

秦巴山区3种人参属药用植物叶绿体基因组特征分析

彭毅丹^{1,2}, 马楠^{1,2}, 叶奕含^{1,2}, 刘亦欣¹, 谭启怡¹, 陈莹^{1,2}, 张雨曲^{1,2}, 杨新杰^{1,2*}

(1. 陕西中医药大学药学院, 西安 712046; 2. 陕西省秦岭中草药应用开发工程技术研究中心, “秦药”研发重点实验室, 西安 712046)

摘要:以秦巴山区3种人参属药用植物为研究对象,运用生物信息学技术分析其叶绿体基因组特征及密码子使用偏好性等,明确珠子参、羽叶三七和秀丽假人参3种药用植物叶绿体基因组特征与系统发育关系,为珠子参药材的分子鉴定提供依据。结果表明:(1)3种人参属药用植物的叶绿体基因组为典型的四分体结构,序列全长为156 071~156 104 bp,总GC含量为38.10%,基因组大小相似度较高。(2)均注释到133个基因,包括88个蛋白编码基因、37个tRNA基因和8个rRNA基因。(3)3种人参属药用植物叶绿体密码子使用偏好性相似,密码子第3位碱基以A/U结尾为主,密码子使用模式在受到突变影响的同时,主要受到自然选择的影响。(4)系统发育结果表明3种人参属药用植物的亲缘关系较近,且秀丽假人参同羽叶三七亲缘关系更近。该研究表明秀丽假人参同珠子参基源植物之间存在近源关系,这项发现对于珠子参中药材资源的开发利用具有重要意义,并为进一步研究人参属物种的分类、系统发育和进化机制提供重要依据。

关键词:珠子参,羽叶三七,秀丽假人参,叶绿体基因组,结构特征,密码子偏好性

Analysis of the chloroplast genome characteristics of three medicinal plants of *Panax* in the Qinba mountains

PENG Yidan^{1,2}, MA Na^{1,2}, YE Yihan^{1,2}, LIU Yixin¹, TAN Qiyi¹, CHEN Ying^{1,2}, ZHANG Yuqu^{1,2}, YANG Xinjie^{1,2*}

(1. College of Pharmacy, Shaanxi University of Chinese Medicine, Xi'an 712046, China; 2. Shaanxi Qinling Application Development and Engineering Center of Chinese Herbal Medicine, Key Laboratory for Research of “Qin Medicine” of Shaanxi Administration of Traditional Chinese Medicine, Xi'an 712046, China)

Abstract: Based on three medicinal plants of *Panax* in the Qinba Mountain area, this study analyses the chloroplast genome characteristics and codon usage preferences by using bioinformatics techniques to clarify the phylogenetic relationship between the chloroplast genome characteristics of three medicinal plants, *P. japonicus* var. *major*, *P. japonicus* var. *bipinnatifidus*, *P. pseudoginseng* var. *elegantior*, and to provide a basis for molecular identification of *Panax* majoris rhizom. The results were as follows: (1) The chloroplast genomes of three species of *Panax* were typical tetrad structure, the genome size was between 156 071 to 156 104 bp, and the total GC-content was 38.10% with a high degree of genome size similarity; (2) 133 coding genes were annotated, including 88 protein coding genes (PCGs), 37 tRNA genes, and 8 rRNA genes with a high degree of genome size similarity; (3) The chloroplast codon usage preferences of the

基金项目: 陕西省科技厅项目(2021SF-360); 中国医学科学院医学与健康科技创新工程项目(2021-I2M-1-032); 陕西省优秀中医药人才研修项目(陕中医药函[2022]102号)。

第一作者: 彭毅丹(2000-), 女, 硕士研究生, 主要从事中药资源学研究, (E-mail)p117816@163.com。

***通信作者:** 杨新杰, 博士, 副教授, 主要从事中药质量控制、中药资源学研究, (E-mail)xxx211xxx@126.com。

three medicinal plants were comparable, with codon 3 bases predominantly ending in A/U. Despite being affected by mutations, the codon usage patterns of the three medicinal plants were found to be mainly influenced by natural selection. (4) The phylogenetic results showed that the three *Panax* species are closely related, and *P. pseudoginseng* var. *elegantior* is closely related to *P. japonicus* var. *bipinnatifidus*. This study presents a significant relationship between *P. pseudoginseng* var. *elegantior* and two original pharmacopeial plants of *Panax majoris* rhizom. This finding has great importance for the development and utilization of resources from *Panax majoris* rhizom, and provides an important basis for further studies on the classification, phylogeny and evolutionary mechanisms of *Panax* species.

Key words: *Panax japonicus* var. *major*, *P. japonicus* var. *bipinnatifidus*, *P. pseudoginseng* var. *elegantior*, chloroplast genome, structural characteristics, codon preference

人参属 (*Panax*) 为五加科 (Araliaceae) 植物分支之一, 本属约有 8 种, 分布于亚洲东部、喜马拉雅山脉地区, 中南半岛, 北美洲, 我国有 7 种 (1 种为引进栽培) (中国科学院《Flora of China》编委会, 2013)。人参属植物均可入药, 大都是我国传统的名贵中药材。在秦巴山区, 主要分布着 3 种人参属植物, 分别为珠子参 (*P. japonicus* var. *major*)、羽叶三七 (*P. japonicus* var. *bipinnatifidus*) 和秀丽假人参 (*P. pseudoginseng* var. *elegantior*), 在《中国植物志》中均记载为五加科人参属假人参 (*P. pseudoginseng*) 的变种 (中国科学院《中国植物志》编委会, 1978)。这 3 种人参属药用植物的根状茎外形特征极为相似, 均为串珠状, 同时三者均可作为珠子参商品出售。珠子参和羽叶三七被 2020 版《中国药典》收录为中药材“珠子参”的基源植物进行药用, 由于秀丽假人参在形态学和分布上与珠子参、羽叶三七具有相似性, 在当地 3 种药用植物常出现混淆及误用情况。因此, 亟需开展一定的研究工作来说明秀丽假人参能否作为珠子参的替代品或扩大来源。随着人参属叶绿体基因组数量的增加, 同属其他物种的叶绿体基因组也越来越容易组装, 这为 3 种人参属药用植物叶绿体基因组的组装提供了条件。

叶绿体是植物细胞中普遍存在的一种质体, 是植物进行光合作用的场所, 作为绿色植物所特有的细胞器, 本身拥有着一套比较完整的遗传体系即叶绿体基因组 (cp DNA), 能够独立指导叶绿体在植物体内进行复制等工作 (Camiolo et al., 2015)。植物物种之间的叶绿体基因组变异现象, 在研究植物群体遗传关系和物种鉴定方面均起着重要作用。近年来, 测序技术的不断更新和生物信息学工具的快速发展对叶绿体基因组学的研究起到了重要的推动作用。已有学者提出, 叶绿体基因组研究可为人参属物种鉴定、起源、进化、遗传多样性分析和资源保护与利用提供基础 (刘潮, 2022)。

