RESEARCH ARTICLE Legume Research- An International Journal, Volume 46 Issue 11: 1431-1438 (November 2023)

Complete Chloroplast Genome and Comparative Analysis of *Entada phaseoloides* (Fabaceae)

Tingting Hu[1], Si Chen[2], Pei Xu[3], Daoyuan Zhou[1], Qingsong Zhu[4]

10.18805/LRF-744

ABSTRACT

Background: *Entada phaseoloides* (Fabaceae) is a large woody climber that is found widely in southern China and other tropical and subtropical areas worldwide. The genus *Entada* contains ~30 species and *E. phaseoloides* is most commonly found in China. The *E. phaneroneura* and *E. pervillei* are endangered species. Previous studies had focused on medicinal components, transcriptional regulation and nuclear genomes. Phylogenetic relationships within the *Entada* are poorly understood and the chloroplast genome of *Entada* has not been disclosed. In this study, we performed short-read sequencing of *E. phaseoloides* and assembled and analyzed its chloroplast genome.

Methods: Dry specimen leaves of *E. phaseoloides* were subjected to DNA extraction and sequenced using the Illumina Novoseq platform. The chloroplast genome was assembled using Get Organelle, annotated using CPGAVAS2 and Geneious Prime. Long repeat and SSR analysis were performed using the Reputer and MISA software, respectively. Phylogenetic analyses were performed using IQTREE and MrBayes software.

Result: The complete chloroplast genome of *E. phaseoloides* was 159,963 bp in length and had a quadripartite structure with large single copy of 89,972 bp and a small single copy of 19,309 bp separated by inverted repeats of 25,341 bp. A total of 112 genes in *E. phaseoloides* comprised 78 protein-coding genes, 30 transfer RNA genes and 4 ribosomal RNA genes. We carried out phylogenetic analysis based on homologous protein-coding genes among 21 species derived from Fabaceae. We found that the phylogeny was largely congruent with prior hypotheses about the position of *E. phaseoloides* in evolutionary branches. The *E. phaseoloides* had a closer relationship with the *P. africanum*.

Key words: Chloroplast genome, *Entada phaseoloides*, Fabaceae, Phylogeny.

INTRODUCTION

Entada phaseoloides (L.) Merr (1914) is a large woody vine in the Fabaceae family that grows in southern China and other tropical and subtropical areas worldwide. The genus *Entada* contains about 30 species and *E. phaseoloides* is most commonly found in China (Ohashi *et al.*, 2010). *Entada* populations are declining due to over-harvesting and habitat destruction, with *E. phaneroneura* and *E. pervillei* being endangered species (www.iplant.cn/rep/protlist/4?key=Entada). The seeds and stems of *E. phaseoloides* are very large and are often used as ornamental items and the seeds are often used as ornaments. The roots of *Entada* have nitrogen-fixing bacteria that are capable of symbiotic nitrogen fixation. Therefore, *E. phaseoloides* can be used to help restore nitrogen-deficient soils and improve and protect the environment (Diabate *et al.*, 2005). Additionally, the stem of *E. phaseoloides* is widely used in traditional medicine because of its remarkable pharmacological activity. Its main bioactive components are triterpenoid saponin compounds and the representative saponins are oleanane-type triterpenoid saponins containing seven sugar chains (Liao *et al.*, 2020).

There are few phylogenetic studies on *Entada*. In 2003, researchers analyzed the phylogenetic relationships of 134 Mimosoideae species based on *trnL* intron, *trnK* intron and *matK* sequences, demonstrated that the Tribe Mimoseae forms a paraphyletic grade in which are embedded both Acacieae and Ingeae and showed that *Entada* is closely related to *Piptadeniastrum* (Luckow *et al.*, 2003). the chromosomal-level genome of *E. phaseoloides* was reported and the evolution of its triterpenoid saponins biosynthetic genes was revealed (Lin *et al.*, 2022). Although chloroplast genomes of most Fabaceae species have been published in recent years (Kim *et al.*, 2016; Souza *et al.*, 2019; Su *et al.*, 2021), studies on the chloroplast genome and phylogeny of *Entada* are still relatively lacking.

[1]Forestry College, Xinyang Agriculture and Forestry University, Xinyang, 464000, China.
[2]Xinyang Forestry Scientific Research Institute, Xinyang, 464000, China.
[3]College of Life Sciences, China Jiliang University, Hangzhou, 310018, China.
[4]Horticultural College, Xinyang Agriculture and Forestry University, Xinyang, 464000, China.

Corresponding Author: Qingsong Zhu, Horticultural College, Xinyang Agriculture and Forestry University, Xinyang, 464000, China. Email: zqs43@126.com.

How to cite this article: Hu, T., Chen, S., Xu, P., Zhou, D. and Zhu, Q. (2023). Complete Chloroplast Genome and Comparative Analysis of *Entada phaseoloides* (Fabaceae). Legume Research. 46(11): 1431-1438. doi: 10.18805/LRF-744.

