



The Chloroplast Genome of *Passiflora edulis* (Passifloraceae) Assembled from Long Sequence Reads: Structural Organization and Phylogenomic Studies in Malpighiales

Luiz A. Cauz-Santos^{1†}, Carla F. Munhoz^{1†}, Nathalie Rodde², Stephane Cauet², Anselmo A. Santos^{1,3}, Helen A. Penha^{1,4}, Marcelo C. Dornelas⁵, Alessandro M. Varani⁴, Giancarlo C. X. Oliveira¹, Hélène Bergès² and Maria Lucia C. Vieira^{1*}

OPEN ACCESS

Edited by:

Michael Eric Schranz,
Wageningen University and Research
Centre, Netherlands

Reviewed by:

Erlin Pang,
Beijing Normal University, China
Freek T. Bakker,
Wageningen University and Research
Centre, Netherlands

*Correspondence:

Maria Lucia C. Vieira
mlcvieir@usp.br

† These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Plant Genetics and Genomics,
a section of the journal
Frontiers in Plant Science

Received: 05 July 2016

Accepted: 27 February 2017

Published: 10 March 2017

Citation:

Cauz-Santos LA, Munhoz CF,
Rodde N, Cauet S, Santos AA,
Penha HA, Dornelas MC, Varani AM,
Oliveira GCX, Bergès H and
Vieira MLC (2017) The Chloroplast
Genome of *Passiflora edulis*
(Passifloraceae) Assembled from
Long Sequence Reads: Structural
Organization and Phylogenomic
Studies in Malpighiales.
Front. Plant Sci. 8:334.
doi: 10.3389/fpls.2017.00334

¹ Departamento de Genética, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, Brazil, ² Institut National de la Recherche Agronomique, French Plant Genomic Resource Center, Castanet-Tolosan, France, ³ FuturaGene Brasil Tecnologia Ltda., São Paulo, Brazil, ⁴ Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, Brazil, ⁵ Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Brazil

The family Passifloraceae consists of some 700 species classified in around 16 genera. Almost all its members belong to the genus *Passiflora*. In Brazil, the yellow passion fruit (*Passiflora edulis*) is of considerable economic importance, both for juice production and consumption as fresh fruit. The availability of chloroplast genomes (cp genomes) and their sequence comparisons has led to a better understanding of the evolutionary relationships within plant taxa. In this study, we obtained the complete nucleotide sequence of the *P. edulis* chloroplast genome, the first entirely sequenced in the Passifloraceae family. We determined its structure and organization, and also performed phylogenomic studies on the order Malpighiales and the Fabids clade. The *P. edulis* chloroplast genome is characterized by the presence of two copies of an inverted repeat sequence (IRA and IRB) of 26,154 bp, each separating a small single copy region of 13,378 bp and a large single copy (LSC) region of 85,720 bp. The annotation resulted in the identification of 105 unique genes, including 30 tRNAs, 4 rRNAs, and 71 protein coding genes. Also, 36 repetitive elements and 85 SSRs (microsatellites) were identified. The structure of the complete cp genome of *P. edulis* differs from that of other species because of rearrangement events detected by means of a comparison based on 22 members of the Malpighiales. The rearrangements were three inversions of 46,151, 3,765 and 1,631 bp, located in the LSC region. Phylogenomic analysis resulted in strongly supported trees, but this could also be a consequence of the limited taxonomic sampling used. Our results have provided a better understanding of the evolutionary relationships in the Malpighiales and the Fabids, confirming the potential of complete chloroplast genome sequences in inferring evolutionary relationships and the utility of long sequence reads for generating very accurate biological information.

Keywords: chloroplast genome, *Passiflora*, single molecule real-time (SMRT) sequencing, phylogenomics, Malpighiales, Fabids

INTRODUCTION

Malpighiales is an order of flowering plants that belongs to the clade Eurosids I, also known as Fabids (The Angiosperm Phylogeny Group, 2009). This large order includes 42 families, more than 700 genera, and contains approximately 16,000 species forming an extremely diverse group of plants in terms of their morphological and ecological aspects (The Angiosperm Phylogeny Group, 2009; Wurdack and Davis, 2009). The Passifloraceae family is a member of the Malpighiales (Judd et al., 2008) and consists of some 700 species of herbaceous or woody vines, shrubs and trees, classified in around 16 genera, and almost all its members belong to the large and variable genus *Passiflora*, popularly known as passion flowers or passion fruits (Feuillet, 2004). There are around 520 species of *Passiflora*, the majority of which are distributed pantropically; the most derived and specialized species are distributed in the Neotropics and Africa (Ulmer and MacDougal, 2004).

The genus *Passiflora* has long attracted considerable attention due to its economic value, broad geographic distribution, and remarkable species diversity, particularly with regard to flower morphology. The main economic value lies in the production of passion fruit juice, an essential exotic ingredient in juice blends. Furthermore, some species are of great ornamental value or used in phytotherapeutic remedies (Ramaiya et al., 2014). *Passiflora* seed oil is also well-suited for use as a regenerative ingredient in cosmetics products.

In Brazil, passion fruit cultivation began relatively recently and has earned the country an outstanding position as the world's top producer of yellow passion fruit (*Passiflora edulis*). It is an outcrossing, diploid ($n = 9$) (Cuco et al., 2005) species with perfect self-incompatibility (Bruckner et al., 1995; Rêgo et al., 2000), and insect-pollinated flowers. Genetic (Moraes et al., 2005; Oliveira et al., 2008) and molecular-based studies have been carried out in our laboratory (Santos et al., 2014; Munhoz et al., 2015), which is able to satisfy the needs of a wide range of breeders to boost passion fruit crop production and fruit quality.

The nuclear genome size of *P. edulis* (expressed in 1C) was estimated at 1.258 (Yotoko et al., 2011) or 1,595 pg (Souza et al., 2004). More recently, a very efficient strategy for obtaining initial insight into the content of this particular genome involved the sequencing of the terminal regions (BAC-ends) of a representative number of BAC clones selected at random from a *P. edulis* genomic library (Santos et al., 2014). This library was constructed by and deposited at INRA-CNRGV and covers around six times the genome length of *P. edulis*¹. Our group was able to characterize some 10,000 high-quality sequences (100 to 1,255 bp) and identify reads likely to contain repetitive mobile elements and SSRs, and to estimate the GC-content of the reads. Approximately one 10th of the BAC end-sequences

contained protein sequences, and gene ontology terms were assigned to most of them. Finally, we were able to map a number of BAC-end pair sequences to intervals of *Arabidopsis thaliana*, *Vitis vinifera* and chiefly to *Populus trichocarpa* chromosomes, representing regions of potential microsynteny. Additionally, a number of BAC clones were identified as containing chloroplast DNA sequences (Santos et al., 2014).

The chloroplast genome usually occurs in multiple copies within the organelle. It consists of fairly long circular or linear DNA molecules, normally ranging from 120 to 180 kb in angiosperms. It has a quadripartite structure characterized by the presence of two copies of a large IRA and IRB separating a SSC and a LSC region. There is a typical gene partitioning pattern with about 80 protein coding genes in addition to tRNA and rRNA coding genes (Sugiura, 1992; Yang et al., 2010). It also contains 20 group II introns (Barkan, 2004), that derived from a class of mobile elements that are thought to be ancestors of spliceosomal introns and eukaryotic retrotransposons (Lambowitz and Zimmerly, 2011). Though highly conserved, changes in the composition of chloroplast genomes may occur, and rearrangements or even gene losses have been documented (Tangphatsornruang et al., 2011; Li et al., 2013).

Chloroplast DNA, specifically non-coding sequences, has been used extensively to investigate phylogenetic relationships in plants (Shaw et al., 2005), including *Passiflora* species (Muschner et al., 2003; Yockteng and Nadot, 2004). Chloroplast genes such as *rbcl*, *matK*, *ndhF*, *atpB*, and *rps2* have been used in evolutionary studies at higher taxonomic levels. Currently, it is possible to generate entire chloroplast genomes as well as entire chloroplast gene sequences and both can be used simultaneously to determine phylogenies (Martin et al., 2013). For instance, the relationships between wild and domestic species within the genus *Citrus* were elucidated based on a phylogenetic analysis of 34 entire chloroplast genomes (Carbonell-Caballero et al., 2015). Very recently, the complete chloroplast genome sequences were used to infer phylogenetic relationships in the *Quercus* genus (Yang et al., 2016).

The development of NGS technologies has provided highly efficient, low-cost DNA sequencing platforms that produce large volumes of short reads (Shin et al., 2013). However, more recently, third generation sequencing technologies producing longer DNA reads have begun to emerge, including the Pacific Biosciences single molecule real-time (SMRT) sequencer that became available in 2011². For example, using PacBio sequence data it was possible to generate an entire chloroplast genome assembled into a single large contig with a high degree of accuracy and at a much greater depth of coverage due to longer read lengths (Ferrari et al., 2013).

In this study, we present the complete nucleotide sequence and the organization of the chloroplast genome of *P. edulis*, the first report on the Passifloraceae family. We were able to localize genes, introns and intergenic spacers, as well as repetitive elements, and to compare the cpDNA of *P. edulis* with that of other of phylogenetically close species, searching

Abbreviations: AIC, Akaike information criterion; BAC, bacterial artificial chromosome; BES, BAC-end sequencing; BI, Bayesian inference; bp, base pair; cp genome, chloroplast genome; GTR, general time reversible model; IR, inverted repeat; IRA, inverted repeat A; IRB, inverted repeat B; LSC, large single copy; ML, maximum likelihood; NGS, next generation sequencing; SSC, small single copy; SSR, simple sequence repeat; TVM, transversional model.

¹<http://cnrgv.toulouse.inra.fr/Library/Passiflora>

²<http://www.pacificbiosciences.com/>

for syntenic regions and possible sequence rearrangements. Moreover, the order Malpighiales was investigated using the available entire chloroplast genomes of members of the four families that compose this order. Finally, a phylogenomic analysis was performed based on a set of chloroplast genes, with the aim of describing species relationships within the Fabids.

MATERIALS AND METHODS

Plant Material

The 'IAPAR-123' passion fruit (*P. edulis*) accession described in Carneiro et al. (2002) was used in the present study. The other complete cp genomes and cp gene sequences were downloaded from GenBank. A list of species and GenBank accession numbers are provided in the Supplementary Table S1.

Sequencing and Subsequent Assembly of *Passiflora edulis* Chloroplast DNA Using the Pacbio RS II Platform

The large-insert BAC library of *P. edulis* was constructed and maintained at the French Plant Genomic Resources Centre³. It was previously accessed using the BES approach and comparative genome mapping, and two clones (Pe69Q4G9 and Pe85Q4F4) were found to match the *Arabidopsis thaliana* chloroplast genome (Santos et al., 2014). In the present study, these clones were selected so that their entire inserts could be sequenced. The DNA was then isolated using the Nucleobond Xtra Midi Plus kit (Macherey-Nagel, Düren, Germany) following the manufacturer's instructions, after growing the bacterial clones on LB medium (100 mL) containing chloramphenicol as the selective marker (12.5 µg/mL).

