ORIGINAL ARTICLE

Molecular evolution of chloroplast genomes in Monsteroideae (Araceae)

Claudia L. Henriquez[1] · Abdullah[2] · Ibrar Ahmed[3] · Monica M. Carlsen[4] · Alejandro Zuluaga[5] · Thomas B. Croat[4] · Michael R. McKain[6]

Abstract

Main conclusion **This study provides broad insight into the chloroplast genomes of the subfamily Monsteroideae. The identified polymorphic regions may be suitable for designing unique and robust molecular markers for phylogenetic inference.**

Abstract Monsteroideae is the third largest subfamily (comprises 369 species) and one of the early diverging lineages of the monocot plant family Araceae. The phylogeny of this important subfamily is not well resolved at the species level due to scarcity of genomic resources and suitable molecular markers. Here, we report annotated chloroplast genome sequences of four Monsteroideae species: *Spathiphyllum patulinervum*, *Stenospermation multiovulatum*, *Monstera adansonii*, and *Rhaphidophora amplissima*. The quadripartite chloroplast genomes (size range 163,335–164,751 bp) consist of a pair of inverted repeats (25,270–25,931 bp), separating a small single copy region (21,448–22,346 bp) from a large single copy region (89,714–91,841 bp). The genomes contain 114 unique genes, including four rRNA genes, 80 protein-coding genes, and 30 tRNA genes. Gene features, amino acid frequencies, codon usage, GC contents, oligonucleotide repeats, and inverted repeats dynamics exhibit similarities among the four genomes. Higher rate of synonymous substitutions was observed as compared to non-synonymous substitutions in 76 protein-coding genes. Positive selection was observed in seven protein-coding genes, including *psbK*, *ndhK*, *ndhD*, *rbcL*, *accD*, *rps8*, and *ycf2*. Our included species of Araceae showed the monophyly in Monsteroideae and other subfamilies. We report 30 suitable polymorphic regions. The polymorphic regions identified here might be suitable for designing unique and robust markers for inferring the phylogeny and phylogeography among closely related species within the genus *Spathiphyllum* and among distantly related species within the subfamily Monsteroideae. The chloroplast genomes presented here are a valuable contribution towards understanding the molecular evolutionary dynamics in the family Araceae.

Keywords Araceae · Monsteroideae · Chloroplast genome · Polymorphic regions · Gene evolution · Phylogeny

---

Claudia L. Henriquez and Abdullah have contributed equally.

✉ Claudia L. Henriquez
  chenriquez@ucla.edu; claudiahenriquez901@gmail.com

✉ Abdullah
  abd.ullah@bs.qau.edu.pk

[1] University of California, Department of Ecology and Evolutionary Biology, Los Angeles, USA

[2] Department of Biochemistry, Faculty of Biological Sciences, Quaid-I-Azam University, Islamabad 45320, Pakistan

[3] Alpha Genomics Private Limited, Islamabad 45710, Pakistan

[4] Missouri Botanical Garden, St. Louis, MO, USA

[5] Departamento de Biología, Universidad del Valle, Calle 13, 100-00 Cali, Colombia

[6] The University of Alabama, Department of Biological Sciences, Tuscaloosa, AL, USA

Introduction

The chloroplast is a double membrane-bounded organelle (Cooper 2000). Chloroplasts contain their own DNA and replicate independently from the nuclear genome (Palmer 1985). This important organelle plays a role in photosynthesis and sustains life on earth (Daniell et al. 2016). Chloroplast genomes exhibit a circular quadripartite structure based on the arrangement of three important regions (Palmer 1985; Daniell et al. 2016; Mehmood et al. 2019; Abdullah et al. 2020). The inverted repeat (IRa and IRb) regions are situated between the large single copy region (LSC) and the small single copy region (SSC) (Palmer 1985; Daniell et al. 2016; Abdullah et al. 2019a; Yu et al. 2019a). However, quadripartite structure has not been observed in the chloroplast genome of various species such as Pinaceae (Wu et al. 2011), Cephalotaxaceae (Yi et al. 2013), Taxodiaceae (Hirao et al. 2008), Taxaceae (Zhang et al. 2014), Fabaceae (Sabir et al. 2014), and Cactaceae (Sanderson et al. 2015) due the loss of one or two IRs, whereas very short IRs are also reported in Pinaceae (Zeb et al. 2019). Moreover, linear chloroplast genome structure has also been reported along with the circular chloroplast genome (Oldenburg and Bendich 2016). The size of chloroplast genomes ranges from 107 kb (*Cathaya argyrophylla*) to 218 kb (*Pelargonium*) (Daniell et al. 2016). The chloroplast genome contains up to 120–130 genes including tRNA genes, rRNA genes, and protein-coding genes (Palmer 1985; Ahmed et al. 2012; Daniell et al. 2016; Iram et al. 2019; Abdullah et al. 2020).

The structure of chloroplast genomes is conserved regarding gene content, intron content, and gene organisation (Palmer 1985; Daniell et al. 2016; Shahzadi et al. 2019; Mehmood et al. 2019). However, events of gene loss, intron loss, gene rearrangements, and conversion of functional genes to pseudogenes has also been reported (Ahmed et al. 2012; Menezes et al. 2018; Hu et al. 2019; Liu et al. 2019a). Contraction and expansion of the inverted repeats occur frequently in chloroplast genomes and lead to origination of pseudogenes, duplication of genes, or conversion of duplicate genes to single copy (Ahmed et al. 2012; Menezes et al. 2018; Abdullah et al. 2019a, 2020; Liu et al. 2019a). Many mutational events occur within the chloroplast genome including inversions, oligonucleotide repeats, microstructural changes, InDels (insertions–deletions), and substitutions (Xu et al. 2015; Abdullah et al. 2019b; Shahzadi et al. 2019; Mehmood et al. 2019). Large-scale gene rearrangements have also been reported in some plant lineages, including Marattiaceae (Roper et al. 2007), Equisetaceae (Karol et al. 2010), Fabaceae (Schwarz et al. 2015), Geraniaceae (Marcussen and Meseguer 2017), Linaceae (Lopes et al. 2018), Passifloraceae (Rabah et al. 2019; Shrestha et al.

2019), and many non-photosynthetic plant species (Wicke et al. 2016).

Chloroplast genomes are inherited maternally in most angiosperms or paternally in some gymnosperms (Neale and Sederoff 1989; Daniell et al. 2016). Unlike the nuclear genome, chloroplasts lack meiotic recombination (Palmer 1985; Daniell et al. 2016). These properties, along with adequate levels of polymorphism, make it a suitable molecule for studies of evolution such as phylogeography, population genetics, phylogenetics, molecular evolution, and genome evolution (Ahmed et al. 2012, 2013; Li et al. 2013; Ahmed 2014; Henriquez et al. 2014; Xu et al. 2015; Marcussen and Meseguer 2017; Li and Zheng 2018; Zhai et al. 2019). Recently, several studies used either complete chloroplast genome sequences for inferring phylogenies (Feng et al. 2019; Zhai et al. 2019) or followed the alternate approach of Ahmed et al. (2013) and identified highly suitable polymorphic loci for designing unique markers for barcoding and phylogenetic studies in several plant lineages (Bi et al. 2018; Menezes et al. 2018; Shahzadi et al. 2019; Abdullah et al. 2019a, 2020).

The plant family Araceae is a large and ancient monocot plant family. This family belongs to the order Alismatales and comprises 118 describe genera and 3414 species, whereas 132 genera and 5946 species are expected (Boyce and Croat 2018). This family is unique among angiosperms based on its diverse morphology, ecology, and wide distribution from tropical to temperate regions (Gunawardena and Dengler 2006; Cabrera et al. 2008). This important family has been subdivided into eight subfamilies (Cabrera et al. 2008; Nauheimer et al. 2012) in which subfamily Monsteroideae is considered the third largest family with ca. 369 described species and ca. 700 estimated species (Boyce and Croat 2018). It comprises mostly hemiepiphytic or epiphytic plants restricted to the tropics, with three intercontinental disjunctions (Zuluaga et al. 2019). Monsteroideae is part of one of the earlier diverging lineages in Araceae and may help to provide a clearer picture of the evolution of the family (Zuluaga et al. 2019). Several studies based on plastid and nuclear markers inferred the phylogeny of this important subfamily, but still, the phylogeny of certain clades and genera is unresolved (Tam et al. 2004; Cabrera et al. 2008; Cusimano et al. 2011; Henriquez et al. 2014; Zuluaga 2015). A recent study inferred the phylogeny of 126 species of subfamily Monsteroideae based on five plastid and two nuclear markers which revealed the low polymorphism and low efficacy of these markers for species-level phylogenetic reconstruction of tropical Araceae (Zuluaga et al. 2019). For inferring the phylogeny of plant lineages with complex taxonomy, uses of specific and suitable polymorphic markers are required (Daniell et al. 2016). The comparative analyses of chloroplast genomes of subfamily Monsteroideae might be appropriate for identification of suitable loci for designing

cost-effective, unique, and robust markers. However, the chloroplast genomes of only two Monsteroideae species are reported, both from the same genus including *Spathiphyllum cannifolium* (Liu et al. 2019b) and *Spathiphyllum kochii* (KR270822). These genomic resources are insufficient for determination of suitable polymorphic loci for designing cost-effective markers with high-resolution potential.

The recent phylogenetic inference, based on chloroplast and nuclear markers, of the 126 species from various genera of subfamily Monsteroideae shows that species of the four genera are distinctly related, including *Spathiphyllum*, *Stenospermation*, *Monstera,* and *Rhaphidophora* (Zuluaga et al. 2019). To broaden the genomic resources and uncover the molecular diversity of the subfamily, we selected one species from each of four diverse genera of subfamily Monsteroideae as comparative analyses of the chloroplast genome sequences of diverse species are helpful in the identification of suitable polymorphic loci for designing of unique markers. The chloroplast genomes of *Spathiphyllum patulinervum*, *Stenospermation multiovulatum*, *Monstera adansonii*, and *Rhaphidophora amplissima* were assembled and annotated. These chloroplast genomes will be helpful in understanding the evolutionary dynamics and in the elucidation of chloroplast genome structure of subfamily Monsteroideae. The comparative analyses of these species enabled us to get insight into the evolutionary patterns, and molecular evolution of subfamily Monsteroideae. These resources also enabled us to identify suitable polymorphic loci for designing cost-effective, robust, and unique markers which could provide high-resolution potential for inferring phylogenies of subfamily Monsteroideae even at the species level.

