



OPEN ACCESS

**Edited by:**

Abdullah.,  
Quaid-i-Azam University, Pakistan

**Reviewed by:**

Peter Poczai,  
University of Helsinki, Finland  
Chuanyuan Deng,  
Fujian Agriculture and Forestry  
University, China  
Xiu-Qun Liu,  
Huazhong Agricultural University,  
China  
Jibran Tahir,  
The New Zealand Institute for Plant  
and Food Research Ltd.,  
New Zealand

**\*Correspondence:**

Xiu-Zhen Cai  
changshacz@126.com  
Guang-Wan Hu  
guangwanhu@wbqcas.cn

† These authors have contributed  
equally to this work

**Specialty section:**

This article was submitted to  
Plant Systematics and Evolution,  
a section of the journal  
Frontiers in Plant Science

**Received:** 03 December 2021

**Accepted:** 31 January 2022

**Published:** 24 February 2022

**Citation:**

Ren J, Tian J, Jiang H, Zhu X-X,  
Mutie FM, Wang VO, Ding S-X,  
Yang J-X, Dong X, Chen L-L, Cai X-Z  
and Hu G-W (2022) Comparative  
and Phylogenetic Analysis Based on  
the Chloroplast Genome  
of *Coleanthus subtilis* (Tratt.) Seidel,  
a Protected Rare Species  
of Monotypic Genus.  
*Front. Plant Sci.* 13:828467.  
doi: 10.3389/fpls.2022.828467

# Comparative and Phylogenetic Analysis Based on the Chloroplast Genome of *Coleanthus subtilis* (Tratt.) Seidel, a Protected Rare Species of Monotypic Genus

Jing Ren<sup>1,2,3†</sup>, Jing Tian<sup>2,3,4†</sup>, Hui Jiang<sup>2,3,4</sup>, Xin-Xin Zhu<sup>5</sup>, Fredrick Munyao Mutie<sup>2,3,4</sup>, Vincent Okelo Wang<sup>2,3,4</sup>, Shi-Xiong Ding<sup>2,3,4</sup>, Jia-Xin Yang<sup>2,3,4</sup>, Xiang Dong<sup>2,3,4</sup>, Ling-Ling Chen<sup>1</sup>, Xiu-Zhen Cai<sup>1\*</sup> and Guang-Wan Hu<sup>2,3,4\*</sup>

<sup>1</sup> College of Life Sciences, Hunan Normal University, Changsha, China, <sup>2</sup> CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China, <sup>3</sup> Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan, China, <sup>4</sup> University of Chinese Academy of Sciences, Beijing, China, <sup>5</sup> College of Life Sciences, Xinyang Normal University, Xinyang, China

*Coleanthus subtilis* (Tratt.) Seidel (Poaceae) is an ephemeral grass from the monotypic genus *Coleanthus* Seidl, which grows on wet muddy areas such as fishponds or reservoirs. As a rare species with strict habitat requirements, it is protected at international and national levels. In this study, we sequenced its whole chloroplast genome for the first time using the next-generation sequencing (NGS) technology on the Illumina platform, and performed a comparative and phylogenetic analysis with the related species in Poaceae. The complete chloroplast genome of *C. subtilis* is 135,915 bp in length, with a quadripartite structure having two 21,529 bp inverted repeat regions (IRs) dividing the entire circular genome into a large single copy region (LSC) of 80,100 bp and a small single copy region (SSC) of 12,757 bp. The overall GC content is 38.3%, while the GC contents in LSC, SSC, and IR regions are 36.3%, 32.4%, and 43.9%, respectively. A total of 129 genes were annotated in the chloroplast genome, including 83 protein-coding genes, 38 tRNA genes, and 8 rRNA genes. The *accD* gene and the introns of both *clpP* and *rpoC1* genes were missing. In addition, the *ycf1*, *ycf2*, *ycf15*, and *ycf68* were pseudogenes. Although the chloroplast genome structure of *C. subtilis* was found to be conserved and stable in general, 26 SSRs and 13 highly variable loci were detected, these regions have the potential to be developed as important molecular markers for the subfamily Pooideae. Phylogenetic analysis with species in Poaceae indicated that *Coleanthus* and *Phippisia* were sister groups, and provided new insights into the relationship between *Coleanthus*, *Zingeria*, and *Colpodium*. This study presents the initial chloroplast genome report of *C. subtilis*, which provides an essential data reference for further research on its origin.

**Keywords:** *Coleanthus subtilis*, chloroplast genome, comparative analysis, phylogeny, monotypic genus

## INTRODUCTION

*Coleanthus subtilis* (Tratt.) Seidel is a rare grass in the monotypic genus *Coleanthus* Seidl, which can be recognized by its rosette-like arrangement of stems, the wide leaf sheaths and curved leaves (Richert et al., 2016). It has a wide but disjunctive distribution area and has been recorded in west-central Europe, southern Norway, Russia, China, United States, and Canada (Richert et al., 2014). It occurs mainly on wet and muddy habitats, growing along streams or rivers (Taran, 1994). Its secondary habitats are artificial ponds and reservoirs, where changes in water level expose bare and moist surfaces that give the seeds the opportunity to germinate (Hejný, 1969; Woike, 1969; Richert et al., 2014). It is an ephemeral grass whose life cycle lasts only a few weeks and requires high levels of moisture and nutrients from germination to reproduction. Moreover, in order to germinate, a diurnal temperature difference of at least 20°C is necessary (Hejný, 1969; Richert et al., 2014). Destruction of favorable habitats in regions such as Europe have threatened the survival of *C. subtilis*. With the development of fisheries and tourism, ponds and reservoirs are becoming increasingly populated with anglers, which affects the secondary habitat of *C. subtilis*. Furthermore, the frequency and timing of ponds and reservoirs drainage also influence the reproductive cycle of *C. subtilis*, as prolonged periods without drainage may limit seed germination and result in failure to renew the seed bank (Richert et al., 2014). The strict conditions for reproduction combined with habitat destruction have led to a sharp decline in the populations of *C. subtilis*, hence it is protected at both national and international levels. For example, it is listed in Annexes II and IV of the Habitats Directive by the European Union Organization and is also documented in Appendix I of the Berne Convention (John et al., 2010). Besides, *C. subtilis* is considered a species in need of conservation in other countries, such as the Czechia (Grulich, 2012) and North America (Catling, 2009). In China, it also has been listed as a second-class national key protected wild plant<sup>1</sup>.

*Coleanthus subtilis* has long been of interest to researchers due to its special distribution pattern, strict habitat requirements and unique inflorescence structure (Kurchenko, 2006). *C. subtilis* has a remarkable ability to reappear in its previous habitats after long time intervals, which may be related to the hypothesis that its seeds can remain viable in the soil for decades (Hejný, 1969; Richert et al., 2014). For example, it was rediscovered in 2001 at Volkhov Shoal, where it was mistakenly thought to have been extinct for 70 years (Yurova, 2001). In 2021, we found it in Harbin after an interval of nearly 100 years. In addition, *C. subtilis* was collected on the banks of the Yangtze River in Wuhan, where its distribution has never been recorded before. The factors responsible for this particular distribution pattern are unclear.

Based on morphological studies, *C. subtilis* was once considered a member of the tribe Agrostideae because of its distinctive inflorescence, which has flowers aggregated in bunches and with staminodes (Gnutikov et al., 2020). In addition, it has been placed near the genera *Alopecurus* and *Mibora*, although it does not share common features with these two

(Gnutikov et al., 2020). However, some researchers believe that there is a close relationship between the genus *Coleanthus* and the genus *Phippisia* because of the similarities in morphology and ecological preferences (Tzvelev, 1976; Gnutikov et al., 2020). Soreng et al. (2003) proposed a new subfamily called Puccinelliinae based on molecular phylogenetic analysis and more thorough morphological examination of Poaceae, which are characterized by thin membranous lemmas with hyaline apex and glabrous margins. Hoffmann et al. (2013) placed *C. subtilis* in the Puccinelliinae using DNA sequence data of the ribosomal internal transcribed spacer (ITS). Subsequently, the subtribe Puccinelliinae was renamed as Coleanthinae after the addition of *Coleanthus* (Soreng et al., 2015). The use of chloroplast genes or fragments (*matK*, *ndhF*, and *trnL-trnF*) to explore the phylogenetic position of *C. subtilis* showed that it is most closely related to the genus *Phippisia* and that both are sister groups to other genera in the subtribe Coleanthinae of the subfamily Pooideae (Gnutikov et al., 2020), but opinions differ on the composition of this subtribe (Soreng et al., 2003, 2015; Gnutikov et al., 2020; Tkach et al., 2020). The whole chloroplast genomes provide more complete genetic information than single gene fragments to enable better discovery of interspecific genetic resources and understanding of evolutionary history (Wariss et al., 2018). However, to date, no studies have explored the phylogenetic position of *C. subtilis* with the help of complete chloroplast genomes, which affects our comprehensive understanding of its phylogeny.

Compared with nuclear and mitochondrial genome, chloroplast genomes are characterized by moderate nucleotide substitution rates, structural simplicity and uniparental inheritance (Burke et al., 2012; Ruhfel et al., 2014; Yang et al., 2019), which makes them ideal resources in phylogenetic studies at different levels and a common tool for species identification (Chen et al., 2018; Yu et al., 2021). Its structure is relatively stable and contains a large amount of genetic information, which is considered a valuable data resource for solving complex evolutionary relationships (Parks et al., 2009; Moore et al., 2010; Oldenburg and Bendich, 2016). At the same time, it has a promising future in molecular marker studies, as some genes are often used in DNA barcoding for species identification, such as *rbcL* and *matK* (Hollingsworth, 2011). In addition, chloroplast genomes have been widely used in plant genetic diversity and conservation studies, since they can provide more complete genetic information compared to individual gene fragments, therefore facilitating better resolution of evolutionary relationships among species (Wariss et al., 2018). Next-generation sequencing (NGS) technology provides an efficient and cost-effective method for chloroplast genome assembly, which greatly enriches chloroplast genome information and provides sufficient data for plant phylogenetic studies (Cronn et al., 2008; Tangphatsornruang et al., 2010). Despite this, the chloroplast genome of *C. subtilis* has not been reported to date, which limits its development of genetic information discovery and phylogenetic studies.

Therefore, the purpose of this study is to (a) provide the first report on the chloroplast genome of the genus *Coleanthus* and conduct a comparative genomic analysis with other species in

<sup>1</sup><https://www.forestry.gov.cn/>

the subfamily Pooideae; (b) make the first attempt to reconstruct the phylogeny of the subfamily Pooideae based on chloroplast genome information to explore the phylogenetic position of *C. subtilis*; (c) identify highly variable loci to provide useful information for future development of molecular markers in *C. subtilis*.

## MATERIALS AND METHODS

### Sampling, Extraction, and Genome Sequencing

The materials of *Coleanthus subtilis* were collected from Harbin, China, in June 2021, and subsequently deposited in the Herbarium of the Wuhan Botanical Garden (HIB), Chinese Academy of Sciences (China), with herbarium number ZXX21129. For drying and long-term preservation of molecular samples, fresh leaves were preserved in silica gel (Chase and Hills, 2019). The complete genomic DNA of *C. subtilis* chloroplast was extracted using a modified CTAB procedure (Allen et al., 2006) and then sequenced at Novogene Co., Ltd. (Beijing, China) with Illumina paired-end technology platform. Purified high-quality genomic DNA was broken into short fragments of approximately 350 bp, and paired-end (PE) libraries were constructed by adding A-tails, PCR amplification and other steps, followed by sequencing in 150 bp paired-end mode on an Illumina HiSeq 2500 platform. The final number of raw reads obtained was 36,062,743 and that of clean reads after filtering was 35,335,540. The raw data has been uploaded to the NCBI database (BioProject ID: PRJNA802068).

### Assembly and Annotation of Chloroplast Genome

Get Organelle v1.7.4 (Jin et al., 2020) was used to assemble the chloroplast genome with default parameters. The low-quality reads and adapters were first filtered, then a *de novo* assembly performed, and the results were further purified to generate the complete chloroplast genomes. The results were visualized with Bandage (Wick et al., 2015). The Plastid Genome Annotator (PGA) software (Qu et al., 2019) was used to perform the annotation of the entire chloroplast genome, and in addition to using *Amborella trichopoda* as the reference genome, some Poaceae species were also selected to enhance the credibility of the annotation results. Furthermore, to ensure the accuracy of the annotation results, the genome was also annotated simultaneously with the help of GeSeq online tool<sup>2</sup> (Tillich et al., 2017).

The check of annotated genes was implemented in the software Geneious-v10.2.3 (Kearse et al., 2012), which was used to further verify and refine the annotation results and to manually correct errors detected in gene annotation. Special attention was paid to some genes located at the boundaries and the highly variable genes, such as *ndhF*, *ndhK*, *ycf2*, *accD*, etc. The circular chloroplast genome map of *Coleanthus subtilis* was drawn and

visualized using OGDraw online tool<sup>3</sup> (Greiner et al., 2019). Lastly, the annotated sequence was submitted to GenBank on the NCBI website, with an accession number OL692806.

### Comparative Analysis of the Chloroplast Genome

The chloroplast genome characteristics of *Coleanthus subtilis* were analyzed in Geneious-v10.2.3 software by comparing chloroplast genomes with those of Poaceae species downloaded from the NCBI database (Supplementary Table 1). A total of 24 species representing 10 subtribes (5 tribes) were used for the comparative analysis of chloroplast genomes. Additionally, to determine genomic divergence among these species, genomic similarity analysis was performed using the Glocal alignment program (shuffle-LAGAN mode) in mVISTA (Brudno et al., 2003; Frazer et al., 2004) with *C. subtilis* as the reference. The SC/IR boundary analysis was done using the IRscope (Amiryousefi et al., 2018) to observe the contraction and/or the expansion of the genes at the borders. For the codon usage bias analysis, MEGA 7.0 software (Kumar et al., 2016) was chosen to calculate relative synonymous codon usage (RSCU) values based on the coding sequences (CDS regions).

### Analysis of Repeats and Nucleotide Diversity

The REPuter tool<sup>4</sup> (Kurtz et al., 2001) was used to identify repeats including forward, reverse, palindrome, and complement sequences. When the Hamming distance is equal to 3, the length and identity of repeats are limited to  $\geq 30$  bp and  $> 90\%$ , respectively. The simple sequence repeats (SSRs) were analyzed using the MISA (Beier et al., 2017) with the basic repeat setting: a threshold of 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, respectively. The DnaSP-v5.10 software (Librado and Rozas, 2009) was used to calculate nucleotide variability ( $P_i$ ) values and variable sites using the aligned chloroplast genome sequences with a window length of 600 bp and a step size of 200 bp.

### Substitution Rate Analysis

The EasyCodeML program in PAML package (Gao et al., 2019) was utilized to identify positive sites in protein-coding genes to quantify selection pressure. This software provided four site models (M0 vs. M3, M1a vs. M2a, M7 vs. M8, and M8a vs. M8), Bayes Empirical Bayes (BEB) analysis (Yang et al., 2005) and Naive Empirical Bayes (NEB) analysis were performed in each model to measure the loci with positive selection pressure.

### Phylogenetic Analysis

To understand the phylogenetic position of *Coleanthus subtilis* in the family Poaceae and its affinities with other species, a phylogenetic tree was reconstructed using the Maximum Likelihood (Felsenstein, 1981) and Bayesian Inference analysis (Huelsenbeck et al., 2001). This was based on 76 shared protein-coding genes of the Chloroplast genome from a total of 53

<sup>2</sup><https://chlorobox.mpimp-golm.mpg.de/geseq.html>

<sup>3</sup><https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>

<sup>4</sup><https://bibiserv.cebitec.uni-bielefeld.de/reputer>

species from 26 genera in Poaceae, with *Acidosasa purpurea* as the outgroup (**Supplementary Table 1**). Each protein-coding sequence was first aligned in the software MAFFT-v7.409 (Kato and Standley, 2013), followed by removing the stop codon and discarding the bad fragment with the Gblock program (Talavera and Castresana, 2007) and later concatenated using the concatenated in-built PhyloSuite program (Zhang et al., 2020). ML analysis in IQ-tree and BI analysis in MrBayes were used to infer phylogenetic relationships. The best-fit models for each of the two analyses were found in Model Finder (Kalyaanamoorthy et al., 2017) according to the Bayesian Information Criterion (BIC), and the most suitable model for Bayesian analysis was detected as GTR + F + I + G4, while GTR + F + R3 was used for the Maximum Likelihood analysis. Subsequently, the BI tree was constructed by the software MrBayes-3.2.6 (Ronquist et al., 2012) for 1,000,000 generations, sampling every 1000 generations, and the software IQ-TREE was implemented to construct the ML tree with bootstrap replications of 1000 (Lam-Tung et al., 2015). The phylogenetic trees were visualized in the software Figtree-v1.4.4<sup>5</sup>. Both phylogenetic trees were combined manually using AI software based on consistent topological structures. The results were imported into the software Figtree-v1.4.4 to view the generated phylogenetic trees and to enhance their visualization. Considering the consistent topology, the phylogenetic trees constructed by both methods were manually combined in the AI software.

## RESULTS

### Chloroplast Genome Features

The chloroplast genome of *Coleanthus subtilis* is 135,915 bp in size and consists of four regions that together form a loop structure. These four regions are the large single copy region (LSC) of 80,100 bp, a small single copy region (SSC) of 12,757 bp, and two inverted repeat regions (IR) of 21,529 bp in length, respectively. In addition, a pair of inverted repeat regions separate the two single-copy regions (**Figure 1** and **Table 1**). GC content varies in different regions of the chloroplast genome. The highest GC content of 43.9% was found in the IR regions of *C. subtilis*, while the two single copy regions had 36.3% (LSC) and 32.4% (SSC) (**Table 1**).

A total of 129 genes were annotated in the chloroplast genome of *C. subtilis*, with 83 protein-coding genes (PCGs), 38 tRNA genes, and 8 rRNA genes. In addition, the *accD* gene was found missing in the chloroplast genome, while *ycf1*, *ycf2*, *ycf15*, and *ycf68* were pseudogenes (**Table 2**). These genes were divided into three groups based on their different functions. Nineteen genes were observed to replicate in the inverted repeat regions, seven of which were PCGs (*ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, *rps15*, *rps19*), eight were tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnH-GUG*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*) and the remaining four genes were rRNA (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*). In addition, the largest number of genes in the LSC region was 82, while only 11 genes were located in the SSC region. More

interestingly, all rRNA genes were distributed in the IR regions (**Supplementary Table 2**). We identified 15 genes containing one intron in *C. subtilis*, with six being tRNAs and nine being PCGs. It's important to highlight that *trnK-UUU* had the longest intron with 2480 bp, which completely wrapped the *MatK* gene. Meanwhile, the *ycf3* gene contained two codons with lengths of 774 bp and 726 bp (**Supplementary Table 3**).

The chloroplast genome of *C. subtilis* showed high similarities with other Poaceae species in terms of genome length and structure, GC content and gene number. The complete genomes length varied from 133608 bp (*Colpodium humile*) to 137370 bp (*Stipa purpurea*), LSC from 78636 bp (*Colpodium humile*) to 81252 bp (*Brachypodium stacei*), SSC from 12390 bp (*Zingeria biebersteiniana*) to 12842 bp (*Stipa purpurea*), and IR from 20831 bp (*Melica mutica*) to 22917 bp (*Briza maxima*) (**Table 1**). The overall GC content was around 38.5%, and each of the four regions also differed only insignificantly. In particular, gene number and composition were almost identical in 24 species, with only the *trnL-UAA* gene missing in *Bromus vulgaris*. In addition, no structural rearrangements were found in any of them (**Figure 2**).

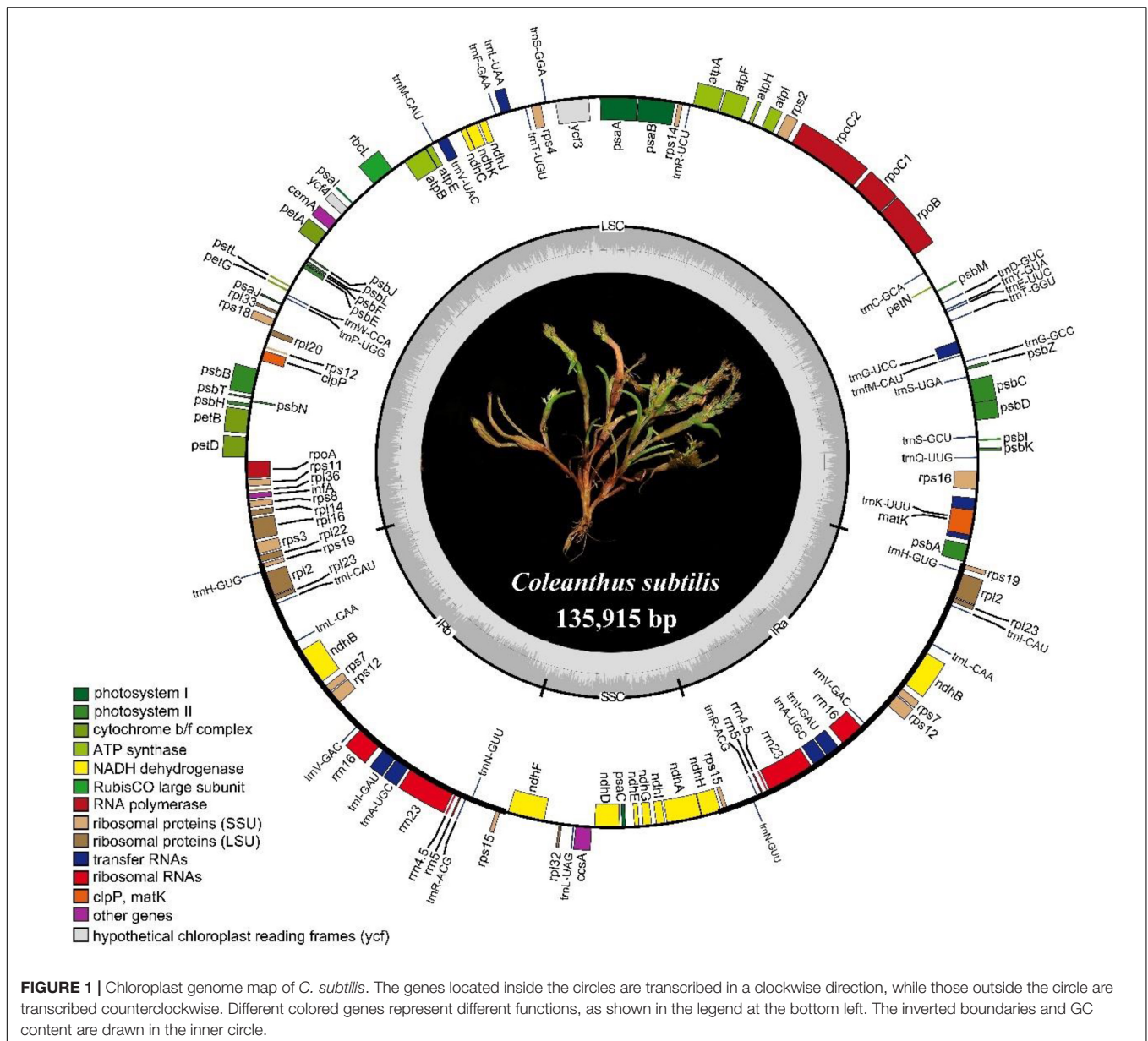
### Junction Characteristics

To observe the variation of IR boundaries, we did a comparative analysis of the junction structure based on the chloroplast genomes of *Coleanthus subtilis* and 23 other Poaceae species (**Figure 3**). The results showed that their boundary features were similar, the genes found at the nodes were mainly *rpl22*, *rps19*, *rps15*, *ndhF*, *ndhH*, and *psbA*. The *rps19* and *rps15* genes were replicated and fully embedded in the IR region, with lengths of 13–46 bp and 293–479 bp from the two IR/LSC boundaries, respectively. The *ndhF* genes were located entirely on the left of the IRb/SSC and were 27 to 122 bp from this boundary. Also, the *ndhH* gene occupied the IRa/SSC junction and was overwhelmingly located within the SSC region, with only a small portion of 156 to 316 bp extending into the IRa region. It should be noted that the *ndhH* gene of *Colpodium humile* was slightly shorter in length and was therefore completely encapsulated in the SSC. In addition, *Brachypodium stacei* and *Briza maxima* showed significant differences in boundary characteristics from the other species. It was clearly observed that the IR regions of *Brachypodium stacei* were contracted, resulting in the distribution of the *rps19* gene originally located in this region to the LSC. However, the IR region of *Briza maxima* expanded, wrapping the *rpl22* that should have been located in the LSC.

### Similarity Analysis of Chloroplast Genomes

Whole sequence alignment of the chloroplast genomes of 24 Poaceae species was performed to detect the differences that exist in their structures (**Figure 4**). The annotation of *Coleanthus subtilis* were used as a reference. The chloroplast genomes of these species were largely identical in terms of the number and arrangement of genes. However, some highly variable regions were still detected, such as *rbcl-psaI*, *psbE-petL*, *trnD-GUC-psbM*, *trnG-UCC-trnT-GGU*, *rpl32-trnL-UAG* and other

<sup>5</sup><https://www.figtreeasia.com/>



intergenic regions. Overall, the non-coding regions showed a higher potential for variation compared to the coding regions. Although the protein-coding regions were relatively conserved, larger variants were observed in the *rpoC2*, *infA*, *cemA* and *matK* genes. Besides, variations were also presented in some genes located at the IR/SC boundary, such as *rps19* and *ndhF*. However, the rRNA and tRNA sequences were highly conserved, where genes such as *rrn16*, *rrn23*, *trnV-GAC*, and *trnR-ACG* were almost unchanged. At the same time, IR regions of these species were minimally altered and significantly more conserved than the two single-copy regions.

## Codon Usage Analysis

There were 19838 codons eventually found in chloroplast genome of *Coleanthus subtilis*. Methionine and Tryptophan amino acids

were encoded by a single codon, AUG and CGG, respectively. However, the remaining amino acids were encoded by two to six codons and showed a clear preference for codon usage (**Figure 5**). The most abundant amino acid in the *C. subtilis* was leucine 2135 (10.76%). Conversely, the least abundant amino acid was cysteine 218, which accounted for only 1.10% of the total. Meanwhile, among the six codons encoding leucine, UUA had the highest RSCU value of 2.10, which indicated that it had a high preference and was the most commonly used codon. Interestingly, most of the codons with RSCU values greater than 1 had A/U as the terminal codon, while those with C/G as the terminal codon usually had RSCU values less than 1.

The RSCU values of the five species were compared in order to understand the differences in their codon usage

**TABLE 1** | Features of the chloroplast genomes of *C. subtilis* and other Poaceae species.

Species	Genome length (bp)				GC content (%)				Gene number			
	Total	LSC	SSC	IR	Total	LSC	SSC	IR	Total	PCG	tRNA	rRNA
<i>Coleanthus subtilis</i>	135915	80100	12757	21529	38.3	36.3	32.4	43.9	129	83	38	8(4)
<i>Phippsia algida</i>	135613	79805	12750	21529	38.3	36.3	32.4	43.9	129	83	38	8(4)
<i>Puccinellia nuttalliana</i>	135353	79594	12727	21516	38.3	36.3	32.6	43.9	129	83	38	8(4)
<i>Sclerochloa dura</i>	135086	79683	12609	21397	38.3	36.2	32.6	43.9	129	83	38	8(4)
<i>Zingeria biebersteiniana</i>	135943	80121	12390	21491	38.3	36.2	32.7	43.9	129	83	38	8(4)
<i>Agrostis gigantea</i>	136705	80604	12769	21666	38.5	36.4	32.8	44.0	129	83	38	8(4)
<i>Alopecurus japonicus</i>	136408	80511	12835	21531	38.3	36.2	32.4	43.9	129	83	38	8(4)
<i>Ammophila breviligulata</i>	136726	80711	12701	21657	38.6	36.5	32.9	44.1	129	83	38	8(4)
<i>Anthoxanthum odoratum</i>	135551	79626	12671	21627	38.2	36.1	32.5	43.9	129	83	38	8(4)
<i>Avena barbata</i>	135946	80111	12625	21605	38.5	36.4	32.6	44.0	129	83	38	8(4)
<i>Brachypodium stacei</i>	136330	81252	12666	21206	38.6	36.6	32.7	44.1	129	83	38	8(4)
<i>Briza maxima</i>	136823	79707	12722	22917	38.3	36.2	32.6	43.7	129	83	38	8(4)
<i>Bromus vulgaris</i>	136934	80964	12566	21702	38.3	36.3	32.3	43.9	128	83	37	8(4)
<i>Calamagrostis pickeringii</i>	136682	80660	12688	21667	38.6	36.5	32.9	44.0	129	83	38	8(4)
<i>Castellia tuberculosa</i>	133798	78819	12497	21241	38.4	36.3	32.6	43.9	129	83	38	8(4)
<i>Colpodium humile</i>	133608	78636	12474	21249	38.3	36.2	32.7	43.9	129	83	38	8(4)
<i>Festuca altissima</i>	135272	79826	12598	21424	38.4	36.4	32.8	43.9	129	83	38	8(4)
<i>Hierochloa odorata</i>	136395	80645	12466	21642	38.5	36.4	33.0	44.0	129	83	38	8(4)
<i>Lolium multiflorum</i>	135175	79848	12485	21421	38.3	36.1	32.4	43.9	129	83	38	8(4)
<i>Melica mutica</i>	134710	80478	12570	20831	38.5	36.5	32.8	44.0	129	83	38	8(4)
<i>Phalaris coerulescens</i>	135794	79728	12760	21653	38.5	36.4	32.9	44.0	129	83	38	8(4)
<i>Phleum alpinum</i>	135568	80009	12823	21368	38.4	36.3	32.6	44.0	129	83	38	8(4)
<i>Poa diaphora</i>	135466	79629	12685	21576	38.3	36.2	32.4	43.8	129	83	38	8(4)
<i>Stipa purpurea</i>	137370	81202	12842	21663	38.8	36.9	32.9	44.0	129	83	38	8(4)

**TABLE 2** | List of the annotated genes in the chloroplast genomes of *C. subtilis*.

Category	Groups of genes	Name of genes
Self-replication	Ribosomal RNA	<i>rrn4.5<sup>c</sup>, rrn5<sup>c</sup>, rrn16<sup>c</sup>, rrn23<sup>c</sup></i>
	Transfer RNA	<i>trnA-UGC<sup>a,c</sup>, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnI-M-CAU, trnG-GCC, trnG-UCC<sup>a</sup>, trnH-GUG<sup>c</sup>, trnI-CAU<sup>c</sup>, trnI-GAU<sup>a,c</sup>, trnK-UUU<sup>a</sup>, trnL-CAA<sup>c</sup>, trnL-UAA<sup>a</sup>, trnL-UAG, trnM-CAU, trnN-GUU<sup>c</sup>, trnP-UGG, trnQ-UUG, trnR-UCU, trnR-ACG<sup>c</sup>, trnS-UGA, trnS-GCU, trnS-GGA, trnT-GGU, trnT-UGU, trnV-UAC<sup>a</sup>, trnV-GAC<sup>c</sup>, trnW-CCA, trnY-GUA</i>
	Small subunit of ribosome	<i>rps2, rps3, rps4, rps7<sup>c</sup>, rps8, rps11, rps12<sup>a,c</sup>, rps14, rps15<sup>c</sup>, rps16<sup>a</sup>, rps18, rps19<sup>c</sup></i>
	Large subunit of ribosome	<i>rpl2<sup>a,c</sup>, rpl14, rpl16<sup>a</sup>, rpl20, rpl22, rpl23<sup>c</sup>, rpl32, rpl33, rpl36</i>
	RNA polymerase subunits	<i>rpoA, rpoB, rpoC1, rpoC2</i>
Photosynthesis	Photosystem I	<i>psaA, psaB, psaC, psal, psaJ, ycf3<sup>b</sup>, ycf4</i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbl, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunits of cytochrome	<i>petA, petB<sup>a</sup>, petD<sup>a</sup>, petG, petL, petN</i>
	ATP synthase	<i>atpA, atpB, atpE, atpF<sup>a</sup>, atpH, atpI</i>
	NADH-dehydrogenase	<i>ndhA<sup>a</sup>, ndhB<sup>a,c</sup>, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
Other genes	Rubisco large subunit	<i>rbcL</i>
	Translational initiation factor	<i>infA</i>
	Maturase K	<i>matK</i>
	Envelope membrane protein	<i>cemA</i>
	Proteolysis	<i>clpP</i>
	Cytochrome c biogenesis	<i>ccsA</i>

<sup>a</sup>Genes with one intron. <sup>b</sup>Genes with two introns. <sup>c</sup>Two gene copied in IR regions.

(Figure 6). For one amino acid, the sum of the RSCU values of all codons involved in its encoding was almost equal. Also, the RSCU values of the same codons were almost

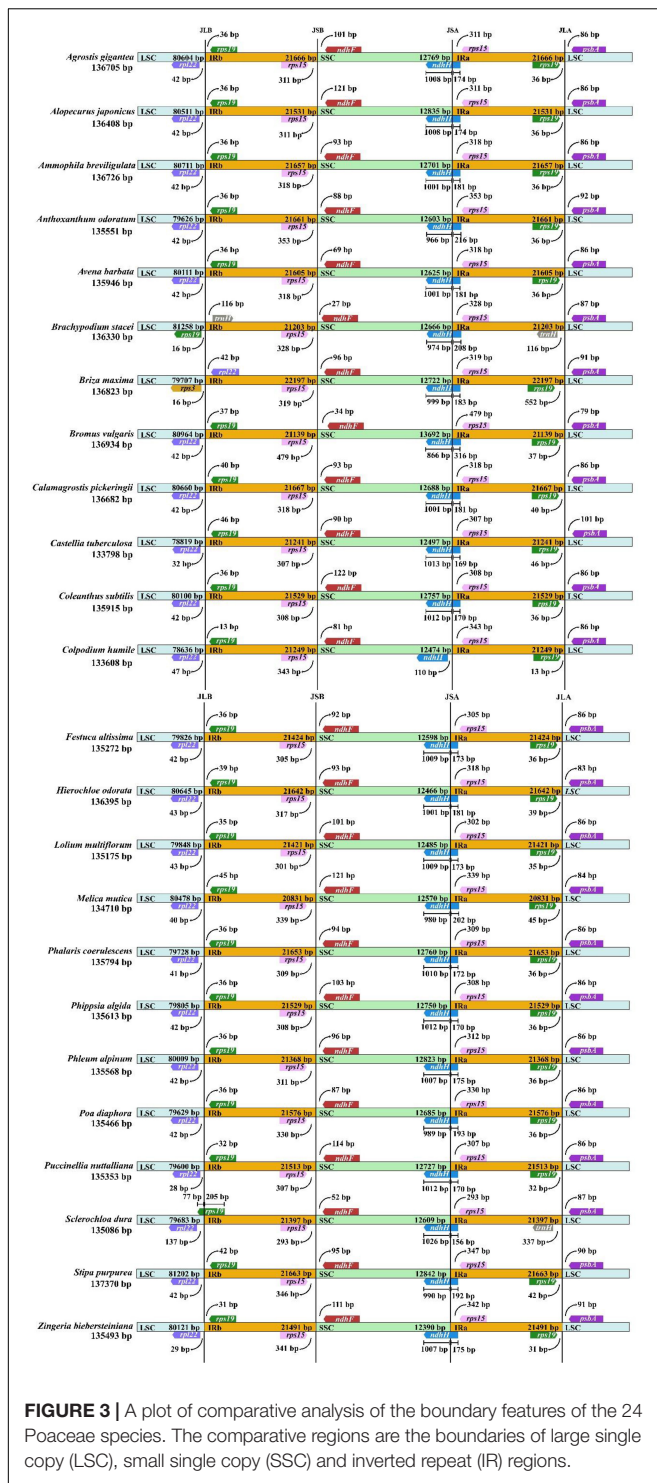
identical in these species, indicating that their codon usage habits were more stable and hardly change (Figure 6 and Supplementary Tables 4, 5).



## Repeat Analysis

We detected only palindromic and forward repeats in chloroplast genomes of *Coleanthus subtilis* and its related species, where the proportion of forward repeats was

higher than that of palindromic repeats (Figure 7A and Supplementary Table 6). Most of repeats were 30–34 bp in length and were mainly distributed in the LSC region (Figures 7B,D and Supplementary Tables 7, 8). Also, the CDS



**FIGURE 3** | A plot of comparative analysis of the boundary features of the 24 Poaceae species. The comparative regions are the boundaries of large single copy (LSC), small single copy (SSC) and inverted repeat (IR) regions.

regions contained most of the repeats, followed by the IGS regions (Figure 7C and Supplementary Table 9). Some repeats were also shared between IGS, CDS, tRNA, and intron regions.

A total of 26 SSRs were detected in *C. subtilis*, while 28, 28, 30, and 33 microsatellites were found in *Phippsia algida*, *Puccinellia nuttalliana*, *Sclerochloa dura*, and *Zingeria bebeersteiniana*,

respectively (Figure 8A and Supplementary Table 10). These SSRs were classified into five types, namely Mono-, di-, tri-, tetra-, and penta-nucleotides repeats. The mono-nucleotide repeats accounted for 50.34% of the 145 microsatellites and were the most abundant SSR types in the five species, followed by tetra-nucleotide repeats (26.21%). Most of the microsatellites were distributed in the LSC region and consisted of A/T motifs (Figures 8B,C and Supplementary Tables 11, 12).

## Nucleotide Diversity ( $P_i$ ) and Selection Pressure Analysis

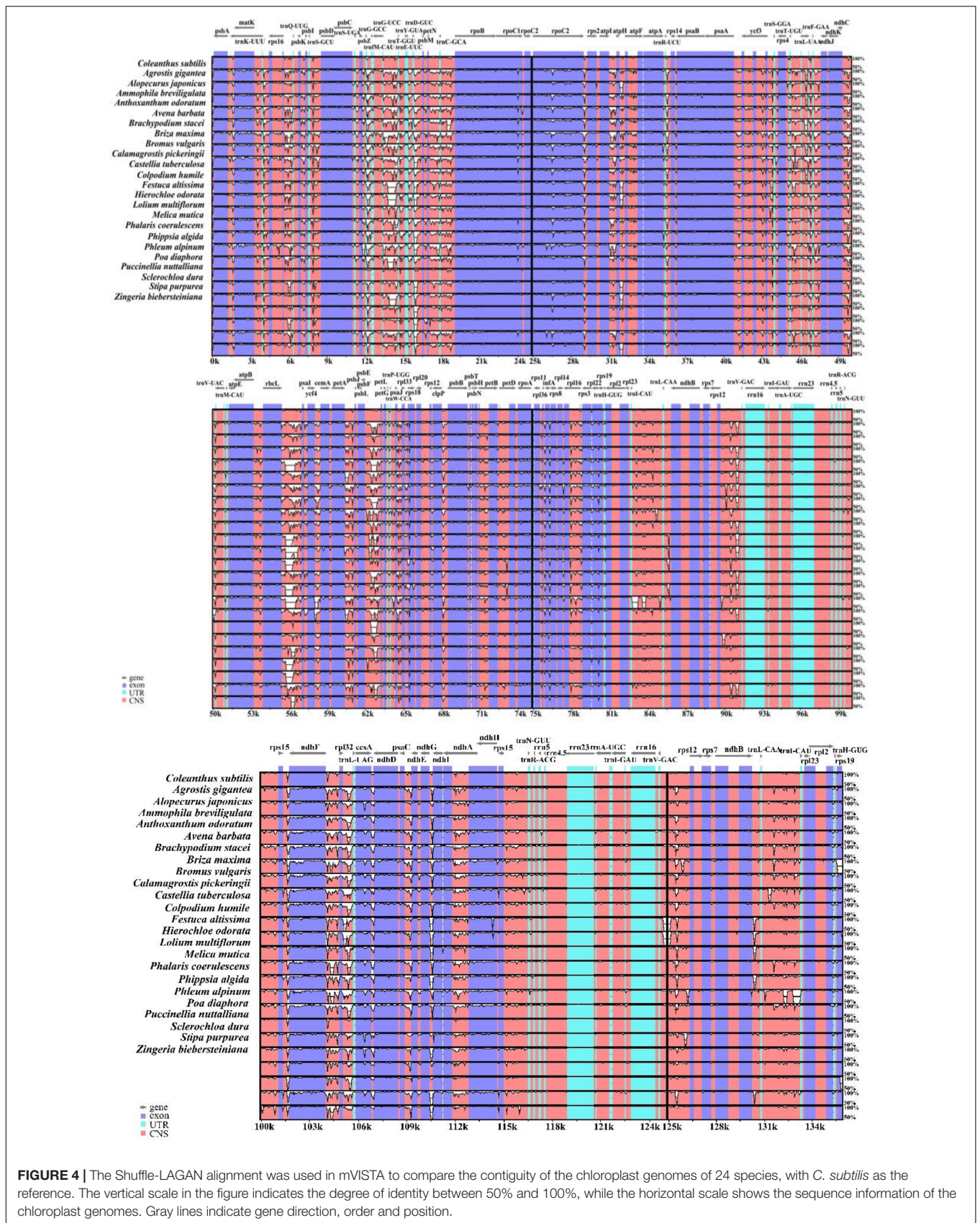
To comprehensively understand the sequence divergence of the chloroplast genomes of *Coleanthus subtilis* and its related species, we calculated  $P_i$  values for nucleotide diversity.  $P_i$  values fluctuated between 0 and 0.0697, with a mean value of 0.02172 (Figure 9). We identified 13 polymorphic regions (*matK*, *trnK-UUU/rps16*, *rps16/trnQ-UUG*, *trnG-UCC/trnT-GGU*, *trnT-GGU/trnE-UUC*, *petN/trnC-GCA*, *trnC-GCA/rpoB*, *rps4/trnL-UAA*, *trnL-UAA/ndhJ*, *ndhC/trnV-UAC*, *ndhF*, *ndhF/rpl32*, and *ndhA*) with nucleotide diversity  $>0.05$ , 10 of which were intergenic spacer regions and the remaining three were protein-coding regions. Meanwhile, no highly variable loci were detected in the IR regions and the nucleotide diversity values were significantly lower than those in the single copy regions (Figure 9 and Supplementary Table 13).

In this study, dN/dS values were calculated based on 76 CDS regions with site models in EasyCodeML. According to the M8 model, only the *atpF* gene possessed a significant positive site in the BEB approach (Table 3). Meanwhile, a total of 45 loci corresponding to 21 genes were identified in the NEB method, of which 13 genes (*atpA*, *atpF*, *atpI*, *csaA*, *clpP*, *infA*, *ndhA*, *ndhD*, *ndhK*, *rbcl*, *rpoA*, *rps16*, and *rps3*) had a significant positive site. In addition, the *ndhF*, *psaA*, *psaB*, *psbC*, and *rpoC1* genes contained two significant positive selection loci, while the *cemA* and *matK* genes were detected with three and eight loci under positive selection, respectively. Moreover, the *rpoC2* gene was found to have the highest number of positive selection sites, including 11 significant positive sites.

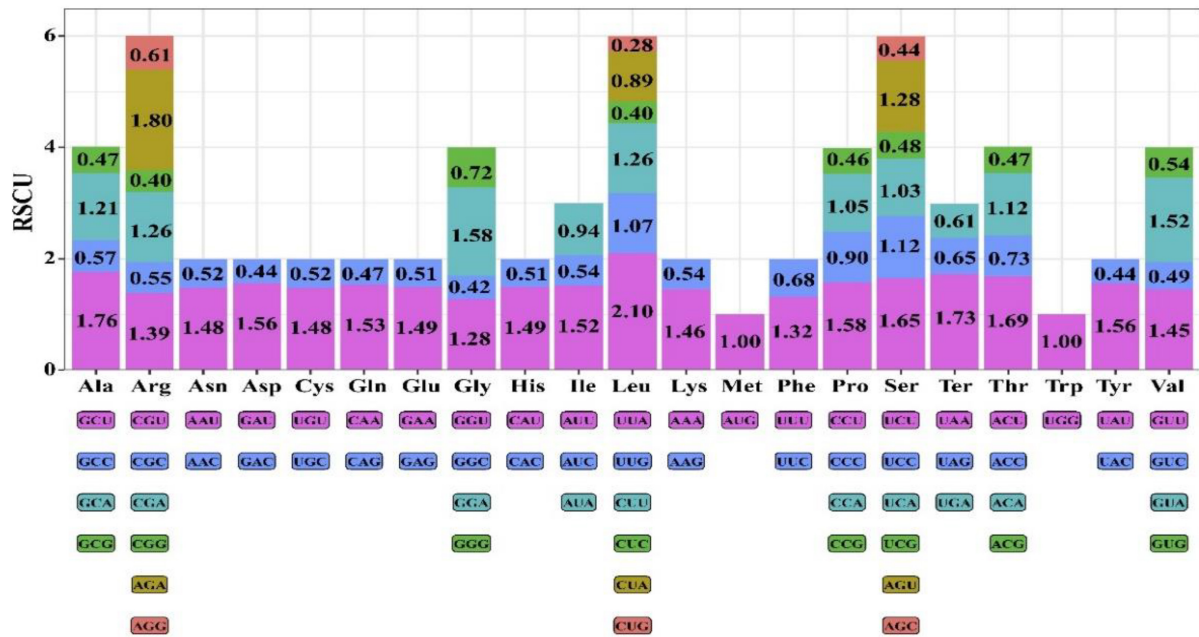
## Phylogenetic Analysis

In the current study, we utilized the protein-coding regions of chloroplast genomes for the first time to explore the phylogenetic position of *Coleanthus subtilis*. The topologies of the phylogenetic trees generated with maximum likelihood (ML) and Bayesian analysis (BI) were identical, with generally high branch bootstrap values and posterior probabilities. Based on consistent topologies, we showed the phylogenetic tree represented by the ML method (Figure 10). The 53 species representing 26 genera were divided into ten subtribes and six tribes. Among them, *Coleanthus* was placed in the big clade containing *Phippsia*, *Puccinellia*, *Sclerochloa*, and *Zingeria*, which were components of the subtribe Coleanthinae. In addition, *C. subtilis* formed a sister branch with the genus *Phippsia*, while this branch was also sister to other taxa of this subtribe (BS = 100, PP = 1). The genus *Colpodium* was nested in the subclade Loliinae and had a sister relationship with the genus *Castilleja* (BS = 100, PP = 1).

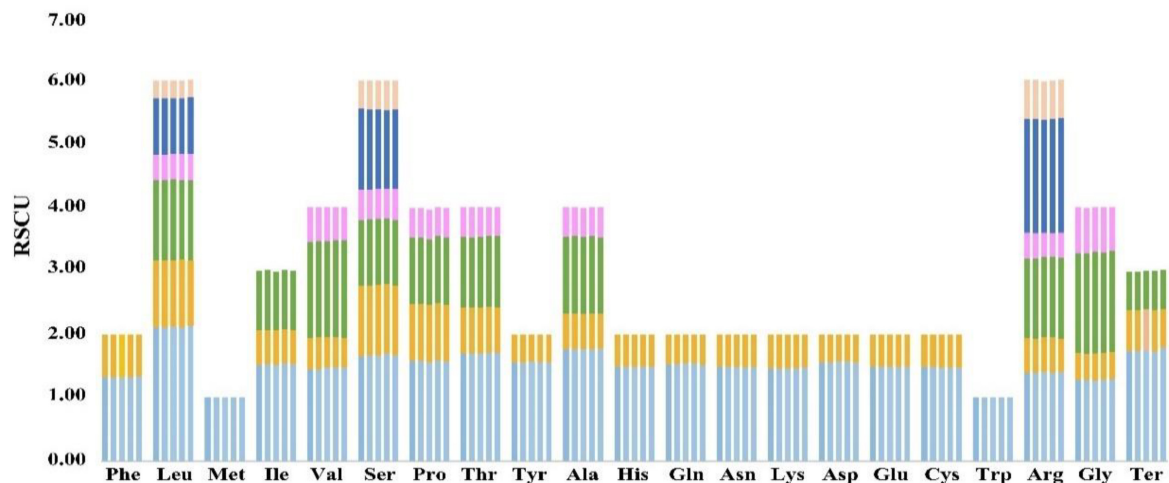




**FIGURE 4 |** The Shuffle-LAGAN alignment was used in mVISTA to compare the contiguity of the chloroplast genomes of 24 species, with *C. subtilis* as the reference. The vertical scale in the figure indicates the degree of identity between 50% and 100%, while the horizontal scale shows the sequence information of the chloroplast genomes. Gray lines indicate gene direction, order and position.



**FIGURE 5** | Relative synonymous codon usage (RSCU) values for amino acids and stop codons of the 76 protein-coding regions of *C. subtilis*. The colors of the histograms correspond to the colors of the codons.



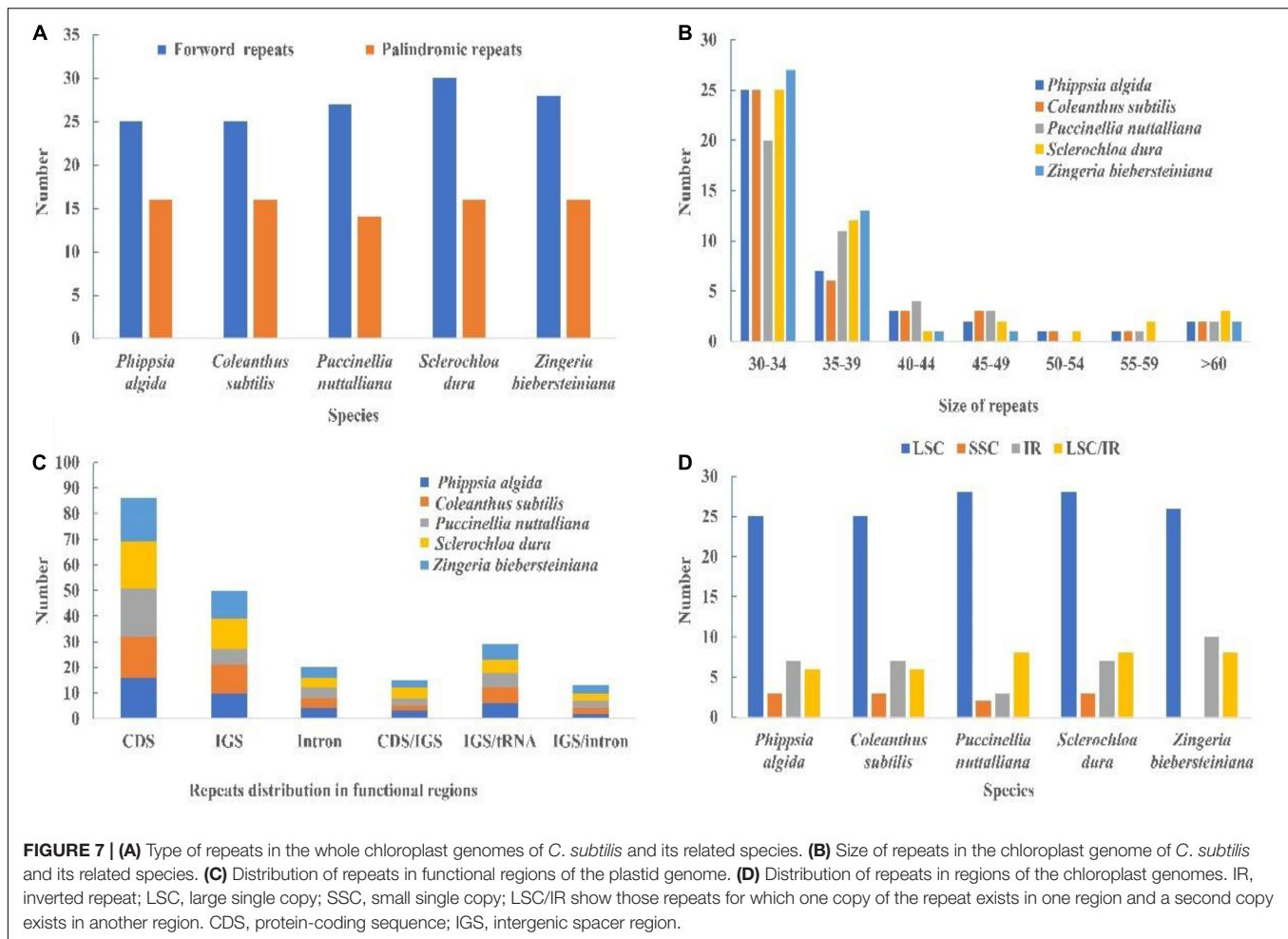
**FIGURE 6** | Comparative analysis plots of RSCU values for the five species. Each amino acid corresponds to five histograms, and their heights represent the RSCU value. The histogram from left to right is *Coleanthus subtilis*, *Phippsia algida*, *Puccinellia nuttalliana*, *Sclerochloa dura*, and *Zingeria biebersteiniana*.

## DISCUSSION

### Plastome Comparison of *Coleanthus subtilis* and Other Species Within Pooideae

The chloroplast genome of *Coleanthus subtilis* exhibited a tetrad structure of 135915 bp in length, which is similar to the length and structural characteristics of cp genomes of other higher plants (Jansen et al., 2005; Daniell et al., 2016). We found

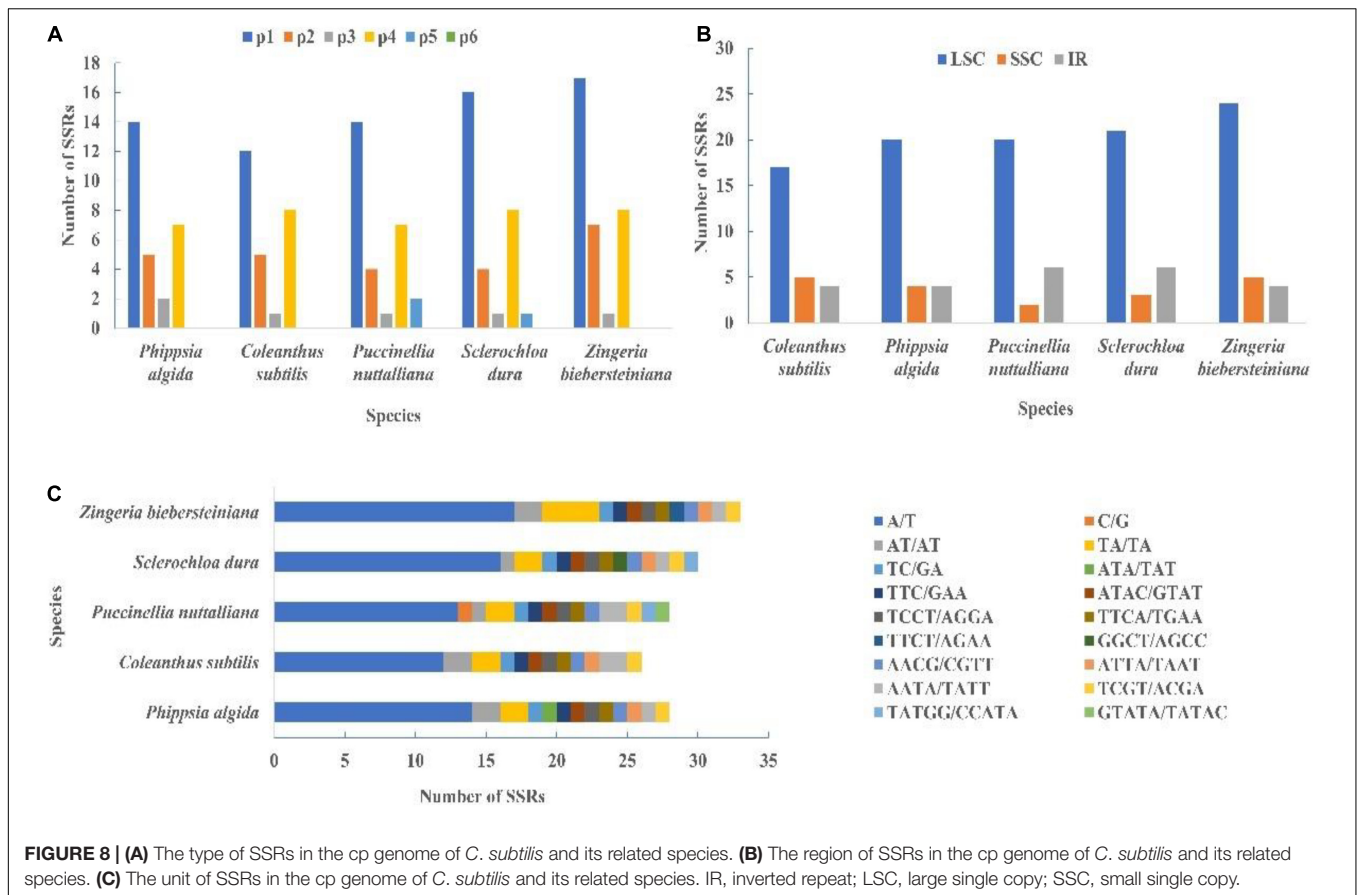
that the GC content in the cp genome of Pooideae species was unevenly distributed, with the IR regions having a higher GC content than the two single copy regions. This may be attributed to the fact that four rRNA genes with high GC content were located in the IR regions, which supported the speculation of previous studies (Mardanov et al., 2008; Gao et al., 2009; Wanga et al., 2021). The *accD* gene has been lost within the cp genomes of Pooideae species, while *ycf1*, *ycf2*, *ycf15*, and *ycf68* were pseudogenes, which is a relatively common phenomenon



in Poaceae (Huang et al., 2017). There is a correlation between gene loss and evolution, and some studies suggest that it may be an adaptive strategy with positive effects on survival and reproduction (Xu and Guo, 2020). In addition, we also found *trnL-UAA* gene loss in *Bromus vulgaris*. Pseudogenization of tRNA (*trnT-GGU*) has also been observed in the Asteraceae family (Abdullah et al., 2021a). Sixteen intron-containing genes were detected in 24 species in which introns of *rpoC1* and *clpP* genes were lost. Besides, the *trnK-UUU* has the longest intron that completely wraps the *matK* gene, a result that has been reported in other studies (Li X. et al., 2019; Souza et al., 2020). The *rpoC1* gene has been reported to contain introns in most land plants (Ohyama et al., 1986; Kugita et al., 2003). However, deletion of the *rpoC1* intron was observed in some angiosperm lineages, such as most Poaceae and some species of the families Fabaceae, Cactaceae, and Aizoaceae (Downie et al., 1996; Wallace and Cota, 1996; Huang et al., 2017). Our study on the subfamily Pooideae further confirm that the absence of the *rpoC1* intron is universal in the Poaceae. Similarly, the *clpP* gene usually contained two introns. Nevertheless, both introns have been lost in *Pinus* and some species from the genera *Oenothera*, *Silene*, and *Menodora* (Lee et al., 2007; Huang et al., 2017). Also, it was demonstrated that the loss of *clpP* introns were present in all

Poaceae species (Guisinger et al., 2010), which was supported by our findings. This study revealed that genomic structure, gene content and total GC content were significantly similar or identical within 24 genera from Pooideae, which were consistent with the genus *Blumea* and the families Solanaceae, Malvaceae, and Araceae (Abdullah et al., 2020b,c, 2021b).

Length variation in the IR region of the chloroplast genome was a common phenomenon during the evolution of land plants, which has led to the formation of diverse boundary features (Yang et al., 2010; Wang et al., 2017; Ding et al., 2021). The study demonstrated that boundary genes in the species of the subfamily Pooideae were mainly *rpl22*, *rps19*, *rps15*, *ndhF*, *ndhH*, and *psbA*, which differ from the boundaries of *Clethra* and *Blumea* species (Abdullah et al., 2021b; Ding et al., 2021). In general, the subfamily Pooideae shared many similarities at the nodes, which further endorsed the idea that the boundary features were relatively stable among closely related species (Liu et al., 2018). This phenomenon has also been observed in the subfamily Asteroideae (Abdullah et al., 2021b). However, distinct junction characteristics also existed in related species, such as *Brachypodium stacei* and *Briza maxima*. The present study found that although both were species of the subfamily Pooideae, they formed different boundary features due to noticeable



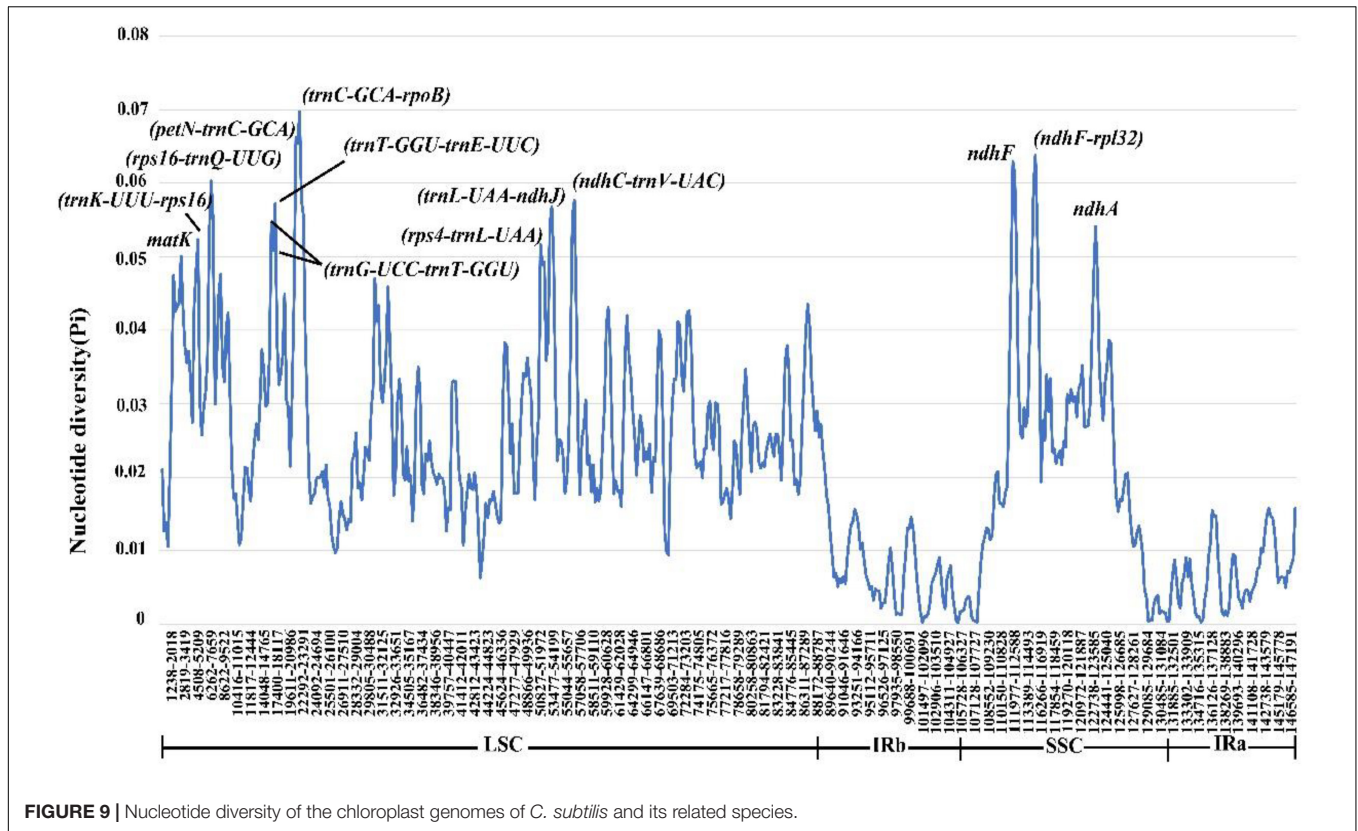
contraction or expansion of the IR Regions, respectively. The same findings were also noted in the genera *Pelargonium* and *Psilotum* (Chumley et al., 2006; Grewe et al., 2013; Sun et al., 2013).

The results of the mVISTA analysis showed that the coding regions were more conserved than the non-coding regions in the cp genomes of the subfamily Pooideae, and the two single copy regions showed higher variation potential than the IR regions. These two findings agreed with previous studies in other plant taxa (Gu et al., 2016; Xu et al., 2017; Alzahrani et al., 2020). We detected some highly variable non-coding regions, such as *rbcl-psaI*, *psbE-petL*, *trnD-GUC-psbM*, and *rpl32-trnL-UAG*. Despite the relative conservation of the protein-coding regions, variations were also observed in *rpoC2*, *infA*, *cemA*, and *matK* genes. The highly variable regions detected in this study were promising to be developed as specific DNA barcodes for the subfamily Pooideae, which has positive implications for the identification of species. In addition, the high GC content might be one of the reasons for less variation in tRNA sequences and IR regions, which further demonstrates the significance of GC content in maintaining sequence stability (Necsulea and Lobry, 2007; Kim et al., 2019).

The codon usage preference is closely related to gene expression and affects protein and mRNA levels in the genome (Zhou et al., 2013; Lyu and Liu, 2020). The most abundant amino acid in the *C. subtilis* was leucine 2135 (10.76%), which has also

been frequently reported in the chloroplast genomes of other angiosperms (Jian et al., 2018; Somaratne et al., 2019). More interestingly, most codons ending in A/U have RSCU values greater than 1, while those ending in C/G are less than 1. This pattern also applies to the preference of codon usage in other plants (Wang et al., 2018; Liu X.Y. et al., 2020).

Oligonucleotide repeats are very common in plastid genome and are thought to be a proxy for identifying mutational hotspots (Ahmed et al., 2012; Lee et al., 2014; Abdullah et al., 2020a,d; Liu Q. et al., 2020). In the present study, we detected both forward and palindromic repeats, mostly distributed in the LSC region. Additionally, most of the repeats were 30–40 bp in length, which was similar to those found in other species (Chen et al., 2018; Li D.M. et al., 2019; Wu et al., 2020). Simple sequence repeats (SSRs) were often used as a molecular marker to explore population relationships and evolutionary history due to its polymorphism, co-dominance and reliability (Oliveira et al., 2006; Sonah et al., 2011; Gao et al., 2018). A total of five types of SSRs were detected in the cp genomes of *C. subtilis* and its related species, of which mono-nucleotide repeats were the most common. Similarly, the most abundant SSR type in the genus *Quercus* was also mono-nucleotide repeats (Yang et al., 2016). However, there are other possibilities, such as tri-nucleotide repeats occurring most frequently in *Urophysa* (Xie et al., 2018). Furthermore, this study not only found that most SSR types were mono-nucleotide repeats, but they had A/T preference. This phenomenon can

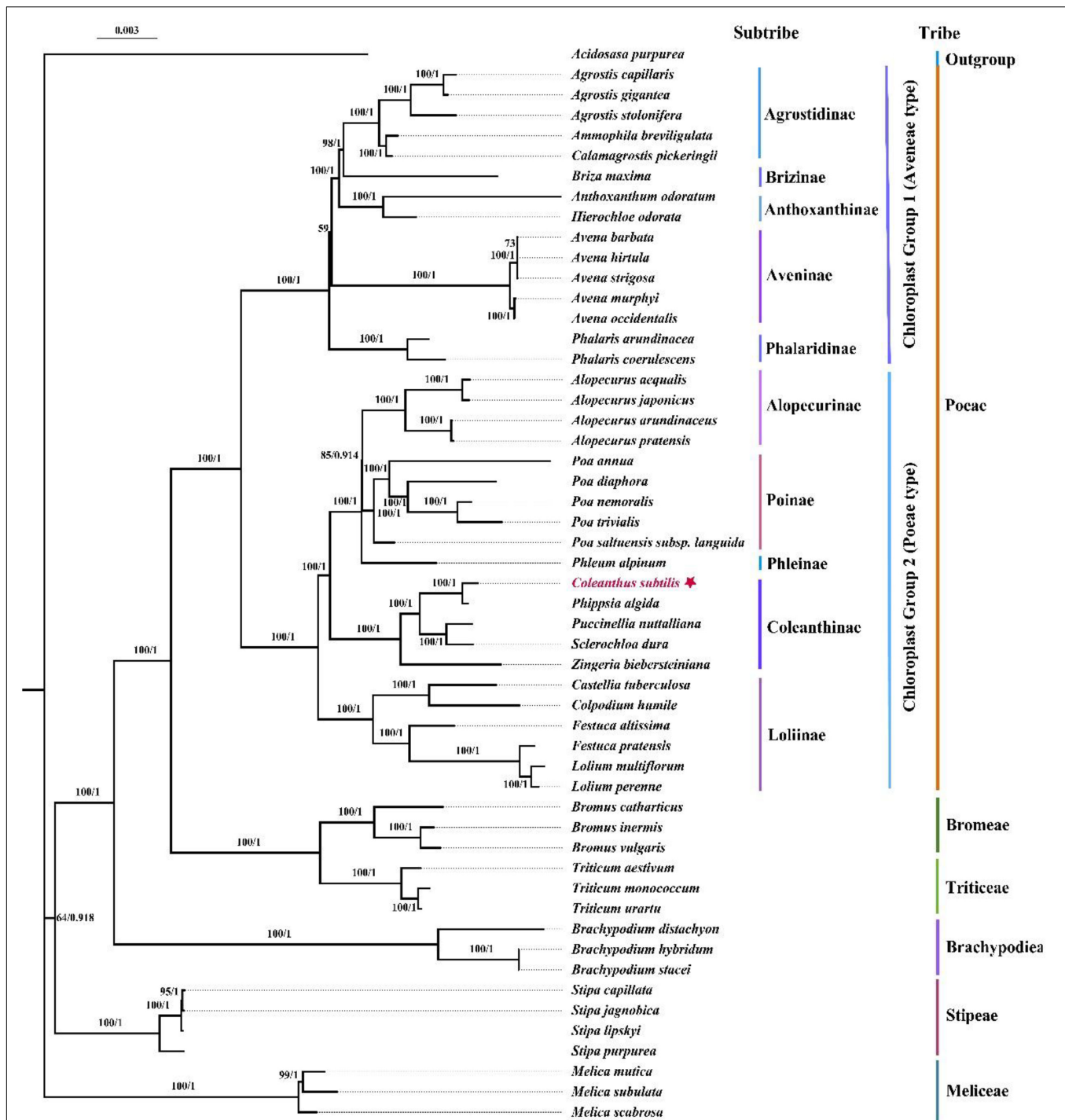


**FIGURE 9** | Nucleotide diversity of the chloroplast genomes of *C. subtilis* and its related species.

**TABLE 3** | dN/dS ratios of the chloroplast genomes of *C. subtilis* and its related species.

M8	Gene	Region	Selected sites	Pr(w > 1)	Number of sites
Naive Empirical Bayes (NEB)	<i>atpA</i>	LSC	133 A	0.958*	1
	<i>atpF</i>	LSC	1247 A	1.000**	1
	<i>atpI</i>	LSC	1646 L	0.998**	1
	<i>ccsA</i>	SSC	1824 F	0.963*	1
	<i>cemA</i>	LSC	1989 L/2043 R/2092 F	0.969*/0.958*/0.986*	3
	<i>clpP</i>	LSC	2370 E	0.964*	1
	<i>infA</i>	LSC	2521 H	0.966*	1
	<i>matK</i>	LSC	2558 C/2599 S/2711 I/2713 I/2805 L/2807 L/2882 V/2937 Q	0.958*/0.954*/0.951*/0.956*/ 0.956*/0.965*/0.953*/0.970*	8
	<i>ndhA</i>	SSC	3322 G	0.954*	1
	<i>ndhD</i>	SSC	4425 I	0.995**	1
	<i>ndhF</i>	SSC	5107 S/5119 I	0.952*/0.951*	2
	<i>ndhK</i>	LSC	6446 K	0.959*	1
	<i>psaA</i>	LSC	7436 L/7917 V	0.954*/0.960*	2
	<i>psaB</i>	LSC	8195 S/8372 L	0.954*/0.950*	2
	<i>psbC</i>	LSC	10060 F/10233 F	0.951*/0.953*	2
	<i>rbcl</i>	LSC	11311 V	0.955*	1
	<i>rpoA</i>	LSC	12867 G	0.959*	1
	<i>rpoC1</i>	LSC	14684 G/14758 R	0.967*/0.959*	2
	<i>rpoC2</i>	LSC	15415 N/15426 I/15775 A/15836 F/15841 S/15860 K/15891 K/15907 F/16026 E/16204 Q/16260 A	0.965*/0.968*/0.998**/0.958*/0.953*/ 0.962*/0.962*/0.964*/0.964*/ 0.961*/0.965*	11
	<i>rps16</i>	LSC	16812 T	0.967*	1
<i>rps3</i>	LSC	17399 E	0.955*	1	
Bayes Empirical Bayes (BEB)	<i>atpF</i>	LSC	1247 A	0.974*	1

The level of significance is indicated by the number of "\*", where "\*" represents significant, and "\*\*\*" indicates highly significant.



**FIGURE 10 |** Phylogenetic trees were generated based on the 76 shared protein-coding sequences of 53 species using maximum likelihood (ML) and Bayesian (BI) methods. The ML tree and BI tree have a consistent topology. The ML bootstrap values/Bayesian posterior probabilities are displayed on the nodes. To make *Coleanthus subtilis* more visible, it was marked with a star.

also be observed in numerous other taxa (Wheeler et al., 2014; Munyao et al., 2020).

We identified 13 polymorphic regions (*matK*, *trnK-UUU/rps16*, *rps16/trnQ-UUG*, *trnG-UCC/trnT-GGU*,

*trnT-GGU/trnE-UUC*, *petN/trnC-GCA*, *trnC-GCA/rpoB*, *rps4/trnL-UAA*, *trnL-UAA/ndhJ*, *ndhC/trnV-UAC*, *ndhF*, *ndhF/rpl32*, and *ndhA*) with nucleotide diversity >0.05, mainly located in the LSC region. In addition, the nucleotide diversity

values within the IR regions were significantly lower than those in the single copy regions, which is consistent with the pattern found in previous studies (Li D.M. et al., 2019; Ding et al., 2021). The dN/dS analysis was regarded as one of the most popular and reliable measures to quantify selective pressure (Kryazhimskiy and Plotkin, 2008; Mugal et al., 2014). We performed a selection pressure analysis on different genera of the subfamily Pooideae, and the result indicated that there are some genes under positive selective pressure, which was crucial for understanding the evolutionary history of these genera. The positively selected genes identified were nearly identical to those previously reported for other species in the family Poaceae, and our findings further support the plausibility of these loci (Piot et al., 2018). Furthermore, these genes are associated with photosynthesis, self-expression and regulatory activity (Piot et al., 2018), which has a positive effect on understanding the mechanisms of selection pressure generation.

## Phylogenetic Analysis

In the current study, the 76 protein-coding regions of the chloroplast genome were used for the first time to explore the phylogenetic position of *Coleanthus subtilis*. The reconstructed phylogenetic tree divided the 53 species into ten subtribes and six tribes, which coincided with the broad framework of the Poaceae phylogeny (Soreng et al., 2017; Saarela et al., 2018; Tkach et al., 2020). Phylogenetic analysis strongly demonstrated that *C. subtilis* formed a sister branch with the genus *Phippisia* (BS = 100, PP = 1), which further justified the results of previous morphological treatments and phylogenetic studies based on chloroplast fragments (Tzvelev, 1976; Soreng et al., 2015; Gnutikov et al., 2020). Moreover, our data revealed that *Colpodium* was nested in the subtribe Loliinae and was particularly closely related to the genus *Castellia*, while *Zingeria* was located in the subtribe Coleanthinae (BS = 100, PP = 1). This finding differed from that of earlier studies and provided a new perspective on the relationships between *Colpodium*, *Zingeria* and Coleanthinae. Some previous studies suggested that the genera *Zingeria* and *Colpodium* are sister groups and rather distantly related to the subtribe Coleanthinae, forming a branch known as the two-chromosome grasses (Rodionov et al., 2008; Kim et al., 2009). At the same time, these two genera were considered as constituent members of Coleanthinae (Soreng et al., 2015). However, apart from the fact that *Zingeria* belongs to the subtribe Coleanthinae, our results do not support the previously reported relationship between *Colpodium*, *Zingeria* and Coleanthinae. This work will not only contribute to further insight into the phylogenetic position of *C. subtilis* and the composition of the subtribe Coleanthinae, but also provide valuable chloroplast genomic information for future exploration of the origin and differentiation between *C. subtilis* and its related species at the cp genome level.

## CONCLUSION

In this study, the complete chloroplast genome of *Coleanthus subtilis* was reported and comparative and phylogenetic analyses

with its closely related species revealed, as well as differences in their genomic structure and composition. Although the chloroplast genome of *C. subtilis* is relatively conserved, 26 SSRs and 13 highly variable loci were detected, which could be developed as important genetic markers. The reconstructed phylogenetic tree further confirmed the sister relationship between *Coleanthus* and *Phippisia*, and also provided new insights into the relationship between *Coleanthus*, *Zingeria* and *Colpodium*. In addition, since *C. subtilis* is rare and legally protected, the genetic information is important for its breeding and conservation. Equally important, the mechanisms that lead to the unique distribution pattern of *C. subtilis* are unknown, which makes the species of great research value. Our results will enrich data and provide a useful reference for further research on the origin and distribution of *C. subtilis*.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## AUTHOR CONTRIBUTIONS

G-WH and X-ZC designed the topic. JR, JT, and XD participated in the sample collection. X-XZ was the first to discover *Coleanthus subtilis*, which inspired us to do research on it. And he assisted in the process of collecting samples. JR analyzed the chloroplast genome data and wrote the manuscript. JT designed the protocol and conducted the experiment. HJ, S-XD, J-XY, and L-LC provided guidance and assistance during the analysis of the data. Also, FM and VW provided valuable comments in writing the article. All authors contributed to this study and approved the final submitted manuscript.

## FUNDING

This work was supported by grants from the National Science & Technology Fundamental Resources Investigation Program of China (Grant Number 2019FY101800) and National Natural Science Foundation of China (Grant Number 31970211).

## ACKNOWLEDGMENTS

We are grateful to National Wild Plant Germplasm Resource Center for kindly help in field investigation and data collection.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.828467/full#supplementary-material>

## REFERENCES

- Abdullah, Henriquez, C. L., Croat, T. B., Poczai, P., and Ahmed, I. (2020a). Mutational dynamics of aroid chloroplast genomes II. *Front. Genet.* 11:610838. doi: 10.3389/fgene.2020.610838
- Abdullah, Henriquez, C. L., Mehmood, F., Carlsen, M. M., Islam, M., Waheed, M. T., et al. (2020b). Complete chloroplast genomes of *Anthurium huixtlense* and *Pothos scandens* (Pothoideae, Araceae): unique inverted repeat expansion and contraction affect rate of evolution. *J. Mol. Evol.* 88, 562–574. doi: 10.1007/s00239-020-09958-w
- Abdullah, Mehmood, F., Shahzadi, I., Waseem, S., Mirza, B., Ahmed, I., et al. (2020c). Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. *Genomics* 112, 581–591. doi: 10.1016/j.ygeno.2019.04.010
- Abdullah, Mehmood, F., Shahzadi, I., Ali, Z., Islam, M., Naeem, M., et al. (2020d). Correlations among oligonucleotide repeats, nucleotide substitutions, and insertion–deletion mutations in chloroplast genomes of plant family Malvaceae. *J. Syst. Evol.* 59, 388–402. doi: 10.1111/jse.12585
- Abdullah., Mehmood, F., Heidari, P., Rahim, A., Ahmed, I., and Poczai, P. (2021a). Pseudogenization of the chloroplast threonine (*trnT-GGU*) gene in the sunflower family (Asteraceae). *Sci Rep* 11:21122. doi: 10.1038/s41598-021-00510-4
- Abdullah, Mehmood, F., Rahim, A., Heidari, P., Ahmed, I., and Poczai, P. (2021b). Comparative plastome analysis of *Blumea*, with implications for genome evolution and phylogeny of Asteroideae. *Ecol. Evol.* 11, 7810–7826. doi: 10.1002/ece3.7614
- Ahmed, I., Biggs, P. J., Matthews, P. J., Collins, L. J., Hendy, M. D., and Lockhart, P. J. (2012). Mutational dynamics of aroid chloroplast genomes. *Genome Biol. Evol.* 4, 1316–1323. doi: 10.1093/gbe/evs110
- Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S., and Thompson, W. F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* 1, 2320–2325. doi: 10.1038/nprot.2006.384
- Alzahrani, D. A., Yaradua, S. S., Albokhari, E. J., and Abba, A. (2020). Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. *BMC Genomics* 21:393. doi: 10.1186/s12864-020-06798-2
- Amiryousefi, A., Hyvonen, J., and Poczai, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34, 3030–3031. doi: 10.1093/bioinformatics/bty220
- Beier, S., Thiel, T., Munch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198
- Brudno, M., Malde, S., Poliakov, A., Do, C. B., Couronne, O., Dubchak, I., et al. (2003). Glocal alignment: finding rearrangements during alignment. *Bioinformatics* 19(Suppl. 1), i54–i62. doi: 10.1093/bioinformatics/btg1005
- Burke, S. V., Grennan, C. P., and Duvall, M. R. (2012). Plastome sequences of two New World bamboos—*Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)—extend phylogenomic understanding of *Bambusoideae*. *Am. J. Bot.* 99, 1951–1961. doi: 10.3732/ajb.1200365
- Catling, P. M. (2009). *Coleanthus Subtilis* (Poaceae), new to northwest territories, and its status in North America. *Rhodora* 111, 109–119. doi: 10.3119/08-8.1
- Chase, M. W., and Hills, H. H. (2019). Silica gel: an ideal material for field preservation of leaf samples for DNA studies. *Taxon* 40, 215–220. doi: 10.2307/1222975
- Chen, X., Zhou, J., Cui, Y., Wang, Y., Duan, B., and Yao, H. (2018). Identification of *Ligularia* herbs using the complete chloroplast genome as a super-barcode. *Front. Pharmacol.* 9:695. doi: 10.3389/fphar.2018.00695
- Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L., et al. (2006). The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* 23, 2175–2190. doi: 10.1093/molbev/msl089
- Cronn, R., Liston, A., Parks, M., Gernandt, D. S., Shen, R., and Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res.* 36:e122. doi: 10.1093/nar/gkn502
- Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17:134. doi: 10.1186/s13059-016-1004-2
- Ding, S. X., Dong, X., Yang, J. X., Guo, C. C., Cao, B. B., Guo, Y., et al. (2021). Complete chloroplast genome of *Clethra fargesii* Franch., an original sympetalous plant from Central China: comparative analysis, adaptive evolution, and phylogenetic relationships. *Forests* 12:441. doi: 10.3390/f12040441
- Downie, S. R., Llanas, E., and KatzDownie, D. S. (1996). Multiple independent losses of the *rpoCl* intron in angiosperm chloroplast DNAs. *Syst. Bot.* 21, 135–151. doi: 10.2307/2419744
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17, 368–376. doi: 10.1007/BF01734359
- Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279. doi: 10.1093/nar/gkh458
- Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., and Ho, S. Y. W. (2019). EasyCodeML: a visual tool for analysis of selection using CodeML. *Ecol. Evol.* 9, 3891–3898. doi: 10.1002/ece3.5015
- Gao, L., Yi, X., Yang, Y. X., Su, Y. J., and Wang, T. (2009). Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC Evol. Biol.* 9:130. doi: 10.1186/1471-2148-9-130
- Gao, X., Zhang, X., Meng, H., Li, J., Zhang, D., and Liu, C. (2018). Comparative chloroplast genomes of *Paris* Sect. *Marmorata*: insights into repeat regions and evolutionary implications. *BMC Genomics* 19(Suppl. 10):878. doi: 10.1186/s12864-018-5281-x
- Gnutikov, A. A., Nosov, N. N., Punina, E. O., Probatova, N. S., and Rodionov, A. V. (2020). On the placement of *Coleanthus subtilis* and the subtribe Coleanthinae within Poaceae by new molecular phylogenetic data. *Phytotaxa* 468, 243–274. doi: 10.11646/phytotaxa.468.3.2
- Greiner, S., Lehwar, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238
- Grewé, F., Guo, W., Gubbels, E. A., Hansen, A. K., and Mower, J. P. (2013). Complete plastid genomes from *Ophioglossum californicum*, *Ptilotum nudum*, and *Equisetum hyemale* reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. *BMC Evol. Biol.* 13:8. doi: 10.1186/1471-2148-13-8
- Grulich, V. (2012). Red List of vascular plants of the Czech Republic: 3rd edition. *Preslia* 84, 631–645.
- Gu, C., Tembrock, L. R., Johnson, N. G., Simmons, M. P., and Wu, Z. (2016). The complete plastid genome of *Lagerstroemia fauriei* and loss of rpl2 Intron from *Lagerstroemia* (Lythraceae). *PLoS One* 11:e0150752. doi: 10.1371/journal.pone.0150752
- Guisinger, M. M., Chumley, T. W., Kuehl, J. V., Boore, J. L., and Jansen, R. K. (2010). Implications of the plastid genome sequence of *typha* (typhaceae, poales) for understanding genome evolution in poaceae. *J. Mol. Evol.* 70, 149–166. doi: 10.1007/s00239-009-9317-3
- Hejny, S. (1969). *Coleanthus subtilis* (Tratt.) seidl in der Tschechoslowakei. *Folia Geobotanica Phytotaxonomica* 4, 345–399. doi: 10.1007/bf02854697
- Hoffmann, M. H., Schneider, J., Hase, P., and Roser, M. (2013). Rapid and recent world-wide diversification of bluegrasses (Poa, Poaceae) and related genera. *PLoS One* 8:e60061. doi: 10.1371/journal.pone.0060061
- Hollingsworth, P. M. (2011). Refining the DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19451–19452. doi: 10.1073/pnas.1116812108
- Huang, Y. Y., Cho, S. T., Haryono, M., and Kuo, C. H. (2017). Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. *PLoS One* 12:e0179055. doi: 10.1371/journal.pone.0179055
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294, 2310–2314. doi: 10.1126/science.1065889
- Jansen, R. K., Raubeson, L. A., Boore, J. L., DePamphilis, C. W., Chumley, T. W., Haberle, R. C., et al. (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. *Mol. Evol.: Producing Biochem. Data, Part B* 395, 348–384. doi: 10.1016/s0076-6879(05)95020-9
- Jian, H. Y., Zhang, Y. H., Yan, H. J., Qiu, X. Q., Wang, Q. G., Li, S. B., et al. (2018). The complete chloroplast genome of a key ancestor of modern roses, *Rosa chinensis* var. *spontanea*, and a comparison with congeneric species. *Molecules* 23:389. doi: 10.3390/molecules23020389



- Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21:241. doi: 10.1186/s13059-020-02154-5
- John, H., Roland, A., Günther, A., Richert, E., Kugler, J., Miekley, B., et al. (2010). *Die Bergwerksteiche der Revierwasserlaufanstalt Freiberg als Lebensraum Einer Einzigartigen Teichbodenvegetation – Gebietsgeschichte und Vegetationsökologie als Basis für Nachhaltigen Naturschutz [Online]*. Available online at: <http://www.dbu.de/OPAC/ab/DBU-AbschlussberichtAZ-24796.pdf> (accessed November 10, 2021).
- Kalyanamoorthy, S., Bui Quang, M., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software Version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Kim, E. S., Bol'sheva, N. L., Samatadze, T. E., Nosov, N. N., Nosova, I. V., Zelenin, A. V., et al. (2009). The unique genome of two-chromosome grasses *Zingeria* and *Colpodium*, its origin, and evolution. *Genetika* 45, 1506–1515. doi: 10.1134/S1022795409110076
- Kim, H. T., Pak, J. H., and Kim, U. S. (2019). The complete chloroplast genome sequence of *Crepidiastrum lanceolatum* (Asteraceae). *Mitochondrial DNA Part B-Resour.* 4, 1404–1405. doi: 10.1080/23802359.2019.1598799
- Kryazhimskiy, S., and Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genet.* 4:e1000304. doi: 10.1371/journal.pgen.1000304
- Kugita, M., Kaneko, A., Yamamoto, Y., Takeya, Y., Matsumoto, T., and Yoshinaga, K. (2003). The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. *Nucleic Acids Res.* 31, 716–721. doi: 10.1093/nar/gkg155
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis Version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Kurchenko, E. I. (2006). Synflorescence of *Coleanthus subtilis* (Poaceae). *Botanicheskii Zhurnal (Moscow & St.-Petersburg)* 91, 200–205.
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633
- Lam-Tung, N., Schmidt, H. A., von Haeseler, A., and Bui Quang, M. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Lee, H. L., Jansen, R. K., Chumley, T. W., and Kim, K. J. (2007). Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol. Biol. Evol.* 24, 1161–1180. doi: 10.1093/molbev/msm036
- Lee, J., Kang, Y., Shin, S. C., Park, H., and Lee, H. (2014). Combined analysis of the chloroplast genome and transcriptome of the Antarctic vascular plant *Deschampsia antarctica* Desv. *PLoS One* 9:e92501. doi: 10.1371/journal.pone.0092501
- Li, D. M., Zhao, C. Y., and Liu, X. F. (2019). Complete chloroplast genome sequences of *Kaempferia Galanga* and *Kaempferia Elegans*: molecular structures and comparative analysis. *Molecules* 24:474. doi: 10.3390/molecules24030474
- Li, X., Zuo, Y., Zhu, X., Liao, S., and Ma, J. (2019). Complete chloroplast genomes and comparative analysis of sequences evolution among seven *Aristolochia* (Aristolochiaceae) medicinal species. *Int. J. Mol. Sci.* 20:1045. doi: 10.3390/ijms20051045
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Liu, L., Wang, Y., He, P., Li, P., Lee, J., Soltis, D. E., et al. (2018). Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. *BMC Genomics* 19:235. doi: 10.1186/s12864-018-4633-x
- Liu, Q., Li, X., Li, M., Xu, W., Schwarzscher, T., and Heslop-Harrison, J. S. (2020). Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. *BMC Plant Biol.* 20:406. doi: 10.1186/s12870-020-02621-y
- Liu, X. Y., Li, Y., Ji, K. K., Zhu, J., Ling, P., Zhou, T., et al. (2020). Genome-wide codon usage pattern analysis reveals the correlation between codon usage bias and gene expression in *Cuscuta australis*. *Genomics* 112, 2695–2702. doi: 10.1016/j.ygeno.2020.03.002
- Lyu, X., and Liu, Y. (2020). Nonoptimal codon usage is critical for protein structure and function of the master general amino acid control regulator CPC-1. *mBio* 11, e2605–e2620. doi: 10.1128/mBio.02605-20
- Mardanov, A. V., Ravin, N. V., Kuznetsov, B. B., Samigullin, T. H., Antonov, A. S., Kolganova, T. V., et al. (2008). Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *J. Mol. Evol.* 66, 555–564. doi: 10.1007/s00239-008-9091-7
- Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4623–4628. doi: 10.1073/pnas.0907801107
- Mugal, C. F., Wolf, J. B., and Kaj, I. (2014). Why time matters: codon evolution and the temporal dynamics of dN/dS. *Mol. Biol. Evol.* 31, 212–231. doi: 10.1093/molbev/mst192
- Muniao, J. N., Dong, X., Yang, J. X., Mbandi, E. M., Wanga, V. O., Oulo, M. A., et al. (2020). Complete chloroplast genomes of *Chlorophytum comosum* and *Chlorophytum gallabatense*: genome structures, comparative and phylogenetic analysis. *Plants (Basel)* 9:296. doi: 10.3390/plants9030296
- Necsulea, A., and Lobry, J. R. (2007). A new method for assessing the effect of replication on DNA base composition asymmetry. *Mol. Biol. Evol.* 24, 2169–2179. doi: 10.1093/molbev/msm148
- Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., et al. (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia Polymorpha* chloroplast DNA. *Nature* 322, 572–574. doi: 10.1038/322572a0
- Oldenburg, D. J., and Bendich, A. J. (2016). The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication. *Curr. Genet.* 62, 431–442. doi: 10.1007/s00294-015-0548-0
- Oliveira, E. J., Padua, J. G., Zucchi, M. I., Vencovsky, R., and Vieira, M. L. C. (2006). Origin, evolution and genome distribution of microsatellites. *Genet. Mol. Biol.* 29, 294–307. doi: 10.1590/S1415-47572006000200018
- Parks, M., Cronn, R., and Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 7:84. doi: 10.1186/1741-7007-7-84
- Piot, A., Hackel, J., Christin, P. A., and Besnard, G. (2018). One-third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta* 247, 255–266. doi: 10.1007/s00425-017-2781-x
- Qu, X. J., Moore, M. J., Li, D. Z., and Yi, T. S. (2019). PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15:50. doi: 10.1186/s13007-019-0435-7
- Richert, E., Achtziger, R., Dajdok, Z., Günther, A., Heilmeyer, H., Hubner, A., et al. (2016). Rare wetland grass *Coleanthus subtilis* in Central and Western Europe - current distribution, habitat types, and threats. *Acta Societatis Botanicorum Poloniae* 85:3511. doi: 10.5586/asbp.3511
- Richert, E., Achtziger, R., Günther, A., Hübner, A., Olias, M., and John, H. (2014). *Das Scheidenblütgras (Coleanthus subtilis) – Vorkommen, Ökologie und Gewässermanagement*. Dresden: Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie (Hrsg.), 52S.
- Rodionov, A. V., Kim, E. S., Nosov, N. N., Raiko, M. P., Machs, E. M., and Punina, E. O. (2008). Molecular phylogenetic study of the genus *Colpodium* sensu lato (Poaceae: poae). *Ecol. Genet.* 6, 34–46. doi: 10.17816/ecogen6434-46
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., and Burleigh, J. G. (2014). From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* 14:23. doi: 10.1186/1471-2148-14-23
- Saarela, J. M., Burke, S. V., Wysocki, W. P., Barrett, M. D., Clark, L. G., Craine, J. M., et al. (2018). A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. *PeerJ* 6:e4299. doi: 10.7717/peerj.4299
- Somarathne, Y., Guan, D. L., Wang, W. Q., Zhao, L., and Xu, S. Q. (2019). The complete chloroplast genomes of two *Lespedeza* species: insights into codon

- usage bias, RNA editing sites, and phylogenetic relationships in desmodieae (Fabaceae: papilionoideae). *Plants (Basel)* 9:51. doi: 10.3390/plants9010051
- Sonah, H., Deshmukh, R. K., Sharma, A., Singh, V. P., Gupta, D. K., Gacche, R. N., et al. (2011). Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. *PLoS One* 6:e21298. doi: 10.1371/journal.pone.0021298
- Soreng, R. J., Peterson, P. M., Davidse, G., Judziewicz, E. J., Zuloaga, F. O., Filgueiras, T. S., et al. (2003). *Catalogue of New World Grasses (Poaceae): IV. Subfamily Pooideae*. Vol. 48. Washington, DC: National Museum of Natural History, Dept of Botany, 1–730.
- Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Teisher, J. K., Clark, L. G., et al. (2017). A worldwide phylogenetic classification of the Poaceae (Gramineae) II: an update and a comparison of two 2015 classifications. *J. Syst. Evol.* 55, 259–290. doi: 10.1111/jse.12262
- Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Zuloaga, F. O., Judziewicz, E. J., et al. (2015). A worldwide phylogenetic classification of the Poaceae (Gramineae). *J. Syst. Evol.* 53, 117–137. doi: 10.1111/jse.12150
- Souza, U. J. B. D., Vitorino, L. C., Bessa, L. A., and Silva, F. G. (2020). The complete plastid genome of *Artocarpus camansi*: a high degree of conservation of the plastome structure in the family moraceae. *Forests* 11:1179. doi: 10.3390/f11111179
- Sun, Y. X., Moore, M. J., Meng, A. P., Soltis, P. S., Soltis, D. E., Li, J. Q., et al. (2013). Complete plastid genome sequencing of Trochodendraceae reveals a significant expansion of the inverted repeat and suggests a Paleogene divergence between the two extant species. *PLoS One* 8:e60429. doi: 10.1371/journal.pone.0060429
- Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164
- Tangphatsornruang, S., Sangsrakru, D., Chanprasert, J., Uthaisaisanwong, P., Yoocha, T., Jomchai, N., et al. (2010). The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. *DNA Res.* 17, 11–22. doi: 10.1093/dnares/dsp025
- Taran, G. S. (1994). Taran G.S. Flood-plain ephemeratum of middle Ob' – a new class for Siberia, Isoëto-Nanojuncetea Br.-Bl. et Tx. 1943 on the northern border of expansion. *Siberian J. Ecol.* 1, 578–582.
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391
- Tkach, N., Schneider, J., Doering, E., Woelk, A., Hochbach, A., Nissen, J., et al. (2020). Phylogenetic lineages and the role of hybridization as driving force of evolution in grass supertribe Pooideae. *Taxon* 69, 234–277. doi: 10.1002/tax.12204
- Tzelev, N. N. (1976). On the origin of the arctic grasses (Poaceae). *Botanicheskij Zhurnal (Moscow & Leningrad)* 61, 1354–1363.
- Wallace, R. S., and Cota, J. H. (1996). An intron loss in the chloroplast gene *rpoC1* supports a monophyletic origin for the subfamily Cactoideae of the Cactaceae. *Curr. Genet.* 29, 275–281. doi: 10.1007/BF02221558
- Wang, L., Xing, H., Yuan, Y., Wang, X., Saeed, M., Tao, J., et al. (2018). Genome-wide analysis of codon usage bias in four sequenced cotton species. *PLoS One* 13:e0194372. doi: 10.1371/journal.pone.0194372
- Wang, W., Yu, H., Wang, J., Lei, W., Gao, J., Qiu, X., et al. (2017). The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae). *Int. J. Mol. Sci.* 18:2288. doi: 10.3390/ijms18112288
- Wanga, V. O., Dong, X., Oulo, M. A., Mkala, E. M., Yang, J. X., Onjalalaina, G. E., et al. (2021). Complete chloroplast genomes of *Acanthochlamys bracteata* (China) and *Xerophyta* (Africa) (Velloziaceae): comparative genomics and phylogenomic placement. *Front. Plant Sci.* 12:691833. doi: 10.3389/fpls.2021.691833
- Wariss, H. M., Yi, T. S., Wang, H., and Zhang, R. (2018). The chloroplast genome of a rare and endangered species *Salweenia bouffordiana* (Leguminosae) in China. *Conserv. Genet. Resour.* 10, 405–407. doi: 10.1007/s12686-017-0836-8
- Wheeler, G. L., Dorman, H. E., Buchanan, A., Challagundla, L., and Wallace, L. E. (2014). A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Appl. Plant Sci.* 2:1400059. doi: 10.3732/apps.1400059
- Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383
- Woike, S. (1969). Beitrag zum Vorkommen von *Coleanthus subtilis* (Tratt.) Seidl (feines Scheidenblütgras) in europa. *Folia Geobotanica Phytotaxonomica* 4, 401–413. doi: 10.1007/bf02854698
- Wu, L., Nie, L., Xu, Z., Li, P., Wang, Y., He, C., et al. (2020). Comparative and phylogenetic analysis of the complete chloroplast genomes of three *Paeonia* section moutan species (Paeoniaceae). *Front. Genet.* 11:980. doi: 10.3389/fgene.2020.00980
- Xie, D. F., Yu, Y., Deng, Y. Q., Li, J., Liu, H. Y., Zhou, S. D., et al. (2018). Comparative analysis of the chloroplast genomes of the chinese endemic genus *Urophyssa* and their contribution to chloroplast phylogeny and adaptive evolution. *Int. J. Mol. Sci.* 19:1847. doi: 10.3390/ijms19071847
- Xu, C., Dong, W., Li, W., Lu, Y., Xie, X., Jin, X., et al. (2017). Comparative analysis of six *Lagerstroemia* complete chloroplast genomes. *Front. Plant Sci.* 8:15. doi: 10.3389/fpls.2017.00015
- Xu, Y. C., and Guo, Y. L. (2020). Less is more, natural loss-of-function mutation is a strategy for adaptation. *Plant Commun.* 1:100103. doi: 10.1016/j.xplc.2020.100103
- Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., et al. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS One* 5:e12762. doi: 10.1371/journal.pone.0012762
- Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., and Zhao, G. (2016). Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front. Plant Sci.* 7:959. doi: 10.3389/fpls.2016.00959
- Yang, Z., Wang, G., Ma, Q., Ma, W., Liang, L., and Zhao, T. (2019). The complete chloroplast genomes of three Betulaceae species: implications for molecular phylogeny and historical biogeography. *PeerJ* 7:e6320. doi: 10.7717/peerj.6320
- Yang, Z., Wong, W. S., and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118. doi: 10.1093/molbev/msi097
- Yu, X., Wang, W., Yang, H., Zhang, X., Wang, D., and Tian, X. (2021). Transcriptome and comparative chloroplast genome analysis of *Vincetoxicum versicolor*: insights into molecular evolution and phylogenetic implication. *Front. Genet.* 12:602528. doi: 10.3389/fgene.2021.602528
- Yurova, E. A. (2001). Floristic findings in Novgorodskaya Oblast'. *Botanicheskij Zhurnal (Moscow & St.-Petersburg)* 86, 154–155.
- Zhang, D., Gao, F. L., Jakovic, I., Zou, H., Zhang, J., Li, W. X., et al. (2020). PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 20, 348–355. doi: 10.1111/1755-0998.13096
- Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J. M., et al. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 495, 111–115. doi: 10.1038/nature11833

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ren, Tian, Jiang, Zhu, Mutie, Wanga, Ding, Yang, Dong, Chen, Cai and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.