Around 1.5 µg of each individual cpDNA were pooled together with other *P. edulis* BAC inserts (12 in total) for the construction of an SMRT library using the standard Pacific Biosciences (San Francisco, CA, USA) preparation protocol for 10 kb libraries. The pool was then sequenced in one SMRT Cell using the P4 polymerase in combination with C2 chemistry, following the manufacturer's standard operating procedures and using the Pacific Biosciences PacBio RS II platform. Sequencing was performed by GATC Biotech⁴.

The reads were assembled following a hierarchical genome assembly process HGAP workflow (Chin et al., 2013), and using the SMRT[®] Analysis (v2.2.0) software suite for HGAP implementation. Reads were first aligned by the PacBio long read aligner or BLASR (Chaisson and Tesler, 2012) against the complete genome of *Escherichia coli* strain K12 substrain DH10B (GenBank: CP000948.1). The *E. coli* reads, as well as low quality reads (minimum read length of 500 bp and minimum read quality of 0.80), were removed from the data set. Filtered reads were then preassembled to yield long, highly accurate sequences. To perform this step, smallest and longest reads were separated from each other to correct read errors by mapping single-pass reads to longest reads (seed reads), which represent the longest

portion of the read length distribution. Next, the sequences were filtered against vector (BAC) sequences, and the Celera assembler was used to assemble data and obtain draft assemblies. The last step of the HGAP workflow is performed in order to significantly reduce the remaining InDel and base substitution errors in the draft assembly. The Quiver algorithm was used for this purpose. It is a quality-aware consensus algorithm that uses rich quality scores (QV scores) embedded in Pacific Biosciences' bas.h5 files. Once the polished assembly was obtained, each BAC sequence was individualized by matching its paired-end sequences to the assembled sequences using BLAST. Read coverage was assessed by aligning the raw reads on the assembled sequences with BLASR.

Obtaining the Complete Chloroplast Genome

The sequences obtained from both inserts (Pe69Q4G9 and Pe85Q4F4) were aligned using ClustalX software (Larkin et al., 2007) to obtain a single contig. Specific primers were designed at the sequence ends of this contig in order to find out whether the circular chloroplast genome was complete. PCR reactions were then performed using a 9700 thermal cycler (Applied Biosystems, Foster City, CA, USA) in reaction mixtures containing 20 ng template DNA (*P. edulis* accession 'IAPAR-123'), 1× buffer, 1 mM MgCl₂, 0.2 mM of each dNTP, 0.3 µM of the forward and reverse primers, 1.2 U Go Taq Flex DNA polymerase (Promega, Madison, WI, USA), and ultra-pure water to bring the final volume up to 20 µL. The thermal profile for amplification was: 95°C for 5 min, 35 cycles at 95°C for 40 s, 55°C for 40 s and 72°C for 1 min, followed by a final 8 min incubation at 72°C. The amplified fragments were checked on 1% (w/v) agarose gel with a 100 bp molecular size standard Invitrogen (Carlsbad, CA, USA). The PCR product was purified using the Wizard[®] SV Gel kit and PCR Clean-Up System (Promega), and then used as a template for the sequencing reaction based on the Sanger method. It was subjected to capillary electrophoresis in the ABI Prism 3100 sequencer (Applied Biosystems). The sequence of the PCR product was aligned with the single contig to obtain the complete sequence of the cp genome.

Genome Annotation

The cp genome was preliminarily annotated using the DOGMA (Dual Organellar GenoMe Annotator) online program (Wyman et al., 2004), with default settings to identify coding sequences (CDS), rRNAs and tRNAs based on the Plant Plastid Code and BLAST homology searches, followed by manual corrections for start and stop codons, and intron positions. All tRNA genes were further confirmed using the tRNAscan-SE online search server. Pseudogenes were classified based on the loss of parts in their sequences or by the presence of internal stop codons. The circular genome map was designed by the GenomeVx program (Conant and Wolfe, 2008). Codon usage frequencies and the relative synonymous codon usage (RSCU) were calculated for all exons of the protein-coding genes using DAMBE 5 (Xia, 2013). Pseudogenes were not included in this analysis.

³http://cnrgv.toulouse.inra.fr/fr/library/genomic_resource/Ped-B-Flav

⁴<http://www.gatc-biotech.com>

Comparative Analysis of Chloroplast Genomes

To examine the expansion of the IR borders, the IR-LSC and IR-SSC boundaries with full annotations for the adjacent genes were manually analyzed across 11 sequenced species related to *P. edulis*, totalling 12 comparisons. These species are members of the families that compose the order Malpighiales: *P. edulis* (Passifloraceae), *P. trichocarpa* and *Salix purpurea* (Salicaceae), *Hevea brasiliensis*, *Manihot esculenta*, *Jatropha curcas* and *Ricinus communis* (Euphorbiaceae), *Hirtella physophora*, *Licania heteromorpha*, *Couepia guianensis*, *Parinari campestris* and *Chrysobalanus icaco* (Chrysobalanaceae) (Supplementary Table S1).

In addition, a multiple alignment with all available entirely sequenced cp genomes of Malpighiales species (22 in the total) (Supplementary Table S1), including *P. edulis*, was run in progressive Mauve v.2.4.0 (Darling, 2004). Briefly, this method identifies conserved genomic regions, rearrangements and inversions in conserved regions, and the sequence breakpoints of these rearrangements across multiple genomes.

Next, to validate the three inversions found in the passion fruit cp genome, a pair of primers was designed to anneal the 5' and 3' ends of the boundaries of each inversion. The amplicons of the expected size and the corresponding sequences should cover each border both downstream and upstream. PCR and Sanger sequencing were performed as described above.

Identification of Repeated Elements

REPuter (Kurtz et al., 2001) was used to identify direct and palindromic repeated elements, based on the following criteria: minimum repeat size ≥ 30 bp and sequence identities $\geq 90\%$ (Hamming distance equal to 3).

Microsatellites or SSRs that consist of tandemly arranged repeats of short DNA motifs (1–6 bp in length) were predicted using MISA (MIcroSATellite)⁵. This tool allows the identification and localization of microsatellites in genome sequences, including organelle genome sequences. The criteria used to search SSR motifs were set as follows: motifs between one and six nucleotides long, with a minimum repeat number defined as 10, 5, and 4 units for mono-, di-, and trinucleotide SSRs, respectively, and three units for each tetra-, penta-, and hexanucleotide SSRs.

Phylogenomic Studies

We performed two phylogenomic studies. The first was restricted to the Malpighiales based on the available entire chloroplast genomes of members of the four families that compose this order (22 species in total, Supplementary Table S1). In the second, a set of chloroplast genes was used to infer the relationships within the Fabids (42 species in total, Supplementary Table S1).

Entire chloroplast genomes of Malpighiales representing the families Passifloraceae, Salicaceae, Euphorbiaceae, and Chrysobalanaceae were used as the ingroup in phylogenomic comparisons, and the cp genome of *V. vinifera* (Vitaceae, Vitales) was used as the outgroup in order to root the phylogenetic tree. The data set consisting of 23 taxa was aligned at nucleotide

level in server-based program MAFFT version 7.221, using the FFT-NS-2 algorithm with default settings. To generate the alignment, inverted sequences detected in the cp genomes of *P. edulis* and *H. brasiliensis* were reversed and the respective positions adjusted. The raw alignment was manually corrected in BioEdit (Hall, 1999) and further processed in GBLOCKS 0.91b (Castresana, 2000) in order to remove regions containing gap positions, with a minimum block length of five, and maximum number of contiguous non-conserved positions of eight. The resulting alignment was analyzed in jModelTest software version 2.1.7 (Darriba et al., 2012) to determine the optimal model of molecular evolution and gamma rate heterogeneity using the AIC.

Maximum likelihood analysis was performed using PAUP version 4.0a146 (Swofford, 2002), based on the transversional (TVM) substitution model, gamma distribution of rate heterogeneity with five discrete categories (+G). To estimate the level of support for the ML topology, bootstrap analysis was performed on 1,000 replicates. BI was run in MrBayes, version 3.1.2 (Ronquist and Huelsenbeck, 2003) but based on the GTR +G model, the second model chosen by jModelTest taking the AIC values into account. The Markov chain Monte Carlo (MCMC) algorithm was run for 5,000,000 generations, sampling one tree every 100 generations. The first 25% of trees were discarded as burn-in to estimate the values of posterior probabilities. Convergence diagnostics were monitored on the basis of an average standard deviation of split frequencies below 0.01, potential scale reduction factor (PSRF) values close to 1.0, and ESS values above 200. Trees were visualized using FigTree version 1.4.0.

A set of 43 nucleotide sequences of homologous protein-coding chloroplast genes from 42 species representing all orders of the Fabids clade (Rosales, Fagales, Cucurbitales, Fabales, Malpighiales, Celastrales, Oxalidales, Zygophyllales) were used as the ingroup in the phylogenomic comparisons. *V. vinifera* (Vitaceae, Vitales) was chosen to serve as outgroup to produce a rooted tree. A list of the chloroplast gene sequence sources is provided in Supplementary Table S1.

First, each protein-coding gene sequence was aligned using ClustalW with default settings and the raw alignments manually corrected in program BioEdit (Hall, 1999) and further processed in GBLOCKS (Castresana, 2000), excluding gap positions from the data set, with a minimum block length of five, and a maximum number of contiguous non-conserved positions of eight. Next, all individual filtered alignments were concatenated into a single alignment matrix. Both conserved and variable nucleotide positions in the alignment matrix were analyzed in MEGA6 (Tamura et al., 2013).

jModelTest software version 2.1.7 (Darriba et al., 2012) was used to determine the optimal model of molecular evolution and gamma rate heterogeneity based on AIC. Both ML and BI were used to infer the phylogenomic relationships within the Fabids clade. ML analysis was performed using RAxML version 8.2.4 (Stamatakis, 2014). The GTR model of nucleotide substitution was selected for ML analysis, taking into account the gamma distribution of rate heterogeneity with five discrete categories (+G). To estimate the support of the ML topology, a

⁵<http://pgrc.ipk-gatersleben.de/misa/>

bootstrap analysis was performed on 1,000 replicates. BI was run on MrBayes software, version 3.1.2 (Ronquist and Huelsenbeck, 2003) with the GTR +G model. The MCMC algorithm ran for 5,000,000 generations, sampling one tree every 100 generations. The first 25% of trees were discarded as burn-in to estimate the values of posterior probabilities. Convergence diagnostics were monitored on the basis an average standard deviation of split frequencies below 0.01, PSRF values close to 1.0, and ESS values above 200. Trees were visualized using FigTree version 1.4.0.

RESULTS AND DISCUSSION

Data Output from the PacBio RS II Platform and Assembly of Chloroplast Genome Sequences

Following extraction of reads containing only chloroplast genome sequence data and subsequent error correction, 2,340 PacBio RS reads from the clone insert Pe69Q4G9 were recovered, ranging from 500 to 22,458 bp, and containing a total of 94,052 bp assembled into a contig. The average depth of coverage of the consensus sequence was 96 \times and the average GC content was 37%. The final quality of the assembly corresponded to a nominal QV of 48.48 (approximately 99.999% accuracy). Similarly, 4,972 reads from the clone insert Pe85Q4F4 were recovered, ranging from 500 to 26,035 bp, and containing a total of 91,155 bp assembled into a contig. The average depth of coverage of the consensus sequence was 172 \times and the average GC content was 38%. The final quality of the assembly corresponded to a QV of 48.57.