当前, 关于药用植物珠子参的叶绿体基因组已有研究 (孙嘉苓, 2021), 但是对于秦巴山区 3 种人参属药用植物的叶绿体基因组特征比对和系统进化关系等尚不明确。该研究通过高通量测序技术 (Wang et al., 2018) 对秦巴山区 3 种人参属药用植物进行分析, 探讨珠子参、羽叶三七和秀丽假人参的叶绿体基因组特征、密码子偏好性以及亲缘关系, 以期为人参属中药材资源的合理开发利用和 3 种药用植物的分子鉴别工作提供科学依据。

1 材料与方法

1.1 实验材料

珠子参、羽叶三七和秀丽假人参 3 种药用植物的新鲜健康叶片均采自陕西省宝鸡市眉县红河谷 (107° 47' 29.3928" E、34° 0' 54.2844" N, 海拔 2 276.65 m)。叶片经纯水清洗擦干后用锡纸包裹放入液氮速冻, 之后于 -80 °C 冰箱保存。凭证标本存放在陕西中医药大学, 凭证编号为 202020203、202020204、202020205。

1.2 基因组DNA提取及测序

采用植物 DNA 提取试剂盒 (TIANGEN 公司, 北京) 提取总 DNA, 使用琼脂糖凝胶电泳和 Qubit3.0 检测 DNA 的质量。检测合格的总 DNA 利用 NovSeq6000 Illumina 高通量测序平台进行序列测定。通过 CLC Genomics Workbench 22.0 软件过滤原始数据 (raw data), 得到待组装序列 (clean data)。

1.3 叶绿体基因组组装和注释

以近缘物种竹节参 (Genebank ID: KP036469.1) 的叶绿体基因组序列为参考, 使用 Get Organelle (v1.6.4) 软件对获得的 Clean data 进行组装, 采用 CPGAVAS2 注释叶绿体基因

组。注释完成后,通过 BankIt 向 NCBI 提交 3 组序列,使用在线细胞器基因组绘图软件 Chloroplot(<https://irscope.shinyapps.io/Chloroplot/>)绘制 3 种药用植物叶绿体基因组物理图谱。

1.4 重复序列分析

利用 MISA 在线软件(<https://webblast.ipk-gatersleben.de/misa/>)对 3 组数据进行 SSR 位点分析,设置单核苷酸至六核苷酸的最少重复性数目为 10、5、4、3、3、3,两个 SSR 之间的距离 ≥ 100 bp。

1.5 叶绿体基因组比较分析

使用 mVISTA 在线工具(<https://genome.lbl.gov/vista/mvista/submit.shtml>)比较人参属 6 个物种的叶绿体基因组序列,输入文件为原始 fasta 格式的核苷酸序列文件和注释文件。

1.6 密码子偏好性及最优密码子分析

在序列长度 ≥ 300 bp 并剔除重复基因的基础上筛选珠子参、羽叶三七及秀丽假人参各自的编码蛋白,3 种药用植物分别获得 53 条满足条件的 CDS 序列。使用 Codon W 软件(<https://sourceforge.net/projects/codonw/>)和 CUSP 在线软件(<https://www.bioinformatics.nl/cgi-bin/emboss/cusp>)分析上述序列,得到密码子不同位置及总 GC 含量,同义密码子第 3 位 GC 含量(GC_{3s})、有效密码子数(ENC)及相对同义密码子使用度(RSCU)。同时,以 ENC 数值为基准,高表达基因的 ENC 值较低,低表达基因的 ENC 值较高,将人参属 3 种药用植物的密码子按照由大到小的顺序进行排列,从高低两端各取 10%的基因分别建立高、低表达基因库,计算两个基因库每个密码子的差值 $\Delta RSCU$,将同时符合高频即 $RSCU > 1$ 与高表达即 $\Delta RSCU \geq 0.08$ 的密码子作为最优密码子(Yu et al., 2023)。

1.7 密码子偏好性影响因素分析

1.7.1 中性绘图

以 GC_3 值为 X 轴, GC_{12} 值(GC_1 与 GC_2 平均值)为 Y 轴作图。图中回归曲线系数趋近 1 时,密码子偏好性主要受到碱基突变压力的影响,反之主要受到自然选择的影响(毛立彦等, 2022)。

1.7.2 ENC 分析

使用 R 语言绘制以 GC_{3s} 含量为 X 轴, ENC 实测值为 Y 轴的二维散点图,同时添加表征 ENC 期望值发展趋势的标准曲线,其中 $ENC \text{ 期望值} = 2 + GC_{3s} + 29/[GC_{3s}^2 + (1 - GC_{3s})^2]$, $ENC \text{ 比值} = (ENC \text{ 期望值} - ENC \text{ 实测值}) / ENC \text{ 期望值}$ (Zhang et al., 2022), ENC 比值代表基因和标准曲线之间的距离,值越小,表示该基因受自然选择的影响也越小。

1.7.3 PR2-plot 分析

统计密码子第 3 位上各碱基含量,以 $A_3/(A_3+T_3)$ 值为纵坐标、 $G_3/(G_3+C_3)$ 值为横坐标绘制散点图。若各编码基因在 4 个区域内均匀分布,则第 3 位嘌呤(A、G)和嘧啶(C、T)的使用频率相等,密码子偏好性主要受突变影响(王飞等, 2022)。

1.8 系统发育分析

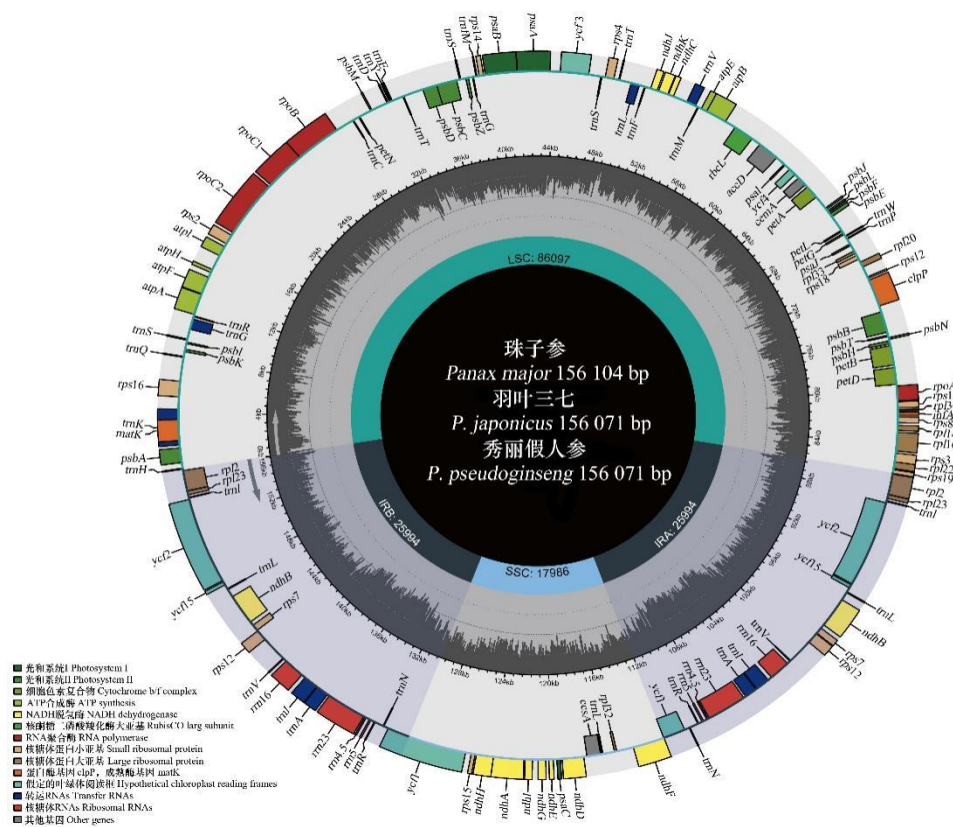
从 NCBI 下载 16 个五加科植物的叶绿体全基因组序列,与上述组装所得的珠子参(Genebank ID: OL543604.1)、羽叶三七(Genebank ID: OL543605.1)和秀丽假人参(Genebank ID: OL543606.1)共计 19 个物种,使用 MEGA 11 软件进行多重序列比对。基于邻接法(Neighbor-joining method, NJ)和双参数模型(Two-parameter model)构建系统发育树。通过 Bootstrap 法检验各分支的支持率,重复 1 000 次。

