

Chromosomal-Level Reference Genome of the Neotropical Tree *Jacaranda mimosifolia* D. Don

Mingcheng Wang^{1,*}, Lei Zhang², and Zhiqiang Wang^{1,*}

¹Institute for Advanced Study, Chengdu University, China

²Key Laboratory of Ecological Protection of Agro-pastoral Ecotones in the Yellow River Basin, National Ethnic Affairs Commission of the People's Republic of China, College of Biological Science & Engineering, North Minzu University, Yinchuan, China

*Corresponding authors: E-mails: wangmcyao@163.com; wangzhiqiang@cdu.edu.cn.

Accepted: 26 April 2021

Abstract

Jacaranda mimosifolia D. Don is a deciduous tree widely cultivated in the tropics and subtropics of the world. It is famous for its beautiful blue flowers and pinnate compound leaves. In addition, this tree has great potential in environmental monitoring, soil quality improvement, and medicinal applications. However, a genome resource for *J. mimosifolia* has not been reported to date. In this study, we constructed a chromosome-level genome assembly of *J. mimosifolia* using PacBio sequencing, Illumina sequencing, and Hi-C technology. The final genome assembly was ~707.32 Mb in size, 688.76 Mb (97.36%) of which could be grouped into 18 pseudochromosomes, with contig and scaffold N50 values of 16.77 and 39.98 Mb, respectively. A total of 30,507 protein-coding genes were predicted, 95.17% of which could be functionally annotated. Phylogenetic analysis among 12 plant species confirmed the close genetic relationship between *J. mimosifolia* and *Handroanthus impetiginosus*. Gene family clustering revealed 481 unique, 103 significantly expanded, and 16 significantly contracted gene families in the *J. mimosifolia* genome. This chromosome-level genome assembly of *J. mimosifolia* will provide a valuable genomic resource for elucidating the genetic bases of the morphological characteristics, adaption evolution, and active compounds biosynthesis of *J. mimosifolia*.

Significance

Jacaranda mimosifolia D. Don is an ornamental plant with great potential in environmental monitoring, soil quality improvement, and medicinal applications. Here we present a chromosome-level genome assembly of *J. mimosifolia*. This high-quality genome assembly will provide a valuable genomic resource for elucidating the genetic bases for elucidating the genetic bases of the morphological characteristics, adaption evolution, and active compounds biosynthesis of *J. mimosifolia*.

Key words: *Jacaranda mimosifolia*, genome assembly, PacBio sequencing, gene family evolution.

Introduction

Jacaranda mimosifolia D. Don is a deciduous tree belonging to the Bignoniaceae family that originated in South America (Mostafa et al. 2014; Ragsac et al. 2019). It is now widely cultivated in the tropics and subtropics throughout the world (Gentry 1992). It is commonly used as an ornamental plant due to its beautiful blue flowers and pinnate compound

leaves (Mostafa et al. 2014; Hendra et al. 2019). In addition to its horticultural value, *J. mimosifolia* was found to be a useful biomonitor of soil metals and atmospheric metals (Olowoyo, van Heerden, and Fischer 2010; Olowoyo, Van Heerden, Fischer, and Baker 2010) and can also play an important role in soil quality improvement (Rigal et al. 2020). A few active compounds, including jacaric acid (Takagi and

