University of São Paulo
“Luiz de Queiroz” College of Agriculture

Chloroplast genome evolution in *Passiflora*

Luiz Augusto Cauz dos Santos

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding.

Piracicaba
2020

Luiz Augusto Cauz dos Santos
BSc in Biotechnology

Chloroplast genome evolution in *Passiflora*

versão revisada de acordo com a Resolução CoPGr 6018 de 2011

Advisor:
Professor **MARIA LUCIA CARNEIRO VIEIRA**

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding.

Piracicaba
2020

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Santos, Luiz Augusto Cauz dos

Chloroplast genome evolution in *Passiflora* / Luiz Augusto Cauz dos Santos. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2020.

150 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Genoma cloroplastidial 2. Rearranjos genômicos 3. Filogenômica 4. *Passiflora* I. Título

DEDICATORY

A Deus,

A minha familia, em especial aos meus avós

Luiz e Beatriz Cauz (In memorian).

ACKNOWLEDGEMENTS

À Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ/USP), e ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela concessão das bolsas de estudo e pelo suporte financeiro ao projeto (processos DD: 2017/04216-9; BEPE: 2018/06939-0; BEPE: 2019/09210-4; BIOTA: 2017/11815-6).

Em especial, à Professora Maria Lucia Carneiro Vieira, pelas oportunidades oferecidas e pela confiança em mim depositada. Agradeço imensamente sua orientação, supervisão e apoio científico neste trabalho.

Ao Professor Cássio van den Berg, pela sua contribuição e por todos os conhecimentos partilhados durante o desenvolvimento deste trabalho.

À Dra. Hélène Bergès e a toda equipe do CNRGV (Centre National de Ressources Génomiques Végétales) do INRA, Toulouse (France), por toda a colaboracao e auxilio no desenvolvimento deste trabalho.

À Dra. Maria Imaculada Zucchi e ao seu grupo de pesquisa pela colaboração nos experimentos.

A todos os docentes do Programa de Pós-Graduação em Genética e Melhoramento de Plantas da ESALQ, pela atenção, dedicação e disposição em compartilharem seus conhecimentos. Em particular ao Professor Giancarlo Conde Xavier Oliveira, pelos ensinamentos e apoio científico nas análises filogenéticas e evolutivas.

Ao Carlos Alberto de Oliveira, pela amizade e pelo suporte técnico durante a realização deste trabalho.

Aos profissionais dos Herbários da UEFS (HUEFS) e da UNEMAT (HERBAM), em especial a Dra Teonildes Sacramento Nunes pelo auxílio na identificação das espécies de *Passiflora*.

Ao Dr. Maurízio Vecchia curador da Coleção Nacional Italiana de Passifloras pela cessão de exemplares de *Passiflora*.

Aos amigos do laboratório de Genética Molecular de Plantas Cultivadas da ESALQ pelo ensinamento e paciência.

De uma forma singular, à minha família, aos meus pais Iara e Amélio, a minha irmã Larissa e aos meus tios Luiz Oreste e Cristina, por sempre apoiarem as minhas escolhas e por também, de certa forma, permitirem que esse sonho fosse possível.

Por fim, a todos aqueles que de alguma forma contribuíram para a minha formação e com o desenvolvimento deste trabalho deixo os meus sinceros agradecimentos.

SUMMARY

ACKNOWLEDGEMENTS **4**

RESUMO **9**

ABSTRACT **10**

LIST OF FIGURES **11**

LIST OF TABLES **13**

LIST OF ABBREVIATIONS **14**

1. INTRODUCTION **15**

REFERENCES **18**

2. THE CHLOROPLAST GENOME OF *PASSIFLORA EDULIS* (PASSIFLORACEAE) USING SEQUENCING AND ASSEMBLY APPROACHES FROM LONG SEQUENCE READS: STRUCTURAL ORGANIZATION AND PHYLOGENOMIC STUDIES IN MALPIGHIALES **21**

ABSTRACT **21**

2.1. INTRODUCTION 21

2.2. MATERIAL AND METHODS 24

2.2.1. Plant material 24

2.2.2. Sequencing and subsequent assembly of *Passiflora edulis* chloroplast DNA using the Pacbio RS II platform 24

2.2.3. Obtaining the complete chloroplast genome 25

2.2.4. Genome annotation 26

2.2.5. Comparative analysis of chloroplast genomes 26

2.2.6. Identification of repeated elements 27

2.2.7. Phylogenomic studies 27

2.3. RESULTS AND DISCUSSION 29

2.3.1. Data output from the PacBio RS II platform and assembly of chloroplast genome sequences 29

2.3.2. Organization of the *P. edulis* chloroplast genome 30

2.3.3. Comparative analysis of chloroplast genomes 35

2.3.4. Analysis of repeated elements 39
2.3.5. Phylogenomic studies 41
2.4. Conclusions 48
REFERENCES 49
3. EVOLUTIONARY HISTORY OF CHLOROPLAST GENOMES: A CASE STUDY IN POINT REVEALS A REPERTORY OF REARRANGEMENTS AND THE LOSS OF AN INVERTED REPEAT REGION IN *PASSIFLORA* 55
ABSTRACT 55
3.1. Introduction 55
3.2. Material and Methods 57
3.2.1. Plant material 57
3.2.2. Intact chloroplast isolation in sucrose gradient and cpDNA extraction 59
3.2.3. Chloroplast genome sequencing and assembly 59
3.2.4. Chloroplast genome annotation, identification of repeated elements and comparative analysis 61
3.2.5. Phylogenomic studies in *Passiflora* 61
3.3. Results 62
3.3.1. Chloroplast genome structure features in Passifloraceae 62
3.3.2. A repertory of rearrangments: inversions, expansions/loss of IR regions and gene losses 69
3.3.3. Phylogenomic studies in *Passiflora* 72
3.4. Discussion 75
3.4.1. A repertory of cpDNA rearrangements concomitant with *Passiflora* evolution and taxonomical implications 75
3.5. Conclusions 82
REFERENCES 84
4. PLASTID PHYLOGENOMICS SUPPORT A NEW INFRAGENERIC CLASSIFICATION IN *PASSIFLORA* 93
ABSTRACT 93
4.1. Introduction 93
4.2. Material and Methods 95

4.2.1. Taxon sampling for chloroplast genome sequencing 95
4.2.2. Chloroplast (cp) DNA obtainment 95
4.2.3. CpDNA sequencing and assembly 95
4.2.4. Chloroplast genome annotation and comparative analysis 96
4.2.5. Plastid phylogenomic studies 96
4.2.6. 35S rDNA sequence assembly and phylogenetic analysis 97
4.3. RESULTS 98
4.3.1. Chloroplast genome structural features and gene content 98
4.3.2. Comparative studies in *Deidamioides* based on plastome structures 99
4.3.3. The plastid phylogenomic study 102
4.3.4. The nuclear phylogeny based on complete 18S/26S gene sequences 104
4.4. DISCUSSION 106
4.5. CONCLUSIONS 110
REFERENCES 111
5. CONCLUSIONS AND FUTURE PERSPECTIVES 117
APPENDICES 119

RESUMO

Evolução do genoma cloroplastidial em *Passiflora*

Os genomas de cloroplasto (cpDNAs ou plastomas) são altamente conservados em Angiospermas. Embora rearranjos tenham sido observados em algumas espécies, os mecanismos que levam a estes rearranjos ainda são pouco elucidados. Inicialmente, nosso gupo de pesquisa descreveu a organização estrutural do plastoma de *Passiflora edulis* (Passifloraceae) e rearranjos foram identificados. Particularmente, convém ressaltar que o gênero *Passiflora* apresenta diferentes padrões de herança cloroplastidial e incompatibilidade citonuclear. Para investigar a história evolutiva dos cpDNAs em *Passiflora* e possíveis implicações na classificação taxonômica infragenérica, foi obtido um total de 35 genomas cloroplastidiais completos, amostrando espécies dos diferentes subgêneros: *Astrophea*, *Decaloba*, *Deidamioides* e *Passiflora*. A organização dos cpDNAs mostrou-se não usual, com uma grande variação em tamanho (~ 60 kb entre o menor e o maior) e estruturas altamente rearranjadas. Além disso, ao mesmo tempo em que grandes expansões das regiões repetidas invertidas (IR) foram identificadas, o extremo oposto, a perda de uma IR foi detectada pela primeira vez em *Passiflora*, um evento raro em angiospermas. Um repertório de rearranjos, como inversões e perdas de genes, também foi constatado, tornando *Passiflora* um dos poucos grupos com uma complexa evolução de genomas cloroplastidiais. Interessantemente, um alto número de rearranjos foi detectado no subgênero *Decaloba*, no qual ocorre herança biparental de cloroplastos. Uma análise de genômica comparativa revelou diferentes estruturas de plastomas de acordo com a classificação dentro do subgênero *Deidamioides*. Além disso, a análise filogenômica baseada em genes plastidiais resultou em uma árvore de alto suporte, com posicionamento polifilético das espécies de *Deidamioides*. Combinando os resultados, sugere-se elevar o status da seção *Tryphostemmatoides* (*Deidamioides*) para subgênero *Tryphostemmatoides*. Por fim, além da contribuição deste trabalho para elucidar a história evolutiva de *Passiflora*, nossos resultados recomendam *Passiflora* como um excelente modelo para o estudo da evolução dos genomas cloroplastidiais.

Palavras-chave: Genoma cloroplastidial; Rearranjos genômicos; Filogenômica; *Passiflora*.

ABSTRACT

Chloroplast genome evolution in *Passiflora*

Chloroplast genomes (cpDNAs or plastomes) are highly conserved in Angiosperms. Although rearrangements have been observed in some lineages, the mechanisms that lead to rearrangements are still poorly elucidated. Our research group initially reported the structural organization of the plastome of *Passiflora edulis* (Passifloraceae) and rearrangments were identified. It is particularly worth noting that *Passiflora* presents different patterns of chloroplast inheritance and cytonuclear incompatibility. In order to investigate cpDNA evolutionary history in *Passiflora* and its possible implications for infrageneric taxonomic classification, a total of 35 complete chloroplast genomes were obtained, sampling for species of subgenera *Astrophea*, *Decaloba*, *Deidamioides* and *Passiflora*. The organization of the cpDNAs was uncommon, with large variation in size (~60 kb between shortest and longest), and highly rearranged genome structures. In addition, although large inverted repeat (IR) expansions were identified, the exact opposite (loss of an IR) was detected for the first time in *Passiflora*. This is indeed a rare event in angiosperms. A repertory of rearrangements, such as inversions and losses of genes was also found, making *Passiflora* one of the few groups with complex chloroplast genome evolution. Interestingly, a high number of rearrangements was detected in subgenus *Decaloba*, in which biparental chloroplast inheritance occur. A comparative genomic analysis revealed different organizational plastome structures, in accordance with the taxonomic classification within the subgenus *Deidamioides*. In addition, the plastid phylogenomics resulted in a highly supported tree with polyphyletic placement of *Deidamioides* species. Based on the combined results, we suggest elevating the status of the section *Tryphostematoides* (*Deidamioides*) to subgenus *Tryphostematoides*. Finally, apart from the contribution of this work to elucidating evolutionary history of *Passiflora*, our results also show that *Passiflora* provides a good model for the study of chloroplast genome evolution.

Keywords: Chloroplast genome; Genome rearrangements; Phylogenomics; *Passiflora*.

LIST OF FIGURES

Figure 1. *Passiflora edulis* chloroplast genome map. Genes are represented as boxes inside or outside the large circle to indicate clockwise (inside) or counter clockwise (outside) transcription. The color of the gene boxes indicates the functional group to which the gene belongs. The thickened lines of the smaller circle indicate IR (inverted repeat) regions. The inner most circle denotes the GC content across the genome. LSC: large single-copy region; SSC: small single-copy region. Intron-containing genes are marked with '*'; Pseudogenes are marked with 'Ψ'. Inv.: Inversion. 32

Figure 2. Comparison of the border positions of LSC, SSC and IR regions among chloroplast genome sequences from 12 species of the order Malpighiales. 37

Figure 3. Synteny and rearrangements detected in Malpighiales chloroplast genome sequences using the Mauve multiple-genome alignment program. A sample of 12 species is shown. Color bars indicate syntenic blocks and connecting lines indicate correspondence blocks. 38

Figure 4. The number of SSR motifs found in the *Passiflora edulis* chloroplast genome, taking into account sequence complementarities. 41

Figure 5. Phylogenetic tree of the order Malpighiales inferred from the complete nucleotide sequence of the chloroplast genome. Maximum Likelihood analysis was applied to the TVM + G model, whereas Bayesian inference (BI) was applied the GTR + G model. The bootstrap values for maximum likelihood and posterior probability for BI are indicated above each node. *Vitis vinifera* was used as outgroup to produce a rooted tree. The scale bar indicates the number of nucleotide substitutions per site. 43

Figure 6. Phylogenetic tree of the fabid clade inferred from the nucleotide sequences of cp protein coding genes of 43 plant species. Both Maximum Likelihood and Bayesian inference (BI) were applied to the GTR + G model. The bootstrap values for maximum likelihood and posterior probability for BI are indicated above each node; a dash indicates that the support nodes obtained were < 50% or < 0.50. *Vitis vinifera* was used as outgroup to produce a rooted tree. The scale bar indicates the number of nucleotide substitutions per site. Protein coding genes used in the analysis: *atpA, atpB, atpE, atpH, atpI, ccsA, cemA, matK, ndhC, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK, petA, petG, petL, petN, psaA, psaB, psaC, psaJ, psbA, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, rbcL, rpl14, rpl36, rpoB, rpoC2, rps11, rps14, rps2, rps3, rps4, rps8*. 46

Figure 7. Species of *Passiflora* collected in scientific expeditions in Brazil. A) *Passiflora loefgrenii*; B) *Passiflora capsularis*; C) *Passiflora edmundoi*; D) *Passiflora rhamnifolia*; E) *Passiflora cristalina*; F) *Passiflora contracta*. 58

Figure 8. Gene losses, IR expansions/contractions mapped onto the cladogram of the Bayesian inference with plastid genes. *Species analyzed by Shrestha et al. (Shrestha et al. 2019). The chloroplast genomes of *P. capsularis* and *P. costaricensis*, the smallest in *Passiflora* due to the loss of an inverted repeat (IR) region. Genes are represented as boxes inside or outside the large circle to indicate clockwise (inside) or counter clockwise (outside) transcription. The flower image of *P. costaricensis* was kindly provided by Maurizio Vecchia, 2005. 66

Figure 9. Synteny and structural rearrangements detected in the chloroplast genomes of 11 *Passiflora* species in addition to the Passifloraceae, *Dilkea retusa* and *Mitostemma brevifilis*,

and the Salicaceae *Populus. trichocarpa*. Colored bars indicate syntenic blocks and connecting lines indicate the correspondence between blocks. 69

Figure 10. IR boundary comparison of 11 *Passiflora* species in addition to the Passifloraceae, *Dilkea retusa* and *Mitostemma brevifilis*, and the reference species *Populus trichocarpa*. The red dotted lines indicate the border of the cp genome regions, and the colored boxes indicate the gene structure. 71

Figure 11. Bayesian phylogenetic reconstruction of *Passiflora* evolutionary history based on 68 chloroplast protein-coding genes. 74

Figure 12. Evolutionary history of chloroplast genome structure in the genus *Passiflora.* The structure of the chloroplast genomes was plotted on a tree representing the evolution of the *Passiflora* genus. In the tree, each subgenus was differentiated by colors: blue for *Astrophea*, green for *Decaloba,* red for *Deidamioides,* orange for *Passiflora,* and violet for *Tetrapathea*. In the comparison of the cp genomes, the boxes and dotted lines indicate the direction of the structural inversions, and the red arrows indicate the direction of IR expansions/contractions.*Indicates that *Mitostemma brevifilis* shares the same structure with *Dilkea retusa*, and the species of the *Passiflora* subgenus obtained in this study share the same structure with *P. alata*. [¥]Indicates the cp genomes obtained from the Genbank database. 78

Figure 13. Synteny and structural rearrangements detected in the chloroplast genomes of 6 *Passiflora* species from the subgenus *Deidamioides* according with their sections. Colored bars indicate syntenic blocks and connecting lines indicate the correspondence between blocks. 101

Figure 14. Phylogenetic reconstruction of *Passiflora* evolutionary history based on 68 chloroplast protein-coding genes. The support of the nodes was indicated by bootstrap values (BS) for Maximum likelihood and posterior probability for Bayesian inference. * indicates BS < 50. 103

Figure 15. The 35S rDNA cistron structure of *Passiflora critalina* (A). The Bayesian inference phylogenetic tree of the *Passiflora* genus based on the nuclear 18S/26S genes (B) and based on 68 chloroplast protein-coding genes (C). * indicates posterior probability < 0,8.. 105

LIST OF TABLES

Table 1. Number of species in the four subgenera of *Passiflora* genus and respective subdivisions. 15

Table 2. Gene content of the *Passiflora* edulis chloroplast genome according to respective categories. 30

Table 3. Codon usage and codon-anticodon recognition pattern of the *Passiflora edulis* chloroplast genome. 33

Table 4. Type, location and size (in bp) of repeated elements found in the *Passiflora edulis* chloroplast genome (igs, intergenic spacer; ψ, pseudogene). 39

Table 5. Summary of the 20 Passifloraceae chloroplast genomes sequenced. 64

Table 6. Summary of the short direct-repeats identified in the cp genomes of 18 *Passiflora* species and the passifloraceae, *Dilkea retusa* and *Mitostemma brevifilis*. 68

Table 7. Summary of the 14 *Passiflora* chloroplast genomes investigated. 99

LIST OF ABBREVIATIONS

BAC	Bacterial artificial chromosome
cp	Chloroplast
CNRGV	Centre national de ressources génomiques végètales
INRA	Institut National de la recherche agronomique
IR	Inverted repeat
LSC	Large single copy
SSC	Small single copy
SSR	Simple sequence repeat
BS	Bootstrap
PP	Posterior probability

1. INTRODUCTION

The Malpighiales belongs to the large group of Rosids, specially the Eurosid I clade, also known as Fabids. It is composed of approximately 16,000 species classified into 36 families, which in turn show a great morphological and ecological diversity and constitute approximately 6% of all Angiosperms (Davis et al., 2005; The Angiosperm Phylogeny Group, 2016). The family Passifloraceae belongs to the order Malpighiales, and has approximately 700 species with a Neotropical distribution, being found mainly in the Americas, but also occurring in Asia, South Africa, New Zealand and Australia (Feuillet, 2004; Ulmer & MacDougal, 2004).

Passiflora is the largest genus in the family Passifloraceae, comprising of about 530 species, popularly known as passion fruits or passionflowers (Ulmer & MacDougal, 2004). Classical taxonomy studies have subdivided the genus based on floral and vegetative attributes. The classification proposed by Killip (1938) established 22 subgenera; however, a new taxonomical classification based on morphological traits has reduced the number of subgenera to four: *Astrophea*, *Decaloba*, *Deidamioides* and *Passiflora* (Ulmer & MacDougal, 2004), with *Passiflora* being the subgenus with the largest number of species (Table 1). Recently, Pérez and D'Eeckenbrugge (2017) confirmed the subdivision into four subgenera. Although based only on qualitative morphological analyses, the authors highlighted the fragile position of the subgenus *Deidamioides*.

Table 1. Number of species in the four subgenera of *Passiflora* genus and respective subdivisions.

Subgenera	No. of species	No. of Supersections	No. of Sections	No. of Series
Passiflora	236	6	24	3
Decaloba	214	8	5	
Astrophea	57	2	5	2
Deidamioides	13		6	

Phylogenetic studies based on DNA sequences (Muschner et al., 2003, 2012; Hansen et al., 2006) also confirmed the subdivision of the genus into four subgenera, with a very early separation of *Astrophea* from the other subgenera. The following clades have been consistently recovered as monophyletic, *Astrophea*, *Decaloba* and *Passiflora*, whereas the

position of the subgenus *Deidamioides* could not be solved (Muschner et al., 2003, 2012), suggesting further analyses, for example, using phylogenomic approaches.

Passiflora genus emerged at 42.9 Mya in the Eocene period (Sader et al., 2019) the same period corresponding to that of the oldest seed fossil of Passifloraceae found in Colombia (Martínez, 2017). Regarding the *Passiflora* geographic distribution, despite its predominant distribution in neotropics, mainly in South America, 22 species from supersection *Disemma* (*Decaloba*) have been found in southeast Asia, as well as in Australia and New Zealand (Krosnick and Freudenstein, 2005).

The species of *Passiflora* present great diversity not only in vegetative structures but also in their genome sizes, for example, *P. alata* and *P. picturata* (both from the subgenus *Passiflora*) show large nuclear genomes (1C = 2,208 pg and 1C = 2,1272 pg, respectively) whereas *P. organensis* (*Decaloba*) shows a small nuclear genome (1C = 0.212 pg) (Yotoko et al., 2011). Despite this, little is known about the structural organization and composition of *Passiflora* genomes. For this purpose, our research group has built at CNRGV (INRA, Toulouse, France) a genomic library of *P. edulis* inserted into BACs (Ped-BFlav). The exploitation of this library by sequencing some 10,000 BAC-end sequences (BES) and using a comparative mapping approach (Santos et al., 2014) allowed us to generate the chloroplast (cp) genome sequence of *P. edulis*, the first cp-genome of Passifloraceae. In addition, we have detected some cp-DNA rearrangements e.g. inversions in the *P. edulis* cp-DNA molecule (Cauz-Santos et al., 2017).

Contextually, the chloroplast genome (plastome) structure can be circular or linear, ranging in size from 120 to 180 kb in Angiosperms. This genome has a quadripartite structure consisting of two copies of inverted repeats (IRa and IRb) separating two single copy regions, a large (LSC) and a small (SSC) region. Approximately 100 protein coding genes can be found in the cp-genomes, as well as tRNA and rRNA coding genes (Sugiura, 1992; Yang et al., 2010).

In different species of Angiosperms, the order of the genes in their cp-genomes is highly conserved (Jansen & Ruhlman, 2012). Although rearrangements are rare events, inversions of large segments have been identified in some species, including *Hevea brasiliensis* (Malpighiales, Euphorbiaceae), which contains a 30 kb inversion in the LSC region (Tangphatsornruang et al., 2011).

The plastome analyses have allowed scientists to describe inversion blocks, which have been used as phylogenetic markers (Cosner et al., 2004; Wu & Chaw, 2014), detect DNA duplications and deletions (Martinez-Alberola et al., 2013) and even recognize gene

transfer between the cp- and nuclear genomes (Park et al., 2015). In the present study, for the first time, it was possible to detected inversions in the cp-genome of *Passiflora edulis*, but these inversions were not verified in the available cp-genomes of phylogenetically related species (Cauz-Santos et al., 2017). This result allowed us to hypothesize that this cp-DNA organization is particular to species from *Passiflora* genus.

The analysis of the cp-DNA sequences also enables scientists to infer phylogenetic relationships and define the evolutionary course for many species from a set of chloroplast genes as, for instance, in *Musa acuminata* (Martin et al., 2013) or even using plastome complete sequences. Recently, this thematic area has been named ‘phylogenomics’ (Carbonell-Caballero et al., 2015). Using both approaches, for instance, we were able to obtain high-support phylogenetic trees that helped us to understand the evolutionary relationships within the order Malpighiales and the Fabid clade. All the results from this study were published (Cauz-Santos et al., 2017) and the scientific article is reproduced in the first Chapter of this thesis.

Therefore, we are continuing the study of the cp-genome organization in *Passiflora* by sequencing different species. Our hypothesis is that the rearrangements in the *Passiflora* cp-genomes has occurred after its diversification. In addition, we decide to investigate whether there are rearrangement specific to each subgenus, and if they could be used in the taxonomical and evolutionary studies. We intend not only to provide important information on the evolution of the genus, but also suggest *Passiflora* as an excellent model for studying the evolution of chloroplast genomes.

This present doctoral thesis is structured in four chapters. This first introduces readers about the genus *Passiflora* and chloroplast genomes. The second chapter deals with the chloroplast genome structure of *Passiflora edulis*, unveiling for the first time the rearrangements observed in a Passifloraceae plastome. The chloroplast genome of different species of the four subgenera of *Passiflora* was analyzed in the third chapter, revealing highly rearranged plastomes and the loss of an IR region in two species of the *Decaloba* subgenus. Finally, in the fourth chapter, the evolutionary relationships in *Passiflora* were studied using plastid phylogenomics, which enabled us to suggest a new infrageneric classification.

REFERENCES

Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., and Dopazo, J. (2015). A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the Genus *Citrus*. *Molecular Biology Evolution*, 32, 2015–2035.

Cauz-Santos, L. A., Munhoz, C. F., Rodde, N., et al. (2017). The chloroplast genome of *Passiflora edulis* (Passifloraceae) using sequencing and assembly approaches from long sequence reads: structural organization and phylogenomic studies in Malpighiales. *Frontiers in Plant Science*, 10, 334.

Cosner, M. E., Raubeson, L. A., Jansen, R. K. (2004). Chloroplast DNA reattangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evolutionary Biology*, 4, 27.

Davis, C. C., Campbell, O., Kenneth, J., et al. (2005). Explosive radiation of Malpighiales supports a mid-Cretaceous origin of modern tropical raiforests. *American Naturalist* 165, 36-65.

Feuillet, C. (2004). “Passifloraceae (Passion flower family),” in *Flowering plants of the Neotropics*, eds. N. Smith, S. A. Mori, A. Henderson, D. W. Stevenson, and S. D. Heald (Oxford: Princeton University Press & New York Botanical Garden), 286–287.

Hansen, A. K., Gilbert, L. E., Simpson, B. B., et al. (2006). Phylogenetic Relationships and chromosome number evolution in *Passiflora*. *Systematic Botany*, 31, 138–150.

Jansen, R. K., and Ruhlman, T. A. (2012). “Plastid genomes of seed plants” in *Genomics of chloroplasts and mitochondria* (Springer Netherlands), p. 103–126.

Killip, E. P. (1938). The American species of Passifloraceae. Chicago: Chicago Field Museum.

Krosnick, S.E., Freudenstein, J. V., 2005. Monophyly and floral character homology of Old world *Passiflora* (subgenus *Decaloba*: supersection *Disemma*) . Syst. Bot. https://doi.org/10.1600/0363644053661959

Martin, G., Baurens, F.-C., Cardi, C., Aury, J.-M., and D’Hont, A. (2013). The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): Insight into plastid monocotyledon evolution. *PLoS ONE* 8, e67350.

Martinez-Alberola, F., del Campo, E. M., Lazaro-Gimeno, D., et al. (2013). Balanced gene losses, duplications and intensive rearrangements led to an unusual regularly sized genome in *Arbutus unedo* chloroplasts. *PLoS ONE*, 8, 11.

Martínez, C., 2017. Passifloraceae seeds from the late eocene of Colombia. Am. J. Bot. https://doi.org/10.3732/ajb.1700224

Muschner, V. C., Lorenz, A. P., Cervi, A. C., et al. (2003). A first molecular phylogenetic analysis of *Passiflora* (Passifloraceae). *American Journal of Botany*, 90, 1229–1238.

Muschner, V. C., Zamberlan, P. M., Bonatto, S. L., Freitas, L. B. (2012). Phylogeny, biogeography and divergence times in *Passiflora* (Passifloraceae). *Genetics and Molecular Biology*, 35, 1036–1043.

Park, A., Jansen, E. K., Park, S. (2015). Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictroideae. *BMC Plant Biology*.

Pérez, J. O., D'Eeckenbrugge, G. C. (2017). Morphological characterization in the genus *Passiflora* L.: an approach to understanding its complex variability. *Plant Systematics and Evolution*, 303, 531–558.

Sader, M.A., Amorim, B.S., Costa, L., Souza, G., Pedrosa-Harand, A., 2019. The role of chromosome changes in the diversification of *Passiflora* L. (Passifloraceae). Syst. Biodivers. https://doi.org/10.1080/14772000.2018.1546777

Santos, A., Penha, H., Bellec, A., Munhoz, C. F., Pedrosa-Harand, A., Bergès, H., et al. (2014). Begin at the beginning: A BAC-end view of the passion fruit (*Passiflora*) genome. *BMC Genomics* 15, 816.

Sugiura, M. (1992). The chloroplast genome. *Plant Molecular Biology* 19, 149–168.

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

The Angiosperm Phylogeny Group (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of Linnean Society* 181, 1–20.

Ulmer, T. and MacDougal, J. M. (2004). *Passiflora: Passionflowers of the world*. Portland: Timber Press.

Wu, C. S. and Chaw, S. M. (2014). Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. *Plant Biotechnology Journal*, 12, 344-353.

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., et al. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE*, 5, e12762.

Yotoko, K. S. C., Dornelas, M. C., Togni, P. D., Fonseca, T. C., Salzano, F. M., Bonatto, S. L., et al. (2011). Does variation in genome sizes reflect adaptive or neutral processes? New clues from *Passiflora*. *PLoS ONE*, 6, e18212.

2. THE CHLOROPLAST GENOME OF *PASSIFLORA EDULIS* (PASSIFLORACEAE) USING SEQUENCING AND ASSEMBLY APPROACHES FROM LONG SEQUENCE READS: STRUCTURAL ORGANIZATION AND PHYLOGENOMIC STUDIES IN MALPIGHIALES[1]

ABSTRACT

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

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

2.1. Introduction

Malpighiales is an order of flowering plants that belongs to the clade Eurosids I, also known as Fabids (The Angiosperm Phylogeny Group, 2009). This large order includes 42 families, more than 700 genera, and contains approximately 16,000 species forming an extremely diverse group of plants in terms of their morphological and ecological aspects (The Angiosperm Phylogeny Group, 2009; Wurdack and Davis, 2009). The Passifloraceae family is a member of the Malpighiales (Judd et al., 2008) and consists of some 700 species of herbaceous or woody vines, shrubs and trees, classified in around 16 genera, and almost all its

---

[1] The text corresponds to that of the article published by Cauz-Santos et al. (2017) in *Frontiers in Plant Science*, 10, 334.

members belong to the large and variable genus *Passiflora*, popularly known as passion flowers or passion fruits (Feuillet, 2004). There are around 520 species of *Passiflora*, the majority of which are distributed pantropically; the most derived and specialized species are distributed in the Neotropics and Africa (Ulmer and MacDougal, 2004).

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

In Brazil, passion fruit cultivation began relatively recently and has earned the country an outstanding position as the world's top producer of yellow passion fruit (*Passiflora edulis*). It is an outcrossing (Ferreira et al., 2010), diploid (n = 9), (Cuco et al., 2005) species with perfect self-incompatibility (Bruckner et al., 1995; Rêgo et al., 2000), and insect-pollinated flowers. Genetic (Moraes et al., 2005; Oliveira et al., 2008) and molecular-based studies have been carried out in our laboratory (Munhoz et al., 2015; Santos et al., 2014), which is able to satisfy the needs of a wide range of breeders to boost passion fruit crop production and fruit quality.

The nuclear genome size of *P. edulis* (expressed in 1C) was estimated at 1.258 (Yotoko et al., 2011) or 1,595 pg (Souza et al., 2004). More recently, a very efficient strategy for obtaining initial insight into the content of this particular genome involved the sequencing of the terminal regions (BAC-ends) of a representative number of BAC clones selected at random from a *Passiflora edulis* genomic library (Santos et al., 2014). This library was constructed by and deposited at INRA-CNRGV and covers around six times the genome length of *P. edulis* (http://cnrgv.toulouse.inra.fr/Library/Passiflora). Our group was able to characterize some 10,000 high-quality sequences (100 to 1,255 bp) and identify reads likely to contain repetitive mobile elements and simple sequence repeats, and to estimate the GC-content of the reads. Approximately one tenth of the BAC end-sequences contained protein sequences, and gene ontology terms were assigned to most of them. Finally, we were able to map a number of BAC-end pair sequences to intervals of *Arabidopsis thaliana*, *Vitis vinifera* and chiefly to *Populus trichocarpa* chromosomes, representing regions of potential microsynteny. Additionally, a number of BAC clones were identified as containing chloroplast DNA sequences (Santos et al., 2014).

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

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

The development of next generation sequencing (NGS) technologies has provided highly efficient, low-cost DNA sequencing platforms that produce large volumes of short reads (Shin et al., 2013). However, more recently, third generation sequencing technologies producing longer DNA reads have begun to emerge, including the Pacific Biosciences Single Molecule Real-Time (SMRT) sequencer that became available in 2011 (http://www.pacificbiosciences.com/). For example, using PacBio sequence data it was possible to generate an entire chloroplast genome assembled into a single large contig with a high degree of accuracy and at a much greater depth of coverage due to longer read lengths (Ferrarini et al., 2013).

In this study, we present the complete nucleotide sequence and the organization of the chloroplast genome of *Passiflora edulis*, the first report on the family Passifloraceae. We were able to localize genes, introns and intergenic spacers, as well as repetitive elements, and

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

2.2. Material and Methods

2.2.1. Plant material

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

2.2.2. Sequencing and subsequent assembly of *Passiflora edulis* chloroplast DNA using the Pacbio RS II platform

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

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

Biosciences PacBio RS II platform. Sequencing was performed by GATC Biotech (http://www.gatc-biotech.com).

The reads were assembled following a hierarchical genome assembly process (HGAP workflow (Chin et al., 2013), and using the SMRT® Analysis (v2.2.0) software suite for HGAP implementation. Reads were first aligned by the PacBio long read aligner or BLASR (Chaisson and Tesler, 2012) against the complete genome of *Escherichia coli* strain K12 substrain DH10B (GenBank: CP000948.1). The *E. coli* reads, as well as low quality reads (minimum read length of 500 bp and minimum read quality of 0.80), were removed from the data set. Filtered reads were then preassembled to yield long, highly accurate sequences. To perform this step, smallest and longest reads were separated from each other to correct read errors by mapping single-pass reads to longest reads (seed reads), which represent the longest portion of the read length distribution. Next, the sequences were filtered against vector (BAC) sequences, and the Celera assembler was used to assemble data and obtain draft assemblies. The last step of the HGAP workflow is performed in order to significantly reduce the remaining InDel and base substitution errors in the draft assembly. The Quiver algorithm was used for this purpose. It is a quality-aware consensus algorithm that uses rich quality scores (QV scores) embedded in Pacific Biosciences' bas.h5 files. Once the polished assembly was obtained, each BAC sequence was individualized by matching its paired-end sequences to the assembled sequences using BLAST. Read coverage was assessed by aligning the raw reads on the assembled sequences with BLASR.

2.2.3. Obtaining the complete chloroplast genome

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

CA, USA). The PCR product was purified using the Wizard® SV Gel kit and PCR Clean-Up System (Promega), and then used as a template for the sequencing reaction based on the Sanger method. It was subjected to capillary electrophoresis in the ABI Prism 3100 sequencer (Applied Biosystems). The sequence of the PCR product was aligned with the single contig to obtain the complete sequence of the cp genome.

2.2.4. Genome annotation

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

2.2.5. Comparative analysis of chloroplast genomes

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

In addition, a multiple alignment with all available entirely sequenced cp genomes of Malpighiales species (12 in the total), including *P. edulis*, was run in progressive Mauve v.2.4.0 (Darling, 2004). Briefly, this method identifies conserved genomic regions,

rearrangements and inversions in conserved regions, and the sequence breakpoints of these rearrangements across multiple genomes.

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

2.2.6. Identification of repeated elements

REPuter (Kurtz et al., 2001) was used to identify direct and palindromic repeated elements, based on the following criteria: minimum repeat size ≥ 30 bp and sequence identities ≥ 90% (Hamming distance equal to 3). Simple sequence repeats (SSRs) were predicted using MISA (MIcroSAtellite, http://pgrc.ipk-gatersleben.de/misa/). The criteria for SSRs search were set as follows: minimum repeat number defined as ten, five and four units for mono-, di-, and trinucleotide SSRs, respectively, and three units for each tetra-, penta-, and hexanucleotide SSR.

