

三叶崖爬藤叶绿体基因组的组装与序列分析

蒋 明¹, 王军峰², 应梦豪¹, 杨如棉¹, 马佳莹¹

1. 台州学院生命科学学院, 浙江 椒江 318000

2. 丽水市林业科学研究院, 浙江 丽水 323000

摘 要: **目的** 以药用植物三叶崖爬藤为材料, 在利用高通量技术测定 DNA 序列的基础上, 对叶绿体基因组进行组装和序列分析, 为进一步开展群体遗传学和多样性研究奠定基础。**方法** 利用 HiSeq X Ten 测定三叶崖爬藤的 DNA 序列, 用 NOVOPlasty 组装叶绿体基因组, 在基因注释的基础上开展序列分析。**结果** 三叶崖爬藤叶绿体基因组的全长为 160 189 bp, GC 值 37.5%, 具 1 个典型的四分区域结构, 包括 1 个大单拷贝区、1 对反向重复区和 1 个小单拷贝区, 序列长度分别为 88 184、26 519、18 967 bp。三叶崖爬藤的叶绿体基因组共有 133 个基因, 其中编码蛋白基因、rRNA 基因与 tRNA 基因的数量分别为 88、8、37 个。位于反向重复区和小单拷贝区的 ycf1 基因 3'端发生缺失, 是 1 个假基因。**结论** 三叶崖爬藤的叶绿体基因组的组装和序列分析为后续开展群体遗传学和遗传多样性研究奠定了基础。

关键词: 三叶崖爬藤; 叶绿体基因组; 组装; DNA 序列; 序列分析

中图分类号: R282.12 **文献标志码:** A **文章编号:** 0253-2670(2020)02-0461-08

DOI: 10.7501/j.issn.0253-2670.2020.02.024

Assembly and sequence analysis of *Tetrastigma hemsleyanum* chloroplast genome

JIANG Ming¹, WANG Jun-feng², YING Meng-hao¹, YANG Ru-mian¹, MA Jia-ying¹

1. College of Life Science, Taizhou University, Jiaojiang 318000, China

2. Lishui Institute of Forestry, Lishui 323000, China

Abstract: **Objective** The complete chloroplast genome of medicinal plant *Tetrastigma hemsleyanum* sequenced by high-throughput technologies was assembled for the sequence analysis to provide evidence for its population genetics and diversity studies. **Methods** HiSeq X Ten was used to sequence DNA of *T. hemsleyanum*, and the chloroplast genome was assembled by NOVOPlasty. Sequence analysis was performed based on gene annotation results. **Results** The complete chloroplast genome of *T. hemsleyanum* was 160 189 bp in length with a GC content of 37.5%. The chloroplast genome exhibited a typical quadripartite structure, including a large single copy region, a pair of inverted repeats, and a small single copy, and the sequence lengths were 88 184 bp, 26 519 bp, and 18 967 bp, respectively. The chloroplast genome harbored 133 genes, including 88 protein-coding genes, eight rRNA genes, and 37 tRNA genes. The ycf1 gene located at the border of IR/SSC was proved to be a pseudogene, with its 3' end truncated. **Conclusion** Sequence assembly and analysis of *T. hemsleyanum* chloroplast genome provide new insights into future studies on both population genetics and genetic diversity.

Key words: *Tetrastigma hemsleyanum* Diels et Gilg; chloroplast genome; assembly; DNA sequence; sequence analysis

崖爬藤属 *Tetrastigma* (Miq.) Planch 隶属于葡萄科 (Vitaceae), 是一类木质或草质藤本植物, 全世界约 100 多种, 我国有 45 种^[1]。其中的三叶崖爬藤 *Tetrastigma hemsleyanum* Diels et Gilg 为草质藤本植物, 又名三叶青、金线吊葫芦和丝线吊金钟等, 分布于浙江、江苏、安徽、福建和湖南等省, 生于山

坡灌丛、溪旁或林下^[2]。三叶崖爬藤四季常绿、花型奇特、果色诱人, 是一种具有一定开发前景的野生观赏植物。三叶崖爬藤具有较强的耐阴性, 可用作阴湿环境的绿化配置植物, 也可用于室内盆栽^[3]。三叶崖爬藤的块根或全草均可入药, 具有清热解毒、活血止痛和祛风化痰等功效, 用于治疗毒蛇

收稿日期: 2019-09-03

基金项目: 台州市 211 人才工程经费资助

作者简介: 蒋 明 (1973—), 男, 浙江嵊州人, 博士, 教授, 硕士生导师, 研究方向为植物逆境生物学及其分子调控。

E-mail: jiangming1973@139.com

咬伤、百日咳、支气管炎、肺炎、咽喉炎、肝炎、小儿高热惊厥和肿瘤等^[4-6]。三叶崖爬藤富含黄酮类、还原糖和甾体类等成分,近年来,科研人员已从其全草或块根中鉴定出多个具有药用效果的化合物^[7]。目前,三叶崖爬藤的相关研究主要集中在栽培、化学成分鉴定和药理学研究等方面,有关该植物叶绿体基因组组装、基因组成和序列分析等未见报道。