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Itabashi 1981), jacaranone (Xu et al. 2003), phenylethanoids (Rana et al. 2013; Hendra et al. 2019), and flavonoids (Naz et al. 2020), are commonly isolated from different tissues of *J. mimosifolia*. Thus, this species has a great potential in medicinal applications due to its antioxidant, antimicrobial, and hypotensive properties (Mostafa et al. 2014). To date, studies of *J. mimosifolia* at the molecular level have been limited to karyotypes (Collevatti and Dornelas 2016; Cordeiro et al. 2016), plastid genomes (Zhao et al. 2019), inter-simple sequence repeat (ISSR) markers (Escandón et al. 2003), and very few nuclear genes (Kram et al. 2008). In a previous study, *J. mimosifolia* was confirmed to be diploid ($2n = 2x = 36$) (Collevatti and Dornelas 2016). However, the genome of *J. mimosifolia* has not yet been sequenced or reported. A recent study reported the genome sequence of its close relative *Jacaranda copaia*, which was found to have up to 5,869 scaffolds and an N50 size of 1.02 Mb (Burley et al. 2021). With the rapid development of sequencing technologies, more and more researchers are performing chromosome-level genome assembly using single-molecule real-time (SMRT) and chromosome conformation capture (Hi-C) technologies (Bickhart et al. 2017; Mahajan et al. 2018; Zhang et al. 2020). Here, we generated a chromosome-level genome assembly of *J. mimosifolia* by integrating PacBio sequencing, Hi-C sequencing, and Illumina sequencing. The high-quality genome assembly will provide a valuable genomic resource for elucidating the genetic bases of the morphological characteristics, adaptation evolution, and active compounds biosynthesis of *J. mimosifolia*. It will also provide a reliable reference for studying the evolutionary history of Bignoniaceae and Lamiales.

Results and Discussion

Genome Sequencing, Assembly, and Quality Assessment

We generated a total of 149.21 Gb PacBio reads, 60.79 Gb Illumina reads, and 89.86 Gb Hi-C reads (supplementary table S1, Supplementary Material online). Before assembly, we obtained a genome size estimation of 739 Mb with a heterozygosity of 0.23% based on 19-mer frequency distribution analysis (supplementary fig. S1, Supplementary Material online), which was very close to the result obtained from the flow cytometry analysis (supplementary fig. S2, Supplementary Material online). The PacBio reads were preliminarily assembled into 218 contigs, with a contig N50 of 17.54 Mb and a total length of 707.32 Mb. These contigs were then anchored into 18 pseudochromosomes based on the Hi-C data (supplementary fig. S3 and table S2, Supplementary Material online). Among the 218 contigs, 101 were grouped into pseudochromosomes with a total length of 688.76 Mb, accounting for 97.36% of the entire genome assembly (fig. 1A and supplementary table S2, Supplementary Material online). The contig and scaffold N50 of the final assembly were 16.77 and

39.98 Mb (table 1 and supplementary table S3, Supplementary Material online), respectively. About 1,588 (98.39%) of the 1,614 conserved BUSCO genes were completely covered in our assembly (supplementary table S4, Supplementary Material online), suggesting the high completeness of the *J. mimosifolia* genome assembly.

Genome Annotation

We estimated that 401.81 Mb (56.80%) of the *J. mimosifolia* genome was composed of repetitive elements (supplementary table S5, Supplementary Material online). Long terminal repeats (LTRs) were the most abundant repeat class, accounting for 49.50% of repetitive sequences representing 28.12% of the genome. Within the LTR family, *Copia* and *Gypsy* were the two most abundant subfamilies, representing 14.87% and 12.34% of the genome, respectively.

A total of 30,507 protein-coding genes were predicted, with mean transcript size of 3,442 bp, an average coding sequence size of 1,231 bp, and an average of 5.24 exons per gene. The annotated protein-coding genes covered 91.82% of the 1,614 conserved BUSCO genes (supplementary table S4, Supplementary Material online), indicating the high completeness of the gene annotation. Overall, 29,035 (95.17%) of the protein-coding genes could be annotated by at least one database, and 24,261 (79.53%) genes could be classified by gene ontology (GO) terms (supplementary table S6, Supplementary Material online).

Evolution of Gene Families

A total of 194 single-copy orthogroups were identified across the 12 plant genomes (supplementary table S7, Supplementary Material online). As expected, the species tree constructed through coalescent-based analysis on the basis of these single-copy genes confirmed that *J. mimosifolia* was mostly related to *Handroanthus impetiginosus* (fig. 1B). The divergence time between *J. mimosifolia* and *H. impetiginosus* was estimated to be 39.9 million years ago (Ma), slightly after the separation between Bignoniaceae and Pedaliaceae at 48.7 Ma.

