



OPEN ACCESS

EDITED BY

Daniel Pinero,
National Autonomous University of Mexico,
Mexico

REVIEWED BY

Tao Zhou,
Xi'an Jiaotong University, China
Zhechen Qi,
Zhejiang Sci-Tech University, China

*CORRESPONDENCE

Ruisen Lu
lurs@cnbg.net

[†]These authors have contributed equally to this work

RECEIVED 12 October 2023

ACCEPTED 27 November 2023

PUBLISHED 12 December 2023

CITATION

Hu K, Chen M, Li P, Sun X and Lu R (2023) Intraspecific phylogeny and genomic resources development for an important medical plant *Dioscorea nipponica*, based on low-coverage whole genome sequencing data. *Front. Plant Sci.* 14:1320473. doi: 10.3389/fpls.2023.1320473

COPYRIGHT

© 2023 Hu, Chen, Li, Sun and Lu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Intraspecific phylogeny and genomic resources development for an important medical plant *Dioscorea nipponica*, based on low-coverage whole genome sequencing data

Ke Hu^{1†}, Min Chen^{1,2†}, Pan Li³, Xiaoqin Sun^{1,2,4} and Ruisen Lu^{1,2*}

¹Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, ²Jiangsu Key Laboratory for the Research and Utilization of Plant Resources, Nanjing, China, ³Laboratory of Systematic & Evolutionary Botany and Biodiversity, College of Life Sciences, Zhejiang University, Hangzhou, China, ⁴Jiangsu Provincial Science and Technology Resources Coordination Platform (Agricultural Germplasm Resources) Germplasm Resources Nursery of Medicinal Plants, Nanjing, China

Dioscorea nipponica Makino, a perennial twining herb with medicinal importance, has a disjunctive distribution in the Sino-Japanese Floristic Region. It has a long history in traditional Chinese medicine, with demonstrated efficacy against various health conditions. However, the limited genomic data and knowledge of genetic variation have hindered its comprehensive exploration, utilization and conservation. In this study, we undertook low-coverage whole genome sequencing of diverse *D. nipponica* accessions to develop both plastome (including whole plastome sequences, plastome-derived SSRs and plastome-divergent hotspots) and nuclear genomic resources (including polymorphic nuclear SSRs and single-copy nuclear genes), as well as elucidate the intraspecific phylogeny of this species. Our research revealed 639 plastome-derived SSRs and highlighted six key mutational hotspots (namely CDS *ycf1*, IGS *trnL-rpl32*, IGS *trnE-trnT*, IGS *rps16-trnQ*, Intron 1 of *clpP*, and Intron *trnG*) within these accessions. Besides, three IGS regions (i.e., *ndhD-cssA*, *trnL-rpl32*, *trnD-trnY*), and the intron *rps16* were identified as potential markers for distinguishing *D. nipponica* from its closely related species. In parallel, we successfully developed 988 high-quality candidate polymorphic nuclear SSRs and identified 17 single-copy nuclear genes for *D. nipponica*, all of which empower us to conduct in-depth investigations into phylogenetics and population genetics of this species. Although our phylogenetic analyses, based on plastome sequences and single-copy nuclear genes revealed cytonuclear discordance within *D. nipponica*, both findings challenged the current subspecies classification. In summary, this study developed a wealth of genomic resources for *D. nipponica* and enhanced our understanding of the intraspecific phylogeny of this species, offering valuable insights that can be instrumental in the conservation and strategic utilization of this economically significant plant.

KEYWORDS

Dioscorea nipponica, plastome-derived markers, polymorphic nSSRs, SCNGs, intraspecific phylogeny

1 Introduction

Dioscorea nipponica Makino, a perennial twining herb of the genus *Dioscorea* belonging to the family Dioscoreaceae, is disjunctively distributed across the Sino-Japanese Floristic Region (Obidiegwu et al., 2020; Yang et al., 2022). This plant is characterized by its slender and cylindrical aerial stem, alternate simple palmate leaves with an anisometric triangular shallow, medium, or deep crack along the leaf edge, and unisexual yellowish-green flowers that droop like small bells (Ding and Gilbert, 2000; Ou-Yang et al., 2018). As it matures, it produces dry capsules that are yellow, obovate-elliptic, and prismatic in shape with winged edges (Ding and Gilbert, 2000). *Dioscorea nipponica* has a long history of use in traditional Chinese medicine, where it has proven effective against various conditions, including rheumatoid arthritis, Kashin-Beck disease, sprains, bruises, chronic bronchitis, and cough (Ou-Yang et al., 2018; Yang et al., 2022). Modern pharmacological research has also revealed the multifaceted attributes of *D. nipponica*, showcasing its wide-ranging benefits, such as its anti-inflammatory, anti-tumor, analgesic, antitussive, calming, and phlegm-dispelling properties (Wu et al., 2023). Notably, recent scientific studies have isolated both fat-soluble and water-soluble steroid saponins from rhizomes of *D. nipponica*, attributing most of its pharmacological effects to saponins and sapogenins (Ou-Yang et al., 2018). Additionally, its aboveground parts contain over ten types of phenanthrene derivatives, further contributing to its medicinal value (Lu et al., 2010; Li et al., 2017). Despite its medicinal significance, a pivotal research gap lies in the limited availability of genomic information, impeding the exploration of new bioactive compounds and a comprehensive understanding of their synthesis. Furthermore, although demand is increasing with growing awareness, the development and utilization of *D. nipponica* resources have been slow (Ou-Yang et al., 2018). The species faces challenges meeting market demand due to the gradual depletion and unrestrained exploitation of wild resources (Chen et al., 2007; Ou-Yang et al., 2018). Therefore, there is an urgent need for additional molecular markers to enhance efforts related to conservation, utilization, and breeding of this economically significant species.

To date, the taxonomic classification and intraspecific phylogenetic relationship of *D. nipponica* have remained subjects of ongoing debate and controversy. Various studies have proposed different perspectives, leading to conflicting perspectives on the subspecies delineation. For instance, Ding and Gilbert (2000) proposed a division of *D. nipponica* into two distinct subspecies, drawing upon considerations such as chromosome count, cork layer characteristics, and geographical distribution. The original subspecies, *D. nipponica* subsp. *nipponica*, is predominantly located in the northern reaches of the Qinling Mountain range, characterized by a rhizome with an easily peelable cork layer and a chromosome count of 20. In contrast, *D. nipponica* subsp. *rosthornii*, the second subspecies, features a persistent cork layer and a chromosome count of 40, primarily inhabiting the southern region of the Qinling Mountains in central China. However, Gao et al. (2008) presented an alternative perspective, suggesting a closer

relationship between *D. nipponica* subsp. *rosthornii* and *D. althaeoides* than with *D. nipponica* subsp. *nipponica*. Their proposition advocated for the elevation of *D. nipponica* subsp. *rosthornii* to the status of an independent species, a viewpoint further corroborated by the chemotaxonomic analysis conducted by Li et al. (2020). Our recent phylogenetic analysis of *Stenophora* species/subspecies based on complete plastome sequences provided strong support for the monophyly of *D. nipponica* subsp. *nipponica* and *D. nipponica* subsp. *rosthornii* (Hu et al., 2023). Nevertheless, it is crucial to acknowledge that all previous studies have been limited by their restricted taxon sampling, typically involving just one specimen each for *D. nipponica* subsp. *nipponica* and *D. nipponica* subsp. *rosthornii* (e.g., Gao et al., 2008; Hu et al., 2023), or relying on a limited set of genetic markers, such as *matK*, *rbcL*, and *trnL-F* (e.g., Gao et al., 2008). Therefore, obtaining a more extensive range of diverse samples from various geographic regions is crucial, with a specific focus on covering both the northern and southern areas of the Qinling Mountains and expanding the scope beyond China's borders. Additionally, employing stronger molecular markers, such as both plastomes and multiple nuclear loci, is vital for conducting intraspecific phylogenetic analysis of this species.

The advent of next-generation sequencing (NGS) technologies has ignited a profound revolution in the acquisition of genome-scale data across a broad spectrum of plant species (Dodsworth, 2015; Lu et al., 2022). Within the realm of NGS methodologies, low-coverage whole genome sequencing, commonly known as genome skimming, has emerged as a notably cost-effective and efficient approach (Straub et al., 2012; Twyford and Ness, 2017; Jin et al., 2020). This technique employs shallow sequencing of genomic DNA, strategically capturing a significant fraction of the genome, with a particular emphasis on high-copy elements such as ribosomal DNA, the plastome, and nuclear repeats like simple sequence repeats (SSRs) and transposable elements (Straub et al., 2012; Dodsworth, 2015; Nevill et al., 2020; Lu et al., 2022). Plastome sequences, among these molecular markers, have proven immensely valuable in plant species identification and phylogenetic studies, thanks to their distinctive characteristics: the absence of recombination, low rates of nucleotide substitutions, small effective population sizes, and typically uniparental inheritance (Birky et al., 1983; Lu et al., 2021). Moreover, recent studies have demonstrated the capacity to recover single-copy nuclear genes (SCNGs) by leveraging multiple assembled nuclear sequences derived from low-coverage whole genome sequencing data (e.g., Liu et al., 2021; Zhou et al., 2022), offering a promising resource for plant phylogenetic analyses.

