

分类号
密级

UDC



学位论文

基于干旱及缺氧候选基因的 生态适应研究：以川滇高山栎为例

(中文题名)

Study on ecological adaptation based on drought and hypoxia
related candidate genes in trees: *Quercus aquifoliodes* Rehd. Et Wils
as an example

(英文题名)

王玉焄

(作者姓名)

指导教师	杜芳 副教授
申请学位级别	硕士
学科专业名称	生态学
研究方向	分子生态学
论文提交日期	2018年4月
论文答辩日期	2018年6月
学位授予日期	

答辩委员会主席：

评阅人：

北京林业大学

本论文研究工作受国家自然科学基金项目（项目编号：**41671039**；项目名称：**气候变化与物种进化历史持续性研究：以高山栎为例**）的资助。

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京林业大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签 名：_____ 日 期：_____

关于论文使用授权的说明

本人完全了解北京林业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

(保密的论文在解密后应遵守此规定)

签 名：_____ 导师签名：_____ 日 期：_____

摘要

川滇高山栎(*Quercus aquifoliodes* Rehd. Et Wils)是一种常绿硬叶栎属(*Quercus*)高山栎组(Sect. *Heterobalanus* (Oerst.) Menits)植物, 广泛分布于我国西南的高海拔地区, 对严酷的生态环境有较强适应力。本研究基于干旱与缺氧相关的候选基因, 使用景观基因组学的研究方法, 对川滇高山栎的生态适应进行了研究。我们挑选了 65 个候选基因, 使用高通量测序的方法, 从 60 个川滇高山栎种群, 587 个个体中鉴定出 381 个高质量的单核苷酸多态性(single nucleotide polymorphism, SNP)位点。首先我们使用所有 SNP 数据和中性的 SNP 数据, 并结合已经发表的 8 个 SSR 位点, 共三套遗传数据对川滇高山栎的遗传多样性和遗传结构进行比较分析。结果表明川滇高山栎在其分布区内分为两个枝系, 分别是分布于西藏东部地区的 Tibet 分枝和分布于横断山脉和四川西部山区的 HDM-WSP 分枝。基于所有 SNP 数据的枝系间分化更加明显, 这可能是对不同生态环境的适应造成的。基于中性 SNP 和 SSR 数据集的分析结果表明, Tibet 分枝的遗传多样性显著小于 HDM-WSP 分枝, 但基于所有 SNP 的分析则没有检测到两个枝系间的遗传多样性存在显著的差异。其次, 我们分别使用两种基于 F_{ST} 异常值的检测方法(BAYESCAN 和 FDIST2)和两种基于等位基因频率与环境梯度相关性的检测方法(BayEnv 和 LFMM), 共检测到来自 56 个候选基因的 310 个异常 SNP。分别检测到对 Tibet 分枝、HDM-WSP 分枝和所有种群的生态适应起到关键作用的适应性 SNP, 并找出影响两个分枝的特有和共有的基因, 这些基因可以用于比较两个分枝生态适应内在机理的异同。最后本研究使用两种线性的方法: Mantel 检验和冗余分析(redundancy analysis, RDA)以及两种非线性的方法: 梯度森林(Gradient Forests, GF)和广义相异建模(Generalized Dissimilarity Modelling, GDM)来分析川滇高山栎的遗传变异对环境梯度的响应模式。结果表明: 无论是在两个分枝还是在所有种群尺度下, 地理因素始终是影响川滇高山栎遗传分化的最主要因素。基于距离矩阵相关性的分析方法(Mantel 检验和 GDM)同时检测到一月降水量(prec01)在 HDM-WSP 分枝和所有种群中对遗传距离的影响最大, 存在显著的环境隔离(IBE, isolation by environment)模式。另外 GF 的累积重要性曲线图表明, 经度范围在东经 96-98 度之间、一月降水量在 12-14 毫米之间、六月降水量在 160-180 毫米之间时, 对遗传变异的影响最大, 而其他环境因子对遗传变异的影响比较平缓。本研究基于从候选基因中得到的 SNP 位点和已发表的 SSR 数据, 分析了川滇高山栎的遗传多样性及遗传结构, 而且对川滇高山栎

的适应性 SNP 位点进行了鉴定,同时研究了川滇高山栎对不同环境因素的响应模式,这为更加深入地研究川滇高山栎的生态适应及制定相应的保护策略提供基础理论依据,为研究其他森林物种的生态适应研究提供了参考。

关键词：候选基因；单核苷酸多态性；景观基因组学；气候变化；生态适应；川滇高山栎；壳斗科

Study on ecological adaptation based on drought and hypoxia related candidate genes in trees: *Quercus aquifolioides* Rehd. Et Wils as an example

Master Candidate: Wang Yuyao

(Specialty of Ecology)

Directed by Associate Professor Du Fang

ABSTRACT

Quercus aquifolioides Rehd. Et Wils is an evergreen oak species belonging to *Quercus* Sect. *Heterobalanus* (Oerst.) Menits. It widely distribute in south-west China with a high altitude range and adapt to harsh different environments. In this study, different methods of landscape genomic was used to study the ecological adaptation of *Q. aquifolioides* based on drought and hypoxia related candidate genes. In this study, 65 candidate genes were selected from the published database from other oak species and primers were designed, verified by Sanger sequencing in *Q. aquifoliesuoides*. Next-Generation pool Sequencing were used to genotyping all the candidate genes in 587 individuals from 60 populations of the species and 381 high-quality single nucleotide polymorphisms (SNPs) were identified. Firstly, three sets of genetic data including all SNP dataste, neutral SNP datasets and eight published SSR loci were used to conduct comparative analysis on the genetic diversity and genetic structure of *Q. aquifolioides*. Results showed that the *Q. aquifolioides* can be divided into two lineages, which are Tibet lineage distributed in the eastern of Tibet and HDM-WSP lineage distributed in Hengduan Mountains and the western Sichuan Province. The divergency of lineages based on all SNP dataset is more obvious, which may be caused by adaptation to different ecological environments. The genetic diversity of Tibet lineage was significantly smaller than HDM-WSP lineage based on analysis of neutral SNP and SSR datasets. However, the genetic diversity based on all SNPs did not detect significant differences between two lineages. Secondly, using two methoed bast on F_{ST} outlier (BAYESCAN and FDIST2) and two method based on the correlations between allele frequency and environmental gradient correlation (BayEnv and LFMM), a total of 310 non-neutral SNPs were detected from 56 candidate genes. The adaptive SNPs that play a key role in the ecological adaptation of Tibet lineage, HDM-WSP lineage and all populations were detected. And the specific and common genes that affect the two lineages were identified. These genes can be used to compare the differences and similarities of the internal mechanisms of ecological adaptation between two lineages. Finally, two linear methods, Mantel test and redundancy analysis (RDA) and two non-linear methods, Gradient Forests (GF) and Generalized Dissimilarity Modelling (GDM) were used to analyze the response pattern of genetic variation with environmental gradients of *Q.*

aquifolioides. The results showed that geographical factors, whether in two lineages or at all populations, are the most important factors that affects the genetic differentiation of *Q. aquifolioides*. The precipitation in January (prec01) was detected as an most robust factors impact on genetic distance and exist an significant isolation-by-environment (IBE) model in HDM-WSP lineage and all populations by two distance matrix correlation analysis methods (Mantel test and GDM) at the same time. In addition, the cumulative importance curve of GF indicates that allelic composition changes sharply within the extent of 96 - 98°E of longitude, 12 - 14 mm of precipitation in January and 160 - 180mm of precipitation in June, whereas changes along other environmental variables occur gradually. In this study, based on SNPs detected from candidate genes and published SSR loci we detected the genetic diversity and genetic structure of the species, investigated the adaptive SNPs and explored the response patterns to different environmental factors of *Q. aquifolioides*, all of which provides a basic theoretical basis for the ecological adaptation of *Q. aquifolioides* and might provides new insights for the landscape genomic studies for other tree species.

Key words: candidate genes; single nucleotide polymorphisms; landscape genomics; ecological adaptation; *Quercus aquifolioides*; Fagaceae

目录

摘要	I
ABSTRACT.....	III
1 引言	1
1.1 生态适应	1
1.1.1 生态适应研究概述.....	1
1.1.2 景观基因组学的概述.....	3
1.1.3 景观基因组学的研究方法.....	4
1.2 川滇高山栎的研究进展.....	8
1.3 本文的研究内容、研究目的及意义.....	10
2 对川滇高山栎的生态适应研究.....	11
2.1 实验材料与研究方法.....	11
2.1.1 采样策略	11
2.1.2 下载候选基因及设计引物.....	15
2.1.3 设计标签(barcode)序列	16
2.1.4 实验方法	19
2.1.5 处理测序结果	26
2.1.6 群体遗传学分析.....	27
2.1.7 景观遗传学分析.....	29
2.2 研究结果	32
2.2.1 测序及鉴定 SNP 结果.....	32
2.2.2 遗传结构	32
2.2.3 遗传多样性	32
2.2.4 异常 SNP 检测结果	38
2.2.5 环境与遗传变异相关性分析结果.....	43
3 讨论	56
3.1 遗传多样性与遗传结构.....	56
3.2 异常 SNP 位点	56
3.3 遗传变异应对气候变化的模式.....	56
4 结论与展望	58
4.1 结论	58
4.2 展望	58
参考文献	60
个人简介	69
导师简介	71
成果	73
致谢	75

1 引言

1.1 生态适应

1.1.1 生态适应研究概述

快速的气候变化正在影响着物种的分布范围，甚至威胁到一些物种的生存(Hughes, 2000; Parmesan, 2006; Parry *et al.*, 2007)。这种影响对树木尤为显著，因为树木的灭绝不仅仅失去一个物种，而且对整个生态系统及其多样性都会产生深远的负面影响(Hughes *et al.*, 2008; Kremer *et al.*, 2012; Sork *et al.*, 2013)。

植物在长期演化过程中有三种应对气候变化的特有模式：(1)短期内改变生活史或者通过表型可塑性产生形态变化；(2)通过远距离迁移的方式扩散至更适宜其生存的地区；(3)通过生态适应(ecological adaptation)演化，获得对环境适合度更高的新基因型，避免物种局部或大面积灭绝(Riordan *et al.*, 2016)。目前关于植物应对气候变化模式的研究主要集中在表型可塑性和物种扩散两个方面。但是对于生活周期较长、固着生长的木本植物来说，第一种和第二种模式的影响非常有限，而是更多地通过适应性演化来应对气候变化。

生态适应又叫局地适应(local adaptation)，是生物体在长期的自然选择进化过程中，随着环境生态因子的改变，生物体能够在自身的形态、结构、生理生化特性甚至是遗传物质水平上产生改变而提高适应能力(李俊清, 2017)。生态适应可以区分为趋同适应和趋异适应。所谓趋同适应，是指长期生活在相似的环境条件下，亲缘关系较远的不同物种也能够表现出相似的性状；而趋异适应是指同一物种由于长期生长在不同环境条件下，产生不同的适应方向而出现不同的性状。同一广幅种的不同种群分布在多样的环境中，选择压力作用在性状上，例如物候、耐寒性、耐旱性，可能会使其在分布范围内的不同空间有所差异。因此，不同栖息地的种群会形成不同的生态适应，也就是说，在每个地点，当地的种群的适合度(fitness)会比其他种群高(Kawecki & Ebert, 2004)。

近三十年来，越来越多的研究使用同质园(common garden)和交互移植实验(reciprocal transplant experiment)在表型上对植物的生态适应进行分析。交互移植实验是将来自不同生境的种群进行交互种植，通过分析每个研究地点上来自不同生境的种群的生长、繁殖情况来检测其对不同生境的适应情况(Briggs & Walters, 2016; Joshi *et al.*, 2001; Santamaría *et al.*, 2003)。这种生态方法也说明了自然选择在种群分化和物种形成中的作用(Hendry *et al.*, 2007)。与此同时，林业领域长期以来一直在不同环境中使用同质园实验来比较不同种源树木的生长和其他特性，使评估种群差异的遗传学基础成为可能(de Villedenreuil *et al.*, 2016; Matyas, 1996; Savolainen *et al.*, 2007; Sork *et al.*, 2013)。这些方法可以通过确定合适的种质资源来源来提高种植树木的生长和经济价值，从而改善对森林种群的管理(Grattapaglia *et al.*, 2009; Langlet, 1971)。因此同质园

和交互移植实验对理解自然选择如何塑造表型的地理差异至关重要。

生态适应是在长期自然选择过程中形成的，因此探究分子水平上的自然选择，是研究遗传水平生态适应的基础。自从《Genetics of the Evolutionary Process》一书(Dobzhansky & Dobzhansky, 1970)用不同的选择模式来诠释不同的进化驱动力以后，各种分子水平的自然选择模式及检验方法被提出。在经典的遗传学中，二倍体种群中同一基因座上的两个不同等位基因被定义为 A 和 a，其中 a 为突变型。在没有选择作用的条件下，三种基因型的适合度相等，频率相等，即 $F_{AA} = F_{Aa} = F_{aa}$ 。当受到选择的作用时，如果 $F_{AA} < F_{Aa} < F_{aa}$ ，即 a 表现为有利突变，因具有最大适合度而被直接固定，此时称为定向选择(directional selection)；如果 $F_{AA} > F_{Aa} > F_{aa}$ ，即 a 表现为有害突变，因适合度小而被从群体中清除，此时称为净化选择(purifying selection)；如果 $F_{AA} < F_{Aa} > F_{aa}$ ，即杂合子为有利变异，具有最高适合度，因此该群体在该基因座上不仅有利突变被平衡选择长期保留下来，并且还会保留两个甚至更多的等位基因数以维持最高的种群杂合度，此时称为平衡选择(balancing selection)。其中，定向选择和平衡选择统称为正向达尔文选择或正选择(positive Darwin selection)，起维持有利变异的作用；净化选择又被称为负选择(negative selection)，起清除有害变异的作用。

事实上，自然选择在基因组上表现出来的遗传效应涵盖了基因重组(包括基因交换与基因转换)与多基因位点连锁的复杂现象，也就是说在所研究的目标位点附近的基因也有可能受到选择的牵连。因此，针对自然选择中所考察的两种突变—有利突变和有害突变，分别有“搭车效应”(hitchhiking effect)或者“选择性清除”(selective sweep)，和“背景选择”(background selection)来对应描述这种复杂的遗传现象。“搭车效应”(Smith & Haigh, 1974)指的是在有利突变产生后被正选择固定的过程中，与之连锁的中性位点的变异也被固定；因此，当完全的“选择性清除”发生时，在没有重组的情况下，群体里所有个体都只享有该有利突变及其连锁位点的变异所形成的这一种单倍型，而原有的其他单倍型都被清除掉(Kaplan *et al.*, 1989)。因此，“搭车效应”或“选择性清除”会减低受选择位点及其连锁位点的变异程度。“搭车效应”的强度取决于选择压力的强度及该区域的重组率(Kaplan *et al.*, 1989)。在低重组区，连锁较紧密，因而“搭车效应”就越强；在重组密集区，这种连锁关系可能被打断，使得“搭乘”被弱化。类似地，“背景选择”(Charlesworth *et al.*, 1993)指的是负选择在清除有害突变时，也会随之清除与其连锁的中性位点的变异，因此，该群体在该研究区域的变异程度也会降低。

为了检测最近发生的自然选择，一种复合的位点频谱(SFS, site frequency spectrum)测试，DHEW (基于 Tajima's D , Fu & Wu's H 和 Ewens-Watterson 单倍型检测的复合中性检测)已经被证明对种群历史动态(demographic)和种群结构不那么敏感(Zeng *et al.*, 2007)。Nielsen 等(Nielsen *et al.*, 2005)引入复合似然比(CLR, composite likelihood ratio)来模拟等位基因频率频谱来检测选择信号。基于遗传分化(F_{ST})的方法已经广泛地应用

于检测异常位点作为指示选择作用的信号(Akey *et al.*, 2002; Beaumont, 2005; Beaumont, 1996; Excoffier *et al.*, 2009; Foll & Gaggiotti, 2008; Lewontin & Krakauer, 1973)。一些其他的方法, 包括基于连锁不平衡(LD, linkage disequilibrium)和基于扩展的单倍体杂合度(EHH, extended haplotype heterozygosity)用来检测最近发生的选择性清除(Sabeti *et al.*, 2007; Tang *et al.*, 2007)。遗传数据和环境变量之间的关联分析已经有效地应用于检测候选基因是否受到选择(Coop *et al.*, 2010; Grivet *et al.*, 2010; Parisod & Christin, 2008)。基于建模的方法, 例如近似贝叶斯计算(ABC, Approximate Bayesian Computation)使用中性位点建立种群历史动态模型, 通过检测候选基因是否适合模型中的特殊概括统计量的后验分布来检测其是否受到选择(Keller *et al.*, 2011)。使用 HKA (Hudson–Kreitman–Aguadé)检测(Hudson *et al.*, 1987)、MK (McDonald–Kreitman)检测(McDonald & Kreitman, 1991)以及由它们发展而来的最大似然 HKA (MLHKA)检测(MLHKA, Wright & Charlesworth, 2004)、MK 泊松随机场(MKPRF, McDonald–Kreitman Poisson Random Field) 检测(Bustamante *et al.*, 2002)可以比较种内多态性和种间分化, 从而检测历史的选择信号。

1.1.2 景观基因组学的概述

景观遗传学是一种比较新颖的对适应性遗传变异进行研究的工具(Holderegger *et al.*, 2006; Manel *et al.*, 2003)。中性遗传标记可以被用来鉴定中性遗传变异的地理模式, 也可以与环境梯度相关, 作为检测自然选择的初步证据(Endler, 1986; Westfall & Conkle, 1992)。尽管它们可以帮助确定哪些区域可能经历类似的选择压力, 但是它们经常受到种群历史动态的影响。对大量扩增片段长度多态性位点(AFLP, amplified fragment length polymorphism)进行统计分析引发了从大量中性位点中鉴定异常位点(outlier loci)的趋势(Beaumont & Nichols, 1996; Bonin *et al.*, 2005; Joost *et al.*, 2007), 特别是当结合这些异常位点和环境梯度的相关性进行分析时(Balding, 2006)。另一方面, 第二代测序技术(NGS, Next-Generation Sequencing)的迅速发展促进了鉴定适应性遗传变异的基因组工具的开发, 并促进从功能基因中获得更好的局地适应证据(Coop *et al.*, 2010; Hancock & Di Rienzo, 2008; Stinchcombe & Hoekstra, 2008; Storz, 2005)。

基因组工具的涌现以及测序成本的降低引发了一波新的研究浪潮。这些研究可以绘制适应性遗传变异的空間模式, 同时也能识别受到局地适应特征影响的基因(Anderson *et al.*, 2011; Franks & Hoffmann, 2012; Savolainen *et al.*, 2013; Stapley *et al.*, 2010)。基因组信息可以为使用同质园和交互移植实验的研究增加价值, 而不是取代它们来研究表型的遗传基础。基因组信息通过提供新颖的数据类型来加深我们对以下内容的理解: 适应性进化、对响应全球变化模式的塑造、最有效地管理受到威胁或濒临灭绝的种群或物种的保护策略。此外, 通过在自然种群中沿着环境梯度对个体进行采样, 基因组工具创造了研究 DNA 序列的景观模式的机会, 绘制了自然选择的基因组特征的空间图谱, 并具有了更高的清晰度和分辨率。

此时，景观基因组学被提出，通过对全基因组的遗传变异对生物的空间模型进行明确的研究(Holderegger *et al.*, 2006; Manel *et al.*, 2010; Sork & Waits, 2010)。传统的景观遗传学有选择性地使用中性的遗传标记，关注景观特征对动物运动的影响，植物通过种子和花粉的基因运动，以及遗传变异的地理模式(Cushman & Landguth, 2010; Holderegger & Wagner, 2008; Manel *et al.*, 2003; Segelbacher *et al.*, 2010)。然而，全基因组序列的可用性——即使是非模式系统——为景观的“基因组”方法创造了机会。它可以同时检测种群历史动态、迁移和选择的作用。要做到这一点，需要一个非常重要的步骤，即确定与涉及局地适应的表型相关的基因组区域。景观基因组学的强大之处在于它能够收集大量的样本，将表型、基因型和局地环境信息整合在一起(Sork *et al.*, 2013)。

景观基因组学可以利用任何类型的基因组标记，如 AFLP 和 SNP (single-nucleotide polymorphisms, 单核苷酸多态性)。越来越多的研究发现，SNP 更适合用于景观基因组学的分析。通过分析采样个体间的大量 SNP，结合基因型、表型和环境数据，可以鉴定种群对环境的响应模式。全基因组关联研究(GWAS, Genome-Wide Association Studies)可以通过将 SNP 与特定性状联系起来来识别适应性变异(Hancock & Di Rienzo, 2008)。这种方法对于模式物种(例如拟南芥)尤其有效。在这些物种中，性状已经被映射到基因组上(Atwell *et al.*, 2010; Bergelson & Roux, 2010; Fournier-Level *et al.*, 2011)。而对于非模式物种，如大多数树种，需要利用环境梯度与基因型之间的联系来识别潜在的受选择位点。目前，由于使用限制性内切酶得到的简化的 DNA 文库进行测序的方法的出现，使用高分辨率全基因组扫描(genome-wide scans)来识别非模式物种受选择的位点是可行的(Davey *et al.*, 2011; Elshire *et al.*, 2011; Miller *et al.*, 2007)。这些方法可以在整个基因组中以相对较低的成本获得成千上万的 SNPs。虽然 GWAS 和简化测序的方法可能只是得到一小部分变异，却可以识别多个区域来揭示大部分变异(Eckert & Dyer, 2012; Parchman *et al.*, 2012)。此外，还可以结合对候选基因的靶向测序(Gnrirke *et al.*, 2009)来提高检测功能基因中变异位点的等位基因频率是否与环境相关的效率。简而言之，通过使用大量的样本树木和大量的全基因组 SNP，景观基因组对遗传变异的分析可以为自然种群如何受到自然选择的影响提供最优的证据(Sork *et al.*, 2013)。

1.1.3 景观基因组学的研究方法

在自然种群中，选择作用的基因组特征可以通过对景观基因组进行统计检验来揭示，这些检验关注种群间的不同选择作用或遗传和环境梯度之间的联系(Sork *et al.*, 2013; Tiffin & Ross-Ibarra, 2014)。许多景观基因组学方法可以作为检测选择作用的证据。

使用生态基因组数据(典型地由成千上万个 SNP 组成)对适应性遗传多样性的空间预测的方法有两个挑战。第一个挑战是发现由于局地适应而产生变异的候选 SNP，同时避免出现由进化历史导致的全基因组变异(例如种群大小改变，基因流，范围扩

张)造成的假阳性。对于这一挑战,有两种被广泛认同的鉴定 SNP 的方法:(1)种群间极端的等位基因频率分化(F_{ST})(Excoffier *et al.*, 2009; Foll & Gaggiotti, 2008; Lewontin & Krakauer, 1973); (2)等位基因频率和环境梯度之间是否存在强烈的联系(Coop *et al.*, 2010; Frichot *et al.*, 2013; Joost *et al.*, 2007; Manel *et al.*, 2012)。这两类方法目前均处于快速的理论发展阶段,尤其强调改进模型的灵敏度同时试图将假阳性尽可能降低(Narum & Hess, 2011; Villemereuil *et al.*, 2014)。第二个挑战涉及到将一组由以上方法鉴定到的候选位点转换为景观水平的对适应性变异的预测。相对于从基因组中鉴定局地适应性 SNP 的方法的快速发展,将基因组数据转化为局地适应空间推断的工具仍处于初级发展阶段。

F_{ST} 异常值检测是一种非常流行的方法,它使用基因组数据,通过检测比预期值更高或更低的 F_{ST} 来识别经历空间分化选择的基因。 F_{ST} 是一种对种群间等位基因频率变异进行标准化度量的工具(Wright, 1949)。 F_{ST} 异常值检测的基础是识别异常高的 F_{ST} (歧化选择)或异常低的 F_{ST} (平衡选择)。然而,非选择性的进化力量也会影响基因在基因组上的分布,从而影响 F_{ST} 值的分布。基因流会降低种群间的差异,而遗传漂变会增加种群间的差异。同一基因组中大多数基因都受到这些中性过程带来的相对均等的影响,因此大多数中性基因的 F_{ST} 大致相同。然而,在被细分的种群中,检测到的 F_{ST} 可能与预期不同,导致即使是中性的基因,它们的 F_{ST} 也会有很大的差异。 F_{ST} 异常值检测试图解释这种 F_{ST} 的中性变异,并确定哪个变异位点的 F_{ST} 足够大或足够小来揭示选择作用的重要证据。异常值检测的挑战是,在没有选择作用的情况下,什么样的 F_{ST} 值是我们期望的(Lotterhos & Whitlock, 2014)。

大量的 F_{ST} 异常值检测方法已经被开发,并广泛地应用于候选基因的发现(Beaumont & Balding, 2004; Beaumont & Nichols, 1996; Excoffier *et al.*, 2009; Foll & Gaggiotti, 2008; Vitalis *et al.*, 2001)。这些方法分为两类,一类以 FDIST2 为代表,模拟一个特定的种群历史动态作为一个零假设分布模型来检测显著性(Beaumont & Nichols, 1996),另一类是以 BAYESFST (Beaumont & Balding, 2004)和 BAYESCAN (Foll & Gaggiotti, 2008)为代表,假设样本从一个共同的祖先分化出来。前一种方法的一个问题是,得到的结果可能对特定的种群历史动态非常敏感,但是真正的历史动态不被我们所知。例如我们在使用 FDIST2 时假设一个岛屿模型(island model),然而这在现实的种群中经常会得到假阳性结果,因为此模型比现实的种群动态模型简单得多(Lotterhos & Whitlock, 2014)。作为替代,一些贝叶斯方法得到发展,如 BAYESFST 和 BAYESCAN。贝叶斯方法假设在任何中性结构种群模型下的基因频率可以近似于一个多项式狄利克雷分布(Beaumont, 2005)。狄利克雷分布描述在种群历史的大范围下的基因频率(Beaumont, 2005; Charlesworth, 2010)。

另一种鉴定受选择 SNP 的方法基于选择压力下受选择的遗传位点会和环境梯度显著相关(Coop *et al.*, 2010; Frichot *et al.*, 2013; Joost *et al.*, 2007)。基因型-环境关联的

方法鉴定到基因位点如果与一个或多个环境变量存在强烈的相关关系,这些位点就被解释为受到潜在的选择作用。

另外一种方法考虑由种群历史动态和空间影响产生的种群间等位基因频率协方差,并在 BayEnv 中得以实现(Coop *et al.*, 2010)。此方法使用贝叶斯马尔可夫链蒙特卡罗分析方法,包括两个步骤。首先使用所有遗传位点(或者被确定为中性的一部分位点)利用一个各种群等位基因频率的协方差矩阵来预测种群结构。第二步,使用一个结合了经验主义协方差矩阵的模型来检测每个基因位点(或感兴趣的位点)的等位基因频率与每个环境变异之间的相关性。该方法返回贝叶斯因子(Bayes Factor)来权衡支持此相关性证据的强度。

潜在因素混合模型(LFMM, Latent factor mixed models)是非常灵活和通用的模型,它提供了另外一种检测等位基因频率和环境变量之间的方法,同时考虑种群结构(Frichot *et al.*, 2013)。该模型可以被看做是一个结合回归的近似主成分分析。其计算速度比 BayEnv 更快。使用 Tracey-Widom 检测,该模型可以为每个情境设置的 K 值(潜在因素的数量)最大为 15。此方法使用显著性检验作为衡量每个位点的等位基因频率与每个环境变量之间相关性的强弱。

将多元环境因子与空间遗传结构集合在一起,模拟遗传变异的分布模型,目前处于早期发展阶段,但在未来几年可能会越来越多地被使用,尤其是试图预测遗传变异在未来气候变化下的响应(Fitzpatrick & Keller, 2015)。

到目前为止,将物种水平的适应性变化纳入空间生物地理模型的努力主要集中在物种分布模型上(SDM, species distribution models; Elith & Leathwick, 2009)。许多研究使用 SDM 用于探索亚种、进化枝、基因型或遗传分组的地理分布预测(Banta *et al.*, 2012; Espíndola *et al.*, 2012; Jay *et al.*, 2012; Pearman *et al.*, 2010)。虽然 SMD 可以预测种内变异的分布,但是在传统的 SDM 框架下,对遗传多样性的空间模拟仍然存在很大的挑战。SMD 不仅不能解释种群内跨空间的基因组变异连续的多维的性质,而且识别和单独模拟大量的基因位点对于计算是一个严峻的挑战。例如,全基因组测序或通过产生的成千上万 SNP 位点进行基因分型(Elshire *et al.*, 2011; Narum *et al.*, 2013)。虽然从理论上讲,可以使用 SDM 对每个 SNP 的分布单独建模,但是在实践中,大量的位点数量使得分析和解释每一个模型变得不切实际。除了计算负担外,每个位点需要重复许多次计算来确保可靠的统计推断(Wisz *et al.*, 2008)。

为了解决以上问题,在群落水平将所有物种整合在一起同时被一组共同的环境因子建模的方法得到重视。类似的方法包括约束排序技术(constrained ordination techniques)如典范对应分析(CCA, canonical correspondence analysis; Ter Braak, 1986)和冗余分析(RDA, redundancy analysis; Legendre & Legendre, 2012)作为线性的工具,已经被常用于种群遗传学中来分析几十到数千个 SNP 在一组环境变量下的组成变异(Lasky *et al.*, 2012; Sork *et al.*, 2010; Sork & Waits, 2010)。一种对分布在美国加州的橡