叶绿体是绿色植物光合作用的重要场所,它利用太阳能和二氧化碳转换成化学能与氧气,在维系地球生命系统中起着重要作用^[8]。叶绿体具有一整套用于光合作用、能量代谢、蛋白质合成及氮、硫同化相关的基因,分布在大小为 120~180 kb 的环状基因组上^[9-11]。陆生植物的叶绿体基因组结构通常十分保守,由 1 个大单拷贝区域 (large single copy, LSC)、1 个短单拷贝区域 (small single copy, SSC) 和 2 个反向重复区域 (inverted repeat, IR) 组成^[12-14]。叶绿体基因组在植物分类、进化研究、物种鉴定和多样性研究等方面有着广泛的用途,随着测序技术的不断改进、组装软件的推陈出新及测序成本的降低,完成测序的叶绿体基因组数量逐年增长^[15-18]。近年来,药用植物叶绿体基因组测序方面取得了很大的进展,多种植物的叶绿体基因组已完成测序和分析,如曼陀罗 *Datura stramonium* L.、长柄石竹 *Dianthus superbus* L.、蛇足石杉 *Huperzia serrata* Thunb. ex Murray、连翘 *Forsythia suspensa* Thunb. 和铁皮石斛 *Dendrobium officinale* Kimura et Migo 等^[19-23]。本研究以三叶崖爬藤为材料,利用高通量方法测定 DNA 序列,并对该植物的叶绿体基因组进行组装和分析,为后续开展群体遗传学和遗传多样性研究奠定基础。

1 材料与仪器

1.1 材料

三叶崖爬藤叶片采自浙江舟山朱家尖,植株生长在海边林下,伴生植物有野桐 *Mallotus tenuifolius* Pax.、椿叶花椒 *Zanthoxylum ailanthoides* Sieb. et Zucc.、桉木 *Eurya japonica* Thunb. 和五节芒 *Miscanthus floridulus* Warb. ex Schum. et Laut. 等。叶片装入取样袋后带回实验室,用无菌水冲洗数次,晾干后置于 -80 °C 冰箱备用。

1.2 仪器

SANYO MDF-382E 型超低温冰箱;艾本德 Eppendorf 移液枪;Illumina HiSeq X Ten 测序仪;

ThinkPad P52 移动工作站(配置 64 G 内存);Covaris S2 超声波破碎仪;NanoDrop 微量紫外-可见分光光度计;伯乐 C1000 型 PCR 仪 (Bio-Rad 公司,美国);北京六一 DYY-12 型电泳仪和电泳槽 (北京市六一仪器厂);伯乐 Gel Doc XR⁺ 凝胶成像系统 (Bio-Rad 公司,美国);超净工作台 (苏州安泰空气技术有限公司)。

2 方法

2.1 DNA 的提取和文库构建

三叶崖爬藤叶片基因组 DNA 的提取采用 CTAB 法, DNA 的完整性用电泳检测,纯度用 NanoDrop 微量紫外-可见分光光度计测定。用 Covaris 超声波破碎仪将基因组 DNA 片段化,经末端修复、加 A 尾、加测序接头、纯化及 PCR 扩增等过程,最后形成测序文库。

2.2 高通量测序

文库质量检测合格后,在 Illumina HiSeq X Ten 上进行高通量测序,测序采用双端 (pair ends, PE) 策略,读长为 PE150,共获得 2.18 G 原始数据。利用 NGS QC Toolkit v2.3.3 对 reads 进行质量控制,运行自带的 perl 程序去除接头和低质量的序列,参数采用默认值。共得到 7 268 045 条 clean reads, Q20 达 96.75%, Q30 为 91.26%,可用于后续的拼接和注释。

2.3 叶绿体基因组的拼接和注释

叶绿体基因组的拼接采用 NOVOPlasty 程序^[24],插入片段大小设为 300 bp,基因组大小为 120 000~200 000 bp,最大内存设为 40 G,拼接在 ThinkPad P52 移动工作站上进行。利用 Dual Organellar GenoMe Annotator (DOGMA, <http://dogma.ccbb.utexas.edu/>) 对序列进行注释,参数为默认值,最后进行手动调整^[25]。tRNA 用 tRNAscan-SE (<http://lowelab.ucsc.edu/tRNAscan-SE/>)^[26]和 ARAGO RN^[27]预测。注释完成后,用 Geneious 11.1.5 生成 GenBank 格式的文件,再利用在线工具 OrganellarGenomeDraw (OGDRAW, <http://ogdraw.mpimp-golm.mpg.de>) 绘制叶绿体结构图^[28]。

2.4 边界的 PCR 验证和测序

序列的边界用 PCR 方法进行鉴定,共设计 4 对引物,分别是 SYQUP1/SYQDN1: 5'-AACTGCTCAGCAACAGTCGG-3'/5'-AAGGGGGTGGTATTGCTCCT-3'; SYQUP2/SYQDN2: 5'-CGCAATGGACGATTCCATCGT-3'/5'-GGCAGAATACCGT-

