

五加科植物叶绿体基因组结构与进化分析

宋 菊, 龙月红, 林丽梅, 尹 峰, 邢朝斌*

华北理工大学生命科学学院, 河北 唐山 063210

摘 要: 目的 获取五加科叶绿体基因组的结构特征以及进化关系。方法 以已测序完成的 10 个属的 20 种五加科植物叶绿体基因组为研究对象, 系统比较基因组间的差异, 分析 4 个 IR 边界扩张与收缩情况, 并以近缘物种当归为外类群, 使用 MEGA 4.0 构建系统进化树, 分析物种间的亲缘关系。结果 20 种五加科植物的叶绿体基因组大小差异较小, 最大差仅 1 909 bp。各物种均出现基因替换现象, 由 *cemA* 基因替换 *ycf10* 基因, 且数量存在一定差异, 主要由 tRNA 引起。通过比较发现, 五加科物种的 4 个 IR 边界比较保守, 仅野三七、三七和穗序鹅掌柴的边界基因进入 IR 区的长度与其他物种的差异较大。以当归为外类构建的系统进化树, 各节点的支持率较高, 清晰地反映了各物种间的亲缘关系。结论 叶绿体基因组涵盖的信息量大, 可以用于分析关系较近与进化较快物种的系统发生问题。

关键词: 五加科; 叶绿体基因组; 刺五加; 人参; 三七; 穗序鹅掌柴; 系统进化

中图分类号: R282.12

文献标志码: A

文章编号: 0253-2670(2017)24-5070-06

DOI: 10.7501/j.issn.0253-2670.2017.24.002

Analysis on structure and phylogeny of chloroplast genomes in *Araliaceae* species

SONG Ju, LONG Yue-hong, LIN Li-mei, YIN Feng, XING Zhao-bin

College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China

Abstract: Objective To obtain the structural characteristics and phylogenetic relationships of the chloroplast genome in *Araliaceae* species. **Methods** we used 20 chloroplast genomes which have sequenced as materials, they were from 20 species belonging to 10 genus of *Araliaceae*. Analysis the differences of genomes and the dilation or shrink of four boundaries for IR, we used MEGA 4.0 to build the phylogenetic tree with *Angelica gigas* of sibling species as the outgroup and analysis their phylogenetic relationships. **Results** There was a small difference among the chloroplast genomes size, and the largest difference is 1 909 bp. All of species had existed gene replacements, *cemA* replaced *ycf10*, and number of genes existed some differences, they were mainly caused by tRNA. The four boundaries of IR was relatively conservative, only *Panax vietnamensi*, *Panax notoginseng* and *Schefflera delavayi* were special, their boundary genes were in IR. All nodes of the phylogenetic tree of *Araliaceae* which was based on *Angelica gigas* were of high supports, and the tree had good resolution to reflect the genetic relationships among *Araliaceae*. **Conclusion** Chloroplast genomes have a lot of information, it can be used to analysis phylogeny among the species which are affinity or faster evolution.

Key words: *Araliaceae*; chloroplast genomes; *Eleutherococcus senticosus* (Rupr. et Maxim.) Maxim.; *Panax ginseng* C. A. Mey.; *Panax notoginseng* (Burkill) F. H. Chen ex C. Chow; *Schefflera delavayi* (Lour.) Harms; phylogeny

叶绿体是绿色植物进行光合作用的细胞器, 具有合成蛋白质、淀粉、色素等功能, 普遍存在于真核自养生物中, 尤其是藻类和陆生植物, 其基因组可自主遗传^[1]。在被子植物中, 叶绿体基因组比较保守, 其结构大多为双链环形, 具有 4 个区域, 即 2 个序列相同方向相反的重复区 (inverted repeats, IRB/IRA)、1 个大单拷贝区 (large single copy, LSC)

