

分类号

密级

UDC



北京林业大学

学术型硕士学位论文

基于全基因组重测序的川滇高山栎和刺叶高山栎适应性渐渗研究

Adaptive introgression of *Quercus aquifolioides* and *Quercus spinosa* based on whole-genome resequencing

秦伟

指导教师 杜芳教授

学院 生态与自然保护学院

学科专业 生态学

研究方向 修复生态学

二〇二四年六月四日

独创性声明

本人声明所呈交的论文是本人在导师指导下独立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： 秦伟 日 期： 2024年06月04日

关于学位论文使用授权的声明

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：研究生在校攻读学位期间论文工作的知识产权单位属北京林业大学；学校有权保留并向国家有关部门或机构送交论文的纸质版和电子版，允许学位论文被查阅、借阅和复印；学校可以将学位论文的全部或部分内容公开或编入有关数据库进行检索，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。

签 名： 秦伟 导师签名： 杜芳 日 期： 2024年06月04日

答 辩 委 员 会 成 员 信 息

	姓 名	职 称	工 作 单 位
主 席	卢存福	教授	北京林业大学
委 员	李景文	教授	北京林业大学
	侯继华	教授	北京林业大学

摘要

适应性渐渗 (Adaptive introgression) 是快速气候变化背景下新等位基因的重要来源, 甚至可能比新发突变 (*de novo* mutation) 和持续性变异 (standing variation) 更重要。近年来的研究已经证明了自然界中普遍存在适应性渐渗, 尤其是在长寿命植物中。因此从全基因组水平解析适应性渐渗的遗传机制对于探究物种的适应性遗传分化和局地适应机制至关重要。本论文针对 18 个川滇高山栎 (*Quercus aquifolioides* Rehd. Et Wils.) 种群的 24 个个体和 18 个刺叶高山栎 (*Quercus spinosa* David ex Franchet) 种群的 25 个个体进行了全基因组重测序, 测序深度为 $24.16\times - 47.37\times$ 。经过质量控制和严格的过滤后, 最终获得了川滇高山栎的 358 661 个和刺叶高山栎的 346 281 个高质量单核苷酸多态性位点 (Single nucleotide polymorphisms, SNPs)。种群基因组学分析揭示了两个栎属物种表现出高水平的物种内部的东-西谱系遗传分化: 川滇高山栎西谱系 (*Q. aquifolioides*_West, *Qa_W*) 和川滇高山栎东谱系 (*Q. aquifolioides*_East, *Qa_E*); 刺叶高山栎西谱系 (*Q. spinosa*_West, *Qs_W*) 和刺叶高山栎东谱系 (*Q. spinosa*_East, *Qs_E*)。同时明确了由西向东“走出青藏高原 (Out-of-QTP)”的生物地理学模式。比较不同种群对之间的 SNPs 等位基因频率, 发现刺叶高山栎物种内的遗传分化高于川滇高山栎和刺叶高山栎物种间遗传分化。使用基因型-环境关联分析 (Genotype-environment associations, GEAs) 揭示了地理隔离 (Isolation-by-distance, IBD) 加速了刺叶高山栎谱系间适应性分化, 发现了青藏高原种群和非青藏高原种群表现出不同的局地适应 (Local adaptation) 模式: 温度和地理因素是驱动青藏高原种群的适应性遗传变异最主要的环境因素, 而非青藏高原种群的适应性遗传变异主要选择性压力是降水。全基因组水平的 f_{dM} 分析表明在川滇高山栎和 *Qs_W* 之间存在 1 461 个潜在的渐渗区域和一系列响应生物和非生物胁迫的功能基因。结合 f_{dM} 和 GEAs 分析发现 435 个参与川滇高山栎和 *Qs_W* 之间适应性渐渗的功能基因。这些功能基因均匀的分布在 12 条染色体上, 拥有更长的序列长度和内含子长度。

进一步研究物种间适应性渐渗位点, 发现参与物种间适应性渐渗基因中的 3-ketoacyl ACP synthase 2 (*KAS2*) 基因区域内表现出低水平的连锁不平衡 (Linkage disequilibrium, LD) 和遗传分化, 以及高水平的核苷酸多样性。其中 *KAS2* 基因内含子区域上一个假定 DNA 顺式调控原件 (GT1CONSENSUS) 中的适应性变异位点

(3_82758545) 的两种主要的等位基因在不同种群中表现出了明显的频率差异。其中 -A 等位基因主要集中在青藏高原种群中, -G 等位基因主要集中在非青藏高原种群中。基于两种等位基因, 构建了含 -A 和 -G 等位基因的质粒, 之后转入农杆菌 (*Agrobacterium tumefaciens*) 后制备成菌液。通过利用烟草的瞬时表达特性在烟草 (*Nicotiana benthamiana*) 中注射农杆菌菌液发现注射含有 -G 等位基因质粒的农杆菌菌液的区域在设置的所有环境条件下均表现出明显的细胞坏死现象。但是, 仅在黑暗、低温或强紫外线环境下, 注射含有 -A 等位基因质粒的农杆菌菌液的区域表现出明显的细胞坏死现象。

在 2050 年和 2070 年 RCP 2.6 和 RCP 8.5 条件下, 梯度森林模型预测发现川滇高山栎的脆弱性区域主要在青藏高原的北部和西部, 刺叶高山栎的脆弱性区域主要在横断山脉地区。

本研究结合种群遗传学、景观基因组学和分子功能实验验证方法揭示了两种栎属物种之间的适应性渐渗机制。为快速气候变化背景下, 森林树种的局地适应、适应性遗传分化、管理和保护提供了新的见解。

关键词: 适应性渐渗, 遗传分化, 局地适应, 走出青藏高原, 景观基因组学

Adaptive introgression of *Quercus aquifolioides* and *Quercus spinosa*
based on whole-genome resequencing

Master Candidate Wei Qin

(Ecology)

Directed by Fang Du

Abstract

Adaptive introgression is an important source of new alleles in the context of rapid-climate change, perhaps even more frequent than *de novo* mutation and standing variation. Recent studies have demonstrated the prevalence of adaptive introgression in nature, especially in long-lived plant species. Therefore, it is crucial to analyze the genetic mechanism of adaptive introgression at the genome-wide level in order to investigate the adaptive divergence and local adaptation. This study sampled 24 individuals from 18 *Quercus aquifolioides* Rehd. Et Wils. populations and 25 individuals from 18 *Quercus spinosa* David ex Franchet populations for whole-genome resequencing. All individuals sequencing depth ranged from 24.16× - 47.37×. After quality control and stringent filtration, we finally obtained 358 661 single nucleotide polymorphisms (SNPs) in *Q. aquifolioides* and 346 281 SNPs in *Q. spinosa*. Population genomics analysis revealed high degree of intra-specific east-west divergence between the two *Quercus* species (*Q. aquifolioides*: *Qa_W* and *Qa_E*; *Q. spinosa*: *Qs_W* and *Qs_E*), and identified biogeographic pattern of "Out-of-QTP" from west-to-east. Comparison of allele frequencies of SNPs between population pairs revealed that genetic differentiation within *Q. spinosa* was higher than that between *Q. aquifolioides* and *Q. spinosa*. Geographic isolation accelerated adaptive divergence among *Q. spinosa* lineages, as revealed using genotype-environment associations (GEAs). QTP populations and non-QTP populations exhibited different patterns of local adaptation: temperature and geography were mainly environmental factor driving adaptive genetic variation in QTP populations, whereas the main source of selective pressure for adaptive genetic variation in non-QTP populations was precipitation. The whole-genome f_{AM} analyses indicated the 1 461 potential adaptive introgression regions and a series of functional genes involved in biotic and abiotic stresses between *Q. aquifolioides* and *Qs_W*. Combining f_{AM} and GEAs analysis identified 435 functional genes that involved in adaptive introgression between *Q. aquifolioides* and *Qs_W*. These functional genes are evenly distributed on 12 chromosomes and have long sequence length and intron length.

Further study of adaptive introgression revealed that 3-ketoacyl ACP synthase 2 (*KAS2*) gene involved in adaptive introgression exhibited low levels of linkage disequilibrium (LD) and genetic differentiation, as well as high levels of nucleotide diversity. Two major allele genotypes of an adaptive variant (3_82758545) in a DNA cis-regulatory progenitor (GT1CONSENSUS) in *KAS2* showed significant frequency differences. A allele was mainly distributed in QTP populations and G allele was mainly concentrated in non-QTP populations. Based on the two allele genotypes, an experiment was constructed using two types of plasmids, which were later transferred into *Agrobacterium tumefaciens*.

Agrobacterium sap was injected into *N. benthamiana* utilizing the transient expression properties of *N. benthamiana*, the result found that *Agrobacterium* sap containing the G allele plasmid exhibited significant cell necrosis in rejection regions, including control and treatment experiments. However, *Agrobacterium* sap containing A allele plasmid exhibited significant cell necrosis only in control experiments.

The gradient forest model predicted that the vulnerability region of *Q. aquifolioides* is mainly in the north and west of the QTP, and that of *Q. spinosa* is mainly in the Hengduan Mountains in 2050 and 2070 under RCP 2.6 and RCP 8.5..

This study combined population genetics, landscape genomics methods and molecular function experimental validation methods to reveal adaptive introgression between two *Quercus* species. It will provide new insights into local adaptation, adaptive divergence, management and conservation of forest tree species in the context of rapid-climate change.

Key Words: adaptive introgression, adaptive divergence, local adaptation, Out-of-QTP, landscape genomic

目录

1 引言	1
1.1 基因渐渗	1
1.1.1 基因渐渗基本概念.....	1
1.1.2 基因渐渗的遗传特征.....	1
1.2 适应性渐渗	1
1.2.1 适应性渐渗基本概念.....	1
1.2.2 森林树种中适应性渐渗的研究案例.....	1
1.3 青藏高原	2
1.3.1 青藏高原范围.....	2
1.3.2 青藏高原历史地质发展.....	2
1.3.3 青藏高原物种间渐渗研究实例.....	3
1.4 川滇高山栎和刺叶高山栎基本概述.....	3
1.4.1 栎属物种基本介绍.....	3
1.4.2 栎属物种中的适应性渐渗.....	4
1.4.3 川滇高山栎和刺叶高山栎研究进展.....	4
1.5 本文的研究内容，研究目的及意义.....	5
1.5.1 研究内容	5
1.5.2 研究目的及意义.....	5
2 材料方法	6
2.1 样品采集	6
2.2 重测序数据过滤和获取.....	8
2.3 种群遗传学分析	9
2.3.1 遗传多样性.....	9
2.3.2 种群结构	9
2.3.3 系统发育树构建.....	9
2.3.4 种群历史动态.....	9
2.3.5 基因组岛检测.....	10
2.3.6 基因渐渗	10
2.4 景观基因组学分析.....	11
2.4.1 生物气候因子的获取.....	11
2.4.2 基因型-环境关联分析	11
2.4.3 遗传结构和环境梯度的多元关系.....	11
2.4.5 分子功能实验验证.....	12
2.4.6 未来命运预测.....	13
3 研究结果	14
3.1 全基因组重测序和变异位点检测.....	14
3.2 物种内和物种间基因组特征.....	14
3.2.1 遗传多样性分析.....	14

3.2.2 遗传结构	14
3.2.3 系统发育关系.....	14
3.2.4 种群历史	16
3.2.5 遗传分化	17
3.3 基因组渐渗区域检测.....	19
3.4 适应性遗传变异位点检测.....	26
3.5 遗传变异与地理、环境关系.....	27
3.6 适应性渐渗基因组特征.....	29
3.7 <i>KAS2</i> 基因内适应性变异与生态适应.....	33
3.8 梯度森林预测	36
4 讨论	38
4.1 川滇高山栎和刺叶高山栎种群历史动态和遗传分化模式.....	38
4.2 适应性渐渗选择信号.....	39
4.3 应对气候变化与保护的启示.....	40
5 结论和展望	42
5.1 结论	42
5.2 展望	43
参考文献	44
附录	51
个人简介	67
导师简介	68
获得成果目录清单	69
致谢	70

符号说明

- AIC: Akaike information criterion, 池信息量准则
CTAB: Cetyltrimethylammonium bromide, 十六烷基三甲基溴化铵法
CV: Cross validation, 交叉验证
D_{XY}: Absolute divergence, 绝对遗传分化
FDR: False discovery rate, 错误发现率
F_{ST}: Relative divergence, 相对遗传分化
GDM: Generalized dissimilarity modeling, 广义相异建模
GEAs: Genotype-environment associations, 基因型-环境关联分析
GF: Gradient forest, 梯度森林
GO: Gene ontology, 基因本体论
HW: Hardy-Weinberg, 哈迪-温伯格
IBD: Isolation-by-distance, 地理隔离
IBE: Isolation by environment, 环境隔离
ILS: Incomplete lineage sorting, 不完全谱系分选
LD: Linkage disequilibrium, 连锁不平衡
LFMM: Latent factor mixed model, 潜在因素混合模型
MAF: Minor allele frequency, 最小等位基因频率
MCMC: Markov chain monte carlo, 马尔可夫链蒙特卡罗
ML: Maximal likelihood, 最大似然法
Mya: Million years ago, 百万年前
N_e: Effective population size, 有效种群大小
NGS: Next generation sequencing, 二代测序
NJ: Neighbour joining, 邻接法
Out-of-QTP: Out-of- Qinghai-Tibet Plateau, 走出青藏高原
PCA: Principal components analysis, 主成分分析
PSMC: Pairwise sequentially markovian coalescent, 成对序列马可夫共祖先
Qa_E: *Q. aquifolioides*_East, 川滇高山栎东谱系
Qa_W: *Q. aquifolioides*_West, 川滇高山栎西谱系
Qs_E: *Q. spinosa*_East, 刺叶高山栎东谱系
Qs_W: *Q. spinosa*_West, 刺叶高山栎西谱系
QTP: Qinghai-Tibet Plateau, 青藏高原
RCP: Representative concentration pathway, 典型浓度路径
RDA: Redundancy analysis, 冗余分析
RI: Reproductive isolation, 生殖隔离
SFS: Site frequency spectrum, 位点频谱
SNPs: Single nucleotide polymorphisms, 单核苷酸多态性位点
SSRs: Simple sequence repeats, 简单重复序列标记
 π : Nucleotide diversity, 核苷酸多样性

1 引言

1.1 基因渐渗

1.1.1 基因渐渗基本概念

渐渗 (Introgression) 是由 Anderson 和 Hubricht^[1] 首先提出, 指物种之间的杂交后代可能会与亲本物种产生回交, 随着时间的推移, 可能导致 DNA 从一个物种持续性的转移到另一个物种中^[1-3]。一般来说, 渐渗更可能发生在亲缘关系密切的物种之间^[4,5]。近年来, 在多数生物物种中已经检测到了广泛的渐渗, 尤其是在长寿命森林树种中^[6-9]。这些在物种间不断发现的最近或更古老的渐渗事件正在颠覆生殖隔离阻碍基因在物种之间交流的传统观念^[3]。

1.1.2 基因渐渗的遗传特征

研究物种间渐渗的遗传特征一直是生态与进化领域的焦点问题。基因组中的渐渗区域往往会表现出适合度下降, 即更低的遗传负荷 (Genetic load), 并且渐渗位点的保留与否和渐渗信号的强度大小会受到多种因素的制约^[8]。基因密度高的区域往往观察到的渗入 DNA 频率较低, 可能是因为它的保留会干扰上下游基因的功能。此外, 基因组中低重组率的区域往往会表现出更少的 DNA 渗入, 因为重组的频率不足以将杂交种中有害的基因与渐渗的 DNA 片段分离^[3,8]。最后, 渐渗主要发生在中性等位基因上, 在导致生殖隔离 (Reproductive isolation, RI) 的等位基因很少^[10]。而具有“全局优势 (Globally advantageous)” 的等位基因更容易发生基因渐渗^[5,11]。

1.2 适应性渐渗

1.2.1 适应性渐渗基本概念

大多数情况下物种之间渐渗带来的结果往往都是中性的, 但渐渗也可能带来正面的影响, 通过渐渗而快速获得适应性等位基因从而提升个体的适应能力的方式, 被定义为“适应性渐渗 (Adaptive introgression)”^[3,7,12-14]。不同于缓慢的新发突变 (*de novo* mutation) 和持续性变异 (Standing variation), 适应性渐渗已被证实是许多物种适应性进化过程中快速获得遗传变异的重要方式。通过物种间持续性的适应性渐渗扩展了可用于应对选择性压力的遗传变异库^[15], 实现了对于物种局地适应 (Local adaptation) 能力的增强和生态位的扩张^[10,15,16]。同时, 一些研究还将渐渗描述为是对基因组的“入侵”^[4], 物种间持续性的适应性渐渗可能驱动了物种间共享多态性的增加, 从而削弱基因组水平的遗传分化, 进而有可能逆转物种之间的生殖隔离^[17]。

1.2.2 森林树种中适应性渐渗的研究案例

具有巨大的生物量和广泛分布的森林树种为其他生物提供了食物和庇护, 但森林树种容易受到干扰, 比如气候变化, 人为干扰等。通过分析森林树种之间的适应性渐渗机理对了解物种的局地适应具有重要的意义^[18]。近年来, 随着测序技术的发展和

方法的出现,研究者可以从全基因组水平(Whole-genome)检测森林树种中适应性渐渗的潜在区域^[3]。例如: Fu 等^[7]使用在亚洲同域分布的麻栎(*Quercus acutissima*)和栓皮栎(*Q. variabilis*)两个栎属物种的全基因组重测序数据进行 ABBA-BABA 检测发现了 80 944 个单核苷酸多态性位点(Single nucleotide polymorphisms, SNPs)位于两个栎属物种的基因组渐渗区域中,研究证明了适应性渐渗可以通过引入 DNA 顺式调控元件上的等位基因变异,从而实现响应生物和非生物胁迫相关的基因的调节,进而增强物种适应快速变化的环境的能力; Liu 等^[8]使用了 7 个杨属物种(响叶杨(*Populus adenopoda*)、银白杨(*P. alba*)、山杨(*P. davidiana*)、琼岛杨(*P. qionghdaoensis*)、圆叶杨(*P. rotundifolia*)、欧洲颤杨(*P. tremula*)和美洲颤杨(*P. tremuloides*))的全基因组重测序数据,通过对比适应性渐渗的基因组区域和基因组背景(除适应性渐渗以外的基因组区域)的遗传多样性指数,发现物种间相对遗传分化(Relative divergence, F_{ST})和绝对遗传分化(Absolute divergence, D_{XY})更低,物种内核苷酸多样性(Nucleotide diversity, π)更高^[8]; Guo 等^[9]以青藏高原东南部杂交起源的高山松(*Pinus densata*)及其亲缘种油松(*P. tabulaeformis*)和云南松(*P. yunnanensis*)为研究对象,利用外显子组捕获测序技术分析物种间的渐渗事件发现在高山松与云南松的生态分布过渡区,检测到 57 849 (3.4%) 个 SNPs 存在异常的渐渗模式,表明了这些 SNPs 位点可能在适应性或生殖隔离中发挥着重要作用。研究结果说明了生态选择在物种过渡区的基因组异质性和遗传屏障的形成中发挥了重要作用^[9]。

1.3 青藏高原

1.3.1 青藏高原范围

被誉为“世界屋脊”和地球“第三极”的青藏高原(Qinghai-Tibet Plateau, QTP)是世界上面积最大和海拔最高的高原。青藏高原西起帕米尔高原,东至横断山脉东部,南至喜马拉雅山脉南缘,北至昆仑山和祁连山北侧^[19,20]。目前,已知青藏高原至少有 1 500 个属和 12 000 种维管植物,其中特有种占 20%以上,是世界生物多样性热点地区之一^[21-23],尤其是喜马拉雅山脉和横断山脉,是北半球物种多样性最丰富的地区,包含许多特有属,例如:杜鹃花属(*Rhododendron*)、马先蒿属(*Orobanchaceae*)、紫堇属(*Papaveraceae*)、龙胆属(*Gentianaceae*)、嵩草属(*Cyperaceae*)、针茅属(*Poaceae*)、报春花属(*Primulaceae*)和雪莲属(*Asteraceae*)等^[21-24]。

1.3.2 青藏高原历史地质发展

青藏高原地区以其地形多样性而闻名,存在一些世界上最崎岖的山脉^[25,26]。在历史发展过程中,青藏高原持续性的地质活动和伴随的气候振荡共同塑造了多样化的环境和栖息地^[21,27]。喜马拉雅山脉从 56 个百万年前(Million years ago, Mya)的古新世(Paleocene)晚期的海拔约 1 km 缓慢增长到中新世(Miocene)早期(21 - 19 Mya)的 2.3 km,然后在中新世中期(~15 Mya)迅速上升到高海拔(~5.5 km),超过冈底

斯山脉。现阶段的海拔高度已达到约 6 km^[28]。此外，青藏高原从古新世到第四纪一直存在着气候振荡。例如，在中新世早期，由于北边的青藏高原的地质抬升和中亚的温度下降，东亚季风在该地区形成^[29,30]。在第四纪，在高海拔的青藏高原地区出现了明显的气候振荡，并伴随着温度升高和温度降低的循环^[31]。

1.3.3 青藏高原物种间渐渗研究实例

青藏高原是大多数温带类群的起源地，因此该地区被称为新物种的“摇篮”^[32,33]。青藏高原范围内明显的高山与深谷在很短的直线距离上呈现交替出现的模式，形成了独特的“天空岛”^[28,34-37]。由此而产生的地理限制一直被认为是该地区物种内部和物种之间遗传分化的关键因素^[20,22]。

强大地理屏障导致的空间尺度上的地理阻隔在整个青藏高原是普遍存在的，但是近年来基于种群遗传学（Population genetics）方面的研究却发现了当地物种之间依旧普遍存在物种间杂交渐渗事件，例如：种群遗传学分析证明了分布在高海拔的滇南山杨（*P. rotundifolia*）和低海拔的山杨（*P. davidiana*）存在着持续的基因流^[38]；采用最大似然法和贝叶斯法构建了主要分布于北温带高山地区的梅花草属（*Parnassia* L.）的系统发育树发现横断山脉和毗邻山区的梅花草属物种中可能存在广泛的杂交和渐渗事件^[39]；运用系统地理学分析揭示了青藏高原地区的深谷中被高山隔离的一种特有的使君子科（Combretaceae）种群之间存在由于种子介导的基因流^[40]。目前，在青藏高原地区内几乎所有研究的物种多样性较高的类群中都存在种间杂交个体甚至杂交群体，这使得青藏高原地区成为研究物种间适应性渐渗的天然实验室^[20,27]。

1.4 川滇高山栎和刺叶高山栎基本概述

1.4.1 栎属物种基本介绍

栎属（*Quercus* L.）广泛分布于东南亚，中美洲以及北美洲地区，叶片形态多样化，一般雌雄同株，风媒传粉，能够频繁的杂交^[23,41,42]。栎属大约包括 400 - 500 个物种，是北半球温带与亚热带森林生态系统中适应性强和物种丰富的重要树木属之一^[43,44]。寿命长和世代周期长的栎属物种表现出对环境的广泛耐受性，是生态学、进化生物学和基因组学等研究的模式属^[45,46]，是研究森林树种遗传多样性、种群历史动态、杂交渐渗和局地适应等的理想材料^[23,42-44]。近年来，随着栎属物种基因组数据的不断积累和发展，研究者可以在有限的成本下获得高质量的全基因组水平的遗传数据，从而探究栎属物种中全基因组水平的难题^[47,48]。Leroy 等^[6]在夏栎（*Q. robur*）参考基因组测序组装完成的基础上，通过全基因组重测序数据解析了欧洲四种栎树物种（无梗栎（*Q. petraea*），*Q. robur*，柔毛栎（*Q. pubescens*），土瑞栎（*Q. pyrenaica*））在气候变下的局地适应的遗传基础，发现冰川后期环境变化引发的生物和非生物因素共同塑造了欧洲温带栎属物种繁殖障碍的现象，这些驱动因素在持续的气候变化中发挥了重要的作用，因为地中海地区的两种栎属物种（*Q. pyrenaica* 和 *Q. pubescens*）正在向北

迁移, 在更北的纬度地区与当地温带栎属物种 (*Q. petraea* 和 *Q. robur*) 发生遗传物质的交流; Fu 等^[7]测序和组装了麻栎基因组, 并使用全基因组重测序技术研究了麻栎和栓皮栎的 219 个个体的渐渗模式、重组率、适应性变异和多个同域种群的基因表达差异, 发现生活在相似环境或相近地理距离的近缘种更趋向于通过渐渗交换相同基因组区域内的遗传变异, 而这些变异因为基因组区间内重组率被抑制而更容易被自然选择完整地保留下来。研究还发现适应性渐渗可能通过引入顺式调节元件上的等位基因变异, 特别是通过转座元件插入, 实现对与应激相关的基因的表达水平的调节作用, 进而促进栎属物种生态适应^[7]; Liang 等^[49] 使用夏栎为参考基因组, 对麻栎、栓皮栎和小叶栎 (*Q. chenii*) 的 101 个个体进行重测序, 获取了 1 600 多万 SNP 研究这三个物种的分化时间始于中新世晚期, 其后各物种在更新世冰期经历了一系列的瓶颈效应, 证实选择性清除和背景选择是导致基因组变异的重要原因。

1.4.2 栎属物种中的适应性渐渗

栎属物种之间一直存在着广泛的杂交渐渗^[50,51], 是研究杂交渐渗的经典材料。栎属物种数量的可变性凸显了在一个表现出高度形态多样性的属的内部划定物种界限的挑战, 分类学家有时将其描述为“botanic horror”^[6,51-53]。遗传标记也已经证实了这些分类学上对于栎属物种的担忧。近年来, 一些研究已经使用全基因组重测序数据, 从全基因组水平上证实了一些栎属物种之间发生了适应性渐渗, 例如: 欧洲两种白栎物种 (*Q. petraea* 和 *Q. robur*)^[6]; 东亚麻栎组两种栎属物种 (*Q. acutissima* 和 *Q. variabilis*)^[7]等。

1.4.3 川滇高山栎和刺叶高山栎研究进展

川滇高山栎 (*Quercus aquifolioides* Rehd. et Wils.) 和刺叶高山栎 (*Quercus spinosa* David ex Franchet), 是冬青栎组 (*Quercus* sect. *Ilex* Loudon) 常绿树种, 广泛分布于东喜马拉雅-横断山脉及其邻近地区^[54]。近年来, 随着测序技术发展和测序成本的下降, 关于川滇高山栎和刺叶高山栎的种群遗传学和景观基因组学方面的研究取得了许多的成果, 例如: Feng 等^[55]使叶绿体基因片段和 SSRs 数据推测了川滇高山栎 (*Q. aquifolioides*)、刺叶高山栎 (*Q. spinosa*) 和毛脉高山栎 (*Q. rehderiana*) 的遗传结构和种群历史动态, 并阐明了气候振荡和环境异质性对其种群动态历史的影响; Du 等^[26]将简单重复序列标记 (Simple sequence repeats, SSRs) 和叶绿体基因片段数据相结合检测到川滇高山栎存在三个不同的遗传谱系, 其中目前西藏地区形成的川滇高山栎的 Tibet 谱系起源于 8.6 Mya, 之后部分种群迁徙到在横断山脉和川西高原后定殖, 形成了两个谱系 (HDM 谱系和 WSP 谱系); Du 等^[23]和 Liu 等^[42]分别使用 SNPs 和 SSRs 遗传数据集进行分析发现这两个栎属物种内部和物种之间都表现出了非对称的基因流, 其物种内部和物种之间的基因流主要是从川滇高山栎到刺叶高山栎, 从西部谱系到东部谱系, 由此推断出两个栎属物种从中国西南地区的青藏高原向中国东部地区的迁徙并定殖的生物地理学模式, 即“走出青藏高原 (Out-of-QTP)”; Ju 等^[56]使用