Materials and methods

Samples collection, DNA extraction, and sequencing

Fresh, healthy leaf tissues were collected from the Araceae Greenhouse at the Missouri Botanical Garden in St. Louis, Missouri from four Monsteroideae species: *S. patulinervum*, *S. multiovulatum*, *M. adansonii*, and *R. amplissima*. We used 100 mg fresh leaves for whole-genomic DNA extraction and performed two extractions per taxon using Qiagen DNeasy Minikit (Qiagen, Germantown, Maryland, USA). The DNA of each extraction was eluted in 125 μL elution buffer. The quantity and quality of DNA were confirmed by 1% agarose gel electrophoresis and Nanodrop (ThermoScientific, Delaware, USA). The libraries were constructed according to the manufacturer's protocol of Illumina TruSeq kits (Illumina, Inc., San Diego, California) in the Pires lab at the University of Missouri,

Columbia. The qualified libraries were sequenced from single end 100 bp reads using the Illumina HiSeq 2000 at the University of Missouri DNA Core. The sequencing of whole-genomic DNA by HiSeq2000 with 100 bp short-read length produces 3.36 GB (12.87 million short reads) in *R. amplissima* to 9.47 GB (36.28 million reads) in *M. adansonii*.

Short-read data-quality analyses and submission to Sequence Read Archive

The quality of short reads were analysed by FastQC (Andrews 2017) in the Galaxy portal (https://usegalaxy.org). Due to a file size limit of 2 GB in Galaxy, fastq.gz files for samples with concatenated fastq.gz files larger than 2 GB were uploaded individually (*M. adansonii* and *S. multiovulatum*). To compare the quality of the raw data across samples, a MultiQC (Ewels et al. 2016) analysis was also performed in the Galaxy portal. These analyses confirmed the quality of the raw reads with average Phred score 35.19–37.85. The raw data of all four species were submitted to the National Center for Biotechnology Information (NCBI) under Sequence Read Archive (SRA) number PRJNA547622.

Genome assembly and annotations

Chloroplast genomes were assembled using Fast-Plast v. 1.2.2 pipeline (https://github.com/mrmckain/Fast-Plast) under default settings. The reads were first clean by Trimmomatic v. 0.36 (Bolger et al. 2014). The reads of chloroplast origin were extracted from clean reads by mapping to available Alismatales plastomes packaged using bowtie2 v. 2.2.9 (Langmead and Salzberg 2012) with the "very-sensitive-local" parameter. We used SPAdes v.3.9.0 (Bankevich et al. 2012) with various k-mers to assemble the extracted mapped reads. The contigs generated by SPAdes v.3.9.0 were assembled using afin v. 1.0 (https://github.com/afinit/afin) with three iterations of 150, 50, and 50 loops, an initial contig trimming of 100 base pairs, a 20, 15, and 10 (per iteration) overlap of reads to contigs, a minimum coverage of 2, 1, and 1 reads per loop, and the full set of trimmed reads from the Trimmomatic phase. In case of assembly of complete genome in single contig, Perl code was used to orientate and identify single copy regions and inverted repeat regions. If assembly of the complete chloroplast genome was not gained from the contigs generated with SPAdes v.3.9.0 by afin v. 1.0. Then, Sequencher (Genecodes, Ann Arbor, MI, USA) was used with the clean short reads to bridge gaps as in McKain et al. (2016). The mapping of short reads to assemble genome was performed after the step of Sequencher using bowtie2 v. 2.2.9. The chloroplast assembly was verified through a coverage analysis conducted in Jellyfish2 v. 2.2.6 (Marçais and Kingsford 2011) under

default parameters. The threshold of 25-mer abundance was used to map a 25-bp sliding window of coverage across the chloroplast genome of each species to determine misassembled regions if any. In case of identification of any misassembled regions in the assembly, the chloroplast genome was reassembled by repeating all the steps from afin v.1.0. After obtaining the assembled chloroplast genome, the clean raw reads were once again mapped to the final assembled genome using bowtie2 v. 2.2.9. After mapping of reads to assemble genome, Pilon v. 1.21 (Walker et al. 2014) was used to identify and fix any potential assembly issues and identify some minor mislabelled base call variants. The program Pilon is usually used for improvement and removal of small errors that exist in the assembled genome. Hence, we accepted the output of Pilon as the final assemblies.

The coverage depth of the final assembled chloroplast genome was again performed with Bowtie2 v.2.2.9 and ranged from 43.1X (*S. patulinervum*) to 449.9X (*S. multiovulatum*). The detail of the quantity and quality of raw reads, number of chloroplast genome reads, coverage depth, and NCBI accession number is provided in Table 1.

The newly assembled chloroplast genome was annotated using GeSeq (Tillich et al. 2017), whereas the tRNA genes were further verified by ARAGORN v1.2.38 (Laslett and Canback 2004) and tRNAscan-SE v.2.0.3 (Lowe and Chan 2016). The start and stop codons of protein-coding genes that were identified by GeSeq were further confirmed by manual visualisation as well as by blasting with homologues genes in Geneious R8.1 (Kearse et al. 2012). The stop codon of gene was also confirmed by analysing translation of each protein-coding genes in Geneious R8.1 (Kearse et al. 2012). A gene was declared pseudogene if contained internal stop codon as compared to other homologues genes or exists as a truncated/partial copy of a gene. GB2Sequin (Lehwark and Greiner 2019) was used to generate five-column tab-delimited annotation file for submission of the chloroplast genome of each species to GenBank at the National Centre for Biotechnology Information (NCBI) with specific accession numbers (Table 1). Fully annotated plastomes of circular diagram were drawn by OrganellarGenomeDRAW (OGDRAW) (Lohse et al. 2007).

Comparative analyses, determination of polymorphic loci, and phylogenetic inference

Geneious R8.1 (Kearse et al. 2012) was used for comparison of genomic features and for determination of codon usage and amino acid frequency. IRscope (Amiryousefi et al. 2018) was used for the analyses of inverted repeat region contraction and expansion at the junctions of chloroplast genomes. The Geneious R8.1 (Kearse et al. 2012) integrated Mauve alignment (Darling et al. 2004) was used to analyse chloroplast genome organisation and gene arrangement based on analyses of collinear blocks.

The rate of synonymous (Ks) and non-synonymous (Ka) substitutions and their ratio (Ka/Ks) of 76 protein-coding genes were also determined. The MAFFT alignment of each protein-coding genes was exported in FASTA format from Geneious R8.1 (Kearse et al. 2012) after removal of stop codon and analysed in DnaSP v. 5.10.01 (Rozas et al. 2017) following the previous studies (Choi et al. 2018; Kim et al. 2019; Abdullah et al. 2020). We used the *S. kochii* as reference for all the species of subfamily Monsteroideae and results were interpreted as: Ka/Ks > 1 indicate positive selection, < 1 indicate purifying selection, and Ka/Ks = 1 indicate neutral selection.

REPuter (Kurtz et al. 2001) program was employed to identify oligonucleotide repeats: palindromic, reverse, forward, and complementary. The parameters for repeats determination were set as to identify repeats pair of ≥ 30 bp with minimum similarity index of 90%.

Suitable polymorphic regions were determined using two different approaches of chloroplast genome comparison. At the family level, we aligned all the reported four species of subfamily Monsteroideae using multiple alignment of MAFFT (multiple alignment using fast Fourier transform) (Katoh et al. 2005) and compared all protein-coding genes, conserved IGS (intergenic spacer regions) and conserved intronic regions following Abdullah et al. (2019b) and Shahzadi et al. (2019). Most of the IGS and intronic regions showed high level of polymorphism and produce high missing data (> 5%) due to large number of InDels and inversions as compared to substitutions. Hence, these regions were not considered suitable for the

Table 1 Quantity and quality of the sequencing data and coverage depth of the assembled genomes

Sample name	Data in GB	Whole genome Sequencing reads	Phred score	Chloroplast genome reads	Coverage mean	Coverage max	NCBI accession
Monstera adansonii	9.47	36,278,235	35.19	1,852,168	113.1	1720	MN046888
Rhaphidophora amplissima	3.36	12,870,626	37.63	247,952	151.8	269	MN046885
Stenospermation multiovulatum	11.17	42,788,539	37.04	737,289	449.9	666	MN046893
Spathiphyllum patulinervum	3.62	13,898,684	37.85	71,004	43.1	110	MN046890

phylogenetic inference of the subfamily Monsteroideae and were discarded from the list of suitable polymorphic loci. The inversions also provide false results of phylogenetic relationships (Menezes et al. 2018). The number of InDels and substitutions of each region were counted manually and divided on the length of the alignment to find percentage diversity of each region. At the genus level, we compared *S. patulinervum* (reported in current study) and *S. kochii* (KR270822) (downloaded from NCBI) using MAFFT pairwise alignment following Abdullah et al. (2019a). In the pairwise alignment, we compared each protein-coding sequence, intronic region, and intergenic spacer region to identify suitable polymorphic loci for designing of unique and robust markers. The average diversity of each region has been determined by dividing the number of substitutions and InDels by the size of the alignment of each region. We also removed the inversion from the alignment to avoid false results.

We used a total of 18 species in inferring the phylogeny including 14 downloaded species from NCBI. The details of the species are provided in Table S1. We used *Acorus americanus* from family Acoraceae to root the tree. The phylogenetic relationships were inferred based on complete chloroplast genomes following Abdullah et al. (2019a) after removal of IRa region from each genome. The IQ-tree (Nguyen et al. 2015; Kalyaanamoorthy et al. 2017; Hoang et al. 2018) program was used for the reconstruction of the phylogenetic tree with default parameters including 1000 replication and 1000 iteration along with best fit model TVM + F + I + G4. The TreeDyn program was used to improve visualization of the phylogenetic tree (Dereeper et al. 2008).