和 1 个小单拷贝区 (small single copy, SSC)^[2]。一般情况下, 基因组大小为 120~180 kb, 有 100~120 类编码基因, 包含 70~88 类蛋白编码基因, 30~32 类 tRNA, 4 类 rRNA^[3]。

五加科是伞形目下的一个大科, 包含 2 个亚科 45 个属, 有 1 400 多种植物, 多为珍贵的药用植物, 如人参 *Panax ginseng* C. A. Mey.、三七 *Panax notoginseng*

收稿日期: 2017-09-13

基金项目: 国家自然科学基金项目 (31570683); 华北理工大学培育基金资助 (SP201508); 华北理工大学研究生创新项目 (2017S12)

作者简介: 宋 菊 (1992—), 女, 在读硕士研究生, 研究方向为生物信息在药用植物中的应用。

*通信作者 邢朝斌 (1975—), 男, 教授, 研究方向为分子生药学、药用植物细胞工程。

Tel: (0315)3725859 Fax: (0315)3726341 E-mail: xzbheuu@126.com

(Burkill) F. H. Chen ex C. Chow、刺五加 *Eleutherococcus senticosus* (Rupr. et Maxim.) Maxim.、刺楸 *Kalopanax septemlobus* (Thunb.) Koidz. 等, 具有抗肿瘤、血糖调节、抗衰老等多种药理作用^[4], 受到人们的极大关注。叶绿体的基因数量、基因顺序以及结构组成都相对稳定^[5]; 所含的信息量大, 进化速率适中^[6]; 基因组序列长度较短, 易获取, 且没有核基因与线粒体基因复杂^[7]; 使其成为植物系统发育分析的最佳选择, 为物种间的亲缘关系分析提供可靠的信息。

本研究通过对五加科 20 种植物的叶绿体基因

组的比较分析, 揭示了叶绿体基因组的结构特征与 IR 区的边界扩张与收缩问题, 展示了 20 个物种的进化关系, 为进一步研究该科物种的遗传背景、药用植物鉴定、资源保护与开发利用奠定下基础。

1 材料

GenBank 中已发表的五加科植物叶绿体基因组。

2 方法

2.1 数据的收集

在 GenBank 中检索已发表叶绿体基因组的五加科植物, 共计 20 个物种 (表 1), 下载基因组序列, 获取注释信息。

表 1 所选植物叶绿体基因组

Table 1 Chloroplast genomes in present study

科	属	种	拉丁学名	GenBank 登录号
五加科	五加属	刺五加	<i>Eleutherococcus senticosus</i>	JN637765.1
		无梗五加	<i>Eleutherococcus sessiliflorus</i>	KT153019.1
		五加	<i>Eleutherococcus gracilistylus</i>	KT153020.1
	人参属	人参	<i>Panax ginseng</i>	KF431956.1
		三七	<i>Panax notoginseng</i>	NC_026447.1
		西洋参	<i>Panax quinquefolius</i>	KT028714.1
		野三七	<i>Panax vietnamensis</i>	KU059178.1
		屏边三七	<i>Panax stipuleanatus</i>	NC_030598.1
		大叶三七	<i>Panax japonicus</i>	KX247146.1
		鹅掌柴属		
	鹅掌柴属	鹅掌柴	<i>Schefflera heptaphylla</i>	NC_029764.1
		穗序鹅掌柴	<i>Schefflera delavayi</i>	NC_022813.1
	槲木属	波缘槲木	<i>Aralia undulata</i>	NC_022810.1
		辽东槲木	<i>Aralia elata</i>	KT153023.1
	树参属	黄漆木	<i>Dendropanax morbifer</i>	NC_027607.1
		树参	<i>Dendropanax dentiger</i>	NC_026546.1
	八角金盘属	八角金盘	<i>Fatsia japonica</i>	NC_027685.1
	刺楸属	刺楸	<i>Kalopanax septemlobus</i>	NC_022814.1
	梁王茶属	梁王茶	<i>Metapanax delavayi</i>	NC_022812.1
	掌叶树属	浅裂掌叶树	<i>Brassaiaopsis hainla</i>	NC_022811.1
	马蹄参属	马蹄参	<i>Diplopanax stachyanthus</i>	KP318983.1
伞形科	当归属	当归	<i>Angelica gigas</i>	KX118044.1