树的 CCA 分析表明, 环境梯度与多元基因数据之间存在相关关系, 甚至控制了空间自相关之后依然得到相同的结果, 从而得到选择作用影响种群遗传结构的证据, 并且鉴定到相关性最强的气候变量(Sork & Waits, 2010)。如果在模型设计中包含空间自相关, CCA 假设群落和环境变量之间不存在因果关系, 可以控制种群动态历史的影响。RDA 和偏 RDA (*p*RDA, partial RDA)作为一种类似于多元回归的分析, 是一种可以替代 CCA 并且更有力方法, 可以在分析一部分空间影响的时候控制其他因素的影响(Boutet *et al.*, 2009)。然而这类技术很少用于预测群落未来的分布。

第二类基于群落层次的多元统计模型是非线性的, 包括广义相异建模(GDM, Generalized Dissimilarity Modelling; Ferrier *et al.*, 2007)和梯度森林(GF, Gradient Forests; Ellis *et al.*, 2012)。这两个新方法同样可以应用于大规模基因组数据集, 独立地或者与基因组扫描方法结合(基于 F_{ST} 异常值或基因与环境的相关性), 从而获得对局部适应遗传变异的分布的推断。当使用遗传数据, GDM 和 GF 相对于其他方法可以: (1) 探索基因和环境之间的非线性相关关系; (2) 处理大型基因组数据集; (3) 预测适应性遗传变异的分布。GDM 和 GF 的一个特别强大的特点是, 模型可以被映射到未来气候的场景中, 从而预测气候变化对生物多样性的潜在影响, 以及这些影响在空间上的变化。更重要的是基因组数据的群落水平建模提供了一个可行的解决方案, 以代替 SMD (它假设一个物种内的所有种群都对环境梯度做出相同的反应, 并在整个基因组水平上对适应性遗传变异进行预测, 以应对气候变化)(Fitzpatrick & Keller, 2015)。

GDM 是一个排列矩阵回归的非线性延伸, 模拟两个采样点之间成对的生物相异性与环境和地理变异的相异性的非线性函数。GDM 为每一个预测因子提供了一种独特的单调 I-样条(I-spline)转换函数, 在基因组中描述了遗传距离随梯度变化的速度和幅度, 同时控制其他变量, 并使用百分比偏差解释模型的拟合程度(Ferrier *et al.*, 2007; Fitzpatrick *et al.*, 2013)。这种基于距离的方法具有相当大的独特优势, 它能够解释由种群动态引起的遗传变异的模式, 如距离隔离、生态阻力隔离、生态抵抗或奠基者效应。其次, 由于响应变量是成对的遗传距离矩阵, 所以理论上讲, GDM 可以容纳任何数量和以任何频率出现的 SNP。此时的计算限制由样品位点的数量决定。

GF 是一个非参数机器学习回归树方法——随机森林(random forests, Breiman, 2001)——的群落水平的延伸。相比于 GDM 间接地使用采样点之间的相异性来模拟遗传位点的空间变异, GF 直接模拟遗传变异的组成变化。GF 不是基于距离、曲线拟合的方法, 相反, GF 使用机器学习算法将生物数据分成不同的箱子(例如等位基因频率的不同范围), 沿着每个环境变量会划分多个范围。沿着环境梯度移动, 尽管环境梯度在改变, 但是我们可能会观察到基因频率几乎不变, 而在其他部位, 等位基因变化明显。GF 可以处理预测因子之间复杂的关系和相互作用, 但是这些预测因子不包括地理距离。GF 也提供了一种方法来检查单个 SNP 对环境梯度的响应, 但是随着

SNP的数目变得非常大，它的计算能力也会受到限制。

1.2 川滇高山栎的研究进展

壳斗科(Fagaceae) 是北半球亚热带森林和温带森林的重要成分之一，栎属(*Quercus*) 约 450 种，是壳斗科中最大的属(Willis, 1973)。我国栎类资源丰富，包含 120 多种，栎类林的面积占中国天然林总面积的近 14%。从栎属系统树的分析中能够得出高山栎组是栎亚属中最原始的类群，同时，它是分布最广泛的常绿硬叶栎属植物(周浙昆, 1992)。国内学者对国产栎亚属展开了比较系统的研究(郑万钧, 1983)，将其分类为四个组：高山栎组、巴东栎组、麻栎组和槲栎组。周浙昆根据叶结构和叶表皮特征对上述分类系统进行了部分调整，将槲子栎单独划分出来成立槲子栎组，最终将国内的栎属植物分成五个组(Zhou *et al.*, 1995)。但是高山栎组植物的地位不存在争议。高山栎组植物发源于中新世(Miocene)以前的青藏高原的东南部(李俊清, 1996)，是一类孑遗物种。在中新世之前，这类植物在群落处于弱势地位，混生于常绿阔叶林中。中新世之后，随着青藏高原和喜马拉雅山脉的不断隆升，古地中海退却，气候变得寒冷、干燥。该地区的常绿阔叶林不能适应气候的变化而衰退。高山栎组植物由于自身拥有防寒、耐旱等生理特性(被厚毛、硬叶和双层上表皮等性状)，很快便适应了这种气候的变化，因而成为这个地区生存竞争的胜利者(Zhou *et al.*, 1995)。也就是说，该类植物具有广泛的适应性和合适的生长对策，因此能够在喜马拉雅山的抬升过程中，得以保存下来，并适应青藏高原复杂的生态环境，形成广泛的分布区(李俊清, 1996)。

但是丰富的栎树资源目前并未得到充分利用，其中一个重要原因是非模式植物较低的关注度和遗传资源的匮乏。近年来，栎类被农民作为培养香菇的锻木和薪炭林被大量砍伐，且栎材不受重视，导致栎类天然林较少，尤其是大径材很少(端木, 1994)。栎类的遗传资源正在减少，危及资源的可持续利用。因此对栎类植物尤其是高山栎进行生态适应及分子机理的研究，为栎类植物遗传多样性的保护及长期合理利用栎属植物遗传资源提供重要的遗传学理论依据。

喜马拉雅山东部和邻近的中国西南部的山脉是世界上主要的生物多样性热点地区之一，约有 12,000 种植物，其中包括 3500 种地方性的物种(Myers *et al.*, 2000)。地质学和气候环境的独特的结合也使得该地区的物种在经过长期的演变和持续之后，形成较高的物种多样性水平。该地区地形多样，有着世界上最崎岖的山脉，海拔高度梯度超过 5000 米。这种地形是由发生在新近纪(Neogene)或更早时期的异常强烈的造山运动造成的(Lippert *et al.*, 2014; Renner, 2016)。喜马拉雅山东部和相邻山脉构成了延伸的青藏高原(QTP, Qinghai-Tibet Plateau)。它们的主要隆起发生在距今 1000 万年前(Mulch & Chamberlain, 2006)，在上新世晚期达到了顶峰的高度(Sun *et al.*, 2011)。

川滇高山栎(*Quercus aquifolioides* Rehd. et Wils)主要分布于我国的喜马拉雅-横断山地区，为壳斗科栎属高山栎组(Sect. *Heterobalanus* (Oerst.) Menits)植物，是一种重要

的生态和经济型树种(周浙昆, 1992)。其范围跨越西藏的东部, 四川省的西部和云南省的北部, 一直延伸到贵州, 缅甸和不丹的邻近地区(Du *et al.*, 2017)。该区域与被定义为中国-喜马拉雅森林亚区(Sino-Himalayan forest subkingdom; Akey *et al.*, 2002; Qiu *et al.*, 2011)的植物区高度一致, 因此川滇高山栎可以看做是该区域的代表树种。川滇高山栎在海拔 1900 米至 4300 米均有分布, 对不同的环境有较强的适应力, 是该区域的优势种和建群种。

川滇高山栎为常绿乔灌木, 乔木可高达 30 m, 胸径 1m 以上, 在干旱阳坡和经常樵采的地方常形成灌木丛林(王国严和徐阿生, 2008)。雌雄同株。叶片椭圆形或倒卵形, 长 2.5-7cm, 全缘或有刺状锯齿, 背面被褐色鳞秕及星状毛, 侧脉 6-8 对。壳斗碗形, 包坚果 1/2 以下, 坚果卵形或长卵形, 直径通常为 1-1.5 cm, 部分可达到 2cm (陈焕镛和黄成就, 1998)。该种根系发达, 萌生能力强, 适应范围广, 具有良好的防寒保水的特性(杨玉坡等, 1992)。

针对川滇高山栎, 国内外学者做了很多相关的研究。之前对川滇高山栎的研究主要集中在生长特点与生物量的研究方面, 而针对遗传和分子生物学的研究还比较少, 因此, 近几年, 对川滇高山栎的分子生物学和生态学领域开展了比较系统的量化研究。李进和陈可咏(1997; 1998)利用同工酶标记研究了西藏川滇高山栎的遗传结构和不同海拔高度的群体遗传多样性变化, 发现随海拔的升高, 其遗传变异水平呈低-高-低的分布。高海拔的干寒气候对川滇高山栎的生长极为不利, 并由于生理或生态上的原因如花期延迟、地形阻隔、植被结构等, 缺少与低海拔群体的基因交流, 导致了高海拔川滇高山栎群体遗传结构的简单和单调; 但低海拔过度的湿热环境也同样不利于川滇高山栎与其它树种的生存竞争。Toumi 和 Lumaret (2001)利用同工酶标记对地中海地区四种常绿栎属树种的研究发现, 由于川滇高山栎分布范围较广, 生境梯度较大, 从而拥有更高的等位基因丰富度以适应不同的环境条件。Zhang 等(2007)利用六对微卫星引物对五个不同海拔川滇高山栎群体的遗传多样性进行研究, 结果表明由海拔梯度形成的选择性压力对川滇高山栎群体的遗传变异影响并不明显。Wang 等(2009)对中国西南地区林分更新管理和林分年龄对川滇高山栎遗传结构的影响的研究, 发现川滇高山栎的遗传多样性在老龄林分中明显高于其在幼龄林分中的水平。Du 等(2017)使用叶绿体 DNA 和核 SSR 两种分子标记, 对川滇高山栎分布区内 58 个自然种群进行研究, 揭示了新近纪和第四纪(Quaternary)气候动荡对川滇高山栎的进化动态和目前分布模式的影响。Meng 等(2017)对高山栎组植物的谱系地理学研究提出了利用高山栎对高寒环境的适应来揭示高山栎地理分布格局形成的成因, 即在青藏高原的抬升过程中, 高山栎逐渐适应了该地区逐渐抬升以后逐步变冷的环境, 从伴生种逐渐演变为优势种的过程, 并利用高山栎遗传谱系的分歧时间对喜马拉雅-横断山的抬升进行界定, 从生物学的视角为该地区的地质演变提供了证据。而最近对同组近缘种刺叶栎(*Quercus spinosa*)的研究将其在分布区内划分为两个进化枝, 该分化来自新近纪气候

变化和造山运动导致的地理隔离。另外，该研究对刺叶栎的冰期避难所进行预测(Feng *et al.*, 2017)。

1.3 本文的研究内容、研究目的及意义

本研究使用 65 个与能够在高山栎组植物中成功扩增的胁迫相关的细胞核候选基因，在川滇高山栎的 60 个种群，587 个个体中进行扩增，使用一种双端加 barcode 的混合 PCR 产物进行建库的方法(pool-sequencing)，通过高通量测序进行 SNP 鉴定(见下文，王玉垚等, 2017)，进而使用景观基因组学的方法比较川滇高山栎种内、种群间和种群内的遗传多样性和遗传结构，了解其核基因水平的遗传变异和分化式样，尤其是气候因子如何影响其遗传结构和物种的命运。相对于大多数研究使用全基因组扫描的方法来发现与气候梯度显著相关的 SNP，由此来推定受选择的基因区域(Bashalkhanov *et al.*, 2013; Eckert *et al.*, 2010)，基于候选基因的方法将注意力集中在被认为参与生态适应的基因序列上，可以为生态适应的研究提供更直接有效的证据(Alberto *et al.*, 2013; González-Martínez, *et al.*, 2006)。而树木由于生长周期长、多为异交且基因组非常大的特点，使得全基因组扫描的方法受到限制(Neale & Savolainen, 2004)。而随着高通量测序方法的越来越成熟，基于候选基因的景观基因组学方法，越来越多地应用在研究林木物种的生态适应上(e.g. Sork *et al.*, 2016; Rellstab, *et al.*, 2016; Roschanski, *et al.*, 2016)。

本研究的目的是(1)鉴定川滇高山栎对环境和气候适应性进化的关键性基因；(2)利用传统一代与高通量二代结合的混合测序建立川滇高山栎的遗传资源库；(3)整合利用遗传和环境因素来研究栎类植物应对气候变化的模式。

本研究的实施具有以下几个可行性：首先，选择的功能基因为双亲遗传，完善地保留了该物种的遗传信息；另外，栎属中已有成功测序完转录组(Durand *et al.*, 2010)和基因组(Plomion *et al.*, 2016)，很容易获得大量与环境因子相关的核基因分子标记。本项目的实施比经典的谱系地理、景观遗传学研究更进一步，从景观基因组水平探讨物种应对气候变化的微观过程。

2 对川滇高山栎的生态适应研究

2.1 实验材料与研究方法

2.1.1 采样策略

通过查阅中国数字标本馆 (<http://www.cvh.ac.cn/>)、中国植物图像库 (<http://www.plantphoto.cn/>)等网络数据库和地方植物志、文献等记载的川滇高山栎的分布,于2013年至2015年对中国西南川滇高山栎的野生分布区进行了实地考察以及样品采集。在川滇高山栎的野生分布区内,我们选择距离大于30公里的采样点作为一个种群。每个种群内随机选择5-15个样本个体(保证个体间隔大于10米),采集其健康、新鲜的叶片样本,擦拭干净后放入样品袋,统一编号后入加硅胶干燥剂对样品进行干燥保存。

经过多次野外调查及采样,本研究选择60个川滇高山栎的野生种群,分布于四川省、云南省、西藏自治区等整个分布区。使用GPS定位系统对每个种群进行定位。每个种群中选择7-11个个体,一共587个个体用于本研究。将采集到的叶片样本带回实验室,使用硅胶干燥剂将其完全干燥,并置于通风处保存,以备后续DNA提取等实验。图2-1和表2-1展示了采样种群的分布地点、代号、样本量、地理坐标等信息。图2-1中的灰色阴影代表川滇高山栎的分布范围。图2-2展示了后续实验分析的具体流程(王玉焄等,2017)。

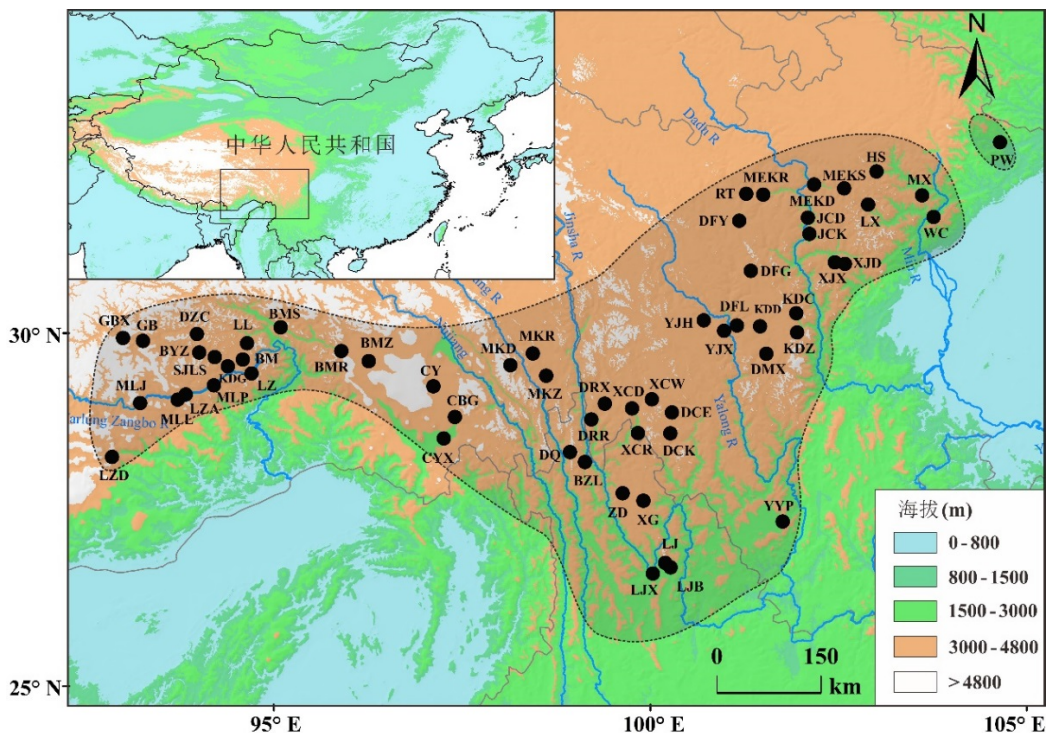


图 2-1 川滇高山栎采样点分布图

Fig. 2-1 Sample distribution of *Q. aquifolioides*

表 2-1 川滇高山栎样品采集信息表

Table 2-1 Sample information of *Quercus aquifolioides*

编号	分布地点	经度 (N)	纬度 (E)	样本量	等温性(*100) bio03	干季平均气温(°C) bio09	一月降水(mm) prec01	六月降水(mm) prec06
LZD	西藏隆子县斗玉乡	93.02	28.37	10	41.09	0.77	3.00	97.00
GBX	西藏工布达江县江达乡	93.16	29.92	10	41.61	-2.47	2.00	93.00
MLJ	西藏米林县甲格村	93.38	29.07	11	42.67	0.75	1.00	87.00
GB	西藏工布达江县	93.42	29.88	10	41.34	-1.35	2.00	98.00
MLL	西藏米林县里龙乡	93.87	29.12	10	43.42	2.37	2.00	112.00
SJLS	西藏色季拉山	93.98	29.18	7	43.38	1.63	2.00	115.00
DZC	西藏大昭村	94.12	29.97	10	41.40	-5.22	3.00	113.00
KDG	西藏卡定沟	94.15	29.73	10	41.48	-1.18	2.00	110.00
LZA	西藏米林县林芝机场	94.34	29.30	10	43.15	2.17	2.00	127.00
BYZ	西藏八一镇	94.35	29.67	10	43.20	1.33	2.00	123.00
LZ	西藏林芝县	94.52	29.55	10	39.80	-3.42	3.00	115.00
BM	西藏波密县	94.72	29.64	10	38.56	-8.28	4.00	112.00
LL	西藏鲁朗县	94.77	29.85	10	38.28	-3.95	2.00	126.00
MLP	西藏米林县派镇	94.83	29.46	10	38.60	-1.22	2.00	136.00
BMS	西藏波密县索通村	95.21	30.06	10	39.19	3.15	6.00	170.00
BMR	西藏波密然乌镇	95.99	29.75	10	38.75	-1.23	5.00	149.00
BMZ	西藏波密中坝村	96.35	29.62	10	37.73	-2.62	5.00	136.00
CY	西藏察隅县古玉乡	97.19	29.29	10	38.83	-4.65	8.00	122.00
CYX	西藏察隅县雄久村	97.32	28.61	10	41.21	5.42	12.00	256.00
CBG	西藏察隅县慈巴沟	97.46	28.89	10	39.99	1.95	9.00	174.00
MKD	西藏芒康县登巴村	98.18	29.57	10	40.77	-6.05	4.00	95.00

MKR	西藏芒康县如美镇	98.47	29.72	10	43.74	-4.90	3.00	91.00
MKZ	西藏芒康县孜许村	98.65	29.43	10	41.53	-3.30	4.00	89.00
DQ	云南德钦县	98.96	28.44	10	40.41	-2.35	8.00	68.00
BZL	云南奔子栏乡	99.15	28.31	10	37.52	-1.83	8.00	77.00
DRR	四川得荣县日龙乡	99.23	28.86	10	39.46	-3.15	5.00	84.00
DRX	四川得荣县-乡城县	99.41	29.06	10	43.64	0.02	4.00	94.00
ZD	云南中甸	99.64	27.90	10	36.96	-0.47	11.00	85.00
XCD	四川乡城县-得荣县	99.76	29.00	10	39.36	-1.90	3.00	100.00
XCR	四川乡城县然乌温泉	99.84	28.68	10	37.06	-5.83	4.00	94.00
XG	云南香格里拉	99.91	27.80	10	38.36	-1.83	9.00	91.00
XCW	四川乡城县无名山	100.02	29.12	10	42.58	-7.72	3.00	103.00
LJX	云南丽江雄古景区	100.03	26.86	10	45.19	7.60	15.00	97.00
LJ	云南丽江高山植物园	100.19	27.00	10	39.14	2.58	12.00	96.00
DCK	四川稻城县卡龙	100.26	28.68	10	41.40	-3.28	3.00	111.00
LJB	云南丽江白沙村	100.26	26.94	10	41.47	4.40	13.00	97.00
DCE	四川稻城县俄洛村	100.28	28.95	10	42.13	-5.18	2.00	112.00
YJH	四川雅江红龙乡	100.69	30.15	10	36.36	-5.62	6.00	157.00
YJX	四川雅江	100.95	30.01	10	40.43	-0.95	1.00	131.00
DFL	四川道孚县龙灯乡	101.12	30.08	10	40.36	0.90	1.00	146.00
DFY	四川道孚县银恩乡	101.15	31.44	10	42.54	-3.38	1.00	138.00
RT	四川壤塘县	101.24	31.79	9	42.38	-4.38	3.00	138.00
DFG	四川道孚县葛卡乡	101.30	30.79	8	39.58	-6.10	3.00	134.00
KDD	四川康定县折多山	101.42	30.07	6	39.24	-3.02	2.00	142.00
MEKR	四川马尔康县松岗镇	101.46	31.78	7	40.23	-1.40	1.00	153.00
DMX	四川康定县明希乡	101.50	29.71	10	42.90	-1.45	3.00	138.00

YYP	四川盐源县至平川垭口	101.71	27.53	10	40.11	0.08	1.00	158.00
KDC	四川康定县中古村	101.89	30.24	10	43.12	-0.53	2.00	156.00
KDZ	四川康定县折多塘村	101.90	29.99	10	36.44	-1.52	3.00	161.00
JCD	四川金川县杜岗山	102.04	31.48	9	42.92	2.13	2.00	132.00
JCK	四川金川县卡撒乡	102.06	31.27	10	42.08	3.70	2.00	129.00
MEKD	四川马尔康县大郎脚沟	102.12	31.91	10	44.16	0.03	3.00	129.00
XJX	四川小金县下磨子村	102.39	30.90	10	40.37	-1.27	2.00	146.00
MEKS	四川马尔康县梭磨乡	102.51	31.86	10	41.42	-3.30	3.00	133.00
XJD	四川小金县达维乡	102.52	30.88	10	41.41	-8.05	3.00	152.00
LX	四川理县	102.82	31.65	10	36.52	-3.98	4.00	129.00
HS	四川黑水县	102.93	32.08	10	41.72	-1.22	5.00	124.00
MX	四川茂县	103.52	31.77	10	35.06	-2.87	4.00	123.00
WC	四川汶川县	103.67	31.49	10	38.27	-0.08	8.00	129.00
PW	四川平武县	104.53	32.46	10	34.93	4.80	4.00	105.00

2.1.2 下载候选基因及设计引物

从已经公开发表欧洲栎树功能基因组网站 EvolTree (<http://www.evoltree.eu/>)中的候选基因数据库中选取 180 个 *Q. petraea*/*Q. robur* 的抗干旱、寒冷的基因，并统计其详细的功能特征(网站内的描述)。

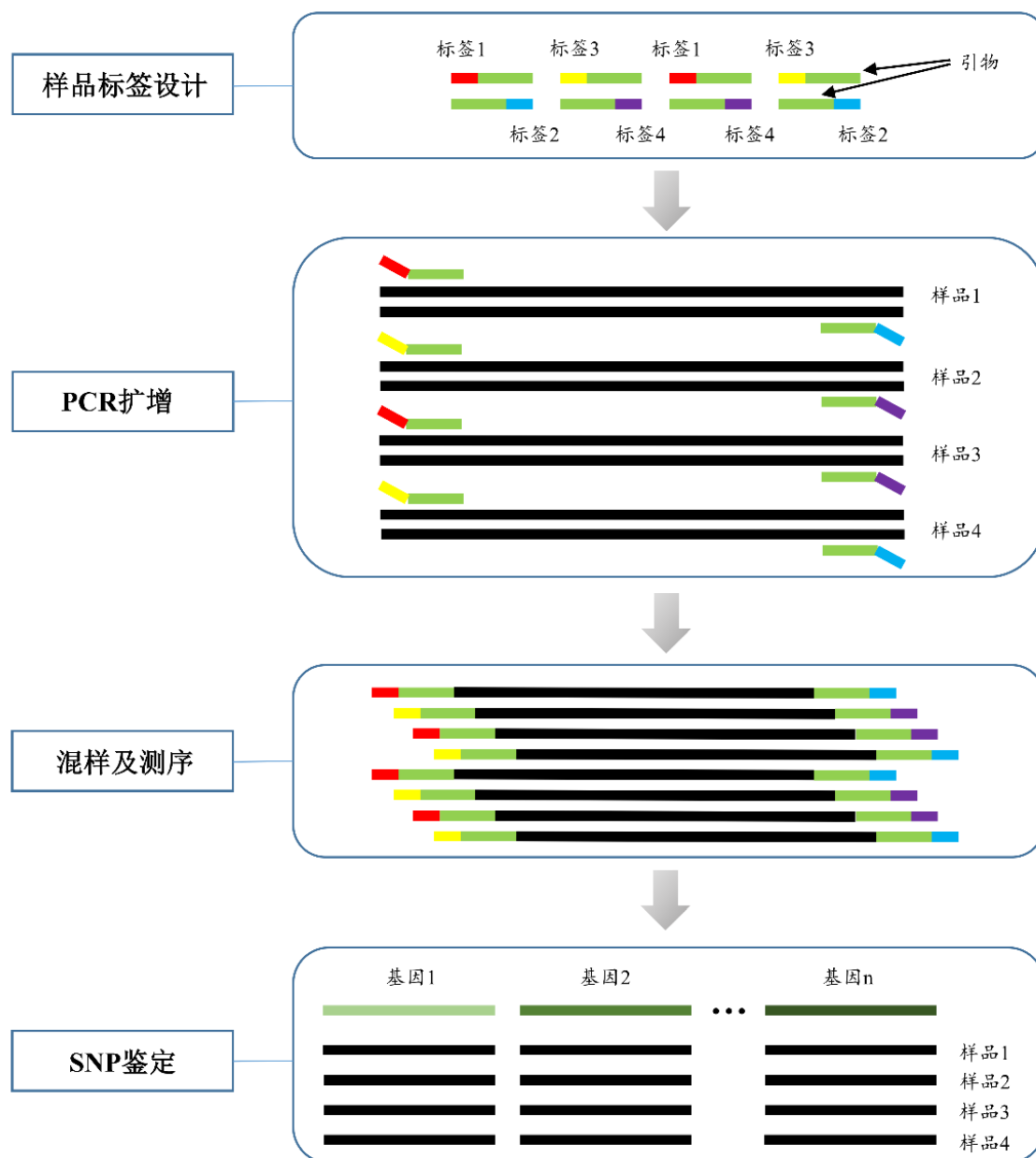


图 2-2 混样测序方法的流程图

Fig. 2-2 Flow chart of pool-sequencing method

使用 Primer3web (<http://primer3.ut.ee/>; Untergasser *et al.*, 2012)对下载到的 180 个候选基因进行引物设计。Primer3web 是一个操作简单的在线引物设计工具，只需将模板序列复制进去，并且设定简单的参数，例如产物序列长度范围、引物长度、退火温

度(T_m)等,便会得到多条符合条件的引物,并且展示其重要参数,包括引物长度、 T_m 、G/C 含量等。我们根据这些信息,挑选最完美的引物,将序列信息交由上海生工公司,合成实验用的双端引物试剂。

引物设计的操作比较容易,但是我们依然强调以下几点设计引物的原则(尤超等, 2011; 任亮等, 2005; 张新宇和高燕宁, 2004):

- (1) 要在序列的保守区设计引物;
- (2) 引物长度应在 16~30 个碱基之间,过短会降低引物的特异性,容易形成错配,过长会增加延伸温度,超过 DNA 聚合酶的最适反应温度(74°C);
- (3) 上下游引物的 T_m 越近越好,方便设定 PCR 反应程序;
- (4) 引物中四种碱基随机分布,不要出现连续三个碱基相同的情况;
- (5) 引物 G/C 含量要适中,并且前后引物 G/C 含量不要相差太多;
- (6) 上下游引物之间不要出现连续四个碱基的互补配对,以避免出现引物二聚体;
- (7) 引物自身不要有连续三个碱基的互补配对,以免出现由自身折叠引起的发卡结构,影响引物与模板的结合;
- (8) 扩增片段长度尽量小于测序长度。如使用 Pair-End 250 文库进行测序,测序长度为双端一共 500bp,则扩增片段长度应小于 500bp,以避免中间的碱基无法被检测到,造成数据丢失。

