Comparative chloroplast genomes of Annonaceae species: Enlargement of plastomes in size, IR region, and gene content

Yangying Gan
Institute of Agricultural Economics and Information, Guangdong Academy of Agricultural Sciences, Key Laboratory of Urban Agriculture in South China, Ministry of Agriculture and Rural Affairs

Xiaojing Liu
Institute of Agricultural Economics and Information, Guangdong Academy of Agricultural Sciences, Key Laboratory of Urban Agriculture in South China, Ministry of Agriculture and Rural Affairs

Jingyao Ping (pingjnyao@foxmail.com)
College of Life Sciences, Sun Yet-sen University

Caixia Peng
Horticulture Center, South China Botanical Garden, Chinese Academy of Sciences

Research Article

Keywords: Annonaceae, plastomes, IR expansion, inversion, SSRs

Posted Date: November 29th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2303224/v1

Abstract

In recent years, the evolution and phylogeny of plastid genomes have attracted much attention from scholars all over the world. Annonaceae is the largest family in Magnoliales with the greatest diversity among and within genera. Thus comparative analyses of its plastomes will be informative. In this study, the complete chloroplast genome of *Miliusa glochidioides* was sequenced by next-generation sequencing technology. Using two Magnoliaceae species as out-group, a comparative analysis of the existing 13 Annonaceae plastomes was conducted, and a phylogenetic relationship was constructed based on four methods. Our results show that the Annonaceae plastomes have great variation in the evolution process. The genome size is between 159kb and 202kb while the gene content ranges from 127 to 165. The number of genes in the IR region is between 5 and 39. Plastomes underwent significant structural rearrangements, including one inversion and multiple large-scale expansion (6-20kb) in the IR region, and shrinkage and inversions in the SSC region. Compared with Magnoliaceae, two fragments (*pafII-atpE* and 5'-*rps12-psbJ*) were inversed in Annonaceae. The phylogenetic relationship based on 78 common protein genes showed that *Cananga odorata* was located at the base of Annonaceae. Annonoideae was a monophyletic group, and *Chieniodendron hainanense* was located inside the Malmeoideae. A total of 737 simple sequence repeats (SSRs) were detected in the study, and the distribution of SSRs varied from species. It is mainly dominated by A/T bases (mononucleotide) and is located in the intergenic region. In conclusion, the plastomes of Annonaceae have undergone great variation during evolution, especially the large-scale expansion of the IR region. Our study provides more information for studying the plastome evolution of Annonaceae.

Introduction

Annonaceae is the largest pantropical family of trees and lianas in the early-divergent order of Magnoliales, with 108 genera and about 2,300 species (Chatrou et al., 2012; Guo et al., 2017). It is the most diverse family and contributes significantly to tree diversity in rain forests around the world (Punyasena et al., 2008; Couvreur et al., 2011; Gan et al., 2015; Gan & Xu, 2018; Lei et al., 2022). However, its remarkable diversity has also brought great difficulty and controversy to its phylogenetic research (Chaowasku, et al., 2018; Chaowasku, et al., 2015; Guo et al., 2017; Maas et al., 2015; Ortiz-Rodriguez et al., 2018; Saunders et al., 2020; Xue & Tan, 2016; Xue et al., 2021; Xue et al., 2020).

In recent years, Plastid genomes (plastomes) are used more frequently in phylogenetic evolution studies. Compared with the nuclear genome, it has small genome size, large copy number and ease to obtain, and compared with plastid DNA markers it is more informative (Li, Hu et al., 2021; Raubeson & Jansen, 2005; Tonti-Filippini et al., 2017). Moreover, it has been found that the evolution of plastome was often accompanied by gene loss or pseudogenization (Chaw et al., 2018; Li, Zhao et al., 2021; Ruhlman & Jansen, 2014), inversion of sequence fragments (Chaw et al., 2018; Ping, Feng, Li, et al., 2021; Raubeson and Jansen, 1992), or expansion and contraction of IR regions (Cauz-Santos et al., 2020; Mower & Vickrey, 2018; Ping Li et al., 2021; Zhu et al., 2016). Thus, by comparing features such as genome size, gene capacity, boundary displacement, and structural rearrangement, it is possible to shed new light on phylogenetic studies and provide rich information for the evolution process of plastomes. Up till now, however, only 13 complete plastid genomes of Annonaceae have been published, one of which has not yet been confirmed, leading to very limited knowledge of the chloroplast genome of Annonaceae.

Miliusa glochidioides Handel-Mazzetti Sinensia is an Annonaceae evergreen shrub distributed in the forested slopes below 900 meters in Guangxi and is endemic to Guangxi, China. In this study, we sequenced the complete chloroplast genomes of *Miliusa glochidioides* and conducted a comparative analysis of the existing 13 plastomes of Annonaceae, expecting to find some evidence for the reconstruction of phylogenetic relationship of Annonaceae and reveal the evolutionary laws of chloroplast genome.

Materials And Methods

Sampling and sequencing

Fresh leaves of *Miliusa glochidioides* were collected in South China Botanical Garden, Chinese Academy of Sciences (E113°36', N23°18'). Total DNA was extracted using E.Z.N. A.® Soil DNA Kit (OMEGA). Illumina NovaSeq6000 platform was used to carry out pair-end sequencing with a 150 bp reading length. The sequences were spliced using the stitching software NOVOPlasty V4.2 (Dierckxsens et al., 2017) and GetOrganelle V1.7.0 + (Jin, Yu et al., 2020) for the optimal assembly results. Combined with software PGA (Qu et al., 2019) and Geseq (Tillich et al., 2017), the assembled plastome was annotated. The *M. glochidioides* complete plastome was uploaded to the NCBI database through the BankIt platform with accession number OM047203.

Sequence data

The complete plastomes sequences were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/nuccore/?term=), including 12 Annonaceae and 2 Magnoliaceae (as out-group), using the online website GeSeq-Annotation of Organellar Genomes (https://chlorobox.mpimp-golm.mpg.de/geseq.html) to re-annotate the genome and the online site OGDRAW-Draw Organelle Genome Maps (https://chlorobox.mpimp-golm.mpg.de/ OGDraw.html) to map the plastomes (Greiner et al., 2019). The sequences were imported into Genious Prime 2022.0.1 (Kearse et al., 2012) for statistical genome structure information and screening of common genes.

Construction of phylogenetic relationships

Based on the tandem dataset of common genes, a Neighbor-Joining (NJ) tree was constructed with MEGA7.0 software (Kumar et al., 2018); the PAUP 4.0 software was to construct Maximum-Parsimony (MP) trees (Swofford, 2002); Maximum-Likelihood (ML) trees were constructed using RaxmlGUI2 (Stamatakis, 2014) software with 1000 bootstrap and GTRGAMMAI; Mrbayes v3.2.0 (Huelsenbeck & Ronquist, 2001) software was used to construct Bayesian-Inference (BI) trees.

Detection of SSRs

MISA (Microsatellite Identification Tool) online website (https://webblast.ipk-gatersleben.de/misa/index.php?action=1) predicts SSRs, the parameters are set to default settings, SSR motif length corresponds to the minimum number of repetitions are 1-10, 2-6, 3-5, 4-3, 5-3, 6-3. When the distance between two SSRs is less than 100 bp, it can be used as a compound SSR (Beier et al., 2017).

Results

Structural character of the plastomes

The *Miliusa glochidioides* plastome is 159,789bp in size, with a typical quadripartite structure that consists of four parts: a large single copy (LSC) region (88,782bp), a small single copy (SSC) region (18,949bp), and two inverted repeats (IRs) regions (both are 26,029bp), encoding a total of 129 genes including 84 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. The GC content is 36.7% (Figure 1).

