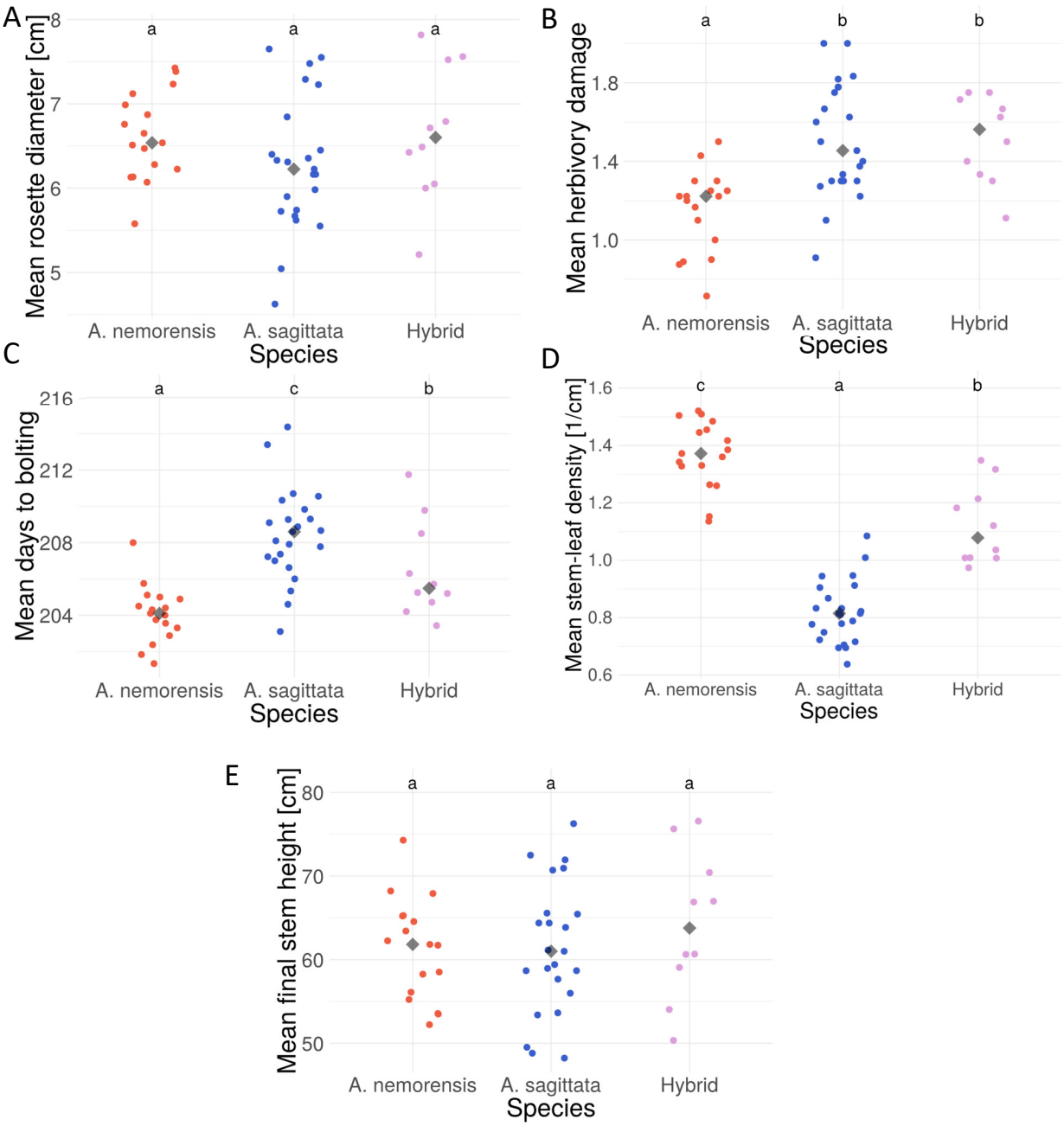

Figure 30: **Significant variation in morphology, phenology and stress tolerance among species.** Points show the per accession means of each trait. The gray diamond represents the median of each species. Distinct letters on top among groups indicate significant (p<0.05) differences in pairwise comparisons based on generalized linear hypothesis tests.

For number of seeds (Figure 31 A), species label (LRT X^2 = 12.33, df=1, p<0.001) was a significant predictor, but not accession (LRT X^2 = 46 df=42, p=0.29). Broad-sense heritability was 0.04. Per-accession means for seed number ranged from 484 to 3753. Notably, both values were measured in hybrids indicating strong variance among hybrid accessions. Mean number of seeds was 2157 (+/- 400) for *A. nemorensis*, 1640 (+/- 524) for *A. sagittata* and 1505 (+/- 892) for hybrids. *A. nemorensis* was significantly different from the other two groups.

Number of siliques (Figure 31 B) also had species label (LRT X^2 = 10.1, df=1, p<0.01) but not accession (LRT X^2 = 49 df=42, p=0.18) as a significant predictor. Broad sense heritability was 0.07. Per accession means ranged from 19 to 85. Mean number of siliques was 45 (+/- 7) for *A. nemorensis*, 36 (+/- 10) for *A. sagittata* and 48 (+/- 17) for hybrids. *A. sagittata* differed significantly from the other two groups.

For seed weight (Figure 31 C), neither species label (LRT X^2 = 0.18 df=1, p=0.67) nor accession (LRT X^2 = 54 df=42, p=0.09) were significant predictors. Broad-sense heritability was smaller than 0.01. Per-accession mean seed weight ranged from 0.0004 to 0.15. Mean seed weight was 0.006 g (+/- 0.003 g) in *A. nemorensis*, 0.0071 g (+/- 0.002 g) in *A. sagittata* and 0.008 g (+/- 0.004) in hybrids.

For silique length (Figure 31 D) both species label (LRT X^2 = 14.24 df=1, p=0.0002) and accession (LRT X^2 = 81 df=42, p=0.0003) were significant predictors. Broad-sense heritability was 0.27. Per-accession means ranged from 2.8 cm to 5.9 cm. Mean silique length was 4.0 cm (+/- 0.33 cm) for *A. nemorensis*, 4.6 cm (+/- 0.56 cm) for *A. sagittata* and 3.7 cm (+/- 0.53 cm) for hybrids. *A. sagittata* differed significantly from the other two groups.

The large number of siliques but low number of seeds in hybrids suggests reduced seeds per silique in hybrids, potentially due to fertility problems. Indeed, species label was a significant predictor for seeds per silique (LRT X^2 = 22.67 df=2, p<0.0001), but accession was not (LRT X^2 = 39.1 df=42, p=0.78). Broad-sense heritability was 0.1. Per-accession means for seeds per silique ranged from 12.75 to 55.3. Mean number of seeds per silique was 46.2 (+/- 4.4) for *A. nemorensis*, 43.3 (+/- 5.7) for *A. sagittata* and 29.63 (+/- 9.9) in hybrids and hybrids were significantly different from the parental species.

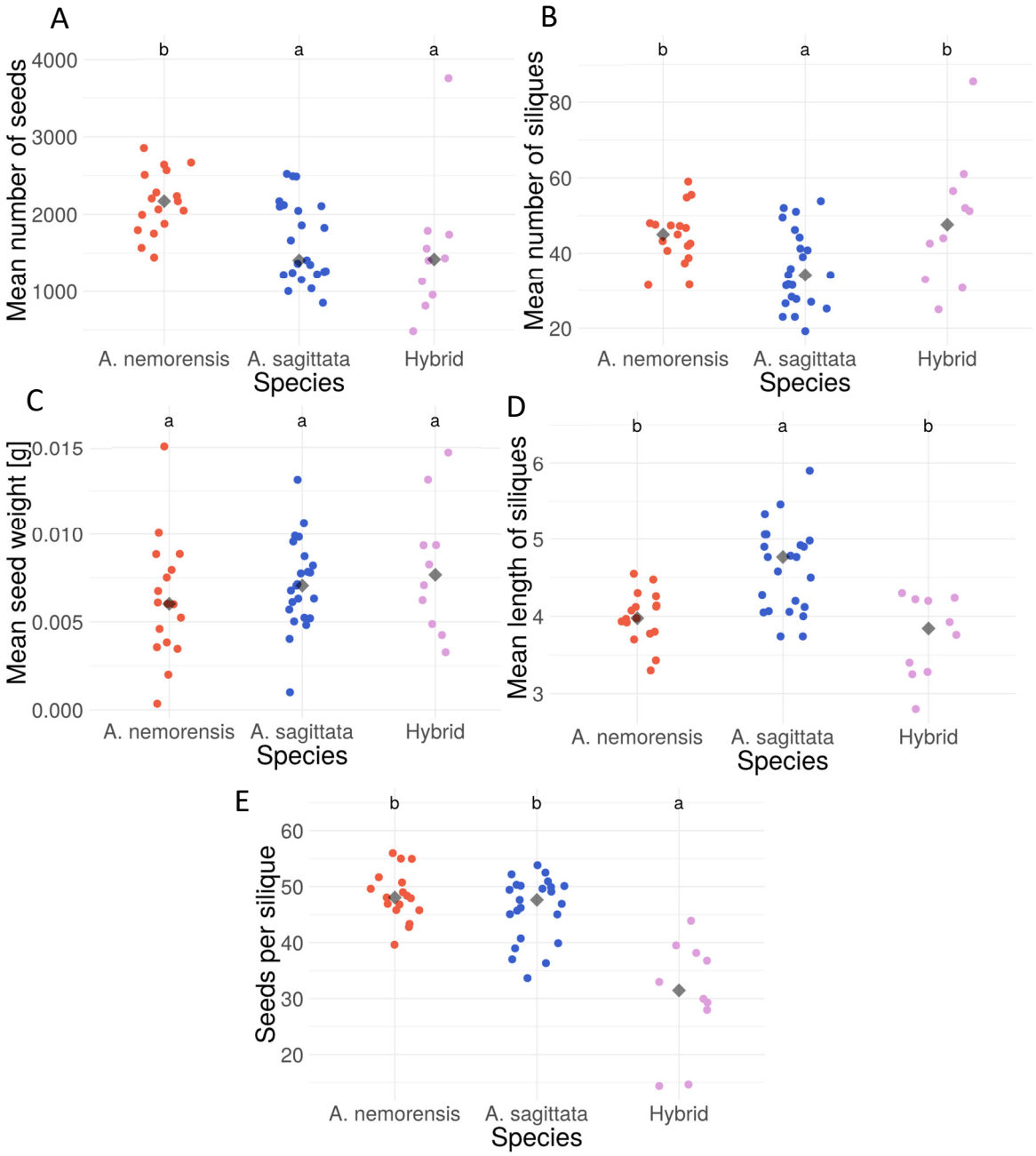

Figure 31: **Significant phenotypic variation in fitness related traits among species and hybrids.** Points show the per accession means of each trait. The gray diamond represents the median of each species. Distinct letters on top among groups indicate significant ($p<0.05$) differences in pairwise comparisons based on generalized linear hypothesis tests.

In summary, the most striking phenotypic difference between the two species was the stem leaf density. Additionally, *A. nemorensis* and *A. sagittata* showed small but significant differences in most other phenotypes: *A. nemorensis* bolted and flowered earlier, showed reduced herbivory damage and increased rosette size. While the phenotypes of hybrids were on average intermediate between their parental species, they showed increased variance.

A. nemorensis also showed increased fitness compared to *A. sagittata*. Interestingly, hybrid accessions had fewer seeds per silique, suggesting potential problems in seed formation due to incompatibilities. However, the total number of seeds in hybrids was not reduced.

To characterize co-variance among all measured genotypes, I calculated all pairwise correlations for each species (Figure 32). I found 11 significant correlations (p<0.05) in *A. nemorensis* and 28 significant correlations in *A. sagittata*. I found the strongest correlation between the number of siliques and the number of seeds in both *A. nemorensis* (r=0.93, p=9.7e-27) and *A. sagittata* (r=0.94, p<0.001), thus both traits give a similar estimate of fitness. Fitness (seeds) was correlated in both species with rosette size (*A. nemorensis* r=0.46, p=4.4e-03; *A. sagittata* r=0.32, p=2.5e-02) and final height (*A. nemorensis* r=0.66, p=8.2e-08; *A. sagittata* r=0.54, p=1.5e-07) and in *A. sagittata* also with herbivory (r=-0.33, p=). Silique length and the number of seeds per silique was correlated in both species (*A. nemorensis* r=0.5, p=7.5e-04; *A. sagittata* r=0.65, p=9.7e-12). Yet, only in *A. sagittata* it silique length also correlated with the number of seeds (r=0.47, p=2.4e-05), the number of siliques (r=0.31, p=4.0e-02), rosette size (r=0.41, p=6.0e-04), final height (r=0.49, p=9.7e-06), bolting time (r=-0.31, p=4.4e-02) and leafdensity (r=-0.31,p=4.4e-02).

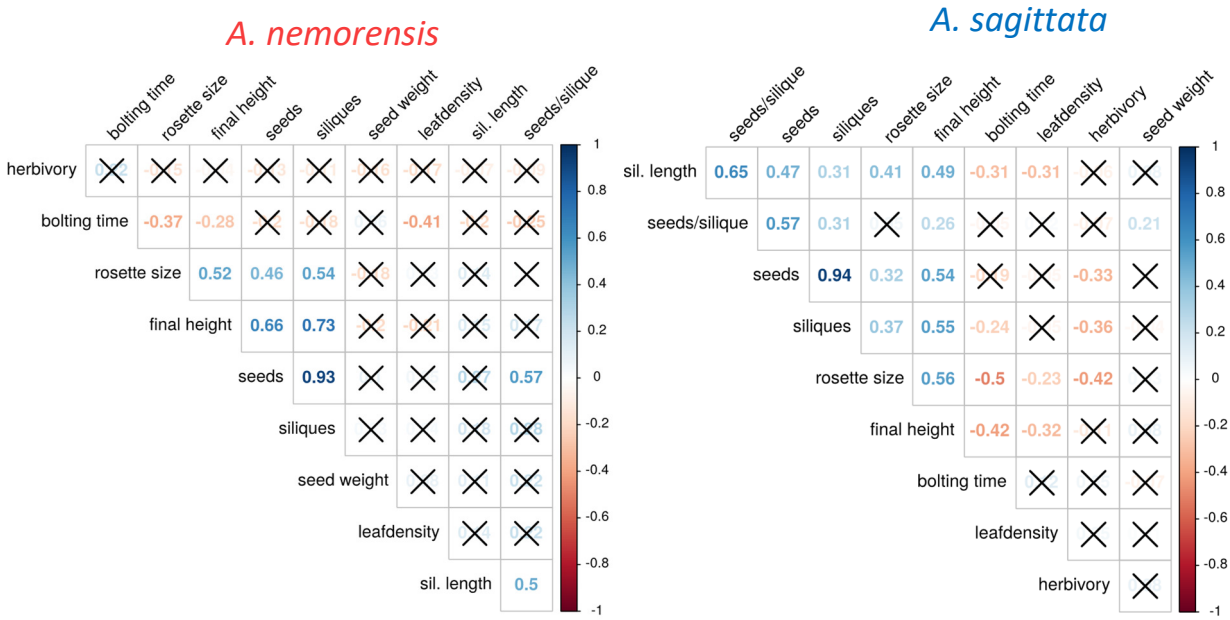

Figure 32: **Correlation matrices of all measured phenotypes for each species.**
Numbers in squares are Pearson correlation coefficients. Crossed-out squares are non-significant (adjusted p>0.05). Phenotypes are clustered based on their correlation.

3.4.2 SUBMERGENCE EXPERIMENTS

The goal of the two submergence experiments was to quantify flooding tolerance in the two *Arabis* species as well as outgroup species with different known levels of flooding resistance: *Rorippa palustris* – high tolerance; *Arabidopsis thaliana* – low tolerance. In models for survivorship for both experiments, both the treatment and the species-treatment interaction term were significant (Table 11). For the second experiment, I included the effect of the rosette diameter before submergence on survivorship, but it had no significant effect (Table 11). The reactions to the flooding treatment of all tested species, except *A. sagittata* and hybrids, were significantly different from each other in both experiments.

Table 11: **Overview of statistical results of survival after submergence.**
Treatment (treat) is the flooding duration. Species (spec) is the term describing the species. The table summarizes results from different tests, thus test statistics differ. R. palustris was model intercept, thus treat:spec(speciesname) terms describe comparisons to it. Significant terms are written in bold. Species abbreviations: nemo – *A. nemorensis*, sag – *A. sagittata*, ath – *A. thaliana*, hyb – *Arabis hybrid*

Model	Term	Test-statistic	df	p
Exp. 1 (glm)	**treatment**	LR X^2=369.40	1	<2.2e-16***
Exp. 1 (glm)	**treat:spec**	LR X^2=191.86	3	<2.2e-16***
Exp. 1 (glm)	**treat:spec(nemorensis)**	t=-9.5	-	2.3e-16***
Exp. 1 (glm)	**treat:spec(sagittata)**	t=-9.97	-	<2e-16***
Exp. 1 (glm)	**treat:spec(thaliana)**	t=-4.75	-	5.59e-06***
Exp. 1 (glht)	**treat:spec(sag) - treat:spec(nemo) == 0**	z=-3.46	-	0.001**
Exp. 1 (glht)	**treat:spec(nemo) - treat:spec(ath) == 0**	z=3.85	-	<0.001***
Exp. 1 (glht)	**treat:spec(ath) - treat:spec(sag) == 0**	z=-3.36	-	0.002**
Exp. 2 (glm)	**treatment**	LR X^2=118.40	1	<1e-08***
Exp. 2(glm)	**treat:spec**	LR X^2=25.91	3	<1e-6***
Exp. 2 (glm)	RosetteDiameter_beforeFlood	LR X^2=0.16	1	0.69
Exp. 2 (glm)	**treat:spec(nemorensis)**	t=-2.4	-	0.02*
Exp. 2 (glm)	**treat:spec(sagittata)**	t=-3.87	-	0.0002***
Exp. 2 (glm)	**treat:spec(hybrids)**	t=-4.1	-	0.0001***

Exp. 2 (glht)	**treat:spec(sag) - treat:spec(nemo) == 0**	z=-3.0	-	0.006**
Exp. 2 (glht)	**treat:spec(nemo) - treat:spec(hyb) == 0**	z=2.83	-	<0.012*
Exp. 2 (glht)	**treat:spec(sag) - treat:spec(hyb) == 0**	z=0.9	-	0.63

In both experiments, *R. palustris* showed the highest flooding resistance, by far, with 100% survival even after 6 weeks of flooding (Table 11). In contrast, *A. thaliana* showed the lowest flooding resistance in the first experiment, with less than 15% survivors after one week of submergence and no survivors after 2 weeks. The two *Arabis* species gave relatively similar results across both experiments. Both showed lower flooding tolerance than *R. palustris*. However, *A. nemorensis* constantly showed a higher survivorship than *A. sagittata*, which was statistically significant. In the first experiment, the difference between the species is maximal after two weeks of flooding; in the second after 4 weeks. Interestingly, I observed the same pattern in both experiments: from week 3 to week 4 the survivorship of *A. sagittata* further declines at a similar rate as before (and reaching 0); in contrast, *A. nemorensis* survivorship does not change between week 3 and week 4 but stays around 25% and only declines in week 5. The hybrid accessions a behave in the same way as *A. sagittata*. In the second experiment, I tested the effect of the rosette diameter before submergence on survivorship, but it had no effect.

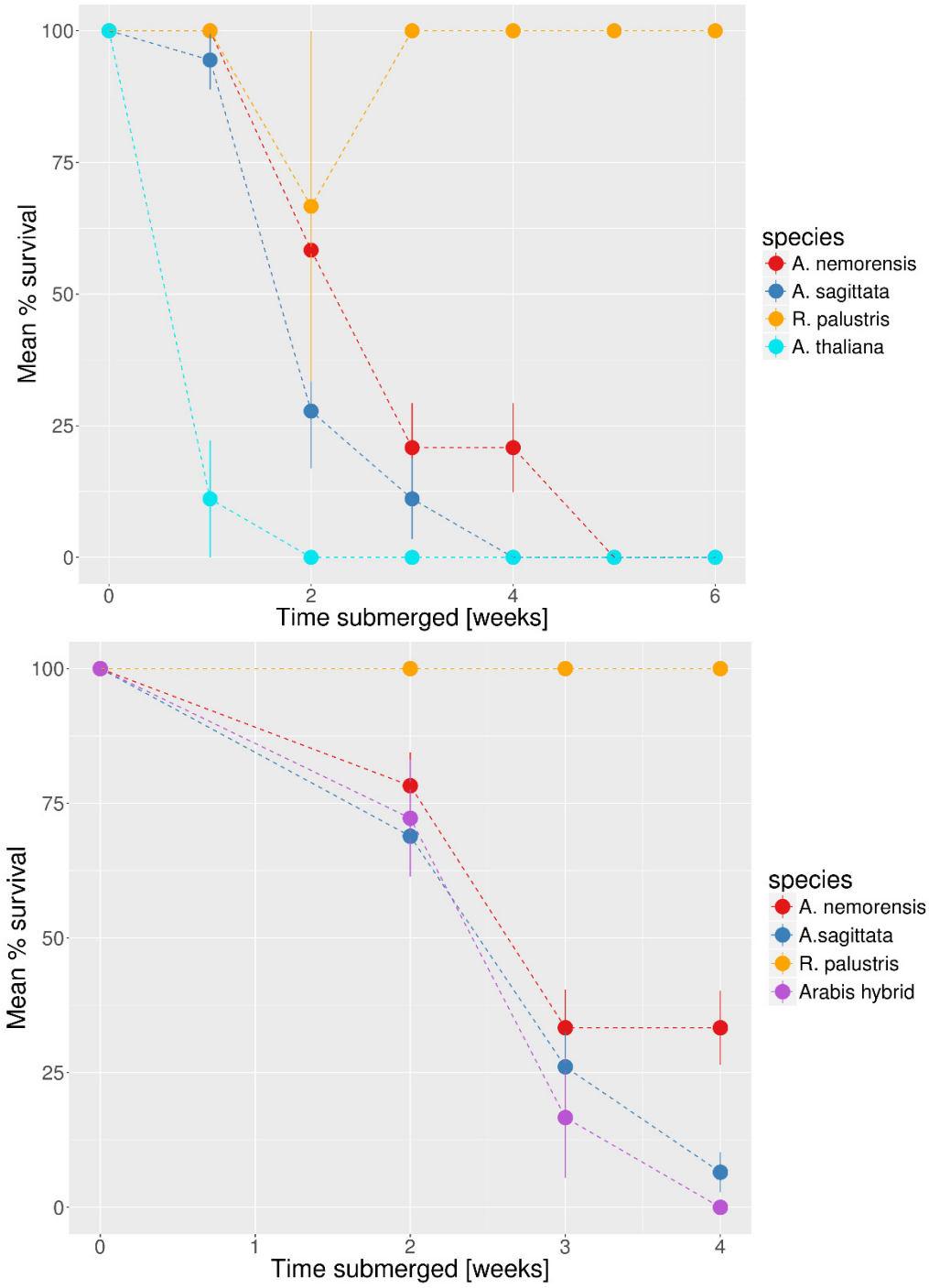

Figure 33: ***A. nemorensis* is more flooding resistant than *A. sagittata***
Plots show the results of two independent experiments. The x-axis shows the duration of the submergence period. The y-axis is the percentage of survivors. Each point shows the mean per species and error bars indicate the standard error. The treatment-species interaction term was significant for all species pairs except, *A. sagittata* and *Arabis* hybrid (see Table 11 for details).

