

Complete Chloroplast Genome Sequence of a Major Warm-Season Turfgrass Species, Centipedegrass [*Eremochloa ophiuroides* (Munro) Hack.]: Genome Characterization, Comparative and Phylogenetic Analysis

Junqin Zong

Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden,
Mem. Sun Yat-Sen

Haoran Wang

Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden,
Mem. Sun Yat-Sen

Jingjing Wang

Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden,
Mem. Sun Yat-Sen

Hailin Guo

Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden,
Mem. Sun Yat-Sen

Jingbo Chen

Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden,
Mem. Sun Yat-Sen

Dandan Li

Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden,
Mem. Sun Yat-Sen

Ling Li

Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden,
Mem. Sun Yat-Sen

Jianxiu Liu

Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden,
Mem. Sun Yat-Sen

Jianjian Li (✉ lijianjian2013@yahoo.com)

Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden,
Mem. Sun Yat-Sen

Research Article

Keywords: Eremochloa ophiuroides, chloroplast, genome structure, gene content, phylogeny

Posted Date: February 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-199791/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Chloroplast (cp) genome sequence data could provide valuable information for molecular taxonomy and phylogenetic reconstruction among plant species and individuals. However, as one of the most important warm-season turfgrasses widely used in USA and China, cp genome characteristics and phylogenetic position of centipedegrass (*Eremochloa ophiuroides*) were poorly understood.

Results: In this study, we determined the complete chloroplast genome sequences of *E. ophiuroides* using high-throughput Illumina sequencing technology. The circle pseudomolecule for *E. ophiuroides* cp genome is 139,107 bp in length, and has a typical quadripartite structure consisting of a pair of inverted repeat (IR) regions of 22,230 bp each separated by a large single copy (LSC) region of 82,081 bp and a small single copy (SSC) region of 12,566 bp. The nucleotide composition of *E. ophiuroides* cp genome is asymmetric with an overall A + T content of 61.60%. It encodes a total of 131 gene species, composed of 20 duplicated genes within the IR regions and 111 unique genes including 77 protein-coding genes (PCGs), 30 transfer RNA (tRNA) genes and four ribosome RNA (rRNA) genes. Analysis of the repetitive sequences revealed that *E. ophiuroides* cp genome contains 51 tandem repeats including 29 forward, 20 palindromic and 2 reverse repeats, and 197 simple sequence repeats (SSRs) which were mainly composed of adenine (A) and thymine (T) bases. Comparison of the *E. ophiuroides* complete cp genome with the genomes of other seven Gramineae species showed a high degree of collinearity among Gramineae plants. Phylogenetic analysis showed that *E. ophiuroides* was closely related to *E. ciliaris* and *E. eriopoda*, and was placed in a clade with the two *Eremochloa* species and *Mnesithea helferi* within the subtribe Rottboelliinae, which clarified evolutionary status of *E. ophiuroides* in tribe Andropogoneae and also authenticated the current taxonomy of the tribe Andropogoneae.

Conclusions: The present study provides the complete structure of the *E. ophiuroides* cp genome, and preliminarily ascertains the phylogenetic position of *E. ophiuroides* in tribe Andropogonodae. This will be of value to grass taxa identification, phylogenetic resolution, population structure and biodiversity, novel gene discovery and functional genomic studies for the genus *Eremochloa*.

Background

Centipedegrass [*Eremochloa ophiuroides* (Munro) Hack.] is one of the most important perennial warm-season (C_4) grass species that originated in China, and is mainly distributed in East Asia, South-east Asia, and the eastern and southern United States [1–3]. *E. ophiuroides* is widely used as a turfgrass in tropical and subtropical regions, and also as a pasture grass in some countries and regions of East Asia [4–5]. Now it has been becoming an increasingly popular turfgrass for its excellent adaptation to infertile soils and a range of climatic conditions, as well as for its resistance to the main biotic and abiotic stresses [2, 6]. It also has great potential for commercial applications as a low-maintenance turfgrass for its exceptional advantages of low management and fertilization requirements [2, 3]. Nevertheless, due to the serious reduction of wild grassland and wasteland areas caused by rapid urbanization process in China, as well with the habitat destruction of the native-grasses resulted from artificial farming and overgrazing,

the natural populations of *E. ophiuroides* have been experiencing a sharp decline. So it is urgent to establish scientific strategies to protect and conserve the resources of *E. ophiuroides* in its main distribution areas.

Chloroplast (cp) is the key organelle responsible for photosynthesis and carbon fixation in green plants, and is involved in the biosynthesis of amino acids, fatty acids, hormones, vitamins, nucleotides, pigments and other secondary metabolites [7, 8]. It contains both highly conserved genes fundamental to plant life and more variable regions which are informative over broad time scales. Therefore, cp genome sequence data can provide valuable information for molecular taxonomy and phylogenetic reconstruction among plant species and individuals [9, 10], which could contribute greatly to plant breeding and conservation strategies. In addition, cp DNA has the characteristics of non-recombinant nature, low mutation rates, and uniparental inheritance, which makes it significant in giving insights into plant evolution and developing applications for biotechnological breeding [11, 12]. In recent years, with the rapid development of next-generation sequencing (NGS) technology, it is now more convenient and relatively inexpensive to obtain cp genome sequences and implement whole genome-based phylogenomics [13]. In contrast to previous studies done with a single or a few cp loci-based approaches, using the complete cp genome sequence now provides a unique opportunity to investigate related species evolution based on whole-genome comparison [14, 15].

As an important turfgrass, turf quality of *E. ophiuroides*, to a great extent, depends on its biomass, color and green color retention that are highly correlated with photosynthetic efficiency. Like most of other turf grass species, *E. ophiuroides* prefers to grow in open sunny places. Therefore, it is hoped that the study of photosynthesis-related chloroplast genes could provide basis for genetic breeding of the turf grass [16]. Analyzing and characterizing the cp genome of a turf grass would provide essential information to improve the turf quality and facilitate the development of a plastid transformation system in this turf plant [17]. However, despite being one of the most important warm-season turfgrasses being introduced into the USA by Frank Meyer one century ago [18], the research on *E. ophiuroides* cp genome is lagging behind. Although a set of cp genome sequence data of *E. ophiuroides* was submitted to NCBI GeneBank in 2017, there is no relevant report to interpret it in detail, which has hindered our understanding and progress of *E. ophiuroides* evolution, species identification, germplasm conservation, genetic engineering, and other related researches.

In the present study, we sequenced the *E. ophiuroides* cp genome using Illumina technology, assembled the complete cp genome sequence of *E. ophiuroides*, and performed detailed phylogenetic analyses on the basis of complete cp genome sequence information. We also analyzed the fully assembled cp genome of *E. ophiuroides* and compared it to seven related species of Gramineae. The main purposes of this study were to investigate the complete structure of the *E. ophiuroides* cp genome, to explore the phylogenetic position of *E. ophiuroides* in tribe Andropogonodae, and also to provide basic data for further molecular studies related to grass taxa identification, phylogenetic resolution, population structure and biodiversity, novel gene discovery and functional genomic studies in the genus *Eremochloa*.

Results

Genome assembly and structure analysis

A total of 19,101,863 clean reads (approximately 5.73 Gb) were obtained from the *E. ophiuroides* leaf library. After performing de novo and reference-guided assembly with minor modifications, these reads finally were integrated into a complete circular pseudomolecule that was 139,107 bp in length (GeneBank accession: MT806102). Thus, the sequencing depth of the *E. ophiuroides* cp genome was expected to be more than $41,000 \times$ (Fig. 1, Table 1). Since 738,323 reads were mapped to the assembled cp genome, the average coverage reached to more than $1500 \times$ of the *E. ophiuroides*.

