

分类号:

密级:

兰州大学

研 究 生 学 位 论 文

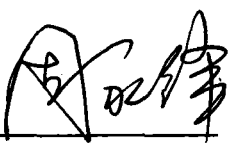
论文题目 (中文)	高山松复合体的气候适应
论文题目 (外文)	Local Adaptation to Climate in <i>Pinus densata</i> Complex
研 究 生 姓 名	周永锋
学 科 、 专 业	生态学、分子生态学
研 究 方 向	松柏类植物分子遗传育种
学 位 级 别	博 士
导师姓名、职称	刘建全 教授
论 文 工 作 起 止 年 月	2009 年 9 月至 2014 年 12 月
论 文 提 交 日 期	2014 年 10 月
论 文 答 辩 日 期	2014 年 12 月
学 位 授 予 日 期	

校址: 甘肃省兰州市

原创性声明

本人郑重声明：本人所呈交的学位论文，是在导师的指导下独立进行研究所取得的成果。学位论文中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人或集体已经发表或撰写过的科研成果。对本文的研究成果做出重要贡献的个人和集体，均已在文中以明确方式标明。

本声明的法律责任由本人承担。

论文作者签名： 

日期： 2014年12月10日

关于学位论文使用授权的声明

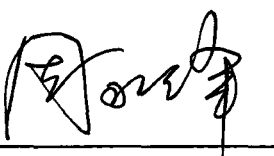
本人在导师指导下所完成的论文及相关的职务作品，知识产权归属兰州大学。本人完全了解兰州大学有关保存、使用学位论文的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人授权兰州大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本学位论文。本人离校后发表、使用学位论文或与该论文直接相关的学术论文或成果时，第一署名单位仍然为兰州大学。

本学位论文研究内容：

☐可以公开

☐不宜公开，已在学位办公室办理保密申请，解密后适用本授权书。

（请在以上选项内选择其中一项打“√”）

论文作者签名： 
日期： 2014年12月10日

导师签名： 
日期： 2014年12月12日

高山松复合体的气候适应

中文摘要

气候变化是人类面临的重大挑战。动物可以通过迁徙或者对新气候环境的适应,而植物则主要依靠本地气候适应来快速响应全球气候变化。松柏类植物经历了自从白垩纪早期(约 300 百万年)的全球气候动荡,它们适应了气候环境的动态变化并形成了现有的约 670 个物种,与之伴随的是种群大小的动态变化以及种群之间的反复隔离-接触过程。松柏类植物的气候适应和物种形成同时进行,因此,近源的松柏类植物是研究植物如何适应于气候变化,以及物种形成的理想材料。同时,由于松柏类植物在地球碳固定中的作用显著,研究松柏类植物的气候适应也为人类如何应对未来的气候变化提供依据。

本论文以高山松以及其可能的亲本种油松和云南松(高山松复合体)为研究材料。三个物种均为有着重要经济价值和生态意义的中国特有种。研究人员通过应用叶绿体,线粒体以及核基因遗传标记初步从遗传学的角度表明高山松可能是通过油松和云南松同倍体杂交形成的。然而这三个物种之间尚未形成完全的生殖隔离,物种间的人工授粉成功率很高。同时,由于它们领域分布,物种形成的谱系筛选过程中伴随着不同时期的种间基因流。高山松复合体的物种形成过程可能比目前所掌握的更复杂,将来应用全基因组中性变异,并采样尽可能多的自然居群,可能有助于我们更好的去模拟该物种形成过程。三个物种适应于不同气候环境,在歧化选择的作用下,适应性分化相应的基因组岛屿因连锁逐渐扩展至整个基因组,因此三个物种之间的分化也是一个生态物种形成过程。另一个更有趣的问题是:为什么高山松能够适应而其亲本种无法适应藏高原的寒旱气候环境?生理生态证据表明高山松在抗旱和抗寒的某些生理性状上优于其亲本种,然而它的分子基础至今尚未触及。

近代的本地选择和古代的平衡选择可能促进种群的本地气候适应。研究表明,在模式植物中,本地选择起着关键作用,然而松柏类植物经历了更久的全球气候变化,古代平衡选择也可能有其贡献,但目前的研究主要集中在近代选择的检测。本地适应相关的基因座位可能源于种群中的固有变异或者新产生的变异,在单个物种中很难区别二者,但在高山松中,我们可以在检测受本地选择影响的基因位点的同时确定位点是否也存在其亲本种中,从而区分基于固有变异和新变异的本地选择。高山松能适应于青藏高原的寒旱气候可能是归功于杂种优势,来源于两

个亲本种的基因之间或者基因调控之间的相互作用从而可能产生优于亲本的表型特征。

我们采集了覆盖三个物种分布范围的 68 个自然居群，每个居群 10-15 个个体。所有个体都对 4 个叶绿体和 2 个线粒体基因片段进行了测序，并同时对其 27 个居群的 25 个气候适应相关的候选基因和 12 个参照基因进行测序。应用群体遗传学和景观遗传学的方法，我们检测了三个物种与外类群所在谱系的古代自然选择，物种内的本地选择。通过比较适应性位点和中性位点的种间共享以及空间分布模式我们检测了来源于不同亲本的适应性渗入。同时我们比较了适应性，中性，父系和母系基因型的种内和种间分布模式，从而区分基于固有变异和新变异的本地适应，并检测父系和母系基因流在适应性渗入中的作用。全球气候变化影响物种分布范围内各个居群的适应性模式，尤其是边缘居群，我们也着重检测了边缘居群本地适应相关的基因位点。

结果表明近代的自然选择和古代的平衡选择同时作用于高山松复合体的本地适应。本研究在多个与气候相关的候选基因如 *agp4*, *aqua-MIP*, *dhn7*, *dhn9* 和 *lp3-1* 检测到了近代正选择的作用。在三个物种中，特异的基因座位受到了近代选择的影响而基本相同的一组基因受到了古代选择的影响，这可能是由于三个物种共享较长的进化历史的原因。在油松最南和最北的边缘居群中，不同的单核苷酸多态位点被检测到本地适应的信号，这些位点可能呈条件中性。高山松中检测到的适应性变异位点或与亲本共享或为高山松所特有，新的变异和固有变异同时贡献于高山松居群的本地适应，来自油松和云南松的祖先多态性或者适应性渗入促进了高山松适应于青藏高原的寒旱环境。叶绿体单倍型分布模式显示种内居群主要被花粉流连接起来，同时可能传递适应性变异。种间的适应性共享可能是由于种子流所导致的，但是这种共享也可能由于不完全谱系筛选所致。

总之，本论文促进了我们对进化生物学的理解，尤其是在全球气候变化的大背景下，物种如何适应于气候变化。同时，该研究检测到了多个与本地气候适应相关的基因可以用于分子标记辅助育种，也揭示了种群中性以及父系和母系遗传结构，为将来松树分子育种以及景观规划提供科学依据。

关键词：群体遗传学，景观遗传学，自然选择，气候适应，适应性渗入，条件中性，气候变化，松柏类

Local Adaptation to Climate in *Pinus densata* Complex

Abstract

Global climate change has been challenging the world. Animals can adopt migration and climatic adaptation while plants mainly depend on local adaptation to face climate change. Coniferous species have experienced past climate oscillations since the Carboniferous Period (360-286 million years ago, MYA). During this time, they have adapted to different climatic conditions and have given rise to new species while undergoing many demographic events, such as range and population size expansions, bottlenecks and variable gene flow. Conifers is therefore becoming a great study system for investigating how plants adapt to climate change. Moreover, conifers play a dominant role in global carbon fixation. Investigations of climatic adaptation in conifers have its potential to shed light on how people will act on future global warming.

I used *Pinus densata* from the Qinghai-Tibetan Plateau (QTP) and its potential parental species *Pinus tabulaeformis* and *Pinus yunnanensis* as a system to detect molecular signals of climatic adaptation in the species complex. Previous investigations suggested that *P. densata* might be a result of homoploid hybrid speciation between *P. tabulaeformis* and *P. yunnanensis*, using cytoplasmic and nuclear genetic variation. However, reproductive barriers between the species are still incomplete. As the species are distributed in parapatry, lineage sorting in the speciation continuum is complicated by introgression. The processes of speciation in *P. densata* complex might be more complicated than what previously known. The species have adapted to different environmental conditions by divergent selection shaping underlying genes. The genomic islands under local selection can extend gradually to the whole genome by linkage disequilibrium. This process is also in line with ecological speciation. Anyway, there is another interesting question about how *P. densata* can have better fitness than its parental species and can adapt to the drought and cold environmental conditions at the QTP. Physiological and ecological studies have suggested that *P. densata* had better fitness for drought stress and cold tolerance at some physiological traits than its parental species, but by far the genetic basis of the possible hybrid vigor is poorly known.

Both recent local selection and ancient balancing selection contribute to climatic adaptation. In model organisms, climatic adaptation have been studied using population genetic analyses on genome-wide data sets, which suggested that recent local selection played a major role. However, as conifers have a rich experience of past climate change, ancient balancing selection might also favored climate change adaptation in the species, but previous studies have been mainly focused on detecting molecular footprints of recent selection. Local selection for adaption use either new mutations or alleles from standing genetic variation, It is hard to distinguish between the two when the investigation is only based on single species. However, in the *P. densata* complex, we can detect whether the adaptive allele is presented in its parental species (standing genetic variation) or only specific in *P. densata* (new mutation). Heterosis or hybrid vigor might explain *P. densata* adapting to the drought and cold enviroments in the QTP, interactions between the alleles or gene regulations from both parental species may enable the species to harbor better fitness compared to its parental species.

We sampled 68 natural populations across the ranges of the three species, 10-15 individual for each population. We sequenced all the individuals using four cp and two mt gene fragments. We also sequenced 25 climate-related candidate genes and 12 reference loci for 27 populations chosen across the species' ranges. Using population genetic, landscape genetic and biogeographic approaches, we detected signals of ancient at the lineages of the species and the outgroup belonging to *strobis*, and recent selection within species especially for marginal populations. We then compared patterns of intra- and interspecific distribution of cp, mt, neutral nuclear and adaptive nuclear genetic variation to distinguish selection on new mutations and standing genetic variation, and to detect possible adaptive introgression, as well as to investigate the roles of cp and mt gene flow in adaptive evolution.

Our results showed that 1) both recent local selection and ancient balancing selection contribute to climatic adaptation in the species; 2) climate-related candidate genes, such as *agp4*, *aqua-MIP*, *dhn7*, *dhn9* and *lp3-1* were found to be under recent positive selection in *P. densata*; 3) recent differential selection but ancient shared selection signals were detected because of the shared evolutionary history between the species; 4) in line with conditional neutrality, different sets of SNPs were detect under local selection at the southern and northern marginal populations in *P. tabulaeformis*;

5) both new mutations and standing genetic variation have contributed to local climatic adaptation in *P. densata*; 6) adaptive introgression or beneficial ancestral polymorphisms from both parental species favors climatic adaptation of *P. densata* to the QTP; 7) co-specific populations have been linked by cp gene flow which might have also transferred adaptive alleles across populations. Shared alleles for climatic adaption between the species might have been caused by mt gene flow or incomplete lineage sorting.

The results in this thesis are valuable for understanding evolution in general, especially for forest genetics. The results can also be used for other applied purposes such as tree breeding and climate change adaptation.

Key words: population genetics, landscape genetics, natural selection, adaptive introgression, conditional neutrality, climate change, conifers

目 录

中文摘要.....	I
Abstract.....	III
目 录.....	6
第一章 前言.....	1
1.1 群体遗传学 (Population Genetics).....	1
1.1.1 哈迪-温伯格平衡(The Hardy-Weinberg Equilibrium).....	1
1.1.2 中性理论 (The Neutral Theory).....	2
1.1.3 溯祖理论 (The Coalescent Theory).....	3
1.1.4 自然选择 (Natural Selection).....	3
1.2 景观遗传学 (Landscape Genetics).....	6
1.2.1 地理隔离和生态隔离 (Isolation by Geography and Isolation by Ecology, IBD and IBE).....	6
1.2.2 频度梯度变异 (Clinal Variation).....	7
1.3 气候变化适应 (Adaptation to Climate Change).....	9
1.4 同倍体杂交物种形成 (Homoploid Hybrid Speciation).....	10
1.5 松柏类植物 (Coniferous Species).....	11
1.5.1 油松 (Chinese pine, <i>Pinus tabulaeformis</i> Carr.).....	12
1.5.2 高山松 (High mountain pine, <i>Pinus densata</i> Mast.).....	12
1.6 本研究的目的 (Study questions).....	13
第二章 材料方法.....	15
2.1 研究材料.....	15
2.1.1 油松.....	15
2.1.2 高山松.....	15
2.1.3 云南松.....	16
2.2 采样以及气候环境数据收集.....	16
2.2.1 采样.....	16
2.2.2 气候环境数据.....	18
2.3 测序.....	18
2.3.1 松树胚乳单倍体 DNA 提取.....	18
2.3.2 基因选取和引物设计.....	20
2.3.3 PCR 扩增, 产物纯化以及测序.....	20
2.4 数据处理.....	25
2.4.1 序列比对和注释.....	25
2.4.2 遗传多样性分析.....	25
2.4.3 连锁与重组.....	26
2.4.4 群体遗传结构.....	26
2.4.5 三个进化时间尺度检测选择信号.....	27
2.4.6 景观遗传分析.....	30

2.4.6.1 物种分布模拟(Species distribution modeling).....	30
2.4.6.2 环境因子和基因频率的关联分析.....	31
第三章 研究结果.....	32
3.1 核苷酸多样性 (Intraspecific genetic diversity)	32
3.2 种间分化 (Interspecific genetic divergence)	34
3.3 近代的本地选择信号.....	36
3.4 古代选择信号.....	38
3.5 基因流.....	39
3.6 物种分布模拟.....	40
第四章 讨论.....	43
4.1 核苷酸多态性.....	43
4.2 连锁不平衡.....	44
4.3 适应性渗入 (Adaptive introgression) 以及高山松适应于青藏高原的旱寒环境.....	45
第五章 总结与展望.....	48
参考文献.....	50
在学期间的研究成果.....	57
致 谢.....	58

第一章 前言

1.1 群体遗传学 (Population Genetics)

群体遗传学主要研究群体中等位基因频率的分布与变化, 主要因素包括自然选择(Natural selection), 遗传漂变(Genetic drift), 基因突变(Mutation), 基因流(Gene flow), 连锁重组(Linkage and Recombination), 交配系统(Breeding System), 群体历史(Demography)和群体结构(Population structure)等。群体遗传学研究试图解释自然种群的本地适应(Local adaptation) 和近缘物种间的物种形成(Speciation) 等进化过程。

群体遗传学是一门新兴学科, 从达尔文发表《物种起源》(Darwin 1859) 和孟德尔遗传定律(Mendel 1866)的发现开始, 到上世纪二十到三十年代, 该学科的奠基人: Sewall Wright, John Burdon Sanderson Haldane 和 Ronald Aylmer Fisher 创立了基本理论基础, 其中大部分目前仍然被广泛引用 (Provine 1971)。但是, 直到上世纪六十年代, 我们才获得等位酶和蛋白测序的群体数据, 后来到 1980s, DNA 测序技术被用于群体遗传学数据收集, 大量研究问题随之产生, 因此促进了该学科的大力发展。群体遗传学不同于其他生物学科, 长期以来, 它的理论规律的发展先于实验研究 (Veuille & Slatkin 2002)。但是, 随着测序技术的发展, 大量的群体基因组数据涌现, 群体遗传学的理论研究已远远赶不上实验研究的发展(Hedrick 2011)。我们需要更切合实际情况的理论模型, 同时也需要精确的计算生物学方法将变异位点从测序错误中区分出来。在大数据时代的背景下, 群体遗传学将迎来它的又一个黄金发展时期。

1.1.1 哈迪-温伯格平衡(The Hardy-Weinberg Equilibrium)

哈迪(Godfrey Hardy)-温伯格(Wilhelm Weinberg)平衡是二十世纪初期(1908年)被提出的, 后来被用于群体遗传学和数量遗传学。简言之, 该平衡认为在没有进化的情况下基因频率在群体中保持不变。假设某基因座位的等位基因 A 和 a 在群体基因库中基因频率分别为 p 和 q, $p + q = 1$, 那么下一代基因型 AA, Aa 和 aa 的频率分别为 p^2 , $2pq$, q^2 , $p^2 + 2pq + q^2 = 1$ 。应用哈迪-温伯格平衡可以计算一代到下一代的基因型频率。但是该平衡时建立在几个严格的条件之上的, 如群体无限大, 无自然选择, 无变异, 无迁入和迁出以及群体里的个体有相等的生殖。现实情况往往偏离这些假设, 但该原理仍然被广泛应用于群体遗传学中。比如, 在检测数量性状位点基因定位(QTL mapping)时, 往往需要先滤过偏离哈迪-

温伯格平衡的位点。用剩余位点进行进一步分析,但是这种滤过很可能把很关键的数量性状关联位点排除在外。在群体基因组中,沿着染色体,遗传多样性分别被连锁不平衡(Linkage disequilibrium)和哈迪-温伯格平衡从 X 轴和 Y 轴控制。

1.1.2 中性理论 (The Neutral Theory)

1968 年,日本学者木村(Kimura 1968)等根据蛋白质中氨基酸和核苷酸的置换速率,以及这样的置换造成核酸、蛋白质的改变并不影响生物大分子功能的事实,提出了分子进化的中性突变随机漂变学说(The neutral mutation-random drift theory),简称中性理论。该学说认为:生物大分子层次上的大量进化改变以及物种中的大多数变异不是由自然选择作用于有利突变而引起,而是在连续的突变压之下由选择上呈中性或近中性的突变等位基因随机漂变固定所造成的。进化是“中性突变”在自然群体中随机漂变的结果(Kimura 1968)。

中性学说的本质不是强调分子的突变型是严格意义上的选择中性,而是在于他们的命运在很大程度上是由随机遗传漂变所决定的(Kimura 1968)。它既可以解释种内存在的高水平的多态现象,也可以解释核苷酸或氨基酸取代随时间积累的事实,为分子遗传学提供了一种很重要很有价值的理论框架。成为目前被广泛接受的分子进化理论。

达尔文的进化论则认为通过自然选择淘汰不利变异,保留有利变异,即适者生存,使生物发展变化(Darwin 1859)。中性理论通过随机漂变决定保留和淘汰。从分子水平来看,大多数的变异不引起蛋白质序列的改变,呈中性变异,从个体和群体水平来看,某些类群产生特定的功能去适应其生境,这些特征对应的基因组片段也被称为基因组岛屿(Genomic islands),它们可能只占基因组上的很小的一段或者几段区域。自然选择理论跟中性理论二者并不矛盾,更不对立,只是各自的侧重不同。前者侧重于个体、群体即宏观水平。另一个侧重于分子即微观水平,二者相互补充,更好的揭示了生物进化的现象和本质。在现实操作中,我们往往以基因组变异作为中性背景去检测可能受自然选择的基因组区域。

随着群体基因组研究的发展,越来越多的证据表明每个核苷酸变异都有其功能意义,非编码区的变异可能通过改变分子结合特性而改变基因调控网路(The ENCODE Project Consortium 2012)。某些适应性特征不仅可能是通过氨基酸序列的改变引起的,某些非编码的‘沉默的’突变可能改变基因的转录调控,从而改变基因量或表达模式,进一步改变物种的适应性。因此,在大数据时代的今天,中性理论可能需要进一步的修饰。但是这种非编码区的功能性突变可能只集中在很

小的适应性基因组岛屿，大多数基因组区域可能还呈现遗传中性。

1.1.3 溯祖理论 (The Coalescent Theory)

随着中性理论的进一步发展，溯祖理论于 1982 年应运而生 (Kingman 1982)。溯祖理论基于 Wright-Fisher 模型 (Fisher 1930; Wright 1931) 应用统计学方法回溯基因变异过程，去回答是什么进化过程导致的我们掌握的种群序列变异数据。它是一种基于序列数据的群体遗传过程的分析方法。溯祖理论与系统树 (Genealogical trees) 不同，系统树描述一段 DNA 序列的历史，而溯祖理论揭示群体进化过程，例如有效种群 (Effective population size, N_e ，在突变和漂变的平衡时，群体基因多样性： $\theta = 4N_e\mu$ ， μ 代表变异速率) 大小的改变：瓶颈效应或种群扩张。基于溯祖理论，遗传漂变是两段等位 DNA 序列在上世代共享最近共同祖先 (The most recent common ancestor, MRCA) 的几率。最近共同祖先描述了居群中 n 条序列的最初情形，从 MRCA 开始，群体历史按照时间推移而前进。由于突变，后代序列之间的分化越来越多，这个过程的相反过程就是溯祖。溯祖的方法被广泛用于群体历史的模拟，例如用 ms (Hudson 2002, 一个很经典的模拟软件) 可以输出特定参数 [如有效种群大小 N_e 和代数 (generation) 时间 T] 先验分布 (Prior) 下的群体数据。

近年来，溯祖方法得到了更进一步的发展，应用最大似然法 (如 IMA, Wakeley & Hey 1998; Nielsen & Wakeley 2001; Hey 2010) 或者贝叶斯谱系采样法 (Approximate Bayesian Computation, ABC) (Beaumont *et al.* 2002; Kuhner 2009; Csilléry *et al.* 2010; Sunnåker *et al.* 2013) 进一步研究种群历史和物种形成过程。

1.1.4 自然选择 (Natural Selection)

虽然自然选择被认为是进化过程的主要动力，但是衡量其具体作用往往是一个复杂的过程。因为不同生物的适应性特征各异，比如，植物的开花时间，种子质量和数量，动物的个体大小和疾病状况。假设一个居群中的个体有特定数量的基因变异，简单来说，基因变异+环境变异→表型变异→适应变异，所有这些适应性因素被叫做相对适应性 (Relative fitness)，它指的是不同基因型传到下一代的相对能力。1930 年，Ronald Fisher 发表了 *The Genetical Theory of Natural selection*，这被群体遗传学家认为是达尔文选择和遗传学的一个里程碑。该书奠基了达尔文选择的遗传学理论基础。由于表型的改变往往是渐进的过程，需要多代来产生，同时，特定表型往往被多基因和多环境因子调控，理解表型适应通常很困难。到目前为止，仍有很多问题亟待回答，比如：特定适应性特征是单基因，

还是多基因控制的？是新的基因突变还是固有变异引起的某一特征的种群适应？等等。

在分子水平，选择信号的检测大都基于中性检测，中性理论建立在突变-漂变平衡的基础上，违背中性理论往往被看作是自然选择的作用。在实际操作中，在分子水平检测选择信号会被种群历史的混杂信号复杂化(Wright & Gaut 2005)。然而，种群历史影响整个基因组的基因多样性模式，而选择只作用于基因组上的特定区域(Genomic islands)。不同的中性选择方法可以用来检测不同进化时间尺度的自然选择信号(图 1-1, Niesen 2005; Hohenlohe *et al.* 2010)。显著低于基因组背景的基因多样性(Genetic diversity) + 显著高于基因组背景的居群间分化(Genetic differentiation, F_{ST}) + 显著高于基因组背景的连锁不平衡水平(Linkage Disequilibrium, LD) = 强的正选择信号(Strong positive selection or selective sweep)(Niesen 2005)。在一个种群内，近期的自然选择可以用变异位点频谱(Site frequency spectrum, SFS, Tajima 1989)，居群间分化水平(F_{ST} , Beaumont 2005; Foll & Gaggiotti 2008)以及基因连锁水平进行检测(图 1-1, Niesen 2005; Hohenlohe *et al.* 2010)。而在一个物种谱系内，多个近亲物种之间的选择信号可以通过比较种内基因多样性和种间分化水平进行检测(图 1-1)。例如 Hudson-Kreitman-Aguad'e (HKA) 检验(Hudson *et al.* 1987)，McDonald-Kreitman (MK) 检验(McDonald & Kreitman 1991)以及它们的扩展：最大似然 HKA 检验[the maximum likelihood HKA tests (Wright & Charlesworth 2004)]和 MK 泊松随机场检验[the McDonald-Kreitman Poisson Random Field (MKPRF) test (Bustamante *et al.* 2002)]。

