

# 北柴胡、紫花阔叶柴胡叶绿体全基因组解析及柴胡属药用植物叶绿体基因组比较与系统发育分析

张明英, 张雨曲, 李依民, 高 静, 沈 霞, 杨新杰, 颜永刚, 王昌利, 张 岗\*

(陕西中医药大学药学院/陕西省秦岭中草药应用开发工程技术研究中心, 陕西 西安 712046)

**摘要:** 柴胡属 (*Bupleurum* L.) 是伞形科 (Apiaceae) 中具有重要经济价值的药用类群。本研究利用 Illumina HiSeq X Ten 平台测序获得北柴胡 (*B. chinense* DC.) 和紫花阔叶柴胡 (*B. boissieuanum* H. Wolff) 的叶绿体全基因组序列, 对其进行了组装、注释和特征分析, 并与同属已发表的叶绿体全基因组进行了比较和系统发育分析。北柴胡和紫花阔叶柴胡叶绿体全基因组大小分别为 155 458、155 800 bp, 均为由一个大单拷贝区 (large single copy, LSC; 85 343、85 804 bp)、一个小单拷贝区 (small single copy, SSC; 17 495、17 410 bp) 和一对反向重复区 (inverted repeat, IRa/IRb; 26 310、26 293 bp) 构成的环状四分体结构; 两者分别注释得到 129 个基因, 包括 84 个蛋白编码基因、37 个 tRNA 基因和 8 个 rRNA 基因; 此外, 两者重复序列的类型与分布模式相似, 但数量有所差异。比较基因组学分析结果表明, 柴胡属植物叶绿体全基因组大小、结构、GC 含量及基因组成和排列顺序等在种内、种间均高度保守, IRs 区未出现明显扩张或收缩; 序列的种间变异高于种内, 非编码序列 (包括基因间区和内含子) 变异高于编码基因序列, LSC 和 SSC 区序列变异高于 IRs 区; 此外, 筛选到 11 条核苷酸多样性较高的种间高变异序列, 分别位于 LSC 和 SSC 区。系统发育分析结果强烈支持柴胡属为单系, 其中, 北柴胡同种不同个体聚为一支, 并与紫花鸭跖柴胡 (*B. commelynoideum* H. Boissieu) 亲缘关系最近, 而紫花阔叶柴胡与三岛柴胡 (*B. falcatum* L.) 亲缘关系更近。本研究将为柴胡属药用植物的分类鉴定、系统发育及资源开发利用等相关研究提供基础。

**关键词:** 柴胡属; 叶绿体基因组; 简单重复序列; 序列变异; 系统发育

中图分类号: R931

文献标识码: A

文章编号: 0513-4870(2021)02-0618-12

## Complete plastid genomes of *Bupleurum chinense* DC. and *B. boissieuanum* H. Wolff, with comparative and phylogenetic analyses of medicinal *Bupleurum* species

ZHANG Ming-ying, ZHANG Yu-qu, LI Yi-min, GAO Jing, SHEN Xia, YANG Xin-jie,  
YAN Yong-gang, WANG Chang-li, ZHANG Gang\*

(College of Pharmacy and Shaanxi Qinling Application Development and Engineering Center of Chinese Herbal Medicine, Shaanxi University of Chinese Medicine, Xi'an 712046, China)

**Abstract:** *Bupleurum* L. (Apiaceae) is an economically important genus, in which many species are of medicinal value. In this study, the complete plastid genomes (plastomes) of *B. chinense* DC. and *B. boissieuanum* H. Wolff were sequenced and their characteristics were investigated. Comparative and phylogenetic analyses were conducted

收稿日期: 2020-08-28; 修回日期: 2020-09-23.

基金项目: 陕西省自然科学基金基础研究计划 (2019JQ-876); 陕西中医药大学校级科研课题 (2020GP34); 陕西中医药大学博士科研启动经费 (104080001); 陕西中医药大学思逸人才工程 (141306200105); 陕西中医药大学“秦药”品质评价及资源开发学科创新团队项目 (2019-QN01); 2019 年医疗服务与保障能力提升补助资金项目 (中医药事业传承与发展部分第四次全国中药资源普查项目) (财社【2019】39 号); 陕西省高校青年杰出人才支持计划项目。

\*通讯作者 Tel: 86-29-38185165, E-mail: jay\_gumling2003@aliyun.com

DOI: 10.16438/j.0513-4870.2020-1419

with other published *Bupleurum* plastomes. The complete plastomes of *B. chinense* and *B. boissieuanum* were 155 458 and 155 800 bp in length, and both exhibited the typical quadripartite circular structure consisting of a large single copy region (LSC, 85 343 and 85 804 bp), a small single copy region (SSC, 17 495 and 17 410 bp), and a pair of inverted repeat regions (IRa/b, 26 310 and 26 293 bp), respectively. A total of 129 genes, including 84 protein-coding genes, 37 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes were identified from each of the two plastomes. Repeat sequences detected were similar in types and distribution patterns, but the numbers were slightly different. Comparative analyses revealed that the *Bupleurum* plastomes were highly conserved in length, structure, the guanine and cytosine (GC) content, and gene content and order, both intraspecifically and interspecifically, and no obvious expansion or contraction of the inverted repeat regions occurred. Sequence variation was lower within the same species than among different species, noncoding sequences (including intergenic regions and introns) showed a higher divergence than the protein-coding sequences, and sequences in the LSC and SSC regions were more divergent than those in the IR regions. In addition, 11 sequences with higher nucleotide diversity among species were detected in the LSC and SSC regions. All studied *Bupleurum* species were inferred forming a monophyletic group with a 100% bootstrap value. *Bupleurum chinense* and *B. boissieuanum* were phylogenetically closest to *B. commelynoideum* and *B. falcatum*, separately, with all three *B. chinense* accessions clustered into a distinct clade. These results provide genetic information for further species identification, phylogenetic resolution, and will assist in exploration and utilization of medicinal *Bupleurum* species.

**Key words:** *Bupleurum*; plastid genome; simple sequence repeat; sequence divergence; phylogeny

柴胡属 *Bupleurum* L. 是伞形科 Apiaceae 一个具有重要药用和经济价值的类群。中国分布有柴胡属植物约 42 种 16 变种<sup>[1]</sup>, 其中 36 种 (包括一些变种及变型) 均可入药使用<sup>[2-4]</sup>。中药材柴胡 (*Bupleuri Radix*) 在《神农本草经》中被列为“上品”, 具有疏散退热、疏肝解郁、升举阳气之功效, 是最常用的大宗药材之一。《中华人民共和国药典》(2020 年版, 一部)<sup>[5]</sup>规定, 中药材柴胡来源于伞形科柴胡属药用植物柴胡 (北柴胡) *B. chinense* DC. 或狭叶柴胡 (红柴胡) *B. scorzonifolium* Willd. 的干燥根。药理学研究揭示柴胡属药用植物含有皂苷、挥发油、黄酮、甾醇等活性成分, 具有抗炎、抗肿瘤、调节神经和免疫等作用<sup>[6-8]</sup>, 因而在现代药物开发研究中被广泛使用。

柴胡属也是分类和系统发育研究中的一个“困难”类群。染色体核型<sup>[9]</sup>、果实形态<sup>[10,11]</sup>、花粉形态<sup>[12]</sup>以及 ITS、*trnH-psbA*、*matK*、*rbcL* 和 *rps16* 等 DNA 分子标记<sup>[13-16]</sup>为柴胡属植物的分类鉴定与系统学研究提供了重要依据。然而, 虽然柴胡属的单系性已经得到分子证据的支持, 但属内不同物种间的系统发育关系目前仍然存在争议<sup>[17,18]</sup>。高通量测序技术和生物信息学分析方法的快速发展使得叶绿体全基因组序列在解析低阶元“困难”植物类群系统发育关系研究中得到广泛应用, 并且在解决柴胡属植物种间系统发育关系问题中的优势已初见端倪<sup>[17]</sup>。目前柴胡属北柴胡<sup>[17,19]</sup>、红柴胡<sup>[20]</sup>、紫花阔叶柴胡 *B. boissieuanum* H. Wolff<sup>[21]</sup>、三岛柴胡 *B. falcatum* L.<sup>[22]</sup>、紫花鸭跖柴胡 *B. commelynoideum* de Boiss.<sup>[17]</sup>、竹叶柴胡 *B. marginatum* Wall. ex DC.<sup>[23]</sup>、龙血

树柴胡 *B. dracaenoides* Huan C. Wang, Z. R. He & H. Sun<sup>[24]</sup>和长茎柴胡 *B. longicaule* Wall. ex DC.<sup>[25]</sup>等物种的叶绿体基因组已有研究报道, 但不同种间及同种不同个体间叶绿体比较基因组学和系统发育相关研究还鲜有报道。

