热带亚热带植物学报
JOURNAL OF TROPICAL AND SUBTROPICAL BOTANY

滇黄精叶绿体全基因组序列及其密码子使用偏性分析

石乃星, 谢平选, 李立, 文国松

引用本文:
石乃星,谢平选,李立,文国松. 滇黄精叶绿体全基因组序列及其密码子使用偏性分析[J]. 热带亚热带植物学报, 2022, 30(3): 336–348.

在线阅读 View online: https://doi.org/10.11926/jtsb.4472

热带亚热带植物学报 2022, 30(3): 336 ~ 348
Journal of Tropical and Subtropical Botany

滇黄精叶绿体全基因组序列及其密码子使用偏性分析

石乃星[1], 谢平选[2], 李立[3], 文国松[1*]

(1. 云南农业大学农学与生物技术学院，昆明 650201；2. 广东药科大学中药学院，广州 510006；3. 西北大学生命科学学院，西安 710069)

摘要：为探究滇黄精(*Polygonatum kingianum*)叶绿体全基因组特征和密码子使用偏性，利用第二代测序技术对滇黄精嫩叶进行测序，再经组装与注释后得到其叶绿体基因组全序列，通过 MISA、EMBOSS 和 CodonW 等软件对滇黄精叶绿体全基因组的 SSR 位点、系统发育及密码子偏好性进行分析。结果表明，滇黄精完整叶绿体基因组长度为 155 852 bp，基因组平均 GC 含量为 37.7%，其大、小单拷贝区(LSC)长度分别为 84 633 和 185 25 bp，反向重复区长度为 26 347 bp，注释了 132 个基因，包括 86 个蛋白编码基因、38 个 tRNA 基因和 8 个核糖 rRNA 基因。叶绿体基因组中共有 69 个 SSR 位点，绝大多数属于单碱基重复的 A/T 类型。系统发育分析表明滇黄精与格脉黄精(*P. tessellatum*)亲缘关系近，可能与分布地域有关。密码子偏好性分析表明，滇黄精叶绿体基因组密码子使用模式受到自然选择影响大于突变因素，最终确定 9 个最优密码子。因此，滇黄精叶绿体基因组遗传结构和系统发育位置及其密码子偏倚的分析，为叶绿体基因工程研究提供理论依据。

关键词：滇黄精；叶绿体基因组；密码子使用偏性；最优密码子

doi: 10.11926/jtsb.4472

Analysis of Chloroplast Genome Characteristics and Codon Usage Bias of *Polygonatum kingianum*

SHI Naixing[1], XIE Pingxuan[2], LI Li[3], WEN Guosong[1*]

(1. *College of Agronomy and Biotechnology*, *Yunnan Agricultural University*, Kunming 650201, China; 2. *College of Traditional Chinese Medicine*, *Guangdong Pharmaceutical University*, Guangzhou 510006, China; 3. *School of Life Science*, *Northwest University*, Xi'an 710069, China)

Abstract: In order to explore the chloroplast genome characteristics and codon bias of *Polygonatum kingianum*, the next generation sequencing technology was utilized to sequence the young leaves, and the complete chloroplast genome sequence was obtained after assembly and annotation. SSR loci, phylogeny and codon preference of the chloroplast genome were analyzed by MISA, Emboss and Codonw software. The results showed that the length of complete chloroplast genome sequence of *P. kingianum* was 155 852 bp, including a pair of inverted repeats of 26 347 bp that were separated by large and small single copy regions (LSC, 84 633 bp and SSC, 18 525 bp). A total of 132 genes were annotated in the chloroplast genome, including 86 protein-coding genes, 8 rRNA genes and 38 tRNA genes. The average GC content of chloroplast genomes was 37.7%. A total of 69 SSR loci were detected, most of which belonged to single-base repeat A/T type. Phylogenetic analysis showed that *P. kingianum* was closer to *P. tessellatum* than other species, which may be related to their geographical distribution. The chloroplast genome codon usage pattern was more influenced by natural selection than mutation, and 9 codons were identified as the optimal codon. Therefore, these would provide important reference information for exploring the genetic relationship and the improvement of exogenous genes in *P. kingianum*.

Received: 2021-06-24 **Accepted:** 2021-09-14
This work was supported by the National Natural Sciences Foundation of China (Grant No. 81360611).
SHI Naixing, Male, interesting in evaluation and germplasm innovation of medicinal plant resources. E-mail: shi15587221483@163.com
* Corresponding author. E-mail: wengs@163.com

Key words: *Polygonatum kingianum*; Chloroplast genome; Codon usage bias; Optimal codon

Polygonatum kingianum, also known as jiejiegao and xianren rice in China, belongs to the perennial herb of *Polygonatum* Mill in Asparagaceae. Its wild resources are widely distributed in southwest China. In addition, it is engaged in relevant artificial cultivation and planting industries in Yunnan, Guizhou, Sichuan[1–3]. *Polygonatum kingianum* is one of the source plants of Rhizoma Polygonati which has high medicinal and edible value[4]. As a medicinal plant, *P. kingianum* is recorded in various national medical books. Chinese ancient medical book (Ming Yi Bie Lu, Han Dynasty, 220—450 AD) listed Rhizoma Polygonati as the top grade. Modern pharmacological studies have found their main chemical components have anti-aging, anti-tumor, immune enhancement, sterilization and anti-inflammatory effects[5].

As an essential subcellular organelle of plants and algae, previous study has found that chloroplast is not only the main place for photosynthesis, but also participates energy transformation[6]. In addition, chloroplasts also have relatively independent genome. In most angiosperms, the chloroplast genome belongs to the maternal inheritance, which has the characteristics of stable structure, conserved coding region sequence, rich information[7–8]. The complete chloroplast genome has been widely used in plant system evolution[9–11], related species identification and genetic diversity analysis[12], chloroplast genetic engineering, etc[13].

Codon, also known as triplet code, as a bridge connecting nucleic acid and protein[14], is the basic unit of biological genetic information transmission. In the case of mutation pressure, natural selection and genetic drift, prokaryotic and eukaryotic organisms generally tend to use one or more specific synonymous codons called synonymous codon usage bias (CUB)[15]. Through the analysis of species codon usage bias, the optimal codon can be determined, which can improve the efficiency and accuracy of related gene expression products, infer the function and expression mode of unknown genes, and provide scientific basis for exploring species relationship and genetic evolution[16–17].

Until now, the codon usage bias of some species has been analyzed[18–20], but the research on the codon usage preference of *P. kingianum* has not been reported yet. In the present study, Illumina Hiseq was used technologies for complete chloroplast genome sequence of *P. kingianum*. Based on this, we analyzed the sequence characteristics and codon usage bias of chloroplast genome. It was of great significance to provide a scientific reference for the application and investigation of the chloroplast genome in *P. kingianum*.

1 Materials and methods

1.1 Material collection and sequencing

Polygonatum kingianum was collected in Tengchong City of China. The total DNA was extracted from 100 mg of fresh and healthy leaves using the modified CTAB method. Then the complete cp genome was sequenced by using Illumina Hiseq 2000 sequencing platform. The reference specimen (Ji et Wang 2) was deposited at the Herbarium of Kunming Institute of Botany, Chinese Academy of Sciences.