2 结果与分析

2.1 秦巴山区 3 种人参属叶绿体基因组结构与特征

珠子参、羽叶三七和秀丽假人参的叶绿体基因组均为双链环状分子,典型的四分体结构。叶绿体基因组大小为 156 071~156 104 bp,包括一对反向重复区(25 989~25 994 bp, inverted repeats, IRs),一个大单拷贝区(86 097~86 114 bp, large single-copy, LSC)和一个小单拷贝区(17 986~18 012 bp, small single-copy, SSC)(见图 1)。基因组的 GC 含量是判断物种亲缘关系的重要指标,3 种人参属药用植物叶绿体基因组的 GC 含量均为 38.10%, IRs、LSC 和 SSC 区的 GC 含量相同(见表 1),可见三者的叶绿体基因组大小存在一定差异,但差异不明显。

叶绿体基因组注释结果表明, 3 种药用植物的基因编码总数均为 133 个, 包含 88 个蛋白质编码基因(CDS)、37 个 tRNA 基因和 8 个 rRNA 基因(见表 2)。其中 7 个 tRNA 基因 (*trnL-CAA*、*trnV-GAC*、*trnI-GAU*、*trnA-UGC*、*trnR-ACG*、*trnN-GUU*、*trnI-CAU*) 和 4 个 rRNA 基因 (*rrn23S*、*rrn4.5S*、*rrn16S*、*rrn5S*) 以及 8 个 CDS 基因 (*rpl2*、*rpl23*、*rps12*、*rps7*、*ndhB*、*ycf1*、*ycf15*、*ycf2*) 在 IRs 区出现 1 次重复。此外 *trnK-UUU*、*trnG-UCC*、*trnL-UAA*、*trnV-UAC*、*trnI-GAU*、*trnA-UGC*、*trnI-GAU*、*rpl16*、*rpl2*、*rps16*、*rpoC1*、*atpF*、*ndhA*、*ndhB*、*petB*、*petD* 基因各含有 1 个内含子, *rps12*、*ycf3*、*ClpP* 基因各含有 2 个内含子。



基因的颜色差异表示不同的功能;叶绿体 GC 含量分布为内圈的灰色区域。
The color difference of genes indicate different functions; the chloroplast GC content is the gray area of the inner sphere.

图 1 人参属 3 种药用植物叶绿体基因组图谱
Fig. 1 Chloroplast genome map of three medicinal plants of *Panax*

表 1 人参属 3 种药用植物叶绿体基因组特征分析			
Table 1 Characterization of the chloroplast genomes from the three medicinal plants of <i>Panax</i>			
物种 Species	珠子参 <i>Panax major</i>	羽叶三七 <i>P. japonicus</i>	秀丽假人参 <i>P. pseudoginseng</i>
基因组大小 Genome size (bp)	156 104	156 071	156 071
大单拷贝区 LSC (bp)	86 114	86 097	86 097
小单拷贝区 SSC (bp)	18 012	17 986	17 986
反向重复区 IRA/B (bp)	25 989	25 994	25 994

GC含量	38.10	38.10	38.10
GC Content (%)			
总基因数量	133	133	133
Total number of genes			
tRNAs/rRNAs 数量	37/8	37/8	37/8
Number of tRNAs/rRNAs			
编码蛋白基因数量	88	88	88
Number of protein-coding genes			
大单拷贝区 GC 含量	36.30	36.30	36.30
GC content of LSC (%)			
小单拷贝区 GC 含量	32.20	32.20	32.20
GC content of SSC (%)			
反向重复区 GC 含量	43.00	43.00	43.00
GC content of IRs (%)			

表 2 人参属 3 种药用植物叶绿体基因组基因类型

Table 2 Chloroplast genome gene types of three medicinal plants of *Panax*

基因类别 Gene category	基因类群 Gene group	基因名称 Gene name	数量 Number
自我复制基因 Self replication	核糖体RNA基因 rRNA genes	<i>rrn16S, rrn23S, rrn4.5S, rrn16S, rrn5S, rrn5S, rrn4.5S, rrn23S</i>	8
	转运RNA基因 tRNA genes	<i>trnH-GUG, trnK-UUU*, trnQ-UUG, trnS-GCU, trnG-UCC*, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnG-GCC, trnM-CAU, trnS-GGA, trnT-UGU, trnL-UAA*, trnF-GAA, trnV-UAC*, trnM-CAU, trnW-CCA, trnP-UGG, trnI-CAU, trnL-CAA, trnV-GAC, trnI-GAU*, trnA-UGC*, trnR-ACG, trnN-GUU, trnL-UAG, trnN-GUU, trnR-ACG, trnA-UGC*, trnI-GAU*, trnV-GAC, trnL-CAA, trnI-CAU</i>	37
	核糖体大亚基 Large subunit of ribosome	<i>rpl14, rpl16*, rpl2*, rpl2*, rpl20, rpl22, rpl23, rpl23, rpl32, rpl33, rpl36</i>	11
	核糖体小亚基 Small subunit of ribosome	<i>rps11, rps12**, rps12**, rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7, rps7, rps8</i>	14
光合作用基因 Genes for photosynthesis	RNA聚合酶亚基 RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>	4
	ATP合成酶 ATP synthase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>	6
	光合系统 II 基因 Photosystem II gene	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ,</i>	15
	NADH脱氢酶 NADH dehydrogenase	<i>ndhA*, ndhB*, ndhB*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	12
	细胞色素复合物 Cytochrome complex	<i>petA, petB*, petD*, petG, petL, petN</i>	6
	光合系统 I 基因 Photosystem I gene	<i>psaA, psaB, psaC, psaI, psaJ, ycf3**, ycf4</i>	7
	二磷酸核酮糖羧化酶大亚基 Large subunit of Rubisco	<i>rbcL</i>	1
	乙酰辅酶A羧化酶亚基 Acetyl-CoA carboxylase beta subunit	<i>accD</i>	1
	C型细胞色素合成 C-type cytochrome synthesis	<i>ccsA</i>	1
	囊膜蛋白基因 Envelop membrane protein	<i>cemA</i>	1
其他基因 Other genes	ATP依赖蛋白酶亚基P基因 ATP-dependent protease subunit P	<i>ClpP**</i>	1
	翻译起始因子 Translation initiation factor	<i>infA</i>	1
	成熟酶K基因 matK	<i>matK</i>	1

未知功能基因	Maturase K gene		
Genes of unknown function	假定叶绿体阅读框		
	Hypothetical chloroplast reading frames	<i>ycf1, ycf1, ycf15, ycf15, ycf2, ycf2</i>	6

注: *. 一个内含子; **. 两个内含子。
Note: *. One intron; **. Two introns.

