A species unique to China—The complete chloroplast genome sequence of Eomecon chionantha Hance and phylogenetic relationships analysis

Linfang Huang (lfhuang@implad.ac.cn)
CAMS PUMC IMPLAD: Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development
Yujing Miao
CAMS PUMC IMPLAD: Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development
Guoshuai Zhang
CAMS PUMC IMPLAD: Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development
Xinke Zhang
CAMS PUMC IMPLAD: Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development
Huihui Zhang
CAMS PUMC IMPLAD: Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development
Junbo Xie
CAMS PUMC IMPLAD: Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development
Rui Zeng
Southwest Minzu University
Zhi Zhang
CAMS PUMC IMPLAD: Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development

Research Article

Keywords: Eomecon chionantha Hance, Chloroplast genome, Phylogenetic relationship, Endemic species of China

Posted Date: May 12th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2881998/v1

Abstract

Eomecon chionantha Hance (*EC*) is a unique species in China with high medicinal value. Ethnic minorities in China, such as the Miao and Tujia ethnic groups, have a long history of using blood herbs for treatment, and *EC* has been used for this purpose for centuries. However, despite its long history of use, we have no knowledge of the chloroplast genome of *EC* Therefore, this study reports *EC*'s complete chloroplast genome information to better develop and protect this unique plant species in China. The complete chloroplast genomic information indicates that *EC*'s chloroplast DNA (CPDNA) (178,808 bp) contains 99 protein-coding genes, including 8 rRNAs, 37 tRNAs. We have discovered 54 SSRs, most of which are single nucleotide adenine-thymidine (A-T) repeats. Comparative analysis of codons, repeats, and genomic sequences have found that the CPDNA of *EC* is highly conserved. According to our phylogenetic tree results, *EC* is closely related to four species. Through K2-P analysis, we have identified five hypervariable regions, including *ycf*4-*cem*A, *ycf*3-*trn*S-GGA, *trn*C-GCA-*pet*N, *rpl*32-*trn*L-UAG, and *psb*I-*trn*S-UGA. In summary, this study has reported, for the first time, the complete chloroplast genome of the unique single genus plant *EC* in China. This provides a more scientific basis for further development and utilization of this species and is conducive to an in-depth understanding of plant species diversity from a genomic perspective.

1. Introduction

Eomecon chionantha Hance (*EC*) (GenBank OQ737801)is a perennial herb of the poppy family, unique to China as the only species in *Eomecon* genus. In ethnic minorities such as the Tujia and Miao nationalities in China, *EC* is commonly used as a whole herb medicine, with therapeutic effects for throat swelling and pain, traumatic injuries, and hemostasis, as well as insecticidal properties(Lu et al. 2020). Modern pharmacological studies have found that *EC* is rich in alkaloids, triterpenes, and lipids, which have anti-inflammatory, antibacterial, enzyme-inhibitory, and anticancer effects. *EC* grows at an altitude of 1,400-1,800m and is widely distributed in southern China, including Guangxi, Sichuan, Guizhou, and other regions(Peng et al. 2011; Yang et al. 2023c; Yang et al. 2023d).

The chloroplast (CP) is an essential organelle for green plants and plays a crucial role in photosynthesis. Research into the complete CP genome of cyanobacteria and *Arabidopsis* indicates that plant CPs originated from cyanobacteria via endo symbiogenesis(Raven and Allen 2003). CPDNA is usually double-stranded in higher plants, although recent studies have found that a portion of the CPDNA is circular. CPDNA comprises four typical regions: two inverted repeats (IRA and IRB) separated by a large and small single-copy region (LSC and SSC)(Boltenkov and Artyukova 2023; Wonok et al. 2023; Yang et al. 2023a; Yang et al. 2023b). An in-depth understanding of the CP genome can aid in comprehending and safeguarding valuable medicinal plants. Variations in the sequence and structure of CP genomes indicate significant differences among the species. The research of the genus level can help us better understand each plant's role in phylogeny. The variation in CP gene sequences provides substantial evidence for understanding the relationship between different species(Ahmed et al. 2021; Bhaskar et al. 2021; Habib et al. 2021; Toshino et al. 2022).

We sequenced, assembled, and annotated the complete CP genome in order to gain a deeper understanding of this Chinese endemic species of plant, *EC* By conducting sequence difference analysis, repeat analysis, and codon analysis, we identified the genomic connections and differences between this genus and related species. This study enriches plant genome research, provides a more comprehensive reference for analyzing plant diversity, and fills the gap in the CP genome of *EC*.

2. Materials and methods

2.1 Plant material and DNA extraction

The *EC* samples were collected from Leigong Mountain(26°37' N, 108°15'E), Guizhou Province, China. It was identified by Huang Linfang, a researcher at the Institute of Medicinal Plants, Chinese Academy of Medical Sciences, as *EC* and stored in an ultra-low temperature freezer (-80 ℃). The CTAB method extracts total DNA from the *EC (*Arseneau et al. 2017).

2.2 Sequencing, assembly and annotation

Genomic DNA sequencing was carried out using the Illumina Hiseq 2500 platform (Novogene Technologies, Inc., Beijing, China). We selected NOVOPlasty v.3.8.3 to assemble high-quality reads(Chen et al. 2022a; Chen et al. 2022b; Giorgashvili et al. 2022; Godeiro and Zhang 2021), annotated the genome using CpGAVAS2(Miao et al. 2022; Shi et al. 2019), and finally drew the genome map with OGDRAW.

2.3 Codon Usage and Repeat Sequences

The *EC* fasta file and the Codon W1.4.2 program were in the same folder, obtain the usage frequency of the codon through Codon W1.4.2 and output the usage of the codon based on the RSCU value in the form of a picture through Rstudio(Bi et al. 2023). The RSCU value serves as a measure of the disparity between observed and expected usage of codons. A value of RSCU greater than 1 suggests that the frequency of codon usage exceeds the predicted level(Desingu et al. 2022; Li et al. 2023b; Tang et al. 2022; Wang et al. 2023b). VMATCH was used to find dispersed repeat sequences, and PREPuter was used to predict the four kinds of repeat sequences: the forward (F), palindromic (P), reverse (R), and complement (C).

2.4 Genome comparison

The gb files were imported to the IRscope(https://irscope.shinyapps.io/irapp/), downloaded from NCBI, of the 15 most closely related species based on the maximum likelihood tree (ML) results constructed from phylogenetic relationships(Kamra et al. 2023; Shi et al. 2023; Xiao et al. 2022; Xie et al. 2023), Analyzed the quadripartite junction sites of chloroplast genes and obtain a comparison.(Amiryousefi et al. 2018; Zheng et al. 2020).

Since *EC* is an endemic plant of a single genus and single species endemic to China, four species were selected for plastid comparative analysis with *EC* in order to explore the differences between *EC* and the genomes of neighboring species according to the results of phylogenetic analysis (*Macleaya microcarpa*, *Coreanomecon hylomeconoides*, *Hylomecon japonica*, *Chelidonium majus*), five species fasta files were imported into Mega11 and then aligned using MAFFT, available at https://mafft.cbrc.jp/alignment/server/index.html. The resulting alignment file was imported into DNAsp5.0 (set the parameters Window Length to 600 sites, Step size to 200 sites)(Çelik et al. 2022; Jin et al. 2023; Wang et al. 2022; Yang et al. 2022).