2.2 数据统计

利用 Excel 统计 20 种五加科植物叶绿体基因组的大小、GC 量、LSC、SSC、IR 的长度等基本结构特征信息。对于未知 LSC/SSC/IRs 边界信息的植物叶绿体基因组 (如无梗五加、屏边三七、辽东槲木等) 使用 Blast 进行自身序列比对, 先确认 IRs 的长度与边界, 最后获取 LSC 和 SSC 的长度信息。

2.3 系统进化树的构建与分析

使用 ClustalX 2.0 做多重序列比对, 选择默认参数, 序列间的差异度使用 Jukes-Cantor 校正。采

用 MEGA 4.0 中的 NJ (neighbor-joining) 法构建系统进化树, Bootstrap method 进行重复检验, 重复 1 000 次, 选择伞形科当归 (表 1) 作为外类群。

3 结果与分析

3.1 五加科叶绿体基因组结构基本特征

20 种五加科植物的叶绿体基因组大小不尽相同, 八角金盘最小, 长 155 613 bp, 马蹄参最大, 长 157 522 bp, 相差约 2 kb。其中 LSC 占总长度的绝大部分, 其差异也最大, 大小在 86 028~87 640 bp, 最大者是马蹄参, 最小者是波缘槲木, 相差约

1.5 kb。SSC 在 17 867~18 247 bp, 仅相差 380 bp。2 个 IR 区长度为 25 629~26 180 bp, 各物种间的具有差异。大部分物种的 IR 区长度相同, 仅有西洋参的 IRB 与 IRA 不相同, 两者之间相差 4 个碱基。

GC 量均分布在 38.00%左右, 同属之间, 差异较小, 五加属的 3 个叶绿体基因组 GC 量完全一致, 鹅掌柴属差异最大, 为 0.10%。不同属间的最大差异为 0.32% (表 2)。

表 2 五加科叶绿体基因组结构特征

Table 2 Structural feature of chloroplast genomes in Araliaceae

物种	基因组大小/bp	LSC/bp	SSC/bp	IR/bp	GC/%	tRNA	rRNA	蛋白编码基因	总基因数
刺五加	156 768	86 755	18 153	25 930	37.95	37	8	87	132
无梗五加	156 730	86 603	18 213	25 957	37.95	37	8	87	132
五加	156 770	86 729	18 175	25 933	37.95	37	8	87	132
人参	156 355	86 129	18 078	26 074	38.08	36	8	86	129
三七	156 387	86 111	18 024	26 126	38.08	37	8	87	132
西洋参	156 359	86 124	18 079	26 076/26 080	38.08	36	8	87	131
野三七	155 993	86 118	17 989	25 943	38.05	36	8	86	129
屏边三七	156 064	86 137	18 195	25 866	38.03	37	8	86	131
大叶三七	156 063	86 112	18 173	25 889	38.07	37	8	86	131
鹅掌柴	156 685	86 609	18 146	25 965	37.93	29	8	86	123
穗序鹅掌柴	156 341	86 123	18 146	26 036	37.83	37	8	87	132
波缘橐木	156 333	86 028	18 089	26 180	38.10	38	8	87	133
辽东橐木	156 220	86 263	18 111	25 923	38.08	37	8	87	132
黄漆木	156 366	86 475	18 125	25 883	37.99	37	8	87	132
树参	156 687	86 680	18 247	25 880	37.96	37	8	87	132
八角金盘	155 613	86 488	17 867	25 629	37.91	37	8	87	132
刺楸	156 413	86 466	18 119	25 914	37.95	37	8	87	132
梁王茶	156 343	86 361	18 130	25 926	37.93	37	8	87	132
浅裂掌叶树	156 459	86 566	18 021	25 936	37.92	37	8	87	132
马蹄参	157 522	87 640	18 182	25 850	37.78	37	8	87	132

