

款冬叶绿体基因组序列特征及其系统发育分析

净易尧¹, 胡本祥^{1,2}, 陈晓颖¹, 姬海月^{1,2}, 罗瑶¹, 史嘉周¹, 王帮庆³, 张岗¹,
高静¹, 杨冰月¹, 彭亮^{1*}

(1. 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, “秦药”研发重点实验室, 陕西 西安 712046; 2. 陕西国际商贸学院, 陕西 西安 712046; 3. 汉中市科技资源统筹中心, 陕西 汉中 723000)

摘要: 款冬 *Tussilago farfara* L. 为菊科植物 Compositae 款冬属 *Tussilago* 多年生草本植物, 其干燥花蕾及叶具有良好的生物活性成分, 在我国及欧洲地区药用历史悠久。本文对款冬全叶绿体基因组特征、序列重复、结构变异和系统发育进行了研究, 利用 Illumination 技术对款冬叶绿体基因组进行测序后, 进一步通过组装、注释获得了完整的款冬叶绿体基因组, 随后开展了反向重复-大单拷贝/小单拷贝区的收缩和扩张分析、基因组序列变异等一系列的特征分析, 并利用从 NCBI 下载的 13 种同科植物序列构建了邻接系统发育树。结果表明, 款冬叶绿体基因组总 GC 含量为 37.4%, 长度为 150 300bp; 共注释得到 125 个基因, 其中包括 82 个蛋白编码基因、35 个 tRNA 和 8 个 rRNA; 共检测出 148 个简单重复序列 (simple sequence repeats, SSR) 位点, 相对同义密码子使用度显示, 在 64 种密码子中, 有 31 种密码子的使用度 > 1, 密码子偏好 A/U 结尾; 边界分析显示款冬在内的 7 种菊科植物叶绿体基因组均高度保守大单拷贝 (large single-copy, LSC) 区和一个小单拷贝 (small single-copy, SSC) 区序列变异高于反向重复 (inverted repeat, IR) 区; 系统发育分析中, 样品款冬与已报道的榆林产款冬叶绿体基因组支持率呈 100%, 聚为一小支, 并与橐吾属植物复序橐吾、全缘橐吾等构成一个单系分支。本研究获得了高质量的款冬叶绿体全基因组, 并对其基因组特征、密码子偏好性、简单重复序列 SSR 特性、SC/IR 边界、序列变异和系统发育等进行了全面的分析, 可为该药用植物的物种鉴定、遗传多样性分析及资源开发提供依据。

关键词: 款冬; 菊科植物; 叶绿体基因组; 序列分析; 系统发育

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2023)11-3439-10

Phylogenetic analysis of chloroplast genome of *Tussilago farfara* L.

JING Yi-yao¹, HU Ben-xiang^{1,2}, CHEN Xiao-ying¹, JI Hai-yue^{1,2}, LUO Yao¹, SHI Jia-zhou¹,
WANG Bang-qing³, ZHANG Gang¹, GAO Jing¹, YANG Bing-yue¹, PENG Liang^{1*}

(1. Key Laboratory for Research of Qin medicine of Shaanxi Administration of Traditional Chinese Medicine, Shaanxi Qinling Application Development and Engineering Center of Chinese Herbal Medicine, College of Pharmacy, Shaanxi University of Chinese Medicine, Xi'an 712046, China; 2. Shaanxi Institute of International Trade & Commerce, Xi'an 712046, China; 3. Hanzhong Science and Technology Coordination Center, Hanzhong 723000, China)

Abstract: *Tussilago farfara* L. is a perennial herb of *Tussilago* genus in the Compositae family. Its dried buds and leaves have good biological activities and have a long history of medicinal use in China and Europe. In this paper, we investigated the whole chloroplast genome characteristics, sequence duplication, structural variation and phylogeny of the *Tussilago farfara* L. After sequencing the *Tussilago farfara* L. chloroplast genome using Illumination technology, the complete *Tussilago farfara* L. chloroplast genome was further obtained by assembly

收稿日期: 2023-04-10; 修回日期: 2023-08-10.

基金项目: 国家自然科学基金资助项目 (82204578); 陕西省科技厅项目 (2021JQ-733); 公益性行业 (中医药) 科研专项经费项目 (201507002); 陕西中医药大学校级课题 (2020GP06); 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).

*通讯作者 E-mail: ppengliang@126.com

DOI: 10.16438/j.0513-4870.2023-0437

and annotation, followed by a series of inverted repeat-large single copy/small single copy region contraction and expansion analysis, genome sequence variation, etc. The sequences of 13 homologous plants downloaded from NCBI were used to construct a neighbor-joining phylogenetic tree. The results showed that the total GC content of the chloroplast genome was 37.4% and the length was 150 300 bp; 125 genes were annotated, including 82 protein-coding genes, 35 tRNAs and 8 rRNAs; 148 (simple sequence repeats, SSR) loci were detected, and the relative synonymous codon usage showed that 31 codons out of 64 codons had a usage of >1. In the phylogenetic analysis, the chloroplast genomes of the seven species of Asteraceae, including the Yulin *Tussilago farfara* L., were highly conserved, and the sequence variation of the (large single-copy, LSC) and (small single-copy, SSC) regions was higher than that of the (inverted repeat, IR) region. This is in general agreement with the reported phylogeny of Yulin *Tussilago farfara* L. In this study, we obtained a high quality chloroplast genome and analyzed its genome characteristics, codon preference, SSR characteristics, SC/IR boundary, sequence variation and phylogeny, which can provide a basis for species identification, genetic diversity analysis and resource development of this medicinal plant.