1.2 Plastome assembly, annotation, and comparison

First, we assembled the complete chloroplast genome with a reference-based assembly strategies with GetOrganelle using the complete chloroplast genome sequence of *P. kingianum* (NCBI reference sequence: MN934979) as reference. Then, the assembly was edited and annotated according to the reference in Geneious V 10.2[21]. We generated a physical map of the cp genome using Organellar GenomeDRAW[22]. Finally, the complete chloroplast genome of *P. kingianum* was submitted to the NCBI (Accession: MW788495).

1.3 Microsatellite analysis

Perl scripts from MISA were used to perform SSR identification with the default parameters (http://pgrc.ipk-gatersleben.de/misa/). The identification criteria were as follows: mono-nucleotide repeat motifs with

at least 10 repeats, di-nucleotide repeat motifs with 5 repeats, trinucleotide repeat motifs with four repeats, tetra-, penta- and hexa-nucleotide repeat motifs with three repeats. Compound SSRs were defined as those with a < 100-nt interval between two repeat motifs[23–24].

1.4 Phylogenetic analysis

In order to explore the evolutionary relationships of *P. kingianum*, the whole chloroplast genome sequences of two genera of *Heteropolygonatum* and *Polygonatum* from NCBI along with the obtained chloroplast genome sequence in the present study, were analyzed for phylogenetic analysis. All of the chloroplast genome sequences were aligned with MAFFT[25] implemented in Geneious (10.0.5), and a maximum-likelihood phylogenetic analysis was performed in RAxML[26] under the GTR-GAMMA model with 1 000 bootstrap replicates.

1.5 Codon composition and optimal codon analysis

Based on the consideration of reducing sample error and accurately counting the number of effective codons, we eliminated the sequences with length less than 300 bp, duplicate genes, and coding sequences (CDS) containing stop codon. Fifty-three CDS sequences (start codon: ATG; stop codons: TAA, TAA, TGA and TAG) of the chloroplast genome of *P. kingianum* were used as research samples to analyze the CUB ultimately[27]. In order to analyze the rule of gene base composition, Codon W 1.4.2 software was used for the analysis relative synonymous codon usage (RSCU). We also used the CUSP and CHIPS models in the online software EMBOSS to analyze the GC content of the first base of codon (GC_1), the second base of codon (GC_2), the third base of codon (GC_3), total GC content (GC), and effective number of codons (ENC). The Pearson correlation analysis of the above parameters was carried out using SPSS 24.0 software.

The RSCU value of codon more than 1 are determined to be high frequency codons[28]. Then taking an ENC as the preferred standard, five genes with the highest ENC values and the lowest ENC values in 51 chloroplast genes were regarded as the high and low expression groups. The RSCU values of 2 datasets were calculated and compared by ΔRSCU (RSCU in high and low expression groups). The codons satisfying both ΔRSCU > 0.08 defined as high expression codons. Finally, by combining high frequency and high expression codons, the optimal codons was defined for the chloroplast genome of *P. kingianum*[29].

1.6 Neutrality plot analysis

A neutral graph was drew to research the influence of mutation pressure and natural selection on the chloroplast codon usage pattern of *P. kingianum*. GC_3 values were regarded as abscissa; the average values of GC_1 and GC_2 of each gene were seen as GC_{12}, which were ordinate. The correlation analysis of GC_3 and GC_{12} will be helpful to make scientific judgments on the main factors affecting codon preference[30]. If there is a significant correlation between the two data and the regression coefficient is close to 1, it means that the codon preference is mainly affected by mutation pressure. On the contrary, it indicates that base composition preference is mainly affected by selection pressure.

1.7 ENC-GC3s plot analysis

The effective number of codon (ENC) is a measure of the degree of species independent synonymous codon bias in genes. Its value ranges from 20 to 61, which is negatively correlated with the CUB[31].

ENC-plot analysis can intuitively judge gene codon preference factors. With the GC3s values as horizontal ordinate and ENC values as longitudinal coordinate, two-dimensional scatter plot was drawn[31]. Standard curve formula: $ENC_{exp}=2+GC3s+29/[GC3s^2+(1-GC3s)^2]$[32]. This curve shows the functional relationship between ENC and GC3s only under mutation pressure conditions. The ENC ratio distribution can quantify the results obtained by ENC-plot and clarify how far away or close each gene point is from the curve. The expression：ENC ratio=(ENC expected value−ENC actual value)/ENC expected value[33].

1.8 Analysis of PR2-bias plot

Based on the analysis of the composition of the

four bases (A, T, C, G) at the third position of the chloroplast genome of the *P. kingianum*, we used $G_3/(G_3+C_3)$ and $A_3/(A_3+T_3)$ as the horizontal and vertical coordinates for analysis. The PR2-bias plot, which analyzes the nucleotide compositions at the third position of codons, are usually used to estimate the effects of mutation pressure and natural selection by analysing the AT bias and GC bias[34]. Through the vector emitted by the center of the plane, we can judge the degree and direction of the 4 kinds of bias[35].

2 Results

2.1 Characteristics of complete chloroplast genome sequence

The length of the complete chloroplast genome sequence of *P. kingianum* was 155 852 bp, with GC content of 37.7%. It contained a pair of inverted repeats (IRs, 26 347 bp each), a small single copy (SSC, 18 525 bp) and large single copy region (LSC, 84 633bp). The average GC content of IR, LSC and SSC regions were 43%, 31.6% and 35.7%, respectively. Annotation results showed that there were 132 genes in the chloroplast genome in *P. kingianum*, including 85 protein-coding genes, 38 tRNA genes and 8 ribosomal rRNA genes. The GC contents of these three types of genes were 38.1%, 53.2% and 55.3%, respectively. The coding region of the gene was 90 607 bp, accounting for 40.3% of the entire chloroplast genome (Fig. 1). The coding genes families of *P. kingiantum* chloroplast were involved in four aspects: photosynthesis, self-replication, biosynthesis and unknown function. Table 1 shows the gene functions and groups in the cp genome. Compared with the chloroplast genomes of other species in *Polygonatum* genus, such as *P. zanlanscianense* (155 609 bp)[36] and *P. humile* (156 082 bp)[37], all cp genomes shared the same gene order and structure, which displayed a high degree of similarity.

2.2 SSR analysis of chloroplast genome

With MISA analysis, a total of 69 SSRs were identified in the chloroplast genome of *P. kingianum*. Examination of all SSR loci in the genome showed that the majority of the SSR loci were located in the LSC region, with a number of 50 (72.46%). There were 11 (15.94%) located within the SSC region and the least number of SSRs located in the IR region, with only 8 (11.59%) (Fig. 2). The types of SSRs differed greatly in the number of repeats. The number of mononucleotide repeats was the largest, with 43, and all repeating units were A/T. There were 15, 5, 10, and 2, dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide repeats, respectively; while hexanucleotide repeats were not observed. In terms of SSR repeat unit types, tetranucleotide repeat units were the most common, followed by the dinucleotide, trinucleotide repeat units and, finally, the mononucleotide, pentanucleotide repeat units (Table 2).

2.3 Phylogenetic analysis

The phylogenetic graph revealed that most of species of *Polygonatum* were clustered into a monophyletic clade with a high bootstrap value (Fig. 3). It indicated that the resulting phylogenetic tree we constructed was relatively robust. The sequence of *P. kingianum* in this study and *P. huanum* (*P. huanum* is a synonym of *P. kingianum*) from GenBank clustered together with a support rate of 100%. It showed that the conclusion of morphological identification was supported by molecular evidence. These samples were clearly divided into three branches. Branch I included 8 species (*P. urceolatum*, *P. punctatum*, *P. stewardtianum*, *P. oppositifolium*, *P. tessellatum*, *P. huanum*, *P. kingianum*). Branch II composed 5 species (*P. yunnanense*, *P. arisanense*, *P. humile*, *P. biflorum*, *P. cyrtonema*). The other two species (*Heteropolygonatum altelobatum* and *H. ginfushanicum*) were classified into branche III. The verticillate leaf type of *Polygonatum* are clustered in a clade, and alternate phyllotaxis species are clustered in anther clade.