SNPs 数据进行 $\delta a\delta i$ 模型推测出川滇高山栎、刺叶高山栎和毛脉高山栎的有效种群大小 (Effective population size, N_e) 随着时间的变化趋势相似, 从中新世 (Miocene) 中期到上新世 (Pliocene) 晚期, 三个物种的 N_e 增加, 而从更新世 (Pleistocene) 到全新世 (Holocene), N_e 发生了收缩。研究还发现历史地质活动和伴随的气候动荡导致了三个栎属物种在新近纪 (Neogene) 的适应性遗传分化^[56]; Ma 等^[57]采用重测序的核基因组和质体基因组 SNPs 揭示了川滇高山栎和四川北部的刺叶高山栎种群在系统发育树上聚为了一支, 推测是由于两个物种间的渐渗导致的。

1.5 本文的研究内容, 研究目的及意义

1.5.1 研究内容

本研究基于全基因组重测序数据首先估算了川滇高山栎和刺叶高山栎谱系间和物种间的遗传多样性、遗传结构, 并推测了 N_e 、基因流以及分化时间。之后, 使用滑动窗口法计算了 F_{ST} 、XP-CLR 和 f_{dM} , 检测全基因组中的“基因组岛 (Genetic islands)”和潜在的渐渗区域。此外, 本研究还结合 f_{dM} 和基因型-环境关联分析 (Genotype-environment associations, GEAs) 方法试图检测参与物种间适应性渐渗的功能基因, 并探究了这些基因在全基因组中分布情况和基因组特征。随后将基于 *KAS2* 基因内一个适应性位点的两种等位基因 (A allele 和 G allele) 构建的两种类型质粒转入农杆菌 (*Agrobacterium tumefaciens*) 后制备成菌液。通过在本氏烟草 (*Nicotiana benthamiana*) 中注射农杆菌菌液, 利用烟草的瞬时表达特性, 解析两种等位基因在不同环境条件下对基因表达量的调控作用, 从而探究适应性渐渗对物种局地适应的影响。基于适应性 SNPs 数据, 运用梯度森林 (Gradient forest, GF) 模型预测了在快速气候变化背景下的川滇高山栎和刺叶高山栎的脆弱性区域。

1.5.2 研究目的及意义

本研究基于全基因组重测序数据进行了种群遗传学、景观基因组学和分子功能实验验证方法研究川滇高山栎和刺叶高山栎全基因组水平的适应性渐渗, 旨在:

1. 探究川滇高山栎和刺叶高山栎的种群历史动态;
2. 解析川滇高山栎和刺叶高山栎的适应性渐渗, 并对选择的候选基因上的适应性位点进行分子实验验证;
3. 预测川滇高山栎和刺叶高山栎在快速气候变化背景下的脆弱性区域。

2 材料方法

2.1 样品采集

川滇高山栎主要分布在青藏高原及其邻近区域，刺叶高山栎在青藏高原和中国的中、东部地区都广泛分布。本研究采集了 18 自然种群的 24 个川滇高山栎样品 (Qa_W : 8 个, Qa_E : 16 个) 和 18 个自然种群的 25 个刺叶高山栎样品 (Qs_W : 12 个, Qs_E : 13 个), 这些自然种群代表了两个栎属物种在中国地区的大致分布范围。川滇高山栎的采样种群全部位于青藏高原地区, 刺叶高山栎的 9 个自然种群采集于青藏高原地区, 另外 9 个种群采集于中国中、东部地区。每个种群相隔采样距离至少 30 km, 采样个体之间的间隔距离至少 100 m。川滇高山栎和刺叶高山栎样品的采样点的详细位置信息和采样点的具体信息见图 2.1 和表 2.1。

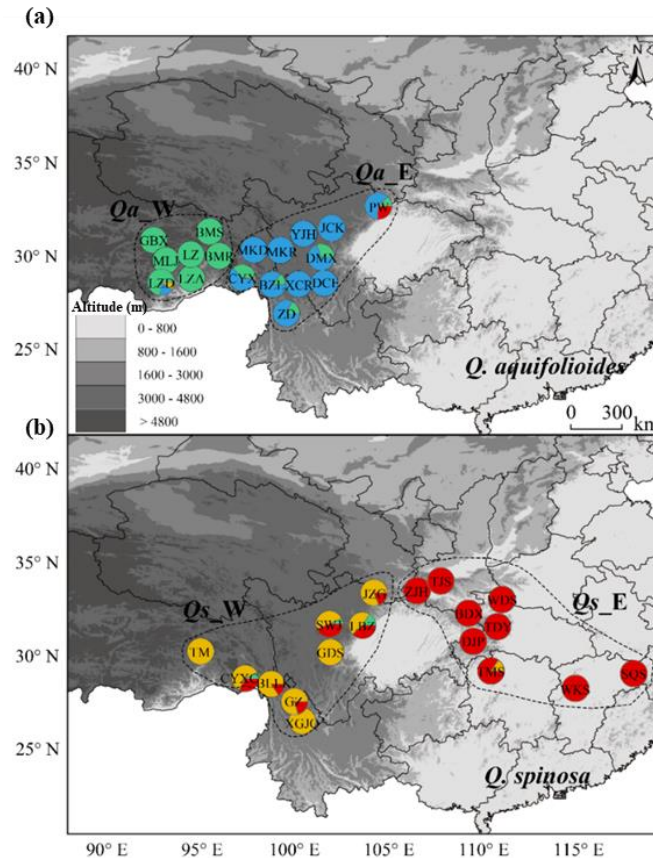


图 2.1 川滇高山栎 (a) 和刺叶高山栎 (b) 地理采样点和遗传结构, 饼图中 4 种不同的颜色代表了祖先成分, 虚线框表示了采样范围

Figure 2.1 Sampling sites and genetic clusters of *Q. aquifolioides* (a) and *Q. spinosa* (b), the four different colors represented ancestral components in bar charts and the dashed boxes indicated the sampling range

表 2.1 川滇高山栎和刺叶高山栎样品采集信息
Table 2.1 Sample information of *Q. aquifolioides* and *Q. spinosa*

种群编号 (Population code)	分布地点 (Sites)
川滇高山栎	
GBX	西藏工布达江县江达乡
LZA	西藏米林县林芝机场
MLJ	西藏米林县甲格村
LZ	西藏林芝县
LZD	西藏隆子县斗玉乡
BMR	西藏波密然乌镇
BMS	西藏波密县索通村
CYX	西藏察隅县雄久村
MKD	西藏芒康县登巴村
MKR	西藏芒康县如美镇
BZL	云南奔子栏乡
ZD	云南中甸
XCR	四川乡城县然乌温泉
DCE	四川稻城县俄洛村
YJH	四川雅江红龙乡
DMX	四川康定县明希乡
JCK	四川金川县卡撒乡
PW	四川平武县
刺叶高山栎	
TM	西藏波密县通麦
CYXC	西藏察隅县县城
BLLK	西藏察隅县白隆拉卡
XGJQ	云南省丽江雄谷景区
GZ	云南省香格里拉高山植物园
GDS	四川康定县郭达山
SWJ	四川金川县撒瓦脚
LBZ	四川汶川县萝卜寨
JZG	四川九寨沟县
ZJH	陕西勉县张家河林场
TJS	陕西周至县铁甲树
BDX	陕西平利县八道乡
TDY	湖北神农架土地垭
DJP	湖北神农架断疆坪
WDS	湖北十堰市武当山
TMS	湖南张家界天门山
WKS	江西武宁县钨矿山
SQS	江西玉山县三清山

2.2 重测序数据过滤和获取

使用 TIANGEN 植物基因组 DNA 提取试剂盒（天根，中国）对所有个体叶片进行了 DNA 提取，对提取失败或者质量不佳的样品再次使用十六烷基三甲基溴化铵法（Cetyltrimethylammonium bromide, CTAB）提取合格质量的 DNA^[58]。所有质量合格的 DNA 根据二代测序（Next generation sequencing, NGS）文库构建要求检测 DNA 的浓度，DNA 总量大于 1 μ g 的样品在 DNBSEQ 平台进行 DNA 小片段文库构建。在 BGISEQ-500 平台对每个合格的文库进行全基因组重测序。采用双端（Paired-end）150 bp 的测序策略，每个个体的测序深度为 30 \times 。然后将获取的原始序列数据（Raw reads）使用 SOAPnuker ver. 1.5.4 进行质量控制^[59]，具体质量控制原则如下：

- （1）去除包含测序接头的 reads；
- （2）去除低质量碱基比例（base 质量小于或等于 5）大于 50% 的 reads；
- （3）去除未知碱基（N 碱基）比例大于 10% 的 reads。

使用序列比对软件 BWA（Burrows-Wheeler Aligner，<http://bio-bwa.sourceforge.net/bwa.shtml>）^[60]将经过上述质量控制得到的无错数据（Clean data）比对到川滇高山栎参考基因组上，具体参数为：“-t 4 -M -R”。数据比对完成后，对数据中重复的 reads 进行删除，同时去除 mapQ 值小于 30 的 reads（参数设置：samtools view -q 30）。为了降低后续变异检测的错误率，使用 GATK（<https://software.broadinstitute.org/gatk/>）^[61]过滤重复比对的 reads。之后，使用 GATK 软件对每个个体进行 SNPs 变异检测，具体参数为：-glm BOTH -T stand_call_conf 50.0 -dcov 1000 -A Coverage -A AlleleBalance。最终，所有样品的数据储存在.vcf 文件中，主要包含所有样品的 SNPs 和插入/缺失（Insertions-deletions, Indels）。对所有的变异位点进行硬过滤（hard filtration），具体参数设置为：-T VariantFiltration -R genome.fa -V AllSample.genotype.raw.snp.vcf.gz --filterExpression "MQ0 >= 4 andand ((MQ0 / (1.0 * DP)) > 0.1)" --filterName "HARD_TO_VALIDATE" --filterExpression "DP < 5" --filterName "LowCoverage" --filterExpression "QUAL < 30.0" --filterName "VeryLowQual" --filterExpression "QUAL > 30.0 andand QUAL < 50.0" --filterName "LowQual" --filterExpression "QD < 1.5" --filterName "LowQD" -o AllSample.genotype.filter.snp.vcf.gz。经过一系列过滤后我们仅保留 SNPs 数据于 VCF 文件中进行后续分析。

VCF 文件中会存在低质量的 SNPs，需要进一步过滤来减少对后续分析的影响，具体过滤标准如下：

- （1）去除缺失率 > 10%（-geno 0.1）或最小等位基因频率（Minor allele frequency, MAF）< 0.05（-maf 0.05）的 SNPs；
- （2）去除每个物种中与哈迪-温伯格（Hardy-Weinberg, HW）平衡显著偏差小于 0.0001（-hwe 0.001）的 SNPs；

(3) 去除基因型的质量得分 (GQ) < 30 的 SNPs;

(4) 为了减少强连锁不平衡 (Linkage Disequilibrium, LD) 对 SNPs 的影响, 在 PLINK v.2.0 中 (<http://www.cog-genomics.org/plink2/>)^[62] 使用参数“-indep 50 10 0.2”过滤强连锁 SNPs 位点。

2.3 种群遗传学分析

2.3.1 遗传多样性

运用 VCFTOOLS ver. 0.1.15^[63] 进行 10 kbp 非重叠滑动窗口法 (window size = 10 kbp, window step = 10 kbp) 计算川滇高山栎和刺叶高山栎每个谱系的全基因组水平的 π 和 Tajima' D^[64]。此外, 为了评估每个谱系的基因组水平的 LD, 使用 PopLDdecay ver. 3.40^[65] 计算 100 kbp 非重叠窗口中成对 SNP 之间的平方相关系数 (r^2), 平均所有 100 kbp 窗口内的结果后进行可视化。

2.3.2 种群结构

基于马尔可夫链蒙特卡罗 (Markov chain monte carlo, MCMC) 算法的 ADMIXTURE ver. 1.3.0^[66] 软件可以模拟后验分布, 进而推测种群等位基因频率和祖先比例。该分析 K 值设置为 1 - 10, 计算交叉验证 (Cross validation, CV) 误差中最小的 CV 值对应的 K 值为最优的聚类枝, 并在 R 语言中绘制百分比堆积图。使用 R 语言中的“adegenet”程序包^[67] 执行的主成分分析 (Principal components analysis, PCA) 分析种群的遗传结构, 然后在 R 语言中对前两个主成分 (PC1, PC2) 绘制散点图。

2.3.3 系统发育树构建

本研究使用 VCF2Dis ver. 1.42 (<https://github.com/BGI-shenzhen/VCF2Dis>) 软件以邻接法 (Neighbour joining, NJ) 构建系统发育树。同时, 还使用 TreeMix ver. 1.13^[68] 以最大似然法 (Maximal likelihood, ML) 构建系统发育树, 具体参数为: “-se -bootstrap -k 1000 -m 5”, 然后使用 R 语言中“OptM”软件包推断最优的 m 值。最后利用 iTOL (Interaction Tree of Life; <https://itol.embl.de/>) 在线工具可视化上述分析构建的 NJ 树和 ML 树。

2.3.4 种群历史动态

使用成对序列马可夫共祖先 (Pairwise sequentially markovian coalescent, PSMC)^[69] 对川滇高山栎和刺叶高山栎每个谱系进行种群历史动态重建, 具体参数为: - N25 - t15 - r5 - p " 4 + 25 * 2 + 4 + 6"。本研究选择了每个谱系中测序深度最大的个体执行 PSMC 分析, 以保证对 N_e 的估算的准确性。对于每个个体的 PSMC 分析进一步进行了 100 次 bootstrap 估计。这里参考了 Sork 等^[70] 的标准, 设置世代周期 (Generation time) 为 30 年, 突变率 (Substitution per site per year) 为 1.0×10^{-8} 。

由于 PSMC 对于距今一万年以内的 N_e 的估算稳定性较差, 本研究进一步使用了基于种群位点频谱 (Site frequency spectrum, SFS) 的 Fastsimcoal2^[71] 分析试图模拟了

种群进化历史。首先，easySFS.py 脚本 (<https://github.com/isaacovercast/easySFS>) 被用于将 VCF 格式的文件转化为位点频谱文件格式 (.sfs)。这里设置了 8 个种群进化历史模型，所有的预设模型开始于川滇高山栎和刺叶高山栎的共同祖先种群。为了获取每个预设模型的参数模拟结果，对每个模型进行 50 次迭代，每次迭代模拟 10 万次，最后选取最大似然值 (MaxlogEst_hood) 最大的一次迭代结果作为该模型的参数模拟结果。同时，通过计算池信息量准则 (Akaike information criterion, AIC) 权重来确定最优模型。最后，使用最优模型生成 100 组模型参数文件，对 100 个生成模型再次进行 40 模型运行计算 95 % 置信区间。

2.3.5 基因组岛检测

为了估算全基因组水平谱系间和物种间的遗传分化，使用 10 kbp 非重叠滑动窗口法比较不同种群对之间的 SNPs 等位基因频率，以此来检测川滇高山栎和刺叶高山栎之间，川滇高山栎谱系之间 (Qa_W vs. Qa_E) 和刺叶高山栎谱系之间 (Qs_W vs. Qs_E) 在整个基因组中的“基因组岛 (Genetic islands)”。同时计算了每个窗口的 F_{ST} 和 XP-CLR 统计量。本研究标准化了 F_{ST} 分析中的每个窗口的 F_{ST} ($Z - F_{ST}$)，并定义在基因组中 $Z - F_{ST} \geq 2$ 的窗口为“基因组岛”。对于 XP-CLR 统计量，本研究定义 XP-CLR 统计量前 5 % 的窗口为“基因组岛”。尽管两种方法都可以检测物种内部和物种之间的全基因组中的遗传分化区域，但是两种方法同时检测到的重叠区域更可能是遗传分化的足迹。使用 R 语言中“topGO”^[72] 程序包对这些重叠的“基因组岛”进行基因本体论 (Gene ontology, GO) 分析，并将错误发现率 (False discovery rate, FDR) 小于 0.05 的定义为显著的 GO 富集分析结果。此外，本研究还计算了其他多组种群对之间的 F_{ST} ，进一步探究物种内和物种间的遗传分化。这些种群对包括了川滇高山栎的所有种群和刺叶高山栎的所有谱系之间的比较 ($Q. aquifolioides$ vs. Qs_W , $Q. aquifolioides$ vs. Qs_E)，刺叶高山栎的所有种群和川滇高山栎的所有谱系之间的比较 ($Q. spinosa$ vs. Qa_W , $Q. spinosa$ vs. Qa_E)，川滇高山栎的所有谱系和刺叶高山栎的所有谱系之间的比较 (Qa_W vs. Qs_W , Qa_W vs. Qs_E , Qa_E vs. Qs_W , Qa_E vs. Qs_E)，以及青藏高原地区种群和非青藏高原种群之间的比较 (QTP vs. non-QTP)。

2.3.6 基因渐渗

Dsuite^[73] 软件可以运用来区分基因渐渗和不完全谱系分选 (Incomplete lineage sorting, ILS)。首先，使用 Fbranch 程序计算每个系统发育分支的 f -branch 统计量。然后，使用 Dtrios 程序估计 D 和 f_4 -ratio 统计量，最后使用重叠滑动窗口法 (window size = 50 SNPs, window step = 20 SNPs) 计算全基因组中每个窗口的 f_{dM} 值，从而检测基因组中的潜在渐渗区域。依据 Morales-Cruz 等^[74] 的计算方法，将前 X % 的 f_{dM} 值的窗口定义为潜在渐渗区域，其中 X 选择 f_4 -ratio 统计量估计的渐渗比例决定。为了解析基因组中潜在的渐渗区域的特征，研究中通过比较基因组中的潜在渐渗区域和整个基因组背景的物种间的 F_{ST} , D_{XY} 值和物种内的 π 值，并可视化了渐渗区域内的 SNPs

在 1 Mb 窗口内的密度分布。使用 R 语言中“topGO”^[72]程序包对渐渗区域内的功能基因进行了 GO 富集分析，以此来确定渐渗区域中是否能够富集到应对非生物胁迫的结果。

2.4 景观基因组学分析

2.4.1 生物气候因子的获取

本研究所使用的生物气候数据均来自于 WorldClim (<http://worldclim.org>)^[75]数据库。这些气候数据包括 1970 - 2000 年分辨率为 30 seconds 的 19 个生物气候变量(Bio01 - Bio19)。其中 Bio01 - Bio11 代表与温度相关的生物气候变量，Bio12 - Bio19 代表与降水相关的生物气候变量。此外，本研究还下载了 2050 年和 2070 年不同路径浓度 (Representative concentration pathway, RCP) 下的 19 个气候数据，包含了 RCP 2.6 和 RCP 8.5 两种碳排放情景。

2.4.2 基因型-环境关联分析

使用两种基因型-环境关联分析 (Genotype-environment associations, GEAs) 来识别基因组中受到环境变化影响的适应性遗传变异位点和功能基因，分别是潜在因素混合模型 (Latent factor mixed model, LFMM)^[76]和冗余分析 (Redundancy analysis, RDA)^[77]。

LFMM 分析是通过 R 语言中“LEA”^[78]程序包实现，可以考虑种群遗传结构带来的影响，能识别生物气候变量和等位基因频率之间的显著关系，从而确定受到环境变化影响的异常 SNPs。该分析运用的气候变量为对 19 个生物气候因子进行了 PCA 分析后的 PC1，是关于与最暖季度平均气温 (Bio10)、年降水量 (Bio12) 以及最干月份降水量 (Bio14)。其余参数设置为 MCMC 进行 10 次独立运算，并将 iterations 设置为 100000，burnin 设置为 50000。

RDA 分析运用 R 语言中“vegan”^[79]程序包进行检测受到环境变化影响的遗传变异位点。该分析的气候变量通过使用 R 语言中“usdm”^[80]包对 19 个气候变量进行方差膨胀因子 (VIF) 分析，删除了共线性阈值大于 0.7 的生物气候变量，最终保留了 6 个生物气候变量，分别是：等温线 (Bio03)、年温年较差 (Bio07)、最湿季度平均温度 (Bio08)、最湿月份降水量 (Bio13)、最干月份降水量 (Bio14) 和降水量季节性变化 (Bio15)。

这两种方法都评估了每个 SNP 与生物气候变量之间的相关性，本研究中将 $FDR \leq 0.05$ 的 SNPs 推定为可能是由于环境因素导致的遗传变异。然后在 R 语言中可视化了两种 GEA 方法共同检测到的适应性 SNPs 在 1 Mb 窗口内的密度分布，并对存在这些 SNPs 的功能基因进行了 GO 富集分析，从而解析物种的局地适应 (Local adaptation) 机制。

2.4.3 遗传结构和环境梯度的多元关系

(1) 线性分析

Mantel 和偏 Mantel 测试被用于检测青藏高原种群和非青藏高原种群中存在的地理隔离 (Isolation by distance, IBD) 和环境隔离 (Isolation by environment, IBE)^[81] 对于遗传变异的影响。偏 Mantel 测试允许通过控制环境因素的影响来检测 IBD, 通过控制地理因素的影响来检测 IBE。分析过程中所需的遗传距离矩阵 (F_{ST} 矩阵) 通过 genepop 软件实现, 地理距离矩阵是使用 R 语言中“geosphere”^[82] 程序包计算而来, 环境距离矩阵是通过 R 语言中“ecodist”^[83] 程序包计算而来。此外, 研究中还探究了青藏高原种群和非青藏高原种群之间的遗传距离矩阵、地理距离矩阵和环境距离矩阵之间的相关性。

(2) 非线性分析

在 R 语言中使用“gdm”程序包实现的广义相异建模 (Generalized dissimilarity modeling, GDM)^[84] 可以解析青藏高原种群和非青藏高原种群内部的遗传变异与地理因素或生物气候变量之间的非线性关系。GDM 分析允许通过样条拟合 (I-spline) 估算等位基因频率沿环境梯度变化的比率, 并根据重要性百分比解释模型拟合的程度^[85]。

2.4.4 适应性渐渗

结合 f_{dM} 值和 GEAs 检测方法, 本研究试图检测基因组中参与物种间适应性渐渗的功能基因。这些参与适应性渐渗的功能基因都在基因组的渐渗区域内且至少含有一个适应性变异位点, 然后在 R 语言中可视化了这些参与适应性渐渗的功能基因的序列长度和位置分布状况。接下来比较这些功能基因与基因组中其他类型基因的长度, 并分析这些功能基因中适应性 SNPs 数量与基因自身长度、基因内所有内含子长度总和以及基因内所有外显子长度总和之间的关系。此外, 研究中还提取了参与适应性渐渗的基因的适应性 SNPs 上、下游 20 bp 的碱基序列, 使用 Plant Cis-Regulatory DNA Elements Database 在线植物顺式调控 DNA 元件数据库 (<https://www.dna.affrc.go.jp/PLACE/?action=>) 预测植物中可能的 DNA 顺式调控元件。

2.4.5 分子功能实验验证

本研究选择了 *KAS2* (*evm.model.scaffold92.49*) 一个内含子中 GT1CONSENSUS 调控元件上的一个适应性位点 (3_82758545) 探究该适应性遗传变异如何影响物种应对来自环境变化的选择性压力。首先, 计算了 *KAS2* 基因内的物种间 F_{ST} 和 D_{XY} , 以及物种内 π 值。使用未经过连锁不平衡过滤的 SNPs 数据集, 计算了 *KAS2* 基因及其周围区域内所有两两 SNPs 之间的连锁不平衡强度, 并在 R 语言中进行可视化。将包含两种基因型的顺式调控元件 (GAAAAT, GAAGAT) 通过五次串联重复后, 分别拼接到 35S mini promoter 和 *INF1* 基因。最终构建了两种类型的质粒: (GAAAAT)₅ - 35S mini :: *INF1* 和 (GAAGAT)₅ - 35S mini :: *INF1*。将两种类型质粒分别转入农杆菌 (*Agrobacterium tumefaciens*) EHA105 感受态细胞, 之后浸入培养基 (10 mM MgCl₂, 10 mM MES, 200 mM acetosyringone) 中重悬。将农杆菌菌液在室温下避光培养 3 h,

用 1 mL 无针注射器分别将含有(GAAAAT)₅ - 35S mini :: *INF1* 的 *A. tumefaciens* 质粒的农杆菌菌液和含有(GAAGAT)₅ - 35S mini :: *INF1* 的 *A. tumefaciens* 质粒的农杆菌菌液注射到新鲜的本氏烟草 (*Nicotiana benthamiana*) 叶片两侧。为了研究两种基因型与光照、低温和紫外线的关系, 设计了六种类型的实验条件:

- (1) 16 h 光照 (25°C) / 8 h 避光 (20°C) (对照组);
- (2) 24 h 避光 (20°C);
- (3) 16 h 光照 (25°C) / 8 h 避光 (20°C) 持续 30 h, 之后 24 h 避光 (20°C);
- (4) 24 h 避光 (20°C) 持续 30 h, 之后 16 h 光照 (25°C) / 8 h 避光 (20°C);
- (5) 16 h 光照 (25°C) / 8 h 避光 (4°C);
- (6) 24 h 紫外线辐射 (25°C)。

每种环境的实验至少进行三次实验重复。

2.4.6 未来命运预测

在 R 语言中使用“gradientForest”^[86]程序包实现的非参数机器学习回归树方法的梯度森林 (Gradient forest, GF) 用于估算未来气候变化对川滇高山栎和刺叶高山栎生态分布区域内的遗传偏移 (Genetic offset), 该分析是基于当前气候和未来气候之间的欧几里得距离计算遗传偏移, 并能利用 ArcGIS 10.2 绘制两个栎属物种生态地理分布图。本研究运用此分析同时预测了 2050 年和 2070 年在两种不同代表性碳排放情景 (Representative concentration pathway, RCP) 下的川滇高山栎和刺叶高山栎的遗传偏移区域, 包括了低排放情景 RCP 2.6 和高排放情景 RCP 8.5。其中分析过程中所使用的生物气候变量是基于 GDM 分析结果中分别对青藏高原种群和非青藏高原种群的生态适应影响最大的生物气候变量作为背景变量。此外, 本研究还通过去除适应性渐渗快速获取的遗传变异作为输入文件执行 GF 模型预测, 探究适应性渐渗如何驱动物种应对快速的气候变化。

3 研究结果

3.1 全基因组重测序和变异位点检测

川滇高山栎和刺叶高山栎的测序深度范围在 24.2× - 47.4×，并且 85 % 以上个体的测序深度超过 30 ×。在测序过程中，准确度在 99 % 以上 (Q20) 的碱基识别率为 94.2% - 97.3%，平均值为 95.4%。测序数据的 GC 含量范围为 36.6 % - 41.1 % (附录 F)。使用 SOAPnuke ver. 1.5.4 进行硬过滤 (Hard filtration) 后，共计产生了约 1.85 Tb 的原始数据，共计获得了 2 170 823 个 Indels 和 18 039 605 个 SNPs，其中 Indels 中包含 1 073 905 个插入和 1 087 088 个缺失。最终分别在川滇高山栎和刺叶高山栎中保留了 358 661 个 SNPs 和 346 281 个 SNPs 用于后续分析。