Key words: *Tussilago farfara* L.; Compositae; chloroplast genome; sequence analysis; phylogeny

款冬 *Tussilago farfara* L. 系菊科 Compositae 款冬属 *Tussilago* 多年生草本植物, 乃单属单种。款冬的花蕾和叶片具有止咳、润肺、化痰等作用, 在欧洲和我国传统中医药学中广泛用于治疗呼吸道疾病^[1]。2020版《中华人民共和国药典》规定, 款冬的干燥花蕾为其入药部位。款冬花临床多用其蜜炙品, 具有解毒的作用^[2]。迄今, 已从款冬花和叶中分离和鉴定出约 150 种化合物, 包括倍半萜类化合物、三萜类化合物、黄酮类化合物、酚酸、生物碱等^[3], 具有抗炎、抗氧化、抗癌等多种药理活性^[4], 临床多用于治疗咳嗽痰多、劳嗽咳血以及咳嗽等症^[5], 具有极高的药用和经济价值。

同时, 款冬作为一种蜜源植物, 广泛分布在中国各地的药用苗圃、山谷、湿地和森林中^[6]。在我国, 款冬野生资源主产于东北、华北、西北等地区; 国外地区则主要分布在印度、伊朗、巴基斯坦、俄罗斯、西欧等地区。目前, 款冬花商品多为栽培品, 甘肃、陕西、山西等地为主产区^[7]。据报道, 2020年甘肃省款冬花产量达 1 100 吨, 占全国总产量的 70%~80%^[8]。款冬花现代研究则多集中于质量评价^[8]、化学成分^[9]、药理作用^[10]等方面, 关于其种质资源及遗传信息的研究较少, 仅见 1 篇关于榆林地区款冬的叶绿体基因组报道, 尚缺乏其密码子偏好、简单重复序列 (simple sequence repeats, SSR)、边界分析和基因组序列变异等系统的叶绿体基因组分析。

叶绿体是植物和藻类中的细胞器, 在植物细胞功能中发挥着重要作用, 包括在光合作用、固碳和应激反应方面^[11]。而叶绿体基因组是一种双链圆形 DNA 分子, 具有保守的四分体结构, 包括一个大单拷贝区域 (large single-copy, LSC)、一个小单拷贝区域 (small single-copy, SSC) 和一对反向重复区域 (inverted

repeat, IR)^[12]。作为绿色植物特有的半自主型细胞器, 叶绿体拥有源于母系遗传的独立基因组。相比于核基因组, 叶绿体基因组结构稳定, 具有分子进化速率适宜、序列高度保守、基因密集度高等优点^[13], 可为高等植物的物种形成、种质资源的遗传多样性以及基因工程等研究提供了重要的理论依据。随着小基因组高通量测序技术的发展以及序列解析方法的开发, 目前已有药用篇蓄^[14]、毛重楼^[15]、桃儿七^[16]等多种植物的叶绿体基因组被先后报道, 为药用植物的分类、鉴定和资源开发提供了有效的技术手段。

基于此, 本研究用高通量测序技术, 以款冬花主产区甘肃省定西市陇西县的栽培种款冬为原材料, 对款冬叶绿体开展了全基因组测序、组装、注释及序列变异和结构特征解析; 同时, 选取菊科 13 种植物与款冬进行系统发育分析, 对其亲缘关系进行分析与比较, 以期对款冬后期的资源开发与利用、系统发育及物种鉴定等研究奠定理论基础。

材料与方法

材料 款冬样品采自甘肃省定西市陇西县, 经陕西中医药大学胡本祥教授鉴定为菊科款冬 *Tussilago farfara* L., 取款冬的新鲜叶片, 液氮速冻, 存于-80 °C 冰箱, 用于 DNA 提取。植物凭证样本保存于陕西中医药大学标本馆 (TF20211201)。

基因组 DNA 提取与测序 取新鲜款冬液氮冷冻, 研磨后存于-80 °C 环境。通过改良的 CTAB 方法分离 cpDNA^[17]。采用琼脂糖凝胶电泳检测 DNA 质量。本项目拟采用超声技术对全基因组进行处理, 通过末端修复、加 A 尾、加测序接头、纯化、PCR 扩增等方法^[18], 构建文库, 并通过 Illumina 高通量测序平台 HiSeq X Ten

测序^[19], 获得序列原始数据。数据分析之前, 对 Reads 进行质量控制。质控后得到的高质量 Clean Data, 以 FASTQ 格式提供。

叶绿体全基因组序列组拼接与注释 选取 SPAdes 进行拼接。处理完后的高质量 Clean Data 使用 SPAdes v3.11.1^[20] 拼接软件对优化序列进行多个 Kmer 参数的拼接, 长度设置为 107、117 和 127。后使用软件 Spades 对叶绿体的 Reads 进行组装。使用在线网站 CHLOROBOX^[21] 对组装好的叶绿体基因组进行注释, 并以已发表的款冬叶绿体基因组序列^[22] (GenBank 登录号: MK756018) 为参考。使用 OGDRAW 软件^[23] 绘制叶绿体全基因组图谱。最终注释的叶绿体基因组存放在 GenBank 中, 登录号为 OQ029303。

叶绿体基因组特征分析 为消除氨基酸组成对密码子使用的影响, 采用 MEGA11^[24] 分析了同义密码子使用量、相对同义密码子使用值、碱基组成和密码子含量的变化特征。由于重复序列的插入、重排或缺失都会影响叶绿体基因组的长度和顺序, 且对叶绿体基因组进化方面有很重要的作用, 利用 REPuter 软件^[25] (<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, 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%。串联重复序列利用 Tandem repeats finder 软件^[26] (<https://tandem.bu.edu/trf/trf.html>) 进行检测, 参数设置选择默认值。使用 SSRHunter 软件 (<http://www.biosoft.net>) 鉴定 cp 基因组中的简单序列重复序列^[27,28]。参数设置为单核苷酸至六核苷酸 8、5、4、3、3、3。SC/IR 边界使用 IRSCOPE^[29] 进行作图分析。通过 mVISTA^[30] (<https://genome.lbl.gov/vista/mvista/submit.shtml>) 做全基因组对比, 分析时勾选全局对比, 以已公布的另一款冬叶绿体基因组做参考, 对菊科 7 种叶绿体基因组序列做同源性对比分析。