Thus, to expand genomic resources available for *D. nipponica* and deepen our insights into its intraspecific phylogeny and taxonomy, we conducted low-coverage whole genome sequencing on diverse *D. nipponica* accessions originating from various regions, including both the northern and southern areas of the Qinling Mountains, as well as Japan. Our objectives encompassed the following key aspects: 1) identification of plastome-derived markers, including whole plastome sequences, plastome-derived SSRs and plastome-divergent hotspots; 2) development of genome-wide nuclear markers, encompassing polymorphic nuclear SSRs

(PolynSSRs) and SCNGs; and 3) reconstruction of the phylogenetic relationships among *D. nipponica* accessions based on both plastome and SCNG data. These findings will offer valuable support for the conservation and strategic utilization of this economically significant species.

2 Materials and methods

2.1 Plant materials, DNA extraction and genome sequencing

A sample of eight *D. nipponica* accessions originating from diverse regions, including Beijing (BJ), Gansu (GS), Henan (HeN), Hunan (HuN), Jining (JN) and Zhejiang (ZJ) Provinces of China, as well as Fukushima (FD) and Nagano (NI) Prefectures of Japan was used in this study (Table 1). The sampling strategy was designed to encompass the primary geographical distribution of this species, including both the northern and southern areas of the Qinling Mountains, as well as Japan, and was carried out under the special permission granted by the Institute of Botany, Jiangsu Province and Chinese Academy of Sciences. For each accession, fresh, healthy leaf samples were collected from a wild mature individual, and dried with silica-gel. The voucher specimens were deposited at Herbarium of Institute of Botany, Jiangsu Province and Chinese

Academy of Sciences (NAS). Genomic DNA was extracted from approximately 50 mg of silica-dried leaf samples using DNasecure Plant Kit (Tiangen Biotech, Beijing, China), in accordance with the manufacturer's prescribed protocol. DNA quantity and purity were checked by spectrophotometry and agarose gel electrophoresis.

For each accession, a barcoded paired-end library with an insert size of 350 bp was constructed using NEBNext Ultra DNA Library Prep Kit for Illumina. Following the ligation of indexed adapters, these indexed DNA libraries were pooled and subjected to paired-end sequencing in a single lane of HiSeq X Ten. Subsequently, Trimmomatic v.0.36 (Bolger et al., 2014) was employed to eliminate adaptor sequences, contamination, and low-quality reads from the raw data. This filtering process yielded approximately 4.5 Gb of clean data for each accession. The entire sequence library preparation, genome sequencing, and raw data filtering procedures were conducted by Novogene Bioinformatics Technology Co., Ltd., located in Beijing, China.

2.2 Plastome assembly and annotation

The clean reads from each accession were utilized for plastome sequence assembly using GetOrganelle v.1.7.6 (Jin et al., 2020), with the following parameters: -R 15 -k 21,45,65,85,105 -F embplant_pt. Subsequently, all assembly graphs underwent thorough visual

TABLE 1 Summary of characteristics of eight *Dioscorea nipponica* plastomes.

Characteristics	BJ	GS	HeN	HuN	JN	ZJ	FD	NI
Locality	Beijing, China	Gansu, China	Henan, China	Hunan, China	Jining, China	Zhejiang, China	Fukushima, Japan	Nagano, Japan
Latitude (N)/Longitude (E)	40°02' / 115°28'	35°33' / 106°31'	35°10' / 112°35'	29°24' / 110°29'	41°56' / 126°24'	30°11' / 119°33'	37°38' / 140°05'	35°50' / 137°54'
Total clean reads	30,324,374	31,352,778	23,737,170	24,959,536	32,495,028	31,790,950	33,758,516	28,825,392
Mapped reads	2,699,436	885,479	937,679	1,206,815	463,218	742,377	791,711	553,670
Plastome coverage (×)	2631	863	914	1176	451	723	771	539
Total plastome length (bp)	153,917	153,918	153,917	153,918	153,918	153,963	154,076	154,074
LSC	83,982	83,983	83,982	83,983	83,983	84,028	84,108	84,103
SSC	18,919	18,919	18,919	18,919	18,919	18,919	18,908	18,911
IR	25,508	25,508	25,508	25,508	25,508	25,508	25,530	25,530
Total GC content (%)	37.20%	37.20%	37.20%	37.20%	37.20%	37.20%	37.20%	37.20%
LSC	35.00%	35.00%	35.00%	35.00%	35.00%	35.00%	35.00%	35.00%
SSC	31.20%	31.20%	31.20%	31.20%	31.20%	31.20%	31.30%	31.30%
IR	43.00%	43.00%	43.00%	43.00%	43.00%	43.00%	43.00%	43.00%
Total number of genes	114	114	114	114	114	114	114	114
PCGs	80	80	80	80	80	80	80	80
tRNA genes	30	30	30	30	30	30	30	30
rRNA genes	4	4	4	4	4	4	4	4
Duplicated genes	19	19	19	19	19	19	19	19

examination using Bandage v.0.8.1 (Wick et al., 2015). To further assess the precision of plastome assemblies, two methods were employed including mapping Illumina reads back to the plastome sequence of each *D. nipponica* accession using CLC Genomics Workbench v.23.0.4 (<https://digitalinsights.qiagen.com>), and comparing our assemblies to NOVOPlasty (Dierckxsens et al., 2017) assemblies. The initial annotation of the newly assembled plastomes was conducted with MAFFT v.7 plugin (Katoh and Standley, 2013) within Geneious Prime® 2022.0.1 (<https://www.geneious.com>). This annotation involved aligning the newly assembled plastomes to the previously published plastome of *D. nipponica* (OQ525997), and transferring reference annotations to these newly assembled plastomes. Following this initial annotation, a meticulous manual verification was conducted to ensure the accuracy of exon/intron boundaries and the precise locations of start/stop codons. Finally, high-resolution circular plastome maps were generated using the web-based tool OrganellarGenomeDRAW (OGDRAW) v.1.3.1 (Greiner et al., 2019).

2.3 Characterization of plastome-derived SSRs

The MISA-web application (Beier et al., 2017) was utilized to identify simple sequence repeats (SSRs) within the eight newly assembled plastome sequences of *D. nipponica*. SSR identification criteria were as follows: a minimum threshold of 10 repeat units for mononucleotide SSRs, 5 for dinucleotide SSRs, 4 for trinucleotide SSRs, and 3 for tetra-, penta-, and hexanucleotide SSRs, respectively. Data visualization was performed by the OmicStudio tools at <https://www.omicstudio.cn/tool> (Lyu et al., 2023).

2.4 Identification of plastome-divergent hotspots

To identify mutational hotspots in *D. nipponica* plastomes, three steps were undertaken: firstly, aligning these eight newly assembled plastome sequences using the MAFFT v.7 plugin (Katoh and Standley, 2013) in Geneious Prime® 2022.0.1; secondly, manually extracting coding regions (CDS), introns, and intergenic spacers (IGS) that met specific criteria: (a) the total number of mutation > 0; and (b) an aligned length >200 bp; and finally, subjecting all extracted regions to DnaSP v.6.12.03 (Rozas et al., 2017) for calculating nucleotide diversity (π).

To further develop potential molecular markers for distinguishing *D. nipponica* from its closely related species, i.e., *D. collettii*, *D. gracillima*, *D. villosa*, and *D. zingiberensis* (Hu et al., 2023), a total of six plastome sequences, including two for *D. nipponica* (BJ and FD, representing Chinese and Japanese accession, respectively), and one each for *D. collettii* (OQ525992), *D. gracillima* (OQ525995), *D. villosa* (KY085893), and *D. zingiberensis* (OQ526000) were analyzed. Nucleotide diversity of CDS, introns and IGS regions was computed using the same method described above.