The nucleotide sequences of the 13 plastid genomes range from 159kb to 202kb. The overall GC content is 38.7- 39.6%. The number of coding genes is 127-165, and the number of protein-coding genes is 84-116. The eight species of Annonoideae have the largest plastomes (178.1-201.7kb), IR region (42.1- 64.6kb), SSC region (2.9-3.7kb) and the largest number of genes (140-165), protein-coding genes (94-116) and tRNAs (36-41). The GC content of the IR region is higher than that of the single-copy region in all other species. The *Annona* has the largest plastomes and IR region, corresponding to the smallest LSC region and largest gene content (Table 2).

All species encode a total of 81 protein-coding genes, of which 78 are shared (Table S1). The *accD* is lost in *A. reticulata* and *A. muricata*; *ycf1*5 exists only in *F. polyanthum*, *U. macrophylla*, *A. pilosus*, *M. alba*, and *L. chinense*; *ycf68* exists only in *F. polyanthum* and *A. pilosus*. The protein-coding gene content of the IR region varies in different species (Figure 2). There are 37 protein-coding genes in the IR region of *A. cherimola* and *A. reticulata*, including part of *psbA* that entered the IR region, 32 in *A. muricata*, 25 in *Fissistigma* and *Uvaria*, 15 in *Artabotrys*, 14 in *Cananga*, and five in the other four species. In the out-group (the red branch), there are seven genes in the IR region, including part of *ycf1* that entered the IR region. The IR region of all species contains five protein-coding genes (*rpl2*, *rpl23*, *ycf2*, *ndhB*, and *rps7*) and 3'-*rps12*.

Structural rearrangement of the plastomes

The structures of the plastomes of the two Magnoliaceae species are completely identical, and the IR region contained *rpl2-trnN* and a partial sequence fragment of *ycf1* (Figures 2 and 3). In Annonaceae, *ycf1* is completely located in the SSC region in the basal taxa, and in the IR region in Annonoideae. There are three species (P. verrucipes, *M. glochidioides*, *G. suaveolens*) contain *rpl2-trnN* fragment only in the IR region, while in *C. hainanense* the fragment is inversed and in the remaining species the IR regions are significantly expanded.

Compared with Magnoliaceae, there are two inversions (*pafII-atpE* and 5'-*rps12-psbJ*) in Annonaceae and large-scale expansion of the IR region in several species. The IR region of the Annonoideae expands to SSC by 16kb. The extent of IR region expansion to LSC region is different in different species. Among them, *A. cherimola* and *A. reticulata* expand by 20kb. *A. muricata* expand by 14kb, *F. oldhamii*, *F. polyanthum*, *U. macrophylla* and *C. odorata* expand by 6kb. *Annona* expanded by about 36kb. Compared with other species in Annonaceae, the *ndhF-rpl32* fragment is an inversion in *A. pilosus*, *F. oldhamii*, *A. reticulata*, and *A. muricata* (Figure 3). For the first time, we observed that the 5'-*rps12* was completely located in the IR region. Since the IR region could not be identified in the plastome of *G.*

suaveolens, there might be some errors in the sequence. Therefore, in this study, the sequence of *G. suaveolens* was only used as a reference.

Phylogenetic relationships of sampled species

With *M. alba* and *L. chinense* as the out-group, the phylogenetic relationship constructed by the maximum likelihood method and Bayesian method was highly supported, and the support of all branches was 100 (Bootstrap Support) or 1 (Posterior Probability). The results support that *C. odorata* as the base of the Annonaceae, and the Annonoideae as a monophyletic branch. *Artabotrys* at the base of Annonoideae, and *Annona* as a sister group of *Fissistigma*+*Uvaria*. *C. hainanense* was located inside the Malmeoideae (Figure 4a).

Distribution pattern of SSRs

We detected a total of 737 SSRs comprising 6 nucleotide types and 13 motif types in the sampled species, which means each species has an average of SSRs ranging from 36 to 77 (Figure 4b, Appendices S2 and S3). The nucleotide types were dominated by mononucleotide SSRs (83%-98%), followed by dinucleotide (0-10%). And the minimum number of tetranucleotide SSRs (only 3) was found in *A. cherimola*, *C. hainanense*, and *C. odorata* (Table S2). The motif types were dominated by A/T base (83%-98%), which existed in all species. AT/AT followed and was present in all species except *M. glochidioides*. *C. odorata* had the largest number of SSRs, including 5 nucleotide types and 6 motif types. A. Hexapetalus contains only A or T bases SSRs (Table S3).

SSRs were mainly located in IGS (64%-86%), but some of them were located in coding sequences (3%-25%) such as *atpB*, *ndhF*, *rps15*, *rps18*, *rps19*, *rpoA*, *rpoB*, *rpoC2*, *infA*, *ycf1*, and *ycf2*, or introns (7%-22%) such as *atpF*, *ndhA*, *rpl16*, *rps16*, *pafI*, *clpP*, *trnI-GAU*, *trnL-UAA*, *trnV-UAC*, and *trnG-UCC* (Figure 4c, Appendices S3). AT/AT was found in the intron of atpF in *A. cherimola*, *A. reticulata*, and *P. verrucipes*.

We also counted the number of SSRs in four partitions on the genome (Figure 5). The number of the SSRs located in the LSC region was the highest (471, 63.9%), followed by the IR region (208, 28.2%) and the SSC region (58, 7.9%). In genus *Annona*, the number of the SSRs located in the IR region were higher than or equal to the that in the LSC region. In other species, the number of the SSRs located in the LSC region is the highest. In Annonoideae, the number of SSRs located in the SSC region is the minimum because of the sharp reduction of this region (0-4). It is obvious that the SSRs in the IR region of the species with IR expansion are much higher than those of other species (16-28). In addition, we found that the inversion fragments (*pafII-atpE* and 5'-*rps12-psbJ*) have 16 SSRs in *C. odorata*, which is much higher than in other species (4-11). Four species (*A. reticulate*, *A. muricata*, *F. oldhamii*, *A. pilosus*) have inversion fragment (*ndhF-rpl32*) but the number of SSRs located in the fragment was not increased.

Discussion

Enlarged plastomes and IR regions

We observe the gradually enlarged plastomes of Annonaceae, from 159kb to 201.9kb. The plastomes size of Annonoideae exceeds 160kb, and the size of the IR region is also between 42kb and 64kb. This is well beyond the size of the IR region of most land plants (15kb to 30kb) (Zhu et al., 2016). The enlargement of Annonoideae plastomes is mainly related to the expansion of the IR region, a single IR region expanded by about 20kb-30kb. Although the SSC region was reduced and contained only two protein-coding genes (*ndhF* and *rpl32*), it was reduced by only about 15kb. In addition, we note that the expansion of the IR region in the Annonaceae is dynamic. In the basal group (*Cananga*), only 6kb was extended to LSC. In Annonoideae, they expanded 20kb, 14kb or 6kb to the LSC region, and 16kb to the SSC region. These suggest that the expansion of the IR region occurred in the early evolution of Annonaceae, but the mechanism of expansion might be different. Large expansions of IR regions have been observed in multiple lineages (Chumley et al., 2006; Dugas et al., 2015; Zhu et al., 2016). Some studies suggested that IR region expansion was related to the Poly A region (Goulding et al., 1996; Dugas et al., 2015). Unlike the massive expansion of IR regions, IR regions are lost in the plastomes of some plant lineages and algae (Cai et al., 2017; Cauz-Santos et al., 2020; Jin, Wicke et al., 2020; Karnkowska et al., 2018; Ruhlman et al., 2017).

Changes in gene content

Due to the large expansion of the IR region, there was a large variation in the gene content of Annonaceae (127-165). In Annonoideae, 10 to 32 protein-coding genes enter the IR region in Annonoideae, which directly increases the total number of genes by 20 to 64. Notably, in *Annona*, 5'-*rps12* enters the IR region, which is rarely reported before. In previous studies, it has been found that 3'-*rps12* enters or leaves

the IR region as the IR region expands or contracts, and it has been shown that 3'-*rps12* in the IR region has reduced substitution rates and a more conserved sequence signature (Ping, Li et al., 2021; Ping, Feng, Hao et al., 2021). Here in *Annona*, 5'-*rps12* also enters the IR region, and whether its substitution rates also changes needs to be further explored.