3.2 物种内和物种间基因组特征

3.2.1 遗传多样性分析

10 kbp 非重叠滑动窗口法计算的川滇高山栎和刺叶高山栎的遗传多样性结果显示 Qa_E 谱系表现出最高水平的 π ， Qs_E 谱系的 π 值最低，各谱系的 Tajima' D 统计量均为正，表明两个栎属物种最近均经历了种群收缩 (图 3.1)。此外，100 kbp 滑动窗口法计算的连锁不平衡 (LD) 发现 Qa_E 和 Qs_W 表现出快速的 LD 衰退模式， Qa_W 和 Qs_E 谱系表现出缓慢的 LD 衰减模式 (图 3.1)。

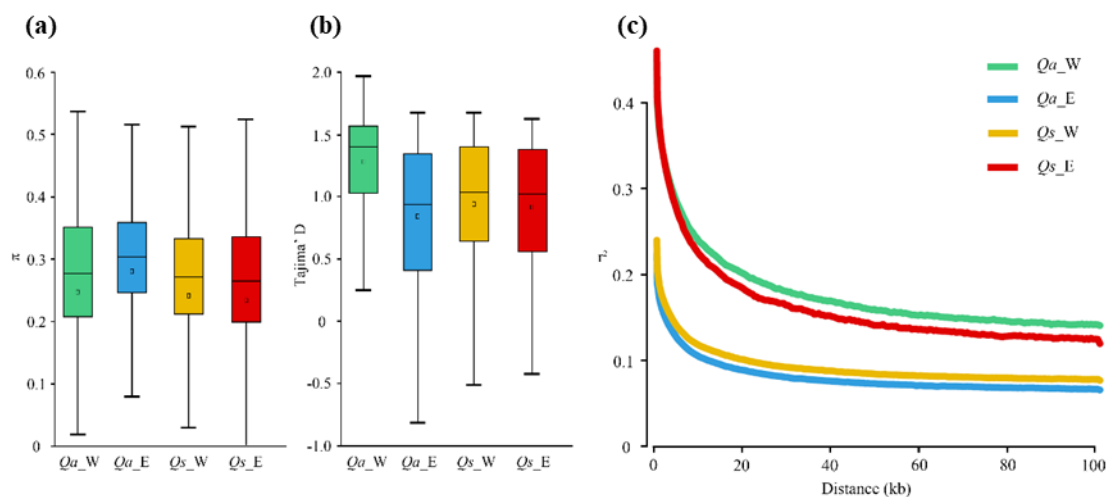


图 3.1 川滇高山栎和刺叶高山栎每个谱系的遗传多样性。10 kbp 滑动窗口法计算的 (a) π 和 (b) Tajima' D; (c) 使用 100 kbp 非重叠滑动窗口计算的每个谱系的连锁不平衡
Figure 3.1 Genetic diversity of *Q. aquifolioides* and *Q. spinosa*. (a) π and (b) Tajima' D of each lineage of *Q. aquifolioides* and *Q. spinosa* was calculated using 10 kbp non-overlapping sliding window; (c) Linkage disequilibrium (LD) for each lineage was calculated using a 100 kbp non-overlapping sliding window

3.2.2 遗传结构

ADMIXTURE 分析结果显示 K 等于 2 是川滇高山栎和刺叶高山栎物种的最优聚

类枝；当 K 等于 3 时，刺叶高山栎种群内部进一步分化为 Qs_W 和 Qs_E 两个谱系；当 K 为 4 时，川滇高山栎和刺叶高山栎内部均被分化为两个谱系 (Qa_W , Qa_E ; Qs_W , Qs_E) (图 3.2)。值得注意的是，当 K 等于 4 时，只有 Qs_E 位于青藏高原地区以外的地区 (non-QTP)，而其他三个谱系 (Qa_W , Qa_E , Qs_W) 位于青藏高原地区 (QTP)。基于遗传距离矩阵的主成分分析 (PCA) 的结果与上述的 ADMIXTURE 的结果一致。PC1 (15.0 %) 很好的区分了川滇高山栎和刺叶高山栎，以及川滇高山栎内部的两个谱系，PC2 (8.0 %) 进一步区分了刺叶高山栎的两个谱系。

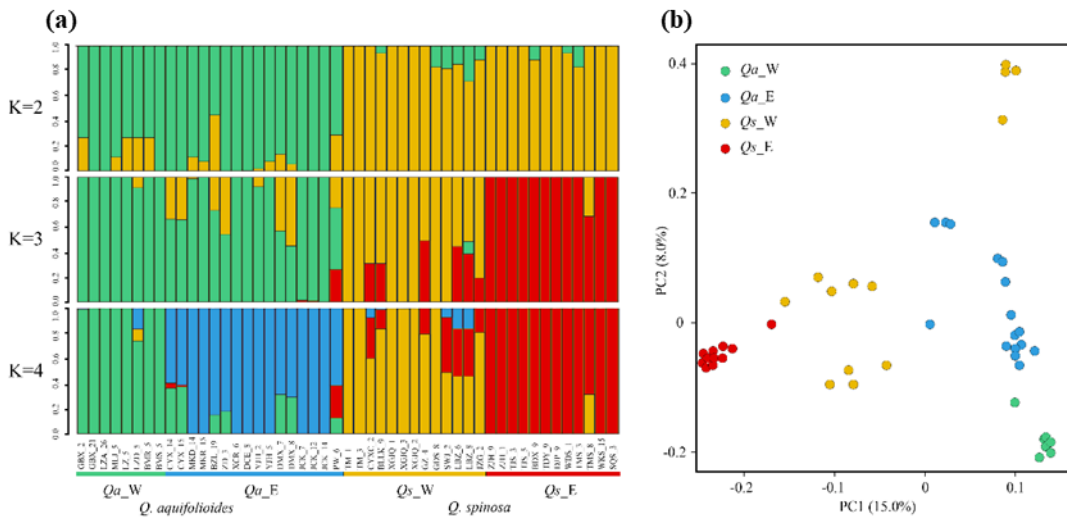


图 3.2 川滇高山栎和刺叶高山栎的遗传结构。(a) ADMIXTURE 种群结构，每个条形代表一个个体，条形部分的颜色取决于假设 K 为 2、3 或 4 时的祖先比例；(b) 主成分分析，每个点表示一个个体，每种颜色表示一种谱系

Figure 3.2 The genetic structure of *Q. aquifolioides* and *Q. spinosa*. (a) Model-based population assignment by ADMIXTURE analysis for K=2 to 4. Each bar represents a single individual, with portions of the bar colored depending on the ancestry proportions estimated assuming K=2, K=3 or K=4 (b) Principal component analysis, each dot indicates an individual and each color indicates a lineage

3.2.3 系统发育关系

NJ 树和 ML 树结果进一步支持了上述分析的遗传结构，两种系统发育树都将所有个体最终聚类为四个支系，有少数个体具混合的遗传成分；进一步分析 TREEMIX 的结果，发现最佳的 TREEMIX 模型 ($m = 4$) 显示了从川滇高山栎到刺叶高山栎，从 Qa_W 谱系到 Qa_E 谱系和从 Qs_W 谱系到 Qs_E 谱系的基因流 (图 3.3)。

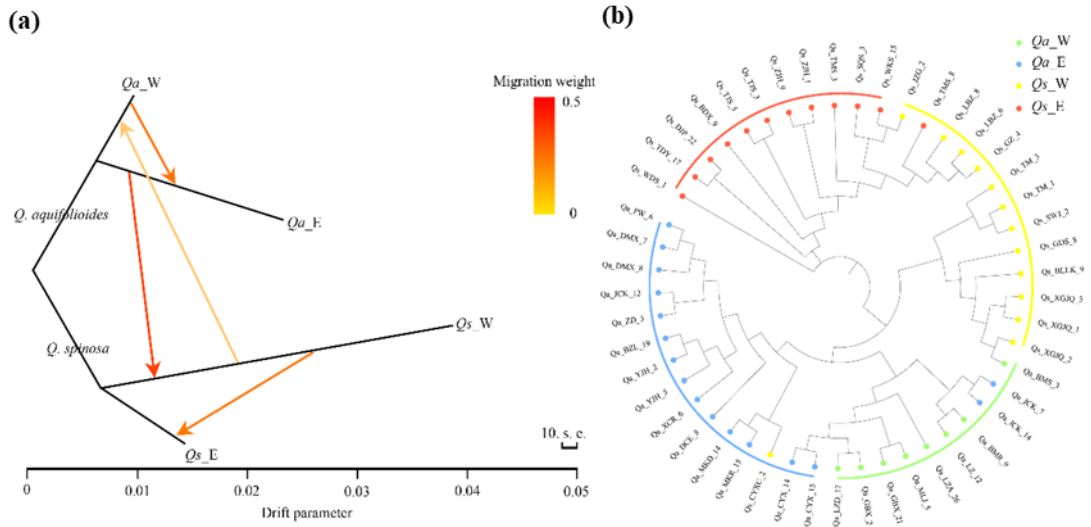


图 3.3 川滇高山栎和刺叶高山栎的系统发育树。(a) TREEMIX 推断出的最大似然系统发育树，箭头的方向和颜色表示基因流的方向和大小；(b) VCF2Dis 构建的邻接树，每个点表示一个个体，每种颜色表示一个谱系

Figure 3.3 Phylogenetic tree of *Q. aquifolioides* and *Q. spinosa*. (a) Maximum-likelihood (ML) tree inferred by TREEMIX. Arrows indicate the direction of gene flow and are colored according to the gene flow. (b) Neighbour-joining (NJ) tree constructed by VCF2Dis. Each dot indicates an individual and each color indicates a lineage

3.2.4 种群历史

PSMC 模型推测的川滇高山栎和刺叶高山栎每个谱系 N_e 的变化的结果显示四个谱系中新世 (Miocene) 开始经历了明显的种群收缩事件；在上新世 (Pliocene)，*Qa_E* 和 *Qs_W* 谱系的 N_e 保持稳定状态，而 *Qa_W* 和 *Qs_E* 谱系则出现了轻微的种群收缩事件 (图 3.4)。

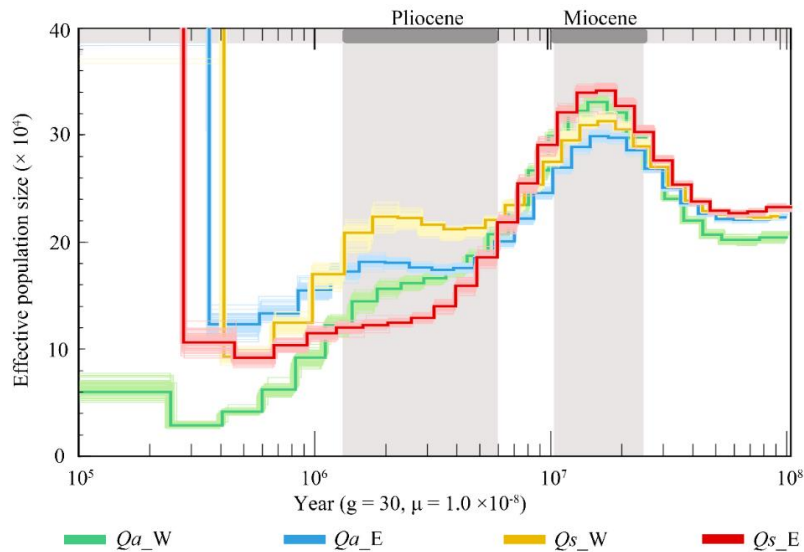


图 3.4 PSMC 推测的每个谱系的有效种群大小随时间的变化。世代周期 (g) 设置为 30 年/世代，突变率 (μ) 设置为 1.0×10^{-8} 碱基/年。横轴表示时间，纵轴表示有效种群大小

Figure 3.4 Effective population size of each lineage was inferred using PSMC. The generation time (g) of 30 years and mutation rate (μ) of 1.0×10^{-8} per site and generation. The x axis represents time and the y axis represents effective population size

Fastsimcoal2 模拟的种群进化历史模型中最佳拟合模型为 Model 7 (MaxEstLhood = -4724991.758, MaxObsLhood = -4401106.725, 附录 A, 附录 G), 模型结果显示川滇高山栎和刺叶高山栎共同祖先种群大约在 14.88 Mya (95% CI = 11.62 - 15.93 Mya) 发生物种分化事件; 随后, 在约 8.98 Mya (95% CI = 7.25 - 10.03 Mya), 刺叶高山栎祖先种群发生了物种内部谱系分化, 形成两个独立的谱系 (Qs_W , Qs_E); 在约 3.04 Mya (95% CI = 2.63 - 3.97 Mya), 川滇高山栎祖先种群也发生了物种内部谱系分化, 形成 Qa_W 和 Qa_E 两个独立谱系 (图 3.5, 附录 H)。该模型中对古代和近代基因流的估算结果显示最高水平的基因流是从 Qa_E 谱系到 Qs_W 谱系的近代基因流 (per-generation migration rate = $3.4e-5$)。

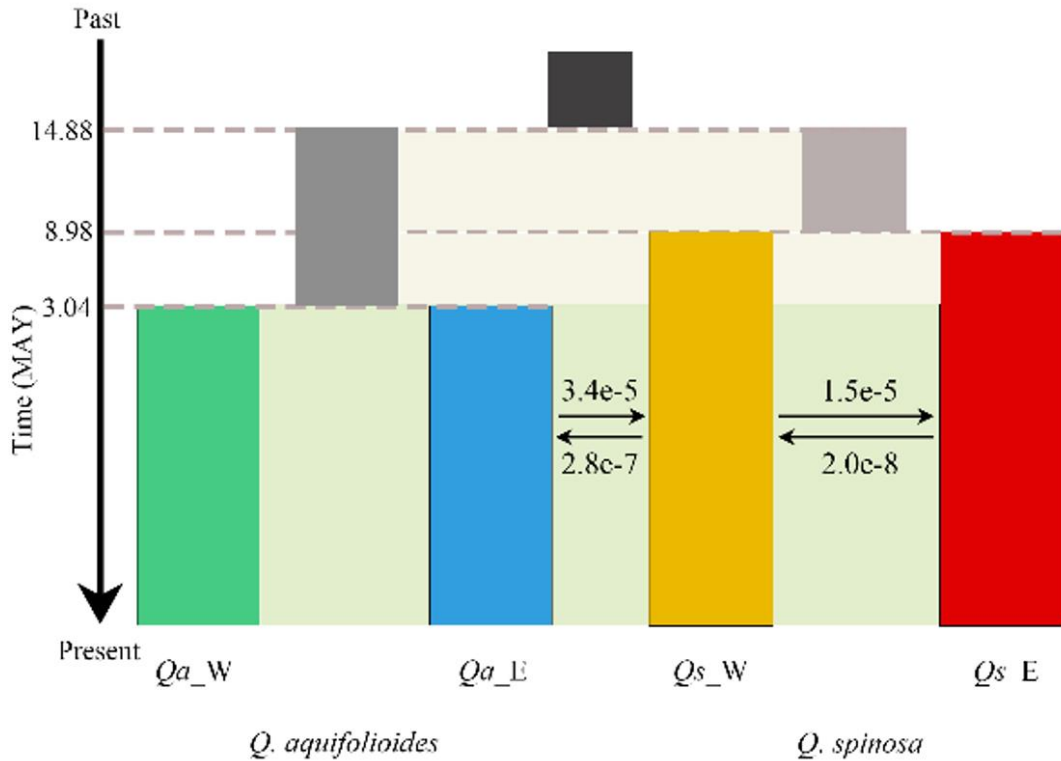


图 3.5 川滇高山栎和刺叶高山栎的种群进化历史。箭头和数字分别表示基因流的方向和大小, 浅灰色和深灰色分别表示川滇高山栎和刺叶高山栎的祖先种群, 点估计和 95% 置信区间的参数估计见附录 H

Figure 3.5 Schematic of the demographic scenario modeled of *Q. aquifolioides* and *Q. spinosa* using fastsimcoal2. Arrows and numbers indicate the direction and size of gene flow, and the ancestral populations were shown in light and dark grey for *Q. aquifolioides* and *Q. spinosa*, the point estimation and 95% confidence intervals of demographic parameters were shown in Appendix H

3.2.5 遗传分化

基于 10 kbp 非重叠滑动窗口方法检测不同种群对之间的“基因组岛”结果显示基

因组每个窗口的 F_{ST} 分别在 *Q. aquifolioides* vs. *Q. spinosa*、*Qa_W* vs. *Qa_E* 和 *Qs_W* vs. *Qs_E* 检测出 3 929、4 063 和 4 710 个高遗传分化区域，XP-CLR 方法分别在 *Q. aquifolioides* vs. *Q. spinosa*、*Qa_W* vs. *Qa_E* 和 *Qs_W* vs. *Qs_E* 检测出 3 477, 3 499 个, 3 032 个 10 kbp 高遗传分化区域 (图 3.6)。值得注意的是, F_{ST} 和 XP-CLR 同时在 *Q. aquifolioides* vs. *Q. spinosa*、*Qa_W* vs. *Qa_E* 和 *Qs_W* vs. *Qs_E* 检测出 744 个, 1 013 个, 988 个高遗传分化区域。这些重叠的高遗传分化区域更有可能代表了谱系之间和物种之间“基因组岛 (Genetic islands)”。

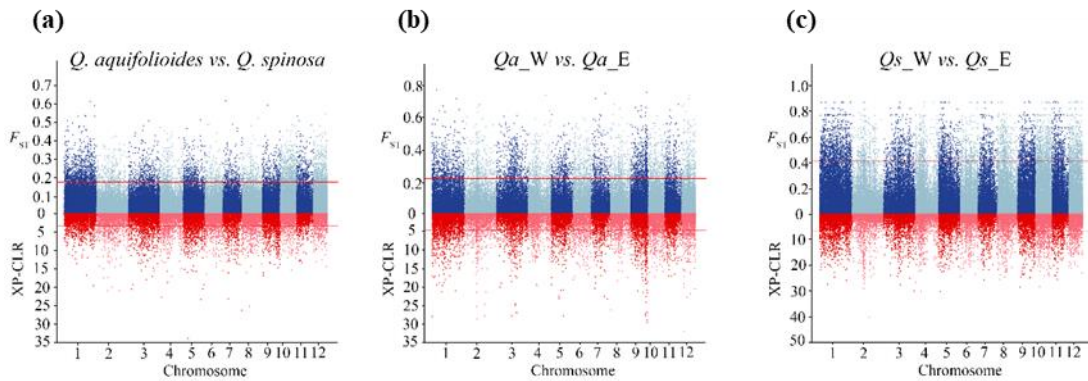


图 3.6 基于 10 kbp 非重叠滑动窗口法计算的 F_{ST} 和 XP-CLR 结果。(a) 川滇高山栎和刺叶高山栎, (b) 川滇高山栎谱系之间和 (c) 刺叶高山栎谱系之间。红色水平线为 F_{ST} 和 XP-CLR 的阈值

Figure 3.6 The result of F_{ST} and XP-CLR was calculated using the 10 kbp non-overlapping sliding window method. (a) *Q. aquifolioides* and *Q. spinosa*, (b) lineages of *Q. aquifolioides* and (c) lineages of *Q. spinosa*. Red horizontal lines are thresholds for F_{ST} and XP-CLR

对 F_{ST} 和 XP-CLR 方法同时检测出的三组“基因组岛”中的功能基因分别进行 GO 富集分析, 结果中富集到了许多参与物种生长发育的结果, 例如 GO:0008152 (metabolic process)、GO:1901137 (carbohydrate derivative biosynthetic process)、GO:0015979 (photosynthesis)、GO:0006952 (defense response) 等 (附录 I~K)。

比较多组种群对之间的 F_{ST} 结果显示物种间的遗传分化水平 (*Q. aquifolioides* vs. *Q. spinosa*: $F_{ST} = 0.07$), 刺叶高山栎内部谱系之间的表现出更高的遗传分化 (*Qs_W* vs. *Qs_E*: $F_{ST} = 0.15$) (图 3.7)。此外, 在所有 12 组种群对之间的 F_{ST} 计算结果中, *Qs_E* 谱系与其他种群之间均 (*Qs_E* vs. *Qa_W* = 0.22, *Qs_E* vs. *Qa_E* = 0.16, *Qs_E* vs. *Qs_W* = 0.15, *Qs_E* vs. *Q. aquifolioides* = 0.15, *Qs_E* vs. QTP = 0.13) 表现出了更高的遗传分化。

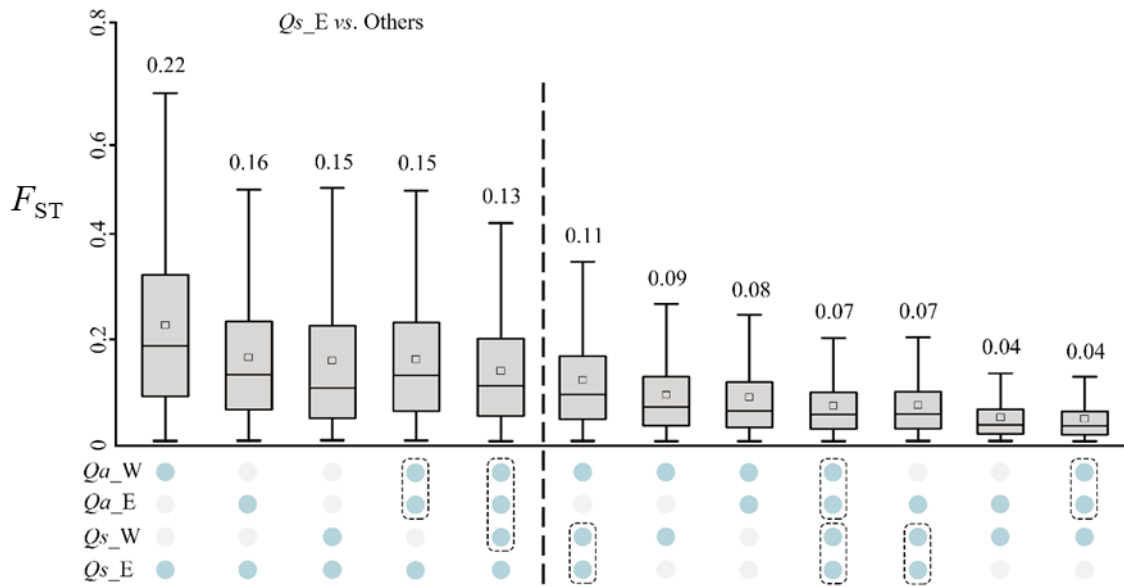


图 3.7 基于 10 kbp 非重叠滑动窗口法计算的不同种群对之间的 F_{ST} 。数字为平均 F_{ST} ，虚线框表示将两个或三个谱系作为一个整体

Figure 3.7 The result of F_{ST} was compared by different population pairs based on the 10 kbp non-overlapping sliding window method. Numbers are mean F_{ST} and dashed boxes indicate two or three lineages as a group

3.3 基因组渐渗区域检测

Dsuite 中 Fbranch 程序计算的 f-branch (f_b) 统计量显示 *Qs_E* 谱系比川滇高山栎与 *Qs_W* 之间表现出更强的渐渗；与 *Qa_W* 谱系相比，刺叶高山栎与 *Qa_E* 之间表现出更强的渐渗（图 3.8）。其中，最强的渐渗是从 *Qa_W* 谱系的 PW 种群到 *Qa_W* 谱系的 YJH 种群。此外，显著的 D 和 f_4 -ratio 统计量 ($D=0.013$, f_4 -ratio=0.110) 也发现川滇高山栎和 *Qs_W* 之间存在基因渐渗。使用重叠滑动窗口方法 (window size = 50 SNPs, window step = 20 SNPs) 计算每一个窗口内的 f_{dM} 值进一步在川滇高山栎和 *Qs_W* 谱系之间发现了 1 461 个潜在的渐渗区域（图 3.9）。

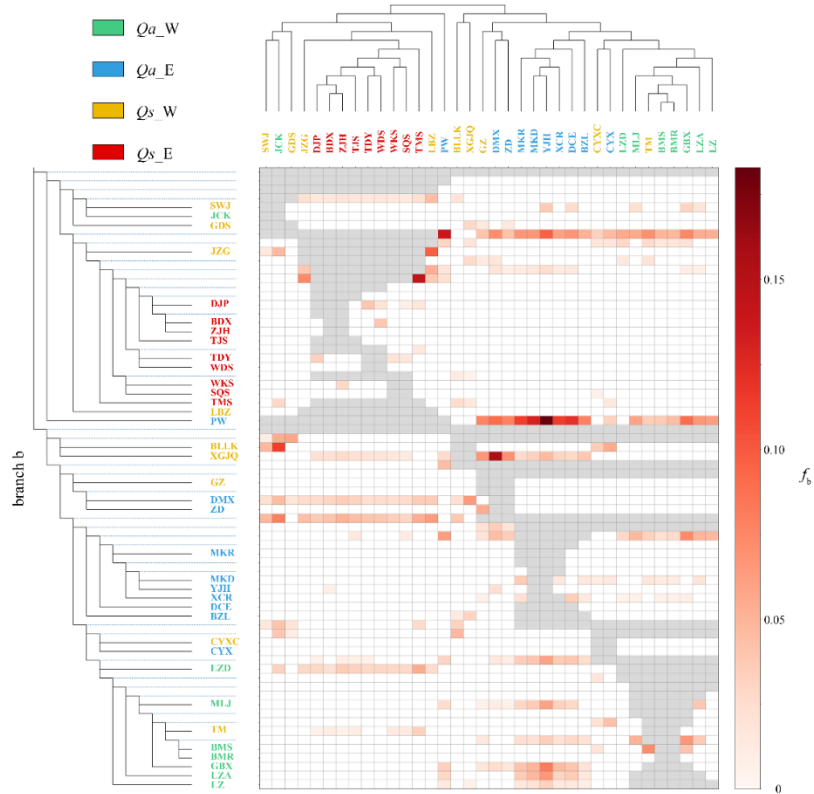


图 3.8 f-branch (f_b) 统计量表示了从 y 轴上的树分支到 x 轴上的树分支的可能的基因流。红色表示基因流的水平高，白色表示基因流水平低，灰色表示没有估算出基因流
 Figure 3.8 The f-branch (f_b) statistic showed possible gene flow from the branch of the tree on the y axis to the species or population on the x axis. The red color indicates high levels of gene flow, white color indicates low levels of gene flow, and grey color indicates that gene flow is not estimated

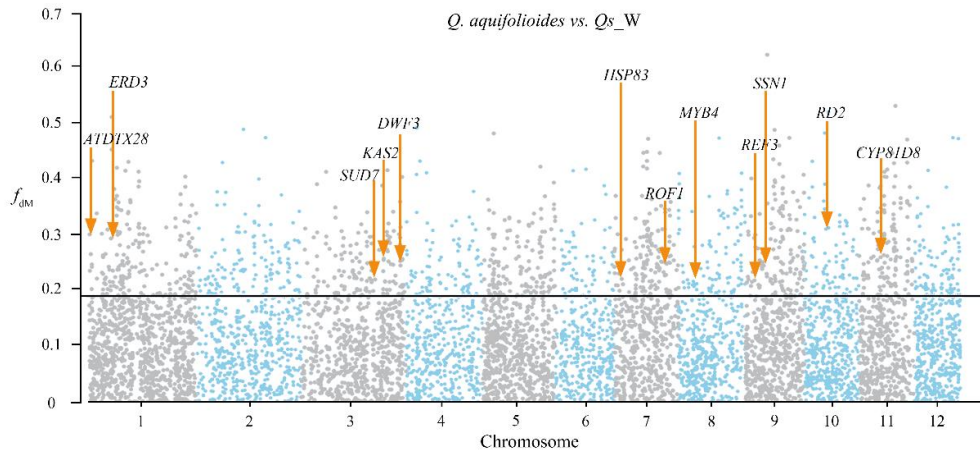


图 3.9 川滇高山栎和刺叶高山栎西谱系之间的 f_{DM} 值。黑色线条以上的区域表示潜在的渐渗区域，红色箭头显示了渐渗区域内参与局地适应的功能基因
 Figure 3.9 The f_{DM} values between *Q. aquifolioides* and *Qs_W*. The region above the black line was indicated potential introgression region, red arrows showed functional genes involved in local adaptation in the introgression region