It is worth noting the usefulness of the HGAP workflow as a solution for successfully resolving long repeat regions, as already pointed out by Chin et al. (2013). The high quality sequence data generated by the PacBio RS II Platform, added to its capability to assemble long reads, allowed us to obtain a single contig for each clone insert and then the complete cp genome of *P. edulis*. Both contigs were aligned and merged into a single long contig of 151,362 bp with an overlapping region of 33,848 bp. The amplification reaction using the primer pair designed to anneal at the extremities of the large contig and genomic DNA from *P. edulis* accession 'IAPAR-123' as a template produced an amplicon of 325 bp, which was subsequently sequenced and aligned with the long contig sequence. Thus, a 44-nucleotide sequence was obtained and added for closing the gap to produce the complete sequence of the circular molecule of 151,406 bp. The chloroplast genome sequence of *P. edulis* has been submitted to the NCBI GenBank and has received the accession number KX290855.

The chloroplast genomes of *Potentilla micrantha* (Ferrarini et al., 2013) and *Aconitum barbatum* var. *puberulum* (Chen et al., 2015) were entirely sequenced using the PacBio RS II Platform, and both studies described the advantages of using this method. For instance, Ferrarini et al. (2013) highlight the generation of a single contig covering the entire cp genome of *P. micrantha*, whereas Illumina HiSeq2000 sequencing data have lower genome coverage (some regions have zero or very low

coverage) and the resulting assembly consisted of seven contigs. Similarly, Chen et al. (2015) emphasize the low level of errors (~0.0027%) in PacBio reads, since after applying the necessary filter, the result is an error-corrected consensus read with a higher intra-molecular accuracy. Interestingly, even in the absence of any other biological information on the target species, the authors stated that it took less than half an hour to finish the genome assembly step, confirming that long read lengths are superior, especially for *de novo* assemblies.

Organization of the *P. edulis* Chloroplast Genome

Chloroplast genome annotation resulted in the identification of 105 unique genes, including 30 tRNAs, 4 rRNAs and 71 protein coding genes (Table 1). The molecule has a typical quadripartite structure characterized by the presence of two copies of an IRA and IRB each of 26,154 bp (34.6% in total) separating a SSC region of 13,378 bp (8.8%) and a LSC region of 85,720 bp (56.6%). Genes arising from duplication events created perfect IRs, each containing the same 16 genes, totalling 120 genes in the complete chloroplast DNA molecule (Figure 1).

Protein coding genes constitute 37.3% of the chloroplast genome (75 genes, totalling 56,428 bp), in addition to tRNA and rRNA coding genes that constitute 1.8% (37 genes totalling 2,801 bp) and 6.0% (eight genes totalling 9,048 bp) respectively of the chloroplast genome. Introns constitute 10.7% (totalling 16,155 bp), while intergenic regions and pseudogenes form the remaining 44.2% (totalling 66,974 bp). The amount of GC was estimated at 37%.

tRNA genes are spread throughout the genome and encode the 20 amino acids incorporated into proteins. One is present in the SSC region, 21 in the LSC, and 7 in the IRs, which also have 5 protein-coding genes and 4 rRNA genes. A total of 18,799 codons represent the coding capacity of the 71 protein-coding genes (Table 2). The most frequent amino acid was found to be leucine (2,026 codons corresponding to 10.8% of total DNA) and the least frequent cysteine (218 codons corresponding to 1.2% of total DNA).

Fifteen unique genes (nine protein- and six tRNA-coding genes) have introns, and two introns were found in only one gene, *ycf3*. The largest intron occurs in the *trnK-UUU* gene (2,524 bp) in which the *matK* gene (1,506 bp) is inserted, and the smallest occurs in the *rps12* gene (537 bp). The number of introns identified in the cp genome of *P. edulis* is similar to that of other species. For instance, in *S. purpurea* (Wu, 2015) there are 17 intron-containing genes, three of them containing two introns, including *ycf3*. Similar figures have been found for *R. communis* (Rivarola et al., 2011), whereas in *H. brasiliensis* (Tangphatsornruang et al., 2011), 23 introns are inserted in 22 cp genes.

The *matK* gene was found in the intronic region of the *trnK-UUU* gene in *P. trichocarpa* (Tuskan et al., 2006), *H. brasiliensis* (Tangphatsornruang et al., 2011), and *M. esculenta* (Daniell et al., 2008). This gene encodes the unique maturase found in the plastid genomes of land plant species, but it does not contain the reverse transcriptase domain and therefore is not

TABLE 1 | Gene content of the *Passiflora edulis* chloroplast genome according to respective categories.

Category	Gene
Subunits of photosystem I	<i>psaA, psaB, psaC, psal, psaj</i>
Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbl, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
Subunits of cytochrome b/f complex	<i>petA, petB^a, petD^a, petG, petL, petN</i>
Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
Large subunit of rubisco	<i>rbcl</i>
Subunits of NADH-dehydrogenase	<i>ndhA^a, ndhB^{a,b}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
Proteins of large ribosomal subunit	<i>rpl2^{a,b}, rpl14, rpl16^a, rpl23^b, rpl32, rpl33, rpl36</i>
Proteins of small ribosomal subunit	<i>rps2, rps3, rps4, rps8, rps11, rps12^{a,b,c}, rps14, rps15, rps18, rps19^b</i>
Subunits of RNA polymerase	<i>rpoA, rpoB, rpoC1^a, rpoC2</i>
Cytochrome c biogenesis	<i>ccsA</i>
Maturase	<i>Matk</i>
Protease	<i>clpP</i>
Envelope membrane protein	<i>cemA</i>
Conserved hypothetical genes	<i>ycf3^a, ycf4</i>
Ribosomal RNAs	<i>rrn4.5^b, rrn5^b, rrn16^b, rrn23^b</i>
Transfer RNAs	<i>trnA-UGC^{a,b}, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnM-CAU, trnG-UCC^a, trnG-GCC, trnH-GUG, trnI-CAU^b, trnI-GAU^{a,b}, trnK-UUU^a, trnL-CAA^b, trnL-UAA^a, trnL-UAG, trnM-CAU, trnN-GUU^b, trnP-UGG, trnQ-UUG, trnR-ACG^b, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC^b, trnV-UAC^a, trnW-CCA, trnY-GUA</i>

^aIntron-containing gene; ^btwo gene copies in the IRs; ^cgene divided into two independent transcription units.

able to promote intron mobility. The *clpP* gene has no intron in *P. edulis*, although it does have two introns in the following conordinal-related species: *P. trichocarpa* (Tuskan et al., 2006), *S. purpurea* (Wu, 2015), *H. brasiliensis* (Tangphatsornruang et al., 2011), *M. esculenta* (Daniell et al., 2008), *J. curcas* (Asif et al., 2010), and *R. communis* (Rivarola et al., 2011). Similar intron losses have already been reported in legumes, almost exclusively in the clade known as the IR-lacking clade (IRLC) (Jansen et al., 2008). Coincidentally, the *clpP* gene is located at the beginning of first the inversion found in the *P. edulis* cp genome (**Figure 1**), which may lead to intron losses.

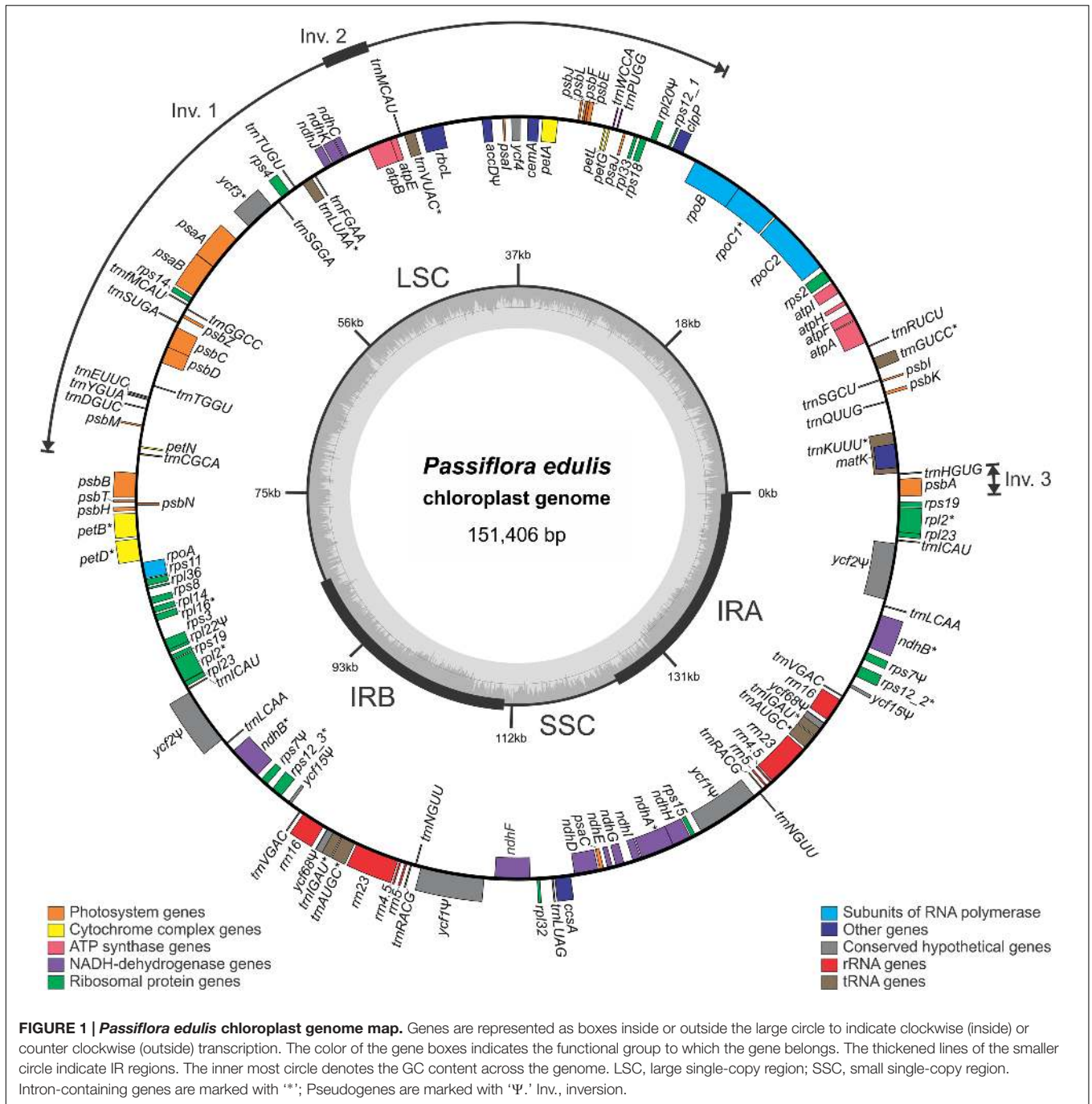
Group I and II introns have been derived from mobile genetic elements, which explains why they are lost or gained in the evolution of chloroplast genomes (Barkan, 2004). Both cyanobacteria and algae, as well as land plant species, share a single group I intron in the *trnL-UAA* gene, which is therefore considered the more ancestral (Simon et al., 2003). Usually, the *atpF* gene shows a conserved group II intron; however, it is absent in the *atpF* gene of *P. edulis*. This loss was also reported in other species of *Passiflora* (Jansen et al., 2007). Daniell et al. (2008) have suggested an association between C-to-T substitutions (at nt position 92) and the loss of the intron in *M. esculenta* and in other *atpF* gene sequences of Malphigiales, implying that recombination between an edited mRNA and the *atpF* gene may be a possible mechanism for this intron loss.

