TÚLIO GOMES PACHECO

THE COMPLETE PLASTID GENOME SEQUENCE OF *Passiflora cincinnata*: GENOME REARRANGEMENTS, MASSIVE PLASTID GENE LOSSES AND IMPLICATIONS TO GENOME-PLASTOME INCOMPATIBILITY

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fisiologia Vegetal, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2016

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
P116c
2016

Pacheco, Túlio Gomes, 1989-
The complete plastid genome sequence of *Passiflora cincinnata* : genome rearrangements, massive plastid gene losses and implications to genome-plastome incompatibility / Túlio Gomes Pacheco. – Viçosa, MG, 2016.
v, 73f. : il. (algumas color.) ; 29 cm.

Orientador: Marcelo Rogalski.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 51-73.

1. *Passiflora cincinnata. 2.* Genômica. 3. Evolução. 4. Plastomas. 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

THE COMPLETE PLASTID GENOME SEQUENCE OF *Passiflora cincinnata*: GENOME REARRANGEMENTS, MASSIVE PLASTID GENE LOSSES AND IMPLICATIONS TO GENOME-PLASTOME INCOMPATIBILITY

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fisiologia Vegetal, para obtenção do título de *Magister Scientiae*.

APROVADA: 20 de julho de 2016

_____ _____

Wagner Campos Otoni | Humberto Josué de Oliveira Ramos

_____

Marcelo Rogalski
(Orientador)

AGRADECIMENTOS

Aos meus pais, Sirley e Heleno, e minha irmã, Gabriella, pelo apoio e amor incondicional, durante toda minha vida. À Giovanna, minha namorada, pela paciência, companheirismo, amor e apoio em todas as horas.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Fisiologia Vegetal por todo suporte e auxílio acadêmico e também pela oportunidade.

À CAPES pela concessão de bolsa de estudos.

Ao Prof. Marcelo Rogalski, pela orientação, atenção, paciência e também pela amizade.

À Prof[a]. Karla Yotoko pelas valiosas contribuições neste trabalho.

Aos professores Wagner Campos Otoni e Humberto Josué de Oliveira Ramos por aceitarem integrar a banca examinadora deste trabalho.

A todos os professores da Fisiologia Vegetal pelo conhecimento compartilhado. A todos os técnicos do Departamento de Biologia Vegetal pelo auxílio e suporte.

Aos amigos do Grupo Rogalski que ajudaram diretamente neste trabalho. A todos os amigos do Laboratório de Fisiologia Molecular de Plantas, atuais e egressos, pelo convívio e amizade.

Aos amigos de Ponte Nova e de Viçosa de longa data, e também aos amigos da Biologia (UFV), da Disciplina BVE 687 (2005/2) e do WoW, pela amizade, atenção e carinho.

A todos que ajudaram direta ou indiretamente para a realização deste trabalho, muito obrigado.

ÍNDICE

RESUMO..... iv

ABSTRACT..... v

1. Introduction……... 1

2. Material and methods……... 8

2.1. Plant material, chloroplast isolation and plastid DNA extraction……... 8

2.2. Plastid genome sequencing, assembling and annotation…… 8

2.3. Comparative analysis of genome structure…….. 8

2.4. Repeats sequence analysis and RNA editing sites prediction……. 9

2.5. Phylogenetic inference……. 9

2.6. Pairwise distance analysis and estimation of synonymous (dS) and nonsynonymous (dN) substitution rates……. 11

3. Results…… 12

3.1. Genome size and gene content of *P. cincinnata* plastid genome…… 12

3.2. Comparative analysis of genome structure and IR boundary…….. 15

3.3. Repeat sequence analysis and RNA editing prediction……... 18

3.4. Phylogenetic inference……. 23

3.5. Pairwise distance analysis and estimation of synonymous (dS) and nonsynonymous (dN) substitution rates…… 25

4. Discussion…….. 28

4.1. *P. cincinnata* plastome shows specific high divergence of genes and unprecedented gene losses in photosynthetic higher plants……... 28

4.2. Multiple inversions in *P. cincinnata* plastome…… 43

4.3. Genome size reduction *versus* IR expansion…… 44

4.4. Repeats sequence analysis……... 45

4.5. RNA editing sites prediction……... 46

4.6. Phylogenetic position of *P. cincinnata* (Passifloraceae) within Malpighiales… 48

5. Conclusion and outlook..... 49

6. References..... 51

RESUMO

PACHECO, Túlio Gomes, M. Sc., Universidade Federal de Viçosa, julho de 2016. **Sequência completa do genoma plastidial de *Passiflora cincinnata*: rearranjos no genoma, múltiplas perdas de genes plastidiais e implicações na incompatibilidade genoma-plastoma.** Orientador: Marcelo Rogalski.

A organização, ordem e conteúdo gênico de genomas plastidiais (plastomas) são bastante conservados em angiospermas, porém há exceções a esta regra. Este parece ser o caso do gênero *Passiflora*, pois há evidências de perdas não usuais de genes plastidiais para espécies deste gênero. Porém, nenhum plastoma de *Passiflora* foi publicado até o momento, o que dificulta estudos a respeito da evolução do plastoma deste grupo. Da mesma forma, o estudo das causas da incompatibilidade entre o genoma nuclear e plastoma, apresentada por alguns híbridos interespecíficos de *Passiflora*, tem se mantido obscuro devido à falta de sequências plastidiais disponíveis no banco de dados. Assim, visando começar a preencher estas lacunas e ainda permitir a caracterização de marcadores genéticos plastidiais e a construção de vetores para transformação plastidial em *Passiflora*, o presente trabalho teve como objetivo o sequenciamento, montagem, análise e caracterização do plastoma de *Passiflora cincinnata*. Os dados indicam uma massiva perda de genes plastidiais essenciais para a viabilidade celular (*infA*, *rps7*, *rps16*, *rpl20*, *rpl22*, *ycf1* e *ycf2*), os quais, muito provavelmente, foram transferidos para o núcleo e seus produtos são importados pelos plastídios. Este genoma mostrou alta taxa de substituição de nucleotídeos para os genes *accD* e *clpP*. Apesar da alta divergência, a sequência traduzida destes genes mantém a maioria dos domínios funcionais previstos para as proteínas que codificam e com isso a funcionalidade dos mesmos não pode ser descartada. Além disso, múltiplas inversões também foram detectadas no plastoma de *P. cincinnata*, mudando a ordem de vários genes. Em conjunto, os dados sugerem uma incomum evolução do plastoma de *P. cincinnata*, caracterizada por perdas gênicas, inversões no genoma e presença de genes com aceleradas taxas de substituição de nucleotídeos. Assim, é possível sugerir que esta instabilidade do genoma e perda de genes essenciais possa estar relacionada com a incompatibilidade entre núcleo e plastoma observada em híbridos de *Passiflora*. Por fim, a sequência completa do plastoma de *P. cincinnata*, obtida neste trabalho torna viável a transformação plastidial nesta espécie, visando aplicações biotecnológicas, além de estudos evolutivos e de genética funcional.

ABSTRACT

PACHECO, Túlio Gomes, M. Sc., Universidade Federal de Viçosa, July, 2016. **The complete plastid genome sequence of *Passiflora cincinnata*: genome rearrangements, massive plastid gene losses and implications to genome-plastome incompatibility**. Adviser: Marcelo Rogalski.

The plastid genome (plastome) organization, gene content and order is well conserved in most angiosperms, but there are some exceptions. The *Passiflora* genus is one of those exceptions, because there are evidences of some unusual plastid gene losses to species of this genus. However, none plastome of *Passiflora* has been published to date, making studies related to the evolution and putative high instability of plastome in this group difficult. In parallel, the study of the causes of nucleus-plastome incompatibility, observed in interspecific hybrids of *Passiflora*, has remained obscure due to the lack of plastid sequences in the database. In the context, starting to fill these gaps and to enable the characterization of plastid genetic markers and the construction of vectors for plastid transformation in *Passiflora*, the aim of the present study was the sequencing, assembly, analysis and characterization of complete *P. cincinnata* plastome. The data indicate a massive loss of plastid genes that are essential for cell viability (*infA*, *rps7*, *rps16*, *rpl20*, *rpl22*, *ycf1* and *ycf2*), which very likely were functionally transferred to the nucleus and its products are imported into plastid. This genome also showed a high rate of nucleotide substitution in several genes, such as *accD* and *clpP*. Despite this high divergence, the translated amino acid sequences of these genes retain most of functional domains predicted indicating that they can still encode functional proteins. In addition, multiple inversions were detected in the *P. cincinnata* plastome, changing the order of several genes. Taken together, the data suggest a markedly uncommon evolution of *P. cincinnata* plastome, characterized for gene losses, multiple inversions and genes with accelerated nucleotide substitution rates. Thus, it is possible to suggest that the genomic instability and essential genes losses, observed here, may be related to the genome-plastome incompatibility observed in *Passiflora* hybrids. This relation can be established and investigated of an accurate manner with the sequencing of other *Passiflora* plastomes. Finally, the complete plastome sequence of *P. cincinnata* obtained in this work enables the plastid transformation to this species, aiming biotechnology applications and studies of evolution and functional genetics.

1. Introduction

The evolutionary origin of plastid organelle is based on endosymbiotic process, in which a eukaryotic cell engulfed a cyanobacterium ancestor. Following this process, the new eukaryotic plant cell with novel biochemistry pathways (e.g. photosynthesis), started to be controlled by three genetic compartments: the nucleus, mitochondria and plastids (Timmis et al., 2004; Bock and Timmis, 2008; Zimorski et al., 2014; Rogalski et al., 2015). After the endosymbiosis, the plastid genome (plastome) underwent a dramatic reorganization, including the loss of dispensable genes, elimination of ambiguous genetic information between nucleus and plastome, transfer of plastid genes to the nucleus, importation of products (proteins) of these transferred genes into plastid and a complex interaction between nuclear and plastid gene products (Timmis et al., 2004; Bock e Timmis, 2008; Zimorski et al., 2014; Rogalski et al., 2015). These processes resulted in a drastic reduction of plastome size and gene content during the evolution of plant cell. Currently, the plastome size of green land plants ranges from 120 to 220 kb, and it encodes about 120-130 genes (Bock and Timmis, 2008; Zimorski et al., 2014; Rogalski et al., 2015). In contrast, the genome of a current cyanobacterium contains about 3200 genes (Kaneko et al., 1996), illustrating the massive gene losses experienced by plastome during the land plants evolution (Bock, 2015; Rogalski et al., 2015).

In general, land plant plastome contains a quadripartite structure, including a large single-copy region flanked on each side by inverted repeats (IRs) with a small single-copy region (SSC) joining the repeats; and this genome molecule encodes basically two groups of genes. The first group comprises components for the photosynthetic machinery such as photosystem I (PSI), photosystem II (PSII), the cytochrome b6/f complex and the ATP synthase (Shinozaki et al., 1986; Bock, 2001). The second group includes the genes required for plastid gene expression such as subunits of a plastid-encoded RNA polymerase (PEP), rRNAs, tRNAs, and ribosomal proteins. The first group of genes encoding components of photosynthetic machinery is not essential for cell survival and can be mutated individually under heterotrophic conditions as in vitro culture on sucrose-containing medium (Ruf et al., 1997; Bock, 2001; Hager et al., 2002). However, several components of gene expression machinery are essential for cell viability and their mutation leads to lethality (Ahlert et al., 2003; Rogalski et al., 2006, 2008; Fleischmman et al., 2011; Alkatib et al., 2012). Moreover, plastid genomes encode several other essential genes involved in other cellular function

than photosynthesis such as *ycf1* (encodes a subunit of the TIC/TOC machinery related to the plastid protein import apparatus; Kikuchi et al., 2013), *ycf2* (a gene of unknown function; Drescher et al., 2000), *clpP* (encodes the proteolytic subunit of the ATP-dependent Clp protease related to plastid protein homeostasis; Shikanai et al., 2001; Kuroda and Maliga, 2003), *accD* (encodes the β-carboxyl transferase subunit of a 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 is transcribed by plastid-encoded plastid RNA polymerase showing a dependence of plastid translation to be expressed and it is the precursor of δ-aminolevulinic acid involved in 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 these exceptions, such as Campanulaceae (Cosner et al., 1997, 2004; Haberle et al., 2008), Ericaceae (Fajardo et al., 2013), Geraniaceae (Chumley et al., 2006; Weng et al., 2013) and Fabaceae (Guo et al., 2007; Cai et al., 2008; Sabir et al., 2014; Schwarz et al., 2015). Several species from these families already have the plastome sequenced, assembled and available in NCBI database (http://www.ncbi.nlm.nih.gov/genome/browse/), allowing us to find and to study the uncommon plastome characteristics present in these lineages. Currently, the plastid database contains approximately 850 complete plastome sequences, whereas several angiosperm families and groups remain unexplored or poorly studied. Therefore, it is presumable that several unusual plastome characteristics (related to gene content, gene order and genome structure) and evolution patterns may be found in some of these lineages.

In the context of plastid genomic studies, an interesting angiosperm lineage that remains poorly explored is the Passifloraceae family. This family belongs to Malpighiales order and contains about 20 genera, with more expressiveness for *Passiflora* genus, and approximately 600 species. Brazil is a center of diversity of Passifloraceae family with 4 genera and about 150 species (Junqueira et al., 2005; Souza and Lorenzi, 2012). South America is considered the center of diversity of the *Passiflora* genus and contains approximately 95% of all species (Vanderplank, 1996; Nakasone and Paull, 1998). The *Passiflora* genus has several species used as resources for ornamental, medicinal and food purposes (Faleiro et al., 2005; Cerqueira-Silva et al.,

2014a). This genus can be organized into four subgenera: *Astrophea, Decaloba, Deidamioides* and *Passiflora* (Feuillet and MacDougal, 2004; Cerqueira-Silva et al., 2014a). Despites various economic applications of *Passiflora* genus and the high richness of genera and species in the Passifloraceae family, there is none plastome available, sequenced and assembled of any species belonging to this family in the plastid genome database. Interestingly, there are several additional characteristics that make this family a very interesting target for studies related to plastid genomes. For example, there are evidences in the literature suggesting an uncommon plastome evolution in this family (Downie et al., 1994; Jansen et al., 2007, 2011; Blazier et al., 2016). The main evidence was provided by Jansen et al. (2007), who observed genetic divergences in the plastome sequence of *Passiflora biflora* in comparison with other angiosperm plastomes. These authors suggested the loss of some genes and intron from conserved genes (Hansen et al., 2006). The examples mentioned above bring strong evidences for a drastic evolutionary variation in the plastome of *Passiflora* species in comparison with most angiosperms. However, the lack of complete plastome sequences of species from Passifloraceae impedes advanced studies related to gene and intron losses, and likewise derails the detection of putative rearrangements, accumulation of repetitive elements and analyzes of regions with acceleration rates of nucleotides substitution in the plastome of this genus.

The plastid genomic in *Passiflora* has still other very interesting aspect, which is related to interaction between nuclear and plastid gene products. In general, plastome contains genes that encode subunits of multiprotein complexes, such as photosynthetic apparatus, gene expression machinery and other enzymatic complexes. The products of these genes (proteins) need to interact with the products encoded by the nucleus to assembly functional multi-protein complexes into the plastids. Thus, nuclear and plastid genes belonging to the same protein complex co-evolve and the appearance of a mutation in a plastid gene requires a compensatory mutation in the nuclear gene that interacts with it, and vice versa. Consequently, functional and regulatory interaction between nuclear and plastid gene products (proteins) prevent the independent evolution of these genetic compartments adjusting harmoniously the evolution of plant cell (Greiner and Bock, 2013). The essential adjustment of plant cell genomes is apparent during interspecific hybridization, which induces a formation of multi-protein complexes constituted of nuclear and plastid proteins. If the two hybrid parental lines suffered different selection pressures affecting the cell genomes, it can result in severe

failures in the hybrid development and, in extreme cases, in the cell lethality (Levin, 2003; Herrmann et al., 2003; Schmitz-Linneweber, 2005; Greiner et al., 2011). The phenomenon of cellular incompatibility, between nucleus (genome) and plastome, is easily observed when a photosynthetic multi-protein complex composed of nuclear and plastid subunits is affected resulting in different phenotypes such as variegated and pale green to white (Greiner et al., 2011). Different cases of genome-plastome incompatibility in the nature were already observed in several angiosperm families, as summarized by Greiner et al. (2011). The first report of genome-plastome incompatibility in *Passiflora* genus was observed by Mráček (2005) who demonstrated that several interspecific hybrids between *P. menispermefolia* and several other *Passiflora* species display biparental inheritance of plastids and a variegated phenotype. Given its uncommon characteristics related to plastid evolution and plastid inheritance, the *Passiflora* genus is a very interesting plant model to nucleus-plastome interaction studies, but the genetic reasons of the incompatibility are totally unknown and have to be elucidated. In other plant groups, several causes of genome-plastome incompatibility were elucidated by comparative analysis of plastome sequences of parental plants. In hybrids of *Pisum* genus, the incompatibility is likely related to the high divergence of *accD* gene, which affects the synthesis of fatty acids in plastids (Bogdanova et al., 2015). In *Oenothora*, it is associated to the deletion of a promotor region, between the *clpP* and *psbB* genes, affecting the expression of photosynthesis-related genes (Greiner et al., 2008). The same situation is observed in experimental strategies *in vivo* using cybrids containing nuclear genome of *Atropa belladonna* and plastids of *Nicotiana tabacum*, which showed genome-plastome incompatibility. The reason of the incompatibility is the failure in one RNA editing site of *atpA* gene (encodes the α subunit of plastid ATP synthase; Shinozaki and Sugiura, 1986), showing a differential co-evolution of nuclear and plastid genes between members of Solanaceae family (Schmitz-Linneweber, 2005).

As we can see in the examples previously mentioned, the study of the reasons of the incompatibility showed in *Passiflora* hybrids will become feasible only after the sequencing of plastid genomes from several species of this genus. In addition to gene order, gene content and gene functionality, it is also necessary to analyze the regulatory sequences (5' and 3' untranslated regions), present in this genome given that they can also be determinants of the incompatibility. The suggested loss of essential genes (Downie et al., 1994; Jansen et al., 2007, 2011) and the apparent high variability of

plastome in *Passiflora* genus (Hansen et al., 2006), may be important characteristics that can be related to the incompatibility genome-plastome. However, the indication of the cause of incompatibility in an accurate manner is possible only with access of complete plastome sequences. Moreover, the available of plastome sequence of several *Passiflora* species, in database, would allow also the selection of species containing differences in the plastome (i.e. in coding and/or regulatory sequences) to generate new interspecific hybrids and to evaluate the role of these differences in the nucleus-plastome interactions.

Plastome sequences allow also the identification and to characterization of simple sequence repeats (SSR markers) of plastome, which have vast potential for use in genetic studies such as interspecific and intraspecific variability, gene flow, genetic divergence and genetic diversity (Powell et al., 1995). Cerqueira-Silva et al (2014b) characterized some nuclear SSRs in three *Passiflora* species (*P. edulis*, *P. cincinnata* and *P. setacea)*. These authors emphasized also the low available of genetic information, especially lack of plastid SSRs for different *Passiflora* species, which limits the studies of genetic variability in this genus. This gap will be filled with the available of complete plastome sequences of *Passiflora* species. Furthermore, plastid genome sequences may support phylogenetic studies with different approaches. In general, plastid genes have a low nucleotide substitution rate and, consequently, are excellent markers to phylogenetic studies at family, order and higher taxa level (Jansen et al., 2007; Moore et al., 2010; Vieira et al., 2014a; Rogalski et al., 2015). In addition, the plastome also contains genes and intergenic regions with high mutation rate that may be used to infer phylogenetic relationships between tribes, genus and lower taxa (Shaw et al., 2005, 2007; Rogalski et al., 2015).

Finally, the plastid genome sequence enables the plastid transformation of *Passiflora* species. The plastid transformation offers several advantages in comparison with nuclear transformation such as high transgene expression levels, multigene stacking in synthetic operons, precisely targeted insertion of transgenes via homologous recombination and absence of epigenetic transgene silencing (Bock et al., 2015; Jin and Daniell, 2015; Rogalski et al., 2015). To correct transgene or mutation insertion via homologous recombination, it is needed a previous knowledge of plastome sequence of the target species, given that a specific transformation vector is required containing homologous regions to the plastid DNA, flanking the transgene (or mutation) that will be inserted (Bock, 2015; Rogalski et al., 2015). Furthermore, the availability of

plastome sequence allows also a safety selection of the exact insertion site preventing coding or regulatory sequence disruption by insertion of the transgene, which would affect the expression of other endogenous plastid genes (Bock, 2013). Several examples of applications of this technology are available in the literature (Maliga, 2004; Maliga and Bock, 2011; Bock, 2015; Jin and Daniell, 2015; Daniell et al., 2016). Plastid genetic engineering has several biotechnology applications and allows also studies related to basic research as functional genetics to analyze the essentiality of highly divergent plastid genes in *Passiflora*.

In view of what was discussed above, the plastome sequencing of *Passiflora* species could fill gaps in several types of studies and applications. For this purpose, *Passiflora cincinnata* Mast. was selected as a good candidate to be the first *Passiflora* species with the complete plastome sequenced and assembled. It is a wild *Passiflora* species, popularly known as maracujá-do-mato (Bernacci et al., 2003) and belongs to the *Passiflora* subgenus (Killip, 1938). *P. cincinnata* is a native species of the Brazilian Caatinga and has nutritional (Kiill et al., 2010), medicinal and ornamental applications (Bernacci et al., 2003). It is also used as gene source in breeding programs for increasing resistance to pathogens (Oliveira and Ruggiero, 2005). Moreover, several protocols for *in vitro* regeneration of *P. cincinnata* are available in the literature, enabling it as candidate species for plastid transformation (Dornelas and Vieira, 1993; Reis, 2005; Lombardi et al., 2007; Silva et al., 2011)

In this context, the present work aimed the sequencing, assembling and characterization of *P. cincinnata* complete plastome, as starting point to studies of genome evolution, plastid genetic markers and nucleus-plastome interaction in the *Passiflora* genus. The complete plastome sequence of *P. cincinnata* reported here showed unusual losses and putative pseudogenezation of several essential genes, suggesting the occurrence of various events of functional transfer of plastid genes to the nucleus. This genome also showed alterations in gene order caused for multiple inversions and also accelerated nucleotide substitution rates for several genes. Several SSRs and other repetitive regions were characterized and curiously these repetitive elements were concentrated predominantly in the *accD/rbcL* intergenic region, a highly divergent region. Taken together, the results showed an uncommon evolution of *P. cincinnata* plastome if compared with most angiosperms. Its specific characteristics, as gene losses, genome rearrangements and high nucleotide substitution rates suggest the existence of a high instability of *Passiflora* plastomes. The instability and specific gene

losses can directly or indirectly be the reasons of genome-plastome incompatibility observed in this genus.