2.2.7. Phylogenomic studies

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

Entire cloroplastidial genomes of Malpighiales representing the families Passifloraceae, Salicaceae, Euphorbiaceae and Chrysobalanaceae were used as the ingroup in phylogenomic comparisons, and the cp genome of *Vitis vinifera* (Vitaceae, Vitales) was used as the outgroup in order to root the phylogenetic tree. The data set consisting of 23 taxa was aligned at nucleotide level in server-based program MAFFT version 7.221, using the FFT-NS-2 algorithm with default settings. To generate the alignment, inverted sequences detected in the cp genomes of *P. edulis* and *H. brasiliensis* were reversed and the respective positions adjusted. The raw alignment was manually corrected in BioEdit (Hall, 1999) and further

processed in GBLOCKS 0.91b (Castresana, 2000) in order to remove regions containing gap positions, with a minimum block length of five, and maximum number of contiguous non-conserved positions of eight. The resulting alignment was analysed in jModelTest software version 2.1.7 (Darriba et al., 2012) to determine the optimal model of molecular evolution and gamma rate heterogeneity using the Akaike Information Criterion (AIC).

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

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

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

jModelTest software version 2.1.7 (Darriba et al., 2012) was used to determine the optimal model of molecular evolution and gamma rate heterogeneity based on AIC. Both ML and BA were used to infer the phylogenomic relationships within the Fabid clade. ML analysis was performed using RAxML version 8.2.4 (Stamatakis, 2014). The general time

reversible (GTR) model of nucleotide substitution was selected for ML analysis, taking into account the gamma distribution of rate heterogeneity with five discrete categories (+G). To estimate the support of the ML topology, a bootstrap analysis was performed on 1,000 replicates. BI was run on MrBayes software, version 3.1.2 (Ronquist and Huelsenbeck, 2003) with the GTR +G model. The Markov Chain Monte Carlo (MCMC) algorithm ran for 5,000,000 generations, sampling one tree every 100 generations. The first 25% of trees were discarded as burn-in to estimate the values of posterior probabilities. Convergence diagnostics were monitored on the basis an average standard deviation of split frequencies below 0.01, potential scale reduction factor (PSRF) values close to 1.0, and ESS values above 200. Trees were visualized using FigTree version 1.4.0.

2.3. Results and Discussion

2.3.1. Data output from the PacBio RS II platform and assembly of chloroplast genome sequences

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

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

which was subsequently sequenced and aligned with the long contig sequence. Thus, a 44-nucleotide sequence was obtained and added for closing the gap to produce the complete sequence of the circular molecule of 151,406 bp. The chloroplast genome sequence of *Passiflora edulis* has been deposited to the NCBI GenBank and has received the accession number NC_034285.

The chloroplast genomes of *Potentilla micranta* (Ferrarini et al., 2013) and *Aconitum barbatum var. puberulum* (Chen et al., 2015) were entirely sequenced using the PacBio RS II Platform, and both studies described the advantages of using this method. For instance, Ferrarini et al., 2013 highlight the generation of a single contig covering the entire cp genome of *P. micranta*, whereas Illumina HiSeq2000 sequencing data have lower genome coverage (some regions have zero or very low coverage) and the resulting assembly consisted of seven contigs. Similarly, Chen et al. (2015) emphasize the low level of errors (~0.0027%) in PacBio reads, since after applying the necessary filter, the result is an error-corrected consensus read with a higher intra-molecular accuracy. Interestingly, even in the absence of any other biological information on the target species, the authors stated that it took less than half an hour to finish the genome assembly step, confirming that long read lengths are superior, especially for de novo assemblies.

2.3.2. Organization of the *P. edulis* chloroplast genome

Chloroplast genome annotation resulted in the identification of 105 unique genes, including 30 tRNAs, 4 rRNAs and 71 protein coding genes (Table 2). The molecule has a typical quadripartite structure characterized by the presence of two copies of an inverted repeat sequence (IRA and IRB) each of 26,154 bp (34.6% in total) separating a small single-copy (SSC) region of 13,378 bp (8.8%) and a large single-copy (LSC) region of 85,720 bp (56.6%). Genes arising from duplication events created perfect IRs, each containing the same 16 genes, totalling 120 genes in the complete chloroplast DNA molecule (Figure 1).

Table 2. Gene content of the *Passiflora edulis* chloroplast genome according to respective categories.

Category	Gene
Subunits of Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
Subunits of Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
Subunits of cytochrome b/f complex	*petA*, *petB*[a], *petD*[a], *petG*, *petL*, *petN*

Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*
Large subunit of Rubisco	*rbcL*
Subunits of NADH-dehydrogenase	*ndhA*[a], *ndhB*[a,b], *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
Proteins of large ribosomal subunit	*rpl2*[a,b], *rpl14*, *rpl16* [a], *rpl23*[b], *rpl32*, *rpl33*, *rpl36*
Proteins of small ribosomal subunit	*rps2*, *rps3*, *rps4*, *rps8*, *rps11*, *rps12*[a,b,c], *rps14*, rps15, *rps18*, *rps19*[b]
Subunits of RNA polymerase	*rpoA*, *rpoB*, *rpoC1*[a], *rpoC2*
Cytochrome c biogenesis	*ccsA*
Maturase	*matk*
Protease	*clpP*
Envelope membrane protein	*cemA*
Conserved hypothetical genes	*ycf3*[a], *ycf4*
Ribosomal RNAs	*rrn4.5*[b], *rrn5*[b], *rrn16*[b], *rrn23*[b]
Transfer RNAs	*trnA-UGC*[a,b], *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-UCC*[a], *trnG-GCC*, *trnH-GUG*, *trnI-CAU*[b], *trnI-GAU*[a,b], *trnK-UUU* [a], *trnL-CAA*[b], *trnL-UAA* [a], *trnL-UAG*, *trnM-CAU*, *trnN-GUU*[b], *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*[b], *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*[b], *trnV-UAC* [a], *trnW-CCA*, *trnY-GUA*

[a]Intron-containing gene; [b]Two gene copies in the IRs; CGene divided into two independent transcription units.

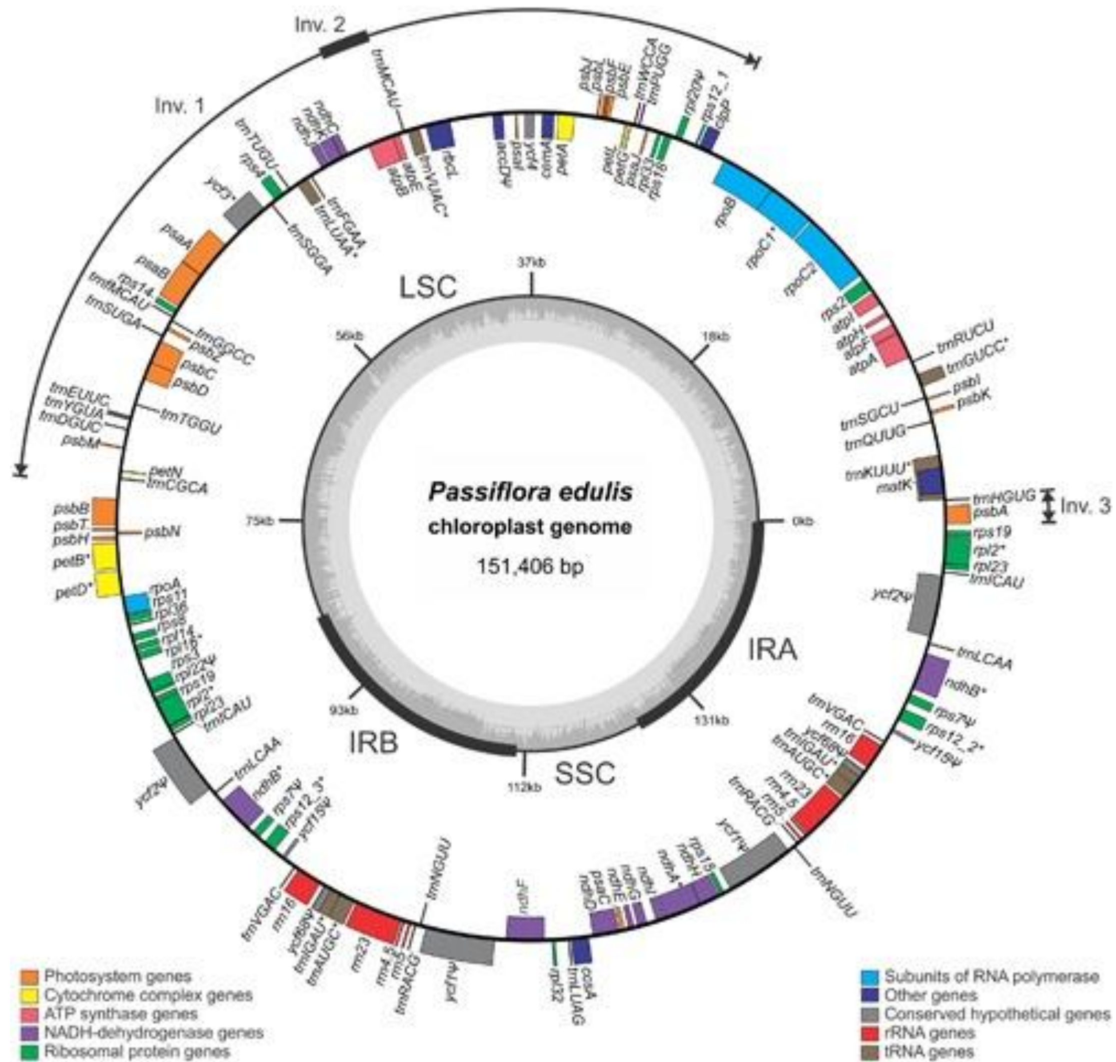


Figure 1. *Passiflora edulis* chloroplast genome map. Genes are represented as boxes inside or outside the large circle to indicate clockwise (inside) or counter clockwise (outside) transcription. The color of the gene boxes indicates the functional group to which the gene belongs. The thickened lines of the smaller circle indicate IR (inverted repeat) regions. The inner most circle denotes the GC content across the genome. LSC: large single-copy region; SSC: small single-copy region. Intron-containing genes are marked with '*'; Pseudogenes are marked with 'Ψ'. Inv.: Inversion.

Protein coding genes constitute 37.3% of the chloroplast genome (75 genes, totaling 56,428 bp), in addition to tRNA and rRNA coding genes that constitute 1.8% (37 genes totaling 2,801 bp) and 6.0% (8 genes totalling 9,048 bp) respectively of the chloroplast genome. Introns constitute 10.7% (totaling 16,155 bp), while intergenic regions and pseudogenes form the remaining 44.2% (totaling 66,974 bp). The amount of GC was estimated at 37%.

tRNA genes are spread throughout the genome and encode the 20 amino acids incorporated into proteins. One is present in the SSC region, 21 in the LSC, and 7 in the IRs,

which also have 5 protein coding genes and 4 rRNA genes. A total of 18,799 codons represent the coding capacity of the 71 protein coding genes (Table 3). The most frequent amino acid was found to be leucine (2,026 codons corresponding to 10.8% of total DNA) and the least frequent cysteine (218 codons corresponding to 1.2% of total DNA).

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

Table 3. Codon usage and Codon-anticodon recognition pattern of the *Passiflora edulis* chloroplast genome.

Codon	Amino acid	Number	RSCU [a]	% [b]	tRNA	Codon	Amino acid	Number	RSCU [a]	% [b]	tRNA
GCU	A	513	1.727	43.2		CCA	P	225	1.108	27.7	*trnP-UGG*
GCG	A	150	0.505	12.6		CCC	P	158	0.778	19.5	
GCC	A	206	0.694	17.3		CCU	P	325	1.601	40.0	
GCA	A	319	1.074	26.9	*trnA-UGC* [c]	CCG	P	104	0.512	12.8	
UGU	C	162	1.486	74.3		CAA	Q	523	1.59	79.5	*trnQ-UUG*
UGC	C	56	0.514	25.7	*trnC-GCA*	CAG	Q	135	0.41	20.5	
GAU	D	524	1.53	76.5		AGA	R	302	1.529	28.5	*trnR-UCU*
GAC	D	161	0.47	23.5	*trnD-GUC*	AGG	R	93	0.471	8.8	
GAG	E	202	0.441	22.0		CGA	R	250	1.508	23.6	
GAA	E	715	1.559	78.0	*trnE-UUC*	CGC	R	80	0.483	7.6	
UUU	F	716	1.351	67.5		CGG	R	88	0.531	8.3	
UUC	F	344	0.649	32.5	*trnF-GAA*	CGU	R	245	1.478	23.2	*trnR-ACG*
GGU	G	471	1.323	33.1		AGC	S	107	0.542	7.8	*trnS-GCU*
GGG	G	233	0.654	16.4		AGU	S	288	1.458	21.1	
GGC	G	169	0.475	11.9		UCA	S	256	1.053	18.7	*trnS-UGA*
GGA	G	551	1.548	38.7	*trnG-GCC*	UCC	S	223	0.918	16.3	*trnS-GGA*
CAC	H	102	0.462	23.1	*trnH-GUG*	UCG	S	114	0.469	8.3	
CAU	H	340	1.538	76.9		UCU	S	379	1.56	27.7	
AUU	I	834	1.515	50.5		ACC	T	170	0.688	17.2	*trnT-GGU*
AUA	I	506	0.919	30.6	*trnI-CAU*	ACA	T	308	1.247	31.2	*trnT-UGU*
AUC	I	311	0.565	18.8	*trnI-GAU* [c]	ACG	T	97	0.393	9.8	
AAA	K	704	1.529	76.4	*trnK-UUU* [c]	ACU	T	413	1.672	41.8	

AAG	K	217	0.471	23.6		GUU	V	388	1.399	35.0	
CUA	L	278	1.214	13.7	*trnL-UAG*	GUG	V	145	0.523	13.1	
CUC	L	120	0.524	5.9		GUC	V	139	0.501	12.5	*trnV-GAC*
CUG	L	97	0.424	4.8		GUA	V	437	1.576	39.4	*trnV-UAC*[c]
CUU	L	421	1.838	20.8		UGG	W	327	1	100	*trnW-CCA*
UUA	L	735	1.324	36.3	*trnL-UAA*[c]	UAC	Y	124	0.363	18.2	*trnY-GUA*
UUG	L	375	0.676	18.5	*trnL-CAA*	UAU	Y	559	1.637	81.8	
AUG	M	464	1	100	*trn(f)M-CAU*	UGA	*	15	0.584	19.5	
AAC	N	186	0.464	23.2	*trnN-GUU*	UAG	*	24	0.935	31.2	
AAU	N	615	1.536	76.8		UAA	*	38	1.481	49.4	

[a]Relative Synonymous Codon Usage
[b]Codon frequency (in %) per amino acid
[c]Intron-containing tRNA genes
*Stop codon

The *matK* gene was found in the intronic region of the *trnK-UUU* gene in *Populus trichocarpa* (Tuskan et al., 2006), *H. brasiliensis* (Tangphatsornruang et al., 2011) and *Manihot esculenta* (Daniell et al., 2008). This gene encodes the unique maturase found in the plastid genomes of land plant species, but it does not contain the reverse transcriptase domain and therefore is not able to promote intron mobility. The *clpP* gene has no intron in *P. edulis*, although it does have two introns in the following conordinal-related species: *P. trichocarpa* (Tuskan et al., 2006), *S. purpurea* (Wu, 2015), *H. brasiliensis* (Tangphatsornruang et al., 2011), *M. esculenta* (Daniell et al., 2008), *Jatropha curcas* (Asif et al., 2010), and *R. communis* (Rivarola et al., 2011). Similar intron losses have already been reported in legumes, almost exclusively in the clade known as the IR-lacking clade (IRLC) (Jansen et al., 2008). Coincidently, the *clpP* gene is located at the beginning of first the inversion found in the *P. edulis* cp genome (Figure 1), which may lead to intron losses.

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

Two protein coding genes show alternative initiation codons. The 'GTG' triplet was found in the *rps19* and *ndhD* genes. This triplet was also found in the *rps19* gene of *J. curcas* (Asif et al., 2010), *H. brasiliensis* (Tangphatsornruang et al., 2011) and *Cynara cardunculus* (Curci et al., 2015). The *rps12* gene is a trans-spliced gene consisting of three exons: the first exon (5'-*rps12*) is located in the LSC region, far from the other two located in the IRs. This kind of organization was observed in *P. trichocarpa*, *S. purpurea*, *H. brasiliensis*, *M. esculenta*, *J. curcas* and *R. communis*. Interestingly the *infA* gene, which codes for translation initiation factor 1, and the *rps16* gene, which codes for a S16 ribosomal protein are absent in the *P. edulis* cp genome, as previously reported in other species of *Passiflora* (Jansen et al., 2007). These genes were also lost or are non-functional in related species *M. esculenta* (Daniell et al., 2008), *J. curcas* (Asif et al., 2010) and in seven species of Salicaceae (Wu, 2015). It is worth noting that Passifloraceae are known to have chloroplast gene and intron losses (Daniell et al., 2008; Hansen et al., 2006; Jansen et al., 2007).

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

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

2.3.3. Comparative analysis of chloroplast genomes

As commonly done in studies of border expansion in the inverted repeats region, the IR-LSC and IR-SSC boundaries with full annotations for the adjacent genes were reexamined

across 11 sequenced species closely related to *P. edulis*. Note that the *rps15* gene of *P. edulis* is located at the end of the SSC region, expanding 20 bp towards the IRA region. It is substituted in other species by the *ycf1* gene, i.e. this gene spans SSC and IRA and there is a pseudogene copy in the IRB region. However, in *P. edulis*, both copies of the *ycf1* gene are of the same size and are full in IR regions (Figure 2). In addition, the *psbA* gene is the first in the LSC region due to a small inversion that is further discussed below.

In accordance with the genome comparison results and based on the 22 members of Malpighiales with available chloroplast genomes, *P. edulis* shows a large inversion of 46,151 bp in the LSC region, between the genes *clpP* and *trnC-GCA*, and a second smaller inversion of 3,765 bp between the genes *trnM-CAU* and *atpB*, located in the medial region of the first inversion. A third inversion of 1,631 bp is located at the beginning of the LSC region containing the *psbA* and *trnH-GUG* genes (Figure 1 and 3). All these genomic features are shown in the sequence alignment results for 12 species representing each genus of the Passifloraceae, Salicaceae, Euphorbiaceae and Chrysobalanaceae families (Figure 3).

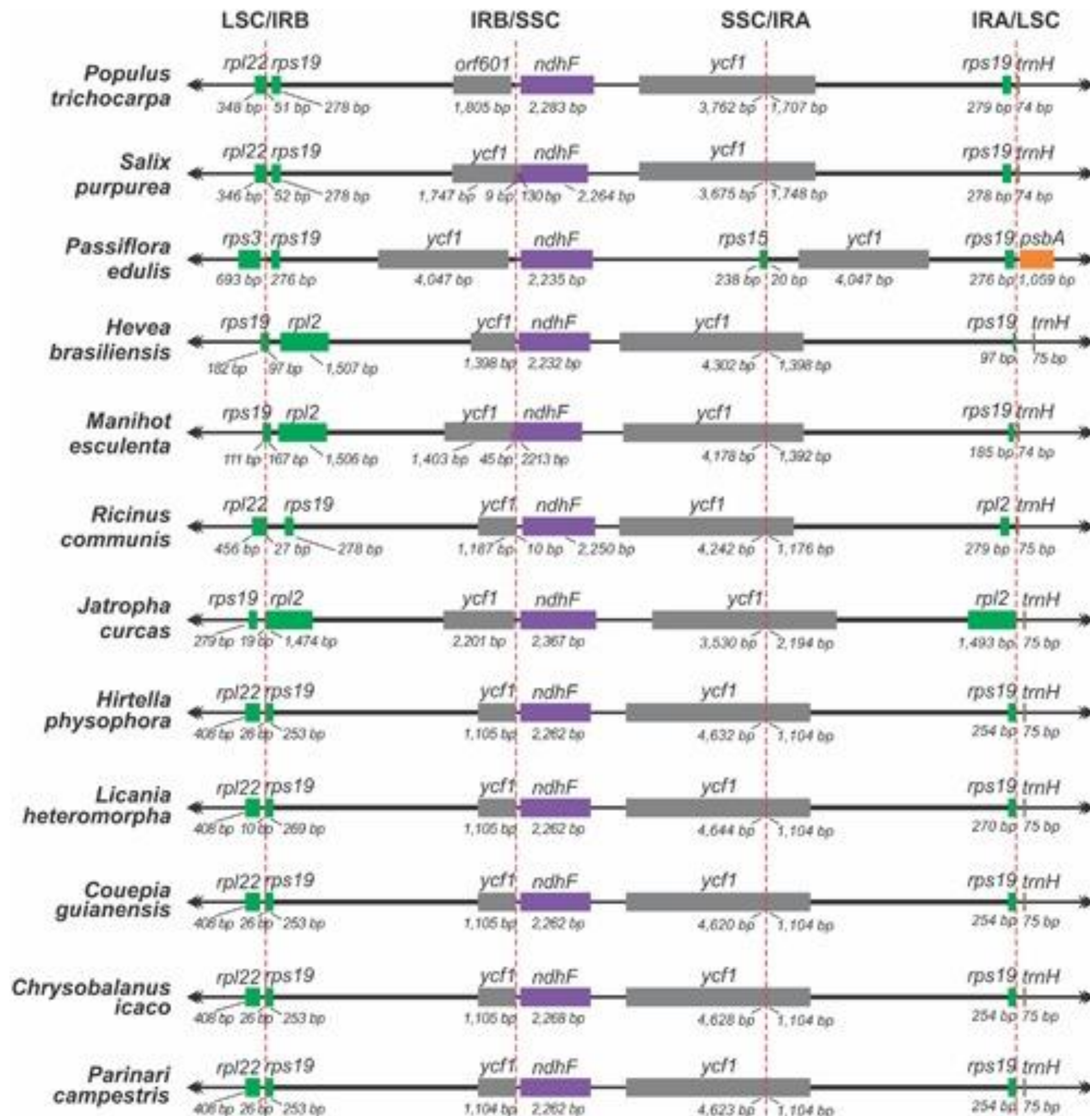


Figure 2. Comparison of the border positions of LSC, SSC and IR regions among chloroplast genome sequences from 12 species of the order Malpighiales.

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

same plant family, as occurs in Salicaceae (Wu, 2015) and Chrysobalanaceae (Malé et al., 2014), also members of the Malpighiales order. The existence of rearrangements in segments of cp genomes may be useful as phylogenetic markers within genera or even within families, becoming a potential tool for understanding the evolution of plant species. Therefore, the complete sequencing of new chloroplast genomes will allow a higher accuracy in evolutionary studies of the inversions in the genus *Passiflora*.

Figure 3. Synteny and rearrangements detected in Malpighiales chloroplast genome sequences using the Mauve multiple-genome alignment program. A sample of 12 species is shown. Color bars indicate syntenic blocks and connecting lines indicate correspondence blocks.

2.3.4. Analysis of repeated elements

We were able to identify 36 repetitive elements, all in the LSC region. These repeats were found predominantly in intergenic regions. However, each of the members of a particular repeat was identified in the coding sequences of the *psaA* and *psaB* genes respectively; one member of other repeat is located at the beginning of the *psbI* gene sequence and the other in an intergenic region. No introns were found to contain repeated elements, and approximately 60% of repetitive elements are within a 2,513-bp region between the *accD* pseudogene and the *rbcL* gene, indicating that these elements are distributed in a peculiar arrangement in *P. edulis*. The abundance of repeated elements in this region might possibly have changed the nucleotide sequence of the *accD* pseudogene, rendering it non-functional. The repeat unit length ranged from 34 to 178 bp, and each repeat showed two copies. Forward (or direct) repeats and forward-tandem repeats (when the repeats are presented immediately one beside the other, there occurring sometimes an overlapping of both repeat units) were predominant, and only three were found to be palindromes (or reverse-complemented) (Table 4). Repetitive sequences are substrates for recombination and cp genome rearrangements (Milligan et al., 1989), and the number and distribution of these sequences vary from one species to another. For instance, the cp genome of *J. curcas* contains 72 repetitive sequences that are distributed in intergenic regions, introns and coding sequences (Asif et al., 2010), whereas in *V. vinifera*, 36 repetitive sequences were found, one in the coding regions of the *psaA* and *psaB* genes (Jansen et al., 2006), as in *P. edulis*, but 58% were palindromes and 12 exist in the *ycf2* gene.

Table 4. Type, location and size (in bp) of repeated elements found in the *Passiflora edulis* chloroplast genome (IGS, Intergenic spacer; Ψ, pseudogene).

Type	Location	Size (in bp)
Forward-tandem	IGS: *matK-psbK*	43
Forward	IGS: *psbK-psbI*, *psbI*	49
Palindrome	IGS: *psbI-atpA*	47
Forward-tandem	IGS: *psbI-atpA*	44
Palindrome	IGS: *psbI-atpA*	34
Forward	IGS: *rpl33-psaJ*, *psaI-accD*	47
Forward	IGS: *accD* Ψ-*rbcL*	66
Forward	IGS: *accD* Ψ-*rbcL*	58
Forward	IGS: *accD* Ψ-*rbcL*	54
Forward-tandem	IGS: *accD* Ψ-*rbcL*	41
Forward-tandem	IGS: *accD* Ψ-*rbcL*	73

Forward	IGS: *accD* Ψ-*rbcL*	61
Forward-tandem	IGS: *accD* Ψ-*rbcL*	74
Forward	IGS: *accD* Ψ-*rbcL*	39
Forward	IGS: *accD* Ψ-*rbcL*	46
Forward	IGS: *accD* Ψ-*rbcL*	56
Forward	IGS: *accD* Ψ-*rbcL*	49
Forward	IGS: *accD* Ψ-*rbcL*	67
Forward	IGS: *accD* Ψ-*rbcL*	46
Forward	IGS: *accD* Ψ-*rbcL*	65
Forward	IGS: *accD* Ψ-*rbcL*	46
Forward	IGS: *accD* Ψ-*rbcL*	43
Forward-tandem	IGS: *accD* Ψ-*rbcL*	58
Forward-tandem	IGS: *accD* Ψ-*rbcL*	124
Forward-tandem	IGS: *accD* Ψ-*rbcL*	99
Forward-tandem	IGS: *accD* Ψ-*rbcL*	74
Forward	IGS: *accD* Ψ-*rbcL*	49
Palindrome	IGS: *rbcL-atpE*	62
Forward-tandem	IGS: *ycf3-psaA*	41
Forward-tandem	IGS: *ycf3-psaA*	57
Forward	*psaA*, *psaB*	79
Forward-tandem	IGS: *psaB-rps14*	44
Forward-tandem	IGS: *petN-psbB*	71
Forward-tandem	IGS: *petN-psbB*	178
Forward	IGS: *petN-psbB*	117
Forward	IGS: *petN-psbB*	56

We found 85 SSRs (microsatellites), ranging in size from 10 to 106 bp, 67 of which (79%) consisted exclusively of A/T. In terms of motifs, we found 43 (50.6%) mono-, 21 (24.7%) di-, 7 (8.2%) tri-, 10 (11.8%) tetra-, and 4 (4.7%) hexanucleotides (Figure 4). SSRs were found mainly in non-coding regions, 68 in intergenic regions and five in pseudogenes/intronic regions. Of the 12 SSRs found in gene sequences, two interrupted elements composed of trinucleotides (GAC/ACG) were found in the *clpP* gene. In addition, the *ndhA* gene contained one mono-(A) and one tetranucleotide (AAAT), the latter also seen in the *ndhD* gene sequence. Genes *ndhF*, *rpoB*, and *rpoC1* contained one mononucleotide each (T or A), whereas *rpoC2* contained four SSRs, three mono- (T), and one dinucleotide (AT).

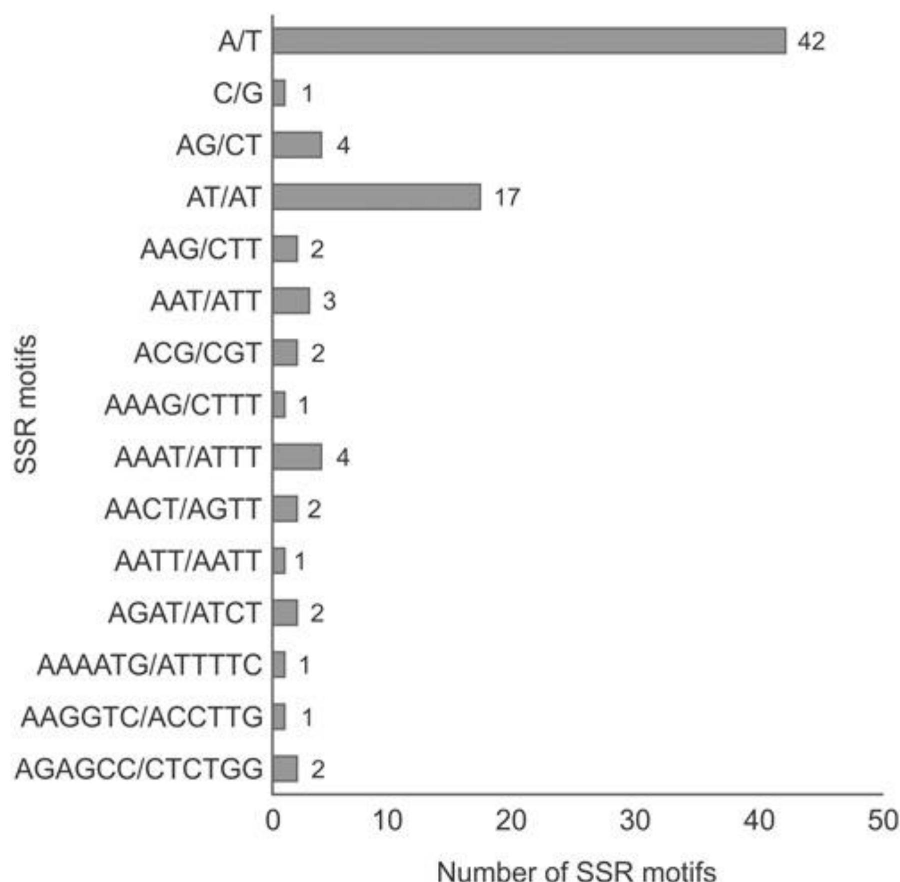


Figure 4. The number of SSR motifs found in the *Passiflora edulis* chloroplast genome, taking into account sequence complementarities.

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

2.3.5. Phylogenomic studies

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

Both Maximum Likelihod (ML) and Bayesian inference (BI)-based methods resulted in a similar tree (Figure 5), and all species were clustered in Malpighiales, in conformity with the APG II system. The nodes resulting from ML analysis were consistently supported in the tree. The bootstrap values at most of the nodes (15 of 20 nodes) were ≥ 85%. In the Bayesian inference the value of the average standard deviation of split frequencies was 0.0001 after 5,000,000 generations. The PSRF values were close to 1.0 (ranging between 0.999 and 1.000), and the ESS values were above 200 (ranging between 4,909 and 29,555) for all the parameters. All these values indicated that the analysis has reached convergence. The nodes resulting from BI analysis were also highly supported (PP = 1 for 17 nodes). *P. edulis* has been placed near the Salicaceae (*Salix* and *Populus* species) with strong support (BS = 100 %; PP = 1), but distant from members of the Chrysobalanaceae, that in turn are closer to the Euphorbiaceae members studied herein.

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

The Salicaceae species were clustered into two clades, both highly supported (BS = 100 %; PP = 1). The most related were *Salix purpurea* and *S. suchowensis* (BS = 100 %; PP = 1), but positioned at some distance from those species is *S. interior* (BS = 100 %, PP =1). *Populus* species were split into two groups, the first incorporating *Populus alba*, *P. tremula* and *P. euphratica* (BS = 78 %; PP = 1), and the second *P. balsamifera*, *P. trichocarpa* and *P. fremontii* (BS = 100 %; PP = 1). The placement of *Salix* and *Populus* species is similar to that recently shown in the phylogenetic tree built from the cp DNA sequences of seven Salicaceae species (Wu, 2015).

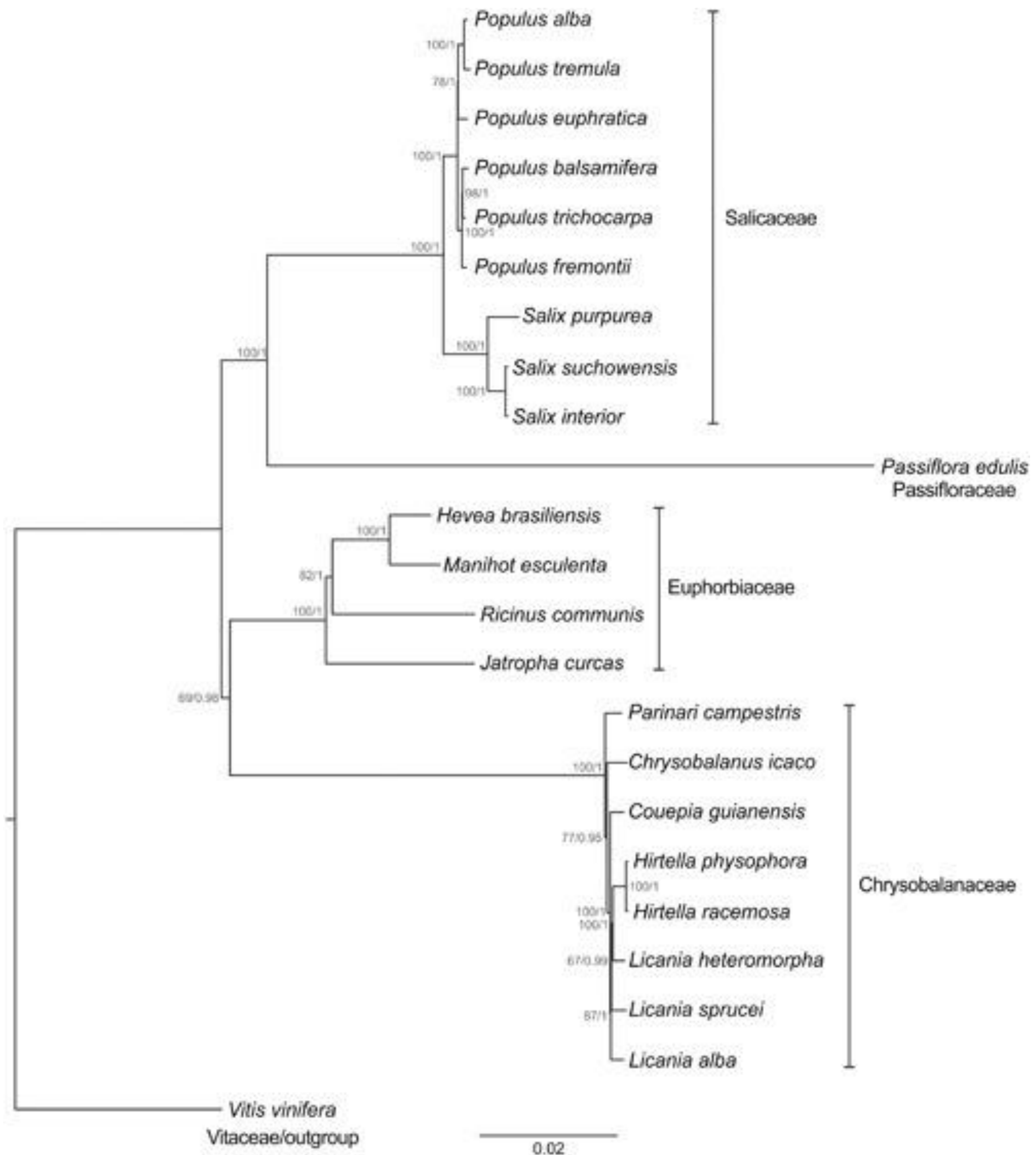


Figure 5. Phylogenetic tree of the order Malpighiales inferred from the complete nucleotide sequence of the chloroplast genome. Maximum Likelihood analysis was applied to the TVM + G model, whereas Bayesian inference (BI) was applied the GTR + G model. The bootstrap values for maximum likelihood and posterior probability for BI are indicated above each node. *Vitis vinifera* was used as outgroup to produce a rooted tree. The scale bar indicates the number of nucleotide substitutions per site.

