

分类号

密级

UDC



北京林业大学

学术型硕士学位论文

常绿和落叶栎类叶绿体基因功能研究

Chloroplast genome function on evergreen and deciduous oaks

原明星

指导教师 杜芳 教授

学院 生态与自然保护学院

学科专业 生态学

研究方向 修复生态学

二〇二四年 六 月 四 日

# 独创性声明

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

签 名： 原明星 日 期： 2024年06月04日

# 关于学位论文使用授权的声明

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

签 名： 原明星 导师签名： 杜芳 日 期： 2024年06月04日

## 答 辩 委 员 会 成 员 信 息

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

## 摘要

栎属植物包含常绿和落叶的叶习性并表现出不同的生存策略。一般来说落叶植物表现出更强的获取性策略，有较高的光合速率，以高养分含量和高累积量为补偿，降低叶片寿命；常绿植物表现出较为保守的生存策略，具有更长叶片寿命，有可能在冬季温度允许的范围内进行光合作用，增加光合作用的时间，弥补较低的光合速率。植物进行光合作用的主要场所是叶绿体，许多光合作用所必须的蛋白质也都由叶绿体基因组所编码，这些蛋白质的结构和功能可能会影响不同叶习性植物的光合作用效率。但对于常绿和落叶植物叶绿体基因组的差异、特定的叶绿体基因如何影响光合效率以及是否有基因通过影响叶绿体的结构和功能来调节光合效率等问题的相关研究甚少。

栎属植物在北半球有丰富的物种多样性，适应性强，是重要的生态树种。本研究首先对 14 种栎属物种，包括七种常绿栎：檀子栎(*Quercus baronii* Skan)、铁橡栎(*Q. coccoferoides* Hand.-Mazz.)、巴东栎(*Q. engleriana* Seem.)、矮高山栎(*Q. monimotricha* Hand.-Mazz.)、高山栎(*Q. semecarpifolia* Sm.)、灰背栎(*Q. senescens* Hand.-Mazz.)、毛脉高山栎(*Q. rehderiana* Hand.-Mazz.)和七种落叶栎：白栎(*Q. fabri* Hance)、大叶栎(*Q. griffithii* Hook. f. & Thoms. ex Miq.)、蒙古栎(*Q. mongolica* Fisch. ex Ledeb.)、枹栎(*Q. serrata* Thunb.)、短柄枹栎(*Quercus serrata* var. *brevipetiolata* (A.DC.) Nakai)、黄山栎(*Q. dentata* subsp. *stewardii* (Rehder) A.Camus)和云南波罗栎(*Q. yunnanensis* Franch.)的叶绿体基因组基本特征进行比较分析。发现 14 种栎属植物的叶绿体基因组变异主要集中在非编码区，反向重复区的变异最小并且不同叶习性的栎属物种之间存在显著差异，有多处变异是同种叶习性的栎树所共有的。结合美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)中公开的 41 个栎属物种的叶绿体基因组序列，通过最大似然法和邻接法构建了系统发育树。利用 PAML (Phylogenetic Analysis by Maximum Likelihood) 软件对 72 个共有叶绿体基因进行正选择检测，位点模型和分支位点模型均证明 *ycf1* 基因为正选择基因，因此本文着重对该基因功能进行研究。研究发现 *ycf1* 基因部分正选择序列在落叶栎属物种中均为谷氨酰胺 (Gln, Q)、甘氨酸 (Gly, G)、和 3 个苯丙氨酸 (Phe, F) 组成，即 QGFFF，而在常绿栎属物种中除 QGFFF 外，还存在精氨酸 (Arg, R)、赖氨酸 (Lys, K)、天冬酰

胺 (Asn, N)、亮氨酸 (Leu, L) 和缬氨酸 (Val, V) 组成的氨基酸序列, 即包括 QGFFF 和 RKNLV 两种类型的氨基酸序列。利用 Alphafold2 软件对该段正选择位点附近的氨基酸序列进行三维结构预测, 发现常绿和落叶栎属物种该段序列在结构上存在显著差异, 可能会影响该基因编码蛋白的功能。

在拟南芥中的研究已经证明 *ycf1* 基因编码的 Tic214 蛋白是叶绿体外膜转运子 (translocon at the outer chloroplast membrane, TOC) 和叶绿体内膜转运子 (translocon at the inner chloroplast membrane, TIC) 复合体的组成成分, 且是该复合体唯一由叶绿体编码的亚基, 在植物细胞核编码的蛋白质运输到叶绿体的过程中起重要作用。进入叶绿体的前体蛋白具有 N 端转运肽, 帮助蛋白识别和转运, 可能与 TOC-TIC 复合体存在相互作用。因此本论文利用酵母双杂实验验证不同叶习性栎属物种转运肽与 Tic214 蛋白的部分序列之间的相互作用, 发现落叶栎属物种对应的氨基酸序列 QGFFF 的相互作用更强, 而常绿栎属物种对应的氨基酸序列 RKNLV 则相互作用相对较弱。

基于以上实验证据本论文提出了可能影响不同叶习性栎属物种光合作用效率的模型: 落叶和常绿栎 *ycf1* 基因氨基酸序列的差异可能使进入叶绿体的蛋白质与特定蛋白 Tic214 的相互作用强度发生变化, 导致不同叶习性栎树在相同时间内进入叶绿体的蛋白质的数量不同, 从而使得光合作用速率不同。丰富了非模式物种中叶绿体基因功能的研究。

**关键词:** 栎树, 叶绿体基因组, 比较分析, 正选择, 基因功能研究, 酵母双杂

# Chloroplast genome function on evergreen and deciduous oaks

Master Candidate Mingxing Yuan

(Ecology)

Directed by Prof. Dr. Fang Du

## Abstract

The genus *Quercus*, contain evergreen and deciduous leaf habits, exhibiting different survival strategies. Generally, deciduous plants exhibit stronger acquisition strategies, higher photosynthetic rates, compensating for lower leaf life with high nutrient content and accumulation. Evergreen plants exhibit a more conservative survival strategy, own longer leaf life, and can carry out photosynthesis within the range of winter temperatures, increasing the duration of photosynthesis and compensating for lower photosynthetic rates. Chloroplasts are the main sites of photosynthesis in plants. The chloroplast genome encodes many proteins necessary for photosynthesis, which can affect the photosynthetic efficiency of plants with different leaf habits. However, there is a lack of research on the differences in chloroplast genomes between evergreen and deciduous species, for example how chloroplast genes affect photosynthetic efficiency, and whether there are genes regulate photosynthetic efficiency by affecting the chloroplast structure and function. Additionally, *Quercus* species are rich in species diversity and strong in adaptability in the entire Northern Hemisphere. They are important ecological tree species. In this study, we firstly compared the basic characteristics of chloroplast genomes of 14 *Quercus* species, including seven evergreen oak species: *Quercus baronii* Skan, *Q. cocciferoides* Hand.-Mazz., *Q. engleriana* Seem., *Q. monimotricha* Hand.-Mazz., *Q. semecarpifolia* Sm., *Q. senescens* Hand.-Mazz., and *Q. rehderiana* Hand.-Mazz., and seven deciduous oak species: *Q. fabri* Hance, *Q. griffithii* Hook. f. & Thoms. ex Miq., *Q. mongolica* Fisch. ex Ledeb., *Q. serrata* Thunb., *Q. serrata* var. *brevipetiolata* (A.DC.) Nakai, *Q. dentata* subsp. *stewardii* (Rehder) A.Camus and *Q. yunnanensis* Franch.et. It was found that the chloroplast genome variations of the 14 *Quercus* species were mainly concentrated in the non-coding region, and the reverse repeat region had the least variation. There were significant differences between *Quercus* species with different leaf habits, and many variations were common to oaks with the same leaf

habit. Combined with the chloroplast genome sequences of 41 *Quercus* species published in the National Center for Biotechnology Information, a phylogenetic tree was constructed by maximum likelihood and neighbour joining method. We used PAML software detect positive selection in 72 common chloroplast genes by site and branch site model. The results showed a positive selection gene, *ycf1*. It was found that the positive selection sequence of *ycf1* gene in deciduous species is composed of glutamine (Gln, Q), glycine (Gly, G), and three phenylalanine (Phe, F), namely QGFFF. In evergreen species, there are two sequence types, one is QGFFF, the other composed of arginine (Arg, R), lysine (Lys, K), asparagine (Asn, N), leucine (Leu, L) and valine (Val, V), RKNLV. AlphaFold2 software was used to predict the tertiary structure of the amino acid sequence attached to the positive selection site. It was found that there are significant differences in the structure of the sequence in evergreen and deciduous oak species, Tic214 protein encoded by *ycf1* gene is a component of the plant protein transport channel complex translocon at the outer chloroplast membrane (TOC)-translocon at the inner chloroplast membrane (TIC) complex, and it is the only subunit encoded by chloroplast as suggested by studies in model plant *Arabidopsis thaliana*., N-terminal transport peptides are required for the translocation of preproteins into the chloroplast, helping protein recognition and transport, and may interact with TOC-TIC complex. Therefore, this study used yeast two-hybrid system to verify the interaction between transport peptides and sequences of Tic214 protein. It was found that the amino acid sequence in deciduous oaks, QGFFF had stronger interaction, while the amino acid in evergreen oaks, RKNLV had relatively weaker interaction. Based on the above experimental evidence, a model affect the photosynthesis efficiency of different leaf habits of oak species was proposed: the difference in amino acid sequences of *ycf1* gene in deciduous and evergreen oak species may change the interaction strength between proteins entering chloroplasts and specific protein Tic214, resulting in different numbers of proteins entering chloroplasts at the same time in different leaf habits of oak species, and thus different photosynthesis rates. This study enriches the study of chloroplast gene function in non-model species.

**Key words:** oaks, chloroplast genome, comparative analysis, positive selection, gene function studies, yeast two-hybrids

## 目录

1 引言 .....	1
1.1 叶绿体基因组学研究.....	1
1.1.1 叶绿体功能与起源.....	1
1.1.2 叶绿体基因组结构及特征.....	1
1.1.3 蛋白互作研究进展.....	2
1.2 栎属植物叶绿体基因组学研究.....	3
1.2.1 栎属常绿和落叶植物特征.....	3
1.2.2 栎属植物叶绿体基因组学研究.....	4
1.3 本文的研究内容, 研究目的及意义.....	6
1.3.1 研究内容 .....	6
1.3.2 研究目的及意义.....	6
2 不同叶习性栎属物种叶绿体基因组序列比较.....	7
2.1 材料与方法 .....	7
2.1.1 叶绿体基因组数据获取.....	7
2.2 数据分析 .....	9
2.2.1 叶绿体基因组序列基本特征分析.....	9
2.2.2 重复序列与序列特异性分析.....	9
2.2.3 密码子偏好性分析.....	9
2.2.4 系统发育分析.....	10
2.2.5 选择压力分析.....	10
2.3 研究结果 .....	11
2.3.1 基本特征分析.....	11
2.3.2 重复序列分析.....	12
2.3.3 序列多样性分析.....	14
2.3.4 密码子偏好性分析.....	17
2.3.5 系统发育分析.....	19
2.3.6 选择压力分析.....	21
3 基因功能研究 .....	24
3.1 实验材料与试剂 .....	24
3.1.1 实验材料及试剂.....	24
3.1.2 实验试剂配制.....	24
3.2 研究方法 .....	25
3.2.1 生物信息学分析.....	25
3.2.2 酵母双杂实验.....	26

3.3 研究结果 .....	28
3.3.1 正选择位点氨基酸保守性分析 .....	28
3.3.2 <i>ycf1</i> 基因的生物信息学分析 .....	29
3.3.3 <i>ycf1</i> 基因功能验证 .....	39
4 讨论 .....	43
4.1 栎属叶绿体基因组序列差异比较 .....	43
4.2 栎属叶绿体基因组系统发育 .....	43
4.3 <i>ycf1</i> 基因功能验证 .....	44
5 结论和展望 .....	46
5.1 结论 .....	46
5.2 展望 .....	47
参考文献 .....	48
附录 .....	54
个人简介 .....	71
导师简介 .....	72
获得成果目录清单 .....	73
致谢 .....	74

# 1 引言

## 1.1 叶绿体基因组学研究

### 1.1.1 叶绿体功能与起源

叶绿体主要集中在叶肉细胞中是植物进行光合作用所必须的细胞器。叶绿体具有双层膜结构，外膜可以渗透小的有机分子，内膜渗透性低存在多种光合色素，含有高浓度的叶绿素，可以捕获能量将二氧化碳和水转换为糖类并释放氧气<sup>[1]</sup>。在植物免疫方面，叶绿体参与水杨酸、茉莉酸等防御分子的合成，在植物抵御病原体入侵上发挥了重要作用<sup>[2]</sup>。此外，叶绿体还是许多代谢物的合成场所，包括脂肪酸、氨基酸和植物激素等<sup>[3]</sup>。在面对环境胁迫，如干旱和高盐胁迫时叶绿体能够作为环境传感器，感知细胞外部的变化，并将信号传递给其他细胞器，从而协助植物适应和应对这些挑战<sup>[3,4]</sup>。

叶绿体虽然具有成套的遗传信息，但其中大部分基因转移到了细胞核中，只保留了 100 多个基因，包括少量 rRNA 基因和一整套 tRNA 基因<sup>[5]</sup>。叶绿体中大多数的蛋白质由细胞核编码，在细胞质中合成再运输到叶绿体中<sup>[6]</sup>。虽然叶绿体含有核糖体，能合成蛋白质，但发挥完整功能仍然需要细胞核的协助，是半自主性细胞器<sup>[7,8]</sup>，而叶绿体自身所具有的这套遗传物质被称为叶绿体基因组或叶绿体 DNA。

### 1.1.2 叶绿体基因组结构及特征

1986 年烟草的叶绿体基因组成功测序，拉开了叶绿体基因组研究的序幕<sup>[9]</sup>。截止 2023 年 NCBI (National Center for Biotechnology Information) 数据库中公布的叶绿体基因组数量接近 13 000 个。丰富的叶绿体基因组数据促进了物种鉴定、物种进化和叶绿体基因功能的研究。

大量研究表明植物的叶绿体基因组大小多在 100-217kb 之间，具有保守的四分体结构，包括大单拷贝区 (Large single copy region, LSC)、小单拷贝区 (Small single copy region, SSC) 和一对反向重复区 (Inverted repeat, IR)。但并非所有植物的叶绿体基因组都会呈现典型的四分体结构，如红豆杉科和豆科的一些植物缺失一个 IR 区<sup>[10,11]</sup>，IR 区的扩张和收缩可能是影响不同物种间叶绿体基因组大小的原因<sup>[12]</sup>。

叶绿体基因组在基因数量和基因顺序上具有保守性，编码了许多参与光合作用和其他代谢过程的蛋白质。具体来说，叶绿体基因组包含 110-130 个基因<sup>[13]</sup>，可分为四类。第一类是与转录翻译有关的基因<sup>[14]</sup>，包含核糖体小亚基，如 *rps2*；核糖体大亚基，如 *rpl2*；核糖体 RNA 基因，如 *rrn16*；RNA 聚合酶基因，如 *rpoA*；转运 RNA 基因，如 *trnC-GCA*。第二类是与光合作用有关的基因，包含 ATP 合成酶，如 *atpA* 基因；

NAD(P)H 脱氢酶基因, 如 *ndhD*; 细胞色素有关基因, 如 *petA*; 光系统 I、II 有关的基因, 如 *psaA*、*psbA* 基因; 以及 Rubisco 大亚基 *rbcL* 基因<sup>[14]</sup>。第三类是与生物合成相关的基因, 如 *accD*、*ccsA*、*matK* 等基因。第四类是一些未知功能的基因, 如 *ycf1* 基因<sup>[14,15]</sup>。

对叶绿体基因组的研究促进了叶绿体基因工程的发展。叶绿体基因工程通过对叶绿体基因进行编辑和转基因研究可改良植物的农艺性状, 增强其抗病性、抗逆性和提高产量等, 也可研发一些生物医学免疫的产品。Kumar 等<sup>[16]</sup>将甜菜中的碱脱氢酶 (BADH) 基因利用胡萝卜叶绿体基因组转化后发现酶活性和生长速度均提高, 转基因胡萝卜的耐盐水平增强。Boyhan 等<sup>[17]</sup>将霍乱毒素亚基与人胰岛素共同转化到烟草和生菜叶绿体基因组中, 得到高水平表达的胰岛素, 降低了生产成本。

### 1.1.3 蛋白互作研究进展

随着高通量测序技术的发展, 已构建出复杂的蛋白互作网络, 揭示了蛋白质如何在细胞内协同工作。同时, 利用核磁共振光谱学和冷冻电子显微镜等技术, 已经解析了许多蛋白质的三维结构, 有利于对蛋白互作更深入的研究。随着生物信息学和信息技术的发展, 通过开发各种模型和算法来预测蛋白质的相互作用, 为理解复杂的蛋白互作提供了有力的支持<sup>[18]</sup>。蛋白相互作用的实验方法包括酵母双杂 (Yeast two-hybrid system)、免疫共沉淀 (Co-Immunoprecipitation, Co-IP)、GST pull-down、荧光共振能量转移 (Förster Resonance Energy Transfer, FRET) 和等离子共振技术 (Surface Plasmon Resonance, SPR)<sup>[19]</sup>等。以上技术中酵母双杂技术应用最为广泛, 该方法无需大型设备, 且成本较低, 方便在实验室中直接进行<sup>[20-22]</sup>。酵母双杂技术通过利用转录因子 (GAL4) 的特性, 将两个蛋白分别与 DNA 结合结构域 (DNA binding domain, BD) 和转录激活结构域 (activation domain, AD) 融合, 然后将融合蛋白转入酵母细胞。如果两个蛋白存在相互作用, 它们将促使这两个融合蛋白在空间上接近, 从而可以激活报告基因, 酵母细胞生长; 反之, 则不会激活报告基因转录<sup>[20]</sup>。

利用酵母双杂技术, 科学家们已经成功鉴定出多个与叶绿体基因组表达调控相关的蛋白质。例如, Chi 等<sup>[23]</sup>利用酵母双杂交技术筛选出了与 Delayed Greening 1 (DG1) 蛋白相互作用的调控因子 sigma factor 6 (SIG6), 发现 SIG6 与 DG1 相互作用, 在调节叶绿体转录过程中起着重要作用。此外, Chen 等<sup>[24]</sup>通过酵母双杂交方法筛选了 Yellow Leaf 1 (YL1) 蛋白的互作蛋白, 并利用双分子荧光互补实验验证了 YL1 与 AtpB 蛋白间的相互作用。这些研究结果不仅揭示了 YL1 在水稻叶绿体发育中的作用, 也为叶绿体功能性基因的调控机制提供了新的视角。除了鉴定与叶绿体基因组表达调控相关的蛋白质外, 酵母双杂技术还可以用于研究叶绿体基因组的组装和复制过程<sup>[25]</sup>。

酵母双杂交技术在研究叶绿体基因组中的应用前景广阔, 有望借助此技术探讨更多与叶绿体基因组表达调控相关蛋白质的功能。这将有助于深入了解叶绿体基因组的

表达调控机制，为植物生长发育和光合作用提供理论依据。

## 1.2 栎属植物叶绿体基因组学研究

### 1.2.1 栎属常绿和落叶植物特征

栎属植物北半球广泛分布，约四百到五百种，北美和亚洲为其多样性中心<sup>[26]</sup>。栎属的分类复杂，Denk 等<sup>[26]</sup>总结的栎属植物最新分类研究将栎属分为两个亚属：栎亚属（subgenus *Quercus*）和麻栎亚属（subgenus *Cerris*）（图 1.1）。栎亚属包括白栎组（sect. *Quercus*）、红栎组（sect. *Lobatae*）、中间栎组（sect. *Protobalanus*）、活栎组（sect. *Virentes*）和本都栎组（sect. *Ponticae*）五个组。麻栎亚属包括三个组，分别是冬青栎组（sect. *Ilex*）、青冈栎组（sect. *Cyclobalanopsis*）和麻栎组（sect. *Cerris*）。栎属植物叶习性包括常绿，落叶和半常绿（图 1.1）<sup>[26,27]</sup>。常绿是指叶片寿命大于一年，叶子全年存在，如川滇高山栎（*Q. aquifolioides*）；落叶是指植物在干旱或寒冷季节开始时叶子掉落，并且每年的大部分时间都没有叶子，如槲栎（*Q. aliena*）；半常绿是指植物叶片通常全年存在或有短暂的无叶期，如檀子栎（*Q. baronii*）<sup>[28,29]</sup>。

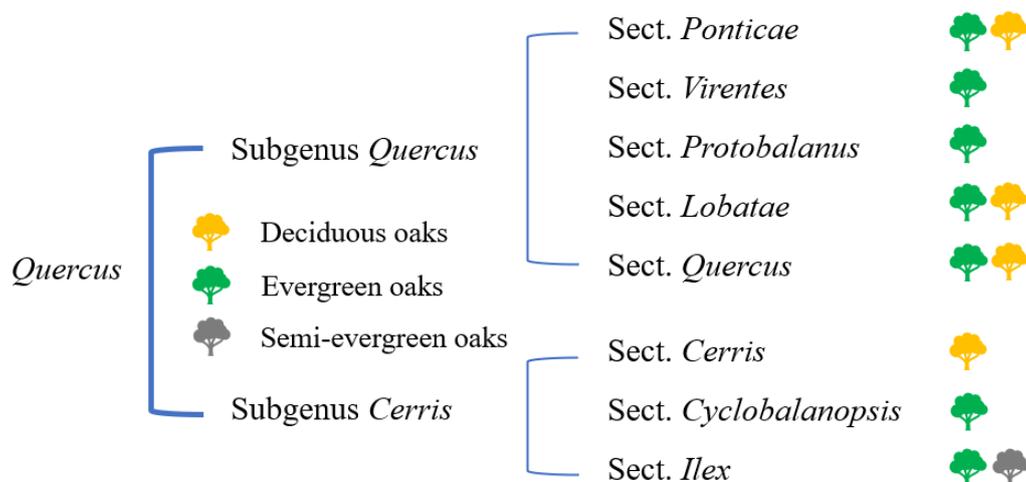


图 1.1 栎属物种不同亚属和组中栎树的叶习性

Figure 1.1 Leaf habits of *Quercus* species in different subgenera and sections

落叶植物表现出更强的获取性策略，叶片氮和磷含量较高，光合速率高，光合氮利用效率较高，以高养分含量和累积量为补偿降低叶片寿命，同时通过减少蒸腾的表面积和呼吸损失抗干旱和避免冰冻<sup>[30-34]</sup>。常绿植物则表现出较为保守的策略，其光合速率和光合氮利用效率低于落叶叶片，但叶片寿命更长，有可能在冬季温度允许的范围内进行光合作用从而增加光合作用时间，弥补光合速率<sup>[30-34]</sup>。另外，常绿植物会经历落叶所没有的代谢和生态压力，会投入更多成本来减弱不利环境带来的影响或者威慑食草动物等。因此，常绿植物叶片的初始构建成本高于同等大小的落叶植物叶片，

叶子更厚，单位面积叶片质量更大，而落叶植物则表现出更高的潜在生长率，更高的比叶面积和更高的光合速率<sup>[30-34]</sup>。半常绿植物叶片特征通常介于常绿植物和落叶植物之间，本文着重讨论常绿和落叶栎的叶绿体基因组结构和功能差异。

