BMC Plant Biology

RESEARCH

Open Access

Comparative genomics and phylogenetic analysis of mitochondrial genomes of *Neocinnamomum*

Wen Zhu[1], Di Zhang[2], Wenbin Xu[3], Yi Gan[4], Jiepeng Huang[5], Yanyu Liu[5], Yunhong Tan[2], Yu Song[5*] and Peiyao Xin[1*]

Abstract

Background *Neocinnamomum* plants are considered a promising feedstock for biodiesel in China, due to the richness in long-chain fatty acids (LCFAs) found in their seeds. However, the mitochondrial genome (mitogenome) of this genus has not yet been systematically described, and the exploration of species relationships within this genus using mitogenome sequences is also an uncharted territory. This has hindered our understanding of mitogenome diversity and the evolutionary relationships within *Neocinnamomum*.

Results In this study, a total of 24 individuals representing seven distinct taxa from the genus *Neocinnamomum* were subjected to Illumina sequencing, and the species *N. delavayi* was sequenced using Oxford Nanopore sequencing technology. We successfully assembled the mitogenome of *N. delavayi*, which is 778,066 bp in size and exhibits a single circular structure. The analysis identified 659 dispersed repeats, 211 simple sequence repeats (SSRs), and 30 tandem repeats within the mitogenome. Additionally, 37 homologous fragments, totaling 9929 bp, were found between the mitogenome and the plastid genome (plastome). The codons of 41 protein-coding genes (PCGs) had a preference for ending in A/T, and the codon usage bias of the majority of these genes was influenced by natural selection pressures. Comparative genomic analysis revealed low collinearity and significant gene rearrangements between species. Phylogenetic analysis resulted in the classification of *Neocinnamomum* into six distinct clades, contradicting previous findings which based on complete plastomes and nuclear ribosomal cistron (nrDNA). In the PCGs of 24 individuals, 86 mutation events were identified, which included three indels and 83 SNPs. Notably, the *ccmC* gene underwent positive selection in pairwise comparisons of three species pairs. Furthermore, 748 RNA editing sites were predicted within the PCGs of the *N. delavayi* mitogenome.

Conclusions This study enriches our knowledge of the mitogenomes in the family Lauraceae, and provides valuable data and a foundation for genomic evolution research, genetic resource conservation, and molecular breeding in *Neocinnamomum*.

Keywords Lauraceae, Oil plant, Mitochondrial genome, Homologous fragments, Phylogenomic inconsistencies

*Correspondence:
Yu Song
songyu@gxnu.edu.cn
Peiyao Xin
xpytgyx@163.com
Full list of author information is available at the end of the article

Background

Neocinnamomum H. Liu, a member of the family Lauraceae, is a group of evergreen shrubs or small trees that inhabit Asian tropical and subtropical regions [1]. Of the seven known species in this genus, six have been discovered in China [2]. The genus is distinguished by its small, conical buds with thick, often pubescent scales; the leaves are alternate; umbellate inflorescences comprising single or multiple flowers; the perianth tube is short and small, with six tepals divided into inner and outer whorls; the fruit is a berry-like drupe, elliptical or globose in shape [2]. The branches, leaves, and barks of many species in this genus are rich in essential oils, making them important oil plants and valuable medicinal resources [3]. It has been reported that the seeds of Lauraceae mainly produce medium-chain fatty acids (MCFAs), whereas the seeds of *Neocinnamomum* exclusively contain long-chain fatty acids (LCFAs), a content that is extremely unusual in the entire Lauraceae and even Magnoliidae, and has been proposed as a possible biodiesel source in China [4, 5].

Early phylogenetic studies relied on the fossil record and morphological and physiological comparisons of species to construct the primary framework for their evolutionary history [6]. *Neocinnamomum* was initially distinguished from other genera in the family Lauraceae by Liou [7] based on the position of the four cells in the anther [7]. However, this characteristic was not universally present within the genus, leading other scholars to reclassify *Neocinnamomum* under *Cinnamomum*. Kostermans [8] subsequently reinstated *Neocinnamomum* as a distinct genus, because of the peculiar inflorescence, the thick, fleshy, obconical, shallow fruit cup with persistent, enlarged tepals and the distichous leaves. Nevertheless, morphological and anatomical evidence has not fully elucidated the interspecific relationships within the genus or the intergeneric relationships within the family. The application of molecular data in phylogenetic studies has dramatically expanded over the past two decades. The pioneering molecular investigation of the genus *Neocinnamomum* conducted by Chanderbali et al. [9], leveraging plastid fragment sequences and the nuclear ribosomal barcode 26S rDNA to elucidate the phylogenetic relationships within Lauraceae. This study revealed that *Neocinnamomum*, *Cassytha*, and *Caryodaphnopsis* were clustered into a single clade, with *Neocinnamomum* and *Cassytha* in a sisterly relationship; however, the length of their respective branches hindered the authority of the conclusions. To ascertain the intergeneric relationships among these taxa, Rohwer et al. [10] utilized the *trnK* intron, incorporating the *matK* gene, for phylogenetic analysis. By employing Bayesian inference (BI), they discovered that *Neocinnamomum* is more closely clustered with *Caryodaphnopsis* than with *Cassytha*, a finding that concurs with prior morphological studies. Subsequently, Wang et al. [11] conducted a molecular phylogenetic study using sequences from three molecular markers: *trnK*, *psbA–trnH*, and ITS. The study indicated that *Neocinnamomum* constitutes a monophyletic group, marking the first confirmation of this taxonomic status. Nevertheless, the aforementioned studies were either constrained by a limited sample size or the information provided by the sequence fragments was insufficient. As a result, their conclusions did not provide a resolution to the phylogenetic questions at hand.

Recently, Song et al. [12, 13] employed the complete plastid genome (plastome) sequences to delineate the evolutionary pattern and phylogenetic positioning of *Neocinnamomum* within Lauraceae, and established the tribe Neocinnamomeae. Other researchers have utilized complete plastome and nuclear ribosomal cistron (nrDNA) sequences to explore the interspecific relationships within *Neocinnamomum*, thereby exposing inconsistencies between nuclear and plastomes in the context of phylogenetic analyses [14–16]. However, to date, no studies have been reported on the mitochondrial genomes (mitogenome) of *Neocinnamomum*, and there are no published accounts of utilizing mitogenome sequences to reconstruct interspecific relationships within this genus.

Plant mitochondria are fundamental in biological processes like energy generation, cell differentiation, and death, which are closely related with plant development, adaptability, and reproductive capacity [17]. The size variation of plant mitogenomes is immense, with significant structural rearrangements occurring even among closely related species [18, 19]. The highly variable intergenic regions are filled with repetitive sequences, which may lead to frequent structural rearrangements, gene losses, and an amount of endogenous and exogenous DNA transfers [20, 21]. Concurrently, a multitude of RNA editing events take place during the process of gene transcription [22]. In contrast to these rapid and dynamic features, plant mitochondrial genes exhibit an exceptionally low nucleotide substitution rate [23]. The marked inconsistency in the pace of evolution between plant mitogenomes and genes presents numerous enigmas in the evolutionary narrative of mitochondria. This complexity renders the genome assembly process challenging, especially when dealing with highly rearranged genomes. Nevertheless, it is this very complexity that positions mitochondria as an ideal subject for studying the evolutionary process of genome complexity, providing valuable insights into our understanding of genome evolution.

In this study, we completed the sequencing and assembly for the first mitogenome of *N. delavayi*, and conducted comprehensive and in-depth analyses based on it. This study aims to: (1) Elucidate the characteristics and organization of the mitogenome; (2) Examine the transfer of genetic material between organelles in *N. delavayi* and the conservation of gene order among mitogenomes; (3) Explore species relationships and identify discrepancies in phylogenetic reconstructions across different genomic sequences; and (4) Investigate the adaptive evolution of mitochondrial genes within *Neocinnamomum*. Through these endeavors, a more profound comprehension of the features, genetic diversity, and adaptive evolution of the mitogenomes of *Neocinnamomum* was obtained. Phylogenetic relationships constructed based on the mitogenome sequences fill a gap in the evolutionary research of this genus, providing new insights into its phylogenetic framework.

Results

Features of the mitogenome of *N. delavayi*

The complete mitogenome of *N. delavayi* was constructed as a single circular molecular of 778,066 bp in length (Fig. 1A). The genome's overall base composition consists of A: 206,785 bp (26.58%), T: 206,476 bp (26.54%), G: 181,271bp (23.3%), C: 183,534 bp (23.59%), yielding in a G+C content of 46.89%. It harbors a total of 66 functional genes, including 41 that encode proteins involved in the respiratory chain, genetic system, and ATP synthase, as well as 22 genes that encode tRNAs and three that encode rRNAs (Additional file 2: Table S1). The 41 PCGs varied in length, ranging from 225 bp for *atp9* to 2001 bp for *nad5*. The combined length of all PCGs was 34,263 bp, which represents about 4.4% of the whole mitogenome (Additional file 2: Table S2). There are three tRNA genes that contain multiple copies: *trnfM-CAU* and *trnM-CAU* each have two copies, and *trnP-UGG* has three copies. Ten genes share a total of 19 cis-splicing introns. Among these genes, five (*ccmFC*, *nad1*, *rpl2*, *rps3*, and *rps10*) have one each, *cox2* and *nad5* have two each, *nad2* and *nad4* have three each, and *nad7* has four. Furthermore, the whole plastome of *N. delavayi* was successfully assembled into a circular molecular structure, while the nrDNA was assembled into a separate linear molecular structure, with lengths of 150,842 bp and 6753 bp, respectively.

Repeats elements analysis

Dispersed repeats, simple sequence repeats (SSRs), and tandem repeats have been detected within the mitogenome of *N. delavayi* (Fig. 1B). There were 659 dispersed repeats identified, each extending 30 bp or above in length. The repeats encompassed 370 forward repeats, 287 palindromic repeats, a pair of reverse repeats, and no complementary repeats (Fig. 1C, and Additional file 2: Table S3). Approximately 64.64% of the repetitive sequences have lengths between 30 and 39 bp. Additionally, 94.99% of the repetitive sequences are less than 100 bp. Only 33 pairs of repetitive sequences have lengths that are equal to or higher than 100 bp (Fig. 1D). In the meanwhile, a total of 211 SSRs with lengths ranging from 10 to 20 bp were discovered. These SSRs consisted of 40 mononucleotide (18.96%), 44 dinucleotide (20.85%), 33 trinucleotide (15.64%), 81 tetranucleotide (38.39%), ten pentanucleotide (4.74%), and three hexanucleotide (1.42%) (Fig. 1E and Additional file 2: Table S4). Among the six SSR types, the three most frequent were tetranucleotide repeats, dinucleotide repeats, and mononucleotide repeats, with their major repeat unit types being AAAG/CTTT, AC/GT, and A/T, respectively (Fig. 1F). Additionally, 30 tandem repeats with ≥73% matches and lengths between 10 and 54 bp were identified in this genome (Fig. 1C and Additional file 2: Table S5). The total lengths of the dispersed repeats, SSRs, and tandem repeats were 67,024 bp, 2521 bp, and 1324 bp, respectively, accounting for 8.61%, 0.324%, and 0.17% of the overall length of the mitogenome. Remarkably, the mapping of all repeats showed that none were situated on genes.

Analysis of codon usage bias

In *N. delavayi* mitogenome, 41 PCGs were extracted, identifying 11,421 codons, with an average of approximately 279 codons per PCG (Additional file 2: Table S6). These codons are classified into 64 types, with 61 codons corresponding to the 20 various amino acids, while the remaining three codons serve as stop signal codons (Fig. 2A). The most abundant amino acids encoded are serine (Ser, 9.76%), leucine (Leu, 9.55%), isoleucine (Ile, 7.55%), and arginine (Arg, 7.36%). Excluding stop codons, the least represented amino acids in the mitogenome were tryptophan (Trp, 1.35%) and cysteine (Cys, 1.44%). The ATG codon, which serves as the start codon for encoding formyl-methionine, was the most often detected start codon. A value of RSCU higher than 1 indicates a preference for certain codons. Of the 34 codons with RSCU values of 1 or greater, 16 ended with U and 13 with A, together accounting for 85.29% of these codons. Conversely, only five codons ended with either C or G. Additionally, 90 percent of the codons with RSCU values less than 1 terminated with the nucleotides C or G. Among the 64 codons analyzed, the TAA codon, which encodes alanine, had the highest RSCU value at 1.69, whereas the CAC codon, encoding histidine, had the lowest value at 0.44.

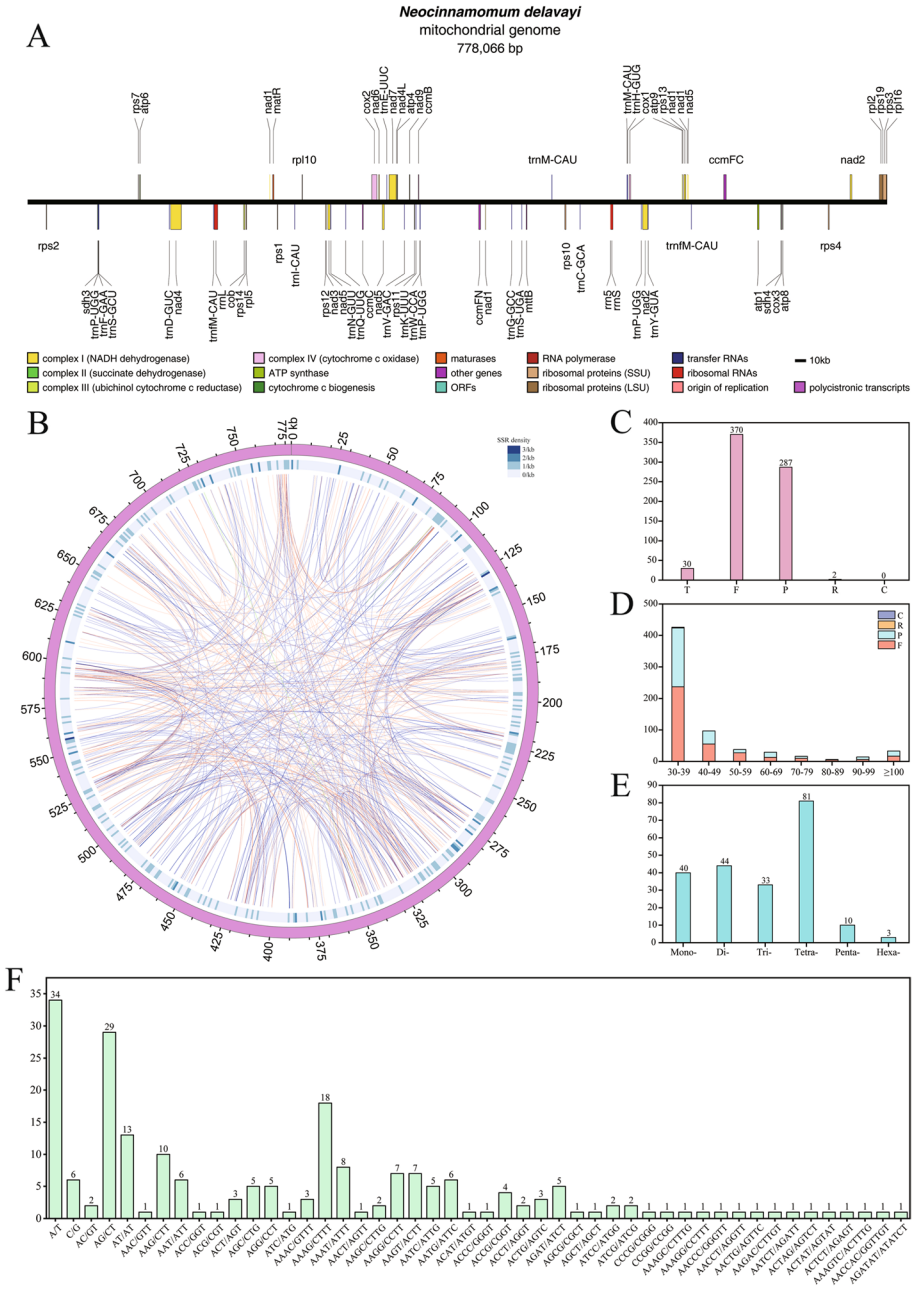

Fig. 1 **A** Mitogenome map of *N. delavayi*. **B** The distribution of repeated elements. The inside lines of various colors show the presence of distributed repeats with lengths equal to or higher than 30 bp. The color blue indicates F, orange represents P, and red represents R. The inner circular map displays the concentration of SSRs. T, F, P, R, and C represent tandem, forward, palindromic, reverse, and complement repeats, respectively. **C** The type and quantity of repeats. **D** Count of dispersed repetitive elements in various length ranges. **E** Number of distinct forms of SSR. **F** Number of distinct SSR repeat units

In the 41 PCG sequences of the *N. delavayi* mitogenome, the GC content ranges from 36.63% (*nad4L*) to 53.1% (*matR*). The content of GC1, GC2, and GC3 was determined and recorded in Fig. 2B and Additional file 2: Table S7. The results showed that GC1 had the highest values, ranging from 38.12% (*atp8*) to 58.38% (*atp1*), with an average of 48.83%. GC2 values were slightly lower, ranging from 34.71% (*rps10*) to 56% (*atp9*), with an average of 44.82%. GC3 values were the lowest, ranging from 28.21% (*rps13*) to 58.06% (*matR*), with an average of 38.68%. The mean GC content at various codon locations was below 50%, suggesting a bias towards A/T bases and

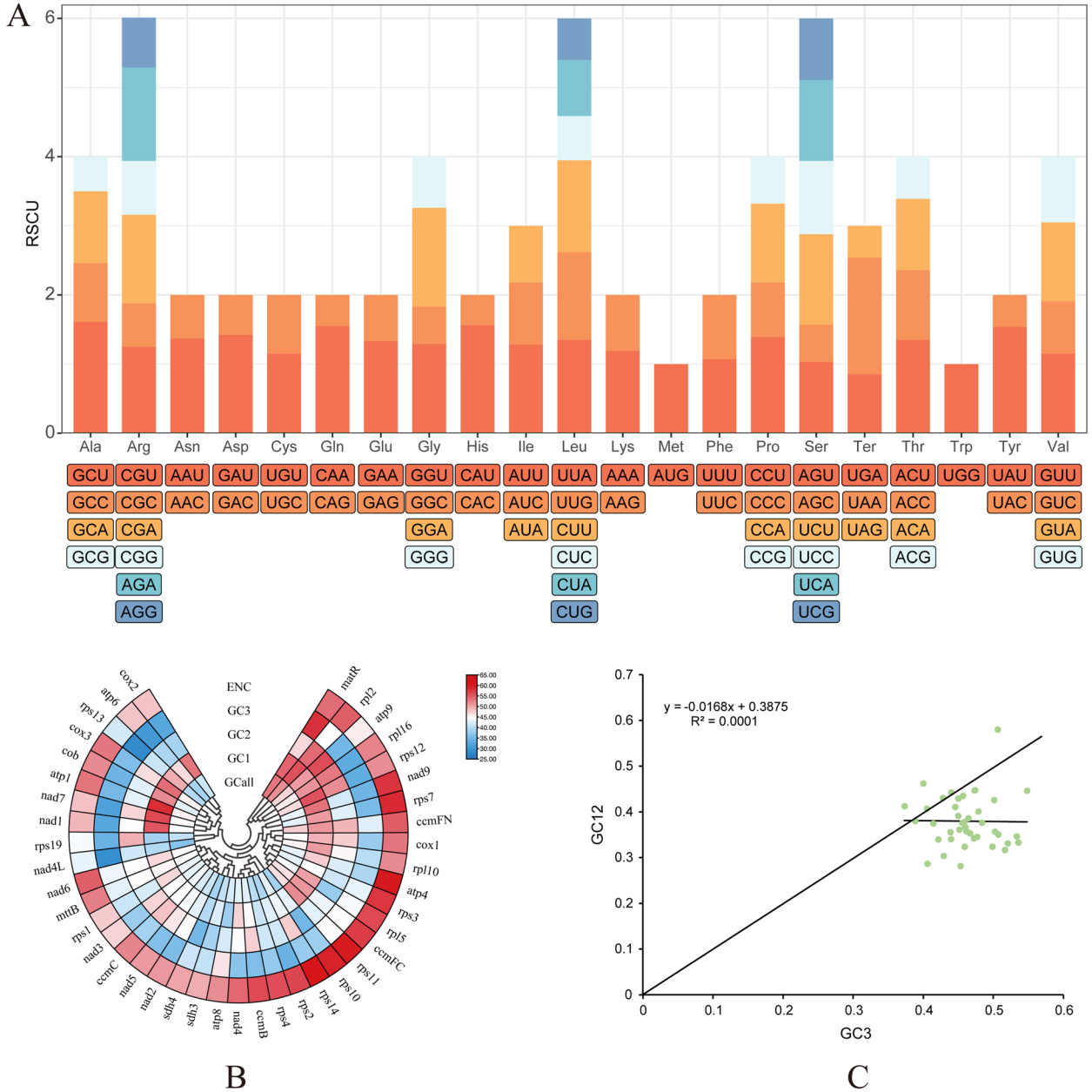

Fig. 2 **A** Codon usage patterns of the *N. delavayi* mitogenome. **B** GC content of different positions of the codon. **C** Performing a plot analysis of the GC12 and GC3 neutrality for the whole coding DNA sequence in the mitogenome of *N. delavayi*

A/T-ending codons in the *N. delavayi* PCGs. The neutrality plot analysis was used to determine the link between GC12 and GC3, revealing a minor negative correlation that was not statistically significant. The regression value of −0.0168 indicates a modest association between GC12 and GC3. This suggests that codon use bias is mostly controlled by natural selection (Fig. 2C).

The ENC values for the 41 PCGs varied from 41.79 (*rps13*) to 61 (*rps14*), with an average value of 53.54. This value is higher than 35, suggesting a little preference for certain synonymous codons in the genes analyzed (Additional file 2: Table S7). The ENC score of 61 for *rps14* indicates that this gene does not display any bias in the utilization of synonymous codons. The ENC-plot offers a concise and easily understandable representation of the codon use bias seen in genes. The findings indicate that eight genes exhibit codon use bias mostly due to mutation, since they are either above or near the standard

curve (Additional file 1: Figure S1). All the remaining genes are positioned below the standard curve, suggesting that the codon use bias may be affected by natural selection.

Identification of mitochondrial plastid DNAs (MTPTs)

The BLAST analysis of the mitogenome and plastome of *N. delavayi* revealed the presence of 40 homologous fragments. Among them, three fragments were found to have duplicate copies in the mitogenome and were consistently recognized in the plastome. Following the elimination of redundant sequences, there were 37 homologous fragments remaining, totaling 9929 bp. These fragments accounted for 6.29% of the plastome (Fig. 3 and Additional file 2: Table S8). The sizes of the 37 homologous fragments varied greatly, ranging from 32 to 1325 bp, resulting in a difference of about 41.4 fold. Out of these, 23 homologous fragments were found in the LSC region of the plastome, with a combined length of 5597 bp, which accounts for 6.09% of the overall length of the LSC. 14 homologous fragments were identified in the IR regions, totaling 4332 bp (9.61% of the IR length).

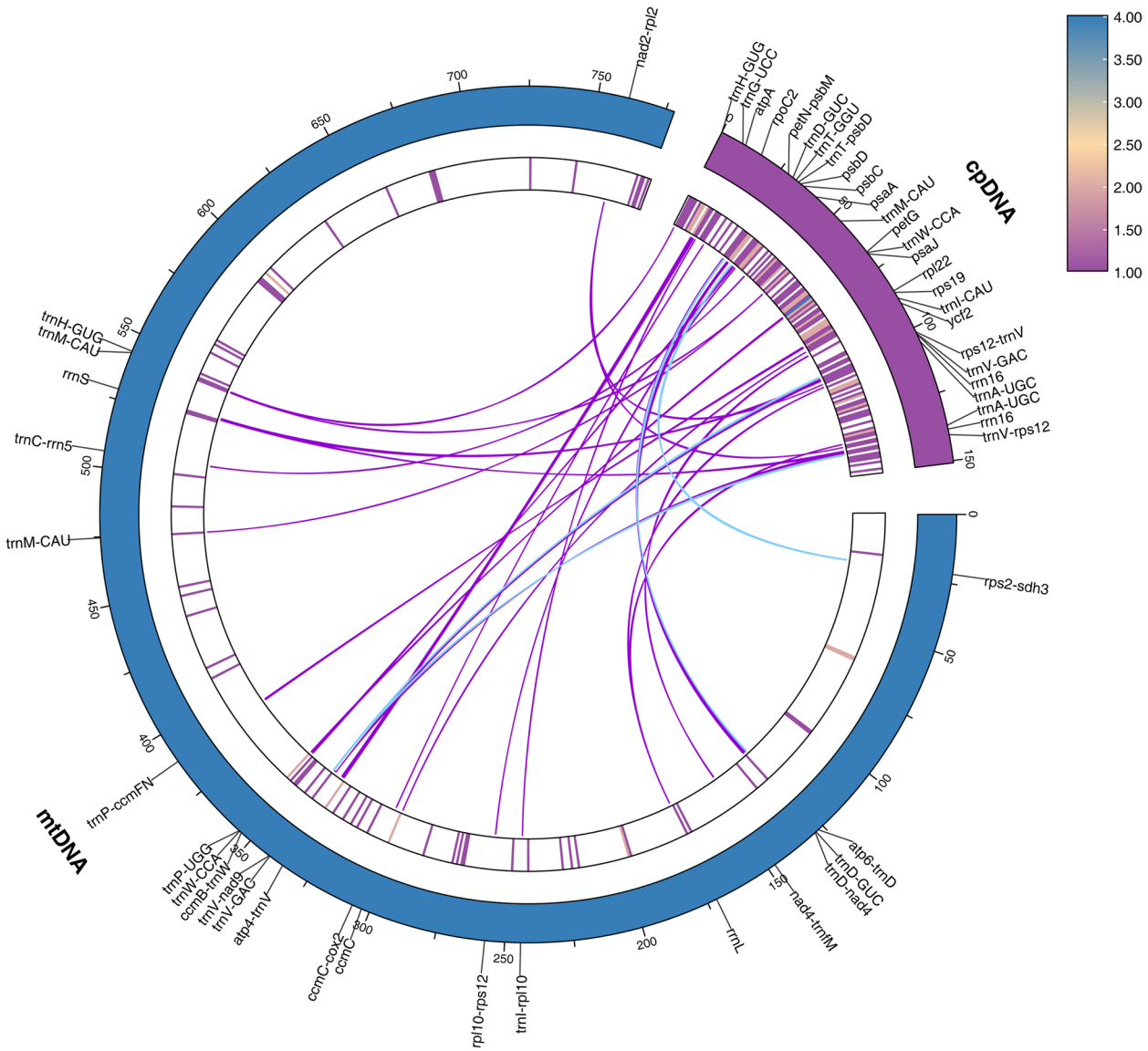

Fig. 3 Homologous fragments in both the mitogenome and plastome of *N. delavayi*. The outside arcs in blue and purple correspond to the mitogenome and plastome, respectively. The inner arcs show the density of gene locations that relate to each other, in units of 1 kb. The internal curves depict homologous fragments, with purple curves representing genes and blue curves representing intergenic regions. Additional file 2: Table S8 contains more specific information

However, no homologous fragment was found in the SSC region. Annotation of all homologous fragments identified 29 fragments in the plastome that were either complete or incomplete genes, including double-copy genes. The set of genes identified consisted of 11 PCGs (nine of which were incomplete and two were complete), 14 tRNA genes (10 of which were complete and four were incomplete), and four incomplete rRNA genes. Out of the 29 genes found in the plastome, only 12 corresponded to homologous fragments which contain genes in the mitogenome. These include six complete genes and four incomplete genes (Fig. 3 and Additional file 2: Table S8). The remaining genes may have been lost or undergone pseudogene alterations in the mitogenome.

Analysis of collinearity among *N. delavayi* and other two species

Both coding and non-coding regions exhibited frequent rearrangement events (Additional file 1: Figure S2). This research primarily emphasizes the reorganization of PCGs rather than the whole sequence of the mitogenome. The mitogenomes of *Caryodaphnopsis henryi*, *Hernandia nymphaeifolia*, and *N. delavayi* showed significant rearrangements of their PCGs, with a considerable variance in the number of conserved gene clusters. There are 15 genes between *H. nymphaeifolia* and *N. delavayi* that form six conserved gene clusters (*cob-rps14-rpl5*, *rps12-nad3*, *nad4-nad4L*, *nad9-rps11*, *sdh4-cox3-atp8*, and *rps19-rps3-rpl16*) (Fig. 4). These clusters are not only shared between *N. delavayi* and *C. henryi* but four of them have expanded: *nad9-rps11*, *rps12-nad3*, and *rps19-rps3-rpl16* have each gained one gene, whereas *sdh4-cox3-atp8* has acquired two. In addition to the six shared conserved gene clusters, *N. delavayi* vs. *C. henryi* had five more conserved gene clusters than *H. nymphaeifolia* vs. *N. delavayi*, averaging about two genes per conserved cluster. Although *N. delavayi* and *C. henryi* are more closely related, they still exhibit significant rearrangements in the order of their gene clusters. Additionally, within these species, some gene clusters, such as *rps12-nad3*, exhibit intra-cluster gene order variations.

Phylogenetic analysis of *Neocinnamomum*

A 33,899 bp matrix was obtained by concatenating 41 trimmed PCGs, and the phylogenetic relationships of *Neocinnamomum* were reconstructed using this matrix with *C. henryi* as the outgroup. The phylogenetic relationships among the species were determined using both Maximum likelihood (ML) and BI analysis, and most relationships had high internal support (Fig. 5A). The BI tree exhibited congruence with the topology derived from the ML analysis at the species-level. The 24 individuals of *Neocinnamomum* were divided into six clades, which corresponded to seven distinct taxa.

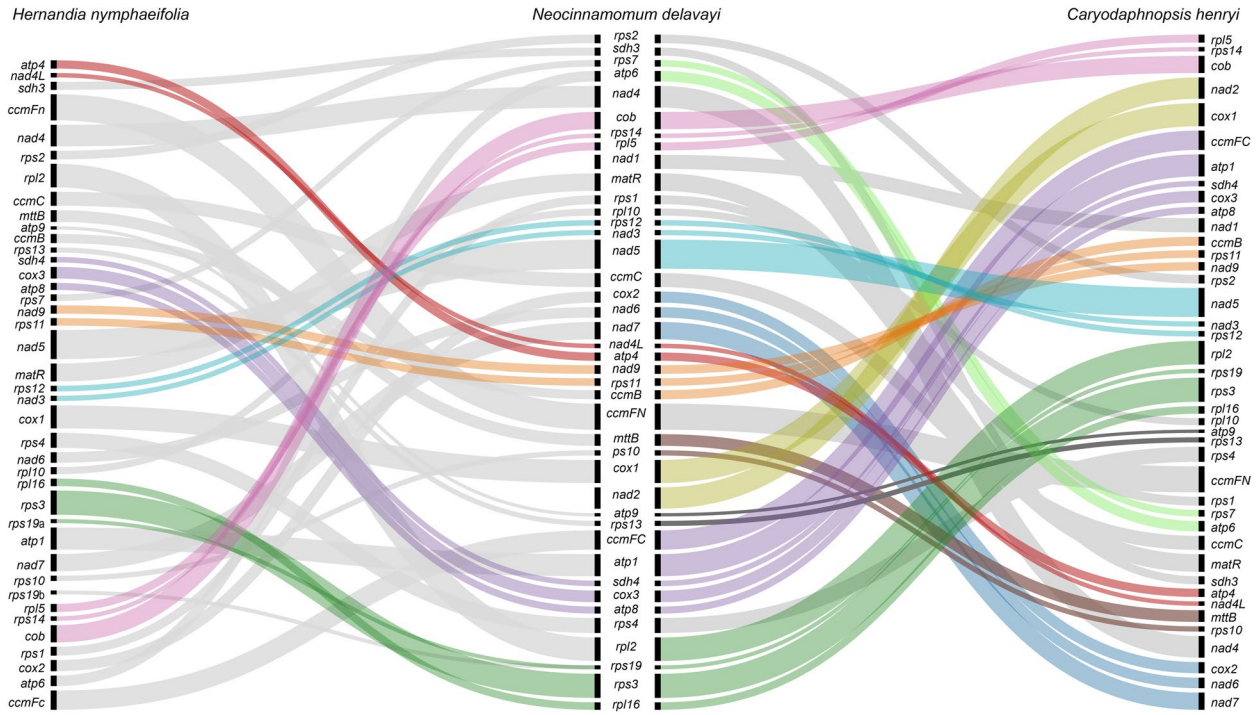

Fig. 4 The colinearity among genes in the mitogenomes of *N. delavayi*, *H. nymphaeifolia*, and *C. henryi*. Different colored lines represent different gene clusters, and gray lines represent genes that are not syntenic. Black bold lines of different lengths indicate the length of each gene

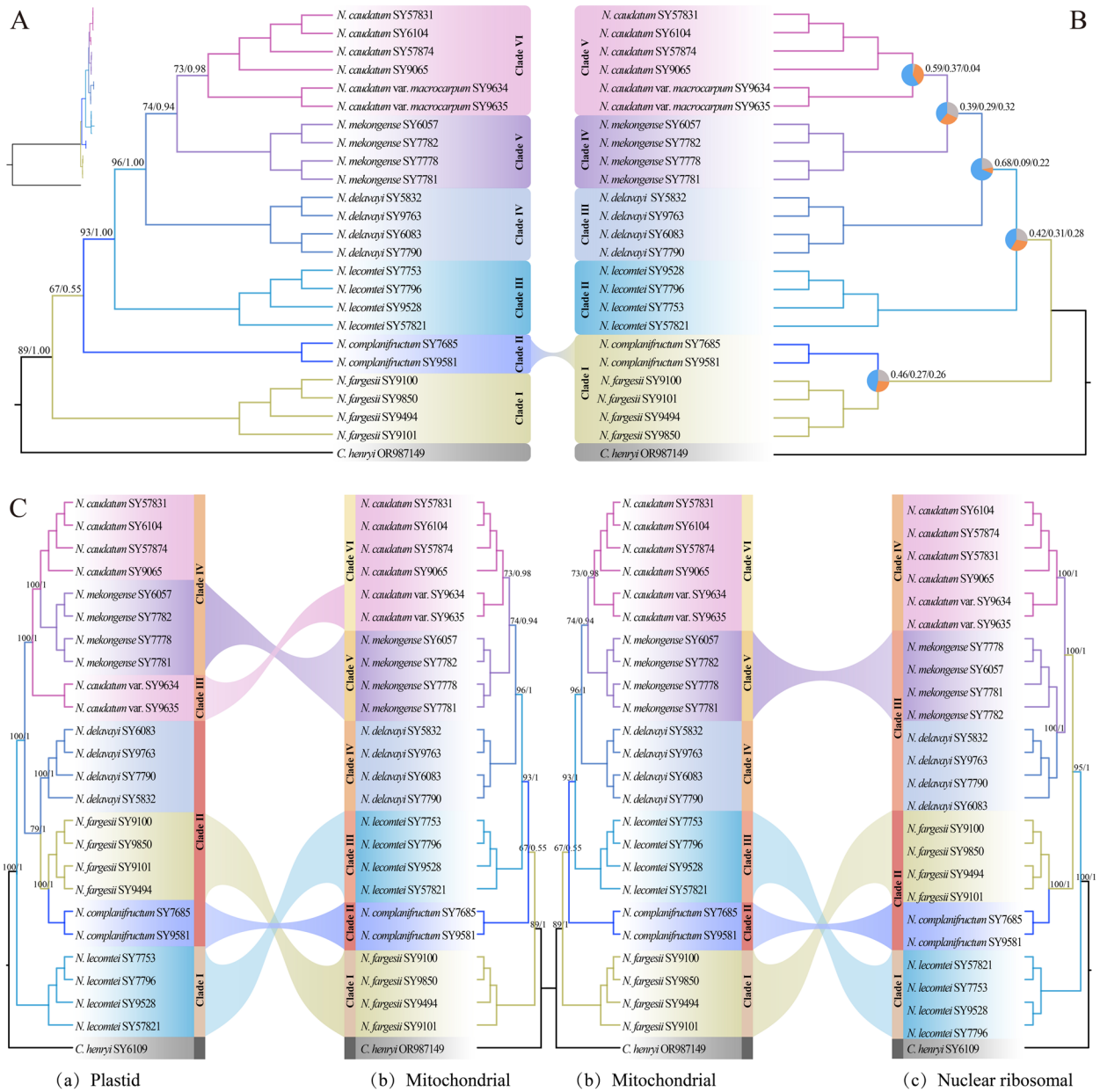

Fig. 5 **A** The molecular phylogenetic trees of 24 individuals from seven species of *Neocinnamomum* were constructed using mitochondrial sequences and analyzed using ML and BI methods. **B** Coalescent ASTRAL analysis. The pie charts represent the distribution of quartet support for different topologies. The main topology (q1) is represented by the blue color, the first alternative topology (q2) is represented by the yellow color, and the second alternative topology (q3) is represented by the grey color. The numbers shown on the nodes correspond to the q1/q2/q3 values. The concatenation and coalescent trees are rooted using the mitochondrial sequence of *C. henryi*. **C** Phylogenetic relationships of *Neocinnamomum* constructed based on different genomic sequences. These trees were created using plastid, mitochondrial, and nuclear ribosomal information, and the ML and BI approaches were employed for reconstruction. (**a**) *Neocinnamomum* phylogenetic trees reconstructed by Cao et al. [15]; (**b**) The phylogenetic trees reconstructed from the 41 PCGs of *Neocinnamomum* mitochondria; (**c**) *Neocinnamomum* phylogenetic trees based on nrDNA reconstruction by Li et al. [16]

Clade I (ML-BS = 89%, BI-PP = 1) consisted of four individuals of *N. fargesii*. This clade formed a sister group to the other five clades. Clade II (ML-BS = 67%, BI-PP = 0.55) consisted of two individuals of *N. complanifructum*. Clade III (ML-BS = 93%, BI-PP = 1), Clade IV (ML-BS = 96%, BI-PP = 1), and Clade V (ML-BS = 74%, BI-PP = 0.94) each consisted of four individuals each of *N. lecomtei*, *N. delavayi*, and *N.*

mekongense, respectively. Clade VI (ML-BS = 74%, BI-PP = 0.94) consisted of two individuals of *N. caudatum* var. *macrocarpum* and four individuals of *N. caudatum*, which are sister groups. Clades I, III, and IV had strong support in both ML and BI analyses, whereas Clades V and VI shown high support in BI analysis and moderate support in ML analysis. On the other hand, Clade II had the least amount of support in both ML and BI analysis.

Coalescent ASTRAL analysis was implemented to further explore the phylogenetic relationships of *Neocinnamomum* (Fig. 5B). Five major clades of *Neocinnamomum* were determined. Clades II, III, IV, and V from the main topology obtained by coalescent analysis were consistent with the concatenation analysis. The main difference is that, in the coalescent analysis, *N. fargesii* and *N. complanifructum* clustered together as a monophyletic branch, forming sister groups. In contrast, the concatenation analysis showed that *N. fargesii* and *N. complanifructum* did not cluster together as a clade or formed a sister relationship, but showed a polytomous structure.

Angiosperms possess three distinct genomes: the nuclear, plastid, and mitochondrial genomes. These genomes have been widely employed by researchers for phylogenetic studies. Here, the topological consistencies and discrepancies of several phylogenetic trees for *Neocinnamomum* were assessed (Fig. 5C). These trees were constructed utilizing sequence differences from various genome sources. The relationships between species inferred from both ML and BI methods were consistent in each genomic phylogenetic tree. In the mitochondrial phylogenetic tree, *Neocinnamomum* was divided into six clades, with *N. fargesii* occupying the most basal position. Conversely, in both the plastome and nrDNA trees, *Neocinnamomum* was split into four clades, with *N. lecomtei* at the basal position. In the plastome tree, *N. complanifructum* and *N. fargesii* formed a sister group, which, along with *N. delavayi*, comprised Clade II. However, in the mitochondrial tree, *N. complanifructum*, *N. fargesii*, and *N. delavayi* were differentiated into Clades II, I, and IV, respectively. In the nrDNA tree, *N. complanifructum* and *N. fargesii* clustered into Clade II, suggesting a closer phylogenetic relationship, while *N. delavayi* and *N. mekongense* are sisters clustered into Clade III. In both the mitochondrial and nrDNA trees, *N. caudatum* var. *macrocarpum* and *N. caudatum* were closely related and clustered. In contrast, the plastome tree displayed *N. caudatum* var. *macrocarpum* as a separate clade, with *N. mekongense* and *N. caudatum* as sister clades.

Nucleotide polymorphism (π) and mutation analysis of PCGs

The π values of 41 PCGs from 24 individuals ranged from 0 to 0.00302. The average π value was 0.00081, suggesting that the variations among the PCGs were insignificant. There were 34 PCGs with polymorphic nucleotide loci, the three PCGs with the highest π values were *rps7*, *ccmC*, and *rps2*, and there were seven PCGs had a π value of 0 (Fig. 6). Most PCGs had π values less than the mean, and only 18 PCGs (43.9%) had π values greater than the mean.

A total of 86 mutation events, including three indels and 83 substitutions, were identified in the sequences of 41 PCGs among the 24 *Neocinnamomum* individuals (Additional file 2: Table S9 and Table S10). The seven PCGs with π values of 0 did not have any mutation events, as expected. Of the three indels, two were in *ccmC*, and one was in *cob*, with lengths ranging from 1 to 17 bp. The 83 SNPs included two transitions (Ts) (*atp9* and *nad2*) and 81 transversions (Tv), with a Ts to Tv ratio of 1:40.5. Out of all the Tv, five were between T and A, one was between C and G, and the remaining 75 were associated with alterations in GC content (Additional file 2: Table S10). Further analysis of substitution events revealed that 51 out of the 83 sites were non-synonymous, indicating changes in amino acid sequences, while the remaining 32 were synonymous, implying no alterations in the encoded amino acid. Among these substitution events, 21 SNP sites were associated with NADH dehydrogenase, and 24 SNP sites were associated with ribosomal proteins (SSU).

Prediction of RNA editing events

A total of 748 putative RNA editing sites were detected in 41 PCGs, and all the observed changes in nucleotides during RNA editing were from cytidine (C) to thymine (T) (Additional file 2: Table S11). Among all PCGs, *nad4* had the highest number of sites of 65, whereas *rpl2*, *rps7*, *rps11*, *rps14*, and *sdh3* had the fewest potential sites, each with only two. They exhibited a difference of 32.5 fold (Fig. 7). The result highlights the significant variation in the number of editing sites across different genes. The proportion of RNA editing sites was calculated based on codon positions to assess whether there was a preference for certain codon locations. The analysis indicated that 253 (33.82%) editing sites were at the first codon position, 462 (61.76%) at the second, and 33 (4.41%) at the third. Furthermore, it was observed that nine genes associated with NADH dehydrogenase (*nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7*, and *nad9*) comprised 301 (40.24%) of the potential editing sites. An additional analysis was performed to examine the possible RNA

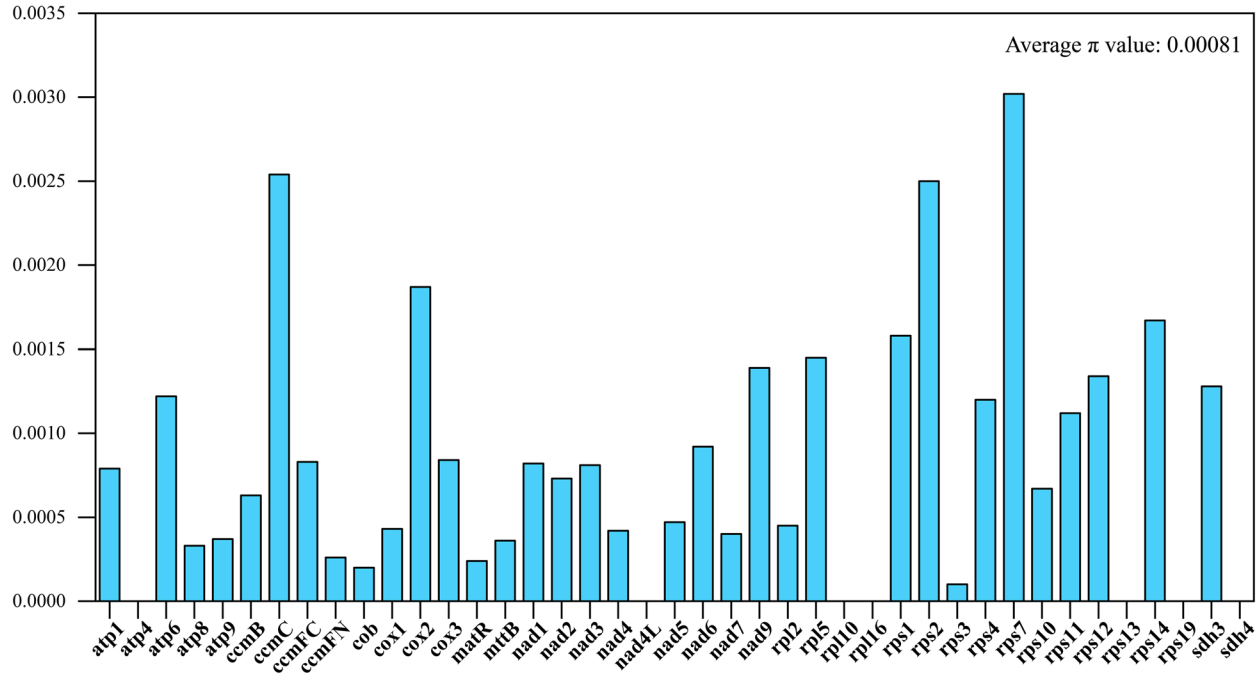

Fig. 6 The nucleotide polymorphism (π) values of the 41 PCGs among the 24 *Neocinnamomum* individuals

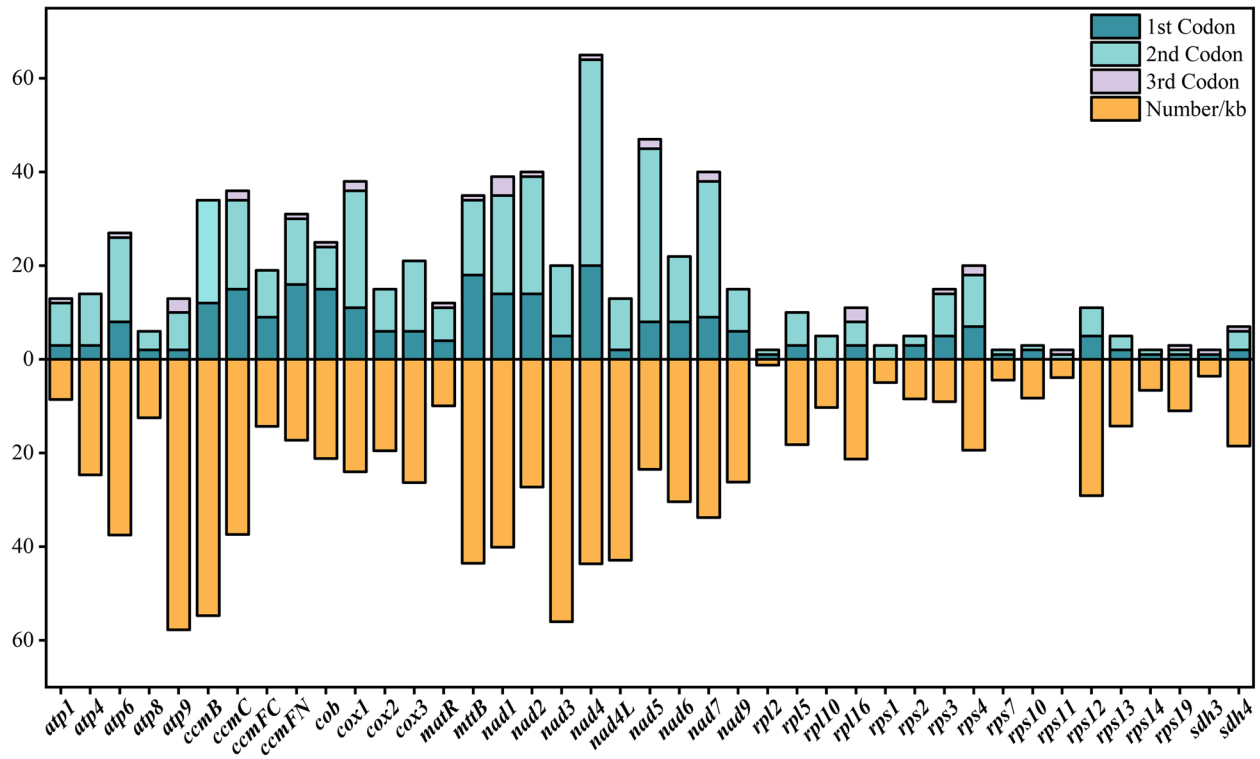

Fig. 7 Predicted number of RNA editing sites for PCGs in the mitogenome of *N. delavayi*. The first codon is represented by dark blue, the second codon by light blue, the third codon by purple, and the density of sites is shown by orange, measured in number per kb

editing sites at the start and stop codons of all PCGs, but no editing sites were found at these positions. Moreover, the frequency of potential editing sites within PCGs was statistically analyzed. The results revealed that *atp9* exhibited the highest density of editing sites, with a rate of 57.78/kb, including 13 editing sites. In contrast, *rpl2* showed the lowest density, with a rate of 1.23/kb and only two editing sites (Fig. 7).

Analysis of selection pressure on the mitochondrial PCGs of *Neocinnamomum*

In the 24 individual mitogenomes of seven taxa of *Neocinnamomum*, seven PCGs without mutation sites were excluded, and the nucleotide substitution rates of the remaining 34 PCGs were calculated at the gene level. A total of 276 pairwise comparisons were made, with Ka values ranging from 0 to 0.00882. Among these comparisons, seven genes (*atp8*, *atp9*, *cox1*, *nad3*, *nad7*, *rps3*, and *rps14*) had Ka values of 0 (Additional file 2: Table S12). On the other hand, Ks values varied from 0 to 0.01379, and there were 11 genes with Ks values of 0 (Additional file 2: Table S13). The results for Ka and Ks were consistent with the mutation results, verifying their accuracy (Additional file 2: Table S9). In the calculation of Ka/Ks, it was found that in some cases, the Ka value was not 0, while the Ks value was 0, resulting in non calculable Ka/Ks ratio. For accuracy, these cases were excluded from the statistical analysis. Therefore, the Ka/Ks values in the 276 groups ranged from 0 to 1.32591, only 17 of 34 PCGs had numerical values, and only *ccmC* underwent positive selection, the rest were purified selection, and no neutral selected genes were found (Additional file 2: Table S14). One representative individual from each species was selected for interspecific comparison of the 17 PCGs (Fig. 8). The positive selection of the *ccmC* gene was not present between every pair, but was observed in the comparisons of *N. delavayi* (SY7790) vs. *N. caudatum* (SY57831), *N. caudatum* (SY57831) vs. *N. mekongense* (SY6057), and *N. caudatum* (SY57831) vs. *N. caudatum* var. *macrocarpum* (SY9634) (Fig. 8). The presence of *N. caudatum* (SY57831) in all three comparisons is noteworthy.

