scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A high-quality chromosome-level genome assembly of the traditional Chinese medicinal herb *Zanthoxylum nitidum*

Yanxia Zhu, Guiyu Tan, Qingsong Dong, Baoyou Huang, Yude Peng & Jianping Jiang ✉

Dried roots of *Zanthoxylum nitidum* (2n=2x=70, family *Rutaceae*) was referred as "*Liang-Mian-Zhen*" in Chinese, acting as a valuable species due to its notable pharmacological activities. Herein, we combined PacBio HiFi data together with Hi-C mapping technology to construct a chromosome-scale reference genome assembly for *Z. nitidum*. The assembly reached a length of 2.24 Gb, successfully anchoring 99.31% of sequences onto 35 pseudo-chromosomes. Among these, 26 chromosomes achieved telomere-to-telomere assembly, and 11 chromosomes were gap-free. The contigs N50 and scaffolds N50 reached 57.61 Mb and 78.00 Mb, respectively. Transposable elements comprised 81.44% of the *Z. nitidum* genome, and over 78% of them were long-terminal repeat retrotransposon elements. Furthermore, 32,737 protein-coding genes were identified and 99.38% of all were functionally annotated. The completeness of the genome assembly and final gene sets reached 97.83% and 96.47% based on Benchmarking Universal Single-Copy Orthologs (BUSCO), respectively. Taken together, our results provided a high-quality chromosome-level assembly of *Z. nitidum* genome and will be a valuable resource that will facilitate breeding varieties with higher alkaloids content.

Background & Summary

The dried roots of *Zanthoxylum nitidum* (Roxb.) DC., referred to as *Zanthoxyli Radix* or "*Liang-Mian-Zhen*" in Chinese, constitute a traditional Chinese medicinal herb with a millennium-long history of therapeutic application[1,2]. *Z. nitidum* (2n=2x=70) is a shrub in the *Rutaceae* family of the *Zanthoxylum* genus, grows in dry habitats and exhibits extensive distribution across southern China, including provinces such as Guangxi, Guangdong, Yunnan, and Taiwan[3]. Owing to its rich content of alkaloids such as chelerythrine, sanguinarine, berberine, and magnoflorine, all belonging to the benzophenanthridine alkaloid class, *Z. nitidum* exhibits notable pharmacological activities, including anti-inflammatory, anti-tumor, antiviral, and antibacterial properties[1,4]. *Z. nitidum* has attracted widespread international attention due to its significant medicinal potential.

Benzophenanthridine alkaloids, an important subclass of isoquinoline alkaloids, are primarily found in plants belonging to the *Rutaceae*, *Papaveraceae*, and *Ranunculaceae* families, such as *Z. nitidum*, *Papaver somniferum*, *Corydalis yanhusuo*, and *Sanguinaria canadensis*[5–7]. The berberine bridge enzyme (BBE) genes, serving as a key catalyst in crucial reactions, play crucial roles in the production of benzophenanthridine alkaloids[8,9]. For instances, the overexpression of the BBE genes from *Pseudomonas fluorescens* resulted in an increased level of benzophenanthridine alkaloids in California poppy, while concurrently inhibiting the expression of the BBE gene family led to a reduced nicotine phenotype in tobacco[10,11]. Berberine bridge enzyme is a flavin-containing oxidase belonging to the FAD-linked oxidase superfamily (SCOPe d.58.32), first identified and named by Rink and Biihm in 1975[12,13]. Advancements in whole genome sequencing technologies have facilitated the successful identification of over 4,000 genes encoding BBE-like enzymes in more than 100 plants, including species such as *Papaver somniferum*, *Nicotiana tabacum*, *Andrographis paniculata*, and *Jatropha curcas*[14–17]. However, the absence of a high quality reference genome has restricted our comprehension of the distribution and evolutionary characteristics of the BBE family within the *Z. nitidum*.

National Center for TCM Inheritance and Innovation, Guangxi Botanical Garden of Medicinal Plants, Naning, 530023, China. ✉e-mail: jiangjianping818@126.com

