GBE

De Novo Chromosome-Level Assembly of the Endangered *Pilocarpus Microphyllus* Highlights Genomic Resources for Conservation and Sustainable Pilocarpine Extraction

Mariane Brom Sobreiro [1], Leandro Magalhães [1], Leandro Araújo Argolo [1], Camila Leitão Nacif [1], Carlos Eduardo de Araújo Batista [1], Débora de Morais Cordeiro [1], Renato Renison Moreira-Oliveira [1], Guilherme Oliveira [1], Alexandre Aleixo [1], Carolina da Silva Carvalho [1], Cecílio Frois Caldeira [1], Sibelle Torres Vilaça [1,*], Amanda F. Vidal [1,*]

[1]Environmental Genomics, Instituto Tecnológico Vale, Belém, PA, Brazil

*Corresponding authors: E-mails: sibelle.vilaca@itv.org; amanda.vidal@itv.org.

Accepted: March 10, 2026

Abstract

Pilocarpus microphyllus (jaborandi) is an endangered plant species with significant bioeconomic relevance, as it is the main known source of pilocarpine, an alkaloid widely used in the treatment of glaucoma and other diseases. Here, we present a functionally annotated, haplotype-phased, chromosome-level genome for *P. microphyllus,* combining PacBio HiFi and Hi-C sequencing. The final genome assembly spans 2.6 Gb anchored into 22 chromosomes across 95 scaffolds, with a scaffold N50 of 120.9 Mb and a BUSCO completeness score of 99.66%. We obtained 28,319 unique protein-coding loci, of which 28,090 were functionally annotated to the RefSeq database. Repetitive sequences constituted 88.98% of the total genome length. This near-T2T genome provides a robust molecular foundation for investigating the pilocarpine biosynthetic pathway and supports future population-level studies, thereby contributing to improved management and conservation strategies.

Key words: jaborandi, conservation, bioeconomy, pilocarpine, reference genome.

Significance

Pilocarpus microphyllus (jaborandi) is a plant of great importance because it produces pilocarpine, a compound widely used to treat glaucoma and other medical conditions. Most of pilocarpine is extracted from natural populations, and Brazil is the only country that supplies pilocarpine worldwide. Due to the high commercial demand, jaborandi has been overharvested and classified as endangered in Brazil since 1992. Despite its importance, key questions about how pilocarpine is produced and accumulated in jaborandi leaves remain unanswered, largely due to a lack of genomic information. In this study, we present a high-quality reference genome for *Pilocarpus microphyllus*, which will support future research and conservation efforts for this species.

Introduction

The *Pilocarpus* genus comprises a group of tropical American shrubs popularly known as jaborandi. It includes 17 species, of which 15 are found in Brazil (Pirani 2015). *Pilocarpus microphyllus* Stapf ex Wardlew. (Rutaceae: Sapindales) is a jaborandi species with the greatest bioeconomic relevance since it accumulates the highest rates of pilocarpine in the leaves (Pinheiro 1997). Pilocarpine is a well-known imidazole alkaloid with pharmaceutical properties, widely used for the treatment of glaucoma, stimulation of sweat and lachrymal glands, and prevention of xerostomia (Gornitsky et al. 2004; Sidhu 2014).

Pilocarpus microphyllus is endemic to Brazil and has been extensively overexploited over the years due to the harvesting of its leaves for pilocarpine extraction, as this genus is the only known natural source of this alkaloid (Caldeira et al. 2017). Pilocarpine is mostly obtained from wild populations, and Brazil is its sole global supplier (CNI 2014). As a result of uncontrolled extractivism and Amazon deforestation, *P. microphyllus* has been included on Brazil's official list of endangered plant species since 1992 (IBAMA 1992). This species (2n = 44) is a perennial shrub that typically grows between 1 and 6 m in height and inhabits hot, humid climates (Pirani and Devecchi 2018) (Fig. 1a). It is commonly found in the forest understory and is adapted to nutrient-poor soils and rocky outcrops (Homma and Menezes 2014).