2. Material and methods

2.1. Plant material, chloroplast isolation and plastid DNA extraction

Plants of *Passiflora cincinnata* were cultivated under greenhouse conditions in the Department of Plant Biology, Federal University of Viçosa, MG, Brazil. Young leaves were collected and kept for 96 hours at 4°C to reduce starch and sugar content from chloroplasts. The chloroplast isolation and plastid DNA extraction procedures were carried out as described by Vieira et al. (2014b).

2.2. 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. Libraries were sequenced using MiSeq Reagent Kit v3 (600 cycles) on Illumina MiSeq Sequencer (Illumina Inc., San Diego, California, USA) at the Federal University of Paraná, Brazil. The paired-end reads (2 x 300bp) obtained in sequencing was submitted to *de novo* assembly using the CLC Genomics Workbench 8.0 software. Initial annotation of the *P. cincinnata* plastid genome was performed using Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004). From this initial annotation, putative starts, stops codons, and intron positions were determined based on comparisons to homologous genes from other plastid genomes. The tRNA genes were further verified by using tRNAscan-SE (Lowe and Eddy, 1997). A physical map of the plastid circular genome was drawn using OrganellarGenomeDRAW (OGDRAW) (Lohse et al., 2013), followed by manual modifications and adjustments.

2.3. Comparative analysis of genome structure

The Mauve Genome Alignment v2.3.1 (MAUVE) software (Darling et al., 2003) and the Nucleotide MUMmer (NUCmer) Perl script in MUMmer 3.0 (to dot-plot analysis) (Kurtz et al., 2004) were used to comparison of plastid genome structures between *P. cincinnata* and other representative species belonging to Malpighiales order, as *Licania alba*, *Hevea brasiliensis*, *Manihot esculenta*, *Viola seoulensis* and *Populus alba*, as well as *Arabidopsis thaliana*, belonging to Brassicales order and used as external representative and reference.

2.4. Repeat sequence analysis and RNA editing sites prediction

Simple sequence repeats (SSRs) in the *P. cincinnata* plastome were detected 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. Additional tandem repeats were analyzed using Tandem Repeats Finder (TRF) (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 as 50 and 500, respectively. REPuter (Kurtz et al., 2001) was used to localize IRs and count forward versus reverse complement (palindromic) within *P. cincinnata* plastome. The settings for the minimal repeat size was 30 bp and the identity of repeats was no less than 90% (hamming distance = 3). All of the repeats found were manually verified, and the nested or redundant results were posteriorly removed.

Potential RNA editing sites in protein-coding genes of *P. cincinnata* plastome were predicted by the program Predictive RNA Editor for Plants (PREP) suite (Mower, 2009), that use 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.*

2.5. Phylogenetic inference

Aiming to infer the phylogenetic position of *P. cincinnata* (family Passifloraceae), the complete plastid genome of other species within Malpighiales order, including the families Chrysobalanaceae (representants of five genera), Euphorbiaceae (four genera), Salicaceae (two genera), and Violaceae (one genus), were obtained from NCBI plastid genome database. Moreover, we used also for phylogenetic analysis the complete plastid genomes of other orders within Fabids: Rosales (representants of two families), Cucurbitales, Fabales e Fagales (one family for each). As an out-group, it was included a species of Malvids, *Arabidopsis thaliana* (Brassicales: Brassicaceae). The table 1 shows the complete plastid genomes used in the phylogenetic inference.

Nucleotide sequences of each protein-coding gene common to the plastid genomes of all species above listed were extracted and aligned individually using the software MUSCLE (Edgar, 2004) implemented in Mega 6.0 (Tamura et al., 2003).

Putative pseudogenes of *P. cincinnata* were removed from phylogenetic analysis, resulting in a set of 64 protein-coding genes. The software Sequence Matrix 1.7.8 (Vaidya et al., 2010) was used to concatenate the genes and a total sequence of 50799 nucleotides was obtained. In addition the software Partition Finder (Lanfear et al., 2012) was used to search for the best set of parameters to be optimized during the phylogenetic search for the best tree. Bayesian inference was performed using MrBayes version 3.2 (Ronquist et al., 2012), with 10 million generations of two runs of four Markov Chains, three hot and one cold in each run. The software Tracer 1.6 (http://tree.bio.ed.ac.uk/software/tracer/) was used to check the parameters convergence and the software FigTree 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize the consensus tree.

Table 1: Species used in the phylogenetic inference of the position of *P. cincinnata*.

Species	**Subclass**	**Clade**	**Order**	**Family**	**GenBank#**
Chrysobalanus icaco	rosids	fabids	Malpighiales	Chrysobalanaceae	NC_024061
Couepia guianensis	rosids	fabids	Malpighiales	Chrysobalanaceae	NC_024063
Hirtella physophora	rosids	fabids	Malpighiales	Chrysobalanaceae	NC_024066
Licania alba	rosids	fabids	Malpighiales	Chrysobalanaceae	NC_024064
Parinari campestris	rosids	fabids	Malpighiales	Chrysobalanaceae	NC_024067
Hevea brasiliensis	rosids	fabids	Malpighiales	Euphorbiaceae	NC_015308
Jatropha curcas	rosids	fabids	Malpighiales	Euphorbiaceae	NC_012224
Manihot esculenta	rosids	fabids	Malpighiales	Euphorbiaceae	NC_010433
Ricinus communis	rosids	fabids	Malpighiales	Euphorbiaceae	NC_016736
Populus alba	rosids	fabids	Malpighiales	Salicaceae	NC_008235
Salix interior	rosids	fabids	Malpighiales	Salicaceae	NC_024681
Viola seoulensis	rosids	fabids	Malpighiales	Violaceae	NC_026986
Morus mongolica	rosids	fabids	Rosales	Moraceae	NC_025772
Prunus persica	rosids	fabids	Rosales	Rosaceae	NC_014697
Cucumis melo	rosids	fabids	Cucurbitales	Cucurbitaceae	NC_015983
Glycine max	rosids	fabids	Fabales	Fabaceae	NC_007942
Castanea mollissima	rosids	fabids	Fagales	Fagaceae	NC_014674
*Arabidopsis thaliana**	rosids	malvids	Brassicales	Brassicaceae	NC_000932

*Out-group

2.6. Pairwise distance analysis and estimation of synonymous (dS) and nonsynonymous (dN) substitution rates

Pairwise distance was calculated based on the nucleotide sequences of 71 protein coding genes, including putative pseudogenes, identified in the plastome of *P. cincinnata*, comparing them with those of all species present in Table 1. The sequences of each gene were aligned individually by MUSCLE (Edgar, 2004) implemented in Mega 6.0 (Tamura et al., 2003), with Pairwise deletion set to Gaps / Missing data treatment. To each alignment was calculated the best model of nucleotide substitution of according the jmodeltest2 software (Darriba et al., 2012) and the matrix of pairwise distance was generated using the PAUP* version 4.0b5 (Swofford, 2002). Furthermore, for each of the group of genes mentioned above, the pairwise dS and dN values between *P. cincinnata* and all the species presented in the Table 1 were estimated using Mega 6.0 software (Tamura et al., 2003) under the Kumar model (Kimura 2-parameter).

All the genes and genomes used here, to comparative analysis with *P. cincinnata*, were obtained from NCBI and are shown accordingly in the Table 1. In addition, the sequences of the *rps18* gene, from *Pisum sativum* (NC_014057.1) and *Secale cereale* (NC_021761.1), and the *accD* gene, from *P. sativum* and *Medicago truncatula* f. *tricyla* (AGR83925.1), were also obtained from NCBI plastid genome database.

3. Results

3.1. Genome size and gene content of *P. cincinnata* plastid genome

The Illumina MiSeq reads obtained and submitted to *de novo* assembly resulted in a high genome average coverage, approximately 3,274-fold. The complete plastid genome of *P. cincinnata* is a circular DNA molecule of 150,783 bp and exhibits the general quadripartite structure, which is typical in most angiosperms. It includes a large single-copy region (LSC) of 86,264 bp flanked on each side by inverted repeats of 25,674 bp (IRs) with a small single-copy region (SSC) of 13,171 bp joining the repeats (Figure 1). In comparison with other representative species of Malpighiales order, the total size of *P. cincinnata* plastome is relatively reduced, with a predominant reduction in the SSC region (Table 2). The GC content determined for the *P. cincinnata* plastome is 37.1%, which is very similar to other flowering plants.

The *P. cincinnata* plastome is predicted to encode 106 unique genes, of which 14 are completely duplicated and the rps12 is partially duplicated in the IRs, resulting in a total of 121 genes (Table 3 and Figure 1). These genes include 72 unique protein-coding genes (of which 4 are completely and one is partial duplicated), 30 unique tRNA genes (6 are duplicated in the IRs) and 4 unique rRNA genes (all completely duplicated in the IRs). The *infA*, *rps16* and *rpl22* genes are absent in *P. cincinnata* plastome. The intron of *atpF* gene and the two introns of *clpP* gene are also absent in this genome. The *rps7* and *rpl20* genes are putative pseudogenes given that they contain premature stop codons into the coding sequence, one and three, respectively. Other pseudogenes identified in this genome were the *ycf1* and *ycf2*. The *ycf1* gene contains several internal stop codons and lacks approximately 2 kb of its coding-sequence. The *ycf2* gene contains premature stop codons, and additionally, it lost the start codon and about 3.2 kb of its coding sequence in comparison with other rosid species.

Furthermore, the *ndhB* and *rps18* genes also present uncommon structures in their coding sequencing in *P. cincinnata* plastome. The *ndhB* gene has an unusual start codon (AGG), seeing that plastid genes of higher plants contain mostly ATG and sometimes GTG as start codon. However, it is possible that in *P. cincinnata* plastome the translation of *ndhB* gene starts in the next ATG into the coding region (approximately 50 pb downstream of the conserved position of the start codon), and consequently, it could be functional. In other hand, the *rps18* gene in *P. cincinnata* plastome is also likely functional, but it contains a relative large expansion closed to the

C-terminus sequence, in comparison with other species belonging to Malpighiales order and fabid clade.

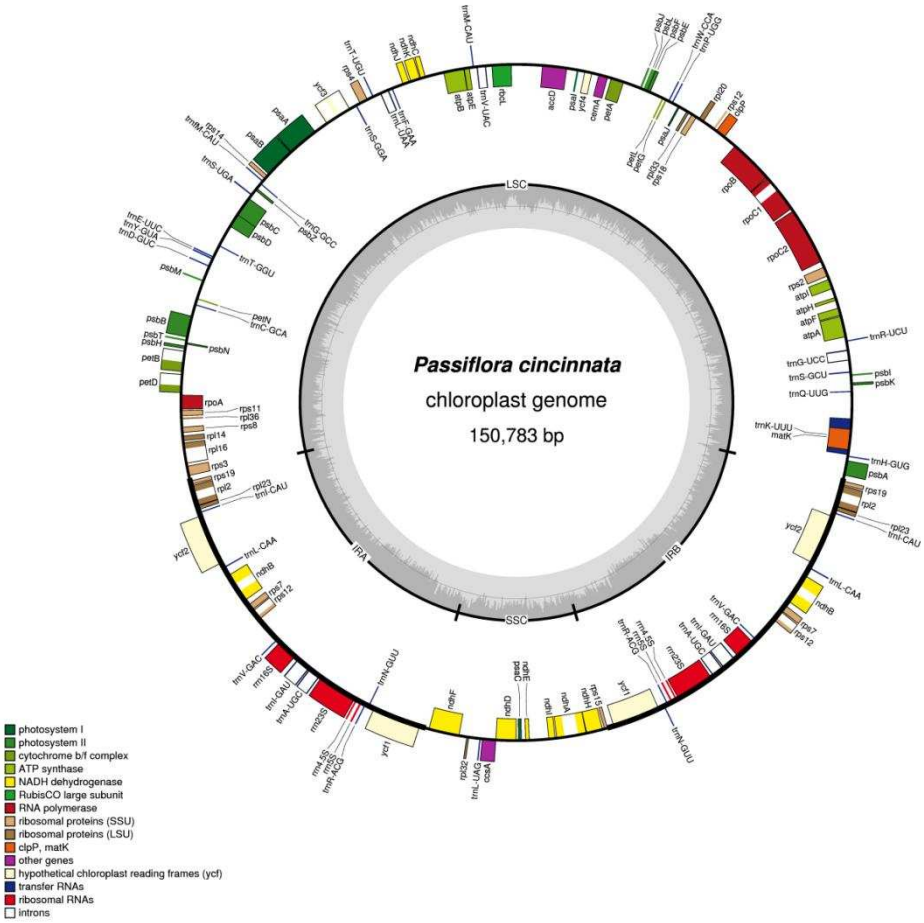

Figure 1. Genome organization and gene content of *Passiflora cincinnata* plastome. Two inverted repeat regions IR_A and IR_B, divide the rest of the circular DNA molecule into large (LSC) and small (SSC) single copy regions. Genes drawn inside the circle are transcribed clockwise, and genes drawn outside are expressed counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray in the inner circle corresponds to GC content, and the lighter gray represents the AT content.

Table 2. Comparison of plastome size between *Passiflora cincinnata* and representative species from other families of Malpighiales order.

Specie	Family	Size (bp)	LSC (bp)	SSC (bp)	IR (bp)
Passiflora cincinnata	Passifloraceae	150,783	86,264	13,171	25,674
Populus alba	Salicaceae	156,505	84,618	16,567	27,660
Viola seoulensis	Violaceae	156,507	85,691	18,008	26,404
Manihot esculenta	Euphorbiaceae	161,453	89,295	18,250	26,954
Licania alba	Chrysobalanaceae	162,467	88,927	19,740	26,900

Table 3: List of genes identified in *Passiflora cincinnata* plastome.

Group of gene	Name of gene
Self-replication	
Ribosomal RNA genes	*rrn16*[b]*; rrn23*[b]; *rrn5*[b]*; rrn4.5*[b]
Transfer RNA genes	*trnA* –UGC[ab]; *trnC* –GCA; *trnD* –GUC; *trnE* –UUC; *trnF* –GAA; *trnfM* –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
Small subunit of ribosome	*rps2; rps3; rps4*; *rps8; rps11; rps12*[ac]; *rps14; rps15*; *rps18; rps19*[b]
Large subunit of ribosome	*rpl2*[ab]; *rpl14; rpl16*[a]; *rpl23*[b]*; rpl32; rpl33; rpl36*
DNA-dependent RNA polymerase	*rpoA; rpoB; rpoC1*[a]; *rpoC2*
Genes for photosynthesis	
Subunits of photosystem I (PSI)	*psaA; psaB; psaC; psaI; psaJ; ycf3*[a]; *ycf4*
Subunits of photosystem I (PSI)	*psbA*; *psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM; psbN; psbT; psbZ*
Subunits of cytochrome	*petA; petB*[a]*; petD*[a]*; petG; petL; petN*
Subunits of ATP synthase	*atpA; atpB; atpE; atpF*; *atpH; atpI*
Subunits of NADH dehydrogenase	*ndhA*[a]; *ndhB*[ab]; *ndhC; ndhD; ndhE; ndhF; ndhG; ndhH*; *ndhI; ndhJ; ndhK*
Large subunit of Rubisco	*rbcL*
Others genes	
Maturase	*matK*
Envelope membrane protein	*cemA*
Subunit of acetyl-CoA carboxylase	*accD**
C-type cytochrome synthesis gene	*ccsA*
ATP-dependent Protease	*clpP**
Pseudogenes	*rps7*[b]*; rpl20; ycf1*[b]*; ycf2*[b]
Genes absent	*rps16; rpl22; infA*

[a]Gene containing introns; [b]Duplicated gene; [c]Partial duplicated gene; *highly divergent gene;

3.2. Comparative analysis of genome structure and IR boundary

The structure of *P. cincinnata* plastome was here compared with plastomes of other representative species of Malpighiales order and *Arabidopsis thaliana* (as an external reference species), via the multiple genome alignment analysis produced by MAUVE software. This alignment showed the presence of rearrangements in *P. cincinnata* plastome, which are not present in other species of Malpighiales order (Figure 2). In the same way, we perform also a dot plot analysis using the plastome of the same species against the *P. cincinnata* plastome for an additional exploration of these rearrangements (Figure 3). One of these rearrangements was characterized to be an inversion of approximately 1.6 kb, in the beginning of the LSC region, that changed the order and direction of the coding region of *trnH*-GUG and *psbA* genes (Figure 2; Figure 3A-F, number1). Furthermore, there is still a significantly complex rearrangement in the *P. cincinnata* plastome corresponding to a large inversion of approximately 47 kb in the LSC region, extending from *trn*C-GCA gene till *clp*P gene, whereas a sequence of 3 kb approximately in the middle of this inversion, from *trnV*-UAC to *atpB* genes, remains in the same direction as observed in the other species analyzed here (Figure 2; Figure 3A-F, number 2). Besides theses rearrangements in the *P. cincinnata* plastome, reported by the first time here, another genome rearrangement was already described in *Hevea brasilienses* (Tangphatsornruang et al., 2011) and may be visualized in both analyses, multiple genome alignment and dot-plot analysis done here (Figure 2; Figure 3C, number 3).

The IR boundaries were compared between *P. cincinnata,* 5 species representatives of Malpighiales order and *A. thaliana*. It is possible to visualize a dynamic process of contraction/expansion of IR in the Malpighiales species (Figure 4). The IRa/LSC boundaries are very similar between *A. thaliana, H. brasiliensis* and *Viola seoulensis*, although a small IR contraction may be visualized in the latter. An expansion of IR in the IRa/LSC boundaries may be visualized in all of the other species in analysis, predominantly, in *Populus alba* and *P. cincinnata* in which this expansion included a full *rps19* in their IRs. In relation to the SSC/IRs boundaries, they are very similar between *A. thaliana, Licania alba* and *V. seoulensis*. An expansion of IR in this boundary including a relative largest portion of *ycf1* gene may be observed in *H. brasilienses, Manihot esculenta, P. alba* and *P. cincinnata.* In this latter, the *ycf1* gene was degenerated into a pseudogene and the IR expansion was of such magnitude that

included the full sequence of the *ycf1* pseudogene and more an intergenic region (Figure 4).

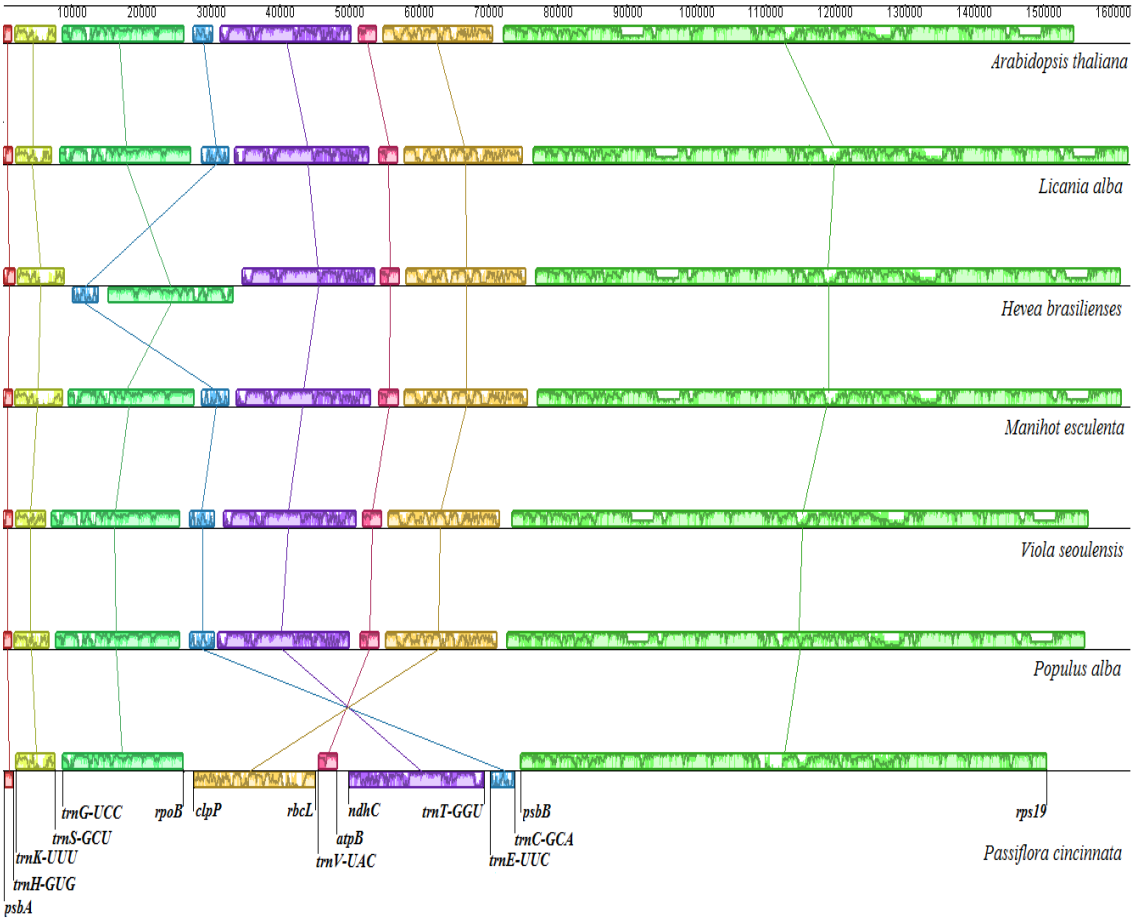

Figure 2. Comparison of plastid genome structures between *P. cincinnata* and other representative species of Malpighiales order with *A. thaliana*, as an external reference member, produced by MAUVE. The boxes above the line represent the gene complex sequences in clockwise direction and the boxes below the line demonstrate those sequences in the opposite direction. The genes described at the bottom of boxes indicate the genes localized at the boundaries of the gene complex (gene groups) that suffered rearrangements in the plastome of *P. cincinnata* plastome in comparison to other species used in this manuscript.