系统发育分析 从 NCBI (<https://www.ncbi.nlm.nih.gov>) 下载菊科橐吾属植物复序橐吾 *Ligularia jaluensis* (NC 039383)、蹄叶橐吾 *Ligularia fischeri* (NC 039352)、无缨橐吾 *Ligularia biceps* (OK 448480)、狭苞橐吾 *Ligularia intermedia* (NC 039382)、全缘橐吾 *Ligularia mongolica* (MF 539932)、兔儿伞属植物兔儿伞 *Syneilesis aconitifolia* (NC 061374)、千里木属植物矮千里木 *Dendrosenecio keniensis* (NC 036832)、埃尔贡千里木 *Dendrosenecio elgonensis* (MK 756018)、款冬 *Tussilago farfara* L. (OQ029303)、已发表款冬属植物款冬 (MW760850)、向日葵属植物向日葵 *Helianthus annuus* (NC 007977)、银胶菊属植物银胶菊 *Parthenium argentatum* (GU 120098), 同时选取忍冬属植物忍冬 *Lonicera*

japonica (NC 026839) 为外类群, 共 13 个物种完整的叶绿体基因组序列, 利用 MAFFT version 7 软件进行序列多重比对, 输出注释好的文件, 检查所得结果并进行校验; 采用邻接法分析系统演化关系。用 MEGA11 软件生成系统发育树, 自展值^[31] 设为 1 000。

结果与分析

1 款冬的叶绿体基因测序

款冬叶绿体基因组图 (图 1) 显示, 款冬叶绿体基因组结构为典型的 4 分结构, 双链环状, 由一段大的单拷贝区和一段小的单拷贝区, 以及两段序列相同、方向相反的两个反向重复区 (IRa 和 IRb) 组成。款冬的叶绿体基因组全长为 150 300 bp, GC 值为 37.44%, LSC、SSC 和 IR 序列的长度分别为 82 503、18 111、24 847 bp (表 1)。其中 LSC、SSC 区域的 GC 含量分别为 35.6% 和 30.6%, 但 IR 区域的 GC 含量则为 43.0%。基因注释结果表明, 款冬叶绿体基因组共有 125 个基因得到注释, 有与植物光合作用相关的基因、与自我复制相关的基因, 以及一些功能未知的基因。其中包含有 35 个 tRNA、8 个 rRNA 以及 25 个核糖体蛋白大小亚基基因。部分基因存在拷贝, 如表 2 所示。其中存在 1 个内含子的基因有 19 个, 存在 2 个内含子的基因是 *clpP* 和 *ycf3*。

2 密码子偏好性分析

款冬共由 64 条蛋白编码, 总长为 71 701 bp, GC 含量为 38.39%, 共有 43 158 个密码子参与。除终止密码子外, 20 种氨基酸由其他密码子编码而来。其中亮氨酸使用最为频繁, 其次是异亮氨酸, 半胱氨酸使用次数最少, 其数量分别为 4 285 (9.93%), 3 763 (8.72%), 564 (1.31%)。密码子偏好性 (relative synonymous codon usage, RSCU) 分析结果表明, 在所示的 64 种密码子中, 有 31 种密码子的 RSCU > 1, 占总量的 65.37%, 其中 28 种以 A/U 结尾, 3 种以 G/C 结尾 (表 3)。

3 长重复序列分析

共检测到 32 个同时满足长度不小于 30 bp 且重复序列间相似度大于 90% 两个条件的散在重复序列, 包括 19 个回文重复和 13 个正向重复, 其长度范围为 30~50 bp。检测到 25 条长度范围为 9~27 bp 的串联重复序列。串联重复序列在 4 个区域均有分布, 其中, LSC 区有 12 条, 重复次数均为 2 次, SSC 区有 5 条, 重复次数依次为 2、1、3、2 和 2 次, 两个 IR 区各 4 条, 重复次数分别为 5、1、2 和 2 次。

4 SSR 结果分析

在款冬叶绿体基因组中鉴定到五种类型的核苷酸重复序列, 分别为单、二、三、四、五核苷酸重复序列, SSRs 位点共 148 个, 其中单、二、三、四、五核苷酸数量

Table 2 Annotation information of the chloroplast genomes of *Tussilago farfara* L. (2): Number of copies of multi-copy genes; *: One intron; **: Two introns

Gene group	Gene name	Number
tRNA	<i>trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnE-UUC*, trnF-GAA, trnG-GCC, trnH-GUG, trnI-GAU*, trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG, trnM-CAU(4), trnN-GUU(2), trnP-UGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-CGA*, trnS-GCU, trnS-GGA, trnS-UGA, trnT-UGU, trnV-GAC(2), trnW-CCA, trnY-GUA</i>	35
rRNA	<i>rrn16S(2), rrn23S(2), rrn4.5S(2), rrn5S(2)</i>	8
Small subunit of ribosome	<i>rps11, rps12**(2), rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7(2), rps8</i>	14
Large subunit of ribosome	<i>rpl14, rpl16*, rpl2*(2), rpl20, rpl22, rpl23(2), rpl32, rpl33, rpl36</i>	11
RNA polymerase	<i>rpoA*, rpoB, rpoC1*, rpoC2</i>	4
NADH dehydrogenase	<i>ndhA*, ndhB*(2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	12
Photosystem I	<i>psaA, psaB, psaC, psaJ</i>	4
Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbM, psbN, psbT, psbZ</i>	14
Cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>	6
ATP synthase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>	6
Large subunit of rubisco	<i>rbcL</i>	1
Maturase	<i>matK</i>	1
Protease	<i>clpP**</i>	1
Envelope membrane protein	<i>cemA</i>	1
Subunit of acetyl-carboxylase	<i>accD</i>	1
C-type cytochrome synthesis	<i>ccsA</i>	1
Translation initiation factor	<i>infA</i>	1
Proteins of unknown function	<i>yef2*(2), yef3**, yef4</i>	4
Total		125

Table 3 Codon usage in *Tussilago farfara* L.

