RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2024.083

ScienceAsia 50 (4): 2024: ID 2024083: **1–9**

Genome survey and complete chloroplast genome structure of a bamboo species (*Bambusa grandis*) provide a useful basis in genomic research

Yu Liu[a], Huanwen Xu[b], Xing Liu[a], Xiaowen Li[a], Guanju Chen[a], Xin Wei[a], Yueying Wang[a], Chuan Jin[a], Jinwang Wang[a,*], Haitao Xia[a,*]

[a] Wenzhou Key Laboratory of Resource Plant Innovation and Utilization, Zhejiang Institute of Subtropical Crops, Zhejiang Academy of Agricultural Sciences, Wenzhou, Zhejiang 325005 China

[b] Southern Zhejiang Key Laboratory of Crop Breeding, Wenzhou Vocational College of Science and Technology (Wenzhou Academy of Agricultural Sciences), Wenzhou, Zhejiang 325006 China

*Corresponding authors, e-mail: kingwwang@163.com, xiaht@zaas.ac.cn

Received 25 Aug 2023, Accepted 13 Aug 2024
Available online 19 Sep 2024

ABSTRACT: *Bambusa grandis* is a bamboo species with high ecological and economic value. In this study, we sequenced the *B. grandis* chloroplast (cp) genome and reconstructed the phylogeny of the genus *Bambusa*. The *B. grandis* cp genome was 139,443 bp in length, which included 136 unique genes and had a GC content of 38.91%. The usage rates of three codons (ATG, TTA, and AGA) were particularly high, indicating codon-biased usage. Multi-genome comparative analyses indicated that the cp genome was highly conserved across species; the contraction and expansion of the IR region appeared to be the primary driver of IR/SC boundary variation across *Bambusa* species. Four regions with high-level polymorphism were discovered with the potential to be used as molecular markers. Phylogenetic analysis indicated that *B. grandis* is most closely related to *B. basihirsuta*, *B. cornigera*, and *B. intermedia*. In addition, we also carried out genome survey sequencing of *B. grandis* and identified genome-wide microsatellite motifs. The results revealed an estimated genome size for *B. grandis* of 1.18 Gb with 5.63% heterozygosity and a repeat ratio of 55.80%. A total of 326,184 SSRs were identified from the genome survey assembly, most of which were mononucleotide motifs with a frequency of 43.84%. The findings will be beneficial to future research on the systematics, genetic diversity, and evolutionary history of the genus *Bambusa* as well as the conservation and sustainable utilization of *B. grandis* genetic resources.

KEYWORDS: *Bambusa*, *Bambusa grandis*, chloroplast genome, phylogeny

INTRODUCTION

The economically and ecologically important bamboo *Bambusa grandis* is widely distributed in Guangxi province, China, and has been introduced to southeastern China. *B. grandis* grows primarily in stony floodplains and warm, humid drainage basins, where its stands act as barriers to soil erosion. Furthermore, *B. grandis* provides food in the form of nutritious rhizome buds and building materials in the form of culm wood [1]. To date, most large-scale studies on the classification of *B. grandis* have been based on morphological traits such as culm wood characteristics, culm length, leaf texture, and leaf shape [1]. However, despite the significance of *B. grandis* with its high ecological and economic value, genomic research is still incomplete; the genomic characteristics and chloroplast genome structure are less known. Therefore, more genomic information will help solve practical problems in taxonomy and tree breeding [2, 3].

In this study, the *B. grandis* cp genome was assembled, and multi-genome comparative analyses in the genus *Bambusa* were performed. In addition, we also carried out genome survey sequencing of *B. grandis* and identified genome-wide microsatellite motifs. These results will be beneficial to research on genetic diversity, genetic map construction, and variety identification of *B. grandis*.

MATERIALS AND METHODS

Samples and DNA extraction

The *B. grandis*, introduced by our research team from Nanning, Guangxi province in 2014, has thrived since its introduction to Zhejiang province. Fresh, disease-free *B. grandis* leaves were collected from Cangnan county (120°20′ E, 27°26′ N), Wenzhou, Zhejiang, China. The voucher specimen (2021D20) was deposited in the Zhejiang Institute of Subtropical Crops, Zhejiang Academy of Agricultural Sciences, China. Genomic DNA was isolated using a modified CTAB method [4].

DNA sequencing, assembly, and validation

Shotgun libraries (150 bp) were constructed with genomic DNA and sequenced using the BGISEQ-500 platform. Clean reads were obtained after quality control of raw reads using default parameters of Fastp 0.21 software. The chloroplast genome was assembled using the Noveplastys software with the complete chloro-

plast sequences of all *Bambusa* species in GenBank serving as seed sequences. K-mer settings of 49, 59, and 69 were tested, and the results showed that the assembly sequences were consistent and automatically circularized. The assembled reads were then compared to the assembled circular chloroplast genome using Geneious Prime software, no gap regions were found, and the generated consensus sequence matched the chloroplast genome sequence assembled by Noveplastys. This concordance indicates that the assembly result was accurate [5, 6]. The *B. grandis* cp genome was annotated with Blast, Aragorn, and Hmmer 3 using the sequences of closely related species (*Bambusa latiflora* (Munro) Kurz) for comparison [7–9]. The circular cp genome map was created with Organellar Genome Draw 1.3.1 [10]. The fully assembled cp genome was submitted to GenBank (NC_068816). Jellyfish 2.2.10 was used to perform K-mer analysis with a K-value of 21. Based on the 21-mer distribution, GenomeScope was used to estimate genome size, heterozygosity rate, and repeat content [11]. Velvet 1.2.10 software was used to *de novo* assemble the *B. grandis* genomes with the starting sequence [12]. Augustus software was used for gene prediction of the *B. grandis* genome sequence [13]. Gene functional annotation was performed using BLAST by searching the NR, eggNOG, COG, Pfam, KEGG, GO, and CAZy databases [7].

