

分类号

密级

UDC



北京林业大学

学术型硕士学位论文

不同叶习性栎属物种间的叶绿体基因组比较研究

Comparison of chloroplast genomes of different leaf habit oak species

许阳

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 恢复生态学

二〇二一年 六月 十六日

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： 许阳 日 期： 2021年6月16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： 许阳 导师签名： 杜杰 日 期： 2021年6月16日

论文资助项目

本论文研究工作受国家自然科学基金项目（项目编号：42071060；项目名称：栎类生态适应的分子机理研究）的资助，在此表示感谢！

摘要

壳斗科栎属(*Quercus*)包含约 400-500 个物种, 是北半球重要的生态物种。栎属植物由于表型相似、网状进化和频繁杂交使得其系统发育研究充满挑战。东亚栎属已有的系统发育研究绝大多数仅涉及白栎组、冬青栎组和麻栎组这三个组(对应了中国植物志中的栎属/栎亚属)或仅涉及青冈栎组这一个组(对应了中国植物志中的青冈属/青冈亚属), 但缺乏对中国栎属四个组整体的系统发育关系和适应性进化的研究。不同的生态环境使得栎属内产生了不同的适应性进化, 例如进化出两种不同的叶习性: 常绿和落叶, 为理解自然选择在驱动叶绿体基因进化中的作用以及识别近缘物种适应性进化的关键基因提供了一个很好的模型。

本研究利用高通量测序技术对中国分布的 14 种栎树, 包括七种常绿栎树: 高山栎(*Quercus semecarpifolia* Smith)、灰背栎(*Quercus senescens* Hand.-Mazz)、槿子栎(*Quercus baronii* Skan)、铁橡栎(*Quercus cocciferoides* Hand.-Mazz)、毛脉高山栎(*Quercus rehderiana* Handel-Mazzetti)、矮高山栎(*Quercus monimotricha* Hand.-Mazz)以及巴东栎(*Quercus engleriana* Seem)和七种落叶栎树: 枹栎(*Quercus serrata* Murray)、短柄枹栎(*Quercus serrata* var. *brevipetiolata*)、云南波罗栎(*Quercus yunnanensis* Franchet)、大叶栎(*Quercus griffithii* Hook. f. et Thoms. ex Miq)、黄山栎(*Quercus stewardii* Rehd)、白栎(*Quercus fabri* Hance)以及蒙古栎(*Quercus mongolica* Fischer ex Ledebour)进行了全基因组测序、叶绿体组装拼接及基因注释, 得到了 14 个栎属物种完整的叶绿体基因组。结合实验室已发表三个物种: 槲栎(*Quercus aliena* Blume)、川滇高山栎(*Quercus aquifolioides* Rehd. et Wils)和刺叶高山栎(*Quercus spinosa* David ex Franchet)的叶绿体基因组进行比较分析。分析结果显示 17 种中国本土栎属植物的叶绿体基因组, 长度范围在 160,415 bp 到 161,415 bp 之间, 具有典型的四分体结构。每个叶绿体基因组均包含 86 个蛋白质编码基因, 40 个转运 RNA 基因和四个核糖体 RNA。重复分析显示, 重复序列在 17 种栎树植物叶绿体基因组中的数量和分布基本保守, 但在部分常绿栎中没有检测到反向和互补序列。对比分析发现, 单拷贝区(single copy, SC)比反向重复(inverted repeat, IR)区包含更多核苷酸变异(Pi)值高的位点。同时, 同种叶习性的物种其基因组结构存在着一定的相似性, 但是在

落叶栎和常绿栎之间，基因组结构又有着明显的差异。通过最大似然法（Maximum Likelihood, ML）、邻接法（Neighbor Joining, NJ）及贝叶斯法（Bayesian Inference, BI）对 34 个东亚栎属物种(17 个本实验室拼接物种, 17 个 NCBI 下载物种。其中常绿 18 种, 落叶 16 种)的叶绿体全基因组数据集以及同源序列数据集分别重建系统发育树得到了高度相似的系统发育关系。与之前基于核基因得到的对全球栎属植物的分类系统一致, 34 种东亚栎树被划分为两个亚属(subgenus *Quercus*, subgenus *Cerris*)和四个组(Sect. *Quercus*, Sect. *Ilex*, Sect. *Cerris*, Sect. *Cyclobalanopsis*, Sect. *Quercus*)。手动筛选出 34 个东亚栎树的 72 个共有基因, 利用 PAML 软件中的 Codeml 程序依次对以上共有基因进行正选分析, 为提高结果的准确度我们同时利用了位点模型(site model)和分支位点模型(branch-sitemodel)模型。两个模型同时检测出正选择基因: *ndhF*。其中 branch-site 模型从 15 个常绿种的 *ndhF* 基因中检测出正选择位点, 考虑到大多数的常绿栎生活在海拔较高的石灰岩基质区域(包括栎属中分布海拔最高的高山栎组植物), “适应较高的紫外辐射强度、缺氧、低温以及干旱等环境胁迫”可能是栎属叶绿体水平适应性进化的重要遗传基础。

本研究重测序了 14 种栎属基因组并组装了他们的叶绿体基因组, 为栎属叶绿体基因组结构、系统发育关系以及适应性进化提供了有力支撑。叶绿体正选择研究将促进对栎属叶绿体基因组多样性和复杂性的理解, 同时也为理解自然选择在驱动叶绿体基因进化中的作用以及识别近缘物种适应性进化的关键基因提供了新思路。

关键词: 栎属, 叶绿体基因组, 叶习性, 系统发育, 正选择

Comparison of chloroplast genomes of different leaf habit oak species

Master Candidate: Yang XU

(Ecology)

Supervised by Prof. Dr. Fang DU

Abstract

Quercus contains about 500 species, which are important ecological species in the northern Hemisphere. Phenotypic similarity, reticulated evolution and frequent hybridization make the *Quercus* phylogeny studies full of challenges. Previous oak phylogenetic studies in EastAsia either included section *Quercus*, section *Ilex* and section *Cerris* or only included section *Cyclobalanopsis*, and there was no phylogenetic studies including all the four sections in China. Different environments may have resulted in different adaptive evolution strategies in *Quercus*, such as two different leaf habits: evergreens and deciduous. The different habitat types and adaptation patterns of the *Quercus* groups make them ideal models to understand the role of natural selection in driving the evolution of chloroplast genome and identify key genes for the adaptive evolution of closely related species.

In this study, high-throughput sequencing technology were used to sequence and assemble there chloroplast genome in 14 oak species of China: Seven evergreen oaks: *Quercus semecarpifolia* Smith, *Quercus senescens* Hand.-Mazz, *Quercus baronii* Skan, *Quercus cocciferoides* Hand.-Mazz, *Quercus rehderiana* Handel-Mazzetti, *Quercus monimotricha* Hand.-Mazz, and *Quercus engleriana* Seem; Seven deciduous oaks: *Quercus serrata* Murray, *Quercus serrata* var. *brevipetiolata*, *Quercus yunnanensis* Franchet, *Quercus griffithii* Hook. f. et Thoms. ex Miquel, *Quercus stewardii* Rehd, *Quercus fabri* Hance and *Quercus mongolica* Fischer ex Ledebour. Comparative analyses were conducted combining with other three plastid genomes obtained in our laboratory: *Quercus aquifolioides* Rehd. et Wils, *Quercus spinosa* David ex Franchet and *Quercus aliena* Blume. The results showed that lengths of these chloroplast genomes were ranging from 160,415 bp to 161,415 bp, with a typical tetrad structure with small changes in CG content. Each chloroplast genome contains 86 protein-

coding genes, 40 transfer RNA genes and 4 ribosomal RNA genes. Repeat analysis showed that the number and distribution of various repeats in the chloroplast genomes of 17 oak species were conserved. Comparative analysis revealed that the single copy region (SC) region contained more loci with high nucleotide variation (Pi) values than that of the inverted repeat (IR) region. The phylogenetic tree constructed by the maximum likelihood method (ML), the neighbor-joining method (NJ) and the Bayesian method (BI) based on the whole chloroplast genomes and shared protein-coding genes of the selected 34 species showed almost the same topological structures. Consistent with the previous classification system for *Quercus* plants based on nuclear markers, 34 *Quercus* species in East Asia were divided into two subgenera (subgenus *Quercus*, subgenus *Cerris*) and 4 groups (Sect. *Quercus*, Sect. *Ilex*, Sect. *Cerris*, Sect. *Cyclobalanopsis*). Manually screened out the 72 homologous genes of 34 East Asian oaks (including all of the four groups, with 18 evergreen oaks and 16 deciduous oaks), and used Codeml program in PAML to detect each homologous genes for positive selection analysis. In order to improve the accuracy of the results, we use both the site model and branch-site model. The two models simultaneously detect out the adaptive evolving genes: *ndhF*. Among them, the branch-site model detects positive selection sites from 15 evergreen oaks' *ndhF* genes, taking into account that most of the evergreen oaks live in the limestone matrix areas with higher altitudes, "adaptation to environmental stresses such as high ultraviolet radiation intensity, hypoxia, low temperature, drought, etc." may be important genetic basis for the adaptive evolution of *Quercus* chloroplast.

In this study I sequenced and assembled 14 *Quercus* chloroplast genomes, which further increased the number of chloroplast genomes in the genome database. Moreover, I also studied chloroplast genome structure, phylogenetic relationship and adaptive evolution of *Quercus*, and the results would help to understand the diversity and complexity of *Quercus* chloroplast genome, and provide new insights into understanding the role of natural selection in driving chloroplast gene evolution.

keywords: *Quercus*, chloroplast genome, leaf habit, comparative analysis, phylogeny positive selection

目录

1 引言.....	1
1.1 植物叶绿体基因组.....	1
1.1.1 叶绿体基因组特征.....	1
1.1.2 叶绿体基因组的结构及其特征.....	1
1.2 栎属植物叶绿体基因组研究进展.....	2
1.2.1 栎属基本信息.....	2
1.2.2 栎属叶绿体全基因组研究进展.....	3
1.3 本论文的研究内容、研究目的及意义.....	4
2 材料与方法.....	6
2.1 样本采集.....	6
2.2 基因组 DNA 提取与测序.....	9
2.2.1 总 DNA 提取.....	9
2.2.2 基因组总 DNA 电泳、质量检测.....	10
2.3 叶绿体基因组测序、拼接及注释.....	11
2.3.1 基因组测序及叶绿体基因组拼接.....	11
2.3.2 叶绿体基因组注释.....	11
2.4 叶绿体全基因组基本特征分析.....	11
2.4.1 基本特征分析.....	11
2.4.2 重复序列分析.....	12
2.5 基因组差异比较.....	13
2.6 栎属叶绿体基因组系统发育分析.....	13
2.7 栎属叶绿体基因组选择压分析.....	15
2.7.1 栎属叶绿体基因组共有基因序列筛选.....	15
2.7.2 栎属叶绿体基因组正选择计算.....	15
3 结果与分析.....	17
3.1 栎属叶绿体基因组的基本特征.....	17
3.1.1 栎属叶绿体基因组基本特征分析.....	17
3.1.2 栎属叶绿体全基因组基因功能分析.....	21
3.1.3 栎属叶绿体全基因组重复序列分析.....	23
3.1.4 栎属叶绿体全基因组序列多样性.....	35
3.2 栎属叶绿体基因组系统发育分析.....	41

不同叶习性栎属物种间的叶绿体基因组比较研究

3.3 17 种栎树叶绿体基因组的正选择计算	43
3.3.1 栎属叶绿体基因组共有基因序列筛选	43
3.3.2 栎属叶绿体基因组正选择计算	44
4 讨论	48
4.1 栎属叶绿体基因组序列进化	48
4.2 栎属叶绿体基因组序列差异比较	48
4.3 栎属叶绿体基因组系统发育	49
4.4 正选择分析	50
5 结论与展望	51
5.1 结论	51
5.2 展望	52
参考文献	53
附表	64
个人简介	82
导师简介	84
成果目录清单	86
致谢	88

1 引言

1.1 植物叶绿体基因组

1.1.1 叶绿体基因组特征

叶绿体(Chloroplast)为绿叶植物和藻类等光合真核生物(photosynthetic eukaryotes)的特征细胞器,在有光照的条件下可以将水和二氧化碳经光合作用转换成碳水化合物并释放出氧气(Dyall *et al.*, 2004)。目前主流观点认为叶绿体起源于古老的蓝细菌(cyanobacteria),这类原核生物通过内共生作用(endosymbiosis)在吞噬它们的真核细胞中与宿主细胞协同进化,因此叶绿体具备独立于宿主细胞核的遗传物质及整套遗传信息的存储和表达机制(Martin *et al.*, 1905; Howe *et al.*, 2003; 于慧敏和解相林, 2007)。

上世纪 80 年代,烟草(*Nicotiana tabacum* Linnaeus) (Shinozaki *et al.*, 1986)和地钱(*Marchantia polymorpha* Linnaeus) (Umesono *et al.*, 1984)的叶绿体全基因组序列首次被报导,标志着叶绿体基因组研究的开始。此后,陆续有其他物种的叶绿体全基因组全长序列被公布。21 世纪初,下一代测序/高通量(Next-generation sequencing, NGS/High-throughput Sequencing)技术诞生。测序技术的发展推动了叶绿体全基因组测序工作的进行,公布的叶绿体基因组序列数量呈指数型增长,这也显著地促进了叶绿体基因组学以及其他相关领域的研究(Du *et al.*, 2015)。

1.1.2 叶绿体基因组的结构及其特征

随着测序技术的发展,叶绿体基因组的获取方法不断改进,叶绿体基因组的研究得到了迅速发展(张悦, 2019)。大量研究表明大多数的陆地植物叶绿体 DNA 一般是双链、闭合环状的 DNA 分子(Izawa, 1963),大多数双链环状的叶绿体基因组由四个区域构成:大单拷贝区(Large single copy region, LSC)、小单拷贝区(Small single copy region, SSC)和两个反向重复区(Inverted repeat, IR),这一对 IR(Inverted repeat A, IRA; Inverted repeat B, IRB)区被 LSC 区和 SSC 区分割开来(Marechal & Brisson, 2010)。IR 区的收缩/扩张可以导致整个叶绿体序列长度的变化,一般来说叶绿体基因组的长度在 100-218 kb 之间。叶绿体基因组在结构和基因顺序上具有较高的保守性,大多数植物的叶绿体基因组编码了 110-130 个基因(Kugita *et al.*, 2003; Gao *et al.*, 2010; 陈琴怡, 2017)。根据功能可以将它们分为四类:

- (1) 第一类是与叶绿体基因组转录翻译有关的基因,包括核糖体大亚基、核糖体小亚基、核糖体 RNA 基因、依赖于 DNA 的 RNA 聚合酶基因以及转运 RNA 基

因；

- (2) 第二类是与光合作用有关的基因，分别是 ATP 合酶亚基、NADH 脱氢酶基因、细胞色素亚基、光系统 I 亚基、光系统 II 亚基以及大亚基；
- (3) 第三类是与生物合成相关基因，包括乙酰辅酶 A 羧化酶亚基 *accD*、囊膜蛋白基因 *cemA*、细胞色素 C 合成基因、依赖 ATP 蛋白酶 P 基因亚基、翻译起始因子基因以及成熟酶基因。
- (4) 未知功能基因 *ycf1*、*ycf2*、*ycf3* 以及 *ycf4* 等。
- (5) 叶绿体基因组包含大量的遗传信息，在进化和分子生态研究领域具有独特优势 (Du *et al.*, 2015)。第一，叶绿体基因组的核酸替换率(nucleotide substitution)适中，约为核基因组的约五分之一，是植物线粒体基因组的约五倍左右。不同区域核苷酸替代速率差异显著，受功能保守的限制，蛋白编码区核苷酸替代较慢，非编码序列则表现出较高的核苷酸替代速率(Clegg *et al.*, 1994)；第二，叶绿体基因组结构与基因组成相对保守，能够在很大种群范围内保证同源性(谢海坤等，2017)；第三，叶绿体基因多为单亲遗传，保证了基因在物种之间的直系同源性(Birky *et al.*, 1983)。基于以上优点目前叶绿体基因组信息已经成为高等植物比较基因组学、系统发生、物种鉴定和群体遗传学研究的良好对象和重要数据来源，对其序列解析日益受到人们重视。

1.2 栎属植物叶绿体基因组研究进展

1.2.1 栎属基本信息

栎属(*Quercus* L.)在全球有约 400-500 种左右(Nixon, 1993; Manos *et al.*, 1999; Denk *et al.*, 2018)，在被子植物中表现出惊人的物种丰富度。因其在北半球森林生态系统占据着最高的生物量，栎属被认为是北半球的重点物种(Manos *et al.*, 1999; Kremer *et al.*, 2012; Cavender, 2016)。美洲中部和墨西哥是近代栎属多样化的中心，近期研究表明它们起源于北美栎树分支，并通过快速适应和生态位转换使其多样性迅速提高(Hipp *et al.*, 2018)。尽管栎树物种在生态上意义极其重要，但它们的分类从 19 世纪开始就一直是一个挑战，至今仍在继续(Chasse, 2018)。这一分类学挑战主要归因于栎树物种复杂的进化史，其特征是表型相似、网状进化和频繁杂交(Burger, 1975; Nixon, 1993; Manos *et al.*, 1999; Denk *et al.*, 2010; Simeone *et al.*, 2013; Eaton *et al.*, 2015; McVay *et al.*, 2017; Crowl *et al.*, 2020)。这一挑战衍生出各类基于形态学或分子数据的分类方案。传统研究(Zhou, 1992; Pu *et al.*, 2002; Deng *et al.*, 2014)根据表型特征将中国栎类分为两个属或亚属：栎属/栎亚属(subgenera *Quercus*)和青冈属/青冈亚属(subgenera *Cyclobalanopsis*)，其中栎属/栎亚属根

据其数量性状又被分为 5 大类群, 分别为巴东栎组(Sect. *Engleriana*)、高山栎组(Sect. *Brachylepides*)、槲栎组(Sect. *Quercus*)、槲子栎组(Sect. *Echinolepides*)和麻栎组(Sect. *Aegilops*) (彭焱松, 2007)。在更新的分类中, 栎属被分为栎亚属(subgenera *Quercus*)和麻栎亚属(subgenera *Cerris*), 分别对应于新大陆栎属和旧大陆栎属。其中栎亚属包括五个组: 白栎组(Sect. *Quercus*)、红栎组(Sect. *Lobatae*)、中间栎组(Sect. *Protobalanus*)、活栎组(Sect. *Virentes*)和包含两个特有栎树物种的本都栎组(Sect. *Ponticae*)。麻栎亚属由三个组组成, 其中冬青栎组(Sect. *Ilex*)和麻栎组(Sect. *Cerris*)仅分布在欧亚大陆, 而青冈栎组(Sect. *Cyclobalanopsis*)完全分布在亚洲(Denk *et al.*, 2017; Hipp *et al.*, 2020)。

我国大约有 130 种栎树, 分布西起喜马拉雅东至台湾岛, 从中国东北一直延伸到海南岛的亚热带和温带地区, 被认为是欧亚栎树物种的多样性中心(Zhou *et al.*, 1992; Xu *et al.*, 2013)。化石记录和最近发表的分子研究表明冬青组和青冈组的常绿栎类可能起源于中国西南部(Zhou, 1993; Denk *et al.*, 2018; Jiang *et al.*, 2019)。

中国栎树在生态和进化上很重要, 但其适应性进化模式和系统发育关系需要进一步探索。对一些广泛分布的中国栎属物种进行的系统地理学研究表明过去地质时期的地理分布和气候变化可能影响了这些物种的种内和种间的遗传变异(e.g. Zeng *et al.*, 2011; Chen *et al.*, 2012; Xu *et al.*, 2014; Zeng *et al.*, 2015; Du *et al.*, 2017; Jiang *et al.*, 2019; Ju *et al.*, 2019; Chen *et al.*, 2020)。此外, 对中国栎树物种的研究表明(Yang *et al.*, 2018; Jiang *et al.*, 2019), 对不同环境的适应可能会影响栎树的遗传多样性和物种更替。

1.2.2 栎属叶绿体全基因组研究进展

目前栎属叶绿体基因组在 NCBI 的上传记录已达上千条, 中国是栎属多样性分布中心之一(Denk *et al.*, 2018; Oldfield and Eastwood, 2007), 但与栎属进化相关的研究却并不多, 尤其是在新的属以下分类方式提出后, 使得以往大多数的研究都局限在一个或数个组内, 或者是狭义上的栎亚属内。例如 Liu 等(2020)测序组装并注释了白栎组的短柄栎、辽东栎以及蒙古栎结合另外 11 种壳斗科植物进行比对分析, 通过比较分析检测了 11 个高度变异的区域(*psbA*、*matK/rps16*、*rps16*、*trnS-GCU/trnG-GCC*、*trnR-UCU/atpA*、*trnT-GGU/psbD*、*ndhJ*、*ndhJ/ndhK*、*accD*、*ndhF* 和 *ycf1*), 这些区域可用作分子标记。在白栎组之间的几个比较组中 *rps12*、*rpoC2* 和 *ycf1* 基因的 ω 大于 1, 并且在六个白栎组植物和栓皮栎的(*Quercus variabilis* Bl.)比较中, *petA* 基因受到显著的正选择。系统发育分析显示, 六种白栎组植物与北美红栎(*Quercus rubra* L.)归为一类, 形成一个单一的分支。Yin 等(2018)利用二代测序技术测序、组装并注释了川滇高山栎(*Quercus aquifolioides* hehd.)第一个叶绿体全基因组序列, 结合栎属已报道的其他三个种的叶绿体基因组进行比较, 以研究这四个栎属叶绿体基因组的共同结构模式和序列差异的热点区域, 检验选

择压力是否作用于蛋白质编码基因，并重建了这四个栎属物种的系统发育关系。Yang 等 (2016)对四种栎属植物的完整叶绿体基因组进行了测序，并检索了一个已发表的栎属植物叶绿体基因组进行了比较分析。这五个叶绿体基因组的长度从 161,072 bp 到 161,237 bp 不等，它们的基因数量和顺序以及 GC 含量与其他壳斗科植物相似。同时作者发现了 19 个相对高度可变的区域，以及四个受到正选择的基因(*ndhA*、*ndhK*、*petA* 和 *ycf1*)这些区域可能可以成为栎属系统发育研究的标记。

1.3 本论文的研究内容、研究目的及意义

本研究利用高通量测序技术对我国分布的 14 种栎树：常绿栎树：高山栎(*Quercus semecarpifolia* Smith)、灰背栎(*Quercus senescens* Hand.-Mazz)、檀子栎(*Quercus baronii* Skan)、铁橡栎(*Quercus cocciferoides* Hand.-Mazz)、毛脉高山栎(*Quercus rehderiana* Handel-Mazzetti)、矮高山栎(*Quercus monimotricha* Hand.-Mazz)以及巴东栎(*Quercus engleriana* Seem)，落叶栎树：枹栎(*Quercus serrata* Murray)、短柄枹栎(*Quercus serrata* var. *brevipetiolata*)、云南波罗栎(*Quercus yunnanensis* Franchet)、大叶栎(*Quercus griffithii* Hook. f. et Thoms. ex Miq)、黄山栎(*Quercus stewardii* Rehd)、白栎(*Quercus fabri* Hance)以及蒙古栎(*Quercus mongolica* Fischer ex Ledebour)进行测序、叶绿体基因组拼接及注释，得到了其完整的叶绿体基因组，其中白栎的叶绿体基因组是首次被测序并公布。我们结合三种已发表的栎树叶绿体基因组：槲栎、川滇高山栎以及刺叶高山栎，对所有 17 个物种的叶绿体基因组的基本特征进行解析，并比较了两种习性栎树叶绿体基因组结构的差异。结合 NCBI 上下载的另外 17 个种的栎树叶绿体基因组，通过 ML 法、NJ 法及 BI 法三种方法对所有 34 种栎树（涵盖了东亚地区分布的四个组）的叶绿体基因组同源序列构建了东亚栎属 4 个组的系统发育树。最后，利用 PAML 软件中 Cdeml 程序对 34 个物种（其中常绿栎树 18 种，落叶栎树 16 种）的共有蛋白质编码基因进行正选择计算，为提升结果的可靠性我们同时使用了位点模型(site model)和分支位点模型(branch-site model)模型。

本研究旨在：

- (1) 组装获得 14 种栎树的叶绿体全基因组序列并对其叶绿体基因组基因信息进行组装和注释；
 - (2) 通过重复序列分析及叶绿体基因组结构差异比较，探讨栎属植物叶绿体基因组结构多样化机制与叶习性之间的潜在联系；
 - (3) 利用叶绿体基因组构建东亚栎属植物系统发育树并探讨栎属物种的进化关系；
 - (4) 通过正选择分析探究选择压力以及适应性进化在不同叶习性栎类上的作用痕迹；
- 本研究组装并注释了 14 种栎树的叶绿体基因组，有利于今后开展对栎属属内、组

