

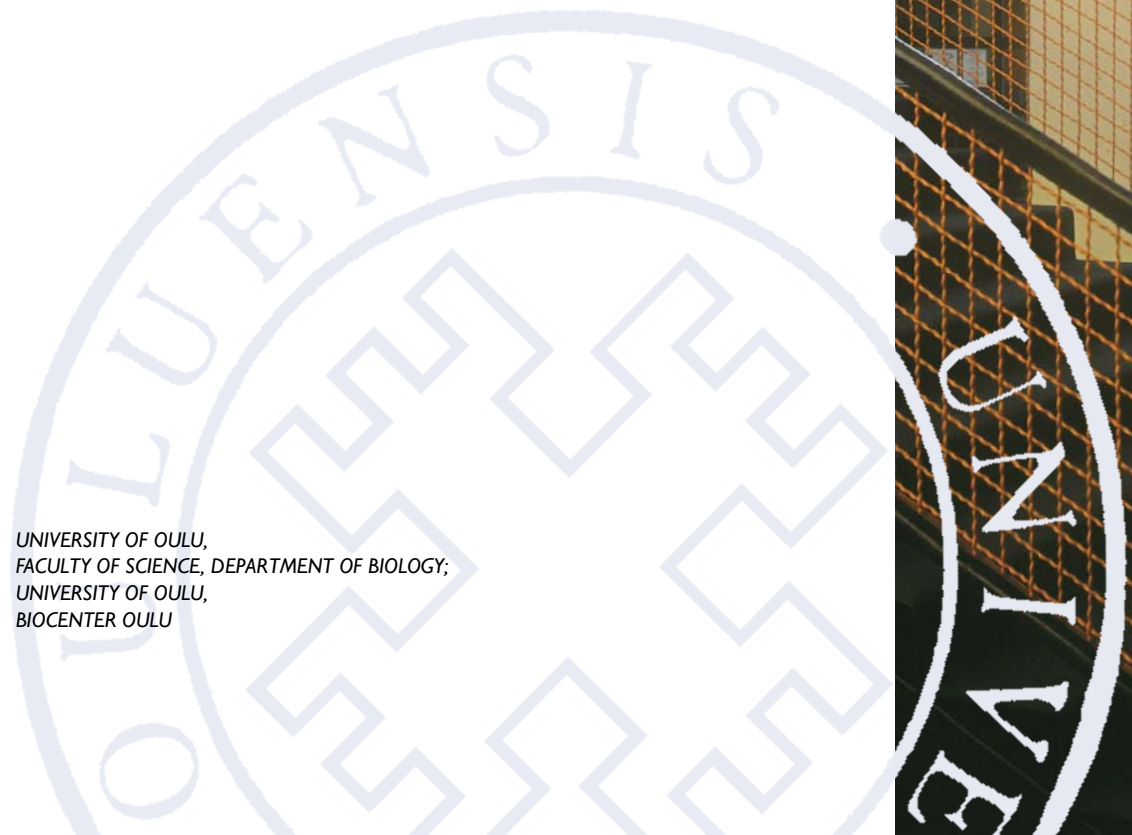
*Johanna Leppälä*

THE GENETIC BASIS OF  
INCIPIENT SPECIATION  
IN *ARABIDOPSIS LYRATA*

UNIVERSITY OF OULU,  
FACULTY OF SCIENCE, DEPARTMENT OF BIOLOGY;  
UNIVERSITY OF OULU,  
BIOCENTER OULU

A

SCIENTIAE RERUM  
NATURALIUM





ACTA UNIVERSITATIS OULUENSIS  
A Scientiae Rerum Naturalium 568

*JOHANNA LEPPÄLÄ*

**THE GENETIC BASIS OF INCIPIENT  
SPECIATION IN *ARABIDOPSIS LYRATA***

Academic dissertation to be presented with the assent of  
the Faculty of Science of the University of Oulu for public  
defence in Kaljusensali (Auditorium KTK 112), Linnanmaa,  
on 4 February 2011, at 12 noon

UNIVERSITY OF OULU, OULU 2011

Copyright © 2011  
Acta Univ. Oul. A 568, 2011

Supervised by  
Professor Outi Savolainen  
Docent Helmi Kuittinen

Reviewed by  
Professor Teemu Teeri  
Docent Xiao-Ru Wang

ISBN 978-951-42-9356-6 (Paperback)  
ISBN 978-951-42-9357-3 (PDF)  
<http://herkules.oulu.fi/isbn9789514293573/>  
ISSN 0355-3191 (Printed)  
ISSN 1796-220X (Online)  
<http://herkules.oulu.fi/issn03553191/>

Cover Design  
Raimo Ahonen

JUVENES PRINT  
TAMPERE 2011

**Leppälä, Johanna, The genetic basis of incipient speciation in *Arabidopsis lyrata***

University of Oulu, Faculty of Science, Department of Biology, P.O.Box 3000, FI-90014  
University of Oulu, Finland; University of Oulu, Biocenter Oulu, P.O.Box 5000, FI-90014  
University of Oulu, Finland

*Acta Univ. Oul. A 568, 2011*

Oulu, Finland

***Abstract***

The study of speciation has been an area of primary interest in evolutionary biology from Darwin to the present day. Understanding the processes that give rise to new species requires knowledge on how reproductive isolation develops between diverging populations. An irreversible and therefore important component of reproductive isolation is intrinsic postzygotic isolation. Postzygotic incompatibilities often manifest themselves through hybrid inviability or sterility, and distort allelic transmission ratios in hybrid progenies. The genetic basis of such incompatibilities has often been found to be negative interactions between two or more loci, also known as Bateson-Dobzhansky-Muller incompatibilities. During the last decade some genes involved in this type of incompatibilities have been identified, but especially in plants our knowledge remains scarce. In this thesis I examined whether intrinsic postzygotic isolation had developed between allopatric populations of a perennial, outcrossing plant; *Arabidopsis lyrata*.

The studied populations of *A. lyrata* were found to be genetically highly differentiated. In  $F_2$  progenies of crosses between populations many molecular markers reveal non-Mendelian genotype ratios, that is, show transmission ratio distortion (TRD). By contrast, TRD was found to be nearly absent in progenies of within population crosses. The degree of TRD clearly increased with genetic distance between the crossed populations, and origin for TRD was often in  $F_1$  gamete formation, instead of  $F_2$  zygotic level. The genetic basis of TRD appeared due to interactions between nuclear loci, and between nuclear and cytoplasmic factors.

In addition to transmission ratio distortion, reduced male fertility and cytoplasmic male sterility was found in the  $F_2$  hybrids between *A. lyrata* subspecies *petraea* and *lyrata*. Quantitative trait loci for reduced male fertility were polymorphic within populations, and dependent on cytoplasm. Thus, the findings in this thesis underline the role of cytonuclear interactions, and the possibility of development of genic incompatibilities through genomic conflicts due to divergence likely unrelated to local adaptation.

**Keywords:** *Arabidopsis lyrata*, genetic linkage map, hybrid incompatibility, population divergence, speciation, transmission ratio distortion



## **Leppälä, Johanna, Alkavan lajiutumisen geneettinen tausta *Arabidopsis lyrata***

Oulun yliopisto, Luonnontieteellinen tiedekunta, Biologian laitos, PL 3000, 90014 Oulun yliopisto; Oulun yliopisto, Biocenter Oulu, PL 5000, 90014 Oulun yliopisto

*Acta Univ. Oul. A 568, 2011*

Oulu

### ***Tiivistelmä***

Lajiutumistutkimus on ollut yksi evoluutiobiologian suurimmista kiinnostuksen kohteista Darwinin ajoista nykypäivään. Jotta voisimme ymmärtää uusien lajien syntymiseen johtavia prosesseja, on ensin ymmärrettävä, miten lisääntymisisolaatio kehittyy erilaistuvien populaatioiden välille. Lajien välisten risteytysten jälkeläiset ovat usein steriilejä tai elinkyvyttömiä. Tämä perustuu kahden erilaistuneen populaation perimän yhteensopimattomuuteen. Tällaiset sisäsynyiset hedelmöityksen jälkeiset yhteensopimattomuudet ovat tärkeä osa lisääntymisisolaatiota, sillä ne ovat peruuttamattomia. Ne aiheuttavat vääristymiä genotyyppien lukusuhteissa risteytysjälkeläistöissä. Posttsgoottisen yhteensopimattomuuden taustalla on usein kahden tai useamman geenilokuksen välinen negatiivinen interaktio, jota kutsutaan myös Bateson-Dobzhansky-Muller -inkompatibiliteetiksi. Viime vuosikymmenen aikana on tunnistettu muutamia geenejä, jotka ovat mukana negatiivisissa interaktioissa, mutta kasveista tietoa geneettisistä tekijöistä on vielä vähän. Väitöstutkimuksessa selvitettiin onko geneettisiä posttsgoottisia yhteensopimattomuuksia kehittynyt kasvilajin sisällä (*Arabidopsis lyrata*) allopatristen populaatioiden välille.

Tutkitut kasvipopulaatiot olivat geneettisesti hyvin erilaistuneita. Näiden populaatioiden välisten jälkeläistöjen genotyypeissä havaittiin yleisesti paljon mendelistisistä lukusuhteista poikkeavia lukusuhteita. Genotyypejä tarkasteltiin myös populaation sisäisissä risteytyksissä, joiden jälkeläistöissä periytyminen oli lähes mendelististä. Väitöstutkimuksessa havaittiin, että mitä geneettisesti erilaisemmat populaatiot risteytettiin, sitä suurempi oli ei-mendelistisen periytyksen osuus. Tutkimukset osoittivat lukusuhteiden vääristyvän jo  $F_1$ -hybridin tuottaessa sukupuolisoluja, eikä niinkään muodostuneissa uusissa  $F_2$ -tsygooteissa. Taustalla olevien geneettisten yhteensopimattomuuksien havaittiin vaikuttavan tuman geenien välillä, mutta myös tuman ja soluliman geneettisten tekijöiden välillä.

Mendelistisistä lukusuhteista poikkeamisen lisäksi tutkimuksessa havaittiin siitepölyn alentunut elinkyky ja solulimaan liittyvä koirassteriliteetti  $F_2$ -jälkeläistössä *A. lyrata* alalajien *petraea* ja *lyrata* välillä. Lokukset, jotka vaikuttivat siitepölyn alentuneeseen elinkykyyn, olivat polymorfisia populaatioiden sisällä ja vuorovaikutuksessa soluliman geneettisten tekijöiden kanssa. Nämä tutkimustulokset korostavat tuman ja soluliman geenien vuorovaikutuksen tärkeyttä lajien välisten lisääntymisesteiden synnyssä. Ne myös osoittavat, että yhteensopimattomuus voi kehittyä mahdollisesti ilman luonnonvalinnan vaikutusta.

**Asiasanat:** *Arabidopsis lyrata*, geneettinen kytkentäkartta, lajiutuminen, populaatioiden erilaistuminen, risteymien yhteensopimattomuus, transmissiosuhteiden vääristymä





## Acknowledgements

First of all I would like to thank my supervisor Outi Savolainen who offered me this most interesting research topic which has greatly grown and developed during my thesis work. Thank you for your expert guidance on scientific issues and academic life in general. I would also like to thank my other supervisor Helmi Kuittinen for her valuable opinions on research issues and in guiding me into the wonderful world of genetic mapping and TRD.

I am grateful to professor Teemu Teeri and docent Xiao-Ru Wang for efficiently reviewing this thesis. My co-authors, especially Jesper Bechsgaard, Folmer Bokma and Marie-Hélène Muller, are warmly thanked for collaboration, and all those discussions on scientific and non-scientific matters. I acknowledge financial support by the Research Council for Biosciences and Environment, Finnish Graduate School in Population Genetics and Biocenter Oulu.

I wish to thank the former and current members of the Plant Genetics Group in Oulu – I cannot imagine a better atmosphere for working than what you have created during the years. I have had excellent help and guidance in the lab. Most of all I would like to thank Meeri Otsukka who superbly multitasked with all the lab work I gave to her – thanks Meeri, you are the best! Also Soile Alatalo created that wonderful spirit in the lab! It has been a great pleasure to work with you and thanks for all the help during the years. I would also like to thank Hannele Parkkinen for skillful and efficient work with microsatellites and sequences, as well as Asta Airikainen, Saana Remula and Terhi Teräväinen for their help in the lab.

The experiments in the greenhouse would not have been accomplished without the dedication and help from Ulla Kemi, Jukka Kiiskilä, Riikka Syrjänen and Tuomas Toivainen. Thanks also to the staff of the Botanical Gardens of Oulu, especially Annikki Kestilä, Tuula Kangas, Aino Hämäläinen, Tuomas Kauppila and Marko Hyvärinen.

Then I would like to thank the people who shared an office with me... and not just for sharing the office but also for sharing a full spectrum of emotions during these years! Thanks to my friends Päivi Leinonen and Ulla Kemi, but also to all the others who have made their, shorter or longer, stay in the same office with me – Anne, Annaleena, Jarkko, Saana, Sanna, Anu and Eeva. I am grateful for having had an opportunity to work (and make friends) with wonderful people in the genetics group in Oulu: Tanja, Sonja, Bénédicte, Mari, Matti, Lumi... and all the others!

Finally, my warmest thanks go to my family and friends. Thanks for being there for me. Kiitos.