In the second experiment, I characterized the recovery of plants by measuring several phenotypes on surviving plants starting two weeks after de-submergence: number of leaves two weeks after de-submergence, rosette diameter two weeks after de-submergence,

average leaf emergence rate from week and average rosette diameter growth rate from week two to four, dry weight and root length four weeks after de-submergence.

For the number of leaves two week after de-submergence (Figure 34 A), I found a significant treatment (LRT X^2 = 12.99, df=1, p=0.0003) and species effect (LRT X^2 = 11.23, df=1, p=0.0008), but interaction (LRT X^2 = 0.74, df=1, p=0.38). Interestingly, plants submerged for four weeks had most leaves in both species. On average *A. nemorensis* had 6.4 leaves and *A. sagittata* 5.2 leaves. The number of leaves before submergence had no effect (LRT X^2 = 0.03, df=1, p=0.85).

For rosette diameter two weeks after de-submergence (Figure 34 B), I also found a significant treatment (LRT X^2 = 24.17, df=1, p<0.0001) and species effect (LRT X^2 = 5.26, df=1, p=0.02), but no interaction (LRT X^2 = 0.005, df=1, p=0.93). The diameter generally decreased with treatment duration. On average, rosette diameter was 30.6 for *A. nemorensis* and 36.9. for *A. sagittata*.

For leaf emergence rate (Figure 34 C), I found a significant effect of treatment (LRT X^2 = 7.67, df=1, p=0.005) and species (LRT X^2 = 8.50, df=1, p=0.003) on this trait, but no interaction between species and treatment. In *A. nemorensis* the average emergence rate was 2.38 and in *A. sagittata* 2.05. The number of leaves two weeks after de-submergence had no effect (LRT X^2 = 0.8, df=1, p=0.36) and there was no significant block effect (LRT X^2 = 15.67, df=15, p=0.4).

For rosette growth rate (Figure 34 D), I found a significant effect of the treatment (LRT X^2 = 4.24, df=1, p=0.04) and of the rosette diameter after two weeks (LRT X^2 = 7.04, df=1, p=0.008), which correlated negatively with the growth rate.

For root length four weeks after de-submergence (Figure 34 E), only treatment had a significant effect (LRT X^2 = 38.26, df=1, p<0.00001), but not species (LRT X^2 = 0.02, df=1, p=0.89) and treatment-species interaction (LRT X^2 = 0.92, df=1, p=0.34). The root length generally decreased with treatment duration.

For rosette dry weight four weeks after de-submergence (Figure 34 F), only treatment had a significant effect (LRT X^2 = 6.82, df=1, p=0.009), but not species (LRT X^2 = 0.33, df=1, p=0.56) or the treatment-species interaction (LRT X^2 = 0.39, df=1, p=0.53). Dry weight was similar after two and three weeks of submergence but decreased after four weeks.

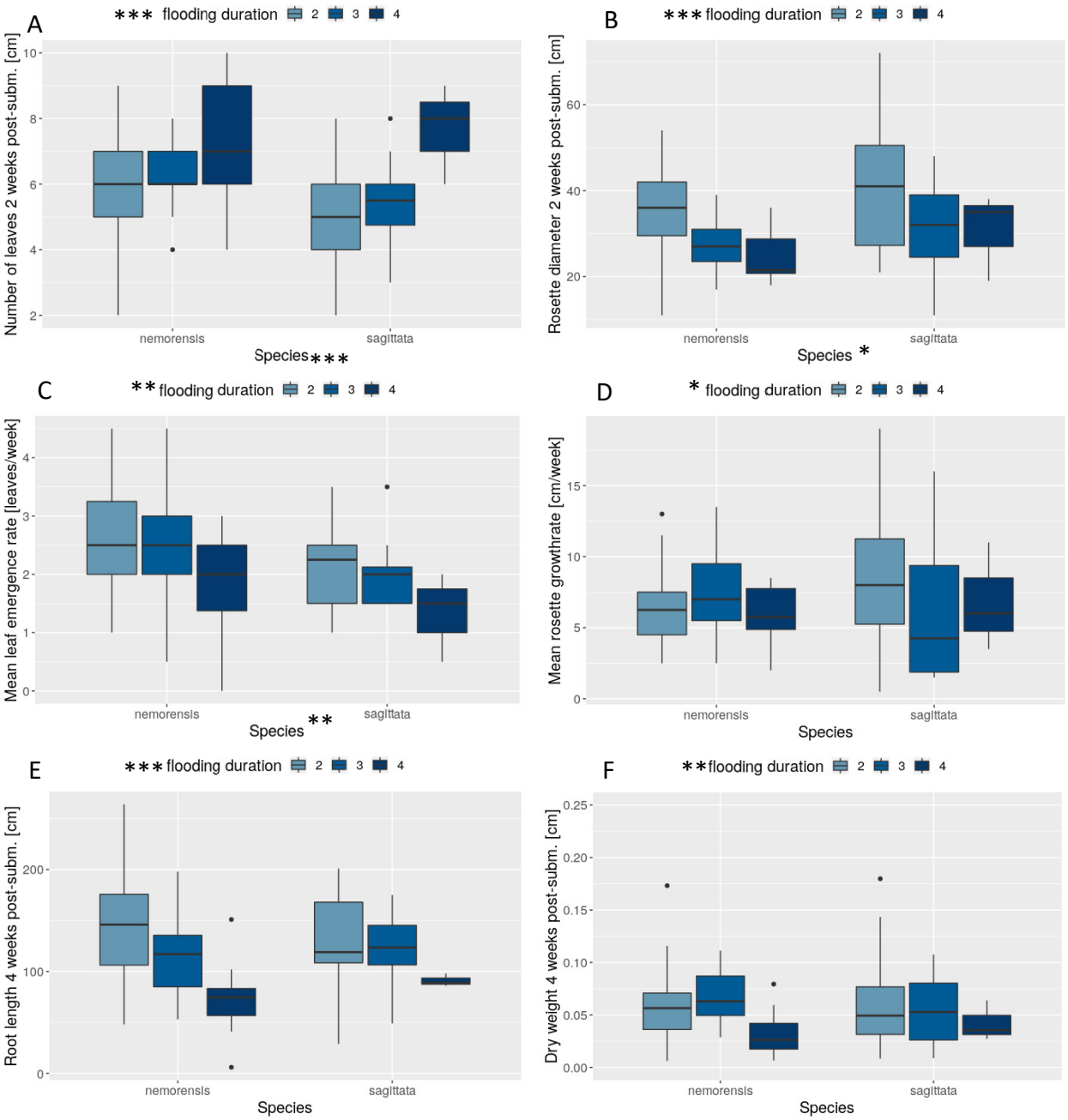

Figure 34: **Strong treatment but small species effect on plant growth after de-submergence**
Boxplots showing effect of flooding duration and species on growth-related traits in survivors of the submergence treatment. Stars next to the factors species and flooding duration indicate the respective significance of the effect: * p<0.05, ** p<0.01, ***p<0.001.

3.5 CHAPTER 5 – PATTERNS OF GENE-FLOW BETWEEN *A. NEMORENSIS* AND *A. SAGITTATA*

3.5.1 SIGNATURES OF INTERSPECIFIC INTROGRESSION

In chapter 2 I found that *A. nemorensis* and *A. sagittata* form natural hybrids and that these hybrids most likely back-cross with both parents but preferentially with *A. sagittata* (Figure 19). Thus, hybridization could facilitate interspecific gene-flow, which could increase the adaptive potential of one or both species. To test this, I conducted whole-genome re-sequencing of a total of 35 accessions, from three populations of *A. sagittata* and seven populations of *A. nemorensis* (see methods).

To characterize relative rates of gene flow among all plausible pairs of populations, I compared estimates of D calculated across the whole genome for three different phylogenies (Figure 35). For the phylogeny with *A. sagittata* Rhine population as P2 (donee), all other *A. sagittata* individuals as P1 (control), and all *A. nemorensis* individuals as P3 (donor), D was 0.22 and significant for both Jack-Knife-test (p = 7.4e-07) and Wilcox-Test (W = 25171000, p = 5.3e-05), indicating increased gene-flow from *A. nemorensis* specifically into the sympatric Rhine population of *A. sagittata* Rhine population. For the phylogeny with *A. sagittata* Adl-1 as P2, *A. sagittata* Lob as P1, and all *A. nemorensis* individuals as P3 D was 0.07 and significant for the Jack-Knife-test (p = 0.004) and the Wilcox test (W = 24738000, p = 0.02*)*, but its low value provided limited evidence for preferential gene flow into these allopatric populations. For the phylogeny with *A. nemorensis* non-Rhine individuals as P2, *A. nemorensis* Rhine individuals as P1 and *A. sagittata* Lob as P3, D was 0.15 and significant for both Jack-Knife-test (p = 5.07e-11) and Wilcox-test (W = 25902000, p = 0.0003), indicating stronger gene-flow from *A. sagittata* into allopatric *A. nemorensis* non-Rhine populations than into the Rhine population.

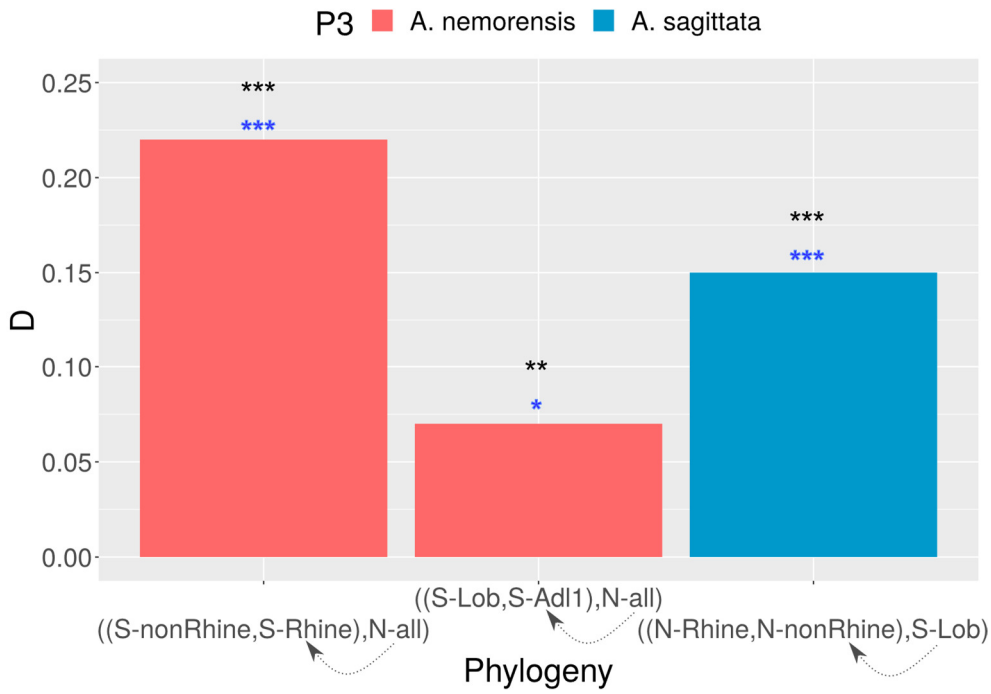

Figure 35: **Significant, genome-wide signatures interspecific gene-flow**
Estimates of D based on whole-genome ABBA and BABA counts for three different phylogenies. Phylogenies are represented in Newick-format below each bar: ((P1, P2), P3). Arrows indicate the gene-flow direction for positive values of D. Bars are colored by the donor species (P3). Stars above each bar indicate significance levels of the Jack-Knife-test of the D statistic(black) and Wilcox-Test between ABBA and BABA counts (blue): ***p < 0.001, **p<0.01, *p<0.05.

Because D can be biased in small genomic windows, I calculated f_D along the genome in 50-kb-windows to locate candidate regions for introgression. For this approach I tested two additional phylogenies, switching P1 and P2 (*A. sagittata* Lob and Adl-1, and *A. nemorensis* Rhine and non-Rhine), to detect introgressions even in populations with relatively less gene-flow. Distributions of f_D (Figure 36) differed significantly among all donee populations ($p<0.05$; pairwise Kolgomorov-Smirnov-Test). Mean f_D estimates were, in decreasing order, 0.02 for *A. sagittata* Rhine, 0.015 for *A. sagittata* Adl-1, 0.011 for *A. sagittata* Lob, 0.007 for *A. nemorensis* non-Rhine and 0.004 for *A. nemorensis* Rhine. Thus, introgression is more frequent from *A. nemorensis* into *A. sagittata* than vice versa and it is most frequent in the Rhine population of *A. sagittata,* which is the only population currently sympatric to *A. nemorensis*. Note that interspecific hybrids found in this sympatric population were removed from the analysis. Thus, f_D and D underestimates the degree of ongoing gene flow between the two species.

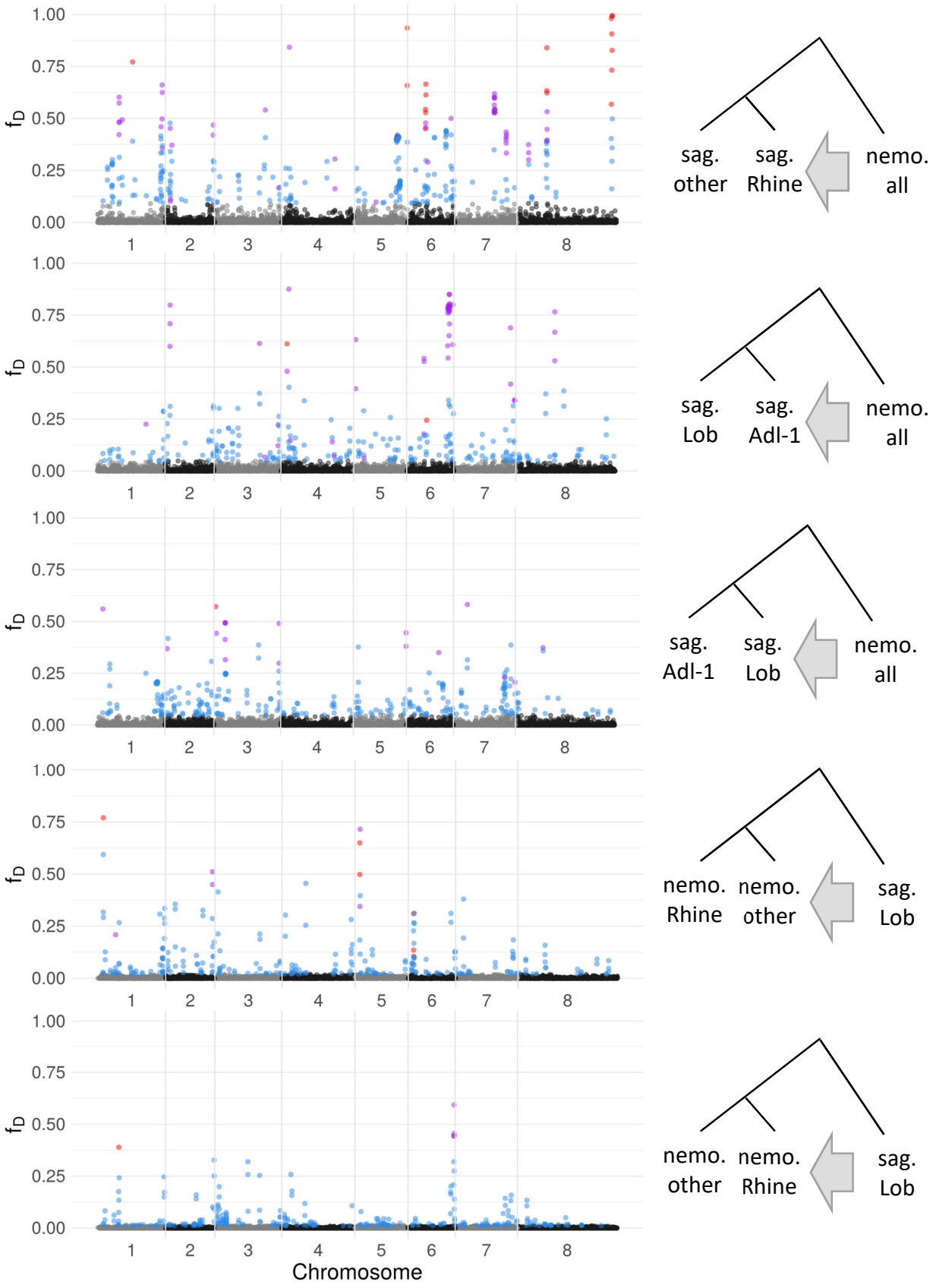

Figure 36: **Interspecific gene flow is asymmetric and strongest in the hybrid zone.**

Manhattan plots of fD calculated in 50 kb windows with 25 kb overlap. Each Manhattan plot shows fD calculated for the respective phylogeny depicted on the right. Arrows indicate the direction of gene-flow. Manhattan plots were arranged in decreasing order of introgression frequency. Points in Manhattan plots are colored according to the outlier level: blue – 1, purple – 2, red – 3. The higher the level, the better the evidence for introgression (see methods).

To select candidate regions for introgression from the f_D distribution, I defined three levels of confidence for introgressed regions. The first was threshold was the 95% quantile of the distribution. The second also required P2 to be more genetically similar to P3 than P1. The third required q95.1.100 (Racimo et al., 2015) to be greater than 0.9 (see methods for details) on top of the first two criteria. The number of outliers in level 2 and 3 differed among populations, agreeing with the differences in the mean f_D (Table 12).

Table 12: **Distribution of outliers for each introgression target**

	Outlier level 1	**Outlier level 2**	**Outlier level 3**
sag. Rhine	231	53	19
sag. Adl-1	226	48	2
sag. Lob	258	17	1
nemo. Non-Rhine	244	5	6
nemo. Rhine	249	4	1

3.5.2 COMPLEX PATTERNS OF INTERSPECIFIC INTROGRESSION

In each candidate region, I manually identified introgressed blocks and in which accessions they were present. This analysis revealed a variety of introgression patterns, of which I will hereafter highlight a few examples, focusing on introgression from *A. nemorensis* to *A. sagittata*. Within a 450 kb region on chromosome eight I identified three distinct introgressed blocks, with sizes ranging from 55 kb to 92 kb, completely fixed in the *A. sagittata* Rhine population (Figure 37, top). Additionally, I detected a region on chromosome 8 with a small introgression of about 30 kb nested within two distinct larger introgressions of about 200 kb in the Rhine population (Figure 37, bottom). The small introgression was fixed in the population, while the large introgressions were not.

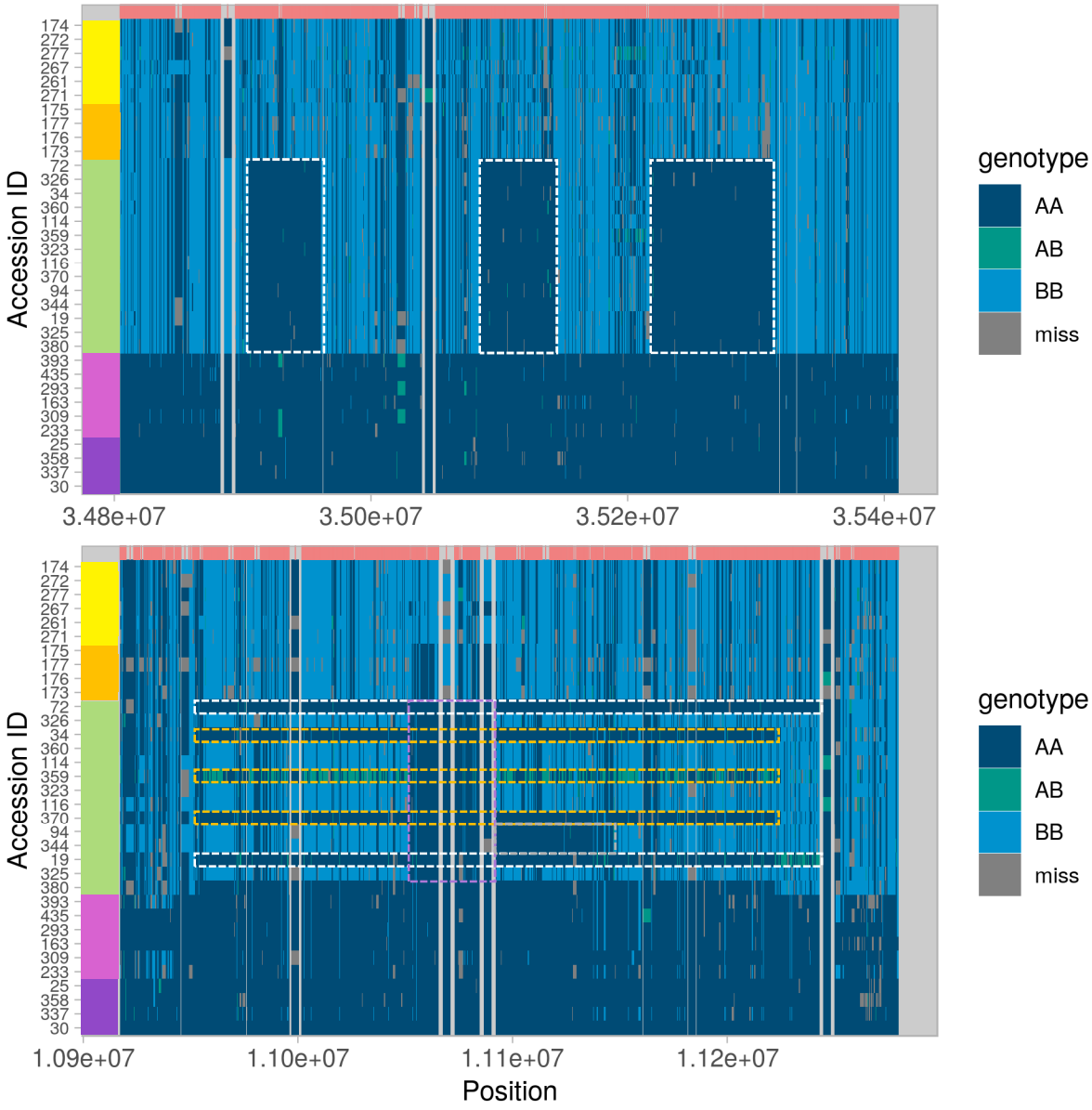

Figure 37: **Examples of complex introgression patterns in Rhine *A. sagittata***
Each plot depicts the genotype composition of all lines in a genomic region. Each SNP is colored according to its genotype. Genotypes are polarized so that accession 30 (*A. nemorensis* Rhine) is always genotype AA. For better visibility of isolated SNPs, SNP width was extended to up to 100th of the region width. Marker positions are indicated by red bars on top of each plot. Colored bars on the left side of the plot indicate the population membership of the accessions: (from top to bottom) yellow – *A. sagittata* Lob, orange – *A. sagittata* Adl-1, green – *A. sagittata* Rhine, violet – *A. nemorensis* scattered, purple – *A. nemorensis* Rhine. Different introgressed blocks in the same region were highlighted with dashed rectangles in different colors. Top) Three fixed introgressed blocks in close proximity in *A. sagittata*. Bottom) Nested introgressions in *A. sagittata* Rhine.

