OULU 2018
A 715

ACTA UNIVERSITATIS OULUENSIS

Tuomas Hämälä

ECOLOGICAL GENOMICS IN *ARABIDOPSIS LYRATA*

LOCAL ADAPTATION, PHENOTYPIC DIFFERENTIATION AND REPRODUCTIVE ISOLATION

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF SCIENCE

ACTA UNIVERSITATIS OULUENSIS
A Scientiae Rerum Naturalium 715

TUOMAS HÄMÄLÄ

ECOLOGICAL GENOMICS IN *ARABIDOPSIS LYRATA*

Local adaptation, phenotypic differentiation and reproductive isolation

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Health and Biosciences of the University of Oulu for public defence in Kaljusensali (KTK112), Linnanmaa, on 24 May 2018, at 12 noon

UNIVERSITY OF OULU, OULU 2018

Acta Univ. Oul. A 715, 2018

Supervised by
Professor Outi Savolainen
Docent Helmi Kuittinen
Professor Mikko J. Sillanpää

Reviewed by
Professor Laura Rose
Professor Bengt Hansson

ISBN 978-952-62-1906-6 (Paperback)
ISBN 978-952-62-1907-3 (PDF)

ISSN 0355-3191 (Printed)
ISSN 1796-220X (Online)

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2018

Hämälä, Tuomas, Ecological genomics in *Arabidopsis lyrata*. Local adaptation, phenotypic differentiation and reproductive isolation

University of Oulu Graduate School; University of Oulu, Faculty of Science
Acta Univ. Oul. A 715, 2018
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

A central goal in evolutionary biology is to identify the ecological and genetic mechanisms that give rise to adaptation and speciation. Importantly, a large body of theoretical work has modelled the adaptive evolution under selection, migration and drift. Yet to test these predictions on an empirical level has proven a challenging task. The aim of my thesis is to explore outstanding questions in local adaptation and reproductive isolation using natural populations of *Arabidopsis lyrata*: How does differential selection lead to adaptive divergence in the face of gene flow and drift? What traits underlie both short- and large-scale adaptive differentiation? And what reproductive barriers are involved in incipient speciation?

By combining whole-genome based demography simulations with a multi-year reciprocal transplant experiment, I confirmed that alpine and lowland populations of *A. lyrata* are adapted to their local environments despite high gene flow and strong drift. Patterns of trait differentiation, supported by analysis of phenotypic selection, further suggested that flowering traits have contributed to the adaptive divergence.

Selection patterns at the sequence level confirmed that the genetic architecture underlying the local adaptation conforms to theory: populations under higher levels of gene flow had fewer adaptive loci that were also found in areas of reduced recombination. Although most selection outliers were population specific, indicating conditional neutrality, a small proportion showed potential for genetic trade-offs (antagonistic pleiotropy). The analysis also revealed important traits and biological processes linked to alpine and lowland adaptation.

The role of seed germination in large-scale adaptation and reproductive isolation was also studied. Populations representing the European and North American subspecies exhibited germination patterns consistent with adaptive differentiation. Comparisons against first- and second-generation hybrids then indicated that genetic incompatibilities impede germination of the hybrid seeds. Furthermore, genetic mapping helped to clarify the genetic basis of these phenotypic traits.

Taken together, the three studies in this thesis highlight the value of combining traditional organismal methods with next-generation genomics, by providing novel insights into processes underlying adaptation and speciation.

Keywords: Arabidopsis, evolutionary genomics, gene flow, genetic drift, life-history evolution, local adaptation, natural selection, reproductive isolation

Hämälä, Tuomas, Ekologinen genomiikka idänpitkäpalolla (*Arabidopsis lyrata*). Paikallissopeutuminen, fenotyyppinen erilaistuminen ja lisääntymisesteet
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Luonnontieteellinen tiedekunta
Acta Univ. Oul. A 715, 2018
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Evoluutiobiologian keskeinen tehtävä on sopeutumiseen ja lajiutumiseen johtavien prosessien selvittäminen. Vaikka evoluutiovoimien – luonnonvalinnan, geenivirran ja geneettisen satunnaisajautumisen – vaikutusta adaptiivisen muuntelun määrään on mallinnettu laajalti, teoreettisten ennusteiden tarkastelu empiirisellä tasolla on usein osoittautunut haastavaksi. Tässä väitöskirjatyössä pyrin vastaamaan eräisiin paikallissopeutumisen ja lajiutumisen kannalta tärkeisiin kysymyksiin, hyödyntäen kasvilaji idänpitkäpalkoa (*Arabidopsis lyrata*) malliorganismina: Kuinka luonnonvalinta johtaa paikallissopeutumiseen tilanteessa, jossa geenivirta samankaltaistaa eriytyvien populaatioiden perimää? Mitkä ominaisuudet vaikuttavat adaptiiviseen erilaistumiseen eri etäisyyksillä olevien populaatioiden välillä? Sekä millaiset lisääntymisesteet johtavat alkavaan lajiutumiseen?

Yhdistämällä genomisekvensointiin perustuvan demografia-analyysin ja monivuotisen siirtoistutuskokeen, selvitin kuinka idänpitkäpalkopopulaatiot ovat sopeutuneet elinympäristöihinsä runsaasta geenivirrasta ja voimakkaasta satunnaisajautumisesta huolimatta. Fenotyyppisen muuntelun ja kelpoisuuden yhteys vahvisti myös, että kukkimisominaisuudet ovat vaikuttaneet populaatioidenväliseen adaptiiviseen erilaistumiseen.

Valinnan merkkien etsiminen sekvenssitasolla osoitti, että havaintoni paikallissopeutumisen geneettisestä arkkitehtuurista tukevat teoreettisia ennusteita: populaatioista, joihin kohdistuu voimakasta geenivirtaa, löytyi vähemmän adaptiivisia lokuksia ja ne olivat keskittyneet matalamman rekombinaation alueille. Suurin osa adaptiivisista lokuksista löytyi ainoastaan yhdestä populaatiosta, ollen näin todennäköisesti valinnan alla ainoastaan tietyssä elinympäristössä. Pieni osuus lokuksista vastasi kuitenkin harvoin havaittua tilannetta, jossa hajottava valinta on johtanut eri alleelien runsastumiseen populaatioissa, joita yhdistää geenivirta.

Tutkin myös, miten itämisajan muuntelu vaikuttaa sopeutumiseen pitkällä aikavälillä. Eurooppalaista ja pohjoisamerikkalaista alalajia edustavat populaatiot itivät tavalla, joka viittaa adaptiiviseen erilaistumiseen. Ensimmäisen ja toisen hybridisukupolven siementen vertailu paljasti lisäksi, että geneettiset yhteensopimattomuudet haittaavat hybridien itämistä, toimien näin lisääntymisesteenä. Geenikartoitus auttoi myös selventämään näiden itämisominaisuuksien geneettistä taustaa.

Tämän väitöskirjan kolme osatyötä korostavat miten perinteisten yhteiskenttäkokeiden ja uuden sukupolven genomimenetelmien yhdistelmä voi tuottaa arvokasta lisätietoa sopeutumisen ja lajiutumisen mekanismeista.

Asiasanat: evolutiivinen genomiikka, geenivirta, geneettinen satunnaisajautuminen, idänpitkäpalko, lisääntymisesteet, luonnonvalinta, paikallissopeutuminen

Acknowledgements

I would first like to thank my main supervisor Outi Savolainen for introducing me to population genetics during my undergrad studies and later on to topics of local adaptation and speciation. I am also grateful for getting the opportunity to work on this interesting Ph.D. project. My additional supervisors Helmi Kuittinen and Mikko Sillanpää have also been important sources of knowledge and inspiration throughout my studies.

My deepest thanks go to all the people who were involved in the thesis work. Päivi Leinonen, Tiina Mattila, Ulla Kemi and Jaromir Guzinski contributed to the germination experiment and QTL mapping. Tuomas Toivainen, Jaakko Tyrmi, Soile Alatalo, Hannele Parkkinen, Aino Hämäläinen, Tiina Huhtakangas, Rami-Petteri Apuli, Elina Haataja and Margarita Takou collected seeds or measured plants for the field experiments. Hans Stenøien and David Remington provided valuable information and seed material at the start of the experiments. I am especially grateful for Marit and Lars Bakkom who hosted the experiment in Lom and tirelessly helped with the measurements in Norway. The expert advice of Jaakko Tyrmi and Tiina Mattila was invaluable for conducting the sequence analysis. Laura Rose and Bengt Hansson are acknowledged for their careful review of this thesis. I further thank the members of my follow-up group, Jouni Aspi, Annamari Markkola and Ulf Lagercrantz, as well the current and former members of the Plant Genetics Group, for their support during this time.

Finally, thanks to my family and friends, whose support and acceptance over the years has been essential for accomplishing this goal. Kiitos.

Oulu, April 2018 Tuomas Hämälä

Abbreviations

β	directional selection gradient
γ	quadratic selection gradient
π	nucleotide diversity
AP	antagonistic pleiotropy
BDMI	Bateson-Dobzhansky-Muller incompatibility
CN	conditional neutrality
ddRAD	double-digest restriction site-associated DNA
d_{XY}	absolute measure of genetic differentiation
F_1	first hybrid generation
F_2	second hybrid generation
F_{ST}	relative measure of genetic differentiation
GLMM	generalized linear mixed model
GO	gene ontology
LRT	likelihood-ratio test
m	migration rate
N_e	effective population size
PBS	population branch statistic
Q_{ST}	measure of quantitative trait differentiation
QTL	quantitative trait loci
s	selection coefficient
SFS	site frequency spectrum
TRD	transmission ratio distortion

Original publications

This thesis is based on the following publications, which are referred throughout the text by their Roman numerals:

I Hämälä T, Mattila TM, & Savolainen O (2018) Local adaptation and ecological differentiation under selection, migration and drift in *Arabidopis lyrata*. Manuscript.

II Hämälä T & Savolainen O (2018) Local adaptation under gene flow: Recombination, conditional neutrality and genetic trade-offs shape genomic patterns in *Arabidopsis lyrata*. Manuscript.

