

婆婆纳叶绿体基因组测序与特征分析

贾守宁¹, 李翔², 鲍海娟², 李生新², 陈文娟¹, 赵国福¹, 王双玺¹, 徐智玮¹,
王久利^{2,3}, 李啟恩^{4,5*}

(1. 青海省中医院中药研究所, 中国青海 西宁 810099; 2. 青海民族大学 生态环境与资源学院, 中国青海 西宁 810007;
3. 青海省特色经济植物高值化利用重点实验室, 中国青海 西宁 810007; 4. 青海大学 藏医学院, 中国青海 西宁 810016;
5. 青海大学 藏药研究中心, 中国青海 西宁 810016)

摘要: 婆婆纳(*Veronica polita* Fr.)是一种重要的藏药材,本研究基于高通量测序方法完成了婆婆纳叶绿体基因组的测序,并运用生物信息学方法分析了其结构、基因构成等特征,随后利用叶绿体基因组的蛋白质编码序列探讨了婆婆纳的系统发育关系。结果表明:1) 婆婆纳的完整叶绿体基因组总长为 150 191 bp, GC 含量为 37.9%,具有典型的四分体环状结构,大单拷贝(large single copy, LSC)区、小单拷贝(small single copy, SSC)区和反向重复(inverted repeat, IR)区的长度分别为 81 847 bp、17 414 bp 和 25 456 bp; 2) 婆婆纳叶绿体基因组含有 134 个基因,包括 88 个蛋白质编码基因(protein coding gene, PCG)、8 个核糖体 RNA (ribosomal RNA, rRNA)基因、38 个转运 RNA (transfer RNA, tRNA)基因,其中假基因 2 个; 3) 在婆婆纳叶绿体基因组中共检测到 26 529 个密码子,其中 32 种密码子的相对同义密码子使用度(relative synonymous codon usage, RSCU)值小于 1, 30 种密码子的 RSCU 值大于 1, 2 种密码子的 RSCU 值等于 1; 4) 婆婆纳属植物的 IR 边界高度保守,无明显的收缩或扩张; 5) 婆婆纳属植物叶绿体基因组之间的变异区域多在基因间隔区; 6) 婆婆纳与阿拉伯婆婆纳(*V. persica*)的亲缘关系较近。本研究为婆婆纳属物种的鉴定和遗传多样性研究提供了基础资料。

关键词: 婆婆纳; 藏药; 叶绿体基因组; 测序; 特征

中图分类号: Q81, Q943.2

文献标志码: A

文章编号: 1007-7847(2023)05-0426-09

Sequencing and Characterization of Chloroplast Genome of *Veronica polita*

JIA Shouning¹, LI Xiang², BAO Haijuan², LI Shengxin², CHEN Wenjuan¹,
ZHAO Guofu¹, WANG Shuangxi¹, XU Zhiwei¹, WANG Jiuli^{2,3}, LI Qien^{4,5*}

(1. Institute of Chinese Medicine, Qinghai Provincial Hospital of Traditional Chinese Medicine, Xining 810099, Qinghai, China; 2. College of Ecological Environment and Resources, Qinghai Minzu University, Xining 810007, Qinghai, China; 3. Qinghai Provincial Key Laboratory of High-value Utilization of Characteristic Economic Plants, Xining 810007, Qinghai, China; 4. Tibetan Medical College, Qinghai University, Xining 810016, Qinghai, China; 5. Tibetan Medicine Research Center, Qinghai University, Xining 810016, Qinghai, China)

Abstract: *Veronica polita* Fr. is an important Tibetan medicinal herb. Herein, the chloroplast genome of *V. polita* was sequenced based on high-throughput sequencing, and its basic characteristics were analyzed using bioinformatics methods. The phylogeny of *V. polita* was studied using the protein coding sequences of chloroplast genomes. The results were as follows: 1) The complete chloroplast genome of *V. polita* is 150 191 bp in length, with a GC content of 37.9%. The lengths of large single copy (LSC), small single copy (SSC) and inverted repeat (IR) regions are 81 847 bp, 17 414 bp and 25 456 bp, respectively; 2) The chloroplast genome contains 134 genes, including 88 protein coding genes (PCGs), 8 ribosomal RNA (rRNA) genes and 38 trans-

收稿日期: 2022-07-04; 修回日期: 2022-10-01; 网络首发日期: 2022-12-23

基金项目: 青海省中藏医药科研创新项目(J2020005); 青海民族大学校级项目(2022-JYQN-002, 2021XJG17)

作者简介: 贾守宁(1966—), 男, 河南开封人, 主任药师, 主要从事中药资源研究; *通信作者: 李啟恩(1982—), 男, 青海贵德人, 博士, 教授, 主要从事藏族药资源、炮制和药理研究, E-mail: qienli@outlook.com。

port RNA (tRNA) genes, including 2 pseudogenes; 3) A total of 26 529 codons were detected in the chloroplast genome. Among them, 32 types of single codons have relative synonymous codon usage (RSCU) values less than 1, 30 types have RSCU values greater than 1, and 2 types have RSCU values equal to 1; 4) The IR boundary regions of *Veronica* are highly conservative, without obvious contraction or expansion; 5) Most variations in chloroplast genomes of *Veronica* plants exist in intergenic regions; 6) *V. polita* is most closely related to *V. persica*. This study laid the foundation for species identification and genetic diversity research of the genus *Veronica*.