With reference to Euphorbiaceae, *M. esculenta* was placed near to *H. brasiliensis* (BS = 100 %; PP = 1), similar to the findings published by Xi et al. (2012) after examining a set of 82 chloroplast genes. Our findings placed *J. curcas* as the species most distant from the other Euphorbiaceae (BS = 100 %; PP = 1), occupying an onward position relative to *R. communis* (BS = 82 %; PP = 1). The placement of these species has weak support in two previous topologies based on 62 (Su et al., 2014) and 60 chloroplast genes (Kong and Yang,

2016). Remarkably, our study resolved the node positions in the Euphorbiaceae with strong support in both analyses (ML and BA). These findings also demonstrate the significance of using complete cp genomes in phylogeny reconstructions, since not only the genes but also non-coding sequences are examined. These sequences are informative, particularly at low taxonomic levels, due to their rapid evolution. In contrast, protein coding genes evolve at a relatively slow rate (Folk et al., 2015).

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

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

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

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

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

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

The members of the order Oxalidales (*Averrhoa carambola* and *Oxalis latifolia*) were placed as sister to a clade comprising the members of the order Celastrales (*Elaeodendron orientale* and *Euonymus americanus*) with strong support (BS = 100 %; PP = 1). Thus, our studies indicate that Malpighiales shares more homologies with Celastrales than with Oxalidales, based on both methods used to infer species relationships within the Fabids.

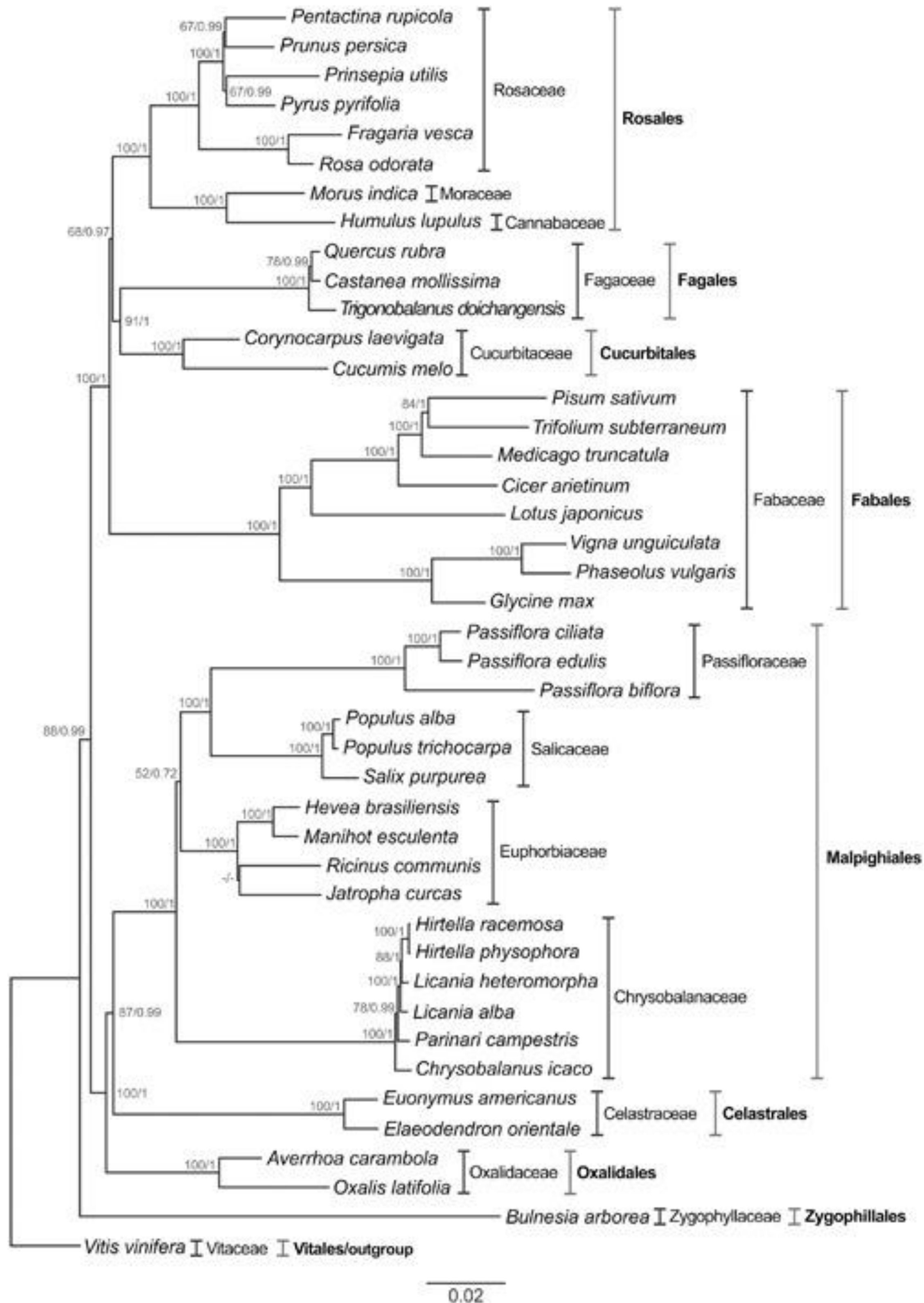


Figure 6. Phylogenetic tree of the fabid clade inferred from the nucleotide sequences of cp protein coding genes of 43 plant species. Both Maximum Likelihood and Bayesian inference (BI) were applied to the GTR + G model. The bootstrap values for maximum likelihood and posterior probability for BI are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

Previously, Hilu et al. (2003) reported a feasible relationship among the orders of the COM clade, based on sequence alignments of *matK*, a cp gene, but with weak support using maximum parsimony (BS = 60%). According to Sun et al. (2015), positioning the orders within the COM clade remains a great challenge. However, our results confirm other findings, for instance placing Celastrales nearer to Malpighiales than Oxalidales (Bell et al., 2010; Zhang and Simmons, 2006). The same scenario was described by Matthews and Endress, 2005 observing floral structures and describing their implications for systematics in Celastrales. The use of a large set of genes is important for defining genomic relationships; otherwise the results could be interpreted as relating to the evolutionary course of particular genes.

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

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

2.4. Conclusions

In this study, it was possible to obtain the first complete sequence of a chloroplast genome for the family Passifloraceae using the SMRT sequencing method, which proved highly effective for generating the biological data and efficient in assembling the chloroplast genome of *Passiflora edulis*. Chloroplast genomes of the order Malpighiales were compared, and although found to be highly conserved, some genomes such as *P. edulis* have undergone rearrangements during evolution, showing that there is diversity in the structure of plastid genomes in angiosperms. Definitely, complete chloroplast genomes or a significant set of their genic sequences are clearly useful in phylogenetic studies. Additionally, our results have opened the way for future phylogenomic studies on Passifloraceae.

REFERENCES

Asif, M. H., Mantri, S. S., Sharma, A., Srivastava, A., Trivedi, I., Gupta, P., et al. (2010). Complete sequence and organisation of the *Jatropha curcas* (Euphorbiaceae) chloroplast genome. *Tree Genet. Genomes* 6, 941–952. doi:10.1007/s11295-010-0303-0.

Bardon, L., Chamagne, J., Dexter, K. G., Sothers, C. A., Prance, G. T., and Chave, J. (2013). Origin and evolution of Chrysobalanaceae: insights into the evolution of plants in the Neotropics. *Bot. J. Linn. Soc.* 171, 19–37. doi:10.1111/j.1095-8339.2012.01289.x.

Barkan, A. (2004). "Intron splicing in plant organelles," in *Molecular biology and biotechnology of plant organelles*, eds. H. Daniell, and C. D. Chase (Dordrecht: Kluwer Academic Publishers), 281–308.

Bell, C. D., Soltis, D. E., and Soltis, P. S. (2010). The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* 97, 1296–1303. doi:10.3732/ajb.0900346.

Bruckner, C. H., Casali, V. W. D., de Moraes, C. F., Regazzi, A. J., and da Silva, E. A. M. (1995). Self-incompatibility in passion fruit (*Passiflora edulis* Sims). *Acta Hortic.*, 45–58. doi:10.17660/ActaHortic.1995.370.7.

Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., and Dopazo, J. (2015). A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the Genus *Citrus*. *Mol. Biol. Evol.* 32, 2015–2035. doi:10.1093/molbev/msv082.

Carneiro, M. S., Camargo, L. E. A., Coelho, A. S. G., Vencovsky, R., Júnior, R. P. L., Stenzel, N. M. C., et al. (2002). RAPD-based genetic linkage maps of yellow passion fruit (*Passiflora edulis* Sims. f. *flavicarpa* Deg.). *Genome* 45, 670–678. doi:10.1139/g02-035.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552.

Chaisson, M. J., and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 13, 238. doi:10.1186/1471-2105-13-238.

Chen, X., Li, Q., Li, Y., Qian, J., and Han, J. (2015). Chloroplast genome of *Aconitum barbatum var. puberulum* (Ranunculaceae) derived from CCS reads using the PacBio RS platform. *Front. Plant Sci.* 6, 42. doi:10.3389/fpls.2015.00042.

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi:10.1038/nmeth.2474.

Conant, G. C., and Wolfe, K. H. (2008). GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24, 861–862. doi:10.1093/bioinformatics/btm598.

Cuco, S. M., Vieira, M. L. C., Mondin, M., and Aguiar-Perecin, M. L. R. (2005). Comparative karyotype analysis of three *Passiflora* L. species and cytogenetic characterization of somatic hybrids. *Caryologia* 58, 220–228.

Curci, P. L., De Paola, D., Danzi, D., Vendramin, G. G., and Sonnante, G. (2015). Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other Asteraceae. *PLoS ONE* 10, e0120589. doi:10.1371/journal.pone.0120589.

Daniell, H., Wurdack, K. J., Kanagaraj, A., Lee, S.-B., Saski, C., and Jansen, R. K. (2008). The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron. *Theor. Appl. Genet.* 116, 723–737. doi:10.1007/s00122-007-0706-y.

Darling, A. C. E. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi:10.1101/gr.2289704.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi:10.1038/nmeth.2109.

Ferrarini, M., Moretto, M., Ward, J. A., Šurbanovski, N., Stevanović, V., Giongo, L., et al. (2013). An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC Genomics* 14, 670. doi:10.1186/1471-2164-14-670.

Ferreira, T. G. T., Penha, H. A., Zucchi, M. I., Santos, A. A., Hanai, L. R., Junqueira, N., et al. (2010). Outcrossing rate in sweet passion fruit based on molecular markers: Outcrossing rate in sweet passion fruit. *Plant Breed.* 129, 727–730. doi:10.1111/j.1439-0523.2009.01744.x.

Feuillet, C. (2004). "Passifloraceae (Passion flower family)," in *Flowering plants of the Neotropics*, eds. N. Smith, S. A. Mori, A. Henderson, D. W. Stevenson, and S. D. Heald (Oxford: Princeton University Press & New York Botanical Garden), 286–287.

Folk, R. A., Mandel, J. R., and Freudenstein, J. V. (2015). A Protocol for targeted enrichment of intron-containing sequence markers for recent radiations: A phylogenomic example from *Heuchera* (Saxifragaceae). *Appl. Plant Sci.* 3, 1500039. doi:10.3732/apps.1500039.

Grassi, F., Labra, M., Scienza, A., and Imazio, S. (2002). Chloroplast SSR markers to assess DNA diversity in wild and cultivated grapevines. *Vitis* 41, 157–158.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Hansen, A. K., Gilbert, L. E., Simpson, B. B., Downie, S. R., Cervi, A. C., and Jansen, R. K. (2006). Phylogenetic relationships and chromosome number evolution in *Passiflora*. *Syst. Bot.* 31, 138–150. doi:10.1600/036364406775971769.

Hilu, K. W., Borsch, T., Muller, K., Soltis, D. E., Soltis, P. S., Savolainen, V., et al. (2003). Angiosperm phylogeny based on *matK* sequence information. *Am. J. Bot.* 90, 1758–1776. doi:10.3732/ajb.90.12.1758.

Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., dePamphilis, C. W., Leebens-Mack, J., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci.* 104, 19369–19374. doi:10.1073/pnas.0709121104.

Jansen, R. K., Kaittanis, C., Saski, C., Lee, S.-B., Tomkins, J., Alverson, A. J., et al. (2006). Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. *BMC Evolitionary Biol.* 6, 32.

Jansen, R. K., Saski, C., Lee, S.-B., Hansen, A. K., and Daniell, H. (2011). Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): Evidence for at least two independent transfers of rpl22 to the nucleus. *Mol. Biol. Evol.* 28, 835–847. doi:10.1093/molbev/msq261.

Jansen, R. K., Wojciechowski, M. F., Sanniyasi, E., Lee, S.-B., and Daniell, H. (2008). Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Mol. Phylogenet. Evol.* 48, 1204–1217. doi:10.1016/j.ympev.2008.06.013.

Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F., and Donghue, M. J. eds. (2008). *Plant systematics: a phylogenetic approach*. 3rd ed. Sunderland, Mass: Sinauer Associates.

Kong, W., and Yang, J. (2016). The complete chloroplast genome sequence of *Morus mongolica* and a comparative analysis within the Fabidae clade. *Curr. Genet.* 62, 165–172 doi:10.1007/s00294-015-0507-9.

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi:10.1093/nar/29.22.4633.

Lambowitz, A. M., and Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* 3, a003616–a003616. doi:10.1101/cshperspect.a003616.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi:10.1093/bioinformatics/btm404.

Li, X., Zhang, T.-C., Qiao, Q., Ren, Z., Zhao, J., Yonezawa, T., et al. (2013). Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). *PLoS ONE* 8, e58747. doi:10.1371/journal.pone.0058747.

Malé, P.-J. G., Bardon, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., et al. (2014). Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Mol. Ecol. Resour.* 14, 966–975. doi:10.1111/1755-0998.12246.

Martin, G., Baurens, F.-C., Cardi, C., Aury, J.-M., and D'Hont, A. (2013). The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): Insight into plastid monocotyledon evolution. *PLoS ONE* 8, e67350. doi:10.1371/journal.pone.0067350.

Matthews, M. L., and Endress, P. K. (2005). Comparative floral structure and systematics in Celastrales (Celastraceae, Parnassiaceae, Lepidobotryaceae). *Bot. J. Linn. Soc.* 149, 129–194. doi:10.1111/j.1095-8339.2005.00445.x.

Melotto-Passarin, D. M., Tambarussi, E. V., Dressano, K., De Martin, V. F., and Carrer, H. (2011). Characterization of chloroplast DNA microsatellites from *Saccharum* spp and related species. *Genet. Mol. Res.* 10, 2024–2033. doi:10.4238/vol10-3gmr1019.

Milligan, B. G., Hampton, J. N., and Palmer, J. D. (1989). Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Mol. Biol. Evol.* 6, 355–368.

Moraes, M. C., Gerald, I. O., Matta, F. P., and Vieira, M. L. C. (2005). Genetic and phenotypic parameter estimates for yield and fruit quality traits from a single wide cross in yellow passion fruit. *HortScience* 40, 1978–1981.

Munhoz, C. F., Santos, A. A., Arenhart, R. A., Santini, L., Monteiro-Vitorello, C. B., and Vieira, M. L. C. (2015). Analysis of plant gene expression during passion fruit- *Xanthomonas axonopodis* interaction implicates lipoxygenase 2 in host defence. *Ann. Appl. Biol.* 167, 135–155. doi:10.1111/aab.12215.

Muschner, V. C., Zamberlan, P. M., Bonatto, S. L., and Freitas, L. B. (2003). Phylogeny, biogeography and divergence times in *Passiflora* (Passifloraceae). *Genet. Mol. Biol.* 35, 1036–1043.

Oliveira, E. J., Vieira, M. L. C., Garcia, A. A. F., Munhoz, C. F., Margarido, G. R. A., Consoli, L., et al. (2008). An integrated molecular map of yellow passion fruit based on simultaneous maximum-likelihood estimation of linkage and linkage phases. *J Am Soc Hortic Sci* 133, 35–41.

Provan, J., Powell, W., and Hollingsworth, P. M. (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol. Evol.* 16, 142–147.

Ramaiya, S. D., Bujang, J. S., and Zakaria, M. H. (2014). Assessment of total phenolic, antioxidant, and antibacterial activities of *Passiflora* species. *Sci. World J.* 2014, 1–10. doi:10.1155/2014/167309.

Rêgo, M. M., Rêgo, E. R., Bruckner, C. H., Da Silva, E. A. M., Finger, F. L., and Pereira, K. J. C. (2000). Pollen tube behavior in yellow passion fruit following compatible and incompatible crosses. *Theor. Appl. Genet.* 101, 685–689. doi:10.1007/s001220051531.

Rivarola, M., Foster, J. T., Chan, A. P., Williams, A. L., Rice, D. W., Liu, X., et al. (2011). Castor bean organelle genome sequencing and worldwide genetic diversity analysis. *PLoS ONE* 6, e21743. doi:10.1371/journal.pone.0021743.

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi:10.1093/bioinformatics/btg180.

Santos, A., Penha, H., Bellec, A., Munhoz, C. F., Pedrosa-Harand, A., Bergès, H., et al. (2014). Begin at the beginning: A BAC-end view of the passion fruit (*Passiflora*) genome. *BMC Genomics* 15, 816. doi:10.1186/1471-2164-15-816.

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., et al. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142–166. doi:10.3732/ajb.92.1.142.

Shin, S. C., Ahn, D. H., Kim, S. J., Lee, H., Oh, T.-J., Lee, J. E., et al. (2013). Advantages of Single-Molecule Real-Time sequencing in high-GC content genomes. *PLoS ONE* 8, e68824. doi:10.1371/journal.pone.0068824.

Simon, D., Fewer, D., Friedl, T., and Bhattacharya, D. (2003). Phylogeny and self-splicing ability of the plastid tRNA-Leu group I Intron. *J. Mol. Evol.* 57, 710–720. doi:10.1007/s00239-003-2533-3.

Souza, M. M., Palomino, G., Pereira, T. N. S., Pereira, M. G., and Viana, A. P. (2004). Flow cytometric analysis of genome size variation in some *Passiflora* species. *Hereditas* 141, 31–38.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033.

Su, H.-J., Hogenhout, S. A., Al-Sadi, A. M., and Kuo, C.-H. (2014). Complete chloroplast genome sequence of Omani Lime (*Citrus aurantiifolia*) and comparative analysis within the Rosids. *PLoS ONE* 9, e113049. doi:10.1371/journal.pone.0113049.

Sugiura, M. (1992). The chloroplast genome. *Plant Mol. Biol.* 19, 149–168. doi:10.1007/BF00015612.

Sun, M., Soltis, D. E., Soltis, P. S., Zhu, X., Burleigh, J. G., and Chen, Z. (2015). Deep phylogenetic incongruence in the angiosperm clade Rosidae. *Mol. Phylogenet. Evol.* 83, 156–166. doi:10.1016/j.ympev.2014.11.003.

Svistoonoff, S., Benabdoun, F. M., Nambiar-Veetil, M., Imanishi, L., Vaissayre, V., Cesari, S., et al. (2013). The independent acquisition of plant root nitrogen-fixing symbiosis in Fabids recruited the same genetic pathway for nodule organogenesis. *PLoS ONE* 8, e64515. doi:10.1371/journal.pone.0064515.

Swofford, D. L. (2002). *PAUP*: Phylogenetic analysis using parsimony (and other methods)*. Sunderland, MA: Sinauer Associates.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi:10.1093/molbev/mst197.

Tangphatsornruang, S., Uthaipaisanwong, P., Sangsrakru, D., Chanprasert, J., Yoocha, T., Jomchai, N., et al. (2011). Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene* 475, 104–112. doi:10.1016/j.gene.2011.01.002.

The Angiosperm Phylogeny Group (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi:10.1111/j.1095-8339.2009.00996.x.

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604. doi:10.1126/science.1128691.

Ulmer, T., and MacDougal, J. M. (2004). *Passiflora: Passionflowers of the world*. Portland: Timber Press.

Wang, H., Moore, M. J., Soltis, P. S., Bell, C. D., Brockington, S. F., Alexandre, R., et al. (2009). Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci.* 106, 3853–3858. doi:10.1073/pnas.0813376106.

Wu, Z. (2015). The new completed genome of purple willow (*Salix purpurea*) and conserved chloroplast genome structure of Salicaceae. *J. Nat. Sci.* 1, e49.

Wurdack, K. J., and Davis, C. C. (2009). Malpighiales phylogenetics: Gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. *Am. J. Bot.* 96, 1551–1570. doi:10.3732/ajb.0800207.

Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255. doi:10.1093/bioinformatics/bth352.

Xi, Z., Ruhfel, B. R., Schaefer, H., Amorim, A. M., Sugumaran, M., Wurdack, K. J., et al. (2012). Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc. Natl. Acad. Sci.* 109, 17519–17524. doi:10.1073/pnas.1205818109.

Xia, X. (2013). DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* 30, 1720–1728. doi:10.1093/molbev/mst064.

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., et al. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE* 5, e12762. doi:10.1371/journal.pone.0012762.

Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., and Zhao, G. (2016). Comparative analysis of the complete chloroplast genomes of five *Quercus* species. Front. Plant Sci. 7, 959. doi: 10.3389/fpls.2016.00959

Yockteng, R., and Nadot, S. (2004). Phylogenetic relationships among *Passiflora* species based on the glutamine synthetase nuclear gene expressed in chloroplast (*ncpGS*). *Mol. Phylogenet. Evol.* 31, 379–396. doi:10.1016/S1055-7903(03)00277-X.

Yotoko, K. S. C., Dornelas, M. C., Togni, P. D., Fonseca, T. C., Salzano, F. M., Bonatto, S. L., et al. (2011). Does variation in genome sizes reflect adaptive or neutral processes? New clues from *Passiflora*. *PLoS ONE* 6, e18212. doi:10.1371/journal.pone.0018212.

Zhang, L.-B., and Simmons, M. P. (2006). Phylogeny and delimitation of the Celastrales inferred from nuclear and plastid genes. *Syst. Bot.* 31, 122–137. doi:10.1600/036364406775971778.

3. EVOLUTIONARY HISTORY OF CHLOROPLAST GENOMES: A CASE STUDY IN POINT REVEALS A REPERTORY OF REARRANGEMENTS AND THE LOSS OF AN INVERTED REPEAT REGION IN *PASSIFLORA*[2]

ABSTRACT

Chloroplast genomes (cpDNA) in angiosperms are usually highly conserved. Although rearrangements have been observed in some lineages, the mechanisms that lead to rearrangements are still poorly elucidated. For instance, some groups with different patterns of chloroplast inheritance, such as *Passiflora*, also exhibit highly rearranged cpDNA structures. In the present study, we obtained 20 new chloroplast genomes to investigate cpDNA evolutionary history in *Passiflora* (Passifloraceae), a genus with cytonuclear incompatibility and biparental chloroplast inheritance. *Passiflora* cpDNAs vary in size considerably, with ~50 kb between shortest and longest. Large inverted repeat (IR) expansions were identified, and at the extreme opposite, the loss of an IR was detected for the first time in *Passiflora*, a rare event in angiosperms. The loss of an IR region was detected in *P. costaricensis*, a species in which occasional biparental chloroplast inheritance has previously been reported. A repertory of rearrangements, such as inversions and gene losses were detected, making *Passiflora* one of the few groups with complex chloroplast genome evolution. Intramolecular recombination of repeats and tRNAs appears to be the mechanism responsible for some inversions. We also performed a phylogenomic study based on all the available cp genomes and our analysis implies that there is a need to reconsider the taxonomic classifications of some species in the group.

Keywords: Chloroplast genome (plastome) rearrangements, *Passiflora*, inverted repeat (IR) loss, *Passiflora* evolution.

3.1. Introduction

In higher plants, most chloroplast (cp) genomes have a quadripartite structure consisting of two copies of inverted repeats (IRs) separating two single copy regions, and a large (LSC) and a small (SSC) region (Sugiura 1992; Yang et al. 2010). Although highly conserved in their structural organization, cp genomes may exhibit deletions, including gene loss, and rearrangements, such as IR inversions and expansions or retractions. These rearrangements are well documented, for instance, in *Hevea brasiliensis* (Euphorbiaceae) (Tangphatsornruang et al. 2011) and in species of riverweeds (Podostemaceae) (Bedoya et al. 2019), both families belonging to the Malpighiales. Some other plant families, such as Fabaceae (Cai et al. 2008), Campanulaceae (Haberle et al. 2008), Pinaceae and other conifers

---

[2] The text corresponds to that of the article by Cauz-Santos et al. (2020) submitted to *Genome Biology and Evolution*.

(Wu et al. 2011a,b), and Geraniaceae (Guisinger et al. 2011; Weng et al. 2014) also exhibit a high number of rearrangements in their cp genomes.

However, the mechanisms which lead to rearrangements in plastid genomes (plastomes) are still poorly elucidated. Among these mechanisms, intramolecular homologous recombination mediated by the presence of repeat structures at the boundaries of the rearranged region plays a role in bringing about structural changes (Ogihara et al. 1988; Wu et al. 2011a), as well as recombination between tRNA genes (Haberle et al. 2008). In addition, foreign DNA insertions (large ORFs, for instance) have led to extensive cp genome rearrangements in Campanulaceae *sensu stricto* (Knox 2014). Interestingly, for some angiosperms in which highly rearranged cpDNAs occur, occasional biparental chloroplast inheritance has been reported, as in Campanulaceae (Cosner et al. 2004; Haberle et al. 2008; Barnard-Kubow et al. 2017), and Geraniaceae (Metzlaff et al. 1981; Chumley et al. 2006; Weng et al. 2014).

In the genus *Passiflora* (Passifloraceae, Malpighiales), the results of artificial intraspecific crosses revealed the prevalence of maternal and the potential for biparental chloroplast inheritance, whereas interspecific crosses in turn showed the occurrence of paternal chloroplast inheritance (Muschner et al. 2006; Hansen et al. 2007). *Passiflora* could be regarded as an excellent model for studying the evolution of chloroplast genomes. Apart from the potential for different patterns of chloroplast inheritance, the genus has been described as having a syndrome of features related to chloroplast genome changes (Shrestha et al. 2019). Rearrangements in the cpDNA structure, such as inversions and gene losses, have been reported for the first time in *P. edulis*. In this study, high-quality sequences were assembled from long reads (Cauz-Santos et al. 2017). In later studies, inversions, gene losses and extensions of IR regions were also reported in different *Passiflora* species (Rabah et al. 2019; Shrestha et al. 2019).

To place this in context, *Passiflora* is the richest genus in the family Passifloraceae, consisting of some 520 species, popularly known as passionflowers, with great diversity in the size and shape of flowers pollinated by insects, hummingbirds or bats. The geographical distribution of passionflowers is mainly Neotropical, and they are found mainly in the South and Central Americas. However, occurrences have been documented in Southeast Asia, Oceania and Australia (Ulmer & MacDougal 2004). The Brazilian biomes (Amazon, Caatinga, Cerrado, Atlantic forest, Pampa and Pantanal) harbor 147 species, including 85 that are endemic to the country. Some species are cultivated for medicinal (e.g. P. incarnata) and ornamental purposes due to their exotic flowers, but most are for fresh fruit consumption and

industrialized juice production (e.g. *P. edulis*). Morphological characters have been used to subdivide the classical intrageneric division of the genus into 22 subgenera (Killip 1938). Nowadays, the most well-accepted classification has reduced the number of subgenera to four: *Astrophea* (57 species), *Decaloba* (214), *Deidamioides* (13) and *Passiflora* (236) (Ulmer & MacDougal 2004). Phylogenetic studies have also confirmed the subdivision of *Passiflora* into four subgenera (Hansen et al. 2006; Muschner et al. 2003, 2012). However, the position of subg. *Deidamioides* is poorly resolved and it has been recognized as a paraphyletic group (Muschner et al. 2003, 2012), suggesting that further analysis is needed using phylogenomic approaches and larger volumes of data (e.g. a set of chloroplast genes).

In our study, we sequenced, assembled and annotated 18 chloroplast genomes representing the four subgenera in the genus *Passiflora*, along with the Passifloraceae *Dilkea retusa* and *Mitostemma brevifilis*, and extensive sequence rearrangements were found. We were able to address the following questions: Are the frequency and type of cpDNA sequence rearrangements particular to each subgenus? And what is the significance of these rearrangements for *Passiflora* intrageneric classification?

3.2. Material and Methods

3.2.1. Plant material

Most of the *Passiflora* species used in this study were field-collected in the following Brazilian biomes: Amazon, Atlantic rainforest, Cerrado (Brazilian savannas) and Caatinga (Figure 7). In total, 18 species were analyzed, belonging to the four subgenera: *Astrophea* (3), *Decaloba* (5), *Deidamioides* (2) and *Passiflora* (8). Note that *P. costaricensis* (*Decaloba*), a native species of Central America, was obtained from the Italian National Collection of *Passiflora*. In addition, two species of Passifloraceae (*Dilkea retusa* and *Mitostemma brevifilis*) were studied and used in comparative studies (APPENDIX C). All plant accessions are registered at Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (SISGEN, Brazil).

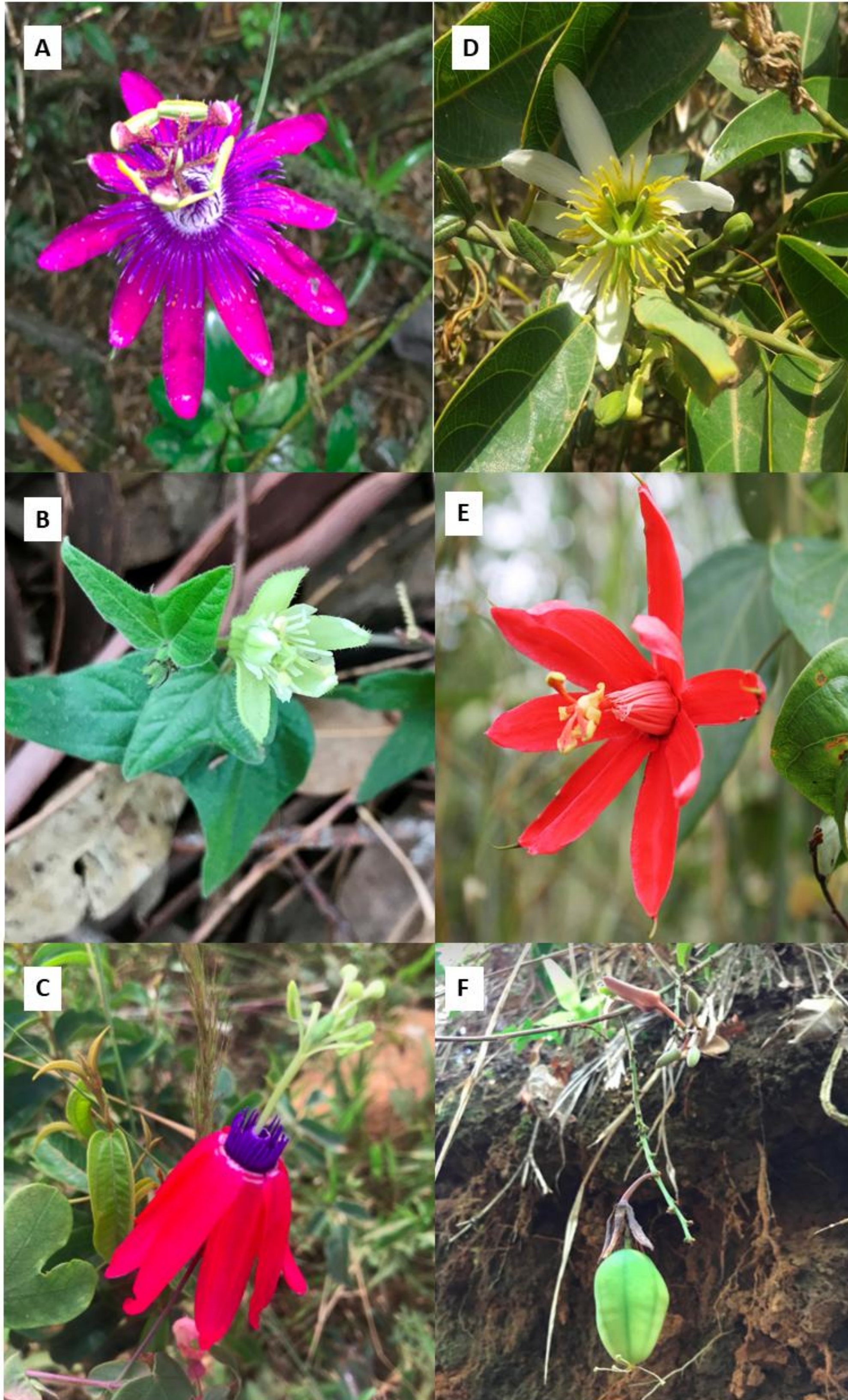


Figure 7. Species of *Passiflora* collected in scientific expeditions in Brazil. A) *Passiflora loefgrenii*; B) *Passiflora capsularis*; C) *Passiflora edmundoi*; D) *Passiflora rhamnifolia*; E) *Passiflora cristalina*; F) *Passiflora contracta*.

3.2.2. Intact chloroplast isolation in sucrose gradient and cpDNA extraction

Intact chloroplast organelles were isolated using the sucrose gradient method (Takamatsu et al. 2018). Some 20 g of fresh leaves from each plant were frozen in liquid nitrogen and macerated. The material was resuspended in 200 mL of isolation buffer (50 mM Tris-HCl pH 8.0, 0.35 M sucrose, 7 mM EDTA, 5 mM 2-mercaptoethanol, 0.1% BSA) and incubated for 10 min in the dark. The suspension was filtered through two layers of gauze and then two layers of Miracloth (Merck), and the filtrate centrifuged at 1000× g for 10 min. Finally, the pellet was washed in 50 mL of isolation buffer and centrifuged at 1,000× g for 10 min.

The pellet was resuspended in 5 mL of isolation buffer and the suspension slowly laid out in a 20/45% sucrose density batch gradient in 50 mM Tris-HCl (pH 8.0), 0.3 M sorbitol and EDTA 7 mM. The gradient was centrifuged at 2000× g for 30 min. The green band formed at the interface containing the intact chloroplasts was collected, diluted with three volumes of isolation buffer and centrifuged at 3,000× g for 10 min to obtain the purified chloroplasts in the pellet.

The pellet was resuspended in 2% CTAB buffer to promote lysis. The suspension was then incubated at 65 °C for 1 h with stirring. The supernatant was extracted 2× with an equal volume of chloroform:isoamyl alcohol (24:1) and centrifuged at 10,000× g for 20 min. An equal volume of isopropanol was added to the upper layer and incubated at 20 °C for 1 h. Finally, the aqueous phase was centrifuged at 10,000× g for 20 min, and the cpDNA pellet washed with ethanol (70%), dried and resuspended in 40 µL of TE buffer.

3.2.3. Chloroplast genome sequencing and assembly

Illumina sequencing libraries were constructed using a total of 100 ng of input cpDNA and the Nextera DNA Flex library Kit, following the manufacturer's instructions. The sequencing was performed on the Illumina NextSeq platform (Fundação Hemocentro de Ribeirao Preto, Brazil) in two distinct runs, the first containing 6 species using paired-end sequencing (2× 76 bp), and the second 13 species using paired-end sequencing (2× 150 bp) (APPENDIX D). The paired-end reads were initially trimmed and filtered in Trimmomatic v. 0.39 (Bolger et al. 2014). The adapter trimming was performed based on ILLUMINACLIP:NexteraPE-PE.fa:2:30:10:2, and using the following quality filtering

parameters: sliding window of 10:20, leading of 20, trailing of 20, and minimum length of 36 bp.

The filtered reads were *de novo* assembled in NOVOPlasty v. 3.8.3 (Dierckxsens et al. 2017), using the protein-coding genes *psbA*, *rbcL* and *rpoB* from the *P. edulis* cp genome (NC_034285.1) as seeds. The k-mer sizes used in the assembly varied from 23 to 39. (APPENDIX D). In order to obtain the final assembly, and taking into account that there were overlapping regions, the cp sequences were merged in Geneious v. 2019.1.2 (https://www.geneious.com) using the "*de novo* Assembly" function. Finally, the correctness of the assembly was checked, and coverage estimated using the "Map to reference" function to map the paired-end raw reads onto the final assembled cp genome.

The plastid DNA sequence of *P. costaricensis* was obtained by long read sequencing. Barcoded, large-insert (10 kb) libraries were constructed using 150 ng of pure high molecular weight DNA. Sequencing was performed in an SMRT cell using P6 polymerase with C4 chemistry on the PacBio RSII instrument at the Uppsala NGI Platform (Uppsala University, Sweden). The data was filtered to obtain high-quality reads (reads with quality < 0.75 and length < 500 bp were removed). The sequences were assembled using the CANU assembler with default parameters (Koren et al. 2017). Subsequently, the chloroplast contigs were extracted in Geneious by mapping the resulting contigs to the *Passiflora* complete cp genomes obtained in this study. A final assembled cp genome of *P. costaricensis* was obtained in Geneious by joining the contigs using the "*de novo* Assembly" function (APPENDIX D). Finally, the raw reads were mapped onto the final contig using the "Map to reference" function.

