



HAL
open science

A genome sequence resource for the genus *Passiflora* , the genome of the wild diploid species *Passiflora* *organensis*

Zirlane Portugal Costa, Alessandro Mello Varani, Luiz Augusto Cauz-santos,
Mariela Analía Sader, Helena Augusto Giopatto, Bruna Zirpoli, Caroline
Callot, Stephane Cauet, Willian Marande, Jessica Luana Souza Cardoso, et al.

► To cite this version:

Zirlane Portugal Costa, Alessandro Mello Varani, Luiz Augusto Cauz-santos, Mariela Analía Sader, Helena Augusto Giopatto, et al.. A genome sequence resource for the genus *Passiflora* , the genome of the wild diploid species *Passiflora organensis*. *Plant Genome*, 2021, 10.1002/tpg2.20117 . hal-03438880

HAL Id: hal-03438880

<https://hal.inrae.fr/hal-03438880v1>

Submitted on 22 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

ORIGINAL RESEARCH

A genome sequence resource for the genus *Passiflora*, the genome of the wild diploid species *Passiflora organensis*

Zirlane Portugal Costa¹  | Alessandro Mello Varani²  | Luiz Augusto Cauz-Santos^{1,7}  |
 Mariela Analía Sader³  | Helena Augusto Giopatto⁴  | Bruna Zirpoli³  |
 Caroline Callot⁵ | Stephane Cauet⁵  | Willian Marande⁵ | Jessica Luana Souza
 Cardoso¹  | Daniel Guariz Pinheiro²  | João Paulo Kitajima⁶ |
 Marcelo Carnier Dornelas⁴  | Andrea Pedrosa Harand³  | Helene Berges⁵  |
 Claudia Barros Monteiro-Vitorello¹  | Maria Lucia Carneiro Vieira¹ 

¹ Dep. de Genética, Escola Superior de Agricultura “Luiz de Queiroz”, Univ. de São Paulo, Piracicaba 13418–900, Brazil

² Dep. de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, Univ. Estadual Paulista, Jaboticabal 14884–900, Brazil

³ Dep. de Botânica, Univ. Federal de Pernambuco, Recife 50670–901, Brazil

⁴ Dep. de Biologia Vegetal, Instituto de Biologia, Univ. Estadual de Campinas, Campinas 13083–862, Brazil

⁵ Institut National de la Recherche Agronomique, Centre National de Ressources Génomique Végétales, Castanet-Tolosan 31326, France

⁶ Mendelics Análise Genômica AS, São Paulo 04013–000, Brazil

⁷ Present address: Dep. of Botany and Biodiversity Research, Univ. of Vienna, Vienna 1030, Austria

Correspondence

Maria Lucia Carneiro Vieira, Departamento de Genética, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, 13418–900, Piracicaba, Brazil.

Email: mlcvieir@usp.br

Assigned to Associate Editor Laura Rossini.

Abstract

The genus *Passiflora* comprises a large group of plants popularly known as passion-fruit, much appreciated for their exotic flowers and edible fruits. The species (~500) are morphologically variable (e.g., growth habit, size, and color of flowers) and are adapted to distinct tropical ecosystems. In this study, we generated the genome of the wild diploid species *Passiflora organensis* Gardner by adopting a hybrid assembly approach. *Passiflora organensis* has a small genome of 259 Mbp and a heterozygosity rate of 81%, consistent with its reproductive system. Most of the genome sequences could be integrated into its chromosomes with cytogenomic markers (satellite DNA) as references. The repeated sequences accounted for 58.55% of the total DNA analyzed, and the *Tekay* lineage was the prevalent retrotransposon. In total, 25,327 coding genes were predicted. *Passiflora organensis* retains 5,609 singletons and 15,671 gene families. We focused on the genes potentially involved in the locus determining

Abbreviations: ANI, average nucleotide identity; CDS, coding regions; FISH, fluorescence in situ hybridization; GO, Gene Ontology; K, synonymous nucleotide substitution rate; LTR, long terminal repeat; LTR-RT, long terminal repeat–retrotransposon; MIKC, MADS intervening keratin-like and C-terminal; mya, million years ago; NADH, nicotinamide adenine dinucleotide hydrogen; SI, self-incompatibility; SLG, S-locus glycoproteins; SRK, S-locus receptor kinases; TE, transposable element; WGD, whole-genome duplication.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *The Plant Genome* published by Wiley Periodicals LLC on behalf of Crop Science Society of America

self-incompatibility and the MADS-box gene family, allowing us to infer expansions and contractions within specific subfamilies. Finally, we recovered the organellar DNA. Structural rearrangements and two mitoviruses, besides relics of other mobile elements, were found in the chloroplast and mt-DNA molecules, respectively. This study presents the first draft genome assembly of a wild *Passiflora* species, providing a valuable sequence resource for genomic and evolutionary studies on the genus, and support for breeding cropped passionfruit species.

1 | INTRODUCTION

Passiflora (Passifloraceae, Malpighiales) is a neotropical genus encompassing hundreds of species with widespread distribution in the American continent, including the Amazonian and Andean regions. Population pressure in all these regions is high, raising considerable concern over pollinator decline (e.g., bees, bats, and hummingbirds) and conservation of *Passiflora* diversity. There are four main subgenera: *Astropheia* (57 species), *Decaloba* (220 species), *Deidamioides* (13 species), and the *Passiflora* subgenus (240 species), which contains several self-incompatible species (Ulmer & MacDougal, 2004). Although some 50 species are traditionally cropped in temperate regions (e.g., the United States), most are grown in tropical climates to produce fruit (Cerqueira-Silva et al., 2015; Zerbini et al., 2008).

Passiflora species morphology is highly variable, with a 10-fold variation in genome size (1C DNA content = 0.212 pg in *Passiflora organensis* Gardner up to 2.68 pg in *Passiflora quadrangularis* L. (Souza et al., 2004; Yotoko et al., 2011) that is not attributable to polyploidy. Although it is the species with the smallest genome species, *P. organensis* (subgenus *Decaloba*) seems to be evolving by diversification of different repeat types; the largest, *P. quadrangularis* (subgenus *Passiflora*), evolved by accumulating retrotransposons, especially *Angela* and *Tekay* elements, which comprise most of its genome (Sader et al., 2021).

The publication of several draft plant genomes has undoubtedly provided valuable resources for the scientific communities conducting top-level research in their respective fields. For instance, advanced sequencing technologies have resulted in near-complete, high-quality chromosome-scale genome assemblies at minimal cost (Pham et al., 2020). However, we are only just beginning to investigate the genomic information on underutilized crops (e.g., Gioppato et al., 2019).

Against this backdrop, a genomic library was built covering around six times the genome length of the sour passionfruit (*Passiflora edulis* Sims) (Santos et al., 2014), whose nuclear genome size was estimated at 1.258 pg (expressed as 1C) or 1,232 Mbp (Yotoko et al., 2011). On the basis of

gene richness, we selected and completely sequenced over 100 inserts. Sequence data were assembled from long sequence reads, and structural sequence annotation resulted in the prediction of some 1,900 genes (Munhoz et al., 2018). The dataset was also subjected to transposable element (TE) discovery and characterization of long terminal repeat (LTR)–retrotransposon (LTR-RT) evolutionary lineages. Most of the TEs were located in intergenic spaces (~70%), although some overlapped genes. Long terminal repeat retroelements predominated, consisting mainly of *Gypsy* elements, with overrepresentation of the *RLG_peDel* (or *Tekay* lineage according to Neumann et al. [2019]), and single elements of *P. edulis* were obtained for the first time (Costa et al., 2019). Additionally, Sader, Dias, et al. (2019), who used a *Gypsy* element sequence as a probe for in situ hybridization, found a dispersed and uniform distribution pattern along *P. edulis* chromosomes, implying that they were abundant and could have significantly influenced the genome size. Other information on the nuclear genome of *P. edulis* was derived from low-coverage sequencing data, facilitating microsatellite marker development (Araya et al., 2017) and providing cytogenomic markers, especially satellite sequences that are possibly associated with 5S and 35S rDNA or subtelomeres (Pamponét et al., 2019).

Very recently, a ~1,341.7-Mbp chromosome-scale genome assembly of the purple passionfruit (*P. edulis*) was reported (Xia et al., 2021), with 98.91% of the assembly assigned to nine pseudochromosomes. The genome harbors 23,171 protein-coding genes, and most of the assembled sequences are repetitive, with a predominance of LTRs, confirming our previous findings. Moreover, important gene families were identified by integrating transcriptomic and metabolomic analyses, centered on genes involved in the synthesis of volatile organic compounds and providing insights for improving the flavor of fresh fruit.

This interesting scenario inspired us to sequence the whole genome of the wild diploid species *P. organensis*, which has the smallest known *Passiflora* genome, but whose overall morphology is very representative of the genus (Figure 1). A draft genome assembly was generated on the basis of a combination of the genomic and transcriptomic data associated

with cytogenomic markers. The availability of a reference sequence will greatly facilitate further progress in comparative genomics and molecular genetics, geared towards assisting passion fruit breeding initiatives. It will also serve as a valuable tool for evolutionary analyses of *Passiflora*, providing a scaffold for genome assembly, variant calling, RNA read alignment, chromosome molecular mapping, gene annotation, and functional analysis, especially for comparison with the purple passionfruit genome (Xia et al., 2021).

Passiflora organensis and the cultivated species *P. edulis* belong to the same genus but different subgenera (*Decaloba* and *Passiflora*, respectively). As expected, they share some common physiological and phenological characteristics (both are perennial lianas and flower in response to longer days), but diverge in others. *Passiflora organensis* produces small fruits with almost no juice, whereas *P. edulis* produces large fruits with juice-rich arillate seeds. In this respect, the availability of both genomes will further our understanding of important domesticated fruit traits and the underlying mechanisms.

Finally, it is well known that some *Passiflora* species have a self-incompatibility system controlled by sporophytic and gametophytic mechanisms, and/or have coevolved with butterflies of the subfamily Heliconiinae. The occurrence of structures mimicking heliconiine eggs, the production of extrafloral nectar to attract the predators of *Passiflora* herbivores, and the wide diversity of defense compounds, including cyanogenic glucosides [reviewed in Krist (2020)], are strong arguments in favor of generating an additional reference genome to facilitate functional studies.

2 | MATERIAL AND METHODS

A single *P. organensis* plant was used for DNA isolation. It was field-collected in the Biological Reserve of Alto da Serra de Paranapiacaba (Santo André, São Paulo; 23.77309 S, 46.3017 W). This municipality is located in the Atlantic forest biome of Southeastern Brazil. The plant was kept in a greenhouse at our laboratory (Piracicaba, São Paulo; 22.71027 S, 47.63286 W). Its accession number (AF82F07) is registered in the Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado, Brazil.

2.1 | Cytogenomics

Root tips obtained from potted plants were pretreated with 2 mM 8-hydroxyquinoline for 4.5 h at 10 °C, fixed in ethanol–acetic acid (3:1 v/v), and stored in a fixative at –20 °C. The root tips were then digested in a solution containing 2% cellulase and 20% pectinase (w/v) for 90 min at 37 °C, and chromosome preparations were performed according to Carvalho and Saraiva (1993). Slides were selected after staining with 2 µg

Core Ideas

- *Passiflora organensis* has a small genome of 259 Mbp.
- We estimate that the repetitive content of the *P. organensis* genome is approximately 59%.
- We predicted 25,327 protein-coding genes and genes involved in self-incompatibility.
- Organelle DNAs were recovered; two mitoviruses and relics of mobile elements occurred in mtDNA.
- This study provides a promising resource for comparative genomic analysis of *Passiflora*.

ml⁻¹ 4',6-diamidino-2-phenylindole (Sigma) in 50% glycerol (Cabral et al., 2006).

Four satDNA probes were used for fluorescent in situ hybridization (FISH) experiments (Sader et al., 2021) (PorSat01-161, PorSat04-1800, PorSat07-1004, and PorSat09-1200) and labeled with 5-amino-propargyl-2'-deoxyuridine 5'-triphosphate and cyanine red fluorescent dye (GE Healthcare). The FISH procedure applied to mitotic chromosomes was essentially as described in Fonsêca et al. (2010). The hybridization mix consisted of 50% (v/v) formamide, 10% (w/v) dextran sulfate, 2× SSC, and 2 to 5 ng µl⁻¹ of the probes. The slides were denatured for 5 min at 75 °C and hybridized for 24 h at 37 °C. The final stringency was 76%. The slides were mounted with Vectashield (Vector) containing 2 µg ml⁻¹ 4',6-diamidino-2-phenylindole. Images were captured on an epifluorescence microscope (Leica Microsystems) running Leica QFISH software and a DMLB microscope running Leica Las-AF software. For final processing, the images were uniformly adjusted for brightness and contrast only, using Adobe Photoshop version 10.0. Three complete metaphase cells were used to measure the short and long arms of the individual chromosomes identified by satDNA probes with the Adobe Photoshop measurement tool. Chromosomes were identified and classified according to their length and arm ratio, as described in Guerra (1986).

2.2 | DNA isolation, library construction, and sequencing of *P. organensis* DNA

Genomic DNA was extracted from young leaves to obtain high-quality DNA for long- and short-read sequencing. For PacBio long-read sequencing, high molecular weight DNA was extracted via a modified version of the cetyltrimethylammonium bromide method adapted from Murray and Thompson (1980) and Carlson's lysis buffer (Carlson et al., 1991), followed by DNA purification via the Qiagen Genomic



FIGURE 1 Macromorphology of *Passiflora organensis*. The habit is shown in the upper panel. The middle images show the top (left) and lateral views of a longitudinal section (right) of a flower at anthesis. The lower panels show a leaf from the abaxial (left) and adaxial views (middle), and a fruit from the lateral view (right). Bars: 1.5 cm

Tip 500/G protocol (Qiagen, Germany) following the manufacturer's instructions. DNA quality and quantity were checked in Nanodrop and Qubit (Thermo Fisher Scientific), and DNA fragment sizes were analyzed by pulsed-field electrophoresis and in a Fragment Analyzer (Agilent Technologies).