与整个基因组背景区域（除渐渗区域外的整个基因组区域）相比，这些潜在的基因组渐渗区域表现出更低的物种间 F_{ST} 和 D_{XY} ，以及更高的物种内 π 值（图 3.10）。

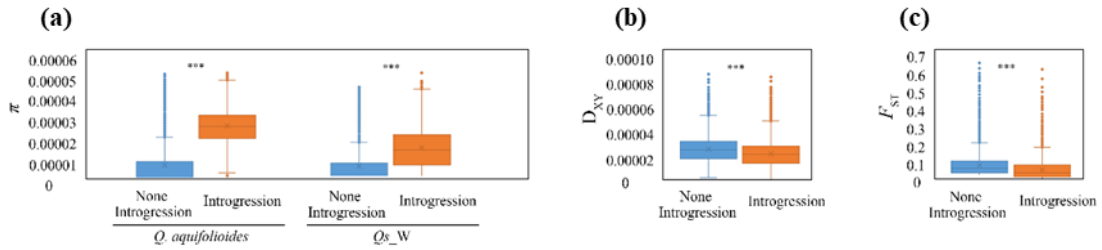


图 3.10 潜在渐渗的基因组区域（红框）和基因组背景区域（蓝框）之间的 (a) π , (b) D_{XY} 和 (c) F_{ST} 的比较。其中纵坐标显示对应的遗传多样性大小

Figure 3.10 Comparison of (a) π , (b) D_{XY} and (c) F_{ST} between potential introgression genomic regions (red boxes) and genomic background regions (blue boxes). The y axis shows the size of the corresponding genetic diversity values

川滇高山栎和 *Qs_W* 渐渗区域内的 SNPs 可视化结果显示这些 SNPs 均匀的分布在 12 条染色体上，并未表现出在任何一条染色体上的显著富集，但是 SnpEff 注释的结果发现这些位于渐渗透区域内的 SNPs 主要集中在基因间区和基因的内含子等区域（图 3.11，图 3.12）。

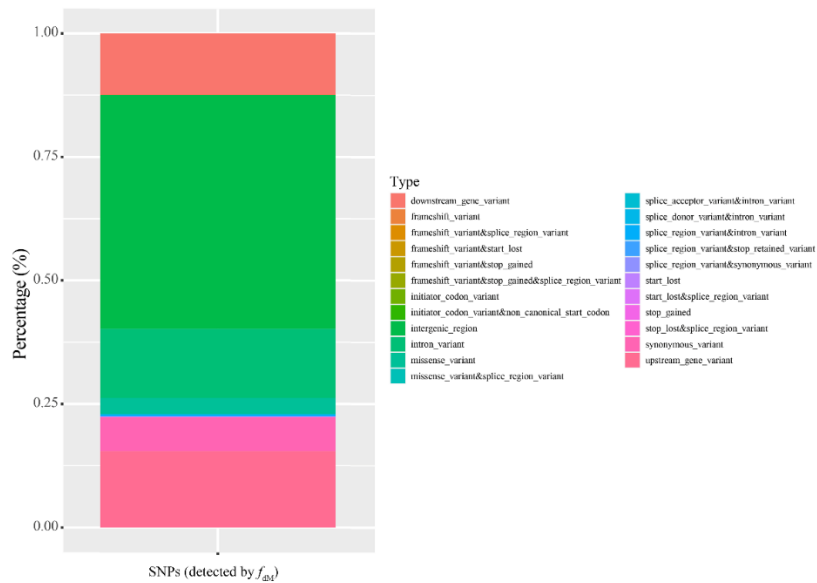


图 3.11 f_{dM} 值检测到的 SNPs 的位置分布统计

Figure 3.11 The positional distribution of SNPs were detected by f_{dM}

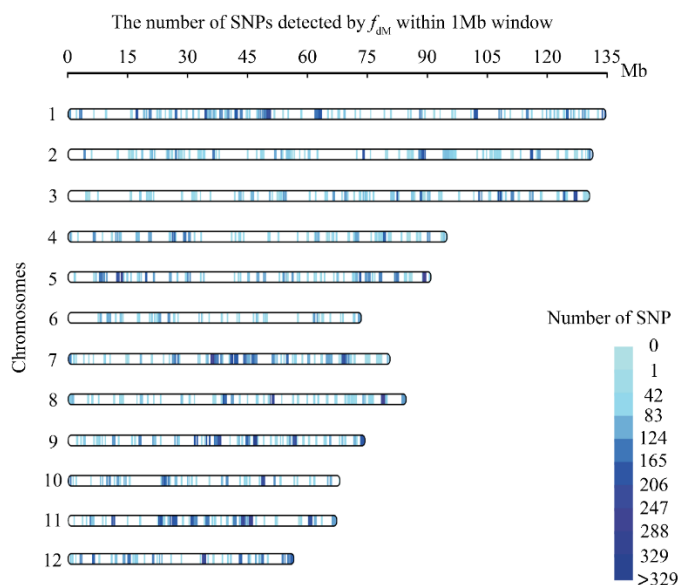


图 3.12 f_{dM} 值检测到的 SNPs 在全基因组上 1 Mb 窗口内的数量
 Figure 3.12 The number of SNPs was detected by f_{dM} within 1 Mb window across whole genome

川滇高山栎和 Q_s_W 渐渗区域内的功能基因的 GO 富集分析结果显示了一些与非生物胁迫相关的 GO 富集结果，例如：DNA repair (GO:0006281)、photosynthesis and light harvesting (GO:0009765)、recognition of pollen (GO:0048544)、reproductive processes (GO:0022414) 和 defense response (GO:0006952) 等 (图 3.13)。

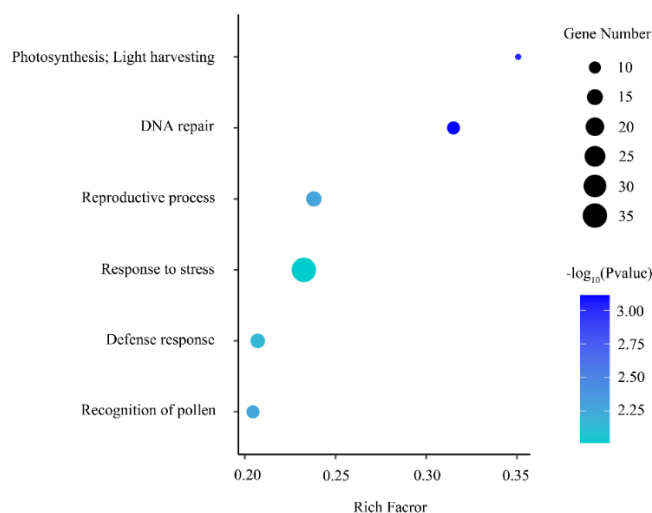


图 3.13 f_{dM} 检测到的基因的 GO 富集结果。仅显示了 $p < 0.05$ 的与非生物胁迫相关的结果
 Figure 3.13 The result of GO enrichment for genes using f_{dM} . Only those of results associated with abiotic stress with $p < 0.05$ are shown

在这些基因组渐渗区域内，一些基因在先前的研究中已被报道能够响应极端环境引发的选择性压力，例如： $ERD3$ (*evm8.model.Scaffold256.52*) [87]、 $HSP83$ (*evm8.model.Scaffold724.36*) [88]、 $KAS2$ (*evm8.model.Scaffold92.49*) [89] 和 $ROF1$

(*evm8.model.Scaffold1105.1*)^[90], 它们参与应对低温或高温的胁迫; *ATDTX28* (*evm8.model.Scaffold224.8*)^[91]和 *RD2* (*evm8.model.scaffold308.48*)^[92]被证明能够参与水分调节机制 (表 3.1)。此外, *SUD7* (*evm8.model.Scaffold237.21*)^[93]、*DWF3* (*evm8.model.Scaffold175.11*)^[94]、*SSN1* (*evm8.model.Scaffold32.115*)^[95]、*REF3* (*evm8.model.Scaffold350.22*)^[96]和 *CYP81D8* (*evm8.model.Scaffold961.11*)^[97]参与了应对强光辐射胁迫和盐分胁迫; *MYB4* (*evm8.model.Scaffold1.69*) 转录因子表明转录调控可能参与了对非生物胁迫的响应 (表 3.1)。

表 3.1 全基因组渐渗区域内响应非生物胁迫的基因的功能注释
 Table 3.1 Functional annotation of genes responsive to abiotic stress within the introgression region across whole genome

基因 (Gene)	同源基因 (Homologs)	<i>p</i> 值 (<i>p</i> -value)	描述 (Description)
<i>evm8.model.Scaffold224.8</i>	<i>ATDTX28</i>	0	Response to water deprivation
<i>evm8.model.Scaffold256.52</i>	<i>ERD3</i>	2.00E-143	Response to temperature stimulus
<i>evm8.model.Scaffold237.21</i>	<i>SUD7</i>	3.00E-81	cellular response to light intensity
<i>evm8.model.Scaffold92.49</i>	<i>KAS2</i>	0	Response to cold
<i>evm8.model.Scaffold175.11</i>	<i>DWF3</i>	2.00E-27	response to light stimulus, response to water deprivation
<i>evm8.model.Scaffold724.36</i>	<i>HSP83</i>	4.00E-143	Response to heat
<i>evm8.model.Scaffold1105.1</i>	<i>ROF1</i>	1.00E-27	Response to heat

研究结果

<i>evm8.model.Scaffold1.69</i>	<i>MYB4</i>	9.00E-72	negative regulation of DNA-templated transcription
<i>evm8.model.Scaffold350.22</i>	<i>REF3</i>	6.00E-79	Response to light stimulus
<i>evm8.model.Scaffold32.115</i>	<i>SSN1</i>	3.00E-141	DNA repair
<i>evm8.model.Scaffold308.48</i>	<i>RD2</i>	4.00E-10	Response to desiccation
<i>evm8.model.Scaffold961.11</i>	<i>CYP81D8</i>	2.00E-147	Response to oxidative stress, Response to salt stress, Response to water deprivation

3.4 适应性遗传变异位点检测

本研究使用的两种 GEAs 方法检测青藏高原种群和非青藏高原种群中由于环境因素而驱动适应性遗传变异位点。LFMM 分析结果显示分别在青藏高原种群和非青藏高原种群中检测出了 16 552 和 9 659 个适应性 SNPs, RDA 方法在青藏高原种群和非青藏高原种群中 23 433 和 9 565 个适应性 SNPs (图 3.14)。两种 GEAs 方法在青藏高原种群检测出共同的 6 600 个适应性 SNPs, 在非青藏高原种群检测出共同的 662 个适应性 SNPs (图 3.14)。Snpeff 注释结果发现青藏高原种群和非青藏高原种群的适应性 SNPs 分别涉及 2 159 和 309 个功能基因。GO 富集分析结果显示出了青藏高原种群中存在一些促进局地适应 (Local adaptation) 的结果, 例如: response to external stimulus (GO:0009605)、defense response (GO:0006952) 和 cellular response to stress (GO:0033554), 然而在非青藏高原种群中并没有富集到与非生物胁迫相关的 GO 富集结果。

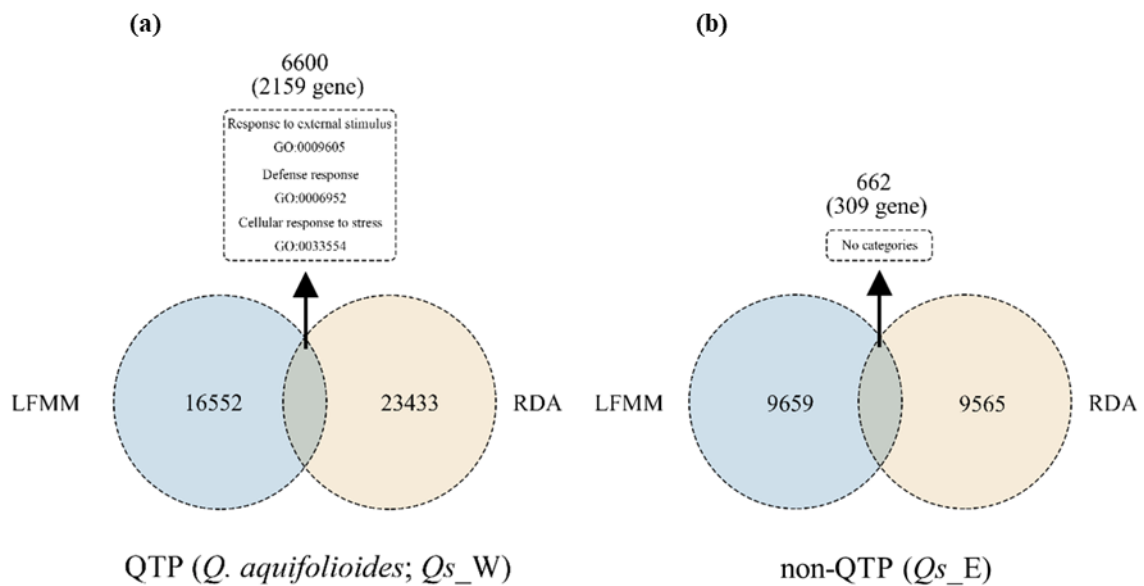


图 3.14 两种 GEAs 方法检测适应性遗传变异位点的结果。(a) 青藏高原种群和 (b) 非青藏高原种群中检测到的共同的适应性 SNPs。虚线框显示了含有这些异常 SNPs 的基因的 GO 结果, 仅显示了 $p < 0.05$ 的与非生物胁迫相关的 GO 结果

Figure 3.14 The result of adaptive SNPs were detected using two GEAs methods. (a) Overlapping adaptive SNPs were detected in (a) QTP populations and (b) non-QTP populations. The dashed boxes showed the results of GO enrichment for genes containing these outlier SNPs, only GO categories associated with abiotic stress ($p < 0.05$) are shown

青藏高原种群的适应性 SNPs 在基因组中的位置和密度分布结果表明适应性 SNPs 与上述川滇高山栎和 *Qs_W* 的基因组渐渗区域的 SNPs 基因组特征相似, 主要集中在基因的调控区域, 并且没有在任何染色体上显著富集 (图 3.15, 图 3.16)。

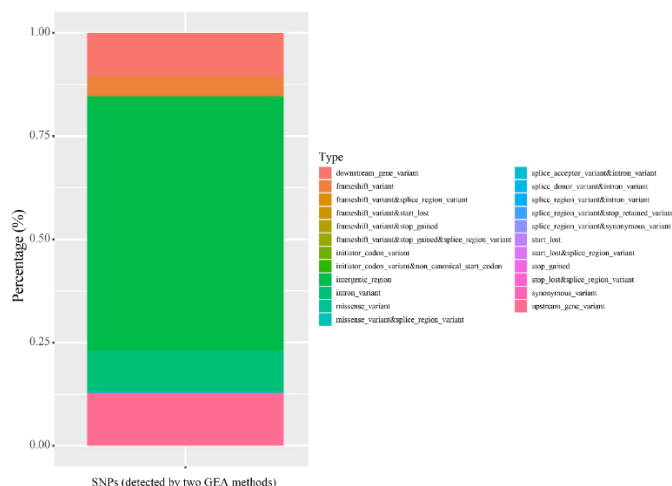


图 3.15 两种 GEA 方法检测到的共同 SNPs 的位置分布统计
 Figure 3.15 The positional distribution of SNPs were detected simultaneously by both GEA methods

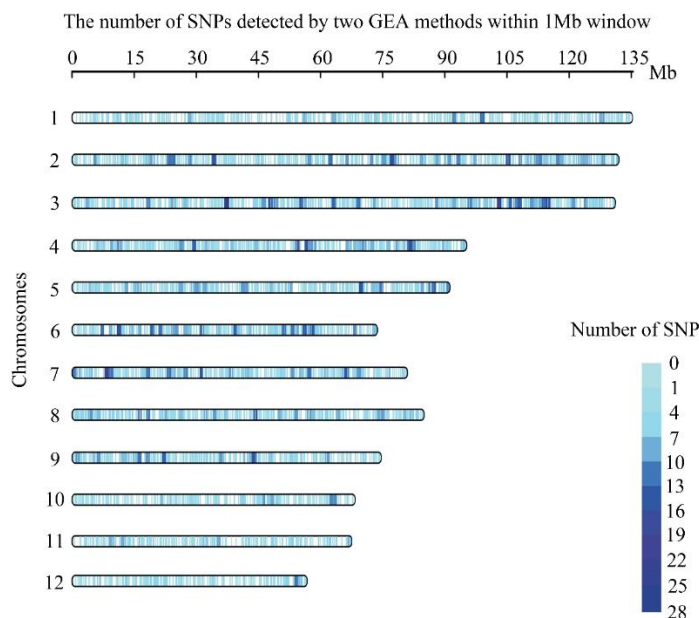


图 3.16 两种 GEA 方法共同检测到的 SNPs 在全基因组上 1 Mb 窗口内的数量
 Figure 3.16 Number of SNPs detected simultaneously by both GEA methods within 1 Mb window across whole genome.

3.5 遗传变异与地理、环境关系

Mantel 检验结果显示在青藏高原种群和非青藏高原种群中没有显著的 IBD (QTP: Mantel's $r = -0.02$, $P = 0.61$; non-QTP: Mantel's $r = 0.31$, $P = 0.14$) 和 IBE (QTP: Mantel's $r = 0.13$, $P = 0.38$; non-QTP: Mantel's $r = 0.13$, $P = 0.38$) (表 3.2)。然而, 通过控制地理因子和控制环境因子的偏 Mantel 检验结果显示在青藏高原种群 (Mantel's $r = -0.14$, $P = 0.82$) 和非青藏高原种群 (Mantel's $r = -0.23$, $P = 0.87$) 中也未检测出显著的 IBE, 但是在非青藏高原种群中具有显著的 IBD (Mantel's $r = 0.36$, $P = 0.04$), 在青藏高原种群中没有显著的 IBD (Mantel's $r = -0.03$, $P = 0.65$) (表 3.2)。

表 3.2 青藏高原和非青藏高原种群 Mantel 和偏 Mantel 检验结果
Table 3.2 Results of Mantel test and partial Mantel test in QTP and non-QTP populations

	QTP (<i>Q. aquifolioides</i> ; <i>Qs_W</i>)		non-QTP (<i>Qs_E</i>)	
	Mantel's <i>r</i>	<i>P</i>	Mantel's <i>r</i>	<i>P</i>
Mantel test				
地理隔离 (IBD)	-0.01853	0.6129	0.309371	0.139
环境隔离 (IBE)	0.131436	0.3773	0.131436	0.3773
partial Mantel test				
地理隔离 (IBD)	-0.03326	0.654	0.35975	0.0441
环境隔离 (IBE)	-0.14013	0.8232	-0.23218	0.87

此外，比较青藏高原种群和非青藏高原之间遗传距离、地理距离和环境距离矩阵结果显示三者之间存在着线性关系，随着青藏高原种群和非青藏高原之间遗传距离的增加，它们的地理距离和环境距离也表现出增加的趋势（图 3.17）。

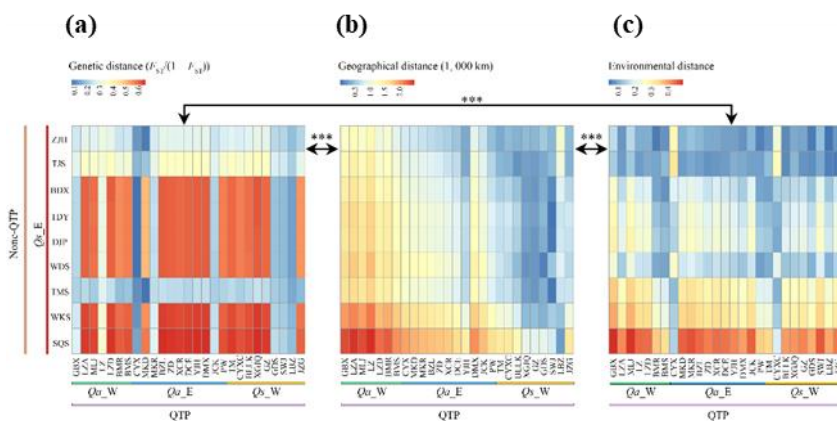


图 3.17 青藏高原和非青藏高原种群的遗传、地理和环境数据计算的欧几里得距离矩阵。(a) 遗传距离矩阵，(b) 地理距离矩阵和 (c) 环境距离矩阵。红色表示欧几里得距离远，蓝色表示欧几里得距离近，双向箭头上方的“***”表示 $p \leq 0.05$

Figure 3.17 Euclidean distance matrices based on genetic, geographic and environmental data for QTP populations and non-QTP populations. (a) Genetic distance matrix, (b) geographic distance matrix and (c) environmental distance matrix. Red color indicates a long Euclidean distance and blue color indicates a short Euclidean distance, the "***" above the double-edged arrows indicates $p \leq 0.05$

GDM 分析结果显示平均气温日较差 (Bio02) 是青藏高原种群适应性遗传变异最重要的影响因子 (Importance = 0.138)，其次是地理因素 (Importance = 0.114) (图 3.18)。整体来说，与温度相关的生物气候变量 (Bio02, Bio05, Bio03, Bio04) 是青藏高原种群局地适应机制中最重要的适应性遗传变异驱动力，其次是与地理因素和与降水相关的生物气候变量 (Bio18, Bio19, Bio15)。在非青藏高原种群中，最湿季度降水量 (Bio17) 是适应性变异的主要驱动力 (Importance = 0.561)，其它的与温度相关的生物气候变量 (Bio10, Bio03, Bio07, Bio04, Bio01) 贡献度相对较低，地理因素也没有展示出显著的重要性 (图 3.18)。

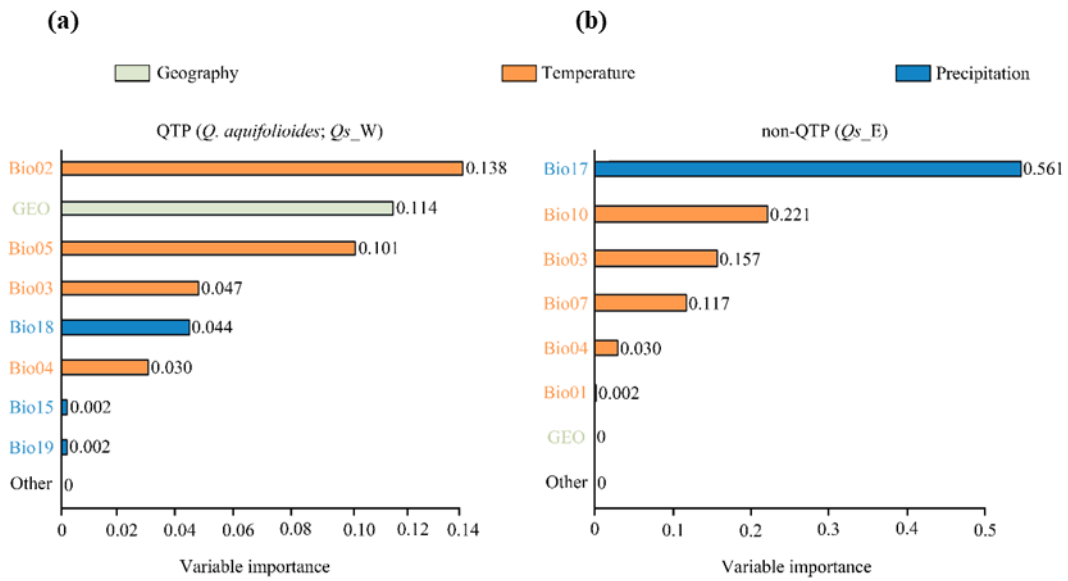


图 3.18 广义相异模型 (GDM) 分析的 (a) 青藏高原种群和 (b) 非青藏高原种群环境变量重要性排序。红色表示与温度相关的生物气候变量，蓝色表示与降水相关的生物气候变量，绿色表示地理变量

Figure 3.18 Importance of environmental variables based on GDM for (a) QTP and (b) non-QTP populations. Orange color indicates bioclimatic associated with variable temperature, blue color indicates bioclimatic variable associated with precipitation, and green color indicates geographic variables

3.6 适应性渐渗基因组特征

结合 f_{aM} 值和 GEAs 方法 (LFMM, RDA) 检测参与川滇高山栎和刺叶高山栎之间适应性渐渗的功能基因结果显示在川滇高山栎和 Qs_W 之间检测到了 435 个参与适应性渐渗的功能基因，这些功能基因在 12 条染色体上均匀分布 (图 3.19)。

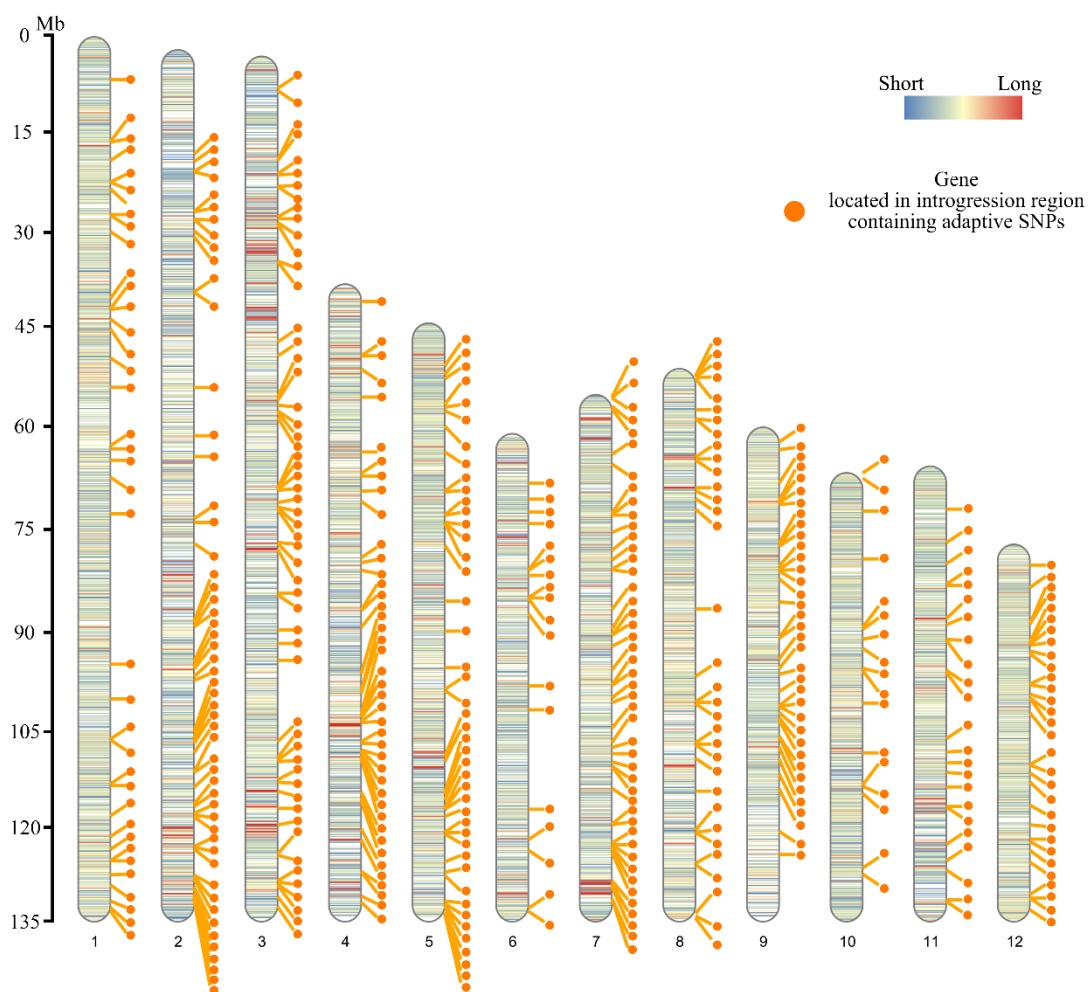


图 3.19 12 条染色体上每个基因的长度可视化结果。橙色圆表示参与适应性渐渗的基因，红色表示基因的序列更长，蓝色表示基因的序列更短

Figure 3.19 The length of each gene across 12 chromosomes. Orange circles indicate genes involved in adaptive introgression, red color indicates gene with longer sequences and blue indicates gene with shorter sequences