2.4 Codon usage bias of chloroplast genome

2.4.1 Codon composition and optimal codon

From the chloroplast genome of *P. kingianum*, 51 CDSs suitable for analysis of CUB were selected (Table

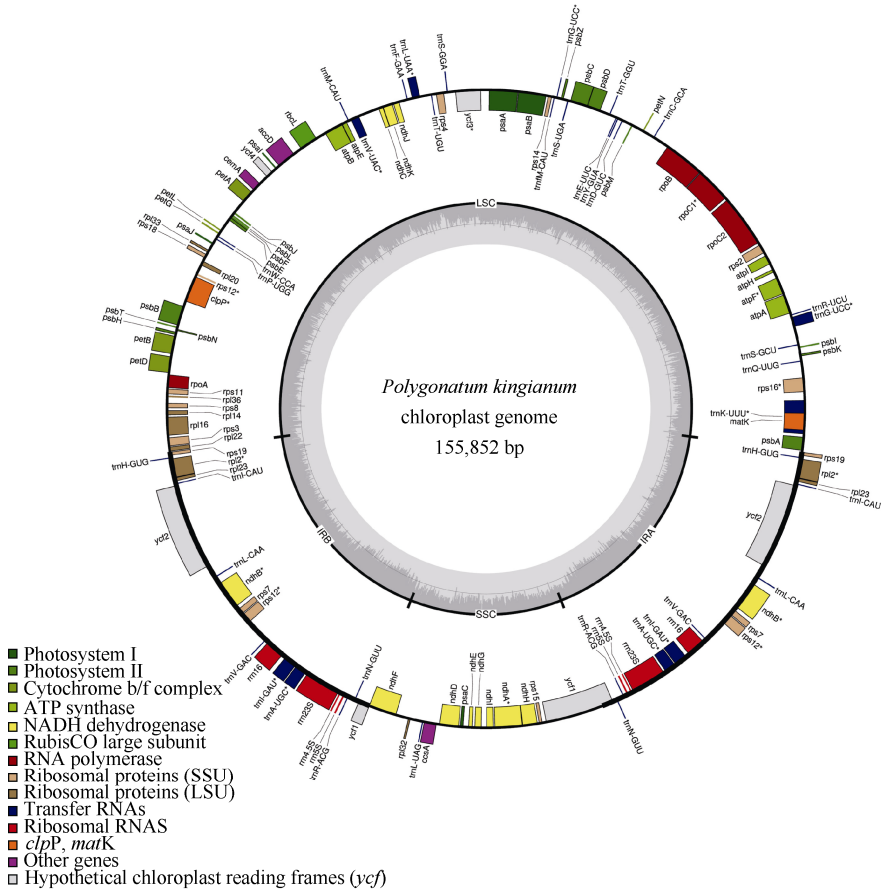

Fig. 1 Gene map of the chloroplast genome of *Polygonatum kingianum*

Table 1 List of identified genes in cp genomes of the *Polygonatum kingianum*

Category	Group	Name
Self-replication	Ribosomal RNA genes	*rrn*4.5 × 2, *rrn*5 × 2, *rrn*16 × 2, *rrn*23 ×
	Transfer RNA genes	*trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*F-GAA, *trn*G-UCC*, *trn*G-GCC, *trn*H-GUG × 2, *trn*K-UUU*, *trn*L-UAA*, *trn*L-UAG, *trn*M-CAU, *trn*P-UGG, *trn*Q-UUG, *trn*R-UCU, *trn*S-GCU, *trn*S-GGA, *trn*S-UGA, *trn*T-UGU, *trn*T-GGU, *trn*V-UAC*, *trn*Y-GUA, *trn*W-CCA, *trn*fM-CAU, *trn*A-UGC* × 2, *trn*I-CAU × 2, *trn*I-GAU* × 2, *trn*L-CAA × 2, *trn*N-GUU × 2, *trn*R-ACG × 2, *trn*V-GAC × 2
	Ribosomal protein (small subunit)	*rps*2, *rps*3, *rps*4, *rps*7 × 2, *rps*8, *rps*11, *rps*12** × 2, *rps*14, *rps*15, *rps*16*, *rps*18, *rps*19 × 2
	Ribosomal protein (large subunit)	*rpl*2* × 2, *rpl*14, *rpl*16*, *rpl*20, *rpl*22, *rpl*23 × 2, *rpl*32, *rpl*33, *rpl*36
	RNA polymerase	*rpo*A, *rpo*B, *rpo*C1*, *rpo*C2
Photosynthesis	Subunits of photosystem I	*psa*A, *psa*B, *psa*C, *psa*I, *psa*J, *ycf*3**, *ycf*4
	Subunits of photosystem II	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*N, *psb*T, *psb*Z
	Subunits of cytochrome	*pet*A, *pet*B*, *pet*D*, *pet*G, *pet*L, *pet*N
	Subunits of ATP synthase	*atp*A, *atp*B, *atp*E, *atp*F*, *atp*H, *atp*I
	Large subunit of Rubisco	*rbc*L
	Subunits of NADH dehydrogenase	*ndh*A*, *ndh*B* × 2, *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K
Others	Maturase	*mat*K
	Envelope membrane protein	*cem*A
	Subunit of acetyl-CoA	*acc*D
	Synthesis gene	*ccs*A
	ATP-dependent protease	*clp*P**
	Component of TIC complex	*ycf*1 × 2
Unknown function	Conserved open reading frames	*ycf*2 × 2

× 2: Two gene copies in IR regions; *: With one intron; **: With two introns.

Table 2 Repeat type, number and frequency of SSRs in complete chloroplast genome of *Polygonatum kingianum*

Repeat	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
A/T	–	–	–	–	–	–	–	24	10	5	1	1	1	1	43
AG/CT	–	–	3												3
AT/AT	–	–	8	2		2									12
AAT/ATT	–	3	1												4
AGC/CTG	–	1													1
AAAT/ATTT	3														3
AATC/ATTG	3														3
AATG/ATTC	3														3
AATT/AATT	1														1
AAACG/CGTTT	2														2

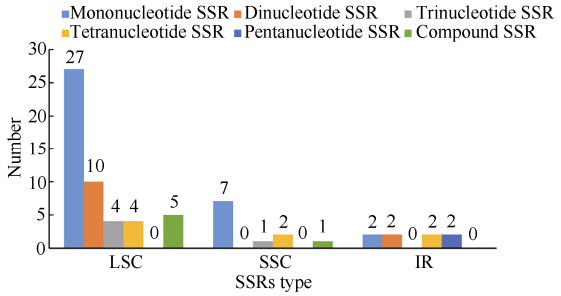

Fig. 2 Distribution of various types of SSRs in LSC, SSC and IR regions

Polygonatum are clustered in a clade, and alternate phyllotaxis species are clustered in anther clade.