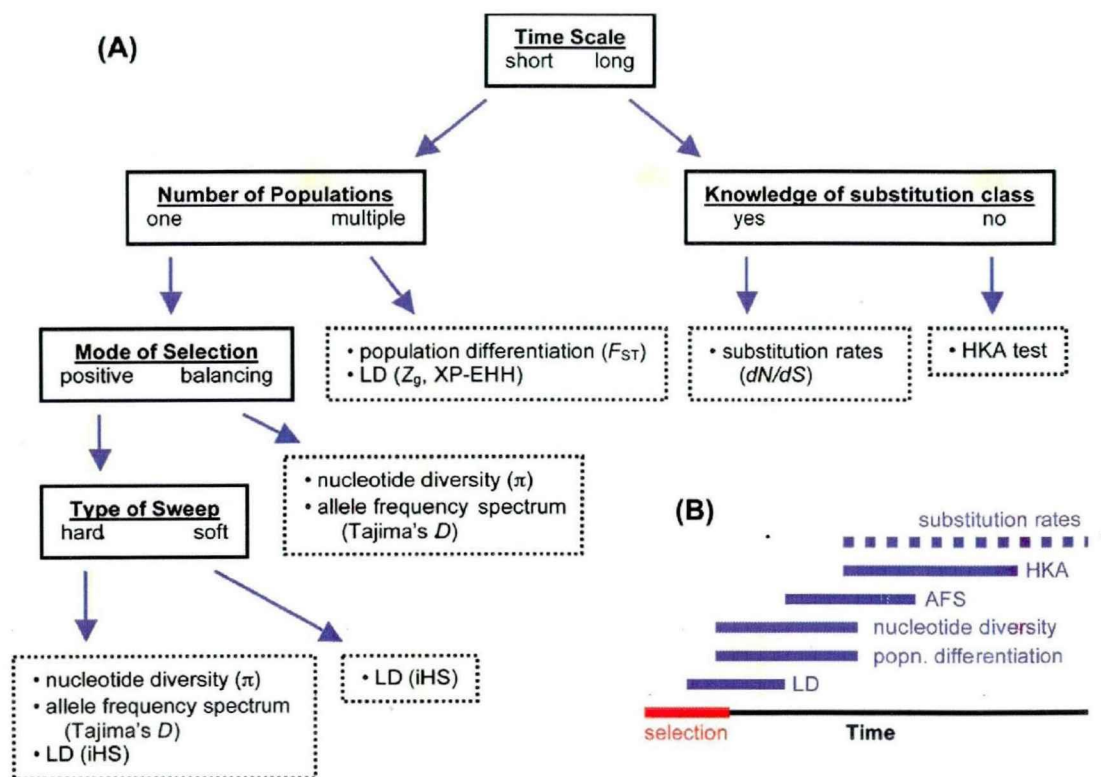


图 1-1 (来自 Hohenlohe *et al.* 2010, *International Journal of Plant Science*) A 图, 实线框显示生物学中应用群体基因组学去检测选择的各种情况, 虚线框显示对应情况下最适合的检测方法。B 图显示各种检测方法所检测的不同时间尺度的选择信号。

F_{ST} 异常值法 (F_{ST} outliers) 被广泛用在群体基因组数据中去检测可能的近期选择信号。显著高于基因组背景的 F_{ST} 值可能因为正选择的作用, 而显著低于基因组背景的 F_{ST} 可能是平衡选择所引起的。基因组扫描的方法主要有两种: 一种是基于岛屿模型的 Bayescan (Foll & Gaggiotti 2008), 另一种是分层岛屿模型 (Hierarchical island model, Excoffier *et al.* 2009) 方法。前一种方法适合无多层遗传结构的种群, 或者两个群体之间, 而后一种则适合有复杂种群遗传结构的物种。但是, 有时候选择可能只作用于某一个居群, 比如边缘居群往往受到很大选择压力。这种情况下, 我们可以用另一种叫做居群特定分化 (F_{FT} , differentiation between a focal population and all the other populations, Yang 1998) 来检测特定居群的本地适应。最近发表的 hapFLK 综合了 F_{ST} 和 LD 信息, 建立在分层岛屿模型之上。它能够更好用于检测本地选择信号 (Fariello *et al.* 2013)。

1.2 景观遗传学 (Landscape Genetics)

景观遗传学是一门综合应用群体遗传学,景观生态学和空间统计学的方法去研究环境异质性和环境变化对动植物群体的基因流和基因多样性分布的影响的新兴学科。它在生物多样性保护,入侵动植物防治和景观生态规划以及濒危动植物的保护管理等方面有着很广泛的应用前景。简言之,目前景观遗传学的研究主要有如下三个方面:

1. 基于 GIS 技术和现有气候环境数据的物种潜在分布区域的模拟。如最大熵模型 (Maximum entropy model) (Phillips *et al.* 2006; Phillips & Dudik 2008)。该方法可应用在农林业规划以及生物多样性保护中。
2. 探讨生态隔离和地理隔离对基因流和基因多样性分布的影响 (IBE 和 IBD,例如 Wang *et al.* 2013)。这对于回答生物多样性以及物种形成等一些基本问题有重要意义。
3. 在全球变化的背景下,景观遗传学的一个主要挑战是研究适应性变异如何在景观各异的生境之间传播。结合基因组扫描 (Genome Scan,如 Foll and Gaggiotti, 2008) 和基因频率沿着环境梯度或在不同生境类型而变化去识别和特定环境因子的显著相关的变异位点 (Clinal variation, Barton 1999; Savolainen *et al.* 2013)。对适应性变异的景观遗传学研究可能回答群体如何响应全球气候变化,以及物种对气候的适应性进化能否赶上全球变暖的速度等问题?

1.2.1 地理隔离和生态隔离 (Isolation by Geography and Isolation by Ecology, IBD and IBE)

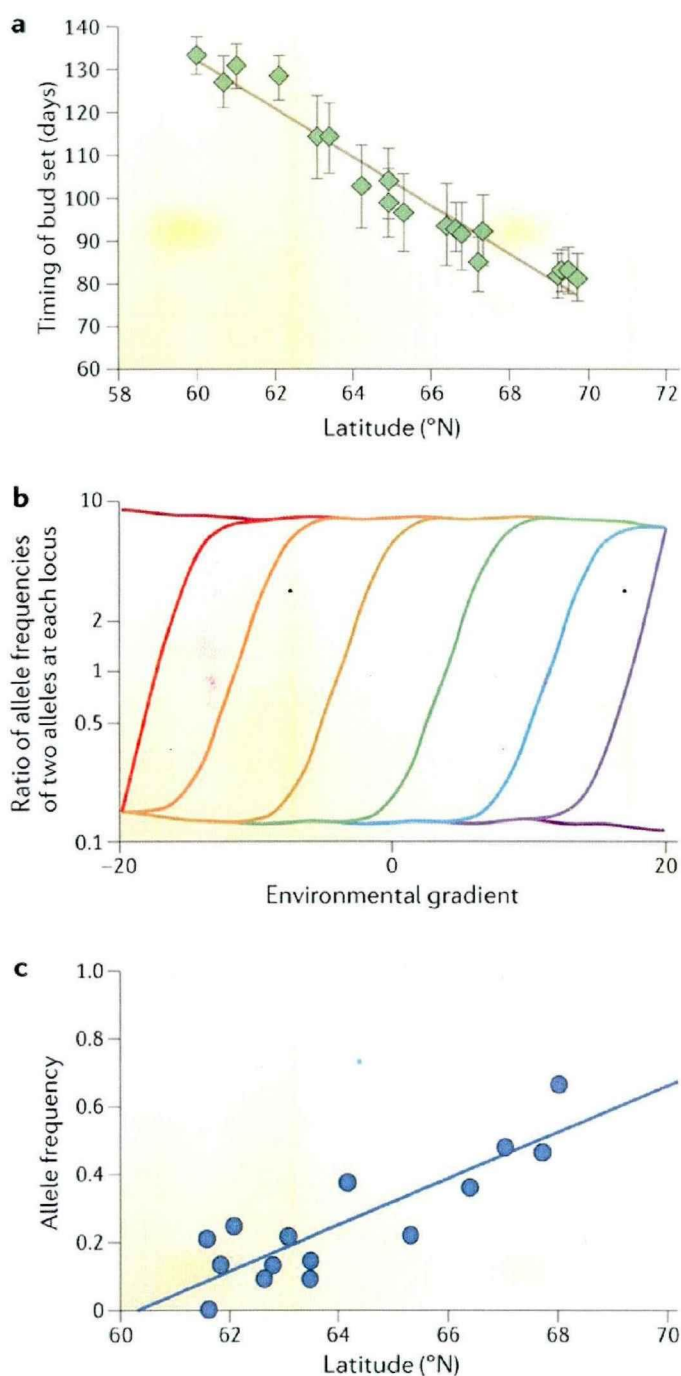
认识促进种群遗传分化的因素是生态学和进化生物学的一个长期目标 (Wright 1921; Mayr 1963; Coyne & Orr 2004)。遗传变异模式常常揭示由两种景观生态方式引起的空间变异的基因流: 1) 空间隔离的群体可能经历地理隔离 (Isolation-by-distance, IBD, Wright 1943), 景观障碍或者是地理距离减少了基因交流; 2) 不同生态环境中的群体之间的基因流因受到自然选择和本地适应的限制 (Dobzhansky 1936), 即生态隔离或者环境隔离 (Isolation-by-Environment, IBE, Wang & Summer 2010)。地理隔离预测地理距离跟遗传分化水平相关, 因为复杂的地理结构或者距离有可能阻断群体间个体的相互扩散。而生态隔离预测生态差异程度跟遗传分化水平相关, 因为生态差异越大, 受到的趋异选择压力越大

(Crispo *et al.* 2006; Lee & Mitchell-Olds 2011)。当然，地理隔离跟生态隔离不是互相排斥的，两者可能一起导致空间变化的基因流模式(e.g. Coyne & Orr 2004; Crispo *et al.* 2006; Thorpe *et al.* 2008)。地理隔离跟种群历史相关并可能导致种群间的生殖隔离，引起异域物种形成(Allopatric speciation)，而生态隔离跟本地适应性相关。

长期以来，研究所集中在研究地理隔离导致的遗传分化，而生态隔离是最近被提出的，它可能在物种形成中起重要作用(Ecological Speciation, Nosil 2012)。同时研究 IBD 和 IBE 对遗传分化的影响有助于我们理解景观和环境是如何影响遗传变异的空间分布的，同时可能为景观设计和动植物保护提供参考。

1.2.2 频度梯度变异 (Clinal Variation)

适应性特征跟生物或者非生物因子的相关性是由于空间变化的自然选择(Spatially varying natural selection)引起的本地适应性，而不是由于种群历史过程的结果，证明这一点很重要。频度梯度变化的变异往往是由选择所导致的。例如果蝇个体大小的纬度梯度，果蝇的生理节律，许多植物的生物气候特征（比如开花时间和芽点产生）。但是在复杂环境中，往往很难区分选择和种群历史的作用，在不同的物种中，难度不同。比如在随机交配的拥有巨大种群的物种中，例如某些松柏类，频度梯度很强，因为受到了更有效自然选择作用（图 2-1, $\gamma = 2NeS$, 其中 Ne 是有效种群大小， S 是选择系数。e.g. Gossmann *et al.* 2012）。模拟显示适应性等位基因频率沿着环境梯度在很小地理范围内快速变化（图 2-1b, Barton 1999），然而在挪威云杉中，研究人员发现等位基因沿着纬度梯度在很大的地理范围内变化(图 2-1c)。我们可能需要更切合实际的模型去模拟等位基因频度梯度变异的模式和过程。



Nature Reviews | Genetics

目前较常用的识别频度梯度变化的方法有两种：基于贝叶斯模型的 Bayenv (1.0 版, Coop *et al.* 2010; 2.0 版, Gunther & Coop 2012) 和潜在因素混合模型 (latent factor mixed model, LFMM, Frichot *et al.* 2013)。二者均可以把选择作用从种群历史的背景中分离出来,从而用于发现适应性等位基因的频度梯度变化。Bayenv 方法需要根据其功能把序列数据分为两类,其中跟感兴趣的适应性特征无关一类(参考基因或者基因组背景)用于计算群体间的相关关系 (Population history), 这个群体关系矩阵将被当做背景用于选择信号的检测。而跟该特征相关的一类(候选基因, 基于基因功能分析的先知)将用于群体矩阵的基础上检

图 1-2 表型和基因型的频度梯度变化。(来自 Savolainen *et al.* 2013, Nature Reviews Genetics) 其中图 a 显示欧洲赤松芽点萌发时间与纬度梯度显著相关 (Mikola

1982) ; 图 b, 理论模拟相关位点等位基因频率沿梯度在很小的地理范围内快速变化 (Barton 1999) ; 但是, 图 c, 在挪威云山中, 候选基因的的等位基因频率在很大的地理范围内沿着纬度梯度变化 (Chen *et al.* 2012) 。

测选择信号。而 LFMM 方法则不用将基因分类。因为在大多数非模式之物中, 基因功能尚不清楚; 我们所掌握的仅有少数适应性特征的基因调控网络, 例如生

物钟(Lankinen1986), 动物某类营养物质新城代谢以及植物开花(Olsson & Ågren 2002)等; 大量证据表明基因调控可能对本地适应很重要, 比如转录因子结合区(Transcription factor binding sites)加强子(enhancer)和启动子区(promoter)的变异所导致调控因子的结合性质改变, 从而很大程度的改变基因产物和相关基因功能网络, 这些调控区往往在基因之外, 应用分类法, 可能排除了很重要的选择信号(Kasinathan *et al.* 2014)。这种情况下, 应该用 LFMM 的方法。通过比较关联分析和基因分化异常值的结果, 得到潜在的气候适应的基因变异。

1.3 气候变化适应 (Adaptation to Climate Change)

全球气候变化快速变化威胁着生物多样性。在过去的三十年中, 大量物种的分布由于气候变化而发生变化, 到 2050 年, 大约有 15%-36%的物种将趋于灭绝(Thomas *et al.* 2004)。因为他们不能适应新的气候环境, 生境片段化(Fragmentation)或者适宜生境逐渐减小。某些裸子植物拥有广泛的连续分布, 由于气候变暖, 北方边缘居群往往能够得到来自最近南方居群的基因流协助(Assisted gene flow, 携带对温暖适应的等位基因)而南方边缘居群则得不到这种适应性基因流的协助(Savolainen 2011; Aitken *et al.* 2013)。所以对南方边缘居群, 应采取迁地保护等策略等其他方法。

通过识别跟环境变化相关的适应性变异, 我们可以更好的理解群体如何适应于它所在的栖息地环境, 同时预测物种将如何应对未来的环境变化(Hoffman and Sgro 2011; Franks & Hoffmann 2012)。我们可以通过比较种内遗传多样性和物种间分化水平来检测历史气候变化对生物进化的长期效应。通过比较基因多样性, 群体间分化水平, 变异位点频谱以及连锁不平衡模式, 我们也可以看到近期的, 种群内的气候变化适应。考虑到种群历史和自然选择都可以引起上述参数偏离中性预期。群体基因组数据可以有效的区分背景噪音和可能的适应性变异。科学家们在蒺藜状苜蓿(*Medicago truncatula*)群体基因组数据发现气候适应是多基因控制的, 这些基因受到较弱的正选择压力(Soft sweeps)而不是长期的平衡选择(Balancing selection)(Yoder *et al.* 2014)。但是, 在裸子植物中, 我们没有可用的参考基因组。另一种被林业遗传学家应用的方法是去比较一组跟适应性特征相关的候选基因(Candidate genes)和另一组跟该特征不相关的参考基因(Reference genes, 如 Keller *et al.* 2013; Kujala & Savolainen 2012; Chen *et al.* 2012)。

气候变化是森林树种适应性进化的最重要的驱动力之一(Aitken *et al.* 2008; Richardson *et al.* 2009; Alberto *et al.* 2013)。在好多森林树种中, 研究人员发现了

候选基因的等位基因频率跟气候因子梯度显著相关。例如芽点的出现和对寒冷的适应 (Holliday *et al.* 2010; Olson *et al.* 2013; Chen *et al.* 2012), 与降雨, 干旱和温度相关 (Bower & Aitken 2008; Eckert *et al.* 2010; Ma *et al.* 2010; Prunier *et al.* 2011; Prunier *et al.* 2012; Hall *et al.* 2011; Keller *et al.* 2011; Mosca *et al.* 2012)。在另一些广布于不同气候条件的生境中的树种中, 研究人员识别了作用于气候适应相关基因的选择信号(如 Gonzalez-Martinez *et al.* 2006; Pyhäjärvi *et al.* 2007; Eckert *et al.* 2009; Grivet *et al.* 2009; Wachowiak *et al.* 2009)。

同时, 基于中性遗传的分子标记和溯祖方法去研究区域内动植物种群的遗传结构、组成和有效种群历史动态变化过程不但对于理解过去该地区环境变化历史具有重要价值(Hewitt 2000), 而且为全球气候变化背景下的生物多样性保护, 以及物种如何通过改变分布模式来响应未来气候变化提供依据。历史气候变化, 尤其是第四纪冰期-间冰期的震荡导致物种退缩至避难所, 而在间冰期, 从这些避难所发生了多次回迁或扩张到物种目前的分布区, 这些反复迁移和扩张共同塑造和奠定了现代物种格局和遗传多样性分布式样(Abbott *et al.* 2000; Avise 2000)。

考虑到全球气候的快速变化, 综合动植物适应性进化以及种群动态历史的研究将可能为生物多样性保护以及生态规划等领域提供科学依据, 这也是景观遗传学的主要研究内容。

1.4 同倍体杂交物种形成 (Homoploid Hybrid Speciation)

同倍体物种形成指的是通过杂交产生同倍体新物种的过程。因为杂种一般被认为具有居中的生态性质 (Coyne & Orr 2004) 很难形成与其亲本种之间的生态隔离, 所以之前科学家认为同倍体杂交形成的物种很稀少。但是, 在杂交过程中, 通过基因重组可能产生优于亲本的基因型, 从而使得杂种能够迁移到亲本种无法生存的生态环境并促进杂种和亲本种之间的生殖隔离 (Gross & Rieseberg 2005)。同时, 由于杂交导致的染色体重组可能促进杂种和亲本种之间的生殖隔离 (Rieseberg *et al.* 1995; Rieseberg 1997)。

高山松被认为是杂交物种形成的一个典型例子, 在高山松由云南松和油松杂交产生的过程中, 基因重组赋予了高山松能够适应亲本种之外的青藏高原的极端环境 (Wang & Szmidt 1994; Wang *et al.* 2011; Gao *et al.* 2012; Zhao *et al.* 2014)。高山松对青藏高原环境的适应过程和物种形成过程同时发生, 因此该过程也符合生态物种形成 (Nosil 2012)。生理生态证据表明高山松在抗旱和抗寒的某些生理

性状上方面优于其亲本种 (Ma *et al.* 2010; 2013), 但是目前我们尚不清楚高山松如何建立优于其亲本的气候适应性, 从而建立三个物种间生殖不相容 (reproductive incompatibility)。

1.5 松柏类植物 (Coniferous Species)

松柏类植物起源于石炭纪 (360-286 Mya, Miller 1977), 它是裸子植物中最丰富的类群, 包括 8 科, 68 个属, 大约 630 个物种, 广泛分布于北半球以及较南方的高山寒冷地区, 代表了大于 80% 的全球生物量 (Neale & Kremer 2011)。松柏类植物有着非常重要的经济和生态价值。它们很好的适应了寒冷和干旱环境, 也见证了 300 万年至今的气候变化, 因此是研究植物对气候变化响应的理想材料。研究它们的适应性进化和种群历史有利于物种的保护, 育种以及对物种气候变化适应的理解。

其中松属 (*Pinus*) 包括大约 110 个种, 他们以不同的模式广泛分布于北半球。松属松亚组中, 20 个物种分布于亚欧大陆。其中欧洲赤松 (*Pinus sylvestris*) 分布最广, 连续分布区横跨亚欧大陆, 而其他物种主要以异域或者领域的方式分布。有 12 个物种在中国分布 (Richardson 2000)。每个物种都有特定的分布区。例如高山松分布于高海拔的青藏高原, 油松分布于较干旱的中国北方。在中国的东南 (马尾松和黄山松) 和东南沿海地区 (赤松和黑松) 分别有两对松树重叠分布。松属植物有三套基因组: 分别是父系遗传的叶绿体, 母系遗传的线粒体和双亲遗传的核基因 (Wagner *et al.* 1987)。这为我们提供了很好的进化生物学研究系统。

松树的基因组很大 (大约 20-40Gb, $2n = 24$), 是人类基因组的大约 7-13 倍, 而且结构很复杂, 存有大量的重复序列。即使在基因组测序高速发展的今天, 测序和注释松树基因组依旧很苦难。目前质量较好的基因组仅有火炬松 (*Pinus teada*), 欧盟多个项目组正在联合测序欧洲赤松的基因组 (<http://procogen.wordpress.com/>)。火炬松的基因大概为 22Gb, 其中包含约 82% 的重复序列, 应用不同的方法, 注释了大约 5 万个基因 (其中约 16000 有较高置信度) (Neale *et al.* 2014; Wegrzyn *et al.* 2014; Zimin *et al.* 2014)。松柏类植物有大约 1500 个特有基因家族, 其中很多跟调控有关 (Wegrzyn *et al.* 2014)。鉴于松柏类基因组的大量重复序列, 推测基因调控作用在松柏类中可能更复杂也更重要。松柏类基因组研究将促进松柏类育种的发展, 基因组选择 (Genomic selection) 等更有效的分子育种方法很快被应用于松柏类植物遗传育种 (Neale & Kremer 2011)。

1.5.1 油松 (Chinese pine, *Pinus tabulaeformis* Carr.)

油松属于松科(Pinaceae), 松属(*Pinus*), 松组松亚组(Section *Pinus*, subsection *Pinus*)。油松是中国特有种, 是一种典型的抗旱树种, 广泛分布于中国北方于东经 102-122 度和北纬 32-43 度之间(Wu 1995), 油松具有重要的生态和经济价值。尤其在中国北方, 油松是主要绿化树种之一。

生物地理学研究表明油松的分布受到了第四纪冰期气候变化的影响, 冰期到来时, 油松分布可能分布于五个冰期避难所(Glacial refugia), 到了间冰期, 它又扩散到适宜的生境(Chen *et al.* 2008)。跟大多数其他的松柏类一样, 本地适应性是油松的一个重要特征, 但是, 目前仍没有相关研究去深入研究其分子基础。因此, 油松的遗传结构知识应该应用在育种和绿化中。同时, 最近, 研究人员发现表观遗传(Epigenetics)在本地适应中起重要作用(Cortijo *et al.* 2014)。育种和绿化应优先考虑本地种质和本地育苗。

1.5.2 高山松 (High mountain pine, *Pinus densata* Mast.)