Some introgressed regions appeared to be shared among *A. sagittata* populations, and thus not specific to the sympatric population. On chromosome 2 I detected a region with multiple overlapping introgressions of different size (~100 - 300 kb) shared between *A. sagittata* Rhine and Adl-1, with multiple heterozygous genotypes in the Rhine population (Figure 38, top).

Additionally, I found a small introgression of about 30 kb shared between Lob and Rhine (Figure 38, bottom). The introgression was fixed in Lob and about 5 kb shorter than in Rhine. The introgressions were genetically highly similar among population differing only by 0 to 2 SNPs and sharing a private SNP, suggesting common origin.

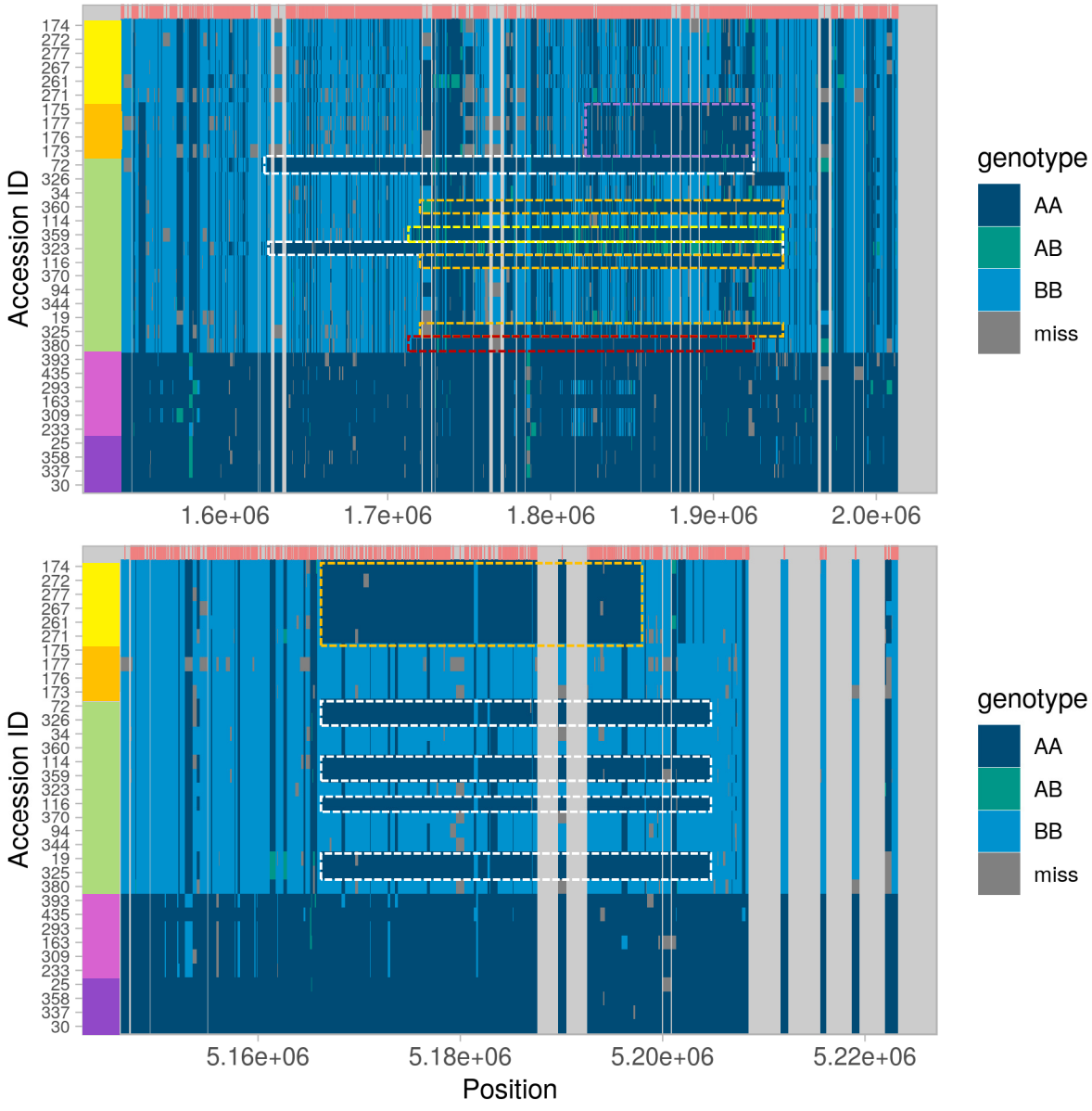

Figure 38: **Shared introgressed regions among *A. sagittata* populations**
Each plot depicts the genotype composition of all lines in a genomic region. Each SNP is colored according to its genotype. Genotypes are polarized so that accession 30 (*A. nemorensis* Rhine) is always genotype AA. For better visibility of isolated SNPs, SNP width was extended to up to 100th of the region width. Marker positions are indicated by red bars on top of each plot. Colored bars on the left side of the plot indicate the population membership of the accessions: (from top to bottom) yellow – *A. sagittata* Lob, orange – *A. sagittata* Adl-1, green – *A. sagittata* Rhine, violet – *A. nemorensis* scattered,

purple – *A. nemorensis* Rhine. Different introgressed blocks in the same region were highlighted with dashed rectangles in different colors. Top) Overlapping introgressions in A. sagittata Rhine and Adl-1. Bottom) Shared introgression in *A. sagittata* Rhine and Lob.

3.5.3 ANALYSIS OF HAPLOTYPES SUGGESTS COMPLEX HISTORY OF GENE-FLOW

Ongoing hybridization and increased rates of introgression in the *A. sagittata* Rhine population, suggest that this population is a major source of interspecific introgression. Yet, introgressed fragment of *A. nemorensis* were sometimes also detected in allopatric *A. sagittata* populations, indicating that introgression has taken place before the formation of the Rhine population. To tease apart these two scenarios, I created haplotype-networks for each of the 16 introgressed regions including all *A. nemorensis* haplotypes and all introgression haplotypes in *A. sagittata* (Figure 39). In seven regions all introgressed haplotypes had the best match to one of the Rhine *A. nemorensis* haplotypes. In three regions all introgressed haplotypes had the best match to one the non-Rhine *A. nemorensis* haplotypes. In the remaining six regions, introgressed haplotypes matched to both Rhine and non-Rhine haplotypes or matched ambiguously. Thus, a majority of introgression appear to originate from the Rhine population.

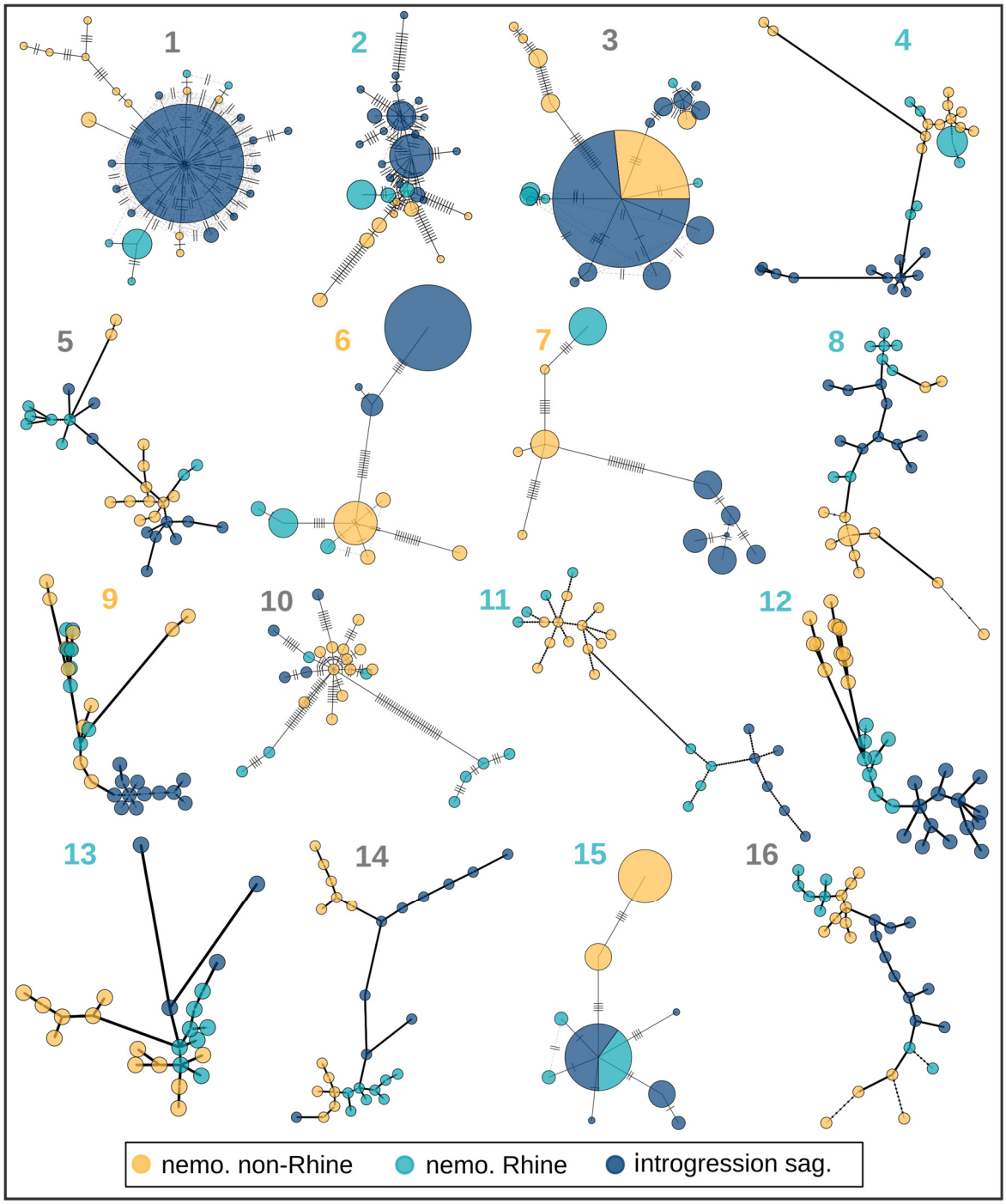

Figure 39: **Sympatric *A. nemorensis* is the main source of introgression in Rhine *A. sagittata***
Haplotype networks for sixteen *A. nemorensis* introgression regions in A. sagittata, including all potential donor haplotypes and introgression haplotypes. Haplotypes are represented by circles whose size represents the relative frequency. Haplotypes are connected by paths with mutation represented as dashes or dots. Haplotypes are colored by species/populations, according to the legend. Numbers above each network are colored according to the best *A. nemorensis* match for introgressed haplotypes, according to the legend and gray if matching is ambiguous.

I also created haplotypes networks for a total of ten introgressed regions in the two *A. sagittata* populations, currently isolated from *A. nemorensis* (Figure 40). Here, only in three regions all introgressed haplotypes had the best match to one of the Rhine *A. nemorensis* haplotypes. In the remaining six regions all introgressed haplotypes had the closest match to one of the non-Rhine *A. nemorensis* haplotypes. These results suggest that the Rhine population was also involved in ancestral introgression, but that recent introgressions in the Rhine lineages might be responsible for the observed increase in introgression frequency.

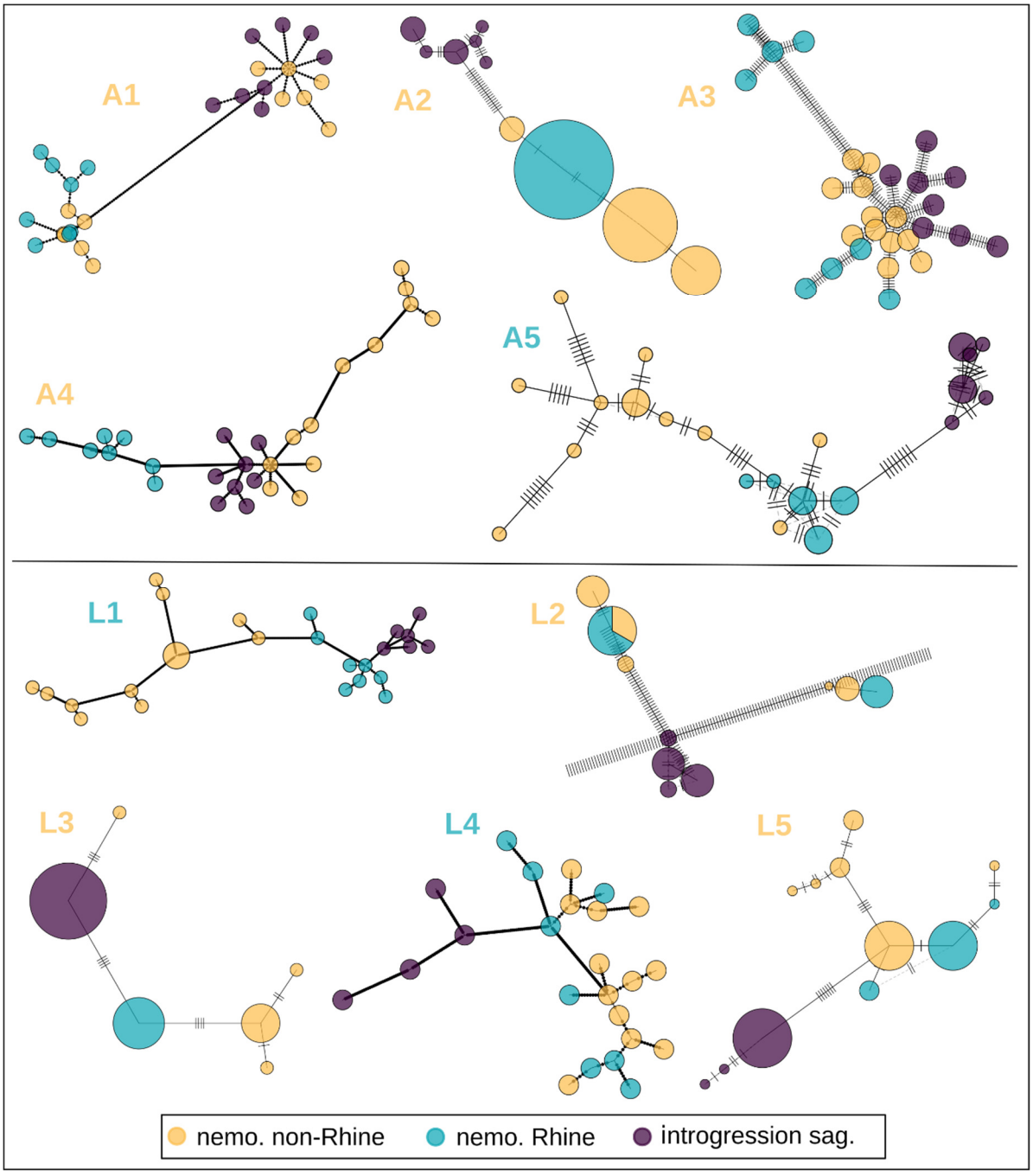

Figure 40: **Ancestral introgression in allopatric *A. sagittata* populations from mixed sources**
Haplotype networks for ten *A. nemorensis* introgression regions in *A. sagittata*, including all potential donor haplotypes and introgression haplotypes. Haplotypes are represented by circles, whose size represents the relative haplotype frequency. Haplotypes are connected by paths with mutation represented as dashes or dots. Haplotypes are colored by species/populations, according to the legend. Labels above each network are colored according to the best *A. nemorensis* match for introgressed haplotypes, according to the legend and gray if matching is ambiguous. Letters in labels indicate population: A – Adl-1, L – Lob.

4 DISCUSSION

4.1 NATURAL CO-VARIATION OF STOMATAL TRAITS AND WATER-USE EFFICIENCY IN *A. THALIANA*

I used high-throughput confocal imaging to characterize stomata patterning in over 31,000 images from 870 samples collected from 330 *A. thaliana* genotypes. The custom high-throughput pipeline could characterize stomata density and stomata size with a reliable accuracy, confirmed by high correlation with manual measurements. Broad-sense heritability and pseudo-heritability estimates for stomata density, which are 30% and 58%, respectively, are slightly lower than in a previous report of manually counted stomata diversity across a smaller sample chosen to maximize genetic diversity (Delgado et al., 2011). Despite the clear impact of environmental (random) variance on both observed phenotypes, stomata size and stomata density showed a strong negative correlation. This is consistent with earlier reports of studies manipulating regulators of stomata development (Doheny-Adams, Hunt, Franks, Beerling, & Gray, 2012; Franks et al., 2015), but also with studies analyzing stomatal trait variation in a wide range of species (Franks & Beerling, 2009; Hetherington & Woodward, 2003).

Both stomata development and reactions to drought stress are being intensively investigated in *A. thaliana* (Bergmann & Sack, 2007; Krasensky & Jonak, 2012; Pillitteri & Torii, 2012; Verslues, Govinal Badiger, Ravi, & M. Nagaraj, 2013). Mutants in stomata density or size have recently been shown to have a clear impact carbon physiology (Franks et al., 2015; Hepworth, Doheny-Adams, Hunt, Cameron, & Gray, 2015; J. Hughes et al., 2017; S. S. Lawson, Pijut, & Michler, 2014; Masle, Gilmore, & Farquhar, 2005; Yoo et al., 2010; H. Yu et al., 2008).Yet, the relevance of natural variation in stomatal patterning for facing local limitations in water availability, had not been documented in this species so far. To examine the impact of variation in stomatal patterning on natural variation in WUE, I obtained concomitant measures of morphological and physiological variation. Analysis of this data in combination with genome-wide patterns of nucleotide diversity resulted in two major findings: i) the decrease in stomata size associates with an increase in WUE in *A. thaliana* and ii) this pattern of co-variation has a genetic basis. This shows that, in *A. thaliana*, variation in stomata size has the potential to be involved in the optimization of physiological processes controlling the trade-

off between growth and water-loss. Interestingly, in the close relative *A. lyrata ssp. lyrata*, stomata were observed to grow smaller in experimental drought compared to well-watered conditions, which coincided with increased WUE (Paccard, Fruleux, & Willi, 2014). This suggest that the consequences of decreased stomata size are conserved in the genus.

The extensive genomic resources available in *A. thaliana* allowed me to investigate the genetic basis of trait variation and co-variation, with the help of GWAS (Atwell et al., 2010). Much is known about the molecular pathways that control the differentiation of stomata in *Arabidopsis thaliana*, providing a set of candidate genes expected to control genetic variation in stomata patterning (Bergmann & Sack, 2007; Pillitteri & Torii, 2012). However, I did not detect any genomic region that associated with stomata density at a p-value beyond the Bonferroni-significance threshold. For stomata size, there was only one significant association on chromosome 4, albeit with very low minor allele frequency in a gene that has not been reported previously in stomata development. GWAS studies can detect small-effect loci only if they segregate at high frequency, whereas rare alleles only give detectable signals when they are of large effect (Korte & Farlow, 2013; Wood et al., 2014). Given that variance for both stomata size and stomata density is clearly heritable, the genetic variants controlling these traits are not causing strong association signals in GWAS. Nevertheless, I can conclude that variation in stomata patterning is controlled by a combination of i) alleles of moderate effect size segregating at frequencies too low to be detected by GWAS, and/or ii) alleles segregating at high frequency but with effect size too small to be detected and/or iii) rare alleles of small effect. In addition, it is possible that the effect of associated loci is weakened by epistatic interactions among loci. In *A. thaliana*, the genetic architecture of natural variation in stomata traits is therefore not caused by a handful of large effect variants but complex and polygenic.

While variation for stomata size and density is likely shaped by a complex genetic architecture that hindered QTL detection, I detected two regions in the genome that associated significantly with carbon isotope discrimination. Three previous QTL-mapping analyses, including one between locally adapted lines from Sweden and Italy, identified 16 distinct QTLs controlling $\delta^{13}C$ (Juenger et al., 2005; McKay et al., 2008; Mojica et al., 2016). One of these is caused by a rare allele in the root-expressed gene *MITOGEN ACTIVATED PROTEIN KINASE 12* (*MPK-12)*, (Campitelli, Des Marais, & Juenger, 2016; Juenger et al., 2005). While QTL-mapping approaches can only reveal the variance shown by the parental lines, GWAS approaches fail

to detect rare alleles unless they have a very strong impact. It is therefore not surprising that the loci that stand out in GWAS do not overlap with the QTL previously mapped. In fact, one of the mapping populations used the parental genotype Cvi-0, a genotypic and phenotypic outlier.