比较参与适应性渐渗的基因的自身长度与其他基因 (Background, Gene containing adaptive SNPs, Gene located in introgression region) 的长度结果发现发生适应性渐渗的基因表现出最长的碱基序列信息, 其次是参与局地适应的基因和位于基因组渐渗区域的基因 (图 3.20)。

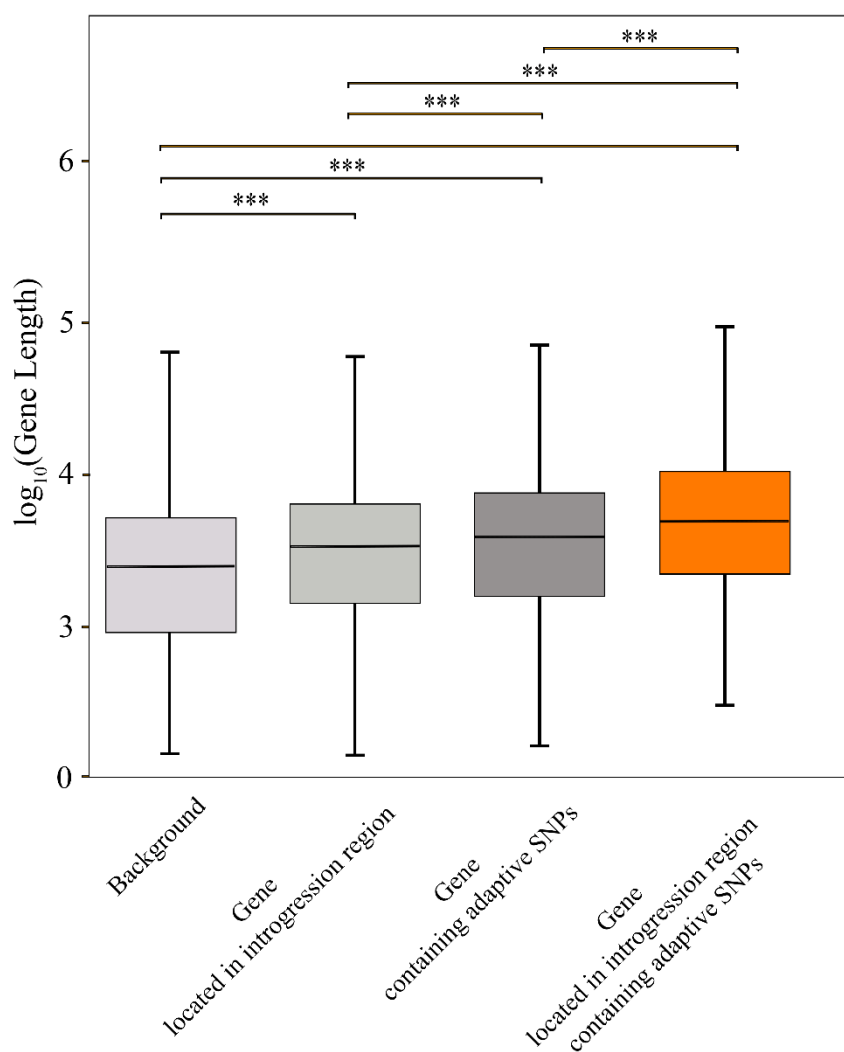


图 3.20 发生适应性渐渗的基因的长度与基因组背景，发生渐渗的基因和存在适应性变异的基因的长度比较

Figure 3.20 Comparison of the length of the genomic background, genes that occurred introgression, genes containing adaptive SNPs, and genes that occurred adaptive introgression

参与适应性渐渗的基因的自身长度 (Gene length)、内含子长度总和 (Intron length) 和外显子长度总和 (CDS length) 与基因自身含有的适应性 SNPs 的数量之间的线性分析结果显示基因自身长度 ($r=0.22, p<0.01$) 或基因自身内含子长度总和 ($r=0.15, p<0.01$) 与适应性 SNPs 数量存在显著正相关，但是并没有显示基因自身外显子长度总和与适应性 SNPs 数量 ($r=-0.07, p=0.14$) 存在显著的线性关系 (图 3.21)。

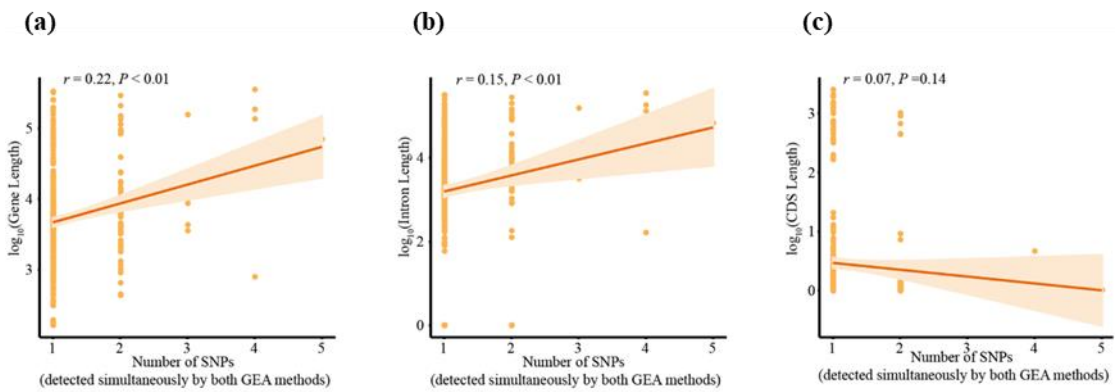


图 3.21 发生适应性渐渗的基因的适应性 SNP 数量与 (a) 基因长度, (b) 内含子长度和 (c) 外显子长度的线性分析

Figure 3.21 Linear analysis of the number of adaptive SNPs and (a) gene length, (b) intron length and (c) exon length for genes that involved in adaptive introgression

使用参与适应性渐渗的功能基因中位于上游和下游的适应性 SNPs 进行 DNA 顺式调控元件的预测结果显示在所有的 247 个适应性 SNPs 中, ~60% (148) 的 SNPs 被预测位于已知的植物 DNA 顺式调控元件中, 例如: SEF3MOTIFGM、CAATBOX1、ARR1AT、RAV1AAT、CACTFTPPCA1 等 (图 3.22)。此外, ~17% (42) 的 SNPs 被预测位于与环境因素 (光照: Light, 温度: Temperature, 降水: Precipitation, 氧气: Oxygen, 盐分: Salt) 相关的 DNA 顺式调控元件中, 例如: GT1CONSENSUS、ANAERO4CONSENSUS、GATABOX、GT1CONSENSUS、TBOXATGAPB 等 (图 3.22)。

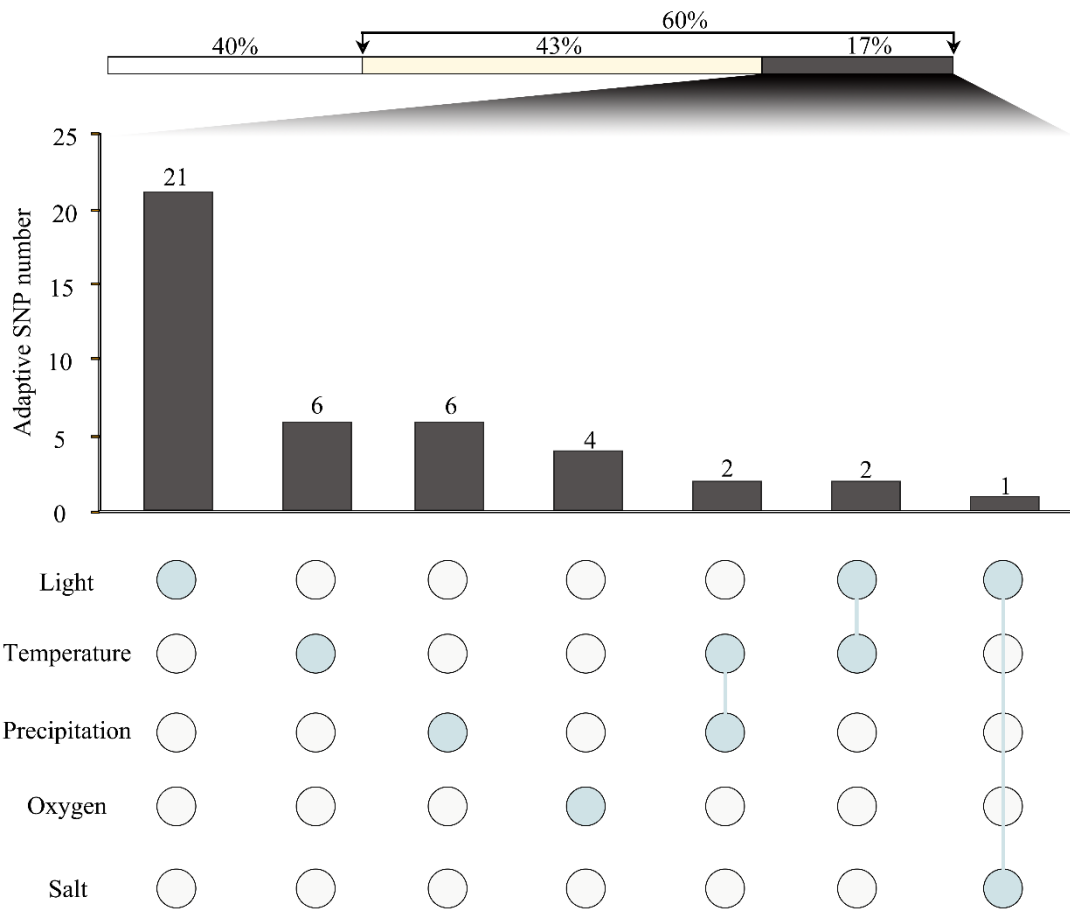


图 3.22 使用 Plant Cis-Regulatory DNA Elements Database 在线植物顺式调控 DNA 元件数据库对适应性渐渗基因上游和下游适应性 SNPs 进行调控元件预测的结果。在所有的 247 个 SNPs 中，~60% (148) 的 SNPs 被预测位于已知的植物 DNA 顺式调控元件中，~17% (42) 的 SNPs 被预测位于与环境相关的 DNA 顺式调控元件中

Figure 3.22 Results of regulatory element prediction using Plant Cis-Regulatory DNA Elements Database for adaptive introgression genes upstream and downstream of adaptive SNPs. Of all 247 SNPs, ~60% (148) were predicted to be located in known cis-regulatory elements of plant DNA, and ~17% (42) were predicted to be located in cis-regulatory elements in response to environmental factors

3.7 *KAS2* 基因内适应性变异与生态适应

参与物种间适应性渐渗的功能基因在物种应对快速气候变化而带来的选择性压力过程具有增强物种局地适应的能力，例如 *KAS2* 基因在植物中已被证明参与应对来自低温的胁迫^[90,98]。*KAS2* 基因结构上由 7 个外显子和 6 个内含子组成。在本研究中，*KAS2* 基因中一个适应性 SNP (3_82758545) 位点在 LFMM 和 RDA 分析中被共同检测出来，LD 分析显示该适应性 SNP 与周围区域的 SNPs 连锁不平衡强度相对较低 (图 3.23)。在 *KAS2* 基因中，川滇高山栎和 *Qs_W* 之间的遗传分化 (F_{ST} , D_{XY}) 相对于青藏高原种群和非青藏高原种群之间 (QTP vs. non-QTP) 更低，非青藏高原种群内 π 值相对于川滇高山栎、*Qs_W* 和青藏高原种群更高 (图 3.24)。

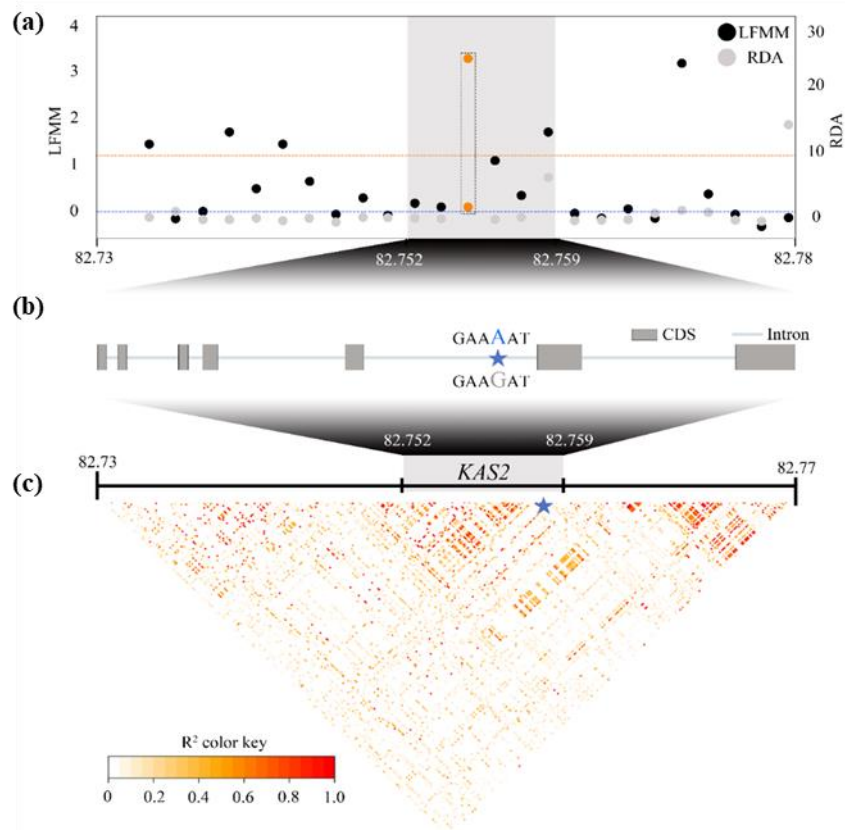


图 3.23 *KAS2* 基因特征。(a) *KAS2* 基因及其周围区域的 SNPs 的 LFMM 和 RDA 结果；(b) *KAS2* 基因结构和候选位点的两种等位基因；(c) *KAS2* 基因及其周围区域的连锁不平衡。横轴上相同位置的两个不同颜色的点表示同一个 SNP 的 LFMM 和 RDA 结果，橙色点和蓝色五角星表示候选位点

Figure 3.23 Characterization of the *KAS2* gene. (a) LFMM and RDA results of SNPs in the *KAS2* gene and its surrounding region; (b) *KAS2* gene structure and two allele genotypes of candidate loci; (c) Linkage disequilibrium in the *KAS2* gene and its surrounding region. Two different color dots at the same position on the horizontal axis indicate the LFMM and RDA results for the same SNP, orange dots and blue pentagrams indicate candidate loci

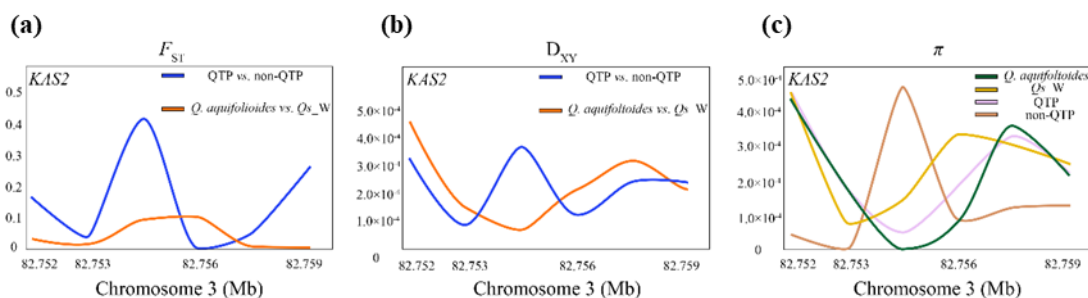


图 3.24 *KAS2* 基因在物种间 (a) F_{ST} , (b) D_{XY} , 以及物种内 (c) π 的结果。横轴表示 *KAS2* 基因在三号染色体上的位置，纵轴表示遗传多样性指数大小
Figure 3.24 The result of inter-species (a) F_{ST} , (b) D_{XY} , and intra-species (c) π values within the *KAS2* gene. The x axis indicates the position of the *KAS2* gene on three chromosome and the y axis indicates the value of the genetic diversity index

本研究通过计算候选位点 (3_82758545) 两种基因型在每个种群中的等位基因频

率发现-A 等位基因 (A allele) 主要集中在青藏高原种群中, -G 等位基因 (G allele) 主要存在于非青藏高原种群中 (图 3.25)。此外, 我们还发现-A 等位基因频率从高海拔青藏高原地区向低海拔的中国东部地区, 等位基因频率呈现下降的趋势, 但是-G 等位基因频率呈现上升的趋势。从温度变化的角度分析, 从寒冷的青藏高原地区到温暖的中国中、东部地区, -A 等位基因频率逐渐降低, -G 型等位基因频率逐渐上升。

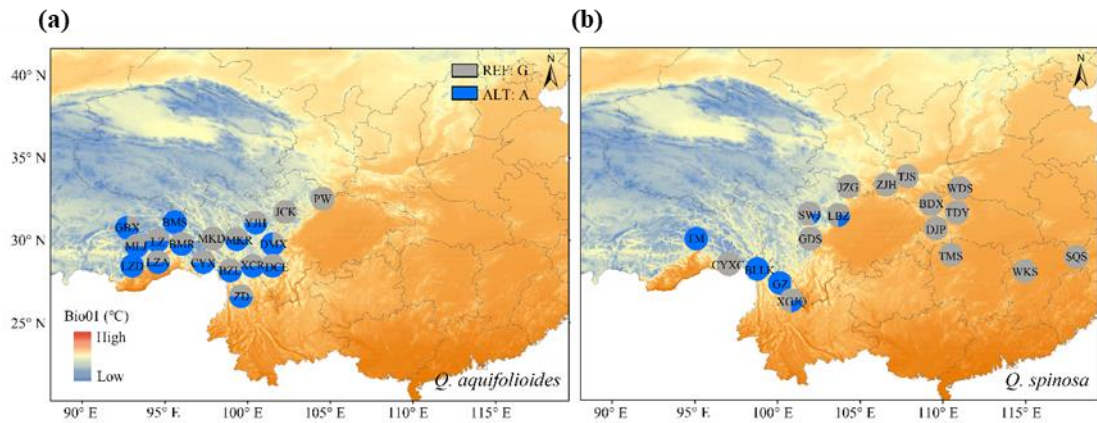


图 3.25 *KAS2* 基因内含子中一个适应性 SNP 在 (a) 川滇高山栎和 (b) 刺叶高山栎每个种群中的等位基因频率变化结果。蓝色表示-A 等位基因, 灰色表示-G 等位基因, 地图背景为年平均温度 (Bio01)

Figure 3.25 Allele frequency of an adaptive SNP located in the intron of *KAS2* gene in each population of (a) *Q. aquifolioides* and (b) *Q. spinosa*. The blue color indicates A allele, grey color indicates G allele, and the map background is mean annual temperature (Bio01)

将分别含有(GAAAAT)₅ – 35S mini :: *INFI* 和(GAAGAT)₅ – 35S mini :: *INFI* 质粒的农杆菌菌液注射到新鲜的烟草叶片左、右两侧结果表明-A 等位基因表现出明显的与光照时间、低温和紫外线辐射的关系 (附录 E) (图 3.26)。当将注射完毕的烟草安放在一直黑暗的条件下培养 (白天温度 25°C / 夜晚温度 20°C) 30 h 后, 注射含有 (GAAAAT)₅ – 35S mini :: *INFI* 质粒的农杆菌菌液的区域出现明显的细胞坏死现象。此外, 将注射完毕的烟草先培养在白天 (温度: 25 °C, 光照时间: 16 h) / 夜晚 (温度: 20 °C, 光照时间: 0 h) 环境下 30 h 后转入黑暗条件下 (白天温度 25°C / 夜晚温度 20°C) 后继续培养, 大约 30 h 后注射含有(GAAAAT)₅ – 35S mini :: *INFI* 质粒的农杆菌菌液的区域依然会出现明显的细胞坏死现象。反之, 当开始是在黑暗条件下培养 30 h, 注射区域出现细胞坏死现象, 之后转入白天 (温度: 25 °C, 光照时间: 16 h) / 夜晚 (温度: 20 °C, 光照时间: 0 h) 环境下培养, 原先的细胞坏死区域范围出现继续扩张的趋势。当将实验样品培养在低温 (温度: 25 °C, 光照时间: 16 h) / 夜晚 (温度: 4 °C, 光照时间: 0 h) 环境条件下或者持续性的紫外线照射条件 (白天: 25 °C / 夜晚: 20 °C) 下, 注射含有(GAAAAT)₅ – 35S mini :: *INFI* 质粒的农杆菌菌液的区域都会出现明显的细胞坏死现象。但是, 如果一直在正常 (温度: 25 °C, 光照时间: 16 h) / 夜晚 (温度: 20 °C, 光照时间: 0 h) 环境条件下培养, 注射含有(GAAAAT)₅

-35S mini :: *INFI* 质粒的农杆菌菌液的区域始终不会出现细胞坏死现象。对于注射含有(GAAGAT)₅ - 35S mini :: *INFI* 质粒的农杆菌菌液的区域，无论在上述哪种环境条件下都可以观察到明显的细胞坏死现象，并且细胞坏死的范围和速度并没有在任一种特定环境下增加或者抑制的结果。

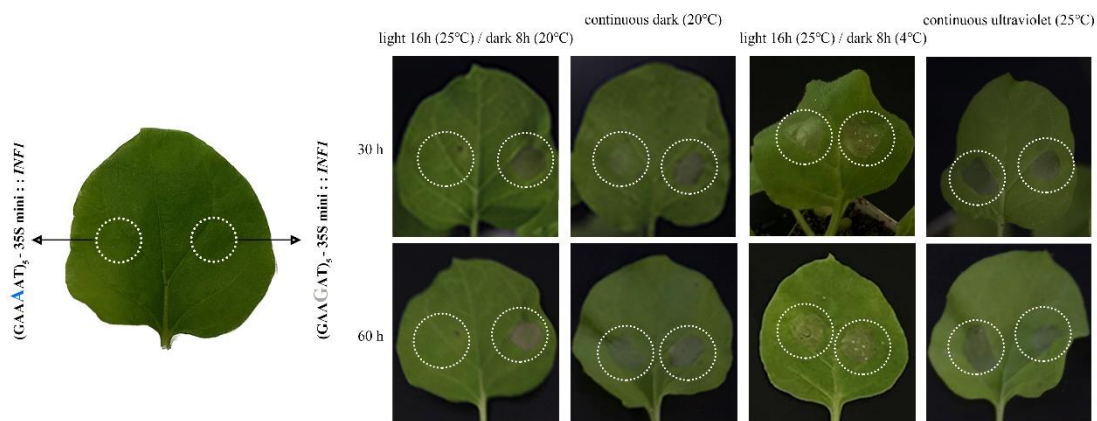


图 3.26 四种环境条件下两种基因型农杆菌侵染烟草叶片的反应结果。在每片叶子上，左边的虚线圈为注射((GAAAAT)₅ - 35S mini:: *INFI*)的区域，右边的虚线圈为注射((GAAAAT)₅ - 35S mini:: *INFI*)的区域

Figure 3.26 Reaction of *Nicotiana benthamiana* leaves was injected by agrobacteria containing two genotypes under four environmental conditions. The dotted circles on the left side of the leaves were the region of injection ((GAAAAT)₅ - 35S mini:: *INFI*) and the right side of the leaves were the region of injection ((GAAAAT)₅ - 35S mini:: *INFI*)

注：图片从左到右四种环境条件分别为：light 16 h (25°C) / dark 8 h (20°C)：16 h 光照 (25°C) / 8 h 避光 (20°C) (对照组)；continuous dark：24 h 避光 (20°C) (避光)；light 16 h (25°C) / dark 8 h (4°C)：16 h 光照 (25°C) / 8 h 避光 (4°C) (低温)；continuous ultraviolet：24 h 紫外线辐射 (25°C) (强紫外线)

3.8 梯度森林预测

在 2050 年和 2070 年的低碳排放 (RCP 2.6) 和高碳排放 (RCP8.5) 情景下，梯度森林预测使用 GDM 分析的结果中最重要的生物气候变量作为环境变量预测青藏高原种群和非青藏高原种群遗传偏移结果表明在 Bio02 作为气候环境变量时，川滇高山栎在青藏高原地区的遗传偏移 (Genetic offset) 最低，其次是 *Qs_W*，最后是 *Qs_E*；然而，Bio17 作为气候环境变量时，川滇高山栎还是在青藏高原地区呈现遗传偏移低，*Qs_W* 在华中和华东地区的遗传偏移较高 (图 3.27，附录 B-D)。而对于 *Qs_E* 谱系，华东沿海地区表现出最高的遗传偏移。最后，当不考虑川滇高山栎和 *Qs_W* 之间适应性渐渗驱动的适应性遗传变异时，通过上述 GF 预测证实了适应性渐渗增强 *Qs_W* 谱系对于横断山脉及其周围区域内环境的适应能力。上述的 GF 预测在 2050 和 2070 年 RCP 2.6 和 RCP 8.5 碳排放情景下都得到了同样的结果。

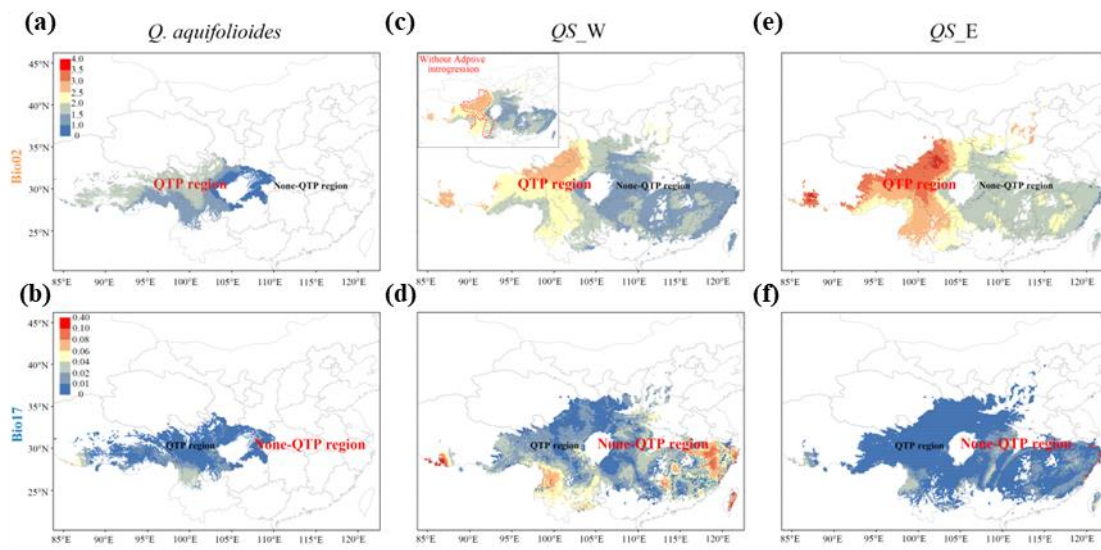


图 3.27 在 2050 年 RCP2.6 碳排放量情景下，两种栎属物种在平均气温日较差（Bio02）和最湿季度降水量（Bio17）气候条件下的遗传偏移。（a, b）川滇高山栎，（c, d）刺叶高山栎西谱系和（e, f）刺叶高山栎东谱系