Codon	Amino acid	Number	Relative synonymous codon usage	Codon	Amino acid	Number	Relative synonymous codon usage
GCA	Ala	685	1.12	CCA	Pro	522	1.21
GCC	Ala	398	0.65	CCC	Pro	297	0.69
GCG	Ala	282	0.46	CCG	Pro	281	0.65
GCU	Ala	1 070	1.76	CCU	Pro	632	1.46
UGC	Cys	186	0.66	CAA	Gln	1 109	1.50
UGU	Cys	378	1.34	CAG	Gln	369	0.50
GAC	Asp	326	0.43	AGA	Arg	828	1.88
GAU	Asp	1 174	1.57	AGG	Arg	315	0.71
GAA	Glu	1 483	1.52	CGA	Arg	566	1.28
GAG	Glu	475	0.49	CGC	Arg	157	0.36
UUC	Phe	892	0.71	CGG	Arg	207	0.47
UUU	Phe	1 631	1.29	CGU	Arg	573	1.30
GGA	Gly	1 140	1.51	AGC	Ser	249	0.43
GGC	Gly	372	0.49	AGU	Ser	654	1.14
GGG	Gly	547	0.72	UCA	Ser	732	1.28
GGU	Gly	967	1.28	UCC	Ser	507	0.88
CAC	His	296	0.57	UCG	Ser	325	0.57
CAU	His	746	1.43	UCU	Ser	973	1.70
AUA	Ile	1 141	0.91	ACA	Thr	662	1.23
AUC	Ile	811	0.65	ACC	Thr	418	0.78
AUU	Ile	1 811	1.44	ACG	Thr	214	0.40
AAA	Lys	1 537	1.47	ACU	Thr	862	1.60
AAG	Lys	553	0.53	GUA	Val	861	1.48
CUA	Leu	573	0.80	GUC	Val	305	0.53
CUC	Leu	309	0.43	GUG	Val	307	0.53
CUG	Leu	272	0.38	GUU	Val	847	1.46
CUU	Leu	915	1.28	UGG	Trp	805	1.00
UUA	Leu	1 298	1.82	UAC	Tyr	354	0.44
UUG	Leu	918	1.29	UAU	Tyr	1 261	1.56
AUG	Met	1 029	1.00	UAA	Stop*	312	1.01
AAC	Asn	489	0.54	UAG	Stop*	252	0.82
AAU	Asn	1 336	1.46	UGA	Stop*	362	1.17

Table 4 Types and amounts of SSRs in the chloroplast genomes of the *Tussilago farfara* L.

SSR type	Repeat type	Number	Proportion/%
Mono	A/T	123	83.11
	C/G	2	1.35
Di	AT/AT	8	5.41
Tri	AAG/CTT	1	0.68
	AAT/ATT	3	2.03
Tetra	ATC/ATG	1	0.68
	AAAG/CTTT	1	0.68
	AAAT/ATTT	3	2.03
	AACT/AGTT	1	0.68
	AATC/ATTG	1	0.68
	AATT/AATT	2	1.35
Penta	AGAT/ATCT	1	0.68
	AATTC/AATTG	1	0.68

trnH 基因之间, 其他植物的边界均处于 *rpl2* 与 *trnH* 基因之间。

以上分析结果可以看出, 菊科植物叶绿体基因组的 IR 边界区中存在一定的差异, 但总体来说, IR 区的收缩与扩张幅度较小, 叶绿体基因组较为保守。

6 款冬叶绿体基因组序列变异分析

以登录号为 MK756018 的榆林产款冬叶绿体基因

组作注释, 结合边界分析图, 选取样品款冬、矮千里木、埃尔贡千里木、兔儿伞、全缘橐吾、狭苞橐吾、无纓橐吾这 7 种植物使用 mVISTA 在线工具进行叶绿体基因组全序列对比分析 (图 3)。7 种菊科植物叶绿体基因组的基因区间组成较为一致, 差异性较小。从四大区段来看, LSC 区差异性最大, 变异程度最高, 而 IRa 区则差异性最小, 最为保守。从非基因编码区和基因编码区来看, 非基因编码区变异程度较高, 基因编码区较为保守, 但在 *ycf1*、*ycf2*、*ycf3*、*ndhF* 和 *ndhD* 等基因编码区变异程度较大, 存在显著差异。

7 系统发育分析

根据本研究中的款冬 (OQ29303) 序列, 选择已发表的菊科植物为内类群, 来自川续断目忍冬科忍冬属植物忍冬的叶绿体基因组作为外类群, 采用邻接法 (NJ法) 构建了 13 种植物的系统发育树。结果表明, 12 种植物互相之间支持率均达 100%, 与忍冬区分开来; 样品款冬花与已报道的榆林款冬聚为一支, 支持率 100%; 复序橐吾、蹄叶橐吾、无纓橐吾、狭苞橐吾、全缘橐吾与兔儿伞可聚为次级单系分支; 除复序橐吾、蹄叶橐吾和无纓橐吾这两个节点的支持率分别为 90% 和 83%, 其