The cp genome of *E. ophiuroides* exhibited a typical quadripartite structure, consisting of a pair of inverted repeat (IR) regions of 22,230 bp each with 56.00% AT, separated by a large single-copy (LSC) region of 82,081 bp with 63.76% AT and a small single-copy (SSC) region of 12,566 bp with 67.32% AT (Fig. 1). The global AT content in the *E. ophiuroides* cp genome was 61.60% (Table 1).

Gene annotation

A total of 131 genes were annotated in the sequenced *E. ophiuroides* cp genome, of which 20 genes are duplicated in the IR regions and 111 are unique, including four rRNA genes, 30 tRNA genes, 77 protein-coding genes (Fig. 1, Table 2). Most of the unique genes contained no introns, but one intron was found in each of six tRNA genes and eight protein-coding genes, and two introns were found in each of two protein coding genes. In the gene function analysis, the 111 unique genes were classified into four categories, including genes associated with photosynthesis (44 genes), self-replication (59 genes), other functions (5 genes), and genes of unknown function (3 genes) (Table 2).

Codon preference analysis

There were 3 termination codons and 63 codons encoding 20 amino acids in the cp protein-coding genes of the *E. ophiuroides* (Fig. 2, Table S1), and most of the amino acid codons, except tryptophan, had codon preferences. A total of 31 codon preferences were identified from all the codons, which involved 18 amino acids and one stop codon. According to the synonymous codon preference partitions, 70.97%, 12.90% and 16.13% of all preferred codons exhibited high (RSCU value > 1.3), moderate ($1.2 \leq \text{RSCU value} \leq 1.3$), and low preferences ($1.0 < \text{RSCU value} < 1.2$), respectively.

Repeat structure and SSR analysis

Repeated analysis revealed 29 forward repeats, 20 palindromic repeats, and two reverse repeats in the *E. ophiuroides* cp genome (Table S2). The forward repeat units were 30–242 bp long, and almost all the forward repeats were located at LSC regions except for four located in IR regions and one in SSC region.

Similar to the forward repeats, the majority of palindromic repeat units were 30–242 bp in length and distributed in LSC regions, with one of them was 22,230 bp long as an exception. Alternatively, for the reverse repeats, two in total were less than 35 bp in length and were detected in LSC regions.

The presence of microsatellites was also detected in the cp genome of the *E. ophiuroides* (Table S3). A total of 197 simple sequence repeats (SSRs) loci were identified, including 191 perfect SSRs and six complex SSRs. Among these perfect SSRs, 129 (67.54%) SSRs were mononucleotide repeats, and the rest (in descending order of abundance) were tri- (46, 24.08%), tetra- (10, 5.24%), di- (5, 2.62%), and penta- (1, 0.52%) nucleotide repeats. Majority of mononucleotides composed of poly A (polyadenine) (43.41%) and poly T (polythymine) (50.39%) repeats, whereas poly G (polyguanine) and poly C (polycytosine) repeats were rather rare (6.20%). For the dinucleotides, four out of five were AT/TA repeat motifs, and the rest one was TC repeat. Among the trinucleotide and tetranucleotide repeat motifs, A/T contents in trimers and tetramers were 71.01% and 55.00%, respectively. One pentanucleotide was ATAAA repeat.

IR contraction and expansion and genome collinearity

The exact IR boundary positions and their adjacent genes of the *E. ophiuroides* and the other eight species from families Gramineae and Cruciferae were compared (Fig. 3). In the cp genome of the three *Eremochloa* species, the IR boundary positions and their adjacent genes were exactly the same. IRa/SSC and IRb/SSC junctions were found within the gene *ndhH* and the gene *ndhF* respectively, and correspondingly the *ndhH* pseudogene (1 bp), the *ndhF* pseudogene (29 bp) was detected at the IRa/SSC boundary and the IRb/SSC border, whereas no pseudogene was observed at the IRa/LSC and IRb/LSC boundaries. As for *Sorghum* and *Zea*, they were found to have exactly the same IR boundary position and the adjacent genes, and to have almost unanimous IR situations with *Eremochloa* only except for the positions of the genes *rpl22* (57bp to IRb) and *psbA* (88bp to IRa) in LSC. *Setaria italic* had IR boundary positions similar to that of *Sorghum* or *Zea*. IR boundary positions and the adjacent genes of *Oryza sativa* and *Brachypodium distachyon* were generally consistent with that of *Eremochloa* plants, even though slight differences were observed, such as the position divergence of the gene *ndhF* adjacent to IRb/SSC junction, the pseudogene *ndhH* length variations detected in IRa regions of the two species. Unlike the Gramineae species, IRb/LSC and IRa/SSC junctions of *Arabidopsis thaliana* were within the gene *rsp19* and the gene *ycf1* respectively, while IRb/SSC junction was within both the *ycf1* gene and the *ndhF* gene.

The mauve alignment for the eight species revealed that all the cp genomes formed locally collinear blocks (LCBs). In particular, the gene order of the three *Eremochloa* cp genomes was highly conserved compared with that of other plant species (Fig. 4). *Eremochloa* and *Sorghum* had the highest chloroplast genome homologies, while the order of cp gene loci was highly consistent among the other Gramineae plant genomes. This demonstrates that the chloroplast genome has a high homology among Gramineae plants.

Phylogenetic relationship

Phylogenetic relationships of species in the tribe Andropogoneae and taxonomic statuses of *E. ophiuroides* and other species in the same tribe were systematically classified through Maximal likelihood (ML) analysis of the newly sequenced and published complete cp sequences. 49 published or available complete cp genome sequences and a newly sequenced *E. ophiuroides* cp sequence were combined in this study. Thus, we reconstructed a phylogenetic tree of tribe Andropogoneae using a total of 50 complete cp genome sequences which were selected from 46 different species in Andropogoneae (three species with two cp sequences) and one species (*Arundinella deppeana*) in Arundinelleae used as an outgroup (Table S4). RAxML analysis produced a phylogenetic tree which fully supported *E. ophiuroides* to be closely related with *Eremochloa ciliaris* and *Eremochloa eriopoda* with 100% bootstrap values, and the three *Eremochloa* species, together with *Mnesithea helferi*, form one monophyletic group corresponding to subtribe Rottboelliinae (Fig. 5).

It is noteworthy that a larger number of species traditionally classified as the same subtribe do not form a group. From the phylogenetic analysis, only four monophyletic groups, corresponding to subtribe Saccjaromae, Sorghinae, Andropogoninae and Rottboelliinae, can be retrieved from the tree, while quite a few non-monophyly be formed. A typical instance is the placement of *Germainia capitata* (Germainiinae) as sister to *Pogonatherum paniceum* (Incertae sedis) with the same branch length. Similar cases can be found for the placement of *Dimeria ornithopoda* (Dimeriinae) as sister of *Eulaliopsis binata* (Saccharinae), and the placement of *Rottboellia cochinchinensis* (Rottboelliinae) as sister of *Coix lacryma-jobi* (Coicinae). In addition, *Heteropogon triticeus* and *Cymbopogon flexuosus*, two species in Anthistiriinae, were clustered to a sister clade of Andropogoninae; *Kerriochloa siamensis*, one species of Ischaeminae, was constrained as sister to Incertae sedis.

Discussion

Genome size and gene identification

The size of the *E. ophiuroides* cp genome was found to be 139,107 bp, similar to those cp genomes in Panicoideae subfamily, which range from 138 Kb in *Setaria viridis* [19] to 141 Kb in *Saccharum officinarum* [20], but larger than those of other sequenced cp genomes in Chloridoideae, Pooideae, and Oryzoideae subfamilies with not more than 137 Kb in length (Table S5). Since the average size of publicly available Poaceae cp genomes is 137,091 bp [21], *E. ophiuroides* is of average size within Panicoideae and of large size within Poaceae (Gramineae). The cp DNA of *E. ophiuroides*, like that of most angiosperms, is circular with a typical quadripartite structure containing a pair of IRs separated by LSC and SSC regions. The overall AT content of the *E. ophiuroides* cp genome was 61.6%, which is similar to that of most Gramineae plants (~61%, Table S5).