3). The average GC content (GC_{all}) of these 51 CDs sequences was 38.28% (30.77%−45.08%). The GC contents at different positions of the codons varied, with the average GC contents at the first position (GC_1) being 46.84% (31.79%−58.54%), the second (GC_2) being 39.38% (26.61%−55.40%), and the third (GC_3) being 28.61% (20.90%−37.23%), indicating that the codons in the chloroplast genome of *P. kingianum* preferred to end with A or U. The statistics of the ENC values of 51 CDs sequences (Table 2) showed that the average value of the ENC of all genes was 48.21 (range from 39.83 to 60.11), suggesting a weak codon preference.

Pearson correlation analysis showed that the average GC content (GC_{all}) at all codon locations was significantly correlated with GC_1, GC_2 and GC_3. GC_1

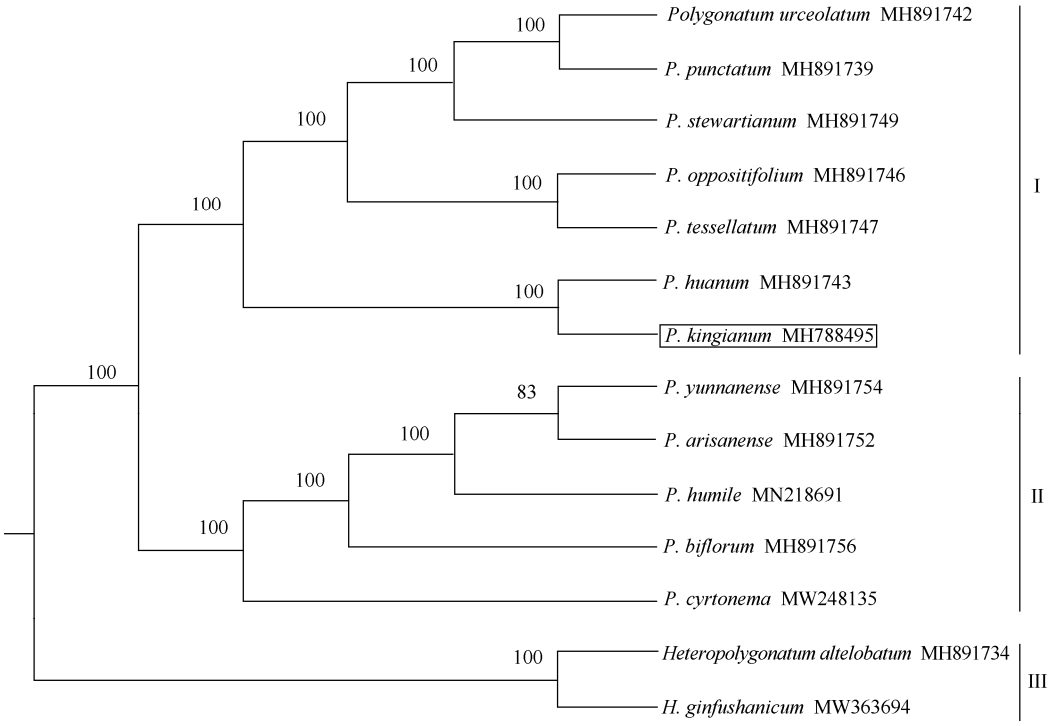

Fig. 3 Phylogenetic tree constructed using the maximum likelihood method based on chloroplast genome sequence

Table 3 ENC value and GC content in different positions of codons in chloroplast 51 CDS of *Polygonatum kingianum*

Gene	GC_1 /%	GC_2 /%	GC_3 /%	GC_{all} /%	GC3s	ENC	Gene	GC_1 /%	GC_2 /%	GC_3 /%	GC_{all} /%	GC3s	ENC
accD	38.79	37.17	28.69	34.88	0.251	46.492	*psbA*	49.72	43.22	32.77	41.90	0.284	39.830
atpA	55.51	39.17	26.18	40.29	0.246	44.131	*psbB*	55.01	46.95	29.67	43.88	0.258	48.078
atpB	57.06	41.53	32.46	43.68	0.306	49.293	*psbC*	54.22	46.20	31.65	44.02	0.279	46.374
atpE	52.59	41.48	31.85	41.98	0.295	51.470	*psbD*	52.26	43.22	32.49	42.66	0.281	45.278
atpF	47.57	35.68	30.81	38.02	0.285	44.010	*rbcL*	57.29	43.33	28.33	42.99	0.252	47.238
atpI	48.79	36.29	26.21	37.10	0.235	46.452	*rpl14*	56.10	36.59	22.76	38.48	0.203	44.350
ccsA	31.79	37.96	29.01	32.92	0.244	49.147	*rpl16*	50.00	54.41	26.47	43.63	0.200	45.052
cemA	39.91	26.61	29.61	32.05	0.256	51.217	*rpl20*	38.98	43.22	29.66	37.29	0.268	53.473
clpP	58.54	36.1	30.73	41.79	0.269	57.759	*rpl22*	41.80	39.34	27.05	36.07	0.233	47.289
matK	41.46	29.37	23.22	31.35	0.211	47.108	*rpoA*	46.06	35.86	30.61	37.51	0.287	51.501
ndhA	44.81	37.98	23.77	35.52	0.208	43.732	*rpoB*	49.91	38.45	28.03	38.80	0.258	47.989
ndhB	42.27	40.12	31.51	37.96	0.277	46.865	*rpoC1*	50.51	37.12	30.42	39.35	0.283	51.481
ndhC	51.24	34.71	28.10	38.02	0.216	48.909	*rpoC2*	46.16	37.84	27.93	37.31	0.264	49.174
ndhD	41.42	36.49	30.77	36.23	0.261	51.560	*rps2*	43.04	42.62	31.22	38.96	0.277	51.220
ndhE	36.27	32.35	30.39	33.01	0.268	57.097	*rps3*	46.61	34.39	24.43	35.14	0.221	44.939
ndhF	38.67	37.45	24.83	33.65	0.207	44.685	*rps4*	48.02	38.61	29.21	38.61	0.279	51.701
ndhG	42.37	33.9	20.90	32.39	0.178	44.105	*rps7*	53.21	45.51	23.72	40.81	0.213	47.021
ndhH	49.75	37.06	25.89	37.56	0.207	45.883	*rps8*	39.85	40.60	25.56	35.34	0.228	41.137
ndhI	41.99	35.36	24.86	34.07	0.229	49.924	*rps11*	56.12	55.40	23.74	45.08	0.211	48.250
ndhJ	47.17	38.99	33.33	39.83	0.291	56.030	*rps14*	45.54	47.52	31.68	41.58	0.299	40.673
ndhK	42.80	41.63	24.51	36.32	0.219	46.302	*rps18*	36.27	43.14	25.49	34.97	0.232	42.376
petA	53.56	35.91	29.1	39.53	0.278	49.691	*ycf1*	36.19	29.43	26.68	30.77	0.235	47.921
petB	48.62	41.28	30.73	40.21	0.252	47.187	*ycf2*	41.89	34.92	37.23	38.01	0.346	52.972
petD	50.29	38.01	25.15	37.82	0.215	45.235	*ycf3*	46.75	40.24	31.95	39.64	0.296	60.106
psaA	51.53	43.28	33.16	42.65	0.291	49.327	*ycf4*	44.32	41.62	31.35	39.10	0.267	50.765
psaB	48.03	42.99	33.33	41.45	0.290	48.853							

was significantly correlated with GC_2, but neither of them reached a significant level with GC_3, showing that the composition of the first and second base of the codon was similar, and there was a great difference from the third base. The correlation of ENC with GC_1 and GC_2 did not reach significance, but it was highly significant with GC_3, indicating that the third base composition can significantly influence codon usage bias (Table 4). The correlation coefficient between ENC and codon counts (CC) was 0.079, which did not reach a significant level, indicating that the CC has a very weak influence on ENC. That is, in this study, the effect of gene length on codon bias analysis was not significant.