2.5 Phylogenetic analysis and intergenic spacer regions (IGS) analysis

In order to better understand the phylogenetic process of *EC*, we selected all species of the poppy family. *Nigella damascene*(NC041537) is an outgroup that reconstructs a phylogenetic tree. Except for *EC*, information for other species was downloaded from NCBI. After extracting 80 CDSs, they were aligned and linked with Phylsuite and MAFFT (v 7.450). The maximum likelihood (ML) method of RAxML v8.2.8 was utilized to construct a phylogenetic tree, with 1,000 bootstrap analyses performed(Zhang et al. 2023).

The intergenic spacer regions (IGS) in these species were analyzed by calculating pairwise distances using the Kimura two-parameter (K2-p) model, which was included in the EMBOSS 6.5.7 package.

2.6 Non-synonymous (Ka) and synonymous (Ks) substitution rate analysis

To investigate the evolution of protein-coding genes in five species, we used PhyloSuite to extract a total of 80 protein genes. We processed these genes through paml4.9 to obtain the Ka/Ks value for each protein (Kuang et al. 2022; Li et al. 2022; Pavan-Kumar et al. 2022). We then displayed these values in the form of a box graph using Rstudio. Finally, we ranked the proteins by their Ka/Ks value and displayed the top-ranked proteins.

3. Results

3.1 The chloroplast genomes features of *EC.*

After the chloroplast sequencing, assembly, and annotation, we obtained complete information about the chloroplast of *EC*. The total length of the CPDNA of *EC* is 178,808 bp. The CP region of a plant is generally divided into four parts. The LSC of *EC* is the longest, with a length of 88,300 bp. The SSC region is the shortest, with a length of 148 bp(Fig. 1). The IRa and IRb regions are 45,180 bp in length.

The cp genome of *EC* contains 99 protein-coding genes (PCG), 37 transfer RNA (tRNA) genes, and 8 ribosomal RNA (rRNA) genes. Among them, 4 rRNAs(*rrn*16s, *rrn*23s, *rrn*4.5s, and *rrn*5s), 6 tRNAs (*trn*L-CAA, *trn*V-GAC, *trn*A-UGC, *trn*R-ACG, *trn*N-GUU, *trn*L-UAG), and 19 encoding-proteins (*ndh*A, *ndh*B, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *psa*C, *rpl*2, *rpl*23, *rpl*32, *rps*15, *rps*7, *ccs*A, *ycf*1, *ycf*15, *ycf*2) contain two repeating units, the tRNA (*trn*E-UUC) contains three repeating units, and *trn*M-CAU contains four repeating units (Table 2). For introns and exons of genes, 21 genes contain an intron, including 7 tRNAs (*trn*K-UUU, *trn*S-CGA, *trn*L-UAA, *trn*E-UUC×2 *trn*A-UGC×2) and 14 protein-coding genes (*rps*16, *atp*F, *rpo*C1, *pet*B, *pet*D, *rpo*A, *rpl*16, *rpl*22, *rpl*2×2 ndhB×2. *ndh*A×2), one protein-coding gene (*ycf*3) contain two introns.

Table 1
Starting and ending sites of the *EC* cp genome

Species	Name	Start (bp)	End(bp)
Eomecon chionantha	IR	88,301	133,480
Eomecon chionantha	IR	133,629	178,808
Eomecon chionantha	SSC	133,481	133,628
Eomecon chionantha	LSC	1	88,300

Table 2
Gene composition of *EC* chloroplast genome

Category of genes	Group of genes	Name of genes
RNA	rRNA	*rrn*16s*×2*,*rrn*23s*×2*,*rrn*4.5s*×2*,*rrn*5s*×2*
	tRNA	*trnH*-GUG,*trnK*-UUU,*trnQ*-UUG,*trnS*-GCU,*trnS*-CGA,*trnR*-UCU,*trnC*-GCA,*trnD*-GUC,*trnY*-GUA, *trnT*-GGU,*trnS*-UGA,*trnG*-GCC,*trnS*-GGA,*trnT*-UGU,*trnL*-UAA,*trnF*-GAA,*trnW*-CCA,*trnP*-UGG, *trnL*-CAA*×2*,*trnV*-GAC*×2*,*trnA*-UGC*×2*,*trnR*-ACG*×2*,*trnN*-GUU*×2*,*trnL*-UAG*×2, trnE*-UUC*×3, trnM*-CAU*×4*
photosynthesis	Subunits of ATP synthase	*atp*A, *atp*B, *atp*E, *atp*F, *atp*H, *atp*I
	Subunits of photosystem II	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*N, *psb*T, *psb*Z, *ycf*3
	Subunits of NADH-dehydrogenase	*ndh*A×2, *ndh*B×2, *ndh*C, *ndh*D×2, *ndh*E×2, *ndh*F×2, *ndh*G×2, *ndh*H×2, *ndh*I×2, *ndh*J, *ndh*K
	Subunits of cytochrome b/f complex	*pet*A, *pet*B, *pet*D, *pet*G, *pet*L, *pet*N
	Subunits of photosystem I	*psa*A, *psa*B, *psa*C×2, *psa*I, *psa*J
	Subunit of rubisco	*rbc*L
Self replication	Large subunit of ribosome	*rpl*14, *rpl*16, *rpl*2×2, *rpl*20, *rpl*22, *rpl*23×2, *rpl*32×2, *rpl*33, *rpl*36
	DNA dependent RNA polymerase	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2
	Small subunit of ribosome	*rps*11, *rps*12×2, *rps*14, *rps*15×2, *rps*16, *rps*18, *rps*19, *rps*2, *rps*3, *rps*4, *rps*7×2, *rps*8
Other genes	Subunit of Acetyl-CoA-carboxylase	*acc*D
	c-type cytochrom synthesis gene	*ccs*A×2
	Envelop membrane protein	*cem*A
	Protease	*clp*P
	Translational initiation factor	*inf*A
	Maturase	*mat*K
Unkown	Conserved open reading frames	*ycf*1×2, *ycf*15×2, *ycf*2×2, *ycf*4

Table 3
Genes with intron in the *EC* chloroplast genome and length of exons and introns.

Gene	Strand	Start	End	ExonI	IntronI	ExonII	IntronII	ExonII
trnK-UUU	-	1,932	4,471	37	2,468	35		
*rps*16	-	5,520	6,,624	40	840	225		
*trn*S-CGA	+	10,755	11,522	31	675	62		
*atp*F	-	13,432	14,728	145	742	410		
*rpo*C1	-	22,693	25,507	453	754	1,608		
*ycf*3	-	45,437	47,386	124	690	230	753	153
*trn*L-UAA	+	50,394	50,984	35	506	50		
*pet*B	+	78,844	80,274	6	783	642		
*pet*D	+	80,468	81,692	8	721	496		
*rpo*A	-	81,916	82,981	939	40	87		
*rpl*16	-	85,248	86,652	9	997	399		
*rpl*22	-	87,456	88,034	431	57	91		
*rpl*2	-	88,439	89,919	385	662	434		
*ndh*B	-	98,609	100,841	775	700	758		
*trn*E-UUC	+	106,423	107,434	32	940	40		
*trn*A-UGC	+	107,499	108,372	37	801	36		
*ndh*A	+	120,586	122,683	553	1,006	539		
*ndh*A	-	144,426	146,523	553	1,006	539		
*trn*A-UGC	-	158,737	159,610	37	801	36		
*trn*E-UUC	-	159,675	160,686	32	940	40		
*ndh*B	+	166,268	168,500	775	700	758		
*rpl*2	+	177,190	178,670	385	662	434		

3.2 Repetitive sequence and codon usage analysis

By studying the CP genome of *EC*, it was revealed that a total of 54 SSRs were detected. Out of these SSRs, only one was a dinucleotide and trinucleotide, while the largest number was a single nucleotide (A/T)(Luo et al. 2023; Nemati et al. 2023; Pei et al. 2023; Savoia et al. 2023). Repeat sequences were detected, and it was found that *EC* had only three types of repeat sequences: Palindromic (P), Forward (F), and Reverse (R). No Complex (C) sequences were detected. The number of repetitions for each repeat sequence type was as follows: Forward (F) had the highest number of repetitions with 25, followed by Palindromic (P) with 24, and Reverse (R) with the lowest at 7(Fig. 2).