Codon usage pattern analysis

Protein-coding genes > 300 nucleotides in length were extracted to analyze the relative synonymous codon usage (RSCU) with CodonW 1.4.4. RSCU reflects codon usage bias with values approaching 1 indicating that all synonymous codons encoding the same amino acid are used with equal frequency [14].

Sequence divergence, genome comparison, and repeat sequence analyses

A comparative genomic analysis was conducted between the *B. grandis* cp genome and the cp genomes of 10 other *Bambusa* species (*B. oldhamii*, *B. odashimae*, *B. rutile*, *B. boniopsis*, *B. basihirsuta*, *B. bicicatricata*, *B. stenoaurita*, *B. sinospinosa*, *B. beecheyana* var. *pubescens*, and *B. variostriata*). The cp genomes were compared using mVISTA [15] in shuffle-LANGAN mode with default parameters and *B. oldhamii* used as a reference. Nucleotide diversity (Pi) was determined using DnaSP 5.0 with a window size of 600 bp and a locus-to-locus distance of 200 bp [16]. IRSCOPE was used to analyze the IR/SC boundaries between the 4 primary regions [17].

Phylogenetic analysis

A phylogenetic analysis of the genus Bambusa was conducted using the newly-constructed *B. grandis* cp genome and the cp genomes of 26 other *Bambusa* species downloaded from the NCBI database (*B. rigida*, *B. ventricosa*, *B. oldhamii*, *B. odashimae*, *B. pervariabilis*, *B. rutila*, *B. multiplex*, *B. boniopsis*, *B. flexuosa*, *B. basihirsuta*, *B. cornigera*, *B. intermedia*, *B. bicicatricata*, *B. contracta*, *B. stenoaurita*, *B. teres*, *B. tulda*, *B. vulgaris*, *B. sinospinosa*, *B. lapidea*, *B. arnhemica*, *B. bambos*, *B. oliveriana*, *B. beecheyana*, and *B. variostriata*, and *B. latiflora*) with *Dendrocalamus barbatus* (NC_050747) and *D. brandisii* (NC_050763) used as the outgroup. The complete cp genomes were aligned using MAFFT 7.450 with default settings [18]. The maximum likelihood (ML) tree was generated in IQTree v2.2.0 software with 5000 ultrafast bootstrap replicates and 1000 replicates of the Shimodaira–Hasegawa-like (SH-like) approximate likelihood ratio test. The optimal model was HKY+F+I, chosen according to BIC by ModelFinder. Meanwhile, a Bayesian inference tree was produced using MrBayes v3.2.7, based on Markov Chain Monte Carlo (MCMC) runs for 2,000,000 generations, employing the HKY+F+I model of nucleotide substitution, as determined by ModelFinder using BIC [19].

Microsatellite motif detection

MISA 2.1 was used to perform genome-wide SSR identification [20]. The parameters were set as follows: the minimum repetition values of dinucleotide to hexanucleotide motif repetition units were set at 6, 5, 4, 4, and 4, respectively.

RESULTS AND DISCUSSION

Assembly and assessment of *B. grandis* cp genome

A total of 640,704,808 bp of clean reads were obtained from BGISEQ-500 sequencing with a Q20 value of 97.75%. The *B. grandis* cp genome was 139,443 bp in length, exhibited a classical quadripartite circular structure, and contained a pair of IR (IRa and IRb) regions (21,795 bp) separated by one large single copy (LSC) region (82,954 bp) and one small single copy (SSC) region (12,899 bp) (Fig. 1). The GC content of the IR region (44.21%) was higher than that of the LSC (37.01%) and SSC (33.22%) regions, and the average whole-genome GC content was 38.91%. This result is similar to those reported from other flowering plants [21] and may result from the presence of rRNA genes in IR regions [22]. All genes contained within the *B. grandis* cp genome were functionally annotated. In all, the *B. grandis* cp genome contained 136 predicted functional genes, including 88 protein-coding genes, 8 rRNA genes, and 40 tRNA genes.

Codon usage bias

The codon preference was determined in order to analyze the gene base composition. In all, the CDS of the *B. grandis* cp genome encoded 20,713 amino acids. The most abundant codon was leucine, comprising 2,234 codons (10.79%), followed by isoleucine