The gene and intron contents in the *E. ophiuroides* cp DNA are basically identical to those of rice [22, 23], wheat [24], maize [25], sorghum [26] and other grasses [21, 27–29], with 77 protein-coding genes, 30 tRNA

genes and four rRNA genes. Among the 111 unique genes, 14 contain one intron (six tRNA and eight protein-coding genes) and two (*rps12* and *ycf3*) possess two introns. For all identified genes, 59 fragments are related to self-replication and 44 genes are associated with photosynthesis. Of the 44 photosynthesis related genes, five genes encode photosystem I components (*psaA, B, C, I, J*), 15 genes are related to photosystem II, and six genes (*atpA, B, E, F, H, I*) are responsible for ATP synthase and the other 18 genes encode electron transport chain components. A similar pattern of protein-coding genes is also present in *Oryza sativa* [30], *Oryza glaberrima* [31] and *Oryza minuta* [23].

Repeat sequence

The nucleotide sequences of most organism genomes contain many different types of repetitive sequences, such as short tandem repeats, interspersed repeats or spaced repeats. These repeat elements are either dispersed throughout the genome or within a short region of the genome [32]. The mismatching on slip chains and inappropriate recombination of repetitive sequences may lead to the occurrence of sequence variation and DNA rearrangement [33, 34]. Interspersed repetitive sequences (IRS) are a kind of repeats interspersed in genome DNAs and are potential resource to revealing gene rearrangements and losses during evolution [35, 36]. It usually includes forward, palindromic, reverse and complement repeats. In the present study, many forward and palindromic repeats, and a few reverse repeats were detected in *E. ophiuroides* cp genome sequences, and most of them were distributed in LSC regions of the genome. Similar findings were also reported in other plant species, such as *Swertia mussotii* [37], *Oryza minuta* [23]. This reflects the common characteristics of the IRSs in most of plant cp genomes.

SSRs, also called microsatellites, are known to be more informative and are very abundant and evenly distributed in angiosperm plastomes [38]. Because of their abundance, high rate of polymorphism, ubiquitous distribution throughout the genome, and high extent of allelic diversity, SSRs have been extensively used as versatile DNA-based markers in plant genetic and genomic research [39]. The motif type, length and abundance of SSRs are the main characteristics of microsatellites [40]. Besides complex SSRs, five types of perfect SSRs (mono-/di-/tri-/tetra-/penta- nucleotide repeats) were detected in the *E. ophiuroides* cp genome sequences. The most abundant SSR motif was mononucleotide repeats followed by trinucleotide and tetranucleotide repeats in detected SSRs. This result is not completely consistent with other findings that showed mono- and di-nucleotides are the most frequent SSR types in plant cp genomes [41–43], but is consistent with the report in Lythraceae [44] and *Magnolia polytepala* [12], and is also in accord with the finding of SSR mining from the *E. ophiuroides* RNA-seq data although mononucleotide repeat was omitted in that study [45]. Whether mononucleotide SSRs or polynucleotide SSRs detected in the present study, most of them were rich in A/T content. This is consistent with the existing chloroplast SSR reports [46–48].

IR contraction and expansion

IRs is prominent feature of most angiosperm cp genomes. Expansion and contraction of IR region boundaries is the main reason for size variations in the cp genome and plays an important role in species evolution [49]. In the present study, a detailed comparison on four junctions (Fig. 3), i.e., JLA (junction line between LSC and IRa), JLB (junction line between LSC and IRb), JSA (junction line between SSC and IRa) and JSB (junction line between SSC and IRb), between the two IRs (IRa and IRb) and the two single-copy regions (LSC and SSC) was performed among *E. eriopoda*, *E. ciliaris*, *S. bicolor*, *Z. mays*, *S. italica*, *O. sativa*, *B. distachyon* with regard to *E. ophiuroides* by carefully analyzing the exact IR border positions and adjacent genes. The IR region of *E. ophiuroides* was 22,230 bp in length, which was in medium length of the nine compared species from 20,804 bp to 22,783bp. This implies that some IR expansion and contraction may occur in the *E. ophiuroides* cp genome. JLA is between *rps19* and *rpl22*, and JLB is located between *rps19* and *psbA* in all eight Gramineae species. Both of the distances between *rps19* and JLA, between *rps19* and JLB are 35 bp in all three *Eremochloa* species, *S. bicolor* and *Z. mays*, which are shorter than that in other three Gramineae species; the distance between *rpl22* and JLA in three *Eremochloa* species is shorter than that in *S. bicolor* and *Z. mays*, but is longer than that in the other species, while the distance between *psbA* and JLB in three *Eremochloa* species is longer than that in the other Gramineae species. The *ndhF* gene traverses the SSC and IRa regions, with 29 bp located in the IRa region for all the C₄ plants including three *Eremochloa* species, *S. bicolor*, *Z. mays* and *S. italic*, but it is located in the SSC region for C₃ plants of *O. sativa* and *B. distachyon* revealed in the present study, or of *O. minuta* reported by Asaf et. (2017)[23]. The *ndhH* gene traverses the SSC and IRb regions, with approximately 1,181 bp located in the IR region and only 1 bp in the IRb region for all species except for *O. sativa* and *B. distachyon*. This is accord with most reported findings in Gramineae plants [23]. This hints that variation in JSA border caused by IR expansion or contraction might result in the difference between C₃ and C₄ plant cp genomes. Our results also demonstrated that size variation of cp genomes resulted from IR contraction and expansion is a common feature during evolution of Gramineae plants, although structural organization and gene order of Gramineae cp genomes are highly conserved [50].

Phylogenetic analysis

The tribe Andropogoneae includes over 1,200 species in ca. 90 genera, and is a primary component of grasslands and savannahs that dominate tropical and subtropical regions throughout the world [51, 52]. Recently, a number of phylogenetic and evolutionary studies have been implemented for the tribe Andropogonodae using complete chloroplast genomes [52–55]. Although *E. ophiuroides* is an important member in genera *Eremochloa* of the tribe Andropogoneae, it has not been included in these studies, which restricts illuminating its evolutionary relationships to other Andropogoneae species. Our molecular phylogenetic tree based on sequences of complete cp genomes revealed that *E. ophiuroides* was closely related to *E. ciliaris* and *E. eriopoda*, and their placement in a clade with *Mnesithea helferi* is highly supported with bootstrap values of 100% within the subtribe Rottboelliinae (Fig. 5). This is congruent with the traditional morphology-based taxa of Rottboelliinae, indicating that the classification of subtribe Rottboelliinae is generally reasonable.

In addition, from our results, the Rottboelliinae, Saccjaromae, Sorghinae and Andropogoninae are typically monophyletic groups, which reflect the agreement between molecular phylogeny and traditional morphology-based taxonomy. However, some non-monophylies of subtribes were recognized in the current molecular phylogeny. In the present study, *Germainia capitata* (Germainiinae) was placed as sister to *Pogonatherum paniceum* (Incertae sedis), *Dimeria ornithopoda* (Dimeriinae) as sister to *Eulaliopsis binata* (Saccharinae), and *Rottboellia cochinchinensis* (Rottboelliinae) as sister to *Coix lacryma-jobi* (Coicinae), which are congruent with previous results for these species [52–55]. Another typical non-monophyletic area in the tree is the placement of *Heteropogon triticeus* and *Cymbopogon flexuosus* (two species in Anthistiriinae) in a clade with Andropogoninae species, and the similar result has actually been reported [52]. However, it is worth mentioning that *Sorghastrum nutans* and *Eulalia aurea* were not clustered as sister clades in the current study, which is incongruent with previously reported results [53–55]. This is mainly due to the fact that more extensive species (50 complete cp genome data of 47 different species) in the tribe Andropogoneae were used for phylogenetic analysis in the present study. Considering the same monophyletic clades clustered between *Rottboellia cochinchinensis* and *Coix lacryma-jobi*, *Germainia capitata* and *Pogonatherum paniceum*, *Dimeria ornithopoda* and *Eulaliopsis binata*, and the different monophyletic clades formed from *Eulalia aurea* and *Eulaliopsis binata* displayed in this study, combined with previously reported phylogenetic relationships between these species [53–55], future more sampling with better balancing of ingroup Rottboelliinae, Coicinae, Germainiinae, Incertae sedis, Dimeriinae, Saccharinae should be considered so as to better address questions of subtribal monophylies.