3.2 基因数量分析

总基因在 123~133 个(表 2), 13 个植物有 132 个基因, 3 个有 131 个基因, 波缘橐木基因树最多, 有 133 个, 而鹅掌柴的基因最少, 有 123 个。其中 rRNA 最保守, 数量均为 8 个。蛋白编码基因主要为 87 个, 仅有 5 个物种是 86 个。它们在进化过程中均出现了基因替换现象, 由 *cemA* 基因替换了原有 *ycf10* 基因 (在某些物种中注释为 ORF229)。仅有 4 个物种发生了基因丢失, 大叶三七和野三七还丢失 *ycf15* 基因, 西洋参丢失 *petD* 基因, 鹅掌柴丢失 *rps16* 基因和 *psbN* 基因。

蛋白编码基因总体差异较小, 因此总基因数的差异主要是由 tRNA 基因引起。大部分物种的 tRNA 是 37 个, 波缘橐木最多 38 个, 鹅掌柴最少 29 个, 具有差异的主要基因是 *trnG* 与 *trnI*。

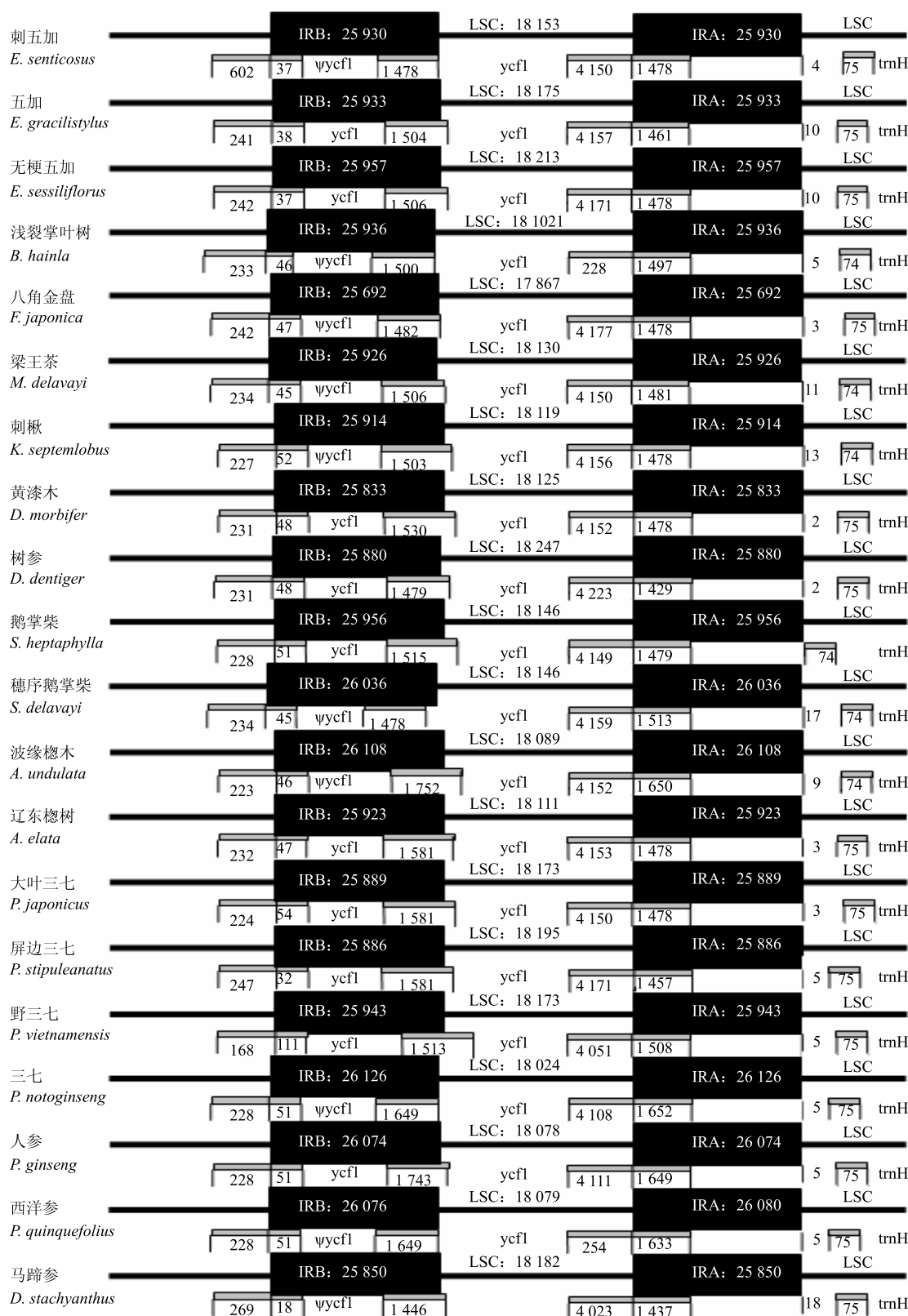
3.3 IR 边界变化分析

叶绿体基因组为环形结构, IR 与 LSC 和 SSC 存在 4 个边界, 即 IRB-LSC、IRB-SSC、IRA-SSC

和 IRA-LSC。在基因组进化过程中, IR 边界会发生扩张与收缩, 使某些基因进入 IR 区或单拷贝区。通过研究发现, 五加科叶绿体基因组的 4 个边界相对较保守(图 1)。所有物种的 IRB-LSC 边界均在 *rps19* 基因内, 且 *rps19* 基因序列大部分在 LSC 区, 野三七的 *rps19* 基因在 IRB 区的序列长度为 111 bp, 其余仅有 16~54 bp。