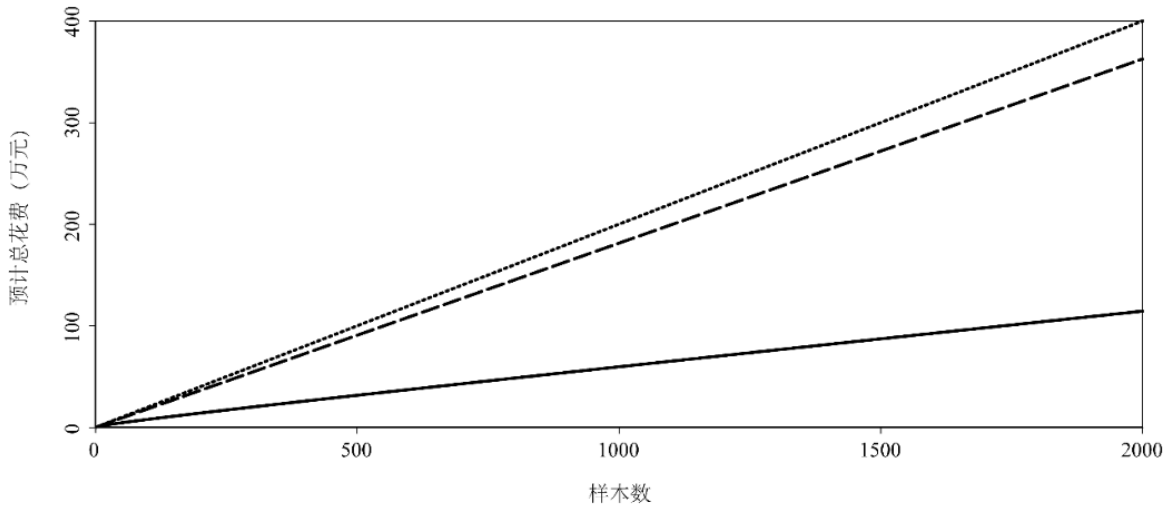
由于可能出现特殊情况无法完全满足以上原则,可以适当放松条件,但要尽可能多地遵守以上原则。

2.1.3 设计标签(barcode)序列

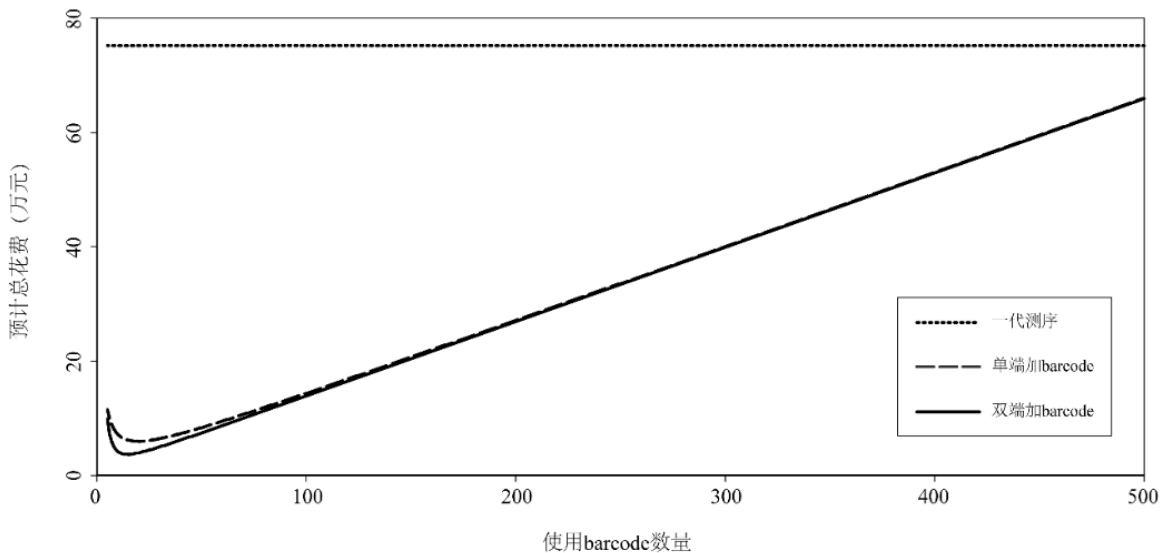
在经过验证且具有多样性的基因的引物一端或两端加上不同的 barcode 序列,分别用于扩增不同的样本个体,可以使不同个体的序列在混样及二代测序后通过 barcode 序列进行区分(图 2-2)。此步是整个实验流程的关键所在,通过此步骤使几十甚至上百个样本个体的 PCR 产物可以混合到一个文库中进行测序,极大地提高效率,降低了建库成本。当使用单端 barcode 时,需要使用的 barcode 数量须等于样本数量(假设建立一个测序文库);使用双端 barcode 时,用 $m + n$ 条 barcode,便可用于区分 $m \times n$ 个样本(前后两个 barcode 逐一配对,配合使用)。使用单端 barcode 时,仅需要再合成单端的“引物+barcode”序列,另一端可使用验证引物时用到的引物。而使用双端 barcode 时,则需要再合成两端的“引物+barcode”序列。虽然单端 barcode 的方法只需要合成一端的“引物+barcode”,但是一个“引物+barcode”只能区分一个样本个体,因此当样本数量比较多时,需要合成的“引物+barcode”也会很多,而双端“引物+barcode”配合使用,则不需要太多数量。

我们预测了假设的三种测序方法的总费用:一代 Sanger 测序、单端加 barcode 的二代高通量测序和双端加 barcode 的二代测序(图 2-3a)。总花费包括引物合成费用、“引物+barcode”合成费用、PCR 费用、建库费用和测序费用。以 100 个基因为例,

需要设计 100 对引物，基因序列长度设定为 500bp；引物长度假设为 20bp，barcode 长度假设为 6bp，合成费用为 0.5 元/bp；PCR 费用预计 5 元/反应（包括实验用到的试剂、耗材等）；二代测序建库费用以 1000 元为例，测序深度为 50 倍，测序费用为 500 元/G；一代测序的费用不包括“引物+barcode”合成费用和建库费用，测序费用为 15 元/反应。从图 2-3a 可以看出，随着样本数量的增多，一代测序的费用始终最高，而双端加 barcode 的二代测序费用始终最低。



a 假设建立1个测序文库, barcode使用数量随着样本数量的变化而变化



b 假设样本数量为500, 测序文库的数量随barcode的使用数量的变化而变化

图 2-3 预测三种测序方法(一代测序、单端加 barcode 的二代测序和双端加 barcode 的二代测序)的总费用

Fig. 2-3 The estimation of total costs using three sequencing methods (Sanger sequencing, NGS with Single-End barcode, NGS with Pair-End barcodes)

但是，以上预测是建立在只有一个测序文库的前提之下的。我们可以建立多个文库，以减少 barcode 的使用数量，从而减少“引物+barcode”的巨额合成费用。我们假设使用 500 个样本，因此建库的数量会根据 barcode 使用数量的改变而改变。例如使用 10 条单端 barcode，这时需要建立 50 个测序文库，当使用 20 条单端 barcode 时，只需要建立 25 个测序文库。双端 barcode 也是类似的情况，例如当使用 5 对双端 barcode 时，需要建立 20 个测序文库，当使用 10 对双端 barcode 时，只需要建立 5 个测序文库。此时再对总费用进行估测(图 2-3b)。由于样本数量固定，PCR 费用不变，因此我们在进行本项预测时排除 PCR 费用。从图 2-3b 可以看出，一代测序仍然是费用最高的方法；两种二代测序方法花费差别不大，但双端加 barcode 测序方法的最低费用依然比单端加 barcode 方法的最低费用低两万元左右(单端 barcode 数量为 25 时，费用最低，为 59250 元；双端 barcode 数量为 15 时，费用最低，为 36750 元)。这两种方法相比，barcode 使用数目一样的情况下，“引物+barcode”的合成费用一致，引物合成费用、测序费用以及没有被统计的 PCR 费用均一致，唯一的区别为建库费用。因此，这种可以区分更多样本个体而建立更少测序文库的双端加 barcode 的测序方法更加节约成本。然而，从图 2b 中可以看出，barcode 使用数量并不是越多越好。在设计实验时，我们需要根据实际情况，进行费用预估，使用适当数量的 barcode 进行后续实验。

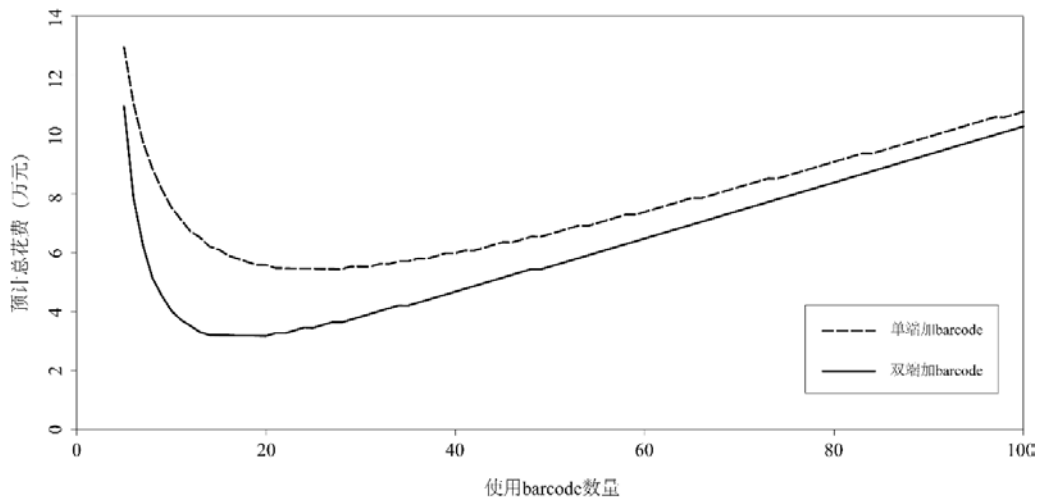


图 2-4 预测使用不同数目的 barcode 所需要的总费用

Fig. 2-4 The estimation of total costs using different number of barcodes

本实验选择 587 个样本个体，65 个候选基因(经过引物筛选得到，见下文)进行后续实验。针对这一具体情况，提前对实验的花费进行了预估，包括引物合成费用，“引物+barcode”合成费用，PCR 费用，建库费用和测序费用(图 2-4)。从图 2-4 我们得出结论：使用 20 条(即 10 对，表 2-2)barcode 用于双端加 barcode 的二代测序方法，最节省开支。此时，前后两段 barcode 两两组合配对，共有 100 种组合方式，在 PCR 扩

增时，分别与 100 个样本对应进行实验，这样每对 barcode 便与对应的样本 DNA 序列连接在一起，在混样测序后可用于区分此 100 个样本的序列。

Barcode 序列的设计原则与引物设计原则类似：(1) barcode 长度一般为 6bp，每个 barcode 之间至少有 3 个碱基的差异，以进行有效地区分；(2) 嘌呤和嘧啶的组成大致相当；(3) 避免连续三个碱基的重复；(4) 避免形成引物二聚体；(5) 避免与引物一起形成发卡结构等。

表 2-2 用于辨别样本个体的标签序列

Table 2-2 Barcode sequences used for distinguishing each samples

barcode 编号	F 端序列	R 端序列
bar01	AATAAT	ATCGGT
bar02	AATTGA	AGCTAT
bar03	AATGTG	ATCTTA
bar04	AAGATA	AGCCGA
bar05	AAGTAG	ACCTGG
bar06	AAGGCT	AGGTGC
bar07	AACACG	ACGCGT
bar08	AACGAA	ACGTCA
bar09	AACCTT	AGGCCG
bar10	ATCCAG	ATGACC

2.1.4 实验方法

2.1.4.1 主要试剂及仪器

主要试剂：

无水乙醇 (Ethanol anhydrous)	上海生工
三氯甲烷 (Trichloromethane)	北京化工
β -巯基乙醇 (β -Mercaptoethanol)	北京 MYM
异丙醇 (isopropanol)	上海生工
氢氧化钠 (Sodium hydroxide)	上海生工
乙二胺四乙酸 (Ethylenediaminetetraacetic acid, EDTA)	上海生工
三(羟甲基)氨基甲烷 (Tris)	上海生工
乙酸 (Acetic Acid)	上海生工
1M Tris-buffer solution (pH = 8.0)	上海生工
十六烷基三甲基溴化铵 (Cetyltrimethyl Ammonium Bromide, CTAB)	上海生工
氯化钠 (Sodium chloride)	上海生工
植物基因组 DNA 提取试剂盒	TIANGEN
2 × Taq PCR Master Mix	AIDLAB
荧光 M13 引物试剂	上海生工

琼脂糖 (Regular AGAROSE G-10)

Biowest

Loading buffer

主要试剂配制：

(1) 10M NaOH

NaOH	40g
ddH ₂ O	96mL

加入 ddH₂O 定容至 100mL，灭菌后保存。

(2) 0.5M EDTA (pH=8.0)

EDTA	73.08g
ddH ₂ O	300mL

加入 10M NaOH 调节 pH 至 8.0，加入 ddH₂O 定容至 500mL，灭菌后保存。

(3) CTAB 提取液

CTAB	4g
NaCl	16.364g
1M Tris-buffer solution (pH=8.0)	20mL
0.5M EDTA (pH=8.0)	8mL
ddH ₂ O	70mL

溶解后加入 ddH₂O 定容至 200mL，灭菌后保存

(4) 50 × TAE (pH=8.0)

Tris	242g
乙酸	57.1mL
0.5M EDTA (pH=8.0)	100mL

加入 ddH₂O 定容至 1L，混匀后保存，使用前稀释为 1× TAE 工作液。

主要设备仪器：

不同量程微量移液器	Eppendorf
电子天平	Sartorius BSA-124S
微量酸度计	奥立龙 PHS-25C
研磨仪	Etsch MM400
恒温水浴锅	NUOHAI XMTD-204
高速冷冻离心机	SCIOLOGEX D3024R
电热恒温鼓风干燥箱	一恒 DHG-9023A
微波炉	Galanz P70D20TP-C6
电泳仪	君意 JY200C
凝胶成像系统	君意 JY04S-3C
超微量紫外分光光度计	Thermo NanoDrop 8000
低速离心机	赫西 TD5A

PCR 扩增仪	东胜 ETC811
漩涡混合器	Kylin-Bell VORTEX-5
冰柜	Haier BC/BD-429H
高压灭菌锅	TOMY SX-500
制冰机	GRANT XB100-FZ

2.1.4.2 基因组总 DNA 提取

本研究对采集到的川滇高山栎的每个个体的经过硅胶干燥后的叶片进行总 DNA 提取。首先使用 TIANGEN 植物基因组 DNA 提取试剂盒对川滇高山栎的干燥叶片进行 DNA 提取。对于提取失败的个体再使用改良 CTAB (Doyle, 1987)提取 DNA。

2.1.4.3 SSR 分型

由于川滇高山栎的遗传分化不仅会来自于生态适应,也可能来自于种群历史动态,如基因流和遗传漂变。因此,利用中性的 SSR 分子标记对遗传结构进行分析,可以为使用非中性位点研究生态适应提供参考依据。

(1) 筛选 SSR 引物

从文献中查找从其他栎树中开发的 SSR 引物 34 对,交由上海生工生物公司合成实验用引物试剂,挑选 32 个样本个体全基因组 DNA 作为模板进行扩增。经过 2%琼脂糖凝胶电泳检测,筛选出 8 对成功扩增的引物(表 2-3)进行批量 PCR 扩增。扩增反应体系及反应程序如下:

扩增体系(15 μ L):

2 \times Taq PCR Master Mix	7.5 μ L
引物-F	0.24 μ L
引物-R	0.24 μ L
样本 DNA	1 μ L
ddH ₂ O	6.02 μ L

扩增程序:

95 $^{\circ}$ C 预变性 10mins	
95 $^{\circ}$ C 变性 30s	} 35 次循环
56 $^{\circ}$ C 退火 30s	
72 $^{\circ}$ C 延伸 30s	
72 $^{\circ}$ C 延伸 10min	
15 $^{\circ}$ C 保存	

(2) 批量 PCR 扩增

将成功筛选的引物 F 端序列送至上海生工生物公司,合成 5'端加 M13 序列的实验用引物试剂,配以 M13 荧光引物接头,以所有 587 个样本个体的基因组 DNA 作为模板,进行批量 PCR 扩增。由于荧光 M13 序列遇光容易失去荧光效果,因此配制

PCR 反应体系时进行避光处理。反应程序与上步相同，扩增体系如下(15 μ L)：

2 \times Taq PCR Master Mix	7.5 μ L
引物-F+M13	0.06 μ L
引物-R	0.24 μ L
荧光 M13 引物	0.24 μ L
样本 DNA	1 μ L
ddH ₂ O	5.96 μ L

将 PCR 扩增产物交由上海生工生物公司 ABI PRISM X3730XL 测序平台进行毛细管电泳测序。

2.1.4.4 候选基因引物筛选

由于从数据库中下载的候选基因并非来自川滇高山栎，以它们为模板设计的引物并不一定可以在川滇高山栎的基因组 DNA 中扩增成功。为了避免在正式大批量 PCR 扩增实验中出现实验失败的情况，我们选择 8 个川滇高山栎样本的基因组 DNA 作为扩增模板，对设计的候选基因 PCR 引物进行预扩增验证，以确保其可用性。若扩增失败，则尝试重新设计引物。将成功扩增的 PCR 产物由上海生工生物公司 ABI X3730XL 测序平台进行 Sanger 测序。使用 MEGA 7.0 (Kumar *et al.*, 2016) 比对测序结果序列，保留存在变异的候选基因。我们从 180 个候选基因中设计、筛选得到 65 对引物(表 2-4)，用于后续实验。此步骤使用的扩增体系及扩增程序如下：

扩增体系(25 μ L)：

2 \times Taq PCR Master Mix	12.5 μ L
引物-F	1 μ L
引物-R	1 μ L
样本 DNA	1 μ L
ddH ₂ O	9.5 μ L

扩增程序：

95 $^{\circ}$ C 预变性 10mins

95 $^{\circ}$ C 变性 35s	} 35 次循环
T _m $^{\circ}$ C 退火 30s	
72 $^{\circ}$ C 延伸 40s	

72 $^{\circ}$ C 延伸 10min

15 $^{\circ}$ C 保存

表 2-3 可用于川滇高山栎扩增的 SSR 引物序列信息

Table 2-3 Primer information of SSR which could be used for PCR amplifying of *Q. aquifolioides*

基因座	F 端引物	R 端引物	基序	等位基因长度范围 (bp)	参考文献
MSQ13	TGGCTGCACCTATGGCTCTTAG	ACACTCAGACCCACCATTTTTC	TC	190-222	Dow <i>et al.</i> , 1995
ssrQpZAG16	CTTCACTGGCTTTTCCTCCT	TGAAGCCCTTGTC AACATGC	TC	138-182	Steinkellner <i>et al.</i> , 1997
ssrQrZAG7	CAACTTGGTGTTCCGATCAA	GTGCATTTCTTTTATAGCATTAC	AG	110-148	Kampfer <i>et al.</i> , 1998
ssrQrZAG11	CCTTGAACCTCGAAGGTGTCCTT	GTAGGTCAA AACATTGGTTGACT	AG	243-281	Kampfer <i>et al.</i> , 1998
ssrQrZAG30	TGCTCCGTCATAATCTTGCTCTGA	GCAATCCTATCATGCACATGCACAT	GA	162-226	Kampfer <i>et al.</i> , 1998
ssrQrZAG87	TCCCACCACTTTGGTCTCTCA	GTTGTC AGCAGTGGGATGGGTA	TC	97-111	Kampfer <i>et al.</i> , 1998
ssrQrZAG96	CCCAGTCACATCCACTACTGTCC	GGTTGGGAAAAGGAGATCAGA	AG	162-218	Kampfer <i>et al.</i> , 1998
PIE_271	CACACTCACCAACCCTACCC	GTGCGGTTGTAGACGGAGAT	AG	200-226	Durand <i>et al.</i> , 2010

表 2-4 可用于川滇高山栎扩增的候选基因引物和功能

Table 2-4 Primer and annotation of candidate genes which could be used for PCR amplifying of *Quercus aquifolioides*

基因座	F 端引物	R 端引物	功能
DF_Qa003	G TTCACCATGACCCTCAA	GCACATATTATTGCCAGTTCTA	cytochrome P450
DF_Qa017	T TTTTGATACAAGACTAGCCTCT	AGGACCTCCAACATACCC	UDP-glucose 6-dehydrogenase
DF_Qa019	A CTTC AAATGGCACAGCA	AACGCCCTTGAGAAACT	hexose transporter 1
DF_Qa020	G TAGGACACCTTCACCCAT	CAGCCTCAGAACCAAATC	9-cis-epoxycarotenoid dioxygenase
DF_Qa026	T GTCTGGGTGCAAAGTAG	ATTGTCTGGATGATGGGT	zinc finger protein, putative
DF_Qa028	C TCGTGGTGTTAGGTCAG	ACATCAGGCATGGAAGAT	aspartic proteinase
DF_Qa030	A CTGGGTTTAAGGGAGGT	TACAAGCATCAAGGCAGA	proton gradient regulation 5
DF_Qa031	G AGCCCAATAAGAAGTCAA	CCATGTCATCTCCACCAT	CBF/DREB1 transcription factor 2
DF_Qa035	C CGAGTAAAGCGAAAGCC	CGACGAAACCGAAGGAAG	Xylem serine proteinase 1 precursor
DF_Qa037	A AAGGGGCAAGAAGAAGC	GAACCACAAATTTACCGT	glucose-6-phosphate dehydrogenase

DF_Qa038	CACCCTTCATCATCACCA	GATTCCTATTAGGCTTCCT	fiddlehead-like protein
DF_Qa039	ACTATGCGGGTGC GTTGG	CAGCCTGAAGGAGTTGATGTAA	amino acid permease
DF_Qa040	TCGGCTACACTATCACTG	GTCCTCCACCTTCATTTT	AAP6 (AMINO ACID PERMEASE 6)
DF_Qa041	TTTGTGAGCCAGTGCTGC	GAGACTGCGATTCGACCA	o-methyltransferase
DF_Qa045	CATCTCACCTGGACGACA	ACTCTTTCACTGCCTAAAC	NAC domain protein NAC4
DF_Qa047	CACTCCCTGCTGACCTAA	ATCCTCAACAGCCATTCC	lipoxigenase
DF_Qa051	TGATGATGGGTATGGTT	GCAAAAGTGT CACAAGGAT	AMP dependent CoA ligase
DF_Qa052	CCTGTCCCTCAATGTAAG	GTCTGTAACCATGCCAAC	glutathione S-transferase
DF_Qa054	ACGGATGCAGTTGGTAAC	TCCAAACCTGAAGCAAAG	amino acid binding protein
DF_Qa055	ACGGATGCAGTTGGTAAC	TCCAAACCTGAAGCAAAG	flavonoid 3-hydroxylase
DF_Qa056	TGCCCTGAAATCCATAAT	GCTTCTCCGAGACTGTTG	indole-3-acetic acid 14 transcription factor
DF_Qa057	GAGGCTTGCCACTCACTC	TGGGCTTCCAACCACATCA	calcium dependent protein kinase-gb
DF_Qa059	AAACCAGAAGAGCAACAT	TGCCTTTTCTATCATACT	cellulose synthase
DF_Qa064	GCAGTCCATGACTTGGGATA	GAGCCTTTGTTGAGATTGTG	multicopper oxidase
DF_Qa070	GAGGAAGGTGAAGGGAGA	GCCAGCCAAGGTAACAAA	IBR10 (INDOLE-3-BUTYRIC ACID RESPONSE 10)
DF_Qa072	TTCAGATGCTGACCCAGAT	AAGGTCGATCATCATTGTAAAG	methionine-tRNA synthetase
DF_Qa077	GAGTTGTCAAGCCTTTGG	GTGGTTGGCTCAGGTTAT	laccase, putative
DF_Qa079	CATCTAAGTAAGGGTGAGC	AGAGGCCAGAGTAGTGAA	mitotic control protein dis3
DF_Qa081	TGGACTGCTGAGGAGGTT	TATTTGGTCGCGGTAATT	abscisic acid responsive elements-binding protein 2
DF_Qa082	GACCTGCTTCTTACACGA	GTCTGCCACTGGACATCA	Eukaryotic initiation factor-5
DF_Qa083	ACTCCCTGACATTTACATTC	ACGCTGGTGACTAATCCT	4-coumarate:coA ligase 2
DF_Qa084	TGCACTCGAAATCCTCAC	TGCCAGCATCAAACCTCTT	white-brown-complex ABC transporter family
DF_Qa089	TGTGACAGGAAAGCCCATA	CTGCCAGTTCTTTCTTTATGC	GDH2
DF_Qa091	TGGGATGGGTTGTAGTGC	CCAAAGGCTACTGGCTGA	beta-ketoacyl-coa synthase family protein
DF_Qa093	CCTACTGGTGGTGGAAATC	CCAAGATATTGTCGTGCC	PS60
DF_Qa094	GCTTAAC TTTGAATCGTGG	TTGCAGCTACTCTAATACTTGTG	Density-regulated protein
DF_Qa095	TTCTGGGCTTGCTTGTGCT	TCCGCAGTCTTCAATACCTTC	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein
DF_Qa096	TAGCATTCACTGCCACCAT	CACTAGCCTCGACTCAT	glucose-methanol-choline (gmc) oxidoreductase

DF_Qa100	CCATGACTCCAGGTGGTAC	CGTTCATCTCAGACATTTCG	PREDICTED: similar to ring domain containing protein isoform 2
DF_Qa102	TCTTGAACGGAGAAGTGG	TCCCATACTCAGCACCAG	aconitase-iron regulated protein 1
DF_Qa103	TGTGGCACTGTTGTAAGA	GCTACTTGAAACCCTTTG	4-coumarate-CoA ligase
DF_Qa104	AGGGATGCTAAGATGGAC	ACCTTCGTAGACCTGGAT	heat shock protein 70-3
DF_Qa105	GGCAGGTTACTACATTCCAG	CTAGGTGCAGTGAGTCCAAA	cytochrome P450 monooxygenase CYP84A16
DF_Qa109	TCCACCTCTATTTCGCTCAG	CAAAGGGCACGGTCTATT	glycosyltransferase family GT8
DF_Qa113	AACAGCAGAAGTTGGGTA	CTTACGATTGGTGGAAGG	aspartic proteinase
DF_Qa118	AGAGGGCACGAACACCAC	CAGGACCTTTGTAACCAT	glutaredoxin
DF_Qa119	AAAACAGGGCAAGTAGAAT	GTACTGGGTTGAGCGTCTA	small heat-shock protein
DF_Qa120	ATTAACCCTCAAACCATG	CTCAGGTGCTTTCCAGTC	26S proteasome non-ATPase regulatory subunit
DF_Qa127	AAACACTGCTCTTGTAAGGGAC	CTACAGGGCTGGGATTGC	plasma membrane major intrinsic protein
DF_Qa129	TAGATCCATTGATGGTGGGG	TCGGGACAACAACATAGACAGT	phospholipase D alpha
DF_Qa132	ATCTAGTGCGAGTTCGTTCA	ACAATTTTGGCTGCCTTG	stress-associated protein 1
DF_Qa134	GCTGTAGGGCGTGCTGAA	CAAGGGTAGTCCTTACTGGATG	mitogen-activated kinase kinase kinase alpha
DF_Qa137	ATAGTGAGGCCGAAGAAA	ACAGAACCAGGCGATACA	phosphoprotein phosphatase
DF_Qa142	AATCCTTCCTTAGGCTCT	CTGTTGATGCTGGTTATG	r2r3-myb transcription factor
DF_Qa143	GTGAAACCTCCGTCTATC	AAACCACCAAGGCTGAAT	short chain alcohol dehydrogenase
DF_Qa145	TGCGGGTAAGTGGTGGAG	CCATTGTGGATTGCGTCA	pyruvate decarboxylase
DF_Qa146	TTCACTCGTGTCCTCC	AGATCCAGCAGCCCCTCC	50S ribosomal protein L7/L12
DF_Qa150	CCTTATTCTATCTGGAGCAG	TGGAACCACCCATCAACC	long-chain acyl-CoA synthetase 4
DF_Qa151	TGGTACAGAAACCGCATT	TGTGAATCGCTGGAGTGT	phosphoenolpyruvate carboxylase
DF_Qa159	GATTGCCGTTCTTCTAG	AACAGCCAACACTTCTT	O-glucosyltransferase 2
DF_Qa163	AGAAGCCATAAACTCCAG	ATCGCTACATAGCAGATACA	sucrose synthase
DF_Qa167	CCCTTTTACCAAACCATA	GTGCCATCATCAAACAAC	R2R3 Myb30 transcription factor
DF_Qa171	CCAATCAACTGCCTAA	GGATTTTCGTTTACCACCA	heavy metal P-type ATPase
DF_Qa172	AGAAGGAAGGGCAGAGTC	ATAACCATCGGCAGCATT	putative transaldolase
DF_Qa180	GGAGCATGGACAGTTATG	GACCAGCTATCAAAGACC	Myb-related transcription factor

2.1.4.5 候选基因 PCR 扩增及高通量测序

每对筛选成功的引物均与 10 对 barcode (表 2-2)连接, 组成 10 对“引物+barcode”组合, 交由上海生工生物公司合成实验用引物试剂。将每一个候选基因的 10 对“引物+barcode”两两组合, 共 100 种组合方式, 分别与 100 个样本对应进行 PCR 扩增。经过大批量 PCR 扩增、电泳检测、补样等步骤, 从 100 个样本个体的所有候选基因的 PCR 产物中分别抽取 1.5 μ L, 均匀混合后, 交由清华大学生物医学测试中心 Illumina MiSeq 高通量测序平台进行建库测序。建库类型为 Pair End 250。本研究一共用到 587 个样本个体, 因此共建立了 6 个测序文库, 实际每个文库分别混合了 100、100、99、96、96、96 个样本个体的 PCR 产物。PCR 扩增体系及扩增程序与引物筛选时用到的体系和程序一致。

2.1.5 处理测序结果

2.1.5.1 质量控制

由于建立了 6 个测序文库来进行测序, 因此测序的结果是 6 对 fastq 文件, 每对文件代表从一个测序文库中测得的双端序列, 每条序列数据长度为 250bp。每对 fastq 文件包含了约 100 个样本个体的序列。为了方便后续的分析, 需要将每个样本个体的序列划分为独立的文件。根据双端 barcode 对应的样本信息, 编写 perl 脚本命令, 在 linux 系统中可将每个样本个体的序列分开, 共获得 587 对包含单个样本序列的 fastq 文件。