The two QTL reported here on chromosomes 2 and 4 add two novel loci, raising the number of genomic regions known to impact $\delta^{13}C$ in *A. thaliana* to 18. The novel loci are locally frequent. Individuals carrying the minor alleles of both loci are almost exclusively from Southern Sweden and display significantly higher $\delta\ ^{13}C$ than other Southern Swedish accessions. However, no climatic factor significantly correlated with the allelic states of these QTLs. This suggests that other factors, like soil composition, play a role in drought adaptation. Alternatively, locally adapted alleles may not yet be fixed within the region.

Interestingly, the accessions with the minor allele associating with high $\delta\ ^{13}C$ in both QTL did not show decreased stomata size compared to other accessions. Multi-trait GWAS confirmed that these QTL are associated with $\delta\ ^{13}C$ variants that are independent of genetic variation for stomata patterning. Therefore, stomata patterning is only one of the traits contributing to the optimization of WUE. A large array of molecular and physiological reactions is indeed known to contribute to tolerance to drought stress (Krasensky & Jonak, 2012; Verslues et al., 2013). The close vicinity of the chromosome 2 QTL to a non-synonymous mutation in a gene encoding an LEA protein, known to act as a chaperone when cells dehydrate, suggests one possible mechanism by which WUE might be optimized independently of stomata size and density (Candat et al., 2014; Eriksson, Kutzer, Procek, Gröbner, & Harryson, 2011; Reyes et al., 2005). Variation in rates of proline accumulation in the presence of drought stress or in nutrient acquisition in the root are also among the physiological mechanism that appear to have contributed to improve drought stress tolerance in this species (Campitelli et al., 2016; Kesari et al., 2012).

4.2 THE ROLE OF STOMATAL TRAITS AND WATER-USE EFFICIENCY IN LOCAL ADAPTATION

Phenotypic variation for stomata patterning and water-use efficiency is not uniformly distributed throughout the species range. All three phenotypes reported in this study were significantly differentiated between the five broad regions defined in our sample of 330

genotypes. I compared phenotypic and nucleotide levels of divergence to evaluate the putative role of past selective events in shaping the distribution of diversity we report (T. Leinonen et al., 2013; Whitlock & Guillaume, 2009). Because these regions are not equally distant, F_{ST}/Q_{ST} comparisons averaged over all populations may mask local patterns of adaptation (T. Leinonen et al., 2013). I therefore measured Q_{ST} between pairs of regions and compared them to the distribution of pairwise F_{ST}, using permutations to establish the significance of outlier Q_{ST}. For all three traits, differentiation between some regions was stronger than expected from genome-wide patterns of diversity, suggesting local adaptation. This is further supported by my finding that stomata density and stomata size correlated with climatic PCs, which are most strongly driven by temperature, humidity, solar radiation, and historic drought regimen.

The strongest Q_{ST}-F_{ST} differences were found across regional pairs including Central Europe. Particularly, WUE is significantly differentiated between Central Europe and Spain as well as both Swedish regions, due to low WUE in Central Europe. It is tempting to speculate that the significantly lower WUE observed in Central Europe results from selection for life cycling at latitudes where two life cycles can be completed each year, as high WUE is usually associated with a reduction in photosynthetic rate (Blum, 2009; Field et al., 1983; Kimball et al., 2014). Interestingly, Central Europe and Southern Sweden are significantly differentiated for all three traits and Southern Sweden and Spain are significantly differentiated for both stomata traits. Combined with the fact that Swedish genotypes show the highest values for WUE, this suggests that stomata size is involved in drought adaptation of Swedish accessions. This result is somewhat counterintuitive because Sweden is not known to be a region experiencing intense drought. However, our result is supported by an independent study showing that Northern and Southern Swedish genotypes maintain photosynthetic activity under terminal drought stress longer than other, especially Central and Western European, accessions (Exposito-Alonso et al., 2018). Additionally, locally adapted genotypes from Northern Sweden (which showed high WUE in our study, as well) have been shown to display higher WUE than Italian genotypes (Mojica et al., 2016).

This regional difference in *A. thaliana* further coincides with the satellite measurements of historic drought regimen, which show that Sweden is a region where drought frequency is changing throughout the season: it is relatively more frequent in the early growing season

(spring) than in the late growing season (summer). Drought episodes occurring earlier in the growth season may favor the evolution of drought avoidance traits (e.g. morphological or physiological stress adaptations) over that of escape strategies mediated by e.g. seed dormancy (Kooyers, 2015; Passioura, 1996). Indeed, in Northern Europe, increased negative co-variation between flowering time and seed dormancy suggested that the narrow growth season imposes a strong selection on life-history traits (Debieu et al., 2013). In Southern European regions, *A. thaliana* appeared to rely on escape strategies provided by increased seed dormancy (Kronholm et al., 2012). Taken together, this suggests that decreased stomata size and, consequently, increased $\delta^{13}C$ have contributed to adaptation to water limitations in spring in a region where the narrow growth season leaves no room for escape strategies. Indeed, both stomata size and $\delta^{13}C$ associate with historic drought regimen. For $\delta^{13}C$, however, this association disappears when genetic population structure is included as a covariate. This indicates that local adaptation for WUE might have contributed to shape current population structure.

Finally, the coarse regional contrasts used in the present study cannot resolve patterns of local adaptation occurring at a fine-grained scale within regions (as e.g. local adaptation to specific soil patches). In fact, I observe most variation for all three phenotypes within regions. It is therefore possible that the magnitude of adaptive differentiation across the species' European range was underestimated, which could further explain why Q_{ST} / F_{ST} differences did not co-vary with environmental divergence in our dataset.

This work provides a comprehensive description of the variation in stomata size and density that segregates throughout the European range of *A. thaliana*. It shows that stomata size covaries with water-use efficiency and may contribute to local adaptation. Further, it shows that the variation in these traits has a largely polygenic basis and thus it provides an example how local adaptation can shape complex traits in a selfing species, suggesting that polygenic adaptation is possible, even when genetic drift is high and recombination very low. However, in these conditions population genetics techniques might not be able to detect signatures of selection and genetics approaches might be more powerful.

4.3 FREQUENT CO-OCCURRENCE AND MISIDENTIFICATION OF THREE *ARABIS* SIBLING SPECIES

The *Arabis hirsuta* species aggregate is a group of 12 closely related species with complex phylogenetic relationships. Three of these species, *A. nemorensis*, *A. sagittata* and *A. hirsuta,* share a highly similar morphology and are difficult to distinguish on morphology alone (Schmeil et al., 2016). Although the species have different ecological preferences (Karl & Koch, 2014), their distributions partially overlap, e.g. in Germany (Jalas & Suominen, 1994). Thus, the co-occurrence of these species is possible in ecological conditions allowing establishment of both species.

Indeed, in the Upper Rhine populations sampled for the restoration genetics project, I detected two independent genetic clusters, likely representing separate species. Based on ITS sequences, ploidy and morphological descriptors, the first one appears to represent *bona fide A. nemorensis* individuals and the second genetic cluster likely represents *A. sagittata*. I did not anticipate the presence of the *A. sagittata* in flood-plain habitats because it is predominantly found in warm and dry habitats (Hand & Gregor, 2006). However, agricultural land-use and flood regulation have considerably modified the flood-plain ecosystem: flooding is contained by a dyke and the ground water level has decreased (Hölzel & Otte, 2001). Man-made modifications of the environment can change selection regimes and facilitate the establishment of non-native species (J. E. Byers, 2002; Crooks, Chang, & Ruiz, 2011; Fukasawa, Miyashita, Hashimoto, Tatara, & Abe, 2013; Tyrrell & Byers, 2007). Thus, it is possible that the migration of *A. sagittata* into the flood-plain ecosystem was facilitated by the decreased frequency and severity of flooding caused by human river regulation. Indeed, the outcome of my flooding experiments show that *A. sagittata* from the Upper Rhine populations can tolerate flooding for several weeks (in contrast to *A. thaliana*), suggesting that it is somewhat adapted to the floodplain environment. Additional experiments are needed to test whether this flooding resistance is specific to the Upper Rhine populations or universal in *A. sagittata*. If resistance is universal a certain degree flooding resistance might be ancestral to *A. nemorensis* and *A. sagittata*. If resistance is specific to the Rhine population, interspecific gene-flow (see Chapter 5) might have facilitated the floodplain adaptation of this population. Moreover, *A. sagittata* is more frequent than *A. nemorensis* in the sample, which raises the concern that this species may be in the process of displacing *A. nemorensis*. This trend could

be enhanced by climate change, which may lead to conditions that further favor the xero-thermophylic *A. sagittata*.

The surprising finding of two species in the Rhine sites demonstrates that species identification based on morphology can be very problematic in this genus, even in sites that were intensively monitored in ecological studies (Burmeier et al., 2011; Hölzel & Otte, 2003, 2004; Mathar, Kleinebecker, & Hölzel, 2015). If misidentification of species is frequent in flora reports on flood-plain environments, the remaining *A. nemorensis* population in Central Europe might be even smaller than is currently assumed and co-occurrence of sister species might be a more common phenomenon than expected. To investigate this, I visited 25 potential *A. nemorensis* populations in Germany and Austria, which were identified by research in flora databases, herbaria and communication with local botanists. Strikingly, *A. nemorensis* was present in less than half of the visited sites, suggesting that the population in Germany is in further decline. Alternatively, extinction of these populations might only be temporary, as *A. nemorensis* build up a long-term seed bank (Hölzel & Otte, 2004) and is known to disappear from the above ground vegetation only to reappear one or two years later (Mathar et al., 2015). Further, other species were identified as *A. nemorensis* in almost half of the sites, highlighting the value of molecular methods for species identification in complex taxa. Especially in important cases, e.g. when selecting populations for conservation or restoration, complementary molecular species determination is critical. However, so called DNA-barcoding might not be sufficient to reliably delimit species borders especially when gene-flow is occurring (Krishna Krishnamurthy & Francis, 2012). Thus, analyses using genome-wide markers might be required. In such species complex, it is necessary to study all interbreeding species in multiple sympatric and allopatric populations with distinct evolutionary histories so that distribution of ecological specialization can be defined. Only then is it possible to design meaningful conservation goals.

The frequent co-occurrence of species and close geographic proximity of populations of different species, suggests little differentiation for climatic factors among species. Thus, local adaptation to climate within species might be a more important factor. It would be interesting to test whether local adaptation to climate has led to similar adaptations in the sister species. A good candidate phenotype to test this would be flowering time, which was shown to vary with latitude in other Brassicaceae species (Neuffer & Hoffrogge, 1999; Paccard et al., 2014;

Stinchcombe et al., 2004; Toräng et al., 2015). However, this will require a larger set of populations/genotypes from a wider geographic range.

4.4 EFFECTS OF ECOLOGICAL RESTORATION BY HAY-TRANSFER ON GENETIC DIVERSITY

Hay-transfer has proven to be a particularly successful method for establishing new populations of target species in ecological restoration across a variety of herbaceous vegetation types (Coiffait-Gombault, Buisson, & Dutoit, 2011; Hölzel & Otte, 2003; Kiehl et al., 2010; Kiehl & Wagner, 2006; Török et al., 2012). Furthermore, hay-transfer is often seen as the gold standard to preserve local levels of genetic diversity and adaptation (Vander Mijnsbrugge, Bischoff, & Smith, 2010). The latter is probably the main reason, why it is increasingly used in in ecological restoration (Kiehl et al., 2014). One aim of this project was to characterize the level of diversity in the pristine source sites and document the impact of hay-transfer on the genetic diversity in restored sites. Although the initial plan was to focus on *A. nemorensis*, a typical representative of species-rich floodplain meadows, the unanticipated presence of *A. sagittata* in the sample allowed to compare the genetic effects of restoration by hay-transfer on the two species.

Several years after restoration, I did not find a significant difference in genetic diversity between pristine and restored sites for either of the two species. Thus, the hay-transfer method can restore populations with levels of diversity indistinguishable from the source populations in the long-term. This is in agreement with studies on population life-stage structure and dynamics comparing pristine and restored sites of *A. nemorensis/A. sagittata* in the same region (Burmeier et al., 2011). This outcome is particularly remarkable given that *A. nemorensis* has a long-term seed bank, with up to 25,000 germinable seeds*m^{-2}, which was not transferred to restored sites (Burmeier et al., 2011). Although species for which the genetic diversity present in the seed bank tends to differ more strongly from the above ground diversity may fare differently after hay-transfer, we note that populations restored with alternative methods, e.g. spontaneous recolonization (Vandepitte et al., 2012) or propagated seed mixtures (Espeland et al., 2017; Fant, Holmstrom, Sirkin, Etterson, & Masi, 2008), both excluded the seed bank and revealed a reduction of diversity (Mijangos et al., 2015). Thus,

hay-transfer might be superior to other restoration methods not only in restoration success (Hölzel & Otte, 2003; Kiehl et al., 2010) but also in transferring genetic diversity.

4.5 A REFERENCE-GENOME IS NOT REQUIRED TO CHARACTERIZE PATTERNS OF GENETIC DIVERSITY

While RAD-seq and related methods are a cheap tool to acquire genotype information across the genome without need for a reference genome (Elshire et al., 2011; Etter et al., 2011; Peterson et al., 2012), the reliability of *de novo* assembly pipelines has been questioned (Shafer et al., 2016). The availability of a reference genome allowed us to ask whether reference-based read mapping pipelines yielded distinct conclusions from a pipeline based on *de novo* read assembly. Conversely, this also allowed verifying that the use of an *A. nemorensis* reference genome to map *A. sagittata* samples did not bias the outcome of our study. The results from both pipelines were highly correlated. Thus, comparative analysis of sites was reliable with both pipelines. However, the *de-novo*-pipeline (*Stacks*, Catchen et al., 2013) underestimated the amount of genetic diversity compared to the reference-based pipeline. This is in contrast to a previous study comparing different RAD-seq pipelines (Shafer et al., 2016), where *Stacks* produced slightly inflated estimates compared to reference-based pipelines. However, in the same study, other de novo pipelines produced substantially lower estimates of genetic diversity (Shafer et al., 2016). Thus, the magnitude of genetic diversity estimates might depend on the study system and pipelines/parameters used, and caution is advised when comparing these estimates across studies.

In contrast, both pipelines agreed for analyses comparing diversity between species or sites (e.g. F_{ST}, ADMIXTURE). Thus, we conclude that RAD-seq is an efficient tool to characterize the distribution of diversity, even in the absence of a reference genome. We therefore hope that this study will pave the way for exploring how species, with diverse life-history and uncharacterized genomes, will be maintained after hay transfer or how modalities of hay transfer affects not only single species but the balance between multiple species in the community.

4.6 PATTERNS OF GENETIC DIVERSITY IN THREE *ARABIS* SPECIES

4.6.1 LOW GENETIC DIVERSITY IN *A. NEMORENSIS* AND *A. SAGITTATA* AND EXCESS HETEROZYGOSITY IN *A. HIRSUTA*

Genetic diversity was low across sampled populations of *A. nemorensis* and *A. sagittata,* compared to other outcrossing or selfing Brassicaceae species (Alonso-Blanco et al., 2016; Mattila, Tyrmi, Pyhäjärvi, & Savolainen, 2017; Onge, Källman, Slotte, Lascoux, & Palmé, 2011). In fact, genetic distance between species was only on the order of magnitude as diversity found within populations of *Arabidopsis lyrata* (Mattila et al., 2017). Genetic diversity within both species was similar to that reported for *Arabis alpina* populations in Scandinavia (Laenen et al., 2018). As my study populations, these populations are selfing and located on the margin of the species' range (Jalas & Suominen, 1994), two factors often coinciding with lower levels of genetic diversity. Thus, additional sampling in the center of each species' distribution is required to assess the remaining genetic diversity within each species. In fact, the relatively strong divergence of the Bot population, which is the only genotype in our collection originating from the Elb riverbanks, suggests that additional divergent clusters might still exist.

For *A. hirsuta*, genetic diversity within species and genetic distance to the other species was two orders of magnitude higher. However, very high levels of heterozygosity, indicated by a strongly negative F_{IS}, suggest that this species might be an allopolyploid originating from two divergent species. Allopolyploidy prevents intergenomic recombination and results in fixed heterozygosity (Comai, 2005). Thus, the estimates of genetic diversity are not comparable to the sister species.

4.6.2 STRONG POPULATION STRUCTURE IN *A. NEMORENSIS* POPULATIONS

The larger number of sampled populations of *A. nemorensis*, allowed me to study the effects of habitat fragmentation on population structure in more detail. Habitat fragmentation often leads to reduced genetic variation and strong population structure due to increased drift and reduced migration (Lienert, 2004). Most populations, except Rhine and Adl-2, consisted only of a single genotype and thus populations were strongly differentiated, indicating no genetic exchange. In Adl-2, two distinct genotypes were present, but I did not detect admixed individuals, which could be a sign of recent migration. The Rhine population was most diverse,

which might be a consequence of its relatively large size (and reduced genetic drift) or the ecological management in this area.

Based on genetic distance among populations, I identified three genetic clusters. Two of these clusters contained geographically close populations, but the other cluster contained distant populations, suggesting complex ancestral migration patterns. Populations Rhine and Bot did not cluster with any other population, suggesting they represent independent lineages of *A. nemorensis*. Overall, population structure correlated with geographic distance but not climatic differences among populations, suggesting that climate adaptation does not contribute to the distribution of genotypes of *A. nemorensis* in Germany. These results suggest that the degradation of the floodplain habitat has led to strong fragmentation of *A. nemorensis* populations, which in turn has led to a strong reduction of genetic diversity by genetic drift.

Small, isolated populations with low genetic diversity may suffer from increased genetic load and decreased adaptive potential (Bijlsma & Loeschcke, 2012; D. L. Byers & Waller, 1999). Indeed, in a meta-analysis in plants, local adaptation was found to be more frequent in large populations (Leimu & Fischer, 2008). Genetic rescue, i.e. the transfer of non-local material, is often envisaged as a means to preserve endangered species (Breed, Stead, Ottewell, Gardner, & Lowe, 2013; Weeks et al., 2011; Whiteley, Fitzpatrick, Funk, & Tallmon, 2015). In a meta-analysis of genetic rescue projects in small inbred populations, a significant increase in fitness was observed in 92% of cases (Frankham, 2015). This approach is however controversial. Strategies that introduce non-local seeds can decrease population fitness either by introducing maladapted genotypes (Crémieux, Bischoff, Müller-Schärer, & Steinger, 2010; McKay, Christian, Harrison, & Rice, 2005) or by causing outbreeding depression (Frankham et al., 2011). As a compromise, a strategy of mixing regional seeds was recently proposed (Bucharova et al., 2018). My results show that this strategy was incidentally implemented in the examined restoration effort: ADMXITURE analysis showed that pristine *A. sagittata* sites are dominated by one or two ancestral groups, with very low genetic diversity within each group. Yet, in some of the restored sites, admixture of low diversity *A. sagittata* groups took place, leading to a strong increase in genetic diversity and decrease in population structure. ADMIXTURE analysis also revealed that this led to genetic recombination in *A. sagittata,* possibly increasing the adaptive potential of restored sites in this species. Increased genetic connectivity between populations has indeed been shown to help maintain or even increase

genetic diversity in the long term (DiLeo, Rico, Boehmer, & Wagner, 2017). Fitness assays of recombined *A. sagittata* genotypes are needed to verify whether this local admixture has reinforced the establishment of this species in restored floodplain meadows. Interestingly, I find less pronounced population structure in pristine sites of *A. nemorensis* than in *A. sagittata*, suggesting that a higher level of gene flow helps maintain diversity within pristine sites. The restoration of this species will therefore not directly benefit from post-transfer admixture. More so, my results suggest that it is possible that *A. nemorensis* is at increased disadvantage in restored habitats, if admixture and resulting recombination were to favor the emergence of more competitive genotypes were to evolve in *A. sagittata* thanks to admixture.

4.7 EVOLUTIONARY CONSEQUENCES OF HYBRIDIZATION BETWEEN *A. NEMORENSIS* AND *A. SAGITTATA*

Co-occurrence of sister species from the *A. hirsuta* group appears to be relatively common (26% of sites) and hybridization occurred between two pairs of species. Natural hybrids between *A. hirsuta* and *A. nemorensis* appeared to be completely sterile, which is likely due to the difference in ploidy. Crossing experiments between the species also reported complete sterility of F1 hybrids (Titz, 1979). Thus, gene-flow between these two species is unlikely, although not impossible (Chapman & Abbott, 2010). Co-occurrence between *A. nemorensis* and *A. sagittata* was only found in the Rhine populations. While hybrids between *A. nemorensis* and *A. sagittata* have been reported (Novotná & Czapik, 1974), I could not find reports of hybridization as frequent as observed in these populations. Thus, the Rhine population might be a hotspot of genetic exchange between the two species and a source of novel genetic variation for either species. Future studies will have to examine the evolutionary dynamics of genetic exchange to determine the impact of this hotspot on contemporary evolution of *A. nemorensis* and *A. sagittata*.