The PacBio library was constructed using 5 g of DNA, and final sizing was performed to select fragments ranging from 20 to 150 kb. The DNA was then sequenced in eight runs in a single-molecule real-time cell and via P6-C4 chemistry on the PacBio RSII (Pacific Biosciences) at the Uppsala NGI Platform (Uppsala University, Sweden). Finally, the raw

data were filtered to obtain high-quality reads (reads with a quality of <0.75 and a length of <500 bp were discarded). For Illumina short-read sequencing, DNA was extracted by the cetyltrimethylammonium bromide method adapted from Murray and Thompson (1980) and used to construct three PCR-free libraries with the Illumina TruSeq kit. Sequencing was run on an Illumina HiSeq platform with the Rapid Run Version 2 (Illumina) protocol to yield 250-bp paired end reads. The Illumina raw reads were trimmed with Trimmomatic software (Bolger et al., 2014), discarding all reads with a PHRED value (a measure for base quality in DNA sequencing) lower than 24.

We also performed one run of Nanopore sequencing to complete the assembly of the mitochondrial genome. MinION libraries were prepared with the Oxford Nanopore 1D Genomic DNA Ligation kit (SQKLSK109) (Oxford Nanopore Technologies) and MinION protocol (version IDI_S1006_v1_revB_18Apr2016), followed by sequencing on a Nanopore MinION platform (MKE_1013_v1_revAR_11Apr2016).

2.3 | Optical map construction

For BioNano optical genome mapping, ultra-high molecular weight DNA was extracted at the Centre National de Ressources Génomiques Végétales (Toulouse, France), according to the Bionano IrysPrep Reagent Kit protocol (Bionano Genomics). Briefly, fresh leaves (~0.5 g) were used for nucleus isolation and gradient density purification. High molecular weight DNA was then extracted from an agarose plug containing the nuclei isolated and subjected to Proteinase K digestion, followed by RNA digestion. Finally, the DNA was recovered, dialyzed, and then quantified in Qubit (Thermo Fisher Scientific). The DNA was loaded into an Irys chip (Bionano Genomics) and optically mapped in a BioNano Genomics Irys system according to the manufacturer's instructions.

2.4 | RNA extraction for *P. organensis* transcriptome sequencing

Tissues from young leaves, stems, and floral buds were collected and immediately placed in liquid nitrogen. RNA was extracted via the Trizol protocol (Invitrogen) according to the manufacturer's instructions. The quality of the RNA was assessed on the basis of the RNA integrity numbers, computed on a Bioanalyzer 2100 (Agilent Technologies) and also by 1.2% agarose gel electrophoresis. This RNA provided a basis for constructing three libraries, one for each tissue (leaf, stem, and bud), with the Illumina TruSeq Stranded mRNA kit. Libraries were pooled and sequenced at 2× 100 bp in a half lane on the Illumina HiSeq2500 platform at the Center for Functional Genomics of the University of São Paulo, Piracicaba, Brazil.

2.5 | Genome assembly

The genome was assembled in five steps (Supplemental Figure S1). In Step 1, FALCON-Unzip Version 2.2.4 (Chin et al., 2016) was used to assemble the reads of eight PacBio Sequel single-molecule real-time cells (2,502,979 reads). In Step 2, SPACE-LongRead Version 1.1 (Boetzer & Pirovano,

2014) corrected the PacBio reads (489,056) and expanded the assembled scaffolds. In Step 3, the trimmed 2× 250-bp Illumina reads (89,918,970) were used to polish and correct the genome assembly with POLCA Version 3.4.2 (Zimin & Salzberg, 2020). In Step 4, RefAligner (Bionano Genomics) and runBNG Version 2.0 (Yuan et al., 2017) assembled the BioNano optical maps with the previously polished assembly reference. In the final step (Step 5), runBNG used the assembled optical map and polished the assembly for input to the Hybrid Scaffold pipeline (Bionano Genomics). In addition, taking *Populus trichocarpa* Torr. & A.Gray ex Hook. and *Manihot esculenta* Crantz. as references (Supplemental Table S1). RaGOO (Alonge et al., 2019) was used for performing additional reference-guided scaffolding steps (parameters: a minimum unique alignment length of 50,000 bp and a maximum inferred gap size of 10,000 bp). Assembly gaps were closed with GapFiller Version 2.1.2 (Nadalin et al., 2012) and LR_GapCloser Version 1.1 (Xu et al., 2018), for the trimmed Illumina and PacBio reads, respectively. Benchmarking Universal Single-Copy Orthologs Version 4.0.2 (Seppey et al., 2019) and the Embryophyta odb10 database were used to assess genome completeness. Genome assembly quality was also assessed on the basis of the LTR Assembly Index, computed according to Ou et al. (2018).

Repetitive sequences were investigated via the dot-blot method to visualize two sequences and identify closely similar regions after sequence alignment. Default parameters were used for pairwise alignment of the scaffolds run on MUMmer4 (Marçais et al., 2018). This approach allowed us to map the centromeric regions.

GenomeScope Version 1.0.0 (Vurture et al., 2017) and Smudgeplot Version 0.2.4 (Ranallo-Benavidez et al., 2020) were used to estimate the coverage, ploidy, and heterozygosity of the *P. organensis* genome. The diploid model was used to fit the results and default parameters for analyzing data on 89,918,970 Illumina reads.

2.6 | Transcriptome assembly

P. organensis and *P. edulis* transcriptomes generated by the RNAseq approach were assembled with the Trinity tool Version 2.11.0 (Grabherr et al., 2011) with the default parameters. Coding regions within transcripts were identified by the TransDecoder pipeline Version 5.5.0 (<https://github.com/TransDecoder/TransDecoder>).

2.7 | Transposable element discovery

De novo detection of TEs was run in REPET Version 3.0 (Jouffroy et al., 2016; Quesneville et al., 2005), and the

bundled PASTEC tool Version 3.0 (Hoede et al., 2014) used to classify TEs at the order level on the basis of the Wicker hierarchical system (Wicker et al., 2007). The elements identified by REPET were manually curated and characterized into superfamilies and evolutionary lineages with DANTE (Domain-based ANnotation of Transposable Elements) Version 1.0.1, available on the RepeatExplorer server (Novák et al., 2013; <https://repeatexplorer-elixir.cerit-sc.cz/>), and the Viridiplantae database Version 3.0 (Neumann et al., 2019). Transposable element masking was performed by RepeatMasker Version 4.0.7 (<http://www.repeatmasker.org>) with the “-s -cutoff 260” parameter. The TEannot pipeline in the REPET package was also used to generate TE annotations in GFF3 file format.

To estimate when the LTR-RT lineages were inserted into the *P. organensis* genome, sequences were used as input to LTR_FINDER Version 1.0.7 (Xu & Wang, 2007) to search for full-length elements and compute the divergence (K) between LTRs. Insertion time (t) was then calculated for each full-length element according to the formula: $t = K/2r$ (Yin et al., 2015), where t is the insertion time in million years ago (mya), K is the number of nucleotide substitutions per site, and r is the nucleotide substitution rate. A value of 1.5×10^{-8} was assigned to r , as reported for the *Arabidopsis thaliana* (L.) Heyhn. chalcone synthase and alcohol dehydrogenase genes (Koch et al., 2000) used for dating the LTR-RTs in *P. edulis* (Costa et al., 2019). In order to investigate the incorporation of entire genes or fragments thereof into TEs, the sequences were scanned for conserved domains in the Conserved Domain Database Version 3.18 (Marchler-Bauer et al., 2017) public resource using the Conserved Domain Database search interface (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), with default parameters.

2.8 | Genome annotation

For gene prediction, the BRAKER Version 2 pipeline (Hoff et al., 2016) was used for *P. organensis*, inputting the transcriptome and protein data of very closely related species derived from the Benchmarking Universal Single-Copy Orthologs searches and OrthoDB (Kriventseva et al., 2019) (–etp mode and –softmasking modes). The PASA pipeline Version 2.4.1 (Haas et al., 2003) produced spliced alignment assemblies based on the RNA-seq data. Evidence Modeler Version 1.1.1 (Haas et al., 2008) combined with BRAKER2 Version 2.1.5 outputs (augustus.ab.initio, augustus.hints, and braker), and the transcript alignments were then used. Finally, two PASA pipeline iterations updated the Evidence Modeler consensus predictions, adding untranslated region annotations and models for the alternatively spliced isoforms. Functional annotation was performed with OmicsBox Version 1.4.10

(<https://www.biobam.com/omicsbox/>), InterProScan Version 5.46-81.0 (Jones et al., 2014), and EggNOG-mapper Version 1.0.3 (Huerta-Cepas et al., 2017) with the EggNOG database Version 5.0 (Huerta-Cepas et al., 2019). The UniProtKB/TrEMBL (The UniProt Consortium, 2017) Viridiplantae database was used for protein function assignment.

Gene ontology (GO) term enrichment analysis across different sets of genes was run with the MCScanX algorithm (Wang et al., 2012), which classifies genes as singletons and dispersed, proximal, tandem, and segmental or whole-genome duplication (WGD) duplicates according to the copy number and genomic distribution. The following criteria were applied: (a) singletons were single-copy genes, (b) adjacent genes were tandem, (c) genes separated by 10 or fewer genes were proximal, (d) genes separated by more than 10 genes were dispersed, and (e) the anchor genes in collinear blocks were WGD genes. We then identified over-represented GO terms in these gene sets with GOATOOLS software (Klopfenstein et al., 2018), with a corrected p value of $<.01$ as the threshold for determining significant overrepresentation.

In addition, orthogroups were defined to assess common protein-coding genes, and all gene sequences of Malpighiales species (*P. organensis*, *P. trichocarpa*, *Salix purpurea* L., *Ricinus communis* L., and *M. esculenta*) were used as input to OrthoVenn2 (Xu et al., 2019), set to the default parameters. To identify the genes potentially involved in self-incompatibility, BlastP searches using Brassicaceae reference sequences of the S-locus and partial *P. edulis* sequences (Madureira et al., 2014) were run to recognize homologs in the predicted *P. organensis* protein dataset. Sequences were identified on the basis of similarity and a domain structure resembling S-locus glycoproteins (SLG) and S-locus receptor kinases (SRK). The linear arrangements of domains (protein architecture) corroborated Xing et al. (2013) and were identified in the NCBI Conserved Domain Database (Lu et al., 2020). Variant domain architectures were also noted.

2.9 | Annotation of noncoding RNA

All noncoding RNA classes were scanned with a query set from the Rfam database and cmsearch software in Infernal toolkit Version 1.1.2 (Nawrocki & Eddy, 2013). The following tools were used for class-specific annotation of ncRNAs: (a) tRNAscan-SE Version 2.0 (Lowe & Chan, 2016; Lowe & Eddy, 1997) with default parameters to identify tRNA genes, (b) the RNAmmer Version 1.2 (Lagesen et al., 2007) algorithm with the default parameters to identify rRNA genes (5.8S, 18S, and 28S), (c) nucleotide BLAST (Altschul et al., 1990) against miRBase release 22.1 (Kozomara & Griffiths-Jones, 2014) with dust parameters and an e-value of 0.00001. Filtering was based on identity and a query coverage of at least 90%.

2.10 | Whole-genome duplication analysis

Whole-genome duplication events were predicted according to the distribution of synonymous substitutions per synonymous site (K) among paralogs. First, an all-by-all BlastP was run with the default parameters to scan all potential homologous gene pairs within any genome (*P. organensis*, *P. trichocarpa*, *S. purpurea*, *M. esculenta*, and *R. communis*; see Supplemental Table S1). The DupGen_finder pipeline (Qiao et al., 2019) was then run to identify WGD-derived gene pairs on the basis of the MCScanX algorithm (Wang et al., 2012) bundled with the pipeline, and to calculate the K (Qiao et al., 2019) (see https://github.com/qiao-xin/Scripts_for_GB).

To assess the genome similarities among related Malpighiales species we used the entirely sequenced and well-annotated genomes of *P. trichocarpa*, *S. purpurea* (both Salicaceae), and *M. esculenta* and *R. communis* (both Euphorbiaceae) (Supplemental Table S1). MCScan was run with the default parameters according to the instructions [[https://github.com/tanghaibao/jcvi/wiki/MCscan-\(Python-version\)](https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version))]. Searching each genome for homologous groups of protein-coding genes was based on the best five hits with the best nonself match in each target genome that met an e-value threshold of 0.00001.

For constructing the matrix of average nucleotide identity (ANI) across coding regions, information was extracted from the BLAST output using Custom Perl scripts and the Malpighiales genomes available in Phytozome database Version 13 (Goodstein et al., 2012) (*P. trichocarpa*, *Populus deltoides*, *S. purpurea*, *M. esculenta*, *R. communis*, and *Linum usitatissimum* L.). *Arabidopsis thaliana* was used as the outgroup (Supplemental Table S1).

2.11 | MADS-box genes in *P. organensis*

By using the gene prediction pipeline combined with protein predictions output by Augustus Version 3.2.3 (Keller et al., 2011), we isolated the candidate *P. organensis* homologs in MADS-box genes with the BlastP tool (Altschul et al., 1997) and *A. thaliana* sequences as baits. The MADS-box sequences from *A. thaliana*, *P. trichocarpa*, and *Vitis vinifera* L. were obtained from the PlantTFDB database (Jin et al., 2017). Amino acid sequences were used for aligning the sequences on the basis of the L-INS-i algorithm in MAFFT Version 7 (Kato & Standley, 2013). The alignments were recurrently inspected, then misaligned or nonaligned *P. organensis* sequences were examined further and either discarded or manually curated with Bioedit (Hall, 1999). The final alignment was fitted to the model with ModelFinder (Kalyaanamoorthy et al., 2017) implemented in IQ-TREE Version 1.5.4 (Nguyen et al., 2015). The best fitting model, according to

the Akaike criterion, was the Jones–Taylor–Thornton substitution model with empirical base frequencies and the discrete gamma model with four categories, which was used for phylogenetic reconstruction by maximum likelihood with IQ-TREE Version 1.5.4. and 1,000 replicates in the bootstrap test.