2.2 简单重复序列（SSR）分析

珠子参、羽叶三七以及秀丽假人参叶绿体基因组中各检测到 40、38 和 38 个 SSR（见图 2），三者的单核苷酸重复基元均为最多，分别是 22、17 和 17 个；四核苷酸重复基元数目居中,均为 8 个；二核苷酸重复基元分别是 4、8 和 8 个；三核苷酸和五核苷酸重复基元均为 3 个、2 个和 2 个；珠子参中六核苷酸重复基元未检测到，但在羽叶三七以及秀丽假人参中分别检测到 1 个。3 种药用植物中检测到的 SSR 大部分以 A/T、AT/AT、AAT/ATT、AAAT/ATTT、AATT/AATT、AAAAT/ATTTT 为重复单元，各自的占比分别为 62.50%、76.32%、76.32%，表明 3 种药用植物中的叶绿体 SSR 偏好使用 A/T 碱基。另发现仅珠子参中含有 C/G 类型和 AAAAT/ATTTT 类型的 SSR 位点，与其他两个物种明显不同。

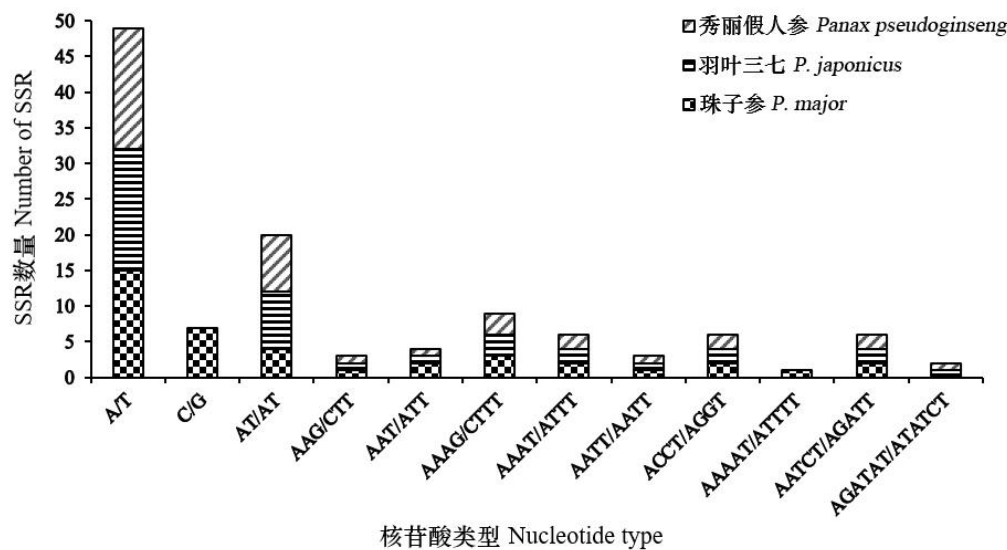
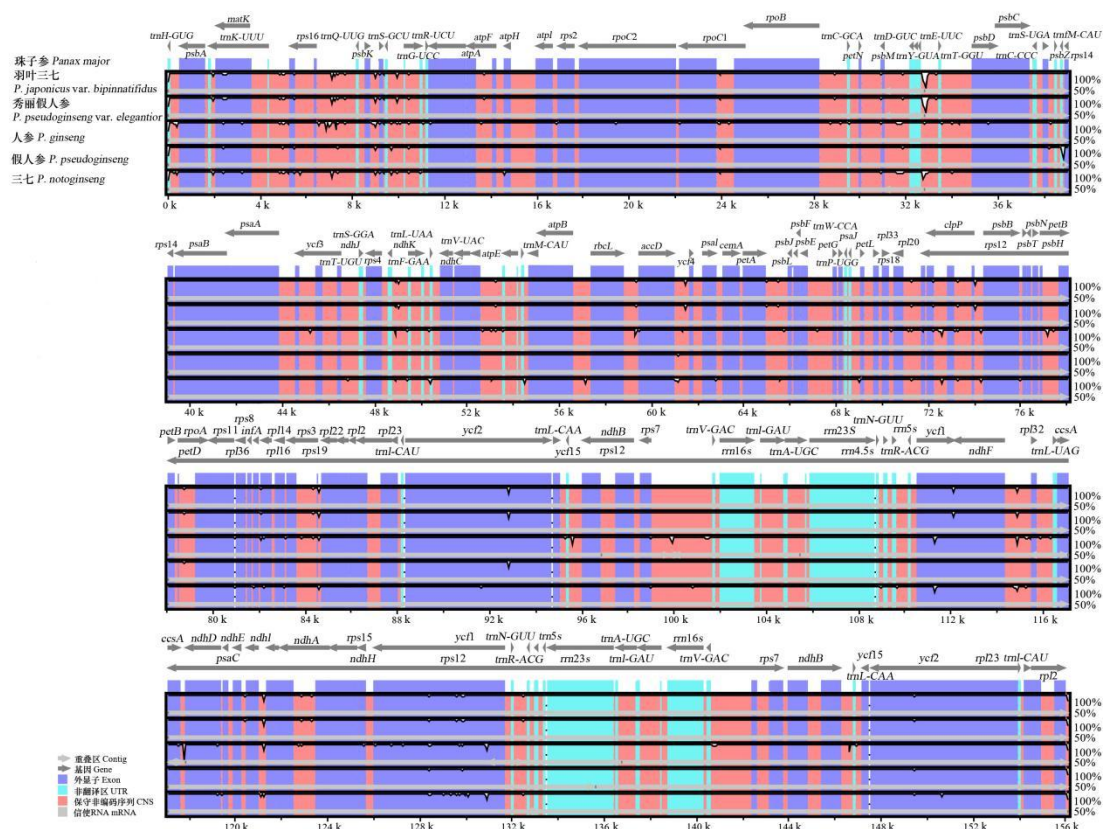


图 2 人参属 3 种药用植物的核苷酸类型
Fig. 2 Nucleotide types of three medicinal plants of *Panax*

2.3 叶绿体基因组的比较分析

以珠子参叶绿体基因组为参考序列得到全序列比对结果（见图 3）。人参属 6 个物种具有高度的相似性，分化程度较低，其中非编码区的序列差异明显大于编码区，IR 区相比于 SC 区更保守，说明人参属内的保守性、稳定性较高。6 个物种存在较大序列差异的区域为 *trnH-GUG~trnR-UCU*、*rpoC1~trnT-GGU*、*ndhF~trnL-UAG*、*rps12*、*rpl22* 和 *ycf1*。同时与珠子参相比，秀丽假人参与羽叶三七的序列相似度更高。



灰色箭头表示基因方向；纵轴表示相似性百分比（50%~100%）。
Grey arrows indicate gene orientation; vertical axis indicates percentage of similarity (50%-100%).