Gene family clustering analysis identified a total of 29,545 gene families comprising 354,790 genes among *J. mimosifolia* and 11 other species, of which 5,118 families were present in all 12 species. Among the 12 species, 481 gene families consisting of 2,034 genes were unique to *J. mimosifolia* (supplementary table S7, Supplementary Material online). GO enrichment analysis revealed that these unique genes were mainly related to "cellular process," "pteridine-containing compound metabolic process," "isoflavonoid biosynthetic process," "positive regulation of response to water deprivation," and "zinc ion binding" (supplementary fig. S4, Supplementary Material online). We also identified 103 gene families comprising 1,268 genes that were significantly

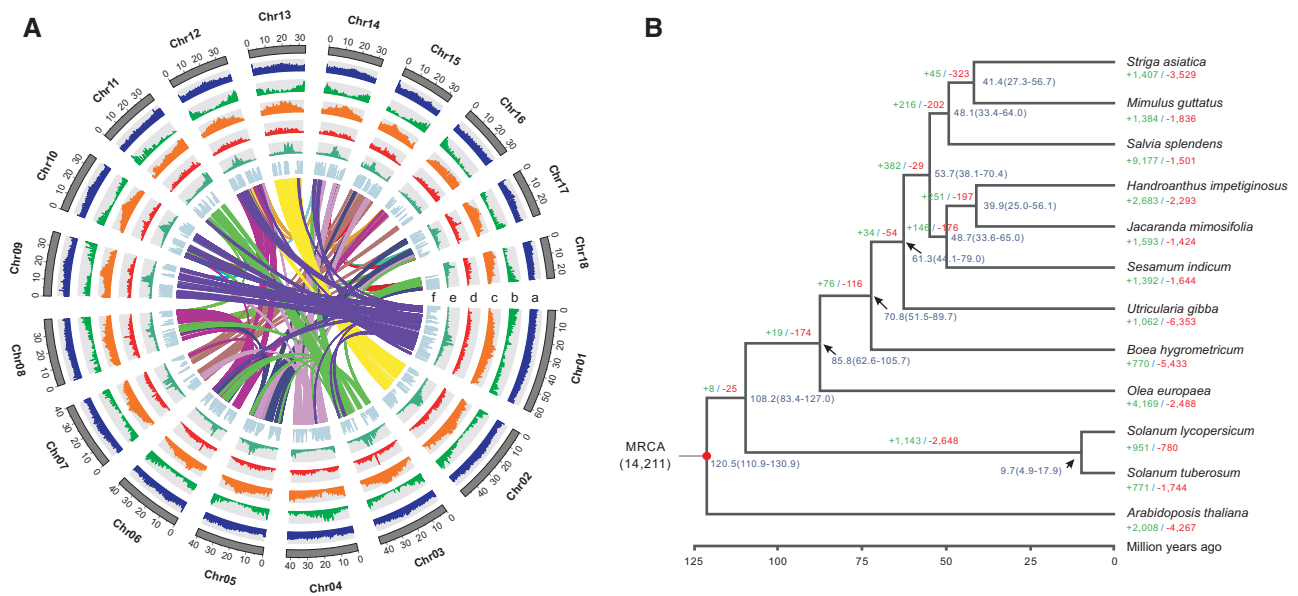


FIG. 1.—Features of the *J. mimosifolia* genome (A) and phylogeny and gene family evolution of *J. mimosifolia* (B). (A) Genome features in non-overlapping windows of 500 Kb across the *J. mimosifolia* genome. Tracks from outside to inside are as follows: (a) GC content, (b) gene density, (c) repeat density, (d) LTR/Gypsy density, (e) LTR/Copia density, and (f) density of collinear blocks between *J. mimosifolia* and *H. impetiginosus*. (B) A species tree on the basis of 194 single-copy orthogroups from 12 plant species. Gene family expansion and contraction are denoted in green and red numbers, respectively. Numbers on nodes represent the inferred divergence times with 95% confidence intervals. Red dots represent the calibration time of *A. thaliana*–*S. tuberosum* divergence.

