

Genomic Analysis of Plastid–Nuclear Interactions and Differential Evolution Rates in Coevolved Genes across Juglandaceae Species

Yang Yang¹, Evan S. Forsythe^{2,3}, Ya-Mei Ding^{1,4}, Da-Yong Zhang ¹, and Wei-Ning Bai^{1,*}

¹State Key Laboratory of Earth Surface Processes and Resource Ecology, and Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, China

²Department of Biology, Oregon State University-Cascades, Bend, Oregon, USA

³Integrative Biology Department, Oregon State University, Corvallis, Oregon, USA

⁴South China Botanical Garden, The Chinese Academy of Sciences, Guangdong, China

*Corresponding author: E-mail: baiwn@bnu.edu.cn.

Accepted: 25 July 2023

Abstract

The interaction between the nuclear and chloroplast genomes in plants is crucial for preserving essential cellular functions in the face of varying rates of mutation, levels of selection, and modes of transmission. Despite this, identifying nuclear genes that coevolve with chloroplast genomes at a genome-wide level has remained a challenge. In this study, we conducted an evolutionary rate covariation analysis to identify candidate nuclear genes coevolving with chloroplast genomes in Juglandaceae. Our analysis was based on 4,894 orthologous nuclear genes and 76 genes across seven chloroplast partitions in nine Juglandaceae species. Our results indicated that 1,369 (27.97%) of the nuclear genes demonstrated signatures of coevolution, with the Ycf1/2 partition yielding the largest number of hits (765) and the ClpP1 partition yielding the fewest (13). These hits were found to be significantly enriched in biological processes related to leaf development, photoperiodism, and response to abiotic stress. Among the seven partitions, AccD, ClpP1, MatK, and RNA polymerase partitions and their respective hits exhibited a narrow range, characterized by dN/dS values below 1. In contrast, the Ribosomal, Photosynthesis, Ycf1/2 partitions and their corresponding hits, displayed a broader range of dN/dS values, with certain values exceeding 1. Our findings highlight the differences in the number of candidate nuclear genes coevolving with the seven chloroplast partitions in Juglandaceae species and the correlation between the evolution rates of these genes and their corresponding chloroplast partitions.

Key words: Juglandaceae, coevolution, evolutionary rate covariation, plastid, orthologous genes.

Significance

Our study provides new insights into the coevolutionary dynamics between plant nuclear and plastid genomes. By applying evolutionary rate covariation (ERC) analysis to nine Juglandaceae species, we detected candidate nuclear genes that exhibited signatures of coevolution with plastid partitions. Our results contribute to the advancement of our understanding of the interaction between these genomes and highlight the potential utility of ERC analysis for uncovering coevolved and cofunctional genes in plants.

Introduction

A significant portion of the proteins required for chloroplast function are encoded in the nucleus and are transported into the chloroplast, where they interact with the chloroplast genome and its gene products (Sloan et al. 2014; Rockenbach et al. 2016; Weng et al. 2016; Williams et al. 2019; Forsythe et al. 2021). The interaction between the chloroplast and nuclear genomes has been well documented in numerous plant species (Huang et al. 2014; Kawabe et al. 2018; Li et al. 2021) and is crucial for overall fitness, as demonstrated by the frequent involvement of plastid–nuclear incompatibilities in reproductive isolation (Schmitz-Linneweber et al. 2005; Greiner et al. 2011; Bogdanova et al. 2015; Barnard-Kubow et al. 2016; Zupok et al. 2021).

The detection of evolutionary rate covariation (ERC) through analysis of the correlation between evolution rates of nuclear and plastid genes has proven to be a valuable approach to identify coevolving and cofunctional genes (i.e., groups of proteins that contribute to a shared function via physical, epistatic, or regulatory interactions). The ERC method, which is based on the premise that functionally related genes that coevolve exhibit correlated changes in their evolution rates across phylogenies, has been widely utilized in studies of fungi, insects, and mammals (Osada and Akashi 2012; Barreto and Burton 2013; Clark et al. 2013; Yan et al. 2019; Steenwyk et al. 2022). However, the application of ERC analysis to identify nuclear candidate genes with coevolution in plant lineages has been limited, with only a few studies conducted in Geraniaceae and *Silene* species (Sloan and Taylor 2012; Weng et al. 2016; Williams et al. 2019). In a recent study, Forsythe et al. (2021) applied ERC analysis using whole nuclear and plastid genome data to major clades of angiosperms and identified hundreds of nuclear genes that could interact with plastid proteins, demonstrating the potential of ERC analysis as a powerful tool for investigating plastid–nuclear interactions in plants.

The family Juglandaceae, belonging to the order Fagales, is comprised of three subfamilies, Rhoipteleoideae (*Rhoiptelea*), Engelhardioideae (*Engelhardtia*, *Oreomunnea*, *Alfaroa*), and Juglandoideae (*Carya*, *Platycarya*, *Cyclocarya*, *Juglans*, and *Pterocarya*) (Lu et al. 1999; Zhang et al. 2021). The species within Rhoipteleoideae and Engelhardioideae are primarily distributed in subtropical and tropical forests, whereas those within Juglandoideae are prevalent in temperate deciduous forests of the Northern Hemisphere (Lu et al. 1999; Zhang et al. 2021). Previous studies have identified a few plastid genes that have undergone positive selection within the Juglandaceae. For instance, Xu et al. (2021) reported a positive selection of the plastid gene *ycf1* in three species of Asian butternuts, and Hu et al. (2016) observed that five plastid genes (*matK*, *ycf1*, *accD*, *rps3*, and *rpoA*) experienced positive selection between sect. *Cardiocaryon* and

sect. *Dioscaryon* in *Juglans*. These plastid genes, which have undergone positive selection, are likely to have elevated rates of protein evolution. In recent years, high-quality, chromosome-level nuclear genomes have been published for several Juglandaceae species. The availability of these genomes facilitates the accurate identification of orthologous nuclear genes, which is a crucial step in detecting ERC. Thus, the abundance of genomic data and the occurrence of plastid genes with elevated evolution rates make Juglandaceae an ideal group for investigating plastid–nuclear interactions through ERC analysis.

In this study, we aimed to investigate the potential plastid–nuclear interactions and the ERC in the Juglandaceae family. To this end, we applied the method of Forsythe et al. (2021) to identify potential novel plastid–nuclear interactions and investigate any differences in the number and evolution rates of nuclear genes that exhibit significant signatures of ERC with the plastid genome (i.e., “ERC hits”). We partitioned the plastid genome into seven functional partitions, including *AccD*, *MatK*, *ClpP1*, Photosynthesis, Ribosomal, RNA polymerase, and *Ycf1/2*, and performed parallel ERC analyses comparing each of these partitions against all nuclear genes. Our results found that 27.97% of the nuclear genes demonstrated signatures of coevolution, and these genes were significantly enriched in the biological processes related to leaf development, photoperiodism, and abiotic stress response.