Discussion

Characteristics and comparison of mitogenomes

Plant mitogenomes are characterized by abundant repetitive sequences and intricate structures, making the assembly process particularly complicated, especially when using short reads. The present study has assembled the complete mitogenome of *N. delavayi* through the integration of Illumina and Oxford Nanopore sequencing reads, representing the first mitogenome sequence obtained for the genus *Neocinnamomum*. The *N. delavayi* mitogenome exhibits a typical master circular molecule consistent with many higher plants. It is 778,066 bp in length, which is 242,261 bp longer than the *H.*

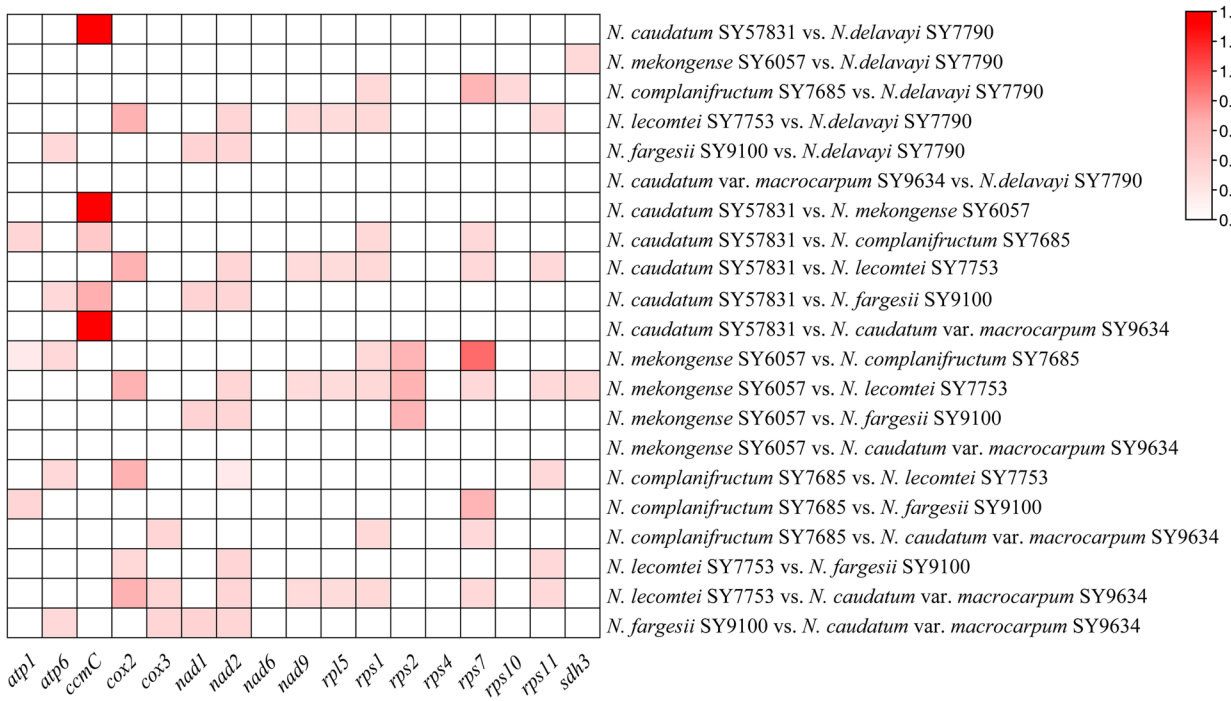

Fig. 8 Heatmap that shows the pairwise Ka/Ks ratios among 17 PCGs in the seven *Neocinnamomum* mitogenomes

nymphaeifolia mitogenome (535,805 bp), and 389,963 bp shorter than the mitogenome of *C. henryi* (1,168,029 bp). This demonstrates a significant variation in mitogenome size within the family and within the order Laurales. The variations in genome size among *N. delavayi*, *C. henryi*, and *H. nymphaeifolia* mainly stem from changes in the length of non-coding regions rather than the number of genes. As an example, the total length of 66 genes in the *N. delavayi* mitogenome is 41,429 bp, which is only 9 bp shorter than the combined length of 65 genes in *C. henryi* and 40 bp longer than the combined length of 83 genes in *H. nymphaeifolia*. The mitogenome of *N. delavayi* has 41 annotated PCGs, which is similar to *C. henryi* but distinct from *H. nymphaeifolia*. According to previous research, the mitogenome of the common ancestor of angiosperms had 41 PCGs, and these functional genes were generally stable and conserved [24]. The genome annotation revealed the conservation of functional genes in *Neocinnamomum*. Skippington et al. (2015) performed a comparative study of PCGs among the mitogenomes of 34 angiosperms that had been sequenced. They discovered that almost all species possessed 24 core PCGs, while the other 17 flexible genes varied across species [25]. There is a possibility that these 17 variable genes have been lost or transferred to the nuclear genome, although there have been no reports of such occurrences in the family Lauraceae. However, it is worth noting that gene migration is a common phenomenon throughout plant evolution [26].

The prevalence of repetitive sequences in plant mitogenomes is a crucial driver of their structural evolution, contributing to their large size, complex organization, and dynamic nature [27]. In the mitogenome of *N. delavayi*, dispersed repeat sequences accounted for 8.61% of the genome, while the lengths of dispersed repeats in *H. nymphaeifolia* and *C. henryi* were 65,068 bp (12.14%) and 260,206 bp (22.28%), respectively. The differences between them potentially lead to structural differences. For instance, *N. delavayi* and *H. nymphaeifolia* have similar proportions of repetitive sequences, and their structures are relatively simple, consisting of single circular or linear structures. In contrast, *C. henryi* has the highest proportion of repetition, and its structure is more complex, composed of a large linear contig and a small circular contig. Although repetitive sequences have a significant impact on genome size, a large number of non-repetitive sequences appear to contribute more. Additionally, the impact of DNA migration should not be overlooked, especially between mitochondrial and nuclear genomes.

The mitogenome of *N. delavayi* contains 33 pairs (5.01%) of dispersed repeats that are at least 100 bp long. These repetitive sequences are prevalent and active in recombination, and large repeats may originate from small repeats [28]. Wynn and Christensend's study suggests that the shortest repeats (less than 50 bp) rarely participate in homologous recombination events, while medium-length repeats, usually around 100 bp, undergo recombination in mutants with damaged genomes or DNA maintenance issues, but typically not in non-stressed, non-mutant plants [27]. Recombination is considered a key genetic factor contributing to the frequent reorganization of mitogenomes in angiosperms. This phenomenon has been extensively studied in *Arabidopsis*, and it is believed that the rearrangement is caused by the recombination of repetitive sequences [29, 30].

Upon comparative analysis of *N. delavayi*, *C. henryi*, and *H. nymphaeifolia*, numerous changes in gene order were observed. These changes reveal an interesting pattern: gene sequences exhibit higher conservation between closely related species, while being less conserved between distantly related species. Typically, species that are closely related tend to retain a greater number of syntenic gene blocks [31, 32]. For instance, the closer relationship between *N. delavayi* and *C. henryi*, as compared to *H. nymphaeifolia*, is reflected by the sharing of more conserved gene clusters. Recombination rates in seed plants vary by at least 600-fold, with some studies suggesting that environmental stress and nuclear gene variation may contribute to mitogenome rearrangements [33–35]. *Neocinnamomum* is distributed solely in Asia, while *Caryodaphnopsis*, as amphi-Pacific genus, is found in both Southeast Asia and South America, and *Hernandia* is widely distributed across tropical coastal regions [36]. We speculate that these diverse geographical distributions and environmental changes may be partial reasons for the observed differences in gene rearrangements among them, a hypothesis that has been validated in Rosaceae plants [37]. Moreover, given that mitochondria are more susceptible to oxidative damage than the nucleus, homologous recombination repair is considered a crucial pathway for mitochondrial genomic repair, which may also contribute to inter-genomic gene rearrangements and reduced collinearity [38, 39]. Currently, data on the mitogenomes of Lauraceae plants are relatively limited, necessitating the collection of additional suitable samples to further investigate the phenomenon of rearrangements within this family.

DNA transfer between organelles

Significant genetic material exchange takes place between the genomes of organelles and the nuclear genome [40]. This transfer of genetic material between organelles primarily occurs through intracellular gene transfer (IGT) and horizontal gene transfer (HGT) between species [41, 42]. The study explored the migration events of genetic material between the plastome and mitogenome of *N.*

delavayi, identifying 37 homologous fragments totaling 9929 bp. Due to the strong genetic conservation of the plastome, the presence of heterologous DNA within the plastome is rare, but the transfer of plastid DNA sequences to mitochondria occurs frequently [43, 44]. Therefore, the identified homologous fragments likely represent one-way transfer events from the plastid to mitochondria. However, this is not absolute, as there have been reports of mitochondrial gene expression products being transported to plastids [45]. These migration sequences are scattered throughout the genome and do not exhibit site specificity, possibly resulting from non-homologous or random recombination events. The function of plastid genes after transfer to mitochondria is still uncertain. In the mitogenome of *N. delavayi*, the identified plastid PCGs, such as *atpA*, *petG*, *psbC*, and *psbD*, have lost their coding functions. However, some tRNA genes, such as *trnD-GUC* and *trnM-CAU*, still retain their functions. The loss of gene functions can be attributed to three main reasons: (1) Genes with specialized functions are difficult to express after transfer; (2) The sequences controlling transcription and translation in mitogenome and plastome are different; and (3) Some plastid genes undergo changes after transfer, resulting in functional loss [17, 46]. While the amount of wandering DNA in the mitogenome is not large, it undeniably enhances its intricacy.

Phylogenetic inconsistencies

The reconstruction of phylogenetic relationships using mitogenome sequences has been demonstrated as a viable approach [47, 48]. In this study, we conducted the first exploration of the phylogenetic relationships among species of the genus *Neocinnamomum* by employing mitogenome fragments. The tree topologies derived from ML and BI methods were found to be consistent, with high support values, attesting to the reliability of the findings. A moderate support was observed for the *N. complanifructum* branch, which represents the sole difference between the concatenated tree and the coalescent tree. Such inconsistencies in phylogenetic relationships seem to be a frequent occurrence in genomic studies [49, 50]. The reasons for these inconsistencies are multifaceted, with research suggesting that concatenation methods may not adequately accommodate the heterogeneity of gene trees, potentially leading to inconsistencies in phylogenetic inferences [51]. By comparing the phylogenetic relationships of *Neocinnamomum* reconstructed from different genomic datasets, significant phylogenetic conflicts were observed. Similar inconsistencies have been reported in other angiosperms, such as Poales [52], Rosidae [53], and *Diplostephium* [54]. Different genome types may lead to distinct phylogenetic relationships due to their varying genetic mechanisms and evolutionary rates. For instance, plastid and mitochondrial genomes typically exhibit maternal inheritance patterns and evolve at a slower rate, while nrDNA shows biparental inheritance and evolves more rapidly. The three kinds of phylogenetic trees for *Neocinnamomum* exhibit three different patterns. The phylogenetic conflict between the mitochondrial and plastid genomes of *Neocinnamomum* cannot be solely attributed to differences in evolutionary rates; it may also result from plastid capture events or gene migration. Additionally, incomplete lineage sorting (ILS) and hybridization have also been recognized as causes of phylogenetic inconsistencies [55, 56]. The overlapping geographical distribution and flowering times of *Neocinnamomum* facilitate widespread hybridization and introgression. The inconsistent phylogenetic relationships reflect the complex evolutionary history among species within *Neocinnamomum*, which may include potential hybridization events and genetic material flows. This holds significant scientific implications for understanding the origin, evolution, and conservation of these species. Further reconstruction of species relationships within *Neocinnamomum* using single-copy or low-copy nuclear genes may be capable of unraveling these mysteries.

Adaptive evolution of *Neocinnamomum*

The nucleotide polymorphisms of 41 PCGs in 24 individuals of *Neocinnamomum* were notably low, with an average π value of 0.00081. In addition, it was observed that seven genes did not exhibit any mutation sites. This indicates that the mitochondrial genes of *Neocinnamomum* are highly conserved and undergo modest evolution, aligning with the conservation pattern observed in the majority of plant mitogenomes [57]. Genetic diversity heavily relies on individual mutations, so mutation sites within genes were calculated. The findings align with nucleotide polymorphism results, showing that genes with more mutation sites exhibit higher π values. It was found that the mutation events in the 41 PCGs of *Neocinnamomum* were characterized by very rare indel events, a highly unbalanced number of transition and transversion SNPs, and dominant non-synonymous substitutions. The low number of indels may indicate higher mitogenome stability or that indel mutations are functionally less adaptive, therefore eliminated by natural selection [58, 59]. Despite the notably slow sequence evolution of plant mitogenomes, there are significant variations in evolutionary rates among different lineages [60]. The plant mitogenomes possess a distinctive repair process called homologous recombination repair, which might potentially play a role in maintaining a low mutation rate [39, 61].

The Ka/Ks ratio is a tool for assessing the selection pressure and can indicate whether genes have been subject to selection pressures during evolution [62]. Out of the 41 PCGs of *Neocinnamomum*, the majority have a Ka/Ks ratio below 1, indicating high conservation and purifying selection. During the process of evolution, the mutation rate was intentionally maintained at a low level in order to avoid the build-up of detrimental mutations and ensure the stability of gene function. No genes exhibiting neutral selection were detected. However, a Ka/Ks ratio of almost 1 was observed for the *rps7* gene in the comparison between *N. mekongense* and *N. complanifructum*, suggesting the potential for neutral selection in certain cases. In addition, the gene *ccmC* involved in cytochrome *c* biogenesis experienced positive selection (Ka/Ks > 1), similar to the positive selection found in *Prunus anserina* [63], *Eichhornia crassipes* [64], and *Triticum aestivum* cv. Chinese Yumai [65]. The *ccmC* gene encodes a protein essential for mitochondrial cytochrome *c* biosynthesis, which is a key component of the respiratory chain responsible for electron transport in mitochondria and is crucial for proper mitochondrial function and cell survival [66]. Specific variants of this gene may enable plants to better withstand adverse conditions such as high temperatures, drought, or low nitrogen, increasing their likelihood of survival in natural selection.

RNA editing is a process that occurs after transcription, when nucleotides are inserted, deleted, or substituted in the mRNA, resulting in changes to the genetic code [67]. This study identified 748 potential editing sites in the 41 PCGs of the *N. delavayi* mitogenome, a number less than that found in the ancient plant *Liriodendron tulipifera* [68]. The distribution of editing sites varied across various codon positions and genes among the 41 PCGs, which seems to be a typical event in the mitogenomes of higher plants. The majority of the potential editing sites were located in the first two codons, with a concentration on the second position, corresponding to the pattern seen in other plant species [69]. RNA editing in the first two codons mostly results in non-synonymous modifications. These alterations often enhance the hydrophobicity of the protein, which is directly associated with the function and stability of the protein [70, 71]. The occurrence of editing sites is limited at the third codon position, leading to alterations to the synonymous codons. Although non-synonymous editing in mitochondrial genes appears to be preserved by natural selection, there is no clear evidence of synonymous editing being maintained by natural selection [70, 72, 73]. A correlation analysis revealed a moderate relationship between gene length and the number of editing sites ($R = 0.61$, $p = 2.2e{-}05$) (Additional file 1: Figure S3). Previous researches have indicated a correlation between the number of editing sites and the rate of gene evolution, with rapidly evolving genes tending to have fewer editing sites [70, 74, 75]. This study performed an initial examination of RNA editing. However, additional investigation is necessary due to the following reasons: (1) The editing sites were only predicted using the database, which may have specific limitations; and (2) There are some particular dynamic editing sites during plant growth and development.

Impact of cytoplasmic effects on seed oil content in *Neocinnamomum*

Cytoplasmic effects (CEs) have been demonstrated to influence a diverse array of agronomic traits across a wide range of crops, including yield, height, leaf width, and protein quality in monocots such as rice and maize, as well as yield, fresh weight, dry weight, protein content, and seed oil content (SOC) in dicots like soybean and rapeseed [76–81]. SOC holds significant agricultural, nutritional, and economic importance. Variability in SOC exists among different species of *Neocinnamomum*, with mutations and positively selected genes within their mitogenomes potentially exerting direct or indirect influences on SOC. As observed in oilseed rape, the mitochondrial-encoded *orf18* significantly increases SOC by enhancing ATP production [82]. The abundance of potential RNA editing sites in *N. delavayi* mitogenome could regulate mitochondrial gene expression post-transcriptionally, thereby affecting protein synthesis and function, including key enzymes involved in fatty acid synthesis and metabolism, thus influencing SOC. This study has laid a solid foundation for further exploration, which can now employ techniques such as gene chips and RNA-Seq to functionally validate the mutations and positively selected genes within the mitogenomes of *Neocinnamomum*. By considering the impact of CEs on SOC, breeding strategies can be optimized. For instance, selecting parent plants with high SOC potential for hybridization can enhance the SOC in the offspring.

Conclusion

This study successfully assembled the first mitogenome of *Neocinnamomum*, *N. delavayi*, using reads from Illumina and Oxford Nanopore sequencing technologies. The mitogenome has a size of 778,066 bp and presents a single circular structure. Although its size differs from that of other known mitogenomes in the family Lauraceae, the GC content and gene number remain relatively stable. The research found frequent genetic information exchange among the organelles of *N. delavayi*. We reconstructed the phylogenetic relationships of *Neocinnamomum* based on 41 PCGs, dividing 24 individuals from seven taxa into six and five clades using concatenation and coalescence methods, respectively. However, these

results were inconsistent with those constructed using complete plastomes and nrDNA sequences. Selection pressure analysis revealed that the *ccmC* gene was under positive selection, suggesting its potential critical role in plant adaptation to the environment. In addition, the study involved analyses of repetitive sequences, codon usage, collinearity, nucleotide polymorphism, mutations in PCGs, as well as the prediction of RNA editing sites. In summary, this research has deepened our understanding of the mitogenome in the genus *Neocinnamomum* and provided valuable data and a foundation for genomic evolution research, genetic resource conservation, and molecular breeding.

Methods

Samples and DNA sequencing

Leaves from 24 individuals of seven *Neocinnamomum* species, either fresh or dried using silica gel, were gathered, and the plant materials were identified by Professor Song Yu. Voucher specimens for these 24 individuals were deposited at the Herbarium of Guangxi Normal University (GXNU), Guangxi, China (Table 1). Fruits of four *Neocinnamomum* species are shown in Fig. 9. Genomic DNA extraction from both fresh and silica-dried leaves was performed using the CTAB method. The high-quality DNA was fragmented using ultrasound for second-generation sequencing. Subsequently, it was purified, repaired at the ends, and joined to adaptors to create short-insert (500 bp) libraries, following the instructions provided by the manufacturer (Illumina). At the Kunming Institute of Botany, China, the DNA samples were subjected to sequencing using a Genome Analyzer (Illumina HiSeq 2500). The sequencing process, which followed the tagging and pooling of DNA samples into a single lane, generated more than 4.0 Gb of raw data for each sample.

The mitogenome of *N. delavayi* was obtained using second- and third-generation hybrid assembly methods. Thus, both second-generation and third-generation sequencing of *N. delavayi* were performed. For third-generation sequencing, the purified large fragments of DNA were ligated to adaptor using the SQK-LSK109 kit from Oxford Nanopore Technology to construct the library. The DNA sequencing process was conducted

Table 1 Collected species of *Neocinnamomum*

No	Species	Herbarium	Voucher	Geographic origin	Sequencing no
1	*N. caudatum*	GXNU	SY36779	Yunnan, China	SY57831
2	*N. caudatum*	GXNU	SY98601	Yunnan, China	SY9065
3	*N. caudatum*	GXNU	SY37109	Guangxi, China	SY6104
4	*N. caudatum*	GXNU	SY36812	Yunnan, China	SY57874
5	*N. complanifructum*	GXNU	SY36737	Yunnan, China	SY7685
6	*N. complanifructum*	GXNU	SY33249	Yunnan, China	SY9581
7	*N. delavayi*	GXNU	SY37088	Guangxi, China	SY6083
8	*N. delavayi*	GXNU	SY36746	Sichuan, China	SY7790
9	*N. delavayi*	GXNU	SY35377	Sichuan, China	SY9763
10	*N. delavayi*	GXNU	YangQE2905	Yunnan, China	SY5832
11	*N. fargesii*	GXNU	SY34404	Chongqing, China	SY9100
12	*N. fargesii*	GXNU	SY34386	Chongqing, China	SY9101
13	*N. fargesii*	GXNU	SY34386	Chongqing, China	SY9494
14	*N. fargesii*	GXNU	SY34404	Chongqing, China	SY9850
15	*N. caudatum* var. *macrocarpum*	GXNU	D097	Guangxi, China	SY9634
16	*N. caudatum* var. *macrocarpum*	GXNU	D098	Guangxi, China	SY9635
17	*N. mekongense*	GXNU	SY34909	Yunnan, China	SY7778
18	*N. mekongense*	GXNU	SY34907	Yunnan, China	SY7782
19	*N. mekongense*	GXNU	SY34538	Yunnan, China	SY7781
20	*N. mekongense*	GXNU	SY37047	Yunnan, China	SY6057
21	*N. lecomtei*	GXNU	SY36768	Hainan, China	SY57821
22	*N. lecomtei*	GXNU	XY25	Hainan, China	SY7753
23	*N. lecomtei*	GXNU	SY34485	Hainan, China	SY9528
24	*N. lecomtei*	GXNU	SY25	Hainan, China	SY7796

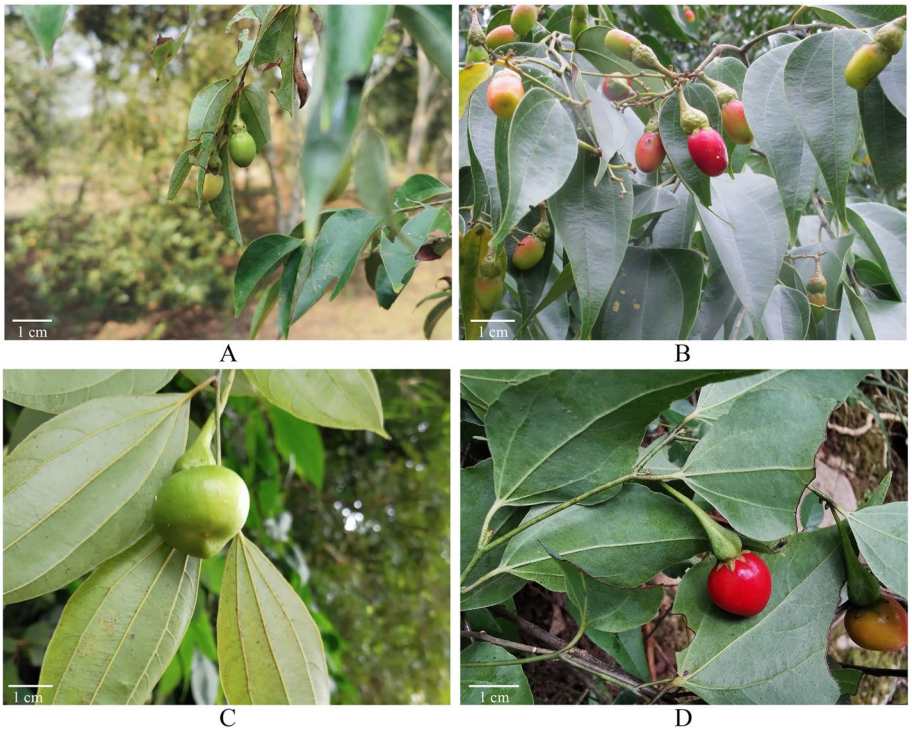

Fig. 9 Fruits of four *Neocinnamomum* species (**A** *N. mekongense*; **B** *N. caudatum*; **C** *N. complanifructum*; **D** *N. fargesii*)

utilizing the Oxford Nanopore sequencing method on the promethION platform at the Grandomics company. This process yielded 21.7 Gb of raw data with an average read length of 27,555 bp.

Mitogenome assembly and annotation

The mitogenomes of *Neocinnamomum* samples were initially assembled via GetOrganelle v1.7.5 [83]. In order to resolve the repeats-tangled graph assembled by short reads alone, long reads sequenced for N. delavayi were utilized to assemble a complete mitogenome. The long reads obtained from Nanopore third-generation sequencing were first filtered employing Porechop v0.2.4 [84]. The clean reads were aligned with the scaffolds assembled from the second-generation data by BLAST 2.9.0 + [85] with an -evalue parameter of 1e-200, and the subset of long sequences that are similar to the mitochondrial sequences was acquired. Subsequently, the long reads obtained from third-generation sequencing and the extended short reads obtained from second-generation sequencing were used together for hybrid assembly. This assembly was performed by the Unicycler v0.4.8 pipeline [86], which also includes a polishing phase by Pilon v1.24 [87], in order to correct remaining errors in sequences by mapping short reads back to the hybrid assembly. In the final assembly, the mitochondrial sequence of *N. delavayi* was obtained.

The mitogenome of *N. delavayi* was initially annotated by GeSeq v2.03 [88]. The annotation was then adjusted by comparing it to the mitogenome annotations of closely related species from NCBI employing PhyloSuite v1.2.2 [89]. The tRNA boundaries were corroborated by applying tRNAscan-SE [90] results from GeSeq. The genome map of *N. delavayi* was drawn by OGDRAW [91]. The genome sequence has been submitted to the National Center for Biotechnology Information (NCBI, accession number: PQ443074).

Repetitive sequence analysis and codon usage bias

To identify potential SSRs, tandem repeats, and dispersed repeats, the MISA [92], the Tandem Repeats Finder v4.09 [93], and the REPuter [94] were utilized, respectively. For SSRs, the minimum repeat threshold was set at 10 for mononucleotides, 5 for dinucleotides, 4 for trinucleotides, 3 for tetranucleotides, pentanucleotides, and hexanucleotides. Tandem repeats were detected using default settings, while for dispersed repeats, parameters were set as follows: Maximum Computed Repeats = 5000, Minimal Repeat Size = 30, and Hamming Distance = 3. Tandem repeats were detected using the default settings. The parameters for dispersed repeats were set as follows: Maximum Computed Repeats = 5000, Minimal Repeat Size = 30, and Hamming Distance = 3. The results were

visualized using Circos v0.69.9 [95] and Origin v2022 (OriginLab Inc., Massachusetts, USA).

PCGs within the *N. delavayi* mitogenome were isolated using PhyloSuite v1.2.3 [89] and examined. The software CodonW v1.4.2 and the web tool CUSP [96] were used to analyze codon usage preference parameters. These parameters include ENC, RSCU, and GC content (GC overall, GC1, GC2, GC3, GC3s). The GC plot was visualized using TBtools v2.096 [97], and the scatter plot of GC12 and GC3 was constructed using Microsoft Excel v2021.

Identification of mitochondrial plastid DNAs (MTPTs) and synteny analysis

In order to discover probable homologous sequences that were transferred between the plastome and mitogenome of *N. delavayi*, BLASTN 2.12.0 + [98] was used to compare the two organelles with the parameter E-value set to 1e-5. Geneious Prime v2023.2.1 [99] was used to locate homologous fragments in previously annotated organelle genomes for annotation and examination, and the Circos [95] module in TBtools v2.096 [97] was used for visualization.

In addition, the genetic sequences of *H. nymphaeifolia* (accession number: ON023262) and *C. henryi* (accession number: OR987149), which is closely related to *N. delavayi*, [36] were acquired from NCBI, and analysis of mitogenome collinearity was performed by Mauve v20150226 [100]. The gene rearrangement results were visualized on the RAWGraphs 2.0 website [101].

Phylogenetic analysis

The sequences of mitogenome PCGs from 24 individuals representing seven taxa of *Neocinnamomum* and the outgroup *C. henryi* were extracted using BLASTN [85] and mview v1.67 [102], with *N. delavayi* (SY7790) as the reference. The PCG sequences were aligned using MAFFT v7.520 [103] and trimmed employing trimAl v1.2 [104]. Subsequently, the concatenated PCGs were used to rebuild the phylogenetic trees of *Neocinnamomum*. The optimal DNA substitution models were assessed for ML and BI approaches using ModelFinder v2.2.2 [105]. IQ-TREE v2.2.2.6 [106] and MrBayes v3.2.6 [107] were employed for ML and BI analyses, with models of "TVM + F + I" and "GTR + F + I", running 5,000 bootstrap replicates and 20 million MCMC generations, respectively. ASTRAL analysis was conducted in IQ-TREE and collapsed branches with < 10% bootstrap support in ASTRAL-III v5.7.8 [108], resulting in the final phylogeny. FigTree v1.4.4 was used for tree visualization and modification.

Analysis of nucleotide polymorphism (π), mutation events, and RNA editing site identification

The sequences of 41 PCGs from the mitogenomes of 24 individuals representing seven taxa of *Neocinnamomum* were manually examined. DnaSP v6.12.03 [109] was then employed to compute the nucleotide polymorphism (π) values and mutation events of each PCG. Mutation events were manually verified to ensure accuracy. The Deepred-Mt tool [110] has been employed to predict RNA editing sites in the 41 PCGs of *N. delavayi*. The prediction threshold value of 0.95 was chosen. An R script was used to perform a correlation test between gene length and the amount of RNA editing sites, with 'Spearman' selected as the test method.

Selective pressure analysis

Selective pressure analysis on the mitogenomes of 24 individuals from seven *Neocinnamomum* taxa was performed by calculating Ka, Ks, and Ka/Ks ratios for 34 PCGs (excluding seven PCGs without any mutations). Stop codons were removed, and two methods were applied to ensure precision. KaKs_Calculator v2.0 [111] and the Nei-Gojobori model [112] in MEGA v11.0.13 [113] were utilized. A Ka/Ks ratio < 1 indicated purifying selection, a ratio of 1 signified neutral evolution, and a ratio > 1 suggested positive selection [62]. Genes with a Ka/Ks ratio of 0 were excluded, and the remaining Ka/Ks ratios were visualized with a heat map in TBtools v2.096 [97].

Abbreviations

MCFA	Medium-chain fatty acid
LCFA	Long-chain fatty acid
SSRs	Simple sequence repeats
PCGs	Protein-coding genes
ENC	Effective number of codons
RSCU	Relative synonymous codon usage
ML	Maximum likelihood
BI	Bayesian inference
MTPTs	Mitochondrial plastid DNAs
LSC	Large single copy
IR	Inverted repeat
SSC	Small single copy
Indels	Insertions and deletions
SNPs	Single-nucleotide polymorphisms
nrDNA	Nuclear ribosomal DNA
ILS	Incomplete lineage sorting
Ka	Nonsynonymous substitution rate
Ks	Synonymous substitution rate
Ka/Ks	Non-synonymous/Synonymous mutation ratio

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06238-x.

Additional file 1: Figure S1. ENC-GC3 plot analysis of 41 PCGs in *N. delavayi*. Figure S2. Collinear analysis among *C. henryi*, *H. nymphaeifolia*, and *N.*

delavayi. Figure S3. Correlation analysis of number and length of gene RNA editing sites.

Additional file 2: Table S1. Gene composition in the mitogenome of *N. delavayi*. Table S2. Mitochondrial gene length of *N. delavayi*. Table S3. Dispersed repeat sequences in the mitogenome of *N. delavayi*. Table S4. SSRs in the mitogenome of *N. delavayi*. Table S5. Tandem repeat sequences in the mitogenome of *N. delavayi*. Table S6. Codon usage bias of mitochondrial PCGs of *N. delavayi*. Table S7. GC content of different positions of codon in the mitogenome of *N. delavayi*. Table S8. Homologous segments between organelles of *N. delavayi*. Table S9. The number and type of SNPs in mitochondrial PCGs of *Neocinnamomum*. Table S10. The mutation statistics of 24 individuals of *Neocinnamomum*. Table S11. Predicted RNA editing sites in 41 PCGs. Table S12. The Ka values for pairwise comparisons of 24 individuals of the genus *Neocinnamomum*. Table S13. The Ks values for pairwise comparisons of 24 individuals of the genus *Neocinnamomum*. Table S14. The Ka/Ks values for pairwise comparisons of 24 individuals of the genus *Neocinnamomum*.

Additional file 3: The matrix of combined mt genes sequences.

Acknowledgements
We express our gratitude to the journal editors and anonymous reviewers.

Authors'contributions
W.Z., Y.S., and P.Y.X. designed the research. D.Z., J.P.H., and Y.Y.L. contributed to the data analysis. W.B.X., Y.G., and Y.H.T. provided materials. W.Z. wrote the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by Yunnan Province landscape architecture first-class discipline construction fund; First-rate (A) Discipline Landscape Architecture Construction Funding of Yunnan Province, China; Key Technologies Research for the Germplasm of Important Woody Flowers in Yunnan Province (No. 202302AE090018); Special Program for Technology Bases and Talents of Guangxi (No. GuikeAD23026281); the National Natural Science Foundation of China (No. 32260060); the Project of the Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences (No. Y4ZK111B01) and Funding from Yunnan Province Science and Technology Department (No. 202203AP140007).

Data availability
The matrix of combined mt genes sequences data can be found in Additional file 3. The assembled mitochondrial genome sequence was submitted to the National Center for Biotechnology Information (NCBI) with accession number PQ443074 (https://www.ncbi.nlm.nih.gov/).

Declarations

Ethics approval and consent to participate
This study's material collections and experimental research complied with relevant institutional, national, and international guidelines and legislation. No specific permissions or licenses were required.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Engineering Technology Research Center of National Forestry and Grassland Administration on Southwest Landscape Architecture, Southwest Forestry University, Kunming, Yunnan 650224, China. [2]Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences & Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China. [3]Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China. [4]College of Advanced Agricultural Science Zhejiang A&F University, Hangzhou, Zhejiang 311300, China. [5]Key Laboratory of Ecology of Rare and Endangered Species and Environmental Protection (Ministry of Education) & Guangxi Key Laboratory of Landscape Resources Conservation and Sustainable Utilization in Lijiang River Basin, Guangxi Normal University, Guilin, Guangxi 541004, China.

Received: 7 October 2024 Accepted: 11 February 2025
Published online: 06 March 2025

References

1. WFO (2024): *Neocinnamomum* H.Liu. http://www.worldfloraonline.org/taxon/wfo-4000025650. Accessed on: 26 July 2024
2. Li XW LJ, Huang PH, Wei FN, Cui HB, van der Werff H: Lauraceae. In: Wu ZY, Raven PH, Hong DY (Eds) Flora of China. vol. 7: Beijing: Science Press; St. Louis: Missouri Botanical Garden Press.; 2008:102–254.
3. Jena S, Ray A, Mohanta O, Das PK, Sahoo A, Nayak S, Panda PC: *Neocinnamomum caudatum* Essential Oil Ameliorates Lipopolysaccharide-Induced Inflammation and Oxidative Stress in RAW 264.7 Cells by Inhibiting NF-kappaB Activation and ROS Production. Molecules. 2022;27(23):8193.
4. Wang XD, Xu CY, Zheng YJ, Wu YF, Zhang YT, Zhang T, Xiong ZY, Yang HK, Li J, Fu C, et al. Chromosome-level genome assembly and resequencing of camphor tree (*Cinnamomum camphora*) provides insight into phylogeny and diversification of terpenoid and triglyceride biosynthesis of *Cinnamomum*. Hortic Res. 2022;9:uhac216.
5. Gan Y, Song Y, Chen Y, Liu H, Yang D, Xu Q, Zheng Z. Transcriptome analysis reveals a composite molecular map linked to unique seed oil profile of *Neocinnamomum caudatum* (Nees) Merr. BMC Plant Biol. 2018;18(1):303.
6. Thomas RH. Molecular Evolution and Phylogenetics. Heredity. 2001;86(3):385–385.
7. Liou H: *Lauracées de Chine et d'Indochine: contribution à l'étude systématique et phytogéographique*: Hermann & Cie; 1934.
8. Kostermans AJ. A monograph of the genus *Neocinnamomum* Liou Ho. Reinwardtia. 1974;9(1):85–96.
9. Chanderbali AS, van der Werff H, Renner SS. Phylogeny and Historical Biogeography of Lauraceae: Evidence from the Chloroplast and Nuclear Genomes. Ann Missouri Bot Gard. 2001;88(1):104–34.
10. Rohwer JG, Rudolph B. Jumping Genera: The Phylogenetic Positions of Cassytha, Hypodaphnis, and *Neocinnamomum* (Lauraceae) Based on Different Analyses of *trnK* Intron Sequences. Ann Mo Bot Gard. 2005;92(2):153–78.
11. Wang Zh, Li J, Conran JG, Li HW. Phylogeny of the Southeast Asian endemic genus *Neocinnamomum* H. Liu (Lauraceae). Plant Systematics Evol. 2010;290(1–4):173–84.
12. Song Y, Yu WB, Tan Y, Liu B, Yao X, Jin J, Padmanaba M, Yang JB, Corlett RT. Evolutionary Comparisons of the Chloroplast Genome in Lauraceae and Insights into Loss Events in the Magnoliids. Genome Biol Evol. 2017;9(9):2354–64.
13. Song Y, Yu WB, Tan YH, Jin JJ, Wang B, Yang JB, Liu B, Corlett RT. Plastid phylogenomics improve phylogenetic resolution in the Lauraceae. J Syst Evol. 2019;58(4):423–39.
14. Ren S, Song Y, Zhao M, Xu W. The plastid genome sequence of *Neocinnamomum delavayi* (Lec.) Liou. Mitochondrial DNA B Resour. 2019;4(2):3711–2.
15. Cao Z, Yang L, Xin Y, Xu W, Li Q, Zhang H, Tu Y, Song Y, Xin P. Comparative and phylogenetic analysis of complete chloroplast genomes from seven *Neocinnamomum* taxa (Lauraceae). Front Plant Sci. 2023;14:1205051.
16. Li Q, Yang L, Yu Q, Xu W, Xin Y, Song Y, Xin P. Nuclear DNA-based phylogenetic analysis of *Neocinnamomum* species. Plant Genetic Resources: Characterization and Utilization. 2023;21(4):323–30.
17. Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. Proc Natl Acad Sci U S A. 2004;101(51):17747–52.
18. Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol. 2010;27(6):1436–48.

19. Tang M, Chen Z, Grover CE, Wang Y, Li S, Liu G, Ma Z, Wendel JF, Hua J. Rapid evolutionary divergence of *Gossypium barbadense* and *G. hirsutum* mitochondrial genomes. BMC Genomics. 2015;16:770.
20. Mackenzie S, McIntosh L. Higher plant mitochondria. Plant Cell. 1999;11(4):571–86.
21. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 2003;424(6945):197–201.
22. Takenaka M, Verbitskiy D, van der Merwe JA, Zehrmann A, Brennicke A. The process of RNA editing in plant mitochondria. Mitochondrion. 2008;8(1):35–46.
23. Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc Natl Acad Sci U S A. 2000;97(13):6960–6.
24. Mower JP, Sloan DB, Alverson AJ. Plant Mitochondrial Genome Diversity: The Genomics Revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ, editors. Plant Genome Diversity, vol. 1. Vienna: Springer Vienna; 2012. p. 123–44.
25. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. Proc Natl Acad Sci U S A. 2015;112(27):E3515-3524.
26. Adams KL, Qiu YL, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci U S A. 2002;99(15):9905–12.
27. Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. G3 (Bethesda). 2019;9(2):549–59.
28. Andre C, Levy A, Walbot V. Small repeated sequences and the structure of plant mitochondrial genomes. Trends Genet. 1992;8(4):128–32.
29. Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA. Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. Plant Cell. 2007;19(4):1251–64.
30. Arrieta-Montiel MP, Shedge V, Davila J, Christensen AC, Mackenzie SA. Diversity of the *Arabidopsis* mitochondrial genome occurs via nuclear-controlled recombination activity. Genetics. 2009;183(4):1261–8.
31. Li Y, Liu H, Steenwyk JL, LaBella AL, Harrison MC, Groenewald M, Zhou X, Shen XX, Zhao T, Hittinger CT, Rokas A. Contrasting modes of macro and microsynteny evolution in a eukaryotic subphylum. Curr Biol. 2022;32(24):5335–43.
32. Huang Y, Wang H, Huo S, Lu J, Norvienyeku J, Miao W, Qin C, Liu W. Comparative Mitogenomics Analysis Revealed Evolutionary Divergence among Neopestalotiopsis Species Complex (Fungi: *Xylariales*). Int J Mol Sci. 2024;25(6):3093.
33. Cole LW, Guo W, Mower JP, Palmer JD. High and Variable Rates of Repeat-Mediated Mitochondrial Genome Rearrangement in a Genus of Plants. Mol Biol Evol. 2018;35(11):2773–85.
34. Shedge V, Davila J, Arrieta-Montiel MP, Mohammed S, Mackenzie SA. Extensive rearrangement of the *Arabidopsis* mitochondrial genome elicits cellular conditions for thermotolerance. Plant Physiol. 2010;152(4):1960–70.
35. Virdi KS, Wamboldt Y, Kundariya H, Laurie JD, Keren I, Kumar KRS, Block A, Basset G, Luebker S, Elowsky C, et al. MSH1 Is a Plant Organellar DNA Binding and Thylakoid Protein under Precise Spatial Regulation to Alter Development. Mol Plant. 2016;9(2):245–60.
36. Yang S, Huang J, Qu Y, Zhang D, Tan Y, Wen S, Song Y. Phylogenetic incongruence in an Asiatic species complex of the genus *Caryodaphnopsis* (Lauraceae). BMC Plant Biol. 2024;24(1):616.
37. Sun M, Zhang M, Chen X, Liu Y, Liu B, Li J, Wang R, Zhao K, Wu J. Rearrangement and domestication as drivers of Rosaceae mitogenome plasticity. BMC Biol. 2022;20(1):181.
38. Yakes FM, Van Houten B. Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. Proc Natl Acad Sci U S A. 1997;94(2):514–9.
39. Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. DNA repair and the stability of the plant mitochondrial genome. Int J Mol Sci. 2020;21(1):328.
40. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 2004;5(2):123–35.
41. Filip E, Skuza L. Horizontal gene transfer involving chloroplasts. Int J Mol Sci. 2021;22(9):4484.
42. Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. Nat Rev Genet. 2015;16(8):472–82.
43. Thorsness PE, Weber ER. Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. Int Rev Cytol. 1996;165:207–34.
44. Gandini CL, Sanchez-Puerta MV. Foreign Plastid Sequences in Plant Mitochondria are Frequently Acquired Via Mitochondrion-to-Mitochondrion Horizontal Transfer. Sci Rep. 2017;7:43402.
45. Gallois JL, Achard P, Green G, Mache R. The *Arabidopsis* chloroplast ribosomal protein L21 is encoded by a nuclear gene of mitochondrial origin. Gene. 2001;274(1–2):179–85.
46. Straub SC, Cronn RC, Edwards C, Fishbein M, Liston A. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (apocynaceae). Genome Biol Evol. 2013;5(10):1872–85.
47. Hu H, Sun P, Yang Y, Ma J, Liu J. Genome-scale angiosperm phylogenies based on nuclear, plastome, and mitochondrial datasets. J Integr Plant Biol. 2023;65(6):1479–89.
48. Song Y, Yu QF, Zhang D, Chen LG, Tan YH, Zhu W, Su HL, Yao X, Liu C, Corlett RT. New insights into the phylogenetic relationships within the Lauraceae from mitogenomes. BMC Biol. 2024;22(1):241.
49. Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol. 2009;24(6):332–40.
50. Krak K, Caklova P, Chrtek J, Fehrer J. Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (*Hieracium*, Asteraceae). Heredity (Edinb). 2013;110(2):138–51.
51. Song S, Liu L, Edwards SV, Wu S. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc Natl Acad Sci U S A. 2012;109(37):14942–7.
52. Wu H, Yang JB, Liu JX, Li DZ, Ma PF. Organelle Phylogenomics and Extensive Conflicting Phylogenetic Signals in the Monocot Order Poales. Front Plant Sci. 2021;12:824672.
53. Sun M, Soltis DE, Soltis PS, Zhu X, Burleigh JG, Chen Z. Deep phylogenetic incongruence in the angiosperm clade Rosidae. Mol Phylogenet Evol. 2015;83:156–66.
54. Vargas OM, Ortiz EM, Simpson BB. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). New Phytol. 2017;214(4):1736–50.
55. Rieseberg LH, Soltis D. Phylogenetic consequences of cytoplasmic gene flow in plants. 1991:65–84.
56. Maddison WP, Wiens JJ. Gene Trees in Species Trees. Syst Biol. 1997;46(3):523–36.
57. Palmer JD, Herbon LA. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol. 1988;28(1–2):87–97.
58. Garcia-Diaz M, Kunkel TA. Mechanism of a genetic glissando: structural biology of indel mutations. Trends Biochem Sci. 2006;31(4):206–14.
59. Leushkin EV, Bazykin GA. Short indels are subject to insertion-biased gene conversion. Evolution. 2013;67(9):2604–13.
60. Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol. 2007;7:135.
61. Boesch P, Weber-Lotfi F, Ibrahim N, Tarasenko V, Cosset A, Paulus F, Lightowlers RN, Dietrich A. DNA repair in organelles: Pathways, organization, regulation, relevance in disease and aging. Biochim Biophys Acta. 2011;1813(1):186–200.
62. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18(9):486.
63. Liu Q, Wu Z, Tian C, Yang Y, Liu L, Feng Y, Li Z. Complete mitochondrial genome of the endangered *Prunus pedunculata* (Prunoideae, Rosaceae) in China: characterization and phylogenetic analysis. Front Plant Sci. 2023;14:1266797.

64. He X, Qian Z, Gichira AW, Chen J, Li Z. Assembly and comparative analysis of the first complete mitochondrial genome of the invasive water hyacinth. Eichhornia crassipes Gene. 2024;914:148416.
65. Cui P, Liu H, Lin Q, Ding F, Zhuo G, Hu S, Liu D, Yang W, Zhan K, Zhang A, Yu J. A complete mitochondrial genome of wheat (*Triticum aestivum* cv. Chinese Yumai), and fast evolving mitochondrial genes in higher plants. J Genet. 2009;88(3):299–307.
66. Kitazaki K, Nomoto Y, Aoshima A, Mikami T, Kubo T. A mitochondrial gene involved in cytochrome *c* maturation (*ccmC*) is expressed as a precursor with a long NH2-terminal extension in sugar beet. J Plant Physiol. 2009;166(7):775–80.
67. Gott JM, Emeson RB. Functions and mechanisms of RNA editing. Annu Rev Genet. 2000;34:499–531.
68. Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 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.
69. Zheng P, Wang D, Huang Y, Chen H, Du H, Tu J. Detection and analysis of C-to-U RNA editing in rice mitochondria-encoded ORFs. Plants (Basel). 2020;9(10):1277.
70. Cuenca A, Petersen G, Seberg O, Davis JI, Stevenson DW. Are substitution rates and RNA editing correlated? BMC Evol Biol. 2010;10:349.
71. Jobson RW, Qiu YL. Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? Biol Direct. 2008;3:43.
72. Edera AA, Gandini CL, Sanchez-Puerta MV. Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. Plant Mol Biol. 2018;97(3):215–31.
73. Mower JP. Modeling sites of RNA editing as a fifth nucleotide state reveals progressive loss of edited sites from angiosperm mitochondria. Mol Biol Evol. 2008;25(1):52–61.
74. Parkinson CL, Mower JP, Qiu YL, Shirk AJ, Song K, Young ND, DePamphilis CW, Palmer JD. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol Biol. 2005;5:73.
75. Cho Y, Mower JP, Qiu YL, Palmer JD. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc Natl Acad Sci U S A. 2004;101(51):17741–6.
76. Tao D, Hu F, Yang J, Yang G, Yang Y, Xu P, Li J, Ye C, Dai L. Cytoplasm and cytoplasm-nucleus interactions affect agronomic traits in japonica rice. Euphytica. 2004;135(1):129–34.
77. Tang Z, Yang Z, Hu Z, Zhang D, Lu X, Jia B, Deng D, Xu C. Cytonuclear epistatic quantitative trait locus mapping for plant height and ear height in maize. Mol Breeding. 2012;31(1):1–14.
78. Allen JO. Effect of teosinte cytoplasmic genomes on maize phenotype. Genetics. 2005;169(2):863–80.
79. Shi CH, Zhu J, Wu JG, Yang XE, Yu YG. Analysis of embryo, endosperm, cytoplasmic and maternal effects for heterosis of protein and lysine content in indica hybrid rice. Plant Breeding. 2008;118(6):574–6.
80. Azizinia S. Combining ability analysis of yield component parameters in winter rapeseed genotypes (*Brassica napus* L.). J Agric Sci. 2012;4(4):87.
81. Liang HZ, Wang SF, Wang TF, Zhang HY, Zhao SJ, Zhang MC. Genetic analysis of embryo, cytoplasm and maternal effects and their environment Interactions for isoflavone content in soybean [*Glycine max* (L.) Merr.]. Agric Sci China. 2007;6(9):1051–9.
82. Liu J, Hao W, Liu J, Fan S, Zhao W, Deng L, Wang X, Hu Z, Hua W, Wang H. A Novel Chimeric Mitochondrial Gene Confers Cytoplasmic Effects on Seed Oil Content in Polyploid Rapeseed (*Brassica napus*). Mol Plant. 2019;12(4):582–96.
83. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21(1):241.
84. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom. 2017;3(10):e000132.
85. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
86. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.
87. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9(11):e112963.
88. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
89. Zhang D, Gao F, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.
90. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021;49(16):9077–96.
91. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–64.
92. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.
93. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
94. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
95. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.
96. EMBOSS. http://imed.med.ucm.es/Tools/EMBOSS/. Accessed 25 May 2024.
97. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y, Xia R. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Mol Plant. 2023;16(11):1733–42.
98. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol. 2000;7(1–2):203–14.
99. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
100. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394–403.
101. RAWGraphs 2.0. https://app.rawgraphs.io/. Accessed 22 June 2024.
102. Brown NP, Leroy C, Sander C. MView: a web-compatible database search or multiple alignment viewer. Bioinformatics. 1998;14(4):380–1.
103. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
104. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.
105. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
106. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
107. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.
108. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 2018;19(Suppl 6):153.
109. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol. 2017;34(12):3299–302.
110. Edera AA, Small I, Milone DH, Sanchez-Puerta MV. Deepred-Mt: Deep representation learning for predicting C-to-U RNA editing in plant mitochondria. Comput Biol Med. 2021;136:104682.

111. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics. 2010;8(1):77–80.
112. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986;3(5):418–26.
113. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38(7):3022–7.

Publisher's Note