

TÚLIO GOMES PACHECO

**PLASTID GENOMICS IN NEGLECTED AND UNCONVENTIONAL BOTANICAL
FAMILIES AS TOOL FOR GENETIC, EVOLUTIONARY AND PHYLOGENETIC
ANALYSES**

Thesis presented to the Universidade Federal de Viçosa as a part of requirements of the Plant Physiology Graduate Program for obtention of the degree of Doctor Scientiae.

Adviser: Marcelo Rogalski

Co-Adviser: Amanda de Santana Lopes

VIÇOSA – MINAS GERAIS

2020

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

Pacheco, Túlio Gomes, 1989-
P116p Plastid genomics in neglected and unconventional botanical
2020 families as tool for genetic, evolutionary and phylogenetic
analyses / Túlio Gomes Pacheco. – Viçosa, MG, 2020.
171 f. : il. (algumas color.) ; 29 cm.

Texto em inglês.

Orientador: Marcelo Rogalski.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. *Passiflora cirrhiflora*. 2. Genômica. 3. Evolução.
4. Cloroplastos. I. Universidade Federal de Viçosa.
Departamento de Biologia Vegetal. Programa de Pós-Graduação
em Fisiologia Vegetal. II. Título.

CDD 22. ed. 583.626

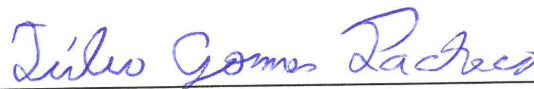
TÚLIO GOMES PACHECO

PLASTID GENOMICS IN NEGLECTED AND UNCONVENTIONAL BOTANICAL
FAMILIES AS TOOL FOR GENETIC, EVOLUTIONARY AND PHYLOGENETIC
ANALYSES

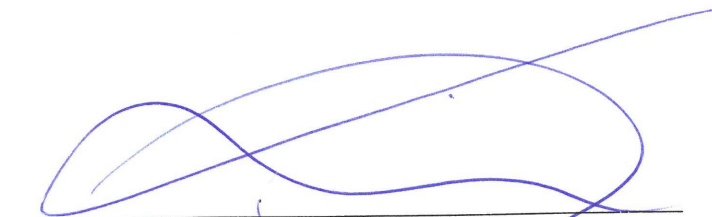
Thesis presented to the Universidade Federal
de Viçosa as a part of requirements of the
Plant Physiology Graduate Program for
obtention of the degree of *Doctor Scientiae*.

APPROVED: December 22, 2020

Assent:



Túlio Gomes Pacheco
Author



Marcelo Rogalski
Adviser

To my parents for the support.

To Giovanna, Jaina e Azshara for all affection.

ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. And also by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (2018-2020). Thus, thanks to CAPES and CNPq for the scholarship granted.

To the prof. Marcelo Rogalski, for eight years and half of mentorship with a lot of dedication, friendship, monitoring and patience.

To all my friends from the Plant Molecular Physiology Lab (current and formers) for the very good interaction and discussions about this and other studies. Among them, a special thanks for my co-adviser, Amanda de Santana Lopes, for the friendship, scientific discussions and direct collaboration during my academic trajectory. Thanks also to Maria Carolina and Daniel Oliveira for the friendship and for the discussions and support during this year of remote work.

To the Universidade Federal de Viçosa (UFV) and Plant Physiology Graduate Program for the opportunity of professional growth.

To the Núcleo de Análises de Biomoléculas (Nubiomol) of the UFV for providing the CLC Genomics Workbench software.

To the profs. Miguel Pedro Guerra (UFSC), Rubens Onofre Nodari (UFSC), Leila do Nascimento Vieira (UFPR), Emanuel Maltempi de Souza (UFPR) and Fábio de Oliveira Pedrosa (UFPR) for supporting the conduction of the present work. To the prof. Wagner Campos Otoni for the valuable collaborations, mainly in the works with *Passiflora* and *annato*.

To the committee who gently accepted to participate of my PhD exam.

To all that contributed directly or indirectly, for this work and for my professional and personal growth, thank you so much!

ABSTRACT

PACHECO, Túlio Gomes, D.Sc, Universidade Federal de Viçosa, December, 2020. **Plastid genomic in neglected and unconventional botanical families as tool for genetic, evolutionary and phylogenetic analyses.** Adviser: Marcelo Rogalski. Co-adviser: Amanda de Santana Lopes.

Normally, the plastid genome (plastome) of land plants contains 100-130 genes, which are organized in a DNA molecule of 120-160 kb. The plastome sequencing molecule has been important for studies related to genetic, evolution, and phylogenetic analyses. Plastid genes are useful to unveil phylogenetic relationships between families and botanical orders, while fast-evolving regions (i.e., introns and intergenic spacers) are a great source of molecular markers to use in phylogenetic analyses of low taxa and genetic diversity analysis. Also, the analysis of complete plastomes allows us to uncover unusual features such as large inversions, gene losses, the active selection acting on plastid genes, and the prediction of RNA editing sites. However, plastomes of several angiosperms families have remained poorly studied. Thus, the objective of this study was the sequencing of the first complete plastomes of the families Tropaeolaceae (*Tropaeolum pentaphyllum*) and Bixaceae (*Bixa orellana*); and contribute to a better understanding of the plastome evolution in the genus *Passiflora* (Passifloraceae) via plastome sequencing of six species of the subgenus *Passiflora* and *Passiflora cirrhiflora* of the subgenus *Deidamioides*. This work reports a detailed characterization of the plastomes, including the analysis of several parameters: structure of the plastomes, gene content, genetic divergence, molecular markers, RNA editing sites, and phylogenetic aspects. *B. orellana* showed some specific features if compared with other Malvales, such as positive selection acting on *psaI* gene and 11 specific RNA editing sites. Similarly, *T. pentaphyllum* showed specific extensions of the *matK* and *rpoA* genes, and two specific RNA editing sites. The species of the subgenus *Passiflora* showed highly divergent genes (i.e. *clpP* and *accD*) and a significant polymorphism of gains and losses of RNA editing sites. In contrast, *P. cirrhiflora* and other species of the subgenus *Deidamioides* showed a conserved set of RNA editing sites. In general, the phylogenetic analyses performed here showed well-supported trees, but some inconsistencies remained to be elucidated by using a larger number of sampled taxons. A large amount of genomic data was provided by this study, which is useful in various analyzes related to genetics, evolution, biotechnology, and conservation of the species studied here and other related taxa.

Keywords: Organelle DNA. Molecular markers. Plastome evolution.

RESUMO

PACHECO, Túlio Gomes, D.Sc, Universidade Federal de Viçosa, dezembro de 2020. **Genômica plastidial em famílias botânicas não convencionais e negligenciadas como uma ferramenta para análises genéticas, evolutivos e filogenéticas.** Orientador: Marcelo Rogalski. Coorientadora: Amanda de Santana Lopes.

Em geral, o genoma plastidial (plastoma) de plantas terrestres contém cde 100-130 genes, organizados em uma molécula de DNA de 120-160 kb. O sequenciamento do plastoma é importante para estudos genéticos, filogenéticos e evolutivos. Os genes plastidiais são úteis para análises filogenéticas entre famílias e ordens, enquanto regiões como íntrons e regiões intergênicas são fontes de marcadores para uso em filogenias entre espécies/gêneros e em estudos de diversidade genética. Análise do plastoma permite a detecção de inversões e perdas gênicas, a seleção atuante nos genes plastidiais, e a predição de sítios de edição de RNA. No entanto, os plastomas de várias famílias de angiospermas permanecem pouco estudados. Assim, o objetivo deste estudo foi sequenciar pela primeira vez o plastoma completo de uma espécie das famílias Tropaeolaceae (*Tropaeolum pentaphyllum*) e Bixaceae (*Bixa orellana*) e contribuir para o melhor entendimento da evolução plastidial no gênero *Passiflora*, por meio do sequenciamento do plastoma de espécies do subgênero *Passiflora* e uma espécie do subgênero *Deidamioides* (*Passiflora cirrhiflora*). Este trabalho apresenta uma caracterização detalhada da estrutura dos plastomas, do conteúdo e da divergência gênica, de marcadores, dos sítios de edição de RNA e da filogenia de cada táxon aqui sequenciado. *B. orellana* apresentou algumas características específicas como seleção positiva atuando no gene *psaI* e 11 sítios de edição de RNA específicos. Paralelamente, *T. pentaphyllum* mostrou extensões específicas nos genes *matK* e *rpoA*, e a predição de dois sítios de edição de RNA específicos. As espécies do subgênero *Passiflora* mostraram sequências altamente divergentes dos genes *clpP* e *accD*, e um polimorfismo significativo relacionado ao ganho e perda de sítios de edição de RNA. Por outro lado, *P. cirrhiflora* e outras espécies do subgênero *Deidamioides* apresentaram grande conservação de sítios de edição. No geral, as análises filogenéticas realizadas aqui mostraram árvores bem suportadas, porém algumas incongruências permaneceram para ser elucidadas com uma maior amostragem de taxóns. Neste estudo foi gerada uma grande quantidade de dados genômicos que poderá ser utilizados em várias análises relacionadas à genética, evolução, biotecnologia e conservação das espécies aqui estudadas e de outros taxóns relacionados.

Palavras-chave: DNA de organela. Marcadores moleculares. Evolução do genoma plastidial.

SUMMARY

General introduction.....	9
References.....	15
Chapter I.....	23
Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (<i>Bixa Orellana</i> L.), the Amazonian commercial species of natural dyes	
Abstract.....	24
Introduction.....	25
Materials and Methods.....	26
Results.....	27
Discussion.....	33
Conclusion.....	39
References.....	39
Supplementary Material.....	44
Chapter II.....	59
Phylogenetic and evolutionary features of the plastome of <i>Tropaeolum pentaphyllum</i> Lam. (<i>Tropaeolaceae</i>)	
Abstract.....	60
Introduction.....	60
Material and Methods.....	62
Results.....	62
Discussion.....	70
Conclusions.....	74
References.....	74
Supplementary Material.....	78
Chapter III.....	91
Plastome sequences of the subgenus <i>Passiflora</i> reveal highly divergent genes and specific evolutionary features	
Abstract.....	92

Introduction.....	92
Materials and Methods.....	93
Results.....	95
Discussion.....	102
Conclusions.....	104
References.....	105
Supplementary Material.....	108
Chapter IV.....	146
The complete plastome of <i>Passiflora cirrhiflora</i> A. Juss.: structural features, RNA editing sites, hotspots of nucleotide diversity and molecular markers within the subgenus <i>Deidamioides</i>	
Abstract.....	147
Introduction.....	147
Materials and Methods.....	148
Results.....	149
Discussion.....	155
Conclusion.....	158
References.....	158
Supplementary Material.....	162
General conclusions.....	170

General introduction

The plastids are essential organelles to plant cell metabolism, harboring primary metabolic reactions (i.e., photosynthesis), and the synthesis of various metabolites such as fatty acids, starch, amino acids, pigments, and several hormone precursors (Tetlow et al., 2004; Aldridge et al., 2005; Rogalski and Carrer, 2011). Plastids originate from proplastids, present in meristematic cells; and can differentiate into several types, based on their morphology, function, and development stage, such as chloroplasts (where photosynthesis takes place), amyloplasts (starch storage), chromoplasts (carotenoid-rich), proteoplasts (protein storage), elaioplasts (oil storage), and xyloplasts (involved in secondary growth) (Leister and Pesaresi, 2005; Pyke, 2007; Pinar and Mizrahi, 2018).

The evolutionary origin of the plastid is based on an endosymbiotic event, in which a eukaryotic cell (already containing mitochondria) engulfed a cyanobacterial ancestor, generating a new cell containing three intracellular genetic compartments: the nucleus, mitochondrial and plastid (Timmis et al., 2004). After the endosymbiotic process, in this new intracellular environment, the plastid genome (plastome) has undergone an extensive genetic reorganization, including a drastic reduction of the size and gene content. This reduction is due mainly to the loss of dispensable and ambiguous genes and transfer of plastid genes to the nucleus (Timmis et al., 2004; Bock and Timmis, 2008; Zimorski et al., 2014; Rogalski et al., 2015). The transferred genes became nuclear genome-encoded, but their products (proteins) are imported by plastids, generating a complex genetic interaction between nuclear and plastid gene products. Nowadays, the plastome of most land plants contains a conserved set of 100-130 genes, organized in a genome of 120-160 kb (Wicke et al., 2011; Tonti-Filippini et al., 2017). On other hand, the genome of typical cyanobacteria species encodes more than 3200 genes (Kaneko et al., 1996; Bock, 2015), illustrating the massive gene loss experienced by the plastome during the plant evolution.

A typical angiosperm plastome is normally subdivided into four parts: two inverted repeats (IRs), separated by two single-copy regions, a large (LSC), and a small region (SSC) (Bock, 2007, 2015). This genome encodes genes related to photosynthesis (large subunit of Rubisco and subunits of the photochemical apparatus – photosystems I and II, cytochrome b6/f, and ATP synthase), gene expression (plastid-encoded RNA polymerase – PEP, ribosomal proteins, tRNAs, and rRNAs), intron maturation (the maturase K, which participates in the splicing of plastid introns of the IIA group), synthesis of fatty acids (a subunit of the Acetyl-CoA carboxylase), protein import (a subunit of the TIC/TOC complex of the plastid

membrane), and protein homeostasis (a subunit of the ATP-dependent Clp protease) (Bock, 2007; Kikuchi et al., 2013; Barthet et al., 2020).

The plastome in most angiosperms has uniparental inheritance (i.e., maternal inheritance) lacking parental recombination (Bock, 2007). Furthermore, plastid genes generally show a low rate of nucleotide substitution, due mainly to the high ploidy of the plastome in each plastid, favoring the diminishing of mutations, through the process of gene conversion (Khakhlova e Bock, 2006; Bock, 2007). These features point to the plastid genes as great molecular markers for phylogenetic studies involving families, order, and other high taxa (Jansen et al., 2007; Moore et al., 2010; Vieira et al., 2014; Rogalski et al., 2015; Lopes et al., 2017). On other hand, introns and intergenic regions generally have higher mutation rates (if compared with coding regions) and thus can be utilized to clarify phylogenetic relationships at low taxonomic levels, such as tribes, genus, and species (Shaw et al 2005, 2007; Rogalski et al., 2015). For example, if a sampled taxa show a conserved structure containing the same gene order and similar gene content, the complete plastome sequence of them can be used in a phylogenomic approach, searching to resolve deep-level phylogenetic relationships (Xi et al., 2012; Wambugu et al., 2015; Lopes et al., 2018a and b).

Simple sequence repetitions (SSRs) are abundant across the plastome and recognized as excellent molecular markers due to their high polymorphism (Provan et al., 2001). Once the complete plastid sequence of a species is available, hundreds of SSRs with high potential of mutation can be detected and characterized (Vieira et al., 2014, 2015; Lopes et al., 2018 a and b). These markers can be accessed for analyzes of inter and intraspecific variability, which is of great value for several genetic studies such as gene flow, bottleneck effects, structure and divergence of populations, cytoplasmic diversity, hybrid origin, and phylogeographic analyses (Powell et al., 1995; Provan et al., 2001; Ebert and Peakall, 2009; Rogalski et al., 2015; Tamburino et al., 2020). Besides SSRs, plastomes can be used to analyze other polymorphic markers such as SNPs (single nucleotide polymorphism) and indels (insertion/deletion), which are also useful genetic resources to population genetic, molecular breeding, and phylogenetic studies (Lopes et al., 2019; Achakkagari et al., 2020; Li et al., 2020; Teshome et al., 2020; Tamburino et al., 2020).

The plastome structure, gene order, and content are well conserved in most angiosperms. However, some families of this group are characterized by containing unusual plastome features such as several inversions in the LSC region, large events of IR contraction and expansion, and losses of conserved genes and introns. Among these families, we can highlight Cactaceae, Campanulaceae, Ericaceae, Fabaceae, Geraniaceae, Passifloraceae, and Plantaginaceae (Cai et

al., 2008; Haberle et al., 2008; Guisinger et al., 2011; Sanderson et al., 2015; Schwarz et al., 2015; Zhu et al., 2016; Rabah et al., 2019; Shrestha et al., 2019; Solórzano et al., 2019). Rearrangements and gene losses are features usually detected through sequencing, assembly and annotation of complete plastome sequences. These sequences allow detailed characterization of evolutionary events, in several taxonomic levels such as species, genera, tribes, families, and orders (Rogalski et al., 2015; Vieira et al., 2016; Lopes et al., 2017, 2018a and b). Furthermore, the availability of these sequences enables the investigation of evolutionary pressures acting on plastid genes. This kind of inference can be carried out by evaluating the rates of synonymous and non-synonymous substitutions in protein-coding genes (Hu et al., 2015; Piot et al., 2018; Lopes et al., 2018a). Non-synonymous (dN) mutations are those that alter the encoded amino acid, while the synonymous substitutions (dS) do not alter. The dN/dS ratio can indicate three selection types: negative (or purifying) selection ($dN/dS < 1$), which favors the amino acid conservation, positive (or Darwinian) selection ($dN/dS > 1$) favoring the amino acid change, and neutral (or balancing) selection (dN equal to dS) (Oleksyk et al., 2010; Goodswen et al., 2018). Thus, negative selection acts to maintain the protein function, while the positive selection is likely related to adaptation to environmental/ecological conditions (Piot et al., 2018). Selection analyses allow us to obtain insights concerning plastid molecular changes and their possible correlation with plant physiology and adaptation to different environmental conditions (Hu et al., 2015; Piot et al., 2018; Lopes et al., 2018a).

The sequence of plastid genes has also been used for the prediction, detection and comparison of RNA editing sites, a post-transcriptional mechanism commonly found in plastid transcripts of land plants (Takenaka et al., 2013). This mechanism can create start/stop codons and restore evolutionarily conserved amino acids by modifying the mRNA sequence before its translation (Tsudzuki et al., 2001; Takenaka et al., 2013). The RNA editing in spermatophyte corresponds to changes from cytidine (C) to uridine (U), while in ferns and mosses U to C changes are also observed (Takenaka et al., 2013; Ichinose and Sugita, 2016). Several species belonging to different taxa were already analyzed concerning the presence and specificity of RNA editing sites (Bock et al., 1997; Tillich et al., 2005; Kahlau et al., 2006; Asif et al., 2010; He et al., 2016; Lopes et al., 2018a and b). Several sites are recurrent in different plant lineages, while others are specific to lineages or species. This indicates that the plastid RNA editing is very dynamic and a fast-evolving feature, which includes recent events of gain and/or loss of editing sites among different land plant taxa.

Another interesting aspect of plastid genomics is related to the interaction between nuclear and plastid gene products. Plastome contains several genes that encode subunits of

multiprotein complexes (MPCs). The product of these genes must interact with the products of nuclear-encoded genes to assemble a functional MPC into the plastids. Consequently, plastid and nuclear subunits of the same MPC must co-evolve to enable essential cellular processes. Mutations that eventually emerge in a plastid-encoded subunit have to be compensated by mutations in the nuclear-encoded subunit of the complex (Greiner and Bock, 2013). The importance of this fine adjustment between the nucleus and plastome is evident during experiments aiming at the search for interspecific hybrids. If the parental lines undergo different selection pressures affecting differently their cell genetic compartments, it can result in several failures in the hybrid plants during growth and development or even in cell lethality (Levin, 2003; Schmitz-Linneweber, 2005; Greiner et al., 2011). The phenomenon of incompatibility between the nucleus (genome) and plastome is easily observed when a photosynthetic MPC is affected resulting in hybrids with variegated, pale green or white phenotypes (Greiner et al., 2011). The complete plastome sequence is a basic tool to elucidate the genetic cause of this kind of cellular incompatibility. Examples of this approach include a study in *Pisum sativum* hybrids, in which the incompatibility was related to the high divergence of the *accD* gene (Bogdanova et al., 2015). In *Oenothera*, it was associated with the deletion of a promoter region, localized between the *clpP* and *psbB* genes, affecting the expression of photosynthesis-related genes (Greiner et al., 2008). Another classic example was observed in cybrids of *Atropa belladonna* and *Nicotiana tabacum* experimentally generated to contain the nucleus of *A. belladonna* and the plastids of *N. tabacum*. The incompatibility was determined by the failure to edit an RNA editing site of the *atpA* gene (encodes a subunit of plastid ATP synthase) (Schmitz-Linneweber et al., 2005), which conferred albinism to the plants.

Lastly, in addition to the several genetic and evolutionary studies, complete plastome sequences are also of great importance for genetic transformation aiming at both basic research and biotechnology. Plastid transformation has been widely used to reveal the function of several plastid genes (Rogalski et al., 2006, 2008; Fleischmann et al., 2011; Alkatib et al., 2012), to engineer metabolic pathways and to express foreign proteins of industrial and pharmaceutical interest (Rogalski and Carrer, 2001; Apel and Bock, 2009; Lu et al., 2013; Bock, 2015). It is very important to know the plastome sequence before transformation because this technique takes the advantage of the homologous recombination mechanism existing in plastids. Therefore, specific transformation vectors should be engineered based on the plastid sequences, which are flanking the genes of interest aiming at precise integration via two homologous recombination events (Bock, 2015; Rogalski et al., 2015).

Here, it is presented the complete plastome sequences of representative species of three different angiosperm families: Bixaceae (*Bixa orellana* L.), Tropaeolaceae (*Tropaeolum pentaphyllum* Lam.), and Passifloraceae (seven species of *Passiflora* L.). In this work several evolutionary and genetic aspects of the plastome of each taxon are discussed in detail. Some aspects are shared among the species of the three families, but several questions are specifically addressed to the different taxa studied here in the following chapters.

1. *Bixa orellana* – Bixaceae

Annatto (*B. orellana*) is a species native to the Amazon forest in Brazil and belongs to the family Bixaceae (order Malvales) (Dequigiovanni et al., 2014; Le Péchon and Gigord, 2014). This species has elevated economic importance worldwide due mainly to the accumulation of pigments in the seeds (bixin and norbixin), which are sources of natural dyes (Vilar et al., 2014). The seed extract is used in food, cosmetic and pharmaceutical industries because it does not alter the flavor and it is not toxic (Raddatz-Mota et al., 2017). Here, it is presented the complete plastome sequence of *B. orellana*, the first plastome of Bixaceae to be fully sequenced and characterized in detail. The data and analyses generated allowed the detection of new putative molecular markers to this socioeconomically important species. The knowledge of a Bixaceae plastome and comparative analysis with other families also allowed to bring new insights concerning the evolution of the plastome within Malvales and contributed to the understanding of phylogenetic relationships among the families of this order.

2. *Tropaeolum pentaphyllum* – Tropaeolaceae

Tropaeolaceae (Brassicales) is composed of approximately 100 species, distributed in the neotropics, mainly in high altitude regions of South America (Souza and Lorenzi, 2008). Cronquist (1988) divided this family into three genera: *Magallana*, *Tropaeastrum*, and *Tropaeolum*. More recently, molecular analyses suggested the existence of a single genus, *Tropaeolum* (Andersson and Andersson, 2000). Five species of the genus *Tropaeolum* (*T. brasiliense*, *T. orthoceras*, *T. pentaphyllum*, *T. sanctae-catharinae*, and *T. warmingiam*) occur in Brazil. *T. warmingiam* is restricted to the south and southeast (Souza and Lorenzi, 2008). In Brazil, *T. pentaphyllum* is popularly known as batata crem or crem and constitutes an unconventional food plant (Kinupp and Lorenzi, 2014). Crem leaves and flowers can be used in salads, while the tubers can be consumed as processed and canned foods (Braga et al., 2018). In addition to food uses, *T. pentaphyllum* is also used as ornamental, due to the durability and beauty of its flowers. Moreover, it is used for medicinal purposes given that its tubers are considered antiscorbutic and depurative (Mors et al., 2000; Kinupp, 2011). This species is in

the vulnerable category in the list of endangered species from Brazil, mainly due to the loss of natural habitat caused by deforestation to increase of agricultural areas (Kinupp, 2011).

Genomic studies are essential for the detection and characterization of molecular markers, which can be used for analyses of genetic diversity in natural populations/germplasm accesses and conservation strategies. However, genetic studies with *T. pentaphyllum* are scarce. Thus, in this present work is reported the complete plastome sequence of this species, the first plastome of the family Tropaeolaceae to be fully sequenced. The plastome of *T. pentaphyllum* was molecularly characterized in detail and allowed the detection of several SSRs markers and hotspots of nucleotide polymorphism, which are useful data for genetic e evolutionary studies in this species. This work also brings insights into the phylogeny and plastome evolution within Brassicales, including polymorphism of RNA editing sites and the nucleotide divergence of plastid genes.

3. Genus *Passiflora* – subgenus *Passiflora*

Passiflora is the largest genus of the family Passifloraceae (MacDougal and Feuillet, 2004) and is majority distributed in the neotropics (Rocha et al., 2020). The widely accepted division of this genus includes five subgenera: *Astrophea*, *Decaloba*, *Deidamioides*, *Passiflora*, and *Tetrapathea* (MacDougal and Feuillet, 2004; Krosnick et al., 2009, 2013)

The subgenus *Passiflora* includes the species with the highest economic importance, including uses for food (fruit consumption), ornamental, medicinal, and cosmetic purposes (Krosnick et al., 2013; Cerqueira-Silva et al., 2014; Rocha et al., 2020). Recently, several plastomes of different species of the genus *Passiflora* were sequenced (Cauz-Santos et al., 2017; Rabah et al., 2018; Shrestha et al., 2019), which revealed various unusual evolutionary features such as diverse rearrangements, and massive gene and introns losses. These unusual evolutionary features vary among and within the subgenera.

Several cases of nucleus-plastome incompatibility were already demonstrated in interspecific hybridization of the subgenus *Passiflora* (Mráček, 2005; Ocampo, 2016). However, the genetic candidates of this incompatibility were still not elucidated. In this work, it is presented complete plastome sequences of six species of the subgenus *Passiflora*: *P. elegans*, *P. maliformis*, *P. malacophylla*, *P. mucronata*, *P. incarnata*, and *P. cincinnata*. These data in combination with the plastomes already available in the database, allowed us to identify and characterize molecular markers, gene sequences, RNA editing sites, and rearrangements. The data and analyses indicate some putative candidates for nucleus-plastome incompatibility that are discussed in detail in Chapter III.

4. Genus *Passiflora* – *Passiflora cirrhiflora* (subgenus *Deidamioides*)

The subgenus *Deidamioides* is the poorest one concerning the species number of the genus *Passiflora*, possessing 14 species found in Central and South America (Krosnick et al., 2013). This subgenus is also poorly studied concerning plastome genetics and evolution. Three species of the subgenus *Deidamioides* were completely sequenced and analyzed regarding gene content, structure, and phylogenetic position (Shrestha et al., 2019). These authors showed a polyphyletic origin of this subgenus and several divergent evolutionary features among them, regarding plastome structure and gene content. Aiming to complement the study with this subgenus and to investigate other evolutionary and genetic aspects of *Deidamioides* plastomes (e.g. RNA editing sites distribution and presence of hotspots of nucleotide polymorphism), the present work sequenced and analyzed the complete plastid genome of *P. cirrhiflora*, a species native from the Amazonian forest (Killip, 1938).

References

- Achakkagari, S.R., Kyriakidou, M., Tai, H.H., Anglin, N.L., Ellis, D., Strömvik, M.V. (2020) Complete plastome assemblies from a panel of 13 diverse potato taxa. *PLoS ONE* 15, e0240124.
- Aldridge, C., Maple, J., Moller, S. G. (2005) The molecular biology of plastid division in higher plants. *J. Exp. Bot.* 56, 1061–1077.
- Alkatib, S., Scharff, L.B., Rogalski, M., Fleischmann, T.T., Matthes, A., Seeger, S., Schöttler, M.A., Ruf, S., Bock, R. (2012) The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet.* 8, e1003076.
- Andersson, L., Andersson, S. (2000). A molecular phylogeny of tropaeolaceae and its systematic implications. *Taxon* 49, 721–736
- Apel, W., Bock, R. (2009) Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene-to-provitamin A conversion. *Plant Physiol.* 151, 59–66.
- Asif, M.H., Mantri, S.S., Sharma, A., Srivastava, A., Trivedi, I., Gupta, P., Mohanty, C.S., Sawant, S.V., Tuli, R. (2010) Complete sequence and organisation of the *Jatropha curcas* (Euphorbiaceae) chloroplast genome. *Tree Genet. Genomes* 6, 941–952.
- Barthet, M.M., Pierpont, C.L., Tavernier, E.K. (2020) Unraveling the role of the enigmatic MatK maturase in chloroplast group IIA intron excision. *Plant Direct.* 4, e00208.
- Bock, R. (2007) Structure, function, and inheritance of plastid genomes. In: *Cell and Molecular Biology of Plastids*. Bock, R. (Ed). *Top. Curr. Genet.* 19, 524p.
- Bock, R. (2015) Engineering Plastid Genomes: Methods, Tools, and Applications in Basic Research and Biotechnology. *Annu. Rev. Plant Biol.* 66, 211–241.

- Bock, R., Albertazzi, F., Freyer, R., Fuchs, M., Ruf, S., Zeltz, P., Maier, R.M. (1997) Transcript editing in chloroplasts of higher plants. In: Schenk, H.E.A., Herrmann, R.G., Jeon, K.W., Müller, N.E., Schwemmler, W. (eds) *Eukaryotism and symbiosis*. Springer, Berlin, pp 123–137.
- Bock, R., Timmis, J.N. (2008) Reconstructing evolution: Gene transfer from plastid to the nucleus. *BioEssays*. 30, 556–566.
- Bogdanova, V.S., Zaytseva, O.O., Mglinets, A.V., Shatskaya, N.V., Kosterin, O.E., Vasiliev, G.V. (2015) Nuclear-Cytoplasmic Conflict in Pea (*Pisum sativum* L.) Is Associated with Nuclear and Plastidic Candidate Genes Encoding Acetyl-CoA Carboxylase Subunits. *PLOS ONE*. 10, e0119835.
- Braga, V.B., Vieira, M.M., Barros, I.B.I. (2018) Nutritional potential of leaves and tubers of crem (*Tropaeolum pentaphyllum* Lam.). *Rev. Nutr.* 31, 423–432.
- Cai, Z., Guisinger, M., Kim, H.G., Ruck, E., Blazier, J.C., McMurtry, V., Kuehl, J.V., Boore, J., Jansen, R.K. (2008) Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J. Mol. Evol.* 67, 696–704.
- 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. *Front. Plant. Sci.* 8,1-17.
- Cerqueira-Silva, C.B.M., Santos, E.S.L., Vieira, J.G.P., Mori, G.M., Jesus, O.N., Corrêa, R.X., and Souza, A.P. (2014). New Microsatellite Markers for Wild and Commercial Species of *Passiflora* (Passifloraceae) and Cross-Amplification. *Appl. Plant Sci.* 2, 1300061.
- Cronquist, A. (1988) *The evolution and classification of flowering plants*. Bronx, New York.
- Dequigiovanni, G., Ramos, S.L.F., Zucchi, M.I., Bajay, M.M., Pinheiro, J.B., Fabri, E.G., Bressan, E.A., Veasey, E.A. (2014) Isolation and characterization of microsatellite loci for *Bixa orellana*, an important source of natural dyes. *Genet. Mol. Res.* 13:9097–9102.
- Ebert, D., Peakall, R. (2009) Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Mol. Ecol. Resour.* 9, 673–690.
- Fleischmann, T.T., Scharff, L.B., Alkatib, S., Hasdorf, S., Schöttler, M.A., Bock, R. (2011). Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell.* 23, 3137–55.
- Goodswen, S.J., Kennedy, P.J., & Ellis, J.T. (2018). A Gene-Based Positive Selection Detection Approach to Identify Vaccine Candidates Using *Toxoplasma gondii* as a Test Case Protozoan Pathogen. *Front. Genet.* 9, 332.
- Greiner, S., Bock, R. (2013) Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays*. 35, 354–365.

- Greiner, S., Rauwolf, U., Meurer, J., Herrmann, R.G. (2011) The role of plastids in plant speciation: Plastids and Speciation. *Mol. Ecol.* 20, 671–691.
- Greiner, S., Wang, X., Herrmann, R.G., Rauwolf, U., Mayer, K., Haberer, G., Meurer, J. (2008) The complete nucleotide sequences of the 5 genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: II. A microevolutionary view using bioinformatics and formal genetic data. *Mol. Biol. Evol.* 25, 2019–2030.
- Guisinger, M.M., Kuehl, J.V., Boore, J.L., Jansen, R.K. (2011) Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* 28, 583–600.
- Haberle, R.C., Fourcade, H.M., Boore, J.L., Jansen, R.K. (2008) Extensive Rearrangements in the Chloroplast Genome of *Trachelium caeruleum* Are Associated with Repeats and tRNA Genes. *J. Mol. Evol.* 66, 350-361.
- He, P., Huang, S., Xiao, G., Zhang, Y., Yu, J. (2016) Abundant RNA editing sites of chloroplast protein-coding genes in *Ginkgo biloba* and an evolutionary pattern analysis. *BMC Plant Biol.* 16, 257.
- Hu, S., Sablok, G., Wang, B., Qu, D., Barbaro, E., Viola, R., Li, M., Varotto, C. (2015) Plastome organization and evolution of chloroplast genes in Cardamine species adapted to contrasting habitats. *BMC Genomics* 16, 306
- Ichinose, M., Sugita, M; (2016) RNA Editing and Its Molecular Mechanism in Plant Organelles. *Genes* 8, 5.
- Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., 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.* 104, 19369–19374.
- Kahlau, S., Aspinall, S., Gray, J.C., Bock, R. (2006) Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J. Mol. Evol.* 63, 194-207.
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., et al. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* 3, 109–136.
- Khakhlova, O., Bock, R. (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* 46, 85-94.
- Kikuchi, S., Bédard, J., Hirano, M., Hirabayashi, Y., Oishi, M., Imai, M., Takase, M., Ide, T., Nakai, M. (2013) Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* 339, 571–574.

- Killip, E.P. (1938) The American species of Passifloraceae. Publ. Field Mus. Nat. Hist. Bot. Ser. 19, 1–613.
- Kinupp, V.F., Lorenzi, H. (2014) Plantas Alimentícias Não-Convencionais (PANC) no Brasil: guia de identificação, aspectos nutricionais e receitas ilustradas. Plantarum, Nova Odessa, p 768
- Kinupp, V.F., Lisbôa, G.N., Barros, I.B.I. (2011) *Tropaeolum pentaphyllum*, Batata-crem. In: Coradin, L., Siminski, L.C., Reis, A. (eds) Espécies nativas da flora brasileira de valor econômico atual ou potencial: plantas para o futuro—Região Sul. Ministério do Meio Ambiente, Brasília, pp 243–250
- Krosnick, S.E., Ford, A.J., Freudenstein, J.V. (2009) Taxonomic revision of *Passiflora* subgenus *Tetrapathea* including the monotypic genera *Hollrungia* and *Tetrapathea* (Passifloraceae), and a new species of *Passiflora*. *Syst Bot* 34, 375–385
- Krosnick, S.E., Porter-Utley, K.E., MacDougal, J.M., Jørgensen, P.M., McDade, L.A. (2013) New insights into the evolution of *Passiflora* subgenus *Decaloba* (Passifloraceae): phylogenetic relationships and morphological synapomorphies. *Syst Bot* 38, 692–713.
- Leister, D., Pesaresi, P. (2005) The genomic are of chloroplast research. *Annu. Plant Rev.* 3, 1–29.
- Le Péchon, T., Gigord, L.D.B. (2014) On the relevance of molecular tools for taxonomic revision in Malvales, Malvaceae s.l., and Dombeyoideae. In: Besse P (ed) *Molecular plant taxonomy. Methods in molecular biology (methods and protocols)*, vol 1115. Humana Press, Totowa.
- Li, Qj., Su, N., Zhang, L., Tong, R., Zhang, X., Wang, J., Chang, Z., Zhao, L., Potter, D. (2020) Chloroplast genomes elucidate diversity, phylogeny, and taxonomy of *Pulsatilla* (Ranunculaceae). *Sci. Rep.* 10, 19781.
- Levin, D.A. (2003) The cytoplasmic factor in plant speciation. *Syst. Bot.* 28, 5–11.
- Lopes, A.S., Pacheco, T.G., Santos, K.G., Vieira, L.N., Guerra, M.P., Nodari, R.O., Souza, E.M., Oliveira, F.P., Rogalski, M. (2017) The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Rep.* 37, 307–328.
- Lopes, A.S., Pacheco, T.G., Nimz, T., Vieira, L.N., Guerra, M.P., Nodari, R.O., Souza, E.M., Oliveira, F.P., Rogalski, M. (2018a) The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae. *Planta.* 247, 1–20.
- Lopes, A.S., Pacheco, T.G., Vieira, L.N., Guerra, M.P., Nodari, R.O., Souza, E.M., Pedrosa, F.O., Rogalski, M. (2018b) The *Crambe abyssinica* plastome: Brassicaceae phylogenomic analysis, evolution of RNA editing sites, hotspot and microsatellite characterization of the tribe Brassiceae. *Gene* 671, 36–49.
- Lopes, A.S., Pacheco, T.G., Silva, O.N., Magalhães-Cruz, L., Balsanelli, E., Maltempo de Souza, E., de Oliveira, P.F., Rogalski, M. (2019) The plastomes of *Astrocaryum aculeatum* G.

- Mey. and A. murumuru Mart. show a fip-fop recombination between two short inverted repeats. *Planta* 250, 1229–1246.
- Lu, Y., Rijzaani, H., Karcher, D., Ruf, S., Bock, R. (2013) Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proc. Natl. Acad. Sci.* 110, E623–E632.
- MacDougal, J.M., Feuillet, C. (2004) Systematics. In: Ulmer T, MacDougal JM (eds) *Passiflora: passionflowers of the world*. Timber, Portland, pp 27–31.
- Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G., Soltis, D.E. (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci.* 107, 4623–4628.
- Mors, W.B., Rizzini, C.T., Pereira, N. (2000) *A medicinal plants of Brazil*. Reference Publications, Michigan, 501p.
- Mráček, J. (2005) Investigation of genome–plastome incompatibility in *Oenothera* and *Passiflora*. PhD Thesis, Ludwig Maximilians-University, Munich, 110 pp.
- Ocampo, J., Arias, J.C., Urrea, R. (2016) Interspecific hybridization between cultivated and wild species of genus *Passiflora* L. *Euphytica*. 209, 395–408.
- Oleksyk, T.K., Smith, M.W., O'Brien, J. (2010) Genome-wide scans for footprints of natural selection. *Phil. Trans. R. Soc. B.* 365, 185–205.
- Piot, A., Hackel, J., Christin, P-A., Besnard, G. (2018) One-third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta* 247, 255–266.
- Pinard, D., E, Mizrachi. (2018) *Unsung and understudied: plastid involved in secondary growth*. *Curr. Opin. Plant Biol.* 42, 30–36.
- Powell, W., Morgantet, M., Andre, C., McNicol, J.W., Machray, G.C., Doyle, J.J., Tingey, S.V., Rafalski, J.A. (1995) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr. Biol.* 5, 1023–1029.
- Provan, J., Powell, W., Hollingsworth, P.M. (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol. Evol.* 16, 142–147.
- Pyke, K. (2007) Plastid biogenesis and differentiation. In: *Cell and Molecular Biology of Plastids*, Bock R (Ed.) *Top. Curr. Genet.* Springer-Verlag, Berlin Heidelberg.
- Rabah, S.O., Shrestha, B., Hajrah, N.H., Sabir, M.J., Alharby, H.F., Sabir, M.J., Alhebshi, A.M., Sabir, J.S.M., Gilbert, L.E., Ruhlman, T.A., Jansen, R.K. (2019) *Passiflora* plastome sequencing reveals widespread genomic rearrangements. *J. Syst. Evol.* 57, 1–14.
- Raddatz-Mota, D., Pérez-Flores, L.J., Carrari, F., Mendoza-Espinoza, J.A., de León-Sánchez, F.D., Pinzón-López, L.L., Godoy-Hernández, G., Rivera-Cabrera, F. (2017) *Achiote (Bixa orellana L.): a natural source of pigment and vitamin E*. *J. Food Sci. Technol.* 54, 1729–1741.

- Rocha, D.I., Batista, D.S., Faleiro, F.G., Rogalski, M., Ribeiro, L.M., Mercadante-Simões, M.O., Yockteng, R., Silva, M.L., Soares, W.S., Pinheiro, M.V.M., Pacheco, T.G., Lopes, A.S., Viccini, L.F., Otoni, W.C. (2020) *Passiflora* spp. Passionfruit. In: Litz, R.E., Alfaro, F.P., Hormaza, J.I. (Org.) *Biotechnology of Fruit and Nut Crops*, 2 ed. CABI, Oxfordshire, pp 381–408
- Rogalski, M., Carrer, H. (2011) Engineering plastid fatty acid biosynthesis to improve food quality and biofuel production in higher plants: Plastid fatty acid biosynthesis. *Plant Biotechnol. J.* 9, 554–564.
- Rogalski, M., Karcher, D., Bock, R. (2008) Superwobbling facilitates translation with reduced tRNA sets. *Nat. Struct. Mol. Biol.* 15, 192–198.
- Rogalski, M., Ruf, S., Bock, R. (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Res.* 34, 4537–4545.
- Rogalski, M., Vieira, L. do N., Fraga, H.P., Guerra, M.P. (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front. Plant Sci.* 6.
- Sanderson, M.J., Copetti, D., Búrquez, A., Bustamante, E., Charboneau, J.L.M., Eguiarte, L.E., Kumar, S., Lee, H.O., Lee, J., McMahon, M., Steele, K., Wing, R., Yang, T-J., Zwickl, D., Wojciechowski, M.F. (2015) Exceptional reduction of the plastid genome of Saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat. *Am. J. Bot.* 102, 1115–1127
- Schmitz-Linneweber, C., Kushnir, S., Babiychuk, E., Poltnigg, P., Herrmann, R.G., Maier, R.M. (2005) Pigment Deficiency in Nightshade/Tobacco Cybrids Is Caused by the Failure to Edit the Plastid ATPase -Subunit mRNA. *Plant Cell* 17, 1815–1828.
- Schwarz, E.N., Ruhlman, T.A., Sabir, J.S.M., Hajrah, N.H., Alharbi, N.S., Al-Malki, A.L., Bailey, C.D., and Jansen, R.K. (2015) Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids: Parallel inversions and *rps16* losses in legumes. *J. Syst. Evol.* 53, 458–468.
- Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E., Small, R.L. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142–166.
- Shaw, J., Lickey, E.B., Schilling, E.E., Small, R.L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94, 275–288.
- Shrestha, B., Weng, M.L., Therio, 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*. *Mol. Phylogenet. Evol.* 138, 53–64.
- Solórzano, S., Chincoya, D.A., Sanchez-Flores, A., Estrada, K., Díaz-Velásquez, C.E., González-Rodríguez, A., Vaca-Paniagua, F., Dávila, P., Arias, S. (2019) De novo assembly

discovered novel structures in g of plastids and revealed divergent inverted repeats in *Mammillaria* (Cactaceae, Caryophyllales). *Plants* 8, 392.

Souza, V.C., Lorenzi, H. (2008) *Botânica Sistemática: Guia ilustrado para identificação das famílias de Fanerógamas nativas do Brasil*. Instituto Plantarum, Nova Odessa.

Tamburino, R., Sannino, L., Cafasso, D., Cantarella, C., Orrù, L., Cardi, T., Cozzolino, S., D'Agostino, N., Scotti, N. (2020) Cultivated Tomato (*Solanum lycopersicum* L.) Suffered a Severe Cytoplasmic Bottleneck during Domestication: Implications from Chloroplast Genomes. *Plants* 9, 1443.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., Brennicke, A. (2013) RNA Editing in Plants and Its Evolution. *Annu. Rev. Genet.* 47, 335-352.

Teshome, G.E., Mekbib, Y., Hu, G., Li, Z.Z., Chen, J. (2020) Comparative analyses of 32 complete plastomes of *Tef* (*Eragrostis tef*) accessions from Ethiopia: phylogenetic relationships and mutational hotspots. *PeerJ.* 8, e9314.

Tetlow, I. J., Morell, M. K., Emes, M. J. (2004) Recent developments in understanding the regulation of starch metabolism in higher plants. *J. Exp. Bot.* 55, 2131–2145.

Tillich, M., Funk, H.T., Schmitz-Linneweber, C., Poltnigg, P., Sabater, B., Martin, M., Maier, R.M. (2005) Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J.* 43, 708–715.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* 5, 123–135.

Tonti-Filippini, J., Nevill, P.G., Dixon, K., Small, I. (2017) What can we do with 1000 plastid genomes? *Plant J. Cell Mol. Biol.* 90, 808–818.

Tsudzuki, T., Wakasugi, T., Sugiura, M. (2001) Comparative analysis of RNA editing sites in higher plant chloroplasts. *J. Mol. Evol.* 53, 327–332

Vieira, L. do N., Faoro, H., Rogalski, M., Fraga, H.P. de F., Cardoso, R.L.A., de Souza, E.M., de Oliveira Pedrosa, F., Nodari, R.O., Guerra, M.P. (2014) The Complete Chloroplast Genome Sequence of *Podocarpus lambertii*: Genome Structure, Evolutionary Aspects, Gene Content and SSR Detection. *PLoS ONE* 9, e90618.

Vieira, L. do N., Dos Anjos, K.G., Faoro, H., Fraga, H.P., Greco, T.M., Pedrosa, F. de O., de Souza, E.M., Rogalski, M., de Souza, R.F., Guerra, M.P. (2015) Phylogenetic inference and SSR characterization of tropical woody bamboos tribe Bambuseae (Poaceae: Bambusoideae) based on complete plastid genome sequences. *Curr. Genet.* 62, 443-53.

Vieira, L. do N., Rogalski, M., Faoro, H., Fraga, H.P.F., Anjos, K.G., Picchi, G.F.A., Nodari, R.O., de Pedrosa, F.O., Souza, E.M., Guerra, M.P. (2016) The plastome sequence of the endemic Amazonian conifer, *Retrophyllum piresii* (Silba) C.N.Page, reveals different recombination events and plastome isoforms. *Tree Genet. Genomes.* 12.

Vilar, D., Vilar, M. S., de Lima e Moura, T. F., Raffin, F. N., de Oliveira, M. R., Franco, C. F., de Athayde-Filho, P. F., Diniz, M., & Barbosa-Filho, J. M. (2014) Traditional uses, chemical constituents, and biological activities of *Bixa orellana* L.: a review. *Sci. World. J.* 2014, 857292

Wambugu, P.W., Brozynska, M., Furtado, A., Waters, D.L., Henry, R.J. (2015) Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences. *Sci. Rep.* 5.

Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Müller, K.F., Quandt, D. (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297.

Xi, Z., Ruhfel, B.R., Schaefer, H., Amorim, A.M., Sugumaran, M., Wurdack, K.J., Endress, P.K., Matthews, M.L., Stevens, P.F., Mathews, S., Davis, C.C. (2012) Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc. Natl. Acad. Sci.* 109, 17519–17524.

Zimorski, V., Ku, C., Martin, W.F., and Gould, S.B. (2014) Endosymbiotic theory for organelle origins. *Curr. Opin. Microbiol.* 22, 38–48.

Zhu, A., Guo, W., Gupta, S., Fan, W., Mower, J.P. (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol.* 209, 1747–1756.

Chapter I

Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (*Bixa Orellana* L.), the Amazonian commercial species of natural dyes

Túlio Gomes Pacheco¹, Amanda de Santana Lopes¹, Gélia Dinah Monteiro Viana¹, Odyone Nascimento da Silva¹, Gleyson Morais da Silva¹, Leila do Nascimento Vieira², Miguel Pedro Guerra², Rubens Onofre Nodari², Emanuel Maltempi de Souza³, Fábio de Oliveira Pedrosa³, Wagner Campos Otoni⁴, Marcelo Rogalski^{1*}

¹ Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa-MG, Brazil.

² Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal, Programa de Pós-graduação em Recursos Genéticos Vegetais, Universidade Federal de Santa Catarina, Florianópolis-SC, Brazil.

³ Departamento de Bioquímica e Biologia Molecular, Núcleo de Fixação Biológica de Nitrogênio, Universidade Federal do Paraná, Curitiba-PR, Brazil.

⁴ Laboratório de Cultura de Tecidos Vegetais, Departamento de Biologia Vegetal, BIOAGRO, Universidade Federal de Viçosa, Viçosa-MG, Brazil.

*Corresponding author.

E-mail address: rogalski@ufv.br

Published in:

Planta (2019) 249, 563-582

<https://doi.org/10.1007/s00425-018-3023-6>



Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (*Bixa orellana* L.), the Amazonian commercial species of natural dyes

Túlio Gomes Pacheco¹ · Amanda de Santana Lopes¹ · Gélia Dinah Monteiro Viana¹ · Odyone Nascimento da Silva¹ · Gleyson Morais da Silva¹ · Leila do Nascimento Vieira² · Miguel Pedro Guerra² · Rubens Onofre Nodari² · Emanuel Maltempi de Souza³ · Fábio de Oliveira Pedrosa³ · Wagner Campos Otoni⁴ · Marcelo Rogalski¹

Received: 16 June 2018 / Accepted: 1 October 2018

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Main conclusion The plastome of *B. orellana* reveals specific evolutionary features, unique RNA editing sites, molecular markers and the position of Bixaceae within Malvales.

Annatto (*Bixa orellana* L.) is a native species of tropical Americas with center of origin in Brazilian Amazonia. Its seeds accumulate the apocarotenoids, bixin and norbixin, which are only found in high content in this species. The seeds of *B. orellana* are commercially valued by the food industry because its dyes replace synthetic ones from the market due to potential carcinogenic risks. The increasing consumption of *B. orellana* seeds for dye extraction makes necessary the increase of productivity, which is possible accessing the genetic basis and searching for elite genotypes. The identification and characterization of molecular markers are essential to analyse the genetic diversity of natural populations and to establish suitable strategies for conservation, domestication, germplasm characterization and genetic breeding. Therefore, we sequenced and characterized in detail the plastome of *B. orellana*. The plastome of *B. orellana* is a circular DNA molecule of 159,708 bp with a typical quadripartite structure and 112 unique genes. Additionally, a total of 312 SSR loci were identified in the plastome of *B. orellana*. Moreover, we predicted in 23 genes a total of 57 RNA-editing sites of which 11 are unique for *B. orellana*. Furthermore, our plastid phylogenomic analyses, using the plastome sequences available in the plastid database belonging to species of order Malvales, indicate a closed relationship between Bixaceae and Malvaceae, which formed a sister group to Thymelaeaceae. Finally, our study provided useful data to be employed in several genetic and biotechnological approaches in *B. orellana* and related species of the family Bixaceae.

Keywords Bixaceae · Cytoplasmic inheritance · Plastid SSRs · Polymorphism hotspots · Gene divergence · Plastid RNA editing

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00425-018-3023-6>) contains supplementary material, which is available to authorized users.

✉ Marcelo Rogalski
rogalski@ufv.br

¹ Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa, MG, Brazil

² Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal, Programa de Pós-graduação em Recursos Genéticos Vegetais, Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil

³ Departamento de Bioquímica e Biologia Molecular, Núcleo de Fixação Biológica de Nitrogênio, Universidade Federal do Paraná, Curitiba, PR, Brazil

⁴ Laboratório de Cultura de Tecidos Vegetais, Departamento de Biologia Vegetal, BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG, Brazil

Introduction

Annatto (*Bixa orellana* L.) is a native species of tropical Americas, with probable center of origin in Brazilian Amazonia, which contains the greatest diversity of this species (Clement et al. 2010; Moreira et al. 2015). The seeds of *B. orellana* accumulate apocarotenoid, bixin and norbixin, which are oil- and water-soluble natural dyes, respectively (Giuliano et al. 2003; Rivera-Madrid et al. 2013). The pigments of *B. orellana* seeds have been used since pre-Columbian period as textile, body and food dyes (Giuliano et al. 2003; de Araújo Vilar et al. 2014). Brazil is the major producer of *B. orellana* seeds with approximately 70% of world production, which originates mostly from family farmers in Northeast Brazil (Moreira et al. 2015). The seeds of *B. orellana* are the main commercial source of these natural dyes (i.e. bixin and norbixin), which are found in other species such as saffron, grape and *Costus pictus* but at very low levels which are not viable for commercialization (Ramamoorthy et al. 2010; Annadurai et al. 2012). Therefore, *B. orellana* seeds are commercially valued by the food industry because its dyes replace synthetic ones from the market due to potential carcinogenic risks, which increased its demand during the last century (Leal and Clavijo 2010; Guesmi et al. 2013).

The unique attractive feature of *B. orellana* seeds to accumulate dyes has aroused the interest of several areas of research to understand, characterize and increase the pigment biosynthesis in *B. orellana* or other targeted organisms of interest. Studies in several areas have been carried out with this purpose, including the morphological characterization of fruits, seed productivity and bixin content (Mantovani et al. 2013), identification of the genes involved in the biosynthesis of apocarotenoids (Rivera-Madrid et al. 2013; Cárdenas-Conejo et al. 2015; Sankari et al. 2016), expression pattern of pigment-related genes (Rodríguez-Ávila et al. 2011; Rivera-Madrid et al. 2013; Cárdenas-Conejo et al. 2015), molecular characterization of pigments (Giuliano et al. 2003; Ramamoorthy et al. 2010; Rivera-Madrid et al. 2016), transfer of bixin biosynthesis to bacteria (Bouvier et al. 2003), in vitro organogenesis and somatic embryogenesis (Parimalan et al. 2011; Da Cruz et al. 2015; de Matos et al. 2016), polyploidy induction (Portela de Carvalho et al. 2005) and genetic transformation of *B. orellana* plants (Zaldívar-Cruz et al. 2003; Giridhar and Parimalan 2010; Parimalan et al. 2011; Zhai et al. 2014; Sankari et al. 2016).

The increasing consumption of *B. orellana* seeds for dye extraction makes it necessary to improve characteristics of interest of the species as well as the search for elite genotypes to increase plant productivity. Since *B. orellana* is a cross-pollinated plant species, a high heterozygosity

is induced in natural plants, which implicates directly in plant productivity if the plants are propagated by seeds (Rivera-Madrid et al. 2006; Teixeira da Silva et al. 2018). The identification and characterization of molecular markers are essential to analyse the genetic diversity of natural populations and establish suitable strategies for conservation, domestication, germplasm characterization and genetic breeding. Several molecular markers were developed for *B. orellana* such as isoenzymes (Carvalho et al. 2005) and microsatellites (SSRs) (Dequigiovanni et al. 2014, 2018), but all of them are based on nuclear sequences. Extranuclear and uniparental inheritance of plastome makes it a useful source of molecular markers, particularly intergenic spacers (IGSs) and introns, which show high rates of mutation in comparison with coding sequences (Rogalski et al. 2015; Smith 2015; Park et al. 2017). Plastid SSR markers have been used in several genetic studies in natural populations (Provan et al. 2001; Ebert and Peakall 2009; Wheeler et al. 2014a). High-polymorphic plastid sequences are useful data to study natural population and germplasm collections regarding genetic diversity, haplotypes, lineages, demography and plant material origin (Tsai et al. 2015; Wambulwa et al. 2016; Roy et al. 2016).

Complete plastome sequences allow the understanding of several evolutionary events in plants such as gene content and function (Rogalski et al. 2015; Daniell et al. 2016), recombination events (Gurdon and Maliga 2014; Vieira et al. 2016b), genome rearrangements (Lopes et al. 2018c; Ruhlman et al. 2017), RNA editing sites (Labiak and Karol 2017; Lopes et al. 2018a), gene transfer to the nucleus (Lu et al. 2017; Bock 2017), positive selection (Piot et al. 2018; Lopes et al. 2018b) and phylogenetic relationships (Comer et al. 2015; Vieira et al. 2016a; Barrett et al. 2016). Complete plastome sequences are also important for plastid transformation because plastomes of higher plants are extremely dense in gene content and have small intergenic regions containing the untranslated regions (5'UTRs and 3'UTRs), which control gene expression and RNA stability (Shinozaki et al. 1986; Wakasugi et al. 2001). Therefore, the site for the correct insertion of the transgenes into the plastome may represent a serious limitation for stable transgene expression into the plastid DNA (Bock 2013; Krech et al. 2013; Daniell et al. 2016; Fuentes et al. 2017). The plastid transformation is a very interesting and attractive tool for biotechnological applications (Daniell et al. 2016; Zhang et al. 2017) and basic research regarding plastid gene-function (Rogalski et al. 2006, 2008; Alkatib et al. 2012). Plastid genetic engineering has been used efficiently to manipulate several metabolic pathways such as carotenoids, tocopherols, artemisinic acid, astaxanthin (Bock 2015; Daniell et al. 2016; Fuentes et al. 2017) and fatty acid biosynthesis (Rogalski and Carrer 2011). It is a very interesting technology to study and/or to manipulate the apocarotenoid biosynthesis

such as norbixin and bixin given that several efficient protocols for organogenesis and somatic embryogenesis are available for *B. orellana* (Paiva Neto et al. 2003; Parimalan et al. 2011; Da Cruz et al. 2015; de Matos et al. 2016).

B. orellana belongs to the family Bixaceae, which contains four genera (*Amoreuxia*, *Bixa*, *Cochlospermum* and *Diego-dendron*) encompassing 23 species (The Plant List 2013). The genus *Bixa* has five species *B. arborea*, *B. excelsa*, *B. orellana*, *B. platycarpa* and *B. urucurana*. The family Bixaceae belongs to order Malvales, which includes other nine families such as Cistaceae, Cytinaceae, Dipterocarpaceae, Malvaceae, Muntingiaceae, Neuradaceae, Sarcocaulaceae, Sphaerosepalaceae and Thymelaeaceae distributed among seven monophyletic lineages (Le Péchon and Gigord 2014). However, the relationships of Bixaceae into the order Malvales and the genera into the family Bixaceae are still molecularly unclear (Fay et al. 1998; Savolainen et al. 2000; Soltis et al. 2000). Therefore, complete plastome sequences of species belonging to the family Bixaceae will be basic and fundamental information to reveal the relationship between Bixaceae and other families of Malvales and the species belonging to Bixaceae since the current complete lack of plastome sequences of this family in organelle database.

Here, we reported the plastome sequence of annatto (*B. orellana* L.), the first species of the family Bixaceae to have the plastome completely sequenced and analysed in detail. The plastome of *B. orellana* is a circular DNA molecule of 159,708 bp with a typical quadripartite structure. Additionally, we identified 112 unique genes and one pseudogene (*InfA*). Moreover, 57 RNA editing sites were predicted for *B. orellana* genes, of which 11 sites seem to be unique for this species. Furthermore, 312 SSRs were identified in the *B. orellana* plastome, which can be used for several polymorphism analyses aiming different genetic approaches in *B. orellana* and family Bixaceae. Finally, our plastid phylogenomic analyses, using the plastome sequences available in the plastid database for species/families belonging to order Malvales, resolved phylogenetic incongruences present in the order Malvales with high support and indicate a closed relationship between Bixaceae and Malvaceae, which formed a sister group to the family Thymelaeaceae. Taken together, our data bring new molecular markers, several structural and evolving features useful for genetic, evolutionary and biotechnological studies in *B. orellana* and Bixaceae.

Materials and methods

Plant material and plastid-DNA extraction

Bixa orellana fresh leaves were collected and kept for 1 week at 4 °C to decrease the starch content. The leaves

were collected from a *B. orellana* plant genotype 2, which is maintained in an experimental field at the Federal University of Viçosa, Viçosa-MG, Brazil (Mantovani et al. 2013). This genotype has been used in several experiments related to tissue culture given that it is highly responsive to somatic embryogenesis (Paiva Neto et al. 2003; Portela de Carvalho et al. 2005; de Matos et al. 2016). The chloroplast isolation and plastid-DNA extraction were carried out according to Vieira et al. (2014).

Plastid genome sequencing, assembling and annotation

Sequencing library was prepared with approximately 1 ng of plastid DNA using the sample preparation kit NexteraXT (Illumina Inc., San Diego, CA), according to the manufacturer's instructions. The obtained library was sequenced using MiSeq Reagent Kit v3 (600 cycles) on Illumina MiSeq Sequencer (Illumina Inc., San Diego, California, USA). The paired-end reads sequenced (3,177,428 reads with average length of 255.9 bp) were trimmed under the threshold with probability of error < 0.05. The trimmed reads (3,057,961 reads, average length of 250.5 bp) were de novo assembled using CLC Genomics Workbench 8.0.2 software (CLC Bio, Aarhus, Denmark). The three contigs used for assembling of *B. orellana* plastome ranged from 5044 to 2104 of average coverage. The initial annotation of the *B. orellana* plastome was carried out using Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004) and BLAST searches. From this initial annotation, putative start codons, stop codons and intron positions were determined based on comparisons to homologous genes in other plastid genomes. All tRNA genes were further verified using tRNAscan-SE (Lowe and Eddy 1997). A physical map of the circular plastome was drawn using OrganellarGenomeDRAW (OGDRAW) (Lohse et al. 2013), followed by manual modification. The complete plastome sequence of *B. orellana* was deposited in the GenBank database under accession number MH025909.

Analysis of genome structure, repeat sequences and RNA editing sites

The comparison of plastome structure between *B. orellana* and other representatives of Malvales was carried out using the Mauve Genome Alignment v2.3.1 (MAUVE) software (Darling et al. 2004).

Simple sequence repeats (SSRs) of the *B. orellana* plastome were characterized using the MIncroSAteLLite (MISA) Perl script (Thiel et al. 2003), with thresholds of eight repeat units for mononucleotide SSRs, four repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta- and hexanucleotide SSRs. Additional tandem repeats were analysed using the Tandem Repeats Finder (TRF) software

(Benson 1999) with parameter settings of 2, 7 and 7 for match, mismatch, and indel, respectively. The minimum alignment score and maximum period size were set to 50 and 500, respectively. The exact localization of inverted repeats (IRs) and the analysis of forward versus reverse complement (palindromic) within *B. orellana* plastome were carried out using the software REPuter (Kurtz et al. 2001). The setting for minimal repeat size was 30 bp and the identity of repeats was determined to be no less than 90% (hamming distance = 3). All repeats found in the plastome were manually verified and nested or redundant results were removed.

Potential RNA editing sites in protein-coding genes of *B. orellana* plastome were predicted using the program Predictive RNA Editor for Plants (PREP) suite (Mower 2009) that use the reference of 35 plastid genes for detecting RNA editing sites in plastomes. The cutoff value was set to 0.8 and the reference genes were *accD*, *atpA*, *atpB*, *atpF*, *atpI*, *ccsA*, *clpP*, *matK*, *ndhA*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *petB*, *petD*, *petG*, *petL*, *psaB*, *psaI*, *psbB*, *psbE*, *psbF*, *psbL*, *rpl2*, *rpl20*, *rpl23*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps2*, *rps8*, *rps14*, *rps16* and *ycf3*. Aiming the evolutionary comparison with members of family Malvaceae, RNA editing sites were also predicted for several available species belonging to this family, using the same parameters.

Phylogenomic inference

The inference of the *B. orellana* phylogenetic position within Malvales was carried out using two phylogenomic approaches. Firstly, whole plastome sequences of species representing each genus of families Bixaceae, Malvaceae and Thymelaeaceae belonging to the order Malvales, available in the plastid database of GenBank (with exception of *Cytinus hypocistis*, a holoparasitic species containing a drastically reduced plastome), were selected and extracted. The plastome of *Carica papaya* (order Brassicales) was also extracted and used as an outgroup species. The GenBank accession number of each taxon used here is shown in the Supplementary Table S1. A LSC inversion present in *A. sinensis* plastome was reoriented to the position found in the plastomes of other species. The SSC regions were changed (if necessary) to the same orientation as seen in the *B. orellana* plastome. Furthermore, the IR_B was withdrawn to prevent overrepresentation of the IR sequences. The alignment of plastome sequences was done using the software MAFFT v.7 (Katoh and Standley 2013) and the best substitution model (GTR + I + G) was selected using the program jModelTest v.2.1.7. The Bayesian inference was performed using MrBayes version 3.2 (Ronquist et al. 2012), with 1 million generations of two runs of four Markov Chains, three hot and one cold for each run. Furthermore, a maximum likelihood tree was also reconstructed with the same dataset, using IQTREE v1.6.6 (Nguyen et al. 2015)

and 1000 non-parametric bootstrap replications were used to assess branch support. Lastly, the consensus trees generated by these methods were visualized in FigTree v.1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Analysis of synonymous (dS) and non-synonymous (dN) substitution rates of plastid genes

Pairwise dS and dN substitution rates between *B. orellana* and other representatives of Malvales (shown in the Supplementary Table S1), were estimated to 73 protein-coding genes commonly found in the plastomes of all species analysed here. Firstly, the sequences of each gene were aligned individually by MUSCLE (Edgar 2004) implemented in MEGA 6.0 (Tamura et al. 2013), with pairwise deletion set to gaps/missing data treatment. To each alignment were then calculated the dS and dN values using MEGA under the Kumar model (Kimura 2-parameter).

Results

Size, organization and gene content of *B. orellana* plastome

The *B. orellana* plastome is a circular DNA molecule of 159,708 bp in length and exhibits the general quadripartite structure, which is typical in most angiosperms. It includes a large single-copy region (LSC) of 89,377 bp flanked on each side by inverted repeats (IRs) of 25,356 bp with a small single-copy region (SSC) of 19,619 bp joining the IRs (Fig. 1). The total size of plastome and its regions is very similar between *B. orellana* and other representative species of Malvales, with *D. kiusiana* being an exception (Table 1). The overall GC content determined in the *B. orellana* plastome is 36.4%, which resembles other flowering plants.

It contains 112 unique genes, of which 16 are completely duplicated, and the *rps12* gene is partially duplicated in the IRs given that its transcripts consist of two exons in the IRs and one exon in the LSC, and undergo a trans-splicing event during mRNA maturation. The *B. orellana* plastome contains a total of 129 genes (Table 2), which include 78 unique protein-coding genes (five are completely and one is partially duplicated), 30 unique tRNA genes (seven duplicated) and 4 unique rRNA genes (all of them duplicated). Among the 112 unique genes, 15 possess one intron (six tRNA genes and nine protein-coding genes) and two contain two introns (*clpP* and *ycf3* genes). Additionally, this plastome contains a pseudogene (*infA*), which was lost from plastome in several lineages of angiosperms. A specific feature of *B. orellana* plastome is the loss of *atpF* intron (Table 2), whereas it is highly conserved in other species of Malvales as well as in most angiosperms.

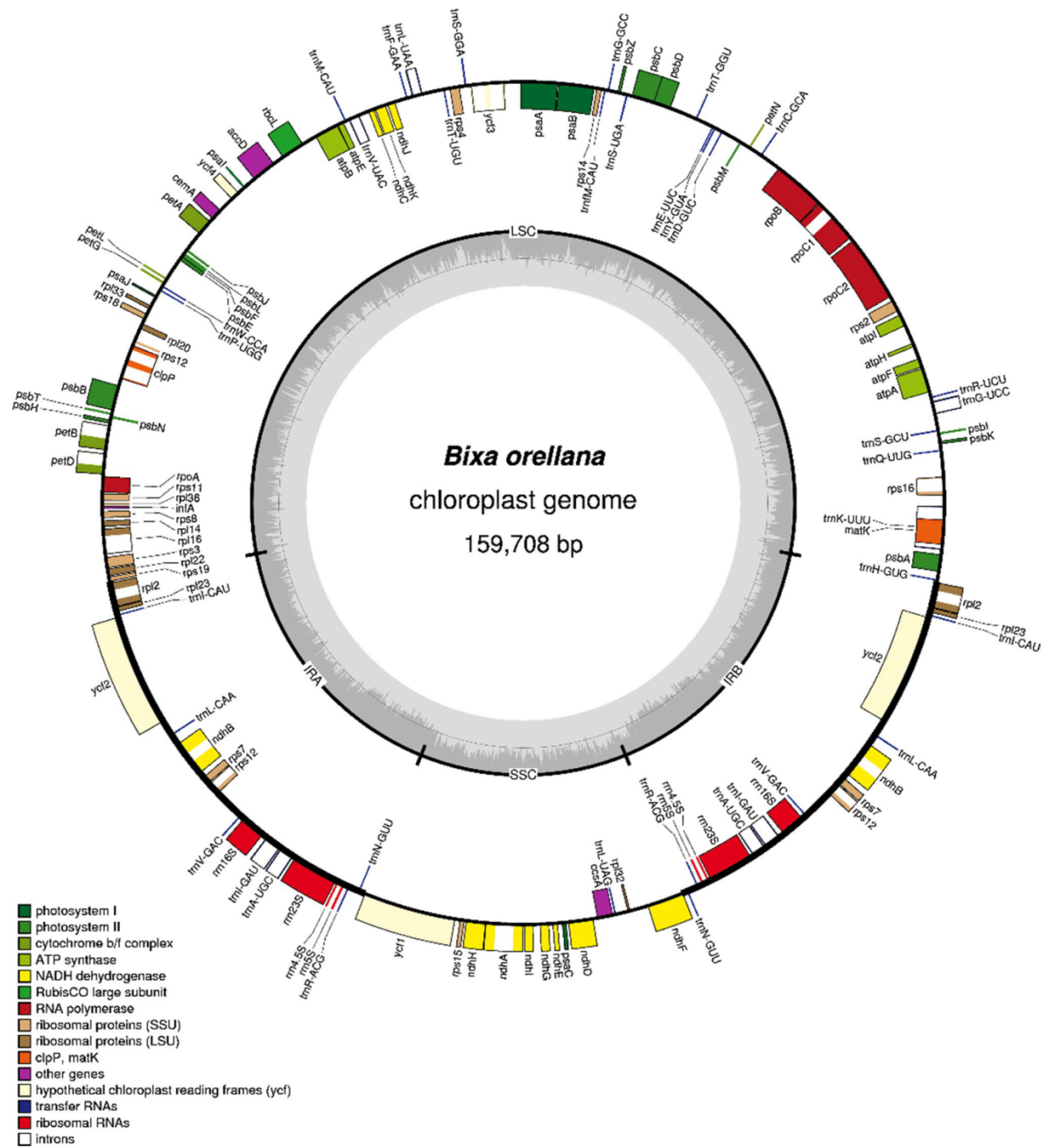


Fig. 1 Gene map and genome organization of *B. orellana* plastome. Two inverted repeat-regions IR_A and IR_B divide the rest of the circular DNA molecule into large (LSC) and small (SSC) single-copy regions. Genes drawn inside the circle are transcribed clockwise and genes drawn outside are expressed counterclockwise. Genes belong-

ing to different functional groups are color-coded. The darker gray in the inner circle corresponds to GC content, while the lighter gray corresponds to AT content. Dotted circle corresponds to 50% of AT/GC content

The structure of *B. orellana* plastome was compared with other representatives of Malvales, via the multiple genome alignment analysis produced by the software MAUVE

(Fig. 2). This alignment revealed a high conservation of plastid genome structure and gene order between *B. orellana* and plastomes belonging to other species of Malvales. The only

Table 1 Comparison of plastome size between *B. orellana* and other six species of Malvales

Species	Family	Size (bp)	LSC (bp)	SSC (bp)	IR (bp)
<i>Bixa orellana</i>	Bixaceae	159,708	89,377	19,619	25,356
<i>Firmiana pulcherrima</i>	Malvaceae	159,556	88,444	19,960	25,576
<i>Gossypium hirsutum</i>	Malvaceae	160,301	88,806	20,280	25,602
<i>Tilia mandshurica</i>	Malvaceae	162,796	91,127	20,371	25,649
<i>Theobroma cacao</i>	Malvaceae	160,619	89,333	20,194	25,546
<i>Daphne kiusiana</i>	Thymelaeaceae	171,491	85,028	2,681	41,891
<i>Aquilaria sinensis</i>	Thymelaeaceae	159,565	87,415	19,780	26,185

Table 2 List of genes identified in the *B. orellana* plastome

Group of gene	Name of genes
Gene expression machinery	
Ribosomal RNA genes	<i>rrn16^b</i> ; <i>rrn23^b</i> ; <i>rrn5^b</i> ; <i>rrn4.5^b</i>
Transfer RNA genes	<i>trnA</i> –UGC ^{a,b} ; <i>trnC</i> –GCA; <i>trnD</i> –GUC; <i>trnE</i> –UUC; <i>trnF</i> –GAA; <i>trnM</i> –CAU; <i>trnG</i> –UCC ^a ; <i>trnG</i> –GCC; <i>trnH</i> –GUG; <i>trnI</i> –CAU ^b ; <i>trnI</i> –GAU ^{a,b} ; <i>trnK</i> –UUU ^a ; <i>trnL</i> –CAA ^a ; <i>trnL</i> –UAA ^b ; <i>trnL</i> –UAG; <i>trnM</i> –CAU; <i>trnN</i> –GUU ^b ; <i>trnP</i> –UGG; <i>trnQ</i> –UUG; <i>trnR</i> –ACG ^b ; <i>trnR</i> –UCU; <i>trnS</i> –GCU; <i>trnS</i> –UGA; <i>trnS</i> –GGA; <i>trnT</i> –UGU; <i>trnT</i> –GGU; <i>trnV</i> –GAC ^b ; <i>trnV</i> –UAC ^a ; <i>trnW</i> –CCA; <i>trnY</i> –GUA
Small subunit of ribosome	<i>rps2</i> ; <i>rps3</i> ; <i>rps4</i> ; <i>rps7^b</i> ; <i>rps8</i> ; <i>rps11</i> ; <i>rps12^{ac}</i> ; <i>rps14</i> ; <i>rps15</i> ; <i>rps16^a</i> ; <i>rps18</i> ; <i>rps19</i>
Large subunit of ribosome	<i>rpl2^{ab}</i> ; <i>rpl14</i> ; <i>rpl16^a</i> ; <i>rpl20</i> ; <i>rpl22</i> ; <i>rpl23^b</i> ; <i>rpl32</i> ; <i>rpl33</i> ; <i>rpl36</i>
DNA-dependent RNA polymerase	<i>rpoA</i> ; <i>rpoB</i> ; <i>rpoC1^a</i> ; <i>rpoC2</i>
Genes for photosynthesis	
Subunits of photosystem I (PSI)	<i>psaA</i> ; <i>psaB</i> ; <i>psaC</i> ; <i>psaI</i> ; <i>psaJ</i> ; <i>ycf3^a</i> ; <i>ycf4</i>
Subunits of photosystem II (PSII)	<i>psbA</i> ; <i>psbB</i> ; <i>psbC</i> ; <i>psbD</i> ; <i>psbE</i> ; <i>psbF</i> ; <i>psbH</i> ; <i>psbI</i> ; <i>psbJ</i> ; <i>psbK</i> ; <i>psbL</i> ; <i>psbM</i> ; <i>psbN</i> ; <i>psbT</i> ; <i>psbZ</i>
Subunits of cytochrome b6/f	<i>petA</i> ; <i>petB^a</i> ; <i>petD^a</i> ; <i>petG</i> ; <i>petL</i> ; <i>petN</i>
Subunits of ATP synthase	<i>atpA</i> ; <i>atpB</i> ; <i>atpE</i> ; <i>atpF</i> ; <i>atpH</i> ; <i>atpI</i>
Subunits of NADH dehydrogenase	<i>ndhA^a</i> ; <i>ndhB^{ab}</i> ; <i>ndhC</i> ; <i>ndhD</i> ; <i>ndhE</i> ; <i>ndhF</i> ; <i>ndhG</i> ; <i>ndhH</i> ; <i>ndhI</i> ; <i>ndhJ</i> ; <i>ndhK</i>
Large subunit of Rubisco	<i>rbcL</i>
Other genes	
Maturase	<i>matK^a</i>
Envelope membrane protein	<i>cemA</i>
Subunit of acetyl-CoA carboxylase	<i>accD</i>
C-type cytochrome synthesis gene	<i>ccsA</i>
Subunit of ATP-dependent protease	<i>clpP^a</i>
Component of TIC complex	<i>ycfI</i>
Genes of unknown function	<i>ycf2^b</i>
Pseudogene	<i>infA</i>

^aGenes containing introns^bDuplicated genes^cPartial duplicated gene

exception was an inversion between *rbcL* and *rps12* genes detected in the LSC region in the *A. sinensis* (Thymelaeaceae) plastome (yellow box, Fig. 2). Interestingly, *Daphne kiusiana*, also belonging to the family Thymelaeaceae, did not show the same inversion, but presented an increase of the size of plastome mainly due to a large expansion of its IRs region (Table 1 and blue box in the Fig. 2; described in Cho et al. 2017). On the other hand, the features of the plastome such as the size of IRs (Table 1) and its boundaries are very

similar between *B. orellana* and the species belonging to the family Malvaceae. These regions show only small nucleotide variations in the intergenic regions next to the IR-junctions (Fig. 3). These variations include also the presence of small part of the *rps19* gene in the IRs of *F. pulcherrima* and *T. cacao* and a minor part of the *ycfI* gene in the IRs of *G. hirsutum* and *T. mandshurica*, whereas in other species analysed here, the full sequence of these genes are completely localized outside of the IRs.