Submitted: 22-03-2023 **Accepted:** 11-07-2023 **Online:** 02-08-2023

With the current development of sequencing technology, the study of chloroplast genomes will provide a solid basis for understanding the phylogeny among species. Thus, using data from the Illumina Novoseq platform, we assembled and examined the chloroplast genome of *E. phaseoloides*. Our main objectives are as follows: (1) to analyze the structural features of complete chloroplast genome of *E. phaseoloides*; (2) to analyze simple sequence repeats (SSRs) and repeat sequences; and (3) to infer the phylogenetic position of *E. phaseoloides*.

MATERIALS AND METHODS

Plant material, DNA extraction and sequencing

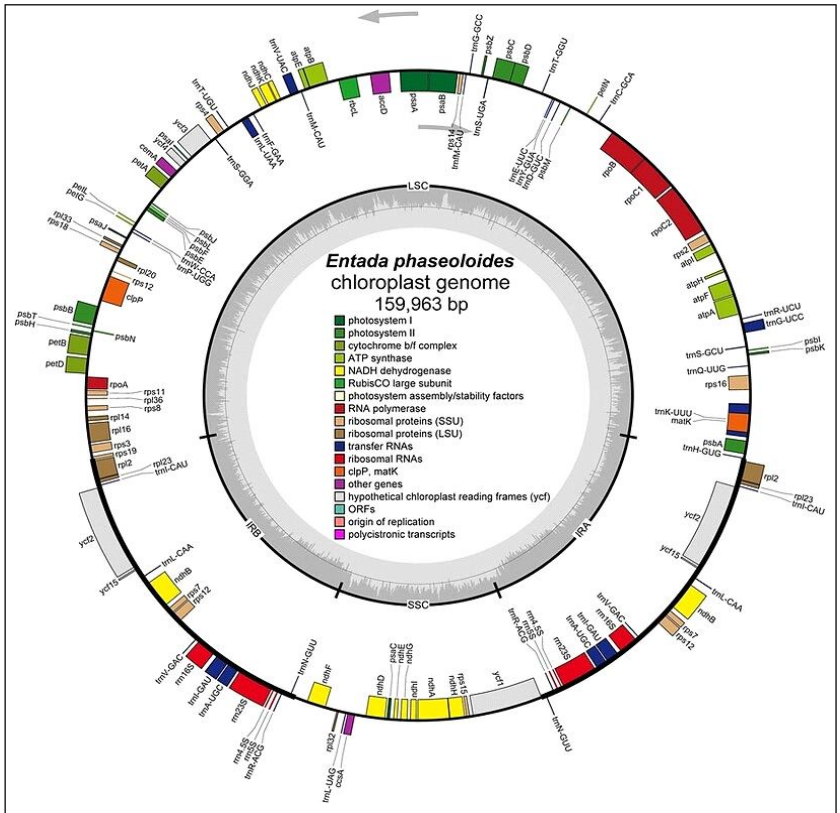
Dry specimen leaves of *E. phaseoloides* were collected from Xishuangbanna, Yunnan Province, China (22.01°N, 100.79°E). Our experimental studies, including the collection of plant material, were in accordance with institutional, national or international guidelines. The sample was deposited at the Herbarium of the Xinyang Agriculture and Forestry University (voucher number: EP001, hutt0716@163.com). Whole genomic DNA was extracted using the CTAB method (Doyle and Doyle, 1987). DNA library of next generation sequencing with an insert size of 300 bp was constructed and sequenced using the Illumina Novoseq 6000 platform, yielded ~5 Gb of raw data and low-quality data were removed to obtain clean data.

Genome assembly and annotation

De novo genome assembly from the clean data of short-reads was accomplished using GetOrganelle v.1.7.5 (Jin *et al.*, 2020). The parameters applied for plastome were '-R 15 -k 21,45,65,85,105 -F embplant_pt'. The chloroplast genome was annotated by using CPGAVAS2 (Shi *et al.*, 2019), PGA (Qu *et al.*, 2019) and Geneious Prime v. 2022.2.2 with a reference genome (*Piptadeniastrum africanum*, GenBank: MZ274093.1). GB2sequin was then used to confirm the annotation results (Tillich *et al.*, 2017) and genome maps were drawn by OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html).

Repeat analysis

The simple sequence repeats (SSRs) were identified using the online website MISA (https://webblast.ipk-gatersleben.de/misa/), including mono-, di-, tri-, tetra-, penta- and hexa-nucleotides with minimum numbers of 10, 5, 4, 3, 3 and 3, respectively (Beier *et al.*, 2017). Additionally, REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) was used to calculate palindromic repeats, forward repeats, reverse repeats and complementary repeats with the following settings: minimal repeat size of 30bp (Kurtz *et al.*, 2001).



The thick lines on the outer complete circle identify the inverted repeat regions (IRa and IRb). The innermost track of the chloroplast genome shows the GC content. Genes on the outside and inside of the map are transcribed in clockwise and counter directions, respectively.