Figure 3.27 The results of genetic offset of two *Quercus* species using Bio02 and Bio17 climatic variables in 2050 under RCP 2.6. (a, b) *Q. aquifolioides*, (c, d) *Qs_W* and (e, f) *Qs_E*

4讨论

尽管青藏高原区域内一直存在着明显的地理限制,但在该区域内依然存在普遍的适应性渐渗^[20,36,99]。一方面,作为一种快速获得适应性遗传变异的重要方式,适应性渐渗可以增强物种应对生物或非生物胁迫的能力,甚至可以驱使物种向新的地区迁移,进而扩展物种原有的生态位。另一方面,一些研究认为物种之间的渐渗是生殖隔离的“破坏者”,它可以增加全基因组水平的物种间共享多态性,进而削弱全基因组水平的物种间遗传分化程度^[10,15,16]。本研究结合种群遗传学和景观基因组学研究方法系统的探究了物种间适应性渐渗的遗传机制发现适应性渐渗加速了物种间和谱系间适应性遗传分化,以及提升了物种局地适应的能力。本研究结果为研究森林树种之间的适应性渐渗机制奠定了基础,为快速气候变化背景下森林树种的管理和保护措施的制定提供了新的见解。

4.1 川滇高山栎和刺叶高山栎种群历史动态和遗传分化模式

青藏高原地区高度的异质性景观造成了川滇高山栎和刺叶高山栎都出现了物种内部东-西谱系分化^[23,42,100]。同时,这种物种内部谱系分化现象在青藏高原内的其他物种中同样存在,例如:高山松(*Pinus densata*)^[101]、岷江柏木(*Cupressus chengiana*)^[38]和互叶醉鱼草(*Buddleja alternifolia*)^[102]。早中新世(Miocene)以来,青藏高原地区发生了持续性的地质隆起,并伴随着气候振荡^[20,103],这给当地物种创造了多样化的生态环境和栖息地,同时也给川滇高山栎和刺叶高山栎的“Out-of-QTP”提供了契机^[22,27]。最近的研究发现这两个物种内部和物种之间非对称的历史基因流方向是从川滇高山栎到刺叶高山栎,从西谱系到东谱系。并以此表明了这两个栎属物种从青藏高原地区迁徙到当今的中国中、东部的“Out-of-QTP”生物地理学模式^[23,42]。但在物种水平上,这两个栎属物种走出青藏高原的生物地理学模式尚不明确^[26,42,100]。基于系统发育分析结果发现物种间和谱系间的基因流是从西向东的。同时,种群历史动态分析结果发现在大约 14.88 Mya,川滇高山栎的一部分祖先种群从西藏地区分离出来,迁徙并定殖到当今的横断山脉地区,形成了刺叶高山栎。持续性的地质活动和伴随的气候振荡导致了刺叶高山栎和川滇高山栎分别在 8.98 和 3.04 Mya 发生了谱系分化。横断山脉地区的刺叶高山栎的部分种群迁徙到当今的秦岭和中国东部地区,并定殖在该地区形成了 *Qs_E* 谱系,在西藏地区的川滇高山栎的部分种群迁移至横断山脉和川西高原,形成了 *Qa_E* 谱系。类似的生物地理学过程在青藏高原地区内的兔耳草属(*Lagotis* spp.)^[104],香青属(*Anaphalis* spp.)^[105]和火绒草属(*Leontopodium* spp.)^[106]中同样存在,甚至可以迁移到更远的中亚、欧洲和北美等地区。

比较川滇高山栎和刺叶高山栎物种间和谱系间的 SNPs 等位基因频率结果发现刺叶高山栎谱系间,以及青藏高原种群和非青藏高原之间表现出比物种间更高的遗传分

化,这可能是青藏高原与外界之间强大的地理屏障造成的。位于青藏高原东南缘的横断山脉由多个南北走向的山脊组成,其南部的早期地质隆升可以追溯到古新世(Paleocene)晚期甚至始新世(Eocene)时期^[20,107]。高大的山体限制了青藏高原内、外的种群的基因交流。青藏高原与非青藏高原种群之间遗传距离、地理距离和环境距离矩阵的线性关系也显示 IBD 对青藏高原种群和非青藏高原种群之间基因流的存在限制。在青藏高原地区,尽管存在明显的“天空岛(Sky islands)”,但很多种群基因组学研究结果显示青藏高原地区的物种之间会频繁发生基因流^[36,37,65,99]。Mantel 和偏 Mantel 检验的结果也解释了青藏高原种群内没有存在显著的 IBD 和 IBE,同时系统发育分析估算的基因流主要是从 Qa_E 谱系向 Qs_W 谱系。基因流在调节生物物种形成过程中起着关键性的作用,它通过阻碍遗传分化和使正向突变传播而将种群聚集在一起^[108]。Ma 等^[57]使用全基因组 SNPs 数据证实了川滇高山栎和刺叶高山栎之间存在杂交渐渗。本研究中的显著 D 和 f_{AM} 统计量不仅确定了两个物种之间的渐渗事件,同时揭示了川滇高山栎和 Qs_W 之间存在 1461 个潜在的渐渗区域。物种间的持续性的渐渗增加了种间共享多态性^[17],降低了物种间的遗传分化,并可能进一步启动新的杂交物种形成过程^[20,109-112]。

4.2 适应性渐渗选择信号

GDM 结果表明温度是驱动青藏高原种群适应性遗传变异最重要环境因素,而降水是非青藏高原种群适应性遗传变异的最主要驱动力。此外,与非青藏高原种群相比,青藏高原种群中富集到了更多的适应性 SNPs 和参与非生物胁迫的基因,这表明青藏高原种群可能比非青藏高原种群包含更多潜在的适应性遗传变异,可能是川滇高山栎和 Qs_W 之间的适应性渐渗导致的^[3,111,112]。渐渗造成的影响往往是中性或者负面的,也存在正向的渐渗促进物种的局地适应的能力^[3,7]。了解物种之间的适应性渐渗的遗传机制,对于解决物种适应性遗传分化和局部适应机制的至关重要。Liu 等^[8]研究多个杨属物种之间的杂交渐渗发现基因组中的潜在渐渗区域往往表现出更低的物种间 F_{ST} 和 D_{XY} , 和更高的物种内 π 值。然而,适应性渐渗的基因组特征可能比中性或有害的等位基因渐渗更为复杂。本研究结合 f_{AM} 和 GEAs 方法发现了 435 个参与川滇高山栎和 Qs_W 适应性渐渗的功能基因。与基因组中的其他基因对比,这些参与适应性渐渗的功能基因表现出更长的碱基序列。最近的研究也证实了基因长度与基因的表达水平之间存在一定关系^[113]。序列更长的基因往往会提供更多的突变位点,这为物种内部的适应性遗传变异提供了更多的机会。适应性遗传变异的存在削弱了环境变化带来的选择性压力,从而有助于物种在特定环境中的生存和繁殖^[8,23,114-116]。一些碱基序列长度大的基因被解释可能与基因自身内含子存在关系^[113]。比较这些参与适应性渐渗的适应性 SNPs 的数量于基因自身碱基序列长度或基因内所有内含子长度总和,发现了它们之间存在着显著的线性关系。

在检测到的 435 个参与适应性渐渗的功能基因中, 其中部分基因的功能与植物开花时间调控有关。物种之间开花时间的不一致是导致的合子前生殖隔离的重要原因, 从而影响物种之间遗传物质交换^[117]。此外, 其中还发现了一些基因已报道在植物物种中能够参与应对生物或非生物胁迫。这些基因都直接或间接的增强了对快速气候变化的适应能力。类似的研究比如: 北美地区的向日葵 (*Helianthus annuus*) 和小向日葵 (*H. debilis*) 物种之间的适应性渐渗影响了小向日葵多个与生物和非生物环境相关的表型性状, 并可能有助于将小向日葵的分布范围扩展到美国的得克萨斯州中部^[118]; 巨柏 (*Cupressus gigantea*) 和干香柏 (*C. duclouxiana*) 之间存在广泛的基因流, 并且检测到了 16 个参与适应性渐渗功能基因, 这些功能基因可能在北部干香柏适应高海拔地区的多种环境压力时发挥了重要作用; 换句话说, 分布于低海拔的干香柏通过向分布于高海拔的巨柏“取经”, 征服了更寒冷更干燥的横断山脉的南部地区^[119]。广泛分布的卡西亚松 (*Pinus kesiya*) 以及与其邻域分布的云南松 (*P. yunnanensis*) 研究发现物种间的渐渗对种子的萌发有显著影响, 可能增强了它们对不同环境下的生态适应能力^[120]。

本研究中发现的响应低温环境的 *KAS2* 基因^[90,98]参与了川滇高山栎和 *Qs_W* 之间的适应性渐渗。*KAS2* 基因在 *Q. aquifolioides* vs. *Qs_W* 与 QTP vs. non-QTP 比较中表现出更低的 F_{ST} 和 D_{XY} , *KAS2* 基因在非青藏高原种群与川滇高山栎、*Qs_W* 和青藏高原种群比较中表现出更高的 π 。此外, 由于 *KAS2* 基因及其周围区域内两两 SNPs 位点之间的低水平的 LD, 可以推测导致位于 *KAS2* 基因内含子中适应性变异 (3_82758545) 的真正原因^[121]。用携带质粒的农杆菌侵染新鲜的烟草叶片, 发现 -A 等位基因在黑暗条件、低温环境下或者持续性紫外线照射条件下, 会导致基因表达量的上升。然而在室温 (25°C/20°C) 或正常光照条件下, -A 等位基因并未表现出对于基因表达量的调控作用, 这可能是由于青藏高原地区特定的环境条件所导致。川滇高山栎和 *Qs_W* 之间的适应性渐渗导致了 *KAS2* 基因内含子区域内的适应性变异, 从而增强或者抑制的 *KAS2* 基因在不同种群或谱系中表达水平的差异。通过调节基因表达水平, 而非直接改变氨基酸的种类或基因的功能, 来提高物种对局部适应的能力的方式已经在其他栎属物种中被发现^[7]。本研究发现渐渗区域内的 SNPs 和 GEAs 方法检测到的 SNPs 都主要集中在调控区, 如基因间区和内含子区域。内含子中的 SNPs 可能产生可变剪接的 mRNA, 而 5'- 或 3'-非翻译 mRNA 区域的遗传变异可能会影响 RNA 的稳定性或加工。此外, 影响调控因子结合的基因非蛋白质编码区的 SNPs 可能导致 SNPs 等位基因表达失衡^[122,123]。这样的结论同样在其它森林树种局地适应机制的研究中得到了证实^[7,121]。

4.3 应对气候变化与保护的启示

在全球气候快速变化的背景下, 森林树种更容易受到快速气候变化的影响^[121,124]。

近年来,景观基因组学的最新进展可以将适应性遗传变异和景观特征联系起来用以检测适应性选择信号,因此为森林树种的局部适应机制和保护策略提供了新的见解^[23,124-126]。测序技术和新分析方法(如:GF和GDM)的出现可以预测物种在未来环境条件的脆弱性区域。本研究基于适应性SNPs使用GF预测发现川滇高山栎的脆弱性区域主要在青藏高原的北部和西部,刺叶高山栎的脆弱性区域主要在横断山脉地区,这些脆弱性区域的种群在未来可能受到的威胁较其他种群更大。同样的方法在其他森林树种中已被证明是有效的,例如:欧洲山杨(*Populus tremula*)^[127],侧柏(*Platycladus orientalis*)^[128]和香杨(*Populus koreana*)^[121]。这种景观基因组学预测方法有助于了解物种在气候变化下可能的脆弱性区域,并反映了当地种群的适应情况,从而为森林树木的保护提供了新的见解^[85]。同时对森林生态系统的可持续管理至关重要,并为设计采样策略、针对区域进行保护或保护区设计以及预测环境变化的区域影响提供了重要资源^[85,124]。

5 结论和展望

5.1 结论

本研究通过采集 18 个种群的 24 个川滇高山栎个体和 18 个种群的 25 个刺叶高山栎个体，以川滇高山栎作为参考基因组，通过一系列严格的过滤步骤，最终获取了川滇高山栎的 358 661 个 SNPs 和刺叶高山栎的 346 281 个 SNPs。基于 ADMIXTURE、PCA 和系统发育分析揭示了川滇高山栎和刺叶高山栎物种内和物种间的遗传结构和系统发育关系；计算了每个谱系的遗传多样性指数 (π , Tajima' D, LD)；使用 PSMC 和 Fastsimcoal2 推测了两个栎属物种每个谱系的 N_e 的变化、基因流事件和遗传分化时间；基于 10 kbp 非重叠滑动窗口法计算了 *Q. aquifolioides* vs. *Q. spinosa*、*Qa_W* vs. *Qa_E* 和 *Qs_W* vs. *Qs_E* 检测基因组中的“遗传分化岛 (Genetic islands)”；在 Dsuite 中使用重叠滑动窗口法计算 f_{AM} 值，检测物种间的潜在渐渗区域；基于 LFMM 和 RDA 两种 GEAs 方法检测青藏高原和非青藏高原种群的适应性 SNPs 和功能基因，同时整合遗传信息和环境因素进行线性和非线性分析；结合 f_{AM} 和 GEAs 方法检测参与物种间适应性渐渗的功能基因，并分析参与适应性渐渗的基因的遗传特征；使用携带目标基因质粒的农杆菌侵染新鲜的烟草叶片，探究参与介导适应性遗传变异的因素；使用 GF 模型预测两种碳排放情景下 (RCP 2.6 和 RCP 8.5) 2050 年和 2070 年川滇高山栎和刺叶高山栎的分布区的脆弱性区域，以及适应性渐渗对于物种应对快速气候变化的影响。基于上述的分析内容得到了如下的结论：

(1) 川滇高山栎和刺叶高山栎都存在明显的东-西谱系分化，物种内部和物种之间都存在着非对称的基因流。基因流方向主要是从川滇高山栎到刺叶高山栎，从西谱系到东谱系。强的基因流是从 *Qa_E* 谱系到 *Qs_W* 谱系；

(2) 两个栎属物种的四个谱系在 1 亿年前到 10 万年前可能经历了相似的历史事件。川滇高山栎和刺叶高山栎祖先种群大约在 14.88 Mya 发生物种分化事件。随后，在约 8.98 Mya，刺叶高山栎形成两个谱系 (*Qs_W*, *Qs_E*)；在约 3.04 Mya，川滇高山栎祖先种群形成 *Qa_W* 和 *Qa_E* 两个谱系；

(3) F_{ST} 和 XP-CLR 在 *Q. aquifolioides* vs. *Q. spinosa*、*Qa_W* vs. *Qa_E* 和 *Qs_W* vs. *Qs_E* 中检测到共同的 744、1 013 和 988 个“遗传分化岛 (Genetic islands)”。此外，刺叶高山栎的物种内部谱系之间 (*Qs_W* vs. *Qs_E*) 的遗传分化高于物种之间 (*Q. aquifolioides* vs. *Q. spinosa*) 的遗传分化；

(4) 川滇高山栎和 *Qs_W* 之间存在 1 461 渐渗区域，这些区域中存在许多参与应对生物或非生物胁迫的功能基因，例如：*ERD3*、*HSP83*、*KAS2*、*ROF1*、*ATDTX28*、*RD2*、*SUD7*、*DWF3*、*SSN1*、*REF3*、*CYP81D8* 和 *MYB4* 等；

(5) LFMM 和 RDA 方法在青藏高原种群和非青藏高原种群检测到共同的 6600 和 662 个适应性 SNPs，这些 SNPs 涉及了 2 159 和 662 个基因。GO 富集分析在青藏高

原种群中富集到一些与非生物胁迫相关的结果，例如：GO:0009605、GO:0006952、GO:0033554 等；

(6) 川滇高山栎和 *Qs_W* 之间存在 435 个参与适应性渐渗的功能基因，这些功能基因拥有更长的碱基序列；

(7) *KAS2* 基因内的适应性位点 (3_82758545) 的 -A 等位基因 (A allele) 主要集中在青藏高原种群中，-G 等位基因 (G allele) 主要集中在非青藏高原种群中。实验结果发现了候选位点的 -A 等位基因会在黑暗、低温或者持续性紫外线条件下促进基因表达量的增加；

(8) GF 模型预测发现川滇高山栎的脆弱性区域主要在青藏高原的北部和西部，刺叶高山栎的脆弱性区域主要在横断山脉地区。

5.2 展望

本研究结合了种群遗传学和景观基因组学研究方法系统的分析了川滇高山栎和刺叶高山栎之间的适应性渐渗对于物种内部和物种之间遗传分化、局地适应的影响。但依然存在一些尚未解决的问题：

(1) 本研究对物种间适应性渐渗的研究主要是针对功能基因，但是对于全基因组中基因间区却忽略了。这些区域往往存在着一些调控元件，可以调控基因的表达水平。本研究也发现适应性遗传变异主要集中在调控区域，下一步需要重点关注基因间区中的适应性渐渗遗传机制。

(2) 川滇高山栎和刺叶高山栎物种之间的适应性渐渗不仅会影响基因组中遗传多样性水平和遗传分化，同时促进了物种的适应性进化。本研究主要针对的是适应性渐渗与遗传变异之间的关联，而忽略其对于表型变化的影响，下一步可以结合 f_{dM} 、基因型-表型关联分析方法，探究适应性渐渗对于栎属物种叶片表型性状的影响。

(3) 对于 *KAS2* 基因和特定 SNP (3_82758545) 的功能验证都是通过携带相应质粒的农杆菌注射到烟草叶片进行的间接证明，还未在栎属物种本身完成相应的功能验证。下一步可以借助同质园实验和转录组数据，实现对于基因功能的验证和基因表达量的测定。

(4) 本研究对于两个栎属物种的未来预测仅使用了一种气候数据模型 (BCC-CSM1-1)，后续可进行不同数据模型预测结果的相结合揭示更为精确的未来生存状况。此外，研究中未考虑植物自身的迁移情况等带来的影响，可能与真实的情况存在偏差。

参考文献

- [1] Anderson E, Hubricht L. Hybridization in *Tradescantia*. III. The evidence for introgressive for introgressive hybridization[J]. *American Journal of Botany*, 1938, 25(6): 396-402.
- [2] Anderson E. Introgressive hybridization[M]. John Wiley and Sons, New York, 1949.
- [3] Aguillon S M, Dodge T O, Preising G A, et al. Introgression[J]. *Current Biology*, 2022, 32(16): R865-R868.
- [4] Mallet J. Hybridization as an invasion of the genome[J]. *Trends in Ecology and Evolution*, 2005, 20(5): 229-237.
- [5] Harrison R G, Larson E L, Mallet J. Hybridization as an invasion of the genome[J]. *Trends in Ecology and Evolution*, 2005, 20(5): 229-237.
- [6] Leroy T, Louvet J M, Lalanne C, et al. Adaptive introgression as a driver of local adaptation to climate in European white oaks[J]. *New Phytologist*, 2020, 226(4): 1171-1182.
- [7] Fu R, Zhu Y, Liu Y, et al. Genome-wide analyses of introgression between two sympatric Asian oak species[J]. *Nature Ecology and Evolution*, 2022, 6(7): 924-935.
- [8] Liu S, Zhang L, Sang Y, et al. Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across *Populus* genome[J]. *Molecular Biology and Evolution*, 2022, 39(2): msac008.
- [9] Guo J F, Zhao W, Andersson B, et al. Genomic clines across the species boundary between a hybrid pine and its progenitor in the eastern Tibetan Plateau[J]. *Plant Communications*, 2023, 4(4) 100574.
- [10] Suarez-Gonzalez A, Lexer C, Cronk Q C B. Adaptive introgression: a plant perspective[J]. *Biology Letters*, 2018, 14(3): 20170688.
- [11] Whitney K D, Randell R A, Rieseberg L H. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*[J]. *The American Naturalist*, 2006, 167(6): 794-807.
- [12] Milne R I, Abbott R J. Origin and evolution of invasive naturalized material of *Rhododendron ponticum* L. in the British Isles[J]. *Molecular Ecology*, 2000, 9(5): 541-556.
- [13] Martinsen G D, Whitham T G, Turek R J, et al. Hybrid populations selectively filter gene introgression between species[J]. *Evolution*, 2001, 55(7): 1325-1335.
- [14] Uecker H, Setter D, Hermisson J. Adaptive gene introgression after secondary contact[J]. *Journal of Mathematical Biology*, 2015, 70(7): 1523-1580.
- [15] Tigano A, Friesen V L. Genomics of local adaptation with gene flow[J]. *Molecular Ecology*, 2016, 25(10): 2144-2164.
- [16] Barton N H. The role of hybridization in evolution[J]. *Molecular Ecology*, 2001, 10(3): 551-568.
- [17] Fraïsse C, Belkhir K, Welch J J, et al. Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels[J]. *Molecular Ecology*, 2016, 25(1): 269-286.
- [18] Isabel N, Holliday J A, Aitken S N. Forest genomics: Advancing climate adaptation, forest health, productivity, and conservation[J]. *Evolutionary Applications*, 2019, 13(1): 3-10.
- [19] Zhang Y, Li B, Zheng D. Datasets of the boundary and area of the Tibetan Plateau[J]. *Acta Geographica Sinica*, 2014, 69(Suppl.1): 65-68.
- [20] Wu S, Wang Y, Wang Z, et al. Species divergence with gene flow and hybrid speciation on the Qinghai-Tibet Plateau[J]. *New Phytologist*, 2022, 234(2): 392-404.
- [21] Wu Z Y, Wu S G. A proposal for a new floristic kingdom (realm)-the E. Asiatic kingdom, its delimitation and characteristics. In: Zhang AL, Wu SG (eds) *Proceedings of the first international*

symposium on floristic characteristics and diversity of east Asian plants[M]. Beijing: China Higher Education Press, 1996.

[22] Wen J, Zhang J, Nie Z L, et al. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau[J]. *Frontiers in Genetics*, 2014, 5(4): 73798.

[23] Du F K, Wang T, Wang Y, et al. Contrasted patterns of local adaptation to climate change across the range of an evergreen oak, *Quercus aquifolioides*[J]. *Evolutionary Applications*, 2020, 13(9): 2377-2391.

[24] Sun H, Zhang J, Deng T, et al. Origins and evolution of plant diversity in the Hengduan Mountains, China[J]. *Plant Diversity*, 2017, 39(4): 161.

[25] Favre A, Päckert M, Pauls S U, et al. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas[J]. *Biological Reviews*, 2015, 90(1): 236-253.

[26] 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.

[27] Liu J Q, Duan Y W, Hao G, et al. Evolutionary history and underlying adaptation of alpine plants on the Qinghai-Tibet Plateau[J]. *Journal of Systematics and Evolution*, 2014, 52(3): 241-249.

[28] Ding L, Spicer R A, Yang J, et al. Quantifying the rise of the Himalaya orogen and implications for the South Asian monsoon[J]. *Geology*, 2017, 45(3): 215-218.

[29] Zachos J, Pagani M, Sloan L, et al. Trends, rhythms, and aberrations in global climate 65 Ma to present[J]. *Science*, 2001, 292(5517): 686-693.

[30] Spicer R A. Tibet, the Himalaya, Asian monsoons and biodiversity-In what ways are they related?[J]. *Plant diversity*, 2017, 39(5): 233-244.

[31] Mao K S, Wang Y, Liu J Q. Evolutionary origin of species diversity on the Qinghai-Tibet Plateau[J]. *Journal of Systematics and Evolution*, 2021, 59(6): 1142-1158.

[32] López-Pujol J, Zhang F M, Sun H Q, et al. Centres of plant endemism in China: places for survival or for speciation?[J]. *Journal of Biogeography*, 2011, 38(7): 1267-1280.

[33] Jia D R, Abbott R J, Liu T L, et al. Out of the Qinghai-Tibet Plateau: evidence for the origin and dispersal of Eurasian temperate plants from a phylogeographic study of *Hippophaë rhamnoides* (Elaeagnaceae)[J]. *New Phytologist*, 2012, 194(4): 1123-1133.

[34] Myers N, Mittermeier R A, Mittermeier C G, et al. Biodiversity hotspots for conservation priorities[J]. *Nature*, 2000, 403(6772): 853-858.

[35] Marchese C. Biodiversity hotspots: A shortcut for a more complicated concept[J]. *Global Ecology and Conservation*, 2015, 3: 297-309.

[36] He K, Jiang X. Sky islands of southwest China. I: an overview of phylogeographic patterns[J]. *Chinese Science Bulletin*, 2014, 59(7): 585-597.

[37] Luo D, Yue J P, Sun W G, et al. Evolutionary history of the subnival flora of the Himalaya-Hengduan Mountains: First insights from comparative phylogeography of four perennial herbs[J]. *Journal of Biogeography*, 2016, 43(1): 31-43.

[38] Li J L, Zhong L L, Wang J, et al. Genomic insights into speciation history and local adaptation of an alpine aspen in the Qinghai-Tibet Plateau and adjacent highlands[J]. *Journal of Systematics and Evolution*, 2021, 59(6): 1220-1231.

[39] Xia M Z, Li Y, Zhang F Q, et al. Reassessment of the phylogeny and systematics of Chinese *Parnassia* (Celastraceae): A thorough investigation using whole plastomes and nuclear ribosomal DNA[J]. *Frontiers Plant Science*, 2022, 13: 855944.

