OULU 2014
A 641

ACTA UNIVERSITATIS OULUENSIS

A
SCIENTIAE RERUM
NATURALIUM

Yongfeng Zhou

DEMOGRAPHIC HISTORY AND CLIMATIC ADAPTATION IN ECOLOGICAL DIVERGENCE BETWEEN TWO CLOSELY RELATED PARAPATRIC PINE SPECIES

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

ACTA UNIVERSITATIS OULUENSIS
A Scientiae Rerum Naturalium 641

YONGFENG ZHOU

DEMOGRAPHIC HISTORY AND CLIMATIC ADAPTATION IN ECOLOGICAL DIVERGENCE BETWEEN TWO CLOSELY RELATED PARAPATRIC PINE SPECIES

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Health and Biosciences of the University of Oulu for public defence in the OP auditorium (L10), Linnanmaa, on 5 December 2014, at 12 noon

UNIVERSITY OF OULU, OULU 2014

Acta Univ. Oul. A 641, 2014

Supervised by
Professor Outi Savolainen

Reviewed by
Professor Xiao-Ru Wang
Doctor Takahito Shikano

Opponent
Doctor Rémy Petit

ISBN 978-952-62-0675-2 (Paperback)
ISBN 978-952-62-0676-9 (PDF)

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

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2014

Zhou, Yongfeng, Demographic history and climatic adaptation in ecological divergence between two closely related parapatric pine species.
University of Oulu Graduate School; University of Oulu, Faculty of Science, Department of Biology; Biocenter Oulu
Acta Univ. Oul. A 641, 2014
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Both demographic histories and natural selection complicate the speciation process. There is a need to jointly study the effects of natural selection on so called magic traits that can cause reproductive isolation such as climatic adaptation, and its interaction with neutral demographic histories. Closely related incipient coniferous species offer us a great system for this effort.

I used genetic variation at one set of climate-related candidate genes and another set of reference loci and cytoplasmic genomic fragments of two closely related parapatric pine species: *Pinus massoniana* Lamb. and *Pinus hwangshanensis* Hisa. Population genetic analyses were used to measure genetic variation and detect signals of ancient and recent selection. Speciation parameters including migration rates and divergence times at candidate genes and reference loci were compared under the Isolation with migration model. Hierarchical Approximate Bayesian Computation (ABC) was used to define demographic and speciation models. Intra- and interspecific genetic variation at cytoplasmic and nuclear intronic sequences were compared between parapatric populations and allopatric populations to distinguish the effects of introgression and incomplete lineage sorting in generating shared genetic variation between the species.

The results showed that ancient selection were shared by the lineages leading to the species while recent selection has been species-specific. Candidate genes had significant lower migration rates compared to reference loci. Recent differential climatic selection might counteract against gene flow at underlying genes, which therefore favors divergence between the two pines through ecological speciation. Shared mitotypes were randomly distributed across species' ranges, which therefore supported the incomplete lineage sorting hypothesis, but the shared nuclear intronic variation distributed more frequently in parapatric populations than in allopatric populations, supported the introgression hypothesis. ABC and species' distribution modeling also supported the secondary gene flow model. The three genomes had different rates of mutation and gene flow might mirror different phases of the speciation continuum. The results in this thesis are valuable for understanding evolution in general and for other applied purposes such as tree breeding and climate change adaptation.

Keywords: climatic adaptation, conifer, demographic history, ecological speciation, gene flow, incomplete lineage sorting, introgression, Pinus hwangshanensis, Pinus massoniana

Zhou, Yongfeng, Demografian ja ilmastoon sopeutumisen vaikutus kahden lähisukuisen mäntylajin ekologisessa erilaistumisessa.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Luonnontieteellinen tiedekunta, Biologian laitos; Biocenter Oulu
Acta Univ. Oul. A 641, 2014
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Luonnonvalinta ja populaatioiden historian demografia tekevät lajiutumisesta monimutkaisen tapahtumaketjun. Luonnonvalinnan ja demografisten tekijöiden vuorovaikutusta on paras tutkia samanaikaisesti kun tarkastellaan lajiutumiseen vaikuttavia ominaisuuksia. Tällaisia ovat esimerkiksi ilmastoon sopeutumiseen liittyvät ominaisuudet. Lähisukuiset havupuulajit tarjoavat erinomaiset mahdollisuudet tähän työhön. Tutkin geneettistä muuntelua yhtäältä ilmastosopeutumiseen liittyvissä ns. ehdokasgeeneissä ja toisaalta neutraaleiksi oletetuissa verrokkigeeneissä sekä sytoplasman genomeissa kahdessa lähisukuisessa mäntylajissa *Pinus massoniana* Lamb. ja *Pinus hwangshanensis* Hisa, joiden populaatiot esiintyvät joskus erillään toisistaan (allopatrisesti), toisinaan vierekkäin (parapatrisesti). Mittasin muuntelun määrää ja etsin merkkejä valinnan vaikutuksesta. Vertasin erilaisia lajiutumismallien parametrejä verrokki- ja ehdokasgeeneissä. Käytin simulaatioita etsiäkseni parhaat demografiset ja lajiutumiseen liittyvät mallit. Vertasin kloroplastien ja mitokondrioiden genomien sekvenssien lajinsisäistä ja lajien välistä muuntelua allopatrisissa ja parapatrisissa populaatioissa tutkiakseni onko lajien yhteinen muuntelu seurausta siitä että lajien eriytymisestä on kulunut vain vähän aikaa vai siitä että sen jälkeen on tapahtunut geenivirtaa. Kauan sitten tapahtunut valinta on vaikuttanut samalla tavalla kumpaankin lajiin, osin koska tutkimus kohdistui myös niiden yhteiseen edeltäjälinjaan. Äskettäinen valinta taas oli suuremmassa määrin kummallekin lajille ominaista. Viime aikojen ilmastoon liittyvä valinta on voinut vähentää geenivirtaa ehdokasgeeneissä, mikä voisi edistää ekologista lajiutumista. Tuman DNA:n muuntelu jakautuminen tuki sitä mahdollisuutta että lajien yhteinen geneettinen muuntelu johtuu äskettäisestä geenivirrasta, ei vain siitä että lajiutuminen on niin varhaisessa vaiheessa. Mitokondrioiden geeneissä lajeilla yhtä paljon yhteistä muuntelua sekä allopatrisissa että parapatrisissa populaatioissa, mikä tukee sen sijaan eriytymisen jälkeistä epätäydellistä muuntelun erilaistumista. Eri genomit heijastavat lajiutumisprosessin eri vaiheita. Väitöskirjan tulokset ovat osaltaan tuottaneet uutta tietoa lajiutumisesta ja valinnasta. Lisäksi niillä on merkitystä ilmastomuutoksen vaikutusten ymmärtämisessä ja metsänjalostuksessa.

Asiasanat: demografia, ekologinen lajiutuminen, geenivirta, havupuu, ilmastoon sopeutuminen, introgressio, Pinus hwangshanensis, Pinus massoniana, populaation historia

To my dearest grandparents and parents

Acknowledgements

The experiment part was conducted at School of Life Sciences, Lanzhou University and data analyses and writing was done at Department of Biology, University of Oulu, during 2010-2014.

First, I wish to express my deepest gratitude to Professor Outi Savolainen for the precious opportunity to learn the basic skills for scientific research and eventually to prepare doctoral thesis under her excellent supervision. I thank Professor Jianquan Liu from Lanzhou University most warmly for his co-supervision in designing the experiments and revising the manuscripts.

Professor Mikko Sillanpää and Doctor Lumi Viljakainen from University of Oulu and Professor Martin Lascoux from Uppsala University were my Follow-up group members. Thank you for useful discussions and following up my studies and thesis progresses.

Special thanks go to Professor Dr. Xiao-Ru Wang from Umea University and Dr. Takahito Shikano from University of Helsinki for reviewing my thesis at a very tight schedule.

I am most grateful to past and current members of Plant Genetics group at University of Oulu and Molecular Ecology group at Lanzhou University. Aleksia Vaattovaara helped me start-up when I firstly arrived Oulu. Tuomas Toivainen, Esa Aalto, Sonja Kujala, Jaakko Tyrmi, Komlan Avia, Helmi Kuittinen, Jaromir Guzinski, Tiina Mattila and Tanja Pyhäjärvi have given me useful suggestions in data analyses. Bin Tian, Xinmin Tian collected samples for the projects. Three master students, Zuyao Jiang, Guangpeng Ren and Lirui Zhang had worked on this projects. This thesis would not have been finished without your help, Thank you very much!

Many thanks to my Chinese friends in Finland: Zufan Zhang, Zitong Li, Gaowei Shi, Tonghe Liu, Qilai Huang, Wenying Yan, Jindong Shan, Hongmin Tu, Ping Gao, Kai Zhang, Ruizhu Lin, Yuehong Yang and Gonghong Wei. It was great to party, go shopping, go fishing and speak Chinese with you. I could not imagine how I can overcome the dark polar winter days without the pleasure times being with you. I also want thank my other friends in China and Finland, thank you for always bringing me happiness and hope.

Furthermore, I want to thank Guili Wu and Soile Alatalo for their excellent work as lab technicians. I also want to thank colleagues and administrative people in both University of Oulu and Lanzhou University for the supportive working atmosphere.

Finally, my warmest thanks go to my girlfriend Qi and my family. Thanks for always being there for me! Thank you!

The research was financially supported by grants from the National Natural Science Foundation of China. My stay in Finland was supported by the Chinese Scholarship Council (CSC NO. 2010618044), the funding from the Center for International Mobility (CIMO, Finland) and the FP7 ProcoGen (FP7-289841) project and the Department of Biology, Faculty of Sciences, University of Oulu.

Oulu, August 2014 Yongfeng Zhou

Abbreviations

θ	population mutation rate
π	nucleotide diversity
μ	mutation rate
bp	base pair
PCR	polymerase chain reaction
CTAB	Cetyltrimethyl ammonium bromide
PVP	polyvinylpyrrolidone
SNP	single nucleotide polymophism
N_e	effective population size
LD	linkage disequilibrium
ILS	incomplete lineage sorting
IBD	isolation by distance
IBE	isolation by environment
cp	chloroplast
mt	mitochondial
SNE	standard neutural equilibrium, i.e. the Wright-Fisher model
HKA	Hudson-Kreitman-Aguadé test
MLHKA	maximum likelihood Hudson-Kreitman-Aguadé test
MK	McDonald–Kreitman test
MKprf	McDonald-Kreitman Poisson Random Field test
IM	Isolation with migration
ABC	Approximate Bayesian Computation
MCMC	Markov Chain Monte Carlo
MYA	million years ago
DHEW	Compound neutrality test based on Tajima's *D*, Fu & Wu 's *H*, and the Ewens–Watterson haplotype test

List of original publications

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

I **Zhou YF**, Zhang LR, Liu JQ, Wu GL, Savolainen O (2014) Climatic adaptation and ecological divergence between two closely related pines species in Southeast China. Molecular Ecology, 23: 3504-3522.

II **Zhou YF**, Abbott RJ, Jiang ZY, Du, FK, Milne RI, Liu JQ (2010) Gene flow and species delimitation: a case study of two pine species with overlapping distributions in southeast China. Evolution, 64: 2342-2352.