Fig 1: Chloroplast genome map of *E. phaseoloide*.

Relative synonymous codon usage and IR boundary analysis

The relative synonymous codon usage (RSCU) was calculated using an online cloud platform (cloud.genepioneer.com). Furthermore, comparisons between the borders of the IR, SSC and LSC regions were generated using IRscope (Amiryousefi *et al.*, 2018).

Phylogenetic analysis

The chloroplast genomes of 18 Fabaceae species and one Polygalaceae species were downloaded from GenBank. The *Polygala tenuifolia* was used as outgroup. We extracted and aligned 78 common protein-coding genes from the genome annotation files using PhyloSuite v. 1.2.2 (Zhang *et al.*, 2020), then the 78 aligned sequences were concatenated. Based on the matrix of concatenate sequences, a phylogenetic tree was constructed using the Maximum Likelihood (ML) method and Bayesian inference (BI) method. The ML tree was implemented in IQ-TREE v. 2.1.2 (Nguyen *et al.*, 2014) and the best model was inferred from ModleFinder (Kalyaanamoorthy *et al.*, 2017), the bootstrap analysis was performed with 1000 replicates. Tree visualization was achieved in Figtree v. 1.4.3 (https://github.com/rambaut/figtree/releases).

RESULTS AND DISCUSSION

Chloroplast genome characterization of *E. phaseoloides*

The complete chloroplast genome map of *E. phaseoloides* (GenBank number: OQ558908), was a circular molecule with a length of 159,963 bp and the GC content of 36.30% (Fig 1). It had a four-region structure comprising a large single copy, a small single copy and two inverted repeats. The LSC and SSC regions were 89,972 bp and 19,309 bp, respectively, while IRa and IRb regions were 25,341 bp each (Table 1). The length of the coding region was 66,765 bp and represented 41.74% of the whole genome.

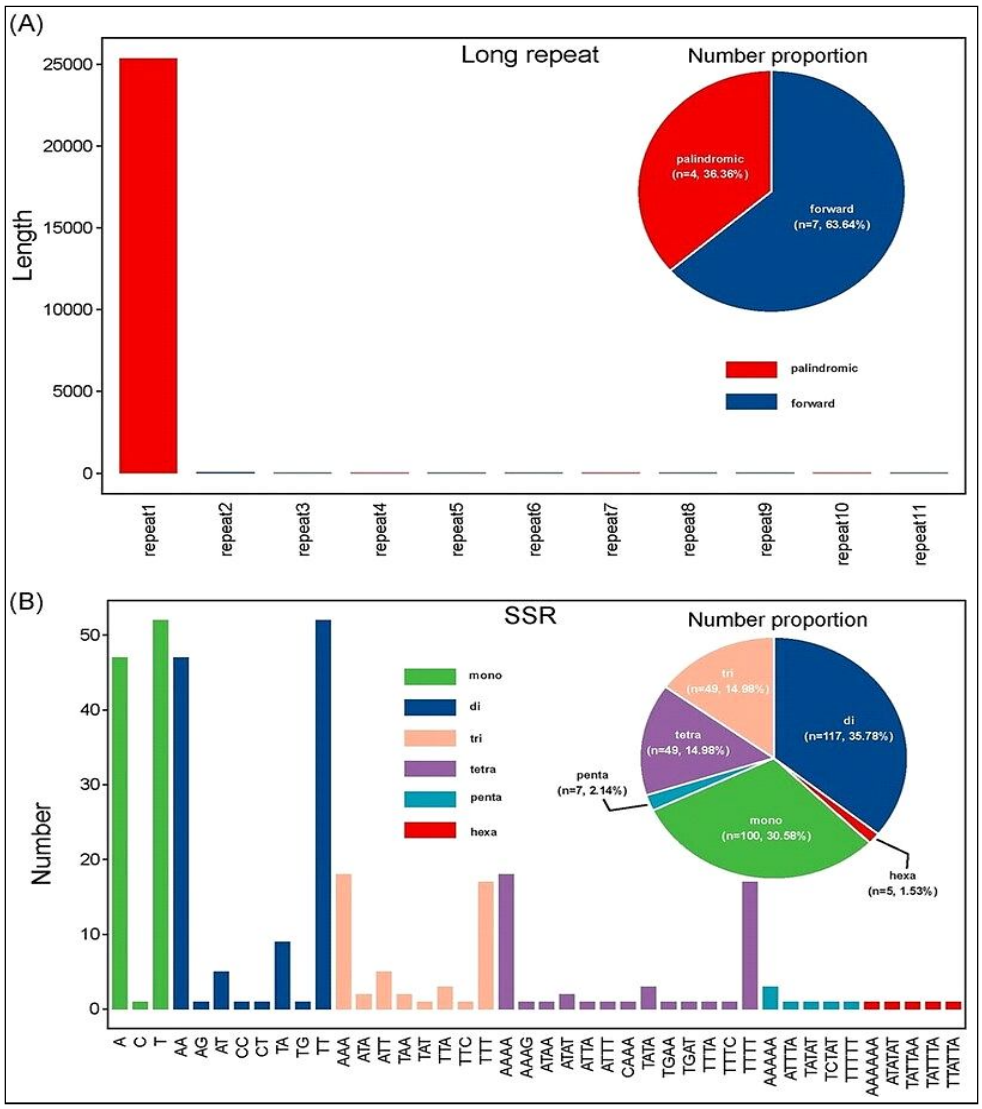


Fig 2: Long repeats (A) and SSR statistics (B) in *E. phaseoloide* chloroplast genome.