Based on codonW's processing of *EC* files, we obtained the codon usage times and RSCU values for 20 amino acids in *EC*. In order to better describe the ratio of actual and expected codon usage(Fig. 3), we selected the RSCU value to study and describe the usage pattern of 20 amino acids. If the RSCU value is greater than 1, the frequency of a specific codon is higher than that of other synonymous codons, and the frequency of codons is higher than expected. Through analysis of the use of the RSCU value of the *EC (*Li et al. 2023b; Tang et al. 2022; Zhang et al. 2022b), it was found that in the cp genome, there are three amino acids with the highest RSCU value of 6, namely Leu, Ser, and Arg, and each of the three amino acids contains six codons. In Leu, there are three codons with RSCU > 1, namely UUA, UUG, and CUU; In Ser, there are also three codons with RSCU > 1, namely UCU, UCC, and UCA; In

Arg, there are four codons with RSCU > 1, namely AGA, AGG, GGA, and GGG. We found that the *EC* contains 59,602 common codons encoding protein genes, with the top three appearing times being Phe-UUU (2,300), Lys-AAA (2,179), and Asn-AAU (2,109).

3.3 Nucleotide polymorphism

We studied nucleotide polymorphisms between *EC* and different genera. To do this, we aligned the sequences of five species using the MAFFT website and exported them in Fas format. Next, we imported the output file into DnaSP v5.0 for nucleotide polymorphism analysis(Fig. 4). Our analysis revealed that the genetic sequence of *EC* is relatively conserved, with only two loci showing a Pi > 0.1(Wonok et al. 2023). These loci correspond to the regions of *psbN-petB* and *ndhH-ndhD*, respectively. Additionally, we used the online platform mVISTA to compare the sequences of the five species(Li et al. 2021; Liu et al. 2022; Wu et al. 2022). By using *EC* as a reference, we identified a highly specific sequence between 115K-134K that did not match any of the neighboring species(Supplement Fig. 1).

3.4 Inverted repeats contraction and expansion

Studying the variation in the IR region can be useful in elucidating the phylogenetic relationships between species, as the contraction and expansion of the inverted repeat (IR) can alter the length of the CPDNA(Yang et al. 2023b).

EC is a species that belongs to the genus *Eomecon* and is found only in China. In order to investigate its genetic relationships with other species, we conducted a study of the cp genome(Fig. 5). We selected 15 other species based on phylogenetic clustering and analyzed them together with *EC.* Using IR analysis, we compared the genomes of *EC* and its neighboring species in four regions of the genome: the large single copy (LSC) and inverted repeat (IR) regions, as well as the junctions between them (JLB, JSB, JSA, and JLA). Our results showed that *EC* shares some similarities with other species in the LSC and IRb regions, including the presence of genes such as *rpl*22, *rpl*2, and *rps*19 at the JLB locus. However, we also found differences between *EC* and its neighbors, particularly in the IRa region, where genes such as *trn*H, *psb*A, and *rpl*2 were present. Interestingly, we observed that the length of the SSC region was significantly shorter in all four regions of the genome, at 148 bp. Further analysis revealed that this was due to an expansion of the IR region, which caused multiple copies of genes to be inserted into the SSC region, leading to its contraction. This finding sheds new light on the evolution of the chloroplast genome in this region and provides a potential explanation for the unique features of *EC*'s genome.

3.5 Phylogenetic analysis

PhyloSuite was used to extract proteins from 31 species and establish a maximum likelihood tree (ML)(Yi et al. 2023). *EC* is an endemic species of a single genus and a single species unique to China, so on the ML tree, *EC* is grouped into a single category(Li et al. 2023a; Wang et al. 2023a). Although *EC* is a separate category within the poppy family, it is closely related to four species (*Macleaya microcarpa, Coreanomecon haloconoids, Hylomecon japonica, Chelidonium majus*). All nodes have high boot support(Fig. 6).

3.6 Kimura's two-parameter (K2-P) analysis

We chose five species for the k2-p model based on phylogenetic analysis, which revealed that the genus *Eomecon* is closely related to four other genera(Fig. 7). Highly variable regions were utilized to distinguish species with close phylogenetic relationships(Toshino et al. 2022). Analysis of the genetic distance of the five selected species identified potential molecular markers in certain regions. From screening 82 intergenic regions in the five species using the K2-p model, we identified five gene regions (*ycf*4-*cem*A, *ycf*3-*trn*S-GGA, *trn*C-GCA-*pet*N, *rpl*32-*trn*L-UAG, and *psb*I-*trn*S-UGA) with the highest genetic distance. These regions can serve as potential markers for future molecular marker development.

3.7 Selective pressures analysis

Changes in the DNA sequence are influenced by natural selection and can be assessed by the ratio of non-synonymous (amino acid substitution, Ka) to synonymous (Ks) substitutions (Ka/Ks)(Zhang et al. 2022a)(Fig. 8). A Ka/Ks value of less than 1 indicates a high synonymous substitution rate, suggesting that the amino acid at this site remains unchanged and the probability

of mutation in this fragment is low. Conversely, a Ka/Ks value of greater than 1 suggests a high rate of non-synonymous substitution, indicating changes in amino acids and a higher possibility of mutation in this fragment. In this study, most of the Ka/Ks values obtained were less than 1, implying that synonymous nucleotide substitution is more common in the protein-coding region during natural selection. Out of the 80 protein-coding genes examined, only *clp*P, *psb*K, and *ycf*2 had a Ka/Ks value greater than 1, indicating that these three genes have the highest non-synonymous substitution and are under positive selection pressure(Lin et al. 2023). Mutations in these sites may lead to the future accelerated development of this species.

4. Discussion

Papaveraceae plants have high ornamental value, with a global distribution of 38 genera and over 700 species, mainly distributed in the northern temperate zone and in West Asia, Central Asia, East Asia, and other regions.

In this research, we focused on *EC*, a unique plant species in China. We sequenced, assembled, and annotated *EC* CP data and downloaded complete plastid data from NCBI for 30 species. We aligned them through the MAFFT website and conducted in-depth research on the sequencing results using various software and models such as cpjavas2, mega11, DNAspv5.0 and so on. The genetic differences, high variation regions, variation sites, and phylogenetic relationships of *EC* were analyzed. The results showed that the plastid size of *EC* was 178,808 bp, with a typical tetrad structure of angiosperms.

In SSR testing, we found that among the 54 SSRs of *EC*, most of them are single nucleotides (A/T). Most of the SSRs are located in non-coding genes, and few are located in protein-coding genes. The CP structure of *EC* is the same as ordinary green plants, divided into four partitions (LSC, SSC, IRa, and IRb). The length of the LSC region is the longest of the four partitions, and due to the expansion of the IR region, the original gene on the SSC has multiple copies, resulting in extremely short SSCs and long IR The CP genome size of species from the same part or genus is extremely uniform, but *EC* is unique to China and belongs to a single genus and single species, giving the CP genome of *EC* its own unique characteristics.