Umeå, Sweden, December 2010

Johanna Leppälä

## Abbreviations

BDMI	Bateson-Dobzhansky-Muller incompatibility
CAPS	cleaved amplified polymorphic sequence
CMS	cytoplasmic male sterility
MYA	million years ago
PCR	polymerase chain reaction
PPR	pentatricopeptide repeat
SNP	single nucleotide polymorphism
TRD	transmission ratio distortion
QTL	quantitative trait locus



## List of original articles

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

- I Muller M-H, Leppälä J & Savolainen O (2008) Genome wide effects of postglacial colonization in *Arabidopsis lyrata*. *Heredity* 100(1): 47–58.
- II Kuittinen H, de Haan AA, Vogl C, Oikarinen S, Leppälä J, Koch M, Mitchell-Olds T, Langley CH & Savolainen O (2004) Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. *Genetics* 168(3): 1575–1584.
- III Leppälä J, Bechsgaard JS, Schierup MH & Savolainen O (2008) Transmission ratio distortion in *Arabidopsis lyrata*: Effects of population structure and the S-locus. *Heredity* 100 (1): 71–78.
- IV Leppälä J & Savolainen O Nuclear-cytoplasmic interactions reduce male fertility in hybrids of *Arabidopsis lyrata* subspecies. In press.
- V Leppälä J, Bokma F & Savolainen O Genetic basis of transmission ratio distortion in crosses between populations of *Arabidopsis lyrata*. Manuscript.



# Contents

<b>Acknowledgements</b>	<b>7</b>
<b>Abbreviations</b>	<b>9</b>
<b>List of original articles</b>	<b>11</b>
<b>Contents</b>	<b>13</b>
<b>1 Introduction</b>	<b>15</b>
1.1 Bateson-Dobzhansky-Muller -incompatibilities .....	16
1.2 Isolating barriers and the importance of reproductive isolation .....	17
1.3 Postzygotic isolation .....	18
1.3.1 Intrinsic postzygotic isolation.....	18
1.3.2 Genic incompatibilities.....	19
1.3.3 Chromosomal rearrangements.....	21
1.3.4 Transmission ratio distortion and reproductive isolation.....	22
1.3.5 Asymmetry in reciprocal crosses.....	23
1.4 Speciation rates and forces driving speciation .....	25
1.5 The Genus <i>Arabidopsis</i> .....	27
1.5.1 <i>Arabidopsis lyrata</i> and its history.....	27
1.6 Aims of the study .....	28
<b>2 Materials and methods</b>	<b>31</b>
2.1 Plant material and crosses .....	31
2.2 Phenotyping .....	32
2.3 Molecular methods.....	32
2.3.1 Microsatellite variation.....	32
2.3.2 Gene-based markers .....	33
2.4 Statistical methods .....	34
2.4.1 Population differentiation.....	34
2.4.2 Genetic linkage map construction .....	34
2.4.3 QTL mapping .....	35
2.4.4 Analysis of transmission ratio distortion .....	35
<b>3 Results and discussion</b>	<b>37</b>
3.1 Evolutionary history of <i>Arabidopsis lyrata</i> .....	37
3.2 Transmission ratio distortion in within and between population crosses .....	39
3.3 Reproductive isolation in <i>Arabidopsis lyrata</i> .....	44
3.4 Genetics of reduced hybrid male fertility and cytoplasmic male sterility .....	46

<b>4 Conclusions</b>	<b>49</b>
<b>References</b>	<b>51</b>



# 1 Introduction

150 years after Charles Darwin's seminal publication *On the Origin of Species* (1859) biologists are still puzzled by the problem of speciation. How do new species arise? What is the genetic basis of the emergence of new species? How do the genes involved in speciation diverge—through natural selection, drift, or some other mechanism?

To study speciation one must define what a species is. At first sight it seems evident how to define species e.g. based on differences in their morphology. For describing species based on similar characteristics in individuals' morphology, a typological species concept was developed by the very early taxonomists. A species concept is something artificial that scientific community has made to be able to categorize things seen in nature. For typological species categorization, phenotypic species differences are important, but they usually become visible when time to the most recent common ancestor is long. Timescales for speciation are often from thousands to millions of years (Price 2007), although there is evidence for rapid speciation e.g. in the cichlid fishes (Seehausen 2002). It is difficult to define species boundaries based solely on morphology, especially when time to the most recent common ancestor is short. One needs some other criteria, than only morphology, to work with. During modern history different species concepts have accumulated and prevailed depending on the purpose of the categorization. Especially at the end of the 20<sup>th</sup> century many new species concepts were developed (reviewed in Coyne & Orr 2004). The most widely used species concept is the biological species concept (Mayr 1942). This concept defines that a species is a group of interbreeding individuals. Ernst Mayr (1942) formulated biological species concept as follows: "Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups". In this thesis the biological species concept is used.

When speciation process is ongoing, or when species have only recently become reproductively separated, it may be very difficult to distinguish whether there are two species. True species may under certain circumstances interbreed and form a somewhat stable hybrid zone (Johnston *et al.* 2004). Of plant families studied by Ellstrand *et al.* (1996), 16% to 34% contained at least one reported hybrid. Therefore the biological species concept is sometimes relaxed: many species are accepted to be true species even though they may hybridize with other species – as long as they are able to maintain their genetic integrity and not fuse into other species.

But how do these reproductively isolated units emerge? It took almost 80 years from the publication of Darwin's book for geneticists to develop a model which explains how hybrid sterility or inviability could develop without the problem of hybrids falling into fitness valleys. This model was first described by William Bateson (1909) but this paper remained largely unknown until highlighted by Orr (1996). Although it seemed that Bateson, in his later years, was not fully convinced how species formed, still the model he presented in 1909 was analogous to the later models independently presented by Dobzhansky and Muller. Theodosius Dobzhansky came at the end of 1930's into the same conclusions as Bateson earlier (Dobzhansky 1936) and during 1940's Herman J. Muller refined the model of genic speciation (Muller 1942), which later on became known as Dobzhansky-Muller model.

### **1.1 Bateson-Dobzhansky-Muller -incompatibilities**

In the Bateson-Dobzhansky-Muller -model the simplest scenario is that incompatibility develops between two loci. First, one population becomes divided into two e.g. by a geographical barrier and these populations gain new alleles through mutation. When these populations become secondarily sympatric, so that they could potentially reproduce, there may be new alleles at different loci that are not compatible with each other. New alleles might also originate in just one lineage so that incompatibility is formed between ancestral and derived alleles. Incompatibilities may be more complicated (polygenic epistatic interactions) than this simplest-case scenario, as shown in *Drosophila* flies (Tao *et al.* 2003). Theory predicts that the number of incompatibilities increases faster than linearly in time (Orr 1995), although this theory is argued to be over-simplified and has been challenged by Gourbiere and Mallet (2010) but also supported by data from *Drosophila* (Matute *et al.* 2010) and *Solanum* (Moyle & Nakazato 2010).

The work of Dobzhansky and Muller was carried out with *Drosophila* flies but genetic incompatibilities appear to be universal and have been found in a wide range of organisms from plants to animals and yeasts (Lee *et al.* 2008, Lowry *et al.* 2008, Presgraves 2010). Some other mechanisms than BDM-incompatibilities also contribute to formation of hybrid inviability or sterility. For example in yeasts chromosomal rearrangements and antirecombination, the inability of diverged chromosomes to form crossovers, seem far more common causes of reproductive isolation than BDM incompatibilities (Greig 2009). Chromosomal inversions have also been observed in other species to influence the viability of

gametes of a hybrid heterozygous for an inversion. This reduced viability of gametes causes lower hybrid fitness and can act as a reproductive barrier between individuals with different chromosomal arrangements. An alternative model for chromosomal speciation was introduced more recently and it describes the reduction of hybrid fitness resulting from the reduced recombination between inverted and noninverted regions of chromosomes (reviewed in Hoffmann & Rieseberg 2008). Chromosomal speciation, especially in the form of polyploidy, is more common in plants than in animals (Ayala & Coluzzi 2005). The general view of the genetic background of speciation processes is still based on a relatively few studies. Only a few speciation genes involved in BDM-incompatibilities have been identified to the molecular level so far and almost all knowledge comes from model organisms such as *Drosophila* (Ting *et al.* 1998, Barbash *et al.* 2003, Brideau *et al.* 2006, Phadnis & Orr 2009, Bayes & Malik 2009) or mouse (Mihola *et al.* 2009). In plants studies have focused on agriculturally important species like rice (reviewed in Ouyang *et al.* 2010) and tomato (Moyle & Graham 2005, Moyle & Nakazato 2008) or on the model species *Arabidopsis thaliana*, in which some incompatibility genes have been identified (Bomblies *et al.* 2007, Bikard *et al.* 2009). Nevertheless, some intensive efforts have been made to understand speciation processes in natural, economically unimportant, species like *Mimulus* (Wu *et al.* 2008) and *Iris* (Arnold *et al.* 2004). With current sequencing techniques the discovery of speciation genes will become more feasible and quick advances are to be expected in this field.

## **1.2 Isolating barriers and the importance of reproductive isolation**

For a new species to develop, reproductive isolation must form. It may be acquired gradually over a long period of time or it can be reached rapidly. A classical example of rapid gain of reproductive isolation, and speciation, is polyploid speciation (reviewed in Ramsey & Schemske 1998). The new species can become reproductively isolated from its progenitors immediately because of different ploidy levels of parents and hybrids.

Rapid gain of reproductive isolation may also be achieved by behavioural change e.g. host-plant shift for mating. This has happened in *Rhagoletis* flies of which apple and hawthorn races have been reproductively isolated for a relatively short time after an introduction of a new host plant, apple, in North America (Feder *et al.* 1994).

Isolating barriers can be divided into pre-mating and post-mating barriers. The example of polyploid speciation belongs to the latter group if failure of new polyploid species to reproduce with either of its progenitors is due to failure of chromosomal pairing in meiosis. Post-mating barrier may also form if produced hybrids are not viable or if they have very low fitness in their environment e.g. by being phenotypically average compared to their parents and therefore not attractive to mates. In contrast, the host shift in *Rhagoletis* is an example of a pre-mating barrier that prevents actual mating from happening. In plants flowering time difference would also serve as a pre-mating barrier. Lowry *et al.* (2008) analysed 19 pairs of plant taxa and concluded that on average pre-zygotic isolation is typically stronger than post-zygotic isolation in plants. However, in a recent review Widmer *et al.* (2009) underline the importance of viewing reproductive isolation as consisting of several pre- and post-zygotic barriers.

### **1.3 Postzygotic isolation**

Postzygotic isolation refers to a barrier that acts after the hybrid is formed. It can be classified as a post-mating barrier but it does not include barriers that act between mating and zygote formation. Postzygotic isolation may become visible in hybrids as viability problems or problems in hybrid fertility. Postzygotic barrier may also form if the hybrids produced have very low fitness in their environment by having an intermediate phenotype compared to the parents e.g. so that they do not attract pollinators or have different flowering time than either parental species and therefore lack, or have fewer, possibilities to mate.

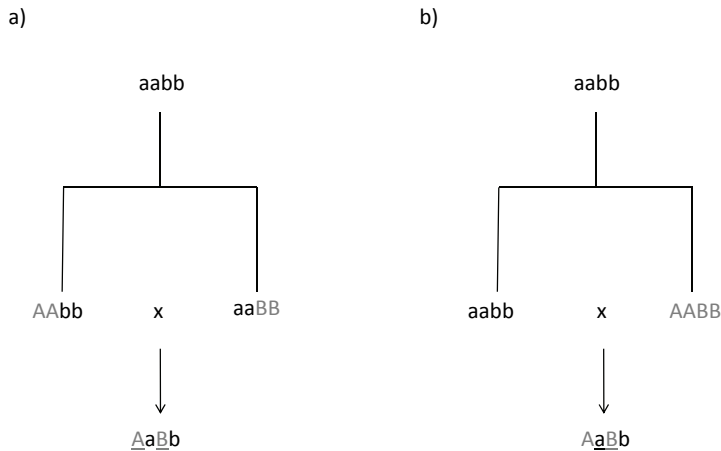
#### **1.3.1 Intrinsic postzygotic isolation**

Intrinsic postzygotic isolation refers to a barrier that appears due to some features of the hybrid independent of the environment. These barriers may affect the hybrid's viability – they may cause death of all individuals or just be lethal in some individuals. For instance, in *Drosophila* hybrids specifically male killing hybrid incompatibility factors are known (Presgraves *et al.* 2003). Another way of intrinsic postzygotic isolation to manifest itself is through reduced fertility of hybrids. It is feasible that hybrids do not have any viability problems but then suffer from partially reduced fertility or even total sterility. Haldane's rule summarises the observation, frequently made in *Drosophila*, but also in other species, that the heterogametic sex suffers more from hybrid problems than

homogametic sex (Haldane 1922). This could be due to incompatibilities being recessive and being exposed in the heterogametic sex or male fertility related genes could be evolving faster e.g. due to sexual selection (Wu & Davis 1993). In plants both male and female functions have been observed to experience fertility reductions in *Helianthus*, *Mimulus* and *Solanum* (Rieseberg 2000, Fishman & Willis 2001, Moyle & Graham 2005). However, between population hybrids of *Arabidopsis thaliana* have showed low fertility due to problems only in male gametogenesis (Törjék *et al.* 2006).

### **1.3.2 Genic incompatibilities**

Hybrid inviability or reduced fertility may be caused by several underlying mechanisms. The most common is thought to be Bateson-Dobzhansky-Muller incompatibility that appears between two or more loci. It may develop when two lineages become allopatric and different new alleles are fixed in different populations. When coming into secondary contact these new alleles may be incompatible with each other (Figure 1a) and cause either failure in hybrid's gamete production or complete abortion of a hybrid zygote. The incompatibility may also develop between ancestral and derived alleles, if e.g. in a two-locus scenario new alleles develop only in one of the lineages. At the secondary contact, incompatibility is found between an ancestral allele maintained in the first lineage and a new derived allele from the second lineage (Figure 1b). The data on the prevalence of derived-derived or ancestral-derived incompatibilities is still scarce. Ancestral-derived incompatibility has been found to cause hybrid male sterility in *Drosophila* (Masly *et al.* 2006) as well as hybrid inviability in three sister species of *Drosophila* (Cattani & Presgraves 2009). In most of the cases, when hybrid incompatibility genes have been identified, it has not been demonstrated in which lineage the derived allele/s originated.