图 3 人参属叶绿体基因组比较分析

Fig. 3 Comparative analysis of chloroplast genomes of *Panax*

2.4 密码子偏好性分析

2.4.1 人参属 3 种药用植物密码子基本特征

3 种人参属药用植物密码子不同位置的 GC 含量存在差异，均呈现出 $GC_1 > GC_2 > GC_3$ 的趋势（见表 3）， GC_3 含量为 30.59%、29.77% 和 29.77%，表明密码子第 3 位偏好于 A 或者 U 结尾。有效密码子数（ENC）的取值范围为 20~61，该值越接近 20，密码子偏性越强。3 种药用植物的 ENC 值分别为 49.30、47.89 和 49.23，均明显 > 45 ，表明三者的叶绿体密码子使用偏好性较弱。

表 3 人参属 3 种药用植物叶绿体密码子基本特征

Table 3 Basic chloroplast codon characteristics of three medicinal plants of *Panax*

物种名 Species name	GC ₁ (%)	GC ₂ (%)	GC ₃ (%)	GC _{all} (%)	ENC
珠子参 <i>Panax major</i>	45.71	39.12	30.59	38.47	49.30
羽叶三七 <i>P. japonicus</i>	47.27	39.63	29.77	38.89	47.89
秀丽假人参 <i>P. pseudoginseng</i>	47.27	39.63	29.77	38.89	49.23

2.4.2 最优密码子的确认

3 种人参属药用植物共有 59 个同义密码子（见表 4），其中 $RSCU > 1$ 的高频密码子分别为 29、28 和 28 个，且均以 A/U 结尾。3 种药用植物中 $\Delta RSCU \geq 0.08$ 的高表达优越密码子数分别为 23、23 和 24（见表 5）。基于以上数据得到 3 种药用植物的最优密码子并做交集处理，得到人参属 3 种药用植物共有 13 个最优密码子：GAA、GUA、UGU、

AAA、AUU、CAA、CCU、UUA、GUU、UCU、ACU、GCU、CGU，其中 5 个以 A 结尾，其余均以 U 结尾。

表 4 人参属 3 种药用植物各氨基酸 RSCU 值

Table 4 RSCU values of amino acids in three medicinal plants of *Panax*

珠子参 <i>Panax major</i>			羽叶三七 <i>P. japonicus</i>			秀丽假人参 <i>P. pseudoginseng</i>		
氨基酸 Amino acid	密码子 Codon	RSCU	氨基酸 Amino acid	密码子 Codon	RSCU	氨基酸 Amino acid	密码子 Codon	RSCU
Phe	UUU	<u>1.31</u>	Phe	UUU	<u>1.31</u>	Phe	UUU	<u>1.31</u>
	UUC	0.69		UUC	0.69		UUC	0.69
Leu	UUA	<u>1.82</u>	Leu	UUA	<u>1.92</u>	Leu	UUA	<u>1.81</u>
	UUG	<u>1.27</u>		UUG	<u>1.24</u>		UUG	<u>1.27</u>
	CUU	<u>1.23</u>		CUU	<u>1.25</u>		CUU	<u>1.21</u>
	CUC	0.45		CUC	0.39		CUC	0.46
	CUA	0.84		CUA	0.81		CUA	0.84
Ile	CUG	0.4	Ile	CUG	0.39	Ile	CUG	0.42
	AUU	<u>1.4</u>		AUU	<u>1.43</u>		AUU	<u>1.39</u>
	AUC	0.64		AUC	0.62		AUC	0.65
	AUA	0.96		AUA	0.94		AUA	0.96
Val	GUU	<u>1.46</u>	Val	GUU	<u>1.48</u>	Val	GUU	<u>1.43</u>
	GUC	0.52		GUC	0.49		GUC	0.54
	GUA	<u>1.41</u>		GUA	<u>1.45</u>		GUA	<u>1.42</u>
	GUG	0.61		GUG	0.59		GUG	0.61
Ser	UCU	<u>1.65</u>	Ser	UCU	<u>1.71</u>	Ser	UCU	<u>1.64</u>
	UCC	0.95		UCC	0.96		UCC	0.95
	UCG	0.63		UCG	0.59		UCG	0.64
Pro	CCU	<u>1.46</u>	Pro	CCU	<u>1.53</u>	Pro	CCU	<u>1.48</u>
	CCC	0.75		CCC	0.77		CCC	0.75
	CCA	<u>1.16</u>		CCA	<u>1.08</u>		CCA	<u>1.13</u>
	CCG	0.63		CCG	0.63		CCG	0.64
Thr	ACU	<u>1.52</u>	Thr	ACU	<u>1.59</u>	Thr	ACU	<u>1.52</u>
	ACC	0.76		ACC	0.75		ACC	0.77
	ACA	<u>1.24</u>		ACA	<u>1.2</u>		ACA	<u>1.22</u>
	ACG	0.49		ACG	0.46		ACG	0.49
Ala	GCU	<u>1.75</u>	Ala	GCU	<u>1.77</u>	Ala	GCU	<u>1.75</u>
	GCC	0.65		GCC	0.63		GCC	0.63
	GCA	<u>1.14</u>		GCA	<u>1.16</u>		GCA	<u>1.15</u>
	GCG	0.46		GCG	0.44		GCG	0.47
Tyr	UAU	<u>1.55</u>	Tyr	UAU	<u>1.59</u>	Tyr	UAU	<u>1.53</u>
	UAC	0.45		UAC	0.41		UAC	0.47
His	CAU	<u>1.54</u>	His	CAU	<u>1.54</u>	His	CAU	<u>1.54</u>
	CAC	0.46		CAC	0.46		CAC	0.46
Gln	CAA	<u>1.48</u>	Gln	<u>CAA</u>	<u>1.5</u>	Gln	CAA	<u>1.48</u>
	CAG	0.52		CAG	0.5		CAG	0.52
Asn	AAU	<u>1.48</u>	Asn	AAU	<u>1.52</u>	Asn	AAU	<u>1.47</u>
	AAC	0.52		AAC	0.48		AAC	0.53
Lys	AAA	<u>1.44</u>	Lys	AAA	<u>1.48</u>	Lys	AAA	<u>1.42</u>
	AAG	0.56		AAG	0.52		AAG	0.58
Asp	GAU	<u>1.56</u>	Asp	GAU	<u>1.58</u>	Asp	GAU	<u>1.55</u>
	GAC	0.44		GAC	0.42		GAC	0.45
Glu	GAA	<u>1.51</u>	Glu	GAA	<u>1.53</u>	Glu	GAA	<u>1.51</u>
	GAG	0.49		GAG	0.47		GAG	0.49
Cys	UGU	<u>1.4</u>	Cys	UGU	<u>1.45</u>	Cys	UGU	<u>1.38</u>
	UGC	0.6		UGC	0.55		UGC	0.62