2.5 Development of polymorphic nuclear SSRs

Low-coverage whole genome sequencing data from each *D. nipponica* accession were firstly mapped onto the genome sequences of *D. alata* (Bredeson et al., 2022) and *D. zingiberensis* (Li et al., 2022) to exclude mitochondria and chloroplast reads, using BWA-MEM v.0.7.17 (Li, 2013). The resulting Binary Alignment/Map (BAM) data, which exclusively contained nuclear reads, were *de novo* assembled into scaffolds using the SOAPdenovo2 program (Luo et al., 2012), a de Bruijn graph-based assembly program. Subsequent to nuclear scaffolds generation, the identification of candidate polymorphic nuclear SSRs (PolynSSRs) were conducted using CandiSSR pipeline (Xia et al., 2016), with default parameters. The data visualization was also accomplished using the OmicStudio tools (Lyu et al., 2023).

2.6 Identification of single-copy nuclear genes

To retrieve Angiosperms353 target genes (Johnson et al., 2019) within the genome of *D. nipponica*, low-coverage whole genome sequencing data from each *D. nipponica* accession were independently subjected to HybPiper v. 2.1.6 (Johnson et al., 2016) for assembling sequences for each gene, with all settings at default. Briefly, paired-end clean reads from each accession were mapped to target genes with bwa v. 0.7.17 (Li and Durbin, 2009). Subsequently, mapped reads were organized into distinct directories and assembled into contigs using SPAdes v.3.13.1 (Bankevich et al., 2012). These assembly contigs were then aligned to their associated target sequences using Exonerate v.2.2 (Slater and Birney, 2005). Finally, the recovered gene sequences were extracted using the HybPiper script called retrieve_sequences.py. Additionally, recovery statistics were generated using the two Python scripts, namely get_seq_lengths.py and hybpiper_stats.py, which are included in the HybPiper pipeline (Johnson et al., 2016). The resulting gene sequences shared across all eight *D. nipponica* accessions were imported into Geneious Prime® 2022.0.1, and aligned individually with the multiple alignment plugin MAFFT v.7 (Katoh and Standley, 2013). After excluding those alignments with pairwise identity below 90%, the remaining alignments were concatenated into a supermatrix for nuclear phylogenetic analyses.

2.7 Phylogenetic analyses within *D. nipponica*

For plastome phylogenetic analyses, maximum likelihood (ML) and Bayesian inference (BI) analyses were conducted based on two distinct datasets: complete plastome sequences and a set of 80 shared protein coding regions present in all eight accessions examined in this study (Table 1), with *D. zingiberensis* (OQ526000) as an outgroup. Both complete plastome sequences and protein coding sequences were aligned using MAFFT v.7 plugin

(Katoh and Standley, 2013) in Geneious Prime® 2022.0.1. The best-fitting substitution model, GTR + I + G, for each dataset was determined based on the Akaike Information Criterion (AIC) as computed by jModelTest v.2.1.4 (Darriba et al., 2012). The ML analyses were performed using RAxML v.8.2.12 (Stamatakis, 2014) in the CIPRES Science Gateway v.3.3 (<http://www.phylo.org/portal2/>), with 1000 bootstrap replications. The BI analyses were conducted using MrBayes v.3.2.7 (Ronquist et al., 2012), comprising two independent runs of 1×10^6 generations. Each run employed four independent Markov chain Monte Carlo (MCMC) chains, consisting of one cold chain and three heated chains, with a sampling frequency of 1000 trees. The first 1000 trees were discarded as ‘burn-in’, and the remaining trees were used to construct a majority-rule consensus tree and estimate the posterior probabilities (PPs).

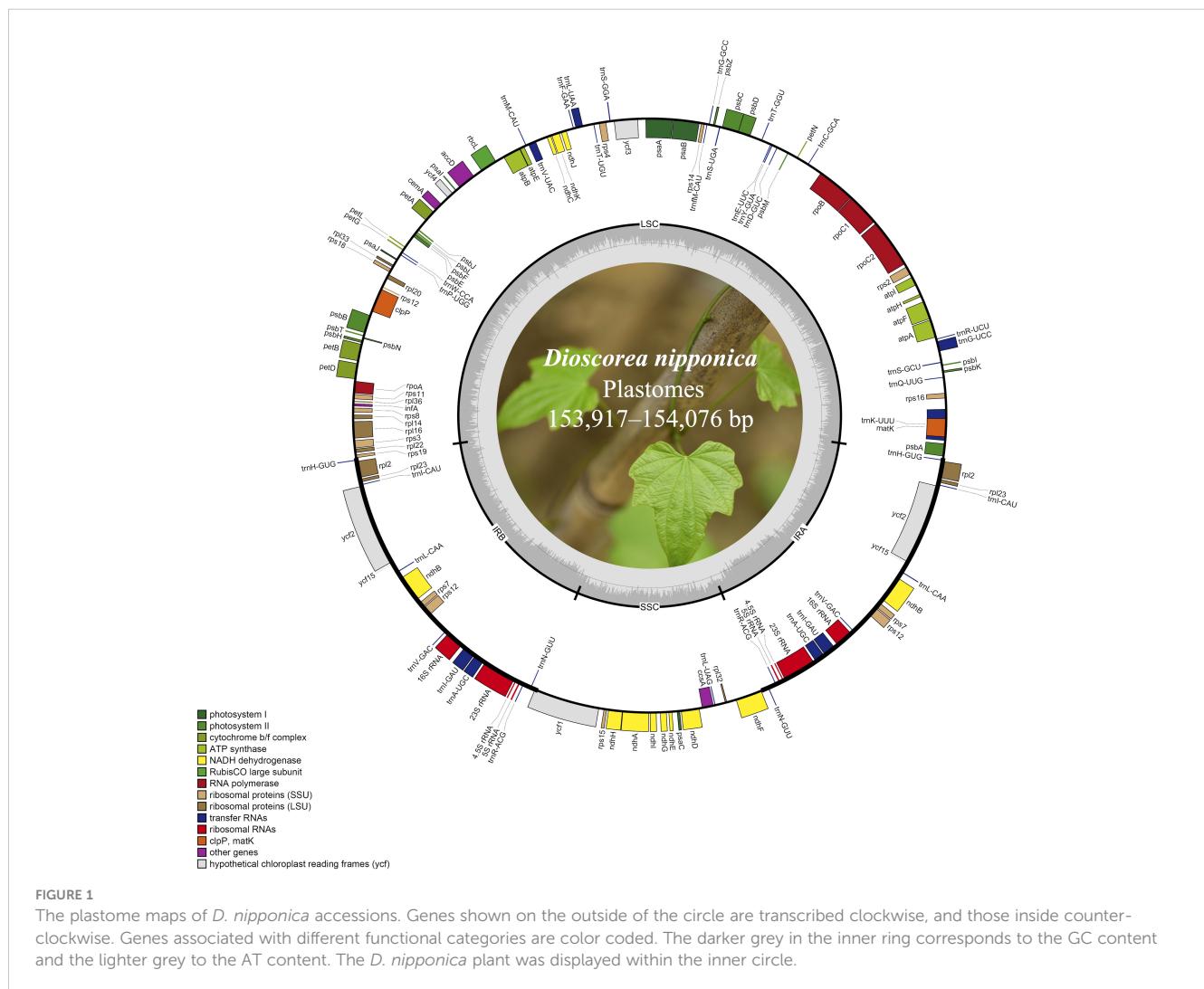
For nuclear phylogenetic analyses, clean reads of *D. zingiberensis* (Cheng et al., 2021) were mapped to the 17 single-copy nuclear genes (17-SCNGs) shared by the eight *D. nipponica* accessions (as detailed in the Results) to yield the 17-SCNGs of *D. zingiberensis*, through CLC Genomics Workbench v.23.0.4 (<https://digitalinsights.qiagen.com>). Subsequently, the 17-SCNGs shared by the eight *D. nipponica* accessions and *D. zingiberensis*, which served

as the outgroup, were aligned using MAFFT v.7 plugin (Katoh and Standley, 2013) in Geneious Prime® 2022.0.1. Both ML and BI analyses were implemented following the same methods described as above for plastome sequences.

3 Results

3.1 Plastome features of *D. nipponica*

Illumina paired-end (150 bp) sequencing produced 23,737,170–33,758,516 clean reads for these eight *D. nipponica* accessions (Table 1). The coverage depths resulting from mapping the Illumina reads to the plastome sequences ranged from 451 \times (JN) to 2631 \times (BJ) (Table 1). Additionally, both GetOrganelle and NOVOPlasty generated identical plastome assemblies. Taken together, these results indicated a high-quality and accuracy of our plastome assemblies. The whole plastome sequences of the eight *D. nipponica* accessions exhibited a narrow range in size, spanning from 153,917 bp (BJ and HeN) to 154,076 bp (FD) (Figure 1; Table 1). The plastome of *D. nipponica* maintained the typical circular quadripartite structure, consisting of a pair of inverted



repeat (IR) regions (25,508–25,530 bp) separated by a large single copy (LSC) region (83,982–84,108 bp), and a small single copy (SSC) region (18,908–18,919 bp) (Table 1). Notably, the lengths of the IR regions in six Chinese accessions were consistent at 25,508 bp, representing a 22 bp reduction when compared to the IR regions in the two Japanese accessions measuring 25,530 bp each (Table 1). The GC content of whole plastome sequences (37.20%), LSC (35.00%) and IR (43.00%) regions were identical across the eight *D. nipponica* accessions, while in the SSC region, the GC content in six Chinese accessions (31.20%) was slightly lower than that in two Japanese accessions (31.30%) (Table 1).