III **Zhou YF**, Duvaux L, Zhang LR, Ren GP, Savolainen O, Liu JQ (2014) Incomplete lineage sorting versus introgression: origin of shared genetic variation between two closely related pines with overlapping distributions. Manuscript.

Author contributions

Paper	Study design	Experiments	Data analyses	Manuscript preparation
I	**YZ**, JL, OS	**YZ**, LZ, GW	**YZ**	**YZ**, OS, JL
II	**YZ**, JL, RA	**YZ**, ZJ	**YZ**	**YZ**, RA, FD, JL
III	**YZ**, OS, JL	**YZ**, LZ, GR	**YZ**, LD	**YZ**, LD, OS, JL

Yongfeng Zhou (YZ), Jianquan Liu (JL), Outi Savolainen (OS), Richard J Abbott (RA), Ludovic Duvaux (LD), Lirui Zhang (LZ), Zuyang Jiang (ZJ), Fang Du (FD), Guili Wu (GW), Guangpeng Ren (GR).

Contents

1 Introduction

Understanding the speciation continuum has been a long-standing goal for evolutionary biologist. Speciation is by-products of various processes between diverging populations, such as development of genomic incompatibilities (Bateson 1909; Dobzhansky 1936; Muller 1939), genetic drift (Rice & Hostert 1993; Coyne & Orr 2004), hybridization or polyploidy (Mallet 2007; Soltis & Soltis 2009; Abbott *et al.* 2013) and divergent natural selection (i.e. ecological speciation, Nosil 2012). In line with Darwin's perspective (Darwin 1859), ecological speciation has recently been indicated as a common speciation model (Nosil 2012). Local adaptation is evolutionary processes by which populations become better suited to their own environments through genetic change. When populations inhabit heterogeneous environments, selective pressure on traits such as phenology, soil, and climate may differ spatially across the distribution range (Kawecki & Ebert 2004). Climate is one of most important drivers for adaptive evolution in forest trees (Aitken *et al.* 2008; Alberto *et al.* 2013). Climatic selection has contributed greatly to genetic differentiation among populations within species (Grivet *et al.* 2011; Kujala & Savolainen 2012; Chen *et al.* 2012), but we still poorly know whether climatic selection can also give rise to reproductive barrier between populations from heterogeneous environments through ecological speciation. Closely related incipient coniferous species are well adapted to their environment and on the way of speciation, and therefore offer a unique opportunity to study the effects and interplays of gene flow and selection during divergence. However, the investigation of gene flow is complicated by incomplete lineage sorting, and the confusion of both processes can lead to inaccurate inference on evolutionary histories (e.g. Rieseberg *et al.* 1999; Charlesworth *et al.* 2005; Sousa & Hey 2013; Huerta-Sánchez *et al.* 2014).

1.1 Local adaptation to climate

Adaptation is evolutionary process by which populations become better suited to their own environments through genetic change. When populations inhabit heterogeneous environments, selective pressure on traits such as phenology, cold hardiness and drought tolerance may differ spatially across the distribution range. As a result, populations may become locally adapted, meaning that at each site, the local population has higher fitness than other introduced populations would have (Kawecki & Ebert 2004). Detecting the footprints of local selection at sequence level is difficult, because demographic histories such as population expansions, bottlenecks and subdivisions generate similar patterns of genetic variation (Tajima 1989; Zeng *et al.* 2007). However, local selection functions at restricted genomic islands underlying specific selective pressures while demographic histories affect genome-wide variation and differentiation (Wright & Gaut 2005; Excoffier 2009).

As the detection of selection at the molecular level is often complicated by confounding effects of demographic histories, several approaches should be used in complementary ways to ascertain the targeted genes (e.g. Wright & Gaut, 2005). In detecting recent selection, a compound site frequency spectrum (SFS) test, DHEW has been proved to be relatively insensitive to demography or population structure (Zeng *et al.* 2007). Nielsen *et al.* (2005) introduced the composite likelihood ratio method (CLR) to model the allele frequency spectrum to detect signals of selection. Genetic differentiation (F_{ST}) based methods have been widely used in indicating the outliers that have been targeted by local selection (Lewontin & Krakauer 1973; Beaumont & Nichols 1996; Akey *et al.* 2002; Beaumont 2005; Foll & Gaggiotti, 2008; Excoffier *et al.* 2009). However, F_{ST} is highly variable over loci under neutrality and depends on the allele frequency of the SNPs before selection (De Mita *et al.* 2013). Other kinds of methods based on linkage disequilibrium (LD) and extended haplotype heterozygosity (EHH) was designed to detect recent selective sweeps (Sabeti *et al.* 2007; Tang *et al.* 2007). Recently, powerful methods based on the maximum frequency of derived mutations (MFDM) to examine the extent of unbalance of the population tree of a locus (Li 2011). The MFDM method is free from the confounding effects of demography and population structure with a very high statistical power. Association analyses between genetic data and environmental variants have been efficiently used to detect selection at candidate genes with using neutral makers as control (e.g. Parisod & Christin 2008, Grivet *et al.* 2011; Coop 2010). Model based methods, such as Approximate Bayesian Computation (ABC) have also been used in detecting selection by examining the fit of candidate genes statistics to posterior distributions of specific summary statistics of appropriate demographic models built using neutral makers (e. g. Keller *et al.* 2011). Footprints of ancient selection can be detected by comparing within species polymorphism and among species divergence using the Hudson–Kreitman–Aguadé (HKA) tests (Hudson *et al.* 1987), the McDonald–Kreitman (MK) tests (McDonald & Kreitman 1991) and their extensions: maximum likelihood HKA tests (MLHKA, Wright & Charlesworth 2004) and McDonald-Krietman Poisson Random Field (MKPRF) test (Bustamante *et al.* 2002).

Climate is one of the most important drivers for adaptive evolution in forest trees (Aitken *et al.* 2008; Alberto *et al.* 2013). Local adaptation to climate had been extensively investigated in *Populus* (e.g. Ma *et al.* 2011; Keller *et al.* 2011; Keller *et al.* 2012; Evans *et al.* 2014). There are much evidence of climatic selection in coniferous trees (e.g. Savolainen *et al.* 2007; Kujala & Savolainen 2012; Chen *et al.* 2012; Chen *et al.* 2014). Several candidate genes in conifers have been suggested to be under local selection, such as drought-stress in *Pinus taeda* (Gonzalez-Martinez *et al.* 2006), *Pinus pinaster* (Eveno *et al.* 2008; Grivet *et al.* 2009; Grivet *et al.* 2011) and *Pinus halepensis* (Grivet *et al.* 2011), and cold tolerance in *Pinus sylvestris* (Pyhäjärvi *et al.* 2007; Wachowiak *et al.* 2009; Kujala & Savolainen 2012) and *Picea abies* (Chen *et al.* 2012; 2014). It has been shown that genetic variation at climate-related candidate genes in many conifers is associated with climatic variables (Eckert *et al.* 2010, Grivet *et al.* 2011; Chen *et al.* 2012). Reciprocal transplant experiments have shown that populations of the *Pinus densata* with a hybrid origin have the highest fitness at its native sites at Qinghai-Tibetan Plateau rather than at its parents' ranges (Zhao *et al.* 2014).

1.2 Demographic histories

Neutral genomic variation offers us the data to track population histories. Population genetic methods provide opportunities for evaluating demographic hypotheses of constant population size through time, of population expansions or bottlenecks, of the rate of growth or decline, and of migration and population size changes in subdivided populations. The core of many of these methods is Monte Carlo approaches based on coalescent theory. With the development of genome sequencing technologies, big data sets from multiple populations will improve the estimations of demographic parameters for complicated models (e.g. Nadachowska-Brzyska *et al.* 2013).

Natural populations have been affected by past climate change. Populations were restricted to one or subdivided into several refugee during glacial periods. At interglacial periods, population recolonized to their preferred habitats with potential secondary contact among isolated populations (Taberlet & Cheddadi 2002; Hewitt 2000). Demographic parameters inference had its potential to shed light on the conservation strategies for endangered species and historical biogeographic changes. Demographic inferences can also be used to study speciation scenarios between closely related species, but it becomes complicated when there is much structure among conspecific populations.

With the development of population genetics, approaches such as Approximate Bayesian Computation (ABC, Beaumont *et al.* 2002, Kuhner 2009, Csillery *et al.* 2010, Sunnaker *et al.* 2013) and Isolation with migration model (IMa, Wakeley & Hey 1998; Nielsen & Wakeley 2001; Hey 2010) have been successfully used to define demographic and speciation histories. ABC allows inference from complex models by substituting data simulated under a given a model under investigation for its respective intractable likelihood function (Beaumont *et al.* 2002) while IMa is based on full likelihood.

Inferring the demographic histories of speciation with gene flow is complicated by incomplete lineage sorting and introgression. IMa (Wakeley & Hey 1998; Nielsen & Wakeley 2001; Hey 2010) and coalescent genealogy samplers using ABC (Beaumont *et al.* 2002) can be used in test alternative evolutionary hypotheses. Because allele frequency changes due to genetic drift in different populations should not be correlated (Reich *et al.* 2009; 2012), new methods such as the *f*3, the *f*4-ratio and the *D* statistics have been successfully used in inferring signals, proportions and directions of mixture events (Reich *et al.* 2009; Green *et al.* 2010; Durand *et al.* 2011; Patterson *et al.* 2012; Jeong *et al.* 2014). Admixture induced LD decay with genetic distance has also been successfully used in dating historical mixture events in human populations (Moorjani *et al.* 2011; Patterson *et al.* 2012; Loh *et al.* 2013). When the geographic distribution information is available, these two hypotheses can be distinguished by comparing patterns of genetic variation between neighboring populations and among distantly located ones (Muir & Schlotterer 2005, Huerta-Sanchez *et al.* 2014). Introgression is expected to happen between neighboring populations, which therefore result in higher levels of intraspecific genetic diversity and lower levels of interspecific genetic differentiation in geographic neighboring populations relative to distantly located ones, because introgression should bring foreign diversity and reduce genetic differentiation (Petit & Excoffier 2009).

1.3 Ecological speciation with gene flow

Adaptation and speciation are closely related (Darwin, 1859). As a result of selection for local adaptation, differentiation and sometimes reproductive isolation can occur (Mayr 1963; Michalak *et al.* 2001; Lenormand 2002). This process is called ecological speciation (Rundle & Nosil 2005; Sobel *et al.* 2010; Nosil 2012; Arnegard *et al.* 2014). The signal of selection in ecological speciation can be revealed by evaluating the correlation between levels of adaptive phenotypic and genetic differentiations across pairwise population comparisons (Funk *et al.* 2011). In early phases of ecological speciation, divergence can be hastened by reducing of gene flow due to spatial separation (i.e. isolation by distance, IBD) and/or natural selection against maladapted immigrants (i.e. isolation by environment, IBE) (Barton & Bengtsson 1986; Gavrilets & Cruzan 1998; Andrew *et al.* 2012). The effects of IBD can be accessed by using correlation analysis between genetic distances and geographic distances while the effects of IBE can be accessed by correlation between genetic distances and environmental variables.