Results

Comparison of chloroplast genomic feature in subfamily Monsteroideae

The size of the chloroplast genome ranged from 163,335 bp (*R. amplissima*) to 164,751 bp in *S. patulinervum*. The size of LSC ranged from 89,714 (*R. amplissima*) to 91,841 bp (*S. patulinervum*), SSC ranged from 21,448 bp (*S. multiovulatum*) to 22,346 bp (*S. kochii*), and the size of each IR region ranged from 25,270 bp (*S. kochii*) to 25,931 bp (*S. multiovulatum*) (Table 2). All the species showed conserved intron, gene content, and gene organisation. The circular map of chloroplast genomes and collinear blocks (LCBs) of Mauve alignment confirmed the high similarities in these species (Figs. 1, 2). The LCBs also revealed similarity in gene arrangement and chloroplast genome organisation. The average GC content of the chloroplast genomes was 36% and revealed a high extent of similarity. However, the GC content showed variation among the three main regions of the chloroplast. IR regions showed high GC content compared to the LSC and SSC regions. All the species have

Table 2 Comparison of chloroplast genomes of *Monstera adansonii*, *Rhaphidophora amplissima*, *Stenospermation multiovulatum*, *Spathiphyllum patulinervum*, and *Spathiphyllum kochii*

| Characteristics	Monstera adansonii	Rhaphidophora amplissima	Stenospermation multiovulatum	Spathiphyllum patulinervum	Spathiphyllum kochii
Size (base pair; bp)	163,643	163,335	163,822	164,751	163,368
LSC length (bp)	89,999	89,714	90,512	91,841	90,471
SSC length (bp)	22,162	22,053	21,448	21,600	22,346
IR length (bp)	25,741	25,784	25,931	25,655	25,270
Number of genes	114	114	114	114	114
Protein-coding genes	80	80	80	80	80
tRNA genes	30	30	30	30	30
rRNA genes	4	4	4	4	4
Duplicate genes	17	17	17	17	17
Intron containing genes	18	18	18	18	18
GC content					
Total (%)	36.1	36.1	35.9	36.1	36.3
LSC (%)	34.6	34.6	34	34.4	34.5
SSC (%)	28.5	28.6	29	29.4	29.2
IR (%)	42.1	42.0	42.4	41.8	42.7
CDS (%)	37.9	37.9	37.9	37.9	38
rRNA (%)	55	55	55	55	55
tRNA (%)	53.1	53.1	53	53.1	52.4
All genes (%)	39.4	39.4	39.1	39.3	39.5
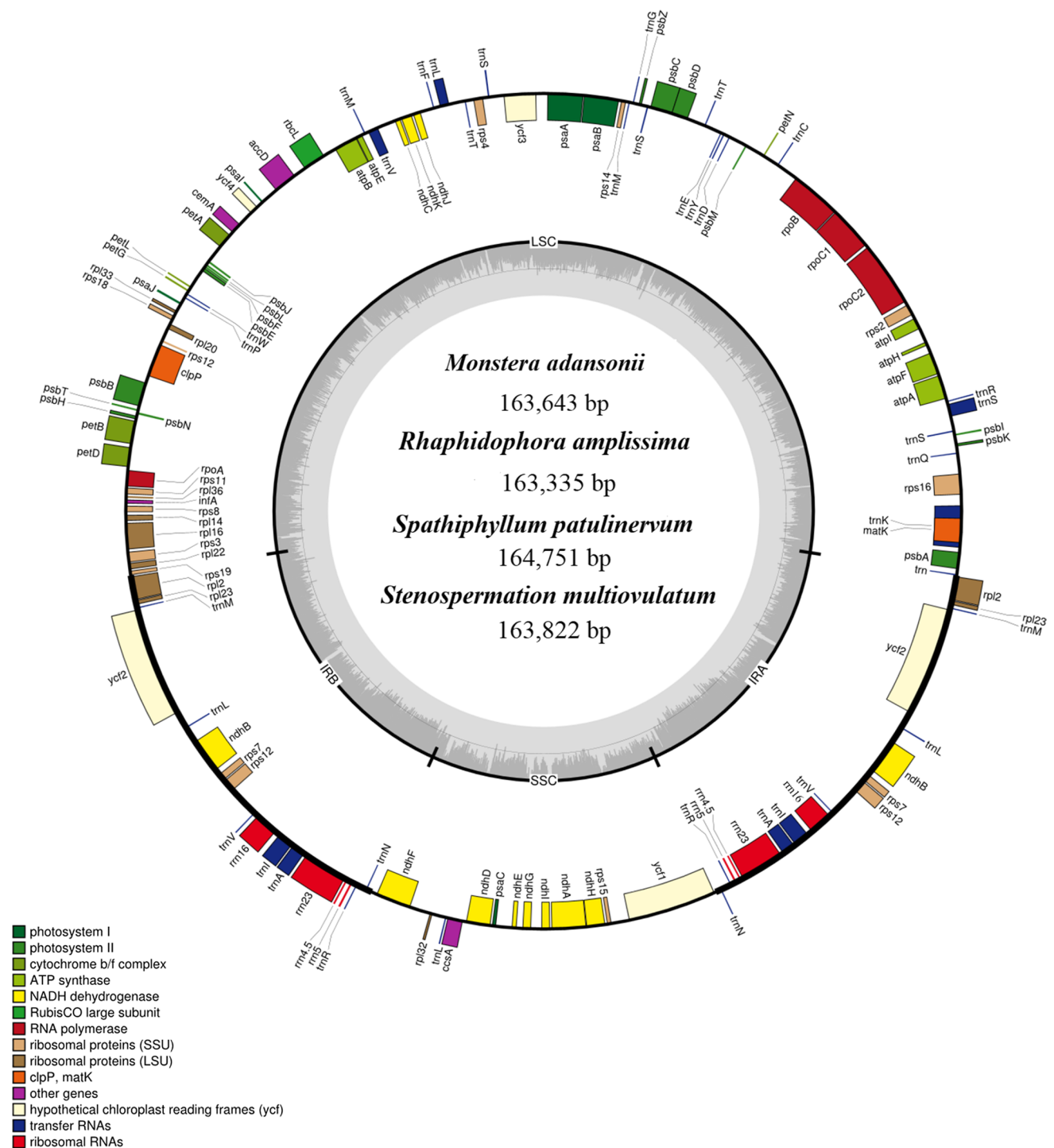
Fig. 1 Circular map of Monsteroideae chloroplast genomes. Genes transcribed counter clockwise are present inside of the circle. Genes transcribed clockwise are present outside of the circle. The colour of the genes correspond to the function of the genes

114 unique genes that included 30 tRNA genes, four rRNA genes, and 80 protein-coding genes (Table 2, Fig. 1). In IR regions, 17 genes were present and duplicated that included seven tRNA genes, four rRNA genes, and six protein-coding genes. We found 18 intron containing genes including six

tRNA genes and 12 protein-coding genes. Among intron containing genes, two tRNA genes and three protein-coding genes contained introns (Table S2). The *infA* gene exists as a pseudogene in all species. The *ycf1* gene was found functional in all species at junction of SSC/IRa. However, a
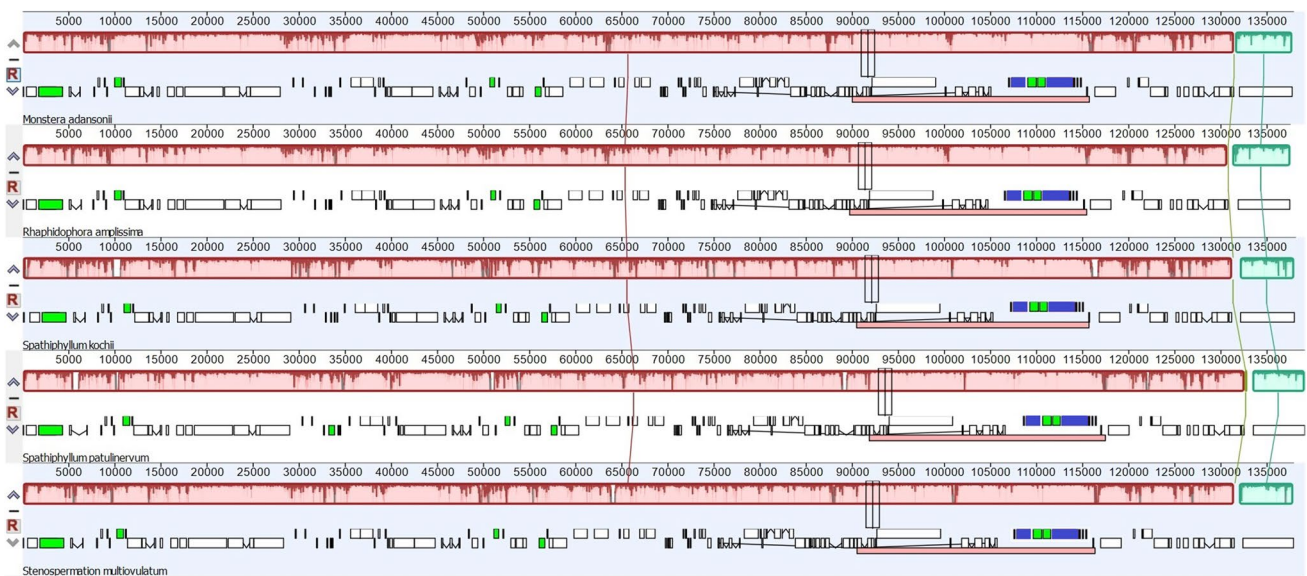
Fig. 2 Collinear block (LCBs) analyses of Monsteroideae genome. The colours of LCBs represent genes. The white colours represent protein-coding genes, black colours represent tRNA genes, green colour represent intron containing tRNA genes, and blue colour represent rRNA genes. The LCBs analyses revealed high level of similarities among the Monsteroideae species
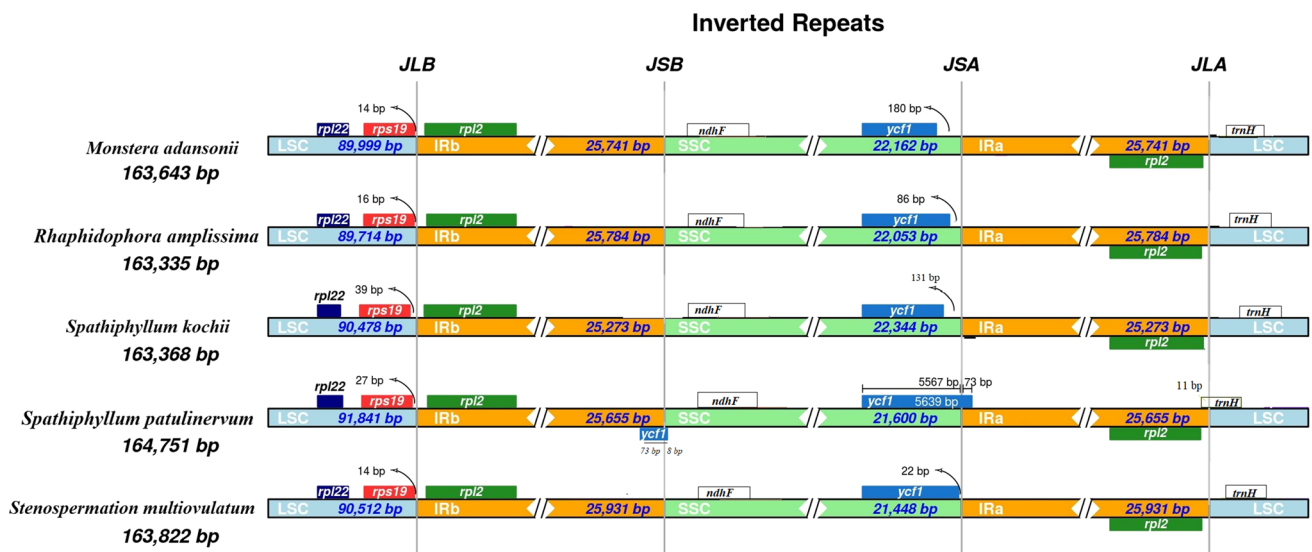

