

The complete chloroplast genome sequence of *Helwingia himalaica* (Helwingiaceae, Aquifoliales) and a chloroplast phylogenomic analysis of the Campanulidae

Xin Yao^{1,2}, Ying-Ying Liu³, Yun-Hong Tan¹, Yu Song^{1,2} and Richard T. Corlett¹

¹ Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Xishuangbanna, Yunnan, China

² University of Chinese Academy of Sciences, Beijing, Beijing, China

³ Key Laboratory of Dai and Southern medicine of Xishuangbanna Dai Autonomous Prefecture, Yunnan Branch Institute of Medicinal Plant, Chinese Academy of Medical Sciences, Jinghong, Yunnan, China

ABSTRACT

Complete chloroplast genome sequences have been very useful for understanding phylogenetic relationships in angiosperms at the family level and above, but there are currently large gaps in coverage. We report the chloroplast genome for *Helwingia himalaica*, the first in the distinctive family Helwingiaceae and only the second genus to be sequenced in the order Aquifoliales. We then combine this with 36 published sequences in the large (c. 35,000 species) subclass Campanulidae in order to investigate relationships at the order and family levels. The *Helwingia* genome consists of 158,362 bp containing a pair of inverted repeat (IR) regions of 25,996 bp separated by a large single-copy (LSC) region and a small single-copy (SSC) region which are 87,810 and 18,560 bp, respectively. There are 142 known genes, including 94 protein-coding genes, eight ribosomal RNA genes, and 40 tRNA genes. The topology of the phylogenetic relationships between Apiales, Asterales, and Dipsacales differed between analyses based on complete genome sequences and on 36 shared protein-coding genes, showing that further studies of campanulid phylogeny are needed.

Subjects Genomics, Plant Science

Keywords Asterids, Campanulidae, Phylogeny, Plastomes, Yunnan

INTRODUCTION

Complete chloroplast genome sequences (plastomes) have been very useful for understanding phylogenetic relationships in angiosperms at the family level and above, and have been used to resolve previously recalcitrant nodes (*Barrett et al., 2016*). However, there are currently large gaps in the coverage of orders and families. Within the ‘very large, very old, and very widespread’ subclass Campanulidae (*Beaulieu, O’Meara & Donoghue, 2013*; also known as ‘Asterids II’), complete chloroplast genome sequences are currently

Submitted 19 July 2016

Accepted 30 October 2016

Published 29 November 2016

Corresponding authors

Xin Yao, yaoxin@xtbg.org.cn

Richard T. Corlett,

corlett@xtbg.org.cn

Academic editor

Katharine Howell

Additional Information and
Declarations can be found on
page 15

DOI 10.7717/peerj.2734

© Copyright

2016 Yao et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

available only for 74 species (out of c. 35,000), in six families (out of 29) and four orders (out of seven). Within the campanulid order Aquifoliales, plastome sequences are currently known only from the large, monogeneric family Aquifoliaceae (Yao *et al.*, 2016).

Helwingia is the only genus in the campanulid family Helwingiaceae. It includes four species distributed in eastern Asia, from the Himalayas to Japan: *Helwingia chinensis* Batalin, *H. himalaica* Hook. f. & Thomson ex C.B. Clarke, *Helwingia japonica* (Thunb.) F. Dietr., and *Helwingia omeiensis* (W.P. Fang) H. Hara & S. Kurosawa (The Plant List, 2013; Wu, Raven & Hong, 2005). On current evidence, Helwingiaceae is sister to the Neotropical monogeneric family Phyllonomaceae (The Angiosperm Phylogeny Group, 2016), with which it shares an inferior ovary, epiphyllous inflorescence, and epigynous disc nectary (Ao & Tobe, 2015). These two small, highly disjunct, families are in turn sister to the near-cosmopolitan, but also monogeneric, family Aquifoliaceae.

Chloroplasts originated from free-living cyanobacteria via endosymbiosis and contain their own genome, which is circular and 76–217 kb in length (Hinsinger & Strijk, 2015; Zhang & Gao, 2016). Because of its abundance in plant cells and ease of sequencing, chloroplast DNA (cpDNA) has been widely utilized in studies of plant taxonomy and evolution (Kress *et al.*, 2005; Kress & Erickson, 2007; Newmaster, Fazekas & Ragupathy, 2006; Chase *et al.*, 2007; Taberlet *et al.*, 2007). The small size, single unit, haploid nature, and highly conserved genomic structure of cpDNA also make it useful for species identifications (Yang *et al.*, 2013). Moreover, the many copies per cell mean that useable fragments of the chloroplast genome are more likely to survive in dried herbarium specimens than are nuclear sequences, making direct comparisons with the genome of the type specimen potentially possible (Xu *et al.*, 2015).

The Helwingiaceae's current position in the order Aquifoliales, subclass Campanulidae (The Angiosperm Phylogeny Group, 2016), came after previous placements in the Cornaceae (Cronquist, 1981; Cronquist, 1988) and Araliaceae (Hutchinson, 1964; Hutchinson, 1973), and was based on molecular phylogenetic studies using *rbcL* (Morgan & Soltis, 1993), 18S rDNA and *rbcL* (Soltis & Soltis, 1997), and *ndhF* (Olmstead *et al.*, 2000). Sequencing the chloroplast genome will facilitate the development of additional chloroplast markers for identification and phylogenetic studies within the family, as well as providing a basis for future studies on the phylogenetics and biogeography of the order Aquifoliales. Beaulieu, Tank & Donoghue (2013) suggest that the initial divergence within this order took place in Australasia in the Cretaceous, with an early expansion into South America and Asia where *Phyllonoma* and *Helwingia*, respectively, persist today, while *Ilex* has spread more widely. In the absence of a fossil record for the two small families, a higher resolution phylogeny is needed to assess this hypothesis. The wider phylogenetic relationships among campanulid orders have been investigated in several studies, using chloroplast markers only (Beaulieu, Tank & Donoghue, 2013; Wikström *et al.*, 2015) or combined with nuclear ribosomal genes (ITS, 18S or 26S) (Tank & Donoghue, 2010; Beaulieu, O'Meara & Donoghue, 2013; Magallón *et al.*, 2015), but not yet with complete chloroplast genomes.

Here, we first explore the structure of the chloroplast genome in the Helwingiaceae using *H. himalaica*. We then investigate the phylogenetic relationships in the Campanulidae by using the complete genome sequences and the protein-coding genes shared between *H. himalaica* and other published genomes.

MATERIALS AND METHODS

Helwingia himalaica is distributed from Nepal through northern India to southwestern China. Plants materials used in this study were intact, fresh, young leaves collected in Bingzhongluo county of Yunnan province (28.015306°N, 98.607944°E). The specimen has been deposited in the herbarium of the Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (HITBC). Total genomic DNA was extracted from fresh leaves using a modified CTAB method (Doyle, 1987; Yang, Li & Li, 2014). Each amplification was performed in 25 μ L of a reaction mixture containing 1 \times PrimeSTAR GXL buffer (10 mM Tris-HCl (pH 8.2), 1 mM MgCl₂, 20 mM NaCl, 0.02 mM EDTA, 0.02 mM DTT; 0.02% Tween 20, 0.02% Nonidet P-40, and 10% glycerol); 1.6 mM of dNTPs, 0.5 μ M of each primer; 1.25 U of Prime-STAR GXL DNA polymerase (TAKARA BIO INC., Dalian, China), and 30–100 ng of DNA template. The amplification was conducted using 94 °C for 1 min, 30 cycles of 98 °C for 10 s and 68 °C for 15 min, followed by a final extension step at 72 °C for 10 min. The purified Polymerase chain reaction (PCR) product was fragmented and used for constructing the short-insert (500 bp) libraries according to the manufacturer's manual (Illumina). DNA of each sample was then indexed by tags and pooled together in one lane in an Illumina Hiseq 2000 to sequence (Yang, Li & Li, 2014).

Raw reads were filtered by quality control software NGSQCToolkit v2.3.3 (Patel & Jain, 2012) to obtain high quality Illumina data (cut-off value for percentage of read length = 80, cut-off value for PHRED quality score = 30) and vector- and adaptor-free reads. Filtered reads were assembled into contigs in CLC Genomics Workbench v.8 (<http://www.clcbio.com>) by the de novo method using a *k*-mer of 63 and a minimum contig length of 1 kb. Outputted contigs were aligned with the chloroplast genome of the asterid *Camellia yunnanensis* (GenBank accession number: KF156838), which was the most similar genome identified via BLAST (<http://blast.ncbi.nlm.nih.gov/>), and ordered according to the reference genome. Genes in the assembled chloroplast genome were predicted using Dual Organellar GenoMe Annotator (DOGMA) (Wyman, Jansen & Boore, 2004). The chloroplast genome was assembled using aligned contigs in Geneious v. 8.1.7 (<http://www.geneious.com>, Kearse et al., 2012). Junctions between large single-copy (LSC)/inverted repeats (IRs) and small single-copy (SSC)/inverted repeats (IRs) were validated by Sanger sequencing of PCR-based products (Table S1).

The assembled genome was annotated using the DOGMA database (Wyman, Jansen & Boore, 2004), then manually edited for start and stop codons. Genome maps were drawn in OGDRAW v.1.2 (Lohse et al., 2013). The annotated chloroplast genome has been submitted to GenBank (accession number: KX434807). REPuter was used to detect and assess repeats, including forward match, reverse match, complement match, and palindromic match repeats (Kurtz et al., 2001). Phobos v3.3.12 was used to detect simple

sequence repeats (SSRs) under default parameters (Mayer, Christoph, Phobos 3.3.11, 2006–2010; http://www.rub.de/spezzoo/cm/cm_phobos.htm). Mauve v. 2.4.0 was used for determining the chloroplast genome rearrangements among the campanulid families (*Darling et al., 2004*).

A matrix of chloroplast genome sequences, including *H. himalaica*, 36 other campanulid species, and *Coffea arabica* as an outgroup (EF044213 in GenBank) (Table 1), was aligned using MAFFT (*Katoh & Standley, 2013*) and manually edited where necessary. These 37 campanulid species represent all families and major clades within the Campanulidae that had complete chloroplast genome sequences in GenBank. Unambiguously aligned DNA sequences were used for phylogeny construction. Phylogenies were constructed by maximum likelihood (ML), Bayesian Inference analyses (BI), and maximum parsimony (MP).

ML analyses were conducted in RAxML version 8.2.8 (*Stamatakis, 2014*), using the GTACAT approximation. Convergence of the bootstrap was tested in RAxML using a posteriori bootstrapping analysis. BI analysis was conducted using MrBayes version 3.2.6 (*Ronquist et al., 2012*) and the best substitution model ('TVM+G') tested by AIC in jModelTest version 2.1.10 (*Darriba et al., 2012*). Four independent Markov Chain Monte Carlo algorithms were calculated for 10,000,000 generations and sampled every 1,000 generations. Potential Scale Reduction Factor (PSRF) values were used to determine convergence in BI using MrBayes version 3.2.6. All PSRF values were 1, indicating that these analyses converged. The first 25% of calculated trees was discarded as burn-in and a consensus tree constructed using the remaining trees. MP analysis was conducted PAUP version 4.0a150 (http://people.sc.fsu.edu/~dswofford/paup_test/), using the heuristic searches with tree bisection-reconnection (TBR) branch swapping and the 'Multrees' option in effect. Bootstrap analysis was conducted with 1,000 replicates with TBR branch swapping.

In addition, 36 protein-coding genes (Table 1) shared across all the 37 campanulid species were selected to build the phylogeny. ML analyses were conducted in RAxML version 8.2.8 (*Stamatakis, 2014*), using the GTACAT approximation. Convergence of the bootstrap was tested in RAxML using a posteriori bootstrapping analysis. BI analysis was conducted using MrBayes version 3.2.6 (*Ronquist et al., 2012*) and the best substitution model ('GTR+I+G') tested by AIC in jModelTest version 2.1.10 (*Darriba et al., 2012*). Methods for phylogeny construction using the 36 protein-coding genes follow the description above. PSRF values were used to determine convergence in BI using MrBayes version 3.2.6. All PSRF values were 1, indicating that these analyses converged. The first 25% of calculated trees was discarded as burn-in and a consensus tree constructed using the remaining trees. MP analysis was conducted in PAUP version 4.0a150 (http://people.sc.fsu.edu/~dswofford/paup_test/), using the heuristic searches with TBR branch swapping and the 'Multrees' option in effect. Bootstrap analysis was conducted with 1,000 replicates with TBR branch swapping.

Table 1 List of campanulid species (and the outgroup, *Coffea arabica*) and their accession numbers in GenBank included in the phylogenetic analyses of whole chloroplast genomes.

Species	Accession number in NCBI	Family	Order	Length (bp)	Coding gene	tRNA	rRNA	GC (%)	LSC (bp)	SSC (bp)	IRs (bp)
<i>Angelica acutiloba</i>	KT963036	Apiaceae	Apiales	147,074	85	35	8	37.5	93,367 (63.48)	17,573 (11.95)	36,134 (24.57)
<i>Anthriscus cerefolium</i>	GU456628	Apiaceae	Apiales	154,719	85	37	8	37.4	84,768 (54.79)	17,551 (11.34)	52,400 (33.87)
<i>Bupleurum falcatum</i>	KM207676	Apiaceae	Apiales	155,989	84	37	8	37.7	85,870 (55.05)	17,518 (11.23)	52,601 (33.72)
<i>Crithmum maritimum</i>	HM596072	Apiaceae	Apiales	158,355	88	37	8	37.6	85,230 (53.82)	27,993 (17.68)	55,986 (35.35)
<i>Daucus carota</i>	DQ898156	Apiaceae	Apiales	155,911	85	43	8	37.7	84,244 (54.03)	17,571 (11.27)	54,096 (34.70)
<i>Foeniculum vulgare</i>	KR011054	Apiaceae	Apiales	153,628	85	37	8	37.6	86,659 (56.41)	17,470 (11.37)	49,499 (32.22)
<i>Ligusticum tenuissimum</i>	KT963039	Apiaceae	Apiales	158,500	88	37	8	37.6	84,875 (53.55)	17,661 (11.14)	55,964 (35.31)
<i>Ostericum grosserratum</i>	KT852844	Apiaceae	Apiales	147,282	83	36	8	37.5	93,185 (63.27)	17,663 (11.99)	36,434 (24.74)
<i>Petroselinum crispum</i>	HM596073	Apiaceae	Apiales	152,890	84	37	8	37.8	86,116 (56.33)	17,508 (11.45)	49,266 (32.22)
<i>Tiedemannia filiformis</i> subsp. <i>greenmannii</i>	HM596071	Apiaceae	Apiales	154,737	85	37	8	37.3	84,585 (54.66)	17,140 (11.08)	53,012 (34.26)
<i>Dendropanax dentiger</i>	KP271241	Araliaceae	Apiales	156,687	87	37	8	38.0	86,680 (55.32)	18,247 (11.65)	51,760 (33.03)
<i>Hydrocotyle verticillata</i>	HM596070	Araliaceae	Apiales	153,207	85	37	8	37.6	84,352 (55.06)	18,739 (12.23)	50,116 (32.71)
<i>Kalopanax septemlobus</i>	KC456167	Araliaceae	Apiales	156,413	87	37	8	37.9	86,467 (55.28)	18,118 (11.58)	51,828 (33.14)
<i>Panax ginseng</i>	AY582139	Araliaceae	Apiales	156,318	87	37	8	38.1	86,114 (55.09)	18,070 (11.56)	52,134 (33.35)
<i>Ilex delavayi</i>	KX426470	Aquifoliaceae	Aquifoliales	157,671	95	40	8	37.6	87,000 (55.18)	18,436 (11.69)	52,234 (33.13)
<i>Ilex latifolia</i>	KX426465	Aquifoliaceae	Aquifoliales	157,610	95	40	8	37.6	86,952 (55.17)	18,429 (11.69)	52,228 (33.14)
<i>Ilex</i> new species	KX426469	Aquifoliaceae	Aquifoliales	157,611	95	40	8	37.6	86,948 (55.17)	18,434 (11.70)	52,227 (33.14)
<i>Ilex polyneura</i>	KX426468	Aquifoliaceae	Aquifoliales	157,621	95	40	8	37.6	87,064 (55.24)	18,435 (11.70)	52,122 (33.07)

(Continued)

Table 1 (continued).

Species	Accession number in NCBI	Family	Order	Length (bp)	Coding gene	tRNA	rRNA	GC (%)	LSC (bp)	SSC (bp)	IRs (bp)
<i>Ilex pubescens</i>	KX426467	Aquifoliaceae	Aquifoliales	157,741	95	40	8	37.6	87,109 (55.22)	18,436 (11.69)	52,238 (33.12)
<i>Ilex szechwanensis</i>	KX426466	Aquifoliaceae	Aquifoliales	157,822	95	40	8	37.7	87,204 (55.25)	18,513 (11.73)	52,182 (33.06)
<i>Ilex wilsonii</i>	KX426471	Aquifoliaceae	Aquifoliales	157,918	95	40	8	37.6	87,266 (55.26)	18,432 (11.67)	52,222 (33.07)
<i>Helwingia himalaica</i>	KX434807	Helwingiaceae	Aquifoliales	158,362	94	40	8	37.7	87,810 (55.45)	18,560 (11.72)	51,991 (32.83)
<i>Artemisia frigida</i>	JX293720	Asteraceae	Asterales	151,076	87	37	8	37.5	82,740 (54.77)	18,392 (12.17)	49,944 (33.06)
<i>Aster spathulifolius</i>	KF279514	Asteraceae	Asterales	149,510	87	37	8	37.7	81,961 (54.82)	17,972 (12.02)	49,577 (33.16)
<i>Centaurea diffusa</i>	KJ690264	Asteraceae	Asterales	152,559	90	36	8	37.7	83,596 (54.80)	18,487 (12.12)	50,476 (33.09)
<i>Chrysanthemum indicum</i>	JN867592	Asteraceae	Asterales	151,129	85	35	8	37.4	82,885 (54.84)	18,376 (12.16)	49,868 (33.00)
<i>Cynara cornigera</i>	KP842707	Asteraceae	Asterales	152,550	87	37	8	37.7	83,580 (54.79)	18,660 (12.23)	50,310 (32.98)
<i>Lactuca sativa</i>	DQ383816	Asteraceae	Asterales	152,772	86	44	8	37.5	84,105 (55.05)	18,599 (12.17)	50,068 (32.77)
<i>Lashtenia burkei</i>	KM360047	Asteraceae	Asterales	150,944	67	25	7	37.4	82,193 (54.45)	18,271 (12.10)	50,480 (33.44)
<i>Parthenium argentatum</i>	GU120098	Asteraceae	Asterales	152,803	57	17	8	37.6	84,593 (55.36)	18,900 (12.37)	49,310 (32.27)
<i>Praxelis clematidea</i>	KF922320	Asteraceae	Asterales	151,410	84	32	8	37.2	85,311 (56.34)	18,559 (12.26)	47,540 (31.40)
<i>Adenophora remotiflora</i>	KP889213	Campanulaceae	Asterales	171,724	82	37	8	38.8	105,555 (61.47)	11,295 (6.58)	54,874 (31.95)
<i>Campanula takedimana</i>	KP006497	Campanulaceae	Asterales	169,551	83	36	8	38.8	102,320 (60.35)	7,747 (4.57)	59,484 (35.08)
<i>Hanabusaya asiatica</i>	KJ477692	Campanulaceae	Asterales	167,287	82	37	10	38.8	104,955 (62.74)	8,578 (5.13)	53,754 (32.13)
<i>Trachelium caeruleum</i>	EU090187	Campanulaceae	Asterales	162,321	83	44	10	38.3	100,110 (61.67)	7,661 (4.72)	54,550 (33.61)
<i>Kolkwitzia amabilis</i>	KT1966716	Caprifoliaceae	Dipsacales	156,875	81	38	8	38.4	90,137 (57.46)	18,846 (12.01)	47,892 (30.53)
<i>Lonicera japonica</i>	KJ1170923	Caprifoliaceae	Dipsacales	155,078	81	39	8	38.6	88,858 (57.30)	18,672 (12.04)	47,548 (30.66)
<i>Coffea arabica</i>	EF044213	Rubiaceae	Gentianales	155,189	85	45	8	37.4	85,164 (54.88)	18,207 (11.73)	51,818 (33.39)

Note:

Numbers in parentheses in the LSC, SSC and IRs columns are the percentage of the total length.

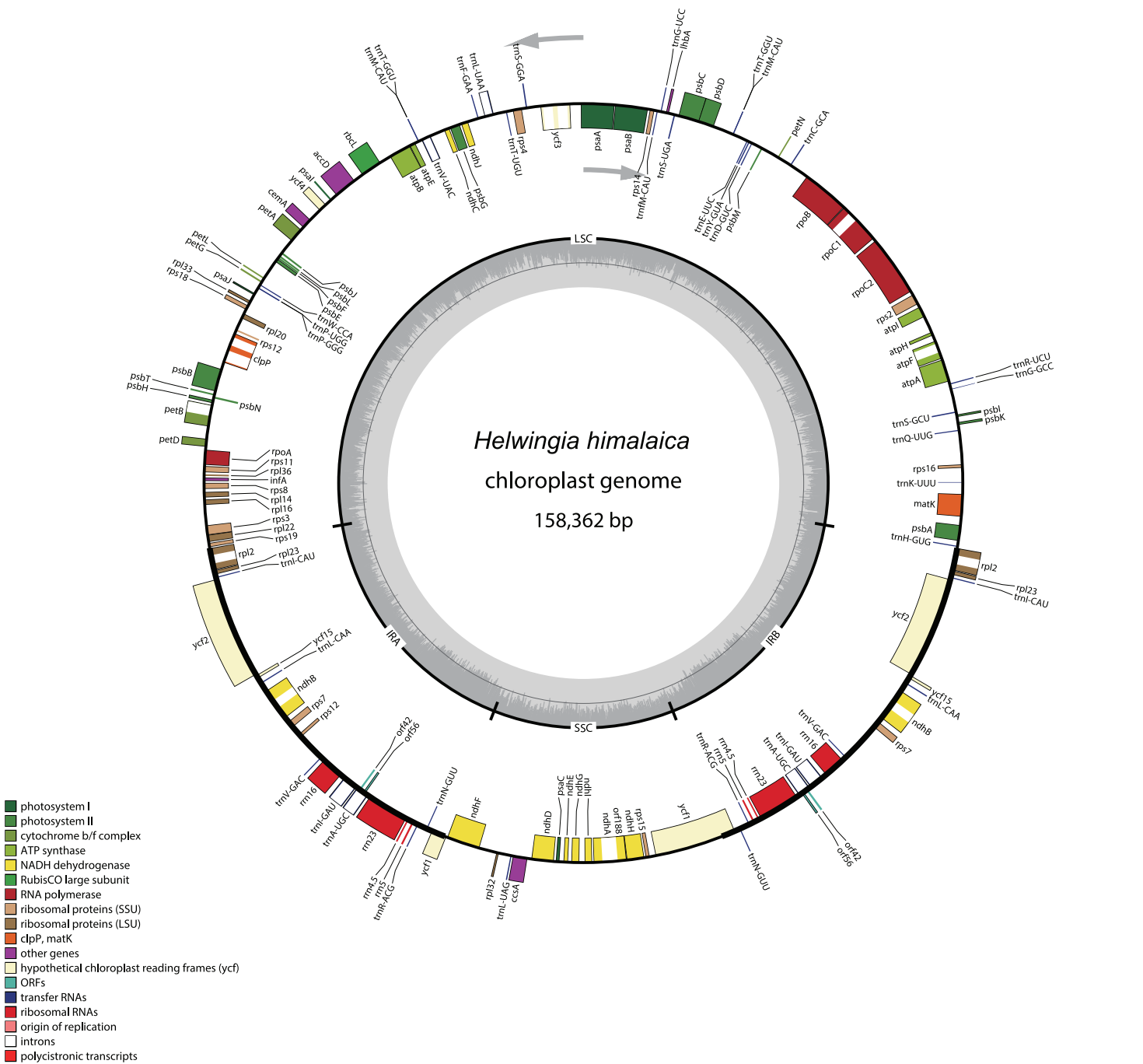


Figure 1 Circular gene map of the *Helwingia himalaica* chloroplast genome. Genes placed outside of the outer circle are transcribed in clockwise direction whereas genes inside are transcribed in counterclockwise direction. Different colours refer to genes from different functional groups. The area in darker gray in the inner circle indicates GC content while the lighter gray indicates AT content.

RESULTS

Genome features

The total length of the chloroplast genome is 158,362 bp. Its quadripartite structure includes an LSC with 87,810 bp and SSC with 18,560 bp, separated by a pair of IR

Table 2 List of genes in the chloroplast genome of *Helwingia himalaica*.

Category	Groups of gene	Name of genes
Protein synthesis and DNA-replication	Transfer RNAs	<i>trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnI-M-CAU, trnG-GCC, trnG-UCC, trnH-GUG, trnK-UUU, trnL-UAA, trnM-CAU, trnQ-UUG, trnP-GGG, trnP-UGG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-UAC, trnW-CCA, trnY-GUA, trnA-UGC(×2), trnI-CAU(×2), trnI-GAU(×2), trnL-CAA(×2), trnL-UAG, trnN-GUU(×2), trnR-ACG(×2), trnV-GAC(×2)</i>
	Ribosomal RNAs	<i>rrn16(×2), rrn23(×2), rrn4.5(×2), rrn5(×2)</i>
	Ribosomal protein small subunit	<i>rps16, rps2, rps14, rps4, rps18, rps12(×2), rps11, rps8, rps3, rps19, rps7(×2), rps15</i>
	Ribosomal protein large subunit	<i>rpl33, rpl20, rpl36, rpl14, rpl16, rpl22, rpl2(×2), rpl23(×2), rpl32</i>
	Subunits of RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>
Photosynthesis	photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbG, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, lhba</i>
	Cytochrome b/f complex	<i>petA, petB, petD, petG, petL, petN</i>
	ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
	NADH-dehydrogenase	<i>ndhA, ndhB(×2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Large subunit rubisco	<i>rbcl</i>
Miscellaneous group	Translation initiation factor	<i>infA</i>
	Acetyl-CoA carboxylase	<i>accD</i>
	Cytochrome c biogenesis	<i>ccsA</i>
	Maturase	<i>matK</i>
	ATP-dependent protease	<i>clpP</i>
	Inner membrane protein	<i>cemA</i>
Pseudogene unknown function	Conserved hypothetical chloroplast ORF	<i>ycf3, ycf4, ycf2(×2), ycf15(×2), orf42(×2), orf56(×2), ycf1(×2), orf188</i>

regions with lengths of 25,996 bp (Fig. 1). The GC content is 37.7% (see Table 1 in Yao *et al.*, 2016). A total of 102 unique genes were detected in the chloroplast genome, of which 20 were duplicated in IR regions. Totally, 94 protein-coding genes (76 unique) encode proteins acting in processes related to photosynthesis, the genetic system, and some currently unknown functions (e.g. *ycf*). In addition, 40 genes (26 unique) encode for tRNAs and eight genes for rRNAs (Table 2). All eight rRNA genes are in IR regions. One *ycf1* gene is a functional pseudogene as it is on the border between the SSC and IRa region. Gene *rps19* is outside the IRb region at the LSC-IRb junction and *rpl2* is fully included in the IRa region. Five genes (*atpF*, *rpoC1*, *rpl2*, *ndhB* and *ndhA*) have one intron and two genes have two introns (*ycf3*, *clpP* and *rps12*).

Repeated sequences and SSR

Thirty repeated sequences were detected, with lengths ranging from 18 to 43 bp and sequence identity more than 90% (Table 3). Among them, 19 repeated sequences were

Table 3 List of repeated sequences in the chloroplast genome of *Helwingia himalaica*.

Repeat length (bp)	Repeat A start site	Repeat A location*	Repeat A region	Repeat B start site	Repeat B location	Repeat B region	Repeat type**
43	0	<i>rpl2(trnH-GUG)</i>	IRa	87797	<i>rps19(rpl2)</i>	IRb	P
30	9030	<i>trnS-GCU</i>	LSC	47728	<i>trnS-GGA</i>	LSC	P
27	45989	<i>ycf3 intron2</i>	LSC	124368	<i>ndfA intron</i>	SSC	F
26	43	<i>rpl2(trnH-GUG)</i>	LSC	87772	<i>rps19(rpl2)</i>	IRb	P
26	10811	<i>trnG-GCC(trnR-UCU)</i>	LSC	10840	<i>trnG-GCC(trnR-UCU)</i>	LSC	P
26	33886	<i>trnT-GGU(psbD)</i>	LSC	33912	<i>trnM-CAU(psbD)</i>	LSC	F
26	91380	<i>ycf2</i>	IRb	154796	<i>ycf2</i>	IRa	F
23	61775	<i>accD</i>	LSC	61786	<i>accD</i>	LSC	F
21	9036	<i>trnS-GCU</i>	LSC	37766	<i>trnS-UGA</i>	LSC	F
21	37766	<i>trnS-UGA</i>	LSC	47731	<i>trnS-GGA</i>	LSC	P
21	38950	<i>trnM-CAU</i>	LSC	69860	<i>trnP-UGG</i>	LSC	F
20	38564	<i>lhbA(trnG-UCC)</i>	LSC	38581	<i>lhbA(trnG-UCC)</i>	LSC	F
20	49313	<i>trnT-UGU(trnL-UAA)</i>	LSC	49333	<i>trnT-UGU(trnL-UAA)</i>	LSC	F
19	385	<i>trnH-GUG(psbA)</i>	LSC	412	<i>trnH-GUG(psbA)</i>	LSC	P
19	6791	<i>rps16(trnQ-UUG)</i>	LSC	6817	<i>rps16(trnQ-UUG)</i>	LSC	F
19	8756	<i>psbI</i>	LSC	38919	<i>trnG-UCC(trnfM-CAU)</i>	LSC	P
19	10620	<i>trnG-GCC</i>	LSC	38738	<i>trnG-UCC</i>	LSC	F
19	15636	<i>atpH(atpI)</i>	LSC	15653	<i>atpH(atpI)</i>	LSC	F
19	34034	<i>trnT-GGU(psbD)</i>	LSC	111456	<i>orf56(trnR-ACG)</i>	IRb	R
19	34034	<i>trnT-GGU(psbD)</i>	LSC	134727	<i>trnR-ACG(trnA-UGC)</i>	IRa	C
19	53790	<i>ndhC(trnV-UAC)</i>	LSC	81476	<i>rpoA</i>	LSC	P
19	59571	<i>rbcL</i>	LSC	59590	<i>rbcL(accD)</i>	LSC	F
18	4719	<i>trnK-UUU</i>	LSC	66845	<i>petA(psbI)</i>	LSC	R
18	5785	<i>rps16(trnQ-UUG)</i>	LSC	34036	<i>trnT-GGU(psbD)</i>	LSC	R
18	6349	<i>rps16(trnQ-UUG)</i>	LSC	90529	<i>ycf2</i>	IRb	F
18	6349	<i>rps16(trnQ-UUG)</i>	LSC	155655	<i>ycf2</i>	IRa	P
18	9101	<i>trnS-GCU</i>	LSC	37836	<i>trnS-UGA</i>	LSC	F
18	40424	<i>psaB</i>	LSC	42639	<i>psaA</i>	LSC	F
18	40973	<i>psaB</i>	LSC	43197	<i>psaA</i>	LSC	F
18	57793	<i>atpB(rbcL)</i>	LSC	121171	<i>ndhD(psaC)</i>	SSC	C

Notes:

* *rpl2(trnH-GUG)* means spacer between *rpl2* and *trnH-GUG*, etc.

** P means palindromic match, F means forward (direct) match, R means reverse match, and C means complement match.

dispersed in intergenic regions, 10 in genes, and one in introns. There were 16 forward repeats, nine palindromic repeats, three reverse repeats, and two complement repeats, and 21, 2 and 7 repeats were detected in the LSC, SSC and IRs, respectively. A total of 813 SSRs were found, including 289 mononucleotides, 35 dinucleotides and 70 trinucleotides (Fig. S1). In mononucleotide SSRs, thymine and adenine made up 92% (266). In dinucleotide SSRs, we found repeated units consisting of TA/AT and GA/AG, but no GC/CG and TC/CT repeats.

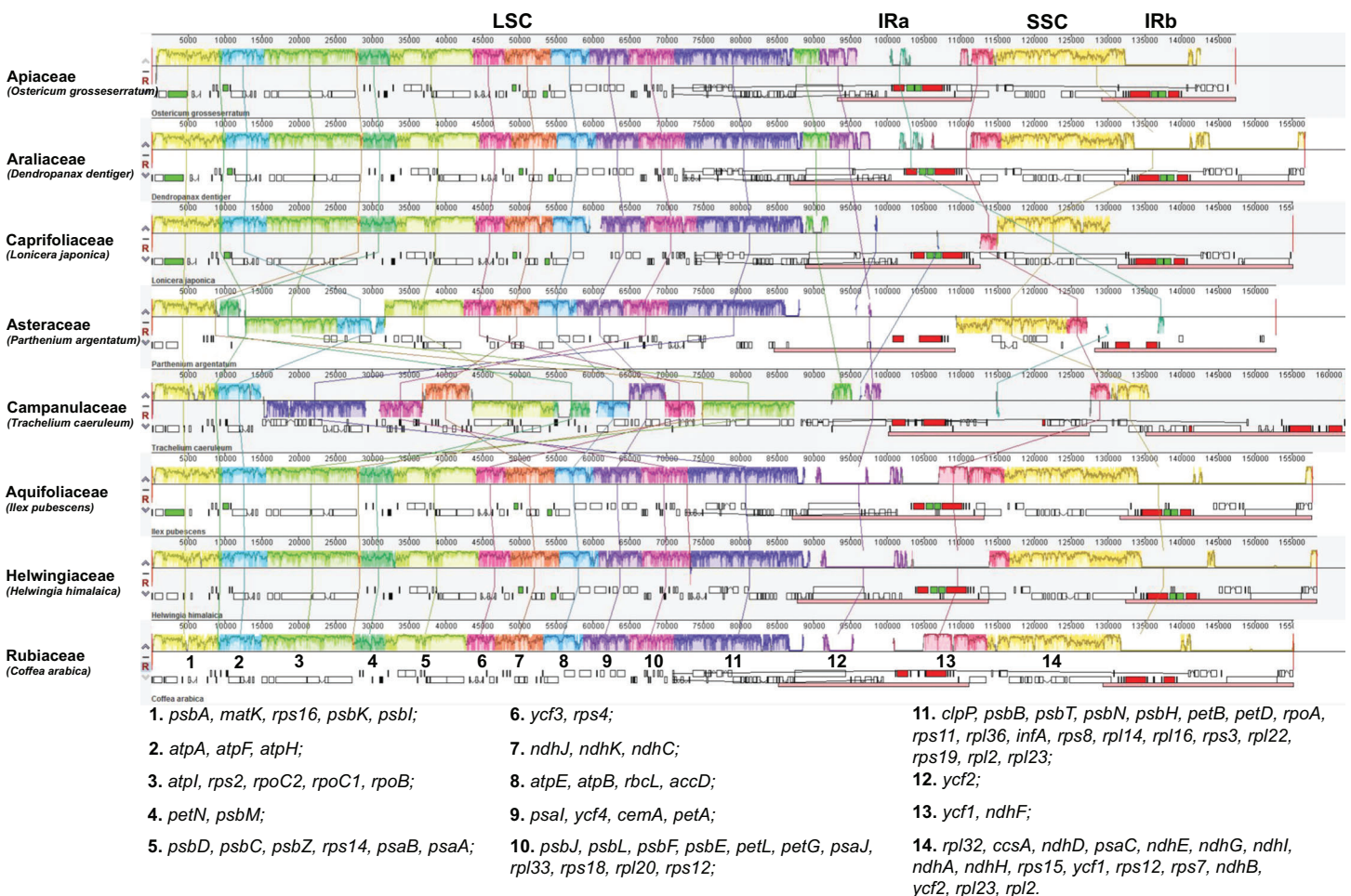


Figure 2 Gene arrangement map of chloroplast genome alignment of seven representative species from seven campanulid families and *Coffea arabica* (as a reference) determined by Mauve software (Darling et al., 2004). The polyline in the blocks indicates sequence similarity among these eight species. Line linking gene blocks among the eight species with same colour indicates ortholog. Gene blocks above are transcribed clockwise and those below are transcribed counterclockwise. The coding genes in the 14 main gene blocks are listed under the figure.

Genome rearrangement in the Campanulidae

Genome alignment among seven species from the seven campanulid families with known chloroplast genomes revealed massive gene rearrangement, especially in the LSC (Fig. 2). Moreover, all four Campanulaceae species had longer genomes and LSCs, and shorter SSCs, compared with other campanulid species (Table 1). The IR in the chloroplast genome of some Apiaceae (*Angelica acutiloba*, *Foeniculum vulgare*, *Ostericum grosseserratum* and *Petroselinum crispum*) was contractive (Table 1; Fig. 2). Even though the lowest number of coding genes in any campanulid species was 57 (*Parthenium argentatum*), only 36 coding genes were shared across all the campanulid families (Table 4), Kumar et al. (2009) which indicates many gene losses or gains had occurred. The number of tRNA ranged from 17 (*P. argentatum*) to 44 (*Lactuca sativa* and *Trachelium caeruleum*), while the number of rRNA was usually eight (Table 1).

Table 4 The 36 protein-coding genes shared by the 37 campanulid species and used for construction of the protein-coding gene phylogeny.

Gene	Length (bp)	Gene	Length (bp)	Gene	Length (bp)
<i>atpA</i>	1,539	<i>psaB</i>	2,205	<i>psbT</i>	144
<i>atpH</i>	246	<i>psaC</i>	246	<i>rbcL</i>	1,458
<i>atpI</i>	744	<i>psaI</i>	113	<i>rpl14</i>	417
<i>cemA</i>	708	<i>psaJ</i>	135	<i>rpl20</i>	415
<i>ndhC</i>	363	<i>psbA</i>	1,062	<i>rpl32</i>	207
<i>ndhD</i>	1,516	<i>psbD</i>	1,062	<i>rpl33</i>	207
<i>ndhE</i>	306	<i>psbF</i>	120	<i>rpl36</i>	114
<i>ndhJ</i>	477	<i>psbH</i>	222	<i>rps2</i>	747
<i>petA</i>	963	<i>psbI</i>	111	<i>rps4</i>	618
<i>petG</i>	114	<i>psbK</i>	186	<i>rps8</i>	435
<i>petL</i>	96	<i>psbM</i>	117	<i>rps11</i>	418
<i>psaA</i>	2,253	<i>psbN</i>	132	<i>rps18</i>	336

Phylogenetic analyses of the Campanulidae

The phylogeny produced from the analysis of 37 complete chloroplast genomes is well-supported, but while the results from ML and BI are congruent, the phylogeny from MP is not (Figs. 3A and 3B). With ML and BI, Aquifoliales are basal, Asterales are the next branch, and the Dipsacales are sister to the Apiales. The six families with multiple species are all well-supported (Fig. 3A). In the MP phylogeny, however, the Dipsacales are sister to the Asterales, and the Apiales are the next branch (Fig. 3B).

The phylogeny based on 36 shared protein-coding genes has a consistent family-level topology in analyses with BI, ML, and MP. The Aquifoliales are still basal, but the Dipsacales are the next branch, and the Asterales are sister to the Apiales (Fig. 4). However, within the Asteraceae, the topology from BI is different from those from ML and MP, and MP also did not resolve the relationships of *Ilex wilsonii* and *Ilex szechwanensis* (Fig. 4C).

DISCUSSION

The only other published chloroplast genomes in the order Aquifoliales are for seven species of *Ilex* in the family Aquifoliaceae (Yao *et al.*, 2016). The length of the *Helwingia* genome (158,362 bp) is similar to that of *Ilex* (157,610–157,918 bp) (see Table 1 in Yao *et al.*, 2016). *Helwingia* (94 genes) has two fewer protein coding genes than *Ilex* (96) while both have the same number of tRNA (40) and rRNA (eight) genes.

Minor differences among the published chloroplast genomes are common, including gene loss or gain events, but these do not separate *H. himalaica* from the others. Both *H. himalaica* and *Helianthus annuus* have *ycf15* (Timme, 2009), but *Guizotia abyssinica* does not (Dempewolf *et al.*, 2010). *H. himalaica* and *L. sativa* have *trnE* (Kanamoto *et al.*, 2004) but *H. annuus* and *G. abyssinica* do not. Gene *rps16* has one intron in *G. abyssinica*, *H. annuus*, and *L. sativa*, but none in *H. himalaica*, while gene *ycf3* has two introns in *H. himalaica*, *G. abyssinica*, and *H. annuus*, but none in *L. sativa*. Gene *ycf15* is between

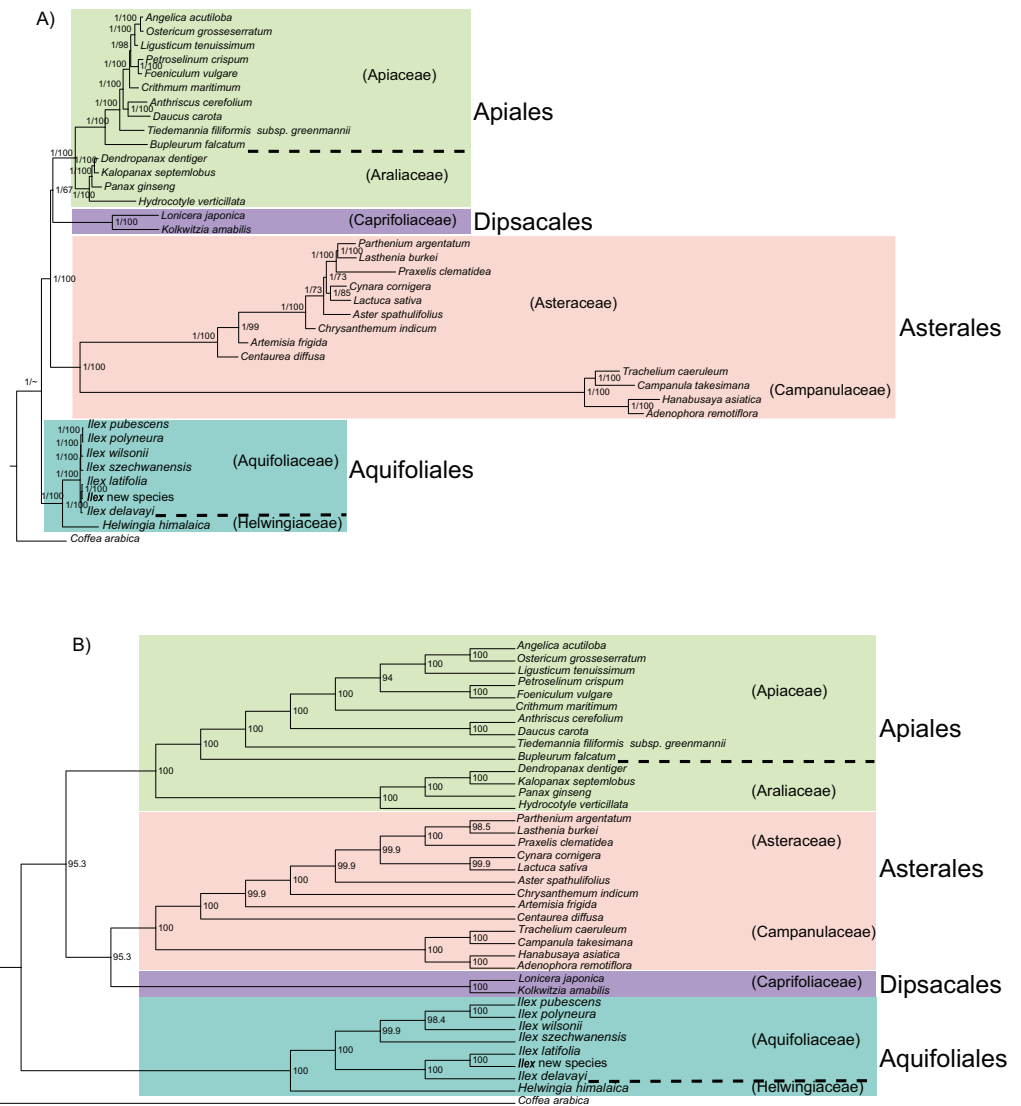


Figure 3 Phylogeny of 37 campanulid species using their complete chloroplast genomes. In subgraph (A) numbers near nodes (on left) indicate the Bayesian posterior probability and numbers near nodes (on right) indicate the maximum likelihood bootstrap values for each clade present in the 50% majority-rule consensus tree. In subgraph (B) numbers near nodes indicate the maximum parsimony bootstrap values for each clade present in the 50% majority-rule consensus tree.

ycf2 and *trnL-CAA* in *H. himalaica*, but between *rps7* and *trnV-GAC* in *H. annuus*. Gene *ndhF* is in the IRb-SSC junction in both *L. sativa* and *H. himalaica*, but in the SSC-IRa junction in *G. abyssinica* and *H. annuus*. The lengths of the published chloroplast genomes for the Campanulaceae range from 162,321 bp (*T. caeruleum*, Haberle et al., 2008) to 171,724 bp (*Adenophora remotiflora*, Kim et al., 2016) and are longer than those of other campanulid species. Kim et al. (2016) attribute this longer length to expansion occurring in the IR and LSC regions as well as the gene arrangements.

The many mononucleotide SSRs identified in *H. himalaica* are potentially useful for studies of the evolutionary history of populations (Khadivi-Khub et al., 2014; Chae et al., 2014). The dominance of A/T in mononucleotide SSRs in *Helwingia* is similar to

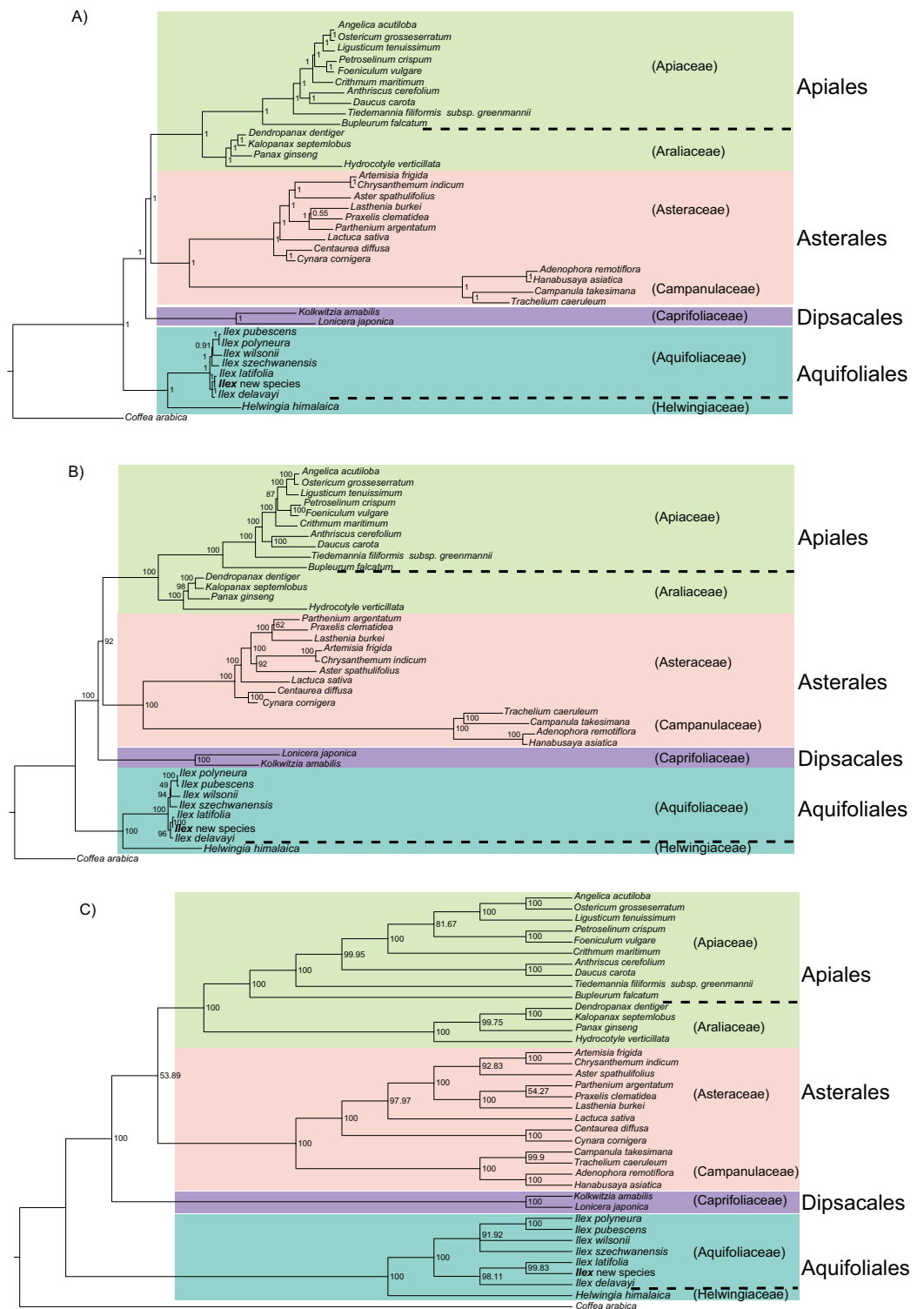


Figure 4 Phylogeny of 37 campanulid species using their 36 shared coding genes. In subgraph (A) numbers near nodes indicate the Bayesian posterior probability. In subgraph (B) numbers near nodes indicate the maximum likelihood bootstrap values for each clade present in the 50% majority-rule consensus tree. In subgraph (C) numbers near nodes indicate the maximum parsimony bootstrap values for each clade present in the 50% majority-rule consensus tree.

other published studies ([Huang et al., 2014](#); [Kuang et al., 2011](#)). It has been suggested that repeated sequences play an important role in genomic rearrangement and sequence variation in chloroplast genomes ([Huang et al., 2014](#); [Yang et al., 2013](#)). Approximately 63% of repeats were found in intergenic regions which are often also divergent hotspot regions (e.g. [Yao et al., 2015](#); [Huang et al., 2014](#)), showing the potential of these regions for the development of new phylogenetic markers for species identification in *Helwingia* and related genera in the Aquifoliales.

Massive rearrangements in the chloroplast genome have been identified in the Campanulaceae in comparison with other campanulid families ([Fig. 2](#)). Except for gene block 1 and 2, most gene blocks in the LSC have been rearranged, including changes in gene order and transcribing direction ([Fig. 2](#)). The chloroplast gene rearrangement in Campanulaceae was first identified in *T. caeruleum*, and inferred as the effects of recombination of repeats or tRNA genes ([Haberle et al., 2008](#)). Comparing with other angiosperm chloroplast genomes, more repeats and tRNA genes occurred near rearrangement endpoints in this species. The positive connection between rearrangement and repeated sequences has also been found in other plants, like *Arbutus unedo* ([Martínez-Alberola et al., 2013](#)), Geraniaceae ([Weng et al., 2013](#)), *Vaccinium macrocarpon* ([Fajardo et al., 2013](#)) and cupressophytes ([Wu & Chaw, 2014](#)). However, the effects of these chloroplast gene rearrangements on plant physical functions still need more study.

The phylogenetic trees based on complete chloroplast genomes are incongruent with those from the protein-coding genes. Aquifoliales are basal in all phylogenetic analyses, but the phylogenetic relationships among the Asterales, Apiales and Dipsacales differ in different analyses ([Figs. 3](#) and [4](#)). The phylogeny based on complete chloroplast genomes using BI and ML methods found that the Apiales are sister to the Dipsacales ([Fig. 3A](#)), which agrees with recent phylogenies for this subclass based on other markers ([Beaulieu, Tank & Donoghue, 2013](#); [Wikström et al., 2015](#); [Chen et al., 2016](#); [The Angiosperm Phylogeny Group, 2016](#)). However, using the MP method with the same data resulted in a phylogeny with the Asterales sister to the Dipsacales ([Fig. 3B](#)). The phylogenies based on protein-coding genes found that the Apiales are sister to Asterales with all three methods, although the topology within the Asteraceae differed between BI and the other two methods ([Fig. 4](#)). Three orders (of seven) and 22 families (of 29) in the subclass Campanulidae could not be included in our analyses because there are no published complete chloroplast genomes for these clades, which emphasizes the need for increased coverage of angiosperm orders and families in future studies of chloroplast genomes.

CONCLUSION

We report the chloroplast genome of *H. himalaica* as the first in the Helwingiaceae and the second genus in the Aquifoliales. It has the typical quadripartite circular structure, including an LSC with 87,810 bp and an SSC with 18,560 bp, separated by a pair of IR regions with 25,996 bp. In total, 142 genes were detected in this genome, consisting of 94 protein-coding genes, 40 tRNA, and eight rRNA. Repeated sequences are mainly

distributed in intergenic regions. Comparisons among the available chloroplast genomes within the campanulids reveal massive chloroplast gene rearrangement in the Campanulaceae. The phylogenetic relationships among Apiales, Asterales and Dipsacales were incongruent between phylogenetic results produced from complete chloroplast genomes and the 36 shared protein-coding genes. The topology within Asteraceae also varied, which shows that further studies are still needed in these three orders. The results of this study will facilitate understanding of not only the family Helwingiaceae and its relationships with other taxa in the Aquifoliales, but also phylogenetic relationships within the angiosperms at higher levels.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Jing Yang, Juan-Hong Zhang, Chun-Yan Lin and Ji-Xiong Yang from the Kunming Institute of Botany, Chinese Academy of Sciences, for their help with experiments.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by grants from the 1000 Talents Program (WQ20110491035). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
1000 Talents Program: WQ20110491035.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Xin Yao conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Ying-Ying Liu performed the experiments, analyzed the data, reviewed drafts of the paper.
- Yun-Hong Tan performed the experiments, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Yu Song performed the experiments, analyzed the data, reviewed drafts of the paper.
- Richard T. Corlett conceived and designed the experiments, wrote the paper, reviewed drafts of the paper.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:
GenBank [KX434807](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.2734#supplemental-information>.

REFERENCES

- Ao C, Tobe H. 2015. Floral morphology and embryology of *Helwingia* (Helwingiaceae, Aquifoliales): systematic and evolutionary implications. *Journal of Plant Research* 128(1):161–175 DOI 10.1007/s10265-014-0672-9.
- Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L, Medina J, Pires JC, Santos C, Wm. Stevenson D, Zomlefer WB, Davis JI. 2016. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytologist* 209(2):855–870 DOI 10.1111/nph.13617.
- Beaulieu JM, O'Meara BC, Donoghue MJ. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62(5):725–737 DOI 10.1093/sysbio/syt034.
- Beaulieu JM, Tank DC, Donoghue MJ. 2013. A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. *BMC Evolutionary Biology* 13(1):80 DOI 10.1186/1471-2148-13-80.
- Chae WB, Hong SJ, Gifford JM, Rayburn AL, Sacks EJ, Juvik JA. 2014. Plant morphology, genome size, and SSR markers differentiate five distinct taxonomic groups among accessions in the genus *Miscanthus*. *GCB Bioenergy* 6(6):646–660 DOI 10.1111/gcbb.12101.
- Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, Seberg O, Jørgensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough T, Kelly L, Wilkinson MJ. 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon* 56(2):295–299.
- Chen ZD, Yang T, Lin L, Lu LM, Li HL, Sun M, Liu B, Chen M, Niu YT, Ye JF, Cao ZY, Liu H-M, Wang X-M, Wang W, Zhang J-B, Meng Z, Cao W, Li J-H, Wu S-D, Zhao H-L, Liu Z-J, Du Z-Y, Wang Q-F, Guo J, Tan X-X, Su J-X, Zhang L-J, Yang L-L, Liao Y-Y, Li M-H, Zhang G-Q, Chung S-W, Zhang J, Xiang K-L, Li R-Q, Soltis DE, Soltis PS, Zhou S-L, Ran J-H, Wang X-Q, Jin X-H, Chen Y-S, Gao T-G, Li J-H, Zhang S-Z, Lu A-M, China Phylogeny Consortiumjse. 2016. Tree of life for the genera of Chinese vascular plants. *Journal of Systematics and Evolution* 54(4):277–306 DOI 10.1111/jse.12219.
- Cronquist A. 1981. *An Integrated System of Classification of Flowering Plants*. New York: Columbia University Press.
- Cronquist A. 1988. *The Evolution and Classification of Flowering Plants*. New York: New York Botanical Garden.
- Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14(7):1394–1403 DOI 10.1101/gr.2289704.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9(8):772 DOI 10.1038/nmeth.2109.
- Dempewolf H, Kane NC, Ostevik KL, Geleta M, Barker MS, Lai Z, Stewart ML, Bekele E, Engels JMM, Cronk QC, Rieseberg LH. 2010. Establishing genomic tools and resources for *Guizotia abyssinica* (Lf) Cass.—the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome. *Molecular Ecology Resources* 10(6):1048–1058 DOI 10.1111/j.1755-0998.2010.02859.x.

- Doyle JJ. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19:11–15.
- Fajardo D, Senalik D, Ames M, Zhu H, Steffan SA, Harbut R, Polashock J, Vorsa N, Gillespie E, Kron K, Zalapa JE. 2013. Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. *Tree Genetics & Genomes* 9(2):489–498 DOI 10.1007/s11295-012-0573-9.
- Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *Journal of Molecular Evolution* 66(4):350–361 DOI 10.1007/s00239-008-9086-4.
- Hinsinger DD, Strijk JS. 2015. Complete chloroplast genome sequence of *Castanopsis concinna* (Fagaceae), a threatened species from Hong Kong and South-Eastern China. *Mitochondrial DNA* 18:1–2 DOI 10.3109/19401736.2015.1110800.
- Huang H, Shi C, Liu Y, Mao S-Y, Gao L-Z. 2014. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evolutionary Biology* 14(1):151 DOI 10.1186/1471-2148-14-151.
- Hutchinson J. 1964. *The Genera of Flowering Plants. Dicotylédones*. Vol. 1. Oxford: Clarendon Press.
- Hutchinson J. 1973. *The Families of Flowering Plants: Arranged According to a New System Based on their Probable Phylogeny*. Oxford: Clarendon Press, 968.
- Kanamoto H, Yamashita A, Okumura S, Hattori M, Tomizawa KI. 2004. The complete genome sequence of the *Lactuca sativa* (lettuce) chloroplast. In: *Plant and Cell Physiology*. Vol. 45. England: Oxford University Press, S39–S39.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30(4):772–780 DOI 10.1093/molbev/mst010.
- 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 DOI 10.1093/bioinformatics/bts199.
- Khadivi-Khub A, Zamani Z, Fattahi R, Wünsch A. 2014. Genetic variation in wild *Prunus* L. subgen. *Cerasus* germplasm from Iran characterized by nuclear and chloroplast SSR markers. *Trees* 28(2):471–485 DOI 10.1007/s00468-013-0964-z.
- Kim K-A, Cheon K-S, Jang S-K, Yoo K-O. 2016. Complete chloroplast genome sequence of *Adenophora remotiflora* (Campanulaceae). *Mitochondrial DNA Part A* 27(4):2963–2964 DOI 10.3109/19401736.2015.1060461.
- Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE* 2(6):e508 DOI 10.1371/journal.pone.0000508.
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 102(23):8369–8374 DOI 10.1073/pnas.0503123102.
- Kuang D-Y, Wu H, Wang Y-L, Gao L-M, Zhang S-Z, Lu L. 2011. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome* 54(8):663–673 DOI 10.1139/g11-026.
- Kumar S, Hahn FM, McMahan CM, Cornish K, Whalen MC. 2009. Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification

- of DNA barcodes to differentiate *Parthenium* species and lines. *BMC Plant Biology* **9**(1):131 DOI 10.1186/1471-2229-9-131.
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001.** REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* **29**(22):4633–4642 DOI 10.1093/nar/29.22.4633.
- Lohse M, Drechsel O, Kahlau S, Bock R. 2013.** OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* **41**(W1):W575–W581 DOI 10.1093/nar/gkt289.
- Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. 2015.** A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**(2):437–453 DOI 10.1111/nph.13264.
- Martínez-Alberola F, del Campo EM, Lázaro-Gimeno D, Mezquita-Claramonte S, Molins A, Mateu-Andrés I, Pedrola-Monfort J, Casano LM, Barreno E. 2013.** Balanced gene losses, duplications and intensive rearrangements led to an unusual regularly sized genome in *Arbutus unedo* chloroplasts. *PLoS ONE* **8**(11):e79685 DOI 10.1371/journal.pone.0079685.
- Morgan DR, Soltis DE. 1993.** Phylogenetic relationships among members of Saxifragaceae *sensu lato* based on *rbcl* sequence data. *Annals of the Missouri Botanical Garden* **80**(3):631.
- Newmaster SG, Fazekas AJ, Ragupathy S. 2006.** DNA barcoding in land plants: evaluation of *rbcl* in a multigene tiered approach. *Botany* **84**(3):335–341 DOI 10.1139/b06-047.
- Olmstead RG, Kim K-J, Jansen RK, Wagstaff SJ. 2000.** The phylogeny of the Asteridae *sensu lato* based on chloroplast *ndhF* gene sequences. *Molecular Phylogenetics and Evolution* **16**(1):96–112 DOI 10.1006/mpev.1999.0769.
- Patel RK, Jain M. 2012.** NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* **7**(2):e30619.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012.** MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**(3):539–542 DOI 10.1093/sysbio/sys029.
- Soltis D, Soltis P. 1997.** Phylogenetic relationships in Saxifragaceae *sensu lato*: a comparison of topologies based on 18S rDNA and *rbcl* sequences. *American Journal of Botany* **84**(4):504.
- Stamatakis A. 2014.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9):1312–1313 DOI 10.1093/bioinformatics/btu033.
- Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E. 2007.** Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35**(3):e14 DOI 10.1093/nar/gkl938.
- Tank DC, Donoghue MJ. 2010.** Phylogeny and phylogenetic nomenclature of the Campanulidae based on an expanded sample of genes and taxa. *Systematic Botany* **35**(2):425–441 DOI 10.1600/036364410791638306.
- The Angiosperm Phylogeny Group. 2016.** An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**(1):1–20 DOI 10.1046/j.1095-8339.2003.t01-1-00158.x.
- The Plant List. 2013.** Version 1.1. Available at <http://www.theplantlist.org/> (accessed 1 January).
- Timme RE. 2009.** *A Comparison of the First Two Sequenced Chloroplast Genomes in Asteraceae: Lettuce and Sunflower*. Berkeley: Lawrence Berkeley National Laboratory.

- Weng M-L, Blazier JC, Govindu M, Jansen RK. 2013. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. *Molecular Biology and Evolution* 31(3):645–659 DOI 10.1093/molbev/mst257.
- Wikström N, Kainulainen K, Razafimandimbison SG, Smedmark JEE, Bremer B. 2015. A revised time tree of the Asterids: establishing a temporal framework for evolutionary studies of the coffee family (Rubiaceae). *PLoS ONE* 10(5):e126690 DOI 10.1371/journal.pone.0126690.
- Wu CS, Chaw SM. 2014. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. *Plant Biotechnology Journal* 12(3):344–353 DOI 10.1111/pbi.12141.
- Wu ZY, Raven PH, Hong DY. 2005. *Flora of China (Apiaceae through Ericaceae)*. Vol. 14. Beijing: Science Press.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20(17):3252–3255 DOI 10.1093/bioinformatics/bth352.
- Xu C, Dong W, Shi S, Cheng T, Li C, Liu Y, Wu P, Wu H, Gao P, Zhou S. 2015. Accelerating plant DNA barcode reference library construction using herbarium specimens: improved experimental techniques. *Molecular Ecology Resources* 15(6):1366–1374 DOI 10.1111/1755-0998.12413.
- Yang J-B, Li D-Z, Li H-T. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Molecular Ecology Resources* 14(5):1024–1031 DOI 10.1111/1755-0998.12251.
- Yang JB, Yang SX, Li HT, Jang J, Li DZ. 2013. Comparative chloroplast genomes of Camellia species. *PLoS ONE* 8(8):e73053.
- Yao X, Tan Y-H, Liu Y-Y, Song Y, Yang J-B, Corlett RT. 2016. Chloroplast genome structure in *Ilex* (Aquifoliaceae). *Scientific Reports* 6:28559 DOI 10.1038/srep28559.
- Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. 2015. The first complete chloroplast genome sequences in Actinidiaceae: genome structure and comparative analysis. *PLoS ONE* 10(6):e129347 DOI 10.1371/journal.pone.0129347.
- Zhang Q-J, Gao L-Z. 2016. The complete chloroplast genome sequence of desert poplar (*Populus euphratica*). *Mitochondrial DNA Part A* 27(1):721–723 DOI 10.3109/19401736.2014.913159.