Several efforts have been made to identify alternative chemical sources of pilocarpine, but our limited understanding of its biosynthetic pathway has hindered significant progress (Andreazza et al. 2015; Caldeira et al. 2017). Driven by the high demand for pilocarpine and the decreasing availability of raw material from natural sources, a company launched a jaborandi domestication program, establishing a plantation with approximately 15 million plants. However, evidence indicates that *Pilocarpus* cultivation as a crop markedly reduces the pilocarpine content in its leaves, underscoring the need for comprehensive genomic and functional analyses to clarify the molecular mechanisms underlying pilocarpine biosynthesis (Caldeira et al. 2017).

In this context, providing a high-quality, chromosome-scale, and well-annotated reference genome for this species represents a critical step toward advancing biochemical research, promoting its sustainable use, and supporting its conservation, as such genomic resources are essential for the conservation of endangered species and for enabling population-level studies (Formenti et al. 2022b). In particular, the genome of *P. microphyllus* will also provide the molecular basis for the pilocarpine route studies and contribute to management efforts. To address the pressing need for jaborandi genomic resources, we present here the first functionally annotated, haplotype-phased, chromosome-level genome of *P. microphyllus,* generated using PacBio HiFi and Hi-C sequencing.

Results and Discussion

Genome Assembly

A 47X HiFi sequencing coverage was used for contig-level genome assembly, combined with 70X Hi-C sequencing coverage for chromosome-level scaffolding. Using a k-mer profile from GenomeScope, we estimated a genome size of 2.65 Gb and 1.76% heterozygosity (Fig. 1b). Our final assembly spans 2.62 Gb across 220 contigs and 95 scaffolds, with a scaffold N50 of 120.9 Mb and a contig N50 of 38.6 Mb (Fig. 1c and Table 1). We obtained 22 chromosome sequences, in agreement with the jaborandi's karyotype of 2n = 44, with 16 chromosomes showing telomeric repeats in both ends (Fig. 1d and Table S1). The assembly achieved a BUSCO completeness of 99.66%, including 84.78% single-copy and 14.88% duplicated orthologs, based on the eudicotis_odb10 database (Fig. 1c).

Genome Annotation

We determined that 2,331,861,174 bp of sequences in the genome are repetitive sequences, constituting 88.98% of the total genome length (Fig. S1a). Among these repetitive sequences, long terminal repeat sequences (LTRs) are the most abundant, comprising 1,639,404,537 bp (approximately 62.56% of the genome length) (Fig. S1a). Within LTRs, the elements Gypsy and Copia were the most abundant (Fig. S1a). This result aligns with the transposable element (TE) abundance in other Rutaceae species, such as the orange jessamine *Murraya paniculata* (Yang et al. 2023), and *Citrus*-related genomes (Wu et al. 2014; Giraud et al. 2025). Although *M. paniculata* showed a similar abundance of LTR elements Copia and Gypsy to that of jaborandi, in the genomes of *Citrus*-related species, Copia elements were more abundant than Gypsy. Considering the Kimura-2-parameter distance (Fig. S1b), we observed two peaks of TE expansion, corresponding to LTR elements. However, we observed very few intact elements (~0% distance), which may indicate a lack of recent insertions due to few active TEs in the jaborandi genome.

We predicted a total of 28,319 protein-coding genes, with an average gene length of 5,628 bp. Genes comprised, on average, 5.5 exons and 3.6 introns, with mean exon and intron lengths of 301 and 697 bp, respectively. Additionally, 89% of the annotated transcripts are multiexonic. Gene annotation completeness was assessed with BUSCO analysis using the longest