expanded in *J. mimosifolia* ($P < 0.01$). GO enrichment analysis revealed that these genes were highly enriched in “brassinosteroid metabolic process,” “defense response,” “pigment biosynthetic process,” and “detoxification” (supplementary fig. S5, Supplementary Material online). By contrast, a total of 16 gene families comprising 71 genes were significantly contracted in the *J. mimosifolia* genome. These genes were highly enriched in “response to symbiotic bacterium,” “mucilage extrusion from seed coat,” “proteolysis,” “plant-type cell wall modification,” and “GTP binding” (supplementary fig. S6, Supplementary Material online).

Conclusions

In this study, we present a high-quality chromosome-level genome assembly of *J. mimosifolia* using PacBio single-molecule long-read sequencing, supplemented with Illumina short-read sequencing and Hi-C sequencing. The final genome assembly was ~ 707.32 , 688.76 Mb (97.36%) of which could be grouped into 18 pseudochromosomes. This assembly is highly contiguous with contig and scaffold N50 values of 16.77 and 39.98 Mb, respectively. A total of 56.80% of the genome was composed of repeat sequences. A total of 30,507 protein-coding genes were predicted, 95.17% of which could be functionally annotated. Phylogenetic analysis on the basis of 194 single-copy orthogroups across 12 plant species confirmed the close genetic relationship between *J. mimosifolia* and *H. impetiginosus*. The divergence time

between these two species was estimated to be 39.9 Ma. Gene family clustering revealed 481 unique, 103 significantly expanded, and 16 significantly contracted gene families in the *J. mimosifolia* genome. The high-quality chromosomal genome assembly of *J. mimosifolia* will provide a valuable genomic resource for elucidating the genetic bases for blue flower color development, the biosynthesis of active compounds, and environmental adaptation as well as for studying the evolutionary history of Bignoniaceae and Lamiales.

Materials and Methods

Sample Preparation and Sequencing

We sequenced a single adult tree of *J. mimosifolia* from Sichuan University in Sichuan Province, southwestern China. Fresh leaves were collected and frozen immediately in liquid nitrogen. High-quality genomic DNA was extracted using the cetyl trimethylammonium bromide (CTAB) method (Doyle and Doyle 1987). For PacBio single-molecule real-time (SMRT) sequencing, at least 10 μg of sheared DNA was used to generate 40-kb SMRTbell libraries according to the manufacturer’s protocol. Subsequently, the SMRTbell libraries were sequenced on the PacBio Sequel II Platform (Pacific Biosciences). A paired-end PCR-free library with an insert size of ~ 350 bp was generated and sequenced on an Illumina HiSeq 2500 platform. For Hi-C sequencing, more than 2 g of young leaves were frozen in liquid nitrogen prior

Table 1Statistics of the assembly and annotation of the *J. mimosifolia* genome

Assembly	
Estimated genome size (by <i>k</i> -mer analysis) (Mb)	739
Total length (Mb)	707.42
GC content (%)	33.87
Contig N50 (Mb)	16.77
Longest contig (Mb)	45.42
Contig number	290
Pseudochromosome number	18
Scaffold N50 (Mb)	39.98
Longest scaffold (Mb)	61.75
Annotation	
Total length of repeats (Mb)	401.81
Number of protein-coding genes	30,507
Mean transcript length (bp)	3,441.77
Mean coding sequence length (bp)	1,230.68
Mean exon length (bp)	234.87
Mean intron length (bp)	521.50
Functionally annotated genes	29,035

to the Hi-C experiment. The Hi-C libraries were then constructed as described previously (Louwers et al. 2009), including chromatin extraction and digestion, DNA ligation, and purification. The DNA was sheared to a mean fragment size of ~350 bp and sequenced on an Illumina HiSeq 2500 platform.