TER	UGA	<u>1.01</u>	TER	UGA	0.8	TER	UGA	0.85
Arg	CGU	<u>1.17</u>	Arg	CGU	<u>1.3</u>	Arg	CGU	<u>1.18</u>
	CGC	0.33		CGC	0.35		CGC	0.34
	CGA	<u>1.27</u>		CGA	<u>1.38</u>		CGA	<u>1.26</u>
	CGG	0.51		CGG	0.47		CGG	0.45
Ser	AGU	<u>1.21</u>	Ser	AGU	<u>1.24</u>	Ser	AGU	<u>1.22</u>
	AGC	0.4		AGC	0.35		AGC	0.39
Arg	AGA	<u>1.94</u>	Arg	AGA	<u>1.81</u>	Arg	AGA	<u>1.95</u>
	AGG	0.78		AGG	0.68		AGG	0.82
Gly	GGU	<u>1.3</u>	Gly	GGU	<u>1.34</u>	Gly	GGU	<u>1.29</u>
	GGC	0.44		GGC	0.45		GGC	0.45
	GGA	<u>1.5</u>		GGA	<u>1.51</u>		GGA	<u>1.51</u>
	GGG	0.76		GGG	0.71		GGG	0.75

注：下划线表示偏好密码子。

Note: Underline indicates preferred codons.

表 5 人参属 3 种药用植物叶绿体基因组最优密码子的确定

Table 5 Determination of optimal coons in the chloroplast gnome of three medicinal plants of *Panax*

氨基酸 Amino acid	密码子 Codon	珠子参 <i>Panax major</i>	羽叶三七 <i>P. japonicus</i>	秀丽假人参 <i>P. pseudoginseng</i>	氨基酸 Amino acid	密码子 Codon	珠子参 <i>P. major</i>	羽叶三七 <i>P. japonicus</i>	秀丽假人参 <i>P. pseudoginseng</i>
Phe	UUU	-	*	-	Tyr	UAU	*	-	*
	UUC	**	-	-		UAC	-	-	-
Leu	UUA	**	***	***	His	CAU	-	-	-
	UUG	-	-	-		CAC	-	-	-
	CUU	*	-	*	Gln	CAA	**	**	**
	CUU	-	-	-		CAG	-	-	-
	CUC	-	-	-	Asn	AAU	-	-	-
	CUA	*	*	*		AAC	*	-	*
	CUG	-	-	-	Lys	AAA	**	**	**
Ile	AUU	**	*	**		AAG	-	-	-
	AUC	-	-	-	Asp	GAU	*	-	*
	AUA	-	-	-		GAC	-	*	-
Val	GUU	**	***	***	Glu	GAA	*	*	*
	GUC	-	-	-		GAG	-	-	-
	GUA	*	***	*	Ala	GCU	***	**	***
	GUG	-	-	-		GCC	-	-	-
Ser	UCU	***	**	***	Arg	GCA	*	-	*
	UCC	*	-	-		GCG	-	-	-
	UCG	-	-	-		CGU	**	***	***
Pro	CCU	***	**	**		CGC	-	-	-
	CCC	-	*	-		CGA	-	**	**
	CCA	-	-	-		CGG	-	-	-
	CCG	-	-	-	Ser	AGU	-	**	*

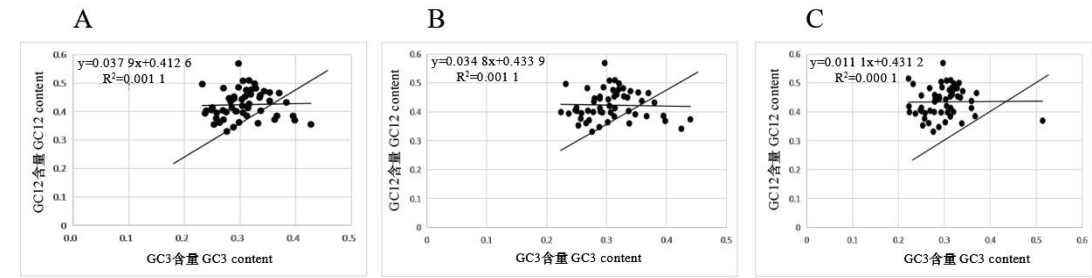
Thr	ACU	***	***	***	Arg	AGC	-	-	-
	ACC	-	*	-		AGA	-	-	-
	ACA	-	-	-		AGG	-	**	-
Cys	ACG	-	-	-	Gly	GGU	-	***	***
	UGU	**	***	*		GGC	-	-	*
	UGC	-	-	-		GGA	-	-	-
TER	UGA	-	-	-	Trp	GGG	-	-	-
	UAA	***	***	**		UGG	*	-	-

注：*. ΔRSCU≥0.08；**. ΔRSCU≥0.3；***. ΔRSCU≥0.8。
 Note:*. ΔRSCU≥0.08；**. ΔRSCU≥0.3；***. ΔRSCU≥0.8.

2.5 密码子偏好性影响因素分析

2.5.1 中性绘图分析

人参属 3 种药用植物的基因大范围地分布在对角线的上方（图 4），以 GC₁₂ 和 GC₃ 拟合的回归曲线斜率在 0.011 1~0.037 9，R²>0，表明人参属 3 种药用植物的叶绿体基因组密码子第 1、2 位碱基突变模式不同于第 3 位，密码子使用偏好性受到自然选择的影响要远大于碱基突变。



A. 珠子参；B. 羽叶三七；C. 秀丽假人参。下同。
 A. *panax major*; B. *Panax japonicus*; C. *Panax pseudoginseng*. The same below.

图 4 人参属 3 种药用植物密码子中性绘图分析

Fig. 4 Codon neutral mapping analysis of three medicinal plants of *Panax*

2.5.2 ENC 绘图分析

人参属 3 种药用植物的叶绿体基因组所包含的各个基因分布范围趋于标准曲线的下方，仅有少数在标准曲线上（图 5），表明自然选择是影响 3 种药用植物密码子偏好性的主要因素。