Two protein-coding genes show alternative initiation codons. The 'GTG' triplet was found in the *rps19* and *ndhD* genes. This triplet was also found in the *rps19* gene of *J. curcas* (Asif et al., 2010), *H. brasiliensis* (Tangphatsornruang et al., 2011), and *Cynara cardunculus* (Curci et al., 2015). The *rps12* gene is a *trans*-spliced gene consisting of three exons: the first exon (5'-*rps12*) is located in the LSC region, far from the other two located in the IRs. This kind of organization was observed in *P.*

trichocarpa, *S. purpurea*, *H. brasiliensis*, *M. esculenta*, *J. curcas*, and *R. communis*. Interestingly the *infA* gene, which codes for translation initiation factor 1, and the *rps16* gene, which codes for a S16 ribosomal protein are absent in the *P. edulis* cp genome, as previously reported in other species of *Passiflora* (Jansen et al., 2007). These genes were also lost or are non-functional in related species *M. esculenta* (Daniell et al., 2008), *J. curcas* (Asif et al., 2010) and in seven species of Salicaceae (Wu, 2015). It is worth noting that Passifloraceae are known to have chloroplast gene and intron losses (Hansen et al., 2006; Jansen et al., 2007; Daniell et al., 2008).

Eight pseudogenes were identified, five in the IRs and three in the LSC region. The nucleotide sequences of the *rpl22*, *ycf1*, and *ycf2* pseudogenes are smaller in length compared to those of the functional genes identified in other genomes. We found a truncated portion of the *rpl22* pseudogene that has also been identified in *Passiflora biflora*, *P. quadrangularis*, and *P. cirrhiflora* (Jansen et al., 2011), but no studies have been performed to demonstrate whether this partial copy is functional or whether there is an *rpl22* functional copy in the nucleus. There is evidence that this gene was exported to the nucleus in some Rosids (*Castanea*, *Prunus*, and *Theobroma*) and *Fagaceae*, and it is assumed that the transfer resulted from two independent events. An additional event may have occurred in *Passiflora* (Jansen et al., 2011).

Only one of the two copies of *ycf1* is a pseudogene in the *H. brasiliensis*, *M. esculenta*, *J. curcas*, and *C. cardunculus* cp genomes, as most of their sequences have been lost. Interestingly, in *P. edulis* both copies lost parts of their sequences, representing non-functional *ycf1* genes (**Figure 1**). On the other hand, the *rpl20*, *rps7*, *ycf15*, and *ycf68* pseudogenes have 5, 1, 3, and 3 internal stop codons, while the *accD* gene was found to have repetitive elements at the 5' end.



Comparative Analysis of Chloroplast Genomes

As commonly done in studies of border expansion in the IRs region, the IR-LSC and IR-SSC boundaries with full annotations for the adjacent genes were reexamined across 11 sequenced species closely related to *P. edulis*. Note that the *rps15* gene of *P. edulis* is located at the end of the SSC region, expanding 20 bp toward the IRA region. It is substituted in other species by the *ycf1* gene, i.e., this gene spans SSC and IRA and there is a pseudogene copy in the IRB region. However, in *P. edulis*, both copies of the *ycf1* gene are of the same size and are full in

IR regions (Figure 2). In addition, the *psbA* gene is the first in the LSC region due to a small inversion that is further discussed below.

In accordance with the genome comparison results and based on the 22 members of Malpighiales with available chloroplast genomes, *P. edulis* shows a large inversion of 46,151 bp in the LSC region, between the genes *ClpP* and *trnC-GCA*, and a second smaller inversion of 3,765 bp between the genes *trnM-CAU* and *atpB*, located in the medial region of the first inversion. A third inversion of

TABLE 2 | Codon usage and codon-anticodon recognition pattern of the *Passiflora edulis* chloroplast genome.

Codon	Amino acid	Number	RSCU ^a	% ^b	tRNA	Codon	Amino acid	Number	RSCU ^a	% ^b	tRNA
GCU	A	513	1.727	43.2		CCA	P	225	1.108	27.7	<i>trnP-UGG</i>
GCG	A	150	0.505	12.6		CCC	P	158	0.778	19.5	
GCC	A	206	0.694	17.3		CCU	P	325	1.601	40.0	
GCA	A	319	1.074	26.9	<i>trnA-UGC</i> ^c	CCG	P	104	0.512	12.8	
UGU	C	162	1.486	74.3		CAA	Q	523	1.59	79.5	<i>trnQ-UUG</i>
UGC	C	56	0.514	25.7	<i>trnC-GCA</i>	CAG	Q	135	0.41	20.5	
GAU	D	524	1.53	76.5		AGA	R	302	1.529	28.5	<i>trnR-UCU</i>
GAC	D	161	0.47	23.5	<i>trnD-GUC</i>	AGG	R	93	0.471	8.8	
GAG	E	202	0.441	22.0		CGA	R	250	1.508	23.6	
GAA	E	715	1.559	78.0	<i>trnE-UUC</i>	CGC	R	80	0.483	7.6	
UUU	F	716	1.351	67.5		CGG	R	88	0.531	8.3	
UUC	F	344	0.649	32.5	<i>trnF-GAA</i>	CGU	R	245	1.478	23.2	<i>trnR-ACG</i>
GGU	G	471	1.323	33.1		AGC	S	107	0.542	7.8	<i>trnS-GCU</i>
GGG	G	233	0.654	16.4		AGU	S	288	1.458	21.1	
GGC	G	169	0.475	11.9		UCA	S	256	1.053	18.7	<i>trnS-UGA</i>
GGA	G	551	1.548	38.7	<i>trnG-GCC</i>	UCC	S	223	0.918	16.3	<i>trnS-GGA</i>
CAC	H	102	0.462	23.1	<i>trnH-GUG</i>	UCG	S	114	0.469	8.3	
CAU	H	340	1.538	76.9		UCU	S	379	1.56	27.7	
AUU	I	834	1.515	50.5		ACC	T	170	0.688	17.2	<i>trnT-GGU</i>
AUA	I	506	0.919	30.6	<i>trnI-CAU</i>	ACA	T	308	1.247	31.2	<i>trnT-UGU</i>
AUC	I	311	0.565	18.8	<i>trnI-GAU</i> ^c	ACG	T	97	0.393	9.8	
AAA	K	704	1.529	76.4	<i>trnK-UUU</i> ^c	ACU	T	413	1.672	41.8	
AAG	K	217	0.471	23.6		GUU	V	388	1.399	35.0	
CUA	L	278	1.214	13.7	<i>trnL-UAG</i>	GUG	V	145	0.523	13.1	
CUC	L	120	0.524	5.9		GUC	V	139	0.501	12.5	<i>trnV-GAC</i>
CUG	L	97	0.424	4.8		GUA	V	437	1.576	39.4	<i>trnV-UAC</i> ^c
CUU	L	421	1.838	20.8		UGG	W	327	1	100	<i>trnW-CCA</i>
UUA	L	735	1.324	36.3	<i>trnL-UAA</i> ^c	UAC	Y	124	0.363	18.2	<i>trnY-GUA</i>
UUG	L	375	0.676	18.5	<i>trnL-CAA</i>	UAU	Y	559	1.637	81.8	
AUG	M	464	1	100	<i>trn(f)M-CAU</i>	UGA	*	15	0.584	19.5	
AAC	N	186	0.464	23.2	<i>trnN-GUU</i>	UAG	*	24	0.935	31.2	
AAU	N	615	1.536	76.8		UAA	*	38	1.481	49.4	

^aRelative synonymous codon usage; ^bcodon frequency (in %) per amino acid; ^cintron-containing tRNA genes. *Stop codon.

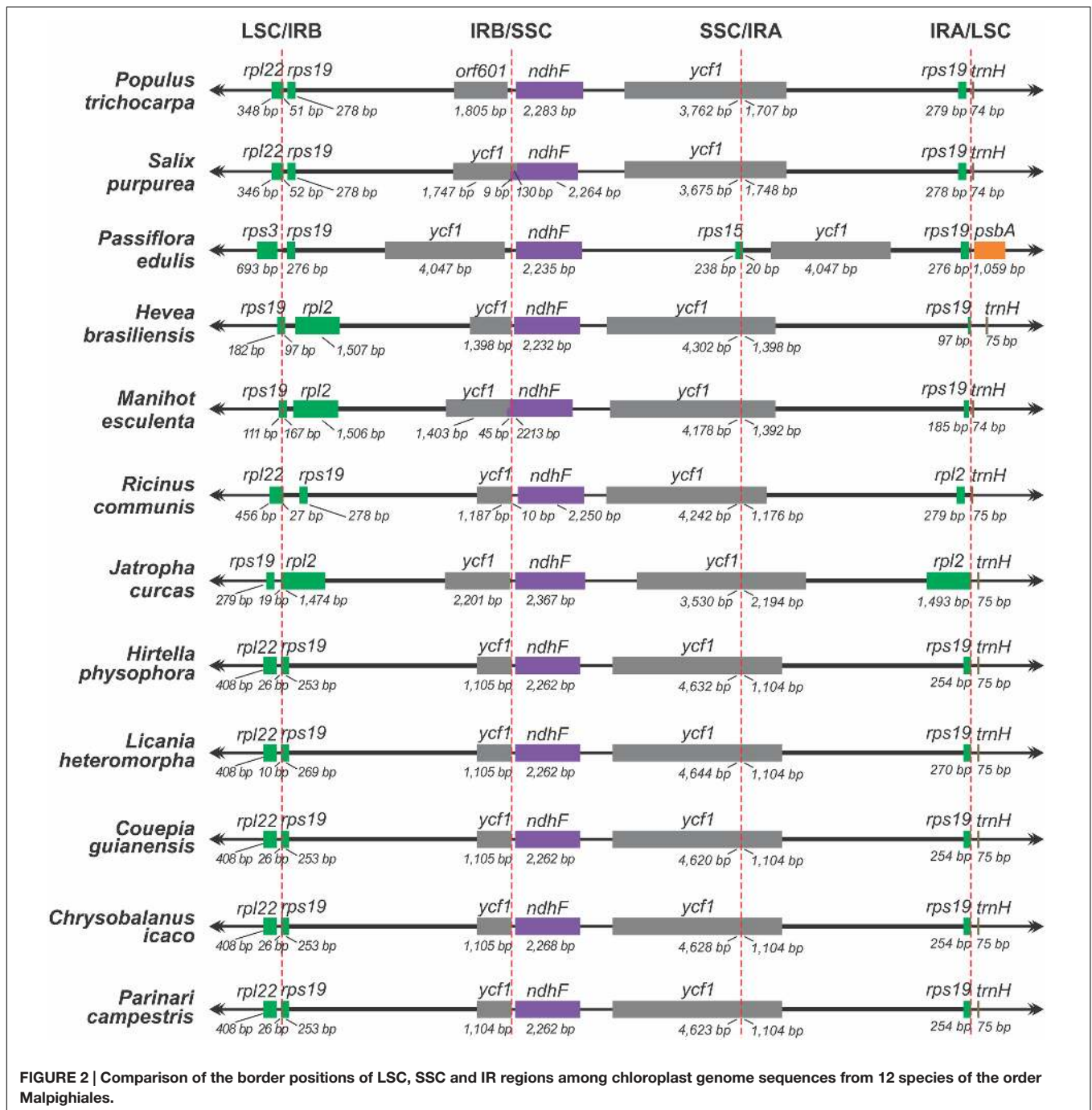
1,631 bp is located at the beginning of the LSC region containing the *psbA* and *trnH-GUG* genes (Figures 1, 3). All these genomic features are shown in the sequence alignment results for 12 species representing each genus of the Passifloraceae, Salicaceae, Euphorbiaceae, and Chrysobalanaceae families (Figure 3).