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

Contents

1 Introduction

Ever since Darwin (1859), natural selection has been recognized as the primary force giving rise to adaptive evolution. Yet the specifics of how selection leads to adaptation and speciation have been under ever increasing discussion. The classical theory by Haldane (1930) and Wright (1931) showed that the effects of selection can be weakened by the influx of foreign alleles (gene flow) and random fluctuations in allele frequencies (genetic drift). How much drift a population experiences is inversely proportional to its effective size (N_e), so that large populations are more efficient at responding to selection (Robertson, 1960). Furthermore, as first demonstrated by Haldane (1930), the strength of selection (s) must exceed the migration rate (m) for locally favored allele to be maintained in a continent-island model. This model was later expanded to two-island system by Maynard Smith (1970) and Bulmer (1972) who found that the beneficial allele will more easily become lost if selection coefficients between the two habitats are not in balance (i.e. $m/s > \alpha/[1-\alpha]$, where α is the ratio of selection coefficients between populations). Since then, several authors have examined the interplay of selection and migration with different demographic and environmental models, and concluded that adaptive evolution under gene flow is conditional on the strength of selection acting above the critical migration threshold (Felsenstein, 1976; Lenormand, 2002).

1.1 Local adaptation

When spatially varying selection overcomes the opposing effects of gene flow and drift, local adaptation may evolve. When this happens, populations have higher fitness at their home environments than any non-local populations introduced to these sites (Kawecki & Ebert, 2004). Requirements for the formation of local adaptation are most easily met when environments vary across latitudes. Indeed, several studies have demonstrated large-scale local adaptation in herbaceous species (Clausen, Keck, & Hiesey, 1940; Colautti & Barrett, 2013; Hall & Willis, 2006; Toräng et al., 2015; Ågren & Schemske, 2012) and in forest trees (Savolainen, Pyhäjärvi, & Knürr, 2007). However, adaptive divergence may also occur among closely adjacent populations if they are found along steep environmental gradients. An extreme form of short-scale local adaptation has been observed at boundaries of mine tip and pasture soil, where plants have developed tolerance to heavy metals despite free gene flow from neighbouring areas (Antonovics & Bradshaw, 1970;

McNeilly, 1968). Selection pressures may vary rapidly also in mountainous environments (Körner, 2007). Microgeoraphical adaptation to altitude specific habitats has been found, for example, in *Festuca eskia* (Gonzalo-Turpin & Hazard, 2009) and in *Boechera stricta* (Anderson, Perera, Chowdhury, & Mitchell-Olds, 2015).

Two main models have been invoked to explain mechanisms of local adaptation at the single-locus level: antagonistic pleiotropy (AP) and conditional neutrality (CN) (Kawecki & Ebert, 2004). Under antagonistic pleiotropy, locus is maintained polymorphic by a fitness trade-off, so that an allele is selected for in one environment and selected against in the alternative environment. Conditional neutrality, on the other hand, means that an allele is under selection (positive or negative) only in one environment, while being neutral in the other (Fig. 1). A major question in ecological genetics concerns the relative roles of AP and CN in local adaptation, and whether individual loci should mirror fitness trade-offs commonly found at the organismal-level (Savolainen, Lascoux, & Merilä, 2013; Tiffin & Ross-Ibarra, 2014; Wadgymar et al., 2017). By searching for correlations between genotypes and fitness, researcher have begun to approach these questions empirically (Anderson, Lee, Rushworth, Colautti, & Mitchell-Olds, 2013; Leinonen, Remington, Leppälä, & Savolainen, 2013; Lowry, Hall, Salt, & Willis, 2009; Ågren, Oakley, McKay, Lovell, & Schemske, 2013). In general, these studies are in agreement with exciting theory (Martin & Lenormand, 2006) by highlight the role of CN over AP in promoting local adaptation across large geographical scales. Yet, more comprehensive answers stemming from genome-wide analysis of selection and demographic histories are currently lacking.

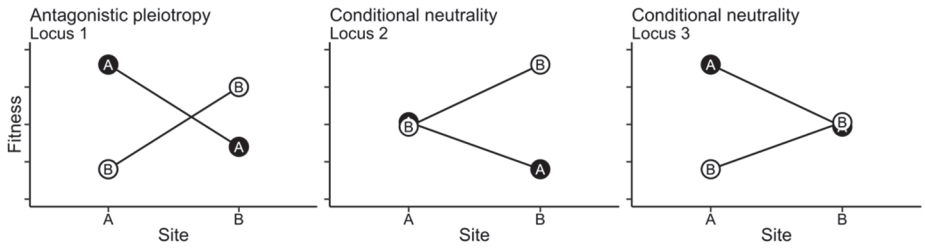

Fig. 1. Maintenance of local adaptation at the single-locus level. Antagonistic pleiotropy: alleles show a fitness trade-off in contrasting sites. Conditional neutrality: allele confers higher (or lower fitness) fitness in one site while being neutral in the other. Opposite pattern is observed at a different locus. Redrawn from Savolainen et al. (2013).

1.1.1 Adaptive variation under selection, migration and drift

The theory on migration-selection balance states that locally beneficial alleles will be lost if selection is not strong enough to overcome the homogenizing effects of the gene flow (Haldane, 1930). Under this scenario of "gene swamping", the overall beneficial allele will be fixed at all populations, preventing adaptive divergence (Lenormand, 2002). Alternatively, under sufficient selection with balanced coefficients between habitats ($\alpha \approx 1$), the frequency change by migration may be cancelled out by selection and the polymorphism is maintained at a selection-migration equilibrium (Felsenstein, 1976). However, in finite populations the probability of reaching this equilibrium is further influenced by genetic drift (Yeaman & Otto, 2011). Alleles with small effect sizes can be easily lost through drift (Yeaman, 2015), so even under low migration rates selection has to be strong for local adaptation to evolve among small populations.

The aforementioned single-locus models can help to elucidate the basic processes underlying divergence with gene flow, but as local adaptation is often based on quantitative traits (Hereford, 2009), more comprehensive theory is needed for realistic modelling. Indeed, quantitative genetic models point towards a more complex interplay between selection, migration and drift (Blanquart, Gandon, & Nuismer, 2012; Hendry, Day, & Taylor, 2001), but even these models cannot reach the complexity of natural systems (Savolainen et al., 2013). The theory of quantitative trait evolution can, however, be useful when predicting the genetic architecture underlying local adaptation with gene flow (Griswold, 2006; Yeaman & Whitlock, 2011). First, a shift towards fewer large effect loci is expected, as they are under stronger selection than small effect ones. Second, loci underlying local adaptation should be concentrated in areas of reduced recombination, because selection is more likely to maintain differences in a set of correlated loci. And third, divergence is more likely to result from genetic trade-offs (antagonistic pleiotropy), because allele beneficial in one environment, while neutral in the other (conditional neutrality), would get fixed in both populations over time. Empirical studies have suggested that large effect loci may be prevalent in short-scale adaptation (Ferris, Barnett, Blackman, & Willis, 2017; Hendrick et al., 2016) and indications of recombination induced clustering of the adaptive loci have also been found in stickleback populations under supposedly high gene flow (Marques et al., 2016; Samuk et al., 2017). Yet, the important prediction that antagonistic pleiotropy is an important driver of adaptive divergence under gene flow is still lacking empirical support.

1.1.2 Common garden and reciprocal transplant experiments

To demonstrate genotype-by-environment interactions for fitness, genotypes from different populations have to be compared in the same conditions. These types of common garden experiments are among the oldest tools used in local adaptation research (Kawecki & Ebert, 2004). Often the easiest way to perform a common garden experiment is to replicate the most important environmental properties in a laboratory and to test individuals from different populations there. In many insects and other mobile animals this approach is the only possible one, but it is also frequently used in plants for practical reasons (Dittmar, Oakley, Ågren, & Schemske, 2014; Quilot-Turion et al., 2013). However, field testing in plants and other organisms that allow for it is often preferable, because only then can individuals be exposed to full extent of natural variation. For example, a common garden approach has been used to get the first genomic view of local adaptation in *Arabidopsis thaliana* (Fournier-Level et al., 2011).

Among different methods to study local adaptation, reciprocal transplant experiments are in many ways the most credible, because they offer a direct estimate of local adaptation by comparing fitness of the local population to that of the introduced population (Blanquart, Kaltz, Nuismer, & Gandon, 2013). Such experiments have demonstrated local adaptation by transplanting populations over long (Clausen, Keck, & Hiesey, 1948; Colautti & Barrett, 2013; Hall & Willis, 2006; Leinonen, Remington, & Savolainen, 2011; Toräng et al., 2015) and short (Gonzalo-Turpin & Hazard, 2009; Sambatti & Rice, 2006) geographical distances. Meta-analyses of published results have further shown that approximately 45% – 70% of reciprocal transplant studies have shown evidence of local adaptation (Hereford, 2009; Leimu & Fischer, 2008). However, the failure to detect these patterns does not necessarily mean that populations are not locally adapted, as seen in a study by Ågren et al. (2013) in which the strength of local selection varied over time. In addition to local adaptation, reciprocal transplant experiments have the potential to reveal phenotypic plasticity (West-Eberhard, 2003) when the same genotypes show different phenotypes in contrasting sites (Merilä & Hendry, 2014).

1.2 Differentiation in quantitative traits

The study of local adaptation commonly involves estimating spatial differences in survival and reproductive output, which serve as a measure for Darwinian fitness (Kawecki & Ebert, 2004). However, this approach alone does not inform us what

traits are under differential selection, and thus contribute to adaptive divergence. In some cases, phenotypic differentiation in quantitative traits can be seen as a mark of selection. For example, in the early 1900s, Turesson (1922) observed considerable variation within plant species grown in common environments, indicating genetic differentiation. Although some differentiation is likely due to selection, not all variation is adaptive. Therefore, more advanced methods are needed to separate the effects of selection from the effects of random drift. A classical way involves finding phenotypic clines across environmental gradients, which are thought to reflect patterns of spatially varying selection (Endler, 1977). Other tools include quantifying the strength of phenotypic selection through regression analysis of the relationship between phenotype and fitness (Lande & Arnold, 1983) or comparing the divergence in quantitative traits (Q_{ST}) to divergence in neutral loci (F_{ST}) (Merilä & Crnokrak, 2001).