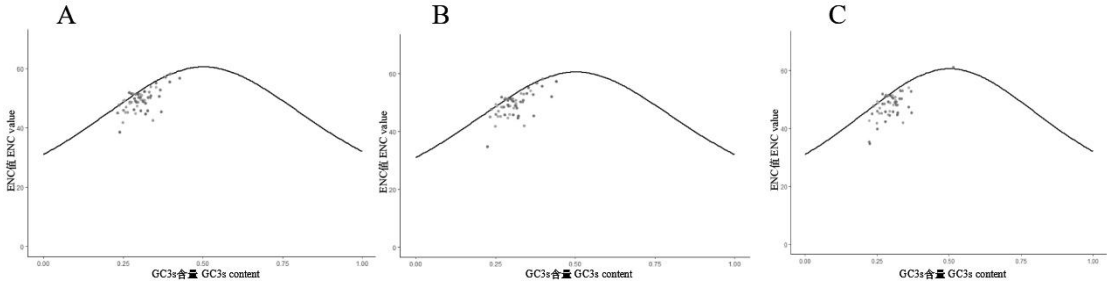


图 5 人参属 3 种药用植物 ENC 绘图分析

Fig. 5 ENC drawing analysis of three medicinal plants of *Panax*

2.5.3 PR2-plot 绘图分析

人参属 3 种药用植物叶绿体基因的分布并不均匀,基因多集中出现在右下方区域(图 6)。3 种药用植物在使用碱基的频率方面, $G>C$, $T>A$, 密码子第 3 位的嘧啶使用频率高于嘌呤, 反映出 3 种药用植物的密码子偏好性同时受到突变和选择压力的影响。

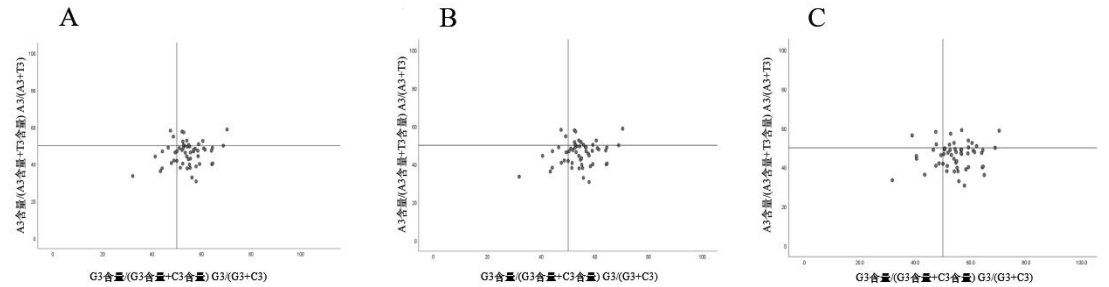


图 6 人参属 3 种药用植物 PR2-plot 绘图分析
Fig. 6 PR2-plot analysis of three medicinal plants of three medicinal plants of *Panax*

2.6 系统发育分析

为了明确秦巴山区 3 种人参属植物之间的亲缘关系, 选取 19 种五加科植物的叶绿体基因组序列进行系统发育关系分析。结果见图 7 (节点上的数字表示各分支的支持率, 支持率为 100%的不显示), 人参属、刺通草属、常春藤属、南鹅掌柴属、通脱木属的植物聚为一大类, 五加属植物被聚为另一大类, 常春藤属的尼泊尔常春藤 (*Hedera nepalensis*) 被单独归为一类。其中人参属与其他属之间呈现出较为明确的发育关系; 同时观察秀丽假人参与珠子参、羽叶三七的相对位置可得到, 秀丽假人参与羽叶三七亲缘关系最近, 其次是珠子参、人参。

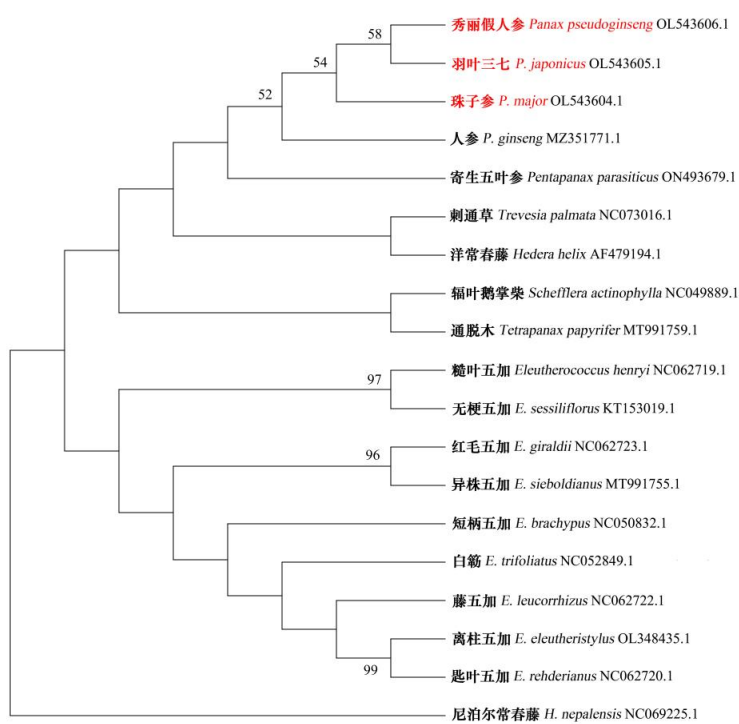


图 7 基于叶绿体基因组构建五加科 19 种植物的 NJ 系统发育树
Fig. 7 Construction of NJ phylogenetic tree of 19 Araliaceae species based on chloroplast genome

3 讨论与结论

该研究对秦巴山区 3 种人参属药用植物叶绿体基因组进行了测序和全序列比对分析。珠子参、羽叶三七和秀丽假人参叶绿体基因组序列全长 156 071~156 104 bp, 大小差异呈现为 $SSC(26\text{ bp}) > LSC(17\text{ bp}) > IRs(5\text{ bp})$, 其中 SSC 区的差异最大, 与岳杰 (2021)

研究一致。3 种药用植物叶绿体基因组总 GC 含量均为 38.10%，同时 IR 区与 SC 区相比 GC 含量最高且较稳定，结构高度保守，与大多数被子植物相似（Zhang et al., 2012）。物种之间亲缘关系越近，密码子使用偏好性越类似（Sahoo et al., 2019）。3 种人参属药用植物叶绿体基因组的密码子偏好以 A/U 结尾且主要受到自然选择的影响，与其他人参属物种密码子偏好性一致（石岩硕和赵永星，2022）。同时表明生活在相似环境中的同属物种因在进化过程中可能会共享类似的密码子使用偏好，以适应共同的生存压力。

植物叶绿体基因组序列中的突变“热点”区域可为物种鉴定提供足够的遗传信息（Li et al., 2018），根据 mVISTA 分析得到的 6 个突变基因片段与植物通用 DNA 条形码（*rbcL*、*matK* 和 *ITS* 等）（Antil et al., 2023）相比表现出更高的变异性，这些片段可用于人参属物种鉴定和系统发育研究。SSR 是一种广泛用于检测植物物种在特定位点和全基因组水平上遗传多态性的工具（Taheri et al., 2018; Ping et al., 2021）。该研究中羽叶三七和秀丽假人参具有相同数量及类型的核苷酸，表明二者之间更近的进化关系。3 种药用植物的核苷酸类型均以单核苷酸重复居多，其次为四核苷酸，且碱基偏好使用 A/T，这与人参属的三七、西洋参等相类似（卢志宏等，2022）。Niu 等（2017）研究表明，A/T 碱基在突变过程中消耗的能量比 G/C 少，这种优势就导致了人参属 3 种药用植物 SSR 和密码子中 A/T 碱基偏好性的结果。同时发现珠子参的 SSR 位点类型与羽叶三七和秀丽假人参相比多出 2 种，结合 3 种人参属药用植物叶绿体基因组结构大小、基因组成及密码子 GC₃ 含量得到，羽叶三七和秀丽假人参始终保持一致并与珠子参区别开来，因此可从分子水平上开发与其他两物种相区分的珠子参专用分子标记。