2.12 | Chloroplast and mitochondrial DNA assembly and annotation

GeSeq (Annotation of Organellar Genomes) (Tillich et al., 2017) with the default settings was used to predict *P. organensis* genes and identify sequence features, including chloroplast DNA protein-coding gene sequences, rRNA, and tRNA, followed by manual correction for start and stop codons and intron positions in GenomeView software. All tRNA genes were further confirmed in tRNAscan-SE (Lowe & Chan, 2016) and the ARAGORN online search server (Laslett & Canback, 2004). Pseudogenes were classified on the basis of the nucleotide losses or the presence of internal stop codons. Finally, the circular chloroplast genome map was constructed in OGDRAW (Greiner et al., 2019).

PacBio long reads were assembled in Canu Version 2.1.1 (Koren et al., 2017) with the default parameters. In total, 1,087,641 genomic reads (nuclear and organelle) were used in the assembly. Five mitochondrial contigs (built from 12,031 reads, i.e., 1.1% of the total genomic reads processed) were manually identified on the basis of the annotation and coverage similarity. Illumina short reads were used to polish the Canu contig mitochondrial sequences by read alignment and consensus extraction. Qiagen CLC Genome Workbench Version 8.5.1 was used to align the short reads against the first raw assembly. Cross_match (Machado et al., 2011) was used to close gaps in the scaffolds manually, generating two molecules, one circular (102,307 bp) and one linear (1,031,229 bp), with an average coverage of approximately 1200-fold of the Illumina reads, 50-fold of the Nanopore reads and 280-fold of the PacBio reads.

For mitogenome annotation, the *P. edulis* genome (MT140634) (Yang & Wang, 2020) was used as the reference, following the protocol described for chloroplast DNA annotation. Gene sequences, intron–exon boundaries, and *trans*-spliced events were manually curated by comparisons with annotated orthologs available in the plant mitochondrial genome database at NCBI (www.ncbi.nlm.nih.gov/genome/organelle). All other genetic elements were detected by manual BlastX and nucleotide BLAST searches against GeneBank, identifying the repeats in Blast2seq. Maps were constructed with the Circos Version 0.69-9 visualization tool (Krzywinski et al., 2009) and manually adjusted with Inkscape (<https://inkscape.org/>).

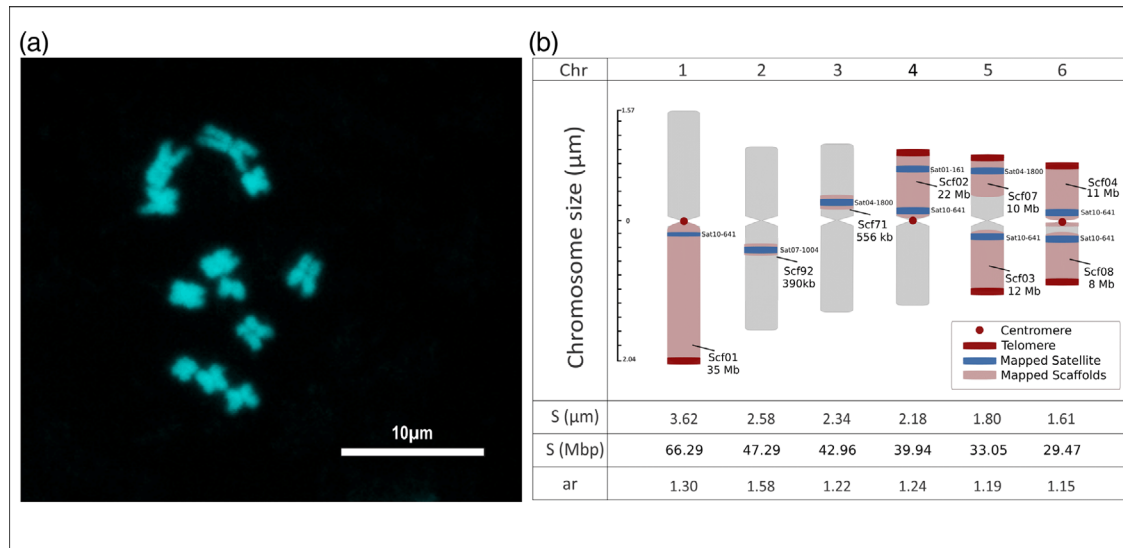


FIGURE 2 Partial integration of the genome sequences of *Passiflora organensis* into its chromosomes. (a) Mitotic metaphase showing $2n = 12$. (b) Ideogram showing the estimated size (S) of each chromosome, the arm ratios (ar), and the anchoring of the assembled pseudochromosomes and scaffolds (Sc) on each chromosome with the satDNAs indicated

3 | RESULTS

3.1 | The *P. organensis* karyotype

Karyotype analysis of *P. organensis* revealed five pairs of metacentric chromosomes and one—the second largest—of submetacentric chromosomes (Figure 2a), ranging from approximately 3.62 to 1.61 μm in size (Figure 2b). We used satellite DNA sequences (Sader et al., 2021) as FISH probes for unambiguous identification of chromosome pairs and to facilitate contig and scaffold anchoring to chromosomes. As a result, satellite DNA PorSat07-1004 was mapped in the long-arm proximal region of chromosome 2, PorSat01-161 in a large short-arm heterochromatic block of chromosome 4, and PorSat04-1800 was found in the short-arm proximal region of chromosome 3 and within the short-arm distal region of chromosome 5 (Figure 2b). Furthermore, when PorSat10-641 was used as a probe, pericentromeric signals were observed in all chromosome pairs, but with lower abundance in chromosome 1.

3.2 | Assembly and annotation of the *P. organensis* genome

We generated a total of 92 Gbp (PacBio and three Illumina sequencing libraries, Supplemental Table S2) to assemble a 259-Mbp genome sequence in 482 contigs on 360 scaffolds with a guanine–cytosine content of 38.3% (Table 1, Supplemental Table S3). Genome sequence coverage was 160×. BioNano data were used to improve the assembly (Supplemental Table S4 and Supplemental Table S5). At least one

TABLE 1 Statistics for the *Passiflora organensis* genome, including some comparisons with those of *P. edulis* (Xia et al., 2021)

Type	Value	<i>P. edulis</i> genome
Assembled genome size (bp)	259,301,974	1,327,182,440
Number of scaffolds	360	9
Longest scaffold	35,942,160 bp	204.53 Mb
Shortest scaffold	4,400 bp	112,417,078 bp
Number of scaffolds > Mb	38	9
Number of scaffolds > 10 Mb	7	9
Scaffold N50 length ^a	8,258,683 bp	140.18 Mb
Scaffold L50 count ^a	9	5
Contig N50 length	2,458,705 bp	3.1 Mb
Guanine–cytosine content (%)	38.3	38.68
Number of protein-coding genes	25,327	23,171
Average gene length	3,200 bp	–
Mean length of exons per gene	304 bp	–
Repeat content (%)	58.55	23.61

^aN50, the sequence length of the shortest contig at 50% of the total genome length; L50, the sequence length of the shortest scaffold at 50% of the total genome length.

end of six scaffolds (01, 02, 03, 04, 07, and 08) exhibited long telomeric repeats (AAACCCT)_n (Figure 2b). By using dot-plot approaches, we identified the centromeric and/or pericentromeric regions (which are difficult to tell apart) for two scaffolds (01 and 02), mapping the complete assembled sequences to the long and short arms of chromosomes 1 and 4, respectively. Three scaffolds (03, 06, and 10) had centromeric-like repeats and another two (05 and the nonmapped scaffold 21) had interstitial telomeric repeats (Supplemental Figure S2, Supplemental Table S6).

The karyotype information, together with mapping of the satellite sequences by FISH and bioinformatic approaches, was used to integrate the scaffolds into *P. organensis* chromosomes. Eight of them have been mapped (01, 02, 03, 04, 07, 08, 71, and 92) (Figure 2; Supplemental Table S6), corresponding to 102.3 Mbp or ~39% of the whole genome. K-mer frequency analysis supported the *P. organensis* genome's diploid state, showing a heterozygosity rate of 81% (Supplemental Figure S3 and Supplemental Figure S4).

Note that we evaluated the possibility of using the available *P. edulis* genome sequence (Xia et al., 2021) to assist in assembling the *P. organensis* genome. However, because of structural chromosomal differences (number and size), it was not possible to anchor the *P. organensis* scaffolds onto fully assembled *P. edulis* chromosomes to reconstruct potential pseudomolecules.

Genome completeness assessment with Benchmarking Universal Single-Copy Orthologs resulted in 1,587 (98.4%) of the expected 1,614 conserved plant orthologs (Supplemental Table S7). The LTR assembly index, a metric that evaluates genome continuity (Ou et al., 2018), was 12.19.

The combination of transcriptomic alignments and ab initio gene predictions allowed us to identify 25,327 genes (31% of the genome). These genes were identified on 233 scaffolds comprising 252 Mbp. The remaining 127 scaffolds 7 Mbp, with the sequence length of the shortest contig at 50% of the total genome length being 101 kbp) consisted of TE sequences. Approximately 21,000 genes were annotated for protein function, gene ontology assignment and InterPro domain. Annotation of noncoding and structural RNA genes revealed 492 tRNAs, at least 122 rRNAs (110 5.8S and six each of 18S and 28S rRNAs) and 83 miRNA genes.

We investigated the presence of genes potentially involved in the self-incompatibility (SI) determining locus. *Brassica* sequences were used as a reference because of their well-known system and their use in a previous investigation of *P. edulis* (Madureira et al., 2014). In total, 54 sequences were recovered from the *P. organensis* dataset, similar to SRK and SLG (Takayama & Isogai, 2005) (Supplemental Table S8). They included 14 SRKs (B_lectin-SLG-PAN_APPLE-TM-KD) and four SLGs (B_lectin-SLG-PAN_APPLE) with a typical domain architecture (Xing et al., 2013). All the others exhibited variant domain structures, which were assumed to correspond to precursors (degenerated products or functions still unknown) (Xing et al., 2013). Note the genome organization laid on Scaffold 6, which harbors a cluster of 10 genes (133,003 bp), including one SRK (scaffold6_gene573) and one SGL (scaffold6_gene572), as detected in *Brassica* (Takayama & Isogai, 2005). The role of these genes in the reproductive system of *P. organensis* requires further investigation.

3.3 | Transposable element discovery

In total, 58.55% of the whole *P. organensis* genome consisted of repetitive elements (Figure 3a; Supplemental Table S9), most belonging to the LTR order (33.81%). Our analysis suggests a massive expansion of the LTR-RT *Tekay* evolutionary lineage (*Gypsy* superfamily), representing up to 54 Mbp of the genome, accompanied by an expansion of long interspersed nuclear elements spanning ~14 Mb. A small fraction (2.79%) of the *P. organensis* genome harbored DNA transposons, and unclassified TEs accounted for up to 15.82%.

Insertion time analysis of the *Copia* and *Gypsy* superfamilies indicated that nearly 50% of the TEs had been inserted into the *P. organensis* genome within the past 0.5 million years, whereas others had appeared between 0.6 and 5.2 mya. Moreover, the insertion time of the *Tekay* lineage, the most predominant of all the TEs, suggests the recent activity of these elements, since the majority appeared from 0.1 to 0.5 mya (Figure 3b).

Some copies of nonautonomous TEs were also identified, including large retrotransposon derivatives and terminal repeat retrotransposons in miniature from the LTR order, and miniature inverted repeat transposable elements derived from DNA transposons. Some are known to lack domains and also to interact with host genes, incorporating fragments or even entire sequences. In fact, nonautonomous *P. organensis* TEs harboring gene sequences were found. Interestingly, all DNA transposons in the Helitron order appeared to incorporate gene fragments, the majority in domains related to resistance genes (Supplemental Table S10).

3.4 | Whole-genome duplication analysis

The distribution of pairwise synonymous substitution rates across the corresponding duplicated genes (paralogs) in conserved genomic blocks was studied in *P. organensis* and the Malpighiales *P. trichocarpa*, *S. purpurea* (both Salicaceae), and *M. esculenta* (Euphorbiaceae). These are related species with entirely sequenced and well-annotated genomes (Supplemental Table S1). Synonymous nucleotide substitution rate-based analysis (Cai et al., 2019) identified WGDs in all species (Figure 4a–d). The Ks distribution were also examined in the transcriptomic data from *P. organensis* (subgenus *Decaloba*) and a combination of three plants of the sour passionfruit (*P. edulis*, subgenus *Passiflora*) (Cai et al., 2019). These species share the ancient genome triplication event (γ whole-genome triplication), exhibiting a second peak that corresponds to a second shared WGD event (Figure 4e,f) beginning in the Eocene, according to Cai et al. (2019).