The plastomes of all eight *D. nipponica* accessions exhibited remarkable uniformity in terms of gene count, content, and arrangement. They shared an identical set of 114 unique genes, comprising 80 protein-coding genes (PCGs), 30 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes. Among these, nineteen unique genes (comprising seven PCGs, eight tRNA genes, and all four rRNAs) were duplicated within the IRs, resulting in a total gene count of 133 (Figure 1; Table S1). Within this set of unique genes, nine PCGs (namely *atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, and *rps16*) and six tRNAs (*trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, and *trnV-UAC*) each contained a single intron, while three PCGs (*clpP*, *rps12*, and *ycf3*) harbored two introns (Figure 1; Table S1).

3.2 Plastome-derived SSRs

The MISA analysis detected a total of 639 simple sequence repeats (SSRs) derived from the plastomes of the eight different *D.*

nipponica accessions. The number of plastome-derived SSRs for each accession ranged from 78 (for ZJ) to 83 (for NI). It appeared that the two Japanese accessions (with 82–83 SSRs) contained more SSRs compared to the six Chinese accessions (with 78–80 SSRs) (Figure 2; Table S2). Among these plastome-derived SSRs, mononucleotide repeats were the most prevalent, varying from 41 (for HuN) to 46 (for NI), followed by dinucleotide repeats (15–16 for each accession) and tetranucleotide repeats (10 for each accession). In contrast, trinucleotide repeats (4 for each accession), pentanucleotide repeats (ranging from 4 to 6 for each accession), and hexanucleotide repeats (3 for each accession) were relatively less common in the *D. nipponica* accessions (Figure 2; Table S2). The most frequent motifs observed were A/T and AT/TA for mono- and dinucleotide repeats, constituting 50.63%–53.01% and 15.00%–15.85% of the total plastome-derived SSRs across all eight *D. nipponica* accessions, respectively. Additionally, a set of at least four plastome-derived SSRs, namely (A/T)₁₆, (A/T)₁₈, (C/G)₁₄, and (AATAT/ATATT)₃, could effectively distinguish between the Chinese and Japanese groups among the eight *D. nipponica* accessions (Figure 2; Table S2).

3.3 Plastome-divergent hotspots

Based on the multiple alignment of eight *D. nipponica* plastomes, a total of 50 regions, comprising 15 coding regions (CDS), 26 intergenic spacers (IGS), and 9 intron regions, were isolated for the purpose of computing nucleotide diversity (π). The π values ranged from 0.01% (CDS *rpoC2*) to 0.19% (IGS *trnL-rpl32*) (Figure 3). Among these regions, CDS *ycf1* stood out as the most

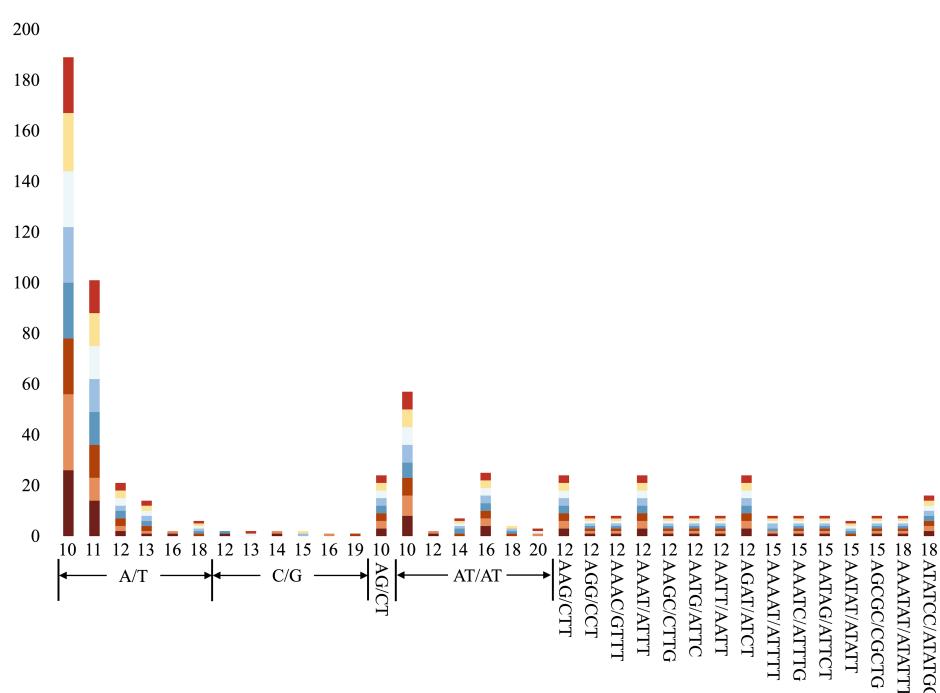


FIGURE 2
The lengths and motif types of plastome-derived SSRs in eight *D. nipponica* accessions.

diverse coding sequence, while the top three intergenic spacers (*trnL-rpl32*, *trnE-trnT*, and *rps16-trnQ*) and two intron regions (Intron 1 of *clpP* and Intron *trnG*) exhibited particularly high levels of variability, each surpassing 0.15% (Figure 3). These six regions hold significant promise for assessing intraspecific genetic variability of *D. nipponica*.

Moving to the inter-specific level, a total of 119 regions (57 CDS, 49 IGS, and 13 introns) were eventually isolated for the calculation of nucleotide diversity, yielding π values ranging from 0.05% (CDS *ycf2*) to 3.05% (IGS *ndhD-cssA*) (Figure 4). Significantly, three IGS regions (i.e., *ndhD-cssA*, *trnL-rpl32*, and *trnD-trnY*), and the intron *rps16* emerged as the four most highly variable regions (with π values exceeding 1.20%) (Figure 4), which could be used for distinguishing *D. nipponica* from its closely related species.

3.4 Polymorphic nuclear SSRs

Utilizing low-coverage whole-genome sequencing data, we generated a range of 1,262,131–1,685,719 scaffolds for each *D. nipponica* accession, with N50 Length varying between 285–317 bp (Table S3). All these scaffolds have been deposited at Dryad under the existing accession number (<https://doi.org/10.5061/dryad.05qfttf96>). Based on these scaffolds, a total of 2417 candidate polymorphic nuclear SSRs (PolynSSRs) were initially generated for *D. nipponica*. After filtering out those low-quality PolynSSRs with a missing rate (MR) ≥ 0.5 and transferability (similarity) $< 95\%$, a subset of 988 high-quality candidate PolynSSRs were remained (Figure 5; Table S4). Within this group of high-quality candidate PolynSSRs, it was possible to design primers for 980 of them, constituting 99.19% of the total (Table

S4). Notably, among 988 high-quality candidate PolynSSRs, dinucleotide repeats (468, 47.37%) were the most abundant, followed by trinucleotide (380, 38.46%) and tetranucleotide (112, 11.34%) repeats (Figure 5). In contrast, pentanucleotide repeats (19, 1.92%) and hexanucleotide repeats (9, 0.91%) were the least frequently encountered types (Figure 5; Table S4).

3.5 Single-copy nuclear genes

Clean reads from these eight accessions were collectively mapped to 351–352 genes, yielding 37 to 117 sequences for each respective accession (Table 2). Notably, among these sequences, 18 were found to be shared across all eight accessions. Following the exclusion of one alignment with pairwise identity below 90%, a supermatrix with a total length of 2154 bp was formed by concatenating the remaining 17 gene sequences. This supermatrix was subsequently utilized for nuclear phylogenetic analyses. However, it's worth emphasizing that only a handful of sequences (6 for BJ, 19 for GS, 3 for HeN, 2 for HuN, 25 for JN, 27 for ZJ, 28 for FD, and 18 for NI, respectively), each with a length comprising at least 50% of the target, were successfully recovered (Table 2).