The total number of unique genes was 112, containing 78 protein-coding genes and 30 tRNA genes and 4 rRNAs. (Table 2). Among the 78 protein-coding genes, 20 genes contained one intron each (*ndhA*, *ndhB*, *petB*, *petD*, *atpF*, *rpl16*, *rpl2*, *rps16*, *atpF* and *rpoC1*) and two genes (*rps12*, *clpP* and *ycf3*) had two introns each. The gene with the largest intron (2,657 bp) was *trnK*-UUU and the *matK* gene was included in this intron.

Table 1: Summary of the complete chloroplast genome of *E. phaseoloides*.

	Characteristic	*Entada phaseoloides*
Plastid genome	Size (base pair, bp)	159,963
	LSC length (bp)	89,972
	SSC length (bp)	19,309
	IR length (bp)	25,341
	Number of genes	112
	Protein-coding genes	78
	rRNA genes	4
	tRNA genes	30
	GC content	36.30%

Repeat analysis

Repeat sequences play a role in the recombination and variation of chloroplast genomes. This chloroplast genome contained 11 long repeats, including 4 palindromic repeats (36.36%) and 7 forward repeats (63.64%) (Fig 2A). These long repeats were at least 30 bp in length, with the longest being 25,341 bp. In population genetic studies, the number and position of repeated DNA motifs (with 1-6 nucleotides) have been routinely employed for the detection of polymorphisms in cp genomes. In the *E. phaseoloided* chloroplast genome, we identified 327 SSRs and most of them consisted of dinucleotide repeats, with mono-, di-, tri-, tetra-, penta- and hexa-nucleotide SSRs accounted for 30.58%, 35.78%, 14.98%, 14.98%, 2.14% and 0.25% of all SSRs, respectively (Fig 2B).

Relative synonymous codon usage (RSCU)

The 78 protein-coding genes were used to determine the RSCU of the *E. phaseoloided* chloroplast genome (Fig 3A). Leucine was the most frequent amino acid (10.52%), whereas cysteine was the least frequent (1.23%) (Fig 3B). The RSCU values in Table S2 showed that half of the codons

Table 2: Gene function of chloroplast genome of *E. phaseoloides*.

Category	Gene group	Gene name
Photosynthesis	Subunits of photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Subunits of photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	Subunits of NADH dehydrogenase	*ndhA*, ndhB*(2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	Subunits of cytochrome b/f complex	*petA, petB*, petD*, petG, petL, petN*
	Subunits of ATP synthase	*atpA, atpB, atpE, atpF*, atpH, atpI*
	Large subunit of rubisco	*rbcL*
Self-replication	Proteins of large ribosomal subunit	*rpl14, rpl16*, rpl2*(2), rpl20, rpl23(2), rpl32, rpl33, rpl36*
	Proteins of small ribosomal subunit	*rps11, rps12**(2), rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7(2), rps8*
	Subunits of RNA polymerase	*rpoA, rpoB, rpoC1*, rpoC2*
	Ribosomal RNAs	*rrn16S(2), rrn23S(2), rrn4.5S(2), rrn5S(2)*
	Transfer RNAs	*trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC*, trnH-GUG, trnI-CAU(2), trnI-GAU*(2), trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG, trnM-CAU, trnN-GUU(2), trnP-UGG, 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*
Other genes	Maturase	*matK*
	Protease	*clpP***
	Envelope membrane protein	*cemA*
	Acetyl-CoA carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsA*
	Conserved hypothetical chloroplast ORF	*ycf1, ycf15(2), ycf2(2), ycf3**, ycf4*

Notes: Gene*: Gene with one intron; Gene**: Gene with two introns; #Gene: Pseudo gene; Gene (2): Number of copies of multi-copy genes.

were > 1 (Fig 3C). It could be seen from the data that tryptophan (UGG) and methionine (AUG) with codon usage bias had an RSCU value of 1.