Primer design and PCR were applied to confirm the positioning of some distinct cpDNA arrangements (APPENDIX E). Amplification reactions were performed using 20 ng template DNA, 1× buffer, 1 mM MgCl2, 0.2 mM of each dNTP, 0.3 µM of the forward and reverse primers, 1.2 U Go Taq Flex DNA polymerase (Promega, Madison, WI, USA), and ultra-pure water added to bring the final volume up to 20 µL. The amplification program was as follows: 95°C for 5 min, 35 cycles at 95°C for 40 s, 55°C for 40 s and 72°C for 1 min, followed by a final 8 min of incubation at 72°C. The amplified fragments were checked on 1% (w/v) agarose gel with a 1000 bp molecular standard Invitrogen ladder (Carlsbad, CA, USA).

3.2.4. Chloroplast genome annotation, identification of repeated elements and comparative analysis

The annotation of cp genomes was carried out in the GeSeq (Organellar Genome Annotation) online program (Tillich et al. 2017) with default settings to identify protein coding gene sequences (CDS), rRNAs and tRNAs based on the chloroplast reference sequences and BLAST homology searches, followed by manual corrections for start and stop codons, and intron positions in GenomeView software. All tRNA genes were further confirmed using the tRNAscan-SE and ARAGORN online search server (Lowe & Chan 2016; Laslett & Canback 2004). Pseudogenes were classified based on nucleotide losses in sequences or the presence of internal stop codons. OGDRAW software was used for constructing the circular cp genome map (Greiner et al. 2019).

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

Two different multiple sequence alignments were run in Progressive Mauve v.2.4.0 (Darling et al. 2010) to identify possible rearrangements in the cpDNA molecules. The first contained 11 cp genomes to represent the diversity in size and structure in each *Passiflora* subgenus, in addition to the cp genomes of *Dilkea retusa* and *Mitostemma brevifilis* (Passifloraceae), with *Populus trichocarpa* (Salicaceae) as a reference. The second alignment contained all 20 cp genomes studied herein. Subsequently, to detect possible expansions or contractions in the IR regions, the IR-boundaries of both LSC and SSC regions with full annotations for the adjacent genes were analyzed across the cp molecules used in the first multiple sequence alignment. In addition to reconstructing the evolutionary history of *Passiflora* chloroplast genomes, we analyzed the plastome structures run in the first multiple sequence alignment against the available Genbank cp genomes of *P. foetida*, *P. tetrandra*, *P. obovata*, *P. auriculata* and *P. biflora* (APPENDIX F).

3.2.5. Phylogenomic studies in *Passiflora*

We performed a phylogenomic study based on all the available cp genomes of passionflowers (49 species in total, 20 generated in this study and 29 species whose cp DNA sequences were imported from the Genbank database, APPENDIX F).

A set of 68 chloroplast protein coding genes was used as the ingroup in a phylogenetic analysis, and the respective gene sequences of *Populus trichocarpa* (Salicaceae) was used as the outgroup, with the aim of obtaining a rooted tree.

A local python pipeline was used to extract each gene in the data set (consisting of 50 taxa), align them individually at nucleotide level in MUSCLE and make an interleaved Nexus matrix consisting of all individual alignments. The resulting matrix was then analyzed in ModelFinder (Kalyaanamoorthy et al. 2017) to determine the best partition scheme and evolutionary models in accordance with the AIC criteria and treating each gene alignment as a charset. Because the partition scheme indicated by ModelFinder is an a priori test based on point estimates, and a single model was selected with GTR+G+I for the whole dataset, we decided to include an alternative model with partition based on codon position with separate parameters in GTR+G+I for each codon position, and then subsequently test the two-partition scheme using Bayes Factors. In both partitioned and non-partitioned analyses, the Markov chain Monte Carlo (MCMC) algorithm was run in MrBayes (Ronquist et al. 2012) for 10,000,000 generations, sampling one tree every 1000 generations and discarding the first 25% of trees as burn-in, to estimate the values of posterior probabilities. The convergence of the runs was monitored based on an average standard deviation of split frequencies below 0.01, potential scale reduction factor (PSRF) values close to 1.0, and ESS values above 1,000. Finally, the phylogenetic trees were visualized using FigTree version 1.4.4.

3.3. Results

3.3.1. Chloroplast genome structure features in Passifloraceae

In *Passiflora*, a large variation in cpDNA sequence length was observed (~55 kb), ranging from 113,114 bp in *P. capsularis* (*Decaloba*) to 167,953 bp in *P. deidamioides* (*Deidamioides*) (APPENDIX G, Table 5). The majority had the typical quadripartite structure; the length of the LSC region ranged from 57,244 bp in *P. suberosa* (*Decaloba*) to 90,064 bp in *P. rhamnifolia* (*Astrophea*), while the SSC region varied from 12,854 in *P. cerradensis* (*Astrophea*) to 13,817 bp in *P. suberosa*. IRs ranged from 21,928 in *P. edmundoi* (*Passiflora*) to 43,626 bp in *P. suberosa* (*Decaloba*). The cpDNA sequence lengths of the Passifloraceae *Dilkea retusa* and *Mitostemma brevifilis* were similar to those of the *Astrophea* species, which

had the shortest SSC region sequenced in this study. The GC content of all 20 cp genomes was very similar (average 36.7%).

Comparing the four subgenera, *Decaloba* was the subgenus with the highest variation, ranging from 113,114 bp in *P. capsularis* and 114,230 in *P. costaricensis*, the smallest cp molecules sequenced herein due to the loss of an IR region (Figure 8), up to 158,313 bp in *P. suberosa*. This loss of the IR region in *P. capsularis* and *P. costaricensis*, particularly the IRa, was confirmed by mapping the raw reads onto the assembled genomes and via PCR (APPENDIX E). In contrast, the largest cp genomes were observed for the two species of the *Deidamioides* subgenus. The *Astrophea* subgenus also had large cpDNA molecules, whereas in the eight species of the *Passiflora* subgenus (which contains the vast majority of the species in the genus) the cpDNA ranged from 142,737 bp (*P. edmundoi*) to 151,920 bp (*P. miniata*).

The cp genomes contained between 102 and 109 unique genes, and this variation is related to the protein-coding genes identified when species were compared (68 to 75 protein-coding genes), since all species were found to have the same tRNA (30) and rRNA (4) gene content. Chloroplast genes are involved in different biological processes, and were annotated accordingly with functional categories. Due to duplication and the emergence of inverted repeat (IR) regions during the evolutionary history of the genus, 18 to 38 genes were found to have two copies, including protein coding genes, tRNAs and all four rRNAs. The total number of genes ranged from 102 (*P. capsularis* and *P. costaricensis*) to 141 (*P. suberosa*), as a result of expansions and retractions of the IR regions, and also the gene losses described below.

Table 5. Summary of the 20 Passifloraceae chloroplast genomes sequenced.

Species	Taxonomic classification Subgenus/Supersections, sections or series	Cp genome size (bp)	LSC (bp)	SSC (bp)	IR (bp)	Total GC %	Number of unique genes	Protein coding genes	tRNAs	rRNAs
Passiflora cerradensis	subgenus *Astrophea*/section *Capreolata*	164,515	84,143	12,854	33,759	36,6	107	73	30	4
Passiflora haematostigma	subgenus *Astrophea*/section *Pseudoastrophea*	163,775	89,717	12,876	30,591	36,5	106	72	30	4
Passiflora rhamnifolia	subgenus *Astrophea*/section *Pseudoastrophea*	162,217	90,064	12,921	29,616	36,4	106	72	30	4
Passiflora candollei	subgenus *Decaloba*/section *Decaloba*	138,081	72,565	13,506	26,005	37,2	104	70	30	4
Passiflora capsularis	subgenus *Decaloba*/section *Xeragona*	113,114	*	*	*	36,1	102	68	30	4
Passiflora costaricensis	subgenus *Decaloba*/section *Xeragona*	114,230	*	*	*	36,1	102	68	30	4
Passiflora suberosa	subgenus *Decaloba*/section *Cieca*	158,313	57,244	13,817	43,626	37,2	103	69	30	4
Passiflora vespertilio	subgenus *Decaloba*/section *Decaloba*	138,456	72,902	13,158	26,196	37,1	104	70	30	4
Passiflora contracta	subgenus *Deidamioides*/section *Tetrastylis*	166,558	87,313	13,513	32,866	36,7	107	73	30	4
Passiflora deidamioides	subgenus *Deidamioides*/section *Deidamioides*	167,953	82,571	13,744	35,819	36,8	108	74	30	4
Passiflora alata	subgenus *Passiflora*/supersection *Laurifolia*/ section *Quadrangularis*	147,773	85,535	13,494	24,372	36,9	105	71	30	4
Passiflora cristalina	subgenus *Passiflora*/supersection *Distephana*	145,054	85,662	13,530	22,931	36,9	104	70	30	4
Passiflora edmundoi	subgenus *Passiflora*/supersection *Stipulata*/section *Kermesinae*	142,737	85,567	13,314	21,928	37,2	105	71	30	4
Passiflora loefgrenii	subgenus *Passiflora*/supersection *Stipulata*/section *Kermesinae*	146,537	86,370	13,267	23,450	37,1	104	70	30	4
Passiflora miniata	subgenus *Passiflora*/supersection *Distephana*	151,920	85,863	13,477	26,290	37,0	105	71	30	4
Passiflora mucronata	subgenus *Passiflora*/supersection *Stipulata*/section *Granadillastrum*	150,839	84,839	13,552	26,224	36,9	104	70	30	4
Passiflora recurva	subgenus *Passiflora*/supersection *Passiflora*/ series *Passiflora*	151,837	85,863	13,504	26,235	37,0	105	71	30	4

Passiflora watsoniana	subgenus *Passiflora*/supersection *Stipulata*/ section *Kermesinae*	146,520	86,139	13,355	23,513	37,0	105	71	30	4
Dilkea retusa	*Dilkea* genus	161,923	88,575	12,686	30,331	36,2	109	75	30	4
Mitostemma brevifilis	*Mitostemma* genus	163,032	88,837	12,695	30,750	36,1	109	75	30	4

* Loss of an IR region

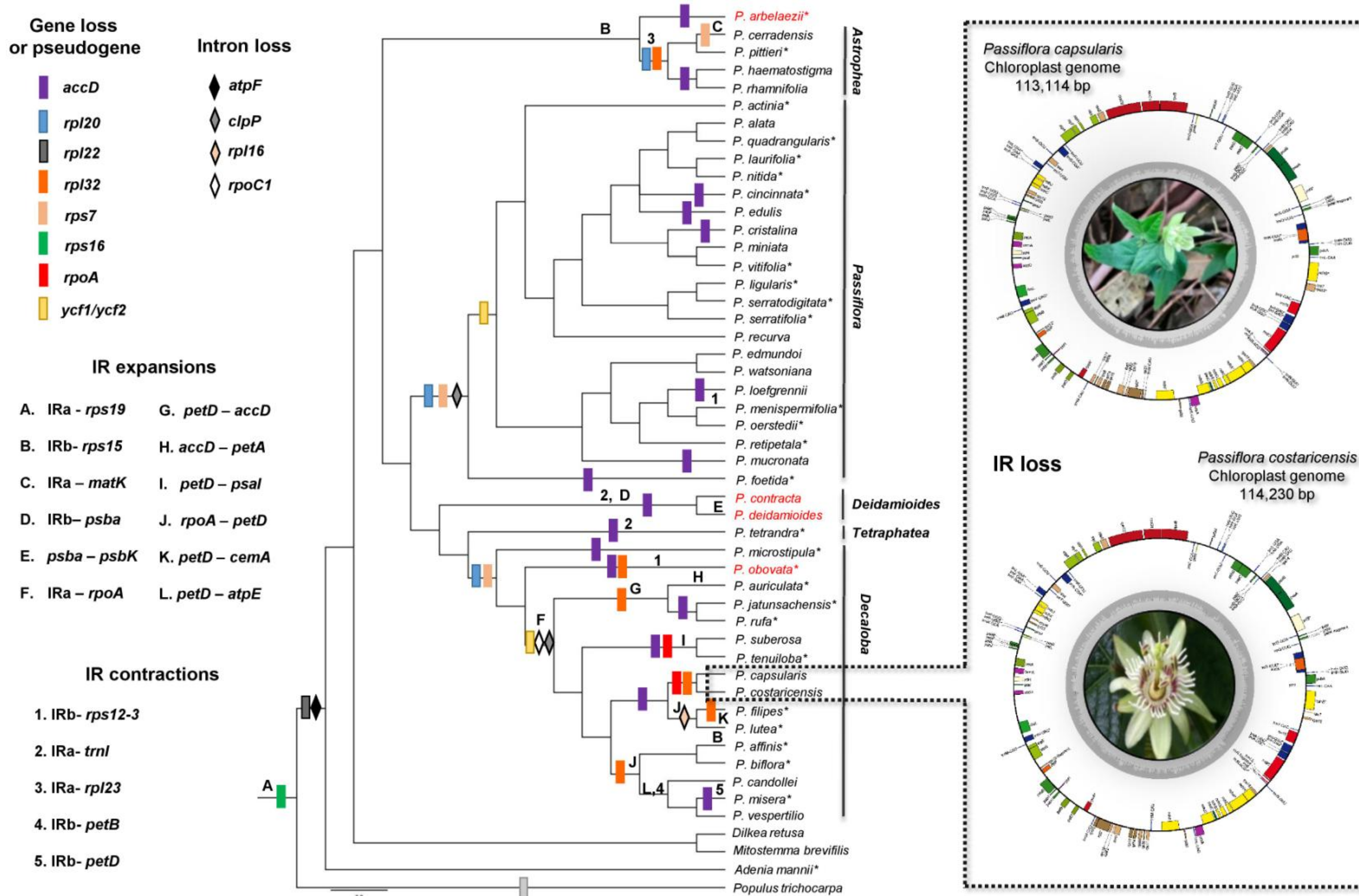


Figure 8. Gene losses, IR expansions/contractions mapped onto the cladogram of the Bayesian inference with plastid genes. *Species analyzed by Shrestha et al. (Shrestha et al. 2019). The chloroplast genomes of *P. capsularis* and *P. costaricensis*, the smallest in *Passiflora* due to the loss of an inverted repeat (IR) region. Genes are represented as boxes inside or outside the large circle to indicate clockwise (inside) or counter clockwise (outside) transcription. The flower image of *P. costaricensis* was kindly provided by Maurizio Vecchia, 2005.

Introns were identified in 12 to 15 sequences of protein coding and tRNA genes, mainly in the *Astrophea* and *Deidamioides* subgenera (15 introns). The intron within the *atpF* gene was not found in the Passifloraceae species analyzed in this study, but it was found in *Populus trichocarpa*, the species used in the comparative analysis (Figure 8). Additionally, the *Astrophea* and *Deidamioides* subgenera harbored an intron in the *clpP* gene, which was not found in the *Decaloba* and *Passiflora* subgenera, revealing the loss of this region. In the *Decaloba* subgenus we found the lowest number of introns, because of losses in the *rpoC1* and *rpl16* genes.

Repetitive sequence analysis detected between 115 (*P. alata*) and 445 (*P. contracta*) repeats (Table 6). The majority were in forward orientation (ranging from 65 repeats in *P. cerradensis* to 286 in *P. haematostigma*), followed by palindromic repeats (ranging from 22 repeats in *P. edmundoi* and *P. loefgrenii* to 183 in *P. contracta*). The length of the repeats varied between 116 bp and 1,070 bp respectively in *P. miniata* and *P. costaricensis*. In most of the species, the highest number of repeat structures was found in the LSC region, with some detected in the IR and SSC regions. Comparing the subgenera, high numbers of repeats were found in *Deidamioides* subgenus (a respective 405 and 445 total repeats in *P. deidamioides* and *P. contracta*) and *Decaloba* subgenus (a respective 199 and 280 total repeats in *P. candollei* and *P. costaricensis*) in which the largest repeat (1,070 bp identified in *P. costaricensis*) possibly corresponds to a piece of the IR region that was lost in this species.

Most of the repeats were found in the intergenic sequences of the LSC or IR regions, but some repeats were located in gene sequences, including *accD*, *clpP*, *psaA*, *psaB, rps18, ycf1*, *ycf2* and *ycf3.* Furthermore, in all species analyzed, a repeat was identified in the *ndhA* gene located in the respective SSC region. Note in particular that a high number of repeats in the *accD* gene and its promoter or intergenic regions was identified in all 20 cp genomes of Passifloraceae, with up to 151 repeats in *P. costaricensis*.

Table 6. Summary of the short direct-repeats identified in the cp genomes of 18 *Passiflora* species and the Passifloraceae, *Dilkea retusa* and *Mitostemma brevifilis*.

Species	Taxonomic classification Subgenus/Supersections, sections or series	Cp genome (bp)	Reputer total number	Palindromic	Forward	Reverse	Complement	Largest repeat (bp)
Passiflora cerradensis	subgenus *Astrophea*/section *Capreolata*	164,515	118	38	65	9	6	193
Passiflora haematostigma	subgenus *Astrophea*/section *Pseudoastrophea*	163,775	399	67	286	28	18	476
Passiflora rhamnifolia	subgenus *Astrophea*/section *Pseudoastrophea*	162,217	398	67	222	56	53	486
Passiflora candollei	subgenus *Decaloba*/section *Decaloba*	138,081	199	87	112			201
Passiflora capsularis	subgenus *Decaloba*/section *Xeragona*	113,114	207	55	155	1		791
Passiflora costaricensis	subgenus *Decaloba*/section *Xeragona*	114,230	280	38	240	2		1070
Passiflora suberosa	subgenus *Decaloba*/section *Cieca*	158,313	279	108	154	9	8	760
Passiflora vespertilio	subgenus *Decaloba*/section *Decaloba*	138,456	238	73	164	1		151
Passiflora contracta	subgenus *Deidamioides*/section *Tetrastylis*	166,558	445	183	262			427
Passiflora deidamioides	subgenus *Deidamioides*/section *Deidamioides*	167,953	405	129	261	15		352
Passiflora alata	subgenus *Passiflora*/supersection *Laurifolia*/ section *Quadrangulares*	147,773	115	33	78	4		123
Passiflora cristalina	subgenus *Passiflora*/supersection *Distephana*	145,054	128	37	87	4		123
Passiflora edmundoi	subgenus *Passiflora*/supersection *Stipulata*/section *Kermesinae*	142,737	185	22	145	18		118
Passiflora loefgrenii	subgenus *Passiflora*/supersection *Stipulata*/section *Kermesinae*	146,537	133	22	104	5	2	137
Passiflora miniata	subgenus *Passiflora*/supersection *Distephana*	151,920	152	35	114	3		116
Passiflora mucronata	subgenus *Passiflora*/supersection *Stipulata*/section *Granadillastrum*	150,839	172	49	97	22	4	372
Passiflora recurva	subgenus *Passiflora*/supersection *Passiflora*/ series *Passiflora*	151,837	158	41	116	1		143
Passiflora watsoniana	subgenus *Passiflora*/supersection *Stipulata*/ section *Kermesinae*	146,520	158	25	128	5		147
Dilkea retusa	*Dilkea* genus	161,923	214	91	110	7	6	106
Mitostemma brevifilis	*Mitostemma* genus	163,032	398	189	209			299

3.3.2. A repertory of rearrangments: inversions, expansions/loss of IR regions and gene losses

The results of the genomic comparison revealed a highly rearranged structure of cpDNAs in *Passiflora*. To detect all rearrangements, *P. trichocarpa* (Salicaceae) was used as a reference because it is a phylogenetically close species with the same cpDNA genome pattern as most angiosperms. The chloroplast genome comparison performed showed different inversions (Figure 9, APPENDIX H). In total, 22 synteny blocks were identified among the cp genomes aligned.

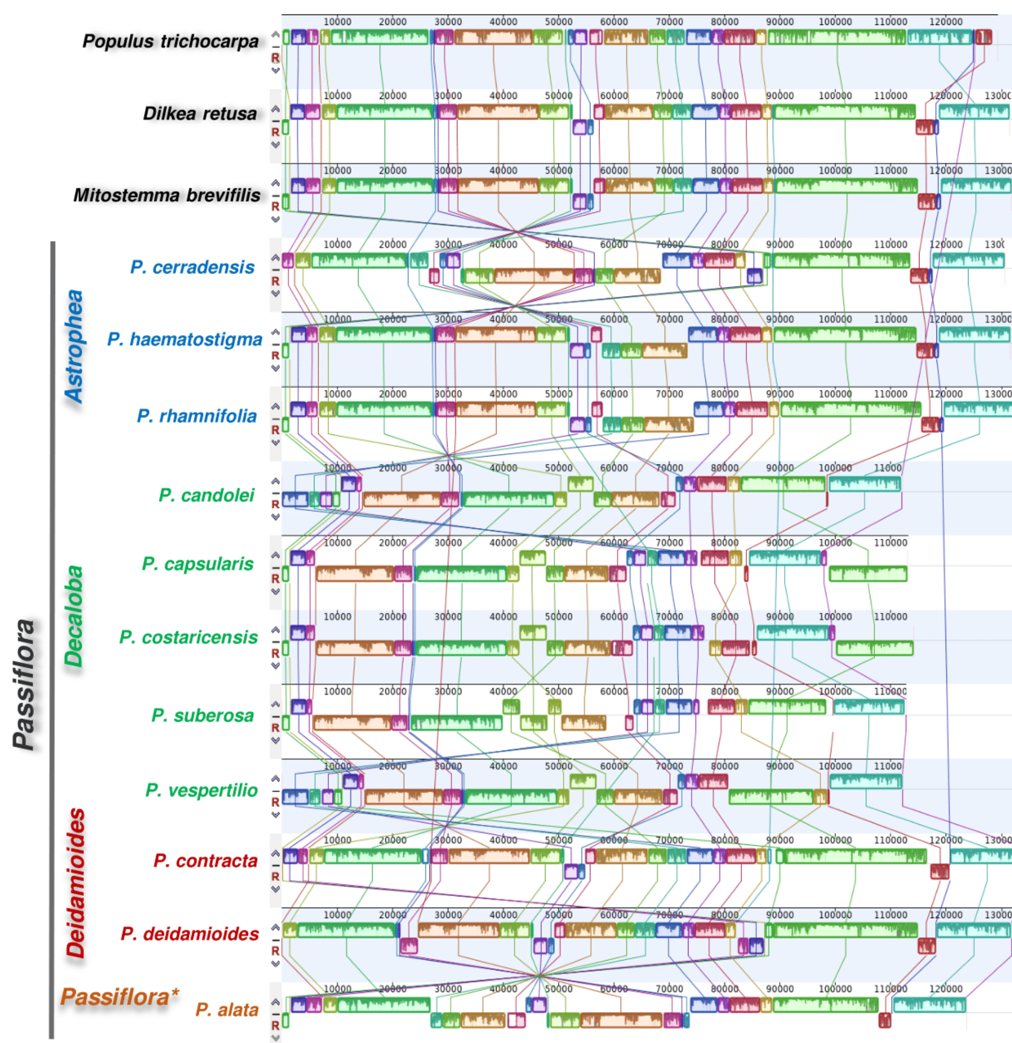


Figure 9. Synteny and structural rearrangements detected in the chloroplast genomes of 11 *Passiflora* species in addition to the Passifloraceae, *Dilkea retusa* and *Mitostemma brevifilis*, and the Salicaceae *Populus. trichocarpa*. Colored bars indicate syntenic blocks and connecting lines indicate the correspondence between blocks.

Inversions were identified in the *Astrophea* subgenus compared to *Dilkea* and *Mitostemma*. They differed only with regard to one inversion (~15 kb) in the LSC region of *P. haematostigma* and *P. rhamnifolia*. In fact, a high level of synteny was found between Passifloraceae *Dilkea* and *Mitostemma,* and Salicaceae *P. trichocarpa*, revealing that fewer rearrangements have occurred, a feature common to most angiosperms. A large inversion in the LSC region flanking the *clpP* gene was found only in *P. cerradensis,* differing from the earlier rearrangement of the *clpP* and *accD* genes in *P. haematostigma* and *P. rhamnifolia*.

The highest number of inversions was found in the *Decaloba* subgenus, with five species exhibiting a large inversion in the LSC region (Figure 9). Small sequence structures were also found inverted in the LSC region of *P. suberosa*. Interestingly, these small inversions are different from the very distinct rearrangement located at the beginning of the LSC region of *P. candollei* and *P verspetilio*, including the *petB* and *clpP* genes. *P. verspetilio* also exhibited a different arrangement in the IR region, with an inversion between the *rpl2* and *rrn5* genes, and this arrangement was confirmed by PCR (APPENDIX E).

The two species of the *Deidamioides* subgenus exhibited a different arrangement in the gene order involving a small region in LSC containing the *matK* gene; in fact, this arrangement occurred by an expansion of the IRs in *P. deidamoides*, as detailed below (Figure 8). *P. deidamoides* also differed from *P. contracta* due to the presence of a small inverted block in the LSC region. Comparing all 20 cp genomes, the species in the subgenus *Passiflora* differed by the presence of a large inversion in the LSC region.

We also examined IR boundaries and detected different expansions and contractions (Figure 10). By comparing them with *P. trichocarpa*, used as a reference, it was possible to detect expansions in the Passifloraceae *D. retusa* and *M. brevifilis*, in which the *ndhH* gene had expanded from the boundary of the SSC region to the IRA region (~ 550 bp) creating a duplicated small fragment copy of the *ndhH* gene in the IRB region. Furthermore, the LSC region contained an expansion including a portion (15 bp) of the *rpl23* sequence up to the IR region, and interestingly this arrangement seems to be particular to *D. retusa* and *M. brevifilis*, and does not occur in the *Passiflora* genus, not even in the reference, *P. trichocarpa*.

In the *Astrophea* subgenus, as in the *Dilkea* and *Mitostemma* genera, an expansion was observed in the SSC/IRA extending into part of the *ndhD* gene, while *P. cerradensis* showed an additional expansion (~ 3.5 kb) of IRA/LSC that includes the *matK* gene, which was confirmed by PCR (APPENDIX E).

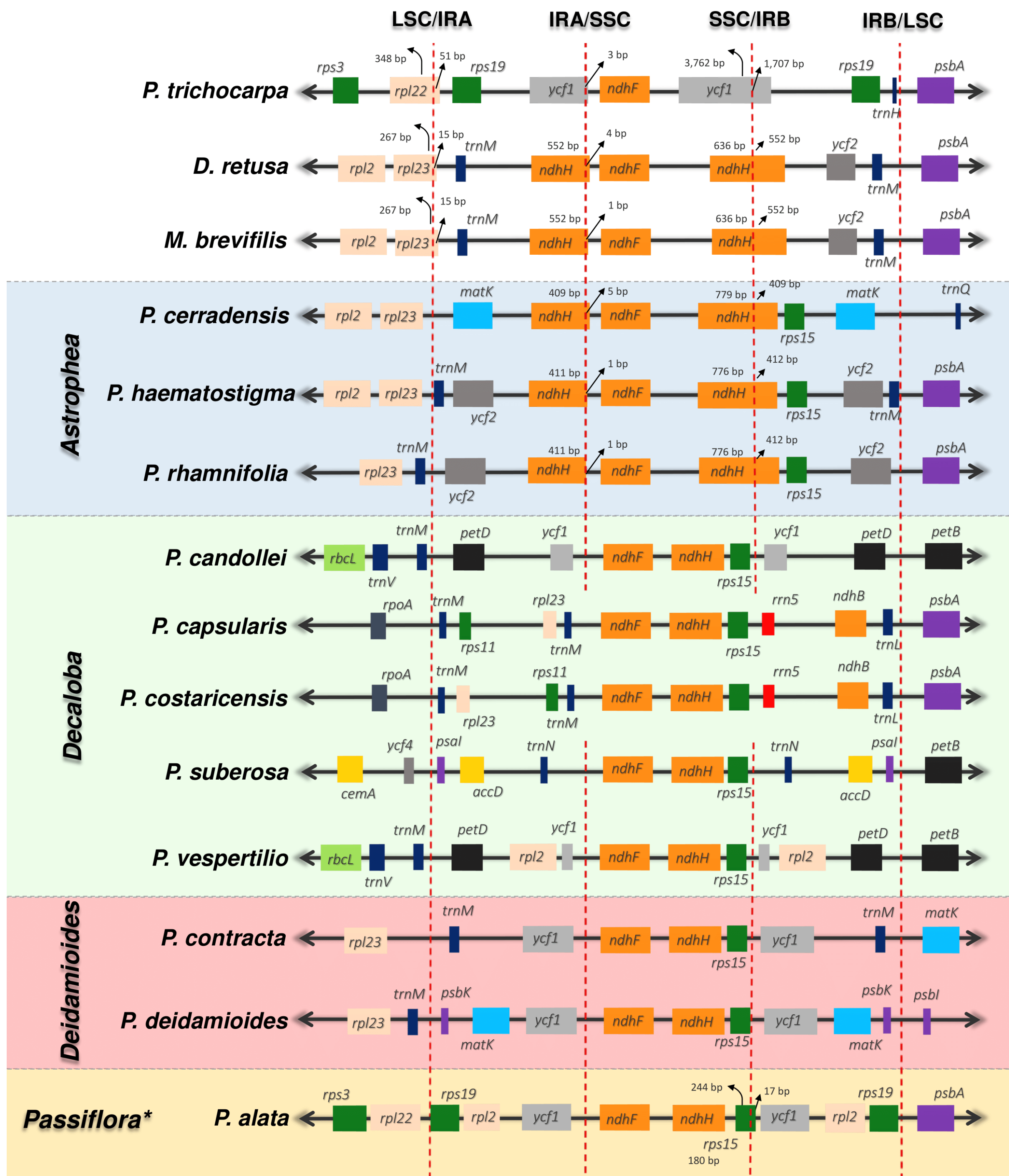


Figure 10. IR boundary comparison of 11 *Passiflora* species in addition to the Passifloraceae, *Dilkea retusa* and *Mitostemma brevifilis*, and the reference species *Populus trichocarpa*. The red dotted lines indicate the border of the cp genome regions, and the colored boxes indicate the gene structure.

Comparing all the species sequenced, *Decaloba* is the subgenus with the highest number of rearrangements related to IR boundaries. For instance, in *P. candollei* and *P. vespertilio,* the expansion of IRA/LSC extends up to the *petD* gene, and in *P. suberosa* this

expansion is even larger, extending up to the *psaI* gene and comprising important genes such as *rbcL*. As a result of this significant expansion, *P. suberosa* has an IR of approximately 43 kb that includes 38 genes, a difference of 17 kb compared to the IRs of other *Decaloba* subgenus species analyzed in this study.

A different expansion of IRB/LSC was observed in *P. deidamioides* (*Deidamioides* subgenus), including the protein coding genes *matK, psbA* and *psbK*, and this expansion was confirmed by amplification via PCR (APPENDIX E). This latter expansion in *P. deidamioides* has an additional fragment length of 3,126 bp compared to the IRB/LSC of *P. contracta* (*Deidamioides* subgenus). Finally, the species that belong to the subgenus *Passiflora* exhibited the smallest rearrangements related to IR boundaries, with part of the *rps15* gene expanding from the SSC to the IRA region (64 bp in *P. edmundoi*).

Comparatively, there were no large variations in the length of the SSC region, and the difference in IR sizes between species was approximately 21.5 kb, whereas in the LSC region it reached 33 kb. These differences were due not only to IR expansions, but also gene and intron losses. All the cp genomes showed gene losses compared to the reference, *P. trichocarpa* (Figure 8). The *rpl22* gene was not found to be complete in any of the species analyzed. Additionally, *accD* was found as a pseudogene, containing just a piece of the sequence or up to 18 stop codons in some species of the different subgenera. Some events may have occurred independently alongside chloroplast genome evolution in the *Passiflora* genus, such as *rpl32* gene loss in some species of the *Decaloba* and *Astrophea* subgenera (Figure 8). In addition, *rpl20* was absent in the cp genomes of the *Astrophea*, *Decaloba* and *Passiflora* subgenera. In the subgenus *Passiflora*, the *rps7*, *ycf1* and *ycf2* genes were found as pseudogenes. Finally, *Decaloba* subgenus exhibited most of the gene losses, over and above the loss of *rps7*, *ycf1* and *ycf2* (the last two were sometimes detected as small pseudogenes), and the loss of *rpoA* was detected in *P. capsularis, P. costaricensis* and *P. suberosa*.

3.3.3. Phylogenomic studies in *Passiflora*

A phylogenomic study was performed based on 68 protein coding genes of 49 Passifloraceae species with available chloroplast genomes. These species of passionflowers were used as the ingroup and *P. trichocarpa* (Salicaceae) was the outgroup to obtain a rooted phylogenetic tree.

The two different partition schemes for Bayesian analysis had no substantial effect on the topology of the resulting trees. The analysis partitioned by codon position was highly favored in relation to the single model based on Bayes Factors (BF= 5448.8), and also showed slightly higher support values in nodes with posterior probabilities (PP) < 1.0. The accuracy of the inferred species phylogeny was strongly supported by the stability of the main clades generated, and the PP and node support values were high, with 97% (42 of 43) reaching 1 (Figure 11).

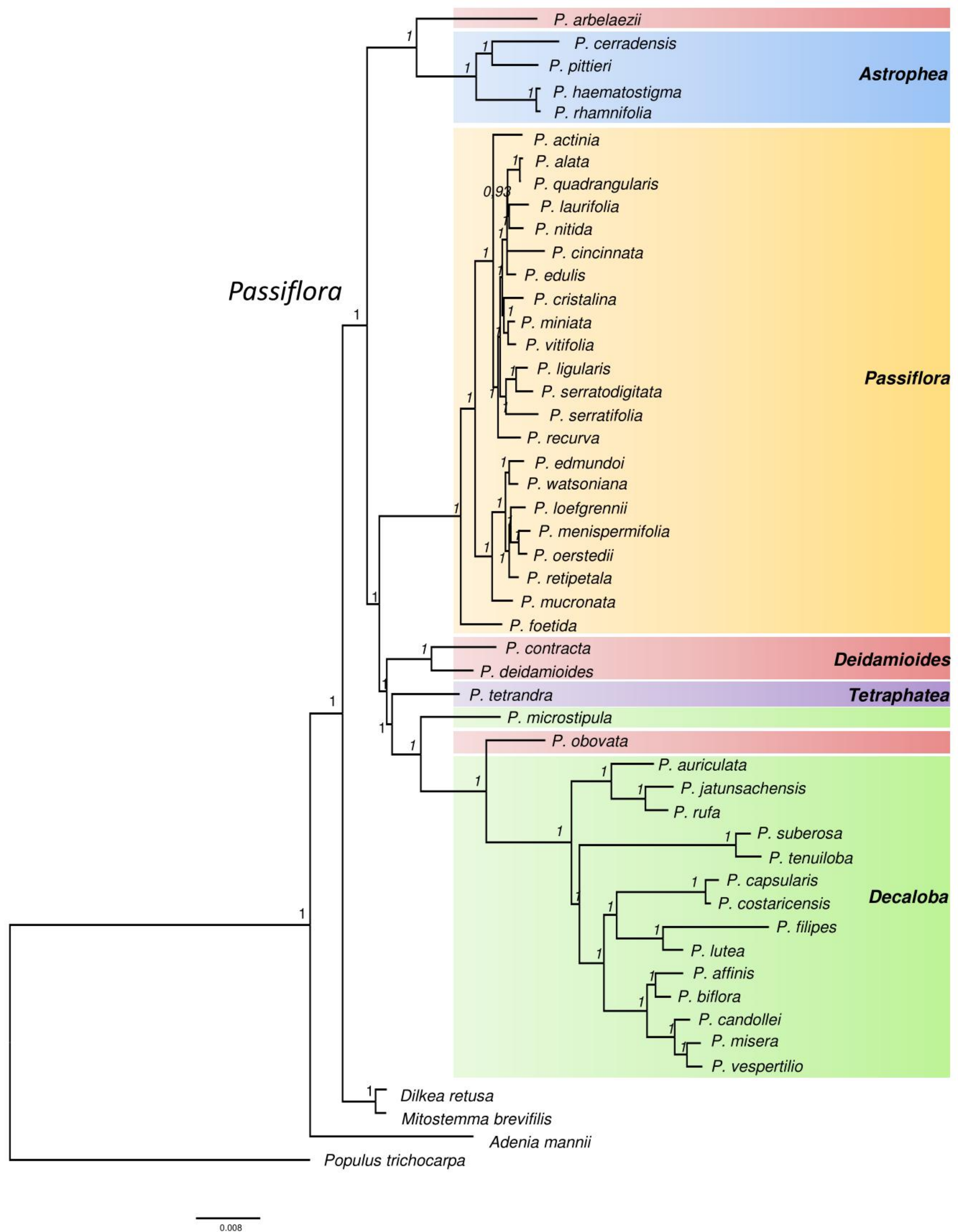


Figure 11. Bayesian phylogenetic reconstruction of *Passiflora* evolutionary history based on 68 chloroplast protein-coding genes.