IGS and nucleotide polymorphism analysis have provided us with a study of regions and sites with high variability potential in *EC.* Using the K-2p model, we selected five regions with high variability, which can provide a reference for future molecular markers and species identification. The analysis of nucleotide polymorphisms has helped us gain a deeper understanding of the possible selection pressures in the *EC* protein genome, providing a prediction of the future direction of EC species and aiding in our understanding of this unique plant species in China.

5. Conclusion

In order to deeply understand the CP structure of *EC*, an endemic plant in China, we sequenced, assembled, and annotated the complete chloroplasts of *EC* and reported the genetic characteristics of the *Eomecon* species in this study. We have discovered the connections and differences between this particular plant and its neighboring genera through various analyses. The expansion of *EC*'s IR region leads to multiple copies of genes on SSC. Therefore, the longer IR and the shorter SSC regions of *EC* are unique characteristics of *EC.* We identified areas and sites that may have high variability in the CPDNA of *EC* through high variation analysis, mVISTA, and IGS analysis, providing reference markers for future molecular recognition and phylogenetic research. This study is the first to report the complete CP genome of a single genus and single species, *EC*, endemic to China, providing more data for the study of species in the poppy family. Through the structural analysis of the chloroplast of *EC*, we gained an in-depth understanding of the characteristics of the *Eomecon* species, which is of great significance for future research on the evolution and variation of the *Eomecon* species.

Declarations

Author Contributions:

Data curation, ZZ, YM, GZ, XZ, HZ, and JX.

Formal analysis, ZZ, and YM.

Methodology, GZ, and XZ.

Resources, LH.

Software, YM And HZ.

Visualization, JX.

Writing original draft, ZZ.

Writing—review and editing, L.H.and R.Z.

All authors have read and agreed to the published version of the manuscript.

Funding:

This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS, 2022-I2M-1-017), National Natural Science Foundation of China (U1812403-1, 82073960,82211540726 and 82274045), the Open Fund of State Key Laboratory of Southwestern Chinese Medicine Resources (SKLTCM2022015), National Science & Technology Fundamental Resources Investigation Program of China (2018FY100701), which are gratefully acknowledged.

Declaration of Competing Interest

No potential conflict of interest was reported by the authors.

Acknowledgments:

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

References

1. Ahmed MS, Datta SK, Saha T, Hossain Z (2021) Molecular characterization of marine and coastal fishes of Bangladesh through DNA barcodes. Ecol Evol 11:3696-3709
2. Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34:3030-3031
3. Arseneau JR, Steeves R, Laflamme M (2017) Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. Mol Ecol Resour 17:686-693
4. Bhaskar R, Das MK, Sharon EA, Kumar RR, R GC (2021) Genetic identification of marine eels (Anguilliformes: Congroidei) through DNA barcoding from Kasimedu fishing harbour. Mitochondrial DNA B Resour 6:3354-3361
5. Bi D, Han S, Zhou J, Zhao M, Zhang S, Kan X (2023) Codon Usage Analyses Reveal the Evolutionary Patterns among Plastid Genes of Saxifragales at a Larger-Sampling Scale. Genes (Basel) 14
6. Boltenkov EV, Artyukova EV (2023) New Approach to the Systematics of the Section Psammiris (Iris, Iridaceae): What Does Chloroplast DNA Sequence Tell Us? Plants (Basel) 12
7. Çelik A, Coşkan S, Morca AF, Santosa AI, Koolivand D (2022) Insight into Population Structure and Evolutionary Analysis of the Emerging Tomato Brown Rugose Fruit Virus. Plants (Basel) 11
8. Chen J, Xie D, He X, Yang Y, Li X (2022a) Comparative Analysis of the Complete Chloroplast Genomes in Allium Section Bromatorrhiza Species (Amaryllidaceae): Phylogenetic Relationship and Adaptive Evolution. Genes (Basel) 13
9. Chen W, Achakkagari SR, Strömvik M (2022b) Plastaumatic: Automating plastome assembly and annotation. Front Plant Sci 13:1011948
10. Desingu PA, Rubeni TP, Sundaresan NR (2022) Evolution of monkeypox virus from 2017 to 2022: In the light of point mutations. Front Microbiol 13:1037598

11. Giorgashvili E, Reichel K, Caswara C, Kerimov V, Borsch T, Gruenstaeudl M (2022) Software Choice and Sequencing Coverage Can Impact Plastid Genome Assembly-A Case Study in the Narrow Endemic Calligonum bakuense. Front Plant Sci 13:779830
12. Godeiro NN, Zhang F (2021) First record of Seira dowlingi (Wray, 1953) (Collembola, Entomobryidae, Seirinae) from China and mitogenome comparison with the New World specimens. Zootaxa 5020:191-196
13. Habib KA, Neogi AK, Rahman M, Oh J, Lee YH, Kim CG (2021) DNA barcoding of brackish and marine water fishes and shellfishes of Sundarbans, the world's largest mangrove ecosystem. PLoS One 16:e0255110
14. Jin X, Guo X, Chen J, Li J, Zhang S, Zheng S, Wang Y, Peng Y, Zhang K, Liu Y, Liu B (2023) The complete mitochondrial genome of Hemigrapsus sinensis (Brachyura, Grapsoidea, Varunidae) and its phylogenetic position within Grapsoidea. Genes Genomics 45:377-391
15. Kamra K, Jung J, Kim JH (2023) A phylogenomic study of Iridaceae Juss. based on complete plastid genome sequences. Front Plant Sci 14:1066708
16. Kuang L, Shen Q, Chen L, Ye L, Yan T, Chen ZH, Waugh R, Li Q, Huang L, Cai S, Fu L, Xing P, Wang K, Shao J, Wu F, Jiang L, Wu D, Zhang G (2022) The genome and gene editing system of sea barleygrass provide a novel platform for cereal domestication and stress tolerance studies. Plant Commun 3:100333
17. Li C, Cai C, Tao Y, Sun Z, Jiang M, Chen L, Li J (2021) Variation and Evolution of the Whole Chloroplast Genomes of Fragaria spp. (Rosaceae). Front Plant Sci 12:754209
18. Li C, Zhou L, Nie J, Wu S, Li W, Liu Y, Liu Y (2023a) Codon usage bias and genetic diversity in chloroplast genomes of Elaeagnus species (Myrtiflorae: Elaeagnaceae). Physiol Mol Biol Plants 29:239-251
19. Li G, Shao J, Wang Y, Liu T, Tong Y, Jansky S, Xie C, Song B, Cai X (2022) Ry(chc) Confers Extreme Resistance to Potato virus Y in Potato. Cells 11
20. Li Q, Luo Y, Sha A, Xiao W, Xiong Z, Chen X, He J, Peng L, Zou L (2023b) Analysis of synonymous codon usage patterns in mitochondrial genomes of nine Amanita species. Front Microbiol 14:1134228
21. Lin Y, Chen W, Yang Q, Zhang Y, Ma X, Li M (2023) Genome-Wide Characterization and Gene Expression Analyses of Malate Dehydrogenase (MDH) Genes in Low-Phosphorus Stress Tolerance of Chinese Fir (Cunninghamia lanceolata). Int J Mol Sci 24
22. Liu Y, Li Y, Feng S, Yan S, Wang J, Huang Y, Yang M (2022) Complete chloroplast genome structure of four Ulmus species and Hemiptelea davidii and comparative analysis within Ulmaceae species. Sci Rep 12:15953
23. Lu M, Li K, He H, Cheng Y, Yang P (2020) Systematic characterization of alkaloids in Eomecon chionantha Hance using ultrahigh-performance liquid chromatography-tandem quadrupole Exactive Orbitrap mass spectrometry with a four-step screening strategy. Rapid Commun Mass Spectrom 34:e8880
24. Luo X, Zhang L, Chen S (2023) Microsatellite Genome-Wide Database Development for the Commercial Blackhead Seabream (Acanthopagrus schlegelii). Genes (Basel) 14
25. Miao Y, Chen H, Xu W, Yang Q, Liu C, Huang L (2022) Structural mutations of small single copy (SSC) region in the plastid genomes of five Cistanche species and inter-species identification. BMC Plant Biol 22:412
26. Nemati Z, Dadkhodaie A, Mostowfizadeh-Ghalamfarsa R, Mehrabi R, Cacciola SO (2023) Genetic Variation of Puccinia triticina Populations in Iran from 2010 to 2017 as Revealed by SSR and ISSR Markers. J Fungi (Basel) 9
27. Pavan-Kumar A, Singh S, Mishra A, Suman S, Gireesh-Babu P, Chaudhari A, Shen KN, Borsa P (2022) Characterization of mitochondrial genome of Indian Ocean blue-spotted maskray, Neotrygon indica and its phylogenetic relationship within Dasyatidae Family. Int J Biol Macromol 223:458-467
28. Pei D, Song S, Kang J, Zhang C, Wang J, Dong T, Ge M, Pervaiz T, Zhang P, Fang J (2023) Characterization of Simple Sequence Repeat (SSR) Markers Mined in Whole Grape Genomes. Genes (Basel) 14
29. Peng F, Liu M, Huang Q, Liu N, Yang H, Sun H, Hu Q, Feng F, Jiang C (2011) Molluscicidal effect of Eomecon chionantha alkaloids against Oncomelania hupensis snails. Southeast Asian J Trop Med Public Health 42:289-296
30. Raven JA, Allen JF (2003) Genomics and chloroplast evolution: what did cyanobacteria do for plants? Genome Biology 4:209
31. Savoia MA, Del Faro L, Turco A, Fanelli V, Venerito P, Montemurro C, Sabetta W (2023) Biodiversity Evaluation and Preservation of Italian Stone Fruit Germplasm (Peach and Apricot) in Southern Italy. Plants (Basel) 12

32. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res 47:W65-w73
33. Shi W, Song W, Chen Z, Cai H, Gong Q, Liu J, Shi C, Wang S (2023) Comparative chloroplast genome analyses of diverse Phoebe (Lauraceae) species endemic to China provide insight into their phylogeographical origin. PeerJ 11:e14573
34. Tang D, Huang S, Quan C, Huang Y, Miao J, Wei F (2022) Mitochondrial genome characteristics and phylogenetic analysis of the medicinal and edible plant Mesona chinensis Benth. Front Genet 13:1056389
35. Toshino S, Yamamoto G, Saito S (2022) Octorhopalona saltatrix, a New Genus and Species (Hydrozoa, Anthoathecata) from Japanese Waters. Animals (Basel) 12
36. Wang C, Tang Y, Ma X, Li H, Li J, Tang H, Liu Z (2023a) First Report of Macrophomina phaseolina Causing Leaf Blight of Cassava in China. Plant Dis
37. Wang H, Liu S, Lv Y, Wei W (2023b) Codon usage bias of Venezuelan equine encephalitis virus and its host adaption. Virus Res 328:199081
38. Wang S, Tian P, Li S, Liu H, Guo X, Huang F (2022) Genetic diversity of transmission-blocking vaccine candidate antigens Pvs25 and Pvs28 in Plasmodium vivax isolates from China. BMC Infect Dis 22:944
39. Wonok W, Sudmoon R, Tanee T, Lee SY, Chaveerach A (2023) Complete Chloroplast Genome of Four Thai Native Dioscorea Species: Structural, Comparative and Phylogenetic Analyses. Genes (Basel) 14
40. Wu L, Nie L, Guo S, Wang Q, Wu Z, Lin Y, Wang Y, Li B, Gao T, Yao H (2022) Identification of Medicinal Bidens Plants for Quality Control Based on Organelle Genomes. Front Pharmacol 13:842131
41. Xiao T, He L, Yue L, Zhang Y, Lee SY (2022) Comparative phylogenetic analysis of complete plastid genomes of Renanthera (Orchidaceae). Front Genet 13:998575
42. Xie ZN, Lao J, Liu H, Zhang WX, He W, Zhong C, Xie J, Zhang SH, Jin J (2023) Characterization of the chloroplast genome of medicinal herb Polygonatum cyrtonema and identification of molecular markers by comparative analysis. Genome
43. Yang C, Zhang N, Wu S, Jiang C, Xie L, Yang F, Yu Z (2023a) A Comparative Analysis of the Chloroplast Genomes of Three Lonicera Medicinal Plants. Genes (Basel) 14
44. Yang J, Chu Q, Meng G, Kong W (2022) The complete chloroplast genome sequences of three Broussonetia species and comparative analysis within the Moraceae. PeerJ 10:e14293
45. Yang L, Deng S, Zhu Y, Da Q (2023b) Comparative chloroplast genomics of 34 species in subtribe Swertiinae (Gentianaceae) with implications for its phylogeny. BMC Plant Biol 23:164
46. Yang P, Cheng YT, Huang X, Huang B, Yi L, He HL, Xie Y (2023c) Identification of a new benzophenanthridine alkaloid from Eomecon chionantha induced necroptosis in breast cancer cells. Nat Prod Res 37:912-918
47. Yang P, Zhou X, Xie Y (2023d) Cytotoxic Effects of the Benzophenanthridine Alkaloids Isolated from Eomecon chionantha on MCF-7 Cells and Its Potential Mechanism. Chem Biodivers 20:e202200871
48. Yi RH, Huang Y, Liu Q, Peng YF, Li D (2023) Dalbergia odorifera, a new host of Colletotrichum tropicale causing anthracnose in China. Plant Dis
49. Zhang L, Li S, Fang X, An H, Zhang X (2022a) Genome-wide analysis of LysM gene family members and their expression in response to Colletotrichum fructicola infection in Octoploid strawberry(Fragaria × ananassa). Front Plant Sci 13:1105591
50. Zhang Y, Zhou J, Tng DYP, Wang S, Wang Y, Peng Y, Liu H, Wang Z (2023) Phylogeny and Systematics of Sassafras (Lauraceae), an Interesting Genus with Disjunct Distributions in Eastern North America and East Asia. Plants (Basel) 12
51. Zhang Z, Zhang DS, Zou L, Yao CY (2022b) Comparison of chloroplast genomes and phylogenomics in the Ficus sarmentosa complex (Moraceae). PLoS One 17:e0279849
52. Zheng S, Poczai P, Hyvönen J, Tang J, Amiryousefi A (2020) Chloroplot: An Online Program for the Versatile Plotting of Organelle Genomes. Front Genet 11:576124