To examine the genetic basis of quantitative trait differentiation, one can use genetic mapping to link phenotype and genotype level variation (Savolainen et al., 2013). Quantitative trait locus (QTL) mapping utilizes recombinant mapping families from crosses between populations. Large number of progeny can be generated using recombinant inbred lines in self-fertilizing species or three-generation crosses in outcrossing species. These can be then genotyped and studied in different environments (including laboratory and experimental fields). Such QTL analyses have been widely employed to study genetic basis of local adaptation and quantitative trait divergence (Anderson et al., 2013; Hall & Willis, 2006; Leinonen et al., 2013; Lowry et al., 2009; Wadgymar et al., 2017). Association mapping, on the other hand, searches for statistical associations between genotypes and phenotypes in natural populations. Compared to mapping families, lower levels of linkage disequilibrium in natural populations often allow for finer scale results, but it also imposes a higher demand for number of individuals and genetic markers. Moreover, association analysis can be complicated by spurious signals caused by genetic structure in the sampling population and unknown relatedness between individuals (McCarthy et al., 2008). Nevertheless, in large-scale studies, genome-wide association mapping (GWAS) has proven to be a powerful tool to infer genetic variation underlying local adaptation (1001 Genomes Consortium, 2016; Fournier-Level et al., 2011; Stanton-Geddes et al., 2013) The emergence of next-generation sequencing has greatly benefitted genetic mapping studies, as reduced-representation sequencing techniques, such as restriction-site associated sequencing (RADseq) (Baird et al., 2008) and genotyping-by-sequencing (GBS)

(Elshire et al., 2011), have made large-scale genotyping cost-effective also in non-model species (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016).

Over the decades, researchers have found ample evidence that selection has shaped life history and morphological traits in several plant species (Linhart & Grant, 1996). Although the number of traits influenced by selection is considerable, those involved in phenology – the timing of life history traits – seem to have a special role in adaptation (Donohue, Rubio De Casas, Burghardt, Kovach, & Willis, 2010; Munguía-Rosas, Ollerton, Parra-Tabla, & De-Nova, 2011; Savolainen et al., 2007). Germination, or the change from seed to seedling, is arguably the first major life history transition in plants. As such, it often determines early success of the seedlings, while also influencing important traits expressed at later life-stages (Donohue et al., 2010). In fact, the timing of germination has had a proven effect on life-time fitness in *A. thaliana* (Donohue et al., 2005), which through geographical variation has also contributed to local adaptation (Postma & Ågren, 2016). Timing of germination is mainly controlled by seed dormancy, a physiological state which prevents viable seeds from germinating even under favourable conditions (Finch-Savage & Leubner-Metzger, 2006), but differences in postdormancy germination may also prove to be adaptive (Hämälä, Mattila, Leinonen, Kuittinen, & Savolainen, 2017). Another significant life history transition is the initiation of flowering. A large body of evidence has shown flowering time variation to be adaptive in several species (Hall & Willis, 2006; Leinonen et al., 2013; Schemske, 1984; Ågren, Oakley, Lundemo, & Schemske, 2017). The observed latitudinal (Stinchcombe et al., 2004) and altitudinal (Montesinos-Navarro, Wig, Xavier Pico, & Tonsor, 2011) clines further suggest that growing season length is often driving the trait divergence.

1.3 Footprints of selection on the DNA level

Although organismal methods can give valuable insights into traits and processes underlying local adaptation, focusing on preselected phenotypes may cause important factors to be overlooked. An alternative, and at best of cases complementary, way is to directly scan the genome of the study organism for signs of selection (Tiffin & Ross-Ibarra, 2014). Common approaches include searching for unexpectedly high divergence between populations (Lewontin & Krakauer, 1973), reduced diversity within populations (Maynard Smith & Haigh, 1974) and excess of rare variants at loci potentially under directional selection (Braverman, Hudson, Kaplan, Langley, & Stephan, 1995). Several statistics have been proposed

for measuring genetic differentiation (Hedrick, 2011), but most selection scan methods are still based on Wright's fixation index F_{ST} (reviewed e.g. in Lotterhos & Whitlock 2015). Despite their ubiquity, F_{ST} based applications can suffer from analytical problems, such as propensity to detect areas where genetic diversity has been reduced without accompanying increase in divergence (Charlesworth, 1998; Cruickshank & Hahn, 2014). An additional issue that affects all selection scans is related to outlier detection. If the demographic history and population structure are unaccounted for, a large proportion of the putatively selected loci can be erroneous (Hoban et al., 2016; Lotterhos & Whitlock, 2015).

An effective way to increase the accuracy of these methods is to model and take into account the demographic history of the study populations (Hoban et al., 2016). A commonly used approach is to infer neutral demography parameters with coalescent modelling (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013; Gutenkunst, Hernandez, Williamson, & Bustamante, 2009), and use the information to build null distributions for outlier detection. Coalescent theory, first introduced by Kingman (1982) and later expanded by Tajima (1983) and Hudson (1991), models how different genetic variants originated from a common ancestor. By considering different population histories, including gene flow, effective population sizes, divergence times and recombination rates, simulated genealogies can be compared against the observed ones to assess the most likely parameter space under which the data originated (Wakeley, 2009). Using such approach, Huber, Nordborg, Hermisson & Hellmann (2014) examined data previously published by Long et al. (2013), and concluded that two thirds of the originally identified selective sweeps were false positives caused by complex demographic histories.

1.4 Reproductive isolation

Over time, the genetic divergence can lead to evolution of reproductive isolation (Coyne & Orr, 2004). Genetic incompatibilities between epistatic alleles, commonly called Bateson-Dobzhansky-Muller incompatibilities (BDMI) (Bateson, 1909; Dobzhansky, 1936; H. J. Muller, 1939), are large drivers of this process. These incompatibilities form when diverging populations accumulate new alleles that are adaptive or neutral in their own genetic environment, but incompatible when in the same genetic environment. Issues emerge when populations come into secondary contact and produce hybrids that have inferior fertility or are not viable – that is, they have lower fitness than pure-population individuals and are therefore

selected against. For example, between species hybrids in *Mimulus* (Fishman & Willis, 2001) and in *Helianthus* (Owens & Rieseberg, 2014) have exhibited inviability consistent with the BDM-model.

The development of BDMIs are often thought to occur between nuclear loci, but the negatively interacting alleles may also be found in the organellar genomes of the mitochondria or chloroplast (Levin, 2003). When the incompatibility forms among nuclear and cytoplasmic genes, the resulting reproductive barrier will appear asymmetrically between the hybrid reciprocals (Burton, Pereira, & Barreto, 2013). Darwin (1859) was among the first to notice that this form of asymmetry is especially common among angiosperms. Indeed, a meta-analysis consisting of 14 angiosperm genera showed that up to 45% of the observed reproductive barriers were not symmetric (Tiffin, Olson, & Moyle, 2001). However, cytonuclear interactions are not the only causes for asymmetry, but incompatibilities stemming from genomic imprinting, maternal gene products or the triploid endosperm (Haig & Westoby, 1991; Turelli & Moyle, 2007) may also lead to similar patterns.

1.5 *Arabidopsis lyrata* as a model for ecological genetics

Arabidopsis lyrata is a perennial and predominantly outcrossing plant species in the family Brassicaceae. It has a wide circumpolar distribution that comprises two subspecies: ssp. *petraea* in Eurasia and ssp. *lyrata* in North-America (Jalas & Suominen, 1994). Both subspecies consist of several isolated populations that are genetically (Mattila, Tyrmi, Pyhäjärvi, & Savolainen, 2017; M-H. Muller, Leppälä, & Savolainen, 2008; Pyhäjärvi, Aalto, & Savolainen, 2012) and phenotypically (Kärkkäinen, Løe, & Ågren, 2004; Quilot-Turion et al., 2013; Remington, Figueroa, & Rane, 2015) differentiated. Hybrids between the subspecies have also exhibited signs of reproductive isolation (Aalto, Koelewijn, & Savolainen, 2013; Leppälä, Bokma, & Savolainen, 2013; Leppälä & Savolainen, 2011), potentially indicating incipient speciation. *A. lyrata* is a close relative of the model species *A. thaliana*, but their differences in distribution and life-history traits have made *A. lyrata* a valuable model to study local adaptation and incipient speciation (Savolainen & Kuittinen, 2011). The high quality reference genome (Hu et al., 2011) and wide array of molecular tools also make it a good system for genetic studies.

Previous studies have shown evidence of large-scale local adaptation among *A. lyrata* populations from Europe and North America (Leinonen et al., 2009, 2011; Vergeer & Kunin, 2013). For example, by reciprocally transplanting individuals from Norway and North Carolina, Leinonen et al. (2011) showed that these

populations have adapted to their native habitats. Furthermore, phenotype selection analysis and QTL mapping indicated that flowering phenology has contributed to the adaptive divergence (Leinonen et al., 2013, 2011). Other organismal studies have also highlighted the role of flowering time variation in *A. lyrata* adaptation (Quilot-Turion et al., 2013; Riihimäki & Savolainen, 2004; Sandring, Riihimäki, Savolainen, & Ågren, 2007), while genetic analyses have found footprints of selection on the underlying genes (Kemi et al., 2013; Mattila et al., 2016; Toivainen, Pyhäjärvi, Niittyvuopio, & Savolainen, 2014).

1.6 Aims of the study

The aim of my thesis is to examine the evolutionary consequences of differential selection: How does it lead into local adaptation despite the opposing effects of migration and drift? When and how does it result in reproductive isolation? And how are quantitative traits affected by it? To approach these questions empirically, I conducted experiments using natural populations of *A. lyrata*.