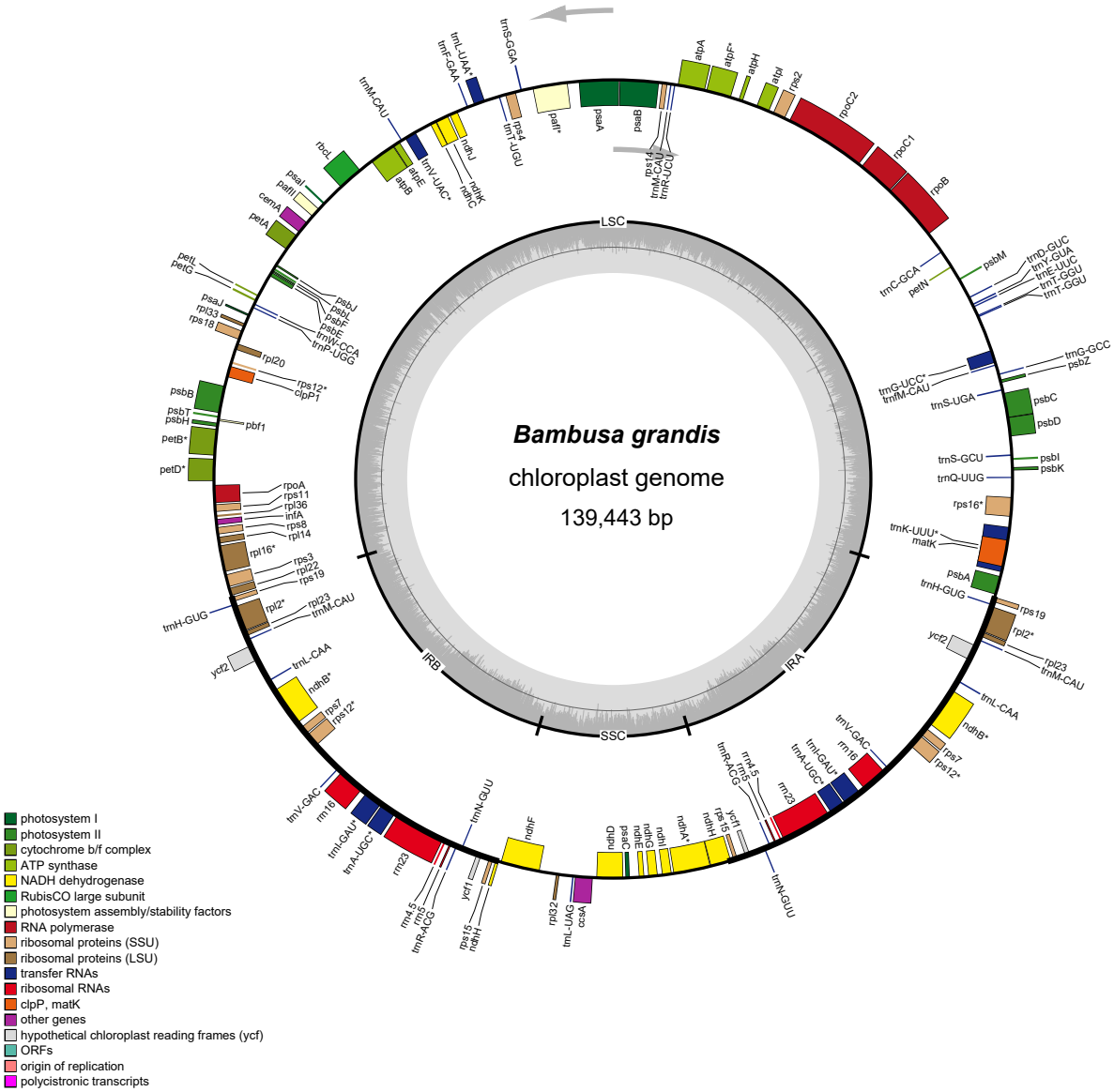

Fig. 1 Gene map of the *B. grandis* cp genome. The forward-coding genes are located on the outside of the circle, and the reverse-coding genes are located on the inside of the circle. The color coding indicates gene function. In the inner circle, dark grey represents the GC content, and light grey represents the AT content.

(1,681 codons, 8.12%), with cysteine being the most infrequent amino acid (225 codons, 1.09%). The RSCU values varied from 0.0165 (ACG and ATA codons in Met) to 3.9424 (ATG codon in Met), and 31 codons with RSCU > 1 were observed across the genome. The usage rates of 3 codons (ATG, TTA, and AGA) were particularly high, indicating codon-biased usage (Fig. 2).

Expansion and contraction of border regions

The IR boundary regions across the *B. grandis* cp genome were examined (Fig. 3), and it was found that the IR/SC boundaries are highly conserved across *Bambusa* species. Across most *Bambusa* species, *rpl22*, *rsp19*, *ndhH*, *rpsl9*, and *psbA* were distributed at the LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC boundaries, respectively. Across all *Bambusa* species except *B. variostriata*, *rpl22* was located in the LSC region, 24-32 bp from the LSC/IRb boundary. In *B. variostriata*, *rpl22* was located far from the LSC and IRb boundaries, which may explain why *rps19* (166 bp) spans the LSC and IRB boundaries in this species. With the exception of *B. variostriata*, *rps19* was located in the IRa region, 37–50 bp from the IRa/LSC boundary. Across all 11 *Bambusa* species, *ndhH* gene crossed the boundary of the IR-SSC region. Overall, the contraction and