高山松是油松和云南松的近缘种, 二者同属于松亚组(Wang *et al.* 1999)。高山松也是中国特有种, 分布于青藏高原以及边缘高海拔区域(平均海拔 3000 米以上)。它的针叶两针或者三针一束, 而云南松的松针为三针一束, 同时其大多数表型特征介于油松和云南松之间, 加上它的分布区域位于油松和云南松之间, 为以上二者不能生存的干旱寒冷的青藏高原, 三者的染色体数目一致($2n = 24$)。基于以上考虑, 早在二十世纪初期, 科学家认为高山松可能来源于北方的油松和西南方的云南松同倍体杂交物种形成(Homoploid Hybrid Speciation, HHS)。

近年来, 科学家们应用细胞质遗传的分子标记(Song *et al.* 2002; 2003; Wang *et al.* 2011), 核基因数据 (Ma *et al.* 2006; Gao *et al.* 2012), 表型数据 (Mao *et al.* 2009; Ma *et al.* 2010) 以及生理生态数据 (Mao *et al.* 2011) 进一步为这一推论提供了新的证据。但是由于近缘松树物种共享很多的遗传多态性, 可能是因为不完全谱系筛选(Incomplete lineage sorting)或者基因渗入(Introgression)(Du *et al.* 2009; Zhou *et al.* 2010; Li *et al.* 2012; Zhou *et al.* 2014 unpublished results), 假如没有分清楚以上二者, 极可能得到错误的结论 (Zhou *et al.* 2014 unpublished results)。不论如何, 考虑到三个物种可能在历史气候变化中反复接触和隔离, 高山松的物种形成过程可能比我们了解的要复杂的多。更切合实际的模型以及基因组数据应该被用来进一步研究高山松的物种形成过程。

另一个非常有趣但很少触及的问题是高山松适应于青藏高原早寒环境的分子基础。其中之一的可能解释是杂种优势 (Heterosis or Hybrid vigor)。杂种优势指的是杂交后代在某些特征上的适应性优于其父母亲本的进化生物学现象, 也被广泛应用在育种中。杂种优势的产生是由于杂交引起的父母亲本基因组间的基因互作, 从而改变了一些适应性相关基因的作用模式, 例如基因-基因互作, 蛋白-基因互作, RNA-DNA 互作以及表观修饰(e.g. Chen *et al.* 2013)。通过比较了三个物种幼苗在干旱和寒冷环境下的适应性, 研究人员发现了高山松在多数生理参数上显著优于其父母亲本(Ma *et al.* 2010; 2013)。随着火炬松, 尤其是更近缘的欧洲赤松的参考基因组的解码, 基于基因组, 表观遗传, RNA 调控以及基因表达的分子以及群体遗传方法可以用来深入研究高山松的杂种优势。另一种可能的解释: 高山松的早寒适应可能是有新的基因变异(New mutation)引起, 而不是适应性渗入(Adaptive introgression)或者固有基因突变(Standing genetic variation)引起的。通过对比早寒相关的候选基因和与早寒无关的参考基因, 可能初步回答以上问题, 找到可能的适应性变异。应用群体遗传学和景观遗传学方法更进一步揭示高山松适应性以及中性遗传结构和分布(Adaptive and neutral genetic structure), 在考虑全球气候变化的同时, 为未来高山松育种和保护提供科学依据。

1.6 本研究的目的 (Study questions)

我们样品采集覆盖了高山松和其亲本种油松和云南松的分布区, 共采集了 68 个 (根据细胞质分子变异的群体遗传结构, 我们选取了其中 27 个用于后续核基因分析) 野生居群的针叶和球果, 大概每个居群 10-15 个个体。根据已发表的基因表达和基因功能分析, 我们筛选并测序了 25 个与干旱和寒冷相关的基因和其他 18 个可能跟干旱和寒冷无关的基因。应用群体遗传学和景观遗传学方法分析了三个物种的早寒气候适应性以及中性的遗传结构和分布模式, 我们拟回答以下几个问题:

1. 那些气候因子是这两种松树分布区域的主要决定因素? 换句话说, 是地理隔离还是其他生态因子隔离导致的群体间基因分化? 借用物种分布的标本记录采样位点以及我们的采样位点的 GPS 信息, 加上气候数据库中不同气候因子的分层数据, 我们用物种潜在分布的模拟的方法去识别那几个气候因子是物种分布最重要决定因素。这些气候因子数据将被用于后续的关联分析。
2. 那些基因或者 SNPs 受到了历史的或者近期的来自于气候变化的选择压力, 促进了各个物种的早寒适应? 应用群体基因分化异常值(F_{ST} outliers,

Lewontin & Krakauer, 1973; Foll & Gaggiotti 2008; Excoffier *et al.* 2009), 以及综合了分层分析和连锁信息的群体基因分化分析 (HapFLK, Fariello *et al.* 2013) 去识别那些基因区域受到本地适应的影响。同时, 通过对主要气候因子跟等位基因频率的关联分析, 用基因分类(候选基因和参考基因, 基于贝叶斯模型的 Bayenv, Gunther and Coop 2013)和不分类[考虑到没有详尽的逐个基因的功能分析, 基于其他松树基因表达的分类也许不准确, 潜在因素混合模型(latent factor mixed model, LFMM), Frichot *et al.* 2013] 两种方法去识别频度梯度变化的基因和 SNPs。通过比较关联分析和基因分化异常值的结果, 得到潜在的气候适应的基因变异。

3. 气候变化首先影响边缘居群。在各个物种的边缘居群中, 那些基因或者 SNPs 受到了本地选择? 最南和最北边的居群中受本地气候选择影响的等位基因相同还是各异? 换句话说, 是平行适应(Parallel adaptation)还是条件中性(conditionality) ?

4. 高山松中受本地气候选择的 SNPs 是高山松固有的还是来源于其亲本? 是新的变异还是固有变异促进了高山松种群的本地气候适应?

5. 三个物种中找到的气候适应的基因是一样的还是不一样的? 换句话说, 能否找到是平行适应(Parallel adaptation)或者趋异进化(Divergent evolution)的证据? 由于油松和云南松是高山松的杂交亲本, 这种‘平行选择’的信号可能是由于适应性渗入(Adaptive introgression)导致的。那些 SNPs 通过适应性渗入促进了高山松适应于青藏高原的寒旱环境?

6. 通过比较叶绿体和线粒体单倍型和适应性和中性等位基因的分布模式, 从而检测种子介导的叶绿体基因流和花粉介导的线粒体基因流在种内本地适应和种间适应性渗入中的作用。

7. 在全球气候变化的背景下, 综上所述结论为油松和高山松提供保护和育种的科学依据。

第二章 材料方法

2.1 研究材料

本论文主要研究高山松以及它的两个亲本种油松和云南松，它们分别分布于青藏高原西南地区，中国北方和云南省有着重要经济和生态意义的中国特有树种。

2.1.1 油松

油松 *Pinus tabulaeformis* Carr，它是松科松组松亚组的常绿乔木(*Pinus* subgenus *Pinus* subsection *Pinus*, Gernandt *et al.* 2005)，树冠呈塔形或广卵形，树皮灰棕色，不规则厚鳞片状开裂。叶2针一束，深绿色，粗硬，长10-15厘米，直径约1.5毫米，边缘有锯齿，两面具气孔，树脂道5-8个或更多，边生于背面。雄球花橙黄色，雌球花绿紫色，种子淡褐色，连刺长1.5-1.8厘米。花期为4—5月，果实约次年10月成熟。球果卵形，长4-9厘米，鳞脐突起，有刺尖。种子生殖。心材淡黄红褐色，边材淡黄白色，纹理直，结构较细密，材质较硬，比重0.4-0.54，富树脂，耐用(Wu 1999)。油松作为我国特有树种，主要分布在中国北方的干旱寒冷地区。油松分布在海拔500-2000米，在青海和四川海拔可达2700米。油松是常被用作绿化树种，也是我国北方广大地区最主要的造林树种之一，有着重要的生态和经济意义。

2.1.2 高山松

高山松 *Pinus densata* Mast 也是松科松组松亚组的常绿乔木(*Pinus* subgenus *Pinus* subsection *Pinus*, Gernandt *et al.* 2005)。树干下部树皮暗灰褐色，深裂成厚块片，上部树皮红色，裂成薄片脱落；针叶2针一束，3针一束或2针3针并存，粗硬，长6-15厘米，径1.2-1.5毫米，微扭曲，两面具气孔线，边缘锯齿锐利；树脂道3-7个，边生。球果卵圆形，种子淡灰褐色，椭圆呈状卵圆形，微扁，种翅呈淡紫色，种子和种翅长约2.5厘米。花期约在5月，球果于第二年10月成熟(Wu 1999)。高山松的分布在青藏高原西南边缘海拔2500~4000m(单株最高分布可达海拔4350m)阳坡，但在过分干燥地区的阴坡和半阴坡也有分布。分布区南端与云南松相接，北端与油松分布地区相交错。高山松在低海拔不存在，在较高海拔比云南松和油松生长得更好，占据异于其亲本种的生态位。高山松对于保卫青藏高原脆弱的生态系统有着重要意义。

2.1.3 云南松

云南松 *Pinus yunnanensis* Faranch: 也是它是松科松组松亚组的常绿乔木 (*Pinus* subgenus *Pinus* subsection *Pinus*, Gernandt *et al.* 2005)。云南松叶为 3 针一束, 细长, 柔软, 树脂道 4-6 个中生或边生, 主要分布于我国云南省海拔 1500-3000 米的广大地区, 云南松对云南省的生态以及经济有着重要意义。

2.2 采样以及气候环境数据收集

我们采集了油松和高山松分布范围内多个群体的样品, 同时应用 GIS 方法和公共数据库 (如 WorldClim) 收集各个样点的气候数据。每个群体采集了 10-20 个个体, 每个个体采集松针 20-30 针和当年的球果 4-10 枚。采样工作主要在 2008-2010 之间 9-11 月种子成熟时完成的。

2.2.1 采样

采样过程遵循了保护性采样的原则, 也征得了相关单位的文件许可。采样点见于表 2-2 和图 2-1。每两个采样个体之间的距离至少大于 100 米以防止采取同一家系的个体。

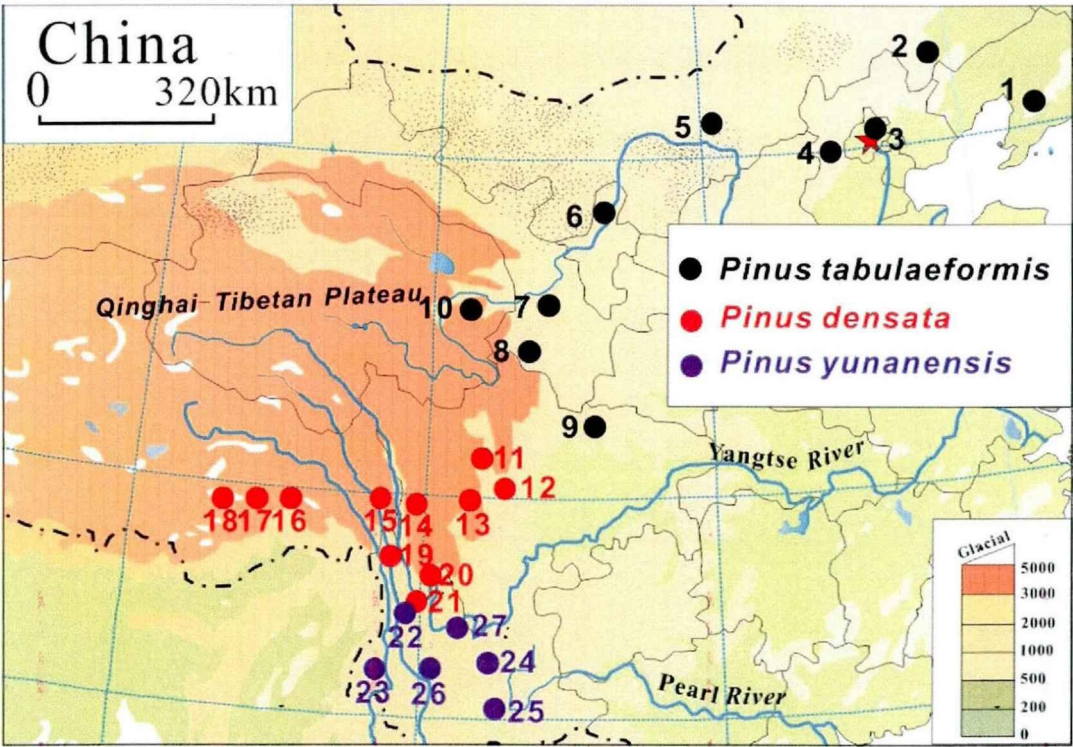


图2-1高山松复合体居群采样点的地理信息。

Fig. 2-1 Geographic information and sampling sites for *P. densata* complex.

表 2-2 采样点的位置和各采样点所取的种子数

Table 2-1 Location of the twenty-two sampled populations of *Pinus tabulaeformis* and the number of seeds analyzed per location in this study.

Code	N	采样点	Location	Latitude (N)	Longitude (E)	Altitude (m)
油松 (<i>Pinus tabulaeformis</i>)						
1	10	辽宁鞍山市千山自然保护区	Qian Mt LN	41°01.439'	123°07.255'	231
2	10	内蒙古翁牛特旗	Wenniuteqi NMG	43°24.930'	119°01.090'	362
3	10	北京市密云县新城子镇崔家峪	Miyun BJ	40.39.286'	117.22.047'	550
4	10	河北省涿鹿县小五台山山涧口	Xiaowutai Mt HB	40.00.290'	115.03.525'	1370
5	10	内蒙古包头市石拐区五当召	Baotou NMG	40.47.302'	110.18.533'	1510
6	10	宁夏贺兰山苏峪口	Helan Mt NX	38°30.761'	105°88.180'	2018
7	10	甘肃省天水市小陇山林场	Tianshui GS	34.20.315'	106.00.744'	1350
8	10	甘肃甘南迭部县电尕镇	Gannan GS	33°59.860'	103°26.020'	2147
9	10	青海坎不拉国家森林公园	Kanbula QH	36°06.299'	101°43.355'	2581
10	10	四川省广元市曾家镇	Guangyuan SC	32°36.678'	106°06.260'	1400
高山松 (<i>Pinus densata</i>)						
11	8	四川省小金县	Xiaojin SC	30°57.155'	102°18.037'	2676
12	8	四川省宝兴县	Baoxin SC	30°24.651'	102°46.351'	2209
13	8	四川省泸定县	Luding SC	29°54.608'	102°13.888'	2357
14	8	四川省乡城县马熊沟	Xiangcheng SC	29°09.061'	99°56.333'	3422
15	6	西藏省芒康县	Mangkang XZ	29°34.297'	98°18.029'	3691
16	8	西藏林芝	Linzhi XZ	29°38.850'	94°24.712'	4491
17	9	西藏省米林县	Milin XZ	29°11.544'	94°8.593'	3531
18	8	西藏省朗县	Langxian XZ	29°1.701'	93°4.405'	3755
19	8	云南省德钦县觉色乡	Deqin YN	28°27.410'	98°51.338'	3234
20	8	云南省香格里拉	Zhongdian YN	27°49.179'	99°26.349'	3284
21	8	云南省丽江	Lijiang YN	26°55.027'	100°16.231'	2603
云南松 (<i>Pinus yunnanensis</i>)						
22	6	云南省香格里拉县白水台	Baishuitai YN	27°29.803'	100°2.176'	2536
23	6	云南省腾冲县腾越镇	Tengchong YN	25°1.656'	98°29.732'	1891
24	6	云南省禄丰县将木菁	Lufeng YN	25°9.647'	102°4.356'	1628
25	6	云南省石屏县牛达镇	Shiping YN	23°42.138'	102°29.591'	1661
26	6	云南省永仁县白马河林场	Yongren YN	26°3.054'	101°40.254'	1534
27	6	云南省楚雄市	Chuxiong YN	25°1.539'	101°31.492'	1966

N, number of individuals sampled; NMG, Neimenggu; LN, Liaoning; BJ, Beijing; HB, Hebei; NX, Ningxia; GS, Gansu; QH, Qinghai; SC, Sichuan; XZ, Xizang; YN, Yunnan.

2.2.2 气候环境数据

我们一共从 WoorldClim(<http://www.worldclim.org/>) 收集了 19 个生物环境因子的数据（表 2-1）。这些气候因子主要关于气温和降雨。同时我们采样过程中记录了各个群体的 GPS 信息。GPS 信息和生物气候因子将被用于模拟油松和高山松的潜在分布区域，同时识别各物种分布区域的主要决定因子，这些关键因子将被用于后续的关联分析去识别呈频度梯度变化的等位基因。

表 2-1 用于物种分布模拟和关联分析的 19 个生物气候因子数据

Table 2-1 Climate data used in species' distribution modeling and association analyses

Name of bio-climate data	
bio1	Annual Mean Temperature
bio2	Mean Diurnal Range: Mean of monthly (max temp - min temp)
bio3	Isothermality: (P2/P7)* 100
bio4	Temperature Seasonality (standard deviation *100)
bio5	Max Temperature of Warmest Month
bio6	Min Temperature of Coldest Month
bio7	Temperature Annual Range (P5-P6)
bio8	Mean Temperature of Wettest Quarter
bio9	Mean Temperature of Driest Quarter
bio10	Mean Temperature of Warmest Quarter
bio11	Mean Temperature of Coldest Quarter
bio12	Annual Precipitation
bio13	Precipitation of Wettest Month
bio14	Precipitation of Driest Month
bio15	Precipitation Seasonality (Coefficient of Variation)
bio16	Precipitation of Wettest Quarter
bio17	Precipitation of Driest Quarter
bio18	Precipitation of Warmest Quarter
bio19	Precipitation of Coldest Quarter

2.3 测序

2.3.1 松树胚乳单倍体 DNA 提取

提取 DNA 的方法有多种：CTAB（十六烷基三甲基溴化铵）法；高盐低 PH 法；蛋白酶 K 法；SDS(十二烷基硫酸钠)法；苯酚法等。SDS 法和 CTAB 法多用于针叶树种基因组总 DNA 的提取。本试验采用改良的 CTAB 法，从萌发的种子

的胚乳中提取单倍体 DNA。所得产物能较好的用于后续分子实验。具体方法如下:

(1) 预处理: 将各个采样点采得的种子, 编号, 于铺有滤纸的培养皿中加入适量双蒸水, 每个个体挑取 2-3 粒饱满的种子分别泡入对应编号的培养皿中(干燥保存的种子种皮坚硬, 胚乳较脆, 容易碎裂, 所以提取胚乳总 DNA 前需进行萌发处理)。25℃培养箱中培养 36 小时以上, (以 48-60 小时为宜)。萌发时间太短则胚乳不易被分离, 尤其是跟内胚层以及胚的分离, 这样得到的用于后续分析的 DNA 中含有二倍体组织, 使得后续研究复杂化。而萌发时间过长, 种子种积累了大量的次生代谢产物, 影响所提取的 DNA 的质量。

(2) 分离胚乳: 培养结束后, 在超净台上剥去内外种皮, 分离胚和胚乳, 把胚和胚乳分别放入对应编号的 1.5ml 的离心管中, 注明编号和日期。将胚乳存入 -70℃ 的冰箱中保存, 便于后续分析。胚放入 4℃ 冰箱及时提取。

(3) 准备工作: 实验中所涉及到的枪头、杵 (于 1.5ml 离心管内腔底部锥形部分相配)、药匙和 BI 都要灭菌。提前 30 分钟打开水浴锅, 调温至 65℃, 将 CTAB 以 1mL 的量分装, 并各加入 10μl β-巯基乙醇编号后放入其中预热。将离心机调至 4℃, 让其自动制冷。取一瓶异丙醇放入 -20℃ 的冰箱冷存待用。

(4) 总 DNA 提取: ①取一泡沫盒, 向其中倒入适量液氮, 将装有胚乳的管放入液氮中 (用镊子操作, 以淹没到管的中部为宜) 冻至不发出吱吱响声, 立即取出, 用已灭菌的杵快速用力研细胚乳, 加入半勺 PVP 粉, 继续研磨至粉末状, 接着加入对应编号 65℃ 预热的 CTAB 研至溶液呈均匀糊状。②研匀后倒回至 2.0ml 离心管 (聚苯乙烯管, 耐氯仿) 中, 65℃, 200rpm 水浴 1h 每隔 3-4min 上下摇动一次。③水浴结束后, 4℃, 12000rpm 离心 10min。④吸取上清约 800μl, 加入 800μl CI (氯仿: 异戊醇=24:1), 于一自制混合仪上适速摇动 10min, 4℃, 12000rpm 离心 10min。⑤重复步骤④, 用 CI 再抽提一次。⑥取上清, 加入预冷的异丙醇 800μl 后于 -20℃ 沉淀过夜。沉淀完之后, 4℃, 12000rpm 离心 6min 沉淀 DNA。⑦小心弃去上清, 加入 500μl 75% 的乙醇, 轻轻弹起沉淀后于 4℃, 12000rpm 离心 6min。⑧重复步骤⑦, 用 75% 的乙醇再洗一次。⑨小心弃去上清, 加入 600μl 100% 的乙醇轻轻弹起沉淀后于 4℃, 12000rpm 离心 6min。⑩弃上清 (小心不要把 DNA 倒掉), 然后将离心管倒扣于滤纸上约 15min, 最后于 37℃ 培养箱中烘至无乙醇为止 (一般以 30 min 为宜)。加入 80-120μl 的 TE (10 mM Tris-HCl (pH8.0), 1 mM EDTA (pH 8.0) 溶解沉淀, 置于 4℃ 冰箱 2h 以上, 然后用

1%琼脂糖凝胶检测 DNA。由于所有样品的 DNA 质量直接影响后续实验的成败, 因为我们试图扩增每个个体的大约 50 个基因片段。我们用 CTAB 法提取的 DNA 只用于引物的筛选, 而群体重测序所用的 DNA 则采用植物 DNA 试剂盒 (QIAGEN DNeasy Plant Mini Kits (QIAGEN, Inc., Valencia, CA, USA)) 提取。种子的物理处理则与上述的改良的 CTAB 法相同[步骤 (1) 到 (3)]。

2.3.2 基因选取和引物设计

由于松树的基因组很大, 在火炬松和挪威云杉基因组发表之前, 可用的基因序列数据有限。由于高山松和油松都显示了抗旱和抗寒的特性, 我们从数据库中主要收集了近缘物种这两方面适应性相关的基因序列。这些基因最初是通过基因表达, 关联分析等方法在近缘物种中识别出来的。同时选取了其他的跟抗寒和抗旱无关的基因序列用作参考基因。总共设计和测试的引物序列有 207 对, 其中在三个物种中成功扩增和测序共 43 对引物。引物序列, 基因功能, 基因结构和参考文献均被列于表 2-2 中。

2.3.3 PCR 扩增, 产物纯化以及测序

扩增反应体系为 25 μ l, 反应体系包括: 10-40 ng DNA, 10 $\times$ Buffer, 2.5mM, MgCl₂, 1.5mM, DNTP, 0.5mM, 正反引物各 2 μ M, DNA 合成酶 0.75U。引物合成以及试剂生产于 TakaRa 公司)。PCR 反应程序: 1.变性, 94 $^{\circ}$ C, 6min; 2.变性 94 $^{\circ}$ C, 30s; 3.退火, 56 $^{\circ}$ C (基因片段特异的退火温度), 30s; 4.延伸, 72 $^{\circ}$ C, 1min; (2-4, 28 cycles); 5.延伸, 72 $^{\circ}$ C, 10min; 6.4 $^{\circ}$ C, 10h。PCR 是用试剂盒实现的 (TIANquick Midi Purification Kit 产于 TIANGEN BIOTECH <BEIJING>)。具体操作遵照生产厂家的说明书。纯化产物采用 2%的琼脂糖凝胶电泳检测, 呈清晰单条带的产物用于后续的序列分析。最初的引物测试的测序反应是在 ABI-3130xl Genetic Analyzer 上进行, 群体重测序是在 ABI-3730xl Genetic Analyzer 上实现的, 使用 ABI 公司提供的试剂。测序反应的 10 μ l 体系包含: 6.21 μ l ddH₂O, 0.64 ITS1a/ITS4(5pM), 0.3 μ l Bigdye, 1.85 μ l 5 $\times$ Sequencing Buffer 和 1.0 μ l 模板 DNA。每个扩增片段都用正反引物进行测序。

表 2-2 候选基因和参考基因的基因功能, 基因结构, 引物序列, 退火温度, 数据库同源序列的序列号以及参考文献。

Table 2-2 Putative functions, gene structures, primer sequences, annealing temperatures, Genbank homologs and source for each amplicon.