Figures

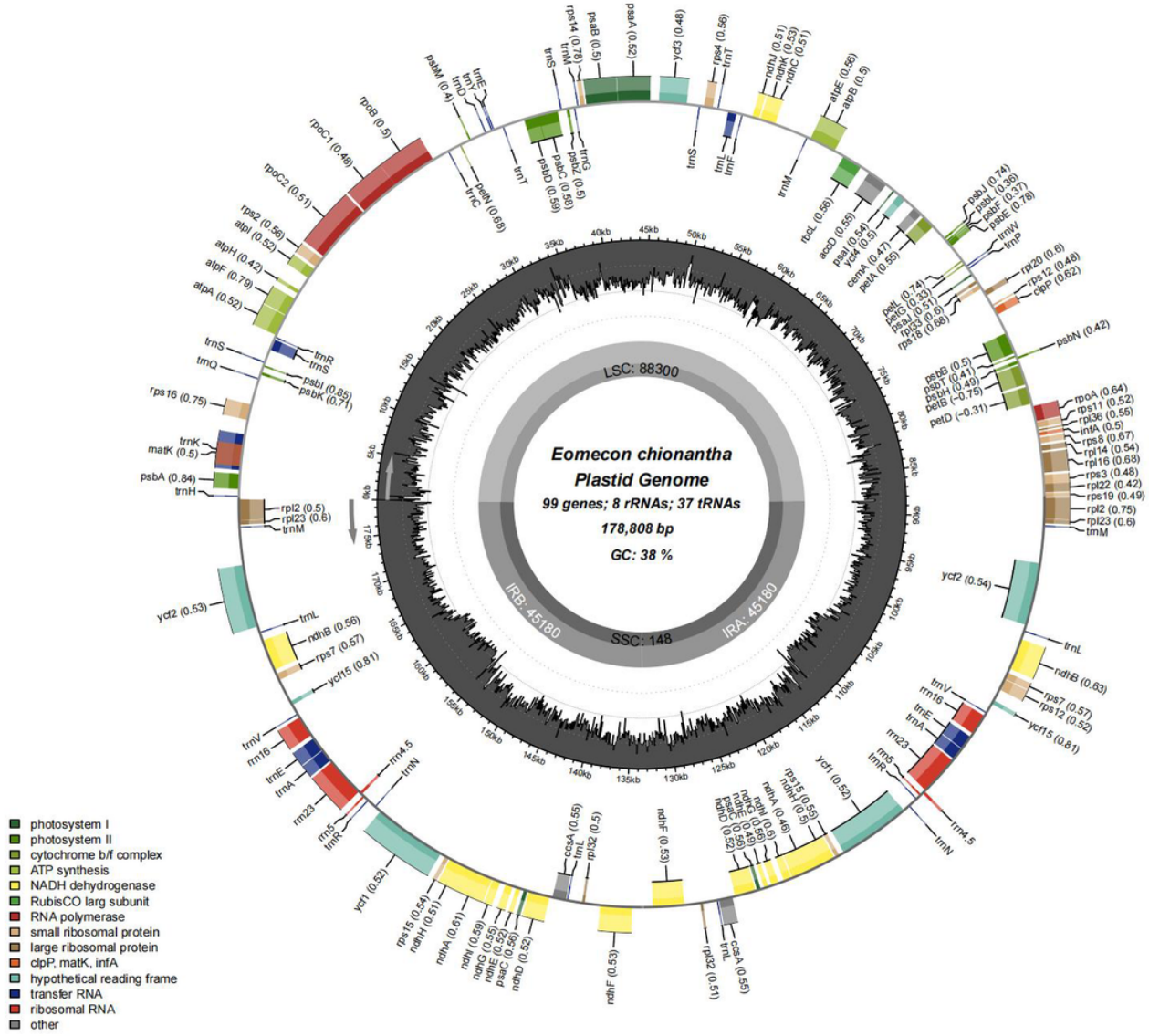
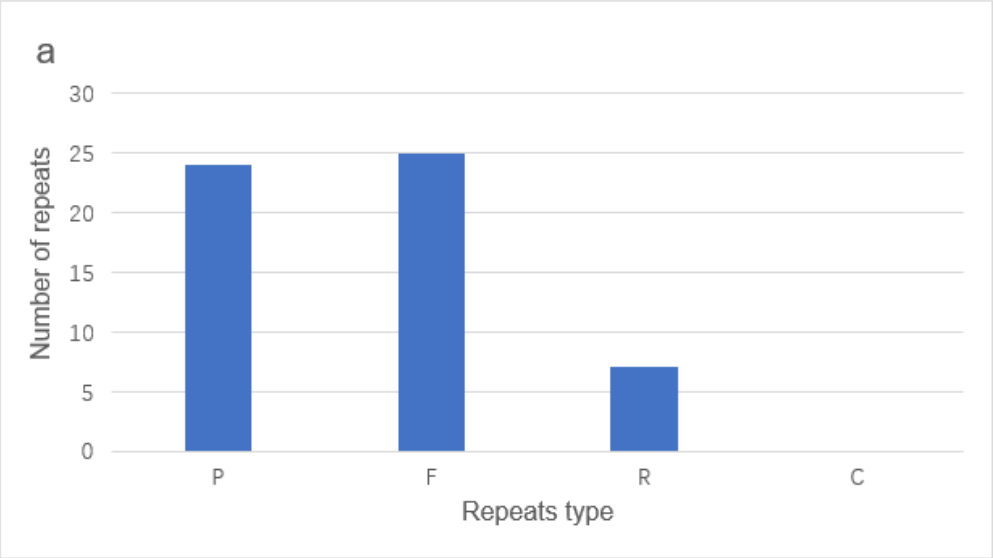


Figure 1

Chloroplast Genome Map of *EC* The genes located outside the circle undergo transcription in a counterclockwise direction, whereas those inside the circle are transcribed in a clockwise direction. Additionally, the colored bars indicate various functional groups.



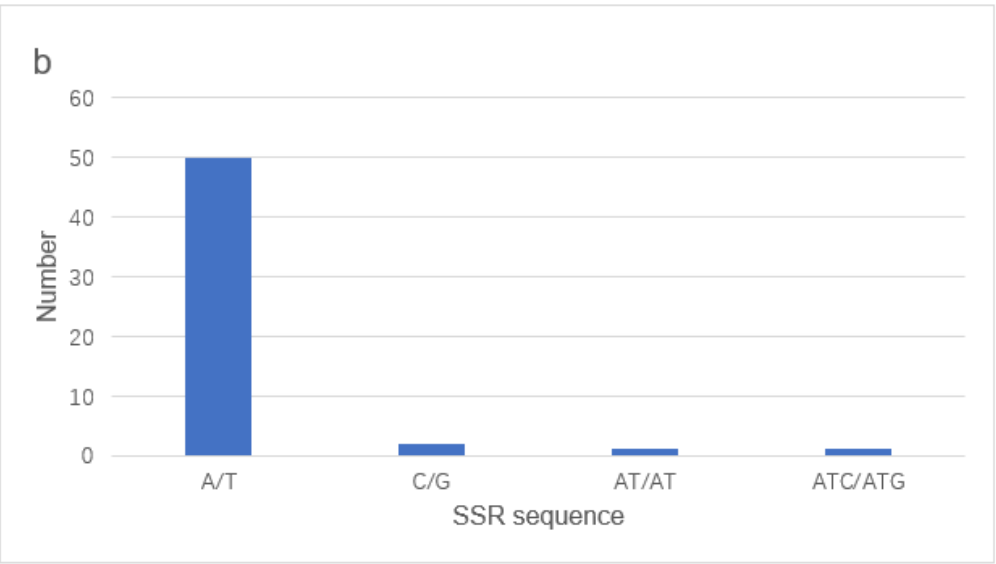


Figure 2

Detection of repetitive sequences and SSRs in the CP genome of *EC* a: Types and numbers of interspersed repeat sequences in *EC* b: Types and quantities of SSRs in *EC.*