由于测序数据是没有经过质量控制, 质量较低的 row data, 因此需要对每一对包含每个样本个体序列的 fastq 文件进行质控处理, 以去除低质量的序列, 并剪掉测序时加上的 Illumina 平台专用的 index 序列。这一步需要用到 trimmomatic-0.36.jar 程序 (Bolger *et al.*, 2014), 并在 linux 系统中运行。

2.1.5.2 使用 GATK pipeline 鉴定 SNP

使用 GATK pipeline (Mckenna *et al.*, 2010)鉴定 SNP 需要在 linux 系统下进行, 因此以下每一个小步骤均不再具体说明。另外这一步会多次用到参考序列。我们在筛选候选基因引物的时候进行了 Sanger 测序, 汇总其测序结果, 合成一个包含 65 个候选基因序列的文件, 作为参考序列文件。

(1) 使用 BWA-MEM (Li & Durbin, 2009)对经过质控处理的 587 对代表每个个体的 fastq 文件进行序列比对。

(2) 使用 samtools (Li, 2011)将上一步生成的包涵比对序列的 sam 格式的文件转换为二进制的 bam 格式文件, 以减轻运算负担。由于 BWA 生成的文件中序列排列顺序与 GATK 在鉴定 SNP 时需要的序列排列顺序不同, 因此接下来需要使用 Picard Tools (<http://broadinstitute.github.io/picard>)对 bam 格式的文件进行重排序处理。

(3) 由于 GATK 不支持对无头文件的变异检测, 因此需要使用 Picard Tools 对 bam 格式文件进行加头处理。加头内容包括序列集 ID 号、序列集文库号、测序平台、测

序平台下级单位名称和样本名称。

(4) 由于在制备文库的过程中,有的序列会被过量 PCR 扩增,这样在比对的时候,这些过量扩增的完全相同的序列就会比对到相同的位置。而这些序列不能作为变异检测的证据,所以要尽量去除这些由 PCR 扩增形成的重复,并生成索引以备下一步使用。此步需要的程序为 Picard Tools 和 samtools。

(5) 由于比对后的序列存在插入/缺失,并且在插入/缺失附件的比对会出现大量的碱基错配,这些碱基的错配会很容易地被误认为 SNP。因此为了将插入/缺失附件的比对错误率降到最低,需要使用 GenomeAnalysisTK.jar 和 samtools 将这些区域重新进行比对,得到新的 bam 格式比对文件。

(6) 使用 GenomeAnalysisTK.jar 进行第一次 SNP 鉴定,并过滤掉低质量的变异位点,得到存放高质量变异位点数据的 vcf 格式文件。

(7) 根据上一步得到的 vcf 文件,使用 GenomeAnalysisTK.jar 中的 BaseRecalibrator 和 PrintReads 功能对第(5)步重新比对得到的 bam 格式文件进行矫正。

(8) 使用 GenomeAnalysisTK.jar 进行第二次 SNP 鉴定,并过滤掉低质量的变异位点,得到存放高质量变异位点数据的 vcf 格式文件。

(9) 使用 GenomeAnalysisTK.jar 将上步得到的 vcf 文件进行过滤,删掉其他类型的变异位点,只留下 SNP。

通过以上九个步骤,使用 GATK pipeline 从测序数据中得到高质量的 SNP 位点。

2.1.5.3 筛选 SNP

使用 VCFtools (Danecek *et al.*, 2011)将次要等位基因频率(minor allele frequency, MAF)小于 2.5%的 SNP 过滤掉(Rellstab *et al.*, 2016),并使用 Excel 删去数据缺失比例大于 50%的 SNP。后续数据分析均基于此 SNP 数据集。

2.1.5.4 毛细管电泳测序获取数据

使用 GenMarker (Hulce *et al.*, 2011)对毛细管电泳测序结果进行读取以获取 SSR 基因型数据。每个基因型结果均由两个人读取。当两个人读取的结果不一致时,我们试图确定一个一致的基因型,并将不能达成一致的基因型数据作为数据缺失。接下来,我们使用 FLEXIBIN Excel 宏(Amos *et al.*, 2007)对 SSR 分型数据进行归一化处理,以简化后续数据分析。

2.1.6 群体遗传学分析

2.1.6.1 遗传结构分析

为确定川滇高山栎的种群遗传结构和评估不同分子标记对种群遗传结构的影响,我们使用 R 语言(Team R Core, 2017)中的“adegenet”函数包(Jombart & Ahmed, 2011),基于 SSR 数据集、包含所有 SNP 的数据集、包含中性 SNP 的数据集对其进行主成分分析(PCA, principal components analysis)。

同时我们使用一种贝叶斯马尔可夫链蒙特卡罗(Bayesian Markov chain Monte

Carlo, MCMC)聚类方法: STRUCTURE v.2.3.3 (Falush *et al.*, 2003; Hubisz *et al.*, 2009; Pritchard *et al.*, 2000), 并基于 SSR 数据集、包含所有 SNP 的数据集、包含中性 SNP 的数据集对其进行评估。每一个数据集均假设为中性遗传标记位点。对于每一个数据集, 我们设置 $K=1-10$ 来预测真实的进化枝数目, 对于每一个 K 值, 我们运行 20 个独立的运算。每一个独立的运算设置 100,000 burn-in 迭代步数和 200,000 MCMC 循环。运算结束后, 使用一种基于 ΔK (Evanno *et al.*, 2005)的方法 Structure Harvester 在线工具 (<http://taylor0.biology.ucla.edu/structureHarvester>, Earl & Vonholdt, 2012)对 STRUCTURE 结果进行评估。选择 ΔK 最大时的 K 值作为最优的进化枝数量, 并在此 K 值的 20 个 STRUCTURE 运算结果中选择 likelihood 值最高的结果, 使用 distruct 1.1 (Rosenberg, 2004)绘制呈现每个种群、每个个体的分组可能性的柱状图。

基于 SSR 和 SNP 数据集的 PCA 和 STRUCTURE 分析均检测到极强的遗传结构, 并将川滇高山栎的遗传结构分成两个分枝, 我们根据其大致的地理位置将其命名为 Tibet lineage 和 HDM-WSP lineage (Hengduan Mountain and western of Sichuan province lineage), 分别包含 17 和 43 个种群。基于这样的分组结果, 我们使用 Arlequin ver3.5 (Excoffier & Lischer, 2010)对三个数据集进行分层分子方差分析(AMOVA, Analysis of molecular variance)来度量枝系之间和种群之间的遗传分化程度。

通过以上三种方法可以检测川滇高山栎遗传结构及分化程度, 并希望比较不同分子标记得到的结果, 分析非中性位点对遗传结构的影响。中性 SNP 数据集由包含所有 SNP 的数据集删去非中性的 SNP 而得到。而非中性的 SNP 由两种基于 F_{ST} 的检测的方法: BAYESCAN 和 FDIST2 检测而获得(详细的使用方法见下文)。我们将两种方法得到不同结果均视为非中性的 SNP, 从而可极大地降低中性 SNP 的假阳性。

2.1.6.2 遗传多样性分析

使用 Stacks (Catchen *et al.*, 2011; Catchen *et al.*, 2013)基于包含所有 SNP 的数据集和中性 SNP 数据集对川滇高山栎的遗传多样性进行评估。使用 GenAlEx 6.5 (Peakall & Smouse, 2012)基于 SSR 数据集计算川滇高山栎的遗传多样性。首先计算每个种群的遗传多样性, 并计算每个枝系中所有种群的平均值作为此枝系的遗传多样性。同时使用 R 语言对两个枝系的遗传多样性进行方差检验。我们选择以下指标来评估川滇高山栎的遗传多样性: P , 每个位点频率最高的等位基因的平均频率; H_O , 观测杂合度; H_E , 期望杂合度; π , 核苷酸多样性; N_E , 有效种群大小; I , 香农指数。

2.1.6.3 基于 F_{ST} 的 outlier 检验

当 SNP 位点的遗传分化明显较高或较低时, 我们认为它正经历歧化选择或平衡选择的作用。基于此原理, 我们使用 BAYESCAN 和 FDIST2 这两种方法检测 SNP 的分化系数 F_{ST} , 以此来判定此 SNP 位点是否收到选择作用的影响。两种方法配合使用以降低检测结果的假阳性。

我们使用基于贝叶斯算法的 BAYESCAN 来预估 F_{ST} 的后验分布。其中重要参数

的设置为: sample size: 5,000; thinning interval: 10; pilot runs: 20; pilot run length: 5,000; additional burn in: 50,000。由于我们鉴定到的 SNP 位于候选基因上, 其受到选择的可能性远高于随机选择的 SNP, 因此我们将 prior odds 设置为 2 (默认值为 10; Csilléry *et al.*, 2014; Roschanski *et al.*, 2016), 并且在运算结果中, 将 posterior odds 大于 2 的 SNP 视为受到选择作用的影响。关于 FDIST2, 我们运行 100,000 次运算, 并使用 95% 置信区间筛选出 F_{ST} 过高或过低的 SNP, 视其受到选择作用的影响。

在每个枝系中, 结合以上两种方法, 筛选由两种方法同时检测到的受到选择的候选基因。比较两个枝系筛选结果的异同, 找到在两个枝系中特有的和共有的受选择的候选基因。通过比较这些基因的功能, 探讨两个枝系中川滇高山栎的生态适应策略。

2.1.7 景观遗传学分析

2.1.7.1 气候因子提取

从 WorldClim Version2 数据库下载(<http://worldclim.org/version2>; Fick & Hijmans, 2017)全球气候因子栅格数据(1970-2000 年的平均值, 30 seconds 分辨率), 包括 19 个生物气候变量以及 12 个月的 7 种气候变量, 共 103 个变量(表 2-5)。使用 ArcMAP 基于经纬度信息对 60 个采样点的气候数据进行提取。使用 R 语言“usdm”函数包(Naimi *et al.*, 2014)中的 vif (Variance Inflation Factor, 方差膨胀因子)函数删除共线性较强的气候因子, 最终得到四个气候因子用于后续的分析(表 2-1)。这四个气候因子分别为: 等温性(bio03), 最湿润季节的平均温度(bio09), 一月降水量(prec01), 六月降水量(prec06)。

2.1.7.2 基于环境梯度相关性的 outlier 检验

当 SNP 位点的遗传变异梯度与环境梯度显著相关时, 我们认为此 SNP 受到该环境变量的选择。我们使用两种基于环境梯度相关性的 outlier 检验方法: BayEnv 和 LFMM (latent factor mixed model)来检测受到环境选择的 SNP。两种方法配合使用以降低检测结果的假阳性。此步选择上文筛选的四个气候因子作为环境梯度变量。

BayEnv 是一种贝叶斯广义线性混合模型, 我们使用它来检测等位基因频率和环境梯度的之间相关性。首先我们使用中性 SNP 数据集计算等位基因频率与气候因子的协方差矩阵, 作为先验分布模型, 用以控制种群历史对遗传变异的影响。其次使用所有 SNP 数据集计算等位基因频率与气候因子之间的协方差, 并与先验分布模型做比较, 得出后验分布, 计算每个 SNP 的 BF 值(Bayes Factor)作为衡量该 SNP 是否受到环境选择的指标。从所有 SNP 数据集中删去由 BAYESCAN 和 FDIST2 检测到的所有受选择的 SNP, 得到中性 SNP 数据集。对中性 SNP 运行 5 次独立的运算, 来塑造中性协方差矩阵, 取其平均值作为先验分布模型用于后续分析。每次独立的运算设置 100,000 步 MCMC。接下来对于所有 SNP, 我们控制中性先验模型(即控制种群历史对遗传变异的影响)对每个气候因子运行 10 次独立的运算, 每次独立的运算设置 100,000 步 MCMC, 检测气候因子与所有 SNP 等位基因频率相关性的后验分布, 取

平均值作为最后的结果。我们选择 $\log_{10}(\text{BF}) > 0.5$ (Jeffreys, 1998) 的 SNP 作为 outlier。

LFMM 是贝叶斯主成分分析的变种，与 BayEnv 最大的不同在于通过潜在因素 (latent factors) 引入种群结构，从而使结果更加准确。我们使用 R 语言中的“LEA”函数包将储存 SNP 位点信息的 vcf 文件格式转换为 lfmm 格式，并运用函数包中的功能将 missing data 补充完整。我们运行 10 个独立的运算以模拟 SNP 等位基因频率与气候因子的相关性，迭代 100,000 步，burn in 50,000 步。在每个枝系中的模拟，设置潜在因素 $K = 1$ ，在所有种群中的运算，设置潜在因素 $K = 2$ 。计算 SNP 等位基因频率与气候因子间每个相关关系的显著性参数 p 值，将 p 值大于 0.05 即 $-\log_{10}(p\text{-value}) > 1.301$ 的 SNP 作为 outlier。

在每个枝系中，结合以上两种方法，找出由两种方法同时检测到的受选择的候选基因。比较两个枝系中，检测到的受选择的基因的异同，找到在两个枝系中特有的和共有的受选择的候选基因。通过比较这些基因的功能，可以探讨两个枝系中川滇高山栎的生态适应策略。

2.1.1.7.3 环境梯度与遗传变异的相关性分析

我们使用四种方法来量化评估环境和空间变异对遗传变异的影响：两种线性的分析方法 Mantel 检验和 RDA 分析以及两种非线性的分析方法 GF 和 GDM。这四种方法均需分别在 Tibet 分枝、HDM-WSP 分枝和所有种群这三个尺度上进行模拟，以比较不同层次、不同枝系内遗传变异受空间变异影响模式的异同。

(1) Mantel 检验

使用 R 语言的“ecodist”函数包(Goslee & Urban, 2007)进行 Mantel 检验以检验地理距离和环境距离对遗传距离的影响。首先我们 Genepop 4.5 (Rousset, 2008) 基于 SNP 数据计算川滇高山栎种群间的遗传分化 F_{ST} 矩阵，并使用 Excel 计算 $F_{ST}/(1 - F_{ST})$ 矩阵以衡量遗传距离。其次使用 R 语言的“geosphere”函数包(Hijmans, 2016)，根据采样点的经纬度数据计算种群间的地理距离矩阵。使用 R 语言的“ecodist”函数将 60 个种群的四个气候因子数据(bio03, bio09, prec01, prec06)整合为一个 Bray-Cutis 矩阵。对地理距离矩阵和遗传分化矩阵进行 Mantel 检验来计算距离隔离 (IBD, Isolation by Distance)，对气候距离矩阵和遗传分化矩阵进行 Mantel 检验来计算环境隔离 (IBE, Isolation by Environment)。同时考虑气候距离的影响以使用偏 Mantel 检验来衡量 IBD，考虑地理距离的影响以使用偏 Mantel 检验来衡量 IBE。

使用 R 语言的“ecodist”函数包进行距离矩阵多元回归分析 (MRM, Multiple Regression on distance Matrices)，相当于多元 Mantel 检验，用来检验遗传距离矩阵与气候距离矩阵之间的多元相关关系。将 60 个种群的四个气候因子数据分别计算 Bray-Cutis 距离矩阵。将五个距离矩阵(四个气候距离矩阵和一个地理距离矩阵)作为自变量，将遗传距离矩阵作为因变量进行 MRM 模拟，得到最优模型。对每个自变量矩阵进行相关性检验 (Spearman's ρ)，同时对整个模型进行方差检验。

接下来我们对四个气候因子距离矩阵和遗传距离矩阵进行 Mantel 检验和偏 Mantel 检验(考虑地理距离的影响)以衡量 IBE。

以上所有 Mantel 检验和 MRM 检验均使用置换检验(permutation test)来评估每一个分析的显著性, 置换排列数量设置为 10,000, 假设检验显著水平设置为 0.05。

(2) 冗余分析

我们使用 R 语言的“vegan”函数包(Oksanen *et al.*, 2017), 基于 SSR 和包含所有 SNP 的数据集, 进行 RDA 和偏 RDA 分析, 以检测遗传变异与气候变异的多元关系, 并比较 SSR 和 SNP 的结果, 分析中性位点和非中性位点受此多元关系影响程度的差异。在这些分析中, 将储存 SNP 数据的 vcf 格式文件经 R 语言中的“LEA”函数包(Frichot & François, 2015)转换为 Ifmm 格式, 作为 RDA 和偏 RDA 的输入文件。而 SSR 数据由 STRUCTURE 格式经“LEA”函数包转换为 Ifmm 格式, 使用 R 语言进行 PCA 后, 选择前 40 个主成分作为 RDA 和偏 RDA 的输入文件。使用上文选择的四个气候变量数据作为 RDA 和偏 RDA 的自变量。考虑到气候数据可能存在明显的空间差异性, 可能与经纬度数据存在强烈的共线性现象, 因此需要消除经纬度的影响来进行偏 RDA 分析。使用置换检验来评估 RDA 和偏 RDA 模型以及模型中的每个坐标的显著性, 置换排列数量均设置为 999, 假设检验显著水平设置为 0.05。

(3) 广义相异建模 (GDM)

GDM 是一种基于距离矩阵的非线性置换排列回归分析, 可以模拟环境距离、地理距离与遗传距离之间的非线性关系。我们使用 R 语言中的“gdm”函数包(Manion *et al.*, 2017), 基于所有 SNP, 进行 GDM 分析, 以检测川滇高山栎遗传距离与气候、地理距离之间的非线性关系。在此项分析中, 利用上一步得到的 Ifmm 格式的遗传数据文件, 计算每个 SNP 位点在每个种群的等位基因频率, 所得矩阵即为代表遗传数据的输入文件。使用上文选择的四个气候变量和经纬度数据作为代表气候、地理变异的输入数据。

(4) 梯度森林

GF 是另一种基于分箱法的非线性多元回归分析, 可以模拟环境、地理变异与遗传变异之间的非线性关系。我们使用 R 语言中的“gradientForest”函数包(Ellis *et al.*, 2012), 基于所有 SNP, 进行 GF 分析, 以检测川滇高山栎遗传变异与气候、空间梯度之间的非线性关系。在此项分析中, 利用上文得到的 Ifmm 格式的遗传数据文件, 作为代表遗传数据的输入文件, 使用上文选择的四个气候变量和经纬度数据作为代表气候、地理变异的输入数据。在不同分枝, 使用样本个体数作为分箱数目。

2.2 研究结果

2.2.1 测序及鉴定 SNP 结果

对候选基因 PCR 产物建立六个 PE250 测序文库进行高通量测序，最终得到 51,623,432 条(每个文库分别为 9,209,776 条, 8,834,784 条, 7,879,664 条, 8,836,664 条, 9,080,872 条, 7,781,672 条)测序片段。使用 GATK pipeline, 从 65 个候选基因中共得到 817 个双等位基因 SNP。使用 VCFtools 保留 MAF 大于 2.5% 的 SNP, 并过滤掉数据缺失大于 50% 的 SNP, 最终得到 381 个高质量的 SNP (来自 56 个候选基因)。其中 351 个 SNP 存在于 Tibet 分枝, 378 个 SNP 存在于 HDM-WSP 分枝。

2.2.2 遗传结构

从图 2-5 可以看出, 基于所有 SNP 的主成分分析可以将所有个体在第一主成分坐标轴上分为明显的两个部分(暖色系和冷色系), 而基于中性 SNP 和 SSR 数据的结果, 虽然也呈现分化为两个部分的趋势, 但是不明显。此外从 PCA 的结果中可以发现, 种群之间的分化极不明显。

STRUCTURE 分析也得到了类似的结果(图 2-6), 即基于所有 SNP 的模拟可以将所有个体聚类为两个分枝, 基于中性 SNP 和 SSR 的模拟同样可以将所有个体聚类为两个分枝, 但不明显。我们将两个分枝分别命名为 Tibet 分枝和 HDM-WSP 分枝, 这两个分枝在地理上分别分布于西藏东部地区和横断山-四川西部地区。

基于 PCA 和 STRUCTURE 的分组情况而进行的 AMOVA 分析支持以上的检测结果(表 2-5)。在所有种群中基于所有 SNP 的分析结果表明遗传分化主要存在于种群内(81.05%, $F_{ST}=0.19$)和枝系间(13.39%, $F_{CT}=0.13$), 而枝系内种群间的分化程度最低(5.56%, $F_{SC}=0.06$); 从中性 SNP 位点的结果中发现, 枝系之间的分化程度有所下降(7.53%, $F_{CT}=0.08$); 基于 SSR 的结果类似, 枝系间的分化程度更低(4.58%, $F_{CT}=0.05$)。此结果印证了上述基于中性 SNP 和 SSR 的 PCA 分析不能将两个枝系明显地分开以及种群间不能分开的结果。

综合上述分析, 表明可以将川滇高山栎分为两个遗传分枝, 这与 Du *et al.* (2017) 基于 SSR 遗传标记的分析结果一致。基于中性 SNP 和 SSR 的分析将川滇高山栎分为两个分枝, 但是不明显, 这表明了该物种进化动态历史(包括分布区地质变化和气候变化导致的分化, 以及普遍存在的花粉基因流对分化的抵消作用)对该物种种内分化的作用。而基于所有 SNP 数据可以将川滇高山栎分为两个明显的遗传分枝, 这表明了选择作用对川滇高山栎的遗传分化具有强烈的影响。

2.2.3 遗传多样性

基于所有 SNP 数据的遗传多样性分析结果显示, Tibet 分枝的遗传多样性低于 HDM-WSP 分枝, 但方差检验结果不显著; 基于中性 SNP 数据的结果则显示, Tibet 分枝的遗传多样性显著地低于 HDM-WSP 分枝; 基于 SSR 数据的结果则显示 Tibet 分

枝的遗传多样性也显著地低于 HDM-WSP 分枝, 且显著性更强(表 2-6)。另外基于 SSR 数据表明 Tibet 分枝的有效种群大小显著低于 HDM-WSP 分枝, 说明 Tibet 分枝经历过更加强烈的遗传漂变。

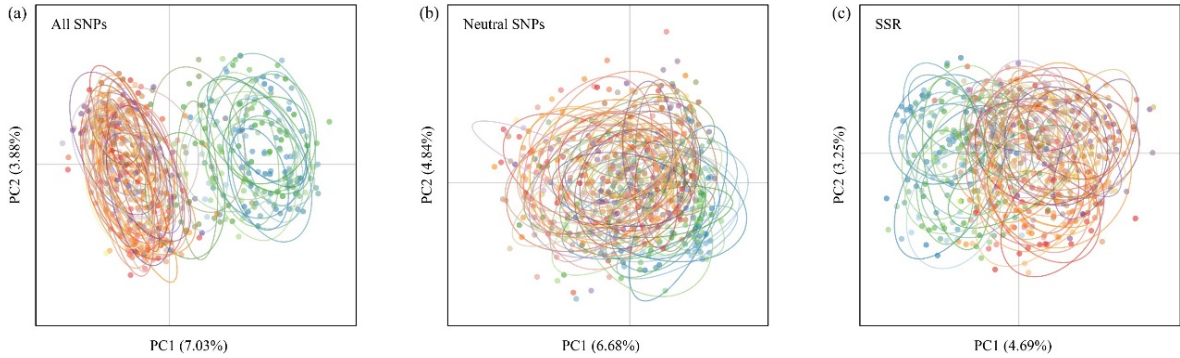


图 2-5 基于(a)所有 SNP、(b)中性 SNP 和(c) SSR 数据的主成分分析结果

Fig. 2-5 Results of PCA based on (a) all SNP dataset, (b) neutral SNP dataset and (c) SSR dataset
注: 每种颜色代表一个种群, 冷色系的点代表 Tibet 分枝的个体, 暖色系的点代表 HDM-WSP 分枝的个体

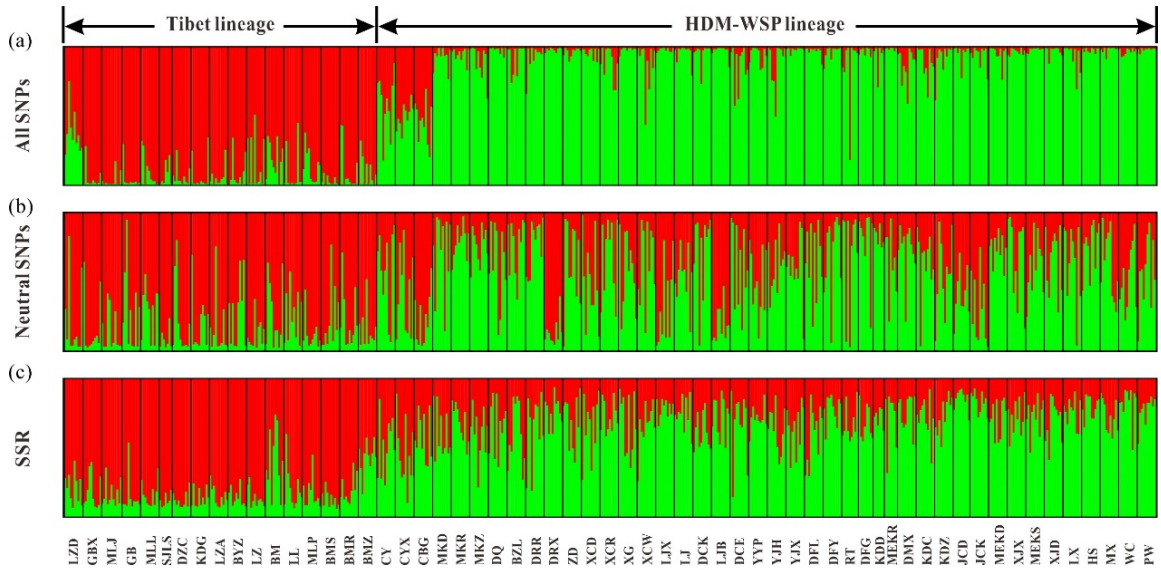


图 2-6 基于(a)所有 SNP、(b)中性 SNP 和(c) SSR 数据的 STRUCTURE 聚类结果

Fig. 2-6 Results of STRUCTURE based on (a) all SNP dataset, (b) neutral SNP dataset and (c) SSR dataset

表 2-5 基于所有 SNP、中性 SNP 和 SSR 数据的分层分子方差分析结果
 Table 2-5 Results of hierarchical AMOVA based on all SNP, neutral SNP and SSR dataset.