间或是种间的叶绿体基因组相关研究，同时本研究也为栎属叶绿体基因组结构及其系统发育关系提供了有力支撑，并且研究结果将促进对栎属叶绿体基因组多样性和复杂性的理解。

2 材料与方法

2.1 样本采集

我们通过对中国数字标本馆以及已经发表的文献对 14 种中国广泛分布的栎亚属物种，即常绿栎树：高山栎(*Quercus semecarpifolia*)、灰背栎(*Quercus senescens*)、槿子栎(*Quercus baronii*)、铁橡栎(*Quercus cocciferoides*)、毛脉高山栎(*Quercus rehderiana*)、矮高山栎(*Quercus monimotricha*)以及巴东栎(*Quercus engleriana*)，落叶栎树：枹栎(*Quercus serrata*)、短柄枹栎(*Quercus serrata* var. *brevipetiolata*)、云南波罗栎(*Quercus yunnanensis*)、大叶栎(*Quercus griffithii*)、黄山栎(*Quercus stewardii*)、白栎(*Quercus fabri*)以及蒙古栎(*Quercus mongolica*)的分布区进行整理，实地考察并采集实验所需的样品。采集的新鲜叶片封装在有变色硅胶干燥剂的信封袋中带回实验室，用于后续 DNA 的提取。详细的采样点信息见（图 2.1，表 2.1）。每个采样点采集一份模式标本用标本夹压制用于物种鉴定，植物材料以及 DNA 样品均储存于本实验室。

表 2.1 14 种栎树采样信息

Table 2.1 Sample information of fourteen *Quercus* species

物种	叶习性	位置	经度	纬度	海拔(m)
白栎	落叶	贵州省都匀市三江堰	106°30'36"E	26°21'0"N	870
枹栎	落叶	陕西平利县药妇山	108°52'12"E	32°7'48"N	890
短柄枹栎	落叶	江苏省连云港市连云港区凰窝	119°28'12"E	34°42'36"N	85
大叶栎	落叶	贵州省遵义市凤凰山公园	106°55'12"E	27°42'00"N	1400
蒙古栎	落叶	辽宁省沈阳市东陵景区	123°35'24"E	41°49'48"N	45
黄山栎	落叶	安徽省黄山市黄山区黄山	118°13'12"E	30°6'36"N	403.3
云南波罗栎	落叶	贵州省凯里市小高山	107°58'48"E	26°34'12"N	1240
矮高山栎	常绿	四川省甘孜州九龙县顺山	101°30'22"E	29°0'25"N	1740
巴东栎	常绿	湖南省张家界天门山	110°29'2"E	29°3'4"N	1410
高山栎	常绿	四川省攀枝花市盐边县格萨拉生态旅游区	101°15'42"E	27°10'5"N	2930
灰背栎	常绿	四川省阿坝州金川县撒瓦脚乡	101°55'28"E	31°34'16"N	1520
毛脉高山栎	常绿	四川省甘孜州泸定县二郎庙	102°14'1"E	29°54'17"N	2780
铁橡栎	常绿	四川省攀枝花苏铁自然保护区	101°35'43"E	26°37'18"N	2730
檀子栎	常绿	甘肃省天水市东岔乡九江峪	106°32'11"E	34°20'53"N	760

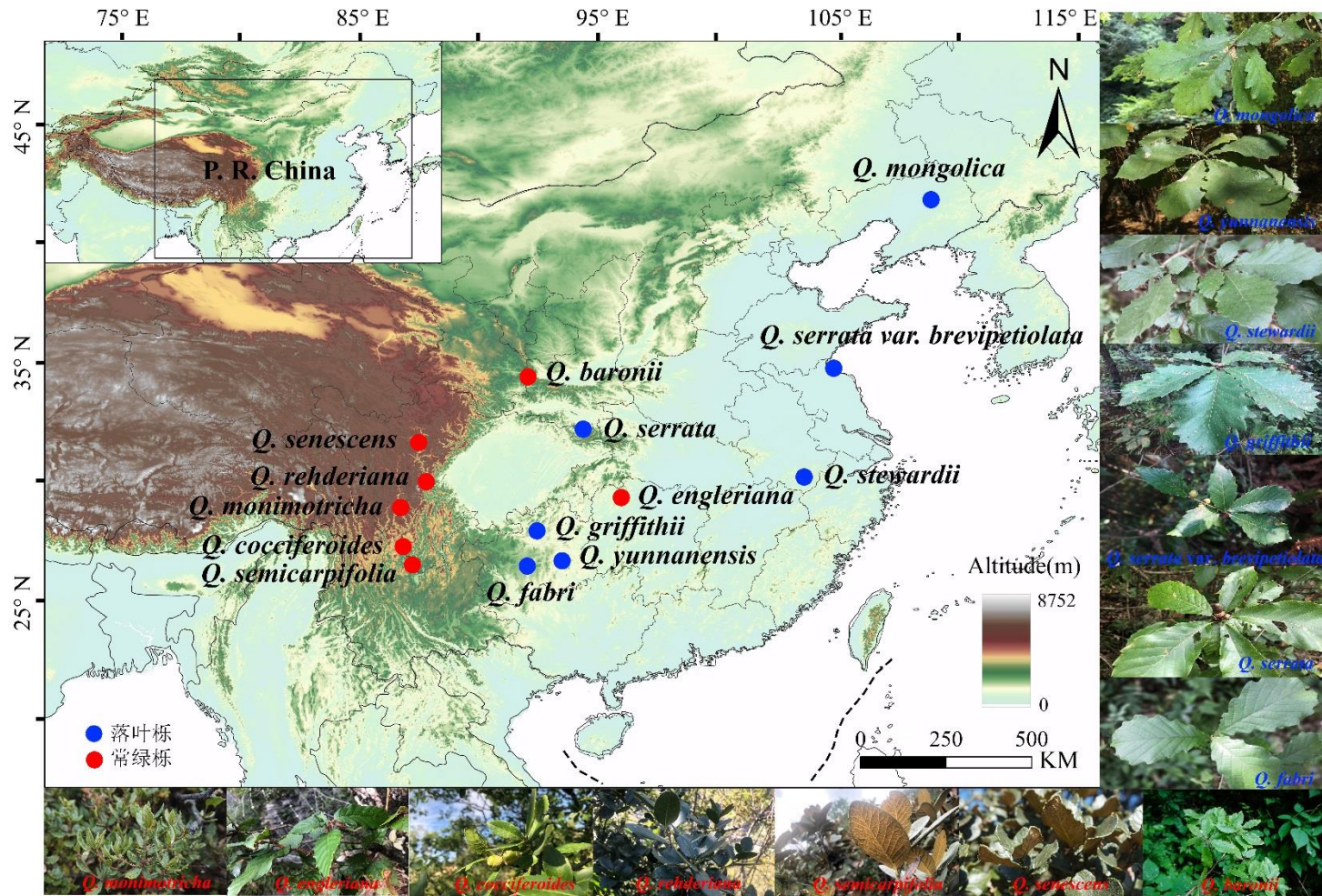


图 2.1 14 种栎树采样信息

Figure 2.1 Sampling sites of fourteen *Quercus* species

2.2 基因组 DNA 提取与测序

2.2.1 总 DNA 提取

使用改良的 CTAB 法提取 14 种栎树基因组 DNA，具体步骤如下：

本研究中对落叶栎使用 TIANGEN 植物基因组 DNA 提取试剂盒进行 DNA 提取，而叶片较硬单宁含量较高的常绿栎使用改良的 CTAB 法进行 DNA 提取 (Doyle, 1987)。两种方法的具体操作步骤如下：

TIANGEN 植物基因组 DNA 提取试剂盒操作步骤：

- (1) 在实验之前先打开水浴锅，将 GP1 放置在 65℃ 的恒温水浴锅预热 30min；
- (2) 接着是实验材料准备阶段，使用电子天平称 0.015±0.005g 的样品，将称取完毕的样品和直径 0.5cm 的小钢珠共同放置在 2ml 的离心管中并做好标记；
- (3) 所有的离心管在研磨之前放入液氮中冰冻 15min，然后将离心管放入研磨仪中，以 50r/s 研磨 5-10min；
- (4) 在通风橱中依次加入 700μL GP1 和 0.7μL 的 β-巯基乙醇；
- (5) 将离心管中的液体摇匀混合，之后在 65℃ 水浴锅中水浴 20min，其间每隔 5min 摇晃一次离心管；
- (6) 在通风橱中加 700μL 的三氯甲烷至离心管并摇晃 10min 以混合；
- (7) 离心管在高速低温离心机中以 12000 rpm，4℃ 的条件下离心 5 min；
- (8) 取一个新的 2ml 离心管，吸取之前离心管中的上清液置于其中，再加入 0.7μL 的 RNA 酶并做好标记，于室温下静置 10min；
- (9) 向静置之后的离心管中加入 700μL 三氯甲烷，并摇晃 10min 以混匀；
- (10) 离心管在高速低温离心机中以 12000 rpm，4℃ 的条件下再次离心 5 min；
- (11) 取一个新的 2ml 离心管，吸取之前离心管中的上清液置于其中，加入 700μL 的 GP2，摇晃 5min 以混匀；
- (12) 将混匀的液体转入吸附柱，吸附柱在高速低温离心机中以 12000 rpm，4℃ 的条件下离心 30s；
- (13) 倒掉吸附柱中的废液，并在吸附柱中添加 500μL GD（缓冲液）；
- (14) 吸附柱在高速低温离心机中 12000 rpm，4℃ 的条件下离心 30s；
- (15) 倒掉吸附柱中的废液，并在吸附柱中添加 600μL PW（漂洗液）
- (16) 吸附柱在高速低温离心机中 12000 rpm，4℃ 的条件下离心 30s；
- (17) 倒掉吸附柱中的废液，再次在吸附柱中添加 600μL PW（漂洗液）；
- (18) 吸附柱在高速低温离心机中 12000 rpm 4℃ 的条件下离心 2min，并倒掉废液；

- (19)将吸附柱在室温条件下放置 10min，之后加入 80 μ L 的 TE（缓冲液），在烘箱中 37 $^{\circ}$ C 的条件下烘干 10min；
- (20)烘干后的吸附柱在高速低温离心机中 12000 rpm, 4 $^{\circ}$ C 的条件下离心 2min, 离心出的 DNA 产物保存在-20 $^{\circ}$ C 冰箱中。

改良的 CTAB 法操作步骤：

- (1) 在实验之前先打开水浴锅,将 GP1 放置在 65 $^{\circ}$ C 的恒温水浴锅预热 30min;
- (2) 使用电子天平称 0.015 \pm 0.005g 的样品, 将称取完毕的样品和直径 0.5cm 的小钢珠共同放置在 2ml 的离心管中并做好标记;
- (3) 所有的离心管在研磨之前放入液氮中冰冻 15min, 然后将离心管放入研磨仪中, 以 50r/s 研磨 5-10min;
- (4) 在通风橱中依次加入 700 μ L GP1, 0.7 μ L 的 β -巯基乙醇和适量的 pvpp;
- (5) 将离心管中的液体摇匀混合, 之后在 65 $^{\circ}$ C 水浴锅中水浴 35min, 其间每隔 5min 摇晃一次离心管;
- (6) 在通风橱中加 700 μ L 的三氯甲烷至离心管并摇晃 10min 以混合;
- (7) 离心管在高速低温离心机中以 12000 rpm, 4 $^{\circ}$ C 的条件下离心 5 min, 取一个新的 2ml 离心管, 吸取之前离心管中的上清液置于其中, 再加入 0.7 μ L 的 RNA 酶并做好标记, 于室温下静置 10min;
- (8) 在通风橱中加 700 μ L 的三氯甲烷至离心管并摇晃 10min 以混合, 离心管在高速低温离心机中以 12000 rpm, 4 $^{\circ}$ C 的条件下离心 5 min;
- (9) 取一个新的 2ml 离心管, 吸取之前离心管中的上清液置于其中, 加入与上清液等体积的异丙醇摇晃混匀, 放置在-20 $^{\circ}$ C 环境下 30min;
- (10)离心管在高速低温离心机中以 12000 rpm, 4 $^{\circ}$ C 的条件下离心 5 min 后倒掉上部液体, 保留底部沉淀;
- (11)加入 1ml 的 70%乙醇, 离心管在高速低温离心机中以 12000 rpm, 4 $^{\circ}$ C 的条件下离心 5 min 后倒掉废液;
- (12)离心管静置于通风橱中 2mins 左右, 待风干后加入 80 μ l TE（缓冲液）, 保存在-20 $^{\circ}$ C 冰箱中。

2.2.2 基因组总 DNA 电泳、质量检测

用 1% 的琼脂糖凝胶对提取的 14 种栎树的 DNA 样本进行初步检测。利用凝胶成像仪(TJY04S-3C)通过观察电泳条带与 marker 的位置来检测 DNA 断片大小以及是否存在明显降解等。接着使用超微量紫外分光光度计(NanoDrop 8000)来检测所提取 DNA 的浓度和纯度。使用吸光度 OD 值在 280nm/260 nm 和 230nm/260 nm 的比值来检测所提取 DNA 的浓度以及是否存在 RNA 或蛋白质污染等（张悦, 2019）。

2.3 叶绿体基因组测序、拼接及注释

2.3.1 基因组测序及叶绿体基因组拼接

将合格的栎树 DNA 样品送往北京百迈客生物科技有限公司，每个样品通过 Illumina-HiSeq2000 (Illumina, USA)平台进行双端测序，序列读长为 150 bp。

公司返回的 14 种栎树的测序数据上传至服务器，通过 CentOS 系统运行 NOVOPlasty2.4.pl (Dierckxsens *et al.*, 2017)对 14 种栎树分别进行 De novo genome 拼接，设置 k-mer 长度为 49 bp，为保证更精确的组装结果，使用已公开发表在 NCBI 上的三个近缘种序列：川滇高山栎(*Quercus aquifolioides*, KP340971)，刺叶高山栎(*Quercus spinosa*, KM841421)及槲栎(*Quercus aliena*, KP301144)作为参考基因组。拼接得到的初步结果使用 samtools (Li *et al.*, 2009)软件转化为 bam 文件后在 Geneious (Kearse *et al.*, 2012)软件对拼接结果进行手工校对并获得 14 个种完整的叶绿体基因组序列。

2.3.2 叶绿体基因组注释

14 个种的完整叶绿体基因组序列上传至在线注释网站 CPGAVAS2 (<http://47.96.249.172:16019/analyzer/home>)进行注释，参考基因组选择川滇高山栎、刺叶高山栎及槲栎的叶绿体全基因组。在线注释后的 sqn 文件在 Sequin 软件中结合 tRNAscan-SE (Lowe & Eddy, 1997) (<http://lowelab.ucsc.edu/tRNAscan-SE/>)验证的 tRNA 基因结果对在线注释的结果进行人工校正。最终的 14 个种的序列信息和注释结果由 sequin 软件编辑整合完成后提交至 NCBI(<https://www.ncbi.nlm.nih.gov/>)，Genbank number 分别为：枹栎 MK922350，短柄枹栎 MK922351，白栎 MK922346，大叶栎 MK922347，蒙古栎 MK922349，黄山栎 MK922352，檀子栎 MW829652，铁橡栎 MW89652，巴东栎 MW829653，矮高山栎 MW829654，毛脉高山栎 MW829655，高山栎 MW829657，灰背栎 MW829657，以及云南波罗栎 MW876863。同时结合本实验室已发表序列槲栎(*Quercus aliena*, KP301144)、川滇高山栎(*Quercus aquifolioides*, KP340971)以及刺叶高山栎(*Quercus spinosa*, KM841421)用于后续的分析。

2.4 叶绿体全基因组基本特征分析

2.4.1 基本特征分析

17 个栎树物种的序列通过检查和注释后导入 Genious R9 软件中统计其基本特征，包括：叶绿体全基因组序列长度，GC 碱基含量；大单拷贝区(Large single copy, LSC)，小单拷贝区(Samll single copy, SSC)，反向重复区(Inverted repeats, IRs)的长度、占总长百分比以及 GC 含量。

统计各类编码基因的数量和顺序,统计蛋白质编码基因(Protein coding genes, PCGs)、转运 RNA (Transfer RNA, tRNA)与核糖体 RNA (Ribosomal RNA, rRNA)的个数和长度。

2.4.2 重复序列分析

重复序列是基因组中不同位置出现的相同或互补性片段,能够为动植物的家系分析、遗传多样性分析等提供一定帮助(张悦, 2019)。重复序列的数量与分布与叶绿体基因组中的基因组重排程度密切相关,尤其是大重复序列(>20bp和>60bp)的比例和数目与基因组重排的程度显著相关(Pombert *et al.*, 2006; Cai *et al.*, 2008; Guisinger *et al.*, 2011; Weng *et al.*, 2013)。

本研究主要对 17 种栎树叶绿体基因组中的:简单重复序列(simple sequence repeat)、回文重复(Palindromic match)、正向重复(Forward match)、反向重复(Reverse match)和互补重复(Complement match),等五类重复序列的数目、长度以及分布情况进行了识别。

2.4.2.1 简单重复序列分析

简单序列重复(simple sequence repeat, SSR),是由长度介于 1-6bp 的重复单元串联组成的重复序列(张悦, 2019; Du *et al.*, 2013)。简单重复序列在复制过程中易发生长度变化,因此通常具有较高的多态性,可作为分子标记并运用在物种遗传分析中(陈琴怡, 2017; 王崇等, 2020),探究 17 种栎树的简单序列重复信息能够为栎属分子标记开发及种群遗传分析工作提供一定的基础。

在 CentOS 系统服务器中运行 MISA.pl (Thiel *et al.*, 2003)程序,其参数设置为,单核苷酸重复单元(Mononucleotide),二核苷酸重复单元(Dinucleotide),三核苷酸重复单元(Trinucleotide),四核苷酸重复单元(Tetranucleotide),五核苷酸重复单元(Pentanucleotide),六核苷酸重复单元(Hexanucleotide)的最小重复次数分别设置为 10, 5, 4, 3, 3, 3。

2.4.2.2 串联重复序列分析

串联重复(Tandem repeats)的形式和简单序列重复相似,但是重复的基本单元稍大(Benson, 1999; 张悦, 2019)。串联重复序列代表了丰富的遗传连锁、定位和个体识别的高度多态性标记,已经成为分子进化研究的有力工具(Brinkmann *et al.*, 1998; Meyer *et al.*, 1995; 高焕和孔杰, 2005)。

17 种栎树的叶绿体通过网页工具 Tandem Repeats Finder(Benson, 1999) (<http://tandem.bu.edu/trf/trf.html>)对串联重复序列进行搜索,设置对比参数 Match、Mismatch 和 Indel 的值分别设置为 2, 7, 7, 重复序列相似度大于 80%, 最大比对长度大于 500 bp。

2.4.2.3 正向、反向、回文以及互补重复序列分析

为了检测出 17 种栎树叶绿体全基因组序列中四种可能的序列重复，即：正向重复(forward/direct match)，反向重复(reverse match)，回文重复(palindromic match) 以及 互 补 重 复 (complement match)。我们利用在线软件 REPuter(<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) (Kurtz et al., 2001)对 17 种栎树叶绿体全基因组进行搜索，参数设置：重复序列相似度大于 80%，Hammung 距离等于 3，最小重复长度大于 30 bp。

2.5 基因组差异比较

为了检测 17 种栎树的序列差异，我们首先在 Ubuntu 系统中，加载 CGView Comparison Tool(CCT)(Grant et al., 2012)导入 17 个种的叶绿体全基因组序列，以榲子栎 *Q. baronii* 的序列作为参考，使用 BLAST 进行比较，BLAST 的结果以图形的形式呈现。

利用软件 Cygwin 中的多重序列比对工具 MAFFT(Standley, 2013)对 17 种栎树的叶绿体基因组序列进行比对，比对分析采用在线比对工具 mVISTA (<http://genome.lbl.gov/vista/index.shtml>)按照 Shuffle-LAGAN 算法(Frazer et al., 2004)，以榲子栎 *Q. bronii* 的序列作为参考，对 17 种栎树的叶绿体基因组的序列进行可视化比较。

我们将每一种栎树的叶绿体全基因组划分为三个数据集(大单拷贝区 LSC，小单拷贝区 SSC，反向重复区 IR)将 17 个物种的三个区域分别进行比对，之后使用 DnaSP 5(Librado et al., 2009)进行滑动窗口分析(Sliding window analysis)以得到每个数据集的核苷酸多样性(Π)值，其中步长设置为 200 bp，窗口长度设置为 600 bp。

最后，我们使用在线分析工具 IRscope (<https://irscope.shinyapps.io/irapp/>)对 17 个物种的各分区长度以及边界信息进行可视化。

2.6 栎属叶绿体基因组系统发育分析

本研究测序拼接并注释了 14 个栎属物种的叶绿体全基因组序列，结合 NCBI 中已经发表的 20 个东亚栎属物种的叶绿体基因序列，共 34 个物种(见表 2.2)构建系统发育树。我们采用两种方式构建数据集：一、直接使用 34 个栎属物种的叶绿体全基因组序列在 Cygwin 软件中的 Mafft 包中对序列进行比对。二、研究利用软件 HomBlocks(Bi et al., 2018)筛选出 34 个栎属物种的同源基因。为了得到更准确的栎属系统发育关系，我们对两组数据集分别通过三种方法：最大似然法

(Maximum Likelihood, ML)、邻接法(Neighbor Joining, NJ)和贝叶斯法(Bayesian Inference, BI)重建系统发育树,。

ML 法和 NJ 法建树的步骤类似, 直接将比对好的序列导入 MEGA X(Sudhir *et al.*, 2018)中, 勾选“Models”中的“Find Best DNA/protein Models (ML)”确定最优碱基替换模型, 之后利用“Phylogeny”中的“Construct/Test Maximum Likelihood Tree”/“Construct/Test Neighbor Joining Tree”进行 1000 次自展支持重复以分别进行 ML 法和 NJ 法的建树。

贝叶斯法建树同样是在 MEGA X 中先确定最优碱基替换模型, 之后将比对好的序列从 MEGA X 中导出为“.nex”格式, 此处不勾选“InterleavedOutput”将导出的 nex 导入到 MrBayes 3.2 (Huelsenbeck & Ronquist, 2001), 使用马尔可夫链-蒙特卡洛方法(Markov Chain Monte Carlo, MCMC)算法, 运行次数为 1, 000, 000 代, 其它参数默认(张悦, 2019)。

表 2.2 栎属叶绿体基因组系统发育分析中用到的物种信息
Table 2.2 Species information used in phylogenetic analysis of *Quercus* chloroplast genome

序号	物种	拉丁名	分组	叶习性	Genebank 号
1	刺叶高山栎	<i>Quercus spinosa</i>	Sect. <i>Ilex</i>	常绿 evergreen	KM841421
2	川滇高山栎	<i>Quercus aquifolioides</i>	Sect. <i>Ilex</i>	常绿 evergreen	KP340971
3	檀子栎	<i>Quercus baronii</i>	Sect. <i>Ilex</i>	常绿 evergreen	MW829651
4	铁橡栎	<i>Quercus cocciferoides</i>	Sect. <i>Ilex</i>	常绿 evergreen	MW829652
5	巴东栎	<i>Quercus engleriana</i>	Sect. <i>Ilex</i>	常绿 evergreen	MW829653
6	矮高山栎	<i>Quercus monimotricha</i>	Sect. <i>Ilex</i>	常绿 evergreen	MW829654
7	毛脉高山栎	<i>Quercus rehderiana</i>	Sect. <i>Ilex</i>	常绿 evergreen	MW829655
8	高山栎	<i>Quercus semicarpifolia</i>	Sect. <i>Ilex</i>	常绿 evergreen	MW829656
9	灰背栎	<i>Quercus senescens</i>	Sect. <i>Ilex</i>	常绿 evergreen	MW829657
10	坝王栎	<i>Quercus bawanglingensis</i>	Sect. <i>Ilex</i>	常绿 evergreen	MK449426.1
11	匙叶栎	<i>Quercus dolicholepis</i>	Sect. <i>Ilex</i>	常绿 evergreen	KU240010.1
12	黄背栎	<i>Quercus pannosa</i>	Sect. <i>Ilex</i>	常绿 evergreen	NC_050963.1
13	乌冈栎	<i>Quercus phillyraeoides</i>	Sect. <i>Ilex</i>	常绿 evergreen	NC_048488.1
14	太鲁阁栎	<i>Quercus tarokoensis</i>	Sect. <i>Ilex</i>	常绿 evergreen	NC_036370.1
15	通麦栎	<i>Quercus tungmaiensis</i>	Sect. <i>Ilex</i>	常绿 evergreen	NC_036936.1
16	赤栎	<i>Quercus acuta</i>	Sect. <i>Cyclobalanopsis</i>	常绿 evergreen	MT742291
17	青冈	<i>Quercus glauca</i>	Sect. <i>Cyclobalanopsis</i>	常绿 evergreen	NC_036930
18	赤皮青冈	<i>Quercus gilva</i>	Sect. <i>Cyclobalanopsis</i>	常绿 evergreen	MK986651
19	槲栎	<i>Quercus aliena</i>	Sect. <i>Quercus</i>	落叶 deciduous	KP301144
20	白栎	<i>Quercus fabri</i>	Sect. <i>Quercus</i>	落叶 deciduous	MK922346.1
21	枹栎	<i>Quercus serrata</i>	Sect. <i>Quercus</i>	落叶 deciduous	MK922350
22	短柄枹栎	<i>Quercus serrata</i> var. <i>brevipetiolata</i>	Sect. <i>Quercus</i>	落叶 deciduous	MK922351
23	大叶栎	<i>Quercus griffithii</i>	Sect. <i>Quercus</i>	落叶 deciduous	MK922347
24	蒙古栎	<i>Quercus mongolica</i>	Sect. <i>Quercus</i>	落叶 deciduous	MK922349