Hybridization of *A. nemorensis* and *A. sagittata* can have different evolutionary consequences depending on the circumstances. The first question is whether hybrids are fertile. RAD-seq analysis of 24 natural hybrids revealed different ancestry proportions of the parental species in hybrid genomes. This suggests that hybrids are not the F1 generation, which would have equal proportions of parental ancestry, and are thus fertile. Artificial crosses I performed recently have shown that F1 hybrids indeed have low levels of fertility and that fertility

segregates among F2 hybrids from sterile to fully fertile (data not shown). The same result was found in an earlier study of artificial crosses in the *A. hirsuta* group (Titz, 1979). Ancestry proportions of natural hybrids are strongly skewed towards *A. sagittata*, suggesting increased back-crossing with this parent. Moreover, several hybrids show very high proportions of *A. sagittata* ancestry, suggesting several rounds of back-crossing have occurred. This directional back-crossing could be due to increased outcrossing rates of *A. sagittata*, which is supported by a lower F_{IS} than *A. nemorensis* and larger, more attractive flowers (personal observation). However, it could also be simply due to a higher number of *A. sagittata* individuals in the population, which increases the opportunity of back-crossing with it. Finally, it is possible that genetic incompatibilities inhibit back-crosses with *A. nemorensis*, which could be tested in artificial crossing experiments in the future, as was done for example in oak (Olrik & Kjaer, 2007). However, back-crossing with *A. nemorensis* cannot be ruled out at this point, as hybrids with more than 50% *A. nemorensis* ancestry were also found. Back-crossing of hybrids over several generations can result in gene-flow between parental species, i.e. the introgression of small genomic fragments (Goulet et al., 2017). Introgression can be a significant source of potentially adaptive variation (Suarez-Gonzalez et al., 2018), which might be critical to survival of species with low levels of standing variation like *A. nemorensis* and *A. sagittata*.

I analyzed whole genome sequences of parental individuals to test for interspecific gene-flow and locate genomic regions of introgression. I found signatures of significant gene-flow among species. Note, that levels of gene-flow might be underestimated here as shared introgressions between all populations, cannot be detected using ABBA-BABA statistics (Green et al., 2010). Thus, additional populations outside of the species' range-overlap should be sampled in the future to improve these estimates. Gene-flow was stronger from *A. nemorensis* to *A. sagittata* than vice versa, as expected from asymmetrical backcross patterns. If *A. sagittata* is indeed replacing *A. nemorensis* in floodplain habitats, the genetic variation gained through geneflow could prove an additional advantage and facilitate the replacement. Asymmetrical gene-flow was also found in another pair of *Brassicaceae* floodplain species, *Rorippa amphibia* and *Rorippa palustris* (W. Bleeker & Hurka, 2001). As in my study, contact between species might have been facilitated by human activity in flood plain habitats, as it was only observed in strongly altered landscapes. In this example individuals carrying introgressions, which was

determined by allozyme and chloroplast markers, showed an altered leaf phenotype compared to pure individuals.

Interestingly, gene-flow between species was not limited to the sympatric population. In *A. sagittata*, introgression frequency was indeed highest in the sympatric population, but I also detected introgressions in both allopatric populations. In *A. nemorensis*, introgression frequency was even higher in the allopatric than sympatric populations. Additionally, analysis of introgressed regions revealed that some introgressed regions were shared among allopatric and sympatric populations. These results suggest that ancestral gene-flow between the species has occurred and that it likely also explains at least part of the introgressions found in the sympatric *A. sagittata* population. Ancestral gene-flow was also found in the *Arabidopsis* genus and thus might be common in closely related *Brassicaceae* species (Novikova et al., 2016). To differentiate between ancestral and recent gene-flow, I conducted haplotype network analysis for introgressed regions of sympatric and allopatric *A. sagittata* populations with the aim of determining the source of each introgressed fragment. My expectation was that if gene-flow is ongoing or recent in the sympatric population, a large proportion of introgressed haplotypes should match best to sympatric *A. nemorensis* haplotypes, compared to allopatric populations. The proportion of regions with haplotypes uniquely matching to *A. nemorensis* Rhine was only 13% higher in the sympatric than the allopatric *A. sagittata* population. Yet, in only 19% of introgressed regions in the sympatric *A. sagittata* population all introgressed haplotypes matched uniquely to allopatric *A. nemorensis* haplotypes, while in the allopatric *A. sagittata* populations, this number was 70%. Moreover, in the sympatric population, in 19% of introgressed regions some haplotypes matched closest to a sympatric and some closest to some allopatric *A. nemorensis* haplotype, suggesting that ancestral and recent introgression have occurred consecutively in the same genomic regions. These results suggest that recent introgression contributed to the increased introgression frequency observed on the sympatric *A. sagittata* population. In future analyses, the relative magnitude of ancestral and recent gene-flow could be estimated using demographic modelling (Payseur & Rieseberg, 2016; Sousa & Hey, 2013).

Moreover, I focused the analysis so far on introgressed regions with relatively high frequency (high f_D), which are more likely to be ancestral. Thus, targeted analysis of low frequency introgressions could reveal more recent introgressions improving the estimate of the

magnitude of recent gene-flow. However, finding these introgressions is more difficult as one cannot rely on f_D peaks.

While my results are clear evidence of introgressive gene-flow between *A. nemorensis* and *A. sagittata*, further research is needed to determine whether it has adaptive value. High introgression frequency was suggested as an indicator for adaptive introgression (Racimo et al., 2017). Indeed, I find several high-frequency introgression regions, especially in the Rhine *A. sagittata* population. However, this is not enough evidence by itself to conclude that introgressions are adaptive, as high-frequency introgressions can also arise from neutral demographic processes (Suarez-Gonzalez et al., 2018). My common garden experiment and flooding experiments have shown that the parental species show divergence in multiple, potentially adaptive phenotype: phenology, morphology, herbivory tolerance and flooding resistance. Thus, introgression has the potential to transfer this adaptive genetic variance between species. However, the link between introgressions and phenotypes must be established. One way to do this is to determine the genetic basis of the phenotypic variation via a QTL-mapping approach of interspecific F2 hybrids and test if it overlaps with introgressed regions. This analysis would also reveal phenotypes showing transgressive segregation, which might have adaptive value in hybrid populations (Goulet et al., 2017). A similar approach was taken to link introgressions to flooding tolerance in the naturally hybridizing *Iris fulva* and *Iris brevicaulis* (N. H. Martin, Bouck, & Arnold, 2006). In this study, flooding tolerance of reciprocal backcross lines was determined in a field experiment in a highly selective (frequently flooded) environment. *I. fulva* survival rates were higher than *I. brevicaulis* and survival rates in back-cross lines was strongly associated with the presence of *I. fulva* alleles in several regions of the genome, demonstrating the potential of introgression to confer flooding tolerance. This study also highlights the value of measuring hybrid fitness *in situ* compared to artificial environments. Thus, *in situ* analysis of artificial hybrid populations should be considered for *Arabis* as well. Population genetics can also be a tool to link introgressions to adaptation. In *Arabidopsis arenosa*, two selective sweep regions conferring adaptation to serpentine soil were shown to originate from the sister species *Arabidopsis lyrata* (B. J. Arnold et al., 2016). Population genetic approaches might have limited power in my study system due to the strong population structure, which can create noisy signals.

4.8 CONCLUSION

The aim of this thesis was to investigate how evolutionary forces shaped natural variation in two selfing *Brassicaceae* species. I detected strong natural variation and signatures of local adaptation in complex traits in the selfing *A. thaliana*, showing that adaptation can shape natural variation in these traits despite strong drift and low recombination. In contrast to the human commensal *A. thaliana*, in the strongly endangered floodplain species *A. nemorensis* genetic diversity was low and populations were highly fragmented. Thus, the potential for adaptation in this species might be strongly reduced. Yet, I found that ecological restoration measures, specifically hay-transfer, can efficiently transfer genetic diversity to restore populations and even increase genetic variation through admixture. Moreover, I found a natural hybrid zone between *A. nemorensis* and its sibling species *A. sagittata*. Hybridization might be a source of novel genetic variation for *A. nemorensis* through interspecific gene-flow. However, my results suggest that gene-flow is strongly asymmetrical. If it were to preferentially increase the adaptive potential of *A. sagittata*, it may even accelerate the replacement of *A. nemorensis.* Under such scenario, adaptive genetic variation derived in *A. nemorensis* might persist even if the species goes extinct. My work thus illustrates the necessity to study species diversity in a global phylogenetic and geographical context in order to determine realistic and efficient goals for species and ecosystem conservation.

5 LITERATURE

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Adrion, J. R., Hahn, M. W., & Cooper, B. S. (2015). Revisiting classic clines in Drosophila melanogaster in the age of genomics. *Trends in Genetics*, *31*(8), 434–444. https://doi.org/10.1016/j.tig.2015.05.006

Ågren, J., Oakley, C. G., McKay, J. K., Lovell, J. T., & Schemske, D. W. (2013). Genetic mapping of adaptation reveals fitness tradeoffs in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *110*(52), 21077–21082. https://doi.org/10.1073/pnas.1316773110

Ågren, J., & Schemske, D. W. (2012). Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. *New Phytologist*, *194*(4), 1112–1122. https://doi.org/10.1111/j.1469-8137.2012.04112.x

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. https://doi.org/10.1101/gr.094052.109

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology*, *26*(2), 420–430. https://doi.org/10.1111/mec.13948

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709. https://doi.org/10.1038/nrg2844

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., … Zhou, X. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*, *166*(2), 481–491. https://doi.org/10.1016/j.cell.2016.05.063

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson, V. J., & Briske, D. D. (1990). Stomatal Distribution, Density and Conductance of Three Perennial Grasses Native to the Southern True Prairie of Texas. *The American Midland Naturalist*, *123*(1), 152–159. https://doi.org/10.2307/2425768

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data (Version 0.11.3). Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, *408*(6814), 796–815. https://doi.org/10.1038/35048692

Arnold, B. J., Lahner, B., DaCosta, J. M., Weisman, C. M., Hollister, J. D., Salt, D. E., … Yant, L. (2016). Borrowed alleles and convergence in serpentine adaptation. *Proceedings of the National Academy of Sciences*, *113*(29), 8320–8325. https://doi.org/10.1073/pnas.1600405113

Arnold, J. B., Daroczi, G., Werth, B., Weitzner, B., Kunst, J., Auguie, B., … London, J. (2017). ggthemes: Extra Themes, Scales and Geoms for “ggplot2” (Version 3.4.0). Retrieved from https://cran.r-project.org/web/packages/ggthemes/index.html

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., … Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, *465*(7298), 627–631. https://doi.org/10.1038/nature08800

Bakker, J. C. (1991). Effects of humidity on stomatal density and its relation to leaf conductance. *Scientia Horticulturae*, *48*(3), 205–212. https://doi.org/10.1016/0304-4238(91)90128-L

Baquero, O. S. (2017). ggsn: North Symbols and Scale Bars for Maps Created with "ggplot2" or "ggmap" (Version 0.4.0). Retrieved from https://CRAN.R-project.org/package=ggsn

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

Barton, N. H. (2010). What role does natural selection play in speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1547), 1825–1840. https://doi.org/10.1098/rstb.2010.0001

Barton, N., & Partridge, L. (2000). Limits to natural selection. *BioEssays*, *22*(12), 1075–1084. https://doi.org/10.1002/1521-1878(200012)22:12<1075::AID-BIES5>3.0.CO;2-M

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis*, *53*(8), 474–485. https://doi.org/10.1002/dvg.22877

Bergmann, D. C., & Sack, F. D. (2007). Stomatal Development. *Annual Review of Plant Biology*, *58*(1), 163–181. https://doi.org/10.1146/annurev.arplant.58.032806.104023

Bierne, N., Roze, D., & Welch, J. J. (2013). Pervasive selection or is it…? why are FST outliers sometimes so frequent? *Molecular Ecology*, *22*(8), 2061–2064. https://doi.org/10.1111/mec.12241

Bijlsma, R., & Loeschcke, V. (2012). Genetic erosion impedes adaptive responses to stressful environments. *Evolutionary Applications*, *5*(2), 117–129. https://doi.org/10.1111/j.1752-4571.2011.00214.x

Bleeker, W., & Hurka, H. (2001). Introgressive hybridization in Rorippa (Brassicaceae): gene flow and its consequences in natural and anthropogenic habitats. *Molecular Ecology*, *10*(8), 2013–2022. https://doi.org/10.1046/j.1365-294X.2001.01341.x

Bleeker, Walter, Schmitz, U., & Ristow, M. (2007). Interspecific hybridisation between alien and native plant species in Germany and its consequences for native biodiversity. *Biological Conservation*, *137*(2), 248–253. https://doi.org/10.1016/j.biocon.2007.02.004

Blum, A. (2009). Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. *Field Crops Research*, *112*(2), 119–123. https://doi.org/10.1016/j.fcr.2009.03.009

Böndel, K. B., Lainer, H., Nosenko, T., Mboup, M., Tellier, A., & Stephan, W. (2015). North–South Colonization Associated with Local Adaptation of the Wild Tomato Species *Solanum chilense*. *Molecular Biology and Evolution*, *32*(11), 2932–2943. https://doi.org/10.1093/molbev/msv166

Bradbury, I. R., Hamilton, L. C., Dempson, B., Robertson, M. J., Bourret, V., Bernatchez, L., & Verspoor, E. (2015). Transatlantic secondary contact in Atlantic Salmon, comparing

microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. *Molecular Ecology*, *24*(20), 5130–5144. https://doi.org/10.1111/mec.13395

Bratteler, M., Lexer, C., & Widmer, A. (2006). Genetic architecture of traits associated with serpentine adaptation of Silene vulgaris. *Journal of Evolutionary Biology*, *19*(4), 1149–1156. https://doi.org/10.1111/j.1420-9101.2006.01090.x

Breed, M. F., Stead, M. G., Ottewell, K. M., Gardner, M. G., & Lowe, A. J. (2013). Which provenance and where? Seed sourcing strategies for revegetation in a changing environment. *Conservation Genetics*, *14*(1), 1–10. https://doi.org/10.1007/s10592-012-0425-z

Browne, L., Ottewell, K., & Karubian, J. (2015). Short-term genetic consequences of habitat loss and fragmentation for the neotropical palm *Oenocarpus bataua*. *Heredity*, *115*(5), 389–395. https://doi.org/10.1038/hdy.2015.35

Browning, B. L., & Browning, S. R. (2009). A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *The American Journal of Human Genetics*, *84*(2), 210–223. https://doi.org/10.1016/j.ajhg.2009.01.005

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics*, *103*(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Browning, S. R., & Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *American Journal of Human Genetics*, *81*(5), 1084–1097.

Bucharova, A., Bossdorf, O., Hölzel, N., Kollmann, J., Prasse, R., & Durka, W. (2018). Mix and match: regional admixture provenancing strikes a balance among different seed-sourcing strategies for ecological restoration. *Conservation Genetics*. https://doi.org/10.1007/s10592-018-1067-6

Bucharova, A., Michalski, S., Hermann, J.-M., Heveling, K., Durka, W., Hölzel, N., … Bossdorf, O. (2017). Genetic differentiation and regional adaptation among seed origins used for grassland restoration: lessons from a multispecies transplant experiment. *Journal of Applied Ecology*, *54*(1), 127–136. https://doi.org/10.1111/1365-2664.12645

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

Burmeier, S., Eckstein, R. L., Donath, T. W., & Otte, A. (2011). Plant Pattern Development during Early Post-Restoration Succession in Grasslands—A Case Study of Arabis nemorensis. *Restoration Ecology*, *19*(5), 648–659. https://doi.org/10.1111/j.1526-100X.2010.00668.x

Busch, V., & Reisch, C. (2016). Population size and land use affect the genetic variation and performance of the endangered plant species Dianthus seguieri ssp. glaber. *Conservation Genetics*, *17*(2), 425–436. https://doi.org/10.1007/s10592-015-0794-1

Byers, D. L., & Waller, D. M. (1999). Do Plant Populations Purge Their Genetic Load? Effects of Population Size and Mating History on Inbreeding Depression. *Annual Review of Ecology and Systematics*, *30*(1), 479–513. https://doi.org/10.1146/annurev.ecolsys.30.1.479

Byers, J. E. (2002). Impact of non-indigenous species on natives enhanced by anthropogenic alteration of selection regimes. *Oikos*, *97*(3), 449–458. https://doi.org/10.1034/j.1600-0706.2002.970316.x

Campitelli, B. E., Des Marais, D. L., & Juenger, T. E. (2016). Ecological interactions and the fitness effect of water-use efficiency: Competition and drought alter the impact of natural MPK12 alleles in Arabidopsis. *Ecology Letters*, *19*(4), 424–434. https://doi.org/10.1111/ele.12575

Candat, A., Paszkiewicz, G., Neveu, M., Gautier, R., Logan, D. C., Avelange-Macherel, M.-H., & Macherel, D. (2014). The Ubiquitous Distribution of Late Embryogenesis Abundant Proteins across Cell Compartments in Arabidopsis Offers Tailored Protection against Abiotic Stress. *The Plant Cell*, *26*(7), 3148–3166. https://doi.org/10.1105/tpc.114.127316

Carlson, J. E., Adams, C. A., & Holsinger, K. E. (2016). Intraspecific variation in stomatal traits, leaf traits and physiology reflects adaptation along aridity gradients in a South African shrub. *Annals of Botany*, *117*(1), 195–207. https://doi.org/10.1093/aob/mcv146

Carlson, S. M., Cunningham, C. J., & Westley, P. A. H. (2014). Evolutionary rescue in a changing world. *Trends in Ecology & Evolution*, *29*(9), 521–530. https://doi.org/10.1016/j.tree.2014.06.005

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Cavalli-Sforza, L. L., & Edwards, A. W. F. (1967). Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, *19*(3 Pt 1), 233–257.

Chapman, M. A., & Abbott, R. J. (2010). Introgression of fitness genes across a ploidy barrier. *New Phytologist*, *186*(1), 63–71. https://doi.org/10.1111/j.1469-8137.2009.03091.x

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*(4), 1289–1303.

Charlesworth, Brian, & Charlesworth, D. (2010). *Elements of Evolutionary Genetics*. Roberts and Company Publishers.

Chater, C., Kamisugi, Y., Movahedi, M., Fleming, A., Cuming, A. C., Gray, J. E., & Beerling, D. J. (2011). Regulatory mechanism controlling stomatal behavior conserved across 400 million years of land plant evolution. *Current Biology: CB*, *21*(12), 1025–1029. https://doi.org/10.1016/j.cub.2011.04.032

Coiffait-Gombault, C., Buisson, E., & Dutoit, T. (2011). Hay Transfer Promotes Establishment of Mediterranean Steppe Vegetation on Soil Disturbed by Pipeline Construction. *Restoration Ecology*, *19*(201), 214–222. https://doi.org/10.1111/j.1526-100X.2010.00706.x

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, *6*(11), 836. https://doi.org/10.1038/nrg1711

Coop, G., Witonsky, D., Rienzo, A. D., & Pritchard, J. K. (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, *185*(4), 1411–1423. https://doi.org/10.1534/genetics.110.114819

Cowan, I. R. (1986). Economics of carbon fixation in higher plants. *On the Economy of Plant Form and Function : Proceedings of the Sixth Maria Moors Cabot Symposium, Evolutionary Constraints on Primary Productivity, Adaptive Patterns of Energy Capture in Plants, Harvard Forest, August 1983*. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=US201301399273

Cowan, I. R., & Farquhar, G. D. (1977). Stomatal function in relation to leaf metabolism and environment. *Symposia of the Society for Experimental Biology*, *31*, 471–505.

Crémieux, L., Bischoff, A., Müller-Schärer, H., & Steinger, T. (2010). Gene flow from foreign provenances into local plant populations: Fitness consequences and implications for biodiversity restoration. *American Journal of Botany*, *97*(1), 94–100. https://doi.org/10.3732/ajb.0900103

Crooks, J. A., Chang, A. L., & Ruiz, G. M. (2011). Aquatic pollution increases the relative success of invasive species. *Biological Invasions*, *13*(1), 165–176. https://doi.org/10.1007/s10530-010-9799-3

Crow, J. F. (2010). Wright and Fisher on Inbreeding and Random Drift. *Genetics*, *184*(3), 609–611. https://doi.org/10.1534/genetics.109.110023

Crutsinger, G. M., Collins, M. D., Fordyce, J. A., Gompert, Z., Nice, C. C., & Sanders, N. J. (2006). Plant Genotypic Diversity Predicts Community Structure and Governs an Ecosystem Process. *Science*, *313*(5789), 966–968. https://doi.org/10.1126/science.1128326

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Daszkowska-Golec, A., & Szarejko, I. (2013). Open or Close the Gate – Stomata Action Under the Control of Phytohormones in Drought Stress Conditions. *Frontiers in Plant Science*, *4*. https://doi.org/10.3389/fpls.2013.00138

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

Delgado, D., Alonso-Blanco, C., Fenoll, C., & Mena, M. (2011). Natural variation in stomatal abundance of Arabidopsis thaliana includes cryptic diversity for different developmental processes. *Annals of Botany*, *107*(8), 1247–1258. https://doi.org/10.1093/aob/mcr060

DiLeo, M. F., Rico, Y., Boehmer, H. J., & Wagner, H. H. (2017). An ecological connectivity network maintains genetic diversity of a flagship wildflower, Pulsatilla vulgaris. *Biological Conservation*, *212*, 12–21. https://doi.org/10.1016/j.biocon.2017.05.026

Doheny-Adams, T., Hunt, L., Franks, P. J., Beerling, D. J., & Gray, J. E. (2012). Genetic manipulation of stomatal density influences stomatal size, plant growth and tolerance to restricted water supply across a growth carbon dioxide gradient. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, *367*(1588), 547–555. https://doi.org/10.1098/rstb.2011.0272

Donath, T. W., Bissels, S., Hölzel, N., & Otte, A. (2007). Large scale application of diaspore transfer with plant material in restoration practice – Impact of seed and microsite limitation. *Biological Conservation*, *138*(1), 224–234. https://doi.org/10.1016/j.biocon.2007.04.020

Drake, P. L., Froend, R. H., & Franks, P. J. (2013). Smaller, faster stomata: scaling of stomatal size, rate of response, and stomatal conductance. *Journal of Experimental Botany*, *64*(2), 495–505. https://doi.org/10.1093/jxb/ers347

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252. https://doi.org/10.1093/molbev/msr048

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Eriksson, S. K., Kutzer, M., Procek, J., Gröbner, G., & Harryson, P. (2011). Tunable membrane binding of the intrinsically disordered dehydrin Lti30, a cold-induced plant stress protein. *The Plant Cell*, *23*(6), 2391–2404. https://doi.org/10.1105/tpc.111.085183

Espeland, E. K., Emery, N. C., Mercer, K. L., Woolbright, S. A., Kettenring, K. M., Gepts, P., & Etterson, J. R. (2017). Evolution of plant materials for ecological restoration: insights from the applied and basic literature. *Journal of Applied Ecology*, *54*(1), 102–115. https://doi.org/10.1111/1365-2664.12739

Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology (Clifton, N.J.)*, *772*, 157–178. https://doi.org/10.1007/978-1-61779-228-1_9

Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H. A., & Weigel, D. (2018). Genomic basis and evolutionary potential for extreme drought adaptation in Arabidopsis thaliana. *Nature Ecology & Evolution*, *2*(2), 352. https://doi.org/10.1038/s41559-017-0423-0

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

Fan, S., Hansen, M. E. B., Lo, Y., & Tishkoff, S. A. (2016). Going global by adapting local: A review of recent human adaptation. *Science*, *354*(6308), 54–59. https://doi.org/10.1126/science.aaf5098

Fant, J. B., Holmstrom, R. M., Sirkin, E., Etterson, J. R., & Masi, S. (2008). Genetic Structure of Threatened Native Populations and Propagules Used for Restoration in a Clonal Species, American Beachgrass (Ammophila breviligulata Fern.). *Restoration Ecology*, *16*(4), 594–603. https://doi.org/10.1111/j.1526-100X.2007.00348.x

Farquhar, G. D., Hubick, K. T., Condon, A. G., & Richards, R. A. (1989). Carbon Isotope Fractionation and Plant Water-Use Efficiency. In P. W. Rundel, J. R. Ehleringer, & K. A. Nagy (Eds.), *Stable Isotopes in Ecological Research* (pp. 21–40). https://doi.org/10.1007/978-1-4612-3498-2_2

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*.

