Comparative and Phylogenetic Analyses of Complete Plastomes of *Aucuba* (Garryaceae)

Xiong Chen

Yunnan Normal University

Linyuan Fan

Yunnan General Administration of Forestry Seeds and Seedlings

Jian Huang

Yunnan General Administration of Forestry Seeds and Seedlings

Guohua Zhou

Chinese Medicinal Resources Co. LTD, Yunnan Baiyao Group

Yuan Huang (✉ huangyuan@mail.kib.ac.cn)

Yunnan Normal University

Research Article

Keywords: Garryales, Lamiids, Molecular markers, Phylogeny, Sequence variations

Posted Date: April 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-403712/v1>

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

[Read Full License](#)

Abstract

Background: *Aucuba* (Garryaceae), which includes approximately 10 evergreen woody species, is a genus endemic to East Asia. Their striking morphological features give *Aucuba* species remarkable ornamental value. Owing to high levels of morphological divergence and plasticity, species definition of *Aucuba* remains perplexing and problematic. Here, we sequenced and characterized the complete plastid genomes (plastomes) of three *Aucuba* species: *Aucuba chlorascens*, *Aucuba eriobotryifolia*, and *Aucuba japonica*.

Results: Comparative analyses revealed that *Aucuba* plastomes are highly conserved in size, structure, gene content, and organization, and exhibit high levels of sequence similarity. We recommend 11 plastid DNA regions as potential DNA barcodes for species identification and genotyping of *Aucuba* germplasm. Phylogenetic reconstruction based on 71 plastid protein-coding genes from taxa encompassing a wide phylogenetic diversity in the lamiids strongly supported the sister relationship between Garryaceae and Eucommiaceae.

Conclusion: Plastome tree revealed the monophyly of Garryales, offering plastid phylogenomic evidence for the acceptance of Garryales as outlined by the updated Angiosperm Phylogeny Group. Under a comparative framework within Garryales, we detected massive plastome arrangements between *Aucuba* and *Eucommia*. In summary, our study provides useful genomic resources for further study of the taxonomy, evolution, conservation, and exploitation of *Aucuba* species.

Background

Aucuba Thunberg is small genus of ~ 10 evergreen woody species distributed in the Eastern Himalayas, China, Korea, Japan, Myanmar, and Vietnam [1]. Although the genus is currently endemic to East Asia, the earliest *Aucuba* fossil, which has been dated to the Eocene, was found in Washington State, USA, suggesting that *Aucuba* had a much broader distribution in the early Cenozoic [2] and its extant, relatively narrow distribution may have resulted from a range contraction triggered by the Neogene global cooling and the Pleistocene glaciations [3]. *Aucuba* is easy to recognize owing to its morphological distinctiveness, but difficult to place taxonomically. The taxonomic affinities of *Aucuba* have been in dispute since the establishment of the genus. Historically, this genus was placed into either the family Cornaceae [4–7] or the monotypic family Aucubaceae [8–11]. Recently, phylogenetic analyses based on DNA sequences revealed a sister relationship between *Aucuba* and *Garrya*, and the two genera are in turn closely related to Eucommiaceae [12–15]. Since *Aucuba* and *Garrya* show high levels of similarity in their morphologies and chemical components, they were grouped together as the family Garryaceae [16, 17]. Together with the monotypic family Eucommiaceae, which includes only one “living fossil” species (*Eucommia ulmoides*), Garryaceae was placed in the order Garryales within lamiids by the Angiosperm Phylogeny Group [16, 17].

Aucuba possesses remarkable horticultural merits. Because of their evergreen habit, spotted and colorful leaves, and showy fruits, *Aucuba* species have been widely introduced and cultivated as garden plants for centuries in East Asia, Europe, and North America [18, 19]. Previous research on *Aucuba* mainly focused on cultivation management, introduction and domestication, phytochemistry, and cytogenetics [18–22]. Genomic resources crucial for plant breeding, however, have received much less attention. In addition, the morphologies of *Aucuba* species are highly divergent and plastic, making morphology-based taxonomy perplexing and problematic [23], and hindering the effective conservation and exploitation of the germplasm. Development of robust and polymorphic molecular markers as DNA barcodes will contribute to accurate and reliable species identification in *Aucuba*.

Chloroplasts are organelles in green plants that perform photosynthesis and the biosynthesis of starch, fatty acids, pigments, and amino acids [24]. Typical angiosperm plastid genomes (plastomes), which consist of a large single copy (LSC, 80–90 kbp) region and a small single copy (SSC, 16–27 kbp) region separated by a pair of inverted repeat (IR, 20–30 kbp) regions, are characterized by circular DNA, uniparentally inherited, and highly conserved in structure and gene content [24–27]. Accordingly, gene content, structural arrangement, gene loss or pseudogenization, cytonuclear gene transfer, and sequence variations can provide informative and valuable resources for elucidating evolutionary relationships and species discrimination in plants [28–31]. With the advent of next-generation DNA sequencing technologies, plastome sequencing has been widely used in recent years to investigate evolutionary relationships among closely related species [28, 31–34] and to develop molecular markers for species identification [35–40] and population genetics [41–44]. Plastome sequencing can also offer genetic information related to synthesis of metabolic compounds, biopharmaceuticals, and tolerance to biotic and abiotic stress [45–49]. The availability and use of complete plastome sequences in biotechnology is likely to increase the performance of cultivated plants in the field [24, 50–52].