多位学者已经对常绿和落叶栎属植物在不同环境下的适应性进行了深入研究。Furze 等<sup>[35]</sup>的研究表明，落叶栎倾向于向非结构性碳水化合物（Non-structural carbohydrate, NSC）浓度较高的方向演化，拥有更高的淀粉浓度和比例。Kaproth 等<sup>[34]</sup>进一步指出，落叶树种展现出较强的耐旱策略，能在干旱环境中生长。相比之下，常绿树种在干旱条件下的叶片和茎枯死率较高，这主要是由于它们在叶片投资上的差异导致的碳损失。Sancho-Knapik 等<sup>[36]</sup>的研究发现，生长季节较短的落叶树种通常具有较低的比叶重（Leaf mass per area, LMA）和叶片厚度，而常绿树种则相反。这些研究结果共同揭示了常绿和落叶栎属植物在适应不同环境压力时的策略差异和形态可塑性。

### 1.2.2 栎属植物叶绿体基因组学研究

目前多个栎属物种的叶绿体基因组序列已成功上传到 NCBI 数据库中，截止到 2024 年 2 月，NCBI 中已公布 647 条栎树叶绿体基因组相关的序列，其中包括 92 个栎属物种完整的叶绿体基因组序列，附表 A 详细统计了目前 NCBI 中已有的叶绿体基因组序列及其分类和叶习性，以及存在的多个版本的 GenBank 号。

栎属物种叶绿体基因组的研究大多先对单个或几个物种完整叶绿体基因组进行组装拼接注释后，再结合栎属的多个物种进行叶绿体基因组比较，构建系统发育树，探讨物种的进化关系，找出高度分化的区域，寻求合适的分子标记，以及受到正选择的基因。但是目前缺乏对受选择基因功能的深入探究。

对植物叶绿体基因组进行组装、拼接和注释是叶绿体基因组研究的基础，通过对单个或多个植物的叶绿体基因组进行组装和注释，为下一步的分析和系统发育奠定了基础。如 Yin 等<sup>[37]</sup>利用 Illumina 平台测序、SOAPdenovo 组装、CpGAVAS 注释了川滇高山栎的叶绿体基因组，然后将川滇高山栎叶绿体基因组与两种落叶栎比较后发现 *atpF* 基因受到正选择，此外位点特异性选择分析还在 12 个基因中发现了正选择位点。Liu 等<sup>[38]</sup>同样利用 Illumina 平台测序、NOVOPlasty 组装、CpGAVAS 注释了蒙古栎、短柄栎和辽东栎三种栎的叶绿体基因组，并结合其他栎属物种比较后发现了 11 个可以作为分子标记的高度分化区域。系统发育分析对于研究栎属物种的进化关系是及其重要的，Yang 等<sup>[39]</sup>基于 54 个栎属物种的叶绿体基因组构建了栎树的系统发育关系，发现青冈栎组和麻栎组存在嵌套关系，冬青栎组在中国的分布、多样性和分子分化可能是青藏高原、横断山脉和喜马拉雅山脉隆升驱动的地理和古气候变化的局部适应的结果。Wang 等<sup>[40]</sup>同样对 4 种青冈栎组植物进行测序后鉴定出 7 个变异相对较高的区域(*rps16*、*ndhk*、*accD*、*ycf1*、*psbZ-trnG-GCC*、*rbcL-accD* 和 *rpl32-trnL-UAG*)，

构建的系统发育树支持栎属分为栎亚属和麻栎亚属，青冈栎组和麻栎组嵌套在冬青栎组当中。Wang 等<sup>[38]</sup>利用 PAML 中的位点模型检测出 11 个编码蛋白质的基因(*atpF*、*NDHA*、*ndhd*、*ndhF*、*ndhK*、*petB*、*petD*、*rbcL*、*rpl22*、*ycf1* 和 *ycf3*)受到正选择。Chen 等<sup>[41]</sup>对 4 种青冈栎组植物进行测序组装，比较分析后检测到 6 个高度可变的热点，其中 *rpoC1*、*clpP* 和 *ycf1* 可作为分子标记。此外，Chen 等<sup>[41]</sup>利用 KaKs\_Calculator 计算后发现 *petA*、*ycf2* 两个基因受到正选择。以上研究对于理解栎属物种的进化机制至关重要。

*ycf* 基因和负责植物光系统 I 电子传递的 *NDH* 基因经常在叶绿体基因组比较研究中鉴定为正选择基因。*NDH* 基因由至少 15 个亚基组成，参与光系统 I 的电子传递循环，有助于减轻叶绿体的氧化应激，对于植物在各种胁迫条件下保持正常光合作用具有重要意义<sup>[42]</sup>。但 *ycf1* 基因的功能却一直没有被明确阐述，对未知的叶绿体基因功能的探究，将极大丰富我们对栎属植物叶绿体基因组功能的探究。

*ycf1* 属于叶绿体中较大的基因之一，但常规的基因功能研究如基因敲除、敲入等实验操作可能会引起植物死亡，无法用来研究植物的 *ycf1* 基因，因此分析该基因的功能具有一定挑战。Drescher 等<sup>[43]</sup>构建突变等位基因，通过同源重组转入烟草叶绿体中后研究发现 *ycf1* 所编码的蛋白对植物的细胞存活至关重要，然而，该基因的具体功能一直未能明确。Nakai 等<sup>[44]</sup>的研究揭示了拟南芥叶绿体 TOC-TIC 复合体中的一个蛋白质 Tic214 是由 *ycf1* 基因编码，是植物叶绿体中蛋白质运输通道的关键组成部分。TOC 是叶绿体外膜转位子，TIC 是叶绿体内膜转位子，TOC-TIC 复合体位于叶绿体膜上，是介导前体蛋白顺利通过的通道，在叶绿体维持正常功能过程中起着重要作用。叶绿体在进化过程中，大多数基因被转移到细胞核中，有 2 000-3 000 多个蛋白需要在细胞质中合成后运输到叶绿体中<sup>[45,46]</sup>。

TOC-TIC 复合体中成分的缺失将导致叶绿体无法继续维持运转，会导致胚胎致死或白化<sup>[47,48]</sup>。TOC 目前已经被确认为 Toc75、Toc159 和 Toc34 家族<sup>[49-51]</sup>，但 TIC 的成分仍存在争议，目前主要提出了两种模式，经典模型认为 Tic110<sup>[52-54]</sup>为核心成分，还包括 Tic62、Tic55、Tic40、Tic32、Tic22、Tic21 和 Tic20<sup>[53,54]</sup>；另一种最新模型认为 Tic 由 Tic214、Tic100、Tic56 和 Tic20 组成，完整的 Tic20/56/100/214 复合体似乎是在胚胎植物（陆地植物）进化时形成的<sup>[44,55]</sup>。

TOC-TIC 复合体在植物细胞中扮演着重要角色，尤其是维持叶绿体蛋白稳定方面。Tic214 是 TIC-TOC 复合体中唯一的叶绿体编码亚基，对叶绿体中蛋白的正常运输起重要作用。抑制 Tic214 蛋白将削弱叶绿体蛋白质的输入，导致这些蛋白质在叶绿体核糖体和蛋白质折叠中无法发挥正常功能<sup>[46]</sup>。此外，Tic214 表达是维持正常叶绿体形态和蛋白平衡所必需的。Tic214 抑制会导致大量细胞肿胀，叶绿体具有高度无序的类囊体超微结构和淀粉颗粒的积累，这些都是经历叶绿体蛋白毒性应激的细胞的典型特征<sup>[45]</sup>。总之，Tic214 蛋白在植物生长发育和进化过程中都扮演着十分重要的角

色，需要不断深入研究其功能。

## 1.3 本文的研究内容，研究目的及意义

### 1.3.1 研究内容

本研究首先利用 55 种栎属植物的叶绿体基因组序列构建了栎属物种系统发育树并探讨栎属物种的进化关系；其次利用 Phylogenetic Analysis by Maximum Likelihood (PAML) 分析了 55 个栎属物种中 72 个共有蛋白编码基因的选择压力，通过位点模型和分支位点模型找出受正选择的基因；由于 *ycf1* 具有五个连续的正选择位点，在常绿和落叶中存在显著差异，因此挑选正选择基因 *ycf1* 进行重点研究，对 *ycf1* 基因的蛋白质基本特征包括蛋白质基本信息、跨膜结构域预测、二级结构和三维结构进行分析，并通过酵母双杂技术对 *ycf1* 基因具体的功能进行验证。

### 1.3.2 研究目的及意义

本研究旨在挖掘常绿和落叶栎属物种叶绿体基因组水平的差异，着重探究受到正选择的基因并对这些基因进行功能研究。本文的意义是探究影响不同叶习性栎属物种光合能力的分子机制，深入植物叶绿体基因组机理研究。对常绿和落叶栎属物种的光合能力的差异研究提供了新的见解，丰富了非模式物种中叶绿体基因功能的研究。

## 2 不同叶习性栎属物种叶绿体基因组序列比较

### 2.1 材料与方法

#### 2.1.1 叶绿体基因组数据获取

本研究涉及 55 种栎属植物的叶绿体基因组，具体信息见表 2.1，其中 14 个栎属植物的叶绿体基因组来自北京林业大学分子生态学实验室<sup>[15]</sup>，其余 41 个物种的叶绿体基因组数据从 NCBI 下载。实验室测序组装的 14 种栎属植物包含 7 种常绿栎和 7 种落叶栎，叶绿体基因组详细信息见表 2.1 前 14 个物种，具体采样信息见图 2.1。

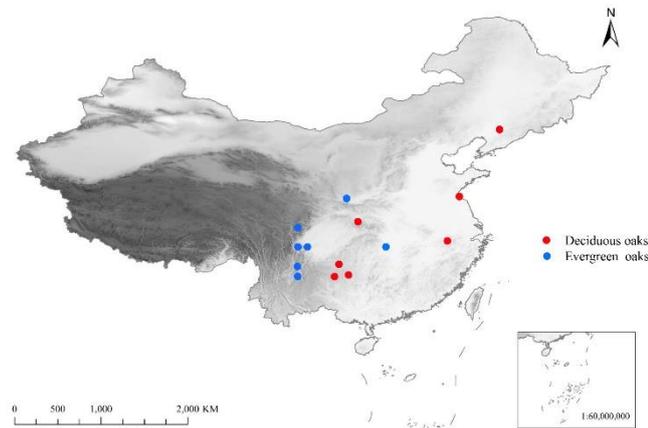


图 2.1 14 个栎属物种的采样点

Figure 2.1 Sampling sites for 14 *Quercus* species

表 2.1 55 个栎属物种叶绿体基因组详细信息表

Table 2.1 Details of chloroplast genomes of 55 *Quercus* species

物种 (Species)	组 (Section)	亚属 (Subgenus)	GenBank	叶习性 (Leaf habit)
<i>Quercus baronii</i>	<i>Ilex</i>	<i>Cerris</i>	MW829651	半常绿
<i>Quercus cocciferoides</i>	<i>Ilex</i>	<i>Cerris</i>	MW829652	半常绿
<i>Quercus engleriana</i>	<i>Ilex</i>	<i>Cerris</i>	MW829653	常绿/半常绿
<i>Quercus monimotricha</i>	<i>Ilex</i>	<i>Cerris</i>	MW829654	常绿
<i>Quercus rehderiana</i>	<i>Ilex</i>	<i>Cerris</i>	MW829655	常绿
<i>Quercus semecarpifolia</i>	<i>Ilex</i>	<i>Cerris</i>	MW829656	常绿
<i>Quercus senescens</i>	<i>Ilex</i>	<i>Cerris</i>	MW829657	常绿

<i>Quercus fabri</i>	<i>Quercus</i>	<i>Quercus</i>	MK922346.1	落叶
<i>Quercus griffithii</i>	<i>Quercus</i>	<i>Quercus</i>	MK922347.1	落叶
<i>Quercus mongolica</i>	<i>Quercus</i>	<i>Quercus</i>	MK922349.1	落叶
<i>Quercus serrata</i>	<i>Quercus</i>	<i>Quercus</i>	MK922350.1	落叶
<i>Quercus serrata</i> var. <i>brevipetiolata</i>	<i>Quercus</i>	<i>Quercus</i>	MK922351.1	落叶
<i>Quercus dentata</i> subsp. <i>stewardii</i>	<i>Quercus</i>	<i>Quercus</i>	MK922352.1	落叶
<i>Quercus yunnanensis</i>	<i>Quercus</i>	<i>Quercus</i>	MW876863	落叶
<i>Quercus acuta</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	MT742291.1	常绿
<i>Quercus chungii</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	MW401633.1	常绿
<i>Quercus delavayi</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	MW450870.1	常绿
<i>Quercus edithiae</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	KU382355.1	常绿
<i>Quercus gilva</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	MK860968.1	常绿
<i>Quercus glauca</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	NC_036930.1	常绿
<i>Quercus multinervis</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	MW450871.1	常绿
<i>Quercus myrsinifolia</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	MN199025.1	常绿
<i>Quercus obovatifolia</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	NC_039972.1	常绿
<i>Quercus saravanensis</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	MW411183.1	常绿
<i>Quercus schottkyana</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	MW450872.1	常绿
<i>Quercus sichourensis</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	NC_036941.1	常绿
<i>Quercus stewardiana</i>	<i>Cyclobalanopsis</i>	<i>Cerris</i>	MN199023.1	常绿
<i>Quercus aquifolioides</i>	<i>Ilex</i>	<i>Cerris</i>	KP340971.1	常绿
<i>Quercus bawanglingensis</i>	<i>Ilex</i>	<i>Cerris</i>	MK449426.1	常绿
<i>Quercus dolicholepis</i>	<i>Ilex</i>	<i>Cerris</i>	KU240010.1	常绿
<i>Quercus franchetii</i>	<i>Ilex</i>	<i>Cerris</i>	MW450869.1	常绿
<i>Quercus pannosa</i>	<i>Ilex</i>	<i>Cerris</i>	NC_050963.1	常绿
<i>Quercus phillyreoides</i>	<i>Ilex</i>	<i>Cerris</i>	NC_048488.1	常绿
<i>Quercus spinosa</i>	<i>Ilex</i>	<i>Cerris</i>	KM841421.1	常绿
<i>Quercus tarokoensis</i>	<i>Ilex</i>	<i>Cerris</i>	NC_036370.1	常绿
<i>Quercus tungmaiensis</i>	<i>Ilex</i>	<i>Cerris</i>	NC_036936.1	常绿
<i>Quercus virginiana</i>	<i>Virentes</i>	<i>Quercus</i>	NC_050972.1	常绿
<i>Quercus acutissima</i>	<i>Cerris</i>	<i>Cerris</i>	NC_039429.1	落叶
<i>Quercus chenii</i>	<i>Cerris</i>	<i>Cerris</i>	NC_039428.1	落叶
<i>Quercus variabilis</i>	<i>Cerris</i>	<i>Cerris</i>	MK105451.1	落叶
<i>Quercus coccinea</i>	<i>Lobatae</i>	<i>Quercus</i>	NC_047481.1	落叶
<i>Quercus palustris</i>	<i>Lobatae</i>	<i>Quercus</i>	MK105461.1	落叶
<i>Quercus rubra</i>	<i>Lobatae</i>	<i>Quercus</i>	JX970937.1	落叶
<i>Quercus aliena</i>	<i>Quercus</i>	<i>Quercus</i>	KP301144.1	落叶
<i>Quercus aliena</i> var. <i>acutiserrata</i>	<i>Quercus</i>	<i>Quercus</i>	MK105452.1	落叶
<i>Quercus dentata</i>	<i>Quercus</i>	<i>Quercus</i>	MK105453.1	落叶
<i>Quercus</i> × <i>fenchengensis</i>	<i>Quercus</i>	<i>Quercus</i>	NC_048513.1	落叶
<i>Quercus gambelii</i>	<i>Quercus</i>	<i>Quercus</i>	MK105457.1	落叶
<i>Quercus lobata</i>	<i>Quercus</i>	<i>Quercus</i>	CM012305.2	落叶
<i>Quercus macrocarpa</i>	<i>Quercus</i>	<i>Quercus</i>	MK105459.1	落叶

<i>Quercus mongolica</i> subsp. <i>crispula</i>	<i>Quercus</i>	<i>Quercus</i>	NC_049877.1	落叶
<i>Quercus petraea</i>	<i>Quercus</i>	<i>Quercus</i>	LT996899.1	落叶
<i>Quercus robur</i>	<i>Quercus</i>	<i>Quercus</i>	LT996900.1	落叶
<i>Quercus stellata</i>	<i>Quercus</i>	<i>Quercus</i>	MK105467.1	落叶
<i>Quercus wutaishanica</i>	<i>Quercus</i>	<i>Quercus</i>	NC_043857.1	落叶

注：栎属物种分组和叶习性参考文献：[26]，[27]；网站：<https://www.treesandshrubsonline.org/>

## 2.2 数据分析

### 2.2.1 叶绿体基因组序列基本特征分析

本论文利用 GeniousR9 从栎属植物叶绿体基因组大小、各分区的大小、总 GC 含量、各分区的 GC 含量和叶绿体基因组中包含各类基因的数量和类型等方面，对 14 种栎属植物的叶绿体基因组基本特征进行比较。

### 2.2.2 重复序列与序列特异性分析

本论文对 14 个栎属物种中的重复序列进行比较，包括简单重复序列、回文重复序列、反向重复序列等。简单重复序列，也叫微卫星序列，是 1-6 碱基的重复单元<sup>[56]</sup>，经常作为分子标记使用。通过在服务器中运行 MISA<sup>[57]</sup>程序来检测简单重复序列，从单核苷酸重复至六核苷酸重复，其最小重复次数参数分别设置为 10、5、4、3、3 和 3。其他重复序列如正向、反向、回文、互补重复序列的检测则利用在线网站 REPuter<sup>[58]</sup> (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>)来进行搜索，参数设置为重复序列一致性需大于 80%，Hamming 距离设置为 3，重复长度要大于 30bp<sup>[58]</sup>。对检测结果筛选单核苷酸、二核苷酸等六种核苷酸序列，得到不同重复类型的数量，如单核苷酸中 A、T、C、G 重复的数量。

为了检测不同叶习性栎属物种叶绿体基因组序列的差异，首先利用 Cygwin 软件中的 MAFFT 比对工具来比对叶绿体基因组，生成 fas 文件。然后采用在线工具 mVISTA 来对序列进行比较，参数设置选择 Shuttle-LAGAN 算法，可视化不同栎属物种的差异<sup>[59]</sup>。同时，利用在线分析工具 IRscope 对 14 种栎属物种的边界区域可视化，重点分析各栎属物种叶绿体基因组边界处的基因<sup>[60]</sup>。

### 2.2.3 密码子偏好性分析

密码子使用偏好性是指在基因翻译过程中，密码子使用情况存在差异，某些密码子的使用次数会多于其它同义密码子<sup>[61]</sup>。本研究首先整理 55 个栎属物种叶绿体基因组的 CDS 序列，删除长度小于 300bp 及重复基因序列，并且确保每个基因都有起始密码子，得到 52 条完整的 CDS 序列。将 52 条 CDS 序列视为一条，放在一个文件中，放在 CodonW 软件的工作目录下，采用软件 CodonW 统计相对同义密码子使用

度 (Relative synonymous codon usage, RSCU)、叶绿体基因组密码子第三位碱基含量 (T3s、C3s、A3s、G3s) 和有效密码子数 (Effective Number of Codon, ENC) 来分析常绿和落叶物种叶绿体基因组的密码子使用偏好性, 探究不同叶习性栎属物种的进化压力和环境适应性。

#### 2.2.4 系统发育分析

为了探究不同栎属物种之间的进化关系, 本研究利用最大似然法 (Maximum Likelihood, ML) 和邻接法 (Neighbor Joining, NJ) 来构建系统发育树。将 55 个叶绿体基因组序列利用 Cygwin 软件中的 Mafft 程序包来进行比对后得到 fas 文件。ML 树利用 Raxml 软件来构建, NJ 树利用 MEGA 来构建, bootstrap 值均设定为 1 000。

#### 2.2.5 选择压力分析

本实验室已有数据结合 NCBI 下载的 41 个栎属物种的叶绿体基因组序列, 共 55 个物种的叶绿体基因组序列, 利用 PAML 软件中的 codeml 程序做正选择分析<sup>[62]</sup>。首先确定 55 个物种中不重复的蛋白编码基因, 并将每个基因对应的 55 个种的序列放在同一文档中。对每个基因的序列进行检查, 确保每个物种每个基因的序列长度均为 3 的倍数, 进行人工校正确保序列正确, 并去除终止密码子。在 MEGA 中对每个基因的文件进行比对, 导出每个基因对应的 fas 文件。将整理好的蛋白编码基因数据集导入服务器中。

首先使用 codeml 程序中的 M0 模型对这些共有蛋白编码基因的同义替换和非同义替换进行初步估计。同义核苷酸替代是碱基发生变化但氨基酸不发生改变; 非同义核苷酸替代是碱基发生变化同时氨基酸也会发生改变。通过计算非同义核苷酸替代和同义核苷酸替代的比值  $\omega$ , 衡量蛋白的选择压力。

本研究利用位点模型 (site-model) 和分支-位点模型 (branch-site model) 来探索这些栎属物种的选择压力。首先, 利用位点模型进行计算时, 选择 M1 vs. M2, M7 vs. M8 两对模型计算位点间的正选择。M1 模型认为  $\omega$  值小于 1 或等于 1, 整个序列中没有正选择位点, M2 模型则在此基础上增加了  $\omega$  值大于 1 的假设认为存在正选择位点; M7 模型假设所有位点的  $\omega$  值都在 0 到 1 之间, 服从 beta 分布, M8 模型类与 M7 类似, 但增加了  $\omega$  值大于 1 的位点, 认为有位点受到正选择。M1 和 M7 模型是研究中的零模型 (null model)。通过计算 M1 和 M2, M7 和 M8 模型的似然值差异, 用卡方检验分析显著性, 算出 P 值, 以 P 值小于 0.05 作为检验的标准, P 小于 0.05 认为存在正选择位点。

分支位点模型用来检测特定的分支或谱系上是否存在正选择位点。同样选择了两个模型, 第一个模型与位点型中的 M2 相同, 假设  $\omega$  值大可大于 1、小于 1 和等于 1; 另一个模型则假设  $\omega$  值固定为 1。分支位点模型需要设置前景枝, 一般将测试是否受

到正选择的分支定义为前景枝。本研究将常绿栎和落叶栎分别设置为前景枝，利用分支位点模型进行计算，由经验贝叶斯法（Bayes empirical Bayes, BEB）识别出后验概率大于等于 0.9 的正选择位点<sup>[62,63]</sup>。

## 2.3 研究结果

### 2.3.1 基本特征分析

统计栎属物种叶绿体基因组整体长度、各分区长度、GC 含量后发现，常绿和落叶栎属植物的基本特征及其相似，但是常绿栎中叶绿体基因组的变化较落叶栎更大。

首先是叶绿体基因组大小极其相似，范围在 161 114bp 到 161 415bp 之间。常绿树种叶绿体基因组大小在 161 114bp 到 161415bp 之间；落叶树种的叶绿体基因组大小在 161 227bp 到 161 247bp 之间。落叶栎物种间的叶绿体基因组大小更接近，变化更小，常绿栎物种间叶绿体基因组的变化更大（图 2.2）。