Although the gene content is variable, the types of protein-coding genes are very conserved. Among the 15 sample species, only *A. reticulata* and *A. muricata* lost the *accD*, and the loss may have occurred independently. In previous studies, the loss or pseudogenization of the *accD* has been reported in angiosperms (*Pelargonium × hortorum* and *Corydalis*) and gymnosperms (*Sciadopitys verticillata*) (Chumley et al., 2006; Chaw et al., 2018). Li et al (2016) found the lost *accD* in *Sciadopitys verticillata* may be complemented by corresponding functions in the nucleus. Li et al (2018) found the presence of some repetitive elements in the *accD* of Cupressophytes could result in an increase in gene length, acceleration of substitution rates, as well as transposition to a rearrangement hotspot region. Similarly, insertion of specific tandem repeats in *accD* was detected in *Tsuga chinensis* and resulted in an increase in gene length (Sudianto et al., 2016).

Inversion event in Annonaceae

Compared with Magnoliaceae, there are two more inversion fragments (*pafII-atpE* and 5'-*rps12-psbJ*) in the plastome of Annonaceae, and more SSRs in the basal group (*Cananga*). From our results, the repetitive sequences seem to play an important role in the structural rearrangement in the early evolutionary period of plastomes. During the evolution of the plastomes in the 13 species of Annonaceae, two regions (the IR region and the ndhF-rpl 32 region) have undergone inversions. And the inversion of IR region was only found in *C. hainanense*, which indicate that the event occurs independently. In addition, four species of Annonoideae have inversions in the SSC region, which is different from the other four species, even they were from the same genus. It is suggested that in the early evolutionary period of Annonoideae (or in the ancestors), the SSC region has been inverted, and along with the evolution process, it is randomly distributed in some groups (Figure 3). Inversion seems to occur frequently during the evolution process. Raubeson & Jansen (1992) found the presence of an inversion of approximately 35kb in the large single-copy region of lycophytes was a strong evidence for its differentiation from the true-leaf plants. Chaw et al. (2018) found during the evolution of gymnosperms, structural rearrangements among different taxa were dominated by inversions.

Phylogenetic relationship of Annonaceae

Due to the great diversity of Annonaceae, their phylogenetic studies have always been controversial (Chatrou et al., 2018; Maas et al., 2015; Mols 2004; Ortiz-Rodriguez et al., 2018; Sauquet et al., 2003). In recent years, molecular evidence has been successfully used to solve some fuzzy classification problems (Guo et al., 2017; Xue et al., 2020; Thomas et al., 2012). In this study, we constructed the phylogenetic relationship of 13 species from Annonaceae based on chloroplast common protein-coding genes. Our results support Annonoideae form a monophyletic branch, *C. odorata* as the base of the Annonaceae, and *C. hainanense* locate inside the Malmeoideae. Among them, *C. hainanense* was classified as a separate category in the NCBI database. This is in consistence with the results of Thomas et al., (2012), who incorporated *Chieniodendron* into *Meiogyne* (Malmeoideae). However, till now, the reported chlorpoplast whole genes of Annonaceae is too limited. More data needs to be added and combined with morphological features to clarify the relationship and evolution process of Annonaceae.

Diversity of SSRs distribution patterns

SSRs are often used to study polymorphisms or as molecular markers (Chmielewski et al. 2015; Fasanella et al. 2020; Huang et al. 2019). Their distribution in the genome of one or several species are often simply described or statistical analyzed, but are seldom used in phylogenetic analysis (Raman et al., 2020; Li, Hu et al., 2021; She et al., 2022). In present study, we analyzed the distribution of SSRs in 13 plastomes of Annonaceae systematically. Our results show that the distribution of SSRs has a significant interspecific difference. In consistence with previous studies, the SSRs are dominated by A/T motifs, and most are located in the IGS region (Gui et al. 2020; Li, Zhao et al. 2021; Ping, Feng, Hao, et al. 2021). However, in some other plants such as Polypodiaceae, the SSRs are dominated by C/G mononucleotides, which were presumed to be an adaptation to rhe environment (Gao et al. 2018; Liu et al. 2021; Schneider et al. 2004). Due to the expansion of the IR region, the distribution of SSRs in the three regions was different in each species. In previous study, however, the distribution of SSRs was reported to be lineage-specific. Ping, Feng, Hao et al. (2021) found the distributions of SSRs in *Cupressus* and *Hesperocyparis* were highly uniform, and suggested that *Callitropsis funebris* was closer to *Cupressus* than other

Callitropsis species. Recently, Zhu et al. (2021) found that the number, type, and localization of SSRs in the plastomes of Cyatheaceae were inter-genus specific, which provide valid evidence for the phylogenetic analysis of Cyatheaceae.

Conclusion

Based on the comparative analysis of the complete chloroplast genomes of 13 species from Annonaceae, we found that during the evolution of Annonaceae, the plastomes had great variation, including genome size, gene content, and the massive expansion of the IR region, etc. However, till now, the reported chlorpoplast whole genes of Annonaceae is still limited, and it is necessary to obtain more plastomes to better understand the evolutionary process of the plastomes of the Annonaceae. At the same time, in the species with the expansion of the IR region, a large number of genes enter the IR region. The evolution mechanism of these genes also need to be revealed.

Declarations

Acknowledgments

We thank professor Jingping Liao and Chunfen Xiao for their assistance with material collection. This work was funded by the National Natural Science Foundation of China (Grant number 31800184).

Author contributions

Yangying Gan designed the study and wrote the manuscript. Jingyao Ping analyzed the data and revised the manuscript. Xiaojing Liu and Caixia Peng analyzed some of the data and revised the manuscript. All authors read and contributed to the final version of the manuscript.

Data Accessibility Statement

Data source is NCBI database: https://www.ncbi.nlm.nih.gov/nuccore/KU563738, MT742547, MT742546, MW136266, MW829282, MH992130, MZ936420, OK216144, MW018366, OM047203, MH924590, MK035708, MN016933, NC_037005, NC_030504

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics*, 33(16), 2583-2585. https://doi.org/10.1093/bioinformatics/btx198.
2. Cai, C., Wang, L. K., Jiang, T., Zhou, L. J., He, P. M. & Jiao, B. H. (2017). Complete chloroplast genome of green tide algae *Ulva flexuosa* (Ulvophyceae, Chlorophyta) with comparative analysis. *PloS One*, 12(9), e0184196. https://doi.org/10.1371/journal.pone.0184196.
3. Cauz-Santos, L. A., da Costa, Z. P., Callot, C., Cauet, S., Zucchi, M. I., Bergès, H., van den Berg, C., & Vieira, M. L. C. (2020). A repertory of rearrangements and the loss of an inverted repeat region in *Passiflora* chloroplast genomes. *Genome biology and evolution*, 12(10), 1841-1857. https://doi.org/10.1093/gbe/evaa155.
4. Chatrou, L. W., Turner, I. M., Klitgaard, B. B., Maas, P. J. & Utteridge, T. (2018). A linear sequence to facilitate curation of herbarium specimens of Annonaceae. *Kew bulletin*, 73(3), 1-10. https://doi.org/10.1007/s12225-018-9764-3.
5. Chatrou L. W., Pirie M. D., Erkens R. H. J., Couvreur T. L. P., Neubig K. M., Abbott J. R., Mols J. B., Maas J. W., Saunders M. K., & Chase M. W. (2012). A new subfamilial and tribal classification of the pantropical plant family Annonaceae informed by molecular phylogenetics. *Botanical Journal of the Linnean Society*, 169 (1), 5-40.
6. Chaw, S. M., Wu, C. S., & Sudianto, E. (2018). Evolution of gymnosperm plastid genomes. *Advances in Botanical Research*, 85, 195-222. https://doi.org/10.1016/bs.abr.2017.11.018.
7. Chaowasku, T., Damthongdee, A., Jongsook, H., Nuraliev, M. S., Ngo, D., Le, H. T., Lithanatudom, P., Osathanunkul, M., Deroin, T., Xue, B**.,** & Wipasa, J. (2018). Genus *Huberantha* (Annonaceae) revisited: erection of *Polyalthiopsis*, a new genus for *H. floribunda*, with a new combination *H. luensis*. *Annales Botanici Fennici*, 55, 121–136.