To aid in gene annotation, we performed RNA sequencing for fresh tissues of the leaf, stem, and peel from the same plant used for genome sequencing using an Illumina HiSeq 2500 platform.

Genome Size Estimation

The genome size of *J. mimosifolia* was estimated using two different approaches. First, a *k*-mer depth distribution ($k = 19$) was obtained with Jellyfish v2.2.9 (Marçais and Kingsford 2011) based on the Illumina short reads, and the final result was plotted in GenomeScope. Second, a flow cytometry analysis was performed using *Glycine max* as a reference standard and propidium iodide as the stain.

Genome Assembly and Assessment

The PacBio reads were self-corrected and de novo assembled using CANU (Koren et al. 2017) with a genomeSize parameter of 730 m. Primary contigs were polished using Pilon v1.22 (Walker et al. 2014) with the filtered PacBio reads. Then, Illumina short reads were mapped to the assembly using BWA v0.7.17 (Li and Durbin 2009) and Pilon was used to improve the contigs. Potential allelic haplotigs were identified and removed with purge_dups (Guan et al. 2020). The contigs were finally anchored into chromosomes using 3D-DNA software (Dudchenko et al. 2017) based on the Hi-C data. The

genome quality was evaluated by examining the coverage of highly conserved genes using BUSCO v3.0.2 (Simão et al. 2015) with the “embryophyta_odb10” ortholog set.

Repeat Annotation

Both homolog-based and de novo approaches were used to determine repetitive elements in the *J. mimosifolia* genome. Initially, a repeat library of green plants (Viridiplantae) was derived from the Repbase database v22.11 using RepeatMasker v4.0.7 (Tarailo-Graovac and Chen 2009). Then a de novo repeat library was generated using RepeatModeler v1.0.11 (Price et al. 2005), with default parameters. Finally, RepeatMasker was used to map the *J. mimosifolia* genome against these two repeat libraries.

Gene Annotation

We performed gene annotation in the repeat-masked genome by using a combination of de novo, homology-based, and transcriptome-based prediction methods. The de novo prediction was conducted using Augustus v3.2.3 (Stanke et al. 2006) with pretrained parameters from *Arabidopsis thaliana*. For the homology-based prediction, protein sequences from five sequenced plant species, namely, *H. impetiginosus* (Silva-Junior et al. 2018), *Olea europaea* (Unver et al. 2017), *Sesamum indicum* (Wang et al. 2014), *Solanum tuberosum* (Potato Genome Sequencing Consortium 2011), and *A. thaliana* (Kaul et al. 2000), were aligned to the *J. mimosifolia* genome using TBLASTN v2.2.26 (Camacho et al. 2009). GeneWise v2.4.1 (Birney et al. 2004) was then used for the accurate prediction of gene models based on the aligned genomic regions. For the transcriptome-based prediction, both de novo and genome-guided approaches were used to assemble the RNA-seq data with Trinity v2.5.1 (Grabherr et al. 2011). All of the assembled transcripts were aligned to the genome assembly, and the gene structures were improved using PASA v2.2.0 (Haas et al. 2008). Finally, all predicted gene models were integrated into a consensus gene set using EVM v1.1.1 (Haas et al. 2008).

Functional annotation of protein-coding genes was achieved with BLASTP against the SwissProt and TrEMBL databases (Bairoch and Apweiler 2000). Protein motifs and domains were annotated using InterProScan v5.31 (Hunter et al. 2009) by searching against multiple publicly available databases. GO IDs for each gene were retrieved with Blast2GO v2.5 (Conesa et al. 2005). The pathways in which the genes might be involved were annotated using the KAAS (KEGG Automatic Annotation Server).