First (I), I studied populations originating from two altitudinal gradients to relate estimates of local adaptation to quantification of migration rates and effective population sizes. Are lowland and alpine populations adapted to their local environments despite ongoing gene flow and small effective population sizes? And does differentiation in flowering traits contribute to local adaptation? To this end, I conducted a reciprocal transplant trial, a common garden trial and a whole-genome based demography analysis.

Second (II), using the same populations, I examined the genetic architecture underlying local adaptation with gene flow. Do I find evidence of antagonistic pleiotropy and are the observed spatial patterns consistent with theory: fewer adaptive loci that are concentrated in areas of reduced recombination? And are the same traits under differential selection in both altitudinal gradient? The experiment was conducted using whole-genome sequencing data.

Third (III), I examined differential selection over large geographical scales by comparing the two subspecies. Has the genetic divergence led to formation of reproductive isolation? Does it affect the viability of the hybrid seeds? And do germination traits show phenotypic and genetic variation consistent with adaptive differentiation? These questions were explored by measuring germination in controlled conditions and by conduction genetic mapping.

2 Material and methods

Materials and methods are described here briefly. More details can be found in the original papers (I–III).

2.1 Study populations and crosses

The experiments included *A. lyrata* individuals from 12 populations. Most of the material was collected from two alpine areas in Norway: Jotunheimen and Trollheimen (I, II). Both areas were represented by four populations: J1 (300 m.a.s.l), J2 (500 m.a.s.l), J3 (1,100 m.a.s.l) and J4 (1,200 m.a.s.l) in Jotunheimen; and T1 (10 m.a.sl.), T2 (210 m.a.s.l), T3 (900 m.a.s.l) and T4 (1,400 m.a.s.l) in Trollheimen. Additionally, populations from Germany (abbreviated as GER in papers I and II), Sweden (abbreviated as SWE in paper II), North Carolina (abbreviated as NC in papers I and II, and as Ma in paper III) and Russia (abbreviated as Kar in paper III) were used.

In paper I, phenotype data were collected from individuals derived from independent full-sib crosses. Sequence data used in papers I and II originated from field collected plants that were used as parents in the crossing program.

A more complex crossing design was used in paper III. Individuals from Karhumäki, Russia (Kar), representing the subspecies ssp. *petraea*, and Mayodan, North Carolina (Ma), representing the subspecies ssp. *lyrata*, were crossed within and between populations to produce a progeny set consisting of pure-populations and reciprocal F_1 and F_2 hybrids. Kar individuals were pollinated with Ma pollen to get F_1 hybrids with Kar cytoplasm (abbreviated as F_1-KarMa) and Ma individuals were pollinated with Kar pollen to get hybrids with Ma cytoplasm (abbreviated as F_1-MaKar). Independent F_1 reciprocals were then crossed to produce F_2 hybrids with different cytoplasms (abbreviated as F_2-KarMa and F_2-MaKar).

2.2 Phenotype experiments

Phenotypic differentiation among populations was studied in papers I and III. Paper I describes a reciprocal transplant experiment and a common garden experiment used to measure fitness and flowering trait differentiation among populations from Norway (eight in total), Germany and North Carolina. Two experimental fields were established in Jotunheimen, Norway. A low-altitude site was situated in Lom,

close to the natural growing site of the J1 population (300 m.a.s.l), and a high-altitude site was situated in Spiterstulen, close to the natural growing site of the J3 population (1,100 m.a.s.l). Four Norwegian populations (J1, J3, T1 and T3), as well as GER and NC, were planted into these fields. Additionally, a third experimental field was established in Oulu, Finland (12 m.a.s.l), which included plants from all eight Norwegian populations and from GER and NC. Plants in all three field sites were followed during three consecutive growing seasons: 2015, 2016 and 2017. Flowering start dates, fruit production and survival were measured in the Norwegian fields, while additional measurements were made in Oulu for flowering shoot lengths, inflorescence numbers, fruit maturation dates and flowering cessation dates.

The experiment in paper III was conducted in controlled conditions. Seeds from the parental populations and from the different hybrid classes (F_1-KarMa, F_1-MaKar, F_2-KarMa and F_2-MaKar) were divided into four groups based on after-ripening and cold stratification treatments: half of the seeds were kept in after-ripening (dry storage in stable conditions) for eight weeks and half for 19 weeks. Seeds were planted into agar plates and half of the plates were exposed to cold conditions (4°C) for four days and half for eight days. These treatments were used to explore seed dormancy variation among the parental and hybrid classes. Plates were then moved into a growth cabinet to induce germination. Plates were inspected daily for 28 days to measure germination time and germination success. The data were used to estimate whether postdormancy germination time varies between Kar and Ma populations, possibly reflecting adaptive differentiation, and whether hybrids have lower germination proportions than pure-populations, suggesting reproductive isolation.

2.3 Next-generation sequencing

Next-generation sequence data were used in all three studies. Papers I and II were mainly based on the same whole-genome data sets. Sampled individuals were from J1 (n = 9), J3 (n = 7), T1 (n = 5), T3 (n = 5) and T4 (n = 9) populations. The T3 data were used only in paper I and the T4 data were used only in paper II. DNA was extracted from fresh leaves with NucleoSpin Plant II kit (Macherey-Nagel) and libraries for Illumina whole-genome sequencing were prepared with NEBNext master mix kit (New England Biolabs). Illumina HiSeq2500 (Institute of Molecular Medicine Finland [FIMM], University of Helsinki) was used to sequence samples from J1, J3, T1 and T4, while samples from T3 were sequenced with Illumina

NextSeq550 (Biocenter Oulu, University of Oulu). Additionally, previously published whole-genome data from J3 (n = 5; total n = 12), GER (n = 6), NC (n = 6; paper I) and SWE (n = 6; paper II) were used (Mattila et al., 2017).

A reduced representation sequencing approach was used in paper III. Seedlings from two reciprocal F_2 families were grown to adults after the germination experiment ended. The following number of individuals were sampled "KarMa1" n = 87, "MaKar1" n = 85, "KarMa2" n = 93 and "MaKar2" n = 89. DNA was extracted from silica dried leaves with DNeasy 96 Plant kits (Qiagen) and double digested with restriction enzymes *MfeI* and *NlaIII* (New England Biolabs). Library preparation for double-digest restriction site-associated DNA (ddRAD) sequencing was conducted according to Peterson, Weber, Kay, Fisher, & Hoekstra (2012). Libraries were sequenced with Illumina HiSeq2500 (FIMM, University of Helsinki). Whole-genome data from Kar (n = 4) and Ma (n = 6) were also used (Mattila et al., 2017).

Similar methods were used to process the sequence data in all three studies. Low quality reads and sequencing adapters were filtered out with Trimmomatic (Bolger, Lohse, & Usadel, 2014). Reads were aligned to *A. lyrata* reference genome (Hu et al., 2011) with bwa-mem (Li & Durbin, 2009). Duplicated reads were removed with Picard tools (http://broadinstitute.github.io/picard/) and indels realigned with GATK (McKenna et al., 2010). Genotype likelihoods used in paper II were estimated with ANGSD (Korneliussen, Albrechtsen, & Nielsen, 2014) and variant calling in paper III was done with Freebayes (Garrison & Marth, 2012).

2.4 Statistical methods

2.4.1 Fitness and phenotype analysis

Individual trait measurements in papers I and III were analysed with general and generalized linear mixed models (GLMM) in the R (R Core Team, 2017) package lme4 (Bates, Maechler, Bolker, & Walker, 2015). Likelihood ratio tests (LRT) were used to define population differentiation of fitness and flowering traits in paper I and of germination traits in paper III by comparing the fit of a full model to a reduced model with pair of populations combined as one category.

Additional analyses were conducted in paper I to estimate fitness and selection differences among populations and experimental fields. The presence of local adaptation was evaluated with aster models (Geyer, Wagenius, & Shaw, 2007;

Shaw, Geyer, Wagenius, Hangelbroek, & Etterson, 2008) by obtaining hierarchical multi-year fitness estimates for each population. Total fitness estimates were inferred from three-year survival, flowering propensity and fruit production. Differences between populations were defined with LRTs, as in the case of individual traits.

Quantitative trait differentiation among the Norwegian populations was estimated with Q_{ST}: $\sigma^2_{GB} / (\sigma^2_{GB} + 2\sigma^2_{GW})$, where σ^2_{GB} and σ^2_{GW} are the between and within population variance components, respectively. The variance components were inferred with Markov chain Monte Carlo based regression models in the R package MCMCglmm (Hadfield, 2010). The whole-genome data from Norwegian populations was then used to estimate a global neutral F_{ST}, which was compared against the Q_{ST} estimates to distinguish the effects of selection ($Q_{ST} > F_{ST}$) from drift ($Q_{ST} = F_{ST}$) (Merilä & Crnokrak, 2001). To quantify selection differences between the field sites, phenotypic selection analysis was conducted with aster models (Shaw & Geyer, 2010). Directional (β) and quadratic (γ) selection gradients were used to define the relationship between flowering start dates and fitness at the three field sites. The slope and curvature of this relationship can give insights into the strength and type of selection (directional, stabilizing or disruptive) acting on the trait (Lande & Arnold, 1983; Shaw & Geyer, 2010).

2.4.2 QTL mapping

Quantitative trait loci (QTL) mapping was used to find genomic areas underlying postdormancy germination time in paper III. Two independent F_2 families were analysed, with 1,520 markers in family 1 (n = 172) and 1,794 markers in family 2 (n = 182). The linkage map was constructed with Joinmap 4.1 (Van Ooijen, 2006) and the QTL were detected with R/qtl (Broman, Wu, Sen, & Churchill, 2003). Haley-Knott regression (Haley & Knott, 1992) with 1,000 permutations was used to find significant correlations between genotypes and phenotypes.