25	黄山栎	<i>Quercus stewardii</i>	Sect. <i>Quercus</i>	落叶 deciduous	MK922352
26	云南波罗栎	<i>Quercus yunnanensis</i>	Sect. <i>Quercus</i>	落叶 deciduous	MW876863
27	锐齿槲栎	<i>Quercus aliena</i> var. <i>acutiserrata</i>	Sect. <i>Quercus</i>	落叶 deciduous	MK105452.1
28	槲树	<i>Quercus dentata</i>	Sect. <i>Quercus</i>	落叶 deciduous	MK105453.1
29	凤城栎	<i>Quercus</i> × <i>fenchengensis</i>	Sect. <i>Quercus</i>	落叶 deciduous	NC_048513.1
30	日本蒙古栎	<i>Quercus mongolica</i> subsp. <i>Crispula</i>	Sect. <i>Quercus</i>	落叶 deciduous	NC_049877.1
31	辽东栎	<i>Quercus wutaishanica</i>	Sect. <i>Quercus</i>	落叶 deciduous	NC_043857.1
32	麻栎	<i>Quercus acutissima</i>	Sect. <i>Cerris</i>	落叶 deciduous	NC_039429.1
33	小叶栎	<i>Quercus chenii</i>	Sect. <i>Cerris</i>	落叶 deciduous	NC_039428.1
34	栓皮栎	<i>Quercus variabilis</i>	Sect. <i>Cerris</i>	落叶 deciduous	MK105451.1

2.7 栎属叶绿体基因组选择压分析

本研究测序拼接并注释了 14 个栎属物种的叶绿体全基因组序列, 结合 NCBI 中已经发表的 20 个 (其中 3 个为本实验室发表) 东亚栎属物种的叶绿体基因序列, 共 34 个物种, 针对其中两种主要的叶习性(常绿栎以及落叶栎), 利用 PAML 软件中 *codeml* 程序进行正选择分析, 旨在探究选择压力以及适应性进化在两种叶习性栎类上作用的痕迹。

2.7.1 栎属叶绿体基因组共有基因序列筛选

将进行系统发育分析的共计 34 个栎属物种的蛋白质编码基因进行统计, 以确定 34 个物种的不重复蛋白质编码基因。对每一个基因进行检查: 确定其序列长度为 3 的倍数, 对在不同物种内长度有差异的基因进行人工校正以保证每个基因在每个物种内的长度尽可能接近。检查完每一个蛋白质编码基因的序列后, 去除其终止密码子并在 MEGA7 软件中利用 ClustalW(Thompson *et al.*, 2003)按照 codon 进行比对, 最后生成的蛋白质编码基因的比对完成数据集导入 CentOS 服务器中。

2.7.2 栎属叶绿体基因组正选择计算

我们使用 PAML 软件包中的 *Codeml* 程序(Yang, 2007)对 34 个栎属物种的叶绿体共有蛋白质编码基因进行了进化分析。我们首先用 M0 模型(model = 0, Nsites = 0, 假设没有位点或分支 dN/dS 变化)对这些基因的同义替换(synonymous substitutions, dS)和非同义替换(non-synonymous substitutions, dN)进行了初步估计, 以检测是否有基因的 dS 过高导致对于 dN/dS 的错误估计。

在此基础之上, 我们以这些叶绿体基因构建了 34 个栎属物种的系统发育树, 并基于位点模型(site-model)和分支位点模型(branch-site model)来探索了这些蛋白编码基因是否在不同栎属类群存在正选择。

位点模型(site-model)假设数据集中不同位点受的选择压力不同, 而不考虑不

同支系间受的选择压力差异。我们分别用 M1 vs. M2, M7 vs. M8 两对位点模型来计算了位点间的正选择;其中 M1 和 M7 为各自模型中的零模型(the null models), 而 M2 和 M8 为替代模型(the alternative models)。替代模型允许特定位点的正选择, 而零模型不允许(Yang *et al.*, 2007)。我们通过对不同模型得出的最大似然值进行似然比检验(the likelihood ratio test, LRT), 与具有相应自由度的零和 χ^2 的混合分布进行比较, 从而判断打破零模型的概率有多大。我们对于计算出的 P 值在 R 中(Team, 2015)进行了 FDR(False Discovery Rate)校正, 并以 Q 值小于 0.05 作为正选择的阈值(Storey *et al.*, 2003)。

为了检测特定位点在特定谱系上发生正选择的证据, 我们进一步使用 Codeml 的分支位点模型进行了分析。其中零模型参数为(model A1)(model=2, Nsites =2, fixed $\omega=1$, $\omega=1$), 替代模型(model A)参数为(model=2, Nsites=2, fixed $\omega_0=0$, $\omega_1=1.5$)。替代模型(the alternative model)假设了四类位点:前两类位点被假设在整个系统发育过程中处于纯化选择($0 \leq \omega_0 \leq 1$)或中性中性选择($\omega_1 = 1$), 而其余两类位点被假设当处于背景支时表示被纯化选择或中性选择($\omega_2 > 1$), 但当它们处于前景支时则表示受到正向选择($\omega_2 > 1$)。零模型将前景支 ω_2 固定为 1。因为分支位点模型需要在数据集中标记前景支, 因此在本研究中我们将 34 个栎属物种按照叶习性分为两类, 第一类是主要包括白栎组 Sect. *Quercus* 以及麻栎组 Sect. *Cerris* 的 16 落叶栎, 第二类是主要包括冬青栎组 Sect. *Ilex* 以及青冈栎组 Sect. *Cyclobalanopsis* 的 18 常绿栎。我们将两类栎分别设置为前景支, 计算了不同模式下的正选择以及由经验贝叶斯法(Bayes empirical Bayes, BEB)识别出后验概率 ≥ 0.95 的正选择位点(Zhang *et al.*, 2005)。其中 LRT 检验和 P/Q 值计算同上所述。

3 结果与分析

3.1 栎属叶绿体基因组的基本特征

3.1.1 栎属叶绿体基因组基本特征分析

14 种栎树的序列组装完成后在 CPGAVAS 2 中进行初步注释（图 3.1），之后在 Sequinv13.05 中人工校正在线注释可能的错误（遗漏 *trnM-CAU*，*rps12* 注释错误等）在 Sequinv13.05 中得到人为校正，同时也对实验室已发表的三种栎（槲栎(*Quercus aliena*, KP301144)，川滇高山栎(*Quercus aquifolioides*, KP340971)以及刺叶高山栎(*Quercus spinosa*, KM841421)进行了重新注释，以避免由人为导致的序列注释差异。全部的 17 种栎树叶绿体基因组的长度范围为 160,415 bp 到 161,415 bp 之间，它们呈典型的四分体结构，相同区域拥有相似的长度(图 3.2 和表 3.1)。17 种栎树基因组的 CG 含量整体变化不大，在 36.8% - 37% 之间（极差 0.3%）。每个物种都包含 86 个蛋白质编码基因，40 个 tRNA 基因和四个 rRNA 基因。在这些基因中 *rps7*, *rps12*, *rpl2*, *rpl23*, *rrn16*, *rrn23*, *rrn4.5*, *rrn5*, *trnV-GAC*, *trnR-ACG*, *trnN-GUU*, *trnN-GUU*, *trnA-UGC*, *trnI-GAU*, *trnL-CAA*, *trnI-CAU*, *ndhB*, *ycf1* 以及 *ycf2* 有两个拷贝；*rpl2*, *atpF*, *ndhA*, *ndhB* 以及 *ndhH* 包含一个内含子；*rps12*, *clpP* 以及 *ycf3* 包含两个内含子；*rps12* 分为两个独立转录单元。17 个物种所包含的基因数量和顺序是一致的，且与其它已发表的壳斗科叶绿体基因组相似 (Alexander *et al.*, 2014; Yang *et al.*, 2017; Liu *et al.*, 2019; Pang *et al.*, 2019; Hu *et al.*, 2019)。

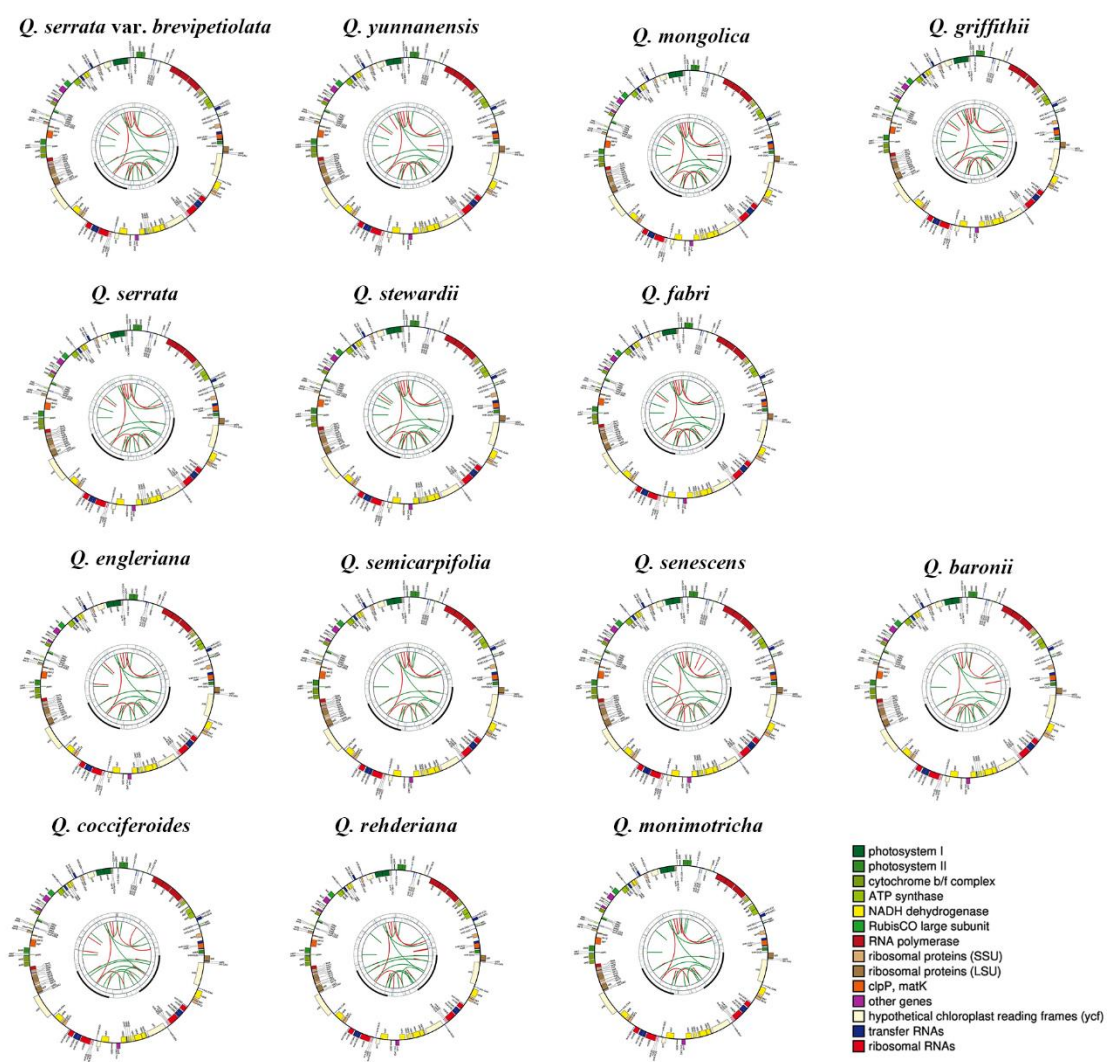


图 3.1 14 个栎树物种在 CPGAVAS 2 中的初步注释结果

Figure 3.1 preliminary annotation results of 14 *Quercus* species in CPGAVAS 2

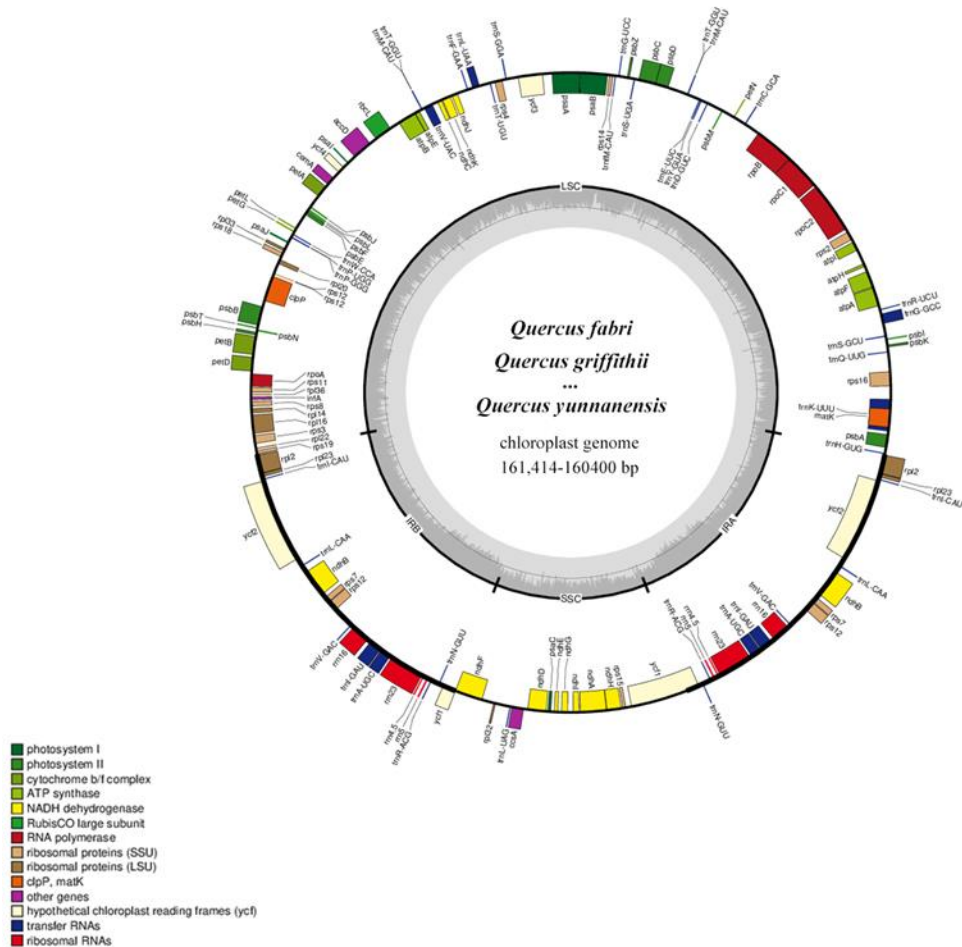


图 3.2 17 种栎属植物的基因图谱

Figure 3.2 Gene map of seventeen *Quercus* species

注：将注释好的栎树叶绿体基因组表示为同心圆。在外圈中外部的基因是逆时针转录的，内部的基因是顺时针转录的。两个反向重复区(IRs)，大单拷贝区(LSC)和小单拷贝区(SSC)的长度以及边界由内部的黑圈表示。

表 3.1 17 种栎属植物叶绿体基因组特征总结

Table 3.1 Summary of the seventeen *Quercus* species chloroplast genome features.

物种名	基因组大小(bp)	编码基因数量(个)	tRNA 数量 (个)	rRNA 数量(个)	LSC 大小(bp)	SSC 大小(%)	IR 大小(%)
	/GC 含量(%)	/大小 (bp)	/大小 (bp)	/大小(bp)	/百分比(%) /GC 含量(%)	/百分比(%) /GC 含量(%)	/百分比(%) /GC 含量(%)
<i>Q. baronii</i>	161114/36.8	86/80103	40/2977	8/9050	90383/56.1/39.6	19045/11.8/30.9	51686/32.1/42.7
<i>Q. spinosa</i>	160825/36.9	86/80812	40/2975	8/9050	90371/56.2/39.6	18732 /11.6/31.2	51722 /32.2/42.72
<i>Q. semicarpifolia</i>	161325/36.8	86/80076	40/2981	8/9050	90341/56.1/39.6	19045/11.8/30.9	51686/32.1/42.7
<i>Q. senescens</i>	161415/36.8	86/80151	40/2982	8/9033	90650/56.2/39.5	19053/11.8/30.8	51712/32.0/42.7
<i>Q. aquifoliodes</i>	160415/37	86/80270	40/2979	8/9048	89822/56/39.7	18893/11.8/31.2	51700/32.2/42.7
<i>Q. cocciferoides</i>	161324/36.9	86/80091	40/2981	8/9050	90625/56.2/39.6	18963/11.8/31.0	51686/32.0/42.6
<i>Q. rehderiana</i>	161332/36.8	86/80077	40/2981	8/9050	90518/56.1/39.5	19040/11.8/31.0	51722/32.1/42.7
<i>Q. monimotriana</i>	161325/36.8	86/80060	40/2909	8/9042	90642/56.2/39.6	18991/11.8/31.0	51690/32.0/42.7
<i>Q. engleriana</i>	161347/36.8	86/80119	40/2909	8/9042	90618/56.2/39.5	19015/11.8/31.0	51712/32.1/42.7
<i>Q. serrata</i> var. <i>brevipetiolata</i>	161266/36.8	86/80059	40/2981	8/9050	90580/56.2/39.6	19034/11.8/31.0	51700/32.1/42.8
<i>Q. yunnanensis</i>	161247/36.8	86/80064	40/2983	8/9050	90564/56.2/39.6	19031/11.8/30.9	51652/32/42.8
<i>Q. aliena</i>	160921/36.9	86/80052	40/2983	8/9048	90258/56.1/34.7	18972/11.8/31	51682/32.1/42.7
<i>Q. griffithii</i>	161346/36.8	86/80075	40/2981	8/9034	90612/56.2/39.6	19082/11.8/30.9	51652/32.0/42.8
<i>Q. serrata</i>	161266/36.8	86/80068	40/2982	8/9050	90580/56.2/34.7	19034/11.8/30.9	51652/32/42.8
<i>Q. stewardii</i>	161262/36.8	86/80088	40/2982	8/9050	90573/56.2/39.6	19037/11.8/30.9	51652/32.0/42.7
<i>Q. fabri</i>	161227/36.8	86/80054	40/2981	8/9050	90539/56.2/39.6	19048/11.8/30.9	51640/32.0/42.8
<i>Q. mongolica</i>	161317/36.8	86/80074	40/2982	8/9050	90605/56.2/39.6	19060/11.8/30.9	51652/32.0/42.8

3.1.2 栎树叶绿体全基因组基因功能分析

植物叶绿体基因组可能有 63-209 个基因,但大多数集中在 110-130 之间,且具有高度保守的组成和排列,包括光合基因、叶绿体转录表达相关基因,以及其他一些蛋白质编码基因(Jansen *et al.*, 2005)。按照基因的功能,17 种栎树叶绿体全基因组编码的基因可以被分为:一、与转录翻译有关的基因,二、与光合作用有关的基因,三、生物合成相关基因,四、功能未知的基因。

3.1.2.1 转录翻译相关基因

在 17 种栎树全基因组基因中,与转录翻译相关的基因有 69 个,分别是核糖体小亚基 *rps*、核糖体大亚基 *rpl*、核糖体 RNA 基因 *rrn*、依赖于 DNA 的 RNA 聚合酶基因 *rpo* 以及转运 RNA 基因 *trn* 具体信息见表 3.2。

3.1.2.2 转录翻译相关基因

在 17 种栎树全基因组基因中,与光合作用相关的基因有 44 个,分别是 ATP 合酶亚基 *atp*、NADH 脱氢酶基因 *ndh*、细胞色素亚基 *pet*、光系统 I 亚基 *psa*、光系统 II 亚基 *psb* 以及 Rubisco 大亚基 *rbcL* 具体信息见表 3.2。

3.1.2.3 生物合成基因

在 17 种栎树全基因组基因中,与生物合成相关的基因有 6 个,分别是乙酰辅酶 A 羧化酶亚基 *accD*、囊膜蛋白基因 *cemA*、细胞色素 C 合成基因 *ccsA*、成熟酶基因 *matK*、以及依赖 ATP 蛋白酶 P 基因亚基 *clpP* 以及翻译起始因子 *infA* 具体信息见表 3.2。

3.1.2.4 未知功能基因

除以上三大类外,还有 *ycf1*、*ycf2*、*ycf2* 以及 *ycf4* 四个功能未知的基因,具体分类情况见表 3.2。

表 3.2 17 种栎树叶绿体全基因组编码基因分类

Table 3.2 Genes in the complete cp genomes of seventeen *Quercus* species

基因功能分类	基因种类	基因名称
转录翻译相关基因	核糖体小亚基	<i>rps2, rps3, rps4, rps7^a, rps8, rps11, rps12^{acd}, rps14, rps15, rps16, rps18, rps19</i>
	核糖体大亚基	<i>rpl2^{ab}, rpl14, rpl16, rpl20, rpl22, rpl23^a, rpl33, rpl32, rpl36</i>
	核糖体 RNA 基因	<i>rrn16^a, rrn23^a, rrn4.5^a, rrn5^a</i>
	依赖于 DNA 的 RNA 聚合酶基因	<i>rpoA, rpoB, rpoC1, rpoC2</i>
	转运 RNA 基因	<i>trnC-GCA, trnH-GUG, trnD-GUC, trnK-UUU, trnQ-UUG, trnE-UUC, trnS-GCU, trnY-GUA, trnG-GCC, trnR-UCU, trnG-UCC, trnT-GGU, trnM-CAU, trnM-CAU, trnS-GGA, trnS-UGA, trnT-UGU, trnL-UAA, trnF-GAA, trnV-UAC, trnM-CAU, trnT-GGU, trnW-CCA, trnP-UGG, trnP-GGG, trnI-CAU, trnL-CAA, trnV-GAC^a, trnI-GAU, trnA-UGC, trnR-ACG^a, trnN-GUU^a, trnL-UAG, trnN-GUU, trnR-ACG, trnA-UGC^a, trnI-GAU^a, trnV-GAC, trnL-CAA^a, trnI-CAU^a</i>
	ATP 合酶亚基	<i>atpA, atpB, atpE, atpF^b, atpH, atpI</i>
	NADH 脱氢酶基因	<i>ndhA^b, ndhB^{ab}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH^b, ndhI, ndhJ, ndhK</i>
	细胞色素亚基	<i>petA, petB, petD, petG, petL, petN</i>
	光系统 I 亚基	<i>psaA, psaB, psaC, psaJ, psal</i>
	光系统 II 亚基	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Rubisco 大亚基	<i>rbcL</i>
生物合成相关基因	乙酰辅酶 A 羧化酶亚基	<i>accD</i>
	囊膜蛋白基因	<i>cemA</i>
	细胞色素 C 合成基因	<i>ccsA</i>
	依赖 ATP 蛋白酶 P 基因亚基	<i>clpP^c</i>
	翻译起始因子	<i>infA</i>
	成熟酶基因	<i>matK</i>
未知功能基因	保守的开放阅读框	<i>ycf1^a, ycf2^a, ycf3^c, ycf4</i>

注：a-反向重复区(IRs)中的双拷贝基因；b-包含一个内含子的基因；c-包含两个内含子的基因；d-分为两个独立转录单元基因。

3.1.3 栎属叶绿体全基因组重复序列分析

3.1.3.1 简单重复序列

通过搜索 17 种栎树叶绿体全基因组序列，榲栌、川滇高山栎、檀子栎、铁橡栎、巴东栎、白栎、大叶栎、蒙古栎、矮高山栎、毛脉高山栎、高山栎、灰背栎、枹栎、短柄枹栎、刺叶高山栎、黄山栎以及云南波罗栎分别检测到 81、76、82、78、76、81、83、84、76、77、76、84、84、84、78、81、83 个 SSR 序列，如表 3.3 所示。17 种栎树的 SSR 类型非常丰富，各类重复类型均存在，其中，单核苷酸重复出现的数量最多，重复次数在 50-56 次之间；五核苷酸、六核苷酸重复类型的数量则较少，值得注意的是：在部分常绿栎（川滇高山栎、巴东栎、矮高山栎、毛脉高山栎、高山栎和灰背栎）中没有检测到六核苷酸重复(表 3.3)。

几乎所有栎树叶绿体基因组内检测到的单核苷酸重复序列均由 A/T 构成，在白栎和川滇高山栎中单核苷酸重复 100%由 A/T 组成；而对于二核苷酸重复序列，17 种栎树的简单重复类型也基本上由 AT/TA 组成(图 3.4)。