Ecological speciation can even occur in the face of gene flow (de Leon *et al.* 2010; Nosil 2012). The process consists of four phases: firstly, genes of ecologically relevant traits are targeted by direct selection and gene exchange at these loci is restricted; secondly, gene exchange is restricted in gene regions linked to loci undergoing divergent selection, which is called divergent hitchhiking. Then, this limitation of migration extended to the whole genome because of linkage disequilibrium, i.e. genome hitchhiking. Eventually, the barrier to gene flow between species evolves before post speciation divergence (Wu 2001; Wu & Ting, 2004; Feder *et al.* 2012). Speciation with gene flow has been commonly indicated in many organisms (e. g. Ma *et al.* 2006; Wachowiak *et al.* 2011 and Zhou *et al.* 2014 in pines; Chen *et al.* 2009; Li *et al.* 2010; Zou *et al.* in spruce; Zhou *et al.* 2012 in frogs; Niemiller *et al.* 2008 in salamanders and Cristescu *et al.* 2012 in Daphnia). And, selection for climatic adaptation has also been indicated in many cases (e.g. Fournier-Level *et al.* 2011; Grivet *et al.* 2011), but it is less clear whether such climatic adaptation is also an important driver of tree speciation through ecological speciation. Divergent selection and/or species-specific selection due to climatic pressures might counteract interspecific gene flow at underlying genes between closely related species with contrasting climatic preferences (Hua & Wiens 2013), and give rise to earlier estimated divergence times and lower estimated migration for these genes than presumably neutral reference genes (which are not related to the traits of interests).

1.4 Coniferous trees as a study system for ecological genomics

Conifers are group of about 670 ecologically and economically important tree species. They have experienced past climate oscillation since the Carboniferous Period (360-286 million years ago, MYA) (Miller, 1977). During this time, they have adapted to different climatic conditions and they have given rise to new species while undergoing many demographic events, such as range and population size expansions, bottlenecks and variable gene flow (e.g. Li *et al.* 2012; Zhou *et al.* 2014). Closely related conifers are still commonly partly sharing their gene pools (Ren *et al.* 2012; Li *et al.* 2010a; Ma *et al.* 2006; Wachowiak *et al.* 2011), meaning that the speciation and adaptation processes are going on simultaneously (e. g. Ma *et al.* 2006; Zhou *et al.* 2014). This history and the closely related species pairs of conifers offer us a precious opportunity to investigate the effects of selection and gene flow and interactions between them during the speciation process.

In conifers, mitochondrial, chloroplast and nuclear genomes were maternally, paternally or bi-parentally inherited though seed, pollen or both, respectively (Wagner *et al.* 1987; Neale & Sederoff 1988). Pollen flow is more efficient in connecting conspecific populations than seed flow (Ennos 1994; Liepelt *et al.* 2002). cp DNAs therefore show species specific genetic diversity while mt DNA genetic diversity is commonly shared among species (Du *et al.* 2009; Zhou *et al.* 2010). Complicated patterns of intra- and interspecific genetic diversity at the three genomes of conifers might be explained by different rates of gene flow and mutation.

Pines have an origin in the Early Cretaceous (Millar 1998). There are about 110 pine species occurring mainly in the northern hemisphere (Richardson 2000; Gernandt *et al.* 2005). There are about 20 hard pines in the subsection *Pinus* that are widely distributed in Eurasia. In particular, there are few hard pines present in Europe but 12 of them occur in China (Richardson 2000). Each species in China is distributed in a specific habitat, for example *Pinus densata* in the high Qinghai-Tibetan plateau (QTP) and *P. tabulaeformis* in the dry northern China, but in southeast China two pairs of species (*P. massoniana* & *P. hwangshanensis* and *P. densiflora* & *P. thunbergii*) have overlapping distributions with different preference of elevations (Fu *et al.* 1999).

Here, we employed a pair of closely related pines, *P. massoniana* and *P. hwangshanensis* (*Pinus* subgenus *Pinus* subsection *Pinus*, Gernandt *et al.* 2005) coexisting in southeast China with different preference of elevations (Fig. 1).

Both species are ecologically and economically important for Southeast China. *P. massoniana* tends to present at altitude below 1000m at the base of the mountains while *P. hwangshanensis* mainly occurs at altitudes above 900m to the treeline of the mountains (Fu *et al.* 1999). Previous investigations based on genetic variation suggested that lineage sorting between the two species is still incomplete (Luo *et al.* 2000; Zhang *et al.* 2014). Within their distribution ranges, no other pine species from the same subsection are present, suggesting that present-day hybridization with other pine species does not occur. Phylogeographic analysis used cpDNA sequence data has been conducted previously on *P. hwangshanensis* (Chiang *et al.* 2006) and *P. massoniana* (Ge *et al.* 2012), which suggested that conspecific populations have weak hierarchical structure. Seed germination rate in the hybrid zones was lower than 50% compared to natural populations of *P. massoniana* and *P. hwangshanensis*, suggesting some level of reproductive barrier evolved between the species (Li *et al.* 2010b). The two closely related pines with overlapping distributions offer us a precious opportunity to investigate the balance between local selection and gene flow during divergence between them and looking for footprints of selection for local adaptation at climate related candidate genes.

1.5 Aim of the study

In this study, we sequenced 25 climatic related candidate genes (drought and cold hardiness) and 12 other loci (not related to climatic adaptation), in 15 and 11 populations across the ranges of *P. massoniana*, *P. hwangshanensis*, respectively. We also sequenced three chloroplast and two mitochondrial gene fragments for a wider sampling (Figure 1 in paper II). Firstly, we compared patterns of intra- and interspecific genetic diversity at cp, mt and nu makers between parapatric and allopatric populations of the two species to distinguish ILS and introgression (paper II and III). Second, we used ABC to define demographic history of each species and speciation models between them with consideration of the complicating processes of ILS and introgression (paper III). Third, multiple approaches were used to detect the signals of recent and ancient selection at candidate genes and reference loci. Population genetic parameters were estimated under isolation with migration model for reference loci and candidate genes (Paper I). In particular, we were addressing the following questions:

1) Is the difference in the size of the geographical distribution of *P. massoiniana* and *P. hwangshanensis* reflected in effective population sizes and thus in patterns

of genetic diversity and linkage disequilibrium (Paper III)? In other words, does *P. massoniana* have higher levels of silent nucleotide diversity and higher rates of recombination than *P. hwangshanenesis* because of its wider distribution and possibly higher *N*e?

2) How do the different rates of pollen and seed flow influence the patterns of cpDNA and mtDNA variation? In particular, what is the influence on species boundaries, through the effects on introgression and linage sorting? With higher rates of pollen flow, does cp DNA variation show clearer species boundaries compared to mt DNA? If yes, do higher rates of gene flow dilute the introgressed foreign alleles or facilitate lineage sorting rates (Paper II)?

3) Does incomplete lineage sorting or introgression explain the shared genetic variation between the species (Paper III)? Given the different opportunities for gene exchange, how does the distribution of nuclear genetic variation in parapatric and allopatric populations compare to that of cytoplasmic genetic variation (Paper III)?

4) Do recent and ancient selection influence the same loci and how do they relate to climate adaptation? (Paper I)?

5) Does differential recent selection between environments give rise to reduced gene exchange? Is the observed pattern of climate related candidate genes and reference loci in line with ecological speciation by divergent climatic selection (Paper I)?

2 Materials and methods

Materials and methods are only described briefly here, more details were included in papers

2.1 Study materials

We collected both needles and cones with mature seeds of 6-15 individuals across 11 and 19 natural populations from *P. hwangshanensis* and *P. massoniana*, respectively. We analyzed all the sampled individuals using cp and mt DNA fragments (Fig.1 in paper I). All the *P. hwangshanensis* populations but only 15 of *P. massoniana* populations were analyzed using nuclear genetic variation (Fig. 1). Two individuals of *Pinus koraiensis* (subgenus *Strobus*) from northeast China (Mohe, Heilongjiang, 53° 03′ N, 122° 22′ E) were sampled for use as an outgroup.

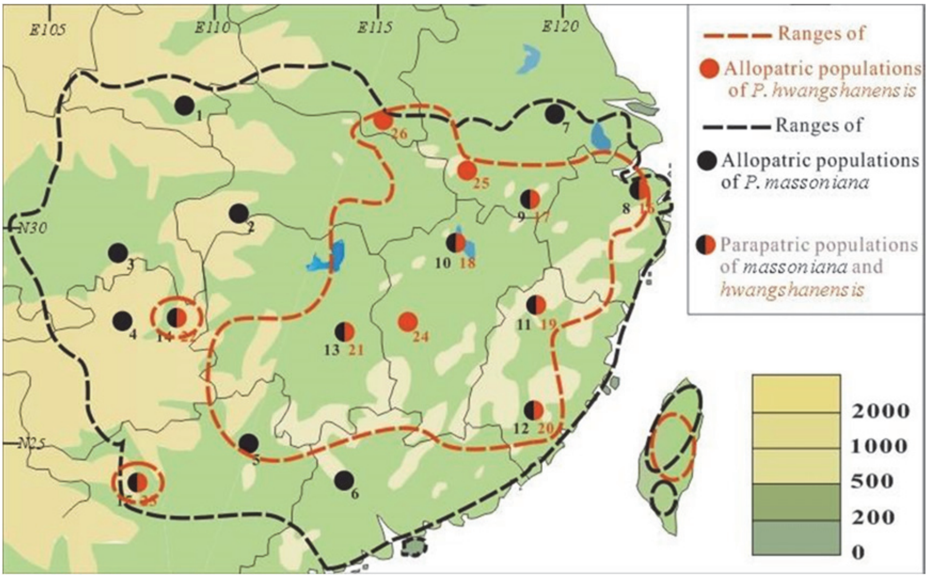


Fig. 1. Geographic distributions and focal populations of *P. massoniana* and *P. hwangshanensis.*

2.2 Experiments

DNA for cytoplasmic genetic analysis (paper II) was isolated from silica-gel dried needles according to a CTAB procedure (Doyle & Doyle 1987) with an electric tissue homogenizer (QIAGEN, manufactured by Retsch). Nuclear DNA for population genetic analysis (paper I and III) was extracted from haploid megagametophytes of germinated seeds using a QIAGEN DNeasy Plant Mini Kits (QIAGEN, Inc., Valencia, CA, USA), with PVP added to the buffer (final concentration, 1%).

We amplified and sequenced two mitochondrial and three chloroplast gene fragments using primers described by previous studies (Gamache *et al.* 2003; Dumolin-Lapegue *et al.* 1997; Small *et al.* 1998; Demesure *et al.* 1995; and Wang *et al.* 1998; Paper II). A set of 25 candidate genes (Table 1 in paper I) were chosen considering previous studies in pines at candidate genes related to climatic adaptation to cold (*P. sylvestris*, Wachowiak *et al.* 2009) and drought (*P. taeda*, Gonzalez-Martinez *et al.* 2006; *P. pinaster*, Eveno *et al.* 2008, Grivet *et al.* 2011). All of these genes were first identified by analyses of gene expression patterns in plants under cold and drought stress (e.g. Joosen *et al.* 2006; Lorenz *et al.* 2006). Similarly, a set of 12 loci (Dvornyk *et al.* 2002; Brown *et al.* 2004; Ma *et al.* 2006; Wachowiak *et al.* 2011; Ren *et al.* 2012) assumed to be neutral were selected for use as reference loci to compare with candidate genes and define demographic histories. Primers were designed using the primer3 software (http://frodo.wi.mit.edu/primer3/) based on available genomic sequences of conifers.