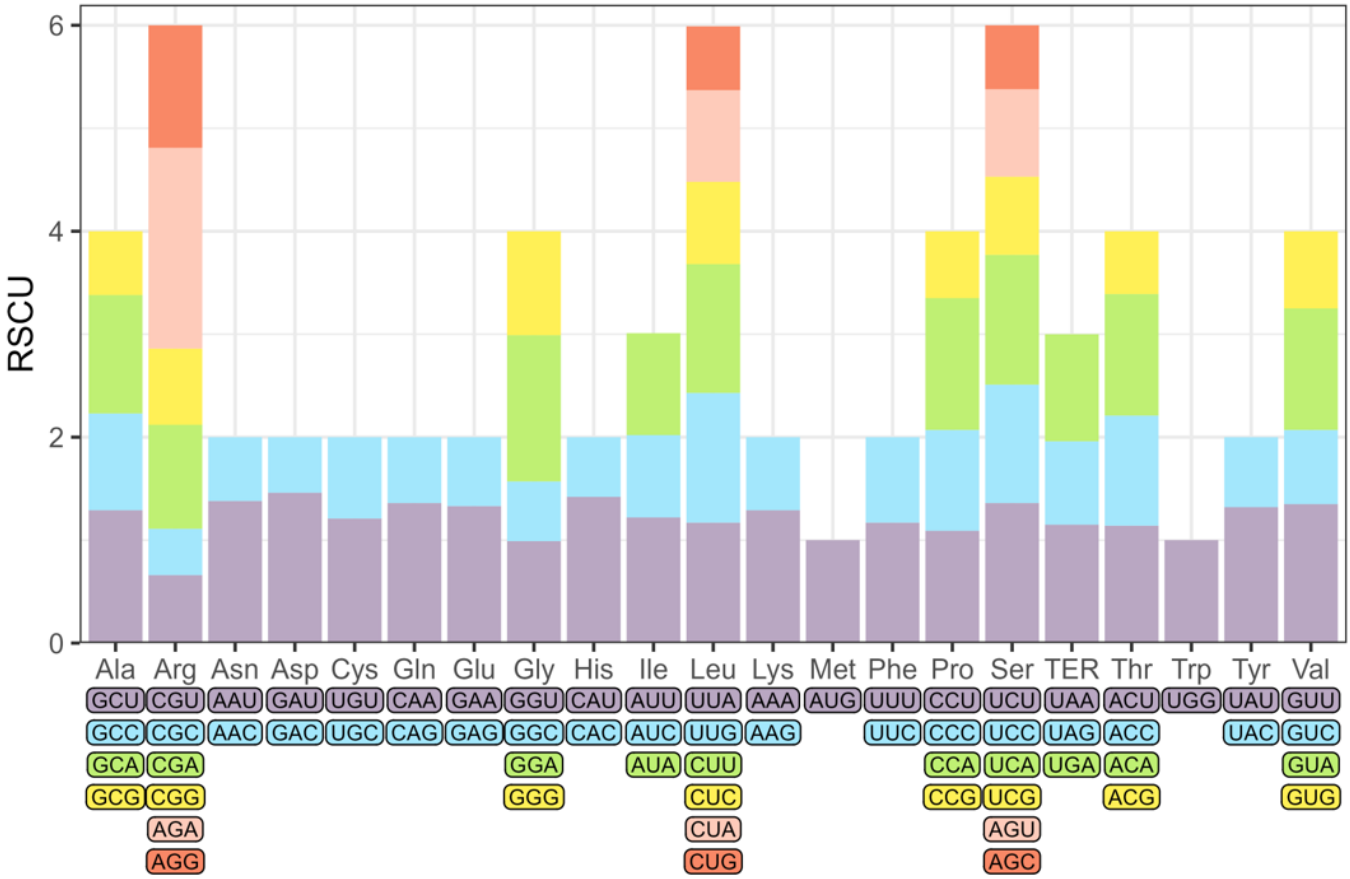


Figure 3

Codon content and stop codon of 20 amino acids in protein-coding genes.

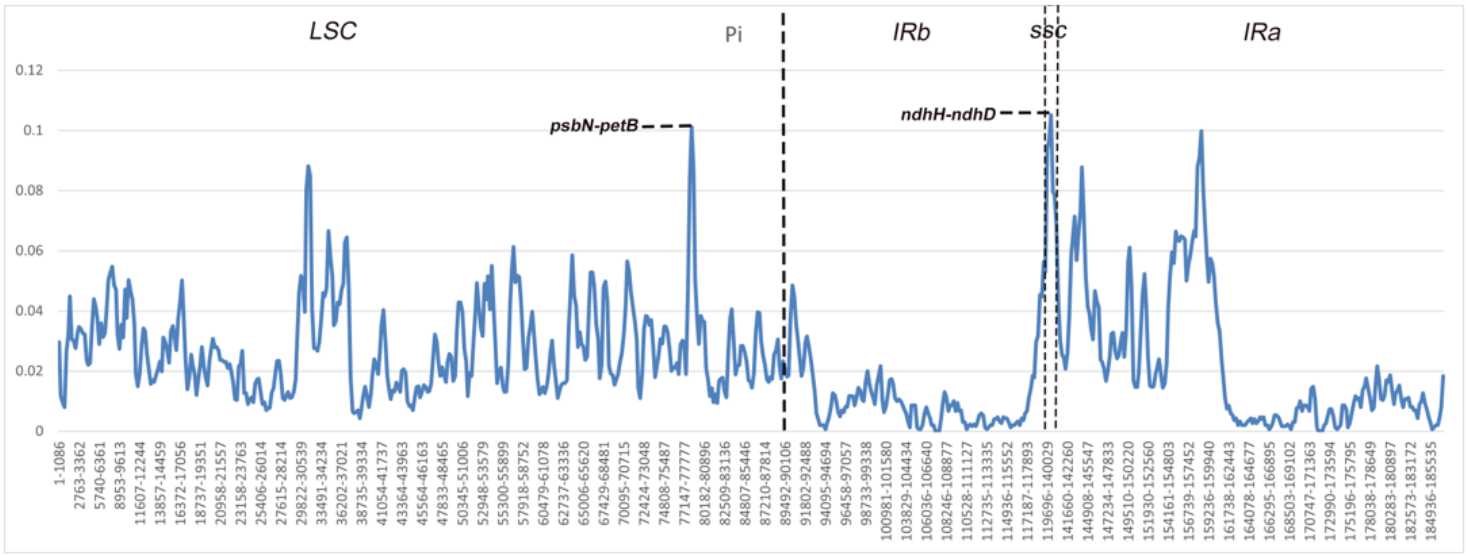


Figure 4

Nucleotide polymorphism distribution of *EC.*

Figure 5

The borders of the IR, SSC, and LSC regions were compared among 16 chloroplast genomes. The JLB indicates the junction of LSC and IRb, JSB denotes the junction of SSC and IRb, JSA represents the junction of SSC and IRa, and JLA signifies the junction of LSC and IRa.