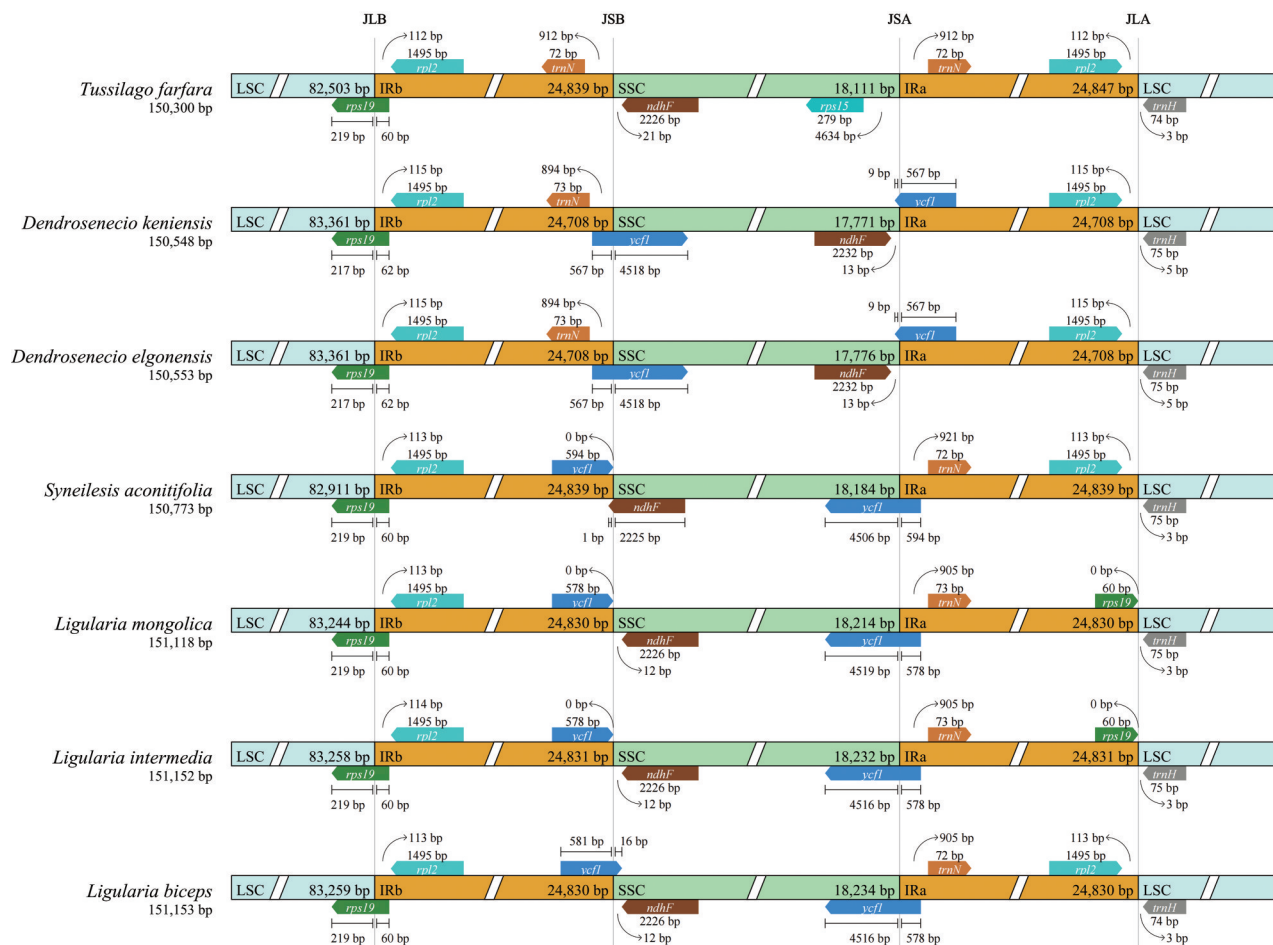


Figure 2 The chloroplast genome boundary analysis of seven plants from Compositae