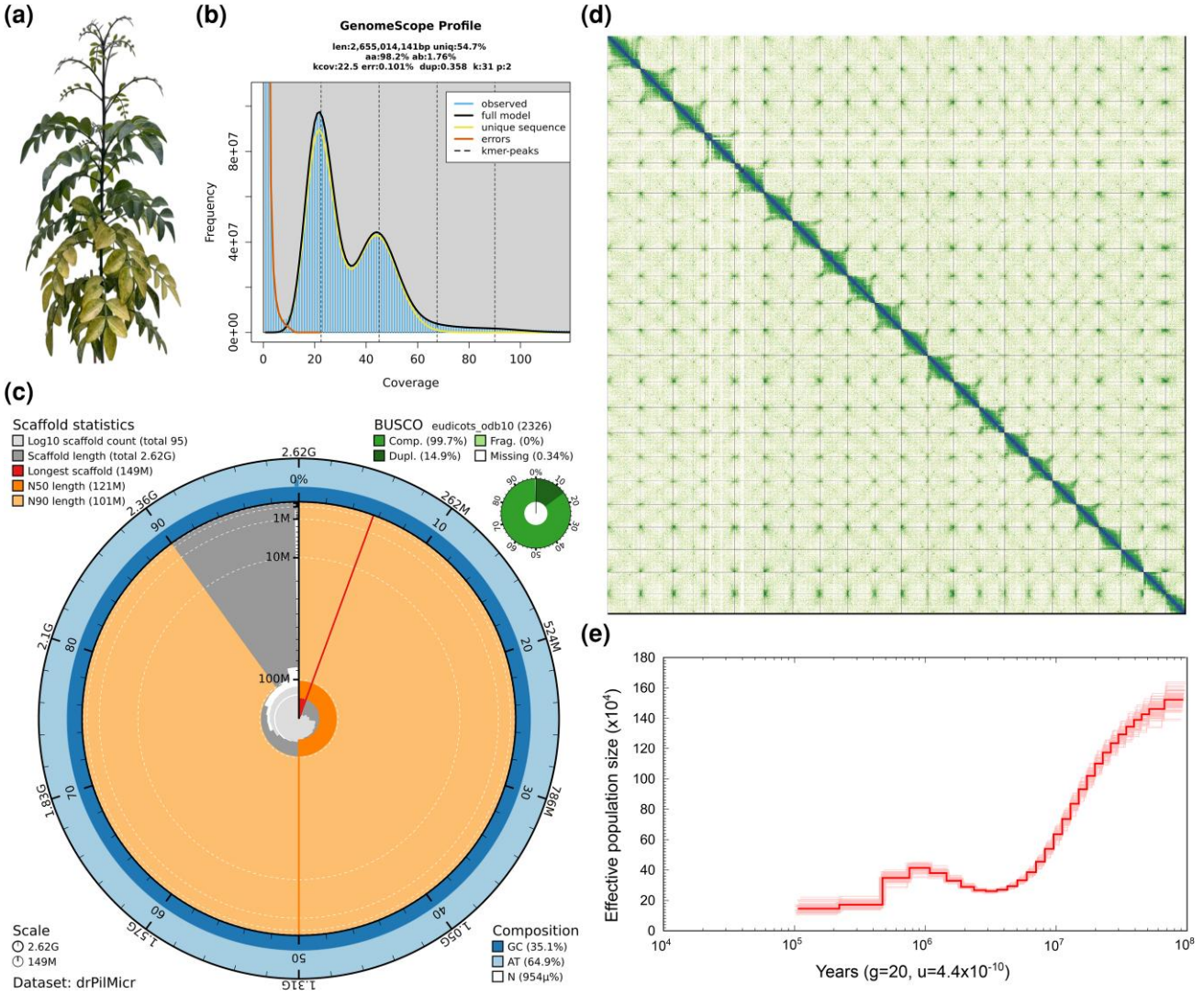

Fig. 1 Chromosome-level genome assembly and evolutionary history of *Pilocarpus microphyllus*. (a) Specimen selected for genome sequencing. (b) GenomeScope k-mer profile. (c) Snail plot for haplotype 1 showing quality metrics. (d) Hi-C contact map of the 22 chromosome-length scaffolds. (e) Past effective population size (Ne) estimates using a PSMC. The dark red line represents the Ne through time, while lighter lines represent bootstrap replicates.

transcript per gene against the eudicotis_odb10 database, revealing 98.8% complete orthologs, including 83.1% single-copy and 15.6% duplicated (Table 1). Functional annotation indicated that 99.19% (28,090 genes) had at least one transcript successfully matched to one or more databases, including EC, KEGG pathway, GO, PFAM, and Panther.

Organellar Genome Assembly

The mitochondrial genome is a single circular sequence with 314,064 bp of length. We annotated 33 protein-coding genes, 3 rRNAs, and 49 tRNAs (Fig. S2). The chloroplast genome is circular with 158,132 bp of length and a quadripartite structure (two inverted repeat regions, one large single copy, and one small single copy). We annotated 80 protein-coding genes, 4 rRNAs and 37 tRNAs (Fig. S3).