Key words: *Veronica polita*; Tibetan medicine; chloroplast genome; sequencing; characterization

(*Life Science Research*, 2023, 27(5): 426–434)

叶绿体是绿色植物进行光合作用的细胞器,其功能的实现需要依赖其相对独立遗传的基因组,即叶绿体基因组(chloroplast genome)^[1]。植物的叶绿体基因组比核基因组或线粒体基因组拥有更高度保守的基因组结构^[2-3]。在大多数植物种类中,叶绿体基因组呈现典型的四分体环状结构,包括一对反向重复(inverted repeat, IR)区(IRA 和 IRB)、一个大单拷贝(large single copy, LSC)区 and 一个小单拷贝(small single copy, SSC)区^[4]。由于叶绿体基因组具有高丰度、单拷贝、缺乏重组和适当的核苷酸进化率等优势,常被用来构建系统发育树,这对于进一步探索多个分类群水平的亲缘关系具有重要意义^[5]。

婆婆纳(*Veronica polita* Fr.)是一年生草本植物,原产于亚洲西南部,现已在世界上大部分地区归化,多生于荒地^[6]。婆婆纳在藏药中称“冬那端赤”,以晒干或新鲜全草入药,具有清热、消炎、消肿、愈伤和止血的功效,甚至可以用来治疗黑色素瘤^[7-9]。在深受中国中医药文化影响的韩国,婆婆纳也被视作有广泛应用价值的药材,并被认为具有显著的抗炎和抗氧化功效^[10]。此外,婆婆纳在农业生产和土壤重金属污染修复等方面也具有巨大的应用潜力^[11-12]。目前,针对婆婆纳的研究多集中在药材特性和植物化学成分方面^[7-10, 13-16],而少有对其遗传信息的研究。本研究基于高通量测序技术对婆婆纳叶绿体基因组进行测序,并分析该基因组的特征,以期为其物种的鉴定和遗传多样性研究提供基础资料。

1 材料与方法

1.1 材料

婆婆纳植株采集于青海省西宁市海湖新区(101.73°E, 36.64°N),由赵国富主任药师鉴定,凭证标本(标本号: VDIDY20210930)保存在青海省中医院标本馆。

1.2 DNA 的提取、测序与组装

使用改良 CTAB (cetyltrimethylammonium bromide)法从新鲜的婆婆纳叶片中提取总 DNA,并用 1%的琼脂糖凝胶电泳检测所提 DNA^[17]。将合格 DNA 样品送往北京诺禾致源科技股份有限公司并采用 Illumina HiSeq X Ten 平台进行测序,测序读长为 PE150。使用 GetOrganelle 程序组装叶绿体基因组,参数设置取默认值^[18]。

1.3 基因组的注释与可视化

利用在线软件 GeSeq 对组装好的叶绿体基因组序列进行基因注释^[9],生成初步注释文件。随后,参考婆婆纳的同属物种阿拉伯婆婆纳(*V. persica*)的叶绿体基因组(GenBank 接收号: KT724052)注释文件进行手工校正。将最终注释好的叶绿体基因组信息提交至 NCBI 的 GenBank 数据库(GenBank 接收号: OL621900)。根据叶绿体基因组的完整序列和注释信息,利用 Chloroplot 软件绘制婆婆纳的环状叶绿体基因组物理图谱^[20]。

1.4 密码子使用偏好性分析

从婆婆纳叶绿体基因组上提取蛋白质编码序列并依序整合,使用 CodonW (version 1.4.2; <http://codonw.sourceforge.net/>; 2022-05-28)分析该基因组的总体密码子使用情况,获得相对同义密码子使用度(relative synonymous codon usage, RSCU)值^[21],并使用在线软件 RSCU-热图(<http://112.86.217.82:9919/#/tool/alltool/detail/301>; 2022-05-28)绘制 RSCU 值的热图。

1.5 婆婆纳属叶绿体基因组的边界分析

选取包括婆婆纳在内的已知叶绿体基因组完整序列的婆婆纳属(*Veronica*)植物,使用在线软件 IRscope 对 IR 区和单拷贝(single copy, SC)区的边界进行分析^[22],参数设置取软件的默认值。

1.6 婆婆纳属叶绿体基因组的序列变异分析

首先通过 Python 3.10.1 将婆婆纳属叶绿体基

因组的 GBK 格式注释文件转换成 mVISTA 软件可识别的注释文件, 然后利用 mVISTA 软件并选用检测基因重排和倒位的全局比对模式(Shuffle-LAGAN)^[23], 以婆婆纳作为参考基因组, 对婆婆纳属的叶绿体基因组进行序列比对, 分析其序列的变异情况。

1.7 系统发育分析

从 GenBank 数据库中下载婆婆纳的 27 个车前科和 3 个玄参科物种的叶绿体基因组序列, 与本研究获得的婆婆纳叶绿体基因组整合到一起, 进行系统发育树的构建。借助 Geneious R11 软件提取 31 个叶绿体基因组中的单个蛋白质编码基因编码序列并进行比对, 选取对比效果良好的 14 个基因(*atpA*、*atpB*、*matK*、*psaA*、*psaB*、*psbA*、*psbB*、*psbC*、*psbD*、*rbcL*、*rpoA*、*rpoB*、*rpoC1* 和 *rpoC2*), 随后依序连接成单个超级基因(supergene), 导出至 TOPALi v2.5 软件^[24], 借助其集成的 MrBayes v3.1.2 程序构建系统发育树^[25]。

2 结果

2.1 叶绿体基因组的基本特征

经测序, 在 Illumina HiSeq X Ten 平台上获得约 4 G 的原始数据(raw data), 共 17 884 744 条原始读数(read), 由这些数据组装获得总长为 150 191 bp 的婆婆纳叶绿体基因组完整序列, GC 含量为 37.9%。婆婆纳的叶绿体基因组为双链环状, 有典型的四分体结构(图 1), 其中一对 IR 区均为 25 456 bp, GC 含量为 43.2%; LSC 区和 SSC 区分别为 81 847 bp 和 17 414 bp, GC 含量分别为 35.9%和 31.6%(表 1)。

表 1 婆婆纳叶绿体基因组的碱基组成
Table 1 Base composition of the *V. polita* chloroplast genome

Region	Percentage of a base or base pair/(%)					
	A	C	G	T	AT	GC
Genome	30.7	19.2	18.7	31.4	62.1	37.9
LSC	31.4	18.3	17.6	32.7	64.1	35.9
SSC	34.4	16.4	15.2	34.0	68.4	31.6
IRA	28.2	22.4	20.8	28.6	56.8	43.2
IRB	28.6	20.8	22.4	28.2	56.8	43.2

婆婆纳叶绿体基因组含有 134 个基因, 包括 88 个蛋白质编码基因(protein coding gene, PCG)、8 个核糖体 RNA (ribosomal RNA, rRNA)基因、38 个转运 RNA (transfer RNA, tRNA)基因, 其中假基因 2 个(*ycf15*), 各基因的功能分类见表 2。该基因组的多个基因含有内含子, 其中含有 1 个内含子