本研究新测序获得北柴胡和紫花阔叶柴胡两个物种的叶绿体全基因组序列, 对其进行组装、注释和特征解析, 并与柴胡属其他已发表的叶绿体基因组进行比较基因组学和系统发育分析, 旨在全面解析北柴胡、紫花阔叶柴胡叶绿体全基因组序列与结构特征; 探讨柴胡属植物叶绿体基因组的整体特征与变异模式, 比较序列的种间、种内变异情况, 并筛选种间高变异序列; 基于叶绿体系统发育基因组学分析解析柴胡属代表药用植物种间系统发育关系。进而为柴胡属药用植物的分类鉴定、保护遗传学及资源开发利用等相关研究提供基础。

## 材料与方法

**实验材料** 野生北柴胡、紫花阔叶柴胡分别采自陕西省宝鸡市太白县桃川镇路坪沟 (33°1'12"N, 107°30'0"E) 和宝河沟 (34°3'0"N, 107°31'48"E), 由陕西中医药大学药学院王继涛高级实验师鉴定。采集到的新鲜幼嫩叶片放入硅胶中快速脱水干燥, 用于基因组 DNA 提取, 凭证标本 (采集号: 610330\_1908、610330\_1905, 采集人: 张明英) 保存于陕西中医药大学中药标本馆。

**基因组 DNA 提取与测序** 利用植物基因组 DNA 提取试剂盒 (TIANGEN, 北京) 提取总 DNA, 1.0% 的琼脂糖凝胶电泳检测 DNA 质量, Qubit 3.0 (Thermo

Fisher Scientific, USA) 荧光定量仪检测 DNA 浓度。检测合格后的基因组总 DNA 构建长度约 350 bp 的插入片段文库, 利用 Illumina HiSeq X Ten 平台进行序列读长为 150 bp 的双端测序 (paired-end sequencing)。利用 NGS QC ToolKit<sup>[26]</sup> 对测序得到的原始序列 (raw reads) 进行质控 (参数设置选择默认值), 过滤去除接头 (adapter) 和低质量序列 (reads), 得到高质量待分析序列 (clean reads)。

**叶绿体全基因组序列组装、拼接与注释** 以 GenBank 数据库中收录的北柴胡叶绿体全基因组序列 (Accession: NC\_046774) 作为序列延伸的种子 (seed), 在 Linux 系统中运行脚本, 利用 GetOrganelle 程序包依次调用 Bowtie2、BLAST、SPAdes 软件, 对 clean reads 进行筛选、比对和组装, 获得序列重叠群 (contigs)<sup>[27]</sup>。利用 Bandage 软件<sup>[28]</sup> 将 contigs 拼接成环并除去冗余序列, 同时将拼接结果与上述 seed (NC\_046774) 序列在 Geneious 软件<sup>[29]</sup> 中进行比对以确定反向重复区 (inverted repeat regions, IRs) 的方向, 初步完成叶绿体全基因组序列的从头组装。再利用 Bowtie2 将原始序列重新映射到组装完成的叶绿体全基因组序列上, 通过检测序列覆盖度来评估和验证大单拷贝区 (large single copy, LSC)、小单拷贝区 (small single copy, SSC) 与两个反向重复区的 4 个边界及各 contigs 拼接处的正确性<sup>[30]</sup>, 最终得到完整的叶绿体全基因组序列。

同样以北柴胡叶绿体全基因组 (NC\_046774) 的注释结果作为参考, 在 Linux 系统中运行脚本, 利用 PGA (plastid genome annotator) 软件<sup>[31]</sup> 对组装完成的叶绿体全基因组序列进行注释, 并在 Geneious 软件中校正注释结果。利用在线软件 tRNAscan-SE (<http://lowelab.ucsc.edu/tRNAscan-SE/>) 确定所有 tRNA 基因的边界。最后, 利用 Organellar Genome DRAW (OGDRAW) 软件 (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) 在线绘制叶绿体全基因组图谱。组装并注释完成后的北柴胡、紫花阔叶柴胡叶绿体全基因组序列利用 sequin 软件编辑并提交至 GenBank 数据库, 检索号分别为 MN756857 和 MN756856。

**重复序列检测与叶绿体全基因组比较分析** 利用 REPuter 软件 (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) 检测叶绿体全基因组序列中的分散重复序列 (dispersed repeats), 参数设置为最小重复序列长度 = 30 bp, 重复序列间的相似度 >90%<sup>[32]</sup>。串联重复序列 (tandem repeats) 利用 Tandem repeats finder 软件 (<https://tandem.bu.edu/trf/trf.html>) 进行检测, 参数设置选择默认值<sup>[33]</sup>。目前, 柴胡属北柴胡 (NC\_046774 和 MN893666)、红柴胡 (MT239475)、紫花阔叶柴胡 (NC\_036017)、紫

花鸭跖柴胡 (MT162552)、竹叶柴胡 (MN968501)、以及三岛柴胡 (KM207676) 和 *B. latissimum* Nakai (NC\_033346) 共 7 种 8 个个体的叶绿体全基因组序列已在 GenBank 数据库开放获取。利用 MISA 软件对本研究新测序获得的北柴胡、紫花阔叶柴胡与上述 8 个个体叶绿体全基因组序列中的简单重复序列 (simple sequence repeats, SSRs) 分别进行检测, 参数设置均为单核苷酸重复单元不少于 10 个, 二核苷酸重复单元不少于 5 个, 三核苷酸和四核苷酸重复单元不少于 4 个, 五核苷酸和六核苷酸重复单元不少于 3 个, 且两个 SSRs 之间的距离不小于 100 bp<sup>[34]</sup>。并对 SSRs 的类型、数量和分布模式分别进行比较分析。

利用 mVISTA 软件在 LAGAN 模型<sup>[35]</sup> 下对本研究新测序获得的北柴胡与紫花阔叶柴胡以及柴胡属上述 8 个已发表个体的叶绿体全基因组序列进行比较分析。同时, 对种间及同种不同个体间叶绿体基因组的 IR/SC 边界位置进行比较, 分析 IRs 区收缩与扩张情况。

**系统发育与序列变异分析** 将包括本研究新测序的北柴胡、紫花阔叶柴胡在内的柴胡属所有 10 个叶绿体全基因组序列构建序列矩阵, 并选取同来自伞形科的药用植物当归属 *Angelica* L. 当归 *A. sinensis* (Oliv.) Diels (NC\_042826) 和藁本属 *Ligusticum* L. 藁本 *L. sinense* Oliv. (NC\_038088) 作为外类群, 利用 RAxML 以最大似然法 (maximum likelihood, ML)<sup>[36]</sup> 进行系统发育重建分析。序列比对利用 MAFFT 软件<sup>[37]</sup> 完成, 并利用 Geneious 软件对比对结果进行必要的手动检查调整; 核苷酸替换模型选择 GTR+G; 系统发育树各分支的靴带支持率 (bootstrap values, BS) 通过进行 1 000 次自展重复分析获得。此外, 利用 MEGA 软件<sup>[38]</sup> 分别计算种间、种内遗传距离 (pairwise distance)。

为了进一步分析柴胡属植物叶绿体全基因组序列间的变异情况, 将所有 7 种 10 个个体及北柴胡同种 3 个个体的叶绿体全基因组序列 (去掉一个 IR 区) 分别构建多序列矩阵, 比对和检查调整之后, 利用 DNAsp 软件<sup>[39]</sup> 进行滑动窗口分析 (sliding window analysis), 计算核苷酸多样性指数 Pi, 并根据分析结果筛选种间高变异序列。滑动窗口 (windows length) 大小设置为 600 bp, 步长 (step size) 为 200 bp。将筛选到的种间高变异序列分别提取出来, 构建联合矩阵, 再次进行系统发育分析, 验证其对于种间划分与系统发育关系的分辨率。方法和参数设置同前。