Conclusion

We present the complete cp genome sequence of *E. ophiuroides* (139,107 bp) in this study. *E. ophiuroides* cp genome possesses circular and quadripartite structure which is well conserved similar to previously reported cp genomes from Gramineae family. A total of 131 genes were annotated in the sequenced *E. ophiuroides* cp genome, of which 44 genes are associated with photosynthesis. Most of *E. ophiuroides* codons encoding amino acids have codon preferences. The location and distribution of repeat sequences was detected, and around 197 SSR loci and 51 repeat sequences were identified in *E. ophiuroides* cp genome. Comparative genomic analysis revealed that *E. ophiuroides* has a high level of collinearity with the other Gramineae cp genomes. Phylogenetic analyses showed that *E. ophiuroides* is most closely related to *E. ciliaris* and *E. eriopoda*, and the three *Eremochloa* species together with *Mnesithea helferi* were placed to a monophyletic group corresponding to subtribe Rottboelliinae, which is completely accord with the traditional morphology-based taxa of Rottboelliinae in tribe Andropogoneae. The cp genome information of *E. ophiuroides* is a useful genetic resource that would be utilized on the conservation genetics, species identification, taxonomic clarification, phylogenetic reconstruction, and molecular breeding in the further studies on *Eremochloa*.

Methods

Plant material, DNA extraction and sequencing

Fresh leaves were sampled from *E. ophiuroides* accession E039, which was collected from Lushan, Jiangxi province and now is deposited in the nursery of the Institute of Botany, Jiangsu Province and Chinese Academy of Sciences. Total genomic DNA was extracted using the EZgeneTM SuperFast Plant Leaves DNA Kit (Biomiga, San Diego, CA, USA) following the manufacturer's protocol. Quality and integrity of the DNA were checked and determined using spectrophotometry and agarose gel electrophoresis, respectively. Average 350 bp paired-end (PE) library was prepared using Illumina TruSeq DNA Sample Prep kit (Illumina Inc.) and was then sequenced on an Illumina's *NovaSeq* 6000 platform.

Data assembly, gene annotation and codon preference analysis

Raw reads were filtered using the base quality control software fastp (version 0.20.0) to obtain high-quality reads. Then a BLAST analysis was performed between the high-quality reads and the reference cp genome (KY432809.1) to extract cp-like reads. The obtained high-quality cp-like reads were further assembled into contigs via de novo assembler SPAdes v3.9.0 [56]. All contigs were sorted and joined into a single draft sequence using NOVOPlasty with the reference genome (KY432809.1) as a template. Cp cyclization and initiation site determination was done by manual processing. Prodigal [57] and hmmer [58] softwares were applied to annotate protein-coding genes and ribosomal RNAs respectively, while transfer RNAs were predicted via Aragorn software [59]. The annotation results were verified using CpGAVAS pipeline and then manually corrected. Finally, the cp genome map was generated using the OrganellarGenomeDRAW tool (OGDRAW) [60].

The codon preference was analyzed by R software. The degree of the codon preference was evaluated by the relative synonymous codon usage (RSCU). The RSCU value was calculated as the ratio between the use frequency and the expected frequency of a particular codon. According to the RSCU theory [61, 62], synonymous codon preference was partitioned into four models artificially: no preference ($\text{RSCU} \leq 1.0$), low preference ($1.0 < \text{RSCU} < 1.2$), moderate preference ($1.2 \leq \text{RSCU} \leq 1.3$), and high preference ($\text{RSCU} > 1.3$).

Comparative analysis and repeat sequence identification

Besides the newly sequenced *E. ophiuroides* cp genome, the complete cp genome sequences of the seven different plant species, including *Eremochloa ciliaris* (KY596146.1), *Eremochloa eriopoda* (KY596134.1), *Sorghum bicolor* (EF115542.1), *Zea mays* (AY928077.1), *Setaria italic* (KJ001642.1), *Oryza sativa* (MG252500.1), and *Brachypodium distachyon* (KU170609.1), were obtained from NCBI Organelle Genome Resources database (<http://www.ncbi.nlm.nih.gov/genome/organelle/>). Collinearity analysis

among the eight species was performed by using Mauve (<http://darlinglab.org/mauve/>) software with default parameters.

Tandem repeats were evaluated using Tandem Repeat Finder program [63] with default settings. Forward, palindromic, reverse and complement repeats were identified using vmatch v2.3.0

(<http://www.vmatch.de/>), and the minimal repeat size setting was greater than 30 bp with a Hamming distance of 3. Microsatellite or simple sequence repeats (SSRs) of one to six nucleotides were detected using the Perl script MISA v1.0 [64], and thresholds of eight, five, three, three, three, and three repeat units were set for mono-, di-, tri-, tetra-, penta- and hexa-nucleotide SSRs, respectively.

Phylogenetic analysis

The phylogenetic analysis was conducted based on *E. ophiuroides* cp genome data including a newly sequenced cp genome (MT806102) in the present study and another one submitted to NCBI GeneBank (KY432809.1) by Gallaher et al. in 2017, together with the cp genomes of 48 Andropogonodae species with strong genetic relationships downloaded from GeneBank, and *Arundinella deppeana* (NC030620) was selected as the outgroup (Table S4). MAFFT [65] and trimAl [66] were used for genome sequences alignment and data set trimmed, respectively. The best substitution model GTR+G was chosen in the jModelTest v2.1.7 [67], and the Randomized Axelerated Maximum Likelihood (RAxML) method was used to infer the phylogenetic relationship with 1000 bootstrap replicates in MEGA 6.0.

Abbreviations

IR: Inverted Repeat; LSC: Large Single Copy; SSC: Small Single Copy; PCG: Protein Coding Gene; tRNA: Transfer RNA; rRNA: Ribosome RNA; SSR: Simple Sequence Repeat; A: Adenine; T: Thymine; G: Guanine; C: Cytosine; cp: Chloroplast; NGS: Next-Generation Sequencing; NCBI: National Center For Biotechnology Information; RSCU: Relative Synonymous Codon Usage; ML: Maximal likelihood; RAxML: Randomized Axelerated Maximum Likelihood; JLA: Junction Line between LSC and IRa; JLB: Junction Line between LSC and IRb; JSA: Junction Line between SSC and IRa; JSB: Junction Line between SSC and IRb; PE: Paired End; OGDRAW: Organellar Genome DRAW

Declarations

Acknowledgements

The authors would like to acknowledge Ms. Aigui Guo for help with the maintenance of centipedegrass plants. The authors also thank the AJE - American Journal Experts Company (www.aje.com) for language editing.

Authors' contributions

J. Zong performed the experiments, analyzed the data, authored drafts of the paper, and approved the final draft; H. Wang and J. Wang performed the experiments, prepared figures and/or tables, and approved the final draft; H. Guo, J. Chen and D. Li performed the experiments, authored or reviewed drafts of the paper, and approved the final draft. J. Liu and J. Li conceived and designed the experiment, and J. Li authored or reviewed drafts of the paper, and approved the final draft.