CACCCATTC-3'; SYQUP3/SYQDN3: 5'-AGGCA-GAATACCGTCACCCAT-3'/5'-ATGTAGGGGGTG-GACGCAT-3'; SYQUP4/SYQDN4: 5'-CATGGTAGT-ACCCTCGTCGACTG-3'/5'-CTTCGCCGCCGTAGT-AAGTAGT-3'。

在 20 μ L PCR 反应体系中, 依次加入 15 μ L ddH₂O、2.0 μ L 10 $\times$ Buffer、0.4 μ L dNTPs、上/下游引物各 0.3 μ L、1.6 μ L DNA 模板 (20 ng/ μ L) 和 0.4 μ L Taq DNA 聚合酶。94 $^{\circ}$ C 预变性 5 min; 94 $^{\circ}$ C 变性 30 s, 56.5 $^{\circ}$ C 复性 45 s, 72 $^{\circ}$ C 延伸 60 s, 共 32 个循环。将 PCR 产物经电泳检测后, 与 p-GEM T-easy 载体 (Promega) 连接, 置于室温 2 h 以完成反应; 在超净工作台上, 将连接产物转化进入大肠杆菌 DH5 α 感受态细胞 (北京全式金生物技术有限公司), 经挑取单菌落、液体培养和菌液 PCR 验证后, 各取 3 份阳性菌液测序。

2.5 系统发育分析

从 NCBI 数据库中下载 12 种植物的叶绿体基因组, 分别为铁青树科 (Olacaceae) 的赤苍藤 *Erythralum scandens* Bl. (登录号 MG661263)、葡萄 *Vitis vinifera* L. (登录号 DQ424856)、夏葡萄 *V. aestivalis* Michx. (登录号 NC_029454)、东北蛇葡萄 *Ampelopsis glandulosa* Mamix var. *brevipedunculata* Trautv. (登录号 KT831767)、槭叶葡萄 *V. acerifolia* Raf. (登录号 NC_035878)、美国圆叶葡萄 *V. coignetiae* Pulliat (登录号 LC333363)、霜葡萄 *V. cordifolia* Lam. (登录号 LC337667)、桑叶葡萄 *V. ficifolia* Bge. (登录号 LC339502)、沙地葡萄 *V. rupestris* Scheele (登录号 LC348382)、甜冬葡萄 *V. cinerea* Engelm. ex Millard (登录号 NC_036336)、葛藟 *V. flexuosa* Thunb. (登录号 LC341205) 和甜山葡萄 *V. monticola* Buckl. (登录号 LC342076) 等。

采用 Find repeats 工具获得 12 种植物叶绿

体基因组的 IR 序列、LSC 和 SSC; 用 Geneious 11.1.5 计算叶绿体基因组的 GC 值。MAFFT 7.388 程序用于叶绿体基因组的多重比对, 比对结果导入 jModelTest 工具用于筛选核苷酸替代矩阵。以赤苍藤叶绿体基因组为外类群, 利用 PhyML 3.1 软件生成最大似然树 (maximum likelihood, ML) [29-30]。

3 结果与分析

3.1 叶绿体基因组的结构

利用 Illumina HiSeq X Ten 测序, 去除接头和低质量的数据后, 共得到 14 536 090 条长度约为 150 bp 的序列。利用 NOVOPlasty 进行组装, 经注释后得到三叶崖爬藤的叶绿体基因组。结果表明, 三叶崖爬藤的叶绿体基因组的测序深度达 537 $\times$, 基因组大小为 160 189 bp, GC 值 37.5% (表 1), 具一个典型的 4 分区域结构, 包括 1 个 LSC、1 对 IR 和 1 个 SSC, 它们的长度分别为 88 185、26 519、18 966 bp (图 1)。IR、LSC 和 SSC 区域的 GC 值存在一定的差异, IR 区域的 GC 值最高, 为 42.9%, LSC 次之, 为 35.5%, SSC 区域的 GC 值最小, 仅 31.7% (表 1)。利用 PCR 克隆了 4 个边界的序列, 经测序后与组装结果进行比对, 发现序列完全一致。