## 结果与分析

### 1 叶绿体全基因组测序、序列组装拼接与注释

北柴胡和紫花阔叶柴胡测序分别得到 38 031 172 条

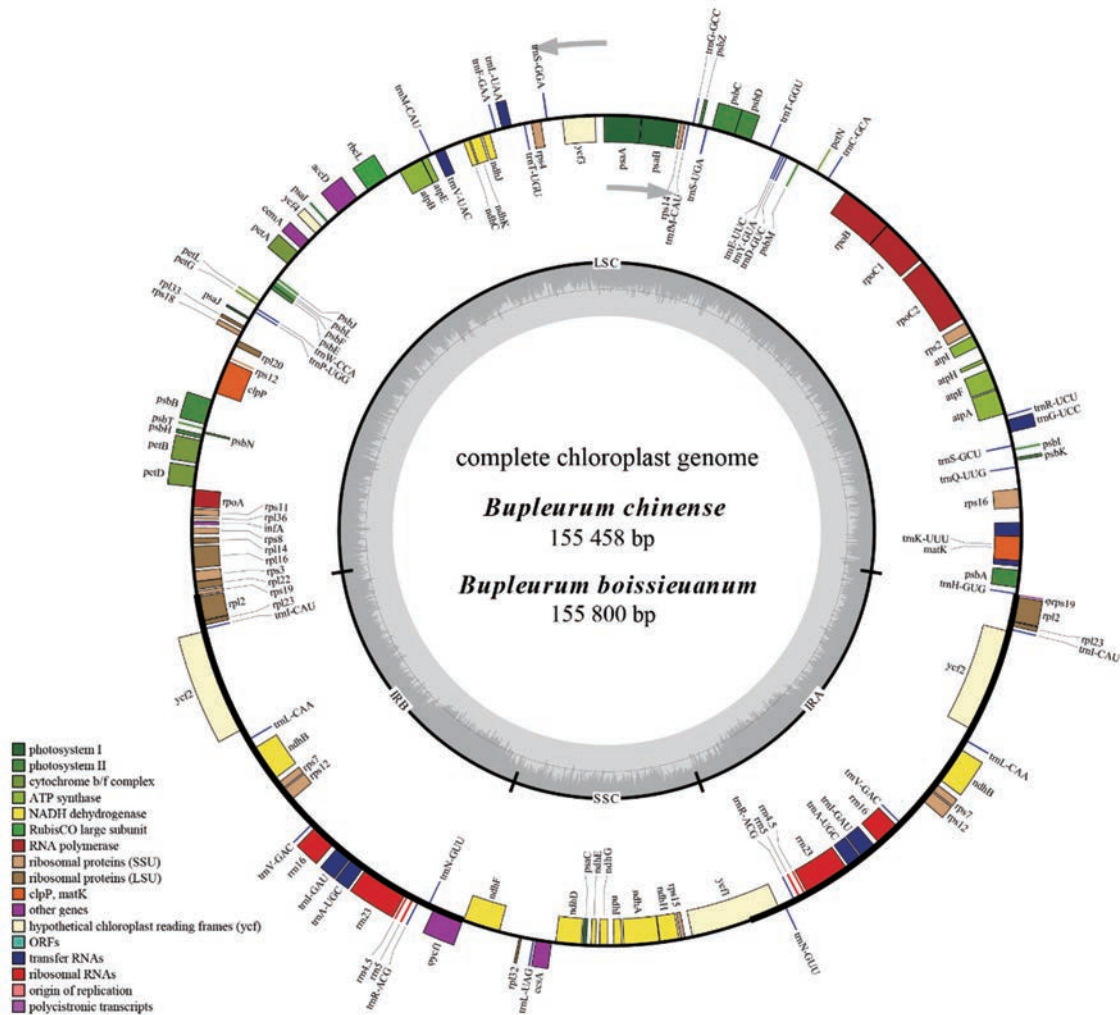
(6.5 Gb) 和 42 649 376 条 (7.29 Gb) raw reads, 过滤去除 adapters 和低质量 reads 后最终分别得到 36 190 390 条和 41 233 108 条 clean reads, 即 clean reads 的比例分别为 95.16% 和 96.68%。组装拼接后两者均得到完整的环状四分体结构叶绿体全基因组, 序列总长度分别为 155 458 bp 和 155 800 bp, 由大单拷贝区 (长度分别为 85 343 bp、85 804 bp)、小单拷贝区 (17 495 bp、17 410 bp) 和一对反向重复区 (26 310 bp、26 293 bp) 组成 (图 1)。两者叶绿体全基因组序列及 LSC 和 IRs 区的 GC 含量均相同, 依次为 37.7%、35.8% 和 42.8%, SSC 区的 GC 含量分别为 31.4% 和 31.5% (表 1)。

在北柴胡和紫花阔叶柴胡叶绿体全基因组序列中分别注释得到 129 个基因, 包括 84 个蛋白编码基因、37 个 tRNA 基因和 8 个 rRNA 基因 (表 2)。其中, 6 个蛋白编码基因 (*rpl2*、*rpl23*、*rps7*、*rps12*、*ndhB* 和 *ycf2*)、7 个 tRNA 基因 (*trnI-CAU*、*trnL-CAA*、*trnV-GAC*、*trnI-GAU*、

*trnA-UGC*、*trnR-ACG* 和 *trnN-GUU*) 和 4 个 rRNA 基因 (*rrn16*、*rrn23*、*rrn4.5*、*rrn5*) 分别在两个反向重复区出现一次重复, 同时, 由于 *rps19* 和 *ycf1* 基因分别横跨 LSC/IRb 和 SSC/IRa 边界, 因此在 IRa 和 IRb 区对应位置分别产生一个假基因  $\phi$ rps19 和  $\phi$ ycf1 (图 1、2)。此外, *rpl16*、*rpl2*、*rpoC1*、*rps16*、*atpF*、*ndhB*、*ndhA*、*petB*、*petD* 和 *trnK-UUU*、*trnG-UCC*、*trnL-UAA*、*trnV-UAC*、*trnI-GAU*、*trnA-UGC* 基因各包含有 1 个内含子, *rps12*、*ycf3* 和 *clpP* 基因各包含 2 个内含子, 同时, *rps12* 基因的 5'端和两个重复的 3'端分别位于 LSC 区和 IR 区, 被分成 2 个独立的转录单元, 为反式剪切 (trans-spliced) 基因。

## 2 重复序列分析

北柴胡和紫花阔叶柴胡叶绿体全基因组序列中分别检测到 33 条和 38 条长度不小于 30 bp 且重复序列间相似度大于 90% 的分散重复序列, 包括正向重复 (forward repeats, F)、反向重复 (reverse repeats, R)、回文



**Figure 1** Plastid genome maps of *B. chinense* and *B. boissieuianum*. Genes inside and outside the circle are transcribed in a clockwise and counter clockwise direction, respectively. Genes are color-coded based on their functions. The grey area in the inner circle indicates the GC content of the plastid genome sequence

**Table 1** Comparison of the plastid genome features of *B. chinense*, *B. boissieuianum* and the other eight accessions from *Bupleurum*

Species	Accession	Genome		LSC		SSC		IR		CDS	
		Length /bp	G+C /%	Length /bp	G+C /%	Length /bp	G+C /%	Length /bp	G+C /%	Length /bp	G+C /%
<i>B. chinense</i>	MN756857	155 458	37.7	85 343	35.8	17 495	31.4	26 310	42.8	77 874	38.2
	MN893666	155 869	37.7	85 772	35.8	17 501	31.4	26 298	42.8	78 288	38.2
	NC_046774	155 545	37.7	85 430	35.8	17 505	31.4	26 305	42.8	77 520	38.2
<i>B. boissieuianum</i>	MN756856	155 800	37.7	85 804	35.8	17 410	31.5	26 293	42.8	77 853	38.2
	NC_036017	156 108	37.7	86 007	35.8	17 495	31.4	26 303	42.8	77 988	38.1
<i>B. falcatum</i>	KM207676	155 989	37.7	85 870	35.8	17 517	31.4	26 301	42.8	77 952	38.1
<i>B. latissimum</i>	NC_033346	155 621	37.6	85 473	35.8	17 548	31.3	26 300	42.8	77 253	38.1
<i>B. marginatum</i>	MN968501	154 991	37.8	84 910	35.9	17 505	31.7	26 288	42.8	77 979	38.2
<i>B. commelynoideum</i>	MT162552	155 629	37.7	85 542	35.8	17 495	31.4	26 296	42.8	78 294	38.2
<i>B. scorzonrifolium</i>	MT239475	155 824	37.7	85 591	35.8	17 599	31.3	26 317	42.8	77 907	38.2

**Table 2** Gene composition in the plastid genomes of *B. chinense* and *B. boissieuianum*. Notes: Bold font means genes duplicated in the IRs regions; <sup>a</sup> represents the gene has two separate transcription units; <sup>b</sup> and <sup>c</sup> indicate genes span the LSC and IRb, SSC and IRa regions, respectively; <sup>d</sup> and <sup>e</sup> indicate genes with one and two introns, separately