Genetic Diversity and Demography

The genome-wide heterozygosity (π) was estimated as 1.10%. There was no significant presence of runs of homozygosity (ROHs), corresponding to only 0.49% of the genome and a maximum ROH length of 0.6 Mb ($F_{ROH} = 0.005$). Given that a 1 Mb threshold for ROH length is used as evidence of inbreeding in recent generations (McQuillan et al. 2008), our estimates show no sign of recent inbreeding despite recent population declines (Monteiro et al. 2022). Our results show that

Table 1 Assembly and annotation summary statistics for the *Pilocarpus microphyllus* genome

Sequencing metrics		
Read type	**Total bases (bp)**	**Coverage**
HiFi	123,700,129,062	~47.20x
Hi-C	182,249,719,686	~69.55x
Assembly metrics		
Measure	**Hap1**	**Hap2**
Total length (bp)	2,620,566,444	2,653,246,189
# Scaffolds	95	214
Gaps	125	149
N count	25,000	29,800
N's per 100 kbp	0.95	1.12
Scaffold N50 (bp)	120,898,695	121,695,687
Scaffold N90 (bp)	101,455,098	101,909,051
Scaffold L50	10	10
Scaffold L90	20	20
Largest scaffold (bp)	148,785,833	151,906,315
# Contigs	220	363
Contig N50 (bp)	38,617,195	30,819,238
Largest contig (bp)	107,134,413	80,921,333
% of assembly in chromosome-length scaffolds	99.76%	99.46%
QV	72.15	69.28
BUSCO completeness (Eudicots_Odb10 *n* = 2326)	C: 99.66% [S: 84.78%; D: 14.88%]; F: 0%; M: 0.34%	C: 99.74% [S: 85.12%; D: 14.62%]; F: 0%; M: 0.26%
GC Content (%)	35.12%	35.13%
Gene annotation statistics		
# Genes		28,319
# Genes with functional annotation		28,090
Total gene length		159,381,658
Mean gene length		5,628
Total CDS length		39,729,591
Mean CDS length		1,403
Mean exons per mRNA		5.5
Mean exon length		301
Mean introns in CDS per mRNA		3.6
Mean intron length		697
BUSCO completeness (Eudicots_Odb10 *n* = 2326)		C: 98.8% [S: 83.1%; D: 15.6%]; F: 0.1%; M: 1.1%

C, Complete; D, Duplicated; F, Fragmented; M, Missing; S, Single Copy.

P. microphyllus has no sign of inbreeding, which may indicate that populations are still large enough to prevent mating between related individuals.

The pairwise sequentially Markovian coalescent (PSMC) analysis, which estimates the past dynamics of effective population size (Ne), revealed high Ne values throughout the species evolutionary history, with a variation in Ne around 1 million years ago (Fig. 1e). Past demographic estimates showed a large population, and the addition of population-level data will increase the resolution of *Ne* estimates to more recent time frames. Overall, our results align with previous studies on jaborandi populations, which have demonstrated high genetic diversity and low inbreeding (Monteiro et al. 2022).

In summary, we provided a high-quality, near telomere-to-telomere genome assembly for *Pilocarpus microphyllus*, being the most complete and contiguous genome available for the entire *Pilocarpus* genus. This annotated genome will serve as a valuable resource for the scientific community, supporting future population genomics analyses and functional transcriptomic studies. These efforts will enhance our understanding of jaborandi genetic diversity and pilocarpine biosynthesis, thereby contributing to the conservation of the species.

Materials and Methods

Sample Acquisition, DNA and RNA Extraction, and Sequencing

Seeds of *Pilocarpus microphyllus*, collected from their natural habitats within the Carajás National Forest

(CNF-Pará, Brazil), were supplied to our institute through a procurement from COEX-Carajás, a cooperative of authorized extractivists permitted to commercialize non-timber forest products sourced from the CNF. Voucher specimens of these plants have been previously deposited in the BHCB herbarium under accession number 158121. These seeds were cultivated under controlled conditions in a growth chamber set to a 12:12 h light–dark photoperiod, with a light intensity of 150 μmol m−2s−1, and a day/night temperature regime of 28 °C/22 °C. We collected fresh samples of leaflets (4 replicates), rachis (3 replicates), root (2 replicates), and stem (2 replicates) from a two-year-old specimen (molecular voucher: ITV30131). This specimen was irrigated daily, cultivated in Carolina Soil®, and received slow-release fertilizer (Osmocote®) every three months.