8. Chaowasku, T., Johnson, D. M., van der Ham, R. W. J. M., & Chatrou, L. W. (2015). *Huberantha*, a substitution name for *Hubera* (Annonaceae-Malmeoideae-Miliuseae). *Kew Bulletin*, 70: 23. doi 10.1007/S12225-015-9571-Z.
9. Chmielewski, M., Meyza, K., Chybicki, I. J., Dzialuk, A., Litkowiec, M. & Burczyk, J. (2015). Chloroplast microsatellites as a tool for phylogeographic studies: The case of white oaks in Poland. *Iforest-Biogeosciences and Forestry*, 8(6), 765. https://doi.org/10.3832/ifor1597-008.
10. Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L. & Jansen, R. K. (2006). The complete chloroplast genome sequence of *Pelargonium × hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular biology and evolution*, 23(11), 2175-2190. https://doi.org/10.1093/molbev/msl089.
11. Couvreur, T. L., Pirie, M. D., Chatrou, L. W., Saunders, R. M., Su, Y. C., Richardson, J. E. & Erkens, R. H. (2011). Early evolutionary history of the flowering plant family Annonaceae: Steady diversification and boreotropical geodispersal. *Journal of Biogeography*, 38(4), 664-680. https://doi.org/10.1111/j.1365-2699.2010.02434.x.
12. Dierckxsens, N., Mardulyn, P. & Smits, G. (2017). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic acids research*, 45(4), e18-e18. https://doi.org/10.1093/nar/gkw955.
13. Dugas, D. V., Hernandez, D., Koenen, E. J., Schwarz, E., Straub, S., Hughes, C. E., Jansen, R. K., Nageswara-Rao, M., Staats, M., Trujillo, J. T., Hajrah N. H., Alharbi, N. S., Al-Malki, A. L., Sabir, J. S. & Bailey C. D. (2015). Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions and accelerated rate of evolution in *clpP*. *Scientific reports*, 5(1), 1-13. https://doi.org/10.1038/srep16958.
14. Fasanella, M., Souto, C. P. & Premoli, A. C. (2020). Preliminary cross-genera transferability of SSRs among threatened South American Cupressaceae. *New Zealand Journal of Botany,* 58(2), 153-166. https://doi.org/10.1080/0028825X.2019.1685552.
15. Gan, Y. Y., Liu, Y., Xu, F. X. (2015). Pollen morphology of selected species from Annonaceae. Grana, 54(4), 271-281.
16. Gan, Y. Y., Xu, F. X. (2018). The coexistence of bicellular and tricellular pollen. *Grana*, 58(2), 129-132.
17. Gao, R., Wang, W. Z., Huang, Q. Y., Fan, R. F., Wang, X., Feng, P., Zhao, G. M., Bian, S., Ren, H. L. & Chang, Y. (2018). Complete chloroplast genome sequence of *Dryopteris fragrans* (L.) Schott and the repeat structures against the thermal environment. *Scientific reports*, 8(1), 1-11. https://doi.org/10.1038/s41598-018-35061-8.
18. Goulding, S. E., Wolfe, K. H., Olmstead, R. G. & Morden, C. W. (1996). Ebb and flow of the chloroplast inverted repeat. *Molecular and General Genetics MGG*, 252(1), 195-206. https://doi.org/10.1007/BF02173220.
19. Greiner, S., Lehwark, P. & Bock, R. (2019). Organellar Genome DRAW (OGDRAW) version 1.3. 1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic acids research*, 47(W1), W59-W64. https://doi.org/10.1093/nar/gkz238.
20. Gui, L., Jiang, S., Xie, D., Yu, L., Huang, Y., Zhang, Z. & Liu, Y. (2020). Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution. *Gene*, 732, 144355. https://doi.org/10.1016/j.gene.2020.144355.
21. Guo, X., Tang, C. C., Thomas, D. C., Couvreur, T. L., & Saunders, R. M. (2017). A mega-phylogeny of the Annonaceae: Taxonomic placement of five enigmatic genera and support for a new tribe, Phoenicantheae. *Scientific Reports,* 7(1), 1-11. https://doi.org/10.1038/s41598-017-07252-2.
22. Huang, C. Y., Yin, Q. Y., Khadka, D., Meng, K. K., Fan, Q., Chen, S. F. & Liao W, B. (2019). Identification and development of microsatellite (SSRs) makers of *Exbucklandia* (Hamamelidaceae) by high-throughput sequencing. *Molecular Biology Reports*, 46, 3381-3386. https://doi.org/10.1007/s11033-019-04800-z.
23. Huelsenbeck, J. P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754-755. https://doi.org/10.1093/bioinformatics/17.8.754.
24. Jin, D. M., Wicke, S., Gan, L., Yang, J. B., Jin, J. J. & Yi, T. S. (2020). The loss of the inverted repeat in the putranjivoid clade of Malpighiales. *Frontiers in plant science*, 942. https://doi.org/10.3389/fpls.2020.00942.
25. Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., DePamphilis, C. W., Yi, T. S. & Li, D. Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome biology*, 21(1), 1-31. https://doi.org/10.1186/s13059-020-02154-5.
26. Karnkowska, A., Bennett, M. S. & Triemer, R. E. (2018). Dynamic evolution of inverted repeats in Euglenophyta plastid genomes. *Scientific reports*, 8(1), 1-10. https://doi.org/10.1038/s41598-018-34457-w.
27. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649. https://doi.org/10.1093/bioinformatics/bts199.

28. Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, 35(6), 1547. https://doi.org/10.1093/molbev/msy096.
29. Lei, J. M., Liang, Z. R., Zhang, H., Lim Y. Q. & Xue, B. E. (2022). Advances in exploitation and utilization of wild fruit resources of Annonaceae in China (in Chinese). *Journal of Fruit Science*, 39(1), 121-130. https://doi.org/10.13925/j.cnki.gsxb.20210301.
30. Li, P. T. & Gilbert, M. G. (2011). Annonaceae. in: Wu, Z.Y., Raven, P.H., & Hong, D.Y., (eds), Flora of China. Science Press, Beijing & Missouri Botanical Garden Press, St. Louis, pp. 672–713.
31. Li, J., Gao, L., Chen, S. S., Tao, K., Su, Y. J. & Wang, T. (2016). Evolution of short inverted repeat in cupressophytes, transfer of *accD* to nucleus in *Sciadopitys verticillata* and phylogenetic position of Sciadopityaceae. *Scientific reports*, 6(1), 1-12. https://doi.org/10.1038/srep20934.
32. Li, J., Su, Y. J. & Wang, T. (2018). The repeat sequences and elevated substitution rates of the chloroplast *accD* gene in cupressophytes. *Frontiers in plant science*, 9, 533. https://doi.org/10.3389/fpls.2018.00533.
33. Li, L., Hu, Y. F., He, M., Zhang, B., Wu, W., Cai, P, M., Huo, D., & Hong, Y, C. (2021). Comparative chloroplast genomes: Insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. *BMC Genomics*, 22, 138. https://doi.org/10.1186/s12864-021-07427-2.
34. Li, X. P., Zhao, Y. M., Tu, X. D., Li, C. R, Zhu, Y. T., Zhong, Z. J., Wu, S. S. & Zhai, J. W. (2021). Comparative analysis of plastomes in Oxalidaceae: Phylogenetic relationships and potential molecular markers. *Plant diversity*, 43(4), 281-291. https://doi.org/10.1016/j.pld.2021.04.004.
35. Liu, S. S, Wang, Z., Su, Y. J & Wang, T. (2021). Comparative genomic analysis of Polypodiaceae chloroplasts reveals fine structural features and dynamic insertion sequences. *BMC Plant Biology,* 21(1), 1-15. https://doi.org/10.1186/s12870-020-02800-x.
36. Maas, P. J. M., Westra, L. Y. T., Guerrero, S. A., Lobão, A. Q., Scharf, U., Zamora, N. A., & Erkens, R. H. J. (2015). Confronting a morphological nightmare: Revision of the Neotropical genus *Guatteria* (Annonaceae). *Blumea-Biodiversity, Evolution and Biogeography of Plants*, 60(1-2), 1-219. https://doi.org/10.3767/000651915X690341.
37. Mols, J. B., Gravendeel, B., Chatrou, L. W., Pirie, M. D., Bygrave, P. C., Chase, M. W. & Keßler, P. J. (2004). Identifying clades in Asian Annonaceae: Monophyletic genera in the polyphyletic *Miliuseae*. *American Journal of Botany*, 91(4), 590-600. https://doi.org/10.1016/j.ympev.2018.01.021.
38. Mower, J. P. & Vickrey, T. L. (2018). Structural diversity among plastid genomes of land plants. *Advances in botanical research*, 85, 263-292. https://doi.org/10.1016/bs.abr.2017.11.013.
39. Ortiz-Rodriguez, A. E., Ornelas, J. F., & Ruiz-Sanchez, E. (2018). A jungle tale: Molecular phylogeny and divergence time estimates of the Desmopsis-Stenanona clade (Annonaceae) in Mesoamerica. *Molecular Phylogenetics and Evolution*, 122, 80-94. https://doi.org/10.3732/ajb.91.4.590.
40. Ping, J. Y., Feng, P. P., Hao, J., Li J. Y., Su, Y. J., & Wang, T. (2021). The molecular evolution pattern of *rps12* gene in gymnosperms (in Chinese). *Chinese Science Bulletin*, 66(24),3182-3193. https://doi.org/10.1360/TB-2020-1528.
41. Ping, J. Y, Feng, P. P., Li, J. Y., Zhang, R. J., Su, Y. J, & Wang, T. (2021). Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*. *Ecology and evolution*, 11(9), 4786-4802. https://doi.org/10.1002/ece3.7381.
42. Ping, J. Y., Li, A. M., Feng, P. P., Zhu, M., Su, Y.J, & Wang, T. (2021). The highly conserved *rps12* gene in ferns provides strong evidence for decreased substitution rates in the inverted repeat region. *Plant Systematics and Evolution*, 307(2), 1-9. https://doi.org/10.1007/s00606-021-01750-7.
43. Punyasena, S. W., Eshel, G. & McElwain, J. C. (2008). The influence of climate on the spatial patterning of Neotropical plant families. *Journal of Biogeography*, 35, 117-130.
44. Qu, X. J., Moore, M. J., Li, D. Z. & Yi, T. S. (2019). PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*, 15(1), 1-12. https://doi.org/10.1186/s13007-019-0435-7.
45. Raman, G., Park, K. T., Kim, J. H. & Park, S. (2020). Characteristics of the completed chloroplast genome sequence of *Xanthium spinosum*: Comparative analyses, identification of mutational hotspots and phylogenetic implications. *BMC Genomics*, 21(1), 1-14. https://doi.org/10.1186/s12864-020-07219-0.
46. Raubeson L. A. & Jansen R. K. Chloroplast genomes of plants. In: Henry R. J., editor. Plant diversity and evolution: Genotypic and phenotypic variation in higher plants. London: CAB International, (2005). p. 45-68. https://doi.org/10.1079/9780851999043.0045.
47. Raubeson, L. A. & Jansen, R. K. (1992). Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science*, 255(5052), 1697-1699. https://doi.org/10.1126/science.255.5052.1697.

48. Ruhlman T. A., Jansen R. K. The plastid genomes of flowering plants. In: Maliga P., editor. Chloroplast Biotechnology: Methods and Protocols. New York: Springer; 2014. p. 3-38. https://doi.org/10.1007/978-1-62703-995-6_1.
49. Ruhlman, T. A., Zhang, J., Blazier, J. C., Sabir, J. S. & Jansen, R. K. (2017). Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *American Journal of Botany*, 104(4), 559-572. https://doi.org/10.3732/ajb.1600453.
50. Saunders, R. M. K., Guo, X., Tang, C. C. (2020). *Friesodielsia subaequalis* (Annonaceae) a new nomenclatural combination following conservation of the generic name against *Schefferomitra*. *Phytotaxa*, 464(2), 183-184.
51. Sauquet, H., Doyle, J. A., Scharaschkin, T., Borsch, T., Hilu, K. W., Chatrou, L. W. & Le Thomas, A. (2003). Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: Implications for character evolution. *Botanical Journal of the Linnean Society*, 142(2), 125-186. https://doi.org/10.1046/j.1095-8339.2003.00171.x.
52. Schneider, H., Schuettpelz, E., Pryer, K. M., Cranfill, R., Magallón, S. & Lupia, R. (2004). Ferns diversified in the shadow of angiosperms. *Nature*, 428(6982), 553-557. https://doi.org/10.1038/nature02361.
53. She, H.-B., Liu, Z.-Y., Xu, Z.-S., Zhang, H.-L., Cheng, F., Wu, J., Wang, X.-W., & Qian W. (2022). Comparative chloroplast genome analyses of *Cultivated spinach* and two wild progenitors shed light on the phylogenetic relationships and variation. *Science Report*, 12, 856. https://doi.org/10.1038/s41598-022-04918-4
54. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312-1313.
55. Sudianto, E., Wu, C. S., Lin, C. P. & Chaw, S. M. (2016). Revisiting the plastid phylogenomics of Pinaceae with two complete plastomes of *Pseudolarix* and *Tsuga*. *Genome Biology and Evolution,* 8(6), 1804-1811. https://doi.org/10.1093/gbe/evw106.
56. Swofford, D. L. (2002). PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4. Sunderland, MA: Sinauer Associates.
57. Thomas, D. C., Surveswaran, S., Xue, B., Sankowsky, G., Mols, J. B., Keßler, P. J. & Saunders, R. M. (2012). Molecular phylogenetics and historical biogeography of the *Meiogyne-Fitzalania* clade (Annonaceae): Generic paraphyly and late Miocene-Pliocene diversification in Australasia and the Pacific. *Taxon*, 61(3), 559-575. https://doi.org/10.1002/tax.613006.
58. Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R. & Greiner, S. (2017). GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic acids research*, 45(W1), W6-W11. https://doi.org/10.1093/nar/gkx391.
59. Tonti-Filippini, J., Nevill, P. G., Dixon, K., & Small, I. (2017). What can we do with 1000 plastid genomes?. *The Plant Journal*, 90(4), 808-818. https://doi.org/10.1111/tpj.13491.
60. Xue, B., Ding, H. B., Yao, G., Shao, Y. Y., Fan, X. J. & Tan, Y. H. (2020). From *Polyalthia* to *Polyalthiopsis* (Annonaceae): Transfer of species enlarges a previously monotypic genus. *PhytoKeys*, 148, 71. https://doi.org/10.3897/phytokeys.148.50929.
61. Xue, B., Tan, Y. H. (2016). Excluding *Miliusa velutina* (Annonaceae) from Flora of China. *Phytotaxa*, 282(2), 166-169.
62. Xue, B., Nurmawati, S., Xu, Y., Li, Y. Q. (2021). Excluding the species *Monoon fragrans* (Annonaceae) from the flora of China. *Phytotaxa*, 487(1), 91-96.
63. Zhu, A. D., Guo, W. H., Gupta, S., Fan, W. S. & Mower, J. P. (2016). Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytologist*, 209(4), 1747-1756. https://doi.org/10.1111/nph.13743.
64. Zhu, M., Feng, P. P., Ping, J. Y., Li, J. Y., Su, Y. J. & Wang, T. (2021). Phylogenetic significance of the characteristics of simple sequence repeats at the genus level based on the complete chloroplast genome sequences of Cyatheaceae. *Ecology and evolution*, 11(20), 14327-14340. https://doi.org/10.1002/ece3.8151.