As expected, the compared Malpighiales species exhibited extensive blocks of microsynteny, showing distinct gene

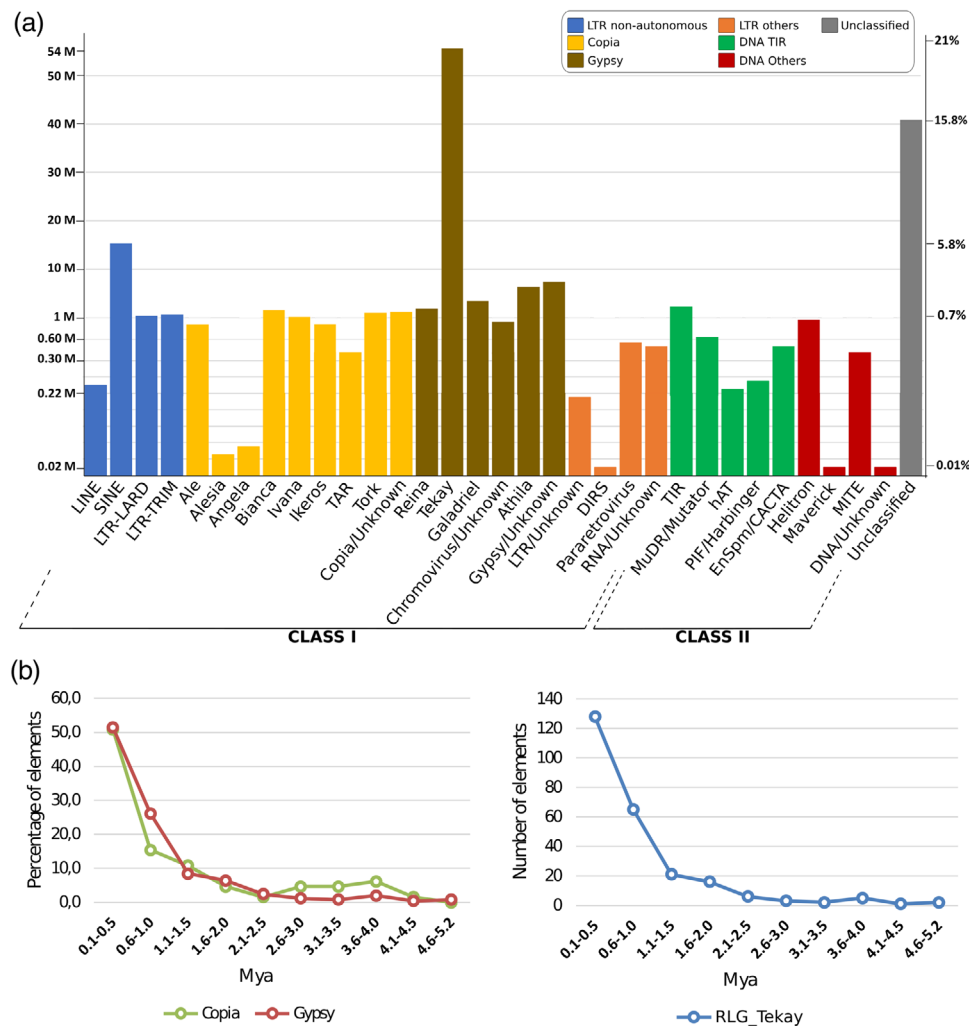


FIGURE 3 Transposable elements recovered in the *Passiflora organensis* genome. (a) The length (in Mbp) (on the left) and percentage occupied by each evolutionary lineage (on the right) were plotted on the y-axis. (b) Estimated insertion times of the retrotransposon evolutionary lineages *Copia*, *Gypsy*, and *RLG_Tekay* in the *Passiflora organensis* genome

retention patterns, fractionation, and alternative deletions of duplicated genes. Some of these rearrangements are shown in Figure 5a–c. In general, the Malpighiales analyzed were collinear in regard to gene orders, supporting the idea of a common ancestor and distinct levels of genome-wide reorganization in each lineage after Malpighiales diversification.

Moreover, ANI between coding regions revealed higher levels of coding region ANI (> 93%) in different Salicaceae species (e.g., *P. trichocarpa*, *P. deltoidea*, and *S. purpurea*; Figure 5d). In contrast, related Euphorbiaceae, such as *M. esculenta* and *R. communis*, exhibited lower coding region ANI (~82%). Unsurprisingly, the Salicaceae, Euphorbiaceae, and Passiflorea species were generally far more similar to each other in terms of coding regions ANI than individual species compared with *L. usitatissimum* (Linaceae), corroborating their phylogenetic distance.

3.5 | Orthogroups and GO enrichment analysis

We used Ortho-MCL to assess orthologous families in each genome and GO enrichment patterns in the complete sequenced genomes of *P. organensis*, *P. trichocarpa*, *S. purpurea*, *M. esculenta*, and *R. communis*. In total, 25,327 coding genes were predicted. *Passiflora organensis* retains 5,609 singletons and 15,671 gene families. Of this total, *P. organensis* shared 12,920 gene families with all the genomes compared (Figure 6), and 981 genes were unique and grouped in 369 families (Supplemental Table S11). The number of members in each family ranged from 2 to 15 genes. The largest family belonged to a hypothetical protein (Cluster 316), followed by a family of proteins (Cluster 415) with DNA-binding domains (GO 0003677).

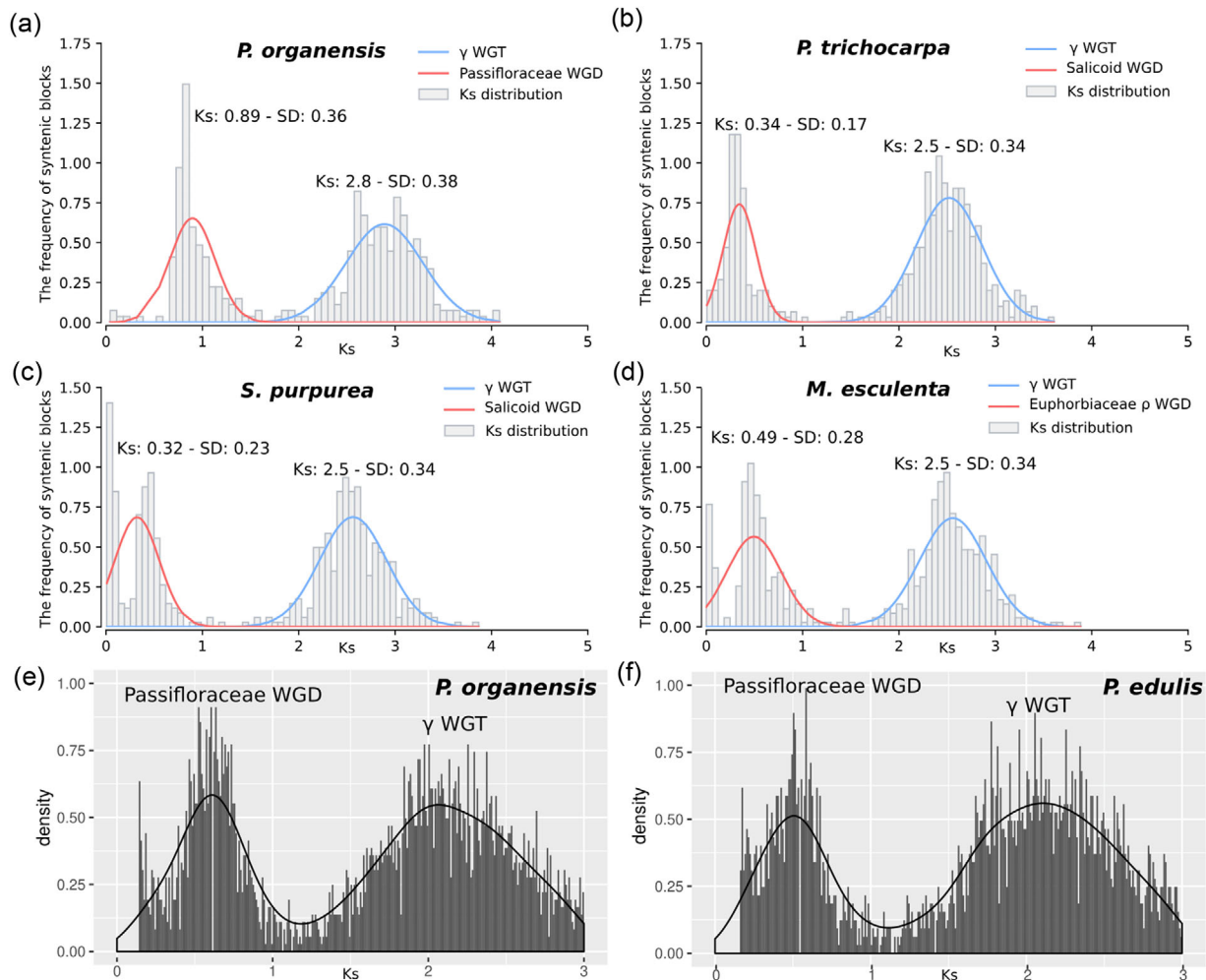


FIGURE 4 Whole-genome duplication (WGD) analysis. The distribution of synonymous nucleotide substitution rates (K) defining the duplication of large chromosomal regions in the genomic data of (a) *Passiflora organensis*, (b) *Populus trichocarpa*, (c) *Salix purpurea*, and (d) *Manihot esculenta*. Distribution of K rates based on transcriptomic data from (e) *P. organensis* and (f) *Passiflora edulis*. All species share the ancient genome triplication event [γ whole-genome triplication (WGT)] associated with the early diversification of the core eudicots

The genomes analyzed shared enriched GO terms mostly related to development, defense, and molecular events such as transcription, translation, and signal transduction (Figure 6). Each species had a particular set of terms assigned for protein-encoding genes in a given functional category. For *P. organensis*, the enriched terms were the biological process medium-chain fatty acid metabolic process (GO:0051791) and the molecular function glutathione transferase activity (GO:0004364), encompassing six proteins. The former may be relevant for seed oil composition studies in *Passiflora* (see Krist, 2020). Glutathione transferase activity has some relevance to various biological processes, from plant development to defense against biotic and abiotic stresses (Gallé et al., 2019). For the complete list of GO annotations, see Supplemental Table S12.

We also investigated whether selection pressure exerted bias on the distribution of genes according to functional categories, taking account of WGD and other duplication events

(dispersed, proximal, and tandem). Most of the genes in each of the function categories to which the GO terms had been assigned were duplicated in large-scale events (WGD). Only functions related to circadian processes (rhythmic process, GO:0048511; circadian rhythm, GO:0007623) were biased towards local tandem duplication events (Supplemental Table S12).

3.6 | The *P. organensis* MADS-box gene family

As an example of how expansions and contractions in gene families have affected the structure of the *P. organensis* genome, we compared it with the MADS-box family members in *A. thaliana*, *P. trichocarpa*, and *V. vinifera* (Figure 7). The MADS-box family of transcription factors is involved in many important developmental processes, and they are well-known for their role in phase transition and establishing the identity

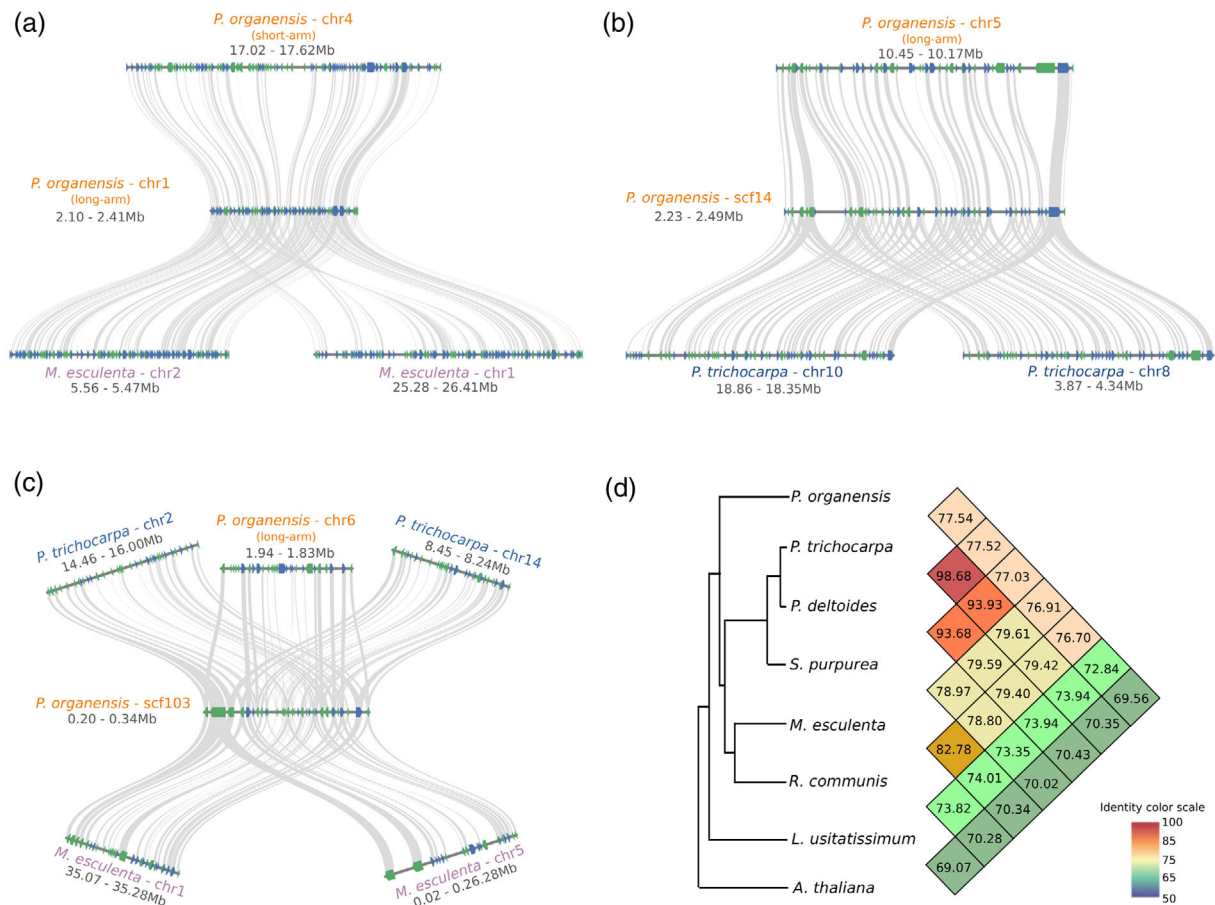


FIGURE 5 Microsynteny and comparative analysis of *Passiflora organensis* and other Malpighiales, showing different patterns of gene retention and loss in different segments of the same genome, and across the genome segments of each species. (a, b) Exclusive and shared microsyntenic segments in the genomes of *P. organensis* and *Manihot esculenta*, and in *P. organensis* and *Populus trichocarpa*: (a) microsynteny between *P. organensis* and *M. esculenta*; (b) microsynteny between *P. organensis* and *P. trichocarpa*. (c) Shared microsyntenic segments in the genomes of *P. organensis*, *M. esculenta*, and *P. trichocarpa*. (d) Average nucleotide identity (ANI) and neighbor-joining tree based on the coding regions of all sequenced Malpighiales genomes available in public databases; *Arabidopsis thaliana* was the outgroup

of floral organs [reviewed by Kramer (2019)]. We chose this gene family because some of its members have already been studied in detail by our group and might have important biological and evolutionary roles in *Passiflora*, especially with regard to reproductive development (Cutri & Dornelas, 2012; Scorza et al., 2017; Silveira et al., 2016). We initially found 104 putative MADS-box family members in *P. organensis*. After manual sequence curation, we ended up with 75 complete sequences that most probably represent the total set of family members in *P. organensis*. Among these sequences, 33 belong to the M-type (Type I) and 42 to the MADS intervening keratin-like and C-terminal (MIKC)-type (Type II). By comparing these sequences with those found in other plant genomes, we were able to infer expansions and contractions within specific subfamilies (see Supplemental Figure S5 and Supplemental Figure S6; Supplemental Table S13 and Supplemental Table S14).