In the current study, we sequenced and characterized the complete plastomes of three *Aucuba* species (i.e., *Aucuba chlorascens*, *Aucuba eriobotryifolia*, and *Aucuba japonica*). Under a comparative framework within Garryales, we detected massive genomic arrangements between *Aucuba* and *Eucommia*. We used plastome DNA sequences to infer the phylogenetic relationships of *Aucuba* with other lamiids. The plastid genomic resources presented here will be beneficial for conservation and exploitation of *Aucuba* species.

Results

General features of *Aucuba* plastomes

Paired-end Illumina sequencing generated over 30 million clean reads for each species. *De novo* assembly yielded three complete *Aucuba* plastomes, each identically encoding 114 unique genes: 80 protein-coding genes, 30 tRNAs, and 4 rDNA (Table S3). The largest plastome was that of *A. japonica* with 158,237 bp (288.577× coverage), followed by *A. eriobotryifolia* with 158,113 bp (345.936× coverage), and *A. chlorascens* with 158,084 bp (636.433× coverage). The fully assembled and annotated plastomes

of *A. chlorascens*, *A. eriobotryifolia*, and *A. japonica* were deposited in the NCBI GanBank database under the accession numbers presented in Table 1.

The newly generated *Aucuba* plastomes exhibited the typical quadripartite structure (Fig. S1), consisting of a pair of IRs (26,008 to 26,143 bp in length) separated by the LSC (87,281 to 87,505 bp in length) region and the SSC (18,544 to 18,550 bp in length) region (Table 1). The overall GC content among these *Aucuba* plastomes was identical (37.7%), which was unevenly distributed in LSC, SSC, and IRs (Table 1). The highest GC content was found in the IR regions (43.0% in *A. eriobotryifolia* and *A. japonica*, 43.1% in *A. chlorascens*), followed by the LSC (35.9%). The lowest GC content was detected in the SSC region (31.5% in *A. chlorascens* and *A. eriobotryifolia*, 32.6% in *A. japonica*).

Table 1
Features of *Aucuba* plastomes newly generated in this study.

	<i>A. chlorascens</i>	<i>A. eriobotryifolia</i>	<i>A. japonica</i>
GenBank Accession	MT338539	MT338540	MT338541
Sequencing coverage (×)	288.577	345.936	536.433
Genome Size (bp)	158,084	158,113	158,237
Large single copy (bp)	87,518	87,281	87,505
Inverted repeats (bp)	26,008	26,143	26,094
Small single copy (bp)	18,550	18,546	18,544
Total number of genes	114	114	114
Coding genes (CDS)	80	80	80
Transfer RNA genes (tRNA)	30	30	30
Ribosomal RNA genes (rRNA)	4	4	4
GC content (%)	Overall	37.7	37.7
	LSC	35.9	35.9
	IR	43.1	43.0
	SSC	31.5	31.5

IR expansion/contraction and genomic rearrangements

We compared the border position of the four plastid regions (LSC, IRA, SSC, and IRB) among four Garryales species (Fig. 1). These species showed similar IR/LSC boundaries; expansion of the IR regions had identically duplicated the *rp12* gene at the IRA/LSC and IRB/LSC junctions in their plastomes. However, the IR/SSC junctions between *Aucuba* and *Eucommia* plastomes exhibited certain differences. Expansion of IR regions into the *ycf1* gene at the IR/SSC boundary occurred in all *Aucuba* species,

leading to an overlap between truncated *ycf1* and *ndhF* at the IRA/SSC junction and a partial duplication of *ycf1* at the IRB/SSC boundary. Comparatively, the IRA/SSC and IRB/SSC boundaries of *E. ulmoides* were located in the intergenic regions between *ycf1* and *ndhF*, and between *rps15* and *ycf1*. We identified three large inversions and an infra-genomic translocation within the LSC of the *E. ulmoides* plastome when compared to other Lamiid plastomes (Fig. 2), including an inversion of ~ 46 kb from *rps16* to *trnT_UGU* (A), an inversion of ~ 17 kb between *trnQ_UUG* and *rps12* (B), and an inversion of ~ 6 kb located between *trnL_UAA* and *trnV_UAC* (C). The latter two sequence regions exchanged positions with each other.

Hyper-variable regions among *Aucuba* plastomes

Overall, we found high levels of sequence similarity among the three *Aucuba* plastomes. Plastome-wide mVISTA analysis (Fig. 3) identified 2,306 variation sites within the 158,856 alignment positions, representing 1.15% variation proportion among *Aucuba* plastomes. Comparatively, the LSC and SSC regions were more variable between species than the two IR regions, and most sequence variations were identified in the intergenic spacers and introns. Sliding window analysis of entire plastomes revealed 11 plastid DNA regions with relatively higher nucleotide diversity than average plastome diversity (Fig. 4), namely, *rps16* intron, *rps16-trnQ*, *rpoB-trnC*, *psbM-trnD*, *atpB-rbcL*, *accD-psaI*, *ycf4-cemA*, *rps3*, *ccsA-ndhD*, *rps15-ycf1*, and *ycf1*. Features of these hyper-variable sequence regions are shown in Table 2.

Table 2
Hyper-variable regions among *Aucuba* plastomes.

Regions	Nucleotide diversity (Pi)	Sequence length (bp)	Variations	Divergence proportion (%)
<i>rps16</i> intron	0.0194	884	19	2.1493
<i>rps16-trnQ</i>	0.0122	1,807	24	1.3282
<i>rpoB-trnC</i> - <i>petN</i>	0.0156	2,163	36	1.6644
<i>psbM-trnD</i>	0.0144	1,207	21	1.7399
<i>atpB-rbcL</i>	0.0122	745	12	1.6107
<i>ycf4-cemA</i>	0.0133	360	10	2.7778
<i>accD-psaI</i>	0.0150	705	16	2.2695
<i>rps3</i>	0.0122	657	7	1.0654
<i>ccsA-ndhD</i>	0.0167	247	11	4.4534
<i>rps15-ycf1</i>	0.0128	388	7	1.8041
<i>ycf1</i>	0.0222	5,688	78	1.3713

Phylogenetic relationships

Phylogenetic reconstruction based on ML and BI methods produced identical tree topologies (Fig. 5). Both the campanulids and lamiids were recovered as strongly supported monophyletic lineages (BS = 100%, PP = 1.00). Within the lamiids, the orders Garryales, Metteniusales, and Icaciniales were resolved as early branching clades (basal lamiids), while the orders Solanales, Boraginales, Gentianales, and Lamiales (core lamiids) were recovered as terminal branches in the phylogenetic tree. Monophyly of the core lamiids was robustly supported by both ML and BI phylogenies (BS = 100%, PP = 1.00), but the relationships among the four core Lamiid orders received low/modest support values: Boraginales was recovered as a sister clade to Gentianales (BS = 46%, PP = 0.83), with these together being sister to the clade of Solanales (BS = 43%, PP = 0.90). Although our data failed to resolve the relationships among Garryales, Metteniusales, and Icaciniales, given their low support values, the monophyly of Garryales (BS = 100%, PP = 1.00) and the sister relationship between Garryaceae and Eucommiaceae (BS = 100%, PP = 1.00) were strongly supported.

Discussion

Plastome comparison

The sizes of *Aucuba* plastomes reported in this study fall with the average size of angiosperm plastomes [25]. Despite the close relationship between Garryaceae and Eucommiaceae, the plastome sizes of *Aucuba* species are ~ 5,000 bp smaller than that of *E. ulmoides* (163,341 bp; [70]). We found that both the LSC (87,281 to 87,505 bp) and SSC (18,544 to 18,550 bp) regions of *Aucuba* plastomes were slightly larger than that of *E. ulmoides* (LSC: 86,592 bp, SSC: 14,191 bp; [70]), while their IR regions (26,008 to 26,143 bp) were ~ 10,000 bp smaller than that of *E. ulmoides* (31,300 bp; [70]). This implies that a relatively smaller IR length is responsible for the size reduction in *Aucuba* plastomes.

IR expansion/contraction, gene loss, and intergenic region variation are proposed to account for plastome size variation [27, 71]. IR expansion/contraction is a commonly observed structural mutation in angiosperm plastomes, which may lead to lineage-specific gain (or loss) or pseudogenization of a small number of genes in these regions [26, 27]. Although the expansion of IRs identically duplicated the *rp12* gene at the IR/LSC boundaries, we observed differences in the IR/SSC junctions between *Aucuba* and *Eucommia*. *Aucuba* plastomes possessed more progressively expanded IRs at the IR/SSC junctions; however, their IR regions were relatively shorter than that of *E. ulmoides*, suggesting that the IR size variation between *Aucuba* and *Eucommia* results from intergenic region variation rather than IR expansion/contraction.

Large rearrangements in the LSC regions suggest that Garryales plastomes are highly divergent in gene organization despite exhibiting high levels of similarity in gene content. Interestingly, the inversions observed in the LSC of the *E. ulmoides* plastome were located flanking the LSC and near to the LSC/IR junctions. This supports the idea that inversions in plastomes might be linked to IR expansion/contraction [72]. Moreover, the regions flanking the inversions contained tRNAs, coinciding with the assumption that tRNA activity most likely triggers inversions in plastomes [73].

Although Garryales as circumscribed by the Angiosperm Phylogeny Group (2009, 2016) consist of only two extant families, both Eucommiaceae (*Eucommia*) and Garryaceae (*Aucuba* and *Garrya*) represent ancient lineages since their divergence was estimated as no later than the Paleocene [74, 75]. The significant difference in their genic arrangement is most likely associated with the long-term divergent evolution of plastomes between *Aucuba* and *Eucommia*. Plastome rearrangements such as inversions and relocation of genes are used as informative markers to explore evolutionary relationships in angiosperms [76]. Since the complete plastome of *Garrya* sequenced by Stull et al. (2015) is still unavailable in the GenBank database [64], further investigation is needed to confirm whether this trait is a molecular synapomorphy for Eucommiaceae.

Potential DNA barcodes

Aucuba is a paleo-endemic genus in East Asia; the earliest fossil belonging to this genus has been dated to the Eocene in North America [2, 3]. However, the high sequence similarity among the three *Aucuba* plastomes indicates that few sequence variations have accumulated since the divergence of these species. This implies that these species may have derived from more recent speciation events, with consequent low levels of sequence evolution. These recently diverged *Aucuba* species therefore represent neo-endemic elements in the flora of East Asia. The evolutionary profiles of *Aucuba* inferred in the current study can deepen our insights into the evolution of the floristic endemism in East Asia.

The diversity and plasticity of morphological characteristics among *Aucuba* species has led to difficulties reconstructing their taxonomy. Plastid DNA sequences, *rbcL*, *matK*, and *psbA-trnH*, are recommended as standard DNA barcodes for plant species discrimination [77]; however, both mVISTA and sliding window analyses revealed that these sequences exhibit relatively low levels of variation among *Aucuba* species. These standard DNA barcodes therefore have limited discriminatory power in *Aucuba*. The complete plastome DNA sequences newly generated in this study provide genomic resources for the development of novel DNA barcodes. Based on plastome-wide analysis of sequence variability, we propose 11 plastid DNA regions harboring relatively high proportions of variable sites. These sequences can serve as reliable and effective DNA barcodes for species identification and germplasm genotyping in *Aucuba*.

Phylogenetic inferences

Complete plastome sequences have been widely used for resolving recalcitrant relationships in phylogenetically challenging taxa [28, 31, 36, 78–80]. As previous phylogenetic analyses based on complete plastome data only involved either *Garrya* [64] or *Eucommia* [63, 70], the monophyly of Garryales recognized by the Angiosperm Phylogeny Group (2009, 2016) needed to be assessed with extended sample size. In this study, the phylogenetic placement of *Aucuba* was inferred by reconstructing phylogenetic relationships based on a large dataset comprising 71 protein-coding genes from 70 plastomes representing 28 families and 11 orders within campanulids and lamiids. Our data strongly support the sister relationship between *E. ulmoides* (Eucommiaceae) and *Aucuba* (Garryaceae), as well as the monophyly of Garryales. This result is consistent with phylogenetic analyses using single- or

multiple-locus DNA sequences [12–15], providing plastid phylogenomic evidence to accept the order Garryales as circumscribed by the Angiosperm Phylogeny Group (2009, 2016).

Although a previous study [64] based on 73 plastid protein-coding genes satisfactorily resolved the relationships within the basal lamiids, our phylogenetic analyses failed to reconstruct robust relationships among Garryales, Icaciniales, and Metteniusales. The utilization of either too few DNA sequences or limited taxon sampling in phylogenetics may result in significant phylogenetic errors [81, 82]. Notably, only six basal lamiid taxa were sampled in the current study in contrast to the 43 representatives included in the phylogenetic analyses of Stull et al. (2015). The weakly supported relationships recovered in this study can most likely be attributed to inadequate sampling of taxa within these basal lamiid orders.

Relationships among the core lamiids remain unresolved and have seldom received strong branch support in previous studies [14, 15, 63, 64, 83]. The current sampling size within this clade (54 taxa) represents a large dataset. Unfortunately, our phylogenetic analyses still failed to resolve the puzzling problem in core lamiids phylogeny. The intricate relationships within the core lamiids might be due to an ancient radiation [64, 84]. Since the sequenced plastomes only represent a small fraction of the real diversity, accumulation of additional plastomes, especially in Boraginales, Solanales, and Gentianales, will facilitate resolution of the phylogenetic puzzle.

Methods

DNA extraction, shotgun sequencing, plastome assembly, and annotation

Samples of *Aucuba chlorascens*, *A. eriobotryifolia*, and *A. japonica* were collected from the Botanical Garden of Kunming Institute of Botany, Kunming, China. The formal identification of the plant material was undertaken by the Herbarium of Kunming Institute of Botany (KUN), and voucher specimens were deposited at KUN (JC-YJ-64, JC-YJ-66, JC-YJ-68). Genomic DNA was isolated from ~ 50 mg silica-gel-dried leaf tissues using the CTAB method [53]. Genomic DNA was fragmented into 500 bp by ultrasonic disruption to construct libraries. Paired-end libraries were prepared according to the manufacturer's protocol (Illumina, San Diego, CA, USA) for sequencing on an Illumina HiSeq 2500 system.

Low-quality reads were removed from raw data using NGS QC Toolkit [54], by setting the cut-off value for percentage of read length to 80 and PHRED quality scores to 30. Filtered reads were used for *de novo* assembly of *Aucuba* plastomes using NOVOPlasty v2.7.0 [55], setting the k-mer size as 30. The DNA sequence of the RuBisCO large subunit of *A. japonica* (GenBank Accession: AY725858) was used as seed for recovery of complete plastomes by iterative extension. Assembled plastomes were annotated using the Dual Organellar Genome Annotator database [56]. Start and stop codons and intron/exon boundaries for protein-coding genes were checked manually. Annotated tRNA genes were further verified using

tRNAscan-SE 1.21 [57] with default parameters. Annotated plastomes were illustrated using the online program OrganellarGenomeDRAW [58].

Comparison of plastomes

The whole plastome DNA sequence of *Eucommia ulmoides* (GenBank Accession: KU204775) was used as a reference. Together with the newly generated *Aucuba* plastomes in this study, IR/LSC and IR/SSC boundaries of Garryales plastomes were compared using Geneious V10.2 [59]. To investigate differences in the Garryales plastomes with respect to those of other lamiids, we progressively aligned the *E. ulmoides* and assembled *Aucuba* plastomes with those of six species representing Icaciniales, Boraginales, Gentianales, Metteniusales, Solanales, and Lamiales, respectively, using the multiple genome alignment software Mauve 2.3.1 [60] with default parameters.

The plastomes newly generated in this study were pairwise aligned using the mVISTA program with LAGAN mode. Subsequently, the annotated plastomes were aligned, and variable sites were identified using MEGA 6.0 software [61]. Nucleotide diversity (Π) among *Aucuba* plastomes was calculated using DnaSP 5.10.01 [62]. The step size was set to 200 bp, with an 800 bp window length.

Phylogenetic analyses

Phylogenetic reconstruction included four complete Garryales plastomes, of which three *Aucuba* plastomes were newly generated in the present study. To investigate the phylogenetic location of *Aucuba* in the lamiidae clade, the complete plastomes of an additional 60 taxa from 22 families and seven orders of the lamiids, as well as seven taxa from six families and four orders of the campanulids, were included in the analyses (Table S1). According to the phylogenetic relationships recovered by previous studies [63, 64], *Stewartia obovate* (Ericales) was selected as the outgroup to root the phylogenetic tree. Seventy-one protein-coding genes (Table S2) commonly shared by these taxa were used for phylogenetic reconstruction. Alignments of these genes were concatenated using MAFFT software [65]. The best-fitting partition scheme and nucleotide substitution models were identified using the program PartitionFinder v2.1.1 [66], using the “greedy” search algorithm.

Bayesian inference (BI) and maximum-likelihood (ML) analyses were employed to reconstruct phylogenetic relationships. The BI analyses were performed in MrBayes v3.2 [67]. Two independent Markov chains, each starting with a random tree, were run simultaneously for one million generations, with trees sampled every 100 generations. Trees from the first 250,000 generations were regarded as “burn in” and discarded. Posterior probability values (PP) were determined from the remaining trees. The ML analyses were performed in RAxML-HPC BlackBox v8.1.24 [68, 69]; 10 independent ML searches were conducted, and branch bootstrap support (BS) was determined by computing 1,000 nonparametric bootstrap replicates.

Declarations

Acknowledgments

We are grateful to Zhangming Wang, Ren Zhao, and Guofeng He for their help in sampling plant materials. This research was financially supported by the National Natural Science Foundation of China (Grant No. 31060052, 31960050), and the Major Project on Biodiversity Conservation of Chinese Ministry of Ecology and Environment.

Authors' Contributions

YH conceived and designed the experiments; LF, JH and GZ collected plant materials; XC, LF performed the experiments and drafted the manuscript; YH, XC, LF revised the manuscript. All authors read and approved the manuscript.

Funding

This research was financially supported by the National Natural Science Foundation of China (Grant No. 31960050), and the Major Project on Biodiversity Conservation of Chinese Ministry of Ecology and Environment.

Availability of data and materials

All sequences used in this study are available from the National Center for Biotechnology Information (NCBI) (Accession numbers: MT338539, MT338540, MT338541)

Ethics approval and consent to participate

Collection of all samples completely complies with national and local legislation permission. No specific permission was required for collecting these plants. . This study did not involve any endangered species or species at risk of extinction according to IUCN Policy Statement, and the plant samples used in the study were not collected from national park or natural reserve. All methods used in this study were in compliance with relevant institutional, national, and international guidelines and legislation.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Xiang QY, Boufford DE. *Aucuba*. In: Wu ZY, Raven P, Hong DY, editors. Flora of China. Beijing & St Louis: Science Press; 2005. p. 222–226.
2. Wehr WC, Hopkins DQ. The Eocene orchards and gardens of Republic. Washington: Washington Geology; 1994.

3. Manchester SR, Chen Z, Anming LU, Uemura K. Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern Hemisphere. *J Syst Evol.* 2009;47:1–42.
4. Wangerin W. Cornaceae. In: Engler A, Engelmänn, editors. *Das Pflanzenreich IV.* Leipzig: Germany; 1910. p. 1–101.
5. Hutchinson J. *Genera Plantarum.* In: Hutchinson J, editors. *The Genera of Flowering Plants (Angiospermae).* UK: Oxford University Press; 1967. p. 497–522.
6. Harms H. Cornaceae. In: Engelmänn, editor. *Nat. Pflanzenfam.* Leipzig: Germany; 1898. p. 250–270.
7. Arthur C. *The evolution and classification of flowering plants.* 2nd ed. New York: Allen Press; 1988.
8. Bremer K, Chase MW, Stevens PF, Anderberg AA, Backlund A, Bremer B, et al. An ordinal classification for the families of flowering plants. *Ann Mo Bot Gard.* 1998;85:531–553.
9. Willis JC, Shaw HKA. Aucubaceae. In: Society AF, editor. *A dictionary of the flowering plants and ferns.* Cambridge: Cambridge University Press; 1973. p. 142–143.
10. Takhtajan AL. Outline of the classification of flowering plants (magnoliophyta). *Bot Rev.* 1980;46:225–359.
11. Bremer B, Bremer K, Chase M, Reveal J, Zmarzty S. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot J Linn Soc.* 2003;141:399–436.
12. Xiang QY, Soltis DE. *rbcL* sequence data define a cornaceous clade and clarify relationships of Cornaceae *sensu lato*. In: Boufford DE, Ohba H, editors. *Sino-Japanese flora - its characteristics and diversification.* Tokyo: Univ Mus Bull; 1998. p. 123–137.
13. Xiang QY, Soltis DE, Soltis M. Phylogenetic Relationships of *Cornus L. Sensu Lato* and Putative Relatives Inferred from *rbcL* Sequence Data. *Ann Mo Bot Gard.* 1993;80(3):723–734.
14. Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis MJ, et al. Angiosperm phylogeny inferred from *18S*rDNA, *rbcL*, and *atpB* sequences. *Bot J Linn Soc.* 2000;133:381–461.
15. Bremer K, Backlund A, Sennblad B, Swenson U, Andreassen K, Hjertson M, et al. A phylogenetic analysis of 100 + genera and 50 + families of euasterids based on morphological and molecular data with notes on possible higher level morphological synapomorphies. *Plant Syst Evol.* 2001;229:137–169.
16. Group, TAP. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc.* 2009;161:105–121.
17. Bremer K, Chase MW, Reveal JL, Soltis DE, Soltis PS, Stevens PF, et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc.* 2016;181:1–20.
18. Ranney TG, Thomasson TH, Neill K, Lynch NP, Weathington M. Ploidy, Relative genome size, and inheritance of spotted foliage in *Aucuba* Species (Garryaceae). *Hortscience.* 2018;53:1271–1274.
19. Hagedoorn AL. *Plant breeding.* London: Crosby Lockwood & Son, Ltd.; 1950. p. 237.
20. Allen EF. *Aucuba japonica*: an example of cytoplasmic inheritance. *Plantsman* 1990;11:244–245.

21. Ohi T, Kajita T, Murata J. Distinct geographic structure as evidenced by chloroplast DNA haplotypes and ploidy level in Japanese *Aucuba* (Aucubaceae). *Am J Bot.* 2003;90:1645–1652.
22. Lehrer JM. Shedding New Light on *Aucuba*. *Am Nurseryman.* 2009;209:38.
23. Song ZP. *Aucuba*. In: Fang WP, Hu WK, editors. *Flora Reipublicae Popularis Sinicae*. Beijing: Science Press; 1990. p. 7–21.
24. Daniell H, Lin C, Yu M, Chang W. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 2016;17:134–134.
25. Wicke S, Schneeweiss GM, De Pamphilis CW, Muller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 2011;76:273–297.
26. Plunkett GM, Downie SR. Expansion and Contraction of the Chloroplast Inverted Repeat in Apiaceae Subfamily Apioideae. *Syst Bot.* 2000;25:648–667.
27. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, et al. The Complete Chloroplast Genome Sequence of *Pelargonium × hortorum*: Organization and Evolution of the Largest and Most Highly Rearranged Chloroplast Genome of Land Plants. *Mol Biol Evol.* 2006;23:2175–2190.
28. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 2009;7:84–84.
29. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci USA.* 2010;107:4623–4628.
30. Moore MJ, Bell CD, Soltis PS, Soltis DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci USA.* 2007;104:19363–19368.
31. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebensmack J, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA.* 2007;104:19369–19374.
32. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer S, Leebensmack J, et al. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol.* 2016;209:855–870.
33. Comer JR, Zomlefer WB, Barrett CF, Davis JI, Stevenson DW, Heyduk K, et al. Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from next-generation sequencing. *Am J Bot.* 2015;102:888–899.
34. Ji Y, Liu C, Yang Z, Yang L, He Z, Wang H, et al. Testing and using complete plastomes and ribosomal DNA sequences as the next generation DNA barcodes in *Panax* (Araliaceae). *Mol Ecol Resour.* 2019;19:1333–1345.
35. Huang Y, Li X, Yang Z, Yang C, Yang J, Ji Y, et al. Analysis of Complete Chloroplast Genome Sequences Improves Phylogenetic Resolution in *Paris* (Melanthiaceae). *Front Plant Sci.* 2016;7:1797–1797.

36. Ji Y, Yang L, Chase MW, Liu C, Yang Z, Yang J, et al. Plastome phylogenomics, biogeography, and clade diversification of *Paris* (Melanthiaceae). *BMC Plant Biol.* 2019;19:1–14.
37. Kane NC, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JMM, et al. Ultra-barcoding in *cacao* (*Theobroma spp.*, Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Am J Bot.* 2012;99:320–329.
38. Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*. *Mol Ecol Resour.* 2015;15:1067–1078.
39. Yang J, Tang M, Li H, Zhang Z, Li D. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol Biol.* 2013;13:84–84.
40. Ji Y, Liu C, Yang J, Jin L, Yang Z, Yang J-B. Ultra-barcoding discovers a cryptic species in *Paris yunnanensis* (Melanthiaceae), a medicinally important plant. *Fron Plant Sci.* 2020;11:411.
41. Lopes ADS, Pacheco TG, Santos KGD, Vieira LDN, Guerra MP, Nodari RO, et al. The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Rep.* 2018;37:307–328.
42. Pacheco TG, Lopes ADS, Viana GDM, Silva OND, Silva GMD, Vieira LDN, et al. Genetic, evolutionary and phylogenetic aspects of the plastome of *annatto* (*Bixa orellana* L.), the Amazonian commercial species of natural dyes. *Planta.* 2019;249:563–582.
43. Sigmon BA, Adams RP, Mower JP. Complete chloroplast genome sequencing of vetiver grass (*Chrysopogon zizanioides*) identifies markers that distinguish the non-fertile ‘Sunshine’ cultivar from other accessions. *Ind Crop Prod.* 2017;108:629–635.
44. Cho K, Cheon KS, Hong SY, Cho JH, Im JS, Mekapogu M, et al. Complete chloroplast genome sequences of *Solanum commersonii* and its application to chloroplast genotype in somatic hybrids with *Solanum tuberosum*. *Plant Cell Rep.* 2016;35:2113–2123.
45. De Cosa B, Moar WJ, Lee S, Miller ME, Daniell H. Overexpression of the Bt cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nat Biotechnol.* 2001;19:71–74.
46. Dufourmantel N, Pelissier B, Garcon F, Peltier G, Ferullo J, Tissot G, et al. Generation of fertile transplastomic *soybean*. *Plant Mol Biol.* 2004;55:479–489.
47. Jin S, Zhang X, Daniell H. Pinellia ternata agglutinin expression in chloroplasts confers broad spectrum resistance against aphid, whitefly, Lepidopteran insects, bacterial and viral pathogens. *Plant Biotechnol J.* 2012;10:313–327.
48. Lee S, Li B, Jin S, Daniell H. Expression and characterization of antimicrobial peptides Retrocyclin-101 and Protegrin-1 in chloroplasts to control viral and bacterial infections. *Plant Biotechnol J.* 2011;9:100–115.
49. Viitanen P, V. Metabolic Engineering of the Chloroplast Genome Using the *Echerichia coli ubiC* Gene Reveals That Chorismate Is a Readily Abundant Plant Precursor for p-Hydroxybenzoic Acid Biosynthesis. *Plant Physiol.* 2004;136:4048–4060.

50. Rogalski M, Vieira LDN, Fraga HPDF, Guerra MP. Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front Plant Sci.* 2015;6:586–586.
51. Vining KJ, Johnson SR, Ahkami A, Lange I, Parrish AN, Trapp SC, et al. Draft Genome Sequence of *Mentha longifolia* and Development of Resources for Mint Cultivar Improvement. *Mol Plant.* 2017;10:323–339.
52. Wurbs D, Ruf S, Bock R. Contained metabolic engineering in tomatoes by expression of carotenoid biosynthesis genes from the plastid genome. *Plant J.* 2007;49:276–288.
53. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 1987;19:11–15.
54. Patel RK, Mukesh J, Liu Z. NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *Plos One.* 2012;7:e30619.
55. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 2017;45:e18.
56. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 2004;20:3252–3255.
57. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 2005;33:686–689.
58. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 2007;52:267–274.
59. Kearse M, Moir R, Wilson A, Stones-Havas S, Drummond A. Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. *Bioinformatics.* 2012;28:1647–1649.
60. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One.* 2010;5:e11147.
61. Tamura K, Stecher G, Peterson DS, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol.* 2013;30:2725–2729.
62. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25:1451–1452.
63. Li H, Yi T, Gao L, Ma P, Zhang T, Yang J, et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat plants.* 2019;5:461–470.
64. Stull GW, De Stefano RD, Soltis DE, Soltis PS. Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. *Am J Bot.* 2015;102:1794–1813.
65. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol.* 2013;30:772–780.

66. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol Biol Evol.* 2016;34:772–773.
67. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003;19:1572–1574.
68. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of Gateway Computing Environments Workshop (GCE 2010). New Orleans, LA, USA. 2010. p. 1–8.
69. Stamatakis A, Hoover P, Rougemont J. A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Syst Biol.* 2008;57:758–771.
70. Wang L, Wuyun T, Du H, Wang D, Cao D. Complete chloroplast genome sequences of *Eucommia ulmoides*: genome structure and evolution. *Tree Genet Genomes.* 2016;12:12.
71. Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M, et al. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc Natl Acad Sci USA.* 1994;91:9794–9798.
72. Bruneau A, Doyle JJ, Palmer JD. A Chloroplast DNA Inversion as a Subtribal Character in the Phaseoleae (Leguminosae). *Syst Bot.* 1990;15:378–386.
73. Walker JF, Zanis MJ, Emery NC. Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *Am J Bot.* 2014;101:722–729.
74. Lemaire B, Vandamme P, Merckx VSFT, Smets E, Dessein S. Bacterial leaf symbiosis in angiosperms: host specificity without co-speciation. *PLoS One.* 2011;6:e24430.
75. Bell CD, Soltis DE, Soltis PS. The age and diversification of the angiosperms re-revisited. *Am J Bot.* 2010;97:1296–1303.
76. Doyle JJ, Doyle JL, Ballenger JA, Palmer JD. The Distribution and Phylogenetic Significance of a 50-kb Chloroplast DNA Inversion in the Flowering Plant Family Leguminosae. *Mol Phylogenet Evol.* 1996;5:429–438.
77. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. *PLoS One.* 2011;6:e19254.
78. Barrett CF, Davis JI, Leebensmack J, Conran JG, Stevenson DW. Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics.* 2013;29:65–87.
79. Barrett CF, Specht CD, Leebensmack J, Stevenson DW, Zomlefer WB, Davis JI, et al. Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical *gingers* (Zingiberales)? *Ann Bot-London.* 2014;113:119–133.
80. Yang L, Yang Z, Liu C, He Z, Zhang Z, Yang J, et al. Chloroplast phylogenomic analysis provides insights into the evolution of the largest eukaryotic genome holder, *Paris japonica* (Melanthiaceae). *BMC Plant Biol.* 2019;19:1–11.

81. Rokas A, Carroll SB. More Genes or More Taxa? The Relative Contribution of Gene Number and Taxon Number to Phylogenetic Accuracy. *Mol Biol Evol.* 2005;22:1337–1344.
82. Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Worheide G, et al. Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. *PLoS Biol.* 2011;9:e1000602.
83. Bremer B, Bremer K, Heidari N, Erixon P, Olmstead RG, Anderberg AA, et al. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Mol Phylogenet Evol.* 2002;24:274–301.
84. Bremer K, Friis EM, Bremer B. Molecular Phylogenetic Dating of Asterid Flowering Plants Shows Early Cretaceous Diversification. *Syst Biol.* 2004;53:496–505.

Figures

Figure 1

Comparison of IR/LSC and IR/SSC boundaries among Garryales plastomes.

Figure 2

Genomic rearrangements detected in the LSC of *Eucommia ulmoides* when compared with plastomes of *Aucuba* and other Lamiid representatives.

Figure 3

Alignment of three *Aucuba* plastomes using mVISTA, showing the percentages of sequence identity (Y axis).

Figure 4

Nucleotide diversity (Pi) in the complete plastomes of three *Aucuba* species. Sliding window analysis with a window length of 800 bp and a step size of 200 bp.

Figure 5

Phylogenetic relationships of *Aucuba* with other Lamiids reconstructed using Bayesian inference (BI) and maximum-likelihood (ML) analyses of 71 plastid protein-coding genes. Numbers above branches indicate posterior probabilities (PP) from the BI analyses and bootstrap percentages (BP) from the ML analyses.

Supplementary Files

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

- [Fig.S1.jpg](#)
- [TableS3.docx](#)
- [Fig.S1.jpg](#)
- [TableS2.docx](#)
- [TableS1.docx](#)
- [TableS1.docx](#)
- [TableS2.docx](#)
- [TableS3.docx](#)