P. trichocarpa grouped as a sister taxon to *Adenia manni*. In addition, *Dilkea* and *Mitostemma* formed a clade at the position *Adenia* + (*Dilkea* + *Mitostemma)* + *Passiflora*, all

with high support in the phylogenetic tree. The *Astrophea* subgenus was placed on the tree as a monophyletic group of species, sister to the remaining species in the genus. The *Astrophea* clade was supported by high posterior probabilities (PP= 1), and was split into two distinct subclades, one containing *P. haematostigma* and *P. rhamnifolia* (species from section *Pseudastrophea*) and the other grouping *P. cerradensis* and *P. pittieri*. However, *P. arbelaezii* (*Deidamioides*) was grouped as a sister taxon to the *Astrophea* species, dismembering the polyphyletic group of *Deidamioides*.

Deidamioides, as currently defined, is a polyphyletic subgenus, and despite the clade formed by *P. contracta* and *P. deidamioides*, other species assigned to *Deidamioides* were placed in different positions on the tree. *P. arbelaezii* was placed as sister to the *Astrophea* subgenus whereas *P. obovata* (*Deidamioides*) was embedded in the *Decaloba* subgenus. Most species from the *Decaloba* subgenus clustered into a monophyletic clade with high support, having as sucessive sister groups *Tetrapathea* and two species of *Deidamioides* (*P. contracta* and *P. deidamioides*). Subgenus *Passiflora* was also recovered as a monophyletic group, but its sections display paraphyletic patterns. This subgenus contains 236 morphological variable species, which are classified in supersections, sections and series (Feuillet & MacDougal 2003).

3.4. Discussion

3.4.1. A repertory of cpDNA rearrangements concomitant with *Passiflora* evolution and taxonomical implications

The cp genomes analyzed herein have the quadripartite structure common to almost all angiosperms. However, a complete loss of one IR was identified in *P. capsularis* and *P. costaricensis*, both species in the *Xeragona* section within the *Decaloba* subgenus (Espinoza et al. 2018). It is worthy of note that this kind of arrangement has not previously been reported in *Passiflora* cpDNAs (Cauz-Santos et al. 2017; Rabah et al. 2019; Shrestha et al. 2019). Bearing in mind that the loss of a complete IR region is rare in angiosperms, this event may have occurred in *Passiflora* before the diversification of the *Xeragona* section.

The loss of a complete IR has been described in few species of Geraniaceae and Cactaceae (Guisinger et al. 2011; Sanderson et al. 2015; Blazier et al. 2016), and in the large group of Fabaceae (legumes) in which the loss of an IR occurred once, leading to a monophyletic group within the subfamily Papilionideae, designated the IR-lacking clade, or

IRLC (Lavin et al. 1990). It was suggested that this IR deletion could provide a means of testing the traditional evidence used to reconstruct Papilionideae phylogeny (Lavin et al. 1990). In addition, the rates of nucleotide substitution in genes of the IRLC papilionoids are generally higher than those in the IR-containing papilionoids, suggesting an association between the loss of an IR and accelerated nucleotide substitution rates (Schwarz et al. 2017).

Palmer and Thompson hypothesized that IR regions could play a part in the stabilization of cp genome structure (Palmer & Thompson 1982). However, in later studies, the authors reported that the deletion of a complete IR does not always lead to rearrangements or cp genome instability (Palmer et al. 1987). Thus, it was considered to be a different pattern of rearrangement that occurs together with other rearrangements (e.g. inversion, gene loss) in the IRLC (Sabir et al. 2014). Interestingly, this pattern was also found in our study. The loss of a complete IR sequence in a *Xeragona* section species is unique, alongside other different rearrangements that occur in the *Decaloba* subgenus, such as inversions, IR expansions and gene and intron losses.

In *Passiflora*, previous studies revealed the occurrence of paternal inheritance in interspecific crosses (Muschner et al. 2006; Hansen et al. 2007). However, the potential for biparental inheritance in intraspecific crosses involving *P. costaricensis* (*Decaloba* subgenus) has also been reported (Hansen et al. 2007). Most surprisingly, when faced with the cp genome structure of *P. costaricensis*, we noted that our results unveiled the loss of an entire inverted repeat region. In other groups of seed plants that exhibit potential for biparental inheritance, a highly rearranged cpDNA structure also occurs, such as in Campanulaceae (Cosner et al. 2004; Haberle et al. 2008; Barnard-Kubow et al. 2017), and Geraniaceae (Metzlaff et al. 1981; Chumley et al. 2006; Weng et al. 2014).

Recent studies have shown that biparental inheritance could promote chloroplast competition mediated by accelerated rates of evolution in repeats located in the *accD* gene regulatory region (Sobanski et al. 2019). In our study, a high number of repeats within the *accD* gene sequence was observed in all species. Additionally, chloroplast biparental inheritance has the potential to restore cytonuclear incompatibility (Barnard-Kubow et al. 2017; Shrestha et al. 2019), and this type of incompatibility has also been reported in the genus *Passiflora* (Mracek 2005). Further studies are required to evaluate whether changes in the inheritance of chloroplast DNA could lead rearrangements in the nucleotide sequence.

As explained, the evolutionary history of the *Passiflora* chloroplast genome revealed a high number of structural rearrangements (Figure 5). Inversions were detected in the subgenus *Astrophea* (*P. haematostigma*), but surprisingly not in the *Dilkea*, *Mitostemma* and

even *Adenia* genera in the family Passifloraceae (Shrestha et al. 2019). This suggests that the inversions in the *Astrophea* subgenus occurred after the separation of the *Passiflora* genus from its ancestors. The other three genera have fewer species (1 to 102) and exhibit cp genomes more closely resembling that of *P. trichocarpa*. It is possible that the abundance of inversions occurred only in *Passiflora* and not in other Passifloraceae.

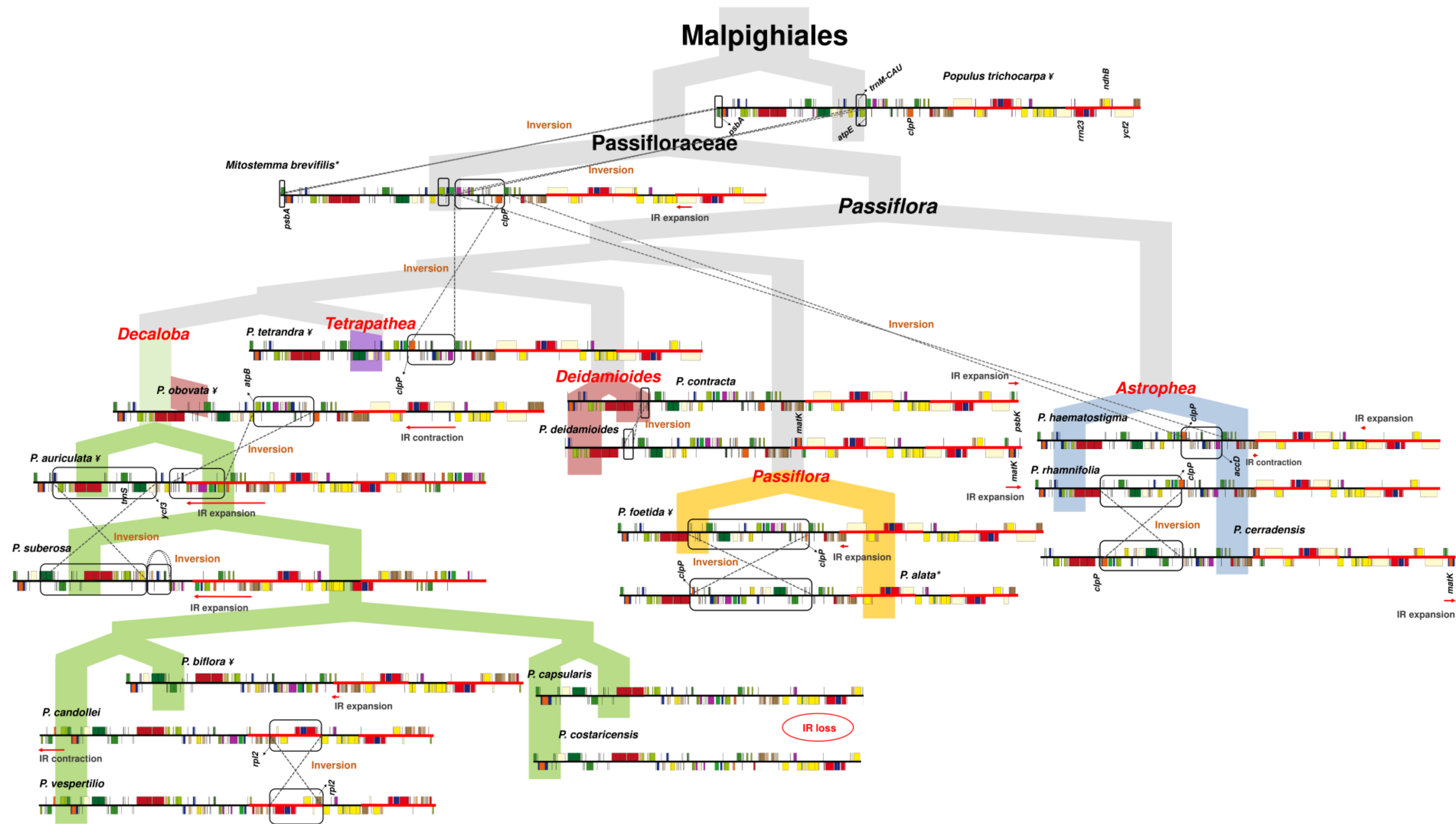


Figure 12. Evolutionary history of chloroplast genome structure in the genus *Passiflora*. The structure of the chloroplast genomes was plotted on a tree representing the evolution of the *Passiflora* genus. In the tree, each subgenus was differentiated by colors: blue for *Astrophea*, green for *Decaloba,* red for *Deidamioides,* orange for *Passiflora,* and violet for *Tetrapathea*. In the comparison of the cp genomes, the boxes and dotted lines indicate the direction of the structural inversions, and the red arrows indicate the direction of IR expansions/contractions.*Indicates that *Mitostemma brevifilis* shares the same structure with *Dilkea retusa*, and the species of the *Passiflora* subgenus obtained in this study share the same structure with *P. alata*. ¥Indicates the cp genomes obtained from the Genbank database.

Some inversions in the *Passiflora* genus seem to be the result of intramolecular recombination of repeats, a mechanism that has been reported to impact the generation of rearrangements (Ogihara et al. 1988; Milligan et al. 1989; Gray et al. 2009; Ruhlman et al. 2017). In our study, short direct-repeat structures in the flanking regions between the *accD* and *clpP* genes were identified in both *P. haematostigma* and *P. rhamnifolia* (*Astrophea*). These kinds of repeats are also present in wheat (Ogihara et al. 1988), Asteraceae (Kim et al. 2005) and Geraniaceae (Guisinger et al. 2011), the latter showing highly rearranged cp genomes similar to the trends observed in *Passiflora*.

In addition, a distinct inversion was found in *P. cerradensis* (*Astrophea*), but since this inversion was flanked by *clpP* and caused repositioning of this gene, in evolutionary terms our results indicate that the rearrangement of *accD* and *clpP* genes in *P. haematostigma* and *P. rhamnifolia* occurred before the unique inversion found in *P. cerradensis*. Interestingly, all the inversions in *Astrophea* are different from those found in the other subgenera. Therefore, it is clear that *Astrophea* underwent further cpDNA changes after its separation from the clade ((*Decaloba, Deidamioides*)*, Passiflora*) (~40 Mya) (Muschner et al. 2012).

As mentioned above, the *Decaloba* subgenus exhibits many different inversions, possibly because of the high number of repeat structures, different modes of chloroplast inheritance, as well as the large IR expansions that are typical of this subgenus. On the other hand, the *Deidamioides* species have a similar cpDNA structure to *Dilkea* and *Mitostemma*, with just one small inversion in the LSC region. However, the *Deidamioides* subgenus differs from the other two Passifloraceae in that it has large IR expansions. Finally, *Passiflora* species exhibit conserved structures and the same rearrangements in the LSC region as those previously described for *P. edulis*, the main cultivated species (Cauz-Santos et al. 2017).

Chloroplast genome IR expansions have been found in different plant groups, including Geraniaceae (Guisinger et al. 2011), Euphorbiaceae (Li et al. 2017), Solanaceae (Amiryousefi et al. 2018) and Bignoniaceae (Thode & Lohmann 2019). *Decaloba* species harbor larger variations, but *P. capsularis* lacks one of the IRs; *P. suberosa* exhibits a large expansion of the IRa/LSC up to the *psaI* gene. The latter IR expansions have already been described in previous studies on the *Decaloba* subgenus (Rabah et al. 2019; Shrestha et al. 2019), except for the lack of IR region in *P. capsularis* and *P. costaricensis*. In *Astrophea* species IRb/SSC has expanded to part of the *ndhH* gene, and remarkably an IRa/LSC expansion was found to include the *matK* gene in *P. deidamioides* and *P. cerradensis*.

However, these expansions do not have a common ancestor and supposedly occurred independently in the *Astrophea* and *Deidamioides* subgenera.

All species of subgenus *Passiflora* exhibited an expansion to the *rps19* gene, like that of *P. trichocarpa*, but this expansion was not found in its sister group, suggesting that independent events occurred in the distinct subgenera. Expansions/contractions and loss of IRs are the main causes of variations in cpDNA sequence length, as in the *Decaloba* subgenus and other plants, such as *Annona* (J. Chris Blazier et al. 2016) and *Lamprocapnos* (Park et al. 2018), as well as in monocots (Wang et al. 2008). Different mechanisms have been proposed to explain IR expansions, such as gene conversion or double-strand DNA breaks (Goulding et al. 1996; Wang et al. 2008).

Distinct genes have been lost during the evolution of cp genomes in the *Passiflora* genus, such as *rpl20* and *rpl22*, the latter absent in all the cp genomes studied herein. Since the chloroplast organelle is responsible for vital processes like photosynthesis, gene loss can impair the efficiency of some metabolic pathways, plant growth and cell survival (Neuhaus & Emes 2000; Kode et al. 2005; Rogalski et al. 2006; Romani et al. 2012). Furthermore, *rpl20* and *rpl22* encode proteins necessary for chloroplast translational apparatus and have been proven to be essential for cell viability in tobacco knockout-mutants (Rogalski et al. 2008; Fleischmann et al. 2011). Similarly, the absence of *rpl22* in legumes is offset by the existence of a functional copy transferred from the chloroplast to the nuclear genome (Gantt et al. 1991).

Gene transfer between chloroplast and nuclear genomes has been described in some species (Millen et al. 2001; Rousseau-Gueutin et al. 2013; Hong et al. 2017), and observed experimentally (Stegemann et al. 2003; Stegemann & Bock 2006; Lloyd & Timmis 2011). Despite the fact that the organelle gene needs a eukaryotic promoter to maintain its functionality in the nuclear genome, studies have revealed that some cp promoters are weakly active in the nucleus, which would render gene transfer viable without a nuclear promoter (Cornelissen & Vandewiele 1989; Wang et al. 2014). The functional transfer of *rpl22* from cp to the nuclear genome has been observed in Fagaceae and it has been suggested that this happens in *Passiflora* (Jansen et al. 2011).

Taxonomically speaking, *Passiflora* is the largest genus of Passifloraceae with approximately 520 species, exhibiting high morphological diversity, especially in flower shape and color, and also in genome size. Although a wide variety of morphological traits has been applied in its traditional taxonomy, the species classification still has unresolved positions, particularly regarding species belonging to the *Deidamioides* subgenus. Molecular

phylogenies did not clarify the position of this subgenus, but they did generally recover *Astrophea, Decaloba* and *Passiflora* as well-supported monophyletic clades based on chloroplast, mitochondrial or nuclear nucleotide sequences (Muschner et al. 2003, 2012; Krosnick et al. 2006, 2013). In addition to the four subgenera described, *Tetrapathea* (Passifloraceae) has been put forward as a new subgenus of *Passiflora*, with three species native to the Old World (Krosnick et al. 2009).

Our findings confirmed the *Astrophea* subgenus as a monophyletic clade, and also grouped together *P. haematostigma* and *P. rhamnifolia*, two species that belong to section *Pseudoastrophea*. Furthermore, the *Astrophea* subgenus was grouped as a sister clade to the species of section *Tryphostemmatoides* in the subgenus *Deidamioides*, confirming previous phylogenetic studies (Muschner et al. 2003, 2012; Krosnick et al. 2013). Based on the divergence times in the *Passiflora* genus, the separations of the clade *Astrophea* plus the *Tryphostemmatoides* section from the other *Passiflora* clades are very ancient, at around 40 Mya (Muschner et al. 2012). Interestingly, some species of the *Astrophea* subgenus are small trees, very different from the lianas that characterize most species of *Passiflora* and *Decaloba*.

The *Deidamioides* subgenus has been recovered as polyphyletic. Only the Brazilian endemic species *P. contracta* and *P. deidamioides* were grouped as a clade. *P. arbelaezii* was sister to the *Astrophea* subgenus, and *P. obovata* placed in the *Decaloba* subgenus, confirming a previously phylogenomic inference (Shrestha et al. 2019). The classical taxonomic studies (Master 1871; Killip 1938) and the recent revision (Feuillet & MacDougal 2003) initially positioned some of these species in the *Deidamioides* subgenus and subsequently the *Decaloba* subgenus. Conversely, *P deidamioides* was initially considered as *Decaloba* (Harms 1923), before the creation of the *Deidamioides* section. *P. obovata* was considered to belong to the subgenus *Plectostemma*, now known as *Decaloba* subgenus (Harms 1923; Killip 1938), and was assigned to the *Deidamioides* subgenus in the classification of Feuillet and MacDougal (Feuillet & MacDougal 2003). However, molecular phylogenies still group this species in the *Decaloba* subgenus (Krosnick et al. 2006, 2013), corroborating our findings. This controversy suggests that the classification of *P. obovata* needs to be revised.

The position of *P. arbelaezii* (*Deidamioides* subgenus, section *Tryphostemmatoides*) as a sister to the *Astrophea* clade in the phylogeny reconstructed herein confirms previous findings (Krosnick et al. 2013). In the past, due to its morphology, *Tryphostemmatoides* was considered a subgenus (Killip 1938), but later it was reduced to a section (Feuillet & MacDougal 2003). Molecular phylogenetic inferences also recovered section

Tryphostemmatoides as a monophyletic clade, although was found to be paraphyletic in relation to other sections of the *Deidamioides* subgenus (Muschner et al. 2012; Krosnick et al. 2013). Our findings also suggest the need for revisiting the taxonomic classification of *Tryphostemmatoides*.

P. tetrandra (subgenus *Tetrapathea*) was grouped as a sister to *Decaloba*, forming the group *Deidamioides* + *Tetrapathea* + *Decaloba*, confirming previous findings (Krosnick et al. 2013; Shrestha et al. 2019).

On the other hand, heteroplasmy, the existence of divergent chloroplasts types, could be problematic and lead to the analysis of paralogous copies in phylogenetic studies (Wolfe & Randle 2004). For this reason, Hansen et al. (2007), discussing the implications of heteroplasmy in *Passiflora,* suggest that caution should be exercised when considering the interpretation of the chloroplast phylogenies. We should also point out that phylogenomics based on cpDNA sequences allow only one view of *Passiflora* evolution, and further reconstructions of phylogenetic trees based on nuclear genes will also be necessary.

Finally, regarding the *Passiflora* subgenus, incongruences were found in the positioning of some species compared with former phylogenetic studies. However, *Passiflora* is the largest subgenus (236 species), and due to the relevance of taxa sampling for phylogeny reconstruction accuracy (see Heath et al. 2008; Nabhan & Sarkar 2012), new analyses of an increased number of species may help elucidate the taxonomy within this subgenus. Interesting results were recently published by Sader et al. (2019) providing a robust and well-resolved time-calibrated phylogeny including some 100 taxa, and by means of phylogenetic comparative methods, they tested the relative importance of polyploidy in *Passiflora* evolution and diversification. According to these authors, changes in chromosome number and genome sizes may have contributed to morphological and ecological traits that explain the pattern of diversification observed in the genus.

3.5. Conclusions

In this study, the analysis of different plastomes revealed highly rearranged structures. These rearrangements, such as inversions, IR expansions/contractions and gene and intron losses, occurred in all the analyzed passionflowers, including species of the four subgenera. We have described, for the first time, the loss of a complete IR region in *Passiflora capsularis* and *Passiflora costaricensis*, both species in the *Xeragona* section within the *Decaloba* subgenus. Bearing in mind that the loss of a complete IR region is rare in

angiosperms, we suggested that this event may have occurred in *Passiflora* before the diversification of the *Xeragona* section.

The largest frequency of rearrangements has happened in the *Decaloba* subgenus. We suggest that gene and intron losses may have occurred multiple times independently alongside the *Passiflora* cp-genome evolution, and that the ancestor cp-genome may possibly have had a structure like those of *Dilkea retusa* and *Mitostemma brevifilis*.

Finally, *Passiflora* could be regarded as an excellent model for studying the evolution of chloroplast genomes, confirming previous reports.

REFERENCES

Amiryousefi A, Hyvönen J, Poczai P. 2018. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae. PLoS One. doi: 10.1371/journal.pone.0196069.

Barnard-Kubow KB, McCoy MA, Galloway LF. 2017. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. New Phytol. doi: 10.1111/nph.14222.

Bedoya AM et al. 2019. Plastid genomes of five species of riverweeds (Podostemaceae): Structural organization and comparative analysis in Malpighiales. Front. Plant Sci. doi: 10.3389/fpls.2019.01035.

Blazier J. Chris et al. 2016. Divergence of RNA polymerase α subunits in angiosperm plastid genomes is mediated by genomic rearrangement. Sci. Rep. doi: 10.1038/srep24595.

Blazier John C. et al. 2016. Variable presence of the inverted repeat and plastome stability in Erodium. Ann. Bot. doi: 10.1093/aob/mcw065.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. doi: 10.1093/bioinformatics/btu170.

Cai Z et al. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J. Mol. Evol. doi: 10.1007/s00239-008-9180-7.

Cauz-Santos LA et al. 2017. The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: Structural organization and phylogenomic studies in Malpighiales. Front. Plant Sci. 8:1–17. doi: 10.3389/fpls.2017.00334.

Chumley TW et al. 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. doi: 10.1093/molbev/msl089.

Cornelissen M, Vandewiele M. 1989. Nuclear transcriptional activity of the tobacco plastid psba promoter. Nucleic Acids Res. doi: 10.1093/nar/17.1.19.

Cosner ME, Raubeson LA, Jansen RK. 2004. Chloroplast DNA rearrangements in Campanulaceae: Phylogenetic utility of highly rearranged genomes. BMC Evol. Biol. doi: 10.1186/1471-2148-4-27.

Darling AE, Mau B, Perna NT. 2010. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 5. doi: 10.1371/journal.pone.0011147.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. doi: 10.1093/nar/gkw955.

Espinoza TEB, Jørgensen PM, MacDougal JM. 2018. A taxonomic revision of *Passiflora* sect. *Xerogona* (Passifloraceae) using principal component analysis. Ann. Missouri Bot. Gard. doi: 10.3417/2017055.

Feuillet C, MacDougal JM. 2003. A new infrageneric classification of *Passiflora* L. (Passifloraceae). J. Newsl. Passiflora Soc. Int. 13:34–35.

Fleischmann TT et al. 2011. Nonessential plastid-encoded ribosomal proteins in tobacco: A developmental role for plastid translation and implications for reductive genome evolution. Plant Cell. doi: 10.1105/tpc.111.088906.

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. 1991. Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J. doi: 10.1002/j.1460-2075.1991.tb07859.x.

Goulding SE, Olmstead RG, Morden CW, Wolfe KH. 1996. Ebb and flow of the chloroplast inverted repeat. Mol. Gen. Genet. doi: 10.1007/BF02173220.

Gray BN, Ahner BA, Hanson MR. 2009. Extensive homologous recombination between introduced and native regulatory plastid DNA elements in transplastomic plants. Transgenic Res. doi: 10.1007/s11248-009-9246-3.

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. doi: 10.1093/nar/gkz238.

Guisinger MM, Kuehl J V., Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. Mol. Biol. Evol. doi: 10.1093/molbev/msq229.

Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. J. Mol. Evol. doi: 10.1007/s00239-008-9086-4.

Hansen a. K et al. 2006. Phylogenetic relationships and chromosome number evolution in *Passiflora*. Syst. Bot. 31:138–150. doi: 10.1600/036364406775971769.

Hansen AK, Escobar LK, Gilbert LE, Jansen RK. 2007. Paternal, maternal, and biparental inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): Implications for phylogenetic studies. Am. J. Bot. doi: 10.3732/ajb.94.1.42.

Harms H. 1923. Beiträge zur Kenntnis der amerikanischen Passifloraceen. I. Repert. Nov. specierum regni Veg. doi: 10.1002/fedr.4870190106.

Heath TA, Hedtke SM, Hillis DM. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. doi: 10.3724/SP.J.1002.2008.08016.

Hong CP et al. 2017. *accD* nuclear transfer of *Platycodon grandiflorum* and the plastid of early Campanulaceae. BMC Genomics. doi: 10.1186/s12864-017-4014-x.

Jansen RK, Saski C, Lee SB, Hansen AK, Daniell H. 2011. Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): Evidence for at least two independent transfers of *rpl22* to the nucleus. Mol. Biol. Evol. doi: 10.1093/molbev/msq261.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods. doi: 10.1038/nmeth.4285.

Killip EP. 1938. *The American species of Passifloraceae*. Chicago Field Museum.: Chicago.

Kim KJ, Choi KS, Jansen RK. 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). Mol. Biol. Evol. doi: 10.1093/molbev/msi174.

Knox EB. 2014. The dynamic history of plastid genomes in the Campanulaceae *sensu lato* is unique among angiosperms. Proc. Natl. Acad. Sci. U. S. A. doi:10.1073/pnas.1403363111.

Kode V, Mudd EA, Iamtham S, Day A. 2005. The tobacco plastid *accD* gene is essential and is required for leaf development. Plant J. doi: 10.1111/j.1365-313X.2005.02533.x.

Koren S et al. 2017. Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. Genome Res. doi: 10.1101/gr.215087.116.

Krosnick SE, Ford AJ, Freudenstein J V. 2009. Taxonomic Revision of *Passiflora* Subgenus *Tetrapathea* Including the Monotypic Genera *Hollrungia* and *Tetrapathea*

(Passifloraceae), and a New Species of *Passiflora*. Syst. Bot. doi: 10.1600/036364409788606343.

Krosnick SE, Harris EM, Freudenstein J V. 2006. Patterns of anomalous floral development in the Asian *Passiflora* (subgenus *Decaloba*: supersection *Disemma*). Am. J. Bot. doi: 10.3732/ajb.93.4.620.

Krosnick SE, Porter-Utley KE, MacDougal JM, Jørgensen PM, McDade LA. 2013. New Insights into the Evolution of *Passiflora* subgenus *Decaloba* (Passifloraceae): Phylogenetic Relationships and Morphological Synapomorphies. Syst. Bot. doi: 10.1600/036364413x670359.

Kurtz S et al. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29:4633–4642. doi: http://dx.doi.org/10.1093/nar/29.22.4633.

Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32:11–16. doi: 10.1093/nar/gkh152.

Lavin M, Doyle JJ, Palmer JD. 1990. Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. Evolution (N. Y). doi: 10.2307/2409416.

Li Z et al. 2017. The complete chloroplast genome sequence of tung tree (*Vernicia fordii*): Organization and phylogenetic relationships with other angiosperms. Sci. Rep. doi: 10.1038/s41598-017-02076-6.

Lloyd AH, Timmis JN. 2011. The origin and characterization of new nuclear genes originating from a cytoplasmic organellar genome. Mol. Biol. Evol. doi: 10.1093/molbev/msr021.

Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. doi: 10.1093/nar/gkw413.

Master MT. 1871. Contributions to the Natural History of the Passifloraceea. Trans. Linn. Soc. London. 27:626–629.

Metzlaff M, Börner T, Hagemann R. 1981. Variations of chloroplast DNAs in the genus *Pelargonium* and their biparental inheritance. Theor. Appl. Genet. doi: 10.1007/BF00275175.

Millen RS et al. 2001. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell. 13:645–658. doi: 10.1105/tpc.13.3.645.

Milligan BG, Hampton JN, Palmer JD. 1989. Dispersed repeats and structural reorganization in subclover chloroplast DNA. Mol. Biol. Evol. doi: 10.1093/oxfordjournals.molbev.a040558.

Mracek J. 2005. Investigation of interspecific genome-plastome incompatibility in *Oenothera* and *Passiflora*. Dr. thesis. 1–118..

Muschner VC et al. 2003. A first molecular phylogenetic analysis of *Passiflora* (Passifloraceae). Am. J. Bot. 90:1229–1238. doi: 10.3732/ajb.90.8.1229.

Muschner VC et al. 2006. Differential organellar inheritance in Passiflora's (Passifloraceae) subgenera. Genetica. doi: 10.1007/s10709-006-7726-4.

Muschner VC, Zamberlan PM, Bonatto SL, Freitas LB. 2012. Phylogeny, biogeography and divergence times in *Passiflora* (Passifloraceae). Genet. Mol. Biol. 35:1036–1043. doi: 10.1590/S1415-47572012000600019.

Nabhan AR, Sarkar IN. 2012. The impact of taxon sampling on phylogenetic inference: A review of two decades of controversy. Brief. Bioinform. doi: 10.1093/bib/bbr014.

Neuhaus HE, Emes MJ. 2000. Nonphotosynthetic metabolism in plastids. Annu. Rev. Plant Biol. 51:111–140.

Ogihara Y, Terachi T, Sasakuma T. 1988. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc. Natl. Acad. Sci. U. S. A. doi: 10.1073/pnas.85.22.8573.

Palmer JD, Osorio B, Aldrich J, Thompson WF. 1987. Chloroplast DNA evolution among legumes: Loss of a large inverted repeat occurred prior to other sequence rearrangements. Curr. Genet. doi: 10.1007/BF00355401.

Palmer JD, Thompson WF. 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell. doi: 10.1016/0092-8674(82)90170-2.

Park S, An B, Park SJ. 2018. Reconfiguration of the plastid genome in *Lamprocapnos spectabilis*: IR boundary shifting, inversion, and intraspecific variation. Sci. Rep. doi: 10.1038/s41598-018-31938-w.

Rabah SO et al. 2019. *Passiflora* plastome sequencing reveals widespread genomic rearrangements. J. Syst. Evol. doi: 10.1111/jse.12425.

Rogalski M, Ruf S, Bock R. 2006. Tobacco plastid ribosomal protein S18 is essential for cell survival. Nucleic Acids Res. doi: 10.1093/nar/gkl634.

Rogalski M, Schöttler MA, Thiele W, Schulze WX, Bock R. 2008. *Rpl33*, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. Plant Cell. doi: 10.1105/tpc.108.060392.

Romani I et al. 2012. Versatile roles of Arabidopsis plastid ribosomal proteins in plant growth and development. Plant J. doi: 10.1111/tpj.12000.

Ronquist F et al. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. doi: 10.1093/sysbio/sys029.

Rousseau-Gueutin M et al. 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. doi: 10.1104/pp.113.214528.

Ruhlman TA, Zhang J, Blazier JC, Sabir JSM, Jansen RK. 2017. Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. Am. J. Bot. doi: 10.3732/ajb.1600453.

Sader, MA et al. 2019. The role of chromosome changes in the diversification of *Passiflora* L. (Passifloraceae). Syst Biodivers Syst Biodivers doi: 10.1080/14772000.2018.1546777.

Sabir J et al. 2014. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. Plant Biotechnol. J. doi: 10.1111/pbi.12179.

Sanderson MJ et al. 2015. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the ndh gene suite and inverted repeat 1. Am. J. Bot. doi: 10.3732/ajb.1500184.

Schwarz EN et al. 2017. Plastome-wide nucleotide substitution rates reveal accelerated rates in Papilionoideae and correlations with genome features across legume subfamilies. J. Mol. Evol. doi: 10.1007/s00239-017-9792-x.

Shrestha B et al. 2019. Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. Mol. Phylogenet. Evol. doi: 10.1016/j.ympev.2019.05.030.

Sobanski J et al. 2019. Chloroplast competition is controlled by lipid biosynthesis in evening primroses. Proc. Natl. Acad. Sci. U. S. A. doi: 10.1073/pnas.1811661116.

Stegemann S, Bock R. 2006. Experimental reconstruction of functional gene transfer from the tobacco plastid geneome to the nucleus. Plant Cell. doi: 10.1105/tpc.106.046466.

Stegemann S, Hartmann S, Ruf S, Bock R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. Proc. Natl. Acad. Sci. U. S. A. 100:8828–8833. doi: 10.1073/pnas.1430924100.

Sugiura M. 1992. The chloroplast genome. Plant Mol. Biol. 19:149–168.

Takamatsu T et al. 2018. Optimized method of extracting rice chloroplast DNA for high-quality plastome resequencing and de novo assembly. Front. Plant Sci. 9:266. doi: 10.3389/FPLS.2018.00266.

Tangphatsornruang S et al. 2011. Characterization of the complete chloroplast genome of Hevea brasiliensis reveals genome rearrangement, RNA editing sites and phylogenetic relationships. Gene. 475:104–112. doi: 10.1016/j.gene.2011.01.002.

Thode VA, Lohmann LG. 2019. Comparative chloroplast genomics at low taxonomic levels: A case study using *Amphilophium* (bignonieae, bignoniaceae). Front. Plant Sci. doi: 10.3389/fpls.2019.00796.

Tillich M et al. 2017. GeSeq - Versatile and accurate annotation of organelle genomes. Nucleic Acids Res. doi: 10.1093/nar/gkx391.

Ulmer T, MacDougal JM. 2004. *Passiflora: passionflowers of the world.* Timber Press: Portland.

Wang D, Qu Z, Adelson DL, Zhu JK, Timmis JN. 2014. Transcription of nuclear organellar DNA in a model plant system. Genome Biol. Evol. doi: 10.1093/gbe/evu111.

Wang RJ et al. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol. Biol. doi: 10.1186/1471-2148-8-36.

Weng ML, Blazier JC, Govindu M, Jansen RK. 2014. Reconstruction of the ancestral plastid genome in geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol. Biol. Evol. doi: 10.1093/molbev/mst257.

Wolfe AD, Randle CP. 2004. Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: Implications for plant molecular systematics. Syst. Bot. doi: 10.1600/0363644042451008.

Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM. 2011. Comparative chloroplast genomes of Pinaceae: Insights into the mechanism of diversified genomic organizations. Genome Biol. Evol. doi: 10.1093/gbe/evr026.

Wu CS, Wang YN, Hsu CY, Lin CP, Chaw SM. 2011. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. Genome Biol. Evol. doi: 10.1093/gbe/evr095.

Yang M et al. 2010. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). PLoS One. 5:1–14. doi: 10.1371/journal.pone.0012762.