的 PCG 包括 *atpF*、*ndhA*、*ndhB*、*petB*、*petD*、*rpl2*、*rpl16*、*rpoC1* 和 *rps16* (表 2); 含有 2 个内含子的 PCG 包括 *rps12*、*clpP* 和 *ycf3*; 含有 1 个内含子的 tRNA 基因包括 *trnA-UGC*、*trnG-UCC*、*trnI-GAU*、*trnK-UUU*、*trnL-UAA*、*trnL-UAG* 和 *trnV-UAC*; rRNA 基因均无内含子。

2.2 相对同义密码子使用度

基于 CodonW 软件开展的 RSCU 值的分析结果显示, 婆婆纳叶绿体基因组中的蛋白质编码序列共含 26 529 个密码子, 其中编码氨基酸的密码子数量为 226 443, 同义密码子数量为 25 355; 总 GC 含量为 38.0%, 以 GC 为第 3 个碱基的密码子的占比为 26.8%。从密码子使用的数量看, 编码亮氨酸(Leu)、丝氨酸(Ser)和精氨酸(Arg)的密码子种类最多, 均为 6 种; 使用频数最高的密码子是编码异亮氨酸(Ile)的 AUU (1 094 个); 使用最少的密码子是终止密码子(TER)的 UGA (14 个)。甲硫氨酸(Met)和色氨酸(Trp)两种氨基酸的密码子不存在使用偏好性(RSCU 值为 1), 其密码子分别为 AUG (同时也是起始密码子)和 UGG。在存在使用偏好性的密码子中, 32 种密码子的 RSCU 值小于 1, 30 种密码子的 RSCU 值大于 1 (表 3、图 2)。

2.3 IR 边界的扩张与收缩表现

婆婆纳属植物的叶绿体基因组为环形, 都有典型的四分体结构, IR 与 LSC 和 SSC 之间存在 4 个边界, 即 JLB (LSC-IRB)、JSB (IRB-SSC)、JSA (SSC-IRA)和 JLA (IRA-LSC)。分析结果显示, 婆婆纳属各物种的边界比较保守, 无明显的扩张或收缩, 婆婆纳属各物种的 4 个边界所处的基因位置高度趋同: JLB 位于 *rps19* 和 *rpl2* 基因之间; JSB 位于 *ndhF* 基因上; JSA 位于 *ycf1* 基因上; JLA 在 *rpl2* 和 *trnH* 基因之间, 并紧邻 *trnH* 基因(图 3)。

2.4 婆婆纳属叶绿体基因组序列比较

以婆婆纳为参考序列的婆婆纳属叶绿体基因组序列的比对分析表明, 该属叶绿体基因组的基因构成和排列顺序高度相似, 但序列层面存在较多变异。rRNA 和 tRNA 的序列比较保守, 特别是 rRNA 均为高度保守区域; 在编码基因区域, 长度较大的基因往往存在易观察到的变异, 出现较高变异的区域包括 *matK*、*atpA*、*rpoC2*、*accD*、*ndhK*、*ndhD*、*ycf1* 和 *ycf2* 等基因。阿拉伯婆婆纳(*V. persica*)与婆婆纳的序列差异较小; 而水苦苣(*V. undulata*)、*V. nakaiana* 和 *V. ovata* 三者与婆婆纳的序列差异较大, 特别是基因间隔区(图 4)。

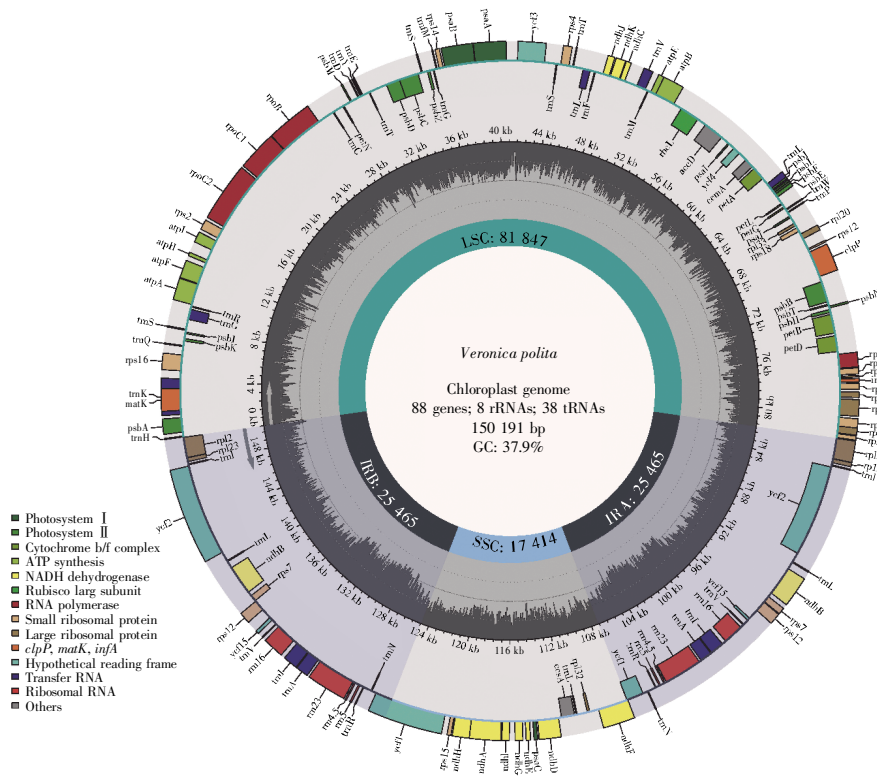


图 1 婆婆纳叶绿体基因组图谱

圈内的基因为顺时针转录, 圈外的基因为逆时针转录; 内部的灰色圈代表 GC 含量。

Fig.1 Chloroplast genome map of *V. polita*

Genes inside the circle are transcribed clockwise, and those outside the circle are transcribed counterclockwise. The inner gray circles represent GC content.