The accuracy of identification of MADS-box family members is shown in Figure 7. The bootstrap values of branches

containing putative *Passiflora* MADS-box family members and recognized MADS-box members from other plant species (including the model species *A. thaliana*) ranged from 75 to 100%, confirming the accuracy of paralog and ortholog identification.

We observed a drastic reduction in the number of members belonging to the M-type subfamily compared with *A. thaliana* and *P. trichocarpa* (Supplemental Figure S6, Supplemental Table S13, Supplemental Table S14). Because of their acknowledged importance in modulating plant development, we focused on the MIKC-type members. The presence of a sequence belonging to TM8 is also noteworthy; this sequence is missing from the *A. thaliana* genome, but present in many other species, including *P. trichocarpa* and *V. vinifera* (Supplemental Figure S6, Supplemental Table S13, Supplemental Table S14). Nonetheless, the group that exhibited the largest expansion within the MIKC-type MADS-box genes was PI/AP3, with six members in *P.*

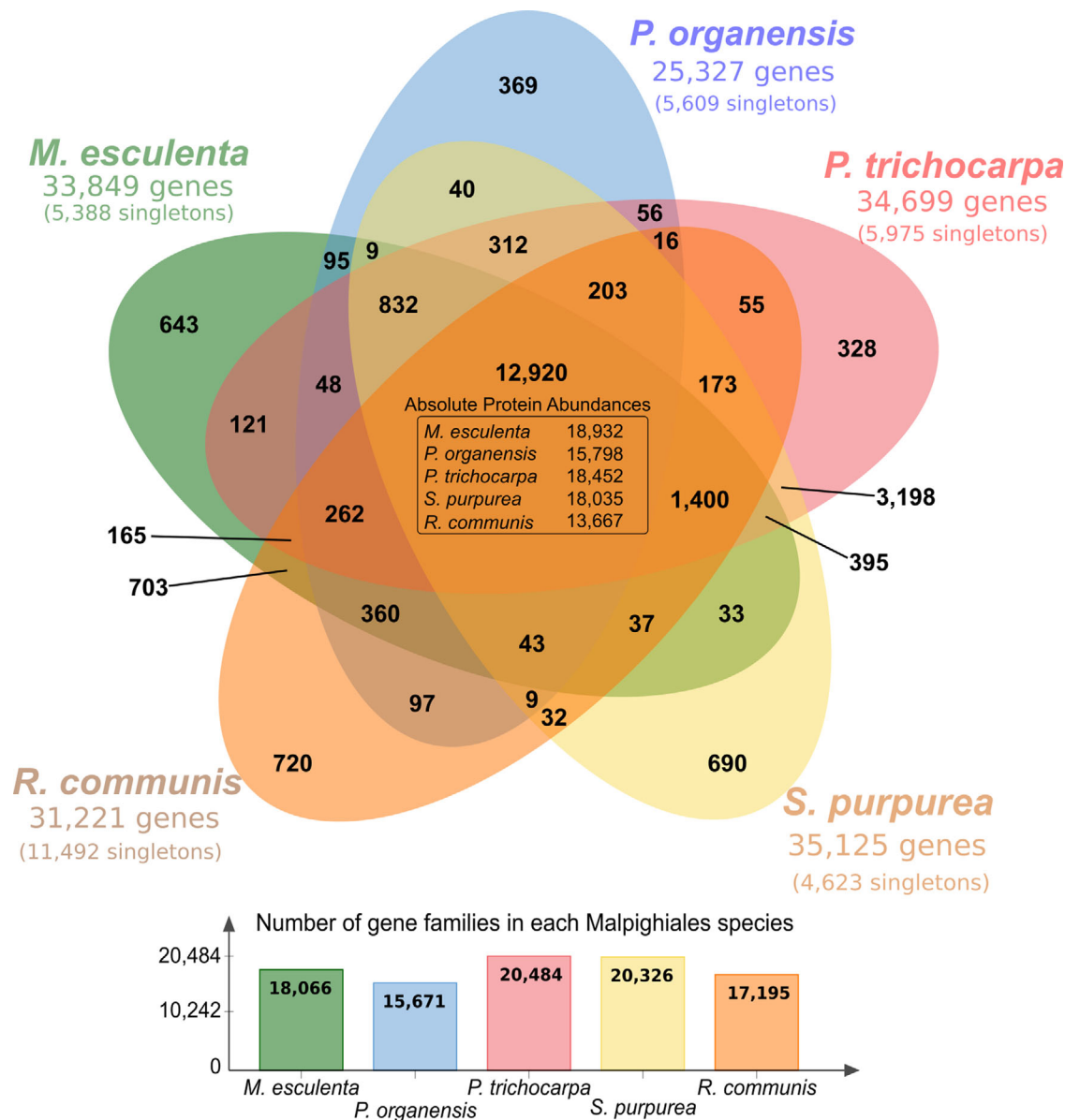


FIGURE 6 Venn diagram showing the distribution of orthologous gene families among the Malpighiales *Passiflora organensis* (Passifloraceae), *Populus trichocarpa*, *Salix purpurea* (both Salicaceae), *Ricinus communis*, and *Manihot esculenta* (both Euphorbiaceae)

organensis but only two in *A. thaliana* and *V. vinifera* (Supplemental Table S13).

3.7 | Chloroplast and mitochondrial genomes

A typical chloroplast DNA circular molecule was recovered with a quadripartite structure spanning 144.959 Mbp. It has 37% guanine–cytosine content and encodes a total of 104 unique genes (Supplemental Table S15, Supplemental Figure S7), 70 of which are protein coding genes, four of which are rRNA genes (4.5S, 5S, 16S, and 23S) and 30 of which are tRNA genes. The *P. organensis* chloroplast genome contains rearrangements that occurred throughout the *Decaloba* sub-

genus's evolution, such as the contraction of inverted repeat regions, and gene and intron losses (Supplemental Figure S8).

We assembled two mitochondrial DNA molecules confirmed throughout with long reads (PacBio and Nanopore) and pairs of short Illumina reads. The linear sequence represents the master mitochondrial DNA molecule comprising 1,031,229 bp (45% guanine–cytosine content), harboring at least one full copy of most canonical mitochondrial protein-encoding genes (32), rRNA genes (two full copies of 5S, and one each of 18S and 26S), and 30 tRNA genes. These genes were interspersed with three mitoviruses comprising a repetitive region of ~1,800 bp, besides relics of other mobile and repetitive elements (Figure 8, Supplemental Table S16, Supplemental Table S17). The mitogenome also

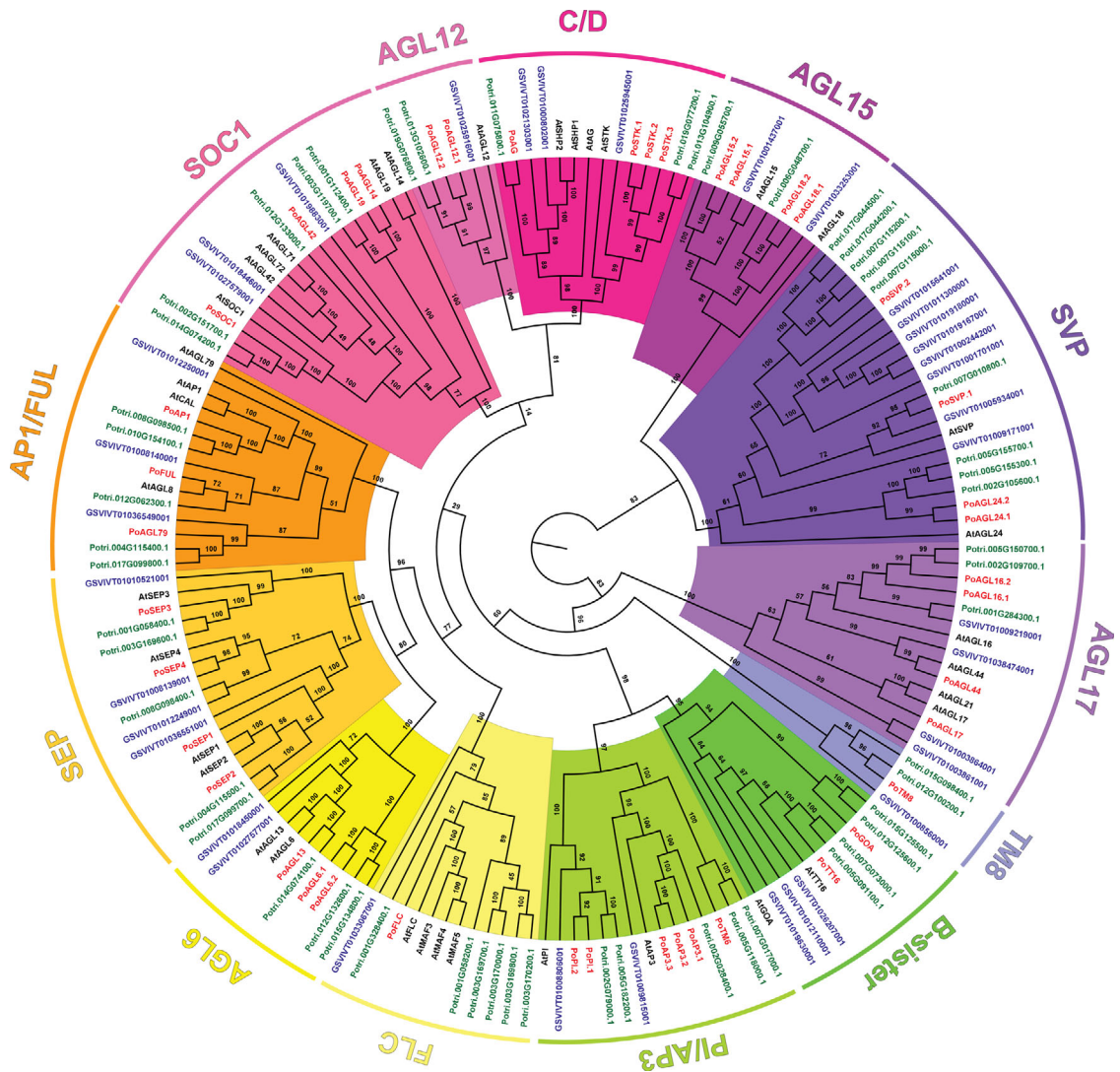


FIGURE 7 Phylogenetic analysis of MADS intervening keratin-like and C-terminal (MIKC)-type MADS-box genes in *Passiflora organensis*, *Arabidopsis thaliana*, *Populus trichocarpa*, and *Vitis vinifera*. The sequences were separated into groups: C and D gene classes (C/D), agamous-like genes (AGL), Flowering Locus c (FLC), sepallata (SEP), apetala 1/fruitfull (API/FUL), suppressor of overexpression of CO 1 (SOC1), tomato mads-8 (TM8), short vegetative phase (SVP), B-sister (the sister group of the B genes), and pistillata/apetala 3 (PI/AP3). Sequences prefixed Po relate to *P. organensis*, At to *A. thaliana*, Potri to *P. trichocarpa*, and GSVIV to *V. vinifera*. The numbers at the nodes are the bootstrap test results

exhibited interspersed sequences of chloroplast DNA. The various repetitive sequences were interspersed within intergenic regions, providing potential sites for rearrangement. In Figure 8, we have highlighted regions of this kind with more than 800 bp and 75% identity.

The genes encoding subunits of the mitochondrial membrane respiratory chain nicotinamide adenine dinucleotide hydrogen (NADH) dehydrogenase Complex I, Nad1 (NADH-ubiquinone oxidoreductase chain 1), Nad2 (NADH-ubiquinone oxidoreductase chain 2), and Nad5 (NADH-ubiquinone oxidoreductase chain 5), exhibited a combination of introns, some of which required *trans*-splicing reactions (Figure 8, Supplemental Table S16, Supplemental Table S17).

The second molecule was circular, comprising 102,307 bp and harboring full versions of two canonical genes missing on the linear molecule (*rsp19*, *rpl10*), besides extra lysine and methionine tRNA. The repeats in the circular version were smaller than 200 bp. In Figure 8, we have highlighted a unique region of 200 bp with an identity higher than 75%.