4. PLASTID PHYLOGENOMICS SUPPORT A NEW INFRAGENERIC CLASSIFICATION IN *PASSIFLORA*

ABSTRACT

Passiflora is the richest genus of the Passifloraceae, characterized by a great diversity in flower size and shape, and these attributes have been used for establishing the taxonomy of this group. However, despite previously morphological and molecular classification, the relationships of some *Passiflora* species remain unsolved. In this study, to increase knowledge on the *Passiflora* evolutionary history, we have obtained the complete chloroplast genome (plastomes) across the different *Passiflora* subgenera in order to reconstruct a new phylogeny. A large variation in plastome size (~ 40Kb) was found, mainly due to expansions of the inverted repeat (IR) regions. In addition, a comparative genomics revealed different organization structures for plastomes in accordance with the taxonomical classification within the subgenus *Deidamioides*. Our study resulted in a strongly supported tree reinforcing the monophyly of the main clades comprising the subgenera *Astrophea*, *Decaloba* and *Passiflora*; however, a polyphyletic placement of species belonging to *Deidamioides* is sustained. A nuclear phylogeny based on complete 18S/26S (35S rDNA) sequences was congruent with the plastid topologies although lower values of support were found. Based on the combined results, we suggest elevating the status of the section *Tryphostematoides* (*Deidamioides*) to subgenus *Tryphostematoides*.

Keywords: Chloroplast genome; Genome rearrangements; Phylogenomics, 35S rDNA, *Passiflora*.

4.1. Introduction

Passifloraceae belongs to the order Malpighiales and is composed of about 950 species of herbaceous or woody vines, shrubs and trees, known as passionflowers (Tokuoka, 2012), and have emerged at 65.5 Mya (Muschner et al., 2012). The richest genus is *Passiflora*, with about 520 described species with great diversity in size and shape of flowers, and a geographic distribution particularly neotropical, being found mainly in the Americas; however, there are occurrences in Southeast Asia, Oceania and Australia (Ulmer and MacDougal, 2004).

Taxonomic studies in *Passiflora* based on morphological characters resulted in an infrageneric classification into 22 subgenera (Killip, 1938). Only after ca. 60 years later the taxonomy of the genus was revisited. Ulmer & MacDougal (2004), taking also into account morphological traits, drastically reduced the Killip's subdivision, proposing four subgenera: *Astrophea*, *Decaloba*, *Deidamioides* and *Passiflora*. Considering the species distribution, subgenus *Passiflora* has the largest number of species (approximately 236) distributed into six supersections and 24 sections, while the subgenus *Deidamioides* is the one with the lowest number of species (13). Recently, Ocampo Pérez & Coppens d'Eeckenbrugge (2017) confirmed the Ulmer & MacDougal's subdivision and the occurrence of subgenera taking into

account morphological traits as descriptors, among them, the size of stems, leaves, and flowers, the shape of bracts, the number and distribution of extrafloral nectaries, the presence of tendrils, the androgynophore sizes and the complexity of corona which, in some species, can show more than three rows of filaments. However, this study pointed out the fragile positioning of the subgenus *Deidamioides*.

The first classification of *Passiflora* based on molecular data using both the nuclear ribosomal internal transcribed spacer (ITS) and plastid *trnL-trnF* intergenic spacer sequences, recovered the monophyly of three subgenera (Muschner et al., 2003) while *Deidamioides* appeared as paraphyletic. The same infrageneric positioning was observed by Yockteng & Nadot (2004) using the nuclear chloroplast-expressed glutamine synthetase gene, *ncpGS*. Subsequently, Muschner et al. (2012), in a phylogenetic and biogeographic study in *Passiflora* using nucleotide sequences from different genomes (plastid *rbcL*, *rps4*, *trnL* intron, *trnL-trnF* spacer, mitochondrial *nad1*, *nad5* and the nuclear portion of the 26S gene), also suggested the paraphyletic position of *Deidamioides*; in the phylogenetic reconstruction, the *Tryphostemmatoides* section appeared as a sister to the *Astrophea* clade.

Krosnick et al. (2009) proposed the addition of a new subgenus, *Tetrapathea,* consisting of three species from Oceania. Later on, the same research group on the basis of nuclear ITS, *ncpGS* gene and plastid *trnF-trnL*, *ndhF* sequences was not able to resolve the correct positioning of the subgenus *Deidamioides* (Krosnick et al., 2013).

The advances provided by the next generation sequencing (NGS) approaches permitted that phylogenomic studies emerged trying to solve the relationships unresolved by using few molecular markers. In this way, chloroplast genomes, a molecule of 120 to 160 kb in length, which encodes approximately 80 unique protein-coding genes, can be applied to generate robust phylogenies (Parks et al., 2009; Song et al., 2019). Therefore, plastid phylogenomics have been extensively used to solve the relationships in different plant groups (Marinho et al., 2019; Moore et al., 2010, 2007; Xi et al., 2012; Zhang et al., 2017). In this scenario, we suggested that this kind of data could be useful to get new insights on the relationships in *Passiflora* (Cauz-Santos et al., 2020).

Therefore, we continue with the study of the *Passiflora* evolutionary history generating 14 additional new plastomes from the four subgenera, but including *Tetrapathea*. A structural comparison between the new and previously described plastomes of *Deidamioides* species was conducted to answer the question: Could chloroplast genome structure may be used for revisiting the taxonomic classification of *Deidamioides*?

In addition, plastid phylogenomics and a phylogenetic study based on the complete 18S/26S rDNA sequences were conducted with the aim of: (i) discovering the position of *Deidamioides* species in the *Passiflora* phylogeny; (ii) testing the utility of 18S/26S sequences for establishing a high-support phylogeny; (iii) testing the congruence of the reconstructed trees by comparing the results based on nuclear and plastid DNA data.

4.2. Material and Methods

4.2.1. Taxon sampling for chloroplast genome sequencing

The present study includes additional 14 species comprising all *Passiflora* subgenera: *Astrophea* (1). *Decaloba* (4), *Deidamioides* (2), *Passiflora* (6) and *Tetrapathea* (1). Most of the species were obtained from the Italian National Collection of *Passiflora*, Ripalta Cremasca, Italy. In addition, *Passiflora organensis* (*Decaloba*) was field-collected in the Atlantic forest biome of the São Paulo State, Brazil, which is registered at Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (SISGEN, Brazil).

4.2.2. Chloroplast (cp) DNA obtainment

Prior to cpDNA extraction, the intact isolation of chloroplast organelles was performed based on liquid nitrogen-sucrose gradient method. Subsequently, the resulted chloroplast pellet was lysed in 2% CTAB buffer, and the protocol was followed by two extractions using chloroform:isoamyl alcohol (24:1), an isopropanol precipitation, and finally washing the pellet in an ethanol solution (70%), that was then dried and resuspended in 40 µL of TE buffer.

4.2.3. CpDNA sequencing and assembly

The cpDNA samples were sequenced using the PacBio SMART strategy that has the advantage of generating long reads. The large-insert (10 kb) libraries were constructed using the Barcode method using 150 ng of pure high molecular weight DNA. The fragment sizes, quality and DNA concentration in the libraries were checked in Fragment Analyzer (Agilent technologies) and Qubit 2.0 Fluorometer (Invitrogen). Sequencing was performed in two

SMRT cells using P6 polymerase with C4 chemistry on the PacBio RSII plataform at the NGI Platform (Uppsala, Sweden).

The raw data was first demultiplexed and subsequently filtered by quality (reads with quality < 0.75 and length < 500 bp were removed). Finally, sequences were assembled and corrected using the CANU assembler (Koren et al., 2017).

4.2.4. Chloroplast genome annotation and comparative analysis

The new complete plastomes were annotated using the GeSeq (Organellar Genome Annotation) online program with default settings in order to identify protein-coding gene sequences (CDS), rRNAs and tRNAs based on cp reference sequences and BLAST homology searches (Tillich et al., 2017). The annotation was followed by manual corrections for start and stop codons, and intron positions in the GenomeView software; the OGDRAW software was used for constructing the circular and linear plastome maps (Abeel et al., 2012; Greiner et al., 2019).

A multiple sequence alignment was conducted using only the species of *Deidamioides* to ecaluate the sinteny levels and identify possible rearrangements. This run was performed in progressive Mauve v.2.4.0 (Darling et al., 2010) and in combination with the new plastomes, the cpDNA sequences previously obtained for *P. arbelaezii* and *P. obovata* (Shrestha et al., 2019), and *P. contracta* and *P. deidamioides* (Cauz-Santos et al., 2020) were included. In this analysis, the *Populus trichocarpa* (Salicaceae) plastome was used as a reference.

4.2.5. Plastid phylogenomic studies

A new phylogenomic study based on all the available plastomes was conducted (63 species in total, 14 generated in this study, 20 described in the 2nd chapter of this thesis and 28 species whose cp DNA sequences were imported from NCBI's Genbank. In addition, to obtain a rooted tree, *Populus trichocarpa* (Salicaceae) plastome was used as outgroup (APPENDIX I).

A set of 68 chloroplast protein-coding genes was also used in the new phylogenetic analysis: *atpA, atpB, atpE, atpF, atpH, atpI, ccsA, cemA, clpP, matK, ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK, petA, petB, petD, petG, petL, petN, psaA*

psaB, psaC, psaI, psaJ, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, rbcL, rpl2, rpl14, rpl16, rpl23, rpl33, rpl36, rpoB, rpoC1, rpoC2, rps2, rps3, rps4, rps8, rps11, rps12, rps14, rps15, rps19, ycf3, ycf4.

The gene sequences were extracted from our data set (consisting of 64 taxa), aligned individually at nucleotide level in MUSCLE (Edgar, 2004), and then all the individual alignments contatenated in an interleaved Nexus matrix. The resulting matrix was then analyzed in ModelFinder (Kalyaanamoorthy et al., 2017) to determine the best partition scheme and evolutionary models according to Akaike Information Criterion (AIC), and treating each gene alignment as a charset.

The Maximum Likelihood (ML) analysis was performed using RAxML version 8.2.4 (Stamatakis, 2014). The GTR+G+I model of nucleotide substitution was selected for whole dataset, and a bootstrap analysis was performed on 1,000 replicates. In the Bayesian inference (BI), the Markov chain Monte Carlo (MCMC) algorithm was run in MrBayes (Ronquist et al., 2012) under the GTR+G+I model of nucleotide substitution, 10,000,000 generations, sampling one tree every 1,000 generations and discarding the first 25% of trees as burn-in, to estimate the values of posterior probabilities. The convergence of the runs was monitored based on an average standard deviation of split frequencies below 0.01, potential scale reduction factor (PSRF) values close to 1.0, and ESS values above 1.0. Subsequently, the phylogenetic trees were visualized using FigTree version 1.4.4 (Rambaut, 2016).

Finally, a phylogenomic study using the complete plastomes of 24 species of subgenus *Passiflora* was performed (APPENDIX I). The sequences were aligned at nucleotide level in server-based program MAFFT version 7.221, using the FFT-NS-2 algorithm with default settings (Katoh and Standley, 2013). The substitution model for BI was estimated in ModelFinder (Kalyaanamoorthy et al., 2017) using AIC. The Bayesian inference was performed in MrBayes, 10,000,000 generations, sampling one tree every 1,000 generations and discarding the first 25% of trees as burn-in. The convergence of the runs was monitored, and the phylogenetic trees were visualized using FigTree version 1.4.4 (Rambaut, 2016).

4.2.6. 35S rDNA sequence assembly and phylogenetic analysis

The complete 35S rDNA sequence was obtained by genome skimming from the sequencing data of 33 *Passiflora* species: 20 of them sequenced in a previous study using the Illumina next generation sequencing strategy (Cauz-Santos et al., 2020), and 13 sequenced in

this study using the PacBio sequencing approach (APPENDIX I). The *de novo* assembly of the 35S for each species was performed in NovoPlasty (Dierckxsens et al., 2017) with regard to the filtered reads generated by Illumina, and in Geneious with regard to filtered reads generated by Pacbio. The annotation was conducted in Geneious and the sequences were extracted to create two datasets and infer the phylogenies based on 18S and 26S complete sequences, respectively.

Sequences were processed in a local pipeline, aligned individually in MUSCLE software (Edgar, 2004), and concatenated in an interleaved matrix. The best evolutionary model was estimated in ModelFinder (Kalyaanamoorthy et al., 2017) in accordance with AIC. The phylogenetic reconstruction was performed based on maximum likelihood (ML) analysis in RA×ML v.8.2.4 software (Stamatakis, 2014). The GTR+G+I substitution model was selected to infer the phylogenetic relationships, and to test the consistency of ML-based phylogenies, we performed a non-parametric bootstrap analysis with 1000 repetitions.

The Bayesian inference was performed using the MrBayes v.3.2.5 software (Ronquist et al., 2012) with the GTR+G+I evolutionary model. Markov chain algorithm (MCMC) was conducted with 10,000,000 generations, sampling one tree every 100 generations, with the first 25% of the samples of discarded trees. The convergence of the analysis was ensured by an average standard deviation of frequencies below 0.01, effective sample size (EES) of all parameters above 200 and the potential scale reduction factor (PSRF) values close to 1.0. All trees were visualized and edited using the program FigTree v.1.4.3 (Rambaut, 2016).

4.3. Results

4.3.1. Chloroplast genome structural features and gene content

The *Passiflora* complete chloroplast genomes investigated herein have the typical quadripartite structure consisting of a large single copy (LSC) region and a small single copy (SSC) region separated by two long inverted repeats (IRs) (APPENDIX K). This organization is found in most angiosperms. A GC content ranging from 36,2% to 37,5% (Table 7) was found. The sizes of the genomes ranged from 132,736 bp to 173,095 bp in *Passiflora adenopoda* (*Decaloba*) and *Passiflora discophora* (*Deidamioides*), respectively. However, while the length of the SSC region was approximately 13 kb for all species sequenced, there was a large variation in the LSC and IR regions. A difference of ~40 kb was observed in the

LSC regions, from the lowest 47,752 bp in *P. intricata* (*Decaloba*) up to 89,230 in *P. rusbyi* (*Astrophea*). Interestingly, comparing the IR regions, the variation was mainly the result of expansions. The major divergence was found in *Decaloba* with a difference of ~25 kb between *P. adenopoda* (19,852 bp) and *P. intricata* (55,553 bp). The cp genome annotation resulted in the identification of 103 and 109 unique genes representing 70 to 75 protein-coding genes, 30 tRNA and 4 rRNA genes.

Table 7. Summary of the 14 *Passiflora* chloroplast genomes investigated.

Subgenus	Species	Cp genome size (bp)	LSC (bp)	SSC (bp)	IR (bp)	GC content %
Astrophea	*Passiflora rusbyi*	163,292	89,230	12,840	30,611	36,6
Decaloba	*Passiflora adenopoda*	132,736	79,859	13,173	19,852	36,9
	Passiflora intricata	171,622	47,752	12,764	55,553	36,5
	Passiflora organensis	147,532	64,505	13,157	34,935	36,6
	Passiflora xiikzozdz	158,237	57,611	13,174	43,726	37,2
Deidamioides	*Passiflora discophora*	173,095	88,767	12,034	35,647	37,0
	Passiflora ovalis	169,222	88,313	13,715	33,597	36,7
Passiflora	*Passiflora chaparensis*	150,928	85,062	13,492	26,187	37,1
	Passiflora garckei	146,299	86,007	13,240	23,526	37,1
	Passiflora palenquensis	149,524	85,083	13,495	25,473	36,9
	Passiflora phoenicea	148,089	85,806	13,495	24,394	36,9
	Passiflora popenovii	147,278	84,239	13,485	24,777	37,0
	Passiflora racemosa	159,892	86,642	13,940	29,625	37,5
Tetrapathea	*Passiflora tetrandra*	160,883	87,005	13,550	30,164	36,2

4.3.2. Comparative studies in *Deidamioides* based on plastome structures

The comparative studies revealed 12 synteny blocks or conserved regions within the *Deidamioides* cp genomes aligned. However, rearrangements characterized by inversions, IR expansions and contractions were detected in this group (Figure 13).

Considering the taxonomical division into the subgenus *Deidamioides*, *P. arbelaezzi* and *P. discophora* are included in section *Tryphostemmatoides*; *P. contracta* and *P. ovalis* are from section *Tetrastylis*; *P. obovata* is from *Myapathanthus*, and *P. deidamioides* is the unique species classified into section *Deidamioides.* In our results, all species from *Deidamioides* share a small inversion in the LSC region (~3kb), however, differences between the plastome structures in each section were clearly observed. The *Tryphostemmatoides* species have a collinearity in their sequences; however, it is the only

section that presents a large inversion (~20 kb) in the LSC region between the *accD* and *clpP* genes, differing from the other *Deidamioides*.

The two species of section *Tetrastylis* have an expansion of IRb to LSC region up to the *psbA* gene, while *P. deidamioides* expanded the IRb to LSC region up to the *psbK* gene, a difference of 3 kb more than that expansion found in the *Tetrastylis* species. *P. deidamioides* also differs from all the others by a small inversion in the LSC region. Finally, *P. obovata*, section *Myapathanthus*, differs from the other *Deidamioides* species from a contraction in the IR region comprising the rRNA genes.

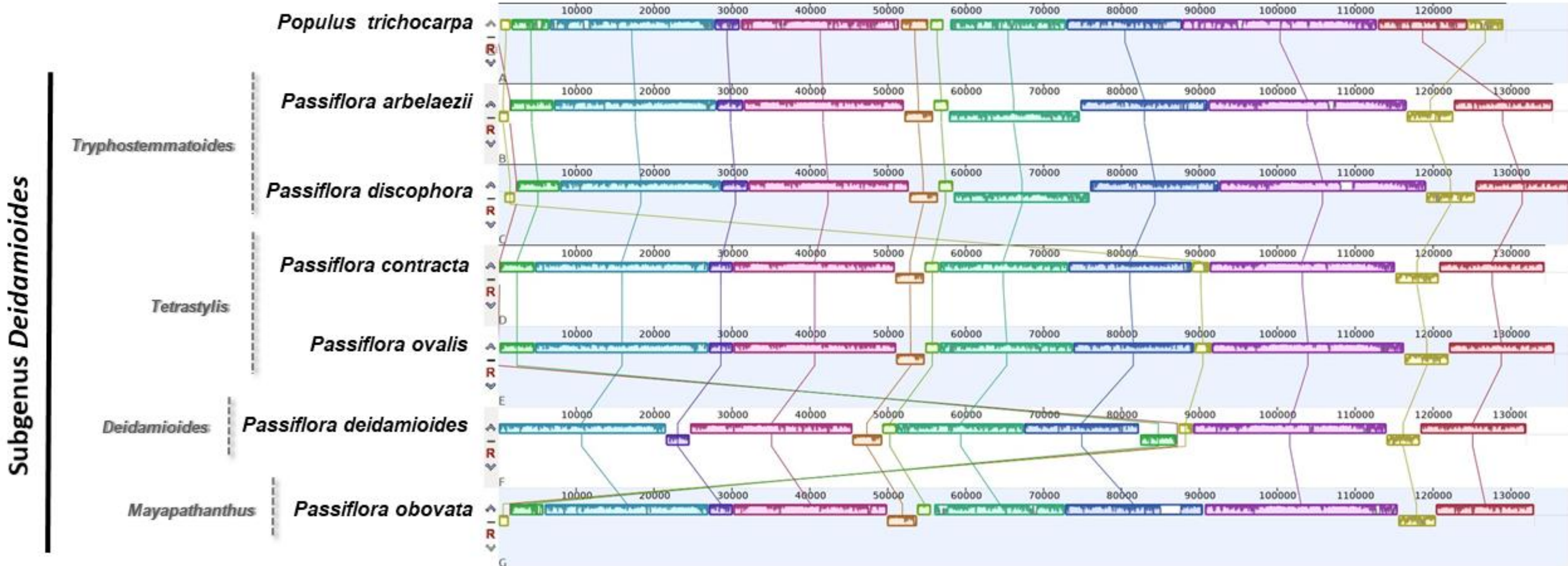


Figure 13. Synteny and structural rearrangements detected in the chloroplast genomes of 6 *Passiflora* species from the subgenus *Deidamioides* according with their sections. Colored bars indicate syntenic blocks and connecting lines indicate the correspondence between blocks.

4.3.3. The plastid phylogenomic study

The final alignment matrix of the plastid protein-coding genes was 59,350 characters long, and the phylogenomic analysis performed with both methods, ML and BI, resulted in trees with similar topology and high support (BS=100, PP=1) for most of the nodes (Figure 14). The species *Adenia manii* (Passifloraceae) clustered as sister of *Populus trichocarpa* (Salicaceae), the reference species used for rooting the tree. In addition, the grouping of the Passifloraceae *Dilkea retusa* and *Mitostemma brevifilis* was observed with high support (BS= 100, PP= 1).

Regarding the formerly described *Passiflora* subgenera, our results revealed the monophyly of *Astrophea*, *Passiflora* and *Decaloba.* In *Astrophea*, a clade consisting of *P. haematostigma* and *P. rhamnifolia* species (section *Pseudoastrophea*) was found, grouping as sister to *P. rusbyi* (PP= 1) from the section *Botryastrophea*. On the other hand, the species from section *Capreolata* (*P. cerradensis* and *P. pittieri*) grouped together (PP= 1). Interestingly, the clade consisting of *Astrophea* species has been recovered. as well as a group sister to the species of *Deidamioides* in section *Tryphostemmatoides* was defined with high support (BS= 100, PP= 1).

These results have shown the polyphyletic positioning of *Deidamioides* species, while *P. discophora* and *P. arbelaezii* (section *Tryphostemmatoides*) formed a clade (PP= 1) which is sister to *Astrophea*, *P. obovata*, another species previously assigned to *Deidamioides*, was embedded in subgenus *Decaloba*. A third clade which includes the type *species* (*P. deidamioides*) and two species in section *Tetrastylis* was recovered, as well as a group sister to *Decaloba* was defined with high support (BS=100, PP=1).

The *Decaloba* species formed a monophyletic group; however, *P. microstipula* grouped with *P. obovata* (*Deidamioides*). A clade with high support (PP= 1) was observed consisting of species of supersection *Auriculata* (*P. jatunsachensis*, *P. rufa*, *P. intricata* and *P. auriculata*). The species of supersection *Cieca* (*P. xiikzodz*, *P. suberosa* and *P tenuiloba*) grouped together and sister to supersection *Bryonoides*. In our study, section *Decaloba* appears as paraphyletic, with the group of *P. filipes* and *P. lutea* as sister to the species of section *Xeragona* (*P. capsularis* and *P costaricensis*). *P. tetrandra,* the type species from the subgenus *Tetrapathea*, grouped as sister to *Decaloba* subgenus.

Our phylogenomic analysis including 28 species from the subgenus *Passiflora* resulted in the monophyly of this clade. However, when comparing the distinct supersections, paraphyly was observed. In the tree (Figure 14), two major clades with high support (BS=

100, PP=1) was found in *Pasiflora* subgenus, one consisted of species from supersection *Stipulata*, and the other constituted of species from different supersections (*Coccinea*, *Distephana*, *Laurifolia*, and *Passiflora*). Although most species from *Stipulata* grouped together as a single cluster, this supersection is paraphyletic with *P. actinia* grouping closer to the species from supersection *Laurifolia* and *Passiflora*.

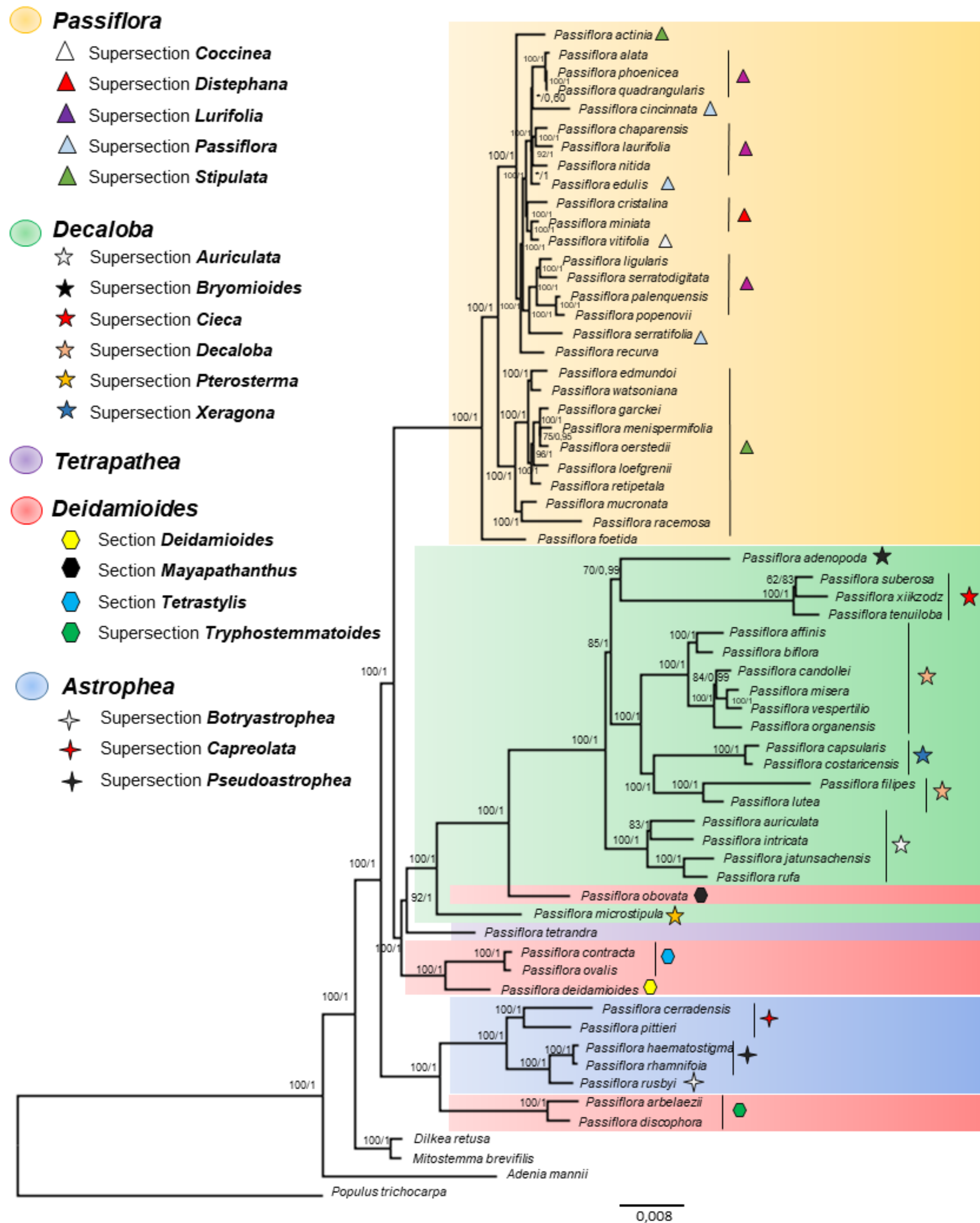


Figure 14. Phylogenetic reconstruction of *Passiflora* evolutionary history based on 68 chloroplast protein-coding genes. The support of the nodes was indicated by bootstrap values (BS) for Maximum likelihood and posterior probability for Bayesian inference. * indicates BS < 50.

Regarding the relationships within supersection *Stipulata*, the placement of species into some sections appeared paraphyletic, *P. loefgrenni* is from section *Kermesinae*, the same of *P. edmundoi* and *P. watsoniana*; however, in our topologies, it was embedded in section *Granadillastrum*, a clade formed by *P garckei*, *P. menispermifolia*, *P. oerstedi* and *P. retipetala*. The supersection *Passiflora* also appeared as polyphyletic: *P. cincinnata*, *P. edulis, P. recurva* and *P. serratifolia* that belong to the same series in this supersection grouped each one with different sub-clades with species from supersection *Laurifolia*. In addition, *P. vitifolia*, from supersection *Coccinea* grouped with *P. miniata* and *P. cristalina* from the supersection *Distephana* with high support (BS= 100, PP= 1).

To compare these results with the clustering into supersections of the *Passiflora* subgenus, an addition phylogenomic analysis was performed based on whole plastome sequences (APPENDIX L). The tree topology was congruent with the one based on the cp genes, with the supersections *Passiflora* and *Laurifolia* grouping as paraphyletic.

4.3.4. The nuclear phylogeny based on complete 18S/26S gene sequences

The sequencing of the *Passiflora* cpDNAs also allowed us to capture of the complete 35S rDNA cistron, and a phylogenetic study was performed using the complete 18S and 26S gene sequences. In our analysis, the phylogenetic tree using both BI and ML methods resulted in similar topologies (Figure 15 B, APPENDIX M A).

The results strongly support the monophyly of subgenus *Astrophea* (BS= 91, PP= 0,99). However, despite the high values that support the grouping of *Passiflora* (BS= 88, PP= 1) and *Decaloba* (BS= 100, PP= 1) species, the monophyly of these subgenera was not observed, since *P. adenopoda* (*Decaloba*) was embedded within the *Passiflora* subgenus. *Deidamioides* also appears paraphyletic, with *P. discophora* as sister to *Astrophea*, and *P. deidamioides* as sister to *P. tetrandra* (*Tetrapathea*). In addition, the clade composed of *P. contracta* and *P. ovalis* (*Deidamioides*) formed a polytomy in the tree.

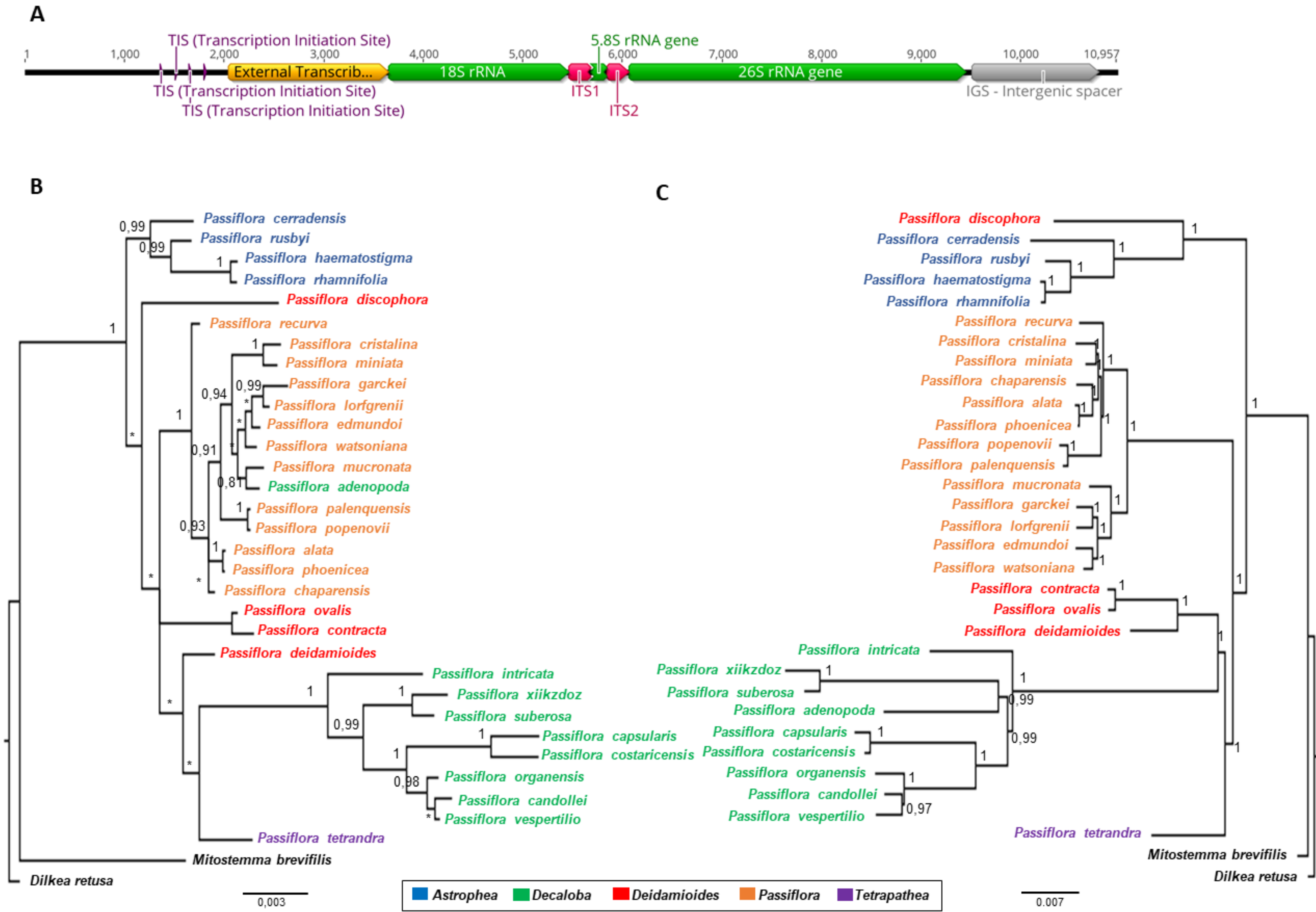


Figure 15. The 35S rDNA cistron structure of *Passiflora critalina* (A). The Bayesian inference phylogenetic tree of the *Passiflora* genus based on the nuclear 18S/26S genes (B) and based on 68 chloroplast protein-coding genes (C). * indicates posterior probability < 0,8.

Subsequently, an analysis using 68 chloroplast protein-coding genes was conducted with the same taxa used for reconstructing the phylogeny based on the 18S/26S sequences. In contrast to the tree obtained using nuclear data, the other tree resulted in high support values for most of the nodes in the topology. The same placement was observed for the *Astrophea* species; however, the plastid gene-based phylogeny strengthened the monophyly of the subgenus *Passiflora* (BS= 100, PP= 1) and *Decaloba* (BS= 100, PP= 1). The relationships within *Deidamioides* still indicate that it may be paraphyletic, with *P. discophora* as sister to *Astrophea* clade. In addition, the incongruence between the trees, generated by two different datasets, was mainly in the clustering into the *Passiflora* subgenus, and differently from the nuclear phylogeny, *P. adenopoda* was placed into *Decaloba* clade (Figure 15).

4.4. Discussion

Chloroplast genomes are known to be highly conservative in structure and gene order in angiosperms. However, some groups present highly rearranged plastomes, as occur in *Passiflora* that is characterized to have a syndrome of features related to plastome changes, as inversions, IR expansions, contractions, gene and intron losses (Cauz-Santos et al., 2017; Rabah et al., 2019; Shrestha et al., 2019).

The plastomes analyzed in the present study and assembled from long sequence reads present large variations between the subgenera, confirming previous findings shown in the 2nd chapter of this thesis. Interestingly, our new results have demonstrated that some species of the *Deidamioides* subgenus also show different plastome structures in comparison to others placed in different taxonomic sections. The most noteworthy case was *Tryphostemmatoides* because of species in this series show an inversion in the LSC region that previously was only described in *Astrophea* species (Cauz-Santos et al., 2020). This evolutionary event possibly occurred after the separation of the clade *Tryphostemmatoides* + *Astrophea* from the other *Passiflora* species. When comparing with *Populus trichocarpa*, a species that has a cpDNA structure very similar to the angiosperm ancestor plastome (Tuskan et al., 2006), the rearrangements shared between all *Deidamioides* plastomes possibly occurred before the diversification of *Passiflora*, since these inversions were also found in the cpDNAs of the Passifloraceae genera, *Dilkea* and *Mitostemma* (Cauz-Santos et al., 2020).

The cpDNA structure and respective rearrangements were described to have a potential use in plant taxonomic analyses (Downie and Palmer, 1992; Jansen and Palmer,

1987; Raubeson and Jansen, 2005). Particularly, the cpDNA gene order has proven to be a good marker for phylogenetic studies in the family Campanulaceae (Cosner et al., 2004), and in legumes resides the most classical example of cpDNA rearrangements used for taxonomic studies. The loss of an IR region was used in the classification of some papilionoids that were then described as an Inverted repeat-lacking clade (IRLC) (Lavin et al., 1990; McMahon and Sanderson, 2006; Wojciechowski et al., 2004).

In our results, the divergences within the different *Deidamioides* taxonomic sections do not only appear when comparing their plastome structures. In view of our findings based on a set of plastid protein-coding genes, *Deidamioides* also appeared as polyphyletic, with the species of *Tryphostemmatoides* appearing as sister to subgenus *Astrophea* with highly support (BS 100%, PP 100%), and *P. obovata* grouping with *Decaloba* species.