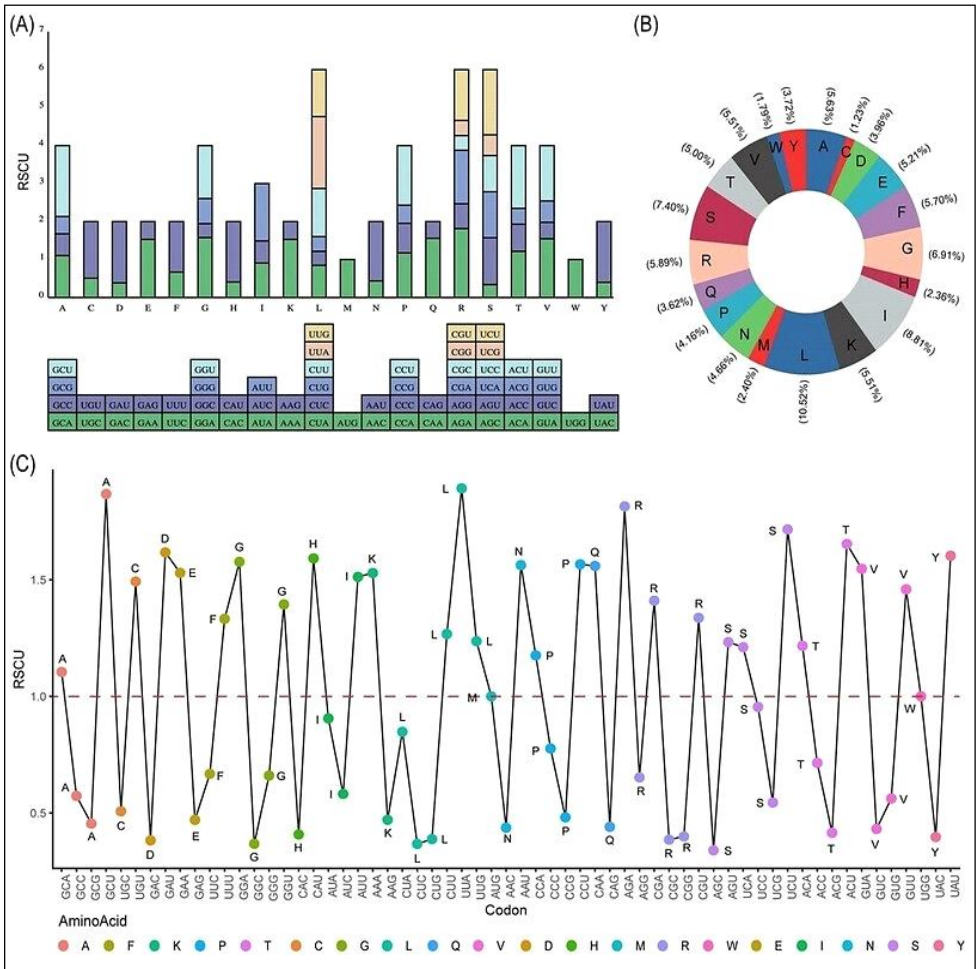
IR boundaries analysis

The comparisons between IR-SC boundaries for the 19 Mimoseae species (Fig 4). In general, the variation in length of the two LSC/SSC regions was lower than that of the IRa/IRb regions. Compared to the chloroplast genomes of other Mimoseae species, the chloroplast genome of *E. phaseoloides* showed a contraction of the IR region and an expansion of the SSC region. The *trnH* gene showed variation in its location in the LSC region. The *ycf1* gene was located within the SSC/IRa boundary in 19 Mimosaceae species, but the length of the expansion of *ycf1* gene into the IRa region in *E. phaseoloides* was 37 bp. Except for *Cylicodiscus gabunensis*, the *ndhF* genes of other species were located in the SSC region. Variations in the location of the *rps19* gene in the IR/LSC border also occurred in the cp genomes. The *rps19* gene spanned the border of LSC/IRb. The *E. phaseoloides*, *Leucaena trichandra* and *Prosopis farcta* had two copies of the *rpl2* gene located in the inverted repeat regions.

Phylogenetic analysis

We used the 78 protein-coding genes for phylogenetic analysis and selected 27 angiosperm species, including 20 Fabaceae species and *Polygala tenuifolia* of Polygalaceae as outgroup. Phylogenetic analysis was performed by maximum likelihood and Bayesian inference. The two phylogenetic trees were topologically similar, with the majority of nodes having 100% bootstrap (BP) values and 1.00 Bayesian posterior probabilities (PP). We found that the phylogeny was largely congruent with prior hypotheses about the position of *E. phaseoloides* in evolutionary branches. The *E. phaseoloides* and *P. africanum* were more closely related and belong to the same group (Fig 5).

This study presents the first chloroplast genome from *E. phaseoloides*. The length of the cp genome in *E. phaseoloides* was similar to that seen in the cp genome of other Mimoseae species. A typical angiosperm chloroplast genome consists of 113 genes, including 79 protein-coding genes, 30 tRNA genes and four rRNA genes (Wicke *et al.*, 2011)



(A) RSCU statistics. The codons are represented by different colors in the histogram.
(B) Proportion of 20 amino acids. (C) RSCU value of each triplet codon.

Fig 3: Codon content of 20 amino acids in all protein-coding genes of the *E. phaseoloide* chloroplast genome.

The *E. phaseoloides* chloroplast genome had a similar number of genes (112 genes), including 78 protein-coding genes, 30 tRNA genes and 4 rRNA genes.