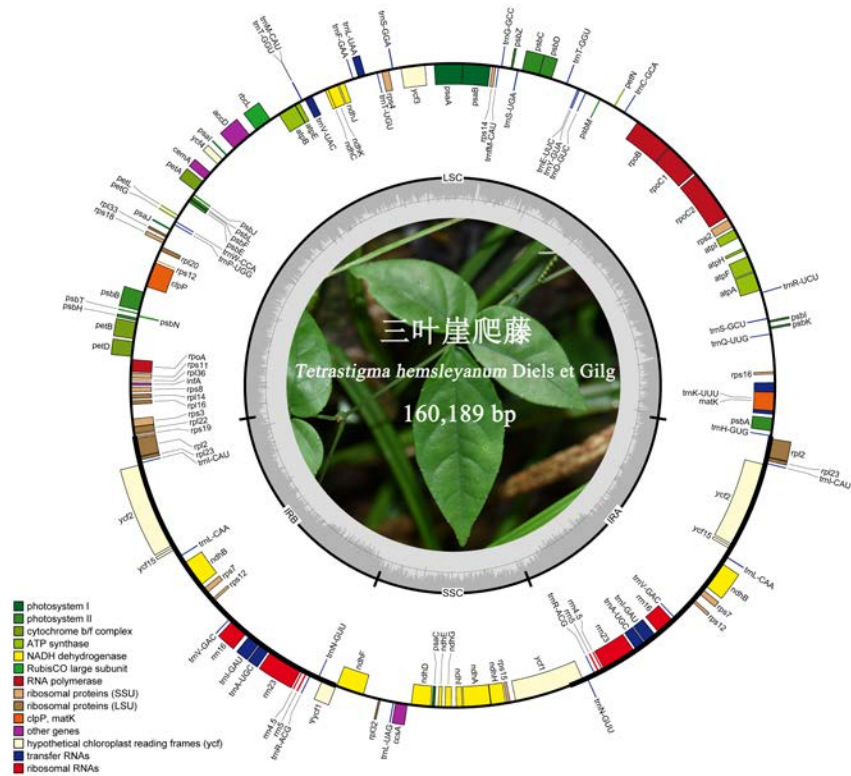
3.2 叶绿体基因的组成和特点

三叶崖爬藤叶绿体基因组上共有 133 个基因, 其中编码蛋白基因、rRNA 基因与 tRNA 基因的数量分别为 88、8、37 个 (表 2)。在 tRNA 中, trnA-UGC、trnI-CAU、trnI-GAU、trnL-CAA、trnN-GUU、trnR-ACG、trnT-GGU 和 trnV-GAC 各有 2 个拷贝; 4 个核糖体 RNA 均有 2 个拷贝, 分别位于反向重复区 IRA 和 IRB。核糖体蛋白大小亚基编码基因中, rps7、rps12、rpl2 和 rpl23 这 4 个基因均有 2 个拷贝, 其余为 1 个拷贝。NADH 脱氢酶亚基基因的 ndhB 及未知功能蛋白基因 ycf2 和 ycf15 的拷贝数均为 2。

表 1 三叶崖爬藤叶绿体基因组碱基组成

Table 1 Base composition of chloroplast genome in *T. hemsleyanum*

区域	A/%	T/%	C/%	G/%	长度/bp	GC/%
LSC	31.6	32.9	18.2	17.3	88 185	35.5
SSC	34.0	34.2	16.7	15.1	18 966	31.7
IRA/IRB	28.6	28.5	20.7	22.2	26 519	42.9
总体	30.9	31.6	19.1	18.4	160 189	37.5



LSC 和 SSC: 大单拷贝区域、小单拷贝区域; IRA 和 IRB: 2 个反向重复区域; 内圈深色部分: GC 含量

LSC and SSC: Large single copy and small single copy; IRA and IRB: two inverted repeats; The darker gray area in inner circle: GC content

图 1 三叶崖爬藤叶绿体基因组

Fig. 1 Map of *T. hemsleyanum* chloroplast genome

大部分叶绿体基因没有检测到内含子, 仅个别基因具 1~2 个含子。其中, *trnA-UGC*、*trnI-GAU*、*trnK-UUU*、*trnL-UAA*、*trnV-UAC*、*rpl2*、*rpoC1*、*ndhA*、*ndhB*、*petB*、*petD* 和 *atpF* 各有 1 个内含子, 而 *ycf3* 和 *clpP* 具 2 个内含子。部分基因之间有序列共用现象, 如 *matK* 基因位于 *trnK-UUU* 基因内, 整个编码区为 *trnK-UUU* 内含子的一部分; *psbD* 基因的 3'端与 *psbC* 基因的 5'端、*ndhC* 基因的 3'端与 *ndhK* 基因的 5'端、*trnM-CAU* 的 3'端与 *trnT-GGU* 的 5'端及 *atpB* 的 3'端与 *atpE* 的 5'端重叠。

3.3 IR 边界的特征

反向重复区域经常出现假基因化, 在三叶崖爬藤叶绿体基因组中, *ycf1* 基因有两份拷贝, 分别位于 *IRB/SSC* 和 *SSC/IRA* 边界(图 2)。位于 *SSC/IRA* 边界的 *ycf1* 基因长度为 5 634 bp, 为正常基因; 但位于 *IRB/SSC* 边界的 *ycf1* 基因全长仅为 1 200 bp, 它的 3'端发生缺失, 在边界右侧+54~+56 位出现一个终止密码子 TAA, 该 *ycf1* 为一个假基因。

SSC/IRB 边界的 *rps19* 基因全长为 279 bp, 虽然被 *LSC/IRB* 分隔成 2 部分, 但基因序列没有发生缺失, 为正常基因; *IRA/LSC* 边界没有检测到基因的存在。

3.4 基因组特征比较分析

三叶崖爬藤及 12 种近缘植物叶绿体基因组的序列长度为 156 154~161 090 bp, 赤苍藤叶绿体基因组的长度最短, 其次为三叶崖爬藤, 东北蛇葡萄的叶绿体基因组最长(表 3)。除三叶崖爬藤和赤苍藤外, 其余植物的叶绿体基因组 GC 值完全一致, 均为 37.4%, 赤苍藤的 GC 值最大, 为 38.0%。利用 Find repeats 程序得到 *IR*、*SSC* 和 *LSC*, 结果表明, 三叶崖爬藤的 *IR* 最长, 达 26 519 bp, 其次为东北蛇葡萄, 为 26 394 bp, 夏葡萄的 *IR* 最短, 为 26 292 bp。13 种植物 *SSC* 的长度为 18 567~19 086 bp, 铁青树科赤苍藤的 *SSC* 最短, 其次为三叶崖爬藤, 其他 11 种植物 *SSC* 的长度差异仅 57 bp。甜山葡萄 *LSC* 的长度最长, 为 89 272 bp, 赤苍藤的最短, 仅 84 779 bp。

