

DOI: 10.11913/PSJ.2095-0837.22231

栗锦焯, 平晶耀, 崔贵峰, 苏应娟, 王艇. 桔梗科 *rps2* 基因簇断裂对进化速率的影响[J]. 植物科学学报, 2023, 41(3): 333-342

Li JY, Ping JY, Cui GF, Su YJ, Wang T. Effects of the broken *rps2* gene cluster on evolutionary rates in Campanulaceae [J]. *Plant Science Journal*, 2023, 41(3): 333-342

桔梗科 *rps2* 基因簇断裂对进化速率的影响

栗锦焯¹, 平晶耀¹, 崔贵峰², 苏应娟^{3,4*}, 王艇^{1*}

(1. 华南农业大学生命科学学院, 广州 510642; 2. 国有南召林场(河南伏牛山国家级自然保护区南召宝天曼管理局), 河南南召 474665; 3. 中山大学生命科学学院, 广州 510275; 4. 中山大学深圳研究院, 广东深圳 518057)

摘要: 保守的 *rps2* 基因簇 (*rps2-atpI-atpH-atpF-atpA*) 在桔梗科的叶绿体基因组中发生了断裂。为探究该基因簇断裂后, 相关基因的进化速率是否发生变化, 本研究选取 50 种桔梗科植物和 2 种睡菜科植物的叶绿体基因组序列进行系统发育关系的构建, 对 *rps2* 基因簇相关基因的进化速率、选择压力以及适应性进化过程进行分析。结果显示, 桔梗科中 *rps2* 基因簇发生断裂的物种均来自风铃草亚科, 且在系统发育树上组成单系分支; 与未断裂物种相比, 断裂物种中相关基因的进化速率均值降低; 基因间的非同义替换速率存在显著差异。适应性进化分析中, 位点模型在 *atpI* 和 *rps2* 基因中检测到正选择位点。本研究结果表明基因簇断裂可能具有系统发育意义, 基因簇断裂后进化速率会发生改变, 不同基因也经历了不同的进化历程。

关键词: *rps2* 基因簇; 桔梗科; 进化速率; 选择压力; 适应性进化

中图分类号: Q943

文献标识码: A

文章编号: 2095-0837(2023)03-0333-10

Effects of the broken *rps2* gene cluster on evolutionary rates in Campanulaceae

Li Jin-Ye¹, Ping Jing-Yao¹, Cui Gui-Feng², Su Ying-Juan^{3,4*}, Wang Ting^{1*}

(1. College of Life Sciences, South China Agricultural University, Guangzhou 510642, China; 2. Nanzhao State-owned Forest Farms (Funiu Mountain National Nature Reserve Nanzhao Baotianman Administration in Henan), Nanzhao, Henan 474665, China; 3. School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China; 4. Research Institute of Sun Yat-sen University in Shenzhen, Shenzhen, Guangdong 518057, China)

Abstract: The conserved *rps2* gene cluster (*rps2-atpI-atpH-atpF-atpA*) is broken in the chloroplast genome of Campanulaceae. To explore whether the evolutionary rates of the related genes changed after the cluster was broken, we selected the chloroplast genome sequences of 50 Campanulaceae and two Menyanthaceae species to construct their phylogenetic relationships, and further analyzed the evolutionary rate, selection pressure, and adaptive evolutionary process of the *rps2* gene cluster-related genes. Results showed that all species with a broken *rps2* gene cluster in Campanulaceae were from Campanuloideae and formed monophyletic branches on the phylogenetic tree. Compared with the unbroken species, the mean evolutionary rate of the related genes decreased in the broken species. There were significant differences in the nonsynonymous substitution rate between genes. In adaptive evolutionary analysis, the site model detected positive-selection sites in the *atpI* and *rps2* genes. Our findings suggest that the fragmentation of the gene cluster holds phylogenetic implications, leading to alterations in evolutionary rates and diverse evolutionary trajectories.

收稿日期: 2022-09-03, 修回日期: 2022-11-09。

基金项目: 国家自然科学基金项目(31770587)。

This work was supported by a grant from the National Natural Science Foundation of China (31770587).

作者简介: 栗锦焯(1998-), 女, 硕士研究生, 研究方向为植物系统发育与分子进化(E-mail: 1204522967@qq.com)。

* 通讯作者(Authors for correspondence. E-mail: suyj@mail.sysu.edu.cn; tingwang@scau.edu.cn)。

ries among individual genes.

Key words: *rps2* gene cluster; Campanulaceae; Evolutionary rate; Selection pressure; Adaptive evolution

叶绿体是绿色植物生命活动的代谢枢纽,为植物进行光合作用和生物合成的半自主性细胞器,具有独立的遗传系统——叶绿体基因组^[1]。关于叶绿体的起源有很多假说,目前广为接受的是内共生起源学说,即叶绿体起源于大约 15 亿年前被异养宿主细胞(真核原生生物)吞并且行光能自养的蓝细菌^[2]。由于早期内共生过程中蓝细菌经历了一系列复杂演化,发生基因丢失和基因转移事件,致使稳定下来的叶绿体基因组长度大大减小,但叶绿体基因组还是保留了少量原核特征,其中包括将基因组织成类似细菌操纵子的多顺反子转录单元,即基因簇^[3]。基因簇与细菌操纵子表面上看都是由功能相关的、共表达的、具有物理连接的基因群组成,但不同的是基因簇可能来源于物种演化过程中的基因复制、新功能化以及基因组的动态重组,而并非由细菌的水平基因转移所引起^[4]。陆地植物的叶绿体基因组通常具有结构稳定、基因顺序和基因含量高度保守的特征,其中一个主要限制因素就是它被组织成了多顺反子转录的基因簇^[5]。

种子植物的叶绿体基因簇很少被破坏,但有报道表明,在被子植物的一些植物类群如牻牛儿苗科^[6]、豆科^[7]和桔梗科^[8]中存在基因簇断裂的现象。在桔梗科植物中,两个叶绿体基因簇被破坏,*clpP-5'-rps12-rpl20*(*clpP*基因簇)和*rps2-atpI-atpH-atpF-atpA*(*rps2*基因簇)。*rps2*基因簇编码功能不相关的多肽,第一个基因(*rps2*基因)编码核糖体蛋白,其他 4 个基因编码 ATP 合酶亚单位^[9]。由于翻译系统和光合系统在蛋白质组分的需求数量上有很大的差异,这两种基因必须经过复杂的调控机制才能实现共同转录^[10]。该基因簇一旦发生断裂,重新分布的片段就需要新的启动子来驱动基因转录,然而如何在新的位置进行转录尚不清楚,基因簇发生断裂后对叶绿体进化的影响也知之甚少^[11]。