**Fig. 1. A schematic model for development of derived-derived (a) and ancestral-derived two-locus incompatibilities, modified from Cattani & Presgraves (2009). Ancestral alleles are in black and derived alleles in grey. The incompatible alleles are underlined in the hybrid.**

However, new insights into the nature of DMIs have been gained when some of these genes causing incompatibilities – so-called speciation genes – have been identified. In *Drosophila* species several genes involved in hybrid viability or sterility are known. Quite often they appear to be fast evolving. Several cases have been described where the locus pairs are involved in maintenance of within species genetic balance, e.g. a transmission ratio distorter and its suppressor (reviewed by Presgraves 2010). Incompatibility could be caused by selfish elements that advance their own transmission and could potentially cause problems when interacting with a genome that has not developed mechanisms to control its transmission. It is also feasible that genes that suppress action of the selfish elements are involved in hybrid incompatibility. There are examples of both. The *Ovd* from *D. pseudoobscura bogotana* (Phadnis & Orr 2009) is the gene causing segregation distortion when in *D. p. pseudoobscura* background, but it also causes male sterility. On the other hand, the incompatibility genes *Lhr* and *Hmr* (*D. simulans* and *D. melanogaster* crosses), and *Zhr* with same species pair,

seem to control regulation of heterochromatic sequences (Brideau *et al.* 2006, Ferree & Barbash 2009).

In plants very few speciation genes have been identified so far. Many studies have managed to map hybrid incompatibility genes to a narrow chromosomal region but the identity of the genes might still be hidden. In the model species *Arabidopsis thaliana* two interesting cases of postzygotic incompatibility have been revealed to genic level. The incompatibilities have been observed between different selfing accessions within the species. In one case the incompatibility establishes itself as hybrid necrosis, as an autoimmunity response is triggered by an interaction of a TIR-NB-LRR class pathogen resistance gene and still unidentified partner gene from the other accession (Bombliet *et al.* 2007). The second case where the gene causing incompatibility is known is caused by differential fate of duplicated HPA gene copies in selfing lineages (Bikard *et al.* 2009). In different accessions one or the other copy has been silenced, which causes embryo lethality in a hybrid that does not receive a functional copy from either of the parents.

### **1.3.3 Chromosomal rearrangements**

Besides BDM incompatibilities, chromosomal rearrangements (peri- or paracentric inversions, chromosomal fusions or reciprocal translocations) have been thought to be a possible mechanism causing reproductive isolation. These rearrangements would cause problems in chromosome pairing in meiosis and yield aneuploid gametes. Already in the beginning of 20<sup>th</sup> century it had been noticed that individuals heterozygous for chromosomal inversions suffered from meiotic problems which often reduced their fertility (Müntzing 1930). Especially in plants evidence for the role of chromosomal rearrangements in the development of reproductive isolation has been found. In plants genome doubling has returned fertility of individuals heterozygous for chromosomal rearrangement (Stebbins 1958). More recently QTLs for pollen viability have been mapped to chromosomal rearrangements between diploid sunflower species *Helianthus annuus* and *H. petiolaris* (Lai *et al.* 2005). Similarly hybrid sterility in *Drosophila pseudoobscura* and *D. persimilis* hybrids associates with fixed chromosomal inversions (Noor *et al.* 2001). Noor and colleagues propose an alternative mechanism for chromosomal inversions to cause RI than meiotic pairing problems. They suggest that hybridization between species would have homogenized other regions of the genome but not the inverted regions. In these

inverted regions selection would not easily eliminate alleles involved in formation of hybrid sterility and they would accumulate close to inversions (reviewed in Hoffmann & Rieseberg 2008). A similar model was also proposed by Rieseberg (2001).

#### **1.3.4 Transmission ratio distortion and reproductive isolation**

Sometimes intrinsic incompatibilities are best discovered when examining genotypic or allelic ratios of hybrid progeny (Harushima *et al.* 2001, Myburg *et al.* 2004). Transmission ratio distortion (TRD) is a statistically significant deviation from Mendelian genotype or allelic expectations. It can be observed at a genotype level in progeny of a cross between two individuals (ab x cd) carrying different alleles at one loci, the expectation is to have 25% of each genotype (ac, bc, ad and bd) in the progeny. At the allelic level it would be expected to have a and b in 1:1 ratio, as well as b and c. Observed TRD may be a sum of several factors acting at the different stages of a hybrid development (reviewed in Korbecka *et al.* 2003, Moyle & Graham 2006). TRD at genotype level can be due to factors influencing only gametophytes or only zygotes or it could be a combination of these two. To get a better understanding of at which developmental stage the genotype ratios get distorted, TRD should be examined at both the gametic and the genotypic level. This allows separating effects at the gamete formation and gametic success rates from zygotic events.

TRD at allelic (gametophyte) level may be caused by recessive genic incompatibilities that make some gametes inviable (this could be demonstrated e.g. as reduced pollen fertility). It could also emerge from an action of a segregation distorter that promotes its own transmission at the expense of the other alleles; this is also called meiotic drive (reviewed in Lyttle 1991). These segregation distorters are often silenced in their native genome but the genes controlling segregation distorters are likely to be missing from other species and meiotic drive is observed in the hybrids. The action of a segregation distorter would not necessarily be seen as reduced fertility in hybrids but only as TRD in marker segregation ratios. Segregation distorter systems are known from plants and animals, and it has been suggested that many of them go unnoticed. It is likely that we detect only segregation distorter systems that stay polymorphic within population because of negative fitness effects of the distorter (Taylor & Ingvarsson 2003).



TRD at allelic level may also be due to interactions between pollen and stigma or due to competitive interactions between gametes (Faris *et al.* 1998). Conspecific pollen precedence may act as a prezygotic species barrier (Howard 1999) but it could also cause TRD in hybrid progenies (Fishman *et al.* 2008). A single individual's gametes would compete with each other because different pollen genotypes may have different pollen tube growth rates or even different ability to germinate on stigma. It has been demonstrated that pollen tube growth rate varies even within a population (Snow & Spira 1991). In *Mimulus* hybrids conspecific pollen precedence has been examined to be the major source of the observed TRD (Fishman *et al.* 2008). In the case of *Mimulus*, conspecific pollen precedence was not directly due to different growth rates of the pollen tubes, but of the style-dependent interaction between male and female tissues (Fishman *et al.* 2008).

When TRD is seen at zygotic level it commonly has a background in genic incompatibilities, often between different loci. In *A. thaliana* it has been found that two loci epistatic interaction causes TRD and reduced fertility in RILs between European and North American accessions (Törjék *et al.* 2006). This study concluded that TRD arises because of highly reduced male fertility in individuals carrying Col-0 allele homozygous on chromosome IV and C24 alleles homozygous on chromosome V. TRD seen at zygotic level may also be present because genic incompatibilities render some zygotes inviable. They may be aborted at different stages of seed development or lose their ability to germinate or survive as a seedling. Zygotic TRD could also result from competitive differences between zygotes on maternal resources or selective embryo abortion (Korbecka *et al.* 2003).

### **1.3.5 Asymmetry in reciprocal crosses**

Hybrid progenies from reciprocal crosses frequently have differences in fertility or viability or reciprocal crosses have different fertilization success (Tiffin *et al.* 2001). If all the genes causing BDMIs were nuclear and had Mendelian inheritance, then the direction of the cross should not have an influence on the fitness of the progeny and we would not see such asymmetric results. But in animals and plants asymmetric postmating isolation is common. Tiffin *et al.* (2001) found asymmetries in all 14 plant genera they studied and in 35–45% of the studied species pairs. Bolnick and Near (2005) also found asymmetric

viability in 17 of the 20 studied species pairs in a clade of fresh water fishes (Centrarchidae).

In heterogametic species, asymmetry could arise from interactions between sex chromosomes and the nuclear genome. E.g. X chromosomes (with XY sex-determination system) from different species could carry different numbers of incompatibilities. Therefore, depending on the origin of the X chromosome of hybrid males, different reductions in male fertility or viability could occur. This type of asymmetric X-linked incompatibilities have been observed e.g. in *Drosophila* (Wu & Beckenbach 1983).

One common mechanism that can explain asymmetric differences in hybrid viability or fertility is cytonuclear interaction. These interactions occur between nuclear genes (in hybrid background) from the other species and either mitochondrial or chloroplast genes from the other species. The role of cytonuclear interactions in speciation have not been studied much but they are of increased interest (Tiffin *et al.* 2001, Levin 2003, Turelli & Moyle 2007). Cytonuclear interactions have been found in animals e.g. in spider mites (Gotoh *et al.* 1995), copepods (Ellison & Burton 2008), wasps (Ellison *et al.* 2008) and newts (Arntzen *et al.* 2009). The interaction is very likely to be between mitochondria and the nuclear genome.

It is frequent to find cytonuclear interactions in plants, especially male sterility (Laser & Lersten 1972). The best known interaction causing asymmetry in fertility of reciprocal crosses is cytoplasmic male sterility (CMS). CMS has been agriculturally used in creating male sterile lineages of crop plants (Schnable & Wise 1998). CMS is also the most common sex determination mechanism in gynodioecious species (McCauley & Olson 2008). Perhaps because of this importance in plant biology, its genetic basis has been studied in many species (maize, petunia, *Brassica*, *Raphanus* and rice; see review by Hanson & Bentolila 2004). All the CMS alleles found so far have been mitochondrial. Most often an open reading frame appears in mitochondria and the transcription of it or accumulation of the novel protein disrupts pollen production (Delph *et al.* 2007). Within a population a nuclear restorer of fertility develops and restores pollen function. CMS alleles causing male sterility can be found at many different loci, but restorers have most often been pentatricopeptide repeat (PPR) containing genes (reviewed in Hanson & Bentolila 2004, see also Barr & Fishman 2010). CMS systems have been found from hermaphroditic plants as *Mimulus guttatus* where cytoplasmic hybrid male sterility was revealed in crosses between *M. guttatus* and *M. nasutus* (Fishman & Willis 2006). The role that CMS could play

in developing isolating barrier between populations depends on many factors. One of the most important would be the way CMS developed. Selfish evolution of cytoplasm could mean benefiting female function and this could allow rapid spread in a population of the CMS and the restorer (Charlesworth 1981). This could also mean introgression of CMS cytoplasm during secondary contact because of advantage it gives to female function. Origin as a byproduct of natural selection or drift would potentially create an asymmetric barrier between populations (see Fishman & Willis 2006 for further discussion).

Different types of maternal effects could cause asymmetric viability or fertility in reciprocal crosses. Maternal effects could be either epigenetic modifications of DNA or transcripts in the developing zygote that form incompatibility with foreign alleles from the other population or species. All maternal-zygotic incompatibilities are expected to be asymmetric (Turelli & Moyle 2007).

In flowering plants, some specific aspects of the development may cause asymmetry in postzygotic isolation. Gametophyte-sporophyte (e.g. pollen-style) interactions could be different depending on which of the species acts as a pollen recipient and which acts as a pollen donor. Interactions could also be at later stage, after fertilization, when the haploid male component could have negative interaction with triploid endosperm. This could cause abortion of the seed because of not developing functional endosperm (Gutierrez-Marcos *et al.* 2003).

#### **1.4 Speciation rates and forces driving speciation**

How rapidly does reproductive isolation develop and does reproductive isolation depend on overall genetic differentiation. In plants these questions have been best studied by Moyle and colleagues (Moyle *et al.* 2004) in three different angiosperm genera. In one of the studied genera development of reproductive isolation (for one measure of prezygotic and postzygotic isolation) was positively correlated with genetic distance (Moyle *et al.* 2004), as observed in many animal groups (Coyne & Orr 1989, Presgraves 2002). In two other genera, no such relationship existed. In another study one group of orchids were found to have positive correlation between postzygotic isolation and genetic distance, although no such trend was observed in prezygotic barriers (Scopece *et al.* 2007). From these studies it is still too early to draw conclusions on trends in plants. Different groups of plants with different mating systems or other life history traits might

also show different patterns. Further studies will be needed to conclude whether postzygotic isolation and genetic distance are positively correlated in plants.

Hybrid inviability seems to develop in different rates depending on the species group. When comparing mammals and birds Fitzpatrick (2004) confirmed earlier results from the 1970's (Wilson *et al.* 1974) that hybrid inviability develops much faster in mammals than in birds. In mammals the estimated time for complete hybrid inviability would be 4 million years whereas this stage of speciation is reached on average in 21 million years in birds (Fitzpatrick 2004). Even if the evolution of hybrid inviability is slow in birds, the rate of speciation is still estimated to be similar on birds and mammals (Avice *et al.* 1998). Coyne and Orr (2004) suggest that prezygotic and extrinsic postzygotic isolation have been more important in species formation in birds than intrinsic barriers. In *Drosophila* complete postzygotic reproductive isolation has been estimated to appear in 0.2 to 2.7 million years, whereas in birds this would be attained in 10.5 million years (Coyne & Orr 2004). For plants the time to complete postzygotic reproductive isolation has not been estimated, as there are no reliable divergence date estimates for the species studied by Moyle *et al.* (2004).

Natural selection has seemed to be the most important driver of development of reproductive isolation (Coyne & Orr 2004, Schluter 2009). Sequence data from identified incompatibility genes show evidence of positive selection (e.g. Ting *et al.* 1998, Barbash *et al.* 2003, Presgraves *et al.* 2003). But it is unlikely that selection has acted directly on the traits involved in the production of reproductive isolation; more likely reproductive isolation might evolve as a byproduct of selection (Coyne & Orr 2004). Whether natural selection leading to reproductive isolation involves local adaptation is still controversial. Some of the recent evidence from development of intrinsic postzygotic hybrid incompatibilities suggests that they may evolve as a byproduct of nearly neutral or selfish genetic changes (Presgraves 2010). The genes identified to be in the background of hybrid fertility or viability problems are often involved in maintaining the inner balance of the genome or then being on the opposite side and causing unbalance, like selfish elements (reviewed in Presgraves 2010). The selfish genes might be advancing their own transmission but at the same time reduce the fitness of the host (like *Ovd* in *Drosophila*, Phadnis & Orr 2009) or these incompatibility genes might be involved in the regulation of heterochromatin (*Lhr* in *Drosophila*, Brideau *et al.* 2006).