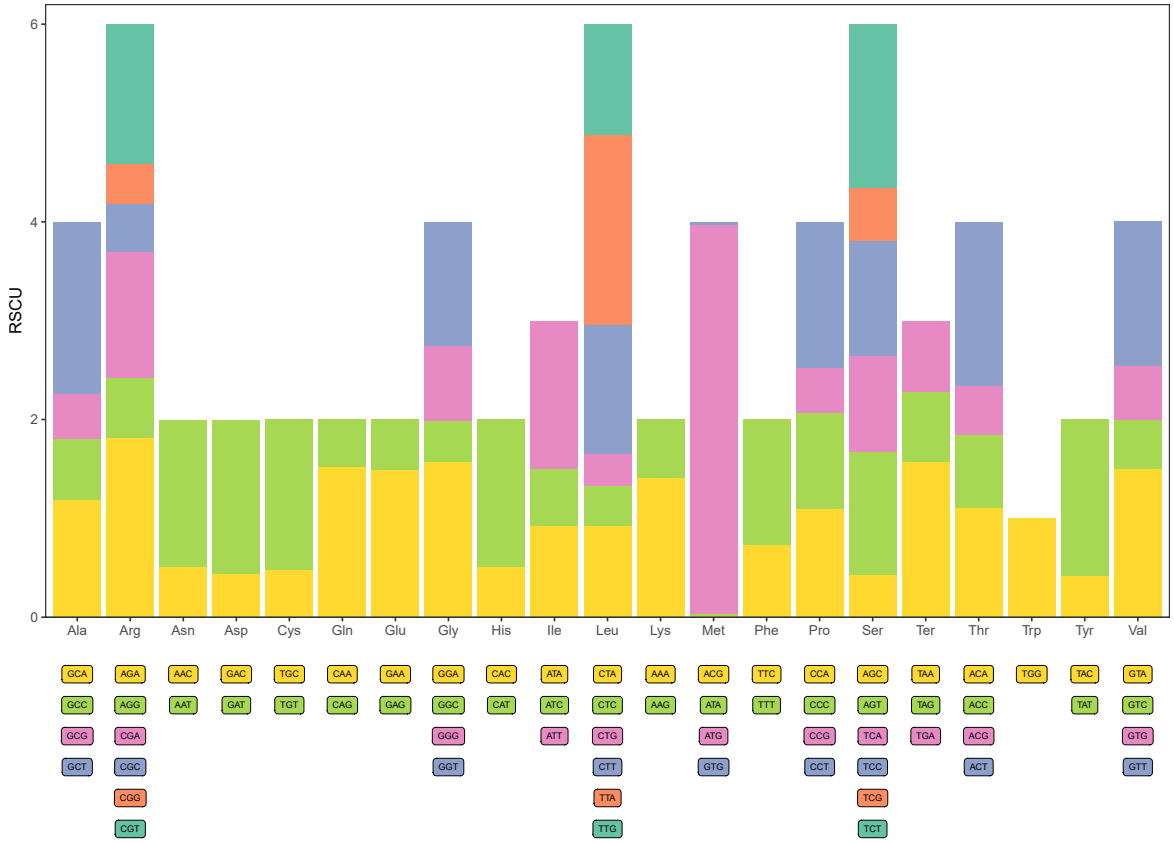

Fig. 2 RSCU of amino acids in the *B. grandis* cp genome. Histogram colors correspond to the colors assigned to each codon.

expansion of the IR region appear to be the primary driver of IR/SC boundary variation across *Bambusa* species.

Comparison of complete cp genomes across *Bambusa* species

The sequences from 11 cp genomes were compared using multi-genome comparative analyses (Fig. 4). The cp genomes of these 11 species ranged from 139,343 to 139,548 bp in length. Overall, we found a high degree of similarity in 11 *Bambusa* species, indicating considerable gene-level structural conservation. A significantly higher level of sequence divergence was exhibited by the noncoding regions than the coding regions (Fig. 4). Additionally, the SC regions exhibited greater interspecific divergence than the IR regions, particularly the SSC region. These results are similar to studies in other angiosperms [23, 24].

To identify divergence hotspots, the Pi values were calculated. The Pi values ranged between 0 and 0.06548. The LSC region exhibited the maximum nucleotide diversity (average Pi = 0.00149), followed by the SSC region (average Pi = 0.00145), and the IR regions (average Pi = 0.00018). Four highly divergent sites (Pi > 0.02) were detected, including *psbK-psbI* (0.02348), *trnG-UCC-trnT-GGU-trnT-GGU* (0.06548), *rpl16* (0.02157), and *ccsA* (0.02563). Three of these sites were located in the LSC region, and only *ccsA* was located in the SSC region. No divergence hotspots were located in IR regions, suggesting that the plastid substitution rates in IRs may be considerably lower than those of SC regions [24]. These 4 divergence hotspots will provide important resources for studies of species identification, phylogeny, and population genetics in *Bambusa*.

Phylogenetic analysis

The cp genome research allows researchers to examine the phylogenetic relationships among species at different taxonomic levels [25, 26]. The systematic classification of the genus *Bambusa* has been controversial with previous morphological studies suggesting that certain *Bambusa* species, including *B. grandis*, should be included in the genus *Dendrocalamus* [1]. However, our results do not support the separation of *B. grandis* from the genus *Bambusa*. The Maximum Likelihood (ML) and Bayesian inference (BI) methods yielded largely consistent results, as depicted in Fig. 5.

