LRF-881
[1-11]

RESEARCH ARTICLE Legume Research- An International Journal

Features and Phylogenetic Analysis of the Chloroplast Genome of the Medicinal Plant *Euchresta tubulosa*

Meng Zhang[1,2], Kefan Cao[3]

10.18805/LRF-881

ABSTRACT

Background: *Euchresta tubulosa* is a medicinal plant in the Fabaceae family, traditionally used in Chinese herbal medicine for treating ailments such as sore throat and esophageal cancer. Despite its pharmacological importance and endangered status, no complete chloroplast genome data had been reported for this species or its genus prior to this study. Chloroplast genomes, being highly conserved and maternally inherited, serve as powerful tools for species identification, phylogenetic inference and genetic diversity assessment in plants.

Methods: Fresh leaves of *E. tubulosa* were collected and genomic DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method. High-throughput sequencing was carried out on the Illumina NovaSeq6000 platform. Clean reads were assembled into a complete chloroplast genome using NovoPlasty, with a related species as the seed reference. Genome annotation was performed using DOGMA and validated manually. Codon usage patterns were analyzed using RSCU values and simple sequence repeats (SSRs) were identified with MISA software. A maximum likelihood phylogenetic tree was constructed using PhyML, based on complete chloroplast genome sequences from 14 related species.

Result: The assembled chloroplast genome of *E. tubulosa* measured 153,960 bp and exhibited a typical quadripartite structure, comprising a large single-copy region (LSC), a small single-copy region (SSC) and two inverted repeats (IRs). The overall GC content was 36.3%. A total of 127 genes were identified, including 44 related to photosynthesis, 72 involved in self-replication, 5 with other biological roles and 6 hypothetical or unknown genes. Codon usage analysis revealed 31 codons with RSCU > 1, showing a strong bias toward A/U-ending codons. A total of 109 SSR loci were detected, the majority of which were located in intergenic spacer regions. Phylogenetic analysis clearly positioned *E. tubulosa* as an independent lineage within the Fabaceae, supported by high bootstrap values. The genomic information provided by this study lays a foundation for future research in molecular taxonomy, phylogenetics and conservation of this endangered medicinal plant.

Key words: Chloroplast genome, *Euchresta tubulosa*, Phylogenetic analysis, SSR markers.

INTRODUCTION

Euchresta tubulosa Dunn, commonly referred to as Hu Dou Lian, is a climbing shrub species within the Fabaceae family and the genus *Euchresta*. This medicinal plant is predominantly found in southern China's provinces of Guangxi, Guangdong and Guizhou, where it typically grows in humid environments at altitudes between 650 and 900 meters. Lin *et al*. 2013 reported that the folk anticancer medicine known as "Hu Dou Lian" primarily refers to *E. tubulosa* and its closely related species (Lin *et al.,* 2013; Wang and Li, 1983). The dried roots of this plant contain a variety of alkaloids and flavonoids, which are widely used to alle*via*te sore throats, gum pain and oral ulcers. Among the Tujia ethnic group inhabiting the Wuling Mountains (including Hunan, Hubei, Sichuan, Chongqing and Guizhou), *E. tubulosa* is traditionally used to treat esophageal and throat cancers (Li *et al.,* 2014; Lo *et al.,* 2002; Toda and Shirataki, 2006). There is even a saying among the Tujia people: "Having Hu Dou Lian at home ensures a good year."

Due to its limited reproductive strategies, specific habitat requirements and low pollination efficiency, the wild population of *E. tubulosa* has declined significantly (Li *et al.,* 2014, Yao *et al.,* 2017, Han *et al.,* 2013). It is now listed as a second-class national key protected wild plant and categorized as an endangered species in the "*China Species Red List*." Current studies on *E. tubulosa* mainly focus on its cultivation, domestication and the extraction of active medicinal compounds. However, despite its medicinal significance, no prior research has addressed the complete chloroplast genome of *Euchresta* species, including *E. tubulosa*.

[1]Annoroad Gene Technology (Beijing) Co., Ltd.Beijing 100176, China.
[2]College of Animal Science and Technology, Henan Agricultural University, Zhengzhou 450046, Henan, China.
[3]College of Grassland Science/Key Laboratory of Grassland Resources of Ministry of Education, Inner Mongolia Agricultural University, Hohhot, Inner Mongolia 010018, China.

Corresponding Author: Kefan Cao, College of Grassland Science/Key Laboratory of Grassland Resources of Ministry of Education, Inner Mongolia Agricultural University, Hohhot, Inner Mongolia 010018, China. Email: caokefan1003@163.com

How to cite this article: Cao, K. and Zhang, M. (2025). Features and Phylogenetic Analysis of the Chloroplast Genome of the Medicinal Plant *Euchresta tubulosa*. *Legume Research*. 1-11. doi: 10.18805/LRF-881.

Submitted: 20-06-2025 **Accepted:** 03-08-2025 **Online**: 24-09-2025

Chloroplast genomes play a central role in photosynthesis, amino acid and secondary metabolite synthesis and are maternally inherited, making them a powerful tool for plant taxonomy, population genetics and phylogenetic analysis. Although the chloroplast genomes of many angiosperms have been sequenced, genomic resources for the genus *Euchresta* remain scarce. Recent studies have demonstrated the value of chloroplast genomes in resolving taxonomic ambiguities in medicinal plants (Cui *et al.* 2012, Wang *et al.* 2020). For example, Zhao *et al*. 2022 sequenced the complete chloroplast genome of *Asarum sieboldii* and reconstructed phylogenetic relationships within the Aristolochiaceae family. Zhou *et al*. 2018 established super DNA barcodes in *Rheum*, while analyzed codon usage patterns in *Alpinia* species.