4 | DISCUSSION

There are many studies on the attributes of *Passiflora* species (Ingale & Hivrale, 2010; de Castro et al., 2018; Cauz-Santos et al., 2020; Krist, 2020) but they have been severely limited by the lack of available sequence data. Even an initial

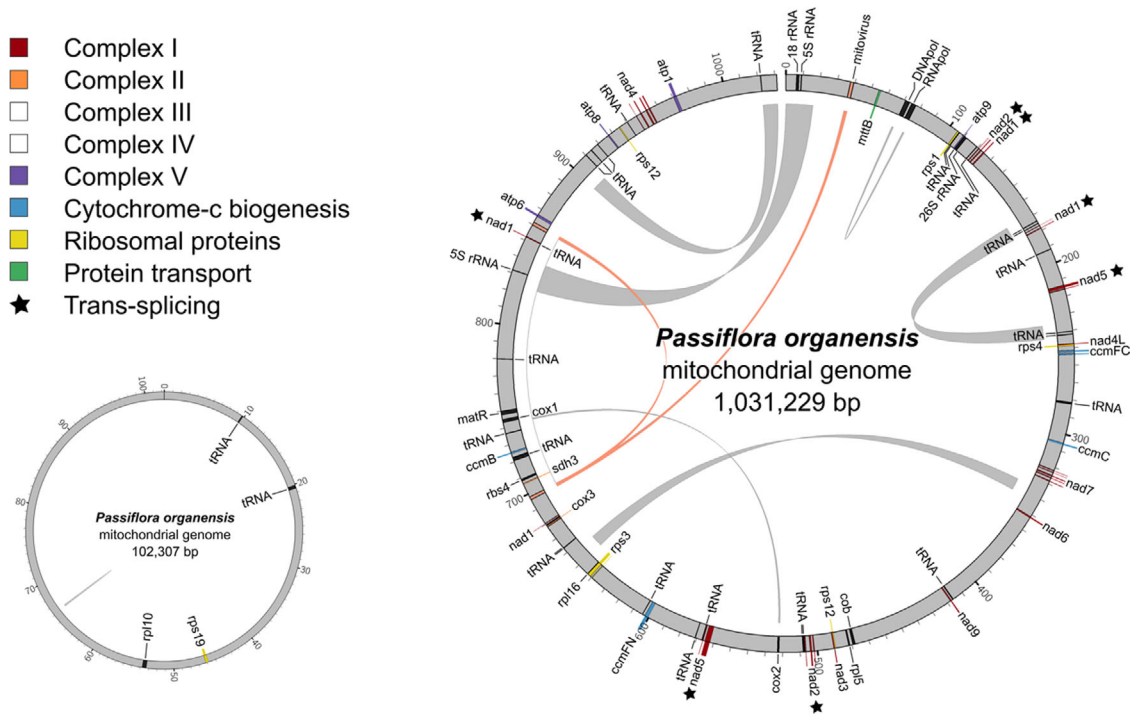


FIGURE 8 Schematic representation of the mitochondrial circular (left) and linear (right) molecules of *Passiflora organensis*. Their main features and gene organization are indicated by bars with gene names, tRNA, and rRNA. Repeats are represented as ribbons inside the circles. For the linear molecule, the ribbons link sequences longer than 800 bp sharing >75% identity. For the circular molecule, only one ribbon links a repeat of 200 bp sharing 89% identity. Stars after the gene name denote genes whose transcripts require *trans*-splicing

draft genome assembly opens up a wide range of possibilities that would simply not be feasible without a reference (Worley et al., 2017). The availability of a draft genome for the wild *P. organensis* (subgenus *Decaloba*) will provide a valuable resource for genetic, genomic, functional, and evolutionary studies of *Passiflora*.

By using a hybrid approach combining Illumina and Pacbio sequencing technologies and Bionano optical maps, we were able to generate a 259-Mbp genome, consistent with the estimate obtained by flow cytometry (Yotoko et al., 2011) and approximately 5.4× smaller than the genome of the purple form of *P. edulis* (Xia et al., 2021). As expected, we identified some conflicts when attempting to correctly map the scaffolds onto chromosome arms, and cytogenetic markers were fundamental for resolving some of them. Eight scaffolds were integrated, accounting for ~102 Mbp (Figure 2, Supplemental Table S6). Centromeres were also located in chromosomes 1, 4, and 6; telomeres in chromosomes 1 and 4 (one end), and 5 and 6 (both ends). Combining these approaches significantly improved the final assembly.

A good-quality assembly was obtained and confirmed by parameters, such as high consensus quality (99.98%), a LTR assembly index of 12.19, and complete conserved plant orthologs reaching 98.4%, a value also obtained for other assemblies (Wei et al., 2020; Zhang et al., 2020).

Herein we present the first complete set of genes from a wild *Passiflora* species, including the MADS-box family of transcription factors, putative genes related to self-incompatibility, and so on. This set of genes could significantly assist teams working on *Passiflora*, facilitating searches for specific genes and different kinds of analysis, such as the genetic changes associated with domestication and comparative studies on other species, including the purple form of *P. edulis* recently made available (Xia et al., 2021). Some of the most abundant gene families in this domesticated species are implicated in the defense response and terpene synthase activity (see Supplemental Figure S9), and many terpenoids exhibit potent toxicity and serve as core components of chemical defenses against herbivores, insect pests, and microbial pathogens.

Consistent with the reproductive mechanisms of several *Passiflora* that behave as outcrossing species, a high level of heterozygosity (81%) was found. Some attempts to self-cross *P. organensis* were unproductive (personal communication, M.C. Dornelas, 2020), suggesting that it is self-incompatible. Self-incompatibility is one of the most important genetic mechanisms for impeding self-fertilization but is poorly understood in *Passiflora*. It has been reported that *Passiflora* has a homomorphic SI system controlled by sporophytic and gametophytic mechanisms (Madureira et al., 2014). We

investigated two of the SI reaction determinants (female) in the SI locus described for Brassicaceae [reviewed in Takayama & Isogai (2005)]. We searched for the SRK gene, which regulates the SI response in the stigma, and SLG, which boosts SI response efficiency (Takasaki et al., 2000) via a modified approach first proposed for *P. edulis* (Madureira et al., 2014). Fifty-four proteins in *P. organensis* have domains commonly detected in the proteins encoded by the S-locus (Xing et al., 2013). We detected a cluster of genes in *P. organensis* with the structural characteristics of the S-locus, on the basis of the proximity of the SRK and SLG candidate genes. Identification of the S-locus would be the first step towards elucidating SI in *Passiflora*. On the basis of the sequences identified herein, it would be possible to detect SI genes in commercialized species e.g., *P. edulis* and *Passiflora alata* Curtis, which could help in breeding improved cultivars.

Transposable elements are important drivers of plant genomes' structure and evolution. For the first time, all TE copies in a *Passiflora* species were identified. Previous studies have identified a collection of complete TEs in *P. edulis*, but only for a gene-rich fraction of the genome (~1%) (Costa et al., 2019). In line with the majority of plant genomes (Orozco-Arias et al., 2019), the most frequent elements in *P. edulis* and *P. organensis* were from the LTR order, particularly the *Tekay* lineage, a sign of possible activity. Intriguingly, long interspersed nuclear elements have significantly proliferated in *P. organensis*, although they are less probable in plants (Wicker et al., 2007). In addition, insertion time analysis suggested that LTR-RTs were recently inserted into the *P. organensis* genome, a pattern similar to that found in *P. edulis* (Costa et al., 2019), possibly indicating that full-length LTR-RT elements have recently become active. Note that repetitive sequences are very difficult to assemble and, consequently, can result in highly fragmented assemblies. This made the challenge we faced even greater (Michael & VanBuren, 2015).

In a comparative analysis of the MADS-box family, we identified 75 complete sequences, 33 of the M-type (Type I) and 42 of the MIKC-type (Type II). A drastic reduction in the number of members belonging to the M-type subfamily was observed. This contraction was previously reported to occur in *Vitis* (Grimplet et al., 2016), but its biological implications remain unexplained. A sequence belonging to the TM8 group was also identified. This sequence is missing from the *A. thaliana* genome but is present in many other species, including *P. trichocarpa* and *V. vinifera* (Supplemental Figure S6, Supplemental Table S13). The biological role of TM8 has been investigated in tomato (*Solanum lycopersicum* L.), where it functions in fruit development (Daminato et al., 2014).

Finally, the PI/AP3 group showed the largest expansion within the MIKC-type MADS-box family (Supplemental Table S13). Expansion of the PI/AP3 group has been reported in Orchidaceae and Zingiberaceae (Bartlett & Specht, 2010;

Mondragón-Palomino & Theißen, 2008). Both plant groups exhibit floral structures that are sui generis (such as the labellum in Orchidaceae and the petaloid stamens in Zingiberaceae). These floral structures are considered evolutionary novelties and their appearance has been correlated with duplications followed by neofunctionalization of genes in the PI/AP3 group (Gioppato & Dornelas, 2019; Hernández-Hernández et al., 2007; Kim et al., 2004; Rijpkema et al., 2010). Accordingly, the corona filaments in the flowers of *Passiflora* species are also considered sui generis floral organs or petaloid stamens (Bernhard, 1999; Hemingway et al., 2011), and it is tempting to speculate that the origin of the *Passiflora* corona might be related to the expansion of the PI/AP3 group of MADS-box genes.

Passiflora organensis has $2n = 12$, as do most species in the *Decaloba* subgenus (De Melo & Guerra, 2003; Hansen et al., 2006; Sader, Amorim, et al., 2019). Nevertheless, apart from the shared γ whole-genome triplication in eudicots, *P. organensis* shares a WGD event with *P. edulis* dated to the Eocene (Cai et al., 2019), which coincides with the origin of the *Passiflora* genus around 42.9 mya (Sader, Amorim, et al., 2019). However, a recently published Ks analysis showed that two WGD events occurred 65 and 12 mya, which may have contributed to the large genome size of *P. edulis* (~1341.7 Mb) (Xia et al., 2021). Further research is required to resolve these discrepancies.

Thus, in contrast to what we proposed earlier (Sader, Amorim, et al., 2019), the genomic data support the idea of $x = 12$ for *Passiflora* (Hansen et al., 2006), dropping to $n = 6$ by descending dysploidy in *P. organensis*. The drop in chromosome number back to $n = 6$ after the ancestral *Passiflora* WGD was accompanied by a reduction in genome size, resembles the genome repatterning observed in *A. thaliana* (Lysak et al., 2006). These rearrangements are supported by the discovery of interstitial telomeric sequences in the two scaffolds. On the basis of a probabilistic model [see Mayrose et al. (2010) for details] that assumes chromosome gains and losses, the predicted base chromosome number was $x = 12$, with a probability of 0.98 (Mayrose et al., 2010). The only possible type of event that would explain the chromosome numbers in the subgenera *Decaloba* is descending dysploidy, as specified herein. Differences in chromosome number and genome size between *P. edulis* and *P. organensis* are associated with important genomic rearrangements, shown here by a comparative analysis that indicated wide macroscale discrepancies and conserved microscale regions (Supplemental Figure S9A,B).

In land plants, the chloroplast DNA is a small, circular molecule but is nevertheless a vital one because it codes for some 100 proteins that, in combination with nuclear-encoded proteins, play a part in important metabolic processes such as photosynthesis (Kleffmann et al., 2004; Sugiura, 1992). Despite the highly conservative nature of

chloroplast DNA in angiosperms, the chloroplast genome of *P. organensis* exhibits rearrangements following the highly dynamic evolution undergone by the chloroplast DNA of *Passiflora* species (Cauz-Santos et al., 2017; 2020; Shrestha et al., 2019). *Passiflora organensis* is placed in the *Decaloba* subgenus, which encompasses a high number of rearrangements already reported for Passifloraceae, with a combination of structural inversions, gene/intron losses, and variations in inverted repeat length caused by inverted repeat expansions and contractions, including the loss of an entire inverted repeat region in some species (Cauz-Santos et al., 2020; Shrestha et al., 2019).

Finally, we explored multiple assemblies of the mitochondrial genome with a combination of long and short reads to capture a sequence that would best represent the mitogenome of *P. organensis*. As reported in a study on lettuce (*Lactuca* spp.) species (Kozik et al., 2019), variants of mitogenome sequences can be recovered in a whole genome sequencing project with long reads. We presented the two molecules assembled in this study, conclusively confirmed by short pair-end Illumina reads and long Pacbio and Nanopore reads. The *P. organensis* mitogenome exhibits the canonical genes essential for respiratory chain functions and protein synthesis, including rRNA and tRNA. However, the complete set must include the two molecules assembled, one larger linear sequence of 1,031,229 bp and one small circular molecule of 102,307 bp. Despite numerous attempts, we did not find any reads combining the two molecules. The fact that there is more than one molecule in plant mitochondria comes as no surprise, since some authors have described one or more molecules, both circular and linear (Kozik et al., 2019). The *P. organensis* mitogenome is among the largest genomes assembled to date compared with other species (Wynn & Christensen, 2019). Only ~5% of the sequence coded for proteins, rRNA, and tRNA. The remaining sequences were immigrants from the chloroplast and other genetic elements, most of which were not found in the *P. edulis* mitogenome (Yang & Wang, 2020).

One of these elements found in the *P. organensis* mitogenome consisted of three copies of an RNA-dependent RNA polymerase of mitoviruses detected in many plant mitochondrial DNAs (Bruenn et al., 2015; Nibert et al., 2018). They are called nonretroviral endogenized RNA virus elements. These are simple viruses carrying only one RNA-dependent RNA polymerase; they are mostly of fungal origin and were potentially acquired by mitogenomic horizontal transfer. Mitogenomes in plant pathogenic fungi are associated with a reduction in virulence towards their hosts. The role of such molecules in plant mitogenomes is still elusive (Nibert et al., 2018). Sequences related to mitoviruses were missing from the published version of the *P. edulis* mitogenome. Note that, in addition to variations in mitogenome intergenic regions, gene organization and the presence of introns involving both *cis*- and *trans*-splicing

were remarkably conserved. The number of large repeats merits further investigation since they may explain recombination events, the size of plant mitochondrial genomes, the recruitment of extra elements, and their contribution to the plant's evolution (Kozik et al., 2019; Wynn & Christensen, 2019).

DATA AVAILABILITY STATEMENT

The Whole Genome Shotgun project has been deposited at DNA Data Bank of Japan, European Nucleotide Archive, and GenBank under accession JAEPCBF000000000. The genome version presented in this paper is JAEPCBF010000000. *Passiflora organensis* genome annotations are available at <https://genomevolution.org/coge/GenomeInfo.pl?gid=59125>.

CONFLICT OF INTEREST

The authors declare that they have no competing interests. Mendelics Análise Genômica provided support only in the form of a salary for J.P.K., but had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Zirlane Portugal Costa: formal analysis, investigation, methodology, writing—original draft. Alessandro Mello Varani: data curation, formal analysis, methodology, writing—original draft. Luiz Augusto Cauz-Santos: data curation, formal analysis, investigation, methodology, resources, writing—original draft. Mariela Analía Sader: investigation, methodology, writing—original draft. Helena Augusto Giopatto: formal analysis, investigation. Bruna Zirpoli: investigation. Caroline Callot: investigation. Stephane Cauet: investigation. Willian Marande: data curation, methodology. Jessica Luana Souza Cardoso: formal analysis. Daniel Guariz Pinheiro: formal analysis, methodology. João Paulo Kitajima: data curation, formal analysis, methodology. Marcelo Carnier Dornelas: conceptualization, formal analysis, funding acquisition, supervision, writing—original draft. Andrea Pedrosa Harand: formal analysis, investigation, resources, supervision, writing—original draft. Helene Berges: conceptualization, supervision. Claudia Barros Monteiro-Vitorello: formal analysis, investigation, methodology, writing—review and editing. Maria Lucia Carneiro Carneiro Vieira: conceptualization, funding acquisition, supervision, writing—review and editing.