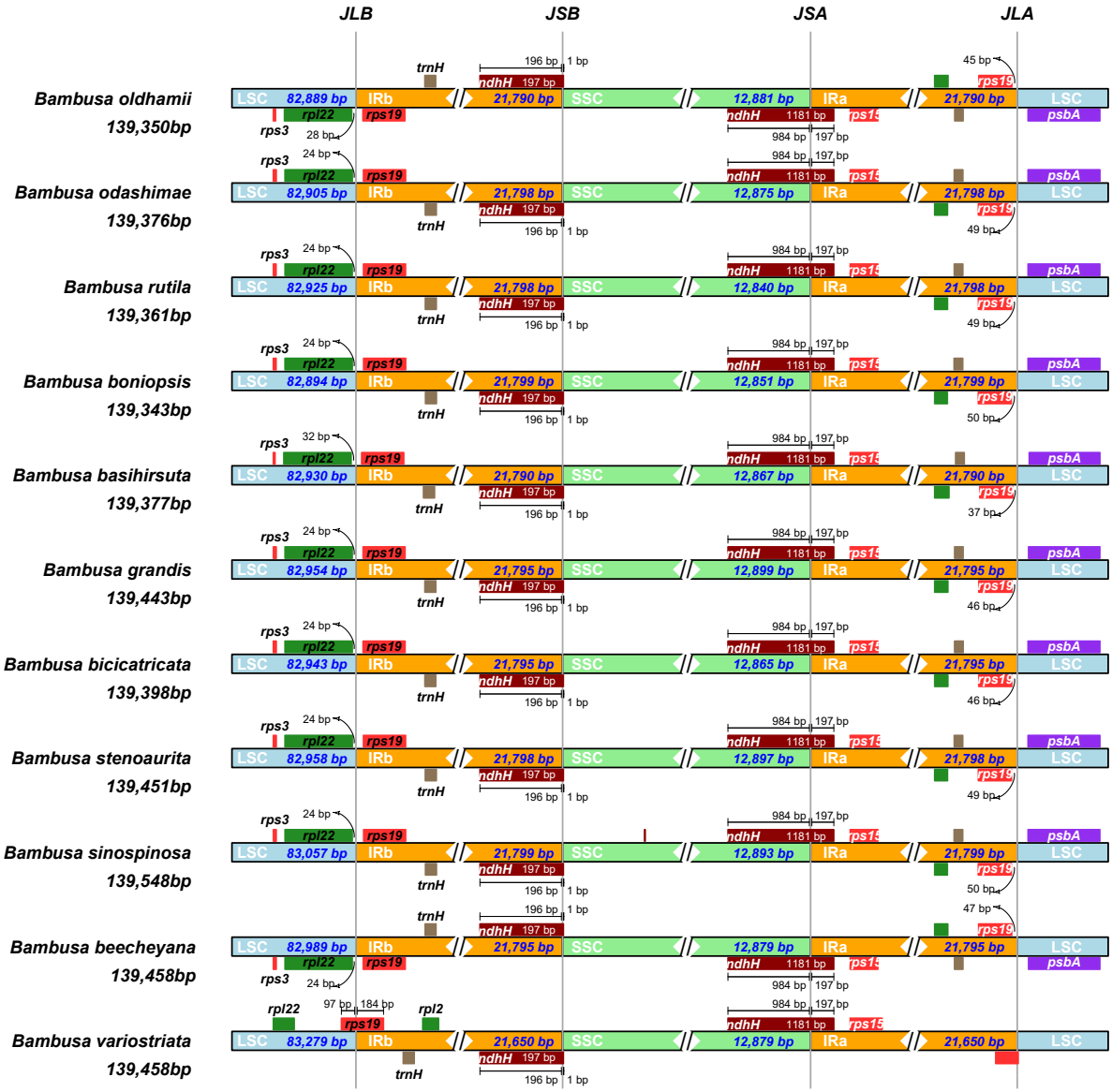

Fig. 3 Comparison of the IR/SC boundaries across 11 *Bambusa* species cp genomes.

In our analysis, which compared the chloroplast (cp) genomes of 26 *Bambusa* species and 2 *Dendrocalamus* species, the phylogenetic tree constructed showed strong support at all 26 nodes for both ML and BI bootstrapping methods. All 26 *Bambusa* species formed 2 monophyletic clades: one clade containing 2 species (*B. beecheyana* var. *pubescens* and *B. variostriata*), and another containing the 24 remaining *Bambusa* species, which contained 4 subbranches. In the new phylogenetic tree, *B. beecheyana* var. *pubescens*, *B. latiflora*, and *B. variostriata* were found to be the most basal species (100/1.00). We also found that *B. grandis* was the sister of *B. basihirsuta*, *B. cornigera*, and *B. intermedia*. Overall, our analysis robustly resolved the phylogenetic relationships among *Bambusa species*.

Genome features and assembly of *B. grandis*

A total of 89.51 Gb of raw reads were obtained from BGISEQ-500 sequencing with a Q20 and Q30 value of 98.33% and 94.79%, respectively. Hence, the genome sequencing date of *B. grandis* is of high quality. Furthermore, we obtained 594,561 scaffolds with a total length of 1,764,468,405 bp. The average length and maximum length were 2,967.7 and 61,664 bp, respectively. The scaffold N50 for the final assembly genome was 5,828 bp. The *B. grandis* genome