Plastid phylogenomics become a potential tool in plant taxonomic classification using both a set of genes (Jansen et al., 2007; Xi et al., 2012) or entire cp genome sequences (Zhang et al., 2011). Actually, the increase in DNA sequence lengths and the use of phylogenomics impact the accuracy of the phylogenies, resulting in an increase of tree resolution at low taxonomic levels (Parks et al., 2009; Wortley et al., 2005).

We also produced a phylogeny based on the complete 18S and 26S rDNA sequences that resulted in some well-supported clades, as the monophyly of *Astrophea* and *Passiflora.* However, in general, lowest values were obtained in comparison with the results from the plastid phylogenomics. The differences are possibly because the nuclear phylogeny was based on two highly conserved gene sequences. In contrast, the plastid phylogenomics were reconstructed based on a set of genes with different polymorphism levels.

In plant phylogenetic studies, the ITS region of 35S rDNA gene is a potential nuclear marker due to its high level of variation (Álvarez and Wendel, 2003; Feng et al., 2016; Jobst et al., 1998; Käss and Wink, 1997). Previously, the ITS region was used to reconstruct the first molecular *Passiflora* phylogenies (Krosnick et al., 2013; Muschner et al., 2003). The high levels of ITS variation could lead to the loss of phylogenetic signal by saturation, mainly in higher differentiated levels as it was reported by Muschner et al. (2003). In this scenario, we decided to use of the complete 18S and 26S rDNA genes which, in turn, resulted in strong support nodes at least for the subgenera *Astrophea* and *Passiflora*. The potential of these gene sequences to infer phylogenies was also suggested by Maia et al. (2014) when analyzing a large sample of angiosperm species; nonetheless, due to the low phylogenetic signal of this region, low support for some clades was obtained. In addition, considering that to generate an assembly we used a combination of sequences from multiple copies of the 35S ribosomal

cistron, the resulting tree need to be interpreted carefully, due to the difficulty in determining the accurate homology for a single copy from 35S rDNA (Fonseca and Lohmann, 2020). To solve this question, the use of nuclear low copy genes in plant phylogeny (Huang et al., 2016; Sang, 2002; Zhang et al., 2012) would be an alternative to propose a nuclear phylogenetic hypothesis for *Passiflora*.

The classical study of Killip (1938) based on morphological traits proposed the existence of 22 *Passiflora* subgenera which posteriorly were drastically reduced to four (Ulmer & MacDougal, 2004). Considering both morphological and ecological information, these authors suggested the existence of *Astrophea*, *Decaloba*, *Deidamioides* and *Passiflora*.

Extraordinary is the morphological diversity of the *Passiflora* species that also lead to a considerable distinction between the different subgenera. For instance, *Decaloba* (ca. 230 species described) are herbaceous vines presenting small fruits and flowers while *Passiflora* (ca. 250) are lianas or herbaceous vines with a corona with multiple rows of filaments, large flowers with great variation in colors and presenting long tubes in some species. In *Deidamioides* (14 species), the species are characterized by the occurrence of tiny stipules, petiolar glands, small bracts and plicate operculum (Ulmer and MacDougal, 2004). *Astrophea* (ca. 60 species), on the other hand, consists of woody lianas or even shrubs and small trees lacking tendrils or presenting short spines. The preference for specific pollinators was reported in *Passiflora* supersection *Tacsonia* in which there are species with flowers adapted for hummingbirds (Abrahamczyk et al., 2014). However, the *Passiflora* flowers are also visited by a variety of pollinators, such as small and large insects and bats, the latter adapted for species from the series *Tetrastylis* (*Deidamioides*) (Buzato and Franco, 1992; Fleming et al., 2009). Considering the *Passiflora* geographic distribution, despite its predominantly distribution in the neotropics, mainly in South America, 22 species from supersection *Disemma* (*Decaloba*) have been found in southeast Asia, as well as in Australia and New Zealand (Krosnick and Freudenstein, 2005).

The necessity of an improvement in the *Deidamioides* classification, cited as a diverse group, was firstly proposed by Feuillet and MacDougal (2003) who have also recognized the subdivisions of the four subgenera into supersections. The first molecular phylogeny based on nuclear and cpDNA markers revealed the existence of three well-supported clades, defined as *Astrophea*, *Decaloba* and *Passiflora*, although not resolving the position of *Deidamioides* (Muschner et al., 2003). The subsequent molecular phylogeny also supported *Deidamioides* as a polyphyletic group (Yockteng and Nadot, 2004), proposing the existence of eight subgenera: *Astrophea*, *Deidamioides*, *Dysosmia*, *Granadilla*,

Plectostemma, *Polyanthea*, *Tetrapathea*, and *Tryphostemmatoides*. In this phylogeny based on the nuclear cp-expressed glutamine synthetase gene, *ncpGS,* the species of *Tryphostemmatoides* appeared in a distinct clade in relation that of other *Deidamioides* belong, and were placed near *Astrophea* species*,* similarly to our results from the plastid phylogenomics.

Hansen et al. (2006) studying the phylogenetic relationships and chromosomal number evolution in *Passiflora* also support the clades *Astrophea*, *Decaloba* and *Passiflora,* and suggested the monophyly of *Deidamioides*. However, these results were limited to taxon sampling, and indicated the monophyly of a clade consisted of section *Tetrastylis* and *Deidamioides*, the same demonstrated in our findings. Muschner et al. (2012) compiling sequences of plastid, nuclear and mitochondrial origin observed a paraphyletic position of subgenus *Deidamioides*, with *P. tryphostemmatoides* clustering as a sister species to *Astrophea* clade, while the other *Deidamioides* species clustered as sister to *Decaloba.* This result was corroborated by Krosnick et al. (2013) findings, in which species from the section *Tryphostemmatoides* also grouped as a sister to the ancestral *Astrophea.*

Besides to the described *Astrophea, Decaloba*, *Deidamioides* and *Passiflora,* Krosnick et al. (2009) proposed an additional subgenus, *Tetrapathea*, which is represented in our study by *P. tetrandra*. Taxonomically speaking, *Tetrapathea* is a genus of the family Passifloraceae; however, due to its phylogenetic position and the distinct morphological characters, as dioecy and variable numbers of stamens and carpels, it was suggested that it would be a new *Passiflora* subgenus. The species are endemic from Australia, and in our analysis *Tetrapathea* was placed as a sister to *Decaloba* corroborating previous phylogenetic findings (Krosnick et al., 2013).

Recently, Sader et al. (2019) using a probabilistic approach implemented in the ChromEvol software revealed that chromosome changes could had an impact on the diversification of *Passiflora* species. While the established basic chromosome number for the genus *Passiflora* (x=6) is the same reported for *Decaloba* species, events of ascending dysploidy could have led to changes from x=6 to x=9 in the subgenus *Passiflora*. Despite the important role of polyploidy in plant speciation (Pelé et al., 2018; Wood et al., 2009), apparently it is not correlated with the diversification in *Passiflora*, since polyploidy has been detected only in a few lineages.

Passiflora genus emerged at 42.9 Mya in the Eocene period (Sader et al., 2019), an age corresponding to the same period for the oldest seed fossil of Passifloraceae found in Colombia, that revealed its presence in the Neotropics at late Eocene (Martínez, 2017). Sader

et al. (2019) also described *Decaloba* as the oldest clade in the genus *Passiflora* (37.04 Mya). In contrast, *Astrophea* shows a more recent diversification (20.59 Mya); however, a very ancient separation of *Astrophea* clade from other *Passiflora* clades was detected (around 40 Mya), corroborating with Muschner et al. (2012) studies that have placed *Tryphostemmatoides* (subgenus *Deidamioides,* series *Tryphostemmatoides*) as a sister to *Astrophea* diverging from this subgenus at 25.4 Mya.

4.5. Conclusions

In this study, the obtainment of new chloroplast genomes allowed us to perform a robust phylogenomics analysis in the genus *Passiflora*. A comparative genomics in the *Deidamioides* subgenus unveiled that each taxonomic section presents a particular cp-genome rearrangement. In this way, these rearrangements could be used as an additional information in classification studies within this subgenus. Furthermore, our phylogenomic analysis corroborates with earlier studies regarding the monophyly of the subgenera *Astrophea*, *Decaloba* and *Passiflora*. Based on the plastome comparative analyses and taking into account the polyphyletic positioning of *Deidamioides* species, we suggest to elevate the status of *Tryphostemmatoides* to subgenus. In this way, we propose an infrageneric classification of *Passiflora* into six subgenera *Astrophea, Decaloba, Deidamioides, Passiflora, Tetraphatea and Tryphostemmatoides.*

REFERENCES

Abeel, T., Van Parys, T., Saeys, Y., Galagan, J., Van De Peer, Y., 2012. GenomeView: A next-generation genome browser. Nucleic Acids Res. 40, 1–10. https://doi.org/10.1093/nar/gkr995

Abrahamczyk, S., Souto-Vilarós, D., Renner, S.S., 2014. Escape from extreme specialization: Passionflowers, bats and the sword-billed hummingbird. Proc. R. Soc. B Biol. Sci. https://doi.org/10.1098/rspb.2014.0888

Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. https://doi.org/10.1016/S1055-7903(03)00208-2

Buzato, S., Franco, A.L.M., 1992. *Tetrastylis ovalis*: A second case of bat-pollinated passionflower (Passifloraceae). Plant Syst. Evol. https://doi.org/10.1007/BF00937450

Cauz-Santos, L.A., Munhoz, C.F., Rodde, N., Cauet, S., Santos, A.A., Penha, H.A., Dornelas, M.C., Varani, A.M., Oliveira, G.C.X., Bergès, H., Vieira, M.L.C., 2017. The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: structural organization and phylogenomic studies in Malpighiales. Front. Plant Sci. 8, 1–17. https://doi.org/10.3389/fpls.2017.00334

Cauz-Santos, L.A., Costa, Z.P., Callot, C., Cauet, S., Zucchi, M.I., Bergès, H., van den Berg, C., Vieira, M.L.C., 2020. Evolutionary history of chloroplast genomes: a case study in point reveals a repertory of rearrangements and the loss of na inverted repeat region in *Passiflora*. Genome Biol. Evol. (submitted)

Cosner, M.E., Raubeson, L.A., Jansen, R.K., 2004. Chloroplast DNA rearrangements in Campanulaceae: Phylogenetic utility of highly rearranged genomes. BMC Evol. Biol. https://doi.org/10.1186/1471-2148-4-27

Darling, A.E., Mau, B., Perna, N.T., 2010. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5. https://doi.org/10.1371/journal.pone.0011147

Dierckxsens, N., Mardulyn, P., Smits, G., 2017. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. https://doi.org/10.1093/nar/gkw955

Downie, S.R., Palmer, J.D., 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Molecular Systematics of Plants. https://doi.org/10.1007/978-1-4615-3276-7_2

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

Feng, S., Jiang, M., Shi, Y., Jiao, K., Shen, C., Lu, J., Ying, Q., Wang, H., 2016. Application of the ribosomal DNA ITS2 region of *Physalis* (Solanaceae): DNA barcoding and phylogenetic study. Front. Plant Sci. https://doi.org/10.3389/fpls.2016.01047

Feuillet, C., MacDougal, J.M., 2003. A new infrageneric classification of *Passiflora* L. (Passifloraceae). J. Newsl. Passiflora Soc. Int.

Fleming, T.H., Geiselman, C., Kress, W.J., 2009. The evolution of bat pollination: A phylogenetic perspective. Ann. Bot. https://doi.org/10.1093/aob/mcp197

Fonseca, L.H.M., Lohmann, L.G., 2020. Exploring the potential of nuclear and mitochondrial sequencing data generated through genome-skimming for plant phylogenetics: A case study from a clade of neotropical lianas. J. Syst. Evol. https://doi.org/10.1111/jse.12533

Greiner, S., Lehwark, P., Bock, R., 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. https://doi.org/10.1093/nar/gkz238

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. https://doi.org/10.1600/036364406775971769

Huang, C.H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M.A., Al-Shehbaz, I., Edger, P.P., Pires, J.C., Tan, D.Y., Zhong, Y., Ma, H., 2016. Resolution of brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msv226

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., 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. https://doi.org/10.1073/pnas.0709121104

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. https://doi.org/10.1073/pnas.84.16.5818

Jobst, J., King, K., Hemleben, V., 1998. Molecular evolution of the internal transcribed spacers (ITS1 and ITS2) and phylogenetic relationships among species of the family Cucurbitaceae. Mol. Phylogenet. Evol. https://doi.org/10.1006/mpev.1997.0465

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods. https://doi.org/10.1038/nmeth.4285

Käss, E., Wink, M., 1997. Phylogenetic relationships in the Papilionoideae (Family Leguminosae) based on nucleotide sequences of cpDNA (rbcL) and ncDNA (ITS 1 and 2). Mol. Phylogenet. Evol. https://doi.org/10.1006/mpev.1997.0410

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010

Killip, E.P., 1938. The American species of Passifloraceae. Chicago Field Museum., Chicago.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. Genome Res. https://doi.org/10.1101/gr.215087.116

Krosnick, S.E., Ford, A.J., Freudenstein, J. V., 2009. Taxonomic revision of *Passiflora* subgenus *Tetrapathea* including the monotypic genera *Hollrungia* and *Tetrapathea* (Passifloraceae), and a new species of *Passiflora*. Syst. Bot. https://doi.org/10.1600/036364409788606343

Krosnick, S.E., Freudenstein, J. V., 2005. Monophyly and floral character homology of Old world *Passiflora* (subgenus *Decaloba* : supersection *Disemma*) . Syst. Bot.

https://doi.org/10.1600/0363644053661959

Krosnick, S.E., Porter-Utley, K.E., MacDougal, J.M., Jørgensen, P.M., McDade, L.A., 2013. New insights into the evolution of *Passiflora* subgenus *Decaloba* (Passifloraceae): Phylogenetic relationships and morphological synapomorphies. Syst. Bot. https://doi.org/10.1600/036364413x670359

Lavin, M., Doyle, J.J., Palmer, J.D., 1990. Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. Evolution. https://doi.org/10.2307/2409416

Maia, V.H., Gitzendanner, M.A., Soltis, P.S., Wong, G.K.S., Soltis, D.E., 2014. Angiosperm phylogeny based on 18S/26S rDNA sequence data: Constructing a large data set using next-generation sequence data. Int. J. Plant Sci. https://doi.org/10.1086/676675

Marinho, L.C., Cai, L., Duan, X., Ruhfel, B.R., Fiaschi, P., Amorim, A.M., van den Berg, C., Davis, C.C., 2019. Plastomes resolve generic limits within tribe Clusieae (Clusiaceae) and reveal the new genus Arawakia. Mol. Phylogenet. Evol. https://doi.org/10.1016/j.ympev.2019.02.005

Martínez, C., 2017. Passifloraceae seeds from the late eocene of Colombia. Am. J. Bot. https://doi.org/10.3732/ajb.1700224

McMahon, M.M., Sanderson, M.J., 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. Syst. Biol. https://doi.org/10.1080/10635150600999150

Moore, M.J., Bell, C.D., Soltis, P.S., Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.0708072104

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. https://doi.org/10.1073/pnas.0907801107

Muschner, V.C., Lorenz, A.P., Cervi, A.C., Bonatto, S.L., Souza-Chies, T.T., Salzano, F.M., Freitas, L.B., 2003. A first molecular phylogenetic analysis of *Passiflora* (Passifloraceae). Am. J. Bot. 90, 1229–1238. https://doi.org/10.3732/ajb.90.8.1229

Muschner, V.C., Zamberlan, P.M., Bonatto, S.L., Freitas, L.B., 2012. Phylogeny, biogeography and divergence times in Passiflora (Passifloraceae). Genet. Mol. Biol. 35, 1036–1043. https://doi.org/10.1590/S1415-47572012000600019

Ocampo Pérez, J., Coppens d'Eeckenbrugge, G., 2017. Morphological characterization in the genus *Passiflora* L.: an approach to understanding its complex variability. Plant Syst. Evol. 303, 531–558. https://doi.org/10.1007/s00606-017-1390-2

Parks, M., Cronn, R., Liston, A., 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. https://doi.org/10.1186/1741-7007-7-84

Pelé, A., Rousseau-Gueutin, M., Chèvre, A.M., 2018. Speciation success of polyploid plants closely relates to the regulation of meiotic recombination. Front. Plant Sci. https://doi.org/10.3389/fpls.2018.00907

Rabah, S.O., Shrestha, B., Hajrah, N.H., Sabir, Mumdooh J., Alharby, H.F., Sabir, Mernan J., Alhebshi, A.M., Sabir, J.S.M., Gilbert, L.E., Ruhlman, T.A., Jansen, R.K., 2019. *Passiflora* plastome sequencing reveals widespread genomic rearrangements. J. Syst.

Evol. https://doi.org/10.1111/jse.12425

Rambaut, A., 2016. FigTree v1.4.3. Mol. Evol. phylogenetics Epidemiol.

Raubeson, L.A., Jansen, R.K., 2005. Chloroplast genomes of plants., in: Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants. https://doi.org/10.1079/9780851999043.0045

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. https://doi.org/10.1093/sysbio/sys029

Sader, M.A., Amorim, B.S., Costa, L., Souza, G., Pedrosa-Harand, A., 2019. The role of chromosome changes in the diversification of *Passiflora* L. (Passifloraceae). Syst. Biodivers. https://doi.org/10.1080/14772000.2018.1546777

Sang, T., 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit. Rev. Biochem. Mol. Biol. https://doi.org/10.1080/10409230290771474

Shrestha, B., Weng, M.L., Theriot, E.C., Gilbert, L.E., Ruhlman, T.A., Krosnick, S.E., Jansen, R.K., 2019. Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. Mol. Phylogenet. Evol. https://doi.org/10.1016/j.ympev.2019.05.030

Song, Y., Yu, W. Bin, Tan, Y.H., Jin, J.J., Wang, B., Yang, J.B., Liu, B., Corlett, R.T., 2019. Plastid phylogenomics improve phylogenetic resolution in the Lauraceae. J. Syst. Evol. https://doi.org/10.1111/jse.12536

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., Greiner, S., 2017. GeSeq - Versatile and accurate annotation of organelle genomes. Nucleic Acids Res. https://doi.org/10.1093/nar/gkx391

Tokuoka, T., 2012. Molecular phylogenetic analysis of Passifloraceae sensu lato (Malpighiales) based on plastid and nuclear DNA sequences. J. Plant Res. https://doi.org/10.1007/s10265-011-0472-4

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.-L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., DePamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J.-C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.-J., Uberbacher, E., Unneberg, P.,

Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y., Rokhsar, D., 2006. The Genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science (80-.). 313, 1596–1604. https://doi.org/10.1126/science.1128691

Ulmer, T., MacDougal, J.M., 2004. Passiflora: passionflowers of the world. Timber Press, Portland.

Wojciechowski, M.F., Lavin, M., Sanderson, M.J., 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family. Am. J. Bot. https://doi.org/10.3732/ajb.91.11.1846

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., Rieseberg, L.H., 2009. The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.0811575106

Wortley, A.H., Rudall, P.J., Harris, D.J., Scotland, R.W., 2005. How much data are needed to resolve a difficult phylogeny?: case study in Lamiales. Syst. Biol. https://doi.org/10.1080/10635150500221028

Xi, Z., Ruhfel, B.R., Schaefer, H., Amorim, A.M., Sugumaran, M., Wurdack, K.J., Endress, P.K., Matthews, M.L., Stevens, P.F., Mathews, S., Davis, C.C., 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proc. Natl. Acad. Sci. 109, 17519–17524. https://doi.org/10.1073/pnas.1205818109

Yockteng, R., Nadot, S., 2004. Phylogenetic relationships among *Passiflora* species based on the glutamine synthetase nuclear gene expressed in chloroplast (ncpGS). Mol. Phylogenet. Evol. https://doi.org/10.1016/S1055-7903(03)00277-X

Zhang, N., Zeng, L., Shan, H., Ma, H., 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. New Phytol. https://doi.org/10.1111/j.1469-8137.2012.04212.x

Zhang, S.D., Jin, J.J., Chen, S.Y., Chase, M.W., Soltis, D.E., Li, H.T., Yang, J.B., Li, D.Z., Yi, T.S., 2017. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. New Phytol. https://doi.org/10.1111/nph.14461

Zhang, Y.J., Ma, P.F., Li, D.Z., 2011. High-throughput sequencing of six bamboo chloroplast genomes: Phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). PLoS One. https://doi.org/10.1371/journal.pone.0020596

5. CONCLUSIONS AND FUTURE PERSPECTIVES

In this thesis, the evolutionary history of *Passsiflora* chloroplast (cp) genomes (plastomes) was investigated. We started by sequencing of the first cp genome in Passifloraceae, precisely of the sour passion fruit, *P. edulis*, that has different structural rearrangements in comparison with other species from Malphighiales order, or even the conserved plastomes found in angiosperms. These changes were inversions, losses of genes and introns and expansions. One could then hypothesize that these events are possibly exclusive to the *Passiflora* genus. As expected, this hypothesis was confirmed when we enlarged our sampling and investigated in depth the plastomes in *Astrophea*, *Decaloba*, *Passiflora* and *Deidamioides*.

The plastome evolutionary history in *Passiflora* is uncommon. Interestingly, this group shows a great diversity in morphology (flowers and vegetative structures) and distinct chromosome numbers (*n*= 6, 7, 9, 11, 12) (discussed in Sader et al., 2019). There is still huge variation in genome sizes (from 0.413 pg on average in *Decaloba* up to 1.311 pg on average in *Passiflora*), and all these attributes contribute for *Passiflora* to be considered as a model in plastome evolutionary studies.

The high number of rearrangements has possibly occurred after the diversification of the genus, since the plastomes of other Passifloraceae species presented less structural changes (e.g. *Dilkea retusa* and *Mitostemma brevifilis*). Sequence rearrangements were mostly inversions of gene order and were distinct in the different subgenera; however, some inversions were congruent with the evolution of *Passiflora* (as shown in Figure 12). In addition to the inversions, gene and intron losses also occurred independently and repeatedly during the evolution process, the same for expansions of IR regions. Most intriguingly is the loss of an IR region, a rare event in angiosperms, which was detected in *P. costaricensis* a species from supersection *Xeragona* in *Decaloba*.

The subgenus *Decaloba* presents a high number of rearrangements, and interestingly is the group which shows cytonuclear incompatibility and chloroplast biparental inheritance, suggesting there is link between these processes. In fact, all these changes possibly occurred in *Passiflora* in a way for rescuing the cytonuclear incompatibility, a process that could lead to speciation.

Regarding the phylogeny, we may conclude that *Deidamioides* is polyphyletic, and suggest that the botanical classification of *Passiflora* needs to consider elevating

Tryphostemmatoides (*Deidamioides* section) to subgenus status. In addition, a revision on the supersections in the *Passiflora* subgenus is also required.

Finally, both the plastid phylogenomics and the 18S/26S rDNA reconstructed phylogeny were congruent in the topologies among subgenera. However, the obtention of new genome sequences in *Passiflora* could open the way for mitochondrial and nuclear-based phylogenies, for example, using low-copy nuclear genes with different levels of variation. Regarding the plastome evolution, the description of other complete cp genomes in the family Passifloraceae would be important in order to reconstruct the ancestral plastome.

APPENDICES

APPENDIX A.

(Corresponds to the Supplementary Table S1 published in Cauz-Santos et al., 2017)

List of complete cp-genomes and cp-gene sequences used in comparative and phylogenomic analysis.

Species	**Genus**	**Family**	**Order**	**Border comparison (Figure 2)**	**Sinteny and rearrangement analysis (Figure 3)**	**Phylogenomics (Complete genomes) (Figure 5)**	**Phylogenomics (Gene sequences) (Figure 6)**	**Genbank acession number**
Pentactina rupicola	*Pentactina*	Rosaceae	Rosales				*	NC_016921.1
Prunus persica	*Prunus*	Rosaceae	Rosales				*	NC_014697.1
Prinsepia utilis	*Prinsepia*	Rosaceae	Rosales				*	NC_021455.1
Pyrus pyrifolia	*Pyrus*	Rosaceae	Rosales				*	NC_015996.1
Fragaria vesca	*Fragaria*	Rosaceae	Rosales				*	NC_015206.1
Rosa odorata	*Rosa*	Rosaceae	Rosales				*	KF753637.1
Humulus lupulus	*Humulus*	Cannabaceae	Rosales				*	NC_028032.1
Morus indica	*Morus*	Moraceae	Rosales				*	NC_008359.1
Castanea mollissima	*Castanea*	Fagaceae	Fagales				*	NC_014674.1
Quercus rubra	*Quercus*	Fagaceae	Fagales				*	NC_020152.1
Trigonobalanus doichangensis	*Trigonobalanus*	Fagaceae	Fagales				*	NC_023959.1
Corynocarpus laevigata	*Corynocarpus*	Curcubitaceae	Curcubitales				*	NC_014807.1
Cucumis melo	*Cucumis*	Curcubitaceae	Curcubitales				*	NC_015983.1
Pisum sativum	*Pisum*	Fabaceae	Fabales				*	NC_014057.1
Trifolium subterraneum	*Trifolium*	Fabaceae	Fabales				*	NC_011828.1
Medicago truncatula	*Medicago*	Fabaceae	Fabales				*	NC_003119.6
Cicer arietinum	*Cicer*	Fabaceae	Fabales				*	NC_011163.1
Lotus japonicus	*Lotus*	Fabaceae	Fabales				*	NC_002694.1
Phaseolus vulgaris	*Phaseolus*	Fabaceae	Fabales				*	NC_009259.1
Vigna unguiculata	*Vigna*	Fabaceae	Fabales				*	NC_018051.1

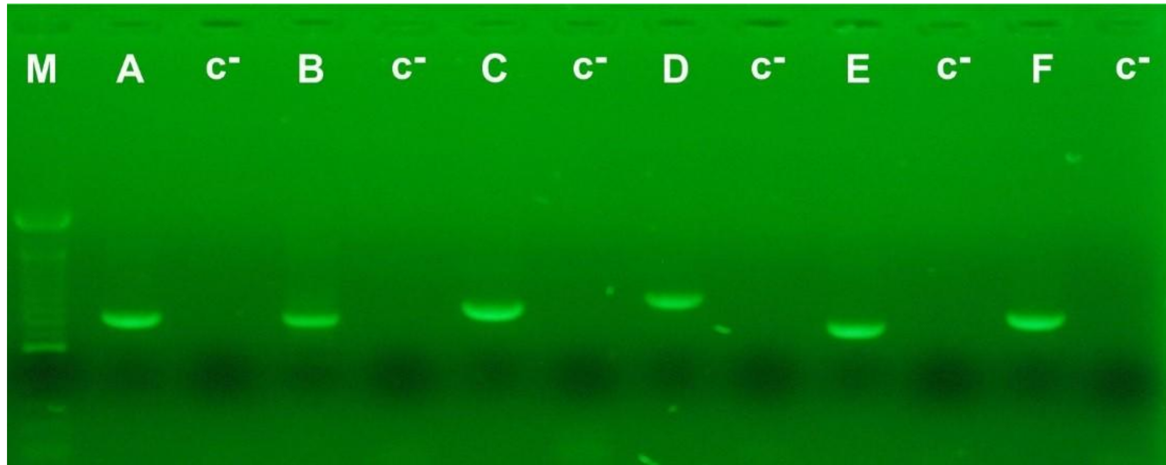
Glycine max	*Glycine*	Fabaceae	Fabales				*	NC_007942.1
Populus alba	*Populus*	Salicaceae	Malpighiales		¥	*	*	NC_008235.1
Populus trichocarpa	*Populus*	Salicaceae	Malpighiales	*	¥*	*	*	NC_009143.1
Populus balsamifera	*Populus*	Salicaceae	Malpighiales		¥	*		NC_024735.1
Populus fremonti	*Populus*	Salicaceae	Malpighiales		¥	*		NC_024734.1
Populus tremula	*Populus*	Salicaceae	Malpighiales		¥	*		NC_027425.1
Populus euphratica	*Populus*	Salicaceae	Malpighiales		¥	*		NC_024747.1
Salix purpurea	*Salix*	Salicaceae	Malpighiales	*	¥*	*	*	NC_026722.1
Salix suchowensis	*Salix*	Salicaceae	Malpighiales		¥	*		NC_026462.1
Salix interior	*Salix*	Salicaceae	Malpighiales		¥	*		NC_024681.1
Passiflora ciliata	*Passiflora*	Passifloraceae	Malpighiales			*		JX661828 to JX665018
Passiflora edulis	***Passiflora***	**Passifloraceae**	**Malpighiales**	*****	**¥***	*****	*****	KX290855
Passiflora biflora	*Passiflora*	Passifloraceae	Malpighiales			*	*	EU017064.1 to EU017137.1
Manihot esculenta	*Manihot*	Euphorbiaceae	Malpighiales	*	¥*	*	*	NC_010433.1
Hevea brasiliensis	*Hevea*	Euphorbiaceae	Malpighiales	*	¥*	*	*	NC_015308.1
Ricinus communis	*Ricinus*	Euphorbiaceae	Malpighiales	*	¥*	*	*	NC_016736.1
Jatropha curcas	*Jatropha*	Euphorbiaceae	Malpighiales	*	¥*	*	*	NC_012224.1
Hirtella physophora	*Hirtella*	Chrysobalanaceae	Malpighiales		¥*	*	*	NC_024066.1
Hirtella racemosa	*Hirtella*	Chrysobalanaceae	Malpighiales		¥	*	*	NC_024060.1
Licania heteromorpha	*Licania*	Chrysobalanaceae	Malpighiales		¥*	*	*	NC_024062.1
Licania alba	*Licania*	Chrysobalanaceae	Malpighiales		¥	*	*	NC_024064.1
Licania sprucei	*Licania*	Chrysobalanaceae	Malpighiales		¥	*		NC_024065.1
Couepia guianensis	*Couepia*	Chrysobalanaceae	Malpighiales		¥*	*		NC_024063.1
Parinari campestris	*Parinari*	Chrysobalanaceae	Malpighiales		¥*	*	*	NC_024067.1
Chrysobalanus icaco	*Chrysobalanus*	Chrysobalanaceae	Malpighiales		¥*	*	*	NC_024061.1
Elaeodendron orientale	*Elaeodendron*	Celastraceae	Celastrales				*	JX661810.1 to JX664998.1
Euonymus americanus	*Euonymus*	Celastraceae	Celastrales				*	EU002160.1 to GQ998211.1
Averrhoa carambola	*Averrhoa*	Oxalidaceae	Oxalidales				*	JX661801.1 to JX664986.1

Oxalis latifolia	*Oxalis*	Oxalidaceae	Oxalidales	*	EU002165.1 to GQ998572.1
Bulnesia arborea	*Bulnesia*	Zygophyllaceae	Zygophyllales	*	EU002159.1 to GQ998067.1
Vitis vinifera	*Vitis*	Vitaceae	Vitales	*	NC_007957.1

¥ Cp genomes available from the order Malpighiales used in the phylogenomic analyses.

APPENDIX B.
(Corresponds to the Supplementary Figure S1 published in Cauz-Santos et al., 2017)

LSC inversions found in *Passiflora edulis* cp-genome validated by PCR and submitted to Sanger sequencing**.**



PCR products run on agarose gel (1%) using primers to anneal inversion boundaries. **A** – inversion 1, 5'end; **B** – inversion 1, 3'end; **C** – inversion 2, 5' end; **D** – inversion 2, 3'end; **E** – inversion 3, 5'end; **F** – inversion 3, 3'end. **c**$^-$ – PCR assay without DNA template. M, 100-bp ladder DNA molecular size marker (Invitrogen®).

APPENDIX C. List of *Passiflora* species, locality of collection, and methodological analyses in which they were included

Subgenus	Species	Locality of collection	Mauve first alignment	Mauve second alignment	IR expansion analysis
Astrophea	*Passiflora cerradensis*	Juara (Mato Grosso)	X	X	X
	Passiflora haematostigma	Sao Bento (rio Grande do Sul)	X	X	X
	Passiflora rhamnifolia	Chapada Diamantina (Bahia)	X	X	X
Decaloba	*Passiflora candollei*	Alta Floresta (Mato Grosso)	X	X	X
	Passiflora capsularis	Ribeirao Grande (São Paulo)	X	X	X
	Passiflora costaricensis	Costa Rica, Central America **	X	X	X
	Passiflora suberosa	Piracicaba (São Paulo)	X	X	X
	Passiflora vespertilio	Alta Floresta (Mato Grosso)	X	X	X
Deidamioides	*Passiflora contracta*	Itabera (Bahia)	X	X	X
	Passiflora deidamioides	Paranapiacaba (São Paulo)	X	X	X
Passiflora	*Passiflora alata*	Piracicaba (São Paulo)	X	X	X
	Passiflora cristalina	Alta Floresta (Mato Grosso)		X	
	Passiflora edmundoi	Palmeiras (Bahia)	X	X	X
	Passiflora loefgrenii	Ribeirao Grande (São Paulo)		X	
	Passiflora miniata	Alta Floresta (Mato Grosso)		X	
	Passiflora mucronata	Itacare (Bahia)		X	
	Passiflora recurva	Palmeiras (Bahia)	X	X	X
	Passiflora watsoniana	Palmeiras (Bahia)		X	
Dilkea	*Dilkea retusa*	Santarém (Pará)	X	X	X
Mitostemma	*Mitostemma brevifilis*	Ribas do Rio Pardo (Mato Grosso do Sul)	X	X	X

* The names of the Brazilian states are given within brackets

** Species obtained from the Italian National Collection of *Passiflora*

APPENDIX D. Assembly report of the *Passiflora* chloroplast genomes as well as of *Dilkea retusa* and *Mitostemma brevifilis*.

			NOVOPlasty					Geneious			
Species	**Illumina Sequencing**	**Total reads**	**Seed sequence**	**K-mer size**	**Number of contigs**	**N50 (bp)**	**Largest contig (bp)**	**Final contig (bp)**	**Mapped reads**	**Coverage**	**Standard deviation**
Passiflora cerradensis	2x150	22,754,226	*psba*	27	8	95,494	95,494	164,515	312,931	273,6	42,4
Passiflora haematostigma	2x150	17,430,270	*psba*	33	4	66,798	78,410	163,775	495,185	406,5	83,7
Passiflora rhamnifolia	2x76	41,545,758	*psba*	33	1	162,217	162,217	162,217	3,159,608	1459,1	416,1
Passiflora candollei	2x150	17,418,848	*psba*	25	3	78,011	78,011	138,081	610,704	596,8	99,3
Passiflora capsularis	2x150	20,764,460	*psba*	33	3	98,968	98,968	113,114	1,385,507	1678,6	265,7
Passiflora suberosa	2x76	39,288,936	*psba*	23	7	50,453	64,935	158,313	3,500,674	1668,4	354,6
Passiflora vespertilio	2x150	24,924,420	*psba*	33	4	83,679	83,679	138,456	1,469,520	1455,2	214,5
Passiflora contracta	2x76	40,087,112	*rpoB*	35	12	52,316	58,232	166,558	1,323,274	589,5	114,6
Passiflora deidamioides	2x76	39,856,504	*rbcL*	25	6	64,351	70,361	167,953	2,337,851	1034,3	244,6
Passiflora alata	2x76	34,408,974	*psba*	25	10	62,093	64,622	147,773	591,986	297,3	69,4
Passiflora cristalina	2x150	22,320,626	*rbcL*	33	3	112,944	112,944	145,054	203,402	190,5	30,7
Passiflora edmundoi	2x150	22,450,584	*psba*	25	4	63,137	66,405	142,737	846,246	801,3	134
Passiflora loefgrenii	2x76	38,113,682	*psba*	33	2	81,924	81,924	146,537	5,879,056	2977,3	586,3
Passiflora miniata	2x150	13,862,202	*psba*	39	5	46,879	67,475	151,920	197,290	174,7	37,4
Passiflora mucronata	2x150	20,276,596	*psba*	23	6	38,580	66,292	150,839	802,974	719,6	112,4
Passiflora recurva	2x150	17,888,470	*rbcL*	23	5	40,041	67,107	151,837	670,343	598,4	108
Passiflora watsoniana	2x150	16,526,196	*rbcL*	39	5	37,941	71,878	146,520	227,824	209,7	36,7
Dilkea retusa	2x150	21,899,744	*rbcL*	27	6	54,747	60,013	161,923	117,425	98,6	17,5
Mitostemma brevifilis	2x150	23,891,314	*psba*	39	6	135,027	135,027	163,032	1,652,427	1379,1	207,3

				Canu			Geneious			
Species	**PacBio Sequencing data (Gb)**	**Total reads**	**Mean read lenght (bp)**	**Number of contigs**	**N50 (bp)**	**Largest contig (bp)**	**cp contigs identified**	**Final contig (bp)**	**Coverage**	**Standard deviation**
Passiflora costaricensis	1,1	247,741	4,441	4,322*	7,061	61,771	4	114,230	538	115

*The high number of contigs and small N50 was a result of nuclear contamination in the chloroplast sample.