Fig. 3 Inverted repeat contraction and expansion analyses of quadripartite junctions of chloroplast genomes among Monsteroideae. Genes represented above the block are transcribed counter clockwise and genes represented below the block are transcribed clockwise. The number represented above the arrow show the distance of the genes from the junctions. Genes that are integrated between the junctions are represented with scale bar and the number shows the extent of base pair by which integration from one region to another region takes place

pseudo-copy of *ycf1* also originated at the junction of IRb/SSC only in *S. patulinervum*, along with functional copy of *ycf1* at junction of SSC/IRa, due to starting in the IR regions instead of its complete presence in the SSC region (Fig. 3).

Analyses of inverted repeat regions contraction and expansion

The inverted repeat regions contraction and expansion revealed high similarities at the junctions of LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC. The *rps19* gene was

completely found in the LSC region, and *rpl2* was completely found in the IRb region at the junction of LSC and IRb. The *ndhF* gene was completely found in the SSC region at the junction of IRb and SSC. At the junction of SSC and IRa, *ycf1* gene was completely located in the SSC region of three species, whereas in one species (*S. patulinervum*), it started from the IR regions and integrated into the SSC region. Hence, it left a pseudogene of 81 bp at the junction of IRb and SSC in *S. patulinervum*. The *trnH* gene showed complete presence in the LSC region in four species except *S. patulinervum* in which *trnH* gene integrated into IRa with 11 bp present at the junction of IRa and LSC (Fig. 3).

Codon usage and amino acid frequency

The relative synonymous codon usage (RSCU) analyses revealed that codons that end with A/T at 3′ end have RSCU ≥ 1 and encode the highest amount of amino acids. The codons that end with C/G at 3′ have RSCU < 1 and encode the lowest amount of amino acids (Table S3). The ATG codon encodes formyl-methionine as a start codon using a specific tRNA-fMet (CAU) for translation initiation and a methionine during translation elongation using the tRNA-Met (CAU). This was the most common start codon in the chloroplast genome of all species. However, other codons were also found as start codons such as ACG (in *rpl2*), ATA (in *cemA*), and GTG (in *rps19*). The analysis of amino acid frequencies revealed that leucine is the most abundant and cysteine the most rare amino acid. In general, we found high similarities in codon usage and amino acids frequency among the four species of subfamilies Monsteroideae (Figure S1).

Rate of evolution of protein-coding genes

The rate of synonymous substitutions (Ks), non-synonymous (Ka) substitutions, and their ratio (Ka/Ks) showed low rates of evolution for all types of protein-coding genes in the chloroplast genome. The synonymous substitutions were found more common than non-synonymous substitutions; therefore, less value of Ka/Ks was observed. We observed low average value of Ks, Ka, and Ka/Ks for different groups of genes as: cytochrome group (Ks = 0.0193, Ka = 0.0015, and Ka/Ks = 0.0784), photosystem I group (Ks = 0.0312, Ka = 0.0036, and Ka/Ks = 0.1145), photosystem II (Ks = 0.0253, Ka = 0.0042, and Ka/Ks = 0.1678), ribosomal small subunit group (Ks = 0.0359, Ka = 0.0063, and Ka/Ks = 0.1753), ATP synthase group (Ks = 0.0201, Ka = 0.0036, and Ka/Ks = 0.1797), NADPH dehydrogenase group (Ks = 0.0238, Ka = 0.0058, Ka/Ks = 0.2434), ribosomal large subunit (Ks = 0.0172, Ka = 0.0046, Ka/

Ks = 0.2693), and RNA polymerase group (Ks = 0.0253, Ka = 0.0073, Ka/Ks = 0.2876). Our data revealed that purifying selection acts on the genes of cytochrome group, photosystem I group, and photosystem II group. The details about the evolution of each gene is provided in Table S4. We found seven genes that showed Ka/Ks ≥ 1 and hence showed positive selection pressure. Genes *ndhK*, *rbcL*, *ycf2*, and *ndhD* showed positive selection in *S. patulinervum*; *accD* and *rps8* in *R. amplissima* and *psbK* in *S. multiovulatum*.

Repeats analyses

REPuter detected four types of oligonucleotide repeats: palindromic, reverse, forward, and complementary. The abundance of the repeats varies among species based on types of repeats. The forward repeats showed abundance in *S. patulinervum*, reverse repeats showed abundance in *S. multiovulatum*, palindromic repeats in *M. adansonii*, and complementary repeats also in *M. adansonii* (Fig. 4a). The size of repeats varies among species, but most of the repeats exist in the range of 35–44 bp (Fig. 4b). The LSC region contained most of the repeats than SSC and IR regions, whereas some repeats were also shared among the different regions of the chloroplast (Fig. 4c). The analysis of repeats distribution based on functional regions of the chloroplast revealed the presence of most of the repeats in intergenic spacer regions (IGS) as compared to protein-coding sequences and introns (Fig. 4d). The detail of the repeats is also provided in Table S5.

Identification of suitable polymorphic loci at subfamily and genus level

The comparison among the species of the subfamily Monsteroideae identified 30 polymorphic regions for phylogenetic inference at the subfamily level, and identified suitable polymorphic loci from IGS regions, intronic regions, and protein-coding sequences (Table 3). Most of the regions were included from IGS regions, namely *trnQ-UUG-psbK*, *trnW-trnP*, *rpoA-rps11*, *petG-trnT*, and *trnC-petN*. We did not include those regions in the list of suitable polymorphic loci which produce high levels of missing data (> 5%) due to multiple InDels and inversions events. The protein-coding regions include *ycf1*, *psbK*, *ccsA*, *accD*, *rbcL*, *matK*, and *ndhF*. The intronic regions of *petL* and *rpoC1* were also included in the suitable polymorphic loci. The sequence of some IGS and intronic regions were partially included in the list of suitable polymorphic loci to avoid those sequences which revealed high polymorphism and showed high level of InDels and inversion. The sequences that were chosen from each region are given in Table 3, whereas the percentage of polymorphism of all the polymorphic protein-coding sequences along some conserved IGS and intronic regions
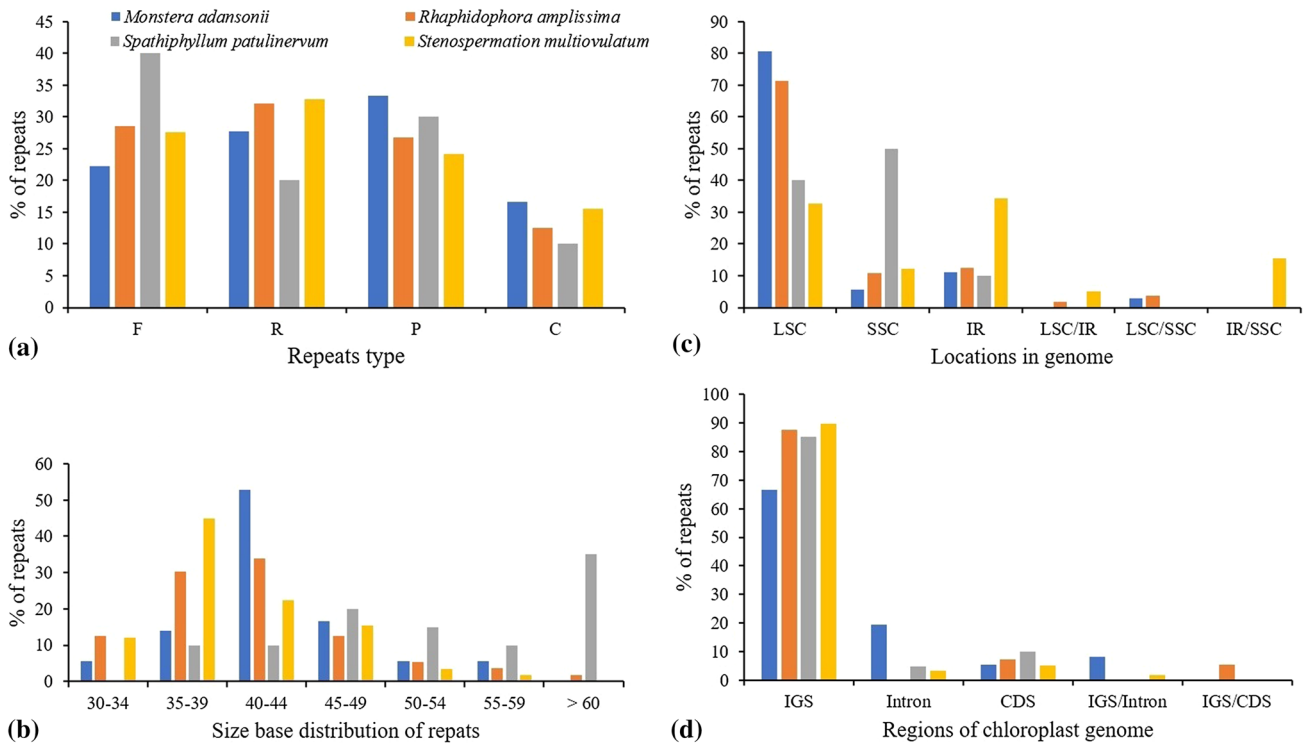
Fig. 4 Oligonucleotide repeats in Monsteroideae chloroplast genomes. **a** Represent four types of repeats: F (forward), R (reverse), P (palindromic), and C (complementary). **b** Represent size of repeats, i.e., 30–34 represent repeat range in size from 30 to 34 bp and so on. **c** Represent distribution of repeats in the three main regions of chloroplast genome: LSC (large single copy), SSC (small single copy), and IR (inverted repeat). The LSC/IR, LSC/SSC, and IR/SSC repre-

