OULU 2013
A 617

ACTA
UNIVERSITATIS OULUENSIS

Esa Aalto

GENETIC ANALYSIS OF DEMOGRAPHY AND SELECTION IN LYRATE ROCKCRESS (*ARABIDOPSIS LYRATA*) POPULATIONS

*UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF SCIENCE,
DEPARTMENT OF BIOLOGY*

ACTA UNIVERSITATIS OULUENSIS
A Scientiae Rerum Naturalium 617

ESA AALTO

GENETIC ANALYSIS OF DEMOGRAPHY AND SELECTION IN LYRATE ROCKCRESS (*ARABIDOPSIS LYRATA*) POPULATIONS

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 Kuusamonsali (Auditorium YB210), Linnanmaa, on 16 December 2013, at 12 noon

UNIVERSITY OF OULU, OULU 2013

Copyright © 2013
Acta Univ. Oul. A 617, 2013

Supervised by
Professor Outi Savolainen
Docent Helmi Kuittinen

Reviewed by
Professor Anneli Hoikkala
Associate Professor Leonie Moyle

ISBN 978-952-62-0321-8 (Paperback)
ISBN 978-952-62-0322-5 (PDF)

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

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2013

Aalto, Esa, Genetic analysis of demography and selection in Lyrate rockcress (*Arabidopsis lyrata*) populations.
University of Oulu Graduate School; University of Oulu, Faculty of Science, Department of Biology
Acta Univ. Oul. A 617, 2013
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Demographic history and selection affect patterns of genetic diversity in nature. Timing of growth, reproduction and dormancy are important traits targeted by natural selection, because they are crucial for survival of plants growing in boreal and temperate climates, as reproduction must occur when conditions are favorable and in outcrossing plants it must be synchronized to assure pollination. In addition to adaptation to local environments, evolutionarily diverged populations may contain genomic incompatibilities that result in sterile hybrids in crosses between populations. In several plant families within population crosses can also lead to male sterile progeny, because of conflict between maternally and bi-parentally inherited genomes.

In this thesis I used DNA-sequence data to estimate the demographic history of nine *Arabidopsis lyrata* populations and present genetic variation in some key flowering time genes, evolution of which natural selection has shaped. By crossing experiments I explored genetics of reproductive fitness in hybrids between divergent populations.

I found that local climatic conditions have resulted in directional selection in addition to the demographic effects of bottlenecks during colonization events. Flowering time genes have reduced diversity compared to reference loci, which indicates selective sweeps. Selection on nucleotide variation in flowering time genes was found in Scandinavian and Icelandic populations that can be explained by selective sweeps at flowering genes when these populations colonized northern habitats after the last glacial maximum.

Cryptic cytoplasmic male sterility was found in a Norwegian population, for which North Carolinian population did not have fertility restorers. It was confirmed that there is only one fertility restorer locus, the genomic location of which was mapped to a 600 kb interval at the top of chromosome two.

Keywords: A. lyrata, cytoplasmic male sterility, demographic history, flowering time, local adaptation, speciation

Aalto, Esa, Geneettinen analyysi demografiasta ja luonnonvalinnasta idänpitkäpalon (*Arabidopsis lyrata*) populaatioissa.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Luonnontieteellinen tiedekunta, Biologian laitos
Acta Univ. Oul. A 617, 2013
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Populaatioiden levittäytymishistoria ja luonnonvalinta vaikuttavat geneettiseen monimuotoisuuteen ja sen vaihteluun genomin eri osissa. Kukkimisen ja kasvun päättämisen ajoitus ovat tärkeitä luonnonvalinnan kohteena olevia ominaisuuksia, sillä ne ovat välttämättömiä kasvien selviytymiselle lauhkeissa ja pohjoisissa ilmastoissa. Paikallisiin olosuhteisiin sopeutumisen lisäksi populaatioiden erilaistuminen voi johtaa genomisiin yhteensopimattomuuksiin, joiden vuoksi populaatioiden väliset risteymät ovat lisääntymiskyvyttömiä. Monilla kasvisuvuilla myös populaation sisäiset risteytykset voivat johtaa koirassteriileihin jälkeläisiin johtuen konfliktista vain äidin puolelta ja molemmilta vanhemmilta periytyvien genomien välillä.

Tässä väitöskirjassa selvitän DNA-sekvenssimuunteluun perustuen yhdeksän idänpitkäpalkopopulaation demografista eli levittäymis- ja populaatiorakennehistoriaa sekä luonnonvalinnan osuutta kukkimisaikaan vaikuttavien geenien evoluutiossa. Risteytyskokeiden avulla tutkin erilaistuneiden populaatioiden risteymäjälkeläisiä selvittääkseni niiden lisääntymiskelpoisuutta ja siihen vaikuttavia geenejä.

Geneettinen muuntelu kukkimisgeeneissä oli vähäisempää kuin vertailugeeneissä, joka on merkki kukkimisgeeneihin kohdistuneesta suuntaavasta valinnasta. Kukkimisgeeneihin kohdistuvaa valintaa löytyi eniten skandinaavisista ja islantilaisesta populaatiosta, mikä on selitettävissä niihin kohdistuneella suuntavalla valinnalla aikana, jolloin kasvit levittäytyivät jääkauden jälkeen pohjoisiin elinympäristöihin.

Norjalaisesta populaatiosta löytyi piilevä sytoplasminen koirassteriliteetti, jolle Pohjois-Carolinan populaatiolla ei ollut hedelmällisyyden palauttavia geenejä. Tutkimus vahvisti, että hedelmällisyyden palauttaa yksi geeni, joka sijaitsee 600 000 emäsparin kokoisella alueella kromosomin kaksi alkupäässä.

Asiasanat: demografia, Idänpitkäpalko, kukkimisaika, lajiutuminen, paikallinen sopeutuminen, sytoplasminen koirassteriliteetti

Acknowledgements

This work would not have been possible without significant effort given by various people. Here I try to remember most of them, but I admit that some names may be missing.

Heidi Aisala, Soile Alatalo, Matti Heino, Sanna Hietala, Johanna Huusko, Kirsi Järvi, Jatta Jääskeläinen, Ulla Kemi, Jukka Kiiskilä, Johanna Kinnunen, Tiina Mattila, Celine Mercier, Meeri Otsukka, Teija Paakkola, Anu Pasanen, Elsa Petit, Saana Remula, Marja-Maija Ristiluoma, Hanna-Riikka Ruhanen, Matti Salmela, Tuomas Toivainen, Aleksia Vaattovaara, Timo Vesimäki, Elisa Vilhunen, Antti Virtanen and Sannakajsa Welmala helped in the greenhouse and in the lab. This enthusiastic team made the whole work possible.

Special thanks belong to Antti Virtanen, with whom I was fortunate to work together by supervising his master thesis and getting lot of progress for this dissertation. The paper I on demography of *A. lyrata* was much driven by Tanja Pyhäjärvi as the first author of the paper. With her help I got well into the statistical methods and R as well as scientific writing, all important skills for the next steps. Hans Peter Koelewijn's expertise on cytoplasmic male sterility was extremely essential for publication of the paper III on that topic. Problems in sequencing the flowering time genes were finally solved together with my colleagues Anne Niittyvuopio and Tuomas Toivainen after initial work by Anne and Susanna Piltonen. At the final part Tiina Mattila and Jaakko Tyrmi participated in analyzing the flowering time and reference genes. It has been really enjoyable to work together with these people as well as other people at the department for which I am very grateful.

Furthermore I am grateful to Professors Anneli Hoikkala and Leonie Moyle for reviewing this thesis. I also thank Jaakko Lumme for many helpful discussions related more or less to the topic and last, but not least the supervisors of this thesis, who were Professor Outi Savolainen and Docent Helmi Kuittinen.

The study was financially supported by the Biosciences and Environment Research Council of the Academy of Finland (projects 132611 and 128114 to Outi Savolainen), the national Population Genetics Doctoral Program (to Esa Aalto) and the Department of Biology of University of Oulu.

Abbreviations

θ	population mutation rate
π	nucleotide diversity
μ	mutation rate
bp	base pair
CMS	cytoplasmic male sterility
MYA	million years ago
N_e	effective population size
NMS	Nuclear male sterility
ORF	open reading frame
PCR	polymerase chain reaction
PPR	pentatricopeptide repeat
TRD	transmission ratio distortion
QTL	quantitative trait locus
Rf	restorer of fertility

List of original papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Pyhäjärvi T, Aalto EA & Savolainen O (2012) Timescales of divergence and speciation among natural populations and subspecies of *Arabidopsis lyrata* (Brassicaceae). Am J Bot 99: 1314–1322.

II Aalto EA, Toivainen T, Niittyvuopio A, Piltonen S, Kuittinen, H & Savolainen O (2013) Selection on flowering time genes in *Arabidopsis lyrata*. (Manuscript).

III Aalto EA, Koelewijn HP & Savolainen O (2013) Cytoplasmic Male Sterility contributes to hybrid incompatibility between subspecies of *Arabidopsis lyrata* G3: Genes, Genomes, Genetics 3: 1727–1740.

Contents

Abstract
Tiivistelmä
Acknowledgements 7
Abbreviations 9
List of original papers 11
Contents 13
1 Introduction 15
1.1 Demography and adaptation ... 15
1.2 Cytoplasmic male sterility (CMS) ... 17
1.3 Evolution of reproductive isolation ... 18
1.3.1 Nuclear genic male sterility ... 18
1.3.2 Transmission ratio distortion (TRD) ... 19
1.3.3 Speciation ... 19
1.4 Study species ... 20
1.4.1 Study populations and questions ... 20
2 Material and methods 23
2.1 Origin of study plants ... 23
2.2 Molecular methods ... 23
2.3 DNA sequence analysis ... 24
2.4 Relationships of the populations ... 24
2.5 Molecular diversity ... 25
2.6 Reproductive isolation ... 25
2.6.1 Crosses ... 25
2.6.2 Growth conditions and phenotypic measurements ... 26
2.6.3 Genotyping of F2 ... 27
3 Results and discussion 29
3.1 Demography of *A. lyrata* ... 29
3.2 Selection on flowering time ... 31
3.3 Cytoplasmic male sterility ... 36
3.4 Transmission ratio distortion (TRD) ... 39
3.5 Reduced pollen fertility in F1 and F2 in Spiterstulen Mayodan cross ... 40
3.6 Male sterility has no effect on female fertility ... 41
3.7 Incipient speciation? ... 43

4 Conclusions **47**
References **49**
Original articles **59**

1 Introduction

Historical events such as the spreading and melting of continental ice sheets have had large effects on the present day populations of the species in high altitudes and over all North Europe. Due to these events many species consist of subpopulations that are only weakly connected with each other or may even be completely isolated. Demographic events during range expansions and reductions may have genome-wide effects on genetic diversity of diverged populations (Excoffier *et al.* 2009, Hewitt 2000, Taberlet *et al.* 1998). An important goal of population genetics and ecological genomics is to understand the relative importance of different evolutionary processes for shaping genetic variation underlying adaptation to different biotic and abiotic conditions in natural populations. It is also important to understand other genomic consequences caused by isolating barriers between populations.

1.1 Demography and adaptation

Knowledge of history of individual populations and the relationships between them is important for studies on adaptation to different habitats, and consequences of random genetic processes caused by divergence. Estimates of divergence times among populations are necessary to account for demography when detecting the footprints of natural selection and for other population genetic analyses. Neutral divergence time estimates are also needed for evaluating how rapidly reproductive isolation between populations emerges (Coyne & Orr 1989, Moyle & Nakazato 2010).

Timing of growth, reproduction and dormancy are important for plants growing in boreal and temperate climate. Differences in these characteristics among populations are likely due to adaptation to local environments (Samis *et al.* 2008, Sandring *et al.* 2007), but they may also evolve as a consequence of random divergence (Masel 2011). Local adaptation on climate has been extensively studied both in *Arabidopsis thaliana* and in *Populus* (e.g. Andres & Coupland 2012, Flowers *et al.* 2009, Fournier-Level *et al.* 2011, Hancock *et al.* 2011, Keller *et al.* 2011, Keller *et al.* 2012). Reciprocal transplant experiments have shown that *A. lyrata* populations are locally adapted and differ in important components of life history (Leinonen *et al.* 2009, Leinonen *et al.* 2011, Riihimäki & Savolainen 2004).