这个结果与其它学者对于被子植物(蔡小彦等, 2015; 盛文涛等, 2017)和裸子植物(Zhang *et al.*, 2020)的研究结果类似：简单重复序列主要是由聚腺嘌呤核苷酸(poly A)和聚胸腺嘧啶核苷酸(poly T)构成，这也导致了叶绿体基因组较强的保守性。

表 3.3 17 种栎属物种的简单重复序列统计

Table 3.3 Summary of the simple sequence repeat in seventeen *Quercus* species

	单核苷酸重复	二核苷酸重复	三核苷酸重复	四核苷酸重复	五核苷酸重复	六核苷酸重复	叶绿体基因组总 SSR 数目
<i>Q. aliena</i>	51	12	6	7	4	1	81
<i>Q. aquifoliodes</i>	51	11	5	6	3	0	76
<i>Q. baronii</i>	53	12	8	6	2	1	82
<i>Q. cocciferoides</i>	54	7	7	6	3	1	78
<i>Q. engleriana</i>	53	10	4	6	3	0	76
<i>Q. fabri</i>	51	11	7	7	4	1	81
<i>Q. griffithii</i>	55	10	7	6	4	1	83
<i>Q. mongolica</i>	56	10	7	6	4	1	84
<i>Q. monimotricha</i>	50	11	7	5	3	0	76
<i>Q. rehderiana</i>	53	8	8	6	2	0	77
<i>Q. semicarpifolia</i>	50	11	7	5	3	0	76
<i>Q. senescens</i>	56	11	8	6	3	0	84
<i>Q. serrata</i>	54	11	7	6	5	1	84
<i>Q. serratavar</i>	54	11	7	6	5	1	84
<i>Q. spinosa</i>	50	11	5	6	5	1	78
<i>Q. stewardii</i>	53	11	7	5	4	1	81
<i>Q. yunnanensis</i>	54	11	7	6	4	1	83

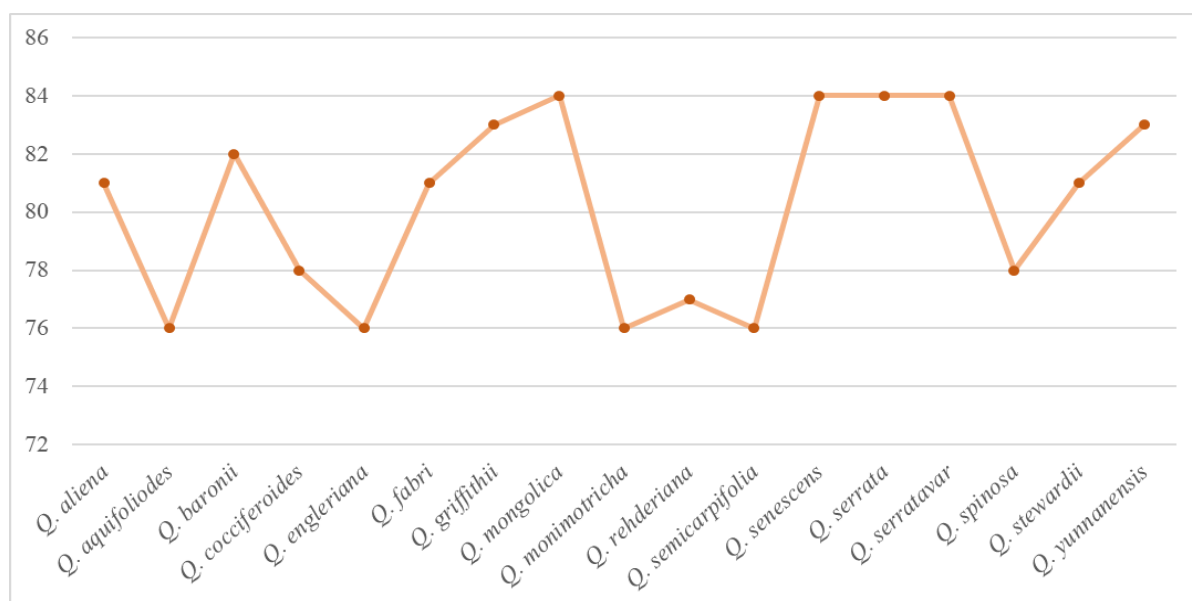


图 3.3 17 种栎属植物叶绿体基因组 SSR 总数折线图

Figure 3.3 Broken line chart of total SSR numbers in chloroplast genomes of 17 *Quercus* species



3.4 17种栎属植物不同类型的简单重复序列数量

Figure 3.4 Number of different simple sequence repeat (SSR) types of 17 *Quercus* species
 注：纵轴为重复类型的个数，横轴为简单序列重复的类型且从左至右长度逐渐增加

3.1.3.2 串联重复序列

在 17 种栎树叶绿体全基因组中，即槲栎、川滇高山栎、槿子栎、铁橡栎、巴东栎、白栎、大叶栎、蒙古栎、矮高山栎、毛脉高山栎、高山栎、灰背栎、枹栎、短柄枹栎、刺叶高山栎、黄山栎以及云南波罗栎叶绿体全基因组序列分别检测出串联重复序列为 8、8、10、10、9、9、11、10、13、16、13、9、9、9、8、

10、11 条，共 173 条。17 种栎树的基本重复单元中最小为 11 bp（黄山栎），最大的为 46 bp(铁橡栎)，重复次数介于 2 至 4 之间。在 17 个种的串联重复长度中 20-29 bp 的最多，有 104 条；值得注意的是：40- 46 bp 之间的串联重复最少，仅在三个常绿栎: 檀子栎(1 条)矮高山栎(3 条)和高山栎(3 条)中存在，同时常绿也包含三个最多的串联重复序列：矮高山栎(13 条)毛脉高山栎（16 条）和高山栎（3 条）（图 3.5 表 3.4）。

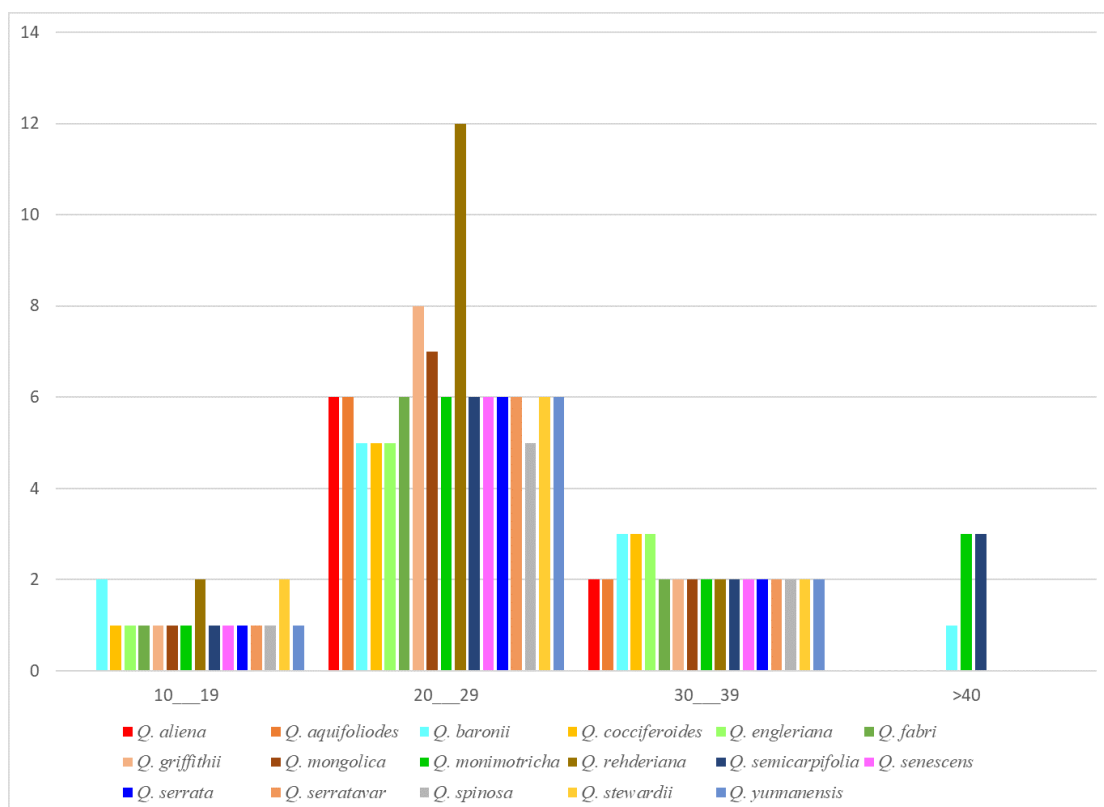


图 3.5 串联重复序列的分布

Figure 3.5 Length distribution of tandem repeat sequences

注：纵轴为串联重复序列的数量

不同叶习性栎属物种间的叶绿体基因组比较研究

表 3.4 17 种栎属植物叶绿体基因组串联重复信息表
Table 3.4 Tandem repeat of chloroplast genome of 17 *Quercus* species

物种	重复大小(bp)	重复次数(次)	匹配率 (%)	参考重复基因	位置
<i>Q. baronii</i>	28	2	100	TAGATGGAACAGAGGGAAATAGATACAT	4796--4852
	14	3	100	ATTACATAATATCC	17620--17661
	35	2	100	TAATTATAATGAATAATATTATATTATTTAATG	35861--35934
	19	2	100	ATTTTATTATTTATATTTA	54384--54423
	20	2	100	TCTTATATATATCGTATAAT	90990--91029
	21	3	82	TTTTTGTCCAAGTACTTCTT	95189--95256
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113527--113587
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	137911--137971
	21	4	81	AAAAAGAGAAGTAACTTAGAC	156243--156331
	20	2	100	ATTATACGATATATATAAGA	160469--160508
<i>Q. spinosa</i>	21	2	100	AAAAAAGTGATACTTAGTATT	14978--15020
	19	2	100	ATTTTATTATTTATATTTA	54343--54382
	20	3	100	TCTTATATATATCGTATAAT	90954--91013
	21	3	82	TTTTTGTCCAAGTACTTCTT	95173--95240
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113511--113571
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	137626--137686
	21	4	81	AAAAAGAGAAGTAACTTAGAC	155959--156047
	20	3	100	ATTATACGATATATATAAGA	160185--160244
<i>Q. semicarpifolia</i>	41	2	98	ATTAATTATTCTTAATAATAATTTTATAGTAATAAAAAATT	11847--11938
	41	3	88	TAATAATAATTTTATAGTAATAAAAAATTATTAATTATTAA	11859--11966
	41	3	83	AATTATTAATTATTAATAATAATTTTATATAATAAAA	11846--11973
	28	2	100	CTTTATTTTCATTTTCGACTCATTATAT	39943--39998

结果与分析

	19	2	100	ATTTTATTATTTATATTTA	54611--54650
	20	2	100	TCTTATATATATCGTATAAT	91251--91290
	21	3	82	TTTTTGTCCAAGTTACTTCTT	95450--95517
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113780--113840
	20	2	100	TCAAATAAACTACTCTTTT	119392--119432
	20	2	100	ACTTTTCTTATTATATTAAT	125150--125190
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138132--138192
	21	4	81	AAAAAGAGAAGTAACTTAGAC	156456--156544
	20	2	100	ATTATACGATATATATAAGA	160682--160721
<i>Q. senescens</i>	27	2	100	ATGTTCTGGGGTTTACATATATTCATA	29945--30000
	19	2	100	ATTTTATTATTTATATTTA	54575--54614
	20	2	100	TCTTATATATATCGTATAAT	91257--91296
	21	3	82	TTTTTGTCCAAGTTACTTCTT	95456--95523
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113795--113855
	20	2	100	ACTTTTCTTATTATATTAAT	125181--125221
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138211--138271
	21	4	81	AAAAAGAGAAGTAACTTAGAC	156544--156632
	20	2	100	ATTATACGATATATATAAGA	160770--160809
<i>Q. aquifoliodes</i>	25	3	89	TTATTTATTTATAAAAACGAGTAG	63467--63528
	25	2	100	AAACGAGTAGTTATTTAGTTTATAA	63480--63528
	20	3	100	TCTTATATATATCGTATAAT	90400--90459
	21	3	82	TTTTTGTCCAAGTTACTTCTT	94619--94686
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	112961--113021
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	137189--137249
	21	4	81	AAAAAGAGAAGTAACTTAGAC	155525--155613
	20	3	100	ATTATACGATATATATAAGA	159751--159810

不同叶习性栎属物种间的叶绿体基因组比较研究

<i>Q. cocciferoides</i>	20	2	100	CTAGAATTATAGAATAGATG	53629--53668
	19	2	100	ATTTTATTATTTATATTTA	54349--54388
	46	2	98	ACTTAACCTATATATAAATAAAAATATATATATAAATAATTTACTAA	77217--77313
	20	2	100	TCTTATATATATCGTATAAT	91047--91086
	21	3	82	TTTTTGTCCAAGTTACTTCTT	95246--95313
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113587--113647
	20	2	100	ACTTTTCTTATTATATTAAT	124955--124995
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	137933--137993
	21	4	81	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	156453--156541
	20	2	100	ATTATACGATATATATAAGA	160679--160718
<i>Q. rehderiana</i>	19	2	96	AATTATATAATAAAAATATA	4942--4987
	22	2	100	TCTCTATATTATATATATTAAT	11275--11319
	22	3	82	TTATTCTCTATATTATATATATTA	11270--11332
	22	2	100	ATACTAAATTAATAAAAATAAT	47369--47413
	20	2	100	CTAGAATTATAGAATAGATG	53716--53755
	19	2	100	ATTTTATTATTTATATTTA	54440--54479
	21	2	100	TCTTATGTGTATATAATTAAT	64428--64469
	20	3	97	TCTTATATATATCGTATAAT	91131--91190
	21	3	82	TTTTTGTCCAAGTTACTTCTT	95350--95417
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113689--113749
	24	2	100	TAAAATACATAACTCTTACTTTAT	119251--119299
	28	2	100	GAATTAATAGTAAAGAACAATTGGTTTT	119557--119615
	20	2	100	ACTTTTCTTATTATATTAAT	125101--125141
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138108--138168
	21	4	81	AAAAAGAGAAGTAACTTAGAC	156441--156529
20	3	97	ATTATACGATATATATAAGA	160667--160726	

结果与分析

<i>Q. monimotricha</i>	41	2	98	ATTAATTATTCTTAATAATAATTTTATAGTAATAAAAAATT	11845--11936
	41	3	88	TAATAATAATTTTATAGTAATAAAAAATTATTAATTATTA	11857--11964
	41	3	83	AATTATTAATTATTAATAATAATTTTATATAATAAAA	11844--11971
	28	2	100	CTTTATTTTCATTTTCGACTCATTATAT	39941--39996
	19	2	100	ATTTTATTATTTATATTTA	54609--54648
	20	2	100	TCTTATATATATCGTATAAT	91249--91288
	21	3	82	TTTTTGCCAAGTACTTCTT	95448--95515
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113778--113838
	20	2	100	TCAAATAAACTACTCTTTT	119390--119430
	20	2	100	ACTTTTCTTATTATATTAAT	125148--125188
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138130--138190
	21	4	81	AAAAAGAGAAGTAACTTAGAC	156454--156542
	20	2	100	ATTATACGATATATATAAGA	160680--160719
	<i>Q. engleriana</i>	21	2	100	AAGAGTGAAGAATAAAAAATA
30		2	100	TCTTGAGAAAGTCTTTCATTTGACTATCAT	60054--60114
19		4	100	TATTATTACTATCTATTAT	78456--78523
20		2	100	TCTTATATATATCGTATAAT	91226--91265
21		3	82	TTTTTGCCAAGTACTTCTT	95425--95492
31		2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113764--113824
31		2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138142--138202
21		4	81	AAAAAGAGAAGTAACTTGGAC	156475--156563
20		2	100	ATTATACGATATATATAAGA	160701--160740
<i>Q. serratavar</i>		29	3	93	TTATTATAACTAATTATAACAAATCTAAC
	19	2	100	ATTTTATTATTTATATTTA	54549--54588
	25	2	100	TTATTATATCACATATACTCACTTC	64833--64882
	20	2	100	TCTTATATATATCGTATAAT	91187--91226

不同叶习性栎属物种间的叶绿体基因组比较研究

	21	3	82	TTTTGTCCAAGTTACTTCTT	95374--95441
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113713--113773
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138074--138134
	21	3	83	AAAAAGAGAAGTAACTTGGAC	156407--156475
	20	2	100	ATTATACGATATATATAAGA	160621--160660
<i>Q. yunnanensis</i>	29	3	93	TTATTATAACTAATTATAACAAATCTAAC	47428--47500
	19	2	100	ATTTTATTATTTATATTTA	54532--54571
	25	2	100	TTATTATATCACATATACTCACTTC	64816--64865
	20	2	100	TCTTATATATATCGTATAAT	91171--91210
	21	3	82	TTTTGTCCAAGTTACTTCTT	95358--95425
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113697--113757
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138055--138115
	21	3	83	AAAAAGAGAAGTAACTTGGAC	156388--156456
	20	2	100	ATTATACGATATATATAAGA	160602--160641
	27	2	97	ATATATATATAATAAATAATATTTAAT	6789--6852
	20	3	100	TCTTATATATATCGTATAAT	90841--90900
<i>Q. aliena</i>	27	2	97	ATATATATATAATAAATAATATTTAAT	6789--6852
	20	3	100	TCTTATATATATCGTATAAT	90841--90900
	21	3	82	TTTTGTCCAAGTTACTTCTT	95048--95115
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113389--113449
	26	3	92	TCAACATTTTATTTCTTACTTATGTG	120277--120341
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	137731--137791
	21	4	81	AAAAAGAGAAGTAACTTAGAC	156066--156154
	20	3	100	ATTATACGATATATATAAGA	160280--160339
<i>Q. griffithii</i>	27	2	100	ATTTTAATTTCTACCATTCTCTTGT	131--184
	29	3	93	TTATTATAACTAATTATAACAAATCTAAC	47459--47531

结果与分析

	19	3	100	ATTTTATTATTTATATTTA	54562--54620
	25	2	100	TTATTATATCACATATACTCACTTC	64863--64912
	20	2	100	TCTTATATATATATCGTATAAT	91219--91258
	21	3	82	TTTTTGCCAAGTTACTTCTT	95406--95473
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113745--113805
	22	2	100	CTAGTTTTTATTCATCAATTTT	120922--120965
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138154--138214
	21	3	83	AAAAAGAGAAGTAACTTGGAC	156487--156555
	20	2	100	ATTATACGATATATATAAGA	160701--160740
<i>Q. serrata</i>	29	3	93	TTATTATAACTAATTATAACAAATCTAAC	47431--47503
	19	2	100	ATTTTATTATTTATATTTA	54549--54588
	25	2	100	TTATTATATCACATATACTCACTTC	64833--64882
	20	2	100	TCTTATATATATATCGTATAAT	91187--91226
	21	3	82	TTTTTGCCAAGTTACTTCTT	95374--95441
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113713--113773
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138074--138134
	21	3	83	AAAAAGAGAAGTAACTTGGAC	156407--156475
	20	2	100	ATTATACGATATATATAAGA	160621--160660
<i>Q. stewardii</i>	29	3	93	TTATTATAACTAATTATAACAAATCTAAC	47426--47498
	19	2	100	ATTTTATTATTTATATTTA	54529--54568
	25	2	100	TTATTATATCACATATACTCACTTC	64812--64861
	11	4	100	TAATTAATTAT	72704--72747
	20	2	100	TCTTATATATATATCGTATAAT	91180--91219
	21	3	82	TTTTTGCCAAGTTACTTCTT	95367--95434
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113706--113766
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138070--138130

不同叶习性栎属物种间的叶绿体基因组比较研究

<i>Q. fabri</i>	21	3	83	AAAAAGAGAAGTAACTTGGAC	156403--156471
	20	2	100	ATTATACGATATATATAAGA	160617--160656
	29	3	93	TTATTATAACTAATTATAACAAATCTAAC	47425--47497
	19	2	100	ATTTTATTATTTATATTTA	54528--54567
	25	2	100	TTATTATATCACATATACTCACTTC	64812--64861
	20	2	100	TCTTATATATATCGTATAAT	91146--91185
	21	3	82	TTTTTGTCCAAGTACTTCTT	95333--95400
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113666--113726
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138041--138101
	21	3	83	AAAAAGAGAAGTAACTTGGAC	156368--156436
<i>Q. mongolica</i>	20	2	100	ATTATACGATATATATAAGA	160582--160621
	27	2	100	ATTTTAATTTCTACCATTCTCTTGTT	131--184
	29	3	93	TTATTATAACTAATTATAACAAATCTAAC	47467--47539
	19	2	100	ATTTTATTATTTATATTTA	54571--54610
	25	2	100	TTATTATATCACATATACTCACTTC	64853--64902
	20	2	100	TCTTATATATATCGTATAAT	91212--91251
	21	3	82	TTTTTGTCCAAGTACTTCTT	95399--95466
	31	2	96	CATTGTTCAACTCTTTGACAACACGAAAAAC	113738--113798
	31	2	96	TTTTTCATGTTGTCAAAGAGTTGAACAATGG	138125--138185
	21	3	83	AAAAAGAGAAGTAACTTGGAC	156458--156526
20	2	100	ATTATACGATATATATAAGA	160672--160711	

3.1.3.3 正向、反向、回文和互补重复

本研究中，我们利用在线软件 REPuter 一共检测到 17 个物种的 693 个重复 (附表 1-4)，包括四种重复类型：回文重复(Palindromic match)、正向重复(Forward match)、反向重复(Reverse match)和互补重复(Complement match)。回文重复和正向重复出现的次数最多，分别被检测到 358 次和 255 次；反向重复和互补重复出现的次数较少，分别被检测到 62 次和 18 次 (表 3.5, 图 3.6)。反向重复和互补重复甚至没有在所有物种中被检测到，在川滇高山栎中没有检测到反向重复序列，在部分常绿栎 (川滇高山栎、铁橡栎、巴东栎、矮高山栎、高山栎和灰背栎) 中没有检测到互补重复序列。大多数的重复序列在 30-71 bp 之间，且分布的区域主要在大单拷贝区和反向重复区 (附表 1-4)。总体而言，这四种重复在这 17 种栎属植物叶绿体基因组中的数量和分布都是十分保守的。

表 3.5 四种重复序列的长度分布

Table 3.5 Distribution of four repetitive sequences

	回文重复	正向重复	反向重复	互补重复	重复总数
<i>Q. aliena</i>	26	16	3	3	48
<i>Q. yunnanensis</i>	20	15	5	1	41
<i>Q. aquifolioides</i>	25	14	0	0	39
<i>Q. cocciferoides</i>	17	14	1	0	32
<i>Q. serrata</i>	20	15	6	1	42
<i>Q. spinosa</i>	23	13	2	1	39
<i>Q. baronii</i>	21	14	2	2	39
<i>Q. engleriana</i>	19	15	7	0	41
<i>Q. fabri</i>	20	15	4	1	40
<i>Q. griffithii</i>	20	17	3	1	41
<i>Q. mongolica</i>	20	16	4	2	42
<i>Q. monimotricha</i>	20	14	3	0	37
<i>Q. rehderiana</i>	22	17	4	1	44
<i>Q. semicarpifolia</i>	20	14	3	0	37
<i>Q. senescens</i>	20	15	3	0	38
<i>Q. serrata</i> var. <i>brevipetiolata</i>	20	15	6	1	42
<i>Q. stewardii</i>	25	16	6	4	51
重复总数	358	255	62	18	693