表 2 婆婆纳叶绿体基因组的基因构成
Table 2 Genes of the *V. polita* chloroplast genome

Gene function	Gene group	Gene name
Self-replication	Small subunit of ribosome	<i>rps16*</i> , <i>rps14</i> , <i>rps11</i> , <i>rps3</i> , <i>rps2</i> , <i>rps19</i> , <i>rps18</i> , <i>rps15</i> , <i>rps4</i> , <i>rps8</i> , <i>rps12**</i> (2), <i>rps7</i> (2)
	rRNA genes	<i>rrn16</i> (2), <i>rrn23</i> (2), <i>rrn4.5</i> (2), <i>rrn5</i> (2)
	Large subunit of ribosome	<i>rpl2**</i> (2), <i>rpl16*</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl22</i> , <i>rpl20</i> , <i>rpl14</i> , <i>rpl23</i> (2), <i>rpl36</i>
	DNA dependent RNA polymerase	<i>rpoC1*</i> , <i>rpoC2</i> , <i>rpoB</i> , <i>rpoA</i>
	tRNA genes	<i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnW-CCA</i> , <i>trnH-GUG</i> , <i>trnT-UGU</i> , <i>trnS-UGA</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnK-UUU*</i> , <i>trnS-GGA</i> , <i>trnS-GCU</i> , <i>trnQ-UUG</i> , <i>trnR-UCU</i> , <i>trnL-UAG*</i> , <i>trnL-UAA*</i> , <i>trnP-UGG</i> , <i>trnI-CAU</i> (2), <i>trnA-UGC*</i> (2), <i>trnL-CAA</i> (2), <i>trnN-GUU</i> (2), <i>trnR-ACG</i> (2), <i>trnV-GAC</i> (2), <i>trnG-UCC*</i> , <i>trnV-UAC*</i> , <i>trnG-GCC</i> , <i>trnT-GGU</i> , <i>trnI-GAU*</i> (2), <i>trnM-CAU</i> , <i>trnY-GUA</i>
Photosynthesis	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF*</i> , <i>atpH</i> , <i>atpI</i>
	Subunits of NADH dehydrogenase	<i>ndhA*</i> , <i>ndhB**</i> (2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Cytochrome b/f complex	<i>petB*</i> , <i>petD*</i> , <i>petA</i> , <i>petN</i> , <i>petL</i> , <i>petG</i>
	Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaJ</i> , <i>psaI</i>
	Photosystem II	<i>psbZ</i> , <i>psbT</i> , <i>psbN</i> , <i>psbM</i> , <i>psbL</i> , <i>psbK</i> , <i>psbJ</i> , <i>psbI</i> , <i>psbD</i> , <i>psbA</i> , <i>psbH</i> , <i>psbF</i> , <i>psbE</i> , <i>psbB</i> , <i>psbC</i>
Others	Subunit of rubisco	<i>rbcL</i>
	Subunit of acetyl-CoA-carboxylase	<i>accD</i>
	Envelope membrane protein	<i>cemA</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	Protease gene	<i>clpP**</i>
	Maturase gene	<i>matK</i>
Unkown function	Translation initiation factor IF-1	<i>infA</i>
	Conserved open reading frame	<i>ycf1</i> (2), <i>ycf15#</i> (2), <i>ycf2</i> (2), <i>ycf3**</i> , <i>ycf4</i>

注: *, 含有 1 个内含子的基因; **, 含有 2 个内含子的基因; #, 假基因; (2), 具有 2 个拷贝的基因。

Notes: *, gene with one intron; **, gene with two introns; #, pseudogene; (2), gene with two copies.

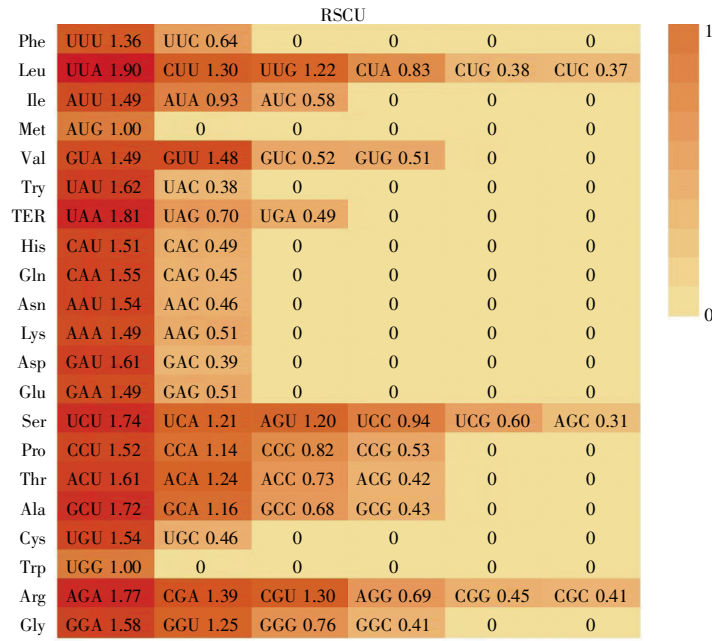


图 2 婆婆纳叶绿体基因组的 RSCU 值热图
Fig.2 Heat map of RSCU values in the *V. polita* chloroplast genome

表 3 婆婆纳叶绿体基因组的相对同义密码子使用度
Table 3 RSCU in the *V. polita* chloroplast genome

Amino acid	Codon	No.	RSCU value	Amino acid	Codon	No.	RSCU value
Phe	UUU	1 010	1.36	Met	AUG	624	1.00
	UUC	480	0.64	Ser	UCU	608	1.74
Leu	UUA	898	1.90		UCC	328	0.94
	UUG	575	1.22	UCA	425	1.21	
	CUU	616	1.30	UCG	211	0.60	
	CUC	175	0.37	AGU	421	1.20	
	CUA	391	0.83	AGC	109	0.31	
	CUG	181	0.38	Pro	CCU	428	1.52
Ile	AUU	1 094	1.49		CCC	230	0.82
	AUC	429	0.58		CCA	321	1.14
	AUA	683	0.93		CCG	148	0.53
Val	GUU	542	1.48	Thr	ACU	542	1.61
	GUC	192	0.52		ACC	247	0.73
	GUA	544	1.49		ACA	419	1.24
Tyr	GUG	185	0.51	Ala	ACG	142	0.42
	UAU	765	1.62		GCU	606	1.72
TER	UAC	178	0.38		GCC	241	0.68
	UAA	52	1.81		GCA	409	1.16
	UAG	20	0.70	GCG	153	0.43	
His	UGA	14	0.49	Cys	UGU	227	1.54
	CAU	474	1.51		UGC	68	0.46
Gln	CAC	154	0.49	Trp	UGG	464	1.00
	CAA	732	1.55		Arg	CGU	344
Asn	CAG	214	0.45	CGC		108	0.41
	AAU	997	1.54	CGA		367	1.39
Lys	AAC	296	0.46	CGG	119	0.45	
	AAA	1 078	1.49	AGA	468	1.77	
Asp	AAG	369	0.51	AGG	182	0.69	
	GAU	863	1.61	Gly	GGU	564	1.25
GAC	209	0.39	GGC		183	0.41	
Glu	GAA	1 011	1.49		GGA	715	1.58
	GAG	343	0.51	GGG	344	0.76	