桔梗科为被子植物中高度进化的一个分支,约 100 属, 2300 ~ 2380 种;广布于全球,在药

用和观赏方面有着重要的经济价值。桔梗科有广义和狭义之分,狭义桔梗科只包括具有传统意义上桔梗科基本特征的风铃草亚科。广义桔梗科包含 5 个亚科,即风铃草亚科、半边莲亚科、线枝草亚科、苣荬亚科和腔柱草亚科^[12]。尽管存在基因成簇分布的进化限制来让叶绿体基因组保持稳定,但在桔梗科的叶绿体基因组中还是发生了多次片段倒位事件,这种大规模的结构重排导致基因位置发生变化,也促使一些基因簇发生断裂。有研究发现,叶绿体基因组中的核苷酸替换速率在不同的谱系中、在反向重复区和单拷贝区之间以及在不同功能基因组之间均有所不同^[13, 14]。基因簇断裂可能会影响基因的转录进程,同时,成簇排列的基因又作为独立的编码蛋白基因,其进化速率可能会因位置的变动而发生改变。本研究对桔梗科典型的单种属(桔梗属)植物—桔梗(*Platycodon grandiflorus* (Jacq.) A. DC)的叶绿体基因组进行了测序,共选取 50 种广义桔梗科植物和 2 种睡菜科植物(附表 1¹⁾),提取其共有蛋白编码序列(Coding sequences, CDS)构建系统发育关系,对 *rps2* 基因簇在桔梗科中的分布情况进行统计;并在系统发育树的背景下,计算 *rps2* 基因簇中各基因的进化速率和选择压力,并检测其经历的正选择作用。

1 材料与方法

1.1 取样及测序

桔梗新鲜叶片采自于河南省南阳县乔端镇自然生长的桔梗植株(33°34' N, 112°6' E),采集后将叶片用锡纸包裹,液氮冷冻后于 -80℃ 保存备用。

样品基因组 DNA 的提取采用 E.Z.N.A.[®] Plant DNA kit (OMEGA) 试剂盒提取基因组,经片段化(300 ~ 500 bp)后进行纯化并构建测序文库。步骤依次为 DNA 末端修复、3' 端加 A、连接测序接头、琼脂糖凝胶电泳法回收目的片段、对目的片段进行 PCR 扩增,最终建成测序文库。使用

1) 如需查阅附件内容请登录《植物科学学报》网站(<http://www.plantscience.cn>)查看本期文章。

FastQC v 0.11.4 软件对建好的文库进行质检，质检合格后采用 Illumina Novaseq™ 平台进行双端测序，测序读长为 150 bp。接下来使用 Trimmomatic v 0.32^[15] 过滤数据得到 clean reads，利用组装软件 NOVOPlasty^[16] 对 clean reads 进行拼接和组装。组装完成的基因组序列利用叶绿体基因组注释软件 PGA^[17] 进行基因注释，参考序列为毛细钟花 (*Codonopsis hongii* Lammer) 的叶绿体基因组全序列，并对所有样本的注释结果进行人工校正。通过在线网站 OGDRAW v 1.3^[18] (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) 绘制叶绿体基因组图谱。最后，通过 BankIt 平台上传序列信息 (<https://www.ncbi.nlm.nih.gov/Web-Sub/>)。

1.2 构建系统发育树

从 NCBI (<https://www.ncbi.nlm.nih.gov>) 数据库下载 2 种睡菜科植物和 49 种桔梗科植物的叶绿体基因组全序列，结合自行组装的桔梗序列，共计 52 种被子植物，涵盖桔梗科 29 个属和睡菜科 2 个属。

利用 Genious prime 2019.0.4^[19] 软件提取 55 条共有 CDS 序列，使用 MEGA 7.0.26^[20] 软件进行比对校正，并建立共有 CDS 数据串联数据集，以睡菜科的两个物种作为外类群构建邻接树 (Neighbor-Joining, NJ)，自展次数 (Bootstrap) 为 1000 次；最大似然 (Maximum likelihood, ML) 树使用 RaxmlGUI2 软件^[21] 构建，设置 GTRGAMMAI 替换模型，1000 自展次数；最大简约 (Maximum parsimony, MP) 树和贝叶斯 (Bayesian inference, BI) 树分别使用 PAUP 4.0 软件和 MrBayes 3.2.6 软件^[22] 进行构建，MrBayes 3.2.6 软件中参数设置 rates 为 invgamma，mcmc ngen 为 1 000 000。综合对比分析得到的 4 种树，结合前人发表的系统发育关系，挑选支持度高的系统发育树用于后续进化分析。

1.3 分子进化速率分析

使用 HyPhy 2.2.4^[23] 软件在系统发育背景下，基于最大似然法计算基因簇内 5 个基因 (*atpA*、*atpF*、*atpH*、*atpI* 和 *rps2*) 的转换率 (Transition rate, *trst*)、颠换率 (Transversion rate, *trsv*)、颠换率与转换率的比值 (*trsv/trst*, *ratio*)、同义替换率 (Synonymous substitution rate, *dS*)、非同

义替换率 (Nonsynonymous substitution rate, *dN*) 和非同义替换率与同义替换率的比值 (*dN/dS*, ω)。转换率、颠换率和 *ratio* 选择核苷酸分区和 HKY85 替换模型计算；同义替换率、非同义替换率以及 ω 选择密码子分区和 MG94 × HKY85 × 3_4 模型计算。利用 SPSS 22 软件的秩和检验进行显著性分析，使用 GraphPad Prism 8 软件绘图。