2.5 Sequence analysis

2.5.1 Demography simulations

Coalescent simulations were used to infer population based demography parameters in papers I and II. Site frequency spectra (SFS) of the derived variants

were estimated for 4-fold degenerate sites with ANGSD (Korneliussen et al., 2014), and the demography models were fitted to these in fastsimcoal2 (Excoffier et al., 2013). Effective diploid population sizes (N_e), population migration rates ($4N_em$) and divergence times in number of generations were estimated. The Akaike information criterion (AIC) was used to define which migration model (no migration, unidirectional migration or bidirectional migration) fits the Norwegian data best.

2.5.2 Selection analysis

Selection analysis was conducted in paper II to find loci underlying the short-scale local adaptation. Selected sites were detected by scanning the chromosomes for higher than neutral differentiation between the populations (Lewontin & Krakauer, 1973). Allele frequencies were inferred directly from the genotype likelihoods using a model by Kim et al. (2011). This method is less biased by uncertain genotypes resulting from low coverage sequencing data, as it takes into account the likelihoods of all genotypes (two homozygotes and a heterozygote in biallelic sites) at each position. Pairwise differentiation was estimated with Hudson's F_{ST} (Bhatia, Patterson, Sankararaman, & Price, 2013; Hudson, Slatkin, & Maddison, 1992) and d_{XY} (Cruickshank & Hahn, 2014; Nei, 1987). To detect the selected lineage, differentiation between the focal populations (J1-J3 and T1-T4) was compared to an outgroup (GER) using a measure called population branch statistic (PBS) (Yi et al., 2010). Genome-wide recombination rates were estimated from the linkage map constructed in paper III, and used in combination with the demography parameters to simulate neutral samples with ms (Hudson, 2002) to obtain null distributions for outlier detection. The observed estimates were compared against quantiles of the simulated distributions to define *p*-values, which were subsequently transformed into *q*-values (Storey & Tibshirani, 2003) to reduce the multiple testing bias. I implemented this PBS scan and outlier detection method into a new C program, PBScan (https://github.com/thamala/PBScan). The population specific outliers were then examined for enrichment of gene ontology (GO) terms (Ashburner et al., 2000) using PANTHER tools (Mi et al., 2017).

2.5.3 TRD mapping

The sequenced F_2 families in paper III were analysed for the presence of transmission ratio distortion (TRD) to map the locations of incompatibility loci.

Under Mendelian segregation, the biallelic markers are expected to have three genotypes (Kar homozygote, KarMa heterozygote and Ma homozygote) that are in ratios 1:2:1. Deviation from these ratios indicates a selective loss of genotypes, most likely caused by genetic incompatibilities between the parental populations. Differences between the hybrid reciprocal can further indicate negative interactions between nuclear loci and uniparentally inherited factors. Significance of the deviation was defined with standard chi-squared tests.

3 Results and discussion

3.1 Local adaptation under high gene flow and strong drift

Demography simulations conducted in papers I and II indicated gene flow between the neighbouring low- and high-altitude populations. The Jotunheimen populations, J1 and J3, exchanged migrants at significantly higher rates than population from Trollheimen: T1 and T3 (I), as well as T1 and T4 (II). Furthermore, the gene flow estimates in all three cases were asymmetric, with migration from high to low altitudes being several times more frequent that from low to high altitudes. The estimated effective population sizes were also small in the Norwegian populations (maximum $N_e < 6{,}000$), suggesting that differential selection has to be strong to overcome the effects of migration and drift. Nevertheless, the reciprocal transplant experiment in paper I showed evidence of local adaptation between the Jotunheimen populations at the level of hierarchical multi-year fitness (Fig. 2A). The examination of fitness components further revealed that local adaptation was attributed to fecundity in the lowland site and to viability in the alpine site. However, the patterns of altitude specific adaptation were not replicated among the Trollheimen populations (Fig. 2B). The experimental fields in Jotunheimen may not be representative of Trollheimen growing sites, and/or the adaptive potential of these populations may be constrained by a combination of short divergence time, small effective population sizes and gene flow.

Earlier reciprocal transplant experiments have documented large-scale local adaptation among *A. lyrata* populations from Europe and North America (Leinonen et al., 2009, 2011; Vergeer & Kunin, 2013). However, when local adaptation forms over these large geographical distances, populations diverge under independent selection pressures. In contrast, patterns described in paper I are the outcome of a fundamentally different process, as local selection in the alpine and lowland populations is continuously challenged by high gene flow and strong drift. Local adaptation under such conditions have rarely been found in nature (Gonzalo-Turpin & Hazard, 2009; Richardson, Urban, Bolnick, & Skelly, 2014; Sambatti & Rice, 2006), and – to the best of my knowledge – never in *Arabidopsis* species. The study in paper I is also the first one to relate estimates of hierarchical multi-year fitness to whole-genome based quantification of migration rates and effective population sizes.

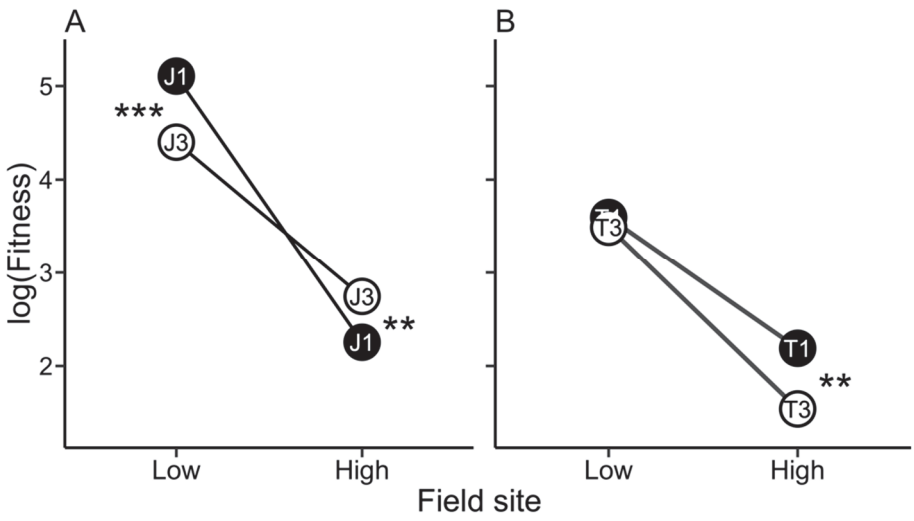

Fig. 2. Total fitness estimates (log-transformed) in the Norwegian field sites (paper I). The hierarchical aster models included three-year survival, flowering propensity and fruit production. Stars indicate significant pairwise differences: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. A: Jotunheimen populations. J1 is local to the low-altitude field and J3 is local to the high-altitude field. B: Trollheimen populations. T1 is the low-altitude population and T3 is the high-altitude population.

3.2 Genetic architecture of local adaptation under gene flow

Assessing the genetic architecture of adaptive traits (i.e. the number of loci, their effect sizes and spatial orientation within the genome) can give insights into evolutionary processes underlying local adaptation (Wadgymar et al., 2017). Importantly, how gene flow shapes the architecture of associated genes has been under recent theoretical discussion (Griswold, 2006; Yeaman & Whitlock, 2011). Theory predicts that when local adaptation forms under gene flow, the genetic architecture should move towards fewer large effect loci that are concentrated in areas of reduced recombination. Antagonistic pleiotropy is also expected to be more common, as alleles neural in one environment would get fixed at all populations over time (Yeaman & Whitlock, 2011). To expand the study of the first paper into genetic level, I conducted whole-genome based selection analyses with PBS in paper II.

Outlier loci in the low-altitude populations J1 and T1 were found in areas of lower recombination rates than in the high-altitude populations J3 and T4. The neutral simulation based outlier detection model also predicted fewer adaptive loci in the low-altitude populations. Based on the theoretical expectations, both observations are in line with higher gene flow going towards the lowland populations. Similar patterns have previously been found in sticklebacks (Marques et al., 2016; Samuk et al., 2017), but as those studies relied on traditional F_{ST} outlier tests that were conducted without quantification of migration rates, comparisons had to be limited to population pairs. In contrast, the combination of demography simulations and PBS analysis in paper II allowed for more in-depth analysis by estimating lineage specific patterns. Furthermore, to the best of my knowledge, corresponding results have not been reported in any other plant taxa.

The use of PBS also made it possible to evaluate whether loci have been under selection in one or in both of the populations. While most outliers were population specific, likely indicating conditional neutrality, a small (~3% [based on 10,000 randomized data sets, ~1.5% of the outliers are expected to be shared by chance]) proportion were found as outliers in both focal lineages, indicating possible antagonistic pleiotropy. To further explore this possibility, the outlier loci were examined for signals of selective sweeps: reduced genetic diversity (Maynard Smith & Haigh, 1974) and excess of rare variants (Braverman et al., 1995). One locus, *XRN2*, exhibited particularly clear patterns of allelic trade-off between the T1 and T4 populations. Both lineages showed low nucleotide diversities and negative Tajima's *D* estimates around the gene, while still being highly diverged from each other (as measured by F_{ST} and d_{XY}). Forward simulations conducted with SLiM 2 (Haller & Messer, 2017) further indicated that the selection patterns have likely formed under two opposing hard sweeps with strong selection coefficients. The traditional method to examine conditional neutrality and antagonistic pleiotropy has been to use genetic mapping on ecologically important traits (Anderson et al., 2013; Fournier-Level et al., 2011; Leinonen et al., 2013; Ågren et al., 2013). However, this approach has the weakness of relying on measurements of pre-selected traits, which can exclude important factors. Furthermore, the causative genes underlying the often-wide QTL intervals have rarely been discovered. In contrast, the sequence analysis based approach used in paper II is independent of phenotypes, while also providing very high resolution. On the other hand, finding footprints of selection from small whole-genome data sets can be difficult, which might partially explain the scarcity of antagonistic pleiotropy signals. Specifically, the lack to power to distinguish selection in both lineages,

likely influenced by the difficulty to detect incomplete, polygenic or soft sweeps (Hermisson & Pennings, 2017; Wellenreuther & Hansson, 2016), might lead to systematic bias against finding signals of antagonistic pleiotropy.