Results

Plastid Genome Features of Juglandaceae

We sequenced, assembled, and annotated the plastid genomes of nine Juglandaceae species, the accession numbers of which are listed in the supplementary material (supplementary table S1, Supplementary Material online). The plastid genome structure of these species was found to be consistent with the typical quadripartite organization, comprising two identical copies of inverted repeats, a long single-copy region, and a short single-copy region (supplementary fig. S1, Supplementary Material online). The plastid genome lengths ranged from 159,730 bp to 161,713 bp (supplementary table S1, Supplementary Material online). The overall GC content of the plastid genomes and the proportion of protein-coding genes varied between 35.9–36.2% and 37.1–37.4%, respectively. The plastid genomes of these species were found to contain a total of 114 genes, including 80 protein-coding genes, 30 tRNA genes, and 4 rRNA genes (supplementary table S1, Supplementary Material online). Additionally, we identified 18 genes with introns, among which six were tRNA genes (*trnA-UGC*, *trnG-UCC*, *trnL-UAA*, *trnI-GAU*, *trnK-UUU*, and *trnV-UAC*) and 12 were protein-coding genes (*atpF*, *clpP*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rps12*, *rpl16*, *rpoC1*,

rps16, and *ycf3*). Out of the 12 protein-coding genes, nine had a single intron, and three (*clpP*, *rps12*, and *ycf3*) had two introns. The absence of any rearrangements among the nine plastid genomes highlights the high level of structural conservation within Juglandaceae (supplementary fig. S2, Supplementary Material online).

Genome-wide ERC Analyses Reveal Correlated Evolution Between Plastid and Nuclear Genes

In accordance with the method proposed by Forsythe et al. (2021), we partitioned the genes encoded in the plastid genome into seven functional categories, AccD, MatK, ClpP1, Photosynthesis, Ribosomal, RNA polymerase, and Ycf1/2 (supplementary table S2, Supplementary Material online). The alignment lengths of the AccD, ClpP1, MatK, and Ycf1/2 partitions were 1,599 bp, 588 bp, 1,521 bp, and 13,011 bp, respectively. The photosynthesis partition, with the longest alignment of 31,959 bp, encompassed 46 genes. The Ribosomal partition consisted of 8,376 bp and 21 genes, while the RNA polymerase partition, comprising four genes, had an alignment length of 10,515 bp. Phylogenetic trees were then constructed for each partition (fig. 1).

Two previous studies on the Juglandaceae family showed that it has an allopolyploid origin (Zhu et al. 2019; Zhang et al. 2020). Subgenome assignments were made for nine Juglandaceae species and 4,894 orthologous nuclear genes were obtained (see Materials and Methods). A genome-wide scan was then conducted to identify potential instances of plastid–nuclear ERC by testing 34,258 pairwise correlations (seven plastid trees $\times$ 4,894 nuclear trees = 34,258 comparisons), based on normalized branch lengths, as described in supplementary table S3, Supplementary Material online. Applying the criteria of Pearson $P \leq 0.05$ and Spearman $P \leq 0.1$, we identified a total of 1,369 significant ERC hits in the seven plastid partitions (supplementary table S4, Supplementary Material online). The number of hits of the seven partitions is as follows, AccD: 55 hits; ClpP1: 13 hits; MatK: 117 hits; Photosynthesis: 284 hits; Ribosomes: 314 hits; RNA polymerase: 80 hits; Ycf1/2: 765 hits. We also applied separate, more stringent, filtering criteria of false discovery rate (FDR)-adjusted Pearson $P \leq 0.05$ or FDR-adjusted Spearman $P \leq 0.05$ to identify the strongest ERC hits, yielding two hits from the MatK partition and 56 hits from the Ycf1/2 partition (supplementary table S5, Supplementary Material online).

Subcellular Localization of ERC Hits

To ask whether our ERC hits are enriched for functional characteristics tied to their coevolution with plastid partitions, we identified orthologous genes in *Arabidopsis* and used these orthologs to map functional annotations from *Arabidopsis* to our Juglandaceae gene families. We analyzed

the ERC hits that are known to directly interact with plastid-encoded proteins, as annotated by the Cytonuclear Molecular Interactions Reference for *Arabidopsis* (CyMIRA) classification (Forsythe et al. 2019). Our analysis revealed a marked increase in the degree of enrichment for plastid localization and interaction in the ERC hits compared to the general classification (approximately 4–8 times more enriched, depending on the plastid partition) (fig. 2A, supplementary table S6, Supplementary Material online). On the other hand, no significant enrichment was observed among the ERC hits for either mitochondrion-targeted proteins or mitochondrion-interacting proteins. These results provide evidence that correlated plastid–nuclear evolution is a widespread phenomenon across nuclear genomes and can be detected through ERC analysis.

Gene Ontology Functional Enrichment of ERC Hits

We performed gene ontology (GO) enrichment analyses, revealing that the ERC hits were overrepresented in cellular components related to the plastid, such as the “plastid envelope” and “intracellular membrane-bounded organelle” (fig. 2B, supplementary table S7, Supplementary Material online). Enrichment analysis of biological processes indicated that the ERC hits were associated with plant growth-related processes, including “shoot system development”, “postembryonic development”, and “phyllome development”, and three metabolic processes, positive regulation of biological processes, and DNA-templated transcription (fig. 2B, supplementary table S7, Supplementary Material online). Additionally, the ERC hits were found to be highly enriched in the terms “response to abiotic stimulus” and “regulation of response to external stimulus”, which may relate to the plant defense system. The molecular functions of the ERC hits were also found to be overrepresented in terms such as “DNA-binding transcription factor activity”, “transcription cis-regulatory region binding”, “protein binding”, “kinase inhibitor activity”, and “oxidoreductase activity” (fig. 2B, supplementary table S7, Supplementary Material online).

As a comparison, we also performed CyMIRA and GO enrichment analyses using the 1,999 mapped genes as the foreground and the full genome as the background. In these analyses, we observed significant enrichment for the same terms which we show in figure 2 (supplementary table S7, Supplementary Material online). This indicates that *Arabidopsis* ortholog mapping may have introduced some degree of bias, resulting in enrichment potentially being tied to our mapping process rather than ERC results in some cases (see Discussion).

dN/dS Ratios for Plastid Partitions and ERC Hits

Since dN/dS test may be unable to detect positive selection on only one or a few substitutions of a long sequence, no matter how strong selection may be (Hahn 2018), we conducted

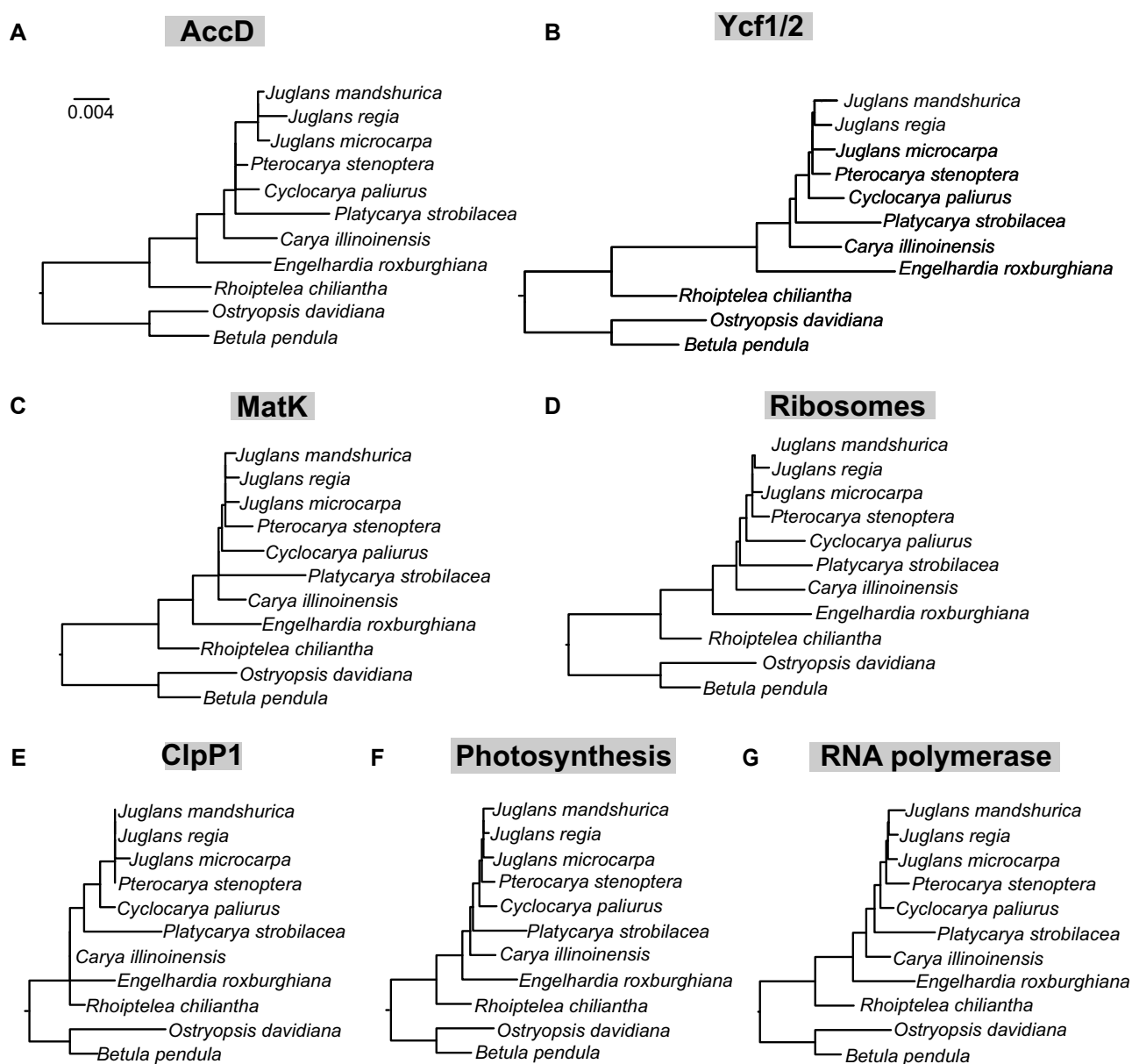


FIG. 1.—Trees based on plastid genome (plastome) partitions. The plastid genes are partitioned into seven functional categories (as described in [supplementary table S4, Supplementary Material](#) online). Branch lengths inferred from nucleotide alignments are shown on the same scale for all trees to highlight differences in rates of nucleotide evolution between partitions. Each plastome partition tree was used for ERC analysis against all nuclear gene trees.

calculations of dN/dS values on individual plastid gene across seven partitions (a total of 76 genes) for each of nine Juglandaceae species. We identified five genes from partition of Photosynthesis, Ribosomal, and Ycf1/2 (*rpl23*, *rpl22*, *rps14*, *psbK*, and *ycf2*) in eight species with dN/dS values exceeding 1, although they did not reach statistical significance (fig. 3A) ([supplementary table S8, Supplementary Material](#) online). All the species for AccD, ClpP1, MatK, and RNA polymerase had a dN/dS value less than 1, showing low variation (fig. 3A). We also found that a total of 53 genes in nine species have a dN/dS value of significantly less than 1, exhibiting

evidence of purifying selection ([supplementary table S8, Supplementary Material](#) online).

Among the 1,369 hits, a total of 987 hits were found to have dN/dS values significantly less than one, indicating the existence of purifying selection ([supplementary table S8, Supplementary Material](#) online). The hits for Photosynthesis and Ribosomes have dN/dS values more than one ([supplementary table S8, Supplementary Material](#) online), but only one hit (onlyS-OG0008075) from the Photosynthesis partition exhibited a significant dN/dS value greater than one in seven species (*Carya*

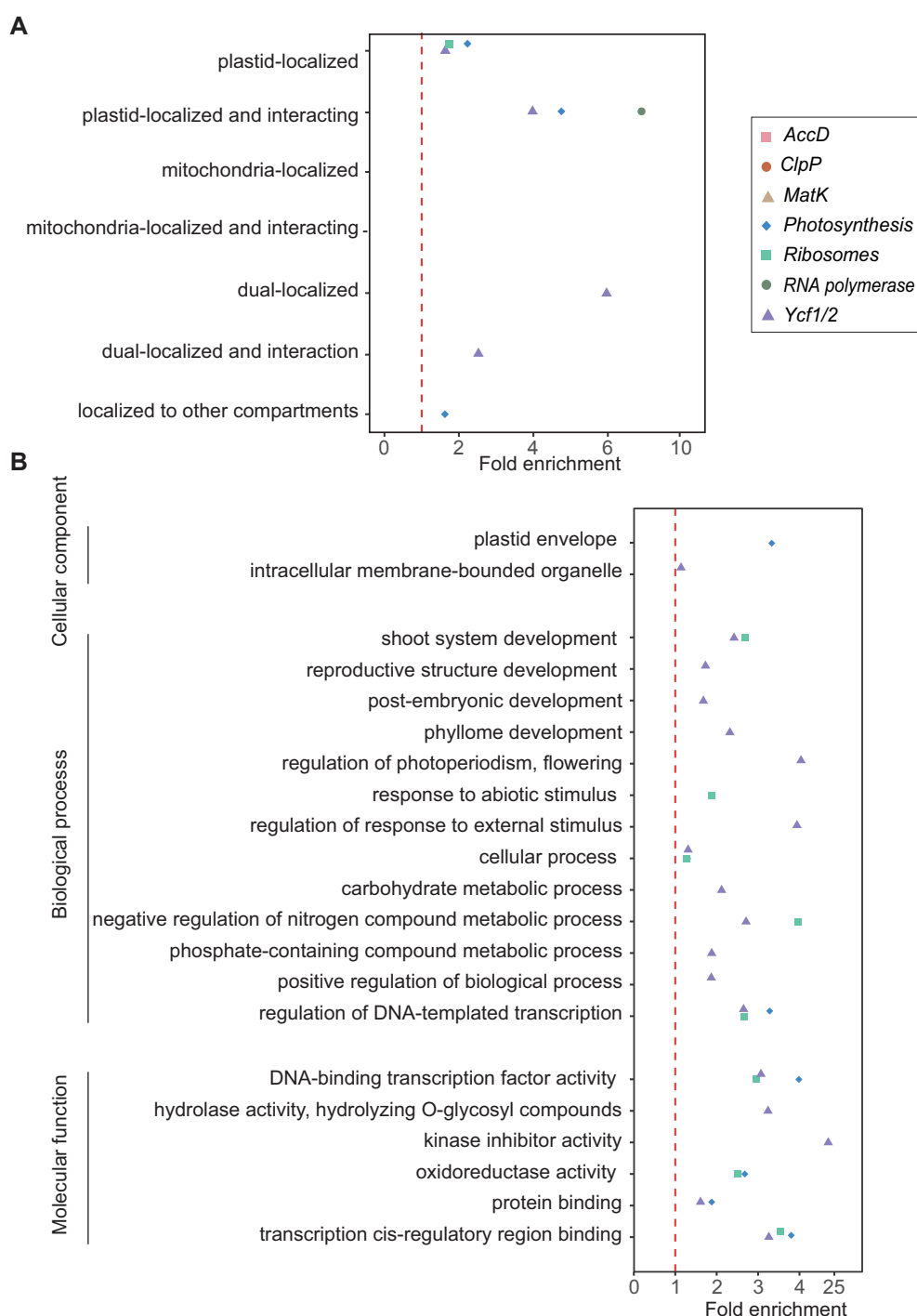


Fig. 2.—Significant results from function enrichment analyses of ERC hits. Enrichment scores are shown only for partitions/categories with significant enrichment. (A) Subcellular localization and cytonuclear interactions of ERC hits. Proteins encoded by genes exhibiting signatures of coevolution with seven chloroplast partitions were analyzed for their localization and interactions, as classified by the CyMIRA database (Forsythe et al. 2019). Categories indicating “interacting” refer to nucleus-encoded proteins predicted to directly physically interact with organelle-encoded proteins. The statistical significance of enrichment/depletion (Fisher’s exact test) is indicated by filled points ($P < 0.05$). (B) GO functional enrichment analyses were performed for ERC hits from seven chloroplast partitions. Categories are grouped by type of GO annotation (cellular component, biological process, and molecular function). Some highly overlapping categories were removed (see [supplementary data, Supplementary Material](#) online for full results). The statistical significance of enrichment/depletion (Fisher’s exact test) is indicated by filled points ($P < 0.05$). P values were corrected for multiple tests using the FDR.

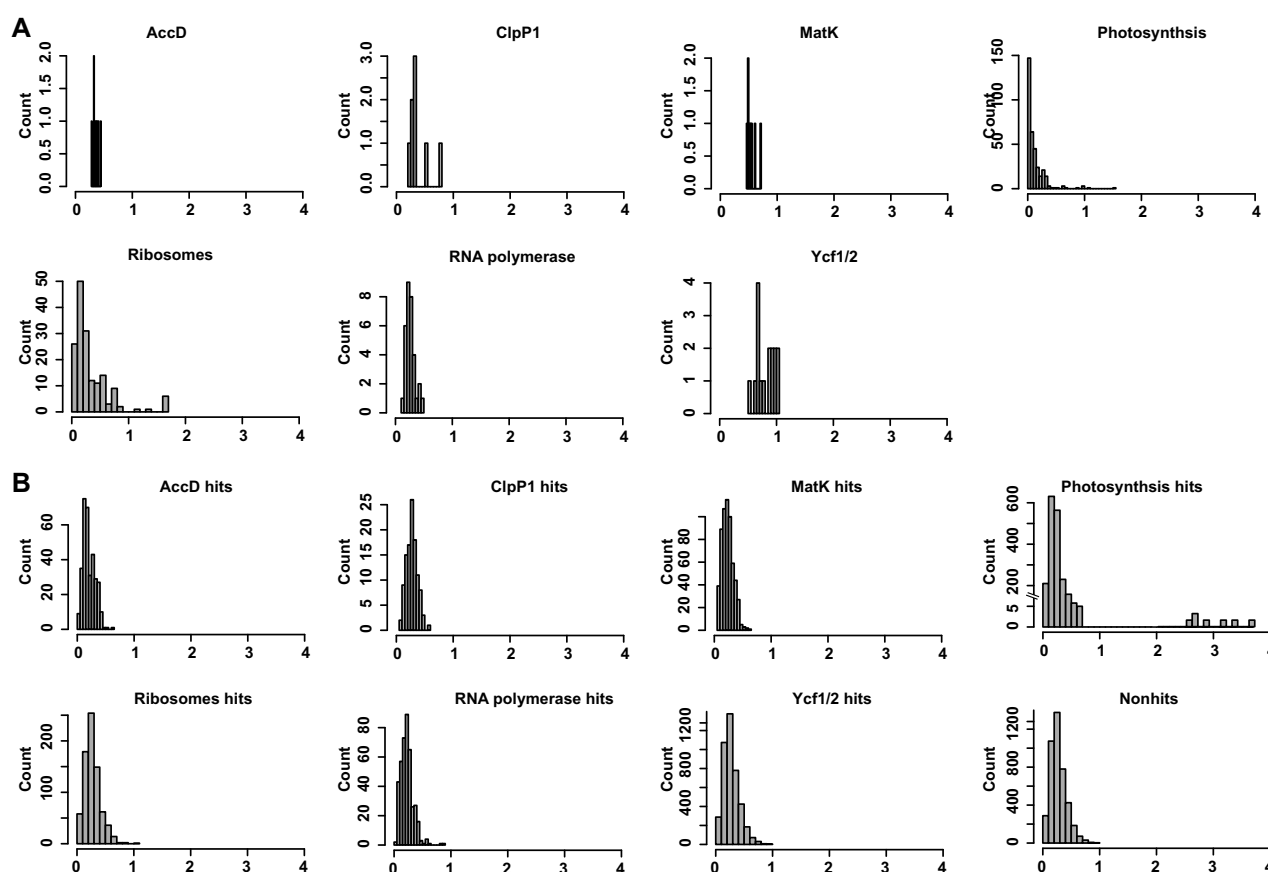


FIG. 3.—The dN/dS distributions of plastid partition, nuclear hits, and nonhits genes for nine Juglandaceae species with *B. pendula* as outgroup. The X axis and Y axis represent the dN/dS value and the number of genes, respectively. A) The dN/dS distributions of plastid genes of seven plastid partitions; B) The dN/dS distributions of nuclear hits (seven partitions) and nonhits.

illinoensis, *Engelhardtia roxburghiana*, *Juglans regia*, *Juglans microcarpa*, *Juglans mandshurica*, *Pterocarya stanoptera*, and *Rhoiptelea chiliantha*), indicating a potential signal of positive selection. We further tested whether dN/dS value of this hit would be caused by relaxed selection and found that only *E. roxburghiana* cannot rule out relaxed negative selection (supplementary table S9, Supplementary Material online).

The nuclear genes that did not exhibit significant signatures of ERC with any plastid partition (referred to as “non-hits”) comprise a total of 3,525 genes. There was a significant difference in the distributions of dN/dS between the hits and nonhits for the six partitions ($P < 0.05$) and marginal significance for the MatK partition ($P = 0.09$) with the rank sum test (Wilcox test).

Discussion

Overcoming Challenges of Gene/Genome Duplication in Plant ERC

ERC has been widely applied to identify novel interactions in animals and fungi (Findlay et al. 2014; Raza et al.

2019; Yan et al. 2019; Steenwyk et al. 2022). However, the application of this approach at broad phylogenetic scales in plants has been limited by the prevalence of gene/genome duplication. The occurrence of gene duplication or loss can lead to erroneous predictions of orthologous nuclear genes (Bansal and Eulenstein 2008; Stolzer et al. 2012; Smith and Hahn 2021; Xiong et al. 2022). Whole-genome duplication events, which are episodic in nature (Bowers et al. 2003; Jiao et al. 2014), tend to result in the subsequent loss of a majority of duplicated genes within a few million years (Lynch and Conery 2000). This differential loss of duplicated copies among species can lead to the presence of “hidden paralogues” or “pseudo-orthologs” in single-copy gene families (Smith and Hahn 2021). Two previous studies on the Juglandaceae family showed that it has an allopolyploid origin (Zhu et al. 2019; Zhang et al. 2020). Fortunately, this same study performed an in-depth analysis of subgenome retention following duplication, minimizing the risk of hidden paralogues. By using the methods described by Ding et al. (2023), which depend on the number of retained ancestrally inherited genes on chromosomes, we were able to

assign each of the homoeologous chromosomes of the nine Juglandaceae species to one of 18 subgenomes and correctly identify 3,819 and 2,468 orthologous nuclear genes from dominant and recessive subgenomes. This allowed us to confidently identify true orthologues, thus sidestepping a common source of error that hinders many plant phylogenomic analyses.

Rate Heterogeneity and Statistical Power in ERC Analyses

The application of the ERC method to examine cytonuclear coevolution in plants encounters a second notable challenge, namely the relatively slow rate of evolution and low levels of heterogeneous evolution of most plastid genes, when compared to nuclear genes or animal mitochondrial genomes (Wolfe et al. 1987). Since ERC relies on rate heterogeneity across a phylogeny (differences in branch lengths/protein evolution across the tree), highly conserved homogenous sequences provide very little power for detecting ERC signatures. However, Gonçalves et al. (2019) showed that plastid genes are not a uniform entity, often exhibiting differential rates of evolution. Indeed, we observe differential rate heterogeneity (branch length) between distinct partitions (supplementary fig. S3, Supplementary Material online).

The dN/dS analysis results for individual plastid genes in Juglandaceae species revealed variations in the rates of evolutionary change among these partitions (fig. 3). Our analysis revealed that 53 plastid genes across the seven partitions and 987 nuclear hits genes exhibited significant evidence of undergoing purifying selection. This is consistent with the idea that most plastid genes and their interacting nuclear genes serve conserved roles in the metabolism of Juglandaceae. In some instances, plastid genes in Photosynthesis partition showed some evidence of evolving under positive selection, although this signature did not reach statistical significance in most cases. In particular, we observed that one hit from the Photosynthesis partition exhibited a significant signal of positive selection. The functional annotation of this gene remains uncharacterized, as an orthologous gene in *Arabidopsis* could not be identified; however, we postulate that this gene could potentially underlie adaptive evolution of photosynthetic mechanisms in Juglandaceae. Further validation is required in the future. The presence of purifying and positive selection may contribute to the observed heterogeneity.

In contrast to Forsythe et al. (2021) which focuses on broadly sampled set of species spanning angiosperms, our study focuses on a single family. This contrast provides an important assessment of the power and sensitivity of ERC applied to closely related species. Despite difference in evolutionary scale, we found that 15 hits shared with Forsythe et al. (2021) (supplementary table S10,

Supplementary Material online). These shared hits provide valuable insights into the functional relationships and the potential for coevolutionary forces to act at many scales of evolution.

Widespread Plastid–Nuclear ERC With Ycf1/2

Consistent with our expectations (discussed above), ERC analyses of the Ycf1/2 partition yield the highest frequency of ERC hits. The striking signatures of evolution of Ycf1/2 genes and coevolution with nuclear genes may be tied to a pivotal role in mediating cytonuclear interactions in Juglandaceae. In photosynthetic eukaryotes, a considerable number of proteins are transferred from the cytosol to the chloroplasts through the concerted action of two translocon complexes located in the outer and inner membranes of the chloroplast envelope, known as the TOC and TIC (translocon at the outer and inner envelope membrane of chloroplasts) complexes, respectively (Ramundo et al. 2020). Ycf1 encodes Tic214, which plays a crucial role in the TIC complex of *Arabidopsis* (Kikuchi et al. 2013; De Vries et al. 2015; Nakai 2015, 2018), enabling the transport of proteins across the inner membrane, the intermembrane space, and the outer membrane, and thus connecting the TOC and TIC proteins (Liu et al. 2023). As such, Ycf1 plays a role in mediating the import of all plastid-targeted nuclear genes. It stands to reason that Ycf1 would be the target of shifts in the selection that impact large suites of interacting plastid–nuclear genes, especially those that are involved in communication between the nucleus and plastid. Pursuing the specific conditions that may have led to these patterns of evolution and coevolution in Ycf1/2 could help reveal the unresolved functions of these essential genes.

Functional Enrichment of ERC Hits

The present study utilized CyMIRA and GO enrichment analyses to investigate the functional implications of the genes identified through the use of ERC. The results revealed that the identified genes associated with plastids are linked to processes involved in leaf and shoot development, as well as photoperiodism regulation. The importance of plastids in photosynthesis, which relies on various nuclear-encoded proteins, has been well established (Nam et al. 2021). It has been estimated that approximately 3,000 nuclear-encoded proteins are localized in plastids (Richly et al. 2003; Richly and Leister 2004; Cui et al. 2006) which could explain the enrichment of ERC hits in these functional categories. Interestingly, our study also found that ERC hits were enriched in terms related to abiotic and external stimulus, which may have implications for the regulation of plant defense mechanisms. Previous studies have highlighted the crucial role played by chloroplasts in plant immune responses, in addition to their photosynthetic functions (Nomura et al. 2012; Serrano et al. 2016; Kretschmer

et al. 2019; Medina-Puche et al. 2020; Kachroo et al. 2021). In response to adverse signals, chloroplasts can trigger the transmission of signals to the nucleus through calcium and reactive oxygen species, which activate defense-related gene expression (Medina-Puche et al. 2020). Juglandaceae species, such as *Juglans cinerea* and *J. regia*, are susceptible to various fungal infections, including butternut canker and anthracnose disease (Michler et al. 2006; Wang et al. 2017). Thus, the present findings suggest that the identified ERC hits may serve as promising candidate genes for transmitting signals from the cell periphery to the chloroplast and from the chloroplast to the nucleus in Juglandaceae species. This could provide valuable insights into retrograde signaling mechanisms in regulating plant defense mechanisms.

It should be noted that our application of ERC in a nonmodel system required that we draw functional annotations from *Arabidopsis*. This degree of separation from the original annotation source reduced our statistical power and may have introduced a layer of directional bias into our functional enrichment analyses (supplementary table S11, Supplementary Material online). Despite these challenges, the functional categories we identify and particularly the individual strong ERC hit genes within those categories represent valuable candidates for revealing the mechanisms that underlie plastid–nuclear coevolution in Juglandaceae species.

In conclusion, the results of ERC analysis presented in this study provide a scalable method to predict gene function and localization, complementing traditional biochemical approaches. Our findings extend the list of putative nuclear genes that may be involved in plastid–nuclear interactions in Juglandaceae. The identification of these candidate genes represents a significant step towards elucidating the molecular mechanisms underlying the interactions between plastids and the nucleus in the Juglandaceae family. Future functional validation studies are necessary to fully characterize these interactions and to enhance our understanding of the biological implications of these processes.