## 1.5 The Genus *Arabidopsis*

In a revision of the genus *Arabidopsis*, the number of species was reduced from 50 to 9 (O’Kane & Al-Shehbaz 1997). Convergence in the morphological characters used for phylogenetic classification has been very high and made resolving the phylogeny difficult. Recently utilization of molecular tools has made it possible to more reliably define the species and genera.

It has been estimated that the most recent common ancestor for model species *A. thaliana* and its outcrossing relative *A. lyrata* has been 5 MYA (Koch *et al.* 2000), but more recent reassessment of fossil and sequence data has suggested more than 10 MY divergence date (Beilstein *et al.* 2010). The more recent radiation of *A. lyrata*, *A. halleri* and *A. arenosa* was estimated to have taken place during recent 2 MY, during Pleistocene, when repeated glaciations were ongoing (Koch & Matschinger 2007), but this also needs to be revised.

The species in the genus *Arabidopsis* seem to be intercrossable and introgression subsequent to speciation has been suggested between *A. lyrata* and *A. halleri* (Ramos-Onsins *et al.* 2004). Natural (*A. suecica*) and artificial hybrids between *A. thaliana* and *A. arenosa* exist (Mummenhoff & Hurka 1994, Comai *et al.* 2000). *A. thaliana* and *A. lyrata* have been proven to produce viable, but male sterile offspring (Nasrallah *et al.* 2000). These hybrids are still capable of producing seeds in backcrosses. Even more distant relative, *Pachycladon cheesemanii*, belonging to the same tribe (Camelineae) as *Arabidopsis*, has been successfully crossed with *A. thaliana*, but the hybrids were practically sterile (Heenan *et al.* 2008). For QTL mapping of heavy metal tolerance *A. halleri* and *A. lyrata* have been successfully crossed (Willems *et al.* 2007). Natural allopolyploids between these two species exist and polyploidization has happened several times independently (Shimizu-Inatsugi *et al.* 2009, Schmickl *et al.* 2010). Even natural hybrid zones between *A. arenosa* and *A. lyrata* have been described to occur in Austria (Wernisch 2007). This history of hybridization, introgression and allopolyploid species formation proves the complexity of speciation processes and ability to intercross in sympatry even after becoming taxonomically accepted true species.

### 1.5.1 *Arabidopsis lyrata* and its history

*Arabidopsis lyrata* has recently become a subject of increasing interest for molecular biologists, especially because of availability of the genomic sequence.

Several studies have found *A. lyrata* populations to be morphologically and genetically differentiated from each other (Jonsell *et al.* 1995, vanTreuren *et al.* 1997, Wright *et al.* 2003, Clauss & Mitchell-Olds 2006, Ross-Ibarra *et al.* 2008). But finding coherence in describing population history of the species has been difficult. Most of the studies indicate strong differentiation between North American ssp. *lyrata* and European ssp. *petraea* (vanTreuren *et al.* 1997, Wright *et al.* 2003, Ross-Ibarra *et al.* 2008), although some have suggested no real subdivision between subspecies (Balaña-Alcaide *et al.* 2006).

Recent research supports a lowland glacial refugium north of the Alps for *A. lyrata* (Clauss & Mitchell-Olds 2006, Koch & Matschinger 2007), from where colonization to Scandinavia could have occurred. Ansell *et al.* (2010) found that allozyme diversity is not reduced in northern European populations, but these results are not consistent with earlier studies with some of the same populations (e.g. Gaudeul *et al.* 2007). Further, the northern populations carry with high frequency chloroplast haplotype that is absent from Central European populations. Therefore they suggested that northern populations would have been established from a different source than from Central Europe. However, the northeastern Karhumäki population is much more differentiated from the Scandinavian populations than the Central European populations.

For eastern North American populations, like Mayodan from North Carolina, there has been uncertainty of the origin of dispersal. Sampling of the *A. lyrata* species complex has not properly covered these eastern parts of United States and comparative studies between European and North American populations have typically had only one or two populations from North America. Recent studies indicate that there has not been trans-Atlantic dispersal to eastern parts of Canada from European *A. lyrata* populations (Schmickl *et al.* 2008) and likely survival of *A. lyrata* in the southeast of North America during Pleistocene glaciations (Schmickl *et al.* 2010). The DNA sequence data indicate that European and North American subspecies have been isolated for at least tens of thousands of years (Ross-Ibarra *et al.* 2008).

## **1.6 Aims of the study**

The development of reproductive isolation is crucial in speciation process. Just one reproductive barrier hardly ever suffices in creating complete reproductive isolation, but usually some of them have a bigger role in the creation of reproductive isolation than others. Isolating barriers that are independent of

environment have been widely studied in animals, especially *Drosophila*. These intrinsic postzygotic barriers are important for origin of new species because they are irreversible. In plants only few studies exist on clarifying the genetic basis of postzygotic incompatibility. In this thesis the aim was to examine the development of postzygotic reproductive isolation between *A. lyrata* populations and its genetic basis. Genetic divergence between populations needed to be estimated (I) so that population pairs with different genetic distance could be formed and compared in their level of TRD. TRD indicates genomic regions that may be incompatible between crossed populations. Studies on TRD, its detailed nature and especially epistatic interactions are almost completely lacking from the plant literature. One aim was to find out how quickly genetic incompatibilities form, if they exist within populations (III) and if they would increase in number with increasing genetic divergence between populations crossed (II, V). Male and female fertility was studied by mapping associated quantitative trait loci (IV), and by observing transmission ratio distortion (TRD) in genetic markers in the F<sub>2</sub> hybrids (V). The nature of these genic incompatibilities was of interest; at which developmental stage they would affect, would they act epistatically and would the interactions be between nuclear genes or also between nuclear and cytoplasmic genes. This was carried out by close examination of TRD in different crosses (V).





## 2 Materials and methods

Only short description of materials and methods used are provided here. More detailed description can be found in the original papers (I-V).

### 2.1 Plant material and crosses

*Arabidopsis lyrata* occupies low-competition habitats in its circumpolar distribution range (Jalas & Suominen 1994, Hoffmann 2005). The distribution is best known for the European subspecies *petraea* which is also the most widely studied of the subspecies. *A. lyrata* ssp. *petraea* has fragmented distribution in central Europe (Germany, Austria, Czech Republic and Hungary) and northern Europe (Sweden, Norway, Iceland and British Isles). The fragmented distribution continues into Russian Karelia, Kola Peninsula and northern Russia. *A. lyrata* ssp. *lyrata* is found from the east coast of North America to the Great Lakes area. The third subspecies *A. lyrata* ssp. *kamchatica* is distributed around eastern Russia, Alaska and western North America, and has recently been discovered to be an allopolyploid between *A. lyrata* and *A. halleri* and is therefore treated as its own species, *A. kamchatica* (Shimizu *et al.* 2005). The seeds used in the experiments described in this thesis, were collected from natural populations of *A. lyrata* ssp. *petraea* and ssp. *lyrata*. For the study of population history (paper I) plants were grown from seeds collected from seven European ssp. *petraea* populations and one North American ssp. *lyrata* population. Between 11 and 30 plants were genotyped per population. The European populations sampled contained both central European, Russian and Scandinavian populations to get an overview of the relationships of the populations and the consequences of the Last Glacial Maximum and the following colonization of Scandinavia.

For creating a genetic linkage map of *A. lyrata* (paper II) two plants from a Swedish population and two plants from a Russian population were crossed pairwise. From there two F<sub>1</sub> individuals were crossed reciprocally to produce the F<sub>2</sub> mapping progenies with different cytoplasm (for crossing design see paper V, Figure 1).

In the study on TRD within population (paper III) eight full-sib families, including from 26 to 70 offspring were used. The parents for initial crosses were collected from Reykjavik, Iceland and the full-sib families were studied earlier by Bechsgaard *et al.* (2004).

For studies on reproductive isolation (papers IV and V) the same crossing design was used as in paper II. For the study on male and female fertility (paper IV) a cross between *ssp. lyrata* (Mayodan, North Carolina, USA) and *ssp. petraea* (Spiterstulen, Norway) was used. The same cross was also examined for marker transmission ratios in paper V. In addition to this cross, TRD in F<sub>2</sub> progeny between two Scandinavian populations (Spiterstulen, Norway and Stubbsand, Sweden) and between a Scandinavian (Spiterstulen) and a Central European (Plech, Germany) population was studied (paper V). To make comparisons between crosses more feasible, one parental population was common in all crosses (Spiterstulen, Norway).

## **2.2 Phenotyping**

In paper IV the population plants, F<sub>1</sub> and F<sub>2</sub> were grown in a greenhouse. Their phenotypes were studied by collecting pollen samples for estimation of male fertility and conducting controlled crosses to estimate female fertility. Pollen samples were dyed with lactophenol-aniline blue (Kearns & Inouye 1993) to enable distinguishing between viable and non-viable pollen. Seeds from crosses were counted and classified into viable or aborted.

## **2.3 Molecular methods**

### **2.3.1 Microsatellite variation**

Microsatellite markers were used in all papers. Microsatellites consist of short repeat motifs, usually of 1 to 6 bases, repeated in tandem. These repeats are usually in intergenic area or in introns of genes and are highly variable within species. Most of the microsatellites used in this thesis have been published earlier (Bell & Ecker 1994, Ponce *et al.* 1999, Clauss *et al.* 2002, Loudet *et al.* 2002) and are derived from *A. thaliana*. Therefore it is feasible that sequence (e.g. structure and number of repeats) of the microsatellites is different in *A. lyrata*, as studied for some of *A. thaliana* derived microsatellites (van Treuren *et al.* 1997). The genetic basis of the observed variation was not important in mapping or TRD studies but could have had an influence on the results in paper I and was studied further there. New microsatellite markers were developed in papers II, IV and V. For paper II the primers were developed based on *A. thaliana* sequences and for

IV and V *A. lyrata* sequence was used. Dinucleotide repeats were searched from the genomic sequences and when long enough repeats were found, primers amplifying 150 to 350 base pairs were designed and tested for polymorphism with F<sub>1</sub> parents of crosses.

In principal the lab procedure for microsatellites was the same in all the experiments. The fragments were amplified with PCR, the quality was checked by a run on an agarose gel, and the labelled PCR fragments were diluted and multiplexed for a fragment analysis run, where differently sized PCR products could be separated. The results from these runs were analyzed and the sizes of the fragments could be identified. The details on the genotyping procedure on specific experiments are described in individual papers.

### **2.3.2 Gene-based markers**

For genetic linkage maps (II, IV, V) gene based markers were developed because their location in *A. lyrata* genome could be estimated based on their location on *A. thaliana*. Before the availability of the *A. lyrata* genome sequence, in paper II, the PCR primers were designed on exons of protein coding genes based on *A. thaliana* sequences. In papers IV and V most of the primers were designed with *A. lyrata* sequence on exons based on *A. thaliana* sequence. Nucleotide sequences of the F<sub>1</sub> parents were acquired by direct sequencing of the PCR products. The sequences were aligned and the chromatograms of polymorphic SNPs were manually checked.

### **CAPS markers**

All the polymorphic sites in the F<sub>1</sub> (II, IV, V) were tested for carrying a restriction endonuclease cutting site. If a cutting site was found in a polymorphic region and a restriction endonuclease would differently cut the two alleles, the PCR products from the F<sub>2</sub> progeny were digested with the endonuclease in question and run on an agarose gel to separate the differently sized fragments. The F<sub>2</sub> genotypes were scored visually from the gel.

### **SNP genotyping**

Surrounding the polymorphic SNPs in the F<sub>1</sub> plants (IV, V) primers were designed in Finnish Genome Center and Sequenom MALDI-TOF (matrix assisted laser

desorption-ionization - Time-of-Flight mass spectrometry) genotyping was performed for the  $F_2$  progenies (Leushner & Chiu 2000). For MALDI-TOF genotyping multiplexed PCR reaction is performed, such that short regions surrounding the selected SNPs are amplified. Then the PCR is cleaned and a multiplexed single nucleotide extension reaction can be performed with mass modified dideoxynucleotides and primers which are designed to anneal their 3' end right next to the selected SNP. These extended primers are then analysed with Sequenom MALDI-TOF and as a result it is possible to recognize which nucleotide has been added in the extension reaction.

## **2.4 Statistical methods**

### **2.4.1 Population differentiation**

Gene diversity within populations and in the whole data set of paper I was counted with program GENETIX (Belkhir *et al.* 2001). Because sample sizes varied between different populations standardization for number of different alleles per population and for number of private alleles was conducted with CONTRIB (Petit *et al.* 1998) and HP-RARE (Kalinowski, 2005), respectively. Deviations from mutation-drift equilibrium were tested with BOTTLENECK (Piry *et al.* 1999). Population differentiation was estimated with Wright's  $F_{ST}$  in GENEPOP (Raymond & Rousset 1995). Because  $F_{ST}$  is highly dependent on within population gene diversity ( $H_s$ ), a standardized  $F'_{ST}$  value was calculated (according to Hedrick 2005) to be compared with  $F_{ST}$ . To further understand the relationship between populations neighbor-joining trees were built in Phylip (Felsenstein 2005).

### **2.4.2 Genetic linkage map construction**

Genetic linkage maps were constructed with software JoinMap 3.0 (Van Ooijen & Voorrips 2001) in papers II and IV. JoinMap can build a linkage map from an outbreeding full-sib family. The program automatically determines the linkage phases of the markers during the estimation of the recombination frequencies. The mapping starts with the most informative pair of markers and proceeds by adding markers one by one. The mapping is conducted in three rounds where the best position for each marker in a linkage group is searched for by comparing

goodness-of-fit of each marker in calculated maps. In the third round all the loci that were removed from the map on the earlier rounds are fitted to the map.