Within the *Fabaceae* family, a growing number of chloroplast genome studies have shed light on genome structure, codon usage bias and evolutionary dynamics of key genera. Notably, the complete chloroplast genome of *Medicago sativa* cv. Qingda no.1 has been sequenced and analyzed, revealing a genome length of 125,637 bp, a relatively high GC content of 38.33% and the absence of a typical quadripartite structure. The study also demonstrated strong A/T-ending codon usage bias and identified 62 SSR loci, highlighting important structural and evolutionary features of alfalfa chloroplast genomes (Ren *et al.* 2023). Other genera, including *Cajanus* (Kaila *et al.* 2016), *Glycine* (Asaf *et al.* 2017) and *Desmodium* (Yen and Park 2022), have also been studied for their plastome characteristics, such as SSR distribution and GC content variation, which offer insights into legume phylogeny and taxonomy. These studies collectively reinforce the importance of plastome data in clarifying taxonomic relationships and evolutionary origins in *Fabacea*.

To address this gap, this study aims to sequence, annotate and analyze the chloroplast genome of *E. tubulosa* using high-throughput sequencing technology (Tai and Tanksley 1990, Zhang *et al.* 2012, Yang *et al.* 2022, Mu *et al.* 2022). The results will provide insights into its genome structure, codon usage and phylogenetic placement, laying a foundation for future research on species identification, genetic diversity, conservation and molecular systematics of this endangered medicinal plant.

MATERIALS AND METHODS

Materials

Leaves of *Euchresta tubulosa* were collected from Fanjingshan in Jiangkou County, Guizhou Province (27.9112°N, 108.7033°E) and taxonomically identified by Professor Shuwen Zhao of Inner Mongolia Agricultural University. Healthy leaf samples were collected into sterile bags, rinsed with tap water followed by 3-4 washes with sterile water, air-dried and subsequently preserved at -80/°C in an ultra-low temperature freezer.

DNA extraction and sequencing

Leaves were rapidly frozen in liquid nitrogen and finely ground into a powder using a mortar and pestle. Genomic DNA was isolated using the cetyltrimethylammonium bromide (CTAB) method as described by (Doyle 1987), followed by purification using a TIANquick Midi Purification Kit (Tiangen Biotech, Beijing, China). The quality and quantity of DNA were assessed using a Nano Drop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and 1.0% agarose gel electrophoresis.

The DNA was fragmented by ultrasonication using a Covaris S220 system (Covaris, Woburn, MA, USA). End repair, A-tailing and adapter ligation were performed following the protocol of (Meyer and Kircher 2010), with minor modifications. The ligated products were purified and enriched *via* PCR amplification using Phusion High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) to construct the sequencing library. The final library was assessed for size distribution using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and quantified using a Qubit 3.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) before sequencing on the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA).

Chloroplast genome assembly

Following the removal of adapter sequences and low-quality reads, high-quality clean reads were acquired. The chloroplast genome of *Euchresta tubulosa* was then assembled using NOVOPlasty (Perl script, default settings), with the chloroplast genome of *Euchresta* japonica serving as the reference seed sequence (Sun *et al.,* 2020).

Genome annotation and comparison

The Geneious Prime Find Repeats plugin was used to detect the inverted repeat (IR) sequences in the chloroplast genome and to define the boundaries of the large single-copy (LSC) and small single-copy (SSC) regions (Kearse *et al.* 2012). Genome annotation was performed using the Dual Organellar GenoMe Annotator (DOGMA, [http://dogma.ccbb.utexas.edu/] (http://dogma.ccbb.utexas.edu/)) and manually adjusted. tRNA genes were predicted using tRNAscan-SE ([http://lowelab.ucsc.edu/tRNAscanSE/] (http://lowelab.ucsc.edu/tRNAsc anSE/)) and ARAGORN tools. Finally, the Organellar Genome DRAW ([https://chlorobox.mpimp-golm.mpg.de /OGDraw.html] (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html)) was used to generate the circular map with the "Tidy up annotation" option selected and other parameters set to default.

Simple sequence repeat (SSR) analysis

SSR loci were analyzed using the Perl script provided by MISA (MIcroSAtellite Identification Tool) (Beier *et al.* 2017). The script was run in the Windows command prompt, setting the minimum number of repeats for mononucleotides, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides to 10, 5, 4, 4, 4 and 4, respectively.

Nucleotide diversity was assessed using DnaSP version 5.0, applying a sliding window approach with a window length of 600 bp and a step interval of 200 bp (Rozas 2009).

Codon usage bias analysis

The relative synonymous codon usage (RSCU) values for genes within the chloroplast genome of *Euchresta tubulosa* were analyzed. Codon with an RSCU greater than 1 was considered a high-frequency codon. Highly expressed genes have lower effective number of codons (ENC) values, while lowly expressed genes have higher ENC values. Genes were ranked based on their ENC values and the top and bottom 5% were selected as the high and low expression gene sets, respectively. The difference in RSCU (ΔRSCU) between the high and low expression gene sets was calculated, with a ΔRSCU greater than 0.08 considered an optimal codon (Duret and Mouchiroud 1999, Liu *et al.* 2020).

Phylogenetic analysis

Complete chloroplast genome sequences from 14 species were retrieved from the NCBI database, including three species of *Lespedeza* (*L. buergeri*, *L. maritima* and *L. bicolor*), one of *Kummerowia* (*K. striata*), three of *Campylotropis* (*C. trigonoclada*, *C. polyantha* and *C. wilsonii*), one of *Christia* (*C. vespertilionis*), one of *Urariopsis* (*U. brevissima*), one of *Pseudarthria* (*P. viscida*) and two species from the genus *Saxifraga* (*Bergenia crassifolia* and *Saxifraga stolonifera*) as outgroups (Katoh *et al.* 2002).

Sequences were aligned using MAFFT version 7.388. The best substitution model was determined by jModelTest and a maximum likelihood (ML) tree was constructed using PhyML with 1,000 bootstrap replicates.

RESULTS AND DISCUSSION

Chloroplast genome structure, gene content and codon usage bias of *E. tubulosa*

The chloroplast genome of *Euchresta tubulosa* exhibits a typical quadripartite structure, comprising a large single-copy (LSC) region, a small single-copy (SSC) region and two inverted repeats (IRs) (Fig 1). The total length is 153,960 bp and the overall GC content is 36.3%. Specifically, the LSC spans 84,107 bp (42.63% GC), SSC 18,053 bp (33.77% GC) and each IR 51,800 bp (29.80% GC), consistent with typical angiosperm chloroplast genome features.