	All SNPs		Neutral SNPs		SSR	
	Percentage of variation (%)	Fixation Indices	Percentage of variation (%)	Fixation Indices	Percentage of variation (%)	Fixation Indices
<i>Tibet lineage</i>						
Among populations	4.62	$F_{ST} = 0.05$	5.74	$F_{ST} = 0.06$	3.41	$F_{ST} = 0.03$
Within populations	95.38		94.26		96.59	
<i>HDM-WSP lineage</i>						
Among populations	7.03	$F_{ST} = 0.07$	9.99	$F_{ST} = 0.10$	5.39	$F_{ST} = 0.05$
Within populations	92.97		90.01		94.61	
<i>All populations</i>						
Between lineages	13.39	$F_{CT} = 0.13$	7.53	$F_{CT} = 0.08$	4.58	$F_{CT} = 0.05$
Among populations within lineages	5.56	$F_{SC} = 0.06$	8.20	$F_{SC} = 0.09$	4.70	$F_{SC} = 0.05$
Within populations	81.05	$F_{ST} = 0.19$	84.27	$F_{ST} = 0.16$	90.72	$F_{ST} = 0.09$

表 2-6 基于所有 SNP、中性 SNP 和 SSR 的遗传多样性分析结果

Table 2-6 Results of genetic diversity based on all SNP, neutral SNP and SSR datasets

种群编号	所有 SNPs				中性 SNPs				SSR			
	P	H_O	H_E	π	P	H_O	H_E	π	N_E	H_O	H_E	I
<i>Tibet lineage</i>												
BM	0.86	0.11	0.19	0.20	0.91	0.05	0.12	0.13	3.98	0.70	0.68	1.46
BMR	0.86	0.11	0.20	0.21	0.91	0.06	0.12	0.13	4.06	0.73	0.67	1.40
BMS	0.87	0.10	0.18	0.19	0.92	0.05	0.12	0.12	3.01	0.72	0.61	1.18
BMZ	0.86	0.11	0.19	0.21	0.92	0.05	0.11	0.12	4.12	0.82	0.74	1.54
BYZ	0.86	0.14	0.19	0.21	0.91	0.06	0.13	0.13	2.98	0.71	0.59	1.20
DZC	0.86	0.11	0.19	0.20	0.92	0.05	0.11	0.12	3.40	0.73	0.60	1.25
GB	0.85	0.11	0.19	0.21	0.89	0.05	0.14	0.16	3.23	0.65	0.59	1.19
GBX	0.87	0.11	0.18	0.19	0.93	0.05	0.10	0.11	3.27	0.69	0.61	1.27
KDG	0.87	0.10	0.18	0.19	0.92	0.05	0.12	0.13	3.08	0.71	0.59	1.16
LL	0.87	0.09	0.18	0.20	0.91	0.07	0.13	0.15	3.84	0.75	0.65	1.36
LZ	0.87	0.09	0.18	0.19	0.91	0.05	0.12	0.13	3.52	0.79	0.67	1.31
LZA	0.87	0.09	0.18	0.19	0.92	0.04	0.12	0.13	3.04	0.66	0.62	1.20
LZD	0.86	0.11	0.19	0.20	0.91	0.05	0.12	0.13	2.72	0.65	0.59	1.15
MLJ	0.87	0.11	0.18	0.19	0.93	0.06	0.10	0.11	2.90	0.65	0.57	1.13
MLL	0.86	0.11	0.20	0.21	0.89	0.05	0.16	0.17	3.37	0.74	0.61	1.27
MLP	0.86	0.11	0.19	0.21	0.91	0.06	0.13	0.13	3.88	0.69	0.64	1.33
SJLS	0.86	0.12	0.18	0.20	0.92	0.06	0.11	0.12	2.97	0.64	0.57	1.12
Mean	0.86	0.11	0.19	0.20	0.91	0.05	0.12	0.13	3.37	0.71	0.62	1.27
<i>HDM-WSP lineage</i>												
BZL	0.87	0.10	0.18	0.19	0.91	0.05	0.13	0.14	4.42	0.70	0.72	1.59
CBG	0.87	0.09	0.18	0.19	0.93	0.05	0.10	0.11	4.23	0.64	0.62	1.36
CY	0.88	0.08	0.17	0.19	0.91	0.03	0.12	0.13	3.93	0.71	0.65	1.44

CYX	0.87	0.11	0.18	0.19	0.91	0.05	0.13	0.14	3.40	0.61	0.61	1.28
DCE	0.86	0.12	0.19	0.21	0.89	0.07	0.15	0.15	5.24	0.71	0.76	1.72
DCK	0.86	0.11	0.19	0.21	0.90	0.05	0.15	0.16	4.65	0.78	0.74	1.61
DFG	0.87	0.11	0.18	0.20	0.89	0.05	0.15	0.17	4.14	0.78	0.70	1.55
DFL	0.85	0.11	0.21	0.22	0.88	0.05	0.16	0.18	4.26	0.73	0.74	1.63
DFY	0.87	0.14	0.19	0.20	0.91	0.07	0.13	0.14	3.92	0.61	0.67	1.45
DMX	0.85	0.13	0.21	0.22	0.88	0.07	0.17	0.18	4.66	0.76	0.73	1.58
DQ	0.86	0.12	0.19	0.20	0.90	0.06	0.13	0.14	4.62	0.76	0.75	1.67
DRR	0.85	0.14	0.21	0.22	0.90	0.07	0.15	0.16	5.24	0.74	0.78	1.78
DRX	0.89	0.11	0.16	0.17	0.93	0.06	0.10	0.10	3.83	0.73	0.72	1.55
HS	0.86	0.12	0.19	0.20	0.90	0.09	0.14	0.15	4.43	0.83	0.75	1.59
JCD	0.88	0.12	0.17	0.19	0.92	0.06	0.11	0.12	4.77	0.79	0.77	1.69
JCK	0.87	0.11	0.17	0.18	0.92	0.05	0.11	0.12	4.82	0.89	0.78	1.74
KDC	0.86	0.14	0.20	0.21	0.90	0.06	0.14	0.15	4.77	0.73	0.72	1.62
KDD	0.86	0.12	0.19	0.21	0.89	0.06	0.16	0.17	3.95	0.75	0.69	1.41
KDZ	0.87	0.10	0.18	0.19	0.92	0.06	0.12	0.13	5.54	0.80	0.79	1.80
LJ	0.88	0.10	0.17	0.18	0.92	0.06	0.12	0.13	4.69	0.78	0.78	1.69
LJB	0.89	0.08	0.16	0.17	0.94	0.02	0.09	0.10	3.97	0.75	0.73	1.60
LJX	0.88	0.08	0.16	0.18	0.93	0.04	0.10	0.11	3.63	0.75	0.67	1.37
LX	0.85	0.14	0.21	0.22	0.91	0.06	0.14	0.14	4.33	0.70	0.74	1.61
MEKD	0.85	0.12	0.20	0.22	0.88	0.06	0.17	0.18	4.91	0.76	0.75	1.68
MEKR	0.86	0.12	0.19	0.21	0.88	0.06	0.16	0.18	4.06	0.70	0.69	1.51
MEKS	0.86	0.12	0.18	0.19	0.90	0.07	0.14	0.15	5.25	0.80	0.75	1.75
MKD	0.86	0.12	0.19	0.21	0.90	0.06	0.14	0.15	4.92	0.72	0.73	1.68
MKR	0.85	0.11	0.20	0.21	0.88	0.06	0.16	0.17	5.26	0.76	0.74	1.70
MKZ	0.86	0.11	0.20	0.21	0.88	0.06	0.17	0.19	4.43	0.78	0.71	1.60
MX	0.86	0.11	0.19	0.20	0.88	0.08	0.16	0.17	4.58	0.78	0.73	1.58

PW	0.86	0.12	0.19	0.20	0.90	0.07	0.15	0.16	4.94	0.71	0.77	1.72
RT	0.85	0.14	0.21	0.22	0.88	0.07	0.16	0.18	4.63	0.68	0.72	1.64
WC	0.87	0.12	0.18	0.19	0.91	0.06	0.12	0.13	5.07	0.66	0.75	1.67
XCD	0.86	0.11	0.20	0.21	0.88	0.06	0.19	0.20	5.09	0.86	0.79	1.78
XCR	0.85	0.11	0.20	0.21	0.88	0.07	0.17	0.18	5.59	0.81	0.77	1.75
XCW	0.86	0.11	0.19	0.20	0.89	0.06	0.15	0.16	5.00	0.78	0.76	1.68
XG	0.87	0.11	0.19	0.20	0.90	0.07	0.15	0.16	4.10	0.70	0.71	1.56
XJD	0.87	0.12	0.18	0.20	0.90	0.08	0.14	0.15	4.20	0.75	0.74	1.59
XJX	0.86	0.13	0.20	0.21	0.88	0.09	0.16	0.17	4.64	0.66	0.71	1.61
YJH	0.85	0.16	0.20	0.21	0.90	0.08	0.15	0.16	5.16	0.88	0.78	1.77
YJX	0.87	0.12	0.18	0.19	0.90	0.07	0.14	0.15	5.59	0.81	0.77	1.81
YYP	0.87	0.14	0.19	0.20	0.90	0.08	0.15	0.16	4.35	0.78	0.75	1.61
ZD	0.87	0.12	0.19	0.20	0.91	0.06	0.14	0.15	4.17	0.81	0.74	1.56
Mean	0.86	0.12	0.19	0.20	0.90	0.06	0.14	0.15	4.59	0.75	0.73	1.62
<i>F</i> -value	0.01	1.74	0.36	0.13	10.86	6.84	11.67	10.47	67.83	5.86	77.19	101.70
<i>P</i> -value	0.946	0.192	0.553	0.722	0.002	0.011	0.001	0.002	< 0.001	0.019	< 0.001	< 0.001
Mean Total	0.86	0.11	0.19	0.20	0.90	0.06	0.14	0.15	4.23	0.74	0.70	1.51

注：*P*：每个位点频率最高的等位基因的平均频率；*H_O*：观测杂合度；*H_E*：期望杂合度； π ：核苷酸多态性；*N_E*：有效种群大小；*I*：香农指数

2.2.4 异常 SNP 检测结果

使用基于 F_{ST} 的异常 SNP 检测方法 BAYESCAN 分别在 Tibet 分枝、HDM-WSP 分枝和所有种群中分别检测到 59、129、和 143 个异常 SNP，分别属于 25、41 和 49 个候选基因，共 208 个异常 SNP，属于 53 个候选基因。FDIST2 分别在 Tibet 分枝、HDM-WSP 分枝和所有种群中分别检测到 70、151、和 172 个异常 SNP，分别属于 28、47 和 51 个候选基因，共 227 个异常 SNP，属于 54 个候选基因(图 2-7; 表 2-6)。这两种方法一共检测到 252 个异常 SNP，属于 54 个候选基因(其余 129 个 SNP 被认为中性 SNP，用于遗传结构和遗传多样性分析)。从图 2-8 可以看出，有 20 个候选基因是这两种方法在 Tibet 分枝共同检测到的非中性基因，有 39 个基因是这两种方法在 HDM-WSP 分枝共同检测到的非中性基因。其中有 1 个非中性基因是在 Tibet 分枝中单独检测到的，14 个非中性基因是在 HDM-WSP 分枝中单独检测到的，16 个非中性基因是在两个分枝中同时检测到的(具体见表 2-7)。

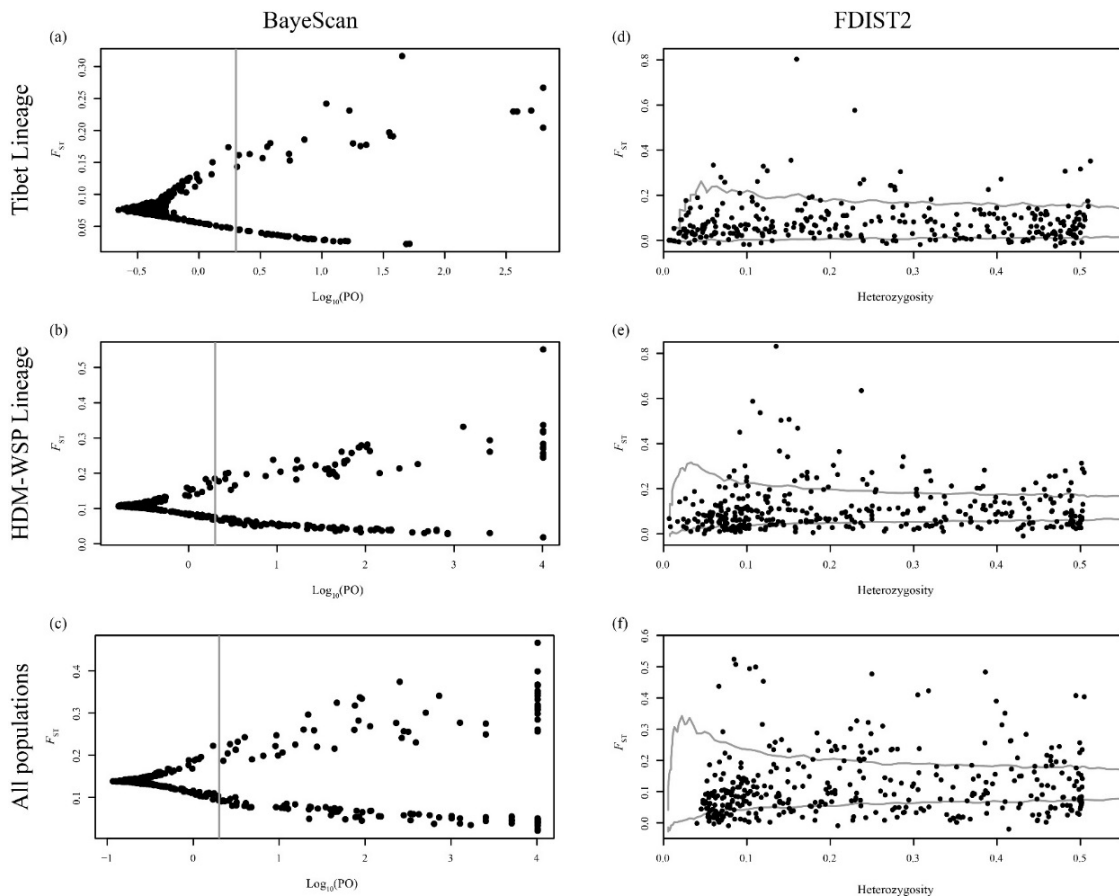


图 2-7 使用(a, b, c) BAYESCAN 和 (d, e, f) FDIST2，基于(a, d) Tibet 分枝、(b, e) HDM-WSP 分枝和(c, f)所有种群的异常 SNP 鉴定结果

Fig. 2-7 Results of outlier SNP identification in (a, d) Tibet lineage, (b, e) HDM-WSP lineage and (c, f) all populations by (a, b, c) BAYESCAN and (d, e, f) FDIST2

注：图 a, b, c 中，竖线右侧的点代表异常 SNP，图 d, e, f 中，两条折线外的点代表异常 SNP

表 2-6 使用四种方法得到的 Tibet 分枝、HDM-WSP 分枝和所有种群中异常 SNP 的数量
Table 2-6 Number of outlier SNPs detected in Tibet lineage, HDM-WSP lineage and all populations

	BayeScan	FDIST2	LFMM	BayEnv	Total
Tibet lineage	59 (25)	70 (28)	127 (48)	11 (9)	177 (49)
HDM-WSP lineage	129 (41)	151 (47)	143 (46)	53 (24)	244 (55)
All populations	143 (49)	172 (51)	123 (39)	31 (16)	241 (55)
Total	208 (53)	227 (54)	206 (54)	65 (28)	310 (56)

注：括号中的数字代表包含异常 SNP 的候选基因的数量。

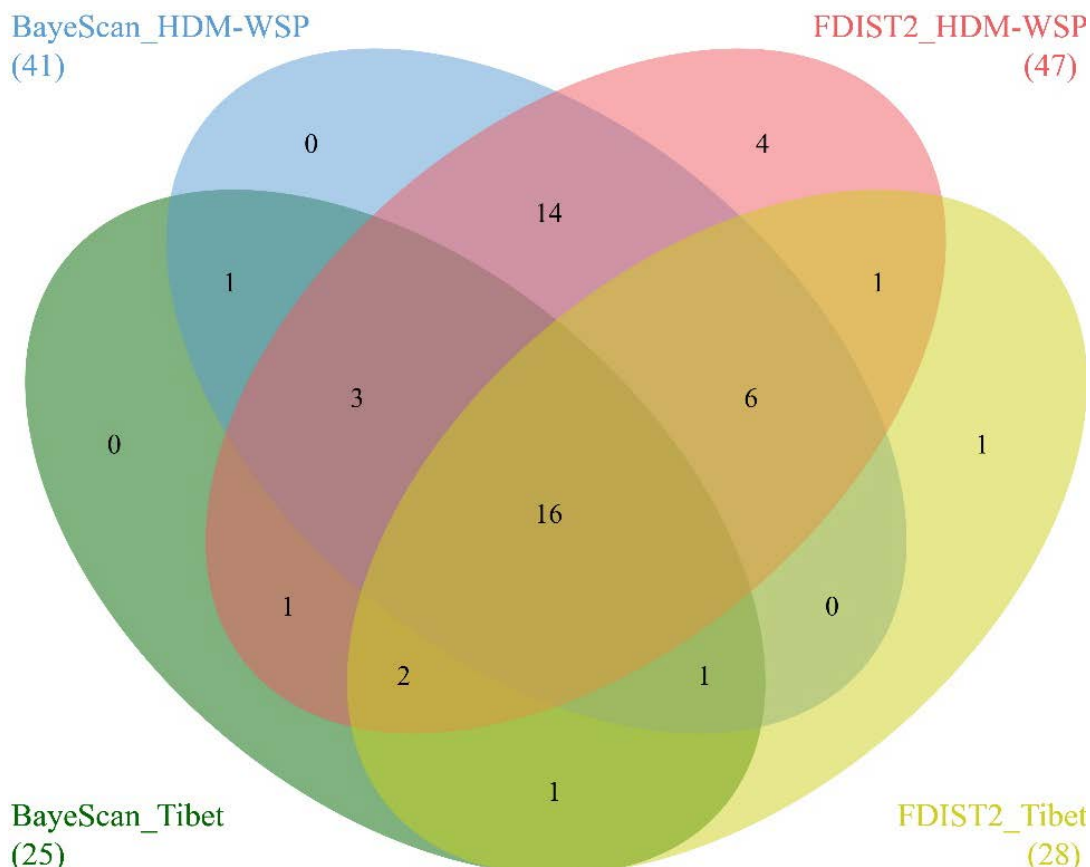


图 2-8 由 BAYESCAN 和 FDIST2 在两个分枝检测到的特有和共有的包含异常 SNP 的基因的数量

Fig. 2-8 Venn diagrams showing the number of unique and shared genes which contain outlier SNPs detected in Tibet and HDM-WSP lineage by BAYESCAN and FDIST2

使用基于遗传-环境相关性的方法 BayEnv 分别在 Tibet 分枝、HDM-WSP 分枝和所有种群中分别检测到 11、53、和 31 个异常 SNP，分别属于 9、24 和 16 个候选基因，共 65 个异常 SNP，属于 28 个候选基因(图 2-9；表 2-6)。LFMM 分别在 Tibet 分枝、HDM-WSP 分枝和所有种群中分别检测到 127、143、和 123 个异常 SNP，分别属于 48、46 和 39 个候选基因，共 206 个异常 SNP，属于 54 个候选基因(图 2-10；表 2-6)。表 2-8 列出了这两种方法分别在四个气候因素下检测到的异常 SNP 的数量。从图 2-11 可以看出，有 9 个候选基因是这两种方法在 Tibet 分枝共同检测到的非中性基

因，有 22 个基因是这两种方法在 HDM-WSP 分枝共同检测到的非中性基因。其中没有任何一个非中性基因是在 Tibet 分枝中单独检测到的，2 个非中性基因是在 HDM-WSP 分枝中单独检测到的，5 个非中性基因是在两个分枝中同时检测到的(具体见表 2-10)。

表 2-7 由 BAYESCAN 和 FDIST2 共同检测到的，Tibet 分枝和 HDM-WSP 分枝共有和特有的非中性的候选基因列表

Table 2-7 List of common and specific non-neutral gene detected from Tibet and HDM-WSP lineage by BayeScan and FDIST2

Gene	Function
<i>Common</i>	
DF_Qa020	9-cis-epoxycarotenoid dioxygenase
DF_Qa030	proton gradient regulation 5
DF_Qa035	Xylem serine proteinase 1 precursor
DF_Qa038	fiddlehead-like protein
DF_Qa051	AMP dependent CoA ligase, putative
DF_Qa077	laccase
DF_Qa079	mitotic control protein dis3
DF_Qa089	GDH2
DF_Qa104	heat shock protein 70-3
DF_Qa113	aspartic proteinase
DF_Qa129	phospholipase D alpha
DF_Qa132	stress-associated protein 1
DF_Qa146	50S ribosomal protein L7/L12
DF_Qa151	phosphoenolpyruvate carboxylase
DF_Qa172	putative transaldolase
DF_Qa180	Myb-related transcription factor
<i>Specific of HDM-WSP lineage</i>	
DF_Qa003	cytochrome P450
DF_Qa019	hexose transporter 1
DF_Qa028	aspartic proteinase
DF_Qa040	AAP6 (AMINO ACID PERMEASE 6)
DF_Qa052	glutathione S-transferase
DF_Qa055	flavonoid 3-hydroxylase
DF_Qa082	Eukaryotic initiation factor-5
DF_Qa084	white-brown-complex ABC transporter family
DF_Qa120	26S proteasome non-ATPase regulatory subunit
DF_Qa127	plasma membrane major intrinsic protein
DF_Qa137	phosphoprotein phosphatase
DF_Qa143	short chain alcohol dehydrogenase
DF_Qa150	long-chain acyl-CoA synthetase 4
DF_Qa167	R2R3 Myb30 transcription factor
<i>Specific of Tibet lineage</i>	
DF_Qa100	PREDICTED: similar to ring domain containing protein isoform 2

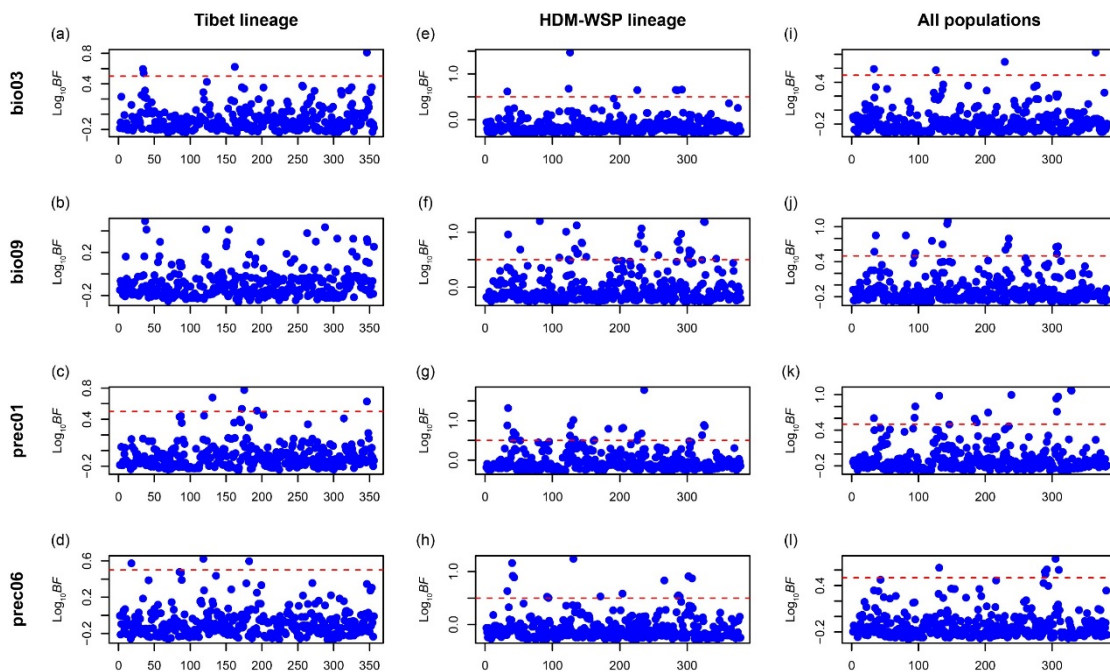


图 2-9 使用 BayEnv, 基于(a-d) Tibet 分枝、(e-h) HDM-WSP 分枝和(i-l)所有种群的异常 SNP 鉴定结果

Fig. 2-9 Results of outlier SNP identification in (a-d) Tibet lineage, (e-h) HDM-WSP lineage and (i-l) all populations by BayEnv

注: 红色虚线上方的点代表异常 SNP

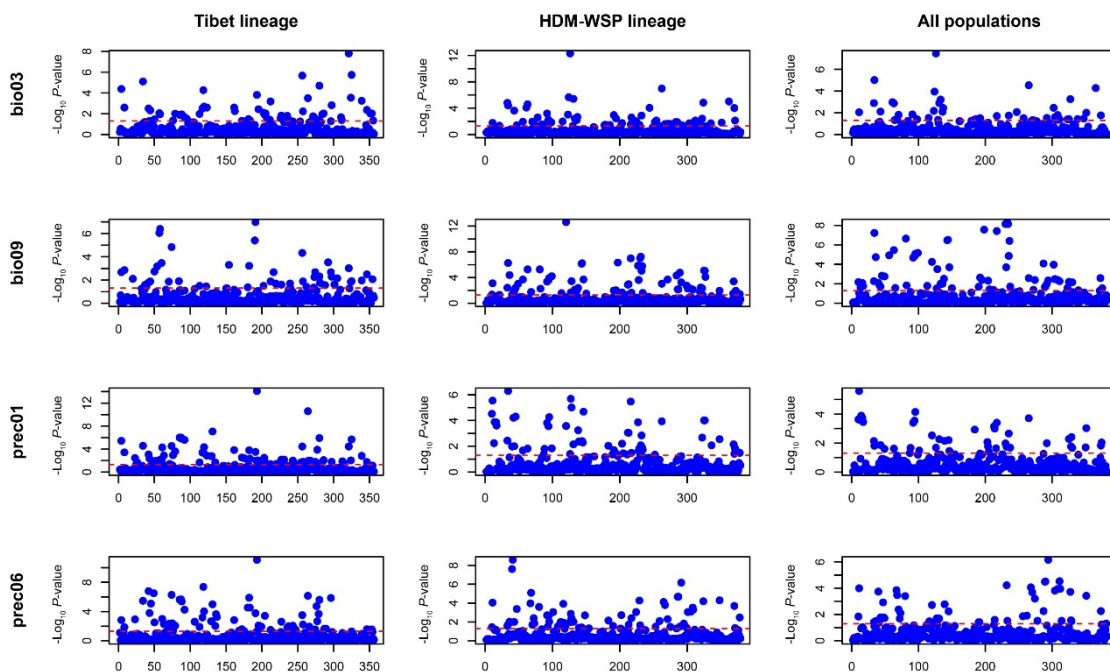


图 2-10 使用 LFMM, 基于(a-d) Tibet 分枝、(e-h) HDM-WSP 分枝和(i-l)所有种群的异常 SNP 鉴定结果

Fig. 2-10 Results of outlier SNP identification in (a-d) Tibet lineage, (e-h) HDM-WSP lineage and (i-l) all populations by LFMM

注: 红色虚线上方的点代表异常 SNP

表 2-8 使用 BayEnv 和 LFMM 得到的 Tibet 分枝、HDM-WSP 分枝和所有种群中与四个环境变量显著相关的异常 SNP 的数量

Table 2-8 Number of outlier SNPs which were significant with each of the four environmental variables in Tibet lineage, HDM-WSP lineage and all populations by BayEnv and LFMM.

	BayEnv			LFMM		
	Tibet	HDM-WSP	All populations	Tibet	HDM-WSP	All populations
bio03	4	7	4	57	63	43
bio09	0	29	14	48	67	67
prec01	5	20	13	72	62	45
prec06	3	14	6	68	65	48

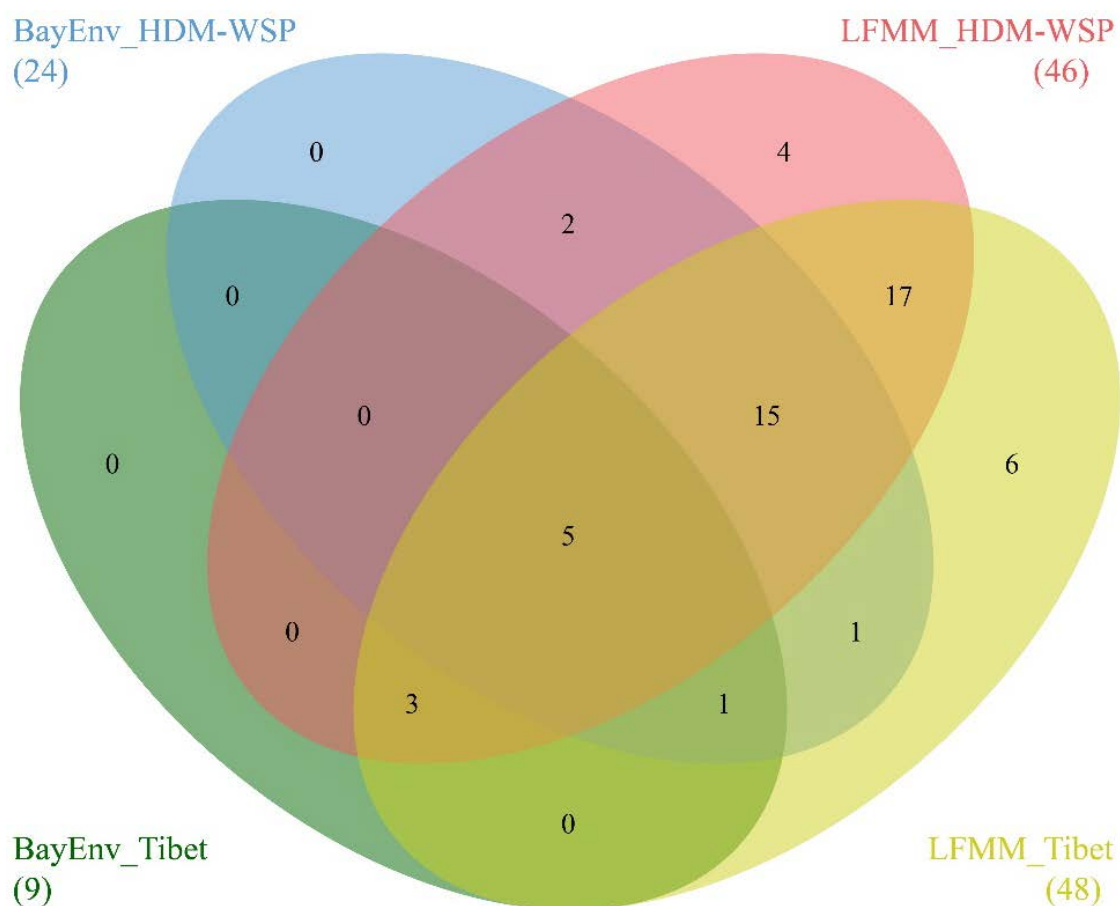


图 2-11 由 BayEnv 和 LFMM 在两个分枝检测到的特有和共有的包含异常 SNP 的基因的数量

Fig. 2-11 Venn diagrams showing the number of unique and shared genes which contain outlier SNPs detected in Tibet and HDM-WSP lineage by BayEnv and LFMM

表 2-10 由 BayEnv 和 LFMM 共同检测到的, Tibet 分枝和 HDM-WSP 分枝共有和特有的非中性的候选基因列表

Table 10 List of common and specific non-neutral gene detected from Tibet and HDM-WSP lineage by BayEnv and LFMM

Gene	Function
<i>Common</i>	
DF_Qa026	zinc finger protein
DF_Qa054	amino acid binding protein
DF_Qa057	indole-3-acetic acid 14 transcription factor like
DF_Qa079	mitotic control protein dis3
DF_Qa083	4-coumarate:coA ligase 2
<i>Specific of HDM-WSP lineage</i>	
DF_Qa064	multicopper oxidase,
DF_Qa150	long-chain acyl-CoA synthetase 4
<i>Specific of Tibet lineage</i>	
Null	Null