- [40] Zhang T C, Comes H P, Sun H. Chloroplast phylogeography of *Terminalia franchetii* (Combretaceae) from the eastern Sino-Himalayan region and its correlation with historical river capture events[J]. *Molecular Phylogenetics and Evolution*, 2011, 60(1): 1-12.
- [41] Du F K., Qi M., Zhang Y Y, et al. Asymmetric character displacement in mixed oak stands[J]. *New Phytologist*, 2022, 236(3): 1212-1224.
- [42] Liu K, Du F K. Population and landscape genetics provide insights into species conservation of two evergreen oaks in Qinghai–Tibet Plateau and adjacent regions[J]. *Frontiers in Plant Science*, 2022, 13: 858526.
- [43] Manos P S, Stanford A M. The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the Northern Hemisphere[J]. *International Journal of Plant Sciences*, 2001, 162(S6): S77-S93.
- [44] Denk T, Grimm G W, Manos P S, et al. An updated infrageneric classification of the oaks: review of previous taxonomic schemes and synthesis of evolutionary patterns[M]. *Oaks physiological ecology. Exploring the functional diversity of genus *Quercus* L.* Cham, Switzerland: Springer, 2017.
- [45] Cavender-Bares J. Diversification, adaptation, and community assembly of the American oaks (*Quercus*), a model clade for integrating ecology and evolution[J]. *New Phytologist*, 2019, 221(2): 669-692.
- [46] Xu X, Dimitrov D, Shrestha N, et al. A consistent species richness–climate relationship for oaks across the Northern Hemisphere[J]. *Global Ecology and Biogeography*, 2019, 28(8): 1051-1066.
- [47] Petit R J, Carlson J, Curtu A L, et al. Fagaceae trees as models to integrate ecology, evolution and genomics[J]. *New Phytologist*, 2013, 197(2): 369-371.
- [48] Kremer A. Microevolution of European temperate oaks in response to environmental changes[J]. *Comptes Rendus. Biologies*, 2016, 339(7-8): 263-267.
- [49] Liang Y Y, Shi Y, Yuan S, et al. Linked selection shapes the landscape of genomic variation in three oak species[J]. *New Phytologist*, 2022, 233(1): 555-568.
- [50] Darwin C. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life[M]. London, Murray, 1872.
- [51] Rieseberg L H, Wood T E, Baack E J. The nature of plant species[J]. *Nature*, 2006, 440(7083): 524-527.
- [52] Darwin, C. On the origin of species: facsimile of the first edition[M]. Cambridge, MA: Harvard University Press, 1859
- [53] Palmer E J. Hybrid oaks of North America[J]. *Journal of the Arnold Arboretum*, 1948, 29(1): 1-48.
- [54] Rodríguez-Rajo F J, Méndez J, Jato V. Factors affecting pollination ecology of *Quercus anemophilous* species in north-west Spain[J]. *Botanical Journal of the Linnean Society*, 2005, 149(3): 283-297.
- [55] Feng L, Li Z H, Zhao G F. Genetic structure and evolutionary history of three alpine sclerophyllous oaks in east Himalaya-Hengduan Mountains and adjacent regions[J]. *Frontiers in Plant Science*, 2016, 7: 221731.
- [56] Ju M, Yang J, Yue M, et al. Speciation patterns of related species under the hybrid zone: A case study of three sclerophyllous oaks in the east Himalaya-Hengduan Mountains[J]. *Molecular Ecology*, 2023, 32(16): 4610-4626.
- [57] Ma X G, Ren Y B, Sun H. Introgression and incomplete lineage sorting blurred phylogenetic relationships across the genomes of sclerophyllous oaks from southwest China[J]. *Cladistics*, 2024.

- [58] Richards E, Reichardt M, Rogers S. Preparation of genomic DNA from plant tissue[J]. *Current Protocols in Molecular Biology*, 1994, 27(1): 2.3. 1-2.3. 7.
- [59] Chen Y, Chen Y, Shi C, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data[J]. *Gigascience*, 2018, 7(1): gix120.
- [60] Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform[J]. *Bioinformatics*, 2010, 26(5): 589-595.
- [61] 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.
- [62] Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses[J]. *The American Journal of Human Genetics*, 2007, 81(3): 559-575.
- [63] Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools[J]. *Bioinformatics*, 2011, 27(15): 2156-2158.
- [64] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism[J]. *Genetics*, 1989, 123(3): 585-595.
- [65] Zhang C, Dong S S, Xu J Y, et al. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files[J]. *Bioinformatics*, 2019, 35(10): 1786-1788.
- [66] Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals[J]. *Genome Research*, 2009, 19(9): 1655-1664.
- [67] Jombart T. adegenet: a R package for the multivariate analysis of genetic markers[J]. *Bioinformatics*, 2008, 24(11): 1403-1405.
- [68] Pickrell J, Pritchard J. Inference of population splits and mixtures from genome-wide allele frequency data[J]. *PLoS Genetics*, 2012, 8(11): e1002967.
- [69] Li H, Durbin R. Inference of human population history from individual whole-genome sequences[J]. *Nature*, 2011, 475(7357): 493-496.
- [70] Sork V L, Cokus S J, Fitz-Gibbon S T, et al. High-quality genome and methylomes illustrate features underlying evolutionary success of oaks[J]. *Nature Communications*, 2022, 13(1): 2047.
- [71] Excoffier L, Dupanloup I, Huerta-Sánchez E, et al. Robust demographic inference from genomic and SNP data[J]. *PLoS Genetics*, 2013, 9(10): e1003905.
- [72] Alexa A, Rahnenführer J. Gene set enrichment analysis with topGO[J]. *Bioconductor Improv*, 2009, 27: 1-26.
- [73] Malinsky M, Matschiner M, Svardal H. Dsuite-Fast D-statistics and related admixture evidence from VCF files[J]. *Molecular Ecology Resources*, 2021, 21(2): 584-595.
- [74] Morales-Cruz A, Aguirre-Liguori J A, Zhou Y, et al. Introgression among North American wild grapes (*Vitis*) fuels biotic and abiotic adaptation[J]. *Genome Biology*, 2021, 22(1): 254.
- [75] 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.
- [76] 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.
- [77] Capblancq T, Forester B R. Redundancy analysis: A Swiss Army Knife for landscape genomics[J]. *Methods in Ecology and Evolution*, 2021, 12(12): 2298-2309.
- [78] Frichot E, François O. LEA: An R package for landscape and ecological association studies[J]. *Methods in Ecology and Evolution*, 2015, 6(8): 925-929.

- [79] Oksanen J, Blanchet F G, Friendly M, et al. Community ecology package: ordination methods, diversity analysis and other functions for community and vegetation ecologists[J]. Version, 2016, 2: 4.
- [80] 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.
- [81] Manthey J D, Moyle R G. Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: A landscape genomics approach[J]. *Molecular Ecology*, 2015, 24(14): 3628-3638.
- [82] Hijmans R J, Williams E, Vennes C, et al. Package 'geosphere'[J]. *Spherical Trigonometry*, 2017, 1(7): 1-45.
- [83] Goslee S C, Urban D L. The ecodist package for dissimilarity-based analysis of ecological data[J]. *Journal of Statistical Software*, 2007, 22: 1-19.
- [84] 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.
- [85] 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.
- [86] Ellis N, Smith S J, Pitcher C R. Gradient forests: calculating importance gradients on physical predictors[J]. *Ecology*, 2012, 93(1): 156-168.
- [87] Song X, Weng Q, Zhao Y, et al. Cloning and Expression Analysis of *ZmERD3* Gene From *Zea mays*[J]. *Iranian Journal of Biotechnology*, 2018, 16(2): e1593
- [88] Conner T W, LaFayette P R, Nagao R T, et al. Sequence and expression of a HSP83 from *Arabidopsis thaliana*[J]. *Plant physiology*, 1990, 94(4): 1689-1695.
- [89] Meiri D, Breiman A. Arabidopsis ROF1 (FKBP62) modulates thermotolerance by interacting with HSP90. 1 and affecting the accumulation of HsfA2-regulated sHSPs[J]. *The Plant Journal*, 2009, 59(3): 387-399.
- [90] Calhoun S, Bell T A S, Dahlin L R, et al. A multi-omic characterization of temperature stress in a halotolerant *Scenedesmus* strain for algal biotechnology[J]. *Communications Biology*, 2021, 4(1): 333.
- [91] Shahzad A, Qian M, Sun B, et al. Genome-wide association study identifies novel loci and candidate genes for drought stress tolerance in rapeseed[J]. *Oil Crop Science*, 2021, 6(1): 12-22.
- [92] He L, Li L, Zhu Y, et al. *BolTLP1*, a thaumatin-like protein gene, confers tolerance to salt and drought stresses in broccoli (*Brassica oleracea* L. var. *Italica*) [J]. *International Journal of Molecular Sciences*, 2021, 22(20): 11132.
- [93] Talloji P, Nehlin L, Hüttel B, et al. Transcriptome, metabolome and suppressor analysis reveal an essential role for the ubiquitin-proteasome system in seedling chloroplast development[J]. *BMC Plant Biology*, 2022, 22(1): 183.
- [94] Hussain Q, Shi J, Scheben A, et al. Genetic and signalling pathways of dry fruit size: targets for genome editing-based crop improvement[J]. *Plant Biotechnology Journal*, 2020, 18(5): 1124-1140.
- [95] Durrant W E, Wang S, Dong X. *Arabidopsis* SNI1 and RAD51D regulate both gene transcription and DNA recombination during the defense response[J]. *Proceedings of the National Academy of Sciences*, 2007, 104(10): 4223-4227.
- [96] Schillmiller A L, Stout J, Weng J K, et al. Mutations in the *cinnamate 4-hydroxylase* gene impact metabolism, growth and development in *Arabidopsis*[J]. *The Plant Journal*, 2009, 60(5): 771-782.

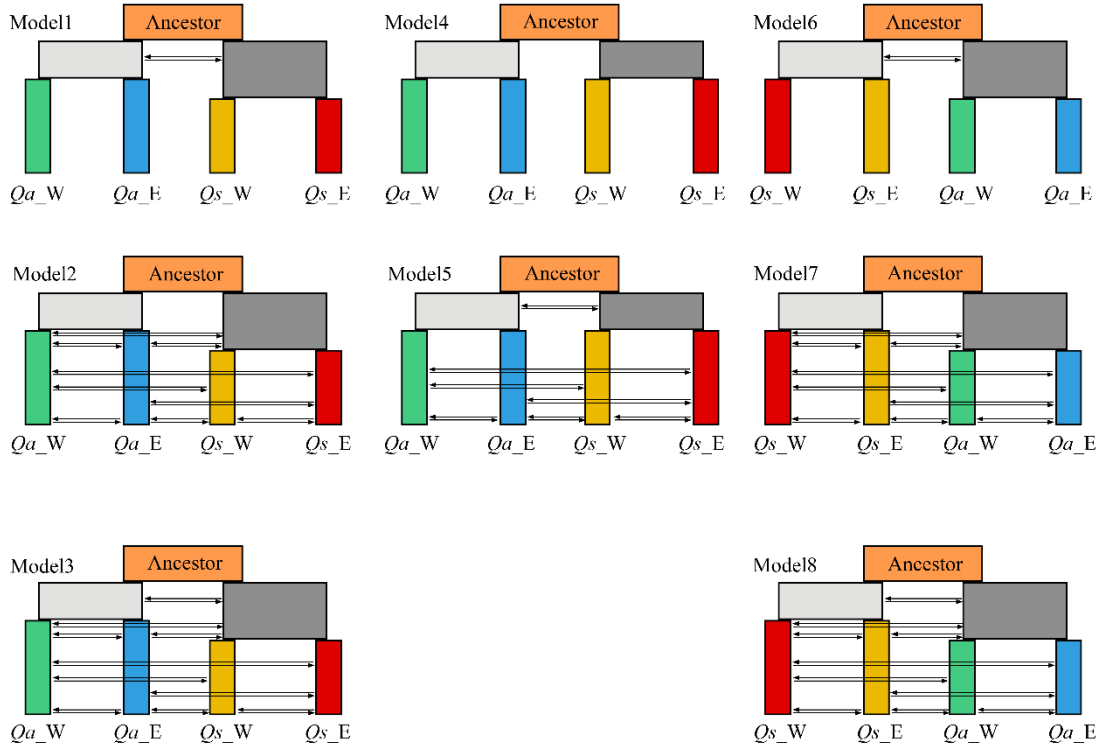
- [97] Panozzo A, Dal Cortivo C, Ferrari M, et al. Morphological changes and expressions of *AOX1A*, *CYP81D8*, and putative *PPF* genes in a large set of commercial maize hybrids under extreme waterlogging[J]. *Frontiers in Plant Science*, 2019, 10: 434224.
- [98] Carlsson A S, LaBrie S T, Kinney A J, et al. A *KAS2* cDNA complements the phenotypes of the *Arabidopsis fab1* mutant that differs in a single residue bordering the substrate binding pocket[J]. *The Plant Journal*, 2002, 29(6): 761-770.
- [99] Hughes C E, Atchison G W. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains[J]. *New Phytologist*, 2015, 207(2): 275-282.
- [100] Ju M M, Feng L, Yang J, et al. Evaluating population genetic structure and demographic history of *Quercus spinosa* (Fagaceae) based on specific length amplified fragment sequencing[J]. *Frontiers in Genetics*, 2019, 10: 460476.
- [101] Zhao W, Sun Y Q, Pan J, et al. Effects of landscapes and range expansion on population structure and local adaptation[J]. *New Phytologist*, 2020, 228(1): 330-343.
- [102] Ma Y P, Wariss H M, Liao R L, et al. Genome-wide analysis of butterfly bush (*Buddleja alternifolia*) in three uplands provides insights into biogeography, demography and speciation[J]. *New Phytologist*, 2021, 232(3): 1463-1476.
- [103] An Z S, John E K, Warren L P, et al. Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan plateau since late Miocene times[J]. *Nature*, 2001, 411(6833): 62- 66.
- [104] Li G, Kim C, Zha H, et al. Molecular phylogeny and biogeography of the arctic-alpine genus *Lagotis* (Plantaginaceae)[J]. *Taxon*, 2014, 63(1): 103-115.
- [105] Nie Z L, Funk V, Sun H, et al. Molecular phylogeny of *Anaphalis* (Asteraceae, Gnaphalieae) with biogeographic implications in the Northern Hemisphere[J]. *Journal of Plant Research*, 2013, 126: 17-32.
- [106] Safer S, Tremetsberger K, Guo Y P, et al. Phylogenetic relationships in the genus *Leontopodium* (Asteraceae: Gnaphalieae) based on AFLP data[J]. *Botanical journal of the Linnean society*, 2011, 165(4): 364-377.
- [107] Gourbet L, Leloup P H, Paquette J L, et al. Reappraisal of the Jianchuan Cenozoic basin stratigraphy and its implications on the SE Tibetan plateau evolution[J]. *Tectonophysics*, 2017, 700: 162-179.
- [108] Morjan C L, Rieseberg L H. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles[J]. *Molecular Ecology*, 2004, 13(6): 1341-1356.
- [109] Choler P, Erschbamer B, Tribsch A, et al. Genetic introgression as a potential to widen a species' niche: insights from alpine *Carex curvula*[J]. *Proceedings of the National Academy of Sciences*, 2004, 101(1): 171-176.
- [110] Taylor E B, Boughman J W, Groenenboom M, et al. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair[J]. *Molecular Ecology*, 2006, 15(2): 343-355.
- [111] Givnish T J. Adaptive radiation versus 'radiation' and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution[J]. *New Phytologist*, 2015, 207(2): 297-303
- [112] Edelman N B, Frandsen P B, Miyagi M, et al. Genomic architecture and introgression shape a butterfly radiation[J]. *Science*, 2019, 366(6465): 594-599.
- [113] Grishkevich V, Yanai I. Gene length and expression level shape genomic novelties[J]. *Genome Research*, 2014, 24(9): 1497-1503.

- [114] Ingvarsson P K, Bernhardsson C. Genome-wide signatures of environmental adaptation in European aspen (*Populus tremula*) under current and future climate conditions[J]. *Evolutionary Applications*, 2019, 13(1): 132-142.
- [115] Heraghty S D, Jackson J M, Lozier J D. Whole genome analyses reveal weak signatures of population structure and environmentally associated local adaptation in an important North American pollinator, the bumble bee *Bombus vosnesenskii*[J]. *Molecular Ecology*, 2023, 32(20): 5479-5497.
- [116] Capblancq T, Lachmuth S, Fitzpatrick M C, et al. From common gardens to candidate genes: exploring local adaptation to climate in red spruce[J]. *New phytologist*, 2023, 37(5): 1590-1605.
- [117] Lowry D B, Rockwood R C, Willis J H. Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*[J]. *Evolution*, 2008, 62(9): 2196-2214.
- [118] Whitney K D, Randell R A, Rieseberg L H. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*[J]. *New phytologist*, 2010, 187(1): 230-239.
- [119] Ma Y, Wang J I, Hu Q, et al. Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex[J]. *Communications Biology*, 2019, 2(1): 213.
- [120] Zhang Z R, Li W Y, Dong Y Y, et al. Geographic cline and genetic introgression effects on seed morphology variation and germination fitness in two closely related pine species in Southeast Asia[J]. *Forests*, 2022, 13(3): 374.
- [121] Sang Y, Long Z, Dan X, et al. Genomic insights into local adaptation and future climate-induced vulnerability of a keystone forest tree in East Asia[J]. *Nature Communications*, 2022, 13(1): 6541.
- [122] Hudson T J: Wanted: regulatory SNPs[J]. *Nature Genetic*, 2003, 33(4): 439-440.
- [123] Liljedahl U, Fredriksson M, Dahlgren A, et al. Detecting imbalanced expression of SNP alleles by minisequencing on microarrays[J]. *BMC Biotechnology*, 2004, 4: 1-10.
- [124] Feng L, Du F K. Landscape genomics in tree conservation under a changing environment[J]. *Frontiers in Plant Science*, 2022, 13: 822217.
- [125] Capblancq T, Fitzpatrick M C, Bay R A, et al. Genomic prediction of (mal) adaptation across current and future climatic landscapes[J]. *Annual Review of Ecology, Evolution, and Systematics*, 2020, 51(1): 245-269.
- [126] Pina-Martins F, Baptista J, Pappas Jr G, et al. New insights into adaptation and population structure of cork oak using genotyping by sequencing[J]. *Global Change Biology*, 2019, 25(1): 337-350.
- [127] Ingvarsson P K, Bernhardsson C. Genome-wide signatures of environmental adaptation in European aspen (*Populus tremula*) under current and future climate conditions[J]. *Evolutionary Applications*, 2020, 13(1): 132-142.
- [128] Jia K H, Zhao W, Maier P A, et al. Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (Cupressaceae)[J]. *Evolutionary Applications*, 2020, 13(4): 665-676.

附录

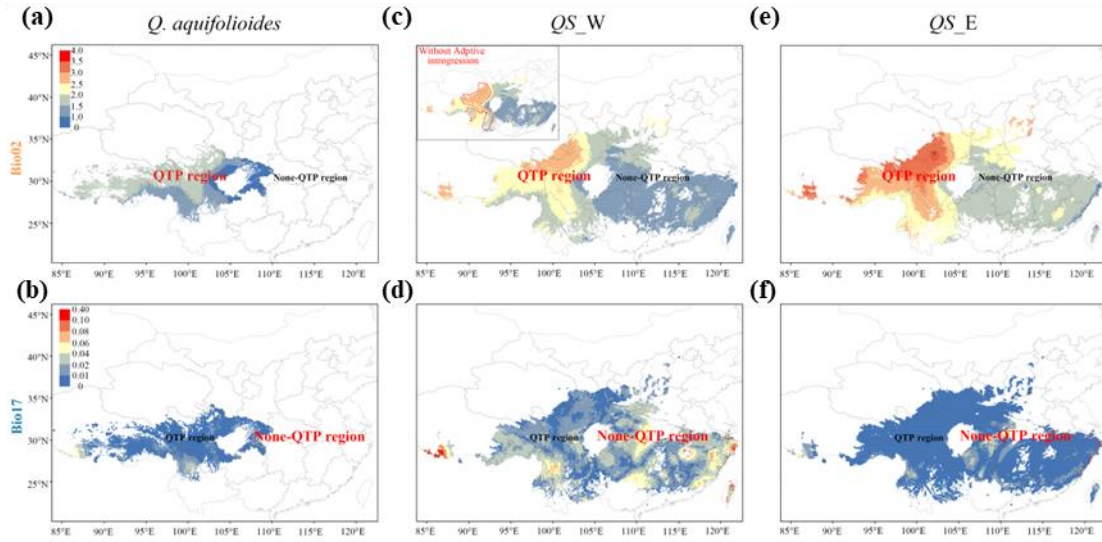
附录 A Fastsimcoal2 模拟的八个种群历史进化历史模型。

Appendix A. Eight population historical evolution models were simulated by the Fastsimcoal2



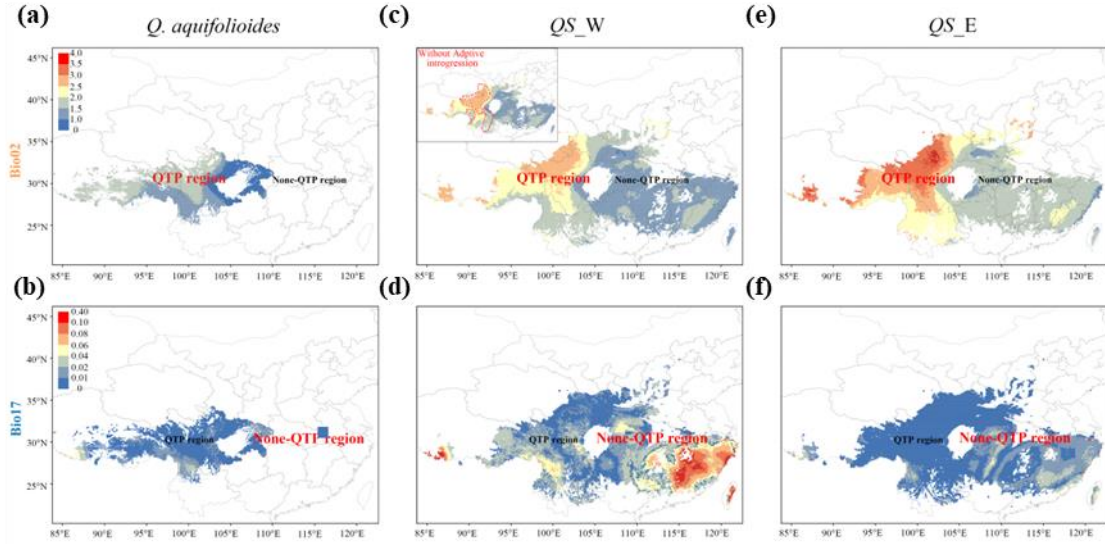
附录 B 在 2050 年 RCP 8.5 碳排放量情景下，两种栎属物种在平均气温日较差（Bio02）和最湿季度降水量（Bio17）气候条件下的遗传偏移。（a, b）川滇高山栎，（c, d）刺叶高山栎西谱系和（e, f）刺叶高山栎东谱系

Appendix B. The results of genetic offset of two *Quercus* species using Bio02 and Bio17 climatic variables in 2050 under RCP 8.5. (a, b) *Q. aquifolioides*, (c, d) *Qs_W* and (e, f) *Qs_E*



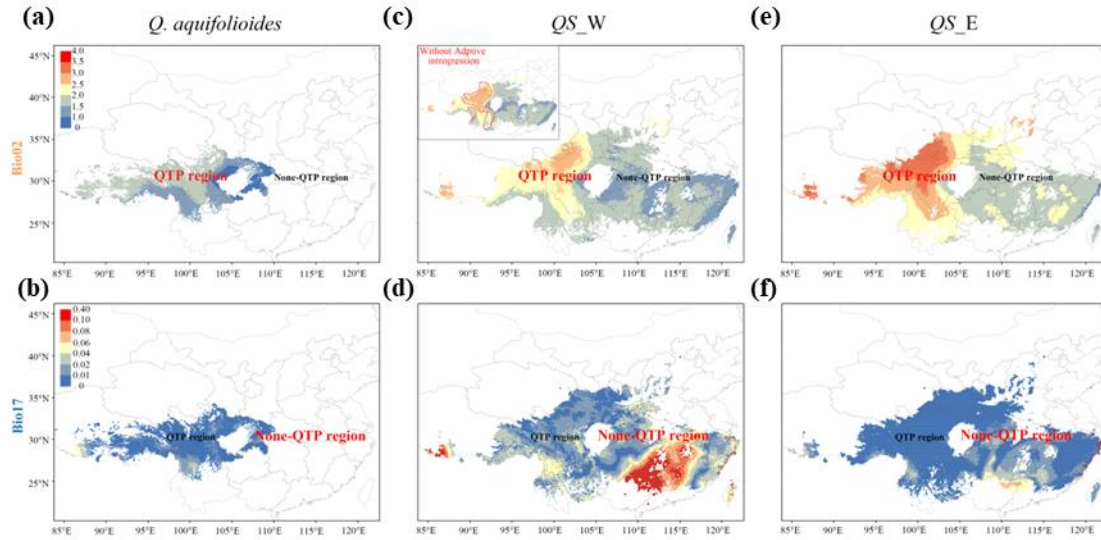
附录 C 在 2070 年 RCP 2.6 碳排放量情景下，两种栎属物种在平均气温日较差（Bio02）和最湿季度降水量（Bio17）气候条件下的遗传偏移。（a, b）川滇高山栎，（c, d）刺叶高山栎西谱系和（e, f）刺叶高山栎东谱系

Appendix C. The results of genetic offset of two *Quercus* species using Bio02 and Bio17 climatic variables in 2070 under RCP 2.6. (a, b) *Q. aquifolioides*, (c, d) *Qs_W* and (e, f) *Qs_E*



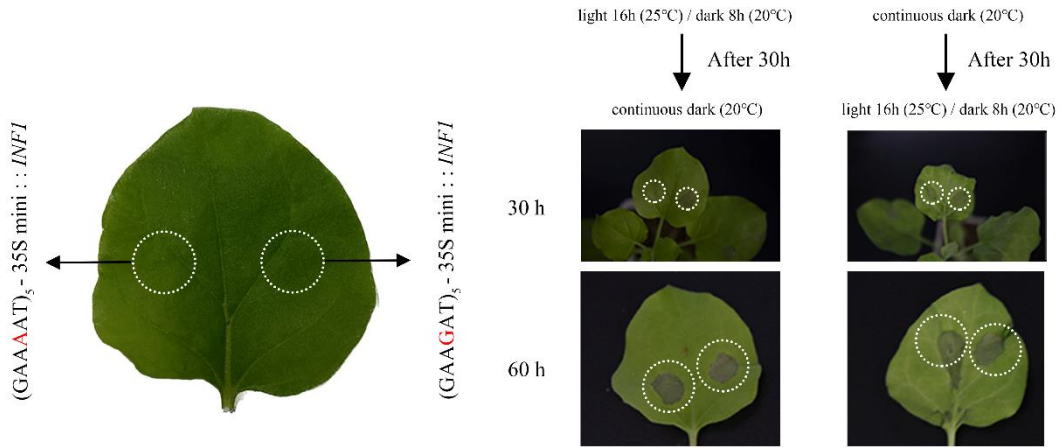
附录 D 在 2070 年 RCP 8.5 碳排放量情景下，两种栎属物种在平均气温日较差（Bio02）和最湿季度降水量（Bio17）气候条件下的遗传偏移。（a, b）川滇高山栎，（c, d）刺叶高山栎西谱系和（e, f）刺叶高山栎东谱系

Appendix D. The results of genetic offset of two *Quercus* species using Bio02 and Bio17 climatic variables in 2070 under RCP 8.5. (a, b) *Q. aquifolioides*, (c, d) *Qs_W* and (e, f) *Qs_E*



附录 E. 两种环境条件下两种基因型农杆菌侵染烟草叶片的反应结果。左边的虚线圈为注射((GAAAAT)₅ - 35S mini:: *INFI*)的区域，右边的虚线圈为注射((GAAAGAT)₅ - 35S mini:: *INFI*)的区域

Appendix E. Reaction of *Nicotiana benthamiana* leaves infiltrated by agrobacteria containing two genotypes under two environmental conditions. The dotted circles on the left side of the leaves were the region of injection ((GAAAAT)₅ - 35S mini:: *INFI*) and the right side of the leaves were the region of injection ((GAAAGAT)₅ - 35S mini:: *INFI*)



附录 F 川滇高山栎和刺叶高山栎测序和比对结果统计
Appendix F. Statistics of sequencing and mapping results of *Q. aquifolioides* and *Q. spinosa*