It is important to point out that the occurrence of these inversions in the cpDNA of *P. edulis* was confirmed by respective amplification reactions and subsequent Sanger sequencing of both boundaries (downstream and upstream), generating the expected sized amplicons and the corresponding sequences (Supplementary Figure S1). These sequence inversions could be typical of the *Passiflora* genus, but at present, this is pure speculation. In accordance with the alignment results (Figure 3), all have the same order and orientation of syntenic blocks, except for *P. edulis* and *H. brasiliensis*, in which there is an inversion of 30,000 bp in the LSC region, between the *trnE-UUC* and *trnR-UCU* genes (Tangphatsornruang et al., 2011). In conclusion, the complete chloroplast genome of *P. edulis*

differs from the others because of three rearrangement events that resulted in inversions of gene block order (Figure 3). Otherwise, chloroplast genomes tend to be conserved and perfectly collinear, especially in the same plant family, as occurs in Salicaceae (Wu, 2015) and Chrysobalanaceae (Malé et al., 2014), also members of the Malpighiales order. The existence of rearrangements in segments of cp genomes may be useful as phylogenetic markers within genera or even within families, becoming a potential tool for understanding the evolution of plant species. Therefore, the complete sequencing of new chloroplast genomes will allow a higher accuracy in evolutionary studies of the inversions in the genus *Passiflora*.

Analysis of Repeated Elements

We were able to identify 36 repetitive elements, all in the LSC region. These repeats were found predominantly in intergenic regions. However, each of the members of a particular repeat was identified in the coding sequences of the *psaA* and *psaB*



genes respectively; one member of other repeat is located at the beginning of the *psbI* gene sequence and the other in an intergenic region. No introns were found to contain repeated elements, and approximately 60% of repetitive elements are within a 2,513-bp region between the *accD* pseudogene and the *rbcL* gene, indicating that these elements are distributed in a peculiar arrangement in *P. edulis*. The abundance of repeated elements in this region might possibly have changed the nucleotide sequence of the *accD* pseudogene, rendering it non-functional. The repeat unit length ranged from 34 to 178 bp,

and each repeat showed two copies. Forward (or direct) repeats and forward-tandem repeats (when the repeats are presented immediately one beside the other, there occurring sometimes an overlapping of both repeat units) were predominant, and only three were found to be palindromes (or reverse-complemented) (Table 3). Repetitive sequences are substrates for recombination and cp genome rearrangements (Milligan et al., 1989), and the number and distribution of these sequences vary from one species to another. For instance, the cp genome of *J. curcas* contains 72 repetitive sequences that are distributed in intergenic regions,



introns and coding sequences (Asif et al., 2010), whereas in *V. vinifera*, 36 repetitive sequences were found, one in the coding regions of the *psaA* and *psaB* genes (Jansen et al., 2006), as in *P. edulis*, but 58% were palindromes and 12 exist in the *ycf2* gene.

We found 85 SSRs (microsatellites), ranging in size from 10 to 106 bp, 67 of which (79%) consisted exclusively of A/T. In terms of SSR motifs, we found 43 (50.6%) mono-, 21 (24.7%) di-, 7 (8.2%) tri-, 10 (11.8%) tetra-, and 4 (4.7%) hexanucleotides (Figure 4). SSRs were found mainly in non-coding regions, 68 in intergenic regions and five in pseudogenes/intronic regions. Of the 12 SSRs found in gene sequences, two interrupted elements composed of trinucleotides (GAC/ACG) were found in the *clpP*

gene. In addition, the *ndhA* gene contained one mono-(A) and one tetranucleotide (AAAT), the latter also seen in the *ndhD* gene sequence. Genes *ndhF*, *rpoB*, and *rpoC1* contained one mononucleotide each (T or A), whereas *rpoC2* contained four SSRs, three mono- (T), and one dinucleotide (AT).

Simple sequence repeats were widely distributed throughout cpDNA molecules. In *P. edulis*, 55 SSRs were found in LSC, 6 in SSC, 10 in the IRA region and 14 in IRB. All sequenced cpDNAs have been reported to contain SSRs and variation is said to occur within the species. This is why these sequences make good genetic markers in population studies and to estimate the relationships between plant taxa (Provan et al., 2001; Grassi et al., 2002; Melotto-Passarini et al., 2011).

TABLE 3 | Type, location, and size (in bp) of repeated elements found in the *Passiflora edulis* chloroplast genome (IGS, intergenic spacer; Ψ , pseudogene).

Type	Location	Size (in bp)
Forward-tandem	IGS: <i>matK-psbK</i>	43
Forward	IGS: <i>psbK-psbl, psbl</i>	49
Palindrome	IGS: <i>psbl-atpA</i>	47
Forward-tandem	IGS: <i>psbl-atpA</i>	44
Palindrome	IGS: <i>psbl-atpA</i>	34
Forward	IGS: <i>rpl33-psaJ, psal-accD</i>	47
Forward	IGS: <i>accD Ψ-rbcL</i>	66
Forward	IGS: <i>accD Ψ-rbcL</i>	58
Forward	IGS: <i>accD Ψ-rbcL</i>	54
Forward-tandem	IGS: <i>accD Ψ-rbcL</i>	41
Forward-tandem	IGS: <i>accD Ψ-rbcL</i>	73
Forward	IGS: <i>accD Ψ-rbcL</i>	61
Forward-tandem	IGS: <i>accD Ψ-rbcL</i>	74
Forward	IGS: <i>accD Ψ-rbcL</i>	39
Forward	IGS: <i>accD Ψ-rbcL</i>	46
Forward	IGS: <i>accD Ψ-rbcL</i>	56
Forward	IGS: <i>accD Ψ-rbcL</i>	49
Forward	IGS: <i>accD Ψ-rbcL</i>	67
Forward	IGS: <i>accD Ψ-rbcL</i>	46
Forward	IGS: <i>accD Ψ-rbcL</i>	65
Forward	IGS: <i>accD Ψ-rbcL</i>	46
Forward	IGS: <i>accD Ψ-rbcL</i>	43
Forward-tandem	IGS: <i>accD Ψ-rbcL</i>	58
Forward-tandem	IGS: <i>accD Ψ-rbcL</i>	124
Forward-tandem	IGS: <i>accD Ψ-rbcL</i>	99
Forward-tandem	IGS: <i>accD Ψ-rbcL</i>	74
Forward	IGS: <i>accD Ψ-rbcL</i>	49
Palindrome	IGS: <i>rbcL-atpE</i>	62
Forward-tandem	IGS: <i>ycf3-psaA</i>	41
Forward-tandem	IGS: <i>ycf3-psaA</i>	57
Forward	<i>psaA, psaB</i>	79
Forward-tandem	IGS: <i>psaB-rps14</i>	44
Forward-tandem	IGS: <i>petN-psbB</i>	71
Forward-tandem	IGS: <i>petN-psbB</i>	178
Forward	IGS: <i>petN-psbB</i>	117
Forward	IGS: <i>petN-psbB</i>	56

Phylogenomic Studies

A comparison of the *P. edulis* cp genome to those of the 22 species representing the four families of the Malpighiales order showed that each of the families (Passifloraceae, Salicaceae, Euphorbiaceae, and Chrysobalanaceae) constitutes a monophyletic group (Supplementary Table S1). The multiple-plantome alignment of the complete cp genome sequence was 208,695 bp in length. Removal of alignment columns containing gaps reduced the alignment length to 118,724 bp (nucleotide positions).

Both maximum likelihood and BI-based methods resulted in a similar tree (Figure 5), and all species were clustered in Malpighiales, in conformity with the APG II system. The nodes resulting from ML analysis were consistently supported

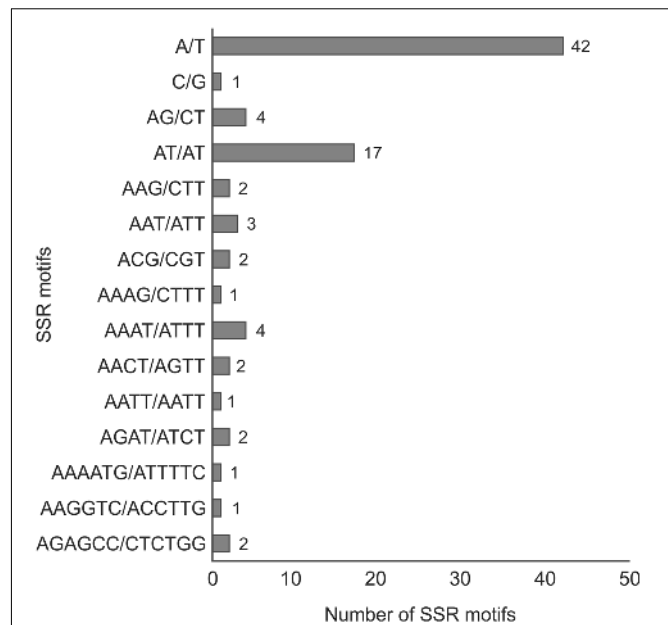
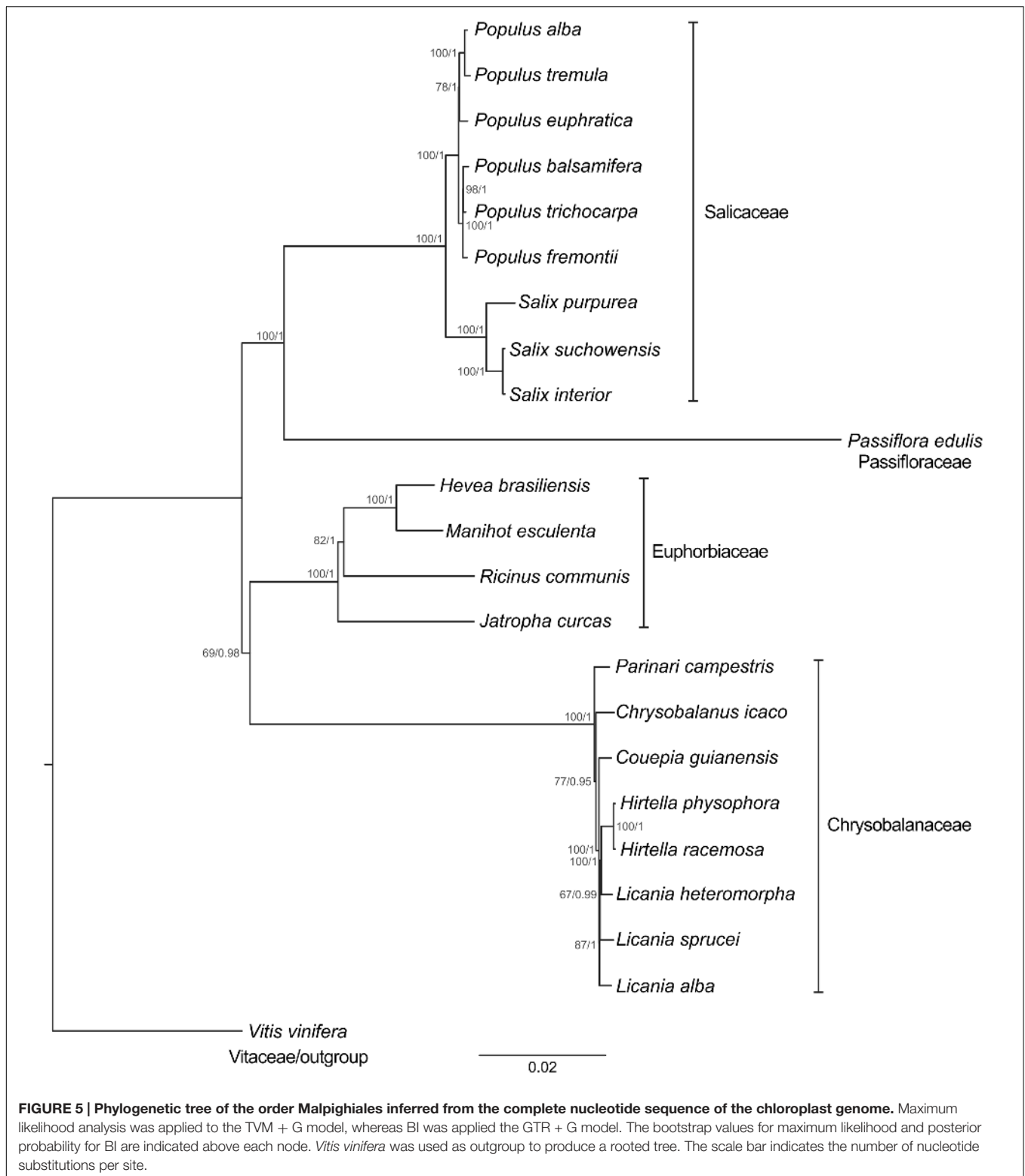


FIGURE 4 | The number of SSR motifs found in the *P. edulis* chloroplast genome, taking into account sequence complementarities. The criteria used to search SSR motifs were set as follows: motifs between one and six nucleotides long, with a minimum repeat number defined as 10, 5, and 4 units for mono-, di-, and trinucleotide SSRs, respectively, and three units for each tetra-, penta-, and hexanucleotide SSRs.

in the tree. The bootstrap values at most of the nodes (15 of 20 nodes) were $\geq 85\%$. In the BI the value of the average standard deviation of split frequencies was 0.0001 after 5,000,000 generations. The PSRF values were close to 1.0 (ranging between 0.999 and 1.000), and the ESS values were above 200 (ranging between 4,909 and 29,555) for all the parameters. All these values indicated that the analysis has reached convergence. The nodes resulting from BI analysis were also highly supported (PP = 1 for 17 nodes). *P. edulis* has been placed near the Salicaceae (*Salix* and *Populus* species) with strong support (BS = 100%; PP = 1), but distant from members of the Chrysobalanaceae, that in turn are closer to the Euphorbiaceae members studied herein.

In Malpighiales, a phylogeny based on plastid, mitochondrial and nuclear gene regions (Wurdack and Davis, 2009) and a recent phylogenomic approach based on a set of 82 chloroplast genes (Xi et al., 2012) indicated the relationship between Passifloraceae and Salicaceae. According to Wurdack and Davis (2009), for instance, most of its members share parietal placentation. Our results confirm this association and the potential of complete chloroplast genome sequences to infer evolutionary relationships. Malpighiales is an order that underwent rapid basal radiation. Therefore the use of large molecular data sets to identify several phylogenetically informative sites is an important step in deducing its course of evolution.

The Salicaceae species were clustered into two clades, both highly supported (BS = 100%; PP = 1). The most related were *S. purpurea* and *S. suchowensis* (BS = 100%; PP = 1), but



positioned at some distance from those species is *S. interior* (BS = 100%, PP = 1). *Populus* species were split into two groups, the first incorporating *Populus alba*, *P. tremula*, and *P. euphratica* (BS = 78%; PP = 1), and the second *P. balsamifera*, *P. trichocarpa*,

and *P. fremontii* (BS = 100%; PP = 1). The placement of *Salix* and *Populus* species is similar to that recently shown in the phylogenetic tree built from the cp DNA sequences of seven Salicaceae species (Wu, 2015).

With reference to Euphorbiaceae, *M. esculenta* was placed near to *H. brasiliensis* (BS = 100%; PP = 1), similar to the findings published by Xi et al. (2012) after examining a set of 82 chloroplast genes. Our findings placed *J. curcas* as the species most distant from the other Euphorbiaceae (BS = 100%; PP = 1), occupying an onward position relative to *R. communis* (BS = 82%; PP = 1). The placement of these species has weak support in two previous topologies based on 62 (Su et al., 2014) and 60 chloroplast genes (Kong and Yang, 2016). Remarkably, our study resolved the node positions in the Euphorbiaceae with strong support in both analyses (ML and BI). These findings also demonstrate the significance of using complete cp genomes in phylogeny reconstructions, since not only the genes but also non-coding sequences are examined. These sequences are informative, particularly at low taxonomic levels, due to their rapid evolution. In contrast, protein-coding genes evolve at a relatively slow rate (Folk et al., 2015).

Regarding the Chrysobalanaceae, the placement of *Licania heteromorpha* near to the species of *Hirtella* with strong support (BS = 100%; PP = 1) makes this genus paraphyletic. *Parinari campestris* was the most basal species in this family and this position is consistent with the findings of Bardon et al. (2013), based on six chloroplast markers and one nuclear marker. Moreover, the placement of the eight Chrysobalanaceae species with strong support is consistent with the phylogenetic hypothesis proposed by Malé et al. (2014) analyzing the complete plastid DNAs of eight species of this tropical family.

It is important to emphasize that our phylogenomic study is the first to take into account all the complete chloroplast genomes of the Malpighiales taxa available in databanks. Malpighiales includes 40 families, which are poorly represented in sequence databanks. Consequently, a limited taxonomic sampling of the order was used in the analysis and this may have contributed to the high support of the nodes in the phylogenetic trees obtained. Therefore, it is not possible to reach any definitive conclusions concerning the monophyly or position of the families within the order without examining the chloroplast genome sequences of members of the other families. However, our study lends support not only to the utility of complete chloroplast genomes to infer phylogenetic relationships, but as well as the enormous utility of long sequence reads from PacBio sequencing and assembly for generating very highly accurate biological information.

Subsequently, the complete nucleotide sequences of 43 chloroplast protein-coding genes of 42 species (including *P. edulis*) that belong to the eight orders of the Fabids clade were compared to reconstruct the phylogenomic relationships based on the GTR +G model (Supplementary Table S1), with *V. vinifera* (Vitales: Rosidae) as the outgroup. After alignment filtering and concatenating the nucleotide sequences into a single matrix, a total of 31,193 nucleotide positions were analyzed. Conservation analysis of positions was then performed and it is estimated that 61.65% correspond to conserved sites and 38.35% to variable sites. To summarize, these rates were estimated in Mega, a program that identifies a site as constant only if at least two sequences contain unambiguous nucleotides. In contrast, a variable site contains at least two types of nucleotide.

The method of analysis (ML or BI) had no substantial effect on the resulting trees. Their topologies were found to be highly similar (Figure 6). The accuracy of the inferred species' phylogeny is strongly supported by the stability of the main clades obtained using different phylogenetic methods (ML and BI), and the best-scoring phylogenomic tree is shown in Figure 6. The nodes resulting from ML analysis were consistently supported in the tree (41 nodes in total). The bootstrap values at most of the nodes were higher than 85% (34 of 41), reaching 100% in some cases (28 nodes). In the BI the value of the average standard deviation of split frequencies was 0.0003 after 5,000,000 generations. The PSRF values were close to 1.0 (range between 0.999 and 1.000), and the ESS values were above 200 (range between 2,564 and 23,705) for all the parameters. All these values indicate that the analysis has reached convergence. The nodes resulting from BI analysis were also well-supported (PP = 1 for 31 nodes).

The Fabids consist of eight orders, which are all represented in this study and have been found to be monophyletic. For instance, *Bulnesia arborea* was the most distant species of the ingroup, thus placing the order Zygophyllales basal to the Fabids clade. This position of Zygophyllales has been previously reported (Wang et al., 2009).

Moreover, we were able to recognize two major monophyletic subgroups with strong support (BS = 100%; PP = 1). The first subgroup includes the orders Rosales, Fagales, Cucurbitales, and Fabales, known as the nitrogen-fixing clade, and the second consists of the orders Malpighiales, Celastrales, Oxalidales known as the COM clade. There is a trend in the literature toward classifying these subgroups as monophyletic (Wang et al., 2009; Su et al., 2014; Kong and Yang, 2016). Although there is morphological divergence among the species comprising the nitrogen-fixing clade, this capacity is shared only by angiosperms that belong to the orders Rosales, Fagales, Cucurbitales, and Fabales (Svistonoff et al., 2013).

The members of the order Oxalidales (*Averrhoa carambola* and *Oxalis latifolia*) were placed as sister to a grade comprising the members of the order Celastrales (*Elaeodendron orientale* and *Euonymus americanus*) with strong support (BS = 100%; PP = 1). Thus, our studies indicate that Malpighiales shares more homologies with Celastrales than with Oxalidales, based on both methods used to infer species relationships within the Fabids. Previously, Hilu et al. (2003) reported a feasible relationship among the orders of the COM clade, based on sequence alignments of *matK*, a cp gene, but with weak support using maximum parsimony (BS = 60%). According to Sun et al. (2015), positioning the orders within the COM clade remains a great challenge. However, our results confirm other findings, for instance placing Celastrales nearer to Malpighiales than Oxalidales (Zhang and Simmons, 2006; Bell et al., 2010). The same scenario was described by Matthews and Endress (2005) observing floral structures and describing their implications for systematics in Celastrales. The use of a large set of genes is important for defining genomic relationships; otherwise the results could be interpreted as relating to the evolutionary course of particular genes.

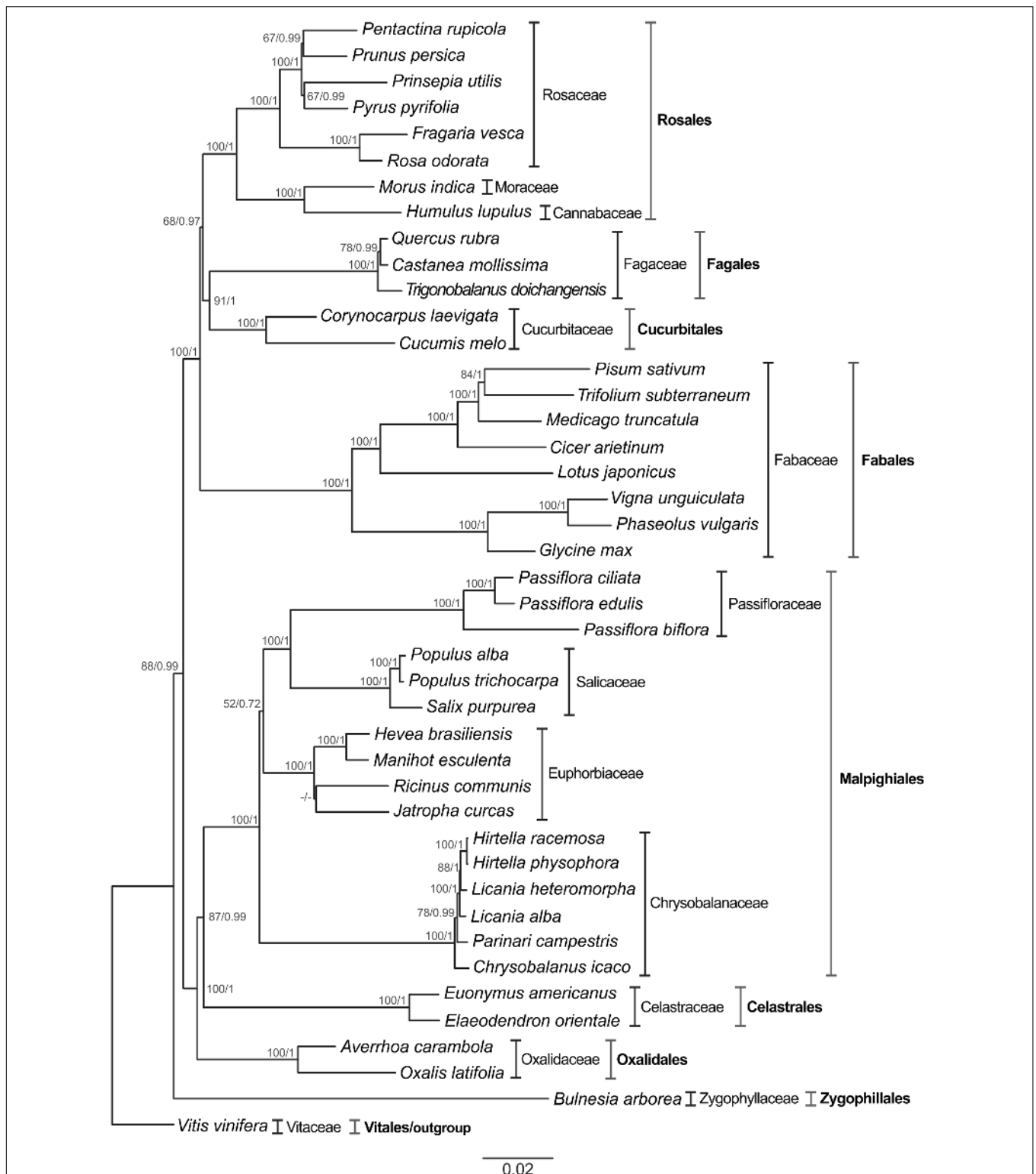


FIGURE 6 | Phylogenetic tree of the Fabids clade inferred from the nucleotide sequences of cp protein-coding genes of 43 plant species. Both maximum likelihood and BI were applied to the GTR + G model. The bootstrap values for maximum likelihood and posterior probability for BI are indicated above each node; a dash indicates that the support nodes obtained were <50% or <0.50. *V. vinifera* was used as outgroup to produce a rooted tree. The scale bar indicates the number of nucleotide substitutions per site. Protein-coding genes used in the analysis: *atpA*, *atpB*, *atpE*, *atpH*, *atpI*, *ccsA*, *cemA*, *matK*, *ndhC*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*, *petA*, *petG*, *petL*, *petN*, *psaA*, *psaB*, *psaC*, *psaJ*, *psbA*, *psbC*, *psbD*, *psbE*, *psbF*, *psbI*, *psbJ*, *psbK*, *rbcL*, *rpl14*, *rpl36*, *rpoB*, *rpoC2*, *rps11*, *rps14*, *rps2*, *rps3*, *rps4*, *rps8*.

It is worth noting that the Passifloraceae species comprise a distinct monophyletic group. In morphological terms, the monophyly of this family is supported by the occurrence of a flower corona (Judd et al., 2008). Additionally, our data suggest that *P. edulis* and *P. ciliata* are in close proximity. Both belong to the subgenus *Passiflora*. In taxonomic terms *P. edulis* is included in the supersection *Passiflora*, section *Passiflora*, and series *Passiflora*; *P. ciliata* is in the supersection *Stipulata*, section *Dysosmia*. However, *P. biflora* belongs to the subgenus *Decaloba* in supersection *Decaloba*, section *Decaloba* (Ulmer and MacDougal, 2004; Hansen et al., 2006). Passifloraceae appeared as a sister group to Salicaceae (BS = 100%; PP = 1) with a distant relationship to Chrysobalanaceae (BS = 52%; PP = 0.72) compared to Euphorbiaceae, confirming other findings based on four chloroplast, six mitochondrial and three nuclear gene sequences (Wurdack and Davis, 2009) and 82 chloroplast genes shared by 58 Malpighiales species (Xi et al., 2012).

However, we were unable to resolve the relationship between *J. curcas* and *R. communis* (BS \leq 50%; PP = \leq 0.50) in our work, confirming previous studies based on 60 chloroplast genes and the maximum likelihood method for estimating phylogeny inferences (Kong and Yang, 2016). Although the current phylogenetic reconstruction of Malpighiales is much improved compared to earlier versions, it is still incomplete, because of the limited taxonomic sampling available, and further phylogenetic and morphological studies are needed, focused especially on relations within Euphorbiaceae, by conducting, for instance, a phylogenomic analysis based on entire chloroplast genome sequences.

CONCLUSION

In this study, it was possible to obtain the first complete sequence of a chloroplast genome for the Passifloraceae family using the SMRT sequencing method, which proved highly effective for generating the biological data and efficient in assembling the chloroplast genome of *P. edulis*. Chloroplast genomes of the order Malpighiales were compared, and although found to be highly conserved, some genomes such as *P. edulis* have undergone rearrangements during evolution, showing that there is diversity

REFERENCES

- Asif, M. H., Mantri, S. S., Sharma, A., Srivastava, A., Trivedi, I., Gupta, P., et al. (2010). Complete sequence and organisation of the *Jatropha curcas* (Euphorbiaceae) chloroplast genome. *Tree Genet. Genomes* 6, 941–952. doi: 10.1007/s11295-010-0303-0
- Bardon, L., Chamagne, J., Dexter, K. G., Sothers, C. A., Prance, G. T., and Chave, J. (2013). Origin and evolution of Chrysobalanaceae: insights into the evolution of plants in the Neotropics. *Bot. J. Linn. Soc.* 171, 19–37. doi: 10.1111/j.1095-8339.2012.01289.x
- Barkan, A. (2004). "Intron splicing in plant organelles," in *Molecular Biology and Biotechnology of Plant Organelles*, eds H. Daniell and C. D. Chase (Dordrecht: Kluwer Academic Publishers), 281–308.
- Bell, C. D., Soltis, D. E., and Soltis, P. S. (2010). The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* 97, 1296–1303. doi: 10.3732/ajb.0900346

in the structure of plastid genomes in angiosperms. Definitely, complete chloroplast genomes or a significant set of their genic sequences are clearly useful in phylogenetic studies. Additionally, our results have opened the way for future phylogenomic studies on Passifloraceae.

AUTHOR CONTRIBUTIONS

LC-S and CFM obtained the complete chloroplast (cp) genome from the BAC sequences, performing all the bioinformatic analyses and drafting the manuscript. LC-S was also responsible for the phylogenomic studies. NR and SC performed the BAC-DNA extraction and assembly of the PacBio sequences, assisted by HB at CNRGV, France. AAS analyzed the BAC-end sequences and identified clones containing cpDNA using the BAC library constructed by HAP. AV assisted with genome annotation and GCXO with the phylogenomic studies, contributing to the discussion of results, together with MCD. MLV conceived the study and drafted the final manuscript, which was read and approved by all authors.

FUNDING

This work was supported by the Brazilian institutions, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant no. 2014/25215-2 and postdoctoral fellowship awarded to CFM) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, scholarship awarded to LC-S).

ACKNOWLEDGMENT

We thank Mr. Steve Simmons for proofreading the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00334/full#supplementary-material>

- Bruckner, C. H., Casali, V. W. D., de Moraes, C. F., Regazzi, A. J., and da Silva, E. A. M. (1995). Self-incompatibility in passion fruit (*Passiflora edulis* Sims). *Acta Hort.* 370, 45–58. doi: 10.17660/ActaHortic.1995.370.7
- Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., and Dopazo, J. (2015). A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the Genus Citrus. *Mol. Biol. Evol.* 32, 2015–2035. doi: 10.1093/molbev/msv082
- Carneiro, M. S., Camargo, L. E. A., Coelho, A. S. G., Vencovsky, R., Júnior, R. P. L., Stenzel, N. M. C., et al. (2002). RAPD-based genetic linkage maps of yellow passion fruit (*Passiflora edulis* Sims. f. *flavicarpa* Deg.). *Genome* 45, 670–678. doi: 10.1139/g02-035
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