Locus	Gene	Length (bp)	Exon/ Intron UTRs	Primers (5'-3')	T _m (°C)	Accession/ GenBank homologs	Reference
<i>Climate-related candidate genes</i>							
		16888	9613 7275				
<i>a3ip2</i>	ABI3-interacting protein 2	1024	269 755	F: AATGCCAGGTTGGTGTTA R: CAGCCTCAATTGCTTTCC	58	EU120040	Pyhäjärvi <i>et al.</i> 2007
<i>agp4</i>	Putative Arabinogalactan/ proline-rich protein	781	532 249	F: TGCCAGCAAAAGCAITTIACCG R: CCCCTGAAATCCGTTACTAC	60	AM501931	Eveno <i>et al.</i> 2008
<i>aqua-MIP</i>	Aquaporin, membrane intrinsic protein	634	263 371	F: TGTCACCTGCCCCAGAGCTATTC R: ATCACAGCCGCTCCAAAAAC	58	AY874569	Gonzalez-Martinez <i>et al.</i> 2006
<i>agp-like</i>	Arabinogalactan-like protein	452	452 0	F: GAAACCCGGGGCTCTGAACT R: CCGCAGCGAGGAAAAACAA	60	AY752407	Brown <i>et al.</i> 2004
<i>araH</i>	Abscisic acid and water-stress-induced protein	401	327 74	F: CGTTAATTCTGGCTCTGC R: GTTTCCTCGTCTCGTTTC	60	EU120040	Wachowiak <i>et al.</i> 2009
<i>araR</i>	Abscisic acid-responsive protein	503	334 169	F: AAGATTGCAATGATGAGTAAAGAA R: CTTGGCCCAITTTGTTGAAGCAGTG	61	GQ262095	Wachowiak <i>et al.</i> 2009
<i>ccoaomt</i>	Caffeoyl-Co-O-methyltransferase	536	287 249	F: GCAGCAGAAAGTGAAGGCTCAGA R: TCTTTCCATCATCGGGCAATG	59	AM502291	Gonzalez-Martinez <i>et al.</i> 2006
<i>Comt</i>	Putative Caffeic-O-methyltransferase	446	446 0	F: CGCTGAACAAGAAGAGGCG R: GTGGCAATGACATGAGGC	60	AJ297297	Eveno <i>et al.</i> 2008
<i>dhn1</i>	Dehydrins	947	528 419	F: TCACCCGCACCCACAGTTC R: TCCGTCGGCTCACATTCA	60	FJ201260	Wachowiak <i>et al.</i> 2009

<i>dhn2</i>	Dehydrins	556	430	F: CTGCAGAGACTGTGCCCTGAGC R: CCAGGGAGCTTTTCCTTATCT	60	EU394116	Eveno <i>et al.</i> 2008
<i>dhn7</i>	Dehydrins	372	270	F: ATTAAGATGGCGGAAGAGCAACAG R: TTGTACCCGAAAGTCCCATTT	62	HQ109167	Wachowiak <i>et al.</i> 2009
<i>dhn9</i>	Dehydrins	808	621	F: GTTCTGTCCATTGTGCTTCTTAT R: CAGCTTCTCTTTGATTTTACCAAG	60	HQ108543	Wachowiak <i>et al.</i> 2009
<i>erd3</i>	Early responsive to dehydration 3	804	565	F: CCCATGCGGTCTTGTATAGT R: ATCCTGGGAAATGAAACG	60	AY874612	Eveno <i>et al.</i> 2008
<i>Gl</i>	<i>GIGANTEA</i>	504	447	F: GCCAGAGCCTCATCAGTTC R: GATGTACTTTCAACACTCATTTGC	60	EU120263	Pyhäjärvi <i>et al.</i> 2007
<i>Glu</i>	Putative Glucan-endo-1,3- bet-glucosidase precursor	960	508	F: CATTCGCTTGGTATTCTG R: CACCCTTCCTTGATGATTT	58	AM502290	Eveno <i>et al.</i> 2008
<i>Hlhl</i>	Helix-loop-helix protein 1A (HLH1)	485	255	F: TCAGTCAGGAGGAAGGAT R: GGATAATAGTCCAGGAGAAC	60	EU120313	Pyhäjärvi <i>et al.</i> 2007
<i>Ino3</i>	Putative Inositol-3-phosphate synthase	676	325	F: AGCGTTCGTGCTTTCTTC R: TTCGAGGGTCAITGTTGT	60	AM774401	Eveno <i>et al.</i> 2008
<i>LEA</i>	Late embryogenesis abundant protein	801	79	F: ATTGTGAGCAGACATGGCCAGTAT R: TATCTGACAGAGTGGGCAGCTTCAT	60	FJ201603	Wachowiak <i>et al.</i> 2009
<i>lp3-1</i>	AB- and WDS induced- gene-1	459	176	F: AAGCACGCAGATAAGAAGGA R: AGAACGGGCCAAATTTTACT	60	GQ262430	Gonzalez-Martinez <i>et al.</i> 2006
<i>MdhA</i>	Malate dehydrogenase	519	190	F: CATATCCGTGATTGGGTGCTT R: CAGTGGCATCCAGTTTTCCT	60	HM853228	Pyhäjärvi <i>et al.</i> 2007

<i>PHYO</i>	Phytochrome A	944	773 171	F: ATTCTAATGGACTGGTAAATGGATG R: ATAAATGCACTCGTTTGCAAAATT	58	EU120580	Garci-Gil <i>et al.</i> 2003
<i>Pod</i>	Peroxidase	441	170 271	F: CCGATGCTGACTCTTCTTAAC R: AGCGAATTGGAGGATGA	60	DQ233171	Ma <i>et al.</i> 2006
<i>ppc</i>	Protein phosphatase 2C	594	408 186	F: CGTTACAGCACCTGCCAGGATTTC R: AACCACTCTCGAATCCCCAC	60	AY874735	Eveno <i>et al.</i> 2008
<i>ppap12</i>	Uncertain, possible wall-associated protein kinase	392	392 0	F: ATTAGCAGGGCATCTGTCTG R: CACGCCCTCTCAITTTTCATC	60	AY874831	Gonzalez-Martinez <i>et al.</i> 2006
<i>RD21A</i>	Cystein protease (pseudotzain)	872	566 306	F: AGGCCTGATCCAGCCAAAG R: ACCGTTGCACTCTTGTTGT	60	EU394109	Eveno <i>et al.</i> 2008
<i>Putative reference loci</i>		7262	3623 3639				
<i>PtIFG2009</i>	Putative 60s ribosomal protein L10A	459	227 232	F: CATTCCTCGCCCTAAGAT R: TTGCCAGCCTTGTGTGAGA	62	DQ233190	Ma <i>et al.</i> 2006
<i>c3h</i>	Coumarate 3- hydroxylase	799	414 385	F: ACGAGTGGTGCCAAACAA R: TGGGTCCTTCCATACAGC	60	AY764608	Brown <i>et al.</i> 2004
<i>c4h2</i>	4- Coumarate: CoA ligase	554	272 282	F: CATGTGCATCCGTAGGCG R: CTCCTTGTCTGAGCATCT	60	AY764926	Brown <i>et al.</i> 2004
<i>CesA2</i>	Cellulose synthase	511	335 176	F: AGATCTTGCTCAATGCCTCG R: CCAAACCTTCACTGTCAATCG	58	AY262821	Krauskopf <i>et al.</i> 2005
<i>GapCp</i>	NAD-dependent glyceraldehyde-3-phosphate dehydrogenase	545	171	F: ACCTTGCCACGATAAACATT	60	AF359129	Dvornyk <i>et al.</i> 2002

			374	R: AGGATAAAGGCACAACAATAC		
<i>LHC44</i>	LHC I type IV chlorophyll-binding protein	539	409 130	F: GGAGCTAGTGAACGGGAGGTG R: GAACGGGCCCTTTCCTGTTA	60	AF359127 Dvornyk et al. 2002
<i>nir</i>	Nitrite reductase	304 0	304 0	F: AAAGCAAGAGCCCTGAAAAT R: CCTGATGGCTCCAAAAGTG	60	EU120466 Palme et al. 2009
<i>PAL</i>	phenylalanine ammoni-lyase	840 0	840 0	F: AATGGGAAGCCTGAGTTT R: AGTTTCTGAGTCAATGGGTA	60	AF353967 Dvornyk et al. 2002
<i>pho</i>	Phosphofructokinase	491	183 308	F: GCAAATGCAAAAAACCATAAC R: AAAATAAGCAGGCAAAAATGAC	60	AF359113 Dvornyk et al. 2002
<i>ptlim-1</i>	LIM transcription factor	439	128 311	F: TGGTTGTGTTATCAGTCCCTCAAAC R: CGGTAACTGCCACTAATAGCCACTT	60	AY670613 Brown et al. 2004
<i>ptlim-2</i>	LIM transcription factor	402	197 205	F: ATCATACCAACCGACCTTGCTTCA R: CTGCCCCAACAGCATAACACTATCTT	60	AY670456 Brown et al. 2004
<i>WD40</i>	WD-40 repeat family protein	1078	143 935	F: ATGGGGGTGCAGCATAAAC R: GGGATGGCAACAACAAAAA	60	FI395960 Willyard et al. 2007

2.4 数据处理

应用群体遗传和景观遗传的方法,我们旨在研究三个物种的遗传多样性模式,例如中性和适应性遗传结构和分布。首先,我们比较油松,高山松以及其他松柏类的遗传多样性;其次,用群体遗传学方法去检测各个时间尺度受到自然选择的基因,去识别平行进化或趋异进化的模式;最后,应用景观遗传学的方法,我们试图发现频度梯度变化的基因位点,进一步调查中性和适应性遗传变异的分布模式。

2.4.1 序列比对和注释

测序结果用 Chromas 软件编辑,然后用软件 MUSCLE (Edgar 2004), MEGA 5.1 和 Clustalx (1.81)进行手工序列比对。只有清晰的单峰被用于后续的分析。基因注释是通过 BLAST 公共数据库完成的。数据库中的同源基因序列号被列于表 2-2 中。

2.4.2 遗传多样性分析

DNA 多态性程度可用多种方法来度量。样品中不同(多态)序列的数目(k)是一种简单的核酸水平的多态性水平的衡量其局限在于 k 值与样本大小直接相关;基因(gene)或单倍型(haplotype)多样性(h 或 H_d)更适合用于多态性的度量(Nei & Tajima 1981),它与等位基因个数和某一座位上第 i 个等位基因的群体频率有关;每个核苷酸座位的分离位点(segregating site)数目(P_s)属于核苷酸水平的多态性指标, $P_s = S/N$,其中, S 代表一组数据中的所有分离位点数目, N 表一组数据中所有核苷酸数目。在不存在选择且达到突变-漂移平衡的群体中(必须符合无限位点模型), P_s 的期望值 $E(P_s) = \alpha\theta$ (Watterson 1975); θ 是一个比 P_s 更基本的遗传变异参数, $\theta = 4\mu N$,其中 N 和 μ 分别代表有效群体大小和每个位点的突变速率,其优点在于它是突变速率和群体大小的积,并且独立于群体大小;另一种更合适的多态性度量是两个序列间每个位点上核苷酸差异的平均值或是核苷酸多样性 π (Nei & Li 1979), $\pi = \sum_{i,j} p_i p_j x_{ij}$, 其中 p_i 和 p_j 是群体中序列 i 和 j 的频率, x_{ij} 是 i 序列和 j 序列之间的差异占序列总长的比例(N)。最常用的是 θ 和 π 。本实验使用 DnaSP 5.10 分别估计了序列间每对碱基的平均差异(π) (Nei & Li 1979)、Watterson's estimates (θ) (Watterson 1975)和单倍型多态性(H_d) (Nei 1987)。同时我们计算了各个物种各基因座位上的同义替代和非同义替代的比率。

2.4.3 连锁与重组

连锁不平衡(Linkage Disequilibrium LD)又被称作等位基因之间的相关(allelic association),指群体内不同位点上,等位基因间的非随机组合式样(Goldstein 2001)。研究人员提出了多种连锁不平衡的参数,其中大部分是衡量两个单倍型之间的相关性。常用的有: D 、 $|D|$ 、 D' 、 $|D'|$ 、 R 、 R^2 。假设基因A、B的出现频率分别为 $P(A)$ 和 $P(B)$,它们组成的单倍型出现的频率为 $P(AB)$,则: $D = P(AB) - P(A)P(B)$; $D' = D/D_{\max}$,如果 D 为正值,则 $D_{\max}$ 是 p_1q_2 和 p_2q_1 中的较小者;如果 D 为负值, $D_{\max}$ 是 p_1q_1 和 p_2q_2 中较小者; $R^2 = D^2 / P(A) * P(B)$ (Hill and Robertson 1968)。 D' 和 R^2 是两种最重要的衡量方法。

如果 $D'=1$,则完全连锁不平衡(Complete Linkage Disequilibrium),即未发生重组,这种情况下两位点的四种单倍型最多出现三种。如果 $D' < 1$,则发生过重组(新发生的突变也会引起这种 D' 值的变化),四种单倍型都有可能出现,但这种 D' 值的相对大小就相对模糊($D' = 0.7$ 和 $D' = 0.3$ 区别模糊)。如果 D' 接近1,则提示两位点间历史上发生重组的可能性很小。但如果 D' 处于中间值则不可用该值比较两位点LD程度的差别,而且在小样本中 D' 值会显著上升。

R^2 代表两位点统计学上的关系,某种程度上可以作为 D' 的补充。 $R^2 = 1$ 则完美连锁不平衡(Perfect Linkage Disequilibrium),两位点没有重组分开,而且等位基因频率相同,这种情况下由两位点构成的四种单倍型在所选样本中只出现两种(ab和AB)。我们只用了信息位点,排除了插入缺失突变以及三突变位点。在突变-漂变模型中,LD随着位点之间距离的衰退可以用非线性回归方程来估算(Hill & Robertson 1968)。我们用非线性最小二乘法估算相邻位点之间的群体重组参数 ρ ($\rho = 4Nec$, N_e 是有效种群大小, c 是重组率)。

2.4.4 群体遗传结构

为了检测油松和高山松的遗传结构和分布,我们运用了贝叶斯群聚方法:STRUCTURE v.2.3.3 (Hubisz *et al.* 2009)。此方法把个体分配到预先定好的 K 个群集中。我们重复运算了6次,每次 K 值均为1-8,10万个循环的预算模拟加上100万个循环的数据收集。我们用 ΔK (Evanno *et al.* 2005)参数来识别最可能的群组数。群体结构的结果是根据最可能的 K 值来解释的。

2.4.5 三个进化时间尺度检测选择信号

我在三个进化时间尺度上检测了自然选择的信号：近期的种群内选择信号；油松和高山松分化过程中（油松-高山松，大约 3 百万年）；两个物种的谱系中的古老选择（红松-油松谱系，红松-高山松谱系，大约 45-85 百万年）。

2.4.5.1 油松和高山松的选择信号以及种内各群体的本地适应

检测种群内的近期选择信号主要建立在序列变异模式和中性期望之间的比较上。本研究采用变异位点频谱(Site Frequency Spectrum, SFS) 相关的中性检验，比如 Tajima's D , Fu and Li's D^* , F^* , Fay and Wu's H 和 Zeng's DHEW 检验。 F_{ST} 异常值法被广泛用于检测本地选择(Local selection), 为此我们用 F_{ST} 异常值的方法检测种内各个居群间的可能存在趋异选择(Divergent selection)信号。

Tajima's D 比较序列分离位点的数量 (the number of segregating sites, θw) 和两两差异 (pairwise differences, $\theta \pi$) (Tajima 1989)。 D 显著大于零可能是由于平衡选择或者种群的瓶颈效应，而 D 值显著小于零则由于强的正选择，群体扩张或者较弱的净化选择所导致的 (Hedrick 2011)。例如人群中 HLA-A 基因的 D 等于 3.69，而全基因组的 D 为接近显著的负值，这说明 HLA-A 基因可能受到平衡选择，而人类群体本身经历了近期的群体扩张。Fu and Li's D^* 比较衍生的单突变和总得衍生突变的数目 Fu and Li's F^* 比较衍生的单突变和序列间的平均成对差异 (Fu & Li 1993)。Fay and Wu's H 比较衍生的等位基因频率和成对差异频率 (Fay & Wu 2000)。上述参数是在 DnaSP v5 (Librado & Rozas 2009) 软件上计算的。DHEW 检验是一种综合了 Tajima's D , Fay and Wu's H 和 Ewens-Watterson (EW Watterson, 1978) 检验的复合检验 (Zeng *et al.* 2007)，它被证明很少受到种群历史的影响。DHEW 检验是在 DH 软件上完成的 (http://zeng-lab.group.shef.ac.uk/wordpress/?page_id=28)。

同时，我们采用了衍生突变的最大频率(maximum frequency of derived mutations, MFDW)去估算近期自然选择发生的可能(Li 2011)。MFDW 检验通过比较进化树的每个基部枝的长度检验自然选择。这种检验很少受到种群历史以及杂交渐渗的影响 (Li *et al.* 2011)。我们用了 MFDW 软件 (<http://www.picb.ac.cn/evolgen/softwares/>) 去检测各个基因座位近期自然选择的可能性。

F_{ST} 异常值法被广泛用于检测近期自然选择 (图 1-1)。显著高于中性期望的

F_{ST} 值是由于本地选择或者歧化选择引起的而显著低于中性期望的 F_{ST} 值是由于平衡选择引起的(e.g. Beaumont & Nichols 1996; Foll & Gaggiotti 2008)。我们应用了三种基于不同模型的 F_{ST} 去检测受到近期本地选择基因位点:

- 1) Bayescan (Foll & Gaggiotti 2008) 模拟一个基于岛屿模型的种群历史的群体特异影响的参数 β 和另一个对选择强度敏感的基因座位特异影响的参数 α , 其中 β 用来描述群体的中性历史, 而 α 在需要座位特异组分来解释观测到的多样性模式时候被保留。贝叶斯因子和后验概率被用来揭示选择和中性模型的相对可能性。我们在同样的条件下分别分析了候选 SNPs 和参考 SNPs。为了控制假阳性结果, 10 次初级运算, 每次 5 万次预算, 接着 10 万次输出结果, 稀疏间隔为每 10 次, 共进行了 1 万次后验估算。我们 5%显著水平的后验频率上修正了错误发现率去识别异常值。
- 2) 基于分层岛屿模型(Hierarchical island model)的异常值法, 由于种群常常显示分层的遗传结构, 这种情况下基于岛屿模型的 Bayescan 方法将带来很多的假阳性和假阴性。考虑到这一点, 基于分层岛屿模型的方法可以更有效的检测复杂遗传结构的种群中的自然选择信号 (Excoffier *et al.* 2009)。高山松和油松的群体呈现较复杂的遗传结构, 因此基于分层岛屿模型的 F_{ST} 异常值法应该能更好的检测到选择信号。
- 3) 基于 LD 和分层岛屿模型的异常值法, 多种检测近期自然选择的方法是建立在 LD 之上的, 因为强的正选择会导致作用位点及邻近多样性降低, 同时基因高度连锁。最新的方法综合了 LD 和 F_{ST} 异常值法, 同时应用分层岛屿模型, 很有效的用于检测近期自然选择信号 (hapFLK, Fariello *et al.* 2013)。

2.4.5.2 油松-高山松中的选择信号 (大约 3 百万年的进化时间尺度)

在较长的进化时间尺度上, 油松-高山松中的选择信号是用种间分化 F_{CT} 异常值检测两种松树之间的歧化选择。我们也用 Hudson-Kreitman-Aguadé(HKA) 检验和最大似然 HKA 检验 (maximum likelihood HKA, MLHKA) 以及 McDonald-Kreitman (MK) 检验以及 McDonald-Kreitman 泊松随机场检验 (McDonald-Kreitman Poisson Random Field analysis, MKPRF)检验油松-高山松中

的各种选择（包括正选择，平衡选择和净化选择，Positive selection, balancing selection and purifying selection）。HKA 和 MK 检验和他们的扩展：MLHKA 和 MKPRF 检验通过比较种内多样性和种间分化来检测各种选择信号（详见 2.4.4.3）。

2.4.5.3 红松-油松谱系，红松-高山松谱系的古老选择（大约 45-85 百万年的进化时间尺度）

在更长的进化时间尺度上红松-油松谱系和红松-高山松谱系上的选择主要用 HKA 和 MK 检验和他们的扩展：MLHKA 和 MKPRF 检验。

McDonald-Kreitman (MK) 检验以及 McDonald-Kreitman 泊松随机场检验 (McDonald-Kreitman Poisson Random Field (MKPRF) analysis)。MK 检验是建立在比较种内和种间的同义和非同义突变 (McDonald & Kreitman 1991)。中性指数 (NI),

$$NI = \frac{\frac{\text{种内非同义多态位点}}{\text{种间非同义固定位点}}}{\frac{\text{种内同义多态位点}}{\text{种间同义固定位点}}} \quad (\text{Rand \& Kann, 1996})$$

沉默变异位点被假定中性，过剩的非同义固定位点被当成是正选择信号，而过剩的非同义多态位点是净化或者平衡选择的信号。我们用红松为外类群在 DnaSP v5 估算了每个种的中性指数。同义和非同义多态和固定位点也被用来在 MKPRF 软件中估算群体选择参数 (Population selection parameter, $\gamma = 2NeS$, Bustamante *et al.* 2002), 其中 Ne 是有效种群大小, S 是选择系数。MKPRF 软件是建立在分层贝叶斯模型上描述特定位点上选择的强度和方向, 同时检测一组基因的总选择趋势。MKPRF 应用蒙特卡洛马尔科夫链从参数的后验分布中采样。我们共运行了 10 次包括 1 万次预预算接着每 10 次采样一次, 每次运算共计采样 1 万次。我们总结了每个座位的选择系数和 95% 置信区间的后验分布。

Hudson-Kreitman-Aguadé (HKA) 检验 和 最大似然 HKA 检验 (maximum likelihood HKA, MLHKA)。HKA 检验 (Hudson *et al.* 1987) 估算多基因座位的种内多态性和种间分化的比率。跟中性位点显著不同的位点被认为是受到选择作用。我们在油松-高山松以及油松-红松以及高山松-红松中在 HKA 软件 (<http://lifesci.rutgers.edu/~heylab/ProgramsandData/Programs/HKA>) 中做了 HKA 检验。

卡平方检验被用来决定统计显著性。由于在所有以上的三次检验中得到了显著性结果。我们进一步应用最大似然HKA检验去识别那个基因受到选择压力的影响。最大似然HKA是HKA的一个扩张,它能够通过比较中性模型和特定位点的选择模型用来在特定位点或者一组基因的选择检验 (Wright & Charlesworth 2004)。首先我们确定了每个参考基因的遗传中性。在参考基因中,我们在中性和选择模型下比较每个参考基因和剩下的参考基因,我们计算了每个位点选择参数K。选择参数K测量相对于种间分化的种内多样性水平, K值显著大于零说明平衡选择导致更高种内多样性, K值显著小于零说明正选择或者净化选择降低种内多态性。标准似然比检验被用来比较特定位点受选择模型和所有位点中性模型(不受选择)(自由度为1)。被评估为中性的参考基因被定义为“已知中性基因”。第二,我们检测了每个候选基因的选择。我们每次加一个候选基因于已知中性基因同时运行中性和选择模型。每个候选基因的参数K和显著性水平是用跟参考基因的中性评估同样的方法。我们在 MLHKA 软件中 (<http://www.yorku.ca/stephenw/StephenI.Wright/Programs.html>) 以不同的初始值 and 分化时间参数T重复运行了5次预测, 每次包括1百万次马尔科夫链。

2.4.6 景观遗传分析

景观遗传学方法越来越多的被用来研究中性和适应性进化对群体间基因流的影响和中性和适应性变异的景观分布(Manel & Holderegger 2013)。我们首先应用物种分布模拟来的调查两个物种的潜在分布区,同时识别两个物种分布的最重要生物和非生物决定因子。这些因子将被用来后续的环境因子和基因频率之间的关联分析。

2.4.6.1 物种分布模拟(Species distribution modeling)

为了识别油松和高山松的潜在分布区以及最重要的分布决定因子,我们应用了基于地理信息系统(GIS)和分层环境数据的最大熵模型(Phillips *et al.* 2006; Phillips & Dudik 2008)去模拟两种松树的潜在分布区。我们应用了 maxent 3.3.3 软件,它假设物种的发生是按照样点均概率分布的。我们的采样点以及标本馆的相关记录的样点 GPS 信息被作输入文件。我们一共从 WorldClim 上下载了 19 个空间分辨率 2.5 米的环境变量(表 3-2)。我们用了 maxent 3.3.3 软件系统设置参数,其中 75%的样点记录用于机器学习,另外 25%用于模型检验。结果中决定物种生态位的主要气候因子被用于后续的景观遗传分析。

2.4.6.2 环境因子和基因频率的关联分析

为了进一步研究物种各个群体沿着不同环境梯度的本地适应, 本文采用了贝叶斯广义线性混合模型(Bayesian generalized linear mixed model, BAYENV, Coop *et al.* 2010; Gunther & Coop 2012) 和潜在因素混合模型 (latent factor mixed model, LFMM, Frichot *et al.* 2013) 去检测关键气候因子与等位基因频率的关联关系, 前者将基因分为参考基因和候选基因, 参考基因用于估算种群历史, 候选基因频率跟气候因子的关联是考虑了种群历史的基础上被检测的; 而后者不用分类, 所有的基因被用于估算种群历史和关联分析。由于松柏类的基因功能分析有限, 为此基于近缘物种基因表达的基因分类可能并不太准确, 所以本文综合以上二者对基因分类进行了分析。

Bayenv 检验候选基因变异频率跟气候因子之间共变超出预期的从参考基因估算而来的群体共变异数矩阵 (Coop *et al.* 2010)。群体差异的共变异数矩阵是基于随机漂变模型下群体间的等位基因频率变化, 与群体‘中性’种群历史相似。我们用参考基因的 SNPs 矩阵去生成‘中性’的种群历史。我们在 BAYENV2 (Gunther & Coop 2012)上共运行了 5 次, 每次包括 50 万的马尔卡夫链。输出结果的平均数作为群体共变矩阵用于后续分析。接着, 我们在用基于参考基因的群体共变系数去控制群体历史的同时检测了候选基因群体特异的等位基因频率跟环境因子的相关性。对于特定候选 SNP, BAYENV2 软件比较一个 SNP 频率跟群体历史相关的空模型和另一个包含了 SNP 频率跟环境因子相关的备选模型。贝叶斯因子是通过备选模型和空模型的比率的后验分布决定的。因为线性模型有时候对异常值无效, Spearman 的基于排序选择模型的 ρ 被用来排除假相关。我们分别分析了每个候选 SNP 并估算了它们的贝叶斯因子和后验优势率。我们一共运行了 10 次, 每次 50 万马尔科夫链。拥有 5%最高贝叶斯因子和 5%最高 Spearman's ρ 值的 SNPs 被理解为支持该频度梯度适应。

潜在因素混合模型(latent factor mixed model, LFMM, Frichot *et al.* 2013) 也是在纠正群体历史的背景的前提下检测等位基因频率跟环境因子相关性。LFMM 首先基于 PCA 和所有群体序列数据估算群体历史, 在这个基础上进一步检测环境因子跟 SNPs 频率的共变。

第三章 研究结果

3.1 核苷酸多样性 (Intraspecific genetic diversity)

为了初步比较三个物种的基因多样性差异，我们应用群体遗传方法分别计算了整个序列，沉默位点以及非同义突变位点的基因多样性。

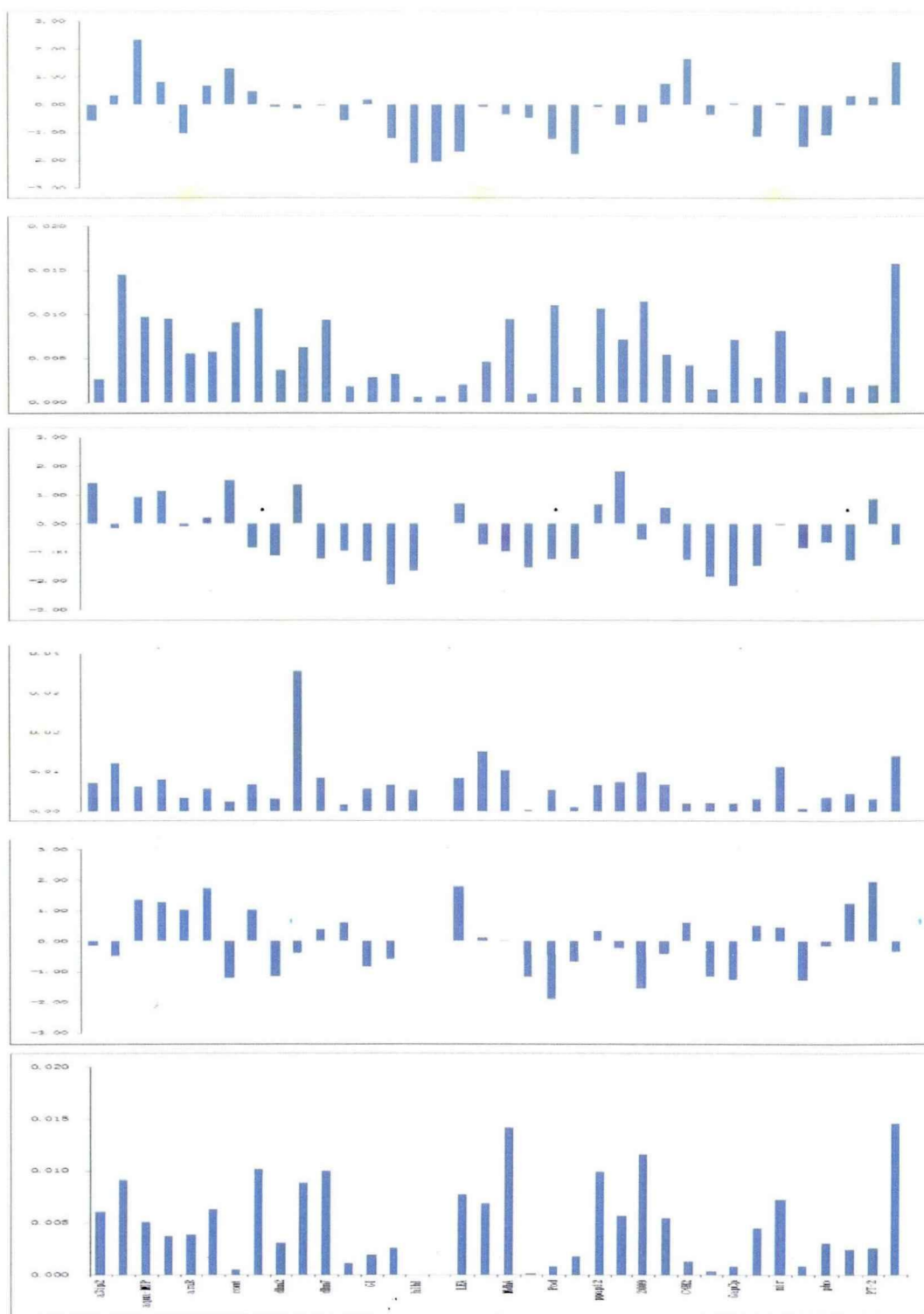
本研究涉及三个物种的 267 个雌配子体（取自 267 个不同的个体），37 个核基因座位，共 6000 多条序列，比对后的片段长度 24389 bp（附件中表 1）。我们在油松，高山松和云南松分别检测到 646,1271 和 429 个信息位点，其中单 WD40 含有的变异位点数最多（附件中表 1）。

在油松中，与他人所作的相似工作相比较（Ma *et al.* 2006），我们得到的结果核苷酸多态性较高（ $\theta_w = 0.00632$, $\pi = 0.00822$ ），同义位点的核苷酸多态性（ $\theta_s = 0.02343$, $\pi = 0.02492$ ）明显大于非同义位点（ $\theta_a = 0.00073$, $\pi = 0.00107$ ）没有发现正选择效应。以油松为例，中核苷酸多态性最高的居群是 ZWL（ $\theta_w = 0.01033$ $\pi = 0.01087$ ），最低的是 BQ（ $\theta_w = 0.00472$ $\pi = 0.00473$ ）。同义位点核苷酸多态性最高的是 ZWL（ $\pi = 0.03064$ ），最低的是 BQ 和 KY（ $\pi = 0.01734$ ）。非同义位点的核苷酸多态性最高的是 KY（ $\pi = 0.00650$ ），最低的是 HB 和 BQ（ $\pi = 0.00000$ ）（表 3-2）。单倍型的数目（H）及其多态性（ H_d ）在居群间也有很大差异，其中最高的是 BQ 和 KY（ $H_d = 1.0000$ ），最低的是 HB（ $H_d = 0.0667$ ），单倍型多态性平均值较高（ $H_d = 0.0856$ ），由于各居群个体数不均，所以对各居群单倍型数目的统计意义不大，但是我们可以粗略发现，HZ 的单倍型数目相对少于其他居群（附件中表 1）。

图 3-1 候选基因和参考基因的基因多样性以及位点频谱统计 Tajima's D

注：从上至下分别为油松的 Tajima's D，基因多样性 π ，高山松的 Tajima's D，基因多样性 π 以及云南松的 Tajima's D，基因多样性 π

Figure 3-1 Genetic diversity and site frequency spectrum for each amplicon in the three species



选择扫荡 (selective sweeps) 会降低相关基因座位的基因多样性, 同时 Tajima's D 呈现负值。如上图 3-1 所示, *erd3*, *Gl*, *Glu*, *Ino3* 和 *pp2c* 在油松和高山松中同时显示较低的基因多样性和负的 Tajima's D 。而在云南松中由于采样个体较少, D 值均不显著偏离中性期望值而且其平均值接近于零。

3.2 种间分化（Interspecific genetic divergence）

为了检测物种之间的分化程度，我们估算了每个基因座位的基因分化值 D_{xy} （图 3-2）。同时我们也计算了四个常用的物种分化参数：固定的分化位点，在一个物种中固定而另一个物种中变异的位点以及两个物种共享的变异位点(表 3-1)。



图 3-2 候选基因和参考基因的种间基因分化

注：A：油松和高山松，B：云南松和高山松，C：油松和云南松的 D_{xy} 值

Figure 3-2 Genetic divergence measured by D_{xy} between the three species

如上图所示,在油松和高山松以及油松和云南松的比较中，comt, dhx9, pho 和 wd40 显示了最高的 D_{xy} 值（图 3-2），与之对应 comt,pp2c 和 pho 拥有种间固定的差异位点（表 3-1）。这些基因可能在该复合体的物种形成中起到重要作用。把所有基因放在一起考虑，油松和云南松以及油松和高山松都有 41 个固定差异位点，而高山松和云南松之间只有两个固定差异位点，这揭示云南松跟高山松可能更近源，图 3-2 显示的物种间分化也支持这一结论。

表 3-1 油松,高山松和云南松的中间分化统计

Table 3-1 Summary statistics of genetic divergence among the three pine species

Locus	<i>P. tabulaeformis</i> & <i>P. densata</i>				<i>P. tabulaeformis</i> & <i>P. yunnanensis</i>				<i>P. densata</i> & <i>P. yunnanensis</i>			
	Fixed	p1	p2	shared	Fixed	p1	p2	shared	Fixed	p1	p2	shared
Climate related candidate genes												
<i>a3ip2</i>	0	3	13	12	0	2	12	13	0	6	6	19
<i>AGP4-2</i>	0	17	15	33	0	30	7	2	0	21	0	27
<i>aqua-MIP</i>	0	9	5	0	0	0	0	9	0	7	1	8
<i>ara</i>	0	5	1	11	0	3	1	3	0	9	1	3
<i>araR</i>	0	11	2	3	0			3	0	1	0	4
<i>ccoaomt</i>	0	7	6	7	0	0	4	4	0	5	0	8
<i>comt</i>	12	5	3	0	12	5	2	0	0	2	1	1
<i>dhn1</i>	0	16	24	25	0	23	11	20	0	30	11	19
<i>dhn2</i>	0	4	8	7	0	6	4	5	0	7	1	8
<i>dhn9</i>	0	13	87	14	0	6	16	11	0	76	2	25
<i>dhn7</i>	0	3	13	12	0	7	7	8	0	14	4	11
<i>erd3</i>	0	7	9	3	0	9	2		0	9	0	3
<i>GI</i>	0		19	4	0	3	2	2	0	9	0	4
<i>Glu</i>	0	2	76	14	0	16	2	0	0	81	1	11
<i>Hlhl</i>	0	4	24	2	0	6	0	0	0	27	0	0
<i>Ino3</i>	0	11	0	0	0		0	0	1	0	0	0
<i>LEA</i>	0	8	24	4	0	3	15	3	0	16	3	14
<i>lp3-1</i>	0		32	10	0	5	5	6	0	31	0	11
<i>MdhA</i>	0	5	15	19	0	8	9	6	0	10	1	24
<i>PHYO</i>	0	3	2	2	0	5		0	0	4	1	0
<i>Pod</i>	0	29	12	6	0	32	1	3	0	16	2	2
<i>pp2c</i>		5	4			4	2	2	0	3	2	2
<i>ppap12</i>	0	9	0	9	0	0	1	8	0	3	3	6
<i>RD21</i>	0	26	6	14	0	28	7	2	0	6	6	13
SUM can	13	234	400	222	13	303	112	141	1	403	46	223
Reference loci												
<i>2009</i>	0	13	2	5	0	14	13	14	0	3	13	15
<i>C3H</i>	0	9	16	6	0	0	13	5	0	6	2	16
<i>CesA2</i>	0		10	3	0	3	0		0	2	0	1
<i>GapCp</i>	0	8	9	9	0	6	1		0	7	1	2
<i>LHCA4</i>	0	7	9	6	0	0	4	3	0	8	0	8
<i>nir</i>	0		8	9	0	4	1	6	0	11	1	6
<i>PAL</i>	0	2	3	2	0	3	4		0	1	1	4
<i>Pho</i>	15	7	5	6	5	9	2	4	0	6	1	5
<i>pt-1</i>	0	0	6	3	0		1	2	0	16	0	3
<i>pt-2</i>	0			2	0		0	2	0	1	0	2
<i>WD40</i>	0	6	62	38	0	5	36	41	0	54	29	46

SUM_ref	15	75	145	105	15	100	75	82	0	143	48	110
Total	41	543	945	549	41	706	299	364	2	949	140	556

3.3 近代的本地选择信号

为了检测近代自然选择信号,我们用了 F_{ST} 异常值法,如下图所示(图 3-3),在油松中的互助居群和其他居群的比较中我们检测到了 3 个显著信号,而他们没有在把所有居群同时考虑时检测到。同时,在比较最南端的广元居群和其他居群时,我们没有检测到显著手选择的 SNPs。这说明边缘居群可能通过不同的位点适应于所在的气候环境。由于最南和最北居群之间相隔很远,基因流没有有效的把它们连接在一起,这些边缘居群本地适应相关位点可能呈条件中性。同时,群体结构分析显示油松中存在较大的群体分化 (Chen *et al.* 2008),分层结构模型应该用于检测近代选择 (Excoffier *et al.* 2009; De Mita *et al.* 2013)。

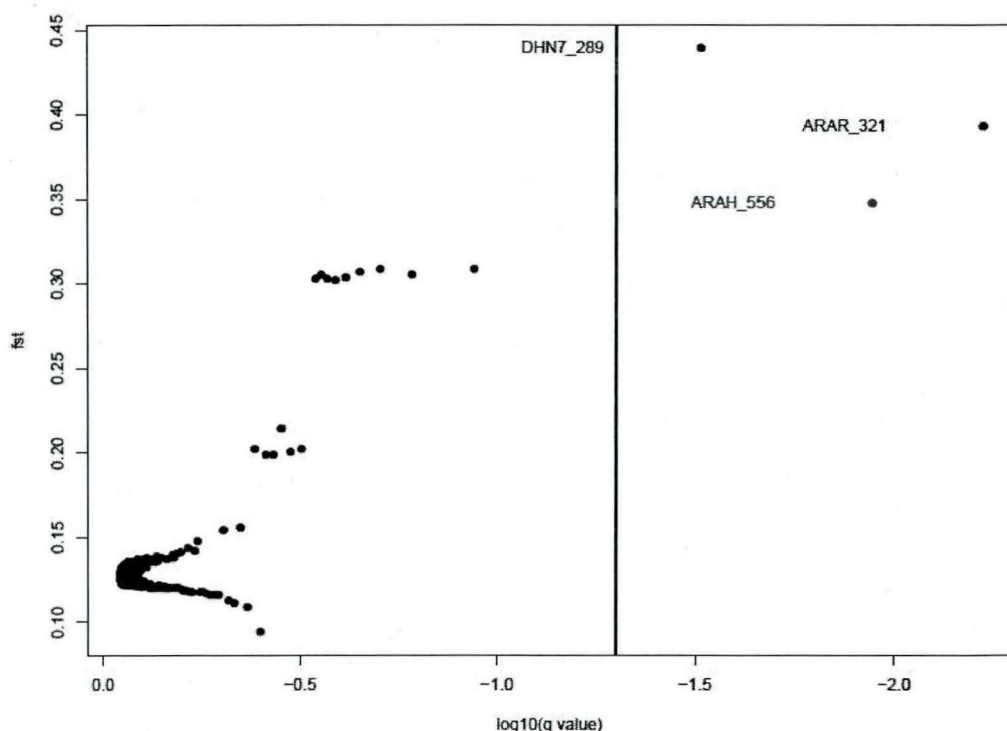


图 3-3 用 F_{ST} 异常值法在油松中检测本地选择信号

Fig. 3-3 F_{ST} outliers detected using Bayescan methods in *P. tabulaeformis*.

同时,我们也有最新的关联分析方法检测了等位基因频率和环境气候因子之间的显著关联关系。如图 3-4 所示,等位基因频率跟物种分布关键决定因子之间

的显著相关关系。

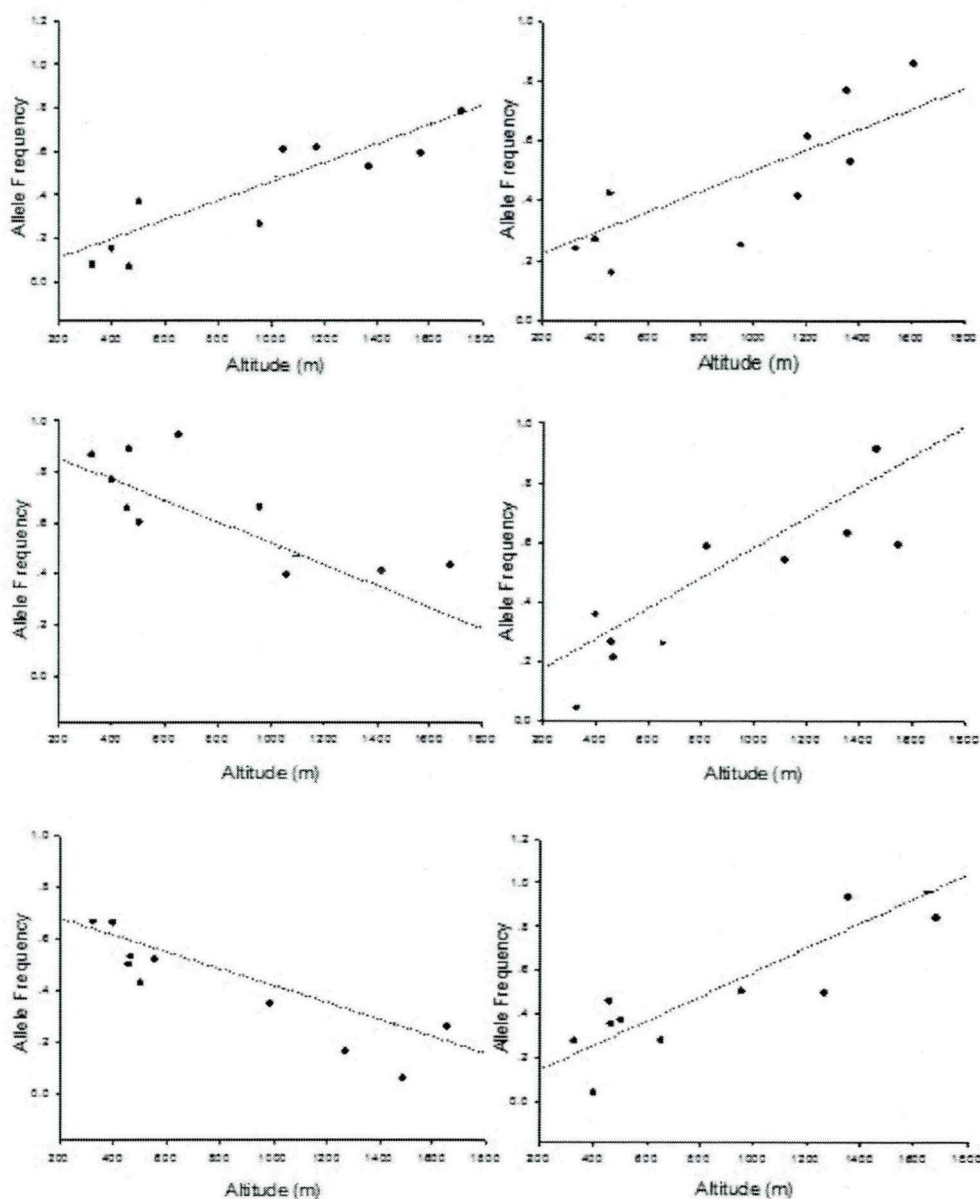


图 3-4 用 BAYENV 法在油松中检测等位基因频率和海拔的关系

Fig. 3-4 Examples of candidate SNPs that showed a significant regression of the transformed allele frequency on transformed altitudes

应用两种方法检测到的 SNPs 只有很少的一部分重合, 这可能说明这两种方法检测不同类型的适应性变异 (De Mita *et al.* 2013)。

在高山松中, 我们用了同样的方法检测到了多个适应性单核苷酸变异, 通过比较这些位点种内和种间的空间分布模式, 我们区分了源于新变异和适应性渗入或者固有

变异的自然选择作用。

3.4 古代选择信号

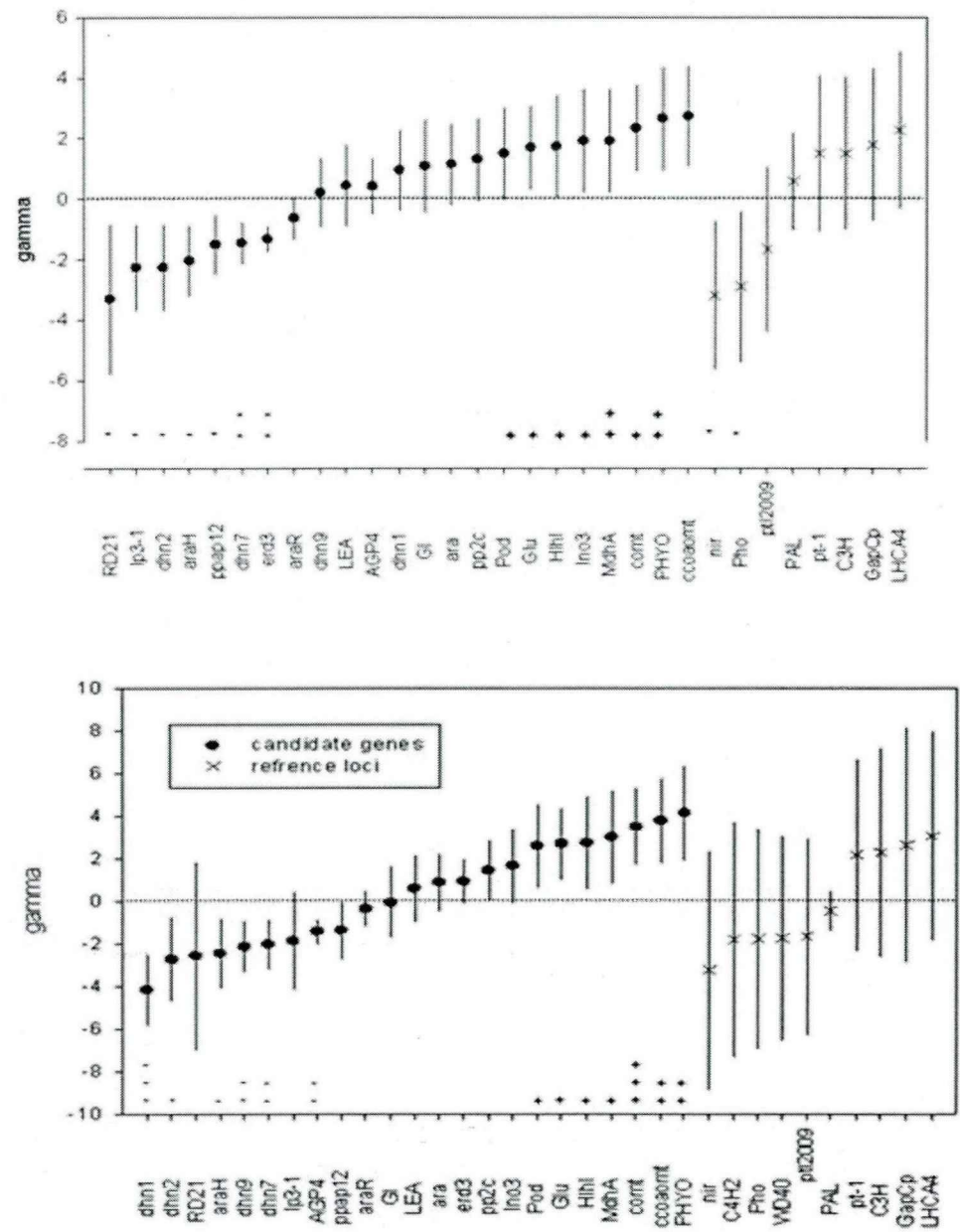


图 3-5 用 MKPRF 方法在油松-红松和高山松-红松谱系中检测到的古代自然选择信号

Figure 3-5 MKPRF analysis showing the mean and 95% credible interval of the posterior

distributions for selection coefficients (γ) at each locus for (a) *P. tabulaeformis* and (b) *P. densata*, a selection coefficient less than zero implies selective constraint at amino acid replacement sites while a coefficient greater than zero implies accelerated fixation of amino acid replacements. γ significantly greater or less than zero were implied using plus sign (+) and minus sign (-), respectively. +/-, $P < 0.05$; ++/--, $P < 0.01$ and +++/---, $P < 0.001$.

考虑到古代平衡选择也可能促进松柏类物种的气候适应, 我们应用 HKA, MK 以及他们的扩充软件检测了古代选择信号。如下图 3-5 所示。多个参考基因显示显著高或者低的选择系数。座位显示显著正或者显著负的 γ 值, 正值表示该基因可能受到了古代的选择扫荡, 而负值表示该基因可能受到古代的平衡选择或者古代净化选择, 而我们同时检测到了 MLHKA 检验的 K 值显著大于 1, 二者同时说明, 这些拥有负 γ 值的基因座位可能受到了古代平衡选择的作用。同时, 我们在油松-红松谱系分支上, 在两个参考基因上 (*pho* 和 *nir*, 图 3-3 中上图) 检测到了古代平衡选择的作用。

通过比较上下图, 我们发现, 两个谱系分支上检测到的受选择的基因基本相同, 同时, 这些基因也在其他近源松树物种, 比如黄山松和马尾松中显示了古代选择信号 (Zhou *et al.* 2014), 因为这些近源物种共享很长的群体历史。

3.5 基因流

因为居群之间以及物种之间的基因流可能促进了种群的气候变化适应。我们用中性遗传的基因座位估计了油松居群之间的基因流。根据单倍型和核苷酸序列数据, 采用不同的方法估测了油松群体间的基因流, 结果见表3-6。根据 Nei (1973) 提出的方法, 由 G_{st} 估测出的基因流 (N_m) 比其它方法的估测值要大。从结果可以看出, 油松种群间有较大的基因流。

表 3-2 个居群间的基因流估测
Table3-2 Estimation of gene flow between the 5 populations

Population		Haplotype Data Information (Nei 1973)		Sequence Data Information (Nei 1982)			Lynch and Hudson, Crease and 1990 1992		Slatkin Maddison	
P 1	P 2	Gst	Nm	DeltaSt	GammaSt	Nm	Nst	Nm	Fst	Nm
GN	TS	0.00030	840.00	0.00091	0.10670	2.09	0.06897	3.37	0.06845	3.40
GN	ZWL	0.03128	7.74	0.00036	0.04157	5.76	-0.09051	-3.01	-0.09012	-3.02
GN	HZ	-0.02917	-8.82	0.00028	0.03676	6.55	-0.06466	-4.12	-0.06455	-4.12
GN	HLS	0.00746	33.25	0.00036	0.04732	5.03	-0.06125	-4.33	-0.06124	-4.33
TS	ZWL	0.03495	6.90	0.00100	0.10892	2.05	0.05812	4.05	0.05796	4.06
TS	HZ	0.03604	6.69	0.00099	0.11970	1.84	0.10892	2.05	0.10865	2.05
TS	HLS	0.03210	7.54	0.00108	0.13065	1.66	0.11496	1.92	0.11411	1.94
HZ	HLS	0.04815	4.94	0.00024	0.03272	7.39	-0.07182	-3.73	-0.07161	-3.74
ZWL	HZ	0.05625	4.19	0.00034	0.04124	5.81	-0.07361	-3.65	-0.07313	-3.67
ZWL	HLS	0.01770	13.88	0.00038	0.04625	5.16	-0.08143	-3.32	-0.08129	-3.33
GN	ZWL	0.03128	7.74	0.00036	0.04157	5.76	-0.09051	-3.01	-0.09012	-3.02
Species		0.00855	28.97	0.00109	0.13482	1.60	-0.09058	-3.01	-0.09059	-3.01

注：只有在较大的居群个较大的居群做了基因流的估计。

油松中的基因流可能主要由于花粉介导的，因为早期基于细胞质遗传标记的分析结果显示叶绿体单倍型没有显著的空间结构，而线粒体单倍型分布显示显著的空间结构（Chen *et al.* 2008）。

3.6 物种分布模拟

应用气候数据和我们以及标本的采样点的地理信息，我们对油松和高山松的分布区域进行了模拟，结果如图 3-6 所示，模拟所得的潜在分布区基本与现实相符。油松主要在中国北方，而高山松主要分布在青藏高原以及周边区域。最早月降雨是油松分布的主要决定因素而海拔和最冷月的气温是高山松分布的主要决定气候因子，我们用这些因子进行下一步的气候因子和基因频率关联分析。

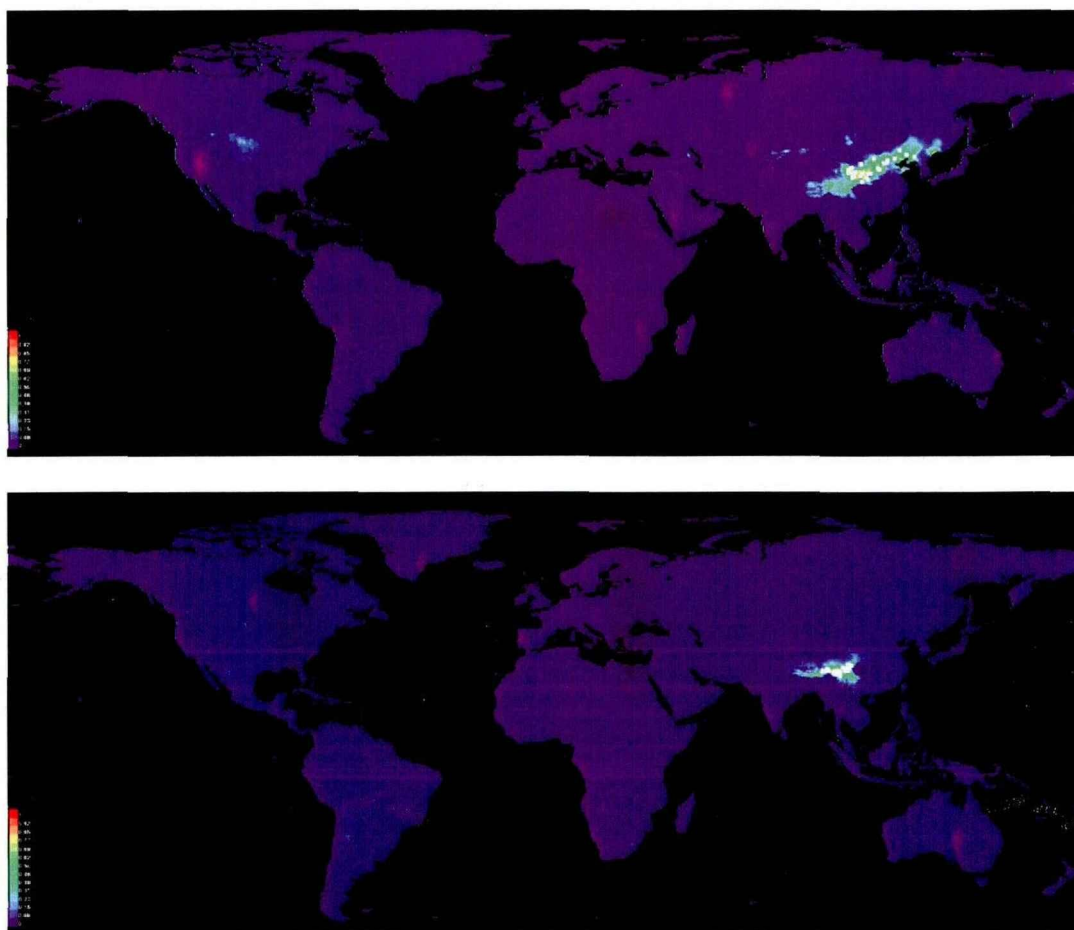


图 3-6 油松(上)和云南松的潜在分布区域模拟结果

Fig. 3-6 Species distribution modeling for *Pinus tabulaeformis* and *Pinus densata*

如表 3-3 所示,各个气候因子对物种分布模拟的贡献各不相同。由于海拔是决定物种分布的重要因子,因此我们在模拟油松和高山松分布范围中时也把海拔作为第 20 个气候因子。总体来说,如上图 3-6 所示,模拟所得的分布范围基本与物种的现实分布范围相同。在高山松中,海拔贡献在物种分布范围模拟中有 52.9% 的贡献。其他气候因子,季节气温,最热月降雨,最冷季气温和季节降雨次之(表 2-1 和表 3-3)。在油松中,海拔的重要性相对较弱,只有 9.6%,而冬季降雨成为最重要的决定因子,含有 20.9% 的相对贡献,最早季均温,最冷月均温以及夏季降雨次之。在现实中,高山松分布于高海拔的青藏高原的寒旱环境,油松主要分布在中国北方,夏季干旱,冬季干旱和寒冷并存。我们进一步分析了各个物种中贡献前五的气候因子之间的相关性,并把他们用在后续的气候因子和等位基因频率的相关性分析中。

表 3-3 物种分布模拟中各个气候因子的相对重要性

Table3-3 Percent contribution and permutation importance of each enviromental variable in *P. densata* and *P. tabulaeformis*

Variable	Percent contribution	Permutation importance	Variable	Percent contribution	Permutation importance
<i>P. densata</i>			<i>P. tabulaeformis</i>		
alt	52.9	5	bio19	20.9	17.2
bio4	22.9	61.2	bio9	13	2.6
bio18	10.6	0.3	bio6	10.1	1.2
bio15	9.9	0	bio18	9.8	0.1
bio11	2.2	27.4	bio1	9.7	1.8
bio19	0.5	2.8	alt	9.6	1.7
bio2	0.4	0.1	bio15	8.3	0.1
bio17	0.2	0	bio3	5.3	11.8
bio8	0.1	0.4	bio4	4.1	0.5
bio3	0.1	2.7	bio14	4.1	57.8
bio1	0.1	0	bio5	3.2	0.2
bio14	0	0	bio13	1.5	0.3
bio6	0	0	bio12	0.2	1.9
bio16	0	0	bio17	0.1	0
bio9	0	0	bio11	0.1	2.8
bio12	0	0	bio2	0.1	0
bio5	0	0.1	bio16	0	0
bio13	0	0	bio7	0	0.1
bio10	0	0	bio10	0	0
bio7	0	0	bio8	0	0

第四章 讨论

4.1 核苷酸多态性

与他人所做的相关工作相比，我们得到了相似的核苷酸多态性估计（见表4-1）。我们用了目前在松柏类植物中最长的序列对三个物种的遗传多样性进行分析。与 Ma *et al.* (2006) 的结果一致，高山松拥有最高的核苷酸多样性（表4-1），这可能与他的空间分布以及杂交物种形成有关。由于高山松与油松和云南松共享大多数的变异位点，因此种间基因流或者祖先变异是可能的原因。与其他欧洲和北美的近源松树相比，油松，高山松和云南松都有显著高的遗传多样性。一方面，如欧洲赤松和火炬松，它们有连续的分布，因此，种内基因流使得核苷酸表变异均一化。而在油松，高山松以及云南松中，居群之间的地理隔离大，物种分布片段化，使得居群之间的分化较高，从而贡献于较高的核苷酸多样性。另一方面，由于高山松生存在青藏高原的寒旱环境中，自然选择，尤其是平衡选择可能促进了核苷酸变异的产生。

表 4-1 核苷酸多态性比较分析

Table 4-1 Comparison of genetic diversity in coniferous species						
Species	loci	Length bp)	θ_{sil}	π_{sil}	Specieswide	Reference
<i>Pinus densata</i>	7	3040	0.0143	0.0122	Yes	Ma <i>et al.</i> (2006)
<i>Pinus pinaster</i>	10	4746	0.0017	0.0024	Yes	Pot <i>et al.</i> (2005)
<i>Pinus radiata</i>	10	4746	0.0008	0.0007	No	Pot <i>et al.</i> (2005)
<i>Pinus</i> <i>abuliformis</i>	7	3040	0.0153	0.0119	Yes	Ma <i>et al.</i> (2006)
<i>Pinus sylvestris</i>	2	4136	0.0030	0.0019	Yes	Garcia-Gil <i>et al.</i> (2003)
<i>Pinus taeda</i>	19	17580	0.0066	0.0064	Yes	Brown <i>et al.</i> (2004)
<i>Pinus taeda</i>	18	10116	0.0079	0.0085	No	Gonzalez-Martinez <i>et al.</i> 2006)
<i>Pinus</i> <i>runnanensis</i>	7	3040	0.0077	0.0095	Yes	Ma <i>et al.</i> (2006)
<i>Pinus</i> <i>abuliformis</i>	37	24389	0.0083	0.0083	Yes	This study
<i>Pinus densata</i>	37	24389	0.010	0.010	Yes	This study
<i>Pinus</i> <i>runnanensis</i>	37	24389	0.0071	0.0073	Yes	This Study

θ_{sil} average θ_w at silent sites; π_{sil} average π at silent sites.

由于随机遗传漂变、选择和基因流都会引起核苷酸多态性的改变,随机遗传漂变作用于所有位点,而选择作用于单个位点以及临近区域 (Galtier *et al.* 2000)。同时,基因内各个间位点发生了较多的连锁(表 4-2),这说明随机遗传漂变可能对这些变异的发生起了较大作用,而 π_a/π_s 值在所有居群均小于 1,

4.2 连锁不平衡

理论上,连锁不平衡值(LD)受多种进化和统计因素影响,例如:选择,重组,基因流,杂交和瓶颈效应等,但是中性检测在所有居群中非显著,说明选择不是影响油松 LD 的主要因素。同时,由于该实验检测到的基因流较小,重组率较低,也不是影响油松 LD 的主要因素,杂交也不会很明显的影响 LD 值(表 4-2)。所以瓶颈效应成了影响油松 LD 值的主要因素。第四纪冰期造成松科植物生境的片断化,从而产生松科植物进化上的瓶颈效应,进一步影响了今天油松的分布格局和基因结构。

表 4-2 ARA 各信息位点间连锁不平衡的统计显著性

Table 4-2 Statistical significance of linkage disequilibria among informative sites of ARA loci

S	246	267	279	303	315	336	367	382	387
8	C=0.317	C=0.671	C=0.003	C=0.013	C=0.584	C=0.000	C=0.000	C=0.000	C=0.336
7	$\chi^2=1.923$	$\chi^2=0.466$	**	*	$\chi^2=0.900$	***B	***B	***B	$\chi^2=1.567$
			$\chi^2=8.644$	$\chi^2=7.219$		$\chi^2=16.004$	$\chi^2=24.505$	$\chi^2=19.899$	
			**	**		***B	***B	***B	
2		C=0.327	C=0.071	C=0.024	C=1.000	C=0.003	C=0.327	C=0.346	C=0.000
4		$\chi^2=1.705$	$\chi^2=4.179$	*	$\chi^2=0.435$	**	$\chi^2=1.705$	$\chi^2=1.497$	***B
6			*	$\chi^2=5.947$		$\chi^2=11.649$			$\chi^2=40.741$
				*		***B			***B
2			C=0.008	C=0.722	C=0.098	C=0.046	C=0.663	C=0.665	C=0.569
6			**	$\chi^2=0.340$	$\chi^2=4.344^*$	*	$\chi^2=0.781$	$\chi^2=0.542$	$\chi^2=1.389$
7			$\chi^2=7.661$			$\chi^2=4.392$			
			**			*			
2				C=0.000	C=0.279	C=0.001	C=0.008	C=0.009	C=0.142
7				***B	$\chi^2=1.956$	***B	**	**	$\chi^2=3.405$
9				$\chi^2=18.783$		$\chi^2=10.767$	$\chi^2=7.661$	$\chi^2=6.727$	
				***B		***B	**	**	
3					C=0.279	C=0.000	C=0.000	C=0.000	C=0.005
0					$\chi^2=1.956$	***B	***B	***B	**
3						$\chi^2=28.663$	$\chi^2=20.395$	$\chi^2=17.908$	$\chi^2=9.064$
						***B	***B	***B	**

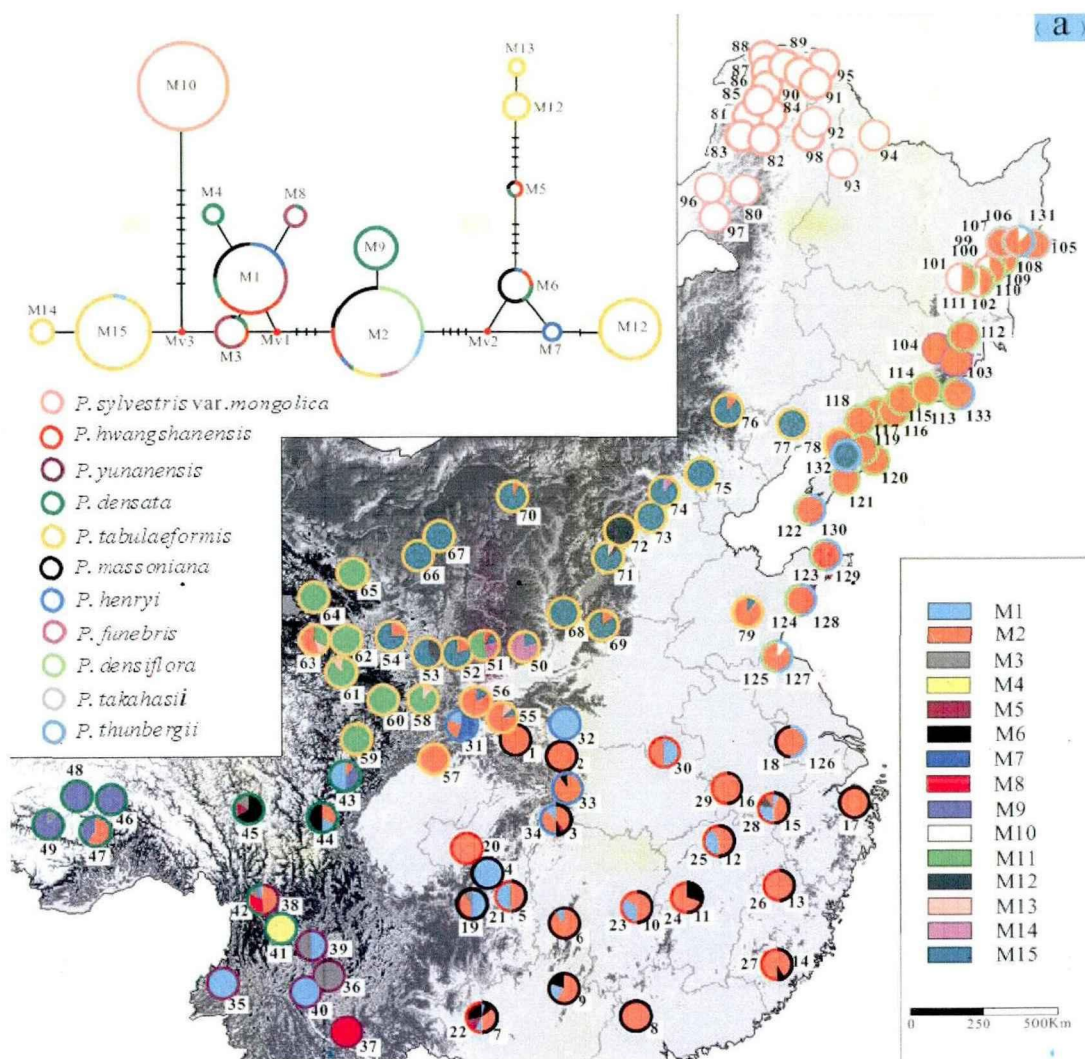
3	C=0.558	C=1.000	C=1.000	C=1.000
1	$\chi^2=1.121$	$\chi^2=0.798$	$\chi^2=0.701$	$\chi^2=0.355$
5				
3		C=0.000	C=0.000	C=0.001
3		***B	***B	***B
6		$\chi^2=18.945$	$\chi^2=15.294$	$\chi^2=15.812$
		***B	***B	***B
3			C=0.000	C=0.569
6			***B	$\chi^2=1.389$
7			$\chi^2=43.902$	
			***B	
3				C=0.570
8				$\chi^2=1.220$
2				

0.01<P < 0.05, *; 0.001<P<0.01, **; P<0.001, ***; B, significant by the Bonferroni procedure; S, Sites; C, Fisher’parameter; χ^2 , Chi-square.

以油松的 ARA 基因座位为例，变异位点之间显示了多对显著的连锁群，这可能是导致该位点核苷酸多样性较低的原因。同时，连锁重组信息直接跟选择和种群历史相关。连锁不平衡分析也可用于检测自然选择和群体动态历史。

4.3 适应性渗入 (Adaptive introgression) 以及高山松适应于青藏高原的旱寒环境

在植物基因组中核 DNA 进化速率最快，叶绿体次之，线粒体最慢 (Wolfe *et al.* 1987)。由于双亲遗传特性，核基因序列提供了更丰富的遗传信息。植物核基因的核苷酸多态性研究最早的是草本植物，尤其是模式植物拟南芥及其近缘种 (Aguade 2001)。这些物种杂合个体的等位基因序列主要依靠克隆得到，成本很高，故被子植物核基因序列的群体遗传学研究最初只局限于几种模式植物。随着测序技术的普及，利用核基因序列研究其它非模式被子植物的群体遗传学陆续报道出来 (Aguade 2001; Schneider *et al* 2001; Zhu *et al.* 2003)。多数松杉类植物雌配子体 (胚乳) 为单倍体，这一特点使得获取松杉类植物的等位基因序列变得容易，加上这些物种在林木中的重要性，近些年来利用核基因序列开展松杉植物的群体遗传和谱系地理研究不断出现。



较早报道的是 Dvornyk 等 (2002) 对欧洲赤松 (*Pinus sylvestris* Linn.) 的研究, Dvornyk 等 (2002) 对欧洲赤松 (*Pinus sylvestris* Linn.) 的研究, 随后, 在其它一些裸子植物, 如 *Cryptomeria japonica* (Kado et al. 2003)、*Pinus taeda* Linn. (Brown et al. 2004; Neale & Savolainen 2004)、*Pseudotsuga menziesii* Franco (Neale & savolainen 2004)、*Picea abies* Karst. (Bouille & Bousquet 2005)、*Pinus pinaster* Ait.、*Pinus radiata* D.Don (Pot et al. 2005)、*Pinus densata* Mast. (Ma et al. 2006) 等物种中, 依据多座位序列的群体遗传学研究陆续被发表。这些研究所用片段都是功能基因座位, 大都涉及到选择和适应的进化问题。

在高山松中, 我们检测到结果表明近代的自然选择和古代的平衡选择同时作用于高山松复合体的本地适应。本研究在多个与气候相关的候选基因如 *agp4*, *aqua-MIP*, *dhn7*, *dhn9* 和 *lp3-1* 检测到了近代正选择的作用。在三个物种中, 特异的基因座位受到了近代选择的影响而基本相同的基因受到了古代选择的影响,

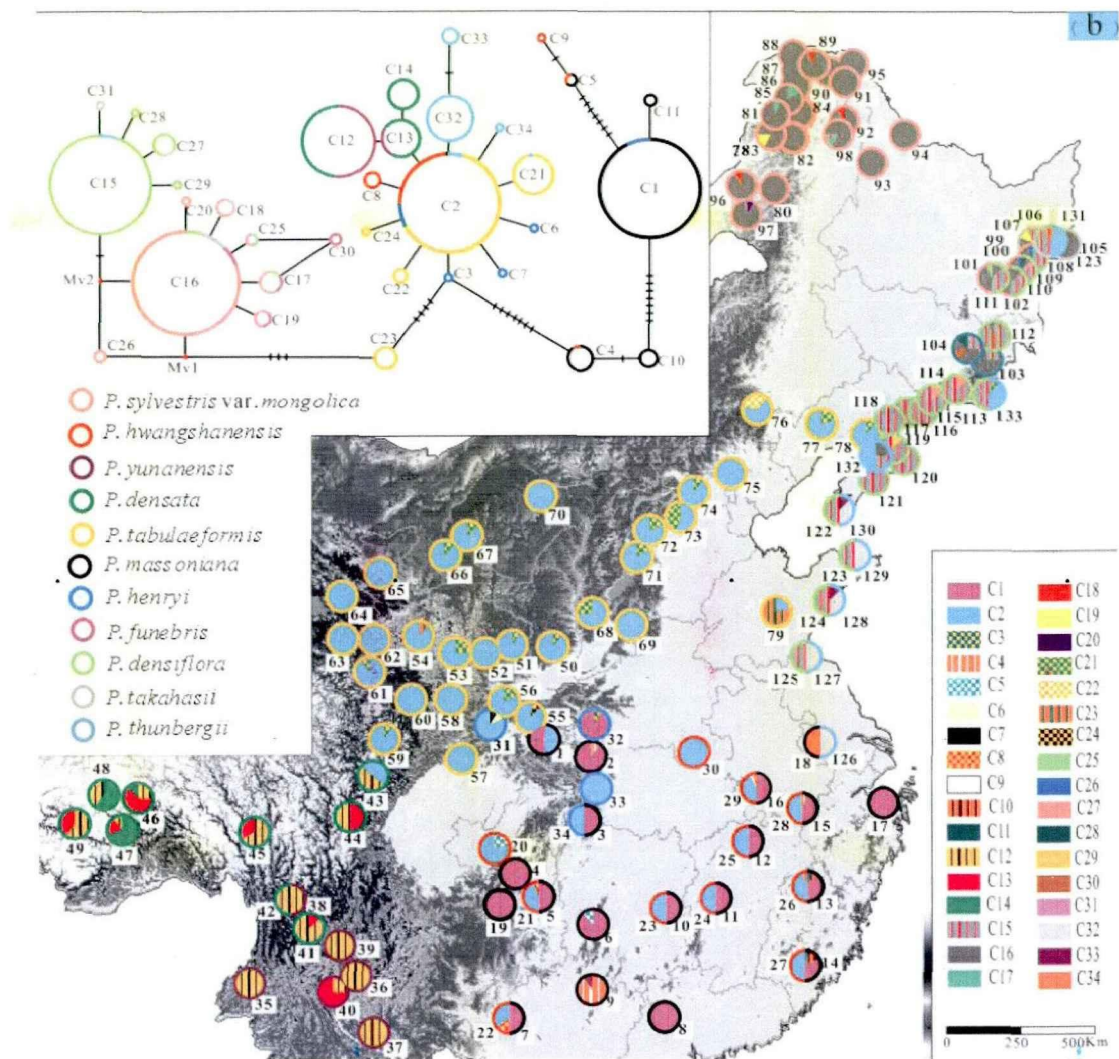


图4-1 油松和云南松以及近缘种的叶绿体（上图）和线粒体（前一页）基因型的分布

Fig. 4-1 Distribution of chloroplast and mitochondrial haplotypes in *Pinus tabulaeformis* and *Pinus densata* and their closely related species

这可能是由于三个物种共享较长的进化历史。在最南和最北的边缘居群中，不同的单核苷酸多态位点被检测到本地适应的信号，这些位点可能呈条件中性。新的变异和固有变异同时贡献于高山松居群的本地适应。来自油松和云南松的适应性渗入促进了高山松适应于青藏高原的寒旱环境。种内居群主要被花粉流连接起来，同时可能传递适应性变异。种间的适应性共享可能是由于种子流所导致的，但是这种共享也可能由于不完全谱系筛选。

第五章 总结与展望

全球气候正在以前所未有的速度变化着,现存生物的进化能不能赶上气候变化的速度?适应和物种形成是进化生物学中两大核心问题。物种多样性是如何形成的?什么因素在物种形成过程中起怎样的作用?以及物种如何适应于它所处的环境?诸多问题亟待回答。松柏类植物起源于石炭纪(360-286 Mya),它们长时间的见证了地球环境变化的过程。研究表明大多数物种的居群在某些关键生理或者生态特征上已适应于其所在环境或者生态位。同时,近缘物种之间存在基因流。它们正在物种形成的过程中。这使得松柏类植物成了研究适应和物种形成的理想材料。松属是松柏类植物的主要类群。我们主要关注松亚组。它们以不同的模式广泛分布于北半球,欧洲赤松的分布区横跨亚欧大陆,而其他物种主要以异域或者领域的方式分布在中国以及其他亚洲地区。到目前为止,仍然有很多待解决的问题:比如是选择作用于新的突变还是固有突变促进物种的气候适应?我们采集了国内分布的该亚组高山松复合体(高山松以及亲本种油松和云南松)的68个天然居群,测序了4个叶绿体片段2个线粒体片段以及37个核基因座位。首先我们检测不同进化尺度上的选择信号,包括群体或者物种内的近代正选择;油松-高山松,高山松-云南松以及云南松-油松等进化分支上的选择(大约分化时间为5百万年左右;更进一步,我们检测红松-高山松,红松-油松以及红松-云南松三个分支上的古代的平衡选择,净化选择和平衡选择。我们补充应用了多种不同的最新的检测发展的检测方法。因为高山松是有油松和云南松杂交形成的,我们可以通过检测高山松中适应性突变是高山松独有的还是来自某个亲本种,从而分辨该物种的气候适应是由于新突变或者固有变异引起的。同时我们应用GENELAND研究了适应性突变的地理分布模式。进一步揭示种内基因流(基因流协助, assisted gene flow)和种间基因渗入(适应性渗入, adaptive introgression)对高山松居群气候适应的贡献。

我们比对了12个参考基因(reference loci)和25个环境适应相关的候选基因(candidate genes)在高山松,油松和云南松中的种内多样性和种间分化。候选基因显示的种间分化显著高于参考基因。而且多种检测方法揭示候选基因经历了选择。候选基因和参考基因在这两个物种分化的过程中分别扮演了领头者(leaders)和跟随者(followers)角色。选择可能促进了候选基因的分化。近代选择信号显示了物种特异性,换句话说,物种间的歧化选择。这种歧化选择可能限制了所在基因组局域以及邻近相连局域的基因交流,促进了物种间的生态物种形成。在油松和高山松的边缘居群和中心居群中,本地选择作用于不同的突变,

说明物种通过不同的机制去适应于极端的边缘环境。高山松中，适应性突变分别为其特有的，与油松共享的，与云南松共享的，这说明新变异和固有变异同时促进了高山松在青藏高原的气候适应。

本研究对我们理解进化生物学现象，尤其是本地适应以及它在同倍体物种形成中的作用。同时，该论文也有广泛的应用前景，尤其在全球气候变化的大背景下，物种的气候变化适应，以及高山松，油松和云南松的分子育种。

参考文献

- Abbott RJ, Smith LC, Milne RI, *et al.* (2000) Molecular analysis of plant migration and refugia in the Arctic. *Science* **289**, 1343-1346
- Aitken SN, Yeaman S, Holliday JA, Wang TL, *et al.* (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**, 95-111.
- Alberto F, Aitken S, Alia A, *et al.* (2013) Evolutionary response to climate change—evidence from tree populations. *Global Change Biology* **19**, 1645-1661.
- Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, London, UK.
- Barton NH (1999) Clines in polygenic traits. *Genet. Res.* **74**, 223-236.
- Beaumont MA 2005. Adaptation and speciation: what can F_{ST} tell us? *Trends Ecol. Evol.* **20**, 435-440.
- Beaumont MA, Nichols RA 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B-Biological Sciences* **263**, 1619-1626.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025-2035.
- Bower AD, Aitken SN (2008) Ecological genetics and seed transfer guidelines for *Pinus albicaulis* (Pinaceae). *American Journal of Botany*, **95**, 66-76.
- Brown GR, Gill GP, Kuntz RJ, *et al.* (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15255-15260.
- Bustamante CD, Nielsen R, Sawyer SA, *et al.* (2002) The cost of inbreeding in *Arabidopsis*. *Nature* **416**, 531-534.
- Chen KM, Abbott RJ, Milne RI, *et al.* (2008) Phylogeography of *Pinus tabulaeformis* Carr. (Pinaceae), a dominant species of coniferous forest in northern China. *Mol Ecol.* **17**, 4276-4288.
- Chen J, Källman T, Ma X, *et al.* (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.
- 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*. Sinauer Associates, Sunderland, MA.
- Crispo E, Bentzen P, Reznick DN, *et al.* (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology* **15**, 49-62.
- Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**, 410-418.
- 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.
- Dobzhansky TH (1936) Studies on hybrid sterility. 11. localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* **21**, 113-135.
- 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.
- Dvornyk V, Sirvio A, Mikkonen M, Savolainen O (2002) Low nucleotide diversity at the *pall* locus in the widely distributed *Pinus sylvestris*. *Molecular Biology and Evolution* **19**, 179-188.
- Eckert AJ, van Heerwaarden J, Wegrzyn JL, *et al.* (2010) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* **185**, 969-982.
- Eckert AJ, Wegrzyn JL, Pande B, *et al.* (2009) Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold-hardiness in coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii*). *Genetics* **183**, 289-298.
- Edgar RC (2004) MUSCLE, multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.
- 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. *Molecular Biology and Evolution* **25**, 417-437.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE, a simulation study. *Mol. Ecol.* **14**, 2611-2620.
- Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285-298.
- Fariello MI, Boitard S, Naya H, *et al.* (2013) Detecting signatures of selection through haplotype differentiation among hierarchically-structured populations. *Genetics* **193**, 929-941.
- Fisher RA (1930) *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford [Variorum edition, Bennett J. H. (Editor), 1999, Oxford University Press, Oxford].
- Foll M, 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.
- Franks SJ, Hoffmann AA (2012) Genetics of climate change adaptation. *Annual Reviews of Genetics* **46**, 185-208.
- Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution* **30**, 1687-1699.
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* **133**, 693-709.
- Gao JB, Wang JF, Mao P, *et al.* (2012) Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. *Mol. Ecol.* **21**, 4811-4827.
- Garci-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Molecular Ecology* **12**, 1195-1206.
- 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, *et al.* (2006) DNA sequence variation and selection of tag singlenucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* **172**, 1915-1926.
- Gossmann TI, Song BH, Windsor AJ *et al.* (2010) Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Molecular Biology and Evolution* **27**, 1822-1832.

- 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 Phytologist* **184**, 1016-1028.
- Gross BL, Rieseberg LH (2005) The ecological genetics of homoploid hybrid speciation. *J. Hered.* **96**, 241-252.
- Gunther T, Coop G (2012) Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205-220.
- Hall D, Ma XF, Ingvarsson PK (2011) Adaptive evolution of the *Populus tremula* photoperiod pathway. *Molecular Ecology* **20**, 1463-1474.
- Hedrick P (2011) *Genetics of populations*. Jones and Bartlett Publishers, Sudbury, Massachusetts, USA.
- Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.
- Hey J (2010) Isolation with Migration models for more than two populations. *Molecular Biology and Evolution* **27**, 905-920.
- Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theoretical Applied Genetics* **38**, 226-231.
- Hoffman AA, Sgro CM (2011) Climate change and evolutionary adaptation. *Nature* **470**, 479-485.
- Hohenlohe PA, Phillips PC, Cresko WA (2010) Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Sciences* **171**, 1059-1071.
- Holliday JA, Ritland K, Aitken SN (2010) Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist* **188**, 501-514.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.
- Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337-338.
- Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153-159.
- Keller SR, Levens N, Ingvarsson PK, et al. (2011) Local selection across a latitudinal gradient shapes nucleotide diversity in balsam poplar, *Populus balsamifera* L. *Genetics* **188**, 941-952.
- Keller SR, Levens N, Olson MS, Tiffin P (2012) Local adaptation in the flowering time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution* **29**, 3143-3152.
- Kimura M (1968) Evolutionary rate at the molecular level. *Nature* **217**, 624-626.
- Kingman JFC (1982) The coalescent. *Stochastic Processes and their Applications* **13**, 235-248.
- Krauskopf E, Harris PJ, Putterill J (2005) The cellulose synthase gene *PrCESA10* is involved in cellulose biosynthesis in developing tracheids of the gymnosperm *Pinus radiata*. *Gene*, **350**, 107-116.
- 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 Genetics & Genomes* **8**, 1451-1467.
- Lankinen P (1986) Geographical variation in circadian eclosion rhythm and photoperiodic adult diapause in *Drosophila littoralis*. *J. Comp. Physiol. A*. **159**, 123-142 .
- Lee CR, Mitchell-Olds T (2011) Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology* **20**, 4631-4642.
- 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
- 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. *Bioinformatics* **25**, 1451-1452.
- Ma F, Zhao CM, Milne R, *et al.* (2010) Enhanced drought-tolerance in the homoploid hybrid species *Pinus densata*: implication for its habitat divergence from two progenitors. *New Phytol.* **185**, 204-216
- Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology and Evolution* **28**, 614-621.
- Mao JF, Wang XR (2011) Distinct niche divergence characterizes the homoploid hybrid speciation of *Pinus densata* on the Tibetan Plateau. *American Naturalist* **177**, 424-439.
- 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. *Molecular Biology and Evolution* **23**, 807-816.
- Ma XF, Hall D, StOnge K, *et al.* (2010) Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. *Genetics* **186**, 1033-1044.
- Mayr E (1963) *Animal Species and Evolution*, Cambridge, Belknap Press of Harvard University Press.
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the ADH locus in *Drosophila*.
- Mikola J (1982) Bud-set phenology as an indicator of climatic adaptation of Scots pine in Finland. *Silva Fenn.* **16**, 178-184
- Miller CN (1977) Mesozoic conifers. *Bot. Rev.* **43**, 217-280.
- Mosca E, Eckert AJ, Di Pierro EA *et al.* (2012) The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular Ecology* **21**, 5530-5545.
- Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* **12**, 111-122.
- Neale DB, Wegrzyn JL, Stevens KA, *et al.* (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* **15**, R59
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA* **76**, 5269-5273.

- Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**, 145-63.
- Nielsen R (2005) Molecular signatures of natural selection. *Ann. Rev. Genet.* **39**, 197-218.
- 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.
- Olson MS, Levens N, Soolanayakanahally RY, *et al.* (2013) The adaptive potential of *Populus balsamifera* L. to phenology requirements in a warmer global climate. *Molecular Ecology* **22**, 1214-1230.
- Olsson K, Ågren J (2002) Latitudinal population differentiation in phenology, life history and flower morphology in the perennial herb *Lythrum salicaria*. *J. Evol. Biol.* **15**, 983-996.
- Palme A, Pyhäjärvi T, Wachowiak W, Savolainen O (2009) Selection on nuclear genes in a *Pinus* phylogeny. *Molecular Biology and Evolution* **26**, 893-905.
- Phillips SJ, Anderson, RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259.
- Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161-175.
- Provine WB (1971) *The Origins of Theoretical Population Genetics*. University Chicago Press: Chicago.
- Prunier J, Gerardi S, Laroche J, *et al.* (2012) Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. *Molecular Ecology* **21**, 4270-4286.
- Prunier J, Laroche J, Beaulieu J, Bousquet J (2011) Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology* **20**, 1702-1716.
- Pyhäjärvi T, Garcia-Gil MR, Knurr T, *et al.* (2007) Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics* **177**, 1713-1724.
- Richardson BA, Rehfeldt GE, Kim MS (2009) Congruent climate-related genecological response from molecular markers and quantitative traits for western white pine (*Pinus monticola*). *International Journal of Plant Sciences* **170**, 1120-1131.
- Richardson DM(Ed) (2000) *Ecology and biogeography of Pinus*. Cambridge University Press.
- Rieseberg LH (1997) Hybrid origins of plant species. *Annu. Rev. Ecol. Evol. Syst.* **28**, 359-389.
- Rieseberg LH, Vanfossen C, Desrochers AM (1995) Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* **375**, 313-316.
- Rosenberg NA, Feldman MW (2002) The relationship between coalescence times and population divergence times. In *Modern Developments in Theoretical Population Genetics*. Edited by Slatkin M, Veuille M. Oxford: Oxford University Press 130-164.
- Savolainen O (2011) The genomic basis of local climatic adaptation. *Science* **334**, 49-50.
- Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**, 807-820.
- Song BH, Wang XQ, Wang XR, *et al.* (2003) Cytoplasmic composition in *Pinus densata* and population establishment of the diploid hybrid pine. *Molecular Ecology* **12**, 2995-3001.
- Sunnåker M, Busetto AG, Numminen E, *et al.* (2013). Approximate bayesian computation. *PLoS*

- computational biology* 9(1), e1002803.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585-595.
- Thomas CD, Cameron A, Green RE, *et al.* (2004) Extinction risk from climate change. *Nature* 427, 145-148.
- 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 Genetics & Genomes*. 5, 117-132.
- Wagner DB, Furnier GR, Saghai-Maroo MA, *et al.* (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 B, Mao JF, Gao J, *et al.* (2011) Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. *Molecular Ecology* 20, 3796-3811.
- Wang IJ, Summers K (2010) Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology* 19, 447-458.
- Wang XR, Szmidi AE (1994) Hybridization and chloroplast DNA variation in a *Pinus* species complex from Asia. *Evolution* 48, 1020-1031.
- Wang XR, Tsumura Y, Yoshimaru H, *et al.* (1999) Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpl20-rps18* spacer, and *trnV* intron sequences. *American Journal of Botany* 86, 1742-1753.
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7, 256-276.
- Watterson GA (1978) "The homozygosity test of neutrality". *Genetics* 88, 405-417.
- Willyard A, Syring J, Gernandt DS, *et al.* (2007) Fossil calibration of molecular divergence infers a moderate mutation rate, recent radiations for *Pinus*. *Molecular Biology and Evolution* 24, 90-101.
- Wright S (1921) Correlation and causation. *Journal of Agricultural Research* 20, 557-585.
- Wright S (1931) Evolution in Mendelian populations. *Genetics* 16, 97-159.
- Wright S (1943) Isolation by distance. *Genetics* 28, 114-138.
- Wu CY (1995). *Vegetation of China* 2nd edn Science Press: Beijing.
- 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, Charlesworth B (2004) The HKA test revisited: a Maximum-Likelihood-Ratio test of the standard neutral model. *Genetics* 168, 1071-1076.
- Yoder JB, Stanton-Geddes J, Zhou P, *et al.* (2014) Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics* 196, 1263-1275.
- Zeng K, Mano SH, Shi SH, Wu CI (2007) Compound tests for the detection of hitchhiking under positive selection. *Molecular Biology and Evolution* 24, 1898-1908.
- Zhao W, Meng J, Wang B, Zhang L, *et al.* (2014) Weak crossability barrier but strong juvenile

selection supports ecological speciation of the hybrid pine *pinus densata* on the tibetan plateau. *Evolution* **10**, 12496.

Zhou YF, Abbott RJ, Jiang ZY, *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 pines species in Southeast China. *Molecular Ecology* **23**, 3504-3522.

在学期期间的研究成果

1. **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. Times cited: 45
2. **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.
3. **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 (under review).
4. **Zhou YF, Zhu HT, Savolainen O, Liu JQ** (2014) Local adaptation to Climate in Chinese pine (*Pinus tabulaeformis*). (draft)
5. **Zhou YF, Ren GP, Savolainen O, Liu JQ** Both new variations and adaptive introgression contribute to climatic adaptation in homoploid hybrid speciated *Pinus densata* to the Qinghai-Tibetan Plateau. (in preparation).
6. **Li Y, Stocks M, Hemmilä S, Källman T, Zhu HT, Zhou YF, Chen J, Liu JQ, Lascoux M** (2010). Demographic histories of four spruce (*Picea*) species of the Qinghai-Tibetan Plateau and neighboring areas inferred from multiple nuclear loci. *Molecular Biology and Evolution* 27: 1001-1014.
7. **Jiang ZY, Peng YL, Hu XX, Zhou YF, Liu JQ** (2011). Cytoplasmic DNA variation in and genetic delimitation of *Abies nephrolepis* and *Abies holophylla* in northeastern China. *Canadian Journal of Forest Research* 41:1555-1561.
8. **Ren GP, Abbott RJ, Zhou YF, Zhang LR, Peng YL, Liu JQ** (2011). Genetic divergence, range expansion and possible homoploid hybrid speciation among pine species in Northeast China. *Heredity* 108: 552-562.
9. **Zhang LR, Peng YL, Ren GP, Zhou YF, Li ZH, Liu JQ** (2011). Population genetic diversity and species divergence of *Pinus massoniana* and *P. hwangshanensis* at two nucleotide loci. *Chinese Journal of Plant Ecology* 35: 531-538. (In Chinese).

致 谢

在此博士学位论文得以完成之际,我首先由衷地感谢导师刘建全教授。从论文的设计,样品采集,到之后学术论文的写作、修改、发表等各个环节,无不凝聚着导师耐心指导与无私帮助的汗水。刘老师为人师表,率先垂范,六年来,我无时不深深感受着老师的人格魅力,让我受益匪浅,这将成为我以后为人处世与从事学术研究的宝贵财富;刘老师学识渊博,勤奋刻苦,精益求精,对自己与学生的知识积累、学术研究都要求严格,这是我们这个时代中难能可贵的精神,是值得我一生学习的榜样!

此外,我要特别感谢赵长明老师,万东石老师,王玉金老师,邓建明老师和吴玉霞老师在学习和工作的帮助。老师们对科研孜孜以求的精神深深的感染着我,也是我今后科研生涯的榜样。吴桂丽老师在实验方面给了我很大帮助,没有吴老师的辛勤工作,我就不可能按进度完成实验数据的收集。在我加入实验室初期,陈康明师兄手把手教我做实验和处理数据,这也为我的论文奠定了基础。蒋祖耀,任广鹏和张丽锐,朱洪涛等同学在本论文实验数据收集方面做了很大贡献,一并感谢。

同时,非常感谢国家留学基金委员会(CSC)资助我为期18个月在芬兰奥卢大学的联合培养。期间合作导师Savolainen Outi教授以及课题组给予了我很大帮助。作为世界知名群体遗传学家, Savolainen教授活跃在教学和科研的第一线,她的群体遗传学和基因组学等课程,让我感受到科学之深,之美。而且每一次讨论都让我受益匪浅。

感谢田斌博士,邱强博士,田兴民博士,张茜博士,王爱兰博士,崔懿奎博士,彭晓丽博士,陈立同博士,毛康珊博士,马飞博士,徐婷婷博士,李忠虎博士,贾东瑞博士,曾军博士,杨鑫等师兄师姐在科研,学习,生活上无私的帮助。也感谢同门学友朱洪涛,刘兵兵,吴丽丽,刘瑞瑞,石大川,王娟,王倩,安家兴,张晓伟,邹嘉宾,王志强,寇一翮,弓娜、高贤良、胡晓晓、郭秋红、王婧、彭艳玲、马涛、胡泉军、张健、马会,周冉、鲁志强、张旭等同学的陪伴和帮助,这份真挚的情谊,将永远珍藏!

非常感谢我的朋友们,谢谢你们一直以来带给我的欢声笑语!尤其我在留学芬兰期间,你们分享了我孤独,谢谢!尤其感谢我的芬兰朋友, Tuomas, Jaro, Komlan, Jaakko, Tiina, Kiitos!

最后，要深深感谢我的家人，没有他们的理解与支持，就没有我完成学业的信心和动力，对于他们，我希望自己能够在以后的生活中加以报答！

周永锋 2014年10月4日于Oulu, Finland

附表 1: 三个物种各个位点的遗传多样性统计

Climate-related candidate genes																									
<i>α3p2</i>	<i>ACIP4-2</i>	<i>equus-ALP</i>	<i>ana</i>	<i>anaR</i>	<i>exonant</i>	<i>com1</i>	<i>dmi1</i>	<i>dmi2</i>	<i>dmi9</i>	<i>dmi7</i>	<i>End5</i>	<i>CJ</i>	<i>Chu</i>	<i>hnh1</i>	<i>hna3</i>	<i>LEA</i>	<i>lpa-1</i>	<i>MdhA4</i>	<i>PHYO</i>	<i>Pod</i>	<i>ppr-c</i>	<i>ppapi12</i>	<i>RD21A</i>	Average	
<i>P. ichikii</i> formis																									
<i>N</i>	125	77	146	94	98	147	119	75	103	76	99	121	123	80	141	77	84	142	84	109	85	130	77	122	105.5
<i>N_s</i>	12	33	61	27	12	29	51	26	11	53	12	12	6	21	2	14	15	31	28	6	20	23	29	30	23.88
<i>Hd</i>	0.663	0.939	0.932	0.909	0.855	0.751	0.891	0.951	0.689	0.986	0.806	0.806	0.569	0.714	0.049	0.393	0.677	0.844	0.959	0.522	0.916	0.552	0.986	0.936	0.765
<i>D</i>	-0.592	0.332	2.323	0.821	-1.042	0.674	1.284	0.462	-0.083	-0.149	-0.045	-0.571	0.182	-1.208	-2.100	-2.062	-1.696	-0.080	-0.377	-0.486	-1.231	-1.780	-0.085	-0.745	-0.344
ALL																									
<i>L</i>	1069	772	680	424	401	545	456	967	551	868	391	1014	504	997	537	731	1283	459	519	948	447	586	395	946	687.50
<i>S</i>	15	50	20	16	13	13	15	43	10	27	15	10	5	26	6	11	18	12	24	5	38	16	17	39	19.33
<i>π</i>	0.0027	0.0145	0.0097	0.0095	0.0056	0.0058	0.0091	0.0106	0.0037	0.0063	0.0094	0.0019	0.0029	0.0032	0.0006	0.0007	0.0020	0.0046	0.0094	0.0010	0.0110	0.0017	0.0106	0.0072	0.0060
<i>θ_w</i>	0.0033	0.0132	0.0053	0.0074	0.0077	0.0043	0.0062	0.0091	0.0035	0.0064	0.0095	0.0023	0.0027	0.0053	0.0029	0.0031	0.0043	0.0047	0.0107	0.0013	0.0171	0.0049	0.0103	0.0090	0.0064
Nonsynonymous																									
<i>L</i>	240.8	378.6	234.9	306.8	230.0	226.0	349.2	391.1	323.8	519.5	206.3	524.1	343.8	386.2	189.7	249.0	57.6	139.3	143.8	572.3	127.3	305.5	305.5	439.7	259.7
<i>S</i>	0	22	4	3	8	0	N.A	N.A	6	12	10	3	3	3	0	3	1	2	0	2	5	8	n.a.	10	5.000
<i>π</i>	0.0000	0.0146	0.0041	0.0024	0.0037	0.0000	0.0009	0.0137	0.0027	0.0044	0.0131	0.0015	0.0028	0.0011	0.0000	0.0006	0.0042	0.0004	0.0000	0.0002	0.0028	0.0007	0.0075	0.0031	0.0039
<i>θ_w</i>	0.0000	0.0118	0.0030	0.0019	0.0083	0.0000	N.A	N.A	0.0036	0.0047	0.0115	0.0013	0.0024	0.0016	0.0000	0.0025	0.0040	0.0026	0.0000	0.0008	0.0078	0.0048	n.a.	0.0048	0.0037
Silent																									
<i>L</i>	823.2	386.4	439.1	116.2	171.0	317.0	100.8	571.0	226.2	346.5	165.7	484.9	160.2	601.8	292.3	480.0	899.4	319.7	372.2	367.7	315.7	287.5	89.5	475.3	367.2
<i>S</i>	15	28	16	13	6	14	N.A	N.A	5	16	5	7	2	23	6	8	17	10	24	3	35	8	n.a.	30	13.86
<i>π</i>	0.0035	0.0147	0.0129	0.0284	0.0081	0.0100	0.0102	0.0086	0.0051	0.0091	0.0048	0.0023	0.0030	0.0046	0.0010	0.0008	0.0019	0.0064	0.0131	0.0023	0.0144	0.0028	0.0213	0.0110	0.0083
<i>θ_w</i>	0.0043	0.0147	0.0066	0.0219	0.0084	0.0079	N.A	N.A	0.0043	0.0094	0.0071	0.0034	0.0034	0.0077	0.0048	0.0034	0.0044	0.0057	0.0148	0.0019	0.0221	0.0051	n.a.	0.0133	0.0083

P. densata																									
atlp2	AGP4-2	aque-ATP	anz	anzR	ecosynt	cont	dln1	dln2	dln9	dln7	Erdb	Cl	Chc	MNH	Ins3	LEA	lpa-1	MdM4	PHYO	Pod	ppsc	ppsp12	RD21A	Average	
N	60	68	80	79	80	51	74	143	96	57	71	61	61	99	76	62	124	67	51	58	70	76	89	99	79.67
Nb	23	24	14	23	6	11	3	28	15	29	21	14	11	20	3	1	31	48	18	4	10	6	23	31	17.38
Hd	0.91	0.84	0.65	0.88	0.68	0.69	0.58	0.72	0.65	0.94	0.81	0.83	0.67	0.84	0.10	0.00	0.94	0.99	0.92	0.17	0.70	0.44	0.95	0.95	0.70
D	1.40	-0.17	0.93	1.14	-0.10	0.20	1.51	-0.84	-1.10	1.36	-1.22	-0.95	-1.31	-2.13	-1.65	0.00	0.70	-0.74	-0.97	-1.51	-1.26	-1.22	0.67	1.81	-0.23
All																									
L	1068	781	654	79	401	545	457	943	556	854	391	1050	504	997	537	731	1283	459	519	948	447	596	395	946	672.54
S	25	48	15	424	4	13	3	45	15	96	25	12	23	92	27	0	30	41	33	4	17	5	9	20	42.75
π	0.00722	0.01219	0.00615	0.00811	0.00340	0.00566	0.00242	0.00682	0.00318	0.03562	0.00848	0.00168	0.00569	0.00661	0.00333	0.00000	0.00833	0.01509	0.01038	0.00028	0.00546	0.00097	0.00662	0.00741	0.007
θw	0.00502	0.01283	0.00463	0.00573	0.00281	0.0053	0.00135	0.00862	0.00525	0.02438	0.01391	0.00253	0.00975	0.01882	0.01138	0	0.00674	0.01904	0.01416	0.00092	0.00889	0.00191	0.00339	0.00468	0.008
Nonsynchronous																									
L	240.77	377.98	236.8	306.83	230	226	351.02	370.28	327.69	502.04	205.79	523.94	343.15	386.11	189.62	249	57.93	139.45	143.86	572.34	127.33	305.5	305.64	439.78	298.410
S	0	22	3	2	0	1	3	N.A	6	49	13	4	N.A	N.A	2	0	2	n.a.	2	1	2	2	5	3	6.100
π	0	0.01033	0.00032	0.00265	0	0.00017	0.00315	0.00991	0.0011	0.02884	0.00872	0.00107	0.00643	0.00261	0.00107	0	0.01641	0.00592	0.00054	0.00006	0.0042	0.00146	0.00353	0.00179	0.005
θw	0	0.01215	0.00256	0.00132	0	0.00098	0.00175	N.A	0.00356	0.02104	0.01397	0.00163	N.A	N.A	0.00215	0	0.0064	n.a.	0.00309	0.00038	0.00326	0.00149	0.00387	0.00147	0.004
Silent																									
L	822.23	396.02	413.2	116.17	171	317	101.98	564.72	228.31	342.96	166.21	485.06	160.85	558.89	291.38	480	767.07	311.55	372.14	367.66	268.67	287.5	89.36	479.22	356.631
S	25	26	12	10	5	12	0	N.A	9	52	12	8	N.A	N.A	25	0	28	n.a.	32	3	16	3	4	17	14.950
π	0.00938	0.01417	0.00955	0.02261	0.00797	0.0096	0	0.00489	0.00617	0.04624	0.00818	0.00237	0.00408	0.00931	0.00815	0	0.00772	0.0192	0.01424	0.00064	0.00608	0.00045	0.01036	0.01238	0.010
θw	0.00652	0.01371	0.00596	0.01742	0.00524	0.00841	0	N.A	0.00767	0.03288	0.01494	0.00352	N.A	N.A	0.01751	0	0.00677	n.a.	0.01911	0.00176	0.01256	0.00237	0.01059	0.00764	0.010

adip2	AGPA-2	capu-ATP	ave	amR	cccount	count	dbr1	dbr2	dbrP	dbr7	End3	GI	Glu	hbl	hcs3	LEA	lpL-1	Mdb4	PHTO	Pod	ppc	ppsp12	RD21A	Average
<i>P. yunnanensis</i>																								
N	23	27	26	23	29	24	44	20	25	20	14	13	28	22	15	38	22	18	15	25	22	17	20	24.79
Mb	14	9	5	3	4	3	14	4	11	12	5	4	8	1	1	8	10	10	2	4	4	6	10	6.58
Hd	0.941	0.787	0.594	0.549	0.719	0.236	0.897	0.626	0.962	0.947	0.78	0.423	0.627	0	0	0.798	0.857	0.935	0.133	0.23	0.561	0.952	0.832	0.63
D	-0.14	-0.47	1.35	1.27	1.01	-1.20	1.04	-1.13	-0.38	0.39	0.61	-0.83	-0.58	0.00	0.00	1.79	0.10	0.03	-1.16	-1.89	-0.66	0.33	-0.23	0.04
Total																								
L	1869	772	654	424	401	545	972	556	854	391	1050	504	997	537	731	1283	459	519	948	447	596	395	946	687.79
S	25	27	9	4	3	8	32	9	27	15	3	4	12	0	0	19	11	25	1	4	4	9	19	11.33
π	0.0661	0.0691	0.0051	0.0037	0.0038	0.0063	0.0102	0.0031	0.0088	0.0100	0.0011	0.0019	0.0026	0.0000	0.0000	0.0077	0.0009	0.0141	0.0001	0.0008	0.0018	0.0099	0.0057	0.0050
θw	0.0063	0.0104	0.0036	0.0026	0.0021	0.0042	0.0076	0.0046	0.0097	0.0114	0.0009	0.0026	0.0031	0.0000	0.0000	0.0050	0.0007	0.0140	0.0003	0.0027	0.0022	0.0093	0.0061	0.0049
Neophytocycas																								
L	240.83	370.74	236.83	306.86	230	226	391.45	327.67	305.14	206.22	450.13	342.81	386.14	189.67	249	57.9	139.39	143.89	572.33	127.33	305.5	306	439.33	255.93
S	0	10	0	0	0	0	N.A	1	9	11	18	3	1	0	0	3	1	1	1	0	2	6	N.A.	3.14
π	0.0000	0.0076	0.0000	0.0000	0.0000	0.0000	0.0107	0.0003	0.0048	0.0131	0.0021	0.0020	0.0013	0.0000	0.0000	0.0141	0.0012	0.0008	0.0002	0.0000	0.0015	0.0084	0.0031	0.0030
θw	0.0000	0.0080	0.0000	0.0000	0.0000	0.0000	N.A	0.0009	0.0055	0.0150	0.0075	0.0028	0.0007	0.0000	0.0000	0.0123	0.0020	0.0020	0.0005	0.0000	0.0022	0.0080	N.A.	0.0031
Silent																								
L	823.17	394.26	413.17	116.14	171.00	317.00	575.55	228.33	342.86	165.78	347.87	161.19	607.86	293.33	480.00	845.10	311.61	372.11	367.67	268.67	287.50	89.00	440.67	355.97
S	25	17	9	4	4	8	N.A	8	18	4	35	1	11	0	0	16	10	24	0	4	2	3	N.A.	9.23
π	0.0079	0.0108	0.0081	0.0136	0.0090	0.0109	0.0099	0.0070	0.0149	0.0060	0.0049	0.0018	0.0034	0.0000	0.0000	0.0073	0.0094	0.0194	0.0009	0.0012	0.0021	0.0150	0.0083	0.0071
θw	0.0082	0.0128	0.0057	0.0093	0.0067	0.0072	N.A	0.0099	0.0162	0.0068	0.0189	0.0020	0.0047	0.0000	0.0000	0.0045	0.0018	0.0188	0.0000	0.0039	0.0023	0.0138	N.A.	0.0073

Reference loci		2009	C3H	C4H2	CesA2	GapCp	LHCA4	nir	PAL	pho	PT-1	PT-2	WD40	aver_ref
P. tabulaeformis														
N	68	80	84	89	68	75	95	115	105	90	76	70	87.7273	
N _h	18	10	6	6	23	12	20	12	20	5	4	22	12.3636	
H _d	0.911	0.888	0.67	0.642	0.792	0.665	0.958	0.6	0.74	0.519	0.487	0.901	0.7156	
Tajima's D	-0.6267	0.76061	1.62376	-0.3805	0.04574	-1.13673	0.05768	-1.50087	-1.07616	0.33232	0.28608	1.53831	-0.1468	
All														
L	461	842	783	511	546	1080	304	917	500	439	402	1104	657.4167	
S	29	15	6	4	17	12	10	14	13	3	3	56	15.1667	
π	0.01151	0.00547	0.00420	0.00156	0.00719	0.00284	0.00814	0.00127	0.00298	0.00185	0.00203	0.01575	0.0054	
θ _w	0.01418	0.00441	0.00265	0.00185	0.00708	0.0044	0.00799	0.00287	0.00498	0.00161	0.0018	0.01077	0.0054	
Nonsynonymous														
L	175.05	315	211.09	263.52	130.02	310.84	234.54	706.42	145.84	99.17	151	108.5	237.5825	
S	1	1	0	2	1	1	7	14	1	0	0	3	2.5833	
π	0.00024	0.00045	0	0.00263	0.00086	0.00009	0.00741	0.00164	0.00013	0	0	0.00156	0.0013	
θ _w	0.00129	0.00075	0	0.0018	0.00161	0.00066	0.00725	0.00373	0.00131	0	0	0.00574	0.0020	
Silent														
L	266.95	483	340.91	245.48	368.98	246.16	68.46	208.58	351.16	337.83	250	969.5	344.7508	
S	27	14	6	2	16	12	3	0	12	3	3	53	12.5833	
π	0.01903	0.00875	0.00683	0.00042	0.00946	0.00633	0.01076	0	0.00418	0.00241	0.00326	0.01735	0.0074	
θ _w	0.02279	0.00681	0.0043	0.00193	0.00905	0.00997	0.01064	0	0.00654	0.00209	0.00289	0.01135	0.0074	

<i>P. densata</i>													
	2009	C3H	C4H2	CesA2	GapCp	LHCA4	nir	PAL	pho	PT-1	PT-2	WD40	aver_ref
<i>N</i>	88	67	68	55	80	60	75	70	64	111	112	104	85.3333
<i>N_h</i>	12	33	11	6	13	14	41	8	17	8	3	38	17.0000
<i>H_d</i>	0.941	0.936	0.61300	0.597	0.504	0.551	0.98	0.44	0.863	0.651	0.621	0.924	0.7184
<i>Tajima's D</i>	-0.54361	0.55719	-1.25641	-1.84567	-2.14501	-1.45963	-0.0317	-0.8369	-0.65549	-1.25564	0.87226	-0.70829	-0.7757
All													
<i>L</i>	463	842	783	511	546	1080	304	917	500	439	402	1104	657.5833
<i>S</i>	18	20	10	13	19	14	16	5	11	19	3	97	20.4167
π	0.00980	0.00687	0.00202	0.00208	0.00194	0.00314	0.01132	0.00071	0.00357	0.00457	0.00313	0.01403	0.0053
θ_w	0.0114	0.00529	0.00377	0.00556	0.00704	0.00534	0.01077	0.00113	0.00466	0.00819	0.00247	0.01736	0.0069
Nonsynonymous													
<i>L</i>	174.96	315.02	211.45	263.55	130	310.83	234.69	706.3	145.89	99.17	151	108.48	237.6117
<i>S</i>	1	N.A.	1	7	2	0	n.a.	5	1	2	0	4	2.3000
π	0.0012	0.00028	0.00027	0.00258	0.0019	0	0.00806	0.00093	0.00183	0.00141	0	0.00157	0.0017
θ_w	0.00166	N.A.	0.00099	0.00581	0.00311	0	n.a.	0.00147	0.00145	0.00382	0	0.00707	0.0025
Silent													
<i>L</i>	281.04	475.98	340.55	245.45	413	250.17	68.31	208.7	351.11	337.83	250	961.52	348.6383
<i>S</i>	17	N.A.	9	6	17	16	n.a.	0	10	17	3	96	19.1000
π	0.01526	0.01124	0.00312	0.00157	0.00196	0.00706	0.02269	0	0.00431	0.00552	0.00503	0.01545	0.0078
θ_w	0.01759	N.A.	0.00552	0.00534	0.00831	0.01372	n.a.	0	0.00602	0.00953	0.00397	0.01914	0.0089

<i>P. yunnanensis</i>													
	2009	C3H	C4H2	CesA2	GapCp	LHCA4	nir	PAL	pho	PT-1	PT-2	WD40	aver_ref
<i>N</i>	26	23	21	21	22	23	22	20	24	48	18	48	29.6667
<i>N_h</i>	10	15	3	2	4	5	13	5	9	6	2	24	8.1667
<i>H_d</i>	0.925	0.917	0.61	0.182	0.333	0.672	0.944	0.616	0.833	0.716	0.523	0.928	0.6833
<i>Tajima's D</i>	-1.53219	-0.40102	0.61678	-1.1285	-1.23957	0.51829	0.47198	-1.26582	-0.15998	1.24872	1.94857	-0.34006	-0.1052
All													
<i>L</i>	463	842	783	511	46	1080	304	917	500	439	402	1104	615.9167
<i>S</i>	27	17	2	1	3	8	7	5	6	3	2	76	13.0833
π	0.01155	0.00548	0.00127	0.00036	0.00079	0.00447	0.00725	0.00090	0.00305	0.00246	0.00260	0.01454	0.0046
θ_w	0.01773	0.00582	0.001	0.00067	0.00151	0.00386	0.00632	0.00154	0.00322	0.00154	0.00145	0.01587	0.0050
Nonsynonymous													
<i>L</i>	175.02	315.07	211.5	263.64	130	310.83	234.52	706.26	145.88	99.17	151	108.47	237.6133
<i>S</i>	N.A.	N.A.	0	1	1	0	4	5	1	0	0	6	1.8000
π	0.00143	0.00089	0	0.00069	0.0007	0	0.00528	0.00117	0.00156	0	0	0.0045	0.0014
θ_w	N.A.	N.A.	0	0.0013	0.00211	0	0.00468	0.002	0.00184	0	0	0.01246	0.0024
Silent													
<i>L</i>	280.98	475.93	340.5	245.36	413	250.17	68.48	208.74	351.13	337.83	250	969.53	349.3042
<i>S</i>	N.A.	N.A.	2	0	2	8	3	0	5	3	2	71	9.6000
π	0.01797	0.00853	0.00207	0	0.00082	0.01005	0.0141	0	0.00368	0.0032	0.00418	0.01567	0.0067
θ_w	N.A.	N.A.	0.00163	0	0.00133	0.00866	0.01202	0	0.00381	0.002	0.00233	0.0165	0.0048

附表 2 三个物种间遗传分化参数统计

Locus	<i>P. tabulaeformis</i> & <i>P. densata</i>						<i>P. tabulaeformis</i> & <i>P. yunnanensis</i>						<i>P. densata</i> & <i>P. yunnanensis</i>					
	<i>S</i>	<i>Dxy</i>	<i>Da</i>	<i>Ka</i>	<i>Ks</i>	<i>Ka/Ks</i>	<i>S</i>	<i>Dxy</i>	<i>Da</i>	<i>Ka</i>	<i>Ks</i>	<i>Ka/Ks</i>	<i>S</i>	<i>Dxy</i>	<i>Da</i>	<i>Ka</i>	<i>Ks</i>	<i>Ka/Ks</i>
<i>α3ip2</i>	28	0.00681	0.00186	0	0.0026	0	27	0.00489	0.0005	0	0	-	31	0.00725	0.00059	0	0.00257	-
<i>ACP-4-2</i>	65	0.01719	0.00377	0.01524	0.0282	0.541385	57	0.01757	0.0056	0.01437	0.03039	0.472853	48	0.01115	0.00041	0.00908	0.01607	0.565028
<i>aquia-MIP</i>	24	0.00914	0.00125	0.0033	0.013	0.253262	19	0.00858	0.0012	0.00318	0.01272	0.25	16	0.00569	0.00006	0.00016	0.00077	0.207792
<i>ara</i>	17	0.00982	0.00101	0.00266	0.0288	0.092265	17	0.00866	0.00203	0.00139	0.02795	0.049732	13	0.0073	0.00138	0.00182	0.02184	0.083333
<i>araR</i>	13	0.00522	0.00072	0.00194	0.012	0.161532	13	0.00552	0.00081	0.00194	0.01484	0.130728	4	0.00371	0.00009	0	0.01157	0
<i>ccozoni</i>	20	0.00835	0.00261	0.00009	0.0107	0.008403	18	0.00936	0.00329	0	0.01267	0	13	0.00622	0.00023	0.00009	0.01411	0.006378
<i>comt</i>	30	0.03637	0.03063	0.03102	0.057	0.543924	29	0.03619	0.0314	0.03079	0.05708	0.539418	4	0.00337	0.0019	0.00439	0	-
<i>dhn1</i>	65	0.01076	0.00227	0.01516	0.0158	0.958887	54	0.0125	0.0022	0.01376	0.02006	0.685942	60	0.01017	0.00186	0.01199	0.00982	1.220978
<i>dhn2</i>	19	0.00358	0.00012	0.00202	0.011	0.18347	15	0.00415	0.00075	0.00202	0.01101	0.18347	16	0.00341	0.00029	0.0007	0.01359	0.051508
<i>dhn9</i>	114	0.02576	0.00496	0.0198	0.0505	0.391924	43	0.00917	0.00177	0.00537	0.02499	0.214886	103	0.02597	0.00375	0.02035	0.05111	0.398161
<i>dhn7</i>	15	0.0117	0.00276	0.01333	0.0003	49.37037	15	0.01708	0.0074	0.01927	0.00097	19.86598	25	0.01167	0.01167	0.01328	0.00123	10.79675
<i>erd3</i>	10	0.00235	0.00057	0.00176	0.0016	1.121019	10	0.00176	0.00102	0.00172	0.00131	1.312977	12	0.00146	0.00006	0.00071	0.00025	2.84
<i>Gl</i>	5	0.0057	0.00143	0.00673	0.0055	1.232601	5	0.00439	0.00199	0.00535	0.00362	1.477901	23	0.00385	0.00004	0.00431	0.00444	0.970721
<i>Glu</i>	26	0.00554	0.00057	0.00208	0.0087	0.239356	26	0.00301	0.00011	0.00125	0.00579	0.215889	92	0.00495	0.0003	0.00195	0.00708	0.275424
<i>HHL</i>	6	0.00302	0.0001	0.00056	0.0008	0.666667	6	0.0003	0	0	0	-	27	0.00277	0.00011	0.00056	0.00084	-
<i>lmo3</i>	11	0.00037	0.00001	0.00031	0.0005	0.596154	11	0.00167	0.0013	0.00031	0.00052	0.596154	0	0.00137	0.00137	0	0	-
<i>LEA</i>	12	0.00768	0.0026	0.02065	0	-	16	0.00705	0.0021	0.01318	0	-	30	0.00818	0.00033	0.01824	0	-
<i>lps-1</i>	11	0.01224	0.00241	0.00355	0.0108	0.327793	11	0.00717	0.00144	0.00085	0.0055	0.154545	41	0.01238	0.00139	0.00375	0.01411	0.265769
<i>MdhA</i>	24	0.01013	0.00022	0.00027	0.0076	0.03548	24	0.0137	0.00191	0.00039	0.01493	0.026122	33	0.01356	0.0013	0.00064	0.01743	0.036718
<i>PHYO</i>	5	0.00072	0.00007	0.00012	0	-	5	0.0007	0.00012	0.00021	0	-	4	0.00022	0	0.00015	0	-

Locus	<i>P. tabulaeformis</i> & <i>P. densata</i>							<i>P. tabulaeformis</i> & <i>P. yunnanensis</i>							<i>P. densata</i> & <i>P. yunnanensis</i>						
	<i>S</i>	<i>Dxy</i>	<i>Da</i>	<i>Ka</i>	<i>Ks</i>	<i>Ka/Ks</i>		<i>S</i>	<i>Dxy</i>	<i>Da</i>	<i>Ka</i>	<i>Ks</i>	<i>Ka/Ks</i>		<i>S</i>	<i>Dxy</i>	<i>Da</i>	<i>Ka</i>	<i>Ks</i>	<i>Ka/Ks</i>	
<i>Pod</i>	33	0.00908	0.00168	0.00476	0.003	1.608108		33	0.00631	0.00123	0.00148	0.0026	0.569231		17	0.00457	0.00144	0.00381	0.00035	10.88571	
<i>ppsc</i>	16	0.00312	0.00179	0.00448	0.0004	12.10811		16	0.00344	0.00169	0.00444	0.00015	29.6		5	0.00148	0.00009	0.0016	0.00022	7.272727	
<i>ppap12</i>	17	0.01145	0.00283	0.00777	0.0255	0.305185		17	0.01296	0.0027	0.00883	0.02869	0.307773		9	0.01632	0.00807	0.00926	0.04288	0.215951	
<i>RL21</i>	39	0.00862	0.00133	0.00287	0.021	0.136862		39	0.01063	0.00404	0.0038	0.02162	0.175763		19	0.00864	0.00196	0.00255	0.02006	0.127119	
Reference loci																					
2009	28	0.01201	0.00161	0.00075	0.0039	0.192802		28	0.01197	0.0004	0.00083	0.00644	0.128882		18	0.01089	0.00021	0.00127	0.00465	0.273118	
<i>C3H</i>	15	0.00759	0.00138	0.00038	0.016	0.02372		15	0.00751	0.002	0.00072	0.02129	0.033819		20	0.00705	0.00087	0.00059	0.0269	0.021933	
<i>C4H2</i>	6	0.00491	0.0018	0.00014	0.0176	0.007964		6	0.0046	0.00186	0	0.01855	0		10	0.00166	0.00001	0.00014	0.00404	0.034653	
<i>CesA2</i>	4	0.00183	0.00001	0.00261	0.0003	10.03846		4	0.00118	0.00022	0.00209	0	-		13	0.00134	0.00012	0.00183	0.00026	-	
<i>GapCp</i>	17	0.00547	0.00087	0.00151	0.0131	0.115004		17	0.00487	0.00093	0.0008	0.0122	0.065574		19	0.00137	0.00001	0.00132	0.00152	0.868421	
<i>LHC44</i>	12	0.00525	0.00231	0.00004	0.0009	0.042553		12	0.00578	0.00247	0.00004	0.00027	0.148148		14	0.00398	0.00017	0	0.00069	0	
<i>nir</i>	10	0.01075	0.00102	0.00836	0.0191	0.438155		10	0.00806	0.00037	0.00677	0.01261	0.536875		16	0.0099	0.00062	0.00725	0.01916	0.378392	
<i>PAL</i>	14	0.00106	0.00007	0.00137	0	-		14	0.00113	0.00005	0.00147	0	-		5	0.00079	-0.00002	0.00103	0	-	
<i>Pho</i>	13	0.03429	0.03101	0.00112	0.0025	0.451613		13	0.03399	0.03096	0.00091	0.00286	0.318182		11	0.00338	0.00008	0.00166	0.00276	0.601449	
<i>pt-1</i>	3	0.00529	0.00208	0.00073	0.0067	0.108793		3	0.00352	0.00136	0	0	-		19	0.00359	0.00007	0.00073	0.00671	-	
<i>pt-2</i>	3	0.00283	0.00025	0	0.0114	0		3	0.00247	0.00016	0	0.01073	0		3	0.00269	-0.00017	0	0.01136	0	
<i>WD40</i>	54	0.01608	0.00138	0.00159	0.0169	0.09425		56	0.02107	0.00593	0.00309	0.0274	0.112774		97	0.02027	0.00611	0.0031	0.02861	0.108354	
Average	22.33	0.009224	0.003093	0.00497	0.0118	0.422794		19.64	0.008692	0.003425	0.004184	0.011382	0.367581		24.72	0.00674	0.00130	0.00356	0.00992	0.359412	
sum	804							707							890						