3.6 Intraspecific phylogenetic relationship of *D. nipponica*

Both the ML and BI analyses, conducted using complete plastome sequences and 80 shared protein coding regions, provided robust support for the division of *D. nipponica* into distinct Chinese and Japanese groups (Figure 6A), with high bootstrap values (BS values = 100%) and posterior probabilities

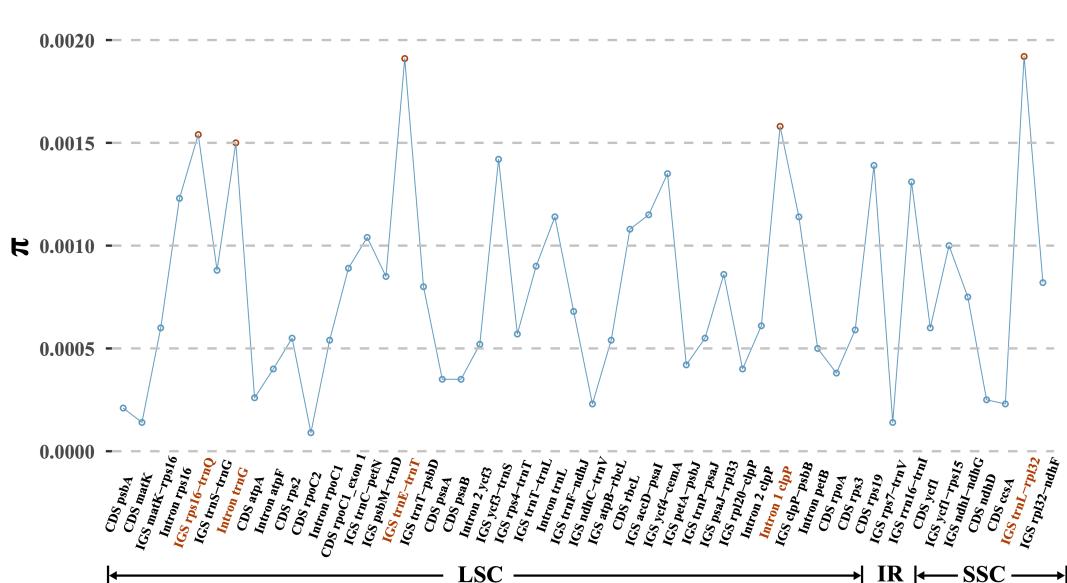
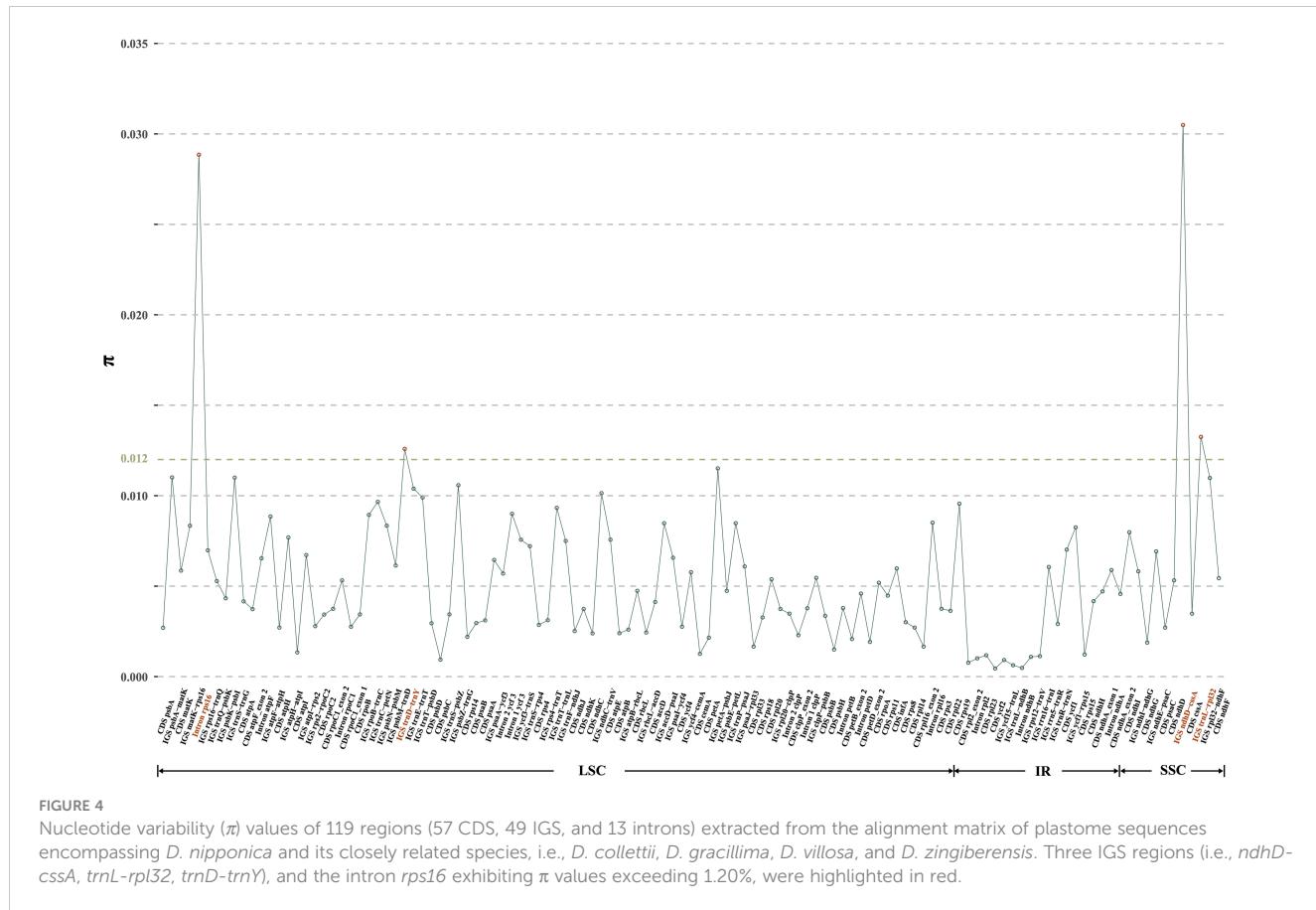


FIGURE 3

FIGURE 3
Nucleotide variability (π) values of 50 regions (15 CDS, 26 IGS, and 9 introns) extracted from the alignment matrix of plastome sequences encompassing eight *D. nipponica* accessions. Three IGS regions (*trnL-rpl32*, *trnE-trnT*, and *rps16-trnQ*), along with two introns (Intron 1 of *clpP* and Intron of *trnG*), exhibiting π values exceeding 0.15% were highlighted in red.



(PPs = 1.0). Notably, no evident genetic distinction was observed within the Chinese group (Figure 6A). However, when phylogenetic analyses were conducted based on 17 single-copy nuclear genes, they indicated that the two Japanese accessions formed a monophyletic group with two Chinese accessions (HuN and ZJ) from Central and South China. These, in turn, constituted a sister group with the other four Chinese accessions (BJ, GS, JN, and HeN) from North China (Figure 6B).

4 Discussion

4.1 Plastome characteristics and plastome-derived markers of *D. nipponica*

Plastomes have become a cornerstone in the realm of plant phylogenetics and evolutionary studies, primarily due to their unwavering preservation of gene order and the striking absence of heteroplasmy and recombination (Birk et al., 1983; Daniell et al., 2016). Notably, plastomes exhibit uniparental inheritance, typically maternal in angiosperms and paternal in gymnosperms, offering a distinctive avenue for unraveling the respective contributions of seed and pollen dispersal to the genetic makeup of natural populations—a perspective enriched when contrasted with nuclear markers (Birk, 1995; Mohammad-Panah et al., 2017). In this study, the investigation of plastome features in eight *D. nipponica* accessions unveiled a striking uniformity in terms of

gene count, content, and arrangement (Figure 1; Table 1), which suggested an enduring evolutionary stability within *D. nipponica*.

The differences in the sizes of inverted repeats (IRs) and the variations observed at four distinct junctions (IRa/SSC, IRa/LSC, IRb/SSC, and IRb/LSC) often significantly influence the overall size of the plastomes and the count of genes, across angiosperm species (Sun et al., 2016). Remarkably, within a single species, the lengths of IR regions were always consistently identical among different individuals or accessions (e.g., *Zizania latifolia*, Lu et al., 2022; *Dioscorea alata*, Lu et al., 2023), playing a crucial role in stabilizing plastomes (Blazier et al., 2016). Our study, in contrast to earlier findings indicating uniform IR length within a specific species, revealed discrepancies in the length of IR regions between Chinese and Japanese accessions (Table 1). This discovery may signify historical divergence or unique evolutionary history of *D. nipponica* accessions in these distinct geographic regions (Jansen and Ruhman, 2012).