其次，常绿栎和落叶栎叶绿体基因组中的 LSC 区、SSC 区和 IR 区四个分区的大小也基本一致。但常绿栎叶绿体基因组物种间的变化较大：如常绿栎物种叶绿体基因组 LSC 区大小范围在 90 341bp 到 90 650bp 之间，落叶栎物种的叶绿体基因组大小范围在 90 539bp 到 90 612bp 之间（图 2.2）。

最后，14 个栎属物种的 GC 含量、其包含的基因种类、数量和顺序也极其接近。14 个栎属物种整个叶绿体基因组总体 GC 含量为 36.8%-36.9%之间，LSC 区的 GC 含量为 34.7%-36.9%之间，SSC 区 GC 含量为 30.9%-31%之间，IR 区 GC 含量为 42.7%-42.8%之间。14 个栎属物种的叶绿体基因组中的 IR 区的 GC 含量最高，而 SSC 区的 GC 含量最低。在基因数量和基因组成上这些栎属物种均为 86 个蛋白编码基因，40 个 tRNA 基因和 4 个 rRNA 基因。



重复都主要是 A/T 重复, C/G 重复较少; 二核苷酸重复也主要是由 AT/TA 重复组成; AAAT/TTTA 重复在重复类型中的占比也相对较高 (图 2.3)。

本论文还对栎类叶绿体基因组的回文重复、正向重复、反向重复和互补重复这四种重复序列进行了统计(图 2.4)。其中, 回文重复出现的次数最多, 范围从 17-25 次, 落叶栎物种的回文重复次数普遍是 20 次, 只有黄山栎 (*Q. dentata* subsp. *stewardii*) 次数较多, 出现了 25 次。常绿栎回文重复出现的次数是 17-21 次。互补重复出现的次数是最少的, 统计的 7 种常绿栎中有 5 种都没有互补重复, 只有槿子栎(*Q. baronii*) 和毛脉高山栎 (*Q. rehderiana*) 中出现过, 次数也只要一两次。反向重复出现的次数也较少, 落叶栎中的反向重复次数整体上高于落叶栎。

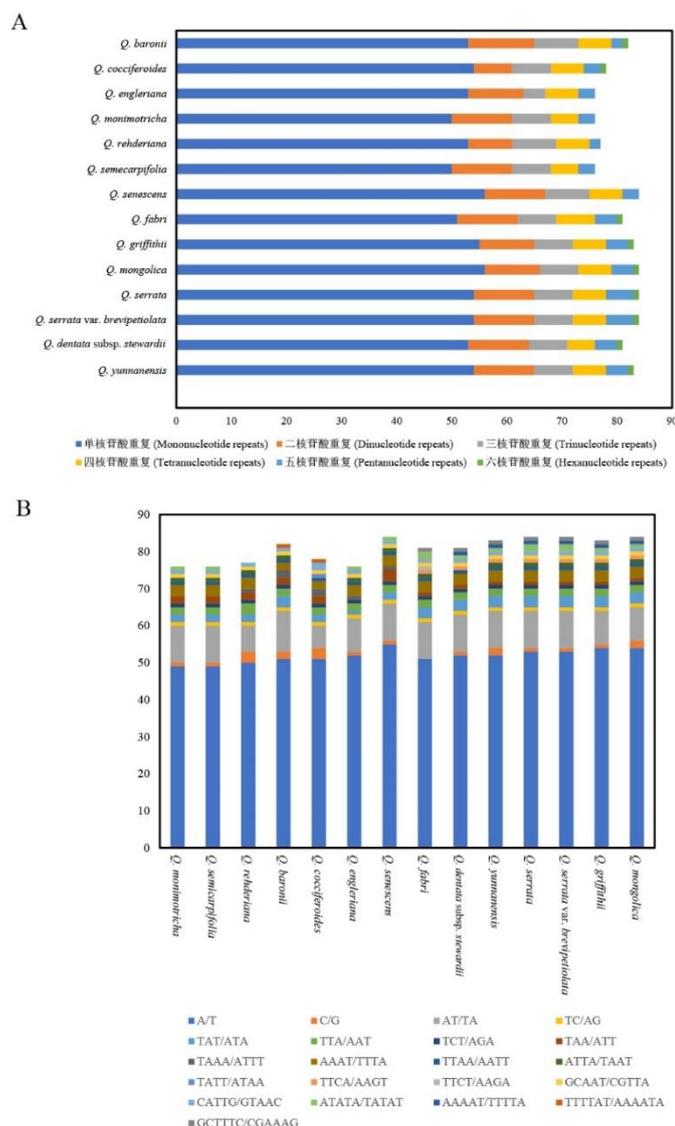


图 2.3 14 个栎属物种简单重复序列类型和数量统计

图 A: 14 个栎属物种 6 种简单重复序列数量统计; 图 B: 14 个栎属物种简单重复序列类型统计

Figure 2.3 Simple repeat sequence type statistics for 14 *Quercus* speciesFigure A: Statistics on the number of six simple repeat sequences of 14 *Quercus* species; Figure B:Statistics on the types of simple repeat sequences of 14 *Quercus* species

注：图 A 横坐标为 6 种简单重复序列的数量,每种颜色为不同类型，纵坐标为不同物种名称；图 B 横坐标为不同物种名称，纵坐标为各种简单重复序列类型的数量

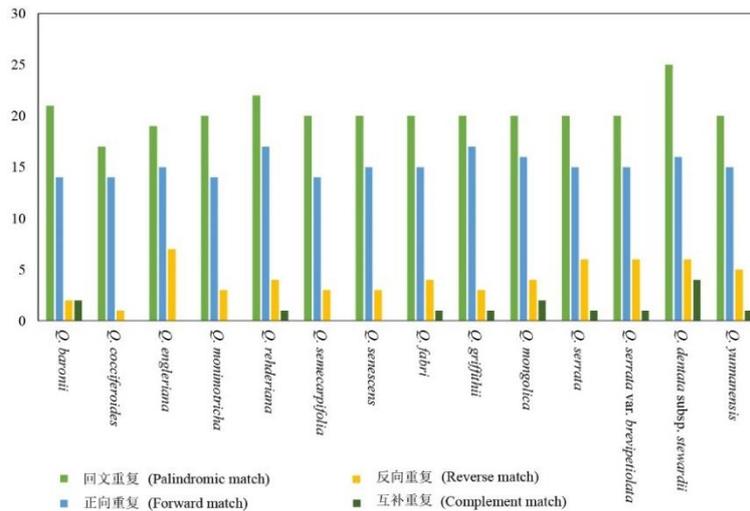


图 2.4 14 个栎属物种四种重复序列数量统计

Figure 2.4 Statistics on the number of four repetitive sequences in 14 *Quercus* species

注：横坐标为不同物种名称，纵坐标为不同重复序列类型的数量

### 2.3.3 序列多样性分析

虽然植物的叶绿体基因组非常保守，但在 IR 区与 LSC 区或 SSC 区的边界经常会发生扩张或收缩。本论文利用 IRscope 对 14 种栎的叶绿体基因组序列进行比较发现不同栎属物种在边界处存在差异(图 2.5)。位于 LSC/IRb 区边界、IRb/SSC 区边界、SSC/IRa 区边界的基因分别是 *rps19* 基因、短拷贝的 *ycf1* 基因和长拷贝的 *ycf1* 基因。*ycf1* 基因有两份拷贝，分别位于 IRb/SSC 区边界和 SSC/IRa 区边界，位于 IRb/SSC 区边界的 *ycf1* 基因拷贝不完整，序列较短，为短拷贝 *ycf1* 基因；SSC/IRa 区边界的 *ycf1* 基因则拷贝正常，为长拷贝 *ycf1* 基因。在 14 种栎属植物中，短拷贝 *ycf1* 基因位于 SSC 区的长度从 15bp-79bp 不等。但 7 种落叶栎的短拷贝 *ycf1* 基因在 SSC 区的长度均较大，大于 42bp。而常绿栎的短拷贝 *ycf1* 基因在 SSC 区的则长度变化较大，铁橡栎 (*Q. cocciferoides*) 中最小是 15bp，灰背栎 (*Q. senescens*) 中最大达到 72bp。SSC/IRa 区边界的长拷贝 *ycf1* 基因同样横跨了两个区域，位于 IRa 区的片段长度从 1042-1067bp 不等。其中，7 种落叶栎长拷贝 *ycf1* 基因位于 IRa 区的片段长度均为 1042bp，而常绿

栎则从 1047bp-1067bp 不等，落叶栎的 *ycf1* 基因长度更为接近。

以榲子栎叶绿体基因组作为参考序列，比对了 14 个栎属物种的叶绿体基因组序列（图 2.6），结果表明，编码区的序列更为保守，差异更小，序列更相似，保守非编码区差异较大，表明基因功能是高度保守。虽然编码区的差异整体较小，但发现 *ycf1* 基因存在较大的差异。IR 区比 SC 区更保守，在 LSC 区变异较多。同时，发现在 14 种栎属物种中，同种叶习性的栎树更为相似，存在多处相同的变异。

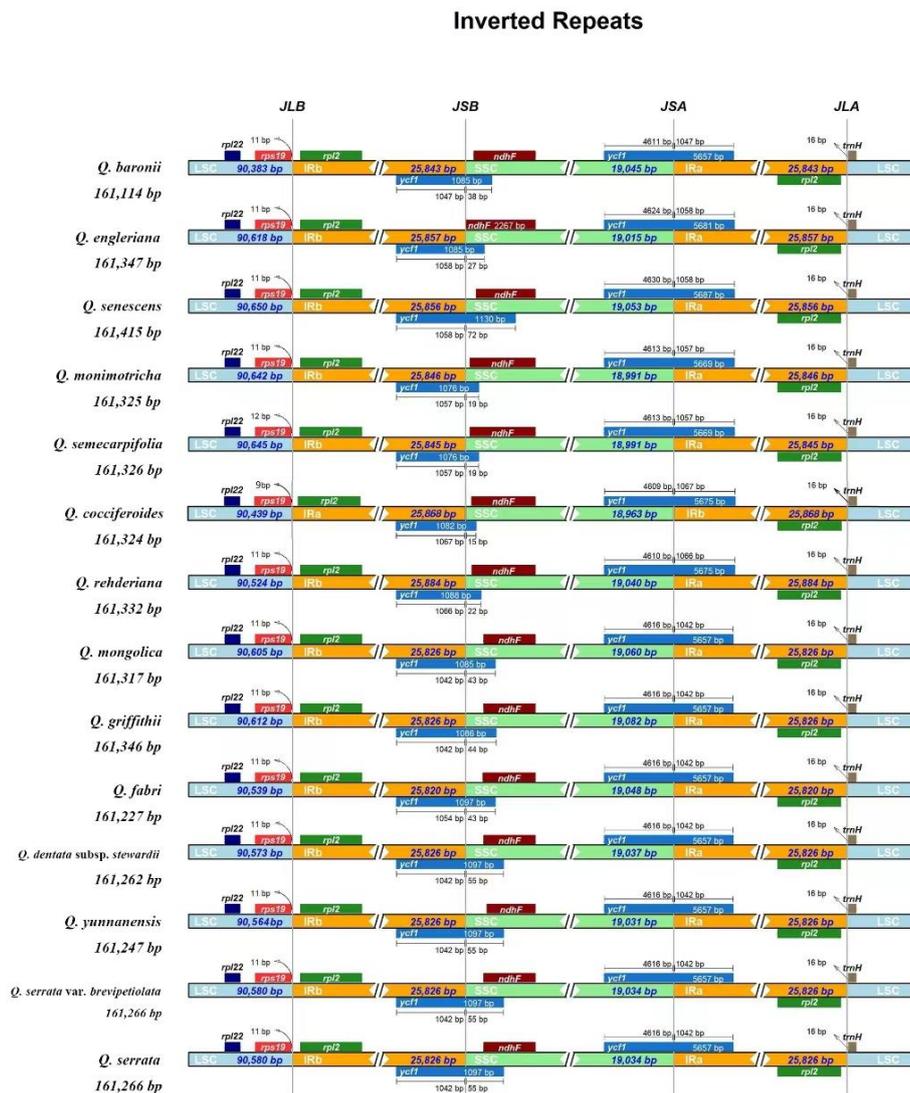


图 2.5 14 个栎属物种叶绿体基因组 LSC 区、SSC 区、IR 区分区边界对比图

Figure 2.5 Comparison of the boundaries of the LSC, SSC and IR regions of the chloroplast genomes of 14 *Quercus* species

注：JLB（junction of LSC and IRb）为 LSC 区与 IRb 区的连接点；JSB（junction of SSC and IRb）为 SSC 区与 IRb 区的连接点；JSA（junction of SSC and IRa）为 SSC 区与 IRa 区的连接点；JLA（junction of LSC and IRa）为 LSC 区与 IRa 区的连接点

# 常绿和落叶栎类叶绿体基因功能研究

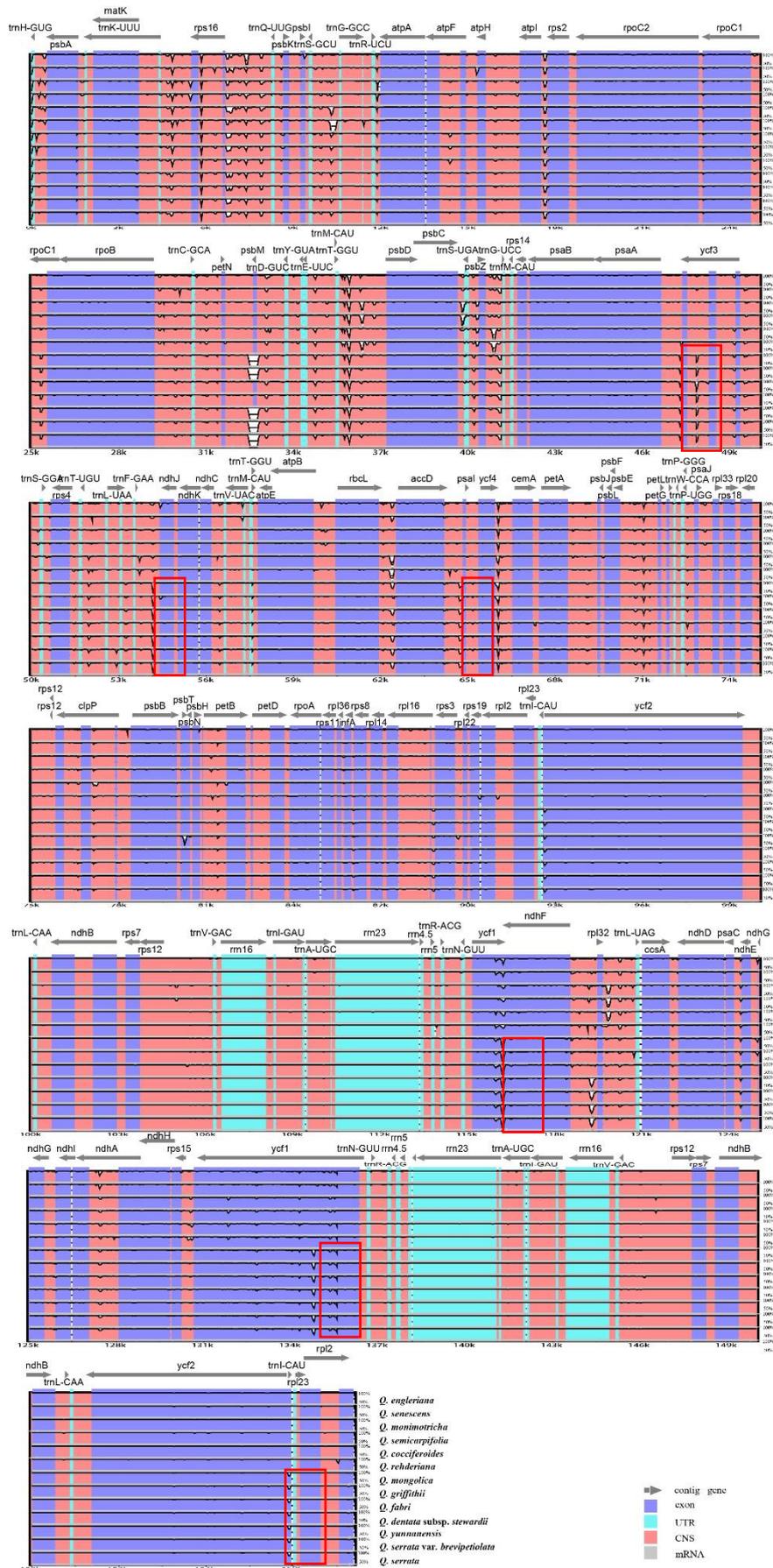




Table 2.2 Chloroplast genome codon third base content and ENC statistics for 14 *Quercus* species

物种 (Species)	T3s	C3s	A3s	G3s	有效密码子数 (ENC)
<i>Q. baronii</i>	0.4651	0.1721	0.4325	0.1812	49.97
<i>Q. cocciferoides</i>	0.4644	0.1725	0.4332	0.1808	49.95
<i>Q. engleriana</i>	0.4642	0.1728	0.4332	0.1809	49.98
<i>Q. monimotricha</i>	0.464	0.1727	0.4327	0.1814	49.98
<i>Q. rehderiana</i>	0.4642	0.1724	0.4333	0.1810	49.94
<i>Q. semecarpifolia</i>	0.4640	0.1727	0.4326	0.1814	49.98
<i>Q. senescens</i>	0.4642	0.1728	0.4332	0.1811	49.98
<i>Q. fabri</i>	0.4645	0.1723	0.4333	0.1807	49.91
<i>Q. griffithii</i>	0.4646	0.1722	0.4331	0.1809	49.92
<i>Q. mongolica</i>	0.4646	0.1722	0.4332	0.1808	49.92
<i>Q. serrata</i>	0.4647	0.1722	0.4334	0.1805	49.90
<i>Q. serrata</i> var. <i>brevipetiolata</i>	0.4647	0.1722	0.4333	0.1805	49.90
<i>Q. dentata</i> subsp. <i>stewardii</i>	0.4647	0.1722	0.4333	0.1806	49.91
<i>Q. yunnanensis</i>	0.4646	0.1721	0.4334	0.1807	49.91

注: A3s、T3s、C3s、G3s 分别表示密码子第三位碱基的 A、T、G、C 含量; A3s: the content of the third base A; T3s: the content of the third base T; C3s: the content of the third base C; G3s: the content of the third base G

表 2.3 栎属物种密码子使用偏性表

Table 2.3 Codon usage bias of *Quercus* species

氨基酸 (amino acid)	密码子 (codon)	相对同义密码子 使用度(RSCU)	氨基酸 (amino acid)	密码子 (codon)	相对同义密码 子使用度 (RSCU)
Phe	UUU	1.325	His	CAU	1.542143
	UUC	0.675		CAC	0.457857
	UUA	1.985		UCU	1.657857
	UUG	1.208571		UCC	1.03
Leu	CUU	1.211429	Ser	UCA	1.163571
	CUC	0.392857		UCG	0.551429
	CUA	0.792143		AGU	1.217857
	CUG	0.404286		AGC	0.379286
Ile	AUU	1.459286	TER	UAA	1.859286
	AUC	0.579286		UAG	0.594286

不同叶习性栎属物种叶绿体基因组序列比较

	AUA	0.963571		UGA	0.55
Met	AUG	1	Trp	UGG	1
	GUU	1.439286		GCT	1.781429
Val	GUC	0.462857	Ala	GCC	0.631429
	GUA	1.532143		GCA	1.103571
	GUG	0.566429		GCG	0.479286
	CCU	1.489286		ACU	1.618571
Pro	CCC	0.829286	Thr	ACC	0.732143
	CCA	1.111429		ACA	1.198571
	CCG	0.569286		ACG	0.45
Tyr	UAU	1.585	Asn	AAU	1.55
	UAC	0.415		AAC	0.45
Gln	CAA	1.555714	Lys	AAA	1.51
	CAG	0.444286		AAG	0.49
Asp	GAU	1.615	Glu	GAA	1.523571
	GAC	0.385		GAG	0.476429
	CGU	1.309286		GGU	1.298571
	CGC	0.422143		GGC	0.482143
Arg	CGA	1.342143	Gly	GGA	1.517857
	CGG	0.448571		GGG	0.697857
	AGA	1.813571	Cys	UGU	1.437143
	AGG	0.663571		UGC	0.562857

### 2.3.5 系统发育分析

比较了 14 种栎属植物的叶绿体基因组后，发现常绿栎与落叶栎之间存在一定的差异。为了进一步深入比较不同叶片习性的栎树叶绿体基因组的差异，从 NCBI 补充下载了 41 个栎属物种的叶绿体基因组（表 2.1）。基于 55 个栎属物种的叶绿体基因组，采用最大似然法和邻接法构建了系统发育树。结果显示，55 个栎属物种分为了栎亚属和麻栎亚属，栎亚属包含白栎组、活栎组和红栎组；麻栎亚属包含青冈栎组、冬青栎组和麻栎组。NJ 树中，麻栎亚属的麻栎组镶嵌在冬青栎组中。ML 树中灰背栎、巴东栎 (*Q. engleriana*)、川滇高山栎和通麦栎 (*Q. tungmaiensis*) 属于冬青栎组但是和栎亚属物种之间的关系更近。所有青冈栎组植物和部分冬青栎组植物在系统发育的基部位置，其余部分的冬青栎组、麻栎组和白栎组等多种栎亚属的物种则位置较后。

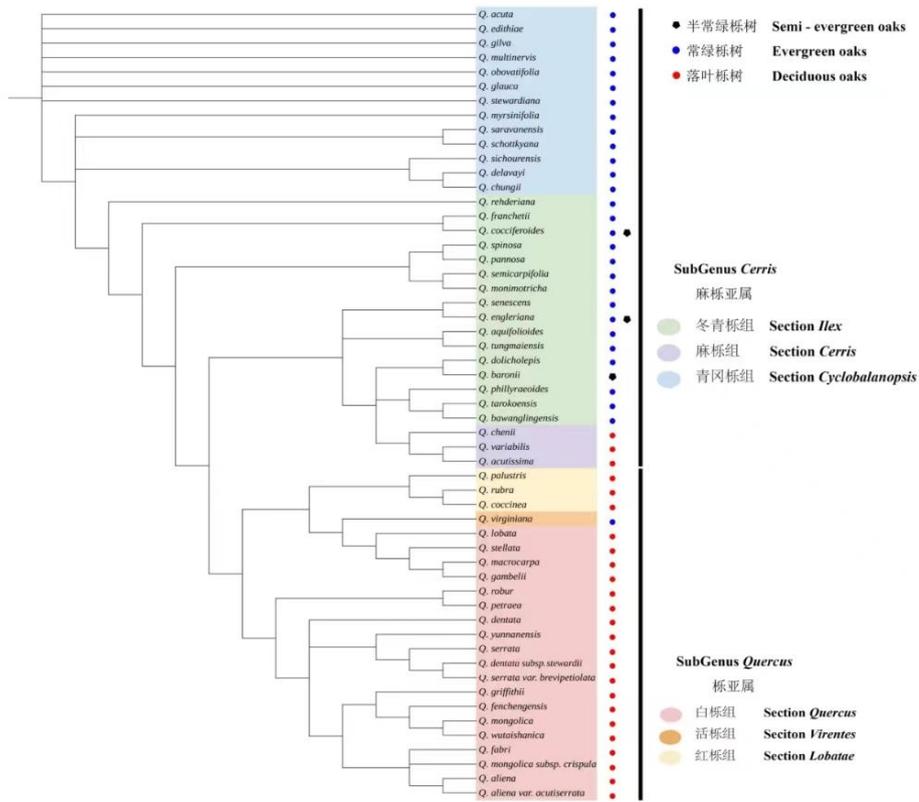


图 2.8 基于 55 个栎属物种叶绿体基因组利用邻接法构建的系统发育树

Figure 2.8 Neighbor-Joining tree constructed based on the chloroplast genomes of 55 *Quercus* species