sent shared repeats, i.e., LSC/IR represent repeats pair in which one copy of repeats is present in LSC and another copy is present in IR. **d** Represent distribution of repeats in functional regions of chloroplast genome; IGS (intergenic spacer regions), Intron (intronic regions), and CDS (protein-coding sequences). The IGS/intron and IGS/CDS represent share repeats between these regions

has been given in Fig. 5a. These polymorphic loci might be suitable for phylogenetic inference of the subfamily Monsteroideae.

We also compared all the regions of chloroplast genomes of *Spathiphyllum* species in pairwise alignment. The average nucleotide differences of intergenic spacer regions (0.0140) were found the highest followed by intronic regions (0.0088) and then by protein-coding sequences (0.0048). We identified 30 high polymorphic loci in which most of the regions belong to IGS including *trnS-GCU-trnG-UCC*, *trnH-psbA*, *atpH-atpI*, *trnP-UGG-psaJ*, *psbK-psbI*, and *psaC-ndhE* (Table 4). The polymorphic loci that belonged to intronic regions include *atpF* and *ndhA,* whereas the polymorphic regions that belonged to protein-coding sequences included *ycf4*, *rpl22*, and *cemA*. The nucleotide differences of complete chloroplast genomes of the *Spathiphyllum* species are given in Fig. 5b, c. These polymorphic loci might be helpful for phylogenetic inference and population genetic studies of the species of genus *Spathiphyllum*.

Phylogenetic relationships among the aroid species

The phylogeny of subfamily Monsteroideae, without problematic species, has been inferred based on complete chloroplast genomes (IRa not included). After removal of InDels, the alignment contained 94,257 nucleotide sites in which 67,230 (71.32%) sites were constant, 6522 sites showed distinct patterns, and 15,500 sites were found to be parsimony informative. The phylogenetic tree of the studied species supports the monophyletic position of five species of subfamily Monsteroideae (Fig. 6).

Discussion

In the current study, we report chloroplast genome sequences of four species of subfamily Monsteroideae. We compared genomic features among the species of Monsteroideae, analysed IRs contraction and expansion, and identified suitable polymorphic loci for designing of suitable molecular markers.

Table 3 Polymorphic loci identified by comparing four species of Monsteroideae

S. no.	Region	Percent diversity	No. of mutations	Alignment length	Missing data (%)
1	*trnQ-UUG-psbK*	7.75	29	374	1.87
2	*trnW-trnP*	7.43	13	175	3.43
3	*rpoA-rps11*	6.94	5	72	0
4	*petG-trnT*	6.56	8	122	0
5	*trnC-petN*	6.25	30	480[a]	2.92
6	*rpl14-rpl16*	6.2	8	129	1.55
8	*ndhC-ndhK*	5.52	19	344[a]	4.65
9	*matK-trnK-UUU*	5.34	41	765	1.44
10	*psaC-ndhE*	5.15	15	291	1.72
11	*ycf1*	4.85	248	5116	0.89
12	*ycf4-petA*	4.50	10	222	1.8
13	*rps12-clpP*	4.49	8	178	0
14	*rbcL-accD*	4.38	34	776	2.32
15	*trnP-psaJ*	4.30	16	372	4.03
16	*psbK*	4.23	8	189	0
17	*petN-psbM*	4.15	35	843[a]	3.08
18	*petL* intron	3.97	30	755	2.52
19	*accD*	3.87	57	1473	0.61
20	*ccsA*	3.81	37	971	0
21	*psbN-psbH*	3.81	4	105	0
22	*psbE-petL*	3.69	44	1194	4.02
23	*petL-petG*	3.68	6	163	0
24	*rps11*	3.36	14	417	0
25	*ndhA* 2nd exon	3.34	18	539	0
26	*ndhF*	3.25	73	2247	0
27	*rbcL*	3.18	45	1413	0.64
28	*rps2*	3.09	22	711	0
29	*rpoC1* intron	2.97	15	505[a]	1
30	*matK*	2.90	45	1551	0.77

[a]This specific portion from the start of region was consider instead of complete region to avoid fast evolving region with high indels and inversions

All the analysed species of subfamily Monsteroideae exhibit conserved chloroplast genomes, and show similarities in gene content, intron content, and chloroplast genome organization. Previous studies of other angiosperm plant lineages have demonstrated both conserved (Choi et al. 2016; Li et al. 2018; Shahzadi et al. 2019; Mehmood et al. 2019) as well as highly polymorphic (Menezes et al. 2018; Abdullah et al. 2019a, 2020; Liu et al. 2019a) chloroplast genomes within specific plant lineages. The chloroplast genomes which we report here are highly conserved. In contrast to our results, in *Amorphophallus*, a genus of the subfamily Aroideae in Araceae, certain events of gene loss were

recently reported (Hu et al. 2019; Liu et al. 2019a). However, species of Monsteroideae show conserved chloroplast genomes similar to previous studies (Choi et al. 2017). The *infA* gene is important as a translation initiation factor. This gene has been reported as absent in many species, either fully deleted or non-functional in multiple independent lineages (Jansen et al. 2007; Ahmed et al. 2012; Abdullah et al. 2020). This gene was found non-functional in all species of Monsteroideae. A functional copy of this gene might be present in the nuclear genome (Jansen et al. 2007).

The contraction and expansion in IRs are considered important evolutionary events which change chloroplast genome size and gene content (Menezes et al. 2018; Abdullah et al. 2020). Previously, the expansion of IRs has been reported in subfamily Lemnoideae of family Araceae which led to duplication of *ycf1* and *rps15* genes (Wang and Messing 2011), whereas the duplication of single copy genes or vice versa have also been reported due to IRs contraction and expansion in two species of Aroideae (Araceae) (Henriquez et al. 2020). In the current study, such a high level of IRs expansion has not been observed and the structure of the chloroplast genomes showed similarities with the chloroplast genome structure of other reported species of family Araceae (Ahmed et al. 2012; Choi et al. 2017). A truncated copy of *ycf1* was observed at the junction of IRb and SSC in *S. patulinervum* along with the existence of one functional copy at the junction of SSC/IRa. The *ycf1* pseudogene has also been reported in other angiosperms, including family Araceae (Ahmed et al. 2012; Choi et al. 2017; Yu et al. 2019b; Shahzadi et al. 2019; Abdullah et al. 2019a, 2020; Henriquez et al. 2020).