通过系统发育分析，可以更好地了解生物体的进化史。《中国植物志》依据 3 种药用植物地上部分的形态特征将其归类为假人参（*Panax pseudoginseng*）的 3 个变种沿用至今。根据 *Flora of China* 的分类，秀丽假人参和羽叶三七同为疙瘩七（*Panax bipinnatifidus*）的变种，但该分类依据未见报道。从 NJ 建树所得的结果来看，秀丽假人参和羽叶三七聚在一支同珠子参互为姊妹群，表明秀丽假人参和羽叶三七之间较近的亲缘关系，该结果可为 *Flora of China* 中的分类提供一定科学依据。因此叶绿体基因组用于阐明人参属物种的系统发育关系是非常有效的，但秀丽假人参能否作为药典规定珠子参基源植物的参考来源还需结合有效成分含量、临床应用效果等多方面指标综合考量。该研究利用生物信息学技术分析了秦巴山区 3 种人参属药用植物的叶绿体基因组特征，得到秀丽假人参和珠子参基源植物之间存在近亲关系，因此有可能在其中发现与珠子参相似或相同的药用成分。这为寻找新的药用植物资源、研发新药以及合理利用珠子参等中药材资源提供了理论基础。

参考文献

- ANTIL S, ABRAHAM JS, SRIPOORNA S, et al., 2023. DNA barcoding, an effective tool for species identification: a review[J]. Mol Biol Rep, 50(1):761-775.
- CAMIOLO S, MELITO S, PORCEDDU A, 2015. New insights into the interplay between codon bias determinants in plants[J]. DNA Res, 22:461-470.
- Editorial Board of Flora of China, Chinese Academy of Sciences, 2013. Flora of China: Vol.13 [M]. Beijing: Science Press:489-491. [中国科学院《Flora of China》编委会, 2013. Flora of China: 13卷[M]. 北京: 科学出版社: 489-491.]
- Editorial board of Flora of China, Chinese Academy of Sciences, 1978. Flora of China Vol.54 [M]. Beijing: Science Press: 181.[中国科学院《中国植物志》编委会, 1978.中国植物志54卷[M]. 北京: 科学出版社: 179.]
- LI W, LIU Y, YANG Y, et al., 2018. Interspecific chloroplast genome sequence diversity and genomic resources in *Diospyros* [J]. BMC Plant Biol, 18(1): 210.
- LIU C, LI M, RENG YY, et al., 2022. Characteristics and evolution of *Panax* chloroplast genomes[J]. Fujian J Agric Sci, 37(7):886-896. [刘潮, 李敏, 任怡园, 等, 2022. 人参属植物叶绿体基因组特征及其进化的研究[J]. 福建农业学报, 37(7): 886-896.]
- LU ZH, TIAN WY, YANG CD, 2022. Chloroplast genome characteristics and codon usage bias analysis of *Panax* Linn[J/OL]. Mol Plant Breed: 1-24. [2023-11-23] <http://134.175.232.43:8085/kcms/detail/46.1068.S.20220720.1016.002.html>. [卢志宏, 田文勇, 杨传东, 2022. 人参属植物叶绿体基因组特征和密码子偏好性分析[J/OL]. 分子植物育种: 1-24. [2023-11-23] <http://134.175.232.43:8085/kcms/detail/46.1068.S.20220720.1016.002.html>.]
- MAO LY, HUANG QW, LONG LY, et al., 2022. Comparative analysis of codon usage bias in chloroplast genomes of seven *Nymphaea* species[J]. J NW F Univ, 37(2):98-107. [毛立彦, 黄秋伟, 龙凌云, 等, 2022. 7种睡莲属植物叶绿体基因组密码子偏好性分析[J]. 西北林学院学报, 37(2): 98-107.]
- NIU ZT, XUE QY, WANG H, et al., 2017. Mutational biases and GC-biased gene conversion affect GC content in the plastomes of *Dendrobium* genus [J]. Int J Mol Sci, 18(11): 2307.
- PING J, FENG P, LI J, et al., 2021. Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*[J]. Ecol Evol, 11(9): 4786-4802.
- SAHOO S, DASS S, RAKSHIT R, 2019. Codon usage pattern and predicted gene expression in *Arabidopsis thaliana*[J]. Gene, 721: 100012.
- SHI YS, ZHAO YX, 2022. Codon preference analysis of chloroplast genome of *Panax* L[J]. Mol Plant Breed, 20(19):6350-6361. [石岩硕, 赵永星, 2022. 人参属植物叶绿体基因组密码子偏好性分析[J]. 分子植物育种, 20(19):6350-6361.]
- SUN JL, 2021. Comparative genomics study of *Panax japonicus* var. *major*, a medicinal plant of the genus Ginseng[D]. Kunming: Kunming University of Science and Technology. [孙嘉苓, 2021. 人参属药用植物珠子参的比较基因组研究[D]. 昆明: 昆明理工大学.]
- TAHERI S, LEE ABDULLAH T, YUSOP MR, et al., 2018. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants[J]. Molecules, 23(2): 399.
- WANG A, WU H, ZHU X, et al., 2018. Species identification of *Conyza bonariensis* assisted by chloroplast genome sequencing. [J]. Front Genet, 11(9): 374.
- WANG F, ZHAO WZ, DONG ZH, et al., 2022. Analysis of chloroplast genome characteristics of *Prinsepia*[J]. Chin J Trop Crops, 43(09):1759-1770. [王飞, 赵文植, 董章宏, 等, 2022. 扁核木属植物叶绿体基因组特征分析[J]. 热带作物学报, 43(9): 1759-1770.]
- YU X, SONG Y R, ZHAO Z N, 2023. The complete chloroplast genome of *Elsholtzia fruticosa* (D. Don) Rehd. (Labiateae), an ornamental plant with high medicinal value[J]. Mitochondrial DNA Part B, 8(3): 336-341.
- YUE J, 2021. Structural and phylogenetic analyses of the complete chloroplast genomes of four species within the genus *panax* Linn[D]. Hanzhong: Shaanxi University of Technology. [岳杰, 2021. 四种人参属植物叶绿体全基因组结构及其系统发育分析[D]. 汉中: 陕西理工大学.]

- ZHANG Y, SHEN Z, MENG X, et al., 2022. Codon usage patterns across seven Rosales species[J]. BMC Plant Biol, 22(1):65.
- ZHANG T, FANG Y, WANG X, et al., 2012. The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes. [J]. PLoS ONE, 7(1): e30531.