2.2.5 环境与遗传变异相关性分析结果

2.2.5.1 Mantel 检验

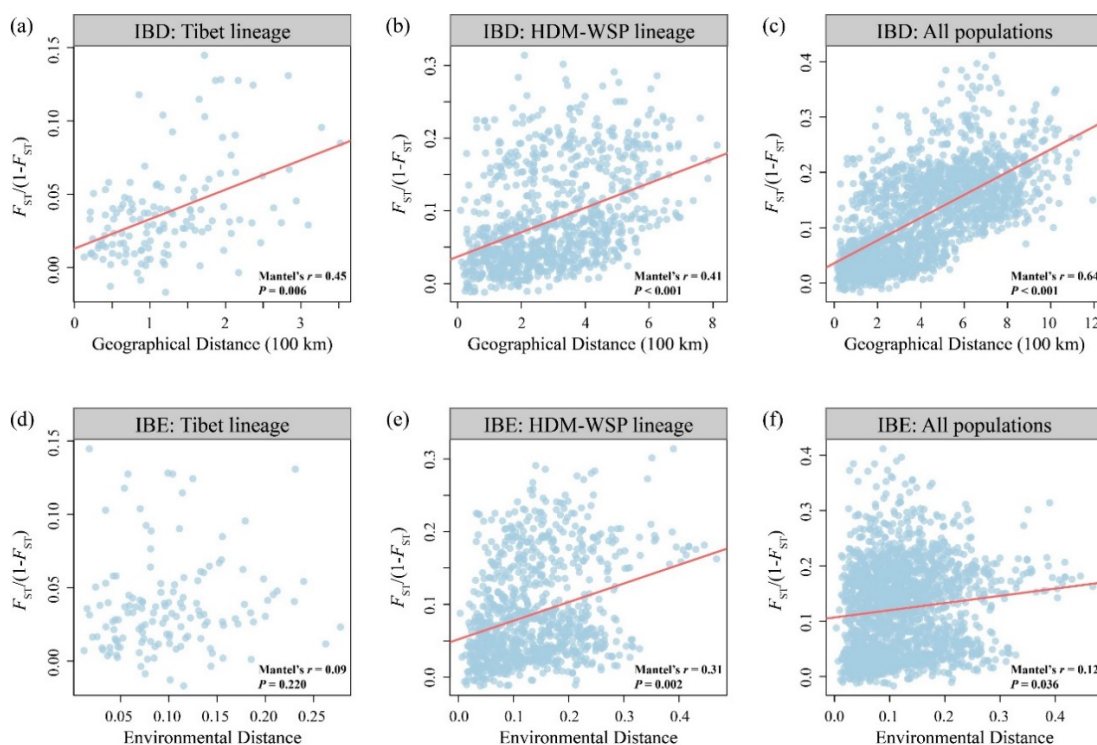


图 2-12 基于所有 SNP 数据, 在(a, d) Tibet 枝系、(b, e) HDM-WSP 枝系和(c, f)所有种群中, 对(a-c)地理距离和遗传距离、(d-f)环境距离和遗传距离进行 Mantel 检验的结果

Fig. 2-12 Results of Mantel tests between (a-c) geographic and genetic distance, (d-f) environmental and genetic distance in (a, d) Tibet lineage, (b, e) HDM-WSP lineage and (c, f) all populations based on all SNP dataset

Mantel 检验显示了在 Tibet 分枝(Mantel's $r = 0.45$, $P = 0.006$)、HDM-WSP 分枝

(Mantel's $r = 0.45$, $P < 0.001$)和所有种群(Mantel's $r = 0.64$, $P < 0.001$)中均检测到了显著的距离隔离, 并在 HDM-WSP 分枝(Mantel's $r = 0.31$, $P = 0.002$)和所有种群(Mantel's $r = 0.12$, $P = 0.036$)中检测到了显著的环境隔离, 但在 Tibet 分枝(Mantel's $r = 0.09$, $P = 0.220$)没有检测到显著的环境隔离(图 2-12)。

偏 Mantel 检验显示了类似的结果: 在 Tibet 分枝(Mantel's $r = 0.49$, $P = 0.011$)、HDM-WSP 分枝(Mantel's $r = 0.34$, $P < 0.001$)和所有种群(Mantel's $r = 0.65$, $P < 0.001$)中均检测到了显著的距离隔离(控制环境距离的影响), 并在所有种群(Mantel's $r = 0.13$, $P = 0.036$)中检测到了显著的环境隔离(控制地理距离的影响), 但在 Tibet 分枝(Mantel's $r = -0.23$, $P = 0.937$)和 HDM-WSP 分枝(Mantel's $r = 0.18$, $P = 0.051$)中没有检测到显著的环境隔离(表 2-11)。

表 2-11 在 Tibet 分枝、HDM-WSP 分枝和所有种群中, 遗传距离和地理距离之间(控制环境距离的影响)以及和环境距离之间(控制地理距离的影响)的偏 Mantel 检验结果

Table 2-11 Results of partial mantel test between genetic distance and geographical distance (controlled the effect of environmental distance) and environmental distance (controlled the effect of geographical distance) in Tibet lineage, HDM-WSP lineage and all populations

	Tibet		HDM-WSP		All populations	
	Mantel's r	P	Mantel's r	P	Mantel's r	P
IBD conditioned with environmental distance	0.49	0.011	0.34	<0.001	0.65	<0.001
IBE conditioned with geographical distance	-0.23	0.937	0.18	0.051	0.13	0.036

表 2-12 Tibet 分枝、HDM-WSP 分枝和所有种群中遗传距离与四个气候因子距离之间的 Mantel 检验和偏 Mantel 检验(控制地理矩阵的影响)结果

Table 2-12 Results of Mantel test and partial Mantel test (controlled the effect of geographical distance) between genetic distance and each of the four bioclimatic distances in Tibet lineage, HDM-WSP lineage and all populations

	Tibet Lineage		HDM-WSP Lineage		All populations	
	Mantel's r	P	Mantel's r	P	Mantel's r	P
<i>Mantel test</i>						
bio03 distance	-0.01	0.528	0.13	0.090	0.07	0.126
bio09 distance	0.04	0.304	-0.05	0.877	-0.02	0.811
prec01 distance	0.04	0.302	0.24	0.006	0.19	0.002
prec06 distance	0.13	0.190	0.22	0.018	0.07	0.147
<i>partial Mantel test</i>						
bio03 distance	-0.20	0.939	0.07	0.197	0.02	0.346
bio09 distance	0.05	0.276	-0.04	0.816	-0.03	0.899
prec01 distance	-0.18	0.908	0.16	0.041	0.22	< 0.001
prec06 distance	-0.16	0.869	0.10	0.163	0.06	0.154

对遗传距离和四个气候因子距离进行 Mantel 检验发现在 HDM-WSP 分枝中, 遗传距离与 prec01 (Mantel's $r = 0.24$, $P = 0.006$)和 prec06 (Mantel's $r = 0.22$, $P = 0.018$)显著相关; 在所有种群中, 遗传距离与 prec01 显著相关(Mantel's $r = 0.19$, $P = 0.002$); 而在 Tibet 分枝中, 没有气候因子与遗传距离显著相关。偏 Mantel 检验发现在 HDM-WSP 分枝中, 遗传距离仅与 prec01 显著相关(Mantel's $r = 0.16$, $P = 0.041$); 在所有种群中, 遗传距离同样仅与 prec01 显著相关(Mantel's $r = 0.22$, $P < 0.001$); 同样在 Tibet 分枝中, 没有气候因子与遗传距离显著相关(表 2-12)。

MRM 结果表明, 在 Tibet 分枝、HDM-WSP 分枝和所有种群中的最优模型均显著($P < 0.05$)。在 Tibet 分枝中, 遗传距离仅与地理距离显著相关(Spearman's $\rho = 0.48$, $P = 0.012$), 而与气候因子均无显著相关关系; 在 HDM-WSP 分枝中, 遗传距离与地理距离(Spearman's $\rho = 0.40$, $P < 0.001$)和 prec01 (Spearman's $\rho = 0.17$, $P = 0.022$)显著相关; 在所有种群中, 遗传距离与地理距离(Spearman's $\rho = 0.66$, $P < 0.001$)和 prec01 (Spearman's $\rho = 0.14$, $P = 0.003$)显著相关(表 2-13)。

表 2-13 Tibet 分枝、HDM-WSP 分枝和所有种群中遗传距离与环境距离之间的距离矩阵多元回归分析结果

Table 2-13 Results of MRM (Multiple Regression on distance Matrices) between genetic distance and environmental distances in Tibet lineage, HDM-WSP lineage and all populations

	Tibet Lineage		HDM-WSP Lineage		All populations	
	ρ	P	ρ	P	ρ	P
geographical distance	0.48	0.012	0.40	<0.001	0.66	<0.001
bio03 distance	-0.04	0.686	0.07	0.301	< 0.01	0.972
bio09 distance	-0.02	0.837	0.03	0.601	0.04	0.140
prec01 distance	-0.13	0.351	0.17	0.022	0.14	0.003
prec06 distance	-0.01	0.939	-0.02	0.830	-0.01	0.734
ρ^2 and P of best model	0.19	0.030	0.27	<0.001	0.46	<0.001

Mantel 检验的结果说明在川滇高山栎分布区内存在显著的地理隔离, 并且为形成遗传分化最主要的形式。在 HDM-WSP 分枝和所有种群中均存在显著的环境隔离, 为形成遗传分化的次要形式。Prec01 也就是一月降水量(可看做干旱季节的降水量)与遗传分化显著相关, 因此为影响遗传分化的最主要气候变量。

2.2.5.2 冗余分析

使用 SNP 和 SSR 数据对川滇高山栎的遗传分化进行 RDA 分析和偏 RDA 分析(控制地理因素的影响), 以评估每个气候因素和地理因素在环境塑造遗传变异的模型中起到的作用(表 2-14; 表 2-15)。每个模型均显著($P = 0.001$)。每个环境变量的特征值和解释度以及方差检验结果均展示于表 2-14 和表 2-15。表 2-16 展示了环境因素的总解释度和地理因素的解释度。

表 2-14 Tibet 分枝、HDM-WSP 分枝和所有种群中基于 SNP 的遗传变异和四个气候因子梯度之间的 RDA 和偏 RDA(控制地理因素的影响)结果总结

Table 2-14 Summary of RDA and *p*RDA between genetic variations based on SNP dataset and four climate gradients in Tibet lineage, HDM-WSP lineage and all populations

Axis	RDA				<i>p</i> RDA			
	Variance	Percentage of variation (%)	<i>F</i>	<i>P</i>	Variance	Percentage of variation (%)	<i>F</i>	<i>P</i>
<i>Tibet Lineage</i>								
bio03	1.28	1.22	2.08	0.001	1.08	1.03	1.75	0.001
bio09	1.09	1.04	1.76	0.001	1.01	0.96	1.65	0.001
prec01	0.84	0.80	1.37	0.023	0.83	0.79	1.36	0.024
prec06	1.11	1.06	1.80	0.001	0.76	0.72	1.24	0.087
whole model			1.75	0.001			1.50	0.001
<i>HDM-WSP Lineage</i>								
bio03	0.61	0.53	2.31	0.001	0.69	0.60	2.65	0.001
bio09	2.28	2.99	8.64	0.001	2.06	1.80	7.96	0.001
prec01	1.15	1.01	4.38	0.001	0.50	0.44	1.94	0.001
prec06	1.19	1.04	4.52	0.001	0.89	0.78	3.42	0.001
whole model			4.96	0.001			4.00	0.001
<i>All populations</i>								
bio03	0.47	0.39	2.33	0.001	0.48	0.39	2.57	0.001
bio09	1.39	1.15	6.94	0.001	1.40	1.16	7.48	0.001
prec01	1.39	1.15	6.95	0.001	0.48	0.40	2.55	0.001
prec06	0.73	0.61	3.64	0.001	0.77	0.64	4.12	0.001
whole model			4.97	0.001			4.18	0.001

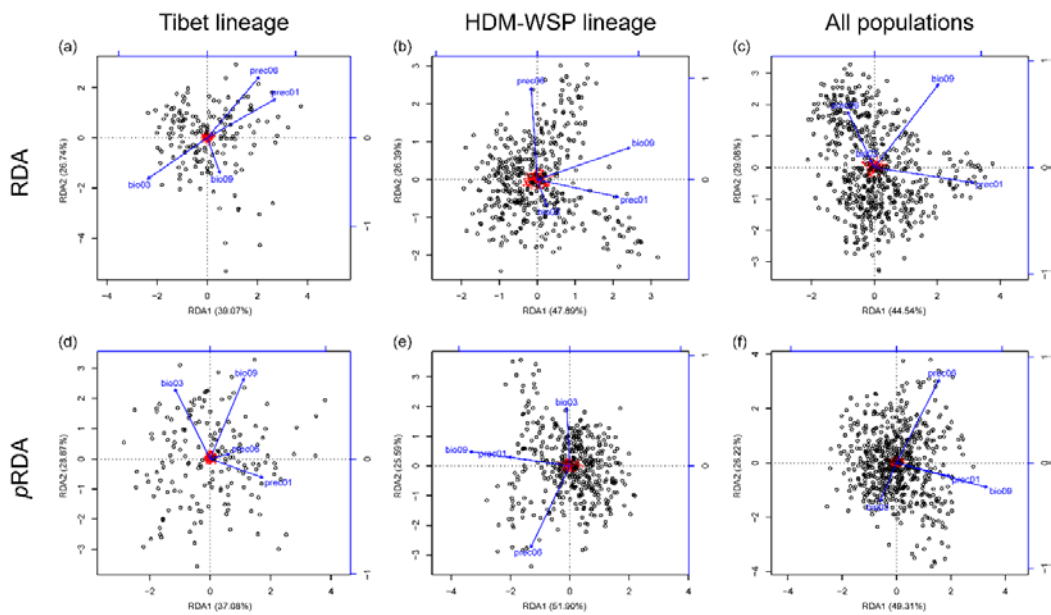


图 2-13 基于所有 SNP 数据，在(a, d) Tibet 枝系、(b, e) HDM-WSP 枝系和(c, f)所有种群中的(a-c) RDA 和(d-f) 偏 RDA 的载荷结果

Fig. 2-13 Loading results of (a-c) RDA and (d-f) *p*RDA in (a, d) Tibet lineage, (b, e) HDM-WSP lineage and (c, f) all populations based on all SNP dataset

对于 SNP 数据, 在 Tibet 分枝中 RDA 的结果显示 bio03、prec06 和 bio09 对遗传分化的影响较大, 偏 RDA 的结果则显示 bio03 和 bio09 的影响较大; 在 HDM-WSP 分枝中, RDA 的结果显示 bio09、prec01 和 prec06 的影响较大, 偏 RDA 的结果则显示 bio09 的影响较大; 在所有种群中, RDA 的结果显示 prec01 和 bio09 的影响较大, 而偏 RDA 的结果则显示 bio09 的影响较大(图 2-13; 表 2-14)。

对于 SSR 数据, 在 Tibet 分枝中 RDA 的结果显示 prec06、prec09 和 bio03 对遗传分化的影响较大, 偏 RDA 的结果则显示 bio09 的影响较大; 在 HDM-WSP 分枝中, RDA 的结果显示 prec06、prec01 和 bio09 的影响较大, 偏 RDA 的结果则显示四种气候因子的影响差别不大; 在所有种群中, RDA 的结果显示 prec01 的影响较大, 而偏 RDA 的结果则显示 prec01、prec06 和 bio9 的影响较大(图 2-14; 表 2-15)。

表 2-15 Tibet 分枝、HDM-WSP 分枝和所有种群中基于 SSR 的遗传变异和四个气候因子梯度之间的 RDA 和偏 RDA(控制地理因素的影响)结果总结

Table 2-15 Summary of RDA and *p*RDA between genetic variations based on SSR dataset and four climate gradients in Tibet lineage, HDM-WSP lineage and all populations

Axis	RDA				<i>p</i> RDA			
	Variance	Percentage of variation (%)	<i>F</i>	<i>P</i>	Variance	Percentage of variation (%)	<i>F</i>	<i>P</i>
<i>Tibet Lineage</i>								
bio03	0.18	1.76	3.06	0.001	0.14	1.33	2.35	0.001
bio09	0.20	1.90	3.30	0.001	0.18	1.77	3.14	0.001
prec01	0.07	0.71	1.24	0.177	0.12	1.19	2.12	0.002
prec06	0.21	1.99	3.46	0.001	0.09	0.88	1.57	0.032
whole model			2.76	0.001			2.30	0.001
<i>HDM-WSP Lineage</i>								
bio03	0.07	0.50	2.15	0.001	0.09	0.66	2.89	0.001
bio09	0.12	0.84	3.59	0.001	0.10	0.68	2.99	0.001
prec01	0.13	0.95	4.07	0.001	0.10	0.69	3.03	0.001
prec06	0.18	1.31	5.64	0.001	0.13	0.95	4.13	0.001
whole model			3.86	0.001			3.26	0.001
<i>All populations</i>								
bio03	0.07	0.49	2.94	0.001	0.05	0.39	2.44	0.001
bio09	0.09	0.64	3.85	0.001	0.08	0.61	3.83	0.001
prec01	0.14	1.05	6.28	0.001	0.11	0.79	4.92	0.001
prec06	0.10	0.78	4.67	0.001	0.09	0.67	4.18	0.001
whole model			4.44	0.001			3.84	0.001

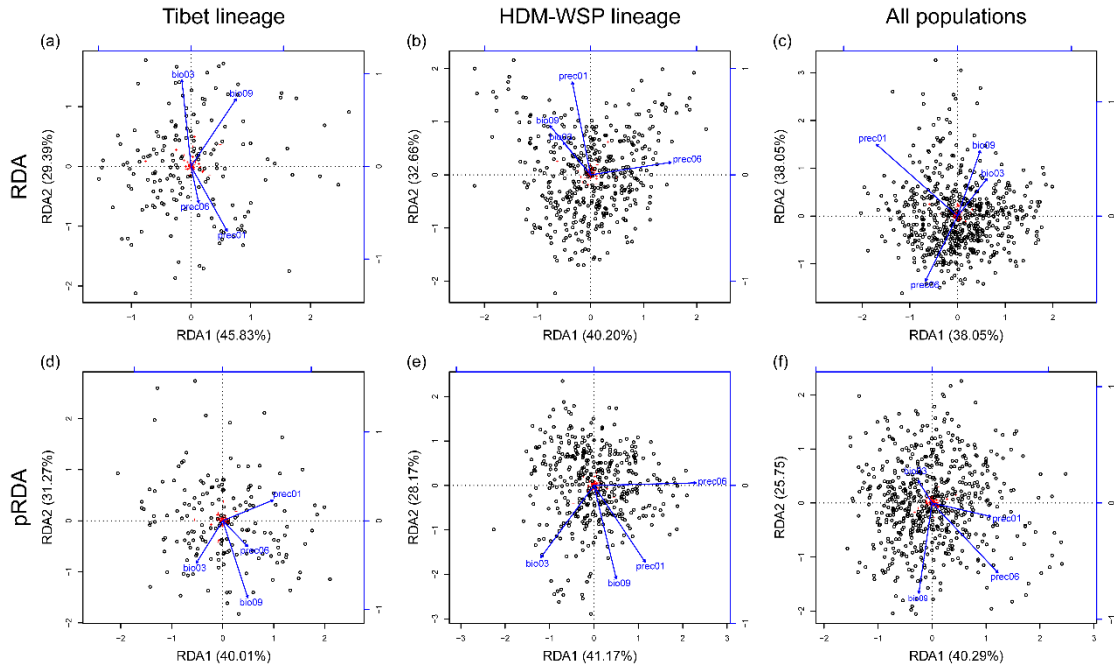


图 2-14 基于 SSR 数据，在(a, d) Tibet 枝系、(b, e) HDM-WSP 枝系和(c, f)所有种群中的(a-c) RDA 和(d-f) 偏 RDA 的载荷结果

Fig. 2-14 Loading results of (a-c) RDA and (d-f) *p*RDA in (a, d) Tibet lineage, (b, e) HDM-WSP lineage and (c, f) all populations based on SSR dataset

表 2-16 Tibet 分枝、HDM-WSP 分枝和所有种群中，RDA 和偏 RDA 得到的环境变量的特征值和解释度

Table 2-16 Eigenvalue and PVE (percentage variance explained) of environmental variables of RDA and *p*RDA in Tibet lineage, HDM-WSP lineage and all populations

	RDA		<i>p</i> RDA			
	Constrained		Conditioned		Constrained	
	Eigenvalue	PVE(%)	Eigenvalue	PVE(%)	Eigenvalue	PVE(%)
SNP						
Tibet	4.33	4.12	2.79	2.66	3.68	3.50
HDM-WSP	5.23	4.57	3.38	2.95	4.14	3.62
All populations	3.97	3.30	8.97	7.46	3.12	2.59
SSR						
Tibet	0.66	6.35	0.42	4.06	0.54	5.18
HDM-WSP	0.51	3.60	0.38	2.72	0.42	2.98
All populations	0.40	2.96	0.67	4.94	0.33	2.45

结合表 2-14、表 2-15、表 2-16，我们发现无论是基于 SNP 数据(其中大多数受到自然选择)还是基于 SSR 数据的分析，地理因素始终是对遗传分化影响最大的环境因子。在控制地理因素的影响后，偏 RDA 的结果与 RDA 有较大的差别，间接反映了地理因素对遗传分化有较大的影响。相对于较小的尺度(Tibet 分枝和 HDM-WSP 分枝)，在大尺度(所有种群)中，地理因素的影响更大。另外基于 SNP 的分析结果表明

bio09 是对遗传分化影响最大的气候因子，而基于 SSR 的结果则表明，四种气候因子均对遗传分化有显著的影响，尤其是 bio9、prec01 和 prec06。

2.2.5.3 梯度森林模型

梯度森林模型的结果显示，地理因素(经度、纬度)是构建模型的最重要环境变量。而四种气候因子的影响差异不大，但 prec06 无论是在 Tibet 分枝、HDM-WSP 分枝，还是在所有种群中，都是重要性最大的气候因子(图 2-15；表 2-17)。

图 2-16、图 2-17 和图 2-18 分别展示了 Tibet 分枝、HDM-WSP 分枝和所有种群中遗传变异沿环境梯度的累积重要性。从这三张图中可以发现，两个分枝和所有种群中，经度梯度在东经 96 度至 98 度之间的遗传组成均变化剧烈。而从采样图和采样表中可以看出，此经度范围恰好为 Tibet 分枝和 HDM-WSP 分枝的分界线，这从另一个角度与遗传结构的划分相互印证。另外，prec06 (六月降水量)在 160 至 180 毫米之间以及 200 毫米至 220 毫米之间时，遗传组成均变化剧烈，而其他的环境因子对遗传变异的影响相对平滑。

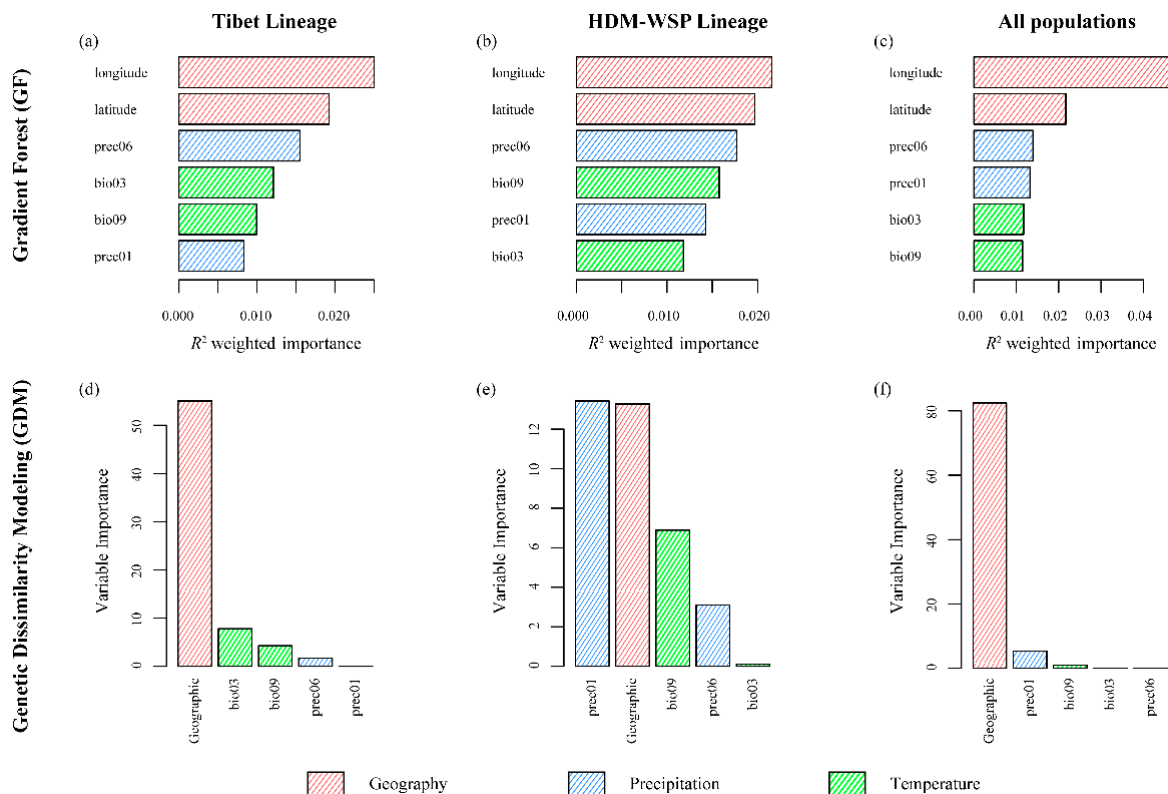


图 2-15 基于 SNP 数据，在(a, d) Tibet 枝系、(b, e) HDM-WSP 枝系和(c, f)所有种群中，对遗传变异和环境梯度进行(a, b, c) GF 和(d, e, f) GDM 模拟的结果

Fig. 2-15 Results of GF (a-c) and (d-f) GDM modeling between genetic variants and environmental gradients in (a, d) Tibet lineage, (b, e) HDM-WSP lineage and (c, f) all populations based on SNP dataset

表 2-17 基于所有 SNP 数据，在 Tibet 分枝、HDM-WSP 分枝和所有种群中，对遗传变异和环境梯度进行 GF 建模的每个环境因子的精度重要性

Table 2-17 Accuracy Importance of each environmental variables of GF modeling between genetic variants and environmental gradients in Tibet lineage, HDM-WSP lineage and all populations based on SNP dataset

Variables	Tibet lineage	HDM-WSP lineage	All populations
latitude	0.02	0.04	0.02
longitude	0.02	0.04	0.02
bio03	0.01	0.02	0.01
bio09	0.01	0.04	0.02
prec01	0.01	0.03	0.01
prec06	0.02	0.03	0.02

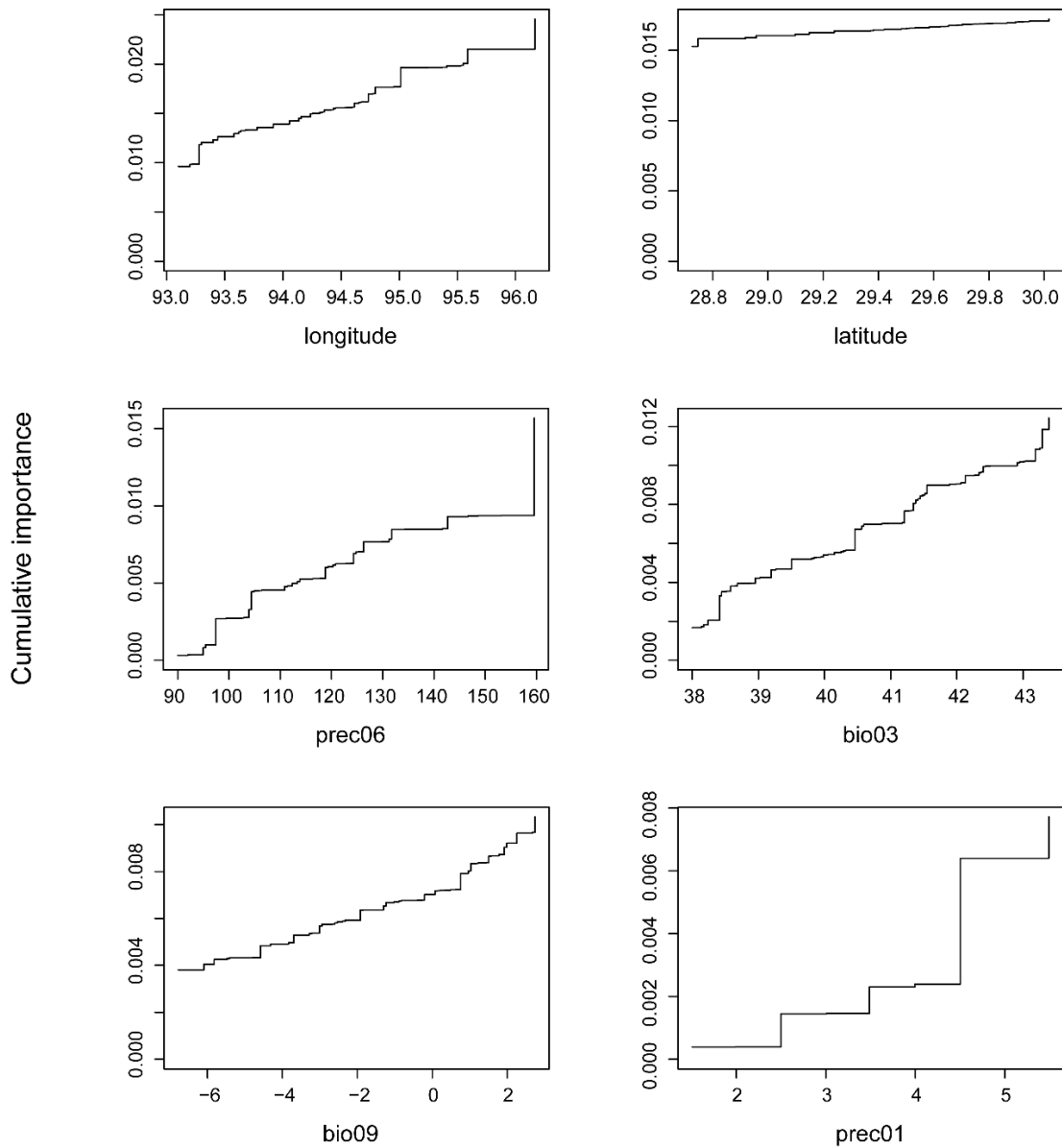


图 2-16 Tibet 分枝遗传变异沿环境梯度的累积重要性

Fig. 2-16 Cumulative importance of genetic variation along environmental gradients in Tibet lineage