The analyses of RSCU provide information about the encoding frequency of codon for an amino acid. The codons that have either A or T at their 3′ end showed high encoding efficacy of the amino acid and mostly have RSCU ≥ 1. Conversely, the codons that have C or G at their 3′ end showed low encoding efficacy and mostly have RSCU < 1. Similar results were previously reported in other angiosperms (Shahzadi et al. 2019; Mehmood et al. 2019; Abdullah et al. 2020). In addition to normal ATG start codon which encodes formyl-methionine (Alkatib et al. 2012a), we also observed alternate start codons, including ACG (in *rpl2* gene), ATA (in *cemA* gene), and GTG (in *rps1* gene). The ACG is converted to AUG by RNA editing (Neckermann et al. 1994), whereas the other alternative codons are also reported in the chloroplast genome of other plant species (Sugiura et al. 1998; Su et al. 2019). Usually, 32 tRNAs are required to read all codons of the mRNA (Crick 1966) and chloroplast genome contains up to 30 tRNAs (Menezes et al. 2018; Abdullah et al. 2020). However, Superwobbling can reduce the required number of tRNAs, whereby a single tRNA species containing a Uridine in the wobble position of the anticodon can read an entire fourfold degenerate codon box
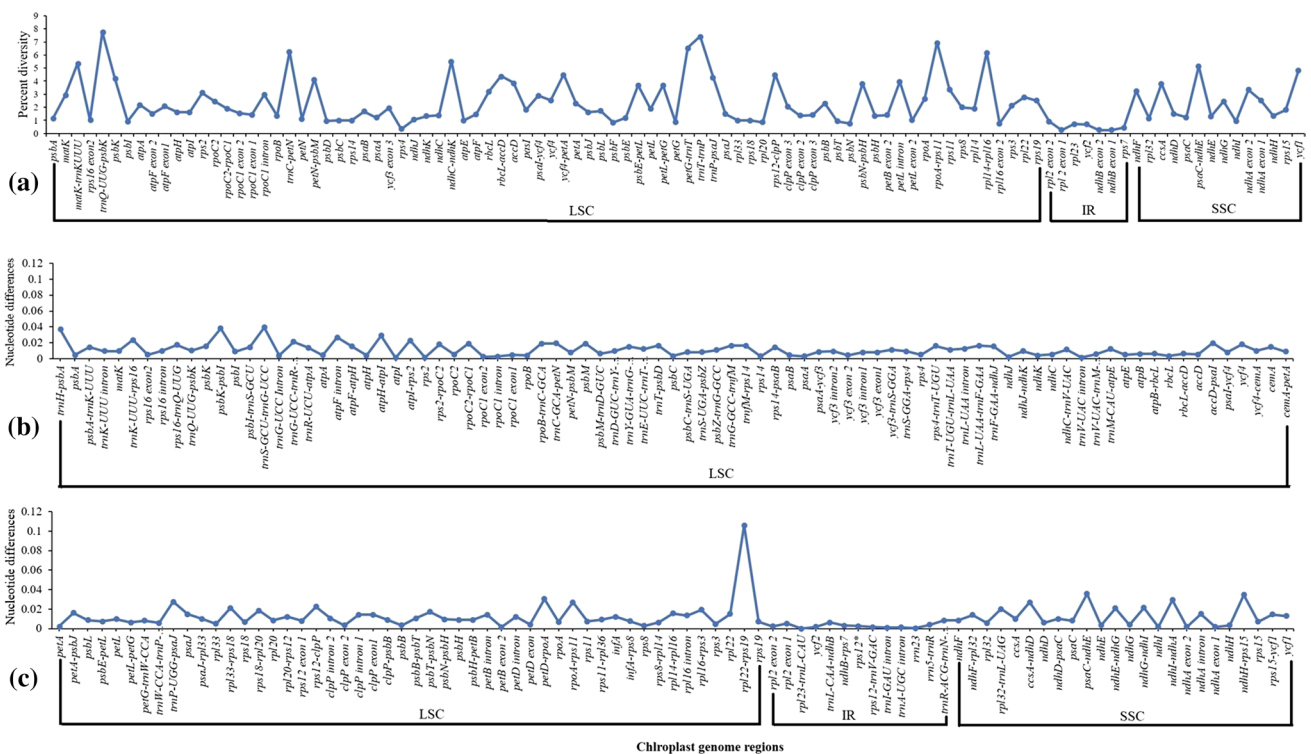
Fig. 5 Suitable polymorphic regions in subfamily Monsteroideae and genus *Spathiphyllum*. **a** The percent diversity of the protein-coding sequences of chloroplast genome regions based on comparison of four species of subfamily Monsteroideae. We included polymorphic protein-coding genes and the conserved intergenic spacer regions and intronic regions. We avoided regions with high indels and inversion events as these are not preferred for inferring of phylogeny at subfamily or family levels. **b**, **c** Nucleotide differences of complete chloro-plast genomes based on comparison of *S. patulinervum* and *S. kochii*. The *X*-axis represents regions of chloroplast genome, whereas the *Y*-axis represents the nucleotide differences. The LSC, IR, and SSC on the *X*-axis indicate those regions which belong to large single copy, inverted repeat, and small single copy, respectively. Conserved regions without nucleotide differences were ignored and were not included in **b** and **c**

(Alkatib et al. 2012b) but also reduce translation efficiency (Rogalski et al. 2008). Moreover, the essential presence of uridine at the wobble position on the gene of tRNA also makes the gene with G on this position not essential for translation (Rogalski et al. 2008). Similar phenomena might exist in Monsteroideae species, enabling only 30 tRNAs to read all the codons in the chloroplast genome. Leucine was the most frequently coded amino acid, whereas cysteine was rarely found. These results are also an agreement with the previous studies of angiosperms (Menezes et al. 2018; Shahzadi et al. 2019; Abdullah et al. 2019a).

The rate of synonymous substitutions (Ks), non-synonymous (Ka) substitutions, and their ratio (Ka/Ks) showed low rates of evolution for all types of protein-coding genes in the chloroplast genomes. The synonymous substitutions were more frequent than the non-synonymous substitutions; Ka/Ks ratio was < 1. These findings agree with various studies of angiosperm chloroplast genomes (Menezes et al. 2018; Shahzadi et al. 2019; Abdullah et al. 2020). However, our study contradicts a recent study of aroid

species in other subfamilies, which reports a higher rate of non-synonymous substitutions compared to synonymous substitutions, and many genes undergoing positive selection (Kim et al. 2019). In agreement with the previous reports in other angiosperms as well as aroids (Choi et al. 2016; Menezes et al. 2018; Piot et al. 2018; Kim et al. 2019; Abdullah et al. 2020), our analyses revealed strong purifying selection on genes which have a role in photo-synthesis. Some genes, including *rbcL*, *psbK*, *accD*, *rps8*, *ndhK*, *ndhD*, and *ycf2,* were found under positive selection which might be due to different types of stresses faced by these species in their respective ecological niches. These genes have also previously been reported to undergo posi-tive selection (Yang et al. 2016; Choi et al. 2018; Yu et al. 2019b; Abdullah et al. 2019a; Kim et al. 2019).

Oligonucleotide repeats play a role in the generation of substitutions, InDels, and inversion (McDonald et al. 2011; Ahmed et al. 2012; Xu et al. 2015; Abdullah et al. 2020). These repeats have been suggested as a proxy for identifi-cation of polymorphic loci (Ahmed et al. 2012, 2013). In

Table 4 Polymorphic regions identified by comparison of *Spathiphyllum patulinervum* and *Spathiphyllum kochii*

| S. no.	Region	Nucleotide differences	No.'s of mutations	Alignment length
1	*trnS-GCU-trnG-UCC*	0.0393	19	483
2	*psbK-psbI*	0.0382	14	366
3	*trnH-psbA*	0.0372	14	376
4	*psaC-ndhE*	0.0357	21	589
5	*atpH-atpI*	0.0290	27	932
6	*trnP-UGG-psaJ*	0.0275	10	363
7	*atpF* intron	0.0267	23	861
8	*trnK-UUU-rps16*	0.0236	19	805
9	*atpI-rps2*	0.0229	6	262
10	*ndhG-ndhI*	0.0214	6	281
11	*ndhE-ndhG*	0.0208	6	289
12	*rpl32-trnL-UAG*	0.0201	21	1044
13	*accD-psaI*	0.0196	17	866
14	*trnC-GCA-petN*	0.0193	17	883
15	*rpoB-trnC-GCA*	0.0190	29	1528
16	*rps18-rpl20*	0.0184	6	326
17	*rps2-rpoC2*	0.0181	5	277
18	*ycf4*	0.0180	10	555
19	*rps16-trnQ-UUG*	0.0175	26	1489
20	*trnT-psbD*	0.0164	19	1159
21	*trnL-UAA-trnF-GAA*	0.0163	6	367
22	*rps4-trnT-UGU*	0.0162	4	247
23	*petA-psbJ*	0.0160	18	1126
24	*atpF-atpH*	0.0156	8	513
25	*trnF-GAA-ndhJ*	0.0154	11	715
26	*ndhA* intron	0.0151	17	1128
27	*rpl22*	0.0150	6	400
28	*cemA*	0.0148	7	477
29	*psbA-trnK-UUU*	0.0144	4	277
30	*rps15-ycf1*	0.0144	6	416

the current study, we reported four types of oligonucleotide repeats, including forward, reverse, palindromic, and complementary repeats. Repeat density was high in IGS and in the LSC region. Most of the repeats ranged between 35 and 44 bp in size. These observations are in agreement with the previous reports (Poczai and Hyvönen 2017; Mehmood et al. 2019; Abdullah et al. 2019a, 2020).

Barcoding and phylogenetic inference can be performed either using complete chloroplast genome or lineage-specific polymorphic loci (Li et al. 2014). Owing to high cost per sequencing of a complete chloroplast genome, the use of lineage-specific polymorphic loci can be a better alternative at times (Li et al. 2014). As all genomic regions are not equally useful for inferring the phylogeny of closely related taxa or resolving taxonomic discrepancies (Daniell et al. 2016), lineage-specific polymorphic regions can serve

these purposes (Ahmed et al. 2013; Ahmed 2014; Li et al. 2014) and have been reported for several species (Li et al. 2018; Menezes et al. 2018; Shahzadi et al. 2019; Abdullah et al. 2019a, 2020). The phylogenetic analysis using a limited number of species in the current study demonstrates the monophyly of subfamily Monsteroideae. Previously, the subfamily Monsteroideae was also identified as monophyletic, however, the species-level phylogeny of subfamily Monsteroideae is not well resolved due to low polymorphism of available molecular markers (Zuluaga et al. 2019). We identified suitable polymorphic regions that might be helpful for designing suitable and unique markers for inferring the phylogeny of subfamily Monsteroideae. In selecting these polymorphic regions we focused on substitution mutations rather than indels, as substitutions are preferred for reconstructing evolutionary history based on maximum-likelihood methods (Ahmed et al. 2013; Ahmed 2014; Menezes et al. 2018; Shahzadi et al. 2019; Abdullah et al. 2019a, 2020). Commonly used molecular markers including *rbcL* and *matK* (Cusimano et al. 2011; Zuluaga et al. 2019) were less polymorphic than many alternatives in our study. The most recent study used *trnC-petN* and partial *ycf1* for inferring the phylogeny in subfamily Monsteroideae (Zuluaga et al. 2019). Our study also suggests that these loci are included among the suitable polymorphic regions. However, use of *ycf1* in phylogenetic reconstruction in Monsteroideae should be done with care and aided by other markers due to large scale and frequent inversions in this gene (Menezes et al. 2018; Zuluaga et al. 2019). The *trnC-petN* locus also showed high incidence of inversions and indels; accordingly, we recommend its partial sequencing. Our findings contradict a recent study (Zuluaga et al. 2019) in which authors reported low polymorphism for *trnT-psbD*, *rps16-trnQ*, *petA-psbJ*, and *psbE-petL* loci. These loci were not included as suitable polymorphic regions in our study at the family-level comparison due to their high level of polymorphism with high production of missing data and not due to low polymorphism.