表 2 三叶崖爬藤叶绿体基因组编码的基因

Table 2 Genes present in chloroplast genome of *T. hemsleyanum*

基因类别	基因名称	数量
转运 RNA	trnA-UGC (×2)*、trnC-GCA、trnD-GUC、trnE-UUC、trnF-GAA、trnM-CAU、trnG-GCC、trnH-GUG、trnI-CAU (×2)、trnI-GAU (×2)*、trnK-UUU*、trnL-CAA (×2)、trnL-UAA*、trnL-UAG、trnM-CAU、trnN-GUU (×2)、trnP-UGG、trnQ-UUG、trnR-ACG (×2)、trnR-UCU、trnS-GCU、trnS-GGA、trnS-UGA、trnT-GGU (×2)、trnT-UGU、trnV-GAC (×2)、trnV-UAC*、trnW-CCA、trnY-GUA	37
核糖体 RNA	rrn4.5 (×2)、rrn5 (×2)、rrn16 (×2)、rrn23 (×2)	8
核糖体蛋白小亚基	rps2、rps3、rps4、rps7 (×2)、rps8、rps11、rps12 (×2)、rps14、rps15、rps16、rps18、rps19	14
核糖体蛋白大亚基	rpl2 (×2)*、rpl14、rpl16、rpl20、rpl22、rpl23 (×2)、rpl32、rpl33、rpl36	11
RNA 聚合酶	rpoA、rpoB、rpoC1*、rpoC2	4
NADH 脱氢酶亚基	ndhA*、ndhB (×2)*、ndhC、ndhD、ndhE、ndhF、ndhG、ndhH、ndhI、ndhJ、ndhK	12
光系统 I 亚基	psaA、psaB、psaC、psaI、psaJ、ycf3**	6
光系统 II 亚基	psbA、psbB、psbC、psbD、psbE、psbF、psbH、psbI、psbJ、psbK、psbL、psbM、psbN、psbT、psbZ	15
细胞色素 b/f 复合物亚基	petA、petB*、petD*、petG、petL、petN	6
ATP 合成酶亚基	atpA、atpB、atpE、atpF*、atpH、atpI	6
Rubisco 大亚基	rbcL	1
成熟酶	matK	1
蛋白酶	clpP**	1
被膜蛋白	cemA	1
乙酰 CoA 羧化酶	accD	1
细胞色素 C 合成酶	ccsA	1
翻译起始因子	infA	1
未知功能蛋白	ycf1、Ψycf1、ycf2 (×2)、ycf4、ycf15 (×2)	7
总计		133

×2-2 个拷贝 Ψ-假基因 *一个内含子 **2 个内含子
×2-two copies Ψ-pseudogene *one intron **two introns

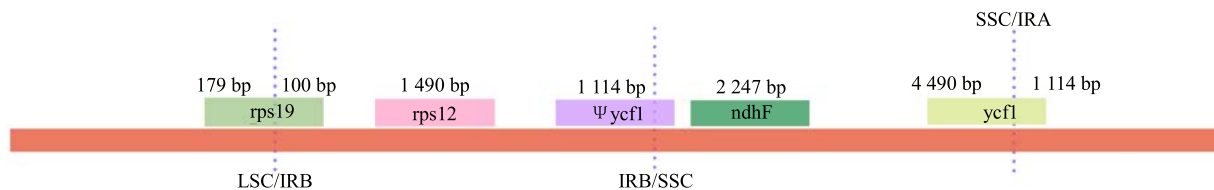


图 2 三叶崖爬藤叶绿体基因组的边界

Fig. 2 Border regions of *T. hemsleyanum* chloroplast genome

表 3 13 种植物叶绿体基因组的特征

Table 3 Chloroplast genome characteristics of 13 plants

植物	叶绿体基因组长度/bp	GC 值/%	LSC 长度/bp	SSC 长度/bp	IR 长度/bp
三叶崖爬藤	160 189	37.5	88 185	18 966	26 519
赤苍藤	156 154	38.0	84 799	18 567	26 394
葡萄	160 928	37.4	89 140	19 072	26 353
夏葡萄	160 913	37.4	89 171	19 086	26 292
东北蛇葡萄	161 090	37.4	89 266	19 035	26 394
槭叶葡萄	161 012	37.4	89 214	19 040	26 374
美国圆叶葡萄	161 012	37.4	89 229	19 065	26 354
霜葡萄	161 018	37.4	89 209	19 081	26 359
桑叶葡萄	160 983	37.4	89 197	19 068	26 354
沙地葡萄	161 008	37.4	89 204	19 046	26 374
甜冬葡萄	161 020	37.4	89 209	19 055	26 373
葛藟	160 927	37.4	89 156	19 059	26 351
甜山葡萄	160 996	37.4	89 272	19 029	26 381