Funding

This study was supported by grants from National Natural Science Foundation of China (32072608), National Project on Sci-Tec Foundation Resources Survey (2017FY100602) and the Opening Foundation of Jiangsu Provincial Key Laboratory for the Research and Utilization of Plant Resources (JSPKLB201841). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

Sequence information of the newly sequenced *E. ophiuroides* cp genome is available in the NCBI database (www.ncbi.nlm.nih.gov/) under the accession number MT806102. The reference cp genome of *E. ophiuroides* and the whole cp genome sequences of 48 species analysed in this study were all downloaded from the NCBI database (www.ncbi.nlm.nih.gov/) with their accession numbers listed in Table S4.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing of interests

The authors declare that they have no competing interests.

References

1. Hanna WW, Burton WG. Cytology, reproductive behavior and fertility characteristics of centipede grass. Crop Sci. 1978;18:835–837.

2. Islam M, Hirata M. Centipedegrass (*Eremochloa ophiuroides* (Munro) Hack.): growth behavior and multipurpose usages. Grassl Sci. 2005;51:183–190.
3. Li J, Guo H, Zong J, Chen J, Li D, Liu J. Genetic diversity in centipedegrass [*Eremochloa ophiuroides*(Munro) Hack.]. Hort Res. 2020;7:4.
4. Liu J, Hanna W, Elsner E. Morphological and seed set characteristics of centipedegrass accessions collected in China. Econ Bot. 2003;57:380–388.
5. Hirata M, Kunieda E, Tobisa M. Preference of cattle grazing conterminous monocultures of centipedegrass (*Eremochloa ophiuroides*) and bahiagrass (*Paspalum notatum*) with contrasting regrowth durations. Anim. Sci. J. 2017;88:909–917.
6. Li J, Zong J, Chen J, Wang Y, Li D, Li L, et al. De novo assembly and comparative transcriptome analysis reveals genes potentially involved in tissue-color changes in centipedegrass (*Eremochloa ophiuroides* [Munro] hack.). Plant Physiol Bioch. 2018;130:345–355.
7. Prabhudas SK, Prayaga S, Madasamy P, Natarajan P. Shallow whole genome sequencing for the assembly of complete chloroplast genome sequence of *Arachis hypogaea* Front Plant Sci. 2016;7:1106.
8. Dobrogojski, J, Adamiec M, Luciński. R. The chloroplast genome: a review. Acta Physiol. Plant. 2020;42:98.
9. Nock CJ, Waters DLE, Edwards MA, Bowen SG, Henry RJ. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol J. 2011;9:328–333.
10. Huang H, Shi C, Liu Y, Mao S, Gao L. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol Boil. 2014;14:151.
11. Takamatsu T, Baslam M, Inomata T, Oikawa K, Itoh K, Ohnishi T. et al. Optimized method of extracting rice chloroplast DNA for high-quality plastome resequencing and *de novo*Front Plant Sci. 2018;9:266.
12. Sun L, Jiang Z, Wan X, Zou X, Yao X, Wang Y, et al. The complete chloroplast genome of Magnolia polytepala: comparative analyses offer implication for genetics and phylogeny of Yulania, Gene. 2020;736:144410.
13. Saina JK, Li Z, Gichira AW, Liao Y. [The complete chloroplast genome sequence of tree of Heaven \(*Ailanthus altissima* \(Mill.\) \(Sapindales: Simaroubaceae\), an important pantropical tree.](#) Int J Mol Sci. 2018;19:929.
14. Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, Müller KF, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA. 2007;104:19369–19374.
15. Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in Araucaria? Mol Ecol Resour. 2015;15:1067–1078.

16. Lessard-Therrien M, Davies TJ, Bolmgren K. A phylogenetic comparative study of flowering phenology along an elevational gradient in the Canadian subarctic. *Int J Biometeorol*. 2014;58:455–462.
17. Sun C, Li J, Dai X, Chen Y. Analysis and characterization of the *Salix suchowensis* chloroplast genome. *J For Res*. 2018;29:1003–1011.
18. Burton GW. The adaptability and breeding of suitable grasses for the Southeastern States. *Adv Agron*. 1951;3:197–241.
19. Wang S, Gao L. Complete chloroplast genome sequence of green foxtail (*Setaria viridis*), a promising model system for C4 photosynthesis. *DNA*. 2015;27:1–2.
20. Takayuki A, Takahiko T, Sakiko T, Hiroaki S, Koh-Ichi K. Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA Res*. 2004;11:93–
21. Diekmann K, Hodgkinson TR, Wolfe KH, van den Bekerom R, Dix PJ, Barth S. Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (*Lolium Perenne*). *DNA Res*. 2009;16:165–176.
22. Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, et al. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol Gen Genet*. 1989;217:185–
23. Asaf S, Waqas M, Khan AL, Khan MA, Kang SM, Imran QM, et al. The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. *Front Plant Sci*. 2017;8:304.
24. Ogihara Y, Isono K, Kojima T, Endo A, Hanaoka M, Shiina T, et al. Chinese spring wheat (*Triticum aestivum*) chloroplast genome: Complete sequence and contig clones. *Plant Mol Biol Rep*. 2000;18:243–253.
25. Maier RM, Neckermann K, Igloi GL, Kössel H. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol*. 1995;251:614–
26. Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK, Luo H, et al. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor Appl Genet*. 2007;115:571–591.
27. Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K. Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA Res*. 2004;11:93–99.
28. Bortiri E, Coleman-Derr D, Lazo GR, Anderson OD, Gu YQ. The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes, *BMC Res Notes*. 2008;1:61.
29. Zhang H, Hall N, Mcelroy JS, Lowe EK, Goertzen LR. Complete plastid genome sequence of goosegrass (*Eleusine indica*) and comparison with other Poaceae. *Gene*. 2017;600:36–43.

30. Zhang T, Hu S, Zhang G, Pan L, Zhang X, Al-Mssallem IS, et al. The organelle genomes of Hassawi rice (*Oryza sativa*) and its hybrid in Saudi Arabia: genome variation, rearrangement, and origins. PLoS One. 2012;7:e42041.
31. Wambugu PW, Brozynska M, Furtado A, Waters DL, Henry RJ. Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences. Sci Rep. 2015;5:13957.
32. Chen S, Chen Y, Sun F, Waterman MS, Zhang X. A new statistic for efficient detection of repetitive sequences. Bioinformatics. 2019;35:4596–4606.
33. Cavalier-Smith T. Chloroplast evolution: secondary symbiogenesis and multiple losses. Curr Biol. 2002;12:62–64.
34. Timme RE, Kuehl JV, Boore JL, Jansen RK. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. Am J Bot. 2007;94:302–312.
35. Weng M, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. Mol Biol Evol. 2014;31:645–659.
36. Brazda V, Lýsek J, Bartas M, Fojta M. Complex analyses of short inverted repeats in all sequenced chloroplast DNAs. Biomed Res Int. 2018;1–10.
37. Xiang B, Li X, Qian J, Wang L, Ma Lin, Tian X, et al. The complete chloroplast genome sequence of the medicinal plant *Swertia Mussotii* using the Pacbio RS II platform. Molecules. 2016;21:1029.
38. Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD. The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome Res. 2008;18:30–38.
39. Zhu H, Song P, Koo DH, Guo L, Li Y, Sun S, et al. Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. BMC Genomics. 2016;17:557.
40. Li J, Zhou R, Endo TR, Stein N. High-throughput development of SSR marker candidates and their chromosomal assignment in rye (*Secale cereale*). Plant Breed. 2018;137:561–572.
41. Vieira LN, Dos Anjos KG, Faoro H, Fraga HP, Greco TM, Pedrosa FO, et al. Phylogenetic inference and SSR characterization of tropical woody bamboos tribe Bambuseae (Poaceae: Bambusoideae) based on complete plastid genome sequences. Curr Genet. 2016;62:443–453.
42. Bi Y, Zhang MF, Xue J, Dong R, Du YP, Zhang XH. Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. Sci Rep. 2018;8:1184.
43. He S, Yang Y, Li Z, Wang X, Guo Y, Wu H. Comparative analysis of four *Zantedeschia* chloroplast genomes: expansion and contraction of the IR region, phylogenetic analyses and SSR genetic diversity assessment. PeerJ. 2020;8:e9132.
44. Gu C, Ma L, Wu Z, Chen K, Wang Y. Comparative analyses of chloroplast genomes from 22 Lythraceae species: inferences for phylogenetic relationships and genome evolution within Myrtales. BMC Plant Biol. 2019;19:281.