循环电子传递中起重要作用<sup>[66-68]</sup>。*ccsA* 基因表达并编码血红素附着于 c 型细胞色素所需的蛋白<sup>[69-71]</sup>。*ycf1* 编码蛋白 Tic214。之前的研究表明，在叶绿体中存在一个由 Tic214、Tic100、Tic56 和 Tic20 组成的功能性 1 兆道尔顿 (Mda) 蛋白易位复合物<sup>[44,55,72]</sup>。最新的研究表明，膜蛋白 Tic214 横跨内膜、间质膜和外膜，是位于内膜、外膜和间质膜的不同蛋白质亚基之间的桥梁<sup>[54]</sup>。

表 2.4 55 个栎属物种分支位点模型正选择分析结果

Table 2.4 Results of the branch-site model positive selection analysis of 55 *Quercus* species

基因 (Gene)	分支-位点模型 (Branch-site model)	-lnL	2 $\delta$ (lnL)	P 值 (P value)	Q 值 (Q value)	正选择位点 (Positively selected sites)
<i>rps2</i>	Branch (Evergreen species)					104-0.981,206-0.609,
	Null	1059.27				234-0.632
	Alternative	1054.89	8.76	3.08E-03	4.44E-02	
<i>ccsa</i>	Branch (Evergreen species)					89-0.996,96-0.515, 144-0.5,166-0.527
	Null	1498.51				
	Alternative	1488.36	20.31	6.60E-06	2.38E-04	
<i>ndhd</i>	Branch (Evergreen species)					28-0.617,56-0.618, 84-0.638,107-0.610,
	Null	2508.85				118-0.643,188-0.946, 217-0.603,279-0.611,
	Alternative	2500.34	17.01	3.72E-05	8.93E-04	504-0.988
<i>ycf1</i>	Branch (Evergreen species)					3-0.910,332-0.525,

不同叶习性栎属物种叶绿体基因组序列比较

Null	9959.86				342-0.942,396-0.522, 649-0.522,740-0.522, 743-0.719,753-0.823, 1409-0.885,1433-0.509, 1474-0.994,1475-1.000, 1476-0.996,1477-0.744,
Alternative	9931.84	56.04	7.12E-14	5.13E-12	1478-0.693,1504-0.503, 1575-0.516,1628-0.902, 1826-0.519

注：最后一列连字符前为由贝叶斯经验贝叶斯 (Bayes empirical bayes , BEB) 分析鉴定出的氨基酸的位置。连字符后面的数字是 BEB 分析下的后验概率 (Posterior probability , PP)

### 3 基因功能研究

#### 3.1 实验材料与试剂

##### 3.1.1 实验材料及试剂

实验中所涉及植物 DNA 来自北京林业大学分子生态学实验室。

实验试剂：

植物 DNA 提取试剂盒	TIANGEN
高纯度质粒小提中量试剂盒 DP107	TIANGEN
微柱浓缩 DNA 凝胶回收试剂盒	ZOMANBIO
Y2H Gold Chemically Competent Cell	上海唯地 WEDI
EZ-TA/Blunt Zero pTOPO II Cloning Kit	康润生物 GenStar
Trans10 Chemically Competent Cell	北京全式金
2×Taq PCR StarMix	Genstar

##### 3.1.2 实验试剂配制

50mg/ml 卡那霉素 (Kana)：按照所需浓度对抗生素进行配置后过滤除菌。首先确定配制的浓度为 50mg/ml，精确称量 10g 卡那霉素粉末，加入足量的蒸馏水溶解，待粉末完全溶解后，再加蒸馏水定容至 200ml。在超净工作台中进行过滤灭菌处理后，将过滤好的卡那霉素溶液分装至 2ml 灭菌离心管中，标注浓度和配制时间，放在-20℃低温贮存。

表 3.1 实验涉及培养基的配制方法

Table 3.1 Experiments involving the preparation of culture media

培养基名称 (Medium)	配方 (Medium component)
LB (Luria-Bertani) 液体培养基 (200ml)	蛋白胨 (Typtone) 2g 酵母提取物 (Yeast extract) 1g 氯化钠 (NaCl) 2g 加蒸馏水定容至 200ml
LB (Luria-Bertani) 固体培养基 (200ml)	蛋白胨 (Typtone) 2g 酵母提取物 (Yeast extract) 1g 氯化钠 (NaCl) 2g 琼脂粉 (Agar) 3g

	加蒸馏水定容至 200ml
SD-Trp-Leu 液体培养基 (二缺液体培养基) (200ml)	酵母氮源 (YNB) 1.34g -Leu/-Trp DO Supplement 0.12g 葡萄糖 (D-Glucose) 4g 加蒸馏水定容至 200ml
SD/-Trp-Leu 固体培养基 (二缺固体培养基) (200ml)	酵母氮源 (YNB) 1.34g -Leu/-Trp DO Supplement 0.12g 葡萄糖 (D-Glucose) 4g 琼脂粉 (Agar) 4g 加蒸馏水定容至 200ml
SD/-Leu-Trp-His-Ade 固体培养基 (四缺固体培养基) (200ml)	酵母氮源 (YNB) 1.34g - Ade/-Leu/-Trp/-His/ DO Supplement 0.12g 葡萄糖 (D-Glucose) 4g 琼脂粉 (Agar) 4g 加蒸馏水定容至 200ml

注：将配制好的培养基进行高压灭菌，温度设置为 121°C，高压灭菌 15min 后再使用

## 3.2 研究方法

### 3.2.1 生物信息学分析

由于本研究主要检测常绿和落叶栎属物种的差异，且位点模型和分支位点模型检测出的正选择基因存在重复，因此，选择了分支位点模型检测出的 4 个正选择基因做深入研究。找出分支位点模型 55 个栎属物种 4 个正选择基因的正选择位点对应的氨基酸后，用在线网站 [weblogo](http://weblogo) 并生成序列标识<sup>[73]</sup>，观察常绿栎和落叶栎正选择位点氨基酸的区别，及不同类型氨基酸出现的频率，分析该位点的序列保守性。利用 ExPASy ProtParam 软件 (<https://web.expasy.org/protparam/>) 对正选择基因的基本理化性质进行统计和预测，包括氨基酸数量、蛋白亲水性和不稳定系数等<sup>[74]</sup>。利用 IPC-Isoelectric Point Calculator (<http://isoelectric.org/>) 计算了正选择位点的等电点<sup>[75]</sup>。利用在线网站 TMHMM<sup>[76]</sup> (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>) 对正选择基因的跨膜结构域进行预测。利用在线网站 SPOMA([https://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_sopma.html](https://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html))对正选择基因的二级结构进行预测。利用 AlphaFold2 预测其三维结构<sup>[77]</sup>，在 pymol 中美化。基于 *ycf1* 基因序列利用 MEGA 软件构建了 55 个栎属物种的 NJ 树和 ML 树，构建 ML 树时，先利用 MEGA 中的 models

菜单中的 Find best DNA/Protein Models(ML)计算出 *ycf1* 基因建树最合适的模型, bootstrap 值均设置为 1 000。

### 3.2.2 酵母双杂实验

细胞核编码的 2 000-3 000 个蛋白质在细胞质中作为前体蛋白合成后, 运输到叶绿体中。这些前体蛋白含有 N 端转运肽, 有助于将这些蛋白引导至叶绿体膜表面<sup>[78,79]</sup>。Tic214 是由叶绿体编码的 TIC 组成成分, 是 TOC-TIC 复合体中唯一的叶绿体编码的亚基。而且 Tic20/56/100/214 组成的 TIC 复合体可能负责主要的光合蛋白运输<sup>[80,81]</sup>。TOC 复合体与新的 TIC 复合体组分 (Tic56, Tic100 和 Tic214) 之间可能存在强烈的物理相互作用<sup>[72,82]</sup>。在拟南芥中, Tic100 还含有两个酸性片段, 可能与前体蛋白带正电荷的转运肽发生短暂的相互作用<sup>[5]</sup>。因此, Tic214 蛋白也可能与转运肽发生相互作用。

常绿和落叶树种氨基酸序列的不同可能会影响 Tic214 蛋白的结构和功能。*ycf1* 基因编码的 Tic214 蛋白是叶绿体 TOC-TIC 通道的组成部分, 可能会影响到叶绿体中蛋白的通过速度, 影响传输效率。本研究假设氨基酸序列的不同可能会影响进入叶绿体的蛋白质与 Tic214 的相互作用强度。由于进入叶绿体中的蛋白质 N 端均含有转运肽, 因此, 选取转运肽序列和常绿落叶栎属物种中 *ycf1* 基因的序列, 观察二者的相互作用强度来判断对传输的效率的影响。

本研究利用酵母双杂技术对蛋白相互作用结果进行验证和比较。实验中所采用的菌株是 Y2H Gold 菌株, 与 pGBKT7 和 pGADT7 两种质粒配套使用。质粒 pGADT7 携带 Leu 基因, 质粒 pGBKT7 携带 Trp 基因。SD/-Trp-Leu 培养基表示缺乏亮氨酸和色氨酸的培养基, 即二缺培养基。SD/-Leu-Trp-His-Ade 培养基表示缺乏亮氨酸、色氨酸、组氨酸和腺嘌呤的培养基, 即四缺培养基。首先构建质粒, 将转运肽与 *ycf1* 基因序列形成不同组合, 将两段序列分别连接到载体 pGBKT7 和 pGADT7 上, 由于 *ycf1* 基因的序列过长, 本研究首先利用了差异较大的部分片段, 并根据 Holbrook 等<sup>[83]</sup>和 Sork 等<sup>[84]</sup>发表的序列在 NCBI 中比对选取合适的转运肽: Rubisco 小亚基序列的转运肽和 Rubisco Activase 酶的转运肽序列。

之后共转化到酵母细胞后在二缺培养基上涂板, 观察酵母菌落的生长, 若长出酵母菌落即对应的质粒转化成功。pGADT7 质粒携带 Leu 基因, 转入酵母菌株后可以合成亮氨酸, pGBKT7 载体携带 Trp 基因, 转入酵母菌株后可以合成色氨酸, 两个质粒都转入酵母菌株, 酵母菌株就可以在对应的二缺板上生长。之后选取单菌落, 在 28 摄氏度的摇床中过夜摇菌, 取出、离心重悬后稀释成不同的浓度梯度的溶液, 滴在四缺培养基上观察生长情况。若存在相互作用, 则 GAL4 可以激活报告基因, 合成组氨酸和腺嘌呤, 结合 pGBKT7 和 pGADT7 中携带的 Leu 基因和 Trp 基因, 酵母菌株可以在四缺培养基上生长。如果没有相互作用, GAL4 无法激活报告基因, 无法合成组

氨酸和腺嘌呤，则酵母菌落无法生长。具体实验步骤如下：

- 1) 提前做好 Carrier DNA、Y2H Gold 感受态、待转化质粒、PEG/LiAC 等试剂，将 Y2H Gold 感受态细胞放在冰上。提前配制好 SD-Trp-Leu 液体培养基、SD-Trp-Leu 固体培养基和 SD-Trp-Leu-Ade-His 固体培养基；
- 2) Carrier DNA 预处理：将 Carrier DNA 水煮 5min 后再冰浴 5min，重复一次后取出备用；
- 3) 取 1.5ml 离心管，加入 100 微升 Y2H Gold 感受态细胞，预冷质粒 5 微克，预处理后的 Carrier DNA 10 微升，PEG/LiAC 500 微升，并轻轻吹打混匀；
- 4) 恢复培养：将上一步加好的产物 30 度水浴 30min，在 15min 时翻转混匀 6-8 次；
- 5) 热激：取出离心管放在 42 度水浴 15min，在 7.5min 时翻转混匀 6-8 次；
- 6) 洗涤：取出离心管，5 000rpm 离心 90s。然后弃上清，用 400 微升 ddH<sub>2</sub>O 重悬，再 5 000rpm 离心 30s，弃上清；
- 7) 涂板：取 50 微升 ddH<sub>2</sub>O 重悬，在二缺固体培养基上涂板；
- 8) 培养：涂板后在 28 度-30 度的培养箱中，倒置培养 48-96h；
- 9) 待二缺固体培养基长出菌落以后，挑选菌落在摇床中过夜培养；
- 10) 在 2ml 离心管中加入 1ml 二缺液体培养基，并挑选单克隆菌落加入离心管中；
- 11) 每个培养基挑 2 个菌落，用封口膜封好，在 28-30 度的摇床摇过夜；
- 12) 从摇床中取出离心管，7 000xg 离心 1min；
- 13) 在超净工作台中倒掉离心管中废液，加入 1ml ddH<sub>2</sub>O 洗涤，并吹打均匀；
- 14) 混匀后从超净工作台中取出，再 7 000xg 离心 1min；
- 15) 在超净工作台中倒出 ddH<sub>2</sub>O，加入 100 微升 0.9%NaCl 溶液重悬；
- 16) 根据实验安排，提前准备多个离心管，均加入 90 微升 0.9%NaCl 溶液；
- 17) 吸 10 微升悬浮好的溶液到加好 90 微升 0.9%NaCl 溶液的离心管中，吹打混匀；
- 18) 吸取上一步骤中重悬好的溶液 10 微升加入 90 微升 0.9%NaCl 溶液的离心管中，吹打混匀；
- 19) 多次重复，吸取上一步骤中重悬好的溶液 10 微升加入 90 微升 0.9%NaCl 溶液的离心管中，吹打混匀，完成对菌液的梯度稀释；
- 20) 从装有不同浓度梯度菌液的离心管中取 3 微升，滴斑，滴在四缺固体培养基上；
- 21) 在 28 度-30 度的培养箱中，倒置培养 3 天，观察生长情况。

### 3.3 研究结果

#### 3.3.1 正选择位点氨基酸保守性分析

利用 4 个基因 (*rps2*, *ccsA*, *ndhd*, *ycf1*) 的正选择位点生成图 3.1 和图 3.2 所示的序列标识。基于 55 个栎属物种的多序列比对结果, 可视化序列相似程度, 序列中每个位置字母高度反映了该位置对应的氨基酸的频率。发现 *rps2* 基因差异较小, *ccsA* 基因位点 89 在常绿树种中存在两种氨基酸类型脯氨酸 (Pro, P) 和苏氨酸 (Thr, T), 而在落叶栎中只有一种氨基酸类型脯氨酸 (Pro, P)。55 种栎属植物 *ndhd* 基因正选择位点 504 在常绿栎中主要为亮氨酸 (Leu, L) 和苯丙氨酸 (Phe, F), 而在落叶栎中主要为亮氨酸 (Leu, L) 和酪氨酸 (Tyr, Y)。 *rps2* 和 *ccsA* 基因的正选择位点在常绿和落叶中比较没有显著的差异。 *ycf1* 基因则存在连续的 5 个正选择位点, 在落叶栎和常绿栎之间存在显著差异: 在落叶栎中一致, 为谷氨酰胺 (Gln, Q)、甘氨酸 (Gly, G)、和 3 个苯丙氨酸 (Phe, F), 即 QGFFF; 而在常绿物种中则除 QGFFF 外, 还存在精氨酸 (Arg, R)、赖氨酸 (Lys, K)、天冬酰胺 (Asn, N)、亮氨酸 (Leu, L) 和缬氨酸 (Val, V) 组成的氨基酸序列, 即 RKNLV。其中谷氨酰胺(Gln, Q)为酸性氨基酸, 甘氨酸(Gly, G)、亮氨酸(Leu, L)、缬氨酸(Val, V)和苯丙氨酸(Phe, F)为中性疏水性氨基酸; 精氨酸(Arg, R)和赖氨酸(Lys, K)为强碱性氨基酸、天冬酰胺(Asn, N)为极性亲水性氨基酸。落叶栎中的氨基酸类型包含酸性和疏水性氨基酸, 而常绿栎中还包含强碱性氨基酸精氨酸和赖氨酸。



图 3.1 55 个栎属物种 *rps2*、*ccsA*、*ndhd* 基因正选择位点序列标识图

Figure 3.1 Sequence logo of positively selected sites for *rps2*, *ccsA*, and *ndhd* genes in 55 *Quercus* species

注: 每个位点的总高度表示氨基酸序列保守性, 每个位点氨基酸的高度表示该位点氨基酸的相对频率

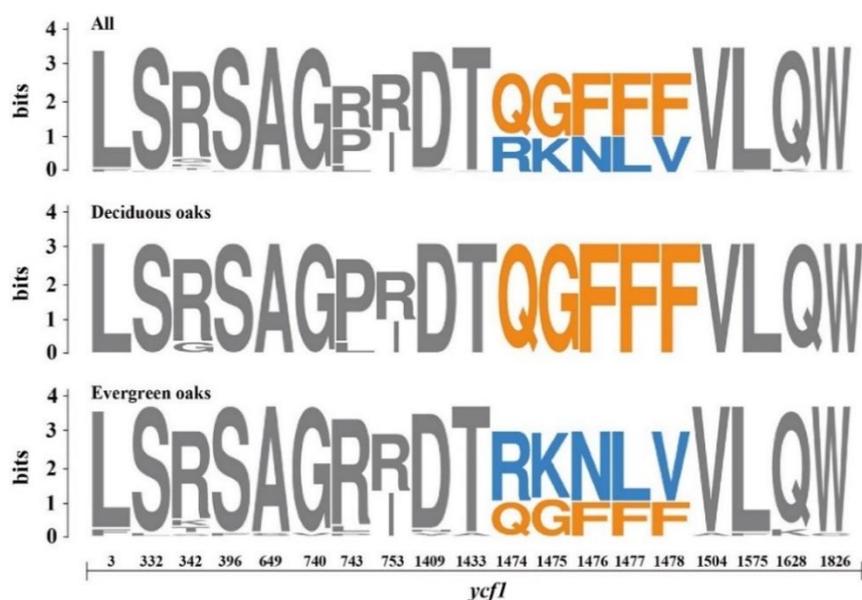


图 3.2 55 个栎属物种 *ycf1* 基因正选择位点序列标识图

Figure 3.2 Sequence logo of positively selected sites for *ycf1* gene in 55 *Quercus* species

注：每个位点的总高度表示氨基酸序列保守性，每个位点氨基酸的高度表示该位点氨基酸的相对频率

### 3.3.2 *ycf1* 基因的生物信息学分析

本研究计算了 55 种栎属植物的 *ycf1* 基因的正选择位点的等电点，发现常绿栎和落叶栎中氨基酸的酸碱性存在差异：落叶栎的等电点明显小于常绿栎（图 3.3），多数落叶栎等电点小于 7，而大多数常绿栎等电点大于 7。

因为序列相似性分析显示差异最大的是 *ycf1* 基因的 5 个连续的正选择位点，所以本研究首先分析这 5 个连续且高度不同的氨基酸位点附近的 62 个氨基酸序列。统计 55 个种中这条氨基酸序列的类型后共发现 6 种氨基酸序列（表 3.2），其中，H1 中常绿栎和落叶栎均有，H2、H3 和 H4 均为落叶栎，H5 和 H6 为常绿栎，H2 和 H5 所包含的物种数量较多，其余序列类型包含的物种数量较少。

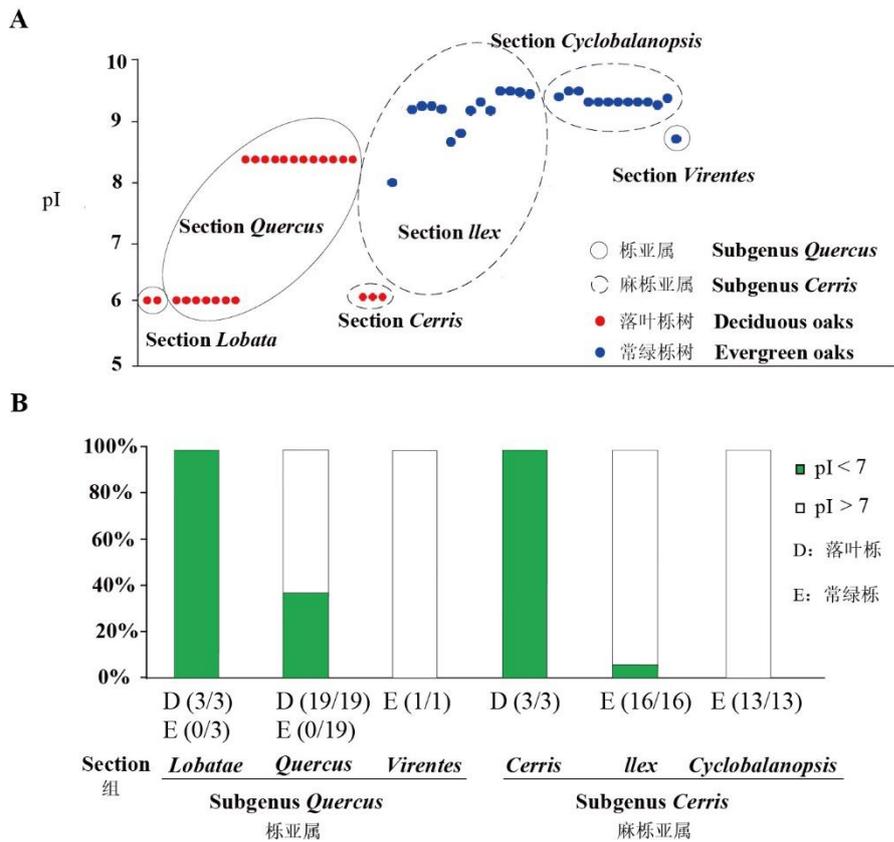


图 3.3 *ycf1* 基因编码蛋白等电点分析

Figure 3.3 Isoelectric point analysis of protein encoded by *ycf1* gene

注：图 A 横坐标为 55 个栎属物种，纵坐标为等电点；图 B：横坐标为不同亚属不同组的栎树，纵坐标为各组中等电点值的统计。白栎组：sect. *Quercus*；红栎组：sect. *Lobatae*；活栎组：sect. *Virentes*

冬青栎组：sect. *Ilex*；青冈栎组：sect. *Cyclobalanopsis*；麻栎组：sect. *Cerris*

图 B 中的数据统计针对于论文中的 55 个栎属物种，D 和 E 代表不同叶习性，D 为落叶习性，E 为常绿习性，括号内是每个组的物种数量以及其中常绿或落叶栎属物种的数量。

表 3.2 55 个栎属物种所包含的 *ycf1* 基因部分序列的类型

Table 3.2 Types of partial sequences of the *ycf1* gene contained in 55 *Quercus* species