Field, C., Merino, J., & Mooney, H. A. (1983). Compromises between Water-Use Efficiency and Nitrogen-Use Efficiency in Five Species of California Evergreens. *Oecologia*, *60*(3), 384–389.

Fishman, L., & Wyatt, R. (1999). Pollinator-Mediated Competition, Reproductive Character Displacement, and the Evolution of Selfing in Arenaria Uniflora (caryophyllaceae). *Evolution*, *53*(6), 1723–1733. https://doi.org/10.1111/j.1558-5646.1999.tb04557.x

Fitzgibbon, J., Beck, M., Zhou, J., Faulkner, C., Robatzek, S., & Oparka, K. (2013). A Developmental Framework for Complex Plasmodesmata Formation Revealed by Large-Scale Imaging of the Arabidopsis Leaf Epidermis. *The Plant Cell Online*, *25*(1), 57–70. https://doi.org/10.1105/tpc.112.105890

Fourcade, Y., Chaput-Bardy, A., Secondi, J., Fleurant, C., & Lemaire, C. (2013). Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. *Molecular Ecology*, *22*(8), 2065–2073. https://doi.org/10.1111/mec.12158

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

Fox, J., Weisberg, S., Friendly, M., Hong, J., Andersen, R., Firth, D., & Taylor, S. (2016). effects: Effect Displays for Linear, Generalized Linear, and Other Models (Version 3.1-2). Retrieved from https://cran.r-project.org/web/packages/effects/index.html

Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., … R-Core. (2018). car: Companion to Applied Regression (Version 3.0-2). Retrieved from https://CRAN.R-project.org/package=car

Frankham, R. (2015). Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, *24*(11), 2610–2618. https://doi.org/10.1111/mec.13139

Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R., & Fenster, C. B. (2011). Predicting the probability of outbreeding depression. *Conservation Biology: The Journal of the Society for Conservation Biology*, *25*(3), 465–475. https://doi.org/10.1111/j.1523-1739.2011.01662.x

Franks, P. J., & Beerling, D. J. (2009). Maximum leaf conductance driven by CO2 effects on stomatal size and density over geologic time. *Proceedings of the National Academy of Sciences*, *106*(25), 10343–10347. https://doi.org/10.1073/pnas.0904209106

Franks, P. J., Drake, P. L., & Beerling, D. J. (2009). Plasticity in maximum stomatal conductance constrained by negative correlation between stomatal size and density: an analysis using Eucalyptus globulus. *Plant, Cell & Environment*, *32*(12), 1737–1748. https://doi.org/10.1111/j.1365-3040.2009.002031.x

Franks, P. J., W. Doheny-Adams, T., Britton-Harper, Z. J., & Gray, J. E. (2015). Increasing water-use efficiency directly through genetic manipulation of stomatal density. *New Phytologist*, *207*(1), 188–195. https://doi.org/10.1111/nph.13347

Fukasawa, K., Miyashita, T., Hashimoto, T., Tatara, M., & Abe, S. (2013). Differential population responses of native and alien rodents to an invasive predator, habitat alteration and plant masting. *Proceedings of the Royal Society of London B: Biological Sciences*, *280*(1773), 20132075. https://doi.org/10.1098/rspb.2013.2075

Gao, S., Bertrand, D., Chia, B. K. H., & Nagarajan, N. (2016). OPERA-LG: efficient and exact scaffolding of large, repeat-rich eukaryotic genomes with performance guarantees. *Genome Biology*, *17*, 102. https://doi.org/10.1186/s13059-016-0951-y

Glémin, S., & Ronfort, J. (2013). Adaptation and Maladaptation in Selfing and Outcrossing Species: New Mutations Versus Standing Variation. *Evolution*, *67*(1), 225–240. https://doi.org/10.1111/j.1558-5646.2012.01778.x

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., … Jaffe, D. B. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences*, *108*(4), 1513–1518. https://doi.org/10.1073/pnas.1017351108

Gómez-Fernández, A., Alcocer, I., & Matesanz, S. (2016). Does higher connectivity lead to higher genetic diversity? Effects of habitat fragmentation on genetic variation and population structure in a gypsophile. *Conservation Genetics*, *17*(3), 631–641. https://doi.org/10.1007/s10592-016-0811-z

Goudet, Jérôme. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Goudet, Jerome, & Jombart, T. (2015). hierfstat: Estimation and Tests of Hierarchical F-Statistics (Version 0.04-22). Retrieved from https://cran.r-project.org/web/packages/hierfstat/index.html

Goulet, B. E., Roda, F., & Hopkins, R. (2017). Hybridization in Plants: Old Ideas, New Techniques. *Plant Physiology*, *173*(1), 65–78. https://doi.org/10.1104/pp.16.01340

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P. M., & Westhoff, P. (2011). Evolution of C4 Photosynthesis in the Genus Flaveria: How Many and Which Genes Does It Take to Make C4? *The Plant Cell*, *23*(6), 2087–2105. https://doi.org/10.1105/tpc.111.086264

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Groszmann, M., Greaves, I. K., Fujimoto, R., James Peacock, W., & Dennis, E. S. (2013). The role of epigenetics in hybrid vigour. *Trends in Genetics*, *29*(12), 684–690. https://doi.org/10.1016/j.tig.2013.07.004

Gruenthal, K. M., Witting, D. A., Ford, T., Neuman, M. J., Williams, J. P., Pondella, D. J., … Larson, W. A. (2014). Development and application of genomic tools to the restoration of green abalone in southern California. *Conservation Genetics*, *15*(1), 109–121. https://doi.org/10.1007/s10592-013-0524-5

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., … Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, *9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., … Townshend, J. R. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, *1*(2), e1500052. https://doi.org/10.1126/sciadv.1500052

Hall, M. C., & Willis, J. H. (2006). Divergent Selection on Flowering Time Contributes to Local Adaptation in Mimulus Guttatus Populations. *Evolution*, *60*(12), 2466–2477. https://doi.org/10.1111/j.0014-3820.2006.tb01882.x

Hamilton, J. A., & Miller, J. M. (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate: Adaptive Introgression and Conservation. *Conservation Biology*, *30*(1), 33–41. https://doi.org/10.1111/cobi.12574

Hamilton, J. A., Okada, M., Korves, T., & Schmitt, J. (2015). The role of climate adaptation in colonization success in Arabidopsis thaliana. *Molecular Ecology*, *24*(9), 2253–2263. https://doi.org/10.1111/mec.13099

Hamrick J. L., & Godt M. J. W. (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1345), 1291–1298. https://doi.org/10.1098/rstb.1996.0112

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

Hand, R., & Gregor, T. (2006). Die Verbreitung von Arabis sagittata In Deutschland. Ergebnisse einer Herbarstudie. *Kochia*, *1*, 21–31.

Haug-Baltzell, A., Stephens, S. A., Davey, S., Scheidegger, C. E., & Lyons, E. (2017). SynMap2 and SynMap3D: web-based whole-genome synteny browsers. *Bioinformatics*, *33*(14), 2197–2198. https://doi.org/10.1093/bioinformatics/btx144

Heinken, T., & Weber, E. (2013). Consequences of habitat fragmentation for plant species: Do we know enough? *Perspectives in Plant Ecology, Evolution and Systematics*, *15*(4), 205–216. https://doi.org/10.1016/j.ppees.2013.05.003

Hepworth, C., Doheny-Adams, T., Hunt, L., Cameron, D. D., & Gray, J. E. (2015). Manipulating stomatal density enhances drought tolerance without deleterious effect on nutrient uptake. *The New Phytologist*, *208*(2), 336–341. https://doi.org/10.1111/nph.13598

Hereford, J. (2009). A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. *The American Naturalist*, *173*(5), 579–588. https://doi.org/10.1086/597611

Hereford, J. (2010). Does selfing or outcrossing promote local adaptation? *American Journal of Botany*, *97*(2), 298–302. https://doi.org/10.3732/ajb.0900224

Hetherington, A. M., & Woodward, F. I. (2003). The role of stomata in sensing and driving environmental change. *Nature*, *424*(6951), 901–908. https://doi.org/10.1038/nature01843

Hölzel, N., & Otte, A. (2001). The impact of flooding regime on the soil seed bank of flood-meadows. *Journal of Vegetation Science*, *12*(2), 209–218. https://doi.org/10.2307/3236605

Hölzel, N., & Otte, A. (2003). Restoration of a species-rich flood meadow by topsoil removal and diaspore transfer with plant material. *Applied Vegetation Science*, *6*(2), 131–140. https://doi.org/10.1111/j.1654-109X.2003.tb00573.x

Hölzel, N., & Otte, A. (2004). Assessing soil seed bank persistence in flood-meadows: The search for reliable traits. *Journal of Vegetation Science*, *15*(1), 93–100. https://doi.org/10.1111/j.1654-1103.2004.tb02241.x

Hopkins, R. (2013). Reinforcement in plants. *New Phytologist*, *197*(4), 1095–1103. https://doi.org/10.1111/nph.12119

Hopkins, R., & Rausher, M. D. (2012). Pollinator-Mediated Selection on Flower Color Allele Drives Reinforcement. *Science*, *335*(6072), 1090–1092. https://doi.org/10.1126/science.1215198

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., & Scheibe, S. (2017). multcomp: Simultaneous Inference in General Parametric Models (Version 1.4-7). Retrieved from https://cran.r-project.org/web/packages/multcomp/index.html

Houle, D., & Kondrashov, A. (2006). Mutation. In *Evolutionary Genetics: Concepts and Case Studies*. Oxford Press.

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., … Guo, Y.-L. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics*, *43*(5), 476–481. https://doi.org/10.1038/ng.807

Huang, H.-R., Yan, P.-C., Lascoux, M., & Ge, X.-J. (2012). Flowering time and transcriptome variation in *Capsella bursa-pastoris* (Brassicaceae). *New Phytologist*, *194*(3), 676–689. https://doi.org/10.1111/j.1469-8137.2012.04101.x

Hufford, K. M., & Mazer, S. J. (2003). Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends in Ecology & Evolution*, *18*(3), 147–155. https://doi.org/10.1016/S0169-5347(03)00002-8

Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*, *11*(6), 609–623. https://doi.org/10.1111/j.1461-0248.2008.01179.x

Hughes, J., Hepworth, C., Dutton, C., Dunn, J. A., Hunt, L., Stephens, J., … Gray, J. E. (2017). Reducing Stomatal Density in Barley Improves Drought Tolerance without Impacting on Yield. *Plant Physiology*, *174*(2), 776–787. https://doi.org/10.1104/pp.16.01844

Jalas, J., & Suominen, J. (1994). *Atlas Florae Europaeae. Distribution of Vascular Plants in Europe 10: Cruciferae (Sisymbrium to Aubrieta)* (Vol. 10). The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki.

Jeffries, D. L., Copp, G. H., Handley, L. L., Olsén, K. H., Sayer, C. D., & Hänfling, B. (2016). Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, Carassius carassius, L. *Molecular Ecology*, *25*(13), 2997–3018. https://doi.org/10.1111/mec.13613

Jiao, W.-B., Accinelli, G. G., Hartwig, B., Kiefer, C., Baker, D., Severing, E., … Schneeberger, K. (2017). Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. *Genome Research*, gr.213652.116. https://doi.org/10.1101/gr.213652.116

Jombart, T., Kamvar, Z. N., Lustrik, R., Collins, C., Beugin, M.-P., Knaus, B., … Calboli, F. (2016). adegenet: Exploratory Analysis of Genetic and Genomic Data (Version 2.0.1). Retrieved from https://cran.r-project.org/web/packages/adegenet/index.html

Juenger, T. E., Mckay, J. K., Hausmann, N., Keurentjes, J. J. B., Sen, S., Stowe, K. A., … Richards, J. H. (2005). Identification and characterization of QTL underlying whole-plant physiology in Arabidopsis thaliana: δ13C, stomatal conductance and transpiration efficiency. *Plant, Cell & Environment*, *28*(6), 697–708. https://doi.org/10.1111/j.1365-3040.2004.01313.x

Kahle, D., & Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, *5*(1), 144–161.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S., Freimer, N. B., … Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, *42*(4), 348–354. https://doi.org/10.1038/ng.548

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., & Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics*, *178*(3), 1709–1723. https://doi.org/10.1534/genetics.107.080101

Karl, R., & Koch, M. A. (2014). Phylogenetic signatures of adaptation: The Arabis hirsuta species aggregate (Brassicaceae) revisited. *Perspectives in Plant Ecology, Evolution and Systematics*, *16*(5), 247–264. https://doi.org/10.1016/j.ppees.2014.06.001

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Kay, K. M., & Schemske, D. W. (2008). Natural Selection Reinforces Speciation in a Radiation of Neotropical Rainforest Plants. *Evolution*, *62*(10), 2628–2642. https://doi.org/10.1111/j.1558-5646.2008.00463.x

Kesari, R., Lasky, J. R., Villamor, J. G., Marais, D. L. D., Chen, Y.-J. C., Liu, T.-W., … Verslues, P. E. (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*(23), 9197–9202. https://doi.org/10.1073/pnas.1203433109

Kiefer, M., Schmickl, R., German, D. A., Mandáková, T., Lysak, M. A., Al-Shehbaz, I. A., … Koch, M. A. (2014). BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. *Plant & Cell Physiology*, *55*(1), e3. https://doi.org/10.1093/pcp/pct158

Kiehl, K., Kirmer, A., Donath, T. W., Rasran, L., & Hölzel, N. (2010). Species introduction in restoration projects - evaluation of different techniques for the establishment of semi-natural grasslands in Central and Northwestern Europe. *Basic and Applied Ecology*, *11*(4), 285–299.

Kiehl, K., Kirmer, A., & Shaw, N. (2014). *Guidelines for Native Seed Production and Grassland Restoration*. Cambridge Scholars Publishing.

Kiehl, K., & Wagner, C. (2006). Effect of Hay Transfer on Long-Term Establishment of Vegetation and Grasshoppers on Former Arable Fields. *Restoration Ecology*, *14*(1), 157–166. https://doi.org/10.1111/j.1526-100X.2006.00116.x

Kim, M., Cui, M.-L., Cubas, P., Gillies, A., Lee, K., Chapman, M. A., … Coen, E. (2008). Regulatory Genes Control a Key Morphological and Ecological Trait Transferred Between Species. *Science*, *322*(5904), 1116–1119. https://doi.org/10.1126/science.1164371

Kimball, S., Gremer, J. R., Barron-Gafford, G. A., Angert, A. L., Huxman, T. E., & Venable, D. L. (2014). High water-use efficiency and growth contribute to success of non-native

Erodium cicutarium in a Sonoran Desert winter annual community. *Conservation Physiology*, *2*(1). https://doi.org/10.1093/conphys/cou006

Kinoshita, T., Ono, N., Hayashi, Y., Morimoto, S., Nakamura, S., Soda, M., … Shimazaki, K. (2011). FLOWERING LOCUS T Regulates Stomatal Opening. *Current Biology*, *21*(14), 1232–1238. https://doi.org/10.1016/j.cub.2011.06.025

Knaus, B. J., & Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53.

Knaus, B. J., Grunwald, N. J., Anderson, E. C., Winter, D. J., & Kamvar, Z. N. (2017). vcfR: Manipulate and Visualize VCF Data (Version 1.4.0). Retrieved from https://cran.r-project.org/web/packages/vcfR/index.html

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., … Wilson, R. K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, *22*(3), 568–576. https://doi.org/10.1101/gr.129684.111

Kogan, F. N. (1995). Application of vegetation index and brightness temperature for drought detection. *Advances in Space Research*, *15*(11), 91–100. https://doi.org/10.1016/0273-1177(95)00079-T

Kooyers, N. J. (2015). The evolution of drought escape and avoidance in natural herbaceous populations. *Plant Science*, *234*, 155–162. https://doi.org/10.1016/j.plantsci.2015.02.012

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 59. https://doi.org/10.1186/1471-2105-5-59

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, *9*, 29. https://doi.org/10.1186/1746-4811-9-29

Korte, A., Vilhjálmsson, B. J., Segura, V., Platt, A., Long, Q., & Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics*, *44*(9), 1066–1071. https://doi.org/10.1038/ng.2376

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

Krasensky, J., & Jonak, C. (2012). Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *Journal of Experimental Botany*, *63*(4), 1593–1608. https://doi.org/10.1093/jxb/err460

Krishna Krishnamurthy, P., & Francis, R. A. (2012). A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation*, *21*(8), 1901–1919. https://doi.org/10.1007/s10531-012-0306-2

Kronholm, I., Picó, F. X., Alonso-Blanco, C., Goudet, J., & de Meaux, J. (2012). Genetic Basis of Adaptation in Arabidopsis Thaliana: Local Adaptation at the Seed Dormancy Qtl Dog1. *Evolution*, *66*(7), 2287–2302. https://doi.org/10.1111/j.1558-5646.2012.01590.x

Laenen, B., Tedder, A., Nowak, M. D., Toräng, P., Wunder, J., Wötzel, S., … Slotte, T. (2018). Demography and mating system shape the genome-wide impact of purifying selection in Arabis alpina. *Proceedings of the National Academy of Sciences*, 201707492. https://doi.org/10.1073/pnas.1707492115

Lasky, J. R., Marais, D., L, D., Lowry, D. B., Povolotskaya, I., McKay, J. K., … Juenger, T. E. (2014). Natural Variation in Abiotic Stress Responsive Gene Expression and Local Adaptation to Climate in Arabidopsis thaliana. *Molecular Biology and Evolution*, *31*(9), 2283–2296. https://doi.org/10.1093/molbev/msu170

Lawson, S. S., Pijut, P. M., & Michler, C. H. (2014). The cloning and characterization of a poplar stomatal density gene. *Genes & Genomics. 36(4): 427-441.*, *36*(4), 427–441. https://doi.org/10.1007/s13258-014-0177-x

Lawson, T., Kramer, D. M., & Raines, C. A. (2012). Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. *Current Opinion in Biotechnology*, *23*(2), 215–220. https://doi.org/10.1016/j.copbio.2011.12.012

Lee, C.-R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., … Nordborg, M. (2017). On the post-glacial spread of human commensal Arabidopsis thaliana. *Nature Communications*, *8*, 14458. https://doi.org/10.1038/ncomms14458

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., … Przeworski, M. (2012). Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? *PLOS Biology*, *10*(9), e1001388. https://doi.org/10.1371/journal.pbio.1001388

Leimu, R., & Fischer, M. (2008). A Meta-Analysis of Local Adaptation in Plants. *PLoS ONE*, *3*(12), e4010. https://doi.org/10.1371/journal.pone.0004010

Leinonen, P. H., Remington, D. L., Leppälä, J., & Savolainen, O. (2013). Genetic basis of local adaptation and flowering time variation in Arabidopsis lyrata. *Molecular Ecology*, *22*(3), 709–723. https://doi.org/10.1111/j.1365-294X.2012.05678.x

Leinonen, T., McCairns, R. S., O'hara, R. B., & Merilä, J. (2013). QST–FST comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics*, *14*(3), 179–190.