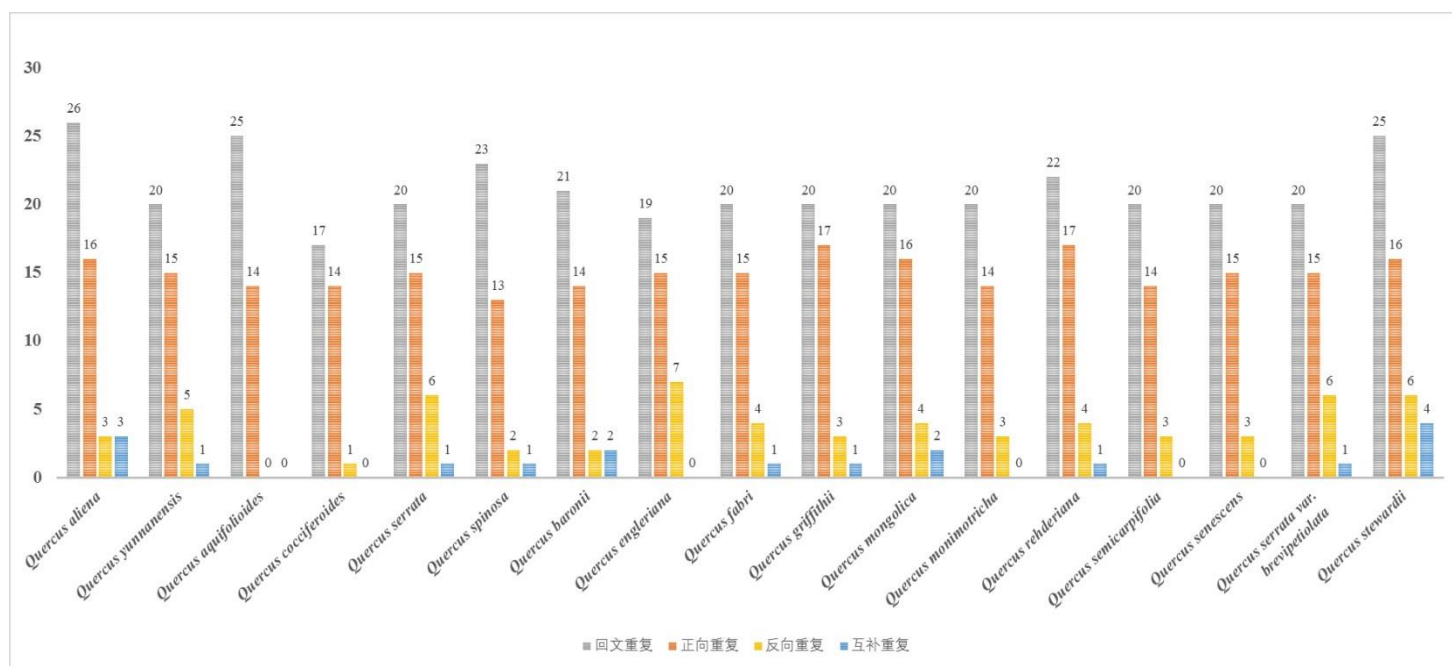


图 3.6 17 种栎属物种中四种重复的分布

Figure 3.6 Distribution of four repeat sequences in 17 *Quercus* species

注：纵坐标表示重复序列的数量

3.1.4 栎属叶绿体全基因组序列多样性

尽管包括基因数量和顺序在内的整个叶绿体四分体结构通常都很保守，但 IR 区域经常会发生扩张或收缩，这种现象在叶绿基因组中称为消长(ebb and flow)(Goulding *et al.*, 1996)。一般来说，扩张或收缩不超过几百个核苷酸。Kim 和 Lee 提出被子植物叶绿体基因组的长度变异主要是由 IR 区和单拷贝边界区的扩张和收缩引起的(Kim *et al.*, 2004)。为了检测 17 种栎树叶绿体基因组序列的差异程度，我们首先使用在线分析工具 IRscope 对 17 个物种的各分区长度以及边界信息进行可视化，17 个物种按照从常绿栎到落叶栎的顺序排列。结果显示，尽管 17 种栎树叶绿体基因组结构和大小都具有高度的保守性，但由于 IR 的收缩/扩张使得 IR/SC 边界在各个物种中表现出了一定差异(图 3.7)。位于 *ycf1* 基因的编码区被 IRB 和 SSC 的边界穿过，*ycf1* 有 4,589-4,630bp 的区域位于 SSC 区，而 *ycf1* 的另外一个稍短的拷贝穿过 IRA，有 72-15bp 的区域位于 SSC 区。在大单拷贝区 LSC 与反向重复区 IRa 的交汇处，*rps19* 与边界在距离在 9-35bp 之间。在大单拷贝区 LSC 和反向重复区 IRb 的交界处，*trnH* 与边界的距离除了川滇高山栎是 30 bp 外，其余 16 个种都为 16 bp。

当关注到不同叶习性物种间序列差异时，我们发现长拷贝的 *ycf1* 基因在常绿栎中变化幅度较大从 5,642bp 到 5,687bp 之间，平均长度为 5,668 bp；而在落

叶柄中长拷贝的 *ycf1* 基因长度则更稳定,除榲栌为 5,654 bp 外其余长度均为 5,657 bp, 平均长度也略小于常绿柄为 5,656 bp。这种长度的差异可以在长拷贝 *ycf1* 基因在 SSC 区的长度中体现出来, 在常绿柄中除了檀子柄(1,047 bp)川滇高山柄(1,043 bp)外, 其余 7 个种的 *ycf1* 基因位于 IRb 区的长度都 \geq 1,057 bp, 而在落叶的 8 个种中这个长度除榲栌(1,058 bp)外其余 7 个种都为 1,042 bp。同样, 对于短拷贝的 *ycf1* 基因, 尽管它的长度没有在哪一种叶习性的柄中显示出一致性, 但是在落叶柄中的平均长度 1,093 bp 却长于在常绿柄中的平均长度 1,089 bp。在常绿柄中除了灰背柄(72 bp)外, 其余 8 个种的 *ycf1* 基因位于 SSC 区的长度都在 40 bp 以下, 而在 8 种落叶柄中这个长度都在 40 bp 以上。

从结果不难看出, 在 17 个种的叶绿体基因组中, 反向重复区(IRa& IRb)的收缩/扩张导致了整个叶绿体基因组长度的变化, 也导致 LSC 和 SSC 的边界位置发生了变动。

Inverted Repeats

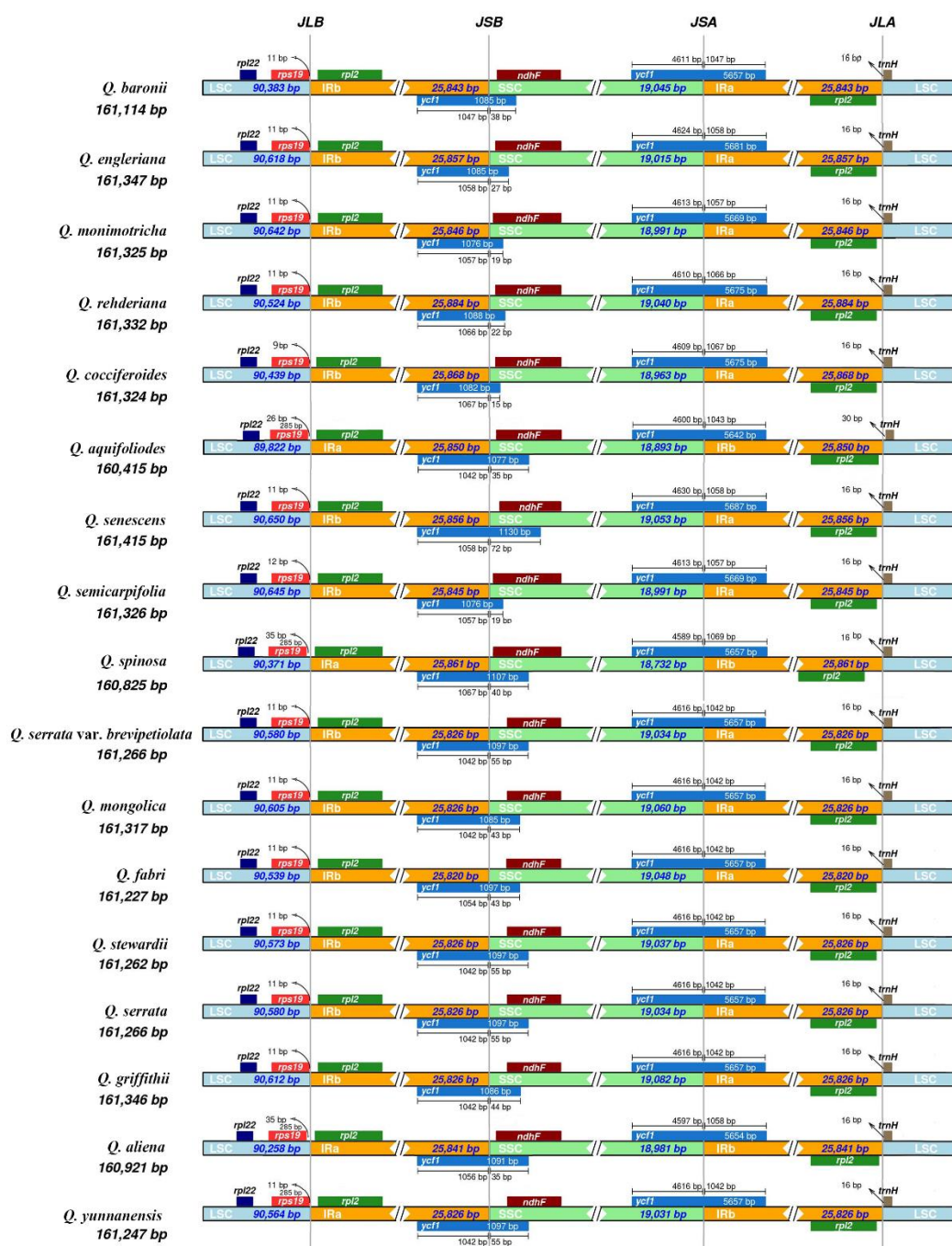


图 3.7 17 个栎叶绿体基因组 LSC、IR 和 SSC 交界处的比较

Figure 3.7 Comparison of the LSC, IRs, and SSC junction regions of seventeen *Quercus* chloroplast genomes

注：基因特征上方的数字是指基因末端与边缘位点之间的距离。

以槲子栎(*Q. baronii*)的叶绿体基因组序列为参考进行了 CCT 分析, mVISTA 分析。从结果可以看出, 17 个栎树物种的叶绿体基因组表现出高度的保守性(图

3.8, 图 3.9), 其中 CCT 图显示 17 个种的序列一致性在 96%以上。mVISTA 分析结果显示, 编码区序列的变异要小于非编码区, 且 SC 区相比 IR 区变异要小, 这与之前的研究结果一致(Liu *et al.*, 2013; Nazareno *et al.*, 2015; Yin *et al.*, 2018)。

在 mVISTA 分析中我们把 17 个种按照从常绿到落叶的顺序, 发现有部分变异区共同出现在同种叶习性的栎中。其中有 6 处差异区域为 8 个落叶种共有, 仅有 2 处为 8 个常绿种共有, 这也许与参考种是常绿种榿子栎有关。

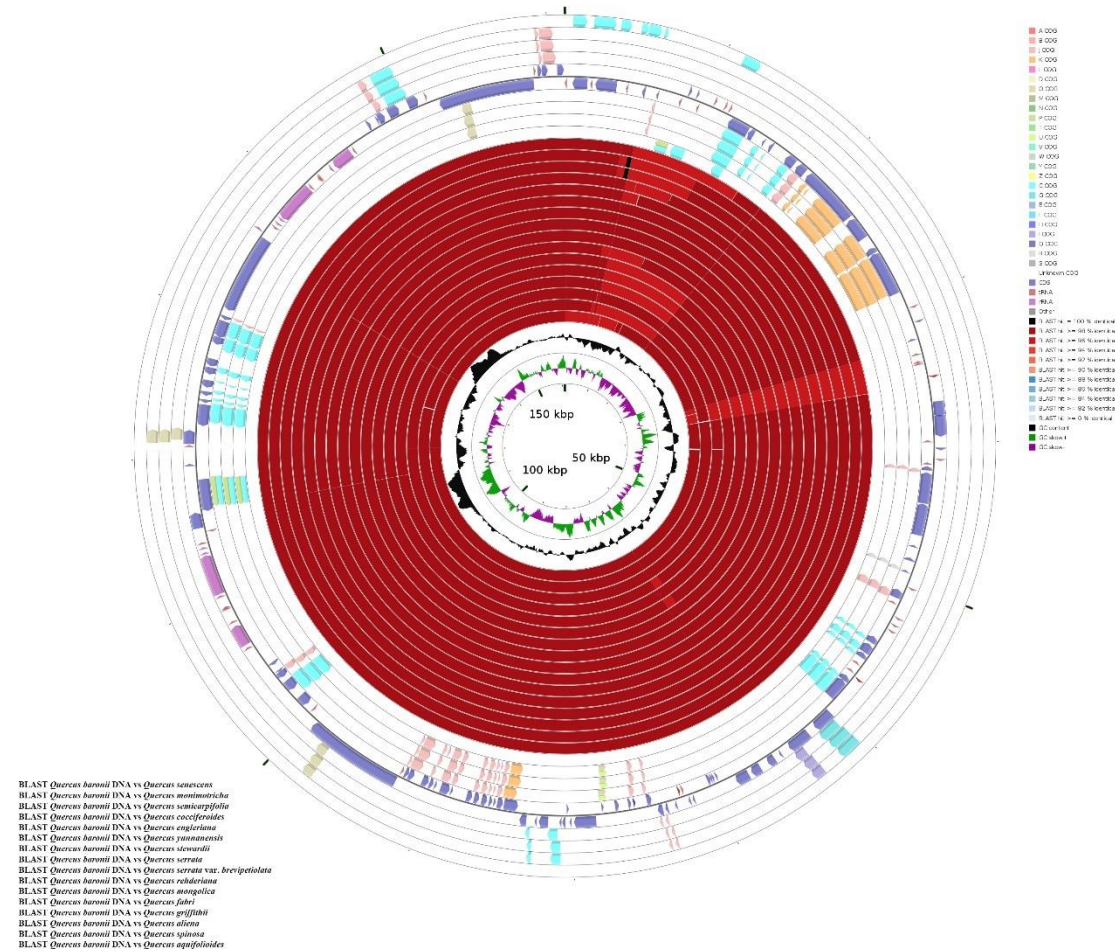


图 3.8 17 种栎属植物 cp 基因组序列的比较。

Figure 3.8 Comparison of seventeen *Quercus* cp genome sequences

注: 最外围的四个环分别显示了编码序列、tRNA 基因、rRNA 基因和其他正向和反向链中的基因。接下来的 16 个环显示了榿子栎和其他 16 个栎属物种叶绿体基因组之间的 blast 结果。再往内的黑环是榿子栎基因组的 GC 含量曲线。最里面的环代表了榿子栎基因组的 GC 偏倚图(GC skew)。GC skew+ (绿色) 就表示 G 的含量大于 C, GC skew- (紫色) 表示 G 的含量小于 C。

不同叶习性栎属物种间的叶绿体基因组比较研究

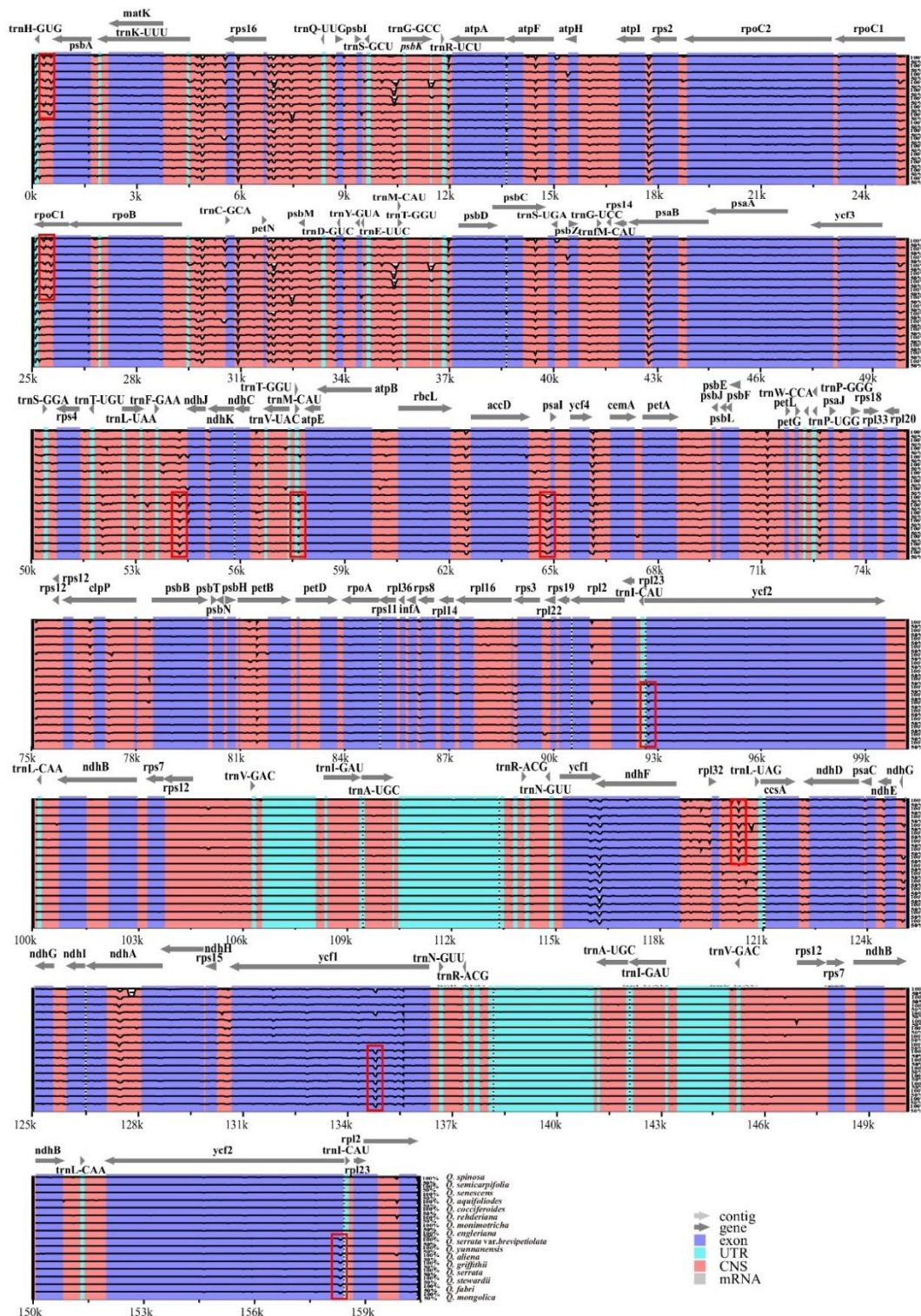


图 3.9 17 种栎属植物叶绿体基因组的序列差异可视化结果

Figure 3.9 Visualization of sequence differences in chloroplast genomes of 17 *Quercus* species

为了研究这 17 种栎属叶绿体基因组之间的核苷酸多样性，我们首先将每一个物种的叶绿体全基因组数据按分区为了 3 段，其中反向重复区 IR 只取其中一段 IRA。计算了基因组的大单拷贝区、小单拷贝区和反向重复区中 600 bp 窗

口(200 bp 步长)内的核苷酸变异性(P_i)值(图 3.10)。在大单拷贝区中数值从 0 到 0.01151 不等, 平均值为 0.002203, 而在小单拷贝区从 0 到 0.01471 不等, 平均值为 0.006173, 反向重复从 0 到 0.00545 不等, 平均值为 0.000685。这些结果表明, 这些基因组区域之间的差异非常小。然而, 我们也在大单拷贝区、小单拷贝区和反向重复区中发现了某些高度可变的区域。在大单拷贝区, 高度可变的区域是 *rpoB*, *rpoB/trnC-GCA*, *trnC-GCA/petN*, *psaB* 和 *psaA* 它们的 $P_i > 0.01$ 。在小单拷贝区中, 高度可变区包括 *ndhF*, *ndhF/rpl32*, *rpl32/trnL-UAG* 和 *ycf1* ($P_i > 0.01$)。在反向重复区中 *rps12/trnV-GAC*, *rrn5*, *trnR-ACG/trnN-GUU* 和 *ycf1*, $P_i > 0.002$ 。其中的五个区域 *trnC-GCA/petN*, *ndhF*, *ndhF/rpl32*, *trnR-ACG/trnN-GUU* 和 *ycf1*, 在其他植物中也被确定为高度可变的(Kim *et al.*, 2004; Dong *et al.*, 2012; Awad *et al.*, 2017; Yin *et al.*, 2018; Li *et al.*, 2021; Liu *et al.*, 2021)。基于我们的结果, 这些可变区显示出作为栎属有用的系统发育标记的来源的巨大潜力。

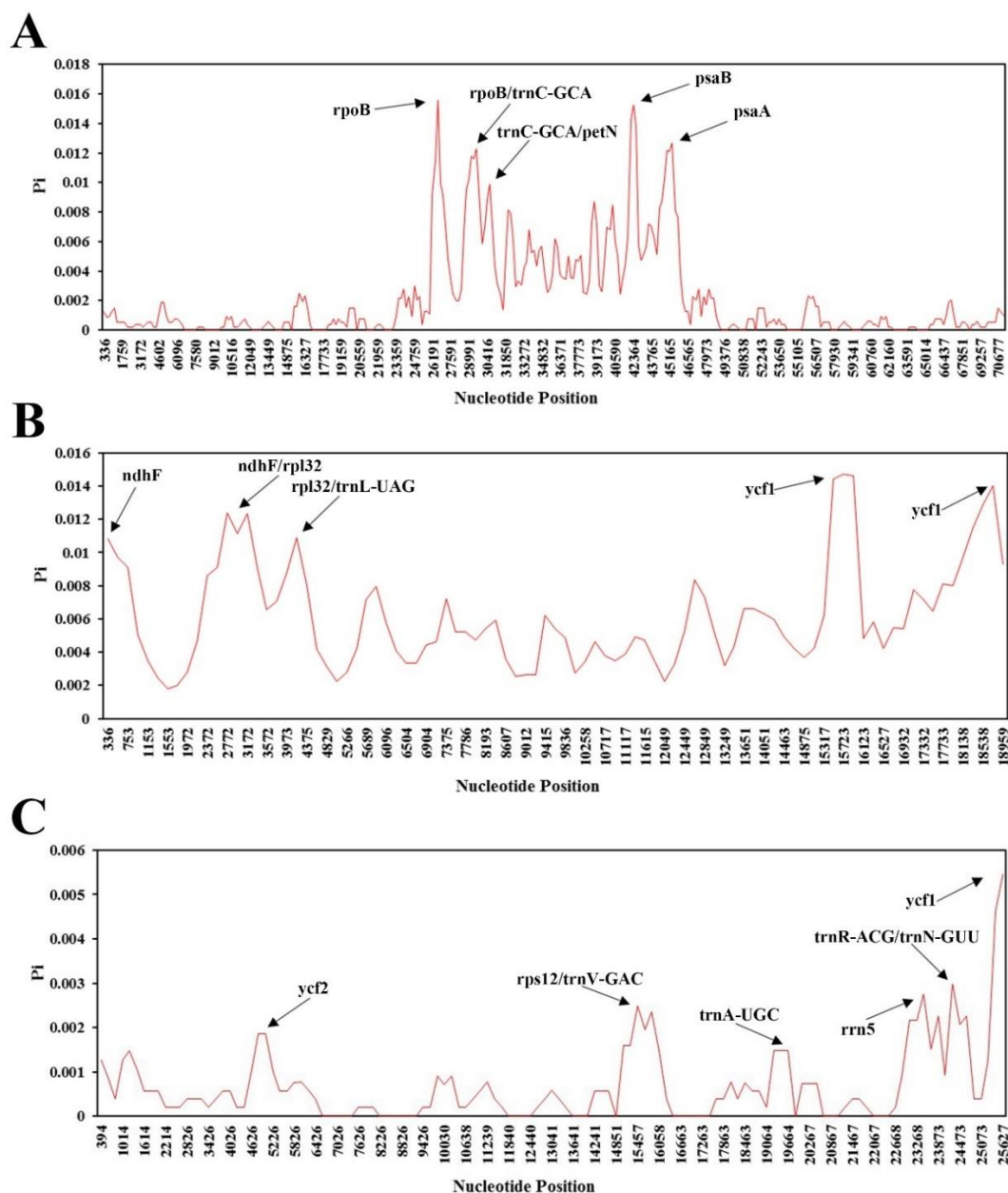


图 3.10 17 种栎树叶绿体基因组序列核苷酸多样性(P_i)值比较分析

Figure 3.10 Comparative analysis of nucleotide variability (P_i) values among the seventeen *Quercus* cp genome sequences.

注: (A)LSC 区域分析; (B)SSC 区域分析; (C)IR 区域分析。(窗口长度: 600 bp, 步长: 200 bp)。X 轴: 窗口中点的位置, Y 轴: 每个窗口的核苷酸多样性。

3.2 栎属叶绿体基因组系统发育分析

本研究通过将栎属 34 个物种的叶绿体基因组处理为两个数据集: 一号数据集使用 34 个物种的叶绿体基因组直接进行比对, 最后比对的长度为 164,576 bp。二号数据集使用 HomBlocks 进行同源序列筛选, 最终生成比对好的同源序列为 97,786 bp。

34 种东亚栎树的两个数据在使用三种方法：ML 法、NJ 法及 BI 法重建的系统发育树基本一致。从图 3.11 可以看出，本研究构建的三个系统发育树中，贝叶斯法分析得到的支持率普遍最高，其次最大简约法分析的支持率普遍较高。从发育树的拓扑结构可以看出，各个组的物种都彼此聚集，尤其是白栎组的 13 个物种单独聚为了一支；同时各个亚属的组也链接十分紧密，白栎组作为栎亚属在东亚的唯一一个组单独聚成一支，麻栎亚属的麻栎组和青冈栎组则镶嵌在冬青栎组中。从分化时间来看，包括所有高山栎在内部分冬青栎组物种和所有青冈栎组植物处于系统发育的基部位置，其余部分的冬青栎组、麻栎组和白栎组物种则处在较后的位置。值得注意的是，蒙古栎 *Quercus mongolica* & 辽东栎 *Quercus wutaishanica* 以及枹栎 *Quercus serrata* & 短柄枹栎 *Quercus serrata* var. *brevipetiolata* 这两对在《*Flora of China*》(Huang *et al.*, 1999)中被合并的物种，在我们的系统发育树中表现出了十分紧密的亲缘关系。

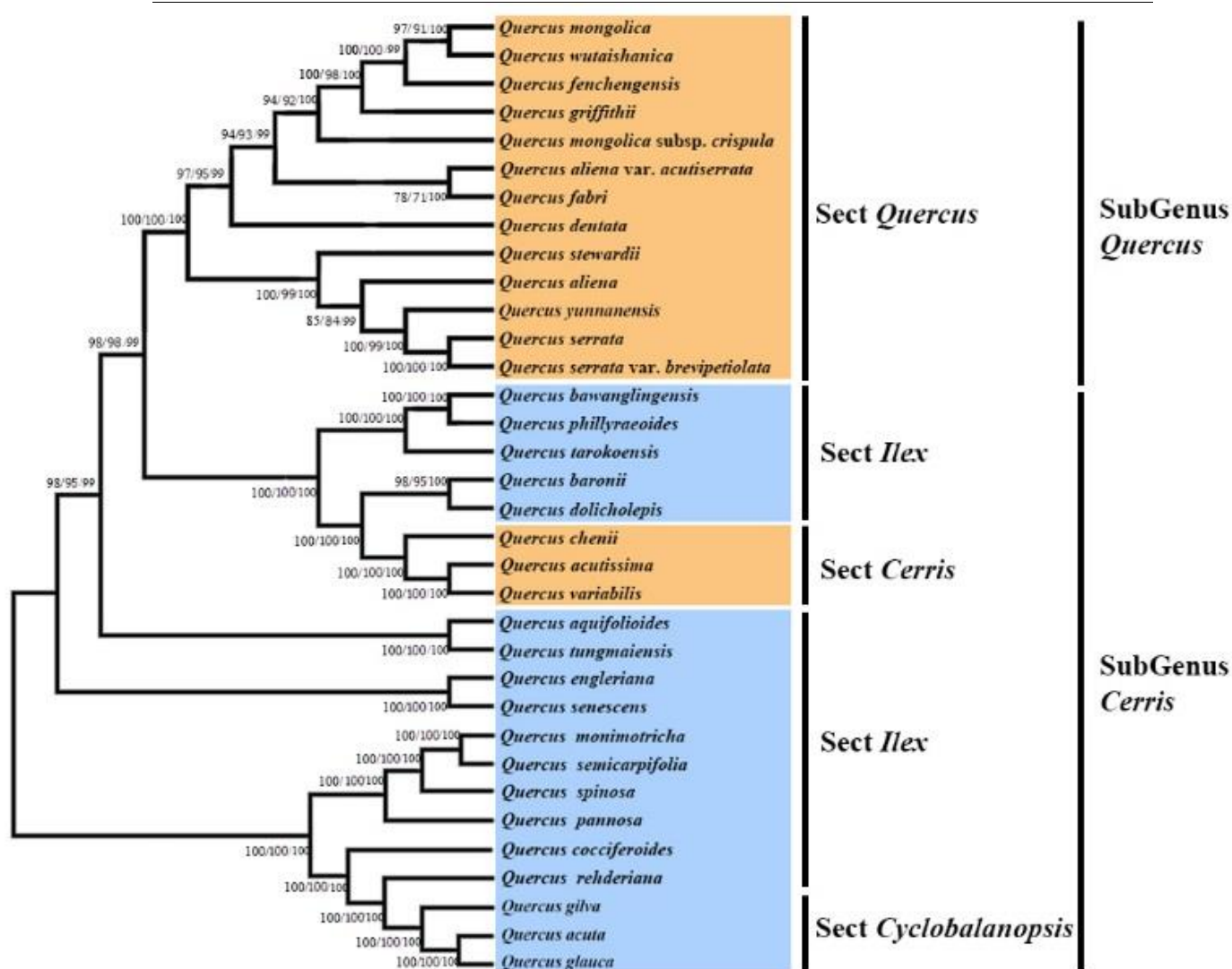


图 3.11 34 个物种叶绿体基因组同源序列构建的 ML、NJ 和 BI 系统发育树
Figure 3.11 ML、NJ、BI phylogenetic tree rebuild by homologous sequences of 34 *Quercus* species cp genomes

注: Bootstrap 值显示在节点上。落叶物种用黄色方框表示, 而常绿物种用蓝色方框表示。图右侧显示的是属以下的分类信息。

3.3 17 种栎树叶绿体基因组的正选择计算

非同义核苷酸替代率(dN)、同义核苷酸替代率(dS)以及它们的比值 ω ($\omega = dN/dS$)被广泛用作加速进化和正向选择或适应性进化的指标。为了研究栎属叶绿体蛋白编码基因的分子进化, 在本研究中我们通过测序、组装拼接及基因注释, 得到了 14 个栎属物种完整的叶绿体基因组, 结合 NCBI 中已经发表的 20 个东亚栎属物种的叶绿体基因序列, 共 34 个物种进行正选择分析。

3.3.1 栎属叶绿体基因组共有基因序列筛选

首先, 确定了 34 个物种的 72 个共有且不重复的蛋白质编码基因(表 3.6)。在 PAML 中利用 *codeml* 程序的基本模型(model = 0, Nsites = 0)的结果显示 72 个共有基因的 dS 值都低于 1, 这确保了足够的系统发育分辨率来准确估计整个系统发育的 dN/dS(Escalona *et al.*, 2017)。

3.3.2 栎属叶绿体基因组正选择计算

为了检测每一个基因的位点特异性正选择, 我们在 *codeml* 中使用两对位点模型(site-model): M1vs M2, M7 vs M8 进行正选择检测。结果显示(表 3.7)在 34 个物种的 72 个共有基因中 *rbcL*, *rpl20*, *ndhF* 和 *ycf1* 通过模型检测、似然比检验以及 FDR 校正之后被检测为正选择基因。其中 *rpl20* 和 *rbcL* 处于大单拷贝区上, *ndhF* 和 *ycf1* 基因处于小单拷贝区上。从基因功能上看 *rpl20* 是转录翻译相关的基因, *ndhF* 和 *rbcL* 是光合作用相关基因, *ycf1* 是未知功能基因, 表明大多数被正选择的基因是与光合作用相关(表 3.6)。

在利用 branch-site model 时, 我们发现在设置常绿物种为前景支后, 光合作用基因 *ndhF* 再次显示出正选择的特征(表 3.8), 受到正选择的位点在巴东栎、坝王栎、乌冈栎、太鲁阁栎、刺叶高山栎、檀子栎、匙叶栎、铁橡栎、黄背栎、高山栎、矮高山栎、灰背栎、赤栎、青冈以及赤皮青冈中都被检测到, 占有研究常绿栎的 15/18。同时另一个被检测为正选择基因的是生物合成相关基因 *ccsA*, 它在设置常绿物种为前景支时, 从匙叶栎、太鲁阁栎、通麦栎以及灰背栎中检测出正选择位点。

表 3.6 34 种栎树叶绿体基因组用于正选择分析的基因

Table 3.6 Genes used for positive selection analysis in chloroplast genomes of 34 *Quercus* species

基因功能分类	基因种类	基因名称
转录翻译相关基因	核糖体小亚基	<i>rps11, rps12^{acd}, rps14, rps15, rps18, rps19, rps2, rps3, rps4, rps7^a, rps8,</i>
	核糖体大亚基	<i>rpl14, rpl16, rpl2^{ab}, rpl20, rpl22, rpl23^a, rpl33, rpl36</i>
	依赖于 DNA 的 RNA 聚合酶基因	<i>rpoA, rpoB, rpoC1, rpoC2</i>
光合作用相关基因	ATP 合酶亚基	<i>atpA, atpB, atpE, atpF^b, atpH, atpI</i>
	NADH 脱氢酶基因	<i>ndhA^b, ndhB^{ab}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH^b, ndhI, ndhJ, ndhK</i>
	细胞色素亚基	<i>petA, petB, petD, petG, petN</i>
	光系统 I 亚基	<i>psaA, psaB, psaC</i>
	光系统 I II 亚基	<i>PsbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Rubisco 大亚基	<i>rbcL</i>
	乙酰辅酶 A 羧化酶亚基	<i>accD</i>
生物合成相关基因	囊膜蛋白基因	<i>cemA</i>
	细胞色素 C 合成基因	<i>ccsA</i>
	依赖 ATP 蛋白酶 P 基因亚基	<i>clpP^c</i>
	成熟酶基因	<i>matK</i>
未知功能基因		<i>ycf1^a, ycf2^a, ycf3^c, ycf4</i>

注：a-反向重复区(IRs)中的双拷贝基因；b-包含一个内含子的基因；c-包含两个内含子的基因；d-分为两个独立转录单元的基因。

表 3.7 34 种栎属植物相应基因正选择位点的位点模型结果
表 3.8 Positively selected sites in corresponding genes of 34 *Quercus* species with site model

基因	功能	位点模型	-lnL	2δ(lnL)	P 值	Q 值	正选择位点
<i>rbcL</i>	ribulose 1, 5-bisphosphate carboxylase/oxygenase large subunit	model7	2100.886	24.97572	3.77E-06	1.36E-04	95-0.942, 219-0.946, 251-20.946, 262-0.942, 328-0.972
		model8	2088.398				
<i>rpl20</i>	ribosomal protein L20	model7	554.934	15.55895	4.18E-04	7.52E-03	105-0.999
		model8	547.155				
<i>ndhF</i>	NADH dehydrogenase subunit 5	model7	3573.115	20.48336	3.57E-05	8.57E-04	24-0.576, 90-0.519, 422-0.897, 457-0.917, 486-0.984, 497-0.903, 501-1.000, 570-0.579, 575-0.504, 576-0.888, 581-0.999, 611-0.887, 655-0.985, 740-0.916
		model8	3562.874				319-0.561, 333-0.559, 337-0.848, 383-0.992, 397-0.553, 417-0.512, 461-0.554, 520-1.000, 522-0.516, 524-0.532, 529-0.937, 542-0.847, 543-0.868, 560-0.505, 579-0.865, 585-0.916, 587-0.916, 650-0.551, 732-0.883, 746-0.999, 750-0.556, 756-0.986, 792-0.948, 895-0.564, 911-0.836, 943-0.517, 946-0.839, 996-0.850, 1395-0.910, 1438-0.524, 1479-0.881, 1480-0.988, 1481-0.891, 1509-0.508, 1545-0.905, 1559-0.982, 1574-0.510, 1580-0.541, 1666-0.557, 1674-0.907, 1697-0.934, 1707-0.842, 1717-0.502, 1766-0.866, 1846-0.555
<i>ycf1</i>	hypothetical chloroplast RF19	model7	9064.58	45.34509	1.42E-10	1.02E-08	
		model8	9041.9				

表 3.8 34 种栎属植物相应基因正选择位点的分支位点模型结果
 Table 3.7 Positively selected sites in corresponding genes of 34 *Quercus* species with branch-site model

基因	功能	分支位点模型	-lnL	2 δ (lnL)	P 值	Q 值	正选择位点
<i>ccsA</i>	cytochrome c biogenesis protein	Branch (Evergreen species)					
		Null	1397.11				89-0.968, 96-0.623, 166-0.626
		Alternative	1390.49	13.24	8.07E-05	5.81E-03	
<i>ndhF</i>	NADH dehydrogenase subunit 5	Branch (Evergreen species)					
		Null	3584.57				486-0.505, 501-0.939, 581-0.505, 655-0.513
		Alternative	3581.49	6.16	8.80E-04	3.17E-02	

4 讨论

4.1 栎属叶绿体基因组序列进化

本研究测序、拼接并注释了 14 种中国本土栎树的叶绿体基因组，长度范围在 160,415 bp 到 161,415 bp 之间。比较分析发现栎属植物的叶绿体基因组的大小、结构、基因内容和 GC 含量都是高度保守的，展示出了较高的序列相似性。简单重复序列分析中，几乎所有栎树叶绿体基因组内检测到的单核苷酸和二核苷酸重复序列基本上由 A 和 T 组成 (A/T; AT/TA) (图 3.4)。该研究结果与其他学者对于被子植物(蔡小彦等, 2015; 盛文涛等, 2017)和裸子植物(Zhang *et al.*, 2020)的研究结果类似: 高等植物叶绿体基因组的简单重复序列主要是由 A 和 T 构成。

本研究检测了 17 种栎属植物叶绿体基因组中的串联重复、正向重复、反向重复、互补重复和回文重复的长度、数量和分布位置情况。结果显示在所有栎物种中回文重复的占比最高 (358 条)，互补重复 (18 条) 和反向重复 (59 条) 的数量最少，值得注意的在部分常绿栎中不存在六核苷酸重复、互补重复和反向重复序列。以往的研究也表明回文重复能够代替反向区进行同源重组，提高基因结构的多样性，进而增加基因组重排的概率，在植物叶绿体基因组中发挥着重要的作用(Zhang *et al.*, 2014; 张悦, 2019)。同时，重复序列的长度将会影响叶绿体基因组的结构和功能 (Yi *et al.*, 2013)。不同于其它植物的研究，从栎属基因组重复序列的长度来看栎属内发现的重复长片段(大于 40 bp)是非常少的(844 个重复中仅出现 73 次)，这似乎是壳斗科叶绿体基因组的一种共性 (杨颜慈等, 2018)。

4.2 栎属叶绿体基因组序列差异比较

尽管植物叶绿体基因组在基因组结构和大小上是高度保守的，但 IR 区的收缩/扩张仍可能会导致基因组长度的变化，同时 IR/SC 边界也会因此发生改变，这一现象在植物叶绿体基因组中比较普遍(Kim and Lee, 2004; Hansen *et al.*, 2007; Wang *et al.*, 2008; Huang *et al.*, 2014;)。我们发现这种收缩/扩张在同种叶习性的栎属植物中存在一定的相似性，这可能因为常绿栎和落叶栎分别适应彼此之间差异明显的环境而产生的适应性进化(Ju *et al.*, 2019; Du *et al.*, 2020)。此前有学者对陆地植物的研究发现 IR 区的变化引起 SC 区和 IR 区内基因的相互转移或基因的增减(Zhu *et al.*, 2016)，但是在我们的研究中并未观察到相似的现象。

通过 CCT、Mvista 以及 Pi 对 17 个栎属叶绿体基因组序列进行比较，发现 SC 区包含较多核苷酸变异(Pi)值高的区域，而 IR 区较少，意味着相比于 SC 区 IR 区具有更低的序列多样性水平。之前的研究已表明在多种植物中由于复制依

赖修复机制(copy-dependent repair mechanism)的存在使得 IR 区的替换速率较低 (Perry and Wolfe, 2002)。在大多数植物中 SC 区的替换速率是 IR 区的几倍(Zhu *et al.*, 2016)。在 mVISTA 分析中我们按照从常绿到落叶的顺序, 发现有部分序列差异区域共同出现在同种叶习性的栎中。其中有 6 处差异区域为 8 个落叶种共有, 仅有 2 处为 8 个常绿种共有, 这可能因为常绿种和落叶种分别适应彼此之间差异明显的环境而产生的适应性进化(Ju *et al.*, 2019; Du *et al.*, 2020)。此外我们还检测到编码区 *ndhF*、*ycf1*、*ycf2* 非编码区 *rps16 - trnQ-UUG*、*psbZ - trnG-UCC*。这些热点区域(divergent hotspot region)提供了有价值的信息, 可以作为发展分子标记用于系统发育研究和栎属物种鉴定的基础。

4.3 栎属叶绿体基因组系统发育

将 34 个栎属植物叶绿体全基因组数据以及它们的同源数据分别以 ML, NJ 以及 BI 法构建系统发育树, 通过不同数据集构建的系统发育树都得到了较高支持率, 而且基于不同数据通过不同方式重建的系统发育关系也大体一致。但是同源数据集得到的结果支持度要高于叶绿体全基因组序列的结果, 这也许是由于在 同源数据中一些高变异的区域(如一些高变的非编码区)被移除从而减少了系统误差。与其它植物中的研究结果类似, 三种构树方法中 BI 法得到的支持率最高, 其次是 ML 法分析的支持率。

栎属植物系统发育研究的困难不仅在于其广泛的杂交渐渗, 也在于不同核标记、叶绿体标记所构建的系统发育关系的结果不一致(Manos *et al.*, 1999; Oh and Manos, 2008; Denk and Grimm, 2010; Simeone *et al.*, 2013; Hubert *et al.*, 2014; Simeone *et al.*, 2016; Vitelli *et al.*, 2017)。我们得出的分类结果与之前研究(Denk and Grimm, 2010; Hubert *et al.*, 2014)中通过核标记得到的全球栎属植物的分类系统一致: 34 种东亚栎被划分为 2 个亚属(subGenus *Quercus*, SubGenus *Cerris*)和 4 个组(Sect. *Quercus*, Sect. *Ilex*, Sect. *Cerris*, Sect. *Cyclobalanopsis*, Sect. *Quercus*)。落叶栎中的 *Quercus* 组中 13 个种彼此聚为一支, 落叶栎中的另一个组 *Cerris* 虽然同样彼此聚集但是嵌套在 *Ilex* 组当中, 没有形成单系。对化石数据的研究(Denk and Tekleva, 2014; Hubert *et al.*, 2014; Simeone *et al.*, 2016)表明 *Cerris* 组是从 *Ilex* 组进化而来, 这也许可以作为以上分类结果的一个证据。这种对于东亚栎属的分组方式虽然与之前的分类方式(巴东栎组, 高山栎组, 榲栎组, 檀子栎组和麻栎组)(彭焱松 2007)不同, 却在最近的研究中得到验证(Yang *et al.*, 2021)。

尤其值得注意的是, 在 *Quercus* 组中的蒙古栎 *Quercus mongolica* & 辽东栎 *Quercus wutaishanica* 以及枹栎 *Quercus serrata* & 短柄枹栎 *Quercus serrata* var. *brevipetiolata* 以及 *Ilex* 组中的刺叶高山栎 *Quercus spinosa* & 黄背栎 *Quercus pannosa* 这三对在 *Flora of China*(Huang *et al.*, 1999)中被合并的物种, 在我们的系

统发育树中表现出了十分紧密的系统发育关系。但是锐齿槲栎 *Quercus aliena* var. *acutiserrata* 作为槲栎 *Quercus aliena* 的变种在我们的研究中并没有表现出较为紧密的系统发育关系，这也仍需要进一步分析。

在我们的研究中没有像核基因数据一样得到 *Ilex* 组的单系分支，这种质核不一致的现象也许是受到了叶绿体捕获的影响，这种现象在之前在其他植物的研究中就有过报道(Petit *et al.*, 2003; Acosta and Premoli, 2010)。在 *Ilex* 组内部可以看到所有的高山栎组和部分巴东栎组植物分布在系统发育树的基部位置，而其余部分的巴东栎组和檀子栎组植物则处在较上的位置。

之前的研究(周浙昆, 1993; 贺金生, 1994)表明高山栎特殊的叶结构可以使其适应高海拔地区干燥、低温、低氧以及强紫外线的恶劣环境。同样是常绿树种的巴东栎组因为不适应这种极端环境而逐渐向温暖的低海拔地区迁移，因此现在巴东栎组的分布区域与高山栎相比明显向东偏移(周浙昆, 1992)。对栎属的种群遗传学和景观基因组学的研究同样发现，栎属的遗传分化很大程度上是受到气候因素和地理因素的影响(Ju *et al.*, 2019; Du *et al.*, 2020)。

4.4 正选择分析

栎属叶片呈现两种不同习性：常绿和落叶，因此提供了一个很好的模型去帮助我们理解自然选择在驱动叶绿体基因进化中的作用以及识别近缘物种适应性进化的关键基因。光合作用是植物重要的能量积累过程，需要多种蛋白参与。利用 site 模型和 branch-site 模型共同对东亚分布的 34 种栎树(常绿 18 种，落叶 16 种)进行正选择分析，其中 site 模型检测出四个正选择基因：*rbcL*, *rpl20*, *ndhF* 和 *ycf1*；branch-site 模型在设定常绿树种为前景支时检测出两个正选择基因：*ndhF* 和 *ccsA*。*ndhF* 在两个模型中都被检测到，其中 branch-site 模型从 15 个(总共 18 个常绿种)常绿种的 *ndhF* 基因中检测出正选择位点，此前的研究表明 *ndh* 基因可能参与叶绿体对光氧化胁迫的保护(Martin *et al.*, 1996)，同时植物抗氧化胁迫能力与其对其他环境条件的抗性也密切相关(Shaaltiel *et al.*, 1988)。考虑到大多数的常绿栎生活在海拔较高的石灰岩基质区域(包括栎属中分布海拔最高的高山栎组植物)(贺金生, 1994)，适应较高的紫外辐射强度、缺氧、低温以及干旱等环境胁迫，可能是栎属叶绿体水平适应性进化的重要遗传基础。

5 结论与展望

5.1 结论

本研究对东亚栎属 14 个种（常绿栎树：高山栎、灰背栎、檀子栎、铁橡栎、毛脉高山栎、矮高山栎以及巴东栎，落叶栎树：枹栎、短柄枹栎、云南波罗栎、大叶栎、黄山栎、白栎以及蒙古栎）进行采样、测序、拼接及注释，最终得到了 14 种栎树完整的叶绿体基因组并上传至 NCBI。其中白栎的叶绿体基因组序列是首次测序并公布。结合本实验室已发表的三种栎树（槲栎、川滇高山栎和刺叶栎）的叶绿体基因组进行了结构比较分析。使用 ML 法、NJ 法及 BI 法三种方法对所选 34 个物种的叶绿体全基因组序列以及同源序列重建了栎属系统发育树。最后利用 PAML 中的 site 模型和 branch-site 模型对 34 个物种的共有不重复蛋白质编码基因进行正选择分析。

主要结论如下：

(1) 17 种栎树叶绿体基因组的长度范围为 160,415 bp 到 161,415 bp，它们呈典型的四分体结构，相同区拥有相似的长度。17 种栎树基因组的 CG 含量整体差异不大，在 36.8% -37%之间(极差 0.3%)。每个物种都包含 86 个蛋白质编码基因、40 个 tRNA 基因和四个 rRNA 基因。在这些基因中，*rps7*, *rps12*, *rpl2*, *rpl23*, *rrn16*, *rrn23*, *rrn4.5*, *rrn5*, *trnV-GAC*, *trnR-ACG*, *trnN-GUU*, *trnN-GUU*, *trnA-UGC*, *trnI-GAU*, *trnL-CAA*, *trnI-CAU*, *ndhB*, *ycf1* 以及 *ycf2* 有两个拷贝；*rpl2*, *atpF*, *ndhA*, *ndhB* 以及 *ndhH* 包含一个内含子；*rps12*, *clpP* 以及 *ycf3* 包含两个内含子；*rps12* 分为两个独立转录单元。17 个物种所包含的基因数量和顺序是一致的，且与其它已发表的壳斗科叶绿体基因组相似。

(2) 本研究统计了 17 个栎属叶绿体基因组简单重复、串联重复、正向重复、反向重复、互补重复和回文重复的长度、数量和分布位置情况。结果显示在所有物种中回文重复的占比最高（358 条），互补重复（18 条）和反向重复（59 条）的数量最少。在部分常绿栎中不存在六核苷酸重复、互补重复和反向重复序列。

(3) 对比分析发现，SC 区包含较多核苷酸变异(Pi)值高的区域，而 IR 区较少。同时，同种叶习性的物种其基因组结构存在着一定的相似性，但是在落叶栎和常绿栎的基因组又有着明显的差异。此外，检测到的高变区(*trnC-GCA/petN*, *ndhF*, *ndhF/rpl32*, *trnR-ACG/trnN-GUU* 和 *ycf1*)可作为后续栎属植物系统发育和 DNA 条形码研究的候选标记。

(4) 通过三种方法对所选 34 个物种的两种数据集构建的系统发育树得到了几乎相同的拓扑结构。与之前基于核标记得到的对全球栎属植物的分类系统一致，34 种东亚栎被划分为 2 个亚属(subgenus *Quercus*, subgenus *Cerris*)和 4 个组(sect.