大部分物种的 IRB-SSC 边界处于 *ycf1* 假基因或 *ycf1* 上, 三七和穗序鹅掌柴的 *ycf1* 假基因在 IRB 内部。所有物种的 IRA-SSC 边界均处于 *ycf1* 基因上, *ycf1* 基因存在于 IRA 内的序列长度在 1 429~1 652 bp, 差异较小。IRA-LSC 边界与 LCS 上的 *trnH* 连续或具有较短的间隔, 间隔为 2~18 bp。

3.4 系统发育分析

20 种五加科植物的叶绿体基因组经过全局比对, 以当归作为外类群, 构制了系统进化树(图 2), 其中 18 个节点, 支持率最低为 70%, 14 个为 100%, 其余 3 个均在 90%以上, 可信度较高。



ψycf1 表示 ycf1 假基因, 数字代表长度; 黑色方框代表基因, 长度与图形大小不成比例, 且基因与 LSC、SSC 和 IR 区不成比例
 ψycf1 is ycf1 pseudogene, the numbers represent length, unit is bp; black squares denote genes, length is disproportional to the size of figures, and genes are also disproportional to the LSC, SSC and IR lengths

图 1 20 个五加科物种叶绿体基因组 4 个 IR 区边界比较
 Fig. 1 Comparison of boundaries of chloroplast genomes from 20 species in Araliaceae