Recent studies have underscored the role of simple sequence repeats (SSRs) in adaptation, survival, and evolution of species (Labbé et al., 2011; Yuan et al., 2021). The differences observed in plastome-derived SSRs between Chinese and Japanese accessions (Figure 2), hold significance in understanding the genetic variations and evolutionary history of *D. nipponica*. It is plausible that geographic isolation, differing environmental factors, and historical events have led to genetic differences between these distinct geographical accessions (Qiu et al., 2011). Further exploration of these variations and their correlation with

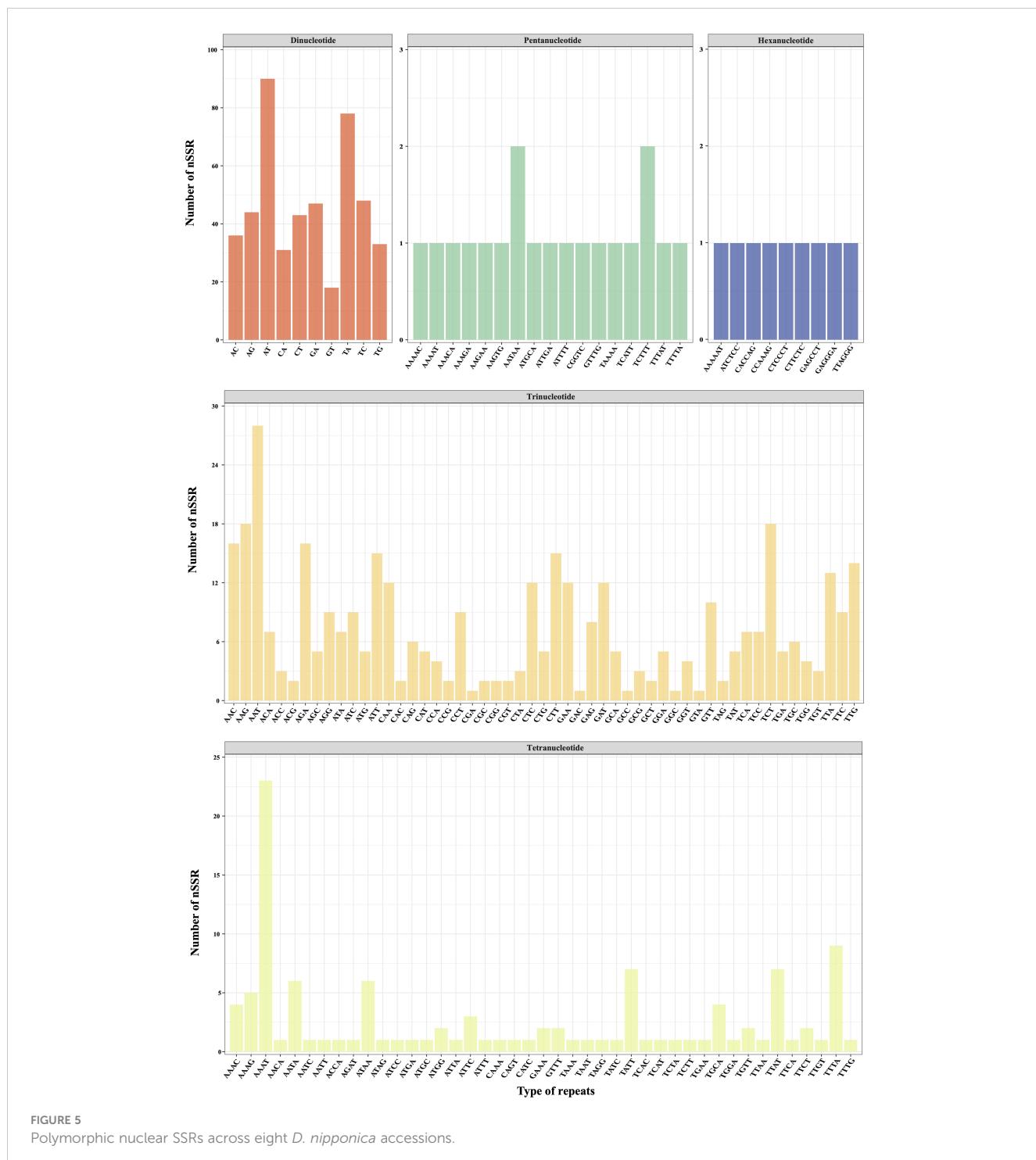


FIGURE 5
Polymorphic nuclear SSRs across eight *D. nippónica* accessions.

historical events and/or environmental factors, particularly within the context of a population genetic framework, could provide valuable insights into the evolutionary history and population dynamics of *D. nipponica*.

Due to their smaller effective population sizes compared to nuclear genomes, as well as the limited gene dispersal via seeds as opposed to pollen-mediated gene flow, plastome-derived markers have the potential to serve as effective indicators for historical bottlenecks, founder effects, and genetic drift (Mohammad-Panah

et al., 2017). Regrettably, when this study was initiated, plastome-derived markers tailored for *D. nippónica* were very limited. The present study successfully identified a significant array of plastome-derived SSRs, ranging from 78 for ZJ to 83 for NI (Figure 2; Table S2), as well as six mutational hotspots (CDS *ycf1*, *trnL-rpl32*, *trnE-trnT*, *rps16-trnQ*, Intron 1 of *clpP*, and Intron *trnG*) (Figure 3). These newly discovered plastome-derived markers hold substantial promise as valuable tools for population genetics investigations and phylogenetic analyses of *D. nippónica*.

TABLE 2 The recovery efficiency of Angiosperms353 target genes in eight *D. nipponica* accessions.

Accession	NumReads	ReadsMap-ped	GenesMap-ped	GenesWithSeq	GenesAt25pct	GenesAt50-pct
BJ	30324374	445246	352	47	14	6
GS	31352778	409836	352	94	44	19
HeN	23737170	325236	352	39	6	3
HuN	24959536	332371	352	37	7	2
JN	32495028	449107	352	116	48	25
ZJ	31790950	506942	352	103	34	27
FD	33758516	540743	352	117	49	28
NI	28825392	482733	351	96	42	18

NumReads: total number of input reads in the *.fastq files provided; ReadsMapped: total number of input reads that mapped to sequences in the target file; GenesMapped: Number of genes in the target file that had reads mapped to their representative sequences; GenesWithSeq: Number of genes with sequences; GenesAt25pct and GenesAt50pct: Number of genes with sequences > 25% and 50% of the mean target length, respectively.

4.2 Nuclear genomic resources for *D. nipponica*

Nuclear SSR markers (nSSRs) have proven to be invaluable tools in various aspects, such as population genetic analyses, identification of germplasm resources, and marker-assisted breeding programs, due to their co-dominant inheritance, adherence to Mendelian inheritance principles, wide genomic distribution, high polymorphism, and verifiable neutrality (Kaldate et al., 2017; Lu et al., 2022). In this study, we made a noteworthy discovery of 988 high-quality candidate polymorphic nSSRs (Figure 5; Table S4), which hold immense potential in shedding light on the genetic diversity and population structure of *D. nipponica*. More importantly, conducting comparative analyses of nSSRs and plastome-derived SSRs may provide complementary and sometimes contrasting perspectives on the genetic structure, differentiation, and gene flow (pollen- and seed-mediated) among *D. nipponica* populations (Mohammad-Panah

et al., 2017). Among these molecular markers, five polymorphic nSSRs (i.e., nSSR_901, nSSR_1065, nSSR_1163, nSSR_1491 and nSSR_2102) showing differentiation between BJ and FD, along with five plastome-derived SSRs were selected for validation using PCR-based Sanger sequencing (Supplementary Data). The sequence validation showed 100% similarity, affirming the accuracy and reliability of these identified molecular markers. Undoubtedly, the genomic resources described here formed a foundational platform for future studies in population genetics, evolution, and breeding of *D. nipponica*, crucial for developing effective conservation strategies, understanding the genetic basis of adaptation, and designing suitable breeding programs.

Single-copy nuclear genes (SCNGs), characterized by bi-parental inheritance, higher evolutionary rates, and numerous unlinked loci (Alvarez et al., 2008; Hojjati et al., 2019), offer significant potential for addressing issues related to hybridization and incomplete lineage sorting, potentially reconciling discrepancies among plastome genes (Small et al., 2004; Doyle, 2022). However, the recovery of SCNGs in

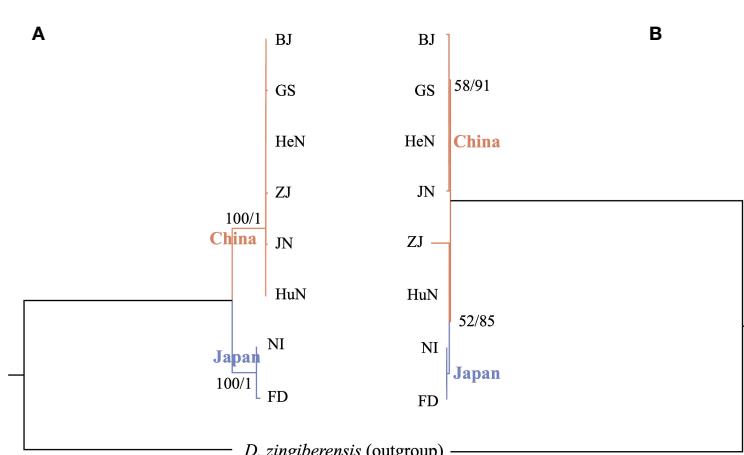


FIGURE 6

Intraspecific phylogenetic trees of *D. nipponica* inferred from (A) complete plastome sequences and 80 shared protein coding regions, as well as (B) 17 single copy nuclear genes, based on the methods of maximum likelihood (ML) and Bayesian inference (BI). The ML bootstrap values/BI posterior probabilities were displayed above the lines.

this study was constrained (Table 2), possibly due to low sequencing coverage. Our previous flow cytometry analysis indicated genome sizes of ~550 Mb and ~1.10 Gb for diploids ($2n = 20$) and tetraploids ($2n = 40$), respectively (detailed data not shown), thus the total clean data generated in this study (Table 1) represented about 8x and 4x coverage of the estimated diploid and tetraploid genomes, respectively. Although the sequencing coverage was somewhat lower than optimal minimum sequencing depth (10x) for high-quality SCNG assembly via low-coverage genome sequencing (Liu et al., 2021), the successful retrieval of 17 SCNGs from diverse *D. nipponica* accessions underscored their potential in resolving phylogenetic relationships within this species. Efforts aimed at obtaining more comprehensive SCNG sequences, possibly through improved sequencing methodologies or increased coverage, will be imperative. Addressing these limitations will undoubtedly enhance the accuracy and robustness of phylogenetic reconstructions, providing a more nuanced understanding of the intraspecific phylogenetic relationships of *D. nipponica*.

4.3 Intraspecific phylogenetic relationship of *D. nipponica*

The taxonomic classification and intraspecific phylogenetic relationships within *D. nipponica* have long been the subject of ongoing debate and controversy (Ding and Gilbert, 2000; Gao et al., 2008; Li et al., 2020; Hu et al., 2023). However, previous taxonomic and phylogenetic studies concerning *D. nipponica* have often grappled with limitations, either relying on short DNA sequences or featuring a restricted sample of taxa, resulting in constrained and sometimes conflicting conclusions (Gao et al., 2008; Li et al., 2020; Hu et al., 2023). Although in this study, phylogenetic analyses based on plastome sequences and concatenated SCNG data, revealed cytonuclear discordance within *D. nipponica*: the former supported the differentiation of *D. nipponica* accessions into distinct Chinese and Japanese groups (Figure 6A), while the latter suggested that the Japanese accessions formed a monophyletic group with two Chinese accessions (HuN and ZJ) from Central and South China (Figure 6B), both findings challenged the current subspecies classification as per the Flora of China (Ding and Gilbert, 2000) and the proposition to elevation of *D. nipponica* subsp. *rosthornii* to the status of an independent species (Gao et al., 2008; Li et al., 2020). This cytonuclear discordance may be attributed to both ancient and recent hybridization events as well as polyploidization occurrences within this species (Ding and Gilbert, 2000). Alternatively, it could be caused by the lack of phylogenetic resolution in the SCNG data. The nuclear gene-based phylogeny is inadequate in offering definitive insights into intraspecies relationships within *D. nipponica* (see Figure 6B), highlighting the inherent challenges of resolving deeper phylogenetic relationships with a limited set of nuclear genes. Thus, to gain a more comprehensive understanding of the taxonomy and evolutionary history of *D. nipponica*, it is imperative to utilize more robust molecular markers, such as nuclear SNPs, and conduct broader sampling across various regions.

5 Conclusions

In this study, we conducted low-coverage whole genome sequencing of diverse *D. nipponica* accessions, to retrieve plastome information, including whole plastome sequences, plastome-derived SSRs and plastome-divergent hotspots, as well as nuclear genomic markers, including polymorphic nuclear SSRs and single-copy nuclear genes. Our findings revealed a striking uniformity in plastome features across diverse *D. nipponica* accessions, with subtle length differences in inverted repeat regions between Chinese and Japanese accessions. A total of 639 plastome-derived SSRs and six divergent hotspots were identified from *D. nipponica* plastomes. Besides, four highly divergent hotspots were developed as potential markers for distinguishing *D. nipponica* from its closely related species. In parallel, 988 high-quality candidate polymorphic nuclear SSRs and 17 single-copy nuclear genes were obtained. The genomic resources identified here will aid in the conservation and strategic utilization of this economically significant plant. Furthermore, the study shed light on the intraspecific phylogenetic relationships of *D. nipponica*, challenging the current subspecies classification and highlighting the need for further taxon sampling and the integration of more robust molecular markers to thoroughly unravel the intraspecific relationship and evolutionary history of this species.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

KH: Software, Writing – original draft. MC: Software, Writing – original draft. PL: Resources, Writing – review & editing. XS: Resources, Writing – review & editing. RL: Conceptualization, Resources, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China (32200192), the Jiangsu Key Laboratory for the Research and Utilization of Plant Resources (JSPKLB202206), the Jiangsu Institute of Botany Talent Fund (JIBTF202102), the Jiangsu Innovative and Entrepreneurial Talent Programme (JSSCBS20211311), and the Jiangsu Forestry Science and Technology Innovation and Promotion Project (LYKJ[2020]08).

Acknowledgments

The authors thank Ichiro Tamaki, Daiki Takahashi and Shota Sakaguchi for their great help in collecting plant materials.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1320473/full#supplementary-material>