We also identified polymorphic regions by comparison of two species of genus *Spathiphyllum*. Here, we selected the high polymorphic loci with alignment ≥ 200 bp for designing of suitable polymorphic markers following Abdullah et al. (2019a, 2020) and Shahzadi et al. (2019). For these comparisons, some of the most commonly employed loci, including *matK*, *rbcL*, and *ndhF* (Alverson et al. 1999; Pfeil et al. 2002; Li et al. 2014) were not essentially the most suitable loci for such comparisons, while other commonly used loci including *ndhA intron* and *trnH-psbA* (Li et al. 2014; Tr et al. 2016; Huang et al. 2019) were among the 30 polymorphic regions in our findings. The identified polymorphic regions by genus- and family-level comparisons showed variations. The region which was found to be highly polymorphic and to produce high levels of missing data (> 5%) and found unfit for phylogenetic inference at the family level
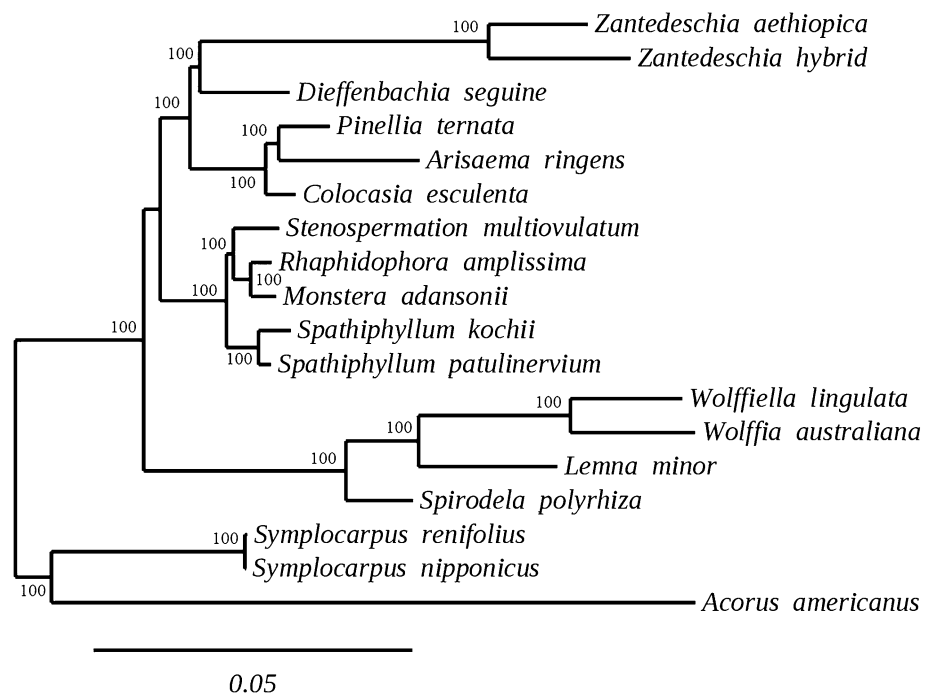
were included in the list of suitable polymorphic loci for phylogenetic inference of genus *Spathiphyllum*. These data also suggest that different levels of polymorphism can be employed at genus (closely related species) and family (more distantly related species) levels for drawing phylogenetic inferences (Menezes et al. 2018; Abdullah et al. 2020).

To conclude, our study provides broad insight into the chloroplast genome structure of subfamily Monsteroideae in which the chloroplast genome of three species was sequenced as first representatives of the genera *Monstera*, *Stenospermation*, and *Rhaphidophora*. Higher synonymous substitutions existed than non-synonymous substitutions and most protein-coding genes showed high purifying selection pressure. The polymorphic regions identified here might be suitable for designing unique and robust markers for inferring the phylogeny and phylogeography among closely related species within the genus *Spathiphyllum* and among distantly related species within the subfamily Monsteroideae.

Compliance with ethical standards

References

Abdullah, Shahzadi I et al (2019a) Comparative analyses of chloroplast genomes among three *Firmiana* species: Identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. Plant Gene 19:100199. https://doi.org/10.1016/J.PLGENE.2019.100199

Abdullah, Waseem S et al (2019b) Comparative analyses of chloroplast genome in *Theobroma cacao* and *Theobroma grandiflorum*. Biologia. https://doi.org/10.2478/s11756-019-00388-8

Abdullah, Mehmood F et al (2020) Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. Genomics 112:581–591. https://doi.org/10.1016/j.ygeno.2019.04.010

Ahmed I (2014) Evolutionary dynamics in taro. PhD dissertation, Massey University, Palmerston North, New Zealand

Ahmed I, Biggs PJ, Matthews PJ et al (2012) Mutational dynamics of aroid chloroplast genomes. Genome Biol Evol 4:1316–1323. https://doi.org/10.1093/gbe/evs110

Ahmed I, Matthews PJ, Biggs PJ et al (2013) Identification of chloroplast genome loci suitable for high-resolution phylogeographic

studies of *Colocasia esculenta* (L.) Schott (Araceae) and closely related taxa. Mol Ecol Resour 13:929–937. https://doi.org/10.1111/1755-0998.12128

Alkatib S, Fleischmann TT, Scharff LB, Bock R (2012a) Evolutionary constraints on the plastid tRNA set decoding methionine and isoleucine. Nucleic Acids Res 40:6713–6724. https://doi.org/10.1093/nar/gks350

Alkatib S, Scharff LB, Rogalski M et al (2012b) The contributions of wobbling and superwobbling to the reading of the genetic code. PLoS Genet 8:e1003076. https://doi.org/10.1371/journal.pgen.1003076

Alverson WS, Whitlock BA, Nyffeler R et al (1999) Phylogeny of the core Malvales: evidence from *ndh*F sequence data. Am J Bot 86:1474–1486. https://doi.org/10.2307/2656928

Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34:3030–3031. https://doi.org/10.1093/bioinformatics/bty220

Andrews S (2017) FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 15 Sept 2019

Bankevich A, Nurk S, Antipov D et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Bi Y, Zhang MF, Xue J et al (2018) Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. Sci Rep 8:1–12. https://doi.org/10.1038/s41598-018-19591-9

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boyce PC, Croat TB (2018) The Überlist of Araceae, totals for published and estimated number of species in aroid genera. https://www.aroid.org/genera/140313uberlist.pdf: Accessed Jan 2020

Cabrera LI, Salazar GA, Chase MW et al (2008) Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. Am J Bot 95:1153–1165. https://doi.org/10.3732/ajb.0800073

Choi KS, Chung MG, Park S (2016) The complete chloroplast genome sequences of three *Veroniceae* species (Plantaginaceae): comparative analysis and highly divergent regions. Front Plant Sci 7:1–8. https://doi.org/10.3389/fpls.2016.00355

Choi KS, Park KT, Park S (2017) The chloroplast genome of *Symplocarpus renifolius* : a comparison of chloroplast genome structure in Araceae. Gene 8:324. https://doi.org/10.3390/genes8110324

Choi KS, Kwak M, Lee B, Park SJ (2018) Complete chloroplast genome of *Tetragonia tetragonioides*: molecular phylogenetic relationships and evolution in caryophyllales. PLoS ONE 13:1–11. https://doi.org/10.1371/journal.pone.0199626

Cooper G (2000) Chloroplasts and other plastids in the cell: a molecular approach, 2nd edn. Sinauer Associates, Sunderland

Crick FHC (1966) Codon-anticodon pairing: the wobble hypothesis. J Mol Biol 19(2):548–555

Cusimano N, Bogner J, Mayo SJ et al (2011) Relationships within the Araceae: comparison of morphological patterns with molecular phylogenies. Am J Bot 98:654–668. https://doi.org/10.3732/ajb.1000158

Daniell H, Lin C-S, Yu M, Chang W-J (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17:134. https://doi.org/10.1186/s13059-016-1004-2

Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704

de Lopes AS, Pacheco TG, dos Santos KG et al (2018) The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within

Malpighiales. Plant Cell Rep 37:307–328. https://doi.org/10.1007/s00299-017-2231-z

Dereeper A, Guignon V, Blanc G et al (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465–W469. https://doi.org/10.1093/nar/gkn180

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Feng Y, Comes HP, Zhou XP, Qiu YX (2019) Phylogenomics recovers monophyly and early Tertiary diversification of *Dipteronia* (Sapindaceae). Mol Phylogenet Evol 130:9–17. https://doi.org/10.1016/j.ympev.2018.09.012

Gunawardena AHLAN, Dengler NG (2006) Alternative modes of leaf dissection in monocotyledons. Bot J Linn Soc 150:25–44. https://doi.org/10.1111/j.1095-8339.2006.00487.x

Henriquez CL, Arias T, Pires JC et al (2014) Phylogenomics of the plant family Araceae. Mol Phylogenet Evol 75:91–102. https://doi.org/10.1016/j.ympev.2014.02.017

Henriquez CL, Abdullah et al (2020) Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). Genomics. https://doi.org/10.1016/j.ygeno.2020.01.006

Hirao T, Watanabe A, Kurita M et al (2008) Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biol. https://doi.org/10.1186/1471-2229-8-70

Hoang DT, Chernomor O, von Haeseler A et al (2018) UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281

Hu H, Liu J, Wang B et al (2019) Characterization of the complete chloroplast genome of *Amorphophallus konjac* (Araceae) and its phylogenetic analysis. Mitochondrial DNA Part B Resour 4:1658–1659. https://doi.org/10.1080/23802359.2019.1606683

Huang Y, Li Z, Wang C et al (2019) *psbE-psbL* and *ndhA* Intron, the promising plastid DNA barcode of *Fagopyrum*. Int J Mol Sci 20:3455. https://doi.org/10.3390/ijms20143455

Iram S, Hayat MQ, Tahir M et al (2019) Chloroplast genome sequence of *Artemisia scoparia*: comparative analyses and screening of mutational hotspots. Plants 8:476. https://doi.org/10.3390/plants8110476

Jansen RK, Cai Z, Raubeson LA et al (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci 104:19369–19374. https://doi.org/10.1073/pnas.0709121104

Kalyaanamoorthy S, Minh BQ, Wong TKF et al (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285

Karol KG, Arumuganathan K, Boore JL et al (2010) Complete plastome sequences of *Equisetum arvense* and Isoetes flaccida: Implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evol Biol. https://doi.org/10.1186/1471-2148-10-321

Katoh K, Kuma KI, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33:511–518. https://doi.org/10.1093/nar/gki198

Kearse M, Moir R, Wilson A et al (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kim S-H, Yang J, Park J et al (2019) Comparison of whole plastome sequences between thermogenic Skunk Cabbage *Symplocarpus renifolius* and nonthermogenic *S. nipponicus* (Orontioideae; Araceae) in East Asia. Int J Mol Sci 20:4678. https://doi.org/10.3390/ijms20194678

Kurtz S, Choudhuri JV, Ohlebusch E et al (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29:4633–4642

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152

Lehwark P, Greiner S (2019) GB2sequin—a file converter preparing custom GenBank files for database submission. Genomics 111:759–761. https://doi.org/10.1016/J.YGENO.2018.05.003

Li B, Zheng Y (2018) Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. Sci Rep 8:9285. https://doi.org/10.1038/s41598-018-27453-7

Li L-F, Wang H-Y, Zhang C et al (2013) Origins and domestication of cultivated banana inferred from chloroplast and nuclear genes. PLoS ONE 8:e80502. https://doi.org/10.1371/journal.pone.0080502

Li X, Yang Y, Henry RJ et al (2014) Plant DNA barcoding: from gene to genome. Biol Rev. https://doi.org/10.1111/brv.12104

Li Y, Zhang Z, Yang J, Lv G (2018) Complete chloroplast genome of seven *Fritillaria* species, variable DNA markers identification and phylogenetic relationships within the genus. PLoS ONE. https://doi.org/10.1371/journal.pone.0194613

Liu E, Yang C, Liu J et al (2019a) Comparative analysis of complete chloroplast genome sequences of four major *Amorphophallus* species. Sci Rep 9:809. https://doi.org/10.1038/s41598-018-37456-z

Liu XF, Zhu GF, Li DM, Wang XJ (2019b) The complete chloroplast genome sequence of *Spathiphyllum cannifolium*. Mitochondrial DNA Part B Resour 4:1822–1823. https://doi.org/10.1080/23802359.2019.1613191

Lohse M, Drechsel O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet 52:267–274. https://doi.org/10.1007/s00294-007-0161-y

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764–770. https://doi.org/10.1093/bioinformatics/btr011

Marcussen T, Meseguer AS (2017) Species-level phylogeny, fruit evolution and diversification history of *Geranium* (Geraniaceae). Mol Phylogenet Evol 110:134–149. https://doi.org/10.1016/j.ympev.2017.03.012

McDonald MJ, Wang WC, Da HH, Leu JY (2011) Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. PLoS Biol. https://doi.org/10.1371/journal.pbio.1000622

McKain MR, McNeal JR, Kellar PR et al (2016) Timing of rapid diversification and convergent origins of active pollination within Agavoideae (Asparagaceae). Am J Bot. https://doi.org/10.3732/ajb.1600198

Mehmood F, Abdullah et al (2019) Characterization of *Withania somnifera* chloroplast genome and its comparison with other selected species of Solanaceae. Genomics. https://doi.org/10.1016/J.YGENO.2019.08.024

Menezes APA, Resende-Moreira LC, Buzatti RSO et al (2018) Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. Sci Rep 8:1–12. https://doi.org/10.1038/s41598-018-20189-4

Nauheimer L, Metzler D, Renner SS (2012) Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. New Phytol 195:938–950. https://doi.org/10.1111/j.1469-8137.2012.04220.x

Neale DB, Sederoff RR (1989) Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in *Loblolly pine*. Theor Appl Genet 77:212–216. https://doi.org/10.1007/BF00266189

Neckermann K, Zeltz P, Igloi GL et al (1994) The role of RNA editing in conservation of start codons in chloroplast genomes. Gene 146:177–182. https://doi.org/10.1016/0378-1119(94)90290-9

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating Maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300

Oldenburg DJ, Bendich AJ (2016) The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication. Curr Genet 62:431–442. https://doi.org/10.1007/s00294-015-0548-0

Palmer JD (1985) Comparative organization of chloroplast genomes. Annu Rev Genet 19:325–354. https://doi.org/10.1146/annurev.ge.19.120185.001545

Pfeil BE, Brubaker CL, Craven LA, Crisp MD (2002) Phylogeny of Hibiscus and the tribe Hibisceae (Malvaceae) using chloroplast DNA sequences of ndhF and the rpl16 intron. Syst Bot 27:333–350. https://doi.org/10.1043/0363-6445-27.2.333

Piot A, Hackel J, Christin PA, Besnard G (2018) One-third of the plastid genes evolved under positive selection in PACMAD grasses. Planta 247:255–266. https://doi.org/10.1007/s00425-017-2781-x

Poczai P, Hyvönen J (2017) The complete chloroplast genome sequence of the CAM epiphyte Spanish moss (*Tillandsia usneoides*, Bromeliaceae) and its comparative analysis. PLoS ONE 12:1–25. https://doi.org/10.1371/journal.pone.0187199

Rabah SO, Shrestha B, Hajrah NH et al (2019) *Passiflora* plastome sequencing reveals widespread genomic rearrangements. J Syst Evol 57:1–14. https://doi.org/10.1111/jse.12425

Rogalski M, Karcher D, Bock R (2008) Superwobbling facilitates translation with reduced tRNA sets. Nat Struct Mol Biol 15:192–198. https://doi.org/10.1038/nsmb.1370

Roper JM, Kellon Hansen S, Wolf PG et al (2007) The complete plastid genome sequence of *Angiopteris evecta* (G. Forst.) Hoffm. (Marattiaceae). Am Fern J 97:95–106. https://doi.org/10.1640/0002-8444(2007)97[95:tcpgso]2.0.co;2

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248

Sabir J, Schwarz E, Ellison N et al (2014) Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. Plant Biotechnol J 12:743–754. https://doi.org/10.1111/pbi.12179

Sanderson MJ, Copetti D, Búrquez A et al (2015) Exceptional reduction of the plastid genome of *Saguaro cactus* (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat. Am J Bot 102:1115–1127. https://doi.org/10.3732/ajb.1500184

Schwarz EN, Ruhlman TA, Sabir JSM et al (2015) Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in Papilionoids. J Syst Evol 53:458–468. https://doi.org/10.1111/jse.12179

Shahzadi I, Abdullah et al (2019) Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae. Genomics. https://doi.org/10.1016/J.YGENO.2019.08.016

Shrestha B, Weng ML, Theriot EC et al (2019) Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus Decaloba. Mol Phylogenet Evol 138:53–64. https://doi.org/10.1016/j.ympev.2019.05.030

Su H-J, Barkman TJ, Hao W et al (2019) Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant Balanophora. PNAS 116:934–943. https://doi.org/10.1073/pnas.1816822116

Sugiura M, Hirose T, Sugita M (1998) evolution and mechanism of translation in chloroplasts. Annu Rev Genet 32:437–459. https://doi.org/10.1146/annurev.genet.32.1.437

Tam S-M, Boyce PC, Upson TM et al (2004) Intergeneric and infrafamilial phylogeny of subfamily Monsteroideae (Araceae) revealed by chloroplast trnL-F sequences. Am J Bot 91:490–498. https://doi.org/10.3732/ajb.91.3.490

Tillich M, Lehwark P, Pellizzer T et al (2017) GeSeq—versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45:W6–W11. https://doi.org/10.1093/nar/gkx391

Tr R, Murali S, Ms F (2016) DNA barcoding of the selected *Artemisia* spp. using the five universal barcodes. Int J Herb Med 4:38–42

Walker BJ, Abeel T, Shea T et al (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9:e112963. https://doi.org/10.1371/journal.pone.0112963

Wang W, Messing J (2011) High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. PLoS ONE. https://doi.org/10.1371/journal.pone.0024670

Wicke S, Müller KF, DePamphilis CW et al (2016) Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. Proc Natl Acad Sci USA 113:9045–9050. https://doi.org/10.1073/pnas.1607576113

Wu CS, Lin CP, Hsu CY et al (2011) Comparative chloroplast genomes of Pinaceae: insights into the mechanism of diversified genomic organizations. Genome Biol Evol 3:309–319. https://doi.org/10.1093/gbe/evr026

Xu J-H, Liu Q, Hu W et al (2015) Dynamics of chloroplast genomes in green plants. Genomics 106:221–231. https://doi.org/10.1016/J.YGENO.2015.07.004

Yang Y, Zhou T, Duan D et al (2016) Comparative analysis of the complete chloroplast genomes of five *Quercus* species. Front Plant Sci 7(959):972. https://doi.org/10.3389/fpls.2016.00959

Yi X, Gao L, Wang B et al (2013) The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. Genome Biol Evol 5:688–698. https://doi.org/10.1093/gbe/evt042

Yu X, Tan W, Zhang H et al (2019a) Complete chloroplast genomes of *Ampelopsis humulifolia* and *Ampelopsis japonica*: molecular structure, comparative analysis, and phylogenetic analysis. Plants 8:410. https://doi.org/10.3390/plants8100410

Yu X, Zuo L, Lu D et al (2019b) Comparative analysis of chloroplast genomes of five *Robinia* species: genome comparative and evolution analysis. Gene 689:141–151. https://doi.org/10.1016/J.GENE.2018.12.023

Zeb U, Dong W, Zhang T et al (2019) Comparative plastid genomics of Pinus species: Insights into sequence variations and phylogenetic relationships. J Syst Evol. https://doi.org/10.1111/jse.12492

Zhai W, Duan X, Zhang R et al (2019) Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. Mol Phylogenet Evol 135:12–21. https://doi.org/10.1016/j.ympev.2019.02.024

Zhang Y, Ma J, Yang B et al (2014) The complete chloroplast genome sequence of *Taxus chinensis* var. mairei (Taxaceae): loss of an inverted repeat region and comparative analysis with related species. Gene 540:201–209. https://doi.org/10.1016/j.gene.2014.02.037

Zuluaga A (2015) Testing the monophyly of *Spathiphyllum*, and the relationship between Asian and tropical American species. Aroideana 38(107):115

Zuluaga A, Llano M, Cameron K (2019) Systematics, biogeography, and morphological character evolution of the Hemiepiphytic subfamily Monsteroideae (Araceae). Ann Missouri Bot Gard 104:33–48. https://doi.org/10.3417/2018269