Total genomic DNA from the leaflets was isolated using the MagAttract HMW DNA Kit (Qiagen). PacBio library was prepared using SMRTbell Express Template Prep Kit 2.0 (Pacific BioSciences) and sequenced on an in-house Sequel IIe (Pacific BioSciences) using SMRT Cell 8 M Tray (Pacific BioSciences). We constructed a Hi-C library from the leaflet sample using the Arima Hi-C 2.0 Kit (Arima), followed by the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs). Hi-C library was sequenced on NextSeq 2000 (Illumina) using a P3 Reagent 300 cycles (2 × 150 bp) (Illumina).

RNA-Seq libraries for the leaflets, rachis, root, and stem were made to aid genome annotation (Sobreiro et al. 2025). Total RNA was extracted using RNAeasy Plant Mini Kit (Qiagen) and quality-evaluated using Qubit (Thermo Fisher Scientific) and TapeStation (Agilent Technologies). RNA Libraries were made using TruSeq Stranded Total RNA with Ribo-Zero Plant (Illumina) and sequenced on NextSeq 2000 (Illumina) using a P3 Reagent 200 cycles (2 × 100 bp) (Illumina).

Nuclear Genome Assembly

The genome was assembled using Pipeasm v1.0.1 (Marques Silva et al. 2026), an automated pipeline developed according to the best practices proposed by the Earth Biogenome Project (EBP) and Vertebrate Genomes Project (VGP) (Blaxter et al. 2025). In summary, Pipeasm comprises six steps: (i) trimming and QC, which is performed using Cutadapt v4.4 (Martin 2011) and Fastp v0.23.4 (Chen 2023); (ii) k-mer profiling using Meryl v1.3 (Rhie et al. 2020), GenomeScope v2.0 (Ranallo-Benavidez et al. 2020), and SmudgePlot v0.3.0 (Ranallo-Benavidez et al. 2020); (iii) assembly using Hifiasm v0.19.6 (Cheng et al. 2021); (iv) assembly statistics with GFAstats v1.3.6 (Formenti et al. 2022a); (v) contaminants removal with FCS Adaptor v0.5.0 and FCS-GX v0.5.0 (Astashyn et al. 2024); (vi) and Hi-C mapping scaffolding using BWA-MEM2 v2.2.1 (Vasimuddin et al. 2019), SAMtools v1.19 (Danecek et al. 2021) and YAHS v1.0 (Zhou et al. 2023). Pipeasm was executed in phased mode, with scaffolding considering Hi-C reads.

Genome curation was performed using a combination of an in-house pipeline and manual inspection with PretextView v1.0.1, guided by synteny with *Arabidopsis thaliana* (GCF_000001735.4), *Citrus sinensis* (GCF_022201045.2), and *Mangifera indica* (GCF_011075055.1). The process was performed in three stages: haplotype, diploid, and final curation. Quality's final assembly was checked using Merqury v1.3 (Rhie et al. 2020), GFAstats v1.3.6 (Formenti et al. 2022a), compleasm v.0.2.7 (Huang and Li 2023), and QUAST v.5.3.0 (Mikheenko et al. 2023). We used the TeloExplorer tool from quarTeT v1.2.5 (Lin et al. 2023), with the "-c plant" option, to identify telomeric sequences, specifying "TTTAGGG" as the typical plant repeat motif.

Genome Annotation

We used EarlGrey v6.3.0 (Baril et al. 2024) to estimate repeat content of the jaborandi genome. EarlGrey was run with the Magnoliopsida search term (-r 3398), the Dfam 3.9 database (Storer et al. 2021), running HELIANO (Li et al. 2024) for Helitron detection (-e yes), and creating a soft-masked genome (-d yes).

Genome annotation was performed with NCBI Eukaryotic Genome Annotation Pipeline v0.4.1 (EGAPx) (Thibaud-Nissen et al. 2016), incorporating RNA-Seq data from leaflets, rachis, root, and stem tissues from jaborandi. We used Another Gtf/Gff Analysis Toolkit v1.5.0 (AGAT) (Dainat 2022) to compute annotation stats and to retain the longest isoform per gene. The completeness of our annotation was assessed by comparing it with the Eudicots lineage (odb10) using BUSCO (Manni et al. 2021). Gene functional annotation was performed using eggnog-mapper v2.1.13 (Huerta-Cepas et al. 2019; Cantalapiedra et al. 2021), considering Viridiplantae as taxonomic scope, all types of Gene Ontology evidence, and queries realigned to PFAM domains.