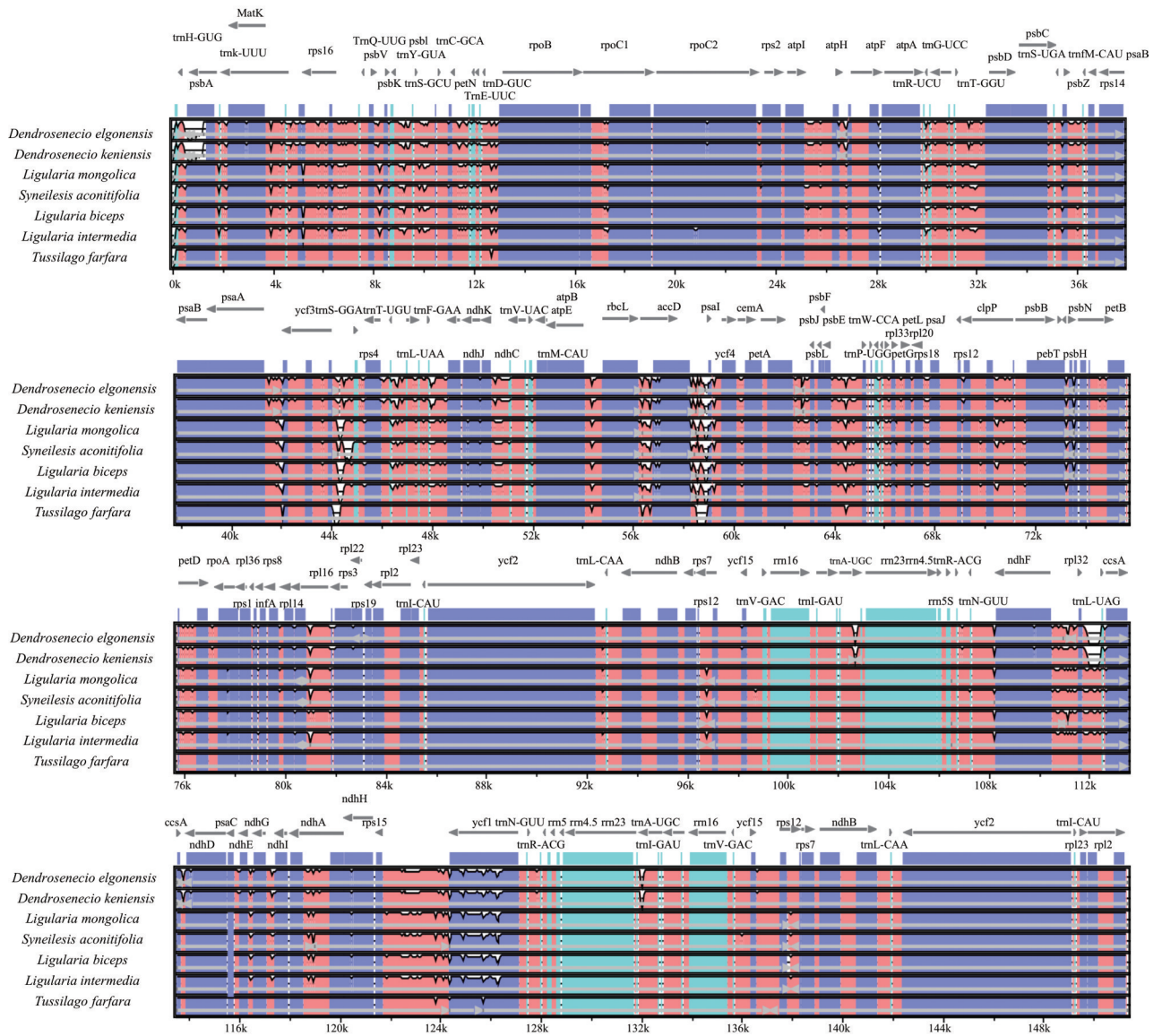


Figure 3 Global alignment analysis of *Tussilago farfara* L. chloroplast genomes

他节点支持率均为 100%; 款冬组与橐吾组构成姐妹类群; 矮千里木、埃尔贡千里木聚为一支, 向日葵与银胶菊聚为一支, 两支再聚为一支, 与所测样品款冬亲缘关系支持率均为 100% (图 4)。

讨论

款冬花药用历史悠久, 始载于《神农本草经》, 是我国常用的中药材之一, 具有润肺下气、止咳化痰的功效, 临床主要用于治疗咳喘痰多、新久咳嗽等症状^[32]。在新型冠状病毒防疫期间, 作为中药汤剂清肺排毒汤中的主要药材, 款冬花对肺炎患者起到了至关重要的作用^[33]。现代研究^[4]发现, 款冬花不仅有止咳平喘的作用, 更有抗炎、抗过敏等药理活性, 遗传信息研究方面虽报道有一个榆林产款冬的叶绿体基因组, 但并未

说明是野生还是栽培品种, 同时也未对基因组密码子偏好、SSR、边界分析和基因组序列变异等系统的叶绿体基因组分析。本文以款冬花商品主产区甘肃省定西市陇西县的栽培种款冬为研究对象, 对款冬叶绿体基因组进行测序、组装和注释, 并对其结构、GC 含量等进行了分析, 获得了款冬的全叶绿体基因组序列。结果表明, 样品款冬叶绿体基因组为典型的 4 分环状结构, 保守性好, 长度为 150 300 bp, 包含 LSC 区域 (82 503 bp), SSC 区域 (18 111 bp), 一对 IR 序列区域 (24 847 bp); 基因组 GC 含量为 37.4%。已发表的榆林产款冬叶绿体基因组长度则为 151 325 bp; LSC 区域 (83 370 bp), SSC 区域 (18 273 bp), 一对 IR 序列区域 (24 841 bp), 基因组总 GC 含量为 37.4%^[22], 从所测数据来看, 两者差异小, 结构相似。样品款冬叶绿体基因组 LSC、SSC 和

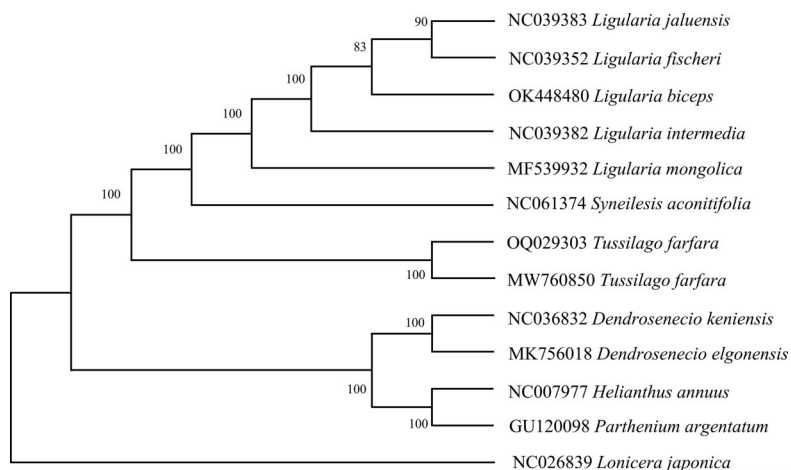


Figure 4 Neighbor-joining (NJ) phylogenetic tree was constructed based on the whole chloroplast genome sequence

IR 区分别占基因组长度的 35.6%、30.6% 和 43.0%; 共编码到 125 个基因, 其中含有一个内含子的有 19 个, 含有两个内含子的有 2 个, 分别为 *clpP* 和 *ycf3* 基因, 而在其他植物叶绿体基因组的研究中也发现 *ycf3* 和 *clpP* 基因含有多个内含子, 可能是因为 *ycf3* 基因需要在光系统 II 复合体中积累, 而额外的内含子可能有利于光合作用的进行, *clpP* 在蛋白质代谢调控过程中发挥着巨大的作用, 额外的内含子可能有利于蛋白质代谢调控^[34-36]。款冬叶绿体基因组大小、结构和组成与已发表的菊科植物结构呈高度相似, 证明菊科植物在进化过程中保守发育^[37]。