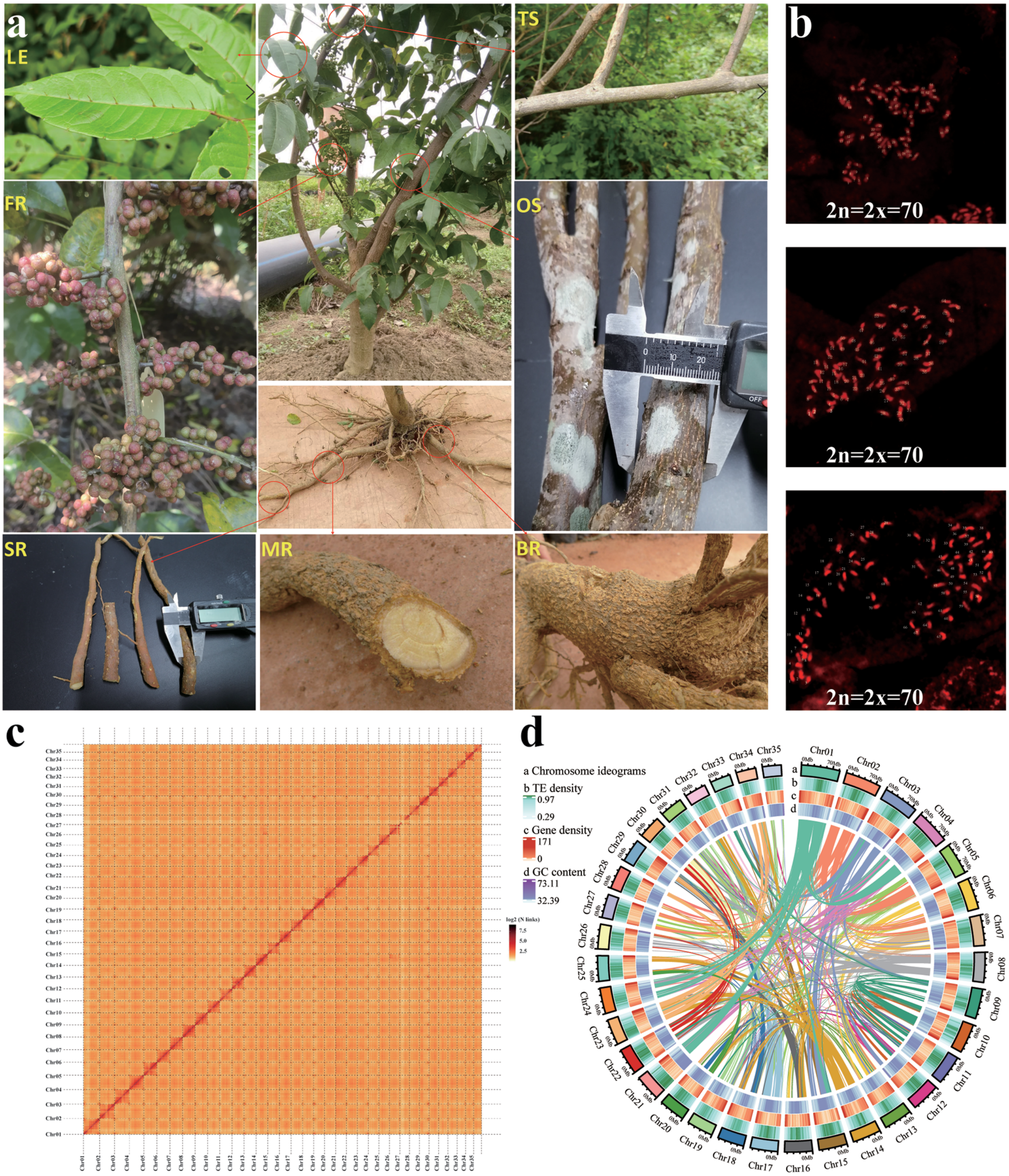

Fig. 1 Morphological characteristics and genomic features of *Zanthoxylum nitidum*. (**a**) Photograph of *Z. nitidum*. The photographs depict different plant tissues: LE - leaves, FR - fruits, TS - tender stems (≥1 and <3 cm in diameter), OS - old stems (≥3 cm in diameter), SR - small roots (≥1 and <2 cm in diameter), MR - medium roots (≥2 cm and <3 cm in diameter), BR - big roots (≥3 cm in diameter). (**b**) Karyotypes of *Z. nitidum* chromosomes using the conventional squashing method. (**c**) Heatmap showing 35 pseudo-chromosomes Hi-C interactions with a resolution of 500 kb. (**d**) Characteristics of the 35 chromosomes of *Z. nitidum*. Tracks a-c display chromosome ideograms, transposable element (TE) density, and gene density, with densities calculated in 2 Mb windows. Track d illustrates syntenic blocks.

The utilization of whole genome sequencing offers valuable assets for examining species evolution, detecting genetic variations, and choosing traits to enhance crops[18,19]. To enhance our comprehensive understanding the genetic basis of *Z. nitidum*, we constructed a high-quality chromosomal-level reference genome through the integration of PacBio HiFi reads and Hi-C data. Subsequently, we systematically annotated the

Metrics	Value
Clean data (Gb)	77.07
Depth (x)	31.51
kmer size	21
Number of kmer	55,291,189,492
Genome size (bp)	2,445,546,787
GC content (%)	35.16
Repeat (%)	78.62
Heterozygous ratio (%)	0.31

Table 1. K-mer analysis of the *Zanthoxylum nitidum* genome by setting k-mer = 21.

Metrics	Value
Total data (bp)	87,759,887,042
Depth (x)*	35.82
Number of subreads	5,799,147
N50 of subreads (bp)	15,502
Mean length of subreads (bp)	15,133
Max length of subreads (bp)	50,449

Table 2. Statistics of PacBio HiFi sequencing data. *Depth was calculated under the estimate of a genome size of 2445.5 Mb.

Data	Value
Statistics of Hi-C data	
Clean data (bp)	257,050,708,824
Total read pairs	862,696,145
GC (%)	36.51
>Q30 (%)	95.18
Mapping stats of Hi-C data	
Total mapped reads	1,353,609,015
Unique mapped read pairs	561,843,892
Percentage of unique mapped read pairs (%)	65.13
Statistics of valid Hi-C data	
Valid interaction pairs	334,199,632
Percentage of valid interaction read pairs (%)	38.74
Dangling End Pairs	155,474,881
Re-ligation Pairs	5,774,303
Self-cycle Pairs	577,539
Dumped Pairs	65,817,537

Table 3. Statistics of Hi-C data and valid data.

functional elements across whole genome, encompassing transposable elements (TEs), protein-coding genes and non-coding RNAs (ncRNAs). In addition, transcriptomic data from seven different tissues of *Z. nitidum* were collected to unveil the functional patterns of the BBE family in the biosynthesis of benzophenanthridine alkaloids. Overall, the chromosomal-level assembly of the *Z. nitidum* genome provides crucial data for future investigations in the evolution of *Rutaceae* family plants. These genomic data can further advance our understanding of the production of benzophenanthridine alkaloids in *Z. nitidum*, offering insights for breeding varieties with higher alkaloid yield.

Methods

Plant materials and karyotype analysis. The samples in this study were collected from 3-years-old individual plants grown at an artificial planting base in Nanning City, Guangxi Province (109°11′E, 22°31′N). The plant was identified by Prof. Yudeng Peng (Guangxi Botanical Garden of Medicinal Plants, China). This study utilized tissue materials from fresh leaves, fruits, roots, and stems for genomic and transcriptomic investigations, with three independent biological replicates collected for each sample type (Fig. 1a). The root tissues were categorized based on their size into large roots (≥3 cm in diameter), medium roots (≥2 cm and <3 cm in diameter), and small roots (≥1 and <2 cm in diameter). The stems were classified based on their size into old stems (≥3 cm in diameter) and tender stems (≥1 and <3 cm in diameter). The chromosome numbers of the *Z. nitidum* genome

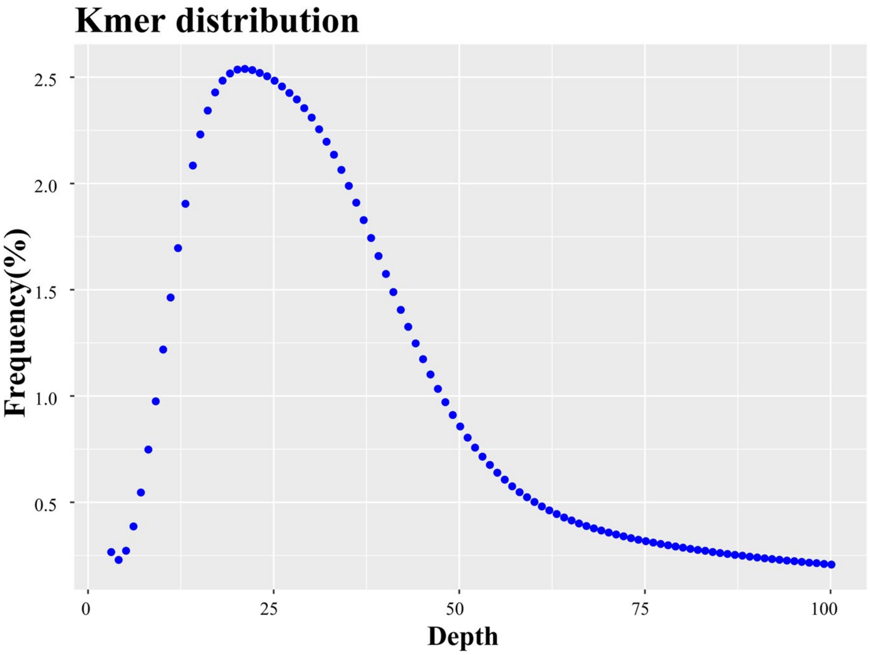

Fig. 2 Estimation of the *Zanthoxylum nitidum* genome size by k-mer analysis. The figure showed the frequency of 21 k-mers, which are 21 bp sequences from clean reads of short-insert-size libraries. We identified 55,291,189,492 k-mers and the main peak of k-mer depth is 20. Genome size can be estimated as (total k-mer number)/(the volume peak). The genome size of the *Z. nitidum* was thus estimated as 2,445 Mb.

Metrics	Value
Total length (bp)	2,241,138,043
Number of contigs	437
N50 of contigs	58,632,861
N90 of contigs	34,730,181
Max length of contigs	85,812,342
GC (%)	35.88
Coverage of genome size estimated by kmer (%)	91.64

Table 4. Statistics of draft assembly of the *Zanthoxylum nitidum* genome using HiFi data.

were determined in root tip cells using the conventional squashing method. Prior to fixation in ethanol-acetic acid (3:1) at 4 °C for 24 hours, the root tips were pre-treated with 2 mM 8-hydroxyquinoline for 3 hours at 24 °C. Subsequently, they were macerated with 1 N HCl at 60 °C for 10 minutes and stained with lactopropionic orcein for 30 minutes at 24 °C. The karyotype analysis results indicated that its somatic cells were diploid, containing 35 pairs of chromosomes (Fig. 1b).

DNA extraction and sequencing. High-quality genomic DNA was extracted from fresh leaves tissues using the cetyltrimethylammonium bromide (CTAB) method. For short reads sequencing, we generated a DNA-seq library with an insert size of 350 bp and conducted paired-end sequencing on the Illumina NovaSeq 6000 platform (Illumina, CA, USA). The generated dataset amounted to approximately 77.07 Gb, with a coverage of around 31.51x (Table 1). For PacBio sequencing, genomic DNA from the same samples was used to construct sequencing libraries. The genomic DNA underwent fragmentation using a g-TUBE centrifuge at 2,000 r.p.m. for 2 min, followed by end-repair, adapter ligation, and exonuclease digestion according to the recommendations from Pacific Biosciences, USA. DNA fragments in the range of 15–20 kb were selected using BluePippin electrophoresis (Sage Sciences). Subsequently, DNA libraries were sequenced on the PacBio Sequel II platform (Pacific Biosciences, USA) utilizing the HIFI model. In total, we obtained 87.76 Gb of PacBio HiFi data with a coverage of 35.82×, encompassing approximately ~5.8 million subreads with an average length of 15.13 kb (Table 2). In the Hi-C experiment recommend by standard protocols, one library was constructed and subsequently sequenced on the Illumina NovaSeq 6000 (Illumina, CA, USA) platform, yielding a total of 257.05 Gb of data (Table 3).

Genome size estimation. The raw short reads underwent quality filtering using fastp (v0.20.0)[20]. Subsequently, the clean reads were employed for genome size estimation. We utilized Jellyfish (v1.1.12) software

Characteristics	Values
Assembled genome features	
Assembled genome size (Gb)	2.24
GC content	35.88
Number of contigs	437
Longest contigs (Mb)	85.81
N50 of HiFi assembly (Mb)	58.63
N90 of HiFi assembly (Mb)	34.73
Pesudochromosomes number	35
Telomere-to-temomere chromosome	26
Gap free chromosomes	11
Seuquence anchored to pseudochromosomes (Gb)	2.23 (99.56%)
Mean number of contigs anchored onto each pseudochromosome	6.37
Number of scaffolds	425
N50 of contigs (Mb)	57.61
N50 of scaffolds (Mb)	61.5
N90 of scaffolds (Mb)	44.96
Genome completeness	
Mapping rate of clean NGS reads	99.72%
Mapping rate of Pacbio HiFi reads	99.87%
BUSCO	97.83%
CEGMA	97.82%
Genome annotation	
Number of predicted protein-coding genes	32,737
Average gene length (bp)	4,480
Percentage of functional annotation genes (%)	99.48
Percentage of repeat sequences (%)	81.44%

Table 5. Summary of *Zanthoxylum nitidum* genome assembly and annotation.

to count the 21-mers and generated the 21-mer frequency curve[21]. The horizontal axis represents k-mer depth, indicating the number of occurrences. The plot illustrates the relationship between k-mer volume and frequency. Using the equation genome size = (total number of k-mers)/(the volume peak), the genome size, of *Z. nididum* was estimated. The heterozygosity and repetitive sequence proportion were also obtained. Utilizing ~32× illumina short-reads for k-mer analysis, the genome size of *Z. nitidum* was estimated to be 2.45 Gb, with a repeat sequence proportion of approximately 78.62% and heterozygosity rate of 0.31% (Fig. 2 and Table 1).

Genome assembly. The high-accuracy HiFi data was assembled using Hifiasm software (version 0.16)[22]. During the assembly process, multiple sets of specific parameter combinations were employed: i) -l=0, -n=3; ii) -l=0, -n=4; iii) -l=2, -n=3; iv) -l=2, -n=4. The criteria for selecting the optimal genome were based on: 1) prioritizing genome sizes that are close; 2) choosing assembly results with higher contigs N50. Ultimately, we selected the genome assembly produced with the parameter combination -l=2, -n=3. Genome assembly based on HiFiasm amounted to 2.24 Gb, comprising 437 contigs, with an N50 of up to 58.63 Mb and a maximum length of 85.81 Mb (Table 4).

Chromosome-level assembly through Hi-C mapping technology. Utilizing the genome-wide spatial interaction information provided by Hi-C data, we refined, clustered, and oriented the contig-level genome assembly, achieving chromosomal-level resolution. The raw sequencing data was processed using HiC-Pro (v3.1.0) to generate a standard interaction matrix[23]. BWA-aln was employed to map the clean data to the contig-level genome assembly[24]. To ensure the accuracy of chromosome construction, only reads that uniquely aligned to the genome and represented valid data were retained. The clustering, ordering, and orientation of sequences resulted in the formation of 35 pseudo-chromosomes, this process executed using Lachesis[25]. The heatmap of intra- and inter-chromosomal interactions was visualized using a customized script, with a resolution set to 500 Kb. The total Hi-C data was 257.05 Gb comprising 862.70 million paired clean reads, with 65.13% uniquely mapping to the contig-level genome (Table 3). A cumulative of 2.23 Gb (99.31%) of sequences were successfully anchored onto 35 pseudo-chromosomes (Fig. 1c). Finally, a high-contiguity chromosome-level genome was obtained, showing scaffold N50 of 61.50 Mb and the maximum chromosome length of 105.18 Mb, with a total size of 2.24 Gb (Fig. 1d and Table 5). The longest chromosome had a length of 105.18 Mb, while the shortest was 42.06 Mb, with 29 of them (82.86%) consisting of no more than five contigs (Table 6). The analysis of centromeres and telomeres was carried out by quarTeT (v1.1.7)[26]. The centromere region lengths of the 35 chromosomes range from 101 Kb to 52.4 Mb (Table 7). Telomeres were identified at least at one end of 34 chromosomes, with 26 chromosomes exhibiting telomeric fragments at both ends (Table 8).

Chromosomes	Number of contigs	Length of chromosomes (bp)	Number of genes
Chr01	11	105,183,316	1,715
Chr02	1	80,063,596	1,370
Chr03	2	85,660,299	1,437
Chr04	2	76,954,204	1,067
Chr05	1	70,666,880	1,186
Chr06	4	75,118,727	975
Chr07	2	68,001,280	870
Chr08	1	66,871,090	1,150
Chr09	2	66,864,123	993
Chr10	4	65,582,310	1,148
Chr11	4	65,450,254	791
Chr12	1	64,064,813	844
Chr13	1	63,252,005	1,134
Chr14	3	63,221,144	1,004
Chr15	4	62,867,951	664
Chr16	3	61,845,942	936
Chr17	1	61,495,432	767
Chr18	1	59,409,268	973
Chr19	9	61,701,785	1,218
Chr20	2	66,086,122	773
Chr21	104	67,717,184	714
Chr22	2	58,061,914	754
Chr23	28	66,769,391	958
Chr24	4	57,577,948	798
Chr25	3	56,908,048	685
Chr26	2	55,233,806	657
Chr27	6	82,297,516	645
Chr28	1	52,585,404	618
Chr29	1	52,428,156	1,259
Chr30	1	50,666,008	582
Chr31	2	50,713,653	712
Chr32	2	52,994,961	576
Chr33	1	44,955,487	582
Chr34	2	44,353,005	765
Chr35	5	42,064,283	438
Total (Ratio %)	223 (49.78%)	2,225,687,305 (99.31%)	31,758 (97.01%)

Table 6. Summary of chromosome level assembly of *Zanthoxylum nitidum* genome based on Hi-C data.

Repetitive sequences annotation. The identification of repeatitive sequences in the *Z. nididum* genome integrated homology-based alignments and *de novo* search methods. The tandem repeat sequences were mainly identified using MISA (v2.1)[27] and Tandem Repeat Finder (TRF, version 409)[28]. To enhance accuracy in our assessment, we employed the Repbase database (version 15.02), a widely acknowledged resource for predicting homologous repetitive sequences in plants. Initially, we employed RepeatModeler2 (v2.0.1) for *de novo* prediction[29], which primarily incorporates two softwares, RECON (v1.0.8)[30], and RepeatScout (v1.0.6)[31]. The classification of predicted results was performed using RepeatClassifier with the assistance of the known database Dfam (v3.5). Subsequently, we employed LTR_retriever (V2.9.0) for dedicated *de novo* prediction of LTR[32], primarily relying on the results predicted by LTRharvest (v1.5.10) and LTR_FINDER (v1.07)[33,34]. Finally, The results from both strategies were merged and redundancies were eliminated to construct a *Z. nididum*-specific repetitive sequences database. RepeatMasker (v4.1.2) was employed for the prediction of transposable element (TE) sequences in the genome[35]. Finally, a total of approximately 1.83 Gb (81.44%) of repetitive sequences was identified using this combined approach of homology-based and *de novo* methods (Table 9). Among them, retrotransposons accounted for 79.13%, and DNA transposons accounted for 2.3%. Long terminal repeat retrotransposons (LTR-RTs) make up over 78% of the *Z. nitidum* genome, with 51.03% attributed to the *Copia* lineage and 8.56% to the *Gypsy* lineage.

Gene model prediction. A combined approach, incorporating homology-based, *ab initio*, and RNAseq-based methods, was employed to identify protein-coding genes in the genome of *Z. nididum*. We utilized *ab initio* programs including Augustus (v3.1.0) and SNAP (version 2006-07)[36,37]. For homologous predictions, protein sequences from *Arabidopsis thaliana*, *Zanthoxylum armatum*, *Citrus grandis*, and *Citrus medica* were downloaded from

Chromosome	Start	End	Length (bp)	Tandem repeat length (bp)	Tandem repeat coverage
Chr01	24,381,750	56,902,529	32,520,780	2,255,824	6.94%
Chr02	28,947,645	53,215,282	24,267,638	2,214,057	9.12%
Chr03	35,106,546	55,198,470	20,091,925	1,892,395	9.42%
Chr04	16,609,951	66,726,135	50,116,185	4,995,084	9.97%
Chr05	13,511,056	43,850,122	30,339,067	1,690,041	5.57%
Chr06	51,813,598	53,267,029	1,453,432	102,416	7.05%
Chr07	51,603,968	51,896,657	292,690	22,730	7.77%
Chr08	23,904,324	48,355,129	24,450,806	1,861,990	7.62%
Chr09	66,665,320	66,779,215	113,896	10,775	9.46%
Chr10	17,353,000	42,651,088	25,298,089	1,886,273	7.46%
Chr11	337,884	1,588,413	1,250,530	168,110	13.44%
Chr12	2,780,568	55,185,130	52,404,563	3,251,516	6.20%
Chr13	28,431,346	56,418,202	27,986,857	1,691,308	6.04%
Chr14	12,898,782	46,836,372	33,937,591	2,184,827	6.44%
Chr15	3,887,466	51,959,316	48,071,851	4,190,315	8.72%
Chr16	37,543,377	38,205,728	662,352	52,224	7.88%
Chr17	14,612,164	50,185,227	35,573,064	2,535,757	7.13%
Chr18	15,276,225	54,534,269	39,258,045	2,043,123	5.20%
Chr19	6,867,235	7,011,757	144,523	9,356	6.47%
Chr20	35,228,423	47,042,189	11,813,767	956,750	8.10%
Chr21	1	36,308,167	36,308,167	3,188,689	8.78%
Chr22	2,132,738	2,670,059	537,322	44,577	8.30%
Chr23	50,240,556	50,440,978	200,423	16,803	8.38%
Chr24	57,449,274	57,551,055	101,782	6,445	6.33%
Chr25	5,656,917	46,538,261	40,881,345	3,048,759	7.46%
Chr26	3,043,863	48,559,729	45,515,867	2,907,902	6.39%
Chr27	6,353,164	42,557,831	36,204,668	2,312,017	6.39%
Chr28	23,780,952	34,482,724	10,701,773	700,343	6.54%
Chr29	42,764,384	42,916,610	152,227	8,879	5.83%
Chr30	8,244,076	44,795,799	36,551,724	2,884,968	7.89%
Chr31	4,902,862	5,072,224	169,363	11,921	7.04%
Chr32	4,933,339	48,009,003	43,075,665	2,322,858	5.39%
Chr33	8,899,439	33,779,367	24,879,929	1,421,635	5.71%
Chr34	40,145,722	40,591,578	445,857	22,460	5.04%
Chr35	3,597,627	37,818,318	34,220,692	2,135,718	6.24%

Table 7. Distribution of centromeres in *Zanthoxylum nitidum*.

Phytozome[38]. Subsequently, the homologous genomic sequences were aligned with corresponding proteins, and precise protein-coding gene models were constructed using GeMoMa (v1.7)[39]. For RNAseq-based predictions, sequencing reads from various tissues of *Z. nididum* were aligned to the reference genome. Transcripts were then identified from these mapping results using GeneMarkS-T (v5.1) and PASA (v2.4.1)[40,41]. EvidenceModeler (v1.1.1) was used to integrate the gene prediction results from three sources, resulting in a non-redundant gene set[42]. In this process, PASA was utilized for terminal exon modification to obtain a more complete gene structure. A total of 32,737 protein-coding genes were identified in *Z. nitidum*, with 31,758 (97.01%) genes located on the 35 pseudo-chromosomes (Tables 6, 10).

Gene function annotation. The protein sequences were mapped to the NR database (https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/) using BLAST (e-values ≤ 1e−5), and the closest hit was selected as the annotation result. In addition, eggNOG, Swiss-Prot, TrEMBL, KOG, and Pfam databases were used for a comprehensive functional annotation of genes[43–45]. The process of alignment was based on Diamond, while InterProScan (v5.34) was employed to meticulously elucidate motifs and domains, which are crucial indicators of protein function[46]. Blast2GO was used to determine the functions and pathways based on the GO and KEGG databases[47]. Among all protein-coding genes, 87.30% of them were successfully annotated in the Gene Ontology (GO) database, and 79.60% of them were annotated in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, providing insights into their functional categorization and enriched biological pathways (Table 11).

Non-coding RNA annotation. The tRNA prediction was executed through the utilization of the tRNAscan-SE program (v1.3.1)[48]. The identification of ribosomal RNAs (rRNAs) was primarily involved using barrnap (v0.9) for obtaining accurate predictions[49]. Prediction of miRNAs, snoRNAs, and snRNAs was carried out based on the Rfam (v14.5) database[50], utilizing Infernal (v1.1) for prediction[51]. *Z. nitidum* genome harbored various types of identified non-coding RNAs, numbering from 159 to 4,746 (Table 12).

ID	Length (bp)	Telomere status	Number of telomere in left	Direction of left	Number of telomere in right	Direction of right
Chr01	85,834,700	both	330	+	2,183	–
Chr02	80,063,596	both	1,102	+	2,228	–
Chr03	79,009,128	both	231	+	854	–
Chr04	76,925,120	no	0		0	
Chr05	70,666,880	both	1,857	+	1,944	–
Chr06	69,393,101	both	890	+	1,559	–
Chr07	68,001,380	left	1,480	+	0	
Chr08	66,871,090	both	1,381	+	2,832	–
Chr09	66,800,052	both	1,971	+	1,367	–
Chr10	65,481,570	right	0		1,625	–
Chr11	65,450,554	both	1,824	+	2,094	–
Chr12	64,064,813	both	1,316	+	1,104	–
Chr13	63,252,005	both	1,067	+	2,120	–
Chr14	63,133,193	left	1,932	+	0	
Chr15	62,838,192	both	848	+	1,742	–
Chr16	61,846,142	both	598	+	1,971	–
Chr17	61,495,432	both	1,712	+	1,918	–
Chr18	59,409,268	both	1,704	+	2,417	–
Chr19	58,931,088	left	2,171	+	0	
Chr20	58,632,861	both	2,031	+	1,020	–
Chr21	58,501,423	right	0		2,160	–
Chr22	58,062,014	right	0		430	–
Chr23	57,652,966	both	2,130	+	1,061	–
Chr24	57,578,248	both	1,958	+	1,478	–
Chr25	56,882,761	both	2,093	+	2,471	–
Chr26	55,018,586	both	2,374	+	1,460	–
Chr27	52,659,341	both	1,657	+	1,881	–
Chr28	52,585,404	both	1,005	+	1,417	–
Chr29	52,428,156	both	2,017	+	1,717	–
Chr30	50,666,008	both	656	+	1,614	–
Chr31	50,527,451	both	1,487	+	2,262	–
Chr32	48,019,206	left	1,915	+	0	
Chr33	44,955,487	both	1,278	+	1,779	–
Chr34	44,082,972	both	2,734	+	1,811	–
Chr35	41,895,113	left	1,612	+	0	

Table 8. Telomere distribution in *Zanthoxylum nitidum*. *Telomere repeat monomer: AAACCCT.

Transcriptomics analysis. RNA isolation and purification were performed on 21 samples obtained from seven tissues (fruits, leaves, tender stems, old stems, big roots, medium roots, and small roots), with three replicates for each tissue (Table 13). After library construction, the quality of the libraries were assessed by Qubit 2.0 and Agilent 2100. Subsequently, these cDNA libraries was performed high-throughput sequencing on the Illumina NovaSeq 6000 platform, producing paired-end reads with a length of 150 bp. This sequencing was carried out by Biomarker Technologies Co., Ltd (Beijing, China). Quality control of the raw sequencing reads, including adapter removal and quality filtering, was performed using Trimmomatic (v0.39)[52]. Clean reads from each tissue were aligned to the high-quality genome of *Z. nitidum* using TopHat (v2.21)[53]. The assembly of mapped reads for each sample was conducted using Cufflinks (v2.2.1)[54]. For the quantification of gene expression levels, RSEM (v1.3.3) was utilized to calculate FPKM[55]. Differential expression genes (DEGs) across different tissues were detected using limma (v3.54.2)[56], incorporating thresholds of a p-value ≤ 0.05 and an absolute log2 (fold change) greater than 1. The number of differentially expressed genes varies considerably between different tissue comparisons, with only 887 differentially expressed genes identified between tender stems and old stems, while 12,139 differentially expressed genes were detected between fruits and medium roots (Table 14).

Data Records

PacBio Hifi reads, illumina paired-end reads, Hi-C reads, and RNA-Seq reads have been deposited in Genome Sequence Archive (GSA: CRA013814, https://ngdc.cncb.ac.cn/gsa)[57]. The genome assembly and annotation of *Zanthoxylum nitidum* are available in Figshare[58] and GenBank[59]. Additionally, the clean paired-end reads, raw counts, and FPKM values from 21 samples have been deposited in the GEO database (accession number: GSE281536)[60], while the differentially expressed genes across various tissues are available on Figshare[61].

Type	Number	Total length (bp)	Percentage of genome (%)
ClassI:Retroelement	1,144,812	1,773,498,832	79.13
LTR-retroelement			
LTR/Copia	522,287	1,143,605,400	51.03
LTR/Gypsy	146,400	191,921,234	8.56
LTR/Others	439,034	425,203,197	18.97
Non LTR-Retroelement			
LINE	26,520	11,243,561	0.5
SINE	10,570	1,525,393	0.07
Other Retroelement	1	47	0
ClassII:DNA transposon	164,077	51,653,244	2.3
hAT	11,992	5,470,063	0.24
CACTA	7,735	1,630,439	0.07
Helitron	1,047	522,536	0.02
Mutator	4,467	2,717,322	0.12
Kolobok	1,272	131,911	0.01
Others	4,501	274,732	0.01
Unknown	133,063	40,906,241	1.83
Unknown	26	1,297	0
Total	1,308,915	1,825,153,373	81.44

Table 9. Annotation of repetitive elements in the *Zanthoxylum nitidum* genome.

Strategy	Software	Species	Gene number
Ab initio			
	Augustus	—	47,688
	SNAP	—	77,268
Homology-based			
	—	*Zanthoxylum armatum*	46,680
	—	*Arabidopsis thaliana*	31,869
	—	*Citrus grandis*	35,253
		Citrus media	35,855
RNA-seq			
	GeneMarkS-T	—	23,190
	PASA	—	26,668
Integration	EVM	—	32,737

Table 10. Statistics of predicted protein-coding genes in the *Zanthoxylum nitidum* genome.

Database	Number of annotated genes	Percentage of annotated genes (%)
All genes	32,737	100
GO	28,579	87.30
KEGG	26,059	79.60
KOG	20,119	61.46
Pfam	29,480	90.05
Swissprot	28,418	86.81
TrEMBL	32,552	99.43
eggNOG	28,880	88.22
NR	32,534	99.38
Total annotated genes	32,568	99.48

Table 11. Functional annotation of predicted protein-coding genes in the *Zanthoxylum nitidum* genome.

Technical Validation

The karyotype analysis results indicated that its somatic cells were diploid, containing 35 pairs of chromosomes (Fig. 1b). The evaluation of the completeness of this genome was performed. Firstly, we utilized the BWA

Type	Number
rRNA	4,746
tRNA	1,036
miRNA	159
snRNA	219
snoRNA	178

Table 12. Prediction of non-coding RNAs in the *Zanthoxylum nitidum* genome.

Tissue	Sample id	Number of Reads	Total Bases(bp)	Q20(%)	Q30(%)	GC(%)
Big root	BR-1	22,193,420	6,642,960,824	97.8	93.84	44.81
Big root	BR-2	21,539,530	6,370,980,688	98.12	94.83	44.6
Big root	BR-3	22,862,627	6,801,730,676	97.89	94.22	44.24
Medium root	MR-1	19,292,207	5,774,605,018	97.56	93.28	44.83
Medium root	MR-2	20,572,765	6,158,911,494	97.87	93.92	44.4
Medium root	MR-3	20,708,023	6,187,708,360	98.22	94.86	44.16
Small root	SR-1	21,628,655	6,474,240,158	97.93	94.09	44.66
Small root	SR-2	21,490,149	6,432,429,070	97.41	92.89	44.61
Small root	SR-3	21,826,816	6,532,972,706	97.87	93.98	44.68
Old stem	OS-1	20,063,346	5,976,428,016	97.33	92.76	43.7
Old stem	OS-2	21,358,335	6,364,107,462	98.07	94.59	43.86
Old stem	OS-3	23,873,404	7,113,747,514	97.48	93.35	44.96
Tender stem	TS-1	25,133,940	7,485,170,266	98.05	94.47	44.25
Tender stem	TS-2	19,549,470	5,812,689,266	98.19	94.93	43.62
Tender stem	TS-3	19,902,982	5,958,057,002	97.25	92.44	44.72
Leave	LE-1	19,062,028	5,707,259,428	97.93	94.02	43.64
Leave	LE-2	19,854,778	5,945,188,182	97.81	93.71	43.57
Leave	LE-3	19,752,357	5,915,798,380	97.68	93.34	43.6
Fruit	FR-1	21,321,337	6,383,174,544	97.77	93.62	43.37
Fruit	FR-2	19,811,759	5,931,693,910	97.9	93.9	43.37
Fruit	FR-3	19,154,454	5,734,396,384	97.95	94.02	43.37

Table 13. Statistics of RNA-seq data of seven *Zanthoxylum nitidum* tissues.

Tissue	Tissue	Number of DEGs
BR	FR	11,075
BR	LE	8,051
BR	MR	5,022
BR	OS	1,538
BR	SR	4,551
BR	TS	3,596
FR	LE	8,738
FR	MR	12,139
FR	OS	7,629
FR	SR	10,514
FR	TS	8,361
LE	MR	10,234
LE	OS	6,163
LE	SR	8,790
LE	TS	7,004
MR	OS	3,310
MR	SR	1,064
MR	TS	5,409
OS	SR	3,086
OS	TS	887
SR	TS	4,743

Table 14. Number of differentially expressed genes between different tissues of *Zanthoxylum nitidum*.

	Illumina short reads	HiFi long reads
Total bases (Gb)	77.07	87.76
Number of reads	514,862,518	5,799,147
Mapped reads	513,431,825	5,791,882
Mapped rate (%)	99.72	99.87
Coverage (≥5X) (%)	99.24	99.59
Coverage (≥10X) (%)	97.14	99.23

Table 15. Statistics of mapping rates of the *Zanthoxylum nitidum* genome assembly.

Number of 458 Core eukaryotic genes (CEGs) present	Percentage of 458 CEGs present (%)	Number of 248 highly conserved CEGs present	Percentage of 248 highly conserved CEGs present (%)
448	97.82	236	95.16

Table 16. Completeness assessment of the assembled genome of the *Zanthoxylum nitidum* using CEMGA.

	Genome sequences	Gene sets
Database	OrthoDB 10	OrthoDB 10
Complete BUSCOs (C)	1,579 (97.83%)	1,557 (96.47%)
Complete and single-copy BUSCOs (S)	1,364 (84.51%)	1,333 (82.59%)
Complete and duplicated BUSCOs (D)	215 (13.32%)	224 (13.88%)
Fragmented BUSCOs (F)	4 (0.25%)	31 (1.92%)
Missing BUSCOs (M)	31 (1.92%)	26 (1.61%)
Total conserved markers	1,614 (100%)	1,614 (100%)

Table 17. Completeness assessment of the assembled genome and predicted gene sets of the *Zanthoxylum nitidum* using BUSCO.

(v0.7.17) software to align short sequences to the assembled genome[24]. Secondly, minimap2 (v2.24) software was employed to align HiFi reads to the assembled genome[62]. The evaluation of the completeness and sequencing coverage uniformity of the assembled genome was based on metrics, such as mapping rate, proportion of genome coverage, and depth distribution. Thirdly, the Core Eukaryotic Genes Mapping Approach (CEGMA) was also employed[63]. Finally, BUSCO (v5.2.2) with the OrthoDB 10 database was used to evaluate the completeness[64]. The alignment efficiency of illumina short-reads to the genome reached was 99.72%, and over 97% of the genome regions exhibited a coverage depth exceeding 10x (Table 15). In addition, the mapping rate of PacBio HiFi long-reads reached 99.87%, with coverage exceeding 10x observed in over 99% of the genome regions (Table 15). In our chromosome-level *Z. nitidum* genome, 448 out of the identified 458 core eukaryotic genes (CEGs) were detected, achieving a completeness rate of 97.82% (Table 16). The Benchmarking Universal Single-Copy Orthologs (BUSCO) database was also utilized for evaluating the completeness, indicating 97.83% completeness for core genes, with only a 1.92% absence (Table 17). All these findings suggested this *Z. nitidum* assembled genome was of high quality, characterized by its high accuracy, completeness, contiguity, and substantial coverage (Table 4).

Code availability

No custom code was used for the curation and/or validation of the dataset. All data processing and analysis were performed using standard software tools and packages as described in the Methods section following their respective protocols and manuals. Below is detailed parameter information about some bioinformatics tools. Fastp: -q 10 -u 50 -y -g -Y 10 -e 20 -l 100 -b 150 -B 150. SOAP: -m 260 -x 440. Jellyfsh: -h 100000. Hifasm: l=2, n=3. LACHESIS:CLUSTER_MIN_RE_SITES=31;CLUSTER_MAX_LINK_DENSITY=2;ORDER_MIN_N_RES_IN_TRUNK=15;ORDER_MIN_N_RES_IN_SHREDS=15. LTRharvest: -minlenltr 100 -maxlenltr 40000 -mintsd 4 -maxtsd 6 -motif TGCA -motifmis 1 -similar 85 -vic 10 -seed 20 -seqids yes. LTR_fnder: -D 40000 -d 100 -L 9000 -l 50 -p 20 -C -M 0.9. Diamond alignment (Orthofnder): e ≤ 1e−3.

Received: 13 May 2024; Accepted: 26 November 2024;
Published online: 02 December 2024

References

1. Hu, J. *et al*. Benzophenanthridine alkaloids from *Zanthoxylum nitidum* (Roxb.) DC, and their analgesic and anti-inflammatory activities. *Chem Biodivers* **3**, 990–995 (2006).
2. Lu, Q. *et al*. *Zanthoxylum nitidum* (Roxb.) DC: Traditional uses, phytochemistry, pharmacological activities and toxicology. *J Ethnopharmacol* **260**, 112946 (2020).
3. Wang, X. *et al*. Distribution survey, phytochemical and transcriptome analysis to identify candidate genes involved in biosynthesis of functional components in *Zanthoxylum nitidum*. *Ind Crop Prod* **150**, 112345 (2020).

4. Yang, G. & Chen, D. Alkaloids from the roots of *Zanthoxylum nitidum* and their antiviral and antifungal effects. *Chem Biodivers* **5**, 1718–1722 (2008).
5. Huang, F.-C. & Kutchan, T. M. Distribution of morphinan and benzo [c] phenanthridine alkaloid gene transcript accumulation in *Papaver somniferum*. *Phytochemistry* **53**, 555–564 (2000).
6. Xu, D. *et al*. Integration of full-length transcriptomics and targeted metabolomics to identify benzylisoquinoline alkaloid biosynthetic genes in *Corydalis yanhusuo*. *Hortic Res* **8**, 16 (2021).
7. Graf, T. N. *et al*. Variability in the yield of benzophenanthridine alkaloids in wildcrafted vs cultivated bloodroot (*Sanguinaria canadensis L*.). *J Agr Food Chem* **55**, 1205–1211 (2007).
8. Dittrich, H. & Kutchan, T. M. Molecular cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of benzophenanthridine alkaloids in the response of plants to pathogenic attack. *P Natl Acad Sci USA* **88**, 9969–9973 (1991).
9. Daniel, B. *et al*. The family of berberine bridge enzyme-like enzymes: A treasure-trove of oxidative reactions. *Arch Biochem Biophys* **632**, 88–103 (2017).
10. Wagner, G. J. & Kroumova, A. B. The Use of RNAi to Elucidate and Manipulate Secondary Metabolite Synthesis in Plants. *Current perspectives in microRNAs (miRNA)*, 431–459 (2008).
11. Facchini, P. J. Regulation of Alkaloid Biosynthesis in Plants. *The alkaloids: chemistry and biology* **63**, 1–44 (2006).
12. Mattevi, A. *et al*. Crystal Structures and Inhibitor Binding in the Octameric Flavoenzyme Vanillyl-alcohol Oxidase: the Shape of the Active-site Cavity Controls Substrate Specificity. *Structure* **5**, 907–920 (1997).
13. Steffens, P., Nagakura, N. & Zenk, M. H. Purification and characterization of the berberine bridge enzyme from Berberis beaniana cell cultures. *Phytochemistry* **24**, 2577–2583 (1985).
14. Pei, L. *et al*. Genome and transcriptome of *Papaver somniferum* Chinese landrace CHM indicates that massive genome expansion contributes to high benzylisoquinoline alkaloid biosynthesis. *Hortic Res* **8** (2021).
15. Edwards, K. D. *et al*. A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics* **18**, 1–14 (2017).
16. Liang, Y. *et al*. Chromosome level genome assembly of *Andrographis paniculata*. *Front Genet* **11**, 701 (2020).
17. Ha, J. *et al*. Genome sequence of *Jatropha curcas L*., a non-edible biodiesel plant, provides a resource to improve seed-related traits. *Plant Biotechnol J* **17**, 517–530 (2019).
18. Bevan, M. W. *et al*. Genomic innovation for crop improvement. *Nature* **543**, 346–354 (2017).
19. Varshney, R. K. *et al*. Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor Appl Genet* **132**, 797–816 (2019).
20. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
21. Marcais, G. & Kingsford, C. Jellyfish: A fast k-mer counter. *Tutorialis e Manuais* **1**, 1–8 (2012).
22. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
23. Servant, N. *et al*. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 1–11 (2015).
24. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
25. Burton, J. N. *et al*. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat Biotechnol* **31**, 1119–1125 (2013).
26. Lin, Y. *et al*. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic Res* uhad127 (2023).
27. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585 (2017).
28. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580 (1999).
29. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *P Natl Acad Sci USA* **117**, 9451–9457 (2020).
30. Bao, Z. & Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res* **12**, 1269–1276 (2002).
31. Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
32. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol* **176**, 1410–1422 (2018).
33. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 1–14 (2008).
34. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265–W268 (2007).
35. Chen, N. Using Repeat Masker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **5**, 4.10. 11–14.10. 14 (2004).
36. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24**, 637–644 (2008).
37. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 1–9 (2004).
38. Goodstein, D. M. *et al*. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* **40**, D1178–D1186 (2012).
39. Keilwagen, J. *et al*. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res* **44**, e89–e89 (2016).
40. Tang, S., Lomsadze, A. & Borodovsky, M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res* **43**, e78–e78 (2015).
41. Haas, B. J. *et al*. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654–5666 (2003).
42. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, 1–22 (2008).
43. Boeckmann, B. *et al*. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31**, 365–370 (2003).
44. Huerta-Cepas, J. *et al*. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**, D309–D314 (2019).
45. Finn, R. D. *et al*. Pfam: clans, web tools and services. *Nucleic Acids Res* **34**, D247–D251 (2006).
46. Jones, P. *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
47. Conesa, A. *et al*. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
48. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955–964 (1997).
49. Loman, T. A novel method for predicting ribosomal RNA genes in prokaryotic genomes. (2017).
50. Griffiths-Jones, S. *et al*. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33**, D121–D124 (2005).
51. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).

52. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
53. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
54. Ghosh, S. & Chan, C.-K. K. Analysis of RNA-Seq data using TopHat and Cufflinks. *Plant Bioinformatics: Methods and Protocols*, 339–361 (2016).
55. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 1–16 (2011).
56. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47–e47 (2015).
57. Zhu, Y. X. *et al.* A high-quality chromosome-level genome assembly of the traditional Chinese medicinal herb *Zanthoxylum nitidum*. *Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRA013814 (2024).
58. Zhu, Y. X. *et al.* Chromosome-level genome assembly and annotation files of *Zanthoxylum nitidum*. *Figshare* https://doi.org/10.6084/m9.figshare.26778394 (2024).
59. NCBI. *GenBank* http://identifiers.org/ncbi/insdc:JBGBDH000000000 (2024).
60. NCBI. *GEO* https://identifiers.org/geo/GSE281536 (2024).
61. Zhu, Y. X. *et al.* Expression dataset of *Zanthoxylum nitidum* and differentially expressed genes across different tissues. *Figshare* https://doi.org/10.6084/m9.figshare.26778331 (2024).
62. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
63. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
64. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

Acknowledgements

This research was funded by the Scientific Research and Technology Development Program of Guangxi (GUIKE AB22035068) and Key Technology Research and Promotion Project of Guangxi Special Medicinal Materials by Guangxi Administration of Traditional Chinese Medicine (GZKJ2314-5).

Author contributions

J.P.J. designed the study and was responsible for providing data. Z.Y.X., D.Q.S., H.B.Y. and P.Y.D. played a crucial role in identifying high-quality resources, gathering, preserving, and cultivating materials of Z. nitidum. T.G.Y. performed the data analyses and visualization. Z.Y.X. drafted the manuscript. J.P.J. revised the manuscript. All authors contributed the final text of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.J.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024