There were 30 codons (RSCU>1) in the chloroplast genomic protein coding sequence of *P. kingianum,* of which only one codon ended with G or C, and the remaining 29 codons ended with A and U, indicating that the chloroplast genome of *P. kingianum* prefers to use Codons ending with A or U; there were 31 codons (RSCU<1), of which 28 ended with G and C, and 3 ended with A and U (Table 5), manifesting that the frequency of codons ending in C and G was relatively low in the chloroplast genome of *P. kingianum*. According to *Δ*RSCU values of codes in high

Table 4 Correlation analysis of each gene's related parameters of *Polygonatum kingianum*

	GC_1	GC_2	GC_3	GC_{all}	ENC
GC_2	0.397**				
GC_3	0.084	0.107			
GC_{all}	0.816**	0.767**	0.423**		
ENC	–0.017	–0.217	0.448**	0.025	
CC	–0.130	–0.246	0.266	–0.115	0.079

**: $P<0.01$.

and low libraries, 26 highly-expressed superior codon were screened out from the chloroplast genome of *P. kingianum* (Table 6). Combining the high expression superior codons with the 30 high-frequency codons described above, nine common codon (UUU, CUU, UCA, CCA, CAU, CAA, AAU, GAU, GGA) were finally identified as the optimal codons for the genome of *P. kingianum*, with 4 codons ending with A, and the remaining 5 codons all end with U, whereas the chloroplast protein encoding genome of *P. kingianum* preferred the codons ending with A and U, especially the codons ending with U.

Table 5 RSCU analysis of CDS in *Polygonatum kingianum*

AA	Codon	Count	RSCU	AA	Codon	Count	RSCU
Phe	UUU	760	1.27	Ser	UCU	449	1.68
	UUC	433	0.73		UCC	267	1.00
Leu	UUA	691	1.92		UCA	321	1.20
	UUG	457	1.27		UCG	153	0.57
	CUU	448	1.25		AGU	337	1.26
	CUC	147	0.41		AGC	78	0.29
	CUA	292	0.81	Pro	CCU	321	1.50
	CUG	119	0.33		CCC	182	0.85
Ile	AUU	877	1.45		CCA	249	1.17
	AUC	359	0.59		CCG	102	0.48
	AUA	582	0.96	Thr	ACU	430	1.65
Met	AUG	493	1.00		ACC	186	0.71
Val	GUU	437	1.51		ACA	318	1.22
	GUC	128	0.44		ACG	110	0.42
	GUA	433	1.50	Tyr	UAU	623	1.60
	GUG	157	0.54		UAC	158	0.40
Asp	GAU	685	1.58	Arg	CGU	285	1.38
	GAC	182	0.42		CGC	66	0.32
Glu	GAA	858	1.51		CGA	277	1.34
	GAG	277	0.49		CGG	96	0.47
Cys	UGU	177	1.53		AGA	393	1.90
	UGC	55	0.47		AGG	121	0.59
Ala	GCU	529	1.82	Gly	GGU	476	1.33
	GCC	176	0.61		GGC	143	0.40
	GCA	347	1.20		GGA	577	1.61
	GCG	108	0.37		GGG	236	0.66
TER*	UAA	22	1.29	His	CAU	417	1.53
	UAG	15	0.88		CAC	127	0.47
	UGA	14	0.82	Gln	CAA	569	1.52
Trp	UGG	384	1.00		CAG	181	0.48
Lys	AAA	791	1.52	Asn	AAU	762	1.53
	AAG	250	0.48		AAC	237	0.47

2.4.2 Neutrality plot

To estimate the extent of mutation pressure as well as natural selection contributed to the CUB of *P. kingianum*, a neutrality plot was constructed based on the GC_{12} and GC_3 (Fig. 4). The range was 0.328 1−0.557 6 and 0.209 0−0.372 3 for GC_{12}, GC_3, respectively. As shown in Fig. 4, most of the points represented by each gene were distributed above the diagonal of the neutrality plot. Only the *ycf*2 gene was close to the diagonal. The Pearson correlation coefficient between GC_{12} and GC_3 was 0.142 (P=0.311>0.05), showing that the correlation between them was not significant. The regression coefficient observed was closer to zero (The slope of the regression line was 0.161 1), which inferred that the GC content in the chloroplast genome of *P. kingianum* was highly

conserved. Natural selection played a remarkably important role in the CUB of *P. kingianum*. And the mutation pressure accounted for a minority of the affecting factors.

Table 6 Optimal codons in chloroplast genome of *Polygonatum kingianum*

AA	Codon	High expression		Low expression		⊿RSCU	AA	Codon	High expression		Low expression		⊿RSCU
		Number	RSCU	Number	RSCU				Number	RSCU	Number	RSCU	
Phe	UUU*	22	1.33	31	1.09	0.24	TER	UAG	1	0.60	1	0.60	0.00
	UUC	11	0.67	26	0.91	–0.24		UGA***	1	0.60	0	0.00	0.60
Leu	UUA	17	1.55	45	2.18	–0.63	Ala	GCU	24	1.68	42	2.40	–0.72
	UUG	13	1.18	24	1.16	0.02		GCC**	10	0.70	5	0.29	0.41
	CUU***	19	1.73	23	1.11	0.62		GCA	13	0.91	21	1.20	–0.29
	CUC**	5	0.45	3	0.15	0.30		GCG***	10	0.70	2	0.11	0.59
	CUA	7	0.64	24	1.16	–0.52	His	CAU*	13	1.53	12	1.41	0.12
	CUG*	5	0.45	5	0.24	0.21		CAC	4	0.47	5	0.59	–0.12
Ile	AUU	30	1.25	54	1.47	–0.22	Gln	CAA*	20	1.54	19	1.31	0.23
	AUC	12	0.50	23	0.63	–0.13		CAG	6	0.46	10	0.69	–0.23
	AUA**	30	1.25	33	0.90	0.35	Asn	AAU***	30	1.58	23	1.00	0.58
Val	GUU	18	1.80	28	2.07	–0.27		AAC	8	0.42	23	1.00	–0.58
	GUC**	6	0.60	3	0.22	0.38	Lys	AAA	19	1.31	30	1.71	–0.40
	GUA	12	1.20	20	1.48	–0.28		AAG**	10	0.69	5	0.29	0.40
	GUG*	4	0.40	3	0.22	0.18	Asp	GAU*	27	1.64	19	1.52	0.12
Ser	UCU	14	1.58	33	2.15	–0.57		GAC	6	0.36	6	0.48	–0.12
	UCC	7	0.79	18	1.17	–0.38	Glu	GAA	35	1.56	45	1.58	–0.02
	UCA***	11	1.25	11	0.72	0.53		GAG	10	0.44	12	0.42	0.02
	UCG**	6	0.68	3	0.20	0.48	Cys	UGU	4	1.14	6	1.50	–0.36
	AGU	13	1.47	23	1.50	–0.03		UGC**	3	0.86	2	0.50	0.36
	AGC	2	0.23	4	0.26	–0.03	Arg	CGU	13	1.30	21	1.68	–0.38
Pro	CCU	7	1.08	27	2.40	–1.32		CGC*	4	0.40	4	0.32	0.08
	CCC***	6	0.92	3	0.27	0.65		CGA	12	1.20	15	1.20	0.00
	CCA***	10	1.54	9	0.80	0.74		CGG**	8	0.80	4	0.32	0.48
	CCG	3	0.46	6	0.53	–0.07		AGA	15	1.50	25	2.00	–0.50
Thr	ACU	9	1.16	21	1.83	–0.67		AGG**	8	0.80	6	0.48	0.32
	ACC*	9	1.16	11	0.96	0.20	Gly	GGU	9	0.90	34	1.70	–0.80
	ACA	9	1.16	10	0.87	0.29		GGC**	8	0.80	8	0.40	0.40
	ACG	4	0.52	4	0.35	0.17		GGA**	18	1.80	28	1.40	0.40
Tyr	UAU	33	1.69	25	1.43	0.26		GGG	5	0.50	10	0.50	0.00
	UAC	6	0.31	10	0.57	–0.26	Met	AUG	23	1.00	24	1.00	0.00
TER	UAA	3	1.80	4	2.40	–0.60	Trp	UGG	13	1.00	22	1.00	0.00