1.4 选择压力及适应性进化分析

运行 PAML 4 软件中 codeml 程序^[24] 的分支模型 (Branch model)、分支位点模型 (Branch-site model) 和位点模型 (Site model)，对 *rps2* 基因簇的相关基因进行进化分析。利用分支模型中最简单的单比率模型 (One-ratio model, M0)、最复杂的自由比率模型 (Free-ratio model, F) 和位于中间的二比率模型 (Two-ratio model, Model 2) 进行似然比检验，选择更优模型。在二比率模型中将 *rps2* 基因簇发生断裂的物种设为前景支，未断裂的物种设为背景支运行分析。分支位点模型同样设置 *rps2* 基因簇发生断裂的物种作为前景支，检测正选择作用对部分位点的影响。采用可以降低假阳性的检验 2 筛选正选择位点，对 MA 模型和无效模型 (ω 值设置为 1) 进行差异显著性检验。位点模型中共包含 4 对嵌套模型：M1a (近中性, Nearly neutral) 与 M2a (正选择, Positive selection) 模型、M0 与 M3 (离散, Discrete)、M7 (β) 与 M8 (β 和 ω) 模型、M8 和 M8a (β 和 $\omega = 1$) 模型。其中 M1a 与 M2a、M7 与 M8 模型用于正选择作用的检测，采用 M1a 与 M2a 检验结果为可能的正选择位点。鉴定正选择位点有 NEB (Native empirical) 和 BEB (Bayes empirical bayes) 两种贝叶斯方法，前者易产生假阳性，本研究主要以 BEB 的结果为准。

将筛选出的可能存在正选择位点的氨基酸序列提交到瑞士生物信息研究所 (Swiss-Model, <https://swissmodel.expasy.org/>)，利用同源建模原理预测编码蛋白的三维结构，查看正选择位点所在的结构位置。

2 结果与分析

2.1 桔梗的叶绿体基因组信息

桔梗的叶绿体基因组为典型的 4 区环状结构，基因组大小为 173 118 bp，大单拷贝 (Large sin-

gle copy, LSC) 区为 79 024 bp, 反向重复 (Inverted repeats, IRs) 区为 43 127 bp, 小单拷贝 (Small single copy, SSC) 区为 7840 bp; 基因组总 GC 含量为 38.1%, 其中 LSC 区 GC 含量为

37.2%, IR 区为 39.5%, SSC 区为 31.0%; 基因组全序列总计 110 个编码基因, 包含 77 个蛋白质编码基因、29 个 tRNA 基因和 4 个 rRNA 基因 (图 1), 提交至 NCBI 的序列登录号为 OP018680。

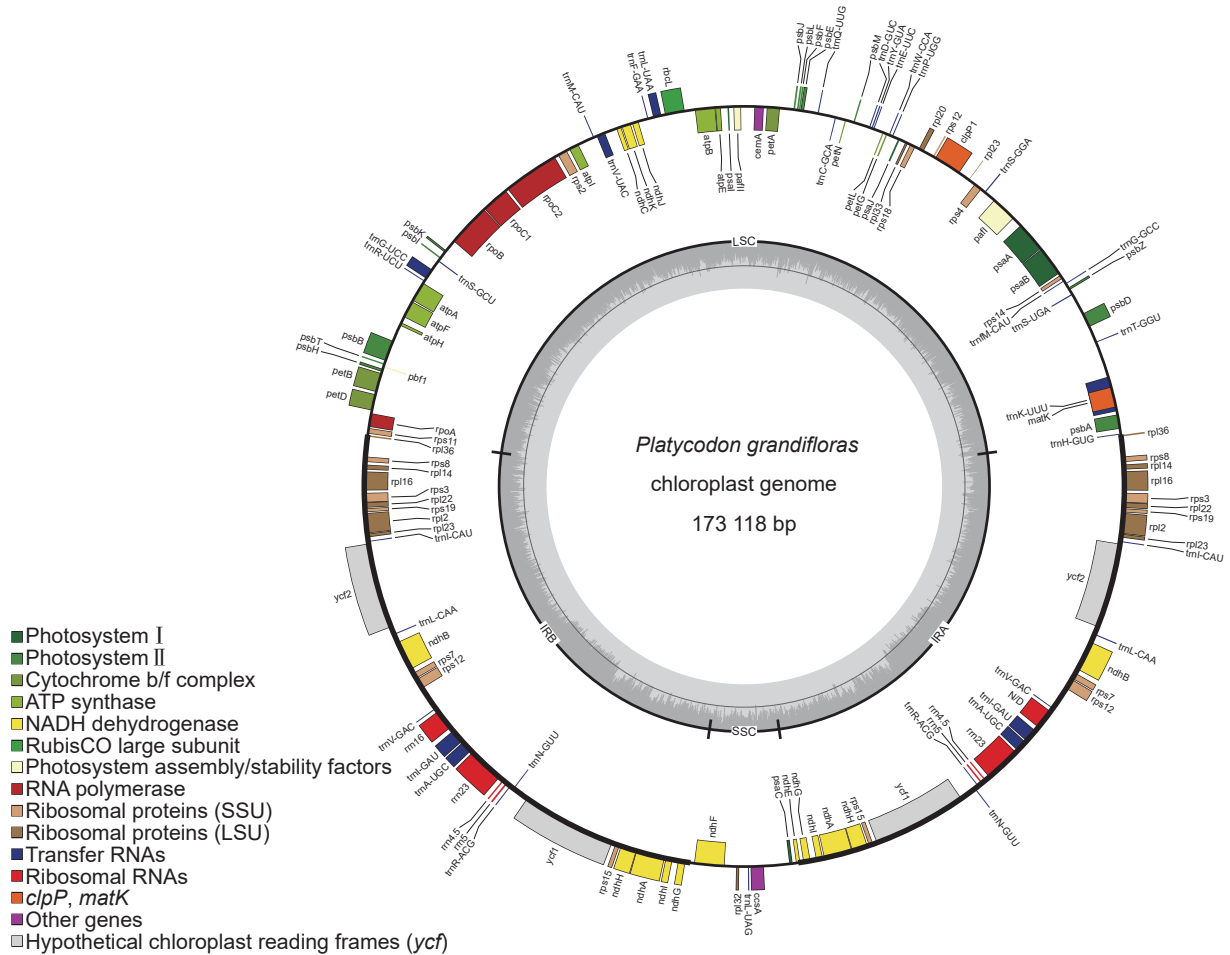


图 1 桔梗的叶绿体基因组图谱

Fig. 1 Chloroplast genome of *Platycodon grandiflorus*

圈内基因按照逆时针方向转录, 圈外基因按照顺时针方向转录。

Genes inside circle are transcribed counter-clockwise, and genes outside circle are transcribed clockwise.