Planta

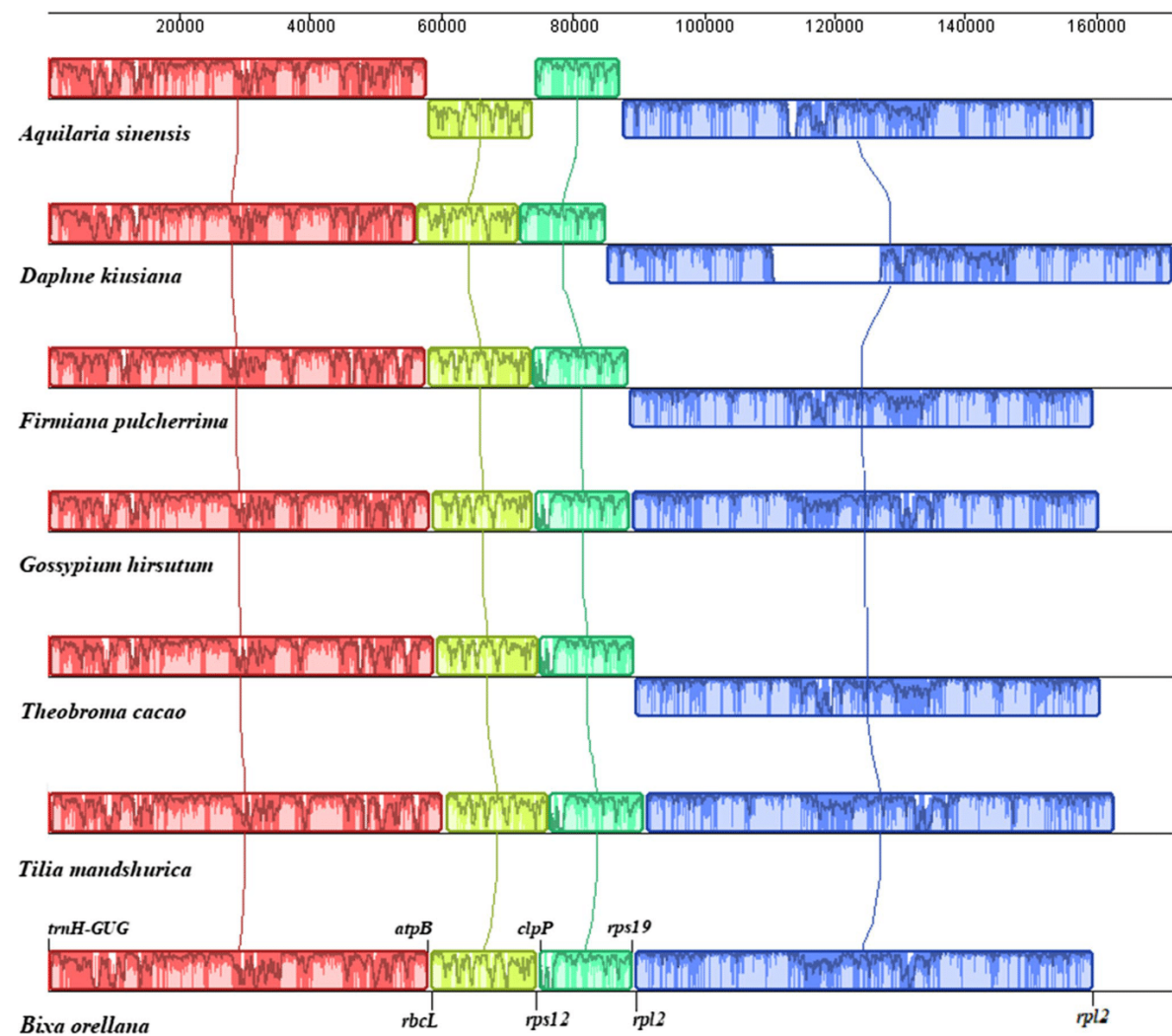


Fig. 2 Comparison of plastome structures between *B. orellana* and other representative species of Malvales, produced by the MAUVE software. The boxes above the line represent the gene complex sequences in clockwise direction and the boxes below the line demonstrate those sequences in the opposite direction. The genes

described at these boxes indicate the genes localized at the boundaries of the gene complex (gene groups). Red, yellow and green boxes correspond to LSC region and blue boxes correspond to both IRs and the SSC region

Repeat sequence analysis

The occurrence, type and distribution of SSRs in the *B. orellana* plastome were extensively analysed here revealing a total of 321 SSRs. Among them, homo- and dipolymers were the most commonly found in this study accounting for 233 and 70, respectively. On the other hand, tri- (5), tetra (14) and pentapolymers (3) occur at low frequency (Table S2). Most of homo- (97.4%) and dipolymer (82.9%) SSRs are constituted by A and T bases. The

size, sequence and location of all SSRs identified here are shown in the Supplementary Table S3. From these polymers, 208 are localized in intergenic spacers (IGS), 36 in coding sequences (CDS) and 39 in introns. The SSRs found in CDS are distributed in 18 genes of which *rpoC2* (7) and *ycf2* (5) contain the highest number. Among the SSRs found in introns, most of them are located in the *clpP* and *rps16* genes, which accounts for 17 and 6 polymers, respectively.

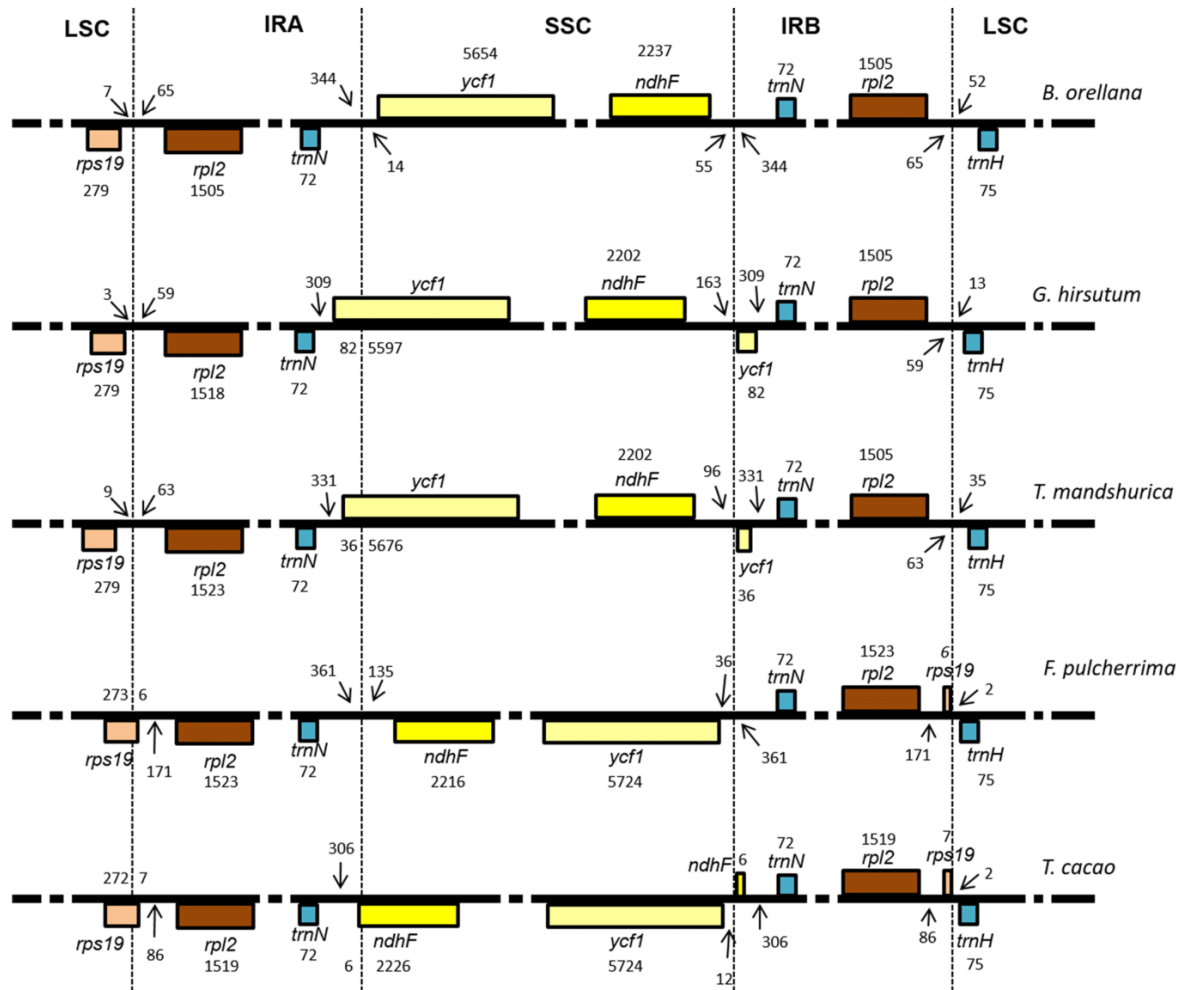


Fig. 3 Comparison of the structure of IR-boundaries between *B. orellana* and four species of family Malvaceae. Genes are represented by boxes with genes above the line being transcribed right to left and those below the line transcribed left to right. Genes belonging to the

same functional groups have the same color. The number of base pairs of the gene or intergenic regions in the IR junctions is indicated the line

Tandem, directed and inverted repeats analysis

Tandem repeats containing more than 30 bp and sequence identity higher than 90% were also characterized here. A total of 19 tandem repeats were found in the *B. orellana* plastome of which 15 are localized in IGS and four in the CDS of *ycf2* gene (Supplementary Table S4). In addition, eight directed repeats and six inverted repeats were found in the *B. orellana* plastome (Supplementary Table S5). Among the directed repeats, six pairs are found in IGS and introns, one pair is located in the CDS of *psaA* and *psaB* genes and another pair is localized in the CDS of *ycf3* gene and the IGS of *rps12/trnV-GAC* genes. Lastly, from the six inverted repeats found here, two pairs are located in the introns of

ycf3 and *ndhA* genes, two pairs were identified in IGS, another pair is localized in the CDS of the *ycf2* and the first intron of *clpP*, and the last pair is part of the *trnS-GGA* and *trnS-GCU* genes.

RNA editing sites prediction

The PREP software identified a total of 57 putative RNA editing sites in the *B. orellana* plastome, which are distributed in 23 genes (Supplementary Table S6). All RNA editing sites occur in the first or the second position of the codon and all changes observed here are from cytidine (C) to uridine (U). Among the genes containing editing sites, the *ndhB* (12 sites), *ndhD* (eight sites), *matK* (five sites),

rpoB (four sites), *atpA* (three sites) and *rps14* (three sites) showed the highest number of editing sites. Other genes with predicted editing sites such as *accD*, *atpB*, *atpF*, *ccsA*, *clpP*, *ndhA*, *ndhF*, *ndhG*, *petB*, *psbE*, *psbF*, *rpl20*, *rpoA*, *rpoC1*, *rpoC2*, *rps2* and *rps16* contain one or two putative editing sites.

The comparison of *B. orellana* plastome with plastomes belonging to eight species of family Malvaceae (Supplementary Table S6, rows 4-11) revealed that 42 out of 57 editing sites predicted for *B. orellana* are shared with these other species. From all specific sites (15) found in the *B. orellana* plastome, 14 of them have a deoxythymidine (T) instead of a deoxycytidine (C) at the referred sites in the plastome of these species, which maintain the conserved amino acid and dispense the need for RNA editing in these codons. The last editing sites predicted here for *B. orellana* convert a serine codon (UCA) to a leucine codon (UUA) in the *matK* gene (amino acid position 403); however, it is not predicted here to occur in other species of family Malvaceae maintaining the serine codons (UCU and UCC) at the referred codons. Curiously, this RNA editing site was a unique predicted here, which encodes different amino acids between *B. orellana* and other species of family Malvaceae. Moreover, another ten RNA editing sites (C-U) were predicted for all

species of family Malvaceae analysed here, except for *B. orellana*, which already has a T instead of a C fixed at these sites in the DNA sequence dispensing the need for editing (Supplementary Table S6).

Phylogenomic inference

The phylogenetic position of *B. orellana* was inferred using whole plastomes of 14 species of Malvales, including the families Bixaceae (represented here by *B. orellana*), Thymelaeaceae (two species) and Malvaceae (11 species distributed in six subfamilies) (Supplementary Table S1). As outgroup, we included a species of Brassicales, *C. papaya* (Caricaceae). Bayesian inference (BI) and maximum likelihood (ML) analyses produced phylogenetic trees with same topology and with lnL equal to $-471,483.7143$ and $-514,854.4539$, respectively (Fig. 4 and Fig. S3). In these phylogenomic reconstructions, *B. orellana* (Bixaceae) was more closely related to the family Malvaceae, with a strong support in our BI analysis (98% of posterior probability, PP; Fig. 4) but with a lower support in the ML tree (62% of bootstrap support, BS; Fig. S3); and the clade composed by the families Bixaceae and Malvaceae formed a sister group to the family Thymelaeaceae.

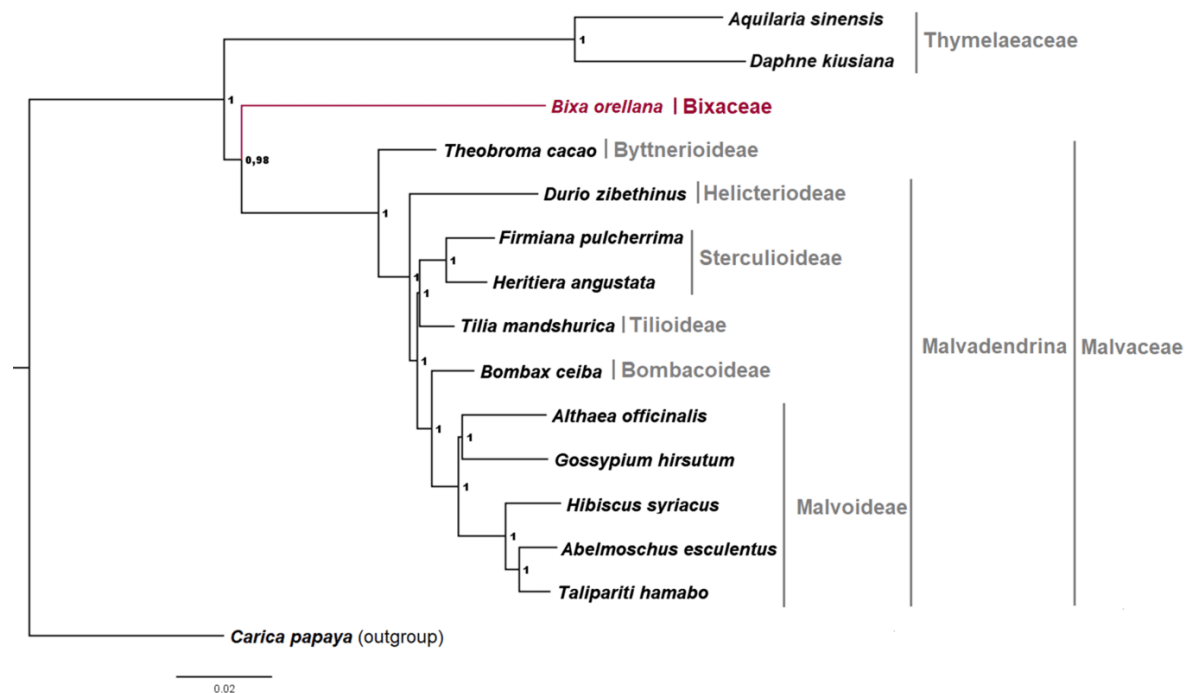


Fig. 4 Bayesian phylogenetic tree based on whole-plastid genomes, showing the position of *B. orellana* (in red) within the order Malvales. Phylogenomic reconstruction was performed using MrBayers with GTR+I+G model. Posterior probabilities (%) are indicated in

front of each node. The branch length is proportional to the inferred divergence level and the scale bar indicates the number of inferred nucleic acid substitutions per site. The subfamilies inside the family Malvaceae are also highlighted

Additionally, the six subfamilies of Malvaceae included in our analysis formed a well-supported monophyletic clade (100% of probability in BI and ML trees). Within the clade Malvaceae, the subfamily Byttnerioideae constituted the early branching lineage, being sister to the clade formed by the other five subfamilies (clade Malvadendrina). Within these five subfamilies, Malvoideae was sister to Bombacoideae and Tilioideae was sister to Sterculioideae, and these four subfamilies constituted a sister group to subfamily Helicterioideae. All clades were strongly supported in our BI and ML trees, showed 100% of probabilities in all of the nodes except for the relationship of the subfamilies Tilioideae and Sterculioideae, which was highly supported in our BI analysis (100% of PP) but was moderately supported in our ML tree (77% of BS). Moreover, the five genera of Malvoideae formed a monophyletic group (with 100% of support in both trees), in which *Hibiscus*, *Abelmoschus* and *Taripariti* formed a sister group to *Gossypium* and *Althaea*. This phylogenetic relationship within subfamily Malvoideae was also highly supported in our analysis, with 100% of probability in all nodes except for one in the ML tree (85% of BS in the *Gossypium* and *Althaea* node).

Synonymous (dS) and non-synonymous (dN) substitution rates

Aiming to analyse the plastid gene divergence between *B. orellana* and other representatives of Malvales (Table S1), synonymous (dS) and non-synonymous (dN) and dN/dS values were estimated for 73 protein-coding genes (Fig. 5). The dN values observed here (first graphic, Fig. 5) were less than 0.05 and similar between the *B. orellana* genes. Similarly, they were also similar between *B. orellana* (filled black circles) and other species of Malvales (open white circles), with few exceptions. The *yef1* and *accD* genes showed the highest dN values in comparison with all other genes, in *B. orellana* (0.14 and 0.11) and other species (0.10 and 0.077), respectively. Other *B. orellana* genes (*psaI*, *rps8*, *rpl20*, *ccsA*, *rps3*, *rpoC2*, *psbK* and *rps15*) also presented dN values above of 0.05, which are higher than those found for the same genes in the other species of Malvales. Outstandingly, the *psaI* gene showed a dN value of 0.11 in *B. orellana*, which is four times higher in comparison with other species of Malvales. The dS values (second graphic, Fig. 5) were also very similar between the genes of *B. orellana* (ranging from 0.01 to 0.22) and the genes of other species of Malvales (ranging from 0.01 to 0.14). Among them we can highlight the *psbI* gene that showed the higher dS value (0.22 in *B. orellana* and 0.15 in other species), whereas *petG* gene presented the smaller dS value (lower than 0.01 in *B. orellana* and in other species). Regarding the dN/dS ratio (third graphic, Fig. 5), most genes showed values lower than 1, as expected of genes under negative selection; however, some

exceptions were observed such as the *yef2*, *rpl23* and *psaI* genes. In the *B. orellana* plastome, the *psaI* gene showed a dN/dS value of 1.92, indicating that this gene may be under positive selection (dN/dS value above of 1), given that its value was lower than 1 (0.63) in other species of Malvales. On the other hand, the *yef2* gene showed a dN/dS value of 1.49 for other species of Malvales and the value was lower than 1 (0.86) in *B. orellana*. Lastly, the *rpl23* gene was a unique case in which dN/dS values were equal or above of 1 in the plastome of *B. orellana* (1.02) and in the plastomes of other species of Malvales (1.60), which suggests that this gene is under relaxed negative selection (or positive selection) in the order Malvales.

Discussion

The *B. orellana* plastome shows small variations in the IR-boundaries and high non-synonymous substitution rates in the *psaI* gene

The *B. orellana* plastome possesses the typical quadripartite structure found in most angiosperm species. All other species of Malvales sequenced to date also have this same structure, with exception of *C. hypocistis*, a holoparasitic species containing a drastically reduced plastome that lacks an IR copy (Roquet et al. 2016). The size of the plastome and its structural regions are very similar to most species of Malvales, with only very small variations (about 200–3000 bp; Table 1). A notable exception is the *D. kiusina* plastome (Thymelaeaceae), which underwent a large IR-expansion resulting in a significant increase in its genome size and a drastic reduction in the SSC region (Table 1 and Fig. 2; also described in Cho et al. 2017). The gene order is also conserved in *B. orellana* and across Malvales, but a unique major inversion is present in the *A. sinensis* (Thymelaeaceae) plastome (yellow box, Fig. 2). Thus, the two species of Thymelaeaceae bear two of the major rearrangements found in the plastomes of Malvales (Wang et al. 2016; Cho et al. 2017). It is also possible to visualize different directions of the blue box between the species analysed here (Fig. 2). However, it is not caused by specific rearrangements, but it is likely originated by the existence of two equimolar states (isomers) of the plastid genome that differ in the orientation of the SSC region between both IRs (Walker et al. 2015). So, the assembling of plastid genomes without preference for the SSC orientation is the reason of this apparent orientation variation in this region.

Regarding the IR-boundaries, they are also highly conserved between *B. orellana* and species of the family Malvaceae; however, several events of small expansion/contraction of the IRs were still detected here (Fig. 3). Normally, most angiosperms include in the edge of IR-boundaries,

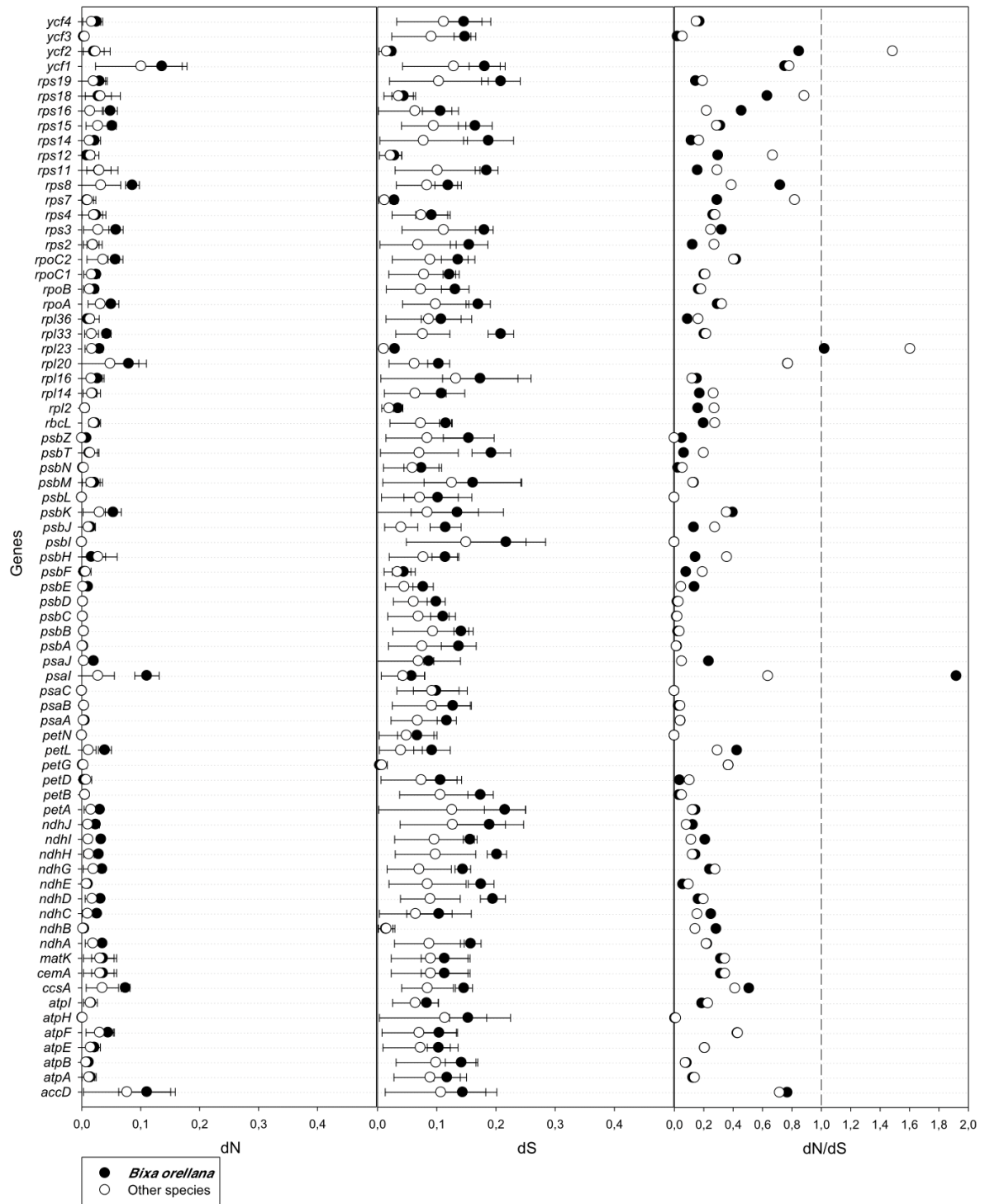


Fig. 5 Non-synonymous (dN), synonymous (dS) and dN/dS values of 73 common plastid protein-coding genes. *B. orellana* is represented by filled circles (black) and the mean of other representatives of Malvaes (see Table 1) is represented by open circles (white)

the *rps19* and *ycf1* genes at the IR/LSC and IR/SSC junctions, respectively (Goulding et al. 1996; Zhu et al. 2016). In *B. orellana*, the full sequence of these genes are localized in the single-copy regions (LSC and SSC) suggesting events of IR-contraction. In the plastomes of *T. cacao* and *F. pulcherrima* a full sequence of *ycf1* gene is also localized in the SSC region; however, a small part of the *rps19* gene is localized in the IRs (Kane et al. 2012; Wang et al. 2018). On the other hand, the plastomes of *G. hirsutum* and *T. mandshurica* have a minor part of the *ycf1* sequence (86 and 32 bp, respectively) in the IRs, while the whole *rps19* gene is localized in the LSC region in these species (Lee et al. 2006; Cai et al. 2015). Therefore, *B. orellana* and the representatives of Malvaceae analysed here seem to have experienced simultaneously common and specific events of IR contraction/expansion resulting in subtle differences in the four IR-boundaries.

The gene number of *B. orellana* plastome resembles that found in most species of Malvales. However, some genes have been lost or have become pseudogenes in some species of this order. Examples of these gene losses include the *rpl22* in *G. hirsutum* and *T. cacao* (Lee et al. 2006; Kane et al. 2012), *clpP* in *F. pulcherrima*, and *rpl32*, *clpP*, *ndhF* and *ndhK* in *D. kiusiana* (Cho et al. 2017), while in *B. orellana*, all these genes show the conserved structure expected for a functional gene. The *B. orellana* plastome contains only one pseudogene (*infA*), which encodes the translation initiation factor 1 being essential in *E. coli* (Cummings and Hershey 1994). The loss of this gene is a common feature found in all species of Malvales sequenced to date and in several other rosids already investigated (Millen et al. 2001). Additionally, nuclear copies of *infA* gene containing plastid transit peptide were identified in some rosids (Millen et al. 2001). Thus, it is likely that in the plastome of *B. orellana* and other species of Malvales, the loss of plastid *infA*-gene was compensated by nuclear copies, which are imported into the plastids.

Concerning the plastid gene evolution in the *B. orellana* plastome, only two genes, *psaI* and *rpl23*, showed exceptional dN/dS values above 1 (Fig. 5), which indicate that most plastid genes in this species are under purifying selection. The *psaI* gene of *B. orellana* showed a dN value of 0.11 and dN/dS value of 1.91, which are approximately four and three times higher in comparison with other species of Malvales, respectively. It suggests that this gene is under positive selection in *B. orellana*. Moreover, the specific amino acid changes in this species include exchange of polarity, always from polar to apolar such as threonine to proline, threonine to isoleucine and serine to phenylalanine (amino acid positions 2, 3 and 28; red arrows, Supplementary Fig. S1). Interestingly, the phenylalanine residue at the position 28 (second arrow, Supplementary Fig. S1) seems to be restored by RNA editing in other species of Malvales (experimentally

confirmed in *G. hirsutum*; Jiang et al. 2012), which convert a serine codon (UCU) to a phenylalanine codon (UUU) in the *psaI* transcript (see Supplementary Table S6). On the other hand, other two modifications of amino acid polarity demonstrated here (first red arrow, Supplementary Fig. S1) are not predicted to be reverted by RNA editing in other species, which suggests specific changes in the hydrophobicity of PsaI protein in *B. orellana*. It is possible that these changes will have an impact in the *psaI* function in this species. This gene encodes the subunit I of the PSI (Photosystem I), a very small protein ranging from 34 to 40 amino acids (37 in *B. orellana*), which is present in cyanobacteria and land plants (Magee et al. 2010; Plöching et al. 2016). The subunit I is involved in trimerization of PSI in cyanobacteria; however, during the evolution of land plants it acquired a novel role in land plants, related to general stabilization of PSI and protection against oxidative damage and proteolytic degradation, mainly under high light and chilling stress (Schöttler et al. 2017). The knockout of this gene via plastid transformation revealed a small reduction of growth only under the previously cited stress conditions, while under normal greenhouse conditions no phenotype was apparent (Schöttler et al. 2017). According to the sequence of this gene in cyanobacteria and several lineages of land plants (Yu et al. 2008), a low selection pressure occurs to conserve the sequence, while the *psaI* gene was lost from the plastome in some lineages such as the genus *Lathyrus* (Magee et al. 2010). Differently, other *psa* genes (*psaA*, *psaB* and *psaC*) that constitute the PSI reaction center and bind to redox cofactors show conserved amino acid sequences across photosynthetic lineages (Yu et al. 2008; and also visualized here in *B. orellana* and in other species of Malvales, Fig. 5). Thus, the evolutionary characteristics of *psaI* gene described here, characterized by shifts of some amino acid residues and changes of polarity, may be the consequence of low selection pressure and/or positive selection, which favor structural changes that influence the protein hydrophobicity and its function.

The *B. orellana rpl23* gene showed a dN/dS value slightly above 1 (1.02); however, it was lower in comparison with the mean value from other species of Malvales (1.60). The value observed here for *rpl23* gene seems to be more associated with sequence variation across species of Malvales than a specific variation which occurred in the *B. orellana* plastome. The dN/dS values (> 1) in *B. orellana* and other species of Malvales also suggest a positive selection acting in this gene. Although these values are approximately one in *B. orellana* (i.e. almost equal values of synonymous and non-synonymous substitutions), it is unlikely that the *rpl23* gene is under neutral selection (absence of both positive and purifying selection, and an indicative of pseudogene) in this species. Firstly, the translated amino acid residues are still conserved in the species of Malvales (Supplementary

Fig. S2). Secondly, the product encoded by this gene, the protein L23, is essential for plastid ribosome functionality and plant cell viability (Fleischmann et al. 2011). Accordingly, this gene is present and functional in most plastomes of angiosperms sequenced to date; however, some exceptional plant lineages lost the *rpl23* gene from the plastomes such as some Caryophyllales (Bubunenko et al. 1994; Logacheva et al. 2008; Sloan et al. 2014; Raman and Park 2015) and flax (Lopes et al. 2018c). In all of those cases, the loss of plastid *rpl23* gene is associated with some compensatory mechanism. In the species of Caryophyllales, the function of the plastid *rpl23* gene seems to be replaced by a eukaryotic-type L23 protein, encoded by the nucleus and imported into the plastids. A similar mechanism was also proposed for flax (Lopes et al. 2018c), although it has not yet been confirmed experimentally. Therefore, the dN/dS value (close to 1) found here for the *rpl23* gene is likely a combination of positive selection acting in some sites and purifying selection acting in others maintaining the functionality of the L23 protein in *B. orellana* and other species of Malvales.

A high number of SSRs are present in the *B. orellana* plastome

The use of plastid simple sequence repeats (SSRs or microsatellites) as genetic markers was introduced by Powell et al. (1995), which reported an extensive level of polymorphism of these sequences in the genera *Pinus* and *Glycine*. This high intraspecific genetic variability of plastid SSRs is present across plastomes of all land-plant lineages (Provan et al. 2001; Wheeler et al. 2014b). Additionally, these sequences have unique features such as lack of allele recombination and uniparental inheritance. Thus, plastid SSRs are recognized as excellent genetic markers to access genetic variability within species and among closely related species. They are useful to a broad range of genetic researches such as studies of population genetics, genetic divergence of plant accessions, germplasm origin and characterization of evolutionary history of native and crop species (Powell et al. 1995; Provan et al. 2001; Wheeler et al. 2014a).

Here, a high number (321) of SSRs (> 8 nucleotides) were characterized in the *B. orellana* plastome. The data observed here are in accordance with other studies, which observed a larger number of mononucleotides and a dominant occurrence in non-coding regions in plastomes of several plant lineages (George et al. 2015; Lopes et al. 2018a). These authors also noted a remarkable abundance of mono and dinucleotides repeats constituted by A and T which is similar to the results found here in the *B. orellana* plastome (i.e. 97.4% mono and 82.9% dinucleotides composed by A/T; Supplementary Table S2). Only one CG dinucleotide was found in the *B. orellana* plastome corroborating with

the rarity of this repeat in plastomes of land plants. This larger number of SSRs found here, mainly those located in non-coding regions, can be employed in intraspecific and interspecific analyses of polymorphism, thus representing a possibility for diverse genetic proposals in *B. orellana* and related species (Supplementary Table S3). The plastid SSRs characterized here can be utilized alone or in combination with nuclear SSRs (Dequigiovanni et al. 2014, 2018), aiming studies of genetic diversity, gene flow, breeding, hybridization, domestication and genetic conservation of *B. orellana*.

The *B. orellana* plastome shows a small number of directed and inverted repeats

In contrast to the high number of SSR markers, only eight directed repeats and six inverted repeats (≥ 30 bp) were found in the *B. orellana* plastome. It is a low number if compared with other species of Malvales, such as *Gossypium hirsutum* (30 directed and 24 inverted repeats; Lee et al. 2006) and *Aquilaria sinensis* (29 directed and 24 inverted repeats; Wang et al. 2016). Most repeats are localized in non-coding regions of *B. orellana* plastome, with only a pair localized completely in the CDS of *psaA* and *psaB* genes (Supplementary Table S5). This pattern of distribution was also reported to other species of Malvales and other angiosperms (Raubeson et al. 2007; Lee et al. 2006; Yang et al. 2013). The role of these repeats is still unknown, but it has been suggested that they could have a functional role (Raubeson et al. 2007). Moreover, other 19 tandem repeats were identified in the *B. orellana* plastome, which are preferentially located in intergenic regions and in the CDS of *ycf2* gene (Supplementary Table S4). The presence of several tandem repeats in the *ycf2* gene was already reported in other plastomes (Jansen et al. 2006; Lee et al. 2006; Asif et al. 2010; Lopes et al. 2018c). Similarly, the functional role of these tandem repeats is poorly studied, although it is suggested that they could play an important role in plastome size, structure and evolution (Dugas et al. 2015).

The *B. orellana* plastome shows unique RNA editing sites

RNA editing is a post-transcriptional mechanism commonly found in plastid transcripts. It can create start and stop codons and restores conserved amino acids modifying the sequence of mRNAs before they are translated (Tsudzuki et al. 2001; Takenaka et al. 2013). In the *B. orellana* plastome 57 putative RNA editing sites were identified using the software PREP (Supplementary Table S6). Of the total RNA editing sites found here, 34 were experimentally validated in at least one angiosperm species such as *G. hirsutum* (Jiang et al. 2012), *Jatropha curcas* (Asif et al. 2010), *Hevea brasiliensis* (Tangphatsornruang et al. 2011), *Arabidopsis*

thaliana (Tillich et al. 2005), *Nicotiana tabacum* (Hirose et al. 1999), *Solanum lycopersicum* (Kahlau et al. 2006), *Atropa belladonna* (Schmitz-Linneweber et al. 2002), *Cocos nucifera* (Huang et al. 2013), *Zea mays* (Maier et al. 1995; Bock et al. 1997) and *Oryza sativa* (Corneille et al. 2000).

The *ndh* genes were predicted to have the majority of RNA editing sites in *B. orellana*. All sites predicted for the *ndhB* gene in *B. orellana* were experimentally confirmed in the species listed before and in the species of family Malvaceae, except for a site localized in the amino acid position 65 (Supplementary Table S6). In *B. orellana*, the cytosine (C) localized in this position is predicted to be edited to a uridine (U), converting a proline codon (CCA) to a conserved serine codon (TCA). On the other hand, the species of family Malvaceae already have a serine codon (UCA) in this position dispensing the need of RNA editing. This RNA editing site can be a new specific site for *B. orellana*. Five out of eight RNA editing sites predicted for the *ndhD* gene in *B. orellana* were already validated in other species of family Malvaceae (amino acid positions: 1, 128, 293, 433 and 437). Among the other three sites found here, two are predicted for *B. orellana* (amino acid positions 359 and 360) representing putative specific sites. The last one (position 16) was also predicted to occur in species of family Malvaceae constituting a new RNA editing site shared with this family and other species of Malvales. Two sites were predicted for the *ndhA* gene in *B. orellana* and other species of family Malvaceae, which were already confirmed experimentally in *G. hirsutum* (Jiang et al. 2012). The genes, *ndhF* and *ndhG*, have also two predicted sites each in *B. orellana*. However, the sites predicted here, at the amino acid positions 196 of *ndhF* and 105 of *ndhG*, were confirmed to be false positives in *C. nucifera* (Huang et al. 2013). The other sites also predicted for *B. orellana* and other species of family Malvaceae require confirmation by experimental validation if they are really new editing sites or other false positives.

The *rpo* genes constitute the second gene class containing a high number of predicted RNA editing sites in the *B. orellana* plastome. All sites predicted here for the genes *rpoB* and *rpoC1* were already validated in some angiosperm species such as *G. hirsutum*, *N. tabacum*, *S. lycopersicum*, *A. belladonna*, *Z. mays* and *C. nucifera* (Maier et al. 1995; Bock et al. 1997; Hirose et al. 1999; Schmitz-Linneweber et al. 2002; Kahlau et al. 2006; Jiang et al. 2012; Huang et al. 2013). However, the sites predicted here for the *rpoA* and *rpoC2* genes still need experimental validation. The two sites predicted for *rpoC2* gene in *B. orellana* do not occur in the species of family Malvaceae, thus being probably unique for *B. orellana* (Supplementary Table S6).

The three RNA editing sites predicted for the *atpA* gene in *B. orellana* were validated in at least one of the angiosperms mentioned before. One editing site was predicted for the *atpB* gene and one for the *atpF* gene (amino acid

position 88). The first is shared with some species of family Malvaceae and the second occurs exclusively in *B. orellana*. Interestingly, the *B. orellana* plastome lacks a conserved RNA editing site localized in the nucleotide position 92 (amino acid position 31) of the *atpF* gene (Supplementary Table S6). Additionally, the intron of this gene is also absent of the *B. orellana* plastome, which contrasts with other species of Malvales that retain the intron and the RNA editing site in the *atpF* gene. Moreover, a strong association between the loss of the intron and a C-to-T substitution in the DNA at the codon position 92 (i.e. loss of the RNA editing site) was already suggested by Daniel et al. (2008). These authors suggest that a recombination between the edited RNA and the *atpF* gene could be the mechanism involved in the loss of the intron. The results found here for the *atpF* gene (i.e. loss of both the intron and the editing site in the position 92) corroborate with this hypothesis and suggest that the same mechanism occurred in the *B. orellana* plastome.

The *matK* gene was predicted to have five RNA editing sites in the *B. orellana* plastome (Supplementary Table S6). Among them, three were validated in other species (amino acid positions 160, 219 and 423) and are shared with other species of family Malvaceae. Nevertheless, the other two sites (amino acid positions 271 and 403) are unique for the *B. orellana* plastome. The *accD* gene has a conserved editing site in the amino acid position 263, which is shared with *B. orellana* and other species of family Malvaceae, and already validated in *G. hirsutum*, *C. nucifera* and *A. thaliana* (Tillich et al. 2005; Jiang et al. 2012; Huang et al. 2013). An additional RNA editing site was predicted in this gene for *B. orellana* and some species of family Malvaceae (amino acid position 466), which was experimentally confirmed in *G. hirsutum* (Jiang et al. 2012). On the other hand, the unique editing site predicted for *ccsA* gene in *B. orellana* was investigated in *C. nucifera* and it was confirmed experimentally to be a false positive (Huang et al. 2013). Thus, it is possible that the same is true in relation to this site in *B. orellana*. The *psbF*, *clpP* and *petB* genes showed only one editing site each, in both *B. orellana* and other species of Malvales, which were already confirmed experimentally in *G. hirsutum* (Jiang et al. 2012). Furthermore, the *psbE* gene also showed an RNA editing site in the *B. orellana* plastome (amino acid position 72), which was already confirmed in other angiosperms (Hirose et al. 1999; Tillich et al. 2005; Tangphatsornruang et al. 2011), but it is dispensable in species of family Malvaceae due to the presence of a T in this codon position in those species.

Some ribosomal protein genes also have putative RNA editing sites (Supplementary Table S6). The *rps14* gene contains the higher number of RNA editing sites. The editing site localized in the amino acid position 27 is conserved between *B. orellana* and other species of family Malvaceae and it was still validated in *G. hirsutum* and in other

angiosperms (Jiang et al. 2012). The RNA editing site at the position 34 seems to be exclusive for *B. orellana* and represents another interesting site to be validated in this species. The last site (amino acid position 50) is predicted to convert a serine (UCG) codon to a leucine (UUG) codon in both *B. orellana* and other species of family Malvaceae, and it was also validated in *G. hirsutum* (Jiang et al. 2012). Interestingly, an editing site in the same amino acid position was also validated in tobacco, whereas in that species the RNA editing converts a proline (CCA) codon to a leucine (CUA) codon (Hirose et al. 1999). Then, in this codon position of *rps14* gene, codons completely different are found in the DNA sequence of tobacco plastome (CCA) and species of Malvales (UCG/UCA). However, the RNA editing site remains conserved in all these species, at the same position, restoring an essential conserved leucine residue. The *rps2* gene showed one predicted RNA editing site in *B. orellana*, which is shared with other species of Malvales and it is experimentally confirmed in *G. hirsutum*. Finally, the *B. orellana* genes, *rpl20* and *rps16*, have one predicted editing site each, which seems to be specific for this species, but their confirmation needs to be analysed experimentally.

Malvales phylogenetic relationship based on whole plastomes

In previous studies related to the phylogeny of Malvales, the position of Bixaceae within Malvales has been ambiguous and not well supported (Fay et al. 1998; Bayer et al. 1999; Soltis et al. 2011; Le Péchon and Gigord 2014). A major discrepancy is associated with the relationship between the families Bixaceae, Malvaceae and Thymelaeaceae. Fay et al. (1998), using the sequence of *rbcL* gene, reconstructed a maximum parsimony tree in which the families Malvaceae and Thymelaeaceae formed a clade that is a sister group to family Bixaceae. In a more broad approach, Soltis et al. (2011) found similar results using the maximum likelihood (ML) method. However, a maximum parsimony tree reconstructed by Bayer et al. (1999), using a combination of sequences of *rbcL* and *atpB* genes, revealed that the family Malvaceae is more related to Bixaceae than to Thymelaeaceae. Here, using whole-plastome sequences as dataset, we reconstructed a Bayesian tree that strongly supports (100% of posterior probability) the closed relationship between the family Bixaceae (represented here by *B. orellana*) and the family Malvaceae, and the position of family Thymelaeaceae as sister to the clade formed by the families Bixaceae and Malvaceae (Fig. 4). Our ML tree was also congruent with these results but it showed only a moderate support value (62%) to the clade formed by Bixaceae and Malvaceae (Fig. S3). It is possible that the addition of other taxa, mainly of other families belonging to the order Malvales, would improve the

support value. Taking into account the importance of plastid genomics, the complete sequences of species belonging to other families of Malvales will contribute significantly to a more complete reconstruction of phylogenetic history of Malvales, including a more precise positioning of Bixaceae within this order.

Our phylogenomic trees also indicate, with 100% of both posterior probability and bootstrap support (Fig. 4 and Fig. S3), the monophyletic origin of family Malvaceae (used here as Malvaceae sensu lato, previously known as core Malvales), in accordance with previous works (Fay et al. 1998; Alverson et al. 1999; Bayer et al. 1999). The subfamilies Tilioideae, Sterculioideae, Helicterioideae, Bombacoideae and Malvoideae, within family Malvaceae, form a clade named Malvadendrina (Bayer et al. 1999). This group was sister to the subfamily Byttnerioideae in previous works, using the sequences of *ndhF* gene (Alverson et al. 1999) and sequences of *rbcL* and *atpB* genes (Bayer et al. 1999). However, in a previous phylogenetic analysis, using the sequence of *rbcL* gene, reconstructed a parsimony tree in which *Tilia* (subfamily Tilioideae) occupied the position of sister to the other Malvaceae subfamilies (Fay et al. 1998). Here, in our trees (Fig. 4 and Fig. S3), we confirmed with strong support that Byttnerioideae is sister to the Malvadendrina clade. Inside Malvadendrina clade, the relationship between some subfamilies has been uncertain and ambiguous (Alverson et al. 1999). In a previous study, *Tilia* (Tilioideae) was sister to the clade formed by all of other Malvadendrina (Bayer et al. 1999). However, in posterior studies, the Helicterioideae subfamily occupied this position (Alverson et al. 1999; Nyffeler et al. 2005). Congruent with the last studies, we found here with strong support that *Durio* (Helicterioideae) is sister to the clade formed by the other Malvadendrina subfamilies. The closed relationship between Malvoideae and Bombacoideae observed here was also confirmed in previous phylogenetic analyses (Alverson et al. 1999; Bayer et al. 1999; Baum et al. 2004; Nyffeler et al. 2005). However, the relationship between these subfamilies with Tilioideae and Sterculioideae were unresolved in the previous studies. Here, we found that Tilioideae and Sterculioideae were more closely related (with strong posterior probability and moderate bootstrap support) and formed a clade that is sister to the clade constituted by Bombacoideae and Malvoideae. Lastly, we confirmed the monophyletic origin of the subfamily Malvoideae, which formed two subclades: one containing three species of the tribe Hibisceae (*H. syriacus*, *A. esculentus* and *T. hamabo*) and other formed by *Gossypium* (tribe Gossypieae) and *Althaea* (tribe Malveae). The relationship between the tribes of Malvaceae is congruent with the results found by Baum et al. (2004).

Conclusion

Here we reported the complete plastome of *B. orellana*, which is the first plastome of the family Bixaceae to be fully sequenced and characterized in detail regarding the gene content, structure, molecular markers and evolutionary features. The *B. orellana* plastome shows a high similarity to the plastome of other species of Malvales. However, some specific structural and evolving features were still detected in the *B. orellana* plastome such as rearrangements at the IR-borders and high rate of non-synonym substitutions in the *psaI* gene. The *infA* gene was the unique pseudogene found in the *B. orellana* plastome, which is also pseudogene or absent in all species of Malvales sequenced to date and even in several other angiosperms. Two interesting evolving features found in the *B. orellana* plastome were the losses of the *atpF* intron and a RNA editing site present in this gene. The occurrence of the RNA editing site is related to the presence of the intron given that both are present in all species of Malvales investigated here and in most angiosperms. Additionally, 57 RNA editing sites were predicted for *B. orellana* genes, of which 42 also were observed in other species of Malvales and 34 of them were already validated in other angiosperms. From all RNA editing sites found here 11 sites seem to be unique for *B. orellana*, thus constituting potential target for experimental validation in this species and/or in other species of this family. Moreover, in the *B. orellana* plastome were identified 312 SSRs, which can be used for polymorphism analyses in several approaches such as population genetic, conservation, germplasm evaluation and breeding. Furthermore, our plastid phylogenomic analyses indicate (with strong PP and moderate BS) a closed relationship between the family Bixaceae (represented by *B. orellana*) and the family Malvaceae, with both families forming a sister group to the family Thymeleaceae. The approach used here also resolved with high support the relationship between subfamilies of Malvaceae, including incongruences found before in this family. The results showed here suggest a great potential of plastid phylogenomics for resolving phylogenetic incongruences found in the family Malvaceae as well as in the order Malvales and other plant lineages. However, the unavailability of plastome sequences from diverse taxa in the plastid database still represents a limitation to use this approach efficiently. It highlights the importance of studies related to plastid genomics in families or taxa that are poorly known. Finally, the data showed here are important in several fields of genetic, evolution, phylogeny, conservation, breeding and biotechnology of *B. orellana* and family Bixaceae.

Author contribution statement TGP, ASL, GDMV, LNV, MPG, RON, EMS, FOP, WCO and MR conceived and designed the research. TGP, ASL, GDMV, ONS, GMS, LNV, EMS, FOP, and MR conducted experiments and analysed the

data. MPG, RON, EMS, FOP, WCO and MR contributed with reagents and materials. TGP and MR wrote the manuscript. All authors read and approved the manuscript.

Acknowledgements This research was support by the National Council for Scientific and Technological Development, Brazil (CNPq, Grant 459698/2014-1). We are grateful to INCT-FBN and for the scholarships granted by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) to GDMV, ONS, GMS and LNV, and those granted by the CNPq to ASL, TGP, MPG, RON, EMS, FOP and WCO.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S et al (2012) The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet* 8:e1003076. <https://doi.org/10.1371/journal.pgen.1003076>
- Alverson WS, Whitlock BA, Nyfeler R, Bayer C, Baum DA (1999) Phylogeny of the core Malvales: evidence from *ndhF* sequence data. *Am J Bot* 86:1474–1486. <https://doi.org/10.2307/2656928>
- Annadurai RS, Jayakumar V, Mugasimangalam RC, Katta MAVSK, Anand S, Gopinathan S, Sarma SP, Fernandes SJ, Mullapudi N, Murugesan S, Rao SN (2012) Next generation sequencing and de novo transcriptome analysis of *Costus pictus* D. Don, a non-model plant with potent anti-diabetic properties. *BMC Genom* 13:1–15. <https://doi.org/10.1186/1471-2164-13-663>
- Asif MH, Mantri SS, Sharma A, Srivastava A, Trivedi I, Gupta P, Mohanty CS, Sawant SV, Tuli R (2010) Complete sequence and organisation of the *Jatropha curcas* (Euphorbiaceae) chloroplast genome. *Tree Genet Genomes* 6:941–952. <https://doi.org/10.1007/s11295-010-0303-0>
- Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L et al (2016) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol* 209:855–870. <https://doi.org/10.1111/nph.13617>
- Baum DA, Smith ST, Yen A, Alverson WS, Nyfeler R, Whitlock BA, Oldham RL (2004) Phylogenetic relationships of Malvaceae (Bombacoideae and Malvoideae; Malvaceae Sensu Lato) as inferred from plastid DNA sequences. *Am J Bot* 91:1863–1871. <https://doi.org/10.3732/ajb.91.11.1863>
- Bayer C, Fay MF, De Bruijn AY, Savolainen V, Morton CM, Kubitzki K, Alverson WS, Chase MW (1999) Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: a combined analysis of plastid *atpB* and *rbcL* DNA sequences. *Bot J Linn Soc* 129:267–303. <https://doi.org/10.1111/j.1095-8339.1999.tb00505.x>
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580. <https://doi.org/10.1093/nar/27.2.573>
- Bock R (2013) Strategies for metabolic pathway engineering with multiple transgenes. *Plant Mol Biol* 83:21–31. <https://doi.org/10.1007/s11103-013-0045-0>

- Bock R (2015) Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. *Annu Rev Plant Biol* 66:211–241. <https://doi.org/10.1146/annurev-arplant-050213-040212>
- Bock R (2017) Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. *Annu Rev Genet* 51:1–22. <https://doi.org/10.1146/annurev-genet-120215-035329>
- Bock R, Albertazzi F, Freyer R, Fuchs M, Ruf S, Zeltz P, Maier RM (1997) Transcript editing in chloroplasts of higher plants. In: Schenk HEA, Herrmann RG, Jeon KW, Müller NE, Schwemmler W (eds) *Eukaryotism and symbiosis*. Springer, Berlin, pp 123–137. https://doi.org/10.1007/978-3-642-60885-8_9
- Bouvier F, Dogbo O, Câmara B (2003) Biosynthesis of the food and cosmetic plant pigment bixin (annatto). *Science* 300:2089–2091. <https://doi.org/10.1126/science.1085162>
- Bubunenko MG, Schmidt J, Subramanian AR (1994) Protein substitution in chloroplast ribosome evolution. A eukaryotic cytosolic protein has replaced its organelle homologue (L23) in spinach. *J Mol Biol* 240:28–41. <https://doi.org/10.1006/jmbi.1994.1415>
- Cai J, Ma P-F, Li H-T, Li D-Z (2015) Complete plastid genome sequencing of four *Tilia* species (Malvaceae): a comparative analysis and phylogenetic implications. *PLoS One* 10:e0142705. <https://doi.org/10.1371/journal.pone.0142705>
- Cárdenas-Conejo Y, Carballo-Uicab V, Lieberman M, Aguilar-Espinosa M, Comai L, Rivera-Madrid R (2015) De novo transcriptome sequencing in *Bixa orellana* to identify genes involved in methylerythritol phosphate, carotenoid and bixin biosynthesis. *BMC Genom* 16:877. <https://doi.org/10.1186/s12864-015-2065-4>
- Carvalho JFRP, Robinson IP, Alfenas AC (2005) Isozymic variability in a Brazilian collection of annatto (*Bixa orellana* L.). *Pesq Agropec Bras* 40:653–660. <https://doi.org/10.1590/S0100-204X2005000700005>
- Cho WB, Han EK, Choi G, Lee JH (2017) The complete chloroplast genome of *Daphne kiusiana*, an evergreen broad-leaved shrub on Jeju Island. *Conserv Genet Resour* 10:103–106. <https://doi.org/10.1007/s12686-017-0774-5>
- Clement CR, Cristo-Araújo M, d'Eeckenbrugge GC, Alves-Pereira A, Picanço-Rodrigues D (2010) Origin and domestication of native Amazonian crops. *Diversity* 2:72–106. <https://doi.org/10.3390/d2010072>
- Comer JR, Zomlefer WB, Barrett CF, Davis JJ, Stevenson DW, Heyduk K, Leebens-Mack JH (2015) Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from next-generation sequencing. *Am J Bot* 102:888–899. <https://doi.org/10.3732/ajb.1500057>
- Corneille S, Lutz K, Maliga P (2000) Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? *Mol Gen Genet* 264:419–424. <https://doi.org/10.1007/s004380000295>
- Cummings HS, Hershey JW (1994) Translation initiation factor IF1 is essential for cell viability in *Escherichia coli*. *J Bacteriol* 176:198–205. <https://doi.org/10.1128/jb.176.1.198-205.1994>
- Da Cruz ACF, Pinheiro MVM, Xavier A, Otoni WC, Costa MGC, Paiva Neto VB, Régio MM (2015) In vitro regeneration of Annatto (*Bixa orellana* L.) plantlets from nodal and internodal adult stem explants. *Acta Hort* 1083:335–346
- Daniell H, Wurdack KJ, Kanagaraj A, Lee SB, Saski C, Jansen RK (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. <https://doi.org/10.1007/s00122-007-0706-y>
- Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* 17:134. <https://doi.org/10.1186/s13059-016-1004-2>
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>
- de Araújo Vilar D, de Araujo Vilar MS, de Accioly Lima e Moura TF et al (2014) Traditional uses, chemical constituents, and biological activities of *Bixa orellana* L.: a review. *Sci World J* 2014:1–11. <https://doi.org/10.1155/2014/857292>
- de Matos EM, Koehler AD, Faria DV, Correia LNF, Moreira VS, de Cruz ACF, Moraes TS, Rocha DI, Soares VLF, Paiva Neto VB, Costa MGC, Otoni WC (2016) Somatic embryogenesis in Annatto (*Bixa orellana* L.). In: Loyola-Vargas Víctor M, Ochoa-Alejo N (eds) *Somatic embryogenesis: fundamental aspects and applications*, vol 1, 1st edn. Springer Nature, Cham, pp 213–231. https://doi.org/10.1007/978-3-319-33705-0_13
- Dequigiovanni G, Ramos SLF, Zucchi MI, Bajay MM, Pinheiro JB, Fabri EG, Bressan EA, Veasey EA (2014) Isolation and characterization of microsatellite loci for *Bixa orellana*, an important source of natural dyes. *Genet Mol Res* 13:9097–9102. <https://doi.org/10.4238/2014.October.31.25>
- Dequigiovanni G, Ramos SLF, Lopes MTG, Clement CR, Rodrigues DP, Fabri EG, Zucchi MI, Veasey EA (2018) New microsatellite loci for annatto (*Bixa orellana*), a source of natural dyes from Brazilian Amazonia. *Crop Breed Appl Biotechnol* 18:116–122. <https://doi.org/10.1590/1984-70332018v18n1n18>
- Dugas DV, Hernandez D, Koenen EJ, Schwarz E, Straub S, Hughes CE, Jansen RK, Nageswara-Rao M, Staats M, Trujillo JT, Hajrah NH, Alharbi NS, Al-Malki AL, Sabir JSM, Bailey CD (2015) Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in clpP. *Sci Rep* 5:16958. <https://doi.org/10.1038/srep16958>
- Ebert D, Peakall R (2009) Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Mol Ecol Resour* 9:673–690. <https://doi.org/10.1111/j.1755-0998.2008.02319.x>
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Fay MF, Bayer C, Alverson WS et al (1998) Plastid rbcL sequence data indicate a close affinity between Diegodendron and Bixa. *Taxon* 47:43–50. <https://doi.org/10.2307/1224017>
- Fleischmann TT, Scharff LB, Alkatib S, Hasdorf S, Schottler MA, Bock R (2011) Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell* 23:3137–3155. <https://doi.org/10.1105/tpc.111.088906>
- Fuentes P, Armarego-Marriott T, Bock R (2017) Plastid transformation and its application in metabolic engineering. *Curr Opin Biotechnol* 49:10–15. <https://doi.org/10.1016/j.copbio.2017.07.004>
- George B, Bhatt BS, Awasthi M, George B, Singh AK (2015) Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr Genet* 61:665–677. <https://doi.org/10.1007/s00294-015-0495-9>
- Giridhar P, Parimalan R (2010) A biotechnological perspective towards improvement of annatto color production for value addition—the influence of biotic elicitors. *Asia Pac J Mol Biol Biotechnol* 18:77–79
- Giuliano G, Salim A, von Lintig J (2003) Carotenoid oxygenases: cleave it or leave it. *Trends Plant Sci* 8:145–149. [https://doi.org/10.1016/S1360-1385\(03\)00053-0](https://doi.org/10.1016/S1360-1385(03)00053-0)

- Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet* MGG 252:195–206. <https://doi.org/10.1007/BF02173220>
- Guesmi A, Hamadi NB, Ladhari N, Sakli F (2013) Sonicator dyeing of modified acrylic fabrics with indicaxanthin natural dye. *Ind Crops Prod* 42:63–69. <https://doi.org/10.1016/j.indcrop.2012.05.022>
- Gurdon C, Maliga P (2014) Two distinct plastid genome configurations and unprecedented intraspecies length variation in the accD coding region in *Medicago truncatula*. *DNA Res* 21:417–427. <https://doi.org/10.1093/dnares/dsu007>
- Hirose T, Kusumegi T, Tsudzuki T, Sugiura M (1999) RNA editing sites in tobacco chloroplast transcripts: editing as a possible regulator of chloroplast RNA polymerase activity. *Mol Gen Genet* MGG 262:462–467. <https://doi.org/10.1007/s004380051106>
- Huang YY, Matzke AJM, Matzke M (2013) Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS One* 8:e74736. <https://doi.org/10.1371/journal.pone.0074736>
- Jansen RK, Kaitanis C, Saski C, Lee SB, Tomkins J, Alverson AJ, Daniell H (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. <https://doi.org/10.1186/1471-2148-6-32>
- Jiang Y, Fan SL, Song MZ, Yu JN, Yu SX (2012) Identification of RNA editing sites in cotton (*Gossypium hirsutum*) chloroplasts and editing events that affect secondary and three-dimensional protein structures. *Genet Mol Res* 11:987–1001. <https://doi.org/10.4238/2012.April.19.4>
- Kahlau S, Aspinall S, Gray JC, Bock R (2006) Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J Mol Evol* 63:194–207. <https://doi.org/10.1007/s00239-005-0254-5>
- Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels M, Cronk JM (2012) Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Am J Bot* 99:320–329. <https://doi.org/10.3732/ajb.1100570>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
- Krech K, Fu HY, Thiele W, Ruf S, Schöttler MA, Bock R (2013) Reverse genetics in complex multigene operons by co-transformation of the plastid genome and its application to the open reading frame previously designated psbN. *Plant J* 75:1062–1074. <https://doi.org/10.1111/tj.12256>
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 29:4633–4642. <https://doi.org/10.1186/gb-2004-5-2-r12>
- Labiak PH, Karol KG (2017) Plastome sequences of an ancient fern lineage reveal remarkable changes in gene content and architecture. *Am J Bot*. <https://doi.org/10.3732/ajb.1700135> (in press)
- Le Péchon T, Gigord LDB (2014) On the relevance of molecular tools for taxonomic revision in Malvales, Malvaceae s.l., and Dombeyoideae. In: Besse P (ed) *Molecular plant taxonomy. Methods in molecular biology (methods and protocols)*, vol 1115. Humana Press, Totowa. https://doi.org/10.1007/978-1-62703-767-9_17
- Leal F, Clavijo CM (2010) Annatto: a natural dye from the tropics. *Chron Hort* 50:34–36
- Lee SB, Kaitanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H (2006) The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms. *BMC Genom* 7:61–72. <https://doi.org/10.1186/1471-2164-7-61>
- Logacheva MD, Samigullin TH, Dhingra A, Penin AA (2008) Comparative chloroplast genomics and phylogenetics of *Fagopyrum esculentum* ssp. ancestrale—a wild ancestor of cultivated buckwheat. *BMC Plant Biol* 8:59. <https://doi.org/10.1186/1471-2229-8-59>
- Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41:W575–W581. <https://doi.org/10.1093/nar/gkt289>
- Lopes AS, Gomes Pacheco T, do Nascimento Vieira L, Guerra MP, Nodari RO, de Maltempi Souza E, de Oliveira Pedrosa F, Rogalski M (2018a) The *Crambe abyssinica* plastome: Brassicaceae phylogenomic analysis, evolution of RNA editing sites, hotspot and microsatellite characterization of the tribe Brassiceae. *Gene* 671:36–49. <https://doi.org/10.1016/j.gene.2018.05.088>
- Lopes AS, Gomes Pacheco T, Nimz T, do Nascimento Vieira L, Guerra MP, Nodari RO, de Maltempi Souza E, de Oliveira Pedrosa F, Rogalski M (2018b) The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analysis of the evolution of plastid genes in Arecaeae. *Planta* 247:1011–1030. <https://doi.org/10.1007/s00425-018-2841-x>
- Lopes AS, Pacheco TG, Santos KG, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Rogalski M (2018c) The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Rep* 37:307–328. <https://doi.org/10.1007/s00299-017-2231-z>
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <https://doi.org/10.1093/nar/25.5.0955>
- Lu Y, Stegemann S, Agrawal S, Karcher D, Ruf S, Bock R (2017) Horizontal transfer of a synthetic metabolic pathway between plant species. *Curr Biol* 27:3034–3041. <https://doi.org/10.1016/j.cub.2017.08.044>
- Magee AM, Aspinall S, Rice DW, Cusack BP, Semon M, Perry AS, Stefanovic S, Milbourne D, Barth S, Palmer JD, Gray JC, Kavanagh TA, Wolfe KH (2010) Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res* 20:1700–1710. <https://doi.org/10.1101/gr.111955.110>
- Maier RM, Necker mann K, Igloi GL, Kössel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol* 251:614–628. <https://doi.org/10.1006/jmbi.1995.0460>
- Mantovani NC, Grando MF, Xavier A, Otoni WC (2013) Avaliação de genótipos de urucum (*Bixa orellana* L.) por meio da caracterização morfológica de frutos, produtividade de sementes e teor de bixina. *Ciência Florestal* 23:355–362. <https://doi.org/10.5902/198050989281>
- Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW (2001) Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13:645–658. <https://doi.org/10.1105/tpc.13.3.645>
- Moreira PA, Lins J, Dequigiovanni G, Veasey EA, Clement CR (2015) The domestication of annatto (*Bixa orellana*) from *Bixa urucurana* in Amazonia. *Econ Bot* 69:127–135. <https://doi.org/10.1007/s12231-015-9304-0>
- Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res* 37:W253–W259. <https://doi.org/10.1093/nar/gkp337>
- Nguyen L-T, Schmidt HA, Haeseler AV, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating

- maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
- Nyffeler R, Bayer C, Alverson WS, Yen A, Whitlock B, Chase MW, Baum DA (2005) Phylogenetic analysis of the *Malvadenrina clade* (Malvaceae s.l.) based on plastid DNA sequences. *Org Divers Evol* 5:109–123. <https://doi.org/10.1016/j.ode.2004.08.001>
- Paiva Neto VB, Ribeiro da Mota T, Otoni WC (2003) Direct organogenesis from hypocotyl-derived explants of annatto (*Bixa orellana*). *Plant Cell Tissue Organ Cult* 75:159. <https://doi.org/10.1023/A:1025063906822>
- Parimalan R, Giridhar P, Ravishankar GA (2011) Enhanced shoot organogenesis in *Bixa orellana* L. in the presence of putrescine and silver nitrate. *Plant Cell Tissue Organ Cult* 105:285–290. <https://doi.org/10.1007/s11240-010-9865-7>
- Park S, Ruhlman TA, Weng ML, Hajrah NH, Sabir JSM, Jansen RK (2017) Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of geranium. *Genome Biol Evol* 9:1766–1780. <https://doi.org/10.1093/gbe/evx124>
- Piot A, Hackel J, Christin PA, Besnard G (2018) One-third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta* 247:255–266. <https://doi.org/10.1007/s00425-017-2781-x>
- Plöschinger M, Torabi S, Rantala M, Tikkanen M, Suorsa M, Jensen PE, Aro EM, Meurer J (2016) The low molecular weight protein Psal stabilizes the light-harvesting complex II docking site of photosystem I. *Plant Physiol* 172:450–463. <https://doi.org/10.1104/pp.16.00647>
- Portela de Carvalho JFR, de Portela Carvalho CR, Otoni WC (2005) In vitro induction of polyploidy in annatto (*Bixa orellana*). *Plant Cell Tissue Organ Cult* 80:69–75. <https://doi.org/10.1007/s11240-004-8833-5>
- Powell W, Morgantet M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV, Rafalski JA (1995) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr Biol* 5:1023–1029. [https://doi.org/10.1016/S0960-9822\(95\)00206-5](https://doi.org/10.1016/S0960-9822(95)00206-5)
- Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol* 16:142–147. [https://doi.org/10.1016/S0169-5347\(00\)02097-8](https://doi.org/10.1016/S0169-5347(00)02097-8)
- Ramamoorthy S, Dossa FP, Kundua K, Satyanarayanab VSV, Kumbar V (2010) Molecular characterization of bixin an important industrial product. *Ind Crops Prod* 32:48–53. <https://doi.org/10.1016/j.indcrop.2010.03.001>
- Raman G, Park S (2015) Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalycinus*, from a comparative genomics perspective. *PLoS One* 10:e0141329. <https://doi.org/10.1371/journal.pone.0141329>
- Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar* and *Ranunculus macranthus*. *BMC Genom* 8:174. <https://doi.org/10.1186/1471-2164-8-174>
- Rivera-Madrid R, Escobedo-GM RM, Balam-Galera E, Vera-Ku M, Harries H (2006) Preliminary studies toward genetic improvement of annatto (*Bixa orellana* L.). *Sci Hortic* 109:165–172. <https://doi.org/10.1016/j.scienta.2006.03.011>
- Rivera-Madrid R, Burnell J, Aguilar-Espinosa M, Rodríguez-Ávila NL, LugoCervantes E, Sáenz-Carbonell LA (2013) Control of carotenoid gene expression in *Bixa orellana* L. leaves treated with norflurazon. *Plant Mol Biol Rep* 31:1422–1432. <https://doi.org/10.1007/s11105-013-0604-1>
- Rivera-Madrid R, Aguilar-Espinosa M, Cárdenas-Conejo Y, Garza-Caligaris LE (2016) Carotenoid derivatives in achioté (*Bixa orellana*) seeds: synthesis and health promoting properties. *Front Plant Sci* 7:1406. <https://doi.org/10.3389/fpls.2016.01406>
- Rodríguez-Ávila NL, Narváez-Zapata JA, Aguilar-Espinosa M, Rivera-Madrid R (2011) Regulation of pigment-related genes during flower and fruit development of *Bixa orellana*. *Plant Mol Biol Rep* 29:43–50. <https://doi.org/10.1007/s11105-010-0207-z>
- Rogalski M, Carrer H (2011) Engineering plastid fatty acid biosynthesis to improve food quality and biofuel production in higher plants: plastid fatty acid biosynthesis. *Plant Biotechnol J* 9:554–564. <https://doi.org/10.1111/j.1467-7652.2011.00621.x>
- Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Res* 34:4537–4545. <https://doi.org/10.1093/nar/gkl634>
- Rogalski M, Schoettler MA, Thiele W, Schulze WX, Bock R (2008) Rpl33, a nonessential plastid encoded ribosomal protein in tobacco, is required under cold stress conditions. *Plant Cell* 20:2221–2237. <https://doi.org/10.1105/tpc.108.060392>
- Rogalski M, do Nascimento Vieira L, Fraga HP, Guerra MP (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front Plant Sci* 6:586. <https://doi.org/10.3389/fpls.2015.00586>
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542. <https://doi.org/10.1093/sysbio/sys029>
- Roquet C, Coissac E, Cruaud C, Boleda M, Boyer F, Alberti A, Gielly L, Taberlet P, Thuiller W, Van Es J, Lavergne S (2016) Understanding the evolution of holoparasitic plants: the complete plastid genome of the holoparasite *Cytinus hypocistis* (Cytinaceae). *Ann Bot* 118:885–896. <https://doi.org/10.1093/aob/mcw135>
- Roy PS, Rao GJN, Jena S, Samal R, Patnaik A, Patnaik SSC, Jambhulkar NN, Sharma S, Mohapatra T (2016) Nuclear and chloroplast DNA variation provides insights into population structure and multiple origin of native aromatic rices of Odisha, India. *PLoS One* 11:e0162268. <https://doi.org/10.1371/journal.pone.0162268>
- Ruhlman TA, Zhang J, Blazier JC, Sabir JSM, Jansen RK (2017) Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *Am J Bot* 104:559–572. <https://doi.org/10.3732/ajb.1600453>
- Sankari M, Hemachandran H, Anantharaman A, Babu S, Madrid RR, Fulzele DP, Siva R (2016) Identifying a carotenoid cleavage dioxygenase 4a gene and its efficient agrobacterium-mediated genetic transformation in *Bixa orellana* L. *Appl Biochem Biotechnol* 179:697–714. <https://doi.org/10.1007/s12010-016-2025-8>
- Savolainen V, Chase MW, Hoot SB, Morton CM, Soltis DE, Bayer C, Fay MF, de Bruijn AY, Sullivan S, Qiu Y-L (2000) Phylogenetics of flowering plants based upon a combined analysis of plastid atpB and rbcL gene sequences. *Syst Biol* 49:306–362. <https://doi.org/10.1093/sysbio/49.2.306>
- Schmitz-Linneweber C, Regel R, Du TG, Hupfer H, Herrmann RG, Maier RM (2002) The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation. *Mol Biol Evol* 19:1602–1612. <https://doi.org/10.1093/oxfordjournals.molbev.a004222>
- Schöttler MA, Thiele W, Belkiss K, Bergner SV, Flügel C, Wittenberg G, Agrawal S, Stegemann S, Ruf S, Bock R (2017) The plastid-encoded Psal subunit stabilizes photosystem I during leaf senescence in tobacco. *J Exp Bot* 68:1137–1155. <https://doi.org/10.1093/jxb/erx009>
- Shinozaki K, Ohme M, Tanaka M et al (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5:2043–2049
- Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, Taylor DR (2014) A recurring syndrome of accelerated plastid genome

- evolution in the angiosperm tribe Sileneae (Caryophyllaceae). *Mol Phylogenet Evol* 72:82–89. <https://doi.org/10.1016/j.ympev.2013.12.004>
- Smith DR (2015) Mutation rates in plastid genomes: they are lower than you might think. *Genome Biol Evol* 7:1227–1234. <https://doi.org/10.1093/gbe/evv069>
- Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoop SB, Fay MF, Axtell M, Swesen SM, Farris JS, Prince LM, Kress WJ, Nixon KC, Farris JS (2000) Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Bot J Linn Soc* 133:381–461. <https://doi.org/10.1m/b0j1.2000.0380>
- Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moorer MJ, Carlswald BS, Bell CD, Latvis M, Crawley S, Black C, Diouf D, Xi Z, Rushworth CA, Gitzendanner MA, Sytsma K, Qiu Y-L, Hilu KW, Davis CC, Sanderson MJ, Beaman RS, Olmstead RG, Judd WS, Donoghue MJ, Soltis PS (2011) Angiosperm phylogeny: 17 genes, 640 taxa. *Am J Bot* 98:704–730. <https://doi.org/10.3732/ajb.1000404>
- Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A (2013) RNA editing in plants and its evolution. *Annu Rev Genet* 47:335–352. <https://doi.org/10.1146/annurev-genet-111212-133519>
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Tangphatsornruang S, Uthapaisanwong P, Sangsrakru D, Chanprasert J, Yoocha T, Jomchai N, Tragoonrun S (2011) Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene* 475:104–112. <https://doi.org/10.1016/j.gene.2011.01.002>
- Teixeira da Silva JA, Dobránszki J, Rivera-Madrid R (2018) The biotechnology (genetic transformation and molecular biology) of *Bixa orellana* L. (achiote). *Planta*. <https://doi.org/10.1007/s00425-018-2909-7> (in press)
- The Plant List (2013) Version 1.1. <http://www.theplantlist.org/>. Accessed 12 June 2018
- Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422. <https://doi.org/10.1007/s00122-002-1031-0>
- Tillich M, Funk HT, Schmitz-Linneweber C, Poltnigg P, Sabater B, Martin M, Maier RM (2005) Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J* 43:708–715. <https://doi.org/10.1111/j.1365-3113X.2005.02484.x>
- Tsai CC, Chou CH, Wang HV, Ko YZ, Chiang TY, Chiang YC (2015) Biogeography of the *Phalaenopsis amabilis* species complex inferred from nuclear and plastid DNAs. *BMC Plant Biol* 15:202. <https://doi.org/10.1186/s12870-015-0560-z>
- Tsuzuki T, Wakasugi T, Sugiura M (2001) Comparative analysis of RNA editing sites in higher plant chloroplasts. *J Mol Evol* 53:327–332. <https://doi.org/10.1007/s002390010222>
- Vieira LN, Faoro H, Fraga HPF, Rogalski M, de Souza EM, Pedrosa FO, Nodari RO, Guerra MP (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. *PLoS One* 9:e84792. <https://doi.org/10.1371/journal.pone.0084792>
- Vieira LN, Dos Anjos KG, Faoro H, Fraga HP, Greco TM, Pedrosa FO, de Souza EM, Rogalski M, de Souza RF, Guerra MP (2016a) Phylogenetic inference and SSR characterization of tropical woody bamboos tribe Bambuseae (Poaceae: Bambusoideae) based on complete plastid genome sequences. *Curr Genet* 62(2):443–453. <https://doi.org/10.1007/s00294-015-0549-z>
- Vieira LN, Rogalski M, Faoro H, Fraga HP, Anjos KG, Picchi GFA, Nodari RO, Pedrosa FO, Souza EM, Guerra MP (2016b) The plastome sequence of the endemic Amazonian conifer, *Retrophyllum piresii* (Silba) CNPage, reveals different recombination events and plastome isoforms. *Tree Genet Genomes* 12:10. <https://doi.org/10.1007/s11295-016-0968-0>
- Wakasugi T, Tsuzuki T, Sugiura M (2001) The genomics of land plant chloroplasts: gene content and alteration of genomic information by RNA editing. *Photosynth Res* 70:107–118. <https://doi.org/10.1023/A:1013892009589>
- Walker JF, Jansen RK, Zanis MJ, Emery NC (2015) Sources of inversion variation in the small single copy (SSC) region of chloroplast genomes. *Am J Bot* 102:1751–1752. <https://doi.org/10.3732/ajb.1500299>
- Wambulwa MC, Meegahakumbura MK, Kamunya S, Muchugi A, Möller M, Liu J, Xu JC, Ranjitkar S, Li DZ, Gao LM (2016) Insights into the genetic relationships and breeding patterns of the African tea germplasm based on nSSR markers and cpDNA sequences. *Front Plant Sci* 7:1244. <https://doi.org/10.3389/fpls.2016.01244>
- Wang Y, Zhan D-F, Jia X, Mei W-L, Dai H-F, Chen X-T, Peng S-Q (2016) Complete chloroplast genome sequence of *Aquilaria sinensis* (Lour.) Gilg and evolution analysis with in the Malvales order. *Front Plant Sci* 7:280. <https://doi.org/10.3389/fpls.2016.00280>
- Wang JH, Cai YC, Zhao KK, Zhu ZX, Zhou RC, Wang HF (2018) Characterization of the complete chloroplast genome sequence of *Firmiana pulcherrima* (Malvaceae). *Conserv Genet Resour* 10:445. <https://doi.org/10.1007/s12686-017-0845-7>
- Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE (2014a) A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Appl Plant Sci* 2:1400059. <https://doi.org/10.3732/apps.1400059>
- Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE (2014b) A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Appl Plant Sci* 2(12):1400. <https://doi.org/10.3732/apps.1400059>
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255. <https://doi.org/10.1093/bioinformatics/bth352>
- Yang J-B, Yang S-X, Li H-T, Yang J, Li D-Z (2013) Comparative chloroplast genomes of *Camellia* species. *PLoS One* 8:e73053. <https://doi.org/10.1371/journal.pone.0073053>
- Yu J, Ma PJ, Shi DJ, Li SM, Wang CL (2008) Homologous comparisons of photosynthetic system I genes among cyanobacteria and chloroplasts. *J Integr Plant Biol* 50:929–940. <https://doi.org/10.1111/j.1744-7909.2008.00679.x>
- Zaldívar-Cruz J, Ballina-Gómez H, Guerrero-Rodríguez C, Avilés-Berzunza E, Godoy-Hernández GC (2003) Agrobacterium-mediated transient transformation of annatto (*Bixa orellana*) hypocotyls with the GUS reporter gene. *Plant Cell Tissue Organ Cult* 73:281–284. <https://doi.org/10.1023/A:1023037108705>
- Zhai B, Clark J, Ling T, Connelly M, Medina-Bolivar F, Rivas F (2014) Antimalarial evaluation of the chemical constituents of hairy root cultures of *Bixa orellana*. *Molecules* 19:756–766. <https://doi.org/10.3390/molecules19010756>
- Zhang J, Khan SA, Heckel DG, Bock R (2017) Next-generation insect-resistant plants: RNAi-mediated crop protection. *Trends Biotechnol* 35:871–882. <https://doi.org/10.1016/j.tibtech.2017.04.009>
- Zhu A, Guo W, Gupta S, Fan W, Mower JP (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol* 209:1747–1756. <https://doi.org/10.1111/nph.13743>

SUPPLEMENTARY MATERIAL

Supplementary Figures

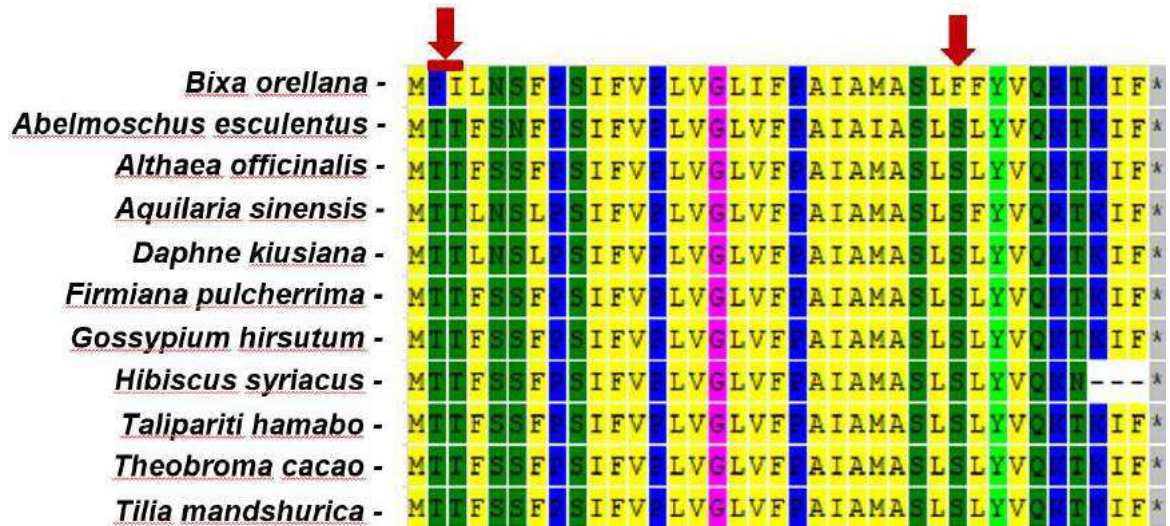


Figure S1. MUSCLE alignment of translated amino acid sequence of *psaI* gene of *B. orellana* and other ten species of Malvales. * indicates stop codon. Red arrows indicate specific substitution for annatto that includes exchange of amino acid polarity. Different background colors indicate amino acids with different biochemical properties.

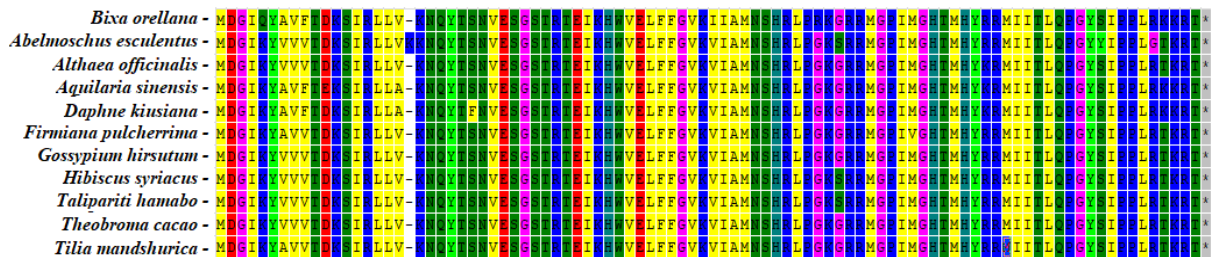


Figure S2. MUSCLE alignment of translated amino acid sequence of *rpl23* gene of *B. orellana* and other ten species of Malvales. * indicates stop codon. Different background colors indicate amino acids with different biochemical properties.

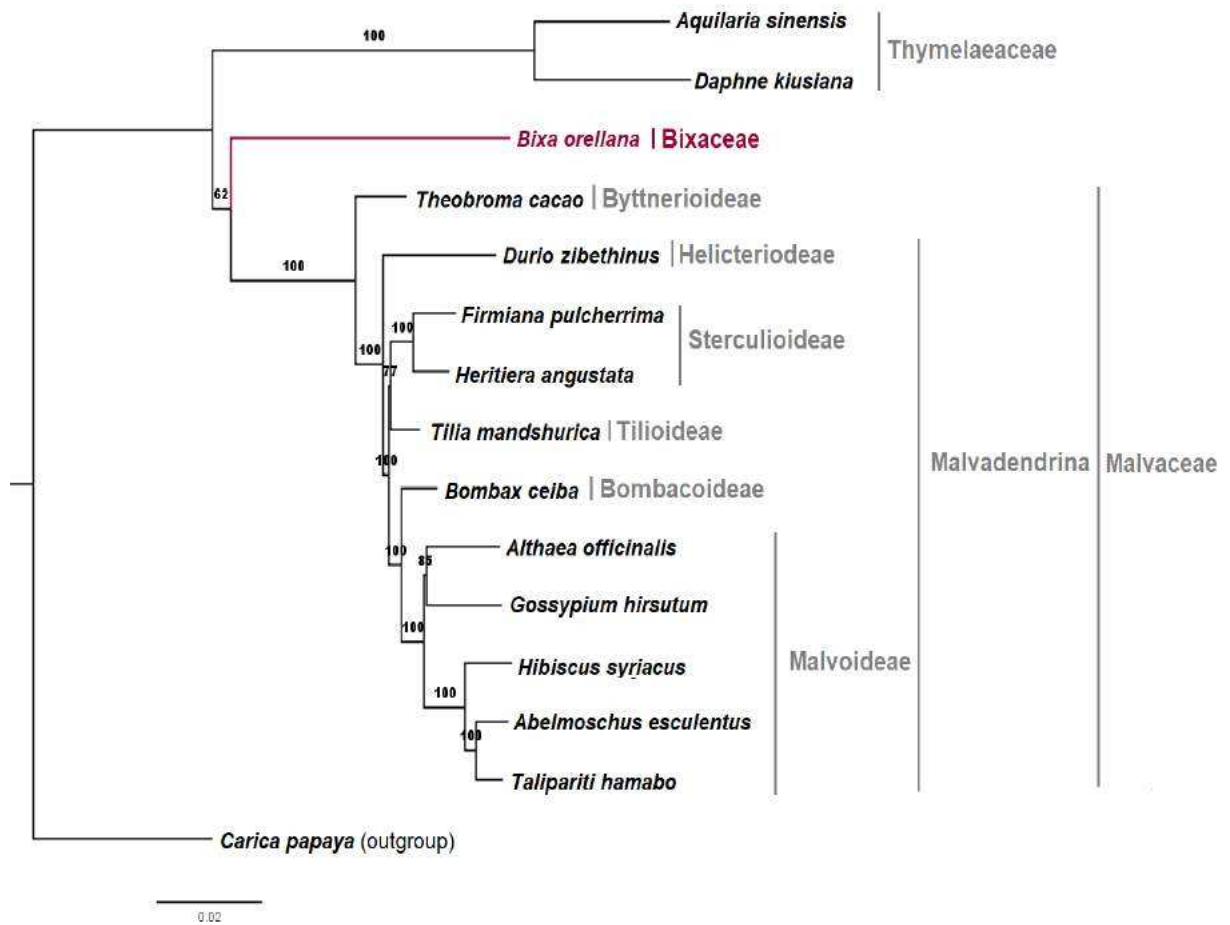


Fig. S3 Maximum likelihood (ML) phylogenomic analysis based on whole plastid genomes, showing the position of *B. orellana* (in red) within the order Malvales. Tree reconstruction was performed using IQTREE software, with TVM+F+R3 model. Numbers (%) associated with branches are ML bootstrap support values. The branch length is proportional to the inferred divergence level. The scale bar indicates the number of inferred nucleic acid substitutions per site. The subfamilies within the family Malvaceae are also highlighted.

Table S3. Distribution of SSR loci in the *B. orellana* plastome.

SSR type	SSR	Size	Start	End	Location
di	(TA)6	12	249	260	trnh-GUG/psbA (IGS)
mono	(A)9	9	479	487	trnh-GUG/psbA (IGS)
mono	(A)8	8	1852	1859	psbA/trk-UUU (IGS)
di	(AT)5	10	1892	1901	psbA/trk-UUU (IGS)
mono	(A)8	8	2062	2069	trk-UUU/matK (IGS)
mono	(T)8	8	2486	2493	matK (CDS)
di	(AT)5	10	4026	4035	matK/trnK-UUU (IGS)
di	(AT)7	14	4041	4054	matK/trnK-UUU (IGS)
mono	(T)8	8	4372	4379	matK/trnK-UUU (IGS)
mono	(A)10	10	4392	4401	matK/trnK-UUU (IGS)
mono	(T)8	8	4778	4785	trnk-UUU/rps16 (IGS)
mono	(A)10	10	4793	4802	trnk-UUU/rps16 (IGS)
mono	(A)10	10	4818	4827	trnk-UUU/rps16 (IGS)
mono	(C)10	10	5463	5472	rps16 (intron)
mono	(A)9	9	5511	5519	rps16 (intron)
di	(CT)4	8	5631	5638	rps16 (intron)
mono	(T)10	10	5901	5910	rps16 (intron)
mono	(T)9	9	6643	6651	rps16/trnQ-UUG (IGS)
mono	(A)9	9	6741	6749	rps16/trnQ-UUG (IGS)
mono	(T)9	9	6980	6988	rps16/trnQ-UUG (IGS)
mono	(A)12	12	7237	7248	rps16/trnQ-UUG (IGS)
mono	(A)10	10	7449	7458	rps16/trnQ-UUG (IGS)
mono	(A)10	10	7864	7873	rps16/trnQ-UUG (IGS)
mono	(A)11	11	8575	8585	trnQ-UUU/psbK (IGS)
mono	(T)8	8	8674	8681	psbK/psbI (IGS)
mono	(T)10	10	8880	8889	psbK/psbI (IGS)
mono	(A)10	10	8891	8900	psbK/psbI (IGS)
mono	(A)8	8	9245	9252	trnS-GCU/trnG-UCC (IGS)
di	(AT)7	14	9289	9302	trnS-GCU/trnG-UCC (IGS)
penta	(ATATT)3	15	9485	9499	trnS-GCU/trnG-UCC (IGS)
tetra	(AAAG)3	12	9526	9537	trnS-GCU/trnG-UCC (IGS)
mono	(A)8	8	9577	9584	trnS-GCU/trnG-UCC (IGS)
mono	(T)10	10	9672	9681	trnS-GCU/trnG-UCC (IGS)
mono	(T)8	8	10400	10407	trnG-UCC (intron)
mono	(T)8	8	10740	10747	trnG-UCC (intron)
mono	(T)8	8	10778	10785	trnG-UCC (intron)
mono	(A)8	8	11149	11156	trnG-UUC/ trnR-UCU (IGS)
penta	(ATTT)3	12	11329	11340	trnR-UCU/atpA (IGS)
mono	(T)11	11	11471	11481	trnR-UCU/atpA (IGS)
mono	(A)8	8	13538	13545	atpF (CDS)
mono	(A)10	10	13700	13709	atpF/atpH(IGS)
mono	(T)8	8	13721	13728	atpF/atpH(IGS)
di	(TA)4	8	13787	13794	atpF/atpH(IGS)
di	(TA)4	8	13803	13810	atpF/atpH(IGS)
mono	(T)10	10	13855	13864	atpF/atpH(IGS)
mono	(T)8	8	13874	13881	atpF/atpH (IGS)

mono	(T)11	11	14060	14070	atpF/atpH (IGS)
mono	(T)9	9	15453	15461	atpH/atpI (IGS)
mono	(T)8	8	15652	15659	atpH/atpI (IGS)
mono	(A)9	9	15775	15783	atpH/atpI (IGS)
mono	(A)8	8	16584	16591	atpI/rps2 (IGS)
mono	(A)10	10	17533	17542	rps2/rpoC2 (IGS)
mono	(A)9	9	17705	17713	rps2/rpoC2 (IGS)
mono	(C)8	8	17820	17827	rpoC2 (CDS)
mono	(T)9	9	17838	17846	rpoC2 (CDS)
mono	(T)9	9	19557	19565	rpoC2 (CDS)
mono	(T)11	11	19729	19739	rpoC2 (CDS)
mono	(A)8	8	19872	19879	rpoC2 (CDS)
mono	(T)9	9	20264	20272	rpoC2 (CDS)
di	(AT)5	10	21111	21120	rpoC2 (CDS)
mono	(A)8	8	23572	23579	rpoC1 (CDS)
mono	(A)8	8	24251	24258	rpoC1 (CDS)
mono	(T)9	9	24269	24277	rpoC1 (CDS)
di	(AT)5	10	28256	28265	rpoB/trnC-GCA (IGS)
mono	(A)9	9	28406	28414	rpoB/trnC-GCA (IGS)
mono	(A)10	10	28541	28550	rpoB/trnC-GCA (IGS)
di	(AG)4	8	28627	28634	rpoB/trnC-GCA (IGS)
mono	(T)9	9	28978	28986	rpoB/trnC-GCA (IGS)
mono	(A)9	9	28988	28996	rpoB/trnC-GCA (IGS)
mono	(T)10	10	29190	29199	rpoB/trnC-GCA (IGS)
mono	(A)9	9	29643	29651	trnC-GCA / petN (IGS)
di	(AT)4	8	29862	29869	trnC-GCA/petN (IGS)
di	(TA)4	8	29878	29885	trnC-GCA/petN (IGS)
mono	(T)12	12	30392	30403	petN/psbM (IGS)
mono	(T)8	8	30777	30784	petN/psbM (IGS)
mono	(T)9	9	30978	30986	psbM/trnD-GUC (IGS)
di	(TA)6	12	31226	31237	psbM/trnD-GUC (IGS)
mono	(T)10	10	31344	31353	psbM/trnD-GUC (IGS)
penta	(AAAGA)3	15	31400	31414	psbM/trnD-GUC (IGS)
mono	(T)8	8	31482	31489	psbM/trnD-GUC (IGS)
mono	(T)11	11	31732	31742	psbM/trnD-GUC (IGS)
mono	(A)10	10	32063	32072	psbM/trnD-GUC (IGS)
mono	(T)9	9	32140	32148	psbM/trnD-GUC (IGS)
tetra	(AAAG)3	12	32247	32258	trnD-GUC/trnY-GUA (IGS)
di	(AT)4	8	32272	32279	trnD-GUC/trnY-GUA (IGS)
mono	(A)8	8	32496	32503	trnD-GUC/trnY-GUA (IGS)
mono	(T)8	8	32538	32545	trnD-GUC/trnY-GUA (IGS)
mono	(A)10	10	32584	32593	trnD-GUC/trnY-GUA (IGS)
di	(GT)7	14	33461	33474	trnE-UUC/trnT-GGU (IGS)
mono	(T)8	7	33475	33481	trnE-UUC/trnT-GGU (IGS)
mono	(T)11	11	33487	33497	trnE-UUC/trnT-GGU (IGS)
di	(AT)9	18	33781	33798	trnE-UUC/trnT-GGU (IGS)
di	(TA)4	8	33808	33815	trnE-UUC/trnT-GGU (IGS)
di	(TA)4	8	33821	33828	trnE-UUC/trnT-GGU (IGS)
mono	(T)8	8	33972	33979	trnT-GGU/psbD (IGS)
di	(AT)5	10	34227	34236	trnT-GGU/psbD (IGS)

di	(AT)4	8	34869	34876	trnT-GGU/psbD (IGS)
di	(AT)5	10	34880	34889	trnT-GGU/psbD (IGS)
mono	(A)10	10	35047	35056	trnT-GGU/psbD (IGS)
mono	(G)9	9	37216	37224	psbC (CDS)
di	(AT)4	8	38116	38123	psbC/ trnS-UGA (IGS)
di	(GA)4	8	38321	38328	psbC/trnS-UGA (IGS)
di	(TA)4	8	38510	38517	psbC/trnS-UGA (IGS)
di	(TA)4	8	38532	38539	psbC/trnS-UGA (IGS)
mono	(A)9	9	39167	39175	psbz/trnG-UCC (IGS)
di	(AT)7	14	39235	39248	psbz/trnG-UCC (IGS)
di	(AT)4	8	39311	39318	psbz/trnG-UCC (IGS)
mono	(T)8	8	44971	44978	psaA/yfc3 (IGS)
mono	(T)9	9	45121	45129	psaA/yfc3 (IGS)
mono	(T)9	9	45180	45188	psaA/yfc3 (IGS)
di	(AT)5	10	45321	45330	psaA/yfc3 (IGS)
mono	(C)8	8	45600	45607	psaA/yfc3 (IGS)
tetra	(TATT)3	12	45690	45701	psaA/yfc3 (IGS)
mono	(T)8	8	45703	45710	psaA/yfc3 (IGS)
mono	(T)11	11	45715	45725	psaA/yfc3 (IGS)
mono	(T)8	8	46681	46688	yfc3 (Intron)
mono	(A)8	8	47130	47137	yfc3 (Intron)
mono	(T)9	9	47652	47660	yfc3 (Intron)
mono	(T)9	9	48026	48034	yfc3 (Intron)
mono	(A)8	8	48504	48511	yfc3/ trnS-GGA (IGS)
mono	(T)8	8	48739	48746	rps4 (CDS)
mono	(T)9	9	49184	49192	rps4/trnT-UGU (IGS)
mono	(T)13	13	49308	49320	rps4/trnT-UGU (IGS)
mono	(A)8	8	49800	49807	trnt-UGU/trnL-UAA (IGS)
tetra	(AATA)3	12	49821	49832	trnt-UGU/trnL-UAA (IGS)
mono	(A)9	9	49845	49853	trnt-UGU/trnL-UAA (IGS)
mono	(A)8	8	49972	49979	trnt-UGU/trnL-UAA (IGS)
di	(TA)4	8	50090	50097	trnt-UGU/trnL-UAA (IGS)
di	(TA)4	8	50100	50107	trnt-UGU/trnL-UAA (IGS)
di	(TA)4	8	50115	50122	trnt-UGU/trnL-UAA (IGS)
mono	(A)8	8	50224	50231	trnt-UGU/trnL-UAA (IGS)
tetra	(TAAT)3	12	50336	50347	trnt-UGU/trnL-UAA (IGS)
mono	(T)8	8	50474	50481	trnt-UGU/trnL-UAA (IGS)
mono	(A)10	10	50507	50516	trnt-UGU/trnL-UAA (IGS)
di	(AT)4	8	50577	50584	trnt-UGU/trnL-UAA (IGS)
mono	(A)10	10	50772	50781	trnt-UGU/trnL-UAA (IGS)
mono	(T)10	10	51187	51196	trnL-UAA (intron)
tetra	(TATT)3	12	51505	51516	trnL-UAA/ trnF-GAA (IGS)
mono	(T)9	7	51517	51523	trnL-UAA/ trnF-GAA (IGS)
mono	(T)11	11	51695	51705	trnL-UAA/ trnF-GAA (IGS)
mono	(T)11	11	51954	51964	trnF-GAA/ ndhJ (IGS)
mono	(T)9	9	51966	51974	trnF-GAA/ ndhJ (IGS)
mono	(A)11	11	52074	52084	trnF-GAA/ ndhJ (IGS)
mono	(T)11	11	52438	52448	trnF-GAA/ ndhJ (IGS)
mono	(T)9	9	53051	53059	ndhJ/ndhK (IGS)
mono	(A)9	9	53087	53095	ndhJ/ndhK (IGS)

mono	(A)12	12	54593	54604	ndhC/trnV-UAC (IGS)
mono	(T)10	10	54632	54641	ndhC/trnV-UAC (IGS)
di	(AT)4	8	54710	54717	ndhC/trnV-UAC (IGS)
di	(AT)5	10	54725	54734	ndhC/trnV-UAC (IGS)
tetra	(TAAA)3	12	55046	55057	trnV-UAC (intron)
mono	(T)11	11	57832	57842	atpB/rbcL (IGS)
mono	(T)9	9	58184	58192	atpB/rbcL (IGS)
mono	(A)11	11	58415	58425	atpB/rbcL (IGS)
di	(AT)4	8	58989	58996	rbcL (CDS)
mono	(A)9	9	60588	60596	rbcL/accD (IGS)
mono	(T)9	9	61124	61132	accD (CDS)
mono	(T)8	8	62442	62449	accD/psaI (IGS)
mono	(A)8	8	62474	62481	accD/psaI (IGS)
di	(TA)4	8	62576	62583	accD/psaI (IGS)
di	(AT)5	10	62599	62608	accD/psaI (IGS)
tri	(ATA)4	12	62787	62798	accD/psaI (IGS)
di	(TA)4	8	62913	62920	accD/psaI (IGS)
tri	(TAT)4	12	62959	62970	accD/psaI (IGS)
mono	(T)11	11	63184	63194	psaI/ycf4 (IGS)
di	(AT)4	8	63219	63226	psaI/ycf4 (IGS)
mono	(T)8	8	63742	63749	ycf4 (CDS)
mono	(T)8	8	64282	64289	ycf4/cemA (IGS)
mono	(T)8	8	64582	64589	ycf4/cemA (IGS)
di	(AT)5	10	64661	64670	ycf4/cemA (IGS)
di	(TA)5	10	64674	64683	ycf4/cemA (IGS)
di	(AT)4	8	64868	64875	ycf4/cemA (IGS)
mono	(T)9	9	64922	64930	ycf4/cemA (IGS)
mono	(A)8	8	64955	64962	ycf4/cemA (IGS)
mono	(A)9	9	65088	65096	ycf4/cemA (IGS)
di	(TC)5	10	65163	65172	cemA (CDS)
mono	(A)8	8	66448	66455	petA (CDS)
mono	(T)10	10	67188	67197	petA/ psbJ (IGS)
mono	(T)13	13	67518	67530	petA/ psbJ (IGS)
tri	(TTA)5	15	67560	67574	petA/ psbJ (IGS)
tetra	(TAAT)3	12	67722	67733	petA/ psbJ (IGS)
mono	(T)9	9	67899	67907	petA/ psbJ (IGS)
mono	(A)9	9	68217	68225	psbJ (CDS)
mono	(T)8	8	69368	69375	psbE/petL (IGS)
mono	(T)10	10	69418	69427	psbE/petL (IGS)
mono	(A)10	10	69503	69512	psbE/petL (IGS)
mono	(A)11	11	69890	69900	psbE/petL (IGS)
mono	(T)9	9	70015	70023	psbE/petL (IGS)
mono	(T)8	8	70028	70035	psbE/petL (IGS)
mono	(A)10	10	70424	70433	petL/petG (IGS)
tetra	(TTAA)3	12	70709	70720	petG/ trnW-CCA (IGS)
mono	(T)8	8	70745	70752	petG/ trnW-CCA (IGS)
mono	(T)8	8	70955	70962	trnW-CCA/trnP-UGG (IGS)
di	(AT)5	10	70997	71006	trnW-CCA/trnP-UGG (IGS)
mono	(T)9	9	71278	71286	trnP-UGG/ psaJ (IGS)
penta	(AATAA)3	15	71518	71532	trnP-UGG/ psaJ (IGS)

mono	(T)8	8	71668	71675	psaJ/rpL 33 (IGS)
mono	(T)9	9	71750	71758	psaJ/rpL 33 (IGS)
mono	(T)8	8	71999	72006	psaJ/rpL 33 (IGS)
di	(TA)5	10	72033	72042	psaJ/rpL 33 (IGS)
di	(TA)5	10	72067	72076	psaJ/rpL 33 (IGS)
mono	(T)9	9	72171	72179	psaJ/rpL 33 (IGS)
di	(TA)4	8	72514	72521	rpl33/rps18 (IGS)
mono	(A)9	9	72562	72570	rpl33/rps18 (IGS)
mono	(T)10	10	72589	72598	rpl33/rps18 (IGS)
tetra	(TTAA)3	12	72632	72643	rpl33/rps18 (IGS)
mono	(A)11	11	72650	72660	rpl33/rps18 (IGS)
mono	(T)8	8	73263	73270	rps18/rpl20 (IGS)
mono	(C)8	8	73552	73559	rpl20 (CDS)
mono	(T)8	8	74442	74449	rpl20/rps12 (IGS)
mono	(T)8	8	74853	74860	rps12/clpP (IGS)
mono	(A)10	10	75160	75169	clpP (intron)
mono	(T)12	12	75327	75338	clpP (intron)
mono	(A)9	9	75350	75358	clpP (intron)
di	(TA)8	16	75392	75407	clpP (intron)
di	(TA)12	24	75410	75433	clpP (intron)
mono	(T)11	11	75466	75476	clpP (intron)
mono	(T)9	9	75512	75520	clpP (intron)
mono	(A)9	9	75570	75578	clpP (intron)
mono	(A)9	9	75674	75682	clpP (intron)
mono	(A)8	8	75784	75791	clpP (intron)
mono	(A)8	8	76168	76175	clpP (intron)
mono	(T)11	11	76259	76269	clpP (intron)
mono	(T)8	8	76292	76299	clpP (intron)
di	(AT)4	8	76333	76340	clpP (intron)
mono	(T)8	8	76546	76553	clpP (intron)
mono	(A)11	11	76639	76649	clpP (intron)
mono	(T)8	8	76723	76730	clpP (intron)
tri	(ATT)4	12	77294	77305	clpP/psbB (IGS)
mono	(A)10	10	81502	81511	petB/petD (IGS)
mono	(A)9	9	81609	81617	petB/petD (IGS)
mono	(T)10	10	81891	81900	petD (intron)
di	(AT)4	8	85064	85071	infA/ rps8 (IGS)
mono	(T)10	10	85557	85566	rps8/ rpl14 (IGS)
mono	(A)12	12	85567	85578	rps8/ rpl14 (IGS)
mono	(A)11	11	86109	86119	rpl14/rpl16 (IGS)
mono	(A)11	11	86648	86658	rpl16 (intron)
di	(AT)5	10	86697	86706	rpl16 (intron)
di	(TA)4	8	86877	86884	rpl16 (intron)
mono	(T)8	8	86958	86965	rpl16 (intron)
mono	(A)9	9	87236	87244	rpl16 (intron)
mono	(T)10	10	87365	87374	rpl16 (intron)
mono	(T)10	10	87675	87684	rpl16 (intron)
mono	(A)9	9	87805	87813	rpl16/rps3 (IGS)
di	(CG)4	8	88026	88033	rps3 (CDS)
di	(TA)4	8	88565	88572	rps3/rpl22 (IGS)

mono	(T)8	8	88723	88730	rpl22 (CDS)
mono	(T)8	8	89316	89323	rps 19 (CDS)
mono	(T)9	9	89348	89356	rps 19 (CDS)
mono	(T)9	9	89391	89399	rps 19 (CDS)
di	(AT)4	8	89986	89993	rpl12 (intron)
di	(GA)4	8	91619	91626	trnI-CAU/ycf2 (IGS)
di	(GA)4	8	91631	91638	ycf2 (CDS)
di	(GA)4	8	92618	92625	ycf2 (CDS)
mono	(A)9	9	94797	94805	ycf2 (CDS)
di	(GA)4	8	94818	94825	ycf2 (CDS)
di	(TA)4	8	98174	98181	ycf2 (CDS)
di	(TA)4	8	99619	99626	trnL-CAA/ndhB (IGS)
di	(AT)4	8	99642	99649	trnL-CAA/ndhB (IGS)
di	(AG)4	8	100412	100419	ndhB (IGS)
mono	(T)10	10	100940	100949	ndhB (intron)
mono	(T)9	9	104156	104164	rps 19 (CDS)
mono	(T)8	8	104336	104343	rps12/trnV-GAC (IGS)
mono	(T)8	8	104351	104358	rps12/trnV-GAC (IGS)
mono	(T)10	10	104364	104373	rps12/trnV-GAC (IGS)
di	(CT)4	8	111689	111696	rrn23 (CDS)
mono	(A)8	8	113444	113451	rrn5/trnR-ACG (IGS)
mono	(T)9	9	113593	113601	rrn5/trnR-ACG (IGS)
di	(TA)4	8	113872	113879	trnR-ACG/ trnN-GUU (IGS)
mono	(T)10	10	113908	113917	trnR-ACG/ trnN-GUU (IGS)
mono	(A)9	9	114710	114718	trnN-GUU/ycf1 (IGS)
mono	(A)8	8	114724	114731	ycf1 (CDS)
mono	(G)8	8	115550	115557	ycf1 (CDS)
mono	(A)9	9	115731	115739	ycf1 (CDS)
mono	(T)8	8	115943	115950	ycf1 (CDS)
mono	(A)9	9	116345	116353	ycf1 (CDS)
mono	(T)9	9	116534	116542	ycf1 (CDS)
mono	(A)9	9	117070	117078	ycf1 (CDS)
mono	(A)8	8	117080	117087	ycf1 (CDS)
mono	(A)8	8	117091	117098	ycf1 (CDS)
mono	(A)8	8	117483	117490	ycf1 (CDS)
mono	(A)8	8	117498	117505	ycf1 (CDS)
mono	(A)8	8	117605	117612	ycf1 (CDS)
mono	(A)9	9	117925	117933	ycf1 (CDS)
mono	(A)8	8	118232	118239	ycf1 (CDS)
mono	(T)8	8	118313	118320	ycf1 (CDS)
mono	(A)9	9	118714	118722	ycf1 (CDS)
mono	(T)8	8	119359	119366	ycf1 (CDS)
mono	(A)8	8	119485	119492	ycf1 (CDS)
tetra	(GTTA)3	12	120397	120408	ycf1 (CDS)
mono	(A)9	9	120691	120699	ycf1/rps15 (IGS)
mono	(A)8	8	120746	120753	rps15 (CDS)
mono	(A)8	8	122994	123001	ndhA (inter)
mono	(A)8	8	123262	123269	ndhA (inter)
tetra	(AGAA)3	12	123460	123471	ndhA (inter)
mono	(A)8	8	123501	123508	ndhA (inter)

mono	(A)8	8	124538	124545	ndhA/ndhI (IGS)
mono	(A)8	8	125182	125189	ndhI/ndhG (IGS)
mono	(T)10	10	125275	125284	ndhI/ndhG (IGS)
di	(TA)4	8	125342	125349	ndhI/ndhG (IGS)
mono	(A)11	11	125368	125378	ndhI/ndhG (IGS)
mono	(A)10	10	125483	125492	ndhI/ndhG (IGS)
mono	(A)9	9	127204	127212	psaC/ndhD (IGS)
mono	(A)8	8	128175	128182	ndhD (CDS)
mono	(A)9	9	128778	128786	ndhD/ccsA (IGS)
tri	(AAT)5	15	128838	128852	ndhD/ccsA (IGS)
di	(TA)6	12	128973	128984	ndhD/ccsA (IGS)
mono	(A)11	11	130010	130020	ccsA/ trnL- UAG (IGS)
mono	(T)8	8	130416	130423	trnL-UAG/rpl32 (IGS)
mono	(T)9	9	130531	130539	trnL-UAG/rpl32 (IGS)
mono	(T)12	12	130867	130878	trnL-UAG/rpl32 (IGS)
mono	(A)8	8	131075	131082	rpl32/ndhF (IGS)
mono	(T)8	8	131154	131161	rpl32/ndhF (IGS)
tetra	(TTAT)3	12	131163	131174	rpl32/ndhF (IGS)
mono	(A)9	9	131220	131228	rpl32/ndhF (IGS)
di	(AT)4	8	131232	131239	rpl32/ndhF (IGS)
mono	(A)9	9	131247	131255	rpl32/ndhF (IGS)
mono	(A)10	10	131391	131400	rpl32/ndhF (IGS)
mono	(A)8	8	131640	131647	rpl32/ndhF (IGS)
mono	(T)11	11	131937	131947	rpl32/ndhF (IGS)
mono	(T)8	8	132056	132063	ndhF (CDS)
mono	(T)8	8	133550	133557	ndhF (CDS)
mono	(T)8	8	133782	133789	ndhF (CDS)

CDS, coding sequences; IGS, intergenic spacers.

Table S4. Distribution of tandem repeats in the *B. orellana* plastome.

Copy number	Consensus length	Start	End	Location
2	18	192	227	accD/psaI (IGS)
2	42	245	331	ycf2 (CDS)
3	17	9349	9399	ycf2 (CDS)
2	34	11293	11357	trnN-GUU/ycf1 (IGS)
2	20	33086	33125	trnE-UUC/ trnT-GGU (IGS)
4	12	34834	34881	trnT-GGU/psbD (IGS)
2	17	35199	35234	trnT-GGU/psbD (IGS)
2	36	35400	35480	trnT-GGU/psbD (IGS)
3	16	45791	45834	psaA/ ycf3 (IGS)
3	13	48401	48437	trnS-GCA/ rps4 (IGS)
3	22	49784	49850	Rps4/ trnT-UGU (IGS)
3	24	54659	54728	ndhC/ trnV-UAC (IGS)
3	18	62589	62645	accD/ psaI (IGS)
4	14	67732	67785	petA/ psbJ (IGS)
2	34	72006	72080	psaJ/ rpl33 (IGS)
4	15	73066	73117	rps18/ rpl20 (IGS)

				ycf2 (CDS)
3	21	94174	94242	
2	22	99620	99664	ycf2 (CDS)
2	32	113059	113120	rrn4.5/ rr5 (IGS)

Table S5. Distribution of directed (D) and inverted (I) repeat loci in the *B. orellana* plastome.

Type	Size (bp)	Repeat 1 (Start)	Repeat 2 (Start)	Repeat 1 (Location)	Repeat 2 (Location)
D	39	47055	103965	Ycf3 (CDS)	Rps12/trnV-GAC (IGS)
D	34	41731	43955	psaB (CDS)	psaA (CDS)
D	34	33781	75391	trnE-UCC/ trnT-GGU (IGS)	clpP (intron)
D	30	23716	86611	rpoC1 (intron)	rps16 (Intron)
D	32	9110	38315	psbI/trnS-GCU (IGS)	psbC/ trnS-UGA (IGS)
D	31	9271	75399	trnS-GCU/trnG-UCC (IGS)	clpP (Intron)
D	30	11460	45689	trnR-UCU/ atpA (IGS)	psaA/ ycf3 (IGS)
D	30	54657	67566	ndhC /trnV-UAC (IGS)	petA/ psbJ (IGS)
I	43	47049	123920	ycf3 (Intron)	ndhA (Intron)
I	30	9112	48199	psbI/ trnS-GCU (IGS)	trnS-GGA (CDS)
I	30	10967	11004	trnG-UCC/ trnR-UCU (IGS)	trnG-UCC/ trnR-UCU (IGS)
I	30	9468	76314	ycf2 (CDS)	clpP (intron)
I	30	47064	123918	ycf3 (intron)	ndhA (Intron)
I	30	235	31220	trnH-GUG/ psbA (IGS)	psbm/ trnD-GUG (IGS)

Table S6. RNA-editing sites predicted in protein-coding genes of the plastome of *B. orellana* and eight species of family Malvaceae. Edited codons are indicated by arrow (\Rightarrow). AA pos, amino acid position. * indicates RNA editing sites putatively unique for *B. orellana*.

Malvaceae										
Gene	AA pos.	<i>B. orellana</i>	<i>A. exaltensis</i>	<i>A. officinalis</i>	<i>F. pulcherrima</i>	<i>G. hirsutum</i>	<i>H. sycocaus</i>	<i>T. hamabo</i>	<i>T. cacao</i>	<i>T. mandshurica</i>
<i>accD</i>	263	UCG(S) \Rightarrow UUG(L)	UCG(S) \Rightarrow UUG(L)	UCG(S) \Rightarrow UUG(L)	UCG(S) \Rightarrow UUG(L)	UCG(S) \Rightarrow UUG(L)	UCG(S) \Rightarrow UUG(L)	UCG(S) \Rightarrow UUG(L)	UCG(S) \Rightarrow UUG(L)	UCG(S) \Rightarrow UUG(L)
	466	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CUU(L)	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CCG(P) \Rightarrow CUG(L)	CCG(P) \Rightarrow CUG(L)
<i>atpA</i>	264	CCU(P) \Rightarrow CUU(L)	CUU(L)	CUU(L)	CUU(L)	CUU(L)	CUU(L)	CUU(L)	CUU(L)	CUU(L)
	305	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)
	383	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)
<i>atpB</i>	135	CCU(P) \Rightarrow UCU(S)	CCC(P) \Rightarrow UCC(S)	UCU(S)	CCU(P) \Rightarrow UCU(S)	UCC(S)	CCU(P) \Rightarrow UCU(S)	CCU(P) \Rightarrow UCU(S)	CCC(P) \Rightarrow UCC(S)	CCC(P) \Rightarrow UCC(S)
	<i>atpF</i>	31	CUA(L)	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)
88		GCA(A) \Rightarrow GUA(V)*	GUA(V)	GUA(V)	GUG(V)	GUA(V)	GUA(V)	GUA(V)	GUG(V)	GUG(V)
<i>ccsA</i>	128	GUA(V)	GCG(A) \Rightarrow GUG(V)	GCG(A) \Rightarrow GUG(V)	GCG(A) \Rightarrow GUG(V)	GCA(A) \Rightarrow GUA(V)	GCG(A) \Rightarrow GUG(V)	GCG(A) \Rightarrow GUG(V)	GCG(A) \Rightarrow GUG(V)	GCG(A) \Rightarrow GUG(V)
	217	ACC(U) \Rightarrow AUC(I)	ACG(U)	ACU(U) \Rightarrow AUU(I)	ACU(U) \Rightarrow AUU(I)	ACU(U) \Rightarrow AUU(I)	ACU(U) \Rightarrow AUU(I)	ACU(U) \Rightarrow AUU(I)	ACU(U) \Rightarrow AUU(I)	ACU(U) \Rightarrow AUU(I)
<i>clpP</i>	187	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	-	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)
	<i>matK</i>	160	CAC(H) \Rightarrow UAC(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAC(H) \Rightarrow UAC(Y)
219		CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)
271	CGG(R) \Rightarrow UGG(W)*	UGG(W)	UGG(W)	UGG(W)	UGG(W)	UGG(W)	UGG(W)	UGG(W)	UGG(W)	UGG(W)
403	UCA(S) \Rightarrow UUA(L)*	UCC(S)	UCU(S)	UCC(S)	UCC(S)	UCC(S)	UCC(S)	UCC(S)	UCC(S)	UCC(S)
423	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)
<i>ndhA</i>	114	UCA(S) \Rightarrow UUA(L)	GCA(A)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	GCA(A)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)
	189	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)

Chapter II

Phylogenetic and evolutionary features of the plastome of *Tropaeolum pentaphyllum* Lam. (Tropaeolaceae)

Túlio Gomes Pacheco¹, Gleyson Morais da Silva¹, Amanda de Santana Lopes¹, José Daniel de Oliveira¹, Juliana Marcia Rogalski², Eduardo Balsanelli³, Emanuel Maltempi de Souza³, Fábio de Oliveira Pedrosa³, Marcelo Rogalski^{1*}

¹ Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa-MG, Brazil

² Núcleo de Ciências Biológicas e Ambientais, Instituto Federal do Rio Grande do Sul, Distrito Engenheiro Luiz Englert, Sertão-RS, Brazil

³ Núcleo de Fixação Biológica de Nitrogênio, Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Curitiba-PR, Brazil

*Corresponding author

E-mail address: rogalski@ufv.br

Published in:

Planta (2020) 252, 17

<https://doi.org/10.1007/s00425-020-03427-w>



Phylogenetic and evolutionary features of the plastome of *Tropaeolum pentaphyllum* Lam. (Tropaeolaceae)

Túlio Gomes Pacheco¹ · Gleyson Morais da Silva¹ · Amanda de Santana Lopes¹ · José Daniel de Oliveira¹ · Juliana Marcia Rogalski² · Eduardo Balsanelli³ · Emanuel Maltempi de Souza³ · Fábio de Oliveira Pedrosa³ · Marcelo Rogalski¹

Received: 17 March 2020 / Accepted: 8 July 2020
 © Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Main conclusion Complete plastome sequence of *Tropaeolum pentaphyllum* revealed molecular markers, hotspots of nucleotide polymorphism, RNA editing sites and phylogenetic aspects

Abstract Tropaeolaceae Juss. ex DC. comprises approximately 95 species across North and South Americas. *Tropaeolum pentaphyllum* Lam. is an unconventional and endangered species with occurrence in some countries of South America. Although this species presents nutritional, medicinal and ornamental uses, genetic studies involving natural populations or promising genotypes are practically non-existent. Here, we report the nucleotide sequence of *T. pentaphyllum* plastome. It represents the first complete plastome sequence of the family Tropaeolaceae to be fully sequenced and analyzed in detail. The sequencing data revealed that the *T. pentaphyllum* plastome is highly similar to the plastomes of other Brassicales. Notwithstanding, our analyses detected some specific features concerning events of IR expansion and structural changes in some genes such as *matK*, *rpoA*, and *rpoC2*. We also detected 251 SSR loci, nine hotspots of nucleotide polymorphism, and two specific RNA editing sites in the plastome of *T. pentaphyllum*. Moreover, plastid phylogenomic inference indicated a closed relationship between the families Tropaeolaceae and Akaniaceae, which formed a sister group to Moringaceae–Caricaceae. Finally, our data bring new molecular markers and evolutionary features to be applied in the natural population, germplasm collection, and genotype selection aiming conservation, genetic diversity evaluation, and exploitation of this endangered species.

Keywords Organelle DNA · Molecular markers · Extranuclear inheritance · Brassicales · Plastid evolution · Neglected crops

Communicated by Anastasios Melis.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00425-020-03427-w>) contains supplementary material, which is available to authorized users.

✉ Marcelo Rogalski
 rogalski@ufv.br

- ¹ Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa, MG, Brazil
- ² Núcleo de Ciências Biológicas e Ambientais, Instituto Federal do Rio Grande do Sul, Distrito Engenheiro Luiz Englert, Sertão, RS, Brazil
- ³ Núcleo de Fixação Biológica de Nitrogênio, Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Curitiba, PR, Brazil

Introduction

The family Tropaeolaceae Juss. ex DC. comprises approximately 95 species of the Americas. The family represents annual or perennial climbing or prostrate herbs that contain rhizomes and tubers with an odor of mustard oils (Andersson and Andersson 2000; Bayer and Appel 2003; Edger et al. 2018). Tropaeolaceae is neotropical and most species occur in areas with a higher altitude of South America, with less frequency and abundance of species in North America (Bayer and Appel 2003; Souza and Lorenzi 2008; Cardinal-McTeague et al. 2016). The first infrageneric classification (Cronquist 1988) divided the family into three genera: *Magallana*, *Trophaeastrum*, and *Tropaeolum*. More recently, the accepted classification based on molecular analyses

suggested the existence of a single genus, *Tropaeolum* (Andersson and Andersson 2000).

Tropaeolaceae presents species of ornamental and gastronomic importance such as *Tropaeolum majus* L. (nasturtium or monks cress), *T. pentaphyllum* (crem) Lam., and *T. tuberosum* Ruiz & Pav. (mashua). The four species, *Tropaeolum pentaphyllum*, *T. peregrinum* L., *T. sanctae-catharinae* Sparre, and *T. warmingianum* Rohrb., are native from Brazil (Souza and Lorenzi 2008). *T. pentaphyllum* has distribution in four South American countries, including Brazil, Uruguay, central and northeast regions of Argentina, and some regions of high land in Bolivia (Sparre 1972; Sparre and Andersson 1991; Fabri and Valla 1998; Souza and Lorenzi 2008; Rix 2010; Kinupp et al. 2011).

In Brazil, *T. pentaphyllum* is popularly known as *batata-crem* or *crem* and represents an unconventional food plant (Sparre 1972; Kinupp et al. 2011; Kinupp and Lorenzi 2014). Generally, the plants grow in well-drained, fertile and organic matter-rich soils (Sparre 1972). *T. pentaphyllum* was listed as one of the plants with the potential for economic use by the project “Plants for the Future—South Region/Brazil” (Kinupp et al. 2011). However, this species is suffering from the loss of natural habitat due to deforestation to increase the areas cultivated with soybeans and cereals. Additionally, the tubers are extracted from natural forest fragments for own consumption and/or sale in small farmer markets (agricultural fairs), which are significantly contributing to reduce natural populations from native forest fragments (Kinupp et al. 2011). Florists and gardeners also use the species as an ornamental resource due to its beautiful flowers and interesting growth habit (Kinupp et al. 2011).

In addition to the tubers, leaves and flowers are also eaten but on a smaller scale. The loss of natural habitat and the indiscriminate use have exerted great pressure on the natural resources of this species; therefore, it integrates the list of endangered species in Brazil. Concerning the tubers, they contain high contents of vitamin C, a fatty acid profile rich in linoleic acid and essential minerals such as sulfur, calcium, and phosphorus (Braga et al. 2018). Also, the tubers have demonstrated antimicrobial, depurative, and antiscorbutic activities (Correa 1984; Mors et al. 2000). Additionally, the tubers have been used to prevent hypercholesterolemia (Kinupp and Lorenzi 2014) and diabetes (Trojan-Rodrigues et al. 2012). Recently, the characterization of dietary constituents and antioxidant capacity of *T. pentaphyllum* leaves, flowers, and tubers show high nutritional quality and antioxidant potential, highlighting the potential of this species as an important source of nutritional compounds (De Bona et al. 2017).

Genetic studies were carried out in some species of the family Tropaeolaceae (Li and Quiros 2001; Ortega et al. 2007; Couvreur et al. 2010; Malice et al. 2010), whereas genetic studies involving *T. pentaphyllum* are practically

non-existent. Genetic studies regarding genotype characterization and genetic diversity of the species will be useful for its conservation and management. Recently, a cytogenetic investigation was carried out by Tolomeotti et al. (2018) comparing with other species of the genus *Tropaeolum*, which revealed that *T. pentaphyllum* is tetraploid containing $2n=4x=28$ chromosomes contrasting to *T. tuberosum*, a tetraploid $2n=4x=52$ (Malice and Baudoin 2009). Genetic studies are essential for the identification and characterization of molecular markers given that they enable the evaluation of the genetic diversity of natural populations and the establishment of suitable strategies for conservation, domestication, and breeding. The plastid genome (plastome), a nonrecombinant inherited DNA molecule, is a suitable source of molecular markers located mainly in the intergenic spacers (IGSs) and introns, where mutation rates are higher in comparison with highly conserved coding sequences (Rogalski et al. 2015; Vieira et al. 2016; Lopes et al. 2018a). Plastid sequences are basic information for several genetic studies such as phylogeographical, population genetics and germplasm collection analyses (Wambulwa et al. 2016; Roy et al. 2016; Stefenon et al. 2019). Among plastid molecular markers, simple sequence repeats (SSRs or plastid microsatellites) are widely recognized as excellent genetic markers due to the extensive level of polymorphism found in these sequences and the typical uniparental inheritance of the plastome (Provan et al. 2001; Wheeler et al. 2014; Rogalski et al. 2015). SSRs consist of short DNA sequences repeated in tandem, present across plastomes of all land plants (Provan et al. 2001; Wheeler et al. 2014; George et al. 2015), and can be assessed for both intraspecific and interspecific genetic studies (Powell et al. 1995; Provan et al. 2001; Rogalski et al. 2015; Stefenon et al. 2019).

Furthermore, complete plastome sequences bring fundamental information to understand evolutionary events and are highly efficient to resolve phylogenetic relationships (Barrett et al. 2016; Lopes et al. 2018b, 2019). Plastomes are usually highly conserved, thereby a unique or few evolutionary events such as structural rearrangements, the presence of gene degeneration, events of gene transfer to the nucleus, and divergence of RNA editing sites enable the characterization of specific lineages (Rogalski et al. 2008; Alkatib et al. 2012; Martin et al. 2014; Vieira et al. 2016; Bock 2017; Lopes et al. 2018c; Pacheco et al. 2020).

The resolution of phylogenetic relationships within the order Brassicales is established by several studies; however, various analyses have omitted some key taxa. Several studies strongly support the sister groups Caricaceae–Morinaceae and Tropaeolaceae–Akaniaceae within Brassicales (Olson 2002a, b; Hall et al. 2004; Cardinal-McTeague et al. 2016). The Akaniaceae–Tropaeolaceae clades are the earliest-diverging families of the order Brassicales (Ronse De Craene and Haston 2006; Cardinal-McTeague et al. 2016;

Edger et al. 2018). The time estimated for the subsequent divergence of the families Tropaeolaceae and Akaniaceae is 75 Mya in the Late Cretaceous (Cardinal-McTeague et al. 2016; Edger et al. 2018). A phylogenetic relationship was estimated among 35 species of Tropaeolaceae based on plastid and nuclear sequences, which suggests a subdivision of the genus *Tropaeolum* into two main sections *Tropaeolum* sect. *Tropaeolum* and T. sect. *Chilensia*. The first one should include the formerly recognized sections *Bicolora*, *Dipetala*, *Mucoronata*, *Schizotrophaeum*, *Serratociliata*, *Tropaeolum*, and *Umbellata*. The second one should include the formerly recognized genera *Magallana* and *Trophaeastrum*, and the formerly accepted T. sect. *Chymocarpus* (Andersson and Andersson 2000). However, the lack of molecular data is a bottleneck to infer accurately the relationships within the family Tropaeolaceae.

Here, we describe the complete plastome of *T. pentaphyllum*, which was molecularly characterized in detail. Our plastid phylogenomic analysis indicated with high support a closed relationship between the families Tropaeolaceae and Akaniaceae, which is congruent with several other phylogenetic studies. Additionally, we identified and characterized several SSR loci and hotspots of nucleotide polymorphism. Moreover, we predicted here two RNA editing sites that seem to be unique to *T. pentaphyllum* plastome. Furthermore, some specific features were also detected in the plastome of *T. pentaphyllum*, such as IR expansion events, and divergent genes related to gene expression machinery. Taken together, our study brings new molecular data useful for genetic studies in natural populations of this species, other species of the family Tropaeolaceae, and order Brassicales given that most studies carried out on species of the order Brassicales focus on the family Brassicaceae.

Material and methods

Plant material and plastid DNA extraction

Young leaves were collected from a single *T. pentaphyllum* plant growing under greenhouse conditions at the Department of Plant Biology, Federal University of Viçosa, Viçosa-MG, Brazil. The leaves were kept in dark for 96 h at 4 °C to reduce starch content in chloroplasts. Chloroplast isolation and plastid DNA extraction were carried out according to Vieira et al. (2014).

Plastid genome sequencing, assembling, and annotation

The sequencing library was prepared with approximately 1 ng of plastid DNA using the sample preparation kit NexteraXT (Illumina Inc., San Diego, CA), according to

the manufacturer's instructions. The obtained library was sequenced using MiSeq Reagent Kit v3 (600 cycles) on Illumina MiSeq Sequencer (Illumina Inc., San Diego, California, USA). The paired-end reads sequenced here (713,228 reads with an average length of 180 bp) were trimmed under the threshold with the probability of error < 0.05. The trimmed reads (712,760 reads, average length of 179.6 bp) were de novo assembled using CLC Genomics Workbench 6.5 software (CLC Bio, Aarhus, Denmark). The contigs used for assembling of the plastid genome ranged from 1360.85 to 479.20 of average coverage. Initial annotation of the genome was carried out using Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004) and BLAST searches. From this initial annotation, putative start and stop codons, and intron positions were determined based on comparisons to homologous genes in other plastid genomes. All tRNA genes were further verified by using tRNAscan-SE (Lowe and Chan. 2016). A physical map of the plastid circular genome was drawn using OrganellarGenomeDRAW (OGDRAW) version 1.3.1 (Greiner et al. 2019). The complete plastome sequence of *T. pentaphyllum* was deposited in the GenBank database under accession number MT210235.

Comparative analysis of genome structure

The Mauve Genome Alignment v2.3.1 (MAUVE) (Darling et al. 2004) was used for a comparison of the plastid genome structure between *T. pentaphyllum* and other representatives of Brassicales. Furthermore, the IR boundaries of the plastome of these species were also compared. The software REPuter (Kurtz et al. 2001) was used to determine the localization of the IR borders.

In all comparative analyses of this work, we used complete plastome sequences available in the plastid genome database of NCBI. Given the low number of complete plastomes available for species of Brassicales, except for the family Brassicaceae (only one species available for the families Moringaceae, Caricaceae, Akaniaceae, and Cleomaceae), our analyses included more species of family Brassicaceae than species from other families. Some families of Brassicales were not sampled due to the total absence of complete plastome sequences available in the database.

Phylogenomic inference

The inference of the phylogenetic position of *T. pentaphyllum* within Brassicales was carried out using a phylogenomic approach. Firstly, whole plastome sequences from species belonging to different families of the order Brassicales, available in the organelle database (Genbank), were selected and extracted. The plastome of *Bixa orellana* (order Malvales) was used as an outgroup species. The GenBank accession number of each taxon used here is shown in

Supplementary Table S1. The IRb was withdrawn to prevent the overrepresentation of IR sequences. The alignment of the plastome sequences was done using MAFFT v.7 (Kato and Standley 2013) and the best-fit evolutionary model (TVM + F + I + G4) was selected following the Bayesian information criterion (BIC) scores computed by ModelFinder (Kalyanamoorthy et al. 2017). A maximum likelihood (ML) estimation was posteriorly conducted using IQTREE v1.6.10 (Nguyen et al. 2015) and 500 non-parametric bootstrap replications were used to assess branch support. Lastly, the consensus tree was visualized using FigTree 1.4.4 (<https://tree.bio.ed.ac.uk/software/figtree/>).

Analysis of synonymous (dS) and nonsynonymous (dN) substitution rates of plastid genes

The pairwise dS and dN substitution rates between *T. pentaphyllum* and other Brassicales representatives (indicated in Supplementary Table S1) were estimated to the 78 protein-coding genes common to the plastome of these species. Firstly, the sequences of each gene were individually aligned by MUSCLE (Edgar 2004) implemented in MEGA 6.0 (Tamura et al. 2013), with pairwise deletion set to gaps/missing data treatment. To each alignment was calculated the dS and dN values using MEGA under the Kumar model (Kimura 2-parameter).

SSR detection, sliding window analysis, and RNA editing prediction

SSRs were detected in the *T. pentaphyllum* plastome using the MicroSatellite (MISA) Perl script (Thiel et al. 2003), with thresholds of eight repeat units for mononucleotide SSRs, four repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta-, and hexanucleotide SSRs.

The hotspots of sequence divergence between *T. pentaphyllum* and *Bretschneidera sinensis* were investigated by sliding window analysis. The whole plastome of these species was aligned using MAFFT v.7 and the sliding window analysis was performed using the DnaSP v.6 software (Rozas et al. 2017). The window length and the step size were set to 400 bp and 100 bp, respectively.

Potential RNA editing sites in protein-coding genes of *T. pentaphyllum* plastome were predicted by the program Predictive RNA Editor for Plants (PREP) suite (Mower 2009), which uses 35 reference genes for detecting RNA editing sites in plastid genomes. The cutoff value was set at 0.8 and the reference genes were *accD*, *atpA*, *atpB*, *atpF*, *atpI*, *ccsA*, *clpP*, *matK*, *ndhA*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *petB*, *petD*, *petG*, *petL*, *psaB*, *psaI*, *psbB*, *psbE*, *psbF*, *psbL*, *rpl2*, *rpl20*, *rpl23*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps2*, *rps8*, *rps14*, *rps16*, and *ycf3*. Aiming evolutionary comparison, RNA editing

sites were also predicted for other Brassicales species, using the same parameters.

Results

Size, organization, and gene content of *T. pentaphyllum* plastome

T. pentaphyllum plastome is a circular molecule of 152,954 bp in length and exhibits the quadripartite structure typical of most flowering plants. It includes a large single-copy region (LSC) of 81,881 bp flanked on each side by inverted repeats (IRs) of 26,597 bp with a small single-copy region (SSC) of 17,879 bp joining the IRs (Fig. 1). Comparatively, this plastome is more similar in structure (gene order) and total size to the *Alyssum desertorum* plastome. Comparatively, these two genomes are smaller than all other representatives of the order Brassicales (Table 1). The LSC region accounts for most size variation visualized among Brassicales plastomes, ranging from 81.5 kb in *A. desertorum* to 88.7 kb in *Carica papaya*. On the other hand, the IR and SSC regions show high similarity in length among the different species of this order. The overall GC content determined to *T. pentaphyllum* plastome is 37.33%, which is slightly larger in comparison with other plastomes of Brassicales (Table 1).

The *T. pentaphyllum* plastome contains 112 unique genes, of which 17 are completely duplicated and three (*rps12*, *rpl22* and *ycf1*) are partially duplicated in the IRs (Table 2). This set of 112 genes includes 78 unique protein-coding genes (six are completely and three are partially duplicated), 30 unique tRNA genes (seven duplicated) and four unique rRNA genes (all of them duplicated). Among the 112 genes, 15 harbor one intron (six tRNA genes and nine protein-coding genes) and two harbor two introns (*clpP* and *ycf3* genes). Additionally, the *infA* gene is absent from this plastome.

The structure of *T. pentaphyllum* plastome was compared with other species of the order Brassicales through multiple genome alignment produced by MAUVE software (Supplementary Fig. S1). It showed a perfect synteny between all species of Brassicales in the analysis (red block, Supplementary Fig. S1), indicating that no structural rearrangements are present in these plastomes. On the other hand, the IR boundaries were also compared between them and some differences were detected. The main one was an IR expansion in the LSC/IRa boundary of *T. pentaphyllum*, which included the complete *rps19* gene and a minor portion of the *rpl22* gene in the IRs of this species. In all other species of Brassicales sequenced so far, the entire *rpl22* is localized in the LSC and only a portion of *rps19* is located in the IRs (Fig. 2). Minor events of IR expansion/contraction can also be

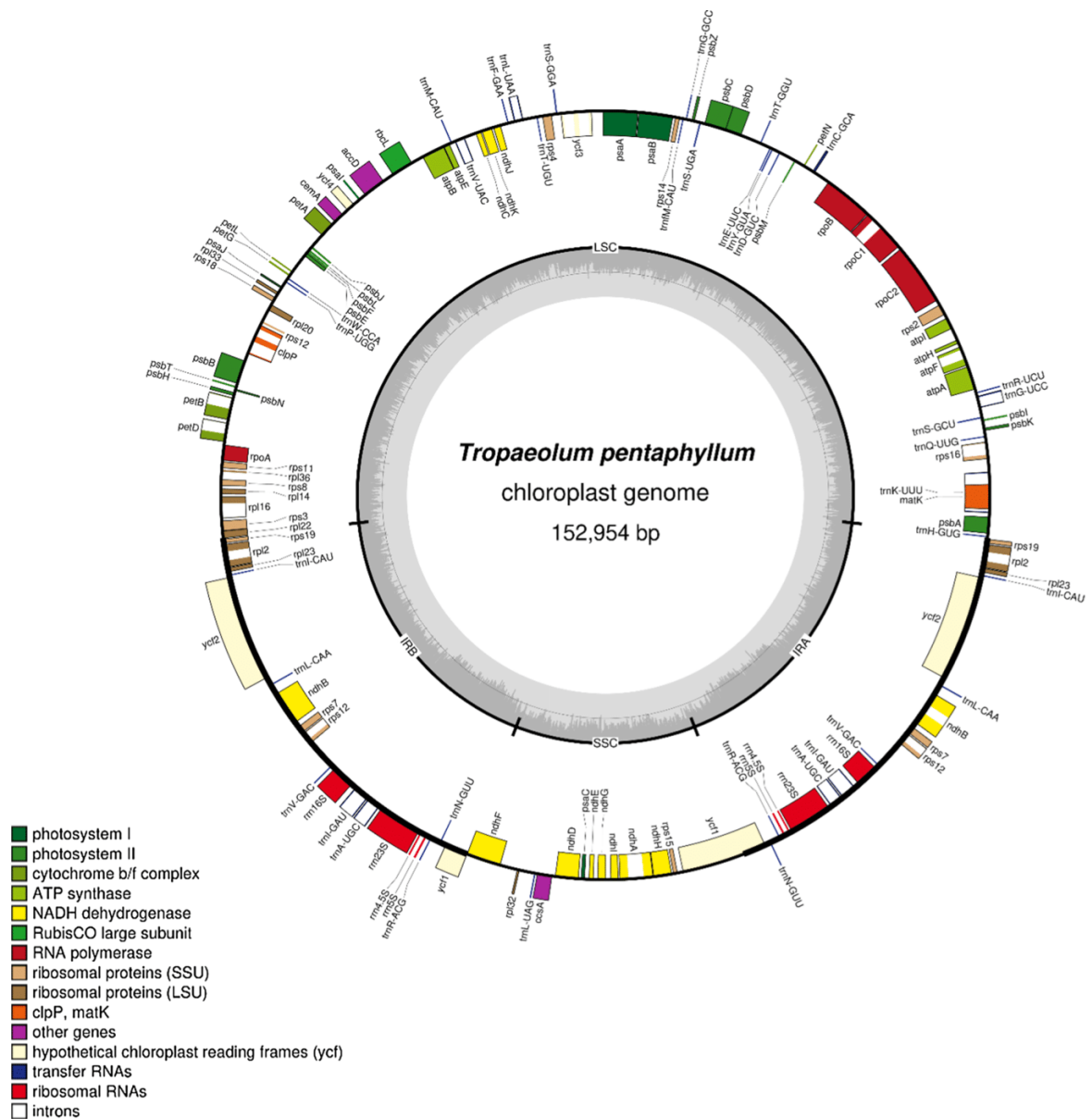


Fig. 1 Gene map and genome organization of *T. pentaphyllum* plastome. Two inverted repeat regions, IR_A and IR_B, divide the circular DNA molecule into large (LSC) and small (SSC) single copy regions. Genes drawn inside the circle are transcribed clockwise and genes

drawn outside are expressed counterclockwise. Genes belonging to different functional groups are color coded. The dark gray in the inner circle corresponds to GC content, while the light gray corresponds to AT content. Dotted circle corresponds to 50% of AT/GC content

observed in the IR_A/SSC boundary, since a small part of the *ndhF* gene is localized in the IRs in *T. pentaphyllum*, *Tarenaya hassleriana*, and *Crambe abyssinica*, whereas in all other species of the order Brassicales, this gene is completely found in the SSC region. Another small

variation corresponds to the portion of the *ycf1* gene that is localized in the IRs, which is larger in *T. pentaphyllum* (1496 bp) in comparison with the other species of Brassicales (ranging from 1171 bp in *Moringa oleifera* to 907 bp in *A. desertorum*).

Table 1 Comparison of plastome size and GC content among *T. pentaphyllum* and other species of Brassicales

Species	Family	Size (bp)	LSC (bp)	SSC (bp)	IR (bp)	GC (%)
<i>Tropaeolum pentaphyllum</i>	Tropaeolaceae	152,954	81,881	17,879	26,597	37.33
<i>Bretschneidera sinensis</i>	Akaniaceae	159,004	86,934	18,798	26,636	37.05
<i>Moringa oleifera</i>	Moringaceae	160,600	88,577	18,883	26,570	36.77
<i>Carica papaya</i>	Caricaceae	160,100	88,749	18,701	26,325	36.89
<i>Tarenaya hassleriana</i>	Cleomaceae	157,688	87,509	18,571	25,804	36.94
<i>Isatis tinctoria</i>	Brassicaceae	156,670	84,779	17,773	26,995	36.51
<i>Crambe abyssinica</i>	Brassicaceae	153,771	83,599	17,782	26,195	36.37
<i>Alyssum desertorum</i>	Brassicaceae	151,677	81,551	17,804	26,161	36.33

Table 2 List of genes identified in the plastome of *T. pentaphyllum*

Group of gene	Name of gene
Gene expression machinery	
Ribosomal RNA genes	<i>rrn16^b; rrn23^b; rrn5^b; rrn4.5^b</i>
Transfer RNA genes	<i>trnA</i> –UGC ^{ab} ; <i>trnC</i> –GCA; <i>trnD</i> –GUC; <i>trnE</i> –UUC; <i>trnF</i> –GAA; <i>trnI</i> –CAU; <i>trnG</i> –UCC ^a ; <i>trnG</i> –GCC; <i>trnH</i> –GUG; <i>trnI</i> –CAU ^b ; <i>trnI</i> –GAU ^{ab} ; <i>trnK</i> –UUU ^a ; <i>trnL</i> –CAA ^b ; <i>trnL</i> –UAA ^a ; <i>trnL</i> –UAG; <i>trnM</i> –CAU; <i>trnN</i> –GUU ^b ; <i>trnP</i> –UGG; <i>trnQ</i> –UUG; <i>trnR</i> –ACG ^b ; <i>trnR</i> –UCU; <i>trnS</i> –GCU; <i>trnS</i> –UGA; <i>trnS</i> –GGA; <i>trnT</i> –UGU; <i>trnT</i> –GGU; <i>trnV</i> –GAC ^b ; <i>trnV</i> –UAC ^a ; <i>trnW</i> –CCA; <i>trnY</i> –GUA
Small subunit of ribosome	<i>rps2; rps3; rps4; rps7^b; rps8; rps11; rps12^{bc}; rps14; rps15; rps16^a; rps18; rps19^b</i>
Large subunit of ribosome	<i>rpl2^{ab}; rpl14; rpl16^a; rpl20; rpl22^c; rpl23^b; rpl32; rpl33; rpl36</i>
DNA-dependent RNA polymerase	<i>rpoA; rpoB; rpoC1^a; rpoC2</i>
Genes for photosynthesis	
Subunits of photosystem I (PSI)	<i>psaA; psaB; psaC; psal; psaJ; ycf3^b; ycf4</i>
Subunits of photosystem II (PSII)	<i>psbA; psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM; psbN; psbT; psbZ</i>
Subunits of cytochrome b ₆ f complex	<i>petA; petB^a; petD^a; petG; petL; petN</i>
Subunits of ATP synthase	<i>atpA; atpB; atpE; atpF^a; atpH; atpI</i>
Subunits of NADH dehydrogenase	<i>ndhA^a; ndhB^{ab}; ndhC; ndhD; ndhE; ndhF; ndhG; ndhH; ndhI; ndhJ; ndhK</i>
Large subunit of Rubisco	<i>rbcL</i>
Others genes	
Maturase	<i>matK</i>
Envelope membrane protein	<i>cemA</i>
Subunit of acetyl-CoA carboxylase	<i>accD</i>
C-type cytochrome synthesis gene	<i>ccsA</i>
ATP-dependent Protease	<i>ClpP^a</i>
Component of TIC complex	<i>ycf1^c</i>
Component of 2-MD AAA-ATPase complex	<i>ycf2^b</i>
Absent genes	<i>infA</i>

^aGenes containing introns, ^bduplicated gene, ^cpartially duplicated genes

Phylogenomic reconstruction

A phylogenomic approach based on whole plastomes was carried out aiming to infer the position of *T. pentaphyllum* within the order Brassicales. This analysis included the families Tropaeolaceae (represented here by *T. pentaphyllum*), Akaniaceae (represented by *Bretschneidera sinensis*), Moringaceae (represented by *M. oleifera*), Caricaceae (represented by *C. papaya*), Cleomaceae (represented by

Tarenaya hassleriana), and Brassicaceae (represented here by 22 species belonging to different tribes) (Supplementary Table S1). As outgroup, we included a species belonging to the order Malvales, *Bixa orellana* (Bixaceae). Maximum likelihood (ML) analysis produced a consensus tree with the log-likelihood (lnL) value of –414,039.669. In this tree reconstruction, *T. pentaphyllum* (Tropaeolaceae) formed a sister group with *B. sinensis* (Akaniaceae). The clade composed of these two families was sister to the group formed

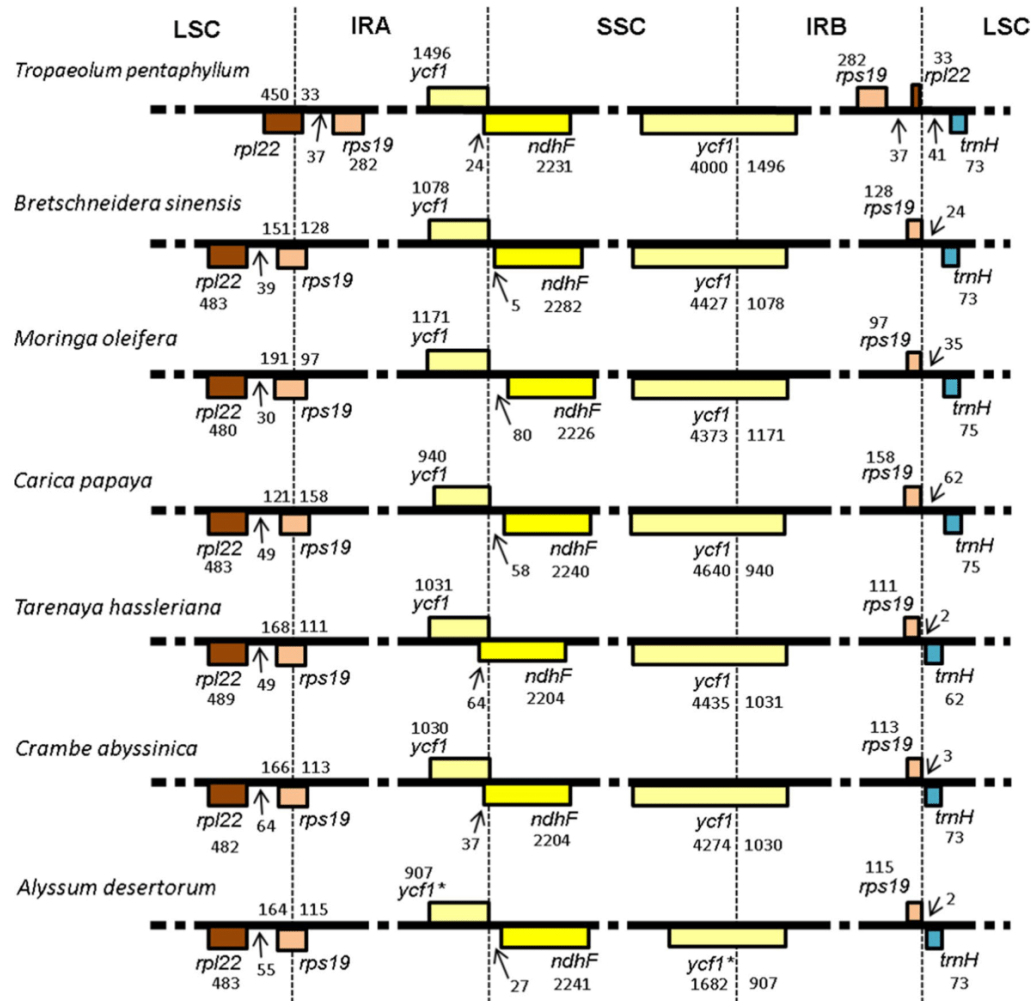


Fig. 2 Comparison of the IR-boundaries between *T. pentaphyllum* and representative species of Brassicales. Genes are represented by boxes. Genes above the line are transcribed from right to left and those below the line are transcribed from left to right. Genes belong-

ing to the same functional groups have the same color. The numbers indicate the lengths (in bp) of the genes or intergenic regions in the IR junctions. Genes and regions lengths are not to scale. *, putative pseudogene

by *C. papaya* (Caricaceae) and *M. oleifera* (Moringaceae). Consequently, the clade composed of these four families was sister to the clade formed by the families Cleomaceae (*T. hassleriana*) and Brassicaceae. The 22 species of the family Brassicaceae included in the tree formed a monophyletic group. All these relationships described above were highly supported, possessing non-parametric bootstrap (BS) values of 100% in our tree (Fig. 3).

Within the family Brassicaceae, all BS support values were 100% except for five nodes, which includes three with 98%, one with 60%, and the last one with 40%. This family is composed of two major clades: one including

only the tribe Aethionemeae and another clade including all other tribes. This second clade is divided into two groups: one including the tribes Alyssopsidae, Camelinae, Cardamineae, Crucihimalayae, Microlepidieae and Lepidieae, and another group including the remaining tribes. The last clade is still subdivided into two subclades: the first composed of the tribes Buniadeae, Anchonieae, Hesperideae and Euclidieae, and the second composed of the tribes Alysseae, Isatideae, Sisymbreae, Brassiceae, Thlaspidiae, Arabideae, Cochlearideae, Anastaticae, Megacarpaceae, and Biscutelleae (Fig. 3).

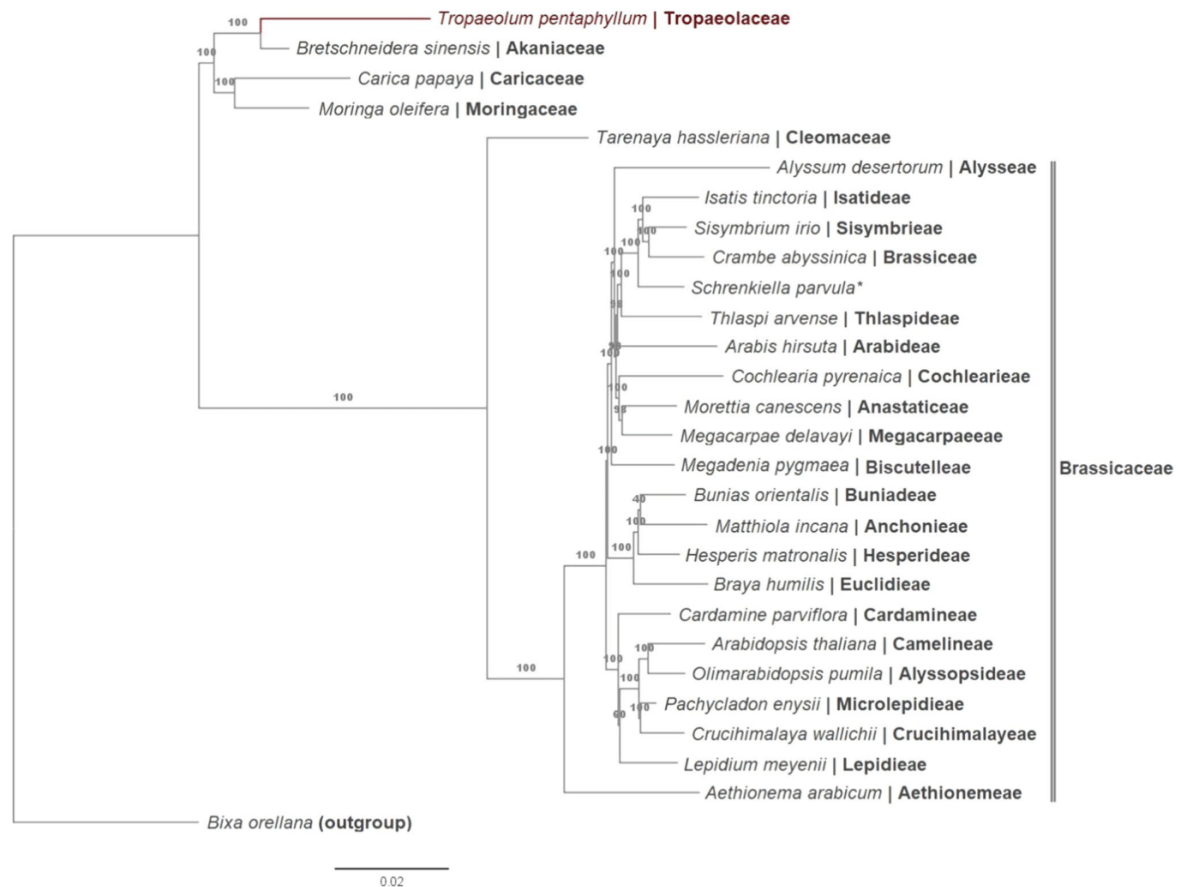


Fig. 3 Maximum likelihood (ML) phylogenomic analysis based on whole plastid genomes shows the position of *T. pentaphyllum* (in red) within the order Brassicales. The families are indicated to the right of the respective species. Moreover, the tribes of the family Brassicaceae are also indicated. Tree reconstruction was performed using

IQTREE software, with TVM+F+I+G4 model. Numbers (%) associated with branches are ML bootstrap support values. The branch length is proportional to the inferred divergence level. The scale bar indicates the number of inferred nucleic acid substitutions per site

Synonymous (dS) and nonsynonymous (dN) substitution rates

Aiming to analyze the molecular evolution of the plastid genes within Brassicales, synonymous (dS), nonsynonymous (dN), and dN/dS values were estimated for 78 protein-coding genes from *T. pentaphyllum* and representative species belonging to different families of Brassicales (species highlighted in Supplementary Table S1). Most genes showed low dN values (below of 0.10) in *T. pentaphyllum* (filled circle, first graphic, Fig. 4). Similarly, the values were very similar to the dN values observed for the other species (open circles; first graphic, Fig. 4). The *ycf1* and *matK* genes showed the highest dN values in comparison with all other genes in *T. pentaphyllum* (0.144 and 0.108, respectively) and the other species analyzed here (0.132 and 0.100, respectively).

Overall, the dS values (second graphic, Fig. 4) were also similar between *T. pentaphyllum* (ranging from 0.010 to 0.296) and the other species analyzed here (ranging from 0.011 to 0.258). The *psaC* gene showed the highest dS value in *T. pentaphyllum* (0.296) and the third higher value in the other species (0.239). On the other hand, the *rpl22* gene showed the third higher value to *T. pentaphyllum* (0.232) and the highest dS value in the other species analyzed here (0.258). Antagonistically, the *ndhB* gene showed the lowest value of dS in *T. pentaphyllum* (0.010) and the other species (0.010). Concerning the dN/dS ratio (third graphic, Fig. 4), all the plastid genes of *T. pentaphyllum* showed values lower than 1, indicating that these genes are under negative selection. The highest values of dN/dS for this species were 0.818 and 0.759 for the *ycf1* and *ycf2* genes, respectively. For the other species, only the *rpl23* gene showed a dN/dS ratio

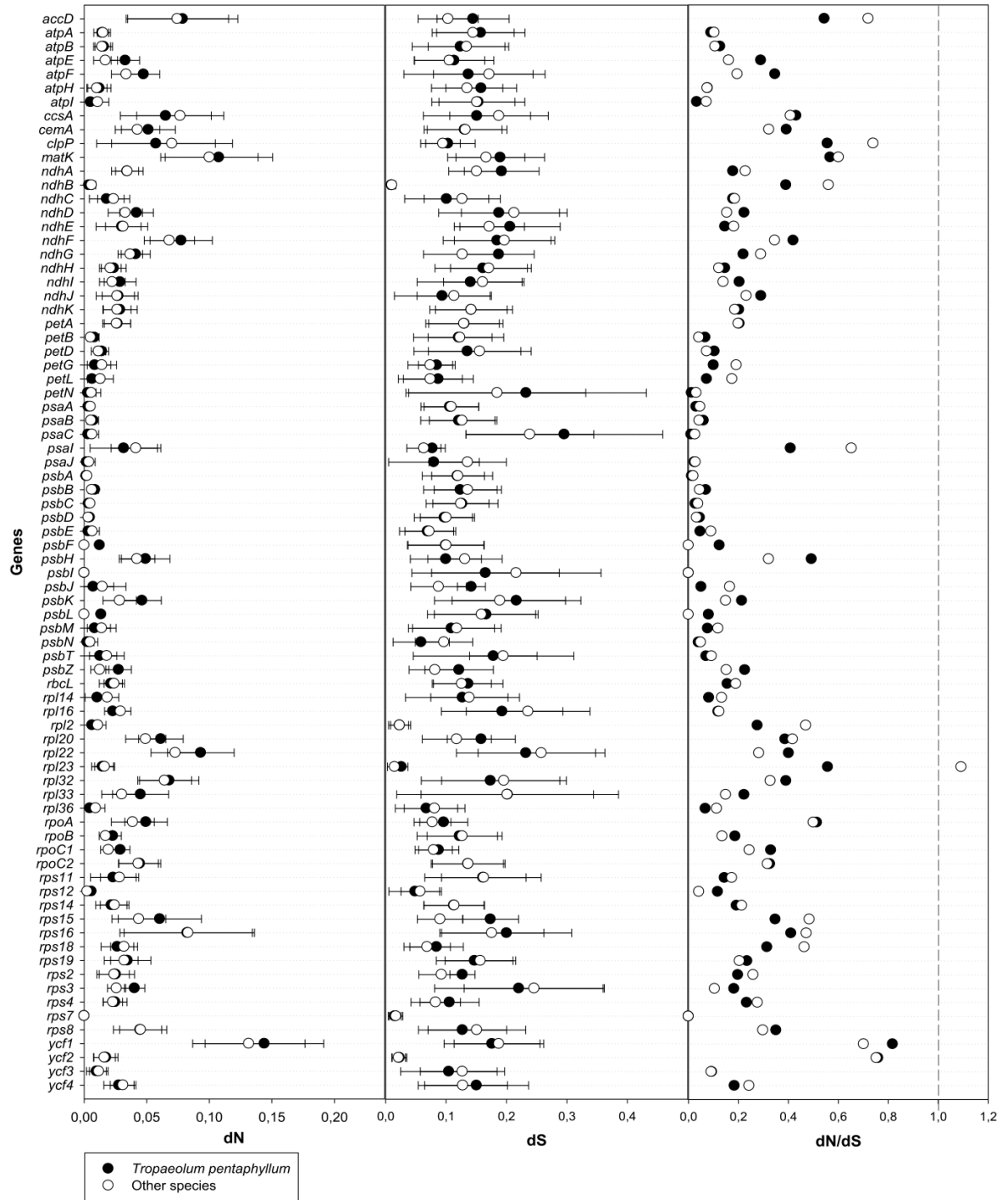


Fig. 4 Nonsynonymous (dN), synonymous (dS) and dN/dS values of 78 common plastid protein-coding genes. *T. pentaphyllum* is represented by filled circles (black) and the mean of other representatives of Brassicales (see Table S1) is represented by open circles (white)

above of 1 (1.093), which suggests that this gene may be under positive selection. The second gene with higher dN/dS value for the other species was the *ycf2* (0.752), which resembles the value observed in the plastome of *T. pentaphyllum* (0.759).

SSR detection and sliding window analysis

The occurrence, type, and distribution of SSRs in the *T. pentaphyllum* plastome were analyzed, revealing a total of 251 repeats. Mono- and dipolymers constitute most of the SSRs identified here, reaching 82.1% and 15.5%, respectively. The remaining 2.4% are constituted by tri- (2), tetra- (3) and hexapolymers (1) (Supplementary Table S2). Most SSRs (91.2%) detected here are composed of A and T bases. The size, sequence, and loci of the 251 SSRs are shown in Supplementary Table S3. From the total found in the plastome of *T. pentaphyllum*, 174 are located in the LSC region, 51 in the SSC, and 26 in the IRs. Additionally, 165 are present in intergenic spacers (IGS), 58 in CDS, and 28 in introns. The IGS with a higher number of SSRs correspond to the spacers between the genes *trnS-GCU* and *trnG-UCC* (13 SSRs), *petA* and *psbJ* (8), *ndhF* and *rpl32* (8), and *trnK-UUU* and *rps16* (7). The SSRs found in the CDS are distributed in 18 genes, of which the *ycf1* (21), *ycf2* (6), *rpoC2* (5), *ndhD* (5), and *ndhF* (4) contain the highest number. Among 28 SSRs found in the introns, most of them are located in the introns of *clpP* (7), *ycf3* (3), *trnL-UAA* (3), and *trnV-UAC*

(3) genes. Lastly, six SSRs (see c*, Supplementary Table S3) are found directly adjacent to each other in the LSC region.

Moreover, based on our sliding window analysis (Fig. 5), nine regions emerged as hotspots of nucleotide polymorphism between *T. pentaphyllum* and *B. sinensis* (species closely related in our phylogeny, Fig. 3). These hotspots are located in the LSC (4) and SSC (5) regions and include two CDS (*ycf1* and *ndhF*) and seven IGS (*trnK-UUU/rps16*, *trnS-GCU/trnG-UCC*, *atpH/atpI*, *petA/psbJ*, *ndhF/rpl32*, *rpl32/trnL-UAG* and *rps15/ycf1*). Several SSRs detected here in *T. pentaphyllum* (Supplementary Table S3) are located within these nine hotspots, with the regions of highest nucleotide diversity containing the highest number of SSRs (21 and 13 in the *ycf1* gene and the *trnS-GCU/trnG-UCC* IGS, respectively).

RNA editing sites prediction

The PREP-CP software identified 56 putative RNA editing sites in the plastome of *T. pentaphyllum*, which are distributed in 19 genes (Supplementary Table S4). All these RNA editing sites occur in the first or the second codon position and all nucleotide changes observed are from cytidine (C) to uridine (U). The genes *ndhB* (12), *ndhD* (8), *ndhA* (5), *matK* (5), *ndhF* (3), *rpoB* (3), and *rpoC2* (3) showed a high number of RNA editing sites. Other genes, such as *accD*, *atpF*, *clpP*, *ndhG*, *petB*, *psaI*, *psbE*, *rpl20*, *rpoA*, *rpoC1*, *rps14* and *rps16*, contain one or two putative RNA editing

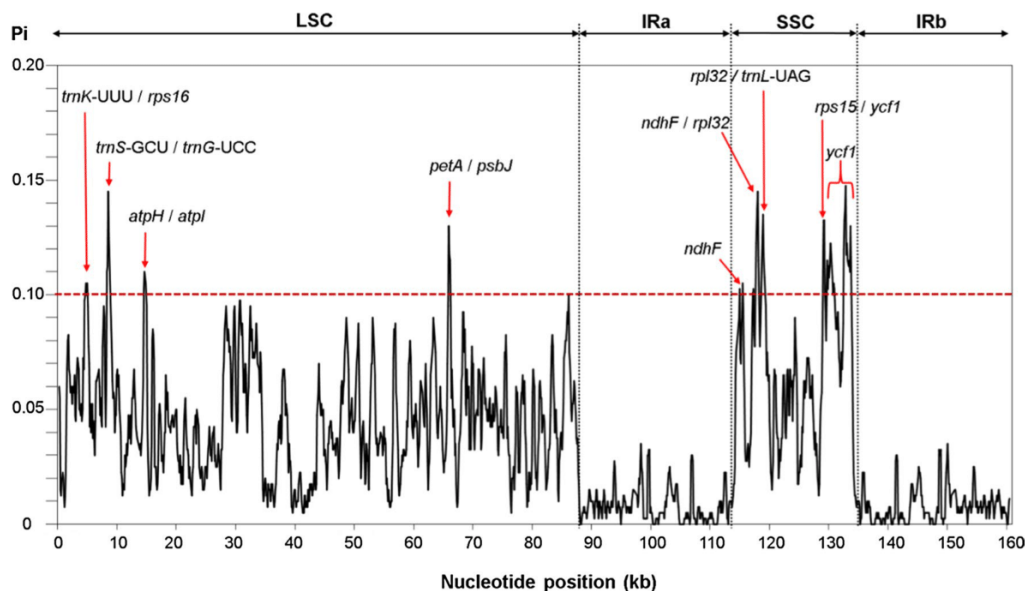


Fig. 5 Sliding window analysis of aligned whole plastomes of *B. sinensis* and *T. pentaphyllum*. The regions with higher nucleotide variability ($Pi > 0.1$; above the red line) are indicated by red arrow. The

intervals corresponding to LSC, IRs and SSC regions are also delimited in the upper x-axis. Pi , nucleotide diversity of each window. Window length, 400pb. Step size, 100pb

sites. Most of the RNA editing sites predicted here change the encoded amino acid from polar to apolar (32 out of 56), and the main change observed here was from serine to leucine (21 sites). Only one RNA editing site changes the amino acid from apolar to polar (proline to serine, amino acid position 72 of the *psbE* gene). Among the other 23 sites, 10 of them change the amino acid from polar positively charged to polar neutral (2, arginine to cysteine; 8, histidine to tyrosine) and 13 do not alter either polarity or charge (apolar to apolar).

From the total of 56 RNA editing sites predicted here to *T. pentaphyllum* plastome, half of them (28) are shared by the species of Brassicales investigated here (Supplementary Table S4). The other 28 sites are absent in one or more of those species, including two RNA editing sites present only in *T. pentaphyllum* (highlighted with + in Supplementary Table S4). One of them is localized in the amino acid position 191 of the *matK* gene, whereas in the DNA of the other species of Brassicales, this site contains a deoxythymidine (T) instead of a deoxycytidine (C) in the editing position, maintaining the conserved amino acid and dispensing the need for RNA editing. The other specific editing site of *T. pentaphyllum* is localized in the amino acid position 332 of the *rpoA* gene, which is inside of the C-terminus extension specifically observed in this species. Additionally, two RNA editing sites, which were predicted for all species of Brassicales analyzed here, are absent in *T. pentaphyllum* plastome (amino acid position 219 of *matK* and 26 of *psbF*; highlighted with * in Supplementary Table S4), which already has a T instead of a C fixed at these sites in the DNA sequence, dispensing the need for RNA editing.

Discussion

The *T. pentaphyllum* plastome is similar in structure and gene content to other Brassicales, but it shows some small variations at the IR-boundaries

The plastome structure and gene order are highly conserved between *T. pentaphyllum* and all other species belonging to the order Brassicales sequenced to date (Table 1 and Supplementary Fig. S1; Guo et al. 2017; Lopes et al. 2018a). The total size of *T. pentaphyllum* plastome is smaller than the other species of the order Brassicales, except for few species of the family Brassicaceae such as *Alyssum desertorum* (Li et al. 2017), and *Lobularia maritima* and *Solms-laubachia eurycarpa* (Guo et al. 2017). The LSC is the region of the plastome with the largest variation in size, while SSC and IR regions are very similar between the species of Brassicales sequenced to date. Concerning the IR borders, they are also overall conserved between these species. However, some events of IR expansion in *T. pentaphyllum* plastome were detected here (Fig. 2). All other Brassicales contain the

rps19 gene partially duplicated in the LSC/IRa boundary and the *rpl22* gene is located in the LSC region, which is also observed in the plastome of several other angiosperms (Goulding et al. 1996; Zhu et al. 2016). On the other hand, an event of IR expansion occurred in the plastome of *T. pentaphyllum* and duplicated the full sequence of *rps19* gene and part of the *rpl22* gene in the IRs. Although it is the first event reported in a species of the order Brassicales, similar events of IR expansion, at the LSC/IRa boundary, were identified in several species belonging to other lineages of angiosperms such as Fabales, Malpighiales, Poales, and Arecales (Zhu et al. 2016; Lopes et al. 2018b). Additionally, the SSC/IRb boundaries of all analyzed Brassicales share the same pattern of partial duplication of *ycf1* in which the minor part of this gene is localized in the IRs and the larger part is localized in the SSC. This pattern is also found in most angiosperms and the shifts of several hundreds of base pairs in this boundary are common to several lineages (Zhu et al. 2016). Here, it is possible to see that the portion of the *ycf1* gene of the *T. pentaphyllum* plastome included in the IRs is larger, about 300–400 bp, compared with other species of Brassicales, indicating that an additional small event of IR expansion occurred in this species.

Molecular evolution of plastid protein-coding genes in *T. pentaphyllum* and within Brassicales

The gene content annotated here to *T. pentaphyllum* is very similar to the other species of Brassicales (Guo et al. 2017; Lopes et al. 2018a; Yang et al. 2019). The only gene lost by the *T. pentaphyllum* plastome was the *infA*, which encodes the translation initiation factor 1 (Cummings and Hershey, 1994). The loss of this gene is a common feature observed in species belonging to Brassicales investigated so far as well as in species of the order Malvales (Pacheco et al. 2019) and several other rosids (Millen et al. 2001). Nuclear copies of the *infA* gene containing a plastid transit peptide were detected in *Arabidopsis* and other rosids. Similar functional transfer of this gene to the nucleus probably occurred in *T. pentaphyllum* and other Brassicales, compensating the loss of the plastid *infA* copy. Additionally, further plastid gene losses were only detected in some species of Brassicaceae family. Among them, the *rps16* gene was lost or degenerated in several genera of this family (Roy et al. 2010; Guo et al. 2017). This gene encodes a ribosomal protein that is essential for cell viability (Fleischmann et al. 2011). In *Populus alba* and *Medicago truncatula*, the losses of the plastid *rps16* copy were compensated by a nuclear-encoded copy of mitochondrial origin that can also target to the plastids (Ueda et al. 2008). It is likely that the species of the family Brassicaceae that lost this gene also possess a nuclear-encoded copy being imported by plastids. Additional plastid gene losses are unique to *Solms-laubachia*

eurycarpa plastome which correspond to losses of the genes of NAD(P)H dehydrogenase-like complex (*ndh* genes) (Guo et al. 2017). The loss of this great number of *ndh* genes is shared by other photosynthetic seed plants, being reported only in few lineages such as conifers (Pinaceae/Gnetales; Wu et al. 2009), *Erodium* genus (Geraniaceae; Blazier et al. 2011), Cactaceae (Sanderson et al. 2015) and Orchidaceae (Lin et al. 2017). If the *ndh* genes were lost or were functionally replaced in these lineages, it remains to be elucidated. Lastly, the *Alyssum desertorum* plastome seems to have lost the functional copy of *ycf1* gene. Despite this loss was not considered by Li et al. (2017), we detected here that the *ycf1* is likely a pseudogene in this species due to the existence of premature stop codons and partial deletion of its coding sequence. Consequently, the predicted *ycf1* gene in *A. desertorum* results in a much smaller ORF (2.59 kb in total, with 0.91 kb localized in the IRs and 1.68 kb in the SSC; Fig. 2) in comparison with other species of Brassicales and other angiosperms (about 5.4 kb; Fig. 2) (Wicke et al. 2011). This gene encodes the subunit Tic214 of the TIC complex, a translocon protein located in the inner membrane of plastids (Kikuchi et al. 2013), which is essential for cell viability (Drescher et al. 2000). However, the plastid *ycf1* gene was also lost or is a pseudogene in other distantly related taxa, such as Poaceae (Guisinger et al. 2010; Vries et al. 2015), Geraniaceae (Guisinger et al. 2011), Ericaceae (Fajardo et al. 2013) and Passifloraceae (Rabah et al. 2019). Thus, it is likely that a compensatory mechanism emerged (such as functional nuclear transfer) in *Alyssum* and other above-mentioned lineages allowing the loss of the plastid *ycf1* gene and absence of organelle/cell damage.

Concerning the gene divergence, the *ycf1* gene was the plastid protein-coding gene with the highest dN value in the plastomes of *T. pentaphyllum* and other Brassicales analyzed here (Fig. 4). This gene also showed the highest dN/dS value in the plastome of *T. pentaphyllum* and the fifth-highest dN/dS value in the plastomes of other species of Brassicales. However, these values were below 1, indicating that the *ycf1* is under negative selection in the plastome of the species of Brassicales, including *T. pentaphyllum*. This gene is among the most divergent genes within the tribe Brassiceae (Brassicaceae; Lopes et al. 2018a) and it is also described as one of the most divergent genes in plastomes of angiosperms (Wicke et al. 2011; Vries et al. 2015).

The second gene with the highest dN value analyzed here is the *matK*. This gene encodes a maturase (MatK) related to the self-splicing process of group IIA introns. It was demonstrated that MatK binding occurs in the introns of seven plastid genes: *trnV*-UAC, *trnI*-GAU, *trnA*-UGC, *trnK*-UUU, *rpl2*, *rps12*, and *atpF* (Zoschke et al. 2010). Recently, another study provided the first direct evidence of MatK splicing activity using an in vitro activity assay, which revealed that heterologous expression of MatK increases

efficiency of group IIA intron self-splicing for the second intron of the *rps12* gene (Barthet et al. 2020). Since MatK controls gene expression of several essential genes related to protein biosynthesis (Rogalski et al. 2006; Alkatib et al. 2012), its function is essential for cell viability. Despite its essential role, the *matK* is also known as a fast-evolving gene presenting high rates of nonsynonymous substitutions compared with other plastid genes in angiosperms (Hilu and Liang 1997; Hilu et al. 2014). A high divergence in the *matK* sequence was also detected in *Crambe abyssinica* and other species of the family Brassicaceae (Lopes et al. 2018a). The evolving features observed in this gene make it an excellent molecular marker to be used in phylogenetic inferences in the order Brassicales and other angiosperm lineages (Hilu et al. 2003, 2014; Worberg et al. 2009; Hall et al. 2011). Although it shows high dN values, its function is maintained in most lineages by restrictive amino acid substitutions and purifying selection (Young and de Pamphilis 2000; Barthet and Hilu 2008). Here, we also detected evidence that the *matK* gene is under purifying selection in the order Brassicales given that calculated dN/dS values were below 1 (about 0.6). Interestingly, a specific change in the *matK* sequence was only detected in *T. pentaphyllum*, which corresponds to an extension of 23 amino acids in the C-terminus of the protein, in comparison with other Brassicales (Supplementary Fig. S2). A similar extension in this region was also observed in conifers and Ginkgo-phyta (Hilu and Alice 1999), but it was not observed before in angiosperms. The extension found in the plastome of *T. pentaphyllum* could have originated from a mutation in a previous stop codon located in a position that is conserved in other species. Posteriorly, an alternative stop codon, located few amino acids downstream, was selected maintaining the integrity and functionality of this essential protein.

In parallel, the *rpoA* and *rpoC2* genes of *T. pentaphyllum* also showed some changes in their C-terminus in comparison with other Brassicales. These genes encode subunits of the plastid-encoded RNA polymerase (PEP), an *Escherichia coli*-like enzyme (Serino and Maliga 1998). The *rpoC2* in *T. pentaphyllum* bears a deletion of 42–54 amino acids in its C-terminal sequence if compared with other representatives of Brassicales (Supplementary Fig. S3, red box). However, it is unlikely that this deletion impairs the protein function since the *rpoC2* of *T. pentaphyllum* maintains all conserved domains (Supplementary Fig. S4). Moreover, the *rpoC2* showed a low value of dN/dS (0.33), indicating that this gene is under negative selection (Fig. 4). On the other hand, the *rpoA* of *T. pentaphyllum* showed a small extension of its C-terminus sequence, consisting of 24 nucleotides (Supplementary Fig. S5). Interestingly, this small extension includes a specific RNA editing site predicted only to *T. pentaphyllum*.

The *rpl23* gene was the exception of our dN/dS analysis. In the plastome of *T. pentaphyllum*, the dN/dS value of this gene was 0.55, indicating that the gene is under negative selection. However, the mean dN/dS value for other species of Brassicales was slightly above 1 (1.09), which suggests that this gene is under neutral or positive selection. This was the only gene in our analysis with a value equal or above 1. Despite the similar values of dN and dS (Fig. 4), it is unlikely that the *rpl23* is under neutral selection in the order Brassicales, once that the absence of positive and negative selections indicates the presence of a pseudogene. Accordingly, the translated amino acid residues of the *rpl23* are still highly conserved between the species of Brassicales analyzed here (Supplementary Fig. S6). Furthermore, this gene encodes a ribosomal protein (L23) that is essential for cell viability (Fleischmann et al. 2011). A functional copy of the *rpl23* is present in most plastomes of angiosperms and only few plant lineages lost this gene from plastome such as some species of Caryophyllales (Bubunenko et al. 1994; Logacheva et al. 2008; Sloan et al. 2012; Sanderson et al. 2015) and flax (Lopes et al. 2018c). However, the loss of this gene in those lineages is associated with some compensatory mechanism. In some species of Caryophyllales, a eukaryotic-type L23 protein, encoded by the nucleus is imported into the plastids (Bubunenko et al. 1994) filling the essential ribosomal function. A similar mechanism was also proposed for flax (Lopes et al. 2018c). Therefore, the dN/dS value close to 1 found here is likely a combination of positive selection acting in some few sites and purifying selection acting in the rest of the gene, maintaining the functionality of the protein. Simultaneously, this gene shows a very low rate of synonymous substitutions between these species (Fig. 4), and few nonsynonymous substitutions, which can result in the high dN/dS value identified in our analysis. Similar results were also reported for *B. orellana* belonging to the order Malvales (Pacheco et al. 2019).

Phylogenetic relationships based on whole plastomes in Brassicales

The plastid phylogenomic reconstructed here showed three well-supported (100% of BS) sister family pairs: Tropaeolaceae–Akaniaceae, Caricaceae–Moringaceae, and Cleomaceae–Brassicaceae. These pairs are also supported in several other phylogenetic studies of the order Brassicales (Ronse De Craene and Haston 2006; Edger et al. 2015; 2018; Cardinal-McTeague et al. 2016; Sun et al. 2016). Additionally, in our phylogenomic tree, the Tropaeolaceae–Akaniaceae pair was sister to Caricaceae–Moringaceae, constituting the early-diverging lineages of the order; and these four families formed a sister group to Cleomaceae–Brassicaceae (with 100% of BS). These results are congruent with the phylogenetic analysis based on transcriptomics of the order

Brassicales (Edger et al. 2015). Similarly, the congruency is also confirmed in trees reconstructed using morphological characters, plastid and nuclear sequences (Ronse De Craene and Haston 2006). However, other studies using plastid and mitochondrial sequences or 72 concatenated plastid-encoded genes revealed that Tropaeolaceae–Akaniaceae constituted the earliest-diverging lineage of the order Brassicales, being sister to the clade formed by all other families of this order (Cardinal-McTeague et al. 2016; Sun et al. 2016; Edger et al. 2018). Here, our phylogeny based on whole-plastomes included only the Brassicales families that have complete sequences of plastid genome available in the organelle database (Genbank). Therefore, the different topological reported here may be explained as being due to our different dataset and/or our small amount of sampling of taxa in comparison with other studies (Cardinal-McTeague et al. 2016; Sun et al. 2016; Edger et al. 2018).

Furthermore, the relationship presented here to the tribes within Brassicaceae is congruent with other recent studies (Huang et al. 2016; Guo et al. 2017; Lopes et al. 2018a). Our tree showed identical topology in comparison with that reconstructed by Lopes et al. (2018a) using a similar dataset (whole-plastomes), except for the use of only a species per tribe and the addition of two new tribes (Buniadeae and Sisymbrieae). Concerning these two tribes, Sisymbrieae was sister to Brassiceae, and Sisymbrieae–Brassicaceae formed a sister group to the tribe Isatideae. The closed relationship between Sisymbrieae and Brassiceae is also in accordance with Huang et al. (2016). Successively, the tribe Buniadeae was sister to Anchonieae (with low BS value, 40%) and these two tribes formed a sister group to the tribe Hesperideae (with 100% of BS support). Mandáková et al. (2017) also reported a close relationship between these three tribes (constituting part of the clade *Hesperis*), whereas in our analysis Hesperideae was sister to Anchonieae and these two tribes formed a sister group to Buniadeae.

Mapping of SSRs and hotspots of nucleotide polymorphism in *T. pentaphyllum* plastome

As far as we know, studies using plastid SSRs have not been yet realized in natural populations of *T. pentaphyllum*. Similarly, the family Tropaeolaceae is also very poorly studied in several aspects concerning population genetics and genetic diversity. Thus, the 251 SSRs detected here in the *T. pentaphyllum* plastome represent a new and rich source of genetic data available to be utilized in pioneer studies in this species such as genetic diversity, population structure, and conservation studies. Most SSRs detected here are monopolymers composed of A or T and occur mainly in intergenic spacers following other reports (Wheeler et al. 2014; George et al. 2015; Lopes et al. 2018a).

Additionally, we also identified nine hotspots of nucleotide polymorphism in the plastome of *T. pentaphyllum* (Fig. 5). Interestingly, we detected a great prevalence of SSRs in these regions, mainly in the two highly divergent regions (i.e., the *ycf1* gene and the spacer between *trnS* and *trnG* genes). These results follow the high level of polymorphism expected for SSR sequences. A similar pattern of SSR distribution was also reported in species of the family Brassicaceae (Lopes et al. 2018a). Finally, the divergent plastome regions identified here represent attractive genetic information to be used in phylogenetic and genetic studies of *T. pentaphyllum* natural population. However, these sequences should be assessed and investigated concerning their potential as polymorphic markers in other species of *Tropaeolum* and family Tropaeolaceae. Several molecular markers detected here are likely shared by other species of Tropaeolaceae. They represent the start point for a reasonable characterization of effective markers to endangered species such as *T. pentaphyllum*.

RNA editing sites are conserved within Brassicales, but two sites were detected only in the plastome of *T. pentaphyllum*

Plastid RNA editing is a post-transcriptional modification (C-to-U and U-to-C conversions) identified in the transcripts of all land plants, except in some lineages of Marchantiales (Takenaka et al. 2013). This RNA maturation step can create start and stop codons or restore conserved amino acids. Thereby, RNA editing can act as an indirect repair mechanism that corrects DNA mutations on the RNA level (Takenaka et al. 2013; Ichinose and Sugita 2017). RNA editing conversions from U to C are absent in angiosperms, occurring only in some plant lineages such as ferns, lycopods, and hornworts (Takenaka et al. 2013; Ichinose and Sugita 2017).

We detected 56 putative RNA editing sites in the plastome of *T. pentaphyllum*. All these sites correspond to C-to-U conversions, occurring in the first or the second positions of the codons as observed in most angiosperms (Takenaka et al. 2013). Most RNA editing sites predicted (32 out of 56) change of the encoded amino acid from polar to apolar, increasing the protein hydrophobicity. This tendency was also reported by other studies, which suggests that the increased hydrophobicity of the proteins could affect their structural features and transmembrane domain specificities (He et al. 2016; Lopes et al. 2018b).

Exactly half of the RNA editing sites predicted here (28 out of 56) are shared by all representatives of Brassicales investigated here (Supplementary Table S4). Concerning the shared sites, 22 were experimentally validated in *Arabidopsis thaliana* (Tillich et al. 2005), which include the sites as follows: *atpF* (amino acid position 31), *ndhB* (50, 156, 196, 204, 246, 249, 277, 279, 419, and 494), *ndhD* (1, 137,

225, 293, and 296), *ndhF* (97), *psbE* (72), *rpoB* (112, 189, and 811) and *rps14* (27). Other RNA editing sites predicted here for *T. pentaphyllum*, but not shared by all Brassicales analyzed here, were also validated in *A. thaliana*, including the site in the *accD* (279), *clpP* (196), and *rps14* (50) genes. Thus, our results in combination with the RNA editing sites already validated in *A. thaliana* demonstrate that the plastid RNA editing sites are highly conserved in Brassicales.

Two RNA editing sites predicted here were only found in the plastome of *T. pentaphyllum*, *matK* (amino acid 191) and *rpoA* (amino acid 332). As previously mentioned here, the *rpoA* gene of *T. pentaphyllum* has a specific C-terminus extension, which included a specific RNA editing site (highlighted in red, Supplementary Fig. S5). This specific site induces the conversion from proline (CCG) to leucine (CUG), which does not change the amino acid polarity. The other specific RNA editing site found in the *T. pentaphyllum* plastome occurs in the *matK* gene converting a histidine (CAU) to a tyrosine (UAU). In all other Brassicales analyzed here, a tyrosine codon is already found in the DNA sequence, dispensing the need for RNA editing. Hence, this new editing site of *T. pentaphyllum* acts probably restoring a conserved tyrosine in the *matK* sequence. However, it is first necessary to experimentally confirm both sites and their occurrence in this species, genus *Tropaeolum*, and consequently, family Tropaeolaceae.

Moreover, we detected two specific losses of RNA editing sites in *T. pentaphyllum*, in comparison with all other species of Brassicales analyzed here. One of them is localized in the *matK* (219), which is predicted to change a histidine (CAU) to a tyrosine (UAU) in the other Brassicales (Supplementary Table S4). The loss of this RNA editing site in the genus *Tropaeolum* was also reported by Tillich et al. (2009). These authors also suggested that this loss occurs in the genus *Batis*, belonging to the family Bataceae (Brassicales). Then, the loss of this RNA editing site occurred putatively at least two times independently during the evolution of Brassicales (i.e. Tropaeolaceae and Bataceae). Tillich et al. (2009) suggested a closed relationship between the rapid evolution of the *matK* gene and the degeneration of its cis-elements leading to a reduction in the RNA editing efficiency, which consequently creates a selective pressure in favor to the mutation C-to-T, resulting in the loss of the RNA editing site. This hypothesis is congruent with our results given that the *matK* gene showed two unshared RNA editing sites in *T. pentaphyllum* (including one loss and one putative gain) when compared with all other Brassicales investigated here (Supplementary Table S4). In addition to that feature, it is also occurring in a faster-evolving gene in our analysis (Fig. 4).

Lastly, the other RNA editing site lost in *T. pentaphyllum* is the *psbF* (26) (Supplementary Table S4). In all other Brassicales investigated here, a serine codon (UUU) in this

position is predicted to be converted to a phenylalanine one (UCU). This RNA editing site was firstly reported to species belonging to genera *Pisum*, *Oenothera*, and *Spinacia* (Bock et al. 1993). However, in *T. pentaphyllum* and some other plant lineages such as Solanaceae and Poaceae (Bock et al. 1993), a phenylalanine codon is already present in DNA sequence in the same position of the coding sequence of *psbF* gene, justifying the absence of the referred RNA editing site.

Conclusions

Here, we report the complete plastome sequence of *T. pentaphyllum*, which represents the first plastid genome of the family Tropaeolaceae to be fully sequenced and characterized in detail. This plastome shows a high similarity to the plastome of other species of Brassicales concerning gene content and genome structure. The specific features of the plastome of *T. pentaphyllum* can be defined by events of IR expansion, full duplication of the *rps19* gene, extension of the C-terminus of *matK* and *rpoA* genes, and partial deletion of the C-terminus of the *rpoC2* gene. Specifically, two RNA editing sites seem to be unique to *T. pentaphyllum* plastome, and occur in two genes, *matK* and *rpoA*. Moreover, we detected 251 SSRs and nine regions as hotspots of nucleotide polymorphism in the plastome of *T. pentaphyllum*. Furthermore, our plastid phylogenomic analysis indicated, with high support, a closed relationship between the families Tropaeolaceae and Akaniaceae, which is congruent with several other phylogenetic studies. These two families formed a sister group to the pair Caricaceae–Moringaceae, and these four families formed a sister group to the pair Cleomaceae–Brassicaceae. However, in several other phylogenetic studies of Brassicales, the pair Tropaeolaceae–Akaniaceae formed the earliest-diverging lineage, being sister to all other families of Brassicales. We speculate that the incongruence observed here is likely related to the low number of families included in our phylogenomic analysis given that complete plastome sequences of several other families are not available in the plastid database. Since the great majority of studies carried out in the order Brassicales are concentrated on the family Brassicaceae, the present work contributes to a better understanding of the plastome evolution within Brassicales as well as the phylogenetic relationships among the different families. Finally, this study serves as a basic tool to understand the genetic diversity and genetic divergence of natural populations of *T. pentaphyllum*, and to enable strategies for genotype characterization, domestication, and conservation of this useful and endangered species.

Author contribution statement TGP, ASL, GMS, EMS, FOP, and MR conceived and designed the research. TGP, GMS,

ASL, JDO, EB, EMS, and MR conducted experiments and analyzed the data. EMS, FOP, and MR contributed with reagents and materials. TGP, ASL, JMR, and MR wrote the manuscript. All authors read and approved the manuscript.

Acknowledgements This research was supported by the National Council for Scientific and Technological Development, Brazil (CNPq—Grants 459698/2014-1, 310654/2018-1 and 436407/2018-3). We are grateful to INCT-FBN and for the scholarships granted by the CNPq to ASL, TGP, JMR, JDO, EB, FOP, EMS, and MR. We are also grateful to the Núcleo de Análise de Biomoléculas (NuBiomol) of the Universidade Federal de Viçosa for providing the software CLC Genomics.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S, Schöttler MA, Ruf S, Bock R (2012) The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet* 8:e1003076. <https://doi.org/10.1371/journal.pgen.1003076>
- Andersson L, Andersson S (2000) A molecular phylogeny of tropaeolaceae and its systematic implications. *Taxon* 49:721–736. <https://doi.org/10.2307/1223973>
- Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L, Medina J, Pires JC, Santos C, Stevenson DW, Zomlefer WB, Davis JI (2016) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol* 209:855–870. <https://doi.org/10.1111/nph.13617>
- Barthel MM, Hilu KW (2008) Evaluating evolutionary constraint on the rapidly evolving gene *matK* using protein composition. *J Mol Evol* 66:85–97. <https://doi.org/10.1007/s00239-007-9060-6>
- Barthel MM, Pierpont CL, Tavernier EK (2020) Unraveling the role of the enigmatic MatK maturase in chloroplast group IIA intron excision. *Plant Direct* 4:e00208. <https://doi.org/10.1002/pld3.208>
- Bayer C, Appel O (2003) Tropaeolaceae. In: Kubitzki K, Bayer C (eds) The families and genera of vascular plants, flowering plants, dicotyledons: Malvales, Capparales and Non-betalain Caryophyllales, 1rd. Springer, Berlin, pp 400–404
- Blazier JC, Guisinger MM, Jansen RK (2011) Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Mol Biol* 76:263–272. <https://doi.org/10.1007/s11103-011-9753-5>
- Bock R (2017) Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. *Annu Rev Genet* 51:1–22. <https://doi.org/10.1146/annurev-genet-120215-035329>
- Bock R, Hagemann R, Kössel H, Kudla J (1993) Tissue- and stage-specific modulation of RNA editing of the *psbF* and *psbL* transcript from spinach plastids—a new regulatory mechanism? *Mol Gen Genet* 240:238–244. <https://doi.org/10.1007/BF00277062>
- Braga VB, Vieira MM, Barros IBI (2018) Nutritional potential of leaves and tubers of crem (*Tropaeolum pentaphyllum* Lam.). *Rev Nutr* 31:423–432. <https://doi.org/10.1590/1678-9865201800400007>

- Bubunenko MG, Schmidt J, Subramanian AR (1994) Protein substitution in chloroplast ribosome evolution. A eukaryotic cytosolic protein has replaced its organelle homologue (L23) in spinach. *J Mol Biol* 240:28–41. <https://doi.org/10.1006/jmbi.1994.1415>
- Cardinal-McTeague WM, Sytsma KJ, Hall JC (2016) Biogeography and diversification of Brassicales: a 103 million year tale. *Mol Phylogenet Evol* 99:204–224. <https://doi.org/10.1016/j.ympev.2016.02.021>
- Correa MP (1984) Dicionário das plantas úteis do Brasil e das exóticas cultivadas. Imprensa Nacional, Rio de Janeiro, pp 669–674
- Couvreur TLP, Franzke A, Al-Shehbaz IA, Bakker FT, Koch MA, Mummehoff K (2010) Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol Biol Evol* 27:55–71. <https://doi.org/10.1093/molbev/msp202>
- Cronquist A (1988) The evolution and classification of flowering plants. Bronx, New York
- Cummings HS, Hershey JW (1994) Translation initiation factor IF1 is essential for cell viability in *Escherichia coli*. *J Bacteriol* 176:198–205. <https://doi.org/10.1128/jb.176.1.198-205.1994>
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>
- De Bona GS, Boschetti W, Bortolin RC, Vale M, Moreira J, de Rios AO, Flôres SH (2017) Characterization of dietary constituents and antioxidant capacity of *Tropaeolum pentaphyllum* Lam. *J Food Sci Technol* 54:3587–3597. <https://doi.org/10.1007/s13197-017-2817-z>
- Drescher A, Ruf S, Calsa T, Carrer H, Bock R (2000) The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J* 22:97–104. <https://doi.org/10.1046/j.1365-3113x.2000.00722.x>
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, Hofberger JA, Smithson A, Hall JC, Blanchette M, Bureau TE, Wright SI, dePamphilis CW, Eric Schranz M, Barker MS, Conant GC, Wahlberg N, Vogel H, Pires JC, Wheat CW (2015) The butterfly plant arms-race escalated by gene and genome duplications. *Proc Natl Acad Sci USA* 112:8362–8366. <https://doi.org/10.1073/pnas.1503926112>
- Edger PP, Hall JC, Harkess A, Tang M, Coombs J, Mohammadin S, Schranz ME, Xiong Z, Leebens-Mack J, Meyers BC, Sytsma KJ, Koch MA, Al-Shehbaz IA, Pires JC (2018) Brassicales phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *Am J Bot* 105:463–469. <https://doi.org/10.1002/ajb2.1040>
- Fabri LT, Valla JJ (1998) Aspectos de la biología reproductiva de *Tropaeolum pentaphyllum* (Tropaeolaceae). *Darwiniana* 36:51–58. <https://doi.org/10.14522/darwiniana.2014.361-4.320>
- Fajardo D, Senalik D, Ames M, Zhu H, Steffan SA, Harbut R, Polashock J, Vorsa N, Gillespie E, Kron K, Zalapa JE (2013) Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. *Tree Genet Genomes* 9:489–498. <https://doi.org/10.1007/s11295-012-0573-9>
- Fleischmann TT, Scharff LB, Alkatib S, Hasdorf S, Schottler MA, Bock R (2011) Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell* 23:3137–3155. <https://doi.org/10.1105/tpc.111.088906>
- George B, Bhatt BS, Awasthi M, George B, Singh AK (2015) Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr Genet* 61:665–677. <https://doi.org/10.1007/s00294-015-0495-9>
- Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet* 252:195–206. <https://doi.org/10.1007/BF02173220>
- Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* 47:59–64. <https://doi.org/10.1093/nar/gkz238>
- Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK (2010) Implications of the Plastid Genome Sequence of *Typha* (Typhaceae, Poales) for Understanding Genome Evolution in Poaceae. *J Mol Evol* 70:149–166. <https://doi.org/10.1007/s00239-009-9317-3>
- Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family geraniaceae: rearrangements, repeats, and codon usage. *Mol Biol Evol* 28:583–600. <https://doi.org/10.1093/molbev/msq229>
- Guo X, Liu J, Hao G, Zhang L, Mao K, Wang X, Zhang D, Ma T, Hu Q, Al-Shehbaz IA, Koch MA (2017) Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics* 18:176. <https://doi.org/10.1186/s12864-017-3555-3>
- Hall JC, Iltis HH, Sytsma KJ (2004) Molecular phylogenetics of core brassicales, placement of orphan genera *Emblingia*, *Forchhammeria*, *Tirania*, and character evolution. *Syst Bot* 29:654–669. <https://doi.org/10.1600/0363644041744491>
- Hall JC, Tisdale TE, Donohue K, Wheeler A, Al-Yahya MA, Kramer EM (2011) Convergent evolution of a complex fruit structure in the tribe Brassiceae (Brassicaceae). *Am J Bot* 98:1989–2003. <https://doi.org/10.3732/ajb.1100203>
- He P, Huang S, Xiao G, Zhang Y, Yu J (2016) Abundant RNA editing sites of chloroplast protein-coding genes in *Ginkgo biloba* and an evolutionary pattern analysis. *BMC Plant Biol* 16:257. <https://doi.org/10.1186/s12870-016-0944-8>
- Hilu KW, Alice LA (1999) Evolutionary implications of *matK* indels in Poaceae. *Am J Bot* 86:1735–1741. <https://doi.org/10.2307/2656671>
- Hilu KW, Liang H (1997) The *matK* gene: sequence variation and application in plant systematics. *Am J Bot* 80:830. <https://doi.org/10.2307/2445819>
- Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW (2003) Angiosperm phylogeny based on *matK* sequence information. *Am J Bot* 90:1758–1776. <https://doi.org/10.3732/ajb.90.12.1758>
- Hilu KW, Black CM, Oza D (2014) Impact of gene molecular evolution on phylogenetic reconstruction: a case study in the rosids (Superorder Rosanae, Angiosperms). *PLoS ONE* 9:e99725. <https://doi.org/10.1371/journal.pone.0099725>
- Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP, Pires JC, Tan D-Y, Zhong Y, MA H (2016) Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol Biol Evol* 33:394–412. <https://doi.org/10.1093/molbev/msv226>
- Ichinose M, Sugita M (2017) RNA editing and its molecular mechanism in plant organelles. *Genes* 8:5. <https://doi.org/10.3390/genes8010005>
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>

- Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M (2013) Uncovering the protein translocator at the chloroplast inner envelope membrane. *Science* 339:571–574. <https://doi.org/10.1126/science.1229262>
- Kinupp VF, Lorenzi H (2014) Plantas Alimentícias Não-Convencionais (PANC) no Brasil: guia de identificação, aspectos nutricionais e receitas ilustradas. Plantarum, Nova Odessa, p 768
- Kinupp VF, Lisbôa GN, Barros IBI (2011) *Tropaeolum pentaphyllum*, Batata-crem. In: Coradin L, Siminski LC, Reis A (eds) Espécies nativas da flora brasileira de valor econômico atual ou potencial: plantas para o futuro—Região Sul. Ministério do Meio Ambiente, Brasília, pp 243–250
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 29:4633–4642. <https://doi.org/10.1186/gb-2004-5-2-r12>
- Li G, Quiros CF (2001) Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. *Theor Appl Genet* 103:455–461. <https://doi.org/10.1007/s001220100570>
- Li Y, LV G, He X, Zhang X, Yang X (2017) The complete chloroplast genome of the spring ephemeral plant *Alyssum desertorum* and its implications for the phylogenetic position of the tribe Alyseae within the Brassicaceae. *Nord J Bot* 35:644–652. <https://doi.org/10.1111/njb.01531>
- Lin C-S, Chen JJW, Chiu C-C, Hsiao HCW, Yang C-J, Jin X-H, Leebens-Mack J, de Pamphilis CW, Huang Y-T, Yang L-H, Chang W-J, Kui L, Wong GK-S, Hu J-M, Wang W, Shih M-C (2017) Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J* 90:994–1006. <https://doi.org/10.1111/tpj.13525>
- Logacheva MD, Samigullin TH, Dhingra A, Penin AA (2008) Comparative chloroplast genomics and phylogenetics of *Fagopyrum esculentum* ssp. *Ancestrale*—a wild ancestor of cultivated buckwheat. *BMC Plant Biol* 8:59. <https://doi.org/10.1186/1471-2229-8-59>
- Lopes AS, Pacheco TG, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Rogalski M (2018a) The *Crambe abyssinica* plastome: Brassicaceae phylogenomic analysis, evolution of RNA editing sites, hotspot and microsatellite characterization of the tribe Brassicaceae. *Gene* 671:36–49. <https://doi.org/10.1016/j.gene.2018.05.088>
- Lopes AS, Pacheco TG, Nimz T, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Rogalski M (2018b) The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae. *Planta* 247:1011–1030. <https://doi.org/10.1007/s00425-018-2841-x>
- Lopes AS, Pacheco TG, Santos KG, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Rogalski M (2018c) The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Rep* 37:307–328. <https://doi.org/10.1007/s00299-017-2231-z>
- Lopes AS, Pacheco TG, Silva ON, Magalhães-Cruz L, Balsanelli E, Maltempo de Souza E, de Oliveira PF, Rogalski M (2019) The plastomes of *Astrocaryum aculeatum* G. Mey. and *A. murumuru* Mart. show a flip-flop recombination between two short inverted repeats. *Planta* 250:1229–1246. <https://doi.org/10.1007/s00425-019-03217-z>
- Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:54–57. <https://doi.org/10.1093/nar/gkw413>
- Malice M, Baudoin JP (2009) Genetic diversity and germplasm conservation of three minor Andean tuber crop species. *Biotechnol Agron Soc Environ* 13:441–448
- Malice M, Bizoux JP, Blas R, Baudoin JP (2010) Genetic diversity of andean tuber crop species in the in situ microcenter of Huanuco, Peru. *Crop Sci* 50:1915–1923. <https://doi.org/10.2135/cropsci2009.09.0476>
- Mandáková T, Hloušková P, German DA, Lysak MA (2017) Monophyletic origin and evolution of the largest crucifer genomes. *Plant Physiol* 174:2062–2071. <https://doi.org/10.1104/pp.17.00457>
- Martin GE, Rousseau-Gueutin M, Cordonnier S, Lima O, Michon-Coudouel S, Naquin D, de Carvalho JF, Ainouche M, Salmon A, Ainouche A (2014) The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. *Ann Bot* 113:1197–1210. <https://doi.org/10.1093/aob/mcu050>
- Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13:645–658. <https://doi.org/10.1105/tpc.13.3.645>
- Mors WB, Rizzini CT, Pereira N (2000) A medicinal plants of Brazil. Reference Publications, Michigan, p 501. [https://doi.org/10.1016/S0378-8741\(01\)00203-3](https://doi.org/10.1016/S0378-8741(01)00203-3)
- Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res* 37:253–259. <https://doi.org/10.1093/nar/gkp337>
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
- Olson ME (2002a) Combining data from DNA sequences and morphology for a phylogeny of Moringaceae (Brassicales). *Syst Bot* 27:55–73
- Olson ME (2002b) Intergeneric relationships within the Caricaceae-Moringaceae clade (Brassicales) and potential morphological synapomorphies of the clade and its families. *Int J Plant Sci* 163:51–65. <https://doi.org/10.1086/324046>
- Ortega O, Duran E, Arbizu C, Ortega R, Roca W, Potter D, Quiros CF (2007) Pattern of genetic diversity of cultivated and non-cultivated mashua, *Tropaeolum tuberosum*, in the Cusco region of Perú. *Genet Resour Crop Evol* 54:807–821. <https://doi.org/10.1007/s10722-006-9160-y>
- Pacheco TG, Lopes AS, Viana GD, Silva ON, Silva GM, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Otoni WC, Rogalski M (2019) Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (*Bixa orellana* L.), the Amazonian commercial species of natural dyes. *Planta* 249:563–582. <https://doi.org/10.1007/s00425-018-3023-6>
- Pacheco TG, Lopes AS, Welter JF, Yotoko KSC, Otoni WC, Vieira LN, Guerra MP, Nodari RO, Balsanelli E, Pedrosa FO, Souza EM, Rogalski M (2020) Plastome sequences of the subgenus *Passiflora* reveal highly divergent genes and specific evolutionary features. *Plant Mol Biol*. <https://doi.org/10.1007/s11103-020-01020-z>
- Powell W, Morgantet M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV, Rafalski JA (1995) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr Biol* 5:1023–1029. [https://doi.org/10.1016/S0960-9822\(95\)00206-5](https://doi.org/10.1016/S0960-9822(95)00206-5)
- Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol* 16:142–147. [https://doi.org/10.1016/S0169-5347\(00\)02097-8](https://doi.org/10.1016/S0169-5347(00)02097-8)

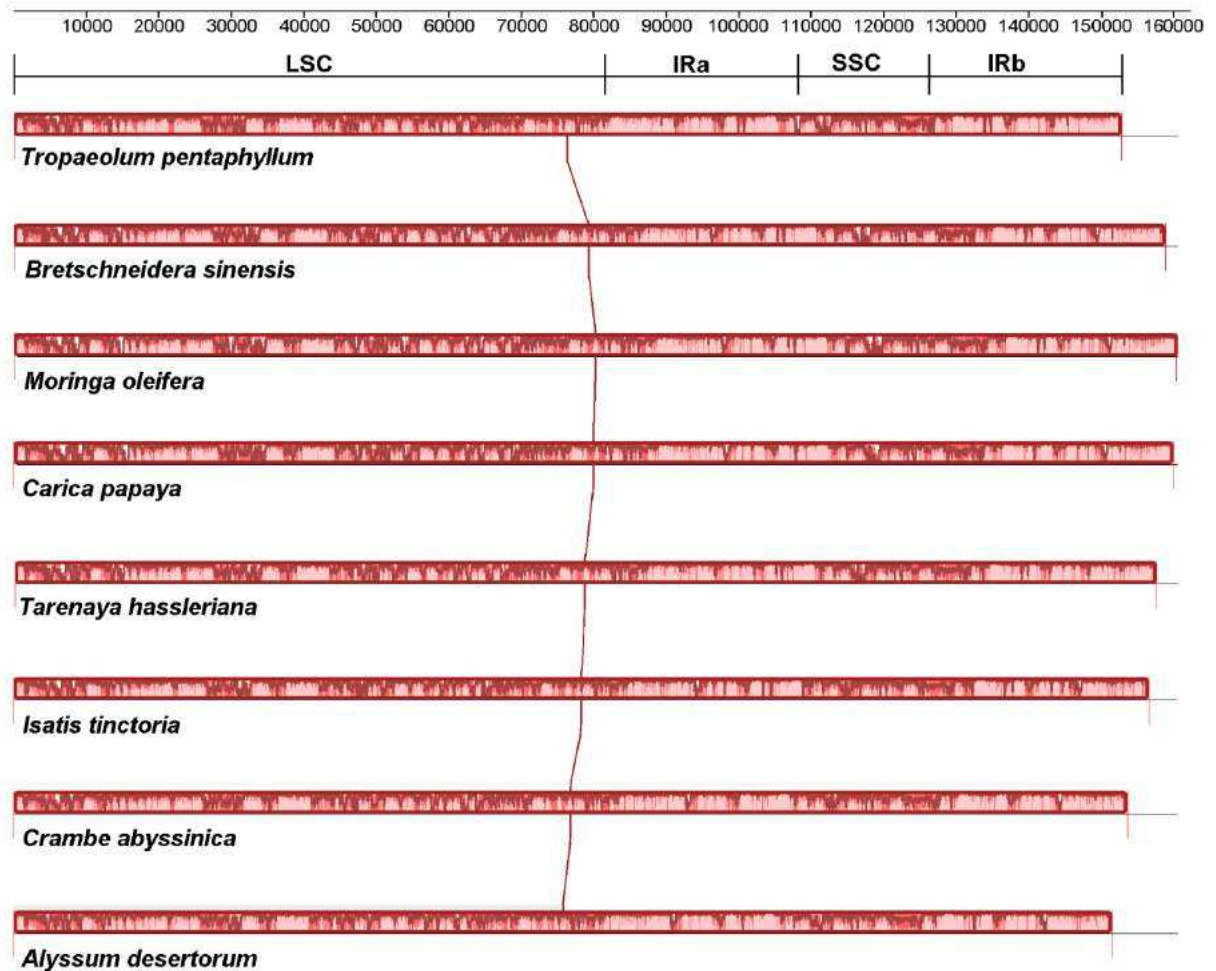
- Rabah SO, Shrestha B, Hajrah NH, Sabir Mumdooh J, Alharby HF, Sabir Mernan J, Alhebshi AM, Sabir JSM, Gilbert LE, Ruhlman TA, Jansen RK (2019) *Passiflora* plastome sequencing reveals widespread genomic rearrangements: *Passiflora* plastome evolution. *J Syst Evol* 57:1–14. <https://doi.org/10.1111/jse.12425>
- Rix M (2010) *Tropaeolum*, botanical illustration, descriptions, phyto-geography, plant cultural practices, plant taxonomy, pollination, South America. *Curtis's Bot Mag* 27:296–300. <https://doi.org/10.1111/j.1467-8748.2010.01706.x>
- Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Res* 34:4537–4545. <https://doi.org/10.1093/nar/gkl634>
- Rogalski M, Schottler MA, Thiele W, Schulze WX, Bock R (2008) Rpl33, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. *Plant Cell* 20:2221–2237. <https://doi.org/10.1105/tpc.108.060392>
- Rogalski M, Vieira LN, Fraga HP, Guerra MP (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front Plant Sci* 6:586. <https://doi.org/10.3389/fpls.2015.00586>
- Ronse De Craene LP, Haston E (2006) The systematic relationships of glucosinolate-producing plants and related families: a clastic investigation based on morphological and molecular characters. *Bot J Linn Soc* 151:453–494. <https://doi.org/10.1111/j.1095-8339.2006.00580.x>
- Roy S, Ueda M, Kadowaki K, Tsutsumi N (2010) Different status of the gene for ribosomal protein S16 in the chloroplast genome during evolution of the genus *Arabidopsis* and closely related species. *Genes Genet Syst* 85:319–326. <https://doi.org/10.1266/ggs.85.319>
- Roy PS, Rao GJN, Jena S, Samal R, Patnaik A, Patnaik SSC, Jambhulkar NN, Sharma S, Mohapatra T (2016) Nuclear and chloroplast DNA variation provides insights into population structure and multiple origin of native aromatic rices of Odisha, India. *PLoS ONE* 11:e0162268. <https://doi.org/10.1371/journal.pone.0162268>
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Sanderson MJ, Copetti D, Búrquez A, Bustamante E, Charboneau JLM, Eguiarte LE, Kumar S, Lee HO, Lee J, McMahon M, Steele K, Wing R, Yang T-J, Zwicli D, Wojciechowski MF (2015) Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the *ndh* gene suite and inverted repeat. *Am J Bot* 102:115–1127. <https://doi.org/10.3732/ajb.1500184>
- Serino G, Maliga P (1998) RNA polymerase subunits encoded by the plastid *rpo* genes are not shared with the nucleus-encoded plastid enzyme. *Plant Physiol* 117:1165–1170. <https://doi.org/10.1104/pp.117.4.1165>
- Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR (2012) Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biol Evol* 4:294–306. <https://doi.org/10.1093/gbe/evs006>
- Souza VC, Lorenzi H (2008) *Botânica Sistemática: Guia ilustrado para identificação das famílias de Fanerógamas nativas do Brasil*. Instituto Plantarum, Nova Odessa
- Sparre B (1972) Tropeoláceas. In: Reitz R (ed) *Flora ilustrada catarinense*. Herbário Barbosa Rodrigues (HBR), Itajaí, p 26p
- Sparre B, Andersson L (1991) A taxonomic revision of the Tropaeolaceae. *Opera Bot* 108:5–139
- Stefenon VM, Klabunde G, Lemos RPM, Rogalski M, Nodari RO (2019) Phylogeography of plastid DNA sequences suggests postglacial southward demographic expansion and the existence of several glacial refugia for *Araucaria angustifolia*. *Sci Rep* 9:2752. <https://doi.org/10.1038/s41598-019-39308-w>
- Sun M, Naeem R, Su J-X, Cao Z-Y, Burleigh JG, Soltis PS, Soltis DE, Chen Z-D (2016) Phylogeny of the Rosidae: a dense taxon sampling analysis. *J Syst Evol* 54:363–391. <https://doi.org/10.1111/jse.12211>
- Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A (2013) RNA editing in plants and its evolution. *Annu Rev Genet* 47:335–352. <https://doi.org/10.1146/annurev-genet-111212-133519>
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422. <https://doi.org/10.1007/s00122-002-1031-0>
- Tillich M, Funk HT, Schmitz-Linneweber C, Poltnigg P, Sabater B, Martin M, Maier RM (2005) Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J* 43:708–715. <https://doi.org/10.1111/j.1365-313X.2005.02484.x>
- Tillich M, Le Sy V, Schulerowitz K, von Haeseler A, Maier UG, Schmitz-Linneweber C (2009) Loss of *matK* RNA editing in seed plant chloroplasts. *BMC Evol Biol* 9:201. <https://doi.org/10.1186/1471-2148-9-201>
- Tolomeotti KRB, Felippi M, Donazzolo J (2018) Cytogenetic characterization of *Tropaeolum pentaphyllum* Lam. *Crop Breed Appl Biotechnol* 18:65–71. <https://doi.org/10.1590/1984-70332018v18n1a9>
- Trojan-Rodrigues M, Alves TLS, Soares GLC, Ritter MR (2012) Plants used as antidiabetics in popular medicine in Rio Grande do Sul, southern Brazil. *J Ethnopharmacol* 139:155–163. <https://doi.org/10.1016/j.jep.2011.10.034>
- Ueda M, Nishikawa T, Fujimoto M, Takanashi H, Arimura S, Tsutsumi N, Kadowaki K (2008) Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. *Mol Biol Evol* 25:1566–1575. <https://doi.org/10.1093/molbev/msn102>
- Vieira LN, Faoro H, Fraga HPF, Rogalski M, Souza EM, Pedrosa FB, Nodari RO, Guerra MP (2014) An improved protocol for intact chloroplasts and cpDNA isolation in Conifers. *PLoS ONE* 9:e84792. <https://doi.org/10.1371/journal.pone.0084792>
- Vieira LN, Rogalski M, Faoro H, Fraga HP, Anjos KG, Picchi GFA, Nodari RO, Pedrosa FO, Souza EM, Guerra MP (2016) The plastome sequence of the endemic Amazonian conifer, *Retrophillum piresii* (Silba) C.N.Page, reveals different recombination events and plastome isoforms. *Tree Genet Genomes* 12:10. <https://doi.org/10.1007/s11295-016-0968-0>
- Vries J, Sousa FL, Bölter B, Soll J, Gould SB (2015) YCF1: a green TIC? *Plant Cell* 27:1827–1833. <https://doi.org/10.1105/tpc.114.135541>
- Wambulwa MC, Meegahakumbura MK, Kamunya S, Muchugi A, Möller M, Liu J, Xu JC, Ranjitkar S, Li DZ, Gao LM (2016) Insights into the genetic relationships and breeding patterns of the African tea germplasm based on nSSR markers and cpDNA sequences. *Front Plant Sci* 7:1244. <https://doi.org/10.3389/fpls.2016.01244>
- Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE (2014) A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Appl Plant Sci* 2:1400059. <https://doi.org/10.3732/apps.1400059>
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* 76:273–297. <https://doi.org/10.1007/s11103-011-9762-4>

- Worberg A, Alford MH, Quandt D, Borsch T (2009) Huerteales sister to Brassicales plus Malvales, and newly circumscribed to include *Dipentodon*, *Gerrardina*, *Huerteia*, *Perrottetia*, and *Tapiscia*. *Taxon* 58:468–478. <https://doi.org/10.1002/tax.582012>
- Wu C-S, Lai Y-T, Lin C-P, Wang Y-N, Chaw S-M (2009) Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. *Mol Phylogenet Evol* 52:115–124. <https://doi.org/10.1016/j.ympev.2008.12.026>
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255. <https://doi.org/10.1093/bioinformatics/bth352>
- Yang Y, Tian Y, He S-L (2019) Characterization of the complete chloroplast genome of *Moringa oleifera* Lam. (Moringaceae), an important edible species in India. *Mitochondrial DNA* 4:1913–1915. <https://doi.org/10.1080/23802359.2019.1611393>
- Young ND, de Pamphilis CW (2000) Purifying selection detected in the plastid gene *matK* and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. *Mol Biol Evol* 17:1933–1941. <https://doi.org/10.1093/oxfordjournals.molbev.a026295>
- Zhu A, Guo W, Gupta S, Fan W, Mower JP (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol* 209:1747–1756. <https://doi.org/10.1111/nph.13743>
- Zoschke R, Nakamura M, Liere K, Sugiura M, Börner T, Schmitz-Linneweber C (2010) An organellar maturase associates with multiple group II introns. *Proc Natl Acad Sci USA* 107:3245–3250. <https://doi.org/10.1073/pnas.0909400107>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Material

Supplementary Figures



Supplementary Fig. S1 Comparison of plastome structures between *T. pentaphyllum* and representative species of Brassicales, produced by MAUVE. This software identified one collinear block corresponding to the whole plastome of these species (red block), indicating that these genomes shared the same gene order. Number in the upper x-axis are nucleotide coordinates of the plastome maps (in base pairs, bp). The positions corresponding to the LSC, IRs and SSC regions of the plastome of *T. pentaphyllum* are also indicated.

```

*           420           *           440           *           460           *
T. pentaphyllum : DSSDSDIIDRFVLRICRNLSHYHSGSS-TKKNLYQIKYILRLSCVKT LARKHKSTVRELFK--RVASDFEL :
B. sinensis      : DSSDSDIIDRFVLRICRNLSHYHSGSSSKKNLYRIKYLRLSCVKT LARKHKSTVRELFK--RLGSDFEL :
C. papaya       : DSSADSDIIDRFVLRICRNLSHYHSGSSSKKNLYRIKYLRLSCVKT LARKHKSTVRELFK--RVGSDFEL :
M. oleifera     : DSSDSDIIDRFVLRICRNLSHYHSGSSSKKNLYRIKYLRLSCVKT LARKHKSTVRELFK--RLGSDFEL :
C. abyssinica  : DSSDSDIIDRFVLRICRNLSHYHSGSSSKKNLYRIKYLRLSCVKT LARKHKSTVRELFK--RLGSDFEL :
T. hassleriana : DSSDSDIIDRFVLRICRNLSHYHSGSSSKKNLYRIKYLRLSCVKT LARKHKSTVRELFK--RVASDFEL :

480           *           500           *           520           *           540
T. pentaphyllum : EEFLTEEEQVLSLIFPRAYSASPRLYRGRWYLDIICNDLVIRE CNWKFFLNDEEIRKKCIHFLYYY : 530
B. sinensis      : EEFLTEEEQVLSLIFPRAYSALRRLYRGRWYLDIICNDLVNHE----- : 506
C. papaya       : EEFLTEEEQVLSLIFPRAYSASRKLYRGRWYLDIICNDLVNHE----- : 506
M. oleifera     : EEFLTEEEQVLSLIFPRAYSASRKLYRGRWYLDIICNDLVNHE----- : 506
C. abyssinica  : EEFLTEEEQVLSLIFPRSDYASKRLYRGRWYLDIICNDLVNHE----- : 502
T. hassleriana : EEFLTEEEQVLSLIFPRGYSACQRLYRGRWYLDIICNDLVNHE----- : 508

```

Supplementary Fig. S2 ClustalW alignment of the partial amino acid sequence of MatK (C-terminal), showing the specific extension (highlighted by the red box) present in *T. pentaphyllum* in comparison with other Brassicales. Black shadowing indicates conserved sequence among all taxa in the alignment, dark gray indicates conserved sequence among all except for one taxon, light gray indicates conserved sequence among four taxa and white indicates nonconserved amino acid. The numbers above the line indicate the amino acid position in the alignment. The number at the end of the line indicates the total length of MatK in each species.

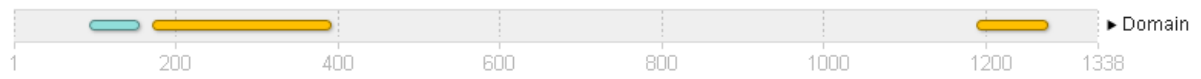
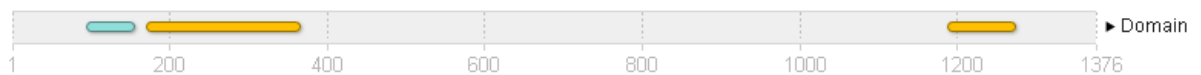
```

*           1280           *           1300           *           1320           *
T. pentaphyllum : RTGRALDEAICYRAILLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMVPAG
B. sinensis      : RTGRALDEAICYRAILLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMI PAG
C. papaya       : RTGRALDEAICYRAILLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMI PAG
M. oleifera     : RTGRALDEAICYRAILLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENI VLGGMIPAG
C. abyssinica  : RTGRALDEAICYRAVLLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGVI PAG
T. hassleriana : RTGRALDEAICYRAILLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGVI PAG

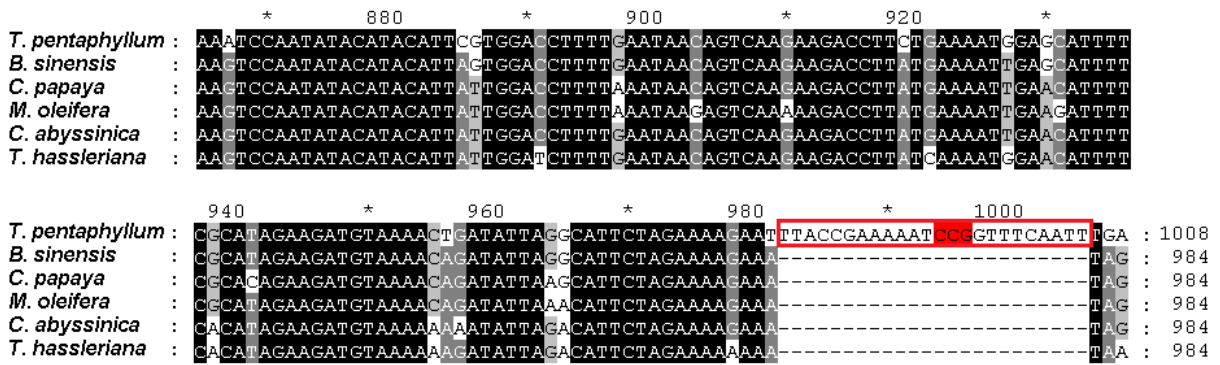
1340           *           1360           *           1380           *
T. pentaphyllum : TGE-KGLVHC SRQ--HTT ----- : 1338
B. sinensis      : TGE-KGLVHC SRQHKTITLLEKKKRNLYLFEGETRDI FEHHRELFDSCTISKNLHNTSERSFIGFNDS : 1392
C. papaya       : TGE-KGLVHC SRQ--HTSITLLETKKRNLYLFEGETRDI FEHHRELFDSCTISKNLHNTSERSFIGLNDS : 1387
M. oleifera     : TGE-KGLVHC SRP---SITLLETKKRNLYLFEGETRDI FEHHRELFDSCTISKNLHNTSERSFIGFNDS : 1388
C. abyssinica  : TGENKGLVHC SRK--HTNITLLEKKTKNLSLLECDMRDIEFHHRELFCDSSILKSAFSKIE----- : 1380
T. hassleriana : TGE-KGLVHC SRQ--HTNITLFEKKRNLSLFECDMRDIEFHHRELFDSCTISKNLHNTSERSFIGFNDS : 1387

```

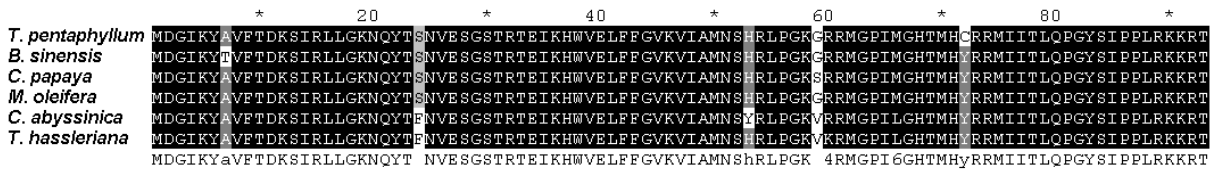
Supplementary Fig. S3 ClustalW alignment of the partial (C-terminal) amino acid sequence of RpoC2, showing the specific deletion (highlighted by the red box) present in *T. pentaphyllum* in comparison with other Brassicales. Black shadowing indicates conserved sequence among all taxa in the alignment, dark gray indicates conserved sequence among all except for one taxon, light gray indicates conserved sequence among four taxa and white indicates nonconserved amino acid. The numbers above the line indicate the amino acid position in the alignment. The number at the end of the line indicates the total length of RpoC2 in each species.

- *Tropaeolum pentaphyllum***- *Bretschneidera sinensis*****- *Carica papaya*****- *Arabidopsis thaliana***

Supplementary Fig. S4 Comparison of the domains of RpoC2 amino acid sequence using the online platform Interpro (Mitchell et al. 2019). Despite the variation in the total size, all the RpoC2 proteins, including that from *T. pentaphyllum*, shared the same domains (blue and yellow).



Supplementary Fig. S5 ClustalW alignment of the C-terminus of rpoA gene, showing the specific extension (highlighted by the red box) present in *T. pentaphyllum* in comparison with other Brassicales. Black shadowing indicates conserved sequence among all taxa in the alignment, dark gray indicates conserved sequence among all except for one taxon, light gray indicates conserved sequence among four taxa and white indicates nonconserved nucleotides. Moreover, red background indicates the RNA editing from proline (CCG) to leucine (CUG) in *T. pentaphyllum*. The numbers above the line indicate the nucleotide position in the alignment. The number at the end of the line indicates the total length of rpoA gene in each species.



Supplementary Fig. S6 ClustalW alignment of the translated amino acid sequence of rpl23 gene between representatives of the order Brassicales, including *T. pentaphyllum*. Black shadowing indicates conserved sequence among all taxa in the alignment, dark gray indicates conserved sequence among all except for one taxon, light gray indicates conserved sequence among four taxa and white indicates nonconserved amino acid. The numbers above the line indicate the amino acid position in the alignment.

Supplementary Tables

Supplementary Table S1 List of species included in the Brassicales phylogenomic.

Species	Tribe	Family	GenBank
<i>Tropaeolum pentaphyllum</i> Lam. ⁺	-	Tropaeolaceae	MT_210235
<i>Bretschneidera sinensis</i> Hemsl. ⁺	-	Akaniaceae	NC_037753.1
<i>Moringa oleifera</i> Lam. ⁺	-	Moringaceae	NC_041432.1
<i>Carica papaya</i> L. ⁺	-	Caricaceae	NC_010323.1
<i>Tarenaya hassleriana</i> (Chodat) Iltis. ⁺	-	Cleomaceae	NC_034364.1
<i>Aethionema arabicum</i> (L.) A.DC.	Aethionemeae	Brassicaceae	NC_034367.1
<i>Alyssum desertorum</i> Stapf	Alysseae	Brassicaceae	NC_034299.1
<i>Olimarabidopsis pumila</i> (Stephan) Al-Shehbaz et al.	Alyssopsidae	Brassicaceae	NC_009267.1
<i>Morettia canescens</i> Boiss.	Anastaticae	Brassicaceae	NC_035514.1
<i>Matthiola incana</i> (L.) R.Br.	Anchonieae	Brassicaceae	NC_034358.1
<i>Arabis hirsuta</i> (L.) Scop.	Arabideae	Brassicaceae	NC_009268.1
<i>Megadenia pygmaea</i> Maxim.	Biscutelleae	Brassicaceae	NC_034357.1
<i>Crambe abyssinica</i> Hochst. ex R.E.Fr. ⁺	Brassicaceae	Brassicaceae	KY883663
<i>Bunias orientalis</i> L.	Buniadeae	Brassicaceae	NC_036111.1
<i>Arabidopsis thaliana</i> (L.) Heynh.	Camelineae	Brassicaceae	NC_000932.1
<i>Cardamine parviflora</i> L.	Cardamineae	Brassicaceae	NC_036964.1
<i>Cochlearia pyrenaica</i> DC.	Cochlearieae	Brassicaceae	NC_029331.1
<i>Crucihimalaya wallichii</i> (Hook.f. & Thomson) Al-Shehbaz et al.	Crucihimalayae	Brassicaceae	NC_009271.1
<i>Braya humilis</i> (C.A.Mey.) B.L.Rob.	Euclidieae	Brassicaceae	NC_035515.1
<i>Hesperis matronalis</i> L.	Hesperideae	Brassicaceae	NC_035511.1
<i>Isatis tinctoria</i> L.	Isatideae	Brassicaceae	NC_028415.1
<i>Lepidium meyenii</i> Walp.	Lepidieae	Brassicaceae	NC_034363.1
<i>Megacarpaea delavayi</i> Franch.	Megacarpaeae	Brassicaceae	NC_034360.1
<i>Pachycladon enysii</i> (Cheeseman ex Kirk) Heenan & A.D.Mitch.	Microlepidieae	Brassicaceae	NC_018565.1
<i>Thlaspi arvense</i> L.	Thlaspidiae	Brassicaceae	NC_034362.1
<i>Sisymbrium irio</i> L.	Sisymbrieae	Brassicaceae	NC_037838.1
<i>Schrenkiella parvula</i> (Schrenk) D.A.German & Al-Shehbaz	Unassigned	Brassicaceae	NC_028726.1
<i>Bixa orellana</i> L. (outgroup)	-	Bixaceae*	NC_041550.1

* Order Malvales; ⁺species used in the dN/dS analysis.

Supplementary Table S2 List of simple sequence repeats (SSRs) identified in the plastome of *T. pentaphyllum*.

SSR sequence	Number of repetitions												Total
	3	4	5	6	7	8	9	10	11	12	13	14	
A/T	-	-	-	-	-	95	41	28	18	5	5	4	196
C/G	-	-	-	-	-	7	2	1	-	-	-	-	10
AC/GT	-	2	-	-	-	-	-	-	-	-	-	-	2
AG/CT	-	7	2	-	-	-	-	-	-	-	-	-	9
AT/AT	-	17	8	3	-	-	-	-	-	-	-	-	28
AAT/ATT	-	-	2	-	-	-	-	-	-	-	-	-	2
AAAT/ATTT	3	-	-	-	-	-	-	-	-	-	-	-	3
AGGGAT/ATCCCT	1	-	-	-	-	-	-	-	-	-	-	-	1
Total													251

Supplementary Table S3 Distribution of SSR loci in the plastome of *T. pentaphyllum*.

SSR type	SSR	Size	Start	End	Location
----------	-----	------	-------	-----	----------

mono	(A)10	10	142	151	trnH-GUG/psbA (IGS)
mono	(T)8	8	1497	1504	psbA/trnK-UUU (IGS)
mono	(A)8	8	1523	1530	psbA/trnK-UUU (IGS)
mono	(A)8	8	3279	3286	matK (CDS)
mono	(T)10	10	3945	3954	matK/trnK-UUU (IGS)
mono	(A)8	8	4201	4208	trnK-UUU/rps16 (IGS)
mono	(A)9	9	4231	4239	trnK-UUU/rps16 (IGS)
mono	(A)8	8	4451	4458	trnK-UUU/rps16 (IGS)
mono	(A)11	11	4645	4655	trnK-UUU/rps16 (IGS)
di	(AT)4	8	4729	4736	trnK-UUU/rps16 (IGS)
mono	(T)9	9	4770	4778	trnK-UUU/rps16 (IGS)
mono	(T)10	10	5003	5012	trnK-UUU/rps16 (IGS)
mono	(A)8	8	5680	5687	rps16 (intron)
mono	(T)8	8	5930	5937	rps16 (intron)
mono	(A)8	8	7038	7045	psbK/psbI (IGS)
c*	(C)10(T)11	21	7153	7173	psbK/psbI (IGS)
mono	(A)8	8	7215	7222	psbK/psbI (IGS)
mono	(A)14	14	7531	7544	psbI (CDS) - psbI/trnS-GCU (IGS)
mono	(A)8	8	7744	7751	trnS-GCU/trnG-UCC (IGS)
mono	(T)9	9	7787	7795	trnS-GCU/trnG-UCC (IGS)
di	(AT)6	12	7808	7819	trnS-GCU/trnG-UCC (IGS)
c*	(AT)5(TA)4	18	7834	7851	trnS-GCU/trnG-UCC (IGS)
mono	(A)9	9	7904	7912	trnS-GCU/trnG-UCC (IGS)
mono	(T)10	10	7959	7968	trnS-GCU/trnG-UCC (IGS)
mono	(T)11	11	8002	8012	trnS-GCU/trnG-UCC (IGS)
tetra	(ATTT)3	12	8044	8055	trnS-GCU/trnG-UCC (IGS)
mono	(A)11	11	8069	8079	trnS-GCU/trnG-UCC (IGS)
mono	(T)11	11	9073	9083	trnG-UCC/trnR-UCU (IGS)
mono	(A)9	9	9190	9198	trnG-UCC/trnR-UCU (IGS)
mono	(A)8	8	9203	9210	trnG-UCC/trnR-UCU (IGS)
di	(TA)4	8	9384	9391	trnR-UCU/atpA (IGS)
tetra	(TTAT)3	12	9485	9496	trnR-UCU/atpA (IGS); atpA (CDS)
mono	(T)8	8	11033	11040	atpA/atpF (IGS)
mono	(A)11	11	11047	11057	atpA/atpF (IGS)
mono	(T)13	13	11628	11640	atpF (intron)
mono	(A)8	8	12384	12391	atpF/atpH (IGS)
di	(TA)5	10	13023	13032	atpH/atpI (IGS)
mono	(A)8	8	13411	13418	atpH/atpI (IGS)
mono	(T)8	8	14511	14518	atpI/rps2 (IGS)
c*	(C)8(A)11	19	14556	14574	atpI/rps2 (IGS)
mono	(T)11	11	15340	15350	rps2/rpoC2 (IGS)
mono	(T)10	10	17512	17521	rpoC2 (CDS)
mono	(A)8	8	17654	17661	rpoC2 (CDS)
mono	(T)9	9	18046	18054	rpoC2 (CDS)
di	(AT)5	10	18884	18893	rpoC2 (CDS)
mono	(A)8	8	21336	21343	rpoC1 (CDS)
mono	(T)12	12	21935	21946	rpoC1 (intron)
mono	(T)9	9	21990	21998	rpoC1 (intron)
mono	(T)8	8	23505	23512	rpoB (CDS)
mono	(T)10	10	25241	25250	rpoB (CDS)
mono	(A)9	9	25908	25916	rpoB/trnC-GCA (IGS)
mono	(A)8	8	26253	26260	rpoB/trnC-GCA (IGS)
mono	(T)10	10	26402	26411	rpoB/trnC-GCA (IGS)
mono	(A)10	10	26715	26724	rpoB/trnC-GCA (IGS)
mono	(A)10	10	26883	26892	rpoB/trnC-GCA (IGS)
mono	(A)9	9	27389	27397	trnC-GCA/petN (IGS)
c*	(C)9(T)12	21	27574	27594	trnC-GCA/petN (IGS)
mono	(T)8	8	27836	27843	petN/psbM (IGS)
mono	(A)8	8	27865	27872	petN/psbM (IGS)
mono	(T)11	11	28455	28465	psbM/trnD-GUC (IGS)

mono	(T)8	8	28536	28543	psbM/trnD-GUC (IGS)
mono	(A)8	8	28685	28692	psbM/trnD-GUC (IGS)
c*	(A)9(AC)4	16	29349	29364	psbM/trnD-GUC (IGS)
mono	(A)8	8	29697	29704	trnD-GUC/trnY-GUA (IGS)
mono	(T)8	8	29738	29745	trnD-GUC/trnY-GUA (IGS)
mono	(A)8	8	29849	29856	trnD-GUC/trnY-GUA (IGS)
mono	(A)9	9	30333	30341	trnE-UUC/trnT-GGU (IGS)
mono	(A)8	8	30668	30675	trnE-UUC/trnT-GGU (IGS)
di	(AT)4	8	30743	30750	trnE-UUC/trnT-GGU (IGS)
di	(AG)4	8	31032	31039	trnT-GGU/psbD (IGS)
di	(AT)5	10	31110	31119	trnT-GGU/psbD (IGS)
di	(TA)4	8	31478	31485	trnT-GGU/psbD (IGS)
mono	(A)8	8	31992	31999	trnT-GGU/psbD (IGS)
mono	(G)8	8	33477	33484	psbC (CDS)
mono	(G)9	9	33747	33755	psbC (CDS)
di	(GA)4	8	34837	34844	trnS-UGA (CDS)
di	(TA)4	8	35022	35029	trnS-UGA/psbz (IGS)
mono	(A)12	12	35068	35079	trnS-UGA/psbz (IGS)
di	(TA)4	8	35606	35613	psbz/trnG-GCC (IGS)
mono	(T)10	10	35934	35943	trnG-GCC/ trnfM-CAU (IGS)
mono	(A)10	10	35945	35954	trnG-GCC/ trnfM-CAU (IGS)
mono	(A)9	9	36039	36047	trnfM-CAU/rps14 (IGS)
di	(AT)5	10	41576	41585	psaA/ycf3 (IGS)
mono	(A)12	12	41776	41787	psaA/ycf3 (IGS)
di	(AT)4	8	42547	42554	ycf3 (Intron)
mono	(T)8	8	42642	42649	ycf3 (Intron)
mono	(A)9	9	43631	43639	ycf3 (Intron)
mono	(A)11	11	43934	43944	ycf3/trnS-GGA (IGS)
c*	(T)8(AT)6	20	44971	44990	rps4/trnT-UGU (IGS)
di	(AT)4	8	45082	45089	rps4/trnT-UGU (IGS)
mono	(T)10	10	45602	45611	trnT-UGU/trnL-UAA (IGS)
mono	(A)8	8	45843	45850	trnT-UGU/trnL-UAA (IGS)
di	(AT)5	10	45926	45935	trnT-UGU/trnL-UAA (IGS)
mono	(A)8	8	46329	46336	trnL-UAA (intron)
di	(TA)4	8	46457	46464	trnL-UAA (intron)
mono	(T)8	8	46517	46524	trnL-UAA (intron)
mono	(T)8	8	46803	46810	trnL-UAA/trnF-GAA (IGS)
mono	(T)8	8	48916	48923	ndhK/ndhC (IGS)
di	(AT)5	10	49839	49848	ndhC/trnV-UAC (IGS)
mono	(T)9	9	49877	49885	ndhC/trnV-UAC (IGS)
mono	(T)8	8	50168	50175	trnV-UAC (intron)
mono	(T)10	10	50225	50234	trnV-UAC (intron)
mono	(A)11	11	50237	50247	trnV-UAC (intron)
mono	(T)11	11	52943	52953	atpB (CDS)
mono	(A)8	8	53037	53044	atpB/rbcL (IGS)
mono	(A)8	8	55332	55339	rbcL/accD (IGS)
mono	(T)8	8	55820	55827	rbcL/accD (IGS)
mono	(T)8	8	56242	56249	accD (CDS)
mono	(T)8	8	57325	57332	accD (CDS)
mono	(T)14	14	57902	57915	psaI/ycf4 (IGS)
mono	(A)13	13	58055	58067	psaI/ycf4 (IGS)
mono	(T)8	8	58897	58904	ycf4/cemA (IGS)
di	(TC)5	10	59306	59315	cemA (CDS)
di	(AT)4	8	60217	60224	petA (CDS)
mono	(A)8	8	60577	60584	petA (CDS)
mono	(T)8	8	61311	61318	petA/ psbJ (IGS)
di	(TA)4	8	61608	61615	petA/ psbJ (IGS)
mono	(T)8	8	61644	61651	petA/ psbJ (IGS)
mono	(A)13	13	61681	61693	petA/ psbJ (IGS)
mono	(T)9	9	61703	61711	petA/ psbJ (IGS)

mono	(A)8	8	61739	61746	petA/ psbJ (IGS)
mono	(T)9	9	61753	61761	petA/ psbJ (IGS)
mono	(A)10	10	61895	61904	petA/ psbJ (IGS)
mono	(A)10	10	62479	62488	psbF (CDS)
mono	(A)8	8	63281	63288	psbE/petL (IGS)
mono	(T)9	9	63519	63527	psbE/petL (IGS)
mono	(A)8	8	63647	63654	psbE/petL (IGS)
mono	(A)9	9	63698	63706	psbE/petL (IGS)
mono	(T)10	10	63743	63752	psbE/petL (IGS)
mono	(A)9	9	64013	64021	psbE/petL (IGS)
mono	(A)8	8	64307	64314	petL/petG (IGS)
mono	(T)9	9	64668	64676	trnW-CCA/trnP-UGG (IGS)
mono	(A)9	9	64723	64731	trnW-CCA/trnP-UGG (IGS)
mono	(T)10	10	64969	64978	trnP-UGG/ psaJ (IGS)
mono	(G)8	8	65449	65456	psaJ/rpL33 (IGS)
mono	(A)9	9	65468	65476	psaJ/rpL33 (IGS)
mono	(T)8	8	65564	65571	psaJ/rpL33 (IGS)
di	(TA)5	10	65959	65968	rpl33/rps18 (IGS)
di	(AT)4	8	65970	65977	rpl33/rps18 (IGS)
mono	(T)8	8	67148	67155	rpl20/rps12 (IGS)
mono	(A)8	8	67167	67174	rpl20/rps12 (IGS)
mono	(A)8	8	67186	67193	rpl20/rps12 (IGS)
mono	(A)8	8	67478	67485	rpl20/rps12 (IGS)
mono	(A)9	9	67918	67926	rps12/clpP (IGS)
mono	(A)8	8	68300	68307	clpP (intron)
mono	(T)10	10	68456	68465	clpP (intron)
mono	(T)10	10	68547	68556	clpP (intron)
mono	(A)11	11	68702	68712	clpP (intron)
mono	(T)9	9	69358	69366	clpP (intron)
mono	(T)8	8	69540	69547	clpP (intron)
mono	(T)9	9	69639	69647	clpP (intron)
mono	(A)10	10	70153	70162	clpP/psbB (IGS)
mono	(T)8	8	72104	72111	psbB/psbT (IGS)
mono	(T)8	8	72250	72257	psbT (CDS)
mono	(T)9	9	73230	73238	petB (intron)
mono	(T)8	8	74393	74400	petB/petD (IGS)
mono	(A)8	8	75788	75795	petD/ rpoA (IGS)
mono	(T)9	9	75925	75933	petD/ rpoA (IGS)
mono	(A)8	8	75936	75943	petD/ rpoA (IGS)
mono	(T)9	9	77517	77525	rps11/rpl36 (IGS)
mono	(T)9	9	78113	78121	rpl36/rps8 (IGS)
mono	(T)11	11	78625	78635	rps8/rpl14 (IGS)
mono	(T)8	8	78719	78726	rps8/rpl14 (IGS)
mono	(A)10	10	79164	79173	rpl14/rpl16 (IGS)
mono	(A)9	9	79191	79199	rpl14/rpl16 (IGS)
mono	(T)8	8	80463	80470	rpl16 (intron)
mono	(A)8	8	80475	80482	rpl16 (intron)
mono	(T)8	8	80745	80752	rpl16/rps3 (IGS)
mono	(T)9	9	82209	82217	rps19 (CDS)
mono	(T)10	10	82248	82257	rps19/rpl2 (IGS)
di	(TA)4	8	82835	82842	rpl2 (intron)
di	(GA)4	8	84454	84461	ycf2 (CDS)
di	(GA)4	8	85441	85448	ycf2 (CDS)
mono	(T)8	8	86416	86423	ycf2 (CDS)
mono	(A)9	9	87584	87592	ycf2 (CDS)
di	(GA)4	8	87605	87612	ycf2 (CDS)
di	(CT)5	10	89662	89671	ycf2 (CDS)
mono	(G)8	8	91547	91554	ycf2/trnL-CAA (IGS)
di	(TA)4	8	91945	91952	trnL-CAA/ndhB (IGS)
di	(AG)4	8	92727	92734	ndhB (CDS)

tetra	(TTAT)3	12	96617	96628	rps12/trnV-GAC (IGS)
mono	(T)11	11	96631	96641	rps12/trnV-GAC (IGS)
hexa	(CTATCC)3	18	96818	96835	rps12/trnV-GAC (IGS)
mono	(C)8	8	98209	98216	trnV-GAC/rrn16 (IGS)
mono	(G)8	8	100473	100480	trnI-GAU (intron)
mono	(T)8	8	100670	100677	trnI-GAU (intron)
di	(CT)4	8	103985	103992	rrn23 (CDS)
mono	(T)9	9	105878	105886	rrn5/trnR-ACG (IGS)
mono	(T)12	12	106123	106134	trnR-ACG/trnN-GUU (IGS)
mono	(T)8	8	106154	106161	trnR-ACG/trnN-GUU (IGS)
mono	(A)13	13	107959	107971	ycf1 (CDS)
mono	(A)10	10	108017	108026	ycf1 (CDS)
mono	(A)8	8	108188	108195	ycf1 (CDS)
mono	(A)8	8	108324	108331	ycf1 (CDS)
mono	(A)8	8	108506	108513	ndhF (CDS)
mono	(A)8	8	109161	109168	ndhF (CDS)
mono	(T)9	9	109227	109235	ndhF (CDS)
mono	(A)9	9	109487	109495	ndhF (CDS)
mono	(T)11	11	110695	110705	ndhF/rpl32 (IGS)
mono	(T)8	8	110789	110796	ndhF/rpl32 (IGS)
mono	(A)8	8	110896	110903	ndhF/rpl32 (IGS)
mono	(T)9	9	111612	111620	ndhF/rpl32 (IGS)
mono	(T)8	8	111627	111634	ndhF/rpl32 (IGS)
di	(TA)4	8	111648	111655	ndhF/rpl32 (IGS)
mono	(A)10	10	111664	111673	ndhF/rpl32 (IGS)
mono	(A)8	8	111717	111724	ndhF/rpl32 (IGS)
mono	(T)10	10	112047	112056	rpl32/trnL-UAG (IGS)
mono	(A)8	8	112352	112359	rpl32/trnL-UAG (IGS)
di	(AT)4	8	112383	112390	rpl32/trnL-UAG (IGS)
mono	(T)8	8	112595	112602	rpl32/trnL-UAG (IGS)
mono	(A)8	8	112857	112864	rpl32/trnL-UAG (IGS)

Supplementary Table S4. RNA-editing sites predicted in plastid protein-coding genes of *T. pentaphyllum* and other species of Brassicales. Amino acids are indicated in parentheses. Putative edited codons are indicated by arrow (\Rightarrow). ⁺, indicated RNA editing sites present only in *T. pentaphyllum*; *, RNA editing sites absent only in this species

Gene	AA position (align)	Codon position	<i>T. pentaphyllum</i> (Tropaeolaceae)	<i>B. sinensis</i> (Akaniaceae)	<i>M. oleifera</i> (Moringaceae)	<i>C. papaya</i> (Caricaceae)	<i>T. hassleriana</i> (Cleomaceae)	<i>A. thaliana</i> (Brassicaceae)
<i>accD</i>	263	1	UAU (Y)	UAU (Y)	UAU (Y)	UAU (Y)	UAU (Y)	CAU (H) \Rightarrow UAU (Y)
	276	2	UCG(S) \Rightarrow UUG(L)	ACG(T)	UCG(S) \Rightarrow UUG(L)	UUG(L)	UCG(S) \Rightarrow UUG(L)	UCG(S) \Rightarrow UUG(L)
	479	2	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CUU(L)
<i>atpA</i>	264	2	CUU(L)	CUU(L)	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CUU(L)	CUU(L)
	305	2	UUA(L)	UUA(L)	UCA(S) \Rightarrow UUA(L)	UUA(L)	UUA(L)	UUA(L)
<i>atpF</i>	31	2	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)	CCA(P) \Rightarrow CUA(L)
<i>ccsA</i>	48	2	AUA(I)	ACA(T) \Rightarrow AUA(I)	AUA(I)	AUA(I)	AUA(I)	AUA(I)
<i>clpP</i>	186	2	GUU(V)	GUU(V)	GUU(V)	GUU(V)	GUU(V)	GUU(V)
	196	1	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	UAU(Y)	UAU(Y)	CAU(H) \Rightarrow UAU(Y)
<i>matK</i>	64	1	GCG(A) \Rightarrow GUG(V)	GCG(A) \Rightarrow GUG(V)	GUG(V)	GUG(V)	GUG(V)	GUG(V)
	160	1	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	UAU(Y)	UAU(Y)	UAU(Y)	UAU(Y)
	166	2	AUC(I)	AUC(I)	AUC(I)	ACC(T) \Rightarrow AUC(I)	AUC(I)	AUC(I)
	191	1	CAU(H) \Rightarrow UAU(Y) ⁺	UAU(Y)	UAU(Y)	UAU(Y)	UAU(Y)	UAU(Y)
	219	1	UAU(Y)*	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)	CAU(H) \Rightarrow UAU(Y)
<i>ndhA</i>	410	2	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)	UCA(S) \Rightarrow UUA(L)
	430	1	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)	CAC(H) \Rightarrow UAC(Y)	UAC(Y)	CAC(H) \Rightarrow UAC(Y)	UAC(Y)
	36	2	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CCU(P) \Rightarrow CUU(L)	CUU(L)	CUU(L)	CUU(L)

	42	2	ACA(T)=>AUA(I)	ACA(T)=>AUA(I)	ACA(T)=>AUA(I)	ACA(T)=>AUA(I)	ACA(T)=>AUA(I)	ACA(T)=>AUA(I)	ACA(T)=>AUA(I)
	114	2	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	GCA(A)
	189	2	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UUA(L)	UUA(L)
	358	2	UCC(S)=>UUC(F)	UCC(S)=>UUC(F)	UCC(S)=>UUC(F)	UCC(S)=>UUC(F)	UCC(S)=>UUC(F)	UUC(F)	UUC(F)
<i>ndhF</i>	97	2	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)
	196	1	CUU(L)=>UUU(F)	CUU(L)=>UUU(F)	CUU(L)=>UUU(F)	CUU(L)=>UUU(F)	CUU(L)=>UUU(F)	CUU(L)=>UUU(F)	CUU(L)=>UUU(F)
	640	2	GUG(V)	GGG(G)	UUG(L)	GCA(A)=>GUA(V)	CAA(Q)	CAA(Q)	CAA(Q)
	711	2	AUA(I)	AUA(I)	AUA(I)	AUA(I)	AUA(I)	AUA(I)	ACA(T)=>AUA(I)
<i>ndhG</i>	44	2	GUU(V)	GUU(V)	GUU(V)	GUU(V)	GUU(V)	GUU(V)	GUU(V)
	56	1	CAU(H)=>UAU(Y)	CAU(H)=>UAU(Y)	CAU(H)=>UAU(Y)	CAU(H)=>UAU(Y)	CAU(H)=>UAU(Y)	CAU(H)=>UAU(Y)	CAU(H)=>UAU(Y)
	105	2	ACA(T)=>ATA(I)	ACA(T)=>ATA(I)	ACA(T)=>ATA(I)	ACA(T)=>ATA(I)	ACA(T)=>ATA(I)	ACA(T)=>ATA(I)	ACA(T)=>ATA(I)
	121	2	UAC(Y)	UAC(Y)	UAC(Y)	UAC(Y)	UAC(Y)	CAC(H)=>UAC(Y)	UAC(Y)
<i>petB</i>	144	1	CGG(R)=>UGG(W)	UGG(W)	CGG(R)=>UGG(W)	UGG(W)	UGG(W)	UGG(W)	UGG(W)
<i>petD</i>	139	2	GUU(V)	GUU(V)	GUU(V)	GUU(V)	GUU(V)	GUU(V)	GCG(A)=>GUU(V)
<i>psaI</i>	28	2	UCU(S)=>UUU(F)	UCU(S)=>UUU(F)	UCU(S)=>UUU(F)	UCU(S)=>UUU(F)	UCU(S)=>UUU(F)	UUU(F)	UCU(S)=>UUU(F)
<i>psbE</i>	72	1	CCU(P)=>UCU(S)	CCU(P)=>UCU(S)	CCU(P)=>UCU(S)	CCU(P)=>UCU(S)	CCU(P)=>UCU(S)	CCU(P)=>UCU(S)	CCU(P)=>UCU(S)
<i>psbF</i>	26	2	UUU(F)*	UCU(S)=>UUU(F)	UCU(S)=>UUU(F)	UCU(S)=>UUU(F)	UCU(S)=>UUU(F)	UCU(S)=>UUU(F)	UCU(S)=>UUU(F)
<i>rpl20</i>	103	2	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UCA(S)=>UUA(L)	UUA(L)	UUA(L)

Chapter III

Plastome sequences of the subgenus *Passiflora* reveal highly divergent genes and specific evolutionary features

Túlio Gomes Pacheco^{1a}, Amanda de Santana Lopes^{1a}, Juliana Fátima Welter¹, Karla Suemy Clemente Yotoko², Wagner Campos Otoni³, Leila do Nascimento Vieira⁴, Miguel Pedro Guerra⁴, Rubens Onofre Nodari⁴, Eduardo Balsanelli⁵, Fábio de Oliveira Pedrosa⁵, Emanuel Maltempi de Souza⁵, Marcelo Rogalski^{1*}

¹ Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa-MG, Brasil.

² Laboratório de Bioinformática e Evolução, Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa-MG, Brasil

³ Laboratório de Cultura de Tecidos Vegetais, Departamento de Biologia Vegetal, BIOAGRO, Universidade Federal de Viçosa, Viçosa-MG, Brasil.

⁴ Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal, Programa de Pós-Graduação em Recursos Genéticos Vegetais, Universidade Federal de Santa Catarina, Florianópolis-SC, Brazil.

⁵ Departamento de Bioquímica e Biologia Molecular, Núcleo de Fixação Biológica de Nitrogênio, Universidade Federal do Paraná, Curitiba-PR, Brasil.

^aTúlio Gomes Pacheco and Amanda de Santana Lopes have contributed equally to this work.

*Corresponding author

E-mail address: rogalski@ufv.br

Published in: **Plant Molecular Biology** (2020) 104, 21-35

<https://doi.org/10.1007/s11103-020-01020-z>



Plastome sequences of the subgenus *Passiflora* reveal highly divergent genes and specific evolutionary features

Túlio Gomes Pacheco¹ · Amanda de Santana Lopes¹ · Juliana Fátima Welter¹ · Karla Suemy Clemente Yotoko² · Wagner Campos Otoni³ · Leila do Nascimento Vieira⁴ · Miguel Pedro Guerra⁴ · Rubens Onofre Nodari⁴ · Eduardo Balsanelli⁵ · Fábio de Oliveira Pedrosa⁵ · Emanuel Maltempi de Souza⁵ · Marcelo Rogalski¹

Received: 21 February 2020 / Accepted: 7 June 2020
 © Springer Nature B.V. 2020

Abstract

Key Message Phylogenetic aspects, hotspots of nucleotide divergence, highly divergent genes, and specific RNA editing sites have been identified and characterized in the plastomes of the subgenus *Passiflora*.

Abstract The genus *Passiflora* comprises more than 500 species across five subgenera: *Astrophea*, *Decaloba*, *Deidamioides*, *Passiflora*, and *Tetrapathea*. The most economically relevant species belong to the subgenus *Passiflora*, whose genetic pool and diversity among wild species remain poorly characterized. Similarly, little is known about the interspecific relationships within the subgenus *Passiflora* and the genetic causes of nuclear-cytoplasmic incompatibility observed in interspecific hybrids. Here, we report the complete nucleotide sequences of six plastomes belonging to species of the subgenus *Passiflora*, with the aim of better understanding the evolution of the plastome in this subgenus. Complete plastome sequences revealed five hotspots of nucleotide polymorphism: three intergenic regions and two coding sequences. Moreover, among 70 RNA editing sites predicted in our analysis for the subgenus *Passiflora*, 38 were not shared by all analyzed species, highlighting their species-specific occurrence. Furthermore, phylogenies based on plastid sequences accurately resolved most relationships between species and suggested a non-monophyletic origin of three super-sections of the subgenus *Passiflora*, previously defined solely based on morphological traits. Finally, our findings identified putative candidates, including predicted RNA editing sites and the coding sequences of *accD* and *clpP* genes, responsible for nuclear-cytoplasmic incompatibility in the interspecific hybrids of *Passiflora*.

Keywords Passifloraceae · Organelle DNA · Genetic incompatibility · Plastome evolution · Polymorphism hotspots

Introduction

The genus *Passiflora* comprises more than 500 species and is the largest within the pantropical family Passifloraceae (MacDougal and Feuillet 2004; Rocha et al. 2020). Its center of diversity in the Neotropics accounts for approximately 95% of all *Passiflora* species in South America, with some

species found also in the Old World, Southeast Asia, Australia, and Pacific Islands (Yockteng et al. 2011). The first infrageneric classification was organized by Killip (1938) based on floral traits and led to division of the genus *Passiflora* into 22 subgenera. The current classification based on morphological characteristics includes only four subgenera: *Astrophea* (DC.) Mast., *Decaloba* (DC.) Rehb., *Deidamioides* (Harms) Killip, and *Passiflora* Feuillet & MacDougal (MacDougal and Feuillet 2004). More recently, phylogenetic analysis has suggested the existence of another subgenus, *Tetrapathea* (Krosnick et al. 2009, 2013).

Species of the genus *Passiflora* are commonly known as passionflowers and are used for food, medicinal, cosmetic, and ornamental purposes (Yockteng et al. 2011; Cerqueira-Silva et al. 2014a; Rocha et al. 2020). All species of economic importance belong to the subgenus *Passiflora*, which is the richest and comprises 45% of entries in the genus

Túlio Gomes Pacheco and Amanda de Santana Lopes contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11103-020-01020-z>) contains supplementary material, which is available to authorized users.

✉ Marcelo Rogalski
 rogalski@ufv.br

Extended author information available on the last page of the article

Published online: 13 June 2020

Springer

Passiflora (MacDougal and Feuillet 2004; Meletti et al. 2005; Yockteng et al. 2011). Wild species of the subgenus *Passiflora* represent a key genetic resource for breeding programs and identification of new traits, such as resistance to biotic/abiotic stress and higher productivity. Nevertheless, they and their accessions are underrepresented within active germplasm banks of passionflowers. Genetic diversity and potential application of the genetic pool of most wild species in breeding programs have remained largely overlooked (Cerqueira-Silva et al. 2014a, b, c). Another challenge concerning the subgenus *Passiflora* is represented by interspecific relationships, which are traditionally subdivided in super-section, section, and series (MacDougal and Feuillet 2004). Molecular phylogenies of the subgenus *Passiflora* have failed to produce well-resolved and well-supported relationships at low-taxon levels. This is because of the small number of taxa sampled and the lack of informative sequences (Hansen et al. 2006; Ramaiya et al. 2014).

Plastome sequences are convenient tools for phylogenetic inferences (Vieira et al. 2016a; Lopes et al. 2018a) and genetic diversity studies using single-nucleotide polymorphisms or single-sequence repeats (Besnard et al. 2011; Rogalski et al. 2015; Qiao et al. 2016; Pacheco et al. 2019). Complete plastome sequences allow the investigation of usual and rare evolutionary events, such as rearrangements, loss of genes and introns, horizontal gene transfer, positive selection, and phylogeography (Wicke et al. 2011; Vieira et al. 2016b; Lopes et al. 2018b, 2019; Stefenon et al. 2019). They are also essential for designing appropriate transformation vectors for biotechnological applications and reverse genetics in plastids (Rogalski et al. 2006, 2008a; Rogalski and Carrer 2011; Alkatib et al. 2012; Daniell et al. 2016). A total of 30 plastomes belonging to different species of the genus *Passiflora* have been completely sequenced to date (Cauz-Santos et al. 2017; Rabah et al. 2019; Shrestha et al. 2019); they include representative species of the subgenera *Astrophea*, *Decaloba*, *Deidamioides*, *Passiflora*, and *Tetrapathea*, which have revealed several rearrangements and losses of genes and introns compared with other angiosperms.

Interspecific hybridization within the subgenus *Passiflora* is a convenient strategy to increase the performance of cultivated species by adding new traits and has encouraged attempts to develop interspecific hybrid plants (Abreu et al. 2009). Nevertheless, many combinations of interspecific crosses are incompatible and have failed to yield viable plants (Bugallo et al. 2011; Conceição et al. 2011; Madureira et al. 2014; Ocampo et al. 2016). Furthermore, Mráček (2005) observed that interspecific crosses involving *Passiflora menispermifolia* and several other species of the subgenus *Passiflora*, such as *P. edulis*, *P. oerstedii*, *P. maliformis*, and *P. actinia*, produced variegated and chlorophyll-deficient plants. Analysis of these hybrid plants revealed that

the plastid genome of *P. menispermifolia* was incompatible with the nucleus of the hybrid (Mráček 2005).

In other angiosperm lineages, instances of nucleus-plastome incompatibility were elucidated by comparative analyses of the plastomes of parental plants. In hybrid plants belonging to the genus *Pisum*, incompatibility was likely related to the high divergence of the *accD* gene, affecting the biosynthesis of fatty acids in plastids (Bogdanova et al. 2015). In the genus *Oenothera*, it was associated with deletion of a promoter region between the *clpP* and *psbB* genes, affecting the expression of photosynthesis-related genes (Greiner et al. 2008). Cybrids bearing the nuclear genome of *Atropa belladonna* and the plastids of *Nicotiana tabacum* exhibited nucleus-plastome incompatibility related to a specific RNA editing site in the transcripts of the *atpA* gene (Schmitz-Linneweber et al. 2005). This last example revealed a differential co-evolution of nuclear and plastid genes between members of the family Solanaceae. Thus, a more detailed molecular analysis of plastomes within the subgenus *Passiflora* will help identify possible factors responsible for nuclear-cytoplasmic incompatibility. Additionally, complete plastome sequences and the characterization of specific molecular markers are pivotal for studies related to plastid inheritance via any of the three patterns (maternal, paternal, and biparental) observed within the subgenus *Passiflora* (Lorenz-Lemke et al. 2005; Muschner et al. 2006; Hansen et al. 2006, 2007).

Here, we report the complete plastome sequences of six species belonging to the subgenus *Passiflora* and the characterization of corresponding regions with high nucleotide polymorphism. The complete sequences allow us to infer the phylogeny based on plastid genes and whole plastomes. Several aspects related to classification (i.e., super-sections and sections), diversity of RNA editing sites, and possible candidates for plastid-nucleus incompatibility observed in interspecific hybridizations within the subgenus *Passiflora* are analyzed and discussed.

Materials and methods

Plant material, chloroplast isolation, and DNA extraction

Plants of *Passiflora cincinnata*, *P. elegans*, *P. incarnata*, *P. malacophylla*, *P. maliformis*, and *P. mucronata* were maintained under greenhouse conditions at the Department of Plant Biology, Federal University of Viçosa, Viçosa-MG, Brazil. Young leaves were collected and kept for 96 h at 4 °C to reduce starch content in chloroplasts. Chloroplast isolation and plastid DNA extraction procedures were carried out as described by Vieira et al. (2014).

Plastome sequencing, assembly, annotation and data archiving statement

Approximately 1 ng of plastid DNA from each species was used to prepare sequencing libraries using the Nextera XT DNA Sample Prep Kit (Illumina Inc., San Diego, CA, USA), according to the manufacturer's instructions. The obtained library was sequenced on the Illumina MiSeq platform. The reads were trimmed (threshold with probability of error < 0.05) and assembled de novo in contigs using the CLC Genomics Workbench 6.5 software (CLC Bio, Aarhus, Denmark). The number of total and trimmed reads and the average coverage of the contigs used to assemble the plastomes are shown in Supplementary Table S1. Gene annotations were carried out first using the program Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004). Then, putative start codons, stop codons, and intron positions were determined based on comparisons with homologous genes of other plastomes at the GenBank database. Posteriorly, all tRNA genes were verified using the tRNAscan-SE server (Lowe and Chan 2016). The physical circular map of plastomes was drawn using OrganellarGenomeDRAW (OGDRAW) (Greiner et al. 2019). The complete nucleotide sequences and gene features of the six sequenced plastomes were deposited in the GenBank database under the accession numbers KY820583.1 (*Passiflora cincinnata*), MN062356 (*P. elegans*), MN062357 (*P. incarnata*), MN062358 (*P. malacophylla*), MN062359 (*P. maliformis*), and MN062360 (*P. mucronata*).

Structural analysis and identification of hotspots of nucleotide divergence

The general structural features of plastomes within the subgenus *Passiflora* were analyzed by progressive Mauve based on whole-plastome alignment using Mauve Genome Alignment 2.3.1 software (Darling et al. 2004). This analysis included the six species sequenced here and another 13 species of this subgenus available from GenBank (Cauz-Santos et al. 2017; Rabah et al. 2019; Shrestha et al. 2019; Supplementary Table S2). *Populus trichocarpa* was used as the outgroup species.

To identify the hotspots of polymorphism within the subgenus *Passiflora*, whole-plastome alignment was conducted using MAFFT v.7 (Kato and Standley 2013). Posteriorly, the aligned plastomes were analyzed by sliding windows using the DnaSP v.5 software (Librado and Rozas 2009). Window length and step size were set to 200 and 50 nucleotides, respectively. Next, we selected the most polymorphic regions and performed a more detailed sliding window analysis, with window length and step size set to 20 and 5 nucleotides, respectively.

Phylogenetic reconstruction of the genus *Passiflora* based on concatenated protein-coding genes

Phylogenetic inference based on concatenated protein-coding genes was carried out to infer relationships within the genus *Passiflora*. The analysis included six species of the subgenus *Passiflora* sequenced here, as well as various species of the subgenera *Passiflora* (thirteen), *Deidamioides* (three), *Decaloba* (eleven), *Astrophea* (one), and *Tetrapathea* (one) available from the organelle database. A species of the genus *Adenia* (family Passifloraceae) and *P. trichocarpa* (family Salicaceae) were used as the outgroup species. The GenBank accession number of each taxon used here is listed in Supplementary Table S2. A total of 66 conserved protein-coding genes were extracted from each species and individually aligned using the Muscle software (Edgar 2004) implemented in Mega 6.0 (Tamura et al. 2013). Subsequently, the aligned genes were concatenated using DnaSP v.5, which resulted in a total sequence of 52,023 bp. The best-fit evolutionary model for each gene was selected following Bayesian Information Criterion scores computed in ModelFinder (Kalyaanamoorthy et al. 2017; Supplementary Table S3). Maximum likelihood (ML) estimation was conducted using IQTREE v1.6.10 (Nguyen et al. 2015), and 500 non-parametric bootstrap replications were used to assess branch support. Moreover, Bayesian inference (BI) was conducted with the same dataset. Partition Finder 2 software (Lanfear et al. 2017) was used to search for the best evolutionary model for each gene (Supplementary Table S4), and BI was performed using the MrBayes version 3.2 software (Ronquist et al. 2012), with one million generations of two runs of four Markov Chains (three hot and one cold per run). Lastly, FigTree v.1.4.2 (<https://tree.bio.ed.ac.uk/software/figtree/>) was used to visualize the consensus tree of both ML and BI.

Phylogenomic inference of the subgenus *Passiflora* based on whole-plastomes was carried out using ML and BI methods. First, inverted repeat B (IRB) was omitted to prevent overrepresentation of IR sequences. Subsequently, the inversion located at the large single copy (LSC) identified in *P. foetida*, when compared with other species of the subgenus *Passiflora*, was reoriented to the same direction as in those species to facilitate analysis (Rabah et al. 2019; Shrestha et al. 2019). Finally, the plastomes of the 19 species of the subgenus *Passiflora* (Supplementary Table S2) were aligned using MAFFT v.7. ML estimation was conducted using IQTREE v1.6.10 with the model GTR + F + R2 (selected by ModelFinder) and 500 non-parametric bootstrap replications. BI was conducted in MrBayes 3.2 using the model GTR + I + G (selected by jModelTest v.2.1.7; Durrani et al. 2012).

Gene divergence inference and identification of putative RNA editing sites in protein-coding genes

The plastomes of the six *Passiflora* subgenus species sequenced here and those of the other 13 from the organelle database contained a set of 72 protein-coding genes. To determine the rate of gene divergence, they were first extracted and individually aligned (codon alignment) with Muscle implemented in Mega 6.0. Posteriorly, phylogenetic reconstruction was performed to assess gene divergence. The search for the best substitution model and BI analysis were performed as described above. Gene divergence was estimated by adding up all total branch lengths that linked the taxonomic units to the common ancestor of the subgenus *Passiflora* based on the 19 species sampled here.

Potential RNA editing sites in protein-coding genes within the subgenus *Passiflora* were predicted by the Predictive RNA Editor for Plants (PREP) suite (Mower 2009). PREP uses the following 35 reference genes to detect possible RNA editing sites in the plastomes: *accD*, *atpA*, *atpB*, *atpF*, *atpI*, *ccsA*, *clpP*, *matK*, *ndhA*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *petB*, *petD*, *petG*, *petL*, *psaB*, *psaI*, *psbA*, *psbB*, *psbE*, *psbF*, *psbL*, *rpl2*, *rpl20*, *rpl23*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps2*, *rps8*, *rps14*, *rps16*, and *ycf3*. The cutoff value was set to 0.8 according to that reported by Lopes et al. (2018a) and Pacheco et al. (2019).

Results

General features of the plastomes within the subgenus *Passiflora*

The choice of *Passiflora* species (*P. cincinnata*, *P. elegans*, *P. incarnata*, *P. malacophylla*, *P. maliformis*, and *P. mucronata*) was based on their availability in our germplasm collection and their morphological characteristics (Mondin et al. 2011; Chitwood and Otoni 2017a, b; Rocha et al. 2020), which indicated variability in leaf morphology, flowers, fruits, and growth patterns.

The sequenced plastomes of the above species showed a typical quadripartite structure that included a pair of IRs (IRA and IRB) and two regions of a single copy, the large single copy (LSC) and the small single copy (SSC) (Fig. 1 and Table 1). The physical map of *P. elegans* as a representative species of the subgenus *Passiflora* is illustrated in Fig. 1, whereas the maps of the other sequenced species are shown in Supplementary Figs. S1–S5. The plastomes ranged from 145,644 bp (*P. elegans*) to 151,337 bp (*P. incarnata*). The size of LSC and SSC regions did not vary significantly between the species, ranging from 84,208 bp (*P. maliformis*) to 86,264 bp (*P. cincinnata*) and 13,167 bp (*P. elegans*) to

13,626 bp (*P. incarnata*), respectively. In contrast, IRs varied significantly from 23,689 bp (*P. elegans*) to 26,337 bp (*P. incarnata*). GC content in the six sequenced species ranged from 36.9% (*P. incarnata*, *P. malacophylla*, and *P. maliformis*) to 37.1% (*P. cincinnata* and *P. elegans*) (Table 1).

Except for specific features found in *P. foetida* and *P. menispermifolia* (Shrestha et al. 2019), overall structural characteristics and gene content of the plastomes sequenced and assembled here (Table 1) were conserved relative to the other species of the subgenus *Passiflora* (Cauz-Santos et al. 2017; Rabah et al. 2019). The plastome of other species in the subgenus *Passiflora* ranges from 133,682 bp (*P. menispermifolia*) to 162,226 bp (*P. foetida*) (Shrestha et al. 2019), with LSC varying from 84,635 bp (*P. foetida*) to 88,369 bp (*P. menispermifolia*), SSC varying from 13,268 bp (*P. oerstedii*) to 24,873 bp (*P. menispermifolia*), and IRs varying from 10,220 bp (*P. menispermifolia*) to 32,062 bp (*P. foetida*) (Cauz-Santos et al. 2017; Rabah et al. 2019; Shrestha et al. 2019). GC content in plastomes within the subgenus *Passiflora* ranges from 36.3% (*P. menispermifolia*) to 37.2% (*P. vitifolia*). The variation in these plastome regions among species of the subgenus *Passiflora* is influenced by the size of intergenic spacers (IGSs), IR contraction events, and different patterns of gene degeneration (Table 1; Cauz-Santos et al. 2017; Rabah et al. 2019; Shrestha et al. 2019).

All plastomes within the subgenus *Passiflora* analyzed here shared a different gene order compared with plastomes belonging to species of related families, such as Salicaceae and other angiosperms. Structurally, plastomes of the subgenus *Passiflora* have a small inversion of 1.6 kb covering the *psbA* and *trnH-GUG* genes at the beginning of the LSC region (LCB1; Supplementary Fig. S6). The *P. foetida* plastome contains a second inversion of approximately 3 kb in the LSC region, extending the sequence from the *trnV-UAC* gene to the *atpB* gene (LCB2; Supplementary Fig. S6). In all other species of the subgenus *Passiflora*, the LCB2 region follows the same orientation observed in *Populus*, which was likely reverted in most species of the subgenus *Passiflora*. Compared with *P. foetida*, all species of the subgenus *Passiflora* contained a large inversion of approximately 47 kb in the LSC region, extending from the *trnC-GCA* gene to the *clpP* gene (LCB4 and LCB5; Supplementary Fig. S6). This inversion included a reversion of the LCB2 sequence to the same direction observed in *Populus*. In addition, the newly sequenced species have expanded IRs including the full sequence of the *ycf1* gene/pseudogene (LCB3; Supplementary Fig. S6).

Our data corroborate the conserved set of genes previously identified within the subgenus *Passiflora* (Rabah et al. 2019; Shrestha et al. 2019). Except for *P. foetida*, all species sequenced to date contain 106 unique genes, which correspond to 72 protein-coding genes, 30 tRNAs, and 4

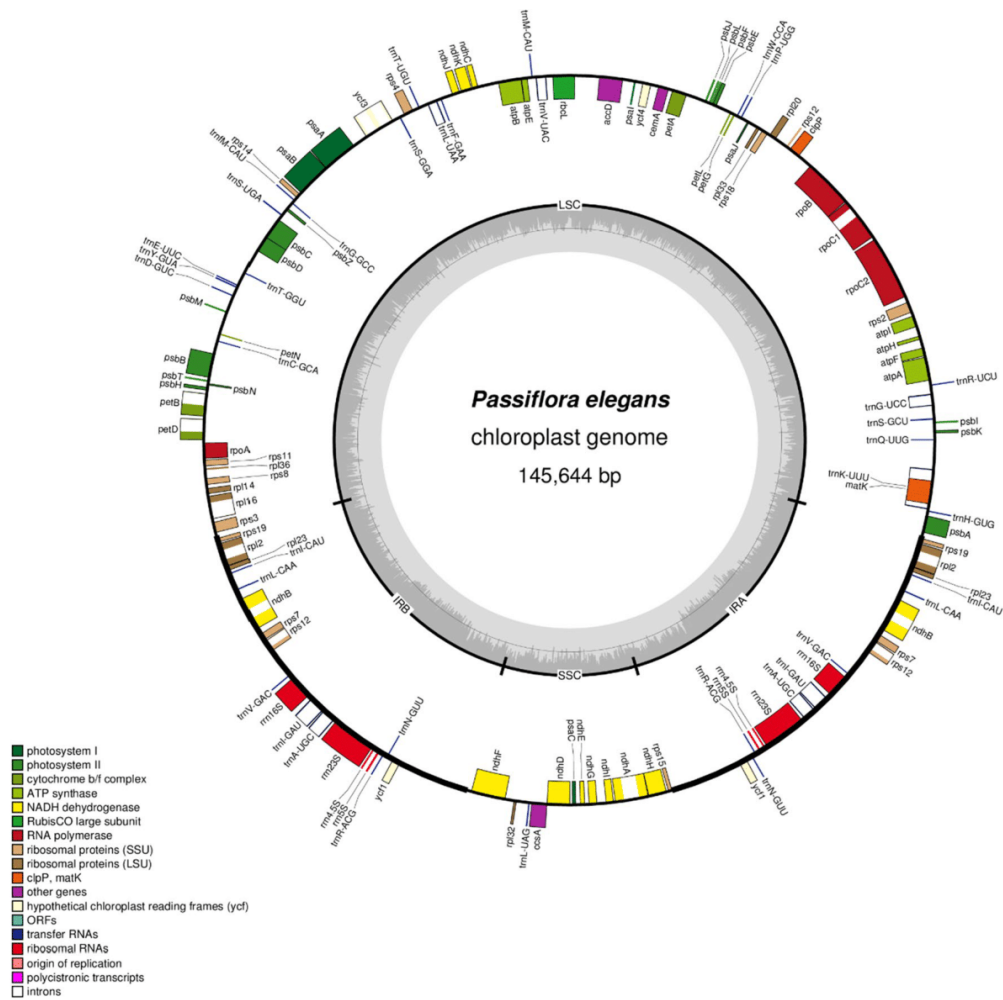


Fig. 1 Gene map of *Passiflora elegans* plastome. Genes drawn inside the circle are transcribed clockwise and genes drawn outside are transcribed counterclockwise. Different functional groups of genes are

color-coded. The darker gray in the inner circle corresponds to GC content, and the lighter gray corresponds to AT content. *LSC* Large single copy, *SSC*, small single copy, *IRA/B* inverted repeat A/B

Table 1 Summary of the plastome characteristics among species within the subgenus *Passiflora* completely sequenced and assembled in this study

Species	Total size (Kb)	LSC (Kb)	SSC (Kb)	IR (Kb)	GC (%)	GenBank
<i>P. cincinnata</i> Mast	150,783	86,264	13,171	25,674	37.1	NC_037690.1
<i>P. elegans</i> Mast	145,644	85,099	13,167	23,689	37.1	MN062356
<i>P. incarnata</i> L	151,337	85,643	13,626	26,337	36.9	MN062357
<i>P. malacophylla</i> Mast	148,933	85,173	13,522	25,119	36.9	MN062358
<i>P. maliformis</i> L	148,868	84,208	13,480	25,590	36.9	MN062359
<i>P. mucronata</i> Lam	150,473	84,859	13,174	26,220	37.0	MN062360

rRNAs (Tables 1 and 2). *P. foetida* contains two additional genes owing to the presence of intact and functional *ycf1* and *ycf2* genes (Shrestha et al. 2019). All other species of the

subgenus contain only pseudogene form of these two genes, except for the plastome of *P. elegans*, in which the pseudogene *ycf2* was completely lost. Two other pseudogenes

Table 2 List of genes identified in the plastomes of the subgenus *Passiflora* completely sequenced and assembled in this study

Group of gene	Name of gene
Gene expression machinery	
Ribosomal RNA genes	<i>rrn16^b; rrn23^b; rrn5^b; rrn4.5^b</i>
Transfer RNA genes	<i>trnA-UGC^{ab}; trnC-GCA; trnD-GUC; trnE-UUC; trnF-GAA; trnG-CAU; trnG-UCC^a; trnG-GCC; trnH-GUG; trnI-CAU^b; trnI-GAU^{ab}; 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-UGA; trnS-GGA; trnT-UGU; trnT-GGU; trnV-GAC^b; trnV-UAC^a; trnW-CCA; trnY-GUA</i>
Small subunit of ribosome	<i>rps2; rps3; rps4; rps8; rps11; rps12^a; rps14; rps15; rps18; rps19^b</i>
Large subunit of ribosome	<i>rpl2^{ab}; rpl14; rpl16^a; rpl23^b; rpl32; rpl33; rpl36</i>
DNA-dependent RNA polymerase	<i>rpoA; rpoB; rpoCI^a; rpoC2</i>
Intron maturase	<i>matK</i>
Genes for photosynthesis	
Subunits of photosystem I (PSI)	<i>psaA; psaB; psaC; psal; psaj; ycf3^a; ycf4</i>
Subunits of photosystem II (PSII)	<i>psbA; psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM; psbN; psbT; psbZ</i>
Subunits of cytochrome <i>b₆f</i>	<i>petA; petB^a; petD^a; petG; petL; petN</i>
Subunits of ATP synthase	<i>atpA; atpB; atpE; atpF; atpH; atpI</i>
Subunits of NADH dehydrogenase	<i>ndhA^a; ndhB^{ab}; ndhC; ndhD; ndhE; ndhF; ndhG; ndhH; ndhI; ndhJ; ndhK</i>
Large subunit of Rubisco	<i>rbcL</i>
Others functions	
Envelope membrane protein	<i>cemA</i>
Subunit of acetyl-CoA carboxylase	<i>accD</i>
C-type cytochrome synthesis	<i>ccsA</i>
Subunit of protease Clp	<i>clpP</i>
Pseudogenes	<i>rpl20; rps7^b; ycf1^b; ycf2^b</i>
Genes absent	<i>infA; rps16; rpl22</i>

^aGenes containing introns; ^bDuplicated gene

are represented by *rps7* and *rpl20*, whereas other genes completely lost from plastomes of the subgenus *Passiflora* include *infA*, *rps16*, and *rpl22*, as well as the introns of *clpP* and *atpF*.

Phylogenies of the genus *Passiflora*

Phylogenetic inferences were carried out based on 66 concatenated plastid protein-coding genes, using both ML and BI on 37 sampled species of the genus *Passiflora* plus one species of the genus *Adenia*. ML and BI inferences produced a consensus tree with identical topology and log-likelihood (lnL) values of $-176,769.6603$ and $-176,776.8088$, respectively (Fig. 2a). This phylogeny expands on the inference performed by Shrestha et al. (2019) by including five additional species of the subgenus *Passiflora* sequenced here (*P. cincinnata* was already included in that study). All nodes except the indicated ones (six nodes) exhibited bootstrap support (BS) and posterior probability (PP) values of 100%. *Adenia mannii* was strongly supported as sister to the genus *Passiflora*. *P. pittieri* (subgenus *Astrophea*) was sister to *P. arbelaezii* (subgenus *Deidamioides*), forming a sister group to the clade containing all other *Passiflora* species. This clade could be divided into two other clades. One was

formed by all species of the subgenus *Passiflora* and supported strongly (100% BS and PP) the monophyletic origin of this subgenus. The other clade was formed by *P. cirrhiflora* and *P. contracta* (*Deidademioides*) and was strongly supported (100% BS and PP) as a sister to the subgenera *Tetrapathea* (represented here by *P. tetrandra*), *Decaloba*, and *Deidamioides* (*P. obovata*). *P. tetrandra* was sister to the subgenus *Decaloba* (94% BS and 100% PP). Within the subgenus *Decaloba*, *P. microstipula* was sister to the clade formed by *P. obovata* (*Deidamioides*) and other species of the *Decaloba* subgenus (100% BS and PP). These results indicate a polyphyletic origin of the subgenus *Deidamioides*, as suggested by one sister species to the subgenus *Astrophea*, two sister species to the subgenera *Tetrapathea* and *Decaloba*, and one within the clade containing species of the subgenus *Decaloba*.

P. foetida was demonstrated to be the early-divergent lineage within the subgenus *Passiflora*. *P. menispermifolia*, *P. oerstedii*, *P. retipetala*, and *P. mucronata* formed a sister group to the clade constituted by all other species of this subgenus. Within this clade, *P. elegans* and *P. actinia* formed a sister group to the clade comprising the remaining species. Among them, *P. incarnata* was sister to two other subclades: the first composed of *P. serratifolia*, *P. maliformis*, *P.*

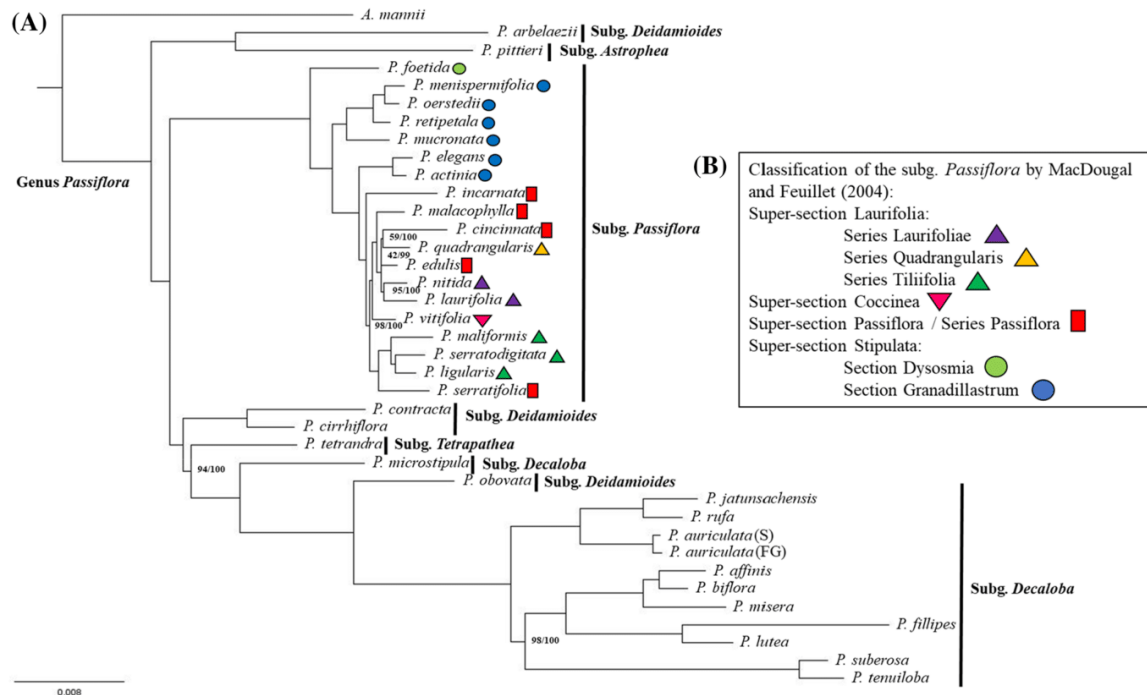


Fig. 2 a Phylogenetic tree of the family Passifloraceae, including *Adenia mannii* and 37 species of the genus *Passiflora*, based on 66 plastid protein-coding genes. *Populus trichocarpa* was used as out-group species to root the tree (omitted from the figure). The tree topology was identical across maximum likelihood (ML) and Bayesian inference (BI) analysis. The numbers in front of the nodes are bootstrap support (BS) values from ML analysis and posterior probabilities (PP) from BI, respectively (BS/PP). The numbers in front of

the nodes with 100% of both BS and PP values were omitted. The branch length is proportional to the inferred divergence level and the scale bar indicates the number of inferred nucleic acid substitutions per site. The subgenera *Astrophea*, *Decaloba*, and *Passiflora* are indicated to the right of the tree. **b** The classification of the subgenus *Passiflora* proposed by MacDougal and Feuillet (2004) is indicated by colored symbols in front of the species in the tree. *S* Suriname, *F* French Guiana

serratodigitata, and *P. ligularis* and the second composed of *P. vittifolia*, *P. malacophylla*, *P. nitida*, *P. laurifolia*, *P. edulis*, *P. cincinnata*, and *P. quadrangularis*. Within the latter clade, *P. nitida* and *P. laurifolia* (95% BS and 100% PP) formed a sister group to the clade composed of *P. quadrangularis*, *P. cincinnata*, and *P. edulis*. This clade was poorly supported by ML analysis (42% BS) but well supported by BI (99% PP). Additionally, *P. edulis* was sister to a clade composed of *P. quadrangularis* and *P. cincinnata*, which was also poorly supported in our ML analysis (53% BS) but well supported by BI (100% PP). All other nodes were well supported, with BS values ranging from 94 to 100%. A comparison of our phylogenetic tree of the subgenus *Passiflora* with the classification of MacDougal and Feuillet (2004) based on morphological features and growth patterns reveals that the following groups were not monophyletic (Fig. 2): super-section *Passiflora* and series *Passiflora* (red squares; Fig. 2b), super-section *Laurifolia* (triangles, Fig. 2b), super-section *Stipulata*, and section *Granadillastrum* (circles; Fig. 2b). On the other hand, the series *Laurifoliae* (two species; purple

triangle; Fig. 2b) and *Tilifolia* (three species; green triangle; Fig. 2b) from the super-section *Laurifolia* formed monophyletic groups in our phylogeny tree.

Lastly, we used concatenated genes to further investigate the relationships that were poorly supported in our tree. To this end, we carried out a phylogenomic analysis of the subgenus *Passiflora* based on whole plastomes. ML and BI analyses generated a consensus tree with lnL values of -295,564.9791 and -296,162.7267, respectively (Supplementary Figs. S7 and S8). The relationships visualized in both phylogenomic trees (ML and BI) confirmed the arrangements inferred in our phylogeny based on plastid genes (Fig. 2). The only exceptions were the positions of *P. edulis*, *P. quadrangularis*, and *P. cincinnata* (red-dashed box; Supplementary Figs. S7 and S8). Based on plastid genes, *P. quadrangularis* was sister to *P. cincinnata* (using both ML and BI), whereas *P. nitida* and *P. laurifolia* formed a sister clade to *P. quadrangularis* with high support values (96% BS and 100% PP; Supplementary Figs. S7 and S8). In the Bayesian phylogenomic tree (Supplementary Fig.

S8), *P. edulis* was sister to *P. cincinnata* (99% PP), and both these species formed a sister group to the clade containing *P. nitida*, *P. laurifolia*, and *P. quadrangularis* (100% PP). However, in the ML phylogenomic tree (Supplementary Fig. S7), *P. nitida*, *P. laurifolia*, and *P. quadrangularis* formed a sister clade to *P. cincinnata* with low support (34% BS), and all four species formed a sister group to *P. edulis*.

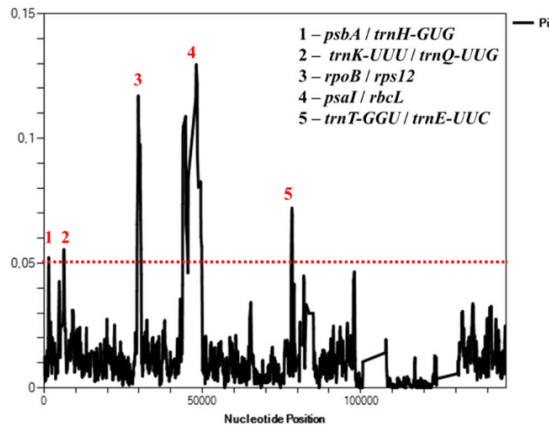


Fig. 3 Sliding window analysis of aligned whole plastomes of the subgenus *Passiflora*. The regions with high nucleotide variability ($P_i > 0.05$) are indicated. P_i , nucleotide diversity of each window. Window length, 200 nt; step size, 50 nt

Hotspots of nucleotide diversity in the plastomes of the subgenus *Passiflora*

To identify fast-evolving plastid sequences within the subgenus *Passiflora*, we performed a sliding window analysis using whole plastomes. IGSs between *psbA* and *trnH-GUG*, *trnK-UUU* and *trnQ-UUG*, and *trnT-GGU* and *trnE-UUC* showed high nucleotide variability ($P_i > 0.05$), but the highest peaks of nucleotide polymorphism were located between *rpoB* and *rps12* ($P_i > 0.1$) and *psal* and *rbcL* ($P_i > 0.15$) (Fig. 3).

Considering that *rpoB/rps12* and *psal/rbcL* IGSs enwrap the coding sequences (CDS) of *clpP* and *accD*, respectively, we used a narrower window that enabled us to identify the most divergent sequences within these hotspots. The *clpP* gene displayed numerous peaks of nucleotide polymorphism, which contrasted with the conserved nucleotides in the surrounding sequences of *rpoB* and *rps12* and even in IGSs (Fig. 4a and Supplementary Fig. S9). Similarly, the *accD* gene exhibited several peaks of nucleotide polymorphism, whereas the surrounding *psal* and *rbcL* genes were highly conserved (Fig. 4b and Supplementary Fig. S10). Although the promoter sequence and most of the *accD* gene were highly divergent, the C-terminus of the gene and the 3'-untranslated region appeared highly conserved, as reported by Rabah et al. (2019). In spite of high nucleotide variability in the promoter region and N-terminus of the *accD* gene, we successfully mapped two putative plastome start codons of *accD* that were conserved in all species of

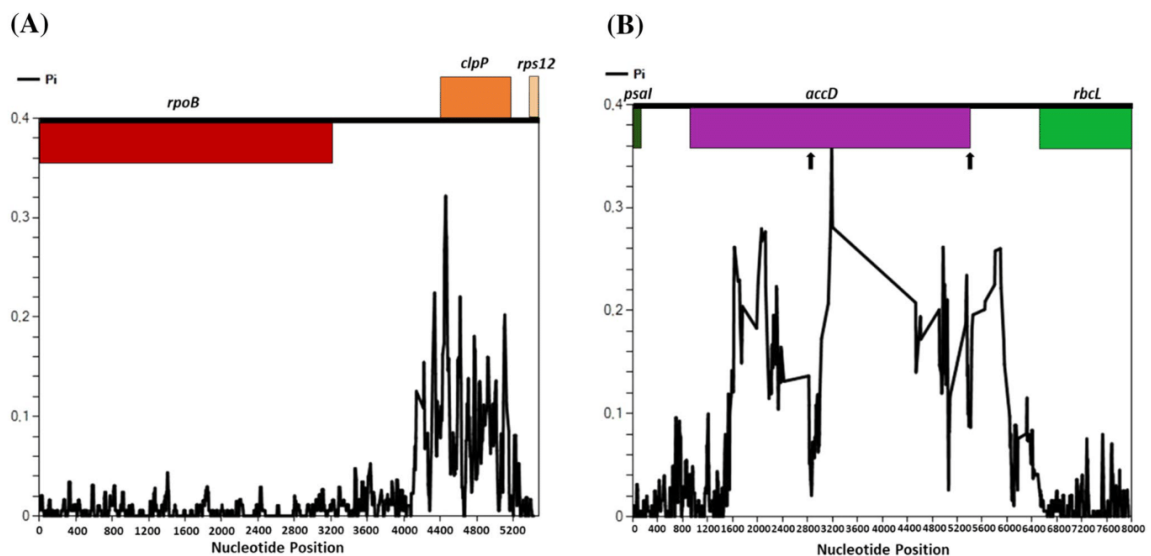


Fig. 4 Sliding window analysis of the subgenus *Passiflora* on selected plastome regions. **a** Plastome region including the *rpoB*, *clpP*, and *rps12* (first exon) genes. **b** Plastome region including the *psal*, *accD*,

and *rbcL* genes. The arrows indicate the position of the two putative start codons located in the *accD* gene. P_i , nucleotide diversity of each window; window length, 20 nt; step size, 5 nt

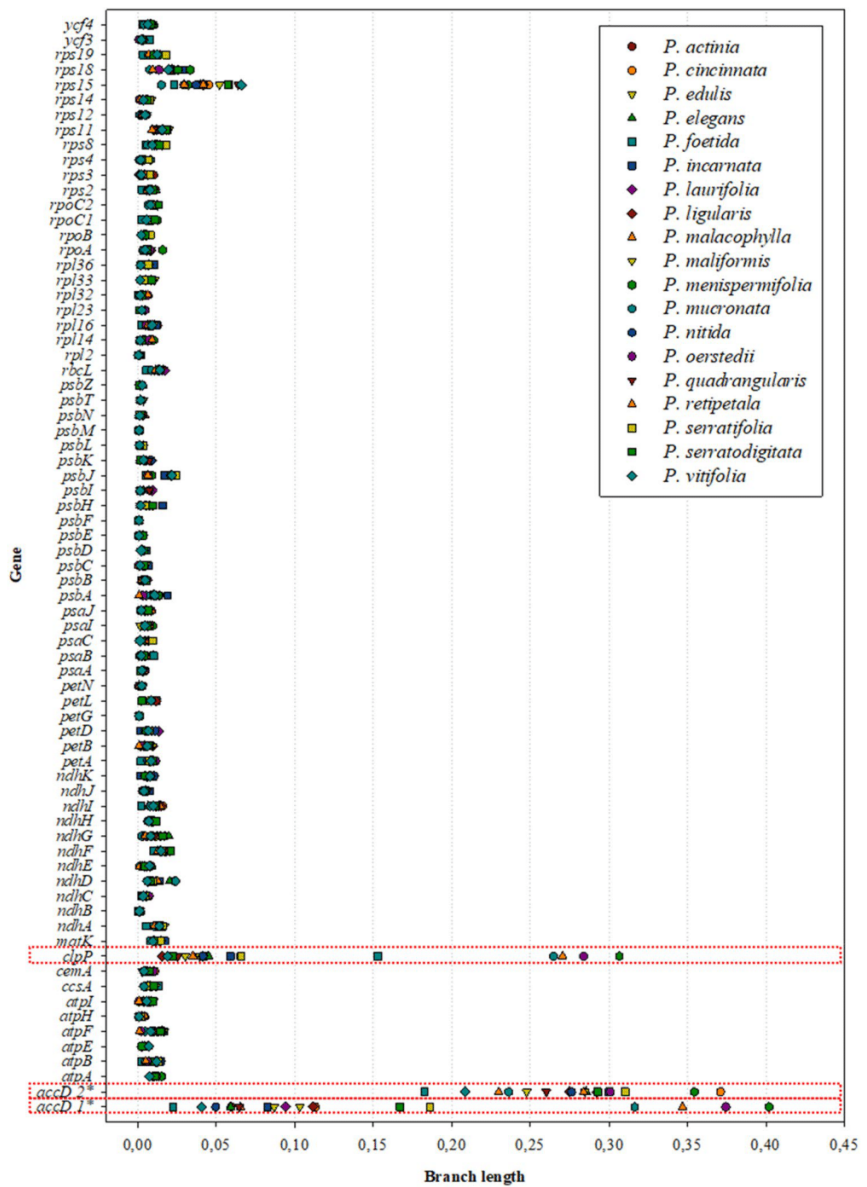
the subgenus *Passiflora* sampled. They were located in small conserved sites between large sequences characterized by elevated nucleotide variability. Valleys in the sequences are indicated by two arrows in Fig. 4b, and the positions of the putative start codons are shown in Supplementary Fig. S10. The first putative start codon would result in short CDS, ranging from 1071 bp in *P. foetida* to 1896 bp in *P. menispermifolia*. The second putative start codon would result in large CDS, ranging from 1620 bp in *P. foetida* to 4236 bp in *P. menispermifolia*. In comparison, the size of the CDS of the *accD* gene is 1497 bp in *P. trichocarpa* (NC_009143.1),

1467 bp in *Arabidopsis thaliana* (NC_000932.1), and 1539 bp in *N. tabacum* (NC_001879.2).

Gene divergence and RNA editing sites in protein-coding genes of the subgenus *Passiflora*

To study plastid gene evolution within the subgenus *Passiflora*, we inferred gene divergence trees of the 72 protein-coding genes based on phylogenetic trees (Fig. 5). Given the existence of two putative start codons within the subgenus *Passiflora*, we considered both possible CDS (*accD 1**

Fig. 5 Divergence of the plastid protein-coding genes among species of the subgenus *Passiflora*. The gene divergence was estimated by the sum of total branch lengths in each gene tree inferred. *accD 1**: *accD* gene starting from the putative start codon 1 (shorter CDS); *accD 2**: *accD* gene starting from the putative start codon 2 (larger CDS)



and *accD 2**; Fig. 5), thus totaling 73 gene sequences. As pointed out by the sliding window analysis, *accD* (irrespective of start codon) and *clpP* were by far the most divergent genes (Fig. 5). Both putative *accD* CDS demonstrated highly variable branch lengths between the species sampled here, with *accD 2** being the most divergent. The *clpP* gene showed similar high variability and segregation into three groups: the first, with large branch lengths and composed of *P. mucronata*, *P. retipetala*, *P. oerstedii*, and *P. menispermiifolia*; the second composed only of *P. foetida* and with intermediate branch lengths; and the third containing relatively short branch lengths and formed by the remaining species. Regarding *accD 1** (Fig. 5), the clade composed of *P. mucronata*, *P. retipetala*, *P. oerstedii*, and *P. menispermiifolia* segregated equally to the group with very large branch lengths. The remaining genes, showing branch length values shorter than 0.018, were grouped according to their function (Supplementary Figs. S11–13).

Next, we described the putative RNA editing sites in protein-coding genes. All RNA editing sites were predicted to occur in the first or second positions of the codons, and all nucleotide changes observed here were from cytidine to uridine. We identified 70 putative RNA editing sites in 21 genes, of which 42 were shared by all species of the subgenus *Passiflora* (Supplementary Table S5). Fourteen sites predicted to be lost in one or more species are highlighted by colored triangles in Fig. 6. Among them, 12 sites were putatively lost owing to the fixation of thymine in the DNA

sequence, which restored the conserved amino acid and eliminated the need for RNA editing. The other two sites, *matK* (190) and *ndhD* (369), were located in non-conserved regions of the protein. In the case of *matK*, for most species the putative RNA editing site was CCA (P) > TCA (S), except for the plastomes of *P. malacophylla* and *P. actinia*, whereby they were ACG (T) and CGG (R), respectively. Similarly, *ndhD* was predicted to be edited as ACT (T) > ATC (I) in all species, except *P. cincinnata*, where it was AAC (N). Conversely, 14 sites were putatively acquired by one or more species and are highlighted by a colored square in Fig. 6. Among these, 13 sites were likely acquired by mutational events that replaced thymine with cytosine in the DNA sequence, making RNA editing necessary to restore the conserved amino acids. The last site, *rpoC1* (681), was only predicted in the plastome of *P. maliformis*, where the codon CCA (P) was putatively edited to TCA (S), whereas all other species displayed the codon CAA (Q), which was not predicted to be edited (Supplementary Table S5).

We plotted the gained and lost RNA editing sites against our phylogenetic tree of the subgenus *Passiflora* (Fig. 6). We also included the insertion of six nucleotides of the third exon of *rps12* identified in some *Passiflora* species (Supplementary Fig. S14). This insertion was absent from the plastome of *Populus* but seemed to originate from a duplication event of the stretch 5'-AAAAAG-3' and was found here in *Adenia* and most *Passiflora* species analyzed, except

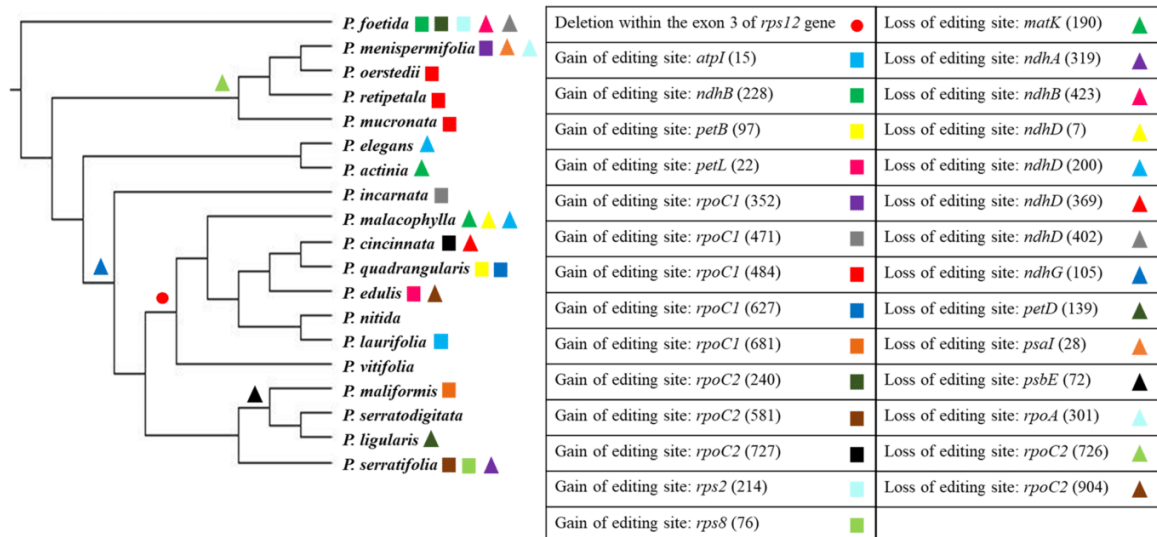


Fig. 6 Differences in the predicted RNA editing sites and *rps12* gene sequence among species of the subgenus *Passiflora* are plotted against the phylogenetic tree of the subgenus *Passiflora* (Fig. 2). The predicted changes in the RNA editing include gains (colorful squares) and losses of sites (colorful triangles). The numbers between paren-

theses indicate the amino acid position of the referred RNA editing sites. The red circle indicates the clade that lost an insertion of six nucleotides (5'-AAAAAG-3') within the exon 3 of the *rps12* gene that is present in all other species of the genus *Passiflora* (Supplementary Figure S14)

for *P. vitifolia*, *P. malacophylla*, *P. laurifolia*, *P. nitida*, *P. edulis*, *P. quadrangularis*, and *P. cincinnata* of the subgenus *Passiflora*. Hence, we suggest that this insertion originated in the common ancestor of *Adenia* and the genus *Passiflora* but was likely lost in the common ancestor of the above seven species (red circle, Fig. 6). Concerning the distribution of RNA editing sites, most putative gains and losses were species-specific. However, some of the predicted losses likely occurred in the common ancestor of some species, such as the sites of the genes *rpoC2* (726), *ndhG* (105), and *psbE* (72) (Fig. 6). The RNA editing site *matK* (190) seems to have been lost at least twice within the subgenus *Passiflora* given that it was shared by *P. actinia* and *P. malacophylla* (green triangle, Fig. 6). The same seems to be the case for the RNA editing site *ndhD* (200), which was lost in the plastome of *P. elegans* and *P. malacophylla* (light blue triangle, Fig. 6). Furthermore, the putative gain of the RNA editing site in *rpoC1* (484) (red square; Fig. 6) was shared by *P. oerstedii*, *P. reitpetala*, and *P. mucronata* but did not appear in the plastome of *P. menispermifolia*. Therefore, we suggest that this site was acquired by the common ancestor of these four species but was posteriorly lost in the plastome of *P. menispermifolia*. *P. foetida* contained the highest number of specific RNA editing site events (three gains and two losses), followed by *P. serratifolia* (two gains and one loss), *P. malacophylla* (three losses), and *P. menispermifolia* (one gain and two losses) (Fig. 6).

Discussion

Plastomes of the subgenus *Passiflora* harbor hotspots of nucleotide diversity useful for genetic, breeding and conservation purposes

Hotspots of nucleotide polymorphisms in plastome sequences are suitable tools for genetic studies and the characterization of germplasm collections (Rogalski et al. 2015; Daniell et al. 2016). Our data and previous reports (Rabah et al. 2019; Shrestha et al. 2019) indicate that with the exception of *P. foetida* and *P. menispermifolia*, the plastomes of the subgenus *Passiflora* share conserved structural features and most genes. Consequently, the low variation among structural markers or plastid genes precludes their application in interspecific or intraspecific genetic studies within this subgenus. A previous study tested the efficiency of some plastid sequences (*rbcL* and *matK* CDS, *trnH*-GUG/*psbA* IGS, and the *trnL*-UAA intron) for DNA barcoding in *Passiflora*, but their low interspecific variability was insufficient to discriminate among plants at the subgenus level (Giudicelli et al. 2015). Our gene divergence analysis revealed low nucleotide diversity in *rbcL* and *matK* CDS within the subgenus *Passiflora*, and the sliding window analysis failed

to identify the *trnL*-UAA intron as a hotspot of nucleotide diversity. Only the *trnH*-GUG/*psbA* IGS was identified in our sliding window analysis as a polymorphic region of the plastome, making it a potential molecular marker. Plastid sequences, including those of the *rps4* gene, *trnL*-UAA intron, and *trnL*-UAA/*trnF*-GAA and *psbA*/*trnH*-GUG IGSs, were previously used to investigate plastid inheritance in intraspecific and interspecific crosses (Lorenz-Lemke et al. 2005; Muschner et al. 2006; Hansen et al. 2007). Although our data suggest that only the *psbA*/*trnH*-GUG IGS is suitable for this purpose, these pioneering studies nevertheless provided evidence of plastid inheritance in such crosses. Appropriate molecular markers will increase the ability to discriminate between plastid inheritance patterns at the subgenus level and direct researchers toward efficient breeding and conservation strategies.

Our analyses revealed that *accD* and *clpP* CDS as well as *trnT*-GGU/*trnE*-UUC, *trnK*-UUU/*trnQ*-UUG, and *psbA*/*trnH*-GUG IGSs showed the highest nucleotide polymorphism in the plastomes of the subgenus *Passiflora*. These sequences can be employed in studies on plastid inheritance, polymorphism in plastid DNA, genetic diversity, genetic divergence in natural populations, and characterization of germplasm collections. Thus, we suggest the use of these highly polymorphic sequences to further investigate plastid genetics and inheritance. Given that *clpP* and *accD* are flanked by conserved genes, it is highly viable to design primers for detecting polymorphism in intraspecific and interspecific crosses and use them as effective molecular markers.

Phylogenies based on plastid sequences reveal inconsistencies within the subgenus *Passiflora* and a non-monophyletic pattern of super-sections

In our phylogenetic trees based on plastid protein-coding genes and either ML or BI, the relationships between the five subgenera of *Passiflora* showed the same pattern reported by Shrestha et al. (2019). The only inconsistencies concerned the position of *P. quadrangularis*, *P. edulis*, and *P. cincinnata*. The differences between our results and those obtained by Shrestha et al. (2019) may be related to the increased number of species used here, which affected the relationships between them.

MacDougal and Feuillet (2004) classified the species belonging to the subgenus *Passiflora* into six super-sections: *Passiflora*, *Stipulata*, *Laurifolia*, *Coccinea*, *Distephana*, and *Tacsonia*. The super-sections *Passiflora* and *Laurifolia* were subdivided into four series, and the super-sections *Stipulata* and *Tacsonia* were classified into five and eleven sections, respectively (MacDougal and Feuillet 2004). Our phylogeny comprised 19 taxa of the subgenus *Passiflora*, including representative species of the super-sections *Coccinea* (one

species), *Laurifolia* (six species), *Passiflora* (five species), and *Stipulata* (seven species). Except for the super-section *Coccinea*, which was only represented by one species, all other super-sections sampled here demonstrated a non-monophyletic pattern. A previous phylogeny of the subgenus *Passiflora*, using only 10 species and based on internal transcribed spacers of nuclear ribosomal DNA, placed two species of the super-section *Stipulata* into different clades in line with our results (Ramaiya et al. 2014). Here, this super-section was paraphyletic and was represented by two sections: *Dysosmia* (*P. foetida*) and *Granadillastrum* (six species). *P. foetida* was sister to all other species of the subgenus *Passiflora* and showed unique features within this subgenus such as the presence of intact *ycf1* and *ycf2* genes, absence of the rearrangement located in the LSC region (47 kb), and specific gain and loss of RNA editing sites. The section *Granadillastrum* was paraphyletic in our phylogeny and included two different clades. The first clade composed of *P. menispermifolia*, *P. oerstedii*, *P. retipetala*, and *P. mucronata*, which shared the loss of only one putative RNA editing site and, except for *P. menispermifolia*, the gain of a specific RNA editing site. The second clade formed by *P. elegans* and *P. actinia*, which was sister to the species of the other super-sections and does not share the same pattern of RNA editing sites identified in those species of the section *Granadillastrum*.

The super-section *Laurifolia* is represented here by the series *Quadrangularis* (one species), *Laurifoliae* (two species), and *Tilifolia* (three species). *P. quadrangularis* (series *Quadrangularis*) was positioned within the clade that included two species of the super-section *Passiflora* and two species of the series *Laurifoliae* (*P. nitida* and *P. laurifolia*). The three species of the series *Tilifolia* formed a monophyletic clade and shared the loss of the RNA editing site *psbE* (72). Moreover, the super-section *Passiflora* (represented here by five species of the series *Passiflora*) was polyphyletic in our phylogeny. Accordingly, we did not predict synapomorphy for this super-section in terms of gain or loss of RNA editing sites. All species of the super-sections *Passiflora* and *Laurifolia* shared the loss of the RNA editing site predicted in the *ndhG* gene. Lastly, loss of the *rps12* insertion constituted a synapomorphy of the clade that included representatives of the super-section *Laurifolia* (series *Quadrangularis* and *Laurifoliae*) and all species of the super-section *Passiflora*, except for the plastome of *P. serratifolia*.

Taken together, our analyses of plastome evolution within the subgenus *Passiflora* strongly suggest that the super-sections *Passiflora*, *Stipulata*, and *Laurifolia* are not monophyletic groups. However, two super-sections (*Distephana* and *Tacsonia*) are absent from our tree. Thus, inclusion of species belonging to these two super-sections, as well as from super-sections and sections/series underrepresented here

(super-section *Coccinea* and series *Quadrangularis*), would be required for a more complete and accurate comparison between the present molecular phylogeny of the subgenus and the classification proposed by MacDougal and Feuillet (2004). A more complete phylogeny could also resolve some of the discrepancies discussed previously. The inconsistencies described by Rabah et al. (2019) among infrageneric relationships in the genus *Passiflora* were mainly because of taxon sampling (Shrestha et al. 2019), and we believe that this could also be the case in our phylogeny. Although we used 19 taxa of the subgenus *Passiflora* in our analyses, the exact phylogenetic position of these species will be properly resolved only by employing more complete plastome sequences, which highlights the importance of plastid genomics for phylogenetic purposes.

Do putative RNA editing sites and fast-evolving genes contribute to nuclear-cytoplasmic incompatibility during interspecific hybridization within the subgenus *Passiflora*?

Our data predicted significant variation in RNA editing sites even among species that were closely related based on phylogenetic inferences. RNA editing affects plastid gene expression and, consequently, could be involved in nuclear-cytoplasmic incompatibility in interspecific hybrids. The RNA editing machinery is fully encoded by the nucleus and composed of site-specific trans-acting factors that are required for correct recognition of particular plastid RNA editing sites (Ichinose and Sugita 2016; Shikanai 2006; Sun et al. 2016). Therefore, most specific trans-acting factors are expected to co-evolve with their respective RNA editing sites (Schmitz-Linneweber et al. 2005). This co-evolution was elegantly illustrated in Solanaceae using a cybrid containing the nucleus of *A. belladonna* and the plastome of *N. tabacum*, whereby the abnormal albino phenotype was determined by a failure to edit a specific site of the ATPase α -subunit transcript of *N. tabacum* and hence impairing photosynthesis (Schmitz-Linneweber et al. 2005). Interestingly, in our analysis of RNA editing sites, we predicted a specific site in the plastome of *P. menispermifolia* located in the *rpoC1* gene. This site was predicted to be converted from A (GCU) to V (GUU), restoring an amino acid conserved in all other species of the subgenus *Passiflora* investigated here. The *rpo* genes (*rpoA*, *rpoB*, *rpoC1*, and *rpoC2*) encode the core subunits of the plastid-encoded RNA polymerase (PEP) (Hajdukiewicz et al. 1997; Swiatecka-Hagenbruch et al. 2008; Börner et al. 2015). A non-functional PEP can lead to a variegated (heteroplasmy) and/or albino (homoplasmy) phenotype (Allison et al. 1996; Krause et al. 2000). Absence of RNA editing in *rpoC1* (352) in the plastome of *P. menispermifolia* could interfere with PEP assembly and, consequently, promote incompatibility in this species

(Mráček, 2005). The other three RNA editing sites predicted in *P. menispermifolia* and *P. oerstedii* (*rpoA*, *rpoC1*, and *psal*) corresponded to losses in the plastome of *P. menispermifolia*. Additional species-specific differences related to the presence or absence of RNA editing sites were identified here, representing putative candidates for incompatibility (Conceição et al. 2011; Ocampo et al. 2016). Lack of RNA editing at these sites could impair plastid gene expression (*rpoA*, *rpoC1*, and *rpoC2*), plastid translation (*rps2* and *rps8*), photosynthesis (*petB*, *petD*, *petL*, and *psbE*), and chlororespiration (*ndhA*, *ndhB*, *ndhD*, and *ndhG*). Among them, we found several essential genes that could lead to embryonic lethality; these include *rps2*, which is essential for cell viability in tobacco (Rogalski et al. 2008b), and *rps8*, which has not been analyzed in plants but is essential in *Escherichia coli* (Shoji et al. 2011). Recently, a clear example emerged from analyzing a specific RNA editing site in the *rps14* gene of *Arabidopsis* (Sun et al. 2018).

We propose *accD* and *clpP* as additional mediators of nucleus-plastome incompatibility. According to our data, the main differences between plastomes of the subgenus *Passiflora* were concentrated in these two genes, which were identified as hotspots of nucleotide divergence. The *accD* gene is essential for cell viability (Kode et al. 2005). It encodes the β -carboxyl transferase subunit of the heteromeric acetyl-CoA carboxylase (ACCase), which catalyzes the first committed step for de novo fatty acid biosynthesis in plastids (Rogalski and Carrer 2011; Salie and Thelen 2016). Rabah et al. (2019) also reported a highly divergent *accD* gene in the genus *Passiflora* marked by the presence of tandem repeats within its sequence. We detected also several insertions composed of repetitive amino acid sequences, which split the *accD* domains. The insertions were highly variable in terms of amino acid residues and size, resulting in widely divergent lengths of the predicted AccD protein (from 535 amino acids in *P. foetida* to 1122 amino acids in *P. menispermifolia*). The *clpP* gene is also essential for cell viability and encodes a proteolytic subunit of ATP-dependent protease (Clp), a large proteolytic complex involved in plastid protein homeostasis (Shikanai et al. 2001; Kuroda and Maliga 2003; Moreno et al. 2018). The elevated divergence of the *clpP* gene observed here is in line with the high substitution rate in this gene within the genus *Passiflora* detected by Shrestha et al. (2019). Both *clpP* and *accD* genes encode plastid subunits for multi-subunit protein complexes formed by nuclear and plastid gene products, which have to co-evolve harmoniously to allow for the correct assembly of subunits and the building of a functional enzymatic complex (Greiner and Bock 2013). In addition, *accD* and *clpP* genes displayed high divergence among species of the genus *Silene*, which was accompanied by rapid evolution of nuclear genes encoding the other subunits of the multi-subunit ACCase and Clp enzymatic complexes (Rockenbach et al. 2016). There is a

strong likelihood of an accelerated co-evolution between highly divergent *accD* and *clpP* plastid genes in *Passiflora* species and their counterparts encoded by the nucleus to build functional enzymatic complexes. Failure to assemble these protein complexes would impair fatty acid biosynthesis and plastid protein homeostasis, leading to cell death (Shikanai et al. 2001; Kuroda and Maliga 2003; Kode et al. 2005). Bogdanova et al. (2015) attributed the failure to obtain plantlets in crosses between wild and domesticated accessions of peas to the high divergence between nuclear and plastid-encoded components of the heteromeric ACCase.

Analysis and characterization of genes that encode the nuclear subunits of ACCase/Clp and experimental validation of the RNA editing sites predicted here will help to determine their involvement in nucleus-plastid incompatibility during interspecific hybridization of *Passiflora*. However, other possible players such as the promoter sequence cannot be discarded given that a complex interplay between nuclear-encoded and plastid-encoded RNA polymerases governs plastid transcription. In the genus *Oenothera*, deletion in a promoter sequence underlined the incompatibility observed in interspecific crosses (Greiner et al. 2008; Greiner and Bock 2013).

Conclusions

Here, we report the complete plastome sequences of six species belonging to the subgenus *Passiflora*. Gene divergence and sliding window analyses revealed that *accD* and *clpP* were the most divergent sequences in 19 species of the subgenus *Passiflora* characterized here. Additionally, we identified three intergenic regions as hotspots of nucleotide divergence, which could foster future genetic analyses. Our phylogenetic inferences based on plastid genes of the genus *Passiflora* were congruent with recent studies. Furthermore, the subgenus *Passiflora* formed a monophyletic and well-resolved clade but exhibited low support for the relationships between *P. quadrangularis*, *P. edulis*, and *P. cincinnata*. The phylogenetic position of these species was inconsistent between our tree based on plastid genes and the phylogenomic tree based on whole-plastomes. A larger taxon sample will resolve these incongruities. The super-sections *Passiflora*, *Stipulata*, and *Laurifolia* formed non-monophyletic groups in our phylogeny of the subgenus *Passiflora*. However, two super-sections proposed for this subgenus were absent from our phylogeny, and the super-section *Coccinea* was only represented by one taxon, suggesting that caution should be employed when interpreting the present results. Moreover, our RNA editing prediction revealed the presence of species-specific RNA editing sites within the subgenus *Passiflora*. Finally, the data allow us to speculate that the predicted diversity of RNA editing

sites and the high divergence of the *accD* and *clpP* genes are putative candidates for nuclear-plastid incompatibility observed in interspecific hybrids. However, rapidly evolving genes and divergent RNA editing sites have yet to be individually analyzed and experimentally confirmed. Taken together, our data point to new molecular markers and evolutionary features applicable to genetic, breeding, and conservation studies of the subgenus *Passiflora*.

Acknowledgements This research was supported by the National Council for Scientific and Technological Development, Brazil (CNPq—Grants 459698/2014-1, 310654/2018-1 and 436407/2018-3). We are grateful to INCT-FBN and for the scholarships granted by the CNPq to ASL, TGP, LNV, WCO, MPG, RON, EB, FOP, EMS, and MR. We are also grateful to the Núcleo de Análise de Biomoléculas (NuBiomol) of the Universidade Federal de Viçosa for providing the CLC Genomics software.

Author contributions TGP, ASL, EMS, FOP, and MR conceived and designed the research. TGP, ASL, JFW, KSCY, EB, EMS, LNV, WCO, and MR conducted experiments and analyzed the data. EMS, FOP, MPG, RON, and MR contributed with reagents and materials. TGP, ASL, WCO, and MR wrote the manuscript. All authors read and approved the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Abreu PP, Souza MM, Santos EA, Pires MV, Pires MM, de Almeida AAF (2009) Passion flower hybrids and their use in the ornamental plant market: perspectives for sustainable development with emphasis on Brazil. *Euphytica* 166:307–315. <https://doi.org/10.1007/s10681-008-9835-x>
- Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S, Schöttler MA, Ruf S, Bock R (2012) The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet* 8:e1003076. <https://doi.org/10.1371/journal.pgen.1003076>
- Allison LA, Simon LD, Maliga P (1996) Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. *EMBO J* 15:2802–2809
- Besnard G, Hernández P, Khadari B, Dorado G, Savolainen V (2011) Genomic profiling of plastid DNA variation in the Mediterranean olive tree. *BMC Plant Biol* 11:1. <https://doi.org/10.1186/1471-2229-11-80>
- Bogdanova VS, Zaytseva OO, Mglinets AV, Shatskaya NV, Kosterin OE, Vasiliev GV (2015) Nuclear-cytoplasmic conflict in pea (*Pisum sativum* L.) is associated with nuclear and plastidic candidate genes encoding acetyl-CoA carboxylase subunits. *PLoS ONE* 10:e0119835. <https://doi.org/10.1371/journal.pone.0119835>
- Börner T, Aleynikova AY, Zubo YO, Kusnetsov VV (2015) Chloroplast RNA polymerases: role in chloroplast biogenesis. *Biochim Biophys Acta* 1847:761–769. <https://doi.org/10.1016/j.bbabi.2015.02.004>
- Bugallo V, Cardone SGT, Pannunzio MJ, Facciuto GR (2011) Breeding advances in *Passiflora* spp. (Passionflower) native to Argentina. *Floriculture Ornamental Biotech* 5:23–34
- Cauz-Santos LA, Munhoz CF, Rodde N, Cauet S, Santos AA, Penha HA, Dornelas MC, Varani AM, Oliveira GCX, Bergès H, 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* 8:1–17. <https://doi.org/10.3389/fpls.2017.00334>
- Cerqueira-Silva CBM, Santos ESL, Jesus ON, Mori GM, Jesus ON, Corrêa RX, Souza AP (2014a) Molecular genetic variability of commercial and wild accessions of passion fruit (*Passiflora* spp.) targeting ex situ conservation and breeding. *Int J Mol Sci* 15:22933–22959. <https://doi.org/10.3390/ijms151222933>
- Cerqueira-Silva CBM, Santos ESL, Jesus GM, Jesus ON, Corrêa RX, Souza AP (2014b) Genetic breeding and diversity of the genus *passiflora*: progress and perspectives in molecular and genetic studies. *Int J Mol Sci* 15:14122–14152. <https://doi.org/10.3390/ijms150814122>
- Cerqueira-Silva CBM, Santos ESL, Vieira JGP, Mori GM, Jesus ON, Corrêa RX, Souza AP (2014c) New microsatellite markers for wild and commercial species of *passiflora* (*Passifloraceae*) and cross-amplification. *Appl Plant Sci* 2:1300061. <https://doi.org/10.1590/S1516-89132013000500009>
- Chitwood DH, Otoni WC (2017a) Divergent leaf shapes among *Passiflora* species arise from a shared juvenile morphology. *Plant Direct* 1:e00028. <https://doi.org/10.1002/pld3.28>
- Chitwood DH, Otoni WC (2017b) Morphometric analysis of *Passiflora* leaves: the relationship between landmarks of the vasculature and elliptical Fourier descriptors of the blade. *Gigascience* 6:1–13. <https://doi.org/10.1093/gigascience/giw008>
- Conceição LDHCS, Souza MM, Belo GO, Santos SF, Freitas JCO (2011) Hybridization among wild passionflower species. *Braz J Bot* 34:237–240. <https://doi.org/10.1590/S0100-84042011000200011>
- Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* 17:134. <https://doi.org/10.1186/s13059-016-1004-2>
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772. <https://doi.org/10.1038/nmeth.2109>
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Giudicelli GC, Mäder G, de Freitas LB (2015) Efficiency of ITS sequences for DNA barcoding in *Passiflora* (Passifloraceae). *Int J Mol Sci* 16:7289–7303. <https://doi.org/10.3390/ijms16047289>
- Greiner S, Bock R (2013) Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* 35:354–365. <https://doi.org/10.1002/bies.201200137>
- Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* 47:W59–W64. <https://doi.org/10.1093/nar/gkz238>
- Greiner S, Wang X, Herrmann RG, Rauwolf U, Mayer K, Haberer G, Meurer J (2008) The complete nucleotide sequences of the 5 genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: II. A microevolutionary view using bioinformatics and formal genetic data. *Mol Biol Evol* 25:2019–2030. <https://doi.org/10.1093/molbev/msn149>
- Hajdukiewicz PT, Allison LA, Maliga P (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J* 16:4041–4048. <https://doi.org/10.1093/emboj/16.13.4041>
- Hansen AK, Escobar LK, Gilbert LE, Jansen RK (2007) Paternal, maternal, and biparental inheritance of the chloroplast genome

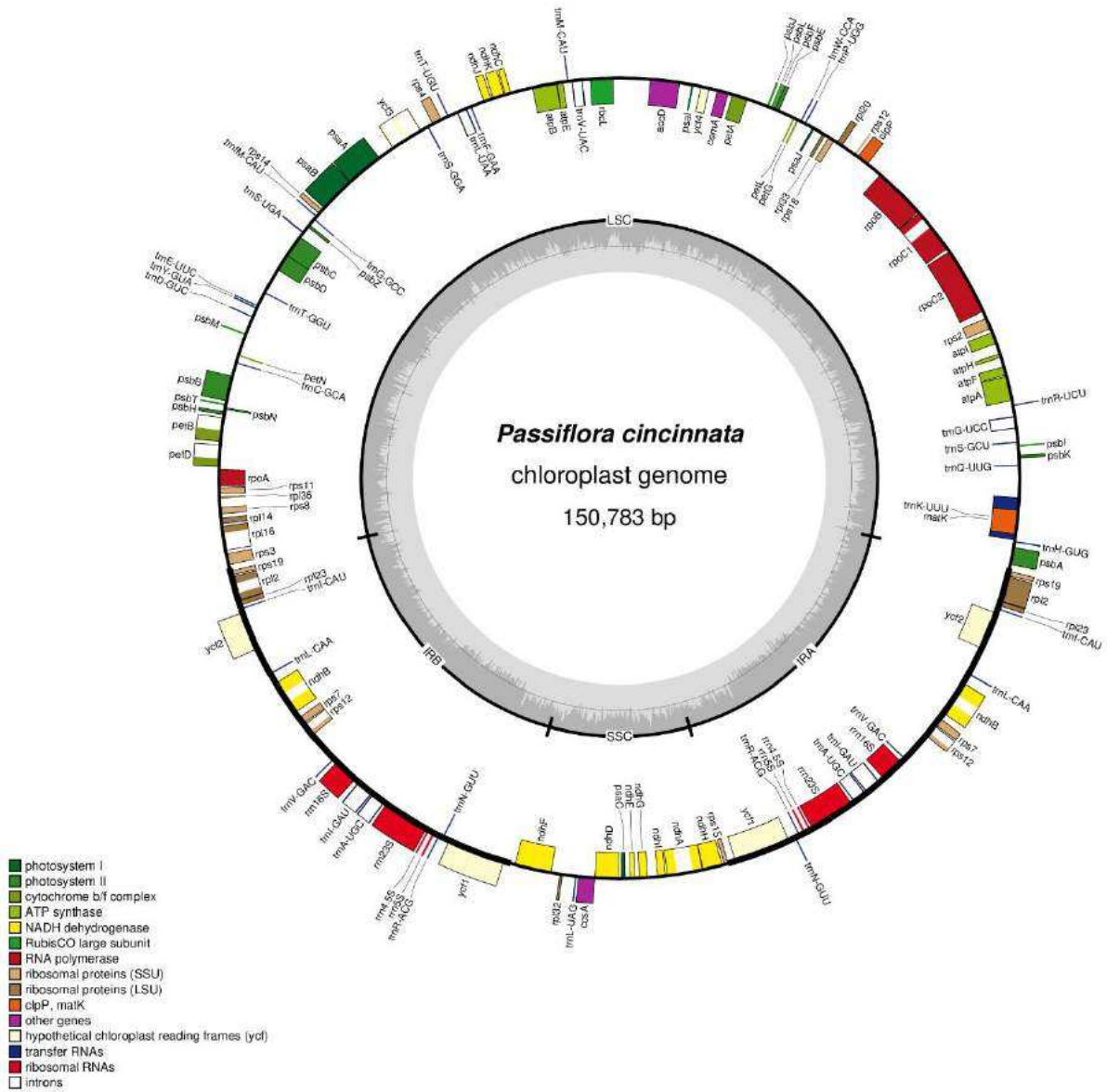
- in *Passiflora* (Passifloraceae): implications for phylogenetic studies. *Am J Bot* 94:42–46. <https://doi.org/10.3732/ajb.94.1.42>
- Hansen AK, Gilbert LE, Simpson BB, Downie SR, Cervi AC, Jansen RK (2006) Phylogenetic relationships and chromosome number evolution in *Passiflora*. *Syst Bot* 31:138–150. <https://doi.org/10.1600/036364406775971769>
- Ichinose M, Sugita M (2016) RNA editing and its molecular mechanism in plant organelles. *Genes*. <https://doi.org/10.3390/genes8010005>
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
- Killip EW (1938) The American species of Passifloraceae. Field Museum of Natural History, Chicago, IL
- Kode V, Mudd EA, Iamtham S, Day A (2005) The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant J* 44:237–244. <https://doi.org/10.1111/j.1365-3113.2005.02533.x>
- Krause K, Maier RM, Kofer W, Krupinska K, Herrmann RG (2000) Disruption of plastid-encoded RNA polymerase genes in tobacco: expression of only a distinct set of genes is not based on selective transcription of the plastid chromosome. *Mol Gen Genet* 263:1022–1030. <https://doi.org/10.1007/pl00008690>
- Krosnick SE, Ford AJ, Freudenstein JV (2009) Taxonomic revision of *Passiflora* subgenus *Tetrapathea* including the monotypic genera *Hollrungia* and *Tetrapathea* (Passifloraceae), and a new species of *Passiflora*. *Syst Bot* 34:375–385. <https://doi.org/10.1600/036364409788606343>
- Krosnick SE, Porter-Utley KE, MacDougal JM, Jørgensen PM, McDade LA (2013) New insights into the evolution of *Passiflora* subgenus *Decaloba* (Passifloraceae): phylogenetic relationships and morphological synapomorphies. *Syst Bot* 38:692–713. <https://doi.org/10.1600/036364413X670359>
- Kuroda H, Maliga P (2003) The plastid *clpP1* protease gene is essential for plant development. *Nature* 425:86–89. <https://doi.org/10.1038/nature01909>
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Brett C (2017) PartitionFinder 2: new methods for selecting partitioned models for evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol* 34:772–773. <https://doi.org/10.1093/molbev/msw260>
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>
- Lopes AS, Pacheco TG, Santos KGD, Vieira LN, Guerra MP, Nodari RO, de Souza EM, Pedrosa FO, Rogalski M (2018a) The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Rep* 37:307–328. <https://doi.org/10.1007/s00299-017-2231-z>
- Lopes AS, Pacheco TG, Silva ON, Cruz LM, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2019) The plastomes of *Astrocaryum aculeatum* G. Mey. and *A. murumuru* Mart. show a flip-flop recombination between two short inverted repeats. *Planta* 250:1229–1246. <https://doi.org/10.1007/s00425-019-03217>
- Lopes AS, Pacheco TG, Vieira LN, Guerra MP, Nodari RO, de Souza EM, Pedrosa FO, Rogalski M (2018b) The *Crambe abyssinica* plastome: Brassicaceae phylogenomic analysis, evolution of RNA editing sites, hotspot and microsatellite characterization of the tribe Brassiceae. *Gene* 671:36–49. <https://doi.org/10.1016/j.gene.2018.05.088>
- Lorenz-Lemke AP, Muschner VC, Bonatto SL, Cervi AC, Salzano FM, Freitas LB (2005) Phylogeographic inferences concerning evolution of Brazilian *Passiflora actinia* and *P. elegans* (Passifloraceae) based on ITS (nrDNA) variation. *Ann Bot* 95:799–806. <https://doi.org/10.1093/aob/mci079>
- Lowe TM, Chan PP (2016) tRNAscan-SE on-line: search and contextual analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54–57. <https://doi.org/10.1093/nar/gkw413>
- MacDougal JM, Feuillet C (2004) Systematics. In: Ulmer T, MacDougal JM (eds) *Passiflora*, passionflowers of the world. Timber, Portland, OR, pp 27–31
- Madureira HC, Pereira TNS, Da Cunha M, Klein DE, de Oliveira MVV, de Mattos L, de Souza Filho GA (2014) Self-incompatibility in passion fruit: cellular responses in incompatible pollinations. *Biologia* 69:574–584. <https://doi.org/10.2478/s11756-014-0353-0>
- Meletti LMM, Soares-Scott MD, Bernacci LC, Passos IRS (2005) Melhoramento genético do maracujá: passado e futuro. In: Faleiro FG, Junqueira NTV, Braga MF (eds) *Maracujá: Germoplasma e Melhoramento Genético*. Embrapa Cerrados, Planaltina, Brazil, pp 55–78
- Mondin CA, Cervi AC, Moreira Gilson RP (2011) Sinopse das espécies de *Passiflora* L. (Passifloraceae) do Rio Grande do Sul. *Brasil R bras Bioc* 9:3–27
- Moreno JC, Martínez-Jaime S, Schwartzmann J, Karcher D, Tillich M, Graf A, Bock R (2018) Temporal proteomics of inducible RNAi lines of Clp protease subunits identifies putative protease substrates. *Plant Physiol* 176:1485–1508. <https://doi.org/10.1104/pp.17.01635>
- Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res* 37:W253–W259. <https://doi.org/10.1093/nar/gkp337>
- Mráček J (2005) Investigation of genome–plastome incompatibility in *Oenothera* and *Passiflora*. PhD Thesis, Ludwig Maximilians-University, Munich
- Muschner VC, Lorenz-Lemke AP, Vecchia M, Bonatto SL, Salzano FM, Freitas LB (2006) Differential organellar inheritance in *Passiflora*'s (Passifloraceae) subgenera. *Genetica* 128:449–453. <https://doi.org/10.1007/s10709-006-7726-4>
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
- Ocampo J, Arias JC, Urrea R (2016) Interspecific hybridization between cultivated and wild species of genus *Passiflora* L. *Euphytica* 209:395–408. <https://doi.org/10.1007/s10681-016-1647-9>
- Pacheco TG, Lopes AS, Viana GDM, Silva ON, Silva GM, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Otoni WC, Rogalski M (2019) Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (*Bixa orellana* L.), the Amazonian commercial species of natural dyes. *Planta* 249:563–582. <https://doi.org/10.1007/s00425-018-3023-6>
- Qiao J, Cai M, Yan G, Wang N, Li F, Chen B, Gao G, Xu K, Li J, Wu X (2016) High-throughput multiplex cpDNA resequencing clarifies the genetic diversity and genetic relationships among *Brassica napus*, *Brassica rapa* and *Brassica oleracea*. *Plant Biotechnol J* 14:409–418. <https://doi.org/10.1111/pbi.12395>
- Rabah SO, Shrestha B, Hajrah NH, Sabir MJ, Alharby HF, Sabir MJ, Alhebshi AM, Sabir JSM, Gilbert LE, Ruhlman TA, Jansen RK (2019) *Passiflora* plastome sequencing reveals widespread genomic rearrangements. *J Syst Evol* 57:1–14. <https://doi.org/10.1111/jse.12425>
- Ramaiya SD, Bujang JS, Zakaria MH (2014) Genetic diversity in *Passiflora* species assessed by morphological and ITS sequence analysis. *Sci World J* 2014:1–11. <https://doi.org/10.1155/2014/598313>

- Rocha DI, Batista DS, Faleiro FG, Rogalski M, Ribeiro LM, Mercadante-Simões MO, Yockteng R, Silva ML, Soares WS, Pinheiro MVM, Pacheco TG, Lopes AS, Viccini LF, Otoni WC (2020) *Passiflora* spp. passionfruit. In: Litz RE, Alfaro FP, Hormaza JI (Org.) Biotechnology of fruit and nut crops, 2 edn. CABI, Oxfordshire, pp 381–408
- Rockenbach K, Havird JC, Monroe JG, Triant DA, Taylor DR, Sloan DB (2016) Positive selection in rapidly evolving plastid-nuclear enzyme complexes. *Genetics* 204:1507–1522. <https://doi.org/10.1534/genetics.116.188268>
- Rogalski M, Carrer H (2011) Engineering plastid fatty acid biosynthesis to improve food quality and biofuel production in higher plants. *Plant Biotechnol J* 9:554–564. <https://doi.org/10.1111/lj.1467-7652.2011.00621.x>
- Rogalski M, Karcher D, Bock R (2008a) Superwobbling facilitates translation with reduced tRNA sets. *Nat Struct Mol Biol* 15:192–198. <https://doi.org/10.1038/nsmb.1370>
- Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Res* 34:4537–4545. <https://doi.org/10.1093/nar/gkl634>
- Rogalski M, Schottler MA, Thiele W, Schulze WX, Bock R (2008b) Rpl33, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. *Plant Cell* 20:2221–2237. <https://doi.org/10.1105/tpc.108.060392>
- Rogalski M, Vieira LN, Fraga HP, Guerra MP (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front Plant Sci* 6:586. <https://doi.org/10.3389/fpls.2015.00586>
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542. <https://doi.org/10.1093/sysbio/sys029>
- Schmitz-Linneweber C, Kushnir S, Babiychuk E, Poltnigg P, Herrmann RG, Maier RM (2005) Pigment deficiency in nightshade/tobacco cybrids is caused by the failure to edit the plastid ATPase α -subunit mRNA. *Plant Cell* 17:1815–1828. <https://doi.org/10.1105/tpc.105.032474>
- Shikanai T (2006) RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol Life Sci* 63:698–708. <https://doi.org/10.1007/s00018-005-5449-9>
- Shikanai T, Shimizu K, Ueda K, Nishimura Y, Kuroiwa T, Hashimoto T (2001) The chloroplast *clpP* gene, encoding a proteolytic subunit of ATP-dependent protease, is indispensable for chloroplast development in tobacco. *Plant Cell Physiol* 42:264–273. <https://doi.org/10.1093/pcp/pce031>
- Shoji S, Dambacher CM, Shajani Z, Williamson JR, Schultz PG (2011) Systematic chromosomal deletion of bacterial ribosomal protein genes. *J Mol Biol* 413:751–761. <https://doi.org/10.1016/j.jmb.2011.09.004>
- Shrestha B, Weng ML, Theriot EC, Gilbert LE, Ruhlman TA, Krosnick SE, Jansen RK (2019) Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. *Mol Phylogenet Evol* 138:53–64. <https://doi.org/10.1016/j.ympev.2019.05.030>
- Stefenon VM, Klabunde G, Lemos RPM, Rogalski M, Nodari RO (2019) Phylogeography of plastid DNA sequences suggests post-glacial southward demographic expansion and the existence of several glacial refugia for *Araucaria angustifolia*. *Sci Rep* 9:2752. <https://doi.org/10.1038/s41598-019-39308-w>
- Sun T, Bentolila S, Hanson MR (2016) The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci* 21(11):962–973. <https://doi.org/10.1016/j.tplants.2016.07.005>
- Sun YK, Gutmann B, Yap A, Kindgren P, Small I (2018) Editing of chloroplast rps14 by PPR editing factor EMB2261 is essential for *Arabidopsis* development. *Front Plant Sci* 9:841. <https://doi.org/10.3389/fpls.2018.00841>
- Swiatecka-Hagenbruch M, Emanuel C, Hedtke B, Liere K, Börner T (2008) Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. *Nucleic Acids Res* 36:785–792. <https://doi.org/10.1093/nar/gkm1111>
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Vieira LN, Dos Anjos KG, Faoro H, Fraga HP, Greco TM, Pedrosa FO, de Souza EM, Rogalski M, de Souza RF, Guerra MP (2016a) Phylogenetic inference and SSR characterization of tropical woody bamboos tribe Bambuseae (Poaceae: Bambusoideae) based on complete plastid genome sequences. *Curr Genet* 62:443–453. <https://doi.org/10.1007/s00294-015-0549-z>
- Vieira LN, Faoro H, Fraga HP, Rogalski M, de Souza EM, de Oliveira PF, Nodari RO, Guerra MP (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. *PLoS ONE* 9:e84792. <https://doi.org/10.1371/journal.pone.0084792>
- Vieira LN, Rogalski M, Faoro H, Fraga HP, Anjos KG, Picchi GFA, Nodari RO, Pedrosa FO, Souza EM, Guerra MP (2016b) The plastome sequence of the endemic Amazonian conifer, *Retrophillum piresii* (Silba) C.N.Page, reveals different recombination events and plastome isoforms. *Tree Genet Genomes* 12:10. <https://doi.org/10.1007/s11295-016-0968-0>
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* 76:273–297. <https://doi.org/10.1007/s11103-011-9762-4>
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255. <https://doi.org/10.1093/bioinformatics/bth352>
- Yockteng R, Deeckenbrugge GC, Souza-Chies TT (2011) *Passiflora*. In: Kole C (ed) Wild crop relatives: genomic and breeding resources. Springer, Berlin, pp 129–171
- Zhu A, Guo W, Gupta S, Fan W, Mower JP (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol* 209:1747–1756. <https://doi.org/10.1111/nph.13743>

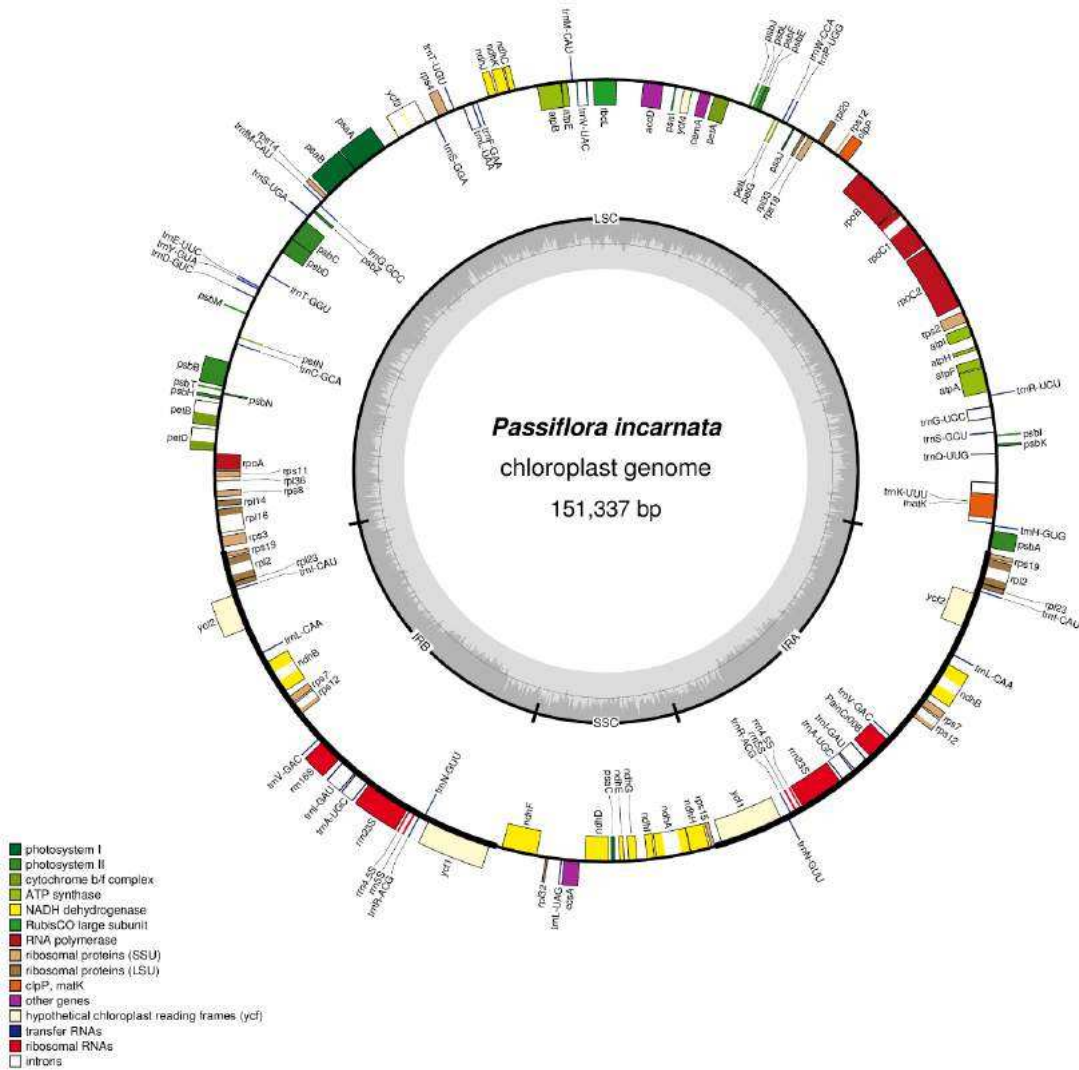
Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Material

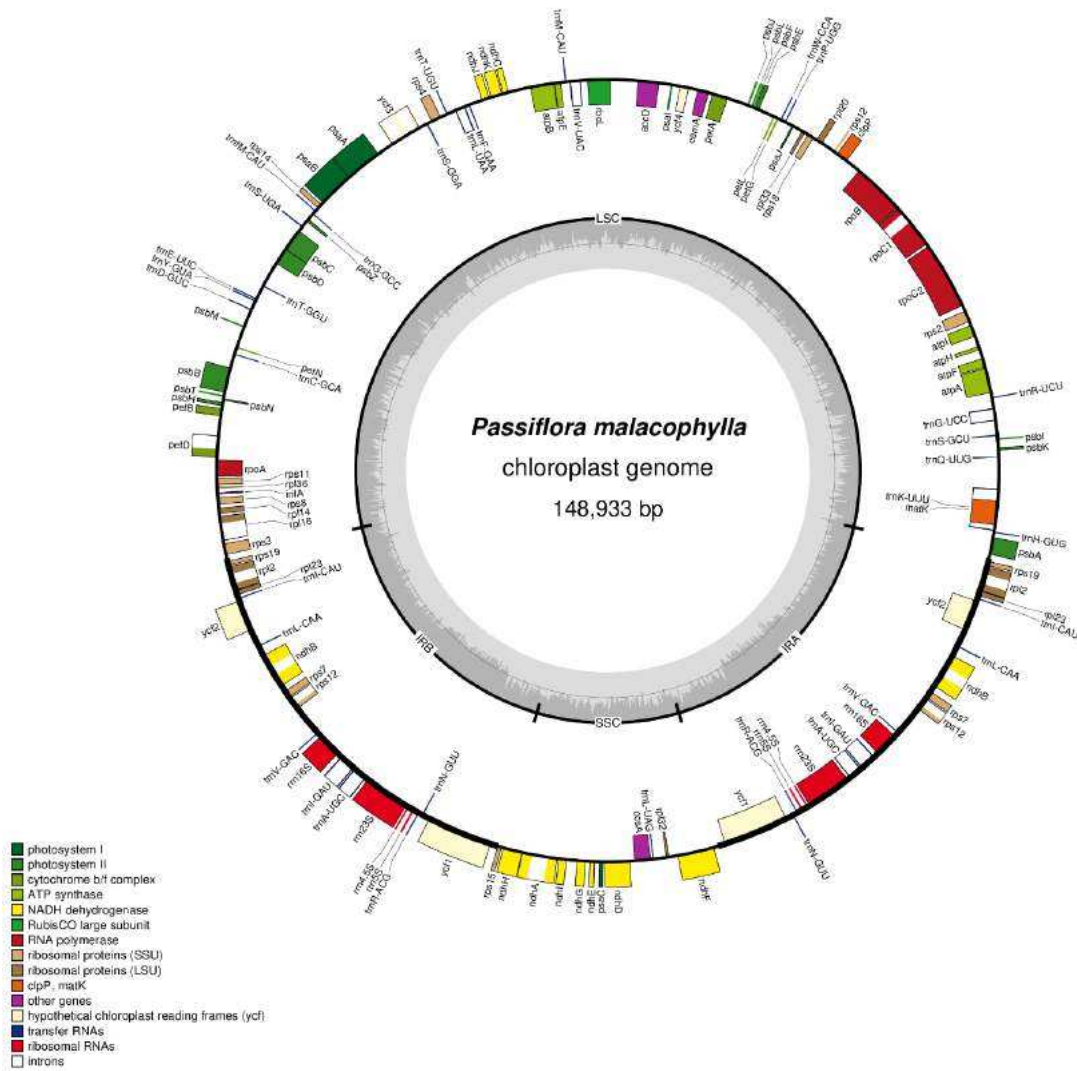
Supplementary Figures



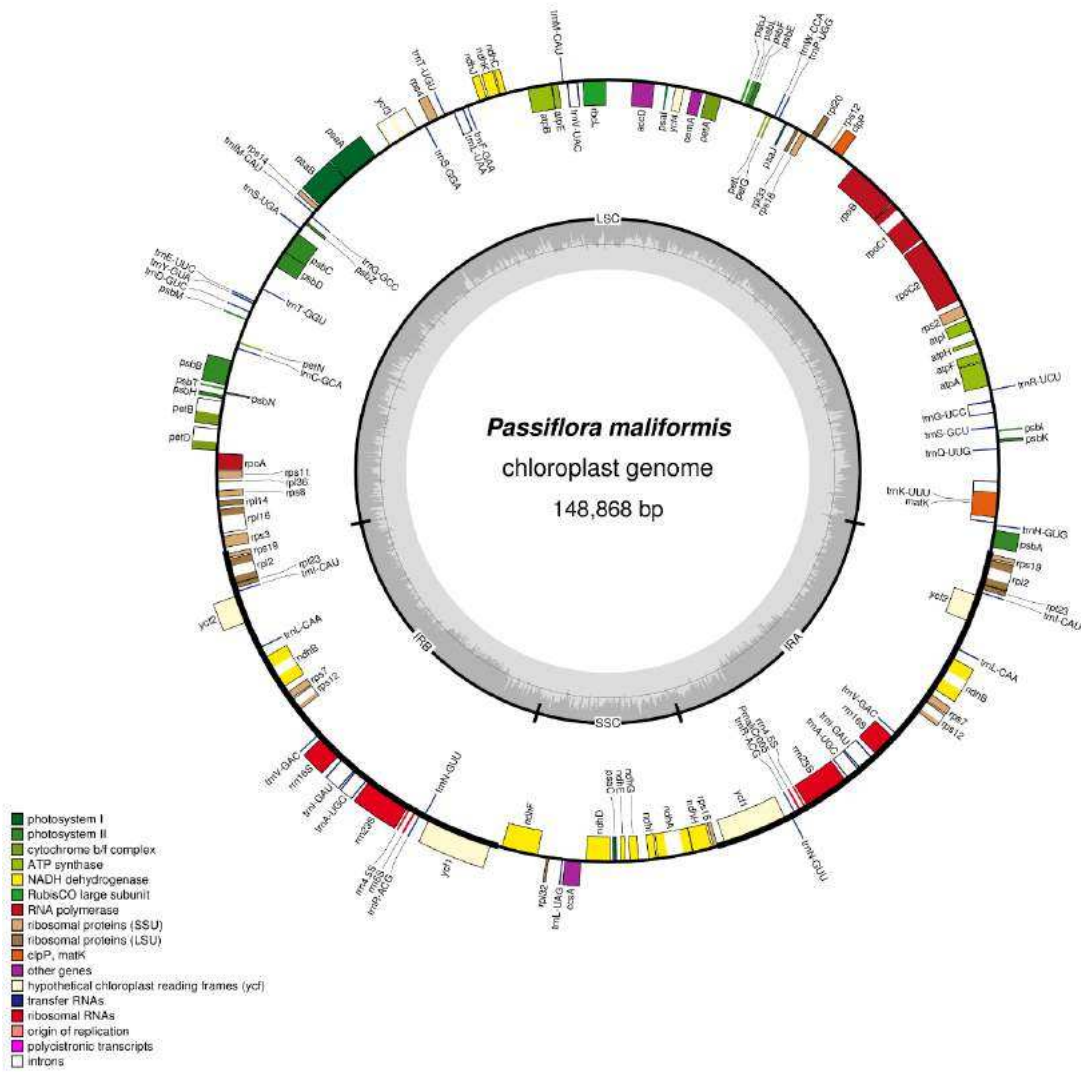
Supplementary Fig. S1 Gene map of *Passiflora cincinnata* plastome. Genes drawn inside the circle are transcribed clockwise and genes drawn outside are transcribed counterclockwise. Different functional groups of genes are color-coded. The darker gray in the inner circle corresponds to GC content and the lighter gray corresponds to AT content. LSC, Large Single Copy; SSC, Small Single Copy; IRA/B, Inverted Repeat A/B



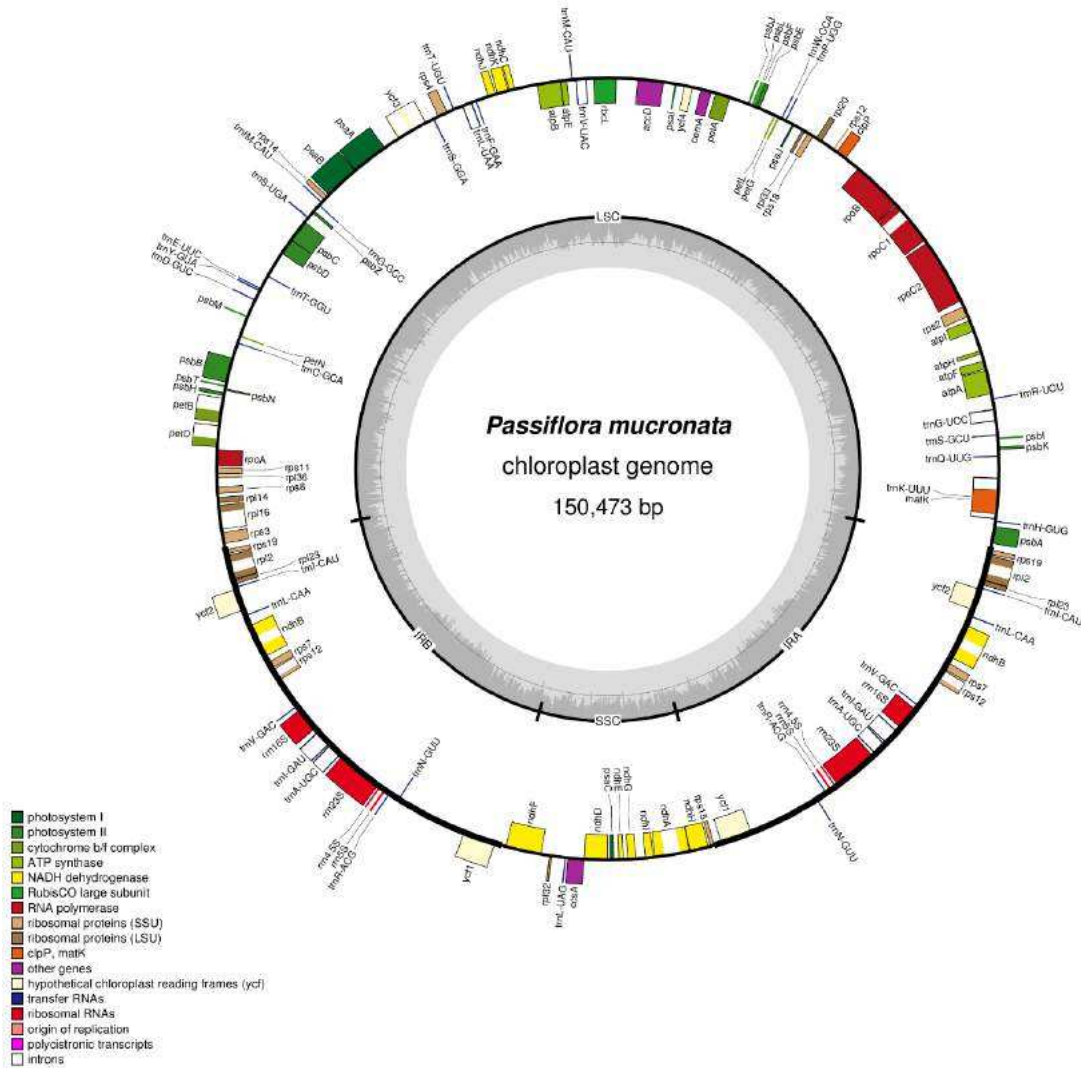
Supplementary Fig. S2 Gene map of *Passiflora incarnata* plastome. Genes drawn inside the circle are transcribed clockwise direction and genes drawn outside are transcribed counterclockwise. Different functional groups of genes are color-coded. The darker gray in the inner circle corresponds to GC content and the lighter gray corresponds to AT content. LSC, Large Single Copy; SSC, Small Single Copy; IRA/B, Inverted Repeat A/B



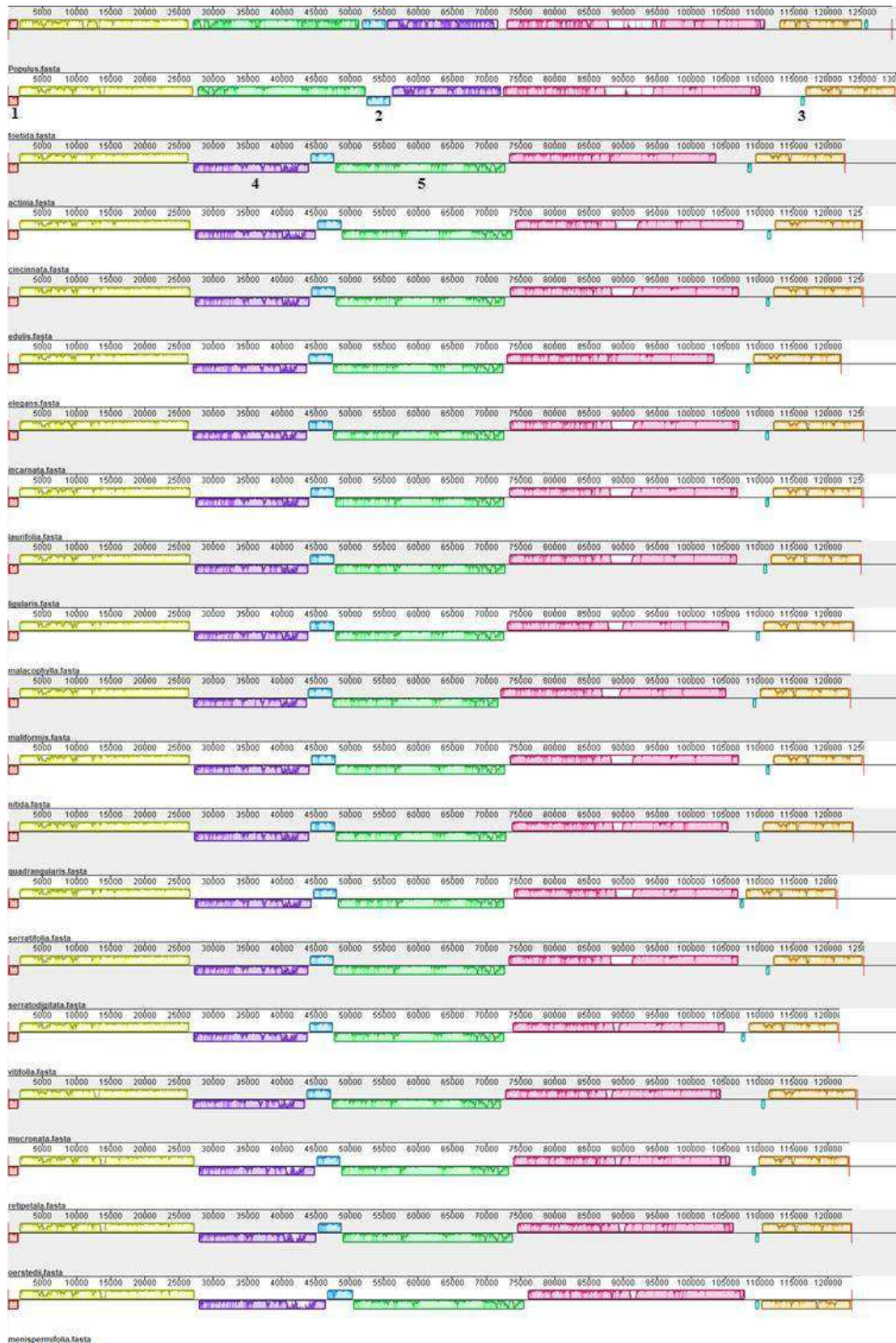
Supplementary Fig. S3 Gene map of *Passiflora malacophylla* plastome. Genes drawn inside the circle are transcribed clockwise and genes drawn outside are transcribed counterclockwise. Different functional groups of genes are color-coded. The darker gray in the inner circle corresponds to GC content and the lighter gray corresponds to AT content. LSC, Large Single Copy; SSC, Small Single Copy; IRA/B, Inverted Repeat A/B



Supplementary Fig. S4 Gene map of *Passiflora maliformis* plastome. Genes drawn inside the circle are transcribed clockwise and genes drawn outside are transcribed counterclockwise. Different functional groups of genes are color-coded. The darker gray in the inner circle corresponds to GC content and the lighter gray corresponds to AT content. LSC, Large Single Copy; SSC, Small Single Copy; IRA/B, Inverted Repeat A/B

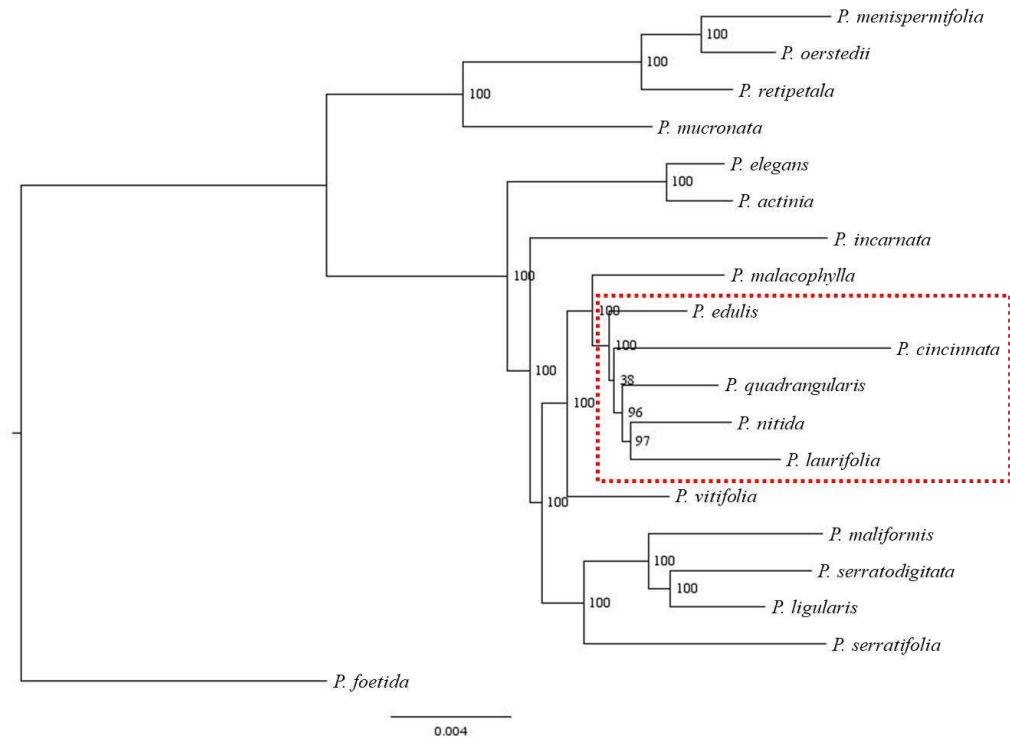


Supplementary Fig. S5 Gene map of *Passiflora mucronata* plastome. Genes drawn inside the circle are transcribed clockwise and genes drawn outside are transcribed counterclockwise. Different functional groups of genes are color-coded. The darker gray in the inner circle corresponds to GC content and the lighter gray corresponds to AT content. LSC, Large Single Copy; SSC, Small Single Copy; IRA/B, Inverted Repeat A/B

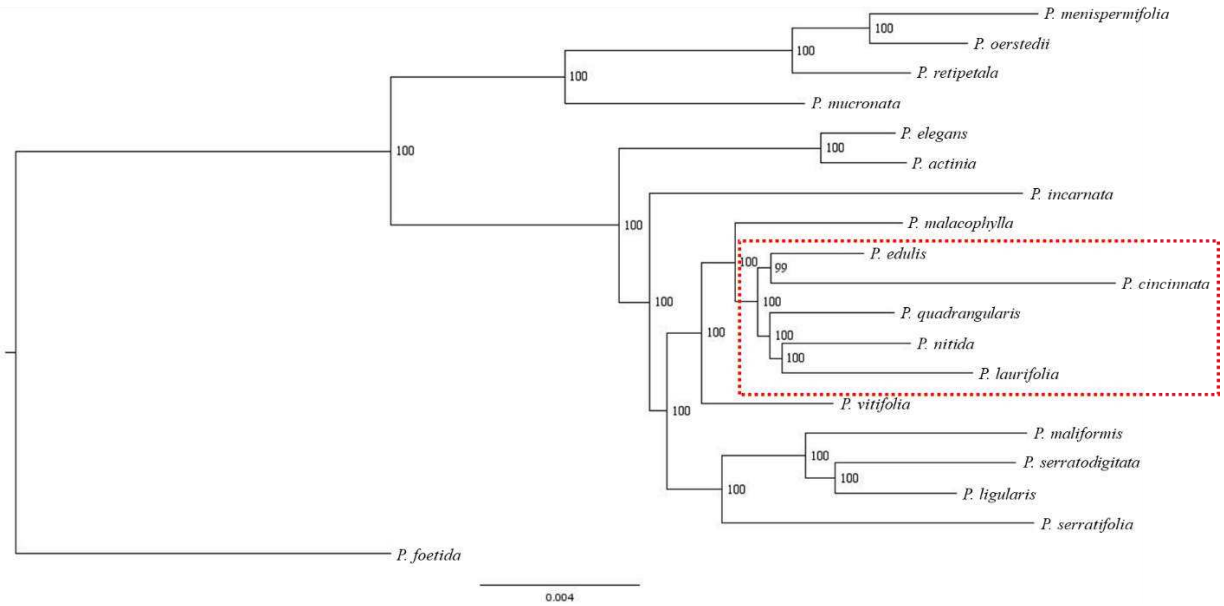


Supplementary Fig. S6 Multiple alignments of *Passiflora* plastomes (subgenus *Passiflora*) and *Populus trichocarpa* as outgroup. Locally collinear blocks (LCBs) are color-coded. The LCBs concerning inversions in *Passiflora* plastomes are illustrated as follows: LCB 1

(including *psbA* and *trnH-GUG* genes); LCB 2 (encompassing from *trnV-UAC* to *atpE* genes); LCB 4 (encompassing from *clpP* to *rbcL* genes); and LCB 5 (encompassing from *ndhC* to *trnC-GCA* genes). The LCB 3 includes the *ycf1* gene that is fully located into the IRs by an expansion event as described by Cauz-Santos et al. (2017) and Rabah et al. (2018)



Supplementary Fig. S7 Phylogenomic tree of the subgenus *Passiflora* based on whole plastomes of 19 taxa using the maximum likelihood (ML) method. The branch length is proportional to the inferred divergence level. The scale bar indicates the number of inferred nucleic substitutions per site. The numbers in front of each node are bootstrap support values (%). The red dashed-box indicates the region of topological incongruence between this tree, the BI tree generated with the same dataset (Supplementary Fig. S8) and the consensus tree based on genes concatenated (Fig. 3)



Supplementary Fig. S8 Phylogenomic tree of the subgenus *Passiflora* based on whole plastomes of 19 taxa using Bayesian inference (BI). The branch length is proportional to the inferred divergence level. The scale bar indicates the number of inferred nucleic substitutions per site. Posterior probabilities (%) are indicated in front of each node. The red dashed-box indicates the region of topological incongruence of this tree in comparison with the ML tree generated with the same dataset (Supplementary Fig. S7) and the consensus tree based on genes concatenated (Fig. 3)

rpoB (partial)

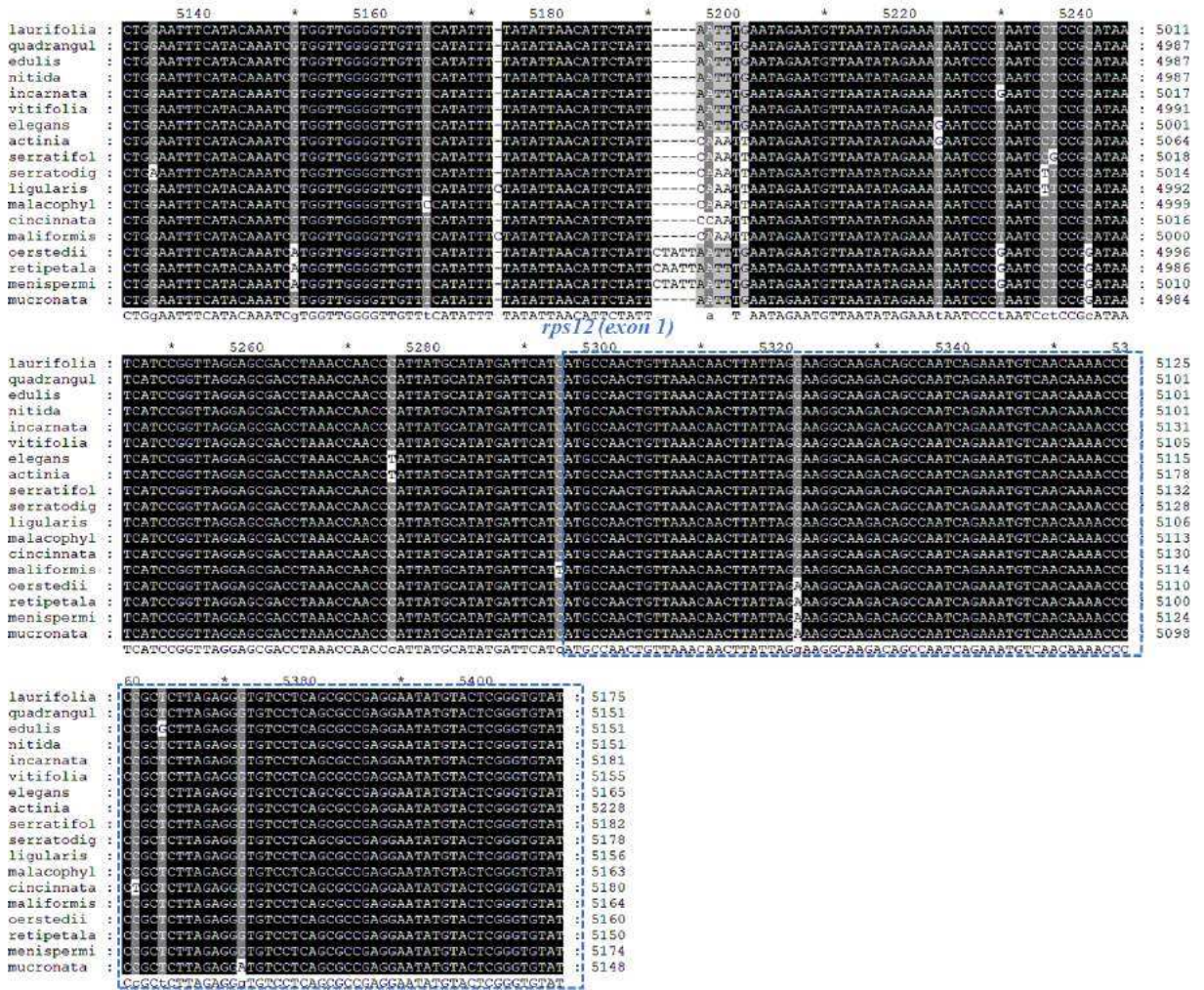
laurifolia	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
quadrangul	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
edulis	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
nitida	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
incarnata	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
vitifolia	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
elegans	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
actinia	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
serratifol	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
serratodig	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
ligularis	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
malacophyl	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
cincinnata	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
maliformis	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
oerstedii	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
retipetala	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
menisperm	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964
micronata	AAATCTGTATATCCATTACTATAGAGGTTCCAGGGAATTCATTAGAGGAATGTTCCAAATAAAAAATGGTTTGTCTTGCATATCCCTTCGCTTTCCAAATTAATCCCTGC	2964

laurifolia	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
quadrangul	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
edulis	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
nitida	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
incarnata	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
vitifolia	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
elegans	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
actinia	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
serratifol	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
serratodig	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
ligularis	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
malacophyl	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
cincinnata	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
maliformis	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
oerstedii	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
retipetala	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
menisperm	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078
micronata	GATACATATAATTCAGAAGATATGTAAGCTTCATATACAGCATCTCTTCTTTATCAAAGTTCTACTAA	3078

laurifolia	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
quadrangul	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
edulis	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
nitida	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
incarnata	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
vitifolia	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
elegans	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
actinia	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
serratifol	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
serratodig	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
ligularis	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
malacophyl	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
cincinnata	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
maliformis	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
oerstedii	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
retipetala	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
menisperm	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192
micronata	ATCTGTACTCTGATTTTGGAAACTTAAAGTCTCTCTTAAAGCCCGTATG	3192

laurifolia	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
quadrangul	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
edulis	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
nitida	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
incarnata	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
vitifolia	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
elegans	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
actinia	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
serratifol	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
serratodig	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
ligularis	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
malacophyl	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
cincinnata	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
maliformis	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
oerstedii	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
retipetala	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
menisperm	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296
micronata	GCATTTCCATCTAAGCATTTTAAATTCOCATTT	3296

laurifolia	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
quadrangul	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
edulis	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
nitida	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
incarnata	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
vitifolia	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
elegans	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
actinia	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
serratifol	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
serratodig	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
ligularis	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
malacophyl	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
cincinnata	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
maliformis	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
oerstedii	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
retipetala	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
menisperm	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400
micronata	AAATTTCTATTCGTTACGGGAATCCATAAAAATTTATCTAA	3400



Supplementary Fig. S9 Clustalw alignment of the plastome region including the *rpoB*, *clpP*, and *rps12* (first exon) genes. The 19 available species of subgenus *Passiflora* (including the species sequenced in this study) were sampled here. The sequences of the *rpoB* (partial) and *rps12* (first exon) genes are highlighted by blue dashed squares. The start and stop codons of the *clpP* gene are indicated by red squares. Black shadowing indicates conserved nucleotides among all taxa; darker gray indicate sequence conserved among all except for one to four taxa; light gray indicate non-conserved sequences in five to seven taxa; and white indicate non-conserved sequences


```

*      *      *      *      *      *
mucronata : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1167
retipetala : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1176
oerstedii : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1199
menisperm : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1193
foetida : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1180
elegans : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1176
actinia : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1176
nitida : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1179
quadrangul : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1179
laurifolia : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1185
malacophyl : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1179
maliformis : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1170
edulis : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1179
ligularis : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1181
vitifolia : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1176
serratedig : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1181
serratifol : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1178
incarnata : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1181
cincinnati : ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTA 1160 1180 1200 1220 1240 1177
          ACCCCACCACTGGTAGGAGTSTAAGGATTGATAA AAAATGACTTTTTAAT C ATTTATAT ATATAAGGGCCAGTATTTTAGCCATTTGCATCAAGCTCAAA

```

```

*      *      *      *      *      *
mucronata : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1281
retipetala : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1290
oerstedii : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1313
menisperm : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1307
foetida : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1294
elegans : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1290
actinia : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1290
nitida : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1293
quadrangul : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1293
laurifolia : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1299
malacophyl : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1293
maliformis : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1284
edulis : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1293
ligularis : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1295
vitifolia : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1290
serratedig : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1295
serratifol : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1292
incarnata : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1295
cincinnati : TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA 1260 1280 1300 1320 1340 1291
          TTTCCTTCGTACATGCGGCTCTCTCCGGAAGCACACACTAGATAAAGGGGTAA TTTGATTGGTAGCATATTCGATCAAAAGAGTATTTCTCACCTACTAAG ATCCGCATA

```

```

*      *      *      *      *      *
mucronata : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1355
retipetala : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1404
oerstedii : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1427
menisperm : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1421
foetida : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1408
elegans : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1404
actinia : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1404
nitida : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1407
quadrangul : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1407
laurifolia : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1413
malacophyl : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1407
maliformis : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1398
edulis : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1407
ligularis : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1409
vitifolia : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1404
serratedig : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1409
serratifol : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1406
incarnata : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1409
cincinnati : CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAAT 1380 1400 1420 1440 1460 1480 1405
          CTACCTCCGATAAATCGAAAAATCCATCACTCCAA TGTCTAGGGAAATGCCGTTATTGACCTAACCCGTTTGAAACAGCTTCGGATAAATCTGTTCCTGTGACAAG ARAA

```

```

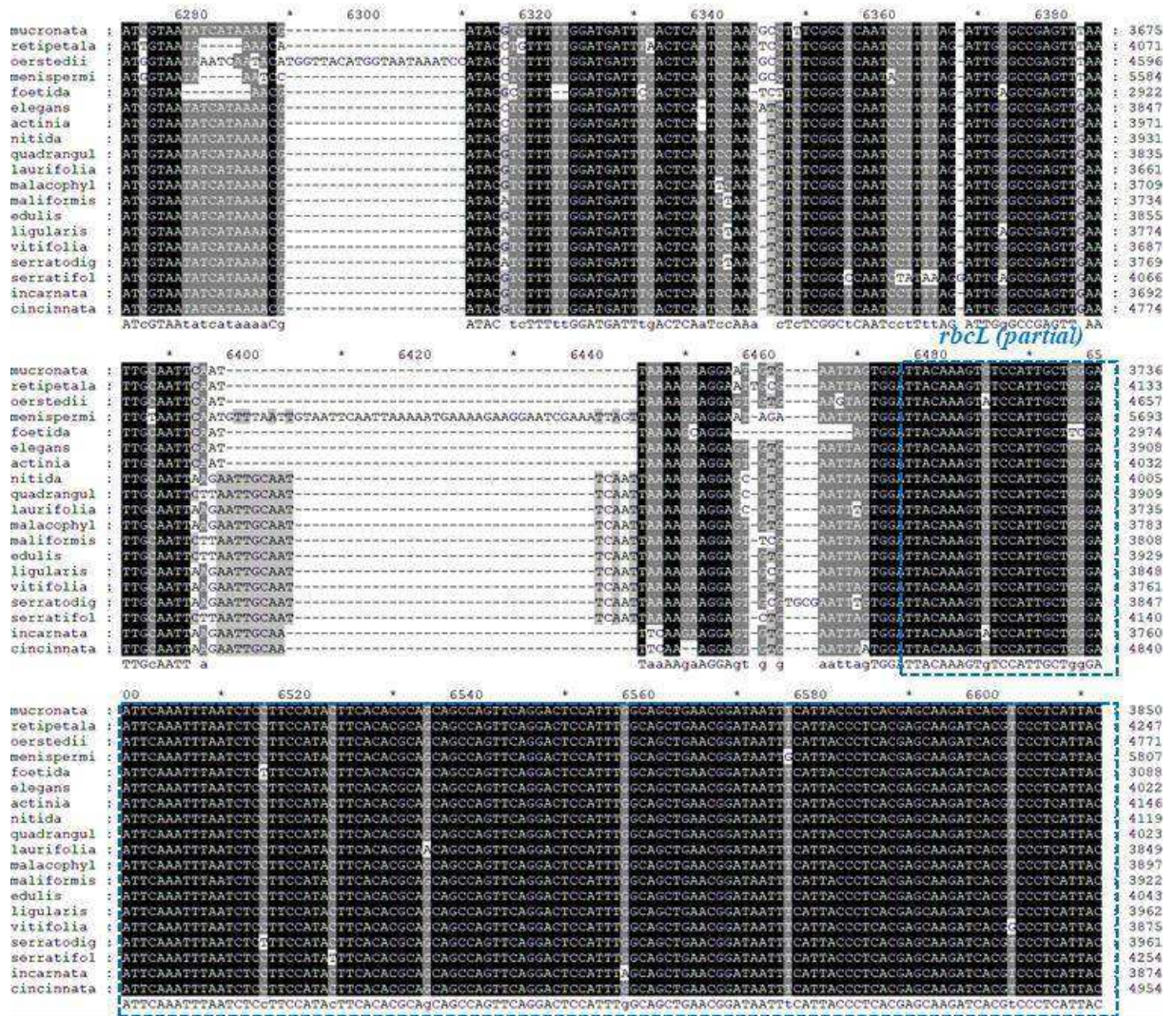
*      *      *      *      *      *
mucronata : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1496
retipetala : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1499
oerstedii : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1535
menisperm : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1522
foetida : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1496
elegans : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1505
actinia : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1505
nitida : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1511
quadrangul : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1511
laurifolia : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1517
malacophyl : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1508
maliformis : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1499
edulis : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1511
ligularis : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1510
vitifolia : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1505
serratedig : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1510
serratifol : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1507
incarnata : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1504
cincinnati : AAGCCGCTCATATAAAGTTTACCTCCAGCCCAACATGCA 1500 1520 1540 1560 1580 1503
          ATGCGCTCtA ATAAGTTTATCTCCAGCCCAACATGCA TtC Tc TtGATct cAACCAtgacAgTCCcCTTcttctctccg atcc ttt CttCtTTTT

```

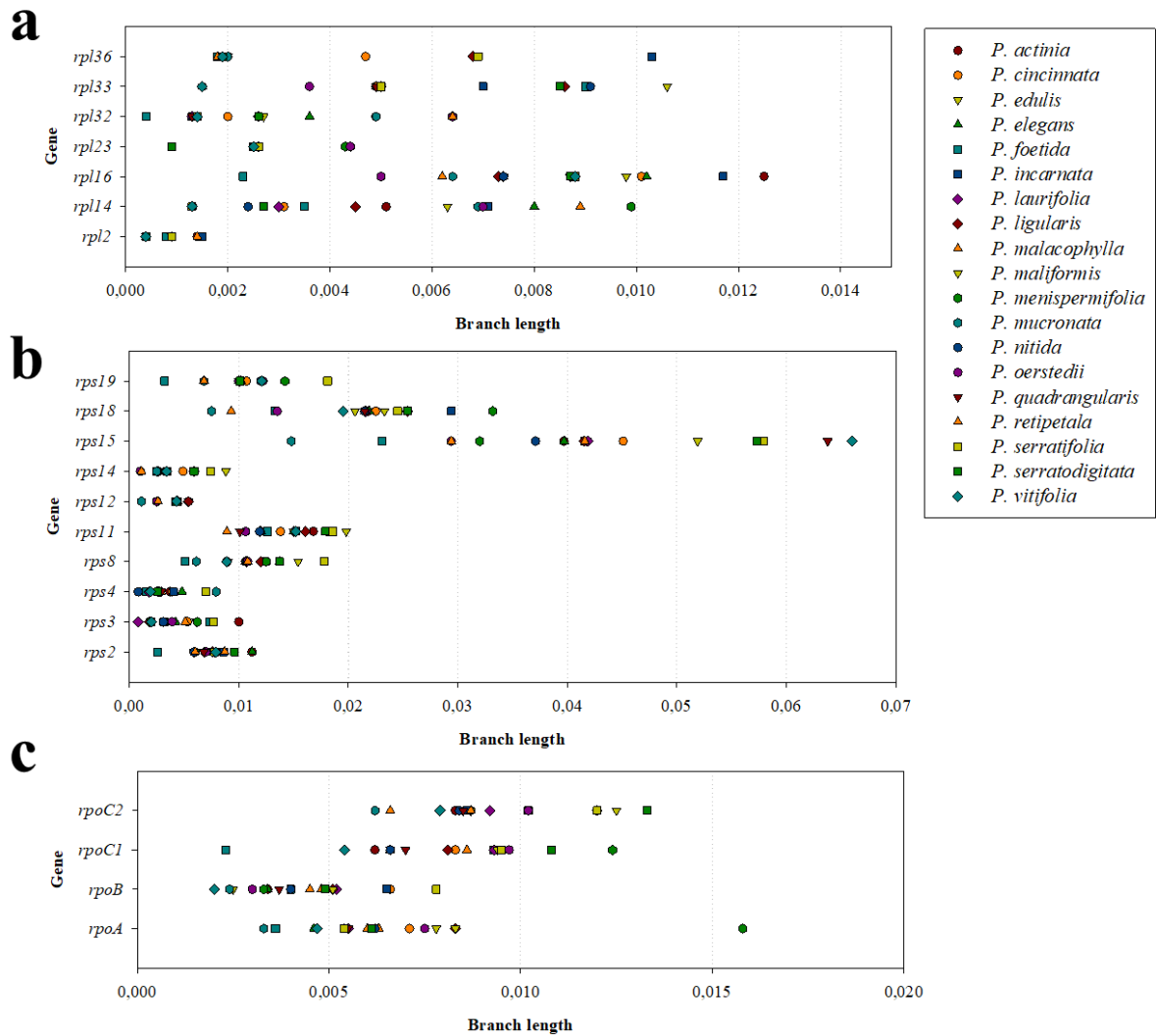
```

*      *      *      *      *      *
mucronata : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1593
retipetala : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1599
oerstedii : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1640
menisperm : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1619
foetida : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1549
elegans : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1578
actinia : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1578
nitida : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1584
quadrangul : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1584
laurifolia : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1590
malacophyl : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1581
maliformis : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1569
edulis : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1584
ligularis : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1580
vitifolia : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1578
serratedig : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1580
serratifol : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1577
incarnata : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1571
cincinnati : TTCTTCTTTT 1600 1620 1640 1660 1680 1700 1582
          ttctcttt ct ct cg atc cggat c tt ttc ccca g atc c tc ttgtcc a

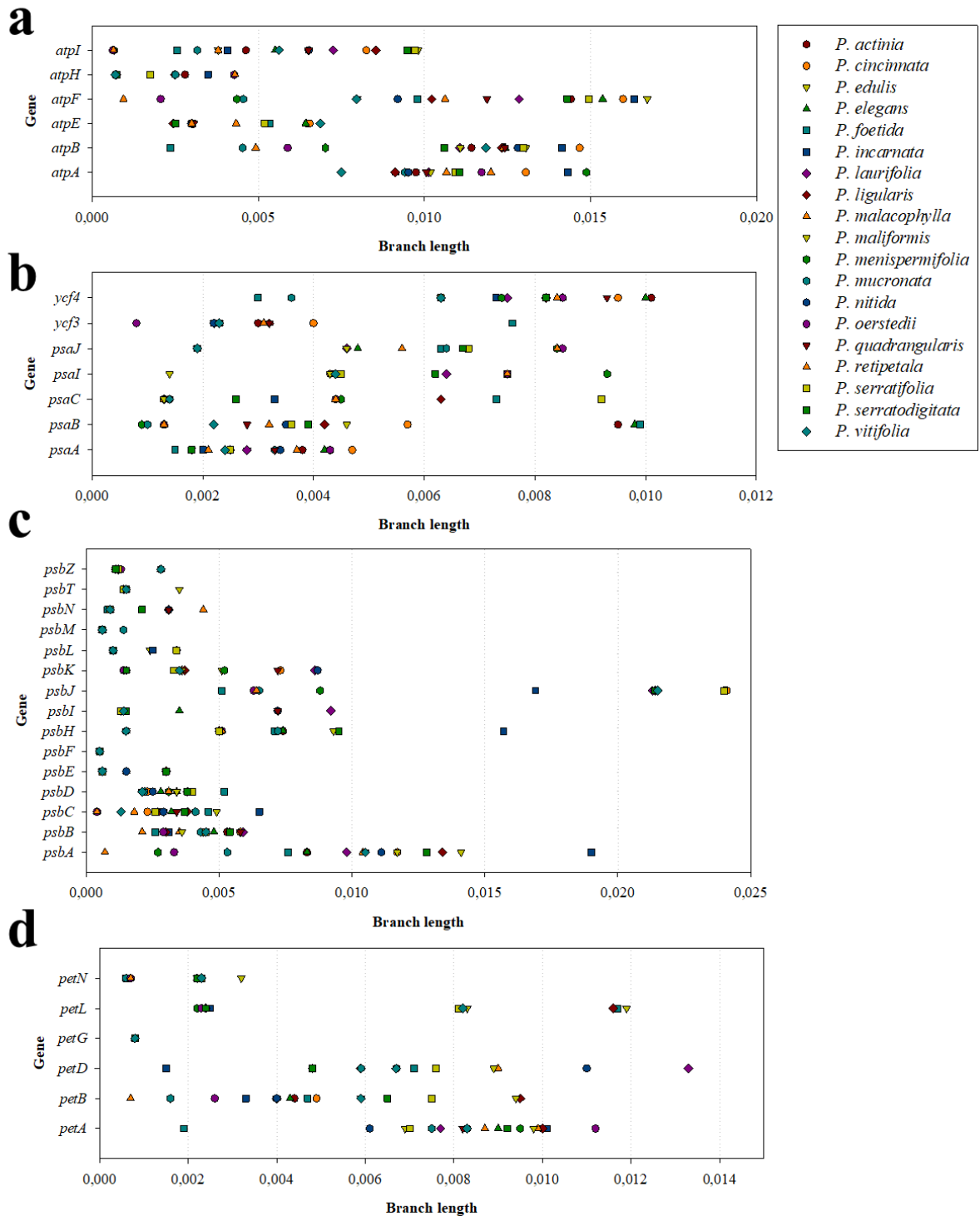
```

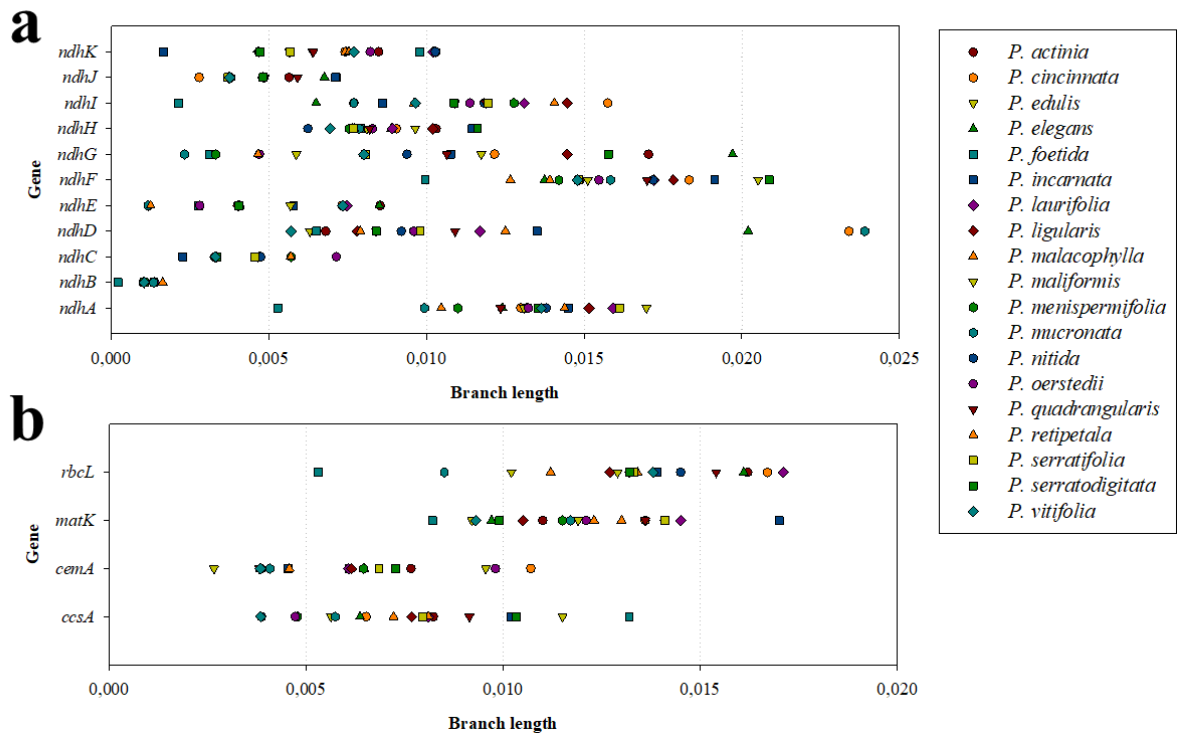
Supplementary Fig. S10 Clustalw alignment of the plastome region including the genes *psaI*, *accD*, and *rbcL*. The 19 available species of subgenus *Passiflora* (including the species sequenced here) were sampled. The sequences of the *psaI* and *rbcL* (partial) genes are highlighted by blue dashed squares. The stop codons of the *accD* genes of each species are indicated by red dashed squares. Two conserved putative start codons are indicated by red squares. Black shadowing indicates conserved sequences among all taxa; darker gray indicate conserved sequences among all except for one to four taxa; light gray indicate non-conserved sequences for five to seven taxa; and white indicate non-conserved sequences



Supplementary Fig. S11 Divergence of plastid genes related to gene expression machinery in the subgenus *Passiflora*. **a** Genes encoding ribosomal proteins of the large subunit of plastid ribosome (*rpl* genes); **b** Genes encoding ribosomal proteins of the small subunit of plastid ribosome (*rps* genes); **c** Genes for plastid-encoded RNA polymerase (*rpo* genes). The gene divergence was estimated by the sum of total branch lengths inferred in each gene tree



Supplementary Fig. S12 Divergence of plastid genes related to photosynthetic machinery in the subgenus *Passiflora*. **a** Genes encoding subunits of ATP synthase (*atp* genes); **b** Genes encoding subunits of photosystem I (*psa* genes); **c** Genes encoding subunits of photosystem II (*psb* genes); **d** Genes encoding subunits of cytochrome b_6f complex (*pet* genes). The gene divergence was estimated by the sum of total branch lengths inferred in each gene tree



Supplementary Fig. S13 Divergence of plastid genes in the subgenus *Passiflora*. **a** Genes encoding subunits of NDH complex (ndh genes); **b** Genes related to other functions (see the **Table 1** for more details). The gene divergence was estimated by the sum of total branch lengths inferred in each gene tree



Supplementary Fig. S14 Alignment of exon 3 of the rps12 gene. The species included in the squares belong to the subgenus Passiflora. The black square encompasses species that lost the “AAAAAG” insertion, while the species of the red square encompasses the species that contain this sequence. Different background colors indicate amino acids with different biochemical properties. Populus was used as the outgroup representing the general sequence observed in other angiosperms


```

*      20      *      40      *      60      *      80      *      100
Populus : MKKCFEHSMSNV--EYRGRRLSKSMONIGPLENTS----- : 35
foetida : MANLNKQKLNPCSSSRGFLDIENPVLG----- : 32
menispermi : MANLNKRLWTPCCSNRGTLDLIEKENP----- : 44
oerstedii : MANLNKRLWTPCCSNRGTLDLIEKDIEN----- : 42
mucronata : MANLNKRLWTPCCSNRGTLDLIESENPA----- : 28
retipetala : MANLNKRLWTPCCSNRGTLDLIEENPG----- : 28
elegans : MANLNKRLWTPCCSNRGTLDLIEENQND----- : 42
actinia : MANLNKRLWTPCCSNRGTLDLIEENQND----- : 50
incarnata : MANLNKRLWTPCCSNRGTLDLIEENKKD----- : 44
malacophyl : MANLNKRLWTPCCSNRGTLDLIEEN----- : 36
vitifolia : MANLNKRLWTPCCSNRGTLDLIEEN----- : 36
laurifolia : MANLNKRLWTPCCSNRGTLDLIEENQKEN----- : 56
nitida : MANLNKRLWTPCCSNRGTLDLIEENPL----- : 48
cinninata : MANLNKRLWTPCCSNRGTLDLIEENKNIETENTVSENNTNNTNTEENNTNTEENNTNTEENNTNTEENNTNTEEN : 96
quadrangul : MANLNKRLWTPCCSNRGTLDLIEEN----- : 36
edulis : MANLNKRLWTPCCSNRGTLDLIEEN----- : 43
maliformis : MANLNKRLWTPCCSNRGTLDLIEENQ----- : 36
serratotidig : MANLNKRLWTPCCSNRGTLDLIEENQNEI----- : 60
ligularis : MANLNKRLWTPCCSNRGTLDLIEENQNEI----- : 70
serratifol : MANLNKRLWTPCCSNRGTLDLIEENQIENENQI----- : 74
putative start  nln k iLwtpCcsn  p6  n
codon 2

```

```

*      120      *      140      *      160      *      180      *      200
Populus : VSEDFILNDTEKNTLNWSSDSSNVDFVGRDIRNLVDDTFLVLRDNKKKDGYSYFDIENQVFGIEN----- : 105
foetida : TINDPSTLDESSFFRANFCYRANFHSIVDDYRSCVEISSSSGSENGSPFSNGSE----- : 91
menispermi : PAREIINDPFFVSTIDNIFYNMLTLPRIENFENLKKNIIDDLIRSVINRARSGGGNEEPSTNNHSD--DQKGENEEPSTNDDSDDDSDSQSEN : 138
oerstedii : PVREIINDPFFVSTIDNIFYNMLTLPRIENFENLKKNIIDDLIRSEIIS--SSGSENEEPSTNDHSDG----- : 109
mucronata : REIINDPFFLNKIESFFHFTFCYRKNYIDNLKDYDLDLIR--SAIANSSRSSENEEPSTNENSD--HKGENEE----- : 100
retipetala : REIINDPFFVSKIDSFFHFTFCYRKNYIDNLKKNIDFIRSEIASSSESENEEPSTNDNSD--QKGENEE----- : 102
elegans : PVSDEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANENQKNGRIENP----- : 115
actinia : QNDIETENPVSDSEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANENQKNGRIENP----- : 131
incarnata : PASEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSADIPSSANKKNEEPNTNENSDAHN----- : 120
malacophyl : SEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANENQKNGRIDPNSD----- : 101
vitifolia : SEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSADIASASSANQKNGRIDPNSD----- : 107
laurifolia : SEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSADIASASSANQKNGRIDPNSD----- : 121
nitida : SEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANQKNGRIDPNSD----- : 118
cinninata : NTNNTTEETENPGSEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANQKNGRIDPNSD----- : 176
quadrangul : SEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANQKNGRIDPNSD----- : 106
edulis : SEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANQKNGRIDPNSD----- : 108
maliformis : SEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANQKNGRIDPNSD----- : 110
serratotidig : SEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANQKNGRIDPNSD----- : 131
ligularis : SEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANQKNGRIDPNSD----- : 144
serratifol : SEIINDPFFRQSAIDGIFVRYENECERKVIDNLDSDIIDDFFDILRSASDIASSANQKNGRIDPNSD----- : 150
e6  DPff s  c5  y  n  r  6  f

```

```

*      220      *      240      *      260      *      280      *      300
Populus : NHFLSKLEKRFSSYWNSSYLKNGSRSDSHYDYSMYDNKYWNNYINSCIYSYLR----- : 161
foetida : NSSPFSSTNDNST--DHKSE----- : 110
menispermi : SSPRPITNDHSDDDSDSQSENSSRPST--DHKSEEDDQKGENEEESTNDSDDDQNGENKEPSTNDHSDDDQNGENKEPSTNDSDDD : 242
oerstedii : KGENGSRPST--DHKSEEDDQKGENEEESTNDSDDDPSTNGRPFWNHSDSDSP--STNGRPFWNHSDSDS : 193
mucronata : PSTENSDSRGONARPTNDHSE--DQKGENEEESTNDSDDDPSTNGRPFWNHSDSDSP--STNGRPFWNHSDSDS : 155
retipetala : PSTENSDSRGONARPTNDHSE--DQKGENEEESTNDSDDDPSTNGRPFWNHSDSDSP--STNGRPFWNHSDSDS : 180
elegans : DNSDHSSTHNGEEDPSTNDSE--ESGSENGSPSTNDHSDDDPSTNGRPFWNHSDSDSP--STNGRPFWNHSDSDS : 172
actinia : HNSDHSSTHNGEEDPSTNDSE--ESGSENGSPSTNDHSDDDPSTNGRPFWNHSDSDSP--STNGRPFWNHSDSDS : 188
incarnata : GENEESTPSTNDSDHNGEEDPSTNDSE--THNCEEEESTSGN-----DESQRENGRPS----- : 176
malacophyl : NHNGEESTPSTNDSDHNGEEDPSTNDSE--THNCEEEESTSGNDESGQNGRPSQNTNEDADAN----- : 169
vitifolia : NHNGEESTPSTNDSDHNGEEDPSTNDSE--THNCEEEESTSGNDESGQNGRPSQNTNEDADITGRHKD----- : 169
laurifolia : NHNGEESTPSTNDSDHNGEEDPSTNDSE--THNCEEEESTSGNDESGQNGRPSQNTNEDADITGRHKD----- : 173
nitida : NHNGEESTPSTNDSDHNGEEDPSTNDSE--THNCEEEESTSGNDESGQNGRPSQNTNEDADITGRHKD----- : 181
cinninata : GETEEPSTNDSDHNGEEDPSTNDSE--DHKSEEDDQKGENEEESTNDSDDDPSTNGRPFWNHSDSDSP--STNGRPFWNHSDSDS : 250
quadrangul : NHNGEESTPSTNDSDHNGEEDPSTNDSE--THNCEEEESTSGNDESGQNGRPSQNTNEDADITGRHKD----- : 169
edulis : NHNGEESTPSTNDSDHNGEEDPSTNDSE--THNCEEEESTSGNDESGQNGRPSQNTNEDADITGRHKD----- : 157
maliformis : GESEEDPSTYENSTHNGEEDPSTNDSE--THNCEEEESTSGNDESGQNGRPSQNTNEDADITGRHKD----- : 179
serratotidig : NHNGEESTPSTNDSDHNGEEDPSTNDSE--THNCEEEESTSGNDESGQNGRPSQNTNEDADITGRHKD----- : 192
ligularis : SDHSSSTYENSTHNGEEDPSTNDSE--THNCEEEESTSGNDESGQNGRPSQNTNEDADITGRHKD----- : 207
serratifol : THNGEEDPSTYENSTHNGEEDPSTNDSE--DHKSEEDDQKGENEEESTNDSDDDPSTNGRPFWNHSDSDSP--STNGRPFWNHSDSDS : 226
n  sd  g  e  pst  nsd  p

```

```

*      320      *      340      *      360      *      380      *      400      *
Populus : ----- : -
foetida : ----- : -
menispermi : SEIPIPRPFFWIWYDSDDDSLSPGPNRINGIPVLNGDSYDGKNGSLFVNDDSDDDSDQLGENSEAPWNGDSYDYNLGTNGSLSWNDHSDDQ----- : 338
oerstedii : PSENNIPRPFWNHSDSDSPSTNNIWP-----FVNDDSDDDSDQLGENSEAPWNGDSYDYNLGTNGSLSWNDHSDDQ----- : 265
mucronata : -----NNDPFDLSNENKPSRIGNDNPDSDRSQNGSAFWNENSDDHKGENEEP----- : 203
retipetala : SDDS--LGPNGFPRNDHSDSLGPNCFPRNDHSDSLGENSEAPRPFWNHSDSDSLGPNCFPRNDHSDSDSPSENGSPS----- : 258
elegans : -----NATSSPFGTNNSPFGKENS--DRPNENIHPFVNVN--PLGNDNSDDSGSQNG-----S----- : 225
actinia : -----NATR--SPFGKENS--DRPNENIHPFVNVN--PLGNDNSDDSGSQNG-----S----- : 234
incarnata : NTNENADINGRHKDDGFSIAYNTNESASGSARRPS--PNTNENADINGRYKDDGLSIDAYYN----- : 238
malacophyl : GSPNASPLRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 220
vitifolia : DGPSTNANSADSPRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 229
laurifolia : NNARPLRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 224
nitida : NNASPLRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 230
cinninata : NSDADADGLNSPLRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 322
quadrangul : NNASPLRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 218
edulis : ANASPLRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 206
maliformis : PDDAPLRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 225
serratotidig : SDDAPLRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 240
ligularis : SDADAPLRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 261
serratifol : DAYNYTNGANSPLRIWVWSSRDANS--DRPNENSHN--VDNVS--LFGNDNSDDSGSQNG----- : 286
ng

```

```

420      *      440      *      460      *      480      *      500      *      520
Populus : -----SQIGIASSLLGSESYSESYISTYLIGESRNSSETGNSRLRUSTNGSIFALRPN----- : 215
foetida : -----NEEKGEN----- : 117
menispermi : -KGENKEPSTNDHSDDDQKGENKEPSTNDHSDDDQKGENKEPSTNDHSDDDQKGENSSPRPSTNDHSDQKGENS PRPSTNDHSDLDQNGENKEPSTNDSDS : 441
oerstedii : -KGENEEPSTNDHSDDDQKGENEEPSTNDHSD--DQKGENEEPSTNDHSDDDQKGENEEPSTNDHSDDDQKGENS PRPSTNDHSDSD : 350
mucronata : -STNENSDDTL-----SPNGRPSRIWTWCSEDTNSDNSPSQARPPWMTWCSTDDNPDNRIPAFWN--YNPDDHEN----- : 273
retipetala : AFWNDNSDDSP-----SENGRPCRITWTCSEDTNSDNSPSQARPPWMTWCSTDDNPDNRIPAFWN--YNPDDHEN----- : 331
elegans : LFQSQNGSPLGND-----NSDDHKGENEEPSTNDNSDDHKGEN----- : 263
actinia : LFQSQNGSPLGNDNYDDSGSQNGSLFQSQNGSLFQSQNGSPLGTSTNDNSDDHKGENEEPSTNDNSDDHKGENEEPSTNDNSDDHKGEN : 323
incarnata : TNGKPNDSPLRIWVWSSLDANYDDRPNIPKVNIGIRNLDLNVGLFVNENSDDHNGENEEPSTNDSSDHNGES EEPSTNENS DAHKGEN : 327
malacophyl : --SLFQSQNG--SLFG-----TSTNDNSDDHKGENEEPSTNDNSDDHKGENEDPSTNDNSDTHNGES DLSTNDNSDDHKGEN : 294
vitifolia : --SLFQSQNG--SLFG-----SLFGTSTNDNSDDHKGENEEPSTNDNSDDHKGENEEPSTNDNSDDHKGEN : 287
laurifolia : --SLFG-----TSTNDNSDDHKGENEEPSTNDNSDDHKGENEEPSTNDNSDDHKGENEEPSTNDNSDDHKGEN : 290
nitida : --SLFQSQNG--SLFQSQNGSLFGTSTNDNSDDHN---EEPSTNDNSDDHKGENEDPSTNDNSDTHNGENEEPSTNDNSDDHKGEN : 309
cincinnata : LLGNDNSDDHGSQNGSLFQSQNGSLFGTSTNDNSDDHKGENEEPSTNDNSDDHKGENEDPSTNDNSDTHNGENEEPSTNDNSDDHKGEN : 414
quadrangul : --SLFQSQNG--SLFQSQNGSLFGTSTNDNSDDHKGENEDPSTNDNSDTHNGESDPSTNDNSDTHNGENEEPSTNDNSDDHKGEN : 300
edulis : --SLFQSQNGSLFG-----TSTNDNSDDHKGENEDPSTNDNSDTHNGESDPSTNDNSDDHKGENEEPSTNDNSDDHKGEN : 280
maliformis : SLFGTSTN-----DNFDDQKGENEEPSTNDNSDTHNGESDPSTNDNSDDHKGENEEPSTNDNSDDQKGEN : 291
serratodig : SLFGTSTN-----DNSDDQKGENEEPSTNDNSDDQKGE P--STNDNSDDQKGENEEPSTNDNSDDQKGEN : 303
ligularis : SLFGTSTN-----DNSDDQKGENEEPSTNDNSDDQKGE-----NEEPSTNDNSDDQKGEN : 311
serratifol : SLFGTSTD-----DSSDTHNGESEDPSTNENS DTHNGE-----SEDPSTNENSDDHNGEN : 336

```

```

*      540      *      560      *      580      *      600      *      620
Populus : ----- :
foetida : ----- :
menispermi : DDDDEIISIPSPFIIITWYDSDDDDSQSLGPNGSAPWNGDSYDSPNLGTNGSLSWNDHSDQKGENSSPRPSTNDHSDQKGENSSPRPSTNDHSDDDQKGEN : 549
oerstedii : -----QKGENEEPSTNDHSDSD : 368
mucronata : ----- :
retipetala : ----- :
elegans : ----- :
actinia : ----- :
incarnata : ----- :
malacophyl : ----- :
vitifolia : ----- :
laurifolia : ----- :
nitida : ----- :
cincinnata : ----- :
quadrangul : ----- :
edulis : ----- :
maliformis : ----- :
serratodig : ----- :
ligularis : ----- :
serratifol : ----- :

```

```

*      640      *      660      *      680      *      700      *      720
Populus : ----- :
foetida : ----- :
menispermi : EEPSTNEDSDDDQKGENSSPRPSTNDSDDDNSPSQNSSPSAFWNDHSDNSPSQNSSPSAFWNDHSDDDSDSQSQNSSPSAFWNDHSDNSPSQNSGPPRIWTWC : 649
oerstedii : -----SPSQNGSPSAFWNDNSDDSPSQNASPSAFWNDNSDD----- : 404
mucronata : ----- :
retipetala : ----- :
elegans : ----- :
actinia : ----- :
incarnata : ----- :
malacophyl : ----- :
vitifolia : ----- :
laurifolia : ----- :
nitida : ----- :
cincinnata : -----STSGNSDADADGSPLAGSSR----- : 434
quadrangul : ----- :
edulis : ----- :
maliformis : ----- :
serratodig : ----- :
ligularis : ----- :
serratifol : ----- :

```

```

*      740      *      760      *      780      *      800      *      820      *
Populus : -----ENLGVTKRKHLLVWQCEHCYGLNYK--FRSKMNI : 252
foetida : --EPSPSRRAIDYSVWVQCEHCYGINVK--FRKSRNI : 156
menispermi : SEDTSDNSPSQNSGPPRIWTWCSEDTSDNSPSQNARESWMWWCSTDDSDVDRNPKPRPDIISYRTERPILIVSALWVCEHCYGVNYKRFKSRNI : 753
oerstedii : -----SPSQNSGPPRIWTWCSEDTSDNSPSQNARESWMWWCSTDDSDSERNPKPRPDIISYSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 500
mucronata : -----EPSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 314
retipetala : --KSTNERKIIDYSDLWVQCEHCYGINYKRFKSRNI : 372
elegans : --EPSTNERKIIDYSVWVQCEHCYGMNYKRFKSRNI : 304
actinia : --EPSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 364
incarnata : --DPSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 368
malacophyl : --DPSTNERKIIDYSVWVQCEHCYGINYK--FRKSRNI : 333
vitifolia : --EPSTNERKIIDYSVWVQCEHCYGINYK--FRKSRNI : 326
laurifolia : --EPSTNERKIIDYSVWVQCEHCYGVNYKRFKSRNI : 331
nitida : --EPSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 350
cincinnata : MWVWSSRDVNSDESEDPTNDNSDTHNGESEDPSTNDFDTHNGETEPEPSTNDFDTHNGETE--EPSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 538
quadrangul : --EPSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 341
edulis : --EPSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 321
maliformis : --EPSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 332
serratodig : --EPSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 344
ligularis : --EPSTNERKIIDYSVWVQCEHCYGINYKRFKSRNI : 352
serratifol : --EP--RKIIDYSVWVQCEHCYGINYKRFKSRNI : 373

```

e pst erk d a WVQCE CYG6NYK FF 346NICE

```

      840      *      860      *      880      *      900      *      920      *
Populus : CCGYHLKMS...SRIELS...IDPGT...M... : 279
foetida : HCGCHLRMS...SRIELS...IDPGT...M... : 196
menispermi : HCGCHLRMS...SRIELS...IDPGT...M... : 821
cerstedii : HCGCHLRMS...SRIELS...IDPGT...M... : 562
mucronata : HCGCHLRMS...SRIELS...IDPGT...M... : 400
retipetala : HCGCHLRMS...SRIELS...IDPGT...M... : 429
elegans : YCGCHLRMS...SRIELS...IDPGT...M... : 372
actinia : HCGCHLRMS...SRIELS...IDPGT...M... : 428
incarnata : HCGCHLRMS...SRIELS...IDPGT...M... : 418
malacophyl : HCGCHLRMS...SRIELS...IDPGT...M... : 373
vitifolia : HCGCHLRMS...SRIELS...IDPGT...M... : 382
laurifolia : YCGCHLRMS...SRIELS...IDPGT...M... : 390
nitida : HCGCHLRMS...SRIELS...IDPGT...M... : 453
cinninata : HCGCHLRMS...SRIELS...IDPGT...M... : 607
quadrangul : HCGCHLRMS...SRIELS...IDPGT...M... : 418
edulis : HCGCHLRMS...SRIELS...IDPGT...M... : 382
maliformis : HCGCHLRMS...SRIELS...IDPGT...M... : 406
serratotdig : HCGCHLRMS...SRIELS...IDPGT...M... : 432
ligularis : HCGCHLRMS...SRIELS...IDPGT...M... : 434
serratifol : HCGCHLRMS...SRIELS...IDPGT...M... : 466
hcgchL46dsperIELsIDPGT... :
      putative start
      codon 1
      d ed

```

```

      940      *      960      *      980      *      1000      *      1020      *      1040
Populus : -----HMFPSID... : 292
foetida : -----SDP... : 240
menispermi : DLSVLDLYGWDSEESDEDLPVLDPMDEAGKEDSIDPGTRDPMDEAGKED... : 925
cerstedii : -----GTRDPMDEAGKEDSIDPGTRDPMDEAGKED... : 628
mucronata : -----DPMDEAGKEDSIDPGTRDPMDEAGKED... : 448
retipetala : -----DPMDEAGKEDSIDPGTRDPMDEAGKED... : 475
elegans : I-----DPGTRDPVDES... : 441
actinia : -----CEDEMS... : 467
incarnata : -----HMS... : 454
malacophyl : -----DPMDEAGKEDSIDPGTRDPMDEAGKED... : 419
vitifolia : -----EAGKED... : 436
laurifolia : -----DSIDPGT... : 444
nitida : -----DSIDPGT... : 507
cinninata : -----SIDPAGKED... : 659
quadrangul : -----DSIDLGT... : 472
edulis : -----DSIDPGT... : 436
maliformis : -----EAGKED... : 448
serratotdig : -----GKED... : 478
ligularis : -----EAGKED... : 476
serratifol : -----DSIDPGT... : 520
      ea ede 6 LDpi 5d
      e eed dew ldsdl

```

```

      *      1060      *      1080      *      1100      *      1120      *      1140
Populus : ----- : 257
foetida : YSYSDDPPNTDPEE... : 1016
menispermi : -----ESEEEKDES... : 684
cerstedii : YSYSDDPPNTDPEE... : 467
mucronata : YSYSDDPPNTDPEE... : 579
retipetala : YSYSDDPPNTDPEE... : 458
elegans : YSYSDDPPNTDPEE... : 484
actinia : YSYSDDPPNTDPEE... : 471
incarnata : YSYSDDPPNTDPEE... : 434
malacophyl : YSYSDDPPNTDPEE... : 451
vitifolia : YSYSDDPPNTDPEE... : 456
laurifolia : YSYSDDPPNTDPEE... : 522
nitida : YSYSDDPPNTDPEE... : 677
cinninata : YSYSDDPPNTDPEE... : 487
quadrangul : YSYSDDPPNTDPEE... : 456
edulis : YSYSDDPPNTDPEE... : 469
maliformis : YSYSDDPPNTDPEE... : 499
serratotdig : YSYSDDPPNTDPEE... : 493
ligularis : YSYSDDPPNTDPEE... : 537
serratifol : YSYSDDPPNTDPEE... :
      ysysddp tdpee

```

```

      *      1160      *      1180      *      1200      *      1220      *      1240
Populus : ----- : 1122
foetida : ESDLDESESEK... : 747
menispermi : -----ESEK... : 530
cerstedii : -----ESEK... : 652
mucronata : SEEDDPEDSS... : 520
retipetala : SEEDDPEDSS... : 542
elegans : SEEDDPEDSS... : 533
actinia : SEEDDPEDSS... : 496
incarnata : SEEDDPEDSS... : 497
malacophyl : SEEDDPEDSS... : 491
vitifolia : SEEDDPEDSS... : 583
laurifolia : SEEDDPEDSS... : 754
nitida : SEEDDPEDSS... : 538
cinninata : SEEDDPEDSS... : 519
quadrangul : SEEDDPEDSS... : 529
edulis : SEEDDPEDSS... : 540
maliformis : SEEDDPEDSS... : 522
serratotdig : SEEDDPEDSS... : 601
ligularis : SEEDDPEDSS... :
serratifol : SEEDDPEDSS... :
      e
      e

```


Supplementary Tables**Supplementary Table S1.** Features of the reads and contigs used for plastome assembly

Species	Subgenus	Total reads	Trimmed reads	Average coverage of the contigs used (lower-higher)
<i>P. cincinnata</i>	Passiflora	3,976,822	3,956,881	1339.88 - 9331.70
<i>P. elegans</i>	Passiflora	248,102	247,542	112.09 - 257.61
<i>P. incarnata</i>	Passiflora	621,142	620,677	174.75 - 740.90
<i>P. malacophylla</i>	Passiflora	634,862	634,585	506.41 - 815.00
<i>P. maliformis</i>	Passiflora	298,568	298,364	180.57 - 282.43
<i>P. mucronata</i>	Passiflora	1,024,072	1,023,409	373.06 - 941.42

Supplementary Table S2. List of species included in the phylogenetic inferences

Espécies	Subgênero	GenBank
<i>P. actinia</i> Hook.	Passiflora	NC_038118.1
<i>P. cincinnata</i> Mast.	Passiflora	NC_037690.1
<i>P. edulis</i> Sims	Passiflora	NC_034285.1
<i>P. elegans</i> Mast.	Passiflora	MN062356
<i>P. foetida</i> L.	Passiflora	NC_043825.1
<i>P. incarnata</i> L.	Passiflora	MN062358
<i>P. laurifolia</i> L.	Passiflora	NC_038121.1
<i>P. ligularis</i> Juss.	Passiflora	NC_038122.1
<i>P. malacophylla</i> Mast.	Passiflora	MN062358
<i>P. maliformis</i> L.	Passiflora	MN062359
<i>P. menispermifolia</i> Kunth	Passiflora	NC_043826.1
<i>P. mucronata</i> Lam.	Passiflora	MN062360
<i>P. nitida</i> Kunth	Passiflora	NC_038123.1
<i>P. oerstedii</i> Mast.	Passiflora	NC_038124.1
<i>P. quadrangularis</i> L.	Passiflora	NC_038126.1
<i>P. retipetala</i> Mast., 1893	Passiflora	NC_038188.1
<i>P. serratifolia</i> L.	Passiflora	NC_038129.1
<i>P. serratodigitata</i> L.	Passiflora	NC_038127.1
<i>P. vitifolia</i> Kunth	Passiflora	NC_038128.1
<i>P. affinis</i> Engelm.	Decaloba	NC_043823.1
<i>P. auriculata</i> Kunth (FG)	Decaloba	NC_038119.1
<i>P. auriculata</i> Kunth (S)	Decaloba	MF807935.1
<i>P. biflora</i> Lam.	Decaloba	NC_038120.1
<i>P. filipes</i> Benth.	Decaloba	NC_043822.1
<i>P. jatunsachensis</i> Schwerdtf.	Decaloba	NC_043813.1
<i>P. lutea</i> L.	Decaloba	NC_043815.1
<i>P. microstipula</i> L.E.Gilbert & J.M.MacDougal	Decaloba	NC_043827.1
<i>P. misera</i> Kunth Kunth, 1816	Decaloba	NC_043821.1
<i>P. rufa</i> Feuillet & J.M.MacDougal	Decaloba	NC_043817.1
<i>P. suberosa</i> L. (USA)	Decaloba	NC_043814.1
<i>P. tenuiloba</i> Engelm.	Decaloba	NC_043816.1
<i>P. cirrhiflora</i> A.Juss.	Deidamioides	MN545921

<i>P. arbelaezii</i> L.Uribe	Deidamioides	NC_043819.1
<i>P. contracta</i> Vitta	Deidamioides	NC_043818.1
<i>P. obovata</i> Killip ex Standl.	Deidamioides	NC_043824.1
<i>P. pittieri</i> Mast.	Astrophea	NC_038125.1
<i>P. tetrandra</i> Banks ex DC.	Tetrapathea	NC_043820.1
<i>Adenia mannii</i> (Mast.) Engl.	-	NC_043791.1
<i>Populus trichocarpa</i> Torr. & A.Gray	-	NC_009143.1

Supplementary Table S3. Plastid genes used in the maximum likelihood phylogenetic analysis, clustered according to the best substitution model calculated by ModelFinder.

Best Model	Genes
K3Pu+F+I	ndhB, psbE, psbF, psbM
TVM+F+G4	atpF, cemA, rpl33, rpl36, rpoC2, rps2, rps3, rps8, rps19
TVM+F+I+G4	atpA, atpB, atpE, atpI, atpH, ccsA, matK, ndhA, ndhC, ndhD, ndhE, ndhF, ndhG, ndhI, ndhH, ndhJ, ndhK, petA, petB, petD, petG, petN, petL, psaA, psaB, psaC, psaI, psaJ, psbA, psbB, psbC, psbD, psbI, psbJ, psbK, psbL, psbN, psbT, psbZ, rpl2, rpl14, rpoB, rpoC1, rps4, rps12, ycf3
TPM3u+F	ycf4
TPM3u+F+G4	rpl23, rps11, rps15
TPM3u+F+I+G4	psbH, rbcL, rps14

Supplementary Table S4. Plastid genes used in the Bayesian phylogenetic inference, clustered according to the best substitution model calculated by PartitionFinder2.

Best Model	Genes
GTR+G	atpF, rpl14, rpl33, rpl36, rps2, rps3, rps4, rps8, rps11, rps14, rps15, rps19, ycf4
GTR+I	ndhB, petN, psbE, psbF, psbN
GTR+I+G	atpA, atpB, atpE, atpI, atpH, ccsA, cemA, matK, ndhA, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK, petA, petB, petD, petG, petL, psaA, psaB, psaC, psaI, psaJ, psbA, psbB, psbC, psbD, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ, rbcL, rpl2, rpl23, rpoB, rpoC1, rpoC2, rps12, ycf3

Chapter IV

The complete plastome of *Passiflora cirrhiflora* A. Juss.: structural features, RNA editing sites, hotspots of nucleotide diversity and molecular markers within the subgenus *Deidamioides*

Túlio Gomes Pacheco¹, Amanda de Santana Lopes¹, José Daniel de Oliveira¹, Wagner Campos Otoni², Eduardo Balsanelli³, Fábio de Oliveira Pedrosa³, Emanuel Maltempi de Souza³, Marcelo Rogalski^{1*}

¹ Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa-MG, Brasil.

² Laboratório de Cultura de Tecidos Vegetais, Departamento de Biologia Vegetal, BIOAGRO, Universidade Federal de Viçosa, Viçosa-MG, Brasil.

³ Departamento de Bioquímica e Biologia Molecular, Núcleo de Fixação Biológica de Nitrogênio, Universidade Federal do Paraná, Curitiba-PR, Brasil.

*Corresponding author

E-mail address: rogalski@ufv.br

Published in:

Brazilian Journal of Botany (2020) 43, 839-853

<https://doi.org/10.1007/s40415-020-00655-y>



The complete plastome of *Passiflora cirrhiflora* A. Juss.: structural features, RNA editing sites, hotspots of nucleotide diversity and molecular markers within the subgenus *Deidamioides*

Túlio Gomes Pacheco¹ · Amanda de Santana Lopes¹ · José Daniel de Oliveira¹ · Wagner Campos Otoni² · Eduardo Balsanelli³ · Fábio de Oliveira Pedrosa³ · Emanuel Maltempi de Souza³ · Marcelo Rogalski¹

Received: 11 April 2020 / Revised: 11 September 2020 / Accepted: 14 September 2020
© Botanical Society of Sao Paulo 2020

Abstract

The family Passifloraceae contains approximately 1000 species distributed in 36 genera. *Passiflora*, a species-rich genus, is divided into five subgenera: *Astrophea*, *Decaloba*, *Deidamioides*, *Passiflora*, and *Tetrapathea*. *Passiflora cirrhiflora* A. Juss. (subgenus *Deidamioides*) is an Amazonian species occurring in North Brazil, Venezuela, Guiana, Suriname, and French Guiana. Plastomes of the subgenus *Deidamioides* have demonstrated a complex pattern of evolution concerning phylogenetic aspects, size, rearrangements, gene losses, and pseudogenization. Therefore, we completely sequenced the plastome of *P. cirrhiflora* and characterized it in detail. The *P. cirrhiflora* plastome is a DNA molecule of 163,365 bp and contains 109 unique genes (75 protein-coding genes, 30 tRNAs, and 4 rRNAs). The *infA* and *rps7* are pseudogenes in *P. cirrhiflora* and the loss of the *rps16* and *rpl22* genes is a common feature shared by all plastomes of the genus *Passiflora* sequenced to date. Comparative analyses revealed a considerable structural variation at the IR borders, which consequently affected plastome size and structure within the subgenus *Deidamioides*. According to our prediction analysis, RNA editing sites are highly conserved within the subgenus *Deidamioides*, which significantly differs from the subgenus *Passiflora*. A total of 230 SSRs and six hotspots of nucleotide polymorphism were detected in the plastome of *P. cirrhiflora*. Taken together, the complete plastome sequence of *P. cirrhiflora* is useful data for several studies in different areas such as phylogeny, genetic, and evolution.

Keywords Cytoplasmic inheritance · Passifloraceae · Plastome evolution · Plastome rearrangements

1 Introduction

The family Passifloraceae contains approximately 1000 species distributed in 36 genera. *Passiflora* is the largest genus and comprises more than 560 species (MacDougal and Feuillet 2004; Rocha et al. 2020). The current infrageneric classification divides the genus *Passiflora* into five subgenera: *Astrophea* (DC.) Mast., *Decaloba* (DC.) Rchb., *Deidamioides* (Harms) Killip, *Passiflora* Feuillet & MacDougal, and *Tetrapathea* (DC.) P.S. Green (MacDougal and Feuillet 2004; Krosnick et al. 2009, 2013). The number of *Passiflora* species are non-uniformly distributed into the subgenera: *Passiflora* (~ 250), *Decaloba* (~ 230), *Astrophea* (~ 60), *Deidamioides* (~ 14) and *Tetrapathea* (~ 3) (Krosnick et al. 2009, 2013; Rabah et al. 2019; Rocha et al. 2020). *Passiflora cirrhiflora* A. Juss. is a native species from the Amazon forest occurring predominantly in North

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40415-020-00655-y>) contains supplementary material, which is available to authorized users.

✉ Marcelo Rogalski
rogalski@ufv.br

- ¹ Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Avenida Purdue, s/nº, Campus Universitário, Edif. CCB II – Centro de Ciências Biológicas II, Viçosa, MG 36570.900, Brazil
- ² Laboratório de Cultura de Tecidos Vegetais, Departamento de Biologia Vegetal, BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG, Brazil
- ³ Departamento de Bioquímica e Biologia Molecular, Núcleo de Fixação Biológica de Nitrogênio, Universidade Federal do Paraná, Curitiba, PR, Brazil

Brazil, Venezuela, Guiana, Suriname, and French Guiana (Killip 1938; Boggan et al. 1997; Hokche et al. 2008).

Normally, land plant plastomes show a quadripartite structure and encode a conserved set of genes divided into two groups. The first group comprises components of the photosynthetic machinery such as photosystem I, photosystem II, the cytochrome b6/f complex, and the ATP synthase. The second group includes the genes required for plastid gene expression such as subunits of a plastid-encoded RNA polymerase, rRNAs, tRNAs, and ribosomal proteins (Rogalski et al. 2015; Daniell et al. 2016). Moreover, plastid genomes encode several essential genes involved in other cellular function than photosynthesis such as *ycf1* (a subunit of the TIC/TOC machinery related to the plastid protein import apparatus; Kikuchi et al. 2013), *ycf2* (a component of 2-MD AAA-ATPase complex; Kikuchi et al. 2018), *clpP* (a proteolytic subunit of the ATP-dependent Clp protease related to plastid protein homeostasis; Kuroda and Maliga 2003), *accD* (β -carboxyl transferase subunit of the eubacteria-like multisubunit acetyl-CoA carboxylase involved in de novo fatty acid biosynthesis; Kode et al. 2005) and *trnE-UUC* (plastid-encoded transfer RNA for glutamate also involved in the control of the plastid nuclear-encoded polymerase and the biosynthesis of chlorophyll and heme group; Schön et al. 1986; Hanaoka et al. 2005).

The plastome structure, gene content, gene order, and gene function are well conserved in most angiosperms; however, there are some exceptions. The literature highlights some angiosperm families as notorious examples of highly rearranged plastomes such as Annonaceae (Blazier et al. 2016), Berberidaceae (Ma et al. 2013), Cactaceae (Solórzano et al. 2019), Campanulaceae (Haberle et al. 2008), Ericaceae (Fajardo et al. 2013), Geraniaceae (Weng et al. 2014), Fabaceae (Schwarz et al. 2015), Passifloraceae (Rabah et al. 2019; Shrestha et al. 2019; Pacheco et al. 2020a), Plantaginaceae (Zhu et al. 2016) and Trochodendraceae (Sun et al. 2013).

Complete plastome sequences bring several molecular markers, which have vast potential to be used in genetic studies in natural populations such as interspecific and intraspecific variability, gene flow, genetic divergence and genetic diversity (Besnard et al. 2011; Rogalski et al. 2015; Qiao et al. 2016). Moreover, plastome sequences support phylogenetic analyses with high accuracy among subfamilies, tribes, genus, and species (Vieira et al. 2016a; Lopes et al. 2018c; Shrestha et al. 2019; Pacheco et al. 2020a). Complete plastome sequences are also useful data for studies concerning evolution of genes (i.e., RNA editing, functionality, pseudogenization, and loss), rearrangements, recombination events, positive selection, and phylogeography (Vieira et al. 2016b; Lopes et al. 2018a; Pacheco et al. 2019; Stefenon et al. 2019).

Phylogenetic analyses containing species of the subgenus *Deidamioides* revealed a complex polyphyletic pattern of relationship among species within the genus *Passiflora* (Shrestha et al. 2019; Pacheco et al. 2020a). The plastomes of the subgenus *Deidamioides* have demonstrated exceptional characteristics related to the structure affecting the size of the regions (LSC, SSC, and IR), rearrangements, gene losses, and pseudogenization (Shrestha et al. 2019). Aiming to contribute to the understanding regarding plastome evolution within the subgenus *Deidamioides*, we completely sequenced and analyzed in detail the plastome of *P. cirrhiflora*. The *P. cirrhiflora* plastome is a DNA molecule of 163,365 bp which is among the sizes observed for the species of this subgenus (ranging from 151,701 to 170,568 bp). The *P. cirrhiflora* plastome differed in the sizes of the IR regions and the number of genes when compared with other species of the subgenus *Deidamioides*. Several evolutionary aspects related to structural rearrangements, gene content, and prediction of RNA editing sites are analyzed, explained and discussed. Furthermore, SSRs, and high polymorphic regions detected in the plastome of *P. cirrhiflora* are useful for diverse genetic studies in this species and the subgenus *Deidamioides*.

2 Materials and methods

Plant material, chloroplast isolation, and plastid DNA extraction – *Passiflora cirrhiflora* plants obtained from seed germination were cultivated under greenhouse conditions in the Department of Plant Biology at the Federal University of Viçosa, Viçosa-MG, Brazil. Young leaves were collected and stored in accordance with Lopes et al. (2018c). The chloroplast isolation and plastid DNA extraction procedures were carried out as described by Vieira et al. (2014).

Sequencing, assembly, annotation, and data archiving statement – The sequencing library was prepared with approximately 1 ng of plastid DNA using the sample preparation kit NexteraXT (Illumina Inc., San Diego, CA), according to the manufacturer's instructions. Libraries were sequenced using MiSeq Reagent Kit v3 (600 cycles) on Illumina MiSeq Sequencer (Illumina Inc., San Diego, California, USA). The paired-end reads (total of 548,214 reads), with an average length of 186.9 bp, were trimmed under the threshold with a probability of error < 0.05. The trimmed reads (547,970 reads, average length of 186.4 bp) were de novo assembled in contigs using CLC Genomics Workbench v6.5 (CLC Bio, Aarhus, Denmark). The four contigs used for assembly of the *P. cirrhiflora* plastome ranged from 446.81 to 212.88 of average coverage. These four contigs were connected with overlapping terminal sequences to yield the entire and circular plastome. Our reads were also mapped to other *Passiflora*

plastomes as references, using CLC Genomics. The assemblies generated by this method were then used to confirm the assembly of the four contigs obtained by de novo method.

Initial annotation of the plastid genome was performed using Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004). From this initial annotation, putative start and stop codons, and intron positions were determined by comparisons with homologous genes from other *Passiflora* plastid genomes available in the Genbank database (Clark et al. 2016). The tRNA genes were further verified using the software tRNAscan-SE (Lowe and Chan 2016). A physical map of the plastid circular genome was drawn by Organellar Genome DRAW (OGDRAW) (Greiner et al. 2019). The complete nucleotide sequence of *P. cirrhiflora* plastome was deposited in the GenBank database under accession number MN545921.

Structural analysis of the *P. cirrhiflora* plastome and comparison with other species of the subgenus *Deidamioides*

– The software Mauve Genome Alignment v2.3.1 (MAUVE) (Darling et al. 2004) was used here to analyze and compare the plastome structure of *P. cirrhiflora* with the other three species of the subgenus *Deidamioides* sequenced previously and published in another study (Shrestha et al. 2019; Table 1). *Populus trichocarpa* (family Salicaceae; Genbank accession number: NC_009143.1; Tuskan et al. 2006) was used as a reference species, representing the general or common plastome structure found in most angiosperms. Furthermore, events of IR expansion and contraction were compared between these species. Firstly, the localization of the IR borders of each species was determined by the software Geseq—Annotation of Organellar Genomes (Tillich et al. 2017). Posteriorly, linear structures of the plastomes were generated using the software OGDRAW (Greiner et al. 2019). Finally, linear structures of the plastomes were used to draw a schematic representation of the events of IR expansion and contraction.

Detection of SSRs, sliding window analysis of hotspots, and prediction of RNA editing sites

– Simple sequence repeats (SSRs) were detected in the plastome of *P. cirrhiflora* using the MicroSATellite (MISA) Perl script (Thiel et al. 2003), with thresholds of eight repeat units for mononucleotide

SSRs, four repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta- and hexanucleotide SSRs. The IRa was omitted of this analysis to prevent the over-representation of the IRs sequences.

The hotspots of nucleotide diversity were also investigated in the plastome of *P. cirrhiflora* using the sliding window analysis by comparison with *P. contracta* (close phylogenetic relationship). Firstly, the plastomes of these two species were reorganized (due to differential IR expansion and contraction events present in *P. contracta*; see Shrestha et al. 2019 or Fig. 3) to start in the *trnH*-GUG gene (in the LSC region); and to end in the *rps15* gene (at the end of the SSC region), omitting the IRa for both species. Additionally, we also removed the *psbA* gene, and small adjacent sequence, present only in the IRb of *P. contracta*. Posteriorly, the whole sequence obtained from the reorganized plastomes (in synteny) was aligned using MAFFT v.7 and the sliding window analysis was performed using the DnaSP v.6 software (Rozas et al. 2017). The window length and the step size were set to 300 bp and 75 bp, respectively.

Potential RNA editing sites in protein-coding genes of *P. cirrhiflora* plastome were predicted by the program predictive RNA editor for plants (PREP) suite (Mower 2009), which uses 35 reference genes for detecting RNA editing sites in plastid genomes. The cutoff value was set at 0.8 and the reference genes included here were *atpA*, *atpB*, *atpF*, *atpI*, *ccsA*, *matK*, *ndhA*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *petB*, *petD*, *petG*, *petL*, *psaB*, *psaI*, *psbB*, *psbE*, *psbF*, *psbL*, *rpl2*, *rpl23*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps2*, *rps8*, *rps14*, and *ycf3*. Aiming evolutionary comparison, RNA editing sites were also predicted for the other three species of the subgenus *Deidamioides*, using the same parameters.

3 Results

General features and gene content of the *P. cirrhiflora* plastome

– The *P. cirrhiflora* plastome is a DNA molecule of 163,365 bp in length and exhibits the quadripartite structure typically found in most angiosperm species. It includes a large single-copy region (LSC) of 86,134 bp flanked on each side by inverted repeats (IRs) of 31,692 bp with a small single-copy region (SSC) of 13,846 bp joining the IRs (Fig. 1).

Table 1 Plastome features among species within the subgenus *Deidamioides*. The species *Passiflora cirrhiflora* was completed sequenced and assembled in this study

Species	Total size (Kb)	LSC (Kb)	SSC (Kb)	IR (Kb)	GC (%)	Number of genes	GenBank
<i>P. arbelaezzi</i> L.Uribe	170,568	87,452	12,764	35,176	36.9	110	NC_043819.1
<i>P. cirrhiflora</i> A. Juss	163,365	86,135	13,846	31,692	36.6	109	MN545921
<i>P. contracta</i> Vitta	166,766	87,597	13,723	32,723	36.8	107	NC_043818.1
<i>P. obovata</i> Killip	151,701	84,697	29,586	18,709	35.7	107	NC_043824.1

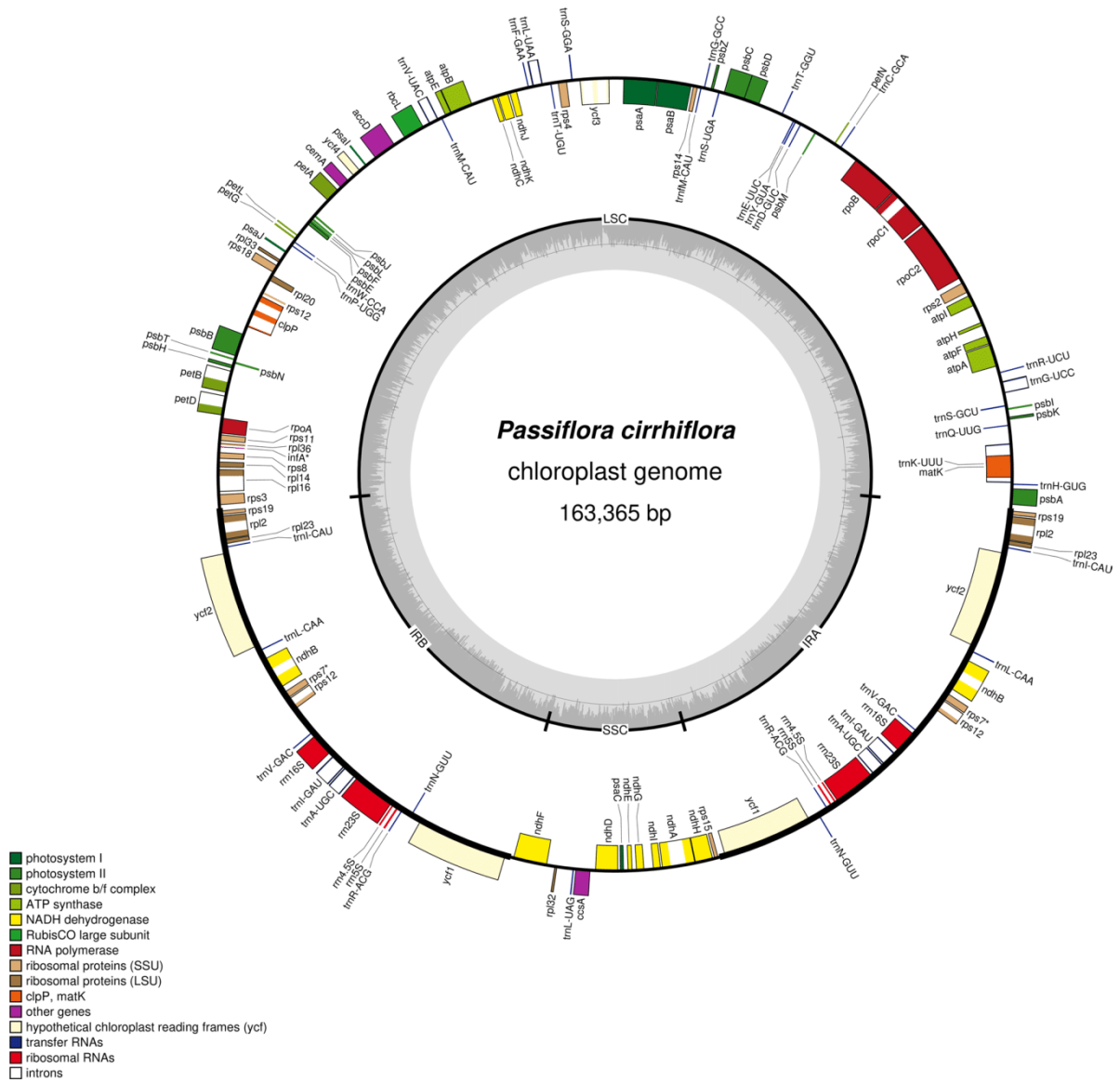


Fig. 1 Gene map of the *Passiflora cirrhiflora* plastome. Genes drawn inside the circle are transcribed in the clockwise direction, and genes drawn outside are transcribed in the counterclockwise direction. Different functional groups of genes are color-coded. The darker gray in the inner circle corresponds to GC content, and the lighter gray corresponds to AT content. LSC, Large Single Copy; SSC, Small Single Copy; IRA/B, Inverted Repeats A/B

If we compare the *P. cirrhiflora* plastome with the plastomes of other species of the subgenus *Deidamioides*, its total size is very similar to *Passiflora contracta* Vitta plastome (166,766 bp), smaller than the plastome of *Passiflora arbelaezii* L.Uribe (170,568 bp) and larger than the plastome of *Passiflora obovata* Killip (151,701 bp) (Table 1). Concerning the size of the plastome regions, we found significant differences between the species of the subgenus *Deidamioides*. The LSC region of *P. cirrhiflora* is larger than *P. obovata*

(84,697 bp), and smaller than *P. arbelaezii* (87,452 bp) and *P. contracta* (87,597 bp). The SSC region of *P. cirrhiflora* is larger than *P. arbelaezii* (12,764 bp) and *P. contracta* (13,723 bp), and smaller than *P. obovata* (29,586 bp). Regarding the IRs of *P. cirrhiflora* plastome, they are larger than *P. obovata* (18,709 bp), and smaller than *P. arbelaezii* (35,176 bp) and *P. contracta* (32,723 bp) (Table 1). The GC content of *P. cirrhiflora* plastome (36.6%) is similar to the values observed for the other species of the subgenus

Deidamioides, *P. arbelaezii* (36.9%), *P. contracta* (36.8%) and *P. obovata* (35.7%) (Table 1). The plastome of *P. cirrhiflora* contains 109 unique genes, whereas the plastomes of *P. arbelaezii*, *P. contracta* and *P. obovata* encode 110, 107 and 107 genes, respectively (Fig. 1 and Table 2). The 109 genes of *P. cirrhiflora* plastome correspond to 75 unique protein-coding genes (six are completely duplicated and one is partially duplicated), 30 unique tRNA genes (seven duplicated) and 4 unique rRNA genes (all duplicated). Fourteen unique genes possess one intron (six tRNA genes and eight protein-coding genes) and two contain two introns (*clpP* and *ycf3*). The *rps16* and *rpl22* genes are absent from the *P. cirrhiflora* plastome as well as in all other species of the genus *Passiflora* (Shrestha et al. 2019). The gene *infA* is a pseudogene in *P. cirrhiflora* due to the presence of premature stop codons in its coding sequences, whereas it was lost in the other species of the subgenus *Deidamioides* (Shrestha et al. 2019). On the other hand, the *rps7* has an intact coding sequence in *P. arbelaezii* (resulting in 110 unique genes in this plastome;

Table 1), whereas in the other species analyzed here it was lost or is a pseudogene. Moreover, the *rpl32* and *rpl20* genes were lost from the plastome of *P. obovata* and *P. contracta* (Shrestha et al. 2019), which result in a total of 107 encoded genes (Table 1). These two genes are intact and apparently functional in *P. cirrhiflora* and *P. arbelaezii* plastomes.

Comparison of plastome structure and IR borders among *P. cirrhiflora* and other species of the subgenus *Deidamioides*

Genome alignment analyzed by MAUVE showed two rearrangements in the plastome of *P. cirrhiflora* in comparison with *Populus trichocarpa* (Fig. 2a). The first corresponds to an inversion of 1.6 kb at the beginning of the LSC region, which changed the order and direction of the *trnH*-GUG and *psbA* genes (1, Fig. 2a). The second one is an inversion in the LSC containing approximately 3 kb from the *trnV*-UAC gene to the *atpB* gene (2, Fig. 2a). A multiple plastome alignment involving *P. cirrhiflora* and the other *Deidamioides* species was also performed in MAUVE (Fig. 2b). All

Table 2 List of genes identified in the plastome of *P. cirrhiflora*

Group of gene	Name of gene
<i>Gene expression machinery</i>	
Ribosomal RNA genes	<i>rrn16^b</i> ; <i>rrn23^b</i> ; <i>rrn5^b</i> ; <i>rrn4.5^b</i>
Transfer RNA genes	<i>trnA</i> –UGC ^{ab} ; <i>trnC</i> –GCA; <i>trnD</i> –GUC; <i>trnE</i> –UUC; <i>trnF</i> –GAA; <i>trnFM</i> –CAU; <i>trnG</i> –UCC ^a ; <i>trnG</i> –GCC; <i>trnH</i> –GUG; <i>trnI</i> –CAU ^b ; <i>trnI</i> –GAU ^{ab} ; <i>trnK</i> –UUU ^a ; <i>trnL</i> –CAA ^b ; <i>trnL</i> –UAA ^a ; <i>trnL</i> –UAG; <i>trnM</i> –CAU; <i>trnN</i> –GUU ^b ; <i>trnP</i> –UGG; <i>trnQ</i> –UUG; <i>trnR</i> –ACG ^b ; <i>trnR</i> –UCU; <i>trnS</i> –GCU; <i>trnS</i> –UGA; <i>trnS</i> –GGA; <i>trnT</i> –UGU; <i>trnT</i> –GGU; <i>trnV</i> –GAC ^b ; <i>trnV</i> –UAC ^a ; <i>trnW</i> –CCA; <i>trnY</i> –GUA
Small subunit of ribosome	<i>rps2</i> ; <i>rps3</i> ; <i>rps4</i> ; <i>rps8</i> ; <i>rps11</i> ; <i>rps12^a</i> ; <i>rps14</i> ; <i>rps15</i> ; <i>rps18</i> ; <i>rps19^b</i>
Large subunit of ribosome	<i>rpl2^{ab}</i> ; <i>rpl14</i> ; <i>rpl16^a</i> ; <i>rpl20</i> ; <i>rpl23^b</i> ; <i>rpl32</i> ; <i>rpl33</i> ; <i>rpl36</i>
DNA-dependent RNA polymerase	<i>rpoA</i> ; <i>rpoB</i> ; <i>rpoC1^a</i> ; <i>rpoC2</i>
<i>Genes for photosynthesis</i>	
Subunits of photosystem I (PSI)	<i>psaA</i> ; <i>psaB</i> ; <i>psaC</i> ; <i>psaI</i> ; <i>psaJ</i> ; <i>ycf3^a</i> ; <i>ycf4</i>
Subunits of photosystem II (PSII)	<i>psbA</i> ; <i>psbB</i> ; <i>psbC</i> ; <i>psbD</i> ; <i>psbE</i> ; <i>psbF</i> ; <i>psbH</i> ; <i>psbI</i> ; <i>psbJ</i> ; <i>psbK</i> ; <i>psbL</i> ; <i>psbM</i> ; <i>psbN</i> ; <i>psbT</i> ; <i>psbZ</i>
Subunits of cytochrome <i>b₆f</i>	<i>petA</i> ; <i>petB^a</i> ; <i>petD^a</i> ; <i>petG</i> ; <i>petL</i> ; <i>petN</i>
Subunits of ATP synthase	<i>atpA</i> ; <i>atpB</i> ; <i>atpE</i> ; <i>atpF</i> ; <i>atpH</i> ; <i>atpI</i>
Subunits of NADH dehydrogenase	<i>ndhA^a</i> ; <i>ndhB^{ab}</i> ; <i>ndhC</i> ; <i>ndhD</i> ; <i>ndhE</i> ; <i>ndhF</i> ; <i>ndhG</i> ; <i>ndhH</i> ; <i>ndhI</i> ; <i>ndhJ</i> ; <i>ndhK</i>
Large subunit of Rubisco	<i>rbcL</i>
<i>Others genes</i>	
Maturase	<i>matK</i>
Envelope membrane protein	<i>cemA</i>
Subunit of acetyl-CoA carboxylase	<i>accD</i>
C-type cytochrome synthesis gene	<i>ccsA</i>
ATP-dependent Protease	<i>clpP^a</i>
Component of TIC complex	<i>ycf1^b</i>
Component of 2-MD AAA-ATPase complex (Kikuchi et al. 2018)	<i>ycf2^b</i>
Pseudogenes	<i>infA</i> ; <i>rps7^b</i>
Genes absent	<i>rps16</i> ; <i>rpl22</i>

^aGenes containing introns

^bDuplicated genes

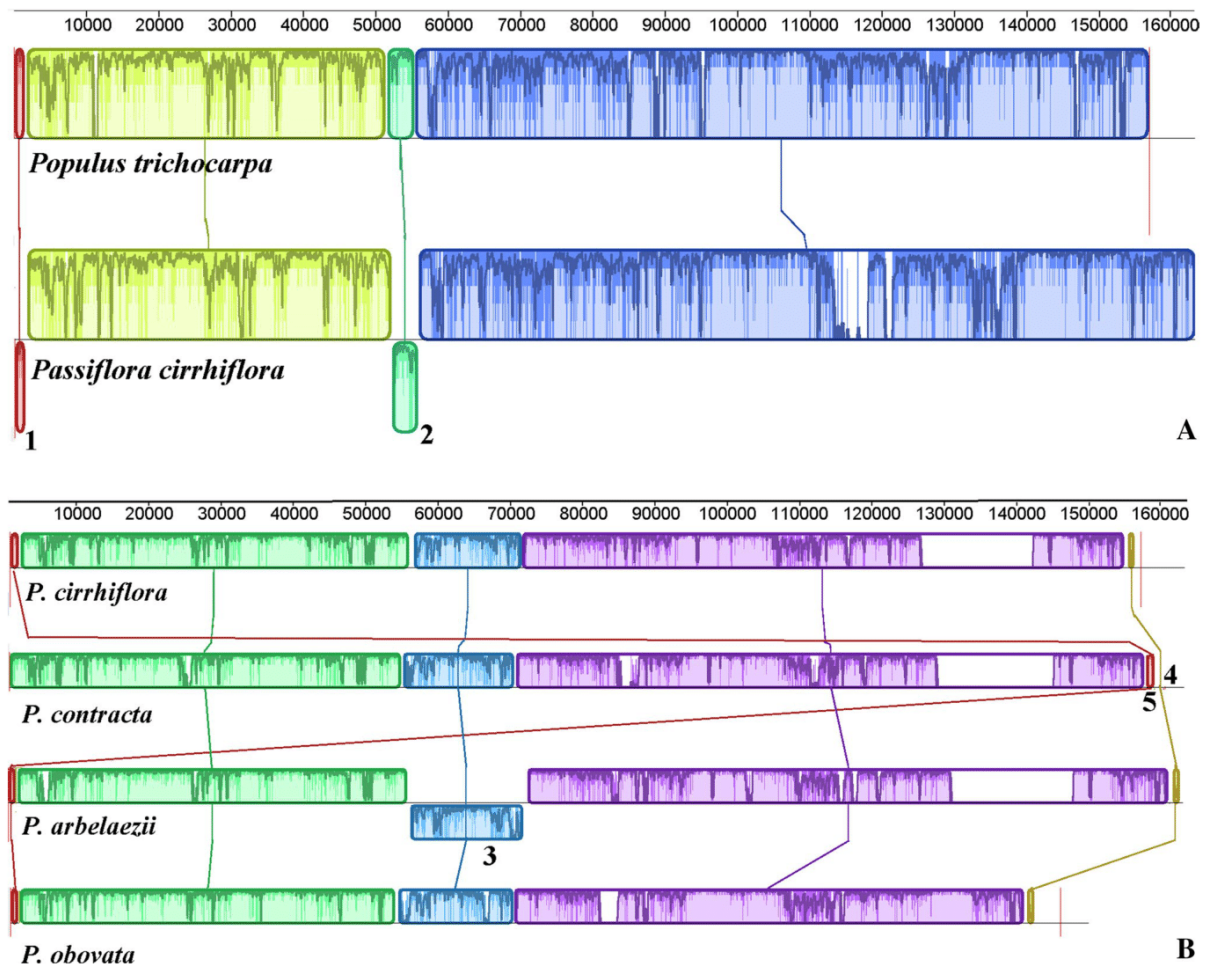


Fig. 2 **a** Alignment among *Passiflora cirrhiflora* and *Populus trichocarpa* (as reference). Locally collinear blocks (LCBs) are color-coded. The LCBs related to inversions are LCB 1 (including *psbA* and *trnH-GUG* genes) and LCB 2 (encompassing from *trnV-UAC* to *atpE* genes). **b** Multiple alignments of four plastomes of the subgenus *Deidamioides*. LCB 3 represents an inversion, encompassing from *accD* to *clpP* genes. The LCBs 4 and 5 are related to contraction and expansion events at the IRa junctions, respectively

species of the subgenus *Deidamioides* share similar rearrangements with *P. cirrhiflora* plastome in comparison with *Populus*. The plastome of *P. contracta* specially showed a high synteny (very similar gene order) with *P. cirrhiflora*, except by two specific changes in the IR borders of the former species (4 and 5, Fig. 2b). Similarly, the plastome of *P. obovata*, except for the contraction of IRs, also showed synteny with *P. cirrhiflora* and *P. contracta*, indicating that no additional inversion is present in these three plastomes. On other hand, a specific and additional inversion was characterized in the LSC region of the *P. arbelaezii* plastome, which encompasses the sequence from the *accD* gene to the *clpP* gene (3, Fig. 2b).

Concerning IR borders, an event of IR expansion was detected in *P. cirrhiflora* compared with the plastome of *Populus*, which resulted in the inclusion of the entire *ycfI* gene in the IRs in this plastome (E1, Fig. 3). This event can be also detected in the plastomes of *P. contracta* and *P. arbelaezii* (E1, blue circle, Fig. 3). Additionally, other two events of IR expansion occurred in these species. The first event included the entire *psbA* gene in the IRs of *P. contracta* (E2, red circle, Fig. 3). The second event is an IR expansion that resulted in the inclusion of the complete sequence of the *rps15* gene and a small fragment of the *ndhH* gene in the IRs of *P. arbelaezii* (E3, green circle, Fig. 3). Moreover, two events of IR contraction were also visualized in two species

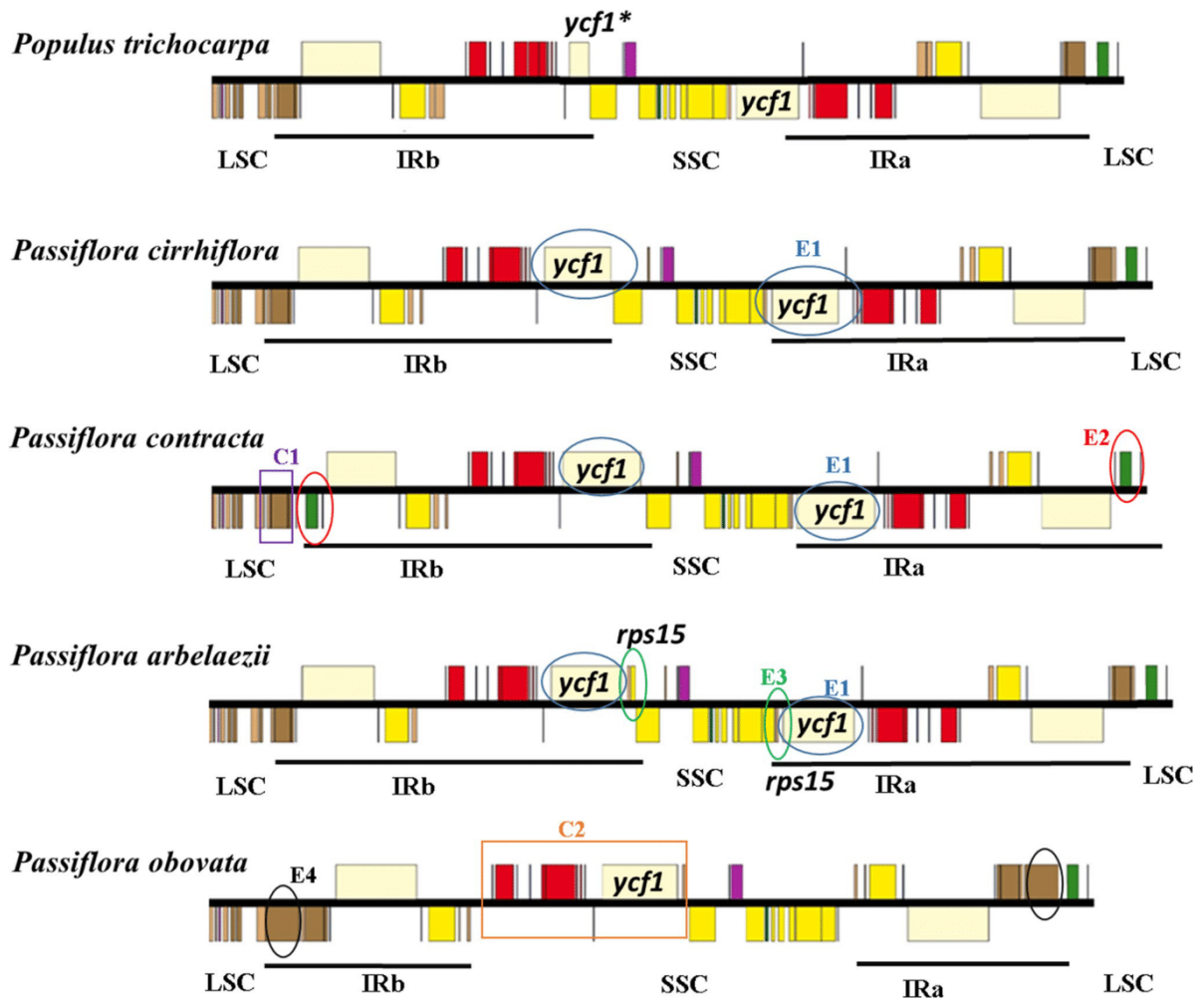


Fig. 3 Schematic representations of IR regions and their boundaries among species of the subgenus *Deidamioides* and *P. trichocarpa* (used as a reference species). The plastid genes correspond to color-coded boxes according to Fig. 1. Boxes above and below the horizontal line represent genes transcribed in forward and reverse directions, respectively. E1 (blue circle) indicates an IR expansion event detected in the *Deidamioides* species, in comparison with *Populus*. This expansion included the full *ycf1* sequence in the IR (blue circle) of these species. E2 (red circle) indicates a specific IR expansion event that included and duplicated the *psbA* in *P. contracta*. This species also shows a specific event of IR contraction (C1, purple square), culminating in the exclusion of the genes *rpl23-rpl2-rps19* of the IRs. E3 (green circle) indicates an IR expansion event that included and duplicated the *rps15* and part of the *ndhH* gene in *P. arbelaezii*. C2 (orange square) indicates a large IR contraction event present in the *P. obovata* plastome, excluding several genes (rRNAs and *ycf1*) from its IRs. Furthermore, a small event of IR expansion (E4, black square) is also characterized in this species, duplicating the whole *rps3* gene. * Fragment of the *ycf1* gene

of the subgenus *Deidamioides*. The first event is detected at the LSC/IRb boundary of *P. contracta*, which culminated in the exclusion of the *rps19-rpl2-rpl23* genes from the IRs (C1, purple square, Fig. 3). The second event is a large IR contraction at the IR/SSC boundary of *P. obovata*, which excluded several genes (including *ycf1* and rRNA genes) from its IRs (C2, orange square, Fig. 3). Furthermore, a specific event of IR expansion also occurred at the LSC/IRb border of this species, extending to the *rps3* gene. As a result of several events of IR expansion and contraction described

here, the plastomes of the four species of the subgenus *Deidamioides* diverge in size (bp) and gene content in their IRs (Table 1 and Fig. 3).

Prediction of RNA editing sites – The PREP software identified 60 putative RNA editing sites in the plastome of *P. cirrhiflora* (Supplementary Table S1). All the RNA editing sites occur in the first or the second codon position and all nucleotide changes were from cytidine (C) to uridine (U).

The *ndh* genes contain most predicted RNA editing sites, totalizing 31 (52%) from 60 predicted sites. These 31 sites are distributed among five genes: *ndhB* (11 sites), *ndhD* (10), *ndhA* (5), *ndhG* (3), and *ndhF* (2). The *rpo* genes constitute the second gene class containing high number of editing sites (11), which are distributed among the genes *rpoC2* (6 sites), *rpoC1* (2), *rpoB* (2), and *rpoA* (1). Besides *rpo* and *ndh* genes, the *matK* (6 sites) gene contains the highest number of RNA editing sites. The other RNA editing sites remaining (12) are distributed among nine genes: *atpA* (2 sites), *aptI* (1), *petD* (1), *psaI* (2), *psbE* (1), *psbF* (1), *rpl2* (1), *rps2* (1) and *rpl14* (2). Most RNA editing sites change the encoded amino acid from polar to nonpolar (31 of 60 sites), specially changing from serine (S) to leucine (L) (17 sites). Five sites change the amino acid from nonpolar to polar (all from proline to serine). From the other 24 sites, 20 do not alter either polarity or charge (nonpolar to nonpolar), and four sites change the amino acid from polar positively charged to polar uncharged (i.e., from histidine to tyrosine and from arginine to cysteine).

Of all the RNA editing sites predicted here to *P. cirrhiflora* plastome, 49 of them (82%) are shared by all species of *Deidamioides* analyzed here (Supplementary Table S1). Concerning the other 11 sites, seven are also shared with *P. contracta* but are absent in *P. obovata* and *P. arbelaezii*, and one site is shared by all species except *P. contracta* (amino acid position 777 of the *rpoC2* gene). The last three sites were only identified in the plastome of *P. cirrhiflora* (highlight in red text, Supplementary Table S1). One of them is localized in the amino acid position 424 of the *atpA* gene and converts a proline codon (CCT) to a serine codon (TCT). The second site is found in the *matK* gene (amino acid position 209) and changes a leucine codon (CTC) to phenylalanine codon (TTC). Finally, the last site is located in the *rpoC2* gene (amino acid position 1264) and converts an alanine codon (GCT) to a valine codon (GTT). On other

hand, the other three species of the subgenus *Deidamioides* already contain a thymidine (T) instead of a cytidine (C) in these three sites, which dismisses the need for RNA editing of these sites.

Identification of SSRs and hotspots of nucleotide divergence

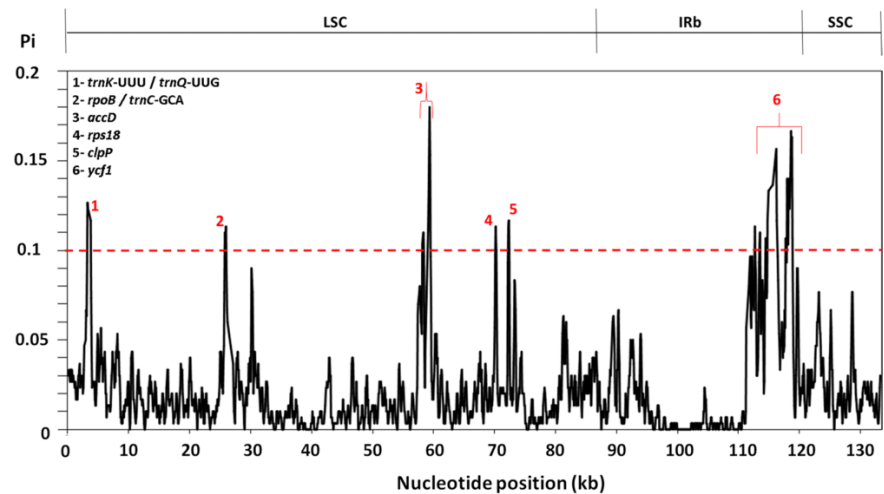
The occurrence, type, size and localization of SSRs in the plastome of *P. cirrhiflora* were analyzed, revealing 230 repeats (Table 3 and Supplementary Table S2). Most SSRs are constituted of mononucleotides (156), followed by di- (61), tri- and hexanucleotides (2). The bases A or T compose 208 (90.5%) of the 230 SSRs detected here (Table 3). Concerning the localization of the SSRs in the plastome of *P. cirrhiflora*, 166 are in the LSC region, 36 in the IRs, and 28 in the SSC region (Supplementary Table S2). Furthermore, 148 SSRs are localized in intergenic spacers (IGSs), 53 in coding sequences (CDS), and 29 in introns. The IGSs with higher number of SSRs correspond to the spacers between the genes: *trnK-UUU/trnQ-UUG* (11 SSRs); *atpF/lptH* (10); *ndhC/atpB* (6); *trnL-UAA/trnF-GAA* (6); *rpoB/trnC-GCA* (5); *trnR-UCU/atpA* (5); *ycf1/ndhF* (5); and *rpl32/trnL-UAG* (5). The SSRs detected in CDS (53) are distributed in 24 different genes, with the higher number occurring in the *ycf1* (10 SSRs), *ycf2* (6), *rpoC2* (4) and *clpP* (4) genes. Finally, among the 29 SSRs found in the introns, most of them are located in the introns of the *clpP* (6 SSRs), *ndhA* (6) and *rpl16* (4) genes.

Our sliding windows analysis (Fig. 4) highlights six plastome regions as hotspots of nucleotide divergence among *P. cirrhiflora* and *P. contracta* (species related phylogenetically in the trees of Shrestha et al. 2019 and Pacheco et al. 2020a). These hotspots correspond to two intergenic regions (*trnK-UUU/trnQ-UUG* and *rpoB/trnC-GCA*) and four genes (*accD*, *rps18*, *clpP*, and *ycf1*). The hotspot detected in the *clpP* gene includes both CDS (exons) and intron sequence

Table 3 List of simple sequence repeats (SSRs) identified in the *P. cirrhiflora* plastome

SSR sequence	Number of repetitions														Total	
	4	5	6	7	8	9	10	11	12	13	14	15	16	17		18
A/T	–	–	–	–	70	38	28	9	2	3	1	–	–	–	1	152
C/G	–	–	–	–	3	1	–	–	–	–	–	–	–	–	–	4
AC/GT	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
AG/CT	10	1	–	–	–	–	–	–	–	–	–	–	–	–	–	11
AT/AT	24	8	11	2	1	3	–	–	–	–	–	–	–	–	–	49
AAT/ATT	6	–	1	–	–	–	–	–	–	–	–	–	–	–	–	7
ACG/CGT	1	–	–	1	1	–	–	–	–	–	–	–	–	–	–	3
AGG/CCT	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
AAAAAG/CTTTTT	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
ACGGCG/CCGTCG	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	1
Total																230

Fig. 4 Sliding window analysis of aligned plastome sequences of *P. cirrhiflora* and *P. contracta*. The regions with higher nucleotide variability ($P_i > 0.1$, above the red line) are indicated by numbers. P_i (y-axis) represents the nucleotide diversity of each window, while the x-axis represents the nucleotide position (in kb) of the midpoint. The intervals corresponding approximately to LSC, IRb and SSC regions, which are delimited in the upper x-axis. Window length, 300pb. Step size, 75pb



of this gene (mainly the first intron). All the six regions identified as hotspots here contain at least one SSR listed in Supplementary Table S2. Three hotspots of *P. cirrhiflora* are SSR-rich regions: the *trnK-UUU/trnQ-UUG* contains 11 SSRs, and the *ycf1* and *clpP* genes (exon + introns) contain 10 SSRs each.

4 Discussion

Differences and similarities in the gene content among *P. cirrhiflora* plastome and other *Deidamioides* species – Shrestha et al. (2019) sequenced and characterized the three first plastomes of *Passiflora* species belonging to the subgenus *Deidamioides* (*P. arbelaezii*, *P. contracta*, and *P. obovata*). These authors also used an unpublished and incomplete draft of the plastome of *P. cirrhiflora* to extract the coding sequences of plastid genes for phylogenetic analyses of the genus *Passiflora* (Shrestha et al. 2019). However, these authors did not include this species in most analyses such as characterization of the plastome length and region sizes, which detail IR borders. We, therefore, presented the complete plastome sequence of *P. cirrhiflora* (Fig. 1), which allows us to characterize this plastome in detail as well as to compare evolutionarily with other species of the subgenus *Deidamioides*.

Most genes commonly found in the plastome of angiosperms (Ruhlman and Jansen 2014) are present in the *P. cirrhiflora* plastome. However, two genes (*rps16* and *rpl22*) are missing and the other two genes (*infA* and *rps7*) have degenerated into pseudogenes in this plastome. The loss or pseudogenization of the *infA*, *rps16* and *rpl22* genes is a common feature shared by all species of the genus *Passiflora* sequenced to date, including all species of *Deidamioides*

(Cauz-Santos et al. 2017; Rabah et al. 2019; Shrestha et al. 2019; Pacheco et al. 2020a). These genes were also lost, and consequently, transferred to the nucleus in other families of the order Malpighiales and/or other angiosperm lineages (Millen et al. 2001; Ueda et al. 2008; Jansen et al. 2011). The functional transfer of these genes to the nucleus is a rational step given that they are essential for cell survival (Cummings and Hershey 1994; Millen et al. 2001; Fleischmann et al. 2011), and loss of function of plastid essential genes leads to lethality (Rogalski et al. 2006, 2008; Alkatib et al. 2012). Accordingly, functional nuclear transfer of the *rps16* and *rpl22* genes was detected in nuclear transcriptome analyses in different species of *Passiflora*, including representatives of subgenera *Deidamioides* (*P. contracta*), *Astrophea*, *Decaloba* and *Passiflora* (Shrestha et al. 2020).

In contrast, the *rps7* gene shows a different status among the species of the subgenus *Deidamioides* (Fig. 5). In *P. cirrhiflora* and *P. contracta*, both gene sequences contain a premature stop codon, which produces a truncated protein isoform. The *rps7* gene is completely intact in *P. arbelaezii* but it is completely absent in *P. obovata*. The absence of this gene is a feature shared by species of the subgenera *Decaloba* (Shrestha et al. 2019). This agrees with the phylogeny of the genus *Passiflora*, in which *P. obovata* is more related to the species of the subgenus *Decaloba* than to other species of the subgenus *Deidamioides* (Shrestha et al. 2019; Pacheco et al. 2020a). Similarly, *P. arbelaezii* is more closely related to *P. pittieri* (subgenus *Astrophea*), which also has a putatively functional copy of the *rps7* gene. Additionally, *P. cirrhiflora* and *P. contracta* form a sister group in phylogenetic analyses (Shrestha et al. 2019; Pacheco et al. 2020a) and showed the same status of pseudogenization of this gene (Fig. 5). This gene encodes the ribosomal protein S7, which is part of the small subunit (30S) of the plastid ribosome (Rogalski et al. 2008; Bieri et al. 2017) and it is

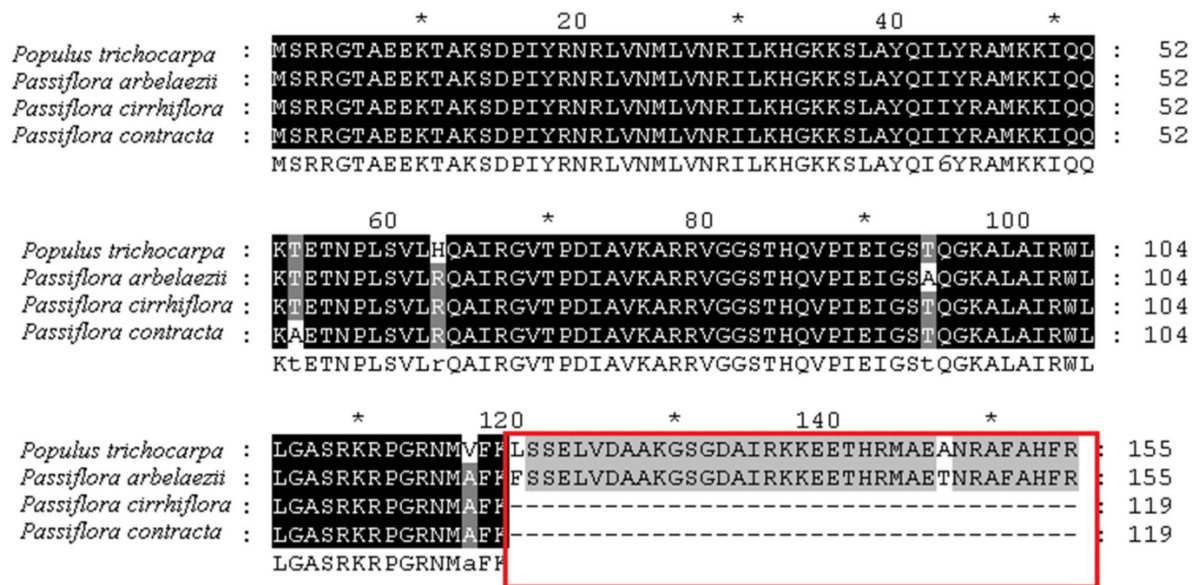


Fig. 5 ClustalW alignment of the translated amino acid sequence of the *rps7* gene, between three species of *Deidamioides* and *Populus trichocarpa* as reference. *Passiflora obovata* is omitted because this gene was completely lost from its plastome. Black shadowing indicates conserved sequence among the four species, dark gray indicates conserved sequence among all except one species, lighter gray indicates conserved sequence between two species and white indicates non-conserved amino acids. Red square highlights the differences in the C-terminus between the intact and truncated forms of the predicted protein

involved in translation initiation as a general and specific factor (Fargo et al. 2001). The loss or pseudogenization of the *rps7* gene is a very rare event in plastomes of photosynthetic angiosperms, which was only reported in two other families, Berberidaceae (Sun et al. 2018) and in Salicaceae (Zhang et al. 2018). Interestingly, a putative functional transfer of the plastid *rps7* was detected in *Passiflora* species that contain both functional (*P. pittieri*) and nonfunctional (*P. contracta*) copies of this gene in the plastome (Shrestha et al. 2020). Due to the close relationship among *P. contracta* and *P. cirrhiflora*, and similar status of the *rps7* between both species (Fig. 5), it is likely that *P. cirrhiflora* also contains the nuclear copy of the *rps7* replacing the plastid copy. Shrestha et al. (2020) suggest that this gene was transferred early during evolution of the genus *Passiflora* (or even in some early stage of the evolution of Malpighiales) and the different status of the *rps7* gene in *Passiflora* is related to differential sequence degradation.

Furthermore, other differences related to plastid gene content among the species of the subgenus *Deidamioides* are associated with the status of the *rpl20* and *rpl32* genes (two essential ribosomal genes) which are intact in *P. cirrhiflora* and *P. arbelaezii* and absent in *P. contracta* and *P. obovata* (Shrestha et al. 2019). Putative functional transfer to the nucleus of the *rpl32* and a substitution of the *rpl20* gene by a nuclear-encoded mitochondrial copy was observed by Shrestha et al. (2020), which includes *P.*

contracta and in *P. arbelaezii*. These authors suggest that these two genes were substituted early by nuclear-encoded copies during *Passiflora* evolution and the species containing functional copies of both genes in their plastomes will likely lose them. Thus, it remains to be elucidated if the plastid *rpl20* and *rpl32* genes are still functional in *P. cirrhiflora* and *P. arbelaezii* or their complete CDS are only evolutionary relics.

Events of IR expansion and contraction among species of the subgenus *Deidamioides* – In comparison with the common structure of angiosperm plastomes, *P. cirrhiflora* shows two inversions in the LSC region (1 and 2, Fig. 2a). These inversions are shared by the other species of the subgenus *Deidamioides* and other subgenera of the genus *Passiflora* (Rabah et al. 2019; Shrestha et al. 2019; Pacheco et al. 2020a). These two rearrangements were also found in *Adenia*, another genus of the family Passifloraceae (Shrestha et al. 2019). Hence, these inversions probably emerged in the common ancestor of the genera *Passiflora* and *Adenia*. *P. arbelaezii* was the only species of the subgenus *Deidamioides* that showed an additional inversion (3, Fig. 2b). This inversion is very similar to that found in the plastome of *P. pittieri* (subgenus *Astropheia*; Rabah et al. 2019; Shrestha et al. 2019).

Expansion and contraction events involving few hundreds of nucleotides occur frequently at the IR boundaries,

whereas large-scale losses or gains at the IR borders are rare events. In most angiosperms, the edge of IR boundaries includes the *ycf1* gene at the IR/SSC junction and the *rps19* gene at the IR/LSC junction (Goulding et al. 1996; Zhu et al. 2016). *Populus* shows this typical IR/SSC border in which the *ycf1* gene is partially duplicated (Fig. 3). On the other hand, the IR boundaries vary significantly between the four *Deidamioides* species (Fig. 3). When compared with the plastome of *Populus*, the *P. cirrhiflora* plastome shows an event of IR expansion, which resulted in the inclusion and duplication of the full sequence of *ycf1* gene. It seems that this expansion event also occurred in the common ancestor of the genera *Passiflora* and *Adenia* given that it is present in species of both genera (Shrestha et al. 2019). In addition to this event, an event of IR expansion in *P. arbelazii* included the *rps15* gene and a small part of the *ndhH* gene, which resulted in the larger IR of the subgenus *Deidamioides* concerning size and gene content (Fig. 3 and Table 1). A very similar IR expansion was detected in *P. pittieri* (subgenus *Astropheae*; Shrestha et al. 2019). Moreover, events of IR expansion including the *ycf1*, *rps15*, and *ndhH* genes were reported in some angiosperm lineages such as *Linum* (family Linaceae) and *Pelargonium* (family Geraniaceae) (Chumley et al. 2006; Weng et al. 2014; Lopes et al. 2018b). Furthermore, *P. contracta* also shows a specific event of IR expansion at the LSC/IRa boundary involving the duplication of the *psbA* gene but also exhibits an event of IR contraction at the LSC/IRb border excluding the *rpl23-rpl2-rps19* genes. These events of IR expansion and contraction in *P. contracta* resulted in IRs with similar size to that found in *P. cirrhiflora*, but with differences in gene content. Finally, the more variable IR of the subgenus *Deidamioides* belongs to *P. obovata*, which experienced a large event of contraction at the IRb/SSC boundary, excluding all the rRNAs genes (Shrestha et al. 2019). This large contraction resulted in smaller IRs and a larger SSC region in the plastome of this species, if compared with all the other species of the subgenus *Deidamioides* (Table 1). Analyzing the borders of the IRs, we also suggest here that the *ycf1* and *rps15* genes were previously included in the IRs of *P. obovata*, by events of IR expansion (like the events E1 and E3 of *P. arbelazii*, Fig. 3), and posteriorly a large event of IR contraction excluded these two genes and several others from the IRa. This hypothesis of duplication and posterior exclusion is based on the current position of these two genes in the plastome of *P. obovata* (i.e., upstream of the *ndhF* gene and close to the typical IRb/SSC boundary), whereas in plastomes of most angiosperms the *rps15* is localized near to the SSC/IRa boundary (i.e., as observed in *Populus*; Fig. 3). Interestingly, a similar IR contraction event occurred independently in *Passiflora menispermifolia*, a species belonging to subgenus *Passiflora* (Shrestha et al. 2019).

RNA editing sites are highly conserved within *Deidamioides* but three sites seem to be unique to *P. cirrhiflora*

– The 60 RNA editing sites predicted here to *P. cirrhiflora* are all alterations from C to T and occur in the first or the second codon position as observed in most angiosperms (Takenaka et al. 2013). Most RNA editing sites predicted here change amino acids from polar to nonpolar, which increases the protein hydrophobicity affecting its structural feature and transmembrane domains (He et al. 2016). This tendency is also reported in several other studies (He et al. 2016; Lopes et al. 2018a, 2019; Pacheco et al. 2020b).

Among RNA editing sites predicted to *P. cirrhiflora*, 49 of them are shared by all the other species of *Deidamioides* analyzed here (*P. contracta*, *P. arbelazii* and *P. obovata*). The results show high conservation of RNA editing sites across the subgenus *Deidamioides* (Supplementary Table S1), despite its polyphyletic origin (Shrestha et al. 2019; Pacheco et al. 2020a). The number of RNA editing sites shared within *Deidamioides* is still higher among the closely related species such as *P. cirrhiflora* and *P. contracta*, totalizing 56 sites. However, three sites predicted here are unique to the *P. cirrhiflora* plastome. These sites change the following amino acids as follows: proline to serine in the *atpA* (424); leucine to phenylalanine in *matK* (209); and alanine to valine in the *rpoC2* (1264). The first change alters the amino acid from nonpolar to polar and the other two changes do not alter the amino acid polarity. In our previous prediction of RNA editing sites in the subgenus *Passiflora* these three editing sites are also absent (Pacheco et al. 2020a). Thus, these RNA editing sites predicted only to the *P. cirrhiflora* plastome probably restore conserved amino acids of the proteins. Accordingly, these conserved amino acids (serine in the *atpA* 424; phenylalanine in the *matK* 209; and valine in the *rpoC2* 1264) are already present in the sequence of these genes in other *Passiflora* species (including all the other *Deidamioides* species) and *Populus trichocarpa*, dispensing the need for editing. Finally, it is necessary to confirm experimentally if these sites are really edited and configure three specific RNA editing sites of *P. cirrhiflora*.

Detection of SSRs and hotspots of nucleotide polymorphism in *P. cirrhiflora* plastome

– Plastid simple sequence repeats (ptSSRs) consist of short DNA sequences repeated in tandem, with a high level of polymorphism, and occur across plastomes of all land plants (Provan et al. 2001; Wheeler et al. 2014; George et al. 2015). These plastid genetic markers can be assessed in both intraspecific and interspecific genetic studies (Powell et al. 1995; Provan et al. 2001; Rogalski et al. 2015). Here, we detected 230 SSRs in the plastome of *P. cirrhiflora*, most of them being mono- and dinucleotides composed by A and T, occurring mainly in

intergenic spacers (Table 3 and Supplementary Table S2). This pattern was also observed in *Passiflora edulis* (Cauz-Santos et al. 2017) and other studies analyzing other angiosperm lineages (Wheeler et al. 2014; George et al. 2015; Lopes et al. 2018c; Pacheco et al. 2020b).

Additionally, we detected six hotspots of nucleotide divergence in the *P. cirrhiflora* plastome. Curiously, most of these regions include gene sequences (*accD*, *clpP*, *rps18* and *ycf1*), and only two of them are localized in IGSs. The *accD* and *clpP* genes are also defined as hotspots of nucleotide polymorphism in the subgenus *Passiflora* (Pacheco et al. 2020a), reinforcing that these genes show an exceptionally high rate of nucleotide divergence within genus *Passiflora* (Shrestha et al. 2019). The *ycf1* gene contains the highest number of peaks of nucleotide divergence in our sliding windows analysis (Fig. 4). This gene was lost in all species of the subgenera *Passiflora* (except for *P. foetida*) and *Decaloba* (Shrestha et al. 2019), but it is intact and putatively functional in all species of the subgenus *Deidamioides*. Corroborating with the high divergence of the *ycf1* in *P. cirrhiflora*, we detected 10 SSRs in its CDS. Other hotspots of nucleotide divergence appointed here are also rich in SSRs, such as *clpP* and the intergenic spacer between the genes *trnK*-UUU and *trnQ*-UUG. This feature (i.e., high nucleotide divergence and presence of SSRs) is expected due to the high level of polymorphism commonly found in SSR sequences. Thus, SSRs and hotspots of nucleotide polymorphism identified here represent attractive genetic information, which is useful for diverse types of genetic studies in *P. cirrhiflora*.

5 Conclusion

The complete plastome sequence of *P. cirrhiflora* reported here and our comparative analyses revealed a considerable structural variation at the IR borders, even among phylogenetically closely related species (*P. cirrhiflora* and *P. contracta*), which consequently resulted in significant differences in gene content and sizes of plastome regions (LSC, SSC, and IRs). Differences between *P. cirrhiflora* and other species of *Deidamioides*, concerning loss, presence and functionality of genes, were also discussed in detail. Moreover, we identified only three RNA editing sites specific to *P. cirrhiflora*, revealing that RNA editing sites are conserved within *Deidamioides*, which contrasts significantly with subgenus *Passiflora*. Furthermore, we detected 230 SSRs and 6 hotspots of nucleotide polymorphism, which represent a rich source of plastid markers to be used in genetic studies. Finally, the availability of the complete plastome of *P. cirrhiflora* is useful for studies in several areas related to phylogeny, genetic, and evolution.

Acknowledgements This research was supported by the National Council for Scientific and Technological Development (CNPq), Brazil (CNPq - Grants 459698/2014-1, 310654/2018-1 and 436407/2018-3). We are grateful to INCT-FBN and for the scholarships granted by the CNPq to ASL, TGP, JDO, WCO, EB, FOP, EMS, and MR. We are also grateful to the Núcleo de Análise de Biomoléculas (NuBiomol) of the Universidade Federal de Viçosa for providing the software CLC Genomics.

Author contribution TGP, ASL, and MR conceived and designed the research. TGP, ASL, JDO, EB, and MR conducted experiments and analyzed the data. WCO, EMS, FOP, and MR contributed with reagents and materials. TGP and MR wrote the manuscript. All authors read and approved the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S, Schöttler MA, Ruf S, Bock R (2012) The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet* 8:e1003076. <https://doi.org/10.1371/journal.pgen.1003076>
- Besnard G, Hernández P, Khadari B, Dorado G, Savolainen V (2011) Genomic profiling of plastid DNA variation in the Mediterranean olive tree. *BMC Plant Biol* 11:1. <https://doi.org/10.1186/1471-2229-11-80>
- Bieri P, Leibundgut M, Saurer M, Boehringer D, Ban N (2017) The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. *EMBO J* 36:475–486. <https://doi.org/10.15252/embj.201695959>
- Blazier JC, Ruhlman TA, Weng ML, Rehman SK, Sabir JSM, Jansen R (2016) Divergence of RNA polymerase α subunits in angiosperm plastid genomes is mediated by genomic rearrangement. *Sci Rep* 6:24595. <https://doi.org/10.1038/srep24595>
- Boggan J, Funck V, Kelloff C (1997) Checklist of the plants of the Guianas (Guyana, Surinam, French Guiana). University of Guyana, Georgetown
- Cauz-Santos LA, Munhoz CF, Rodde N, Cauet S, Santos AA, Penha HA, Dornelas MC, Varani AM, Oliveira GCX, Bergès H, 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* 8:1–17. <https://doi.org/10.3389/fpls.2017.00334>
- Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium × hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol Biol Evol* 23:2175–2190. <https://doi.org/10.1093/molbev/msl089>
- Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2016) Genbank. *Nucl Acids Res* 44:67–72. <https://doi.org/10.1093/nar/gkv1276>
- Cummings HS, Hershey JW (1994) Translation initiation factor IF1 is essential for cell viability in *Escherichia coli*. *J Bacteriol* 176:198–205. <https://doi.org/10.1128/jb.176.1.198-205.1994>

- Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* 17:134. <https://doi.org/10.1186/s13059-016-1004-2>
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>
- Fajardo D, Senalik D, Ames M, Zhu H, Steffan SA, Harbut R, Polashock J, Vorsa N, Gillespie E, Kron K, Zalapa JE (2013) Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. *Tree Genet Genomes* 9:489–498. <https://doi.org/10.1007/s11295-012-0573-9>
- Fargo DC, Boynton JE, Gillham NW (2001) Chloroplast ribosomal protein S7 of *Chlamydomonas* binds to chloroplast mRNA leader sequences and may be involved in translation initiation. *Plant Cell* 13:207–218. <https://doi.org/10.1105/tpc.13.1.207>
- Fleischmann TT, Scharff LB, Alkatib S, Hasdorf S, Schottler MA, Bock R (2011) Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell* 23:3137–3155. <https://doi.org/10.1105/tpc.111.088906>
- George B, Bhatt BS, Awasthi M, George B, Singh AK (2015) Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr Genet* 61:665–677. <https://doi.org/10.1007/s00294-015-0495-9>
- Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet* 252:195–206. <https://doi.org/10.1007/BF02173220>
- Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucl Acids Res* 47:59–64. <https://doi.org/10.1093/nar/gkz238>
- Haberle RC, Fourcade HM, Boore JL, Jansen RK (2008) Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol* 66:350–361. <https://doi.org/10.1007/s00239-008-9086-4>
- Hanaoka M, Kanamaru K, Fujiwara M, Takahashi H, Tanaka K (2005) Glutamyl-tRNA mediates a switch in RNA polymerase use during chloroplast biogenesis. *EMBO Rep* 6:545–550. <https://doi.org/10.1038/sj.embo.7400411>
- He P, Huang S, Xiao G, Zhang Y, Yu J (2016) Abundant RNA editing sites of chloroplast protein-coding genes in *Ginkgo biloba* and an evolutionary pattern analysis. *BMC Plant Biol* 16:257. <https://doi.org/10.1186/s12870-016-0944-8>
- Hokche O, Berry PE, Huber O (2008) Nuevo catálogo de la flora vascular de Venezuela. Fundación Instituto Botánico de Venezuela, Caracas
- Jansen RK, Saski C, Lee SB, Hansen AK, 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. <https://doi.org/10.1093/molbev/msq261>
- Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M (2013) Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* 339:571–574. <https://doi.org/10.1126/science.1229262>
- Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, Nakai Y, Takafuji K, Bédard J, Hirabayashi-Ishioka Y, Mori H, Shiina T, Nakai M (2018) A Ycf2-FtsHi Heteromeric AAA-ATPase complex is required for chloroplast protein import. *Plant Cell* 30:2677–2703. <https://doi.org/10.1105/tpc.18.00357>
- Killip EP (1938) The American species of Passifloraceae. *Publ Field Mus Nat Hist Bot Ser* 19:1–613
- Kode V, Mudd EA, Iamtham S, Day A (2005) The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant J* 44:237–244. <https://doi.org/10.1111/j.1365-3113.2005.02533.x>
- Krosnick SE, Ford AJ, Freudenstein JV (2009) Taxonomic revision of *Passiflora* subgenus *Tetrapathea* including the monotypic genera *Hollrungia* and *Tetrapathea* (Passifloraceae), and a new species of *Passiflora*. *Syst Bot* 34:375–385. <https://doi.org/10.1600/036364409788606343>
- Krosnick SE, Porter-Utley KE, MacDougal JM, Jørgensen PM, McDade LA (2013) New insights into the evolution of *Passiflora* subgenus *Decaloba* (Passifloraceae): phylogenetic relationships and morphological synapomorphies. *Syst Bot* 38:692–713. <https://doi.org/10.1600/036364413X670359>
- Kuroda H, Maliga P (2003) The plastid *clpP1* protease gene is essential for plant development. *Nature* 425:86–89. <https://doi.org/10.1038/nature01909>
- Lopes AS, Pacheco TG, Nimz T, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Rogalski M (2018a) The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae. *Planta* 247:1011–1030. <https://doi.org/10.1007/s00425-018-2841-x>
- Lopes AS, Pacheco TG, Santos KG, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Rogalski M (2018b) The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Rep* 37:307–328. <https://doi.org/10.1007/s00299-017-2231-z>
- Lopes AS, Pacheco TG, Vieira LN, Guerra MP, Nodari RO, de Souza EM, Pedrosa FO, Rogalski M (2018c) The *Crambe abyssinica* plastome: Brassicaceae phylogenomic analysis, evolution of RNA editing sites, hotspot and microsatellite characterization of the tribe Brassiceae. *Gene* 671:36–49. <https://doi.org/10.1016/j.gene.2018.05.088>
- Lopes AS, Pacheco TG, Silva ON, Cruz LM, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2019) The plastomes of *Astrocaryum aculeatum* G. Mey. and *A. murumuru* Mart. show a flip-flop recombination between two short inverted repeats. *Planta* 250:1229–1246. <https://doi.org/10.1007/s00425-019-03217>
- Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucl Acids Res* 44:54–57. <https://doi.org/10.1093/nar/gkw413>
- Ma J, Yang B, Zhu W, Sun L, Tian J, Wang X (2013) The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. *Gene* 528:120–131. <https://doi.org/10.1016/j.gene.2013.07.037>
- MacDougal JM, Feuillet C (2004) Systematics. In: Ulmer T, MacDougal JM (eds) *Passiflora*, passionflowers of the world. Timber, Portland, pp 27–31
- Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13:645–658. <https://doi.org/10.1105/tpc.13.3.645>
- Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucl Acids Res* 37:253–259. <https://doi.org/10.1093/nar/gkp337>
- Pacheco TG, Lopes AS, Viana GDM, Silva ON, Silva GM, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Otoni WC, Rogalski M (2019) Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (*Bixa orellana* L.), the Amazonian commercial species of natural dyes. *Planta* 249:563–582. <https://doi.org/10.1007/s00425-018-3023-6>
- Pacheco TG, Lopes AS, Welter JF, Yotoko KSC, Otoni WC, Vieira LN, Guerra MP, Nodari RO, Balsanelli E, Pedrosa FO, Souza

- EM, Rogalski M (2020a) Plastome sequences of the subgenus *Passiflora* reveal highly divergent genes and specific evolutionary features. *Plant Mol Biol*. <https://doi.org/10.1007/s11103-020-01020-z>
- Pacheco TG, Silva GM, Lopes AS, Oliveira JD, Rogalski JM, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2020b) Phylogenetic and evolutionary features of the plastome of *Tropaeolum pentaphyllum* Lam. (Tropaeolaceae). *Planta*. <https://doi.org/10.1007/s00425-020-03427-w>
- Powell W, Morgantet M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV, Rafalski JA (1995) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr Biol* 5:1023–1029. [https://doi.org/10.1016/S0960-9822\(95\)00206-5](https://doi.org/10.1016/S0960-9822(95)00206-5)
- Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol* 16:142–147. [https://doi.org/10.1016/S0169-5347\(00\)02097-8](https://doi.org/10.1016/S0169-5347(00)02097-8)
- Qiao J, Cai M, Yan G, Wang N, Li F, Chen B, Gao G, Xu K, Li J, Wu X (2016) High-throughput multiplex cpDNA resequencing clarifies the genetic diversity and genetic relationships among *Brassica napus*, *Brassica rapa* and *Brassica oleracea*. *Plant Biotechnol J* 14:409–418. <https://doi.org/10.1111/pbi.12395>
- Rabah SO, Shrestha B, Hajrah NH, Sabir MJ, Alharby HF, Sabir MJ, Alhebshi AM, Sabir JSM, Gilbert LE, Ruhlman TA, Jansen RK (2019) *Passiflora* plastome sequencing reveals widespread genomic rearrangements. *J Syst Evol* 57:1–14. <https://doi.org/10.1111/jse.12425>
- Rocha DI, Batista DS, Faleiro FG, Rogalski M, Ribeiro LM, Mercadante-Simões MO, Yockteng R, Silva ML, Soares WS, Pinheiro MVM, Pacheco TG, Lopes AS, Viccini LF, Otoni WC (2020) *Passiflora* spp. Passionfruit. In: Litz RE, Alfaro FP, Hormaza JI (Org.) *Biotechnology of Fruit and Nut Crops*, 2 ed. CABI, Oxfordshire, pp 381–408
- Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucl Acids Res* 34:4537–4545. <https://doi.org/10.1093/nar/gkl634>
- Rogalski M, Schottler MA, Thiele W, Schulze WX, Bock R (2008) Rpl33, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. *Plant Cell* 20:2221–2237. <https://doi.org/10.1105/tpc.108.060392>
- Rogalski M, Vieira LN, Fraga HP, Guerra MP (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front Plant Sci* 6:586. <https://doi.org/10.3389/fpls.2015.00586>
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Ruhlman TA, Jansen RK (2014) The plastid genomes of flowering plants. *Methods Mol Biol* 1132:3–38. https://doi.org/10.1007/978-1-62703-995-6_1
- Schön A, Krupp G, Gough S, Berry-Lowe S, Kannangara CG, Söll D (1986) The RNA required in the first step of chlorophyll biosynthesis is a chloroplast glutamate tRNA. *Nature* 322:281–284. <https://doi.org/10.1038/322281a0>
- Schwarz EN, Ruhlman TA, Sabir JSM, Hajrah NH, Alharbi NS, Al-Malki AL, Bailey CD, Jansen RK (2015) Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids. *J Syst Evol* 53:458–468. <https://doi.org/10.1111/jse.12179>
- Shrestha B, Weng ML, Therio EC, Gilbert LE, Ruhlman TA, Krosnick SE, Jansen RK (2019) Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. *Mol Phylogenet Evol* 138:53–64. <https://doi.org/10.1016/j.ympev.2019.05.030>
- Shrestha B, Gilbert LE, Ruhlman TA, Jansen RK (2020) Rampant nuclear transfer and substitutions of plastid genes in *Passiflora*. *Genome Biol Evol*. <https://doi.org/10.1093/gbe/evaa123>
- Solórzano S, Chincoya DA, Sanchez-Flores A, Estrada K, Díaz-Velásquez CE, González-Rodríguez A, Vaca-Paniagua F, Dávila P, Arias S (2019) De novo assembly discovered novel structures in g of plastids and revealed divergent inverted repeats in *Mammillaria* (Cactaceae, Caryophyllales). *Plants* 8:392. <https://doi.org/10.3390/plants8100392>
- Stefenon VM, Klabunde G, Lemos RPM, Rogalski M, Nodari RO (2019) Phylogeography of plastid DNA sequences suggests post-glacial southward demographic expansion and the existence of several glacial refugia for *Araucaria angustifolia*. *Sci Rep* 9:2752. <https://doi.org/10.1038/s41598-019-39308-w>
- Sun YX, Moore MJ, Meng AP, Soltis PS, Soltis DE, Li JQ, Wang HC (2013) Complete plastid genome sequencing of trochodendraceae reveals a significant expansion of the inverted repeat and suggests a paleogene divergence between the two extant species. *PLoS ONE* 8:e60429. <https://doi.org/10.1371/journal.pone.0060429>
- Sun Y, Moore MJ, Landis JB, Lin N, Chen L, Deng T, Zhang J, Meng A, Zhang S, Tojibaev KSh, Sun H, Wang H (2018) Plastome phylogenomics of the early-diverging eudicot family Berberidaceae. *Mol Phylogenet Evol* 128:203–211. <https://doi.org/10.1016/j.ympev.2018.07.021>
- Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A (2013) RNA editing in plants and its evolution. *Annu Rev Genet* 47:335–352. <https://doi.org/10.1146/annurev-genet-111212-133519>
- Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of genederived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422. <https://doi.org/10.1007/s00122-002-1031-0>
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq—versatile and accurate annotation of organelle genomes. *Nucl Acids Res* 45:W6–W11. <https://doi.org/10.1093/nar/gkx391>
- Tuskan GA, Difazio S, Jansson S et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604. <https://doi.org/10.1126/science.1128691>
- Ueda M, Nishikawa T, Fujimoto M, Takashi H, Arimura S, Tsutsumi N, Kadowaki K (2008) Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. *Mol Biol Evol* 25:1566–1575. <https://doi.org/10.1093/molbev/msn102>
- Vieira LN, Faoro H, Fraga HP, Rogalski M, Souza EM, Pedrosa FB, Nodari RO, Guerra MP (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. *PLoS ONE* 9:e84792. <https://doi.org/10.1371/journal.pone.0084792>
- Vieira LN, Dos Anjos KG, Faoro H, Fraga HP, Greco TM, Pedrosa FO, de Souza EM, Rogalski M, de Souza RF, Guerra MP (2016a) Phylogenetic inference and SSR characterization of tropical woody bamboos tribe Bambuseae (Poaceae: Bambusoideae) based on complete plastid genome sequences. *Curr Genet* 62:443–453. <https://doi.org/10.1007/s00294-015-0549-z>
- Vieira LN, Rogalski M, Faoro H, Fraga HP, Anjos KG, Picchi GFA, Nodari RO, Pedrosa FO, Souza EM, Guerra MP (2016b) The plastome sequence of the endemic Amazonian conifer, *Retrophyllum piresii* (Silba) C.N.Page, reveals different recombination events and plastome isoforms. *Tree Genet Genomes* 12:10. <https://doi.org/10.1007/s11295-016-0968-0>
- Weng ML, Blazier JC, Govindu M, Jansen RK (2014) Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol Biol Evol* 31:645–659. <https://doi.org/10.1093/molbev/mst257>
- Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE (2014) A review of the prevalence, utility, and caveats of using

- chloroplast simple sequence repeats for studies of plant biology. *Appl Plant Sci* 2:1400059. <https://doi.org/10.3732/apps.1400059>
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255. <https://doi.org/10.1093/bioinformatics/bth352>
- Zhang L, Xi Z, Wang M, Guo X, Ma T (2018) Plastome phylogeny and lineage diversification of Salicaceae with focus on poplars and willows. *Ecol Evol* 8:7817–7823. <https://doi.org/10.1002/ece3.4261>
- Zhu A, Guo W, Gupta S, Fan W, Mower JP (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol* 209:1747–1756. <https://doi.org/10.1111/nph.13743>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SUPPLEMENTARY MATERIAL

Supplementary Table S1 List of RNA editing sites predicted by PREP program in plastid protein-coding genes of *P. cirrhiflora* and other species belonging to the subgenus *Deidamioides*. Red text indicates the predicted RNA editing sites unique to *P. cirrhiflora*, in this analysis. AA pos, amino acid position of the RNA editing site.

Gene	AA Pos	<i>P. cirrhiflora</i>	<i>P. contracta</i>	<i>P. obovata</i>	<i>P. arbelaezii</i>
atpA	354	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ATC (I)	ATC (I)
	424	CCT (P) => TCT (S)	TCT (S)	TCT (S)	TCT (S)
			ACG (T) => ATG (M)		
atpB	214	ATT (I)		ATG (M)	ATG (M)
atpF	121	TTA (L)	TCA (S) => TTA (L)	TTC (P)	TTC (P)
atpI	210	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
matK	180	CTC (L) => TTC (F)	CTC (L) => TTC (F)	CTC (L) => TTC (F)	CTC (L) => TTC (F)
	190	CCG (P) => TCG (S)	CCG (P) => TCG (S)	CCG (P) => TCG (S)	CCG (P) => TCG (S)
	209	CTC (L) => TTC (F)	TTC (F)	TTC (F)	TTC (F)
		CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)
	394				
	397	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
	462	CTT (L) => TTT (F)	CTT (L) => TTT (F)	TTT (F)	TTT (F)
ndhA	23	GCT (A) => GTT (V)	GCT (A) => GTT (V)	GCT (A) => GTT (V)	GCT (A) => GTT (V)
	109	ACT (T) => ATT (I)	ACT (T) => ATT (I)	ACT (T) => ATT (I)	ACT (T) => ATT (I)
	114	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
	140	ACA (T) => ATA (I)	ACA (T) => ATA (I)	ACA (T) => ATA (I)	ACA (T) => ATA (I)
	317	CTC (L) => TTC (F)	CTC (L) => TTC (F)	CTC (L) => TTC (F)	CTC (L) => TTC (F)
ndhB	9	CTC (L) => TTC (F)	CTC (L) => TTC (F)	CTC (L) => TTC (F)	CTC (L) => TTC (F)
	32	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
	138	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)
	178	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)
	186	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)
	228	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)
	231	TCT (S) => TTT (F)	TCT (S) => TTT (F)	TCT (S) => TTT (F)	TCT (S) => TTT (F)
	259	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
	261	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
	401	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)
	476	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)
		ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)
ndhD	1				
	7	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
	9	ACA (T) => ATA (I)	ACA (T) => ATA (I)	ACA (T) => ATA (I)	GCA (A)
	16	TCC (S) => TTC (F)	TCC (S) => TTC (F)	TCC (S) => TTC (F)	TCC (S) => TTC (F)
		CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)
	105				
	182	GCT (A) => GTT (V)	GCT (A) => GTT (V)	GCT (A) => GTT (V)	GCT (A) => GTT (V)
	200	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)

	225	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
	369	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ATC (I)
	469	CTT (L) => TTT (F)	CTT (L) => TTT (F)	CTT (L) => TTT (F)	CTT (L) => TTT (F) GCA (A) => GTA (V)
	472	GTA (V)	GTA (V)	GTA (V)	GTA (V)
ndhF	196	CTT (L) => TTT (F)	CTT (L) => TTT (F)	CTT (L) => TTT (F)	CTT (L) => TTT (F) ACG (T) => ATG (M)
	332	ATG (M) GCC (A) => GTC (V)	ATG (M) GCC (A) => GTC (V)	ATG (M)	ATG (M) GCC (A) => GTC (V)
	644	GTC (V)	GTC (V)	GTC (V)	GTC (V)
ndhG	56	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)
	84	CCA (P) => TCA (S)	CCA (P) => TCA (S)	CCA (P) => TCA (S)	CCA (P) => TCA (S)
	105	ACA (T) => ATA (I) GCG (A) => GTG (V)	ACA (T) => ATA (I) GCG (A) => GTG (V)	ACA (T) => ATA (I) GCG (A) => GTG (V)	ACA (T) => ATA (I) GCG (A) => GTG (V)
petD	139	GTC (V)	GTC (V)	GTC (V)	GTC (V)
psaI	9	CCT (P) => TCT (S)	CCT (P) => TCT (S)	CCT (P) => TCT (S)	CCT (P) => TCT (S)
	28	TCT (S) => TTT (F)	TCT (S) => TTT (F)	TCT (S) => TTT (F)	TCT (S) => TTT (F)
psbB	193	TTT (F)	TTT (F)	CTT (L) => TTT (F)	TTT (F)
psbE	72	CCT (P) => TCT (S)	CCT (P) => TCT (S)	CCT (P) => TCT (S)	CCT (P) => TCT (S)
psbF	26	TCT (S) => TTT (F) GCC (A) => GTC (V)	TCT (S) => TTT (F) GCC (A) => GTC (V)	TCT (S) => TTT (F)	TCT (S) => TTT (F)
rpl2	199	GTC (V)	GTC (V)	ACC (T)	GTC (V)
rpoA	310	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)
rpoB	113	TCT (S) => TTT (F)	TCT (S) => TTT (F)	TCT (S) => TTT (F)	TCT (S) => TTT (F)
	184	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L) CAC (H) => TAC (Y)	TCA (S) => TTA (L) TAC (Y)
	501	TAC (Y)	TAC (Y)	GCG (A) => GTG (V)	TAC (Y)
	900	GTG (V)	GTG (V)	GTG (V)	GTG (V)
	1024	ATC (I)	ATC (I)	ATC (I)	ACC (T) => ATC (I)
rpoC1	14	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
	352	GTT (V)	GTT (V)	GTT (V) CGG (R) => TGG (W)	GCT (A) => GTT (V) TGG (W)
	438	TGG (W)	TGG (W)	TGG (W)	TGG (W)
	483	CGT (R) => TGT (C)	CGT (R) => TGT (C)	CGT (R) => TGT (C)	CGT (R) => TGT (C)
rpoC2	282	TTT (F)	TTT (F)	CTT (L) => TTT (F)	CTT (L) => TTT (F)
	536	CTT (L) => TTT (F)	CTT (L) => TTT (F)	CTT (L) => TTT (F)	CTT (L) => TTT (F)
	726	CTT (L) => TTT (F)	CTT (L) => TTT (F)	CTT (L) => TTT (F)	CTT (L) => TTT (F)
	775	GTC (V) CGG (R) => TGG (W)	GTC (V) TGG (W)	GTC (V) CGG (R) => TGG (W)	GTC (V) CGG (R) => TGG (W)
	777	GTC (V)	GTC (V)	GTC (V)	GTC (V)
	910	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TTG (L)	TCG (S) => TTG (L)
	1264	GCT (A) => GTT (V)	GTT (V)	GTT (V)	GTT (V)
	1266	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
rps2	83	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)
	203	TCG (S)	TCG (S)	TCG (S)	CCG (P) => TCG (S)
rps14	27	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)
	50	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)

Supplementary Table S2 Distribution of SSR loci in the plastome of *P. cirrhiflora*.

SSR n°	SSR type	SSR Sequence	Size	Start	End	Location
1	mono	(T)8	8	1404	1411	psbA/trnH-GUG (IGS)
2	mono	(T)8	8	1421	1428	psbA/ trnH-GUG (IGS)
3	mono	(A)8	8	1732	1739	trnK-UUU (intron)
4	di	(AT)4	8	2224	2231	matK (CDS)
5	mono	(A)10	10	4399	4408	trnK-UUU/trnQ-UUG (IGS)
6	mono	(A)8	8	4447	4454	trnK-UUU/trnQ-UUG (IGS)
7	mono	(T)8	8	4581	4588	trnK-UUU/trnQ-UUG (IGS)
8	mono	(T)9	9	4592	4600	trnK-UUU/trnQ-UUG (IGS)
9	mono	(A)8	8	4857	4864	trnK-UUU/trnQ-UUG (IGS)
10	mono	(A)9	9	4908	4916	trnK-UUU/trnQ-UUG (IGS)
11	tri	(TAT)4	12	5334	5345	trnK-UUU/trnQ-UUG (IGS)
12	mono	(A)8	8	5346	5353	trnK-UUU/trnQ-UUG (IGS)
13	di	(TA)5	10	5354	5363	trnK-UUU/trnQ-UUG (IGS)
14	di	(AT)5	10	5364	5373	trnK-UUU/trnQ-UUG (IGS)
15	di	(AT)7	14	5375	5388	trnK-UUU/trnQ-UUG (IGS)
16	mono	(T)8	8	5958	5965	psbK(CDS)
17	mono	(T)10	10	6425	6434	psbK/psbI (IGS)
18	di	(TA)4	8	7562	7569	trnS-GCU/trnG-UCC (IGS)
19	mono	(T)8	8	8189	8196	trnG-UCC (intron)
20	di	(TA)4	8	8542	8549	trnG-UCC/trnR-UCU (IGS)
21	tri	(TAT)4	12	8629	8640	trnG-UCC/trnR-UCU (IGS)
22	di	(AT)5	10	8685	8694	trnG-UCC/trnR-UCU (IGS)
23	mono	(T)10	10	8850	8859	trnG-UCC/trnR-UCU (IGS)
24	di	(TA)5	10	9063	9072	trnR-UCU/atpA (IGS)
25	di	(TA)5	10	9081	9090	trnR-UCU/atpA (IGS)
26	di	(TA)4	8	9117	9124	trnR-UCU/atpA (IGS)
27	di	(AT)8	16	9299	9314	trnR-UCU/atpA (IGS)
28	mono	(A)10	10	9332	9341	trnR-UCU/atpA (IGS)
29	mono	(A)9	9	10943	10951	atpA/atpF (IGS)
30	mono	(A)8	8	11395	11402	atpF (CDS)
31	di	(AT)4	8	11632	11639	atpF/atpH (IGS)
32	di	(TA)4	8	11673	11680	atpF/atpH (IGS)
33	di	(TA)6	12	11696	11707	atpF/atpH (IGS)
34	mono	(T)8	20	11708	11715	atpF/atpH (IGS)
35	mono	(T)10	10	11741	11750	atpF/atpH (IGS)
36	mono	(T)8	8	11759	11766	atpF/atpH (IGS)
37	mono	(T)10	10	11887	11896	atpF/atpH (IGS)
38	mono	(T)10	10	12974	12983	atpH/atpI (IGS)
39	mono	(T)8	8	13066	13073	atpH/atpI (IGS)
40	mono	(T)13	13	13314	13326	atpH/atpI (IGS)
41	mono	(T)9	9	15125	15133	rps2 (CDS)
42	mono	(T)10	10	15382	15391	rps2/rpoC2 (IGS)
43	mono	(A)9	9	15397	15405	rps2/rpoC2 (IGS)
44	mono	(T)9	9	15525	15533	rps2/rpoC2 (IGS)

45	mono	(T)18	18	17584	17601	rpoC2 (CDS)
46	mono	(T)14	14	17732	17745	rpoC2 (CDS)
47	mono	(A)8	8	17878	17885	rpoC2 (CDS)
48	mono	(T)10	10	18272	18281	rpoC2 (CDS)
49	di	(AT)5	10	19120	19129	rpoC2 (CDS)
50	mono	(A)8	8	21571	21578	rpoC1 (CDS)
51	mono	(T)8	8	21877	21884	rpoC1 (intron)
52	mono	(A)10	10	22185	22194	rpoC1 (intron)
53	mono	(A)9	9	22860	22868	rpoB (CDS)
54	mono	(T)10	10	25452	25461	rpoB (CDS)
55	mono	(A)10	10	26611	26620	rpoB/trnC-GCA (IGS)
56	di	(TA)4	8	26743	26750	rpoB/trnC-GCA (IGS)
57	di	(AT)9	18	26964	26981	rpoB/trnC-GCA (IGS)
58	di	(AT)9	18	27056	27073	rpoB/trnC-GCA (IGS)
59	mono	(A)9	9	27141	27149	rpoB/trnC-GCA (IGS)
60	di	(TA)4	8	27611	27618	trnC-GCA/petN (IGS)
61	mono	(T)9	9	27650	27658	trnC-GCA/petN (IGS)
62	mono	(T)8	8	27691	27698	trnC-GCA/petN (IGS)
63	mono	(A)8	8	28014	28021	petN/psbM (IGS)
64	mono	(T)8	8	28751	28758	petN/psbM (IGS)
65	di	(AT)4	8	29097	29104	petN/psbM (IGS)
66	mono	(A)8	8	30227	30234	psbM/trnD-GUC (IGS)
67	mono	(A)10	10	30592	30601	trnD-GUC/trnY-GUA (IGS)
68	mono	(T)11	11	30649	30659	trnD-GUC/trnY-GUA (IGS)
69	mono	(A)11	11	31552	31562	trnE-UUC/trnT-GGU (IGS)
70	mono	(T)10	10	33032	33041	trnT-GGU/psbD (IGS)
71	mono	(A)10	10	33165	33174	trnT-GGU/psbD (IGS)
72	di	(GA)4	8	36130	36137	trnS-UGA (CDS)
73	mono	(A)8	8	36987	36994	psbZ/trnG-GCC (IGS)
74	di	(TA)6	12	43349	43360	psaA/ycf3 (IGS)
75	di	(AT)4	8	43369	43376	psaA/ycf3 (IGS)
76	mono	(A)8	8	43409	43416	psaA/ycf3 (IGS)
77	mono	(T)8	8	44257	44264	ycf3 (intron)
78	mono	(A)9	9	45506	45514	ycf3/trnS-GGA (IGS)
79	di	(AT)4	8	45614	45621	ycf3/trnS-GGA (IGS)
80	di	(AT)4	8	45634	45641	ycf3/trnS-GGA (IGS)
81	mono	(T)8	8	45712	45719	ycf3/trnS-GGA (IGS)
82	di	(TA)4	8	47024	47031	rps4/trnT-UGU (IGS)
83	di	(TA)6	12	47035	47046	rps4/trnT-UGU (IGS)
84	di	(TA)4	8	47050	47057	rps4/trnT-UGU (IGS)
85	di	(AT)6	12	47471	47482	trnT-UGU/trnL-UAA (IGS)
86	di	(TA)5	10	47583	47592	trnT-UGU/trnL-UAA (IGS)
87	mono	(T)10	10	48262	48271	trnL-UAA (intron)
88	mono	(A)8	8	48282	48289	trnL-UAA (intron)
89	mono	(A)8	8	48511	48518	trnL-UAA/trnF-GAA (IGS)
90	mono	(T)10	10	48609	48618	trnL-UAA/trnF-GAA (IGS)

91	di	(AT)4	8	48913	48920	trnL-UAA/trnF-GAA (IGS)
92	di	(TA)6	12	48925	48936	trnL-UAA/trnF-GAA (IGS)
93	di	(AT)4	8	49007	49014	trnL-UAA/trnF-GAA (IGS)
94	di	(AT)4	8	49032	49039	trnL-UAA/trnF-GAA (IGS)
95	di	(TC)4	8	50080	50087	ndhJ/ndhK (IGS)
96	mono	(T)11	11	50890	50900	ndhK/ndhC (IGS)
97	mono	(T)8	8	51388	51395	ndhC/atpB (IGS)
98	mono	(T)10	10	51710	51719	ndhC/atpB (IGS)
99	tri	(AAT)4	12	52134	52145	ndhC/atpB (IGS)
100	mono	(T)8	8	52333	52340	ndhC/atpB (IGS)
101	mono	(A)10	10	52540	52549	ndhC/atpB (IGS)
102	mono	(T)9	9	52561	52569	ndhC/atpB (IGS)
103	mono	(A)8	8	55398	55405	trnV-UAC (intron)
104	mono	(A)10	10	55514	55523	trnV-UAC (intron)
105	mono	(A)8	8	55726	55733	trnV-UAC/rbcL (IGS)
106	mono	(A)9	9	55818	55826	trnV-UAC/rbcL (IGS)
107	di	(TA)4	8	55948	55955	trnV-UAC/rbcL (IGS)
108	di	(GA)4	8	56627	56634	rbcL (CDS)
109	di	(AT)7	14	57841	57854	rbcL/accD (IGS)
110	tri	(GGA)4	12	59159	59170	accD (CDS)
111	mono	(T)9	9	59961	59969	accD/psaI (IGS)
112	mono	(A)8	8	60357	60364	accD/psaI (IGS)
113	mono	(A)8	8	61136	61143	psaI/ycf4 (IGS)
114	mono	(A)8	8	61254	61261	ycf4 (CDS)
115	mono	(T)8	8	61376	61383	ycf4 (CDS)
116	mono	(T)8	8	61934	61941	ycf4/cemA (IGS)
117	mono	(A)8	8	62020	62027	cemA (CDS)
118	mono	(A)9	9	62211	62219	cemA (CDS)
119	di	(AT)4	8	63204	63211	petA (CDS)
120	mono	(T)8	8	63538	63545	petA (CDS)
121	mono	(A)8	8	63564	63571	petA (CDS)
122	tri	(AAT)4	12	64905	64916	petA/psbJ (IGS)
123	tri	(AAT)4	12	64929	64940	petA/psbJ (IGS)
124	mono	(T)9	9	65162	65170	petA/psbJ (IGS)
125	mono	(T)8	8	66576	66583	psbJ/psbL (IGS)
126	mono	(A)8	8	67001	67008	psbE/petL (IGS)
127	mono	(A)9	9	67174	67182	petL/petG (IGS)
128	mono	(T)9	9	67293	67301	petL/petG (IGS)
129	mono	(A)9	9	67633	67641	trnW-CCA/trnP-UGG (IGS)
130	mono	(T)9	9	67665	67673	trnW-CCA/trnP-UGG (IGS)
131	mono	(A)9	9	67802	67810	trnW-CCA/trnP-UGG (IGS)
132	mono	(A)9	9	67819	67827	trnW-CCA/trnP-UGG (IGS)
133	mono	(A)8	8	69047	69054	psaJ/rpl33 (IGS)
134	mono	(A)8	8	69814	69821	rps18 (CDS)
135	mono	(T)10	10	70631	70640	rpl20/rps12 (IGS)
136	mono	(T)13	13	70754	70766	rpl20/rps12 (IGS)

137	mono	(T)9	9	71164	71172	rpl20/rps12 (IGS)
138	mono	(T)9	9	71314	71322	rpl20/rps12 (IGS)
139	hexa	(TCGCCG)5	30	71953	71982	clpP (CDS)
140	tri	(TCG)7	21	71989	72009	clpP (CDS)
141	tri	(GTC)4	12	72012	72023	clpP (CDS)
142	tri	(GTC)8	24	72036	72059	clpP (CDS)
143	mono	(A)9	9	72327	72335	clpP (intron)
144	mono	(A)8	8	72471	72478	clpP (intron)
145	mono	(A)9	9	72706	72714	clpP (intron)
146	tri	(ATA)4	12	72747	72758	clpP (intron)
147	mono	(A)8	8	73125	73132	clpP (intron)
148	mono	(T)8	8	73441	73448	clpP (intron)
149	mono	(T)8	8	76152	76159	psbB/psbT (IGS)
150	di	(CT)4	8	77338	77345	petB (intron)
151	mono	(T)8	8	79035	79042	petD (intron)
152	mono	(T)8	8	81760	81767	rps11/rpl36 (IGS)
153	mono	(T)11	11	81997	82007	rpl36/rps8 (IGS)
154	mono	(A)9	9	82034	82042	rpl36/rps8 (IGS)
155	mono	(T)10	10	82406	82415	rpl36/rps8 (IGS)
156	mono	(T)9	9	82651	82659	rps8 (CDS)
157	mono	(T)9	9	82936	82944	rps8/rpl14 (IGS)
158	mono	(A)9	9	83477	83485	rpl14/rpl16 (IGS)
159	mono	(A)8	8	83508	83515	rpl14/rpl16 (IGS)
160	mono	(T)9	9	83527	83535	rpl14/rpl16 (IGS)
161	mono	(T)9	9	84501	84509	rpl16 (intron)
162	mono	(T)10	10	84767	84776	rpl16 (intron)
163	mono	(T)9	9	84805	84813	rpl16 (intron)
164	mono	(T)8	8	85010	85017	rpl16 (intron)
165	di	(AT)4	8	85941	85948	rps3/rps19 (IGS)
166	mono	(T)9	9	86552	86560	rps19 (CDS)
167	mono	(T)13	13	86590	86602	rps19/rpl2 (IGS)
168	mono	(T)9	9	86609	86617	rps19/rpl2 (IGS)
169	di	(AG)4	8	88923	88930	trnI-CAU/ycf2 (IGS)
170	di	(AG)4	8	89145	89152	ycf2 (CDS)
171	mono	(G)8	8	90286	90293	ycf2 (CDS)
172	mono	(A)9	9	91558	91566	ycf2 (CDS)
173	di	(GA)4	8	91579	91586	ycf2 (CDS)
174	mono	(T)8	8	92811	92818	ycf2 (CDS)
175	mono	(A)11	11	94912	94922	ycf2 (CDS)
176	di	(TA)6	12	96089	96100	tnrL-CAA/ndhB (IGS)
177	di	(AG)4	8	96853	96860	ndhB (CDS)
178	mono	(T)8	8	98786	98793	ndhB/rps7 (IGS)
179	mono	(T)8	8	102318	102325	trnV-GAC/rrn16 (IGS)
180	mono	(C)8	8	102332	102339	trnV-GAC/rrn16 (IGS)
181	mono	(A)8	8	104083	104090	rrn16/trnI-GAU (IGS)
182	mono	(T)8	8	104785	104792	trnI-GAU (intron)

183	mono	(G)8	8	106002	106009	trnA-UGC (intron)
184	di	(CT)4	8	108114	108121	rrn23 (CDS)
185	mono	(A)8	8	109509	109516	rrn4.5/rrn5 (IGS)
186	mono	(A)11	11	109819	109829	rrn5/trnR-ACG (IGS)
187	di	(AG)5	10	109916	109925	rrn5/trnR-ACG (IGS)
188	di	(TA)5	10	111229	111238	trnN-GUU/ycf1 (IGS)
189	di	(AT)4	8	111307	111314	trnN-GUU/ycf1 (IGS)
190	mono	(A)12	12	111418	111429	ycf1 (CDS)
191	di	(AG)4	8	112216	112223	ycf1 (CDS)
192	hexa	(AAAGAA)4	24	114398	114421	ycf1 (CDS)
193	mono	(A)11	11	114581	114591	ycf1 (CDS)
194	mono	(A)11	11	115070	115080	ycf1 (CDS)
195	di	(AT)4	8	115149	115156	ycf1 (CDS)
196	mono	(A)8	8	115233	115240	ycf1 (CDS)
197	mono	(A)8	8	115272	115279	ycf1 (CDS)
198	mono	(A)8	8	115311	115318	ycf1 (CDS)
199	di	(AC)4	8	116381	116388	ycf1 (CDS)
200	di	(AT)6	12	117481	117492	ycf1/ndhF (IGS)
201	mono	(A)11	11	117696	117706	ycf1/ndhF (IGS)
202	mono	(A)10	10	117802	117811	ycf1/ndhF (IGS)
203	mono	(T)8	8	117968	117975	ycf1/ndhF (IGS)
204	mono	(A)9	9	117981	117989	ycf1/ndhF (IGS)
205	mono	(A)8	8	118724	118731	ndhF (CDS)
206	mono	(T)10	10	118778	118787	ndhF (CDS)
207	di	(AT)4	8	120032	120039	ndhF (CDS)
208	mono	(T)9	9	120758	120766	ndhF/rpl32 (IGS)
209	mono	(A)8	8	121164	121171	rpl32/trnL-UAG (IGS)
210	mono	(A)8	8	121177	121184	rpl32/trnL-UAG (IGS)
211	mono	(A)8	8	121556	121563	rpl32/trnL-UAG (IGS)
212	di	(AT)6	12	121657	121668	rpl32/trnL-UAG (IGS)
213	mono	(T)10	10	121866	121875	rpl32/trnL-UAG (IGS)
214	mono	(T)8	8	122763	122770	ccsA (CDS)
215	mono	(A)9	9	123128	123136	ccsA/ndhD (IGS)
216	di	(TA)6	12	123240	123251	ccsA/ndhD (IGS)
217	mono	(T)10	10	123350	123359	ccsA/ndhD (IGS)
218	mono	(T)8	8	125330	125337	psaC/ndhE (IGS)
219	di	(TA)6	12	127075	127086	ndhG/ndhI (IGS)
220	di	(TA)9	18	127089	127106	ndhG/ndhI (IGS)
221	di	(AT)6	12	127119	127130	ndhG/ndhI (IGS)
222	mono	(A)8	8	128271	128278	ndhA (CDS)
223	di	(AT)4	8	128698	128705	ndhA (intron)
224	mono	(A)8	8	128730	128737	ndhA (intron)
225	mono	(T)12	12	128835	128846	ndhA (intron)
226	mono	(A)10	10	128866	128875	ndhA (intron)
227	mono	(T)8	8	129276	129283	ndhA (intron)
228	mono	(A)8	8	129438	129445	ndhA (intron)

229	mono	(C)9	9	130367	130375	ndhH (CDS)
230	tri	(AAT)6	18	131661	131678	rps15/ycf1 (IGS)

General conclusions

Plastid genome sequencing is valuable to diverse biological fields of knowledge such as evolution, phylogeny, genetic, cytology, and biotechnology. The present work reports the complete plastome sequence of *Bixa orellana* (Bixaceae), *Tropaeolum pentaphyllum* (Tropaeolaceae), and seven species of *Passiflora* L. (six of the subgenus *Passiflora* and *Passiflora cirrhiflora* of the subgenus *Deidamioides*). The two former represented the first complete plastome sequence of Bixaceae and Tropaeolaceae to be sequenced and characterized in detail. If one considers the structure and gene content, the plastomes of *B. orellana* and *T. pentaphyllum* show a high similarity to other plastomes of the families Malvales and Brassicales, respectively. However, some specific evolving features were also detected in the two species.

In *B. orellana*, the *psaI* gene shows exceptional dN/dS values in comparison with other Malvales, indicating that this gene can be under positive selection in this species, which could change structurally and functionally the encoded protein. Other specific features of the of *B. orellana* plastome is related to the loss of the *atpF* intron and the presence of 11 specific RNA editing sites. Concerning the specific features of *T. pentaphyllum* plastome, an IR expansion event was observed in comparison with other Brassicales, including the whole sequence of the *rps19* gene in the IRs. Also, an extension in the C-terminus of *matK* and *rpoA* genes and a partial deletion of the C-terminus of the *rpoC2* gene. Additionally, two specific RNA editing sites were predicted only to this species.

The complete plastome of *Passiflora cirrhiflora* and comparatives analysis shows a significant variation at the IR borders among the species of the subgenus *Deidamioides*. Concerning the gene content, the differences among the species within *Deidamioides* are mainly related to the status of the *rps7*, *rpl20*, and *rpl32* genes. The *rps7* gene has a premature stop codon in *P. cirrhiflora* and *Passiflora contracta*, while it is completely absent in *Passiflora obovata* and intact in *Passiflora arbelaezii*. On other hand, the *rpl20* and *rpl32* are intact and putatively functional in *P. cirrhiflora* and *P. arbelaezii*, but it is absent in *P. contracta* and *P. obovata*. Moreover, three RNA editing sites were predicted here only to *P. cirrhiflora*, while 49 sites were predicted to all the *Deidamioides* analyzed here.

Sliding window and gene divergence analyses among the six species of the subgenus *Passiflora* reported here in combination with other 13 species of this subgenus available in the organelle database, highlighted the *accD* and the *clpP* gene as the regions with the highest peaks of nucleotide divergence in these plastomes. Furthermore, the comparative analysis of RNA

editing sites shows significant interspecific variation between the species analyzed here. These data allowed us to speculate if the high divergence of *clpP* and *accD* genes and the polymorphism of RNA editing sites could be candidates for the nucleus-plastome incompatibility in interspecific hybrids of this subgenus.

Also, this work identified a high number of SSRs in *B. orellana*, *T. pentaphyllum*, and *P. cirrhiflora*. Hotspots of nucleotide divergence were also appointed to the subgenus *Passiflora*, *P. cirrhiflora*, and *T. pentaphyllum*. These plastid molecular markers are useful for several genetic studies with these taxa such as conservation, breeding, and deep-level phylogeny.

The first complete plastome sequence of the families Bixaceae and a Tropaeolaceae reported here enabled a phylogenomic approach of the order Malvales and Brassicales with a higher number of sampled taxa. This analysis revealed that Bixaceae is more closely related to Malvaceae (Malvales). Similarly, the analysis showed that Tropaeolaceae is closely related to Akaniaceae, forming a sister group to Caricaceae-Moringaceae, within Brassicales. The species of the subgenus *Passiflora* formed a monophyletic clade with high support. The relationships among the species of this subgenus were also well resolved, except for relationships between *P. quadrangularis*, *P. edulis*, and *P. cincinnata*. The position of them was also inconsistent between the tree based on concatenated-genes and phylogenomic trees using whole-plastomes. Therefore, it is possible to suggest here that a larger taxon sample is needed to resolve such incongruities. The data provided here bring a huge amount of data that can be applied in several analyses using species related to the taxa sequenced here. Finally, these data are available and represent a rich source of genetic resources to diverse scientific fields.