Tables

Table1 The information of sample species

Family name		Species name	Genbank accession number
Annonaceae			
	Annonoideae	*Annona cherimola* Mill.	KU563738
		Annona reticulata L.	MT742547
		Annona muricata L.	MT742546
		Fissistigma oldhamii (Hemsl.) Merr.	MW136266
		Fissistigma polyanthum (Hook. f. et Thoms.) Merr.	MW829282
		Uvaria macrophylla Roxb.	MH992130
		Artabotrys hexapetalus (Linn. f.) Bhandari	MZ936420
		Artabotrys pilosus Merr. et Chun	OK216144
	Malmeoideae	*Polyalthiopsis verrucipes* (C. Y. Wu ex P. T. Li) B. Xue & Y. H. Tan	MW018366
		Miliusa glochidioides Handel-Mazzetti Sinensia	OM047203
		Greenwayodendron suaveolens (Engl. & Diels) Verdc.	MH924590
		Chieniodendron hainanense (Merr.) Tsiang et P. T. Li	MK035708
	Ambavioideae	*Cananga odorata* (Lamk.) Hook. f. & Thomson	MN016933
Magnoliaceae			
		Magnolia alba DC.	NC_037005
		Liriodendron chinense (Hemsl.) Sargent.	NC_030504

Table 2 Plastomes characteristics

species name	Genome		LSC		SSC		IR		Numeber			
	size	GC content /%	size	GC content /%	size	GC content /%	size	GC content /%	Gene	CDS	tRNA	rRNA
A. cherimola	201,723	39.6	69,771	38.6	2,966	33.6	64,493	40.2	165	116	41	8
A. reticulata	196,038	39.9	69,650	38.6	3,014	33.5	64,621	40.2	163	115	40	8
A. muricata	201,906	39.6	75,339	38.8	3,105	35.1	58,797	40.7	158	110	40	8
F. oldhamii	187,782	38.9	82,584	38	3,074	31.9	51,062	39.8	150	104	38	8
F. polyanthum	189,920	38.7	83,000	38	3,273	31.6	51,936	39.5	152	106	38	8
U. macrophylla	192,782	38.7	83,581	37.9	3,741	31.5	52,730	39.6	151	105	38	8
A. hexapetalus	178,457	38.8	90,803	37.6	3,066	32.2	42,294	40.3	140	94	38	8
A. pilosus	178,195	38.8	90,797	37.6	3,098	32	42,150	40.4	142	96	38	8
P. verrucipes	159,965	39.0	89,030	37.5	18,987	34.2	25,974	43.4	130	84	38	8
M. glochidioides	159,789	39.2	88,782	37.7	18,949	34.4	26,029	43.4	129	84	37	8
G. suaveolens	159,031	39.1	*	*	*	*	*	*	129	84	37	8
C. hainanense	160,497	39.1	89,424	37.6	18,949	34.3	26,062	43.4	127	84	35	8
C. odorata	167,946	39.0	83,620	37.6	20,310	34.4	32,008	42.4	137	93	36	8
M. alba	160,106	39.2	88,137	37.9	18,777	34.2	26,596	43.2	130	85	37	8
L. chinense	159429	39.2	87,766	37.8	18,997	34.3	26,333	43.2	130	85	37	8