3.5 系统发育分析

叶绿体基因组序列用 MAFFT 7.388 程序进行多重比对后,利用 jModelTest 的赤池信息量(akaike information criterion, AIC)进行最佳模型和参数的筛选。结果表明,用于构建 ML 树的最佳模型为 GTR+I+G,模型似然值自然对数的负值($-\ln L$)为 311 493.285 5。利用 PHYML 构建三叶崖爬藤、赤苍藤及 11 种葡萄科植物的系统发生树,结果显示,13 种植物可分为 4 组,蛇葡萄属的东北蛇葡萄、崖爬藤属的三叶崖爬藤和赤苍藤属的赤苍藤分别处于不同的分支,葡萄属的 10 种植物聚为一组(图 3)。葡萄属植物在系统发生树上可分为 2 个亚组,葡萄、美国圆叶葡萄、桑叶葡萄和葛藟处于同一亚组,其余 6 种葡萄属植物聚于另一亚组。

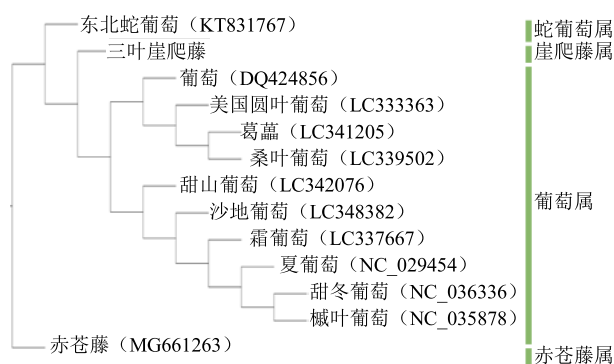


图 3 基于叶绿体全基因组构建的最大似然树状图

Fig. 3 Maximum likelihood tree constructed by complete chloroplast genomes

4 讨论

叶绿体是绿色植物最重要的细胞器之一,是光合作用的场所。自首次从地钱 *Marchantia polymorpha* L. 和烟草 *Nicotiana tabacum* L. 中测定完整的叶绿体基因组以来,完成叶绿体基因组测序的植物已超过 2 500 种^[31]。陆生植物叶绿体的基因组大小通常十分保守,长度为 120~160 kb^[32],如药用植物平贝母 *Fritillaria ussuriensis* Maxim. 和川贝母 *Fritillaria cirrhosa* D. Don 叶绿体基因组大小为 151 524 bp 和 151 083 bp^[17],油桐 *Vernicia fordii* Hemsl. 的叶绿体基因组大小为 161 528 bp^[18],而兰科植物羊耳蒜 *Liparis japonica* Maxim. 的叶绿体基因组长度为 153 687 bp^[33]。银杉 *Cathaya argyrophylla* Chun et Kuang 的叶绿体基因组最小,仅 107 kb,而天竺葵属 *Pelargonium* L'Her. 植物达 218 kb,是目前已知的最大叶绿体基因组^[34-35]。在

本研究中,三叶崖爬藤叶绿体基因组大小为 160 189 bp,介于羊耳蒜和油桐之间。陆生植物中,叶绿体基因的数量也十分保守,通常为 120~140 个,大部分为蛋白质编码基因和 tRNA 基因^[36]。油桐叶绿体基因组中有 114 个基因,其中 81 个为蛋白质编码基因,29 个为 tRNA^[18];地钱有 128 个叶绿体基因,其中的 51 个为蛋白质编码基因,32 个为 tRNA 基因^[31]。在三叶崖爬藤叶绿体基因组中,共有 133 个基因,其中编码蛋白基因与 tRNA 的数量分别为 88 和 37 个。

叶绿体基因组通常为 4 分区域结构,即由 1 个 LSC、1 个 SSC 和 2 个 IR 组成。叶绿体基因组的 IR 比较保守,在陆生植物中,IR 的长度约 15 kb 到 30 kb,4 个 rRNA 基因分布于该区域^[37]。油桐的 IR 大小为 26 819 bp,羊耳蒜叶绿体基因组的 IR 大小为 25 709 bp,地钱的相对较小,仅 10 058 bp^[18,33]。裸子植物的 IR 比被子植物小得多,甚至发生丢失现象,银杉的 IR 长度仅 267 bp,而南方红豆杉 *Taxus chinensis* Mon. var. *mairei* Cheng et L. K. Fu 的 IR 完全丢失^[34,38]。本研究中,三叶崖爬藤的 IR 大小为 26 519 bp,较油桐叶绿体基因组的 IR 小。假基因指序列与正常基因相似,但已丧失功能的 DNA 片段,是功能基因的缺陷拷贝,假基因在核基因组中十分常见^[39]。在叶绿体基因组中,假基因化(pseudogenization)的发生也十分频繁,在银杉叶绿体基因组中,位于 trnE-UUC 与 trnY-GUA 之间的 psbB 及位于 IRA 边界的 ycf2 为假基因^[34];在羊耳蒜叶绿体基因组中,ycf1 和 rps19 各有 2 份拷贝,其中 1 份为假基因^[33];而油桐叶绿体基因组中,ycf15 和 ycf68 为假基因^[18]。本研究中,三叶崖爬藤的 ycf1 有 2 份拷贝,其中位于 IRB/SSC 边界的 ycf1 基因 3'端发生缺失,为假基因。