- Chaisson, M. J., and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 13:238. doi: 10.1186/1471-2105-13-238
- Chen, X., Li, Q., Li, Y., Qian, J., and Han, J. (2015). Chloroplast genome of *Aconitum barbatum* var. *puberulum* (Ranunculaceae) derived from CCS reads using the PacBio RS platform. *Front. Plant Sci.* 6:42. doi: 10.3389/fpls.2015.00042
- Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Conant, G. C., and Wolfe, K. H. (2008). GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24, 861–862. doi: 10.1093/bioinformatics/btm598
- Cuco, S. M., Vieira, M. L. C., Mondin, M., and Aguiar-Perecin, M. L. R. (2005). Comparative karyotype analysis of three *Passiflora* L. species and cytogenetic characterization of somatic hybrids. *Caryologia* 58, 220–228. doi: 10.1080/00087114.2005.10589454
- Curci, P. L., De Paola, D., Danzi, D., Vendramin, G. G., and Sonnante, G. (2015). Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other Asteraceae. *PLoS ONE* 10:e0120589. doi: 10.1371/journal.pone.0120589
- Daniell, H., Wurdack, K. J., Kanagaraj, A., Lee, S.-B., Sasaki, C., and Jansen, R. K. (2008). The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron. *Theor. Appl. Genet.* 116, 723–737. doi: 10.1007/s00122-007-0706-y
- Darling, A. C. E. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109
- Ferrarini, M., Moretto, M., Ward, J. A., Šurbanovski, N., Stevanović, V., Giongo, L., et al. (2013). An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC Genomics* 14:670. doi: 10.1186/1471-2164-14-670
- Feuillet, C. (2004). “*Passifloraceae* (Passion flower family),” in *Flowering Plants of the Neotropics*, eds N. Smith, S. A. Mori, A. Henderson, D. W. Stevenson, and S. D. Heald (Oxford: Princeton University Press), 286–287.
- Folk, R. A., Mandel, J. R., and Freudenstein, J. V. (2015). A protocol for targeted enrichment of intron-containing sequence markers for recent radiations: a phylogenomic example from *Heuchera* (Saxifragaceae). *Appl. Plant Sci.* 3:1500039. doi: 10.3732/apps.1500039
- Grassi, F., Labra, M., Scienza, A., and Imazio, S. (2002). Chloroplast SSR markers to assess DNA diversity in wild and cultivated grapevines. *Vitis* 41, 157–158.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Hansen, A. K., Gilbert, L. E., Simpson, B. B., Downie, S. R., Cervi, A. C., and Jansen, R. K. (2006). Phylogenetic relationships and chromosome number evolution in *Passiflora*. *Syst. Bot.* 31, 138–150. doi: 10.1600/036364406775971769
- Hilu, K. W., Borsch, T., Muller, K., Soltis, D. E., Soltis, P. S., Savolainen, V., et al. (2003). Angiosperm phylogeny based on matK sequence information. *Am. J. Bot.* 90, 1758–1776. doi: 10.3732/ajb.90.12.1758
- Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., dePamphilis, C. W., Leebens-Mack, J., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19369–19374. doi: 10.1073/pnas.0709121104
- Jansen, R. K., Kaitanis, C., Sasaki, C., Lee, S.-B., Tomkins, J., Alverson, A. J., et al. (2006). Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. *BMC Evol. Biol.* 6:32. doi: 10.1186/1471-2148-6-32
- Jansen, R. K., Sasaki, C., Lee, S.-B., Hansen, A. K., and Daniell, H. (2011). Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of rpl22 to the nucleus. *Mol. Biol. Evol.* 28, 835–847. doi: 10.1093/molbev/msq261
- Jansen, R. K., Wojciechowski, M. F., Sanniyasi, E., Lee, S.-B., and Daniell, H. (2008). Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (Leguminosae). *Mol. Phylogenet. Evol.* 48, 1204–1217. doi: 10.1016/j.ympev.2008.06.013
- Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F., and Donghue, M. J. (eds) (2008). *Plant Systematics: A Phylogenetic Approach*, 3rd Edn. Sunderland, MA: Sinauer Associates.
- Kong, W., and Yang, J. (2016). The complete chloroplast genome sequence of *Morus mongolica* and a comparative analysis within the Fabidae clade. *Curr. Genet.* 62, 165–172. doi: 10.1007/s00294-015-0507-9
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633
- Lambowitz, A. M., and Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* 3:a003616. doi: 10.1101/cshperspect.a003616
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Li, X., Zhang, T.-C., Qiao, Q., Ren, Z., Zhao, J., Yonezawa, T., et al. (2013). Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). *PLoS ONE* 8:e58747. doi: 10.1371/journal.pone.0058747
- Malé, P.-J. G., Bardou, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., et al. (2014). Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Mol. Ecol. Resour.* 14, 966–975. doi: 10.1111/1755-0998.12246
- Martin, G., Baurens, F.-C., Cardi, C., Aury, J.-M., and D’Hont, A. (2013). The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): insight into plastid monocotyledon evolution. *PLoS ONE* 8:e67350. doi: 10.1371/journal.pone.0067350
- Matthews, M. L., and Endress, P. K. (2005). Comparative floral structure and systematics in Celastrales (Celastraceae, Parnassiaceae, Lepidobotryaceae). *Bot. J. Linn. Soc.* 149, 129–194. doi: 10.1111/j.1095-8339.2005.00445.x
- Melotto-Passarin, D. M., Tambarussi, E. V., Dressano, K., De Martin, V. F., and Carrer, H. (2011). Characterization of chloroplast DNA microsatellites from *Saccharum* spp and related species. *Genet. Mol. Res.* 10, 2024–2033. doi: 10.4238/vol10-3gmr1019
- Milligan, B. G., Hampton, J. N., and Palmer, J. D. (1989). Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Mol. Biol. Evol.* 6, 355–368.
- Moraes, M. C., Gerald, I. O., Matta, F. P., and Vieira, M. L. C. (2005). Genetic and phenotypic parameter estimates for yield and fruit quality traits from a single wide cross in yellow passion fruit. *HortScience* 40, 1978–1981.
- Munhoz, C. F., Santos, A. A., Arenhart, R. A., Santini, L., Monteiro-Vitorello, C. B., and Vieira, M. L. C. (2015). Analysis of plant gene expression during passion fruit–*Xanthomonas axonopodis* interaction implicates lipoxygenase 2 in host defence. *Ann. Appl. Biol.* 167, 135–155. doi: 10.1111/aab.12215
- Muschner, V. C., Zamberlan, P. M., Bonatto, S. L., and Freitas, L. B. (2003). Phylogeny, biogeography and divergence times in *Passiflora* (Passifloraceae). *Genet. Mol. Biol.* 35, 1036–1043. doi: 10.1590/S1415-47572012000600019
- Oliveira, E. J., Vieira, M. L. C., Garcia, A. A. F., Munhoz, C. F., Margarido, G. R. A., Consoli, L., et al. (2008). An integrated molecular map of yellow passion fruit based on simultaneous maximum-likelihood estimation of linkage and linkage phases. *J. Am. Soc. Hort. Sci.* 133, 35–41.
- Provan, J., Powell, W., and Hollingsworth, P. M. (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol. Evol.* 16, 142–147. doi: 10.1016/S0169-5347(00)02097-8
- Ramaiya, S. D., Bujang, J. S., and Zakaria, M. H. (2014). Assessment of total phenolic, antioxidant, and antibacterial activities of *Passiflora* species. *Sci. World J.* 2014, 1–10. doi: 10.1155/2014/167309
- Rêgo, M. M., Rêgo, E. R., Bruckner, C. H., Da Silva, E. A. M., Finger, F. L., and Pereira, K. J. C. (2000). Pollen tube behavior in yellow passion fruit following compatible and incompatible crosses. *Theor. Appl. Genet.* 101, 685–689. doi: 10.1007/s001220051531
- Rivarola, M., Foster, J. T., Chan, A. P., Williams, A. L., Rice, D. W., Liu, X., et al. (2011). Castor bean organelle genome sequencing and worldwide

- genetic diversity analysis. *PLoS ONE* 6:e21743. doi: 10.1371/journal.pone.0021743
- Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180
- Santos, A., Penha, H., Bellec, A., Munhoz, C. F., Pedrosa-Harand, A., Bergès, H., et al. (2014). Begin at the beginning: A BAC-end view of the passion fruit (*Passiflora*) genome. *BMC Genomics* 15:816. doi: 10.1186/1471-2164-15-816
- Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., et al. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142–166. doi: 10.3732/ajb.92.1.142
- Shin, S. C., Ahn, D. H., Kim, S. J., Lee, H., Oh, T.-J., Lee, J. E., et al. (2013). Advantages of single-molecule real-time sequencing in high-GC content genomes. *PLoS ONE* 8:e68824. doi: 10.1371/journal.pone.0068824
- Simon, D., Fewer, D., Friedl, T., and Bhattacharya, D. (2003). Phylogeny and self-splicing ability of the plastid tRNA-Leu group I Intron. *J. Mol. Evol.* 57, 710–720. doi: 10.1007/s00239-003-2533-3
- Souza, M. M., Palomino, G., Pereira, T. N. S., Pereira, M. G., and Viana, A. P. (2004). Flow cytometric analysis of genome size variation in some *Passiflora* species. *Hereditas* 141, 31–38. doi: 10.1111/j.1601-5223.2004.01739.x
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Su, H.-J., Hogenhout, S. A., Al-Sadi, A. M., and Kuo, C.-H. (2014). Complete chloroplast genome sequence of Omani Lime (*Citrus aurantifolia*) and comparative analysis within the Rosids. *PLoS ONE* 9:e113049. doi: 10.1371/journal.pone.0113049
- Sugiura, M. (1992). The chloroplast genome. *Plant Mol. Biol.* 19, 149–168. doi: 10.1007/BF00015612
- Sun, M., Soltis, D. E., Soltis, P. S., Zhu, X., Burleigh, J. G., and Chen, Z. (2015). Deep phylogenetic incongruence in the angiosperm clade Rosidae. *Mol. Phylogenet. Evol.* 83, 156–166. doi: 10.1016/j.ympev.2014.11.003
- Svistoonoff, S., Benabdoun, F. M., Nambiar-Vetil, M., Imanishi, L., Vaissayre, V., Cesari, S., et al. (2013). The independent acquisition of plant root nitrogen-fixing symbiosis in Fabids recruited the same genetic pathway for nodule organogenesis. *PLoS ONE* 8:e64515. doi: 10.1371/journal.pone.0064515
- Swofford, D. L. (2002). *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods)*. Sunderland, MA: Sinauer Associates.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tangphatsornruang, S., Uthapaisanwong, P., Sangsrakru, D., Chanprasert, J., Yoocha, T., Jomchai, N., et al. (2011). Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene* 475, 104–112. doi: 10.1016/j.gene.2011.01.002
- The Angiosperm Phylogeny Group (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1111/j.1095-8339.2009.00996.x
- Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604. doi: 10.1126/science.1128691
- Ulmer, T., and MacDougal, J. M. (2004). *Passiflora: Passionflowers of the World*. Portland, OR: Timber Press.
- Wang, H., Moore, M. J., Soltis, P. S., Bell, C. D., Brockington, S. F., Alexandre, R., et al. (2009). Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3853–3858. doi: 10.1073/pnas.0813376106
- Wu, Z. (2015). The new completed genome of purple willow (*Salix purpurea*) and conserved chloroplast genome structure of Salicaceae. *J. Nat. Sci.* 1, e49.
- Wurdack, K. J., and Davis, C. C. (2009). Malpighiales phylogenetics: gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. *Am. J. Bot.* 96, 1551–1570. doi: 10.3732/ajb.0800207
- Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255. doi: 10.1093/bioinformatics/bth352
- Xi, Z., Ruhfel, B. R., Schaefer, H., Amorim, A. M., Sugumaran, M., Wurdack, K. J., et al. (2012). Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17519–17524. doi: 10.1073/pnas.1205818109
- Xia, X. (2013). DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* 30, 1720–1728. doi: 10.1093/molbev/mst064
- Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., et al. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE* 5:e12762. doi: 10.1371/journal.pone.0012762
- Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., and Zhao, G. (2016). Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front. Plant Sci.* 7:959. doi: 10.3389/fpls.2016.00959
- Yockteng, R., and Nadot, S. (2004). Phylogenetic relationships among *Passiflora* species based on the glutamine synthetase nuclear gene expressed in chloroplast (ncpGS). *Mol. Phylogenet. Evol.* 31, 379–396. doi: 10.1016/S1055-7903(03)00277-X
- Yotoko, K. S. C., Dornelas, M. C., Togni, P. D., Fonseca, T. C., Salzano, F. M., Bonatto, S. L., et al. (2011). Does variation in genome sizes reflect adaptive or neutral processes? New clues from *Passiflora*. *PLoS ONE* 6:e18212. doi: 10.1371/journal.pone.0018212
- Zhang, L.-B., and Simmons, M. P. (2006). Phylogeny and delimitation of the Celastrales inferred from nuclear and plastid genes. *Syst. Bot.* 31, 122–137. doi: 10.1600/036364406775971778

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer FTB and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Cauz-Santos, Munhoz, Rodde, Cauet, Santos, Penha, Dornelas, Varani, Oliveira, Bergès and Vieira. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.