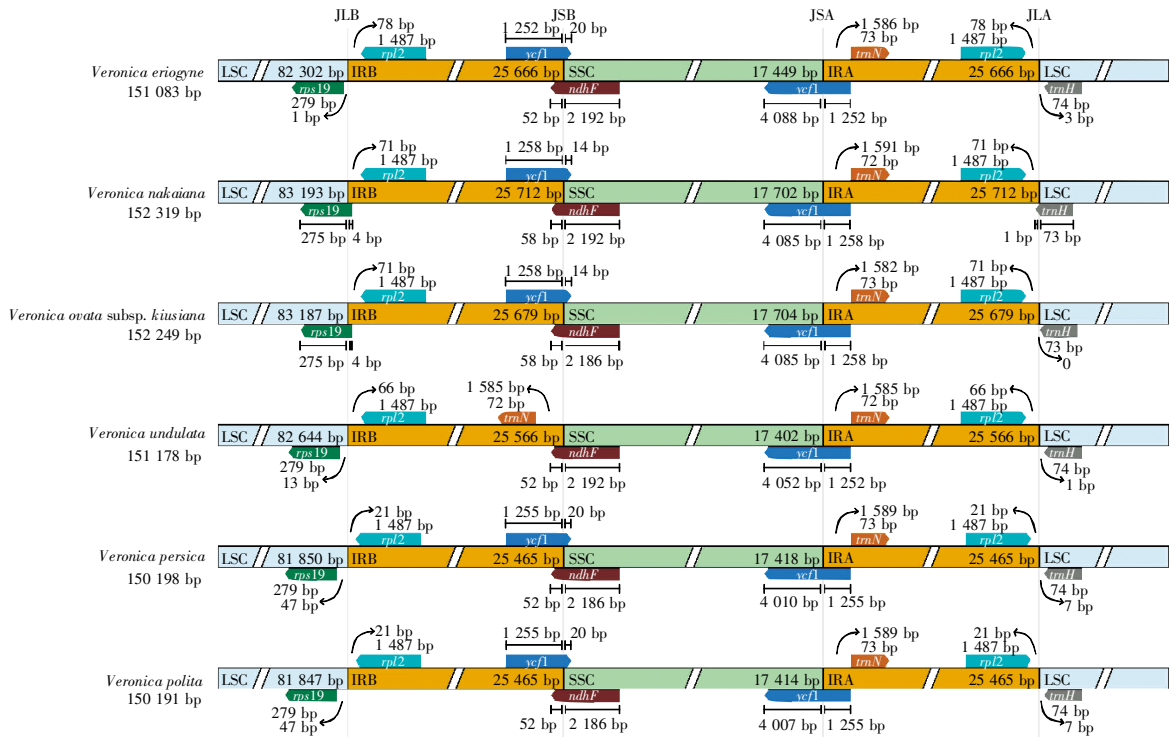


图 3 婆婆纳属各物种之间的 IR 边界比较

Fig.3 Comparison of IR boundaries among species of Veronica

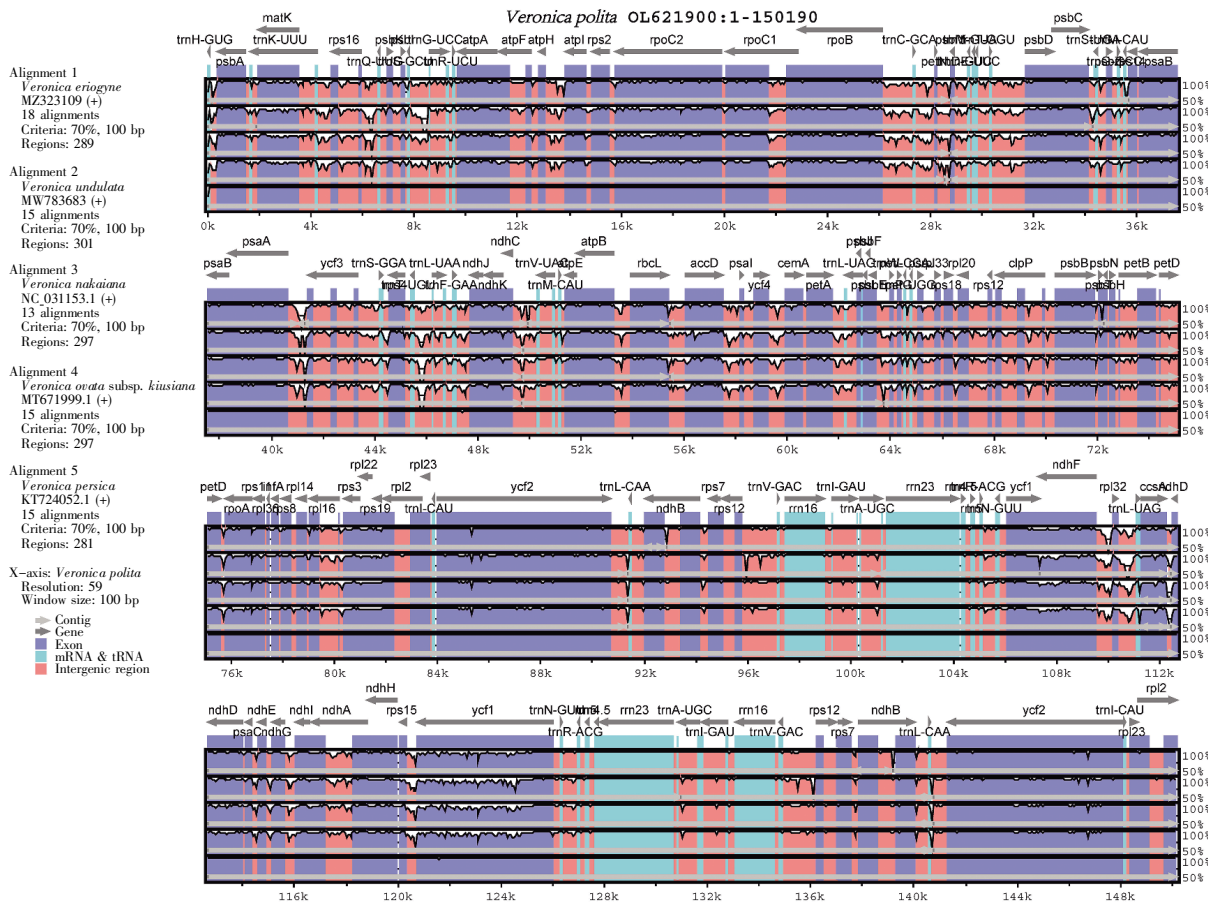


图 4 婆婆纳属植物叶绿体基因组的序列比对

Fig.4 Alignment of chloroplast genome sequences among species of Veronica

2.5 基于叶绿体基因组的系统发育位置

贝叶斯推理法(Bayesian inference, BI)系统发育树(图 5)显示,车前科(Plantaginaceae)的各族植物均聚为相对独立的一支;水八角族(Gratioleae)位于系统发育树的基部;婆婆纳所属的婆婆纳族(Veroniceae)植物以强支持率(100%)聚为一支,说明该族为单系类群。此外,婆婆纳属植物也以强支持率(100%)聚为一支,说明该属可能为单系类群。其中,婆婆纳和阿拉伯婆婆纳以强支持率(100%)聚为一支,表明二者有很近的亲缘关系;婆婆纳、阿拉伯婆婆纳和毛果婆婆纳(*V. eriogyne*)聚为一支,而其他 3 个婆婆纳属物种聚为一支;在婆婆纳族内,胡黄连属(*Neopicrorhiza*)和婆婆纳属聚为一支,而该族的其他类群也以强支持率聚为较大的一支。

3 讨论

本研究基于高通量测序和生物信息分析获得了婆婆纳叶绿体基因组的特征信息,其叶绿体基因组有典型的四分体结构,序列总长为 150 191 bp,极其接近已知的婆婆纳属完整叶绿体基因组的平均长度(150 993 bp)。该基因组的总 GC 含量为 37.9%,其 4 个区段的 GC 含量有所差异(表 1),其中 IR 区的 GC 含量最高(43.2%),原因是该区域有较多富含 GC 的 tRNA(图 1),这与其他婆婆纳属植物^[26]的 GC 分布特征相似。该叶绿体基因组共有 134 个功能基因(表 2),与婆婆纳属其他叶绿体基因组的基因数量和种类^[26-28]一致,表明婆婆纳属植物的叶绿体基因组均比较保守,这也符合陆地高等植物叶绿体基因组的一般特征^[1-2]。本研究绘