ACKNOWLEDGMENTS

This work was supported by the following Brazilian Institutions: Fundação de Amparo à Pesquisa do Estado de São Paulo (Project Numbers: 2017/04216-9, 2017/11815-6, 2018/21469-0, 2018/25242-0, 2019/07838-6), Conselho Nacional de Desenvolvimento Científico e Tecnológico (Project Number: 301017/2019-0; 303965/2015-0;

310804/2017-5), Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (BIC 0370–2.02/20), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Finance Code 001). We are also grateful to Institut National de la Recherche Agronomique and the Centre National de Ressources Génomique Végétales, France, and to Mr. Steve Simons for proofreading the manuscript.

ORCID

Zirlane Portugal Costa  <https://orcid.org/0000-0002-2374-0656>

Alessandro Mello Varani  <https://orcid.org/0000-0002-8876-3269>

Luiz Augusto Cauz-Santos  <https://orcid.org/0000-0003-1694-2433>

Mariela Analía Sader  <https://orcid.org/0000-0001-8188-2217>

Helena Augusto Giopatto  <https://orcid.org/0000-0002-4712-468X>

Bruna Zirpoli  <https://orcid.org/0000-0002-9477-2516>

Stephane Cauet  <https://orcid.org/0000-0003-4985-8940>


Jessica Luana Souza Cardoso  <https://orcid.org/0000-0002-0483-9060>

Daniel Guariz Pinheiro  <https://orcid.org/0000-0001-7062-5936>

Marcelo Carnier Dornelas  <https://orcid.org/0000-0002-6710-3050>

Andrea Pedrosa Harand  <https://orcid.org/0000-0001-5213-4770>

Helene Berges  <https://orcid.org/0000-0002-5492-1062>

Claudia Barros Monteiro-Vitorello  <https://orcid.org/0000-0002-1238-9354>

Maria Lucia Carneiro Vieira  <https://orcid.org/0000-0003-0341-5714>

REFERENCES

- Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology*, 20, 224. <https://doi.org/10.1186/s13059-019-1829-6>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Araya, S., Martins, A. M., Junqueira, N. T. V., Costa, A. M., Faleiro, F. G., & Ferreira, M. E. (2017). Microsatellite marker development by partial sequencing of the sour passion fruit genome (*Passiflora edulis* Sims). *BMC Genomics*, 18, 549. <https://doi.org/10.1186/s12864-017-3881-5>
- Bartlett, M. E., & Specht, C. D. (2010). Evidence for the involvement of *GLOBOSA*-like gene duplications and expression divergence in the evolution of floral morphology in the Zingiberales. *New Phytologist*, 187, 521–541. <https://doi.org/10.1111/j.1469-8137.2010.03279.x>
- Bernhard, A. (1999). Flower structure, development, and systematics in Passifloraceae and in *Abatia* (Flacourtiaceae). *International Journal of Plant Sciences*, 160, 135–150. <https://doi.org/10.1086/314112>
- Boetzer, M., & Pirovano, W. (2014). SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, 15, 211. <https://doi.org/10.1186/1471-2105-15-211>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bruenn, J. A., Warner, B. E., & Yerramsetty, P. (2015). Widespread mitovirus sequences in plant genomes. *PeerJ*, 2015, e876. <https://doi.org/10.7717/peerj.876>
- Cabral, J. S., Felix, L. P., & Guerra, M. (2006). Heterochromatin diversity and its co-localization with 5S and 45S rDNA sites in chromosomes of four *Maxillaria* species (Orchidaceae). *Genetics and Molecular Biology*, 29, 659–664. <https://doi.org/10.1590/S1415-47572006000400015>
- Cai, L., Xi, Z., Amorim, A. M., Sugumaran, M., Rest, J. S., Liu, L., & Davis, C. C. (2019). Widespread ancient whole genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *New Phytologist*, 221, 565–576. <https://doi.org/10.1111/nph.15357>
- Carlson, J. E., Tulsieram, L. K., Glaubitz, J. C., Luk, V. W. K., Kauffeldt, C., & Rutledge, R. (1991). Segregation of random amplified DNA markers in F₁ progeny of conifers. *Theoretical and Applied Genetics*, 83, 194–200. <https://doi.org/10.1007/BF00226251>
- Carvalho, C. R., & Saraiva, L. S. (1993). An air drying technique for maize chromosomes without enzymatic maceration. *Biotechnic and Histochemistry*, 68, 142–145. <https://doi.org/10.3109/10520299309104684>
- Cauz-Santos, L. A., Costa, Z. P., Callot, C., Cauet, S., Zucchi, M. I., Bergès, H., van den Berg, C., & Vieira, M. L. C. (2020). A repertoire of rearrangements and the loss of an inverted repeat region in *Passiflora* chloroplast genomes. *Genome Biology and Evolution*, 12, 1841–1857. <https://doi.org/10.1093/gbe/evaa155>
- Cauz-Santos, L. A., Munhoz, C. F., Rodde, N., Cauet, S., Santos, A. A., Penha, H. A., Dornelas, M. C., Varani, A. M., Oliveira, G. C. X., Bergès, H., & Vieira, M. L. C. (2017). The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: Structural organization and phylogenomic studies in Malpighiales. *Frontiers in Plant Science*, 8, 334. <https://doi.org/10.3389/fpls.2017.00334>
- Cerqueira-Silva, C. B. M., Faleiro, F. G., de Jesus, O. N., dos Santos, E. S. L., & de Souza, A. P. (2015). The genetic diversity, conservation, and use of passion fruit (*Passiflora* spp.). In M. R. Ahuja & S. M. Jain (Eds.), *Genetic diversity and erosion in plants*. Springer.
- Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G. R., Delledonne, M., Luo, C., Ecker, J. R., Cantu, D., Rank, D. R., & Schatz, M. C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, 13, 1050–1054. <https://doi.org/10.1038/nmeth.4035>
- Costa, Z. P., Cauz-Santos, L. A., Ragagnin, G. T., Van Sluys, M.-A., Dornelas, M. C., Berges, H., de Mello Varani, A., & Vieira, M. L. C. (2019). Transposable element discovery and characterization of LTR-retrotransposon evolutionary lineages in the tropical fruit species *Passiflora edulis*. *Molecular Biology Reports*, 46, 6117–6133. <https://doi.org/10.1007/s11033-019-05047-4>

- Cutri, L., & Dornelas, M. C. (2012). Passioma: Exploring expressed sequence tags during flower development in *Passiflora* spp. *Comparative and Functional Genomics*, 2012, 1–11. <https://doi.org/10.1155/2012/510549>
- Daminato, M., Masiero, S., Resentini, F., Lovisetto, A., & Casadoro, G. (2014). Characterization of *TM8*, a MADS-box gene expressed in tomato flowers. *BMC Plant Biology*, 14, 319. <https://doi.org/10.1186/s12870-014-0319-y>
- de Castro, É. C. P., Zagrobelny, M., Cardoso, M. Z., & Bak, S. (2018). The arms race between heliconiine butterflies and *Passiflora* plants—New insights on an ancient subject. *Biological Reviews*, 93, 555–573. <https://doi.org/10.1111/brv.12357>
- De Melo, N. F., & Guerra, M. (2003). Variability of the 5S and 45S rDNA sites in *Passiflora* L. species with distinct base chromosome numbers. *Annals of Botany*, 92, 309–316. <https://doi.org/10.1093/aob/mcg138>
- Fonsêca, A., Ferreira, J., Dos Santos, T. R. B., Mosiolek, M., Bellucci, E., Kami, J., Gepts, P., Geffroy, V., Schweizer, D., Dos Santos, K. G. B., & Pedrosa-Harand, A. (2010). Cytogenetic map of common bean (*Phaseolus vulgaris* L.). *Chromosome Research*, 18, 487–502. <https://doi.org/10.1007/s10577-010-9129-8>
- Gallé, Á., Czékus, Z., Bela, K., Horváth, E., Ördög, A., Csiszár, J., & Poór, P. (2019). Plant glutathione transferases and light. *Frontiers in Plant Science*, 9, 1944. <https://doi.org/10.3389/fpls.2018.01944>
- Gioppato, H. A., & Dornelas, M. C. (2019). When Bs are better than As: The relationship between B-Class MADS-Box gene duplications and the diversification of perianth morphology. *Tropical Plant Biology*, 12, 1–11. <https://doi.org/10.1007/s12042-018-9212-4>
- Gioppato, H. A., da Silva, M. B., Carrara, S., Palermo, B. R. Z., de Souza Moraes, T., & Dornelas, M. C. (2019). Genomic and transcriptomic approaches to understand *Passiflora* physiology and to contribute to passionfruit breeding. *Theoretical and Experimental Plant Physiology*, 31, 173–181. <https://doi.org/10.1007/s40626-018-0134-1>
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, 40, D1178–D1186. <https://doi.org/10.1093/nar/gkr944>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29, 644–652. <https://doi.org/10.1038/nbt.1883>
- Greiner, S., Lehwark, P., & Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, 47, W59–W64. <https://doi.org/10.1093/nar/gkz238>
- Grimplet, J., Martínez-Zapater, J. M., & Carmona, M. J. (2016). Structural and functional annotation of the MADS-box transcription factor family in grapevine. *BMC Genomics*, 17, 80. <https://doi.org/10.1186/s12864-016-2398-7>
- Guerra, M. (1986). Reviewing the chromosome nomenclature of Levan et al. *Revista Brasileira de Genética*, 9, 741–743
- Haas, B. J., Delcher A. L., Mount S. M., Wortman J. R., Smith R. K. J., Hannick L. I., Maiti R., Ronning C. M., Rusch D. B., Town C. D., Salzberg S. L., & White O. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, 31, 5654–5666. <https://doi.org/10.1093/nar/gkg770>
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biology*, 9, R7. <https://doi.org/10.1186/gb-2008-9-1-r7>
- Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95–98
- Hansen, A. K., Gilbert, L. E., Simpson, B. B., Downie, S. R., Cervi, A. C., & Jansen, R. K. (2006). Phylogenetic relationships and chromosome number evolution in *Passiflora*. *Systematic Botany*, 31, 138–150. <https://doi.org/10.1600/036364406775971769>
- Hemingway, C. A., Christensen, A. R., & Malcomber, S. T. (2011). B- and C-class gene expression during corona development of the blue passionflower (*Passiflora caerulea*, Passifloraceae). *American Journal of Botany*, 98, 923–934. <https://doi.org/10.3732/ajb.1100026>
- Hernández-Hernández, T., Martínez-Castilla, L. P., & Alvarez-Buylla, E. R. (2007). Functional diversification of B MADS-box homeotic regulators of flower development: Adaptive evolution in protein–protein interaction domains after major gene duplication events. *Molecular Biology and Evolution*, 24, 465–481. <https://doi.org/10.1093/molbev/msl182>
- Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., & Quesneville, H. (2014). PASTEC: An automatic transposable element classification tool. *PLoS ONE*, 9, e91929. <https://doi.org/10.1371/journal.pone.0091929>
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, 32, 767–769. <https://doi.org/10.1093/bioinformatics/btv661>
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34, 2115–2122. <https://doi.org/10.1093/molbev/msx148>
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., Von Mering, C., & Bork P. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47, D309–D314. <https://doi.org/10.1093/nar/gky1085>
- Ingale, A. G., & Hivrale, A. U. (2010). Pharmacological studies of *Passiflora* sp. and their bioactive compounds. *African Journal of Plant Science*, 4, 417–426
- Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, 45, D1040–D1045. <https://doi.org/10.1093/nar/gkw982>
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30, 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Jouffroy, O., Saha, S., Mueller, L., Quesneville, H., & Maumus, F. (2016). Comprehensive repeatome annotation reveals strong potential