	序列 (Sequence)	数量 (Number)	组 (Section)	物种 (Species)
H1	DKIDKKGLFYLTIHQDQKINPS	19	sect. <i>Quercus</i>	<i>Q. tungmaiensis</i> <i>Q. engleriana</i> <i>Q. senescens</i> <i>Q.</i>
	NQQGFFFDWMGMNNEEILSHPISNL		sect. <i>Lobatae</i>	<i>myrsinifolia</i> <i>Q. baronii</i> <i>Q. coccinea</i> <i>Q. rubra</i> <i>Q. petraea</i>
	ELWFFPEFVILYNTYK		sect. <i>Ilex</i>	<i>Q. lobata</i> <i>Q. palustris</i> <i>Q. aquifolioides</i> <i>Q. robur</i> <i>Q.</i>
			sect. <i>Cyclobalanopsis</i>	<i>gambelii</i> <i>Q. macrocarpa</i> <i>Q. stellata</i> <i>Q. dolicholepis</i> <i>Q.</i> <i>bawanglingensis</i> <i>Q. phillyraeoides</i> <i>Q. tarokoensis</i>
H2	DKIDKKGLFYLTIHQDQKMNP	9	sect. <i>Quercus</i>	<i>Q. dentata</i> <i>Q. aliena</i> var. <i>acutiserrata</i> <i>Q. aliena</i>
	SNQQGFFFDWMGMNNEEILSHPISN			<i>Q. fenchengensis</i> <i>Q. wutaishanica</i> <i>Q. mongolica</i>
H3	LELWFFPEFVILYNTYK	4	sect. <i>Quercus</i>	<i>Q. griffithii</i> <i>Q. mongolica</i> subsp. <i>crispula</i> <i>Q. fabri</i>
	DKIDKKGLFYLTIHQDQKMNP			<i>Q. serrata</i> <i>Q. serrata</i> var. <i>brevipetiolata</i> <i>Q. dentata</i>
H4	SNQQGFFFDWMGMNNEEILSHPISN	3	sect. <i>Cerris</i>	subsp. <i>stewardii</i> <i>Q. yunnanensis</i>
	LELWFFPEFVILYNTYK			<i>Q. acutissima</i> <i>Q. chenii</i> <i>Q. variabilis</i>
H5	DKIDKKGLFYLTIHQDQKINPS	19	sect. <i>Ilex</i>	<i>Q. saravanensis</i> <i>Q. spinosa</i> <i>Q. virginiana</i> <i>Q. edithiae</i>
	NQRKNLVDWMGMNNEEILSHPISNL		sect. <i>Cyclobalanopsis</i>	<i>Q. multinervis</i> <i>Q. obovatifolia</i> <i>Q. glauca</i> <i>Q. stewardiana</i>

ELWFFPEFVILYNTYK

sect. *Virentes*

*Q. acuta* *Q. delavayi* *Q. chungii* *Q. sichourensis* *Q. gilva*

*Q. pannosa* *Q. schottkyana* *Q. cocciferoides* *Q.*

*semecarpifolia* *Q. monimotricha* *Q. franchetii*

DKIDKKGLFYLIHQDQKINPS

H6 NQRKNLVDWMGMNEEILSHPISNL

1

sect. *Ilex*

*Q. rehderiana*

ELWFFPEFAILYNTYK

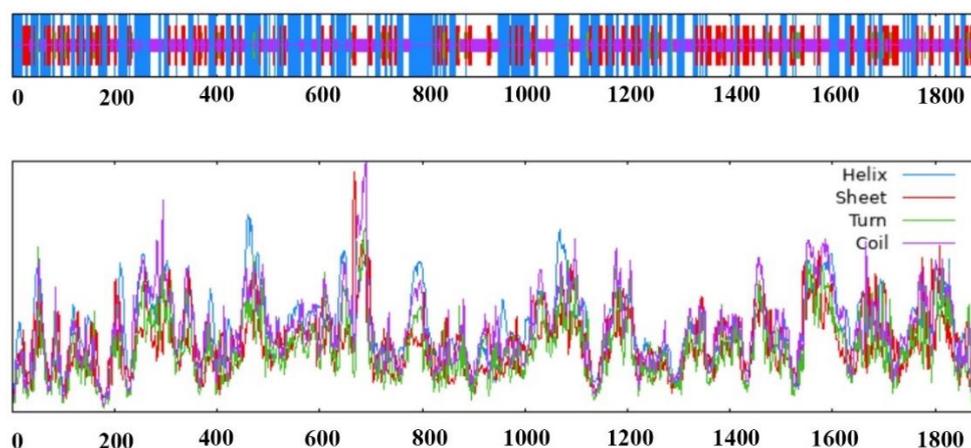
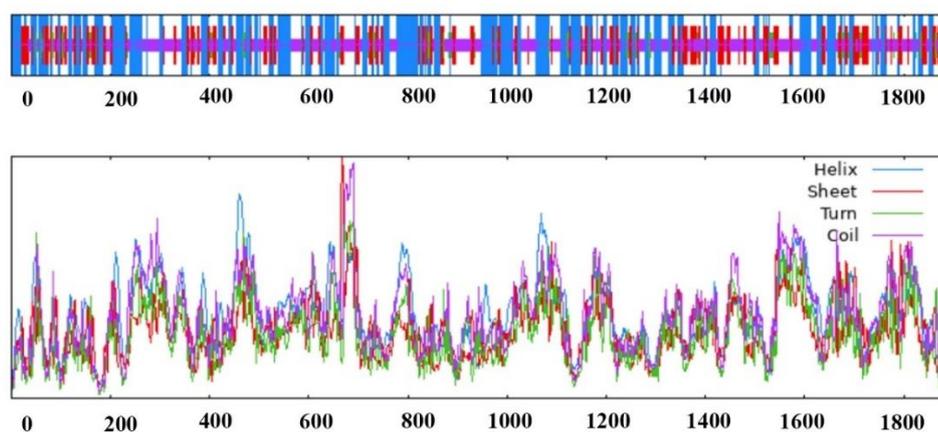
---

由于叶绿体基因组的保守性，选取常绿栎和落叶栎利用 ExPASy ProtParam 软件对 *ycf1* 基因完整序列的基本理化性质如不稳定系数和平均亲水性进行预测。本研究中常绿栎选择刺叶高山栎 (*Q. spinosa*)，落叶栎选择槲栎 (*Q. aliena*)。预测结果显示刺叶高山栎 *ycf1* 基因编码的蛋白理论等电点为 9.59，其中赖氨酸 (Lys) 的数量最多，占比达 11.6%，其次是亮氨酸 (Leu) 和异亮氨酸 (Ile)，占比分别达到了 10.1% 和 9.0%，带正电的残基总数为 327 个，带负电的残基总数为 227 个，不稳定系数为 42.39，平均亲水性为 -0.571，为亲水不稳定蛋白。对槲栎 *ycf1* 基因的基本理化性质预测结果显示 *ycf1* 编码蛋白理论等电点为 9.52，赖氨酸的数量最多占比达 11.6%，亮氨酸和异亮氨酸占比分别达到了 10.3% 和 8.9%，带正电的残基总数为 324 个，带负电的残基总数为 234 个，不稳定系数为 43.4，平均亲水性为 -0.580，为亲水不稳定蛋白 (表 3.3)。

表 3.3 常绿和落叶栎树 *ycf1* 基因的基本理化性质预测Table 3.3 Prediction of basic physicochemical properties of the *ycf1* gene of deciduous and evergreen oak

理化性质 (Physicochemical properties of proteins)	常绿栎树 (Evergreen oak)	落叶栎树 (Deciduous oak)
理论等电点	9.59	9.52
氨基酸总数	1885	1885
不稳定系数	42.39	43.4
平均亲水性	-0.571	-0.580

本论文选择了刺叶高山栎和槲栎为常绿栎和落叶栎的代表，对 *ycf1* 基因的二级结构进行预测，发现常绿栎中  $\alpha$ -螺旋 (Alpha helix) 和无规则卷曲 (Random coil) 较多，其次是延伸链 (Extended strand) 和  $\beta$ -折叠 (Beta turn)。常绿栎中  $\alpha$ -螺旋最多，有 747 个氨基酸，占比 39.6%；无规则卷曲有 697 个氨基酸，占 36.9%；延伸链由 372 个氨基酸组成，占比 19.7%； $\beta$ -折叠只有 69 个氨基酸，占比 3.7% (图 3.4)。落叶栎预测同样发现含有较多的  $\alpha$ -螺旋和无规则卷曲， $\alpha$ -螺旋有 752 个氨基酸，占比 39.89%；无规则卷曲有 715 个氨基酸，占 37.93%；延伸链和  $\beta$ -折叠较少，延伸链由 347 个氨基酸组成，占比 18.41%， $\beta$ -折叠只有 71 个氨基酸，占比 3.77% (图 3.5)。

图 3.4 常绿栎树 *ycf1* 基因二级结构预测Figure 3.4 Prediction of secondary structure of *ycf1* gene in evergreen oak图 3.5 落叶栎树 *ycf1* 基因二级结构预测Figure 3.5 Prediction of secondary structure of *ycf1* gene in deciduous oak

对常绿和落叶栎的跨膜结构域进行预测，落叶栎选择槲栎 (*Q. aliena*)，常绿栎选择刺叶高山栎 (*Q. spinosa*)，利用 TMHMM 对 *ycf1* 基因跨膜结构域进行预测后发现均预测到 6 个跨膜螺旋 (图 3.6)。利用 SMART 预测结构域同样预测到 6 个跨膜结构域 (表 3.4)。且刺叶高山栎 (*Q. spinosa*) 和槲栎 (*Q. aliena*) 结构域预测结果及具体位置完全一致。

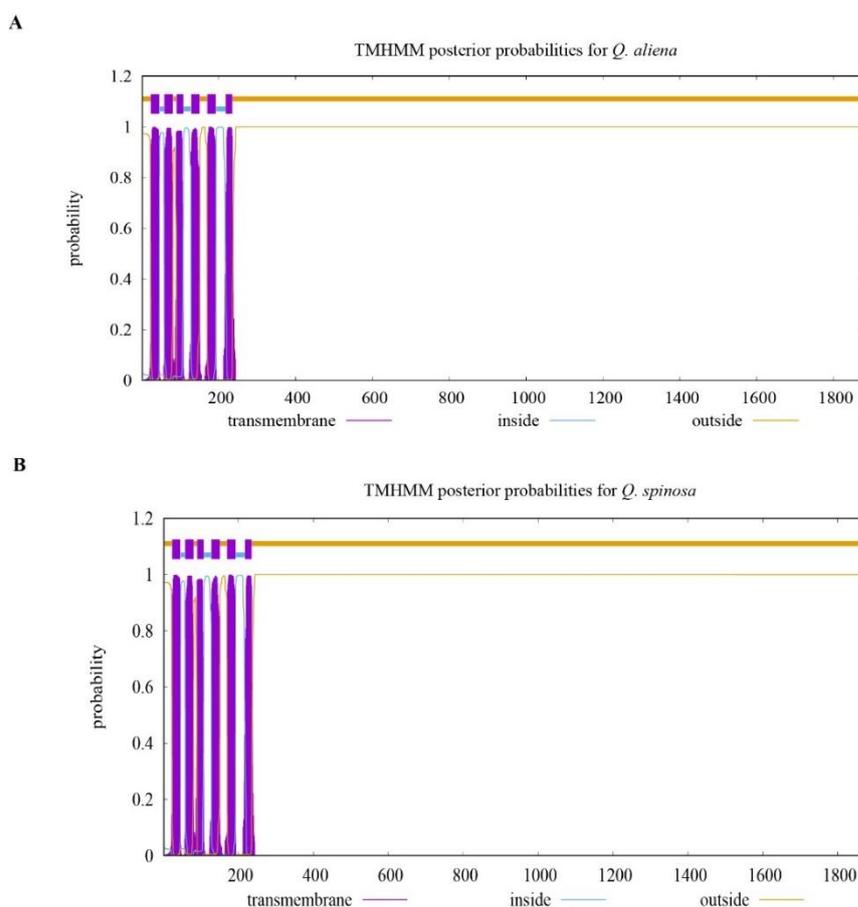


图 3.6 常绿和落叶栎 *ycf1* 基因跨膜结构域预测

图 A: 落叶栎 *ycf1* 基因跨膜结构域预测; 图 B: 常绿栎 *ycf1* 基因跨膜结构域预测

Figure 3.6 Prediction of *ycf1* gene transmembrane domain in evergreen and deciduous oaks

Figure A: Prediction of *ycf1* gene transmembrane domain in deciduous oak; Figure B: Transmembrane domain prediction of *ycf1* gene in evergreen oak

表 3.4 *ycf1* 基因 6 个跨膜结构域的具体位置

Table 3.4 Specific locations of the six transmembrane structural domains of *ycf1* gene

名称	起始	结束
跨膜结构域	23	45
跨膜结构域	58	80
跨膜结构域	90	107
跨膜结构域	128	150
跨膜结构域	170	192
跨膜结构域	218	235

常绿和落叶栎属物种氨基酸序列的差异可能会对其结构产生影响。由于 *ycf1* 基

因的序列较长，没有成功地预测出其完整的三维结构，只对差异较大部分的序列的三维结构利用 AlphaFold2 进行了预测（图 3.7）。对这 6 种序列的三维结构进行预测后发现落叶栎和部分常绿栎的三维结构存在显著差异。H5 和 H6 更加相似，H2、H3 和 H4 更为相似。H1 虽然常绿和落叶栎均有，但结构上与 H2、H3 和 H4 更相似。这可能与氨基酸序列均为 QGFFF 有关。氨基酸类型和结构的改变，可能会影响蛋白质的功能。考虑到 *ycf1* 是一种膜蛋白而且是 TIC 的主要成员，在蛋白质进入叶绿体中起着重要作用。因此，本研究认为氨基酸序列的不同可能会影响进入叶绿体的蛋白质与 Tic214 的相互作用强度。

基于预测的三维结构，对这 6 种序列的静电势也进行了分析，进一步揭示这些结构变化可能带来的影响。结果发现在本论文所研究的氨基酸片段周围，H1、H2、H3 和 H4 带负电，而 H5 和 H6 带正电。由于转运肽带正电，预期与 H1、H2、H3 和 H4 的相互作用可能更强（图 3.7）。

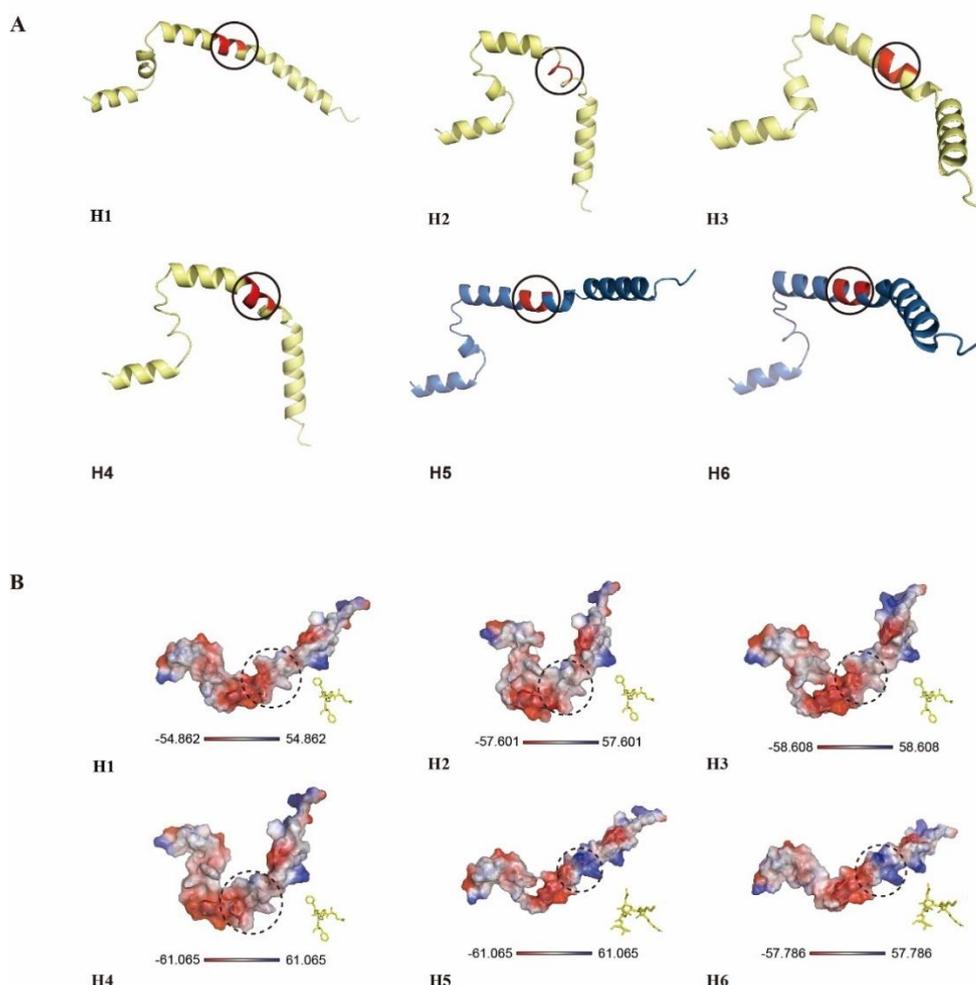


图 3.7 *ycf1* 基因 6 种序列类型三维结构预测及静电势分析

图 A: *ycf1* 基因 6 种序列类型三维结构预测; 图 B: *ycf1* 基因 6 种序列类型静电势分析

Figure 3.7 Three-dimensional structure prediction and electrostatic potential analysis of six sequence types of *ycf1* gene

Figure A: Three-dimensional structure prediction of 6 sequence types of *ycf1* gene; Figure B: Electrostatic potential analysis of 6 sequence types of *ycf1* gene

基于 *ycf1* 基因的序列构建了 55 个栎属物种的系统发育树，并将对应的 *ycf1* 基因 6 种序列类型具体的 62 个氨基酸序列比对后放在系统发育树右侧比较。建树采用了最大似然法 (ML) 和邻接法 (NJ)。ML 树构建前利用 MEGA 计算出最合适的模型是 JTT+G+I+F (图 3.8, 图 3.9)。

通过建树和序列比对分析，发现 RKNLV 和 QGFFF 这两种氨基酸序列类型在栎属植物中表现出明显的聚类趋势。RKNLV 序列对应的栎属植物包含青冈栎组和冬青栎组以及活栎组。QGFFF 序列对应的栎属植物包含冬青栎组、白栎组、麻栎组以及红栎组。冬青栎组和青冈栎组均存在两种序列类型，但冬青栎组两种序列 QGFFF 和 RKNLV 的出现频率均很高，分别为 56% 和 44%。青冈栎组则以 RKNLV 序列为主，只有一个栎属植物小叶青冈 (*Q. myrsinifolia*) 氨基酸序列为 QGFFF。

从常绿和落叶的角度来看，常绿植物中存在两种序列类型，大多数常绿植物对应氨基酸序列均为 RKNLV。而落叶植物对应的氨基酸序列全部为 QGFFF。弗吉尼亚栎 (*Q. virginiana*)，是常绿栎，但属于栎亚属，对应的序列仍然是 QGFFF。

对基于 *ycf1* 基因利用最大似然法和邻接法构建的系统发育树进行比较，发现整体上都被较好的聚类为两个亚属和六个组。基于 NJ 方法构建的系统发育树与基于 ML 方法构建的系统发育树存在一定的差异。特别是麻栎组的位置，NJ 树显示冬青栎组被较好的聚类，而 ML 树显示麻栎组嵌套在冬青栎组中。在 ML 树和 NJ 树中，麻栎组的三个物种和匙叶栎 (*Q. dolicholepis*)、檀子栎 (*Q. baronii*)、乌冈栎 (*Q. phillyraeoides*)、太鲁阁栎 (*Q. tarokoensis*)、坝王栎 (*Q. bawanglingensis*) 的关系均较近。

在对 55 个栎属物种的 *ycf1* 基因 6 种序列类型的 62 个氨基酸序列比对中，还发现了一些特有的变异位点。比如麻栎组的三个物种在该段序列第 41 位氨基酸均为天冬氨酸(Asp)，而剩下的 52 个栎属植物均为组氨酸(His)。云南波罗栎(*Q. yunnanensis*)、枹栎(*Q. serrata*)、短柄枹栎(*Q. serrata* var. *brevipetiolata*) 和黄山栎(*Q. dentata* subsp. *stewardii*) 四个物种则在 62 个氨基酸序列第 60 位氨基酸处均为丝氨酸(Ser)，其余物种均为苯丙氨酸(Phe)。白栎组部分物种该段序列第 19 位氨基酸为甲硫氨酸(Met)，其余的栎属物种则为异亮氨酸(Ile)。

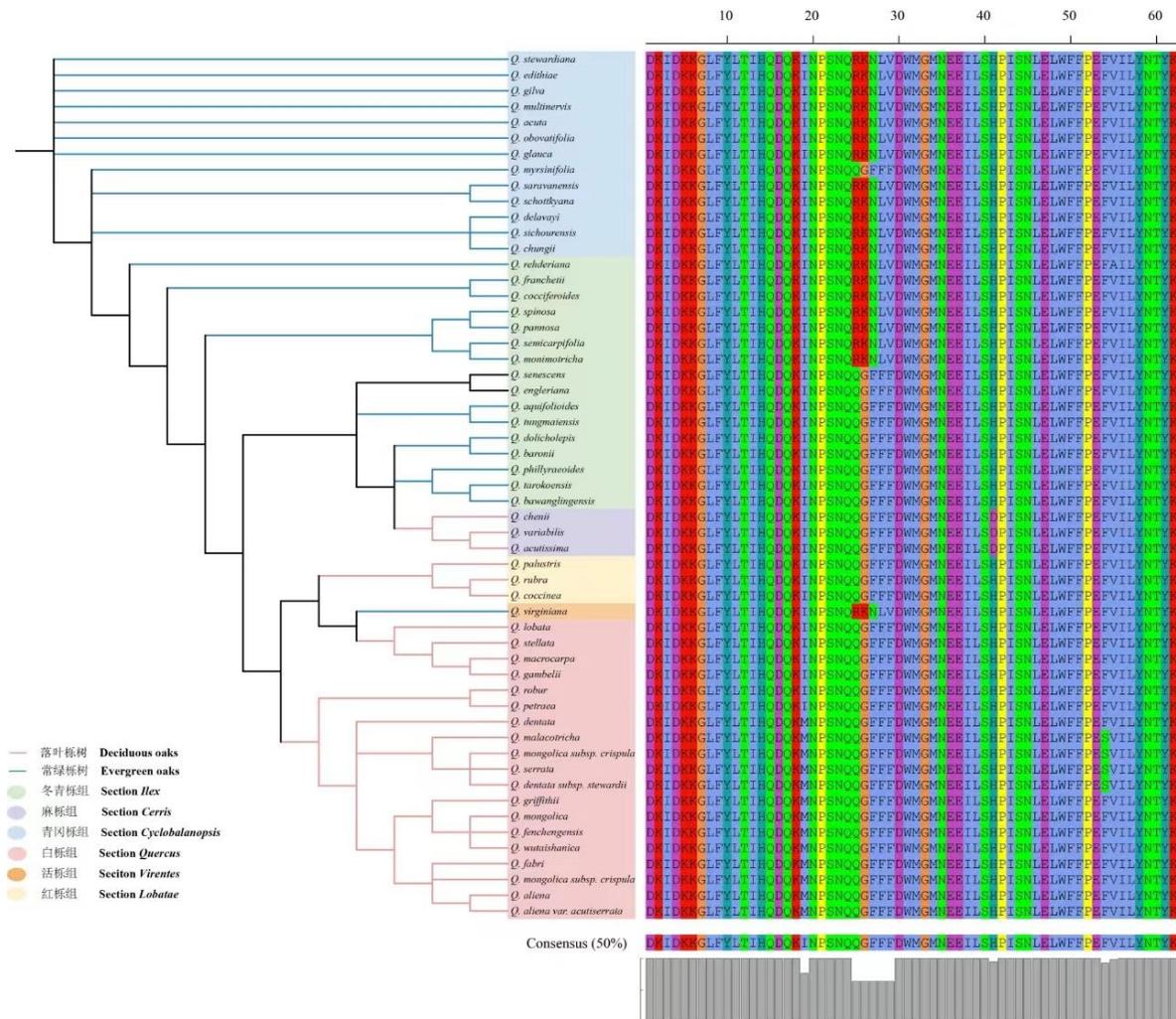


图 3.8 55 个栎属物种 *ycf1* 基因序列邻接法构建的系统发育树

Figure 3.8 Neighbor-Joining tree constructed from *ycf1* gene sequences of 55 *Quercus* species