3.3 Flowering traits contribute to short-scale local adaptation

As flowering phenology is known to influence large-scale local adaptation in several plant species (Hall & Willis, 2006; Leinonen et al., 2011; Ågren et al., 2017), the study in paper I also examined whether differentiation in flowering traits is involved in adaptive divergence between the lowland and alpine populations of *A. lyrata*. Indeed, several lines of evidence converged to point out that differential selection on flowering time contributes to the observed patterns. First, the Q_{ST}-F_{ST} comparisons suggested that population divergence in flowering start dates and shoot lengths is caused by selection. Second, the traits showed clinal variation among the Jotunheimen groups, with high-altitude populations flowering earlier and producing shorter flowering shoots than low-altitude populations. And third, the phenotype selection analysis indicated that the alpine environment imposes stronger selection towards earlier flowering than the lowland environment. These results bear a resemblance to those commonly observed across latitudes (Endler, 1977; Munguía-Rosas et al., 2011; Stinchcombe et al., 2004), suggesting that different growing season lengths are likely underlying the local selection pressures.

3.4 Genes and biological processes underlying local adaptation

Besides evidence of phenotypic selection in paper I, the analysis in paper II revealed footprints of selection at the sequence level. Overall, genes involved in reproduction, either through flower development or the control of vegetative to reproductive transition, were strongly represented among the outlier loci (Table 1). Flower development gene *NGA4* (Lee et al., 2015) localized within the highest-ranking outlier window in the high-altitude population J3, and it was also found among the top outlier loci in the low-altitude population T1. As the corresponding low (J1) and high (T4) pairs did not show elevated differentiation, this result likely reflects independent selection events. Other flower development related genes were also found among the outliers in each population, but most genes hypothesised to be major flowering time regulators in *A. lyrata* [e.g. *FT, FLC, PHYA* and *CO* (Kemi et al., 2013; Leinonen et al., 2013; Mattila et al., 2016; Toivainen et al., 2014)] showed no evidence of differential selection in the present study. However, in the

northern latitudes, the flowering after snowmelt occurs in long days, so it may be primarily governed by other factors than photoperiodism. In fact, the earlier flowering of the alpine populations in paper I could be better explained by faster reproductive development, and therefore selection on genes like *NGA4* (Lee et al., 2015), *GNC* (Richter, Bastakis, & Schwechheimer, 2013) and *VRN1* (Levy, Mesnage, Mylne, Gendall, & Dean, 2002), that were found as outliers in the J3 and T4 populations (Table 1). An interesting exception to this pattern was the blue-light receptor gene *CRY2*, an outlier in the J1 population, which is involved in the long-day mediated flowering response in *A. thaliana* (Andrés & Coupland, 2012). However, the lineage specific selection at this gene could be related to cessation of flowering, for which there may be different critical day lengths even between these neighboring populations (Kemi, 2013).

Although most outlier loci in paper II were population specific, the GO enrichment analysis revealed patterns of adaptive convergence at the level of biological processes. In both lowland populations, many outliers were involved in pathogen defence and vegetative growth, whereas resistance to solar radiation was indicated as an adaptive trait in the alpine populations. The importance of pathogen defence and solar radiation resistance has previously been noted in studies of altitude adaptation in *A. halleri* (Fischer et al., 2013; Kubota et al., 2015) and in *A. thaliana* (Günther, Lampei, Barilar, & Schmid, 2016). However, as the analysis in paper II was done with PBS, the selection could be pinpointed to specific populations, making it possible to separate which traits were involved in lowland and alpine adaptation.

Table 1. Candidate genes from the highest-ranking PBS outlier windows that localized within an annotated gene. Data from paper II.

Population	Gene	Biological process
J1	*SRT2*	Defense response to bacterium
	OTS2	Response to salt stress, vegetative to reproductive development
	GATL1	Cell wall thickening, leaf development
	CRY2	Blue light sensing, photoperiodism, flowering
	EER5	Response to ethylene, root development
J3	*NGA4*	Flower and leaf development
	NIP2	Response to light stimulus
	UBP13	Circadian rhythm, flower development
	GNC	Flower development, cold tolerance, response to light stimulus
	TT5	Response to UV
T1	*CBL5*	Response to salt stress and water deprivation
	NGA4	Flower and leaf development
	XRN2	Regulation of posttranscriptional gene silencing
	SYP122	Response to fungus
	AGL14	Flower development, vegetative to reproductive development
T4	*XRN2*	Regulation of posttranscriptional gene silencing
	RAD23C	Double strand DNA break repair
	ATATG18F	Response to starvation
	LSH2	Response to light stimulus
	VRN1	Vernalization response, flower development

3.5 Germination traits show adaptive potential

Paper III examined the adaptive potential of germination traits using *A. lyrata* populations from two isolated and highly distinctive environments: Karhumäki, Russia and Mayodan, North Carolina. The cold stratification and after-ripening treatments had no discernible effects on germination, suggesting that seed dormancy differences between the groups (Kar and Ma, as well as first and second -generation hybrid reciprocals) are negligible. The postdormancy germination time did, however, show significant difference between the pure-populations, with Kar seeds germinating faster than Ma seeds. This result could indicate adaptive differentiation, as the shorter growing season of the Kar environment may impose stronger selection towards early germination. As indicated by QTL mapping, the differentiation in postdormancy germination was mainly controlled by two large-effect loci. Several candidate genes with known function in postdormancy germination in *A. thaliana* were discovered within the confidence intervals of the

QTL. Three of these genes (*SPY, FHY1* and *MAX2*) exhibited high F_{ST} estimates between the Kar and Ma whole-genome data sets, while including fixed nonsynonymous differences that may underlie the differentiation. As the evolutionary role of germination has mainly been studied in the context of dormancy (Donohue et al., 2005, 2010; Kronholm, Picó, Alonso-Blanco, Goudet, & Meaux, 2012; Postma & Ågren, 2016), the study in paper III provided novel insights into adaptive potential and genetic basis of germination traits beyond the dormancy stage.

3.6 Evidence of reproductive isolation between the *A. lyrata* subspecies

By comparing germination proportions between the hybrids and pure-populations, the study in paper III also revealed evidence of reproductive isolation between the subspecies. The F_1 seeds germinated worse than pure-populations and F_2 seeds germinated worse than F_1 seeds – a pattern consistent with the Bateson-Dobzhansky-Muller model (Coyne & Orr, 2004). Combined with earlier findings of reduced male fertility (Aalto et al., 2013; Leppälä & Savolainen, 2011), these results point towards incipient speciation between the *A. lyrata* subspecies. Furthermore, the reproductive barrier appeared asymmetrically in the first hybrid generation, indicating that uniparentally inherited factors are involved in the BDMIs. Between population crosses in *A. lyrata* have previously exhibited asymmetries in fitness traits (Aalto et al., 2013; Leinonen et al., 2011; Leppälä & Savolainen, 2011), as well as in genetic expression levels (Videvall, Sletvold, Hagenblad, Ågren, & Hansson, 2015), highlighting the importance of non-nuclear inheritance in plant evolution. Sequencing of the F_2 families also allowed for the examination of transmission ratios. Large number of markers deviated from the expected genotype ratios (1:2:1 in biallelic markers), showing further effects of BDMIs. Although the levels of TRD did not correspond with seed inviability proportions, likely reflecting independent incompatibilities, interesting patters were discovered among the hybrid reciprocals. In hybrids with Kar cytoplasm, distorted markers were more or less equally distributed across the genome, whereas in hybrids with Ma cytoplasm, clear clustering of the distorted markers was seen. Chromosomes 3, 4 and 6 showed significant deficit of Kar homozygotes, suggesting incompatibilities between Ma cytoplasm and nuclear loci of Kar origin. These results show that seed inviability can be one of the first reproductive barriers

to form in *Arabidopsis* species and that early acting barriers may manifest only as a selective loss of genotypes.

4 Conclusions

How natural selection leads to adaptation and speciation is a major question in evolutionary biology. Over the years many theoreticians have examined the conditions of adaptive evolution under the effects of selection, migration and drift, but few studies have attempted to tackle these question at the empirical level. In this thesis, I conducted phenotype and sequence based experiments aimed at clarifying the processes that lead to adaptive divergence despite the opposing effects of migration and drift. I also studied the effects of genetic divergence at a large geographical scale, by looking at the role of seed germination in adaptation and reproductive isolation between the *A. lyrata* subspecies.

The demography simulations and multi-year fitness estimates confirmed that *A. lyrata* populations from Jotunheimen, Norway are adapted to their local environments despite ongoing gene flow and small effective population sizes. Patterns of trait differentiation and inferences on phenotypic selection further indicated that flowering traits have contributed to the adaptive divergence.

Selection patterns on the sequence level then showed that the underlying genetic architecture is concordant with theoretical expectations: populations under higher gene flow had fewer predicted selection outliers and they were found in areas of lower recombination rates. Across all populations, patterns at most loci were consistent with conditional neutrality, but a small proportion showed potential for antagonistic pleiotropy. The analysis further suggested directional selection on several flowering related genes, as well as convergent adaptation for alpine and lowland environments at the level of biological processes.

Additionally, the third study revealed that variation in postdormancy germination may be adaptive and that impaired hybrid germination can be one of the first reproductive barriers to arise during divergence. The genetic mapping also indicated that germination time may be under simple genetic control and that genetic incompatibilities between nuclear and organellar genomes can lead to selective loss of genotypes.

Taken together, the three studies in this thesis highlight the value of combining traditional organismal methods with next-generation genomics, by providing novel insights into processes underlying adaptation and speciation.

List of references

1001 Genomes Consortium. (2016). 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*, *166*, 481–491.

Aalto, E., Koelewijn, H. S., & Savolainen, O. (2013). Cytoplasmic male sterility contributes to hybrid incompatibility between subspecies of *Arabidopsis lyrata*. *G3*, *3*, 1727–1740.

Ågren, J., Oakley, C. G., Lundemo, S., & Schemske, D. W. (2017). Adaptive divergence in flowering time among natural populations of *Arabidopsis thaliana*: estimates of selection and QTL mapping. *Evolution*, *71*, 550–564.