Category	Gene group	Gene name
Self-replication	Large subunit of ribosome	<i>rpl33</i> , <i>rpl20</i> , <i>rpl36</i> , <i>rpl14</i> , <i>rpl16<sup>d</sup></i> , <i>rpl22</i> , <b><i>rpl12<sup>d</sup></i></b> , <b><i>rpl23</i></b> , <i>rpl32</i>
	DNA-dependent RNA polymerase	<i>rpoC2</i> , <i>rpoC1<sup>d</sup></i> , <i>rpoB</i> , <i>rpoA</i>
	Small subunit of ribosome	<i>rps16<sup>d</sup></i> , <i>rps2</i> , <i>rps14</i> , <i>rps4</i> , <i>rps18</i> , <b><i>rps12<sup>a</sup></i></b> , <i>rps11</i> , <i>rps8</i> , <i>rps3</i> , <i>rps19<sup>b</sup></i> , <b><i>rps7</i></b> , <i>rps15</i>
	Ribosome RNA genes	<b><i>rrn16</i></b> , <b><i>rrn23</i></b> , <b><i>rrn4.5</i></b> , <b><i>rrn5</i></b>
	Transfer RNA genes	<i>trnH</i> -GUG, <i>trnK</i> -UUU <sup>d</sup> , <i>trnQ</i> -UUG, <i>trnS</i> -GCU, <i>trnG</i> -UCC <sup>d</sup> , <i>trnR</i> -UCU, <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnY</i> -GUA, <i>trnE</i> -UUC, <i>trnT</i> -GGU, <i>trmS</i> -UGA, <i>trmG</i> -GCC, <i>trnM</i> -CAU, <i>trnS</i> -GGA, <i>trnT</i> -UGU, <i>trnL</i> -UAA <sup>d</sup> , <i>trnF</i> -GAA, <i>trnV</i> -UAC <sup>d</sup> , <i>trnM</i> -CAU, <i>trnW</i> -CCA, <i>trnP</i> -UGG, <b><i>trnI</i>-CAU</b> , <b><i>trnL</i>-CAA</b> , <b><i>trnV</i>-GAC</b> , <b><i>trnI</i>-GAU<sup>d</sup></b> , <b><i>trnA</i>-UGC<sup>d</sup></b> , <b><i>trnR</i>-ACG</b> , <b><i>trnN</i>-GUU</b> , <i>trnL</i> -UAG
Photosynthesis	Subunits of ATP synthase	<i>atpA</i> , <i>atpF<sup>d</sup></i> , <i>atpH</i> , <i>atpI</i> , <i>atpE</i> , <i>atpB</i>
	Subunits of NADH-dehydrogenase	<i>ndhJ</i> , <i>ndhK</i> , <i>ndhC</i> , <b><i>ndhB<sup>d</sup></i></b> , <i>ndhF</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhG</i> , <i>ndhI</i> , <i>ndhA<sup>d</sup></i> , <i>ndhH</i>
	Subunits of cytochrome b/f complex	<i>petN</i> , <i>petA</i> , <i>petL</i> , <i>petG</i> , <i>petB<sup>d</sup></i> , <i>petD<sup>d</sup></i>
	Subunits of photosystem I	<i>psaB</i> , <i>psaA</i> , <i>psaI</i> , <i>psaJ</i> , <i>psaC</i>
	Subunits of photosystem II	<i>psbA</i> , <i>psbK</i> , <i>psbI</i> , <i>psbM</i> , <i>psbD</i> , <i>psbC</i> , <i>psbZ</i> , <i>psbJ</i> , <i>psbL</i> , <i>psbF</i> , <i>psbE</i> , <i>psbB</i> , <i>psbT</i> , <i>psbN</i> , <i>psbH</i>
Other genes	Subunit of rubisco	<i>rbcL</i>
	proteins photosystem I assembly	<i>ycf4</i> , <i>ycf3<sup>e</sup></i>
	Subunit of acetyl-CoA-carboxylase	<i>accD</i>
	c-Type cytochrome synthesis gene	<i>ccsA</i>
	Envelop membrane protein	<i>cemA</i>
	Protease	<i>clpP<sup>e</sup></i>
	Translational initiation factor	<i>infA</i>
Maturase	<i>matK</i>	
Proteins of unknown function	<b><i>ycf2</i></b> , <i>ycf1<sup>c</sup></i>	

重复 (palindromic repeats, P) 和互补重复 (complement repeats, C) 序列, 数量依次分别为 16 和 17.0 和 3、16 和 16 以及 1 和 2。同时, 分别检测到 22 条和 30 条长度为 9~49 bp、重复次数为 2~4 次的串联重复序列。SSRs 的数量、类型及其在叶绿体基因组中的分布情况分别见表 3、4。这些 SSRs 主要分布于叶绿体基因组的 LCS 区 (约占总数的 66.67%~75%), 并且主要分布于非编码的基因间区序列中。编码基因序列总长度占叶绿体全基因组全长的 49.64%~50.31%, 而编码基因序列中分布的 SSRs 数量仅占其总数的 13.89%~22.03%。

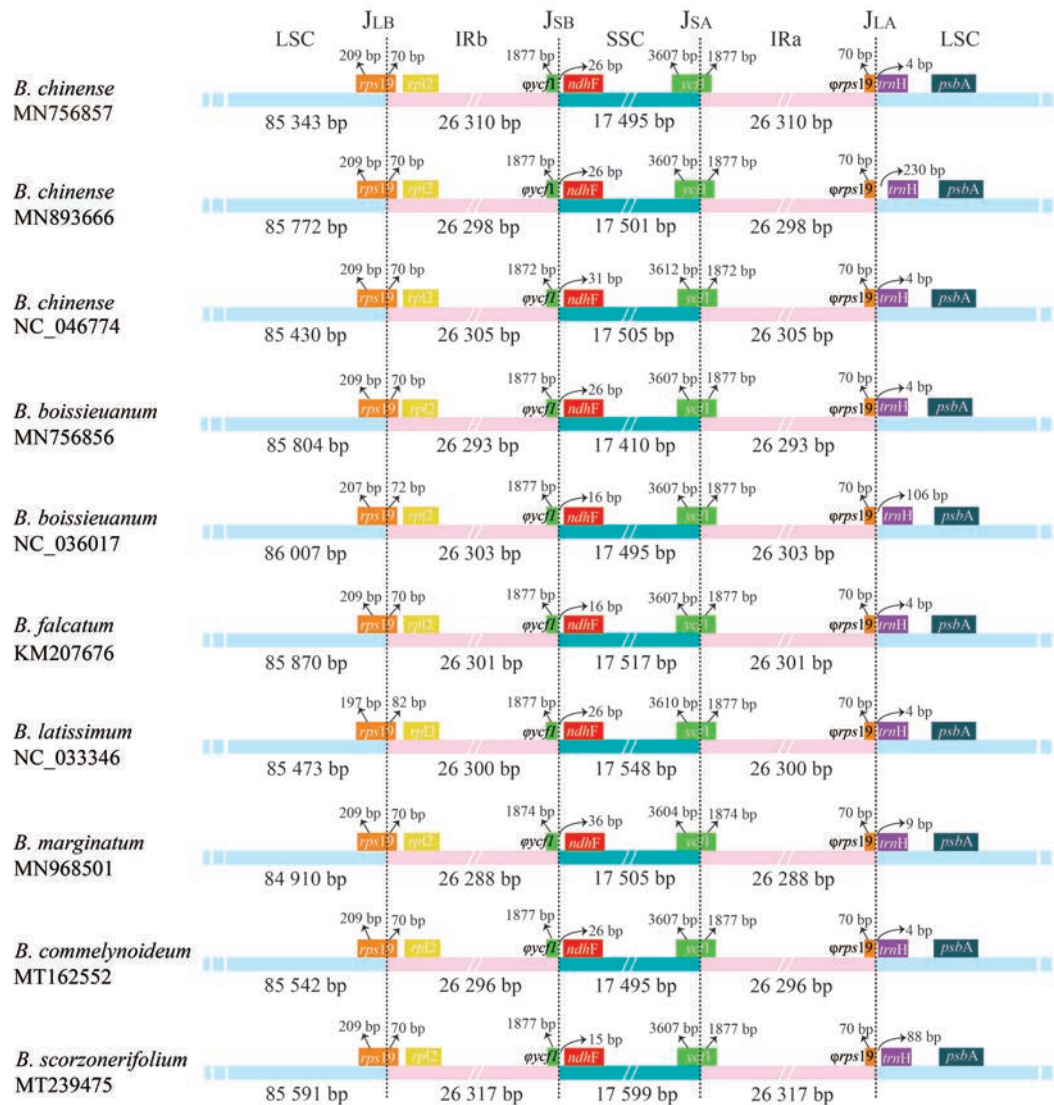
### 3 比较基因组学分析

叶绿体全基因组序列比较分析结果见图 2、3。柴

胡属植物叶绿体全基因组序列所编码的基因类别、数量和排列顺序在种间、种内均高度一致; 序列间的变异主要发生在非编码基因间区, 且种内变异小于种间; 此外, IRs 区没有出现明显扩张或收缩现象。核苷酸多样性分析结果显示, 种间序列的 Pi 值整体变化范围为 0~0.022 2, LSC、SSC 和 IRs 区的平均 Pi 值依次为 0.006 3、0.008 8 和 0.001 1。同时, 筛选到 11 条 Pi > 0.015 的种间高变异序列, 分别位于 LSC 区 (*trnK*-UGG\_ *rps16*、*petN*\_ *psbM*×3、*petA*\_ *psbJ*×2+*psbJ*) 和 SSC 区 (*ndhF*+*ndhF*\_ *rpl32*×2、*ycf1*×3), 见图 4。

### 4 系统发育分析

系统发育建树结果如图 5 所示, 柴胡属所有 10 个



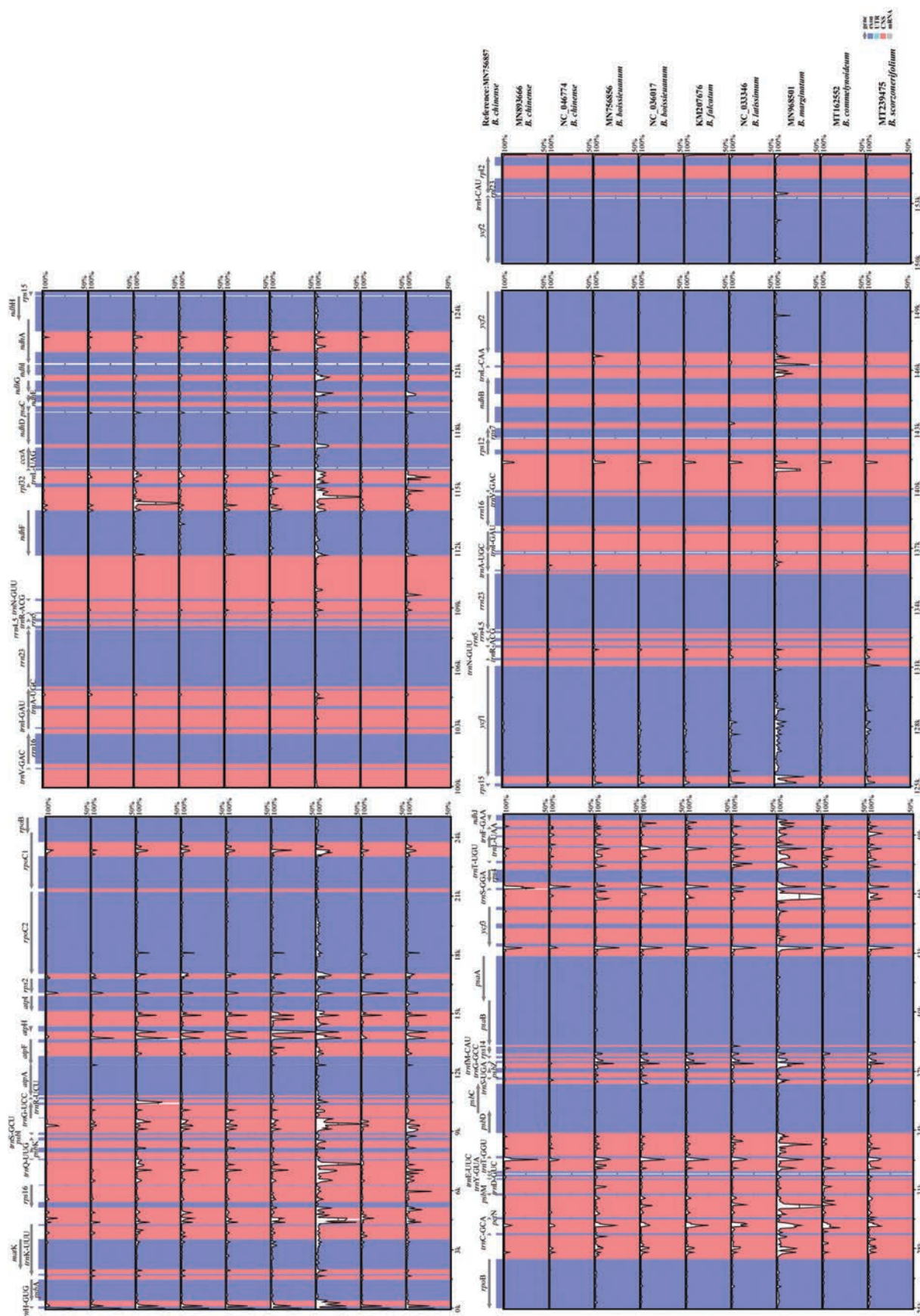
**Figure 2** Comparisons of the borders of LSC, SSC and IRA/b regions among the ten *Bupleurum* plastid genomes. Numbers above represent the distance between the gene ends and the borders sites, and numbers below represent the length of the LSC, SSC and IRA/b regions. The  $\phi$  symbol indicates pseudogenes. This figure is not to scale

**Table 3** Numbers and distribution of the detected SSRs in the plastid genomes of *B. chinense*, *B. boissieuianum* and the other eight accessions from *Bupleurum*

Species	Accession	Total SSRs	SSRs number and percentage/%			
			CDS	LSC	IRs	SSC
<i>B. chinense</i>	MN756857	57	11 (19.30)	42 (73.68)	12 (21.05)	3 (5.26)
	MN893666	66	12 (18.18)	48 (72.73)	14 (21.21)	4 (6.06)
	NC_046774	57	11 (19.30)	41 (71.93)	11 (21.05)	4 (7.02)
<i>B. boissieuianum</i>	MN756856	72	10 (13.89)	53 (73.61)	18 (25.00)	1 (1.39)
	NC_036017	68	13 (19.12)	50 (73.53)	14 (20.59)	4 (5.88)
<i>B. falcatum</i>	KM207676	64	12 (18.75)	48 (75.00)	14 (21.88)	2 (3.13)
<i>B. latissimum</i>	NC_033346	64	10 (15.63)	44 (68.75)	13 (20.31)	7 (10.94)
<i>B. marginatum</i>	MN968501	59	13 (22.03)	40 (67.80)	12 (20.34)	7 (11.86)
<i>B. commelynoideum</i>	MT162552	61	10 (16.39)	44 (72.13)	12 (19.67)	5 (8.20)
<i>B. scorzonifolium</i>	MT239475	57	11 (19.30)	38 (66.67)	12 (21.05)	7 (12.28)

个体以 100% 的支持率聚为一个单系, 其中, 竹叶柴胡、红柴胡和 *B. latissimum* 依次从基部独立分化出来; 北柴胡同种 3 个个体聚为一支, 并与紫花鸭跖柴胡构

成姐妹关系; 紫花阔叶柴胡与三岛柴胡亲缘关系最近, 但紫花阔叶柴胡同种两个不同个体 MN756856 和 NC\_036017 却并未聚类为一支, 而是 NC\_036017 与三岛柴



**Figure 3** Sequence identity among the plastid genomes of *B. chinense* and *B. boissieuanum* and the other eight accessions from *Bupleurum*, with *B. chinense* (MN756857) as reference. Coding and noncoding regions are colored in blue and red, respectively

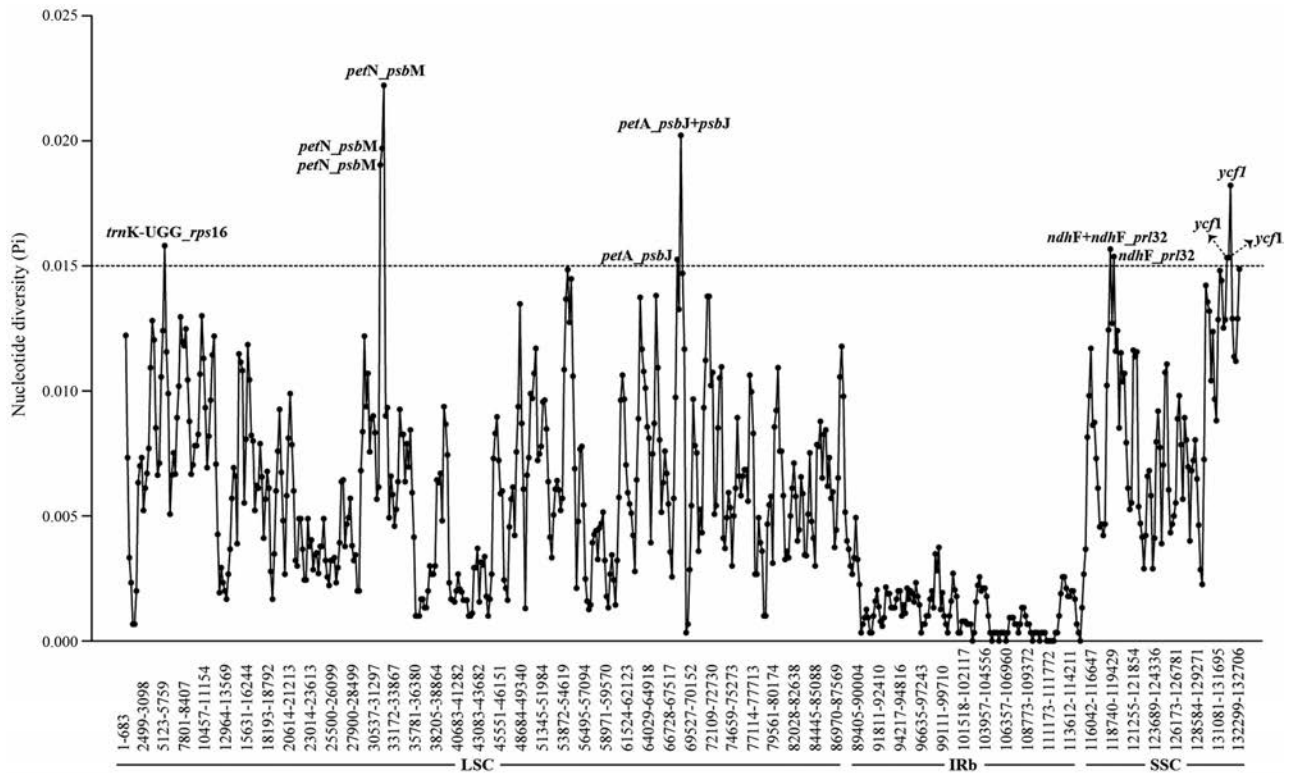


Figure 4 The nucleotide diversity of the ten *Bupleurum* plastid genomes. Window length: 600 bp; step size: 200 bp

Table 4 Types and their proportions of the detected SSRs in the plastid genomes of *B. chinense*, *B. boissieuianum* and the other eight accessions from *Bupleurum*

Accession	SSRs type				
	Mon /%	Di /%	Tri /%	Penta /%	Hexa /%
MN756857	42 (73.68)	8 (14.03)	7 (12.28)	/	/
MN893666	48 (72.72)	8 (12.12)	7 (10.61)	2 (3.03)	1 (1.52)
NC_046774	42 (73.67)	8 (14.03)	7 (12.28)	/	/
MN756856	47 (65.28)	15 (20.83)	7 (9.72)	1 (1.39)	2 (2.78)
NC_036017	51 (75.00)	10 (14.70)	7 (10.29)	/	/
KM207676	47 (73.44)	9 (14.06)	6 (9.38)	/	2 (3.13)
NC_033346	47 (73.44)	9 (14.06)	5 (7.81)	3 (4.69)	/
MN968501	41 (69.49)	11 (18.64)	5 (8.47)	2 (3.39)	/
MT162552	41 (67.21)	9 (14.75)	7 (11.47)	3 (4.92)	1 (1.64)
MT239475	39 (68.41)	11 (19.30)	5 (8.77)	2 (3.51)	/

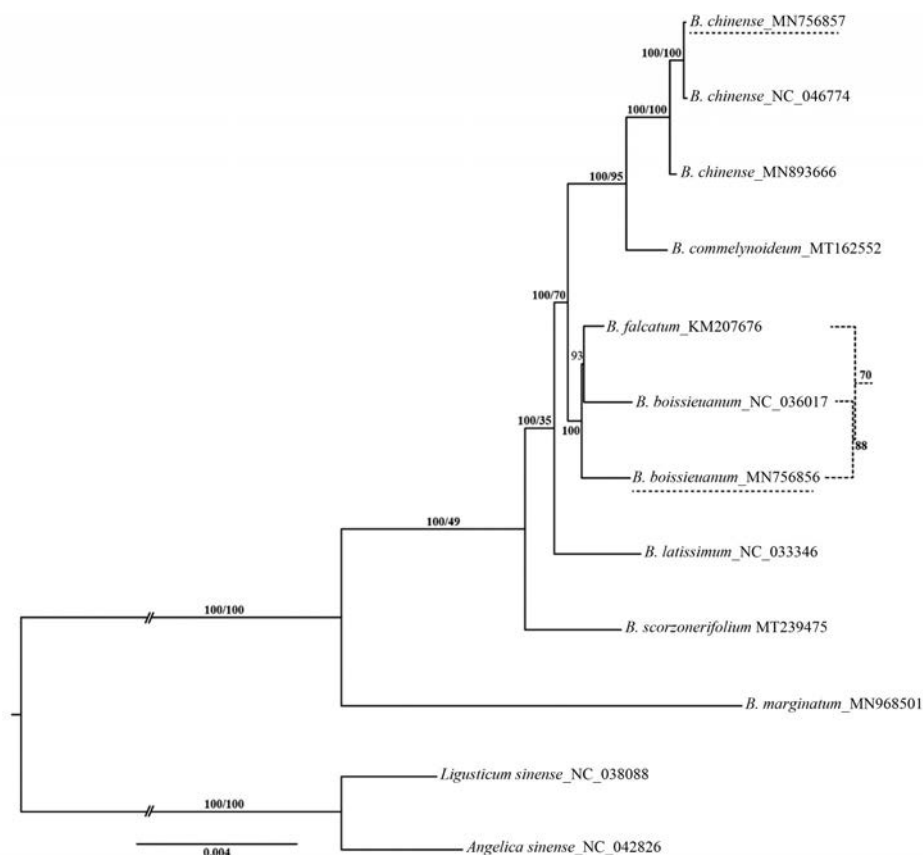
胡以 93% 支持率构成姐妹关系, 再与 MN756856 共同聚为一个单系 (支持率为 100%)。11 个高变异序列联合分析构建的系统发育树干群节点支持率普遍偏低, 同时, 紫花阔叶柴胡 MN756856、NC\_036017 和三岛柴胡三者之间的拓扑关系与叶绿体全基因组序列的建树结果出现冲突, 即 MN756856 和 NC\_036017 以 88% 的支持率构成姐妹关系, 再与三岛柴胡聚为一支 (支持率为 70%)。

### 讨论

本研究对北柴胡和紫花阔叶柴胡叶绿体全基因组

的分析结果表明, 两者均具有被子植物叶绿体基因组典型的环状四分体结构, 所编码的基因类别、数量及排列顺序完全相同, 同时具有高度相似的 GC 含量 (仅 SSC 区有 0.1% 的微小差别)。其中, IRs 区序列的 GC 含量最高, 各区域序列中 GC 含量从高到低依次为 IRs > Genome > LSC > SSC。进一步分析发现, IRs 区分布的 4 个 rRNA 基因序列中较高的 GC 含量 (50.5%~56.5%, 平均 55.3%) 是引起 IRs 区整体 GC 含量偏高的一个重要因素。此外, 北柴胡、紫花阔叶柴胡叶绿体全基因组序列的总长度、CG 含量、基因组成等与柴胡属其他已报道的物种亦高度一致<sup>[17, 19-22]</sup>, 但与同为伞形科的当归属<sup>[40]</sup>、四带芹属 *Tetrataenium* (DC.) Manden.<sup>[41]</sup>等植物却存在不同程度的差异。包括本研究新测序的北柴胡、紫花阔叶柴胡在内的柴胡属 7 种 10 个不同个体叶绿体全基因组、LSC、SSC 和 IRs 区长度的种间差异范围分别为 8~1 117 bp、32~1 097 bp、0~198 bp 和 1~22 bp, 即总长度的差异主要来自于 LSC 区, 而 SSC 和 IRs 区长度相对保守。序列比对分析发现, LSC 区长度差异主要是由非编码基因间区中出现的串联重复序列和短片段插入/缺失所引起。而相比于不同物种间, 北柴胡同种 3 个不同个体间叶绿体全基因组、LSC、SSC 及 IRs 区的长度则更为保守。

被子植物叶绿体基因组的环状四分体结构使得其两个反向重复区与大、小单拷贝区之间存在 4 个边界,



**Figure 5** Maximum likelihood phylogenetic tree of *Bupleurum* inferred utilizing datasets of the complete plastid genome sequences and the concatenated 11 highly divergent sequences (differences in topology were indicated by dotted lines). Numbers beside each node are bootstrap values (BS) inferred from the two datasets

即 LSC/IRb、IRb/SSC、SSC/IRa 和 IRa/LSC。在进化过程中, IRs 区常发生收缩或向单拷贝区扩张现象,这也是导致被子植物不同类群叶绿体基因组长度差异的一个重要原因<sup>[42]</sup>。Downie 等<sup>[43]</sup>研究发现,伞形科植物叶绿体基因组的 LSC/IRb 及 IRa/LSC 边界在不同属间常存在较大变异。本研究对柴胡属 7 种 10 个代表植物个体叶绿体全基因组 IR/SC 边界位置的分析结果显示, IRs 区长度分别为 26 288~26 310 bp,其中, *rps19* 基因横跨 LSC 和 IRb 区,其下游有 70 bp 的序列位于 IRb 区; *ycf1* 基因横跨 SSC 和 IRa 区,下游有 1 871~1 877 bp 的序列分布于 IRa 区,种间、种内均未出现明显的 IRs 区收缩或扩张现象,说明柴胡属植物叶绿体基因组 IRs 区大小和基因组成高度保守。但与伞形科当归属<sup>[40]</sup>、四带芹属<sup>[41]</sup>、天胡荽属 *Hydrocotyle* L.<sup>[43]</sup> 等的植物 IRs 区大小均存在明显差异,与 Downie 等的研究结果一致。

SSRs 也称为微卫星 DNA (microsatellites DNA),是以少数核苷酸(一般 1~6 个)为基本重复单元构成的简单串联重复序列,普遍存在于真核生物基因组中,具有多态性高、分布广泛、两端(侧翼)序列高度保守

等特点<sup>[44]</sup>。此外,分布于植物叶绿体基因组中的 SSRs 还同时兼具单亲遗传模式,被作为分子标记广泛用于群体遗传学<sup>[45]</sup>、谱系地理学<sup>[46]</sup>等相关研究。本研究分析结果表明,柴胡属 7 种 10 个代表植物个体叶绿体全基因组序列中的 SSRs 均以单核苷酸重复(占总数的 65.28%~75.00%)为主要类型,其次为二核苷酸(12.12%~20.83%)和三核苷酸(7.81%~12.28%),且随着拷贝数目增加,SSRs 数量明显减少。单核苷酸和二核苷酸 SSRs 的重复序列类型分别以 A/T(约占单核苷酸 SSRs 总数的 91.49%~100%)和 AT/TA(100%)为主,三核苷酸 SSRs 的重复单元亦主要由 A、T 碱基组合(ATA/AAT/ATT/TAA/TAT/TTA, 60%~80%)构成,进一步验证了叶绿体基因组序列中的 SSRs 主要由 polyA 或 polyT 所构成,而较少出现 C 或 G 串联重复这一结论<sup>[47]</sup>。这些 SSRs 可以为柴胡属药用植物分子遗传学相关研究提供候选分子标记。

基于 mVISTA 的叶绿体全基因组序列比较分析结果(图 2)可以看出,柴胡属植物叶绿体全基因组序列的种内变异小于种间,北柴胡同种 3 个不同个体间叶绿体全基因组序列高度保守;同时, IRs 区序列的变异

显著小于 LSC 和 SSC 区, 且编码序列的变异低于非编码序列 (基因间区和部分内含子), 这与 Li 等<sup>[17]</sup>对柴胡属 5 种植物叶绿体基因组序列的分析结果相同, 也与伞形科其他属<sup>[40, 41]</sup>及其他被子植物多个不同类群叶绿体基因组序列变异模式一致<sup>[48-50]</sup>。此外, 核苷酸多样性分析结果显示, 柴胡属不同物种间及北柴胡同种不同个体间叶绿体全基因组序列的核苷酸多样性指数 Pi 平均值分别为 0.005 6 和 0.000 4, 同时, LSC、SSC 和 IRs 区序列的种间 Pi 平均值依次为 0.006 3、0.008 8 和 0.001 1, 进一步证明柴胡属植物叶绿体全基因组序列在种内和种间均高度保守, 且序列变异主要出现在 LSC 和 SSC 区, 与 mVISTA 分析结果一致。

柴胡属药用植物是中药材的重要来源, 但由于属内物种数量较多, 不同种间甚至同种不同个体间常存在不同程度的形态变异, 导致种间分类鉴定与系统发育关系存在较多问题。本研究基于叶绿体全基因组序列构建的系统发育树中, 除紫花阔叶柴胡 (NC\_036017) 与三岛柴胡之间姐妹关系的支持率为 93% 之外, 其余分支均得到 100% 的支持率; 此外, 7 个不同物种间的遗传距离为 0.001~0.015 (平均为 0.005 1), 显著高于北柴胡同种 3 个不同个体间的遗传距离 (0~0.001, 平均为 0.000 3), 说明柴胡属植物叶绿体全基因组序列包含丰富的种间演化信息位点。然而紫花阔叶柴胡同种两个不同个体 MN756856 和 NC\_036017 却并未聚类为一支, 而是 NC\_036017 与三岛柴胡以 93% 支持率构成姐妹关系, 再与 MN756856 共同聚为一个单系。进一步对这 3 个叶绿体全基因组序列进行了两两分析比较, 结果发现, 紫花阔叶柴胡 MN756856 和 NC\_036017 两者之间存在 293 bp 的变异位点, 遗传距离为 0.002, 而三岛柴胡与 MN756856 和 NC\_036017 之间的变异位点数量分别为 207 bp 和 217 bp, 遗传距离均为 0.001。结合前人基于形态特征、染色体核型和核基因序列的研究结论<sup>[9, 11]</sup>, 推测这一结果可能是由于样品鉴定问题引起。因此, 在保证原始实验材料准确鉴定的前提下, 可以增加种间、种内个体取样密度, 并结合核基因数据进一步进行分析验证。

被子植物叶绿体基因组序列中的一些高变异区 (divergence hotspot regions) 序列往往可以作为物种鉴定及系统发育关系分析等相关研究的分子标记<sup>[51]</sup>。本研究在柴胡属种间共筛选到 11 条 Pi 值 > 0.015 的高变异序列。然而, 尽管这 11 条序列联合分析构建的系统发育树拓扑结构与叶绿体全基因组序列分析结果基本一致, 但多个干群节点的支持率明显降低。这一结果说明有限的分子片段所包含的种间系统发育信息不足, 而叶绿体全基因组序列是解决柴胡属植物种间分

类鉴定和系统发育关系问题一个更为理想的选择。

**作者贡献:** 第一作者张明英负责论文设计、实验、数据分析及论文撰写; 通讯作者张岗负责论文设计、数据分析和论文指导; 李依民、高静、沈霞参与实验及数据分析; 颜永刚、杨新杰参与样品采集; 张雨曲、王昌利参与数据分析。所有作者参与论文修改。

**利益冲突:** 本文所有的作者之间不存在利益冲突。

## References

- [1] She ML, Watson MF. Apiaceae (Umbelliferae) in Flora of China [M]. Beijing: Science Press, 2005, 14: 60-74.
- [2] Huang HQ, Wang XH, Fu H, et al. Research progress on medicinal plant resources of *Bupleurum* L. [J]. Chin Tradit Herb Drugs (中草药), 2017, 48: 2989-2996.
- [3] Pan SL, Shun QS, Bai QM, et al. The Coloured Atlas of the Medicinal Plants from Genus *Bupleurum* in China (中国药用柴胡原色图志) [M]. Shanghai: Shanghai Science and Technology Literature Publishing House, 2002.
- [4] Li SJ. *Bupleurum* L. in Medicinal Flora of China (中国药用植物志) [M]. Beijing: Peking University Medical Press, 2018, 7: 607-647.
- [5] Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China (中华人民共和国药典) [M]. Beijing: China Medical Science Press, 2020, 1: 293.
- [6] Sun P, Li YJ, Wei S, et al. Pharmacological effects and chemical constituents of *Bupleurum* [J]. Mini Rev Med Chem, 2019, 19: 34-55.
- [7] Liu SH, Tsai HP, Guo JL, et al. Anti-inflammatory effect and chemical composition of *Bupleurum chinense* and *Bupleurum kaoi* [J]. Nat Prod Chem Res, 2014, 2: 139.
- [8] Yu JQ, Deng AJ, Qin HL. Distinctive features of chemical composition of *Bupleurum chinense* applicable to original authentication [J]. Anal Methods, 2014, 6: 1067-1075.
- [9] Liang QL, Wang CB, Ma XG, et al. Chromosomal study on Chinese *Bupleurum* L. (Apiaceae) [J]. Plant Sci J (植物科学学报), 2013, 31: 11-22.
- [10] Ma XG, Wang CB, He XJ. Micromorphological features of pericarp surface of *Bupleurum* L. (Apiaceae) in China and its taxonomic significance [J]. Acta Bot Boreal-Occident Sin (西北植物学报), 2010, 30: 1388-1396.
- [11] Wang CB, Ma XG, He XJ. Fruit features of some *Bupleurum* species (Apiaceae) and their systematic implication [J]. Plant Sci J (植物科学学报), 2011, 29: 399-408.
- [12] Wang PL, Pu FD. Pollen morphology of *Bupleurum* L. from Sino-Himalaya and its systematic significance [J]. Chin J Appl Environ Biol (应用与环境生物学报), 1995, 1: 34-43.
- [13] Wang CB, Ma XG, He XJ. A taxonomic reassessment in the Chinese *Bupleurum* (Apiaceae): insights from morphology,

- nuclear ribosomal internal transcribed spacer, and chloroplast (*trnH-psbA*, *matK*) sequences [J]. *J Syst Evol*, 2011, 49: 558-589.
- [14] Wang QZ, Zhou SD, Liu TY, et al. Phylogeny and classification of Chinese *Bupleurum* based on nuclear ribosomal DNA internal transcribed spacer and *rps16* [J]. *Acta Biol Crac Ser Bot*, 2008, 50: 105-116.
- [15] Moon BC, Choo BK, Ji YI, et al. Molecular authentication and phylogenetic relationship of *Bupleurum* species by the rDNA-ITS sequences [J]. *Korea J Herbol*, 2009, 24: 59-68.
- [16] Yuan BC, Li WD, Ma YS, et al. The molecular identification of *Bupleurum* medicinal species and the quality investigation of *Bupleuri Radix* [J]. *Acta Pharm Sin (药学报)*, 2017, 52: 162-171.
- [17] Li J, Xie DF, Guo XL, et al. Comparative analysis of the complete plastid genome of five *Bupleurum* species and new insights into DNA barcoding and phylogenetic relationship [J]. *Plants (Basel, Switzerland)*, 2020, 9: 543.
- [18] Zhou J, Gong X, Downie SR, et al. Towards a more robust molecular phylogeny of Chinese Apiaceae subfamily Apioideae: additional evidence from nrDNA ITS and cpDNA intron (*rpl16* and *rps16*) sequences [J]. *Mol Phylogenet Evol*, 2009, 53: 56-68.
- [19] Zhang F, Zhao ZY, Yuan QJ, et al. The complete chloroplast genome sequence of *Bupleurum chinense* DC. (Apiaceae) [J]. *Mitochondr DNA Part B Resour*, 2019, 4: 3665-3666.
- [20] Zhang F, Yang Z, Wang Z, et al. The complete chloroplast genome sequence of *Bupleurum scorzoniferifolium* Willd. (Apiaceae) [J]. *Mitochondr DNA Part B Resour*, 2020, 5: 1998-1999.
- [21] Wu Y, Zhang TZ, Qiu DY, et al. Complete plastid genome of *Bupleurum boissieuianum*, an endemic herb plant in western China [J]. *Conserv Genet Resour*, 2018, 10: 635-637.
- [22] Shin DH, Lee JH, Kang SH, et al. The complete chloroplast genome of the Hare's ear root, *Bupleurum falcatum*: its molecular features [J]. *Genes*, 2016, 7: 20.
- [23] Deng XD, Liu HZ, Yang YH, et al. The first complete chloroplast genome sequence of the medicinal plant *Bupleurum marginatum* (Apiaceae) [J]. *Mitochondr DNA Part B Resour*, 2020, 5: 1836-1838.
- [24] Zhao ZC, Liu J, Zhou MM, et al. Chloroplast genome characterization of *Bupleurum dracaenoides*, a critically endangered woody species endemic to China, with insights of Apioideae phylogeny [J]. *Gene Rep*, 2020. <https://doi.org/10.1016/j.genrep.2020.100784>.
- [25] Yang LC, Xiong F, Xiao YM, et al. The complete chloroplast genome of *Bupleurum longicaule* var. *strictum*, an annual herb endemic to China [J]. *Mitochondr DNA Part B Resour*, 2020, 5: 899-901.
- [26] Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data [J]. *PLoS One*, 2012, 7: e30619.
- [27] Jin JJ, Yu WB, Yang JB, 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, 4: 256479.
- [28] Wick RR, Schultz MB, Zobel J, et al. Bandage: interactive visualization of *de novo* genome assemblies [J]. *Bioinformatics*, 2015, 31: 3350-3352.
- [29] Kears M, Moir R, Wilson A, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data [J]. *Bioinformatics*, 2012, 28: 1647-1649.
- [30] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2 [J]. *Nat Methods*, 2012, 9: 357-359.
- [31] Qu XJ, Moore MJ, Li DZ, et al. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes [J]. *Plant Methods*, 2019, 15: 50.
- [32] Kurtz S, Choudhuri JV, Ohlebusch E, et al. REPuter: the manifold applications of repeat analysis on a genomic scale [J]. *Nucleic Acids Res*, 2001, 29: 4633-4642.
- [33] Benson G. Tandem repeats finder: a program to analyze DNA sequences [J]. *Nucleic Acids Res*, 1999, 27: 573-580.
- [34] Beier S, Thiel T, Munch T, et al. MISA-web: a web server for microsatellite prediction [J]. *Bioinformatics*, 2017, 33: 2583-2585.
- [35] Frazer KA, Pachter L, Poliakov A, et al. VISTA: computational tools for comparative genomics [J]. *Nucleic Acids Res*, 2004, 32 (suppl\_2): W273-W279.
- [36] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies [J]. *Bioinformatics*, 2014, 30: 1312-1313.
- [37] Katoh K, Toh H. Parallelization of the MAFFT multiple sequence alignment program [J]. *Bioinformatics*, 2010, 26: 1899-1900.
- [38] Tamura K, Stecher G, Peterson D, et al. MEGA6: molecular evolutionary genetics analysis version 6.0 [J]. *Mol Biol Evol*, 2013, 30: 2725-2729.
- [39] Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data [J]. *Bioinformatics*, 2009, 25: 1451-1452.
- [40] Park I, Yang S, Kim WJ, et al. Sequencing and comparative analysis of the chloroplast genome of *Angelica polymorpha* and the development of a novel indel marker for species identification [J]. *Molecules*, 2019, 24: 1038.
- [41] Kang L, Xie DF, Xiao QY, et al. Sequencing and analyses on chloroplast genomes of *Tetrataenium candicans* and two allies give new insights on structural variants, DNA barcoding and phylogeny in Apiaceae subfamily Apioideae [J]. *PeerJ*, 2019, 7: e8063.
- [42] Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants [J]. *DNA Res*, 2004, 11: 247-261.

- [43] Downie SR, Jansen RK. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions [J]. *Syst Biol*, 2015, 40: 336-351.
- [44] Powell W, Morgante M, McDevitt R, et al. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines [J]. *Proc Natl Acad Sci U S A*, 1995, 92: 7759-7763.
- [45] Du Q, Wang B, Wei Z, et al. Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers [J]. *J Hered*, 2012, 103: 853-862.
- [46] Chmielewski M, Meyza K, Chybicki I, et al. Chloroplast microsatellites as a tool for phylogeographic studies: the case of white oaks in Poland [J]. *iForest*, 2015, 8: 765-771.
- [47] Kuang DY, Wu H, Wang YL, et al. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics [J]. *Genome*, 2011, 54: 663-673.
- [48] Yang QQ, Jiang M, Wang LQ, et al. Complete chloroplast genome of *Allium chinense*: comparative genomic and phylogenetic analysis [J]. *Acta Pharm Sin* (药学学报), 2019, 54: 173-181.
- [49] Zhou JG, Chen XL, Cui YX, et al. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species [J]. *Int J Mol Sci*, 2017, 18: 1839.
- [50] Zhu S, Niu Z, Xue Q, et al. Accurate authentication of *Dendrobium officinale* and its closely related species by comparative analysis of complete plastomes [J]. *Acta Pharm Sin B*, 2018, 8: 969-980.
- [51] Li XX, Tan W, Sun JH, et al. Comparison of four complete chloroplast genomes of medicinal and ornamental *Meconopsis* species: genome organization and species discrimination [J]. *Sci Rep*, 2019, 9: 10567.