2.2 系统发育关系

用 4 种方法构建出来的系统发育树中 (附图 1¹⁾), 贝叶斯法和最大似然法的结果一致, 邻接树和最大简约树与前面两种树的差别在于半边莲亚科的部分分支, 这些分支的 *rps2* 基因簇并未发生断裂。贝叶斯树除了两个分支的自展值分别为 0.94 和 0.99 外, 其余分支自展值均为 1, 支持度最高。综合前人研究结果, 最终选取基于贝叶斯树的系统发育关系进行后续分析 (图 2)。

2.3 *rps2* 基因簇在桔梗科中的分布信息

rps2 基因簇 (*rps2-atpI-atpH-atpF-atpA*) 在所有物种中均位于 LSC 区。在桔梗科的 50 个物种中, 39 个物种的 *rps2* 基因簇未发生断裂, 11 个物种发生了断裂, 断裂位置均位于 *atpI* 和 *atpH* 基因之间, 且这 11 个物种均来自风铃草亚科, 在系统发育树上聚为一支 (图 2)。其中, 8 个物种断裂后两个片段的排列方式一致但相距较远, 在系统发育树上聚为一支, 剩余 3 个物种的断裂片段

1) 如需查阅附件内容请登录《植物科学学报》网站 (<http://www.plantscience.cn>) 查看本期文章。

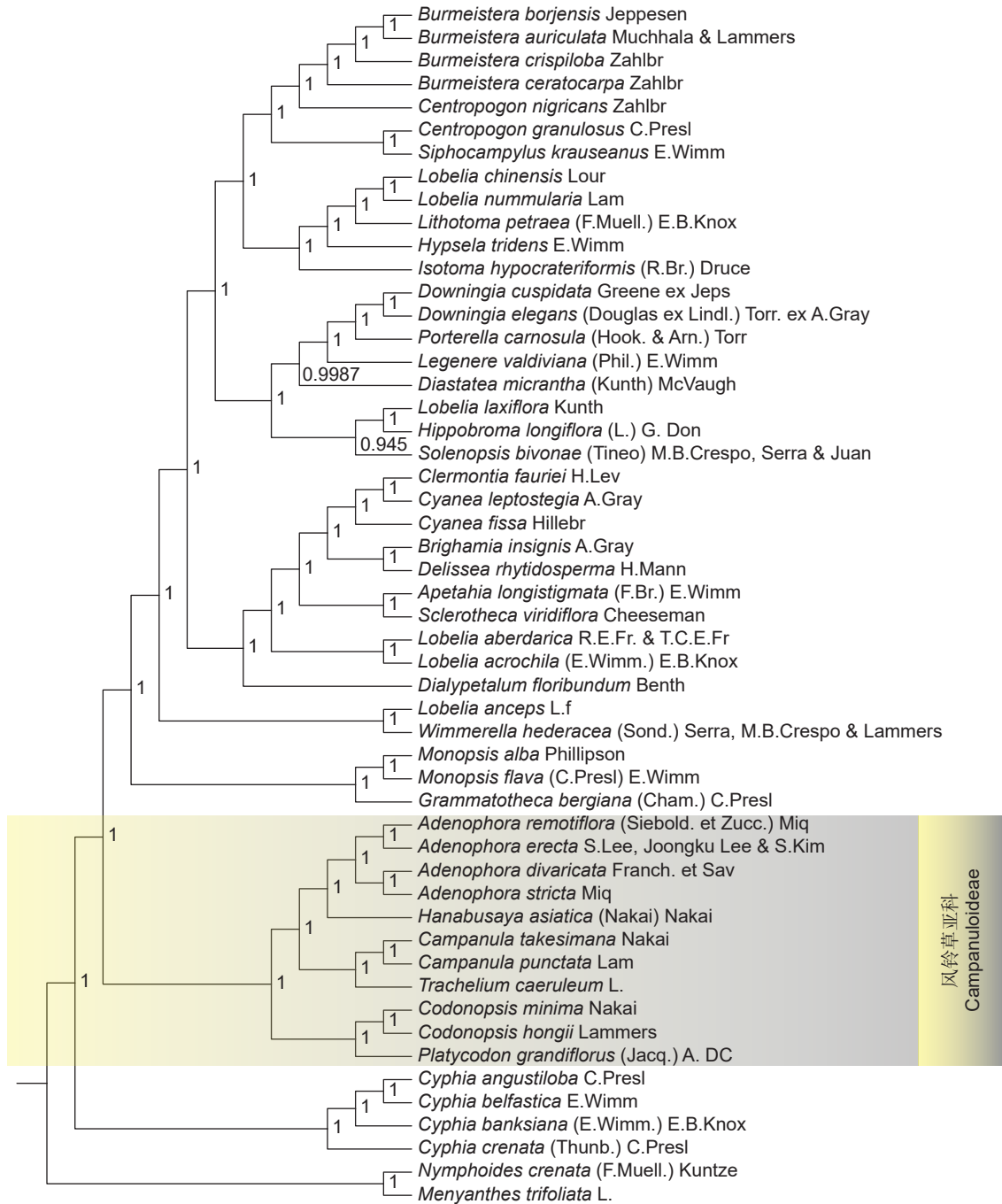


图 2 所选物种构建的贝叶斯树
Fig. 2 Bayesian inference tree of sampled species

分布位置各不相同，转录方向或者排列顺序发生改变，两个片段之间的距离缩短，这 3 个物种在系统发育树上也聚为一支（图 3）。

2.4 *rps2* 基因簇的进化速率分析

对基因簇中 5 个基因的进化速率参数进行分析，根据基因簇是否断裂，将 50 个物种分为未断裂（39 种）和断裂（11 种）两组。在均值上，除了 *atpH* 基因的 ω ，其余未断裂组中相关基因的进

化速率参数均高于断裂组，为断裂组的 1.13 ~ 7.86 倍。曼惠特尼 U 检验（Mann-Whitney U test）结果显示，具有显著差异的有：*atpA* 基因的 *dS* ($P = 0.023$)、*atpH* 基因的 *trst* ($P = 0.040$)、*atpI* 基因的 *ratio* 和 *dS* ($P = 0.038$ 、 $P = 0.026$) 以及 *rps2* 基因的 *trst*、*ratio*、*dS* ($P = 0.011$ 、 $P = 0.047$ 、 $P = 0.010$)（图 4）。