Gene Family Analysis and Phylogenetic Analysis

Gene family clustering analysis was performed using BLASTP and OrthoMCL v2.0.9 (Li et al. 2003) based on the annotated genes of *J. mimosifolia* and 11 other sequenced plant species,

namely, 8 Lamiales (*H. impetiginosus*, *S. indicum*, *Salvia splendens*, *Mimulus guttatus*, *Striga asiatica*, *Utricularia gibba*, *Boea hygrometrica*, and *O. europaea*), 2 Solanales (*S. tuberosum* and *Solanum lycopersicum*), and 1 rosid plants (*A. thaliana*). We identified 194 single-copy orthogroups across the 12 plant genomes. For each orthogroup, protein sequences were aligned using MAFFT v7.313 (Katoh et al. 2002) and conserved sites were extracted from the alignments using Gblocks v0.91b (Castresana 2000). The conserved sequences were concatenated into a single multiple sequence alignment. A maximum-likelihood phylogeny was constructed using RAxML v8.2.11 (Stamatakis 2014) under the PROTGAMMAILGX model with 100 bootstrap replicates. Then MCMCTREE in PAML v4.9e (Yang 2007) was applied to estimate the divergence time among 12 species based on the phylogenetic tree with the main parameters of burnin = 1,000,000, nsample = 20,000, and sampfreq = 150. The divergence time between *A. thaliana* and *S. tuberosum* (111–131 Ma) was obtained from the TimeTree database (<http://www.timetree.org>, last accessed May, 12, 2021) and used as fossil calibration point. Gene family evolution was examined across the time-calibrated phylogenetic tree using CAFÉ v3.1 (De Bie et al. 2006) with the main parameters -p = 0.05 and lambda = -s.

Supplementary Material

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

Acknowledgments

This work was supported by startup funds provided by Chengdu University.

Data Availability

Raw sequencing data and genome assembly have been deposited at the NCBI under the BioProject PRJNA678551 with biosample accession number SAMN16807716. The annotations and predicted peptides are available on FigShare at the link: <https://doi.org/10.6084/m9.figshare.14379086.v1>.

Literature Cited

Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28(1):45–48.

Bickhart DM, et al. 2017. Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat Genet.* 49(4):643–650.

Birney E, Clamp M, Durbin R. 2004. GeneWise and genomewise. *Genome Res.* 14(5):988–995.

Burley JT, Kellner JR, Hubbell SP, Faircloth BC. 2021. Genome assemblies for two Neotropical trees: *Jacaranda copaia* and *Handroanthus guayacan*. *G3 (Bethesda)* 11(2):jkab010.

Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):421–429.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17(4):540–552.

Collevatti RG, Dornelas MC. 2016. Clues to the evolution of genome size and chromosome number in *Tabebuia* alliance (Bignoniaceae). *Plant Syst Evol.* 302(5):601–607.

Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674–3676.

Cordeiro JMP, Lima SAA, Paz SN, Santos MAS, Felix LP. 2016. Karyotype evolution in the genus *Jacaranda* Juss. (Jacarandaeae, Bignoniaceae): chromosome numbers and heterochromatin. *Genet Mol Res.* 15:gmr15048973.

De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22(10):1269–1271.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytoch Bull.* 19:11–15.

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.

Escandón A, De La Torre MP, Acevedo A, Marcucci-Poltri S, Miyajima I. 2005. Anchored ISSR as molecular marker to characterize accessions of *Jacaranda mimosifolia* L. Don. *Acta Hort.* 683:121–127.

Gentry AH. 1992. A synopsis of Bignoniaceae ethnobotany and economic botany. *Ann Mo Bot Gard.* 79(1):53–64.

Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652.

Guan D, et al. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36(9):2896–2898.

Haas BJ, et al. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9(1):R7.

Hendra R, Willis A, Keller PA. 2019. Phytochemical studies on the Australian native plant species *Acacia pycnantha* and *Jacaranda mimosifolia* D. Don. *Nat Prod Res.* 33(14):1997–2003.

Hunter S, et al. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Res.* 37(Database issue):D211–D215.

Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30(14):3059–3066.

Kaul S, et al. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408(6814):796–815.

Koren S, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27(5):722–736.