Organellar Genome Assembly

HiFi long-reads were also used to assemble chloroplast and mitochondrial genomes of *P. microphyllus.* The mitogenome was recovered using OrganPipe v1.0 (Moreira-Oliveira et al. 2025), with *Citrus x aurantiifolia* (PX251582.1) as reference. (Zhou et al. 2024). The plastome was assembled using oatk (Organellar Assembly Toolkit) v1.0 (Zhou et al. 2024), with the Angiosperm plastid profile, k-mer size of 1,001 and coverage

threshold of 30. Final organellar genomes assemblies were better annotated with GESeq online (https://chlorobox.mpimp-golm.mpg.de/geseq.html) (Tillich et al. 2017), with the cpDNA (NC_008334.1) and the mtDNA reference of *C. sinensis* (NC_037463.1).

Genetic Diversity and Demography

To infer the evolutionary history and parameters related to the jaborandi's genetic health, we estimated genetic diversity, inbreeding, and population dynamics. We performed quality control and trimming of Illumina reads using fastp v0.23.4 (Chen 2023), and the parameters "–cut_mean_quality 24 –cut_tail –trim_front1 10 –detect_adapter_for_pe –dont_eval_duplication –length_required 50". The mapping and variant calling were performed using NVIDIA Parabricks v.4.3.2 (Zhu et al. 2025). We mapped the reads to the haplotype 1 using the fq2bam module, which integrates the bwa-mem2 mapping and mark duplicates. Variant calling was performed using haplotycaller, which corresponds to the GATK HaplotypeCaller, followed by genotypeGVCFs. The VCF was filtered for quality and depth (MAPQ>20, mimDP>8, maxDP<100), and kept only biallelic SNPs (-m 2 -M 2). Variant-level hard filters were then applied using the following expression: QD < 2.0 || QUAL < 60.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0. We also removed low mappability regions using a mappability mask generated with GenMap (Pockrandt et al. 2020). Our final dataset consisted of 7,048,519 variants.

Genome-wide heterozygosity was estimated using bcftools v1.3.1 (Danecek et al. 2021). We estimated runs of homozygosity (ROHs) as a proxy for inbreeding. Although we have a single individual, this reference genome can provide an initial estimate of potential past and recent inbreeding within the population. We used ROHan (Renaud et al. 2019) which estimates heterozygosity as Waterson's Theta (Θ_W) from mapped reads to the haplotype 1. Parameters were left as default, except for a rohmu (heterozygosity rate) of 5e-04 and windows size of 500 kb. The fraction of the genome in homozygosity (F_{ROH}) was estimated as the total number of ROHs divided by the total genome length.

To reconstruct the past historical demography of *P. microphyllus*, we used the pairwise sequentially Markovian coalescent (PSMC) method (Li and Durbin 2011). We set the parameter -p as "2+2+25*2+4+6 +10", -t as 5, and performed 100 bootstrap replicates. All other parameters were set as default. We used a mutation rate of 4.4×10^{-10} substitutions/site/generation (Perez-Roman et al. 2022) and a generation time of 20 years (Fernandez et al. 2021).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

The authors would like to thank the workers from COEX-Carajás for their support, and the BHCB herbarium for providing the seeds.

Funding

This work was supported by Vale S.A. (R100603.CJ.05).

Data Availability

PacBio HiFi and Hi-C raw data for *Pilocarpus microphyllus* are available on NCBI under BioProject PRJNA1387192, with accession numbers SRR36466372 and SRR36466371, respectively. The RNA-seq transcriptome data are available at BioProject PRJNA1220646. Final chromosome assembly data for both haplotypes are available at BioProject PRJNA1406783 and PRJNA1406782.

Literature Cited

Andreazza NL, Abreu IN, Sawaya ACHF, Mazzafera P. Characterisation of the membrane transport of pilocarpine in cell suspension cultures of *Pilocarpus microphyllus*. J Plant Physiol. 2015:175:37–47. https://doi.org/10.1016/j.jplph.2014.11.002.