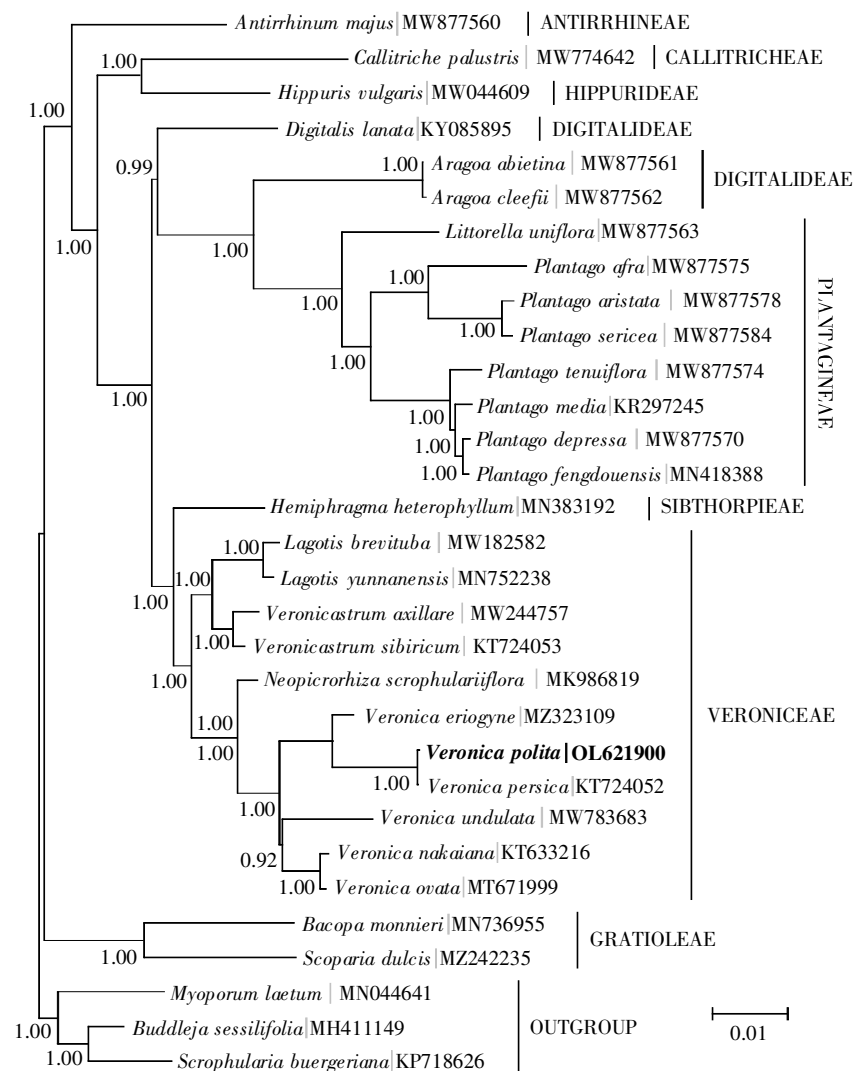


图 5 基于叶绿体基因组蛋白质编码序列的系统发育树

Fig.5 BI tree based on protein coding sequences of the chloroplast genomes

制了婆婆纳叶绿体基因组的物理图谱,并对其所含基因进行了分类,实现了该基因组基本信息的直观展示,以方便后续研究者针对该基因组开展其他相关工作。

密码子的分析结果显示,婆婆纳叶绿体基因组蛋白质编码基因的总 GC 含量(38.0%)略高于完整序列的 GC 含量(37.9%);使用最多的密码子是编码 Ile 的 AUU (1 094 个),而使用最少的密码子是 TER 中的 UGA (14 个)(表 3),这与蛋白质中氨基酸和终止子的数量相关,蛋白质中 Ile 的含量本身就高,而一个基因通常只有一个终止密码子。在所有的密码子中,仅 Met 和 Trp 两种氨基酸密码子不存在使用偏好性(RSCU 值为 1),因为它们仅由一个对应密码子编码。

通过对已知叶绿体基因组的所有婆婆纳属植物进行比较分析,发现它们的叶绿体基因组在基因种类、基因数量、基因组结构等方面的特征均高度保守。婆婆纳属各物种叶绿体基因组的四分体边界均无明显的扩张或收缩(图 3)。不同植物的叶绿体基因组往往伴随着重复区和单拷贝区之间的边界收缩与扩张,这常常会导致基因组长度的变化^[29]。婆婆纳属叶绿体基因组边界的高度保守,导致基因组长度无较大差异,其中最短的毛果婆婆纳(151 083 bp)^[28]和最长的 *V. nakaiana* (152 319 bp)^[26]的基因组长度差仅为 1 236 bp,而在同样具有四分体叶绿体基因组结构的紫堇属内,金球黄堇(*Corydalis boweri*; GenBank 接收号: MZ983401)和紫堇(*C. edulis*; GenBank 接收号: MW110633)^[30]叶绿体基因组完整序列的长度差高达 46 528 bp。序列差异分析(图 4)表明,6 个婆婆纳属植物的叶绿体基因组的基因排列顺序和转录方向相同,总体序列差异不大,但是它们在基因间隔区的序列上广泛存在变异,并且也有一些变异分布在蛋白质编码区域(如 *ycf1*、*ndhF*、*rpoC2* 等),这种兼有保守性和变异性的特征让其很适合开发为分子标记,用于种间遗传多样性或系统进化研究。

基于贝叶斯推理法的系统发育树(图 5)显示,车前科的各族植物为单系类群;水八角族是车前科中的较古老类群;婆婆纳族的胡黄连属和婆婆纳属亲缘关系最近;在婆婆纳属内,婆婆纳、阿拉伯婆婆纳和毛果婆婆纳三者亲缘关系较近,水苦苣、*V. nakaiana* 和 *V. ovata* 三者亲缘关系较近,婆婆纳和阿拉伯婆婆纳的亲缘关系最近,这与婆婆纳属叶绿体基因组的序列比较结果吻合。然而,

以往未涉及婆婆纳的研究表明,与阿拉伯婆婆纳亲缘关系最近的物种是毛果婆婆纳^[28],可见在进行系统发育分析时,一些物种数据的缺失会导致系统发育关系判断的片面。婆婆纳属植物多达约 300 个物种,分子生物学和化学的证据均表明该属具有比较复杂的进化关系^[31-33],而目前已获得叶绿体基因组完整序列的婆婆纳属物种数量极少,很难凭此得出令人信服的婆婆纳属系统发育关系,所以在未来需要开展更多的分子遗传学和植物形态学研究,包括进行婆婆纳属其他物种的叶绿体基因组研究,以为婆婆纳属的系统发育问题提供更多的研究证据。

总的来讲,本研究完成了婆婆纳叶绿体基因组完整序列的测定,运用生物信息学的研究方法揭示了其基本特征,并利用蛋白质编码序列探析了婆婆纳的系统发育关系,可以为婆婆纳属的资源鉴定、遗传多样性等研究提供可用的数据。