Genome annotation identified 127 functional genes, including 44 photosynthesis-related genes, 72 self-replication-related genes, 5 genes with other known functions and 6 genes of unknown function. Among the photosynthesis-related genes, *ndhB* (encoding NADH dehydrogenase subunit) occurs in two copies and contains one intron. The *petB* (cytochrome b/f complex subunit) and *atpF* (ATP synthase subunit) genes each contain one intron (Table 1). For self-replication-related genes, duplicated genes include ribosomal proteins (*rpl2*, *rpl23*, *rps12*, *rps7*), ribosomal RNAs (*rrn16S*, *rrn23S*, *rrn4.5S*, *rrn5S*) and several tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*). Genes with one intron include *rpl16*, *rpl2*, *rpoC1*, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnL-UAA* and *trnV-UAC*, while *clpP* contains two introns. Additionally, the pseudogenes *ycf1* and *ycf2* are both duplicated and *ycf3* has two introns.

Compared with other legumes, the *E. tubulosa* chloroplast genome is relatively large and gene-rich. For instance, *Medicago sativa* cv. Qingda No.1 possesses a smaller genome (125,637 bp) with a distinct structure showing IR region contraction and gene loss such as *ndh* genes (Ren

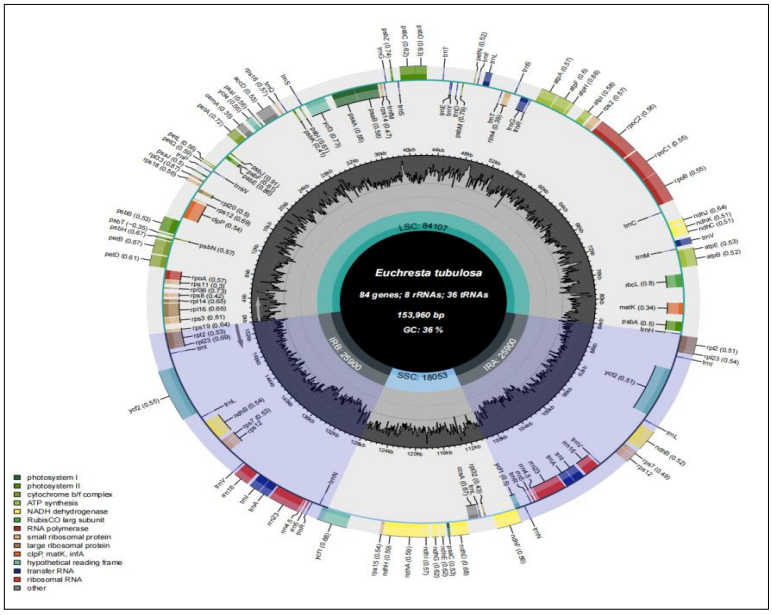

Fig 1: Circularized map of the chloroplast genome of *Euchresta tubulosa*.

Table 1: Gene annotation of the chloroplast genome of *Euchresta tubulosa*.

Gene category	Gene group	Gene name
Genes for photosynthesis	Subunits of photosystemI Subunits of photosystemaII Subunits of NADH dehydrogenase	*psaA, psaB, psaC, psaI, psaJ, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbM, psbN, psbT, psbZ* *ndhA*, ndhB*(2), ndhC, ndhD, ndh E, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	Subunit of cytochromeb/f complex	*petA, petB*, petD*, petG, petL, petN*
	Subunits of ATP synthase	*atpA, atpB, atpE, atpF*, atpH, atpI, rbcL*
	Large subunit of Rubisco	*rpl14, rpl16*, rpl2*(2), rpl20, rpl23(2), rpl32,*
	Large subunit of ribosome	*rpl33, rpl36rps11, rps12**(2), rps14, rps15,*
Self replication	Smal lsubuni tof ribosome	*rps16, rps18, rps19, rps2, rps3, rps4, rps7(2), rps8*
	DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1*, rpoC2, rrn16S(2), rrn23S(2), rrn4.5S(2), rrn5S(2),*
	Ribosomal RNA gene	*trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC*, trnH-GUG, trnI-CAU(2), trnI-GAU*(2), trnL-CAA(2), trnL-UAA*,trnL-UAG, trnM-CAU, trnN-GUU(2), trnP-GGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(2), trnV-UAC*, trnW-CCA, trnY-GUA, trnfM-CAU*
Transfer RNA gene		
	Maturase	*matK*
	Protease	*clpP***
	Envelop membrane proteinn	*cemA*
Other genes	Submit of acetyl-Co A-carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsA*
Genes of unknown function	Conserved open reading frame	*ycf1(2), ycf2(2), ycf3**, ycf4*

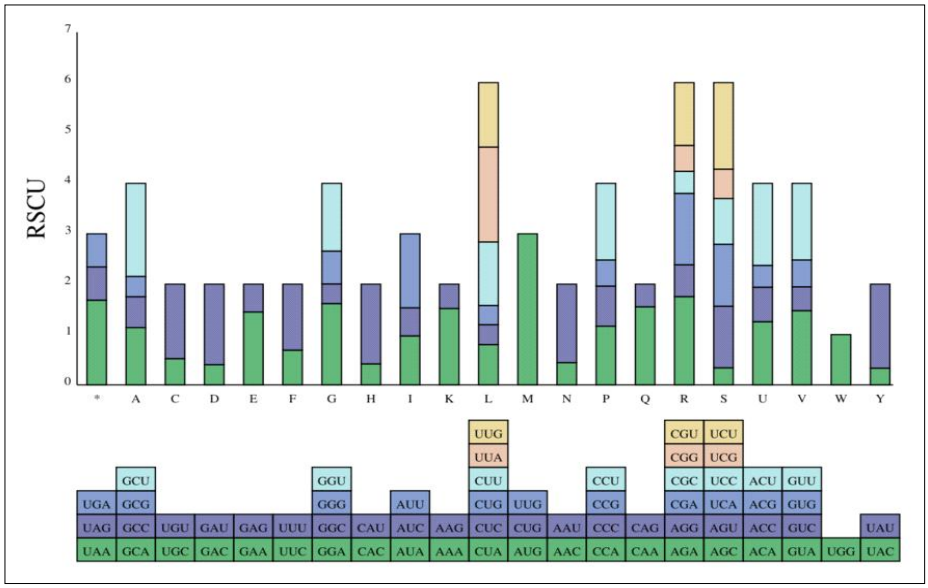

Fig 2: RSCU analysis of each amino acid in *Euchresta tubulosa*.

et al. 2023). This suggests significant evolutionary divergence within Fabaceae. Codon usage analysis of *E. tubulosa* indicates a strong preference for codons ending in A or U, with 29 out of 31 high-frequency codons terminating in either A or U and only two ending in G, consistent with findings from *M. sativa* and other angiosperm chloroplast genomes (Jiang *et al.* 2020).

This AT-rich codon usage bias likely reflects underlying mutational pressure and natural selection, as reported in related Fabaceae species (Zhao *et al.*, 2022). Codon usage bias analysis plays an important role in exploring gene expression efficiency, genome evolution and species adaptation, providing insights into phylogenetic relationships (Chen *et al.,* 2010).

Codon usage bias in the chloroplast genome of *E. tubulosa*

To investigate codon usage bias, we analyzed all chloroplast coding sequences (CDSs) longer than 200 bp. The results showed that leucine (Leu) was the most frequently encoded amino acid, with 2,644 codons representing 10.52% of the total codon count (Table 2). Relative synonymous codon usage (RSCU) values greater than 1 were found for 31 codons, indicating that these codons are used more frequently than expected. Among these, only two codons ended in G, while the remaining 29 codons ended in either A or U (Fig 2).

This pronounced preference for A/U-ending codons suggests an AT-rich bias in the chloroplast genome of *E. tubulosa*. Similar A/U-ending codon usage patterns have

Table 2: Analysis of protein coding region in *Euchresta tubulosa.*

Amino acid	Codon	Number	RSCU	Ratio/%	Amino acid	Codon	Number	RSCU	Ratio/%
*	UAA	52	1.6957	0.38%	Asn	AAC	249	0.433	4.75%
	UAG	19	0.6196			AAU	901	1.567	
	UGA	21	0.6848		Pro	CCA	301	1.1735	4.24%
Ala	GCA	376	1.1368	5.46%		CCC	205	0.7992	
	GCC	206	0.6228			CCG	128	0.499	
	GCG	134	0.4051			CCU	392	1.5283	
	GCU	607	1.8352		Gln	CAA	650	1.5403	3.48%
Cys	UGC	86	0.5292	1.34%		CAG	194	0.4597	
	UGU	239	1.4708		Arg	AGA	412	1.7695	5.77%
Asp	GAC	194	0.4097	3.91%		AGG	137	0.5884	
	GAU	753	1.5903			CGA	336	1.4431	
Glu	GAA	903	1.4755	5.05%		CGC	98	0.4209	
	GAG	321	0.5245			CGG	114	0.4896	
Phe	UUC	477	0.6704	5.88%		CGU	300	1.2885	
	UUU	946	1.3296		Ser	AGC	102	0.3357	7.53%
Gly	GGA	668	1.601	6.89%		AGU	381	1.254	
	GGC	165	0.3954			UCA	370	1.2178	
	GGG	269	0.6447			UCC	275	0.9051	
	GGU	567	1.3589			UCG	175	0.576	
His	CAC	121	0.4137	2.42%		UCU	520	1.7115	
	CAU	464	1.5863		Thr	ACA	391	1.2593	5.13%
Ile	AUA	713	0.9803	9.01%		ACC	212	0.6828	
	AUC	396	0.5445			ACG	127	0.409	
	AUU	1073	1.4753			ACU	512	1.649	
Lys	AAA	958	1.5291	5.17%	Val	GUA	472	1.4693	5.31%
	AAG	295	0.4709			GUC	153	0.4763	
Leu	CUA	350	0.82	10.57%		GUG	170	0.5292	
	CUC	159	0.3725			GUU	490	1.5253	
	CUG	161	0.3772		Trp	UGG	414	1	1.71%
	CUU	534	1.2511		Tyr	UAC	153	0.3438	3.67%
	UUA	821	1.9235			UAU	737	1.6562	
	UUG	536	1.2558						
Met	AUG	566	3	2.34%					
	CUG	0	0						
	UUG	0	0						

also been reported in the chloroplast genomes of *Medicago sativa*, *Glycine max* and other legume species, where the vast majority of preferred codons terminate in A or U (Jiang *et al.* 2020). These consistent findings among Fabaceae species indicate that codon usage bias in their chloroplast genomes is likely driven by mutational pressures and nucleotide composition constraints.

Such codon bias plays an important role in understanding gene expression regulation, organelle genome evolution and species adaptation. It also aids in optimizing gene expression for transgenic research and synthetic biology applications (Zhou *et al.,* 2018).

Identification and distribution of ssrs in the chloroplast genome of e. tubulosa

A total of 109 simple sequence repeats (SSRs) were identified in the *E. tubulosa* chloroplast genome, including 85 mononucleotide repeats, 10 dinucleotide repeats, 1 trinucleotide repeat and 13 multinucleotide repeats (Table 3). No tetranucleotide or other complex SSR types were detected. Most SSRs (n = 80, 73.40%) were distributed in intergenic spacer (IGS) regions. Additionally, 16 SSRs were located in coding regions of genes such as *rps18*, *ycf4*, *psbC*, *rpoC2*, *rpoB*, *atpB*, *matK*, *ndhF* and ycf1-2, with *ycf1-2* containing four SSR loci. Thirteen SSRs were found in intronic regions, including those of *rpl16*, *petD*, *petB*, *clpP*, *ycf3*, *atpF*, *trnV-UAC*, *rpl2* and *ndhA*. Regionally, 74 SSRs were located in the LSC region, 21 in the SSC and 14 in the IRs.

The predominance of SSRs in IGS regions is consistent with previous findings in other legumes such as *Glycine max* and *Cajanus cajan*, where SSRs are frequently localized to non-coding areas (Zhao *et al.* 2022). The observed pattern-primarily short polyA or polyT mononucleotide repeats-matches earlier studies indicating a strong AT-rich bias in chloroplast SSR motifs (Kuang *et al.* 2011). These SSRs, due to their high polymorphism, maternal inheritance and genomic abundance, are highly valuable molecular markers for species identification, genetic diversity analysis and phylogenetic reconstruction (Du *et al.* 2012).

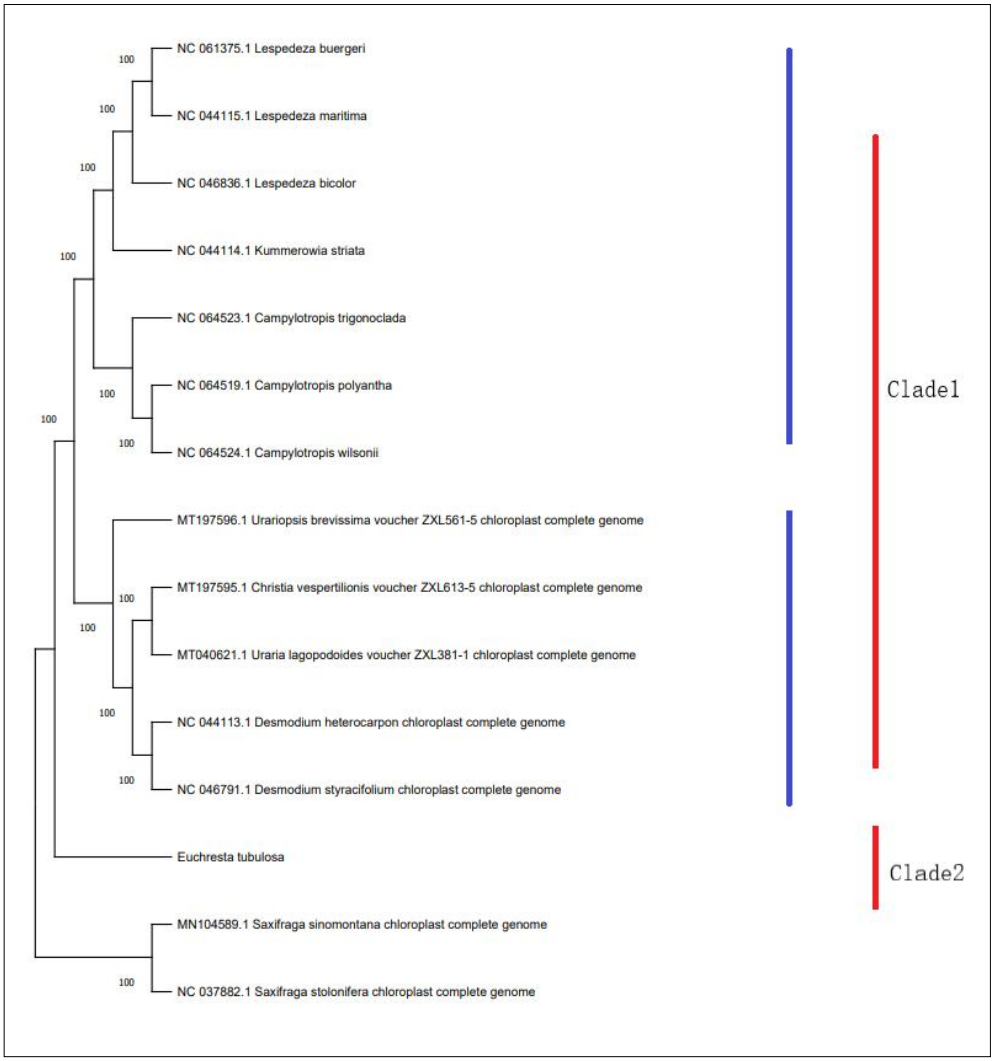

Fig 3: Hylogenetic tree based on chloro plast whole genome sequence.

Table 3: SR information of the chloroplast genome in *Euchresta tubulosa*.

No.	SSRType	SSR	Size	Start	End	Location	Area
1	P1	(A)11	11	768	778	IGS	LSC
2	P1	(A)16	16	1753	1768	rpl16 intron	LSC
3	P1	(T)10	10	2369	2378	rpl16 intron	LSC
4	P1	(T)10	10	3275	3284	IGS	LSC
5	P2	(TA)8	16	3807	3822	IGS	LSC
6	P2	(TA)6	12	4698	4709	IGS	LSC
7	P3	(TAT)6	18	5149	5166	IGS	LSC
8	P1	(T)12	12	6962	6973	IGS	LSC
9	c	(C)11(A)12	23	7928	7950	petD intron	LSC
10	P1	(A)10	10	9702	9711	petB intron	LSC
11	P1	(G)10	10	10827	10836	IGS	LSC
12	P1	(T)11	11	14297	14307	clpP intron	LSC
13	c	(A)17tcaattccatagtatagccgtatgcaatacgca aaagatgcctgtacggttattcaattttatctttg(T)11	96	14697	14792	clpP intron	LSC
14	P2	(AT)9	18	15465	15482	IGS	LSC
15	P1	(T)13	13	16738	16750	rps18	LSC
16	c	(TA)6tgttgttatgttatataatgaaatatatgt(TA)8	58	17143	17200	IGS	LSC
17	P1	(A)12	12	17834	17845	IGS	LSC
18	P1	(T)11	11	18223	18233	IGS	LSC
19	P2	(TA)7	14	18346	18359	IGS	LSC
20	P1	(T)19	19	21668	21686	IGS	LSC
21	P1	(T)12	12	22012	22023	IGS	LSC
22	P1	(A)13	13	22411	22423	IGS	LSC
23	P1	(T)11	11	24902	24912	ycf4	LSC
24	P2	(AT)7	14	25987	26000	IGS	LSC
25	P1	(C)10	10	28132	28141	IGS	LSC
26	P1	(A)10	10	28335	28344	IGS	LSC
27	P1	(T)12	12	28688	28699	IGS	LSC
28	P1	(A)15	15	28940	28954	IGS	LSC
29	P1	(T)10	10	29655	29664	IGS	LSC
30	P1	(A)15	15	29972	29986	IGS	LSC
31	c	(T)14acatttactgatcccagtcttaggaa aaagacacattcttcgg(A)11	68	30497	30564	ycf3 intron	LSC
32	P1	(T)15	15	32358	32372	IGS	LSC
33	P1	(A)11	11	38530	38540	IGS	LSC
34	P1	(A)12	12	39666	39677	IGS	LSC
35	P1	(C)10	10	40634	40643	psbC	LSC
36	P1	(A)10	10	42368	42377	IGS	LSC
37	P1	(T)10	10	42810	42819	IGS	LSC
38	P1	(T)10	10	43078	43087	IGS	LSC
39	P1	(A)14	14	43282	43295	IGS	LSC
40	P1	(T)12	12	43405	43416	IGS	LSC
41	P1	(A)10	10	43702	43711	IGS	LSC
42	P1	(A)10	10	44194	44203	IGS	LSC
43	c	(C)12ttttttttagaggtaagtaaagg ttccgacgaagaaag(A)10	60	45310	45369	IGS	LSC
44	c	(TA)7aat(A)10	27	45960	45986	IGS	LSC
45	P1	(T)10	10	46202	46211	IGS	LSC
46	P1	(A)11	11	46570	46580	IGS	LSC

Table 3: Continue ....

Table 3: Continue ....

47	P1	(A)12	12	47143	47154	IGS	LSC
48	P1	(T)12	12	48376	48387	IGS	LSC
49	c	(A)14tagaaataaaggatagaaaatattaaaac aaaaaggaat(A)10	63	49006	49068	IGS	LSC
50	P1	(T)14	14	50752	50765	IGS	LSC
51	c	(A)11tacttgttatttattatttaaacat gttaaacataaat(A)10	59	50904	50962	IGS	LSC
52	P2	(AT)12	24	52384	52407	IGS	LSC
53	P1	(T)12	12	55087	55098	atpF intron	LSC
54	P1	(T)10	10	56648	56657	IGS	LSC
55	P1	(T)10	10	57055	57064	IGS	LSC
56	c	(T)10(A)10taataataat (A)13gagaccaggtatcttgaaata gaaataaataattgttccgacggaaccttatc ttctaccgaagattggccgttgatacatgatcag gccattcctc(T)10	150	58852	59001	IGS	LSC
57	P1	(T)11	11	59185	59195	rpoC2	LSC
58	P1	(T)11	11	60973	60983	rpoC2	LSC
59	P1	(A)10	10	65161	65170	rpoC1 intron	LSC
60	P1	(A)10	10	68060	68069	rpoB	LSC
61	P1	(T)10	10	68820	68829	rpoB	LSC
62	P1	(T)10	10	69946	69955	IGS	LSC
63	P1	(C)10	10	70866	70875	IGS	LSC
64	P1	(T)14	14	74021	74034	trnV-UAC intron	LSC
65	P1	(T)11	11	74686	74696	IGS	LSC
66	c	(T)10caagtgcagaaacttgggga ccagaagtggtaggatttattctcatattaaa atatatcca(T)10	81	76560	76640	atpB	LSC
67	P1	(T)13	13	76990	77002	IGS	LSC
68	P1	(A)13	13	77110	77122	IGS	LSC
69	P1	(T)11	11	78961	78971	IGS	LSC
70	P2	(TA)6	12	79521	79532	IGS	LSC
71	P1	(A)10	10	79950	79959	IGS	LSC
72	P1	(T)10	10	80289	80298	IGS	LSC
73	P1	(A)10	10	81800	81809	matK	LSC
74	P1	(T)10	10	82444	82453	IGS	LSC
75	P1	(T)11	11	84132	84142	IGS	IRb
76	P2	(TA)7	14	84756	84769	rpl2 intron	IRb
77	c	(TA)7(AT)7	28	94358	94385	IGS	IRb
78	P1	(T)13	13	97830	97842	IGS	IRb
79	P1	(T)13	13	99084	99096	IGS	IRb
80	P1	(G)11	11	100676	100686	IGS	IRb
81	P1	(T)11	11	107964	107974	IGS	IRb
82	P1	(C)10	10	110081	110090	ndhF	SSC
83	P1	(A)13	13	110783	110795	ndhF	SSC
84	P1	(A)10	10	112104	112113	ndhF	SSC
85	P1	(A)12	12	112247	112258	IGS	SSC
86	P1	(T)11	11	112539	112549	IGS	SSC
87	P1	(T)10	10	113341	113350	IGS	SSC
88	P1	(T)10	10	113590	113599	IGS	SSC

Table 3: Continue ....

Table 3: Continue ....

89	P2	(AT)8	16	114630	114645	IGS	SSC
90	P1	(T)11	11	114802	114812	IGS	SSC
91	P1	(T)15	15	116389	116403	IGS	SSC
92	P1	(A)11	11	116743	116753	IGS	SSC
93	P1	(A)13	13	118398	118410	IGS	SSC
94	P1	(A)10	10	119594	119603	ndhA intron	SSC
95	P1	(T)17	17	123370	123386	IGS	SSC
96	P1	(T)10	10	124197	124206	IGS	SSC
97	P1	(T)14	14	124673	124686	IGS	SSC
98	P1	(T)11	11	125647	125657	IGS	SSC
99	c	(A)11ttaatttttttaa(T)11	35	125811	125845	ycf1-2	SSC
100	P1	(T)13	13	125950	125962	ycf1-2	SSC
101	P1	(T)11	11	126347	126357	ycf1-2	SSC
102	P1	(T)13	13	127770	127782	ycf1-2	SSC
103	P1	(A)11	11	130094	130104	IGS	IRa
104	P1	(C)11	11	137382	137392	IGS	IRa
105	P1	(A)13	13	138972	138984	IGS	IRa
106	P1	(A)13	13	140226	140238	IGS	IRa
107	c	(AT)7(TA)7	28	143683	143710	IGS	IRa
108	P2	(AT)7	14	153298	153311	rpl2 intron	IRa
109	P1	(A)11	11	153926	153936	IGS	IRa

Therefore, the SSR loci identified in this study offer potential for future molecular breeding, population genetics and evolutionary biology research in *Euchresta* and related genera.

Phylogenetic analysis of e. tubulosa based on chloroplast genome

To clarify the phylogenetic position of *Euchresta tubulosa*, a phylogenetic tree was constructed using complete chloroplast genome sequences. Species from *Lespedeza* (3), *Kummerowia* (1), *Campylotropis* (3), *Christia* (1), *Urariopsis* (1), *Pseudarthria* (1) and *Saxifraga* (2) were selected as outgroups. The resulting tree showed robust topology with a high bootstrap support of 100%, indicating strong confidence in the inferred relationships.

Within the phylogenetic tree, species of the same genus clustered together, forming two major clades (Clade 1 and Clade 2). Clade 1 was further subdivided into a *Lespedeza* subclade and a *Desmodium* subclade. The *Lespedeza* subclade included *Lespedeza buergeri*, *L. maritima*, *L. bicolor*, *Kummerowia striata*, *Campylotropis trigonoclada*, *C. polyantha* and *C. wilsonii*. The *Desmodium* subclade included *Urariopsis brevissima*, *Christia vespertilionis*, *Uraria lagopodoides*, *Desmodium heterocarpon* and *D. styracifolium*. Notably, *E. tubulosa* formed an independent clade (Clade 2), distinct from the *Desmodieae* tribe, confirming its unique phylogenetic position (Fig 3).

The phylogenetic placement of *E. tubulosa* supports the proposal by Hiroyoshi Ohashi that *Euchresta* may belong to a distinct monogeneric tribe, although it is closely related to the tribe *Sophoreae* (Zhang *et al.,* 2012). These results demonstrate that complete chloroplast genomes are highly effective in resolving taxonomic ambiguities among closely related legume genera (Jiang *et al.,* 2020). Moreover, the strong bootstrap values observed in this study further affirm the utility of chloroplast genome data in molecular systematics.

Our findings not only clarify the systematic position of *Euchresta tubulosa* within *Fabaceae* but also provide foundational data for future research in phylogeny, evolution and taxonomic classification of the genus *Euchresta*.

CONCLUSION

In this study, we sequenced and analyzed the complete chloroplast genome of *Euchresta tubulosa*. The genome exhibits a typical quadripartite structure with a total length of 153,960 bp and 127 annotated genes. Codon usage analysis revealed a strong preference for A/U-ending codons, reflecting an AT-rich bias similar to that observed in other *Fabaceae* species. A total of 109 SSRs were identified, most of which were located in intergenic spacer regions, providing valuable molecular markers for future genetic studies. Phylogenetic analysis based on the complete chloroplast genome clearly distinguished *E. tubulosa* from related genera within the *Desmodieae* tribe, forming an independent clade with strong bootstrap support. This study provides important genomic resources and phylogenetic evidence that enhance our understanding of the evolutionary relationships and systematic position of *Euchresta*, laying a foundation for its conservation, molecular taxonomy and future breeding research.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (32160334).

Disclaimers

The views and conclusions expressed in this article are solely those of the authors and do not necessarily represent the views of their affiliated institutions. The authors are responsible for the accuracy and completeness of the information provided, but do not accept any liability for any direct or indirect losses resulting from the use of this content.

Informed consent

All animal procedures for experiments were approved by the Committee of Experimental Animal care and handling techniques were approved by the University of Animal Care Committee.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this article. No funding or sponsorship influenced the design of the study, data collection, analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Abba, A., Alzahrani, D., Yaradua, S. and Albokhari, E. (2021). Comparative chloroplast genome analysis and evolutionary relationships in some species of asclepiadeae, apocynaceae. *Legume Research-An international Journal.* **44(2):** 145-151. doi: 10.18805/LR-565.

Asaf, S., Khan, A.L., Aaqil, Khan M., Muhammad, I.Q., Kang, S.M., Al-Hosni, K., Jeong, E.J., Lee, K.E. and Lee, I.J. (2017). Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other Glycine species. *PLOS ONE.* **12(8):** e0182281.

Beier, S., Thiel, T., Muench, T., Scholz, U. and Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics.* **33(16):** 2583-2585.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y. and Leon, C. (2010). Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PLoS ONE.* **5(1):** e8613.

Cui, L., Li, Y. and Pan, W. (2012). Research progress in chloroplast genetic engineering. *Biotechnology Bulletin.* **(6):** 1-6.

Doyle, J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* **19**.

Du, Q., Wang, B., Wei, Z., Zhang, D. and Li, B. (2012). Genetic diversity and population structure of chinese white poplar (*Populus tomentosa*) revealed by SSR Markers. *Journal of Heredity.* **103**(6): 853-862.

Duret, L. and Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila* and *Arabidopsis*. *Proceedings of the National Academy of Sciences.* **96(8):** 4482-4487.

Hu, T., Chen, S., Xu, P., Zhou, D. and Zhu, Q. (2023). Complete chloroplast genome and comparative analysis of *entada phaseoloides* (Fabaceae). *Legume Research-An international Journal.* **46(11):** 1431-1438. doi: 10.18805/LRF-744.

Han, F., Lin, M., Zhang, J., Shen, J., Han, R. and Liu, J. (2013). Effects of foliar fertilizer on the growth of Guan'e Shandougen. *Modern Chinese Medicine.* **15(10):** 853-855.

Jiang, W., Guo, M. and Pang, X. (2020). Application of chloroplast genome in identification and phylogenetic studies of medicinal plants. *World Journal of Traditional Chinese Medicine.* **15(5):** 702-708, 716.

Kaila, T., Chaduvla, P.K., Saxena, S., Bahadur, K., Gahukar, S.J., Chaudhury, A., Sharma, T.R., Singh, N.K. and Gaikwad, K. (2016). Chloroplast genome sequence of pigeonpea (*Cajanus cajan* L.) Millspaugh) and *Cajanus scarabaeoides* (L.) Thouars: Genome Organization and Comparison with Other Legumes. *Frontiers in Plant Science.* **7**.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research.* **30(14):** 3059-3066.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. paper presented at the 20th Annual International Conference on Intelligent Systems for Molecular Biology. *Bioinformatics.* **28(12):** 1647-1649.

Kuang, D.Y., Wu, H., Wang, Y-L., Gao, L-M., Zhang, S-Z. and Lu, L. (2011). Complete chloroplast genome sequence of Magnolia kwangsiensis (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome.* **54(8):** 663-673.

Li, H.C., Yuan, D.P. and Liu, Y. (2014). Research progress on chemical constituents in plants of Euchresta J. Benn and their biological activities. *Chinese Traditional and Herbal Drugs.* **45(23):** 3486-3493.

Lin, M., Han, F., Han, R., Zhang, J., Shen, J. and Liu, Z. (2013). Preliminary observation on biological characteristics of *Guan'e Shandougen*. *Hunan Agricultural Sciences.* **(4):** 17-18. 23.

Liu, X.Y., Li, Y., Ji, K.K., Zhu, J., Ling, P., Zhou, T., Fan, L.Y. and Xie, S.Q. (2020). Genome-wide codon usage pattern analysis reveals the correlation between codon usage bias and gene expression in Cuscuta australis. *Genomics.* **112(4):** 2695-2702.

Lo, W.L., Chang, F.R., Liaw, C.C. and Wu, Y.C. (2002). Cytotoxic coumaronochromones from the roots of Euchresta formosana. *Planta Medica.* **68(2):** 146-151.

Mu, Y., Fan, D., Lü, L., Li, X., Lu, J. and Zhang, X. (2022). Codon usage features and phylogenetic comparison of chloroplast genomes in Ranunculaceae and Paeoniaceae. *Bulletin of Botanical Research.* **42(6):** 964-975.

Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols.* **2010(6):** pdb. prot5448.

Ren, Y., Ma, Y., Li, X., Li, X., Yang, G. and Li, P. (2023). Complete chloroplast genome sequence and characteristics analysis of Qingda no.1 alfalfa (*Medicago sativa* L. cv. Qingda no.1). *Czech Journal of Genetics and Plant Breeding.* **59(3):** 160-168.

Rozas, J. (2009). DNA Sequence Polymorphism Analysis Using DnaSP. *Bioinformatics for DNA Sequence Analysis*, Posada D(Ed). Totowa, NJ: Humana Press. pp337-350.

Sun, Y., Ding, B., Guo, C., Han, R., Li, J., Rong, F. and Ding, J. (2020). The complete mitochondrial genome of Holothuria fuscocinerea (Jaeger,1833). *Mitochondrial Dna Part B-Resources.* **5(1):** 33-34.

Tai, T.H. and Tanksley, S.D. (1990). A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. *Plant Molecular Biology Reporter.* **8(4):** 297-303.

Toda, S. and Shirataki, Y. (2006). Inhibitory effect of prenylated flavonoid in *Euchresta japonica* and *Artocarpus heterophyllus* on lipid peroxidation by interaction of hemoglobin and hydrogen peroxide. *Pharmaceutical Biology.* **44(4):** 271-273.

Wang, D. and Li, Z. (1983). Pharmacognostic studies on the herbal medicine *Guan'e Shandougen*. *Chinese Traditional Medicine Bulletin*. **(5):** 8-10.

Wang, X., An, Y. and Xiao, J. (2020). Research progress on the regulation of plastid gene expression and chloroplast development by PPR proteins. *Journal of Plant Physiology.* **56(12):** 2510-2516.

Yin, D., Mu, Y., Li, X., Hua1, M., Li, X., Zhang, P., Zhang, X. and Wang, J. (2024). Comparison of the complete chloroplast genomes in the *astragalus*. *Legume Research-An International Journal.* **47(3):** 420-427. doi: 10.18805/LRF-763.

Yao, S., Cheng, J. and Dong, A. (2017). Thin layer chromatography analysis and preparation of alkaloids from *Guan'e Shandougen (Milkvetch root with tubular calyx)*. *Asia-Pacific Traditional Medicine.* **13(21):** 21-24.

Yang, C., Yao, X., Chen, Z. and Wang, Q. (2022). Structure of the chloroplast genome and phylogenetic relationships among species of *Desmodium in Hedysarum coronarium*. *Bulletin of Botanical Research.* **42(3):** 437-445.

Yen, L.T. and Park, J. (2022). The complete chloroplast genome of *Desmodium styracifolium*. *Mitochondrial DNA Part B.* **7(3):** 513-514.

Zhang, T., Fang, Y., Wang, X., Deng, X., Zhang, X., Hu, S. and Yu, J. (2012). The complete chloroplast and mitochondrial genome sequences of boea hygrometrica: Insights into the evolution of plant organellar genomes. *Plos ONE.* **7(1):** e30531.

Zhao, R., Yin, S., Jiang, C., Xue, J., Liu, C., Cai, X., Xing, Y. and Kang, T. (2022). Comparative analysis of chloroplast genomes of medicinal plants in Aristolochiaceae. *China Journal of Chinese Materia Medica.* **47(11):** 2932-2937.

Zhou, Y., Nie, J., Xiao, L., Hu, Z. and Wang, B. (2018). Comparative chloroplast genome analysis of rhubarb botanical origins and the development of specific identification markers. *Molecules.* **23(11):** 2811.