45. Li J, Guo H, Wang Y, Zong J, Chen J, Li D, et al. High-throughput SSR marker development and its application in a centipedegrass (*Eremochloa ophiuroides* (Munro)) genetic diversity analysis. PLoS One. 2018;13:e0202605.
46. Li P, Lu R, Xu W, Ohi-Toma T, Cai M, Qiu Y, et al. Comparative genomics and phylogenomics of east Asian tulips (*Amana*, Liliaceae). Front. Plant Sci. 2017;8:451.
47. Shen X, Wu M, Liao B, Liu Z, Bai R, Xiao S, et al. Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. Molecules. 2017;22:1330.
48. Gui L, Jiang S, Xie D, Yu L, Huang Y, Zhang Z, et al. Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution. Gene. 2020;732:144355.
49. Yang Z, dos Reis M. Statistical properties of the branch-site test of positive selection. Mol Biol Evol. 2011;28:1217–1228.
50. Frailey DC, Chaluvadi SR, Vaughn JN, Coatney CG, Bennetzen JL. Gene loss and genome rearrangement in the plastids of five Hemiparasites in the family orobanchaceae. BMC Plant Biol. 2018;18:30.
51. Kellogg EA. Tribe Andropogoneae Dumort. in The families and genera of vascular plants vol. XIII, ed. K. Kubitski. Cham, Switzerland: Springer International. 2015;Pp.289–
52. Arthan W, McKain MR, Traiperm P, Welker CA, Teisher JK, Kellogg EA. Phylogenomics of Andropogoneae (Panicoideae: Poaceae) of Mainland Southeast Asia. Syst. Bot. 2017;42:418–431.
53. Burke SV, Wysocki WP, Zuloaga FO, Craine JM, Pires JC, Edger PP, et al. Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). BMC Plant Biol. 2016;16:140.
54. Saarela JM, Burke S, Wysocki WP, Barrett MD, Clark LG, Craine JM, et al. A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. PeerJ. 2018;6:e4299.
55. Lloyd Evans D, Joshi SV, Wang J. Whole chloroplast genome and gene locus phylogenies reveal the taxonomic placement and relationship of *Tripidium* (Panicoideae: Andropogoneae) to sugarcane. BMC Evol Biol. 2019;19:33.
56. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–477.
57. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
58. Eddy SR. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. PLoS Comput Biol. 2008;4:e1000069.
59. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32:11–16.
60. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet.

2007;52:267–274.

61. Zhao J, Qi B, Ding L, Tang X. Based on RSCU and QRSCU research codon bias of F/10 and G/11 Xylanase. J Food Sci Biotechnol. 2010;29:755–764.
62. Zuo L, Shang A, Zhang S, Yu X, Ren Y, Yang M, et al. The first complete chloroplast genome sequences of *Ulmus* species by *de novo* sequencing: Genome comparative and taxonomic position analysis. PLoS One. 2017;12:e0171264.
63. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–580.
64. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583–2585.
65. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–3066.
66. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–1973.
67. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9:772.

Tables

Due to technical limitations, tables are only available as a download in the Supplemental Files section.

Figures

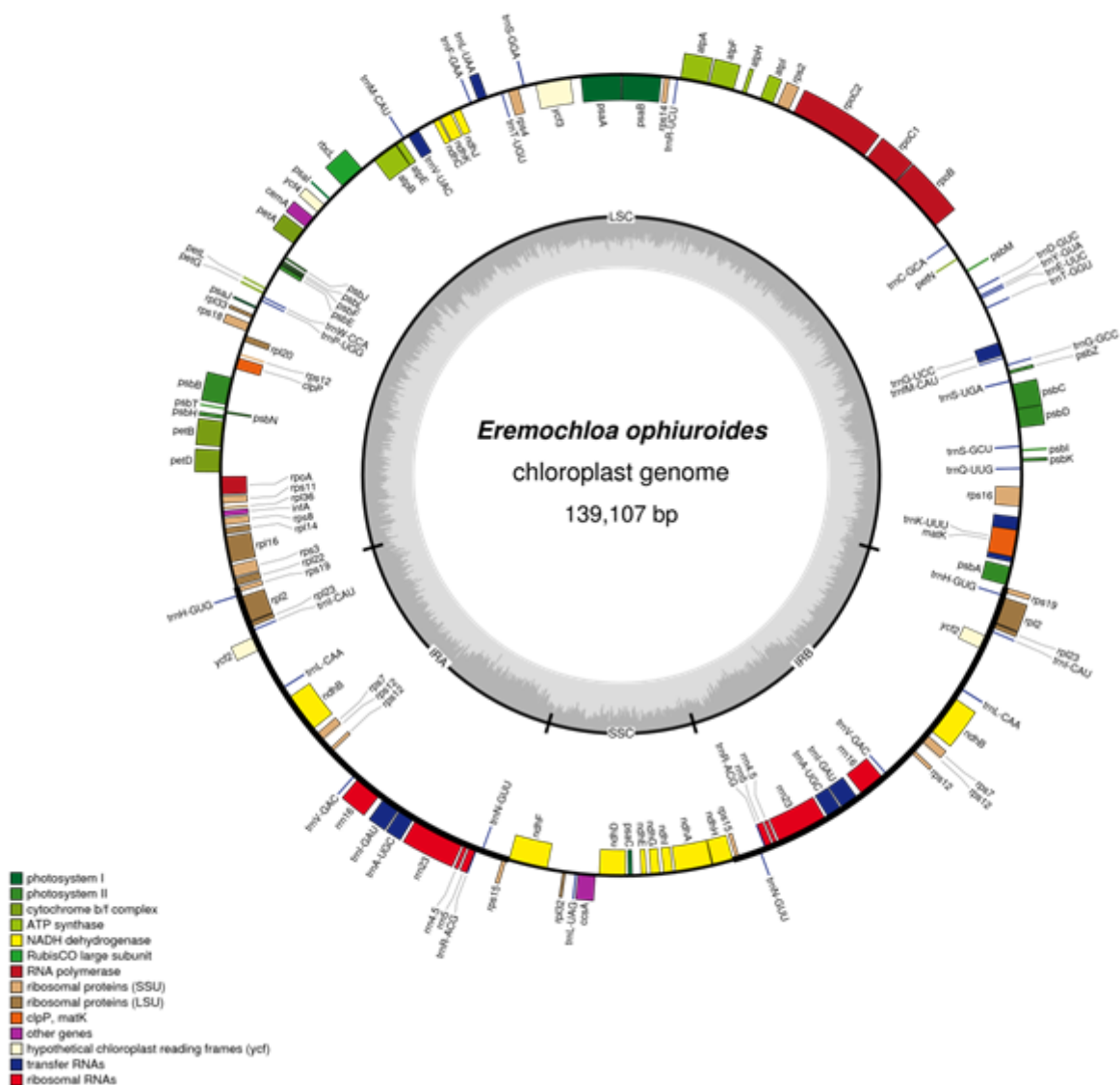


Figure 1

Physical map of *E. ophiuroides* complete chloroplast genome. Genes outside the circle are transcribed clockwise, and genes inside the circle are transcribed counterclockwise. Genes in the same color are in the same functional group. Internal circle of darker gray and lighter gray indicate GC content and AT content, respectively.