Lempe, J., Balasubramanian, S., Sureshkumar, S., Singh, A., Schmid, M., & Weigel, D. (2005). Diversity of Flowering Responses in Wild Arabidopsis thaliana Strains. *PLOS Genetics*, *1*(1), e6. https://doi.org/10.1371/journal.pgen.0010006

Lenssen, J. P. M., Kleunen, M. V., Fischer, M., & Kroon, H. D. (2004). Local adaptation of the clonal plant Ranunculus reptans to flooding along a small-scale gradient. *Journal of Ecology*, *92*(4), 696–706. https://doi.org/10.1111/j.0022-0477.2004.00895.x

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and

SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lienert, J. (2004). Habitat fragmentation effects on fitness of plant populations – a review. *Journal for Nature Conservation*, *12*(1), 53–72. https://doi.org/10.1016/j.jnc.2003.07.002

Linhart, Y. B., & Grant, M. C. (1996). Evolutionary Significance of Local Genetic Differentiation in Plants. *Annual Review of Ecology and Systematics*, *27*(1), 237–277. https://doi.org/10.1146/annurev.ecolsys.27.1.237

Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics : TIG*, *26*(8), 345–352. https://doi.org/10.1016/j.tig.2010.05.003

Lyons, E., & Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. *The Plant Journal*, *53*(4), 661–673. https://doi.org/10.1111/j.1365-313X.2007.03326.x

Lyons, E., Freeling, M., Kustu, S., & Inwood, W. (2011). Using Genomic Sequencing for Classical Genetics in E. coli K12. *PLOS ONE*, *6*(2), e16717. https://doi.org/10.1371/journal.pone.0016717

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics (Oxford, England)*, *20*(16), 2878–2879. https://doi.org/10.1093/bioinformatics/bth315

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), 279–283. https://doi.org/10.1038/nature05706

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Martin, N. H., Bouck, A. C., & Arnold, M. L. (2006). Detecting Adaptive Trait Introgression Between Iris fulva and I. brevicaulis in Highly Selective Field Conditions. *Genetics*, *172*(4), 2481–2489. https://doi.org/10.1534/genetics.105.053538

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

Masel, J. (2011). Genetic drift. *Current Biology*, *21*(20), R837–R838. https://doi.org/10.1016/j.cub.2011.08.007

Masle, J., Gilmore, S. R., & Farquhar, G. D. (2005). The ERECTA gene regulates plant transpiration efficiency in Arabidopsis. *Nature*, *436*(7052), 866. https://doi.org/10.1038/nature03835

Massatti, R., Doherty, K. D., & Wood, T. E. (2018). Resolving neutral and deterministic contributions to genomic structure in <Emphasis Type="Italic">Syntrichia ruralis</Emphasis> (Bryophyta, Pottiaceae) informs propagule sourcing for dryland restoration. *Conservation Genetics*, *19*(1), 85–97. https://doi.org/10.1007/s10592-017-1026-7

Mathar, W., Kleinebecker, T., & Hölzel, N. (2015). Environmental variation as a key process of co-existence in flood-meadows. *Journal of Vegetation Science*, *26*(3), 480–491. https://doi.org/10.1111/jvs.12254

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

McKay, J. K., Christian, C. E., Harrison, S., & Rice, K. J. (2005). "How Local Is Local?"—A Review of Practical and Conceptual Issues in the Genetics of Restoration. *Restoration Ecology*, *13*(3), 432–440. https://doi.org/10.1111/j.1526-100X.2005.00058.x

McKay, J. K., Richards, J. H., Nemali, K. S., Sen, S., Mitchell-Olds, T., Boles, S., … Juenger, T. E. (2008). Genetics of Drought Adaptation in Arabidopsis Thaliana Ii. Qtl Analysis of a New Mapping Population, Kas-1 × Tsu-1. *Evolution*, *62*(12), 3014–3026. https://doi.org/10.1111/j.1558-5646.2008.00474.x

Méndez-Vigo, B., Picó, F. X., Ramiro, M., Martínez-Zapater, J. M., & Alonso-Blanco, C. (2011). Altitudinal and Climatic Adaptation Is Mediated by Flowering Traits and FRI, FLC, and PHYC Genes in Arabidopsis. *Plant Physiology*, *157*(4), 1942–1955. https://doi.org/10.1104/pp.111.183426

Mijangos, J. L., Pacioni, C., Spencer, P. B. S., & Craig, M. D. (2015). Contribution of genetics to ecological restoration. *Molecular Ecology*, *24*(1), 22–37. https://doi.org/10.1111/mec.12995

Mitchell-Olds, T., Willis, J. H., & Goldstein, D. B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*, *8*(11), 845–856. https://doi.org/10.1038/nrg2207

Mojica, J. P., Mullen, J., Lovell, J. T., Monroe, J. G., Paul, J. R., Oakley, C. G., & McKay, J. K. (2016). Genetics of water use physiology in locally adapted Arabidopsis thaliana. *Plant Science*, *251*(Supplement C), 12–22. https://doi.org/10.1016/j.plantsci.2016.03.015

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? *PLoS Biology*, *9*(8), e1001127. https://doi.org/10.1371/journal.pbio.1001127

Muchow, R. C., & Sinclair, T. R. (1989). Epidermal conductance, stomatal density and stomatal size among genotypes of Sorghum bicolor (L.) Moench. *Plant, Cell & Environment*, *12*(4), 425–431. https://doi.org/10.1111/j.1365-3040.1989.tb01958.x

Mummenhoff, K., Franzke, A., & Koch, M. (1997). Molecular phylogenetics of Thlaspi s.l. (Brassicaceae) based on chloroplast DNA restriction site variation and sequences of the internal transcribed spacers of nuclear ribosomal DNA. *Canadian Journal of Botany*, *75*(3), 469–482. https://doi.org/10.1139/b97-051

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia university press.

Neuffer, B., & Hoffrogge, R. (1999). Ecotypic and allozyme variation of Capsella bursa-pastoris and C. rubella (Brassicaceae) along latitude and altitude gradients on the Iberian Peninsula. *Anales Del Jardín Botánico de Madrid*, *57*(2), 299–315. https://doi.org/10.3989/ajbm.1999.v57.i2.204

Nordborg, M. (2000). Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics*, *154*(2), 923–929.

Nordborg, Magnus, Borevitz, J. O., Bergelson, J., Berry, C. C., Chory, J., Hagenblad, J., … Weigel, D. (2002). The extent of linkage disequilibrium in Arabidopsis thaliana. *Nature Genetics*, *30*(2), 190–193. https://doi.org/10.1038/ng813

Novikova, P. Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., … Nordborg, M. (2016). Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics*, *48*(9), 1077–1082. https://doi.org/10.1038/ng.3617

Novotná, I., & Czapik, R. (1974). Studies on the Progenies of Hybrids from the Arabis hirsuta Complex. *Folia Geobotanica & Phytotaxonomica*, *9*(4), 341–357. Retrieved from JSTOR.

Ohsumi, A., Kanemura, T., Homma, K., Horie, T., & Shiraiwa, T. (2007). Genotypic Variation of Stomatal Conductance in Relation to Stomatal Density and Length in Rice (Oryza sativa L.). *Plant Production Science*, *10*(3), 322–328. https://doi.org/10.1626/pps.10.322

O'Leary, S. J., Hollenbeck, C. M., Vega, R. R., Gold, J. R., & Portnoy, D. S. (2018). Genetic mapping and comparative genomics to inform restoration enhancement and culture of southern flounder, Paralichthys lethostigma. *BMC Genomics*, *19*(1), 163. https://doi.org/10.1186/s12864-018-4541-0

Olrik, D. C., & Kjaer, E. D. (2007). The reproductive success of a Quercus petraea × Q. robur F1-hybrid in back-crossing situations. *Annals of Forest Science*, *64*(1), 37–45. https://doi.org/10.1051/forest:2006086

Onge, K. R. S., Källman, T., Slotte, T., Lascoux, M., & Palmé, A. E. (2011). Contrasting demographic history and population structure in Capsella rubella and Capsella grandiflora, two closely related species with different mating systems. *Molecular Ecology*, *20*(16), 3306–3320. https://doi.org/10.1111/j.1365-294X.2011.05189.x

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, *11*(285–296), 23–27.

Ottewell, K. M., Bickerton, D. C., Byrne, M., & Lowe, A. J. (2016). Bridging the gap: a genetic assessment framework for population-level threatened plant conservation prioritization and decision-making. *Diversity and Distributions*, *22*(2), 174–188. https://doi.org/10.1111/ddi.12387

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

Paccard, A., Fruleux, A., & Willi, Y. (2014). Latitudinal trait variation and responses to drought in Arabidopsis lyrata. *Oecologia*, *175*(2), 577–587. https://doi.org/10.1007/s00442-014-2932-8

Palazzo, A. F., & Gregory, T. R. (2014). The Case for Junk DNA. *PLOS Genetics*, *10*(5), e1004351. https://doi.org/10.1371/journal.pgen.1004351

Paradis, E., cre, cph, Blomberg, S., Bolker, B., Brown, J., … Vienne, D. de. (2019). ape: Analyses of Phylogenetics and Evolution (Version 5.3). Retrieved from https://CRAN.R-project.org/package=ape

Paradis, E., Jombart, T., Schliep, K., Potts, A., & Winter, D. (2016). pegas: Population and Evolutionary Genetics Analysis System (Version 0.9). Retrieved from https://cran.r-project.org/web/packages/pegas/index.html

Passioura, J. B. (1996). Drought and drought tolerance. In *Drought Tolerance in Higher Plants: Genetical, Physiological and Molecular Biological Analysis* (pp. 1–5). https://doi.org/10.1007/978-94-017-1299-6_1

Pavlova, A., Beheregaray, L. B., Coleman, R., Gilligan, D., Harrisson, K. A., Ingram, B. A., … Sunnucks, P. (2017). Severe consequences of habitat fragmentation on genetic diversity of an endangered Australian freshwater fish: A call for assisted gene flow. *Evolutionary Applications*, *10*(6), 531–550. https://doi.org/10.1111/eva.12484

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360. https://doi.org/10.1111/mec.13557

Pearce, D. W., Millard, S., Bray, D. F., & Rood, S. B. (2006). Stomatal characteristics of riparian poplar species in a semi-arid environment. *Tree Physiology*, *26*(2), 211–218. https://doi.org/10.1093/treephys/26.2.211

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Pfeifer, B., Wittelsbuerger, U., Li, H., & Handsaker, B. (2018). PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses (Version 2.6.1). Retrieved from https://CRAN.R-project.org/package=PopGenome

Pillitteri, L. J., & Torii, K. U. (2012). Mechanisms of Stomatal Development. *Annual Review of Plant Biology*, *63*(1), 591–614. https://doi.org/10.1146/annurev-arplant-042811-105451

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2015). *nlme: Linear and Nonlinear Mixed Effects Models*. Retrieved from http://CRAN.R-project.org/package=nlme

Platt, A., Horton, M., Huang, Y. S., Li, Y., Anastasio, A. E., Mulyati, N. W., … Borevitz, J. O. (2010). The scale of population structure in Arabidopsis thaliana. *PLoS Genetics*, *6*(2), e1000843. https://doi.org/10.1371/journal.pgen.1000843

Postma, F. M., & Ågren, J. (2016). Early life stages contribute strongly to local adaptation in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *113*(27), 7590–7595. https://doi.org/10.1073/pnas.1606303113

Postma, F. M., Lundemo, S., & Ågren, J. (2016). Seed dormancy cycling and mortality differ between two locally adapted populations of Arabidopsis thaliana. *Annals of Botany*, *117*(2), 249–256. https://doi.org/10.1093/aob/mcv171

Purcell, S. (2009). PLINK (Version 1.07).

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org

Racimo, F., Marnetto, D., & Huerta-Sánchez, E. (2017). Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. *Molecular Biology and Evolution*, *34*(2), 296–317. https://doi.org/10.1093/molbev/msw216

Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews. Genetics*, *16*(6), 359–371. https://doi.org/10.1038/nrg3936

Raven, J. A. (2002). Selection pressures on stomatal evolution. *New Phytologist*, *153*(3), 371–386. https://doi.org/10.1046/j.0028-646X.2001.00334.x

Raven, J. A. (2014). Speedy small stomata? *Journal of Experimental Botany*, *65*(6), 1415–1424. https://doi.org/10.1093/jxb/eru032

Reed, D. H., & Frankham, R. (2003). Correlation between Fitness and Genetic Diversity. *Conservation Biology*, *17*(1), 230–237. https://doi.org/10.1046/j.1523-1739.2003.01236.x

Reich, P. B. (1984). Leaf Stomatal Density and Diffusive Conductance in Three Amphistomatous Hybrid Poplar Cultivars. *New Phytologist*, *98*(2), 231–239. https://doi.org/10.1111/j.1469-8137.1984.tb02733.x

Reitzel, A. M., Herrera, S., Layden, M. J., Martindale, M. Q., & Shank, T. M. (2013). Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology*, *22*(11), 2953–2970. https://doi.org/10.1111/mec.12228

Reusch, T. B. H., Ehlers, A., Hämmerli, A., & Worm, B. (2005). Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(8), 2826–2831. https://doi.org/10.1073/pnas.0500008102

Revelle, W. (2019). psych: Procedures for Psychological, Psychometric, and Personality Research (Version 1.8.12). Retrieved from https://CRAN.R-project.org/package=psych

Reyes, J. L., Rodrigo, M.-J., Colmenero-Flores, J. M., Gil, J.-V., Garay-Arroyo, A., Campos, F., … Covarrubias, A. A. (2005). Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects in vitro. *Plant, Cell & Environment*, *28*(6), 709–718. https://doi.org/10.1111/j.1365-3040.2005.01317.x

Reynolds, L. K., McGlathery, K. J., & Waycott, M. (2012). Genetic Diversity Enhances Restoration Success by Augmenting Ecosystem Services. *PLOS ONE*, *7*(6), e38397. https://doi.org/10.1371/journal.pone.0038397

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*, *83 (Pt 4)*, 363–372.

Rieseberg, Loren H., & Willis, J. H. (2007). Plant Speciation. *Science*, *317*(5840), 910–914. https://doi.org/10.1126/science.1137729

Ripley, B., Venables, B., Bates, D. M., ca 1998), K. H. (partial port, ca 1998), A. G. (partial port, & Firth, D. (2018). MASS: Support Functions and Datasets for Venables and Ripley's MASS (Version 7.3-51.1). Retrieved from https://CRAN.R-project.org/package=MASS

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*, 24–26. https://doi.org/10.1038/nbt.1754

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, *12*(12), 2640–2659. https://doi.org/10.1038/nprot.2017.123

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820. https://doi.org/10.1038/nrg3522

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 595–619. https://doi.org/10.1146/annurev.ecolsys.38.091206.095646

Scheepens, J. F., Rauschkolb, R., Ziegler, R., Schroth, V., & Bossdorf, O. (2017). Genotypic diversity and environmental variability affect the invasibility of experimental plant populations. *Oikos*, *127*(4), 570–578. https://doi.org/10.1111/oik.04818

Schmeil, O., Fitschen, J., Fleischmann, A., Nordt, B., Raab-Straube, E. von, & Vogt, R. (2016). *SCHMEIL-FITSCHEN Die Flora Deutschlands und der angrenzenden Länder: Ein Buch zum Bestimmen aller wildwachsenden und häufig kultivierten Gefäßpflanzen*. Quelle & Meyer.

Schnittler, M., & Günther, K.-F. (1999). Central European vascular plants requiring priority conservation measures – an analysis from national Red Lists and distribution maps. *Biodiversity & Conservation*, *8*(7), 891–925. https://doi.org/10.1023/A:1008828704456

Schoen, D. J., & Brown, A. H. (1991). Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences*, *88*(10), 4494–4497. https://doi.org/10.1073/pnas.88.10.4494

Schopf, J. W., Kudryavtsev, A. B., Czaja, A. D., & Tripathi, A. B. (2007). Evidence of Archean life: Stromatolites and microfossils. *Precambrian Research*, *158*(3), 141–155. https://doi.org/10.1016/j.precamres.2007.04.009

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2016). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*(8), 907–917. https://doi.org/10.1111/2041-210X.12700

Shang, L., Wang, Y., Cai, S., Wang, X., Li, Y., Abduweli, A., & Hua, J. (2016). Partial Dominance, Overdominance, Epistasis and QTL by Environment Interactions Contribute to Heterosis in Two Upland Cotton Hybrids. *G3: Genes, Genomes, Genetics*, *6*(3), 499–507. https://doi.org/10.1534/g3.115.025809

Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., … Kingsley, D. M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, *428*(6984), 717. https://doi.org/10.1038/nature02415

Shen, G., Zhan, W., Chen, H., & Xing, Y. (2014). Dominance and epistasis are the main contributors to heterosis for plant height in rice. *Plant Science*, *215–216*, 11–18. https://doi.org/10.1016/j.plantsci.2013.10.004

Silvertown, J., Servaes, C., Biss, P., & Macleod, D. (2005). Reinforcement of reproductive isolation between adjacent populations in the Park Grass Experiment. *Heredity*, *95*(3), 198–205. https://doi.org/10.1038/sj.hdy.6800710

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31. https://doi.org/10.1186/1471-2105-6-31

Smit, A., & Hubley, R. (2008). RepeatModeler Open-1.0 (Version 1.0). Retrieved from http://www.repeatmasker.org

Smit, A., Hubley, R., & Green, P. (2013). *RepeatMasker Open-4.0*. Retrieved from http://www.repeatmasker.org

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, *14*(6), 404–414. https://doi.org/10.1038/nrg3446

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics (Oxford, England)*, *19 Suppl 2*, ii215-225.

Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Halldorsdottir, S. S., … Schmitt, J. (2004). A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(13), 4712–4717. https://doi.org/10.1073/pnas.0306401101

Suarez-Gonzalez, A., Lexer, C., & Cronk, Q. C. B. (2018). Adaptive introgression: a plant perspective. *Biology Letters*, *14*(3), 20170688. https://doi.org/10.1098/rsbl.2017.0688

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, *19*(1), 131. https://doi.org/10.1186/s13059-018-1520-3

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. *Molecular Ecology*, *15*(2), 343–355. https://doi.org/10.1111/j.1365-294X.2005.02794.x

Tellier, A., Laurent, S. J. Y., Lainer, H., Pavlidis, P., & Stephan, W. (2011). Inference of seed bank parameters in two wild tomato species using ecological and genetic data. *Proceedings of the National Academy of Sciences*, *108*(41), 17052–17057. https://doi.org/10.1073/pnas.1111266108

Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, *25*(10), 2144–2164. https://doi.org/10.1111/mec.13606

Titz, W. (1979). Die Interfertilitätsbeziehungen europäischer Sippen derArabis hirsuta-Gruppe (Brassicaceae). *Plant Systematics and Evolution*, *131*(3), 291–310. https://doi.org/10.1007/BF00984260

Toräng, P., Wunder, J., Obeso, J. R., Herzog, M., Coupland, G., & Ågren, J. (2015). Large-scale adaptive differentiation in the alpine perennial herb *Arabis alpina*. *New Phytologist*, *206*(1), 459–470. https://doi.org/10.1111/nph.13176

Török, P., Miglécz, T., Valkó, O., Kelemen, A., Tóth, K., Lengyel, S., & Tóthmérész, B. (2012). Fast restoration of grassland vegetation by a combination of seed mixture sowing and

low-diversity hay transfer. *Ecological Engineering*, *44*, 133–138. https://doi.org/10.1016/j.ecoleng.2012.03.010

Torres-Martinez, L., & Emery, N. C. (2016). Genome-wide SNP discovery in the annual herb, Lasthenia fremontii (Asteraceae): genetic resources for the conservation and restoration of a California vernal pool endemic. *Conservation Genetics Resources*, *8*(2), 145–158. https://doi.org/10.1007/s12686-016-0524-0

Trabucco, A., & Zomer, R. (2010). Global soil water balance geospatial database. *CGIAR Consortium for Spatial Information, Published Online, Available from the CGIAR-CSI GeoPortal at: Http://Www. Cgiar-Csi. Org (Last Access: January 2013)*.

Tsaftaris, A. S., & Kafka, M. (1997). Mechanisms of Heterosis in Crop Plants. *Journal of Crop Production*, *1*(1), 95–111. https://doi.org/10.1300/J144v01n01_05

Tyrrell, M. C., & Byers, J. E. (2007). Do artificial substrates favor nonindigenous fouling species over native species? *Journal of Experimental Marine Biology and Ecology*, *342*(1), 54–60. https://doi.org/10.1016/j.jembe.2006.10.014

Vandepitte, K., Gristina, A. S., Hert, K. D., Meekers, T., Roldán-Ruiz, I., & Honnay, O. (2012). Recolonization after habitat restoration leads to decreased genetic variation in populations of a terrestrial orchid. *Molecular Ecology*, *21*(17), 4206–4215. https://doi.org/10.1111/j.1365-294X.2012.05698.x

Vander Mijnsbrugge, K., Bischoff, A., & Smith, B. (2010). A question of origin: Where and how to collect seed for ecological restoration. *Basic and Applied Ecology*, *11*(4), 300–311. https://doi.org/10.1016/j.baae.2009.09.002

Verslues, P. E., Govinal Badiger, B., Ravi, K., & M. Nagaraj, K. (2013). Drought tolerance mechanisms and their molecular basis. In tthew A. Jenks & P. M. Hasegawa (Eds.), *Plant Abiotic Stress* (pp. 15–46). https://doi.org/10.1002/9781118764374.ch2

Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, *106*(Supplement 1), 9939–9946. https://doi.org/10.1073/pnas.0901397106

Vrijenhoek, R. C. (1994). Genetic diversity and fitness in small populations. In V. Loeschcke, S. K. Jain, & J. Tomiuk (Eds.), *Conservation Genetics* (pp. 37–53). https://doi.org/10.1007/978-3-0348-8510-2_5

Vu, V. Q. (2011). A ggplot2 based biplot (Version 0.55).

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., … Hoffmann, A. A. (2011). Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications*, *4*(6), 709–725. https://doi.org/10.1111/j.1752-4571.2011.00192.x

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). corrplot: Visualization of a Correlation Matrix (Version 0.84). Retrieved from https://CRAN.R-project.org/package=corrplot

Whalen, M. D. (1978). Reproductive Character Displacement and Floral Diversity in Solanum Section Androceras. *Systematic Botany*, *3*(1), 77–86. https://doi.org/10.2307/2418533

Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution*, *30*(1), 42–49. https://doi.org/10.1016/j.tree.2014.10.009

Whitlock, M. C., & Guillaume, F. (2009). Testing for Spatially Divergent Selection: Comparing QST to FST. *Genetics*, *183*(3), 1055–1063. https://doi.org/10.1534/genetics.108.099812

Whitney, K. D., Ahern, J. R., Campbell, L. G., Albert, L. P., & King, M. S. (2010). Patterns of hybridization in plants. *Perspectives in Plant Ecology, Evolution and Systematics*, *12*(3), 175–182. https://doi.org/10.1016/j.ppees.2010.02.002

Whitney, K. D., Randell, R. A., & Rieseberg, L. H. (2006). Adaptive Introgression of Herbivore Resistance Traits in the Weedy Sunflower Helianthus annuus. *The American Naturalist*, *167*(6), 794–807. https://doi.org/10.1086/504606

Whitney, K. D., Randell, R. A., & Rieseberg, L. H. (2010). Adaptive introgression of abiotic tolerance traits in the sunflower Helianthus annuus. *New Phytologist*, *187*(1), 230–239. https://doi.org/10.1111/j.1469-8137.2010.03234.x

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Retrieved from http://ggplot2.org

Wickham, H. (2016). plyr: Tools for Splitting, Applying and Combining Data (Version 1.8.4). Retrieved from https://CRAN.R-project.org/package=plyr

Williams, A. V., Nevill, P. G., & Krauss, S. L. (2014). Next generation restoration genetics: applications and opportunities. *Trends in Plant Science*, *19*(8), 529–537. https://doi.org/10.1016/j.tplants.2014.03.011

Willing, E.-M., Rawat, V., Mandáková, T., Maumus, F., James, G. V., Nordström, K. J. V., … Schneeberger, K. (2015). Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. *Nature Plants*, *1*(2), 14023. https://doi.org/10.1038/nplants.2014.23

Willoughby, J. R., Sundaram, M., Wijayawardena, B. K., Kimble, S. J. A., Ji, Y., Fernandez, N. B., … DeWoody, J. A. (2015). The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biological Conservation*, *191*, 495–503. https://doi.org/10.1016/j.biocon.2015.07.025

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., … Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*(11), 1173. https://doi.org/10.1038/ng.3097

Wu, C. A., Lowry, D. B., Nutter, L. I., & Willis, J. H. (2010). Natural variation for drought-response traits in the Mimulus guttatus species complex. *Oecologia*, *162*(1), 23–33. https://doi.org/10.1007/s00442-009-1448-0

Yoo, C. Y., Pence, H. E., Jin, J. B., Miura, K., Gosney, M. J., Hasegawa, P. M., & Mickelbart, M. V. (2010). The Arabidopsis GTL1 Transcription Factor Regulates Water Use Efficiency and Drought Tolerance by Modulating Stomatal Density via Transrepression of SDD1. *The Plant Cell Online*, *22*(12), 4128–4141. https://doi.org/10.1105/tpc.110.078691

Yu, G. (2018). scatterpie: Scatter Pie Plot (Version 0.1.2). Retrieved from https://CRAN.R-project.org/package=scatterpie

Yu, H., Chen, X., Hong, Y.-Y., Wang, Y., Xu, P., Ke, S.-D., … Xiang, C.-B. (2008). Activated Expression of an Arabidopsis HD-START Protein Confers Drought Tolerance with Improved Root System and Reduced Stomatal Density. *The Plant Cell*, *20*(4), 1134–1151. https://doi.org/10.1105/tpc.108.058263

Zhao, S., Guo, Y., Sheng, Q., & Shyr, Y. (2015). heatmap3: An Improved Heatmap Package (Version 1.1.1). Retrieved from https://cran.r-project.org/web/packages/heatmap3/index.html

6 APPENDIX

6.1 RAD-SEQ PROTCOL

ADAPTED FROM ETTER ET AL. 2011

This protocol is more compact than the one by (Etter et al., 2011). We recommend to also take a look into the original protocol for additional information.

6.1.1 1. MATERIAL

a. Restriction endonuclease digestion
    i. Restriction enzyme (e.g. KpnI; NEB)
    ii. Genomic DNA: 250ng per sample; min. 12.5 ng/µl
b. P1 Adapter ligation
    i. NEB Buffer 2
    ii. rATP(Promega): 100mM
    iii. P1 Adapter 450nM (see Appendix 1)
    iv. Concentrated T4 DNA Ligase (NEB): 2,000,000 U/ml
c. Purfication
    i. QIAquick or MinElute PCR Purification Kit (Qiagen)
    ii. Ampure XP Beads
d. End Repair
    i. Quick Blunting Kit (NEB)
e. 3`-dA overhang addition
    i. NEB Buffer 2
    ii. dATP (Fermentas): 10mM
    iii. Klenow Fragment (3´ to 5´ exo-, NEB): 5,000 U/ml.
f. P2 Adapter ligation
    i. NEB Buffer 2
    ii. rATP: 100mM
    iii. P2 Adapter 10µM (see Appendix 1)
    iv. Concentrated T4 DNA Ligase
g. RAD tag Amplification Enrichment
    i. Phusion High-Fidelity PCR Master Mix with HF Buffer
    ii. Primer Mix 10µM (see Appendix 1)

6.1.2 2. METHODS

6.1.2.1 2.1 RESTRICTION ENDONUCLEASE DIGESTION:

1. Digest 250 ng of DNA for each sample individually following manufacturer's instructions in 25 µl reaction volume

2. Heat inactivate the enzyme if possible. If not, reaction has to be cleaned using QIAquick column **only** if cut sites are recreated by adapter ligation (depends on barcode sequence)

6.1.2.2 2.3 P1 ADAPTER LIGATION

1. In this step each digestion sample is ligated with an individually barcoded P1 Adapter using the sticky overhang created by the restriction enzyme
2. To each digest add:

   0.5µl 10X NEB Buffer (the same as used for digestion)

   1.5µl Barcoded P1 Adapter 450 nM for KpnI

   0.3µl rATP (100mM)

   0.25µl T4 DNA Ligase (2,000,000 U/ml)

   2.45µl H_2O (ad 30 µl)

Note: Add adapter before the enzyme to avoid re-ligation of the genomic DNA

3. Incubate 30 min @ room temperature

4. Heat-inactivate for 10 min @ 65 °C, let cool to RT afterwards

6.1.2.3 2.4. SAMPLE MULTIPLEXING

1. Combine barcoded samples at desired ratios to create pools that will later be multiplexed with the second adapter (see Appendix 2). For each pool, make 100µl Aliquots. Use one aliquot per pool to complete the protocol and freeze the rest as backup.

The following steps in the protocol are done for each of the pools.

6.1.2.4 2.5 DNA SHEARING

1. For each pool one aliquot (100µl) is fragmented for 7 cycles with 30 s shearing followed by 30 s break
2. Check on tape station: average size should be 500-700bp (predominantly smaller than 1kb)

6.1.2.5 2.6 AMPURE XP CLEANUP AND SIZE SELECTION

1. This step purifies the samples and removes fragments shorter than 200bp, which are mostly adapter dimers
2. Add to each sample 136µl AmpureXP beads and 22µl EB Buffer
3. Mix well by pipetting 10 times
4. Incubate 15 minutes
5. Put on magnet and incubate 10 minutes

6. Remove the liquid carefully without disturbing the beads
7. Wash twice with 200µl Ethanol (riddle tubes carefully)
8. Remove all ethanol and dry for 15 minutes until beads are matte
9. Remove from magnet
10. Add 20µl EB and re-suspend beads
11. Put on magnet and transfer liquid without beads into fresh tube

6.1.2.6 2.7. END REPAIR

1. To eluate from previous step add:
   2.5 µl 10x Blunting Buffer

   2.5 µl dNTP mix (1mM)

   1 µl Blunt Enzyme Mix
2. Incubate at RT for 30 minutes
3. Purify using Ampure beads or QIAquick column and elute in 43µl EB

6.1.2.7 2.8. 3'-DA OVERHANG ADDITION

1. To eluate from previous step add:
   5 µl 10X NEB Buffer 2

   1µl dATP

   3µl Klenow-fragment
2. Incubate at 37°C for 30 min and slowly cool to ambient temperature
3. Heat inactivation for 5 min at 70°C and hold on 4°C

6.1.2.8 2.9. P2 ADAPTER LIGATION

1. To inactivated reaction from previous step add:
   1 µl P2 Adapter (10µM)

   0.5 µl rATP (100mM)

   0.5 µl T4 DNA Ligase
2. Incubate at room temperature for 30 minutes
3. Purify using Ampure beads or QIAquick column and elute in 52µl EB

6.1.2.9 2.10. RAD TAG AMPLIFICATION

1. To determine library quality a test amplification should be done:

   10.5 µl H2O

   12.5 µl Phusion High-Fidelity Master Mix

   1.0 µl Primer Mix (10µM); different mix for each pool as primer contains second barcode (index)

1.0 µl RAD library template (eluate from last step)

Run on Thermocycler with following program:

30 s @ 98° C

16x {10s @ 98° C; 30s @ 65° C; 30s @ 72° C}

5 min 72° C

Analyze the purified product on tapestation next to template to check if amplification worked.

2. If amplification worked, repeat the PCR under the same conditions but in 100µl volume with 4µl template, purify the product using Ampure Beads excluding fragments smaller than 200bp and check the results on tapestation. **Note: If amplification does not work well more template can be used, but more template also introduces more fragments that don't amplify, can't be sequenced and might disturb the sequencing process.**

3. Adjust the number of cycles if needed or try adding more template to the reaction.

4. Mix the indexed pools in desired ratios before sequencing.

6.1.2.10 SEQUENCES

6.1.2.10.1 1. ADAPTERS

RADseq uses modified Illumina adapters with specific overhang for the restriction enzyme cut site. They also include barcodes (XXXXXX) for multiplexing and short random sequences (NNNNN) for identification of PCR duplicates. Each adapter is made from two custom oligos:

P1 with SacII-specific overhang:

5' - ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNXXXXXXG*C -3'

5' - [PHO]XXXXXXNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT - 3'

P2 universal:

5' - [PHO]AGATCGGAAGAGCGAGAACA*A -3'

5'- GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*T -3'

Oligos should be HPLC purified. Best is NGSgrade oligos by Eurofins.

To anneal the complementary oligos, follow these instructions:

1. Prepare 100µM stocks for each oligo in 1X EB
2. Combine complementary oligos at 10µM in 1X Annealing Buffer (AB; 10X AB: 500mM NaCl, 100mM Tris-Cl, pH 7.5-8.0): 80µl AB, 10µl Oligo 1, 10µl Oligo 2
3. Run samples on Thermocycler with the following program: 2.5 min @ 97.5°C; cool to 21° C with -1.5 °C per minute; 1 min @ 21 °C; hold @ 4° C
4. Dilute to desired concentrations in AB and/or freeze as stock solution

Final P1-adapter concentration and amount used in ligation depends on the number of DNA fragments (cut-sites). I used the molarity calculator provided in the following publication: https://doi.org/10.1371/journal.pone.0037135

6.1.2.10.2 2. PRIMERS

RADseq uses modified Illumina primers. The reverse primer contains an index sequence that can be used for multiplexing. Indices are the same as used in Illumina kits.

Primer forward:

5’ – AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG -3’

Primer reverse:

5’ – CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGC -3’

For the protocol prepare 10µM primer mixes with EB, e.g.: 80µl EB, 10µl Primer 1, 10µl Primer 2. Note that a specific primer mix with a different index is needed for each pool.

Primer sequences were taken from (Peterson et al., 2012).

6.2 TABLES

Table S 1: Overview of accessions used for measuring natural variation of stomata traits and water-use efficiency

Accession (1001 genomes ID)	Latitude	Longitude
139	48.5167	-4.06667
410	49.4211	16.3497
430	47	15.5
991	55.3833	14.05
992	55.3833	14.05
997	55.3833	14.05
1002	55.3833	14.05
1061	55.7167	14.1333
1062	55.7167	14.1333
1063	55.7167	14.1333
1066	55.7167	14.1333
1254	59.4333	17.0167
5023	51.3	0.5
5165	51.3	0.4
5353	54.5	-3
5830	56.3333	15.9667
5832	56.3333	15.9667
5865	55.76	14.12
5867	55.76	14.12
5907	49.4112	16.2815
6009	62.877	18.177
6011	62.877	18.177
6012	62.877	18.177
6013	62.877	18.177
6016	62.9	18.4
6019	56.06	14.29
6020	56.06	14.29
6021	56.06	14.29
6023	55.7509	13.3712
6024	55.7509	13.3712
6034	56.1	13.74
6035	56.1	13.74
6036	56.1	13.74
6038	56.1	13.74
6039	56.1	13.74
6040	55.66	13.4
6041	56.0328	14.775
6043	62.801	18.079
6069	62.9513	18.2763

6071	62.9308	18.3448
6073	56.1481	15.8155
6074	56.4573	16.1408
6076	55.6942	13.4504
6085	55.7097	13.2145
6088	56.4666	16.1284
6090	55.6525	13.2197
6092	55.6514	13.2233
6094	55.6494	13.2147
6097	55.6481	13.2264
6098	55.6561	13.2178
6099	55.6575	13.2386
6100	55.6	13.2
6104	55.7	13.2
6105	55.7967	13.1211
6106	55.7931	13.1186
6107	55.7942	13.1222
6108	55.7989	13.1206
6109	55.7936	13.1233
6111	55.7989	13.1219
6112	55.7967	13.1044
6113	55.8078	13.1028
6114	55.8097	13.1342
6115	55.8	13.1367
6118	55.7	13.2
6122	55.8364	13.3075
6124	55.8378	13.3092
6125	55.8403	13.3106
6126	55.8411	13.3047
6128	55.8397	13.2881
6131	55.8369	13.3181
6132	55.8386	13.3186
6133	55.8364	13.2906
6134	55.8383	13.2906
6136	55.9336	13.5519
6137	55.9419	13.5603
6138	55.9403	13.5511
6140	55.9392	13.5539
6142	55.9428	13.5558
6145	55.9497	13.5533
6148	55.9319	13.5508
6149	55.9281	13.5481
6150	55.9261	13.5319

6151	55.6528	13.2244
6153	62.6425	17.7422
6154	62.6422	17.7406
6163	62.6425	17.7356
6169	62.8714	18.3447
6173	62.8717	18.3419
6177	62.6322	17.69
6184	62.8892	18.4522
6188	55.7683	14.1386
6189	55.7686	14.1383
6191	55.7689	14.1375
6192	55.7692	14.1369
6193	55.7694	14.1347
6194	55.7706	14.1342
6195	55.7708	14.1342
6201	55.7719	14.1211
6202	55.7717	14.1206
6203	55.7714	14.1208
6209	62.8836	18.1842
6210	62.8839	18.1836
6217	63.0169	18.3283
6235	62.9611	18.3589
6242	55.7	13.2
6252	55.5796	14.3336
6258	55.5796	14.3336
6268	55.5796	14.3336
6276	55.5796	14.3336
6284	55.5796	14.3336
6296	49.2771	16.6314
6413	56.06	13.97
6424	49.3853	16.2544
6445	49.3853	16.2544
6903	49.4013	16.2326
6915	50.3	6.3
6918	63.0165	18.3174
6924	51.4083	-0.6383
6933	41.59	2.49
6944	51.4083	-0.6383
6945	52.24	4.45
6957	49.42	16.36
6961	38.3333	-3.53333
6970	41.7194	2.93056
6971	41.7194	2.93056

6974	56.0648	13.9707
6982	52.3	9.3
6987	48.0683	7.62551
6989	54.8	-2.4333
6997	51.8333	5.5833
7002	51.3333	6.1
7003	47.5	7.5
7008	52	5.675
7013	52.4584	13.287
7062	50.2981	8.26607
7119	50	8.5
7133	50.1102	8.6822
7147	50.584	8.67825
7158	47	15.5
7169	54.4175	9.88682
7177	49	15
7244	50.001	8.26664
7255	50.95	7.5
7268	52.6969	10.981
7276	50.2	8.5833
7305	53.476	10.6065
7327	41.7833	3.03333
7328	41.7833	3.03333
7343	52.5339	13.181
7346	52.6058	11.8558
7354	56.5	14.9
7382	52.0918	5.1145
8214	49	2
8222	56.0328	14.775
8227	62.7989	17.9103
8231	56.3	16
8234	56.4606	15.8127
8237	55.8	13.1
8240	55.705	13.196
8241	55.9473	13.821
8242	56.1494	15.7884
8247	56.07	13.74
8249	57.7	15.8
8256	56.4	12.9
8259	56.4	12.9
8264	41.6833	2.8
8283	55.76	14.12
8306	56.1	13.74

8307	56.1	13.74
8326	56.0328	14.775
8335	55.71	13.2
8351	60.25	18.37
8369	55.6942	13.4504
8376	62.69	18
8386	58.9	11.2
8422	56.06	14.29
8426	56.06	13.97
9057	56.1	13.9167
9321	62.8622	18.336
9339	57.7133	15.0689
9363	62.9147	18.4045
9369	57.6781	14.9986
9371	63.016	18.3175
9380	55.7488	13.3742
9386	62.806	18.1896
9388	62.806	18.1896
9395	57.5089	15.0105
9399	55.4234	13.9905
9404	55.7491	13.399
9405	55.7491	13.399
9407	55.7491	13.399
9408	56.047	13.9519
9409	56.0573	14.302
9421	55.9745	14.3997
9436	56.1633	14.6806
9442	55.5678	14.3398
9453	57.8009	18.5162
9454	57.8009	18.5162
9470	57.6511	14.8043
9476	55.5796	14.3336
9481	55.4242	13.8484
9506	40.11	-7.47
9517	42.19	-7.8
9519	41.94	2.64
9520	41.7	-3.68
9521	41.43	2.13
9524	42.52	-0.56
9525	42.49	0.54
9526	41.54	2.39
9527	40.37	-5.74
9528	40.94	-1.37

	9531	41.21	-4.54
	9532	42.23	-4.64
	9534	40.05	-4.65
	9535	42.31	3.19
	9539	40.29	-6.67
	9540	41.81	2.34
	9546	40.82	-1.68
	9547	41.67	2.62
	9548	40.4	-5
	9551	42.28	-5.92
	9552	40.87	-4.5
	9553	41.58	-4.71
	9556	39.66	-4.34
	9557	42.46	0.7
	9558	41.57	-5.64
	9561	40.71	-5.04
	9562	41.67	2
	9564	40.45	-1.6
	9565	42.97	-1.23
	9567	42.34	1.3
	9568	42.76	-0.23
	9569	42.87	-6.45
	9573	41.86	2.99
	9577	42.34	2.17
	9578	42.13	-6.7
	9581	42.84	-5.12
	9582	41.48	-1.63
	9584	41.19	-3.58
	9585	43.4	-7.39
	9586	41.03	-3.27
	9587	41.5	-1.88
	9588	42.11	0.6
	9589	41.6	-2.83
	9591	42.86	-3.59
	9593	42.26	-2.99
	9594	42.04	1.01
	9595	40.89	-5.5
	9596	41.95	-7.45
	9597	42.31	-2.53
	9599	42.8	-5.77
	9601	41.85	-1.88
	9602	40.5	-3.96
	9756	44.3	21.08

9772	48.41	8.85
9776	48.43	8.79
9778	48.41	8.84
9784	48.5	8.8
9795	48.5	9.11
9804	48.45	8.87
9806	48.56	9.16
9807	48.6	9.22
9811	48.52	9.05
9819	42.35	-3.03
9820	41	-4.71
9821	41.81	2.49
9822	40.52	-4.02
9824	42.91	-4.91
9825	40.4	-3.88
9826	42.49	-6.71
9833	40.38	-4.21
9834	40.51	-3.9
9835	40.61	-6.57
9836	41.25	-1.32
9838	41.83	-5.38
9839	40.44	-4.27
9840	41.13	-1.43
9841	40.59	-4.15
9843	40.53	-3.92
9844	42.27	0.19
9845	40.48	-3.96
9846	42.31	-3.02
9848	40.11	-5.77
9849	40.65	-4.11
9850	42.86	-0.7
9851	42.96	-6.1
9852	40.46	-3.75
9853	43.33	-5.91
9855	40.57	-3.89
9856	40.51	-4
9857	40.33	-3.8
9858	40.98	-3.8
9859	40.5	-3.88
9860	42.24	-2.62
9861	40.72	-3.21
9864	41.76	2.69
9866	41.89	-2.79

9867	40.94	-3.22
9868	41.78	2.37
9870	41.91	0.17
9874	42.34	-3
9876	41.34	0.99
9877	43.02	-5.6
9878	40.78	-3.62
9880	42.72	-3.44
9881	40.46	-5.32
9882	40.46	-4.26
9883	42.1	-2.56
9885	41.14	-3.68
9886	42.38	1.73
9888	40.93	-3.31
9890	43.16	-5.07
9891	41.93	2.92
9892	42.68	-6.96
9895	41.78	2.57
9897	40.95	-5.63
9898	41.14	-3.58
9899	42.54	0.84
9901	42.27	-2.98
9902	40.71	-3.24
9903	42.31	-3.1
9904	40.95	-3.31
9906	40	4.25
9938	50.65	2.99
9942	41.32	-1.34
9948	41.05	-3.54
9949	42.69	-6.93
9950	42.63	0.76
10005	46.11	21.95
10017	44.34	21.46

ACKNOWLEDGEMENTS

First, I would like to thank Juliette de Meaux for supervising my PhD project and for her excellent support and feedback. Second, I would like to thank my TAC members, Norbert Hölzel and Aurélien Tellier for taking time to discuss my project and giving very helpful feedback. Also, I would like to thank Norbert for helping with the sampling in the Upper Rhine sites and sharing his knowledge about floodplain ecology, and Aurélien for hosting me in his lab for some time, so I could get insights into demographic modelling. I would like to thank the all current and former members of the de Meaux lab, especially Gregor for interesting discussions and helpful feedback; Kirsten for her assistance with experiments and everything else; Kim and Margarita for fun after-work activities, and of course Bene, Fei, Maroua, Ulrike, Martina and Agustin for creating a very nice and enjoyable working environment. I am also very grateful to the gardeners for taking excellent care of my plants and helping me set up experiments. Thank you also to Janine Altmüller, Christian Becker and the rest of the Cologne Center for Genomics team for their collaboration and support on all sequencing projects. I would like to thank my co-authors, Andreas Weber, Tabea Mettler-Altmann, Christian Becker, Grey Monroe, Wen-Biao Jiao, Korbinian Schneeberger, Arthur Korte, Aurélien Tellier and Norbert Hölzel for fruitful collaborations. I would like to thank Markus Koch for insightful discussions about the *Arabis* taxonomy and Eric Schranz for his helping me with genome assembly.

I would especially like to thank the members of my thesis committee Juliette de Meaux, Thomas Wiehe, Hartmut Arndt, Michael Lenhard and Kim Steige for taking their time to review my work.

I am also grateful to my parents for enabling and encouraging me to follow my interests and study Biology and making the great experience of doing a PhD project possible.

Finally, I want to thank my wife, Veronique, for her great support in these past years and of course for an unforgettable sampling trip and helping me with my experiments.

ERKLÄRUNG

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 Prof. Dr. Juliette de Meaux betreut worden.

Köln, 01.11.2019

_____
Hannes Dittberner

Teilpublikationen

Dittberner, H., Korte, A., Mettler-Altmann, T., Weber, A. P. M., Monroe, G., & de Meaux, J. (2018). Natural variation in stomata size contributes to the local adaptation of water-use efficiency in Arabidopsis thaliana. *Molecular Ecology*, *27*(20), 4052–4065.

Dittberner, H., Becker, C., Jiao, W.-B., Schneeberger, K., Hölzel, N., Tellier, A., & Meaux, J. de. (2019). Strengths and potential pitfalls of hay transfer for ecological restoration revealed by RAD-seq analysis in floodplain Arabis species. Molecular Ecology, 28(17), 3887–3901.