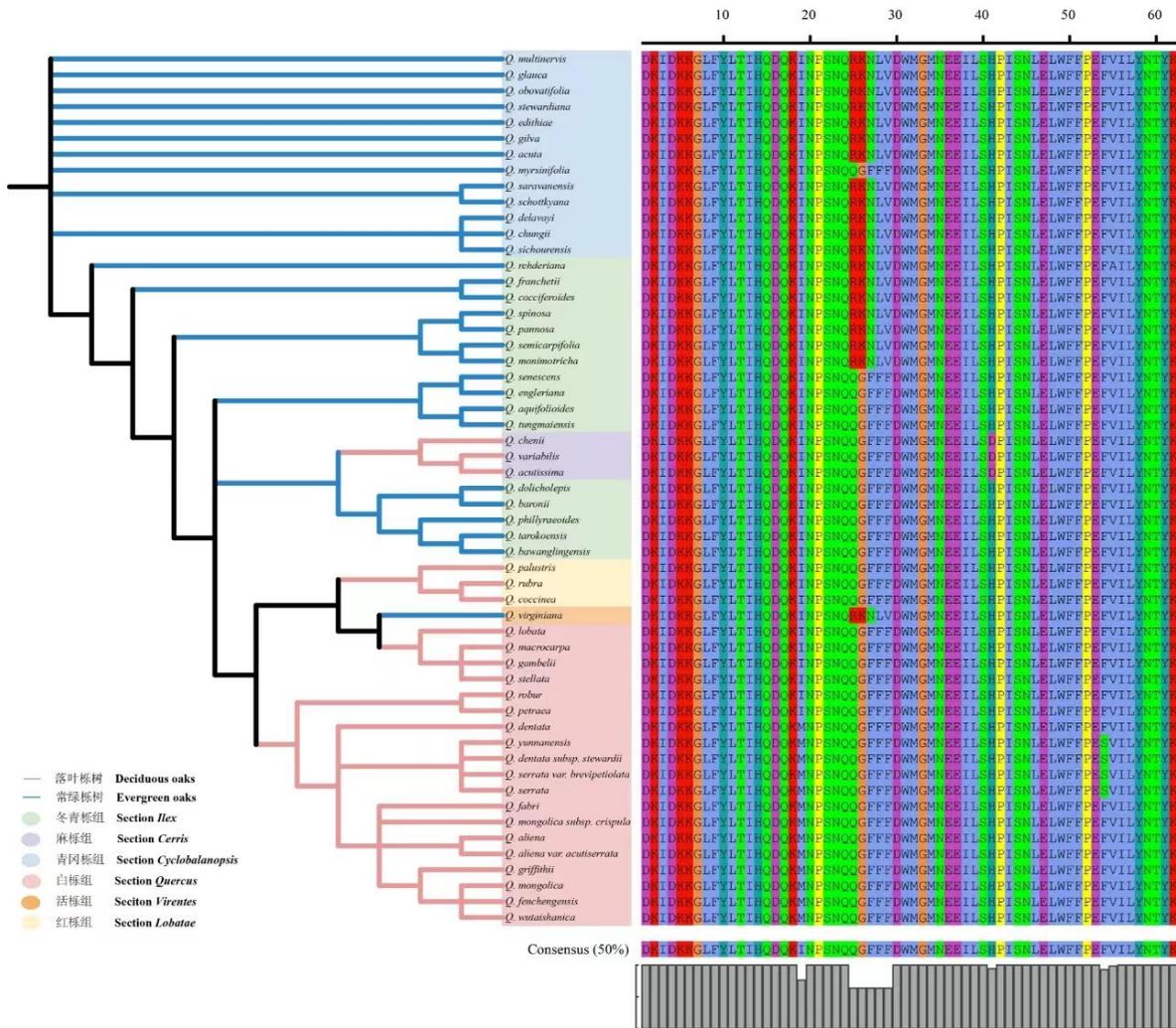
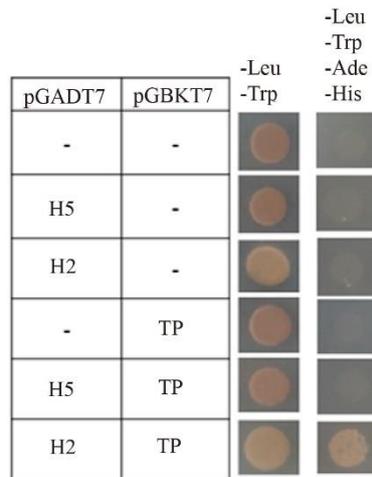


图 3.9 55 个栎属物种 *ycf1* 基因序列最大似然法构建的系统发育树  
 Figure 3.9 Maximum likelihood tree constructed from *ycf1* gene sequences of 55 *Quercus* species

### 3.3.3 *ycf1* 基因功能验证

为了探究落叶栎和常绿栎中叶绿体转运肽与 *ycf1* 蛋白片段的相互作用，首先根据文献，找到了预测的 Rubisco Activase 酶的转运肽序列。利用 Rubisco Activase 酶的转运肽与常绿栎和落叶栎中数量最多的序列类型 H2 和 H5 做了酵母双杂实验验证，将转运肽连接到载体 pGBKT7 上，*ycf1* 的 2 种氨基酸序列连接到载体 pGADT7 上。观察实验结果发现存在显著差异。在四缺培养基上，落叶栎对应的序列存在较强的相互作用，而常绿栎对应的序列，相互作用强度较弱（图 3.10）。

图 3.10 *ycf1* 基因部分序列与 Activaste 转运肽序列酵母双杂实验结果Figure 3.10 Results of yeast two-hybrid assay of partial sequence of *ycf1* gene and activaste transit peptide sequence

注：H5、H2 为 *ycf1* 基因部分序列的两种类型,H5 为常绿栎, H2 为落叶栎; TP 是 Rubisco activaste 酶的转运肽序列

之后下载了 Holbrook 等<sup>[82]</sup>2016 预测的拟南芥 Rubisco 小亚基的转运肽序列, 并在 NCBI 数据库中进行了比对, 确定了两种栎(*Q. aquifolioides*, *Q. lobata*)的 rubisco 小亚基转运肽序列, 共找到三条转运肽序列, 将 Rubisco Activaste 转运肽序列命名为 TP, 将 *Q. aquifolioides*, *Q. lobata* 的 rubisco 小亚基转运肽序列命名为 QATP 和 QLTP。将三条转运肽和统计的 *ycf1* 基因该片段所有类型的序列设计实验进行验证。

将这三条转运肽连接到载体 pGBKT7 上, *ycf1* 基因的 6 种氨基酸序列连接到载体 pGADT7 上。由于 6 种序列类型中不止存在 QGFFF 与 RKNLV 的差异, 因此增加了两种替换后的序列。将 H5 序列中的 RKNLV 替换为 QGFFF, H2 序列中的 QGFFF 替换为 RKNLV。由于 H5 替换后的序列与 H1 相同, 因此没有单独命名。H2 替换后序列命名为 H7。利用 Y2H Gold 酵母感受态将转运肽与 *ycf1* 的不同的组合转化到酵母细胞后进行酵母双杂实验。

二缺和四缺培养基上酵母细胞的生长结果表明, 三组转运肽与该序列均有不同强度的相互作用(图 3.11)。从同一转运肽与不同 *ycf1* 序列来看, H1、H2、H3 和 H4 与转运肽的相互作用较强, H5 的相互作用最弱, H6 的相互作用较弱。即整体上, 落叶栎的相互作用强度要大于常绿栎(图 3.11)。然而, 研究发现 H5 替换后与 H1 相同后, 序列的相互作用较强。H2 替换后 H7 仍然存在相互作用。考虑到 H2 替换后的序列目前并没有发现, 且影响其相互作用的位点可能不止一个。H6 尽管是常绿栎, 却仍然

有相互作用，考虑到只存在一条，可能是特例。总之，QGFFF 氨基酸序列类型比 RKNLV 氨基酸序列有更强的相互作用。从不同转运肽与同一 *ycf1* 基因序列来看，Rubisco 小亚基转运肽的相互作用强度大于 Activate 酶。

由于转运肽与 *ycf1* 基因这段序列之间的相互作用存在差异，叶绿体是植物光合作用所必需的细胞器，且不同叶习性的栎树在叶片性状指标上也存在差异<sup>[33]</sup>，落叶栎叶片有更低的 LMA，更高的光合速率，而常绿栎叶片有更高的 LMA，低的光合速率。因此本论文假设 *ycf1* 基因这段序列对应的氨基酸的差异可能影响到了光合作用中蛋白的转运过程，从而对不同叶习性栎属物种的光合效率产生了影响。

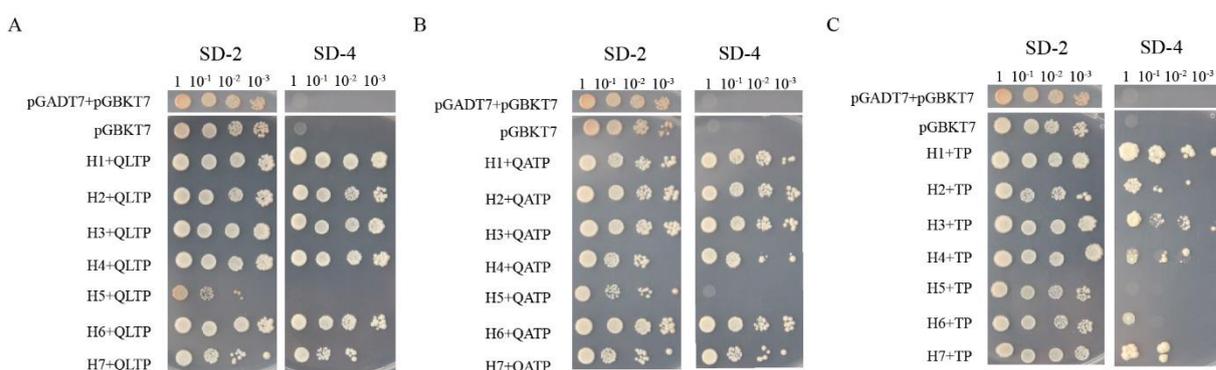


图 3.11 *ycf1* 基因 7 种序列与 3 种转运肽序列酵母双杂实验结果

Figure 3.11 Results of yeast two-hybrid assay of seven sequences of *ycf1* gene and three transit peptide sequences

注：图 A、B、C 为不同转运肽与 *ycf1* 基因序列组合的酵母双杂实验结果

光合作用速率是指植物在单位时间内通过光合作用固定碳的能力，受到多种因素的影响，包括叶绿素浓度、酶活性以及参与光合作用的其他蛋白质的种类和数量。落叶栎与常绿栎在相同时间内进入叶绿体的蛋白质数量可能存在差异。结合之前的分析，本论文提出了一个影响不同叶习性栎树光合作用效率的模型：落叶栎和常绿栎 *ycf1* 基因氨基酸序列的差异可能使进入叶绿体的蛋白质与 *ycf1* 基因编码蛋白 Tic214 的相互作用强度发生变化。导致不同叶习性栎树在相同时间内进入叶绿体的蛋白质的数量不同，从而光合作用速率不同（图 3.12）。在相同条件下，落叶栎有更强的相互作用，所以在相同的时间内可以通过更多的蛋白，拥有更高的光合速率。而常绿栎则相反，相互作用较弱，在相同的时间内通过蛋白的数量较少，光合速率较低。

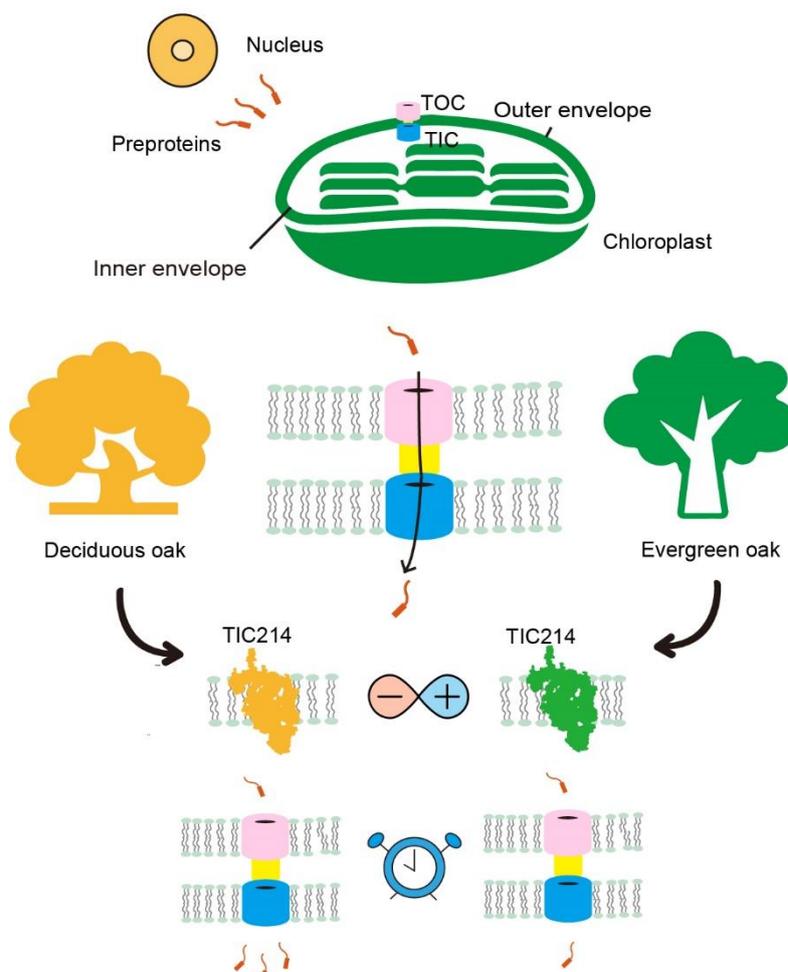


图 3.12 *ycf1* 基因影响不同叶习性栎树光合作用效率的模型

Figure 3.12 Modeling of *ycf1* gene effects on photosynthetic efficiency in oaks with different leaf habit

## 4 讨论

### 4.1 栎属叶绿体基因组序列差异比较

本论文首先对 14 个栎属植物的叶绿体基因组进行了比较,发现叶绿体基因组整体高度保守,基因的顺序、功能、结构均相似。在整体大小和各分区的大小详细比较方面,落叶栎间彼此更为接近,而几个常绿栎间则变化较大。大小上存在的差异主要是由于 IR 区的扩张和收缩引起的,对各分区的边界进行可视化分析后发现,边界处基因的长度变化,同样在落叶栎中较为相似,而常绿栎差异较大。重复序列在物种的遗传变异和进化中起着极其重要的作用,对叶绿体基因组中的多种重复序列进行分析后发现,简单重复序列中单核苷酸重复是最多的,其次是二核苷酸重复,且重复类型多为 A/T 重复和 AT/TA 重复。这与多位研究者在栎属物种和许多被子植物中的研究结果一致<sup>[15,37]</sup>。六核苷酸重复出现是最少的,多个常绿栎属物种中不存在六核苷酸重复。黄婷<sup>[85]</sup>对青冈组植物的研究指出,多个常绿青冈植物没有六核苷酸重复。正向重复、反向重复、回文重复和互补重复,四种重复类型中,反向重复是最少的,在多个常绿树种中甚至没有出现。以榲子栎叶绿体基因组为参考基因组,可视化序列差异后发现,在基因组的差异上,SC 区的变异更大,IR 区的变异更小。由于自然选择,编码区往往表现出更高的保守性,变异较小,但 *ycf1* 基因存在较大差异。同时对这 14 个栎属物种的密码子偏好性进行分析后发现存在较弱的密码子偏好,RSCU>1 的多个密码子中,90%是 A/U 结尾的,对所有密码子统计后也发现,密码子第三位多数为 A/T,少数为 G/C。这与被子植物中的多项研究结果均一致<sup>[86,87]</sup>。

### 4.2 栎属叶绿体基因组系统发育

栎属由于广泛的杂交渐渗、趋同进化等原因,它的系统发育研究一直备受关注。本论文利用 55 个栎属物种的叶绿体基因组数据和 *ycf1* 基因序列利用最大似然法和邻接法分别构建了系统发育树。结果将这 55 种植物分为栎亚属和麻栎亚属两个亚属,麻栎亚属分为青冈栎组、麻栎组、冬青栎组共三组,栎亚属分为白栎组、活栎组和红栎组。比较两种方法构建的系统发育树发现,无论是基于整个叶绿体基因组还是 *ycf1* 基因,ML 树的结果麻栎组嵌套在冬青栎组当中。此前的研究同样支持了麻栎组和青冈栎组嵌套在冬青栎组中。可能是由于这三个部分的祖先谱系之间的不完全谱系分化或渐渗导致了这种复杂的模式<sup>[37]</sup>。这与许多基于质体基因组的研究结果也相似,但与 Hipp 等<sup>[88]</sup>基于 RAD-seq 数据推断的系统发育关系不同。但基于叶绿体基因组和 *ycf1* 构建的 ML 树麻栎组的位置以及冬青栎组部分物种的位置发生了改变。在本研究中,基于栎属物种整个叶绿体基因组和 *ycf1* 基因的深入分析,本论文不仅构建了详细的系统发育树,还结合序列比对结果发现了一些有趣的遗传现象和变异位点。在本论文

所重点研究的 *ycf1* 基因的部分序列中，冬青栎组植物存在 QGFFF 和 RKNLV 两种序列类型，可能是冬青栎组植物适应性进化的结果。部分白栎组物种则在这部分序列的第 19 个氨基酸为甲硫氨酸 (Met)，其余的栎属物种则为异亮氨酸 (Ile)，这也揭示了白栎组内部的遗传多样性。这些发现不仅有助于理解栎属植物的分类和演化关系，还为未来的研究提供了重要的基础数据和参考方向。

### 4.3 *ycf1* 基因功能验证

明确了系统发育关系后，本研究利用位点模型和分支位点模型对 55 个栎属物种的正选择基因进行了检测。基因在进化过程中可能会因为正选择的作用而加速进化。在研究中，发现了多个正选择基因，这些基因的变异和选择压力可能与栎属植物的环境适应性及其在不同生态系统中的角色有关。尤其是 *ycf1* 基因，研究中显示出较多的正选择位点，并且在多项研究中被认为是植物的 DNA 条形码，具有正选择基因的特征。此外，在对常绿和落叶栎属物种的叶绿体基因组序列进行比较分析时，就已经发现 *ycf1* 基因在常绿栎和落叶栎中存在诸多差异。Dong 等<sup>[89]</sup>研究也指出，*ycf1* 编码了 1 800 多个氨基酸，在叶绿体基因组中的 SSC 区具有较高的序列变异性。甚至超过了之前大多数研究中发现的变异较高的 *matK* 基因。*ycf1* 中的两个区域，*ycf1a* 和 *ycf1b*，已被预测在被子物质体基因组的物种水平上具有最高的核苷酸多样性。但对 *ycf1* 基因功能的研究却较少，分析后认为 *ycf1* 是功能基因，对植物生长发育至关重要<sup>[43]</sup>。但对具体的功能一致没有明确的说法。

*ycf1* 基因编码 Tic214 蛋白，在叶绿体蛋白转运过程中扮演重要角色。Kikuchi 等<sup>[48]</sup>通过拟南芥中 Tic20-i 的转基因植物，纯化了与之相互作用的复合物，发现 3 个分子量分别为 214、100 和 56 kD 的蛋白。经证实与 Tic20-i 共同形成稳定的 1-MD 复合物。Kikuchi 和 Nakai 等<sup>[44,48,55,72]</sup>的研究报告详细指出了拟南芥叶绿体的中存在一个功能性的 1 兆达尔顿(Mda)蛋白易位复合体，由 Tic214、Tic100、Tic56 和 Tic20 组成，但不包括 Tic110<sup>[44,48,54,55,72]</sup>。

蛋白转运复合体在叶绿体转运过程中发挥着重要作用，叶绿体在进化过程中，大多数基因被转移到细胞核中，有 2 000-3 000 多个蛋白需要在细胞质中合成后运输到叶绿体中，这需要通过叶绿体蛋白的运输机制来实现。蛋白质的运输是叶绿体正常功能的重要组成部分，缺少运输蛋白或者信号肽序列错误都可能导致光合作用反应的受损或中断<sup>[45]</sup>。TOC-TIC 复合体的形成和功能需要多个蛋白质组分之间密切协作，确保叶绿体内蛋白质的正确转运和定位。细胞质中这些前体蛋白含有引导它们进入叶绿体的 N 端转运肽，有助于叶绿体外膜 (TOC) 和内膜 (TIC) 上的转位点复合体的识别<sup>[90]</sup>。一般来说，各种转运肽之间没有一致的保守序列，但存在一些共同的特征：在大多数情况下，它们带正电荷，含有少量酸性氨基酸残基，也含有丰富的羟基化氨基酸残基，特别是丝氨酸<sup>[91,92]</sup>。

前体蛋白穿过 TOC 复合体时, 会在膜间隙和内包膜中与 TIC 组分结合<sup>[80]</sup>。Tic20/56/100/214 组成的一般 TIC 复合体可能负责主要的光合蛋白运输, 因此可能优先与已知的由 Toc159、Toc33 和 Toc75 组成的光合型 TOC 复合体发生物理和功能相互作用。Kessler 等<sup>[93]</sup>的研究也认为 TOC 复合体与新的 TIC 复合体组分(包括 Tic56, Tic100 和 Tic214/*ycf1*)之间存在强烈的物理相互作用。此外, Tic100 中的两个 MORN 基序可能介导膜之间的连接, 有人提出 MORN 重复序列通过与脂质相互作用来促进膜结合<sup>[82]</sup>。共免疫沉淀实验清楚地揭示了 Tic214 与该藻类中 TIC 和 TOC 复合体的组分相互作用。此外, 衣原体体外导入叶绿体蛋白的研究证实, Tic214 是蛋白转运复合体的一部分, 在导入反应中 ATP 的存在刺激了它与蛋白转运复合体的关联<sup>[5]</sup>。这些研究表明 Tic214 也可能与前体蛋白的转运肽存在相互作用。而常绿和落叶柞氨基酸的差异, 可能会影响 Tic214 蛋白的结构和功能, 影响 Tic214 与转运肽的相互作用。本论文利用酵母双杂实验对 *ycf1* 基因功能进行了验证。实验结果表明不同的氨基酸序列与 *ycf1* 的相互作用强度不同, 可能会影响到叶绿体中蛋白的输送效率, 从而影响植物光合作用的速率, 提出了可能影响不同叶习性柞树光合速率的模型。

*ycf1* 基因在多项研究中显示出较大的差异和多个正选择位点, 表明它在柞属植物的适应性进化中可能发挥了重要作用。*ycf1* 基因中存在的两种类型的氨基酸序列, 有可能影响到蛋白质的功能, 这些功能可能与物种对环境的适应有关, 从而被保留。蛋白转运复合体, 包括 Tic214 蛋白在调节叶绿体的生物发生和稳态中起着直接作用, 帮助植物抵御生物和非生物胁迫, 包括光、盐、干旱和渗透胁迫等<sup>[94]</sup>。这些机制有助于解释常绿柞和落叶柞如何通过不同的叶习性来优化光合作用效率, 以及它们如何适应不同的环境条件。

## 5 结论和展望

### 5.1 结论

本研究首先对 14 个栎属物种的叶绿体基因组进行了全面的比较分析, 随后, 本论文整合了来自 NCBI 的 41 个栎属物种的数据, 通过最大似然和邻接法构建了 55 个栎属物种的系统发育关系。在此基础上, 运用 PAML 软件包中的位点模型和分支位点模型进行了正选择分析, 并对检测出的正选基因进行了生物信息学分析。重点关注了 *ycf1* 基因的功能, 并通过酵母双杂实验对基因功能进行了验证, 提出了可能影响不同叶习性栎属植物光合效率的分子机制。主要结论如下:

1. 从基因组大小, 各分区的大小, 重复序列的类型等方面来看, 落叶栎整体上更稳定, 数值更接近, 变化更小。相比之下, 常绿栎的变化较大。在重复序列方面, 有多个常绿栎不存在六核苷酸重复, 而落叶栎均有。序列相似度分析显示, 落叶栎与常绿栎存在几处区别为同种叶习性的栎树共同存在的。

2. 利用分支位点模型和位点模型对 55 个栎属物种进行正选择分析, 位点模型检测出 *rps12*、*rpl16*、*rpoC2*、*ndhB*、*ndhD*、*ndhF*、*ndhK*、*petD*、*rbcL*、*ccsA*、*ycf1*、*ycf2* 和 *ycf3* 多个基因。分支位点模型以常绿栎作为前景枝, 发现四个受到正选择的基因 *rps2*、*ccsA*、*ndhD* 和 *ycf1*, 以落叶栎作为前景枝则没有发现受到正选择的基因。这表明叶绿体基因组在栎属物种适应环境变化中起到重要作用。

3. *ycf1* 基因在常绿和落叶树种中存在较大差异, 前期基于 14 个栎属物种叶绿体基因组进行比较时, *ycf1* 在边界处变化较大。在序列差异性分析时也可以看到明显差异, 用两种模型做选择压力分析时, *ycf1* 基因均被认为是受正选择的基因。

4. *ycf1* 基因连续 5 个正选择位点在常绿和落叶栎中存在显著差异, 常绿栎中存在 QGFFF 和 RKNLV 两种类型的氨基酸序列, 而落叶栎中只存在 QGFFF 一种类型的氨基酸序列。常绿栎中, 麻栎亚属青冈栎组植物只有一个物种小叶青冈 (*Q. myrsinifolia*) 氨基酸序列为 QGFFF, 其余均为 RKNLV。而冬青栎组植物则两种氨基酸序列都有, 出现频率均较高。弗吉尼亚栎 (*Q. virginiana*) 为常绿栎树, 氨基酸序列 RKNLV, 属于栎亚属中的活栎组。

5. 基于叶绿体全基因组和 *ycf1* 基因构建的 ML 树和 NJ 树存在差异, 两个 ML 树中麻栎组均嵌套在冬青栎组当中。并且构建的两个 ML 树同样存在差异, 麻栎组以及部分冬青栎组植物的位置发生改变。

6. 酵母双杂实验进行功能验证后结果显示 *ycf1* 基因的部分序列与转运肽之间的相互作用在常绿栎和落叶栎中存在差异。在落叶栎 (QGFFF) 中互作更强, 而在常绿栎 (RKNLV) 中互作能力更弱, 可能会影响不同叶习性栎树的光合作用能力。

7. 提出了可能影响不同叶习性栎属物种光合能力的模型: 由于氨基酸序列不同,

导致 Tic214 蛋白与转运肽的互作能力不同，在相同时间内运输蛋白的效率不同，光合能力存在差异。

本研究不仅对栎属物种的叶绿体基因组进行了全面的比较分析，还探索了其系统发育关系和正选择基因，最后聚焦于 *ycf1* 基因的功能研究。通过对 *ycf1* 基因的深入分析，本论文提出了可能影响不同叶习性栎属植物光合效率的分子机制，这是理解栎属物种适应环境和进化过程中的一个重要发现。

## 5.2 展望

本研究中 55 个栎属物种中栎亚属的 5 个组中只包含了 3 个，且活栎组只有一个物种，为了更全面地理解栎属物种的多样性和系统发育关系，未来可以进一步结合栎亚属其他两个组栎属物种的叶绿体基因组数据进行比较，提供一个更全面的视角。

目前已经对 *ycf1* 基因的部分序列利用酵母双杂实验进行了验证，这为后续的研究提供了坚实的基础。可以进一步利用 *ycf1* 基因的完整序列进行实验，以验证其在栎属中的功能和作用机制。对于 *ycf1* 基因的序列过长问题，可以将其分解为多段较短的序列，通过拼接得到解决。从而有效地对 *ycf1* 基因的完整基因进行单独的实验分析，这对于理解其功能和在不同栎属物种中的差异性表达至关重要。具体来讲，可以设计实验，将扩增得到的 *ycf1* 序列前端加入信号肽和荧光信号后转化到烟草中，利用其瞬时表达的特性，观察在相同时间内，不同的 *ycf1* 序列通过的数量。这样能够定量判断 *ycf1* 在两种叶习性的栎树中是否存在差异，对于理解栎树如何适应不同的环境条件具有重要意义。

除此之外，可以利用 *ycf1* 基因进行多种相互作用实验，如免疫共沉淀、GST pull-down、质谱分析、等离子共振技术等，来定性定量地探索 *ycf1* 在栎属植物中的作用和与其他分子的交互作用。还可以尝试通过基因编辑手段来改变常绿和落叶栎中的 *ycf1* 基因，观察其对光合效率的影响。同时这也为培育高光合效率的植物品种提供了可能性。

另外，之前只对一个正选择基因做了详细分析，现在可以对筛选出的其他正选择基因进行更多分析。例如，NDH 基因在叶绿体光合作用循环电子传递中起着重要作用，可能在植物适应性进化中起到关键作用。通过对这些基因的深入分析，可以更全面地理解栎树是如何适应各种环境的。

## 参考文献

- [1] Sato N. Origin and Evolution of Plastids: Genomic View on the Unification and Diversity of Plastids[M]. *The Structure and Function of Plastids*, Wise R R, Hooper J K, Dordrecht: Springer Netherlands, 2006, 75-102.
- [2] Padmanabhan M S, Dinesh-Kumar S P. All hands on deck—the role of chloroplasts, endoplasmic reticulum, and the nucleus in driving plant innate immunity[J]. *Molecular plant-microbe interactions*, 2010, 23(11): 1368-1380.
- [3] Daniell H, Lin C, Yu M, et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering[J]. *Genome Biology*, 2016, 17(1).
- [4] Cackett L, Luginbuehl L H, Schreier T B, et al. Chloroplast development in green plant tissues: the interplay between light, hormone, and transcriptional regulation[J]. *New Phytologist*, 2022, 233(5): 2000-2016.
- [5] Rochaix J D. Chloroplast protein import machinery and quality control[J]. *The FEBS Journal*, 2022, 289(22): 6908-6918.
- [6] Jarvis P, Lopez-Juez E. Biogenesis and homeostasis of chloroplasts and other plastids[J]. *Nature Reviews Molecular Cell Biology*, 2013, 14(12): 787-802.
- [7] Heber U. Protein synthesis in chloroplasts during photosynthesis[J]. *Nature*, 1962, 195: 91-92.
- [8] Sager R, Ishida M R. Chloroplast DNA in *Chlamydomonas*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1963, 50(4): 725-730.
- [9] Shinozaki K, Ohme M, Tanaka M, et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression[J]. *The EMBO Journal*, 1986, 5(9): 2043-2049.
- [10] 张悦. 红豆杉科叶绿体基因组结构多样化研究[D]. 北京: 北京林业大学, 2019.
- [11] Magee A M, Aspinall S, Rice D W, et al. Localized hypermutation and associated gene losses in legume chloroplast genomes[J]. *Genome Research*, 2010, 20(12): 1700-1710.
- [12] 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(4): 247-261.
- [13] Kugita M, Kaneko A, Yamamoto Y, 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.
- [14] Wicke S, Schneeweiss G M, Depamphilis C W, et al. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function[J]. *Plant Mol Biol*, 2011, 76(3-5): 273-297.
- [15] 许阳. 不同叶习性栎属物种间的叶绿体基因组比较研究[D]. 北京: 北京林业大学, 2021.
- [16] Kumar S, Dhingra A, Daniell H. Plastid-expressed betaine aldehyde dehydrogenase gene in carrot cultured cells, roots, and leaves confers enhanced salt tolerance[J]. *Plant Physiology*, 2004, 136(1): 2843-2854.
- [17] Boyhan D, Daniell H. Low-cost production of proinsulin in tobacco and lettuce chloroplasts for injectable or oral delivery of functional insulin and C-peptide[J]. *Plant Biotechnology Journal*, 2011, 9(5): 585-598.
- [18] Lupala C S, Ye Y, Chen H, et al. Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor[J]. *Biochemical and Biophysical Research Communications*,

2022, 590: 34-41.

[19] Olaru A, Bala C, Jaffrezic-Renault N, et al. Surface plasmon resonance (SPR) biosensors in pharmaceutical analysis[J]. *Critical Reviews in Analytical Chemistry*, 2015, 45(2): 97-105.

[20] Fields S, Song O. A novel genetic system to detect protein-protein interactions[J]. *Nature*, 1989, 340(6230): 245-246.

[21] Bendixen C, Gangloff S, Rothstein R. A yeast mating-selection scheme for detection of protein-protein interactions[J]. *Nucleic Acids Research*, 1994, 22(9): 1778-1779.

[22] Bruckner A, Polge C, Lentze N, et al. Yeast two-hybrid, a powerful tool for systems biology[J]. *International Journal of Molecular Sciences*, 2009, 10(6): 2763-2788.

[23] Chi W, Mao J, Li Q, et al. Interaction of the pentatricopeptide-repeat protein DELAYED GREENING 1 with sigma factor SIG6 in the regulation of chloroplast gene expression in *Arabidopsis* cotyledons[J]. *The Plant Journal*, 2010, 64(1): 14-25.

[24] Chen F, Dong G, Wu L, et al. A Nucleus-Encoded Chloroplast Protein YL1 Is Involved in Chloroplast Development and Efficient Biogenesis of Chloroplast ATP Synthase in Rice[J]. *Scientific Reports*, 2016, 6(1).

[25] Wu X X, Mu W H, Li F, et al. Cryo-EM structures of the plant plastid-encoded RNA polymerase[J]. *Cell*, 2024, 187(5): 1127-1144.

[26] Denk T, Grimm G W, Manos P S, et al. An updated infrageneric classification of the oaks: review of previous taxonomic schemes and synthesis of evolutionary patterns[J]. *Oaks physiological ecology. Exploring the functional diversity of genus *Quercus* L.*, 2017: 13-38.

[27] Huang CJ, Zhang YT, Bartholomew B. *Flora of China*, vol. 4 [in English]. Beijing, China: Science Press, 1999, 314-400.

[28] Hipp A L, Manos P S, González Rodríguez A, et al. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity[J]. *The New phytologist*, 2018, 217(1): 439-452.

[29] Schmerler S B, Clement W L, Beaulieu J M, et al. Evolution of leaf form correlates with tropical-temperate transitions in *Viburnum* (Adoxaceae)[J]. *Proceedings of the Royal Society B: Biological Sciences*, 2012, 279(1744): 3905-3913.

[30] Bai K, He C, Wan X, et al. Leaf economics of evergreen and deciduous tree species along an elevational gradient in a subtropical mountain[J]. *AoB plants*, 2015, 7: v64.

[31] Antúnez I, Retamosa E C, Villar R. Relative growth rate in phylogenetically related deciduous and evergreen woody species[J]. *Oecologia*, 2001, 128(2): 172-180.

[32] Wyka T P, Oleksyn J. Photosynthetic ecophysiology of evergreen leaves in the woody angiosperms – a review[J]. *Dendrobiology*, 2014, 72: 3-27.

[33] Baldocchi D D, Ma S, Rambal S, et al. On the differential advantages of evergreenness and deciduousness in mediterranean oak woodlands: a flux perspective[J]. *Ecological Applications*, 2010, 20(6): 1583-1597.

[34] Kaproth M A, Fredericksen B W, Gonzalez-Rodriguez A, et al. Drought response strategies are coupled with leaf habit in 35 evergreen and deciduous oak (*Quercus*) species across a climatic gradient in the Americas[J]. *New Phytologist*, 2023, 239(3): 888-904.

[35] Furze M E, Wainwright D K, Huggett B A, et al. Ecologically driven selection of nonstructural carbohydrate storage in oak trees[J]. *New Phytologist*, 2021, 232(2): 567-578.

[36] Sancho-Knapik D, Escudero A, Mediavilla S, et al. Deciduous and evergreen oaks show

contrasting adaptive responses in leaf mass per area across environments[J]. *New Phytologist*, 2021, 230(2): 521-534.

[37] Yin K, Zhang Y, Li Y, et al. 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.

[38] Liu X, Chang E, Liu J, et al. Comparative analysis of the complete chloroplast genomes of six white oaks with high ecological amplitude in China[J]. *Journal of Forestry Research*, 2021, 32(5): 2203-2218.

[39] Yang Y, Zhou T, Qian Z, et al. Phylogenetic relationships in Chinese oaks (Fagaceae, *Quercus*): Evidence from plastid genome using low-coverage whole genome sequencing[J]. *Genomics*, 2021, 113(3): 1438-1447.

[40] Wang L, Li Y, Zheng S, et al. Complete Chloroplast Genomes of Four Oaks from the Section *Cyclobalanopsis* Improve the Phylogenetic Analysis and Understanding of Evolutionary Processes in the Genus *Quercus*[J]. *Genes*, 2024, 15(2): 230.

[41] Chen X, Li B, Zhang X. Comparison of chloroplast genomes and phylogenetic analysis of four species in *Quercus* section *Cyclobalanopsis*[J]. *Scientific Reports*, 2023, 13(1).

[42] Martín M, Sabater B. Plastid *ndh* genes in plant evolution [J]. *Plant Physiology and Biochemistry*, 2010, 48(8): 636-645.

[43] Drescher A, Ruf S, Calsa T J, et al. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes[J]. *The Plant Journal*, 2000, 22(2): 97-104.

[44] Nakai M. The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts[J]. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2015, 1847(9): 957-967.

[45] Neuhaus H E, Emes M J. Nonphotosynthetic metabolism in plastids [J]. *Annual Review of Plant Biology*, 2000, 51: 111-140.

[46] Ramundo S, Asakura Y, Salome P A, et al. Coexpressed subunits of dual genetic origin define a conserved supercomplex mediating essential protein import into chloroplasts[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(51): 32739-32749.

[47] Hust B, Gutensohn M. Deletion of core components of the plastid protein import machinery causes differential arrest of embryo development in *Arabidopsis thaliana*[J]. *Plant Biology*, 2006, 8(1): 18-30.

[48] Kikuchi S, Bedard J, Hirano M, et al. Uncovering the protein translocon at the chloroplast inner envelope membrane[J]. *Science*, 2013, 339(6119): 571-574.

[49] Schnell D J, Kessler F, Blobel G. Isolation of components of the chloroplast protein import machinery[J]. *Science*, 1994, 266(5187): 1007-1012.

[50] Kubis S, Patel R, Combe J, et al. Functional specialization amongst the *Arabidopsis* Toc159 family of chloroplast protein import receptors[J]. *Plant Cell*, 2004, 16(8): 2059-2077.

[51] Jarvis, P., Chen, L.J., Li, H., et al. An *Arabidopsis* mutant defective in the plastid general protein import apparatus. *Science*, 1998, 282, 100-103.

[52] Akita M, Nielsen E, Keegstra K. Identification of protein transport complexes in the chloroplastic envelope membranes via chemical cross-linking[J]. *Journal of Cell Biology*, 1997, 136(5): 983-994.

- [53] Soll J, Schleiff E. Protein import into chloroplasts[J]. *Nature Reviews Molecular Cell Biology*, 2004, 5(3): 198-208.
- [54] Jin Z, Wan L, Zhang Y, et al. Structure of a TOC-TIC supercomplex spanning two chloroplast envelope membranes[J]. *Cell*, 2022, 185(25): 4788-4800.
- [55] Nakai M. YCF1: A Green TIC: Response to the de Vries et al. Commentary[J]. *The Plant Cell*, 2015, 27(7): 1834-1838.
- [56] Du FK, Xu F, Qu H, et al. 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.
- [57] Thiel T, Michalek W, Varshney R K, et al. 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.
- [58] Kurtz S, Choudhuri J V, Ohlebusch E, et al. REPuter: the manifold applications of repeat analysis on a genomic scale[J]. *Nucleic Acids Research*, 2001, 29(22): 4633-4642.
- [59] Frazer K A, Pachter L, Poliakov A, et al. VISTA: computational tools for comparative genomics[J]. *Nucleic Acids Research*, 2004, 32(Web Server issue): W273-W279.
- [60] Amiryousefi A, Hyvonen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes[J]. *Bioinformatics*, 2018, 34(17): 3030-3031.
- [61] 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]. 杭州: 浙江大学, 2017.
- [62] Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood[J]. *Molecular Biology and Evolution*, 2007, 24(8): 1586-1591.
- [63] Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level[J]. *Molecular Biology and Evolution*, 2005, 22(12): 2472-2479.
- [64] Ramundo S, Rahire M, Schaad O, et al. Repression of Essential Chloroplast Genes Reveals New Signaling Pathways and Regulatory Feedback Loops in *Chlamydomonas* [J]. *The Plant Cell*, 2013, 25(1): 167-186.
- [65] Wang Y, Zhao B, Lu Z, et al. The complete chloroplast genome provides insight into the polymorphism and adaptive evolution of *Garcinia paucinervis*[J]. *Biotechnology, biotechnological equipment*, 2021, 35(1): 377-391.
- [66] Kubicki A, Funk E, Westhoff P, et al. Differential expression of plastome-encoded *ndh* genes in mesophyll and bundle-sheath chloroplasts of the C4 plant *Sorghum bicolor* indicates that the complex I-homologous NAD(P)H-plastoquinone oxidoreductase is involved in cyclic electron transport[J]. *Planta*, 1996, 199(2): 276-281.
- [67] Martín M, Sabater B. Plastid *ndh* genes in plant evolution[J]. *Plant Physiology and Biochemistry*, 2010, 48(8): 636-645.
- [68] Shikanai T. Chloroplast NDH: A different enzyme with a structure similar to that of respiratory NADH dehydrogenase[J]. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2016, 1857(7): 1015-1022.
- [69] Xie Z, Merchant S. The plastid-encoded *ccsA* gene is required for heme attachment to chloroplast c-type cytochromes[J]. *J Biol Chem*, 1996, 271(9): 4632-4639.
- [70] Simon J, Hederstedt L. Composition and function of cytochrome c biogenesis System II[J].

The FEBS Journal, 2011, 278(22): 4179-4188.

[71] Gabilly S T, Hamel P P. Maturation of Plastid c-type Cytochromes[J]. *Frontiers in Plant Science*, 2017, 8: 1313.

[72] Nakai M. New Perspectives on Chloroplast Protein Import[J]. *Plant and Cell Physiology*, 2018, 59(6): 1111-1119.

[73] Crooks G E, Hon G, Chandonia J M, et al. WebLogo: a sequence logo generator[J]. *Genome Res*, 2004, 14(6): 1188-1190.

[74] Wilkins M R, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools in the ExPASy server[J]. *Methods in Molecular Biology*, 1999, 112: 531-552.

[75] Kozłowski L P. IPC 2.0: prediction of isoelectric point and pK a dissociation constants [J]. *Nucleic Acids Research*, 2021, 49(W1): W285-W292.

[76] Moller S, Croning M D, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions[J]. *Bioinformatics*, 2001, 17(7): 646-653.

[77] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold[J]. *Nature*, 2021, 596(7873): 583-589.

[78] Leister D. Chloroplast research in the genomic age[J]. *Trends In Genetics*, 2003, 19(1): 47-56.

[79] Richardson L, Schnell D J. Origins, function, and regulation of the TOC-TIC general protein import machinery of plastids[J]. *Journal of Experimental Botany*, 2020, 71(4): 1226-1238.

[80] Paila Y D, Richardson L, Schnell D J. New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development[J]. *Journal of Molecular Biology*, 2015, 427(5): 1038-1060.

[81] Kikuchi S, Oishi M, Hirabayashi Y, et al. A 1-megadalton translocation complex containing tic20 and tic21 mediates chloroplast protein import at the inner envelope membrane [J]. *The Plant Cell*, 2009, 21(6): 1781-1797.

[82] Zufferey M, Montandon C, Douet V, et al. The novel chloroplast outer membrane kinase KOC1 is a required component of the plastid protein import machinery[J]. *Journal of Biological Chemistry*, 2017, 292(17): 6952-6964.

[83] Holbrook K, Subramanian C, Chotewutmontri P, et al. Functional analysis of semi-conserved transit peptide motifs and mechanistic implications in precursor targeting and recognition [J]. *Molecular Plant*, 2016, 9(9): 1286-1301.

[84] Sork V L, Fitz-Gibbon S T, Puiu D, et al. First draft assembly and annotation of the genome of a california endemic oak *Quercus lobata* nee (fagaceae)[J]. *G3 (Bethesda)*, 2016, 6(11): 3485-3495.

[85] 黄婷. 中国青冈亚属青冈组叶绿体基因组分析和系统发育研究[D]. 南充: 西华师范大学, 2023.

[86] Tangphatsornruang S, Sangsrakru D, Chanprasert J, et al. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships[J]. *DNA Research*, 2010, 17(1): 11-22.

[87] Delannoy E, Fujii S, Colas D F C, et al. 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.

[88] Hipp A L, Manos P S, Hahn M, et al. Genomic landscape of the global oak phylogeny[J]. *New Phytologist*, 2020, 226(4): 1198-1212.

[89] Dong W, Xu C, Li C, et al. *ycf1*, the most promising plastid DNA barcode of land plants[J].

Scientific Reports, 2015. 5(1).

[90] Llamas E, Pulido P. A proteostasis network safeguards the chloroplast proteome[J]. *Essays Biochem*, 2022, 66(2): 219-228.