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 of the United States of America*, *110*, 21077–21082.

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*, 1112–1122.

Anderson, J. T., Lee, C. R., Rushworth, C. A., Colautti, R. I., & Mitchell-Olds, T. (2013). Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, *22*, 699–708.

Anderson, J. T., Perera, N., Chowdhury, B., & Mitchell-Olds, T. (2015). Microgeographic patterns of genetic divergence and adaptation across environmental gradients in *Boechera stricta* (Brassicaceae). *The American Naturalist*, *186*, S60–S73.

Andrés, F., & Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics*, *13*, 627–639.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*, 81–92.

Antonovics, J., & Bradshaw, A. D. (1970). Evolution in closely adjacent plant populations. VIII. Clinal patterns at a mine boundary. *Heredity*, *23*, 507–524.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*, 25–29.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*, e3376.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

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

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting F_{ST}: the impact of rare variants. *Genome Research*, *23*, 1514–21.

Blanquart, F., Gandon, S., & Nuismer, S. L. (2012). The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of Evolutionary Biology*, *25*, 1351–1363.

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology Letters*, *16*, 1195–1205.

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

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H., & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*, *140*, 783–796.

Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*, 889–890.

Bulmer, M. G. (1972). Multiple niche polymorphism. *The American Naturalist*, *106*, 254–257.

Burton, R. S., Pereira, R. J., & Barreto, F. S. (2013). Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematic*, 44, 281–302.

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, *15*, 538–543.

Clausen, J., Keck, D. D., & Hiesey, W. M. (1940). *Experimental studies on the nature of species. I. Effect of varied environments on western North American plants*. *Carnegie Institution of Washington publication* (Vol. 520). Washington: Carnegie Institution.

Clausen, J., Keck, D. D., & Hiesey, W. M. (1948). *In experimental studies on the nature of species. III. Environmental responses of climatic reces of Achillea. Carnegie Institution of Washington publication* (Vol. 518). Washington: Carnegie Institution.

Colautti, R. I., & Barrett, S. C. H. (2013). Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, *364*, 346–366.

Coyne, J. A., & Orr, A. H. (2004). *Speciation*. Sunderland, MA: Sinauer Associates, Inc.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157.

Darwin, C. (1859). *On the origin of species by means of natural selection: or, the preservation of favored races in the struggle for life*. London: J. Murray.

Dittmar, E. L., Oakley, C. G., Ågren, J., & Schemske, D. W. (2014). Flowering time QTL in natural populations of *Arabidopsis thaliana* and implications for their adaptive value. *Molecular Ecology*, *23*, 4291–4303.

Dobzhansky, T. (1936). Studies in hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*, *21*, 113–135.

Donohue, K., Dorn, L., Griffith, C., Kim, E., Aguilera, A., Polisetty, C. R., & Schmitt, J. (2005). The evolutionary ecology of seed germination of *Arabidopsis thaliana*: Variable natural selection on germination timing. *Evolution*, *59*, 758–770.

Donohue, K., Rubio De Casas, R., Burghardt, L., Kovach, K., & Willis, C. G. (2010). Germination, postgermination adaptation, and species ecological ranges. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 293–319.

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*, 1–10.

Endler, J. A. (1977). *Geographic variation, speciation, and clines. Monographs in population biology* (Vol. 10). Princeton, NJ: Princeton University Press.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*, e1003905.

Felsenstein, J. (1976). The theoretical population genetics of variable selection and migration. *Annual Review of Ecology, Evolution, and Systematics*, *10*, 253–2580.

Ferris, K. G., Barnett, L. L., Blackman, B. K., & Willis, J. H. (2017). The genetic architecture of local adaptation and reproductive isolation in sympatry within the *Mimulus guttatus* species complex. *Molecular Ecology*, *26*, 208–224.

Finch-Savage, W. E., & Leubner-Metzger, G. (2006). Seed dormancy and the control of germination. *New Phytologist*, *171*, 501–523.

Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K. K., … Widmer, A. (2013). Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology*, *22*, 5594–5607.

Fishman, L., & Willis, J. H. (2001). Evidence for Dobzhansky-Muller incompatibilites contributing to the sterility of hybrids between *Mimulus guttatus* and *M. nasutus*. *Evolution*, *55*, 1932–1942.

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*, 86–89.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv Preprint*: https://arxiv.org/pdf/1207.3907.pdf

Geyer, C. J., Wagenius, S., & Shaw, R. G. (2007). Aster modeling for life history analysis. *Biometrika*, *94*, 415–426.

Gonzalo-Turpin, H., & Hazard, L. (2009). Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species *Festuca eskia*. *Journal of Ecology*, *97*, 742–751.

Griswold, C. K. (2006). Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. *Heredity*, *96*, 445–453.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*, e1000695.

Günther, T., Lampei, C., Barilar, I., & Schmid, K. J. (2016). Genomic and phenotypic differentiation of *Arabidopsis thaliana* along altitudinal gradients in the North Italian Alps. *Molecular Ecology*, *25*, 3574–3592.

Hadfield, J. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, *33*, 1–22.

Haig, D., & Westoby, M. (1991). Genomic imprinting in endosperm - its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *333*, 1–13.

Haldane, J. B. S. (1930). A mathematical theory of natural and artificial selection. (Part VI, Isolation.). *Mathematical Proceedings of the Cambridge Philosophical Society*, *26*, 220.

Haley, C. S., & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 315–324.

Hall, M. C., & Willis, J. H. (2006). Divergent selection on flowering time contributes to local adaptation in *Mimulus guttatus* populations. *Evolution*, *60*, 2466–2477.

Haller, B. C., & Messer, P. W. (2017). SLiM 2: flexible, interactive forward genetic simulations. *Molecular Biology and Evolution*, *34*, 230–240.

Hämälä, T., Mattila, T. M., Leinonen, P. H., Kuittinen, H., & Savolainen, O. (2017). Role of seed germination in adaptation and reproductive isolation in *Arabidopsis lyrata*. *Molecular Ecology*, *26*, 3484–3496.

Hedrick, P. W. (2011). *Genetics of populations*. Burlington, MA: Jones & Bartlett Learning.

Hendrick, M. F., Finseth, F. R., Mathiasson, M. E., Palmer, K. A., Broder, E. M., Breigenzer, P., & Fishman, L. (2016). The genetics of extreme microgeographic adaptation: an integrated approach identifies a major gene underlying leaf trichome divergence in Yellowstone *Mimulus guttatus*. *Molecular Ecology*, *25*, 5647–5662.

Hendry, A. P., Day, T., & Taylor, E. B. (2001). Population Mixing and the Adaptive Divergence of Quantitative Traits in Discrete Populations: a Theoretical Framework for Empirical Tests. *Evolution*, *55*, 459–466.

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *American Naturalist*, *173*, 579–588.

Hermisson, J., & Pennings, P. S. (2017). Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, *8*, 700–716.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, *188*, 379–397.

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*, 476–483.

Huber, C. D., Nordborg, M., Hermisson, J., & Hellmann, I. (2014). Keeping it local: Evidence for positive selection in swedish *Arabidopsis thaliana*. *Molecular Biology and Evolution*, *31*, 3026–3039.

Hudson, R. R. (1991). Gene genealogies and the coalescent process. In: D. Futuyama & J. Antonovics (Eds.), *Oxford surveys in evolutionary biology* (Vol 7, pp. 1–44). New York: Oxford University Press.

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, *18*, 337–338.

Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, *132*, 583–589.

Jalas, J., & Suominen, J. (1994). *Atlas florae europaea. Distribution of vascular plants in Europe. 10: Cruciferae (Sisymbrium to Aubrieta)*. *Feddes Repertorium* (Vol. 106). Helsinki: The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo.

Kärkkäinen, K., Løe, G., & Ågren, J. (2004). Population structure in *Arabidopsis lyrata*: Evidence for divergent selection on trichome production. *Evolution*, *58*, 2831–2836.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*, 1225–1241.

Kemi, U. (2013). *Adaptation to growing season lenght in the perennial Arabidopsis lyrata*. Oulu: ACTA Universitatis Ouluensis.

Kemi, U., Niittyvuopio, A., Toivainen, T., Pasanen, A., Quilot-Turion, B., Holm, K., … Kuittinen, H. (2013). Role of vernalization and of duplicated *FLOWERING LOCUS C* in the perennial *Arabidopsis lyrata*. *New Phytologist*, *197*, 323–335.

Kim, S., Lohmueller, K. E., Albrechtsen, A., Li, Y., Korneliussen, T., Tian, G., … Nielsen, R. (2011). Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinformatics*, *12*, 231.

Kingman, J. F. C. (1982). On the genealogy of large populations. *Journal of Applied Probability*, *19*, 27.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*, 1471–2105.

Körner, C. (2007). The use of "altitude" in ecological research. *Trends in Ecology and Evolution*, *22*, 569–574.

Kronholm, I., Picó, F. X., Alonso-Blanco, C., Goudet, J., & Meaux, J. de. (2012). Genetic basis of adaptation in *Arabidopsis thaliana*: local adaptation at the seed dormancy QTL *DOG1*. *Evolution*, *66*, 2287–2302.

Kubota, S., Iwasaki, T., Hanada, K., Nagano, A. J., Fujiyama, A., Toyoda, A., … Morinaga, S. I. (2015). A genome scan for genes underlying microgeographic-scale local adaptation in a wild *Arabidopsis* species. *PLoS Genetics*, *11*, e1005361.

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, *37*, 1210–1226.

Lee, B. H., Kwon, S. H., Lee, S. J., Park, S. K., Song, J. T., Lee, S., … Kim, J. H. (2015). The *Arabidopsis thaliana* NGATHA transcription factors negatively regulate cell proliferation of lateral organs. *Plant Molecular Biology*, *89*, 529–538.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, *3*, e4010.

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*, 709–723.