叶绿体基因或基因间隔区常用于系统发育和物种鉴定分析,如 matK、rbcL、accD、ycf1、psbA-trnH 和 trnS-trnG 等^[40-41]。与叶绿体基因或基因间隔区相比,叶绿体基因组包含的信息量较大,而且它的结构、大小和基因组成相对恒定,已广泛应用于系统发育分析、物种鉴定等^[42]。Carbonell-Caballero 等^[43]利用叶绿体基因组解决了柑橘属 *Citrus* L. 野生种和栽培种之间的关系及驯化历程;de Abreu 等^[44]利用叶绿体基因组明确了多穗兰属 *Polystachya* Hook. 的系统发育地位;Yang 等^[45]测定了 5 种兰属 *Cymbidium* Sw. 植物的叶绿体基因组,为该属植物

的分子鉴定奠定了基础; Kang 等^[46]测定和分析了金银花 *Lonicera japonica* Thunb. 的叶绿体基因组, 为该药用植物的鉴定提供了分子依据。在本研究中, 三叶崖爬藤单独聚为一组, 与葡萄属较为接近, 结果与形态学鉴定一致。三叶崖爬藤叶绿体基因组的组装和序列分析, 为后续开展群体遗传学和遗传多样性研究奠定了基础。

参考文献

- [1] 中国科学院中国植物志编辑委员会. 中国植物志 (第48卷) [M]. 北京: 科学出版社, 1998.
- [2] 浙江植物志编委会. 浙江植物志 (第4卷) [M]. 杭州: 浙江科学技术出版社, 1993.
- [3] 陈 斌. 三叶青药用观赏两不误 [J]. 中国花卉园艺, 2015(2): 51-52.
- [4] 李 松, 郭淑珍, 王翠霞. 活血止痛的三叶青 [J]. 首都食品与医药, 2016, 23(6): 64.
- [5] 陈丽芸, 郭素华. 三叶青的化学成分及药理作用研究进展 [J]. 浙江中医药大学学报, 2012, 36(12): 1368-1370.
- [6] 刘培刚, 魏克民. 三叶青生物学、药理学及临床应用相关研究进展 [J]. 中国中医药科技, 2018, 25(6): 927-932.
- [7] 何丽针, 李中杰. 中药三叶青有效化学成分及在临床医学上的意义 [J]. 现代医药卫生, 2013, 29(1): 85-86.
- [8] Daniell H, Lin C S, Yu M, *et al.* Chloroplast genomes: Diversity, evolution, and applications in genetic engineering [J]. *Genome Biol*, 2016, 17(1): 134.
- [9] Bobik K, Burch-Smith T M. Chloroplast signaling within, between and beyond cells [J]. *Front Plant Sci*, 2015, 6: 781.
- [10] Smith D R. Evolution: In chloroplast genomes, anything goes [J]. *Current Biol*, 2017, 27(24): R1305-R1307.
- [11] Zoschke R, Bock R. Chloroplast translation: Structural and functional organization, operational control, and regulation [J]. *Plant Cell*, 2018, 30(4): 745-770.
- [12] Yu X, Zuo L, Lu D, *et al.* Comparative analysis of chloroplast genomes of five *Robinia* species: Genome comparative and evolution analysis [J]. *Gene*, 2018, 689(2019): 141-151.
- [13] Zhao Z, Wang X, Yu Y, *et al.* Complete chloroplast genome sequences of *Dioscorea*: Characterization, genomic resources, and phylogenetic analyses [J]. *Peer J*, 2018, 6: e6032.
- [14] Zhou Y, Nie J, Xiao L, *et al.* Comparative chloroplast genome analysis of rhubarb botanical origins and the development of specific identification markers [J]. *Molecules*, 2018, 23(11): 2811-2820.
- [15] Jin J J, Yu W B, Yang J B, *et al.* GetOrganelle: A simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data [J]. *BioRxiv*, 2018, 12: 256479.
- [16] Verbruggen H, Marcelino V R, Guiry M D, *et al.* Phylogenetic position of the coral symbiont *Ostreobium* (Ulvophyceae) inferred from chloroplast genome data [J]. *J Phycol*, 2017, 53(4): 790-803.
- [17] Park I, Kim W J, Yeo S M, *et al.* The complete chloroplast genome sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don, and comparative analysis with other *Fritillaria* species [J]. *Molecules*, 2017, 22(6): 982.
- [18] Li Z, Long H, Zhang L, *et al.* The complete chloroplast genome sequence of tung tree (*Vernicia fordii*): Organization and phylogenetic relationships with other angiosperms [J]. *Sci Rep*, 2017, 7(1): 1869.
- [19] Yang Y, Dang Y, Li Q, *et al.* Complete chloroplast genome sequence of poisonous and medicinal plant *Datura stramonium*: Organizations and implications for genetic engineering [J]. *PLoS One*, 2014, 9(11): e110656.
- [20] Raman G, Park S. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalycinus*, from a comparative genomics perspective [J]. *PLoS One*, 2015, 10(10): e0141329.
- [21] Guo Z Y, Zhang H R, Shrestha N, *et al.* Complete chloroplast genome of a valuable medicinal plant, *Huperzia serrata* (Lycopodiaceae), and comparison with its congener [J]. *Appl Plant Sci*, 2016, 4(11): 4342-4343.
- [22] Wang W, Yu H, Wang J, *et al.* The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae) [J]. *Inter J Mol Sci*, 2017, 18(11): 2288-2295.
- [23] Konhar R, Biswal D, Debnath M, *et al.* Chloroplast genome sequence annotation of *Dendrobium nobile* (Asparagales: Orchidaceae), an endangered medicinal orchid from Northeast India [J]. *Genome Announc*, 2016, 4(5): e01088-16.
- [24] 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.
- [25] Wyman S K, Jansen R K, Boore J L. Automatic annotation of organellar genomes with DOGMA [J]. *Bioinformatics*, 2004, 20(17): 3252-3255.
- [26] Lowe T M, Eddy S R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic

- sequence [J]. *Nucleic Acids Res*, 1997, 25(5): 955-964.
- [27] Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences [J]. *Nucleic Acids Res*, 2004, 32(1): 11-16.
- [28] Lohse M, Drechsel O, Kahlau S, *et al.* OrganellarGenomeDRAW—A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets [J]. *Nucleic Acids Res*, 2013, 41: W575-W581.
- [29] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability [J]. *Mol Biol Evol*, 2013, 30(4): 772-780.
- [30] Guindon S, Dufayard J F, Lefort V, *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3. 0 [J]. *Syst Biol*, 2010, 59(3): 307-321.
- [31] Ohyama K, Fukuzawa H, Kohchi T, *et al.* Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA [J]. *Nature*, 1986, 322(6079): 572-574.
- [32] Du Y P, Bi Y, Yang F P, *et al.* Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses [J]. *Sci Rep*, 2017, 7(1): 5751.
- [33] Krawczyk K, Wiland-Szymanska J, Buczkowska-Chmielewska K, *et al.* The complete chloroplast genome of a rare orchid species *Liparis loeselii* (L.) [J]. *Conserv Genet Res*, 2018, 10(3): 305-308.
- [34] Lin C P, Huang J P, Wu C S, *et al.* Comparative chloroplast genomics reveals the evolution of *Pinaceae* genera and subfamilies [J]. *Genome Biol Evol*, 2010, 2: 504-517.
- [35] Chumley T W, Palmer 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.
- [36] Zeng X, Liu M M, Li Y Y, *et al.* The complete chloroplast genome of *Anoetochilus roxburghii* [J]. *Mitochondrial DNA A*, 2016, 27(6): 4264-4265.
- [37] Zhu A D, Guo W H, Gupta S, *et al.* Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates [J]. *New Phytol*, 2016, 209(4): 1747-1756.
- [38] Zhang Y Z, Ma J, Yang B X, *et al.* The complete chloroplast genome sequence of *Taxus chinensis* var. *mairei* (Taxaceae): loss of an inverted repeat region and comparative analysis with related species [J]. *Gene*, 2014, 540(2): 201-209.
- [39] 刘国庆, 白音宝力高, 邢永强. 假基因研究进展 [J]. 生物化学与生物物理进展, 2010, 37(11): 1165-1174.
- [40] 夏至, 黄勇, 李贺敏, 等. 基于核基因 ITS 及叶绿体 psbA-trnH 和 trnS-trnG 基因怀地黄栽培起源探讨 [J]. 中草药, 2018, 49(2): 423-430.
- [41] Levin R A. Phylogenetic relationships within Nyctaginaceae tribe Nyctagineae: Evidence from nuclear and chloroplast genomes [J]. *Syst Biol*, 2000, 25(4): 738-750.
- [42] 宋菊, 龙月红, 林丽梅, 等. 五加科植物叶绿体基因组结构与进化分析 [J]. 中草药, 2017, 48(24): 5070-5075.
- [43] Carbonell-Caballero J, Alonso R, Ibanez V, *et al.* A Phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus* [J]. *Mol Bioland Evol*, 2015, 32(8): 2015-2035.
- [44] de Abreu N L, Alves R J V, Cardoso S R S, *et al.* The use of chloroplast genome sequences to solve phylogenetic incongruences in *Polystachya* Hook (Orchidaceae Juss) [J]. *Peer J*, 2018, 6: e4916.
- [45] Yang J B, Tang M, Li H T, *et al.* Complete chloroplast genome of the genus *Cymbidium*: Lights into the species identification, phylogenetic implications and population genetic analyses [J]. *BMC Evol Biol*, 2013, 13: 84-90.
- [46] Kang S J, Park H S, Koo H J, *et al.* The complete chloroplast genome sequence of Korean *Lonicera japonica* and intra-species diversity [J]. *Mitochondrial DNA Part B*, 2018, 3(2): 941-942.