Quercus, sect. *Ilex*, sect. *Cerris*, sect. *Cyclobalanopsis*)。本研究不仅提供了对栎属叶绿体基因组的基本理解,也证明叶绿体系统发育基因组学可以为解决栎属的系统发育关系提供帮助。

(5)利用 site 模型和 branch-site 模型共同对东亚分布的 34 种栎树(常绿 18 种,落叶 16 种)的 72 个共有基因进行正选择分析,其中 site 模型检测出四个正选择基因: *rbcL*, *rpl20*, *ndhF* 和 *ycf1*; branch-site 模型在设定常绿树种为前景支时检测出两个正选择基因: *ndhF* 和 *ccsA*。引人注目的是,大多数正选择基因都是光合作用相关基因, *ndhF* 在两个模型中都被检测到,其中 branch-site 模型从 15 个(总共 18 个常绿种)常绿种的 *ndhF* 基因中检测出正选择位点,以往的研究表明 *ndh* 基因家族可能参与叶绿体对光氧化胁迫的保护,因此,适应高海拔地区较高的紫外辐射强度、缺氧、低温以及干旱等环境胁迫,可能是栎属叶绿体水平适应性进化的重要遗传基础。

5.2 展望

本论文对东亚栎属叶绿体基因组多样化结构、系统发育关系以及选择压力进行了研究,但依然存在一些问题需要解决。

- (1) 本研究使用 34 个栎属植物的叶绿体全基因组数据集和同源数据集构建系统发育树,但是可能还有其它支持度更高的数据集可供选择,例如蛋白质编码区数据集,密码子 1+2 位数据集,去除 IR 区的全基因组数据集。
- (2) 本论文对 34 个栎属物种的 72 个共有基因进行正选择分析,在设置常绿栎树为前景支时检测出了正选择基因 *ndhF* 和 *ccsA*;在设置落叶栎为前景支时未检测出正选择基因。在下一步的研究中,为了挖掘更多不同栎属支系之间的正选择证据,可以考虑更多的前景支设置方式:按亚属分类,按组分类。
- (3) 本文对所检测出的正选择基因的功能提出了猜想,在未来的研究中可以对栎属内正选择基因的功能及作用机理进行更详细的探究。

参考文献

- 蔡小彦, 刘方, 周忠丽, 王星星, 王春英, 王玉红, 王坤波. 棉花叶绿体基因组全序列微卫星分布规律研究[J]. 棉花学报, 2015, 27(6): 570-575.
- 陈琴怡. 两种五加科植物的叶绿体全基因组研究及其系统发育分析[D]. 浙江大学, 2017.
- 贺金生, 陈伟烈, 王勋陵. 高山栎叶的形态结构及其与生态环境的关系[J]. 植物生态学报, 1994, 18(3): 219-227.
- 黄成就, 张永田, 徐永椿. 中国植物志[M]. 北京: 科学出版社, 1998.
- 彭焱松, 陈丽, 李建强. 中国栎属植物的数量分类研究[J]. 武汉植物学研究, 2007, 25(2): 149-157
- 盛文涛, 柴学文, 邹家华, 饶友生. 芦笋叶绿体基因组序列微卫星分布规律研究[J]. 园艺与种苗, 2017, 4: 39-41.
- 王崇, 王连军, 苏文瑾, 雷剑, 柴沙沙, 杨新笋, 张文英. 基于 cpSSR 标记的甘薯品种亲缘关系及遗传多样性分析[J]. 分子植物育种, 2020(5): 1687-1696.
- 谢海坤, 焦健, 樊秀彩, 张颖, 姜建福, 孙海生, 刘崇怀. 基于高通量测序组装“赤霞珠”叶绿体基因组及其特征分析[J]. 中国农业科学, 2017, 50(9): 1655-1665.
- 杨颜慈. 中国栎属植物和壳斗科主要属质体基因组比较分析和系统发育研究[D]. 西北大学, 2018.
- 于惠敏, 解相林. 叶绿体的“内共生”与“基因转移”现象[J]. 生物学通报, 2007, L0: 17-19.
- 张悦. 红豆杉科叶绿体基因组结构多样化研究[D]. 北京林业大学, 2019.
- 周浙昆. 中国栎属的地理分布[J]. 中国科学院研究生院学报, 1993, 10(1):95-108
- 周浙昆. 中国栎属的起源演化及其扩散[J]. 云南植物研究, 1992, 14(3):227-236
- Alexander L.W., Woeste K.E. Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management[J]. Tree Genetics & Genomes, 2014, 10(4): 803-812.
- Awad M, Fahmy R M, Mosa K A, Helmy M, El-Feky F A. Identification of effective DNA barcodes for *Triticum* plants through chloroplast genome-wide analysis. Computational Biology and Chemistry, 2017, 71: 20–31.
- Benson G. Tandem repeats finder: a program to analyze DNA sequences[J]. Nucleic Acids Research, 1999, 27(2): 573-580.
- Birky C W, Maruyama T, Fuerst P. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results[J]. Genetics,

- 1983, 103(3): 513-527.
- Blazier J C, Ruhlman T A, Weng M, Rehman S K, Sabir J, Jansen R K. Divergence of RNA polymerase α subunits in angiosperm plastid genomes is mediated by genomic rearrangement[J]. Scientific Reports, 2016, 6(24595):1-15.
- Brinkmann B, Klintschar M, Neuhuber F, *et al.* Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat.[J]. American Journal of Human Genetics, 1998, 62(6):1408-15.
- Burger W C. The species concept in *Quercus*. TAXON, 1975, 24: 45–50.
- Cai Z, Guisinger M, Kim HG, Ruck E, Blazier JC, McMurtry V, Kuehl JV, Boore J, Jansen RK. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (*Fabaceae*) is associated with numerous repeated sequences and novel DNA insertions. Journal of Molecular Evolution. 67:696–704.
- Cavender-Bares J. Diversity, distribution and ecosystem services of the North American oaks. International Oaks, 2016, 27: 37–48.
- CJ Howe, AC Barbrook, VL Koumandou, RER Nisbet, HA Symington, TF Wightman. Evolution of the chloroplast genome [J]. Philosophical Transactions of the Royal Society B, 2003, 358(1429): 99-107.
- Chasse B. Updated classification of oaks: A summary. International Oaks, 2018, 29: 11–18.
- Chen D M, Zhang X X, Kang H Z, Sun X, Yin S, Du H.M, Yamanaka N, Gapare W, Wu H X, Liu C J. Phylogeography of *Quercus variabilis* based on chloroplast DNA sequence in East Asia: Multiple glacial refugia and mainland-migrated island populations. PLoS ONE, 2012, 7: e47268.
- Chen X D, Yang J, Feng L, Zhou T, Zhang H, Li H M, Bai G Q, Meng X, Li Z H, Zhao G F. Phylogeography and population dynamics of an endemic oak (*Quercus fabri* Hance) in subtropical China revealed by molecular data and ecological niche modeling. Tree Genetics & Genomes, 2020, 16(2):1-13.
- Cheng H, Li J, Zhang H, Cai B, Lin M. The complete chloroplast genome sequence of strawberry (*Fragaria × ananassa* Duch.) and comparison with related species of Rosaceae[J]. Peerj, 2017, 5(10):e3919.
- Chumley T W, Palmer J D, Mower J P, Matthew F H, Calie P J, Boore J L, Jansen R K. The complete chloroplast genome sequence of *Pelargonium × hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants[J]. Molecular Biology and Evolution, 2006, 23(11): 2175-2190.

- Crowl A A, Manos P S, McVay J D, Lemmon A R, Lemmon E M, Hipp A L. Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). *New Phytologist*, 2020, 226: 1158–1170.
- Delannoy E, Fujii S, Colas Des Francs-Small C, Brundrett M, Small I. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes[J]. *Molecular Biology and Evolution*, 2011, 28(7): 2077-2086.
- Deng M, Hipp A L, Song Y G, Li Q S, Coombes, A.; Cotton, A. Leaf epidermal features of *Quercus* subgenus *Cyclobalanopsis* (*Fagaceae*) and their systematic significance. *Botanical Journal of the Linnean Society*, 2014, 176: 224–259.
- Deng M, Jiang X L, Hipp A L, Manos P S, Hahn M. Phylogeny and biogeography of East Asian evergreen oaks (*Quercus* Section *Cyclobalanopsis*; *Fagaceae*): Insights into the Cenozoic history of evergreen broad-leaved forests in subtropical Asia. *Molecular Phylogenetics and Evolution*, 2018, 119: 170–181.
- Denk T, Grimm G W, Manos P S, Deng M, Hipp A L. An updated infrageneric classification of the oaks: Review of previous taxonomic schemes and synthesis of evolutionary patterns. In *Tree Physiology. Oaks Physiological Ecology. Exploring the Functional Diversity of Genus Quercus L*, Gil-Pelegrin, E., Peguero-Pina, J.J., Sancho-Knapik, D., Eds, Springer: Cham, Switzerland, 2017, 13–38.
- Denk T, Grimm G W. The oaks of western Eurasia: Traditional classifications and evidence from two nuclear markers. *TAXON*, 2010, 59: 351–366.
- Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data[J]. *Nucleic Acids Research*, 2017, 45(4): e18.
- Dong W, Liu J, Yu J, Wang L, Zhou S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE*, 2012, 7: e35071.
- Doyle J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue[J]. *Phytochem Bull*, 1987, 19(1): 11-15.
- Drummond A J, Suchard M A, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST1.7[J]. *Molecular Biology and Evolution*, 2012, 29(8): 1969-1973.
- Du F K, Hou M, Wang W T, Mao K S, Hampe A. Phylogeography of *Quercus aquifolioides* provides novel insights into the Neogene history of a major global hotspot of plant diversity in south-west China. *Journal of Biogeogr*, 2017, 44: 294–

- 307.
- Du F K, Lang T, Lu S, Wang Y, Li J, Yin K Q. An improved method for chloroplast genome sequencing in non-model forest tree species[J]. *Tree Genetics & Genomes*, 2015, 11(114):1-14.
- Du F K, Wang T R, Wang Y Y, Ueno S, Lafontaine G D. Contrasted patterns of local adaptation to climate change across the range of an evergreen oak, *Quercus aquifolioides*[J]. *Evolutionary Applications*, 2020, 00: 1-15.
- Du F K, Xu F, Qu H, Feng S S, Tang J J, Wu R L. Exploiting the transcriptome of Euphrates Poplar, *Populus euphratica*(*Salicaceae*) to develop and characterize new EST-SSR markers and construct an EST-SSR database[J]. *Plos One*, 2013, 8(4): e61337.
- Dyall S D, Brown M T, Johnson P J. Ancient invasions: from endosymbionts to organelles[J]. *Science*, 2004, 304(5668): 253-257.
- Eaton, D A R, Hipp A L, González-Rodríguez A, Cavender-Bares J. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, 2015, 69: 2587–2601.
- Edwards A I, Civitello A A, Hammond H A, Caskey C T. DNA Typing and Genetic Mapping with Trimeric and Tetrameric Tandem Repeats[J]. *The American Journal of Human Genetics*, 1991, 49(4): 746-756.
- Escalona T, Weadick C J, Antunes A. Adaptive patterns of mitogenome evolution are associated with the loss of shell scutes in turtles. *Molecular Biology and Evolution*, 2017, 34(10): 2522–36.
- Frazer K A, Pachter L, Poliakov A, *et al.* VISTA: computational tools for comparative genomics [J]. *Nucleic Acids Research*, 2004, 32(2): 273-279.
- Funk H T, Berg S, Krupinska K, Maier U G, Krause K. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*[J]. *BMC Plant Biology*, 2007, 7(1): 45.
- Gao L Z, Liu Y L, Zhang D, Li W, Eichler E E. Evolution of *Oryza* chloroplast genomes promoted adaptation to diverse ecological habitats[J]. *Communications Biology*, 2009, 2, 278.
- Gao L, Su Y, Wang T. Plastid genome sequencing, comparative genomics, and phylogenomics: Current status and prospects[J]. *Journal of Systematics and Evolution*, 2010, 48(2): 77-93.
- Gray M W. The evolutionary origins of organelles [J]. *Trends in Genetics* 5: 294-299.
- Goulding S E, Wolfe K H, Olmstead R G, Morden C W. Ebb and flow of the chloroplast

- inverted repeat. *Molecular and General Genetics*, 1996, 252: 195–206.
- Grant J R, Arantes A S, Stothard P. Comparing thousands of circular genomes using the CGView Comparison Tool[J]. *Bmc Genomics*, 2012, 13(1): 202.
- Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family *Geraniaceae*: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution*. 28:583–600.
- Hill R S, Jordan G J. The evolutionary history of *Nothofagus* (*Nothofagaceae*)[J]. *Australian Systematic Botany*, 1993, 6(2): 111-126.
- Hill R S. Biogeography, evolution and palaeoecology of *Nothofagus* (*Nothofagaceae*): the contribution of the fossil record[J]. *Australian Journal of Botany*, 2001, 49(3): 321-332.
- Hipp A L, Manos P S, Gonzalez-Rodriguez A, Hahn M, Kaproth M, McVay J D, Avalos, S V, Cavender-Bares J. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytologist*. 2018, 217: 439–452.
- Hipp A L, Manos P S, Hahn M, Avishai M, Bodenes C, Cavender-Bares J, Crowl A A, Deng M, Denk T, Fitz-Gibbon S, *et al.* Genomic landscape of the global oak phylogeny. *New Phytologist*. 2020, 226: 1198–1212.
- Hsu C, Wu C, Chaw S. Ancient nuclear plastid dna in the yew family (*Taxaceae*)[J]. *GENOME Biology and Evolution*, 2014, 6(8): 2111-2121.
- Hu H L, Wang L Z, Yang J, Zhang R S, Qin L. The complete chloroplast genome of *Quercus fenchengensis* and the phylogenetic implication[J]. *Mitochondrial DNA Part B*, 2019, 4(2): 3066-3067.
- Huang C C, Chang Y T, Bartholomew B. *Fagaceae*. In: Wu C Y, Raven P H, (Eds.) *Flora of China*[M]. Science Press and Missouri Botanical Garden Press, 1999: 380–400.
- Huelsenbeck J P, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees[J]. *Bioinformatics*, 2001, 17(8): 754-755.
- Jansen R.K, Raubeson, L A, Boore J L, Depamphilis C W, Chumley T W, Haberle R C, Wyman S K, Alverson A J, Peery R, Herman S J. Methods for obtaining and analyzing whole chloroplast genome sequences[J]. *Method Enzymol*, 2005, 395: 348.
- Jiang X L, Gardner E M, Meng H H, Deng M, Xu G B. Land bridges in the Pleistocene contributed to flora assembly on the continental islands of South China: Insights from the evolutionary history of *Quercus championii*[J]. *Molecular Phylogenetics*

- And Evolution, 2019, 132: 36–45.
- Jiang X L, Hipp A L, Deng M, Su T, Zhou Z K, Yan M X. East Asian origins of European holly oaks (*Quercus* Section *Ilex* Loudon) via the Tibet-Himalaya[J]. Journal of Biogeogr, 2019, 46: 2203–2214.
- Ju M M, Feng L, Yang J, Yang Y C, Chen X D, Zhao G F. Evaluating population genetic structure and demographic history of *Quercus spinosa* (*Fagaceae*) based on specific length amplified fragment sequencing[J]. Frontiers in Genetics, 2019, 10: 965.
- Kans J A, Ouellette B F. Submitting DNA sequences to the databases[J]. Methods of Biochemical Analysis, 1998, 39(3): 319.
- Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. Molecular biology and evolution, 2013, 30(4): 772-780.
- Kearse M, Moir R, Wilson A, Ashton B, Meintjes P, Drummond A, *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data[J]. Bioinformatics, 2012, 28(12): 1647-1649.
- Kim K J, Lee H L. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants[J]. DNA Research, 2004, 11: 247–261.
- Kremer A, Abbott A G, Carlson J E, Manos P S, Plomion C, Sisco P, Staton M E, Ueno S, Vendramin G G. Genomics of *Fagaceae*[J]. Tree Genetics & Genomes, 2012, 8: 583–610.
- Kugita M, Kaneko A, Yamamoto Y, Ashton B, Meintjes P, Drummond A, *et al.* The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants[J]. Nucleic Acids Research, 2003, 31(2): 716-721.
- Kurtz S, Choudhuri J V, Ohlebusch E, Chris S, Jens S, Robert G. REPuter: the manifold applications of repeat analysis on a genomic scale[J]. Nucleic Acids Research, 2001, 29(22): 4633-4642.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools[J]. Bioinformatics, 2009, 25(16): 2078-2079.
- Li J, Tang J, Zeng S, Han F, Yu J. Comparative plastid genomics of four *Pilea* (*Urticaceae*) species: insight into interspecific plastid genome diversity in *Pilea*[J]. BMC Plant Biology, 2021, 21(1):25.

- Librado, Rozas. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. [J]. *Bioinformatics* (Oxford, England), 2009, 25(11): 1451.
- Lilly J W, Havey M J, Jackson S A. Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants[J]. *The Plant Cell*, 2001, 13(2): 245-254.
- Liu X, Chang E M, Liu J F, Huang Y N, Jiang Z P. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of *Quercus bawanglingensis* Huang, Li et Xing, a Vulnerable Oak Tree in China[J]. *Forests*, 2019, 10(7): 587.
- Liu X, Chang E, Liu J, Jiang Z. Comparative analysis of the complete chloroplast genomes of six white oaks with high ecological amplitude in China[J]. *Journal of Forestry Research*, 2021(41).
- Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young H A, Feng X, Gu Y Q. Complete Chloroplast Genome Sequences of Mongolia Medicine *Artemisia frigida* and Phylogenetic Relationships with Other Plants. *PLoS ONE*, 2013, 8: e57533.
- Logan D C. The mitochondrial compartment[J]. *Journal of experimental botany*, 2006, 57(6): 1225-1243.
- Lohse M, Drechsel O, Kahlau S, Ralph B. OrganellarGenomeDRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets[J]. *Nucleic Acids Research*, 2013, 41(W1): W575-W581.
- Lowe T M, Eddy S R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. [J]. *Nucleic acids research*, 1997, 25(5): 955-964.
- Izawa A G. The DNA Content of the Chloroplasts of *Acetabularia*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1963, 50(6):1164-1169.
- Manos P S, Cannon C H, Oh S H. Phylogenetic relationships and taxonomic status of the paleoendemic *Fagaceae* of western North America: recognition of a new genus, *Notholithocarpus*[J]. *Madrono*, 2008, 55(3): 181-190.
- Manos P S, Doyle J J, Nixon K C. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (*Fagaceae*)[J]. *Molecular Phylogenetics and Evolution*, 1999, 12(3): 333-349.
- 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.

- Marechal A, Brisson N. Recombination and the maintenance of plant organelle genome stability[J]. *New Phytologist*, 2010, 186(2): 299-317.
- Martin M, Casano L M, Sabater B. Identification of the product of *ndhA* gene as a thylakoid protein synthesized in response to photooxidative treatment. [J]. *Plant & Cell Physiology*, 1996, 37(3):293.
- Martin W, Kowallik K V. Annotated English translation of Mereschkowsky's 1905 paper 'über Natur und Ursprung der Chromatophoren im Pflanzenreiche' [J]. *European Journal of Phycology*, 1999, 34(3): 287-295.
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus[J]. *Proceedings of the National Academy of Sciences*, 2002, 99(19): 12246-12251.
- McVay J D, Hipp A L, Manos P S. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 2017, 284: 0300.
- Meyer E, Wiegand P, Rand SP, Kuhlmann D, Brack M, Brinkmann B. Microsatellite polymorphisms reveal phylogenetic relationships in primates. *Journal of Molecular Evolution*, 1995, 41:10–14
- Nixon K C. Infrageneric classification of *Quercus* (Fagaceae) and typification of sectional names. *Annals of Forest Science*, 1993, 50:25s–34s.
- Oh S H, Manos P S. Molecular phylogenetics and cupule evolution in *Fagaceae* as inferred from nuclear CRABS CLAW sequences[J]. *Taxon*, 2008, 57(2): 434-451.
- Palmer J D. Plastid chromosomes: structure and evolution[J]. *The Molecular Biology of Plastids*, 1991, 7: 5-53.
- Pang X, Liu H, Wu S, Yuan Y, Li H, Dong J, Liu Z, An C, Su Z, Li B. Species Identification of Oaks (*Quercus* L. *Fagaceae*) from Gene to Genome[J]. *International Journal of Molecular Sciences*, 2019, 20(23):5940.
- Perry A S, Wolfe K H. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat[J]. *Journal of Molecular Evolution*, 2002, 55(5): 501-508.
- Pombert JF, Lemieux C, Turmel M. The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes[J]. *Bmc Biology*, 2006,

- 4(1):3-3.
- Posada D, Buckley T R. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests[J]. *Systematic Biology*, 2004, 53(5): 793-808.
- Pu C X, Zhou Z K, Luo Y. A cladistic analysis of *Quercus* (*Fagaceae*) in China based on leaf epidermis and architecture. *Acta Botanica Yunnanica*, 2002, 24: 689–698.
- Shaaltiel Y, Glazer A, Bocion P F, *et al.* Cross tolerance to herbicidal and environmental oxidants of plant biotypes tolerant to paraquat, sulfur dioxide, and ozone[J]. *Pesticide Biochemistry & Physiology*, 1988, 31(1):13-23.
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Sugiura M. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression[J]. *The EMBO Journal*, 1986, 5(9): 2043-2049.
- Simeone M.C, Piredda R, Papini A, Vessella F, Schirone B. Application of plastid and nuclear markers to DNA barcoding of Euro-Mediterranean oaks (*Quercus*, *Fagaceae*): Problems, prospects and phylogenetic implications. *Botanical Journal of the Linnean Society*, 2013, 172: 478–499.
- Sudhir K, Glen S, Li M, *et al.* MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms[J]. *Molecular Biology & Evolution*, 2018(6):6.
- Standley D M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. *Molecular Biology and Evolution*, 2013, 30:772.
- Storey J D, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the national academy of sciences of the united states of america*, 2003, 100(16): 9440–5.
- Taylor F. Implications and Extensions of the Serial Endosymbiosis Theory of the Origin of Eukaryotes[J]. *Taxon*, 1974, 23(2-3):229-258.
- Team C R. R: A Language and Environment for Statistical Computing. *computing*, 2015.
- Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.) [J]. *Theoretical and Applied Genetics*, 2003, 106(3): 411-422.
- Thompson J D, Gibson T J, Higgins D G. Multiple Sequence Alignment Using ClustalW and ClustalX[J]. *Current Protocols in Bioinformatics*, 2003, 00(1): 2.3.1-2.3.22.
- Umesono K, Inokuchi H, Ohyama K, Haruo O. Nucleotide sequence of *Marchantia polymorpha* chloroplast DNA: a region possibly encoding three tRNAs and three

- proteins including a homologue of *E. coli* ribosomal protein S14[J]. *Nucleic Acids Research*, 1984, 12(24): 9551-9565.
- Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*[J]. *Proceedings of the National Academy of Sciences*, 1994, 91(21): 9794-9798.
- Wolfe K H, Li W H, Sharp P M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs[J]. *Proceedings of the National Academy of Sciences*, 1987, 84(24): 9054-9058.
- Xu J, Deng M, Jiang X L, Westwood M, Song Y G, Turkington R. Phylogeography of *Quercus glauca* (*Fagaceae*), a dominant tree of East Asian subtropical evergreen forests, based on three chloroplast DNA interspace sequences. *Tree Genetics & Genomes*, 2014, 11: 805.
- Xu X T, Wang Z H, Rahbek C, Lessard J P, Fang J Y. Evolutionary history influences the effects of water-energy dynamics on oak diversity in Asia. *Journal of Biogeogr*, 2013, 40:2146–2155.
- Yan M X, Xiong Y S, Liu R B, Deng M, Song J J. The application and limitation of universal chloroplast markers in discriminating East Asian evergreen oaks. *Frontiers in Plant Science*, 2018, 9:569.
- Yang J, Vazquez L, Chen X D, Li H M, Zhang H, Liu Z L, Zhao G F. Development of chloroplast and nuclear DNA markers for Chinese oaks (*Quercus* subgenus *Quercus*) and assessment of their utility as DNA barcodes. *Frontiers in Plant Science*, 2017, 8: 816.
- Yang J, Vazquez L, Feng L, Liu Z L, Zhao G F. Climatic and soil factors shape the demographical history and genetic diversity of a deciduous oak (*Quercus liaotungensis*) in Northern China. *Frontiers in Plant Science*, 2018, 9: 1534.
- Yang Y, Tao Z, Zhu J, Zhao J, Zhao G. Characterization of the complete plastid genome of *Quercus tarokoensis*[J]. *Conservation Genetics Resources*, 2017, 10(2): 1-3.
- Yang Y, Zhou T, Qian Z, Zhao G F. Phylogenetic relationships in Chinese oaks (*Fagaceae*, *Quercus*): Evidence from plastid genome using low-coverage whole genome sequencing[J]. *Genomics*, 2021, 113(3):1438-1447.
- Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*. 2007;24(8): 1586–91.
- Yi X, Gao L, Wang B, Su Y J, Wang T. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (*Cephalotaxaceae*): evolutionary comparison of

- Cephalotaxus* chloroplast dnas and insights into the loss of inverted repeat copies in gymnosperms[J]. *Genome Biology and Evolution*, 2013, 5(4): 688-698.
- Yin K, Zhang Y, Li Y, Fang D. Different Natural Selection Pressures on the atpF Gene in Evergreen Sclerophyllous and Deciduous Oak Species: Evidence from Comparative Analysis of the Complete Chloroplast Genome of *Quercus aquifolioides* with Other Oak Species[J]. *International Journal of Molecular Sciences*, 2018, 19(4): 1042.
- Zeng Y F, Liao W J, Petit R J, Zhang D Y. Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. *Molecular Ecology*, 2011, 20: 4995–5011.
- Zeng Y F, Wang W T, Liao W J, Wang H F, Zhang D Y. Multiple glacial refugia for cool-temperate deciduous trees in northern East Asia: The Mongolian oak as a case study. *Molecular Ecology*, 2015, 24: 5676–5691.
- Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol.Biol.Evol*, 2005, 22(12): 2472–9.
- Zhang Y, Ma J, Yang B, Li R, Zhu W, Sun L, Tian J, Zhang L. The complete chloroplast genome sequence of *Taxus chinensis* var. *mairei* (*Taxaceae*): loss of an inverted repeat region and comparative analysis with related species[J]. *Gene*, 2014, 540(2): 201-209.
- Zhang Y, Xu Y, Chen H , *et al.* Comparative Genomic Analysis Reveals the Mechanism Driving the Diversification of Plastomic Structure in *Taxaceae* Species[J]. *Frontiers in Genetics*, 2020, 10: 1295.
- Zhou Z.K. Origin, phylogeny and dispersal of *Quercus* from China. *Acta Botanica Yunnanica*, 1992, 14: 227–236.
- Zhou, Z.K. The fossil history of *Quercus*. *Acta Botanica Yunnanica*, 1993, 15: 21–33. 31.
- Zhu A, Guo W, Gupta S, Fan W, Jeffrey P. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates[J]. *New Phytologist*, 2016, 209(4): 1747-1756.

附表

附表 1 反向重复序列位置信息

	重复大小(bp)	重复起始 1	重复起始 2
<i>Q. aliena</i>	32	10205	118677
	31	10201	118678
	30	10208	118677
<i>Q. yunnanensis</i>	30	72762	72764
<i>Q. aquifolioides</i>	/	/	/
<i>Q. cocciferoides</i>	/	/	/
<i>Q. serrata</i>	30	72718	72720
<i>Q. spinosa</i>	32	55700	118677
<i>Q. baronii</i>	34	6855	11896
	30	6946	8544
	/	/	/
<i>Q. engleriana</i>	/	/	/
<i>Q. fabri</i>	30	72700	72702
<i>Q. griffithii</i>	30	72762	72764
<i>Q. mongolica</i>	31	14370	14374
	30	72754	72756
<i>Q. monimotricha</i>	/	/	/
<i>Q. rehderiana</i>	34	6864	11798
<i>Q. semicarpifolia</i>	/	/	/
<i>Q. senescens</i>	/	/	/
<i>Q. serrata</i> var. <i>brevipetiolata</i>	30	72718	72720
<i>Q. stewardii</i>	41	72703	72705
	33	72702	72711
	32	72712	72714
	30	72703	72716

附表 2 正向重复序列位置信息

	重复大小(bp)	重复起始 1	重复起始 2
<i>Q. aliena</i>	40	90840	90860
	40	160279	160299
	37	6788	6815
	39	48479	104447
	40	104445	126929
	30	113388	113419
	30	137730	137761
	30	43428	45652
	30	48491	104459
	30	104455	126939
30	115641	135508	

附表

	32	95047	95068
	32	156079	156100
	30	9464	39999
	30	11301	41176
	30	120270	120296
<i>Q. yunnanensis</i>	40	54561	54580
	39	48757	104804
	31	47471	47500
	40	104802	127345
	30	113744	113775
	30	138152	138183
	32	130	157
	30	43674	45898
	30	48769	104816
	30	64857	64882
	30	104812	127355
	30	115996	135931
	32	95405	95426
	32	156499	156520
	30	9517	40084
	30	11340	41422
	30	116436	135499
<i>Q. aquifolioides</i>	40	90399	90419
	40	159750	159770
	39	47950	104021
	40	104019	126369
	30	112960	112991
	30	137188	137219
	35	63472	63497
	30	42899	45123
	30	47962	104033
	30	104029	126379
	30	115215	134964
	32	94618	94639
	32	155538	155559
	30	9402	39580
<i>Q. cocciferoides</i>	185	145975	146901
	51	77216	77262
	39	48512	104644
	40	104642	127117
	30	113586	113617
	30	137932	137963
	31	65914	65941
	30	43461	45685

	30	48524	104656
	30	104652	127127
	32	95245	95266
	32	156466	156487
	30	4416	24622
	30	9426	39879
<i>Q. serrata</i>	39	48729	104772
	31	47443	47472
	40	104770	127268
	30	113712	113743
	30	138073	138104
	30	43646	45870
	30	48741	104784
	30	64827	64852
	30	104780	127278
	30	115964	135852
	32	95373	95394
	32	156420	156441
	30	9498	40056
	30	11322	41393
	30	116404	135420
<i>Q. spinosa</i>	40	90953	90973
	40	160184	160204
	39	48510	104573
	30	113510	113541
	30	137625	137656
	33	104571	126989
	30	43455	45679
	30	115763	135403
	32	95172	95193
	32	155972	155993
	30	11315	41203
	30	48522	104585
	30	104581	126999
<i>Q. baronii</i>	39	35860	35895
	39	48542	104587
	40	104585	127092
	30	4794	4822
	30	17619	17633
	30	113526	113557
	30	137910	137941
	30	43490	45714
	30	48554	104599

附表

	30	104595	127102
	30	115779	135688
	32	95188	95209
	32	156256	156277
	30	9527	39887
<i>Q. engleriana</i>	49	78455	78474
	39	48696	104823
	31	60053	60083
	40	104821	127287
	30	78455	78493
	30	113763	113794
	30	138141	138172
	30	43644	45868
	30	48708	104835
	30	104831	127297
	30	116017	135918
	32	95424	95445
	32	156488	156509
	30	9458	40082
	30	11306	41391
<i>Q. fabri</i>	39	48722	104731
	31	47437	47466
	40	104729	127233
	30	113665	113696
	30	138040	138071
	30	43640	45864
	30	48734	104743
	30	64806	64831
	30	104739	127243
	30	115917	135819
	32	95332	95353
	32	156381	156402
	30	9480	40052
	30	11303	41387
	30	116357	135387
<i>Q. griffithii</i>	40	54561	54580
	39	48757	104804
	31	47471	47500
	40	104802	127346
	30	113744	113775
	30	138153	138184
	32	130	157
	30	43674	45898
	30	48769	104816

	30	64857	64882
	30	104812	127356
	30	115996	135932
	32	95405	95426
	32	156500	156521
	30	9517	40084
	30	11340	41422
	30	116436	135500
<i>Q. mongolica</i>	39	48765	104797
	31	47479	47508
	40	104795	127316
	30	113737	113768
	30	138124	138155
	32	130	157
	30	43682	45906
	30	48777	104809
	30	64847	64872
	30	104805	127326
	30	115989	135903
	32	95398	95419
	32	156471	156492
	30	9517	40091
	30	11340	41430
	30	116429	135471
<i>Q. monimotricha</i>	51	11844	11885
	40	104844	127315
	39	48775	104846
	31	39940	39968
	30	113777	113808
	30	138129	138160
	30	43724	45948
	30	104854	127325
	30	116031	135906
	32	95447	95468
	32	156467	156488
	30	9496	40130
	30	11328	41471
	30	48787	104858
<i>Q. rehderiana</i>	40	91130	91150
	40	160666	160686
	39	48615	104748
	31	119556	119584
	40	104746	127267

附表

	30	113688	113719
	30	138107	138138
	30	4421	4422
	30	43542	45766
	30	48627	104760
	30	104756	127277
	30	115942	135884
	32	95349	95370
	32	156454	156475
	30	4946	4965
	30	9488	39957
	30	11240	41286
<i>Q. semicarpifolia</i>	51	11846	11887
	40	104846	127317
	39	48777	104848
	31	39942	39970
	30	113779	113810
	30	138131	138162
	30	43726	45950
	30	104856	127327
	30	116033	135908
	32	95449	95470
	32	156469	156490
	30	9498	40132
	30	11330	41473
	30	48789	104860
<i>Q. senescens</i>	39	48721	104854
	40	104852	127346
	31	29942	29969
	30	113794	113825
	30	138210	138241
	30	43669	45893
	30	48733	104866
	30	104862	127356
	30	116048	135987
	32	95455	95476
	32	156557	156578
	31	76826	119103
	30	9481	40084
	30	11335	41417
	30	24761	24762
<i>Q. serrata. var. brevipetiolata</i>	39	48729	104772
	31	47443	47472
	40	104770	127268

	30	113712	113743
	30	138073	138104
	30	43646	45870
	30	48741	104784
	30	64827	64852
	30	104780	127278
	30	115964	135852
	32	95373	95394
	32	156420	156441
	30	9497	40056
	30	11321	41393
	30	116404	135420
<i>Q. stewardii</i>	33	72703	72714
	39	48724	104765
	31	47438	47467
	40	104763	127264
	30	113705	113736
	30	138069	138100
	30	43641	45865
	30	48736	104777
	30	64806	64831
	30	104773	127274
	30	115957	135848
	32	95366	95387
	32	156416	156437
	30	9496	40052
	30	11321	41389
	30	116397	135416

附表 3 回文重复序列位置信息

	重复大小(bp)	重复起始 1	重复起始 2
<i>Q. aliena</i>	25841	90258	135080
	58	123654	123654
	44	80111	80111
	40	90840	160279
	40	90860	160299
	48	6785	6807
	38	14956	14956
	45	36452	36452
	37	6815	6815
	39	48479	146693
	40	126929	146694
	34	36439	36476

附表

	37	6788	6788
	34	134325	134325
	30	9467	50253
	30	113388	137730
	30	113419	137761
	30	10216	118677
	30	48491	146690
	30	115641	115641
	30	126939	146694
	30	135508	135508
	32	95047	156079
	32	95068	156100
	31	62199	62251
	30	40002	50253
<i>Q. yunnanensis</i>	25826	90612	135519
	71	36521	36521
	60	124031	124031
	44	80498	80498
	38	14994	14994
	39	48757	147114
	40	127345	147115
	36	72760	72760
	34	134748	134748
	30	9520	50531
	30	113744	138152
	30	113775	138183
	30	48769	147111
	30	115996	115996
	30	127355	147115
	30	135931	135931
	32	95405	156499
	32	95426	156520
	31	62513	62565
	30	40087	50531
<i>Q. aquifolioides</i>	25850	89808	134551
	64	123106	123106
	44	79651	79651
	40	90399	159750
	40	90419	159770
	38	14690	14690
	34	5385	5457
	34	133796	133796
	40	118657	118657
	39	47950	146149

	40	126369	146150
	39	36036	36036
	37	6794	6794
	30	9405	49729
	30	112960	137188
	30	112991	137219
	30	47962	146146
	30	72536	72536
	30	115215	115215
	30	126379	146150
	30	134964	134964
	32	61678	61731
	32	94618	155538
	32	94639	155559
	30	39583	49729
<i>Q. cocciferoides</i>	25868	90440	135271
	185	105419	146901
	56	123824	123824
	44	80306	80306
	38	14818	14818
	34	134507	134507
	39	36338	36338
	34	48512	147086
	36	127117	147086
	30	9429	50288
	30	113586	137932
	30	113617	137963
	32	62272	62325
	32	95245	156466
	32	95266	156487
	30	39882	50288
<i>Q. serrata</i>	25826	90580	135440
	71	36495	36495
	60	123967	123967
	44	80459	80459
	38	14966	14966
	39	48729	147035
	40	127268	147036
	36	72716	72716
	34	134669	134669
	30	9501	50503
	30	113712	138073
	30	113743	138104

附表

	30	48741	147032
	30	115964	115964
	30	127278	147036
	30	135852	135852
	32	95373	156420
	32	95394	156441
	31	62483	62535
	30	40059	50503
<i>Q. spinosa</i>	25861	90371	134964
	56	123710	123710
	40	90953	160184
	40	90973	160204
	38	14945	14945
	44	80214	80214
	34	134220	134220
	39	48510	146584
	30	62265	62306
	30	9481	50301
	30	113510	137625
	30	113541	137656
	35	36312	36312
	33	126989	146592
	30	34802	34802
	30	115763	115763
	30	135403	135403
	32	95172	155972
	32	95193	155993
	30	48522	146581
	30	126999	146585
<i>Q. baronii</i>	25843	90383	135271
	60	123782	123782
	44	80247	80247
	38	15006	15006
	34	134505	134505
	39	48542	146871
	40	127092	146872
	39	36344	36344
	30	120195	120195
	30	9530	50338
	30	113526	137910
	30	113557	137941
	30	48554	146868
	30	115779	115779
	30	127102	146872

不同叶习性栎属物种间的叶绿体基因组比较研究

	30	135688	135688
	32	62331	62384
	32	95188	156256
	32	95209	156277
	31	127973	127973
	30	39890	50338
<i>Q. engleriana</i>	25857	90618	135490
	62	124011	124011
	44	80485	80485
	38	14955	14955
	39	48696	147103
	40	127287	147104
	39	36540	36540
	34	134711	134711
	30	9461	50478
	30	113763	138141
	30	113794	138172
	30	48708	147100
	30	116017	116017
	30	127297	147104
	30	135918	135918
	32	62489	62542
	32	95424	156488
	32	95445	156509
	30	40085	50478
<i>Q. fabri</i>	25820	90539	135407
	71	36488	36488
	60	123934	123934
	44	80435	80435
	38	14960	14960
	39	48722	146996
	40	127233	146997
	36	72698	72698
	34	134636	134636
	30	9483	50495
	30	113665	138040
	30	113696	138071
	30	48734	146993
	30	115917	115917
	30	127243	146997
	30	135819	135819
	32	95332	156381
	32	95353	156402

附表

	31	62462	62514
	30	40055	50495
<i>Q. griffithii</i>	25826	90612	135520
	71	36521	36521
	60	124032	124032
	44	80498	80498
	38	14994	14994
	39	48757	147115
	40	127346	147116
	36	72760	72760
	34	134749	134749
	30	9520	50531
	30	113744	138153
	30	113775	138184
	30	48769	147112
	30	115996	115996
	30	127356	147116
	30	135932	135932
	32	95405	156500
	32	95426	156521
	31	62513	62565
	30	40087	50531
<i>Q. mongolica</i>	25826	90605	135491
	71	36528	36528
	60	124003	124003
	44	80491	80491
	38	15001	15001
	39	48765	147086
	40	127316	147087
	36	72752	72752
	34	134720	134720
	30	9520	50539
	30	113737	138124
	30	113768	138155
	30	48777	147083
	30	115989	115989
	30	127326	147087
	30	135903	135903
	32	95398	156471
	32	95419	156492
	31	62503	62555
	30	40094	50539
<i>Q. monimotricha</i>	25846	90642	135479
	56	124016	124016

	38	15064	15064
	44	80516	80516
	34	134711	134711
	40	127315	147083
	39	36563	36563
	39	48775	147082
	30	9499	50560
	30	113777	138129
	30	113808	138160
	30	34912	34912
	30	116031	116031
	30	127325	147083
	30	135906	135906
	32	62551	62603
	32	95447	156467
	32	95468	156488
	30	40133	50560
	30	48787	147079
<i>Q. rehderiana</i>	25668	90524	135664
	215	116196	135451
	56	123969	123969
	44	80380	80380
	38	14907	14907
	40	91130	160666
	40	91150	160686
	34	134683	134683
	39	48615	147069
	40	127267	147070
	39	36429	36429
	30	9491	50389
	30	113688	138107
	30	113719	138138
	30	48627	147066
	30	115942	115942
	30	127277	147070
	30	135884	135884
	32	62385	62438
	32	95349	156454
	32	95370	156475
	30	39960	50389
<i>Q. semicarpifolia</i>	25845	90645	135481
	56	124018	124018
	38	15066	15066

附表

	44	80518	80518
	34	134713	134713
	40	127317	147085
	39	36565	36565
	39	48777	147084
	30	9501	50562
	30	113779	138131
	30	113810	138162
	30	34914	34914
	30	116033	116033
	30	127327	147085
	30	135908	135908
	32	62553	62605
	32	95449	156469
	32	95470	156490
	30	40135	50562
	30	48789	147081
<i>Q. senescens</i>	25856	90650	135559
	62	124045	124045
	44	80506	80506
	38	14978	14978
	39	48721	147172
	40	127346	147173
	39	36540	36540
	34	134774	134774
	30	9484	50503
	30	113794	138210
	30	113825	138241
	30	48733	147169
	30	73373	73373
	30	116048	116048
	30	127356	147173
	30	135987	135987
	32	62532	62585
	32	95455	156557
	32	95476	156578
	30	40087	50503
<i>Q. serrata</i> var. <i>brevipetiolata</i>	25826	90580	135440
	71	36495	36495
	60	123967	123967
	44	80459	80459
	38	14965	14965
	39	48729	147035
	40	127268	147036

不同叶习性栎属物种间的叶绿体基因组比较研究

	36	72716	72716
	34	134669	134669
	30	9500	50503
	30	113712	138073
	30	113743	138104
	30	48741	147032
	30	115964	115964
	30	127278	147036
	30	135852	135852
	32	95373	156420
	32	95394	156441
	31	62483	62535
	30	40059	50503
<i>Q. stewardii</i>	25826	90573	135436
	71	36493	36493
	60	123964	123964
	44	80451	80451
	39	67371	67439
	38	14965	14965
	47	72701	72701
	39	48724	147031
	40	127264	147032
	39	72703	72703
	34	72702	72702
	34	134665	134665
	30	9499	50498
	30	113705	138069
	30	113736	138100
	32	72712	72712
	32	72714	72714
	30	48736	147028
	30	115957	115957
	30	127274	147032
	30	135848	135848
	32	95366	156416
	32	95387	156437
	31	62462	62514
	30	40055	50498

附表

附表 4 互补重复序列位置信息

	重复大小(bp)	重复起始 1	重复起始 2
<i>Q. aliena</i>	31	10210	10210
	31	11777	11777
	31	13455	34155
<i>Q. yunnanensis</i>	30	72762	72762
	32	13494	34185
	31	48435	55940
<i>Q. aquifolioides</i>	/	/	/
<i>Q. cocciferoides</i>	33	13328	34008
<i>Q. serrata</i>	30	72718	72718
	31	11798	11798
	31	76780	76780
	31	76783	76783
	31	13476	34159
<i>Q. spinosa</i>	30	48408	55909
	31	11790	11790
	30	13475	34145
<i>Q. baronii</i>	34	13516	33991
	30	48221	55742
<i>Q. engleriana</i>	31	11782	11782
	37	78464	78472
	37	78472	78483
	30	127595	127596
	35	78455	78464
	35	78455	78483
<i>Q. fabri</i>	32	76774	127592
	30	72700	72700
	31	11779	11779
	31	116663	116663
<i>Q. griffithii</i>	31	13457	34153
	30	72762	72762
	32	13494	34185
	31	48435	55940
<i>Q. mongolica</i>	30	72754	72754
	30	14378	14378
	32	13494	34192
	31	48443	55930
<i>Q. monimotricha</i>	31	11804	11804
	32	13566	34250
	30	48454	55966
<i>Q. rehderiana</i>	31	11740	11740
	32	6824	6824
	33	13418	34106

不同叶习性栎属物种间的叶绿体基因组比较研究

	30	130628	130628
<i>Q. semicarpifolia</i>	31	11806	11806
	32	13568	34252
	30	48456	55968
<i>Q. senescens</i>	31	11811	11811
	33	13489	34182
	30	4911	4911
<i>Q. serrata</i> var. <i>brevipetiolata</i>	30	72718	72718
	31	11797	11797
	31	76780	76780
	31	76783	76783
	31	13475	34159
	30	48408	55909
<i>Q. stewardii</i>	41	72703	72703
	36	72711	72711
	30	72703	72703
	31	11797	11797
	31	13475	34158
	30	48403	55888

个人简介

许阳,男, 1996 年 5 月生, 籍贯江西宜春。2014 年 9 月至 2018 年 7 月就读于江西农业大学林学院林学专业, 并于 2018 年 7 月取得农学学士学位。2018 年 9 月至 2021 年 6 月就读于北京林业大学生态与自然保护学院生态学, 师承杜芳教授, 研究方向为恢复生态学。

硕士阶段主修了: 分子生态学理论方法、高级森林生态学、生态学研究方法、景观生态学, 多元统计分析等课程。完成了硕士期间全部的课程学习, 总学分 26 分。硕士期间的主要研究内容是栎属的叶绿体基因组比较及其选择作用分析, 并以第一作者发表 SCI 论文一篇, 以第二作者发表 SCI 论文一篇。

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。以第一作者或通讯作者发表SCI论文21篇，发明专利1项，论文引用700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；2018年起担任国际栎树协会专业委员会成员和北京生态修复协会理事；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 & Evolution的Associate Editor。

成果目录清单

1. **Xu Y**, Chen H, Qi M, *et al.* The complete chloroplast genome of *Quercus fabri* (Fagaceae) from China [J]. *Mitochondrial DNA Part B*, 2019, 4(2): 2857-2858.
2. Zhang Y, **Xu Y**, Chen H, *et al.* Comparative Genomic Analysis Reveals the Mechanism Driving the Diversification of Plastomic Structure in Taxaceae Species[J]. *Frontiers in Genetics*, 2020, 10:1295.

致谢

落笔至此，意味着我在北京林业大学大学三年的硕士研究生生涯即将结束。2018 年的那个夏天，我拎着一个行李箱只身北上求学，2021 年的这个夏天，我即将离开。回首望去，三年的岁月如同一场难忘的旅行，好的坏的都是我人生中弥足珍贵的风景。纵有万般不舍，但我仍旧心存感激。

杜芳教授是我来到北林之后第一个见到的人，她成为了我的导师，也是我学术和人生的引路人。早在我入学前，杜老师就引荐我到江西农大寇一翱老师那里学习，寇老师从看文献开始教我，让我明白阅读文献在学术科研中的重要性。还记得第一次见到杜老师的时候，很难想象一个如此年轻的女老师居然要管理一个十多人的实验室，更何况每个人的进度和研究内容都不一样，但是她做到了。在我翻看实验记录本的时候发现在第十来页的地方写着“Oak of China”，写这几个单词的时候正是杜老师兴奋地告诉我们最新的林调结果显示栎树是中国天然林占比最高的物种，我们研究非常有意义。当时的我想着，一定要跟着杜老师做出成果来。值得庆幸的是，在我研三这一年我们的网站 www.oakofchina.org 终于问世了，虽然还有很多地方不完善但这一定是非常有意义的一步。我是一个十分不让人省心的人，这三年里没少让杜老师操心，但是她从来没有放弃我，一遍一遍地指导我的工作，指出我的不足。为了让我学会认植物，亲自带我去中国科学院植物研究所北京植物园认识了刘冰老师；为了让我学会整理采样信息和照片，带我认识了台湾师范大学的廖培鈞老师和杨志凯师兄；为了让我学会栎树扦插和种植，带我认识了中国林业科学研究院林业科学研究所的李迎超老师和钱家连师妹；为了学 call SNP 的方法，带我认识了北京林业大学的毛建丰老师，在这里感谢以上提到的老师和同学，感谢您们对我学术和科研上的指导及帮助，感谢！

另外还有两位老师对我的研究起到了非常关键的帮助，他们是北京林业大学的尹康权老师和中国科学院西双版纳植物园的马会老师。还记得和尹老师在微生物所上传数据直到深夜，还记得马老师凌晨还加班给我检查分析的结果，讲解分析的原理，从数据分析到论文的写作和修改，两位老师的巨大付出让我看到了他们对于学术的严谨态度和诲人不倦的教师精神，桃李不言下自成蹊。

感谢同窗的各位同学。感谢实验室各位同学三年来的朝夕相伴：感谢祁敏和我一起采样，帮助我提取 DNA；感谢刘可可可在外业时对我的帮助，在我受伤时对我悉心照顾；感谢张艺在我论文撰写时帮我检查格式，查找错误；感谢杨霖和我一起探讨分析流程和服务器的使用；感谢苏蔚和我一起琢磨栎树的种植方法，和我共同建设网站，尤其是在网站搭建初期付出了大量努力；感谢罗艳君对我生活上的帮助；感谢康艺馨和我一起探讨栎类的分类；感谢韦羽茜在学习生活上对

致谢

我的帮助；感谢已经毕业的师兄师姐：王玉垚、王天瑞、王瑞、刘媛、张悦、苏金源、燕语、李月娟，他们是我学术和生活的榜样。尤其感谢张悦师姐，是她一点一点教会我各种分析，教会我基因组学的分析必须要细致细致再细致。是师姐把我领进了叶绿体基因组的门，没有她就不会有我的各种成果，更没有现在这篇毕业论文，感谢！还要感谢我的室友：刘正霄、吴鑫磊、黄睿智、王鑫和韩尚君，我总是工作到很晚回宿舍，大家都已经睡了，但是大家对我的打扰却没有一句抱怨，感谢！

感谢我的家人。二十多来一直有各位亲人给予我支持和帮助，做我最坚强的后盾，支撑着我砥砺前行。回想起三年研究生一共只回过两次家，那年在江西采样最近的时候离家只有十几公里，但是因为自己是在工作，也没能回家见家人一面。养育之恩当涌泉相报，希望未来的我能打拼出一番事业去回报我的家人对我的养育之恩。

最后，感谢自己。如果不是自己三年前虽然是笔试最后一名但还是执意来北京面试，就不会遇到这么多对自己有知遇之恩的人，就不会发生这么多令人难忘的事，也不会有现在的我。感谢自己一路以来的坚持，不悔梦归处，只恨太匆匆。

感谢所有的相遇，聚散终有时，未来方可期，山水一程，三生有幸。北林的各位，我们更高处再见。