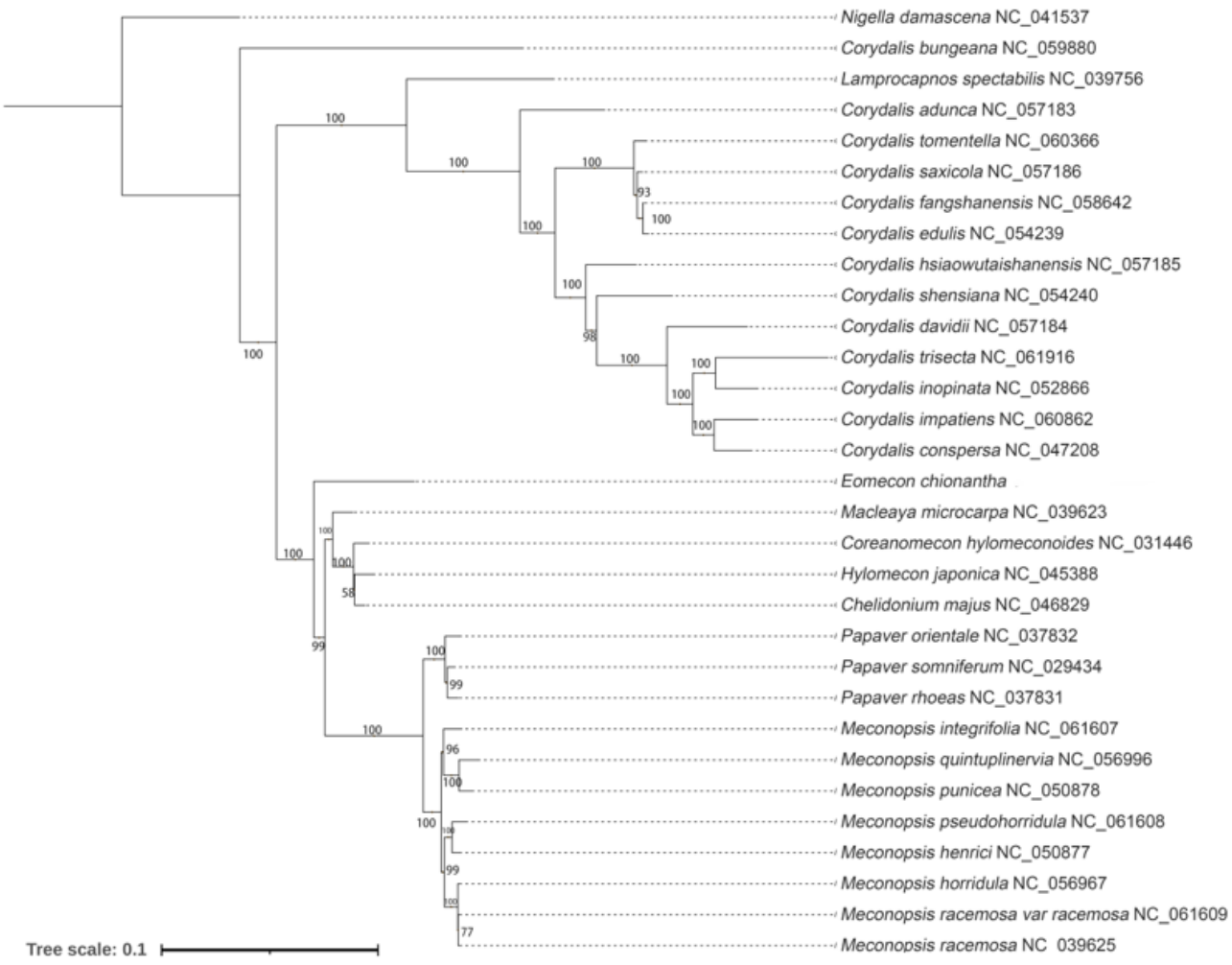


Figure 6

A phylogenetic analysis was performed, and a maximum likelihood tree (ML) was constructed using 31 species of Papaveraceae. *Nigella damascena* was selected as an outgroup, and all nodes in the tree exhibited high bootstrap support.

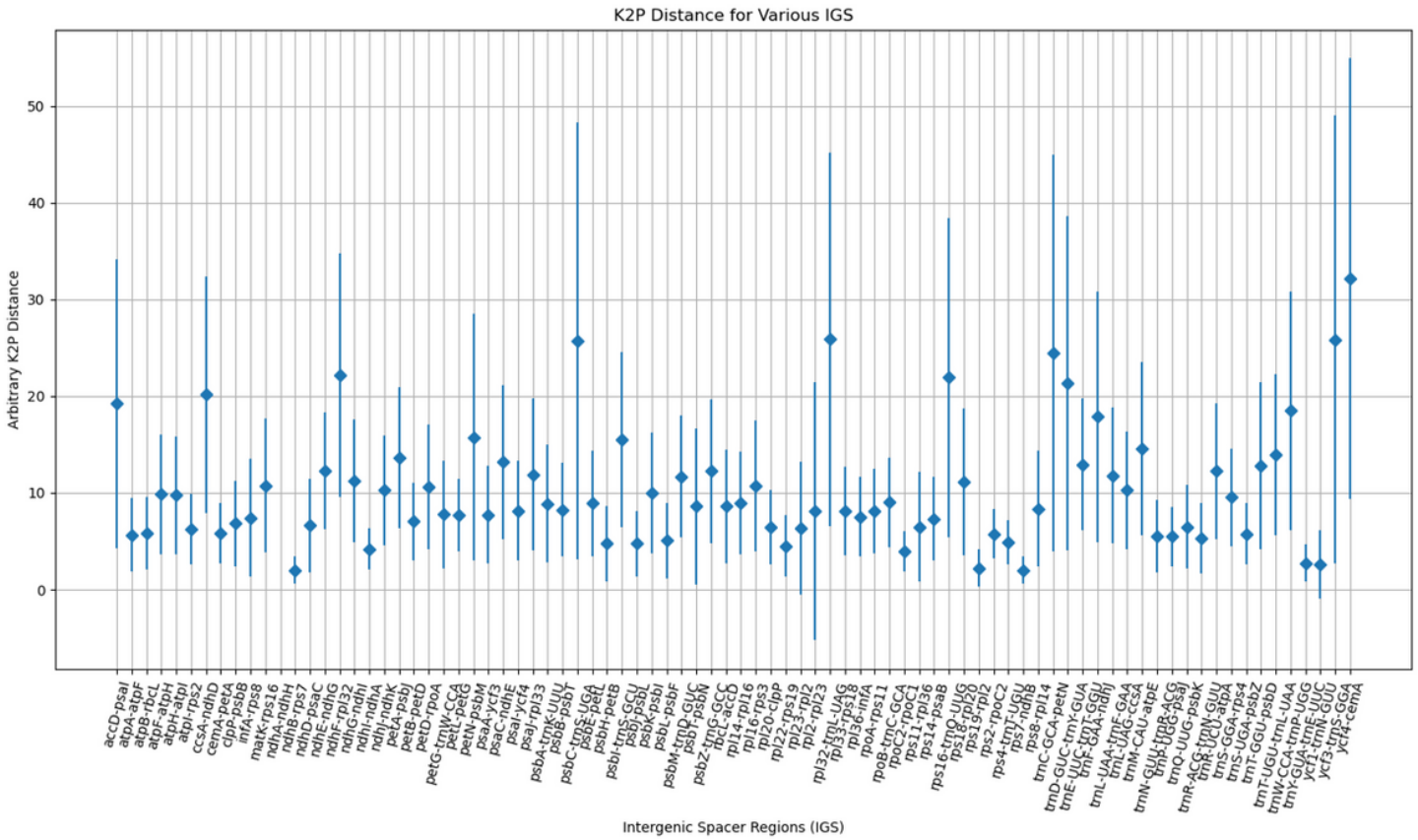


Figure 7

Genetic distance analysis of *EC* and four other related species by K-2p model. The X-axis indicates the IGS regions, and the Y-axis shows the value of K-2p distances.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx