

RESEARCH

Open Access



Phylogenomic and comparative analyses of Coffeae alliance (Rubiaceae): deep insights into phylogenetic relationships and plastome evolution

Sara Getachew Amenu^{1,2†}, Neng Wei^{1,2†}, Lei Wu³, Oyetola Oyebanji^{2,4,5}, Guangwan Hu^{1,6,7},
Yadong Zhou^{1,6,7*} and Qingfeng Wang^{1,6,7*}

Abstract

Background: The large and diverse Coffeae alliance clade of subfamily Ixoroideae (Rubiaceae) consists of 10 tribes, > 90 genera, and > 2000 species. Previous molecular phylogenetics using limited numbers of markers were often unable to fully resolve the phylogenetic relationships at tribal and generic levels. Also, the structural variations of plastomes (PSVs) within the Coffeae alliance tribes have been poorly investigated in previous studies. To fully understand the phylogenetic relationships and PSVs within the clade, highly reliable and sufficient sampling with superior next-generation analysis techniques is required. In this study, 71 plastomes (40 newly sequenced and assembled and the rest from the GenBank) were comparatively analyzed to decipher the PSVs and resolve the phylogenetic relationships of the Coffeae alliance using four molecular data matrices.

Results: All plastomes are typically quadripartite with the size ranging from 153,055 to 155,908 bp and contained 111 unique genes. The inverted repeat (IR) regions experienced multiple contraction and expansion; five repeat types were detected but the most abundant was SSR. The size of the Coffeae alliance clade plastomes and its elements are affected by the IR boundary shifts and the repeat types. However, the emerging PSVs had no taxonomic and phylogenetic implications. Eight highly divergent regions were identified within the plastome regions *ndhF*, *ccsA*, *ndhD*, *ndhA*, *ndhH*, *ycf1*, *rps16-trnQ-UUG*, and *psbM-trnD*. These highly variable regions may be potential molecular markers for further species delimitation and population genetic analyses for the clade. Our plastome phylogenomic analyses yielded a well-resolved phylogeny tree with well-support at the tribal and generic levels within the Coffeae alliance.

Conclusions: Plastome data could be indispensable in resolving the phylogenetic relationships of the Coffeae alliance tribes. Therefore, this study provides deep insights into the PSVs and phylogenetic relationships of the Coffeae alliance and the Rubiaceae family as a whole.

Keywords: Coffeae alliance, Phylogenomic, Plastome, Plastome structural variations (PSVs), Rubiaceae

Introduction

With over 13,000 species belonging to 611 genera and 65 tribes, Rubiaceae, commonly known as coffee, madder, or bedstraw family, is the fourth-largest angiosperm family after Asteraceae, Orchidaceae and Fabaceae [1]. Rubiaceae is cosmopolitan, occupies all the tropical and

*Correspondence: zhouyd@wbgcas.cn; qfwang@wbgcas.cn

[†]Sara Getachew Amenu and Neng Wei contributed equally to this work.

⁷Sino-Africa Joint Research Center (SAJOREC), Chinese Academy of Sciences, Wuhan 430074, Hubei, People's Republic of China
Full list of author information is available at the end of the article



subtropical ecosystems [2]. The family occupies wide range of ecological niches from moist evergreen tropical lowland forests to deserts with altitudes 4000 m above sea level [3]. Also, the family is mostly characterized as tropical trees/shrubs and rarely annual/perennial herbs [4]; hence, often referred to as one of the most biomass-rich and species-rich woody plant families [3]. Despite the eminence of previous studies on the taxonomy of Rubiaceae, still the number of traditional subfamilies that should be recognized is yet to be fully substantiated [5, 6]. However, the recent classification [6] has suggested three subfamilies Rubioideae, Cinchonoideae, and Ixoroideae.

Subfamily Ixoroideae, the largest subfamily has the highest species richness which contain ca. 4000 species in 189 genera and 26 recognized tribes [7, 8]. The Ixoroideae is morphologically characterized by secondary pollen presentation, entire interpetiolar stipules, distorted aestivation, and various succulent fruits [9]. It also includes many well-known economically important genera used for horticultural e.g., *Gardenia* J.Ellis and *Ixora* L. [10] and medicine—e.g., *Ixora*, *Gardenia* and beverages *Coffea* L. [11]. Broadly, two informal alliances are recognized within the Ixoroideae: the Vanguerieae alliance (8 tribes, >37 genera and ca. 1200 species) and the Coffeae alliance (10 tribes, >90 genera and >2000 species) [7, 8]. The Coffeae alliance is an important radiation of the subfamily Ixoroideae and previous taxonomic classification was based on morphological features such as secondary pollen presentation, total interpetiolar stipules, and contorted aestivation [5, 6]. Fruit characters, number of seeds per carpel, and the number of ovules are of low taxonomic use because of high variations within the group [10]. Interestingly, handful of previous studies have made efforts to delimit the species and clarify the phylogenetic relationships using morphological and molecular data at different taxonomic levels [5, 6].

Previous molecular phylogenetic studies based on plastid (*matK*, *ndhF*, *rbcl*, *rps16* intron, *trnS-G*, *trnT-F*, *rbcl*, and *atpB-rbcL*) and nuclear DNA data (nrITS) have shed light on the evolutionary circumstance of the Coffeae alliance [7, 8]. In most reports, some tribes, such as Airospermeae, Alberteae, Augusteae, Bertiereae, Coffeae, Octotropideae, and Pavetteae, have been described as monophyletic while the polyphyletic status of tribe Gardenieae remained obscured. Additionally, the clade support was generally low indicating many unknown relationships and internal resolution. Furthermore, several genera such as *Burchellia* R.Br., *Didymosalpinx* Keay, *Mantalanania* Capuron ex J.-F.Lero and *Monosalpinx* N.Hallé nested within the tribe failed to appear to be related to any of the described tribes [7, 8]. In addition, previous studies have also described multiple genera relative to the tribe Gardenieae e.g., such as *Adenorandia*

Vermoesen, *Aidiopsis* Tirveng, *Alleizettella* Pit, *Fosbergia* Tirveng & Sastre, *Himalrandia* T. Yamaz, *Pseudaidia* Tirveng, *Pseudomantalanania* J.-F.Leroy, and *Sulitia* Merr. Hence, urgent taxonomic reassessment via molecular phylogeny is required to understand the pattern of evolutionary relationships [5, 8]. On the other hand, in the study by Mouly et al., a revised tribal description is given for the Gardenieae by using plastid loci. The Alibertia and Sherbournia groups are recognized at the tribal level as extended Cordiereae and Sherbournieae [8]. Therefore, further studies are needed to confirm the two tribes. Overall, molecular phylogenetic studies of the Coffeae alliance using few markers (plastid regions and nrITS) have often failed to resolve the phylogenetic relationships at tribal/generic levels and yield a well-supported tree [7, 8, 10].

In recent years, the emergence of Next Generation Sequencing (NGS) technology has improved the availability of genome data, including the plastome, to address phylogenetic relationships and understand the potential roles of the plastome structural variations (PSVs) in different plant lineages. Plastome has a quadripartite structure consisting of a large single copy (LSC) region, a small single copy (SSC) region, and a pair of inverted repeat (IR) regions [12, 13]. The PSVs, including the IR contraction and expansion, gene loss and duplication, pseudogenization, and inversions have been substantiated in many angiosperm families, e.g., Geraniaceae [14], Leguminosae [15, 16], and Rubiaceae [1, 11, 17]. A recent study has explored the PSVs within the Coffeae alliance tribes of subfamily Ixoroideae and revealed frequent inversions, gene duplications, gene loss/pseudogenization, IR expansion/contraction, and loss of one IR in the plastomes [11]. Also, previous phylogenetic studies using few plastid loci, nuclear data and a limited number of species could not fully substantiate the tribe-level phylogenetic relationships of the Coffeae alliance tribes [1, 11, 17]. For instance, the sister relationships of the tribes Cordierinae and Octotropideae, as well as that of the tribes Gardenieae with Pavetteae and Sherbournieae were unresolved. Thus, studying further plastomes of the representative species can provide deep insights into the PSVs alongside their phylogenetic implications and clarify the prominent phylogenetic relationships of the Coffeae alliance.

In this study, 71 (40 newly sequenced and assembled and the remainder from the Gene bank) plastome data. We aim to explore the PSVs and reconstruct the phylogenetic tree of the Coffeae alliance using four data matrices: coding sequences (CDs), noncoding sequences (NCDs), protein-coding genes (PCGs), and complete plastome (CP) sequences. The following key challenges were systematically explored: (1)

investigation of structural variation and screening of promising marker(s) from the plastome data to provide a way forward.; (2) Identification of PSV traits and taxonomic implications; (3) Establishment of a well-supported phylogenetic framework from the partitioning schemes of the plastome data, (4) Examination of the relationships of the Coffeae alliance tribe and comparison with previous gene fragment studies of plastid phylogenomics. The results of this study will provide insights into the plastome evolution and phylogenomics of the Coffeae alliance tribes.

Results

Plastome Organization of the Coffeae alliance tribes

Mean coverage of the newly sequenced plastome ranged between $165.7 \times$ (*Pavetta barbertonensis* Bremek. of Pavetteae) and $635 \times$ (*Rubovietnamia aristata* Tirveng. of Gardenieae) (Table S1). The Plastomes of species from the Coffeae alliance tribes exhibited a typical quadripartite structure (Fig. 1) except for *Feretia aeruginescens* Stapf of Octotropideae, in which one of the IRs is absent (Fig. S1). The plastome lengths varied from 129,434 bp in *Feretia aeruginescens* to 155,908 bp in *Rothmannia manganjae* (Hiern) Keay in Gardenieae. Likewise, there were apparent variations in the size of the plastome regions (Table S2); LSC from 83,586 bp (*Pavetta abyssinica* Frese, Pavetteae) to 85,556 bp (*Diplospora dubia* (Lindl.) Masam, Gardenieae), SSC from 15,216 bp (*Diplospora dubia*, Gardenieae) to 18,229 bp (*Fosbergia shweliensis* (J. Anthony) Tirveng. & Sastre, Gardenieae), and IR from 25,138 bp (*Bertiera breviflora* Hiern, Bertiereae) to 25,937 bp (*Feretia aeruginescens*).

Additionally, we observed a marginal difference in the GC content within the range of 36.1% (*Feretia aeruginescens*) to 37.8% (*Rubovietnamia aristata*) (Table S2). Typically, the observed plastomes contain 131 genes, 84 CDS, 37 tRNA, and 4rRNA, Exon 33, and Intron 15, but in the plastome of *Feretia aeruginescens*, 114 genes, 79 CDS, 37 tRNA, and 4 rRNA, Exon 30 and Intron 12 were detected (Table S2). Among the 11 genes with introns detected in our plastomes analyses, five had double introns (*clpP*, *ndhF*, *rpl2*, *rps12*, and *ycf3*), while six had single intron (*ndhA*, *ndhB*, *petB*, *petD*, *rpl16*, and *rpoC1*) (Table 1). In addition, the *rps12* gene trans-spliced where one intron was shared between two exons, which was duplicated in the IR region. Also, the pseudogenization of *ycf1* and *rps19* genes was observed (Fig. S1).

Plastome structural variations in the Coffeae alliance tribes

Extensive IR expansion and contraction were recorded among the Coffee alliance tribes (Fig. 2, Fig. S1). The

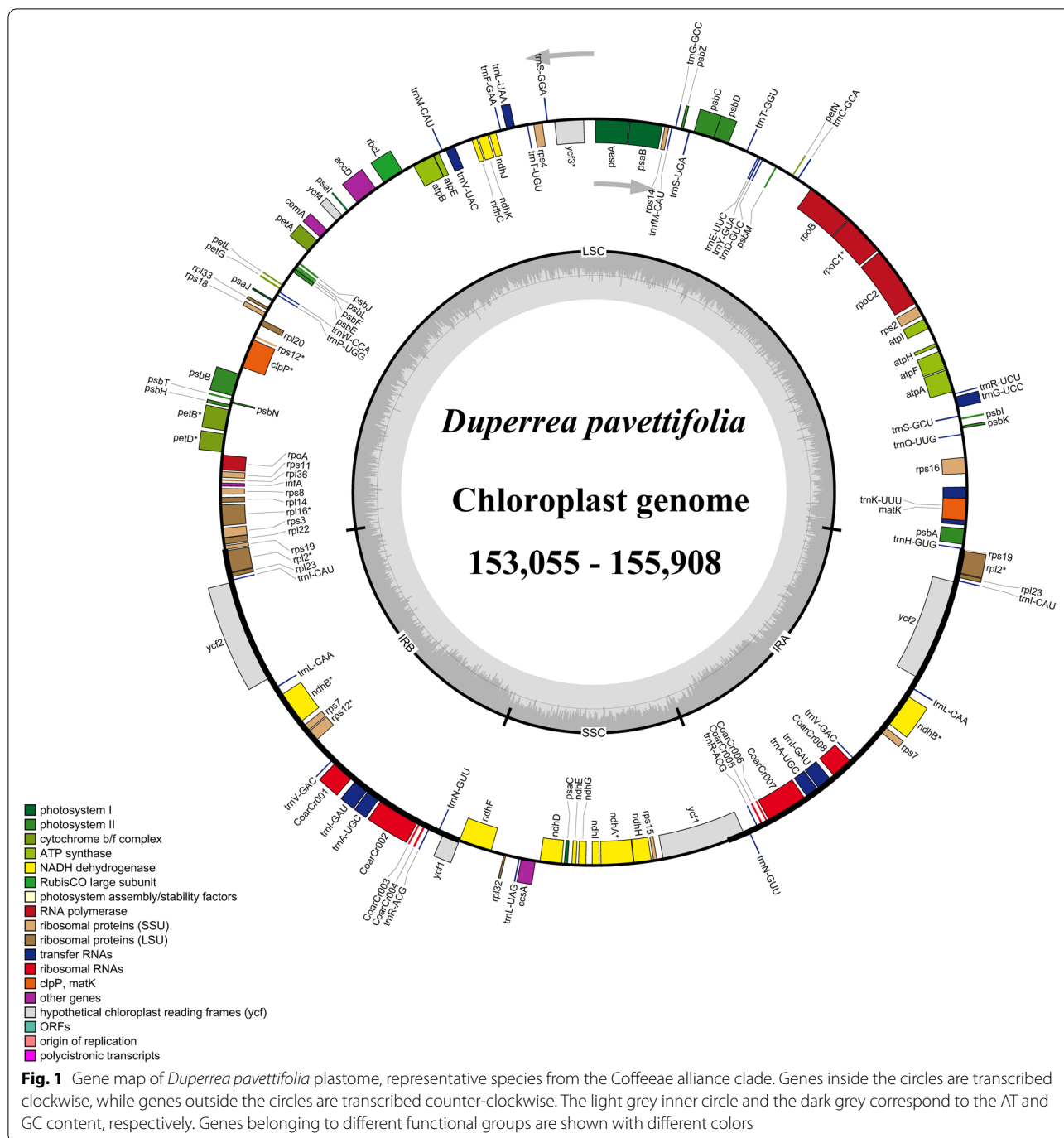
LSC/IRB (JSB) junction, which is located within the *rps19* gene led to the gene duplication in the IRA region with lengths extending from 27 bp in *Diplospora dubia* to 95 bp in *Coffea canephora* Pierre ex A. Froehner. Also, the observed IRB gene duplication of SSC/IRA (JSA) junction largely lies within the *ycf1* gene ranging from 987 bp in *Leptactina leopoldi* Buettner to 1199 bp in *Tarenna drummondii* Bridson. Mainly, the SSC/IRB junction (JSB) is located between the duplicated Ψ *ycf1* (in the IR) and *ndhF* genes (in the SSC) except in few species, where the IRB is partially expanded into the *ndhF* gene in the IR (1 to 22 bp) (Fig. 2). The LSC/IRA junction (JLA) was located between duplicated IRA boundary of Ψ *rps19* (in the IR) and *trnH-GUG* genes (in the LSC). However, the location of the *trnH-GUG* gene varied at the JLA border in different lengths (1 to 3 bp). No significant gene rearrangements were observed in the studied plastomes.

Identification of sequence divergence

It is noteworthy that the mutation and parsimony-informative site (PIS) were different in the observed plastome. A relatively large number of variable regions (mutation hotspots) were identified, where a high rate of nucleotide substitutions may occur. These include two intragenic regions and six gene regions with an informative parsimony site between 0.02977 (*rps16* - *trnQ-UUG*) and 0.05927 (*ycf1*) (Fig. 3). In addition, the informative sites in the IR region was lower than the SC regions (Fig. 3).

Codon usage and repeat analyses in the tribes of the Coffeae alliance

Codon usage and amino acid abundances in the tribes of the Coffeae alliance were similar (Fig. 4a, Fig. S2). The total sequence of encoding protein genes ranged between 74,682 bp in *Empogona ovalifolia* (Coffeae) to 79,249 bp in *Rothmannia manganjae* (Gardenieae). For the protein-coding genes, codons ranged from 22,573 bp in *Feretia aeruginescens* (Octotropideae) to 26,304 bp in *Atractocarpus fitzalanii* (F. Muell.) Puttock (Gardenieae) (Fig. S2). The most abundant amino acid, isoleucine Codon usage, ranged from 5.8 to 8.75% and the least abundant amino acid, cysteine Codon usage, averaged from 1.1 to 1.17% in Coffeae alliance tribes (Fig. S2). The majority of amino acid codons have a bias, although codons AU(T)G and U(T)GG, which encode methionine (Met) and tryptophan (Trp) respectively, both show no codon preferences (RSCU = 1.00). Additionally, most types of preferred synonymous codons (RSCU > 1.00) possessed C- or U-ending codons (Fig. S2). The percentage Effective Number of Codon (Enc) and Codon Bias Index (CBI) ranged from 55.01 to 55.24% and 0.20 to 0.25%, respectively (Table S5).



We recorded five repeat types within the studied plastomes and their percentage distribution was strikingly different (Table S4). We recorded five repeat types within the studied plastomes and their percentage distribution was strikingly different Forward repeats from 4 to 20%, Reverse repeats from 2 to 12%, Palindromic repeats from 13 to 29%, Tandem repeats from 15 to 31% and SSR repeats ranging from 29 to

45% (Table S4). In addition, of the total frequency of the repeats in the sampled group, the Dispersed repeats accounted for approximately 19% (of which 19% were palindromic, 15% Forward, and 7% Reverse), Tandem repeats for 23%, and SSR for 36% (Fig. 4b). Microsatellite analyses of the Coffeae alliance tribes identified a total of 2588 SSRs (Table S4). Our result showed a variable distribution within the tribe ranging

Table 1 Plastome content of Coffeaeae alliance tribes

Gene group	Gene
Subunit of acetyl-CoA	<i>accD</i>
Large subunit of ribosomal proteins	<i>rpl2⁺*</i> , <i>rpl14</i> , <i>rpl16⁺#</i> , <i>rpl20</i> , <i>rpl23*</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
Small subunit of ribosomal proteins	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7*</i> , <i>rps8</i> , <i>rps11</i> , <i>rps12^{++*}</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> , <i>rps18</i> , <i>rps19^{U#}</i>
Ribosomal RNA genes	<i>rrn4.5*</i> , <i>rrn5</i> , <i>rrn16*</i> , <i>rrn23*</i>
Transfer RNA genes	<i>trnA-UGC*</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnfm-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> , <i>trnI-CAU*</i> , <i>trnI-GAU*</i> , <i>trnK-UUU</i> , <i>trnL-CAA*</i> , <i>trnL-UAA</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU*</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG*</i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC*</i> , <i>trnV-UAC</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>
DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1⁺</i> , <i>rpoC2</i>
Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psal</i> , <i>psaJ</i>
Photosystem II	<i>petA</i> , <i>petB⁺</i> , <i>petD⁺</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
Cytochrome b/f complex	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>
Maturase K	<i>matK</i>
Envelope membrane proteins	<i>cemA</i>
subunit of acetyl-CoA	<i>accD</i>
RubisCo large subunit	<i>rbcl</i>
Hypothetical reading frames	<i>ycf1^{U#}</i> , <i>ycf2*</i> , <i>ycf3⁺⁺</i> , <i>ycf4</i>
Protease	<i>clpP⁺⁺</i>
NADH dehydrogenase	<i>ndhA⁺</i> , <i>ndhB⁺*</i> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF⁺⁺</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
c-type cytochrome	<i>ccsA</i>

Note: +, single intron; ++, double introns; *, gene in the IR; Ψpseudogenes; #, duplicated

from 60 in *Himalrandia lichiangensis* (W.W.Sm.) Tirveng. (Gardenieae) to 28 in *Polysphaeria parvifolia* (Octotropideae) (Fig. 4b). Of the six SSR motif types, penta-nucleotide ($n = 7$) was the most common while only one hexa-nucleotide was recorded. The analyses also revealed that the mononucleotide (A/T) motif was present in all the plastomes contrasting the tetra-nucleotide (AGAT/ATCT) motif that was only found in *Mitriostigma axillare* Hochst. (Fig. 4c). The A/T motifs accounted for more than 19%, and the other motif types occurred in less than 10% (Fig. 4c). In addition, mononucleotide repeats accounted for (71%) and hexanucleotides were the least common (Table S4).

Phylogenetic relationships of the Coffeaeae alliance tribes

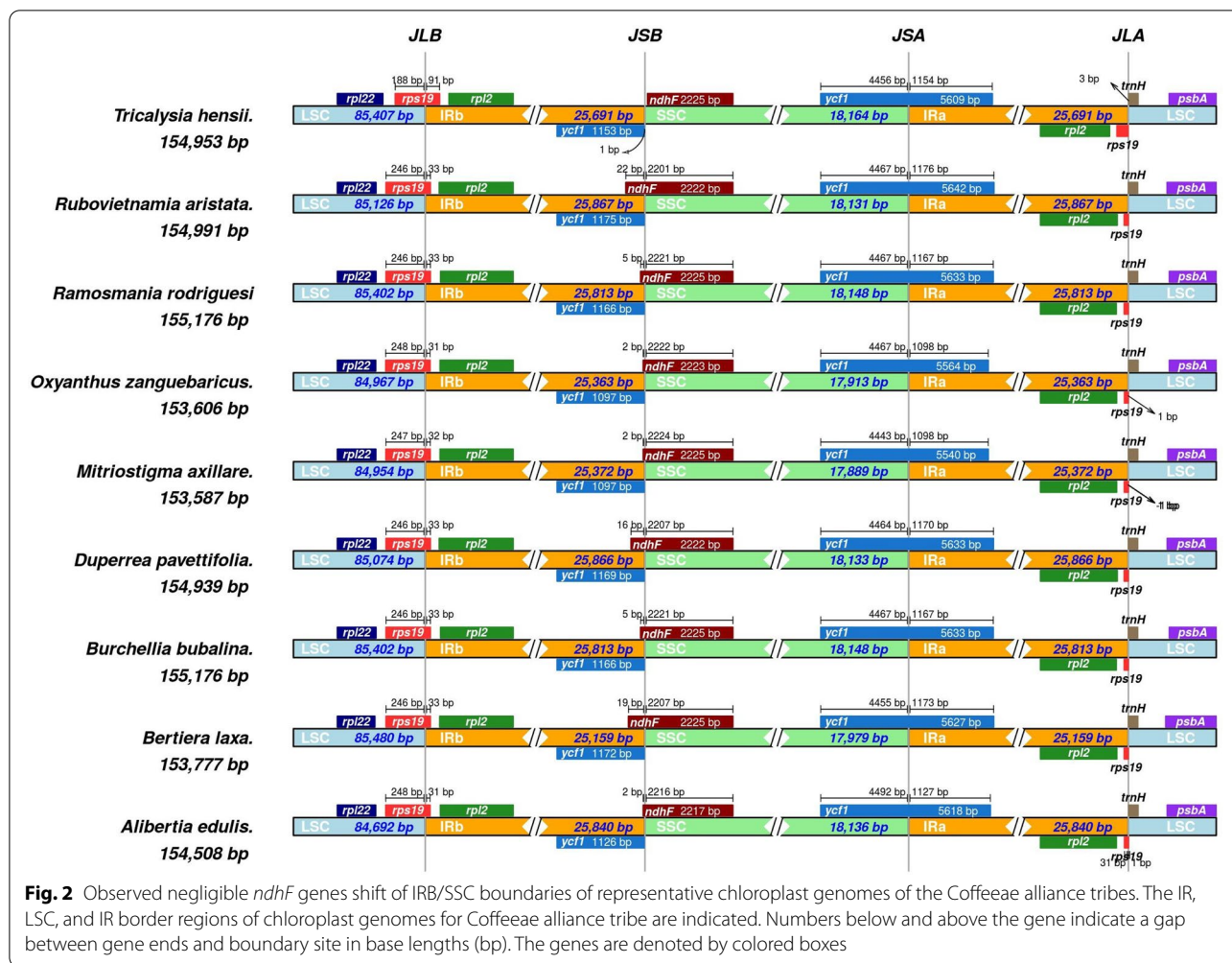
The phylogenies inferred from four data matrices using two tree methods (ML and BI) yielded nearly similar topologies. Our plastid phylogenomic analyses supported the monophyly relationships of the Coffeaeae alliance tribes and resolved relationships of all the deep nodes, with the strong support values (ML = 100%, BI = 1.0, Fig. 5). However, few subclades, namely the tribes Cordiereae + Octotropideae (ML = 47%, BI = 0.5, Fig. 5a) and *Belonophora* — *Empogona* of the tribe Coffeaeae (ML = 59%, BI = 0.8, Fig. 5a) are weakly supported. Also, the analyses presented well-supported sister relationships of the tribe Aiospermaeae to the remaining Coffeaeae alliance tribes.

The tribe Augusteae formed a sister relationship with the tribe Albertieae, as a sister tribe to the lineages consisting of tribes Octotropideae and Cordiereae.

The tribe Cordiereae has different phylogenetic positions within the entire clade of the Coffeaeae alliance from the data matrices, especially the data from the CDs and NCDs compared to other data matrices (Fig. 5). Within the Coffeaeae alliance clade, the phylogenetic position of the tribe Octotropideae in the CP and NCDs trees with weak support values was attributed to a sister relationship with the Cordiereae (Fig. 5a), whereas the monophyletic relationship of this tribe is well-supported in the trees reconstructed from the PG and CDs (Fig. 5b). The analyses supported the sister relationships of the tribes Sherbournieae-Pavetteae clade, sister to the monophyletic tribe Gardenieae with high support values (ML = 100%/BI = 1 Fig. 5b). In addition, our analyses strongly supported the relationships of the tribes Coffeaeae and Berticereae. Our data further revealed the topological differences at genera and species levels, which can easily be observed from *Heinsenias diervilleoides* K.Schum. located in different positions according to the molecular data matrices (Fig. 5).

Discussion

Evolutionary pattern of PSVs in the Coffeaeae alliance tribes
Plastomes structural variations in the subfamily Ixoroideae have previously been documented [11].



Moreover our comprehensive sampling of the plastomes from the Coffeae alliance tribe further enhances the understanding of its plastome evolutionary history. From our analyses, no gene loss was detected hence; there are marginal differences between the sizes of studied plastomes in the Coffeae alliance tribe. Contrary to our results, loss of the *trnH-GUG* gene has been reported in *Tarenna grevei* (Drake) Homolle of the subfamily Ixoroideae [11]. Previous reports have revealed the potential implications of gene loss (single or multiple) in the plastome size variations. It has been suggested in previous studies that gene losses in the plastomes are attributed to the rate of sequence evolution, gene transfer into the nucleus, or substitution of a plastid gene product with a nuclear-encoded protein [18, 19].

In our study, we suspected two pseudogenes (*rps19* and *ycf1*), based on the presence of premature stop codons and their short length. This is synapomorphy to the Coffeae alliance tribes (Fig. S1). The observation is consistent with the previous studies where the pseudogenization

of *rps19* and *ycf1* genes have been detected in many angiosperms [11, 20–22]. Pseudogenization is an evolutionary process in plastid genomes, which might occur during gene transfer to the nucleus, substitution, and nuclear-encoded protein for a plastid gene product source of DNA barcodes [23–25]. Our observation on the trans-spliced *rps12* gene (exon 1 and exons 2) is congruent with the previous report [26]. During genetic evolution events, it has been proposed that exons, which are the gene fragments encoding for protein structural domains are impacted by reorganization into other genes [27]. Additionally, as a result of RNA splicing for *rps12* exon 1 with transcripts from other genes, and polypeptide variations in the plastome have been reported [27]. Therefore, it could be suggested that similar events can explain the trans-splicing events of *rps12* gene in the plastomes members of the Coffeae alliance tribes.

The observed IR loss and occurrences of multiple IR contraction/expansion in the studied plastomes agrees with a previous study of the subfamily Ixoroideae [11].

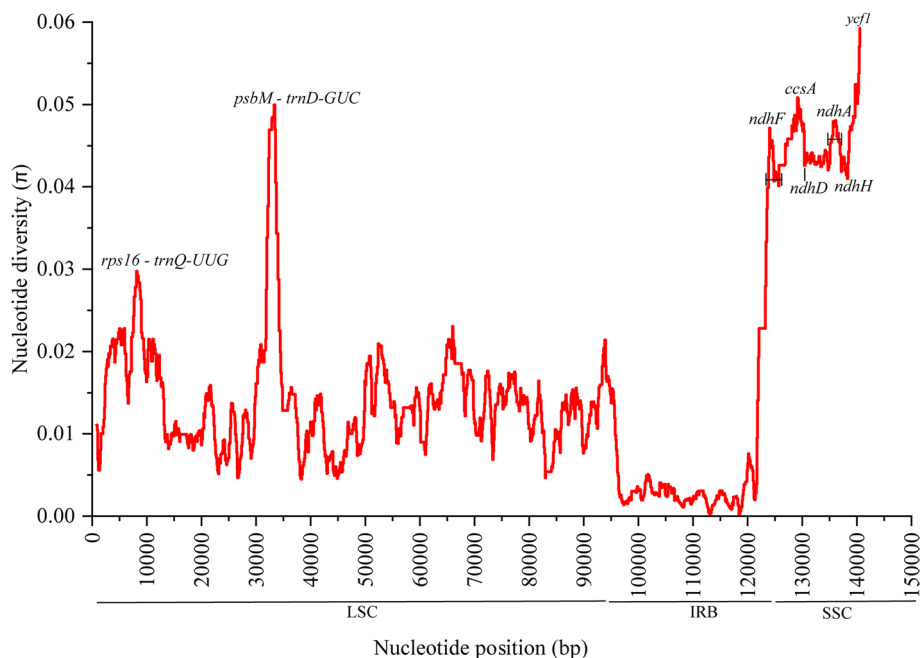


Fig. 3 Sliding-window analysis of nucleotide variability (π) across the sampled 62 plastome sequences of the Coffeaeae alliance including the single copy of IR

In our investigation, IR contraction/expansion varies at the boundaries, and this character is synapomorphy of the Coffeaeae alliance (Fig. S1). For instance, the loss of one IRs or expansion in *Feretia aeruginescens* species resulted in a larger IR size. The evolution of plant lineages is a dynamic process IR contraction/expansion at borders is the lineage with plastome genomes and plays an essential role in revealing the evolution events [28–30]. The terminal IR gene adjacent to the SSC region is highly conserved, however, there were variations at the IR/LSC boundary [31–33]. The rate of gene conversions during cell division/evolution and high concentration of short repeats (AT-rich) are the widely accepted explanation of the IR boundary shifts among several angiosperm lineages [34, 35]. In our study, the *ndhF* gene independently shifts at IRB/SSC border of the Coffeaeae alliance plastomes (Fig. 2). Other previous studies have also revealed that unique boundary shifts of the IR and SSC regions are correlated with the transformations of *ndhF* and *ycf1* genes [35–37]. The IR boundary shifts are widely accepted mechanisms for the plastomes rearrangements, which contributes to the structural diversification in plastomes [35].

Plastome sequencing has been used to infer relationships at all taxonomic levels, for example, to reveal relationships at the lowest level in land plants [36]. Two plastid protein coding genes, *rbcL* and *matK*, and the *psbA-trnH* intergenic spacer, have been recommended

as universal plastid DNA barcodes for land plants [37]. Seven promising plastid DNA barcode (*matK*, *accD*, *rps19*, *ndhF*, *ndhD*, *ccsA* and *ycf1*) have been consistently consider the most variable regions for different plant [38–41]. It was important to determine whether these loci could be used as potential molecular markers to establish relationships between species and populations of Coffeaeae allinese tribes. Our results showed that the IRs had negligible sequence divergence compared to the SC regions, which also occurred in higher plants and possibly due to copy correction between IR sequences by gene conversion [42].

Codon usage has shown that plastomes evolve due to a high frequency of privileged mutations in the coding sequences in the evolution of species [43]. The RSCU values are affected by the mutation factor that affects the codon usage pattern. Thus, Relative synonymous codon usage with value < 1 indicates that codons are used less frequently, > 1 represents codons that are used more frequently, and value = 1 indicates no bias [44]. We found identical trend in the pattern in which the amino acids were encoded for the 62 species (Fig. S2). The most abundant amino acid, isoleucine, and the trend are consistent with other angiosperm plastomes, where leucine and isoleucine were the most abundant codons [43, 45]. The Effective Number of Codons (ENc), which ranges from 20 (one codon per amino acid) to 61 (equal use of synonymous codons) [46], and the Codon Bias Index (CBI),

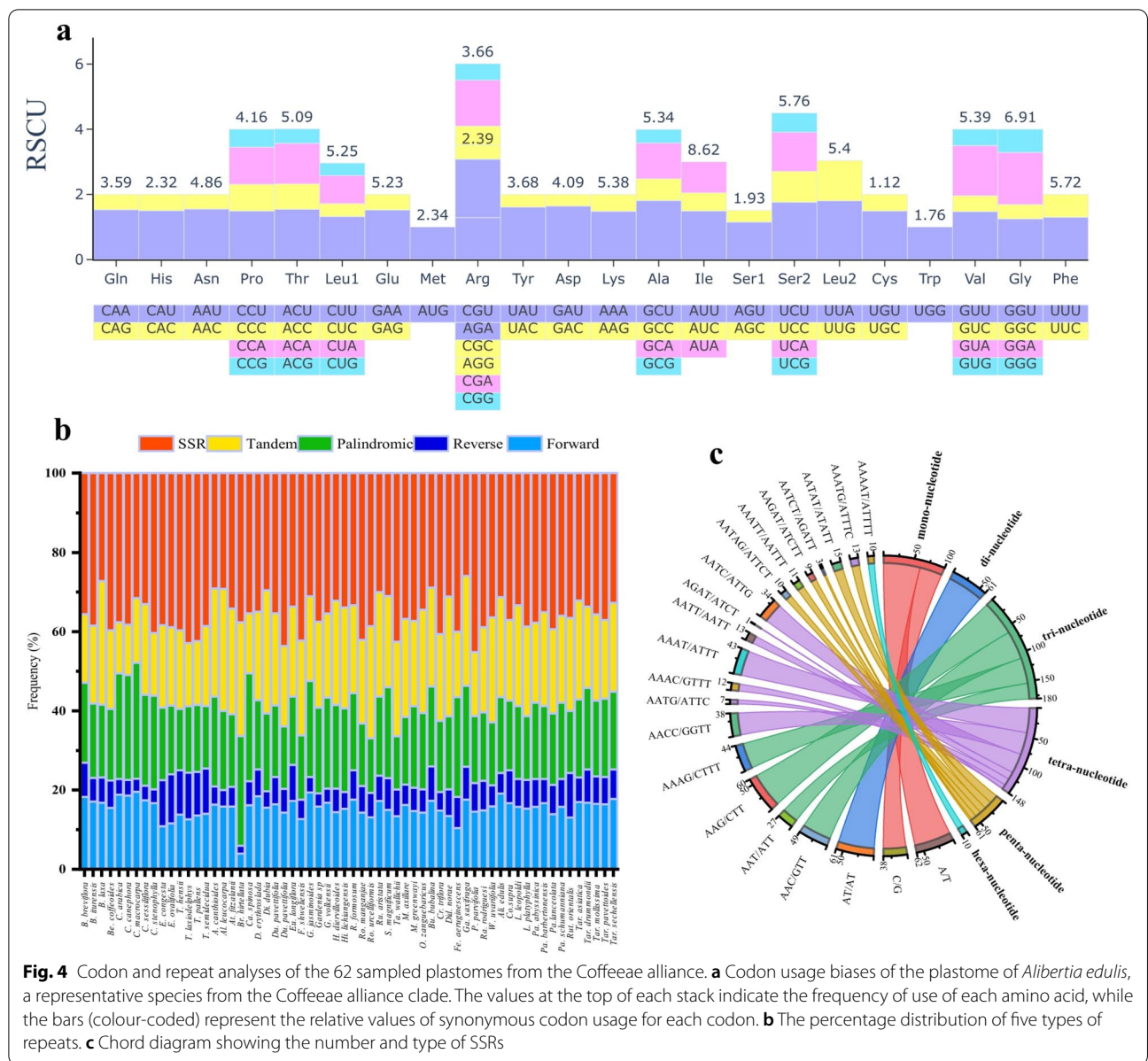
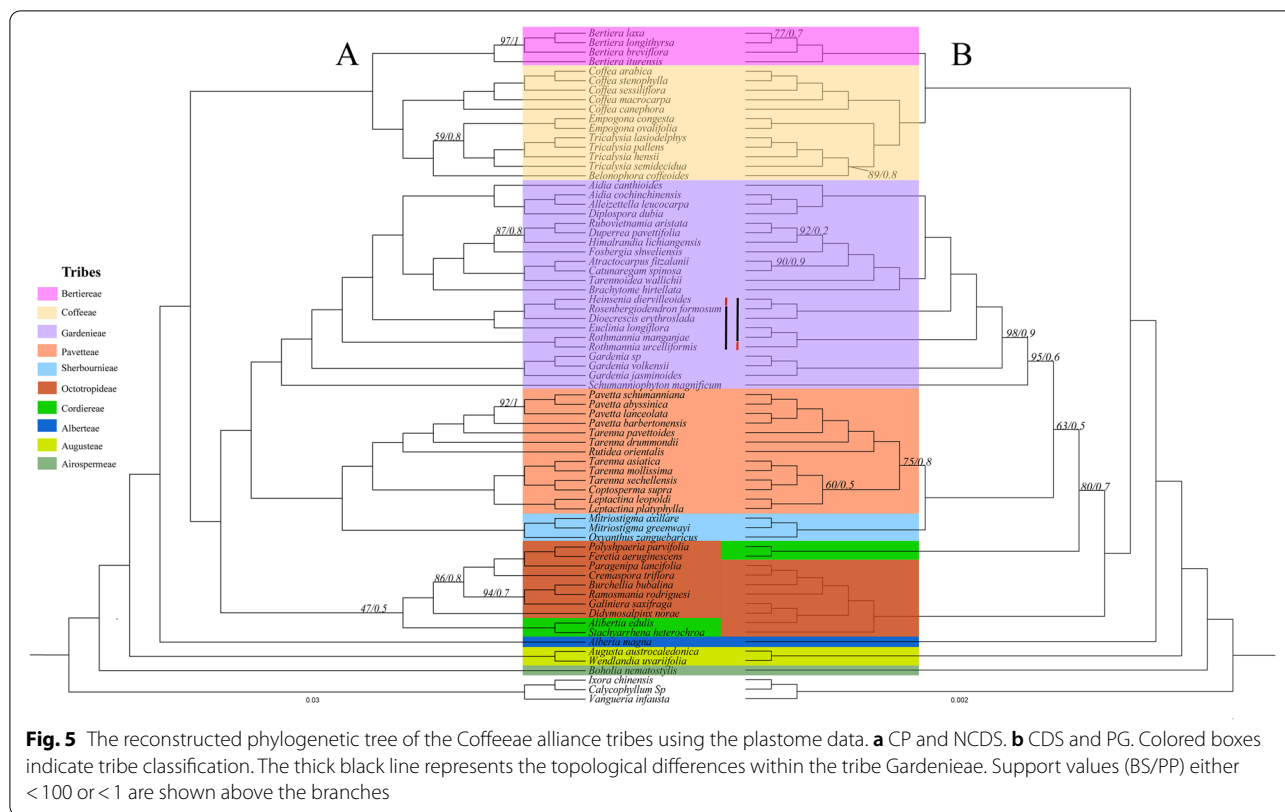


Fig. 4 Codon and repeat analyses of the 62 sampled plastomes from the Coffeaeae alliance. **a** Codon usage biases of the plastome of *Alibertia edulis*, a representative species from the Coffeaeae alliance clade. The values at the top of each stack indicate the frequency of use of each amino acid, while the bars (colour-coded) represent the relative values of synonymous codon usage for each codon. **b** The percentage distribution of five types of repeats. **c** Chord diagram showing the number and type of SSRs

which ranges from 0, no bias, to 1, equal use of all synonymous codons [47]. In our study, the values for ENC (55 to 55.2) and CBI (0.20 to 0.251) differed insignificantly among species of Coffeaeae alliance consistence with other report [48, 49].

The high degree of polymorphism and genome-wide distribution, such as significant variation in motifs and with several repeats, can be studied using SSRs analyses. Especially in determining genetic diversity and population genetic issues such as gene flow, ancestry, and population structure [50]. From the comparative analysis of five repeats in the Coffeaeae alliance tribe, SSRs are predominantly mononucleotide (A/T) motifs.

A similar trend of SSR distribution was also confirmed in the chloroplast genome of other angiosperms [43]. Other downstream analyses, repeats in a system with various rearrangements, deletions, additions, and large inversions within the cp genome are well known [51]. The percentage distributions of Tandem repeats were abundant from our analyses (Fig. 4b). Several studies have shown that a high number of repeats positively correlates with the PSV or plastome size of angiosperms. Noteworthy, a high number of Tandem repeats are important for structural reorganization of the plastome and a wide range of IR expansions/contractions [52].



Phylogenetic relationship of Coffeae alliance tribes

Our results confirm that ML and BI analyses with multiple genes partitioned models (CDs, NCDs, PG, and CP) could resolve the phylogeny relationships of the Coffeae alliance tribes (Fig. 5). The most recent phylogenetic studies have reported improved phylogenetic resolutions of the Coffeae alliance, but with only a limited number of representatives [1, 11, 17]. Again, our study punctuated the importance high taxonomic sampling for phylogenetic relationships as we included more than half of the total genus to achieve deep phylogenetic resolution of the Coffeae alliance tribes. The phylogenetic relationships among key lineages, genera, and tribes were resolved with strong support values (Fig. 5) and are consistent with previous studies [8, 11, 17, 53]. Our multi-locus analyses showed that the monophyletic Airospemeae is the first diverged lineage followed by Augusteae, Alberteae, Cordiereae, Octotropideae, Sherbournieae, Pavetteae, Gardenieae, Coffeae, and Bertiereae (Fig. 5) coinciding with recent findings [1]. Our phylogenetic analyses for all the datasets agree with previous analysis based using CDS, and combined CDS + non CDS on the phylogenetic position of the tribe Augusteae (ML = 100%/BI = 1) with full support non CDS [1].

Despite the conflicting phylogenetic positions of the tribes Cordiereae and Octotropideae (Fig. 5), the sister relationship of both tribes is largely consistent with the recent phylogenetic analyses inferred based on the plastome data [1]. Herein, we revealed the tribes Pavetteae, Sherbournieae, and Gardenieae have monophyletic relationships, including most of the genera within the tribes with the highest support values (ML=100%/BI=1). In previous studies, analyses of CDS, combined with CDS + non CDS resulted in different topologies for the tribe Gardenieae, Sherbournieae and Pavetteae. In addition, the use of plastid data in the tribe Gardenieae poorly supported sister relationship with the tribes Sherbournieae and Pavetteae. However, the tribe Pavetteae showed an unsupported sister relationship to Sherbournieae, which contrasts with the tribes Sherbournieae and Pavetteae. Our phylogenomic analyses have shown that the relationship of the tribe Pavetteae and tribe Sherbournieae is strongly supported (Fig. 5). Besides, this study strongly supports the sister relationship between the tribes Coffeae and Bertiereae, which agrees with the other studies [1, 11, 17, 54].

Previous studies using the morphological characters of the genera (*Alleizettella* Pit, *Fosbergia* Tirveng & Sastre and *Himalrandia* T. Yamaz) have a limited role in the

classification of the tribe Gardenieae [8, 55]. Our phylogenomic analyses reveal a deeper phylogenetic resolution of the Coffeae alliance tribe. We confirmed the phylogenetic placement of *Fosbergia shweliensis*, *Himalrandia lichiangensis*, and *Duperrea pavettifolia* (Kurz) Pit. Within the monophyletic tribe Gardenieae with strong support (ML=87%/BI=0.8; Fig. 5). Interestingly, the presence of morphological features such as terminal inflorescences and absent of Raphides seems to further reinforce the relationships of *Duperrea pavettifolia* and tribe Gardenieae, [8, 55]. In addition, previous phylogenetic studies of the Coffeae alliance exploiting several plastid and nuclear markers have failed to resolve the relationships between several genera (e.g., *Burchellia*, *Didymosalpinx*, and *Galiniera Delile*) and their phylogenetic status with members of the tribes Sherbournieae and Cordiereae [5, 8]. Our finding moderately supported the close relationships of these genera with the tribe Octotropideae whereas the basal lineage, *Didymosalpinx norae* (Swynn.) Keay formed a sister relationship with the remaining genera of the tribes (Fig. 5). Thus, our study provides an improved understanding of the phylogenetic relationships in the Coffeae alliance. Also, the study shows the potential of plastomes in resolving intertribal and intergeneric relationships within the Coffeae alliance tribes. However, further phylogenetic studies of the tribes Cordiereae and Sherbournieae, perhaps integrating additional molecular data and morphological traits would fully clarify the evolutionary relationships of the Coffeae alliance clade.

Conclusion

The study detected several PSVs which have occurred independently across the lineages but without clear taxonomic and phylogenetic implications. Rather, they played an important role in the structural restructuring of the plastome in the Coffeae alliance tribe. Furthermore, we screened promising molecular markers in both the intragenic and coding regions which might be suitable for future phylogenetic studies for the Coffeae alliance clade. We utilized the plastome data to present a well-resolved phylogenetic tree at the tribal and generic levels for the Coffeae alliance tribes. Thus, the results obtained provide significant insight into the PSVs and phylogenetic relationships of the Coffeae alliance and the Rubiaceae family and open a way for robust phylogenetic studies in future.

Materials and methods

Taxon sampling, DNA extraction, and genome sequencing

A total of 71 plastomes were used for this study, including 68 sequences from the alliance group of Coffeae

alliances (representing 44 genera), and three outgroups. Of these, 31 plastomes were obtained from GenBank (www.ncbi.nlm.nih.gov/genbank) and the remaining 40 were newly generated for this study (Table S1). For the newly sequenced taxa, total genomic DNA was extracted from silica-gel dried leaves using plant genomic DNA kit based on the manufacturer's instructions. The genome skimming technique was used to obtain the plastome data using fragmented gDNA data; the library size 350 bp inserts were selected. Sequencing with 2 × 150-bp paired-end (PE) read was performed on the Illumina Hiseq 2500 platform at the Novogene Company (Beijing, China).

Plastomes assembly and annotation

The clean-up Raw and quality checks were performed using the NGS QC Tool Kit [56]. De novo assembly of the cleaned reads was performed in GetOrganelle with K-mer values of 21, 45, 65, 85 and 105 GetOrganelle, using the plastome of *Coffea arabica* L. (NC008535) as reference [57]. Subsequently, the assembled reads were visualized and filtered in Bandage v.0.80 [58] to generate a complete plastome. For incomplete plastomes, we filled the gaps between contigs with consensus sequences of raw reads mapped initially to the reference plastome to obtain the complete plastome. The number of mapped paired-end (PE) reads and the depth of coverage was determined by mapping the paired reads against the plastome using Bowtie2 [59]. The locations of single copies (SC) and IR boundaries in the resequenced plastomes were determined using in Geneious v. 8.1.4 [60]. The find repeat function in Geneious was used to flank the IR regions. Then, the paired reads were reassigned to the assembled plastomes to validate the SC /IR regions with Bowtie2 [59]. Finally, we visualized the read stacks of the reassembled plastomes and compared the labeled SC/IR boundaries in Geneious. The new plastomes were annotated using the Dual Organellar Genome Annotator web interface (DOGMA) [61]. We manually checked the consistency of start/stop codons and intron/exon boundaries in Geneious. The Find ORFs function in Geneious was used to confirm protein-coding gene (PCG) annotations, while the tRNAscan-SE web service was used to determine tRNA genes [62]. The OrganellarGenomeDRAW web server was used to produce circular maps for the newly sequenced plastomes [63]. We also deposited the 40 newly assembled plastome sequences from the Coffeae alliance and outgroup in GenBank (Table S1).

Plastome structural analyses

To investigate the patterns of plastome genomic evolution of the Coffeae alliance clade, 62 annotated plastomes were compared and the length (bp) of the plastome

and its elements (LSC, SSC, and IR), gene contents (number of genes), and GC contents were summarized (Table S2, Fig. S1). Comparative analyses of the SC/IR boundary shifts at four junctions (JLB-LSC/IRB, JSB-IRB/SSC, JSA-SSC/IRA, and JLA-IRA/LSC) of the sampled plastomes was conducted in Geneious and plotted in IRscope [60, 64]. We investigated the potential rearrangements in gene order (e.g., inversion, IR boundary shifts) for all the sampled species of the Coffeae alliance using *Coffea arabica* as the reference. This analysis was performed in Mauve v.2.3.1 [65] using default settings: automatically calculate the seed weight (15) and Locally Collinear Blocks (LCBs) with the minimum LCB score of 30,000 [65].

Sequence divergence, codon usage and repeat analyses

To assess sequence divergence and examine highly variable regions in the plastome, nucleotide diversity (π) was calculated by sliding window analysis using DnaSP v6 [66] with a window size of 1000bp and a step size of 100bp. The level of codon usage bias was determined by analysing the Relative Synonymous Codon Usage RSCU [67]; Effective Number of codon ENc [46]; and the Codon Biased Index CBI [47]; for all PCGs, in DnaSP 6.10 [66]. The average RSCU and amino acid frequency values for each species were plotted using R-script (Fig. S2) [68].