参考文献(References):

- [1] DANIELL H, LIN C S, YU M, *et al.* Chloroplast genomes: diversity, evolution, and applications in genetic engineering[J]. *Genome Biology*, 2016, 17: 134.
- [2] 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 Molecular Biology*, 2011, 76(3): 273-297.
- [3] ASAF S, KHAN A L, AAQIL KHAN M, *et al.* Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other *Glycine* species[J]. *PLoS One*, 2017, 12(8): e0182281.
- [4] EGAMBERDIEV S S, SAHA S, SALAKHUTDINOV I, *et al.* Comparative assessment of genetic diversity in cytoplasmic and nuclear genome of upland cotton[J]. *Genetica*, 2016, 144(3): 289-306.
- [5] ALWADANI K G, JANES J K, ANDREW R L. Chloroplast genome analysis of box-ironbark *Eucalyptus*[J]. *Molecular Phylogenetics and Evolution*, 2019, 136: 76-86.
- [6] HONG D Y, YANG H B, JIN C L, *et al.* Scrophulariaceae[M]// WU Z Y, RAVEN P H. *Flora of China*. Beijing: Science Press, 1998: 76.
- [7] 旦增曲培, 郭肖, 东知多杰, 等. 藏药巴夏嘎本草考证及同名异物形成原因分析[J]. *世界科学技术-中医药现代化(DANZENG Qupei, GUO Xiao, DONGZHI Duo jie, et al. Herbal verification of the Tibetan medicine baxiaga and the analysis of the causes of the formation of homonym of baxiaga)*. *Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology*, 2020, 22(10): 3746-3752.
- [8] 陈静, 格桑顿珠. 藏药巴夏嘎化学成分及其在藏药方中应用的研究进展[J]. *中医药导报(CHEN Jing, GE Sangdunzhu. Study on chemical constituents and prescription of Tibetan medicine *Veronica eriogone* H. Winkl.)*. *Guiding Journal of Traditional Chinese Medicine and Pharmacology*, 2017, 23(10): 46-47.