Flowering time is a central part of plant life history, because reproduction must occur when the abiotic and other conditions are favourable. Further, flowering of outcrossing plants must be synchronized so that pollination is ensured. Responses to photoperiod and other environmental cues and genes or alleles behind these differences may differ between populations. The molecular genetics and physiology of flowering of the plant molecular biology model organism *Arabidopsis thaliana* is being intensively studied. The main pathways that influence the transition from vegetative to inflorescence meristems are known, and many of the genes have been identified (reviewed e.g. in Andres & Coupland 2012, Lagercrantz 2009, Michaels 2009).

Distinguishing between demographic effects and selection is often difficult, because similar patterns of variation can result from both (e.g. Depaulis *et al.* 2003, Li *et al.* 2012, Tajima 1989). However, natural selection generally acts more on some genes than others, but demographic processes affect the entire genome similarly. Thus, it is possible to distinguish between these effects using a large number of loci to get overall distribution against which individual genes can be viewed. A common method is to identify candidate loci, based on their function, which are supposed to be under different selection pressures in different populations. Then these genes are compared to randomly chosen loci where such selection is not expected. One of the challenges in this kind of analysis is the large variance of mutations and stochastic events that generate large differences among loci just by chance (Wright & Gaut 2005). Better than direct comparison of genes is to use reference loci to build up a demographic model, against which candidate genes can be tested (see e.g. Ingvarsson 2008, Keller *et al.* 2011, Li *et al.* 2012). Selected candidate genes can differ from genomic averages in several ways depending on the type of selection. For instance, geographically variable selection can maintain different alleles in separate populations possibly leading to selective sweeps in one or several related populations (e.g. Moeller & Tiffin 2008), and long term adaptive protein evolution leads to excess of amino acid changing nucleotide replacements between species (e.g. Flowers *et al.* 2009, Wright & Andolfatto 2008).

Flowering time genes often evolve under directional and purifying selection. For instance, within species genetic diversity in flowering time related genes is reduced compared to genomic averages as observed in *Populus* (Hall *et al*. 2011, Keller *et al*. 2011, Keller *et al*. 2012). In the present study I expect to find signs of selection on flowering time in *A. lyrata* populations with different expansion histories and growing sites and adaptations to local photoperiods and other

environmental variables in populations growing in very different habitats. Populations that have shared history and similarities in their environment are expected also to share adaptations, while more distant populations may have different adaptations to similar environments. Another prediction is that directional selection in the loci controlling flowering time has reduced within population variation relative to genome wide neutral averages (Muller *et al.* 2008, Ross-Ibarra *et al.* 2008, Wright *et al.* 2003).

1.2 Cytoplasmic male sterility (CMS)

Isolation of outcrossing hermaphrodite plant populations may lead to a cryptic form of cytoplasmic male sterility (CMS), which is a common cause for asymmetry in reproductive isolation (see e.g. review by Rieseberg & Blackman 2010). In most cases CMS is caused by rearrangements of mitochondrial open reading frames (ORFs) that lead to novel chimeric genes. These genes may interact with the mitochondrial respiratory system and cause male sterility (Chase 2007). Male fertility can be restored by fertility restorer (*Rf*) genes located in the nucleus. Both the CMS and *Rf* genes have been identified in a number of crop plants (Hanson & Bentolila 2004, Yamamoto *et al.* 2008) and in *Mimulus* (Barr & Fishman 2010, Case & Willis 2008). The known fertility restorers are mostly members of a class of pentatricopeptide repeat containing (PPR) genes, the common function of which is to modify transcripts in mitochondria (Bentolila *et al.* 2002). Population genetics of CMS has been most extensively studied in natural gynodioecious populations in *Plantago* (DeHaan *et al*. 1997, Koelewijn & vanDamme 1995a, Koelewijn & vanDamme 1995b, Koelewijn 2003, van Damme *et al*. 2004, vanDamme & vanDelden 1982).

An important feature of CMS is that it does not reduce transmission of the maternally inherited mitochondrial genome (Case & Willis 2008, Chase 2006, Hanson & Bentolila 2004) that allows the male-sterility gene to invade a hermaphrodite population. This can lead to a genomic conflict (Cosmides & Tooby 1981). Selection may even favour the invasion, if female plants have increased female fertility. It may be due to reduced energetic investment in male floral organs of females or because females avoid selfing and thus inbreeding depression (Charlesworth & Ganders 1979, Dufay & Billard 2012, Hanson & Bentolila 2004, Shykoff *et al.* 2003). However, even drift alone can lead to fixation of CMS, if females have the same ovule production as hermaphrodites (Charlesworth & Charlesworth 1978, Gouyon *et al.* 1991). Fixation of both

cytoplasmic male sterility (CMS) genes and the corresponding fertility restoring alleles is likely to be very fast, which means that populations can be carriers of different cryptic CMS and matched restorers despite being hermaphroditic.

1.3 Evolution of reproductive isolation

Population differentiation can lead to reproductive isolation between the diverged populations due to several reasons. In plants, hybrid sterility, especially hybrid male sterility, is the most common type of post-mating reproductive isolation (Lowry *et al.* 2008), but other barriers are also widely known (Bomblies & Weigel 2007, Bomblies & Weigel 2010, Lexer & Widmer 2008). The post mating isolating mechanisms can evolve in allopatric populations as so called Bateson-Dobzhansky-Muller (BDM) incompatibilities (Bateson 1909, Dobzhansky 1936, Muller 1939). These are genic incompatibilities causing reduction in hybrid fitness, which can depend on the type of hybrid or the direction of the cross. Asymmetrical incompatibilities have been known for a long time, as Darwin already noticed that hybrids raised from reciprocal crosses generally differ in their degree of sterility (Darwin 1859). More recently Tiffin *et al.* (2001) observed significant asymmetries in reproductive isolation of all three Angiosperm genera they studied. Reproductive isolation can accumulate in several ways and ultimately contribute to the evolution of new species i.e. speciation.

1.3.1 Nuclear genic male sterility

Nuclear genic male sterility (NMS) refers to BDM incompatibilities, which are caused by negative interaction of two or more nuclear loci. NMS factors have not been documented in detail in wild plants. Maybe the best known case is in *Mimulus*, where a dominant *M. guttatus* allele at the *hms1* locus acts in combination with recessive *M. nasutus* alleles at *hms2* causing nearly complete male sterility in hybrids (Sweigart *et al.* 2006). In crop plants there are more examples, as three hybrid male sterility alleles in two loci known in rice (Long *et al.* 2008) and many QTL reducing hybrid seed and pollen fertility by 36–90% found in tomato *Lycopersicon hirsutum* x *L. esculeatum* cross (Moyle & Graham 2005).

1.3.2 Transmission ratio distortion (TRD)

A further signal of possible genomic incompatibilities between allopatric populations is transmission ratio distortion, often found in crosses between species or populations with high divergence. Genetic studies have shown that it is often closely associated with reduced male fertility in QTL mapping populations, due to non-random elimination of specific allele combinations (e.g. Fishman & Willis 2001, Hall & Willis 2005, Harushima *et al.* 2001, Koide *et al.* 2008, Moyle & Graham 2006). TRD is better studied in flies. It is known that in *Drosophila pseudoobscura* a single gene *Overdrive* causes both male sterility and segregation distortion in F1 hybrids between two subspecies (Phadnis & Orr 2009, Phadnis 2011).

1.3.3 Speciation

When populations are isolated from each other for a long time, they accumulate genetic differences that can eventually be sufficient to produce evolutionary independence, i.e., speciation. Based on the biological species concept (Mayr 1942), speciation can be defined as formation of reproductive barriers between populations (Coyne & Orr 2004). Chromosomal rearrangements can create these barriers immediately, and they are known to be common between plant species (Lowry & Willis 2010, Rieseberg 2001), but their relative importance in plant speciation is unclear (Lowry *et al.* 2008). Speciation without chromosomal mutations is often referred to genic speciation (Lexer & Widmer 2008). Many of the known speciation genes do not directly affect reproductive traits (see e.g. review by Presgraves 2010). However, there are exceptions as for instance reciprocal loss of duplicated genes encoding mitochondrial ribosomal protein essential for the later stage of pollen development in rice (Yamagata *et al.* 2010). Speciation genetics and especially post zygotic BDM interactions causing hybrid sterility or lethality are intensively studied in *Drosophila* (e.g. Barbash *et al.* 2000, Barbash *et al.* 2003, Brideau *et al.* 2006, Presgraves *et al.* 2003, Sun *et al.* 2004). When populations diverge more from each other, the number of BDM incompatibilities between them should increase like a snowball with the square of interspecific substitutions or even faster (Orr 1995, Orr & Turelli 2001).

1.4 Study species

The organism used in the present study is the perennial outcrossing self-incompatible plant Lyrate rockcress (*Arabidopsis lyrata*) (OKane and AlShehbaz 1997; Savolainen and Kuittinen 2011). This species is a good model to study evolution of isolated lineages at different time scales, because it consists of separate populations with wide range of divergence, and tools to work with the species in the field, greenhouse and laboratory are well developed. *A. lyrata* has two subspecies representing distinct evolutionary lineages: *ssp. lyrata* in North America and *ssp. petraea* in Eurasia. The phylogenetics of the species has been well studied using cpDNA and limited nuclear gene data (e.g. Koch & Matschinger 2007, Schmickl *et al.* 2010). The two subspecies are highly diverged with microsatellite F_{ST} estimates >0.5 (Muller *et al.* 2008) and F_{ST} for sequence data ranging from 0.3 to 0.6 depending on within population diversity (Ross-Ibarra *et al.* 2008). Northern Europe has been hypothesized to have been clononized from Central Europe after the Pleistocene (Muller *et al.* 2008, Ross-Ibarra *et al.* 2008), but new studies suggest that the present Scandinavian populations may have their origin in a more Northwestern refugia, perhaps on the British isles (Ansell *et al.* 2010, Koch & Matschinger 2007, Schmickl *et al.* 2010). The Russian population in Karhumäki has more eastern origin (Schmickl *et al.* 2010) and thus differs remarkably from the other European populations.

1.4.1 Study populations and questions

Nine populations were used. Seven of them represent subspecies *petraea*: Plech (Germany) and Bohemia (Chech Republic) from central Europe, Reykjavik (Iceland), Stubbsand (Sweden) and Spiterstulen and Lom (Norway) from Northern Europe, Karhumäki (Russia) from Northeastern Europe. The two populations representing the subspecies *lyrata* are Mayodan, North Carolina (USA) and Ithaca N.Y. (USA) from North America. The aim of the study is to characterise the relationships of the study populations and to examine genetic differences between the populations caused by natural selection and stochastic processes during the time of divergence. First estimates for divergence times and patterns among nine *A. lyrata* populations are provided. Using a coalescence-based Bayesian method (Becquet & Przeworski 2007), new divergence time estimates are presented for 10 pairs of populations representing different levels of the population hierarchy. In addition, a traditional distance-based method is used

to give a comprehensive picture of relationships among populations. Then patterns of nucleotide diversity in 19 flowering time genes are described and compared to reference loci chosen without considering their function. The hypothesis is that adaptation is based on responses to different photoperiods and other environmental signals. Due to directional selection, reduced within population variation is predicted (Braverman *et al.* 1995, Hudson 1989, Kaplan *et al.* 1989) in the loci controlling flowering time relative to the genome wide neutral averages (Muller *et al.* 2008, Ross-Ibarra *et al.* 2008, Wright *et al.* 2003).

Despite the high divergence, crosses between *A. lyrata* subspecies produce viable progeny without any apparent difficulties. However, a considerable number of transmission ratio distorted loci have been observed between some of the distant populations (Kuittinen *et al.* 2004, Leppälä & Savolainen 2011). Further, in a cross between Spiterstulen and Mayodan, hybrid male fitness was reduced due to both supposed nuclear genic incompatibilities and a putative CMS (Leppälä & Savolainen 2011). In that study, QTL-mapping detected one locus restoring hybrid fertility and another hidden locus for fertility restoration was suspected. To characterize in more depth the genetics of the hybrid sterility between the subspecies of *A. lyrata*, I present here large crossing experiments between the Norwegian and North Carolinian populations in which the CMS system is confirmed and way of its inheritance is shown. To study the inheritance, examination of TRD across the genome and its impact on the genetic analyses is important, because it affects to predicted genotype ratios. In addition, polymorphisms in CMS and restorers are searched. Finally, the fertility restorer is mapped with a dense set of markers near to the previously identified putative restorer QTL, and its importance in speciation is discussed.

2 Material and methods

Methods are described here only briefly and generally. Original papers should be consulted for more details.

2.1 Origin of study plants

Seeds representing the nine wild *A. lyrata* populations were collected in Plech, Germany 49° 39′ N, 11° 29′ E; Bohemia, Czech Republic 50° 03′ N, 14° 06′ E (by Mark McNair); near Reykjavik, Iceland 64° 09′ N, 21° 58′ W (by Mikkel Schierup); Karhumäki, Russia 62° 55′ N, 34° 25′ E; Stubbsand, Sweden 63° 12′ N, 18° 57′ E; Lom, Norway 61° 50′ N, 8° 24′ E, Spiterstulen, Norway 61° 38′ N, 8° 24′ E, Ithaca, New York, USA. 42° 22′ N, 76° 21′ W and Mayodan, North Carolina, USA 36° 25′ N, 79° 58′ W (by David Remington).

2.2 Molecular methods

Total genomic DNA was extracted from leaves using the FastDNA® Kit and the FastPrep® Instrument (Qbiogene, Inc, CA) or Mag-Bind E-Z 96 (Omega, Bio-tek, USA). Overall, 38 loci were sequenced to study demography and selection on flowering time. For examining demography and for reference to the flowering time genes we sequenced a set of 19 loci from Ross-Ibarra *et al.* 2008 chosen regardless of their putative function. Only partial sequences of the reference genes were obtained. The other 19 sequenced loci were located at flowering time pathways: (*CIRCADIAN CLOC ASSOCIATED 1* (*CCA1*), *CONSTANS* (*CO*), *CRYPTOCHROME 1* (*CRY1*), *CRYPTOCHROME 2* (*CRY2*), *FCA*, *FKC*, *FLOWERING LOCUS T* (*FT*), *FRIGIDA* (*FRI*), *FVE*, *GIGANTEA* (*GI*), *LUMINIDEPENDENCE* (*LD*), *LATE ELONGATED HYPOCOTYL* (*LHY*), *PHYTOCHROME A* (*PHYA*), *PHYTOCHROME B* (*PHYB*), *SUPPRESSION OF CONSTANS OVEREXPRESSION 1* (*SOC1*), *TIMING OF CAB1* (*TOC1*), *VERNALIZATION 1* (*VRN1*), *VERNALIZATION 2* (*VRN2*) and *ZEITLUPE* (*ZTL*)). Sequencing of the reference and flowering time loci is explained in more detail in the original paper I and II of this dissertation, respectively.

2.3 DNA sequence analysis

Sequence data were aligned with the Sequencher 4.0.5 (Gene Codes, Ann Arbor, MI) or with CodonCodeAligner 3.5.7 (CodonCode Corporation, Centerville, MA). MEGA 4 (Tamura *et al.* 2007) was used for editing the sequences. All putative polymorphic sites were finally validated by visual inspection of chromatograms. Sequences were converted to haplotypes and the phase of heterozygous sites was inferred using the program PHASE v.2.1 (Stephens *et al.* 2001, Stephens & Scheet 2005) and also partly verified by sequencing progenies of the individuals.

2.4 Relationships of the populations

To get an overview of the genetic relationships among the nine populations, pairwise F_{ST} values were calculated with DnaSP 5.10 (Rozas *et al.* 2003) using all sites. The average F_{ST} values among the nine populations were calculated to build an NJ tree using program MEGA 3.0 (Kumar *et al.* 2004). Divergence times among populations were estimated using the program MIMAR (Becquet & Przeworski 2007) that uses a Bayesian method based on MCMC to estimate speciation parameters from resequencing data of diverging population pairs. For the MIMAR analyses, only four fold degenerate sites were considered. For each locus and population pair, the number of sampled individuals in each population (n_1 and n_2), the number of unique derived polymorphisms in each population (S_1 and S_2), the number of derived shared polymorphic sites (S_S), and the number of fixed alleles in either population (S_f) were used as summary statistics to inform MIMAR. The *A. thaliana* reference genome sequence (*Arabidopsis* genome initiative 2000) was used to define the derived or ancestral status of each allele. The mutation rate 7×10^{-9} per site per generation (Ossowski *et al.* 2010) was used to estimate population sizes and divergence times. The method and estimation process are explained in more detail in the original paper I. Ten population pairs, Karhumäki–Ithaca, Karhumäki–Mayodan, Karhumäki–Plech, Karhumäki–Stubbsand, Mayodan–Spiterstulen, Mayodan–Plech, Bohemia–Plech, Ithaca–Plech, Plech–Spiterstulen, and Bohemia–Ithaca, were chosen for MIMAR analysis. These pairs were chosen based on geographical coverage of the analysis and because they had reasonable divergence to get enough information for MIMAR to calculate reliable posteriors.

2.5 Molecular diversity

To compare genetic polymorphism between populations and loci, the number of polymorphic sites (S), nucleotide diversity (π) (Tajima 1983), and Watterson's estimate of the scaled mutation rate (θ_W) (Watterson 1975), within *A .lyrata* populations were calculated separately for silent, synonymous and nonsynonymous sites at each locus using MANVA (Heidel *et al.* 2010). In addition, to study selection on protein evolution, ratios of synonymous to nonsynonymous diversity and divergence were calculated as π_a/π_s and K_a/K_s respectively. For each locus departure from the standard neutral model was studied with Tajima's D (Tajima 1989). To inspect whether some of the loci had unexpectedly high or low amounts of nucleotide variation, the Maximum-likelihood-ratio HKA (Hudson *et al.* 1987) test was performed using the the MLHKA software (Wright & Charlesworth 2004). For selected genes the HKA test gives selection parameters (k-values), which indicate the strength of selection and whether it is positive or purifying. To detect loci that may have been undergoing positive selection the McDonald-Kreitman Poisson random field (MKPRF; Bustamante *et al.* 2002) test was used. The MKPRF test is based on a comparison of synonymous and nonsynonymous (replacement) variation within and between species.

2.6 Reproductive isolation

Reproductive isolation between subspecies was studied by crossing experiments between Norwegian population from Spiterstulen (representing *ssp. petraea* and North American population from Mayodan, North Carolina, U.S.A. representing *ssp. lyrata.* Unrelated plants were used in all crossing designs to avoid expression of self-incompatibility and inbreeding depression. Crossing experiments are explained in detail in the original paper III. Here, I only briefly describe the main crosses.

2.6.1 Crosses

Altogether 565 F1 plants belonging to 8 families (4 of both reciprocals) were grown. They were used to compare them with F2s and to produce 8 families of F2s for examining polymorphism of CMS. A large family of F2 plants (745 SpMaF2 with Sp cytoplasm and 1352 MaSpF2 with Ma cytoplasm) was grown to

get precise estimates of male fertility and seed production per silique of these hybrids and to create a genetic linkage map. The parents of this cross were grown from field-collected seeds. One Sp pollen donor was crossed to a Ma pollen recipient to produce F1 hybrids with Ma cytoplasm and another pair of Ma and Sp plants was crossed reciprocally to produce F1 hybrids with both cytoplasms. Two of the F1 hybrids were then crossed reciprocally to yield two sets of F2 individuals with different cytoplasms. To explore whether asymmetry between reciprocal F2s is really due to cytoplasmic factors, and to examine the genetics, we generated four pseudobackcross populations: [MaSp]Sp (n = 188), [MaSp]Ma (259), [SpMa]Sp (82) and [SpMa]Ma (116). These are pseudobackcrosses because they are structured like a backcross but the same plants or closely genetically related individuals were not used twice in the same line.

To get a more detailed view of the genetics of asymmetric male sterility, backcrossing was continued past BC1. Since the CMS appeared only in the [SpMa]Ma (and SpMaF2) plants, only those second generation backcross plants (BC2) that had cytoplasm from Spiterstulen and 7/8 of the nuclear genome from Mayodan were grown. In total there were 1389 of these (SpMa)MaMa pseudobackcross plants. To further examine the inheritance of the CMS and to find possible polymorphism in fertility restorers within Ma population, a third generation pseudobackcrosses were produced with multiple Ma pollen donors. Total 745 of these (SpMa)MaMaMa plants (hereafter BC3) having cytoplasm from Spiterstulen and 15/16 of the nuclear genome from Mayodan were grown. To examine possible within population variation in CMS several F2 and backcross families were generated. Altogether 562 F2 plants from seven maternal lineages and 287 (SpMa)Ma pseudobackcross plants from eight maternal lineages were grown.

2.6.2 Growth conditions and phenotypic measurements

Seeds were germinated, planted in pots in fully randomized order, grown in the greenhouse for 4 weeks, vernalized in cold room at +4 °C for 8 weeks and finally grown to flowering in the greenhouse in long days (20h) under natural midsummer light and temperatures. To assess male fertility, quantity and quality of pollen of the plants was measured. The plants that produced less than 10 pollen grains in the counted subsample were considered male sterile. To obtain seeds two flowers from each plant were pollinated with Sp and another two with Ma pollen. Seeds were classified as good quality (plump) or low quality (shriveled or green).

As overall fitness measures rosette sizes of the plants were measured and the number of flowering shoots counted.

2.6.3 Genotyping of F2

Leaves were collected for genotyping at the end of the experiment and DNA extracted in plate format. The four parental, two F1, and 1880 F2 plants of the main crossing family were genotyped for altogether 96 SNP markers in 75 genes chosen to cover the genome with at 10-20 cM distance between them. The marker density was set higher at the top of chromosome two to fine map *Rf*.

Genetic linkage maps were generated using CRI-MAP (Green *et al.* 1990) version 2.50 and QTL analyses was performed with R/qtl (Broman *et al.* 2003) separately for the both reciprocals. Transmission of each genotype of each marker was inspected separately and Bonferroni corrected significances calculated for comparing the observed ratios to equal transmission of the genotypes.

3 Results and discussion

3.1 Demography of *A. lyrata*

First the relatioships between the populations needed to be investigated. To jointly estimate divergence times of multiple populations is computationally demanding and needs a massive amount of genetic data. Nevertheless, I tried to do that using Approximate Bayesian Computation but the 19 relatively short reference genes were not enough to get reliable posterior estimates simultaneously for multiple parameters for multiple populations. The way to clarify the relationships among populations in North America and Europe was to use a combination of basic NJ trees and model-based divergence time estimates for two populations at once.

As expected, the two Norwegian populations, Lom and Spiterstulen that are separated by only about 50 km were also the closest related according to NJ trees. The third Scandinavian population Stubbsand grouped with these populations as did the fourth northern population from Iceland. These four North European populations formed a clear well supported cluster. However, the relationships among the Scandinavian group and the other populations were less consistent. The Russian Karhumäki population did not group with other northern populations but formed its own separate branch as did the two Central European populations Plech and Bohemia. The two North American populations that belong to *ssp. lyrata* separated clearly from the group of European *ssp. petraea* populations supporting well the separation of the two subspecies and consistent with previous results based on microsatellites, cpDNA, and nuclear DNA sequence data (Muller *et al.* 2008, Schmid *et al.* 2005, Wright *et al.* 2003).

MIMAR analysis gave further information on the divergence times for pairs of populations taking into account the different levels of diversity in each population. The most ancient divergence time estimates among 10 population pairs analyzed were 160 000 and 140 000 generations found in contrasts of North American Mayodan population and European populations Karhumäki and Spiterstulen, respectively. Divergence times between the other North American population, Ithaca, and European populations Karhumäki and Bohemia also were high (130 000 and 100 000 generations, respectively). The generation time of *A. lyrata* is at least 2 years and often longer. Based on the above number of generations and assuming a generation time of 2 years, the divergence time between European and North American *A. lyrata* is 200 000–320 000 years. In

contrast, divergence estimate between Ithaca and Plech was only 69 000 generations, but the posterior distributions of all estimates were wide and thus well overlapping. However, lower divergence time estimates between Ithaca and European *A. l. petraea* populations may be an indication of some more recent western genetic contribution to the Northeastern part of *A. l. lyrata* distribution range or that North American *A. lyrata* populations have otherwise variable population histories. Distribution of variation in cpDNA (Schmickl *et al.* 2010) and microsatellite data (Willi & Määttänen 2010) also show division of North American populations into two potential ancestors. Altogether, the results are in accordance with those by Schmickl *et al.* (2010), who suggested that the North American *A. l. lyrata* populations originated from colonization via Beringia.

Contrasts of Russian Karhumäki population with the Central European Plech and North European Stubbsand both gave a posterior probability peak at 73 000 generations, which is the highest distance among *A. l. petraea* populations. The cpDNA haplotype in Karhumäki is also distinct from the others and it is not found in western and northern European populations, but is present farther east instead (Schmickl *et al.* 2010). These observations support the conclusion that the population in Karhumäki is distinct from the other European populations and has a more eastern origin.

Estimated divergence time between the Central European Plech the and Scandinavian Spiterstulen population was 41 000 generations and among the two Central European populations Bohemia and Plech it was 39 000 generations. Contrasts between the more closely related North European populations failed to converge, because there were not enough (or at all) fixed differences between them, used by MIMAR to estimate the posteriors. The most closely related Scandinavian populations separated maybe as late as at the end of the last glacial period. The exact relationships among central European and Scandinavian populations remain unclear, but new studies suggest that the present Scandinavian populations have their origin in more Northwestern refugia, perhaps British isles, where (Ansell *et al.* 2010, Kivimäki *et al.* 2007, Schmickl *et al.* 2010).

In addition to using very simple two population model, another weakness of this study on the demography of *A. lyrata* is the lack of some populations possibly central for getting a complete picture of the species wide genetic structure. These populations include British Isles, Eastern Austria and other Eastern Central European populations, East Asia and Central and Northwestern North America. Still despite of these deficiencies, this study gives a nice starting point for more detailed analysis of worldwide history of *A. lyrata*.

3.2 Selection on flowering time

The rate of nucleotide evolution after the divergence of *A. lyrata* and *A. thaliana* can be measured as synonymous nucleotide differences between the species. At the same time the rate of long term evolution in the flowering time genes compared to the reference loci can be analyzed. If the mutation rate substantially differs between the two sets of loci, it should be visible in synonymous divergences (K_{syn}) as a significant difference. K_{syn} for reference loci was 0.15 (SD 0.07) and for flowering loci 0.11 (SD 0.03). K_{syn} in flowering genes is on average 28% lower, than in the reference loci, which is only a marginally significant difference.

Demographic effects on genetic diversity can be distinguished by examining patterns of polymorphism in the reference genes. The level of silent nucleotide diversity at the reference loci was compatible with other studies of *A. lyrata* (Balañá-Alcaide *et al.* 2006, Ross-Ibarra *et al.* 2008, Wright *et al.* 2003). Within the subpecies *petraea*, the median estimate of synonymous genetic diversity (θ_s) was higher in the south than in the north. Genetic diversity was reduced to ca. 50% in Scandinavian (Spiterstulen, Stubbsand, and Lom) populations compared to the Central European Plech, the most variable population in this study, confirming previous results (Muller *et al.* 2008, Ross-Ibarra *et al.* 2008, Wright *et al.* 2003). The loss of diversity is likely due to smaller long term population sizes of *A. lyrata* in North Europe explained by post glacial colonization from south to the north. The lowest θ_s among *ssp. petraea* was in Karhumäki. In North American populations, θ_s was even smaller than in Scandinavia reflecting strong bottlenecks during their colonization history.

In contrast to *ssp. petraea*, within *ssp. lyrata* median θ_s was smaller in the southern Mayodan population than in Ithaca in the north. The best explanation for this is the direction of dispersion of *A. lyrata* in North America from the north to the south as supposed by (Schmickl *et al.* 2010). Another reason for low genetic variability in in Mayodan can be that *A. lyrata* is more short-lived there (Leinonen *et al.* 2011), which can increase the effect of drift (Kimura & Ohta 1969).

Colonization history seems to have a significant effect on genetic variation, which has to be taken into account while considering the flowering time genes. In all populations synonymous nucleotide diversity θ_s was much lower in flowering time genes than in the reference loci. Average θ_s in flowering time genes was only 28% of that in reference genes, but divergence from *A. thaliana* in flowering genes was as much as 72% of what was observed in reference loci. The difference

in θ_s values was most striking in Iceland, where there was only 11% of the variation present in flowering time genes compared to the reference genes. Yet in general, variation in flowering time genes was not reduced more in the North of Europe compared to the Central European populations or in Mayodan compared to Ithaca.

The lower K_{syn} in flowering time genes compared to reference loci can be caused by differences in mutation rate, codon usage or base composition of these genes. However, the observed difference in nucleotide diversity between flowering and reference loci is much higher and cannot be explained by these effects only. Nor do differences in demography in different parts of the genome explain the observation, because both sets of loci have been widely sampled throughout the genome. Thus, natural selection as a reason for the reduction in variation of flowering time genes need to be considered. Reduced variation at flowering time related genes compared to genomic averages has also been found other species as *A. thaliana* (Flowers *et al.* 2009), European aspen (*Populus tremula*) (Hall *et al.* 2011) and Balsam poplar (*Populus balsamifera*) (Keller *et al.* 2012), but the difference in these cases has not been as clear as observed here in *A. lyrata*.

The average Tajima's D in the reference loci in all populations was clearly positive, ranging from 0.45 in Bohemia to 1.17 in Stubbsand that indicates more intermediate frequency variants than expected under the standard neutral equilibrium (SNE). This is best explained by decreased effective population size in all populations, after which a new equilibrium have not been reached yet. Substantial variation was evident among loci, Tajima's D values ranging from -2.18 to 3.42. Variation between the flowering and the reference genes would be an indication of natural selection. However, the average Tajima's D values observed did not differ between reference and flowering loci, but there were large differences between individual loci within both sets instead.

Loci with unexpectedly high or low amounts of nucleotide diversity relative to divergence can be examined by the HKA test. I did the test separately for each population, because the main interest was in relatively short term evolution after the split of the *A. lyrata* lineages studied. k values (selection parameters) given by the test for the loci are below or above one if there is less or more than expected diversity, respectively. In Scandinavian populations and in Iceland the model with selection on flowering time genes had significantly higher likelihood compared to neutral model ($P < 0.001$). Both test runs gave similar results. Most of the flowering time genes including eleven photoperiodic pathway genes *CO, CRY2,*

FKF1, GI, TOC1, LHY, PHYA, PHYB and *ZTL*, two of the three vernalization pathway genes *FVE* and *FRI* and one of the three autonomous pathway genes *LD*, showed reduced level of polymorphism compared to expectation. Five genes, *SOC1, FT, FCA, VRN1* and *VRN2*, had k values close to one (0.81-1.29) while polymorphism was much increased in two genes, *CCA* and *CRY1*.

The results were very similar in all the Scandinavian and the Icelandic population reflecting that adaptation to northern conditions dates after the last glacial period at the time when these populations migrated to the north perhaps as a larger and more constant population. However, another North European population, the Russian Karhumäki, has a distinct origin further east and did not show such a difference between the two sets of loci. The lack of significant difference in this population may also be because of low overall genetic variation observed in Karhumäki, which may reduce power of the MLHKA test to detect possible signs of selection. Variation in diversity between the loci studied was also very similar within the two Central European populations, Plech and Bohemia. Yet, it differed much from the pattern observed in the North European populations and no significant difference between flowering and reference loci could be found by the HKA test in these Central European populations.

Synonymous diversity (π_a) was substantially higher than nonsynonymous (π_s) in both sets of genes and in all populations, suggesting purifying selection at most codons. The π_a/π_s ratios for different sets of loci were compatible (individual population means from 0.12 – 0.57 for reference loci and 0.19 – 0.60 for flowering loci). Thus purifying selection seems to act equally on both sets of genes. None of the π_a/π_s ratios for individual loci were statistically different from one ($p > 0.05$), perhaps due to the low number of segregating sites. In 78 cases the π_a/π_s ratio could not even be estimated because of lack of synonymous polymorphisms. Average values for the ratio of synonymous to nonsynonymous divergence from *A. thaliana*, K_a/K_s, ranged from 0.11 to 0.16 for both reference and flowering time loci in all populations, implicating predominantly purifying selection acting on these genes. Thus the data suggest predominant action of purifying selection in most of the loci studied as evidenced by K_a/K_s and π_a/π_s values below one, but it does not exclude the possibility of positive selection in a minority of the coding positions. The population structure can reduce the rate of fixation of mutation by positive selection in *Arabidopsis* (Foxe *et al.* 2008, Gossmann *et al.* 2010). Other recent studies have shown that the genomes of both *A. thaliana* and *A. lyrata* show an excess of nonsynonymous polymorphisms (Bustamante *et al.* 2002, Foxe & Wright 2009) and little evidence of adaptive

substitutions compared for example to *Populus* or *Drosophila* (Hough et al. 2013, Ingvarsson 2010, Keller et al. 2012, Wright & Andolfatto 2008).

In the MKPRF test, twelve of the reference loci showed significant excess of within population nonsynonymous polymorphisms as did also five of the flowering time loci: *CRY1, PHYA, VRN2, CCA1* and *TOC1*. This indicates either balancing or local selection or segregating slightly deleterious mutations. Again, more deviation from neutral evolution was expected in the flowering time genes than in the reference genes, but long term adaptive evolution was found only in one locus, *CO*, by an excess of fixations of nonsynonymous mutations. Because the MKPRF test detects long term processes (Hohenlohe *et al.* 2010), the post glacial events may not be visible in this test. Instead, the importance of *CO* and circadian clock in general in flowering time regulation in *A. lyrata* emphasizes the difference between perennial *A. lyrata* and annual *A. thaliana*, in which temperature sensing is more important in adaptation and timing of flowering (Le Corre *et al.* 2002, Riihimäki *et al.* 2005).

Flowering time genes in general showed signs of directional selection, but some of them can be pointed out as the most interesting for future studies. One gene, *CO,* was the most extreme in both the MLHKA and MKPRF tests, thus showing most significant reduction on synonymous polymorphisms and most significant long term adaptive protein evolution. Transcription of *CO* is controlled by circadian clock and CO is a key protein in sensing photoperiod (Putterill *et al.* 1995). In long days the CO protein promotes expression of *FT* that in turn promotes floral induction. Signatures of adaptive evolution on circadian clock genes have been detected also in many other species as Balsam poplar *Populus balsamifera* (Keller *et al.* 2011, Lagercrantz & Axelsson 2000). Another interesting gene showing far reduced polymorphism in Scandinavia and in Iceland was *PHYA*, which is a key photoreceptor that stabilizes the CO protein (Flowers *et al.* 2009, Putterill *et al.* 1995). *PHYA* has also been shown to be under directional selection in *Populus* (Hall *et al.* 2011).

Purifying selection dominates in the *A. lyrata* genome (Foxe *et al.* 2008), but signs of other selective forces shaping the evolution of flowering time loci also exist. Flowering time is regulated by dozens of genes of which 19 were analysed in this study. Thus, adaptive divergence of populations to flower at the right time in various different environments, is not necessarily due to a single gene difference, but could be due to a combination of several adaptations instead (Leinonen *et al.* 2013, Quilot-Turion *et al.* 2013).

Three reference loci and all flowering time genes had introns allowing a preliminary comparison of the intronic sites between the two sets of loci. Average silent nucleotide diversity θ_{sil} (synonymous and intronic sites) was 0.0052 in reference loci and 0.0032 in flowering time genes. Thus, nucleotide diversity in flowering time gene introns seems to not have decreased as much as that at synonymous sites of exons. Interestingly, the highest average θ_{sil} 0.0148 was observed in a flowering time gene CRY1 and it was high (0.010–0.024) in all populations except Karhumäki. Based on the MLHKA and MKPRF tests, synonymous and nonsynonomous diversities were also increased in *CRY1*, indicating either diversifying or relaxed selection on this gene. *CRY1* is one of the photoreceptors, which receive different wavelength of lights, at the top of the photoperiodic pathway, and are important in the regulation of *CO* (Putterill *et al.* 1995).

Based on these results the photoperiodic pathway and circadian clock appear to have been important when *A. lyrata* diverged from *A. thaliana,* and when the present *A. lyrata* populations adapted to their current environments. The causative nucleotide differences between populations are not necessarily in exons of flowering time genes in which very few nonsynonymous replacements are generally present. Thus, while examining the candidate genes as *CO, PHYA* and *CRY1* further, introns and regulatory areas should also be considered potentially interesting.

Initially, 22 flowering time genes were chosen to be sequenced. Soon it was found that there are two copies of FLC in *A. lyrata* genome, which would make sequencing of this gene extremely challenging and it was omitted. After the first flowering time genes were sequenced, the difference in polymorphism between them and the reference genes was even more striking than presented here. The reason for it was that the first genes finished were those with least polymorphism and thus the easiest to be completely sequenced. Respectively, the genes with high polymorphism and long introns as *VRN1* and *VRN2* were the last finished and adding them to the data set changed the conclusions, because fewer differences between reference and flowering time genes became statistically significant. In addition to *FLC*, the two genes missing from the final data set are *ELF3* and *FPA*, both of which seemed to have lot of variation between and within the populations including indels, repeat number variation and changes in UTR areas, which made primer design and sequencing too difficult to be finished during this project.

Moreover, the reference genes were not chosen completely at random, but instead they were initially selected because they had long exons easy to sequence (Ross-Ibarra *et al.* 2008). For example, this choice leads to overrepresentation of pentatricopeptide repeat containing genes, which are known to be rapidly evolving (Fujii & Small 2011, Fujii *et al.* 2011) and thus more diverse than average genes in genomes. Thus, by unintentionally selecting reference genes with presumably increased variation and sequencing those flowering time genes with least variation, we may detect differences between the two sets of genes without these differences reflecting any real evolutionary forces deviating from genomic averages that would shape the evolution of flowering time genes. However, in the final data set, also most of those flowering time genes that were difficult to sequence are included, and thus I believe that the results presented here are not strongly affected by the bias caused by the reference and flowering genes used. To completely avoid such a bias in future studies, one possibility is to start and finish the candidate genes and samples one by one to confirm that those, which are easy to sequence are not ready to be published before others, and by choosing set of reference loci without any structural preferences.

A similar kind of bias is also possible when new sequencing techniques are used, because by those methods some parts of genome become often sequenced by higher coverage than others and also coverage between samples can vary. In addition to sequencing reaction, difficulties in aligning highly variable but relatively short sequences can induce bias. This may lead to underrepresentation of genomic areas of high variability, though whole genome is evaluated in the comparison.

3.3 Cytoplasmic male sterility

To find if the hybrids between diverged populations show reduction in fertility, I conducted large crossing experiments between Norwegian and North Carolinian populations representing subspecies *petraea* and *lyrata*, respectively. The two parental populations were highly male fertile; the median of three-year average pollen production of the Sp and Ma plants was 75.9 and 109.6 pollen grains per sample, respectively. Already in F1 hybrids pollen production was reduced with a difference observed between reciprocals, F1 plants with Ma cytoplasm producing less pollen. In the F2 generation there was even more difference in pollen numbers between plants with Ma and Sp cytoplasms, Ma cytoplasm leading to 40% higher pollen production. Somewhat less than a quarter, 20.7% of SpMaF2

plants were completely male sterile (MS), a phenotype which was not present in reciprocal MaSpF2 nor in F1.

Assuming a CMS model with one dominant major fertility restorer, I expected that MS plants appear also in backcrosses, but only in [SpMa]Ma backcross hybrids. This indeed was the case: 37.2% of 113 [SpMa]Ma plants that flowered were MS, while other types of backcross progeny remain fully hermaphrodite (H). Thus male sterility segregated only in crosses where the progeny had cytoplasm from Sp and nuclear Ma alleles from both parents. To locate the fertility restorer gene (*Rf)*, I conducted QTL mapping with high marker density on Chr2. The mapping revealed only one QTL for *Rf* located at 2.00–4.35 cM interval corresponding to 750–1345 kb from the top of Chr2. The difference in pollen production between the genotypes was striking, this QTL explaining 27% of the pollen number variation in SpMaF2, and no other cytoplasm-dependent QTLs for pollen number was detected. The genotyping revealed that the Ma alleles were able to restore fertility in 4 out of 10 cases in F2 generation. Based on phenotype ratios of backcrosses, the same 40% fertility restoration was present there as well.

Most of the Ma alleles present in the crosses were able to restore fertility in about 40% of the cases. However, as several Ma plants were used, some variation was also found in the trait. To generate the BC3 generation, I used 11 different Ma pollen donors. Two fathers of these used in crosses with H BC2 mothers produced fewer MS progeny than the other 9 that followed the expected ratio. The result suggests that the deviating fathers that are full siblings either carry restorer alleles that restore fertility more often than 40% of the cases or they may have additional restorer genes.

To find possible polymorphisms in maternal CMS factor(s), I grew 4 half-sib families of both reciprocals of F2 and backcrosses. There were no differences between families with Ma cytoplasm, but the family Sp4 produced more pollen than backcross progeny of the other families in (SpMa)Ma backcrosses (Kruskal-Wallis chi-squared = 18.8463, df = 3, p = 0.0003). The F2 progeny of the SP4 family also seemed to produce more pollen, but the difference was not significant. There were almost no male sterile plants in the family Sp4 that can be explained only, if there is no CMS the Sp4 cytoplasm.

The most likely explanation for the results is that there is no CMS present in the Mayodan population, thus the plants having Ma cytoplasm are always hermaphrodites. In Spiterstulen there is only one CMS – one *Rf* system present, which is carried by most, but not all of the plants in the population. Plants of the

Ma population lack complete restorers for the CMS of Sp, but partial restorers with varying efficiency exist, which can be remnants of an old CMS – restorer system. One CMS – one *Rf* system is what would be expected when crossing diverged populations of hermaphrodite species. In gynodioecious populations, balancing selection can maintain several polymorphic CMS-restorer systems (Dufay *et al.* 2007). Instead in natural populations of hermaphrodites, where both CMS and the corresponding restorers become fixed, the selective advantage of the CMS gene disappears and only possible cost of restoration (see e.g. McCauley & Bailey 2009) remains. This favors mutations causing degeneration of the CMS and later inactivation of its restorers. So, most hermaphrodite populations should lack functional CMS and restorers (Figure 1). If this is the case, cryptic CMS should come up only in some crosses and often only in one reciprocal direction, which has been observed in one of the best studied genus *Mimulus* (Martin & Willis 2010, Wise *et al*. 2011).

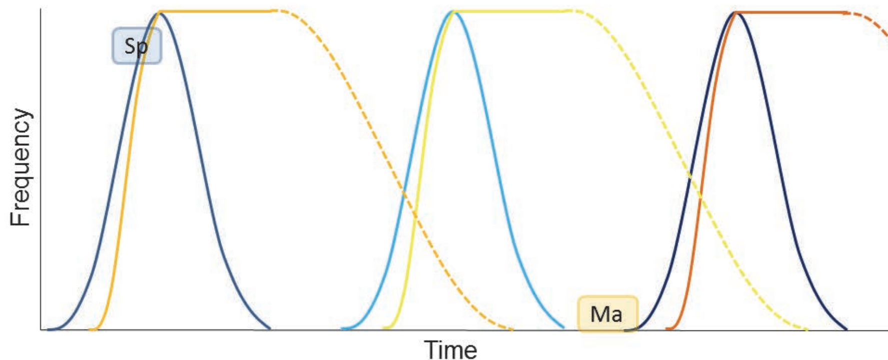

Fig. 1. A schematic of epidemic model of CMS and restorers in functionally hermaphroditic plants. Blue lines represent frequency of CMS types and yellow-orange lines corresponding restorers. Fertile cytoplasm is not shown in the figure. When restorer is fixed, CMS loses its (possible) transmission advantage and may degenerate and disappear by drift or selection against cost of restoration. When frequency of CMS becomes zero, there is no more selection for functional restorer, which allows its degeneration (dashed lines). Hypothetical positions of current states of Sp and Ma populations are presented in the figure.

There are many possibilities to study this CMS issue further. The CMS gene in the Sp cytoplasm could be identified by sequencing mtDNA of the male fertile Sp4 cytoplasm and one or few male sterile cytoplasms. Studying genes expressed

in mitochondria is a well-known way to find candidates for causative DNA rearrangements for CMS (Hanson & Bentolila 2004, Schnable & Wise 1998). Sequencing mtDNA of apparently male fertile Ma cytoplasm for comparison could also be worthwhile. Identification of the fertility restorer gene could also be possible. Within the current 95% Bayes credible interval of the QTL there are only 100 annotated genes. Sequencing this area of selected F2 and 1st, 2nd and 3rd generation backcross hybrids would reveal the exact locus that differs between male sterile and hermaphrodite hybrids. After the causative mutations of CMS and *Rf* are identified, the existence and distribution of them in *A. lyrata* populations could be easily examined. Producing reciprocal F2 hybrid progenies of diverged populations systematically throughout the *A. lyrata* distribution area would reveal possible other cryptic CMS–*Rf* systems present in this species today and help us to understand better the dynamics of CMS in hermaphrodite species.

3.4 Transmission ratio distortion (TRD)

Significant TRD was observed in all eight chromosomes of *A. lyrata* in the F2 Sp x Ma hybrid mapping population. Ma alleles were favored in chromosomes two and four, while Sp alleles were overrepresented in the other six chromosomes. In chromosomes five and six, only single markers were distorted, while other areas of TRD covered from four markers to almost entire chromosomes. In addition, there was a single marker in chromosome four with a surplus of heterozygotes and in chromosome six a marker with deficiency of heterozygotes. The genetic analysis was strongly influenced by the TRD and the genotyping was a necessary component for taking this into account both for examining the number of loci and the issues of partial restoration. Surprisingly, the allele unable to restore CMS was favored as a part of large block in chromosome two with more Ma alleles than expected. The highest TRD maps about 2 cM downstream of the *Rf* locus indicating that the loci are different. The reason for TRD in unknown, but nothing suggests it has something to do with CMS. As recent results by Leppälä *et al.* (2013) indicate, the severe TRD observed here in intersubspecies cross is most likely not present in crosses where more closely related populations are involved. The TRD analysis used here was very cursory. A more detailed view of the TRD could be obtained with the automated computer script designed for the purpose, as was used e.g. by Leppälä *et al.* (2013) (but the same script was incompatible with a data set as large as presented here).

3.5 Reduced pollen fertility in F1 and F2 in Spiterstulen Mayodan cross

The two parental populations Sp and Ma were highly male fertile also, when the fertility was measured as proportion of stained pollen grains in a sample. In 2007 (F1, F2 and backcrosses), 2008 (2nd generation backcrosses and several F2 and 1st generation backcross families) and 2009 (3rd generation backcrosses) experiments Sp plants produced on average 85.8, 91.4 and 93.2% fertile pollen grains (dyed blue in aniline blue staining), respectively. Corresponding pollen fertilities for Ma plants were even higher, 92.6, 93.0 and 92.5%, respectively. When exploring pollen fertility of the hybrids, male sterile plants producing no pollen were excluded. The fertility of the F1 was reduced compared to the parental populations as only 84.2% and 82.3% of pollen grains were fertile in SpMa and MaSp reciprocal F1 hybrids, respectively. In the F2 generation average pollen fertility did not differ from the corresponding F1 generation (proportion of fertile pollen 81.2% and 80.3% in MaSpF2 and SpMaF2, respectively). The mean pollen fertilities in the four backcrosses [MaSp]Ma, [MaSp]Sp, [SpMa]Ma and [SpMa]Sp were 0.87, 0.89, 0.77 and 0.90, respectively. Kruskal-Wallis test detected highly significant ($p < 0.001$) differences between the four groups due to lower pollen fertility in [SpMa]Ma compared to all the others, but no differences in other pairwise comparisons. When pollen fertilities of the four backcrosses were compared to Ma, the reduction in the proportion of fertile pollen was always significant but compared to Sp only the [SpMa]Ma plants differed significantly. The observed reduced pollen viability of F1 and F2 hybrids was symmetric, but in backcrosses pollen viability, in addition to CMS, was dependent on the cytoplasm of the plant. Thus it is evident that there is a cytoplasmic component in this trait, but based on phenotypic data only we cannot exclude possible interaction between nuclear loci.

The mapping of pollen fertility in F2s revealed completely different QTL in reciprocal SpMaF2 and MaSpF2 plants supporting the hypothesis of nuclear cytoplasmic interaction. For MaSpF2 we found altogether 4 QTL, 2 with P value below 0.001 (at chromosomes 2 and 4), and 2 with p value below 0.05 (at chromosomes 7 and 8). The QTL at chromosome 2, explaining 2.3% of total pollen fertility variation, had the highest LOD score of all pollen fertility QTL. The highest probability for the location was at 3.61 cM ≈ 1180 kb that is very close to the QTL for *Rf*. The low pollen quality caused by this locus was due to an interaction between the Ma cytoplasm and one of the two Sp alleles. Average

pollen quality was 0.79 when this Sp allele was present and 0.84 and when it was absent. For SpMaF2 only one QTL with $p < 0.01$ located at chromosome 8 was found.