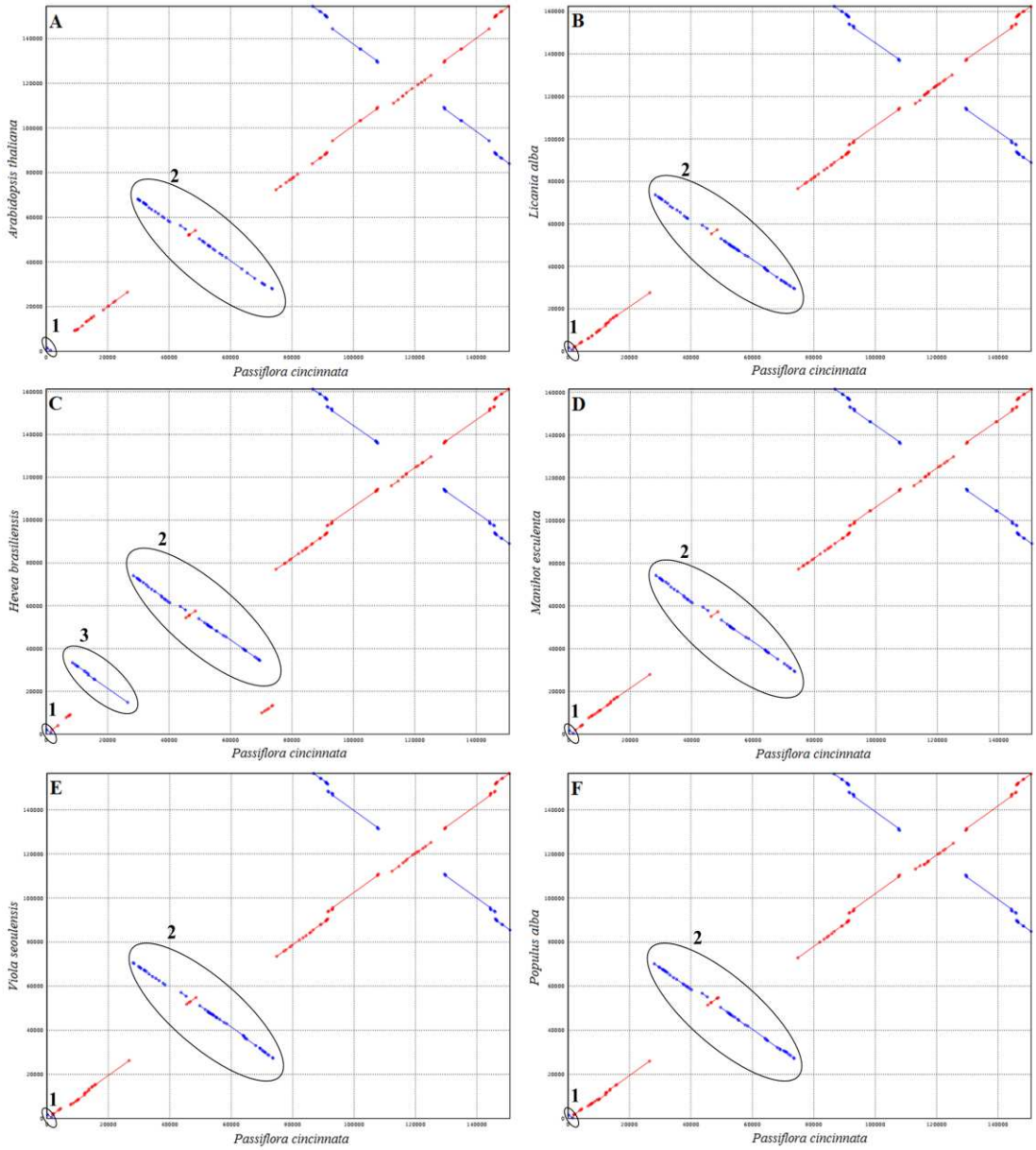

Figure 3. Dot-plot analysis using the *P. cincinnata* plastome (x axis) against other representatives of Malpighiales and *A. thaliana* (y axis). A positive slope denotes that the sequences are in the same orientation. A negative slope denotes that the sequences are in opposite orientation. Sequence in the same direction are indicated in red and inversions found in the genome are blue colored. Highlight circles (1) and (2) in A-F indicate rearrangement present in *P. cincinnata* and (3) in C indicate rearrangement present in *H. brasiliensis* plastome.

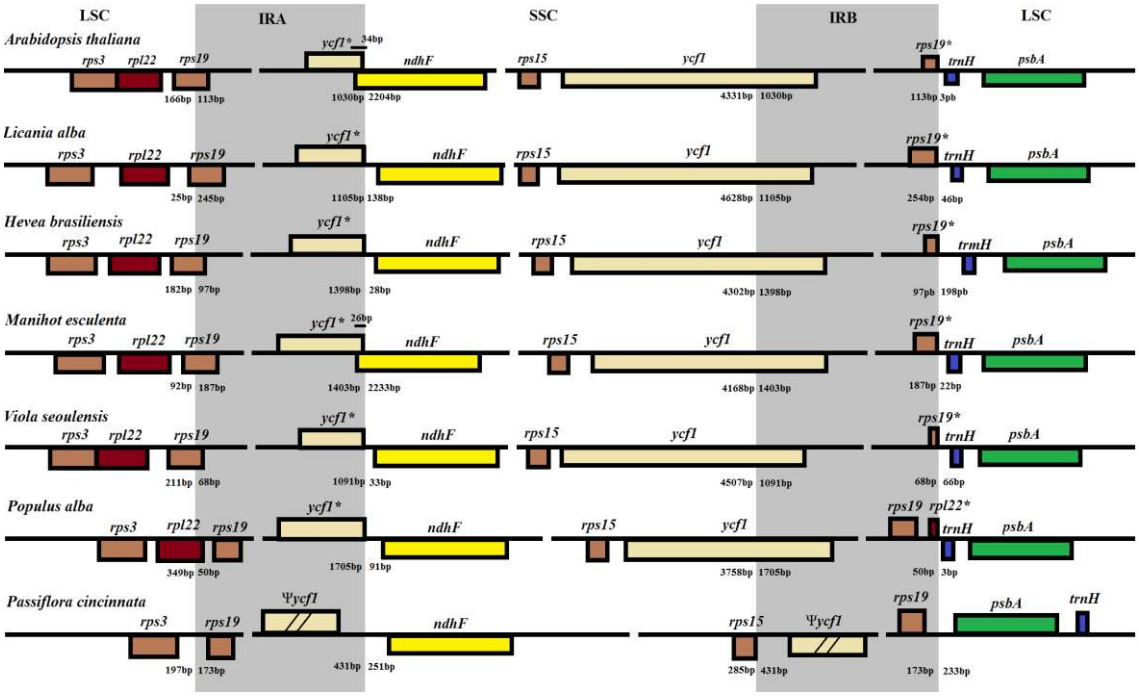

Figure 4. Comparison of the IR structures between plastomes *P. cincinnata*, five other species from Malpighiales and *A. thaliana.* Genes are represented by boxes with genes above the line being transcribed right to left and those below the line transcribed left to right. Genes belonging to the same functional groups have the same color. The number of base pairs of the gene or intergenic regions in the IR junctions is indicated below the line, except to the overlap between *ndhF* and *ycf1*, which is indicated above the line. * indicates partially duplicated genes; ψ indicates pseudogenes.

3.3. Repeat sequence analysis and RNA editing prediction

The occurrence, type and distribution of SSRs in *P. cincinnata* plastome were analyzed. In total, 213 SSRs were identified. Among them, homo- and dipolymers were the most common with, respectively, 141 and 58 occurrences, whereas tri-(3) and tetrapolymers (11) occurred with lower frequency (Table 4). Most of homo- (96,4%) and dipolymers (81%) are constituted by A and T bases. The size, sequence, and location of the tri- and tetrapolymers are shown in table 5. From these 13 polymers identified, 6 are located in the intergenic region between *accD* and *rbcL* genes, 5 in other intergenic regions and 2 in coding sequences.

Tandem repeats with more than 30 pb and with sequence identify of more than 90% have also been examined. We found 51 tandem repeats in the *P. cincinnata* plastome, of which 11 are located in coding regions of *accD* (4), *rps3* (1), *rps18* (1),

rps19 (1), *ndhF* (1), *ndhH* (1), *psbI* (1) and *trnP*-UGG genes (1); 2 are located in the intron of *rpl16* (1) and *trnF*-GAA (1) genes; 5 are located in the pseudogenes *ycf1* (2) and *ycf2* (3); 33 are distributed in the intergenic spacers around the genome (Table 6) and 8 out of them are located in the *accD*/rbcL intergenic region. In addition, 18 direct repeats (≥30 bp) were identified in the *P. cincinnata* plastome (Table 7). Among them, 2 are located in coding sequences of *psaA* and *psaB* genes, 3 are located in coding regions of *accD*, 4 are located in coding regions of *accD* and in the *accD*/rbcL intergenic region, and 9 are located in the *accD*/rbcL intergenic region.

The software Predictive RNA Editor for Plants identified in the *P. cincinnata* plastome a total of 64 putative RNA editing sites distributed in 21 genes (Table 8). Among them, 11 were identified in *ndhB*, 11 in *ndhD*, 7 in *accD*, 6 in *rpoC2*, 5 in *ndhA* and 4 in *matK*.

Table 4. List of simple sequence repeats (SSRs) identified in *P. cincinnata* plastome.

SSR Sequence	**Number of repeats**												**Total**
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	
A/T	-	-	-	-	-	50	43	24	11	4	2	2	136
C/G	-	-	-	-	-	5	-	-	-	-	-	-	5
AG/CT	-	10	1	-	-	-	-	-	-	-	-	-	11
AT/AT	-	33	7	3	4		-	-	-	-	-	-	47
AAT/ATT	-	1	-	-	-	-	-	-	-	-	-	-	1
ACG/CGT	-	-	1	1	-	-	-	-	-	-	-	-	2
AAAG/CTTT	7	-	-	-	-	-	-	-	-	-	-	-	7
AAAT/ATTT	2	-	-	-	-	-	-	-	-	-	-	-	2
AACT/AGTT	1	-	-	-	-	-	-	-	-	-	-	-	1
AGAT/ATCT	1	-	-	-	-	-	-	-	-	-	-	-	1
Total													**213**

Table 5. Distribution of tri- and tetrapolymer SSRs loci in *P. cincinnata* plastome.

SSR type	SSR sequence	Size	Start-End	Localization
tri	(GAC)6	18	27542-27559	*clpP* (CDS)
tetra	(TTCT)3	12	42832-42843	*accD*/*rbcL* (IGS)
tetra	(TTCT)3	12	42895-42906	*accD*/*rbcL* (IGS)
tetra	(TCTT)3	12	42907-42918	*accD*/*rbcL* (IGS)
tetra	(TTCT)3	12	42969-42981	*accD*/*rbcL* (IGS)
tetra	(TCTT)3	12	42992-43004	*accD*/*rbcL* (IGS)
tetra	(TTCT)3	12	43054-43065	*accD*/*rbcL* (IGS)
tri	(ATA)4	12	69771-69782	*trnT*-GGU/*trnE*-UUC (IGS)
tetra	(CTTT)3	12	72842-72853	*psbM*/*petN* (IGS)
tetra	(TTTA)3	12	97902-97913	*rps12*/*trnV*-GAC (IGS)
tetra	(TTAG)3	12	98508-98519	*rps12*/*trnV*-GAC (IGS)
tetra	(AATA)3	12	117212-117223	*ndhD* (CDS)
tetra	(ATAG)3	12	120673-120684	*ndhG*/*ndhI* (IGS)

Table 6. Distribution of tandem repeats in *P. cincinnata* plastome.

Number	Consensus size X Copy Number	Start-End	Localization
1	3 x 18	4654-4703	*trnK*-UUU/*trnQ-UUG* (IGS)
2	3 x 17	6297-6347	*trnQ*-UUG/*psbK* (IGS)
3	2 x 66	7239-7365	*psbK*/*psbI* (IGS/CDS)
4	2 x 13	9598-9622	*trnG*-UCC/*trnR-UCU* (IGS)
5	3 x 13	9875-9917	*trnR*-UCU/*atpA* (IGS)
6	2 x 23	13454-13499	*atpH*/*atpI* (IGS)
7	2 x 15	15652-15681	*rps2*/*rpoC2* (IGS)
8	2 x 18	29596-29631	*rps18* (CDS)
9	2 x 14	31149-31176	*psaJ*/*trnP*-UGG (IGS)
10	2 x 15	31581-31609	*trnP*-UGG/*trnW*-CCA (CDS/IGS)
11	2 x 17	35112-35145	*psbJ*/*petA* (IGS)
12	3 x 14	37338-37375	*cemA*/*ycf4* (IGS)
13	2 x 18	40378-40413	*accD* (CDS)
14	3 x 15	40503-40547	*accD* (CDS)
15	3 x 54	40718-40879	*accD* (CDS)
16	3 x 48	41082-41228	*accD* (CDS)
17	2 x 96	41344-41535	*accD*/*rbcL* (IGS)
18	2 x 24	41545-41590	*accD*/*rbcL* (IGS)
19	2 x 75	41638-41777	*accD*/*rbcL* (IGS)
20	5 x 47	41860-42095	*accD*/*rbcL* (IGS)
21	7 x 27	42296-42490	*accD*/*rbcL* (IGS)
22	3 x 65	42782-42982	*accD*/*rbcL* (IGS)
23	3 x 74	42831-43033	*accD*/*rbcL* (IGS)
24	2 x 15	43293-43322	*accD*/*rbcL* (IGS)
25	2 x 15	49251-49280	*atpB*/*ndhC* (IGS)
26	2 x 21	51448-51489	*ndhJ*/*trnF*-GAA (IGS)
27	2 x 17	52161-52194	*ndhJ*/*trnF*-GAA (IGS)

28	2 x 13	52677-52701	*trnF*-GAA (Intron)
29	2 x 15	53517-53545	*trnL*-UAA/*trnT*-UGU (IGS)
30	2 x 15	53574-53603	*trnL*-UAA/*trnT*-UGU (IGS)
31	2 x 13	57803-57828	*trnT*-UGU/*rps4* (IGS)
32	2 x 22	63066-63109	*psaB*/*rps14* (IGS)
33	2 x 16	64641-64672	*psbZ*/*trnS*-UGA (IGS)
34	2 x 19	69459-69496	*trnT*-GCU/*trnE*-UCC (IGS)
35	2 x 17	69531-69564	*trnT*-GCU/*trnE*-UCC (IGS)
36	2 x 17	69949-69984	*trnT*-GCU/*trnE*-UCC (IGS)
37	2 x 13	77104-77128	*psbH*/*petB* (IGS)
38	2 x 16	84615-84646	*rpl16* (Intron)
39	3 x 24	85720-85794	*rps3* (CDS)
40	2 x 25	86423-86472	*rps19* (CDS)
41	2 x 24	89019-89066	*trnI*-CAU/*ycf2* (IGS/ψ)
42	3 x 18	89257-89310	*ycf2* (ψ)
43	2 x 23	92569-92614	*ycf2* (ψ)
44	2 x 16	98853-98884	*rps12*/*trnV*-GAC (IGS)
45	2 x 21	106712-106753	*rrn4.5*/*rrn5* (IGS)
46	3 x 40	107806-107926	*trnN*-GUU/*ycf1* (IGS)
47	2 x 18	110455-110490	*ycf1* (ψ)
48	3 x 18	110597-110647	*ycf1* (ψ)
49	2 x 15	112170-112200	*ndhF* (CDS)
50	2 x 15	120754-120783	*ndhG*/*ndhI* (IGS)
51	2 x 12	123847-123870	*ndhH* (CDS)

Table 7. List of dispersed repeated sequences in *P. cincinnata* plastome.

First repeat copy		**Second repeat copy**		**Mismatch**	**Type**	**Size (bp)**
Start	**Location**	**Start**	**Location**			
40300	*accD* (CDS)	40432	*accD* (CDS)	-3	F	55
40744	*accD* (CDS)	40852	*accD* (CDS)	0	F	58
41099	*accD* (CDS)	41195	*accD* (CDS)	-3	F	54
41127	*accD* (CDS)	42042	*accD*/*rbcL* (IGS)	-3	F	58
41180	*accD* (CDS)	42047	*accD*/*rbcL* (IGS)	-3	F	53
41184	*accD* (CDS)	41343	*accD*/*rbcL* (IGS)	-3	F	52
41310	*accD* (CDS)	41826	*accD*/*rbcL* (IGS)	0	F	62
41353	*accD*/*rbcL* (IGS)	41449	*accD*/*rbcL* (IGS)	-3	F	60
41353	*accD*/*rbcL* (IGS)	41965	*accD*/*rbcL* (IGS)	-3	F	60
41377	*accD*/*rbcL* (IGS)	41473	*accD*/*rbcL* (IGS)	-3	F	71
41377	*accD*/*rbcL* (IGS)	41989	*accD*/*rbcL* (IGS)	-3	F	71
41402	*accD*/*rbcL* (IGS)	41918	*accD*/*rbcL* (IGS)	-1	F	142
41439	*accD*/*rbcL* (IGS)	41859	*accD*/*rbcL* (IGS)	-3	F	52
41450	*accD*/*rbcL* (IGS)	42014	*accD*/*rbcL* (IGS)	-3	F	54
41859	*accD*/*rbcL* (IGS)	42113	*accD*/*rbcL* (IGS)	-3	F	52
42513	*accD*/*rbcL* (IGS)	42594	*accD*/*rbcL* (IGS)	0	F	54
59422	*psaA* (CDS)	61646	*psaB* (CDS)	-3	F	79
59434	*psaA* (CDS)	61658	*psaB* (CDS)	-2	F	67

CDS, coding sequences; IGS, intergenic spacers. F, forward repeats.

Table 8. RNA editing sites predicted in the *P. cincinnata* plastome by PREP-cp software.

Gene	Nt position	AA position	Codon change	AA change
accD	709	237	**C**CG-**T**CG	P-S
	820	274	**C**CG-**T**CG	P-S
	850	284	**C**CG-**T**CG	P-S
	932	311	C**C**G-C**T**G	P-L
	992	331	C**C**G-C**T**G	P-L
	1093	365	**C**CG-**T**CG	P-S
	1511	504	T**C**A-T**T**A	S-L
atpI	629	210	TC**A**-TC**C**	S-L
clpP	664	222	**C**AT-**T**AT	H-I
matK	538	180	**C**TC-**T**TC	L-F
	568	182	**C**CG-**T**CG	P-S
	1180	394	**C**GG-**T**CG	R-W
	1190	397	T**C**A-T**T**A	S-L
ndhA	68	23	G**C**T-G**T**T	A-V
	326	109	A**C**T-A**T**T	T-I
	341	114	T**C**A-T**T**A	S-L
	419	140	A**C**A-A**T**A	T-I
	949	317	**C**TC-**T**TC	L-F
ndhB	25	9	**C**TC-TTC	L-F
	95	32	T**C**A-T**T**A	S-L
	413	138	C**C**A-C**T**A	P-L
	532	178	**C**AT-**T**AT	H-Y
	557	186	T**C**G-T**T**G	S-L
	692	231	T**C**T-T**T**T	S-F
	776	259	T**C**A-T**T**A	S-L
	782	261	T**C**A-T**T**A	S-L
	1201	401	**C**AT-**T**AT	H-Y
	1268	423	G**C**T-G**T**T	A-V
	1427	476	C**C**A-C**T**A	P-L
ndhD	2	1	A**C**G-A**T**G	T-M
	20	7	T**C**A-T**T**A	S-L
	26	9	A**C**A-A**T**A	T-I
	47	16	T**C**T-T**T**T	S-F
	313	105	**C**GG-**T**GG	R-W
	545	182	G**C**T-G**T**T	A-V
	599	200	T**C**A-T**T**A	S-L
	674	225	T**C**A-T**T**A	S-L
	1204	402	**C**TT-TTT	L-F
	1405	469	**C**TC-**T**TC	L-F
	1454	485	T**C**A-T**T**A	S-L
ndhF	149	50	A**C**A-A**T**A	T-I
	586	196	**C**TT-**T**TT	L-F
ndhG	166	56	**C**AC-**T**AC	H-Y
	250	84	**C**CA-**T**CA	P-S
petD	422	141	G**C**G-G**T**G	A-V
psaI	25	9	**C**CT-**T**CT	P-S
	83	28	T**C**T-T**T**T	S-F
psbE	214	72	**C**CT-**T**CT	P-S
psbF	77	26	T**C**T-TTT	S-F
rpl2	596	199	G**C**C-G**T**C	A-V
rpoA	881	294	T**C**G-T**T**G	S-L
rpoB	338	113	T**C**T-T**T**T	S-F
	551	184	T**C**A-T**T**A	S-L
	1279	427	**C**CT-**T**CT	P-S
rpoC1	41	14	T**C**A-T**T**A	S-L
rpoC2	1606	536	**C**TT-**T**TT	L-F
	2179	727	**C**TT-**T**TT	L-F
	2182	728	**C**CT-**T**CT	P-S
	2311	771	**C**GG-**T**GG	R-W
	2711	904	T**C**G-T**T**G	S-L
	3779	1260	T**C**A-T**T**A	S-L
rps2	248	83	T**C**A-T**T**A	S-L
rps8	403	135	**C**AC-**T**AC	H-Y
rps14	149	50	C**C**A-C**T**A	P-L

3.4. Phylogenetic inference

Bayesian inference (BI) using a data matrix including 64 plastid protein-coding genes of 19 taxa (Table 1) was performed using the best model of nucleotide substitution to each alignment of according PartitionFinder (Table 9). BI generated a phylogenetic tree with a –lnL = -215513.0311 (Figure 5). The BI posterior probability (PP) values were 1.00 for 12 nodes, 0.99 for 2 nodes, and 0.96, 0.66, 0.62 to one node each.