Figures

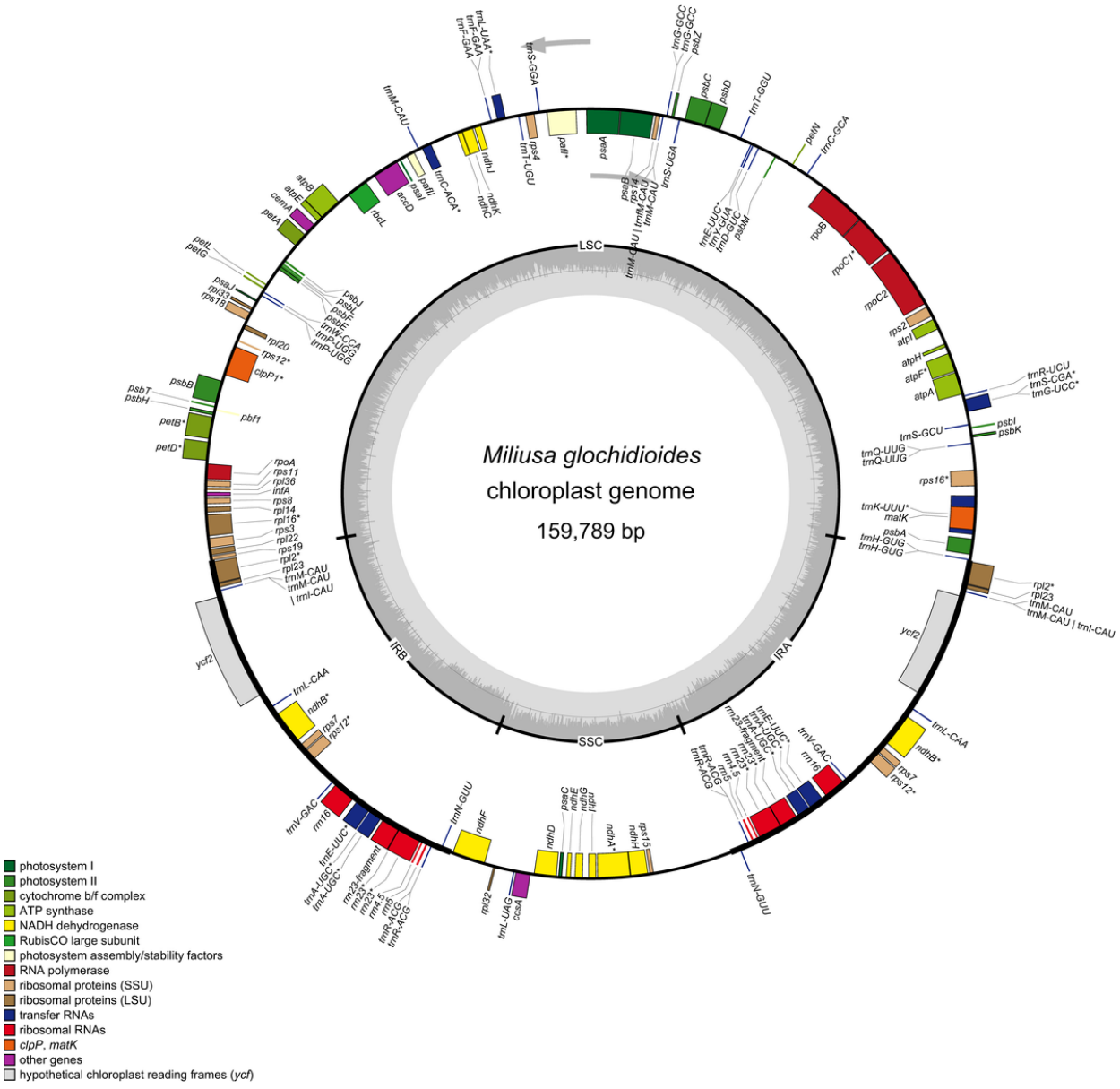

Figure 1

Gene map of *Miliusa glochidioides* plastome

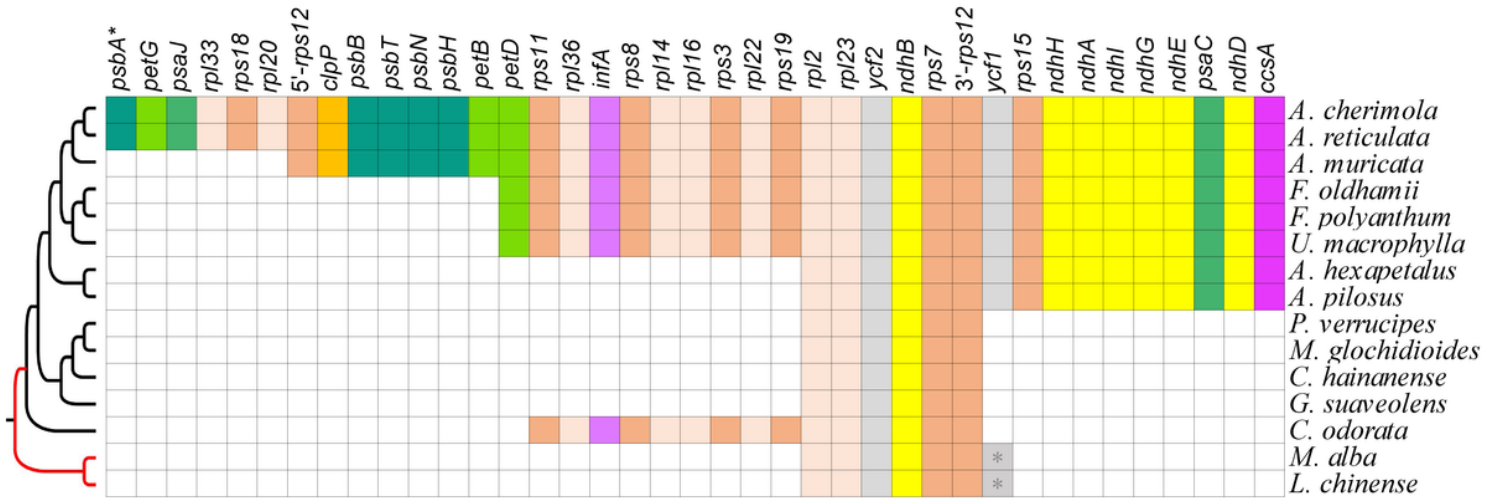

Figure 2

Protein-coding genes in the IR region of each species. There are 37 genes in the IR region of *A. cherimola*and *A. reticulata*, including part of *psbA* that entered the IR region, 32 in *A. muricata*, 25 in *Fissistigma,* and *Uvaria*, 15 in *Artabotrys*, 14 in *Cananga*, and five in the other four genera. In the outgroup (the red branch), there were six genes in the IR region, including part of *ycf1* that entered the IR region

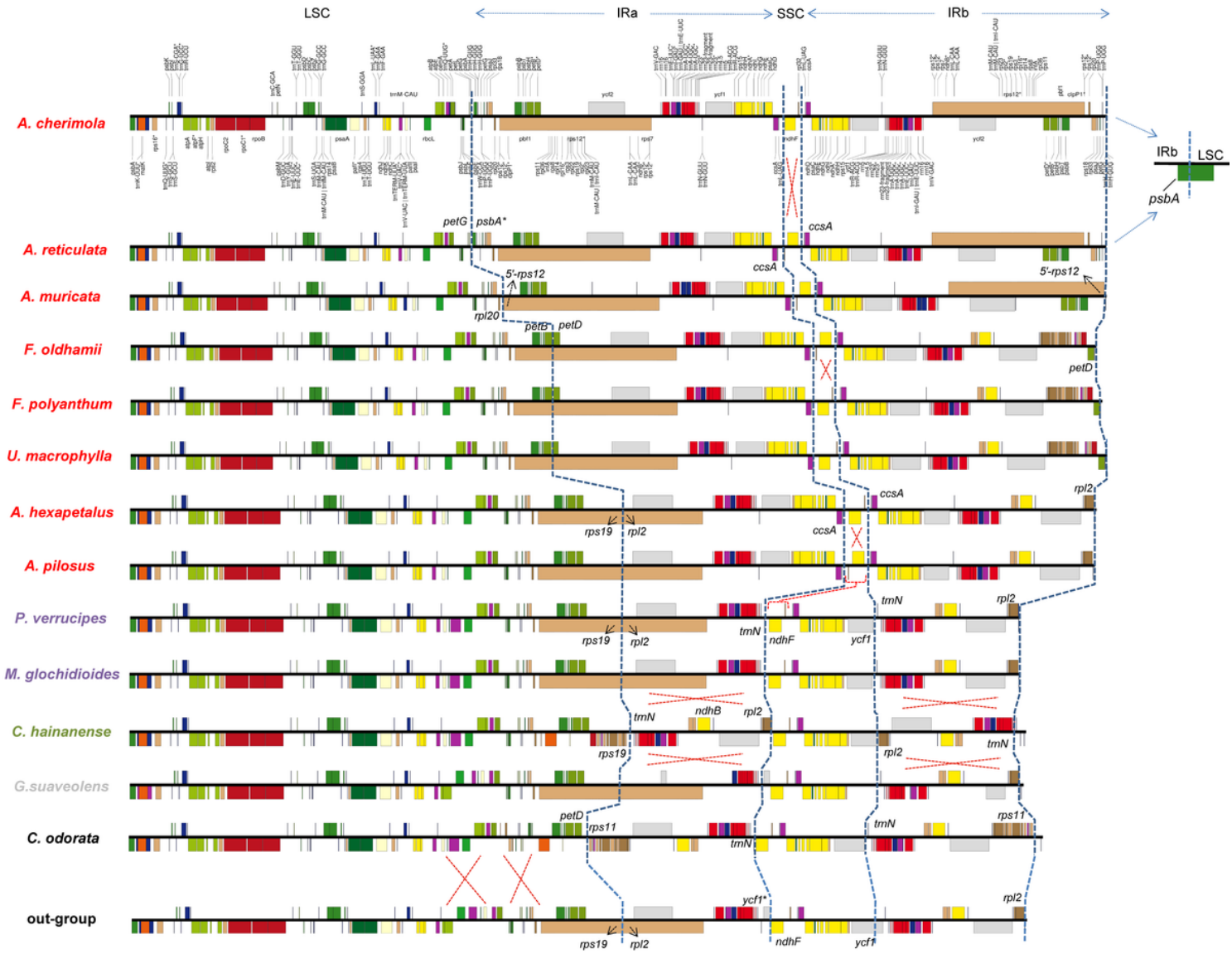

Figure 3

Structural comparison of plastid genomes. Compared with out-groups, the IR region of Annonaceae has been expanded, and the *ndhF-rpl32* fragment undergoes multiple inversions. The inversions of the two fragments happen in *C. odorata* (*pafII-atpE* and 5'-*rps12-psbJ*). In comparison with other species, the IR region of *C. hainanense* was inverted