Counting of fertile and sterile pollen grains was one of the most time consuming parts of this work. The method used for staining pollen grains was incompatible with automated detection of different pollen grains by flow cytometry. However, by choosing the staining carefully and optimizing the flow cytometer analysis could make quantification of male fertility faster and more precise, thus reducing stochastic error and enhancing detection of QTL.

3.6 Male sterility has no effect on female fertility

In the year 2007 experiment seed production of the plants was examined. There were no differences between seed sets produced by Sp pollen (average 7.6. good and 12.5 total seeds/silique) and Ma pollen (7.5. good and 12.2 total seeds/silique). A Kruskall-Wallis test revealed differences between the different kinds of progenies (parents, F1, F2, BC; $p < 0.001$), but the only significant ($p < 0.05$) pairwise differences were the higher number of good seeds produced by Ma control plants than MaSpF1 and (SpMa)Ma hybrids. The latter also produced fewer good seeds than (MaSp)Ma plants.

If MS (female) plants produce more seeds than hermaphrodites, selection favors the spread of CMS in the population. Instead, if seed production of females is not increased, CMS must spread by genetic drift, and there is no potential for stable CMS polymorphism (Charlesworth 1981, Dufay *et al.* 2007). In the current cross, correlation between seed (female) and pollen (male) fertilities was generally weak or absent. There was a positive correlation between seed and pollen production in SpMaF2 population, but not in the other set of hybrids expressing MS, namely (SpMa)Ma plants. No QTL was found for seed fertility, measured as number of good seeds in silique or as proportion of good seeds in silique. Although there was no significant reduction in female fertility in any of the hybrids, the point estimates of seed production were lower in hybrids compared to the parental populations. Measuring female fitness is more prone to environmental effects leading to large variances of the estimates, which reduces the power of tests. In the present study four or more different pollen donors per experiment were used to exclude donor specific effects on seed number.

The number of good seeds per silique is not the best way to measure female fertility, as it ignores possible differences in flower numbers between plants.

Indeed, per flower seed production of female plants is often not increased in gynodioecious populations, while total seed production is (Shykoff *et al.* 2003). Due to the huge number of flowers in our plants, I was not able to count all of them, but I examined the number of flowering shoots and the plant leaf rosette size instead. There was no correlation between pollen production and rosette size or flower shoot number in SpMaF2, but in the (SpMa)Ma backcross plants, pollen production was positive correlated with plant rosette size ($r = 0.313$, $p < 0.001$) and negatively correlated with flower shoot number ($r = 0.221$, $p = 0.019$). However, it is difficult to infer whether plants with more flowering shoots have better female fitness or not. This is because plants having bigger rosette size may be better prepared for the next flowering season and may produce more flowers then. Measuring female fitness precisely would require estimating total seed production over multiple years.

By inspecting our data more carefully, I found a logical explanation for the higher flowering shoot number of the plants producing less pollen: The *Rf* locus is located at the top of chromosome two and plants having two Ma alleles at that location lack functional restorers for the CMS. In the middle of the same chromosome there is a strong QTL for flower shoot number, where Ma homozygotes produce twice as many shoots as the other genotypes (unpublished QTL mapping results of the current cross and personal communication with David Remington, 2012). Thus, the cause of the correlation is not the CMS, but a strong QTL for flower shoot number linked to the *Rf* locus. To conclude, we did not observe any signs of increased seed production of females compared to hermaphrodites.

Reduction of male fertility of the hybrids has evolved prior to any significant female effects. Faster evolution of low hybrid male fitness is common in many animals such as *Drosophila* flies, mainly due to the effects of sex chromosomes (Presgraves 2008), but other reasons also exist. Instead in many plant species both pollen and seed fertility of hybrids are decreased, as in hybrids of *Mimulus guttatus* and *M. nasutus* (Fishman & Willis 2001) or in tomato *Lycopersicon hirsututm x L. esculentum* hybrids, where seed production is reduced even more than pollen viability (Moyle & Graham 2005).

Measuring seed production was the most difficult part of the research for hybrid fertility. First, a lot of homogenous high quality pollen is needed. In the 2007 experiment cloned tissue cultured plants were used as pollen donors to ensure enough pollen with homogenous genotype. Later, ploidy levels of ramets (in tissue culture) of the four of the pollen donors and 25 of their progeny were

analyzed by flow cytometer by Plant Cytometry Services (AG Shijndel, The Netherlands). The results indicated possible tetraploidization of two of the four pollen donor cultures (one Sp and one Ma clone). Luckily, all the 25 progeny tested were diploid, thus showing that there were no effects of polyploidy in our results on pollen fertility. To avoid the issue of polyploidy in years 2008–2009, tissue cultured pollen donors were not used anymore, but population plants grown from seeds instead, which led to introduction of more genetic variability to the seeds also used to generate the next generation back crosses. Another challenge in measuring seed production was the size of well grown plants together with large number of them flowering at the same time. Thus we could not systematically pollinate the same flowers of all the plants, nor did we not succeed to collect all the siliques produced by plants at the optimal time of maturity. Too high density of the plants also led to contact between flowers of neighboring plants resulting in undesirable pollinations.

Just counting number of seeds per silique is not enough for measuring plant seed production, because it does not take into account number of flowers in a plant. I counted number of flowering shoots, but did not estimate number of flowers per shoot, which would have improved the fitness estimate significantly. However, if there are life history differences between the study populations, it makes measuring of reproduction even more complex. Plants from Mayodan put their resources to intensive flowering during the first flowering season, whereas plants from Spiterstulen are real perennials and often concentrate on growing their leaf rosette to prepare flowering in the coming years. Hybrids then express both of these two phenotypes, which leads to high differences in flower production during the first flowering, that is not immediately related to fitness but more a choice of strategy in dividing the resources for several years. I measured rosette sizes of the plants after flowering, which could have been transformed to fitness estimates for subsequent years, but that was not done. Of course, a better strategy would be to follow flowering for several years, if the length of the experiment was not a consideration. The problem may also be avoided by choosing the study populations in such a way that the plants invest flowering rather equally during the first season.

3.7 Incipient speciation?

Speciation in plants is unlikely to be a consequence of a single gene or isolating barrier (Lowry *et al.* 2008, Widmer *et al.* 2009). Post zygotic BDM

incompatibilities are thought not to be a major cause of speciation in plants, because they typically make small contributions to total isolation at late stage and correlate poorly with species boundaries (Rieseberg & Willis 2007). The incompatibilities presented here may not contribute much to speciation of the subspecies in case of possible secondary contact after isolation of 140 000 generations. The average male reproductive fitness of the hybrids is 3/4 compared to parental populations, which mostly consist of male sterility caused by CMS. Seed production of the hybrids was not less than observed in their parents, and in addition (Leinonen *et al.* 2011) observed higher total fitness of hybrids compared to their parents in the Norwegian environment.

More importantly, due to the dynamics of CMS and restorers, they arise and disappear cyclically and spread rapidly through hermaphroditic populations. Hence, secondary contact of hermaphrodite populations having different CMS and restorers would most likely lead to introgression of both CMS and its restorer through the whole hybrid population even faster than would be the fixation rate of a novel CMS mutation (Charlesworth & Ganders 1979, Fishman & Willis 2006). Thus the case is very different compared to reproductive barriers, which cause permanent reduction in hybrid fitness between diverged populations.

On the other hand, we do not know the time periods needed for CMS and *Rf* to evolve through one cycle or length of the different phases presented in the figure 1. It is not possible to predict at which phase (of fig. 1) these populations or subspecies would be at the time of the possible secondary contact, and how many CMS systems there would be at that time, if any. At the subspecies level the appearance of CMS also depends on the populations considered if CMS in *A. lyrata* evolves mainly at the local scale, as it does in *Mimulus* (Fishman & Willis 2006, Martin & Willis 2007, Martin & Willis 2007, Martin & Willis 2010). Because most hermaphrodite populations should lack functional CMS and restorers, hybrid male sterility is likely to be asymmetric. In the case studied here, the ability of Ma restorers to restore CMS of Sp in 40% of the hybrids allows introgression of nuclear genomes between populations even in both directions and thus speeds up formation of a stable hybrid population.

To be effective to initiate speciation with gene flow, lower hybrid fitness should either directly prevent gene flow between populations or lead to increased assortative mating by reinforcement (Feder *et al.* 2012). Thus, the incompatibilities arising in hybrids of *A. l. lyrata* from Mayodan and *A. l. petraea* from Spiterstulen are not necessarily an indication of incipient speciation, but

instead an interesting part of plant metapopulation dynamics (Pannell & Charlesworth 2000).

Biologists have always been interested in the origin of species and huge progress in understanding of speciation has been achieved, but many open questions still remain. Nature is so complex, that often there are similar outcomes caused by different processes. Not only in plants, but also for instance in birds, signs of speciation are sometimes reported with little evidence for the final direction of the process. Often it seems to be forgotten that speciation is the evolutionary process by which new biological species arise. We should remember that new species are not necessarily forming, if there is something unusual observed in the hybrid progeny of two diverged or different looking lineages. If the hybrids do not experience any measurable reduction in their fitness and gene flow between the parental lineages is almost unlimited as in many birds (Brodin *et al.* 2013, Elgvin *et al.* 2011, Hermansen *et al.* 2011), there is no guarantee or even a good reason why these lineages, or their hybrids, should become good species. In these examples there is a secondary contact of two previously allopatric populations, as it would be if the *A. lyrata* populations studied here were to meet again. In case of speciation, the two populations should remain distinct, but there are also other possible results of secondary contact, such as formation of a stable hybrid population or an allele frequency cline or even replacement of one population by the other.

4 Conclusions

Populations of *A. lyrata* are highly differentiated. The basis of the differentiation of European *A. l. petraea* populations is much older than would be the minimum expectation of the divergence after last glacial maximum. The European populations belong to three main groups, which are west, central and east. West European populations include Scandinavia, Iceland and presumably British isles, of which the Scandinavian populations have lost their genetic diversity likely due to bottlenecks during post glacial colonization. Central European populations are very diverse reflecting stable populations that have existed there for long time, whereas little diversity observed in the East European population is likely only a small fraction of the diversity existing in the main Eurasian *A. lyrata* population from which it originates. The origin of North American subspecies *A. l. lyrata* still remains unclear, but the direction of colonization there has evidently been from north to south.

Flowering time genes of *A. lyrata* show reduced levels of nucleotide diversity in all populations due to purifying or directional selection. Signs of selective sweeps in flowering time genes are observed especially in Northwest European populations in which the sweeps have occurred during the last glacial period before or during the expansion of these populations to their current locations. Photoperiodic pathway is important in sensing environmental cues to assure suitable time of flowering in the diverse environments where *A. lyrata* grows.

Outcrossing hermaphrodite *A. lyrata* is vulnerable to spread and fixation of CMS and fertility restorers, which leads to cryptic CMS systems that can be revealed by crossing diverged populations. However, nothing suggests that stable CMS polymorphisms leading to gynodioecious populations would exist in this species. In general, cryptic CMS can be common in many flowering plants, but due to its transient nature it does not have potential to permanently prevent gene flow between populations.

On the whole, high divergence of *A. lyrata* populations has many genetic consequences, part of which are due to stochastic demographic processes, but also many are driven by natural selection. Some of these lead to reduced reproductive success of hybrids made in artificial inter population crosses, but no clear signs of speciation in terms of irreversible post zygotic isolation exists.

References

Andres F & Coupland G (2012) The genetic basis of flowering responses to seasonal cues. Nature Reviews Genetics 13(9): 627–639.

Ansell SW, Stenoien HK, Grundmann M, Schneider H, Hemp A, Bauer N, Russell SJ & Vogel JC (2010) Population structure and historical biogeography of European Arabidopsis lyrata. Heredity 105(6): 543–553.

Balañá-Alcaide D, Ramos-Onsins SE, Boone Q & Aguadé M (2006) Highly structured nucleotide variation within and among Arabidopsis lyrata populations at the FAH1 and DFR gene regions. Molecular Ecology 15(8): 2059–2068.

Barbash DA, Roote J & Ashburner M (2000) The Drosophila melanogaster hybrid male rescue gene causes inviability in male and female species hybrids. Genetics 154(4): 1747–1771.

Barbash DA, Siino DF, Tarone AM & Roote J (2003) A rapidly evolving MYB-related protein causes species isolation in Drosophila. Proceedings of the National Academy of Sciences USA 100(9): 5302–5307.

Barr CM & Fishman L (2010) The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. Genetics 184(2): 455–465.

Bateson W (1909) Heredity and variation in modern lights. In: Seward AC (ed) Darwin and Modern Science. Cambridge, Cambridge University Press: 85–101.

Becquet C & Przeworski M (2007) A new approach to estimate parameters of speciation models with application to apes. Genome Research 17(10): 1505–1519.

Bentolila S, Alfonso A & Hanson M (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proceedings of the National Academy of Sciences U S A 99(16): 10887–10892.

Bomblies K & Weigel D (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. Nature Reviews Genetics 8(5): 382–393.

Bomblies K & Weigel D (2010) Arabidopsis and relatives as models for the study of genetic and genomic incompatibilities. Philosophical Transactions of the Royal Society B-Biological Sciences 365(1547): 1815–1823.

Braverman J, Hudson R, Kaplan N, Langley C & Stephan W (1995) The hitchhiking effect on the site frequency-spectrum of DNA polymorphisms. Genetics 140(2): 783–796.

Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X & Barbash DA (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. Science 314(5803): 1292–1295.

Brodin A, Haas F & Hansson B (2013) Gene-flow across the European crow hybrid zone - a spatial simulation. J Avian Biol 44(3): 281–287.

Broman K, Wu H, Sen S & Churchill G (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19(7): 889–890.

Bustamante C, Nielsen R, Sawyer S, Olsen K, Purugganan M & Hartl D (2002) The cost of inbreeding in *Arabidopsi*s. Nature 416(6880): 531–534.

Case AL & Willis JH (2008) Hybrid male sterility in mimulus (phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. Evolution 62(5): 1026–1039.

Charlesworth D (1981) A further study of the problem of the maintenance of females in gynodioecious species. Heredity 46(FEB): 27–39.

Charlesworth D & Charlesworth B (1978) Population-genetics of partial male-serility and evolution of monoecy and dioecy. Heredity 41(OCT): 137–153.

Charlesworth D & Ganders F (1979) Population-genetics of gynodioecy with cytoplasmic-genic male-sterility. Heredity 43(OCT): 213–218.

Chase C (2006) Genetically engineered cytoplasmic male sterility. Trends Plant Science 11(1): 7–9.

Chase CD (2007) Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends in Genetics 23(2): 81–90.

Cosmides L & Tooby J (1981) Cytoplasmic inheritance and intragenomic conflict. Journal of Theoretical Biology 89(1): 83–129.

Coyne JA & Orr HA (2004) Speciation. Sunderland USA, Sinauer Associates.

Coyne JA & Orr HA (1989) Patterns of speciation in *Drosophila*. Evolution 43(2): 362–381.

Darwin C (1859) The origin of species by means of natural selection. London. John Murray.

DeHaan A, Luyten R, Bak-Schotman T & VanDamme J (1997) The dynamics of gynodioecy in *Plantago lanceolata* L. .1. Frequencies of male-steriles and their cytoplasmic male sterility types. Heredity 79: 453–462.

Depaulis F, Mousset S & Veuille M (2003) Power of neutrality tests to detect bottlenecks and hitchhiking. Journal of Molecular Evolution 57: S190–S200.

Dobzhansky TH (1936) Studies on hybrid sterility. 11. localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics 21: 113–135.

Dufay M, Touzet P, Maurice S & Cuguen J (2007) Modelling the maintenance of male-fertile cytoplasm in a gynodioecious population. Heredity 99(3): 349–356.

Dufay M & Billard E (2012) How much better are females? The occurrence of female advantage, its proximal causes and its variation within and among gynodioecious species. Annals of Botany 109(3): 505–519.

Elgvin TO, Hermansen JS, Fijarczyk A, Bonnet T, Borge T, Saether SA, Voje KL & Saetre G (2011) Hybrid speciation in sparrows II: a role for sex chromosomes? Mol Ecol 20(18): 3823–3837.

Excoffier L, Foll M & Petit RJ (2009) Genetic consequences of range expansions. Annual Review of Ecology Evolution and Systematics 40: 481–501.

Feder JL, Egan SP & Nosil P (2012) The genomics of speciation-with-gene-flow. Trends in Genetics 28(7): 342–350.

Fishman L & Willis JH (2001) Evidence for Dobzhansky-Muller incompatibilities contributing to the sterility of hybrids between *Mimulus guttatus* and M. nasutus. Evolution 55(10): 1932–1942.

Fishman L & Willis JH (2006) A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. Evolution 60(7): 1372–1381.

Flowers JM, Hanzawa Y, Hall MC, Moore RC & Purugganan MD (2009) Population genomics of the *Arabidopsis thaliana* flowering time gene network. Molecular Biology and Evolution 26(11): 2475–2486.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J & Wilczek AM (2011) A Map of local adaptation in *Arabidopsis thaliana*. Science 333(6052): 86–89.

Foxe JP, Dar V, Zheng H, Nordborg M, Gaut BS & Wright SI (2008) Selection on amino acid substitutions in *Arabidopsis*. Molecular Biology and Evolution 25(7): 1375–1383.

Foxe JP & Wright SI (2009) Signature of diversifying selection on members of the pentatricopeptide repeat protein family in *Arabidopsis lyrata*. Genetics 183(2): 663–672.

Fujii S, Bond CS & Small ID (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proceedings of the National Academy of Sciences U S A 108(4): 1723–1728.

Fujii S & Small I (2011) The evolution of RNA editing and pentatricopeptide repeat genes. New Phytologist 191(1): 37–47.

Gossmann TI, Song B, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA & Eyre-Walker A (2010) Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Molecular Biology and Evolution 27(8): 1822–1832.

Gourbiere S & Mallet J (2010) Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". Evolution 64(1): 1–24.

Gouyon P, Vichot F & vanDamme J (1991) Nuclear-cytoplasmic male-sterility - single-point equilibria versus limit-cycles. American Naturalist 137(4): 498–514.

Green P, Falls K & Crook S (1990) Documentation for CRIMAP, Version 2.4. .

Hall D, Ma X & Ingvarsson PK (2011) Adaptive evolution of the *Populus tremula* photoperiod pathway. Molecular Ecology 20(7): 1463–1474.

Hall M & Willis J (2005) Transmission ratio distortion in intraspecific hybrids of Mimulus guttatus: Implications for genomic divergence. Genetics 170(1): 375–386.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F & Bergelson J (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. Science 333(6052): 83–86.

Hanson M & Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16: S154–S169.

Harushima Y, Nakagahra M, Yano M, Sasaki T & Kurata N (2001) A genome-wide survey of reproductive barriers in an intraspecific hybrid. Genetics 159(2): 883–892.

Heidel AJ, Ramos-Onsins SE, Wang W, Chiang T & Mitchell-Olds T (2010) Population history in *Arabidopsis halleri* using multilocus analysis. Molecular Ecology 19(16): 3364–3379.

Hermansen JS, Saether SA, Elgvin TO, Borge T, Hjelle E & Saetre G (2011) Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. Molecular Ecology 20(18): 3812–3822.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405(6789): 907–913.
Hewitt G (2004) Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 359(1442): 183–195.
Hohenlohe PA, Phillips PC & Cresko WA (2010) Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. International Journal of Plant Science 171(9): 1059–1071.
Hough J, Williamsson R, J. & Wright SI (2013) Patterns of selection on plant genomes. Annual Review of Ecology Evolution and Systematics 44: 3.1–3.19.
Hudson RR (1987) Estimating the recombination parameter of a finite population model without selection. Genet Res 50(3): 245–250.
Hudson R (1989) How often are polymorphic restriction sites due to a single mutation. Theoretical Population Biology 36(1): 23–33.
Hudson R, Kreitman M & Aguade M (1987) A Test of neutral molecular evolution based on nucleotide data. Genetics 116(1): 153–159.
Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. Genetics 180(1): 329–340.
Ingvarsson PK (2010) Natural selection on synonymous and nonsynonymous mutations shapes patterns of polymorphism in *Populus tremula*. Molecular Biology and Evolution 27(3): 650–660.
Kaplan N, Hudson R & Langley C (1989) The hitchhiking effect revisited. Genetics 123(4): 887–899.
Arabidopsis genome initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408(6814): 796–815.
Keller SR, Levsen N, Ingvarsson PK, Olson MS & Tiffin P (2011) Local selection across a latitudinal gradient shapes nucleotide diversity in balsam poplar, *Populus balsamifera* L. Genetics 188(4): 941–952.
Keller SR, Levsen N, Olson MS & Tiffin P (2012) Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. Molecular Biology and Evolution 29(10): 3143–3152.
Kimura M & Ohta T (1969) The average number of generations until fixation of a mutant gene in a finite population. Genetics 61: 763–771.
Kivimäki M, Kärkkainen K, Gaudeul M, Loe G & Agren J (2007) Gene, phenotype and function: *GLABROUS1* and resistance to herbivory in natural populations of *Arabidopsis lyrata*. Molecular Ecology 16(2): 453–462.
Koch MA & Matschinger M (2007) Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences U S A 104(15): 6272–6277.
Koelewijn H (2003) Variation in restorer genes and primary sexual investment in gynodioecious *Plantago coronopus*: the trade-off between male and female function. Proceedings of the Royal Society B-Biological Sciences 270(1527): 1939–1945.

Koelewijn H & vanDamme J (1995a) Genetics of male-sterility in gynodioecious *Plantago coronopus* .1. Cytoplasmic Variation. Genetics 139(4): 1749–1758.

Koelewijn H & vanDamme J (1995b) Genetics of male-sterility in gynodioecious *Plantago coronopus* .2. Nuclear Genetic-Variation. Genetics 139(4): 1759–1775.

Koide Y, Onishi K, Nishimoto D, Baruah AR, Kanazawa A & Sano Y (2008) Sex-independent transmission ratio distortion system responsible for reproductive barriers between Asian and African rice species. New Phytologist 179(3): 888–900.

Kuittinen H, de Haan A, Vogl C, Oikarinen S, Leppälä J, Koch M, Mitchell-Olds T, Langley C & Savolainen O (2004) Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. Genetics 168(3): 1575–1584.

Kumar S, Tamura K & Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in bioinformatics 5(2): 150–163.

Lagercrantz U & Axelsson T (2000) Rapid evolution of the family of *CONSTANS LIKE* genes in plants. Molecular Biology and Evolution 17(10): 1499–1507.

Lagercrantz U (2009) At the end of the day: a common molecular mechanism for photoperiod responses in plants? Journal of Experimental Botany 60(9): 2501–2515.

Le Corre V, Roux F & Reboud X (2002) DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: Extensive nonsynonymous variation is consistent with local selection for flowering time. Molecular Biology and Evolution 19(8): 1261–1271.

Leinonen PH, Remington DL, Leppälä J & Savolainen O (2013) Genetic basis of local adaptation and flowering time variation in *Arabidopsis lyrata*. Molecular Ecology 22(3): 709–723.

Leinonen PH, Remington DL & Savolainen O (2011) Local adaptation, phenotypic differentiation, and hybrid fitness in diverged natural populations of *Arabidopsis lyrata*. Evolution 65(1): 90–107.

Leinonen PH, Sandring S, Quilot B, Clauss MJ, Mitchell-Olds T, Agren J & Savolainen O (2009) Local adaptation in European populations of *Arabidopsis lyrata* (Brassicaceae). American Journal of Botany 96(6): 1129–1137.

Leppälä J, Bokma F & Savolainen O (2013) Investigating incipient speciation in *Arabidopsis lyrata* from patterns of transmission ratio distortion. Genetics 194(3): 697–708.

Leppälä J & Savolainen O (2011) Nuclear-cytoplasmic interactions reduce male fertility in hybrids of *Arabidopsis Lyrata* subspecies. Evolution 65(10): 2959–2972.

Lexer C & Widmer A (2008) The genic view of plant speciation: recent progress and emerging questions. Philosophical Transactions of the Royal Society B-Biological Sciences 363(1506): 3023–3036.

Li J, Li H, Jakobsson M, Li S, Sjodin P & Lascoux M (2012) Joint analysis of demography and selection in population genetics: where do we stand and where could we go? Molecular Ecology 21(1): 28–44.

Long Y, Zhao L, Niu B, Su J, Wu H, Chen Y, Zhang Q, Guo J, Zhuang C, Mei M, Xia J, Wang L, Wu H & Liu Y (2008) Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. Proceedings of the National Academy of Sciences U S A 105(48): 18871–18876.