密码子是生物体中连接氨基酸、蛋白质和遗传物质的重要核心元素, 其偏好使用的研究为蛋白质表达及其相应功能的研究提供了可靠的信息^[38]。从款冬叶绿体基因组密码子偏好性来看, 其偏好性最大的密码子为 AGA, 64 种密码子中有 31 种密码子 RSCU > 1, 且 RSCU > 1 的密码子多数以 A/U 结尾, 类似结果在紫花地丁等植物中均有体现, 这可能与叶绿体基因组 AT 含量高的原因有关^[39]。长重复序列是导致叶绿体基因组结构变异的重要因素, 款冬的长重复序列中有 19 个回文重复和 13 个正向重复、25 个串联重复序列为款冬的分子鉴定、物种演化、群体遗传学研究等方面提供了候选分子标记^[40]简单序列重复, 也称为微卫星, 主要存在于基因外部和基因非编码区, 目前多用于群体遗传学研究的分子标记中^[41]。款冬叶绿体基因组中, 检测出的 SSRs 共有 148 个, 其中以单核苷酸重复居多, 为 125 个; 其次为四核苷酸, 为 9 个; 五核苷酸最少, 只有 1 个。这些检测到的 SSRs 可为款冬后续在遗传学研究中提供分子标记, 作为其物种和产地鉴别的有力工具^[42]。此外, 对 7 种菊科植物的叶绿体基因组的 IR/SC 边界进行分析发现, SSC/IR 边界区域差异性较大, LSC/IR 区域

差异变化小, 但总体来说, 整个基因组仍然较为保守。

为了进一步基于款冬叶绿体基因组获得与近源种的进化关系, 探讨分类地位, 本文选取了 12 种菊科植物与款冬 (OQ029303) 叶绿体基因组进行系统发育分析并构建了进化树。结果显示, 样品款冬与已报道的榆林产款冬以 100% 支持率聚为一支, 均与橐吾属植物聚为一个姐妹分支; 样品款冬除复序橐吾、蹄叶橐吾和无纒橐吾这两个节点的支持率为 90%、83% 外, 其他节点均为 100%, 证明样品款冬与所选菊科植物亲缘关系好。所构建的系统发育树阐明了款冬与菊科其他植物的系统发育关系, 也表明其可能用作款冬物种和产区鉴别的超级条形码^[43], 佐证了叶绿体基因组数据在提高系统发育分辨率方面的潜力。

本研究获得了款冬花主产区基原植物款冬的高质量叶绿体基因组, 丰富了款冬的遗传信息资源。同时, 对款冬的叶绿体基因组进行了系统分析与总结, 揭示了其密码子偏好、SSR、边界和基因组序列变异特征, 获得了款冬与其他菊科物种之间的亲缘关系, 为药用植物款冬的种质资源筛选、鉴定、保存及遗传多样性分析提供了分子证据。

作者贡献: 净易尧是本研究的执行人, 并参与论文撰写; 陈晓颖、姬海月、罗瑶、史嘉周参与数据整理及论文初稿的撰写; 胡本祥参与款冬鉴定工作; 张岗、王帮庆、高静、杨冰月参与实验设计和实验结果的分析; 彭亮是项目的构思者及负责人, 指导实验设计、数据分析、论文写作与修改。

利益冲突: 所有作者均声明不存在利益冲突。

References

- [1] Artur A, Bogna O, Agnieszka G, et al. Content of pyrrolizidine alkaloids in the leaves of coltsfoot (*Tussilago farfara* L.) in Poland [J]. Acta Soc Bot Pol, 2013, 82: 289-293.
- [2] Bota VB, Neamtu AA, Olah NK, et al. A comparative analysis