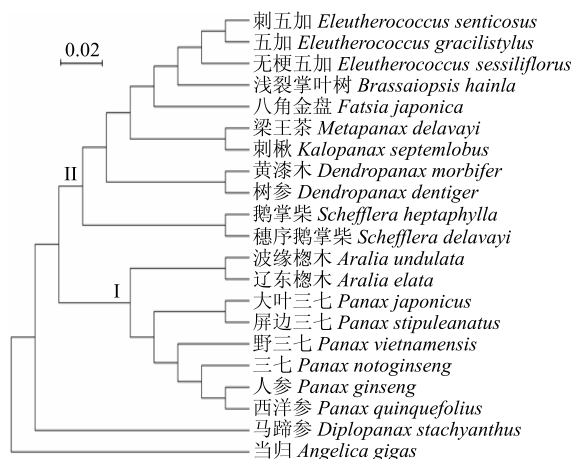


图 2 基于五加科叶绿体全基因组序列构建系统进化树

Fig. 2 Phylogenetic tree based on chloroplast genome sequence of Araliaceae

从图 2 中可以看出, 马蹄参(多蕊木族, 马蹄参属)是五加科植物中最早分离出来的物种, 与其他物种的亲缘关系较远, 其余的 19 种五加科植物聚到 2 支进化枝, I 进化枝和 II 进化枝。人参属(人参族)和槲木属(槲木族)的人参、三七、屏边三七等 8 个物种聚到 I 进化枝, 槲木属的 2 个物种先分离出来, 与人参属的 3 个分支成姐妹关系。II 进化枝则为多蕊木族物种, 包含刺五加、八角金盘、树参等 7 个属的 11 个物种, 其中鹅掌柴属物种较早分离, 与其他 6 个属的 9 个物种形成的分支形成姐妹关系。

4 讨论

被子植物的叶绿体基因组大小在 120~180 kb, IR 区的大小在 20~30 kb^[3]。20 种五加科植物的叶绿体基因组大小在 155~158 kb, IR 区长 25~27 kb, 与被子植物叶绿体基因组特征相吻合。与近缘物种当归相比, 基因组增大, 而 IR 区有所减小。通过比较发现, 20 个基因组大小差异度约为 2 kb, 而 LSC、SSC 与 IR 4 个区域的差异呈现为 LSC (约 1.5 kb) > IR (551 bp) > SSC (380 bp), 因此整个基因组长度的差异主要是由 LSC 的差异引起。

叶绿体基因组上的基因转换能力较高, 能保证 2 个 IR 区序列的一致与稳定, 增强了叶绿体基因组的稳定性和保守性^[8]。本研究, 已测序完成的 20 个五加科植物, 分类于 10 个属中, 属间与属内的叶绿体基因组 IR 区差异都很小, 属内最大差异仅 260 bp, 属间差异为 551 bp。天竺葵属^[9] IR 区长达 75 kb, 柳杉^[10] IR 区仅 114 bp, 豌豆^[11] IR 区丢失, 五加科

叶绿体基因组并未出现这些 IR 区过长或过短甚至丢失的现象, 整体比较稳定与一致。

叶绿体基因组中存在一个普遍的进化现象, 即 4 个 IR 边界发生扩张与收缩事件使相同植物群体或不同植物群体的整个叶绿体基因组大小出现差异或较大差异。在猕猴桃^[12]中, rps19 基因没有进入 IR 区, 整个基因位于 LSC 区, 不是 IRB-LSC 边界基因。而在五加科中, rps19 基因是叶绿体基因组 IRB-LSC 的边界基因, 且该边界均在其内部。rps19 基因进入 IRB 区的序列大部分在 16~54 bp, 仅野三七较长, 111 bp, 整体比十字花本科^[13]短。IRA-SSC 边界与 LSC 上的 trnH 基因存在 0~18 bp 的间隔, 而在单子叶植物中, 如兰科的文心兰和蝴蝶兰, 禾本科的水稻、大麦、剪股颖等^[14]植物叶绿体的这 2 个边界发生扩张, 使 rps19 与 trnH 完全进入 IR 区, 由一个拷贝变成了 2 个拷贝, 而之后大麦与剪股颖的这 2 个边界又发生了收缩事件, 使 rps19 与 trnH 回到了 LSC 区。IRB-SSC 边界和 IRA-SSC 边界几乎都位于 ycf1 或 ycf1 假基因上, 由于边界的扩张, 导致三七和穗序鹅掌柴的 ycf1 假基因全部进入到 IRB 中, 又 IRA-SSC 边界的收缩或扩张, 使 ycf1 在 IRA 中的长度存在一定差异, 这 2 个边界都不同于单子叶植物高粱、玉米、甘蔗等^[14]。这都表明双子叶植物与单子叶植物的叶绿体基因组边界具有较大差异^[15]。

