Article

Phylogenomic Inference Suggests Differential Deep Time Phylogenetic Signals from Nuclear and Organellar Genomes in Gymnosperms

Yu-En Lin [1,2,†] iD **, Chung-Shien Wu** [2] iD **, Yu-Wei Wu** [3] iD **and Shu-Miaw Chaw** [2,*,†] iD

1 Department of Biochemical Science and Technology, National Taiwan University, Taipei 106319, Taiwan; wayne0218888@gmail.com
2 Biodiversity Research Center, Academia Sinica, Nankang Campus, Taipei 11529, Taiwan; d92625002@gmail.com
3 Graduate Institute of Medical Bioinformatics, College of Medical Science and Technology, Taipei Medical University, Taipei 11030, Taiwan; yuwei.wu@tmu.edu.tw
* Correspondence: smchaw@sinica.edu.tw
† These authors contributed equally to this work.

Abstract: The living gymnosperms include about 1200 species in five major groups: cycads, ginkgo, gnetophytes, Pinaceae (conifers I), and cupressophytes (conifers II). Molecular phylogenetic studies have yet to reach a unanimously agreed-upon relationship among them. Moreover, cytonuclear phylogenetic incongruence has been repeatedly observed in gymnosperms. We collated a comprehensive dataset from available genomes of 17 gymnosperms across the five major groups and added our own high-quality assembly of a species from Podocarpaceae (the second largest conifer family) to increase sampling width. We used these data to infer reconciled nuclear species phylogenies using two separate methods to ensure the robustness of our conclusions. We also reconstructed organelle phylogenomic trees from 42 mitochondrial and 82 plastid genes from 38 and 289 gymnosperm species across the five major groups, respectively. Our nuclear phylogeny consistently recovers the Ginkgo–cycads clade as the first lineage split from other gymnosperm clades and the Pinaceae as sister to gnetophytes (the Gnepines hypothesis). In contrast, the mitochondrial tree places cycads as the earliest lineage in gymnosperms and gnetophytes as sister to cupressophytes (the Gnecup hypothesis) while the plastomic tree supports the Ginkgo–cycads clade and gnetophytes as the sister to cupressophytes. We also examined the effect of mitochondrial RNA editing sites on the gymnosperm phylogeny by manipulating the nucleotide and amino acid sequences at these sites. Only complete removal of editing sites has an effect on phylogenetic inference, leading to a closer congruence between mitogenomic and nuclear phylogenies. This suggests that RNA editing sites carry a phylogenetic signal with distinct evolutionary traits.

Keywords: RNA editing; cycads; ginkgo; conifers; gnetophytes; phylogenomics; *Nageia*; cytonuclear incongruence; mitogenome; plastome

1. Introduction

Gymnosperms are one of the two clades of extant seed plants that originated in the Middle-Devonian period about 385 million years ago [1,2]. In contrast to angiosperms, which comprise some 350,000 species, gymnosperms are much less diverse and include only 1174 species [2]. Nonetheless, gymnosperms dominate huge swaths of the Earth's landmass, especially the conifers (Pinaceae and Cupressaceae) in the Northern Hemisphere and most

Araucariaceae and Podocarpaceae genera in the Southern Hemisphere [3]. Moreover, gymnosperms exhibit diverse growth forms, predominantly as trees, such as pines (*Pinus*) and firs (*Abies*). Some species, like junipers (*Juniperus*), adopt a shrub-like form. While rare, a few gymnosperms, notably species within the genus *Gnetum*, exhibit vine-like growth forms. These varied growth forms reflect the adaptability of gymnosperms to a diverse range of ecological niches, from canopies to understories and from rainforests to deserts [4–6]. Consequently, gymnosperms are of great economic and ecological importance in several countries. For example, many *Cupressaceae* (cypress family), including false cypress (*Chamaecyparis*), China fir (*Cunninghamia*), bald cypress (*Taxodium*), and arborvitae (*Thuja*), are valuable as both timber sources and ornamentals. The yew family (*Taxaceae*), including about 30 species in six genera, produces taxane compounds useful in anticancer therapies [7].

Advances in sequencing, alignment, and clustering methods starting in the 1990s prompted a reassessment of the classical taxonomic relationships among gymnosperms using molecular data. These analyses categorized gymnosperms into five widely accepted groups: cycads (*Cycadales*), ginkgo (*Ginkgoales*), gnetophytes (*Gnetales*), pine family (*Pinales* or conifers I), and cupressophytes (conifers II) [8,9]. The phylogenetic relationships among these groups are a subject of extensive debate [10], comparable to the controversies surrounding primitive (or early divergent) dicots, monocots, and eudicots. The gymnosperm group monophyly and the placements of ginkgo and gnetophytes were once highly contentious. The initial molecular phylogenetics studies proposed five alternative hypotheses: (1) The extant gymnosperms comprise a monophyletic group; (2) Cycads and ginkgo are sister groups (i.e., the Ginkgo–cycads hypothesis); (3) *Gnetophytes* are sister to all conifers (Gnetifer hypothesis); (4) *Gnetophytes* are sister to *Pinaceae* (the Gnepines hypothesis); and (5) Gnetophytes are sister to cupressophytes (the Gnecup hypothesis) (Figure 1) [8,9,11].
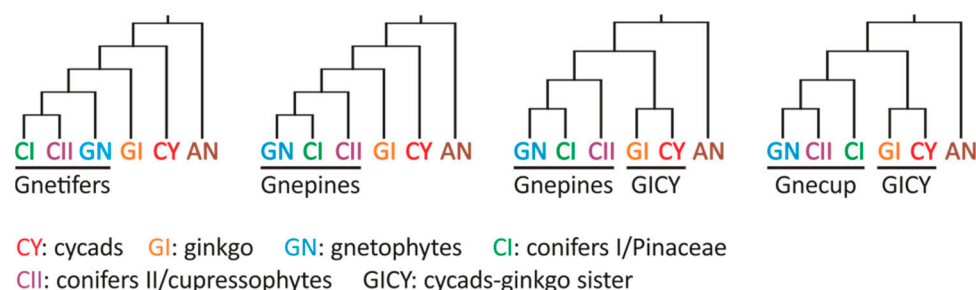

CY: cycads   GI: ginkgo   GN: gnetophytes   CI: conifers I/Pinaceae
CII: conifers II/cupressophytes   GICY: cycads-ginkgo sister

Figure 1. Combinations of the four major hypotheses concerning the phylogenetic relationships among the five extant gymnosperm groups. Angiosperms (AN) were the designated sister group of gymnosperms.

Substantial progress in understanding angiosperm evolution is driven by extensive nuclear phylogenomics studies [12,13]. Although a few recent studies have analyzed the gymnosperm phylogeny using large transcriptomic datasets [14,15], this line of investigation remains relatively underdeveloped. Nuclear genome sequencing is challenging in gymnosperms due to their large genomes (ca. 5–28 Gb [16]), with exceptionally long genes, introns, and many repeated sequences [17–21]. Fortunately, the advancements in sequencing technology over the past five years have made whole-genome assemblies more feasible and cost-effective.

Phylogenomic trees based on whole nuclear genomes are largely inferred from single-copy orthologous genes, excluding the numerous genes including multiple paralogs and thus only accounting for a small, and possibly biased, portion of the phylogenetic signal accumulated in genomes during evolution. Moreover, some phylogenomic trees are constructed from transcriptomes [14,15]. This can lead to several problems: (1) some transcripts

are not in full length, causing sampling bias or alignment errors; (2) transcripts at low levels might lead to inaccurate assembly (or come from DNA contamination [22]); and (3) transcripts represent only a partial set of genes specific to cell types, developmental stages, and growing conditions sampled. To address these issues, we gathered 20 well-assembled seed plant nuclear genomes and incorporated two phylogeny inference strategies, ASTRAL-Pro2 (ASTRAL for Paralogs and Orthologs) and SpeciesRax [23–25], into this study. These recently developed gene tree-based methods can account for gene duplications, losses, transfers, and, to some degree, incomplete lineage sorting by reconciling multiple-copy gene families when building species trees.

In addition to nuclear genomes, organelle markers/genomes have been frequently used to resolve plant phylogenetic relationships at different taxonomic levels [8,9,26]. As the amount of nuclear and organellar data increases, the number of cases of discordance in phylogenies between nuclear and organellar trees, known as cytonuclear incongruence, has likewise grown (see review in Kao et al., 2022) [27]. For instance, plastomes were widely used in green plant phylogenetics for their small size, predominantly uniparental inheritance (mostly paternal in gymnosperms), stable genomic structure, resistance to recombination, and generally low nucleotide substitution rates [28–37]. However, in gymnosperms, plastid–nuclear incongruence has been consistently observed since the earliest study [38]. These discrepancies contributed, among other things, to the continuing controversy over the placement of ginkgo and gnetophytes [9,39–46]. To date, studies on gymnosperm plastome phylogeny have unanimously agreed on the Ginkgo–cycads sisterhood. Nonetheless, the position of gnetophytes is still debated, with earlier studies supporting either the Gnepines or Gnecup hypothesis (see review by Chaw et al., 2018) [30].

In contrast to the extensive studies on gymnosperm phylogeny using plastomes, mitogenomic phylogeny has received significantly less attention. This can be best attributed to the highly variable size, complex structure, genome rearrangements, and intergenomic interactions in plant mitogenomes [47–53]. These complicated genomic features render the assembly and syntenic analysis of mitogenomes difficult. As a result, most previous studies only extracted and combined the 40/41 conserved genes in plant mitogenomes for phylogenetic inference (e.g., Liu et al., 2014) [54]. However, since plant mitogenomes have a slow substitution rate, conserved gene numbers, and rearrangement in some gene groups, for example, *nad5-nad4-nad2* [55–57], it was suggested that plant mitogenomes can harbor ancient phylogenetic signals in bryophytes and other basal land plants [55]. Therefore, in this study, we consider using mitogenomes to help inform gymnosperm phylogeny reconstruction.

Another remarkable feature of gymnosperm mitogenomes is the prevalence of RNA editing sites [58]. RNA editing in plant organelles can correct mutated codons in a transcript to restore conserved amino acids [59–62]. Since mitogenomes evolve slowly, RNA editing sites may contribute to sequence variation, impacting mitophylogeny estimation and potentially causing cytonuclear incongruence. Consequently, many prior studies have tried to evaluate the impact of RNA editing sites on mitogenome phylogenetic inference. Hiesel et al. (1994) favored the use of cDNA rather than DNA to infer phylogeny, as it takes into account the corrected amino acid sequences [63]. Later, Bowe and dePamphilis (1996) found that RNA editing sites did not affect phylogenetic inference and suggested the use of either (but not both) DNA or cDNA for tree building [64]. However, subsequent comparative phylogenetics analyses argued that RNA editing sites carry important phylogenetic signals and advocated for the use of genomic DNA in mitophylogenomics [65–67]. Nevertheless, some studies restore the conserved amino acids or mitigate the effects of RNA editing by changing RNA editing sites from C back to T or even completely removing them [68–75].

In this study, we surveyed reference-level nuclear genomes of 18 gymnosperm species, including 17 published previously (Table 1; as of 31 July 2024), and reported the first high-quality draft genome from *Nageia nagi* in the podocarp family (*Podocarpaceae*, details in Supplementary Information). *Podocarpaceae*, with 19 genera and 187 species, mainly found in the Southern Hemisphere, is the largest and most diverse group in the conifers II clade [76]. Incorporation of the first podocarp genome should improve taxonomic sampling of cupressophytes and fill in the gaps in our understanding of nuclear genomic variation in gymnosperms. We carried out two gene tree-based phylogenetic inference strategies to reconstruct and compare two nuclear phylogenetic trees, and assess the influence of these inference methods on cytonuclear incongruence. Furthermore, we evaluated the impact of RNA editing on mito-nuclear incongruence and consolidated our previous findings regarding the plastid–nuclear incongruent placements of gnetophytes to shed light on the underlying causes of phylogenetic discordance across the five gymnosperm clades.

Table 1. The available and well-annotated gymnosperm nuclear genomes (last accessed: July 2024).

Taxonomic Group	Species	BioProject Accession Number	Assembly Level	Size (Gb)
Conifers I	*Larix kaempferi*	PRJNA587041 [N]	Scaffold	10.9
	Picea abies	PRJEB1822 [N]	Scaffold	12
	Picea sitchensis	PRJNA304257 [N]	Scaffold	20.5
	Pinus albicaulis	PRJNA1034085 [N]	Scaffold	27.6
	Pinus lambertiana	PRJNA174450 [N]	Scaffold	27.6
	Pinus taeda	PRJNA174450 [N]	Scaffold	20.5
	Pseudotsuga menziesii	PRJNA174450 [N]	Scaffold	14.7
Conifers II	*Cryptomeria japonica*	PRJDB13806 [N]	Chromosome	9
	Metasequoia glyptostroboides	PRJCA016596 [C]	Chromosome	8.1
	Nageia nagi *	PRJNA1179671 [N]	Chromosome	4.3
	Sequoia sempervirens	PRJNA542879 [N]	Scaffold	26.5
	Sequoiadendron giganteum	PRJNA541481 [N]	Chromosome	8.1
	Taxus chinensis	PRJNA730337 [N]	Chromosome	10.2
	Torreya grandis	PRJNA938254 [N]	Scaffold	19.1
Cycads	*Cycas panzhihuaensis*	PRJNA734434 [N]	Chromosome	10.5
Ginkgo	*Ginkgo biloba*	PRJCA001755 [C]	Chromosome	9.8
Gnetophytes	*Gnetum montanum*	PRJNA339497 [N]	Scaffold	2.1
	Welwitschia mirabilis	PRJCA004995 [C]	Chromosome	6.8

* First time sequenced and reported in this project. [N]: NCBI, [C]: CNCB.

2. Results

We used nuclear, mitochondrial, and plastid genomic datasets to separately infer species trees. To explore the effect of mitochondrial RNA editing sites on phylogenetic inference, we also gathered a smaller dataset that includes 13 gymnosperm mitogenomes (ME datasets) where C-to-U editing sites in protein-coding genes were verified using transcriptomics (Table S4). Two basal angiosperms, *Amborella* and *Nymphaea*, were used as the outgroups when inferring phylogenetic trees from each dataset.

2.1. Multiple-Copy Nuclear Gene Families Support the Ginkgo–Cycads Sister Relationship and the Gnepines Hypothesis

We were unable to use single-copy orthologs to reconstruct nuclear phylogenetic trees because their number drops drastically with increasing taxon count (Figure S4). We therefore used 10,567 (out of more than 40,000) multi-copy gene families that are common

to the 18 sampled gymnosperm species (Table 1). We used SpeciesRax and ASTRAL-Pro2 methods to reconstruct the species phylogeny, obtaining identical results with both (Figure 2).
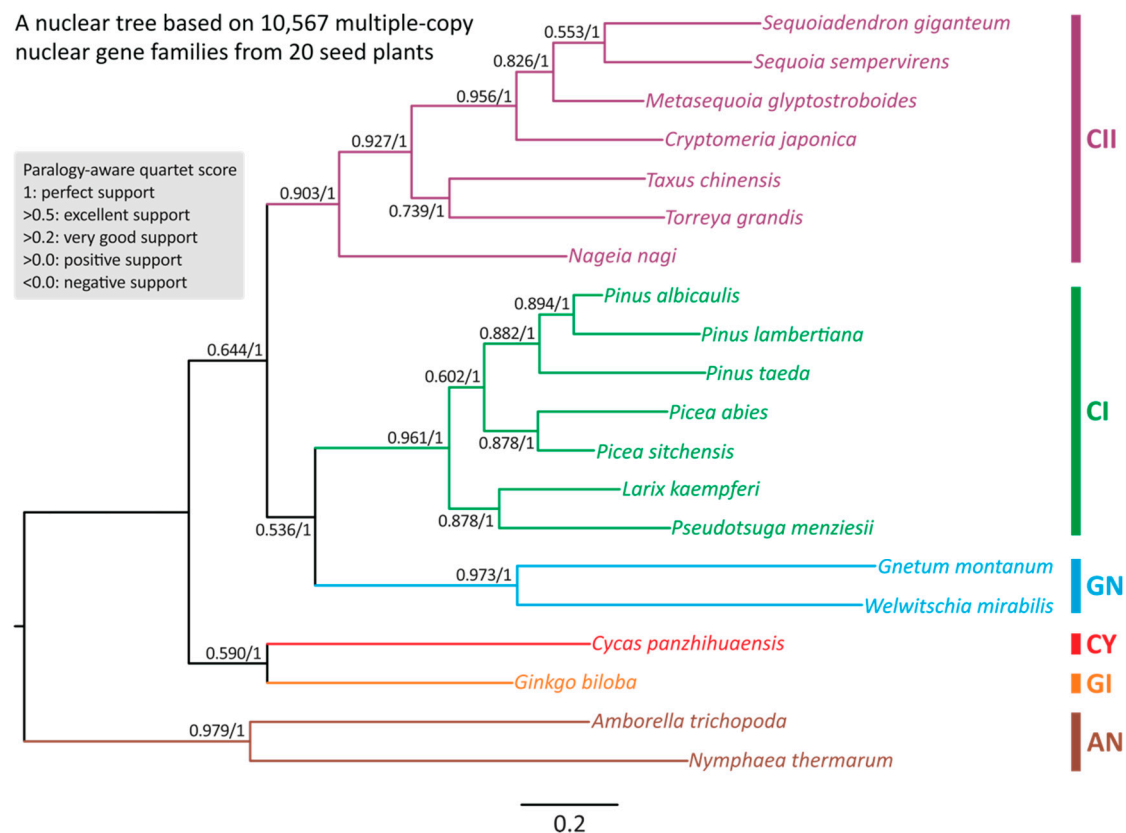

Figure 2. Species trees inferred from 10,567 multiple-copy nuclear gene families across 18 gymnosperms and two angiosperms. The tree framework shown here was constructed using SpeciesRax. Values along branches denote EQPIC scores (before the slashes) and bootstrap support (after the slashes) estimated using SpeciesRax and ASTRAL-Pro2, respectively. AN: angiosperms; GI: ginkgo; CY: cycads; GN: gnetophytes; CI: conifers I; CII: conifers II.

As expected, *Amborella* and *Nymphaea* together form an outgroup clade. At the level of taxonomical orders, our nuclear tree includes two sister clades: the so-called "Ginkgo–cycads sister" and "Gnepines" hypotheses in the gymnosperm phylogeny (Figure 2). SpeciesRax estimates robust extended quadripartition internode certainty (EQP-IC) scores at all nodes, suggesting that the quartets of gene trees also support the two sister clades within the species tree. The strongly supported nodes lead to (1) the Ginkgo–cycads clade (0.59); (2) the three sub-clades: gnetophytes (EQP-IC = 0.973), conifers I (0.961), and conifers II (0.903); and (3) the gnetophytes–conifers I (the Gnepines subclade) (0.536) clade. Notably, clades (1) and (3) also received 100% bootstrap support in ASTRAL-Pro2 (Figure 2). We thus infer that during gymnosperm evolution (1) the Ginkgo–cycads clade was the first to diverge from the other gymnosperm groups and that (2) *Pinaceae* (or conifers I clade) is sister to the gnetophytes rather than the conifers II clade. Therefore, the group commonly known as conifers is not monophyletic.

2.2. Discordant Mitochondrial and Plastid Phylogenomic Trees

The ML tree inferred from both the Mito and Plastid datasets recovered cycads, gnetophytes, conifers I, and conifers II each as a monophyletic clade with strong support (all BS > 99%; Figures S5 and S6). In addition, the mitochondrial tree topology suggests

that (1) cycads diverged first, followed by ginkgo, and that (2) the "Gnepines" topology is well resolved (BS = 90%, Figure S5). In stark contrast, our plastid tree resolves the "Gnecup" topology and the "Ginkgo–cycads" clade with 100% and 98% bootstrap values, respectively (Figure S6). Our results demonstrate that gymnosperm mitochondrial and plastid phylogenies are not only discordant with each other but also differ from the nuclear phylogenomic tree.

2.3. Mitochondrial RNA Editing Sites Influence on Tree Topology

To test if RNA editing sites affect the phylogenetic tree topology, we inferred mitogenomic trees with or without these sites. We needed mitochondrial transcriptome data to find these editing sites. Therefore, the ME datasets only include 13 gymnosperm species where such data are available. Despite the reduction in taxon count, this set still includes representative species from each of the five extant gymnosperm groups. The RNA editing site numbers vary among genomes, ranging from 99 (*Welwitschia*) to 1299 (*Keteleeria*) (Figure S7). RNA editing changes some C nucleotides to U in seed plants. It functions as a repair system for maintaining normal mitochondrial protein function, masking some genomic mutations. The sites involved are thus likely under relaxed constraint and might provide a misleading phylogenetic signal. Replacing these editing sites with missing data ("N") or thymine bases ("T") that reflect their translated sequence would then change the phylogenetic tree topology. However, we do not observe that trees inferred from ME datasets with C-to-N and C-to-T replacements yield identical topologies, both supporting the "Gnecup" clade, and are the same as the full unaltered dataset phylogeny (Figure 3A–C). However, excluding all RNA editing sites resolved the "Gnepines" clade with decreased support (BS = 62%) (Figure 3D). Nevertheless, the placement of cycads remains unchanged.

Because most RNA editing leads to changes in protein sequences, we re-inferred phylogenetic trees using amino acid sequences. We modified the original dataset, in line with the manipulations of the DNA sequences, as follows: (1) amino acids were replaced with "?" if their codons harbored editing sites; (2) all editing sites were recoded as "T" before translation, generating amino acids sequences that are produced in vivo; and (3) positions in the alignments were completely excluded if they contained amino acids affected by either synonymous or nonsynonymous editing in any sampled taxon. Phylogenies estimated from the original and the first two modified datasets are the same (Figure 4A–C). The only difference is that the inference based on the dataset without amino acids affected by RNA editing isolates "Gnepines" into a monophyletic clade with strong support (BS = 94%, Figure 4D). In contrast, the placement of ginkgo is insensitive to the modifications of sampled taxa numbers and RNA editing sites. All mitochondrial trees, inferred from either nucleotide or amino acid sequences, strongly indicate that ginkgo diverged after cycads and is sister to the clade comprising gnetophytes, conifers I, and conifers II (Figures 3, 4 and S5; with all BS = 100%).