Codons encoding the leucine were the most common in the chloroplast genome of *E. phaseoloides*, while those encoding cysteine were the least common. These findings have also been reported in the chloroplast genome of *Balanites aegyptiaca*. Several reports have shown the importance of chloroplast SSRs as reliable molecular markers to discriminate specimens at lower taxonomic levels and study population structure. The *E. phaseoloides* chloroplast genome had 327 SSRs. Dinucleotide AA/TT SSRs were the most frequent. Therefore, we recommended the use of the chloroplast genome for the development of SSR sites and the study of the population genetic level in *E. phaseoloides*.

Although the plastid genome is conserved in angiosperm plants as previously reported, several studies have reported variation in the size and boundaries among IR/LSC and IR/SSC regions and variation in gene location (Al-Juhani *et al*., 2022; Ruhsam *et al*., 2016). In the present study, comparisons between IR-LSC and IR-SSC boundaries in the 19 complete chloroplast genomes of Mimoseae showed clear variation in the inverted repeat region in chloroplast genomes and significant expansion in the IR region in the chloroplast genome of *E. phaseoloides*.

Chloroplast genomes are composed of many efficient genes that can solve phylogenetic problems at different levels of angiosperm taxonomy (Al-Juhani *et al*., 2022; Dong *et al*., 2017). In this study, we found that *E. phaseoloides* was more closely related to *P. africanum*.

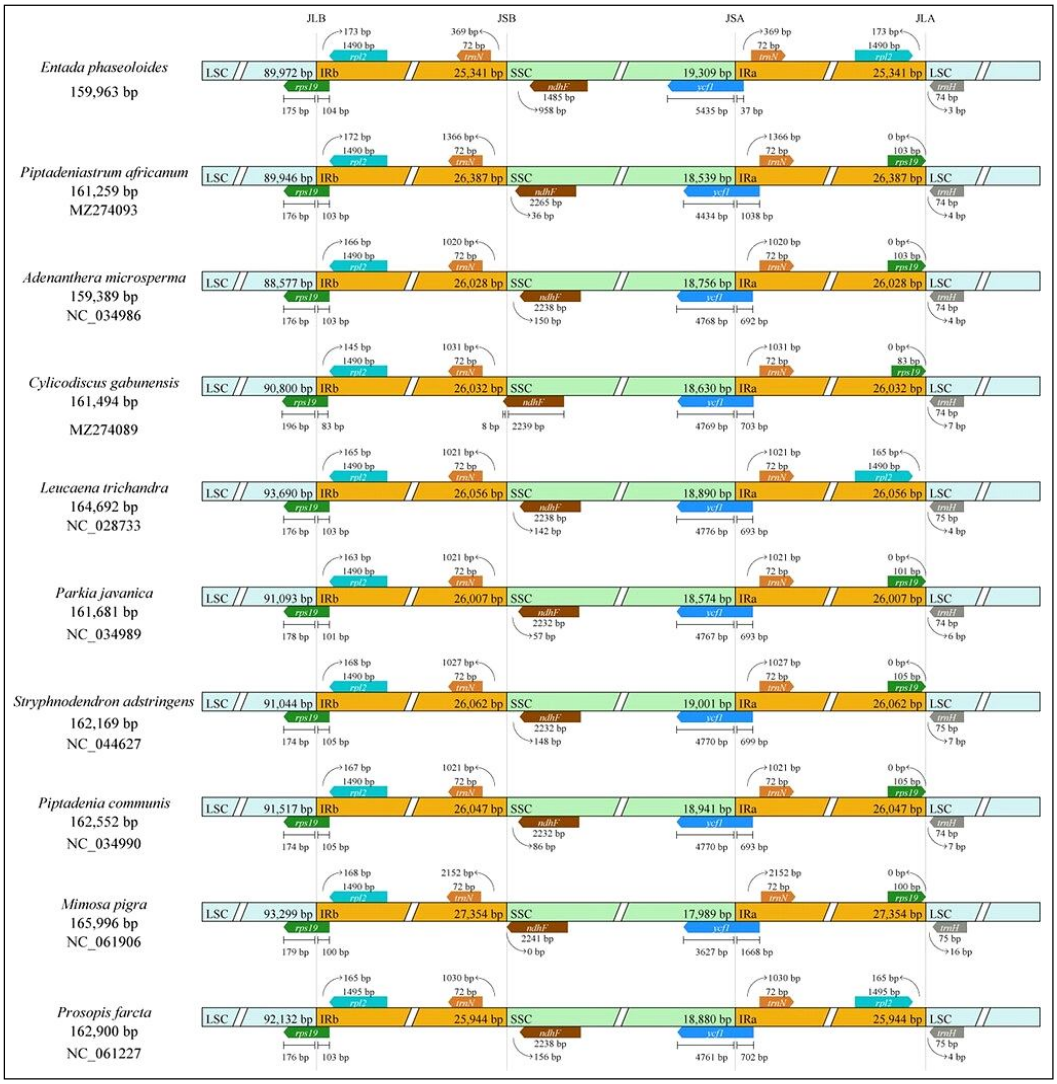


Fig 4: Comparison of the junction sites between the long single copy (LSC, light blue), short single copy (SSC, light green) and inverted repeat (IRa and IRb, orange) regions among the ten Mimoseae chloroplast genomes. JLB (IRb/LSC), JSB (IRb/SSC) JSA (SSC/IRa) and JLA (IRa/LSC) denote the junction sites between each corresponding regions on the chloroplast genome.