- of the anatomy, phenolic profile, and antioxidant capacity of *Tussilago farfara* L. vegetative organs [J]. *Plants (Basel)*, 2022, 11: 1663.
- [3] Wang YD, Yang CD, Jin MJ, et al. Isolation and identification of *Tussilago farfara* leaf spot caused by *Alternaria alternata* in China [J]. *Microb Pathog*, 2022, 172: 105750.
- [4] Chen SJ, Dong L, Quan HF, et al. A review of the ethnobotanical value, phytochemistry, pharmacology, toxicity and quality control of *Tussilago farfara* L. (coltsfoot) [J]. *J Ethnopharmacol*, 2020, 267: 113478.
- [5] Hou AJ, Guo XY, Man WJ, et al. Research progress of chemical compositions and pharmacological effects of Farfarae Flos [J]. *Inf Tradit Chin Med (中医药信息)*, 2019, 36: 107-112.
- [6] Tao HJ. The Collected Notes of Ben Cao Jing (本草经集注) [M]. Beijing: People's Medical Publishing House, 1994: 355.
- [7] Wu D, Zhang ZF, Zhang M, et al. Study on chemical constituents of the flower buds of *Tussilago farfara* [J]. *Chin Pharm J (中国药理学杂志)*, 2008, 43: 260-263.
- [8] Shi LW, Li ML, Cui XW, et al. Comprehensive quality appraisal of flower buds of *Tussilago farfara* from 39 populations in Gansu province and analysis of affecting factors [J]. *Chin Tradit Herb Drugs (中草药)*, 2022, 53: 3784-3792.
- [9] Yang BY, Luo Y, Ji HY, et al. Analysis of volatile organic compounds of raw and honey baked Farfarae Flos based on head-space-gas-chromatography ion-mobility spectrometry [J]. *Chin Tradit Herb Drugs (中草药)*, 2022, 53: 1854-1861.
- [10] Lin X, Li CY, Jia MT, et al. Research progress on chemical constituents and pharmacological effects of Farfarae Flos and quality markers prediction and analysis [J]. *Chin Arch Tradit Chin Med (中华中医药学刊)*, 2023, 41: 131-140.
- [11] Chen CJ, Miao YH, Luo DD, et al. Sequence characteristics and phylogenetic analysis of the *Artemisia argyi* chloroplast genome [J]. *Front Plant Sci*, 2022, 13: 906725.
- [12] Wang WW, Lanfear R. Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants [J]. *Genome Biol Evol*, 2019, 11: 3372-3381.
- [13] Fan SJ, Guo XX. Research and application of chloroplast genome in plants [J]. *J Shandong Norm Univ (山东师范大学学报)*, 2022, 37: 22-31.
- [14] Hu SW, Ding YN, Bi GY, et al. Characterization and phylogenetic analysis of complete chloroplast genome of medicinal plant *Polygonum aviculare* [J]. *Chin Tradit Herb Drugs (中草药)*, 2022, 53: 2776-2785.
- [15] Jiang Y, Yang QS, Wang J, et al. Complete chloroplast genome of *Paris mairei*: characterization and phylogeny [J]. *Chin Tradit Herb Drugs (中草药)*, 2021, 52: 4014-4022.
- [16] Ma LH, Ning JQ, Wang YJ, et al. Comparative genomics on chloroplasts of *Sinopodophyllum hexandrum* [J]. *Chin J Biotechnol (生物工程学报)*, 2022, 38: 3695-3712.
- [17] Mcpherson H, van der MM, Delaney SK, et al. Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree [J]. *BMC Ecol*, 2013, 13: 8.
- [18] Zheng CC, Su X, Liu YP, et al. Characteristics of complete chloroplast genome and phylogenetic analysis of *Cirsium souliei* [J]. *Acta Pharm Sin (药学报)*, 2022, 57: 2224-2233.
- [19] Barry DB. Chloroplast transit peptides: structure, function and evolution [J]. *Trends Cell Biol*, 2000, 10: 440-447.
- [20] Yang XY, Liu C, Zeng ZY, et al. Sequence structure and phylogenetic analysis of the chloroplast genomes of *Alangium chinense* (Lour.) Harms and its different subspecies [J]. *Acta Pharm Sin (药学报)*, 2022, 57: 3229-3239.
- [21] Tillich M, Lehwark P, Morton BR, et al. The evolution of chloroplast RNA editing [J]. *Mol Biol Evol*, 2006, 23: 1912-1921.
- [22] Yi ZD, Lu K, Zhong YD, et al. The complete chloroplast genome sequence of *Tussilago farfara* (Asteraceae) [J]. *Mitochondrial DNA B Resour*, 2022, 7: 528-530.
- [23] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes [J]. *Nucleic Acids Res*, 2019, 47: W59-W64.
- [24] Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11 [J]. *Mol Biol Evol*, 2021, 38: 3022-3027.
- [25] 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.
- [26] Benson G. Tandem repeats finder: a program to analyze DNA sequences [J]. *Nucleic Acids Res*, 1999, 27: 573-580.
- [27] Ping C, Gang G, Chun MY, et al. Data set for transcriptome analysis of *Apocynum venetum* L [J]. *Data Brief*, 2018, 20: 1739-1744.
- [28] Li Q, Wan JM. SSR Hunter: development of a local searching software for SSR sites [J]. *Hereditas (遗传)*, 2005, 27: 808-810.
- [29] Amirousetfi A, Hyvönen J, Poczar P. IRscope: an online program to visualize the junction sites of chloroplast genomes [J]. *Bioinformatics (Oxford, England)*, 2018, 34: 3030-3031.
- [30] Li ZW, Qiu Q, Lang JQ, et al. Sequence analysis of complete chloroplast genome of *Dendrobium heterocarpum* and *Dendrobium trigonopus* [J]. *Chin Tradit Herb Drugs (中草药)*, 2022, 53: 5159-5169.
- [31] Du Q, Wang LQ, Chen ZE, et al. Characterization and phylogenetic analysis of the complete chloroplast genome of *Lycopus europaeus* [J]. *Acta Pharm Sin (药学报)*, 2022, 57: 2206-2215.
- [32] Xu SN, Niu BZ. Sheng Nong's Herbal Classic (神农本草经) [M]. Shijiazhuang: Hebei Science and Technology Publishing house, 1996: 84.
- [33] Fan JX, Qin XM, Li ZY. Mechanism of Farfarae Flos in Qingfei Paidu Decoction against COVID-19 based on network pharmacology and molecular docking [J]. *Chin Tradit Herb Drugs (中草*

- 药), 2020, 51: 2317-2325.
- [34] Zhang YY, Jiao WY, Jiao WR, et al. Analysis of the chloroplast genome of *Incarvillea younghusbandii* Sprague [J]. J Biol Eng (生物工程学报) 2023, 39: 2954-2964.
- [35] Wang X, Feng ZQ, Wang DR, et al. Chloroplast genome in *malus floribunda siebold* [J]. J Biol Eng (生物工程学报), 2022, 38: 3713-3727.
- [36] Zhang B, Yang LQ, Bian F, et al. Cloning and sequence analysis of *clp* protease gene (TaClpP) from wheat *Triticum aestivum* L. [J]. Shandong Agric Sci (山东农业科学), 2014, 46: 1-5.
- [37] Liang FP, Wen XN, Gao HY, et al. Analysis of chloroplast genomes features of Asteraceae species [J]. Genom Appl Biol (基因组学与应用生物学), 2018, 37: 5437-5447.
- [38] Feng Z, Jiang Y, Zheng Y, et al. Codon use bias analysis of chloroplast genome of *Cistanche* [J]. Chin Tradit Herb Drugs (中草药), 2023, 54: 1540-1550.
- [39] Qin N, Zhang NJ, Chen C, et al. Comparison and phylogenetic analysis of chloroplast genomes of *Viola philippica* and other *Viola* L. species [J]. Mol Plant Breed (分子植物育种), 2022, 20: 8099-8111.
- [40] Zhao PY, Zhao W, Hou ZY, et al. Comparison of chloroplast genomic characteristics and phylogenetic analysis between *themedata japonica* and two species of *Themedata* [J]. Chin Herb Med (中草药), 2023, 54: 3261-3272.
- [41] Fernandez PN, Menda N, Edwards JD, et al. The sol genomics network (SGN) --from genotype to phenotype to breeding [J]. Nucleic Acids Res, 2015, 43: 1036-1041.
- [42] Huang HB, He YH, Huang X, et al. Chloroplast whole genome of *ilex rotunda* and its phylogenetic analysis [J]. Agric Res Appl (农业研究与应用), 2022, 35: 7-14.
- [43] Jiang WJ, Guo MY, Pang XH. Application of chloroplast genome in identification and phylogenetic analysis of medicinal plants [J]. World Chin Med (世界中医药), 2020, 15: 702-708, 716.