Within the fabid clade, two major groups were supported: a clade corresponding to the Malpighiales order (Figure 5; gray square) that is a sister group to other clade including the others fabid orders (Rosales, Cucurbitales, Fabales and Fagales). Within the Malpighiales clade, two major groups were formed, the first clade including the Chrysobalanaceae and Euphorbiaceae families and the second clade including Passifloraceae, Salicaceae and Violaceae. In this last clade, Passifloraceae was a sister group with Salicaceae and both these families formed a sister group to Violaceae family.

Table 9. Plastid genes used in phylogenetic inference clustered according to the best set of parameters found in PartitionFinder.

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

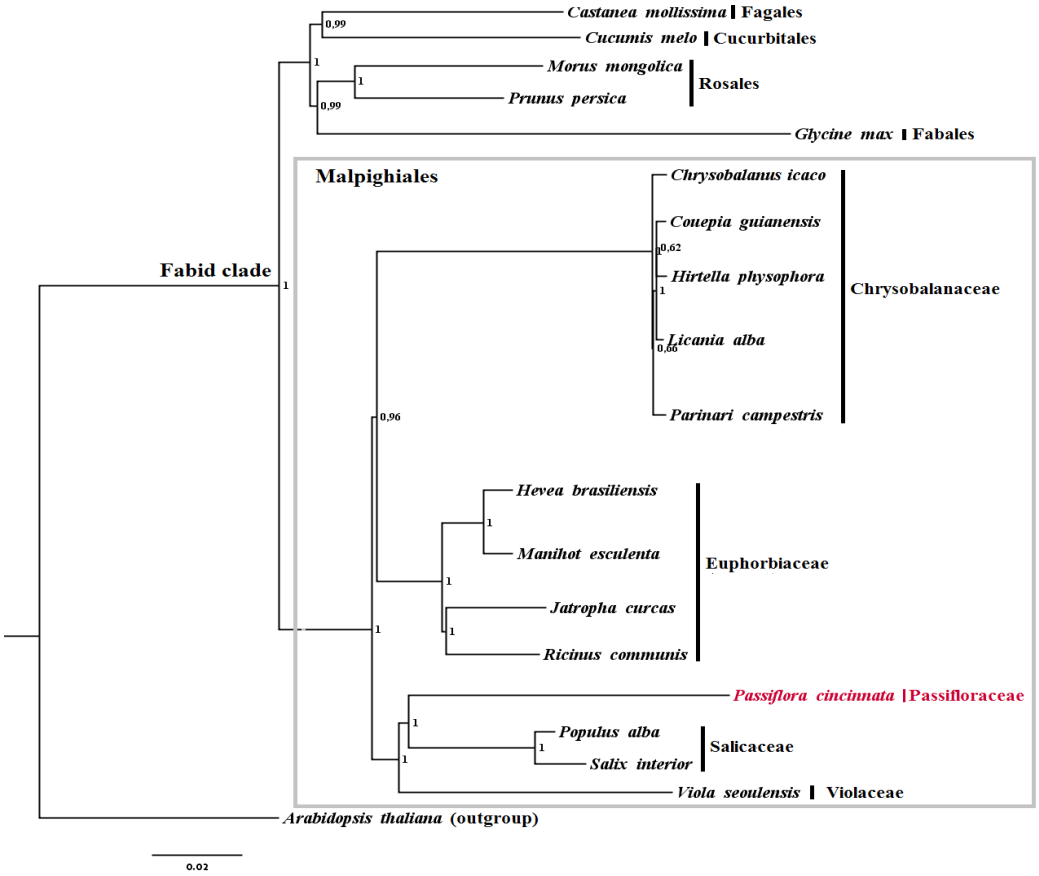

Figure 5. Phylogenetic tree constructed from 64 concatenated plastid protein-coding genes showing the position of *P. cincinnata* within Malpighiales order and Fabids clade. Phylogenetic reconstruction was performed using MrBayes with GTR+I+G and GTR+G models. Posterior probabilities are indicated in 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.

3.5. Pairwise distance and synonymous (dS) and nonsynonymous (dN) substitution rates

Pairwise distance based on nucleotide sequence was estimated using 71 protein-coding genes present in all species listed in the table 1 (Figure 6). These 71 genes include *rpl20*, *ycf1* and *ycf2* (Figure 6; red-colored), which were identified in *P. cincinnata* as putative pseudogenes due to the presence of internal stop codons and loss of portion of their coding regions. The *clpP* and *accD* genes showed a strongly higher value of pairwise distance in comparison with all other genes (including the pseudogenes already identified). The next highest values of substitution rates were visualized, in decreasing order, to the genes *rps11*, *rps15*, *ycf1* (pseudogene), *ccsA*, *matK*, *rpl20* (pseudogene), whereas the *ycf2* pseudogene showed surprisingly low substitution rates.

To additional analysis of gene divergences in *P. cincinnata* plastome, synonymous (dS) and nonsynonymous (dN) and dN/dS values were estimated for the same 71 genes. In general, values of *dN* are higher in *P. cincinnata* in comparison to other species in analysis (Figure 7). Our analyses also revealed that *accD*, *clpP*, ribosomal protein genes and the pseudogenes *ycf1* and *rpl20* showed most significantly differences if we compare them to genes involved in photosynthesis (*ndh*, *psa*, *psb* genes) (Figure 7). Values of dS exhibit similar, although with less significant differences, patterns of variation (Figure 7). The *clpP* gene of *P. cincinnata* showed a dN value (0.36) higher than all other genes and strongly higher that the dN value (0.07) to the *clpP* in the other species in analysis. The second higher value of dN (0.25) was visualized to the *accD* gene of *P. cincinnata* that also showed higher value in comparison with the dN (0.11) of this gene in the other species. Also, *clpP* of *P. cincinnata* showed high value of dN/dS (0.93), lower only that the *ycf1* pseudogene, and higher whether compared with the value found in other species (0.35). Considering the *accD* gene, in the plastome of *P. cincinnata*, it also showed a higher value of dN/dS (0.55), in comparison with those of the other species (0.49). Taken together, these results showed the existence of an exceptional high rate of substitution in several plastid genes of *P. cincinnata* plastid genome, predominantly in *clpP* and *accD*, allowing hypothesize that both gene are in a process of degeneration and could be pseudogenes in this species.

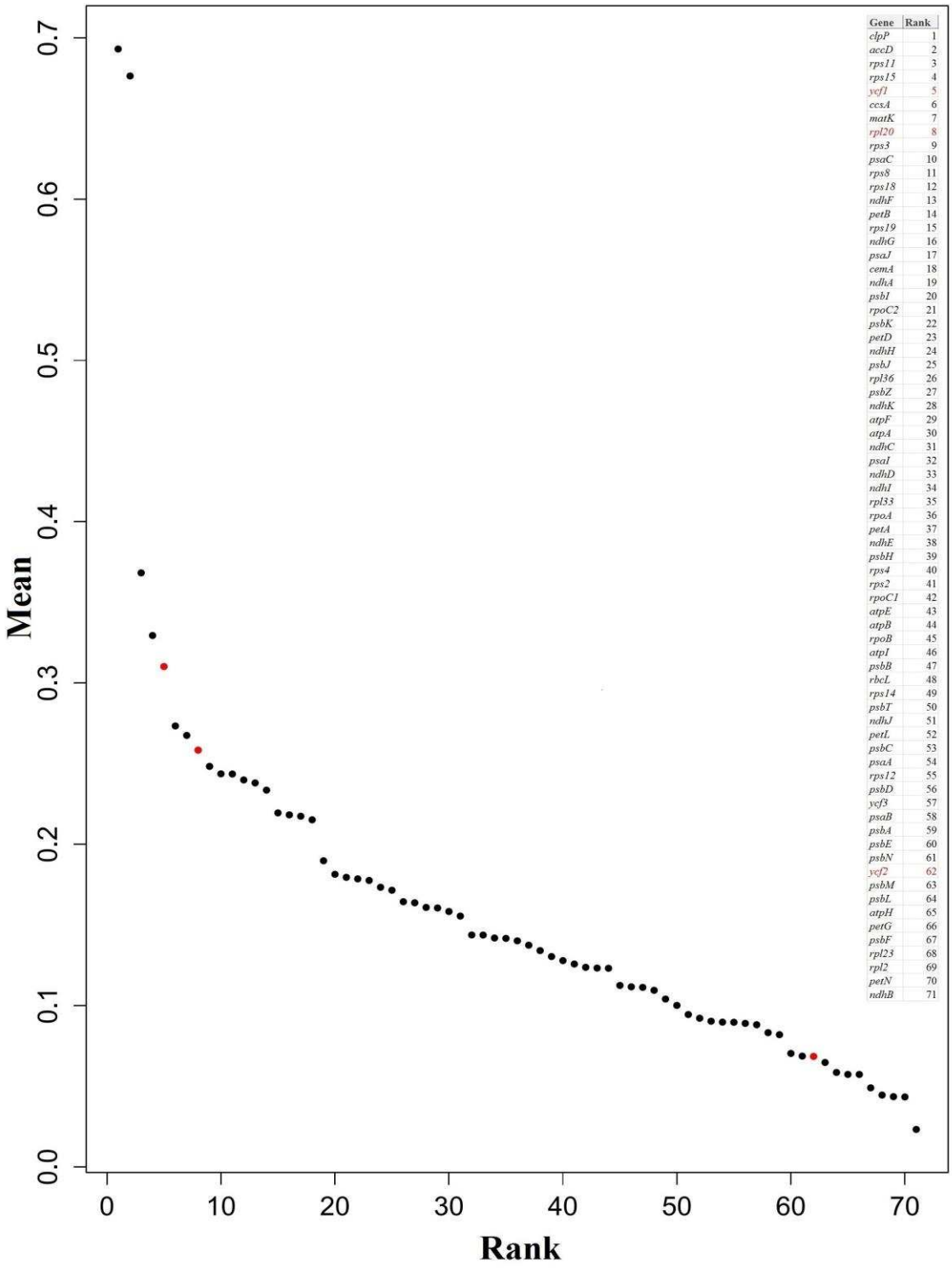

Figure 6. Mean of pairwise distance based on the nucleotide sequences between *P. cincinnata* plastome and all species present in Table 1, which was analyzed for 71 plastid protein-coding genes. The genes were ranked according with their mean values, in decreasing order. Putative pseudogenes previously identified are red colored.

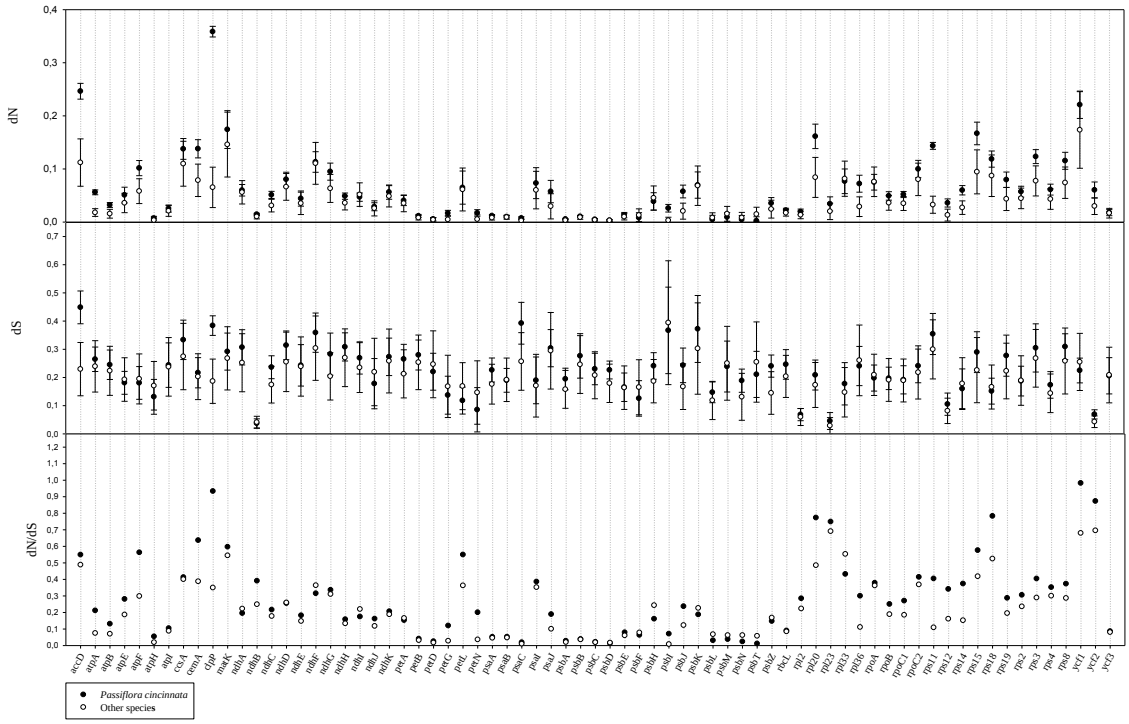

Figure 7. Synonymous (dS) and nonsynonymous (dN) substitution rates, and dN/dS values of 71 common plastid protein-coding genes, in alphabetical order. *P. cincinnata* is represented by filled circles and the mean of the others species (see table 1) is represented by open circles.

4. Discussion

4.1. *P. cincinnata* plastome shows specific high divergence of genes and unprecedented gene losses in photosynthetic higher plants

The *P. cincinnata* plastome encodes 30 tRNA and 4 rRNA genes like most flowering plants (Shinozaki et al., 1986; Sugiura and Wakasugi, 1989; Bock, 2007). Comparing the protein-coding genes, this genome shows an uncommon number of genes losses (or pseudogenization): the *infA*, *rps16* and *rpl22* genes were lost and the *rps7, rpl20, ycf1* and *ycf2* genes were identified as putative pseudogenes in this genome (Figure 1 and Table 3). The *infA* encodes the translation initiation factor 1 and is an essential gene in *Escherichia coli* (Cummings and Hershey, 1994). This gene was reported as an unusually unstable angiosperm plastid gene, given that it has been suffered different events of horizontal gene transfer and has been transferred to the nucleus in several species. This gene was lost from plastome in all Malpighiales and almost all rosid species investigated up to now (Millen et al., 2001). Furthermore, nuclear *infA* copies, containing a plastid transit peptide, were also already identified in different rosid species (Millen et al., 2001). Thus, it is strongly likely that a similar mechanism of i*nfA* transfer to the nucleus has occurred in *P. cincinnata* and other Malpighiales species.

As in *P. cincinnata* plastome, the *rps16* gene was lost almost in all Malpighiales species investigated so far, except in some Euphorbiaceae species. In this family the *rps16* was lost in the plastome of *Jatropha curcas* (Asif et al., 2010), but it remained putatively functional in the other species as rubber tree, castor bean and cassava (Daniell et al., 2008; Rivarola et al., 2011; Tangphatsornruang et al., 2011). The presence of intact *rps16* gene in some Euphorbiaceae species suggests at least three independent losses of this gene across Malpighiales. The *rps16* gene was lost also in several legumes (Doyle et al., 1995; Schwarz et al., 2015) and also in several Brassicaceae species (Roy et al., 2010). Despite this high rate of parallel losses, this gene encodes the ribosomal protein (Rps16) of the small subunit (30S) of plastid ribosome and the gene has been demonstrated to be essential for cellular viability even under heterotrophic conditions (Fleischmann et al., 2011). It has been showed that the loss of plastid *rps16* gene in *Medicago truncatula* and *Populus alba* was compensated by a nuclear-encoded *rps16* copy of mitochondrial origin that can also target the plastids (Ueda et al., 2008). Given the history of *rps16* in Malpighiales, it is likely that other

species in this family, that lost the *rps16* gene from plastome, including *P. cincinnata*, also have this nuclear-encoded *rps16* copy replacing the function of plastid gene.

The *rpl22* gene encodes the ribosomal protein L22 (Rpl22), which assemblies in the large ribosomal subunit of plastid ribosome. It is conserved in plastid genomes of most angiosperms and it is essential for plastid translation, cell survival and plant development (Fleischmann et al., 2011). The loss of the plastid *rpl22* gene was reported in three species of *Passiflora* (*P. biflora, P. quadrangularis* and *P. cirrhiflora*) by Jansen et al. (2011). Here, the loss of this gene was also reported in *P. cincinnata*, suggesting that this is a shared characteristic of *Passiflora* plastomes. Other studies indicated that this gene is also absent (or is a pseudogene) in Fabaceae (Doyle et al., 1995; Martin et al., 2014; Sherman-Broyles et al., 2014) and Fagaceae plastomes (*Castanea* and *Quercus*; Jansen et al., 2011), and that the *rpl22* was independently transferred to the nucleus in each of this families (Jansen et al., 2011). Additionally, it is likely that a third independent event of gene transfer to the nucleus occurred in Passifloraceae with acquisition of a plastid transit peptide sequence, which allows its returning as a functional protein into the plastids, compensating the loss of the *rpl22* gene in the plastome.

The *rps7* gene encodes a protein (S7), normally with 155 amino acids residues, which is part of the plastid ribosome 30S subunit (Shinozaki et al., 1986; Wakasugi et al., 1998). The S7 homolog protein in *E. coli* initiates the assembly process of 30S ribosome subunit (Nowotny and Nierhau, 1988) and the deletion of the *rps7* gene in this organism shows that it is essential for cell viability (Shoji et al., 2011). However its detailed function remains to be elucidated in plants (Fleischmann et al., 2011). The *rps7* gene is highly conserved in plastid genomes of angiosperms and other plant groups. Here, we suggest that the *rps7* is a pseudogene in *P. cincinnata* plastome because it contains a premature stop codon at amino acid position 62 (Figure 8). Apparently, the gene also degenerated into pseudogene in the *Salix interior* (Salicaceae) due to a premature stop codon and several deletions in its coding-sequence (manually checked; Figure 8). In other hand, in *Populus*, other genus of the Salicaceae family, the plastid *rps7* gene encodes a full and functional Rps7 protein (Figure 8), indicating that the losses of this gene in *Passiflora* and *Salix* were two independent evolutionary events. The Rps7 protein assemblies very early directly on ribosomal RNA and interacts with other mid-binding ribosomal proteins being highly conserved in the plastid ribosome (Napper and Culver, 2015). Beyond its structural function, it is involved in translation

initiation as a general or specific factor (Fargo et al., 2001). Due to its functional characteristics related to plastid ribosome and translation initiation it is very likely that *rps*7 gene was transferred to the nucleus as a functional gene in *Passiflora* and *Salix* species, representing an unprecedented gene transfer to the nucleus in angiosperms.

The *rpl20* encodes the protein Rpl20, which is part of the 50S subunit of plastid ribosome. This protein is required for early steps of ribosome assembly (Nowotny and Nierhau, 1980) and it is essential for cell viability and plant development (Rogalski et al., 2008). Curiously, in *P. cincinnata* plastome the *rpl20* gene was identified as pseudogene due to an internal stop codon and a 5 pb insertion, which shifted the reading-frame (producing additional premature stop codons; Figure 9). The loss of this gene was also suggested by Jansen et al. (2007) in *P. biflora* and except for these *Passiflora* species, loss of the *rpl20* was not yet reported in plastomes of photosynthetic plants. It was only found to be absent from the plastid genomes of the parasitic protozoans *Eimeria tenella* (Cai et al., 2003), *Theileria parva* (Gardner et al., 2005), and *Toxoplasma gondii* (Wilson, 2002). Thus, based in the essential function and in the conservation of *rpl20* gene in plants, it is very possible that, in *Passiflora* genus, this gene was transferred to the nucleus and a functional Rpl20 protein is then imported into the plastids.

The *ycf1* and *ycf2* genes are giant plastid open reading frame encoding protein of approximately 1800 and 2280 amino acids, respectively (Downie et al., 1994; Bock, 2007; Kikuchi et al., 2013). They are essential for cell viability given that *ycf1* and *ycf2* mutant plants were not obtained even after several attempts (Drescher et al., 2000). The *ycf1* gene was recently characterized and it encodes the subunit Tic214 of the TIC complex, a translocon protein located at the inner membrane of plastids (Kikuchi et al., 2013), while the *ycf2* gene still has an unknown function. In *P. cincinnata* plastome, the *ycf1* gene was identified as pseudogene due to the presence of several internal stop codons in the coding sequence (5 of them in the 101 first amino acids; Figure 10) and the loss of a larger portion, with approximately 2 kb, of the coding sequence in comparison with rosid species. On the other hand, the *ycf2* gene is also a putative pseudogene in this species due to the lack of the start codon (Figure 11a), the presence of premature stop codons (5 of them between amino acid 769 and 799 position; Figure 11b) and the lack of a larger portion of its coding region, containing about 3.2 kb, in comparison with other rosid species. Previous studies had already suggested the putative loss of *ycf2* and *ycf1* genes in *Passfilora helleri* (Downie et al., 1994) and *P.*

biflora (Jansen et al., 2007), suggesting that the loss of these genes is a common trait in the *Passiflora* genus. The *ycf1* and *ycf2* genes are also absent (or are pseudogenes) in three distantly related taxa: Poaceae (Guisinger et al., 2010; Vries et al., 2015), *Erodium* genus (Geraniaceae; Guisinger et al., 2011) and in cranberry (*Vaccinum macrocarpon*, Ericaceae) (Fajardo et al., 2013). The characterization of *ycf1* and *ycf2* genes in *P. cincinnata* demonstrated that at least four independent events of *ycf1* and *ycf2* losses occurred in photosynthetic angiosperms. Given the essential function of these two genes in eudicotyledon species, it is very likely that *ycf1* and *ycf2* genes were transfer to the nucleus in these species and a functional copy is present in the nucleus compensating the loss of these essential genes from their plastomes. Alternatively, and less likely, other compensatory mechanisms may have emerged in these lineages, allowing them to tolerate the lack of Ycf1 and Ycf2 proteins as suggested in plastids of grasses, in which the existence of an alternative TIC transport system would compensate the lack of the *ycf1* gene (Nakai, 2015).

In addition to the putative pseudogenes showed here, we found other genes in *P. cincinnata* plastome with unusual structures and features. The first case is the *ndhB* gene encoding the B subunit of NAD(P)H dehydrogenase, which is involved in the process of chlororespiration and cyclic electron transfer in chloroplast (Bock, 2007). In the *P. cincinnata* plastome, the *ndhB* gene has an unusual start codon (AGG), which is not conserved in other Malpighiales and rosid species (Figure 12). It is unlikely that AGG can be used as start codon. In plastids, the canonical start codon is ATG, whereas other start codons were also reported in the literature as GTG, TTG and ACG (Hirose and Sugiura, 1997; Sugiura et al., 1998; Hirose et al., 1999). The latest alternative start codon can be restored post-transcriptionally to AUG by RNA editing. The RNA editing can become a nonfunctional gene at level of DNA into a complete functional transcript by restoring initiation, termination and/or conserved codons (Tsudzuki et al., 2001; Takenaka et al., 2013). Therefore, to analyze this possibility, the gene was sequenced again using DNA sequence as control and cDNA synthesized from *ndhB* mRNA from total leaf RNA. The sequences were completely identical, excluding the creation of a new start codon via RNA editing. The pseudogenization of *ndhB* gene is also unlikely because the sequence of this gene is highly conserved between *P. cincinnata* and other Malpighiales and related species (see alignment in Figure 12), showing the lowest value of pairwise distance between the *P. cincinnata* plastid genes (Figure 6) and also it shows low values of dS (0,037) and dN (0,015) (Figure 7). The most probably

hypothesis, in the *P. cincinnata* plastome, is that the *ndhB* gene starts in the next ATG in frame, approximately 50 pb downstream of the AGG codon (see Figure 12, blue arrow), and it is still functional. In some species of Orchidaceae family, the *matK* gene was previously described as pseudogene due to the frame shift mutations, but an alternative start codon in this gene can alleviate the frame shift mutations restoring the ORF and maintain this gene functional (Barthet et al., 2015). Then, we propose here that a downstream ATG became the new start codon in the *ndhB* transcript of *P. cincinnata* and it alleviates the mutation in the original start codon maintaining this gene functional.

The *rps18* gene encodes the S18 plastid ribosomal protein, which is integral part of the small ribosomal subunit (30S) (Harris et al., 1994; Sugiura et al., 1998). This protein is essential for functional ribosomal subunit assembly, and consequently, for plastid translation and cell viability in angiosperms (Rogalski et al., 2006). In *P. cincinnata*, the Rps18 protein contains an C-terminal extension, which increases 22-38 amino acids in the protein structure if compared with other species from Malpighiales, whereas the N-terminal sequence is highly conserved in comparison with most angiosperms (Figure 13). The extended C-terminal sequence of Rps18 in *P. cincinnata* shows more similarity, in terms of length, to species belonging to other orders as *Castanea mollisima* (Fagales) and the monocot *Secale cereale* (Poales) than to species of Malpighiales (Figure 13; Weglöhner et al., 1995; Jansen et al., 2011). Additionally, *S. cereale* Rps18 protein acquired during the evolution additional repeated SKQPFRK motifs in the N-terminal sequence. It is known that the *rps18* gene has acquired several structural modifications in the N-terminus and C-terminus during the evolution of plastome, which are observed in different species, families, order and plant groups (Weglöhner and Subramanian, 1991; Weglöhner et al., 1995). The interaction between different genomes present in the current plant cell and the exchange of genetic material between them can explain the changes found in the plastid ribosome as presence of plastid-specific ribosomal proteins and increase of mass essentially due to additional N- and/or C-terminal in the sequence of ribosomal proteins, if compared with their *E. coli* homologues (Yamaguchi and Subramanian, 2000, 2003; Yamaguchi et al., 2000; Tiller et al., 2012). The *rps18* gene is highly conserved in the plastome of higher plants sequenced until the data and it was only transferred to the nucleus from plastome of parasite organisms including the colorless alga *Euglena longa* and the parasitic protozoans *E. tenella* (Cai et al., 2003), *T. parva* (Gardner et al., 2005) and *T. gondii*

(Wilson, 2002). To eliminate the possibility of stop codon restoration by RNA editing, the DNA sequence and the cDNA synthesized from *rps18* mRNA were sequenced again, which revealed sequences completely identical. As observed in other species, it is likely that this variation in the C-terminal portion does not interfere negatively in the functions of Rps18 in plastid ribosome. Analyzing the DNA sequence, we can conclude that the larger C-terminal of Rps18 in *P. cincinnata* may have originated from the loss of an upstream stop codon in the coding sequence and the *rps18* gene encodes a functional protein in *P. cincinnata* plastome.

Unlike in most angiosperms, the gene *clpP* lost both introns in *P. cincinnata* plastome. This gene encodes the catalytic subunit (ClpP) of the plastid ATP-dependent protease Clp and it was demonstrated *in vivo* that this gene is essential for cell viability (Shikanai et al., 2001; Kuroda and Maliga, 2003). In addition to intron losses, the *clpP* gene is the most divergent gene found in *P. cincinnata* plastome, showing higher values of pairwise distance and dN values in comparison with other plastid genes and *clpP* genes from other rosid species (Figures 6 and 7). Moreover, this gene showed a high dN/dS value (0.93) in *P. cincinnata*, suggesting that this gene can be subject to relaxed purifying selection in this species, which is a strong evidence for a pseudogenization (Wu et al., 1995; Go et al., 2005). Both introns of *clpP* were also lost in several species of different families as Caryophyllaceae (Sloan et al., 2012b, 2014), in two species of Oleaceae (Lee et al., 2007), in grasses and in the *Oenothera* genus (Jansen et al., 2007; Erixon and Oxelman, 2008). Similarly, the intron 1 was lost in several legume species (Jansen et al., 2008; Dugas et al., 2015). Also as observed in *P. cincinnata*, various of these lineages that lost the introns of *clpP* gene also show an elevated substitution rate in this gene (Erixon and Oxelman, 2008; Sloan et al., 2012b, 2014; Dugas et al., 2015). Similarly, the *clpP* is also a putative pseudogene in *Acacia ligulata* (Williams et al., 2015). In comparison as what was observed in *A. ligulata*, several evidences suggest the non-functionality of *clpP* in *P. cincinnata*, notwithstanding *P. cincinnata clpP* lacks frameshift mutations and/or premature stop codons in the coding sequence. If the *clpP* of *P. cincinnata* is indeed a pseudogene, some compensatory mechanism must have emerged in this species allowing it tolerate the loss of this essential plastid gene (Shikanai et al., 2001; Kuroda and Maliga, 2003), such as a possible gene transfer to the nucleus and the replacing of the plastid gene by a nuclear functional copy. On the other hand, the amino acid residues of the ClpP active sites and oligomer interfaces remained

conserved in this species (Figure 14), suggesting that this gene may still be functional in *P. cincinnata* plastome.

As the *clpP* gene, the *accD* gene was also reported as highly divergent in *P. cincinnata* (Figures 6 and 7). This gene encodes the β-carboxyl transferase subunit (AccD) of a eubacteria-like multisubunit acetyl-CoA carboxylase, which is composed of other three nuclear-encoded subunits, AccA, AccB and AccC (Ohlrogge and Browse, 1995; Sasaki and Nagano, 2004; Rogalski e Carrer, 2011). The eubacteria-like acetyl-CoA carboxylase (ACCase) occurs in the plastids and catalyzes the formation of malonyl-CoA from acetyl-CoA during *de novo* fatty acid biosynthesis (Ohlrogge and Browse, 1995; Rogalski and Carrer, 2011). ACCase is intended to be the mainly regulatory enzyme of fatty acid synthesis (Sasaki and Nagano, 2004; Rogalski and Carrer, 2011) and it is essential for cell survival and plant development (Kode et al., 2005). Despite its essential function, the *accD* gene was lost from plastome of several angiosperms lineages, including the grasses (Hiratsuka et al., 1989; Maier et al., 1995; Katayama and Ogihara, 1996), Acoraceae (Goremykin et al., 2005); Campanulaceae (Cosner et al., 1997), Geraniaceae (Chumley et al., 2006), Lobeliaceae (Knox and Palmer, 1999), Oleaceae (Lee et al., 2007) and in *Trifolium* (Fabaceae) (Doyle et al., 1995; Cai et al., 2008; Magee et al., 2010). In the grasses, the plastid *accD* gene has been replaced by a nuclear encoded homomeric ACCase of eukaryotic origin that acquired a plastid transit peptide encoding-sequence (Konishi et al., 1996). In turn, a chimeric nuclear *accD* gene of chloroplast origin was detected in *Trachelium caerulem* (Rousseau-Guetin et al., 2013) and also in species from *Trifolium* genus (Magee et al., 2010), showing a functional replacing the plastid *accD* gene that was lost from plastomes in these lineages. Due to the high divergent and substitution rate (Figures 6 and 7), low conservation of its initial portion and the presence of an 800 pb insertion in the middle of its sequence (principally constituted by tandem repeats, see table 6) with no hits in BLAST search, it is possible that the *accD* is a pseudogene in *P. cincinnata* plastome. If the *accD* gene is a pseudogene in *P. cincinnata*, a nuclear-encoded *accD* with a plastid transit peptide, as seen in *T. caerulum* and *Trifolium* species, should exist to replace the essential function. However, the AccD protein of *P. cincinnata* still retains a well-conserved active domains with conserved residues for acetyl-CoA binding site, CoA-carboxylation catalytic site, and the carboxybiotin-binding site motifs (Figure 15). In BLASTp search, the AccD protein of *P. cincinnata* shows 78-79% of identity and 56-60% of coverage relative to the protein of other Malpighiales species.

Furthermore, the insertions and high substitution rates reported here did not caused frameshifts mutations and/or premature stop codons that interrupted the ORF. Thus, similar to the *clpP* case, the possibility of *accD* gene is still functional in *P. cincinnata* plastome cannot be excluded.

The *P. cincinnata* plastome also show the loss of the intron of *atpF* gene, which encodes the ATP synthase β subunit (Bock, 2007). The loss of this intron was already suggested by Daniell et al. (2008) to the family Passifloraceae and also to other Malpighiales families, such as Euphorbiaceae, Picrodendraceae, Lophopyxidaceae and Phyllanthaceae. These authors also suggested an association between C-to-T substitutions (i.e loss of a RNA-edition site) at nucleotide position 92 and the loss of the *atpF* intron, suggesting that recombination between an edited mRNA and the *atpF* gene may be a possible mechanism for the intron loss. Here, the loss of *atpF* intron in *P. cincinnata* plastome in fact coincided with the presence of a T at nucleotide position 92, confirming the previous hypothesis.

On the other hand, *P. cincinnata* plastome retains the intron of *rpoC1* gene, which encodes the β' subunit of plastid RNA polymerase (Serino and Maliga, 1998), in agreement with the expected to species of subgenus *Passiflora*, and contrasting with the loss of this intron in species belonging to subgenus *Decaloba* (Hansen et al., 2006). In the same way with other gene coding for a subunit of plastid RNA polymerase, *rpoA* gene, which encodes the plastid RNA polymerase β subunit (Serino and Maliga, 1998). The *rpoA* gene is not essential for cell viability under heterotrophic conditions, but the deletion of this gene and other *rpo* genes impair the photosynthetic capacity in mutant plants (Serino and Maliga, 1998), which is consistent with the role of the plastid RNA polymerase in the transcription of photosynthesis-related genes. Interestingly, the *P. cincinnata rpoA* is a well-conserved gene with similar substitution rates found in other rosids (Figure 7) and its amino acid translated sequence showed 93% of identity and 100 % of coverage relative to other Malpighiales species (in BLASTp search), as observed in several *Passiflora* species, such as *P. ciliata*, *P. cirrhiflora* and *P. quadrangularis* (Brazier et al., 2016). However, the *rpoA* gene in *P. biflora*, a species belonging to Decaloba subgenus, shows high sequence divergence, retaining only 30% of amino acid sequence identity in comparison with the *rpoA* gene from other Malpighiales species (Jansen et al., 2007; Blazier et al., 2016).

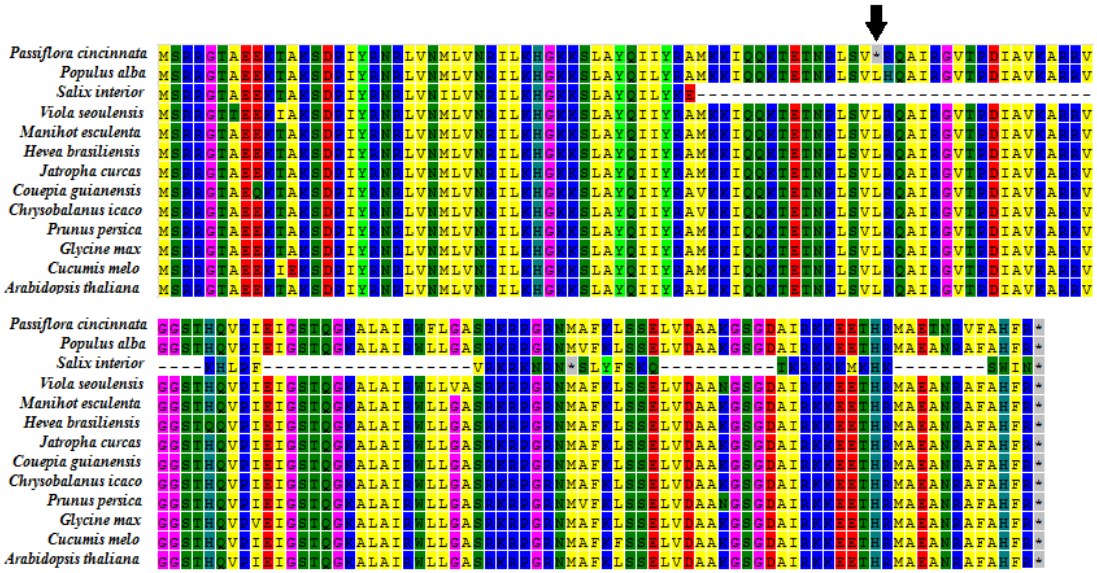

Figure 8: MUSCLE alignment of *rps7* translated amino acid sequence from 9 species of Malpighiales, including *P. cincinnata*, and other 4 rosids species. * indicates stop codon. Black arrow indicates premature stop codon in *rps7* of *P. cincinnata*. Different background colors indicate amino acids with different biochemical properties.

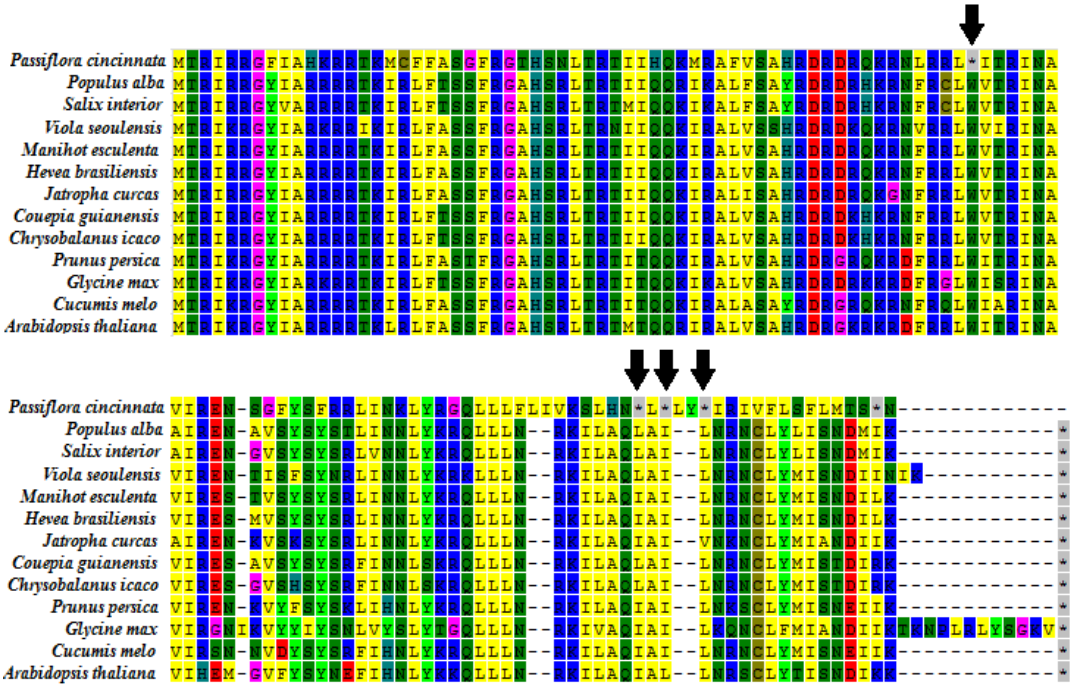

Figure 9: MUSCLE alignment of *rpl20* translated amino acid sequence from 9 species of Malpighiales, including *P. cincinnata*, and other 4 rosids species. *indicates stop codon. Black arrow indicates premature stop codons in *rpl20* of *P. cincinnata*. Different background colors indicate amino acids with different biochemical properties.

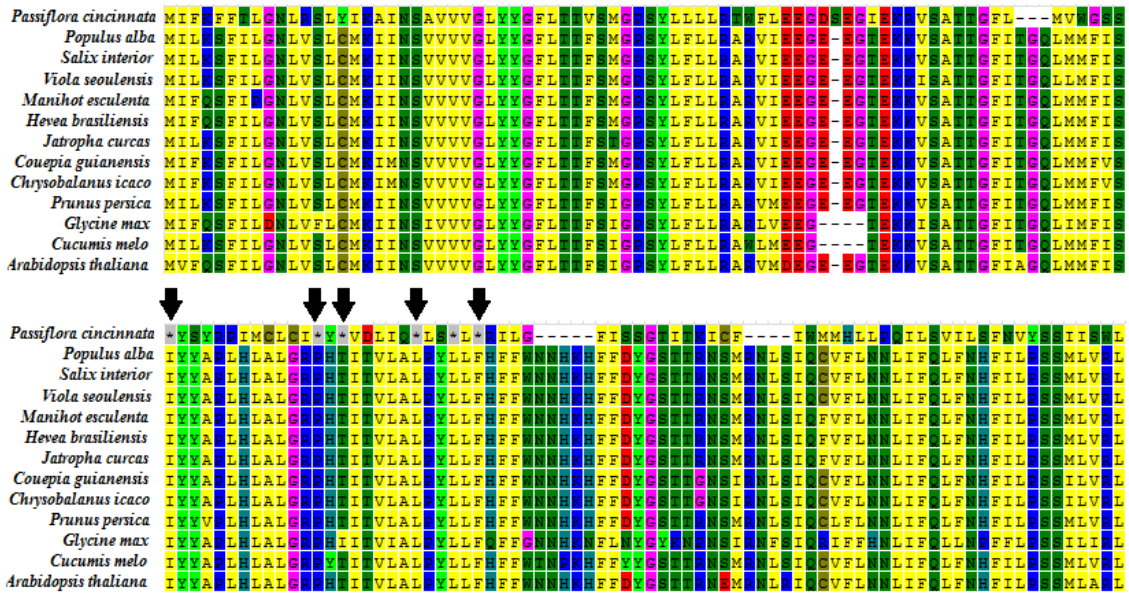

Figure 10. MUSCLE alignment of partial translated amino acid sequence of *ycf1* (first 150 amino acids) from 9 species of Malpighiales, including *P. cincinnata*, and other 4 rosids species. * indicates stop codon. Black arrow indicates premature stop codons in *ycf1* of *P. cincinnata*. Different background colors indicate amino acids with different biochemical properties.

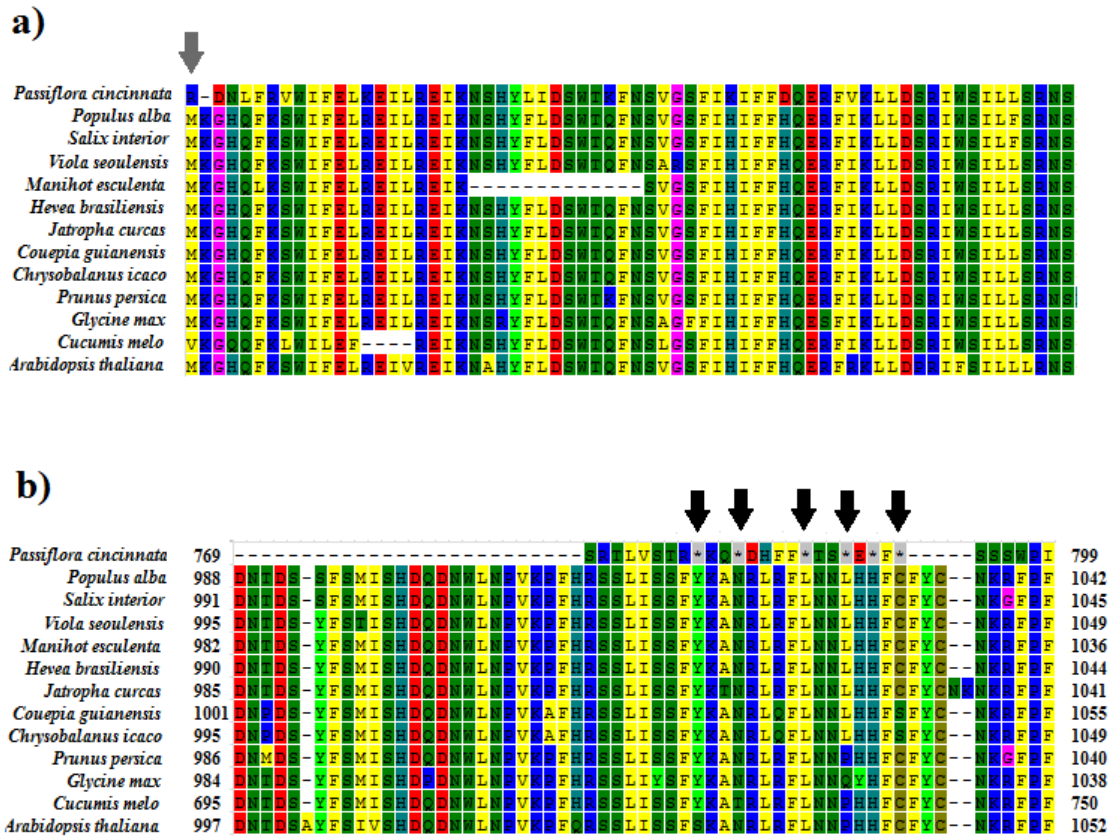

Figure 11. MUSCLE alignment of partial translated amino acid sequences of *ycf2* from 9 species of Malpighiales, including *P. cincinnata*, and other 4 rosids species. **a.** Alignment of the first 66 amino acid sequence of *ycf2*. Gray arrow indicates the lack of a methionine in the start codon position of *P. cincinnata*. **b.** Alingment of the partial middle sequence of *ycf2* amino acids. * indicates stop codons. Black arrow indicates premature stop codons in *ycf2* of *P. cincinnata*. Different background colors indicate amino acids with different biochemical properties.

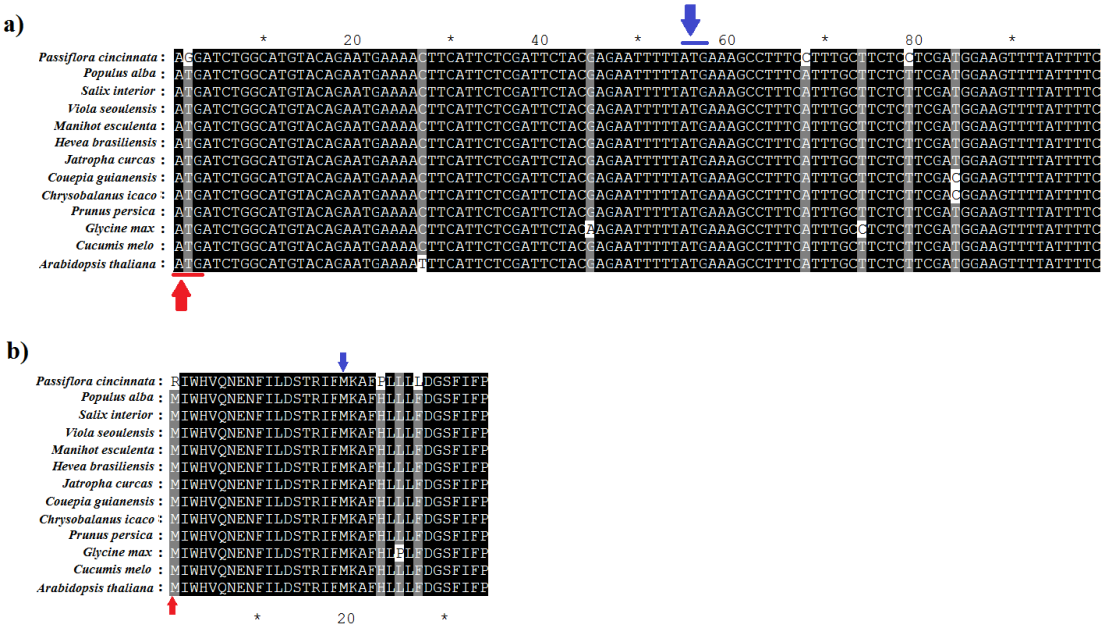

Figure 12. ClustalW (Thompson et al., 1994) alignment of the initial sequences of the *ndhB* gene and subsequent translation products from 9 species of Malpighiales, including *P. cincinnata*, and other 4 rosids species. **a.** Nucleotide alignment showing the consensus initiation codon (red arrow) and the putative new start codon (blue arrow) in *P. cincinnata*. **b.** Translated amino acids of initial sequences of *ndhB* gene, showing the consensus initiation amino acid (red arrow) and the putative new initiation amino acid in *P. cincinnata* (blue arrow). Black shadowing indicates conserved sequence among all taxa in the alignment, gray indicates conserved sequence among all except to one-three taxa, and white indicates nonconserved nucleotide or amino acid.

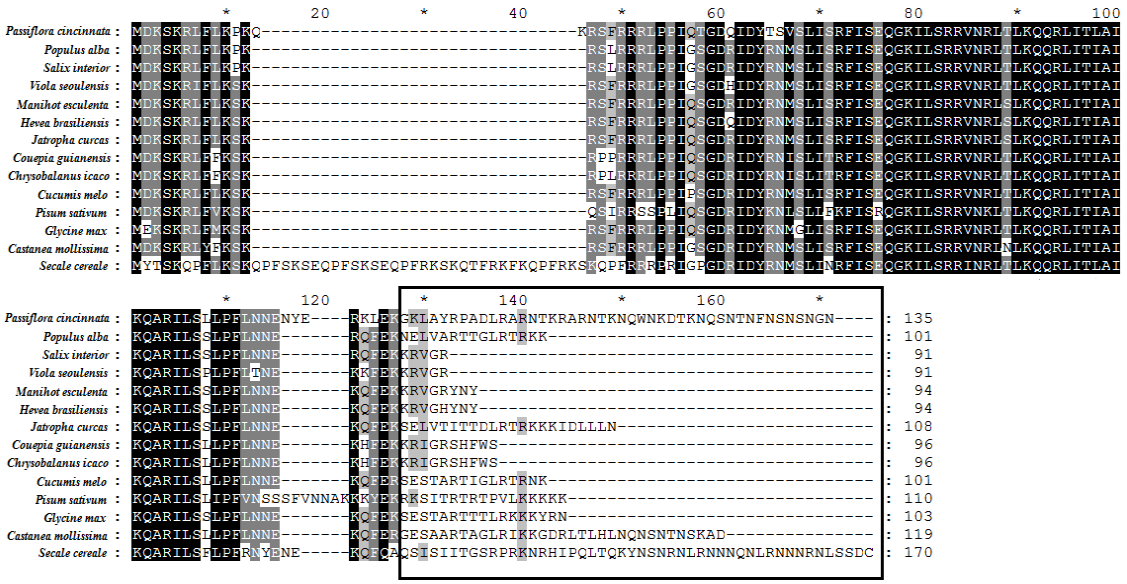

Figure 13. ClustalW alignment of the translated amino acid sequence of the *rps18* gene from nine Malpighiales species (*P.* cincinnata, *P. alba*, *S. interior*, *V. seoulensis*, *M. esculenta*, *H. brasiliensis*, *J. curcas*, *C. guianensis* and *C. icaco*), one Cucurbitales (*C. melo*), two Fabales (*P. sativum* and *G. max*), one Fagales (*C. mollissima*) and one monocot (*S. cereale*). The final portion of the gene showing variable sizes and amino acids composition to the different taxa is highlighted (open box). Black shadowing indicates conserved sequence among all taxa in the alignment, darker gray indicates conserved sequence among all except to one-three taxa, lighter gray indicates less conserved sequence, and white indicates non conserved amino acid.

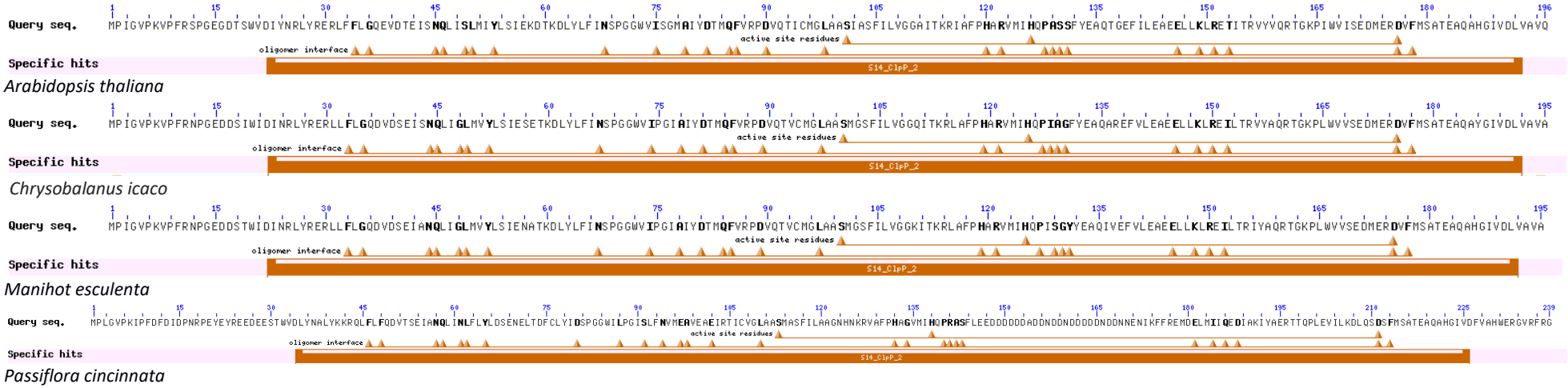

Figure 14. The ClpP amino acid sequence was analyzed using Conserved Domains (NCBI). All the ClpP proteins, including that of *P. cincinnata*, belong to the family S14_ClpP_2. The residues of the active site and of the oligomer interface are highlighted.

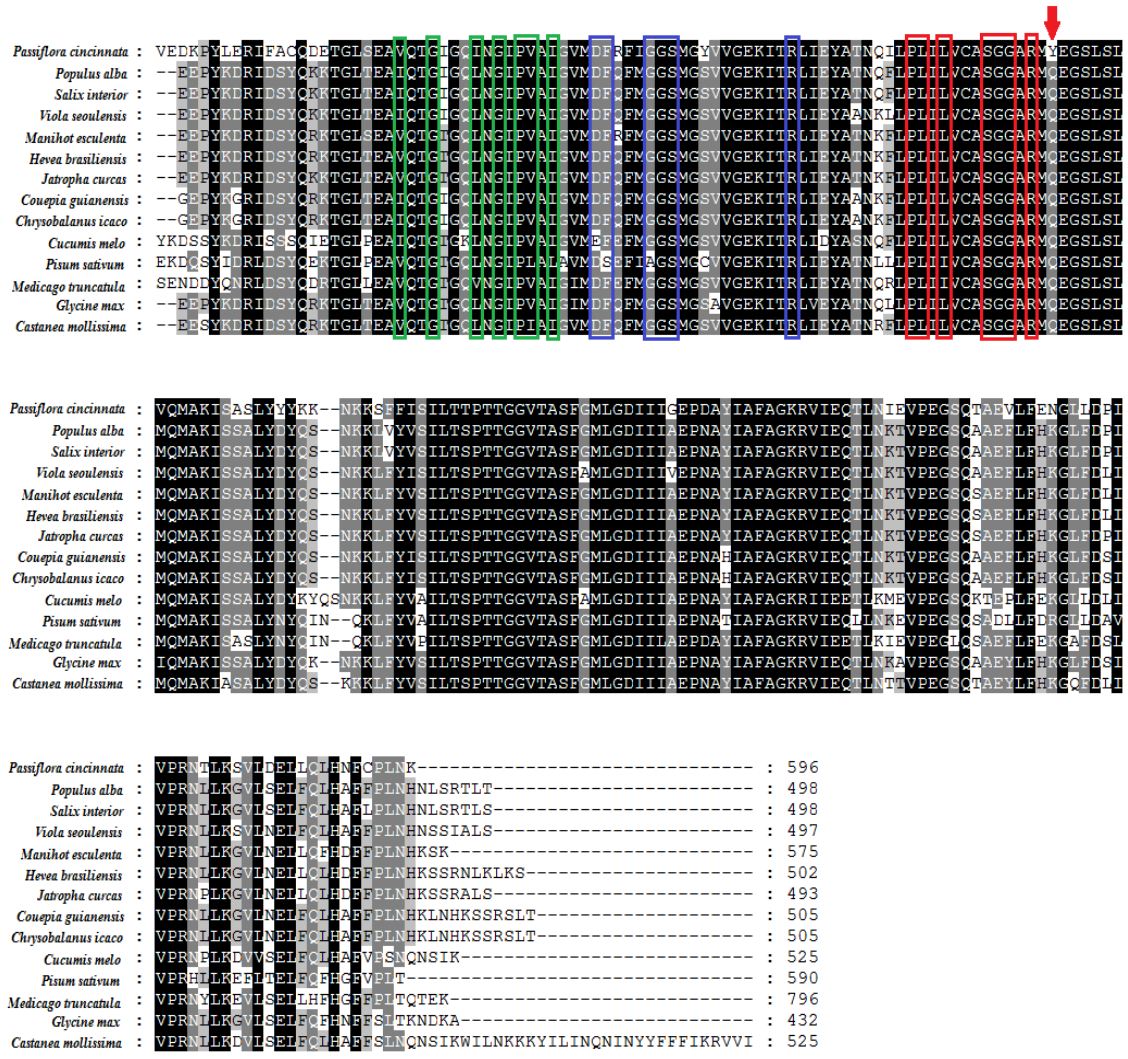

Figure 15. ClustalW alignment of the *accD* C-terminus region sequence in 9 Malpighiales species (including *P. cincinnata*) and other rosid species (*C. melo*, *P.* sativum, *M.* truncatula f. *tricycla*, *G. max* and *C. mollissima*), showing the putative acetyl-CoA binding site (in green), the CoA-carboxylation catalytic site (in blue) and the carboxybiotin-binding site (in red) (Lee et al., 2004; Gurdon and Maliga, 2014). Red arrow indicates a carboxybiotin-binding site conserved among all taxa, except to *P. cincinnata*. Black shadowing indicates conserved sequence among all taxa in the alignment, darker gray indicates conserved sequence among all except to one-three taxa, lighter gray indicates less conserved sequence, and white indicates lack of conservation.

4.2. Multiple inversions in *P. cincinnata* plastome

Gene order in the plastome is typically conserved in majority of land plants (Palmer 1991; Jansen and Ruhlman, 2012). However, rearrangements in the plastome have been reported in several plants groups, including Asteraceae, Campanulaceae, Fabaceae, Geraniaceae, Lobeliaceae, Onagraceae, Pinaceae, Poaceae, Ranunculaceae, Euphorbiaceae and Podocarpaceae (Jansen and Palmer, 1987; Palmer et al., 1987; Hiratsuka et al., 1989; Hachtel et al., 1991; Lindholm and Gustafsson, 1991; Doyle et al., 1992; Hoot and Palmer, 1994; Cosner et al., 1997, 2004; Strauss et al., 1998; Knox and Palmer, 1999; Saski et al., 2005; Chumley et al., 2006; Guo et al., 2007; Lee et al., 2007; Haberle et al., 2008; Tangphatsornruang et al., 2011; Vieira et al., 2014a). The mechanisms implicated in these rearrangements that change gene order and content in plastome, include: inversion, gene duplication, gene/intron loss, insertion/deletion, and IR expansion/contraction (Palmer, 1991; Raubeson and Jansen, 2005). Here, it was reported for the first time in a Passifloraceae species, the presence of multiples rearrangements in the plastome LSC region, which generate significant changes in its gene order**.**

One of these rearrangements was characterized to be a simple inversion of 1.6-1.7 kb in the start of LSC region that changed the order and direction of the coding region of *trnH*-GUG and *psbA* genes (Figure 2; Figure 3A-F, number1). Thus, the *psbA* gene in *P. cincinnata* is considered the first gene in the LSC region, while this position is occupied for *trnH*-GUG in the majority of angiosperms (Figure 1 and 2). Furthermore, it was here reported other rearrangement in the *P. cincinnata* plastome, corresponding to a large inversion of approximately 47 kb in the LSC region, extending from *trnC-GCA* gene till *clpP* gene, whereas a sequence of 3 kb approximately in the middle of this inversion, from *trnV*-UAC to *atpB* genes, remains in the same direction as observed in the other species analyzed here (Figure 2; Figure 3A-F, number 2). It is likely that this rearrangement was caused by multiple inversion events in this region. A possible scenario is that the inversion of this 3 kb sequence occurred previously and thus a new large inversion of 47 kb, in this region, returned the sequence of 3kb to the same direction observed in other species. Alternatively, the inversion of 47 kb may have occurred first and a posterior inversion returned a 3 kb sequence to the same direction as observed in other species. However, scenarios more complex involving a larger series of recombinations yielding inversions and translocations cannot be exclude. The plastome

sequences of other *Passiflora* species could clarify the mechanism, order and phylogenetic origin of these inversion events.

Inversions in plastid genome are generally associated with the presence of short inverted repeats (IRs) sequences flanking the inversion borders or tRNA genes (Hiratsuka et al.; 1989; Hupfer et al. 2000; Pombert et al., 2005, 2006; Vieira et al., 2016b). Here, significant IRs flanking the inversion borders were not found in *P. cincinnata*. In *H. brasiliensis* (the only Malpighiales species with a larger inversion reported prior this work) significant repeats flanking the 30 kb inversion were also not found (Tangphatsornruang et al., 2011). However, it is still possible that the presence of IRs in the borders is the responsible mechanism for inversions reported here to *P. cincinnata*, and subsequent mutations and/or inversions in these regions could have obscured or eliminated these repeated sequences.

Jansen et al. (2007) detected a positive correlation between increased substitution rates, number of gene and intron losses and gene order changes. The *P. cincinnata* plastome strongly corroborate with this relation because it showed a large number of gene and intron losses, gene order changes and elevated substitution rates to several genes (Figure 7). A possible mechanism that explains this correlation involves homologs to the eubacterial *recA* gene, which is involved in homologous recombination and DNA repair in this organism (Guisinger et al., 2008). Homologs to this gene and other nuclear-encoded genes involved in DNA repair were already found in plants and some are target to plastid (Cerutti et al., 1992; Lin et al., 2006; Saotome et al., 2006).

4.3. Genome size reduction *versus* IR expansion

Two copies of larger inverted repeats (~25 kb) are present in most land plant plastomes and they are designated IR_A and IR_B. The sequence found in an IR is completely duplicated in the other IR (Jansen and Ruhlman, 2012). Expansion and contraction events involving few hundreds of base pairs are common to happen at the IR boundaries. Normally, in most angiosperms the edge of boundaries includes the *ycf1* gene at the IR/SSC junction and the *rps19* or *rpl22* genes at the IR/LSC junction (Goulding et al., 1996; Zhu et al., 2016). In the Figure 4 is shown different events of expansion of both boundaries in species of Malpighiales order. Comparing with *Arabidopsis thaliana* plastome, events of IR expansion in the IR_A/LSC junction were observed in *Manihot esculenta*, *Licana alba*, *Populus alba* and *Passiflora cincinnata*, which was described in this study. In the last two species, the expansion includes a full

rps19 in the IRs. In the genus *Salix*, the *rps19* gene is also completely duplicated into the IR. Expansion of IRs in the IR_B/SSC boundary was also observed predominantly in *M. esculenta*, *P. alba* and *P. cincinnata.* The later showed the largest IR expansion in the IR_B/SSC junction, including the full pseudogene *ycf1* (~3.5kb), while in the other species in analysis only 1-1.7kb of *ycf1* is duplicated in the IR (Figure 3). These results suggest an IR expansion tendency, in both of the boundaries of Salicaceae and Passifloraceae species. The IR expansion may have probably originated in the common ancestral of these families or could still be independent evolutionary events.

Curiously, despite of *P. cincinnata* showed the largest expansion events in our IR boundary analysis (Figure 4), the IRs in these species are smaller in total size than those other Malpighiales species (Table 2). The reduction of size could be attributed to the loss of about 3,2kb of the *ycf2* pseudogene. Moreover, *P. cincinnata* contains a quite reduced SSC region, in comparison with other Malpighiales (Table 2). This reduction may be explained by the loss of about 2kb from *ycf1* pseudogene and also for expansion of IR in the IRb/SSC boundary. Finally, the reduction in the SSC and IR regions explain the reduction in total genome size observed in *P. cincinnata* (Table 2).

4.4. Repeat sequence analysis

The use of plastid single sequence repeat (ptSSRs or microsatellites) as plastid markers was firstly introduced by Powell et al. (1995). ptSSRs exhibit length variation and polymorphism, which can be useful tool to analyze the genetic diversity. The ptSSRs smaller than 15 bp have been identified in several plastomes of land plants (Marshall et al., 2001; Provan et al., 2001; Raubeson et al., 2007; Vieira et al., 2014a, 2016a). In the *P. cincinnata* plastome, we identified 213 SSRs (Table 4). The ptSSRs identified here can be assessed for intraspecific and interspecific analyzes of polymorphism (Powell et al., 1995; Vieira et al., 2014a; Rogalski et al., 2015) in *P. cincinnata* accesses and other Passifloraceae species. Most ptSSRs are mononucleotide repeats (homopolymers) and di-, tri-, tetra, pent and hexa-nucleotide repeats are less common (George et al., 2015). It is in accordance with what we found here for *P. cincinnata*. From 213 ptSSRs identified here, 141 were mono- and dionucleotides repeats were the most common with 141 and 58, respectively; whereas tri-(3) and tetra-nucleotides (11) occurred with lower frequency and penta- and hexa-nucleotides repeats were not found in *P. cincinnata* plastome (Table 4). Most tetra-nucleotide repeats (6) found here is localized within of an unique intergenic region (bewteen *accD* and rbcL;

Table 5). Moreover, two trinucleotide repeats are localized within the coding-sequence of *clpP* gene (Table 5), forming an insertion with no hit in BLAST search. This insertion containing repetitive nucleotide and may be related to the high genetic divergence and evolution rate showed here for this gene in *P. cincinnata.*

Through of the TRF software, 51 tandem repeats with repetition units of 12-96 pb were identified in *P. cincinnata* plastome (Table 6). The importance of tandem repeats to plastome structure and evolution is poorly understood, but some reports suggested that this type of repeated sequence play an import role in plastome size, structure and evolution in some flowering-plants lineages (Jo et al., 2011; Sloan et al., 2014; Dugas et al., 2016; Vieira et al., 2016b). Among tandem repeats found here, one is localized in the coding-sequence of *rps18* gene, probably contributing for the C-terminal extension reported to this gene in *P. cincinnata* (Figure 13; Table 6). Furthermore, three tandem repeats were found in the *accD* ORF, constituting an insertion (with no hits in BLAST search) in the middle of its sequence. As seen for the SSR sites, several tandem repeats were also found in the intergenic region between *accD* and *rbcL* (Table 6). Additionally, we identified 18 direct repeats (≥30 bp) in the *P. cincinnata* plastome, using the software Reputer (Table 7). Interestingly, most of these direct repeats were found within *accD* coding-sequence and in the *accD*/*rbcL* intergenic region (Table 7). Taken together, the results showed that, in *P. cincinnata* plastome, the region including the *accD* gene and intergenic sequence between *accD* and *rbcL* genes is extremely rich in repetitive elements. As suggested to the *clpP* gene, this accumulation of repetitive nucleotides, in the sequence-coding and flanking the *accD* gene, may be related with the high substitution rate and the elevated divergence showed for this gene in *P. cincinnata* plastome. In summary, it was showed here several examples of repetitive sequences that could be affecting the structure of plastid genes in *P. cincinnata*.

4.5. RNA editing sites prediction

RNA editing has been found in plastid transcripts from all major lineages of land plants (Freyer et al., 1997). RNA editing can create start and termination codons, and restore conserved amino acids modifying the sequence of mRNAs to be translated, producing a functional protein (Tsudzuki et al., 2001; Takenaka et al., 2013). The PREP software identified 62 putative RNA editing sites in the *P. cincinnata* plastome (Table 8). All the sites occurred in the first or second codon position and all nucleotide changes

observed were from C-to-U, which is the predominant form of RNA editing commonly found in seed plants (Bock, 2000; Takenaka et al., 2013). In other hand, the reverse C-to-U editing is common in other plants groups, such hornworts, ferns and Lycophyta (Yoshinaga et al., 1996; Takenaka et al., 2013; Oldenkott et al., 2014). Among 62 sites identified in the *P. cincinnata* plastome, 26 have been experimentally validated in at least one of the angiosperms species *Jatropha curcas* (Asif et al., 2010), *Hevea brasilienses* (Tangphatsornruang et al., 2011); *Arabidopsis thaliana* (Tillich et al., 2005), *Nicotiana tabacum* (Hirose et al., 1999b), *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 sativus* (Corneille et al., 2000).

Among the 11 editing sites predicted to the *ndhB* transcript (Table 8), 9 out of them were experimentally confirmed in at least one of the 9 species mentioned above. The editing sites in the codon positions 138, 178, 259 and 261 were the most frequent, occurring in all 9 species, whereas the putative editing sites in the codon 9 and 423 have not been validated in any species. The software predicted 11 editing sites in the *ndhD* transcript. Five of them (in the codon positions 1, 7, 200, 225 and 485) were validated in at least three of these species. The edition site in the first codon creates a start codon in the *ndhD* transcripts and it is generally found in several angiosperms lineages. Takenaka et al. (2013) suggest that the RNA editing mechanism has the role of controlling the available pool of mature functional RNA molecules. The *ndhA* gene also contains a predicted editing site (in codon position 114) validated in *N. tabacum* (Hirose et al., 1999b), *A. belladonna* (Schmitz-Linneweber et al., 2002), *A. thaliana* (Tillich et al., 2005) and *C. nucifera* (Huang et al., 2013). These results suggest that the *ndh* genes (predominately the *ndhB* and *ndhD*) are the genes more frequently edited in *P. cincinnata* and most editing sites are well conserved in flowering-plants.

The *rpo* genes are the second gene class with more predicted editing sites in *P. cincinnata* plastome (Table 8). Among these, the editing sites in the codon positions 113 and 184 of the *rpoB* transcript, in the codon position 14 of the *rpoC1* transcript, and in the codon position 904 and 1260 of the *rpoC2* transcript were validated in at least one of the following species: *A. thaliana* (Tillich et al., 2005), *N. tabacum* (Hirose et al., 1999b), *S. lycopersicum* (Kahlau et al., 2006), *A. belladonna* (Schmitz-Linneweber et al., 2002), *C. nucifera* (Huang et al., 2013) and *Z. mays* (Maier et al., 1995; Bock et al., 1997). The editing sites predicted to *rps2*, *rps14*, *psbE*, *atpI* and *clpP* genes were also

validated in at least one of these species. Moreover, 7 editing sites were predicted to the *accD* transcript in *P. cincinnata* (table 8), but only one (in the codon position 504) was experimentally confirmed (Huang et al., 2013). Experimental procedures, such as transcript sequencing, can confirm if the remaining editing sites of *accD* (and also those predicted to other genes and that have not been validated) represent unprecedented editing sites or are false positives.

4.6. Phylogenetic position of *P. cincinnata* (Passifloraceae) within Malpighiales

The Bayesian inference (BI) performed here, using a data matrix including 64 plastid protein-coding genes of 19 taxa, supports, with 1.00 of posterior probability (PP), the monophyly of Malpighiales (Figure 5). The monophylethic origin of this order and its component families is also well supported for several previous works, using different approaches (BI and Maximum Likelihood) and different genes combinations, such plastid genes, plastid intergenic regions and introns, mitochondrial genes and nuclear genes (Davis et al., 2005; Tokuoka and Tobe, 2006; Korotkova et al., 2009; Wurdack and Davis, 2009; Xi et al., 2012).

Here, it was found, with high degree of confidence, that Passifloraceae is more closely related to Salicaceae (1.00 PP) and that both families form a clade that is a sister group to Violaceae family (1.00 of PP). The clade formed for these three families (parietal clade) is also well supported in other phylogeny studies (Korotkova et al., 2009; Wurdack and Davis, 2009; Xi et al., 2012). However, differences in the relationship within parietal clade are found in the literature. In work of Wurdack and Davis (2009), using maximum likelihood (ML) method and a data matrix including plastid (only 4 genes), mitochondrial and nuclear genes, the Passifloraceae family formed a sister group with Violaceae (supported with 80% of boostrap value) and both families formed a clade that was sister to Salicaceae. Similar results were found for Tokuoka and Tobe (2006) using plastid and nuclear genes. In other hand, Xi et al. (2012), using a data matrix including a core of 82 plastid genes and BI e ML methods, found that Passifloraceae and Salicaceae form a well-supported (96% of boostrap value and 1.00 of PP) clade (salicoids), which is sister group to Violaceae. These results are in fine with Korotkova et al. (2009) (using the plastid *petD* intron) and corroborate with the phylogenetic position inferred to Passifloraceae in the present work (Figure 5).

5. Conclusions and outlook

In this work, the complete sequence plastome of *P. cincinnata* was obtained, representing the first *Passiflora* species with plastome full sequenced, assembled and characterized in detail. This genome showed a high number of gene losses of which essential for cell viability (*infA*, *rps7*, *rps16*, *rpl20*, *rpl22*, *ycf1* and *ycf2*). Due to the essential function of these genes for plant cell, it was suggested the occurrence of various events of functional transfer of these plastid genes to the nucleus in this species. Transfer or functional substitution of *infA*, *rps16* and *rpl22* already were reported in other species (Millen et al., 2001; Ueda et al., 2008; Jansen et al., 2011), but it was not still reported to *rps7* and *rpl20*. Thus, the transfer to the nucleus and functional substitution of plastid *rps7* and *rpl20* genes suggested here to *P. cincinnata* is an unprecedented event in land plant plastomes. Moreover, this genome showed highly divergent genes, *accD* and *clpP*, which may have been degenerated into pseudogenes in this species. On the other hand, the translated amino acid sequence of these genes showed the conservation of most functional domains of AccD and ClpP proteins, suggesting that these genes may still be functional. The genome also showed alteration, in comparison with related species, in N- and C-terminus of *ndhB* and *rps18* genes, respectively. Additionally, the loss of conserved introns of *atpF* and *clpP* genes. Multiple inversions were also detected in *P. cincinnata* plastome, which changed the gene order in LSC region. Moreover, events of IR expansions resulted in the inclusion of full *rps19* gene and the full pseudogene *ycf1* in the IR of *P. cincinnata* plastome. Several SSRs were detected in this plastome and may be used to intraspecific and interspecific genetic studies. Curiously, several SSRs and other repetitive sequences were concentrated predominantly in *accD/rbcL* intergenic region. The putative evolutionary and/or functional explanation for the accumulation of these repetitive elements, in this region, still remains to be investigated. The RNA editing sites prediction showed a high number of editing sites in *ndh* and *rpo* genes in *P. cincinnata* according with other angiosperms, whereas several sites predicted here were not reported in other species and need experimental validation. The inferred phylogenetic position of Passifloraceae family, using 64 complete sequences of *P. cincinnata* plastid genes, showed similar topology to recent phylogenetic studies (Korotkova et al., 2009; Xi et al., 2012). Taken together, these results showed an uncommon evolution of *P. cincinnata* plastome, characterized for several gene/introns losses, genome inversions and high nucleotide substitution rates in several genes.

The unusual characteristic reported here to *P. cincinnata* may or not be specific for this species. It is possible that other species of *Passiflora* genus do not share several of the losses and gene structure alterations reported here in *P. cincinnata*. Thus, it is attempting to speculate that the unusual evolutionary events described in this plastome may be related, direct or indirectly, to the nucleus-plastome incompatibility, observed in this genus (Mráček, 2005). The high variability of *accD* gene was already described as cause of genome-plastome incompatibility in inter-specific hybrids of *Pisum* (Bogdanova et al., 2015). Similarly, aiming to study the incompatibility in *Passiflora* genus, the highly divergent *accD* gene, reported in *P. cincinnata*, is a potential candidate to be investigated in other species of this genus. Other candidates could be also the highly divergent *clpP*, the extended *rps18* and genes essential that were lost (such as *ycf1, ycf2* and essential ribosomal protein genes). Thereby advances in the plastid genomic in Passiflora are necessary to build a more accurate and complete study of nucleus-plastome interaction in this genus. These advances may also allow the identification of the phylogenetic origin of several gene losses, rearrangements and repetitive elements found here in *P. cincinnata*.

Finally, the complete plastome sequence of *P. cincinnata* reported here and the advances to obtain efficient protocols of in vitro regeneration (Dornelas and Vieira, 1993; Reis, 2005; Lombardi et al., 2007; Silva et al., 2011), enables the plastid genetic transformation of this species. This technology is a powerful tool to answer several question related to the function and essentiality of the putative pseudogenes described here. Moreover, there are several biotechnology applications of plastid genetic transformation (Maliga and Bock, 2011; Bock, 2015; Jin and Daniell, 2015; Daniell et al., 2016), which may be also applicated to the economically important *Passiflora* genus.

6. References

Ahlert, D., Ruf, S., Bock, R. (2003). Plastid protein synthesis is required for plant development in tobacco. Proc. Natl. Acad. Sci. U. S. A. **100**, 15730-15735.

Alkatib S., Scharff L.B., Rogalski M., Fleischmann T.T., Matthes A., Seeger S., Schöttler M.A., Ruf S., Bock R. (2012b). The contributions of wobbling and superwobbling to the reading of the genetic code. PLoS Genet. **8**, e1003076.

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.

Bagherieh-Najjar, M.B., Vries, O.M.H., Hille, J., Dijkwel, P.P. (2005). Arabidopsis RecQl4A suppresses homologous recombination and modulates DNA damage responses. Plant J. **43,** 789-798**.**

Barthet, M.M., Moukarzel, K., Smith, K.N., Patel, J., Hilu, K.W. (2015). Alternative translation initiation codons for the plastid maturase MatK: unraveling the pseudogene misconception in the Orchidaceae. BMC Evol. Biol. **15**.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27**, 573-580.

Bernacci, L.C., Meletti, L.M.M., Soares-Scott, M.D., Passos, I.R.S., Junqueira, N.T.V. (2005). Espécies de maracujá: caracterização e conservação da biodiversidade. In: Maracujá: germoplastoma e melhoramento genético. Faleiro, F. G., Junqueira, N. T. V., Braga, M. F. (Eds). Embrapa Cerrados, p. 559-586.

Blazier, J.C; Ruhlman, T.A., Weng, M-L., Rehman, S.K., Sabir, J.S.M., Jansen, R. (2016). Divergence of RNA polymerase α subunits in angiosperm plastid genomes is mediated by genomic rearrangement. Sci. Rep. **6**, 24595.

Bock, R., Albertazzi, F., Freyer, R., Fuchs, M., Ruf, S., Zeltz, P., Maier, R.M. (1997) Transcript editing in chloroplasts of higher plants. In: Eukaryotism and symbiosis. Schenk, H.E.A., Herrmann, R., Jeon, K.W., Müller, N.E., Schwemmler, W. (Eds). Springer, Berlin, pp. 123-137.

Bock, R. (2000). Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing. Biochimie. **82,** 549-557.

Bock, R. (2001). Transgenic plastids in basic research and plant biotechnology. J. Mol. Biol. **312**, 425-438.

Bock, R. (2007). Structure, function, and inheritance of plastid genomes. In: Cell and Molecular Biology of Plastids. Bock, R. (Ed). Topics in Current Genetics. 19, 524p.

Bock, R. (2013). Strategies for metabolic pathway engineering with multiple transgenes. Plant Mol. Biol. **83**, 21–31

Bock, R. (2015). Engineering Plastid Genomes: Methods, Tools, and Applications in Basic Research and Biotechnology. Annu. Rev. Plant Biol. **66**, 211–241.

Bock, R., Timmis, J.N. (2008). Reconstructing evolution: Gene transfer from plastids 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.

Cai, X., Fuller, A.L., McDougald, L.R., Zhu, G. (2003). Apicoplast genome of the coccidian *Eimeria tenella*. Gene. **321**, 39-46.

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.

Cerqueira-Silva, C.B.M., Santos, E.S.L., Jesus, O.N., Mori, G.M., Jesus, O.N., Corrêa, R.X., Souza, A.P. (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.

Cerqueira-Silva, C.B.M., Santos, E.S.L., Vieira, J.G.P., Mori, G.M., Jesus, O.N., Corrêa, R.X., Souza, A.P. (2014b). New Microsatellite Markers for Wild and Commercial Species of *Passiflora* (Passifloraceae) and Cross-Amplification. Appl. Plant Sci. **2**, 1300061.

Cerutti, H., Osman, M., Grandoni, P., Jagendorf, A.T. (1992). A homolog of *Escherichia coli* RecA protein in plastids of higher plants. Proc. Natl. Acad. Sci. U. S. A. **89**, 8068-8072.

Chumley, T.W., Palmer, J.D., Mower, J.P., Fourcade, H.M., Calie, P.J., Boore, J.L., Jansen, R.K. (2006). The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol. Biol. Evol. **23**, 2175–2190.

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.

Cosner, M.E., Jansen, R.K., Palmer, J.D., Downie, S.R. (1997) The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. Curr. Genet. **31**, 419-429.

Cosner, M.E., Raubeson, L.A., Jansen, R.K. (2004). Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. BMC Evol. Biol. **4**, 1.

Cummings, H.S., Hershey, J.W. (1994). Translation initiation factor IF1 is essential for cell viability in *Escherichia coli*. J. Bacteriol. **176**, 198–205.

Daniell, H., Wurdack, K.J., Kanagaraj, A., Lee, S.-B., Saski, C., Jansen, R.K. (2008). The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron. Theor. Appl. Genet. **116**, 723–737.

Daniell, H., Lin, C.-S., Yu, M., Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. **17** (1).

Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**, 1394–1403.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods. **9**, 772–772.

Davis, C.C., Webb, C.O., Wurdack, K.J., Jaramillo, C.A., Donoghue, M.J. (2005). Explosive radiation of Malpighiales supports a mid-Cretaceous origin of modern tropical rain forests. Am. Nat. **165**, E36-E65.

Dornelas, M.C., Vieira, M.L.C. (1993). Plant regeneration from protoplast cultures of *Passiflora edulis* var. *flavicarpa* Deg., *P. amethystine* Mikan. And *P. cincinnata* Mast. Plant Cell Rep. **13**, 103-106.

Downie, S.R., Katz-Downie, D.S., Wolfe, K.H., Calie, P.J., Palmer, J.D. (1994). Structure and evolution of the largest chloroplast gene (ORF2280): internal plasticity and multiple gene loss during angiosperm evolution. Curr. Genet. **25**, 367–378.

Doyle, J.J., Davis, J.I., Soreng, R.J., Garvin, D., Anderson, M.J. (1992). Chloroplast DNA inversions and the origin of the grass family (Poaceae). Proc. Natl. Acad. Sci. U. S. A. **89**, 7722-7726.

Doyle, J.J., Doyle, J.L., Palmer, J.D. (1995). Multiple independent losses of two genes and one intron from legume chloroplast genomes. Syst. Bot. **20,** 272-294.

Drescher A, Ruf S, Calsa T Jr, 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.

Dugas, D.V., Hernandez, D., Koenen, E.J.M., Schwarz, E., Straub, S., Hughes, C.E., Jansen, R.K., Nageswara-Rao, M., Staats, M., Trujillo, J.T., Hajrah, N.H., Alharbi, N.S., Al-Malki, A.L., Sabir, J.S.M., Bailey, C.D. (2015). Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in clpP. Sci. Rep. **5**, 16958.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**, 1792–1797.

Erixon, P., Oxelman, B. (2008). Whole-Gene Positive Selection, Elevated Synonymous Substitution Rates, Duplication, and Indel Evolution of the Chloroplast clpP1 Gene. PLoS ONE. **3**, e1386.

Fajardo, D., Senalik, D., Ames, M., Zhu, H., Steffan, S.A., Harbut, R., Polashock, J., Vorsa, N., Gillespie, E., Kron, K., Zalapa, J.E. (2013). Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. Tree Genet. Genomes **9**, 489–498.

Faleiro, F.G., Junqueira, N.T.V., Bragam M.F., Peixoto, J.R. (2005). Germoplasma e melhoramento genético do maracujazeiro – Desafios da pesquisa. In: Maracujá: germoplastoma e melhoramento genético. Faleiro, F. G., Junqueira, N. T. V., Braga, M. F. (Eds). Embrapa Cerrados, p. 187-202.

Feuillet, C., MacDougal, J.M. (2004). A New Infrageneric Classification of *Passiflora* L. (Passifloraceae). Passiflora. **13**, 34-38.

Fleischmann, T.T., Scharff, L.B., Alkatib, S., Hasdorf, S., Schottler, M.A., and 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.

Freyer, R., Kiefer-Meyer, M.-C., Kössel, H. (1997). Occurrence of plastid RNA editing in all major lineages of land plants. Proc. Natl. Acad. Sci. U. S. A. **94**, 6285–6290.

Gardner, M.J., Bishop, R., Shah, T., de Villiers, E.P., Carlton, J.M., Hall, N., Ren, Q., Paulsen, I.T., Pain, A., Berriman, M., Wilson, R.J., Sato, S., Ralph, S.A., Mann, D.J., Xiong, Z., Shallom, S.J., Weidman, J., Jiang, L., Lynn, J., Weaver, B., Shoaibi, A., Domingo, A.R., Wasawo, D., Crabtree, J., Wortman, J.R., Haas, B., Angiuoli, S.V., Creasy, T.H., Lu, C., Suh, B., Silva, J.C., Utterback, T.R., Feldblyum, T.V., Pertea, M., Allen, J., Nierman, W.C., Taracha, E.L., Salzberg, S.L., White, O.R., Fitzhugh, H.A., Morzaria, S., Venter, J.C., Fraser, C.M., Nene, V. (2005). Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. Science. **309**, 134-137.

Go, Y., Satta, Y., Takenaka, O., Takahata, N. (2005). Lineage-specific loss of function of bitter taste receptor genes in humans and nonhuman primates. Genetics. **170**, 313-326.

Goremykin, V.V., Holland, B., Hirsch-Ernst, K.I, Hellwig, F.H. (2005). Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. Mol. Biol. Evol. **22**, 1813-1822.

Goulding, S.E., Olmstead, R.G., Morden, C.W., Wolfe, K.H. (1996). Ebb and flow of the chloroplast inverted repeat. Mol. Gen. Genet**. 252**, 195–206.

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.

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., Bock, R. (2013). Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. BioEssays. **35**, 354–365.

Guisinger, M.M., Kuehi, J., Boore, J. L., Jansen, R.K. (2008). Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. Proc. Natl. Acad. Sci. U. S. A. **105** (47), 18424-18429.

Guisinger, M.M., Chumley, T.W., Kuehl, J.V., Boore, J.L., Jansen, R.K. (2010). Implications of the Plastid Genome Sequence of *Typha* (Typhaceae, Poales) for Understanding Genome Evolution in Poaceae. J. Mol. Evol. **70**, 149–166.

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.

Guo, X., Castillo-Ramírez, S., González, V., Bustos, P., Luís Fernández-Vázquez, J., Santamaría, R., Arellano, J., Cevallos, M.A., Dávila, G. (2007). Rapid evolutionary change of common bean (*Phaseolus vulgaris* L) plastome, and the genomic diversification of legume chloroplasts. BMC Genomics. **8**, 228.

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.

Hager, M., Biehler, K., Illerhaus, J., Ruf, S., Bock, R. (1999). Targeted inactivation of the smallest plastid genome-encoded open reading frame reveals a novel and essential subunit of the cytochrome b(6)f complex. EMBO J. **18,** 5834-5842.

Hajdukiewicz, P.T., Allison, L.A., 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.

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.

Hansen, A.K., Gilbert, L.E., Simpson, B.B., Downie, S.R., Cervi, A.C., Jansen, R.K. (2006). Phylogenetic Relationships and Chromosome Number Evolution in *Passiflora*. Syst. Bot. **31**, 138–150.

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.

Harris, E.H., Boynton, J.E., Gillham, N.W. (1994). Chloroplast ribosomes and protein synthesis. Microbiol Rev. **58**, 700-754.

Hachtel, W., Neuss, A., Stein, J.V. (1991). A chloroplast DNA inversion marks an evolutionary split in the genus *Oenothera*. Evolution. **45**, 1050-1052.

Herrmann, R.G., Maier, R.M., Schmitz-Linneweber, C. (2003). Eukaryotic genome evolution: rearrangement and coevolution of compartmentalized genetic information. Philos. Trans. R. Soc. B Biol. Sci. **358**, 87–97.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y., Li, Y-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K., Sugiura, M. (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. **217**, 185-194.

Hirose, T., Sugiura, M. (1997). Both RNA editing and RNA cleavage are required for translation of tobacco chloroplast ndhD mRNA: a possible regulatory mechanism for the expression of a chloroplast operon consisting of functionally unrelated genes. EMBO J. **16**, 6804–6811.

Hirose, T., Ideue, T., Wakasugi, T., Sugiura, M. (1999a). The chloroplast infA gene with a functional UUG initiation codon. FEBS Lett. **445**, 169–172.

Hirose, T., Kusumegi, T., Tsudzuki, T., Sugiura, M. (1999b). RNA editing sites in tobacco chloroplast transcripts: editing as a possible regulator of chloroplast RNA polymerase activity. Mol. Gen. Genet. **262**, 462–467.

Hoot, S.B., Palmer, J.D. (1994). Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera. J. Mol. Evol. **38**, 274–281.

Huang, Y.-Y., Matzke, A.J.M., Matzke, M. (2013). Complete Sequence and Comparative Analysis of the Chloroplast Genome of Coconut Palm (*Cocos nucifera*). PLoS ONE. **8**, e74736.

Hupfer, H., Swaitek, M., Hornung, S., Herrmann, R.G., Maier, R.M., Chiu, W-L., Sears, B. (2000). Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable *Euoenothera* plastomes. Mol. Gen. Genet. **263**, 581-5.

Jansen, R.K., Palmer, J.D. (1987). A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proc. Natl. Acad. Sci. U. S. A. **84**, 5818-5822.

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., Lee, S-B., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci. U. S. A. **104**, 19369–19374.

Jansen, R.K., Wojciechowski, M.F., Sanniyasi, E., Lee, S.-B., Daniell, H. (2008). Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (Leguminosae). Mol. Phylogenet. Evol. **48**, 1204–1217.

Jansen, R. K., Ruhlman, T. A. (2012). Plastid Genomes of Seed Plantas. In: Genomics of Chloroplasts and Mitochondria: Advances in Photosynthesis and Respiration. Bock, R., Knoop, V. (Eds). Spring Science and Business Media. **35**, 103-126.

Jansen, R.K., Saski, C., Lee, S.-B., Hansen, A.K., 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.

Jin, S., Daniell, H. (2015). The Engineered Chloroplast Genome Just Got Smarter. Trends Plant Sci. **20**, 622–640.

Jo, Y.D., Park, J., Kim, J., Song, W., Hur, CG., Lee, Y.H., Kang, B.C. (2011) Complete sequencing and comparative analyses of the pepper (*Capsicum annuum L*.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. Plant Cell Rep. **30**, 217–229.

Junqueira, N.T.V., Braga, M.F., Faleiro, F.G., Peixoto, J. R., Bernacci, L.C. (2005). Potencial de espécies silvestres de maracujazeiro como fonte de resistência a doenças. In: Maracujá: germoplastoma e melhoramento genético. Faleiro, F. G., Junqueira, N. T. V., Braga, M. F. (Eds). Embrapa Cerrados, p. 81-108.

Kahlau, S., Aspinall, S., Gray, J.C., and 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., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., Tabata, S. (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.

Katayama, H., Ogihara, Y. (1996). Phylogenetic affinities of the grasses to other monocots as revealed by molecular analysis of chloroplast DNA. Curr. Genet. **29**, 572-581.

Knox, E.B., Palmer, J.D. (1999). The chloroplast genome arrangement of *Lobelia thuliniana* Lobeliaceae: expansion of the inverted repeat in an ancestor of the Campanulales. Plant Syst Evol. **214**, 49-64.

Kiill, L.H.P., Siqueira, K.M.M., Araújo, F.P., Trigo, S.P.M., Feitoza, E.A., Lemos, I.B. (2010). Biologia reprodutiva de *Passiflora cincinnata* Mast. (Passifloraceae) na região de Petrolina-PE. Oecologia Aust. **14**, 115–127.

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.

Kode, V., Mudd, E.A., Iamtham, S., Day, A. (2005). The tobacco plastid accD gene is essential and is required for leaf development: Essential plastid gene. Plant J. **44**, 237–244.

Konishi, T., Shinohara, K., Yamada, K., Sasaki, Y. (1996). Acetyl-CoA carboxylase in higher plants: most plants other than gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. Plant Cell Physiol. **37**: 117–122.

Korotkova, N., Schneider, J.V., Quandt, D., Worberg, A., Zizka, G., Borsch, T. (2009). Phylogeny of the eudicot order Malpighiales: analysis of a recalcitrant clade with sequences of the petD group II intron. Plant Syst. Evol. **282**, 201–228.

Kuroda, H., Maliga, P. (2003). The plastid clpP1 protease gene is essential for plant development. Nature. **425**, 86-89.

Kurtz, S., Choudhuri, J.V, Ohlebusch, E., Scheleiermacher, C., Stoye, J., Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. **29** (22), 4633–4642.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. **5**, R12.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S. (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. Mol. Biol. Evol. **29**, 1695–1701.

Lee, H.-L., Jansen, R.K., Chumley, T.W., Kim, K.-J. (2007). Gene Relocations within Chloroplast Genomes of *Jasminum* and *Menodora* (Oleaceae) Are Due to Multiple, Overlapping Inversions. Mol. Biol. Evol. **24**, 1161–1180.

Lee, S.S., Jeong, W.J., Bae, J.M., Bang, J.W., Liu, J.R., Harn, C.H. (2004). Characterization of the plastid-encoded carboxyltransferase subunit (accD) gene of potato. Mol. Cells. **17**, 422-9.

Levin, D.A. (2003). The cytoplasmic factor in plant speciation. Syst. Bot. **28**, 5–11.

Lindholm, J., Gustafsson, P. (1991). A three-step model for the rearrangement of the chloroplast trnK-psbA region of the gymnosperm *Pinus contorta*. Nucleic Acids Res. **19**, 2881-2887.

Lin, Z., Kong, H., Nei, M., Ma, H. (2006). Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. Proc. Natl. Acad. Sci. U. S. A. **103**, 10328-10333.

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.

Lombardi, S.P., Passos, I.R.S., Nogueira, M.C.S., Appezzato-da-Glória, B. (2007). *In vitro* regeneration from roots and leaf discs of *Passiflora cincinnata* Mast. Braz. Arch. Boil. Technol. **50**, 239-247.

Magee, A.M., Aspinall, S., Rice, D.W., Cusack, B.P., Semon, M., Perry, A.S., Stefanovic, S., Milbourne, D., Barth, S., Palmer, J.D., Gray, J.C., Kavanagh, T.A., Wolfe, K.H. (2010). Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res. **20**, 1700–1710.

Maier, R.M., Neckermann, K., Igloi, G.L., Kosel, 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.

Maliga, P. (2004). Plastid transformation in higher plants. Annu. Rev. Plant Biol. **55**, 289–313.

Maliga, P., Bock, R. (2011). Plastid biotechnology: food, fuel, and medicine for the 21st century. Plant Physiol. **155**, 1501–1510.

Marshall, H.D., Newton, C., Ritland, K. (2001). Sequence-repeat polymorphisms exhibit the signature of recombination in lodgepole pine chloroplast DNA. Mol. Biol. Evol. **18**, 2136-2138

Martin, G.E., Rousseau-Gueutin, M., Cordonnier, S., Lima, O., Michon-Coudouel, S., Naquin, D., de Carvalho, J.F., 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.

Martínez-Alberola, F., del Campo, E.M., Lázaro-Gimeno, D., Mezquita-Claramonte, S., Molins, A., Mateu-Andrés, I., Pedrola-Monfort, J., Casano, L.M., Barreno, E. (2013). Balanced Gene Losses, Duplications and Intensive Rearrangements Led to an Unusual Regularly Sized Genome in Arbutus unedo Chloroplasts. PLoS ONE **8**, e79685.

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. U. S. A. **107**, 4623–4628.

Mower, J.P. (2009). The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res. **37**, W253–W259.

Mráček, J. (2005). Investigation of genome–plastome incompatibility in *Oenothera* and *Passiflora*. PhD Thesis, pp. 1–104. Ludwig Maximilians-University, Munich.

Nakai, M. (2015). The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. Biochim. Biophys. Acta BBA - Bioenerg. **1847**, 957–967.

Nakasone, H., Paull, R.E. (1998). Tropical Fruits. Cab Internacional, Wallington, p. 270-291.

Nowotny, V., Nierhaus, K.H. (1980). Protein L20 from the large subunit of *Escherichia coli* ribosomes is an assembly protein. J. Mol. Biol. **137**, 391-399.

Nowotny, V., Nierhaus, K.H. (1988). Assembly of the 30S subunit from *Escherichia coli* ribosomes occurs via two assembly domains which are initiated by S4 and S7. Biochemistry **27**, 7051-7055.

Ohlrogge, J. and Browse, J. (1995). Lipid biosynthesis. Plant Cell, **7**, 957-970.

Oliveira, J. C., Ruggiero, C. (2005). Espécies de Maracujá com potencial agronômico. Espécies de maracujá: caracterização e conservação da biodiversidade. In: Maracujá: germoplastoma e melhoramento genético. Faleiro, F. G., Junqueira, N. T. V., Braga, M. F. (Eds). Embrapa Cerrados, p. 141-158.

Palmer, J.D., Osorio, B., Aldrich, J., Thompson, W.F. (1987). Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. Curr. Genet. **11**, 275-286.

Palmer JD. (1991). Plastid chromosomes: structure and evolution. In: Cell culture and somatic cell genetics in plants, vol. 7, The molecular biology of plastids. Vasil, I.K., Bogorad, L. (Eds). Academic Press, p. 5–53.

Pombert, J.F., Otis, C., Lemieux, C., Turmel, M. (2005). The chloroplast genome sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. Mol. Biol. Evol. **22**, 1903-1918.

Pombert, J.F., Lemieux, C., Turmel, M. (2006). The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. BMC Biol. **4**, 3.

Powell, W., Morgante, 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.

Raubeson, L.A., Jansen, R.K. (2005). Chloroplast genomes of plants. In: Diversity and evolution of plants genotypic and phenotypic variation in higher plants. Henry, R (Ed). CABI Publishing, p. 45-68.

Raubeson, L.A., Peery, R., Chumley, T.W., Dziubek, C., Fourcade, H.M., Boore, J.L., Jansen, R.K. (2007). Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genomics. **8**, 174.

Reis, L.B. (2005). Morfogênese in vitro e transformação genética de maracujazeiros (*Passiflora edulis* f. *flavicarpa* Degener e *P. cincinnata* Masters). PhD Thesis (Plant Physiology), 105p. Federal University of Viçosa, Viçosa-MG, Brazil.

Rivarola, M., Foster, J. Chan, A.P., Williams, A.L., Rice, D.W., Liu, X., Melake-Berhan, A., Creasy, H.H., Puiu, D., Rosovits, M.J., Khouri, H.M., Beckstrom-Sternberg, S.M., Allan, G.J., Keim, P., Ravel, J., Rabinowicz, P. (2001). Castor bean organelle genome sequencing and worldwide genetic diversity analysis. PLoS ONE. **6**, e21734.

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., Karcher, D., Bock, R. (2008a). Superwobbling facilitates translation with reduced tRNA sets. Nat. Struct. Mol. Biol. **15**, 192-198.

Rogalski, M., Schoettler, M.A., Thiele, W., Schulze, W.X., Bock, R. (2008b). Rpl33, a nonessential plastid encoded ribosomal protein in tobacco, is required under cold stress conditions. Plant Cell. **20**, 2221-2237.

Rogalski, M., Carrer, H. (2011). Engineering plastid fatty acid biosynthesis to improve food quality and biofuel production in higher plants. Plant Biotechnol. J. **9**, 54–564.

Rogalski, M., do Nascimento Vieira, L., 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**, 586.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol. **61**, 539–542.

Rousseau-Gueutin, M., Huang, X., Higginson, E., Ayliffe, M., Day, A., Timmis, J.N. (2013). Potential Functional Replacement of the Plastidic Acetyl-CoA Carboxylase Subunit (accD) Gene by Recent Transfers to the Nucleus in Some Angiosperm Lineages. Plant Physiol. **161**, 1918–1929.

Ruf, S., Kossel, H., Bock, R. (1997). Targeted inactivation of a tobacco intron-containing open reading frame reveals a novel chloroplast-encoded photosystem I-related gene. J. Cell Biol. **139**, 95-102.

Saotome, A., Kimura, S., Uchiyama, Y., Morohashi, K., Sakaguchi, K. (2006). Characterization of four RecQ homologues from rice (*Oryza sativa* L. cv. Nipponbare). Biochem. Biophys. Commun. **345**, 1283-1291.

Sasaki, Y., Nagano, Y. (2004) Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. Biosci. Biotechnol. Biochem. **68**, 1175-1184.

Saski, C., Lee, S.-B., Daniell, H., Wood, T.C., Tomkins, J., Kim, H.-G., Jansen, R.K. (2005). Complete Chloroplast Genome Sequence of *Glycine max* and Comparative Analyses with other Legume Genomes. Plant Mol. Biol. **59**, 309–322

Schattner, P., Brooks, A.N., and Lowe, T.M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. **33**, W686–W689.

Schmitz-Linneweber, C., Regel, R., Du, T.G., Hupfer, H., Herrmann, R.G., Maier, R.M. (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.

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.

Schön, A., Krupp, G., Gough, S., Berry-Lowe, S., Kannangara, C.G., Söll, D. (1986). The RNA required in the first step of chlorophyll biosynthesis is a chloroplast glutamate tRNA. Nature. **322**: 281–284.

Schwarz, E.N., Ruhlman, T.A., Sabir, J.S.M., Hajrah, N.H., Alharbi, N.S., Al-Malki, A.L., Bailey, C.D., 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.

Serino, G., Maliga, P. (1998). RNA Polymerase Subunits Encoded by the Plastid rpoGenes Are Not Shared with the Nucleus-Encoded Plastid Enzyme. Plant Physiol. **117**, 1165–1170.

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.

Sherman-Broyles, S., Bombarely, A., Grimwood, J., Schmutz, J., Doyle, J. (2014). Complete Plastome Sequences from *Glycine syndetika* and Six Additional Perennial Wild Relatives of Soybean. G3 (Bethesda). **4**, 2023–2033.

Shikanai, T., Shimizu, K., Ueda, K., Nishimura, Y., Kuroiwa, T., and 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.

Shinozaki, K., Sugiura, M. (1986). Organization of chloroplast genomes. Adv. Biophys. **21**, 57–78.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. **5**, 2043.

Shoji, S., Dambacher, C.M., Shajani, Z., Williamson, J.R., Schultz, P.G. (2011). Systematic Chromosomal Deletion of Bacterial Ribosomal Protein Genes. J. Mol. Biol. **413**, 751–761.

Silva, C.V., de Oliveira, L.S., Loriato, V.A.P., da Silva, L.C., de Campos, J.M.S., Viccini, L.F., de Oliveira, E.J., Otoni, W.C. (2011). Organogenesis from root explants of commercial populations of *Passiflora edulis* Sims and a wild passionfruit species, *P. cincinnata* Masters. Plant Cell Tissue Organ Cult. PCTOC **107**, 407–416.

Sloan, D.B., Alverson, A.J., Wu, M., Palmer, J.D., Taylor, D.R. (2012). Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. Genome Biol. Evol. **306**.

Sloan, D.B., Triant, D.A., Forrester, N.J., Bergner, L.M., Wu, M., Taylor, D.R. (2014). A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Mol. Phylogenet. Evol. **72**, 82–89.

Souza, V.C., Lorenzi, H. (2012). Botânica sistemática: guia ilustrado para identificação das famílias de Angiospermas da flora brasileira, baseado em APG III. 3ª ed. Instituto Plantarum, Nova Odessa, p. 768.

Strauss, S.H., Palmer, J.D., Howe, G.T., Doerksen, A.H. (1998). Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. Proc. Natl. Acad. Sci. U. S. A. **85**, 3898-3902.

Sugiura, M., Hirose, T. and Sugita, M. (1998). Evolution and mechanism of translation in chloroplasts. Annu. Rev. Genet. **32**, 437-459.

Swofford, D.L. PAUP*. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4.0b5. Sunderland, Massachusetts.: Sinauer Associates; 2002.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., Brennicke, A. (2013). RNA editing in Plants and Its evolution. Annu. Rev. Genet. **47**, 335-353.

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.

Tangphatsornruang, S., Uthaipaisanwong, P., Sangsrakru, D., Chanprasert, J., Yoocha, T., Jomchai, N., Tragoonrung, S. (2011). Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. Gene. **475**, 104–112.

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.

Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994). CLUSTAL W: improving the sensitivy of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and qeight matrix choice. Nucleic Acids Res. **22**(22), 4673-4680.

Tiller N., Weingartner M., Thiele W., Maximova E., Schöttler M.A., Bock R. (2012). The plastid-specific ribosomal proteins of *Arabidopsis thaliana* can be divided into non-essential proteins and genuine ribosomal proteins. Plant J. **69**, 302–316

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: Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. Plant J. **43**, 708–715.

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

Tokuoka, T., Tobe, H. (2006). Phylogenetic analyses of Malpighiales using plastid and nuclear DNA sequences, with particular reference to the embryology of Euphorbiaceae sens. str. J. Plant Res. **119**, 599–616.

Tsudzuki, T., Wakasugi, T., Sugiura, M. (2001). Comparative Analysis of RNA Editing Sites in Higher Plant Chloroplasts. J. Mol. Evol. **53**, 327–332.

Vaidya, G., Lohman, D.J., Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics **27**, 171–180.

Vanderplank, J. (1996). Passion Flowers. MIT Press, Massachusetts, p. 224.

Vieira, L. do N., Faoro, H., Rogalski, M., Fraga, H.P. de F., Cardoso, R.L.A., de Souza, E.M., Pedrosa, F de O., Nodari, R.O., and Guerra, M.P. (2014a). 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., Faoro, H., Fraga, H.P. de F., Rogalski, M., de Souza, E.M., de Oliveira Pedrosa, F., Nodari, R.O., Guerra, M.P. (2014b). An Improved Protocol for Intact Chloroplasts and cpDNA Isolation in Conifers. PLoS ONE. **9**, e84792.

Vieira, L. do N., dos Anjos, K.G., Faoro, H., Fraga, H.P. de F., Greco, T.M., Pedrosa, F. de O., de Souza, E.M., Rogalski, M., de Souza, R.F., Guerra, M.P. (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.

Vieira, L. do N., Rogalski, M., Faoro, H., Fraga, H.P. de F., Anjos, K.G, Picchi, F.A., Nodari, R.O., Pedrosa, F. de O., de Souza, E.M., Guerra, M.P. (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.

Vries, J., Sousa, F.L., Bölter, B., Soll, J., Gould, S.B. (2015). YCF1: A Green TIC? Plant Cell. **27**, 1827-1833.

Wakasugi, T., Sugita, M., Tsudzuki, T., Sugiura, M. (1998). Updated gene map of tobacco chloroplast DNA. Plant Mol. Biol. Report. **16**, 231-241.

Weglohner, W., Subramanian, A.R. (1991). A heptapeptide repeat contributes to the unusual length of chloroplast ribosomal protein-S18 - nucleotide sequence and map position of the rpl33-rps18 gene cluster in maize. FEBS Lett. **279**, 193-197.

Weglöhner, W., Kauschmann, A., Subramanian, A.R. (1995). Evolution of the NH2- and COOH-terminal extensions of chloroplast ribosomal protein S18: Nucleotide sequence of pea and rye chloroplast rps 18 genes. Biochem. Mol. Biol. Int. **36**, 265–273.

Weng, M.L., Blazier, J.C., Govindu, M., Jansen, R.K. (2013). Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. Mol. Biol. Evol. **31**, 645-559.

Williams, A.V., Boykin, L.M., Howell, K.A., Nevill, P.G., Small, I. (2015). The Complete Sequence of the *Acacia ligulata* Chloroplast Genome Reveals a Highly Divergent clpP1 Gene. . PLoS ONE . **10**, e0125768.

Wilson, R.J.M. (2002) Progress with parasite plastids. J. Mol. Biol. **319**, 257–274.

Wolfe, K.H., Morden, C.W., Palmer, J.D. (1992). Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc. Natl. Acad. Sci. U. S. A. **89**, 10648–10652.

Wu, C.I., Li, W.H., Shen, J.J., Scarpulla, R.C., Limbach, K.J., Wu, R. (1986) Evolution of cytochrome c genes and pseudogenes. J. Mol. Evol. **23**(1), 61-75.

Wyman, S.K., Jansen, R.K., Boore, J.L. (2004). Automatic annotation of organellar genomes with DOGMA. Bioinformatics. **20**, 3252–3255.

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. U. S. A. **109**, 17519–17524.

Yamaguchi, K., Subramanian, A.R. (2000). The plastid ribosomal proteins. Identification of all the proteins in the 50 S subunit of an organelle ribosome (chloroplast). J. Biol. Chem. **275**, 28466-28482.

Yamaguchi, K., von Knoblauch, K., Subramanian, A.R. (2000). The plastid ribosomal proteins. Identification of all the proteins in the 30 S subunit of an organelle ribosome (chloroplast). J. Biol. Chem . **275**, 28455-28465.

Yamaguchi, K., Subramanian, A.R. (2003) Proteomic identification of all plastid-specific ribosomal proteins in higher plant chloroplast 30S ribosomal subunit. Eur. J. Biochem. **270**, 190-205.

Yoshinaga, K., Kakehi, T., Shima, Y., Iinuma, H., Masuzawa, T., Ueno, M. (1997). Extensive RNA editing and possible double-stranded structures determining editing sites in the *atpB* transcripts of hornwort chloroplasts. Nucleic Acids Res. **25**, 4830-4834.

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.

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