Materials and Methods

Obtaining the Orthologous Nuclear Genes and Phylogenetic Inference

In this study, we acquired the genomic datasets of nine Juglandaceae species, including *Cyclocarya paliurus* (Zheng et al. 2021), *Ca. illinoensis* (Lovell et al. 2021), *E. roxburghiana* (Ding et al. 2023), *J. regia* (Zhu et al. 2019), *J. microcarpa* (Zhu et al. 2019), *J. mandshurica* (Zhang et al. 2022), *Platycarya strobilacea* (Zhang et al. 2019), *Pt. stenoptera* (Zhang et al. 2022), and *R. chiliantha* (Ding et al. 2023), as well as two outgroup species, *Betula pendula* (Salojärvi et al. 2017) and *Ostryopsis davidiana* (Wang et al. 2021). A summary of the species used in the

study is presented in the supplementary material (supplementary table S12, Supplementary Material online). To ensure accuracy in our analysis, the transcript with the longest coding sequence was selected for each gene when multiple transcripts were available.

The Juglandaceae species are of allopolyploid origin (Zhu et al. 2019; Zhang et al. 2020). In lineages of allopolyploid origin, entire parental subgenomes may coexist, with two or more sets of homoeologous chromosomes that differ in gene content and expression patterns. Generally, the dominant subgenome has preferentially retained higher gene content compared with recessive subgenome (Edger et al. 2018). Based on the method of Ding et al. (2023), the gene families for the dominant and recessive subgenomes of the nine Juglandaceae species and two outgroup species (*B. pendula* and *O. davidiana*) were obtained using OrthoFinder (Emms and Kelly 2019) with default parameters (supplementary table S13, Supplementary Material online). The analysis was limited to strict one-to-one orthologous relationships with fewer than three species absent in the gene family and with information for at least one outgroup. This resulted in 3,819 single-copy orthogroups from nine dominant subgenomes and two outgroups, and 2,468 single-copy orthogroups from nine recessive subgenomes and two outgroups. When aligning DNA sequences encoding proteins, nucleotide natural selection occurs at the level of a codon, rather than a single nucleotide. Based on that, we used the perl script PAL2NAL (Suyama et al. 2006) to perform codon-guided nucleotide sequence alignment. Nucleotide alignments were subsequently used for phylogenetic inference and dN/dS analyses. Orthogroups with a gap frequency ("—" character) greater than 50% were excluded from subsequent analysis. Ultimately, 4,894 gene families remained for the construction of gene trees by RAxML (Stamatakis 2014), with 100 bootstrap replicates. Tree inference was performed on nuclear sequence alignments with the following command for each gene, using the generalized time reversible (GTR) + optimization of substitution rates + GAMMA model of rate heterogeneity. The seed for a parsimony search and rapid bootstrapping was provided by the -p and -x arguments, respectively, and the number of bootstrap replicates was specified by the -# argument. The -f argument implemented rapid bootstrap analyses and the best-scoring tree search, and the -T argument indicated the number of threads used for parallel computing. The input file name and output file name were specified by the -s and -n arguments, respectively, while the outgroup was designated by the -o argument.

```
raxmlHPC-SSE3 -s <input file name> -n <output file name> -m GTRGAMMA -o <outgroup> -p 12345 -x 12345 -# 100 -f a
```

Plastid Genome Assembly and Phylogenetic Inference

To obtain the plastid genomes of the nine Juglandaceae species, fresh young leaves were sampled from the wild

and total genomic DNA was extracted from dried leaf tissue using a plant total genomic DNA kit (Omega, Norcross, GA, USA) (detailed sampling information in [supplementary table S14, Supplementary Material](#) online). Whole-genome resequencing was performed using paired-end libraries with an insert size of 350 bp on Illumina NovaSeq 6000 instruments by NovoGene (Beijing, China). The plastid genomes were assembled using the program GetOrganelle (Jin et al. 2020) with default settings and annotated using PGA software (Qu et al. 2019). Ambiguous regions were checked and manually corrected in Geneious Prime (Kearse et al. 2012), and plastid genome maps were generated using OGDRAW (Greiner et al. 2019). In addition, the 80 shared protein-coding genes of 11 species were extracted and aligned using MAFFT v7.47573 (Katoh and Standley 2013) and pal2nal v. 1474 (Suyama et al. 2006). The corresponding gene sequences were either analyzed individually or concatenated from multiple plastid genes to achieve “plastome partition” alignments, following the method of Forsythe et al. (2021). The sequences were optimized using RAXML and branch lengths were optimized on a constraint tree, as indicated by the use of the `-f e` argument in the RAXML command.

```
raxmlHPC-SSE3 -s <input file name> -n <output file name> -g <name of constraint tree file> -o <outgroup> -m GTRGAMMA -p 12345 -f e
```

Inference of Evolutionary Rate Variation

The evolutionary rate variation (ERC) analysis was conducted following the pipeline described by Forsythe et al. (2021), which involved a root-to-tip approach. This approach measures the cumulative length of branches from the species' tips to the node representing the most recent common ancestor of all species tips, providing a phylogeny-based assessment of nucleotide substitution in each lineage. The ERC analysis was performed on all orthologous nuclear gene trees and plastid partition trees, excluding species with missing values from the analysis. To account for lineage-specific differences in genome-wide evolution rates, the branch length of each species was normalized by dividing it by the average branch length across all genes analyzed. The resulting normalized branch length values were utilized for pairwise ERC comparisons.

A correlation analysis was carried out between the branch lengths in the plastid and orthologous nuclear gene trees, calculating Pearson and Spearman correlation coefficients. To account for multiple comparisons, the *P* values were adjusted using the false discovery rate (FDR) method in R. Both Pearson and Spearman *P* values were used as metrics due to their complementing strengths, where Pearson is more sensitive to long-branch lengths, potentially indicative of coaccelerated evolution, and

Spearman is less sensitive to outliers and is generally considered to be more prone to false negatives. Similar to Forsythe et al. (2021), we applied one filtering strategy in the context of identifying large groups of genes with elevated ERC signatures and a separate, more stringent, strategy in the context of identifying the strongest individual ERC hits. For our group-level enrichment analyses, a gene was considered a hit if both (uncorrected) Pearson *P* value ≤ 0.05 and (uncorrected) Spearman *P* value ≤ 0.1 . For individual gene analyses, a gene was considered a strong hit if either the adjusted Pearson *P* value or the adjusted Spearman *P* value (or both) were ≤ 0.05 . Additionally, for the genes in which only one of the two adjusted *P* was ≤ 0.05 , we also required that the raw Pearson and raw Spearman *P* value both be ≤ 0.05 .

CyMIRA and GO Functional Enrichment Analyses

We performed a BLASTP analysis with a stringent *e*-value cutoff of 10^{-10} to identify orthologous genes in *Arabidopsis thaliana* based on the orthologous nuclear genes from the Juglandaceae species. *Arabidopsis* orthologs were successfully mapped for 1,999 of 4,894 (~40.8%) gene families. Subsequently, to assess the functional enrichment of the identified orthologous genes, we integrated localization/interaction annotations from the CyMIRA database and functional annotations from the GO database. In instances where multiple *Arabidopsis* paralogues were present within a gene family, we chose a single paralogue at random to represent the family, and only *Arabidopsis* genes with GO annotations were included in the analysis. The list of cytonuclear interacting protein genes in *Arabidopsis* was obtained from the CyMIRA database (Forsythe et al. 2019). Fold enrichment was calculated as the ratio of the number of observed hits in a particular category to the expected number of hits, which was determined by multiplying the proportion of the background in that category by the number of hits. The statistical significance of the observed fold enrichment was evaluated using Fisher's exact test with the `fisher.test()` function in R. Furthermore, GO enrichment analysis was conducted using the PANTHER web-based tool (<http://geneontology.org/>) (database release from October 13, 2022), and a FDR correction was applied. We performed CyMIRA and GO enrichment analyses using the 1,999 genes as the “foreground” and the full genome as the background.

dN/dS Analysis for Plastid Partitions and ERC Hits

For each plastid gene and ERC hit of each plastid partition, the dN/dS ratio of each of nine Juglandaceae species with *B. pendula* as outgroup species was calculated using KaKs_Calculator 2.0 (Wang et al. 2010), based on the Gamma-MYN algorithm (Wang et al. 2009). The

Gamma-MYN algorithm operates in a pairwise fashion, meaning each dN/dS value is calculated by comparing the outgroup sequence with the sequence of one of the nine Juglandaceae species. We used RELAX (Wertheim et al. 2015) to detect whether species with $dN/dS > 1$ have undergone relaxed negative selection. Then, we performed a rank sum test using *wilcox.test* in R to determine whether there are any significant differences in the distribution of hits and nonhits.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

The authors thank Bo-wen Zhang, Xiao-xu Pang, and Yu Liang for their assistance in data analysis. The authors thank Wei-ping Zhang, Yu Cao, and Rui-min Yu for the sample collection.

Authors' Contributions

W.-N.B. conceived this research. Y.Y. and Y.-M.D. analyzed the data. Y.Y. and W.-N.B. wrote the manuscript. D.-Y.Z. and E.F. revised the manuscript. All authors read and approved the final version of the manuscript.

Funding

The study was supported by the National Natural Science Foundation of China (31421063), the National Key R & D Program of China (2017YFA0605104), the National Science Foundation of the United States (IOS-2114641), the "111" Program of Introducing Talents of Discipline to Universities (B13008), and a key project of State Key Laboratory of Earth Surface Processes and Resource Ecology.

Data Availability

All the new chloroplast genomes in the present study were submitted to the National Center for Biotechnology Information (NCBI) database with accession numbers OP837959–OP837967 (supplementary table S1, Supplementary Material online). The newly resequenced raw reads from five individuals have been deposited in GenBank of submission number SUB12353115 under PRJNA356989.

Literature Cited

Bansal MS, Eulenstein O. 2008. The multiple gene duplication problem revisited. *Bioinformatics* 24:i132–i138.

- Barnard-Kubow KB, So N, Galloway LF. 2016. Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution* 70:2752–2766.
- Barreto FS, Burton RS. 2013. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Mol Biol Evol*. 30:310–314.
- Bogdanova VS, Zaytseva OO, Mglinets AV, Shatskaya NV, Kosterin OE, et al. 2015. Nuclear-cytoplasmic conflict in pea (*Pisum sativum* L.) is associated with nuclear and plastid candidate genes encoding acetyl-coA carboxylase subunits. *PLoS One*. 10:e0119835.
- Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:428–433.
- Clark NL, Alani E, Aquadro CF. 2013. Evolutionary rate covariation in meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. *Genetics* 193:529–538.
- Cui L, Veeraraghavan N, Richter A, Wall K, Jansen RK, et al. 2006. ChloroplastDB: the chloroplast genome database. *Nucleic Acids Res*. 34:D692–D696.
- De Vries J, Sousa FL, Bölter B, Soll J, Gould SB. 2015. YCF1: a green TIC? *Plant Cell*. 27:1827–1833.
- Ding YM, Pang XX, Cao Y, Zhang WP, Renner SS, et al. 2023. Genome structure-based Juglandaceae phylogenies contradict alignment-based phylogenies and substitution rates vary with DNA repair genes. *Nat Commun*. 14:617.
- Edger PP, McKain MR, Bird KA, VanBuren R. 2018. Subgenome assignment in allopolyploids: challenges and future directions. *Curr Opin Plant Biol*. 42:76–80.
- Emms DM, Kelly S. 2019. Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 20:238.
- Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL, et al. 2014. Evolutionary rate covariation identifies new members of a protein network required for *Drosophila melanogaster* female post-mating responses. *PLoS Genet*. 10:e1004108.
- Forsythe ES, Sharbrough J, Havird JC, Warren JM, Sloan DB. 2019. CyMIRA: the cytonuclear molecular interactions reference for *Arabidopsis*. *Genome Biol Evol*. 11:2194–2202.
- Forsythe ES, Williams AM, Sloan DB. 2021. Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. *Plant Cell*. 33:980–997.
- Gonçalves DJP, Simpson BB, Ortiz EM, Shimizu GH, Jansen RK. 2019. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Mol Phylogenet Evol*. 138:219–232.
- Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res*. 47:W59–W64.
- Greiner S, Rauwolf U, Meurer J, Herrmann RG. 2011. The role of plastids in plant speciation. *Mol Ecol*. 20:671–691.
- Hahn MW. 2018. Molecular population genetics. New York: Sinauer Associates. p. 153.
- Hu Y, Woeste KE, Zhao P. 2016. Completion of the chloroplast genomes of five Chinese *Juglans* and their contribution to chloroplast phylogeny. *Front Plant Sci*. 7:1955.
- Huang DI, Hefer CA, Kolosova N, Douglas CJ, Cronk QCB. 2014. Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, *Populus balsamifera* and *P. trichocarpa* (Salicaceae). *New Phytol*. 204:693–703.
- Jiao Y, Li J, Tang H, Paterson AH. 2014. Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell*. 26:2792–2802.

- Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, et al. 2020. Getorganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21:241.
- Kachroo P, Burch-Smith TM, Grant M. 2021. An emerging role for chloroplasts in disease and defense. *Annu Rev Phytopathol.* 59:423–445.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Kawabe A, Nukii H, Furihata HY. 2018. Exploring the history of chloroplast capture in *Arabidopsis* using whole chloroplast genome sequencing. *Int J Mol Sci.* 19:602.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.
- Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, et al. 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* 339:571–574.
- Kretschmer M, Damoo D, Djamei A, Kronstad J. 2019. Chloroplasts and plant immunity: where are the fungal effectors? *Pathogens* 9:19.
- Li HT, Luo Y, Gan L, Ma PF, Gao LM, et al. 2021. Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol.* 19:232.
- Liu H, Li A, Rochaix JD, Liu Z. 2023. Architecture of chloroplast TOC–TIC translocon supercomplex. *Nature* 615:349–357.
- Lovell JT, Bentley NB, Bhattarai G, Jenkins JW, Sreedasyam A, et al. 2021. Four chromosome scale genomes and a pan-genome annotation to accelerate pecan tree breeding. *Nat Commun.* 12:4441.
- Lu A, Stone DE, Grauke LJ. 1999. Juglandaceae. In: Wu ZY and Raven PH, editors. *Flora of China*. Beijing, St. Louis: Science Press and Missouri Botanical Garden Press. p. 277–285.
- lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Medina-Puche L, Tan H, Dogra V, Wu M, Rosas-Diaz T, et al. 2020. A defense pathway linking plasma membrane and chloroplasts and co-opted by pathogens. *Cell* 182:1109–1124.e25.
- Michler CH, Pijut PM, Jacobs DF, Meilan R, Woeste KE, et al. 2006. Improving disease resistance of butternut (*Juglans cinerea*), a threatened fine hardwood: a case for single-tree selection through genetic improvement and deployment. *Tree Physiol.* 26:121–128.
- Nakai M. 2015. The TIC complex uncovered: the alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. *Biochim Biophys Acta Bioenerg.* 1847:957–967.
- Nakai M. 2018. New perspectives on chloroplast protein import. *Plant Cell Physiol.* 59:1111–1119.
- Nam HI, Shahzad Z, Dorone Y, Clowez S, Zhao K, et al. 2021. Interdependent iron and phosphorus availability controls photosynthesis through retrograde signaling. *Nat Commun.* 12:7211.
- Nomura H, Komori T, Uemura S, Kanda Y, Shimotani K, et al. 2012. Chloroplast-mediated activation of plant immune signalling in *Arabidopsis*. *Nat Commun.* 3:926.
- Osada N, Akashi H. 2012. Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. *Mol Biol Evol.* 29:337–346.
- Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods.* 15:50.
- Ramundo S, Asakura Y, Salomé PA, Strenkert D, Boone M, et al. 2020. Coexpressed subunits of dual genetic origin define a conserved supercomplex mediating essential protein import into chloroplasts. *Proc Natl Acad Sci USA.* 117:32739–32749.
- Raza Q, Choi JY, Li Y, O'Dowd RM, Watkins SC, et al. 2019. Evolutionary rate covariation analysis of E-cadherin identifies Raskol as a regulator of cell adhesion and actin dynamics in *Drosophila*. *PLoS Genet.* 15:e1007720.
- Richly E, Dietzmann A, Biehl A, Kurth J, Laloi C, et al. 2003. Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. *EMBO Rep.* 4:491–498.
- Richly E, Leister D. 2004. An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of *Arabidopsis* and rice. *Gene* 329:11–16.
- Rockenbach K, Havird JC, Monroe JG, Triant DA, Taylor DR, et al. 2016. Positive selection in rapidly evolving plastid-nuclear enzyme complexes. *Genetics* 204:1507–1522.
- Salojärvi J, Smolander OP, Nieminen K, Rajaraman S, Safronov O, et al. 2017. Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nat Genet.* 49:904–912.
- Schmitz-Linneweber C, Kushnir S, Babiychuk E, Poltnigg P, Herrmann RG, et al. 2005. Pigment deficiency in nightshade/tobacco cybrids is caused by the failure to edit the plastid ATPase α -subunit mRNA. *Plant Cell.* 17:1815–1828.
- Serrano I, Audran C, Rivas S. 2016. Chloroplasts at work during plant innate immunity. *J Exp Bot.* 67:3845–3854.
- Sloan DB, Taylor DR. 2012. Evolutionary rate variation in organelle genomes the role of mutational processes. Berlin (Germany): Springer. p. 123–146.
- Sloan DB, Triant DA, Wu M, Taylor DR. 2014. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Mol Biol Evol.* 31:673–682.
- Smith ML, Hahn MW. 2021. New approaches for inferring phylogenies in the presence of paralogs. *Trends Genet.* 37:174–187.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Steenwyk JL, Phillips MA, Yang F, Date SS, Graham TR, et al. 2022. An orthologous gene coevolution network provides insight into eukaryotic cellular and genomic structure and function. *Sci Adv.* 8:eabn0105.
- Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, et al. 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28:i409–i415.
- Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.
- Wang Q-H, Fan K, Li D-W, Niu S-G, Hou L-Q, et al. 2017. Walnut anthracnose caused by *Colletotrichum siamense* in China. *Australas Plant Pathol.* 46:585–595.
- Wang Z, Jiang Y, Bi H, Lu Z, Ma Y, et al. 2021. Hybrid speciation via inheritance of alternate alleles of parental isolating genes. *Mol Plant.* 14:208–222.
- Wang DP, Wan HL, Zhang S, Yu J. 2009. Gamma-MYN: a new algorithm for estimating Ka and Ks with consideration of variable substitution rates. *Biol Direct.* 4:20.
- Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. Kaks_calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics.* 8:77–80.
- Weng ML, Ruhlman TA, Jansen RK. 2016. Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. *Genome Biol Evol.* 8:1824–1838.
- Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* 32:820–832.
- Williams AM, Friso G, van Wijk KJ, Sloan DB. 2019. Extreme variation in rates of evolution in the plastid Clp protease complex. *Plant J.* 98:243–259.

- Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA*. 84:9054–9058.
- Xiong H, Wang D, Shao C, Yang X, Yang J, et al. 2022. Species tree estimation and the impact of gene loss following whole-genome duplication. *Syst Biol*. 71:1348–1361.
- Xu LL, Yu RM, Lin XR, Zhang BW, Li N, et al. 2021. Different rates of pollen and seed gene flow cause branch-length and geographic cytonuclear discordance within Asian butternuts. *New Phytol*. 232:388–403.
- Yan Z, Ye G, Werren JH. 2019. Evolutionary rate correlation between mitochondrial-encoded and mitochondria-associated nuclear-encoded proteins in insects. *Mol Biol Evol*. 36:1022–1036.
- Zhang WP, Cao L, Lin XR, Ding YM, Liang Y, et al. 2022. Dead-end hybridization in walnut trees revealed by large-scale genomic sequence data. *Mol Biol Evol*. 39:msab308.
- Zhang BW, Lin-Lin X, Li N, Yan PC, Jiang XH, et al. 2019. Phylogenomics reveals an ancient hybrid origin of the Persian walnut. *Mol Biol Evol*. 36:2451–2461.
- Zhang Q, Ree RH, Salamin N, Xing Y, Silvestro D. 2021. Fossil-Informed models reveal a boreotropical origin and divergent evolutionary trajectories in the walnut family (Juglandaceae). *Syst Biol*. 71:242–258.
- Zhang J, Zhang W, Ji F, Qiu J, Song X, et al. 2020. A high-quality walnut genome assembly reveals extensive gene expression divergences after whole-genome duplication. *Plant Biotechnol J*. 18:1848–1850.
- Zheng X, Xiao H, Su J, Chen D, Chen J, et al. 2021. Insights into the evolution and hypoglycemic metabolite biosynthesis of autotetraploid *Cyclocarya paliurus* by combining genomic, transcriptomic and metabolomic analyses. *Ind Crops Prod*. 173:114154.
- Zhu T, Wang L, You FM, Rodriguez JC, Deal KR, et al. 2019. Sequencing a *Juglans regia* x *J. microcarpa* hybrid yields high-quality genome assemblies of parental species. *Hort Res*. 6:55.
- Zupok A, Kozul D, Schottler MA, Niehorster J, Garbsch F, et al. 2021. A photosynthesis operon in the chloroplast genome drives speciation in evening primroses. *Plant Cell*. 33:2583–2601.

Associate editor: Dr. John Archibald