In summary, we analyzed three comprehensive datasets, a nuclear and two organelle genomes, of gymnosperms separately to examine the causes of cytonuclear incongruence. Using two separate methods, SpeciesRax and ASTRAL-Pro2, we obtained congruent gymnosperm nuclear phylogenomic trees. ML trees from gymnosperm mitogenomes and plastomes recovered some of the same clades but were not completely in agreement with the nuclear genome-based phylogeny. With ME datasets, we also saw that eliminating RNA editing sites from consideration restored "Gen–pines" monophyly. Keeping these sites, regardless of the treatment of their nucleotide or amino acid states, makes this clade paraphyletic.
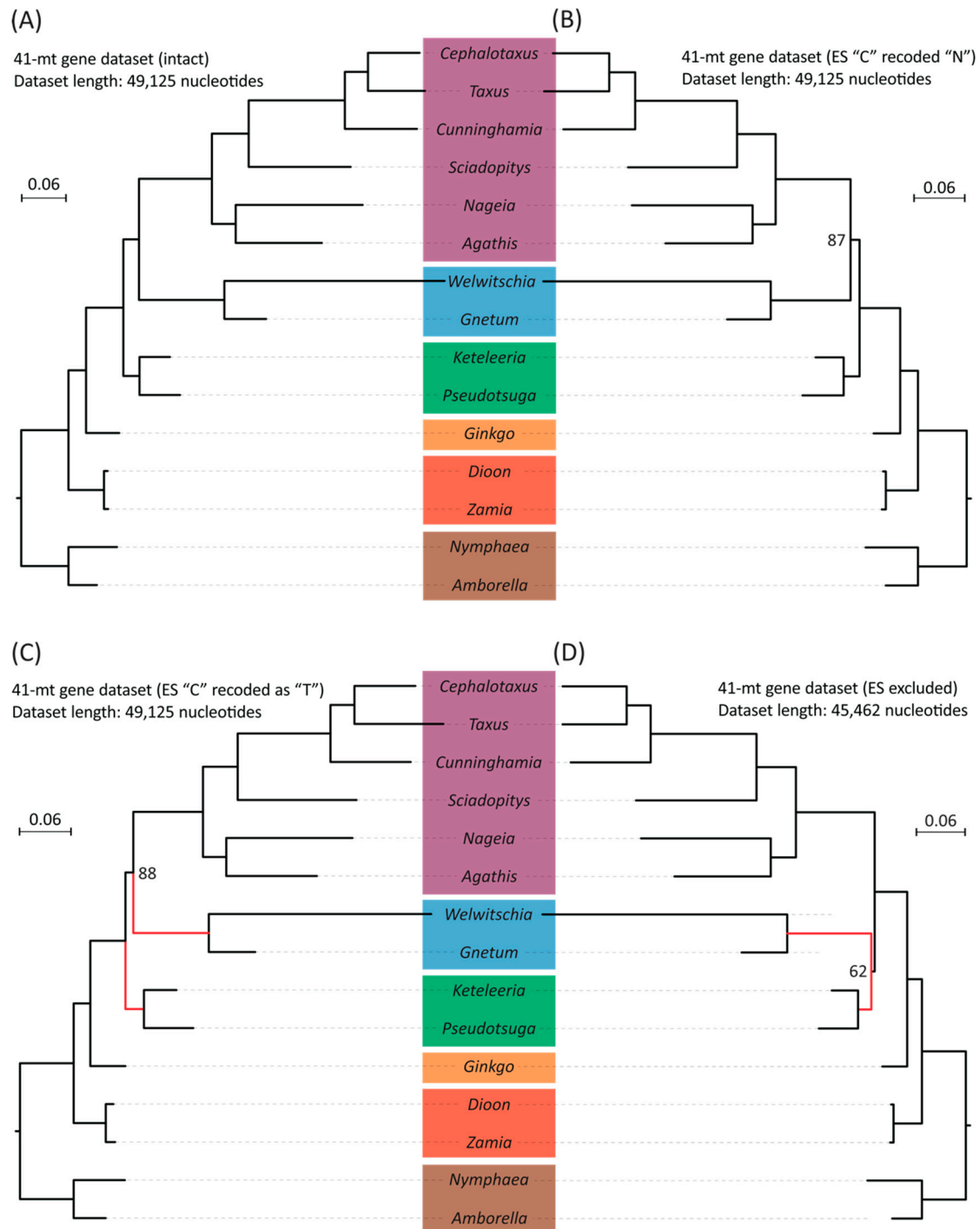
Figure 3. Comparisons of species trees based on four datasets generated from concatenated nucleotide alignments of 41 mt protein-coding genes with and without modifications: (**A**) A maximum likelihood (ML) tree inferred from the dataset without modification. (**B**) An ML tree inferred from the dataset where editing sites were replaced with "N" (=missing data). (**C**) An ML tree inferred from the dataset where edited C nucleotides were recoded as "T" (as in the RNA transcripts). (**D**) An ML tree was inferred from the dataset where alignment positions were removed if they contained an edited nucleotide in any taxon. Bootstrap values are indicated if they are smaller than 100%. The branch length scale bar represents 0.06 substitutions per site. Red lines highlight the placement changes of gnetophytes (here represented by *Welwitschia* and *Gnetum*) between (**C**) and (**D**) trees.
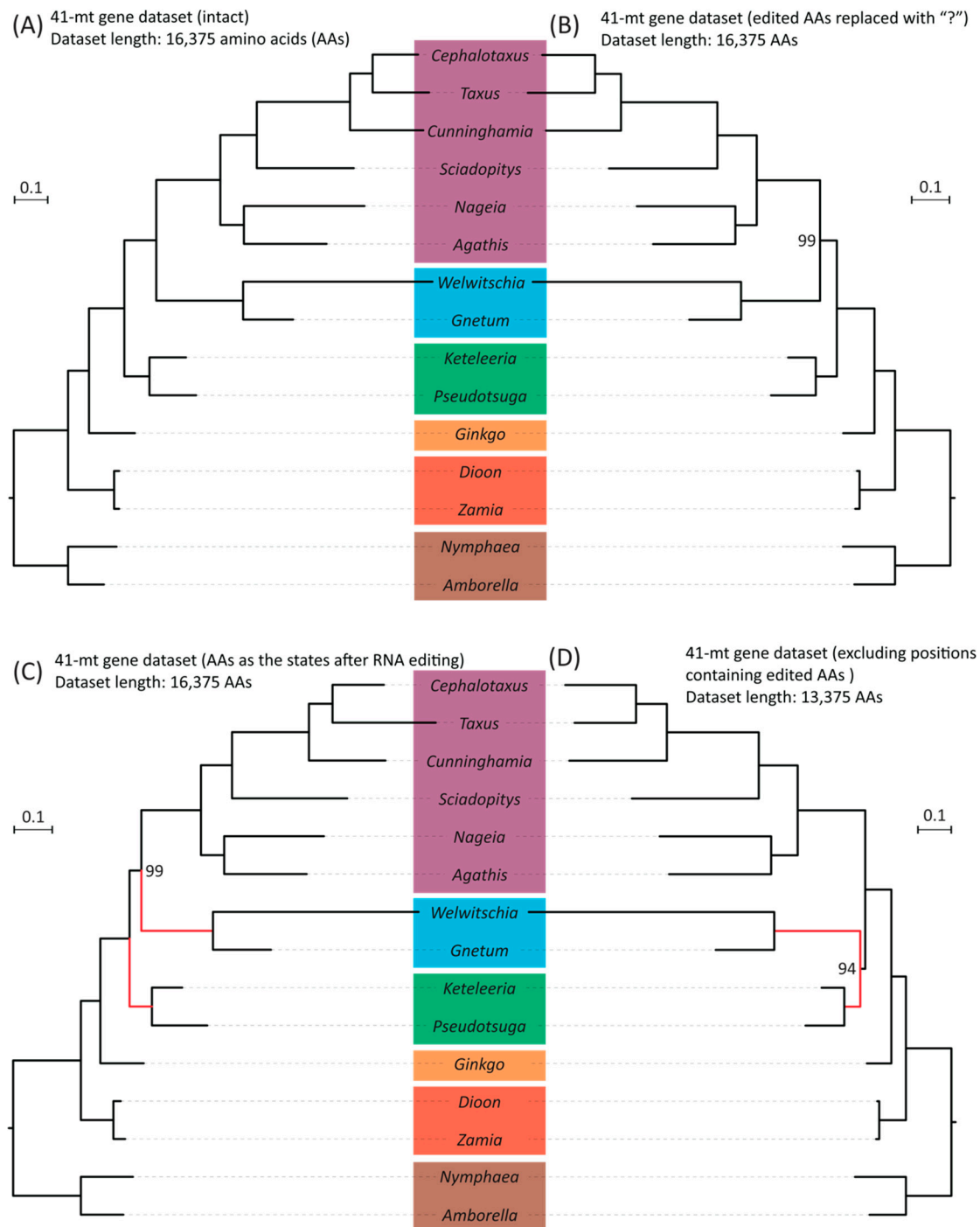
Figure 4. Comparisons of species trees based on four datasets generated from a concatenated amino acid alignment of 41 mt genes with and without modifications: (**A**) A ML tree inferred from the dataset without modification. (**B**) An ML tree inferred from the dataset where amino acids affected by RNA editing (including both synonymous and non-synonymous editing) were replaced with "?". (**C**) An ML tree inferred from the dataset with amino acids recoded according to the state after RNA editing. (**D**) An ML tree inferred from the dataset with the aligned positions containing amino acids affected by editing at either synonymous or non-synonymous sites completely removed. Bootstrap values are indicated if they are smaller than 100%. The branch length scale bar represents 0.1 substitutions per site. Red lines highlight the placement changes of gnetophytes (here represented by *Welwitschia* and *Gnetum*) between (**C**) and (**D**) trees.

3. Discussion

Plant nuclear genomes generally mutate at higher rates than the two organelle (or cytoplasmic) genomes [56]. Additionally, while nuclear genomes undergo sexual reproduction and recombination, organelle genomes are generally uniparentally inherited without sexual recombination and thus less genetically variable. The absence of recombination can lead to a build-up of harmful mutations and eventually the meltdown of cytoplasmic genomes, a phenomenon known as Muller's ratchet [77–79], but this process can be slowed down by the low mutation rate in organellar genomes [56].

Interestingly, some plant species (e.g., *Pelargonium*, *Plantago*, and *Silene*) exhibit extraordinarily accelerated mutation rates in their organellar genomes. In these taxa, biparental inheritance of plastids can occur under mild environmental stress, challenging the long-held belief that organelles are strictly asexual [80]. These discoveries imply that cyto-nuclear phylogenomic incongruence might be common in plants, as nuclear and organellar genomes follow fundamentally different inheritance patterns. In addition, nuclear genes regulate the function and division of organellar genomes.

Earlier studies attributed the cause of cytonuclear incongruence to a variety of processes including long-branch attraction (LBA), incomplete lineage sorting, introgression, gene duplication and loss, distinct organelle inheritance modes, or analytical factors such as sample size and taxon sampling strategies [27,67,81–94]. In this study, we further scrutinize the causes of cytonuclear incongruence in gymnosperms. To do so, we first constructed a nuclear phylogenomic tree with minimized noise from incomplete lineage sorting, gene duplication and loss, and insufficient sample size. Our objective was to reduce systematic errors in the nuclear phylogenetic inference so that we can justify whether the observed incongruence originates from intrinsic evolutionary processes of organellar genomes or from methodological artifacts.

Using the first comprehensive nuclear genome dataset to include *Nageia* (Podocarpaceae), we applied two gene tree-based phylogenetic inference methods: ASTRAL-Pro2 and SpeciesRax. Traditional concatenation methods combine per-gene alignments into a single supermatrix to infer a species tree. Since the concatenation method only works well with accurate orthology inference [95], it often fails when the evolution of genes deviates from that of species due to incomplete lineage sorting or events like duplications, loss, and transfers [96]. On the other hand, gene family tree methods, like ASTRAL-Pro2 and SpeciesRax, preserve the distinct evolutionary history within individual gene families and thus can explicitly account for the underlying processes causing discordance. Specifically, ASTRAL-Pro2 operates under the multispecies coalescent framework, which uses the statistical distribution of gene trees (summarized as quartets) to reduce the influence of incomplete lineage sorting. This method is particularly powerful when multiple-copy nuclear gene families are involved, as it leverages the coalescent process to "average out" the discordance introduced by incomplete lineage sorting. In contrast, SpeciesRax employs a maximum likelihood approach that explicitly models gene duplication, loss, and horizontal transfer events. By incorporating information from both orthologs and paralogs, SpeciesRax is designed to address the complexities of gene family evolution. However, it does not explicitly mitigate the effects of incomplete lineage sorting as the coalescent-based methods do. Together, these two approaches provide a robust alternative to concatenation-based methods by directly modeling the heterogeneous evolutionary processes that shape gene trees [23–25,97].

Using both methods, we obtained identical topologies of the rooted trees on nuclear data. Our results agree with previous studies that include inferences of rooted as well as unrooted trees [9,14,15]. Therefore, nuclear phylogenomics of living gymnosperms has reached a consensus that places ginkgo and cycads as the earliest-diverging lineage, fol-

lowed by two sister groups: Gnepines and cupressophytes (Figure 2). Furthermore, mounting molecular phylogenetic evidence over the past two decades also supports Pinaceae and gnetophyte monophyly. We thus propose replacing the term "conifers" with Conifers I (Pinaceae) and Conifers II (cupressophytes) in future research for clarity, since conifers are consistently paraphyletic.

Our analysis of the gymnosperm plastome phylogeny supports the Ginkgo–cycads topology and the Gnecup topology, in contrast to the nuclear-derived Gnepines topology (Table 2, Figures 2 and S6). Numerous factors have been proposed to interpret such incongruence, including incomplete lineage sorting [98], long-branch attraction (LBA) [41,43,99,100], and chloroplast capture [27,101–105], to name a few. We previously reported significantly accelerated nucleotide substitution rates in gnetophytes, potentially leading to long-branch attraction artifacts in phylogenetic reconstruction [41,43]. Multiple methods that alleviate this artifact restore the Gnepines topology in plastome trees, suggesting that long-branch attraction plays at least a role in mis-inference [41–43,106]. However, additional analyses would be needed to eliminate or confirm other potential sources of phylogenetic inconsistencies between analyses based on nuclear and plastid data.

Table 2. Comparisons of phylogenomic trees inferred from different datasets on controversial clades among the five gymnosperm groups.

Genome Dataset	Ginkgo–Cycads	Gnepines	Gnecup
Nuclear	Yes	Yes	No
Mito	No	Yes	No
Plastid	Yes	No	Yes
ME-genomic	No	No	Yes
ME-protein	No	No	Yes
ME-ES * corrected	No	No	Yes
ME-ES * excluded	No	Yes	No

* ES: RNA editing sites.

Phylogenies reconstructed from mitochondrial datasets differ from both nuclear- and plastid-based trees. In particular, the mitogenomic phylogeny includes cycads as sister to the remaining gymnosperms (Table 2, Figures 2, S5 and S6). Given that RNA editing sites have been proposed as a potential source of phylogenomic incongruence, we further investigated whether modifications of the RNA editing sites at either the genomic or protein level could address the observed mito-nuclear discrepancies. Nevertheless, using transcriptomic-verified ME datasets, our mitogenome species trees show no topological difference before and after the substitution of RNA editing sites (Figure 4A–C), which indicates that the substitution alone is not enough to eliminate the effect of RNA editing, as suggested earlier by other authors [74,75]. This observation is consistent with the notion that the abundance of RNA editing sites correlates with the irregular substitution rates and pronounced nucleotide compositional biases, such as an overrepresentation of pyrimidines, even after substitution [58,62,67,74,75]. In addition, RNA editing frequently produces homoplastic changes, as similar edited states may arise independently in different lineages, thereby introducing convergent noise that confounds the true phylogenetic signal [64]. Moreover, different taxa may exhibit varying efficiencies or patterns of RNA editing. As a result, a uniform substitution does not capture the full extent of the underlying evolutionary variability (Figure S7) [60]. Consequently, only the complete removal of RNA editing sites can mitigate the influence and restore the Gnepines relationship that resembles the nuclear phylogenomic tree topology at both the genomic and protein level (Figures 3D and 4D). Furthermore, there is the possibility that the slow substitution rate of mitogenomes allows them to carry deep historical signals from ancient events, such as hybridization [70,71,85].

This might explain why excluding these sites did not alter the sister relationship of cycads to the remaining gymnosperms (Figures 3D and 4D) [75,107].

In summary, we suggest that the intrinsic properties of RNA editing sites, like their rapid, biased evolution and convergent behavior, contribute to the observed mito-nuclear incongruence, and that the complete exclusion of them is necessary to mitigate misleading effects. Nonetheless, RNA editing sites in gymnosperms are too few to bias the inference from unedited sites that preserve mitogenomes' deep and strong phylogenetic signals from ancient hybridization events. We thus see a general preservation of tree topology even after eliminating editing sites, other than the restoration in the placement of gnetophytes [69]. We also note that the "Gnepines" relationship recovered from the full mitochondrial dataset (Figure S5) was not resolved in the ME datasets that contain only 13 gymnosperm species (Figure 4A). This suggests that a reduction in taxon sampling can also affect mitochondrial tree topology.

4. Materials and Methods

4.1. Data Access

Nuclear genomes and their annotations were downloaded from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/ (accessed on 31 July 2024)), the China National Center for Bioinformation (CNCB, https://ngdc.cncb.ac.cn), and the TreeGenes (https://treegenesdb.org/) databases. Incorporating our newly sequenced *Nageia* draft genome (see Supplementary Information, Figures S1–S4, Tables S1 and S2 [108–117]), we gathered 18 gymnosperm and two angiosperm genomes that were well-annotated and assembled to scaffold/chromosomal levels (Nu dataset; Table 1). We used AGAT 1.4.0 (https://github.com/NBISweden/AGAT (accessed on 31 July 2024)) to retrieve amino acid sequences of annotated genes from these 20 nuclear genomes. We also gathered amino acid sequences from publicly available mitochondrial and plastid genomes/scaffolds, comprising the Mito and Plastid datasets with the former representing 38 (Table S3) and the latter 289 (Table S4) gymnosperm species. Both datasets also include two angiosperms as the outgroups.

4.2. Classification of Nuclear Multiple-Copy Gene Families and Construction of Gene and Species Trees

OrthoFinder v3.0.1b1 was used to identify orthologous genes in the 20 sampled nuclear genomes [118]. Only one single-copy and 43,333 multiple-copy nuclear gene families were obtained. To save on computing time, we discarded gene families that are shared by fewer than 10 taxa (i.e., fewer than half of the 20 sampled taxa). This resulted in 10,567 gene families retained for downstream analyses. After using MUSCLE 3.8.1551 to align the sequences [119], the resulting alignments were used to infer gene trees with ParGenes v1.1.2 and its "-m" option for automatically determining the best-fit model [120]. The generated gene trees were pooled before reconciling into species trees using Astral-Pro 2 v1.15.2.4 or SpeciesRax v2.1.3 [23–25].

4.3. Construction of Mitochondrial and Plastid Trees

The amino acid sequences of mitochondrial and plastid genes were aligned and concatenated into supermetrics using Geneious Prime (www.geneious.com). To better handle sequence heterogeneity in the supermetrics, each gene was designated as a partition to determine the best-fit model. We employed IQtree 2.3.6 for constructing mitochondrial and plastid trees with the "MFP+MERGE" option and 1000 nonparametric bootstrap replicates [121].

5. Conclusions

We present the most comprehensive dataset and analysis of gymnosperm cytonuclear incongruence to date. Our results encompass nuclear phylogenomic analyses incorporating, for the first time, a draft Podocarpaceae (from the Southern Hemisphere Conifers II clade) genome. We also construct mitogenome and plastome trees to investigate the factors underlying the observed nuclear–organellar discrepancies. Our extensive analyses suggest that gymnosperm cytonuclear incongruence is likely due to several factors: (1) distinct organelle genome inheritance modes that increase the influence from ancient phylogenetic signals; (2) insufficient taxon sampling; and (3) inference bias that stems from genomic processes producing misleading evolutionary signals, like RNA editing sites, incomplete lineage sorting, and long-branch attraction. Future studies should carefully account for these factors when interpreting discrepancies among inferences based on disparate data sources.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/plants14091335/s1, Figure S1: Flow cytometry of DNA content in *N. nagi*; Figure S2: Accumulations of scaffolds from the longest to the shortest; Figure S3: Dot-plot analysis of the 13 longest scaffolds; Figure S4: Variation in the number of identified single-copy gene families with different strategies of taxon sampling; Figure S5: A maximum likelihood (ML) tree inferred from concatenated amino acid alignments of 42 mitochondrial genes across 38 gymnosperms; Figure S6: A maximum likelihood (ML) tree inferred from concatenated amino acid alignments of 82 plastid genes across 289 gymnosperms; Figure S7: Variation in the total number of RNA editing sites across the 13 sampled gymnosperm mitogenomes; Table S1: Assembly statistics before and after HiRise scaffolding; Table S2: BUSCO statistics; Table S3: Gymnosperm plastomes used in this study; Table S4. Gymnosperm mitogenomes/scaffolds used in this study.

Author Contributions: S.-M.C. conceived and initiated the study. Y.-E.L. and S.-M.C. sorted, discussed, wrote, and revised the manuscript. Y.-E.L. prepared Supplementary Information and gathered references. C.-S.W. carried out phylogenetic tree analyses, made figures, and gave critical comments. Y.-W.W. performed the assembly and annotation of the *Nageia* genome and provided technical support on genomic analyses. S.-M.C. supervised the experiments and gathered funding. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All of the raw sequence reads used in this study have been deposited in NCBI BioProject under the accession PRJNA1179671.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

NCBI	National Center for Biotechnology Information
CNCB	China National Center for Bioinformation

References

1. Gerrienne, P.; Meyer-Berthaud, B.; Fairon-Demaret, M.; Streel, M.; Steemans, P. Runcaria, a Middle Devonian seed plant precursor. *Science* **2004**, *306*, 856–858. [CrossRef] [PubMed]
2. Yang, Y.; Ferguson, D.K.; Liu, B.; Mao, K.-S.; Gao, L.-M.; Zhang, S.-Z.; Wan, T.; Rushforth, K.; Zhang, Z.-X. Recent advances on phylogenomics of gymnosperms and a new classification. *Plant Divers.* **2022**, *44*, 340–350. [CrossRef] [PubMed]
3. Williams, C. *Conifer Reproductive Biology*; Springer: New York, NY, USA, 2009; Volume 169.
4. Hou, C.; Humphreys, A.M.; Thureborn, O.; Rydin, C. New insights into the evolutionary history of Gnetum (Gnetales). *Taxon* **2015**, *64*, 239–253. [CrossRef]
5. Contreras-Medina, R.; Vega, I.L. On the distribution of gymnosperm genera, their areas of endemism and cladistic biogeography. *Aust. Syst. Bot.* **2002**, *15*, 193–203. [CrossRef]
6. Su, Z.-H.; Zhang, M.-L. Evolutionary history of a desert shrub *Ephedra przewalskii* (Ephedraceae): Allopatric divergence and range shifts in northwestern China. *PLoS ONE* **2016**, *11*, e0158284. [CrossRef]
7. Škubník, J.; Pavlíčková, V.; Ruml, T.; Rimpelová, S. Current perspectives on taxanes: Focus on their bioactivity, delivery and combination therapy. *Plants* **2021**, *10*, 569. [CrossRef]
8. Bowe, L.M.; Coat, G.; DePamphilis, C.W. Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4092–4097. [CrossRef]
9. Chaw, S.-M.; Parkinson, C.L.; Cheng, Y.; Vincent, T.M.; Palmer, J.D. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4086–4091. [CrossRef]
10. Yang, Y.; Yang, Z.; Ferguson, D.K. The Systematics and Evolution of Gymnosperms with an Emphasis on a Few Problematic Taxa. *Plants* **2024**, *13*, 2196. [CrossRef]
11. Nickrent, D.L.; Parkinson, C.L.; Palmer, J.D.; Duff, R.J. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol. Biol. Evol.* **2000**, *17*, 1885–1895. [CrossRef]
12. Zuntini, A.R.; Carruthers, T.; Maurin, O.; Bailey, P.C.; Leempoel, K.; Brewer, G.E.; Epitawalage, N.; Françoso, E.; Gallego-Paramo, B.; McGinnie, C. Phylogenomics and the rise of the angiosperms. *Nature* **2024**, *629*, 843–850. [CrossRef] [PubMed]
13. Zhang, G.; Ma, H. Nuclear phylogenomics of angiosperms and insights into their relationships and evolution. *J. Integr. Plant Biol.* **2024**, *66*, 546–578. [CrossRef] [PubMed]
14. Ran, J.-H.; Shen, T.-T.; Wang, M.-M.; Wang, X.-Q. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Proc. R. Soc. B* **2018**, *285*, 20181012. [CrossRef]
15. Stull, G.W.; Qu, X.-J.; Parins-Fukuchi, C.; Yang, Y.-Y.; Yang, J.-B.; Yang, Z.-Y.; Hu, Y.; Ma, H.; Soltis, P.S.; Soltis, D.E. Gene duplications and phylogenomic conflict underlie major pulses of phenotypic evolution in gymnosperms. *Nat. Plants* **2021**, *7*, 1015–1025. [CrossRef]
16. Murray, B.G. Nuclear DNA amounts in gymnosperms. *Ann. Bot.* **1998**, *82*, 3–15. [CrossRef]
17. Ahuja, M.R.; Neale, D.B. Evolution of genome size in conifers. *Silvae Genet.* **2005**, *54*, 126–137. [CrossRef]
18. Morse, A.M.; Peterson, D.G.; Islam-Faridi, M.N.; Smith, K.E.; Magbanua, Z.; Garcia, S.A.; Kubisiak, T.L.; Amerson, H.V.; Carlson, J.E.; Nelson, C.D. Evolution of genome size and complexity in Pinus. *PLoS ONE* **2009**, *4*, e4332. [CrossRef]
19. Liu, H.; Wang, X.; Wang, G.; Cui, P.; Wu, S.; Ai, C.; Hu, N.; Li, A.; He, B.; Shao, X. The nearly complete genome of Ginkgo biloba illuminates gymnosperm evolution. *Nat. Plants* **2021**, *7*, 748–756. [CrossRef]
20. Wan, T.; Gong, Y.; Liu, Z.; Zhou, Y.; Dai, C.; Wang, Q. Evolution of complex genome architecture in gymnosperms. *GigaScience* **2022**, *11*, giac078. [CrossRef]
21. Zhu, P.; He, T.; Zheng, Y.; Chen, L. The need for masked genomes in gymnosperms. *Front. Plant Sci.* **2023**, *14*, 1309744. [CrossRef]
22. Li, X.; Zhang, P.; Wang, H.; Yu, Y. Genes expressed at low levels raise false discovery rates in RNA samples contaminated with genomic DNA. *BMC Genom.* **2022**, *23*, 554. [CrossRef] [PubMed]
23. Zhang, C.; Scornavacca, C.; Molloy, E.K.; Mirarab, S. ASTRAL-Pro: Quartet-based species-tree inference despite paralogy. *Mol. Biol. Evol.* **2020**, *37*, 3292–3307. [CrossRef] [PubMed]
24. Zhang, C.; Mirarab, S. ASTRAL-Pro 2: Ultrafast species tree reconstruction from multi-copy gene family trees. *Bioinformatics* **2022**, *38*, 4949–4950. [CrossRef]
25. Morel, B.; Schade, P.; Lutteropp, S.; Williams, T.A.; Szöllősi, G.J.; Stamatakis, A. SpeciesRax: A tool for maximum likelihood species tree inference from gene family trees under duplication, transfer, and loss. *Mol. Biol. Evol.* **2022**, *39*, msab365. [CrossRef]
26. Lockwood, J.D.; Aleksić, J.M.; Zou, J.; Wang, J.; Liu, J.; Renner, S.S. A new phylogeny for the genus Picea from plastid, mitochondrial, and nuclear sequences. *Mol. Phylogenet. Evol.* **2013**, *69*, 717–727. [CrossRef]
27. Kao, T.T.; Wang, T.H.; Ku, C. Rampant nuclear–mitochondrial–plastid phylogenomic discordance in globally distributed calcifying microalgae. *New Phytol.* **2022**, *235*, 1394–1408. [CrossRef]

28. Smith, D.R. Mutation rates in plastid genomes: They are lower than you might think. *Genome Biol. Evol.* **2015**, *7*, 1227–1234. [CrossRef]

29. Xiao-Ming, Z.; Junrui, W.; Li, F.; Sha, L.; Hongbo, P.; Lan, Q.; Jing, L.; Yan, S.; Weihua, Q.; Lifang, Z. Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci. Rep.* **2017**, *7*, 1555. [CrossRef]

30. Chaw, S.-M.; Wu, C.-S.; Sudianto, E. Evolution of gymnosperm plastid genomes. In *Advances in Botanical Research*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 85, pp. 195–222.

31. Li, H.-T.; Luo, Y.; Gan, L.; Ma, P.-F.; Gao, L.-M.; Yang, J.-B.; Cai, J.; Gitzendanner, M.A.; Fritsch, P.W.; Zhang, T. Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol.* **2021**, *19*, 232. [CrossRef]

32. Lubna; Asaf, S.; Khan, A.L.; Jan, R.; Khan, A.; Khan, A.; Kim, K.M.; Lee, I.J. The dynamic history of gymnosperm plastomes: Insights from structural characterization, comparative analysis, phylogenomics, and time divergence. *Plant Genome* **2021**, *14*, e20130. [CrossRef]

33. Yang, X.; Zhou, T.; Su, X.; Wang, G.; Zhang, X.; Guo, Q.; Cao, F. Structural characterization and comparative analysis of the chloroplast genome of Ginkgo biloba and other gymnosperms. *J. For. Res.* **2021**, *32*, 765–778. [CrossRef]

34. Lian, C.; Yang, H.; Lan, J.; Zhang, X.; Zhang, F.; Yang, J.; Chen, S. Comparative analysis of chloroplast genomes reveals phylogenetic relationships and intraspecific variation in the medicinal plant Isodon rubescens. *PLoS ONE* **2022**, *17*, e0266546. [CrossRef] [PubMed]

35. Wang, J.; Kan, S.; Liao, X.; Zhou, J.; Tembrock, L.R.; Daniell, H.; Jin, S.; Wu, Z. Plant organellar genomes: Much done, much more to do. *Trends Plant Sci.* **2024**, *29*, 754–769. [CrossRef] [PubMed]

36. Feng, M.; Kong, H.; Lin, M.; Zhang, R.; Gong, W. The complete plastid genome provides insight into maternal plastid inheritance mode of the living fossil plant Ginkgo biloba. *Plant Divers.* **2023**, *45*, 752. [CrossRef]

37. Shrestha, B.; Gilbert, L.E.; Ruhlman, T.A.; Jansen, R.K. Clade-specific plastid inheritance patterns including frequent biparental inheritance in passiflora interspecific crosses. *Int. J. Mol. Sci.* **2021**, *22*, 2278. [CrossRef]

38. Palmer, J.D. Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* **1985**, *19*, 325–354. [CrossRef]

39. Chaw, S.-M.; Zharkikh, A.; Sung, H.-M.; Lau, T.-C.; Li, W.-H. Molecular phylogeny of extant gymnosperms and seed plant evolution: Analysis of nuclear 18S rRNA sequences. *Mol. Biol. Evol.* **1997**, *14*, 56–68. [CrossRef]

40. Palmer, J.D.; Soltis, D.E.; Chase, M.W. The plant tree of life: An overview and some points of view. *Am. J. Bot.* **2004**, *91*, 1437–1445. [CrossRef]

41. Wu, C.-S.; Wang, Y.-N.; Liu, S.-M.; Chaw, S.-M. Chloroplast genome (cpDNA) of Cycas taitungensis and 56 cp protein-coding genes of Gnetum parvifolium: Insights into cpDNA evolution and phylogeny of extant seed plants. *Mol. Biol. Evol.* **2007**, *24*, 1366–1379. [CrossRef]

42. Zhong, B.; Yonezawa, T.; Zhong, Y.; Hasegawa, M. The position of Gnetales among seed plants: Overcoming pitfalls of chloroplast phylogenomics. *Mol. Biol. Evol.* **2010**, *27*, 2855–2863. [CrossRef]

43. Wu, C.-S.; Lin, C.-P.; Hsu, C.-Y.; Wang, R.-J.; Chaw, S.-M. Comparative chloroplast genomes of Pinaceae: Insights into the mechanism of diversified genomic organizations. *Genome Biol. Evol.* **2011**, *3*, 309–319. [CrossRef] [PubMed]

44. Wu, C.-S.; Chaw, S.-M.; Huang, Y.-Y. Chloroplast phylogenomics indicates that Ginkgo biloba is sister to cycads. *Genome Biol. Evol.* **2013**, *5*, 243–254. [CrossRef] [PubMed]

45. Wang, X.-Q.; Ran, J.-H. Evolution and biogeography of gymnosperms. *Mol. Phylogenet. Evol.* **2014**, *75*, 24–40. [CrossRef]

46. Ruhfel, B.R.; Gitzendanner, M.A.; Soltis, P.S.; Soltis, D.E.; Burleigh, J.G. From algae to angiosperms–inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evol. Biol.* **2014**, *14*, 23. [CrossRef]

47. Palmer, J.D.; Herbon, L.A. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *J. Mol. Evol.* **1988**, *28*, 87–97. [CrossRef]

48. Chaw, S.-M.; Chun-Chieh Shih, A.; Wang, D.; Wu, Y.-W.; Liu, S.-M.; Chou, T.-Y. The mitochondrial genome of the gymnosperm Cycas taitungensis contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Mol. Biol. Evol.* **2008**, *25*, 603–615. [CrossRef]

49. Gualberto, J.M.; Mileshina, D.; Wallet, C.; Niazi, A.K.; Weber-Lotfi, F.; Dietrich, A. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* **2014**, *100*, 107–120. [CrossRef]

50. Jackman, S.D.; Coombe, L.; Warren, R.L.; Kirk, H.; Trinh, E.; MacLeod, T.; Pleasance, S.; Pandoh, P.; Zhao, Y.; Coope, R.J. Complete mitochondrial genome of a gymnosperm, Sitka spruce (*Picea sitchensis*), indicates a complex physical structure. *Genome Biol. Evol.* **2020**, *12*, 1174–1179. [CrossRef]

51. Liu, H.; Zhao, W.; Zhang, R.-G.; Mao, J.-F.; Wang, X.-R. Repetitive elements, sequence turnover and cyto-nuclear gene transfer in Gymnosperm Mitogenomes. *Front. Genet.* **2022**, *13*, 867736. [CrossRef]

52. Wu, Z.Q.; Liao, X.Z.; Zhang, X.N.; Tembrock, L.R.; Broz, A. Genomic architectural variation of plant mitochondria—A review of multichromosomal structuring. *J. Syst. Evol.* **2022**, *60*, 160–168. [CrossRef]

53. Xia, C.; Li, J.; Zuo, Y.; He, P.; Zhang, H.; Zhang, X.; Wang, B.; Zhang, J.; Yu, J.; Deng, H. Complete mitochondrial genome of Thuja sutchuenensis and its implications on evolutionary analysis of complex mitogenome architecture in Cupressaceae. *BMC Plant Biol.* **2023**, *23*, 84. [CrossRef] [PubMed]

54. Liu, Y.; Cox, C.J.; Wang, W.; Goffinet, B. Mitochondrial phylogenomics of early land plants: Mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Syst. Biol.* **2014**, *63*, 862–878. [CrossRef] [PubMed]

55. Groth-Malonek, M.; Knoop, V. Bryophytes and other basal land plants: The mitochondrial perspective. *Taxon* **2005**, *54*, 293–297. [CrossRef]

56. Drouin, G.; Daoud, H.; Xia, J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenetics Evol.* **2008**, *49*, 827–831. [CrossRef]

57. Mower, J.P.; Sloan, D.B.; Alverson, A.J. Plant mitochondrial genome diversity: The genomics revolution. In *Plant Genome Diversity Volume 1: Plant Genomes, Their Residents, and Their Evolutionary Dynamics*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 123–144.

58. Wu, C.S.; Chaw, S.M. Evolution of mitochondrial RNA editing in extant gymnosperms. *Plant J.* **2022**, *111*, 1676–1687. [CrossRef]

59. Knoop, V. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell. Mol. Life Sci.* **2011**, *68*, 567–586. [CrossRef]

60. Sloan, D.B. Nuclear and mitochondrial RNA editing systems have opposite effects on protein diversity. *Biol. Lett.* **2017**, *13*, 20170314. [CrossRef]

61. Edera, A.A.; Gandini, C.L.; Sanchez-Puerta, M.V. Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. *Plant Mol. Biol.* **2018**, *97*, 215–231. [CrossRef]

62. Dong, S.; Zhao, C.; Zhang, S.; Wu, H.; Mu, W.; Wei, T.; Li, N.; Wan, T.; Liu, H.; Cui, J. The amount of RNA editing sites in liverwort organellar genes is correlated with GC content and nuclear PPR protein diversity. *Genome Biol. Evol.* **2019**, *11*, 3233–3239. [CrossRef]

63. Hiesel, R.; von Haeseler, A.; Brennicke, A. Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 634–638. [CrossRef]

64. Bowe, L.M.; DePamphilis, C.W. Effects of RNA editing and gene processing on phylogenetic reconstruction. *Mol. Biol. Evol.* **1996**, *13*, 1159–1166. [CrossRef] [PubMed]

65. Szmidt, A.E.; Lu, M.-Z.; Wang, X.-R. Effects of RNA editing on the coxI evolution and phylogeny reconstruction. *Euphytica* **2001**, *118*, 9–18. [CrossRef]

66. Petersen, G.; Seberg, O.; Davis, J.I.; Stevenson, D.W. RNA editing and phylogenetic reconstruction in two monocot mitochondrial genes. *Taxon* **2006**, *55*, 871–886. [CrossRef]

67. Picardi, E.; Quagliariello, C. Is plant mitochondrial RNA editing a source of phylogenetic incongruence? An answer from in silico and in vivo data sets. *BMC Bioinform.* **2008**, *9*, S14. [CrossRef]

68. Bergthorsson, U.; Adams, K.L.; Thomason, B.; Palmer, J.D. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **2003**, *424*, 197–201. [CrossRef]

69. Seberg, O.; Petersen, G.; Davis, J.I.; Pires, J.C.; Stevenson, D.W.; Chase, M.W.; Fay, M.F.; Devey, D.S.; Jørgensen, T.; Sytsma, K.J. Phylogeny of the Asparagales based on three plastid and two mitochondrial genes. *Am. J. Bot.* **2012**, *99*, 875–889. [CrossRef]

70. Richardson, A.O.; Rice, D.W.; Young, G.J.; Alverson, A.J.; Palmer, J.D. The "fossilized" mitochondrial genome of Liriodendron tulipifera: Ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* **2013**, *11*, 29. [CrossRef]

71. Guo, W.; Grewe, F.; Fan, W.; Young, G.J.; Knoop, V.; Palmer, J.D.; Mower, J.P. Ginkgo and Welwitschia mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol. Biol. Evol.* **2016**, *33*, 1448–1460. [CrossRef]

72. Gitzendanner, M.A.; Soltis, P.S.; Wong, G.K.S.; Ruhfel, B.R.; Soltis, D.E. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *Am. J. Bot.* **2018**, *105*, 291–301. [CrossRef]

73. Bell, D.; Lin, Q.; Gerelle, W.K.; Joya, S.; Chang, Y.; Taylor, Z.N.; Rothfels, C.J.; Larsson, A.; Villarreal, J.C.; Li, F.W. Organellomic data sets confirm a cryptic consensus on (unrooted) land-plant relationships and provide new insights into bryophyte molecular evolution. *Am. J. Bot.* **2020**, *107*, 91–115. [CrossRef]

74. Dong, S.S.; Li, H.L.; Goffinet, B.; Liu, Y. Exploring the impact of RNA editing on mitochondrial phylogenetic analyses in liverworts, an early land plant lineage. *J. Syst. Evol.* **2022**, *60*, 16–22. [CrossRef]

75. Dong, S.-S.; Zhou, X.-P.; Peng, T.; Liu, Y. Mitochondrial RNA editing sites affect the phylogenetic reconstruction of gymnosperms. *Plant Divers.* **2023**, *45*, 485. [CrossRef] [PubMed]

76. Christenhusz, M.J.; Byng, J.W. The number of known plants species in the world and its annual increase. *Phytotaxa* **2016**, *261*, 201–217. [CrossRef]

77. Muller, H.J. The relation of recombination to mutational advance. *Mutat. Res./Fundam. Mol. Mech. Mutagen.* **1964**, *1*, 2–9. [CrossRef]

78. Blanchard, J.L.; Lynch, M. Organellar genes: Why do they end up in the nucleus? *Trends Genet.* **2000**, *16*, 315–320. [CrossRef]

79. Khakhlova, O.; Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* **2006**, *46*, 85–94. [CrossRef]

80. Chung, K.P.; Gonzalez-Duran, E.; Ruf, S.; Endries, P.; Bock, R. Control of plastid inheritance by environmental and genetic factors. *Nat. Plants* **2023**, *9*, 68–80. [CrossRef]

81. Renoult, J.P.; Kjellberg, F.; Grout, C.; Santoni, S.; Khadari, B. Cyto-nuclear discordance in the phylogeny of Ficus section *Galoglychia* and host shifts in plant-pollinator associations. *BMC Evol. Biol.* **2009**, *9*, 248. [CrossRef]

82. Huang, D.I.; Hefer, C.A.; Kolosova, N.; Douglas, C.J.; Cronk, Q.C. Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, *Populus balsamifera* and *P. trichocarpa* (Salicaceae). *New Phytol.* **2014**, *204*, 693–703. [CrossRef]

83. Roch, S.; Nute, M.; Warnow, T. Long-branch attraction in species tree estimation: Inconsistency of partitioned likelihood and topology-based summary methods. *Syst. Biol.* **2019**, *68*, 281–297. [CrossRef]

84. Smith, S.A.; Walker-Hale, N.; Walker, J.F.; Brown, J.W. Phylogenetic conflicts, combinability, and deep phylogenomics in plants. *Syst. Biol.* **2020**, *69*, 579–592. [CrossRef] [PubMed]

85. Liu, Y.; Wang, S.; Li, L.; Yang, T.; Dong, S.; Wei, T.; Wu, S.; Liu, Y.; Gong, Y.; Feng, X. The Cycas genome and the early evolution of seed plants. *Nat. Plants* **2022**, *8*, 389–401. [CrossRef] [PubMed]

86. Duan, L.; Fu, L.; Chen, H.-F. Phylogenomic cytonuclear discordance and evolutionary histories of plants and animals. *Sci. China Life Sci.* **2023**, *66*, 2946–2948. [CrossRef] [PubMed]

87. Tamashiro, R.A.; White, N.D.; Braun, M.J.; Faircloth, B.C.; Braun, E.L.; Kimball, R.T. What are the roles of taxon sampling and model fit in tests of cyto-nuclear discordance using avian mitogenomic data? *Mol. Phylogenetics Evol.* **2019**, *130*, 132–142. [CrossRef]

88. Pandey, A.; Braun, E.L. The roles of protein structure, taxon sampling, and model complexity in phylogenomics: A case study focused on early animal divergences. *Biophysica* **2021**, *1*, 87–105. [CrossRef]

89. Cummings, M.P.; Meyer, A. Magic bullets and golden rules: Data sampling in molecular phylogenetics. *Zoology* **2005**, *108*, 329–336. [CrossRef]

90. Pollock, D.D.; Bruno, W.J. Assessing an unknown evolutionary process: Effect of increasing site-specific knowledge through taxon addition. *Mol. Biol. Evol.* **2000**, *17*, 1854–1858. [CrossRef]

91. Young, A.D.; Gillung, J.P. Phylogenomics—Principles, opportunities and pitfalls of big-data phylogenetics. *Syst. Entomol.* **2020**, *45*, 225–247. [CrossRef]

92. Philippe, H.; Brinkmann, H.; Lavrov, D.V.; Littlewood, D.T.J.; Manuel, M.; Wörheide, G.; Baurain, D. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol.* **2011**, *9*, e1000602. [CrossRef]

93. Heath, T.A.; Hedtke, S.M.; Hillis, D.M. Taxon sampling and the accuracy of phylogenetic analyses. *J. Syst. Evol.* **2008**, *46*, 239.

94. Thureborn, O.; Wikström, N.; Razafimandimbison, S.G.; Rydin, C. Plastid phylogenomics and cytonuclear discordance in Rubioideae, Rubiaceae. *PLoS ONE* **2024**, *19*, e0302365. [CrossRef] [PubMed]

95. Altenhoff, A.M.; Glover, N.M.; Dessimoz, C. Inferring orthology and paralogy. In *Evolutionary Genomics: Statistical and Computational Methods*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 149–175.

96. Mendes, F.K.; Hahn, M.W. Why concatenation fails near the anomaly zone. *Syst. Biol.* **2018**, *67*, 158–169. [CrossRef] [PubMed]

97. Yan, Z.; Smith, M.L.; Du, P.; Hahn, M.W.; Nakhleh, L. Species tree inference methods intended to deal with incomplete lineage sorting are robust to the presence of paralogs. *Syst. Biol.* **2022**, *71*, 367–381. [CrossRef] [PubMed]

98. Rose, J.P.; Toledo, C.A.; Lemmon, E.M.; Lemmon, A.R.; Sytsma, K.J. Out of sight, out of mind: Widespread nuclear and plastid-nuclear discordance in the flowering plant genus *Polemonium* (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Syst. Biol.* **2021**, *70*, 162–180. [CrossRef]

99. Bergsten, J. A review of long-branch attraction. *Cladistics* **2005**, *21*, 163–193. [CrossRef]

100. Coiro, M.; Roberts, E.A.; Hofmann, C.-C.; Seyfullah, L.J. Cutting the long branches: Consilience as a path to unearth the evolutionary history of Gnetales. *Front. Ecol. Evol.* **2022**, *10*, 1082639. [CrossRef]

101. Rieseberg, L.H.; Whitton, J.; Randal Linder, C. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Bot. Neerl.* **1996**, *45*, 243–262. [CrossRef]

102. Liston, A.; Parker-Defeniks, M.; Syring, J.V.; Willyard, A.; Cronn, R. Interspecific phylogenetic analysis enhances intraspecific phylogeographical inference: A case study in Pinus lambertiana. *Mol. Ecol.* **2007**, *16*, 3926–3937. [CrossRef]

103. Acosta, M.C.; Premoli, A.C. Evidence of chloroplast capture in south American Nothofagus (subgenus Nothofagus, Nothofagaceae). *Mol. Phylogenetics Evol.* **2010**, *54*, 235–242. [CrossRef]

104. Nauheimer, L.; Boyce, P.C.; Renner, S.S. Giant taro and its relatives: A phylogeny of the large genus Alocasia (Araceae) sheds light on Miocene floristic exchange in the Malesian region. *Mol. Phylogenetics Evol.* **2012**, *63*, 43–51. [CrossRef]

105. Liu, P.-L.; Wen, J.; Duan, L.; Arslan, E.; Ertuğrul, K.; Chang, Z.-Y. *Hedysarum* L. (Fabaceae: Hedysareae) is not monophyletic–evidence from phylogenetic analyses based on five nuclear and five plastid sequences. *PLoS ONE* **2017**, *12*, e0170596. [CrossRef] [PubMed]

106. Zhong, B.; Deusch, O.; Goremykin, V.V.; Penny, D.; Biggs, P.J.; Atherton, R.A.; Nikiforova, S.V.; Lockhart, P.J. Systematic error in seed plant phylogenomics. *Genome Biol. Evol.* **2011**, *3*, 1340–1348. [CrossRef]

107. Zhong, Z.-R.; Li, N.; Qian, D.; Jin, J.-H.; Chen, T. Maternal inheritance of plastids and mitochondria in *Cycas* L. (Cycadaceae). *Mol. Genet. Genom.* **2011**, *286*, 411–416. [CrossRef]

108. Galbraith, D.W.; Harkins, K.R.; Maddox, J.M.; Ayres, N.M.; Sharma, D.P.; Firoozabady, E. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **1983**, *220*, 1049–1051. [CrossRef]

109. Stewart, C.J.; Via, L.E. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *BioTechniques* **1993**, *14*, 748–751.

110. Kolosova, N.; Miller, B.; Ralph, S.; Ellis, B.E.; Douglas, C.; Ritland, K.; Bohlmann, J. Isolation of high-quality RNA from gymnosperm and angiosperm trees. *Biotechniques* **2004**, *36*, 821–824. [CrossRef]

111. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **2017**, *27*, 722–736. [CrossRef]

112. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef]

113. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

114. Zhang, Y.; Park, C.; Bennett, C.; Thornton, M.; Kim, D. Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Res.* **2021**, *31*, 1290–1295. [CrossRef]

115. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008. [CrossRef]

116. Gabriel, L.; Brůna, T.; Hoff, K.J.; Ebel, M.; Lomsadze, A.; Borodovsky, M.; Stanke, M. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **2024**, *34*, 769–777. [CrossRef] [PubMed]

117. Kuznetsov, D.; Tegenfeldt, F.; Manni, M.; Seppey, M.; Berkeley, M.; Kriventseva, E.V.; Zdobnov, E.M. OrthoDB v11: Annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* **2023**, *51*, D445–D451. [CrossRef] [PubMed]

118. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [CrossRef] [PubMed]

119. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]

120. Morel, B.; Kozlov, A.M.; Stamatakis, A. ParGenes: A tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. *Bioinformatics* **2019**, *35*, 1771–1773. [CrossRef]

121. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]