本研究利用五加科植物叶绿体全基因组序列构建了 10 个属 20 个物种的系统进化树, 结果表明, 该进化树的分辨率较高, 各节点也获得了较高的支持率, 每个属之间呈现出较为明确的发育关系。其中不同属的梁王茶与刺楸聚为一枝, 且获得较高支持率, 95%, 说明两者之间的亲缘关系很近, 高于同其他属的物种, 与 Li 等^[16]的研究结果一致。马蹄参位于进化树的基部进化枝, 应是五加科的原始类群物种。

参考文献

- [1] Neuhaus H E, Emes M J. Nonphotosynthetic metabolism in plastids [J]. *Annu Rev Plant Physiol Plant Mol Biol*, 2000, 51(51): 111-140.
- [2] 王 玲, 董文攀, 周世良. 被子植物叶绿体基因组的结构变异研究进展 [J]. *西北植物学报*, 2012, 32(6): 1282-1288.
- [3] Zhang T, Fang Y, Wang X, et al. The complete chloroplast and mitochondrial genome sequences of *Boea*

- hygrometrice*: Insight into the evolution of plant organellar genomes [J]. *PLoS One*, 2012, 7: e30531.
- [4] 任玉新, 金立弟, 韩宇. 五加科植物多糖的药理作用研究进展 [J]. 中国药业, 2012, 21(6): 83-86.
- [5] Nock C J, Waters D J, Edwards M A, *et al.* Chloroplast genome sequence from total DNA for plant identification [J]. *Plant Biotechnol J*, 2011, 9(3): 328-333.
- [6] 邢少辰, Clarke J L. 叶绿体基因组研究进展 [J]. 生物化学与生物物理进展, 2008, 31(1): 21-28.
- [7] 田欣, 李德铎. DNA 序列在植物系统学研究中的应用 [J]. 云南植物研究, 2002, 2(2): 170-184.
- [8] Khakhlova O, Bock R. Elimination of deleterious mutations in plastid genome by gene conversion [J]. *Plant J*, 2006, 46(1): 85-94.
- [9] Chumley T W, Plamer J D, Mower J P, *et al.* The complete chloroplast genome sequence of *Pelargonium × hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants [J]. *Mol Biol Evol*, 2006, 23(11): 2175-2190.
- [10] Hirao T, Watanabe A, Kurita M, *et al.* Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: Diversified genomic structure of coniferous species [J]. *BMC Plant Biol*, 2008, 8: 70.
- [11] Saski C, Lee S B, Daniell H, *et al.* Complete chloroplast genome sequence of *Glycine max* and comparative analysis with other legume genomes [J]. *Plant Mol Biol*, 2005, 59(2): 309-322.
- [12] Yao X H, Tang P, Li Z Z, *et al.* The first complete chloroplast genome sequence in Actinidiaceae: Genome structure and comparative analysis [J]. *PLoS One*, 2015, 10(6): e0129347.
- [13] 李岩, 吕光辉, 张雪妮, 等. 十字花科植物叶绿体基因组结构及变异分析 [J]. 西北植物学报, 2017, 37(6): 1090-1101.
- [14] 唐萍, 阮秋燕, 彭程. 禾本科植物叶绿体基因组结构的系统进化分析 [J]. 中国农学通报, 2011, 27(30): 171-176.
- [15] Chang C C, Lin H C, Lin I P, *et al.* The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications [J]. *Mol Biol Evol*, 2006, 23(2): 279-291.
- [16] Li R, Ma P, Wen J, *et al.* Complete sequencing of five *araliaceae* chloroplast genomes and the phylogenetic implications [J]. *PLoS One*, 2013, 8(10): e78568.