- [9] 刘芳, 张远芳, 李玲环, 等. 婆婆纳抗黑色素瘤物质基础的初步研究[J]. 广西植物(LIU Fang, ZHANG Yuanfang, LI Linghuan, *et al.* Preliminary study on the material basis of anti-melanoma activity of *Veronica didyma*[J]. Guihaia), 2023, 43(1): 60–67.
- [10] AKANDA M R, NAM H H, TIAN W S, *et al.* Regulation of JAK2/STAT3 and NF- κ B signal transduction pathways; *Veronica polita* alleviates dextran sulfate sodium-induced murine colitis[J]. Biomedicine & Pharmacotherapy, 2018, 100: 296–303.
- [11] 晏方勤, 李晓, 焦坤, 等. 邻作婆婆纳对花生苗期主要害虫和天敌种群的影响[J]. 花生学报(YAN Fangqin, LI Xiao, JIAO Kun, *et al.* The effects of neighbouring *Veronica persica* on the population of main pests and natural enemies during peanut seedling stage[J]. Journal of Peanut Science), 2022, 51(1): 23–29.
- [12] 李庚飞, 徐承. 婆婆纳对铅污染耐受性的研究[J]. 北方园艺(LI Gengfei, XU Cheng. Study on the tolerance of *Veronica didyma* Tenore to lead[J]. Northern Horticulture), 2015(8): 70–73.
- [13] 赵云山, 毕雅琼, 雷露静, 等. “巴沙嘎”类蒙药品种整理与研究进展[J]. 中国中药杂志(ZHAO Yunshan, BI Yaqiong, LEI Lujing, *et al.* Variety systematization and research progress of Mongolian medicine “Bashaga”[J]. China Journal of Chinese Materia Medica), 2017, 42(5): 998–1004.
- [14] 王恒, 欧阳辉, 唐辉, 等. 婆婆纳的性状与显微鉴定[J]. 中药材(WANG Heng, OUYANG Hui, TANG Hui, *et al.* Characterization and microscopic identification of *Veronica polita*[J]. Journal of Chinese Medicinal Materials), 2017, 40(2): 325–327.
- [15] 王莉娟, 张樱山, 马瑞瑞, 等. 藏药婆婆纳 HPLC 特征图谱的建立与主成分含量测定[J]. 中兽医医药杂志(WANG Lijuan, ZHANG Yingshan, MA Ruirui, *et al.* Establishment of HPLC characteristic map and determination of principal component in Tibetan medicine *Veronica officinalis*[J]. Journal of Traditional Chinese Veterinary Medicine), 2022, 41(3): 17–22.
- [16] BARREIRA J C M, DIAS M I, ŽIVKOVIĆ J, *et al.* Phenolic profiling of *Veronica* spp. grown in mountain, urban and sandy soil environments[J]. Food Chemistry, 2014, 163: 275–283.
- [17] DOYLE J J, DOYLE J L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue[J]. Phytochemical Bulletin, 1987, 19: 11–15.
- [18] JIN J J, YU W B, YANG J B, *et al.* GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes[J]. Genome Biology, 2020, 21: 241.
- [19] TILLICH M, LEHWARK P, PELLIZZER T, *et al.* GeSeq: versatile and accurate annotation of organelle genomes[J]. Nucleic Acids Research, 2017, 45(W1): W6–W11.
- [20] ZHENG S Y, POCZAI P, HYVÖNEN J, *et al.* Chloroplast: an online program for the versatile plotting of organelle genomes[J]. Frontiers in Genetics, 2020, 11: 576124.
- [21] INGVARSSON P K. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*[J]. Molecular Biology and Evolution, 2007, 24(3): 836–844.
- [22] AMIRYOUSEFI A, HYVÖNEN J, POCZAI P. IRscope: an online program to visualize the junction sites of chloroplast genomes[J]. Bioinformatics, 2018, 34(17): 3030–3031.
- [23] FRAZER K A, PACHTER L, POLIAKOV A, *et al.* VISTA: computational tools for comparative genomics[J]. Nucleic Acids Research, 2004, 32(Suppl. 2): W273–W279.
- [24] MILNE I, LINDNER D, BAYER M, *et al.* TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops[J]. Bioinformatics, 2009, 25(1): 126–127.
- [25] RONQUIST F, HUELSENBECK J P. MrBayes 3: Bayesian phylogenetic inference under mixed models[J]. Bioinformatics, 2003, 19(12): 1572–1574.
- [26] CHOI K S, CHUNG M G, PARK S. The complete chloroplast genome sequences of three Veroniceae species (Plantaginaceae): comparative analysis and highly divergent regions[J]. Frontiers in Plant Science, 2016, 7: 355.
- [27] MAURYA S, DARSHETKAR A M, YI D K, *et al.* Plastome comparison and evolution within the tribes of Plantaginaceae: insights from an Asian gypsyweed[J]. Saudi Journal of Biological Sciences, 2020, 27(12): 3489–3498.
- [28] DANZENG Q P, SUONAN R Q, LI Q E, *et al.* Characterization of the complete plastid genome of *Veronica eriogyne* H. Winkl., a Tibetan medicinal herb[J]. Mitochondrial DNA Part B, 2021, 6(12): 3402–3403.
- [29] HUANG H, SHI C, LIU Y, *et al.* Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships[J]. BMC Evolutionary Biology, 2014, 14: 151.
- [30] LIU Y Y, KAN S L, WANG J L, *et al.* Complete chloroplast genome sequences of *Corydalis edulis* and *Corydalis shensiana* (Papaveraceae)[J]. Mitochondrial DNA Part B, 2021, 6(1): 257–258.
- [31] TASKOVA R M, ALBACH D C, GRAYER R J. Phylogeny of *Veronica*: a combination of molecular and chemical evidence[J]. Plant Biology, 2004, 6(6): 673–682.
- [32] ALBACH D C, MEUDT H M. Phylogeny of *Veronica* in the Southern and Northern Hemispheres based on plastid, nuclear ribosomal and nuclear low-copy DNA[J]. Molecular Phylogenetics and Evolution, 2010, 54(2): 457–471.
- [33] 宋晓琳. 婆婆纳属植物系统学位置探讨[D]. 曲阜: 曲阜师范大学(SONG Xiaolin. Discussion on the Phylogenetic Position of *Veronica*[D]. Qufu: Qufu Normal University), 2021.