Astashyn A, et al. Rapid and sensitive detection of genome contamination at scale with FCS-GX. Genome Biol. 2024:25:60. https://doi.org/10.1186/s13059-024-03198-7.

Baril T, Galbraith J, Hayward A. Earl grey: a fully automated user-friendly transposable element annotation and analysis pipeline. Mol Biol Evol. 2024:41:msae068. https://doi.org/10.1093/molbev/msae068.

Blaxter M, et al. The earth BioGenome project phase II: illuminating the eukaryotic tree of life. Front Sci. 2025:3:fsci.2025.1514835. https://doi.org/10.3389/fsci.2025.1514835.

Brasil. Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA). Portaria n° 37. 1992.

Caldeira CF, et al. Sustainability of Jaborandi in the eastern Brazilian Amazon. Perspect Ecol Conserv. 2017:15:161–171. https://doi.org/10.1016/j.pecon.2017.08.002.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021:38:5825–5829. https://doi.org/10.1093/molbev/msab293.

Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta. 2023:2:e107. https://doi.org/10.1002/imt2.107.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:18:170–175. https://doi.org/10.1038/s41592-020-01056-5.

CNI. Impact study of the adoption and implementation of the Nagoya Protocol on the Brazilian industry. Editorar Multimídia; 2014.

Dainat J. Another Gtf/Gff Analysis Toolkit (AGAT): Resolve interoperability issues and accomplish more with your annotations; 2022. https://github.com/NBISweden/AGAT. [accessed 2025 Dec 17].

Danecek P, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021:10:giab008. https://doi.org/10.1093/gigascience/giab008.

Fernandez E, et al. 2021. *Pilocarpus microphyllus*. International Union for Conservation of Nature (IUCN). The IUCN Red List of Threatened Species. https://doi.org/10.2305/IUCN.UK.2021-1.RLTS.T178713167A178713173.pt.

Formenti G, et al. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. Bioinformatics. 2022a:38:4214–4216. https://doi.org/10.1093/bioinformatics/btac460.

Formenti G, et al. The era of reference genomes in conservation genomics. Trends Ecol Evol. 2022b:37:197–202. https://doi.org/10.1016/j.tree.2021.11.008.

Giraud D, et al. Construction of a comprehensive library of repeated sequences for the annotation of *Citrus* genomes. BMC Genom Data. 2025:26:30. https://doi.org/10.1186/s12863-025-01321-6.

Gornitsky M, et al. Double-blind randomized, placebo-controlled study of pilocarpine to salvage salivary gland function during radiotherapy of patients with head and neck cancer. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2004:98:45–52. https://doi.org/10.1016/j.tripleo.2004.04.009.

Homma AKO, Menezes AJEA. Extrativismo de folhas de jaborandi no município de Parauapebas, estado do Pará. In: Homma AKO, editors. Extrativismo vegetal na Amazônia: história, ecologia, economia e domesticação. Brasília, DF: Embrapa; 2014. p. 119–132 https://www.alice.cnptia.embrapa.br/alice/handle/doc/1016694.

Huang N, Li H. Compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics. 2023:39:btad595. https://doi.org/10.1093/bioinformatics/btad595.

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. 2019:47:D309–D314. https://doi.org/10.1093/nar/gky1085.

Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011:475:493–496. https://doi.org/10.1038/nature10231.

Li Z, Gilbert C, Peng H, Pollet N. Discovery of numerous novel *Helitron* -like elements in eukaryote genomes using HELIANO. Nucleic Acids Res. 2024:52:e79–e79. https://doi.org/10.1093/nar/gkae679.

Lin Y, et al. Quartet: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. Hortic Res. 2023:10:uhad127. https://doi.org/10.1093/hr/uhad127.

Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. Curr Protoc. 2021:1:e323. https://doi.org/10.1002/cpz1.323.

Marques Silva B, et al. Pipeasm: a tool for automated large chromosome-scale genome assembly and evaluation. Bioinforma Adv. 2026:6:vbaf326. https://doi.org/10.1093/bioadv/vbaf326.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011:17:10–12. https://doi.org/10.14806/ej.17.1.200.