rps2 基因簇发生断裂的 11 个物种根据断裂片

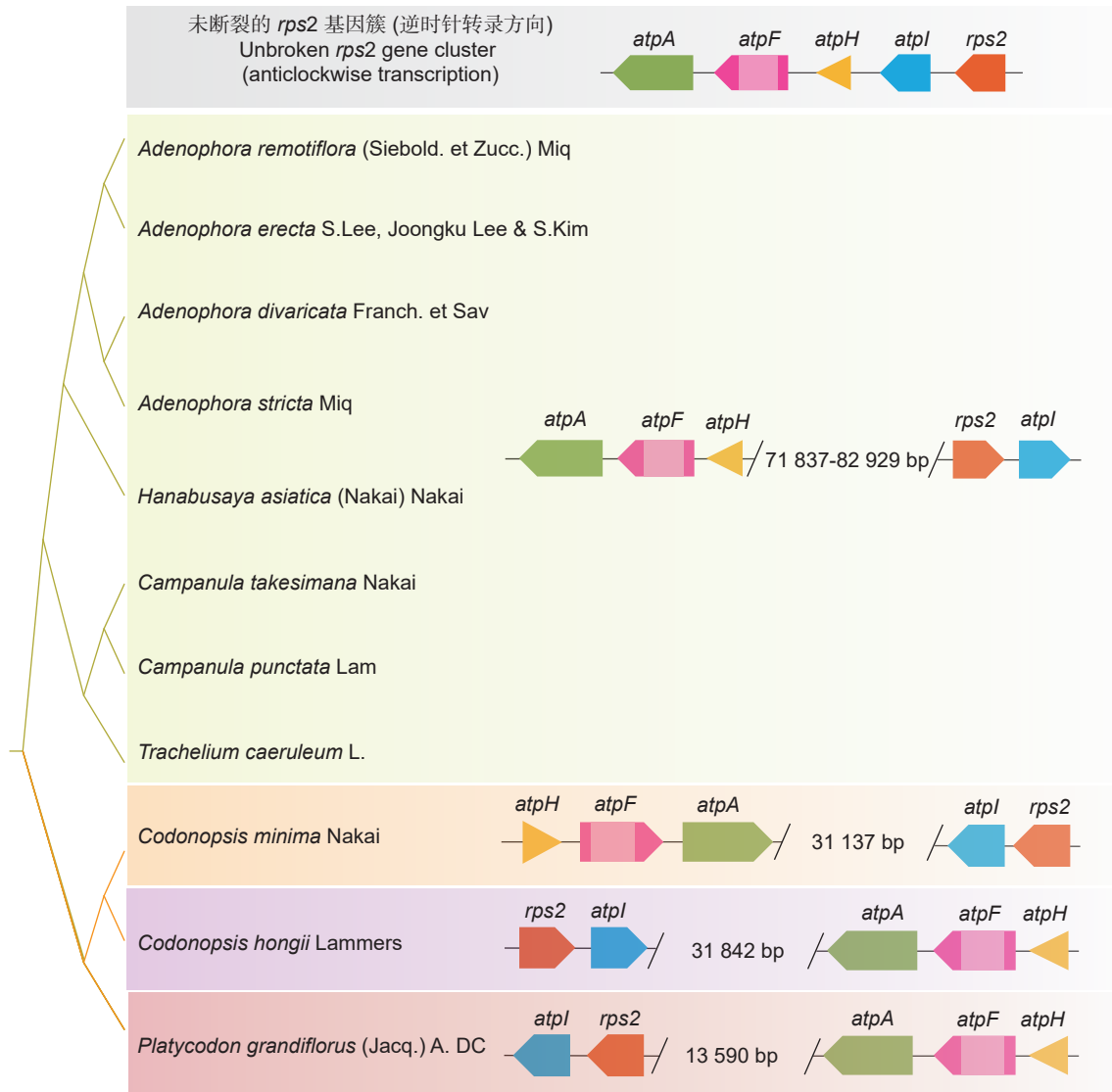


图 3 *rps2* 基因簇在桔梗科中的断裂片段分布
Fig. 3 Distribution of broken *rps2* gene cluster in Campanulaceae

段排布方式可分为 4 组，桔梗 (*Platycodon grandiflorus* (Jacq.) A. DC)、*Codonopsis minima* Nakai 和毛细钟花 (*Codonopsis hongii* Lammers) 的进化速率参数大多高于断裂方式一致的 8 个物种均值 (附图 2¹⁾)。秩和检验结果显示，*atpI* 的 *trst*、*ratio*、 ω 和 *dS* ($P = 0.019$ 、 $P = 0.040$ 、 $P = 0.019$ 、 $P = 0.040$)，*rps2* 的 ω 和 *dS* 差异显著 ($P = 0.019$ 、 $P = 0.019$)。将桔梗、*C. minima* 和毛细钟花 3 个物种与未断裂组进行均值比较，发现这 3 个物种的很多值也高于未断裂组，秩和检验结果显示所有参数均无显著差异 (附图 3²⁾)。对 5 个进化速率参数做秩和检验，结果显示在桔

梗科物种 (50 个) 中，除了 *dS* 无显著差异外 ($P = 0.678$)，其余参数值均存在显著差异 ($P_{trsv} = 0.000$ 、 $P_{trst} = 0.028$ 、 $P_{ratio} = 0.000$ 、 $P_{\omega} = 0.000$ 、 $P_{dN} = 0.000$) (附表 1³⁾)。

2.5 选择压力及适应性进化分析

分支模型中，以断裂组为前景支，未断裂组为背景支，M0 与 Model 2 的似然比检验结果显示，*rps2* 基因在前景支和背景支之间受到的选择压力存在显著差异 ($\omega_{前景支} = 0.875$ ， $\omega_{背景支} = 0.431$ ， $P = 0.029$)；*atpA*、*atpF*、*atpH* 和 *atpI* 基因无显著差异，即每个分支受到的选择压力一致 ($\omega_{atpA} = 0.089$ 、 $\omega_{atpF} = 0.356$ 、 $\omega_{atpH} = 0.026$ 、 $\omega_{atpI} =$

1~3) 如需查阅附件内容请登录《植物科学学报》网站 (<http://www.plantscience.cn>) 查看本期文章。

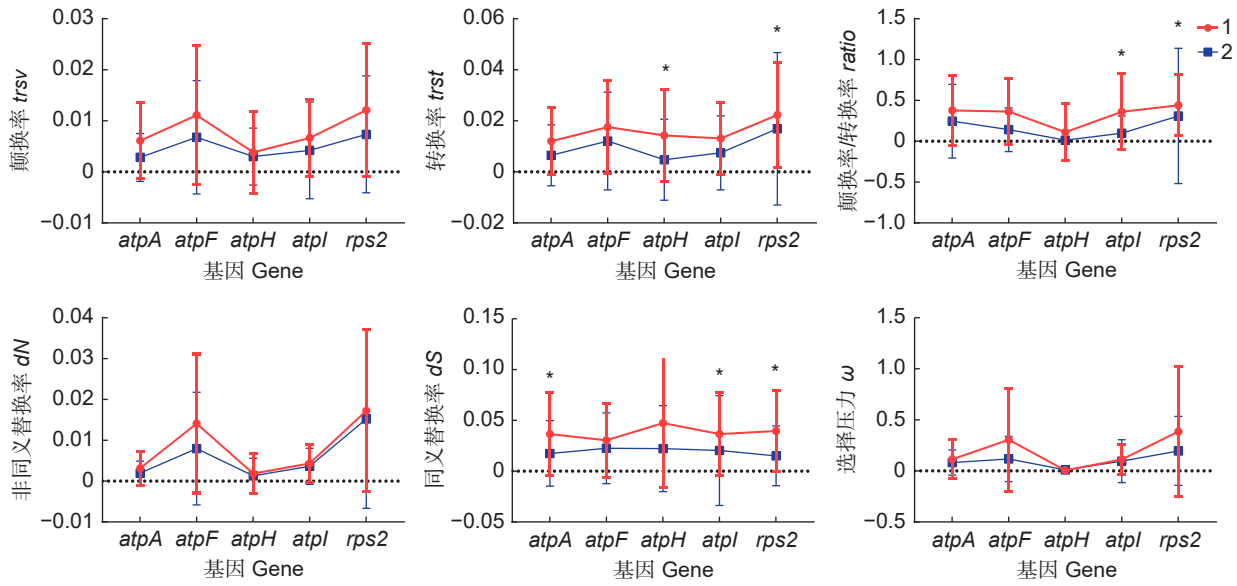


图 4 *rps2* 基因簇未断裂与断裂两组的进化速率均值 ± 标准差比较

Fig. 4 Comparison of mean ± SD of evolutionary rate parameters of *rps2* gene cluster in unbroken and broken groups

1 表示 *rps2* 基因簇未断裂的物种，2 表示 *rps2* 基因簇断裂的物种。* 表示两者的进化速率参数在 $P < 0.05$ 水平上差异显著。
1 is species with unbroken *rps2* gene cluster; 2 is species with broken *rps2* gene cluster. * indicates significant differences at $P < 0.05$ level in evolutionary rate parameters.

0.121)。分支位点模型中前景支与背景支的选择压力均无显著差异，未检测到正选择位点。位点模型中，*atpA*、*atpF*、*atpH* 基因的 M1a 和 M2a 模型的似然比检验结果无显著差异 ($P = 1.000$ 、 $P = 1.000$ 、 $P = 1.000$)，*atpl*、*rps2* 基因的似然比检验结果存在显著差异 ($P = 0.021$ 、 $P = 0.000$)。*atpl* 基因的 M2a 模型中检测到 1 个正选择位点 124H (后验概率 > 95%)；*rps2* 基因中 M2a 模型中检测到 7 个正选择位点 23D、70R、105P、106N、127N、164K、240R，其中 23D、70R、

127N 3 个位点的后验概率超过 99% (附表 2、3¹⁾)。

基于 M1a 与 M2a 的检验结果，将 ATPI 和 RPS2 的蛋白质序列导入到 Swiss-Model 中，以桔梗为例，基于同源建模原理得到蛋白质的三维结构图，*atpl* 基因的 M2a 模型检测到的正选择位点 124H 位于 S (弯曲) 位置。*rps2* 基因的 M2a 模型检测到的正选择位点 23D 和 164K 位于 E (β -折叠)，70R 位于 C (环或无规则卷曲) 位置，105P 和 106N 位于 T (氢键转角) 位置，127N 位于 H (α -螺旋) 位置，240R 位于不规则区域 (图 5)。

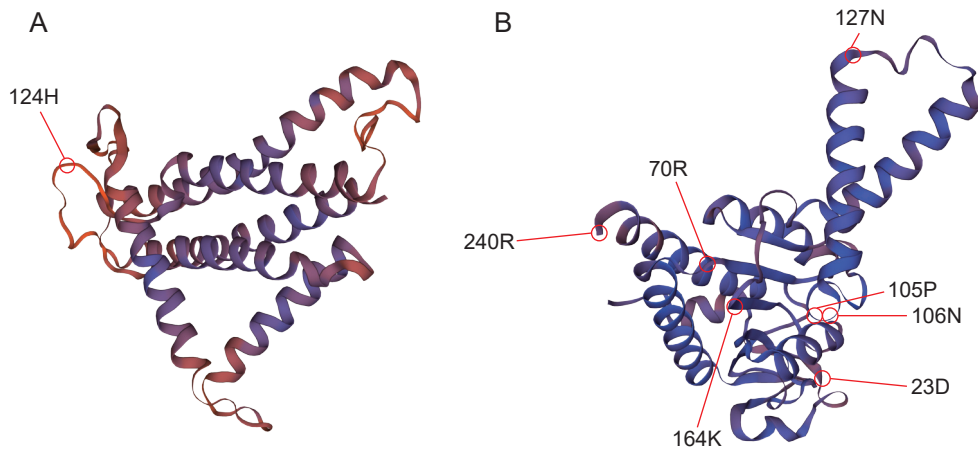


图 5 ATPI (A) 和 RPS2 (B) 蛋白质的三维结构

Fig. 5 Three-dimensional structure of ATPI (A) and RPS2 (B) protein

1) 如需查阅附件内容请登录《植物科学学报》网站(<http://www.plantscience.cn>)查看本期文章。

3 讨论

3.1 桔梗科 *rps2* 基因簇的断裂及其系统发育意义

桔梗科植物叶绿体基因组中有大规模的重组合并现象发生,包括倒位、基因丢失以及 IR 区的收缩/扩张等^[25]。最早在疗喉草 (*Trachelium caeruleum* L.) 的叶绿体基因组中报道过大规模的重排现象,包括 7 ~ 10 个大倒位片段、2 个基因丢失、IR 区的收缩/扩张以及基因簇的断裂等^[26]。重排片段的两端往往伴随着大量重复序列和基因复制,而断裂的基因簇片段与重复序列一样,也分布在重排片段的末端^[27]。保守叶绿体基因簇在序列、基因含量和排列上的差异可能有助于系统发育分析^[28]。在桔梗科中, *rps2* 基因簇发生断裂的位置皆位于 *atpI* ~ *atpH* 基因之间,且发生断裂的物种均来自风铃草亚科,在系统发育树上聚为一支;断裂片段摒弃原本紧密连接的物理位置进行重新排布,重新排布方式相同的 8 个物种在系统发育树上也聚为一支;断裂片段排布方式各不相同的 3 个物种均来自蓝钟花族,在系统发育树上同样聚为一支,说明 *rps2* 基因簇断裂可能具有系统发育意义。本研究结果验证了前人的猜想,同时也为构建桔梗科复杂的系统发育关系提供了一个新思路。另外, *rps2* 基因簇发生断裂的物种均位于桔梗科系统发育关系的基部分支,随着物种的不断进化,断裂现象越来越少,表明 *rps2* 基因簇可能更倾向于成簇转录。

3.2 桔梗科 *rps2* 基因簇与进化速率的关系

研究发现,在细菌和动物线粒体基因组以及被子植物的叶绿体基因组中,核苷酸进化速率和基因组重排率之间存在显著的正相关^[29-31]。究其原因,有人认为在线粒体中这种相关性与 DNA 复制的准确性有关,在叶绿体基因组中则可能是由 DNA 修复机制的异常所引起^[6, 11, 31]。被子植物叶绿体的 DNA 修复中涉及 4 类核编码基因,它们产生抑制重复 DNA 序列间重组的蛋白质,通过防止不正当重组保证基因组的稳定性,一旦编码基因发生突变就会影响错配修复系统的正常运作,使得叶绿体基因组极其不稳定^[32]。那么,这种相关性很有可能是由错配修复系统受到干扰的结果。与其他物种相比, *rps2* 基因簇发生断裂的物种经历了

更多的重排事件,但其相关基因进化速率的均值降低;其中,非同义替换速率无显著差异,同义替换速率在 *atpA*、*atpI*、*rps2* 基因中显著降低。根据断裂片段的排布方式进一步探究,发现桔梗、*C. minima* 和毛细钟花 3 个物种中多个基因的进化速率值均高于未断裂组,而断裂后 *atpA-atpF-atpH* 片段不发生变动,仅 *atpI-rps2* 片段进行重新排布的 8 个物种,其进化速率参数大多为零或接近零,这种结果可能是计算中的假阳性导致,而这也是基因簇断裂后进化速率均值降低的主要原因。同时, *rps2* 基因簇断裂后的进化速率表现出的不一致性变化,可能与断裂片段的重新排布方式或断裂片段间的距离有关。

3.3 桔梗科 *rps2* 基因簇相关基因的适应性进化

生物进化的本质是基因进化^[33]。对叶绿体结构和功能特征的研究表明,质体基因可能不是作为单个位点进化,而是经历着不同的进化力量^[13]。由于基因功能的差异以及各个基因在染色体上的位置不同,基因之间关系复杂,不同基因的核苷酸替换速率有时相差很大,非同义替换速率在基因间的差异比同义替换速率的差异大^[33]。*rps2* 基因簇这 5 个基因的非同义替换率存在显著差异,而同义替换速率并未出现差异,可能是因为 *rps2* 基因簇的 5 个基因分别隶属于两类功能基因。这也体现出了基因间的进化速率异质性,表明进化速率在不同的蛋白编码基因中可能是独立且高度可变的。

叶绿体 ATP 合酶 (CF₁-CF₀ ATP synthase) 主要利用光合作用过程中产生的跨膜质子动力势催化 ATP 的合成,使得光能转化为化学能^[34]。它包含 2 个亚复合物—CF₁ 和 CF₀。CF₁ 主要参与 ATP 的合成或水解,由 α 、 β 、 γ 、 δ 和 ϵ 5 种亚基构成;CF₀ 构成跨膜质子通道的同时也是固定 CF₁ 的支架,由 I ~ IV 4 种亚基构成^[35]。在 *rps2* 基因簇的这 5 个基因中, *atpA*、*atpF*、*atpH* 和 *atpI* 分别编码 CF₁ 的 α 亚基和 CF₀ 的 I、III 和 IV 亚基^[36]。其中 *atpI* 基因中检测到了正选择位点,说明该基因经历了适应性进化,可能会对植物的光合作用及能量转换产生一定的影响。*rps2* 基因是小亚基核糖体蛋白的一种,编码核糖体蛋白 S2,参与核糖体前体的加工和组装,主要与蛋白 S3、S5、S8 交联,同 16S RNA 的 3' 结构域相互作用^[37]。该

基因也是叶绿体基因组中关键的功能基因, 在物种适应不同环境中起着十分重要的作用。

4 结论

桔梗科叶绿体基因组大规模的重排导致 *rps2* 基因簇发生断裂, 这种断裂可能具有一定的系统发育意义。*rps2* 基因簇断裂后, 进化速率均值降低, 但在不同的断裂方式中具有差异性。同时, 基因间进化速率和选择应力的差异可能与基因的功能和结构相关。本研究揭示了叶绿体基因之间广泛的进化速率异质性和基因组结构变异, 为桔梗科叶绿体进化提供了新的见解, 但由于本研究选取物种有限, 得出的结论可能不具备普遍性, 今后还应扩大物种数量, 以获得更准确的结论。

参考文献:

- [1] Zhu B, Qian F, Hou YF, Yang WC, Cai MX, Wu XM. Complete chloroplast genome features and phylogenetic analysis of *Eruca sativa* (Brassicaceae)[J]. *PLoS One*, 2021, 16 (3): e0248556.
- [2] Chan CX, Gross J, Yoon HS, Bhattacharya D. Plastid origin and evolution: new models provide insights into old problems[J]. *Plant Physiol*, 2011, 155 (4): 1552–1560.
- [3] Hsu CY, Wu CS, Chaw SM. Birth of four chimeric plastid gene clusters in Japanese umbrella pine[J]. *Genome Biol Evol*, 2016, 8 (6): 1776–1784.
- [4] 沈双欠, 占传松, 杨陈坤, 罗杰. 植物代谢基因簇的研究进展[J]. *中国科学: 生命科学*, 2021, 51(12): 1734–1746.
Shen SQ, Zhan CS, Yang CK, Luo J. Research progress on plant biosynthetic gene cluster[J]. *Scientia Sinica Vitae*, 2021, 51 (12): 1734–1746.
- [5] Perry AS, Brennan S, Murphy DJ, Kavanagh TA, Wolfe KH. Evolutionary re-organisation of a large operon in adzuki bean chloroplast DNA caused by inverted repeat movement[J]. *DNA Res*, 2002, 9 (5): 157–162.
- [6] Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the Angiosperm family Geraniaceae: rearrangements, repeats, and codon usage[J]. *Mol Biol Evol*, 2011, 28 (1): 583–600.
- [7] Cai ZQ, Guisinger M, Kim HG, Ruck E, Blazier JC, et al. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions[J]. *J Mol Evol*, 2008, 67 (6): 696–704.
- [8] Haberle RC, Fourcade HM, Boore JL, Jansen RK. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes[J]. *J Mol Evol*, 2008, 66 (4): 350–361.
- [9] Drager RG. Structure and transcript processing of a *Euglena gracilis* chloroplast operon encoding genes *rps2*, *atpI*, *atpH*, *atpF*, *atpA* and *rps18*[D]. Tucson: The University of Arizona, 1993: 13–19.
- [10] Stahl DJ, Rodermeil SR, Bogorad L, Subramanian AR. Co-transcription pattern of an introgressed operon in the maize chloroplast genome comprising four ATP synthase subunit genes and the ribosomal *rps2*[J]. *Plant Mol Biol*, 1993, 21 (6): 1069–1076.
- [11] Jansen RK, Ruhlman TA. Plastid genomes of seed plants[M]//Bock R, Knoop V, eds. Genomics of Chloroplasts and Mitochondria. Dordrecht: Springer, 2012: 103–126.
- [12] The Angiosperm Phylogeny Group, Chase MW, Christenhusz MJM, Fay MF, Byng JW, et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV [J]. *Bot J Linn Soc*, 2016, 181 (1): 1–20.
- [13] Zhang X, Sun YX, Landis JB, Lv ZY, Shen J, et al. Plastome phylogenomic study of Gentianeae (Gentianaceae): widespread gene tree discordance and its association with evolutionary rate heterogeneity of plastid genes[J]. *BMC Plant Biol*, 2020, 20 (1): 340.
- [14] 平晶耀, 冯佩沛, 郝静, 栗锦焯, 苏应娟, 王艇. *rps12*基因在裸子植物中的分子进化式样[J]. *科学通报*, 2021, 66(24): 3182–3193.
Ping JY, Feng PP, Hao J, Li JY, Su YJ, Wang T. The molecular evolution pattern of *rps12* gene in gymnosperms[J]. *Chinese Science Bulletin*, 2021, 66 (24): 3182–3193.
- [15] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. *Bioinformatics*, 2014, 30 (15): 2114–2120.
- [16] Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data[J]. *Nucleic Acids Res*, 2017, 45 (4): e18.
- [17] Zhao FQ, Zhao FG, Li T, Bryant DA. A new pheromone trail-based genetic algorithm for comparative genome assembly[J]. *Nucleic Acids Res*, 2008, 36 (10): 3455–3462.
- [18] Greiner S, Lehwark P, Bock R. Organellar Genome DRAW (OGDRAW) version 1.3. 1: Expanded toolkit for the graphical visualization of organellar genomes[J]. *Nucleic Acids Res*, 2019, 47 (W1): W59–W64.
- [19] Kearsley M, Moir R, Wilson A, Stones-Havas S, Cheung M, 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.
- [20] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. *Mol Biol Evol*, 2016, 33 (7): 1870–1874.
- [21] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies[J]. *Bioinformatics*, 2014, 30 (9): 1312–1313.
- [22] Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees[J]. *Bioinformatics*, 2001, 17 (8): 754–755.
- [23] Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies[J]. *Bioinformatics*, 2005, 21 (5): 676–679.
- [24] Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood[J]. *Mol Biol Evol*, 2007, 24 (8): 1586–1591.
- [25] Cosner ME, Raubeson LA, Jansen RK. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes[J]. *BMC Evol Biol*, 2004, 4: 27.
- [26] Cosner ME, Jansen RK, Palmer JD, Downie SR. The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families[J]. *Curr Genet*, 1997, 31 (5): 419–429.
- [27] Haberle RC. Phylogeny and comparative chloroplast genomics of the campanulaceae[D]. Austin: The University of Texas at Austin, 2006: 71–73.
- [28] Gray MW. Origin and evolution of plastid genomes and genes[M]//Bogorad L, Vasil IK, eds. *The Molecular Biology of Plastids*. New York: Academic Press, 1991: 303–330.
- [29] Belda E, Moya A, Silva FJ. Genome rearrangement distances and gene order phylogeny in γ -Proteobacteria[J]. *Mol Biol Evol*, 2005, 22 (6): 1456–1467.
- [30] Shao RF, Downton M, Murrell A, Barker SC. Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects[J]. *Mol Biol Evol*, 2003, 20 (10): 1612–1619.
- [31] Jansen RK, Cai ZQ, Raubeson LA, Daniell H, Depamphilis CW, *et al.* Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns[J]. *Proc Natl Acad Sci USA*, 2007, 104 (49): 19369–19374.
- [32] 文艺. 桔梗科质体基因组进化研究[D]. 北京: 中国科学院大学, 2020: 33–34.
- [33] 李易. 基因进化的同义与非同义替代计算及统计检验的比较分析[J]. *曲靖师范学院学报*, 2006, 25 (6): 1–8.
- Li Y. Comparing and analysing on models of calculation and statistical testing of nosynonymous substitution rate and synonymous substitution rate during gene evolution[J]. *Journal of Qujing Normal University*, 2006, 25 (6): 1–8.
- [34] Hahn A, Vonck J, Mills DJ, Meier T, Kühlbrandt W. Structure, mechanism, and regulation of the chloroplast ATP synthase[J]. *Science*, 2018, 360 (6389): eaat4318.
- [35] 张琳, 彭连伟. 叶绿体ATP合酶生物发生的研究进展[J]. *植物生理学报*, 2019, 55 (6): 703–710.
- Zhang L, Peng LW. Research progress on regulation mechanisms of biogenesis of chloroplast ATP synthase[J]. *Plant Physiology Journal*, 2019, 55 (6): 703–710.
- [36] Miyagi T, Kapoor S, Sugita M, Sugiura M. Transcript analysis of the tobacco plastid operon *rps2/atpl/H/F/A* reveals the existence of a non-consensus type II (NC II) promoter upstream of the *atpl* coding sequence[J]. *Mol Gen Genet*, 1998, 257 (3): 299–307.
- [37] Harris EH, Boynton JE, Gillham NW. Chloroplast ribosomes and protein synthesis[J]. *Microbiol Rev*, 1994, 58 (4): 700–754.

(责任编辑: 李惠英)