*: ⊿RSCU≥0.08; **: ⊿RSCU≥0.3; ***: ⊿RSCU≥0.5.

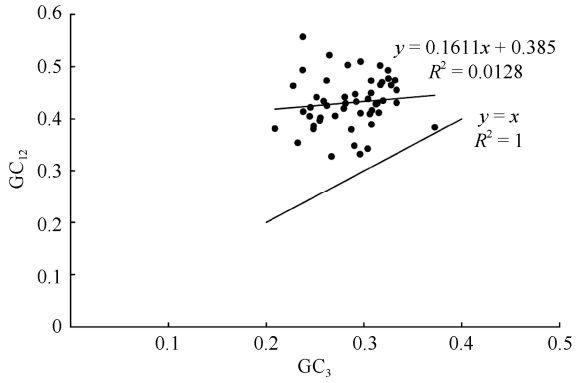

Fig. 4 Neutrality plot analysis

2.4.3 ENC-GC3s plot

The standard curve in the ENC-plot reflected the relationship between ENC and GC3s only when the influence of selection pressure is excluded. Figure 5 showed that some genes in the chloroplast of *P. kingianum* were located near the standard curve. The actual ENC value of this part of the gene was close to the expected ENC value, indicating that the mutation effect was greater than natural selection; while the position of the other part of the gene farther from the

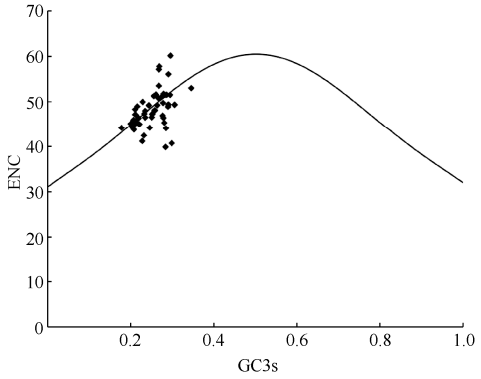

Fig. 5 Analysis of ENC-plot

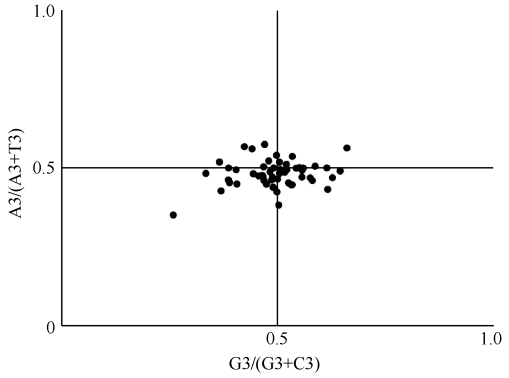

Fig. 6 Analysis of PR2-plot

standard curve represented natural selection factors were stronger than mutations. In order to quantify the closeness of genes to the standard curve, the ENC ratio was used to count the frequency of ENC ratios (Table 7). The results showed that there were only 21 genes with ratios in the range from −0.05 to 0.05, accounting for 41.18% of the total number of genes. That means, most genes were far away from the standard curve, and the codon preference was related to the difference of GC3s, indicating that the codon preference of the chloroplast genome of *P. kingianum* was more affected by selection than mutation.

Table 7 Distribution of ENC ratio

Class	Middle value	Number	Frequence
–0.15–0.05	–0.1	3	0.06
–0.05–0.05	0.0	21	0.41
0.05–0.15	0.1	22	0.43
0.15–0.25	0.2	4	0.08
0.25–0.35	0.3	1	0.02
Total		51	1.00

2.4.4 PR2-plot

PR2-plot analysis showed that the chloroplast genes of *P. kingianum* were scatteredly distributed in the four regions of the chart, and most of the genes were distributed in the lower left part of the chart (Fig. 6), indicating that the frequency of T base in the third codon was higher than that of A base, and that of C base was higher than that of G base, that is, the frequency of pyrimidine was higher than that of purine. If the codon usage pattern is completely caused by mutation, the usage frequency of the four bases should be equal. The biased usage of four bases indicated that the usage pattern of chloroplast codon in *P. kingianum* was not only influenced by mutation, but also by other factors, such as selection pressure.

3 Conclusion and discussion

Polygonatum kingianum is one of medicine and food homologous plants announced by the National Health Commission, PRC. For thousands of years, its medicinal effect and edible value has been widely recognized by Chinese people. Up to date, it has integrated medicinal, edible, ornamental and health function which has extremely high economic value and social benefits. In summary, *P. kingianum* has good development and research prospects[2].

In this study, basic research on the chloroplast of *P. kingianum* was conducted. The assembly annotation results showed that the chloroplast genome of *P. kingianum* was 155 852 bp in length, including one large single-copy region, one small single-copy region and two inverted repeat regions, which is consistent with the typical tetrad structure of the chloroplast genome of most angiosperms[38]. The study found that the chloroplast genome of *P. kingianum* was not much different from the chloroplast genome of other species in the genus *Polygonatum* in sequence length. Simple sequence repeats (SSRs) are an important part of the plant chloroplast genome which play an indispensable role in gene expression, transcription regulation, chromosome construction, and physiological metabolism[39]. After statistical analysis of MISA online software, 69

SSR loci were detected. These SSR loci will be helpful for subsequent research on the population genetics of *P. kingianum*. Studies have shown that *P. kingianum* is a species with large morphological variation, and it is difficult to identify this species only from morphology. The phylogenetic tree results of this study provide reliable DNA molecular evidence support for the morphological identification of *P. kingienum*. Compared with other plants of *Polygonatum* genus, the relationship between *P. kingianum* and *P. tessellatum* is closer, and the geographic distribution of 2 species is basically the same.

Although CUB is affected by many factors, natural selection and mutation are key factors that affect codon usage preference[40]. The results of codon bias in the chloroplast genome of *P. kingianum* showed that the average effective codon ranged from 39.830 to 60.106, which suggested a weak codon preference among these chloroplast genes.

The results of related parameters analysis and neutrality plot showed that the third position of codon had low base composition similarity with the first and second position. The correlation between the first position and the other two positions of the codon was not significant, indicating that the codon bias was subject to a strong degree of selection.

The results of ENC-plot analysis also confirmed the above argument. The ENC-plot graph showed that only a small part of genes distributed near the standard curve, and the actual ENC value of these genes were basically consistent with the theoretical ENC values. It indicated that these codon preferences were greatly affected by mutations. The scattered points of most genes were far from the standard curve. The actual ENC value of this part of the gene was quite different from the theoretical ENC value, indicating that it was more easily affected by the selection.

The analysis result of PR2-plot concluded that at the third codon position of cp genes, pyrimidines (C and T) were used more frequently than purines (A and G), which verified that the codon usage pattern of the chloroplast genes of *P. kingianum* was more affected by selection factors.

Based on the results of the high-frequency and highly expressed codons, nine codons (UUU, CUU, UCA, CCA, CAU, CAA, AAU, GAU, GGA) were obtained as the optimal codons of *P. kingianum* chloroplast genes finally. This preference pattern is consistent with the results of codon bias analysis of chloroplast genes in *Oryza*[41], *Oncidium gower*[42], *Panicum miliaceum*[43], etc. These indicate that *P. kingiantum* prefers to use codons ending in AT like other monocotyledonous plants.

The chloroplast genome sequences are not only most valuable for understanding plant evolution and phylogeny but also have made some achievments in chloroplast transformation technology. However, chloroplast genome has rarely been used in evolution and phylogeny of the *P. kingiantum.* Only three relevant research have been repored[37,44–45]. These reports mainly discussed the phylogenetic relationship and identification of *Polygonatum* genus. Up to now, chloroplast genetic transformation technology has been applied in a variety of plants. Degray, et al[46] transferred the antimicrobial peptide gene *MSI-99* into the chloroplast genome of tobacco by chloroplast transformation technology, and the descendants of transgenic plants showed high antibacterial activity. Chakrabarti, et al[47] incorporated a truncated *Bacillus thuringiensis cry9Aa2* gene in the plastid genome of tobacco to control of potato tuber moth (*Phthorimaea operculella*). But, it has not been found any studies on chloroplast genetic transformation of *Polygonatum* species by far.

In conclusion, we sequenced and analyzed the complete cp genome of *P. kingiantum*, which exhibits conserved structure. The phylogenetic relationship based on the plastid genome data shows that the chloroplast genome as ultra-barcoding has great potential in the identification of *Polygonatum* species. Comprehensive analysis found that the codon bias of the *P. kingiantum* chloroplast genome is weak, and the factors that affect the formation of codon bias in the chloroplast protein-coding genes of *P. kingianum* do not depend on a single factor, but are the result of mutations, selection and many other factors. According

to the codon usage characteristics of its chloroplast genome, we screened out 9 optimal codons for the chloroplast genome of *P. kingiantum*. The current study provides a scientific reference for the identification of germplasm resources, genetic breeding as well as for prediction of unknown functional genes, discovery of new genes, improvement of foreign gene expression in *P. kingianum*.

References

[1] CHEN X Q, LIANG S Y, XU J M, et al. Liliaceae [M]// WU Z Y, RAREN P H. Flora of China, vol. 24. Beijing: Science Press & St. Louis: Missouri Botanical Garden Press, 2000: 223–232.

[2] LIU W, LIN M Y, LIU J J, et al. Progress in study of *Polygonatum kingianum* and research status of *Polygonati rhizoma* [J]. Chin J Exp Trad Med Formulae, 2017, 23(14): 226–234. (in Chinese) doi: 10.13422/j.cnki.syfjx.2017140226.

[3] LI J N, LIU H P, XIONG X W, et al. Growth performance of *Polygonatum kingianum* in *Carya illinoensis* grove [J]. J W China For Sci, 2020, 49(5): 42–46. (in Chinese) doi: 10.16473/j.cnki.xblykx1972.2020.05.007.

[4] Chinese Pharmacopoeia Commission. Chinese Pharmacopoeia of the People's Republic of China, Vol. 1 [M]. Beijing: China Medical Science Press, 2020: 1–319. (in Chinese)

[5] ZHANG J, WANG Y Z, YANG W Z, et al. Research progress in chemical constituents in plants of *Polygonatum* and their pharmacological effects [J]. China J Chin Mat Med, 2019, 44(10): 1989–2008. (in Chinese) doi: 10.19540/j.cnki.cjcmm.20190222.006.

[6] VERMA D, DANIELL H. Chloroplast vector systems for biotechnology applications [J]. Plant Physiol, 2007, 145(4): 1129–1143. doi: 10.1104/PP.107.106690.

[7] JIANG W J, GUO M Y, PANG X H. Application of chloroplast genome in identification and phylogenetic analysis of medicinal plants [J]. World Chin Med, 2020, 15(5): 702–708. (in Chinese) doi: 10.3969/j.issn.1673-7202.2020.05.008.

[8] BI Y F, WEN X, PAN Y H, et al. Application and research progress of chloroplast DNA barcoding in forest trees [J]. Mol Plant Breed, 2020, 18(16): 5444–5452. (in Chinese) doi: 10.13271/j.mpb.018.005444.

[9] WU L W, CUI Y X, NIE L P, et al. The characteristics of complete chloroplast genome sequence and phylogenetic analysis of *Dendrobium moniliforme* [J]. Acta Pharm Sin, 2020, 55(5): 1056–1066. (in Chinese) doi: 10.16438/j.0513-4870.2019-0940.

[10] LI P R, ZHANG S J, LI F, et al. A phylogenetic analysis of chloroplast genomes elucidates the relationships of the six economically important *Brassica* species comprising the triangle of U [J]. Front Plant Sci, 2017, 8: 111. doi: 10.3389/fpls.2017.00111.

[11] ZHANG J W, JIANG Z M, SU H, et al. The complete chloroplast genome sequence of the endangered species *Syringa pinnatifolia* (Oleaceae) [J]. Nord J Bot, 2019, 37(5): 2201–2212. doi: 10.1111/njb.02201.

[12] WANG M H, HU S J, YANG C H, et al. Chloroplast genome analysis of *Liriope spicata* var. *prolifera*, *Ophiopogon japonicus* in Sichuan and Zhejiang [J]. Chin J Exp Trad Med Formulae, 2020, 26(8): 182–191. (in Chinese) doi: 10.13422/j.cnki.syfjx.20200947.

[13] ZHANG J F, LI Y, JIN J J, et al. Recent advances in tobacco chloroplast genetic engineering [J]. Tob Sci Technol, 2017, 50(6): 88–98. doi: 10.16135/j.issn1002-0861.2017.0035. (in Chinese)

[14] WALL D P, HERBECK J T. Evolutionary patterns of codon usage in the chloroplast gene *rbc*L [J]. J Mol Evol, 2003, 56(6): 673–688. doi: 10.1007/s00239-002-2436-8.

[15] QIN H, WU W B, COMERON J M, et al. Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes [J]. Genetics, 2004, 168(4): 2245–2260. doi: 10.1534/genetics.104.030866.

[16] LI Y, DUO J C, XIONG H Y, et al. Codon bias analysis of barley homeodomain-leucine zipper gene family [J]. J Triticeae Crops, 2020, 40(2): 144–153. (in Chinese) doi: 10.7606/j.issn.1009-1041.2020.02.02.

[17] YANG G F, SU K L, ZHAO Y R, et al. Analysis of codon usage in the chloroplast genome of *Medicago truncatula* [J]. Acta Pharm Sin, 2015, 24(12): 171–179. doi: 10.11686/cyxb2015016.

[18] LIU F L, WANG F J, SHAO Z R. Characterization of codon usage in chloroplast genome of *Ectocarpus siliculosus* [J]. Adv Mar Sci, 2012, 30(4): 587–594. doi: 10.3969/j.issn.1671-6647.2012.04.015.

[19] WANG J, WANG T Y, WANG L Y, et al. Assembling and analysis of the whole chloroplast genome sequence of *Elaeagnus angustifolia* and its codon usage bias [J]. Acta Bot Boreali-Occid Sin, 2019, 39(9): 1559–1572. doi: 10.7606/j.issn.1000-4025.2019.09.1559.

[20] ZHOU C Z, ZHU C, LI X Z, et al. Codon usage bias analysis methods of *Camellia sinensis* and its research progress [J]. Mol Plant Breed, 2020, 18(5): 1480–1488. (in Chinese) doi: 10.13271/j.mpb.018.001480.

[21] KEARSE 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(12): 1647–1649. doi: 10.1093/bioinformatics/bts199.

[22] LOHSE M, DRECHSEL O, BOCK R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes [J]. Curr Genet, 2007, 52(5/6): 267–274. doi: 10.1007/s00294-007-0161-y.

[23] ZHAO C Z, QIU J J, AGARWAL G, et al. Genome-wide discovery of microsatellite markers from diploid progenitor species, *Arachis duranensis* and *A. ipaensis*, and their application in cultivated peanut (*A. hypogaea*) [J]. Front Plant Sci, 2017, 8: 1209. doi: 10.3389/fpls.2017.01209.

[24] DENG P C, WANG M, FENG K W, et al. Genome-wide characterization of microsatellites in Triticeae species: Abundance, distribution and evolution [J]. Sci Rep, 2016, 6(1): 32224. doi: 10.1038/srep32224.

[25] KATOH K, STANDLEY D M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability [J]. Mol Biol Evol, 2013, 30(4): 772–780. doi: 10.1093/molbev/mst010.

[26] STAMATAKIS A. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies [J]. Bioinformatics, 2014, 30(9): 1312–1313. doi: 10.1093/bioinformatics/btu033.

[27] XIN Y X, DONG Z H, QU S H, et al. Analysis on codon usage bias of chloroplast genome in *Pyrus betulifolia* Bge. [J]. J Agric Univ Hebei, 2020, 43(6): 51–59. doi: 10.13320/j.cnki.jauh.2020.0112. (in Chinese)

[28] XU Y, JIA R Y, ZHANG Z L, et al. Analysis of synonymous codon usage pattern in duck circovirus [J]. Gene, 2015, 557(2): 138–145. doi: 10.1016/j.gene.2014.12.019.

[29] QIN Z, ZHENG Y J, GUI L J, et al. Codon usage bias analysis of chloroplast genome of camphora tree (*Cinnamomum camphora*) [J]. Guihaia, 2018, 38(10): 1346–1355. doi: 10.11931/guihaia.gxzw201805023. (in Chinese)

[30] YANG X, LUO X N, CAI X P. Analysis of codon usage pattern in *Taenia saginata* based on a transcriptome dataset [J]. Parasite Vector, 2014, 7(1): 527. doi: 10.1186/s13071-014-0527-1.

[31] WRIGHT F. The ‘effective number of codons’ used in a gene [J]. Gene, 1990, 87(1): 23–29. doi: 10.1016/0378-1119(90)90491-9.

[32] YUAN X L, LI Y Q, ZHANG J F, et al. Analysis of codon usage bias in the chloroplast genome of *Dalbergia odorifera* [J]. Guihaia, 2021, 41(4): 622–630. (in Chinese) doi: 10.11931/guihaia.gxzw201906012.

[33] SUEOKA N. Near homogeneity of PR2-bias fingerprints in the human genome and their implications in phylogenetic analyses [J]. J Mol Evol, 2001, 53(4/5): 469–476. doi: 10.1007/s002390010237.

[34] SUEOKA N. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons [J]. J Mol Evol, 1995, 40(3): 318–325. doi: 10.1007/BF00163236.

[35] SUEOKA N. Translation-coupled violation of parity rule 2 in human genes is not the cause of heterogeneity of the DNA G+C content of third codon position [J]. Gene, 1999, 238(1): 53–58. doi: 10.1016/S0378-1119(99)00320-0.

[36] SHI N X, LI L, WANG S Y, et al. The complete chloroplast genome of *Polygonatum zanlanscianense* (*Pampanini*, 1915) (Asparagaceae), an adulterants of *Polygonati rhizoma* [J]. Mitochondrial DNA B, 2021, 6(8): 2420–2421. doi: 10.1080/23802359.2021.1945973.

[37] LEE S Y, ZOU Y L, LIAO W B, et al. The complete chloroplast genome of a traditional medicinal and food plant, *Polygonatum humile* (Asparagaceae, Asparagales) [J]. Mitochondrial DNA B, 2019, 4(2): 3184–3185. doi: 10.1080/23802359.2019.1666044.

[38] ZHANG Y J, LI D Z. Advances in phylogenomics based on complete chloroplast genomes [J]. Plant Div, 2011, 33(4): 365–375. (in Chinese) doi: 10.3724/SP.J.1143.2011.10202.

[39] AI D Y. The meaning of repeat sequences [J]. Chem Life, 2008, 28(3): 343–345. (in Chinese) doi: 10.3969/j.issn.1000-1336.2008.03.031.

[40] SHI Y B, WANG G B, YANG X M, et al. Analysis of codon usage bias of gene factors in *Ginkgo biloba* WRKY family [J]. Mol Plant Breed, 2019, 17(5): 1503–1511. (in Chinese) doi: 10.13271/j.mpb.017.001503.

[41] CHAKRABORTY S, YENGKHOM S, UDDIN A. Analysis of codon usage bias of chloroplast genes in *Oryza* species [J]. Planta, 2020, 252(4): 67. doi: 10.1007/s00425-020-03470-7.

[42] LI D M, LÜ F B, ZHU G F, et al. Analysis on codon usage of chloroplast genome of *Oncidium gower* Ramsey [J]. Guangdong Agric Sci, 2012, 39(10): 61–65. (in Chinese) doi: 10.16768/j.issn.1004-874x. 2012.10.058.

[43] LIU H, WANG M X, YUE W J, et al. Analysis of codon usage in the chloroplast genome of *Broomcorn millet* (*Panicum miliaceum* L.) [J]. Plant Sci J, 2017, 35(3): 362–371. (in Chinese) doi: 10.11913/PSJ.2095-0837.2017.30362.

[44] JIN J, LAO J, ZHONG C, et al. Complete chloroplast genome of a medicinal species *Polygonatum kingianum* in China (Asparagaceae, Asparagales) [J]. Mitochondrial DNA B, 2020, 5(1): 959–960. doi: 10.1080/23802359.2020.1721373.

[45] FLODEN A, SCHILLING E E. Using phylogenomics to reconstruct phylogenetic relationships within tribe Polygonateae (Asparagaceae), with a special focus on *Polygonatum* [J]. Mol Phylogenet Evol, 2018, 129: 202–213. doi: 10.1016/j.ympev.2018.08.017.

[46] DEGRAY G, RAJASEKARAN K, SMITH F, et al. Expression of an antimicrobial peptide *via* the chloroplast genome to control phytopathogenic bacteria and fungi [J]. Plant Physiol, 2001, 127(3): 852–862. doi: 10.1104/pp.127.3.852.

[47] CHAKRABARTI S K, LUTZ K A, LERTWIRIYAWONG B, et al. Expression of the *cry9Aa2 B.t.* gene in tobacco chloroplasts confers resistance to potato tuber moth [J]. Transgen Res, 2006, 15(4): 481. doi: 10.1007/s11248-006-0018-z.