Whole Genome based Insights into the Phylogeny and Evolution of the Juglandaceae

Huijuan Zhou
Northwest A&F University: Northwest Agriculture and Forestry University
Yiheng Hu
Northwestern University
Aziz Ebrahimi
Purdue University
Peiliang Liu
Northwestern University
Keith Woeste
Purdue University
Shuoxin Zhang
Northwest A&F University: Northwest Agriculture and Forestry University
Peng Zhao (pengzhao@nwu.edu.cn)
Northwest University https://orcid.org/0000-0003-3033-6982

Research article

Keywords: Diversification, Divergence time, Genome, Juglandaceae, Phylogenomics, Plastome

Posted Date: May 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-495294/v1

Abstract

Background: The walnut family (Juglandaceae) contains commercially important woody trees commonly called walnut, wingnut, pecan and hickory. Phylogenetic relationships in the Juglandaceae are problematic, and their historical diversification has not been clarified, in part because of low phylogenetic resolution and/or insufficient marker variability.

Results: We reconstructed the backbone phylogenetic relationships of Juglandaceae using organelle and nuclear genome data from 27 species. The divergence time of Juglandaceae was estimated to be 78.7 Mya. The major lineages diversified in warm and dry habitats during the mid-Paleocene and early Eocene. The plastid, mitochondrial, and nuclear phylogenetic analyses all revealed three subfamilies, i.e., Juglandoideae, Engelhardioideae, Rhoipteleoideae. Five genera of Juglandoideae were strongly supported. Juglandaceae were estimated to have originated during the late Cretaceous, while Juglandoideae were estimated to have originated during the Paleocene, with evidence for rapid diversification events during several glacial and geological periods. The phylogenetic analyses of organelle sequences and nuclear genome yielded highly supported incongruence positions for *J. cinerea*, *J. hopeiensis*, and *Platycarya strobilacea*. Winged fruit were the ancestral condition in the Juglandoideae, but adaptation to novel regeneration regimes after the Creaceous-Paleogene boundary led to the independent evolution of zoochory among several genera of the Juglandaceae.

Conclusions: A fully resolved, strongly supported, time-calibrated phylogenetic tree of Juglandaceae can provide an important framework for studying classification, diversification, biogeography, and comparative genomics of plant lineages.

Introduction

Phylogenomics applies genomic data to reconstruct the evolutionary biology of organisms [1–3], including the resolution of evolutionary relationships among and within family clades [4–7], genera, and closely related species [8–10]. Next generation sequencing (NGS) has made the generation of large-scale genomic data easier, cheaper, and greatly increased the availability complete chloroplast genomes [6, 9] and whole genome resequencing data [11]. The plastid genome has provided insight into molecular phylogeny and evolutionary relationships at many taxonomic levels [4, 9, 12–13]. Foundational studies of the Juglandaceae were based on analysis of selected loci [14–16]. Whole genome scale studies can be useful supplements to previous research – and in some cases necessary – to resolve evolutionary relationships where sequence variation is limited by taxonomic level, early divergence, large difference in morphology, rapid speciation or slow genome evolution [7, 17–19].

The walnut family (Juglandaceae) is distributed in both the Old and New World, from North and South America to southeastern Europe, eastern Asia, and Japan, from S10°to N49°[20–27] (Fig. 1). All species of Juglandaceae are perennial woody plants [28–32]. The accepted phylogeny for the Fagales shows the Juglandaceae family is monophyletic and most closely related to the Myricaceae [28–30]. The Juglandaceae are lumped with five other families (Betulacea, Fagacea, Casuarinaceae, Nothofagaceae, and Ticodendraceae) to constitute the order Fagales [28, 31, 33–38].

The Juglandaceae contains ten living genera (*Juglans*, *Pterocarya*, *Cyclocarya*, *Platycarya*, *Carya*, *Annamocarya*, *Engelhardia*, *Alfaroa*, *Oreomunnea*, and *Rhoiptelea*) comprised of ca 50 ~ 60 total species [21, 22, 36–42]. Members of the family are considered some of the most important nut, medicinal, and timber trees. The phylogenetic relationships among and within genera of Juglandaceae are a complex puzzle that has been the subject of numerous studies [16, 24, 30–31, 35–42]. Comparative morphology – primarily interpretation of the floral parts - was used to develop the classically accepted taxonomy and phylogeny of the family [21, 22, 39–45].

Although studies based on a limited number of loci (chloroplast DNA fragments) and fossil evidence have greatly advanced our understanding of Juglandaceae [16, 22–24, 34–35], some relationships within *Juglans*, *Carya*, and *Pterocarya* are weakly supported or conflicting; especially the relationship of *Platycarya* to *Carya*, and the position of *Cyclocarya* and *Pterocarya* in relation to *Juglans* [16, 43]. Other issues include the placement of the Rhoipteleaceae, a monotypic family containing only the species *Rhoiptelea chiliantha* [35, 46–47]. It was placed in the Juglandaceae by APG III (2009) system (Fig. 1) [48]. Similarly, the genus *Annamocarya* contains only one species, *A. sinensis*. Manos and Stone (2001) [26] suggested sectional

recognition within *Carya* might be appropriate for this taxon, but they also found that it shares a number of characteristics with the walnuts (genus *Juglans*).

The evolution of the Juglandaceae remains a difficult problem too; hypothesized to have both ancient and recent extinctions and radiations [21, 49–50], it is species poor. The species that remain, however, are highly divergent in their ecology (wind versus animal-dispersed fruit) [30, 43], and flower development [22, 40].

The primary goal of this study was to increase the resolution of the molecular phylogeny of the Juglandaceae by maximizing the number of taxa sampled and the number of genetic markers [22, 27, 30]. We selected 27 Juglandaceae taxa, slightly more than half of the ~ 50 recognized species from three subfamilies (Engelhardioideae, Juglandoideae, and Rhoipteleoideae), and from seven of the nine worldwide genera, making this the most comprehensive study to date. We used sequence data from matrilineally (chloroplast genomes and mitochondrial protein-coding genes) and biparentally (whole genome re-sequencing of nuclear genome SNPs) inherited DNA to illuminate the evolutionary history of the Juglandaceae. We also reanalyzed phylogenetic relationships of 55 species using ITS (Internal transcribed spacers) sequences. Our goal was to 1) reconstruct the phylogenetic relationships of the family Juglandaceae based on whole chloroplast genomes, whole genome re-sequencing of nuclear genome SNPs (nrSNPs), ITS, and sixteen mitochondrial protein-coding genes (mtCDS), with an eye toward the major unresolved systematic questions in this family, 2) compare the plastid genomes of Juglandaceae, and identify the location and extent of genetic variation in these genomes across within the Juglandaceae, 3) reconstruct a time-calibrated phylogeny of the Juglandaceae based on whole chloroplast genomes, 4) reveal the timing of diversification for important nodes within the family.

Materials And Methods

Taxon Sampling, Genomic DNA Extractions, Library, and Sequencing

We analyzed twenty-seven species of Juglandaceae from seven genera that span the taxonomic, geographic, and morphological range of the family. These were contextualized using published plastomes of nine species of Fagales (include four species for Betulaceae, and five species for Fagaceae), three species of Cucurbitales, and four species of Rosales (Table S1). The voucher specimens were deposited in the herbarium of Key Laboratory of Resource Biology and Biotechnology in Western China (Ministry of Education), Northwest University (Table 1). We collected fresh leaf samples from field, and the samples were stored in air tight bags filled with silica gel desiccant for later DNA extraction.

Table 1
Taxa and voucher information for plant material from Juglandaceae plastomes.

Species	Total Length	GC %	LSC	SSC	IR	Raw reads	Mapped reads	Sequencing Platform	GenBank No.
Carya cathayensis	160,300	36.2	89,715	18,553	26,016	7,391,021	92,365	Illumina Hiseq2500	MH189594
Carya hunanensis	160,397	36.2	89,807	18,532	26,029	9,303,790	94,317	Illumina Hiseq2500	MH188303
Carya illinoensis	160,585	36.2	90,030	18,435	26,060	9,652,336	70,573	Illumina Hiseq2500	MH188302
Carya kweichwensis	159,780	36.3	89,264	18,430	26,043	9,087,431	236,036	Illumina Hiseq2500	MH188301
Carya sinensis	160,195	36.3	89,541	18,538	26,085	13,878,699	420,540	Illumina Hiseq2500	KX703001
Cyclocarya paliurus	160,562	36.1	90,007	18,477	26,039	9,073,816	277,193	Illumina Hiseq2500	KY246947
Engelhardia roxburghiana	161,164	36	90,477	18,531	26,087	10,435,597	261,440	Illumina Hiseq2500	MH188300
Platycarya strobilacea	158,281	36.1	87,990	18,399	25,946	12,345,252	79,584	Illumina Hiseq2500	MH189595
Pterocarya fraxinifolia	160,246	36.2	89,783	18,437	26,013	9,731,800	187,013	Illumina Hiseq2500	MH188291
Pterocarya hupehensis	159,770	36.2	89,229	18,505	26,018	9,605,591	273,616	Illumina Hiseq2500	MH188293
Pterocarya insignis	160,207	36.2	89,728	18,476	26,006	9,805,922	156,606	Illumina Hiseq2500	MH188292
Pterocarya macroptera var. insignis	159,941	36.2	89,517	18,410	26,007	9,164,994	113,662	Illumina Hiseq2500	MH188290
Pterocarya stenoptera	160,202	36.2	89,727	18,433	26,021	11,542,884	186,936	Illumina Hiseq2500	MH188289
Pterocarya tonkinensis	160,096	36.2	89,600	18,482	26,007	7,449,017	160,611	Illumina Hiseq2500	MH188288
Pterocarya stenoptera var. zhijiangensis	160,174	36.2	89,640	18,510	26,012	9,295,098	179,137	Illumina Hiseq2500	MH188304
Juglans ailantifolia	160,353	36.1	89,931	18,376	26,023	12,918,979	145,907	Illumina Hiseq2500	MH188299
Juglans cathayensis	159,730	36.1	89,333	18,351	26,023	13,178,238	344,843	Illumina Hiseq2000	KX671976
Juglans cinerea	160,193	36.2	89,719	18,406	26,034	13,018,105	720,145	Illumina Hiseq2500	MH188298
Juglans hindsii	159,929	36.2	89,597	18,296	26,018	29,975	6,646	Roche 454	MH188297

Note: LSC = Large single copy, SSC = Small single copy, IR = Inverted repeat. Length of regions is given in number of base pairs (bp).

Species	Total Length	GC %	LSC	SSC	IR	Raw reads	Mapped reads	Sequencing Platform	GenBank No.
Juglans major	160,221	36.1	89,766	18,372	26,034	9,407,594	513,705	Illumina Hiseq2500	MH188296
Juglans mandshurica	159,729	36.1	89,331	18,346	26,023	11,805,821	527,970	Illumina Hiseq2000	KX671975
Juglans hopeiensis	159,714	36.1	89,872	18,406	26,036	12,382,845	517,928	Illumina Hiseq2000	KX671977
Juglans microcarpa	160,065	36.2	89,637	18,383	26,022	85,030	22,781	Roche 454	MH188295
Juglans nigra	160,301	36.1	89,840	18,393	26,034	13,178,283	434,844	Illumina Hiseq2500	MH188294
Juglans regia	160,367	36.1	89,872	18,425	26,035	3,511,124	1,713,581	Illumina Miseq	KT963008
Juglans sigillata	160,350	36.1	89,872	18,406	26,036	12,225,897	402,317	Illumina Hiseq2000	KX424843

Note: LSC = Large single copy, SSC = Small single copy, IR = Inverted repeat. Length of regions is given in number of base pairs (bp).

Total genomic DNA was extracted from 200 mg of silica gel-dried leaves using a modified CTAB (cetrimonium bromide) method [51–52]. The DNA concentration was quantified using a NanoDrop spectrophotometer (Thermo Scientific, Carlsbad, CA, USA). A paired-end (PE) library with 350 bp insert size was constructed using the Illumina PE DNA library kit according to the manufacturer's instructions and sequenced using an Illumina Hiseq2500 by Novogene (www.novogene.com, China).

Mitochondrion Protein-coding Genes (mtCDS) Primer Design and PCR Amplification

We investigated genetic variation within mitochondrial protein-coding genes (mtCDS) to evaluate phylogenetic relationships of the Juglandaceae. We used a total of sixteen primers designed from the complete mitochondrion sequence of *Populus tremula* (NCBI accession number: KT337313.1) using Primer3 (Sangon Biotech in Shanghai, China). Primers were targeted to the sequence of mitochondrial protein-coding genes of *P. tremula* (Table S2). PCR amplification was carried out on a SimpliAmp Thermal Cycler (Applied Biosystem, USA) in 20 μL reaction volumes (10 μL 2× PCR Master Mix including 0.1 U Taq polymerase/μL; 500 μM each dNTP; 20 mMTris-HCl (pH8.3); 100 mMKCl; 3.0mM $MgCl_2$ (Tiangen, Beijing, China),0.5μL each primer, 2μL BSA, 2μL of 10 ng/μL DNA). The PCR was programmed for 3 min at 94°C followed by 35 cycles of 15 s at 93°C, 1 min at annealing temperature (Table S2), 30 s at 72°C and extension of 10 min at 72°C. After PCR amplification, fragments were sequenced by Sangon Biotech (Shanghai, China).

Plastomes Assembly and Annotation

The sequenced and assembled plastomes were quality controlled using the NGSQC toolkit v2.3.3 trim tool to remove low quality reads, unknown bases, adapter sequences, and sequencing errors [53]. Short reads were assembled into long contigs using SPAdes Genomic Assembler v3.6.0 [54], followed by manual checking and finishing. The contigs were assembled in Geneious v8.0.2 [55]. To exclude nuclear DNA, we used BLAST to remove contigs that did not align to a reference plastome from *J. regia* (Genbank accession number KT963008) [56]. A reference-based assembly allowed us to reconstruct each of all other species [13].

After we identified the boundaries between the inverted repeats (IR) and the single copy regions, i.e., the Large Single Copy (LSC) and Small Single Copy (SSC) regions, the completed plastomes were annotated using the online software DOGMA based on the *J. regia* reference [56–57]. We manually annotated start and stop codons and other regions of interest using Geneious v8.0.2 [55]. A circular representation of each plastome was visualized in OGDraw [58]. Finally, gene content, order,

and variability were analyzed in Geneious and R [59]. The plastid genomes data were deposited in National Center for Biotechnology Information (NCBI), the accession numbers were KX703001 to KX703038 (Table 1).

Variant Calling

Using paired-end (2 × 150 bp) Illumina sequencing, we obtained a high sequencing depth (> 30×) per sample based on alignment with the *J. regia* reference plastome [56]. After aligning the re-sequencing reads, we processed the alignments to remove duplicate reads and applied a series of quality control filters with the intent of limiting false-positive variants. Sequence reads passing Illumina's quality control filter were aligned using bwa-mem algorithm of BWA v0.7.12 [60] and then mapped to the *J. regia* plastid genome. Only uniquely mapped reads were retained, which removed the repeat region IR. Duplicate reads were removed from individual sample alignments using Picard tools v2.5.0 (https://github.com/broadinstitute/picard) Mark Duplicates function and assigned genomic positions for each accession based on the alignment files generated by SAMtools v0.1.19 [61].

The SNPs (single nucleotide polymorphisms) and small Indels (insertion-deletion) among Juglandaceae plastid genome accessions were identified if they were supported by at least three mapped reads. Following bwa-mem mapping, the rest of the sequencing pipeline was performed using the toolkit GATK v3.5.0 [62]. Reads present in areas surrounding Indels were realigned using the built-in function Indel Realigner, after which SNPs were called using Unified Genotyper. Finally, a series of quality filters were applied to reduce systematic errors, including quality-by-depth ratio (QD) < 10, ReadPosRankSum < − 8.0, depth coverage (DP) ≥ 30, probability of strand bias (FS) > 10.0, SNPs that passed these filters were kept for subsequent analyses. Finally, we use the stats module in the bcftools v1.1 to count SNPs and Indels and calculate Ts/Tv (transition/transversion) rates [63].

In this study, we called the nuclear SNPs from all samples of Juglandaceae (Table S3). The Illumina paired-end reads from each sample were first processed to remove adaptor and low-quality sequences using Trimmomatic [64]. The cleaned unique reads were aligned to the common walnut reference genome version 1.0 (https://treegenesdb.org/FTP/Genomes/Jure/) using BWA [61, 65], and only uniquely mapped reads were retained. Following mapping, genotypes were assigned to each genomic position for each sample based on the alignment files generated by SAMtools [60]. Single nucleotide polymorphism (SNPs) and small indels (insertion and deletion) in the 27 samples were identified using GATK [62]. The redundant reads were then filtered based on the location of clean reads in the reference genome using software Picard (Picard: http://sourceforge.net/projects/picard/). We used GATK's Haploype Caller (local haplotype assembly) algorithm for SNPs and InDels based on each sample.

Partition Strategy and Phylogenetic Analysis

To infer the evolutionary relationships among the 27 Juglandaceae plastomes and to test the phylogenetic signal from different regions of the plastomes, we reconstructed the Juglandaceae phylogeny using the following four datasets based on the exons of protein-coding genes; to avoid large amounts of missing data in the phylogenetic analyses, sixty-one protein coding genes that were shared by all 44 taxa were extracted and aligned (Table S4). Best-fit partitioning schemes and models were selected using the greedy search mode implemented in PartitionFinder v2.1.1 [66] (Table S5).

Plastomes were aligned using default settings in MAFFT v7.245 [67]. The resulting alignments were manually checked in Geneious v8.0.2 [55]. The best-fit nucleotide substitution model for all our plastome data sets was determined (as suggested by Modeltest v3.7 with the Akaike information criterion (AIC) [68–69]. A concatenated data set was analyzed using Bayesian Inference (BI) and Maximum Likelihood (ML) analysis in MrBayes v3.2.6 [70] or RAxML v8.1.24 [71]. BI trees were produced by MrBayes v3.2.6 set at 10,000,000 generations. Two independent Markov chain Monte Carlo (MCMC) chains were run, each with one cold chain and three incrementally heated chains. Trees were sampled every 10,000 generations, with the first 25 % of the trees discarded as burn-in. Stationarity was considered reached when the average standard deviation of split frequencies was < 0.01. The Maximum Likelihood (ML) trees were generated using RAxML v8.1.24 using a GTRGAMMA model [71]. For ML analysis, difference general time reversible models were performed with all data sets. For all analyses, 10 independent ML searches were conducted, bootstrap support was estimated with 1,000 bootstrap replicates, and bootstrap

(BS) proportions were drawn on the tree with highest likelihood score from the 10 independent searches. We generated multiple mtCDS sequence alignments using ClustalX with default parameters [72]. The phylogenetic tree analysis was performed using MEGA7 [73].

For the phylogenetic tree analysis based on nuclear genome data, we selected a total of 1,161,468 SNPs with minor allele frequency (MAF) ≥ 5 % and missing rate per site ≤ 10% for phylogenetic analyses. A Maximum Likelihood (ML) tree was constructed using RAxML v8.1.24 in 1,000 bootstrap replicates [71]. In order to gain a better understanding of the species relationships, we selected 55 species to represent all extant genera in the Juglandaceae for which internal transcribed spacer (ITS) sequence data are available in NCBI (Table S6). We generated multiple ITS sequence alignments using ClustalX with default parameters [72], and a phylogenetic tree analysis using Maximum Likelihood analysis [71].

Divergence-time Estimation and Fossil Calibration

Penalized likelihood (PL) dating analyses were conducted using the treePL v1.0 program [74]. This program allows for better optimization with large trees by combining stochastic optimization with hill-climbing gradient-based methods. To identify the appropriate level of rate heterogeneity in the phylogram, a data-driven cross-validation analysis was conducted with treePL v1.0. One thousand bootstrap replicates with branch lengths were also generated using RAxML v8.1.24 for calculating the confidence age intervals with TreeAnnotator as implemented in BEAST v2.4.5 [75]. The phylogenetic trees were then compiled into a maximum clade credibility tree using Tree Annotator v1.8.0 [76]. The program FigTree v1.3.1 (http://tree.bio.ed.ac.uk) was used to visualize mean node ages and highest posterior density (HPD) intervals at 95% (upper and lower) for each node and to estimate branch lengths and divergence times. Six fossils were also used for calibration (Table S7) [24–25, 28, 77].

Results

Plastome Sequencing, Assembly, and Characteristics of Juglandaceae Plastome

All Juglandaceae plastomes were entirely syntenic, non-recombining circular genomes with conserved gene content and gene order (Table 1, Fig. 2). The Juglandaceae plastome has a mean length of 160,150 bp, and ranged from 158,281 bp (*Platycarya strobilacea*) to 160,585 (*Carya illinoinensis*) with four main parts; a large single-copy region (LSC; 87,900 – 90,477 bp), a small single-copy region (SSC; 18,296 – 18,533 bp), and two inverted repeat regions (IRs; 25,946 – 26,242 bp) (Table 1; Fig. 2; Table S1). The coding regions contained 137 genes, including 81 protein coding genes (eight duplicated in the IR), 33 tRNA genes (seven duplicated in the IRs) and four rRNA genes (four duplicated in the IRs) (Fig. 2). There were four introns (*rpl2*, *rpl16*, *rps16*, and *rpoC1*) located in the IRs region and 13 introns in the LSC region in each of the plastomes (Fig. S1). Seven tRNA genes, *trnI-CAU*, *trnL-CAA*, *trnV-GAC*, *trnI-GAU*, *trnA-UGC*, *trnR-ACG*, and *trnN-GUU* were duplicated and scattered in the inverted repeat (Fig. 2). We aligned each of the protein-coding genes (CDS) of all species. Three potential pseudogenes (*infA*, *rpl22*, and *ycf15*) were identified and their sequence verified using Sanger sequencing (Shagon Biotech, Shanghai, China) (Fig. S2; Table S8).

Variant Analysis

Comparison of the whole chloroplast genome sequences revealed a total of 18,050 SNPs and 2,496 Indels (insertions and deletions), for a total of 6,594 high-quality non-redundant variant positions, or approximately 5.66 SNPs/kb (Table 2; Fig. 2; Fig. S1). A total of 4,228 variant positions (64 %) were found in intergenic regions. The remaining variants affected 88 genes, leaving 41 genes unaffected (Table 2). Several regions were remarkably variable, including *matK* (68.0 SNPs per kb), *ndhD* (56.5), *ndhF* (53.5), *rpoC2* (39.1), *rpoB* (26.5), *accD* (46.8), and *ycf1* (101.5). A total of 1,161,468 SNPs were identified from whole genome sequencing data (Table S3) based on comparison with a *J. regia* reference genome [65]. The SNPs number, mapping ratio, heterozygosity, and heterozygosity ratio ranged from 202,314 to 1,143,008, 17.81–98.45%, 166 to 540,829, and 3.54–54.61%, respectively (Table S3; Fig. S3)

Table 2

Summary of variants from all Juglandaceae genomes based on comparison with *Juglans regia* whole genome sequences.

Species	cp-SNPs	nr-SNPs	cp-Indels	Ts/Tv ratio	Mapped	Heterozygosity	Homozygosity	Het-ratio
Juglans ailantifolia	433	443,059	134	1.02/1.61	37.64%	21704	421355	4.89%
Juglans cathayensis	458	1,143,008	23	1.05/1.61	45.01%	120312	1022696	10.52%
Juglans mandshurica	455	1,059,545	32	0.97/1.63	92.58%	110349	949196	10.41%
Juglans hopeiensis	368	990,423	26	0.83/1.82	61.78%	540892	449531	54.61%
Juglans regia	0	742,382	0	0.00/2.29	96.69%	225592	516790	30.38%
Juglans sigillata	6	1,051,470	2	0.20/2.29	46.02%	421420	630050	40.07%
Juglans hindsii	376	361,913	46	1.12/1.43	68.84%	17962	343951	4.96%
Juglans major	472	943,553	15	0.93/1.65	84.57%	73817	869736	7.82%
Juglans microcarpa	374	949,180	106	1.23/1.66	98.11%	111259	837921	11.72%
Juglans nigra	478	682,382	3	0.94/1.66	69.12%	166	1871	8.14%
Juglans cinerea	482	1,014,615	109	0.99/1.64	71.72%	80304	934311	7.91%
Pterocarya fraxinifolia	506	621,411	103	1.11/1.42	23.93%	37402	584009	6.01%
Pterocarya hupehensis	505	515,964	75	1.17/1.45	36.97%	35974	479990	6.97%
Pterocarya insignis	506	589,290	116	1.00/1.43	98.45%	41877	547413	7.10%
Pterocarya stenoptera	483	653,976	28	1.12/1.39	86.64%	53662	600314	8.20%
Pterocarya tonkinensis	487	547,524	133	1.05/1.44	68.29%	33764	513760	6.16%
Pterocarya stenoptera var. zhijiangensis	495	686,705	123	1.09/1.43	64.74%	45839	640866	6.67%
Pterocarya macroptera var. insignis	513	430,744	153	1.11/1.42	72.96%	23897	406847	5.54%
Carya cathayensis	1,131	253,730	288	1.04/1.43	53.16%	9006	244724	3.54%
Carya hunanensis	1,066	275,372	241	1.14/1.45	84.00%	21179	254193	7.69%
Carya illinoensis	1,100	761,308	268	1.05/2.28	44.90%	263845	497463	34.65%

Note: cp-SNPs = The number of SNPs of chloroplast genomes, nr-SNPs = The number of SNPs of whole genome resequencing, cp-Indels = The number of Indels of chloroplast genomes, Ts/Tv ratio = The transition divided transversion ratio based on chloroplast genomes and whole genome resequencing data respectively, Mapped = the mapped ratio of whole genome resequencing data used common walnut genome sequence data, Het-ratio = The Heterozygosity ratio of each samples based on the whole genome resequencing data.

Species	cp-SNPs	nr-SNPs	cp-Indels	Ts/Tv ratio	Mapped	Heterozygosity	Homozygosity	Het-ratio
Carya kweichwensis	1,102	326,824	24	1.09/1.43	66.91%	23680	303144	7.24%
Carya sinensis (*Annamocarya sinensis*)	1,041	344,322	5	1.01/1.56	69.18%	26548	325860	6.88%
Cyclocarya paliurus	676	483,874	9	0.77/1.43	56.99%	40571	443303	8.38%
Platycarya strobilacea	1,465	319,852	22	1.33/1.49	97.71%	39263	280589	12.27%
Engelhardia roxburghiana	1,959	215,357	166	1.17/1.66	55.74%	10655	204702	4.94%
Rhoiptelea chiliantha	2,126	202,314	179	1.21/1.71	58.22%	11023	196822	4.78%

Note: cp-SNPs = The number of SNPs of chloroplast genomes, nr-SNPs = The number of SNPs of whole genome resequencing, cp-Indels = The number of Indels of chloroplast genomes, Ts/Tv ratio = The transition divided transversion ratio based on chloroplast genomes and whole genome resequencing data respectively, Mapped = the mapped ratio of whole genome resequencing data used common walnut genome sequence data, Het-ratio = The Heterozygosity ratio of each samples based on the whole genome resequencing data.

Phylogenetic Analysis

Based on best-fit partitioning schemes and models, the phylogenies returned from the RAxML and MrBayes analyses using sixty-one chloroplast protein-coding genes with all branches highly supported (Fig. 3a). At the family level, there are six well-supported major clades with all species of the Juglandaceae, Myricaceae, Betulaceae, Fagaceae, and Cucurbitaceae families (Fig. 3b). Within the Fagales, members of the Juglandaceae were closest to the Myricaceae and Betulaceae (Fig. 3b). Species within the Juglandaceae divided into three groups corresponding to the three previously described sub-families (Juglandoideae, Engelhardioideae, and Rhoipteleoideae) with 100 % bootstrap (BS) support based on mtCDS and chloroplast genomes by maximum likelihood (ML) analysis (Fig. 3a, b).

Within the Juglandoideae subfamily, the species divided into five groups, corresponding to the five genera *Carya*, *Platycarya*, *Cyclocarya*, *Pterocarya*, and *Juglans* that were strongly supported as monophyletic (Fig. 3b). The genus *Pterocarya* was most closely related to *Juglans* (Fig. 3). The wheel wingnut (*Cyclocarya paliurus*) is the sole member of its genus in Juglandaceae. It was monophyletic and most closely related to *Pterocarya* based on chloroplast genomes (Fig. 3b; Fig. 4). In *Carya*, Pecan (*C. illinoinensis* a North American species) was joined with the other four species of *Carya* (Asian hickories) with 100% BS. The cladograms supported the current division of *Carya* into two sections (Sect. *Sinocarya*, Asian hickories, i.e., *C. cathayensis*, *C. hunanensis*, *C. kweichowensis*, and *C. sinensis*; and Sect. *Apocarya*, which includes *C. illinoinensis*). We confirmed that the genus *Annamocarya* (*A. sinensis*) is properly within *Carya* and closest to *Sinocarya* Fig. 3) [26]. Our dataset for the genus *Pterocarya* included seven species/subspecies resolved as two groups: *P. stenoptera* and *P. tonkinensis* in a group, and *P. macroptera*, *P. macroptera var. insignis*, *P. fraxinifolia*, *P. hupehensis*, and *P. stenoptera var. zhijiangensis* in another group. The three sections within *Juglans* were well resolved with high bootstrap support (*J. regia* and *J. sigillata* into Sect. *Juglans/Dioscaryon*; *J. mandshurica*, *J. ailantifolia*, and *J. cathayensis* into Sect. *Cardiocaryon*; *J. cinerea*, *J. nigra*, *J. hindsii*, *J. microcarpa*, and *J. major* into Sect. *Trachycaryon* and Sect. *Rhysocaryon*) based on data from both chloroplasts and mitochondria (Fig. 3a, b). Branch lengths for *J. hopeiensis*/*J. mandshurica* and *J. regia*/*J. sigillata* were extremely short, further supporting their recent divergence.

Based on 1,161,468 nuclear SNPs, the phylogenetic analysis showed a generally well-supported clustering topology with high bootstrap values when rooted against *Populus trichocarpa* as outgroup (Fig. 4). The resulting phylogeny identified and provided 100 % support for the three sub-families that we observed in the plastome-based phylogeny of the Juglandaceae

(Fig. 3; Fig. 4): Clade I (Rhoipteleoideae), clade II (Engelhardioideae), and clade III (Juglandoideae). Clade III (Juglandoideae) contained five genera *Platycarya*, *Carya*, *Cyclocarya*, *Pterocarya*, and *Juglans*, however, the relative placement of the three genera, *Carya*, *Platycarya*, and *Cyclocarya* was not consistent in the phylogenies based on the combined Cp and mitochondrial genomes versus the nuclear data. Although we only used one species in *Platycarya*, our results strongly supported the model that *Cyclocarya* and *Platycarya* are monophyletic with long branches and taxa-specific SNPs (Fig. 3; Table S3). Based on nuclear SNPs we found a strong sister relationship of *Cyclocarya* to *Pterocarya* and, secondarily, to *Juglans* (Fig. 4), as suggested by Manos et al. (2007) [16] and Larson-Johnson (2016) [34].

We reconstructed the Bayesian and ML trees based on ITS sequences of 55 Juglandaceae species (Fig. S4). The resulting phylogenetic tree showed that the three subfamilies, Juglandoideae, Engelhardioideae, and Rhoiptelioideae, cluster as monophyletic branches, however, support for the genera within the Juglandoideae was weak (< 50%) (Fig. S4). ITS alone produced cladograms markedly different than accepted topologies.

The Divergence Time and Historical Diversification of Juglandaceae

The stem age of Juglandaceae was estimated at 78.69 Mya (95 % highest posterior density (HPD): 76.58–80.50 Mya). The walnut family diverged from the Myricaceae during the late Cretaceous (Fig. 5). During the Middle Cretaceous to Late Cretaceous, the three subfamilies Rhoiptelioideae, Engelhardoideae and Juglandoideae diverged at 68.64 Mya and 60.65 Mya (95 % HPD: 58.98–70.98 Mya), respectively. The crown age of the genus *Carya* was estimated at 57.88 Mya (95 % HPD: 56. 67-60.32 Mya) during the Late Paleocene, *Platycarya* at 56.99 Mya (95 % HPD: 56.80–58.80 Mya), and *Cyclocarya paliurus* at 55.80 Mya (95 % HPD: 54.30–57.30 Ma). The divergence of *Pterocarya* and *Juglans* was estimated at 47.10 Mya (95 % HPD: 43.93–50.93 Mya) during the Early Eocene. Most genera of Juglandaceae diverged from 50.93 to 61.98 Mya in the relatively warm and dry climate of the Middle Paleocene to the Early Eocene (Fig. 5).

Discussion

Comparison Genomes of Juglandaceae

Both genome size and GC content among Juglandaceae plastomes were consistently more than the median genome size and GC content for land plant plastomes (Table 1). The nucleotide variability (Pi) across all 27 plastomes of Juglandaceae included in this study was 0.00791 (Fig. 2; Fig. S1). Coding regions with the highest variation included *matK*, *atpI*, *rpoC2*, *rps14*, *aacD*, *psaI*, *ycf4*, *cemA*, *rpl33*, *infA*, *rpl22*, *rps19*, *ndhF*, *rpl32*, *ndhD*, *ndhI*, and *ycf1*. Non-coding regions that were most variable were *matK-rps16*, *petN-psbD*, *ndhC-trn*V-UAC, *rbcL-psaI*, *psbE-petL*, and *rpl*14-*ycf*1. These regions of maximum variability will no doubt prove the most informative for phylogenetic studies in the Juglandaceae [6, 12]. Previous studies have identified *rpl22*, *rps19* and *ycf1* genes as the most variable genes in the Juglandaceae plastomes based on high indel density [12]. It was surprising, however, that the LSC region also contained variation, including 2,577 bp differences among Juglandaceae plastomes, while SSC had 237 bp and IR had 296 bp differences among plastomes (Table 1).

Backbone Relationship of Juglandaceae

The phylogeny of family Juglandaceae has been inferred based on microsporogenesis, morphology [22–23], fossils [24, 78], molecular markers [26, 35], and combined data (morphology, fossils, and molecular data) [16]. Several recent studies of phylogeny in the Juglandaceae have included data from plastomes [12, 35–38, 56, 79–80]. The previously recognized subfamilies (Engelhardioideae and Juglandoideae), tribes (Platycaryeae and Juglandeae) and subtribes (Caryinae and Juglandinae) were all strongly supported [26, 30, 35–38]. Our phylogenetic analyses indicated that the Juglandaceae is subdivided into three major clades corresponding to the three subfamilies Rhoipteleoideae, Engelhardioideae, and Juglandoideae [16, 26–27, 29, 34–35, 81] (Fig. 1). The evidence for these three subfamilies can be found from morphology, fossil, and molecular data [16], fruits [43], and flower development [40]. The subfamily Engelhardioideae includes *Engelhardia*, *Oreomunnea*, and *Alfaroa* [22] (Fig. 1). Our results supported the separation of *Alfaropsis* [16] as a separate genus within Engelhardioideae (Fig. S4). The Rhoipteleoideae (*Rhoiptelea chiliantha*) was a basal, monophyletic branch,

which indicated that winged (dry) fruit was an ancestral character for the Juglandaceae (Fig. 3, Fig. 4). The fruits of Myricaceae, the closest relative of the Juglandaceae, are small and fleshy, of a type common among Cretaceous flora [34–38]. The subfamily Rhoipteleoideae has only one species (*Rhoiptelea chiliantha*), which is a threatened and endemic in China [35–38, 47]

The subfamily Juglandoideae includes the commercially important nut-producing trees commonly called walnuts and butternuts (*Juglans*), pecan and hickory (*Carya*) [15, 26, 29] (Fig. 1). The Persian walnut, *Juglans regia*, is one of the major nut crops of the world. Walnuts and hickories are also valuable timber trees [65]. Our plastid phylogenomic analyses fully resolved relationships among the major clades and genera of Juglandoideae (Fig. 3). Within subfamily Juglandoideae, four tribes are recognized (Juglandeae, Cyclocaryae, Platycaryae, and Hicorieae). Our results strongly supported the previously published merger of the genera *Annamocarya* and *Carya* into the genus *Carya* [26, 43]. Five genera, with their subgenera and sections were identified previously [22, 24, 26], i.e., *Carya* (here including *Annamocarya*), *Platycarya*, *Cyclocarya*, *Pterocarya*, and *Juglans*. These five genera resolved in our analysis with 100 % support (Fig. 3). The phylogenetic relationships of the genera of the Juglandaceae reveal that *Carya* retains more primitive characters than *Platycarya* based on chloroplast DNA variation and morphology [81].

In previous studies, it was suggested the genus *Cyclocarya* is sister to genus *Platycarya* [16] based on fossil, chloroplast DNA fragments, and morphological data. Our data also confirm this relationship (Fig. 5). Alternatively, it was suggested by Xiang et al. (2014) that *Platycarya* is sister to *Juglans* based on five chloroplast markers [30], that *Carya* and *Platycarya* are sister groups [30]. Others considered *Cyclocarya* and *Juglans* to be sister groups [28]. Using criteria based on fruit morphology, however, *Carya* and *Juglans* are sister groups [34], this relationship was not confirmed by our DNA-based analysis (Fig. 5), and *Cyclocarya* and *Pterocarya* are sister groups [this relationship was supported in our data (Fig. 3, Fig. 5) [34, 43]. Previously, Smith and Doyle (1995) [81], based on chloroplast DNA and morphological data, concluded that *Platycarya* evolved earlier than *Carya;* our results based on nuclear resequencing (Fig. 4) supported this conclusion. Our results based on sequencing the entire chloroplasts, however, indicated that the differentiation of *Carya* preceded *Platycarya* (Fig. 3, Fig. 5; Fig. S4), as suggested by Zhang et al. (2013), although their differentiation, about 57 Mya, was roughly simultaneous.

The Phylogenetic Relationships within genus of Juglandaceae

Our analyses fully resolved some previously unresolved intergeneric relationships and added additional evidence supporting some of the recently altered generic circumscriptions based on analyses with much more appropriate representation at the species level. The species *C. sinensis* (Chinese Hickory, beaked walnut, or beaked hickory) was resolved into *Carya* (*Annamocarya sinensis*) [82]. The generic circumscription of *Annamocarya* (also *C. sinensis*) has frequently been altered, and many genera have been segregated from or merged with *Carya* [26, 79, 83].

The previously unresolved intrageneric relationships of *Pterocarya* were also resolved with high support. *P. stenoptera var. zhijiangensis* and *P. hupehensis* were clustered together (Fig. 3). These two species are sympatric and *P. stenoptera var. zhijiangensis* may be a subspecies of *P. hupehensis* (Fig. 3, Fig. 5, but see Fig. 4). The taxonomy of sub-species *P. stenoptera var. zhijiangensis* and *P. macroptera var. insignis* conflicted with the previous study of Wu and Raven (1999) [82]. We consider these taxa subspecies based on our data (Fig. 3; Fig. 4, but see Fig. 5), however we did not complete a detailed phylogeny of *Pterocarya* because our sample pool was too small.

Our phylogenomic analyses resolved genus *Juglans* into three well sections (*Cardiocaryon*, *Dioscaryon*, and *Rhysocaryon*) with high support (Fig. 3; Fig. 4). Earlier phylogenies [22, 24] based on limited molecular data sometimes included a fourth section (*Trachycaryon*) containing only the North American species *J. cinerea*. The separation of *Trachycaryon* as distinct from section *Cardiocaryon* was inconsistent with morphology [21–23] and nuclear markers [84–85], but congruent with fossil data [24] and the results of other analyses based on plastid sequences [12, 15]. In our phylogenetic analysis of nuclear genome SNPs, American butternut (*J. cinerea*) has high support (100%) as sister to Section *Cardiocaryon* (Asian butternut, *J. cathayensis*, *J. mandshurica*, and *J. ailantifolia*) (Fig. 4).

Based on sequence data from 16 mtCDS and 61 chloroplast protein-coding genes, our results supported the unification of *J. mandshurica*, *J. ailantifolia*, and *J. cathayensis* within sect. *Cardiocaryon* (Fig. 3; Fig. S4), consistent with a previous conclusion based on genotyping by sequencing data [22, 86]. We also confirmed that the Ma walnut (*J. hopeiensis*) arose from the resent hybridization of *J. regia* and *J. mandshurica* based on both matrilineal and biparental inheritance data (Fig. 3; Fig. 4) [12, 86]. The placement of *J. cinerea* into *Rhysocaryon* (black walnuts) based on plastome sequence was clear (Fig. 3), however, it belongs to *Cardiocaryon* (Asian butternuts) based on nuclear sequences (Fig. 4), and its morphology is consistent with *Cardiocaryon* [12, 15]. In addition, *J. cinerea* can hybridize with members of *Cardiocaryon* and even *Dioscaryon*, but not with *Rhysocaryon* [87]. All other North American *Rhysocaryon* freely hybridize. The discordance between the *J. cinerea* nuclear genome and its plastome is almost certainly the result of a chloroplast capture [15, 31]. It is notable that the chloroplast of *J. cinerea* is not an ancient one (ancestral to the *Rhysocaryon*) but is instead most like *J. nigra* (Fig. 5). Our results indicated that the capture of a *Rhysocaryon* chloroplast by *J. cinerea* capture was relatively recent (Fig. 5). Hybridization and chloroplast capture between *Rhysocaryon* and *Cardiocaryon* apparently played a major role in the diversification of *Juglans*, as it did in other plant families [88–90].

Dating of the Origin and Historical Diversification of Juglandaceae

Stem ages in the Juglandaceae are controversial [13, 16, 28–29, 91]. Most previous studies estimated a stem age of Juglandaceae about 84 Mya in the Cretaceous [24, 28], however, the divergence times for some genera remain uncertain [28–29, 91], as only a few studies have examined the divergence times among the major genera and within the species of the family [16, 29, 91]. The lack of a robust phylogenetic framework and time tree has hindered development of a full understanding of the diversification of Juglandaceae.

The crown ages of Betulaceae, Myricaceae, and Casuarinaceae were 74.0 Mya (66.9–80.3), 90.4 Mya (85.0-94.6), and 82.8 Mya (74.7–88.6), respectively [30]. The crown age of Juglandaceae varied among previous studies, 78 Mya by Manos et al. (2007) [16], 71 Mya by Larson-Johnson (2016) [34], 85.5 Mya by Sauquet et al. (2012) [28] and 79.9 Mya by Xiang et al. (2014). Our results indicated the stem age of Juglandaceae to be during the late Cretaceous (78.58 Mya with 95% HPD: 76.58–80.50 Mya). The major diversification of the family is recorded in the pollen and megafossil record of the early Tertiary (~ 65Mya) at the K-T boundary [24]. The three subfamilies diverged during the Late Cretaceous to Early Palaeocene (60.65–68.64 Mya) (Fig. 5). Our estimates of divergence times among subfamilies and major genera were from 50.93 to 61.98 Mya in warm and dry habitats during the Middle Palaeocene to the Early Eocene (Fig. 5), which is largely consistent with the estimates of Xiang et al. (2014) and Larson-Johnson (2016) [34]. We estimated the divergence time of *Juglans* and *Pterocarya* to have been ~ 47 Mya (Fig. 5; Manos et al. 2007, ~ 55 Mya) [16], and ~ 56 Mya between *Pterocarya* and *Cyclocarya* (Fig. 5; Manos et al. 2007, ~ 59 Mya) [16], however, both Xiang et al. (2014) and Larson-Johnson (2016) estimated a divergence time between *Juglans* and *Pterocarya* of ~ 24 Mya [30, 34], and ~ 18 Mya between *Pterocarya* and *Cyclocarya* [34]. During the end of the Eocene, *Cyclocarya* and *Platycarya* became extinct in North America but survived in Eurasia [24]. Our results indicated *Carya* emerged as an animal-dispersed genus about 58 Mya, considerably earlier than the estimate (44 Mya) of Larson-Johnson (2016) [34], although we agree that the overwhelming majority of winged and wingless fruited *genera* diverged or diversified during the Paleogene, probably reflecting adaptation to changing regeneration regimes[92].

During the early Tertiary to the Neogene there was likely extensive migration and exchange among North Atlantic, North America, western Europe, and Asia [24]. Interestingly, most species within the extant genera diversified between 18.54 and 8.52 Mya in warm and dry environments of the Early Miocene (Fig. 5), a period of especially rapid speciation within *Juglans* and *Pterocarya*. Some closely related species pairs within *Juglans* appear to have diverged relatively recently, under the influence of climate change during the Quaternary glacial period (Fig. 5; Bai et al. 2017). For example, *J. regia* and *J. sigillata*, *J. mandshurica* and *J. hopeiensis*, and *Carya hunanesis* and *C. kweichwensis* (Fig. 5). Overall, the Juglandaceae reflect a complex evolutionary history and diversification affected by changes in geography, distinctive distributions, climate changes, coevolution with animals. Biotic interactions (e.g., pathogens) no doubt also had a role in driving species abundance and distribution [93], but biotic interactions of that type are difficult to detect from current data [35–39].

Conclusion

Our results are a first attempt to use whole genomes to elucidate the characterize sequence divergence and evolutionary history in the Juglandaceae. Evidence of early lineage diversification, hybridization and extinction lead us to predict complex evolutionary histories for the extant species in the Juglandaceae. A fully resolved, strongly supported, time-calibrated phylogenetic tree of Juglandaceae will provide an important framework for studying classification, diversification, biogeography, phenotypic evolution, gene function and comparative genomics of this important family. Our results supported some recently clarified circumscriptions of controversial genera, although our taxonomic sampling is insufficient to stand alone as definitive. Wider plastid phylogenomics, whole genomes (nuclear data), a more complete fossil record, better dating of the fossil record and more studies of morphology will all be needed to fully reconstruct the phylogeny of woody plant families such as the Juglandaceae and other families of Fagales.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The annotated newly chloroplast genomic sequence were deposited into GenBank (MH188288-MH188304, MH189594-MH189595; Details see Table 1).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (32070372; 41471038; 31200500) and Natural Science Foundation of Shaanxi Province of China (2019JM-008).

Authors’ contributions

PZ, HZ, and KW designed and managed the project; HZ, YH, AE, PL, and PZ collected the materials; HZ, YH, PZ, and KW prepared and purified DNA samples; HZ, PZ, PL, YH, and KW performed the data analyses; PZ wrote the paper; HZ, KW, SZ, and FZ revised the paper. All authors read and approved the final manuscript.

Acknowledgements

We thank N. Hou, Y. L. Xu, N. Lin and L. Wang for assisting with taxon sampling. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture and does not

imply its approval to the exclusion of other products or vendors that also may be suitable.

References

1. Daniell, H., Lin, C.S., Yu, M., Chang, W.J. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134.
2. Huang, C.H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M.A., Al-Shehbaz, I., Edger, P.P. et al. 2016. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Biol. Evol.* 33, 394-412.
3. Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C.A., Carretero-Paulet, L., Chang, T.H., Lan, T., Welch, A.J., Juárez, M.J., Simpson, J. et al. 2013. Architecture and evolution of a minute plant genome. *Nature*. 498, 94-8.
4. Bellot, S., Cusimano, N., Luo, S., Sun, G., Zarre, S., Gröger, A., Temsch, E., Renner, S.S. 2016. Assembled plastid and mitochondrial genomes, as well as nuclear genes, place the parasite family Cynomoriaceae in the Saxifragales. *Genome Biol. Evol.* 8, 2214-2230.
5. Song, Y., Yu, W.B., Tan, Y., Liu, B., Yao, X., Jin, J., Padmanaba, M., Yang, J.B., Corlett, R.T. 2017. Evolutionary comparisons of the chloroplast genome in Lauraceae and insights into loss events in the Magnoliids. *Genome Biol. Evol.* 9, 2354-2364.
6. Uribe-Convers, S., Carlsen, M.M., Lagomarsino, L.P., Muchhala, N. 2017. Phylogenetic relationships of *Burmeistera* (Campanulaceae: Lobelioideae): Combining whole plastome with targeted loci data in a recent radiation. *Phylogenet. Evol.* 107, 551-563.
7. Zhang, S..D, Jin, J.J., Chen, S.Y., Chase, M.W., Soltis, D.E., Li, H.T., Yang, J.B., Li, D.Z., Yi. T.S. 2017. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytol.* 214, 1355-1367.
8. Novikova, P.Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., Guggisberg, A., Paape, T., Schmid, K., Fedorenko, O.M. et al. 2016. Sequencing of the genus arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Genet.* 48, 1077-82.
9. Williams, A.V., Miller, J.T., Small, I., Nevill, P.G., Boykin, L.M. 2016. Integration of complete chloroplast genome sequences with small amplicon datasets improves phylogenetic resolution in *Acacia*. *Phylogenet. Evol.* 96, 1-8.
10. Du, Y.P, Bi, Y., Yang, F.P., Zhang, M.F., Chen, X.Q., Xue, J., Zhang, X.H. 2017. Complete chloroplast genome sequences of *Lilium*: insights into evolutionary dynamics and phylogenetic analyses. *Rep.* 7, 5751.
11. Ma, T., Wang, K., Hua, Q., Xia, Z., Wan, D., Wang, Q., Feng, J., Jiang, D., Ahanic, H., Abbottd, R.J. et al. Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *P. Natl. Acad. Sci. USA* E236–E243.
12. Dong, W., Xu, C., Li, W., Xie, X., Lu, Y., Liu, Y., Jin, X., Suo, Z. 2017. Phylogenetic resolution in *Juglans* based on complete chloroplast genomes and nuclear DNA sequences. *F in Plant Sci.* 8, 1148.
13. Hu, Y.H., Woeste, K.E., Zhao, P. 2017. Completion of the chloroplast genomes of five Chinese *Juglans* and their contribution to chloroplast phylogeny. *Plant Sci.* 7, 1955.
14. Stanford, A.M., Harden, R., Parks, C.R. 2000. Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. *J. Bot.* 87, 872-882.
15. Aradhya, M.K., Potter, D., Gao, F., Simon, C.J. 2007. Molecular phylogeny of *Juglans* (Juglandaceae): a biogeographic perspective. *Tree Genet. Genomes.* 3, 363-378.
16. Manos, P.S., Soltis, P.S., Soltis, D.E., Manchester, S.R., Oh, S.H., Bell, C.D., Dilcher, D.L., Stone, D.E. 2007. Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. *Biol.* 56, 412-430.
17. Chen, Z., Grover, C.E., Li, P., Wang, Y., Nie, H., Zhao, Y., Wang, M., Liu, F., Zhou, Z., Wang, X. et al. Molecular evolution of the plastid genome during diversification of the cotton genus. *Mol Phylogenet. Evol.* 112, 268.

18. Bernhardt, N., Brassac, J., Kilian, B., Blattner, F.R. 2017. Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. *BMC Evol. Biol.* 17, 141.
19. Simpson, L., Clements, M.A., Crayn, D.M., Nargar, K. 2017. Evolution in Australia's mesic biome under past and future climates: Insights from a phylogenetic study of the Australian rock Orchids (*Dendrobium speciosum* complex, Orchidaceae). *Phylogenet. Evol.* 118,32-46.
20. Lu, A.M. 1982. On the geographical distribution of the Juglandaceae. *Acta Phytotaxon. Sin.* 20:257-274.
21. Lu, A.M., Stone, D.E., Grauke, L.J. 1999. "Juglandaceae," in Flora of China. Wu Z.Y., Raven P.H., ed. Beijing: Science Press, 277-285.
22. Manning, W.E. 1938. The morphology of the flowers of the Juglandaceae. I. The inflorescence. *J. Bot.* 25, 407-419.
23. Manning WE. 1978. The classification within the Juglandaceae. *Mo. Bot. Gard.* 65:1058-1087.
24. Manchester, S.R., Garden, M.B. 1987. The fossil history of the Juglandaceae. *Mo. Bot. Gard.* 21, 1-137.
25. Manchester, S.R., Dilcher, D.L. 1997. Reproductive and vegetative morphology of *Polyptera* (Juglandaceae) from the Paleocene of Wyoming and Montana. *J. Bot.* 84, 649.
26. Manos, P.S., Stone, D.E. 2001. Evolution, phylogeny, and systematics of the Juglandaceae. *Mo. Bot. Gard.* 88, 231-269.
27. Angiosperm Phylogeny Group (APG IV). 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1-20.
28. Sauquet, H., Ho, S.Y., Gandolfo, M.A., Jordan, G.J., Wilf, P., Cantrill, D.J., Bayly, M.J., Bromham, L., Brown, G.K., Carpenter, R.J. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of Nothofagus (Fagales). *Biol.* 61, 289-313.
29. Zhang, J..B, Li, R.Q., Xiang, X.G., Manchester, S.R., Lin, L., Wang, W., Wen, J., Chen, Z.D. 2013. Integrated fossil and molecular data reveal the biogeographic diversification of the eastern Asian-eastern North American disjunct hickory genus (*Carya*). *PLoS One.* 8, e70449.
30. Xiang, X.G., Wang, W., Li, R.Q., Lin, L., Liu, Y., Zhou, Z.K., Li, Z.Y., Chen, Z.D. 2014. Large-scale phylogenetic analyses reveal Fagalean diversification promoted by the interplay of diaspores and environments in the Paleogene. *Plant Ecol.* 16, 101-110.
31. Zhang, B.W., Xu, L.L., Li, N., Yan, P.C., Jiang, X.H., Woeste, K.E., Lin, K., *et al*. (2019) Phylogenomics reveals an ancient hybrid origin of the Persian walnut. *Biol. Evol.* 11:2451-2461.
32. Zhang JP, Zhang WT, Ji FY, Qiu J, Song XB, Bu DC, et al. High-quality walnut genome assembly reveals extensive gene expression divergences after whole-genome duplication. Plant Biotechnol J. 2020;1-3.
33. Li, R.Q., Chen, Z.D., Lu, A.M., Soltis, D.E., Soltis, P.S., Manos, P.S. 2004. Phylogenetic relationships in Fagales based on DNA sequences from three genomes. *J. Plant Sci.* 165, 311-324.
34. Larson-Johnson, K. 2016. Phylogenetic investigation of the complex evolutionary history of dispersal mode and diversification rates across living and fossil Fagales. *New Phytol.* 209, 418-435.
35. Mu, X.Y., Tong, L., Sun, M., Zhu, Y. X., Wen, J., Lin, Q. W., Liu, B. (2020) Phylogeny and divergence time estimation of the walnut family (Juglandaceae) based on nuclear RAD-Seq and chloroplast genome data. *Phylogenet. Evol.* 147, 106802
36. Song, Y.G., Fragnière, Y., Meng, H.H., Li, Y., Bétrisey, S., Corrales, A., Manchester, S., Deng, M., Jasińska, A.K., Sâm, H.V., Kozlowski, G. (2020) Global biogeographic synthesis and priority conservation regions of the relict tree family Juglandaceae. *Biogeog.* 47, 643-657.
37. Song YG, et al. (2020) Phylogeny, taxonomy, and biogeography of *Pterocarya* (Juglandaceae). Plants 9.
38. Zhang, C.Y., Low, S.L., Song, Y.G., Nurainas, Kozlowski, G., et al. (2020) Shining a light on species delimitation in the tree genus *Engelhardia* Leschenault ex Blume (Juglandaceae). Molecular Phylogenetics and Evolution 152, 106918.
39. Dode, L.A. 1909a. Contribution to the study of the genus *Juglans* (English translation by R.E. Cuendett). *Bulletin of the Society of Dendrology*, France 11, 22-90.

40. Dode, L.A. 1909b. Contribution to the study of the genus *Juglans* (English translation by R.E. Cuendett). *Bulletin of the Society of Dendrology*, France 12, 165-215.
41. Lin, R.Z., Li, R.Q., Lu, A.M., Zhu, J.Y., Chen, Z.D. 2016. Comparative flower development of *Juglans regia*, *Cyclocarya paliurus* and *Engelhardia spicata*: homology of floral envelopes in Juglandaceae. *Bot. J. Linn. Soc.* 181, 279-293.
42. Christenhusz, M.J., Byng, J.W. 2016. The number of known plants species in the world and its annual increase. 261, 201-217.
43. Hermsen, E.J., Gandolfo, M.A. 2016. Fruits of Juglandaceae from the Eocene of South America. *Bot.* 41, 316-328.
44. Trelease, W. 1896. Juglandaceae of the United States. *Bot.Gard. Ann. Rep.* 1896, 25-46.
45. Hemenway, A.F. 1911. Studies on the Phloem of the Dicotyledons. I. Phloem of the Juglandaceae. *Gazette*, 51: 131-135.
46. Kuang, K.Z. 1960. De familia monotypica Rhoipteleaceae. *Bot. Sin.* 9, 43-47.
47. Geng, Y.F., Hu, G.X., Wang, S., Xu, J.C. 2018. Complete chloroplasts genome of the threatened *Rhoiptelea chiliantha* (Juglandaceae s.l.). *Genet. Resour.* https://doi.org/10.1007/s12686-018-1021-4.
48. Angiosperm Phylogeny Group (APG III). 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105-121.
49. Manchester, S.R. 1989. Early history of the Juglandaceae. *Plant Syst. Evol.* 162, 231-250.
50. Gunter, L.E., Kocher,t G., Giannasi, D.E. 1994. Phylogenetic relationships of the Juglandaceae. *Plant Syst. Evol.* 192, 11-29.\
51. Doyle, J.J., Doyle. J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Bull.* 19, 11-15.
52. Zhao, P., Woeste, K.E. 2011. DNA markers identify hybrids between butternut (*Juglans cinerea*) and Japanese walnut (*Juglans ailantifolia* Carr.). *Tree Genet. Genomes.* 7, 511-533.
53. Patel, R.K., Jain, M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One.* 7, e30619.
54. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D. et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Comput. Biol.* 19, 455-477.
55. Kearse, M., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T. et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. 28, 1647-1649.
56. Hu, Y., Woeste, K.E., Dang, M., Zhou, T., Feng, X.J., Zhao, G.F., Liu, Z.L., Li, Z.H., Zhao, P. 2016b. The complete chloroplast genome of common walnut (*Juglans regia*). *DNA Part B.* 1, 189-190.
57. Wyman, S.K., Jansen, R.K., Boore, J.L. 2004. Automatic annotation of organellar genomes with DOGMA. 20, 3252-5.
58. 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 Res.* 41, W575.
59. R Core Team. 2016. *R: a language and environment for statistical computing, version 3.3.1.* Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL http://www.R-project.org/.
60. Li, H., Durbin. R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.
61. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-2079.
62. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297-1303.
63. Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 27, 2987-2993.

64. Bolger, A.M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–212.
65. Martínez-García PJ, Crepeau MW, Puiu D, Gonzalez-Ibeas D, Whalen J, Stevens KA, et al. The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis of nonstructural polyphenols. Plant J. 2016; 87:507-532.
66. Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Biol. Evol.* 34, 772-773.
67. Katoh, K., Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7, improvements in performance and usability. *Biol. Evol.* 30, 772-780.
68. Posada, D., Buckley, T.R. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Biol.* 53, 793-808.
69. Posada, D., Crandall, K.A. 1998. Modeltest: testing the model of DNA substitution. 14, 817-8.
70. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Biol.* 61, 539-542.
71. Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. 30, 1312-1313.
72. Edgar, R.C. 2004. Muscle: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
73. Kumar, S., Stecher, G., Tamura, K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Biol. Evol*. 33, 1870.
74. Smith, S.A., O'Meara, B.C. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. 28, 2689-2690.
75. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J. 2014. BEAST 2: a software platformfor bayesian evolutionary analysis. *PloS Biol.* 10, e1003537.
76. Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Biol. Evol.* 29, 1969-1973.
77. Wing, S.L., Hickey, L.J. 1984. The *Platycarya* perplex and the evolution of the Juglandaceae. *J. Bot.* 71, 388-411.
78. Blokhina, N.I. 2007. Fossil wood of the Juglandaceae: Some questions of taxonomy, evolution, and phylogeny in the family based on wood anatomy. *Paleontological J.* 41, 1040-1053.
79. Y., Chen, X., Feng, X., Woeste, K.E., Zhao, P. 2016a. Characterization of the complete chloroplast genome of the endangered species *Carya sinensis* (Juglandaceae). *Conserv. Genet. Resour.* 8, 1-4.
80. Hu, Y., Yan, J., Feng, X., Dang, M., Woeste, K.E., Zhao, P. 2017. Characterization of the complete chloroplast genome of wheel wingnut (*Cyclocarya paliurus*), an endemic in China. *Genet. Resour.* 9, 1-3.
81. Smith, J.F., Doyle, J. 1995. A cladistic analysis of chloroplast DNA restriction site variation and morphology for the genera of the Juglandaceae. *Am. J. Bot.* 82, 1163-1172.
82. Wu, Z.Y., Raven, P.H. 1999. Flora of China. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press, 139-162.
83. Wade, E.M., Nadarajan, J., Yang, X., Ballesteros, D., Sun, W., Pritchard, H.W. 2016. Plant species with extremely small populations (PSESP) in China: A seed and spore biology perspective. *Plant Divers.* 38, 209-220.
84. Fjellstrom, R.G., Parfitt, D.E. 1995. Phylogenetic analysis and evolution of the genus *Juglans* (Juglandaceae) as determined from nuclear genome RFLPs. *Plant Syst. Evol.* 197, 19-32.
85. Stanford AM, Harden R, Parks CR. Phylogeny and biogeography of *Juglans* (*Juglandaceae*) based on matK and ITS sequence data. Am J Bot. 2000;87:872–882.

86. Zhao, P., Zhou, H., Potter, D., Hu, Y., Feng, X., Dang, M., Feng, L., Zulfiqar, S., Liu, W., Zhao, G., et al. 2018. Population genetics, phylogenomics and hybrid speciation of *Juglans* in China determined from whole chloroplast genomes, transcriptomes, and genotyping-by-sequencing (GBS). *Phylogenet. Evol.* 126, 250–265
87. Woeste, K., Michler, C. 2011. *Juglans*. In Wild Crop Relatives: Genomic and Breeding Resources. Springer Berlin Heidelberg, 77-88.
88. Rieseberg, L.H., Soltis, D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5, 65–84.
89. Wendel, J.F., Doyle, J.J. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. Molecular systematics of plants II. Springer, Boston, MA, 265-296.
90. Stegemann, S., Keuthe, M., Greiner, S., Bock, R. 2012. Horizontal transfer of chloroplast genomes between plant species. *Proc Natl Acad Sci* S. A. 109, 2434-2438.
91. Kou, Y., Cheng, S., Tian, S., Li, B., Fan, D., Chen, Y., Soltis, D.E., Soltis, P.S., Zhang, Z. 2016. The antiquity of *Cyclocarya paliurus* (Juglandaceae) provides new insights into the evolution of relict plants in subtropical China since the late Early Miocene. *Biogeogr.* 43, 351-360.
92. Eriksson, O., Friis, E.M., Löfgren, P. 2000. Seed size, fruit size, and dispersal systems in angiosperms from the early Cretaceous to the late Tertiary. *Nat.* 156, 47-58.
93. Bai, W.N., Yan, P.C., Zhang, B.W., Woeste, K.E., Lin, K., Zhang, D.Y. 2017. Demographically idiosyncratic responses to climate change and rapid Pleistocene diversification of the walnut genus Juglans (Juglandaceae) revealed by whole-genome sequences. *New Phytol.* 217, 1726-1736.
94. Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. 292, 686-93.
95. Zachos, J.C., Dickens, G.R., Zeebe, R.E. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. 451, 279-83.

Figures

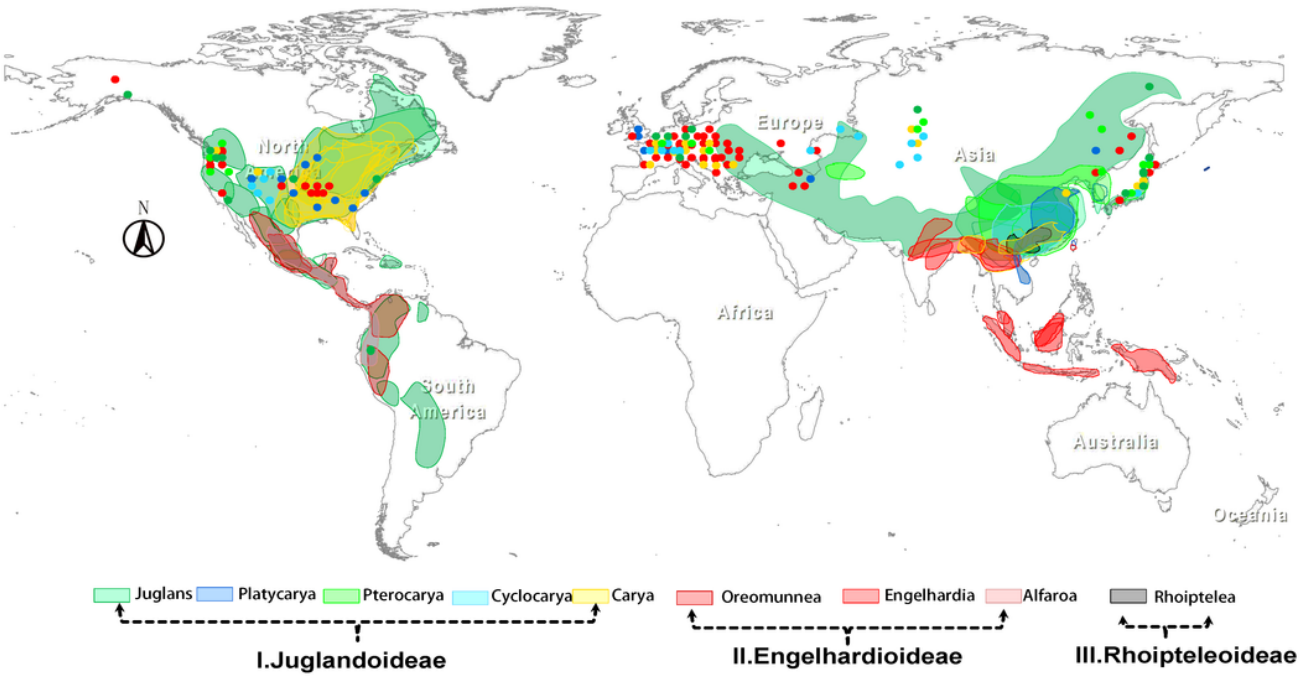

Figure 1

Geographic distribution of modern and fossil members of the Juglandaceae. Shaded regions indicate the modern distribution of eight genera belong to the three subfamilies (Juglandoideae, Engelhardioideae, and Rhoipteleoideae). The map used ArcGIS (version 10.0). The source locations of Juglandaceae fossils used in our analyses are colored dots. Green-Juglans, Blue-Platycarya, Light green-Pterocarya, Light blue-Cyclocarya, Yellow-Carya, Dark red-Oreomunnea, Red-Engelhardia, Light red-Alfaroa, and Black-Rhoiptelea. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

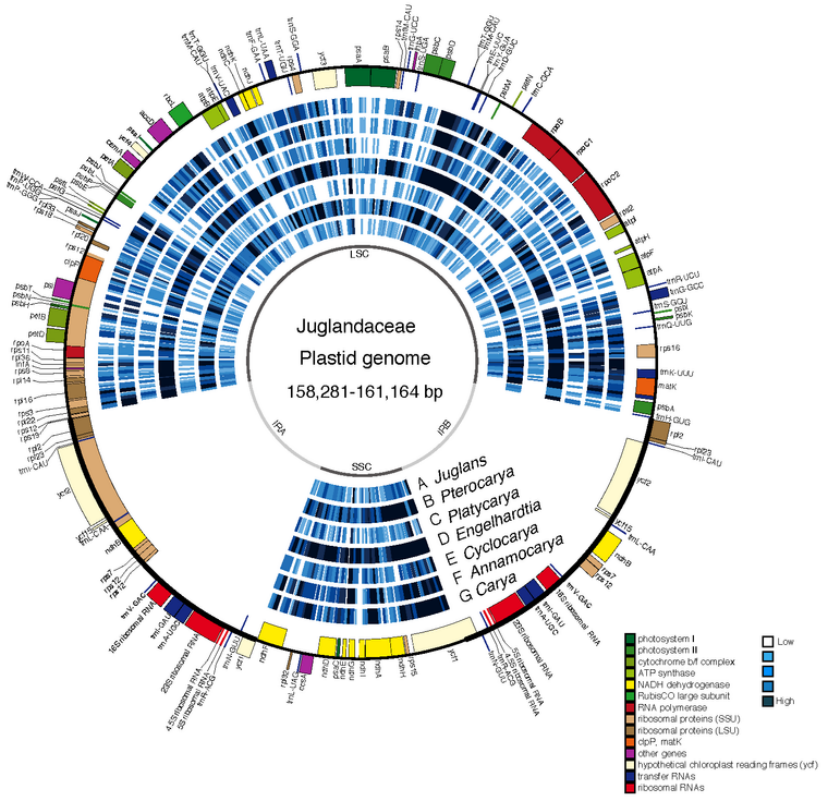

Figure 2

Variability of the family Juglandaceae represented over the circular map of the Juglans regia (NCBI accession number: KT963008; Hu et al. 2016a) chloroplast genome. The two inverted repeat regions (IRa and IRb) separate the large (LSC) and small (SSC) single copy regions, respectively. Genes represented by outside rectangles are on the positive strand, whereas genes represented by inside rectangles are on the negative strand. Density of SNPs is represented by a heatmap that varies from low (white) to high (dark blue). The circle depicts average SNPs density estimated in 350 bp moving windows. Carya=Carya cathayensis, Annamocarya =Annamocarya sinensis, Engelhardia=Engelhardia roxburghiana, Platycarya= Platycarya strobilacea, Pterocarya=Pterocarya fraxinifolia, Juglans=Juglans ailantifolia.

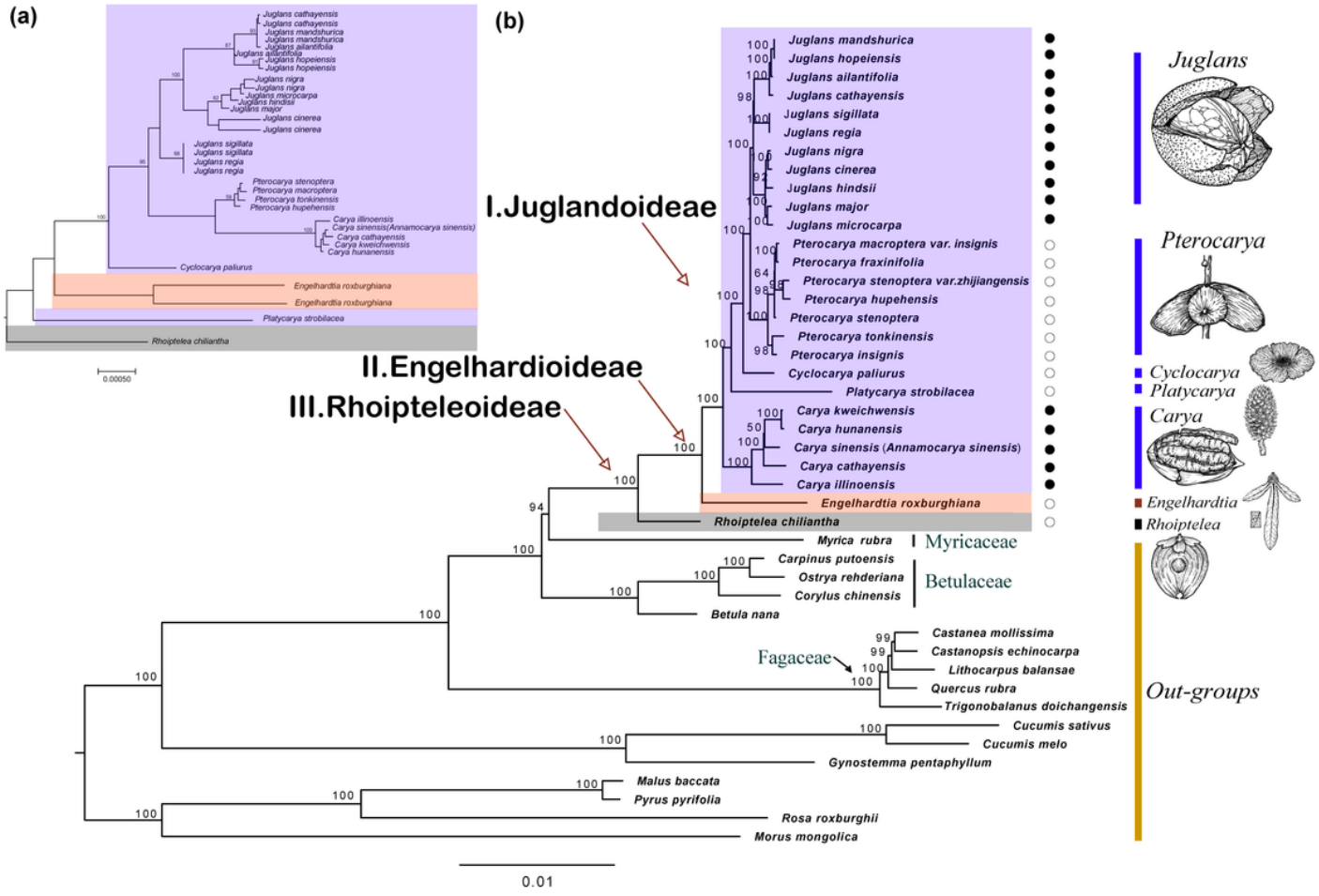

Figure 3

The Maximum Likelihood (ML) phylogenetic trees of Juglandaceae. Trees are based on sixteen mtCDS (a) and sixty one chloroplast protein-coding genes in the chloroplast (b). For both trees, the PartitionFinder method for the best model combinations (Table S5) was inferred by RAxML. Numbers at nodes correspond to ML bootstrap percentages (10,000 replicates). The three subfamilies are indicated with colored shading; Rhoipteleoideae (grey), Engelhardioideae (pink), and Juglandoideae (blue). Fruit morphology is shown using one species from each genus; the black solid circles indicate wingless fruits, hollow circles indicate winged fruits. Details for the outgroups (orange bar) are in Table S1.

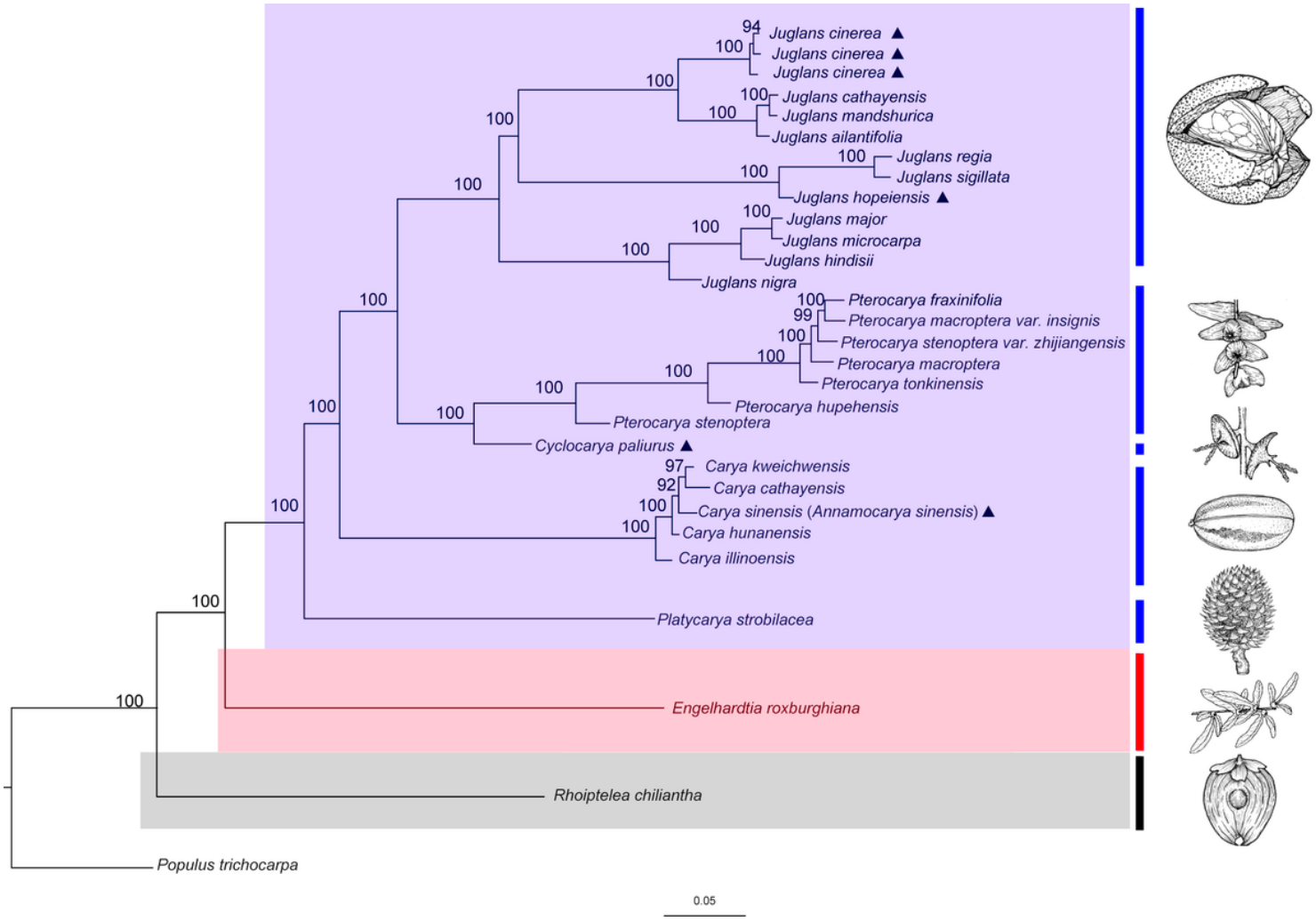

Figure 4

The Maximum Likelihood (ML) phylogenetic tree of Juglandaceae plus outgroup taxa (Populus trichocarpa) based on nuclear SNPs from whole genome resequencing data. Numbers at nodes correspond to ML bootstrap percentages (10,000 replicates). The three subfamilies are indicated with shading (Rhoipteleoideae (grey), Engelhardioideae (red), and Juglandoideae (blue). The fruit of one species from each genus is shown. The triangles indicate taxa with discordance between nuclear and chloroplast phylogeny.

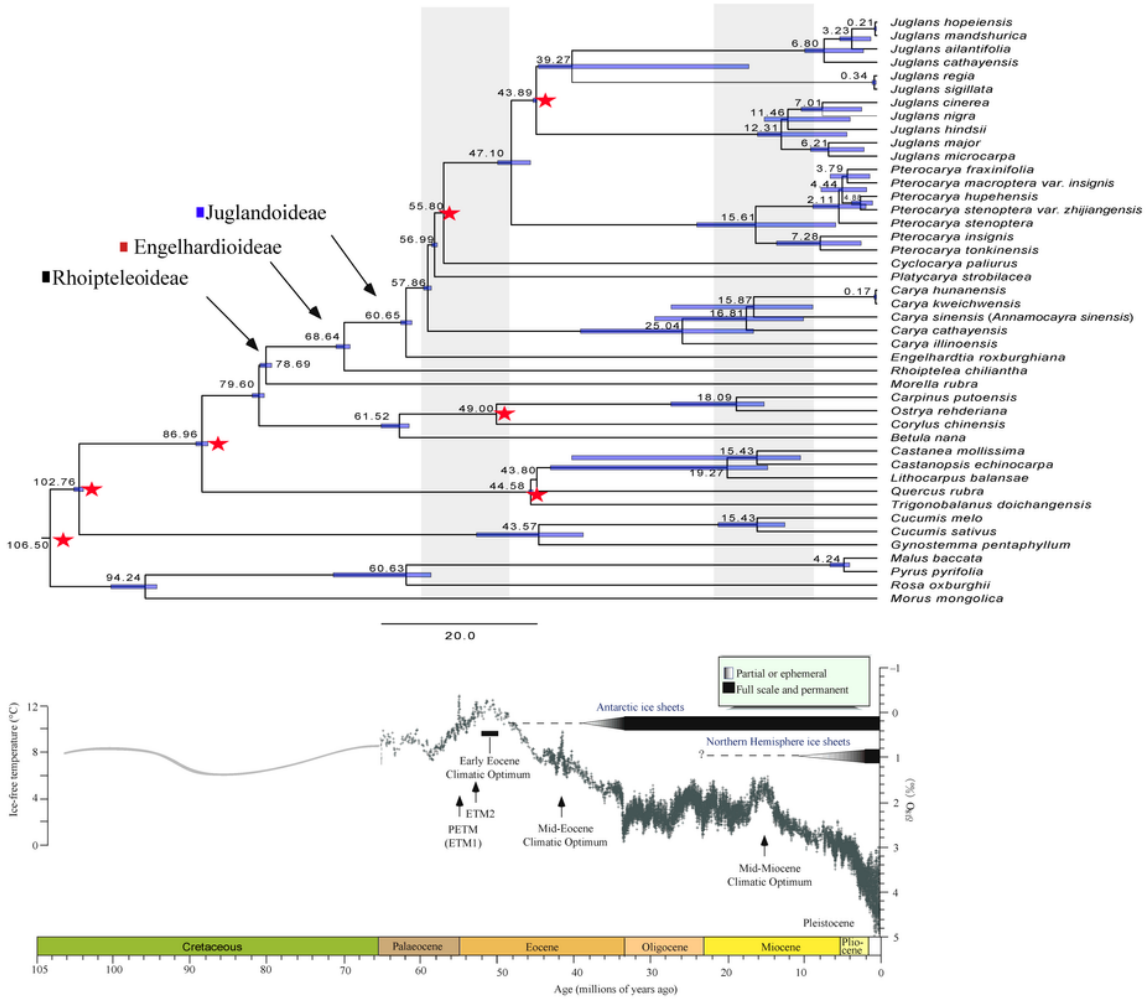

Figure 5

Time-calibrated phylogenetic tree of Juglandaceae. Mean divergence times estimated using a relaxed molecular clock model with 6 fossil priors (red stars). Blue bars across nodes indicate 95 % HPD intervals around the mean divergence time estimates. Nodes are numbered as ages. The genera and subfamilies of Juglandaceae are shown in the figure and the geological time scale is shown below the tree. A stacked deep-sea benthic foraminiferal oxygen-isotope curve shows the evolution of global climate over the last 65 Mya, as modified from Fig. 2 in Zachos et al. (2001, 2008) [94-95]. PETM: Palaeocene–Eocene thermal maximum; ETM: Eocene thermalmaximum; Pl: Pliocene. Reprinted by permission from Macmillan Publishers Ltd: Nature (451, 279-283), copyright (2008).

Supplementary Files

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

- SupplementaryMaterial.rar