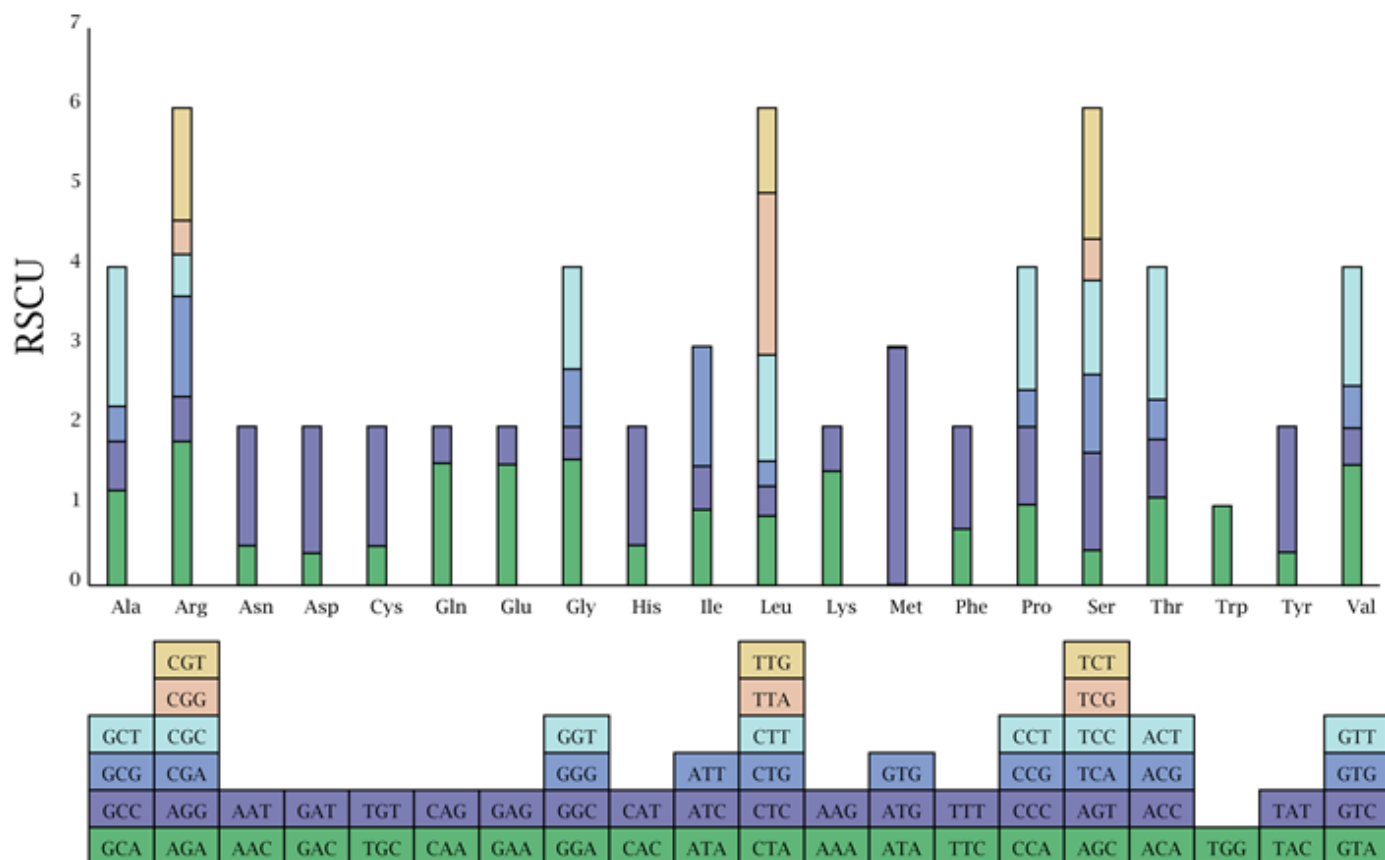


Figure 2

Codon content of 20 amino acids in all cp protein-coding genes of the *E. ophiuroides*.

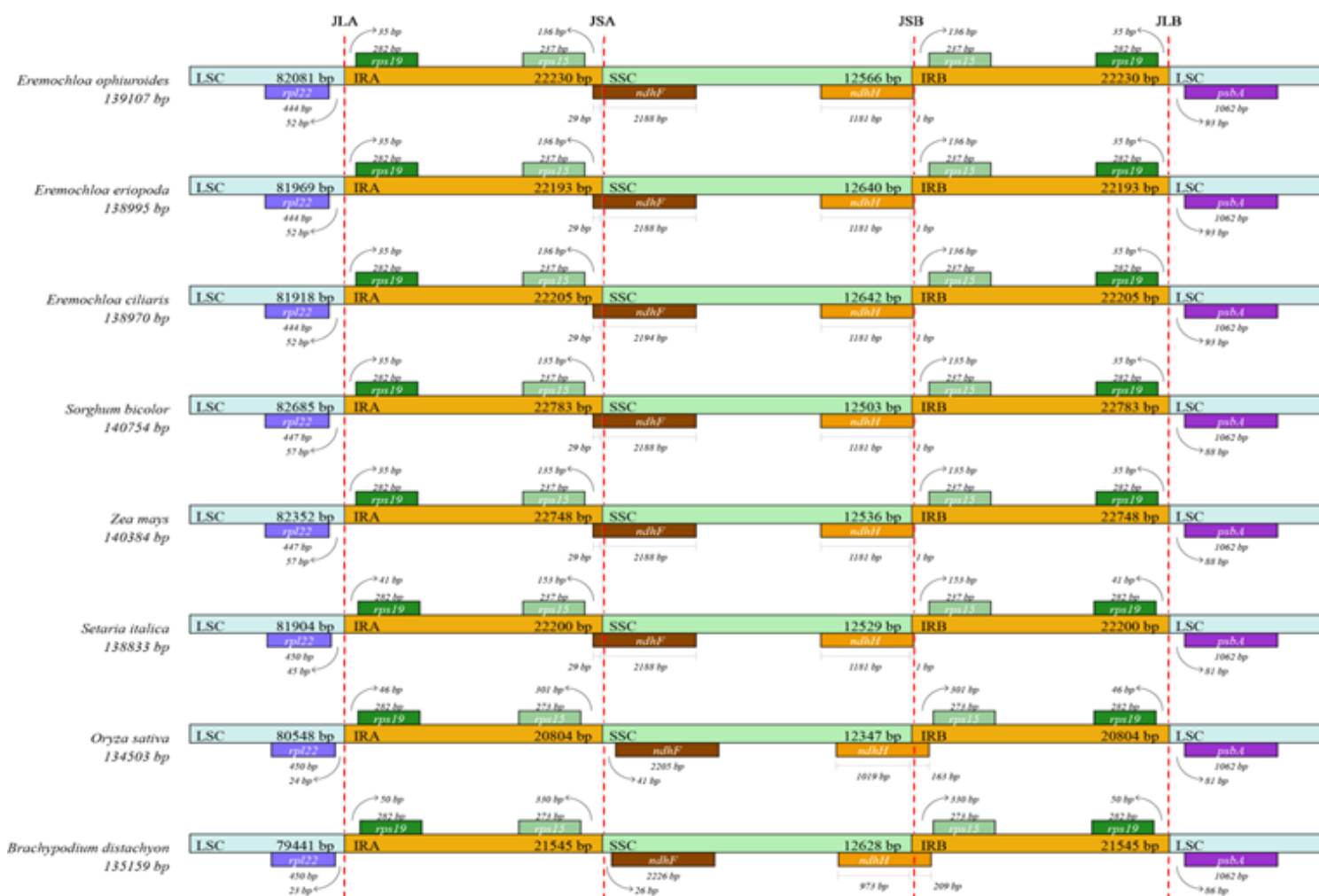


Figure 3

Comparison of the borders of IR regions among the cp genomes of eight plant species. Three colors were used to indicate the LSC, IR and SSC regions, respectively. JLA: junction line between LSC and IRA; JSA: junction line between SSC and IRA; JSB: junction line between SSC and IRB; JLB: junction line between LSC and IRB.

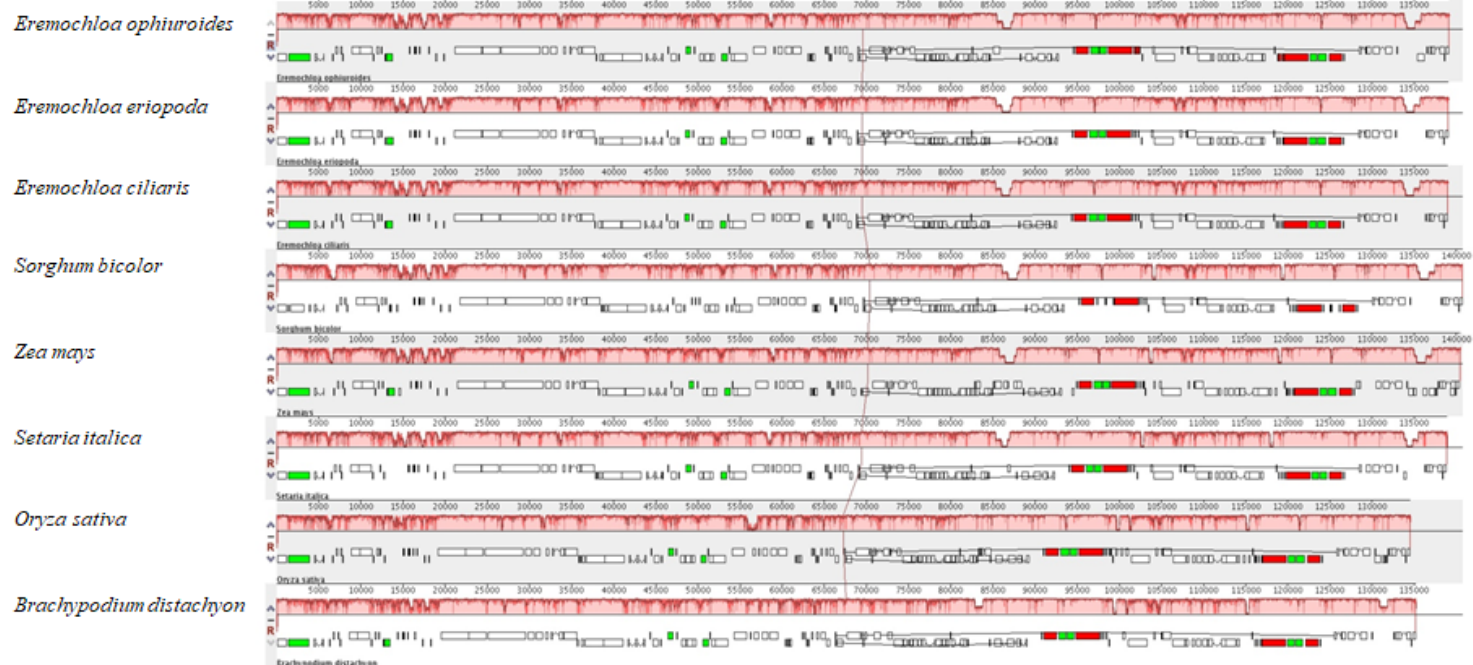


Figure 4

Gene collinearity analysis of the chloroplast genomes from eight Gramineae species. Annotations of protein coding, tRNA and rRNA genes are shown in white, green and red boxes respectively.

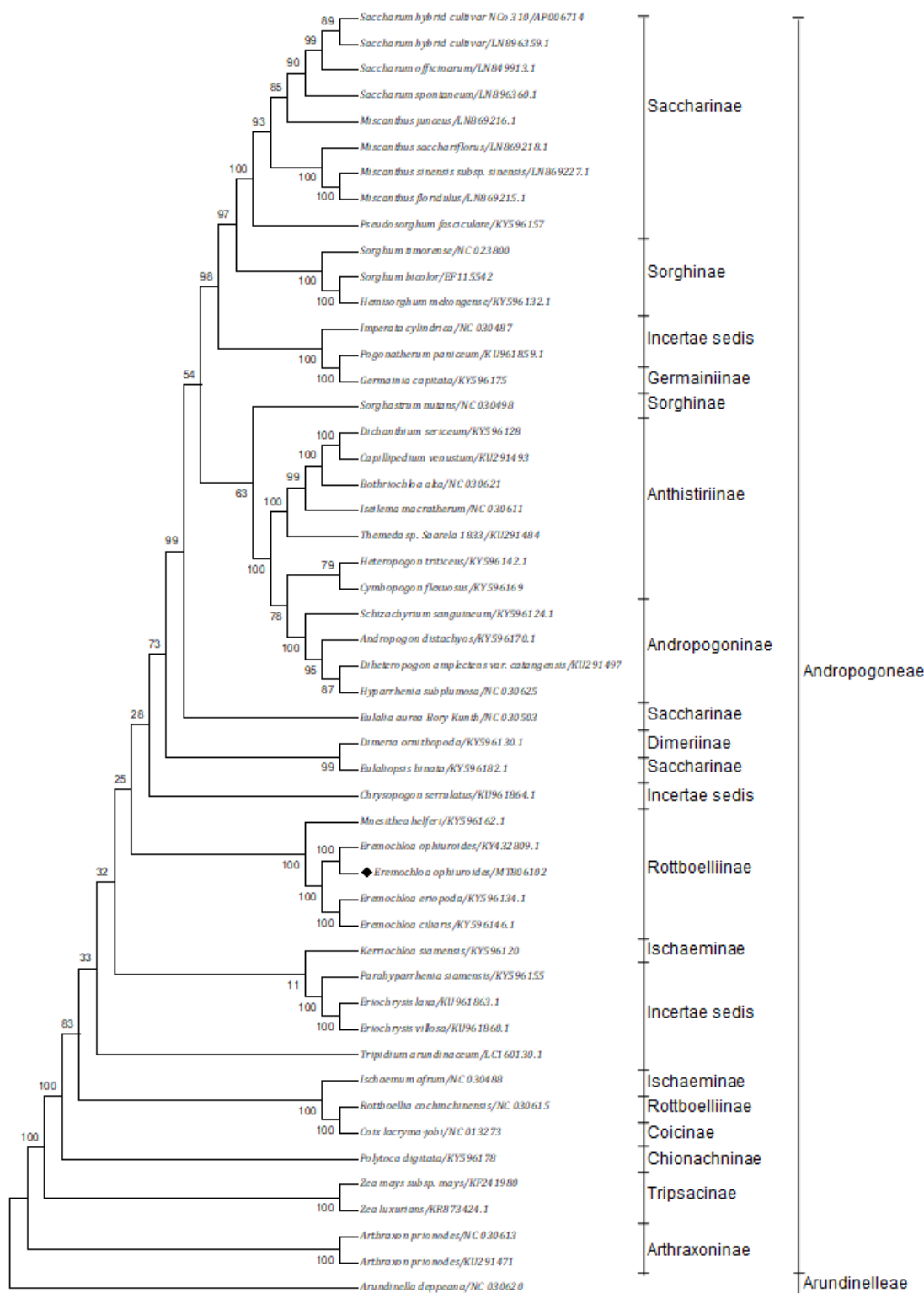


Figure 5

Phylogenetic tree of the Andropogoneae species based on the complete cp genome data by Maximum likelihood (ML). A total of 50 species were used to reconstruct a phylogenetic tree using MEGA6 software, and *Arundinella deppeana* was used as the outgroup. Subtribes and higher taxonomic groupings are indicated.

Supplementary Files

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

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [Table1.xlsx](#)
- [Table2.xlsx](#)