PCR reactions were detailed in paper I and II. PCR products were purified using a TIANquick Midi Purification Kit (Tiangen, Beijing, China) and Sequenced with the forward and reverse PCR primers for all amplicons, using an ABI Prism BigDye Terminator Cycle Version 3.1 Sequencing Kit and an ABI3130xl Genetic Analyzer (Applied Biosystems) at Lanzhou University or an ABI3730xl Genetic Analyzer at Beijing Genome Institute (BGI).

Singletons were verified by resequencing the reamplified fragments from the same individual. Only sequences with single peaks were retained. We aligned DNA sequences using ClustalX (Thompson *et al.* 1997) or MUSCLE (Edgar 2004) implemented in Mega 5 (Tamura *et al.* 2011). All putative polymorphic sites were further confirmed by visual inspection of the chromatograms. Coding and non-coding regions (introns and untranslated regions) were annotated on the basis of National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). Haplotype sequences of each locus was deposited in the GenBank under accession numbers: FJ906698-FJ906722 and KJ921127–KJ921496.

2.3 Statistical analyses

Genetic diversity was compared between the two species at both cytoplasmic and nuclear makers considering demographic histories and effective population size (paper I and III). Hierarchical ABC was employed to define demographic histories of each species and further to model speciation scenarios between the two species using intronic genomic sequences (paper III). Multiple complementary approaches were used to detect ancient and recent selection signals at climate-related candidate genes (paper I). Population genetic parameters were compared between climate-related candidate genes and reference loci based on IM model (Hey 2010) to test the hypothesis that recent species specific selection might have counteracted against gene flow and favored divergence between the two species (paper I). We compared genetic diversity at allopatric and parapatric populations of the two species to investigate the complicated processes of ILS and gene flow as sources of shared genetic variation, accompany with ABC model selection and species distribution modeling (Paper II and III). We also compared patterns of genetic diversity at cp, mt and nuclear makers (Paper II and III).

2.3.1 Inter- and intraspecific genetic diversity

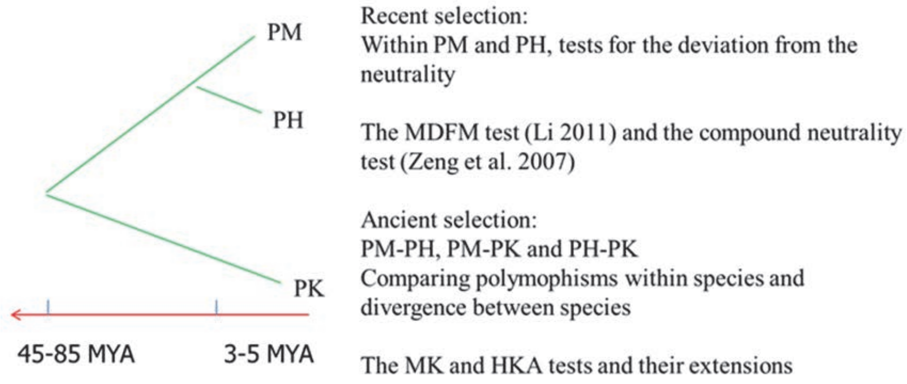
Population genetic statistics including the number of segregating sites (S), nucleotide diversity statistics (π, Nei 1987; θ_{w} , Watterson 1975), the number of haplotypes (K), and haplotypic diversity (H_{d}, Nei 1987) for all sites, silent sites and nosynonymous sites were computed to measure the diversity within the two species. Divergence statistics including average divergence per site for all sites (K_{T}) nonsynonymous sites (K_{a}) and silent sites (K_{s}), number of net nucleotide substitutions per site (D_{a}, Nei 1987) were used to indicate divergence between the two closely related species. Hierarchical genetic differentiation among populations within species (F_{ST}) and between both species (F_{CT}) at all three types of markers were used to measure genetic differentiation. The ratio of replacement and synonymous polymorphism ($\pi_{\mathrm{a}}/\pi_{\mathrm{s}}$) in each species and the ratio of replacement divergence and synonymous divergence (K_a/K_{s}) between both species and between each species and outgroup were calculated. All summary statistics were computed for each locus in the program DnaSP v5 (Librado & Rozas 2009).

Because LD patterns are related to demography and sometimes to selection. The level of LD was measured as the correlation coefficient r^2 (Hill & Robertson 1968) using informative sites. Indels and sites with three-nucleotide variants were excluded from the analysis. Under the mutation-drift-equilibrium model, the decay of LD with physical distance was estimated using non-linear regression of r^2 between polymorphic sites and the distance (in base pairs) between sites as detailed in Pyhajarvi *et al.* (2007). The nonlinear least-squares (nls) estimate of ρ ($\rho = 4N_ec$, where N_e is effective population size, c is the recombination rate) between adjacent sites was fitted by the nls-function implemented in the R statistical package (http: //www. r-project.org).

The genetic structure within and between the two closely related species was assessed at both candidate gens and reference loci using Structure v. 2.3.3 (Hubisz *et al.* 2009). To infer the structure of the sampled populations, the likelihood of each number of clusters, *K*, where 1 _ *K* _ 8, was assessed and allowance made for the correlation of allele frequencies between clusters under an admixture model. We performed 6 runs with 100,000 burnin steps and 1,000,000 iterations. Because each run give similar results, we just retained those with highest likelihoods.

2.3.2 Detecting signals of historical and recent climatic selection

As shown in Fig. 2 and in paper I, signals of natural selection were detected from three different evolutionary timescales: within each species (recent selection), between the two species (which diverged about 3 million years ago, MYA, Table 5 in paper I) and in the lineages leading to these two species (i.e. PM-PK and PH-PK), by comparing them to the outgroup (which diverged 45–85 MYA, Willyard *et al.* 2007).



Recent selection:
Within PM and PH, tests for the deviation from the neutrality

The MDFM test (Li 2011) and the compound neutrality test (Zeng et al. 2007)

Ancient selection:
PM-PH, PM-PK and PH-PK
Comparing polymophisms within species and divergence between species

The MK and HKA tests and their extensions

Fig. 2. Selection test at three different time scales.

Within the species, recent selection signals were detected using the DHEW neutrality test (A compound of Tajima's D, Fay & Wu's H and the Ewens-Watterson (EW) test (based on haplotype frequency spectrum, Watterson 1978). Zeng *et al.* 2007) conducted in the DH software (http://zeng-lab.group.shef.ac.uk/wordpress/?page_id=28) and Maximum Frequency of Derived Mutations (MFDM, Li 2011) tests in MFDM v1.1 software (http://www.picb.ac.cn/evolgen/softwares/softwares.html).

At longer evolutionary time scales, PM-PH, PM-PK and PH-PK in Fig. 2, ancient selection were detected by comparing diversity within species and divergence between the species among multiple-locus, using MK and HKA tests and their extensions (McDonald & Kreitman 1991; Bustamante *et al.* 2002; Hudson *et al.* 1987; Wright & Charlesworth 2004, detailed in Paper I).

2.3.3 Demographic histories

Neutral genetic diversity and recombination are related effective population sizes Ne. We compared genetic diversity at cp, mt and intronic sequences between the species (Paper II and III). We further used hierarchical ABC (detailed in paper III) for inference on demographic histories of each species and speciation model between the species (Fig. 1 in paper III). Within each species, we tested three "single-population models": the standard neutral model (SNM or Wright-Fisher model, Fisher 1930; Wright 1931) with the effective population size *N* as only parameter, a change of population size model (CNe) where the ancestral population size N_a changes to N at time T_c in the past, and a bottleneck model (Bnk) where N is reduced to N_b during $4N_0 \times 0.01$ generations at time Tb. Based on the "single-population models", we inferred the best speciation scenario amongst four "two-population models" (Figure 1b in Paper III), all derived from the 'classical' IM model (Nielsen & Wakeley 2001).

We sequenced a set of 37 loci distributed across the genome (paper I). Some of the loci showed deviation from standard neutral model in either or both of the species when we analyzed full sequences of the genes (Zhou *et al.* 2014). Here, we only used intronic sequences of the loci aiming at reducing the confounding effects of demography and natural selection. Four intronic loci showing Tajima's D values deviating strongly from zero, and therefore potentially affected strongly by selection, were excluded in ABC modeling. We discuss the potential effect of selection on demographic and speciation model inferences in paper III.

2.3.4 Species distribution modelling

Along with ABC inferences of speciation scenarios, we modeled the dynamics of changes in species distributions during the Last Glacial Maximum (LGM; ~21,000 years BP) and the Last inter-glacial (LIG; ~120,000 - 140,000 years BP). We used geographical information system (GIS) based methods with layers of historical and current environmental variables under the maximum entropy model implemented in maxent 3.3.3 (Phillips *et al.* 2006; Phillips & Dudik 2008). GPS data from 123 and 109 localities were used to generate species' distribution modeling of *P. massoniana* and *P. hwangshanensis*, respectively. We used 19 environmental variables from the WorldClim database with spatial resolutions of 2.5 m (Hijmans *et al.* 2005) for the current and the LGM and 1 km for LIG (Otto-Bliesner *et al.* 2008) as environment layers (detailed in Paper III).

2.3.5 ILS and introgression

Speciation with gene flow is common (Nosil 2008). However the process is complicated by introgression and incomplete lineage sorting, both processes generate similar patterns of shared genetic variation. We analyzed cp, mt and intronic genetic variation for both parapatric and allopatric populations of the two species. First, we compared patterns of genetic diversity, population admixture and LD patterns between parapatric and allopatric populations of both species. Additionally, we inferred the occurrence, proportions and dates of admixture of parapatric populations using allele frequency correlation and admixture-induced LD decay based analyses. We then conducted a hierarchical ABC analysis to infer the best demography and speciation scenario between the two species. Along with the demographic modeling, we used ecological niche modeling to track distribution changes of the two species during the Pleistocene climate change (paper III).

3 Results and discussion

I investigated the roles of climatic adaptation, demographic histories and gene flow in the speciation continuum of two closely related parapatric pine species in Southeast China using biogeography, population genetics and landscape genetics. Both species can occur at same mountains with different altitudinal preferences (parapatric populations) or at different mountains at species-specific altitudes (allopatric populations). Based on analyses of intra- and interspecific genetic variation at cp, mt and nuclear sequences, we made the following findings. Both recent positive selection and ancient balancing selection have contributed to climate adaptation in the two pines; Further, multiple SNPs showed significant correlations with altitudes (one of the most important diversifying ecological factors between the species) in the two species in the Bayenv analyses (Coop *et al.* 2010); Third, ancient climatic selection signals were mostly shared by the two lineages leading to the species because of shared evolutionary histories, but recent climatic selection signals were species-specific that select against maladapted migrants from the other species and played an important role in ecological speciation between the two pines; Further, We compared distributions of genetic variation between parapatric populations and allopatric populations considering incomplete lineage shorting (ILS) and introgression as dual causes of shared genetic variation. The results showed that mitotypes were highly shared with no difference between parapatric and allopatric population, supporting the ILS hypothesis; cp DNA types were highly species-specific; few climate adaptation related genes harbored fixed sites between species and shared nuclear genetic variation found more frequent in parapatric populations (supported introgression hypothesis). Considering different rates of gene flow (intraspecific) and mutation, genetic variation at mt, cp and nu genomes might mirror different phases of the speciation continuum between the species.

3.1 Climatic adaptation

Plant adaptation to different climates has recently became a major topic, due to potential effects of present global climate changes on agriculture and biodiversity (Manel *et al.* 2010). Studying climatic adaptation of coniferous trees has its potential to solve problems raised by global warming, since conifers play a dominant role in global carbon fixation (Canovas *et al.* 2007). This study may help to point to loci that partly govern the variation in climatic adaptation.

The d_N/d_S ratio in conifers (*Pinus*-*Picea*, diverged about 140 MYA, e.g. Wang *et al.* 2000) might be higher than that in angiosperms (*Arabidopsis*-*Populus*, diverged about 110 MYA, Bell *et al.* 2010) (Buschiazzo *et al.* 2012). We found high K_a/K_s ratios at candidate genes, but not at reference loci, between the two species studied here (section *Pinus*) and outgroup (*P. koraiensis*, section *Strobus*) (0.344, 0.360 for massoniana-koraiensis and hwangshanensis-koraiensis, respectively) with divergence time of 45 -85 MYA (Willyard *et al.* 2007). Between the two closely related species K_a/K_s was 0.401, (same subsection, diverged within 5 MYA) and π_a/π_s was also high within each species (0.438, 0.429 for massoniana and hwangshanensis, respectively). The climate related candidate genes might be firstly shaped by selection during the past climate change. Some of the candidate genes and one reference locus had K_a/K_s ratios above 1 (Paper I), meaning that the genes might be targeted by positive selection.

The genes indicated under selection using complementary methods were compatible in all three evolutionary scales. Within *P. massoniana*, two candidate genes (*AGP4-2*, *aqua-MIP*) and probably other loci, including eight candidate genes (*a3ip2*, *ccoaomt*, *dhn7*, *dhn9*, *GI*, *Glu*, *PHYO* and *Pod*) and two reference loci (*c3h* and *LHCA4*) were found to be influenced by recent positive selection. Within *P. hwangshanensis*, six candidate genes (*aqua-MIP*, *comt*, *dhn1*, *dhn7*, *Glu* and *PHYO*) and one reference locus (*PAL*) were identified under recent positive selection. Different genes of SNPs were targeted by recent positive selection in the two species (i.e. divergent selection, Paper I). In an analysis comparing the two species, seven candidate genes (*A3ip2*, *ccoaomt*, *dhn1*, *dhn7*, *LEA*, *PHYO*, and *RD21*) and two reference loci (*LHCA4* and *pt-1*) had some signals of selection. In a longer evolutionary scale, between each species and outgroup, similar sets of genes were ascertained under positive selection in the two species (*PHYO*, *dhn1*, *ccoaomt*, *Hlh1*, *Glu*, *POD*, *comt*), the sharing likely due to long common evolutionary history.

Candidate genes, such as *PHYO*, dehydrins (*dhn1* and *dhn7*) *AGP4-2* and *aqua-MIP* have been frequently detected to be under selection. Phytochrome O (*PHYO*): homologues of the Arabidopsis thaliana *PHYA* (Garcia-Gil *et al.* 2003), have a central role in mediating photoperiodic reactions (Smith, 2000; Quail, 2002). *PHYA* was indicated under selective sweeps in northern populations of *Arabidopsis lyrata* (Toivainen *et al.* 2014). Here, *PHYO* was detected under selection in three evolutionary time scales. Dehydrins play a major role in cell protection against desiccation. These proteins are produced in response to any type of stress that causes dehydration at the cellular level, such as cold, drought, or salinity (Close 1997). Arabinogalactan (*AGP4-2*) plays important role in cell wall formation, gene expression studies show downregulation by water stress in *P. pinaster* roots (Dubos *et al.* 2003). Aquaporins (*Aqua-MIP*) are membrane pore proteins (i.e. water channels) that play a critical role in controlling the water content of cells (Cruz *et al.* 1992). Interestingly, the same set of genes has also been found to be under selection in other conifers, such as Gonzalez-Martinez *et al.* 2006 in *P. taeda* (*aqua-MIP*), Eveno *et al.* 2008 (*AGP4-2* and *dhn1*) and Grivet *et al.* 2009 (*dhn1* and *aqua-MIP*), 2011 (*dhn2*) *in P. pinaster*, and Wachowiak *et al.* 2009 in *P. sylvestris* (*dhn1*). These genes might play important roles for climatic adaptation during climate oscillation in conifers.

3.2 Ecological speciation by divergent climatic selection

Species delimitation can be facilitated by the barrier to interspecific gene flow created by selection against maladapted immigrants (Gavrilets & Cruzan 1998). Intraspecific genetic diversity is lower while interspecific genetic divergence is higher at selection-targeted genes: *PHYO*, dehydrins (*dhn1* and *dhn7*), *AGP4-2* and *aqua-MIP* than those of other loci, thus, these genes show higher species identity. Two genes: *dhn1* and *aqua-MIP* can clearly delimit the two closely related pines. Furthermore, the NJ trees of these genes consist of two clear clades, one clade for each species, with limited immigrations (Supporting Figure S1 in paper I).

Whether a species maintains a cohesive unit depends on the homogenizing intraspecific gene flow connecting conspecific populations and efficient selection that prevents interspecific gene flow (introgression). Although all unlinked markers having the same mode of inheritance should experience similar rates of gene flow, selection can greatly modify the allelic frequencies within and between species. We detected multiple candidate genes under selection of climatic pressures, i.e. ecological adaptation. Interspecific gene flow at the targeted genes was limited. It seems selection was effective in preventing alleles from interspecific gene flow. Under the Isolation with migration model, climate related candidate genes show more limited migration than reference loci. The IM model in fact assumes neutral loci. The differences between candidate and neutral genes reflect the fact that some loci deviate from this assumption (Sousa *et al.* 2013).

Closely related pines always occur in ecologically isolated habitats with large population size. Environment related selection might be a major issue in evolving differentiation among closely related species. And efficient intraspecific gene flow might maintain the links among conspecific populations. The interaction between interspecific gene flow and selection is poorly investigated. But a negative relationship between intra- and interspecific gene flow was recently modeled (Currat *et al.* 2008). Intraspecific gene flow might dilute the introgressed alleles. Both divergent selection and intraspecific gene flow have negative effects on interspecific gene flow and play important role in conserving the homogeneity of species. A recent study compared genetic diversity at genomic islands of five pair of closely related species suggested that genomic islands of divergence are due to reduced diversity instead of reduced gene flow (Cruickshank & Hahn 2014). Our results also showed that both D_{xy} and F_{CT} were higher at climate-related candidate genes than reference loci, but our simulations based on IMa analysis showed lower rates of migration at candidate genes than at reference loci.

Association between the selection targeted candidate genes and environmental variation can be investigated in the future, not only species' range wide but also along the elevations. The speciation process of the two closely related pines is still at its early phase, the time since divergence is about $3N_e$ between both species. This is less than the 9-12 N_e required for full lineage sorting between allopatric populations (Hudson & Coyne 2002). Clearly, demographic history, adaptation, and gene flow should be jointly investigated.

3.3 Joint study demography and selection

We obtained the shape of the distribution of neutral diversity and divergence under the best demographic model and speciation scenario (SNM within each species and secondary contact speciation model, Paper III). Simulated distributions under neutrality were then compared with the observed values for the climate-related genes for several statistics (within species SFS Tajima's D and between species F_{CT}, and number of biallelic sites or in other words two allelic variants segregating in the species). Within both species, the candidate genes detected here is a subset of genes detected using the DHEW tests and the MDFM tests.

Table 1. Fitting of interspecific site frequency spectrum (Tajima's *D*) and intraspecific genetic divergence (F_{CT} and number of biallelic sites) at all candidate genes and reference loci to the posterior distributions from approximate Bayesian computation (ABC) demographic models with the highest posterior probability based on intronic sequences.

locus	Tajima's *D*_masso	Tajima's *D*_hwang	F_{CT}	nb of biallelic sites
PtIFG2009	-0.9634	-0.836	-0.8564	0.7286
a3ip2	-0.045*	0.7218	0.1271	0.0968
agp4	-0.0162*	0.5996	0.1551	0.5832
aqua-MIP	-0.003**	0.9944	0.0661	0.269
agp-like	0.6724	0.692	-0.7752	-0.6336
araH	0.2604	-0.1212	0.1806	0.0134*
araR	-1.1472	-1.138	-0.7484	-0.1904
c3h	-0.1176	0.8234	-0.9871	-0.7924
c4h2	-0.256	0.2746	0.1383	-0.8254
ccoaomt	-0.224	0.5752	0.3921	-0.02*
cesA2	-0.432	0.1134	0.0838	0.0972
dhn1	-0.0336*	-0.1288	0.1988	4e-04***
dhn2	0.0328*	-0.4194	0.0941	0.2136
dhn7	-0.2746	-0.707	0.4786	0.2274
dhn9	-0.9792	-0.9798	0.0747	-0.8024
erd3	0.3168	-0.7086	0.3833	-0.1276
GapCP	-0.1558	0.6204	0.1596	-0.1598
GI	-0.3428	-0.933	0.0257*	-1.058
Glu	-0.1322	-0.2656	0.0184*	0.6498
hlh1	-0.3546	0.5128	0.0381*	-0.0112*
Ino3	-0.7576	0.8946	0.226	-0.0306*
LEA	0.3942	-0.499	-0.751	0.8426
LHCA4	-0.3996	-0.3146	0.175	0.5282
lp3-1	0.6864	0.4804	-0.5128	0.0304*
MdhA	0.6112	-0.5242	0.1149	-0.6154
pho	-0.5614	-0.8292	-0.5	-0.1218
PHYO	-0.5196	-0.0562	0.0143*	-0.0992
Pod	0.8346	-0.1994	0.1632	-0.7738
pp2c	-0.454	0.8194	0.2641	0.526
ptlim-1	0.7572	0.135	-0.7299	-0.1046
ptlim-2	0.1082	0.1576	-0.9842	0.097
RD21A	0.3402	0.448	0.1582	0.0168*
WD40	0.9556	0.7116	-0.9829	0.7354
comt	0.9656	-0.2606	-0.5572	-0.6406
nir	0.8562	0.6246	-0.9123	-0.986
PAL	0.8846	-0.0766	-0.8526	-0.7846
ppap12	0.3064	0.021*	-0.7715	0.9838

*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$

Different genes showed deviation from the neutral distributions, which was in line with the results based on other detection methods (Paper I). Between the species, different loci were found to deviate from neutral distribution, except *hlh1* genes shared by both F_{CT} and number of biallelic sites. The genes detected by the HKA and MLHKA tests were a subset of genes detected here.

The methods take into account both demography and selection. However, previous studies have been mostly based on many loci of variable length which brings statistical difficulties and inaccuracies in simulations (Keller *et al.* 2011). Population genomic data based on next generation sequencing might overcome this difficulty. Sliding windows with different length along the chromosomes have the potential to shed light on this question.

3.4 Complicating histories of ILS and introgression in speciation with gene flow

Previously, we investigated cytoplasmic genetic differentiation between the two pines and found out that chrolotypes were highly species-specific while mitotypes were extensively shared (Zhou *et al.* 2010). As shown here, the nu DNA diversity seems to present another picture, i.e. both species shared 38% of polymorphic sites with 0.6% fixed sites (Paper III). As shared mitotypes are randomly distributed among parapatric and allopatric populations, the cause of shared mt DNA variation was thought to be the retention of ancestral polymorphism rather than introgression. However, the distribution of nuclear genetic diversity showed that shared polymorphism occurred preferentially between parapatric populations which therefore is best explained by a secondary contact scenario.

Such conflicts of genetic variation associated with different sex were also observed in other studies in other conifers (Lockwood *et al.* 2013), in bears (Hailer *et al.* 2012), in humans (Poznik *et al.* 2013) and in other organisms (Petit and Excoffier 2009; Toews & Brelsforo 2012). Most of them were explained by sex-specific migration and introgression patterns (Petit and Excoffier 2009; Hailer *et al.* 2012; Toews & Brelsforo 2012), but the lineage sorting process was somehow neglected.

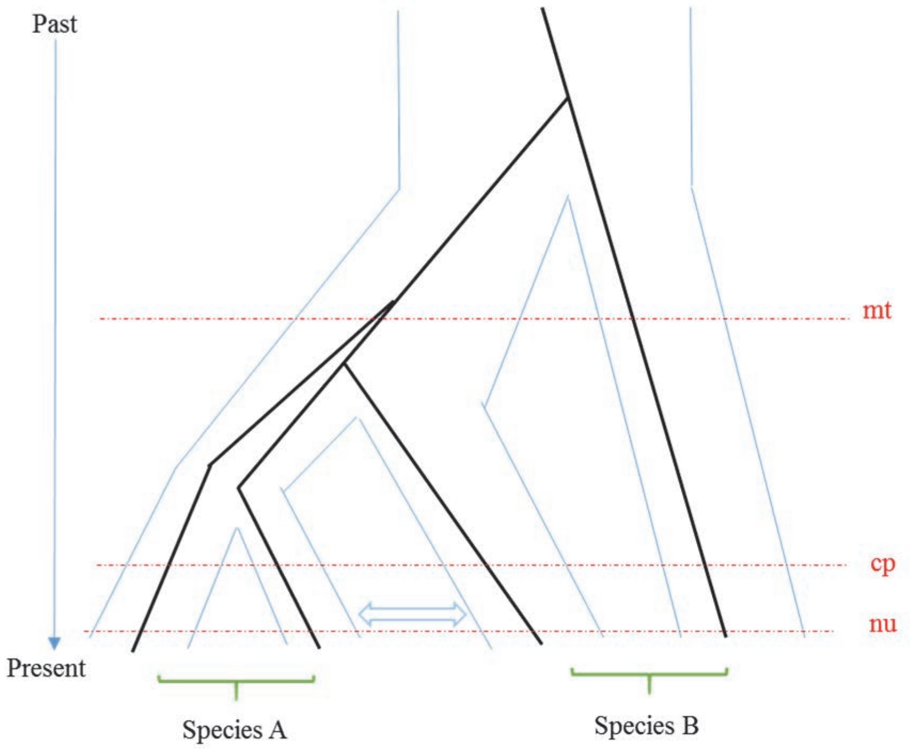


Fig. 3. The diagram shows divergence between species A and B with incomplete lineage sorting and secondary gene flow. Mitochondrial, chloroplast and nuclear genomic variation mirror at different phases of the speciation continuum.

Intraspecific gene flow has its roles in facilitating the rate of species lineage sorting (Zhou *et al.* 2010). DNAs with higher levels of intraspecific gene flow are more efficient in homogenize conspecific population gene pools than DNAs with lower levels of gene flow, which facilitate the fixation of species-specific advantageous alleles and even neutral genetic variation (Rieseberg *et al.* 2006). DNAs associated with different sex harbour different rates of intraspecific gene flow, and such gene flow has a role in diluting introgressed alleles from gene pools of closely related species (Petit & Excoffier 2009). In a simulation study, Currat *et al.* (2008) found that intraspecific gene flow was negatively related with interspecific gene flow. Consequently, DNAs with higher levels of intraspecific gene flow might have been less introgressed by foreign alleles and thus might show less shared polymorphism. In conifers, mitochondrial, chloroplast and nuclear genomes are maternally, paternally or bi-parentally inherited through seed, pollen or both, respectively (Wagner *et al.* 1987; Neale & Sederoff 1988; Mogensen 1996). In addition, pollen flow is more efficient in connecting conspecific populations than seed flow (Liepelt *et al.* 2002). The ratio of pollen flow to seed flow is about 20-50 based on investigations in four pine species (Ennos 1994). In mature forests, foreign alleles from pollen flow are easier than that from seed flow to be involved in the local gene pools. In agreement with these observations, cpDNAs show species specific genetic diversity whereas mtDNA genetic diversity is commonly shared among species (Du *et al.* 2009; Zhou *et al.* 2010; Wang *et al.* 2011).

In parallel to rates of gene flow, nucleotide substitution vary greatly among mitochondrial, chloroplast and nuclear genomes (Wolfe *et al.* 1987). The effective population size of cp, mt and nuDNAs is N, N and $2N$, respectively. Thus, different markers usually reflect evolutionary dynamics at different timescales. Among the three genomes of conifers, the mitochondrial and nuclear genomes harbour the lowest and highest rates of nucleotide substitution and gene flow, respectively. In practice, mitochondrial genetic variation has been used very efficiently to infer population structures and was found shared between closely related coniferous species, while chloroplast makers have shown week population structures but clear species boundaries (Petit *et al.* 2005; Du *et al.* 2009; Zhou *et al.* 2010). As shown in Fig. 3, the three kinds of DNAs from coniferous trees seem to provide sequential and complementary views of an evolutionary continuum, which therefore offer us a great study system in evolutionary biology.

4 Conclusions and future directions

The neutral demographic processes and selection at islands of the genome interact in the speciation continuum. I studied genetic variation at cytoplasmic and nuclear sequences of two closely related parapatric pine species: *P. massoniana* and *P. hwangshanensis*. I found that ancient selection was shared by the lineages leading to the species while recent selection was species-specific. Estimates of migration rates for candidate genes were significantly lower than that for reference loci. Recent differential climatic selection might counteract gene flow and favor divergence between the two pines through ecological speciation. Shared mitotypes were randomly distributed across species' ranges, which therefore supported the incomplete lineage sorting scenario, but the shared nuclear variation distributed more frequent at parapatric population than at allopatric population, supporting the gene flow scenario. ABC and species' distribution modeling also supported the secondary gene flow model. The three genomes had different rates of mutation and gene flow might mirror different phases of the speciation continuum.

The results in this thesis are valuable for understanding evolution in general and for other applied purposes such as tree breeding and climate change adaptation. Climate change drives speciation not only through shaping species' distributions to generate geographic isolation of populations (i.e. niche divergence), but also through divergent climatic selection facilitating the fixation of different beneficial alleles in populations in different environmental conditions (Hua & Wiens 2013). When populations of species recolonize their preferred habitats during interglacial periods, sometimes secondary contact and reinforcement favor reproductive isolation between them (Lukhtanov 2011). Coniferous trees have experienced a long history of climate change since the Carboniferous Period (360-286 million years ago, MYA, Miller, 1977). Both demographic processes caused niche divergence and divergent climate selection might have an influence of their own and interact during speciation of the two closely related pines, and other closely related coniferous species. The investigations based on cp, mt and nuclear sequences with different rates of gene flow and mutation seem to give complementary views of the speciation continuum. The early phases of speciation between the species has been found to show a picture of incomplete lineage sorting as inferred by mitochondrial markers while the later phases has been found to show a picture of secondary contact inferred by nuclear markers. Several climate related genes have been detected to be under local selection for climatic adaptation, for example to different altitudes of the species. These information can also be used for future landscape planning and breeding under the background of global climate change.

Further studies should take advantages of NGS technology and published genomic resources, such as loblolly pine reference genome (Neale *et al.* 2014) to study local adaptation and divergence along altitudes at different mountains, which could shed light on climate change adaptation, convergent adaptation and speciation.

List of references

Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. J Evol Biol 26: 229–246.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12: 1805–1814.

Aitken SN, Yeaman S, Holliday JA, Wang TL, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl 1: 95–111.

Alberto F, Aitken S, Alia A, Gonzáles-Martinez S, Hanninen H, Kremer A, Lefevre F, Lenormand T, Yeaman S, Whetten R, Savolainen O (2013) Evolutionary response to climate change—evidence from tree populations. Glob Chang Biol 19: 1645–1661.

Arnegard ME, McGee MD *et al.* (2014) Genetics of ecological divergence during speciation. Nature 511: 307-311.

Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. Mol Ecol 21(9): 2078-2091.

Barton NH, Bengtsson BO (1986) The barrier to genetic exchange between hybridizing populations. Heredity 57: 357–376.

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

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics 162: 2025–2035.

Beaumont MA (2005) Adaptation and speciation: what can FST tell us? Trends Ecol Evol 20: 435–440.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc London B 263: 1619–1626.

Bell C, Soltis D, Soltis P (2010) The Age and Diversification of the Angiosperms Re-revisited. Am J Bot 97: 1296–1303.

Brown GR, Gill GP, Kuntz R J, Langley CH and Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc Natl Acad Sci USA 101: 15255–15260.

Buschizzo E, Ritland C, Bohlmann J, Ritland K (2012) Slow but not low: genomic comparisons reveals slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. BMC Evol Biol 12:8.

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD *et al.* (2002) The cost of inbreeding in *Arabidopsis*. Nature 416: 531–534.

Canovas FM, Avila C, Canton FR, Canas RA, de la Torre F (2007) Ammonium assimilation and amino acid metabolism in conifers. J Exp Bot 58: 2307-2318.

Charlesworth B, Bartolome C, Noel V (2005) The detection of shared and ancestral polymorphisms. Genet Res 86:149–157.

Chen J, Kallman T, Gyllenstrand N, Lascoux M (2009) New insights on the speciation history and nucleotide diversity of three boreal spruce species and a Tertiary relict. Heredity 104: 3–14.

Chen J, Källman T, Ma X, Gyllenstrand N, Zaina G, Morgante M, Bousquet J, Eckert A, Wegrzyn J, Neale D, Lagercrantz U, Lascoux M (2012) Disentangling the roles of history and local selection in shaping clinal variation in allele frequencies and gene expression in Norway spruce (*Picea abies*). Genetics 191: 865–881.

Chen J, Tsuda Y *et al*. (2014) Clinal Variation at Phenology-Related Genes in Spruce: Parallel Evolution in *FTL2* and *Gigantea*? Genetics 197: 1025-1038.

Chiang YC, Hung KH, Schaal BA, Ge XJ, Hsu TW *et al*. (2006) Contrasting phylogeographical patterns between mainland and island taxa of the *Pinus luchensis* complex. Mol Ecol 15: 765–779.

Close T (1997) Dehydrins: a commonality in the response of plants to dehydration and low temperature. Physiol Plant 100: 291–296.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. Genetics 185: 1411–1423.

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

Cristescu M, Constantin A, Bock D, Caceres C, Crease T (2012) Speciation with gene flow and the genetics of habitat transitions in the *Daphnia pulex* species complex. Mol Ecol 21: 1411–1422.

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

Cruz RT, Jordan WR, Drew MC (1992) Structural changes and associated reduction of hydraulic conductance in roots of *Sorghum bicolour* L. following exposure to water deficit. Plant Physiol 99: 203–212.

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. Trends Ecol Evol 25: 410–418.

Currat M, Manuel R, Petit RJ, and Laurent E (2008) The hidden side of invasion: massive introgression by local genes. Evolution 62: 1908–1920.

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

de Leon LF, Bermingham E, Podos J, Hendry AP (2010) Divergence with gene flow facilitated by ecological differences: within-island locality in Darwin's finches. Phil Trans R Soc B 365: 1041–1052.

De Mita S, Thuillet AC, Gay L *et al.* (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol Ecol 22: 1383–1399.

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol Ecol 4:129–131.

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

Dvornyk V, Sirvio A, Mikkonen M, Savolainen O (2002) Low nucleotide diversity at the pal1 locus in the widely distributed *Pinus sylvestris*. Mol Biol Evol 19: 179–188.

Du KF, Petit RJ, Liu JQ (2009) More introgression with less gene flow: Chloroplast versus Mitochondrial DNA in the *Picea asperata* complex in china, and comparison with other conifers. Mol Ecol 18: 1396–1407.

Dubos C, Le-Provost G, Pot D, Salin F, Lalane C *et al*. (2003) Identification and characterization of water-stress-responsive genes in hydroponically grown maritime pine (*Pinus pinaster*) seedlings. Tree Physiol 23: 169–179.

Dumolin-Lapegue S, Pemonge MH, and Petit RJ (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. Mol Ecol 6:393–397.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. Mol Biol Evol 28: 2239–2252.

Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, Gonzalez-Martinez SC, Neale DB (2010) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). Genetics 185: 969– 982.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. Heredity 72: 250–259.

Evans LM, Slavov GT *et al*. (2014) Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. Nat Genet 46: 1089–1096.

Eveno E, Collada C, Guevara MA *et al*. (2008) Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. Mol Biol Evol 25: 417–437.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. Heredity 103: 285–298.

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

Fisher RA (1930) The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Foll M and Gaggiotti O (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977–993.

Fournier-Level A, Korte A, Cooper MD *et al*. (2011) A map of local adaptation in *Arabidopsis thaliana*. Science 334: 86–89.

Fu L, Li N, Elias TS. Eds. (1999) Pinaceae. In:Wu Z. and Raven PH. Flora of China, Volume 4. Beijing: Science Press; St. Louis: Missouri Botanical Garden Pp. 15–90 (http://www.efloras.org/).

Funk DJ, Egan SP, Nosil P (2011) Isolation by adaptation in *Neochlamisus* leaf beetles: host-related selection promotes neutral genomic divergence. Mol Ecol 20: 4671–4682.

Gamache I, Jaramillo-Correa JP, Payette S, Bousquet J (2003) Diverging patterns of mitochondrial and nuclear DNA diversity in subarctic black spruce: imprint of a founder effect associated with postglacial colonization. Mol Ecol 12:891–901.

García-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Mol Ecol 12: 1195–1206.

Gavrilets S, Cruzan MB (1998) Neutral gene flow across single locus clines. Evolution 52: 1277–1284.

Ge XJ, Hsu TW, Hung KH, Lin CJ, Huang CC *et al.* (2012) Inferring multiple refugia and phylogeographical patterns in *Pinus massoniana* based on nucleotide sequence variation and DNA fingerprinting. PloS one 7: e43717.

Gernandt DS, Lopez GG, Garcia SO, Liston A (2005) Phylogeny and classification of *Pinus*. Taxon 54: 29-42.

Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag single nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. Genetics 172: 1915–1926.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U *et al.* (2010) A draft sequence of the Neandertal genome. Science 328: 710–722.

Grivet D, Sebastiani F, Gonzalez-Martinez SC, Vendramin GG (2009) Patterns of polymorphism resulting from long-range colonization in the Mediterranean conifer Aleppo pine. New Phytol 184: 1016–1028.

Grivet D, Sebastiani F, Alia R, Bataillon T, Torre S *et al.* (2011) Molecular footprints of local adaptation in two Mediterranean conifers. Mol Biol Evol 28: 101–116.

Hailer F, Kutschera VE, Hallstrom BM, Klassert D, Fain SR *et al.* (2012) Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. Science 336: 344–347.

Hey J (2010) Isolation with Migration Models for More Than Two Populations. Mol Biol Evol 27: 905-920.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38: 226–231.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25: 1965–1978.

Hua X, Wiens JJ (2013) How does climate influence speciation? Am Nat 182: 1–12.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9: 1322–1332.

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116: 153–159.

Hudson RR, Coyne J (2002) Mathematical consequences of the genealogical species concept. Evolution 56: 1557–1565.

Huerta-Sánchez E, Jin X, Asan, Bianba Z *et al.* (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194-197.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. Nature 405: 907–913.

Jeong C, Alkorta-Aranburu G *et al.* (2014) Admixture facilitates genetic adaptations to high altitude in Tibet. Nat Commun 5: 3281.

Joosen RVL, Lammers M, Balk PA, Bronnum P, Konings MCJM, Perks M, Stattin E, Van Wordragen MF, van der Geest AHM (2006) Correlating gene expression to physiological parameters and environmental conditions during cold acclimation of *Pinus sylvestris*, identification of molecular markers using cDNA microarrays. Tree Physiol 26: 1297–1313.

Kawecki TJ and Ebert D (2004) Conceptual issues in local adaptation. Ecol Lett 7: 1225-1241.

Keller SR, Levsen N, Ingvarsson PK, Olson MS, Tiffin P (2011) Local selection across a latitudinal gradient shapes nucleotide diversity in balsam poplar, *Populus balsamifera* L. Genetics 188: 941–952.

Keller SR, Levsen N, Olson MS, Tiffin P (2012) Local adaptation in the flowering time gene network of balsam poplar, *Populus balsamifera* L. Mol Biol Evol 29: 3143–3152.

Kuhner MK (2009) Coalescent genealogy samplers: windows into population history. Trends Ecol Evol 24: 86–93.

Kujala ST, Savolainen O (2012) Sequence variation patterns along a latitudinal cline in Scots pine (*Pinus sylvestris*): signs of clinal adaptation? Tree Genet Genomes 8: 1451–1467.

Lenormand T (2002) Gene flow and the limits to natural selection. Trends Ecol Evol 17: 183–189.

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

Li HP (2011) A new test for detecting recent positive selection that is free from the confounding impacts of demography. Mol Biol Evol 28: 365–375.

Li Y, Stocks M, Hemmilä S, Källman T, Zhu HT, Zhou YF, Chen J, Liu JQ, Lascoux M (2010a) Demographic histories of four spruce (*Picea*) species of the Qinghai–Tibetan Plateau and Neighboring areas inferred from multiple nuclear loci. Mol Biol Evol 27: 1001–1014.

Li SX, Chen Y, Gao HD, Yin TM (2010b) Potential chromosomal introgression barriers revealed by linkage analysis in a hybrid of *Pinus massoniana* and *P. hwangshanensis*. BMC Plant Biol 10:37.

Li Z, Zou J, Mao K, Lin K, Li H, Liu J, Källman T, Lascoux M (2012) Population genetic evidence for complex evolutionary histories of four high altitude juniper species in the Qinghai-Tibetan Plateau. Evolution 66: 831–845.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinform 25: 1451–1452.

Liepelt S, Bialozyt R, Ziegenhagen B (2002) Wind-dispersed pollen mediates postglacial gene flow among refugia. Proc Natl Acad Sci USA 99(22): 14590-14594.

Lockwood JD, Aleksic JM, Zou JB, Wang J, Liu JQ, Renner SS (2013) A new phylogeny for the genus *Picea* from plastid, mitochondrial, and nuclear sequences. Mol Phylogenet Evol 69: 717-727.

Loh P-R, Lipson M, Patterson N, Moorjani P, Pickrell JK *et al*. (2013) Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. Genetics 193: 1233–1254.

Lorenz WW, Sun F, Liang C, Kolychev D, Wang H, Zhao X, Cordonnier-Pratt MM, Pratt LH, Dean JFD (2006) Water stress-responsive genes in loblolly pine (*Pinus taeda*) roots identified by analyses of expressed sequence tag libraries. Tree Physiol 26: 1–16.

Lukhtanov VA (2011) Dobzhansky’s rule and reinforcement of prezygotic reproductive isolation in zones of secondary contact. Biol Bull Rev 1:2-12.

Luo SJ, Zou HY (2001) Study on the introgressive hybridization between *pinus hwangshanensis* and *P. massoniana*. Scientia Silvae Sinicae 37: 118–122.

Ma XF, Szmidt AE, Wang XR (2006) Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. Mol Biol Evol 23: 807–816.

Ma XF, Hall D, StOnge K, Jansson S, Ingvarsson PK (2010) Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. Genetics 186: 1033–1044.

Mallet J (2007) Hybrid speciation. Nature 446: 279-283.

Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, Scribner KT, Bonin A, Fortin MJ (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol Ecol 19: 3760-3772.

Mayr E (1963) Animal Species and Evolution. Cambridge: Belknap Press of Harvard University Press.

McDonald JH and Kreitman M (1991) Adaptive protein evolution at the *ADH* locus in *Drosophila*. Nature 351: 652–654.

Michalak P, Minkov I, Helin A, Lerman DN, Bettencourt BR, Feder ME, Korol AB, Nevo E (2001) Genetic evidence for adaptation-driven incipient speciation of *Drosophila melanogaster* along a microclimatic contrast in “Evolution Canyon,” Israel. Proc Natl Acad Sci USA 98: 13195-13200.

Miller CN (1977) Mesozoic conifers. Bot Rev 43: 217–280.

Millar C (1998) Early evolution of pines. In: Richardson D, editor. Ecology and biogeography of *Pinus*. Cambridge (UK): Cambridge University Press. p: 69–91.

Mogensen HL (1996) The hows and whys of cytoplasmic inheritance in seed plants. Am J Bot 83: 383–404.

Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L *et al.* (2011) The History of African Gene Flow into Southern Europeans, Levantines, and Jews. PLoS Genet 7(4): e1001373.