- impact of repetitive elements on tomato ripening. *BMC Genomics*, *17*, 624. <https://doi.org/10.1186/s12864-016-2980-z>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*, 587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780. <https://doi.org/10.1093/molbev/mst010>
- Keller, O., Kollmar, M., Stanke, M., & Waack, S. (2011). A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*, *27*, 757–763. <https://doi.org/10.1093/bioinformatics/btr010>
- Kim, S., Yoo, M. J., Albert, V. A., Farris, J. S., Soltis, P. S., & Soltis, D. E. (2004). Phylogeny and diversification of B-function MADS-box genes in angiosperms: Evolutionary and functional implications of a 260-million-year-old duplication. *American Journal of Botany*, *91*, 2102–2118. <https://doi.org/10.3732/ajb.91.12.2102>
- Kleffmann, T., Russenberger, D., Von Zychlinski, A., Christopher, W., Sjölander, K., Gruissem, W., & Baginsky, S. (2004). The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Current Biology*, *14*, 354–362. <https://doi.org/10.1016/j.cub.2004.02.039>
- Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramírez, F., Vesztrocy, A. W., Naldi, A., Mungall, C. J., Yunes, J. M., Botvinnik, O., Weigel, M., Dampier, W., Dessimoz, C., Flick, P., & Tang, H. (2018). GOA-TOOLS: A Python library for gene ontology analyses. *Scientific Reports*, *8*, 10872. <https://doi.org/10.1038/s41598-018-28948-z>
- Koch, M. A., Haubold, B., & Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Molecular Biology and Evolution*, *17*, 1483–1498. <https://doi.org/10.1093/oxfordjournals.molbev.a026248>
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*, 722–736. <https://doi.org/10.1101/gr.215087.116>
- Kozik, A., Rowan, B. A., Lavelle, D., Berke, L., Schranz, M. E., Michelmore, R. W., & Christensen, A. C. (2019). The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLoS Genetics*, *15*, e1008373. <https://doi.org/10.1371/journal.pgen.1008373>
- Kozomara, A., & Griffiths-Jones, S. (2014). MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, *42*, D68–D73. <https://doi.org/10.1093/nar/gkt1181>
- Kramer, E. M. (2019). Plus ça change, plus c'est la même chose: The developmental evolution of flowers. *Current Topics in Developmental Biology*, *131*, 211–238. <https://doi.org/10.1016/bs.ctdb.2018.11.015>
- Krist, S. (2020). Passion fruit seed oil. In S. Krist, ed. *Vegetable fats and oils* (pp. 535–539). Springer.
- Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, *47*, D807–D811. <https://doi.org/10.1093/nar/gky1053>
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*, 1639–1645. <https://doi.org/10.1101/gr.092759.109>
- Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35*, 3100–3108. <https://doi.org/10.1093/nar/gkm160>
- Laslett, D., & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, *32*, 11–16. <https://doi.org/10.1093/nar/gkh152>
- Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE on-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, *44*, W54–W57. <https://doi.org/10.1093/nar/gkw413>
- Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*, 955–964. <https://doi.org/10.1093/nar/25.5.955>
- Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, *48*, D265–D268. <https://doi.org/10.1093/nar/gkz991>
- Lysak, M. A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., & Schubert, I. (2006). Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 5224–5229. <https://doi.org/10.1073/pnas.0510791103>
- Machado, M., Magalhães, W. C. S., Sene, A., Araújo, B., Faria-Campos, A. C., Chanock, S. J., Scott, L., Oliveira, G., Tarazona-Santos, E., & Rodrigues, M. R. (2011). Phred-Phrap package to analyses tools: A pipeline to facilitate population genetics re-sequencing studies. *Investigative Genetics*, *2*, 3. <https://doi.org/10.1186/2041-2223-2-3>
- Madureira, H. C., Pereira, T. N. S., Da Cunha, M., Klein, D. E., de Oliveira, M. V. V., de Mattos, L., & de Souza Filho, G. A. (2014). Self-incompatibility in passion fruit: Cellular responses in incompatible pollinations. *Biologia*, *69*, 574–584. <https://doi.org/10.2478/s11756-014-0353-0>
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, *14*, e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., ... Bryant, S. H. (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, *45*, D200–D203. <https://doi.org/10.1093/nar/gkw1129>
- Mayrose, I., Barker, M. S., & Otto, S. P. (2010). Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology*, *59*, 132–144. <https://doi.org/10.1093/sysbio/syp083>
- Michael, T. P., & VanBuren, R. (2015). Progress, challenges and the future of crop genomes. *Current Opinion in Plant Biology*, *24*, 71–81. <https://doi.org/10.1016/j.pbi.2015.02.002>

- Mondragón-Palomino, M., & Theißen, G. (2008). MADS about the evolution of orchid flowers. *Trends in Plant Science*, *13*, 51–59. <https://doi.org/10.1016/j.tplants.2007.11.007>
- Munhoz, C. F., Costa, Z. P., Cauz-Santos, L. A., Reátegui, A. C. E., Rodde, N., Cauet, S., Dornelas, M. C., Leroy, P., de Melo Varani, A., Bergès, H., & Vieira, M. L. C. (2018). A gene-rich fraction analysis of the *Passiflora edulis* genome reveals highly conserved microsyntenic regions with two related Malpighiales species. *Scientific Reports*, *8*, 13024. <https://doi.org/10.1038/s41598-018-31330-8>
- Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, *8*, 4321–4325
- Nadalin, F., Vezzi, F., & Policriti, A. (2012). GapFiller: A de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics*, *13*, S8. <https://doi.org/10.1186/1471-2105-13-S14-S8>
- Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*, 2933–2935. <https://doi.org/10.1093/bioinformatics/btt509>
- Neumann, P., Novák, P., & Ho, N. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA*, *10*, 1–17. <https://doi.org/10.1186/s13100-018-0144-1>
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*, 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nibert, M. L., Vong, M., Fugate, K. K., & Debat, H. J. (2018). Evidence for contemporary plant mitoviruses. *Virology*, *518*, 14–24. <https://doi.org/10.1016/j.virol.2018.02.005>
- Novák, P., Neumann, P., Pech, J., Steinhaisl, J., & Macas, J. (2013). RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, *29*, 792–793. <https://doi.org/10.1093/bioinformatics/btt054>
- Orozco-Arias, S., Isaza, G., & Guyot, R. (2019). Retrotransposons in plant genomes: Structure, identification, and classification through bioinformatics and machine learning. *International Journal of Molecular Sciences*, *20*, 3837. <https://doi.org/10.3390/ijms20153837>
- Ou, S., Chen, J., & Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Research*, *46*, e126. <https://doi.org/10.1093/nar/gky730>
- Pamponét, V. C. C., Souza, M. M., Silva, G. S., Micheli, F., De Melo, C. A. F., De Oliveira, S. G., Costa, E. A., & Corrêa, R. X. (2019). Low coverage sequencing for repetitive DNA analysis in *Passiflora edulis* Sims: Cytogenomic characterization of transposable elements and satellite DNA. *BMC Genomics*, *20*, 1–17. <https://doi.org/10.1186/s12864-019-5576-6>
- Pham, G. M., Hamilton, J. P., Wood, J. C., Burke, J. T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J., & Buell, C. R. (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience*, *9*, 1–11. <https://doi.org/10.1093/gigascience/giaa100>
- Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., & Paterson, A. H. (2019). Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biology*, *20*, 38. <https://doi.org/10.1186/s13059-019-1650-2>
- Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., & Anxolabehere, D. (2005). Combined evidence annotation of transposable elements in genome sequences. *PLoS Computational Biology*, *1*, e22. <https://doi.org/10.1371/journal.pcbi.0010022>
- Ranallo-Benavidez, T. R., Jaron K. S., & Schatz M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11*, 1432. <https://doi.org/10.1038/s41467-020-14998-3>
- Rijkema, A. S., Vandenbussche M., Koes, R., Heijmans, K., & Gerats, T. (2010). Variations on a theme: Changes in the floral ABCs in angiosperms. *Seminars in Cell and Developmental Biology*, *21*, 100–107. <https://doi.org/10.1016/j.semcdb.2009.11.002>
- Sader, M., Vaio, M., Cauz-Santos, L. A., Dornelas, M. C., Vieira, M. L. C., Melo, N., & Pedrosa-Harand, A. (2021). Large vs small genomes in *Passiflora*: The influence of the mobilome and the satellitome. *Planta*, *253*, 86. <https://doi.org/10.1007/s00425-021-03598-0>
- Sader, M. A., Amorim, B. S., Costa, L., & Souza, G. (2019). The role of chromosome changes in the diversification of *Passiflora* L. (Passifloraceae). *Systematics and Biodiversity*, *17*, 7–21. <https://doi.org/10.1080/14772000.2018.1546777>
- Sader, M. A., Dias, Y., Costa, Z. P., Munhoz, C., Penha, H., Bergès, H., Vieira, M. L. C., & Pedrosa-Harand, A. (2019). Identification of passion fruit (*Passiflora edulis*) chromosomes using BAC-FISH. *Chromosome Research*, *7*, 299–311. <https://doi.org/10.1007/s10577-019-09614-0>
- Santos, A., Penha, H., Bellec, A., De Munhoz, C., Pedrosa-Harand, A., Bergès, H., & Vieira, M. L. (2014). Begin at the beginning: A BAC-end view of the passion fruit (*Passiflora*) genome. *BMC Genomics*, *15*, 816. <https://doi.org/10.1186/1471-2164-15-816>
- Scorza, L. C. T., Hernandes-Lopes, J., Melo-de-Pinna, G. F. A., & Dornelas, M. C. (2017). Expression patterns of *Passiflora edulis* APETALA1/FRUITFULL homologues shed light onto tendril and corona identities. *EvoDevo*, *8*, 3. <https://doi.org/10.1186/s13227-017-0066-x>
- Seppy, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, *1962*, 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14
- Shrestha, B., Weng, M. L., Theriot, E. C., Gilbert, L. E., Ruhlman, T. A., Krosnick, S. E., & Jansen, R. K. (2019). Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. *Molecular Phylogenetics and Evolution*, *138*, 53–64. <https://doi.org/10.1016/j.ympev.2019.05.030>
- Silveira, S. R., Dornelas, M. C., & Martinelli, A. P. (2016). Perspectives for a framework to understand aril initiation and development. *Frontiers in Plant Science*, *7*, 1919. <https://doi.org/10.3389/fpls.2016.01919>
- Souza, M. M., Palomino, G., Pereira, T. N. S., Pereira, M. G., & Viana, A. P. (2004). Flow cytometric analysis of genome size variation in some *Passiflora* species. *Hereditas*, *38*, 31–38. <https://doi.org/10.1111/j.1601-5223.2004.01739.x>
- Sugiura, M. (1992). The chloroplast genome. *Plant Molecular Biology*, *19*, 149–168. <https://doi.org/10.1007/BF00015612>
- Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A., & Hinata, K. (2000). The S receptor kinase determines self-incompatibility in *Brassica stigma*. *Nature*, *403*, 913–916. <https://doi.org/10.1038/35002628>
- Takayama, S., & Isogai, A. (2005). Self-incompatibility in plants. *Annual Review of Plant Biology*, *56*, 467–489. <https://doi.org/10.1146/annurev.arplant.56.032604.144249>

- The UniProt Consortium. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 45, D158–D169
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, 45, W6–W11. <https://doi.org/10.1093/nar/gkx391>
- Ulmer, T., & MacDougal, J. M. (2004). *Passiflora: Passionflowers of the world*. Timber Press.
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, 33, 2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T.-H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40, e49. <https://doi.org/10.1093/nar/gkr1293>
- Wei, S., Yang, Y., & Yin, T. (2020). The chromosome-scale assembly of the willow genome provides insight into Salicaceae genome evolution. *Horticulture Research*, 7, 45. <https://doi.org/10.1038/s41438-020-0268-6>
- Wicker, T., Sabot, F. F., Hua-Van, A. A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8, 973–982. <https://doi.org/10.1038/nrg2165>
- Worley, K. C., Richards, S., & Rogers, J. (2017). The value of new genome references. *Experimental Cell Research*, 358, 433–438. <https://doi.org/10.1016/j.yexcr.2016.12.014>
- Wynn, E. L., & Christensen, A. C. (2019). Repeats of unusual size in plant mitochondrial genomes: Identification, incidence and evolution. *G3: Genes, Genomes, Genetics*, 9, 549–559. <https://doi.org/10.1534/g3.118.200948>
- Xia, Z., Huang D., Zhang, S., Wang, W., Ma, F., Wu, B., Xu, Y., Xu, B., Chen, D., Zou, M., Xu, H., Zhou, X., Zhan, R., & Song, S. (2021). Chromosome-scale genome assembly provides insights into the evolution and flavor synthesis of passion fruit (*Passiflora edulis* Sims). *Horticulture Research*, 8, 14. <https://doi.org/10.1038/s41438-020-00455-1>
- Xing, S., Li, M., & Liu, P. (2013). Evolution of S-domain receptor-like kinases in land plants and origination of S-locus receptor kinases in Brassicaceae. *BMC Evolutionary Biology*, 13, 69. <https://doi.org/10.1186/1471-2148-13-69>
- Xu, G. C., Xu, T. J., Zhu, R., Zhang, Y., Li, S. Q., Wang, H. W., & Li, J. T. (2018). LR-GapCloser: A tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience*, 8, 1–14. <https://doi.org/10.1093/gigascience/giy157>
- Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y. Q., Coleman-Derr, D., Xia, Q., & Wang, Y. (2019). OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*, 47, W52–W58. <https://doi.org/10.1093/nar/gkz333>
- Xu, Z., & Wang, H. (2007). LTR-FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, 35, W265–W268. <https://doi.org/10.1093/nar/gkm286>
- Yang, J., & Wang, X.-A. (2020). The complete mitochondrial genome of a yellow passion fruit (*Passiflora edulis* Sims. f. *flavicarpa* Deg.) in China and phylogenetic relationships. *Mitochondrial DNA Part B*, 5, 1598–1600. <https://doi.org/10.1080/23802359.2020.1742622>
- Yin, H., Du, J., Wu, J., Wei, S., Xu, Y., Tao, S., Wu, J., & Zhang, S. (2015). Genome-wide annotation and comparative analysis of long terminal repeat retrotransposons between pear species of *P. bretschneideri* and *P. communis*. *Scientific Reports*, 5, 1–15. <https://doi.org/10.1038/srep17644>
- Yotoko, K. S. C., Dornelas, M. C., Togni, P. D., Fonseca, T. C., Salzano, F. M., Bonatto, S. L., & Freitas, L. B. (2011). Does variation in genome sizes reflect adaptive or neutral processes? New clues from *Passiflora*. *PLoS ONE*, 6, e18212. <https://doi.org/10.1371/journal.pone.0018212>
- Yuan, Y., Bayer, P. E., Lee, H.-T., & Edwards, D. (2017). runBNG: A software package for BioNano genomic analysis on the command line. *Bioinformatics*, 33, 3107–3109. <https://doi.org/10.1093/bioinformatics/btx366>
- Zerbini, F. M., Otoni, W. C., & Vieira, M. L. C. (2008). Passionfruit. In C. Kole & T. C. Hall (Eds.), *A compendium of transgenic crop plants. Tropical and subtropical fruit and nuts* (pp. 213–234). Wiley.
- Zhang, J., Yuan, H., Li, Y., Chen, Y., Liu, G., Ye, M., Yu, C., Lian, B., Zhong, F., Jiang, Y., & Xu, J. (2020). Genome sequencing and phylogenetic analysis of allotetraploid *Salix matsudana* Koidz. *Horticulture Research*, 7, 201. <https://doi.org/10.1038/s41438-020-00424-8>
- Zimin, A. V., & Salzberg, S. L. (2020). The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Computational Biology*, 16, e1007981. <https://doi.org/10.1371/journal.pcbi.1007981>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Costa, Z. P., Varani, A. M., Cauz-Santos, L. A., et al. A genome sequence resource for the genus *Passiflora*, the genome of the wild diploid species *Passiflora organensis*. *Plant Genome*. 2021;e20117. <https://doi.org/10.1002/tpg2.20117>