### **2.4.3 QTL mapping**

In paper IV QTL mapping was performed for male and female reproductive traits in reciprocal  $F_2$  progenies. Several free and commercial programs are available for QTL mapping. Heterogeneity of the  $F_1$  parents in our crosses limits the number of available programs to use. R/qtl (Broman *et al.* 2003) is a software package that is freely available and is run on R (R Development Core Team 2010) which also is a freely available environment for statistical computing and graphics. In addition to a wide range of cross types, R/qtl can handle normal, non-normal and binary phenotypic data. Interval mapping can be performed with single, two or multiple QTL-model in R/qtl. In two-QTL analysis different models can be tested and the best model describing the action of the QTLs, (single-QTL model, additive or epistatic two-QTL model), can be defined. R/qtl enables testing if reciprocal crosses are different in a QTL of interest.

### **2.4.4 Analysis of transmission ratio distortion**

TRD was analysed in within population crosses in paper III and in between population crosses in paper V. In paper III a maximum likelihood framework was adopted for analysing TRD (Bechsgaard *et al.* 2004). The marker transmission ratios were studied on genotype, gametic and zygotic level. The analyses were conducted at a single marker level.

In paper V TRD was also analysed at the gametic and the zygotic level, but variable information content of the markers was improved by inferring the four different genotypes for all loci, either by assigning the most likely genotype based on the surrounding markers (in the analysis of two-locus epistasis) or with surrounding markers and by assigning the most likely genotypes to maximize the sum of the likelihood in an equation (paper V, equation 1). The goodness of fit was tested for observed and expected frequencies under null (no TRD), gametic and zygotic model. Likelihood ratio tests were applied to evaluate which of the models best described the observed data. To get a better estimate for the locations of the transmission ratio distorting loci, pseudomarkers were used between the genotyped molecular markers.

Because the most likely genotypes were inferred for each F<sub>2</sub> individual at the genetic marker loci and pseudomarkers, it was possible to obtain an estimate for the F<sub>1</sub> allele transmission, by calculating for every locus how many times each allele was found from the F<sub>2</sub> progeny. Therefore it could be seen which of the F<sub>1</sub> parents (or both) was causing the observed TRD when the gametic model was best explaining the observed F<sub>2</sub> genotypes. It could also be observed which of the alleles was under- or overrepresented compared to expected 1:1 ratio.

TRD may not be simply caused by single, independent loci, but may be due to epistatic interactions between two or more loci. Therefore two locus interactions were also analyzed in paper V by calculating two-locus genotype frequencies, their expected frequencies and compared them with  $\chi^2$ -tests, for all pairs of markers. This was also conducted at gametic level by calculating the observed and expected two-locus allelic combinations for both F<sub>1</sub>-parents separately. From F<sub>2</sub> genotype data it was impossible to continue to explore whether epistatic TRD was due to solely gametic or both gametic and zygotic factors.

## 3 Results and discussion

### 3.1 Evolutionary history of *Arabidopsis lyrata*

A microsatellite study comprising seven European *A. lyrata* ssp. *petraea* populations and one North American ssp. *lyrata* population (paper I) revealed high between population differentiation as well as highly variable within population diversity among studied populations. Gene diversity ( $H_e$ ) was highest in the two Central European populations ( $H_e$  close to 0.55) and lowest in the Russian and North American populations ( $H_e$  close to 0.2). The other summary statistics for within population diversity also describe how most of the variation is found in the Central European populations and how diversity is lower in the north (paper I, table 4). These variable levels of within population diversity might partially reflect differences in the current population sizes but also the recolonization events during and after ice ages when *A. lyrata* spread over new areas after retreating ice sheet. The Central European populations have likely survived in their habitats, or at least the refugia had been in Central Europe (North of the Alps) as suggested by Clauss and Mitchell-Olds (2006). The northern populations have experienced bottlenecks during the ice age and when colonizing the area that was covered by ice. The smaller population sizes are reflected in the reduction of diversity in the north. Our populations were collected in Russia, Sweden, Norway and Iceland, all of which were covered by ice sheets during the Last Glacial Maximum (Clark & Mix 2002). The microsatellite study did not describe clear bottlenecks in the Icelandic and Scandinavian populations which indicate that the populations have colonized Iceland and Scandinavia quickly after the ice retreated so that the microsatellites have had time to return to the mutation-drift equilibrium. A microsatellite study focusing on Norwegian and Swedish populations also found little evidence of recent bottlenecks (Gaudeul *et al.* 2007).

As *A. lyrata* is cold-tolerant and thrives in low-competition areas, it could have had refugia even closer to the ice sheet than in Central Europe, as suggested e.g. by Koch & Matschinger (2007) and Ansell *et al.* (2010). To support their hypothesis, Ansell *et al.* (2010) provided allozyme and chloroplast DNA sequence data where the decrease in genetic diversity towards north, as in our data, was not detected. They also found common chloroplast haplotype for populations from the British Isles and Scandinavia which was absent from Central Europe.

Gaudeul *et al.* (2007) had more extensive population sampling from Norway and Sweden than Ansell *et al.* (2010), and when they compared genotype data from 12 common microsatellites with data from Central Europe (Clauss & Mitchell-Olds 2006), the lower genetic diversity in Scandinavia was clear ( $H_e = 0.27$ , Scandinavia vs.  $H_e = 0.52$ , Central Europe). The population sampling of Scandinavia in Ansell *et al.* (2010) included one Swedish and one Norwegian population shared with Gaudeul *et al.* (2007). The gene diversity ( $H_e$ ) in these two populations was only slightly higher in the microsatellite data (Gaudeul *et al.* 2007) than in the allozyme data (Ansell *et al.* 2010), which suggests that finding similar levels of diversity in north and south, by Ansell *et al.* (2010), could have been due to low gene diversity in their Central European samples. These differences between the conclusions of our study and the study by Ansell *et al.* (2010) may be partly due to different behavior of allozymes and microsatellites as population genetic markers. For microsatellites a repeat number mutation is not usually size restricted (in a sense that there would be an upper or lower limit) and will often result a new allele. In contrast, allozymes are functional proteins, which might be restricted in the possible numbers of amino acid substitutions that can occur without detrimentally affecting the functionality of the protein. But final explanation for the differences between the studies cannot be concluded from the available data. It would be interesting to have some of the populations collected by Ansell *et al.* (2010) to be genotyped with the same set of microsatellites than in the paper I.

The high genetic population structure, as observed in paper I and by several other authors: Jonsell *et al.* 1995, vanTreuren *et al.* 1997, Wright *et al.* 2003, Ross-Ibarra *et al.* 2008, gives further support to the relatively long separate histories for the populations. The estimates for pairwise  $F_{ST}$  varied between 0.072 (two geographically close Norwegian populations) to 0.728 (between Russian and North American population). For between subspecies comparison we had only one North American population representing ssp. *lyrata*. Pairwise  $F_{ST}$  values between this population and the European populations were high; a comparison between Mayodan and the Central European populations gave the lowest values (close to 0.5) whereas in comparisons with Icelandic or Scandinavian populations  $F_{ST}$  was higher than 0.6. The lower pairwise  $F_{ST}$  values between North American and Central European (high gene diversity) populations than between North American and Scandinavian (low gene diversity) were likely due to  $F_{ST}$ 's dependence on within population gene diversity: within population gene diversity ( $H_S$ ) was negatively correlated with  $F_{ST}$  values (paper I, Figure 4). Because of



differences in the level of diversity between populations, the  $F_{ST}$  values were a bit hard to interpret, but when standardized  $F'_{ST}$  values were calculated (paper I, Figure 4) it could be seen that small population size (some Scandinavian and Russian populations) did not always lead into high pairwise  $F'_{ST}$ . Besides the genetic divergence found on this study, the populations have been demonstrated to be morphologically different and locally adapted to their environment (Jonsell *et al.* 1995, Riihimäki & Savolainen 2004, Leinonen *et al.* 2010).

European populations of *A. lyrata* have been more extensively studied than North American (Clauss & Koch 2006) and only one study has included samples from several Russian populations (Schmickl *et al.* 2010). This recent paper pursues to describe evolutionary history of *A. lyrata* species complex by studying sequence variation in nuclear and chloroplast markers over the circumpolar distribution of the species. Shimizu *et al.* (2005) and Shimizu-Inatsugi *et al.* (2009) had found that allotetraploid hybrid species had formed in Eastern Asia and North America several times independently by hybridization event between *A. halleri* and *A. lyrata*. This allotetraploid species was formerly described as a subspecies of *A. lyrata* but is now acknowledged to its own species *A. kamchatica* (Shimizu *et al.* 2005). Besides this clearly separate lineage, *A. kamchatica*, Schmickl *et al.* (2010) identified two clearly separate lineages of *A. lyrata* – the Eurasian and North American. The Eurasian subspecies *petraea* (including studied Russian populations) had higher genetic diversity than the North American subspecies *lyrata*. Our microsatellite data indicated high divergence between these subspecies and Schmickl *et al.* (2010) further confirmed the long-term isolation and likely dispersal from Europe to North America through Beringia. This dispersal route would mean that the Scandinavian and North American populations, studied in paper IV and V, could have been allopatric for tens to hundreds of thousands of years.

### **3.2 Transmission ratio distortion in within and between population crosses**

In paper III TRD within population was examined in eight Icelandic full-sib families and compared with observed TRD of a cross between two *A. lyrata* ssp. *petraea* populations (paper II). Only TRD at gametic level was higher than expected by chance in the within population crosses, although it was still only half of that observed in the between population cross (11% of the tests showed significant departure from the Mendelian segregation in within population crosses

vs. 19% of the tests in between population cross). The observation of some gametic TRD in within population crosses could be due to meiotic drive or gametic competition (e.g. variation in pollen tube growth rates within population) or inbreeding depression, although the parents of the crosses were not likely to be closely related. Inbreeding depression was unlikely a source of TRD in between population crosses (papers II and V) because of highly outcrossing crossing design.

Especially TRD at zygotic level was increased in the between population cross compared with the within population crosses (3% of the tests in within population vs. 11% in between population cross). The comparison between within population and between populations crosses overall showed a higher level of transmission ratio distorted genomic regions in between population cross. This clearly indicates that TRD is not a common phenomenon within population, but more likely appears when genetically more distant individuals (e.g. from different populations or species) are crossed. This could be due to genetic incompatibilities (BDMIs) that accumulate over time due to drift or natural selection. TRD observed e.g. in rice (Harushima *et al.* 2001), *Eucalyptus* (Myburg *et al.* 2004) and tomato (Moyle & Graham 2006) have been interpreted to be due to selection against certain heterospecific allele combinations. TRD has also been detected in other within species (but between different populations) crosses, e.g. from *Mimulus guttatus* (Hall & Willis 2005) and *Silene vulgaris* (Bratteler *et al.* 2006). In both cases genomic divergence was estimated to play a role in the formation of TRD.

In within population crosses (paper III), segregation ratios of the reciprocal crosses were similar, as only 4% of the tests showed different patterns of segregation between reciprocal crosses. In contrast, the reciprocal crosses were quite different from each other in the between population cross (paper II and III). Between 15% and 19% of the tests showed a significant difference in the segregation pattern between reciprocal crosses, depending on whether tested on the gametic or zygotic level (respectively). These higher differences in between population cross could be due to different origin of cytoplasm in the reciprocal progenies. Reciprocal crosses with cytoplasm from different populations also showed differences in the number of TRD regions observed, as well as in locations of the TRD regions (paper V). In the cross between a Russian and a Swedish population (paper II) only two of the eight mapped TRD-loci were in the same chromosomal region and the six other TRD-regions were seen in only one of the reciprocal progenies. The roles of cytonuclear interactions or epigenetic

mechanisms cannot be distinguished in the formation of TRD. When the cytoplasm was from the same population in both reciprocal crosses, the reciprocal  $F_2$  progenies displayed less variation in the intensity of TRD than when the cytoplasm came from a different population (Sp x Pl vs. Sp x Stu and Sp x Ma, Figure 4, paper V). This supports a prominent role of cytoplasm in the emergence of TRD. The important role of cytoplasm in reciprocal crosses, and plant speciation, has been acknowledged in literature (Tiffin *et al.* 2001, Levin 2003).

In paper V the focus was to study in more detail TRD in crosses between populations of *A. lyrata*. The results showed that all three crosses had transmission ratio distorted regions. When comparing the TRD regions between the different crosses (papers II and V), hardly any common regions were found. This implies that if TRD was due to negative heterospecific interactions between loci, the same loci were not likely interacting between populations, even if the same population was crossed with several other populations (as Spiterstulen in paper V). An interesting observation was that chromosomes AL1 and AL6 experienced strong TRD in all of the crosses in paper V, they also had TRD in the cross of paper II and TRD was found on the same chromosomes from a backcross progeny between *A. halleri* and *A. lyrata* (Willems *et al.* 2007). The loci involved in causing TRD were unlikely the same between different crosses because TRD was strongest in different parts of the chromosomes. It is not known why these two chromosomes would be more likely to harbour incompatibilities than the other chromosomes.

Our expectation was to find more TRD (also hybrid incompatibility) with higher genetic divergence between the parents crossed. Actually the number of genic incompatibilities between species has been estimated to increase faster than linearly in time (Orr 1995). Recent studies in *Drosophila* and tomato support this model (Matute *et al.* 2010, Moyle & Nakazato 2010). We found an increase in the proportion of genome in TRD with genetic divergence between the crossed populations (Figure 6, paper V), although with the restricted number of crosses it would be hard to draw firm conclusions on whether the increase was linear or faster than linear. The difference between the two genetically close crosses (Sp x Pl and Sp x Stu) was not very clear, because the level of TRD varied much between the two reciprocal progenies in the Sp x Stu –cross.

In all three crosses (paper V) TRD was mostly found at the gametic level, both at single-locus and two-locus analyses. In only one region zygotic single-locus model gained more support over gametic model for being the stage where TRD occurred (AL6, Sp x Ma –cross). Surprisingly, the only zygotic epistatic

interaction was found in the Sp x Stu –cross (the genetically closest pair crossed) where AL1 and AL6 were interacting. Recessive zygotic BDM-type of incompatibility was observed, as within population homozygotes were interacting negatively (underrepresented) between within population homozygotes of the other population at the other chromosome.