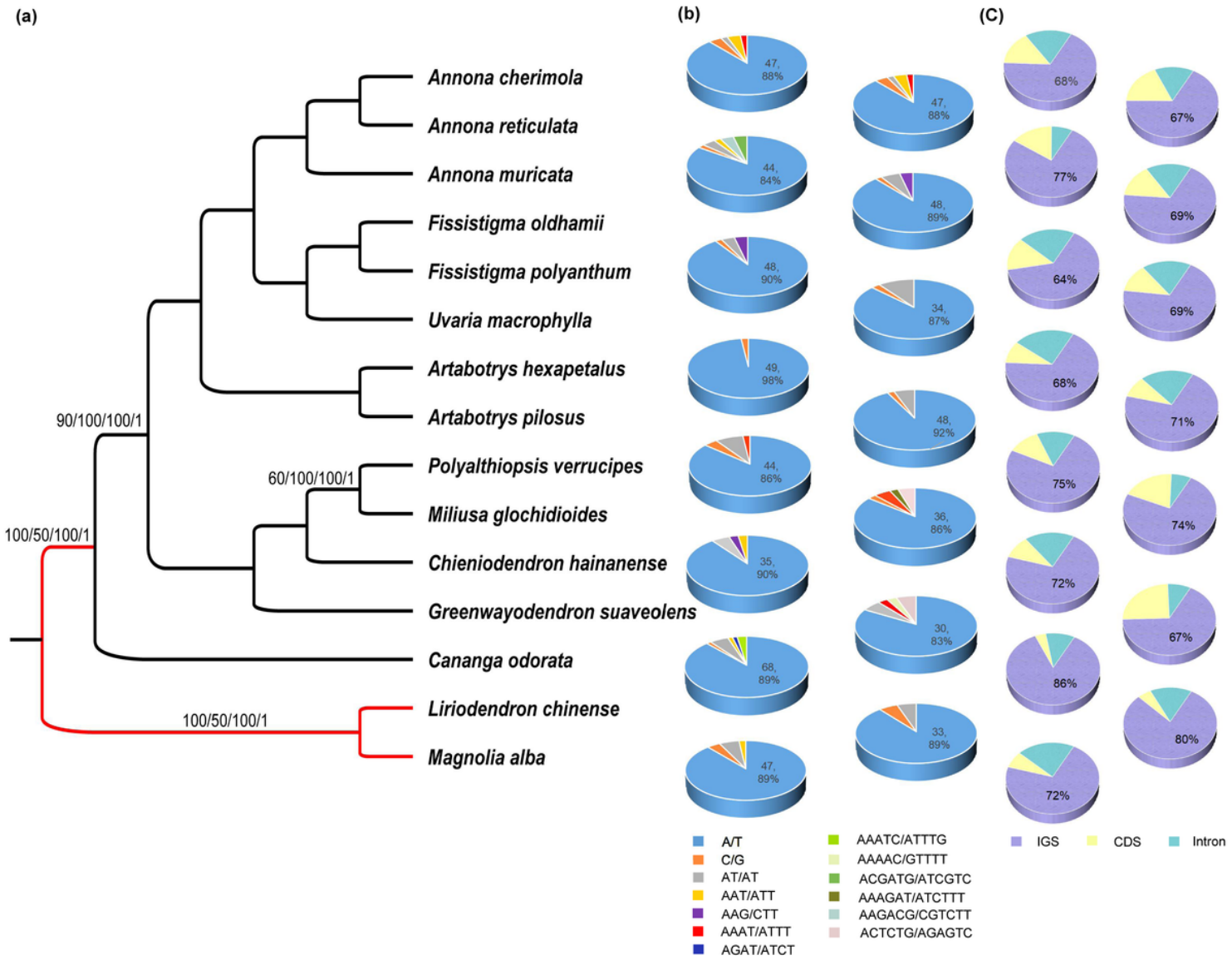

Figure 4

Phylogenetic tree and SSRs analysis. (a) phylogenetic tree constructed with 78 common genes. The red branch was the outgroup, the values on the branch were the branch support of the four methods (NJ/MP/ML/BI), and the support values of the remaining branches were 100/100/100/ 1. (b) the motif types of SSRs in each species are dominated by A/T bases. (c) the location of SSRs in each species on the genome, with the most SSRs located in the IGS

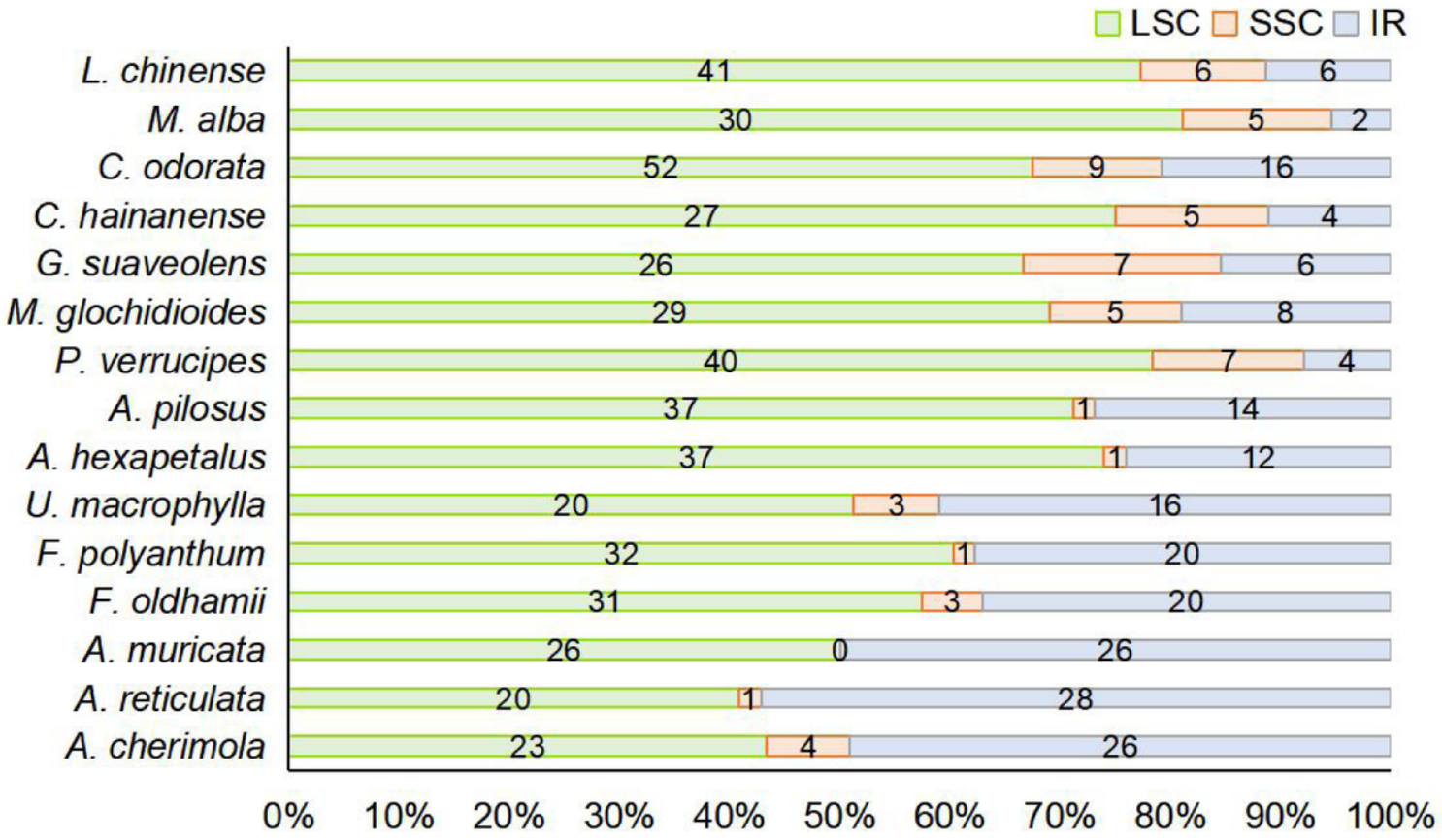

Figure 5

The number of SSRs in the three partitions of each species. Overall, most SSRs were located in the LSC region. Among the species with IR expansion, the SSRs located in this region was much higher than other species

Supplementary Files

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

- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx