

Chromosome-Level Genome Assembly of the Hemiparasitic *Taxillus chinensis* (DC.) Danser

Jine Fu[†], Lingyun Wan[†], Lisha Song, Lili He, Ni Jiang, Hairong Long, Juan Huo, Xiaowen Ji, Fengyun Hu, Shugen Wei^{*}, and Limei Pan^{*}

Guangxi Botanical Garden of Medicinal Plants, Nanning 530023, China

[†]These authors contributed equally to this work.

^{*}Corresponding authors: E-mails: weishugen2@163.com; plimei2006@163.com.

Accepted: 21 April 2022

Abstract

The hemiparasitic *Taxillus chinensis* (DC.) Danser is a root-parasitizing medicinal plant with photosynthetic ability, which is lost in other parasitic plants. However, the cultivation and medical application of the species are limited by the recalcitrant seeds of the species, and even though the molecular mechanisms underlying this recalcitrance have been investigated using transcriptomic and proteomic methods, genome resources for *T. chinensis* have yet to be reported. Accordingly, the aim of the present study was to use nanopore, short-read, and high-throughput chromosome conformation capture sequencing to construct a chromosome-level assembly of the *T. chinensis* genome. The final genome assembly was 521.90 Mb in length, and 496.43 Mb (95.12%) could be grouped into nine chromosomes with contig and scaffold N50 values of 3.80 and 56.90 Mb, respectively. In addition, a total of 33,894 protein-coding genes were predicted, and gene family clustering identified 11 photosystem-related gene families, thereby indicating photosynthetic ability, which is a characteristic of hemiparasitic plants. This chromosome-level genome assembly of *T. chinensis* provides a valuable genomic resource for elucidating the genetic basis underlying the recalcitrant characteristics of *T. chinensis* seeds and the evolution of photosynthesis loss in parasitic plants.

Key words: *Taxillus chinensis*, nanopore sequencing, Hi-C proximity mapping, chromosomal assembly.

Significance

Taxillus chinensis is a hemiparasitic plant with photosynthetic ability and has been difficult to cultivate due to its drought- and cold-sensitive seeds and a poor understanding of its genome. The present study succeeded in constructing a high-quality reference genome. This genome will be a valuable resource for elucidating the evolution of photosynthesis loss and the genetic mechanisms that underlie the recalcitrance of the species' seeds.

Introduction

Taxillus chinensis (DC.) Danser (Loranthaceae; fig. 1A) is a root-hemiparasitic plant found in southern China and Southeast Asia (Liu, Su, et al. 2019). The species has been reported to produce neuroprotective compounds, such as triterpenes, lectins, polysaccharides, and alkaloids (Wong et al. 2012), and possesses great potential for medical application, owing to its antioxidant, antiinflammatory, and

antiproliferative properties (Liu et al. 2012). Indeed, the species is widely used as a traditional Chinese medicine for the treatment of rheumatism, threatened abortion, hypertension, angina pectoris, stroke, and arrhythmia (Li et al. 2017).

The available genomic resources for parasitic plants are very limited. The first genome sequence of a shoot-parasitizing plant that depends on host plants to produce

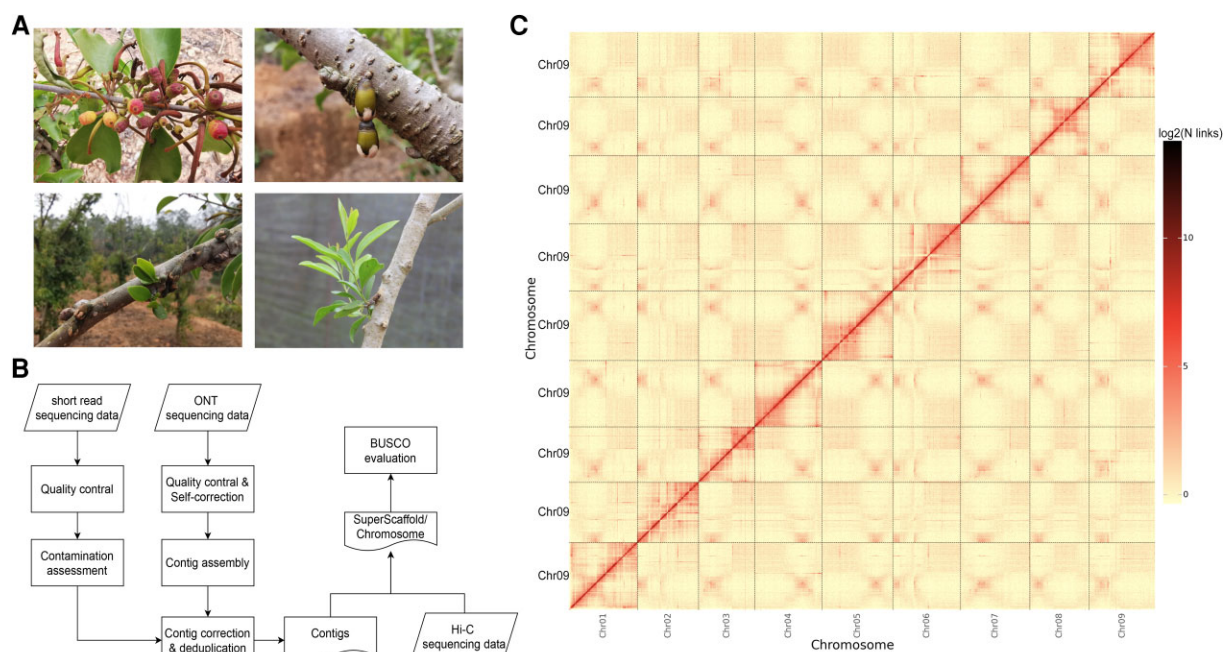


FIG. 1.—Genome assembly of *Taxillus chinensis*. (A) *Taxillus chinensis*. (B) Workflow used to generate the chromosome-level genome assembly. (C) Genome-wide Hi-C heatmap of chromatin interaction counts in 100 kb bins. Only sequences anchored on chromosomes are shown. The abbreviations Chr01–09 represent the nine chromosomes, and the color bar represents the \log_2 value of interaction counts.

photoassimilates was reported for *Cuscuta campestris* Yunck. (Convolvulaceae) in 2018 (Vogel et al. 2018). The genome assembly of the root parasite *Santalum album* (Santalaceae), which instead exhibits strong photosynthetic ability, has been sequenced and deposited in the NCBI database (GCA_002911635.1). Yet, the reference genome of *T. chinensis* has not been reported, despite the fact that its plastome has been sequenced and analyzed phylogenetically (Liu, Zhang, et al. 2019).

The root-hemiparasitic *T. chinensis* has photosynthetic ability (Tesitel et al. 2015) and, therefore, a high-quality genomic reference will provide a valuable resource for the investigation of the evolution of photosynthesis loss in parasitic plants. Besides, *T. chinensis* can only be propagated by seed, and its seed is generally recalcitrant, exhibiting sensitivities to both dehydration and low temperature (Pan et al. 2021), which ultimately hinder the species' utilization. Previous transcriptomic and proteomic studies have investigated the molecular mechanisms associated with the dehydration tolerance of *T. chinensis* (Wei et al. 2017; Pan et al. 2021), and cold stress-related differentially expressed microRNAs (miRNAs) have also been reported (Fu et al. 2021). However, despite the insight these studies have provided into the recalcitrance, the whole-genome sequence of *T. chinensis* is still needed to fully understand the molecular mechanisms involved in the species' seed recalcitrance.

The rapid development of high-throughput sequencing techniques has enabled the generation of chromosome-

level genome assemblies for a variety of species. Thus, the aim of the present study was to use nanopore, short-read, and high-throughput chromosome conformation capture (Hi-C) sequencing to construct a chromosome-level assembly of the *T. chinensis* genome. The generation of a high-quality genome assembly for *T. chinensis* will provide a valuable genetic resource for investigating the evolution of photosynthesis loss in parasitic plants and the species' seed recalcitrance.

Results and Discussion

Genome Assembly

The present study used Oxford Nanopore Technologies (ONT) sequencing technology and Hi-C-assisted genome assembly to generate a chromosome-level genome assembly for *T. chinensis* (fig. 1B). The ONT reads (51.65 Gb) provided $\sim 101\times$ coverage, and the mean long-read length and N50 were 23.14 and 27.31 kb, respectively (supplementary table S1, Supplementary Material online). A total of 216.71 Gb clean short-read sequencing data ($\sim 427\times$ coverage) were used for subsequent polishing.

The contig N50 of the draft genome assembly was about 3.80 Mb (table 1). Hi-C sequencing yielded 95.89 Gb clean reads ($\sim 189\times$ coverage; supplementary table S1, Supplementary Material online), and 86.67% of the Hi-C data were aligned to the draft genome (supplementary

Table 1

Genome Sequencing, Assembly, and Annotation Statistics

	Statistics
Genome assembly and chromosomes construction	
Contig N50 size (bp)	3,797,897
Contig N90 size (bp)	554,497
Maximum contig size (bp)	13,585,695
Scaffold number	434
Scaffold N50 (bp)	56,927,202
Scaffold N90 (bp)	47,601,956
Maximum scaffold size (bp)	59,987,258
Genome size (bp)	521,908,327
Number of chromosomes	9
Total length of chromosomes (bp)	496,429,085
GC content (%)	40.17
Genome quality evaluation	
Proportion of complete BUSCO orthologs (%)	95
Proportion of complete and single-copy BUSCO orthologs (%)	92.4
Proportion of complete and duplicated BUSCO orthologs (%)	2.6
Proportion of fragmented BUSCO orthologs (%)	1.5
Proportion of missing BUSCO orthologs (%)	3.5
Gene annotation	
Number of GO annotation	9,362
Number of KEGG annotation	19,863
Number of KOG annotation	20,225
Number of TrEMBL annotation	28,335
Number of Interpro annotation	26,400
Number of SwissProt annotation	21,376
Number of NR annotation	27,967
Number of all annotated	33,894

table S2, Supplementary Material online). Hi-C-PRO detected 40,286,727 valid read pairs (supplementary table S3, Supplementary Material online), which yielded a final chromosome-level genome assembly of 521.9 Mb, with a scaffold N50 of 56.90 Mb (table 1). The final genome size was close to the estimated genome size by 17-mer analysis (a genome size of 507 Mb and heterozygosity of 0.632%).

The nine chromosomes could be easily distinguished, and the interaction signal intensity around the diagonal of the genome-wide Hi-C heatmap was considerably stronger than that at other positions (fig. 1C), which indicated that the chromosome-level genome assembly was high quality. In addition, BUSCO evaluation indicated that the final genome contained 95% complete genes in the “embryophyta_odb10” ortholog set (table 1), thereby confirming that the genome assembly was complete and of high quality.

Genome Annotation

The identified repetitive sequences (291.23 Mb) comprised 55.8% of the whole-genome assembly (supplementary table S4, Supplementary Material online). Long terminal

repeat (LTR) retrotransposons (50.7%) and DNA elements (3.65%) were the most abundant repeat types (supplementary table S4, Supplementary Material online), which is consistent with the high abundance of LTRs generally observed in the plant kingdom (Gao et al. 2016). Meanwhile, tandem repeats (23.31 Mb) comprised 4.47% of the whole-genome assembly.

A total of 33,894 protein-coding genes, with a mean length of 3,854.56 bp, were predicted through the integration of de novo, homologous, and RNA-seq-based methods (supplementary table S5 and fig. S1, Supplementary Material online). BUSCO assessment indicated that all of the 1,440 genes typically conserved in plants were present (1,351 single-copy and 56 duplicated), thereby indicating high-quality gene annotation, and 93 protein-coding genes were predicted to be photosynthesis related (supplementary table S6, Supplementary Material online). Noncoding RNAs included 48 miRNAs, 537 transfer RNAs (tRNAs), 755 ribosomal RNAs (rRNAs), and 1,042 small nucleolar RNAs (snRNAs; supplementary table S7 and fig. S1, Supplementary Material online).

Gene Family and Domain Identification

A total of 19,426 (57.31%) genes were identified using hmsearch and were clustered into groups (2,280 gene families and 2,164 protein domains). The 20 most abundant gene families included the pentatricopeptide repeat (PPR)-containing proteins PPR, PPR_1, and PPR_2 (supplementary fig. S2, Supplementary Material online), which are reportedly duplicated more often in the genome of the parasitic plant *C. campestris* than in those of other dicots. Eleven family genes related to photosystems I and II (Photo_RC, PsaD, PsaL, PsaN, Psb28, PsbH, Psbl, PsbK, PsbN, PsbQ, PsbR, PsbT, PsbW, PsbX, PsbY, PSI_8, PSI_PsaF, PSI_PsaH, PSI_PSAK, PSII, and PSII_Pbs27) were also identified (supplementary table S8, Supplementary Material online). These findings coincide with the hemiparasitic characteristics of *T. chinensis*, which exhibits photosynthetic ability.

Taxillus chinensis-Specific Genes and Gene Losses

Both shared and unique orthogroups were identified in the *T. chinensis* genome, when compared with genomes of the model organism *Arabidopsis thaliana*, *Malania oleifera* (Santalales), *Cuscuta australis* (Sun et al. 2018), and the shoot-parasitic *C. campestris* (Vogel et al. 2018) (supplementary fig. S3, Supplementary Material online). Gene Ontology (GO) enrichment analysis indicated that shared orthogroups that were absent in *T. chinensis* were associated with a variety of processes, including glucosyltransferase and nutrient reservoir activities (supplementary table S9, Supplementary Material online), whereas the *T. chinensis*-specific genes were significantly

enriched in “mitochondrial RNA metabolism,” “carbohydrate derivative metabolism,” “organic cyclic compound metabolism,” “glycosyl compound metabolism,” “transport,” and “purine ribonucleotide metabolism” (supplementary table S10, Supplementary Material online).

Conclusion

In this study, we present a chromosome-level genome assembly of *T. chinensis* using Nanopore sequencing, supplemented with short-read sequencing and Hi-C sequencing. The final genome assembly was grouped into nine pseudochromosomes with a size of 521.9 Mb. The gene prediction identified multiple genes related to photosystems I and II, coinciding with the hemiparasitic characteristics of *T. chinensis*, which exhibits photosynthetic ability. Furthermore, orthogroups found in *T. chinensis* that were absent from *C. campestris* and *C. australis* were enriched in “chloroplast nucleoid,” “chloroplast stroma,” and “chloroplast” (supplementary table S11, Supplementary Material online), which confirmed that there were differences in the lifestyles of hemiparasitic and parasitic plants and that the latter cannot support themselves by photosynthesis (Sun et al. 2018; Vogel et al. 2018). The high-quality reference *T. chinensis* genome generated in the present study represents the first genomic resource reported for hemiparasitic plants and will facilitate future investigations of the recalcitrance of the species’ seeds and make an evolutionary insight into the mechanisms of photosynthesis loss in parasitic plants more accessible.

Materials and Methods

Sample Collection and DNA Extraction

Tender *T. chinensis* leaves were collected from the Germplasm Resources Nursery of the Guangxi Botanical Garden of Medicinal Plants (Nanning, China) (22°51'2" E and 108°22'44" N latitude, altitude 57 m). Then, genomic DNA was extracted from fresh leaf tissue (200 mg), which had been ground in liquid nitrogen, using Cetyltrimethylammonium Bromide (CTAB) buffer (incubation for 60 min at 65 °C) and was purified using phenol/chloroform/isopentyl (25:24:1), isopropyl alcohol, and ethanol precipitation. The resulting purified DNA was resuspended in Tris–EDTA buffer for subsequent sequencing.

Library Construction and Genome Sequencing

Size selection was performed using BluePippin (Sage Science, Beverly, MA, USA), and 1 µg recovered genomic DNA (20 kb insert size) was subjected to damage repair, end repair, and purification.

A Nanopore sequencing library was prepared from the resulting high-quality DNA using the SQK-LSK109

Ligation Sequencing Kit (ONT, Oxford, UK), according to the manufacturer’s recommendations, evaluated using Qubit, and then sequenced using a MinION long-read sequencer (ONT).

Two short-read sequencing libraries, with insert sizes of 270 and 500 bp, were constructed from the resulting high-quality DNA. The DNA was subject to fragmentation (Covaris, Woburn, MA, USA) and end repair, followed by adaptor ligation, which enabled the formation of circular DNA molecules and subsequent rolling circle amplification to produce DNA nanoballs (DNBs). To prepare the Hi-C library, cells of the sample were treated with formaldehyde to cross-link DNA–protein or protein–protein complexes, and then subject to fragmentation, end repair, purification, and adaptor ligation. The short-read sequencing libraries and Hi-C library were sequenced using the DNBSEQ platform (MGI, Shenzhen, China) in paired-end mode.

Genome Assembly and Assessment

The short-read sequences were filtered using SOAPnuke (v1.6.5, -n 0.01 -q 0.1 -l 20 -Q 2 -M 2 -A 0.5; Chen et al. 2018) to remove low-quality reads and adapter contamination. Based on the short reads, genome size and heterozygosity were estimated using GenomeScope (Vurture et al. 2017) and JELLYFISH (Marçais and Kingsford 2011), respectively.

A draft assembly was generated from the ONT sequencing data using Necat (GENOME_SIZE = 507 Mb; Chen et al. 2021) and polished using Racon (Vaser et al. 2017). A consensus sequence was then constructed from the draft assembly using Medaka (<https://github.com/nanoporetech/medaka>), and the short-read sequence data were used to correct and polish the draft assembly using a pilon (Walker et al. 2014). HaploMerger2 was then used to improve contiguity and reduce duplication.

The contigs were anchored to the chromosomes using Hi-C data. In brief, the Hi-C reads were filtered using SOAPnuke (v1.6.5, -n 0.01 -q 0.1 -l 20 -Q 2 -M 2 -A 0.5), and unique mapped read pairs were selected using the HiC-Pro v2.5.0 pipeline (Servant et al. 2015) to obtain valid interaction pairs. Then, Juicer (Durand et al. 2016) was used to align the sequence against the draft genome assembly, and 3D-DNA (Dudchenko et al. 2017) was used to construct a chromosome-level assembly. Finally, genome quality was evaluated using BUSCO v3 with the “embryophyta_odb10” ortholog set (Simão et al. 2015).

Repeat Sequence Annotation

A de novo repeat library was generated using RepeatModeler (Flynn et al. 2020) and LTRfinder v1.07 (Xu and Wang 2007) with default parameters, and predicted repetitive sequences in the de novo repeat library were identified using RepeatMasker v4.0.7 (Tarailo-Graovac and Chen 2009). At

the same time, homologous prediction of the repeats was performed using RepeatMasker v4.0.7 (Tarailo-Graovac and Chen 2009) and RepeatProteinMasker v4.0.7 (<http://www.repeatmasker.org/cgi-bin/RepeatProteinMaskRequest>) with the Repbase v21.12 database (Bao et al. 2015). The two sets of predicted repeats were then combined to generate nonredundant repetitive sequences. Tandem repeats were identified using Tandem Repeats Finder v4.09 (Benson 1999).

Gene Prediction and Functional Annotation

Gene annotation was performed using Maker v2.31.8 (Holt and Yandell 2011) and protein sequences from six closely related species (*A. thaliana*, *Vitis vinifera*, *Olea europaea*, *Solanum lycopersicum*, *Solanum tuberosum*, and *Fragaria esca*). Three thousand complete predicted genes were used as a training set for de novo prediction with Augustus (Stanke et al. 2006) and SNAP (Johnson et al. 2008). In addition, transcriptomic data of 10 *T. chinensis* seed samples (NCBI accession SRP201073) were combined for auxiliary gene annotation. Briefly, the RNA-seq data were aligned to the genome using HISAT2 v2.1.0 (Kim et al. 2015), assembled using StringTie v1.3.4d (Pertea et al. 2015), and corrected using Pasa_lite (https://github.com/PASAPipeline/PASA_Lite). Based on protein sequences from the six related species, the assembled transcripts, and the Augustus and SNAP models, the annotation data were consolidated using Evidence Modeler (Haas et al. 2008) and Maker (Holt and Yandell 2011). The final consensus gene sets were assessed using BUSCO v3 with the “embryophyta_odb10” ortholog set (Simão et al. 2015).

To obtain the functional annotation, BLAST v2.2.31 (Altschul et al. 1990) was used to align the predicted genes to the nonredundant protein sequences (NRs; Marchler-Bauer et al. 2011), SwissProt (Boeckmann et al. 2003), Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto 2000), eukaryotic orthologous groups of proteins (Tatusov et al. 2003), translation of European Molecular Biology Laboratory EMBL (Boeckmann et al. 2003), InterPro (Apweiler et al. 2001), and GO databases.

For the prediction of noncoding RNA, tRNAs were annotated using tRNAscan-SE v1.3.1 (Lowe and Eddy 1997). Because rRNAs are highly conserved, rRNAs were identified using blastn (Altschul et al. 1990) and the rRNA sequences of related species as a reference. The INFERNAL (<http://infernal.janelia.org/>) software and Rfam database (Griffiths-Jones et al. 2005) were used to predict miRNA and snRNA sequences.

Gene Family and Domain Identification

Gene family and protein domain prediction were performed using HMMER (hmmsearch 3.1b2; Eddy 2011) and the Pfam database (version 34; Mistry et al. 2021),

with the arguments `-domE 1e-3` and an hmm coverage filter (`>45%`) to suppress unreliable domain assignments.

Orthogroup and Functional Enrichment Analysis

Orthogroups were created using OrthoFinder (Emms and Kelly 2019) and genome-wide protein sequences from *T. chinensis* and another four species including *M. oleifera* (Xu et al. 2019), *A. thaliana* (<https://www.arabidopsis.org/>), *C. campestris* (GenBank Accession No. GCA_900332095.2), and *C. australis* (GenBank Accession No. GCA_003260385.1). Orthogroups present in *T. chinensis* but not in the other species were defined as *T. chinensis*-specific genes, whereas those common in the other species but not detected in *T. chinensis* were defined as gene losses. GO enrichment analysis was performed for exclusive orthogroups between *T. chinensis* and two other plants *C. campestris* and *C. australis*, and GO terms with false discovery rates of ≤ 0.05 were defined as significantly enriched.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (82173933, 81860672, 81960695) the Guangxi Natural Science Foundation, China (2021GXNSFBA075037, 2022GXNSFAA035557), the Guangxi Botanical Garden of Medicinal Plants Research and Innovation Team Building Project (GYCH2019008) and the Scientific Research Funding Project of Guangxi Botanical Garden of Medicinal Plants (GYJ202012).

Data Availability

Genomic assembled sequences, ONT raw reads, and raw short reads have been deposited to NCBI database under Bioproject accession no. PRJNA781352.

Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.
- Apweiler R, et al. 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* 29(1):37–40.
- Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27(2):573–580.
- Boeckmann B, et al. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31(1):365–370.

- Chen Y, et al. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7(1):1–6.
- Chen Y, et al. 2021. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun.* 12(1):60.
- Dudchenko O, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356(6333):92–95.
- Durand NC, et al. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3(1):95–98.
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol.* 7(10):e1002195.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20(1):238.
- Flynn JM, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 117(17):9451–9457.
- Fu J, et al. 2021. Identification of microRNAs in *Taxillus chinensis* (DC.) Danser seeds under cold stress. *Biomed Res Int.* 2021:5585884.
- Gao D, Li Y, Kim KD, Abernathy B, Jackson SA. 2016. Landscape and evolutionary dynamics of terminal repeat retrotransposons in miniature in plant genomes. *Genome Biol.* 17:7.
- Griffiths-Jones S, et al. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* 33(Database issue): D121–D124.
- Haas BJ, et al. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9(1):R7.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* 12:491.
- Johnson AD, et al. 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24(24): 2938–2939.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28(1):27–30.
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12(4):357–360.
- Li Y, et al. 2017. Gene losses and partial deletion of small single-copy regions of the chloroplast genomes of two hemiparasitic *Taxillus* species. *Sci Rep.* 7(1):12834.
- Liu C-Y, et al. 2012. Antioxidant, anti-inflammatory, and antiproliferative activities of *Taxillus sutchuenensis*. *Am J Chin Med.* 40(2): 335–348.
- Liu R, et al. 2019. Identification and analysis of cardiac glycosides in Lorantheaceae parasites *Taxillus chinensis* (DC.) Danser and *Scurrula parasitica* Linn. and their host *Nerium indicum* Mill. *J Pharm Biomed Anal.* 174:450–459.
- Liu B, Zhang Y, Shi Y. 2019. Complete chloroplast genome sequence of *Taxillus chinensis* (Lorantheaceae): a hemiparasitic shrub in South China. *Mitochondrial DNA B Resour.* 4(2):3077–3078.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25(5):955–964.
- Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6): 764–770.
- Marchler-Bauer A, et al. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39(Database issue):D225–D229.
- Mistry J, et al. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49(D1):D412–D419.
- Pan L, et al. 2021. Comparative proteomic analysis of parasitic lorchanthus seeds exposed to dehydration stress. *Plant Biotechnol Rep.* 15:95–108.
- Pertea M, et al. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 33(3): 290–295.
- Servant N, et al. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16:259.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19): 3210–3212.
- Stanke M, et al. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34(Web Server issue):W435–W439.
- Sun G, et al. 2018. Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nat Commun.* 9(1): 2683.
- Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinform.* Chapter 4:Unit 4.10.
- Tatusov RL, et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinform.* 4:41.
- Tesitel J, et al. 2015. Integrating ecology and physiology of root-hemiparasitic interaction: interactive effects of abiotic resources shape the interplay between parasitism and autotrophy. *New Phytol.* 205(1):350–360.
- Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27(5):737–746.
- Vogel A, et al. 2018. Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nat Commun.* 9(1): 2515.
- Vurture GW, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33(14):2202–2204.
- Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
- Wei S, et al. 2017. Transcriptome analysis of *Taxillus chinensis* (DC.) Danser seeds in response to water loss. *PLoS One* 12(1): e0169177.
- Wong DZH, Kadir HA, Ling SK. 2012. Bioassay-guided isolation of neuroprotective compounds from *Loranthus parasiticus* against H₂O₂-induced oxidative damage in NG108-15 cells. *J Ethnopharmacol.* 139(1):256–264.
- Xu C-Q et al. 2019. Genome sequence of *Malaria oleifera*, a tree with great value for nervonic acid production. *GigaScience* 8(2):giy164.
- Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35(Web Server issue):W265–W268.

Associate editor: Maud Tenaillon