Most of the TRD appeared at the gametic level in the between population crosses (paper V), which supports the idea that F<sub>2</sub> hybrid viability is not strongly reduced in the hybrids of *A. lyrata*. The interactions leading to TRD in F<sub>2</sub> take place mainly already when F<sub>1</sub> hybrids produce gametes. Alternatively, some TRD could occur due to interactions between pollen and pistil in the F<sub>1</sub>, but mostly before F<sub>2</sub> fertilization. The viability of F<sub>2</sub> hybrids was not observed to be affected in the growth chamber experiment (paper IV) and field experiments show heterosis in Sp x Ma –F<sub>2</sub> progeny's fitness in Spiterstulen, Norway (Leinonen *et al.* 2010). More specifically, we also examined whether the TRD arose in one or both of the F<sub>1</sub> parents, and whether it was observed in both reciprocal crosses. We found two examples of sex-independent TRD (Sp x Ma AL1, Sp x Stu AL1), where both parents were transmitting their alleles to the progeny in unequal proportions independent of whether they acted as pollen donors or recipients. However, in Sp x Ma –cross the effect was stronger in the pollen donors. Strong sex-independent transmission ratio distorter is well known from rice, where the background has been identified to be a single locus with allelic interactions (Koide *et al.* 2008). Even though both of these TRD regions appeared in AL1 they were not caused by the same system; in Sp x Ma –cross region in TRD was in the upper arm of the chromosome and showed an excess of transmission of the Sp allele from both of the F<sub>1</sub> parents, whereas in the Sp x Stu –cross the TRD region was in the lower arm of the chromosome and the parents transmitted Sp allele in deficiency.

In the Sp x Pl –cross we found that the strongest TRD came from a single F<sub>1</sub> transmitting (as a pollen donor and recipient) its Sp –alleles in deficiency. The same chromosomal region was observed to be involved in an epistatic interaction with chromosome AL3 and this interaction was again observed only in this one F<sub>1</sub> –parent. The epistatic interaction was stronger when the F<sub>1</sub> acted as a pollen recipient but it was also visible when the same parent acted as a pollen donor. Gametes carrying alleles from the same population in these two loci were overrepresented and heterozygous combinations were less frequent than expected. A similar interaction was seen in the Sp x Stu –cross (AL2 and AL4), but only when the F<sub>1</sub> parent with Stu –cytoplasm was acting as a pollen donor. These

observations could be due to selection against recombinant gametes, with allelic variation within population (in Sp x Pl – and Sp x Stu –cross distorted segregation ratio observed in only one parent) and involvement of cytoplasmic or epigenetic factors in Sp x Stu –cross.

The most common type of TRD observed in Sp x Ma -cross was single- or two-locus TRD which acted in only one of the reciprocal crosses and was due to segregation distortion in the gametes of one F<sub>1</sub> parent. Because it acted only in one of the F<sub>1</sub> parents, this indicated that the F<sub>1</sub> parents had different alleles i.e. polymorphism within population. It was not possible to distinguish whether the TRD was dependent on sex or cytoplasmic factors. Because reduced male fertility was observed in the F<sub>2</sub> progeny in this cross (paper IV), some of the TRD could be due to different viability of pollen grains of different genotype. Polymorphism within population was also observed in the QTLs for reduced male fertility (see 3.4 for detailed discussion). Most of the TRD in the Sp x Ma –cross was observed in a pollen donor, which suggests differential viability or fertilization success of the pollen grains. In *Arabidopsis thaliana* TRD was found to be associated with reduced male fertility (Törjék *et al.* 2006). Similarly the co-located male fertility QTLs and TRD regions could be in association in *A. lyrata* hybrids, but it would require further studies to identify this. However, not all TRD regions observed in Sp x Ma –cross co-located with male fertility QTLs. The results from paper V reveal the complexity of TRD in *A. lyrata* crosses. Clearly TRD has different genetic basis in different crosses and interactions between nuclear loci and between nuclear and cytoplasmic loci might be involved. We could examine only two-locus interactions whereas more complicated interactions could occur and not be detected (at single- or two-locus level) because they would not necessarily distort segregation ratios as severely as simpler interactions. The genetic background of TRD seen in F<sub>2</sub> hybrids between species (Fishman *et al.* 2001) has been well clarified only in *Mimulus* (Fishman *et al.* 2008). They found that interactions were not between heterospecific allele combinations but TRD was caused by polygenic conspecific pollen precedence, which explained most of the TRD seen in the interspecific F<sub>2</sub> hybrids. The two *Mimulus* species differ in their mating system (outcrossing vs. selfing) which might have lead to sexual selection and coevolution between male and female reproductive genes (Fishman *et al.* 2008). In *A. lyrata* all populations of these studies (papers II, III and V) are mainly outcrossing, which would make the coevolution of male and female reproductive genes equally likely in all populations.

### 3.3 Reproductive isolation in *Arabidopsis lyrata*

In addition to the findings of distorted marker transmission ratios in crosses between populations, further evidence for emergence of reproductive barriers comes from the male and female fertility studies (paper IV). Only male fertility was observed to be reduced in the F<sub>2</sub> progeny of a cross between subspecies (ssp. *petraea* Spiterstulen and ssp. *lyrata* Mayodan) compared to the parental populations. But the reduction in male fertility had occurred in two ways: 17% of the F<sub>2</sub> plants with Spiterstulen cytoplasm were male sterile and regardless of cytoplasm F<sub>2</sub> plants had reduced pollen fertility compared to the parental populations (when the male sterile individuals were excluded from the analysis). The parental populations had a high proportion of fertile pollen (<94%) and the reduction in male fertility was more severe in the F<sub>2</sub> progenies than in F<sub>1</sub>. The F<sub>2</sub> reciprocal crosses differed significantly from each other in proportion of fertile pollen grains (MaSpF<sub>2</sub> = 0.837 vs. SpMaF<sub>2</sub> = 0.736, without male steriles).

The other aspect of male fertility measured was the number of pollen grains produced. Mayodan population produced almost twice as many pollen grains than Spiterstulen, and MaSpF<sub>1</sub> and MaSpF<sub>2</sub> were very close to the mid-parent value. The F<sub>1</sub> and F<sub>2</sub> hybrids with Spiterstulen cytoplasm resembled Spiterstulen parent and produced significantly less pollen grains than the mid-parent value or the reciprocal F<sub>2</sub> progeny. This could be due to nuclear-cytoplasmic incompatibilities or maternal effects. Because Spiterstulen population plants produced the lowest number of pollen grains, this might not be describing hybrid weakness but rather how number of pollen grains produced in a flower could depend on the type of cytoplasm or e.g. maternal effects.

The male sterile individuals were observed only in the F<sub>2</sub> with cytoplasm from Spiterstulen. This could be a case of cytoplasmic male sterility (CMS), which is often seen in plants (Schnable & Wise 1998). It becomes visible when related species are crossed e.g. for agricultural purposes. Another well-known biological phenomenon which involves CMS is gynodioecy. A plant population has hermaphroditic and female plants, and the sex is determined by the cytoplasm and genomic restorer/s of fertility. The CMS observed both in distant crosses between hermaphrodites and in the sex determination of gynodioecious species, is due to an interaction between mitochondria and nuclear genes (reviewed in Schnable & Wise 1998, Hanson & Bentolila 2004). CMS has been rarely observed to result from a cross between two hermaphrodite natural populations of plants. Aside from our findings, earlier work by Fishman & Willis (2006) on

*Mimulus* suggests that mitochondrial haplotypes that cause male sterility also appear in fully hermaphroditic species in nature. In *Mimulus guttatus* the male sterility causing mitochondrial type had been fixed in one population of the 35 studied populations (Case & Willis 2008). In addition to this one CMS type, another mitochondrial rearrangement and cytoplasmic male sterility in interspecific hybrids has been found in *M. guttatus* (Martin & Willis 2010). The observations in our *A. lyrata* cross revealed that CMS had developed in the lineage leading to Spiterstulen population. We do not know how widely the CMS mitotype has spread in Norway or even how common it is within this one population. We did collect bud samples from Spiterstulen natural population (32 plants, 2 buds per plant) and found one individual which produced only inviable pollen grains. In SpMaF<sub>2</sub> the male sterile plants produced hardly any pollen grains whereas this field collected plant produced shrivelled pollen grains. This could be due to different nuclear genetic background of the plant compared to the hybrid F<sub>2</sub>. The finding of this male sterile plant from nature reveals that nuclear fertility restorers have not been fixed within Spiterstulen population. If the CMS mutation increases fitness through female function, then it is likely to spread to fixation (Charlesworth 1981). After development of male fertility restorer(s), the population could either fix the restorer allele(s) and maintain a hermaphroditic state or develop stable dynamics between females and hermaphrodites (gynodioecy). Because female plants are not commonly observed in the Spiterstulen population, the CMS mutation might be of recent origin and still polymorphic, or develop into gynodioecy. Or perhaps the male sterility does not increase female fitness sufficiently and could eventually disappear because of this deleterious effect on male function.

Whether or not CMS would create a long-term barrier to introgression between Spiterstulen and Mayodan (if they could come to secondary contact) cannot be answered without further experiments. The outcome would depend on the evolutionary history of CMS; whether it arose as a by-product of neutral or selective divergence or via selfish cytoplasmic evolution. Fishman and Willis (2006) hypothesised that neutral or selective divergence would lead to longer-term barrier than selfish evolution which could even accelerate introgression because of the female fitness advantage of the CMS cytoplasm. It is acknowledged that CMS is commonly observed in hybrids between species and it may play a role in the formation of reproductive isolation and plant speciation, but its significance as a reproductive barrier is doubtful (Bomblies 2010, Rieseberg & Blackman 2010).

The male fertility reduction that was observed in both reciprocal crosses is likely caused by nuclear genes interacting negatively with genes from a different population (discussed in more detail in 3.4). This type of BDM-incompatibility would reduce introgression of genes in secondary contact but the  $F_1$  individuals were viable and relatively fertile, so the barriers could be formed only between certain regions of the genome. The other regions of the genome would likely freely move from one subspecies to another.

Interestingly, only male fertility was clearly reduced in these *A. lyrata* hybrids, not female fertility or hybrid viability. Both male and female fertility have been reduced in hybrids of many other plant species (e.g. *Mimulus* Fishman & Willis 2001, *Helianthus* Lai *et al.* 2005). However, when Lowry *et al.* (2008) compiled data from 19 flowering plant systems, they found that  $F_1$  male fertility acted as a stronger reproductive isolating barrier than  $F_1$  seed set, which was highly variable depending on the crossed plant species (Lowry *et al.* 2008, Figure 1). The lack of viability and female fertility problems in these hybrids between *A. lyrata* subspecies might be due to recent divergence, consistent with a very early stage of postzygotic reproductive isolation.

### **3.4 Genetics of reduced hybrid male fertility and cytoplasmic male sterility**

The genetic background of reduced male fertility was different in the reciprocal  $F_2$  progenies examined in paper IV. For SpMa $F_2$  four QTLs for proportion of fertile pollen were found and for the reciprocal cross with Ma cytoplasm two QTLs were found (paper IV, Figure 3). The two QTLs in MaSp $F_2$  were acting epistatically (paper IV, Figure 5), whereas the other QTLs were not found to act interactively. One of the QTLs was common between the reciprocal progenies but all other QTLs were in different chromosomal locations. Either cytoplasmic factors or maternal effects must be involved in formation of reduced hybrid male fertility. For number of pollen grains produced no QTLs were found in MaSp $F_2$  and in SpMa $F_2$ , except when male sterile plants were included and then the QTL was located in the same region (AT1G62520) as the restorer of cytoplasmic male sterility, which was mapped as a binary trait (phenotype coded either as sterile or fertile).

Most of the QTLs for pollen quality shared an interesting characteristic: the two different heterozygotes for Ma and Sp alleles were phenotypically different from each other in the  $F_2$  progenies (paper IV, Figure 4). Within a reciprocal cross



these heterozygotes differed genotypically because one carried Sp allele from a pollen donor and Ma allele from a pollen recipient and the other way round. Because one of the homozygotes was also phenotypically similar to the worse heterozygote, this effect in the QTLs seemed to come from a single allele. The polymorphism in hybrid incompatibility alleles within the population suggests that these loci have not been under so strong directional selection to have become fixed. It rather implies that the variation within population might be neutral or due to balancing selection. This has been suggested earlier for the variation found for hybrid sterility in *Chorthippus* grasshoppers (Shuker *et al.* 2005) and *Drosophila* flies (Reed & Markow 2004). In plants studies on monkey-flowers (*Mimulus*) polymorphism for hybrid male sterility was seen in several populations (Martin & Willis 2010).

The number of QTLs observed for male fertility reduction (proportion of fertile pollen grains) was relatively low, although in plant literature the number of QTLs found for hybrid fertility have been lower than in animals, especially *Drosophila* (Moyle & Graham 2005). We might not have detected all the QTLs involved in reduced pollen fertility because of the inherent bias in QTL studies for over-estimating the effects but under-estimating the number of QTLs (Beavis 1994). Low numbers of hybrid sterility QTLs have been found in addition to tomato (Moyle & Graham 2005, Moyle & Nakazato 2008), also in rice (Song *et al.* 2005) and sunflower (Lai *et al.* 2005). Contrasting to these studies, we did not find reduced female fertility, a result that resembles the observations from *Drosophila* (Wu *et al.* 1996, Tao *et al.* 2003).

Of the F<sub>2</sub> hybrids carrying Sp-cytoplasm, 17% showed male sterility. Cytoplasmic male sterility is well known phenomenon and its genetic basis is known from many species and shows characteristic patterns. Commonly an open reading frame (ORF) appears in mitochondrial DNA and the expression of this ORF somehow disrupts pollen production (Hanson & Bentolila 2004). The plant mitochondrion has a multipartite structure so that within one mitochondrial haplotype there are multiple copies of the genome as master molecules, mini-molecules and sublimons (reviewed in McCauley & Olson 2008). This makes intragenomic recombination common and the structure, gene order, of the mitochondrial genome extreme variable. Rearrangements have been commonly observed in mitochondria and these could favour the generation of the mitochondrial counterpart of CMS, an open reading frame, which would then be expressed and disrupt the pollen or stamen production (reviewed in Chase 2007).

The normal function of the mitochondria may be restored by a nuclear gene or genes. All the restorer genes identified so far come from the same gene family, pentatricopeptide-repeat (PPR) genes (Hanson & Bentolila 2004, Chase 2007). In paper IV we map the male fertility restorer in the chromosome AL2, close to marker AT1G62520. In *A. thaliana* the only cluster of PPR genes (Lurin *et al.* 2004) locates in a syntenic region with our QTL for restorer locus. These clustered PPR genes in *A. thaliana* are the closest homologs for CMS restorers in petunia, radish and rice (Lurin *et al.* 2004). Because the gene order is well shared in this region between *A. thaliana* and *A. lyrata* this implies that male sterility observed in this study is likely restored in a similar manner in *A. lyrata* as in other species – with PPR genes. This is a second case where CMS restoration has been observed to happen in the same manner in natural, hermaphroditic populations as with cultivated and gynodioecious plants. The first genetic identification for CMS in natural populations and with hermaphroditic plants was in *Mimulus*, where fertility restorer also mapped to a cluster of PPR genes (Barr & Fishman 2010).

## 4 Conclusions

All the papers in this thesis support the emerging picture of genetically differentiated populations of *Arabidopsis lyrata*, where intrinsic postzygotic reproductive isolation has begun to accumulate between populations in the form of genic incompatibilities. Before these studies species *Arabidopsis lyrata*, and its subspecies *lyrata* and *petraea*, were thought to be fully interfertile (but highly differentiated, justifying the subspecies status). The studies in this thesis, however, show that even within subspecies *petraea* the populations have diverged sufficiently to experience TRD which might be caused by genic incompatibilities. Intrinsic postzygotic incompatibilities are irreversible and give us an example of the genetics of incipient speciation, when genetic speciation studies in plants are still rare. At this stage we do not know if the different subspecies or populations will eventually develop into new species, but for understanding of speciation processes it is helpful to study diverging populations than already established species where genetic incompatibilities accumulated during or after speciation process cannot be distinguished.

Transmission ratio distortion has rarely been comprehensively studied to understand the origins of it and how it could further reveal the genetic basis of reproductive isolation between populations or species. In this thesis TRD in different crosses between *A. lyrata* populations was dissected to the gametic and zygotic components, to find that most of the TRD forms already at the  $F_1$  gametes or in the interaction of  $F_1$  male and female tissues. TRD was found to be due to complicated interactions between nuclear genes (epistatic interactions between two loci) and nuclear and cytoplasmic genes. Further studies would be needed on the role of conspecific pollen precedence and  $F_1$  male and female tissue interactions, to better understand their role compared to BDM incompatibilities.

The TRD studies also give further support for the theory of number of incompatibilities increasing with genetic distance between populations crossed. TRD was almost absent in within population crosses and clearly increased with genetic distance. Also reciprocal crosses were more different in the strength of TRD when cytoplasm came from different populations than when the cytoplasm was the same in both reciprocal crosses. The role of nuclear-cytoplasmic interactions was found to be important in creating incompatibilities between diverging populations. This was observed at the TRD studies but also in hybrid fertility studies.

In this thesis the first findings of reduced male fertility and cytoplasmic male sterility of hybrids between *A. lyrata* populations were demonstrated. The genetic background of cytoplasmic male sterility restoring was found to be likely similar to gynodioecious sex determination system or CMS restoring in cultivated plants, which are genetically well known. This study was the second to prove this mechanism to function on natural, hermaphroditic plants. QTLs for reduced male fertility were found to be different depending on cytoplasm. The QTLs were also polymorphic within populations. Within population polymorphisms were also detected in the formation of TRD between crossed populations. These results underline the possibility that the development of genic incompatibilities might not happen solely as by-products of natural selection (when the alleles would likely be fixed within population) but rather through neutral processes.

## 5 References

- 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.
- Arnold ML, Bouck AC & Cornman RS (2004) Verne Grant and Louisiana Irises: Is there anything new under the sun? *New Phytol* 161(1): 143–149.
- Arntzen JW, Jehle R, Bardakci F, Burke T & Wallis GP (2009) Asymmetric viability of reciprocal-cross hybrids between crested and marbled newts (*Triturus cristatus* and *T. marmoratus*). *Evolution* 63(5): 1191–1202.
- Avise JC, Walker D & Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London B* 265(1407): 1707–1712.
- Ayala FJ & Coluzzi M (2005) Chromosome speciation: Humans, *Drosophila*, and mosquitoes. *Proc Natl Acad Sci USA* 102: 6535–6542.
- Balana-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. *Mol Ecol* 15(8): 2059–2068.
- Barbash DA, Siino DF, Tarone AM & Roote J (2003) A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proc Natl Acad Sci 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, pp.85–101 in Darwin and modern science, edited by A. C. Seward. Cambridge, Cambridge University Press.
- Bayes JJ & Malik HS (2009) Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* sibling species. *Science* 326(5959): 1538–1541.
- Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. Pp. 250–266. *Proc. Forty-Ninth Annu. Corn and Sorghum Industry Res. Conf.*, Washington, D. C.
- Bechsgaard J, Bataillon T & Schierup MH (2004) Uneven segregation of sporophytic self-incompatibility alleles in *Arabidopsis lyrata*. *J Evol Biol* 17(3): 554–561.
- Beilstein MA, Nagalingum NS, Clements MD, Manchester SR & Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*. 107(43): 18724–18728.
- Belkhir K, Borsa P, Chikhi L, Raufaste N & Bonhomme F (2001). GENETIX 4.02, Logiciel Sous Windows Pour La GÉNÉTIQUE Des Populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II: Montpellier, France.
- Bell CJ & Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* 19(1): 137–144.

- Bikard D, Patel D, Le Mette C, Giorgi V, Camilleri C, Bennett MJ & Loudet O (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* 323(5914): 623–626.
- Bolnick DI & Near TJ (2005) Tempo of hybrid inviability in centrarchid fishes (Teleostei : Centrarchidae). *Evolution* 59(8): 1754–1767.
- Bomblies K (2010) Doomed lovers: Mechanisms of isolation and incompatibility in plants. *Annual Review of Plant Biology*, Vol 61 61: 109–124.
- Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL & Weigel D (2007) Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biology* 5(9): 1962–1972.
- Bratteler M, Lexer C & Widmer A (2006) A genetic linkage map of *Silene vulgaris* based on ALFP markers. *Genome* 49: 320–327.
- 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.
- Broman KW, Wu H, Sen S & Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19(7): 889–890.
- Case AL & Willis JH (2008) Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution* 62(5): 1026–1039.
- Cattani MV & Presgraves DC (2009) Genetics and lineage-specific evolution of a lethal hybrid incompatibility between *Drosophila mauritiana* and its sibling species. *Genetics* 181(4): 1545–1555.
- Charlesworth D (1981) A further study of the problem of the maintenance of females in gynodioecious species. *Heredity* 46(FEB): 27–39.
- Clark PU & Mix AC (2002) Ice sheets and sea level of the Last Glacial Maximum. *Quaternary Science Reviews* 21(1–3): 1–7.
- Clauss MJ, Cobban H & Mitchell-Olds T (2002) Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaceae). *Mol Ecol* 11: 591–601.
- Clauss MJ & Koch MA (2006) Poorly known relatives of *Arabidopsis thaliana*. *Trends Plant Sci* 11(9): 449–459.
- Clauss MJ & Mitchell-Olds T (2006) Population genetic structure of *Arabidopsis lyrata* in Europe. *Mol Ecol* 15(10): 2753–2766.
- Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y & Byers B (2000) Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* 12(9): 1551–1567.
- Coyne JA & Orr HA (1989) Patterns of speciation in *Drosophila*. *Evolution* 43(2): 362–381.
- Coyne JA & Orr HA (2004) *Speciation*. Sunderland, MA, USA, Sinauer Associates.
- Darwin C (1859) *On the origin of species*. John Murray, London.
- Delph LF, Touzet P & Bailey MF (2007) Merging theory and mechanism in studies of gynodioecy. *Trends Ecol Evol* 22(1): 17–24.

- Dobzhansky T (1936) Studies on hybrid sterility II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21(2): 113–135.
- Ellison CK, Niehuis O & Gadau J (2008) Hybrid breakdown and mitochondrial dysfunction in hybrids of *Nasonia* parasitoid wasps. *J Evol Biol* 21(6): 1844–1851.
- Ellison CK & Burton RS (2008) Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution* 62(3): 631–638.
- Ellstrand NC, Whitkus R & Rieseberg LH (1996) Distribution of spontaneous plant hybrids. *Proc Natl Acad Sci USA* 93(10): 5090–5093.
- Faris JD, Laddomada B & Gill BS (1998) Molecular mapping of segregation distortion loci in *Aegilops tauschii*. *Genetics* 149(1): 319–327.
- Feder JL, Opp SB, Wlazlo B, Reynolds K, Go W & Spisak S (1994) Host fidelity is an effective pre-mating barrier between sympatric races of the apple maggot fly. *Proc Natl Acad Sci USA* 91(17): 7990–7994.
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) Version 3.6, Distributed by the author. Department of genome sciences, University of Washington, Seattle.
- Ferree PM & Barbash DA (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *Plos Biology* 7(10): e1000234.
- Fishman L, Kelly AJ, Morgan E & Willis JH (2001) A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics* 159(4): 1701–1716.
- Fishman L & Willis JH (2001) Evidence for Dobzhansky-Muller incompatibilities contributing to the sterility of hybrids between *Mimulus guttatus* and *M. nasutus*. *Evolution* 55: 1932–1942.
- Fishman L, Aagaard J & Tuthill JC (2008) Toward the evolutionary genomics of gametophytic divergence: Patterns of transmission ratio distortion in monkeyflower (*Mimulus*) hybrids reveal a complex genetic basis for conspecific pollen precedence. *Evolution* 62(12): 2958–2970.
- Fishman L & Willis JH (2006) A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution* 60(7): 1372–1381.
- Fitzpatrick BM (2004) Rates of evolution of hybrid inviability in birds and mammals. *Evolution* 58(8): 1865–1870.
- Gaudeul M, Stenoien HK & Agren J (2007) Landscape structure, clonal propagation, and genetic diversity in Scandinavian populations of *Arabidopsis lyrata* (Brassicaceae). *Am J Bot* 94(7): 1146–1155.
- Gotoh T, Oku H, Moriya K & Odawara M (1995) Nucleus-cytoplasm interactions causing reproductive incompatibility between two populations of *Tetranychus quercivorus* Ehara-Et-Gotoh (Acari, Tetranychidae). *Heredity* 74: 405–414.
- 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.
- Greig D (2009) Reproductive isolation in *Saccharomyces*. *Heredity* 102(1): 39–44.

- Gutierrez-Marcos JF, Pennington PD, Costa LM & Dickinson HG (2003) Imprinting in the endosperm: a possible role in preventing wide hybridization. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 358(1434): 1105–1111.
- Haldane JBS (1922) Sex-ratio and unidirectional sterility in hybrid animals. *J Genet* 12: 101–109.
- Hall MC & Willis JH (2005) Transmission ratio distortion in intraspecific hybrids of *Mimulus guttatus*: Implications for genomic divergence. *Genetics* 170(1): 375–386.
- Hanson MR & 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.
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* 59(8): 1633–1638.
- Heenan PB, Dawson MI, Smitsen RD & Bicknell RA (2008) An artificial intergeneric hybrid derived from sexual hybridization between the distantly related *Arabidopsis thaliana* and *Pachycladon cheesemanii* (Brassicaceae). *Bot J Linn Soc* 157(3): 533–544.
- Hoffmann AA & Rieseberg LH (2008) Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology Evolution and Systematics* 39: 21–42.
- Hoffmann MH (2005) Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution* 59(7): 1425–1436.
- Howard DJ (1999) Conspecific sperm and pollen precedence and speciation. *Annu Rev Ecol Syst* 30: 109–132.
- Jalas J & Suominen J (1994) *Atlas Florae Europaea*. Distribution of vascular plants in Europe. 10. Cruciferae (Sisymbrium to Aubrieta). Helsinki, Finland, Helsinki University Printing House.
- Johnston JA, Donovan LA & Arnold ML (2004) Novel phenotypes among early generation hybrids of two Louisiana iris species: flooding experiments. *J Ecol* 92(6): 967–976.
- Jonsell B, Kustås K & Nordal I (1995) Genetic variation in *Arabis petraea*, a disjunct species in northern Europe. *Ecography* 18: 321–332.
- Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5(1): 187–189.
- Kearns CA & Inouye DW (1993) *Techniques for pollination biologists*. Niwot, CO, Univ. of Colorado Press.
- Koch MA, Haubold B & Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol* 17(10): 1483–1498.
- Koch MA & Matschinger M (2007) Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 104(15): 6272–6277.



- Koide Y, Ikenaga M, Sawamura N, Nishimoto D, Matsubara K, Onishi K, Kanazawa A & Sano Y (2008) The evolution of sex-independent transmission ratio distortion involving multiple allelic interactions at a single locus in rice. *Genetics* 180(1): 409–420.
- Korbecka G, Klinkhamer PGL & Vrieling K (2002) Selective embryo abortion hypothesis revisited - A molecular approach. *Plant Biology* 4(3): 298–310.
- Lai Z, Nakazato T, Salmaso M, Burke JM, Tang SX, Knapp SJ & Rieseberg LH (2005) Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* 171(1): 291–303.
- Laser KD & Lersten NR (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male sterile angiosperms. *Bot Rev* 38(3): 425-&.
- Lee H, Chou J, Cheong L, Chang N, Yang S & Leu J (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* 135(6): 1065–1073.
- Leinonen PH, Remington DL & Savolainen O (2010) Local adaptation, phenotypic differentiation and hybrid fitness in diverged natural populations of *Arabidopsis lyrata*. *Evolution* doi: 10.1111/j.1558-5646.2010.01119.x.
- Leushner J & Chiu NHL (2000) Automated mass spectrometry: A revolutionary technology for clinical diagnostics. *Molecular Diagnosis* 5(4): 341–348.
- Levin DA (2003) The cytoplasmic factor in plant speciation. *Syst Bot* 28(1): 5–11.
- Loudet O, Chaillou S, Camilleri C, Bouchez D & Daniel-Vedele F (2002) Bay-0 x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *Theor Appl Genet* 104(6–7): 1173–1184.
- 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.
- Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L & Small I (2004) Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16(8): 2089–2103.
- Lyttle TW (1991) Segregation Distorters. *Annu Rev Genet* 25: 511–557.
- 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.
- Masly JP, Jones CD, Noor MAF, Locke J & Orr HA (2006) Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science* 313(5792): 1448–1450.
- Matute DR, Butler IA, Turissini DA & Coyne JA (2010) A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* 329(5998): 1518–1521.
- Mayr E (1942) *Systematics and the origin of species, from the viewpoint of a zoologist*. Cambridge, Harvard University Press.