Kram BW, Bainbridge EA, Perera MAD, Carter C. 2008. Identification, cloning and characterization of a GDSL lipase secreted into the nectar of *Jacaranda mimosifolia*. *Plant Mol Biol.* 68(1–2):173–183.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13(9):2178–2189.

Louwers M, Splinter E, Van Driel R, De Laat W, Stam M. 2009. Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C). *Nat Protoc.* 4(8):1216–1229.

Mahajan S, Wei KHC, Nalley MJ, Gibilisco L, Bachtrog D. 2018. De novo assembly of a young *Drosophila* Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biol.* 16(7):e2006348.

- Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6):764–770.
- Mostafa NM, Eldahshan OA, Singab ANB. 2014. The genus *Jacaranda* (Bignoniaceae): an updated review. *Pharmacogn Commun.* 4(3):31–39.
- Naz R, et al. 2020. GC-MS analysis, antimicrobial, antioxidant, antilipoxygenase and cytotoxic activities of *Jacaranda mimosifolia* methanol leaf extracts and fractions. *PLoS One* 15(7):e0236319.
- Olowoyo JO, van Heerden E, Fischer JL. 2010. Investigating *Jacaranda mimosifolia* tree as biomonitor of atmospheric trace metals. *Environ Monit Assess.* 164(1–4):435–443.
- Olowoyo JO, van Heerden E, Fischer JL, Baker C. 2010. Trace metals in soil and leaves of *Jacaranda mimosifolia* in Tshwane area, South Africa. *Atmos Environ.* 44(14):1826–1830.
- Potato Genome Sequencing Consortium. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355):189–195.
- Price AL, Jones NC, Pevzner PA. 2005. *De novo* identification of repeat families in large genomes. *Bioinformatics* 21(Suppl 1):i351–i358.
- Ragsac AC, Farias-Singer R, Freitas LB, Lohmann LG, Olmstead RG. 2019. Phylogeny of the Neotropical tribe Jacarandae (Bignoniaceae). *Am J Bot.* 106(12):1589–1601.
- Rana A, Bhangalia S, Singh HP. 2013. A new phenylethanoid glucoside from *Jacaranda mimosifolia*. *Nat Prod Res.* 27(13):1167–1173.
- Rigal C, Xu J, Vaast P. 2020. Young shade trees improve soil quality in intensively managed coffee systems recently converted to agroforestry in Yunnan Province. *Plant Soil.* 453(1–2):119–137.
- Silva-Junior OB, Grattapaglia D, Novaes E, Collevatti RG. 2018. Genome assembly of the pink ipê (*Handroanthus impetiginosus*, Bignoniaceae), a highly valued, ecologically keystone neotropical timber forest tree. *Gigascience* 7(1):1.
- 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.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Stanke M, et al. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res.* 34(Web Server issue):W435–W439.
- Takagi T, Itabashi Y. 1981. Occurrence of mixture of geometrical isomers of conjugated octadecatrienoic acids in some seed oils: analysis by open-tubular gas liquid chromatography and high performance liquid chromatography. *Lipids* 16(7):546–551.
- Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics.* 25(1):4–10.
- Unver T, et al. 2017. Genome of wild olive and the evolution of oil biosynthesis. *Proc Natl Acad Sci U S A.* 114(44):E9413–E9422.
- Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
- Wang L, et al. 2014. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* 15(2):R39.
- Xu H, Zhang N, Casida JE. 2003. Insecticides in Chinese medicinal plants: survey leading to jacaranone, a neurotoxicant and glutathione-reactive quinol. *J Agric Food Chem.* 51(9):2544–2547.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Zhang Z, et al. 2020. Improved genome assembly provides new insights into genome evolution in a desert poplar (*Populus euphratica*). *Mol Ecol Resour.* 20(3):781–794.
- Zhao K, et al. 2019. Complete plastome sequence of *Jacaranda mimosifolia* D. Don (Bignoniaceae): a beautiful landscaping tree species. *Mitochondrial DNA B Resour.* 4(2):4111–4112.

Associate editor: Tanja Slotte