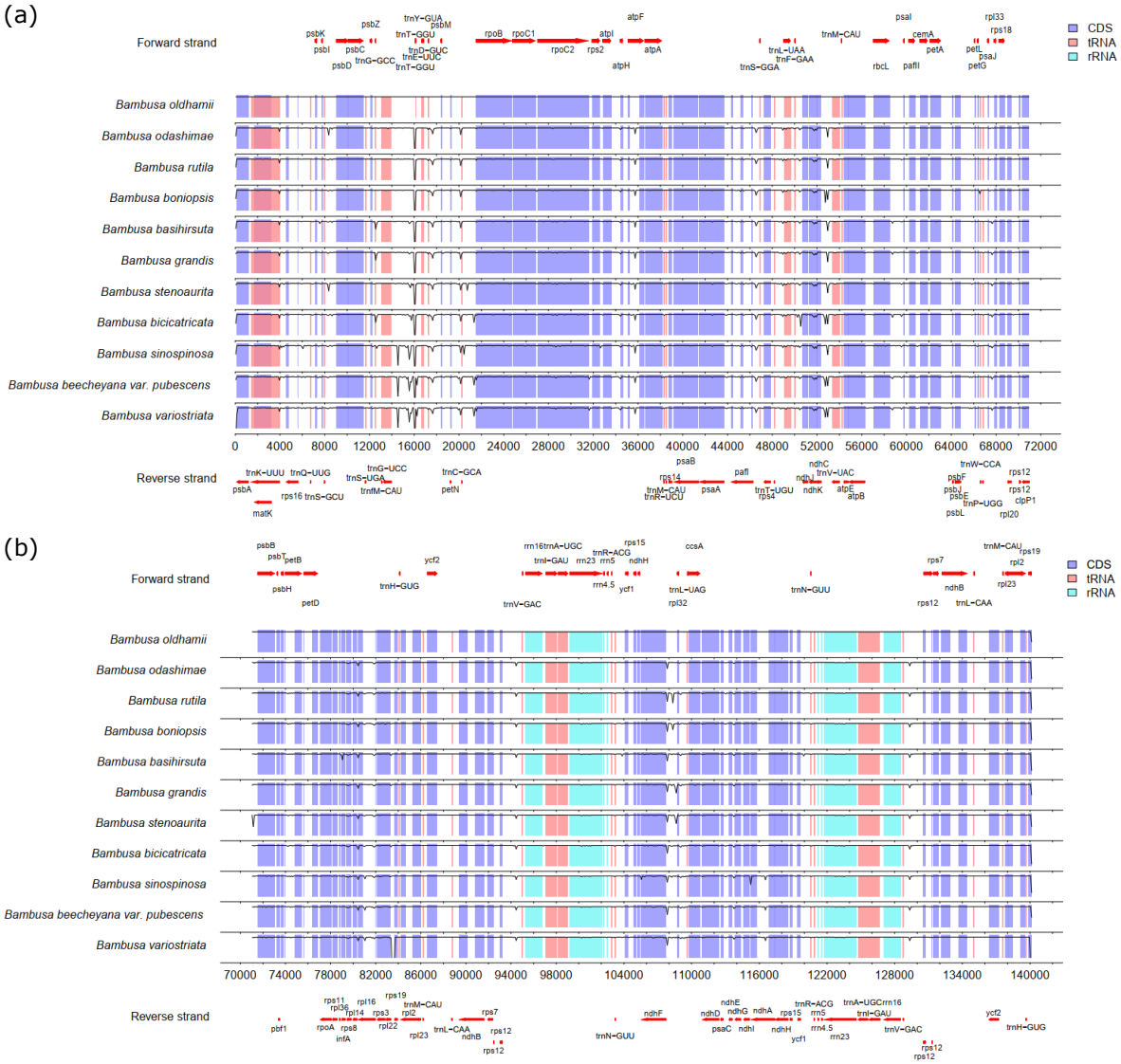

Fig. 4 Variable across homologous regions of *B. grandis* and related *Bambusa* species. The red arrows above each alignment indicate genes and their orientation.

Table 1 Information of the assembled genome sequences of *B. grandis*.

Assembly information	Scaffold
Number of sequence	594,561
Total length (bp)	1,764,468,405
Average length (bp)	2,967.7
Maximum length (bp)	61,664
N50 length (bp)	5,828

size was evaluated using the K-mer method based on BGISEQ-500 short reads, and the result was 1.18 Gb with 55.8% repetitive sequence, 5.63% heterozygosity, and 44.29% GC content (Table 1 and Fig. 6). These genome survey sequencing results offer helpful preliminary information for further whole-genome sequencing and assembly. The genome of *B. grandis* is a complex genome with high levels of heterozygosity and repetition. The use of PacBio long-reading sequencing technology and HiFi assistance to sequence and assemble the reference genome of *B. grandis* would be beneficial for increased continuity, precision, and integrity of high heterozygosity genome assembly [27].

Based on the gene prediction results of *B. grandis*, a comparison was made to protein databases such as COG, GO, KEGG, eggNOG_OGs, Pfam, CAZy, and NR to obtain genome annotation information. The results show that a total of 92,090 genes were annotated with functional information. The average gene length was

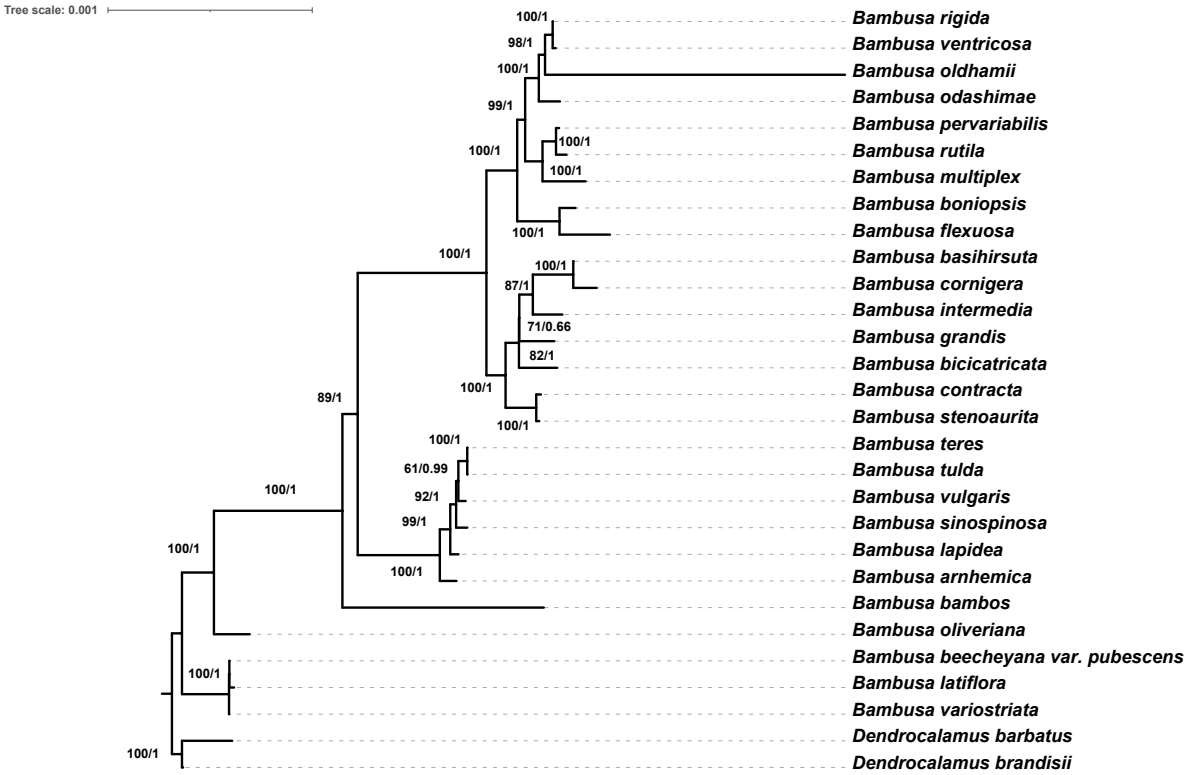

Fig. 5 Phylogenetic tree based on the complete cp genomes of 27 *Bambusa* species. The values preceding the slash ('/') represent the bootstrap support values derived from the Maximum Likelihood method, while the values following the slash represent the bootstrap support values obtained through the Bayesian inference method.

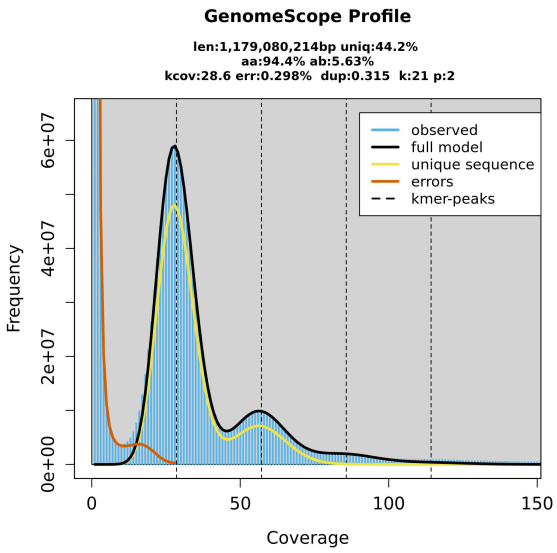

Fig. 6 Distribution of K-mer ($K = 21$).

determined to be 3,002.75 bp with the longest gene length of 116,222 bp. Among them, the NR database annotated the most genes, with 89,036, accounting for 96.68% of the total gene set, followed by the eggNOG database, which annotated a total of 87,144 genes, accounting for 94.63% of the total; the COG and Pfam databases each annotated 77,241 and 76,136 genes, respectively, accounting for 83.88% and 82.68% of the total, respectively; the KEGG and GO databases each annotated 25,215 and 23,391 genes, respectively, accounting for 27.38% and 25.40% of the total, respectively; the CAZy database annotated the fewest genes with only 1,403 genes, accounting for only 1.52% of the gene set.

Microsatellite motif identification

A total of 326,184 microsatellite motifs were identified from 594,561 scaffolds. There are 6 types of motifs, including 1 to 6 nucleotide motifs. Among them, mononucleotide, dinucleotide, and trinucleotide motifs are the most prevalent types of motifs. The proportion of different motifs in repeat units decreased with the increase in nucleotide number, and the number of mononucleotides was 143,001, accounting for 43.84% of the total number. The number of dinucleotides was 134,307, accounting for 41.18%. Trinucleotides and tetranucleotides accounted for 10.84% and 3.33% of the total, respectively. The number of pentanucleotides

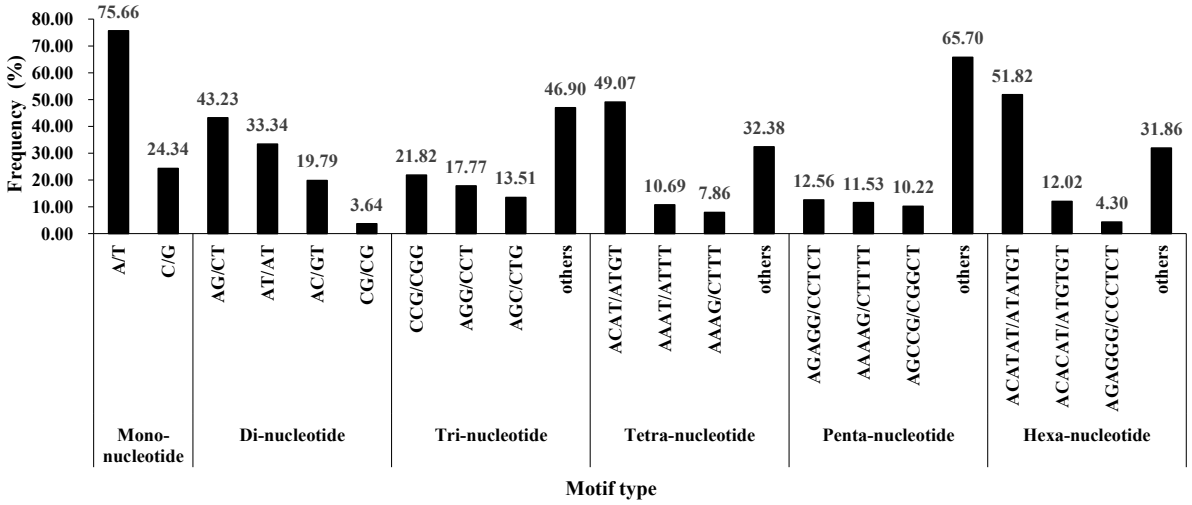

Fig. 7 The distribution and frequency of gSSR motif repeat numbers.

and hexanucleotides accounted for 0.54% and 0.28%, respectively.

In mononucleotides, A/T was the most frequent motif, accounting for 75.66%; among dinucleotides, the most frequent motif was AC/GT (43.23%), followed by AT/AT (33.43%), and AC/GT (19.79%), and other motifs accounted for only 3.64%. Among trinucleotides, the main motif was CCG/CGG, accounting for 21.82%, followed by AGG/CCT and AGC/CTG, accounting for 17.77% and 13.51%, respectively, and the proportion of other motifs was lower than 13.00%. Among tetranucleotides, the main type was ACAT/ATGT, accounting for 49.07%, followed by AAAT/ATTT and AAAG/CTTT, accounting for 10.69% and 7.86%, respectively, and the other motifs accounted for less than 5.00%. Among pentanucleotides, the main type was AGAGG/CCTCT, accounting for 12.56%, followed by AAAAG/CTTTT and AGCCG/CGGCT, accounting for 11.53% and 10.22%, respectively, and the other motifs accounted for less than 10.00%. Among hexanucleotides, the main type was ACATAT/ATATGT, accounting for 51.82%, followed by ACACAT/ATGTGT, accounting for 12.02%, and the proportion of other motifs was lower than 5.00%. It has been suggested that the length variation of the repeating unit is related to the selection pressure. The greater the selection pressure and the lower the copy number, the longer the unit repeats [28]. Therefore, the mutation rate of microsatellites with shorter lengths in the genome is faster, while the mutation rate of long repeat units is slower. In this study, the variation of mononucleotide and dinucleotide motifs was significantly higher than that of other motifs with longer length, indicating that mononucleotide and dinucleotide motif repeats were the most active in the genome of *B. grandis*.

CONCLUSION

In this study, the complete *B. grandis* cp genome was assembled and compared with the cp genomes of several closely related Bambusa species. The complete *B. grandis* cp genome was 139,443 bp in length and exhibited the classical quadripartite circular structure. Multi-genome comparative analyses indicated that the cp genomic sequences exhibited high collinearity and low variation. We found that noncoding sequences exhibited higher divergence than coding sequences. Four regions with high-level polymorphism were discovered with the potential to be used as molecular markers. The phylogenetic analysis resulted in well-supported relationships among all major clades with *B. grandis* most closely related to *B. basihirsuta*, *B. cornigera*, and *B. intermedia*. The genome size of *B. grandis* was 1.18 Gb, and heterozygosity and the repeat ratio were 5.63% and 55.80%, respectively. A total of 326,184 SSRs were identified from the genome survey assembly, and the variation of mononucleotide and dinucleotide motifs was significantly higher than that of other motifs with longer length. The findings will be beneficial to future research on the systematics, genetic diversity, and evolutionary history of the genus *Bambusa* as well as the conservation and sustainable utilization of *B. grandis* genetic resources.

Acknowledgements: This research was funded by the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding, grant number 2021C02070-4; Wenzhou High-level Innovation Team "Coastal Characteristic Plant Innovation and Utilization Project", grant number NY202401; Open Research Subject of Ouhai Science and Technology Innovation Center of Zhejiang Academy of Agricultural Sciences.

REFERENCES

1. Editorial Board of Flora of China, Chinese Academy of Sciences (2006) *Flora of China*. Science Press, Beijing, China.
2. Yang L, Tian J, Xu L, Zhao X, Song Y, Wang D (2022) Comparative chloroplast genomes of six magnoliaceae species Provide new insights into intergeneric relationships and phylogeny. *Biology* **11**, 1279.
3. Gao J, Chen F, Shang S, Shu J, Liu Y, Li W (2020) The complete chloroplast genome sequence of carmine radish (*Raphanus sativus* L.) and its evolutionary implications. *ScienceAsia* **46**, 520–529.
4. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* **19**, 11–15.
5. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, et al (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.
6. Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res* **45**, e18.
7. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
8. Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* **32**, 11–16.
9. Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* **7**, e1002195.
10. Lohse M, Drechsel O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet* **52**, 267–274.
11. Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, 1432.
12. Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829.
13. Keller O, Kollmar M, Stanke M, Waack S (2011) A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**, 757–763.
14. Sharp PM, Li WH (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**, 1281–1295.
15. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **32**, W273–W279.
16. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452.
17. Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **34**, 3030–3031.
18. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evo* **30**, 772–780.
19. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490.
20. Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585.
21. Shen J, Li X, Chen X, Huan X, Jin S (2022) The complete chloroplast genome of *Carya cathayensis* and phylogenetic analysis. *Genes* **13**, 369.
22. Li C, Cai C, Tao Y, Sun Z, Jiang M, Chen L, Li J (2021) Variation and evolution of the whole chloroplast genomes of *Fragaria spp*. (Rosaceae). *Front Plant Sci* **12**, 754209.
23. Huang H, Shi C, Liu Y, Mao SY, Gao LZ (2014) Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol Biol* **14**, 151.
24. Yang Z, Zhao T, Ma Q, Liang L, Wang G (2018) Comparative genomics and phylogenetic analysis revealed the chloroplast genome variation and interspecific relationships of *Corylus* (Betulaceae) species. *Front Plant Sci* **9**, 927.
25. Birky CW (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc Natl Acad Sci USA* **92**, 11331–11338.
26. Wang W, Chen S, Zhang X (2018) Whole-genome comparison reveals divergent IR borders and mutation hotspots in chloroplast genomes of Herbaceous bamboos (Bambusoideae: Olyreae). *Molecules* **23**, 1537.
27. Hon T, Mars K, Young G, Tsai YC, Karalius JW, Landolin JM, Maurer N, Kudrna D, et al (2020) Highly accurate long-read HiFi sequencing data for five complex genomes. *Sci Data* **7**, 399.
28. Jo E, Lee SJ, Choi E, Kim J, Lee SG, Lee JH, Kim JH, Park H (2021) Whole genome survey and microsatellite motif identification of *Artemia franciscana*. *Biosci Rep* **41**, BSR20203868.