- McCauley DE & Olson MS (2008) Do recent findings in plant mitochondrial molecular and population genetics have implications for the study of gynodioecy and cytonuclear conflict? *Evolution* 62(5): 1013–1025.
- Mihola O, Trachtulec Z, Vlcek C, Schimenti JC & Forejt J (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323(5912): 373–375.
- Moyle LC & Graham EB (2005) Genetics of hybrid incompatibility between *Lycopersicon esculentum* and *L. hirsutum*. *Genetics* 169(1): 355–373.
- Moyle LC & Graham EB (2006) Genome-wide associations between hybrid sterility QTL and marker transmission ratio distortion. *Mol Biol Evol* 23(5): 973–980.
- Moyle LC & Nakazato T (2010) Hybrid incompatibility “snowballs” between *Solanum* species. *Science* 329(5998): 1521–1523.
- Moyle LC, Olson MS & Tiffin P (2004) Patterns of reproductive isolation in three angiosperm genera. *Evolution* 58(6): 1195–1208.
- Muller HJ (1942) Isolating mechanisms, evolution and temperature. *Biological Symposia* 6: 71–125.
- Mummenhoff K & Hurka H (1994) Subunit polypeptide composition of Rubisco and the origin of allopolyploid *Arabidopsis suecica* (Brassicaceae). *Biochem Syst Ecol* 22(8): 807–812.
- Müntzing A (1930) Outlines to a genetic monograph of the genus *Galeopsis* with special reference to the nature and inheritance of partial sterility. *Hereditas* 13(2): 185–341.
- Myburg AA, Vogl C, Griffin AR, Sederoff RR & Whetten RW (2004) Genetics of postzygotic isolation in *Eucalyptus*: Whole-genome analysis of barriers to introgression in a wide interspecific cross of *Eucalyptus grandis* and *E. globulus*. *Genetics* 166(3): 1405–1418.
- Nasrallah ME, Yogeewaran K, Snyder S & Nasrallah JB (2000) *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiol* 124(4): 1605–1614.
- Noor MAF, Grams KL, Bertucci LA & Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA* 98(21): 12084–12088.
- O’Kane SL & Al-Shehbaz IA (1997) A synopsis of *Arabidopsis* (Brassicaceae). *Novon* 7: 323–327.
- Orr HA (1995) The population-genetics of speciation - the evolution of hybrid incompatibilities. *Genetics* 139(4): 1805–1813.
- Orr HA (1996) Dobzhansky, Bateson, and the genetics of speciation. *Genetics* 144(4): 1331–1335.
- Ouyang Y, Liu YG & Zhang Q (2010) Hybrid sterility in plant: stories from rice. *Curr Opin Plant Biol* 13(2):186–92.
- Petit RJ, El Mousadik A & Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conserv Biol* 12(4): 844–855.
- Phadnis N & Orr HA (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* 323(5912): 376–379.

- Piry S, Luikart G & Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in effective population size using allele frequency data. *J Hered* 90: 502–503.
- Ponce MR, Robles P & Micol JL (1999) High-throughput genetic mapping in *Arabidopsis thaliana*. *Molecular and General Genetics* 261(2): 408–415.
- Presgraves DC (2002) Patterns of postzygotic isolation in *Lepidoptera*. *Evolution* 56(6): 1168–1183.
- Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nature Reviews Genetics* 11(3): 175–180.
- 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.
- Price T (2007) Speciation in birds. Greenwood Village, Colorado, Roberts & Company Publishers.
- R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Ramos-Onsins SE, Stranger BE, Mitchell-Olds T & Aguadé M (2004) Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. *Genetics* 166(1): 373–388.
- Ramsey J & Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu Rev Ecol Syst* 29: 467–501.
- Raymond M & Rousset F (1995) Genepop (Version-1.2) - Population-genetics software for exact tests and ecumenicism. *J Hered* 86(3): 248–249.
- Reed LK & Markow TA (2004) Early events in speciation: Polymorphism for hybrid male sterility in *Drosophila*. *Proc Natl Acad Sci USA* 101(24): 9009–9012.
- Rieseberg LH (2000) Crossing relationships among ancient and experimental sunflower hybrid lineages. *Evolution* 54(3): 859–865.
- Rieseberg LH (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.
- Riihimäki M & Savolainen O (2004) Environmental and genetic effects on flowering differences between northern and southern populations of *Arabidopsis lyrata* (Brassicaceae). *Am J Bot* 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.
- Schluter D (2009) Evidence for ecological speciation and its alternative. *Science* 323(5915): 737–741.
- Schmickl R, Jørgensen MH, Brysting AK & Koch MA (2008) Phylogeographic implications for the North American boreal-arctic *Arabidopsis lyrata* complex. *Plant Ecology & Diversity* 1(2): 245–254.

- 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.
- Schnable PS & Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci* 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.
- Seehausen O (2002) Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14 600 year history for its cichlid species flock. *Proceedings of the Royal Society of London Series B-Biological Sciences* 269(1490): 491–497.
- Shimizu KK, Fujii S, Marhold K, Watanabe K, Kudoh H (2005) *Arabidopsis kamchatica* (Fisch. ex DC.) K. Shimizu & Kudoh and *A. kamchatica* subsp. *kawasakiana* (Makino) K. Shimizu & Kudoh, new combinations. *Acta Phytotaxonomica et Geobotanica* 56: 165–174.
- Shimizu-Inatsugi R, Lihova J, Iwanaga H, Kudoh H, Marhold K, Savolainen O, Watanabe K, Yakubov VV & Shimizu KK (2009) The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. *Mol Ecol* 18(19): 4024–4048.
- Shuker D.M., Underwood K., King T.M. & Butlin RK (2005) Patterns of male sterility in a grasshopper hybrid zone imply accumulation of hybrid incompatibilities without selection. *Proceedings of the Royal Society of London Series B-Biological Sciences* 272(1580): 2491–2497.
- Snow AA & Spira TP (1991) Pollen vigor and the potential for sexual selection in plants. *Nature* 352(6338): 796–797.
- Stebbins GL (1958) The inviability, weakness, and sterility of interspecific hybrids. *Adv Genet* 9:147–215.
- Tao Y, Xhen SN, Hartl DL & Laurie CC (2003) Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. *Genetics* 164(4): 1383–1397.
- Taylor DR & Ingvarsson PK (2003) Common features of segregation distortion in plants and animals. *Genetica* 117(1): 27–35.
- Tiffin P, Olson MS & Moyle LC (2001) Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society of London Series B-Biological Sciences* 268: 861–867.
- Ting CT, Tsauro SC, Wu ML & Wu CI (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282(5393): 1501–1504.
- Törjék O, Witucka-Wall H, Meyer RC, von Korff M, Kusterer B, Rautengarten C & Altmann T (2006) Segregation distortion in *Arabidopsis* C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *Theor Appl Genet* 113(8): 1551–1561.
- Turelli M & Moyle LC (2007) Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* 176(2): 1059–1088.

- Van Ooijen JW and Voorrips RE (2001) JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands.
- Van Treuren R, Kuitinen H, Kärkkäinen K, Baena-Gonzalez E & Savolainen O (1997) Evolution of microsatellites in *Arabis petraea* and *A. lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol Biol Evol* 14: 220–229.
- Wernisch MM (2007) Molecular systematics, phylogeography and evolution of the genus *Cardaminopsis* HAYEK (Brassicaceae), the closest relatives of the modelplant *Arabidopsis thaliana* (L.) HEYNH. Dissertation, Universität für Bodenkultur, Wien, Austria.
- Widmer A, Lexer C & Cozzolino S (2009) Evolution of reproductive isolation in plants. *Heredity* 102(1):31–38.
- Willems G, Drager DB, Courbot M, Gode C, Verbruggen N & Saumitou-Laprade P (2007) The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp *halleri* (Brassicaceae): An analysis of quantitative trait loci. *Genetics* 176(1): 659–674.
- Wilson AC, Maxson LR & Sarich VM (1974) Two types of molecular evolution - evidence from studies of interspecific hybridization. *Proc Natl Acad Sci USA* 71(7): 2843–2847.
- Wright SI, Lauga B & Charlesworth D (2003) Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Mol Ecol* 12(5): 1247–1263.
- Wu CA, Lowry DB, Cooley AM, Wright KM, Lee YW & Willis JH (2008) *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100(2): 220–230.
- Wu CI & Beckenbach AT (1983) Evidence for extensive genetic differentiation between the sex-ratio and the standard arrangement of *Drosophila pseudoobscura* and *Drosophila persimilis* and identification of hybrid sterility factors. *Genetics* 105(1): 71–86.
- Wu CI & Davis AW (1993) Evolution of postmating reproductive isolation - the composite nature of Haldane rule and its genetic bases. *Am Nat* 142(2): 187–212.
- Wu CI, Johnson NA & Palopoli MF (1996) Haldane's rule and its legacy: Why are there so many sterile males? *Trends Ecol Evol* 11(7): 281–284.



## Original articles

- I Muller M-H, Leppälä J & Savolainen O (2008) Genome wide effects of postglacial colonization in *Arabidopsis lyrata*. *Heredity* 100(1): 47–58.
- II Kuittinen H, de Haan AA, Vogl C, Oikarinen S, Leppälä J, Koch M, Mitchell-Olds T, Langley CH & Savolainen O (2004) Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. *Genetics* 168(3): 1575–1584.
- III Leppälä J, Bechsgaard JS, Schierup MH & Savolainen O (2008) Transmission ratio distortion in *Arabidopsis lyrata*: Effects of population structure and the S-locus. *Heredity* 100 (1): 71–78.
- IV Leppälä J & Savolainen O Nuclear-cytoplasmic interactions reduce male fertility in hybrids of *Arabidopsis lyrata* subspecies. In press.
- V Leppälä J, Bokma F & Savolainen O Genetic basis of transmission ratio distortion in crosses between populations of *Arabidopsis lyrata*. Manuscript.

Reprinted with permission from Macmillan Publishers Ltd (I, III), Genetics Society of America (II) and John Wiley and Sons (IV).

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





ACTA UNIVERSITATIS OULUENSIS  
SERIES A SCIENTIAE RERUM NATURALIUM

552. Bisi, Jukka (2010) Suomalaisen susikonfliktin anatomia
553. Poutanen, Hilikka (2010) Developing the role of human resource information systems for the activities of good leadership
554. Saukkoriipi, Jaakko (2010) Theoretical study of the hydrolysis of aluminum complexes
555. Sirviö, Anu (2010) The role of factors promoting genetic diversity within social insect colonies
556. Tahkokorpi, Marjaana (2010) Anthocyanins under drought and drought-related stresses in bilberry (*Vaccinium myrtillus* L.)
557. Piippo, Sari (2010) Grazing tolerance of biennial meadow plants in relation to resource availability
558. Lappalainen, Niina (2010) The responses of ectohydric and endohydric mosses under ambient and enhanced ultraviolet radiation
559. Luojus, Satu (2010) From a momentary experience to a lasting one : the concept of and research on expanded user experience of mobile devices
560. Siirtola, Antti (2010) Algorithmic multiparameterised verification of safety properties : process algebraic approach
561. Lappi, Anna-Kaisa (2010) Mechanisms of protein disulphide isomerase catalyzed disulphide bond formation
562. Sarala, Marian (2010) Elongation of Scots pine seedlings under blue light depletion
563. Vance, Anthony (2010) Why do employees violate is security policies? : insights from multiple theoretical perspectives
564. Karppinen, Katja (2010) Biosynthesis of hypericins and hyperforins in *Hypericum perforatum* L. (St. John's wort) – precursors and genes involved
565. Louhi, Pauliina (2010) Responses of brown trout and benthic invertebrates to catchment-scale disturbance and in-stream restoration measures in boreal river systems
566. Hekkala, Riitta (2011) The many facets of an inter-organisational information system project as perceived by the actors
567. Niittyvuopio, Anne (2011) Adaptation to northern conditions at flowering time genes in *Arabidopsis lyrata* and *Arabidopsis thaliana*

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

UNIVERSITY OF OULU P.O.B. 7500 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

S E R I E S E D I T O R S

**A**  
**SCIENTIAE RERUM NATURALIUM**

*Professor Mikko Siponen*

**B**  
**HUMANIORA**

*University Lecturer Elise Kärkkäinen*

**C**  
**TECHNICA**

*Professor Hannu Heusala*

**D**  
**MEDICA**

*Professor Olli Vuolteenaho*

**E**  
**SCIENTIAE RERUM SOCIALIUM**

*Senior Researcher Eila Estola*

**F**  
**SCRIPTA ACADEMICA**

*Information officer Tiina Pistokoski*

**G**  
**OECONOMICA**

*University Lecturer Seppo Eriksson*

EDITOR IN CHIEF

*Professor Olli Vuolteenaho*

PUBLICATIONS EDITOR

*Publications Editor Kirsti Nurkkala*

ISBN 978-951-42-9356-6 (Paperback)

ISBN 978-951-42-9357-3 (PDF)

ISSN 0355-3191 (Print)

ISSN 1796-220X (Online)

UNIVERSITY of OULU  
OULUN YLIOPISTO