Leinonen, P. H., Remington, D. L., & Savolainen, O. (2011). Local adaptation, phenotypic differentiation, and hybrid fitness in diverged natural populations of *Arabidopsis lyrata*. *Evolution*, *65*, 90–107.

Leinonen, P. H., Sandring, S., Quilot, B., Clauss, M. J., Thomas, M. O., Ågren, J., & Savolainen, O. (2009). Local adaptation in european populations of *Arabidopsis lyrata* (Brassicaceae). *American Journal of Botany*, *96*, 1129–1137.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology and Evolution*, *17*, 183–189.

Leppälä, J., Bokma, F., & Savolainen, O. (2013). Investigating incipient speciation in *Arabidopsis lyrata* from patterns of transmission ratio distortion. *Genetics*, *194*, 697–708.

Leppälä, J., & Savolainen, O. (2011). Nuclear-cytoplasmic interactions reduce male fertility in hybrids of *Arabidopsis lyrata* subspecies. *Evolution*, *65*, 2959–2972.

Levin, D. A. (2003). The cytoplasmic factor in plant speciation. *Systematic Botany*, *28*, 5–11.

Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R., & Dean, C. (2002). Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control. *Science*, *297*, 243–246.

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

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

Linhart, Y. B., & Grant, M. C. (1996). Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, *27*, 237–277.

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., … Nordborg, M. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics*, *45*, 884–890.

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, *24*, 1031–1046.

Lowry, D. B., Hall, M. C., Salt, D. E., & Willis, J. H. (2009). Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. *New Phytologist*, *183*, 776–788.

Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics*, *12*, e1005887.

Martin, G., & Lenormand, T. (2006). The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution*, *60*, 2413.

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

Mattila, 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*, 2665–2677.

Maynard Smith, J. (1970). Genetic polymorphism in a varied environment. *The American Naturalist*, *104*, 487–490.

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

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, *9*, 356–369.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–303.

McNeilly, T. (1968). Evolution in closely adjacent plant populations III. *Agrostis tenuis* on a small copper mine. *Heredity*, *23*, 99–108.

Merilä, J., & Crnokrak, P. (2001). Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, *14*, 892–903.

Merilä, J., & Hendry, A. P. (2014). Climate change, adaptation, and phenotypic plasticity: The problem and the evidence. *Evolutionary Applications*, *7*, 1–14.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., & Thomas, P. D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research*, *45*, D183–D189.

Montesinos-Navarro, A., Wig, J., Xavier Pico, F., & Tonsor, S. J. (2011). *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude. *New Phytologist*, *189*, 282–294.

Muller, H. J. (1939). Reversibility in evolution considered from the standpoint of genetics. *Biological Reviews of the Cambridge Philosophical Society*, *14*, 261–280.

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

Munguía-Rosas, M. A., Ollerton, J., Parra-Tabla, V., & De-Nova, J. A. (2011). Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured. *Ecology Letters*, *14*, 511–521.

Nei, M. (1987). *Molecular evolutionary genetics*. New York, NY: Columbia university press.

Owens, G. L., & Rieseberg, L. H. (2014). Hybrid incompatibility is acquired faster in annual than in perennial species of sunflower and tarweed. *Evolution*, *68*, 893–900.

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*, e37135.

Postma, F. M., & Ågren, J. (2016). Early life stages contribute strongly to local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 7590–7595.

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

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

R Core Team. (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/*.

Remington, D. L., Figueroa, J., & Rane, M. (2015). Timing of shoot development transitions affects degree of perenniality in *Arabidopsis lyrata* (Brassicaceae). *BMC Plant Biology*, *15*, 1.

Richardson, J. L., Urban, M. C., Bolnick, D. I., & Skelly, D. K. (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology and Evolution*, *29*, 165–176.

Richter, R., Bastakis, E., & Schwechheimer, C. (2013). Cross-repressive interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the control of greening, cold tolerance, and flowering time in Arabidopsis. *Plant Physiology*, *162*, 1992–2004.

Riihimäki, M., & Savolainen, O. (2004). Environmental and genetic effects on flowering differences between northern and southern populations of *Arabidopsis lyrata* (Brassicaceae). *American Journal of Botany*, *91*, 1036–1045.

Robertson, A. (1960). A theory of limits in artificial selection. *Proceedings of the Royal Society London B*, *153*, 234–239.

Sambatti, J. B. M., & Rice, K. J. (2006). Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). *Evolution*, *60*, 696–710.

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology*, *26*, 4378–4390.

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

Savolainen, O., & Kuittinen, H. (2011). *Arabidopsis lyrata Genetics. Genetics and Genomics of the Brassicaceae.* New York, USA: Springer Science+Business Meadia LLC 2011.

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

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, *38*, 595–619.

Schemske, D. W. (1984). Population structure and local selection in *Impatiens pallida* (Balsaminaceae), a selfing annual. *Evolution*, *38*, 817–832.

Shaw, R. G., & Geyer, C. J. (2010). Inferring fitness landscapes. *Evolution*, *64*, 2510–2520.

Shaw, R. G., Geyer, C. J., Wagenius, S., Hangelbroek, H. H., & Etterson, J. R. (2008). Unifying life-history analyses for inference of fitness and population growth. *The American Naturalist*, *172*, E35-47.

Stanton-Geddes, J., Paape, T., Epstein, B., Briskine, R., Yoder, J., Mudge, J., … Tiffin, P. (2013). Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS ONE*, *8*, e65688.

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*, 4712–4717.

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 9440–9445.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*, 437–60.

Tiffin, P., Olson, M. S., & Moyle, L. C. (2001). Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society B: Biological Sciences*, *268*, 861–867.

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology and Evolution*, *29*, 673–680.

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

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*, 459–470.

Turelli, M., & Moyle, L. C. (2007). Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics*, *176*, 1059–1088.

Turesson, G. (1922). The genotypical response of the plant species to the habitat. *Hereditas*, *3*, 211–350.

Van Ooijen, J. W. (2006). *JoinMap 4: Software for the calculation of genetic linkage maps in experimental populations*. Wageningen, Netherlands: Kyazma BV.

Vergeer, P., & Kunin, W. E. (2013). Adaptation at range margins: Common garden trials and the performance of *Arabidopsis lyrata* across its northwestern European range. *New Phytologist*, *197*, 989–1001.

Videvall, E., Sletvold, N., Hagenblad, J., Ågren, J., & Hansson, B. (2015). Strong maternal effects on gene expression in *Arabidopsis lyrata* hybrids. *Molecular Biology and Evolution*, *33*, 984–994.

Wadgymar, S. M., Lowry, D. B., Gould, B. A., Byron, C. N., Mactavish, R. M., & Anderson, J. T. (2017). Identifying targets and agents of selection: innovative methods to evaluate the processes that contribute to local adaptation. *Methods in Ecology and Evolution*, *8*, 738–749.

Wakeley, J. (2009). *Coalescent theory: an introduction*. Greenwood Village, CO: Roberts & Co. Publishers.

Wellenreuther, M., & Hansson, B. (2016). Detecting polygenic evolution: problems, pitfalls, and promises. *Trends in Genetics*, *32*, 155–164.

West-Eberhard, M. (2003). *Developmental plasticity and evolution*. Oxford: Oxford University Press.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–159.

Yeaman, S. (2015). Local adaptation by alleles of small effect. *The American Naturalist*, *186*, S74–S89.

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

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

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., … Wang, J. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, *329*, 75–78.

Original publications

I Hämälä T, Mattila TM, & Savolainen O (2018) Local adaptation and ecological differentiation under selection, migration and drift in *Arabidopis lyrata*. Manuscript.

II Hämälä T & Savolainen O (2018) Local adaptation under gene flow: Recombination, conditional neutrality and genetic trade-offs shape genomic patterns in *Arabidopsis lyrata*. Manuscript.

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

Original publications are not included in the electronic version of the dissertation.

700. Aaramaa, Sanja (2017) Developing a requirements architecting method for the requirement screening process in the Very Large-Scale Requirements Engineering Context

701. Mattila, Tiina (2017) Post-glacial colonization, demographic history, and selection in *Arabidopsis lyrata* : genome-wide and candidate gene based approach

702. Lwakatare, Lucy Ellen (2017) DevOps adoption and implementation in software development practice : concept, practices, benefits and challenges

703. Marttila, Maare (2017) Ecological and social dimensions of restoration success in boreal river systems

704. Hartikainen, Heidi (2017) Malice in Wonderland : children, online safety and the wonderful world of Web 2.0

705. Ventä-Olkkonen, Leena (2017) The characteristics and development of urban computing practices : utilizing practice toolkit approach to study public display network

706. Mononen, Jukka (2017) Korkeasti koulutettujen vammaisten integroituminen ICT-alalle heidän itsensä kokemana : ”Älä anna muille etumatkaa!”

707. Tolonen, Katri (2018) Taxonomic and functional organization of macroinvertebrate communities in subarctic streams

708. Turunen, Jarno (2018) Responses of biodiversity and ecosystem functions to land use disturbances and restoration in boreal stream ecosystems

709. Huusko, Riina (2018) Downstream migration of salmon smolts in regulated rivers : factors affecting survival and behaviour

710. Huusko, Karoliina (2018) Dynamics of root-associated fungal communities in relation to disturbance in boreal and subarctic forests

711. Lehosmaa, Kaisa (2018) Anthropogenic impacts and restoration of boreal spring ecosystems

712. Sarremejane, Romain (2018) Community assembly mechanisms in river networks : exploring the effect of connectivity and disturbances on the assembly of stream communities

713. Oduor, Michael (2018) Persuasive software design patterns and user perceptions of behaviour change support systems

714. Tolvanen, Jere (2018) Informed habitat choice in the heterogeneous world: ecological implications and evolutionary potential

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/

UNIVERSITY OF OULU P.O. Box 8000 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

UNIVERSITY of OULU
OULUN YLIOPISTO

ISBN 978-952-62-1906-6 (Paperback)
ISBN 978-952-62-1907-3 (PDF)
ISSN 0355-3191 (Print)
ISSN 1796-220X (Online)