种群编号 (Population code)	测序深度 (depth)	原始读数 (raw reads)	无错读数 (clean reads)	原始数据 (bp) (raw data)	无错数据 (bp) (clean data)	读数 长度 (bp)	Q20 (%)	GC (%)	比对 read (Mapping read)	比对率 (%) (Mapping ratio)
<i>Qa_LZD_17</i>	27.819X	219383168	212825472	32907475200	31923820800	150	95.13	38.38	214637501	97.71
<i>Qa_GBX_2</i>	37.860X	297204112	289635454	44580616800	43445318100	150	95.41	37.24	293953794	98.34
<i>Qa_GBX_21</i>	39.270X	308811744	300426938	46321761600	45064040700	150	95.48	38.93	288732832	93.21
<i>Qa_MLJ_5</i>	32.874X	259578274	251491496	38936741100	37723724400	150	94.88	37.60	255304548	98.39
<i>Qa_LZA_26</i>	26.461X	209963332	202433570	31494499800	30365035500	150	94.85	37.73	205774543	98.52
<i>Qa_LZ_12</i>	35.861X	297473652	274349092	44621047800	41152363800	150	95.44	37.64	279542837	98.72
<i>Qa_BMR_9</i>	26.694X	209964678	204214316	31494701700	30632147400	150	94.87	37.83	208253034	98.84
<i>Qa_BMS_3</i>	33.633X	266201314	257300196	39930197100	38595029400	150	95.99	37.64	260739646	97.98
<i>Qa_CYX_14</i>	37.016X	293615644	283183644	44042346600	42477546600	150	95.35	37.61	286702832	97.98
<i>Qa_CYX_15</i>	34.572X	274984176	264487044	41247626400	39673056600	150	95.42	37.64	268900769	98.39
<i>Qa_MKD_14</i>	30.841X	244212544	235942986	36631881600	35391447900	150	94.91	37.53	238883057	98.05
<i>Qa_MKR_15</i>	35.711X	284800978	273197366	42720146700	40979604900	150	95.07	37.46	278985092	99.03
<i>Qa_BZL_19</i>	27.447X	216812616	209978532	32521892400	31496779800	150	94.72	37.68	213556436	98.18
<i>Qa_ZD_3</i>	43.160X	351915756	330187906	52787363400	49528185900	150	95.64	37.87	339307657	99.27
<i>Qa_XCR_6</i>	24.160X	195254690	184831208	29288203500	27724681200	150	95.12	37.26	189526594	99.09
<i>Qa_DCE_8</i>	28.223X	224094552	215914040	33614182800	32387106000	150	94.98	37.28	221203731	99.12
<i>Qa_YJH_2</i>	38.813X	307237232	296933602	46085584800	44540040300	150	95.61	37.40	303603641	99.08

附录

Qa_YJH_5	32.843X	257969890	251259264	38695483500	37688889600	150	95.11	37.56	254219146	98.03
Qa_DMx_7	31.492X	247367892	240920886	37105183800	36138132900	150	94.93	37.53	246823137	99.11
Qa_DMx_8	31.960X	250895674	244506206	37634351100	36675930900	150	95.44	37.37	249906247	98.85
Qa_JCK_7	36.657X	288653828	280434602	43298074200	42065190300	150	95.75	37.47	287274376	98.99
Qa_JCK_12	37.912X	300066766	290033146	45010014900	43504971900	150	95.70	37.71	296268091	98.81
Qa_JCK_14	38.966X	306398242	298097130	45959736300	44714569500	150	95.15	37.78	304303485	98.66
Qa_PW_6	33.361X	261721410	255220584	39258211500	38283087600	150	95.09	38.54	236103940	89.83
Qs_TM_1	34.543X	271561268	264265746	40734190200	39639861900	150	95.16	41.11	230946538	84.99
Qs_TM_3	31.324X	243637638	239637564	36545645700	35945634600	150	96.20	37.26	243482023	98.21
Qs_CYXC_2	32.384X	254548784	247746774	38182317600	37162016100	150	95.88	37.66	252158298	98.50
Qs_BLLK_9	37.216X	290214906	284711856	43532235900	42706778400	150	96.27	37.26	290648860	98.77
Qs_GZ_4	32.845X	256654216	251272332	38498132400	37690849800	150	96.69	37.21	256704476	98.88
Qs_XGJQ_1	33.559X	265015824	256733246	39752373600	38509986900	150	96.04	37.42	261197630	98.56
Qs_XGJQ_3	37.341X	293065508	285670630	43959826200	42850594500	150	96.06	37.43	289949375	98.21
Qs_XGJQ_2	34.333X	270003264	262654124	40500489600	39398118600	150	95.96	37.20	268489177	98.82
Qs_SWJ_2	40.212X	312399110	307630914	46859866500	46144637100	150	95.43	36.79	314182966	98.68
Qs_GDS_8	47.371X	368841690	362399486	55326253500	54359922900	150	97.33	36.97	370248570	98.48
Qs_LBZ_6	37.325X	288579388	285544338	43286908200	42831650700	150	94.88	36.57	291006406	98.50
Qs_LBZ_8	28.309X	221849504	216571112	33277425600	32485666800	150	95.43	37.48	220469770	98.51
Qs_JZG_2	35.397X	278015408	270799202	41702311200	40619880300	150	95.77	37.42	276390854	98.53
Qs_ZJH_9	38.679X	301547228	295903288	45232084200	44385493200	150	95.60	38.16	295977651	96.71
Qs_ZJH_1	30.974X	241362366	236960106	36204354900	35544015900	150	94.23	41.12	218117300	89.33

基于全基因组重测序技术川滇高山栎和刺叶高山栎适应性渐渗研究

<i>Qs_TJS_3</i>	36.906X	288389614	282344292	43258442100	42351643800	150	94.88	37.98	282522152	96.65
<i>Qs_TJS_5</i>	36.213X	282522938	277038418	42378440700	41555762700	150	94.86	37.55	280101984	97.62
<i>Qs_BDX_9</i>	30.327X	237466610	232013088	35619991500	34801963200	150	94.77	37.99	235280539	97.96
<i>Qs_DJP_22</i>	33.150X	260376016	253609760	39056402400	38041464000	150	95.00	37.28	259259378	98.56
<i>Qs_TDY_17</i>	33.938X	268269350	259634468	40240402500	38945170200	150	95.24	37.46	264378723	98.23
<i>Qs_WDS_1</i>	37.927X	300314080	290150614	45047112000	43522592100	150	95.38	37.31	297886509	98.98
<i>Qs_TMS_3</i>	30.385X	237348352	232455222	35602252800	34868283300	150	94.76	37.30	237607961	98.62
<i>Qs_TMS_8</i>	36.835X	288423978	281800222	43263596700	42270033300	150	95.14	37.13	288664087	98.60
<i>Qs_WKS_15</i>	36.872X	291287888	282084352	43693183200	42312652800	150	95.13	37.60	287862925	98.49
<i>Qs_SQS_3</i>	34.191X	269303486	261569978	40395522900	39235496700	150	95.34	37.43	266046498	98.08

注 a: Q20: 碱基正确识别率在 99%以上

注 b: GC: 鸟嘌呤和胞嘧啶所占的比率

附录

附录 G 川滇高山栎和刺叶高山栎 Fastsimcoal2 测试的八种模型的结果
 Appendix G. Results of eight models tested by Fastsimcoal2 for *Q. aquifolioides* and *Q. spinosa*

Model	Max(log10(likelihood)) ^a	No. of parameters	AIC ^b	δ^b	Model normalized relative likelihood (w) ^b
Model 1	-4022669	11	18525097	1202898	~0
Model 2	-3771618	27	17368997	46798	~0
Model 3	-3771106	29	17366643	44444	~0
Model 4	-3766921	22	17347356	25157	~0
Model 5	-4088879	8	18830000	1507801	~0
Model 6	-3986943	11	18360573	1038374	~0
Model 7	-3761456	27	17322199	0	1
Model 8	-3762346	29	17326302	4103	~0

注 a: 基于每个模型的 50 次独立运行中的最佳可能性

注 b: AIC_i、 Δ_i 和 w_i 的计算方法如 Excoffier 等 (2013) 所示

附录 H Fastsimcoal2 最佳拟合进化历史模型的 95%置信区间的参数估计
Appendix H. Inferred parameters estimates with 95% confidence intervals for the best fitting demographic history modelled in Fastsimcoal2

参数 (Parameters)	点估计 (Point estimation)	95%置信区间 (95% confidence intervals)	
		下限 (Lower bound)	上限 (Upper bound)
NPOPO	101288	92223	115632
NPOP1	101903	97256	120263
NPOP2	101640	94525	118263
NPOP3	104804	99251	129526
NQS	983376	945252	1015256
NQA	119686	101056	132526
NANC	107864	99232	126462
T1	3040470	2632512	3965261
T2	8979690	7252663	10026260
T3	14877450	1162233	15932321
$Qa_E \rightarrow Qa_W$	2.86E-05	2.82E-05	2.89E-05
$Qs_W \rightarrow Qa_W$	4.22E-09	3.95E-09	4.33E-09
$Qs_E \rightarrow Qa_W$	2.52E-07	2.49E-07	2.53E-07
$Qs_W \rightarrow Qa_E$	2.79E-05	2.76E-05	2.81E-05
$Qs_E \rightarrow Qa_E$	1.04E-05	1.02E-05	1.07E-05
$Qs_E \rightarrow Qs_W$	1.95E-08	1.93E-08	1.97E-08
$Qa_W \rightarrow Qa_E$	1.73E-05	1.71E-05	1.76E-05
$Qa_W \rightarrow Qs_W$	2.16E-05	2.14E-05	2.19E-05
$Qa_W \rightarrow Qs_E$	3.46E-08	3.44E-08	3.48E-08
$Qa_E \rightarrow Qs_W$	3.43E-05	3.40E-05	3.47E-05
$Qa_E \rightarrow Qs_E$	2.01E-08	1.98E-08	2.03E-08
$Qs_W \rightarrow Qs_E$	1.45E-05	1.41E-05	1.47E-05
$Qs_W \rightarrow Qa_W$	5.62E-08	5.60E-08	5.65E-08
$Qs_E \rightarrow Qa_W$	2.66E-08	2.62E-08	2.69E-08
$Qs_E \rightarrow Qs_W$	4.45E-08	4.41E-08	4.47E-08
$Qa_W \rightarrow Qs_W$	8.54E-06	8.51E-06	8.57E-06
$Qa_W \rightarrow Qs_E$	6.99E-08	6.96E-08	7.03E-08
$Qs_W \rightarrow Qs_E$	3.65E-07	3.62E-07	3.68E-07

附录

附录 I 川滇高山栎和刺叶高山栎之间“基因组岛”内基因的 GO 富集分析结果
Appendix I. GO enrichment results of genes located within the "genetic islands" of *Q. aquifolioides* vs. *Q. spinosa*

GO.ID	Term	Annotated	Significant	Expected	Fisher.p	Ontology
<i>Q. aquifolioides</i> vs. <i>Q. spinosa</i>						
GO:0005515	protein binding	1849	43	42.34	6.6E-09	MF
GO:0003700	DNA-binding transcription factor activity	228	10	5.22	0.00003	MF
GO:0003676	nucleic acid binding	1219	32	27.91	0.0003	MF
GO:0003677	DNA binding	489	16	11.2	0.00033	MF
GO:0004672	protein kinase activity	1000	20	22.9	0.00041	MF
GO:0005216	ion channel activity	72	5	1.65	0.00109	MF
GO:0005524	ATP binding	364	9	8.33	0.00436	MF
GO:0003690	double-stranded DNA binding	55	4	1.26	0.00669	MF
GO:0022836	gated channel activity	34	3	0.78	0.00709	MF
GO:0004252	serine-type endopeptidase activity	94	4	2.15	0.0088	MF
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	75	4	1.72	0.01122	MF
GO:0003824	catalytic activity	3977	80	91.06	0.01253	MF
GO:0008375	acetylglucosaminyltransferase activity	22	2	0.5	0.01511	MF
GO:0004180	carboxypeptidase activity	50	3	1.14	0.01695	MF
GO:0003712	transcription coregulator activity	27	2	0.62	0.02232	MF
GO:0016831	carboxy-lyase activity	27	2	0.62	0.02232	MF
GO:0009055	electron transfer activity	77	3	1.76	0.02862	MF
GO:0010181	FMN binding	32	2	0.73	0.03068	MF
GO:0140096	catalytic activity, acting on a protein	1418	33	32.47	0.03181	MF
GO:0000166	nucleotide binding	754	17	17.26	0.03226	MF
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	304	6	6.96	0.04774	MF
GO:0016020	membrane	624	13	13.32	0.0039	CC
GO:0005694	chromosome	53	3	1.13	0.0057	CC
GO:0005856	cytoskeleton	19	2	0.41	0.0097	CC
GO:0000160	phosphorelay signal transduction system	29	3	0.57	0.0012	BP
GO:0007034	vacuolar transport	12	2	0.24	0.0034	BP
GO:0008152	metabolic process	1453	23	28.66	0.0044	BP

基于全基因组重测序技术川滇高山栎和刺叶高山栎适应性渐渗研究

GO:0006401	RNA catabolic process	13	2	0.26	0.0073	BP
GO:0019725	cellular homeostasis	18	2	0.36	0.0077	BP
GO:0005975	carbohydrate metabolic process	95	3	1.87	0.0332	BP
GO:0009150	purine ribonucleotide metabolic process	22	2	0.43	0.043	BP

附录 J *Qa_W vs. Qa_E* “基因组岛”内基因的 GO 富集分析结果
 Appendix J. GO enrichment results of genes located within the "genetic islands" of *Qa_W vs. Qa_E*

GO.ID	Term	Annotated	Significant	Expected	Fisher.p	Ontology
<i>Qa_W vs. Qa_E</i>						
GO:0005515	protein binding	1849	48	47.13	8.50E-11	MF
GO:0003723	RNA binding	213	14	5.43	4.90E-06	MF
GO:0005524	ATP binding	364	12	9.28	0.0002	MF
GO:0008270	zinc ion binding	126	6	3.21	0.0013	MF
GO:0003700	DNA-binding transcription factor activity	228	8	5.81	0.0016	MF
GO:0003677	DNA binding	489	13	12.46	0.0028	MF
GO:0005506	iron ion binding	308	9	7.85	0.0028	MF
GO:0005525	GTP binding	70	4	1.78	0.0045	MF
GO:0008171	O-methyltransferase activity	11	2	0.28	0.0047	MF
GO:0016773	phosphotransferase activity, alcohol group as acceptor	1038	19	26.46	0.0051	MF
GO:0004842	ubiquitin-protein transferase activity	73	4	1.86	0.0052	MF
GO:0003824	catalytic activity	3977	83	101.38	0.0062	MF
GO:0016491	oxidoreductase activity	660	16	16.82	0.0073	MF
GO:0005198	structural molecule activity	222	6	5.66	0.0097	MF
GO:0004252	serine-type endopeptidase activity	94	4	2.4	0.0124	MF
GO:0003676	nucleic acid binding	1219	37	31.07	0.0151	MF
GO:0016798	hydrolase activity, acting on glycosyl bonds	306	5	7.8	0.0188	MF
GO:0003743	translation initiation factor activity	24	2	0.61	0.0216	MF
GO:0036094	small molecule binding	767	22	19.55	0.0265	MF
GO:0016866	intramolecular transferase activity	27	2	0.69	0.027	MF
GO:0005543	phospholipid binding	28	2	0.71	0.0289	MF
GO:0004672	protein kinase activity	1000	16	25.49	0.0298	MF
GO:0015079	potassium ion transmembrane transporter activity	29	2	0.74	0.0308	MF
GO:0010181	FMN binding	32	2	0.82	0.0369	MF
GO:0005622	intracellular anatomical structure	546	10	10.75	0.0055	CC
GO:0016020	membrane	624	11	12.28	0.017	CC
GO:0008152	metabolic process	1453	49	48.97	1.30E-05	BP

基于全基因组重测序技术川滇高山栎和刺叶高山栎适应性渐渗研究

GO:0005975	carbohydrate metabolic process	95	5	3.2	0.0076	BP
GO:0006629	lipid metabolic process	171	7	5.76	0.0133	BP
GO:0045454	cell redox homeostasis	14	2	0.47	0.0137	BP
GO:0016579	protein deubiquitination	17	2	0.57	0.02	BP
GO:1901137	carbohydrate derivative biosynthetic process	53	5	1.79	0.0246	BP
GO:0009165	nucleotide biosynthetic process	18	2	0.61	0.0256	BP
GO:0018193	peptidyl-amino acid modification	26	2	0.88	0.0256	BP
GO:0006855	xenobiotic transmembrane transport	50	3	1.69	0.0268	BP
GO:0006486	protein glycosylation	24	2	0.81	0.0384	BP

附录 K Q_s_W vs. Q_s_E “基因组岛”内基因的 GO 富集分析结果
 Appendix K. GO enrichment results of genes located within the "genetic islands" of Q_s_W vs. Q_s_E

GO.ID	Term	Annotated	Significant	Expected	Fisher.p	Ontology
<i>Qs_W vs. Qs_E</i>						
GO:0005524	ATP binding	364	18	10.95	6.00E-08	MF
GO:0003676	nucleic acid binding	1219	43	36.67	2.00E-07	MF
GO:0005515	protein binding	1849	44	55.61	3.20E-06	MF
GO:0003824	catalytic activity	3977	96	119.62	4.10E-05	MF
GO:0000166	nucleotide binding	754	28	22.68	0.00011	MF
GO:0008237	metallopeptidase activity	38	4	1.14	0.00063	MF
GO:0003677	DNA binding	489	15	14.71	0.00122	MF
GO:0003777	microtubule motor activity	56	4	1.68	0.00276	MF
GO:0003700	DNA-binding transcription factor activity	228	8	6.86	0.00279	MF
GO:0004190	aspartic-type endopeptidase activity	65	4	1.96	0.00473	MF
GO:0003899	DNA-directed 5'-3' RNA polymerase activity	11	2	0.33	0.00558	MF
GO:0003723	RNA binding	213	8	6.41	0.00636	MF
GO:0046873	metal ion transmembrane transporter activity	44	3	1.32	0.01032	MF
GO:0046983	protein dimerization activity	89	4	2.68	0.01406	MF
GO:0004672	protein kinase activity	1000	19	30.08	0.01451	MF
GO:0005506	iron ion binding	308	8	9.26	0.01583	MF
GO:0004252	serine-type endopeptidase activity	94	4	2.83	0.01687	MF
GO:0140657	ATP-dependent activity	78	6	2.35	0.02122	MF
GO:0016831	carboxy-lyase activity	27	2	0.81	0.03195	MF
GO:0008270	zinc ion binding	126	4	3.79	0.04297	MF
GO:0004222	metalloendopeptidase activity	34	2	1.02	0.0487	MF
GO:0005643	nuclear pore	14	3	0.33	0.00028	CC
GO:0016020	membrane	624	15	14.57	0.00056	CC
GO:0005576	extracellular region	24	2	0.56	0.02186	CC
GO:0000145	exocyst	32	2	0.75	0.03746	CC
GO:0008152	metabolic process	1453	47	47.58	3.40E-06	BP
GO:0006754	ATP biosynthetic process	13	3	0.43	0.00071	BP

基于全基因组重测序技术川滇高山栎和刺叶高山栎适应性渐渗研究

GO:0000160	phosphorelay signal transduction system	29	3	0.95	0.00766	BP
GO:0046856	phosphatidylinositol dephosphorylation	11	2	0.36	0.01003	BP
GO:0045454	cell redox homeostasis	14	2	0.46	0.01615	BP
GO:0006355	regulation of transcription, DNA-templated	72	4	2.36	0.01851	BP
GO:0006812	cation transport	80	4	2.62	0.02616	BP
GO:0006364	rRNA processing	19	2	0.62	0.02899	BP
GO:0006952	defense response	63	4	2.06	0.03073	BP
GO:0006886	intracellular protein transport	49	3	1.6	0.0316	BP
GO:0015979	photosynthesis	25	2	0.82	0.04153	BP
GO:0006810	transport	398	14	13.03	0.04597	BP
GO:0006464	cellular protein modification process	107	5	3.5	0.04826	BP

个人简介

秦伟，男，1998年10月生，籍贯安徽省马鞍山市。2017年9月至2021年7月就读于内蒙古农业大学草原与资源环境学院，草业科学专业，并于2018年7月取得农学学术专业。2021年9月至2024年7月就读于北京林业大学生态与自然保护学院，生态学专业，导师为杜芳教授。

硕士期间主修课程：实践生态学、分子生态学理论与方法、恢复生态学专题、保护生物学专题、高级森林生态学、系统与进化生物学等。硕士期间的主要研究内容是基于全基因组重测序数据青藏高原地区两种栎属物种的种群历史动态研究。以共同第一作者发表SCI论文一篇，目前以第一作者投稿CSCD文章一篇。

导师简介

杜芳，女，1981年11月生，甘肃甘南人。2010年6月获得兰州大学生态学博士学位，2010年12月获得法国波尔多第一大学生态学博士学位。现任北京林业大学生态与自然保护学院教授和云南大学生态与环境学院客座教授。

主要致力于森林树种遗传多样性及物种形成的机理研究。主持国家自然科学基金项目三项，参与国家自然科学基金重点项目和科技部重点研发计划各一项。以第一作者或通讯作者发表SCI论文26篇，发明专利1项，论文总计被引用1300余次(谷歌学术统计)。2014年在北京林业大学创建了首个分子生态实验室，为从微观角度解决宏观森林生态学问题搭建了良好的平台。

2009年获法国农业科学学院三大科研进展奖励；2014年获得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、2018年获得梁希青年论文奖；2015年当选北京市科技新星；2019年当选中国植物学会女科学家协会委员；2021年开始担任SCI期刊*Ecology and Evolution*的Associate Editor；2022年联合华东师范大学的王嵘教授与日本和加拿大的学者一起组织了“森林树种保护基因组学(Forest Tree Conservation Genomics)”的专刊；2023年开始担任*Plant Diversity*的Associate Editor。

获得成果目录清单

Luo Y J, Qin W, Yan Y, Yin K Q, Zang R G, Du F. Climate Change Vulnerability and Conservation Strategies for Tertiary Relict Tree Species: Insights from Landscape Genomics of *Taxus cuspidate*[J] (已被 Evolutionary applications 接受, SCI)

秦伟, 杜芳, 尹康权. 景观基因组学在草本植物中的应用与展望[J] 中国草地学报 (投稿中)

致谢

研究生的三年时光即将结束，但是这段回忆却在我的一生里留下了浓墨重彩的一笔。依稀记得是在复试结束的那一天，我正式开始了我的研究生旅途。那一天，我上午怀着忐忑的心结束了我的复试环节。一通来自北京归属地的电话让我紧张的心情得到了舒缓，在没有接起电话的时候，我就感觉那一刻，我对得起自己一年来的辛苦了，对得起在早上六点起床，晚上十点离开教室的自己。我没有辜负冬日走廊里背书的自己。一年的努力在这通电话接起的那一刻得到了最好的回报。

北京林业大学的杜芳老师，给了我一张进入北京林业大学继续学习的通行证。那时候的我，对于这位即将引领着我完成三年科研学习的老师并不熟悉。当天下午，杜芳老师以视频会议的方式给我介绍了实验室的情况，我们相互完成了简短的介绍。虽是一次再简单不过的见面，但是那时候的我也对这个大家庭有了最基本的认识。同时，对于能够快速融入这个大家庭，我充满了期待。由于疫情的原因，我没有能够在暑期来到学校，大概这是我研究生三年生涯最大的遗憾吧。那时候，我每天都希望可以赶紧来到学校，尽快的开展自己的研究课题。杜芳老师为了我的返校工作也提供了很大的帮助，为了安抚我，她安排了实验室的师兄、师姐帮助我提前了解自己的科研方向，帮助我学习相关的理论知识和分析方法。那个暑期，身边上岸的同学都很轻松，他们有时会打趣我，问我怎么这么忙，这是暑假呀！对于他们的玩笑话，我内心毫无波澜，因为我的内心是快乐的，因为我可以更快的开始我的课题，我得到了比他们更多的学习机会。暑假的每一天，我过得都很充实。虽然没有在学校学习，但是我知道此刻的我已经正式踏上了科研的道路。

九月，每年的这个时候，都是离家的日子。那一天，带着父母的期待，我正式坐上了前往北京的列车。我知道，父母是舍不得的，但是为了让自己变得更好，我自私的做出决定。大学是如此，我不顾父母的反对，毅然决然的前往了离家千里的内蒙古。研究生学习亦是如此，我还是没有接受父母的建议，没有选择他们期待的城市继续学业，依然选择了离家很远的北京。在此，我应该向我的父母说一声“对不起”。我一次又一次辜负了你们对我的期盼，但我还是希望你们不要怪我。当然，我也很清楚，你们是不会怪我的，你们只会无条件的包容我，只会无条件的支持我，只会以我为自豪。

三年的学习在 2021 年的秋天正式开始了，在这里，我面对面见到了杜芳老师和实验室的师兄师姐。她们给我的第一印象都很好，很快我就融入了这个大家庭。我和杜芳老师正式见面了，那天我记得您对我说，电脑上的游戏应该卸载了，今后应该把自己的时间多花在科研上。其实，您不知道的是，那些游戏早在我复试结束的时候就卸载了，那些只是弹出的广告。我也没有和您解释，因为我知道，说是没有用的，行动才是证明。那一天，我也感受到了，您对于科研的态度：严谨且专注。在刘可可师兄的帮助下，我开始了自己的课题。杜芳老师口中，可可师兄是最优秀的学生，也是

我们学习的榜样。实际也是如此，他会无条件的帮助我，带着我学习理论知识和分析方法，给我答疑解惑。谢谢你，可可师兄！

很快，一学期的学习结束了，我迎来了我的第一个年终组会。那一次的组会，我受到了老师您对我的表扬。您说我是实验室最勤奋的学生，每天来实验室都很早，中午午休都在学习，晚上也很晚离开实验室。那一刻，真的很开心，因为我的努力被人看到了，收到了您的夸奖。谁不愿意被自己的老师夸奖呢！哈哈，这种被认可的感觉真好！同时，您也指出了我的不足，阅读文献不够细致，与您的沟通不足。对于您的批评，我也很认可。时间继续一点一点的过去，我依旧继续着自己的课题，和您的沟通也越来越频繁了。大多数时候，对于我，您都是很包容的，您会腾出自己的时间，对我的课题提出建议，给我指出其中的不足，给我提供下一步的研究方向。但是，有时候，我们也会爆发一些争吵。那时候我会因为一些意见不合，和您犟嘴，和您耍一些小脾气。那时候的我，心浮气躁，自以为是，过于自我，给您添了不少麻烦。现在想想，挺幼稚的，您都是为我好。您完全可以不问不管，大可以说一些冠冕堂皇的话糊弄我。但是您没有，即使我和您闹脾气，您也没放弃对我的指导。相反，对于我的帮助只增不减。您帮助我邀请到了北京林业大学草业与草原学院尹康权老师和兰州大学生态学院汝大福老师帮助我解决课题上问题。两位老师对我科研上不足提出了宝贵的意见，同时指出了下一步的研究方向。对于两位老师，我感谢你们对我的帮助，感谢你们对我耐心的指导。你们的帮助是无私且真诚的，谢谢你们！对于您，杜芳老师，我为我幼稚的行为向您道歉。三年来，您对我的是帮助，是鼓励，是夸奖，是包容。但是，我回报的您的是幼稚的不理解。谢谢您没有放任我不管，谢谢您让我变得优秀。无论是在我的生活中，还是在我的学习中，您都给予了我很大的帮助。2023年那个除夕夜，我对您说，我未来想去南京发展，但是我不自信。您鼓励我，说我是个很努力的孩子，一定可以的！毕业以后，希望自己不会辜负您对我的期望，更希望自己也能够成为您引以为傲的学生！成为您以后向师弟师妹可以带着笑容谈及的学生！

感谢无条件支持我的父母，感谢包容和给予我帮助的杜芳老师，感谢帮助解决科研困难的尹康权老师和汝大福老师，感谢实验室的同门！三年很短，但是你们对我的帮助只会在我的心中愈发深刻！