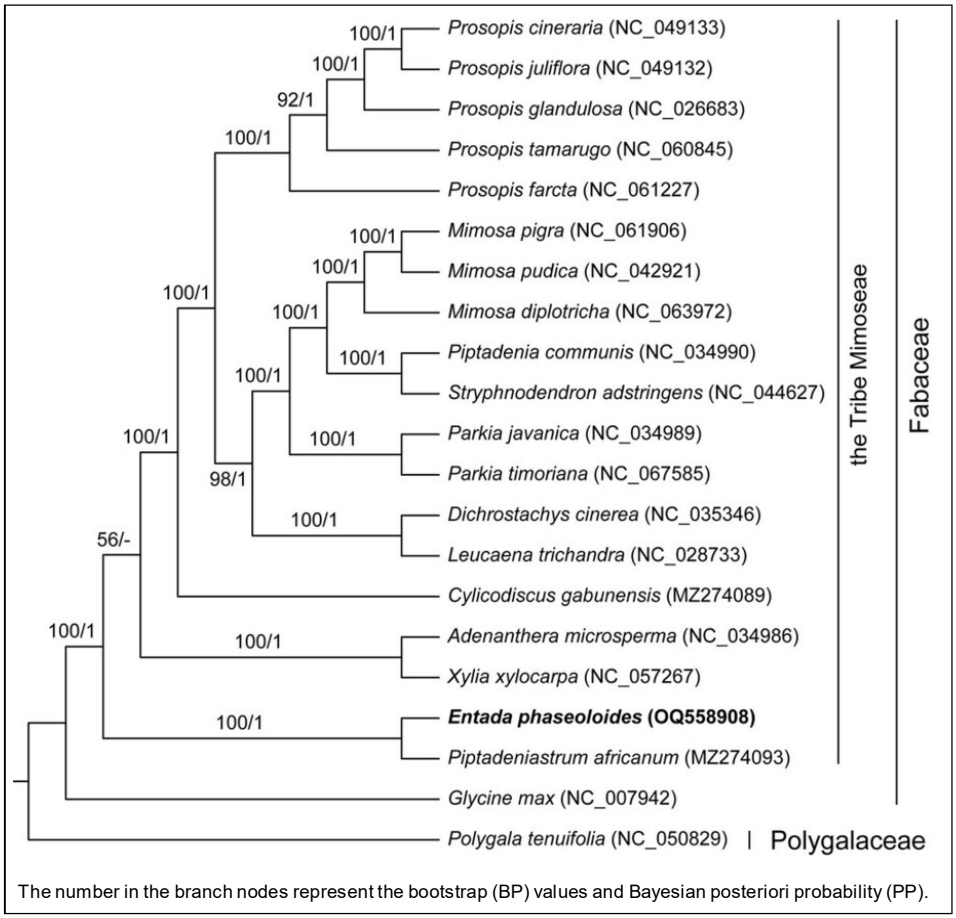


Fig 5: The phylogenetic relationships among *E. phaseoloide* using the maximum likelihood (ML) and Bayesian Inference (BI) methods.

CONCLUSION

In this research, we assembled the complete chloroplast genome of *E. phaseoloides* with 159,963 bp for the first time, consisting of the LSC region of 89,972 bp, the SSC region of 19,309 bp and two copies of IR regions of 25,341 bp. The chloroplast genome contains 112 unique genes, which are 78 PCGs, 30 tRNA genes and 4 rRNA genes. Gene contents and orientation are similar to those found in the chloroplast genome of other Mimoseae species. This study also revealed the distribution of repeated structures and microsatellites along the chloroplast genome of *E. phaseoloides*. We also generated important genomic resources for Mimoseae and *Entada*. Based on 78 protein-coding genes, the phylogenetic tree for 19 Mimoseae species was constructed with good supports. Using a 100% BS and 1.00 PP score, we discovered that the *E. phaseoloides* is more closely related to the *P. africanum*. These results will not only help to clarify the evolutionary study of the *Entada*, but also help to explore more genetic information and better utilize *E. phaseoloides*.

Author contributions

Tingting Hu was primarily responsible for the design of the experiment; Si Chen and Daoyuan Zhou participated in genome assembly and annotation work. Tingting Hu and Qingsong Zhu analyzed and interpreted the data. All authors have read and approved the final manuscript.

Funding

This work was supported by the Youth Science Funds of Xinyang Agriculture and Forestry University(20200109).