McQuillan R, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008:83:359–372. https://doi.org/10.1016/j.ajhg.2008.08.007.

Mikheenko A, Saveliev V, Hirsch P, Gurevich A. WebQUAST: online evaluation of genome assemblies. Nucleic Acids Res. 2023:51:W601–W606. https://doi.org/10.1093/nar/gkad406.

Monteiro WP, et al. Genetic diversity and structure of an endangered medicinal plant species (*Pilocarpus microphyllus*) in eastern Amazon: implications for conservation. Conserv Genet. 2022:23:745–758. https://doi.org/10.1007/s10592-022-01454-6.

Moreira-Oliveira RR, et al. 2025. OrganPipe: An automated tool to facilitate the assembly, annotation, and curation of mitochondrial and chloroplast genomes [preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-5686696/v1.

Perez-Roman E, Borredá C, López-García Usach A, Talon M. Single-nucleotide mosaicism in citrus: estimations of somatic mutation rates and total number of variants. Plant Genome. 2022:15:e20162. https://doi.org/10.1002/tpg2.20162.

Pinheiro CU. Jaborandi (*Pilocarpus* sp., Rutaceae): a wild species and its rapid transformation into a crop. Econ Bot. 1997:51:49–58. https://doi.org/10.1007/BF02910403.

Pirani JR. Lista de Espécies da Flora do Brasil; 2015. http://floradobrasil2015.jbrj.gov.br/jabot/floradobrasil/FB866. [Accessed 2025 Dec 16]

Pirani JR, Devecchi MF. Flora das cangas da Serra dos Carajás, Pará, Brasil: Rutaceae. Rodriguésia. 2018:69:209–217. https://doi.org/10.1590/2175-7860201869119.

Pockrandt C, Alzamel M, Iliopoulos CS, Reinert K. GenMap: ultra-fast computation of genome mappability. Bioinformatics. 2020:36:3687–3692. https://doi.org/10.1093/bioinformatics/btaa222.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020:11:1432. https://doi.org/10.1038/s41467-020-14998-3.

Renaud G, Hanghøj K, Korneliussen TS, Willerslev E, Orlando L. Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. Genetics. 2019:212:587–614. https://doi.org/10.1534/genetics.119.302057.

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020:21:245. https://doi.org/10.1186/s13059-020-02134-9.

Sidhu P. Endangered Jaborandi. Br Dent J. 2014:217:2–3. https://doi.org/10.1038/sj.bdj.2014.555.

Sobreiro MB, et al. Decoding pilocarpine biosynthesis and its roles in *Pilocarpus microphyllus* through a comparative transcriptomics approach. BMC Plant Biol. 2025:25:1024. https://doi.org/10.1186/s12870-025-07087-4.

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob DNA. 2021:12:2. https://doi.org/10.1186/s13100-020-00230-y.

Thibaud-Nissen F, et al. P8008 the NCBI eukaryotic genome annotation pipeline. J Anim Sci. 2016:94:184–184. https://doi.org/10.2527/jas2016.94supplement4184x.

Tillich M, et al. Geseq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017:45:W6–W11. https://doi.org/10.1093/nar/gkx391.

Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for Multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). 20–24 May 2019; Rio de

Janeiro, Brazil; IEEE; 314–324. https://doi.org/10.1109/IPDPS.2019.00041.
Wu GA, et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat Biotechnol. 2014:32:656–662. https://doi.org/10.1038/nbt.2906.
Yang T, et al. Chromosome-level genome assembly of Murraya paniculata sheds light on biosynthesis of floral volatiles. BMC Biol. 2023:21:142. https://doi.org/10.1186/s12915-023-01639-6.
Zhou C, Brown M, Blaxter M, McCarthy SA, Durbin R. 2024. Oatk: a de novo assembly tool for complex plant organelle genomes [preprint]. bioRxiv 10.23.619857. https://doi.org/10.1101/2024.10.23.619857.
Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 2023:39:btac808. https://doi.org/10.1093/bioinformatics/btac808`.
Zhu T, et al. 2025 Aug 04. Parabricks: GPU Accelerated Universal Pan-Instrument Genomics Analysis Software Suite [preprint]. bioRxiv 2025.07.23.666378. https://doi.org/10.1101/2025.07.23.666378

Associate editor: Chuan Ku