References

- Alvarez, I., Costa, A., and Feliner, G. N. (2008). Selecting single-copy nuclear genes for plant phylogenetics: a preliminary analysis for the Senecioneae (Asteraceae). *J. Mol. Evol.* 66, 276–291. doi: 10.1007/s00239-008-9083-7
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198
- Birky, C. W. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci.* 92, 11331–11338. doi: 10.1073/pnas.92.25.11331
- Birky, C. W., Maruyama, T., and Fuerst, P. (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* 103, 513–527. doi: 10.1093/genetics/103.3.513
- Blazier, J. C., Jansen, R. K., Mower, J. P., Govindu, M., Zhang, J., Weng, M. L., et al. (2016). Variable presence of the inverted repeat and plastome stability in Erodium. *Ann. Bot.* 117, 1209–1220. doi: 10.1093/aob/mcw065
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bredeson, J. V., Lyons, J. B., Oniyinde, I. O., Okereke, N. R., Kolade, O., Nnabue, I., et al. (2022). Chromosome evolution and the genetic basis of agronomically important traits in greater yam. *Nat. Commun.* 13, 2001. doi: 10.1038/s41467-022-29114-w
- Chen, F. Q., Fu, Y., Wang, D. L., Gao, X., and Wang, L. (2007). The effect of plant growth regulators and sucrose on the micropropagation and microtuberization of *Dioscorea nipponica* Makino. *J. Plant Growth. Regul.* 26, 38–45. doi: 10.1007/s00344-005-0147-2
- Cheng, J., Chen, J., Liu, X., Li, X., Zhang, W., Dai, Z., et al. (2021). The origin and evolution of the diosgenin biosynthetic pathway in yam. *Plant Commun.* 2, 100079. doi: 10.1016/j.xplc.2020.100079
- Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 1–29. doi: 10.1186/s13059-016-1004-2
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109
- Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45, e18–e18. doi: 10.1093/nar/gkw955
- Ding, Z., and Gilbert, M. G. (2000). “Dioscoreaceae,” in *Flora of China*. Eds. Z. Wu and P. H. Raven (Beijing: Science Press & St. Louis: Missouri Botanical Garden Press), 276–296.
- Dodsworth, S. (2015). Genome skimming for next-generation biodiversity analysis. *Trends Plant Sci.* 20, 525–527. doi: 10.1016/j.tplants.2015.06.012
- Doyle, J. J. (2022). Defining coalescent genes: theory meets practice in organelle phylogenomics. *Syst. Biol.* 71, 476–489. doi: 10.1093/sysbio/syab053
- Gao, X., Zhu, Y. P., Wu, B. C., Zhao, Y. M., Chen, J. Q., and Hang, Y. Y. (2008). Phylogeny of *Dioscorea* sect. *Stenophora* based on chloroplast *matK*, *rbcL* and *trnL*-F sequences. *J. Syst. Evol.* 46, 315–321. doi: 10.3724/SP.J.1002.2008.08007
- Greiner, S., Lehwerk, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, 59–64. doi: 10.1093/nar/gkz238
- Hojjati, F., Adams, R. P., and Terry, R. G. (2019). Discovery of chloroplast capture in *Juniperus excelsa* complex by multi-locus phylogeny. *Phytotaxa* 413, 11–26. doi: 10.11646/phytotaxa.413.1.2
- Hu, K., Sun, X. Q., Chen, M., and Lu, R. S. (2023). Low-coverage whole genome sequencing of eleven species/subspecies in *Dioscorea* sect. *Stenophora* (Dioscoreaceae): comparative plastome analyses, molecular markers development and phylogenetic inference. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1196176
- Jansen, R. K., and Ruhrlman, T. A. (2012). “Plastid genomes of seed plants,” in *Genomics of chloroplasts and mitochondria*. Eds. R. Bock and V. Knoop (Dordrecht: Springer), 103–126.
- Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., DePamphilis, C. W., Yi, T. S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21, 1–31. doi: 10.1186/s13059-020-02154-5
- Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., et al. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* 4, 1600016. doi: 10.3732/apps.1600016
- Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., et al. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Bot.* 68, 594–606. doi: 10.1093/sysbio/syy086
- Kaldate, R., Rana, M., Sharma, V., Hirakawa, H., Kumar, R., Singh, G., et al. (2017). Development of genome-wide SSR markers in horsegram and their use for genetic diversity and cross-transferability analysis. *Mol. Breed.* 37, 1–10. doi: 10.1007/s11032-017-0701-1
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Labbé, J., Murat, C., Morin, E., Le Tacon, F., and Martin, F. (2011). Survey and analysis of simple sequence repeats in the *Laccaria bicolor* genome, with development of microsatellite markers. *Curr. Genet.* 57, 75–88. doi: 10.1007/s00294-010-0328-9
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. doi: 10.48550/arXiv.1303.3997
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, X., Zhao, C., Jing, S., Sun, J., Li, X., Zhao, C., Zhao, P., et al. (2017). Novel phenanthrene and isocoumarin from the rhizomes of *Dioscorea nipponica* Makino subsp. *rosthornii* (Prain et Burkil) CT Ting (Dioscoreaceae). *Bioorg. Med. Chem. Lett.* 27, 3595–3601. doi: 10.1016/j.bmcl.2017.03.095
- Li, X., Wang, Y., Sun, J., Li, X., Zhao, C., Zhao, P., et al. (2020). Chemotaxonomic studies of 12 *Dioscorea* species from China by UHPLC-QTOF-MS/MS analysis. *Phytochem. Anal.* 31, 164–182. doi: 10.1002/pca.2876
- Li, Y., Tan, C., Li, Z. H., Guo, J. Z., Li, S., Chen, X., et al. (2022). The genome of *Dioscorea zingiberensis* sheds light on the biosynthesis, origin and evolution of the medicinally important diosgenin saponins. *Hortic. Res.* 9, 165. doi: 10.1093/hr/uhacl165
- Liu, L. X., Zhang, Y. H., and Li, P. (2021). Development of genomic resources for the genus *Celtis* (Cannabaceae) based on genome skimming data. *Plant Diversity* 43, 43–53. doi: 10.1016/j.pld.2020.09.005
- Lu, R., Chen, M., Feng, Y., Yuan, N., Zhang, Y., Cao, M., et al. (2022). Comparative plastome analyses and genomic resource development in wild rice (*Zizania* spp., Poaceae) using genome skimming data. *Ind. Crop Prod.* 186, 115244. doi: 10.1016/j.indcrop.2022.115244
- Lu, D., Liu, J. P., Li, H. J., and Li, P. Y. (2010). Phenanthrene derivatives from the stems and leaves of *Dioscorea nipponica* Makino. *J. Asian Nat. Prod. Res.* 12, 1–6. doi: 10.1080/10286020903403626

- Lu, R. S., Yang, T., Chen, Y., Wang, S. Y., Cai, M. Q., Cameron, K. M., et al. (2021). Comparative plastome genomics and phylogenetic analyses of Liliaceae. *Bot. J. Linn. Soc.* 196, 279–293. doi: 10.1093/botlinean/boaa109
- Lu, R. S., Hu, K., Zhang, F. J., Sun, X. Q., Chen, M., Zhang, Y. M., et al. (2023). Pan-plastome of greater yam (*Dioscorea alata*) in China: intraspecific genetic variation, comparative genomics, and phylogenetic analyses. *Int. J. Mol. Sci.* 24, 3341. doi: 10.3390/ijms24043341
- Luo, R. B., Liu, B. H., Xie, Y. L., Li, Z. Y., Huang, W. H., Yuan, J. Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1, 18. doi: 10.1186/2047-217X-1-18
- Lyu, F., Han, F., Ge, C., Mao, W., Chen, L., Hu, H., et al. (2023). OmicStudio: a composable bioinformatics cloud platform with real-time feedback that can generate high-quality graphs for publication. *iMeta* 2, e85. doi: 10.1002/imt2.85
- Mohammad-Panah, N., Shabani, N., Khadivi, A., Rahmani, M. S., and Emami, A. (2017). Genetic structure of gall oak (*Quercus infectoria*) characterized by nuclear and chloroplast SSR markers. *Tree Genet. Genomes* 13, 1–12. doi: 10.1007/s11295-017-1146-8
- Neville, P. G., Zhong, X., Tonti-Filippini, J., Byrne, M., Hislop, M., Thiele, K., et al. (2020). Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics. *Plant Methods* 16, 1–8. doi: 10.1186/s13007-019-0534-5
- Obidiegwu, J. E., Lyons, J. B., and Chilaka, C. A. (2020). The *Dioscorea* genus (yam)—an appraisal of nutritional and therapeutic potentials. *Foods* 9, 1304. doi: 10.3390/foods9091304
- Ou-Yang, S. H., Jiang, T., Zhu, L., and Yi, T. (2018). *Dioscorea nipponica* Makino: a systematic review on its ethnobotany, phytochemical and pharmacological profiles. *Chem. Cent. J.* 12, 1–18. doi: 10.1186/s13065-018-0423-4
- Qiu, Y. X., Fu, C. X., and Comes, H. P. (2011). Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Mol. Phylogenet. Evol.* 59, 225–244. doi: 10.1016/j.ympev.2011.01.012
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248
- Slater, G. S. C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinf.* 6, 1–11. doi: 10.1186/1471-2105-6-31
- Small, R. L., Cronn, R. C., and Wendel, J. F. (2004). Use of nuclear genes for phylogeny reconstruction in plants. *Aust. Syst. Bot.* 17, 145–170. doi: 10.1071/SB03015
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Straub, S. C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., and Liston, A. (2012). Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am. J. Bot.* 99, 349–364. doi: 10.3732/ajb.1100335
- Sun, Y., Moore, M. J., Zhang, S., Soltis, P. S., Soltis, D. E., Zhao, T., et al. (2016). Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. *Mol. Phylogenet. Evol.* 96, 93–101. doi: 10.1016/j.ympev.2015.12.006
- Twyford, A. D., and Ness, R. W. (2017). Strategies for complete plastid genome sequencing. *Mol. Ecol. Resour.* 17, 858–868. doi: 10.1111/1755-0998.12626
- Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383
- Wu, C., Zhang, R., Wang, J., Chen, Y., Zhu, W., Yi, X., et al. (2023). *Dioscorea nipponica* Makino: a comprehensive review of its chemical composition and pharmacology on chronic kidney disease. *Biomed. Pharmacother.* 167, 115508. doi: 10.1016/j.bioph.2023.115508
- Xia, E. H., Yao, Q. Y., Zhang, H. B., Jiang, J. J., Zhang, L. P., and Gao, L. Z. (2016). CandiSSR: an efficient pipeline used for identifying candidate polymorphic SSRs based on multiple assembled sequences. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.01171
- Yang, G., Liu, P., Shi, H., Fan, W., Feng, X., Chen, J., et al. (2022). Identification of anti-inflammatory components in *Dioscorea nipponica* Makino based on HPLC-MS/MS, quantitative analysis of multiple components by single marker and chemometric methods. *J. Chromatogr. B.* 1213, 123531. doi: 10.1016/j.jchromb.2022.123531
- Yuan, J., Zhang, X., Wang, M., Sun, Y., Liu, C., Li, S., et al. (2021). Simple sequence repeats drive genome plasticity and promote adaptive evolution in penaeid shrimp. *Commun. Biol.* 4, 186. doi: 10.1038/s42003-021-01716-y
- Zhou, Q. Y., Cai, H. X., Liu, Z. H., Yuan, L. X., Yang, L., Yang, T., et al. (2022). Development of genomic resources for *Wenchengia alternifolia* (Lamiaceae) based on genome skimming data. *Plant Diversity* 44, 542–551. doi: 10.1016/j.pld.2021.09.006