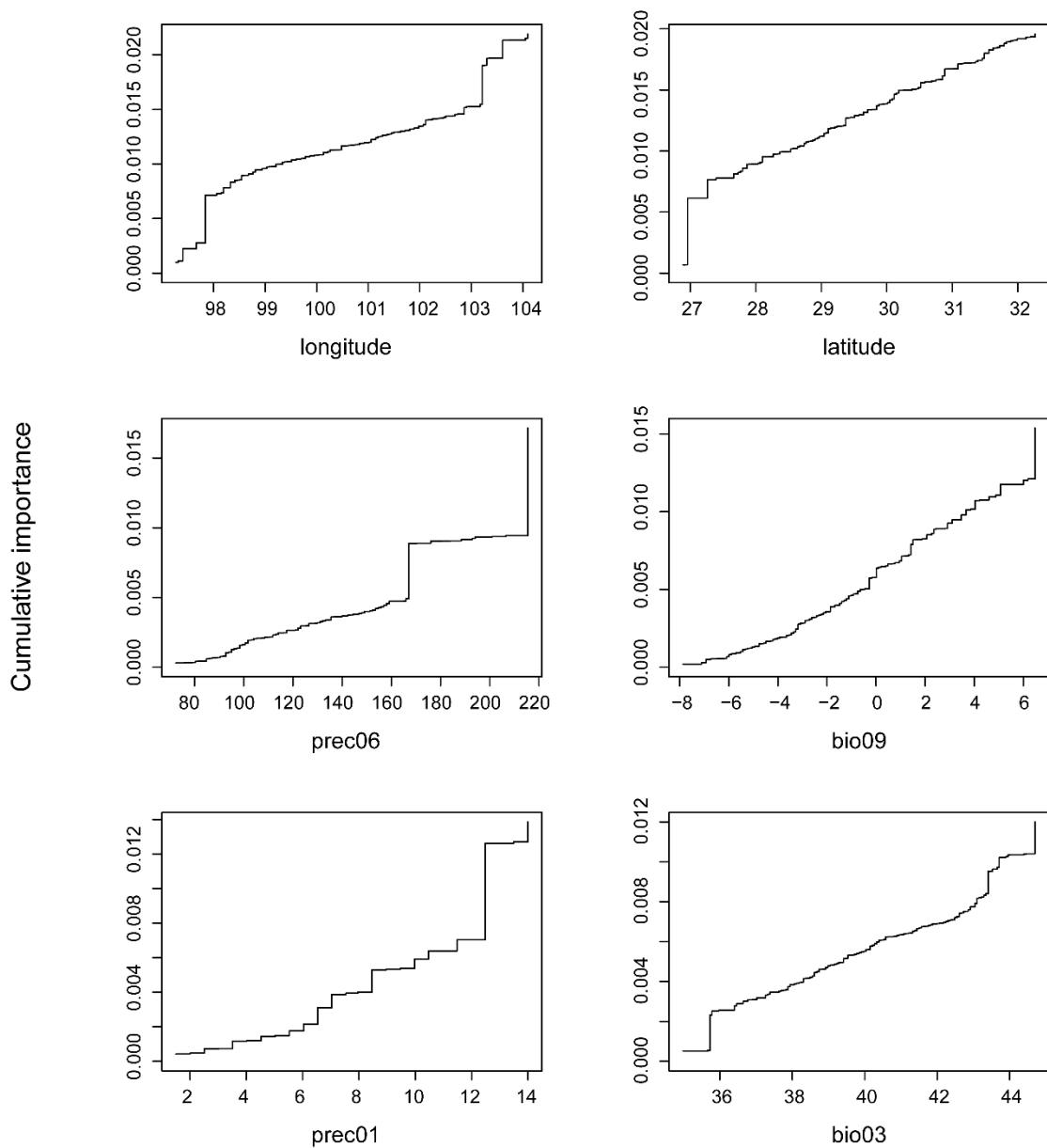


图 2-17 HDM-WSP 分枝遗传变异沿环境梯度的累积重要性

Fig. 2-17 Cumulative importance of genetic variation along environmental gradients in HDM-WSP lineage

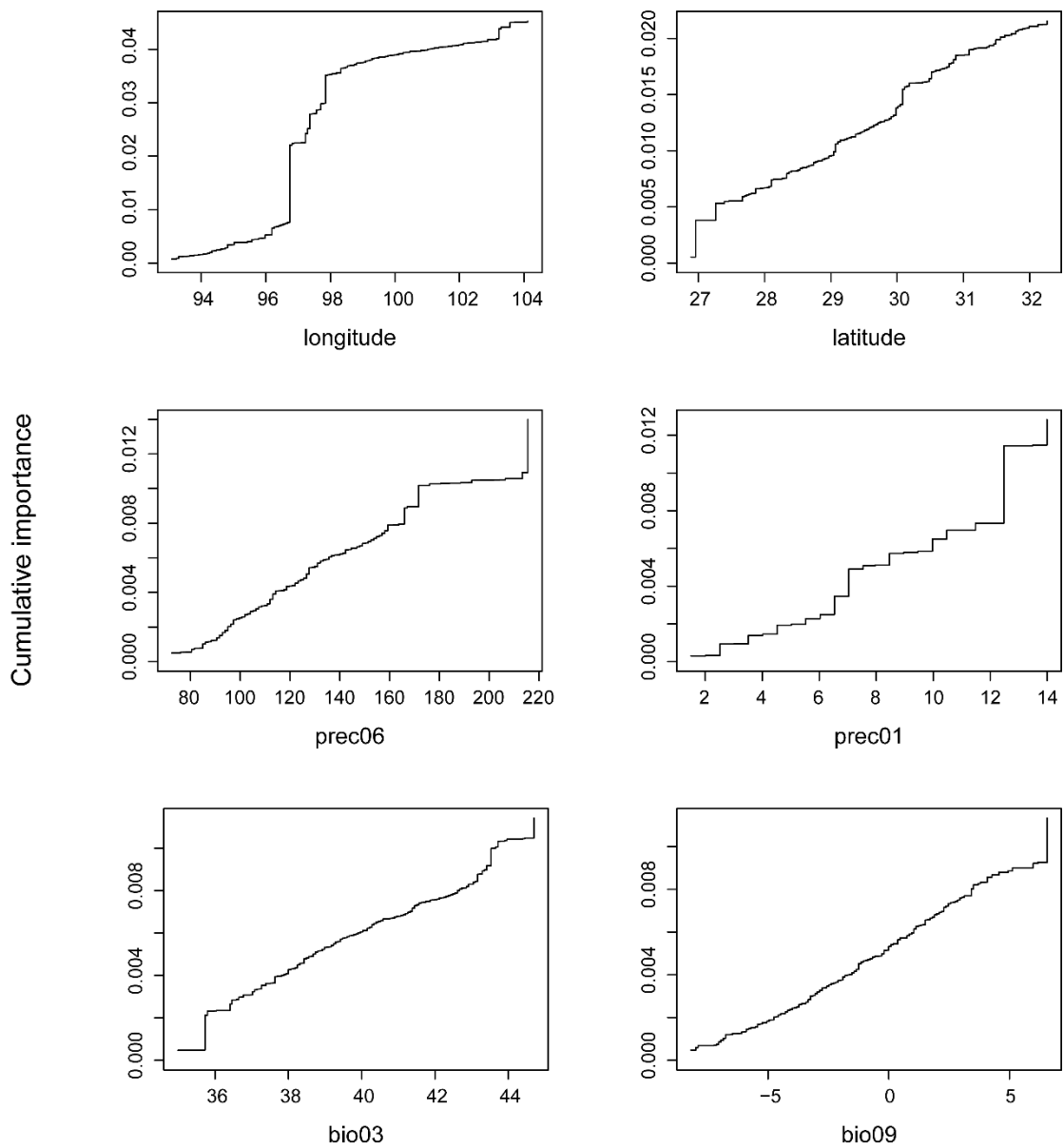


图 2-18 所有种群中遗传变异沿环境梯度的累积重要性

Fig. 2-18 Cumulative importance of genetic variation along environmental gradients in all populations

2.2.5.4 广义相异建模

广义相异建模的结果显示，地理距离是构建模型的最重要环境距离矩阵(图 2-15；表 2-18)。在 Tibet 分枝和所有种群中，距离矩阵的重要程度分别为 55.05% 和 82.43%，远远超过其他环境矩阵的重要性。所有种群中，prec01 的重要性虽然远远小于地理距离，但相对于其他三个气候因子，却处于较高的位置(5.34%)。而在 HDM-WSP 种群中，地理矩阵的重要性依然很高(13.28%)，但是其他环境矩阵也有很高的重要性(如 prec01 为 13.42%，bio9 为 6.89%)，prec01 甚至超过了地理矩阵，成为最重要的环境矩阵变量。因此广义线性模型的结果与同样基于距离矩阵的 Mantel 检验的结果非常相似，即地理距离是形成遗传分化最主要的形式，prec01 在 HDM-WSP 分枝和所有

种群中为影响遗传分化的最主要气候变量。图 2-19、图 2-20 和图 2-21 分别展示了 Tibet 分枝、HDM-WSP 分枝和所有种群中遗传距离沿环境距离梯度变化的 I-样条曲线。由于在 Tibet 分枝中，prec01 的重要性为零，因此在图 2-19 中缺少 prec01 的 I-样条曲线图。

表 2-18 基于所有 SNP 数据，在 Tibet 分枝、HDM-WSP 分枝和所有种群中，对遗传变异和环境梯度进行 GDM 建模的每个环境因子的变量重要性(%)

Table 2-18 Variable Importance of each environmental variables of GDM modeling between genetic variants and environmental gradients in Tibet lineage, HDM-WSP lineage and all populations based on SNP dataset (%)

Variables	Tibet lineage	HDM-WSP lineage	All populations
Geographic	55.05	13.28	82.43
bio03	7.81	0.11	0.11
bio09	4.27	6.89	0.98
prec01	0.00	13.42	5.34
prec06	1.71	3.11	0.04

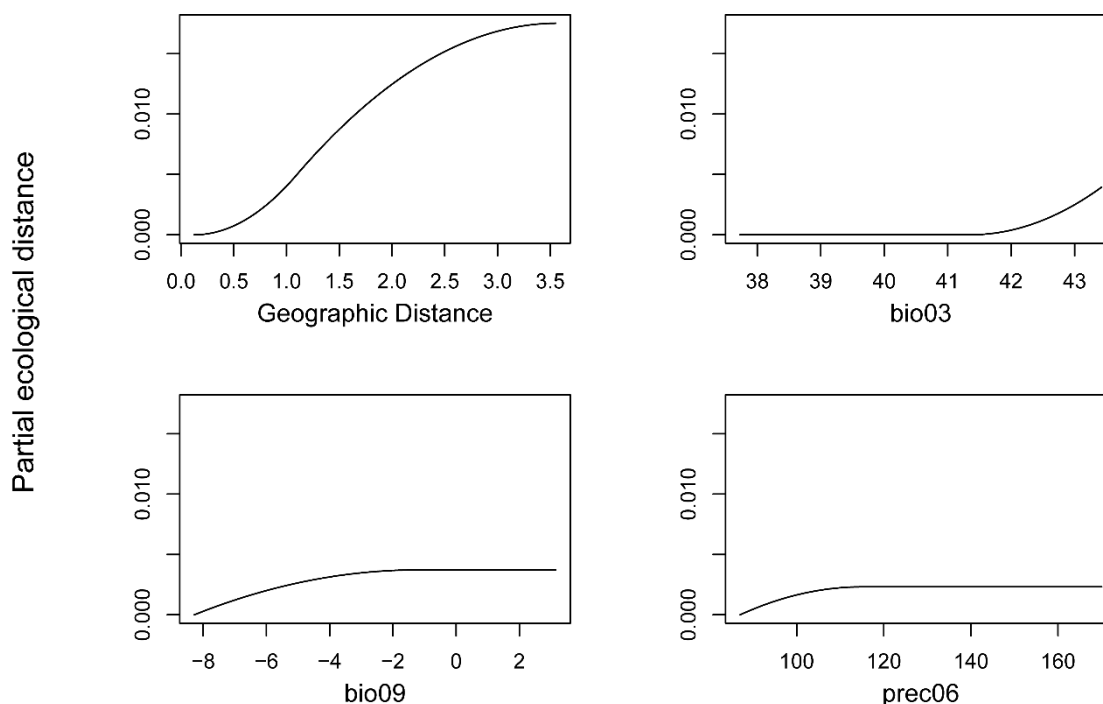


图 2-19 Tibet 分枝遗传距离沿环境距离梯度变化的 I-样条曲线

Fig. 2-19 I-splines showing changes genetic distance along environmental distance gradients in Tibet lineage

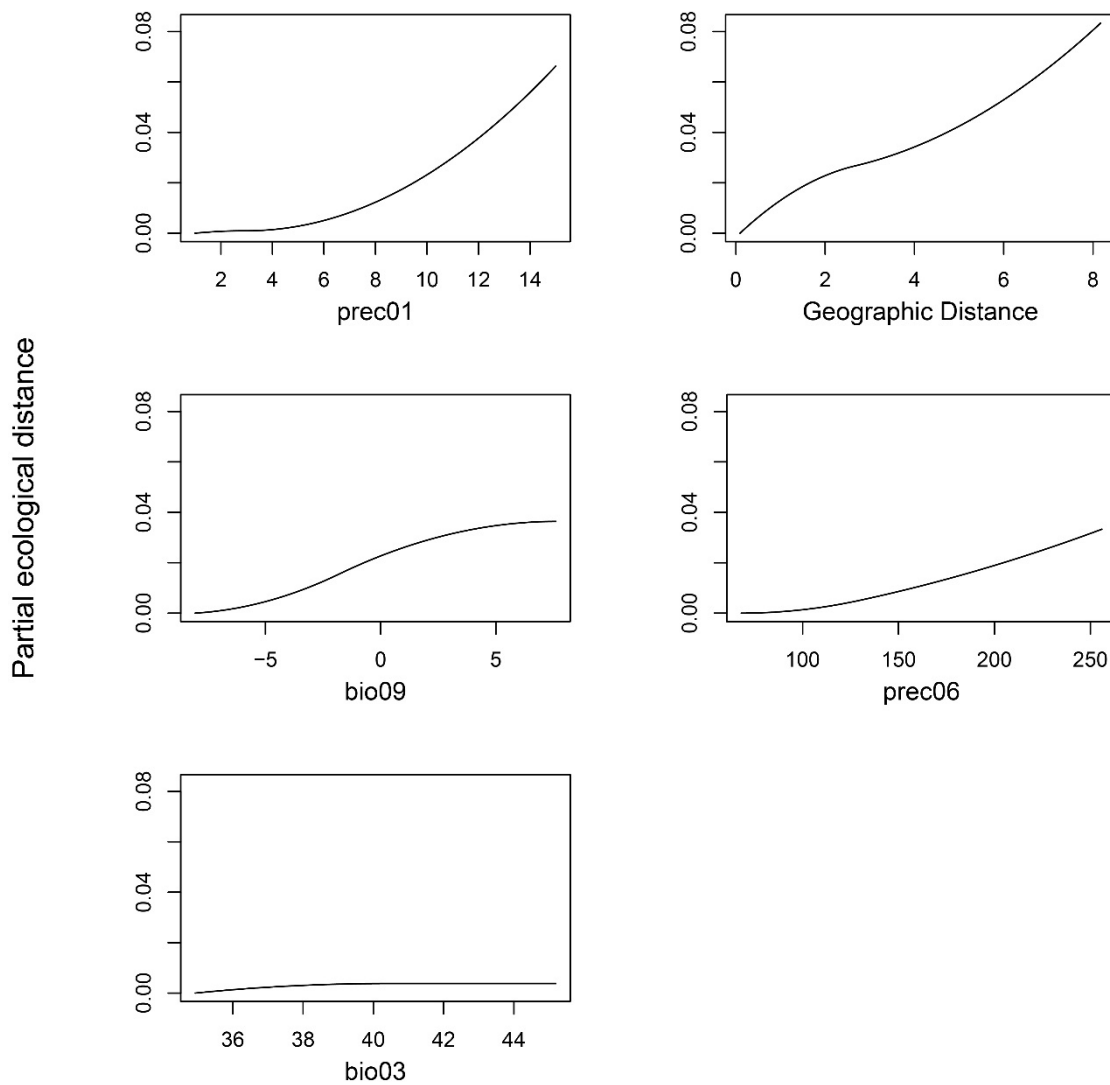


图 2-20 HDM-WSP 分枝遗传距离沿环境距离梯度变化的 I-样条曲线

Fig. 2-20 I-splines showing changes genetic distance along environmental distance gradients in HDM-WSP lineage

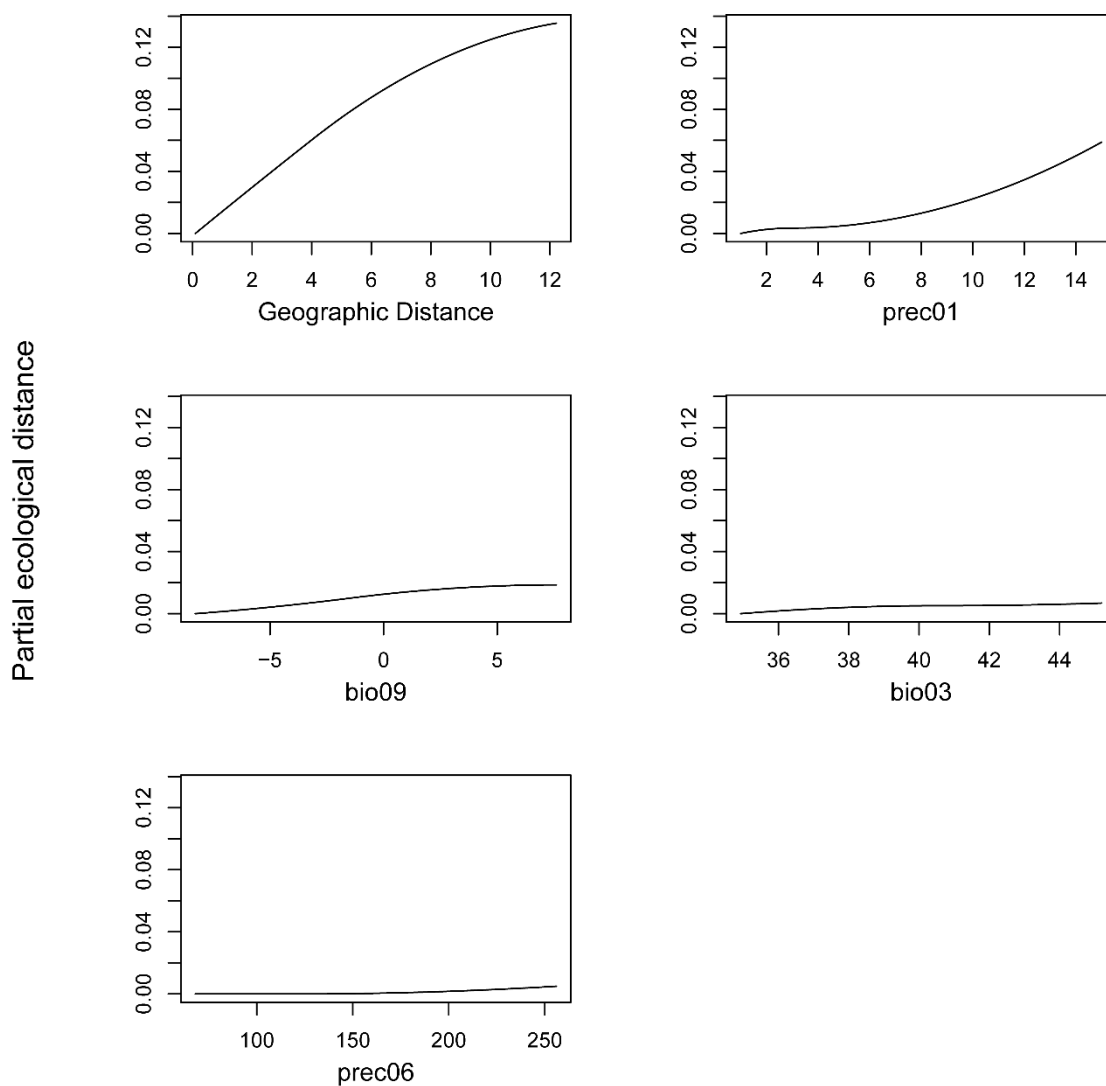


图 2-21 所有种群中遗传距离沿环境距离梯度变化的 I-样条曲线

Fig. 2-21 I-splines showing changes genetic distance along environmental distance gradients in all populations

3 讨论

3.1 遗传多样性与遗传结构

本研究使用65个与胁迫相关的细胞核候选基因，在川滇高山栎的60个种群，587个个体中进行扩增，使用高通量测序的方法鉴定到381个SNP位点，加上8个核SSR位点，对川滇高山栎的遗传结构和遗传多样性进行分析。两套数据对遗传结构的研究得到类似的结果：川滇高山栎在其分布区内存在明显的遗传分化，即以东经106-108度为界线，分布于喜马拉雅东部的种群(Tibet分枝)和分布在横断山脉、四川西部山地(HDM-WSP分支)的川滇高山栎属于两个明显的遗传分枝。这种分化结构来源于种群历史动态，即第三纪、第四纪的气候动荡及喜马拉雅山脉的隆升过程造成的川滇高山栎种群扩张、漂变、迁移、隔离等(Du *et al.*, 2017)。从图2-5、图2-6及表2-5中我们发现，基于中性位点的分析，得到不明显的遗传结构，但基于非中性位点的分析得到了非常明显的遗传结构，因此可以判断，川滇高山栎对不同分布区差异明显的生态条件的适应加剧了枝系间的分化程度。普遍存在的枝系内的花粉基因流促进了基因交流，因而遗传分化主要分布于种群内，在种群间极低。

基于中性位点的分析发现，Tibet分枝的遗传多样性显著低于HDM-WSP分枝，这可能与其经历过遗传漂变有关。第四纪冰期震荡对喜马拉雅山地区的影响非常明显，川滇高山栎的分布区被冰川反复地覆盖，川滇高山栎随之收缩至避难所，当气候变暖后又重新扩张(Meng *et al.*, 2017; Du *et al.*, 2017)。这个过程造成了Tibet分枝的遗传多样性的丧失，尽管突变和基因流不断增加了其多样性，但依旧无法达到HDM-WSP的程度。而基于多数为受选择位点的全部SNP的分析却发现，两个分枝的遗传多样性差异不明显。这是因为非中性位点并未经历冰期的遗传漂变，因此冰期的遗传漂变不会对Tibet的遗传多样性造成影响。

3.2 异常SNP位点

使用基于 F_{ST} 异常值检验和基于等位基因频率-气候梯度相关性的检验方法对381个SNP位点进行鉴定，我们分别检测出对Tibet分枝、HDM-WSP分枝和所有种群的生态适应起到关键作用的适应性SNP。未来我们可以关注其详细的功能，并比较两个分支中非中性SNP影响植物生态适应内在机理的异同，进而为保护两个分枝的川滇高山栎的遗传资源提出不同的针对性策略。在使用BayEnv对非中性SNP进行鉴定时，由于没有得到真正中性的SNP数据集，而是使用BAYESCAN和FDIST2没有鉴定到的SNP位点作为参考模型，因此结果中存在大量假阴性，最终得到的非中性SNP极少。因此仅有的研究需要避免该情况的出现。

3.3 遗传变异应对气候变化的模式

使用多种方法对川滇高山栎应对气候变化的模式进行分析，得到一个一致的结论：

无论是在两个分枝还是在所有种群尺度下，地理因素是影响遗传分化的最主要因素。除了地理因素，几种方法模拟的气候因素对遗传变异的影响差异很大。同为基于距离矩阵的方法，Mantel检验和GDM的结果类似，即在HDM-WSP分枝和所有种群中，川滇高山栎的遗传变异受到一月降水量(prec01)的影响比较显著，而在Tibet分枝中，气候变量的影响不显著。基于SNP数据的RDA的结果表明干季平均气温(bio09)是影响HDM-WSP分枝和所有种群遗传变异的最重要因素，而在Tibet分枝中，四种气候因子对遗传变异的影响差异不大，其中等温性(bio03)为影响最大的气候变量。基于SSR的结果则表明，四种气候因子对遗传分化均存在显著的影响，尤其是干季平均气温(bio09)、一月降水量(prec01)和六月降水量(prec06)。GF结果却表明在Tibet分枝和HDM-WSP分枝和所有种群中，所有气候变量对遗传变异的影响差异不大，其中六月降水量(prec06)是相对影响最大的气候变量。

综上所述，除了一月降水量的距离矩阵对HDM-WSP分枝和所有种群遗传距离的影响最大，其他气候因素对遗传变异都存在显著的影响，但差异不大。另外GF的累积重要性曲线图表明，经度范围在东经96-98度之间、一月降水量在12-14毫米之间、六月降水量在160-180毫米之间时，对遗传变异的影响最大，而其他环境因子对遗传变异的影响比较平滑。此结果可用于预测川滇高山栎的遗传变异对未来气候变化的响应模式。

4 结论与展望

4.1 结论

本研究从65个胁迫相关候选基因中鉴定到381个SNP位点，以此SNP数据集为基础，另外使用8个中性SSR位点，对我国西南地区川滇高山栎种群的遗传多样性、遗传结构及生态适应性进行研究，得到以下结论：

(1) 两种分子标记均检测到川滇高山栎分布区内存在显著的遗传结构，并将其分为Tibet分枝和HDM-WSP分枝，分别分布于西藏东部和横断山脉-四川西部山区。基于所有SNP数据集的分析得到两个分枝之间的结构更加明显，由此可以判断，川滇高山栎对不同分布区差异明显的生态条件的适应加剧了枝系间的分化程度。

(2) 基于中性位点(SSR和中性SNP)的分析发现，Tibet分枝的遗传多样性显著低于HDM-WSP分枝，这可能与其经历过遗传漂变有关。而基于多数为受选择位点的全部SNP位点的分析却发现，两个分枝的遗传多样性差异不明显。这是因为非中性位点并未经历冰期的遗传漂变，因此冰期的遗传漂变不会对Tibet的遗传多样性造成影响。

(3) 使用四种检测方法，共检测到310个非中性的SNP，来自56个候选基因。分别检测出对Tibet分枝、HDM-WSP分枝和所有种群的生态适应起到关键作用的适应性SNP，并找出两个分枝中特有的和共有的基因，比较两个分枝中非中性基因影响植物生态适应内在机理的异同，进而为保护两个分枝的川滇高山栎的遗传资源提出不同的针对性策略。

(4) 无论是在两个分枝还是在所有种群尺度下，地理因素是影响川滇高山栎遗传分化的最主要因素。除了地理因素，四种气候因子对遗传分化的影响差异不大，但基于距离矩阵相关性的分析方法(Mantel检验和GDM)同时检测到一月降水量(prec01)在HDM-WSP分枝和所有种群中对遗传距离的影响最大，存在显著的环境隔离(IBE, isolation by environment)模式。另外GF的累积重要性曲线图表明，经度范围在东经96-98度之间、一月降水量在12-14毫米之间、六月降水量在160-180毫米之间时，对遗传变异的影响最大，而其他环境因子对遗传变异的影响比较平滑。此结果可预测川滇高山栎的遗传变异对未来气候变化的响应模式。

4.2 展望

本论文对川滇高山栎的生态适应进行了深入的研究，但依然存在一些问题需要解决，以下列举了对今后研究的一些展望。

(1) 本研究对候选基因PCR产物进行高通量测序，进而使用生物信息学的方法处理序列数据，鉴定SNP。但结果发现，存在大量的数据缺失，可能的原因是测序深度不够高，或鉴定SNP的过程存在缺陷，因此建议提高测序深度，并优化SNP的鉴定流程，提高数据的使用效率。

(2) 在使用BayEnv对非中性SNP进行检测时, 由于没有得到真正中性的SNP数据集作为参考模型, 因此结果中存在大量假阴性, 最终得到的非中性SNP极少。因此建议设计实验时加入大量中性基因与候选基因同时进行PCR、测序与SNP鉴定, 从而获取中性SNP数据集, 以提高BayEnv的检测效率。

(3) 由于测序技术的不断成熟以及成本的不断降低, 对非模式物种进行全基因组或简化基因组测序, 进而对生态适应进行分析已经成为趋势(Bay *et al.*, 2018)。这种方法可以得到更大规模的数据, 从而对生态适应进行更加精准的分析。

(4) 本研究使用四种方法检测到大量非中性SNP位点, 但仍需要对这些SNP适应生态变异的内在机理进行深入的分析, 从而得到更加直接的证据来对物种生态适应的分子机制进行解释。

(5) 两种非线性的生态-遗传相关性的建模方法GF和GDM可用于预测遗传变异对未来气候变化的响应模式。这两种方法甚至优于物种分布模型的效果(Fitzpatrick & Keller, 2015), 可以在后续的研究中加以利用。