[91] Kovács-Bogdán E, Soll J, Bölder B. Protein import into chloroplasts: The Tic complex and its regulation[J]. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2010, 1803(6): 740-747.

[92] Li H M, Chiu C C. Protein transport into chloroplasts[J]. *Annu Rev Plant Biol*, 2010, 61: 157-180.

[93] Kessler F, Schnell D. Chloroplast biogenesis: diversity and regulation of the protein import apparatus[J]. *Current Opinion in Cell Biology*, 2009, 21(4): 494-500.

[94] Richardson L, Schnell D J. Origins, function, and regulation of the TOC-TIC general protein import machinery of plastids[J]. *Journal of Experimental Botany*, 2020, 71(4): 1226-1238.

## 附录

附录 A NCBI 中已上传栎属物种叶绿体基因组序列详细信息

Appendix A Chloroplast genome sequences details of *Quercus* species have been uploaded in NCBI

物种 (Species)	亚属 (Subgenus)	组 (Section)	叶习性 (Leaf habit)	GenBank
<i>Quercus acrodonta</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MG678019.1
				NC_059815.1
				MW553099.1
<i>Quercus aquifolioides</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	KP340971.1
				NC_026913.1
				KX911971.1
<i>Quercus baronii</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	KT963087.1
				MW829651.1
				NC_029490.1
<i>Quercus cocciferooides</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MG678016.1
				NC_061586.1
				MW829652.1
				OR188390.1
				OR188389.1
				OR188388.1

附录

				KU240010.1
				NC_031357.1
<i>Quercus dolicholepis</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	OR188391.1
				OR188392.1
				OR188393.1
				NC_061593.1
<i>Quercus engleriana</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MG678029.1
				MZ196209.1
				MW829653.1
<i>Quercus gilliana</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MG678007.1
				NC_061578.1
<i>Quercus guyavifolia</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MG678020.1
				NC_061587.1
				NC_082110.1
<i>Quercus longispica</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	OR188400.1
				OR188401.1
				OR188402.1
<i>Quercus utilis</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MG678015.1
				NC_061584.1
<i>Quercus tatakaensis</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MG678036.1
				NC_061585.1
<i>Quercus tarokoensis</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	NC_036370.1

附录

				MF135621.1
				MG678038.1
<i>Quercus spinosa</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	KX911972.1
				KM841421.1
				NC_026907.1
				MG678023.1
<i>Quercus senescens</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	NC_061590.1
				MW829657.1
				MG678017.1
<i>Quercus semecarpifolia</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	NC_061613.1
				MZ196211.1
				MW829656.1
				OK585062.1
<i>Quercus rehderiana</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MG678037.1
				NC_061615.1
				MW829655.1
				MG678025.1
<i>Quercus pannosa</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	NC_050963.1
				MT180076.1
				MW450869.1
<i>Quercus franchetii</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	OR188394.1
				OR188395.1

				OR188396.1
				MG678018.1
				NC_060378.1
				MG678026.1
<i>Quercus phillyraeoides</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	NC_048488.1
				MN882701.1
				MK105462.1
<i>Quercus oxyphylla</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MG678021.1
				NC_061588.1
<i>Quercus pseudosemecarpifolia</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MG678014.1
				NC_061583.1
				NC_061577.1
<i>Quercus monimotricha</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	MW829654.1
				MG678006.1
<i>Quercus tungmaiensis</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	NC_036936.1
				MF593893.1
<i>Quercus acuta</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_054352.1
				MT742291.1
<i>Quercus delavayi</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>		NC_060379.1
				MW450870.1
<i>Quercus gilva</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	-	MK986651.1
				MK860968.1

附录

				MG678009.1
				OP450822.1
				NC_049876.1
				OP450818.1
<i>Quercus glauca</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	-	KX852399.1
				NC_036930.1
				OP450823.1
<i>Quercus hypargyrea</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	MW450871.1
				NC_060381.1
<i>Quercus schottkyana</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿阔叶	MW450872.1
				OQ534360.1
<i>Quercus lamellosa</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_069845.1
				ON497016.1
				MZ382817.1
				OP581003.1
<i>Quercus sessilifolia</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	OQ534361.1
				MG678012.1
				NC_061581.1
				NC_083127.1
<i>Quercus salicina</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	OR458923.1
				NC_085458.1
<i>Quercus poilanei</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	OR835153.1

附录

<i>Quercus oxyodon</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_069209.1 ON258628.1 OQ534363.1 OP450819.1
<i>Quercus myrsinifolia</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	MN199025.1 NC_061611.1 MG678005.1 MG678008.1 OP450821.1
<i>Quercus fleuryi</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_058778.1 MZ502291.1 OP442516.1
<i>Quercus jenseniana</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	MG678011.1 NC_061580.1 NC_061579.1
<i>Quercus neglecta</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	ON258631.1 MG678010.1 NC_061575.1
<i>Quercus glaucoides</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	ON381168.1 OQ534362.1 MG678003.1
<i>Quercus multinervis</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_060380.1

附录

				NC_061576.1
				MG678004.1
<i>Quercus edithiae</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	KU382355.1
				MW628880.1
<i>Quercus ningangensis</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	MG678013.1
				NC_061582.1
				ON303301.1
				MF787253.1
<i>Quercus sichourensis</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_036941.1
				OK638154.1
				OQ998920.1
<i>Quercus kerrii</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_068858.1
				OP679796.1
<i>Quercus rex</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_083083.1
				OQ998921.1
<i>Quercus austrocochinchinensis</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_083081.1
				OQ998918.1
				OQ534365.1
<i>Quercus chungii</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_057248.1
				MW401633.1
<i>Quercus ciliaris</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	MN199024.1
<i>Quercus pachyloma</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	OP442517.1

附录

				NC_068537.1
<i>Quercus stewardiana</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	MN199023.1
				MH899015.1
<i>Quercus acutissima</i>	<i>Cerris</i>	<i>Cerris</i>	落叶	MF593895.1
				NC_039429.1
				MH607377.1
<i>Quercus chenii</i>	<i>Cerris</i>	<i>Cerris</i>	落叶	MF593894.1
				NC_039428.1
				KU240009.1
				CP129457.1
				OR188403.1
				OR188404.1
<i>Quercus variabilis</i>	<i>Cerris</i>	<i>Cerris</i>	落叶	OR188405.1
				MK105451.1
				MK105466.1
				NC_031356.1
				MK105465.1
<i>Quercus agrifolia</i>	<i>Quercus</i>	<i>Lobatae</i>	常绿	OK634019.1
				JX970937.1
<i>Quercus rubra</i>	<i>Quercus</i>	<i>Lobatae</i>	落叶	NC_020152.1
				MK105463.1
<i>Quercus phellos</i>	<i>Quercus</i>	<i>Lobatae</i>	落叶	MZ196210.1

附录

				NC_069212.1
				ON332546.1
<i>Quercus coccinea</i>	<i>Quercus</i>	<i>Lobatae</i>	-	NC_047481.1
				MN308055.1
<i>Quercus palustris</i> isolate BOP010429	<i>Quercus</i>	<i>Lobatae</i>	-	MK105461.1
<i>Quercus wislizeni</i>	<i>Quercus</i>	<i>Lobatae</i>	常绿	OM541583.1
<i>Quercus kelloggii</i> voucher Hartnell College Collection 270	<i>Quercus</i>	<i>Lobatae</i>	落叶	NC_062091.1
				OM541584.1
				NC_062092.1
				NC_039725.1
				MG967555.1
<i>Quercus dentata</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	MG678030.1
				MK105453.1
				MK105454.1
				MK105455.1
				MK693136.1
				NC_061594.1
<i>Quercus fabri</i>	<i>Quercus</i>	<i>Quercus</i>	-	MK922346.1
				MG678031.1
				MK105456.1
<i>Quercus griffithii</i>	<i>Quercus</i>	<i>Quercus</i>	-	NC_061614.1
				MK922347.1

				OR188397.1
				OR188398.1
				OR188399.1
				MG678034.1
				MG678028.1
				NC_061592.1
<i>Quercus yunnanensis</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	MW829658.1
				MW876863.1
				MK922348.1
<i>Quercus virginiana</i>	<i>Quercus</i>	<i>Quercus</i>	常绿或部分落叶	MT916773.1
				NC_050972.1
				MK089571.1
				OK274109.1
				OK589866.1
				MG678033.1
<i>Quercus mongolica</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	MK564083.1
				MK922349.1
				MN602451.1
				NC_043858.1
				MK105460.1
<i>Quercus turbinella</i>	<i>Quercus</i>	<i>Quercus</i>	常绿或部分落叶	NC_081467.1
				OQ835463.1

附录

				KP301144.1
<i>Quercus aliena</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	NC_026790.1
				KU240007.1
				NC_061591.1
				MG678024.1
<i>Quercus serrata</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	MK922350.1
				MN602452.1
				MK105464.1
<i>Quercus aliena</i> var. <i>acutiserrata</i>	<i>Quercus</i>	<i>Quercus</i>	-	KU240008.1
				MK105452.1
				OK274107.1
				MK059753.1
				MK105468.1
<i>Quercus wutaishanica</i>	<i>Quercus</i>	<i>Quercus</i>	-	MN602450.1
				MG678027.1
				NC_043857.1
				OK274108.1
				MG678032.1
<i>Quercus serrata</i> var. <i>brevipetiolata</i>	<i>Quercus</i>	<i>Quercus</i>	-	NC_061595.1
				MK922351.1
				MK105458.1
<i>Quercus dentata</i> subsp. <i>stewardii</i>	<i>Quercus</i>	<i>Quercus</i>	-	MG678022.1

附录

				NC_061589.1
				MK922352.1
				LT996900.1
<i>Quercus robur</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	MG678035.1
				NC_046388.1
				MN562095.1
<i>Quercus petraea</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	LT996899.1
<i>Quercus lobata isolate SW786</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	CM012305.2
				OQ534364.1
<i>Quercus argyrotricha</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_073549.1
				NC_069211.1
<i>Quercus austroglauca</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	ON258630.1
				MK449426.1
<i>Quercus bawanglingensis</i>	<i>Cerris</i>	<i>Ilex</i>	常绿	NC_046583.1
				NC_048513.1
<i>Quercus fenchengensis</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	MN095295.1
<i>Quercus gambelii isolate BOP010431</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	MK105457.1
				ON258629.1
<i>Quercus gambleana</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_069210.1
				OM541585.1
<i>Quercus × morehus</i>	-	-	-	NC_062093.1
<i>Quercus helferiana</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_083082.1

附录

				OQ998919.1
				OR835154.1
<i>Quercus litseoides</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	ON598394.1
				NC_065782.1
<i>Quercu macrocarpa isolate BOP010428</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	MK105459.1
				MK860969.1
<i>Quercus mongolica subsp. crispula</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	NC_049877.1
				MG356785.1
<i>Quercus obovatifolia</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_039972.1
				MK105467.1
<i>Quercus stellata isolate BOP010424</i>	<i>Quercus</i>	<i>Quercus</i>	落叶	NC_060377.1
				MW411183.1
<i>Quercu saravanensis</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	NC_068538.1
				OP450820.1
<i>Quercus shennongii</i>	<i>Cerris</i>	<i>Cyclobalanopsis</i>	常绿	

注：分组及叶习性数据统计来自 NCBI 数据库、Flora of China、中国植物志、文献[26]及网站 <https://www.treesandshrubsonline.org/>、<https://plants.ces.ncsu.edu/>等搜集统计。

## 附录 B 55 个栎属物种位点模型正选择分析结果

Appendix B Results of the site model positive selection analysis of 55 *Quercus* species

基因 (Gene)	功能 (Function)	位点模型 (Site model)	-lnL	$2\delta(\ln L)$	<i>P</i> -value ( <i>P</i> 值)	<i>Q</i> -value ( <i>Q</i> 值)	正选择位点 (Positively selected sites)
		Model7	-6698.11				11-0.957,28-0.948,33-0.948,131-0.948,234-0.948,238-0.948,280-0.987,283-0.948,364-0.976,372-0.947,384-0.947,388-0.949,401-0.948,410-0.948,429-0.948,439-0.948,443-0.947,491-0.949,505-0.947,519-0.948,542-0.948,544-0.948,552-0.948,566-0.947,567-0.947,568-0.947,587-0.947,595-0.948,598-0.947,604-0.948,626-0.947,637-0.948,639-0.949,643-0.948,680-0.948,691-0.948,697-0.947,754-0.949,807-0.948,815-0.948,841-0.947,848-0.948,869-0.957,890-0.953,895-0.949,922-0.948,935-0.975,939-0.948,970-0.948,985-0.948,1003-0.947,1005-0.947,1014-0.948,1028-0.947,1066-0.947,1079-0.949,1124-0.947,1126-0.947,1154-0.948,1158-0.947,1168-0.947,1174-0.947,1184-00.947,1194-0.948,1239-0.948,1265-0.948,1351-0.948,1366-0.947,1376-0.948,1382-0.947,1383-0.947
<i>rpoC2</i>	RNA polymerase beta" subunit	Model8	-6691.84	12.53409	1.90E-03	1.05E-02	
<i>ycf3</i>	Photosystem I assembly protein Ycf3	Model7	-821.203				119-0.999,120-0.826,121-0.997,122-0.940

附录

---

		Model8	-806.036	30.33531	2.59E-07	2.66E-06	
		Model7	-1088.18				
<i>ndhk</i>	NADH dehydrogenase subunit K						15-1.000,54-0.640
		Model8	-1077.57	21.21442	2.47E-05	1.62E-04	
		Model7	-2263.35				
<i>rbcL</i>	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit						95-0.985,219-0.974,226-0.940,251-0.993,262- 0.944,309-0.972,328-0.974
		Model8	-2238.21	50.27083	1.21E-11	2.18E-10	
		Model7	-565.248				
<i>rps12</i>	ribosomal protein S12						13-0.931,39-0.973,79-0.931,116-0.970
		Model8	-555.269	19.95801	4.64E-05	2.78E-04	

		Model7	-794.266				
<i>petD</i>	cytochrome b6/f complex subunit IV						1-0.954,2-1.000,3-0.996
		Model8	-777.507	33.51761	5.27E-08	6.32E-07	
		Model7	-721.959				
<i>rpl16</i>	ribosomal protein L16						1-0.998,2-0.999,3-0.773,4-0.899,119-0.762
		Model8	-709.758	24.40124	5.03E-06	4.02E-05	
		Model7	-9443.34				96-0.958,284-0.958,541-0.958,562-0.958,694-0.958,719-0.958,920-0.958,928-0.989,937-0.958,944-0.958,1087-0.958,1170-0.958,1213-0.958,1247-0.958,1287-0.958,1337-0.958,1850-0.979,1878-0.958,2003-0.958,2239-0.958,2266-0.958
<i>ycf2</i>	hypothetical chloroplast RF21	Model8	-9425.57	35.53901	1.92E-08	2.76E-07	
		Model7	-2140.67				
<i>ndhB</i>	NADH-plastoquinone oxidoreductase subunit 2	Model8	-2110.11	61.13029	5.32E-14	1.92E-12	16-0.915,259-0.997,276-0.915
		Model7	-3909.87				21-0.948,25-0.947,41-0.947,49-0.947,88-0.953,12-0.947,219-0.947,221-0.947,282-0.947,292-0.947,373,0.947,376-0.949,377-0.947,388,0.947,393-0.947,410-0.947,418-0.947,420-0.947,421-0.947,455-0.993,473-0.947,475-0.948,480-0.949,482-0.947,484-0.971,487-0.949,495-0.947,496-0.947,500-0.985,510-0.972,532-0.947,533-0.948,534-0.948,541-0.947,566-0.948,569-0.948,570-0.948,573-0.947,574-0.947,575-
<i>ndhF</i>	NADH-plastoquinone oxidoreductase subunit 5	Model8	-3898.35	23.02921	9.98E-06	7.19E-05	

附录

<i>ccsA</i>	cytochrome c heme attachment protein	Model7	-1498.52				0.946,580-0.987,597-0.947,607-0.948,610-0.946,633-0.948,647-0.947,648-0.947
		Model8	-1483.88	29.27323	4.40E-07	3.96E-06	35-0.952,66-0.952,75-0.952,89-0.988,91-0.952,94-0.978,96-0.952,108-0.952,109-0.953,127-0.952,144-0.952,166-0.953,195-0.952,201-0.952,203-0.952,228-0.978,277-0.953
		Model7	-2509.48				6-0.953,17-0.953,22-0.953,28-0.953,56-0.953,74-0.953,84-0.954,107-0.953,113-0.953,118-0.954,164-0.953,170-0.956,199-0.953, 205-0.953,216-0.953,217-0.953,278-0.953,285-0.953,361-0.953,368-0.956,374-0.954,412-0.953,413-0.988,451-0.953,452-0.953,503-1.000
<i>ndhd</i>	NADH-plastoquinone oxidoreductase subunit 4	Model8	-2479.64	59.68662	1.10E-13	2.64E-12	3-0.953,156-0.943,201-0.942,283-0.942,329-0.945,333-0.995, 335-0.942,339-0.993,340-0.944,369-0.943,379-0.975,380-0.94,383-0.943,385-0.942,386-0.942,391-0.948, 392-0.945,393-0.944,409-0.952,412-0.955,456-0.945,473-0.943,486-0.943,489-0.943, 493-0.944,496-0.945,499-0.944,509-0.942,514-0.993,516-0.945,518-0.945,523-0.976,536-0.948,537-0.973,552-0.950,553-0.942,554-0.944,556-0.944,564-0.942,565-0.943,573-0.948,579-0.953,581-0.950,584- 0.942,588-0.944,596-0.949
		Model7	-9960				
		Model8	-	144.537	0.00E+00	0.00E+00	
<i>ycf1</i>	hypothetical chloroplast RF19	Model8	9887.73	2			

注：最后一列连字符前为由贝叶斯经验贝叶斯 (Bayes empirical bayes , BEB) 分析鉴定出的氨基酸的位置。连字符后面的数字是 BEB 分析下的后验概率 (Posterior probability ,

PP)

## 个人简介

原明星，女，1998年1月生，籍贯山西省吕梁市。2016年9月至2020年7月就读于山西财经大学资源环境学院自然地理与资源环境专业，取得管理学学士学位。2021年9月至2024年7月就读于北京林业大学生态与自然保护学院生态学专业，导师为杜芳教授，研究方向为分子生态学。

硕士期间，主修了分子生态理论方法、高级森林生态学、生态学研究方法、生态学数据分析—R语言、多元统计分析等课程，成绩良好。硕士期间主要进行常绿和落叶栎树叶绿体基因组比较及相关基因功能研究。

## 导师简介

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

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

2009年获法国农业科学学院三大科研进展奖励；2014年获得教育部高等学校科学研究优秀成果奖自然科学奖一等奖(排名第四)；2014、2018年获得梁希青年论文奖；2015年当选北京市科技新星；2019年当选中国植物学会女科学家协会委员；2021年开始担任SCI期刊*Ecology & Evolution*的Associate Editor；2022年联合华东师范大学的王嵘教授与日本和加拿大的学者一起组织了“森林树种保护基因组学(Forest Tree Conservation Genomics)”的专刊；2023年开始担任*Plant Diversity*的Associate Editor。

## 获得成果目录清单

[1] 原明星,尹康权,李俄仁措,尚海忠,杜芳. 叶绿体基因编辑技术:进展与展望[J]. 中国草地学报. (已被中国草地学报期刊接收, CSCD 源刊)

## 致谢

时光荏苒，三年的研究生生涯即将结束。我曾经无数次想象过毕业的场景，真的到来却总觉得不真实。永远记得复试完的一通电话，驱散了我复试完糟糕的情绪，开始对未来三年的生活充满向往。复试完第一次组会见到了导师和实验室的师兄师姐，对实验室有了一些了解，对未来的研究生生活有了更多期待。之后确定了研究方向，在家里和师兄学习了叶绿体基因组的相关分析。2021年7月28日，自己从山西来到北京，可能是我真正意义上的独立的开始。在我刚来的日子里师兄师姐和老师给予了我极大的帮助，让我快速适应了这里的生活。后来开始慢慢接触分子实验，对于实验，我感到陌生而抵触，在老师和师兄师姐的帮助下发现并没有我想象中那样艰难，也越来越得心应手。

这既漫长又短暂的三年，让我接触到了更多优秀的人，也让我见到了更广阔的世界。回顾这三年，我受益良多，我取得的每一点进步都离不开老师的谆谆教诲，同学、朋友的热心帮助和家人的默默支持，谢谢你们！

我衷心感谢我的导师杜芳老师，作为一名跨专业报考的学生，具体研究方向上也是从未涉及过的分子生态学领域，刚入学的我总是感到迷茫和无措，没有别人的专业背景，之前也从未接触过分子实验，但杜芳老师对我学业上的帮助和生活上的关心，让我坚定了读研的信心，坚持走到现在。我不是一个很勤奋，很优秀的人，也不是一个会积极主动和老师交流的人，但杜老师认真负责，会认真分析我的问题，找我交流，帮助我进步。杜老师严谨的治学态度，修改我的学术文章时的专业、认真和细致，仍然深深印在我脑海中，难以忘怀，让我深深敬佩。您传授给我诸多人生的道理也烙印在我心灵深处，让我终身受益。希望老师在未来的日子里工作顺利，身体健康，生活幸福。

感谢北京林业大学草业草原学院尹康权老师三年来对我实验和论文写作上的指导和帮助。老师的耐心指导让我在实验领域进步极大，在论文写作上也给予了我许多帮助，经常提出很多新的想法和思考，让我在我的研究方向上得以不断深入。

感谢实验室的各位同学，这三年的陪伴、支持，从各位同学身上我也学会了很多。感谢师兄许阳，一点点教会我叶绿体基因组的分析，为我这三年的研究打下了坚实的基础。感谢师兄刘可可、师姐祁敏和罗艳君、秦伟对我科研上和生活上的帮助，他们一直是我学习的榜样。感谢师姐张艺一直以来给予的温暖和陪伴，从日常课程学习和论文写作到让我更好的融入实验室，都一直在帮助我。在毕业压力、工作压力让我焦虑迷茫的时候，也给了我极大耐心和鼓励。

感谢我的父母一直以来的关心、支持、理解、信任和期盼，这是我一直以来前行的动力。感谢你们不求回报的付出，尊重我每一个决定，让我内心始终充满能量。这

三年回家很少，都在临近过年的时候，只是短暂的陪伴了你们。希望你们平安健康，也希望未来我是你们的依靠和骄傲。

感谢一直以来陪伴我的朋友们，我们在或不在一个城市从未影响我们之间的距离，是你们让我这三年的生活并不枯燥，一起进步，一起成长，一起成为更好的人。

想感谢的人很多，他们让我慢慢成为更好的我，但也想感谢一直努力的自己，第一次考研失败，也想过选择工作，但对于科研的向往让我选择再来一次。期间也承受了一定的压力，感谢自己的坚持和勇敢。在这三年中，实验总是失败，让我受挫，但一直在调整、尝试，感谢自己的坚持。

我和北林的故事结束于 2024 年的夏天，银杏大道见证着我三年的成长。这个夏天也要和学生时代的自己说再见了，要成为独当一面的大人了，感谢所有的相遇。虽然我仍然不够成熟不够优秀，但我始终在前行。我坚信好事，总会发生在下个转弯！