Data availability statement

The genome sequence data that support the findings of this study are available in GenBank of NCBI (http://www.ncbi.nlm.nih.gov/) under the accession no. OQ558908. The associated BioProject, BioSample and SRA numbers are PRJNA940368, SAMN33564936 and SRR23684266, respectively.

Conflict of interest: None.

REFERENCES

Al-Juhani, W.S., Alharbi, S.A., Al Aboud, N.M., Aljohani, A.Y. (2022). Complete chloroplast genome of the desert date (*Balanites aegyptiaca* (L.) Del. comparative analysis and phylogenetic relationships among the members of Zygophyllaceae. BMC Genomics. 23(1): 626.

Amiryousefi, A., Hyvönen, J., Poczai, P. (2018). IRscope: An online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 34(17): 3030-3031.

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

Diabate, M., Munive, A., De Faria, S.M., Ba, A., Dreyfus, B., Galiana, A. (2005). Occurrence of nodulation in unexplored leguminous trees native to the West African tropical rainforest and inoculation response of native species useful in reforestation. New Phytologist. 166(1): 231-239.

Dong, W., Xu, C., Li, W., Xie, X., Lu, Y., Liu, Y., Jin, X., Suo, Z. (2017). Phylogenetic Resolution in Juglans Based on Complete Chloroplast Genomes and Nuclear DNA Sequences. Frontiers in Plant Science. 8: 1148.

Doyle, J.J., Doyle, J.L. (1987). A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissues. Phytochem Bull. 19: 11-15.

Jin, J., Yu, W., Yang, J., Song, Y., dePamphilis, C.W., Yi, T., Li, D. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology. 21(1): 241.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. Nature Methods. 14(6): 587-589.

Kim, N.R., Kim, K., Lee, S.C., Lee, J.H., Cho, S.H., Yu, Y., Kim, Y.D., Yang, T.J. (2016). The complete chloroplast genomes of two *Wisteria* species, *W. floribunda* and *W. sinensis* (Fabaceae). Mitochondrial DNA A DNA Mapp Seq Anal. 27(6): 4353-4354.

Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research. 29(22): 4633-4642.

Liao, W., Mei, Z., Miao, L., Liu, P., Gao, R. (2020). Comparative transcriptome analysis of root, stem and leaf tissues of *Entada phaseoloides* reveals potential genes involved in triterpenoid saponin biosynthesis. BMC Genomics. 21(1): 639.

Lin, M., Jian, J.B., Zhou, Z.Q., Chen, C.H., Wang, W., Xiong, H., Mei, Z.N. (2022). Chromosome-level genome of Entada phaseoloides provides insights into genome evolution and biosynthesis of triterpenoid saponins. Mol. Ecol. Resour. 22(8): 3049-3067.

Luckow, M., Miller, J., Murphy, D., Livshultz, T. (2003). A phylogenetic analysis of Mimosoideae (Leguminosae) based on chloroplast DNA sequence. pp: 197-220.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution. 32(1): 268-274.

Ohashi, H., Huang, T.C., Ohashi, K. (2010). *Entada* (Leguminosae subfam. Mimosoideae) of Taiwan. Taiwania. 55: 43-53.

Qu, X., Moore, M.J., Li, D., Yi, T. (2019). PGA: A software package for rapid, accurate and flexible batch annotation of plastomes. Plant Methods. 15(1): 50.

Ruhsam, M., Clark, A., Finger, A., Wulff, A.S., Mill, R.R., Thomas, P.I., Gardner, M.F., Gaudeul, M., Ennos, R.A., Hollingsworth, P.M. (2016). Hidden in plain view: Cryptic diversity in the emblematic Araucaria of New Caledonia. Am. J. Bot. 103(5): 888-898.

Shi, L.C., Chen, H.M., Jiang, M., Wang, L.Q., Wu, X., Huang, L.F., Liu, C. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Research. 47(W1): W65-W73.

Souza, U.J.B., Nunes, R., Targueta, C.P., Diniz-Filho, J.A.F., Telles, M.P.C. (2019). The complete chloroplast genome of *Stryphnodendron adstringens* (Leguminosae-Caesal pinioideae): Comparative analysis with related Mimosoid species. Scientific Reports. 9(1): 14206.

Su, C., Duan, L., Liu, P., Liu, J., Chang, Z., Wen, J. (2021). Chloroplast phylogenomics and character evolution of eastern Asian *Astragalus* (Leguminosae): Tackling the phylogenetic structure of the largest genus of flowering plants in Asia. Molecular Phylogenetics and Evolution. 156: 107025.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., Greiner, S. (2017). GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Research. 45(W1): W6-W11.

Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Müller, K.F., Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. Plant Molecular Biology. 76(3-5): 273-297.

Zhang, D., Gao, F., Jakovliæ, I., Zou, H., Zhang, J., Li, W.X., Wang, G.T. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Molecular Ecology Resources. 20(1): 348-355.