(6) 更加深入的研究需要继续进行, 也为对其他地区、其他林木物种的研究提供参考, 同时为森林资源的保护提供理论依据。

参考文献

1. 陈焕镛, 黄成就. 中国植物志: 第 22 卷[M]. 北京: 科学出版社, 1998.
2. 端木. 我国栎属资源的综合利用[J]. 河北林学院学报, 1994, 19(2): 177-181.
3. 李进, 陈可咏. 不同海拔高度川滇高山栎群体遗传多样性的变化[J]. 植物学报: 英文版, 1998, 40(8): 761-767.
4. 李进, 陈可咏. 川滇高山栎群体遗传结构的初步研究[J]. 北京林业大学学报, 1997, 19(2): 93-98.
5. 李俊清. 森林生态学(第三版)[M]. 北京: 高等教育出版社, 2017.
6. 李俊清. 西藏硬叶常绿栎类的多样性与进化[J]. 生态学杂志, 1996, 15(4): 55-58.
7. 任亮, 苏玉虹, 巴彩凤, 等. PCR引物设计技巧[J]. 现代畜牧兽医, 2005, (6): 49.
8. 王国严, 徐阿生. 川滇高山栎研究综述[J]. 四川林业科技, 2008, 29(2): 23-29.
9. 王玉垚, 张悦, 方剑火, 等. 一种基于候选基因研究树木生态适应的方法: 以川滇高山栎为例[J]. 兰州大学学报(自然科学版), 2017, N20170283(已录用, 未发表).
10. 杨玉坡, 李承彪, 管中天. 四川森林[M]. 北京: 中国林业出版社, 1992.
11. 尤超, 赵大球, 梁乘榜, 等. PCR 引物设计方法综述[J]. 现代农业科技, 2011, (17): 48-51.
12. 张新宇, 高燕宁. PCR 引物设计及软件使用技巧[J]. 生物信息学, 2004, 2(4): 15-18.
13. 郑万钧. 中国树木志[M]: 中国林业出版社, 1983.
14. 周浙昆. 中国栎属的起源演化及其扩散[J]. 云南植物研究, 1992, 14(3): 227-236.
15. Akey J M, Zhang G, Zhang K, et al. Interrogating a high-density SNP map for signatures of natural selection[J]. Genome research, 2002, 12(12): 1805-1814.
16. Alberto F J, Derory J, Boury C, et al. Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*[J]. Genetics, 2013, 195(2): 495-512.
17. Amos W, Hoffman J I, Frodsham A, et al. Automated binning of microsatellite alleles: problems and solutions[J]. Molecular Ecology Notes, 2007, 7(1): 10-14.
18. Anderson J T, Willis J H, Mitchell-Olds T. Evolutionary genetics of plant adaptation[J]. Trends in Genetics, 2011, 27(7): 258-266.
19. Atwell S, Huang Y S, Vilhjálmsson B J, et al. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines[J]. Nature, 2010, 465: 627-631.
20. Balding D J. A tutorial on statistical methods for population association studies[J]. Nature Reviews Genetics, 2006, 7(10): 781-791.
21. Banta J A, Ehrenreich I M, Gerard S, et al. Climate envelope modelling reveals intraspecific relationships among flowering phenology, niche breadth and potential range size in *Arabidopsis thaliana*[J]. Ecology Letters, 2012, 15(8): 769-777.
22. Bashalkhanov S, Eckert A J, Rajora O P. Genetic signatures of natural selection in response to air pollution in red spruce (*Picea rubens*, Pinaceae)[J]. Molecular ecology, 2013, 22(23): 5877-5889.
23. Bay R A, Harrigan R J, Le Underwood V, et al. Genomic signals of selection predict climate-driven population declines in a migratory bird[J]. Science, 2018, 359(6371): 83-86.
24. Beaumont M A. Adaptation and speciation: what can F_{ST} tell us?[J]. Trends in ecology & evolution, 2005, 20(8): 435-440.
25. Beaumont M A, Balding D J. Identifying adaptive genetic divergence among populations from genome scans[J]. Molecular ecology, 2004, 13(4): 969-980.
26. Beaumont M A, Nichols R A. Evaluating Loci for Use in the Genetic Analysis of Population

- Structure[J]. Proceedings of the Royal Society of London, 1996, 263(1377): 1619-1626.
27. Bergelson J, Roux F. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*[J]. Nature Reviews Genetics, 2010, 11(12): 867-879.
 28. Bolger A M, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. Bioinformatics, 2014, 30(15): 2114-2120.
 29. Bonin A, Taberlet P, Miaud C, et al. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*)[J]. Molecular Biology and Evolution, 2005, 23(4): 773-783.
 30. Boutet I, Tanguy A, Le Guen D, et al. Global depression in gene expression as a response to rapid thermal changes in vent mussels[J]. Proceedings of the Royal Society of London B: Biological Sciences, 2009: b20090503.
 31. Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author)[J]. Statistical science, 2001, 16(3): 199-231.
 32. Briggs D, Walters S M. Plant variation and evolution[M]: Cambridge University Press, 2016.
 33. Bustamante C D, Nielsen R, Sawyer S A, et al. The cost of inbreeding in *Arabidopsis*[J]. Nature, 2002, 416: 531-534.
 34. Catchen J M, Angel A, Paul H, et al. Stacks: Building and genotyping loci de novo from short-read sequences[J]. G3 Genesgenetics, 2011, 1(3): 171-182.
 35. Catchen J, Hohenlohe P A, Bassham S, et al. Stacks: an analysis tool set for population genomics[J]. Molecular Ecology, 2013, 22(11): 3124-3140.
 36. Charlesworth B. Elements of evolutionary genetics[M]: Roberts Publishers, 2010.
 37. Charlesworth B, Morgan M T, Charlesworth D. The effect of deleterious mutations on neutral molecular variation[J]. Genetics, 1993, 134(4): 1289-1303.
 38. Coop G, Witonsky D, Di Rienzo A, et al. Using environmental correlations to identify loci underlying local adaptation[J]. Genetics, 2010, 185(4): 1411-1423.
 39. Csilléry K, Lalaüe H, Vendramin G G, et al. Detecting short spatial scale local adaptation and epistatic selection in climate - related candidate genes in European beech (*Fagus sylvatica*) populations[J]. Molecular Ecology, 2014, 23(19): 4696-4708.
 40. Cushman S A, Landguth E L. Spurious correlations and inference in landscape genetics[J]. Molecular Ecology, 2010, 19(17): 3592-3602.
 41. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools[J]. Bioinformatics, 2011, 27(15): 2156-2158.
 42. Davey J W, Hohenlohe P A, Etter P D, et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing[J]. Nature Reviews Genetics, 2011, 12(7): 499-510.
 43. de Villemereuil P, Gaggiotti O E, Mouterde M, et al. Common garden experiments in the genomic era: new perspectives and opportunities[J]. Heredity, 2016, 116(3): 249-254.
 44. Dobzhansky T, Dobzhansky T G. Genetics of the evolutionary process[M]: Columbia University Press, 1970.
 45. Dow B D, Ashley M V, Howe H F. Characterization of highly variable (GA/CT)_n microsatellites in the bur oak, *Quercus macrocarpa*[J]. Theoretical and applied genetics, 1995, 91(1): 137-141.
 46. Doyle J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue[J]. Phytochem Bull, 1987, 19(1): 11-15.
 47. Du F K, Hou M, Wang W, et al. Phylogeography of *Quercus aquifolioides* provides novel insights into the Neogene history of a major global hotspot of plant diversity in South-West China[J]. Journal of biogeography, 2017, 44(2): 294-307.

48. Durand J, Bodénès C, Chancerel E, et al. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study[J]. BMC genomics, 2010, 11(1): 570. <https://doi.org/10.1186/1471-2164-11-570>.
49. Earl D A, Vonholdt B M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method[J]. Conservation Genetics Resources, 2012, 4(2): 359-361.
50. Eckert A J, Bower A D, González Martínez S C, et al. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae)[J]. Molecular ecology, 2010, 19(17): 3789-3805.
51. Eckert A J, Dyer R J. Defining the landscape of adaptive genetic diversity[J]. Molecular ecology, 2012, 21(12): 2836-2838.
52. Elith J, Leathwick J R. Species distribution models: ecological explanation and prediction across space and time[J]. Annual review of ecology, evolution, and systematics, 2009, 40: 677-697.
53. Ellis N, Smith S J, Pitcher C R. Gradient forests: calculating importance gradients on physical predictors[J]. Ecology, 2012, 93(1): 156-168.
54. Elshire R J, Glaubitz J C, Sun Q, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species[J]. PloS one, 2011, 6(5): e19379.
55. Endler J A. Natural selection in the wild[M]: Princeton University Press, 1986.
56. Espíndola A, Pellissier L, Maiorano L, et al. Predicting present and future intra-specific genetic structure through niche hindcasting across 24 millennia[J]. Ecology Letters, 2012, 15(7): 649-657.
57. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study[J]. Molecular Ecology, 2005, 14(8): 2611-2620.
58. Excoffier L, Hofer T, Foll M. Detecting loci under selection in a hierarchically structured population[J]. Heredity, 2009, 103(4): 285-298.
59. Excoffier L, Lischer H E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows[J]. Molecular Ecology Resources, 2010, 10(3): 564-567.
60. Falush D, Stephens M, Pritchard J K. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies[J]. GENETICS, 2003, 164(4): 1567-1587.
61. Feng L, Zhang Y, Chen X, et al. Allopatric divergence, local adaptation, and multiple Quaternary refugia in a long-lived tree (*Quercus spinosa*) from subtropical China[J]. bioRxiv, 2017: 112375.
62. Ferrier S, Manion G, Elith J, et al. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment[J]. Diversity and distributions, 2007, 13(3): 252-264.
63. Fick S E, Hijmans R J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas[J]. International Journal of Climatology, 2017, 37(12): 4302-4315.
64. Fitzpatrick M C, Keller S R. Ecological genomics meets community - level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation[J]. Ecology Letters, 2015, 18(1): 1-16.
65. Fitzpatrick M C, Sanders N J, Normand S, et al. Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients[J]. Proceedings of the Royal Society of London B: Biological Sciences, 2013, 280(1768): 20131201.
66. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective[J]. Genetics, 2008, 180(2): 977-993.
67. Fournier-Level A, Korte A, Cooper M D, et al. A map of local adaptation in *Arabidopsis thaliana*[J]. Science, 2011, 334(6052): 86-89.

68. Franks S J, Hoffmann A A. Genetics of climate change adaptation[J]. Annual review of genetics, 2012, 46(12): 185-208.
69. Frichot E, François O. LEA: An R package for landscape and ecological association studies[J]. Methods in Ecology & Evolution, 2015, 6(8): 925-929.
70. Frichot E, Schoville S D, Bouchard G, et al. Testing for associations between loci and environmental gradients using latent factor mixed models[J]. Molecular biology and evolution, 2013, 30(7): 1687-1699.
71. Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing[J]. Nature biotechnology, 2009, 27(2): 182-189.
72. González-Martínez S C, Ersoz E, Brown G R, et al. DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L.[J]. Genetics, 2006, 172(3): 1915-1926.
73. Goslee S C, Urban D L. The ecodist package for dissimilarity-based analysis of ecological data[J]. Journal of Statistical Software, 2007, 22(7): 1-19.
74. Grattapaglia D, Plomion C, Kirst M, et al. Genomics of growth traits in forest trees[J]. Current opinion in plant biology, 2009, 12(2): 148-156.
75. Grivet D, Sebastiani F, Alía R, et al. Molecular footprints of local adaptation in two Mediterranean conifers[J]. Molecular biology and evolution, 2010, 28(1): 101-116.
76. Hancock A M, Di Rienzo A. Detecting the genetic signature of natural selection in human populations: models, methods, and data[J]. Annual review of anthropology, 2008, 37: 197-217.
77. Hendry A P, Nosil P, Rieseberg L H. The speed of ecological speciation[J]. Functional ecology, 2007, 21(3): 455-464.
78. Hijmans R J. Geosphere: spherical trigonometry[J], 2016. URL <http://CRAN.R-project.org/package=geosphere>. R package version, 1.2-24.
79. Holderegger R, Kamm U, Gugerli F. Adaptive vs. neutral genetic diversity: implications for landscape genetics[J]. Landscape Ecology, 2006, 21(6): 797-807.
80. Holderegger R, Wagner H H. Landscape genetics[J]. AIBS Bulletin, 2008, 58(3): 199-207.
81. Hubisz M J, Falush D, Stephens M, et al. Inferring weak population structure with the assistance of sample group information[J]. Molecular Ecology Resources, 2009, 9(5): 1322-1332.
82. Hudson R R, Kreitman M, Aguadé M. A test of neutral molecular evolution based on nucleotide data[J]. Genetics, 1987, 116(1): 153-159.
83. Hughes A R, Inouye B D, Johnson M T, et al. Ecological consequences of genetic diversity[J]. Ecology letters, 2008, 11(6): 609-623.
84. Hughes L. Biological consequences of global warming: is the signal already apparent?[J]. Trends in ecology & evolution, 2000, 15(2): 56-61.
85. Hulce D, Li X, Snyderleiby T, et al. GeneMarker® Genotyping Software: Tools to Increase the Statistical Power of DNA Fragment Analysis[J]. Journal of Biomolecular Techniques Jbt, 2011, 22(Suppl): S35.
86. Jay F, Manel S, Alvarez N, et al. Forecasting changes in population genetic structure of alpine plants in response to global warming[J]. Molecular ecology, 2012, 21(10): 2354-2368.
87. Jeffreys H. The theory of probability[M]: OUP Oxford, 1998.
88. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data[J]. Bioinformatics, 2011, 27(21): 3070-3071.
89. Joost S, Bonin A, Bruford M W, et al. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation[J]. Molecular ecology, 2007, 16(18):

- 3955-3969.
90. Joshi J, Schmid B, Caldeira M C, et al. Local adaptation enhances performance of common plant species[J]. *Ecology Letters*, 2001, 4(6): 536-544.
 91. Kampfer S, Lexer C, Glössl J, et al. Characterization of (GA)_n microsatellite loci from *Quercus robur*[J]. *Hereditas*, 1998, 129(2): 183-186.
 92. Kaplan N L, Hudson R R, Langley C H. The "hitchhiking effect" revisited[J]. *Genetics*, 1989, 123(4): 887-899.
 93. Kawecki T J, Ebert D. Conceptual issues in local adaptation[J]. *Ecology letters*, 2004, 7(12): 1225-1241.
 94. Keller S R, Levensen N, Ingvarsson P K, et al. Local selection across a latitudinal gradient shapes nucleotide diversity in balsam poplar, *Populus balsamifera* L[J]. *Genetics*, 2011, 188(4): 941-952.
 95. Kremer A, Ronce O, Robledo Arnuncio J J, et al. Long-distance gene flow and adaptation of forest trees to rapid climate change[J]. *Ecology letters*, 2012, 15(4): 378-392.
 96. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. *Molecular Biology & Evolution*, 2016, 33(7): 1870-1874.
 97. Langlet O. Two hundred years genecology[J]. *Taxon*, 1971: 653-721.
 98. Lasky J R, Des Marais D L, Mckay J, et al. Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate[J]. *Molecular Ecology*, 2012, 21(22): 5512-5529.
 99. Legendre P, Legendre L F. *Numerical ecology*[M]: Elsevier, 2012.
 100. Lewontin R C, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms[J]. *Genetics*, 1973, 74(1): 175-195.
 101. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data[J]. *Bioinformatics*, 2011, 27(21): 2987-2993.
 102. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform[M]: Oxford University Press, 2009.
 103. Lippert P C, van Hinsbergen D, Dupont-Nivet G. The Early Cretaceous to Present latitude of the central Lhasa-plano: a paleomagnetic synthesis with implications for Cenozoic tectonics, paleogeography, and climate of Asia[J]. *Towards an Improved Understanding of Uplift Mechanisms and the Elevation History of the Tibetan Plateau*. Geological Society of America Special Paper, 2014, 507.
 104. Lotterhos K E, Whitlock M C. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests[J]. *Molecular ecology*, 2014, 23(9): 2178-2192.
 105. Manel S, Gugerli F, Thuiller W, et al. Broad - scale adaptive genetic variation in alpine plants is driven by temperature and precipitation[J]. *Molecular Ecology*, 2012, 21(15): 3729-3738.
 106. Manel S, Joost S, Epperson B K, et al. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field[J]. *Molecular Ecology*, 2010, 19(17): 3760-3772.
 107. Manel S, Schwartz M K, Luikart G, et al. Landscape genetics: combining landscape ecology and population genetics[J]. *Trends in ecology & evolution*, 2003, 18(4): 189-197.
 108. Manion G, Lisk M, Ferrier S, et al. gdm: Generalized Dissimilarity Modeling[J], 2017.
 109. Matyas C. Climatic adaptation of trees: rediscovering provenance tests[J]. *Euphytica*, 1996, 92(1-2): 45-54.
 110. McDonald J H, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*[J]. *Nature*, 1991, 351: 652-654.
 111. Mckenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for

- analyzing next-generation DNA sequencing data[J]. *Genome Research*, 2010, 20(9): 1297-1303.
112. Meng H H, Su T, Gao X Y, et al. Warm-cold colonization: response of oaks to uplift of the Himalaya-Hengduan Mountains[J]. *Molecular ecology*, 2017, 26(12): 3276-3294.
113. Miller M R, Dunham J P, Amores A, et al. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers[J]. *Genome research*, 2007, 17(2): 240-248.
114. Mulch A, Chamberlain C P. Earth science: the rise and growth of Tibet[J]. *Nature*, 2006, 439(7077): 670-671.
115. Myers N, Mittermeier R A, Mittermeier C G, et al. Biodiversity hotspots for conservation priorities[J]. *Nature*, 2000, 403: 853-858.
116. Naimi B, Hamm N A S, Groen T A, et al. Where is positional uncertainty a problem for species distribution modelling?[J]. *Ecography*, 2014, 37(2): 191-203.
117. Narum S R, Buerkle C A, Davey J W, et al. Genotyping-by-sequencing in ecological and conservation genomics[J]. *Molecular ecology*, 2013, 22(11): 2841-2847.
118. Narum S R, Hess J E. Comparison of FST outlier tests for SNP loci under selection[J]. *Molecular ecology resources*, 2011, 11(s1): 184-194.
119. Neale D B, Savolainen O. Association genetics of complex traits in conifers[J]. *Trends in plant science*, 2004, 9(7): 325-330.
120. Nielsen R, Bustamante C, Clark A G, et al. A scan for positively selected genes in the genomes of humans and chimpanzees[J]. *PLoS biology*, 2005, 3(6): e170.
121. Oksanen J, F Blanchet G, Friendly M, et al. *vegan: Community Ecology Package. Ordination methods, diversity analysis and other functions for community and vegetation ecologists. Version 2.4-5. URL <https://CRAN.R-project.org/package=vegan>*[M], 2017.
122. Parchman T L, Gompert Z, Mudge J, et al. Genome - wide association genetics of an adaptive trait in lodgepole pine[J]. *Molecular ecology*, 2012, 21(12): 2991-3005.
123. Parisod C, Christin P A. Genome - wide association to fine - scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae)[J]. *New Phytologist*, 2008, 178(2): 436-447.
124. Parmesan C. Ecological and evolutionary responses to recent climate change[J]. *Annu. Rev. Ecol. Evol. Syst.*, 2006, 37: 637-669.
125. Parry M, Canziani O, Palutikof J, et al. *Climate change 2007: impacts, adaptation and vulnerability*[M]: Cambridge University Press Cambridge, 2007.
126. Peakall R, Smouse P E. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update[J]. *Bioinformatics*, 2012, 28(19): 2537-2539.
127. Pearman P B, D'Amen M, Graham C H, et al. Within - taxon niche structure: niche conservatism, divergence and predicted effects of climate change[J]. *Ecography*, 2010, 33(6): 990-1003.
128. Plomion C, Aury J M, Amselem J, et al. Decoding the oak genome: public release of sequence data, assembly, annotation and publication strategies[J]. *Molecular ecology resources*, 2016, 16(1): 254-265.
129. Pritchard J K, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data[J]. *Genetics*, 2000, 155(2): 945-959.
130. Qiu Y, Fu C, Comes H P. Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora[J]. *Molecular phylogenetics and evolution*, 2011, 59(1): 225-244.
131. Rellstab C, Zoller S, Walthert L, et al. Signatures of local adaptation in candidate genes of oaks

- (*Quercus* spp.) with respect to present and future climatic conditions[J]. *Molecular Ecology*, 2016, 25(23): 5907-5924.
132. Renner S S. Available data point to a 4-km-high Tibetan Plateau by 40 Ma, but 100 molecular - clock papers have linked supposed recent uplift to young node ages[J]. *Journal of Biogeography*, 2016, 43(8): 1479-1487.
133. Riordan E C, Gugger P F, Ortego J, et al. Association of genetic and phenotypic variability with geography and climate in three southern California oaks[J]. *American journal of botany*, 2016, 103(1): 73-85.
134. Roschanski A M, Csilléry K, Liepelt S, et al. Evidence of divergent selection for drought and cold tolerance at landscape and local scales in *Abies alba* Mill. in the French Mediterranean Alps[J]. *Molecular Ecology*, 2016, 25(3): 776-794.
135. Rosenberg N A. dstruct: a program for the graphical display of population structure[J]. *Molecular Ecology Notes*, 2004, 4(1): 137-138.
136. Rousset F. genepop ’007: a complete re-implementation of the genepop software for Windows and Linux[M], 2008: 103, 106.
137. Sabeti P C, Varilly P, Fry B, et al. Genome-wide detection and characterization of positive selection in human populations[J]. *Nature*, 2007, 449: 913-918.
138. Santamaría L, Figuerola J, Pilon J J, et al. Plant performance across latitude: the role of plasticity and local adaptation in an aquatic plant[J]. *Ecology*, 2003, 84(9): 2454-2461.
139. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation[J]. *Nature Reviews Genetics*, 2013, 14(11): 807-820.
140. Savolainen O, Pyhäjärvi T, Knürr T. Gene flow and local adaptation in trees[J]. *Annu. Rev. Ecol. Evol. Syst.*, 2007, 38: 595-619.
141. Segelbacher G, Cushman S A, Epperson B K, et al. Applications of landscape genetics in conservation biology: concepts and challenges[J]. *Conservation genetics*, 2010, 11(2): 375-385.
142. Smith J M, Haigh J. The hitch-hiking effect of a favourable gene[J]. *Genetics Research*, 1974, 23(1): 23-35.
143. Sork V L, Aitken S N, Dyer R J, et al. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate[J]. *Tree Genetics & Genomes*, 2013, 9(4): 901-911.
144. Sork V L, Davis F W, Westfall R, et al. Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change[J]. *Molecular Ecology*, 2010, 19(17): 3806-3823.
145. Sork V L, Squire K, Gugger P F, et al. Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*[J]. *American Journal of Botany*, 2016, 103(1): 33-46.
146. Sork V L, Waits L. Contributions of landscape genetics-approaches, insights, and future potential[J]. *Molecular Ecology*, 2010, 19(17): 3489-3495.
147. Stapley J, Reger J, Feulner P G, et al. Adaptation genomics: the next generation[J]. *Trends in ecology & evolution*, 2010, 25(12): 705-712.
148. Steinkellner H, Fluch S, Turetschek E, et al. Identification and characterization of (GA/CT)_n-microsatellite loci from *Quercus petraea*[J]. *Plant molecular biology*, 1997, 33(6): 1093-1096.
149. Stinchcombe J R, Hoekstra H E. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits[J]. *Heredity*, 2008, 100(2): 158-170.
150. Storz J F. INVITED REVIEW: Using genome scans of DNA polymorphism to infer adaptive

- population divergence[J]. *Molecular ecology*, 2005, 14(3): 671-688.
151. Sun B, Wu J, Liu Y C, et al. Reconstructing Neogene vegetation and climates to infer tectonic uplift in western Yunnan, China[J]. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 2011, 304(3): 328-336.
152. Tang K, Thornton K R, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome[J]. *PLoS biology*, 2007, 5(7): e171.
153. Team R Core. R: A language and environment for statistical computing[J]. *Computing*, 2017.
154. Ter Braak C J. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis[J]. *Ecology*, 1986, 67(5): 1167-1179.
155. Tiffin P, Ross-Ibarra J. Advances and limits of using population genetics to understand local adaptation[J]. *Trends in ecology & evolution*, 2014, 29(12): 673-680.
156. Toumi L, Lumaret R. Allozyme characterisation of four Mediterranean evergreen oak species[J]. *Biochemical systematics and ecology*, 2001, 29(8): 799-817.
157. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3--new capabilities and interfaces[J]. *Nucleic Acids Research*, 2012, 40(15): e115.
158. Villemereuil P, Frichot É, Bazin É, et al. Genome scan methods against more complex models: when and how much should we trust them?[J]. *Molecular Ecology*, 2014, 23(8): 2006-2019.
159. Vitalis R, Dawson K, Boursot P. Interpretation of variation across marker loci as evidence of selection[J]. *Genetics*, 2001, 158(4): 1811-1823.
160. Wang Z, Liu H, Wei N, et al. Effects of stand regeneration management regimes and age on genetic structure of *Quercus aquifolioides* (Sclerophyllous Oak) in Southwestern China[J]. *Forest Science*, 2009, 55(2): 142-148.
161. Westfall R D, Conkle M T. Allozyme markers in breeding zone designation: Springer, 1992: 279-309.
162. Willis J C. A dictionary of the flowering plants and ferns[M]: CUP Archive, 1973.
163. Wisz M S, Hijmans R J, Li J, et al. Effects of sample size on the performance of species distribution models[J]. *Diversity and Distributions*, 2008, 14(5): 763-773.
164. Wright S I, Charlesworth B. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model[J]. *Genetics*, 2004, 168(2): 1071-1076.
165. Wright S. The genetical structure of populations[J]. *Annals of Human Genetics*, 1949, 15(1): 323-354.
166. Zeng K, Mano S, Shi S, et al. Comparisons of site-and haplotype-frequency methods for detecting positive selection[J]. *Molecular biology and evolution*, 2007, 24(7): 1562-1574.
167. Zhang X, Korpelainen H, Li C. Microsatellite variation of *Quercus aquifolioides* populations at varying altitudes in the Wolong Natural Reserve of China[J]. *Silva Fennica*, 2007, 40(3): 407-415.
168. Zhou Z, Wilkinson H, Wu Z. Taxonomical and evolutionary implications of the leaf anatomy and architecture of *Quercus* L. subgenus *Quercus* from China[J]. *Cathaya*, 1995, 7: 1-34.

个人简介

王玉垚，男，共青团员，1993年11月出生，河南济源人。2011-2015年就读于河南农业大学植物保护学院植物科学与技术专业，获得农学学士学位。2015-2018年就读于北京林业大学攻读生态学硕士学位，主要从事群体遗传学和生态适应的研究。硕士期间在兰州大学学报（自然科学版）发表论文一篇。

导师简介

杜芳，女，1981年11月生，甘肃兰州人，兰州大学生态学博士，法国波尔多第一大学环境科学博士。北京林业大学生态学副教授。主要获奖情况：2009年获得法国农业科学学院三大科研进展奖励；2013年获高等学校科学研究成果奖自然科学奖一等奖：松柏类植物的生物地理学和进化研究（第四完成人）；2013年获北京林业大学青年科技奖（自然科学类）；2014年获选“分子生态（Molecular Ecology）”期刊（影响因子6.5）最佳审稿人。2014年获教育部高等学校科学研究优秀成果奖自然科学奖一等奖（排名第四）；2014年获梁希青年论文奖三等奖；2015年当选北京市科技新星；2018年获北京林业大学科技之星；2018年获北京林业大学第六届优秀青年科技论文二等奖。近年来主要从事以分子生物学方法和生态学理论为研究手段揭示我国森林物种的遗传多样性分布及其形成规律并预测在气候环境变迁下森林物种变迁规律的研究。目前主持和参与国家、省部级等科研项目10项，累计科研经费300万元。其中主持国家自然科学基金青年项目、北京市科技新星、教育部博士点新教师基金，中央高校基金科研业务费专项资金等；作为主要参与人参与国家及省部级科研项目两项，学校创新团队项目一项。

成果

硕士在读期间发表文章:

王玉垚, 张悦, 方剑火, 杜芳. 一种基于候选基因研究树木生态适应的方法: 以川滇高山栎为例[J]. 兰州大学学报(自然科学版), 2017, N20170283. (已录用, 未发表)

致谢

时光荏苒，三年的硕士求学生涯即将结束。在学位论文交付之际，回顾研究生三年的生活，要感谢的人太多太多……

首先，我要把最诚挚的感谢献给我的导师，杜芳副教授！三年里，从论文的选题、设计、文献阅读、样品采集、数据处理分析到论文撰写的每个阶段，无不倾注了恩师大量的心血与汗水。这三年来的工作中，杜老师一直以各种方式激发和锻炼我的科研能力，同时她的严谨和极具感染力的科研热情也潜移默化的影响着我的生活和性格，对于科学问题的每次请教，都能得到导师耐心的解答；工作中的成败得失，总是会受到导师关切的鼓励和启发。杜老师渊博的学识、严谨的治学态度、敏锐的科学思路、忘我的工作精神始终影响和教育着我。在科研上，我很庆幸恩师能作为自己三年科研道路上的引路人，为我树立了学习工作方面的一个好榜样；在生活中，恩师对我的言传身教也让我感触颇深并使我终身受益，感谢杜老师早早培养了我自由正直的襟怀。在此，谨向恩师三年来在学习、科研和生活上对我的淳淳教诲、精心培养和亲切关怀表示我最衷心的感谢。千言万语融入一句：感谢恩师！

感谢中科院微生物所的尹康全老师、清华大学的方剑火老师、台湾师范大学的廖培均老师和黄秉宏老师、山东农业大学的王年老师以及北京林业大学生态学学科的李俊清老师、刘艳红老师、王襄平老师、吕佳老师等老师在我的研究过程中给我的细心指导。谢谢实验室的科研助理王宇暹老师在我的工作和生活中对我的支持和鼓励。在此向各位老师奉上我最诚挚的感谢和最美好的祝福！

感谢读研期间陪伴我的侯盟、鲁四海、宋佳、吕岩、刘媛、于涛、苏金源、张悦、李月娟、燕语、王天瑞、王瑞、王肖静、崔亦凡、杨睿等实验室的兄弟姐妹，感谢这三年来他们对我的照顾和帮助。感谢我的室友叶子奇、张鹏、杨雨果、余克非平日的关心和包容。感谢我的异地恋女友孙文静，她是我心中的光！感谢她这三年来对我默默的支持，是她对我的包容，鼓励和不离不弃让我能够坚持走到今天。

最后，感谢父母二十多年来对我的的养育之恩，让我能够在他们的关爱之下健康成长，感谢他们对我无微不至的照顾和关心。感谢我的亲人，在我的学习期间给予的支持和鼓励，我所取得的每一分成绩都与他们息息相关。

最后，向所有关系和帮助过我的朋友和亲人致以最诚挚的感谢！