Muller HJ (1939) Reversibility in evolution considered from the standpoint of genetics. Biol Rev Camb Philos Soc 14: 261–280.

Muir G, Schlötterer C (2005) Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (Quercus spp.) Mol Ecol 14:549–561.

Nadachowska-Brzyska K, Burri R, Olason PI, Kawakami T, Smeds L *et al.* (2013) Demographic Divergence History of Pied Flycatcher and Collared Flycatcher Inferred from Whole-Genome Re-sequencing Data. PLoS Genet 9(11): e1003942.

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Neale DB and Sederoff RR (1988) Inheritance and evolution of conifer organelle genomes. Pp. 251–264 in Genetic manipulation of woody plants. Plenum, New York, USA.

Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D *et al.* (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol 15: R59.

Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence-with-gene-flow in Tennessee cave salamanders (Plethodontidae; Gyrinophilus) inferred from gene genealogies. Mol Ecol 17: 2258–2275.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB *et al.* (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3: 976–985.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation. A Markov chain Monte Carlo approach. Genetics 158: 885–896.

Nosil P (2012) Ecological Speciation. Oxford University Press, Oxford.

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A, Members CLIP (2006) Simulating arctic climate warmth and ice field retreat in the last interglaciation. Science 311: 1751–1753.

Parisod C, Christin PA (2008) Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). New Phytol 178: 436–447.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N *et al.* (2012) Ancient admixture in human history. Genetics 192: 1065–109.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14: 689–701.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. Trends Ecol Evol 24: 386–389.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Modell 190: 231–259.

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31: 161–175.

Poznik GD, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, Snyder M, Quintana-Murci L, Kidd JM, Underhill PA, Bustamane CD (2013) Sequencing Y Chromosomes Resolves Discrepancy in Time to Common Ancestor of Males Versus Females. Science 341: 562–565.

Pyhajarvi T, Garcia-Gil MR, Knurr T, Mikkonen M, Wachowiak W, Savolainen O (2007) Demographic History Has Influenced Nucleotide Diversity in European *Pinus sylvestris* Populations. Genetics 177: 1713–1724.

Quail PH (2002) Phytochrome photosensory signalling networks. Nat Rev Mol Cell Bio 3: 85–93.

Richardson DM (2000) Ecology and biogeography of *Pinus*. Cambridge University Press.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. Nature 461: 489-494.

Reich DE, Patterson N, Campbell D, Tandon A, Mazieres S *et al*. (2012) Reconstructing Native American Population History. Nature 488: 370-374.

Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152:713–727.

Rieseberg LH, Wood TE, and Baack EJ (2006) The nature of plant species. Nature 440:524–527.

Rice W, Hostert E (1993) Laboratory experiments on speciation: what have we learned in 40 years? Evolution 47:1637–1653.

Ren GP, Abbott RJ *et al*. (2012). Genetic divergence, range expansion and possible homoploid hybrid speciation among pine species in Northeast China. Heredity 108(5): 552-562.

Rundle HD and Nosil P (2005) Ecological speciation. Ecol Lett 8: 336–352.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E *et al*. (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449: 913–918.

Savolainen O, Pyhajarvi T, Knurr T (2007) Gene flow and local adaptation in trees. Annu Rev Ecol Evol Syst 38: 595–619.

Smith H (2000) Phytochromes and light signal perception by plants-an emerging synthesis. Nature 407: 585–591.

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am J Bot 85:1301–1315.

Sobel JM, Chen GF, Watt LR, Schemske DW (2010) The biology of speciation. Evolution 64: 295–315.

Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. Annu Rev Plant Biol 60: 561–588.

Sousa VC, Carneiro M, Ferrand N, Hey J (2013) Identifying loci under selection against gene flow in isolation-with-migration models. Genetics 194: 211–233.

Sousa VC, Hey J (2013) Understanding the origin of species with genome-scale data: modelling gene flow. Nat Rev Genet 14: 404–414.

Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C (2013) Approximate bayesian computation. PLoS Comput Biol 9(1): e1002803.

Taberlet P, Cheddadi R (2002) Quaternary refugia and persistence of biodiversity. Science 297: 2009–2010.

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

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28: 2731-2739.

Tang K, Thornton KR, Stoneking M (2007) A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol 5:e171.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acids Res 24: 4876–4882.

Toivainen T, Pyhäjärvi T, Niittyvuopio A, Savolainen O (2014) A recent local sweep at the PHYA locus in the Northern European Spiterstulen population of *Arabidopsis lyrata*. Mol Ecol 23: 1040-1052.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol 21: 3907–3930.

Wachowiak W, Balk PA, Savolainen O (2009) Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold related candidate genes in Scots pine (*Pinus sylvestris* L.). Tree Genet Genomes 5: 117–132.

Wachowiak W, Palmé AE, Savolainen O (2011) Speciation history of three closely related pines *Pinus mugo* (T.), *P. uliginosa* (N.) and *P. sylvestris* (L.). Mol Ecol 20: 1729–1743.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Poppul Biol 7: 256–276.

Watterson GA (1978) The homozygosity test of neutrality. Genetics 88: 405-417.

Wagner DB, Furnier GR, Saghai-Maroof MA, Williams SM, Dancik BP, Allard RW (1987) Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. Proc Natl Acad Sci USA 84: 2097–2100.

Wakeley J, Hey J (1998) Testing speciation models with DNA sequence data. Pp: 157–175 in R. DeSalle and B. Schierwater, eds. Molecular approaches to ecology and evolution. Birkhauser Verlag, Basel.

Wang XQ, Han Y, and Hong DY (1998) A molecular systematic study of Cathaya, a relic genus of the Pinaceae in China. Plant Syst Evol 213:165–172.

Wang XQ, Tank DC, Sang T (2000) Phylogeny and divergence times in Pinaceae: Evidence from three genomes. Mol Biol Evol 17:773-781.

Wang B, Mao JF, Gao J, Zhao W, and Wang XR (2011) Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. Mol Ecol 20: 3796–3811.

Willyard A, Syring J, Gernandt DS, Liston A, Cronn R (2007) Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. Mol Biol Evol 24: 90–101.

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84: 9054–9058.

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

Wright SI and Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol 22: 506-519.

Wright SI and Charlesworth B (2004) The HKA test revisited: a Maximum-Likelihood-Ratio test of the standard neutral model. Genetics 168: 1071–1076.

Wu CI (2001) The genic view of the process of speciation. J Evol Biol 14: 851–865.

Wu CI and Ting CT (2004) Genes and speciation. Nat Rev Genet 5: 114–122.

Zeng K, Mano SH, Shi SH, Wu CI (2007) Comparisons of site- and haplotype-frequency methods for detecting positive selection. Mol Biol Evol 24: 1562–1574.

Zhao W, Meng J *et al*. (2014) Weak crossability barrier but strong juvenile selection supports ecological speciation of the hybrid pine *Pinus densata* on the Tibetan plateau. Evolution 68: 3120–3133.

Zhang D, Xia T, Yan M, Dai X, Xu J *et al*. (2014) Genetic Introgression and Species Boundary of Two Geographically Overlapping Pine Species Revealed by Molecular Markers. PLoS ONE 9(6): e101106.

Zhou WW, Wen Y, Fu J, Xu YB, Jin JQ, Ding L, Min MS, Che J, Zhang YP (2012) Speciation in the *Rana chensinensis* species complex and its relationship to the uplift of the Qinghai–Tibetan Plateau. Mol Ecol 21: 960–973.

Zou JB, Sun YS, Li L, Wang GN, Yue W, Lu ZQ, Wang Q, Liu JQ (2013) Population genetic evidence for speciation patter and gene flow between *Picea wilsonii*, *P. morrisonicola* and *P. neoveitchii*. Ann Bot 12:1829–1844.

Zhou YF, Abbott RJ, Jiang ZY, Du FK, Milne RI *et al*. (2010) Gene flow and species delimitation: a case study of two pine species with overlapping distributions in southeast china. Evolution 64: 2342–2352.

Zhou YF, Zhang LR, Liu JQ, Wu GL, Savolainen O (2014) Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. Mol Ecol 23: 3504–3522.

Original publications

I **Zhou YF**, Zhang LR, Liu JQ, Wu GL, Savolainen O (2014) Climatic adaptation and ecological divergence between two closely related pines species in Southeast China. Molecular Ecology, 23: 3504-3522.

II **Zhou YF**, Abbott RJ, Jiang ZY, Du, FK, Milne RI, Liu JQ (2010) Gene flow and species delimitation: a case study of two pine species with overlapping distributions in southeast China. Evolution, 64: 2342-2352.

III **Zhou YF**, Duvaux L, Zhang LR, Savolainen O, Liu JQ (2014) Incomplete lineage sorting versus introgression: origin of shared genetic variation between two closely related pines with overlapping distributions (manuscript).

Reprinted with permission from John Wiley and Sons.

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

ACTA UNIVERSITATIS OULUENSIS

SERIES A SCIENTIAE RERUM NATURALIUM

626. Langrial, Sitwat (2014) Exploring the influence of persuasive reminders and virtual rehearsal on the efficacy of health behavior change support system

627. Jaakkonen, Tuomo (2014) Intra- and interspecific social information use in nest site selection of a cavity-nesting bird community

628. Päätalo, Heli (2014) Stakeholder interactions in cross-functional productization : the case of mobile software development

629. Koskela, Timo (2014) Interaction in asset-based value creation within innovation networks : the case of software industry

630. Stibe, Agnis (2014) Socially influencing systems : persuading people to engage with publicly displayed Twitter-based systems

631. Sutor, Stephan R. (2014) Large-scale high-performance video surveillance

632. Niskanen, Alina (2014) Selection and genetic diversity in the major histocompatibility complex genes of wolves and dogs

633. Tuomikoski, Sari (2014) Utilisation of gasification carbon residues : activation, characterisation and use as an adsorbent

634. Hyysalo, Jarkko (2014) Supporting collaborative development : cognitive challenges and solutions of developing embedded systems

635. Immonen, Ninna (2014) Glaciations and climate in the Cenozoic Arctic : evidence from microtextures of ice-rafted quartz grains

636. Kekkonen, Päivi (2014) Characterization of thermally modified wood by NMR spectroscopy : microstructure and moisture components

637. Pietilä, Heidi (2014) Development of analytical methods for ultra-trace determination of total mercury and methyl mercury in natural water and peat soil samples for environmental monitoring

638. Kortelainen, Tuomas (2014) On iteration-based security flaws in modern hash functions

639. Holma-Suutari, Anniina (2014) Harmful agents (PCDD/Fs, PCBs, and PBDEs) in Finnish reindeer (*Rangifer tarandus tarandus*) and moose (*Alces alces*)

640. Lankila, Tiina (2014) Residential area and health : a study of the Northern Finland Birth Cohort 1966

641. Zhou, Yongfeng (2014) Demographic history and climatic adaptation in ecological divergence between two closely related parapatric pine species

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

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

ACTA UNIVERSITATIS OULUENSIS

ISBN 978-952-62-0675-2 (Paperback)
ISBN 978-952-62-0676-9 (PDF)
ISSN 0355-3191 (Print)
ISSN 1796-220X (Online)

UNIVERSITY of OULU
OULUN YLIOPISTO