We searched for three repeat types, compositions, and distributions within the sampled 62 plastomes from the Coffeae alliance tribes. Tandem repeat distribution were identified using Tandem Repeats Finder v4.09 [69] with the following settings: alignment parameters match (2), mismatch (7), indels (7), minimum alignment score (80), and maximum period size (500). Disperse repeats (forward, reverse, complementary, and palindromic) were also detected via REPuter [70] with a minimum repeat size of 30bp and a Hamming distance of 1 (Table S3). The simple sequence repeat (SSRs or microsatellites) motifs in the plastomes were searched using the Perl script using MISA v2.0 [71]. We set the minimum numbers (thresholds) of the SSRs for mono-, di-, tri-, tetra-, penta-, and hexanucleotides, to 10, 6, 5, 5, 5, and 5, respectively (Table S3).

Phylogenetic analyses

Phylogenetic analyses included a total of 71 plastomes data (68 ingroups and 3 outgroups). We performed individual gene alignment to overcome poor alignment challenges from the whole genome and generate different data matrices for the phylogenetic analyses. To achieve these, we extracted coding (CDs) and noncoding (NCDs) regions in Phylosuit v1.1.13 [68], followed by an individual gene alignments using MAFFT v7.450 [72] with LINSI algorithm and concatenation in Geneious [60].

To minimize the systematic error, we excluded NCDs loci with less than 80% taxon occupancy and alignment lengths less than 100bp. The data matrices were generated to reconstruct the phylogenetic relationships of the Coffeae alliances: CDs (84 genes), NCDs (122 loci), PG (78 genes), and CP (concatenated CDs and NCDs for all species). Substitution models of the four matrices were determined using Partition Finder v2.1.1 [73], and the best evolutionary models were selected using the Akaike Information Criterion (AIC) (Table S2). The phylogenetic relationships were reconstructed using Maximum Likelihood (ML) and Bayesian Inference (BI) tree methods. The ML analyses were performed in RAxML-HPC2 v.8.1 on XSEDE [74] with the GTRCAT substitution model and 1000 Bootstrap replicates. The BI analyses were inferred using MrBayes v.3.2.6 [75]. We estimating the Bayesian posterior probability (PP) with two independent Markov Chain Monte Carlo (MCMC) runs, via a single cold chain and three hot chains for 10,000,000 generations, while sampling the tree at every 1000 generations. We assessed the MCMC convergence of each parameter for every run based on sufficient effective sample size ($ESS \geq 200$). The first 20% runs were discarded as burn-in using TRACER v.1.7 [76]. The resulting trees (ML and BI) are visualized and edited in Fig tree v.1. 3.1 [77].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-022-03480-5>.

Additional file 1. The Mauve alignment of the 62 coffeae alliance plastomes. The collinear blocks rearrangement of the accessible gene at junction of inverted repeat and small single copy. White colored blocks: protein coding gene, Black colored blocks: tRNA genes, Red coloured blocks: Rrna. For interpretation of references to the colors, see the web version of the article.

Additional file 2. The codon usage bias in all the protein-coding genes of the plastome of the 62 Coffeae alliances tribes.

Additional file 3: Table S1. GenBank accession numbers and coverage of the newly sequenced species from the Coffeae alliance studied.

Additional file 4: Table S2. Summary of plastomes for 63 species Coffeae alliance tribes.

Additional file 5: Table S3. Summary of partitioning scheme for the ML tree and BI analysis.

Additional file 6: Table S4. Distributions of SSR in the Coffeae alliance tribes.

Additional file 7: Table S5. Details of Relative Synonymous Codon Usage in plastome of Coffeae alliance tribes.

Acknowledgments

We thank Dr. Romain Guyot from University de Montpellier, Colombia, for sharing data, Mr. Bashir Bolaji Timiyu from Wuhan Botanical Garden, CAS, China, Dr. Samaila S. Yaradua from King Abdulaziz University, Saudi Arabia, Dr. Endale Tamiru Basore from Zhejiang University, Hangzhou, Duncan Gichuki from Wuhan Botanical Garden, CAS, China and Vincent Wanga from Wuhan Botanical Garden, CAS, China for technical support during our project.

Authors' contributions

Zhou Y., Hu G., and Wang Q. conceived the idea and supervised the project. Sara G., Wei N. performed the experiments. Wei N., Wu L., Oyebanji O., Zhou, Y., and Sara G. analyzed the data and wrote the manuscript. All the authors discussed the results, commented, and revised them. The author(s) read and approved the final manuscript.

Funding

The authors acknowledge the financial support from the National Science Foundation of China (31961143026&31800176) and the Sino-Africa Joint Research Center, CAS, China (Y323771W07&SAJC201322).

Availability of data and materials

All sequences used in for this study can be found in GenBank; the list of accession can be found in Supplementary Table S1.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests

Author details

¹Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, Hubei, People's Republic of China. ²University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China. ³College of Forestry, Central South University of Forestry and Technology, Changsha 410004, Hunan, People's Republic of China. ⁴Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, People's Republic of China. ⁵Department of Botany, Faculty of Science, University of Lagos, Lagos, Nigeria. ⁶Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan 430074, Hubei, People's Republic of China. ⁷Sino-Africa Joint Research Center (SAJOREC), Chinese Academy of Sciences, Wuhan 430074, Hubei, People's Republic of China.

Received: 8 September 2021 Accepted: 15 February 2022

Published online: 26 February 2022

References

- Wikstrom N, Bremer B, Rydin C. Conflicting phylogenetic signals in genomic data of the coffee family (Rubiaceae). *J Syst Evol*. 2020;58(4):440–60.
- Mendoza-Cifuentes H, Celis A, Tovar E, Gonzalez MA. *Lintersemia* (Rubiaceae: Condamineae), a new and enigmatic genus from the Magdalena Medio region of Colombia. *Phytotaxa*. 2020;451(1):1–20.
- Davis AP, Govaerts R, Bridson DM, Ruhsam M, Moat J, Brummitt NA, et al. A global assessment of distribution, diversity, endemism, and taxonomic effort in the Rubiaceae. *Ann Mo Bot Gard*. 2009;96(1):68–78.
- Robbrecht E. Tropical woody Rubiaceae. *Op Bot Belg*. 1988;1(272):599–602.
- Robbrecht E, Manen JF. The major evolutionary lineages of the coffee family (Rubiaceae, Angiosperms). Combined analysis (nDNA and cpDNA) to infer the position of *Coptosapelta* and *Luculia*, and supertree construction based on *rbcL*, *rps16*, *trnL-trnF*, and *atpB-rbcL* data. A new classification in two sub-families, Cinchonoideae and Rubioideae. *Syst Geogr Pl*. 2006;76:85–146.
- Bremer B, Eriksson T. Time tree of Rubiaceae: phylogeny and dating the family, subfamilies, and tribes. *Int J Plant Sci*. 2009;170(6):766–93.
- Kainulainen K, Mouly A, Khodabandeh A, Bremer B. Molecular phylogenetic analysis of the tribe Alberteae (Rubiaceae), with description of a new genus, *Razafimandibsonia*. *Taxon*. 2009;58(3):757–68.
- Mouly A, Kainulainen K, Persson C, Davis AP, Wong KM, Razafimandibson SG, et al. Phylogenetic structure and clade circumscriptions in the Gardenieae complex (Rubiaceae). *Taxon*. 2014;63(4):801–18.
- Puff C, Robbrecht E, Buchner R, De Block P. A survey of secondary pollen presentation in the Rubiaceae. *Opera Bot Belg*. 1996;7:369–402.
- Andreasen K, Bremer B. Combined phylogenetic analysis in the Rubiaceae-Ixoroideae: morphology, nuclear and chloroplast DNA data. *Am J Bot*. 2000;87(11):1731–48.
- Ly SN, Garavito A, De Block P, Asselman P, Guyeux C, Charr JC, et al. Chloroplast genomes of Rubiaceae: comparative genomics and molecular phylogeny in subfamily Ixoroideae. *PLoS One*. 2020;15(4):e0232295.
- Yaradua SS, Alzahrani DA. Complete chloroplast genome sequence of *Justicia flava*: comparative genome analysis and phylogenetic relationships among Acanthaceae. *Biomed Res Int*. 2019;2019:1–17.
- Alzahrani DA, Albokhari EJ, Yaradua SS, Abba A. Comparative analysis of chloroplast genomes of four medicinal Capparaceae species: genome structures, phylogenetic relationships, and adaptive evolution. *Plants*. 2021;10(6):1229.
- Martin GE, Rousseau-Gueutin M, Cordonnier S, Lima O, Michon-Coudouel S, Naquin D, et al. The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. *Ann Bot*. 2014;113(7):1197–210.
- Oyebanji O, Zhang R, Chen SY, Yi TS. New insights into the plastome evolution of the Millettoid/Phaseoloid Clade (Papilionoideae, Leguminosae). *Front Plant Sci*. 2020;11:151.
- Wang YH, Qu XJ, Chen SY, Li DZ, Yi TS. Plastomes of *Mimosoideae*: structural and size variation, sequence divergence, and phylogenetic implication. *Tree Genet Genomes*. 2017;13(2):41.
- Antonelli A, Clarkson JJ, Kainulainen K, Maurin O, Brewer GE, Davis AP, et al. Settling a family feud: a high-level phylogenomic framework for the Gentianales based on 353 nuclear genes and partial plastomes. *Am J Bot*. 2021;108:1143–65.
- Slimp M, Williams LD, Hale H, Johnson MG. On the potential of angiosperms353 for population genomic studies. *Appl Plant Sci*. 2021;9(7):e11419.
- Chu X, Li S, Wang S, Luo D, Luo H. Gene loss through pseudogenization contributes to the ecological diversification of a generalist Roseobacter lineage. *ISME J*. 2021;15(2):489–502.
- Shrestha B, Gilbert LE, Ruhlman TA, Jansen RK. Rampant nuclear transfer and substitutions of plastid genes in *Passiflora*. *Genome Biol Evol*. 2020;12(8):1313–29.
- Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, et al. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoSOne*. 2013;8:e57607.
- Curci PL, De Paola D, Danzi D, Vendramin GG, Sonnante G. Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other Asteraceae. *PLoSOne*. 2015;10:e0120589.
- Menezes AP, Resende-Moreira LC, Buzatti RS, Nazareno AG, Carlsen M, Lobo FP, Kalopothakis et al. Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Sci Rep* 2018;8:1–12.
- Jansen RK, Ruhlman TA. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. *Genomics of chloroplasts and mitochondria*. Dordrecht: Springer Netherlands: Springer; 2012. p. 103–26.
- Williams AV, Boykin LM, Howell KA, Nevill PG, Small I. The complete sequence of the *Acacia ligulata* chloroplast genome reveals a highly divergent *clpP1* gene. *PLoSOne*. 2015;10(9):e0138367.
- Hildebrand M, Hallick RB, Passavant CW, Bourque DP. Trans-splicing in chloroplasts: the *rps12* loci of *Nicotianatabacum*. *Proc Natl Acad Sci*. 1988;85(2):371–6.
- Sharp PA. On the origin of RNA splicing and introns. *Cell*. 1985;42:397–400.
- Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol*. 2016;209(4):1747–56.
- Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics*. 2007;8(1):1–27.
- Wei N, Pérez-Escobar OA, Musili PM, Huang WC, Yang JB, Hu AQ, et al. Plastome evolution in the hyperdiverse genus *Euphorbia* (Euphorbiaceae) using phylogenomic and comparative analyses: large-scale expansion and contraction of the inverted repeat region. *Front Plant Sci*. 2020;15:55.

31. Wu C-S, Lai YT, Lin CP, Wang YN, Chaw SM. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. *Mol Phylogenet Evol.* 2009;52(1):115–24.
32. Plunkett GM, Downie SR. Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. *Syst Bot.* 2000;25(4):648–67.
33. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol Biol Evol.* 2011;28(4):1543.
34. Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, et al. Diversification of Rosaceae since the late cretaceous based on plastid phylogenomics. *New Phytol.* 2017;214(3):1355–67.
35. Dugas DV, Hernandez D, Koenen EJ, Schwarz E, Straub S, Hughes CE, et al. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions and accelerated rate of evolution in *clpP*. *Sci Rep.* 2015;5(1):1–13.
36. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *Am J Bot.* 2014;101(11):1987–2004.
37. Bai HR, Oyeibanji O, Zhang R, Yi TS. Plastid phylogenomic insights into the evolution of subfamily Dialioideae (Leguminosae). *Plant Divers.* 2021;43(1):27–34.
38. Duchene D, Bromham L. Rates of molecular evolution and diversification in plants: chloroplast substitution rates correlate with species-richness in the Proteaceae. *BMC Evol Biol.* 2013;13(1):1–11.
39. Ni L, Zhao Z, Xu H, Chen S, Dorje G. The complete chloroplast genome of *Gentiana straminea* (Gentianaceae), an endemic species to the Sino-Himalayan subregion. *Gene.* 2016;577(2):281–8.
40. Li Z, Long H, Zhang L, Liu Z, Cao H, Shi M, et al. The complete chloroplast genome sequence of tung tree (*Vernicia fordii*): organization and phylogenetic relationships with other angiosperms. *Sci Rep.* 2017;7(1):1–11.
41. Li Y, Jia L, Wang Z, Xing R, Chi X, Chen S, et al. The complete chloroplast genome of *Saxifraga sinomontana* (Saxifragaceae) and comparative analysis with other Saxifragaceae species. *Rev Bras Bot.* 2019;42(4):601–11.
42. Khakhlova O, Bock R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* 2006;46(1):85–94.
43. Saina JK, Li Z-Z, Gichira AW, Liao Y-Y. The complete chloroplast genome sequence of tree of heaven (*Ailanthus altissima* (mill.)) (sapindales: Simaroubaceae), an important pantropical tree. *Int J Mol Sci.* 2018;19(4):929.
44. Barbhuiya PA, Uddin A, Chakraborty S. Analysis of compositional properties and codon usage bias of mitochondrial *CYB* gene in anura, urodela and gymnophiona. *Gene.* 2020;751:144762.
45. Henriquez CL, Mehmood F, Carlsen MM, Islam M, Waheed MT, Poczai P, et al. Complete chloroplast genomes of *Anthurium huixtense* and *Pothos scandens* (Pothoideae, Araceae): unique inverted repeat expansion and contraction affect rate of evolution. *J Mol Evol.* 2020;88(7):562–74.
46. Wright F. The effective number of codons used in a gene. *Gene.* 1990;87(1):23–9.
47. Morton BR. Chloroplast DNA codon use: evidence for selection at the *psbA* locus based on tRNA availability. *J Mol Evol.* 1993;37(3):273–80.
48. Nie S, Yao S, Qin C, Li K, Liu X, Wang L, et al. Kinetics of AOX formation in chlorine dioxide bleaching of bagasse pulp. *Bioresour.* 2014;9(3):5604–14.
49. Gichira AW, Avoga S, Li Z, Hu G, Wang Q, Chen J. Comparative genomics of 11 complete chloroplast genomes of Senecioideae (Asteraceae) species: DNA barcodes and phylogenetics. *Bot Stud.* 2019;60(1):17.
50. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. *Biol Rev.* 2015;90(1):157–66.
51. Keller J, Rousseau-Gueutin M, Martin GE, Morice J, Boutte J, Coissac E, et al. The evolutionary fate of the chloroplast and nuclear *rps16* genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from *Lupinus*. *DNA Res.* 2017;24(4):343–58.
52. 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.
53. Kainulainen K, Razafimandimbison SG, Wikström N, Bremer B. Island hopping, long-distance dispersal and species radiation in the Western Indian Ocean: historical biogeography of the Coffeae alliance (Rubiaceae). *J Biogeogr.* 2017;44(9):1966–79.
54. Toshi J, Davis AP, Dessein S, De Block P, Huysmans S, Fay MF, et al. Phylogeny of *Tricalysia* (Rubiaceae) and its relationships with allied genera based on plastid DNA data: resurrection of the genus *Empogona*. *Ann Mo Bot Gard.* 2009;96:194–213.
55. Robbrecht E, Puff C. Survey of the Gardenieae and related tribes (Rubiaceae). *Bot Jahrb Syst Pflanzengesch Pflanzengeogr.* 1986;108:63–137.
56. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next-generation sequencing data. *PLoS One.* 2012;7:e30619.
57. Jin JJ, Yu W-B, Yang JB, Song Y, Depamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de-novo assembly of organelle genomes. *Genome Biol.* 2020;21(1):1–31.
58. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics.* 2015;31(20):3350–2.
59. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357–9.
60. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, 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.
61. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 2004;20(17):3252–5.
62. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 2005;33:W686–9.
63. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 2013;41(W1):W575–81.
64. Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 2018; 34(17):3030–3031.
65. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004;14(7):1394–403.
66. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol.* 2017;34(12):3299–302.
67. Sharp PM, Li W-H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 1987;15(3):1281–95.
68. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. 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.
69. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27(2):573–80.
70. 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.
71. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet.* 2003;106(3):411–22.
72. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice, and visualization. *Brief Bioinform.* 2019;20(4):1160–6.
73. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. Partition finder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol.* 2017;34(3):771–3.
74. Stamatakis A. RAXML-HPC v.8.1 on XSEDE: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3.
75. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003;19(12):1571–4.
76. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 2018;67(5):901.
77. Rambaut A. Fig Tree version 1.3.1 (2009). [computer program] <http://tree.bio.ed.ac.uk>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.