Lowry DB, Modliszewski JL, Wright KM, Wu CA & Willis JH (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. Philosophical Transactions of the Royal Society B-Biological Sciences 363(1506): 3009–3021.

Lowry DB & Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. Plos Biology 8(9): e1000500.

Martin NH & Willis JH (2007) Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. Evolution 61(1): 68–82.

Martin NH & Willis JH (2010) Geographical variation in postzygotic isolation and its genetic basis within and between two *Mimulus* species. Philosophical Transactions of the Royal Society B-Biological Sciences 365(1552): 2469–2478.

Masel J (2011) Genetic drift. Current Biology 21(20): R837–R838.

Mayr E (1942) Systematics and the origins of species. New York, Columbia university press.

McCauley DE & Bailey MF (2009) Recent advances in the study of gynodioecy: the interface of theory and empiricism. Annals of Botany 104(4): 611–620.

Michaels SD (2009) Flowering time regulation produces much fruit. Current Opinion in Plant Biology 12(1): 75–80.

Moeller DA & Tiffin P (2008) Geographic variation in adaptation at the molecular level: a case study of plant immunity genes. Evolution 62(12): 3069–3081.

Moyle L & Graham E (2006) Genome-wide associations between hybrid sterility QTL and marker transmission ratio distortion. Molecular Biology and Evolution 23(5): 973–980.

Moyle L, Olson M & Tiffin P (2004) Patterns of reproductive isolation in three angiosperm genera. Evolution 58(6): 1195–1208.

Moyle LC & Graham EB (2005) Genetics of hybrid incompatibility between *Lycopersicon esculentum* and *L. hirsutum*. Genetics 169(1): 355–373.

Moyle LC & Nakazato T (2010) Hybrid incompatibility "snowballs" between *Solanum* species. Science 329(5998): 1521–1523.

Muller HJ (1939) Reversibility in evolution considered from the standpoint of genetics. Biological Reviews of the Cambridge Philosophical Society 14: 261–280.

Muller M, Leppälä J & Savolainen O (2008) Genome-wide effects of postglacial colonization in *Arabidopsis lyrata*. Heredity 100(1): 47–58.

Orr HA (1995) The population genetics of speciation: The evolution of hybrid incompatibilities. Genetics 139: 1805–1813.

Orr HA & Turelli M (2001) The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities. Evolution 55(6): 1085–1094.

Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, Shaw RG, Weigel D & Lynch M (2010) The rate and molecular spectrum of Spontaneous mutations in *Arabidopsis thaliana*. Science 327(5961): 92–94.

Pannell J & Charlesworth B (2000) Effects of metapopulation processes on measures of genetic diversity. Philosophical transactions of the Royal Society series B Biological sciences 355(1404): 1851–1864.

Phadnis N (2011) Genetic architecture of male sterility and segregation distortion in *Drosophila pseudoobscura* Bogota-USA hybrids. Genetics 189(3): 1001–1009.

Phadnis N & Orr HA (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. Science 323(5912): 376–379.

Presgraves DC, Balagopalan L, Abmayr SM & Orr HA (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. Nature 423(6941): 715–719.

Presgraves DC (2008) Sex chromosomes and speciation in *Drosophila*. Trends in Genetics 24(7): 336–343.

Presgraves DC (2010) The molecular evolutionary basis of species formation. Nature Reviews Genetics 11(3): 175–180.

Putterill J, Robson F, Lee K, Simon R & Coupland G (1995) The *Constans* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc-finger transcription factors. Cell 80(6): 847–857.

Quilot-Turion B, Leppälä J, Leinonen PH, Waldmann P, Savolainen O & Kuittinen H (2013) Genetic changes in flowering and morphology in response to adaptation to a high-latitude environment in *Arabidopsis lyrata*. Annals of Botany 111(5): 957–968.

Rieseberg L (2001) Chromosomal rearrangements and speciation. Trends in Ecology & Evolution 16(7): 351–358.

Rieseberg LH & Blackman BK (2010) Speciation genes in plants. Annals of Botany 106(3): 439–455.

Rieseberg LH & Willis JH (2007) Plant speciation. Science 317(5840): 910–914.

Riihimäki M, Podolsky R, Kuittinen H, Koelewijn H & Savolainen O (2005) Studying genetics of adaptive variation in model organisms: flowering time variation in *Arabidopsis lyrata.* Genetica 123(1–2): 63–74.

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(7): 1036–1045.

Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D & Gaut BS (2008) Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. Plos One 3(6): e2411.

Rozas J, Sánchez-DelBarrio JC, Messeguer X & Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics (Oxford, England) 19(18): 2496–2497.

Samis KE, Heath KD & Stinchcombe JR (2008) Discordant Longitudinal Clines in Flowering time and Phytochrome C in *Arabidopsis Thaliana*. Evolution 62(12): 2971–2983.

Sandring S, Riihimäki M-, Savolainen O & Agren J (2007) Selection on flowering time and floral display in an alpine and a lowland population of *Arabidopsis lyrata.* Journal of Evolutionary Biology 20(2): 558–567.

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

Schmid K, Ramos-Onsins S, Ringys-Beckstein H, Weisshaar B & Mitchell-Olds T (2005) A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. Genetics 169(3): 1601–1615.

Schnable P & Wise R (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. Trends in Plant Science 3(5): 175–180.

Scopece G, Musacchio A, Widmer A & Cozzolino S (2007) Patterns of reproductive isolation in Mediterranean deceptive orchids. Evolution 61(11): 2623–2642.

Scopece G, Widmer A & Cozzolino S (2008) Evolution of postzygotic reproductive isolation in a guild of deceptive orchids. American Naturalist 171(3): 315–326.

Shykoff J, Kolokotronis S, Collin C & Lopez-Villavicencio M (2003) Effects of male sterility on reproductive traits in gynodioecious plants: a meta-analysis. Oecologia 135(1): 1–9.

Staedler T, Florez-Rueda AM & Paris M (2012) Testing for "snowballing" hybrid incompatibilities in *Solanum*: Impact of ancestral polymorphism and divergence estimates. Molecular Biology and Evolution 29(1): 31–34.

Stephens M & Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. American Journal of Human Genetics 76(3): 449–462.

Stephens M, Smith N & Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 68(4): 978–989.

Sun S, Ting C & Wu C (2004) The normal function of a speciation gene, *Odysseus*, and its hybrid sterility effect. Science 305(5680): 81–83.

Sweigart A, Fishman L & Willis J (2006) A simple genetic incompatibility causes hybrid male sterility in mimulus. Genetics 172(4): 2465–2479.

Taberlet P, Fumagalli L, Wust-Saucy A & Cosson J (1998) Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7(4): 453–464.

Tajima F (1983) Evolutionary relationship of DNA-Sequences in finite populations. Genetics 105(2): 437–460.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3): 585–595.

Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24(8): 1596–1599.

Tiffin P, Olson M & Moyle L (2001) Asymmetrical crossing barriers in angiosperms. Proceedings of the Royal Society of London Series B-Biological Sciences 268(1469): 861–867.

van Damme J, Hundscheid M, Ivanovic S & Koelewijn H (2004) Multiple CMS - restorer gene polymorphism in gynodioecious *Plantago coronopus*. Heredity 93(2): 175–181.

vanDamme J & vanDelden W (1982) Gynodioecy in *Plantago lanceolata* L .1. Polymorphism for plasmon type. Heredity 49(DEC): 303–318.

Watterson G (1975) Number of segregating sites in genetic models without recombination. Theoretical Population Biology 7(2): 256–276.

Widmer A, Lexer C & Cozzolino S (2009) Evolution of reproductive isolation in plants. Heredity 102(1): 31–38.

Willi Y & Määttänen K (2010) Evolutionary dynamics of mating system shifts in *Arabidopsis lyrata*. Journal of Evolutionary Biology 23(10): 2123–2131.

Wise MJ, Vu JV & Carr DE (2011) Potential ecological constraints on the evolution of gynodioecy in *Mimulus guttatus*: Relative fecundity and pollinator behavior in a mixed-sex population. International Journal of Plant Science 172(2): 199–210.

Wright S & Charlesworth B (2004) The HKA test revisited: A maximum-likelihood-ratio test of the standard neutral model. Genetics 168(2): 1071–1076.

Wright S & Gaut B (2005) Molecular population genetics and the search for adaptive evolution in plants. Molecular Biology and Evolution 22(4): 1158–1158.

Wright S, Lauga B & Charlesworth D (2003) Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. Molecular Ecology 12(5): 1247–1263.

Wright SI & Andolfatto P (2008) The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. Annual Review of Ecology Evolution and Systematics 39: 193–213.

Yamagata Y, Yamamoto E, Aya K, Win KT, Doi K, Sobrizal, Ito T, Kanamori H, Wu J, Matsumoto T, Matsuoka M, Ashikari M & Yoshimura A (2010) Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. Proceedings of the National Academy of Sciences USA 107(4): 1494–1499.

Yamamoto MP, Shinada H, Onodera Y, Komaki C, Mikami T & Kubo T (2008) A male sterility-associated mitochondrial protein in wild beets causes pollen disruption in transgenic plants. Plant Journal 54(6): 1027–1036.

Original articles

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

II Aalto EA, Toivainen T, Niittyvuopio A, Piltonen S, Kuittinen H & Savolainen O (2013) Selection on flowering time genes in *Arabidopsis lyrata* ssp. (Manuscript).

III Aalto EA, Koelewijn HP & Savolainen O (2013) Cytoplasmic Male Sterility contributes to hybrid incompatibility between subspecies of *Arabidopsis lyrata.* G3: Genes, Genomes, Genetics 3: 1727–1740.

Reprinted with permission from Botanical society of America (I) and Genetics society of America (III).

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

602. Juntunen, Kaisu (2012) Tieto- ja viestintätekniikan soveltamiseen perustuvat toimintaprosessien uudistukset terveydenhuollossa : sosio-teknis-taloudellinen näkökulma

603. Seppä, Karri (2012) Quantifying regional variation in the survival of cancer patients

604. Kuvaja, Pasi (2012) Software process capability and maturity determination : BOOTSTRAP methodology and its evolution

605. Laanti, Maarit (2012) Agile Methods in large-scale software development organizations : applicability and model for adoption

606. Hokkanen, Juho (2013) Liquid chromatography/mass spectrometry of bioactive secondary metabolites – *in vivo* and *in vitro* studies

607. Kuokkanen, Matti (2013) Development of an eco- and material-efficient pellet production chain—a chemical study

608. Jansson, Eeva (2013) Past and present genetic diversity and structure of the Finnish wolf population

609. Myllykoski, Matti (2013) Structure and function of the myelin enzyme 2′,3′-cyclic nucleotide 3′-phosphodiesterase

610. Lehto, Tuomas (2013) The importance of persuasive systems design in enhancing consumers' perceptions and adoption of health behavior change support systems

611. Hernoux-Villière, Audrey (2013) Catalytic depolymerisation of starch-based industrial waste : use of non-conventional activation methods and novel reaction media

612. Lawrence, Carl (2013) Innovating with information technology in a globalized world : being proactive about culture

613. Ardanov, Pavlo (2013) Priming capacities of endophytic *Methylobacterium* sp. on potato (*Solanum tuberosum* L.)

614. Koskela, Anni (2013) Wolverine habitat selection, diet and conservation genetics

615. Holm, Jana (2013) Catalytic pretreatment and hydrolysis of fibre sludge into reducing sugars

616. Kemi, Ulla (2013) Adaptation to growing season length in the perennial *Arabidopsis lyrata*

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

SERIES EDITORS

A SCIENTIAE RERUM NATURALIUM

Professor Esa Hohtola

B HUMANIORA

University Lecturer Santeri Palviainen

C TECHNICA

Postdoctoral research fellow Sanna Taskila

D MEDICA

Professor Olli Vuolteenaho

E SCIENTIAE RERUM SOCIALIUM

University Lecturer Hannu Heikkinen

F SCRIPTA ACADEMICA

Director Sinikka Eskelinen

G OECONOMICA

Professor Jari Juga

EDITOR IN CHIEF

Professor Olli Vuolteenaho

PUBLICATIONS EDITOR

Publications Editor Kirsti Nurkkala

ISBN 978-952-62-0321-8 (Paperback)
ISBN 978-952-62-0322-5 (PDF)
ISSN 0355-3191 (Print)
ISSN 1796-220X (Online)

UNIVERSITY of OULU
OULUN YLIOPISTO