APPENDIX E. PCR primers used to confirm the chloroplast DNA sequence in five *Passiflora* species

Species	Region within the chloroplast genome	Primer code	Primer sequence 5'-3'
Passiflora capsularis	*ndhB - psbA*	Pcap1_F	ACCCCTCACGTGCGAAATTA
	ndhB - psbA	Pcap1_R	ACCGTCCGAAAAGCTTCCTT
	rps15 - rrna5	Pcap21_F	AGATAGTAGTCGTCGGCGTT
	rps1 5 - rrna5	Pcap21_R	GGGGAAGTCCTGCGGAAAAA
	rps15 - rrna5	Pcap22_F	TCCATTTAACCAAGGCCGGT
	rps15 - rrna5	Pcap22_R	TCCTAGGCGTAGAGGAACCA
	rps15 - rrna5	Pcap23_F	TCTGACTCCCGAATATCCAATTTAC
	rps15 - rrna5	Pcap23_R	GCTCAGGTTTATATTGGCAGCA
	ndhF - rpl23	Pcap31_F	GACACCAAAGAAGAGTTCGGC
	ndhF - rpl23	Pcap31_R	GAATTACCAATGGGGTCGGC
	ndhF - rpl23	Pcap32_F	CGGCCCAATGCTTTATTTCTGT
	ndhF - rpl23	Pcap32_R	CCCACTTCGGTAACTCGTTCT
	ndhF - rpl23	Pcap33_F	GCACTAGCGTTGGTAATGTGG
	ndhF - rpl23	Pcap33_R	GCCTTTTCTTTCTATAAGCCGGT
	rps11 - rpoA	Pcap4_F	TGATTCCTTGCCCTGGCTTC
	rps11 rpoA	Pcap4_R	GCCGATTTAAGGGACCACGA
Passiflora costaricensis	*rps11 - ndhF*	Pcost1_F	GTTTCGCAAGAAAGGCGGTT
	rps11 - ndhF	Pcost1_R	CGGCAGCCGTATCTCTTCTT
	rpoA - rpl23	Pcost2_F	CTTGCCCTGGCTTCAGGTTA
	rpoA - rpl23	Pcost2_R	ACCCGGTTGAAGCGTAATGA
	rps15 - rrn5	Pcost3_F	GGAACAAGCGCTCGTATCTTT
	rps15 - rrn5	Pcost3_R	ACTGTAGGGGAAGTCCTGCG
	ndhB - psbA	Pcost4_F	TAACCCCTCACGTGCGAAAT
	ndhB - psbA	Pcost4_R	ACCGTCCGAAAAGCTTCCTT
Passiflora cerradensis	*matK - rpl23*	Pcerra_F	CGGGAGTCGATGGCTATGAA
	matK - rpl23	Pcerra_R	ACTCCCGAAACATAAGTGGGT
Passiflora deidamioides	*psbK - rpl23*	Pdeid_F	ACCCGGTTGAAGCGTAATGA
	psbK - rpl23	Pdeid_R	ATGCCAGTAATCCCCCTGTTC
Passiflora vespertilio	*ndhF - rpl2*	Pvesp1_F	GCCAGATGAAGGAACGGGAA
	ndhF - rpl2	Pvesp1_R	GAATTACCAATGGGTTCGGCA
	rps15 - rpl2	Pvesp2_F	TAGTAGTCGTCGGCGTTTTTCT
	rps15 - rpl2	Pvesp2_R	AACCCTGTAGACCATCCCCAT
	rrn5 - rps19	Pvesp3_F	TCCTAGGCGTAGAGGAACCA
	rrn5 - rps19	Pvesp3_R	TTCTCCCAATTTATGACCCACCA

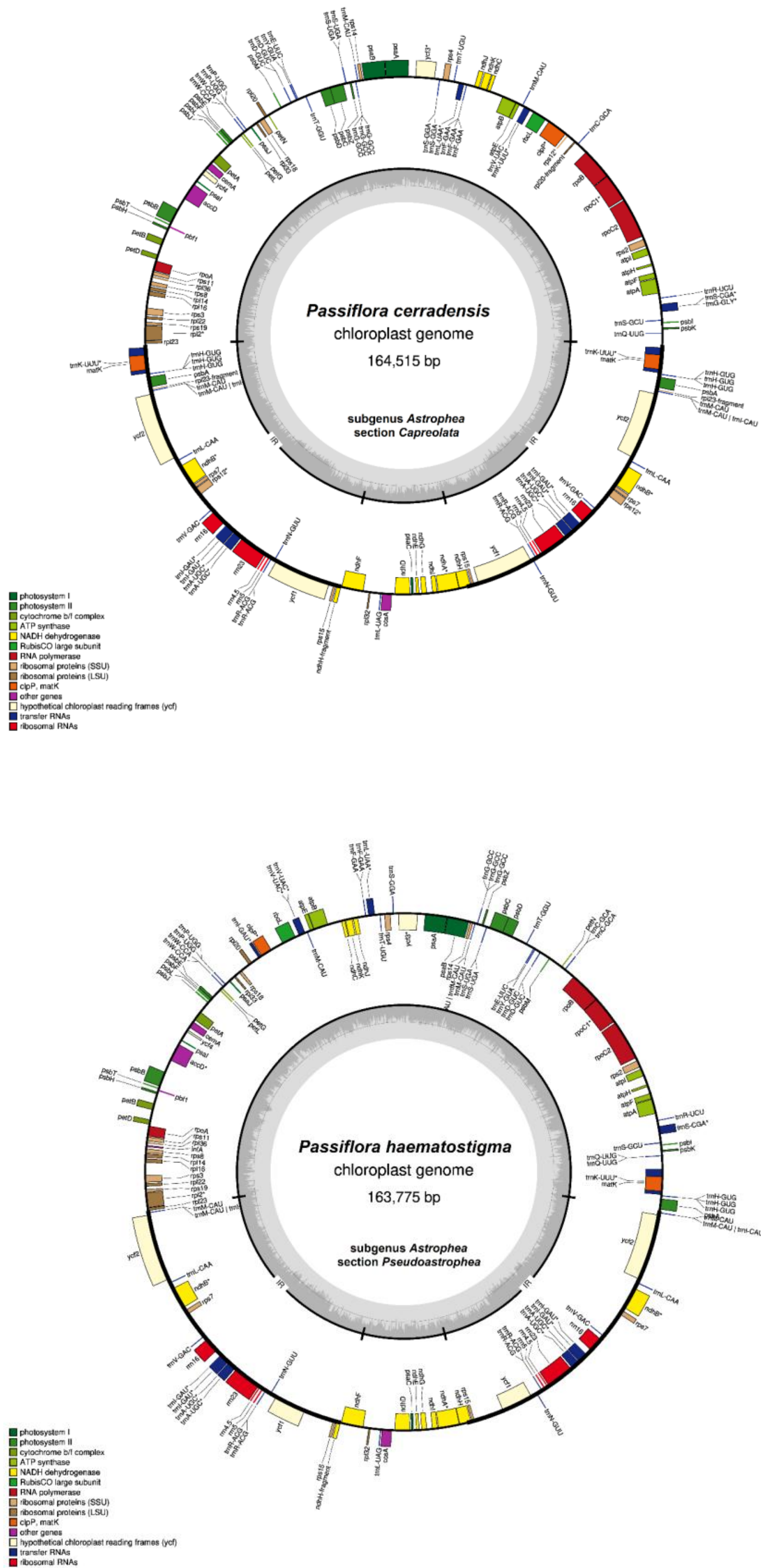
APPENDIX F. List of chloroplast genomes of *Passiflora* species available in NCBI's Genbank database as well as of *Adenia mannii* and *Populus trihocarpa*

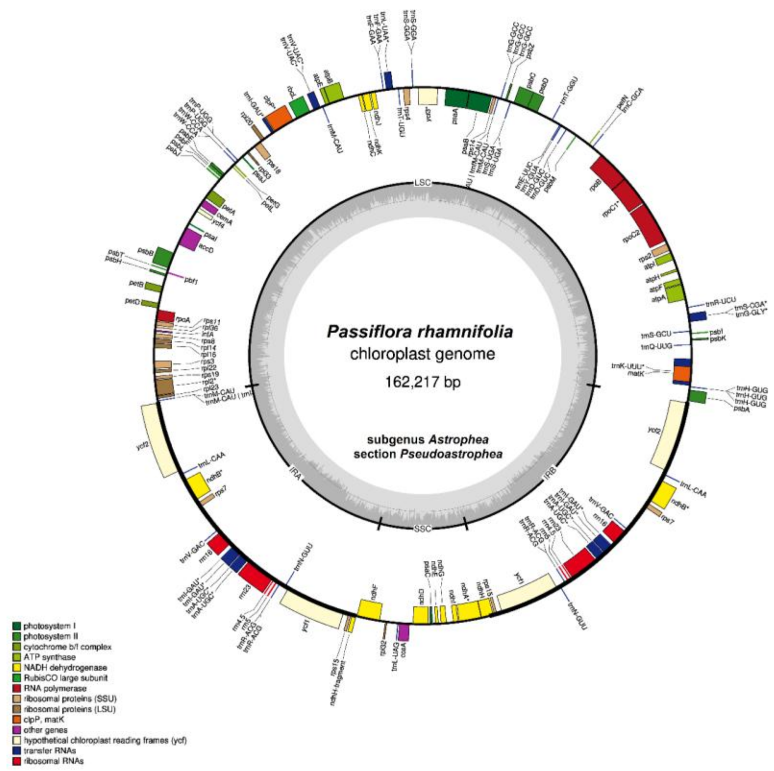
Subgenus	Species	Genbank accession number	Reference
Astrophea	*Passiflora pittieri*	NC_038125.1	Rabah et al. 2019
Decaloba	*Passiflora microstipula*	NC_043827.1	Shrestha et al. 2019
	Passiflora auriculata	NC_038119.1	Rabah et al. 2019
	Passiflora jatunsachensis	NC_043813.1	Shrestha et al. 2019
	Passiflora rufa	NC_043817.1	Shrestha et al. 2019
	Passiflora tenuiloba	NC_043816.1	Shrestha et al. 2019
	Passiflora filipes	NC_043822.1	Shrestha et al. 2019
	Passiflora lutea	NC_043815.1	Shrestha et al. 2019
	Passiflora affinis	NC_043823.1	Shrestha et al. 2019
	Passiflora biflora	NC_038120.1	Rabah et al. 2019
	Passiflora misera	NC_043821.1	Shrestha et al. 2019
Deidamioides	*Passiflora arbelaezii*	NC_043819.1	Shrestha et al. 2019
	Passiflora obovata	NC_043824.1	Shrestha et al. 2019
Tetraphatea	*Passiflora tetrandra*	NC_043820.1	Shrestha et al. 2019
Passiflora	*Passiflora actinia*	NC_038118.1	Rabah et al. 2019
	Passiflora quadrangularis	NC_038126.1	Rabah et al. 2019
	Passiflora laurifolia	NC_038121.1	Rabah et al. 2019
	Passiflora nitida	NC_038123.1	Rabah et al. 2019
	Passiflora cincinnata	NC_037690.1	Pacheco et al. 2017
	Passiflora edulis	NC_034285.1	Cauz-Santos et al. 2017
	Passiflora vitifolia	NC_038128.1	Rabah et al. 2019
	Passiflora ligularis	NC_038122.1	Rabah et al. 2019
	Passiflora serratodigitata	NC_038127.1	Rabah et al. 2019
	Passiflora serratifolia	NC_038129.1	Rabah et al. 2019
	Passiflora menispermifolia	NC_043826.1	Shrestha et al. 2019
	Passiflora oerstedii	NC_038124.1	Rabah et al. 2019
	Passiflora retipetala	NC_038188.1	Rabah et al. 2019
	Passiflora foetida	NC_043825.1	Shrestha et al. 2019
	Adenia mannii	NC_043791.1	Shrestha et al. 2019
	Populus trihocarpa	NC_009143.1	Tuskan et al. 2006

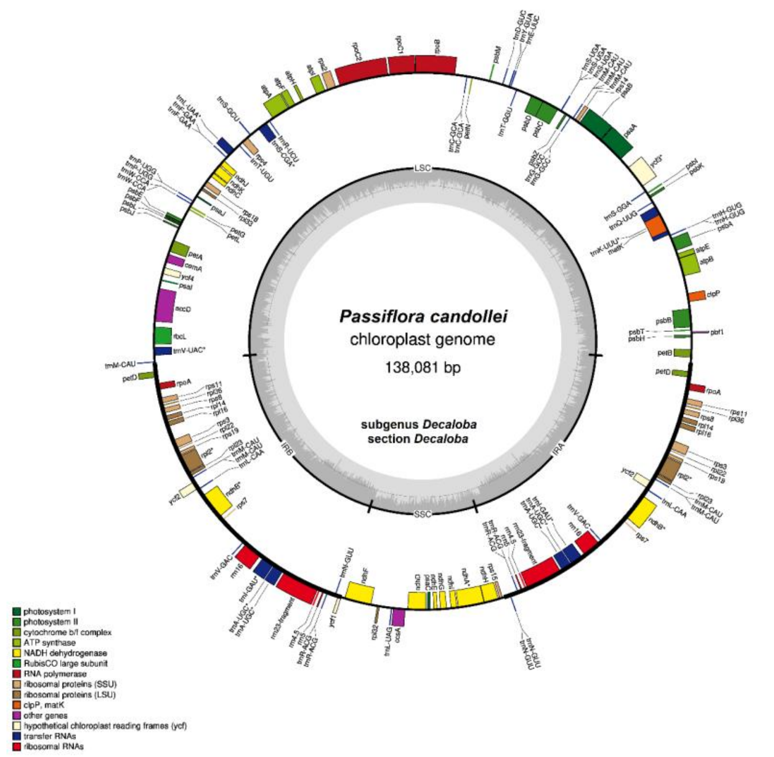
APPENDIX F. Illumina sequencing results of *Passiflora* chloroplast genomes

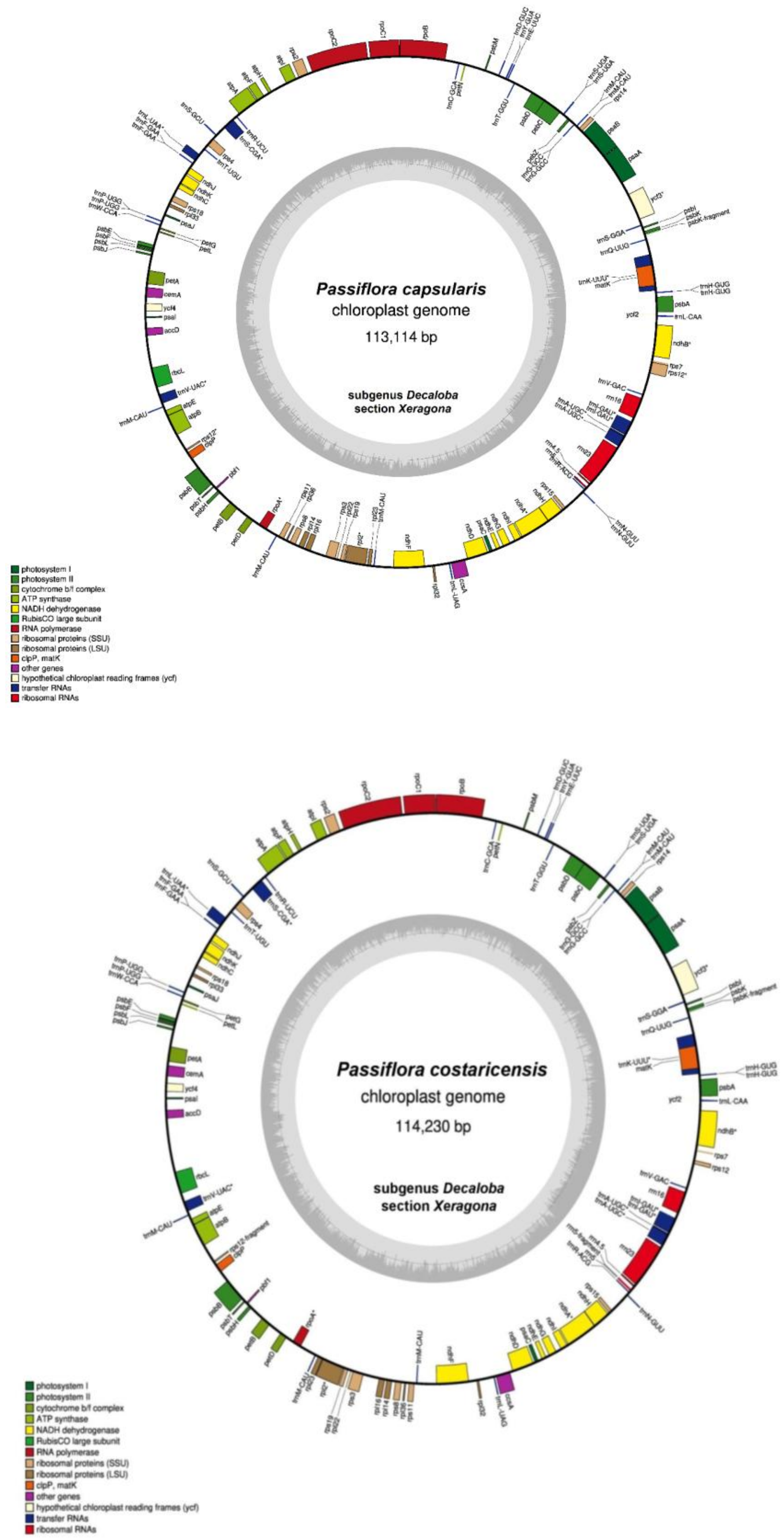
Subgenus	Species	Sequence technology	Data (Gb)	Total no. of reads	cpDNA reads	Cp genome size	Coverage (*X*)
Passiflora	*Passiflora cerradensis*	Illumina NextSeq 550	2,11	22,754,226	319,997	164,276	292
	Passiflora haematostigma	Illumina NextSeq 550	1,63	17,430,270	545,841	161,755	506
	Passiflora rhamnifolia	Illumina NextSeq 550	1,93	41,545,758	3,449,715	162,601	1,612
Decaloba	*Passiflora candollei*	Illumina NextSeq 550	1,65	17,418,848	679,600	138,081	738
	Passiflora capsularis	Illumina NextSeq 550	1,95	20,764,460	1,528,512	113,152	2,026
	Passiflora suberosa	Illumina NextSeq 550	1,84	39,288,936	3,498,296	156,355	1,700
	Passiflora vespertilio	Illumina NextSeq 550	2,30	24,924,420	1,596,101	138,456	1,730
Deidamioides	*Passiflora contracta*	Illumina NextSeq 550	1,84	40,087,112	1,498,172	167,141	681
	Passiflora deidamioides	Illumina NextSeq 550	1,83	39,856,504	2,546,610	167,827	1,152
Passiflora	*Passiflora alata*	Illumina NextSeq 550	1,65	34,408,974	654,895	148,113	335
	Passiflora cristalina	Illumina NextSeq 550	2,09	22,320,626	892,441	145,053	922
	Passiflora edmundoi	Illumina NextSeq 550	2,10	22,450,584	935,400	142,646	983
	Passiflora loefgrenii	Illumina NextSeq 550	1,81	38,113,682	6,500,728	146,320	3,375
	Passiflora miniata	Illumina NextSeq 550	1,32	13,862,202	225,762	151,758	223
	Passiflora mucronata	Illumina NextSeq 550	1,91	20,276,596	884,293	151,259	875
	Passiflora recurva	Illumina NextSeq 550	1,67	17,888,470	746,730	151,976	737
	Passiflora watsoniana	Illumina NextSeq 550	1,57	16,526,196	252,586	146,782	258
	Dilkea retusa	Illumina NextSeq 550	1,99	21,899,744	133,896	161,923	124
	Mitostemma brevifilis	Illumina NextSeq 550	2,20	23,891,314	1,822,971	162,350	1,684

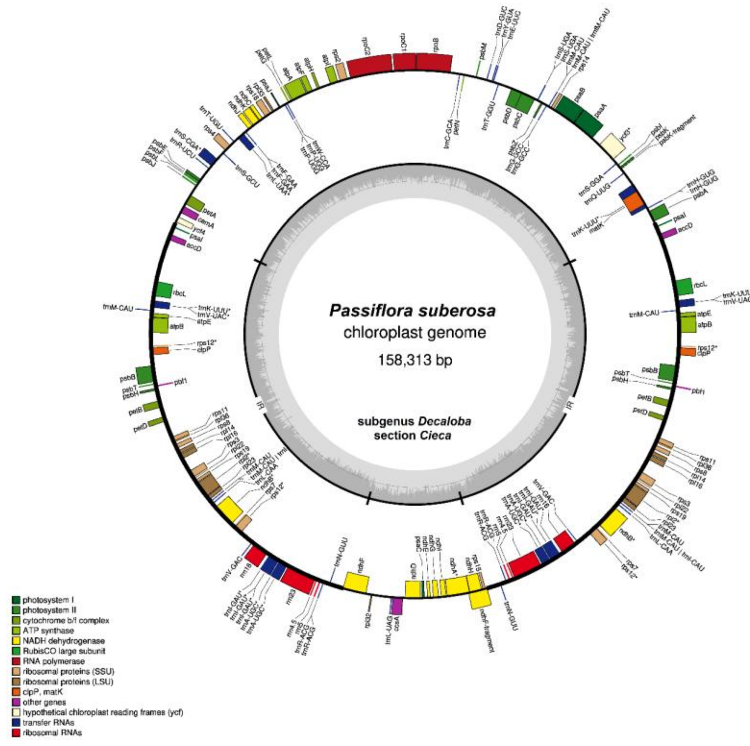
APPENDIX G. Chloroplast genome maps of 18 *Passiflora* species in addition to the Passifloraceae, *Dilkea retusa* and *Mitostemma brevifilis*.

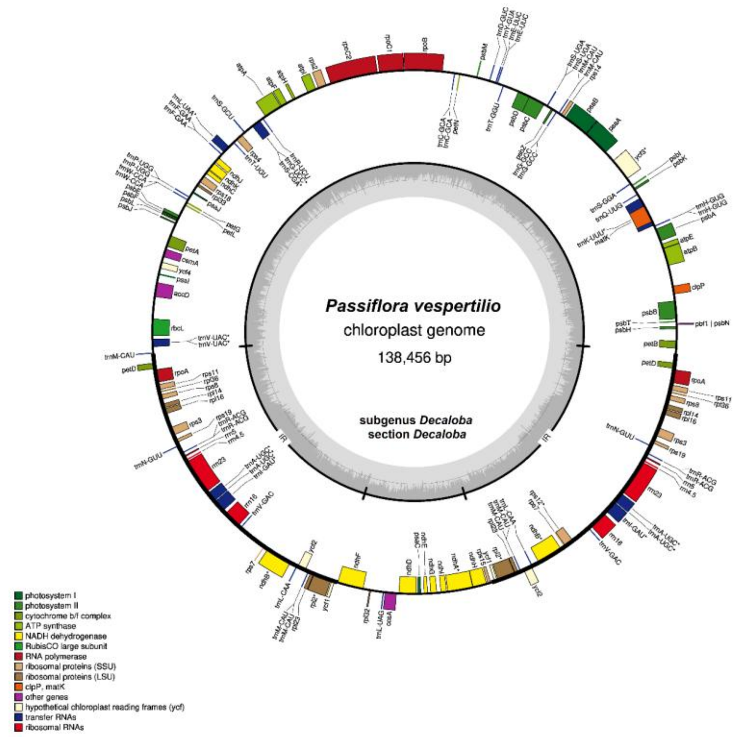


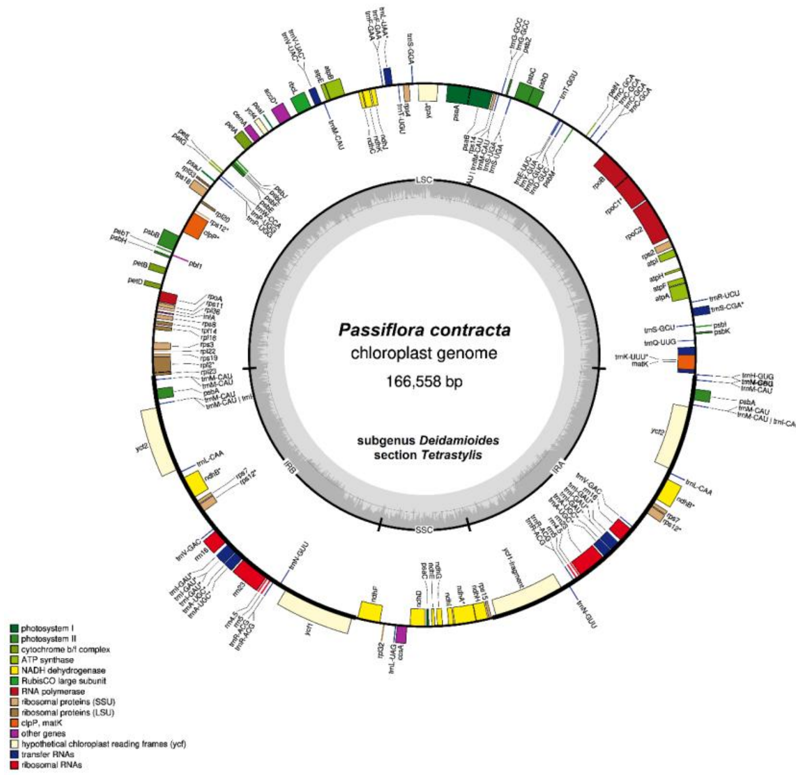

Passiflora rhamnifolia
chloroplast genome
162,217 bp
subgenus Astrophea
section Pseudoastrophea
LSC
IRB
SSC
IRA
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs


Passiflora candollei
chloroplast genome
138,081 bp
subgenus Decaloba
section Decaloba
LSC
IRB
SSC
IRA
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

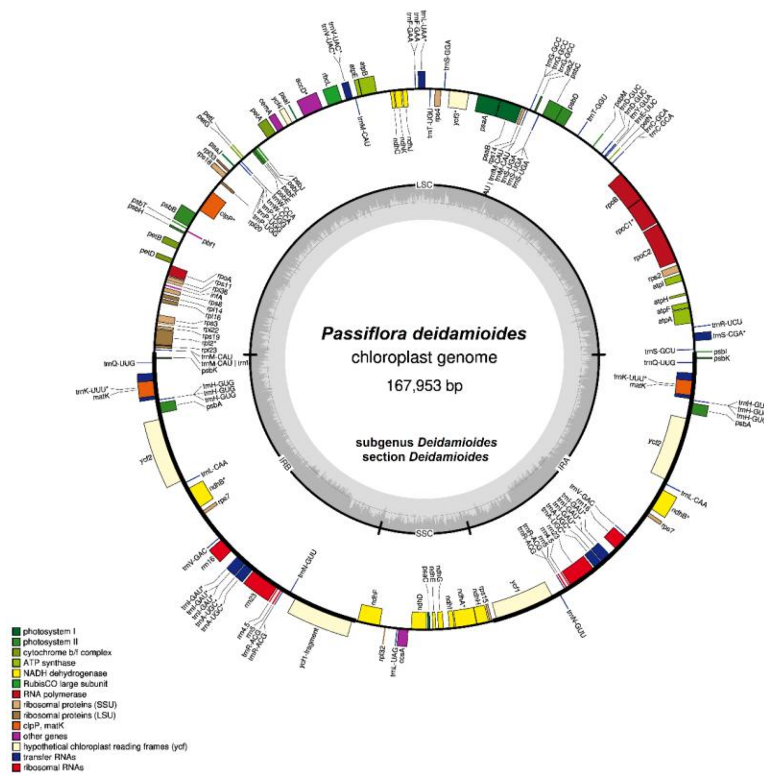

Passiflora capsularis
chloroplast genome
113,114 bp
subgenus Decaloba
section Xeragona
Passiflora costaricensis
chloroplast genome
114,230 bp
subgenus Decaloba
section Xeragona


Passiflora suberosa
chloroplast genome
158,313 bp
subgenus *Decaloba*
section *Cieca*
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

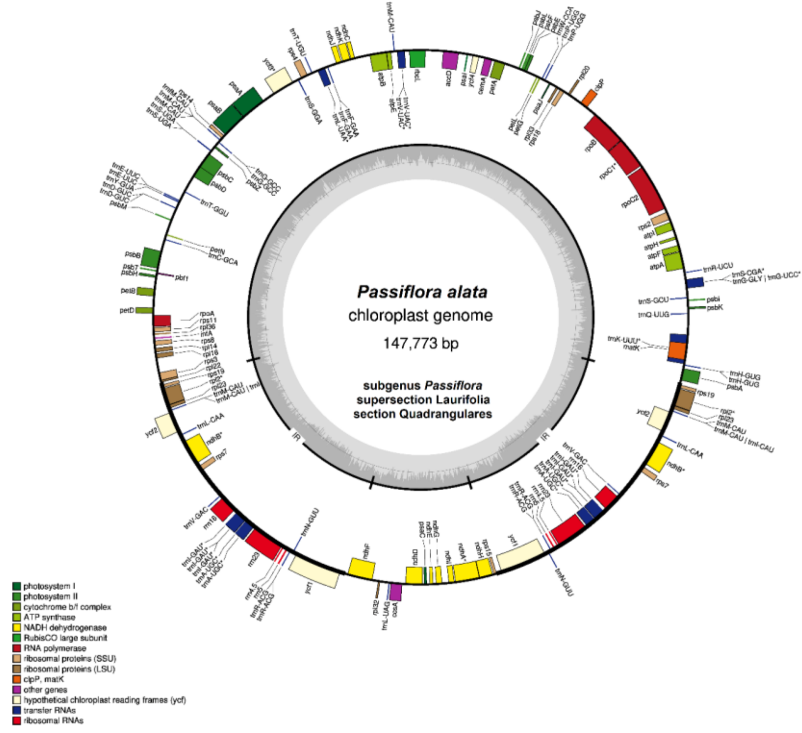

Passiflora vespertilio
chloroplast genome
138,456 bp
subgenus *Decaloba*
section *Decaloba*
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

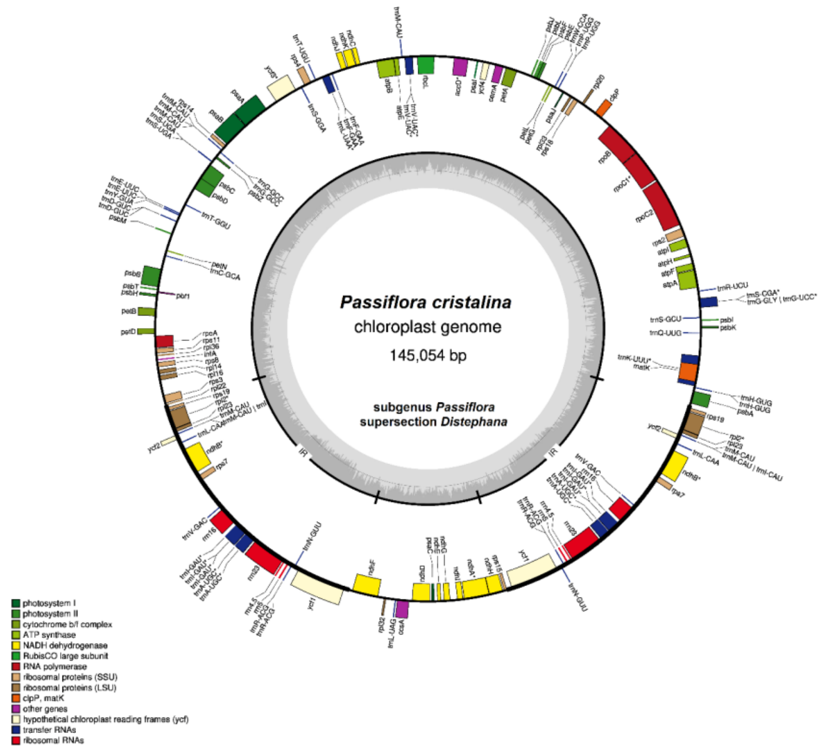


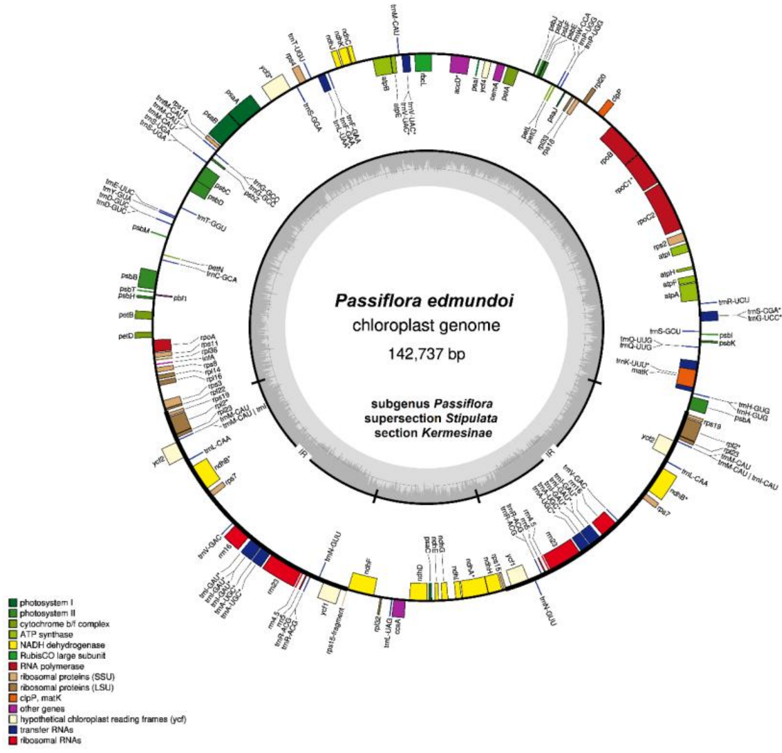
Passiflora contracta
chloroplast genome
166,558 bp
subgenus *Deidamioides*
section *Tetrastylis*
LSC
IRA
IRB
SSC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

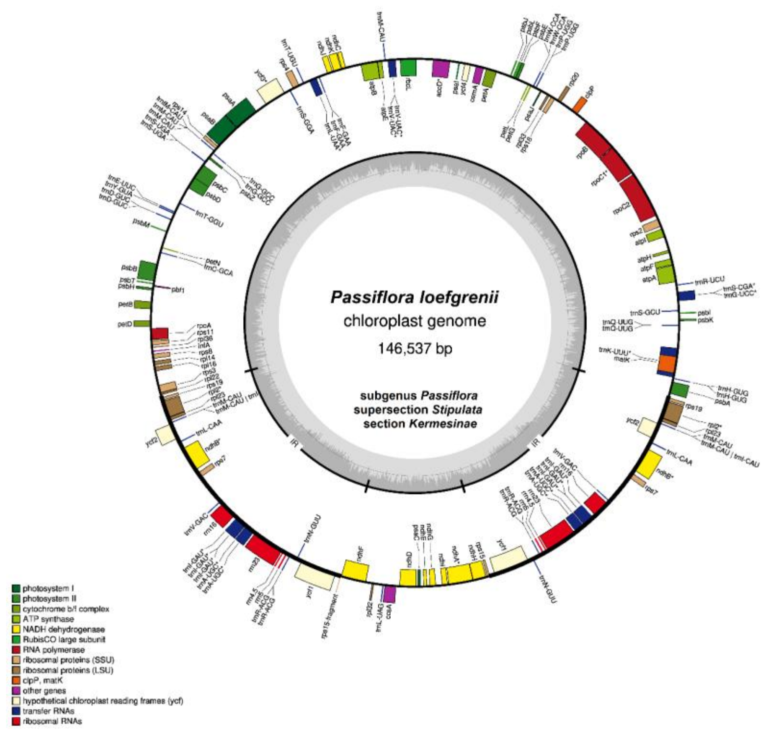


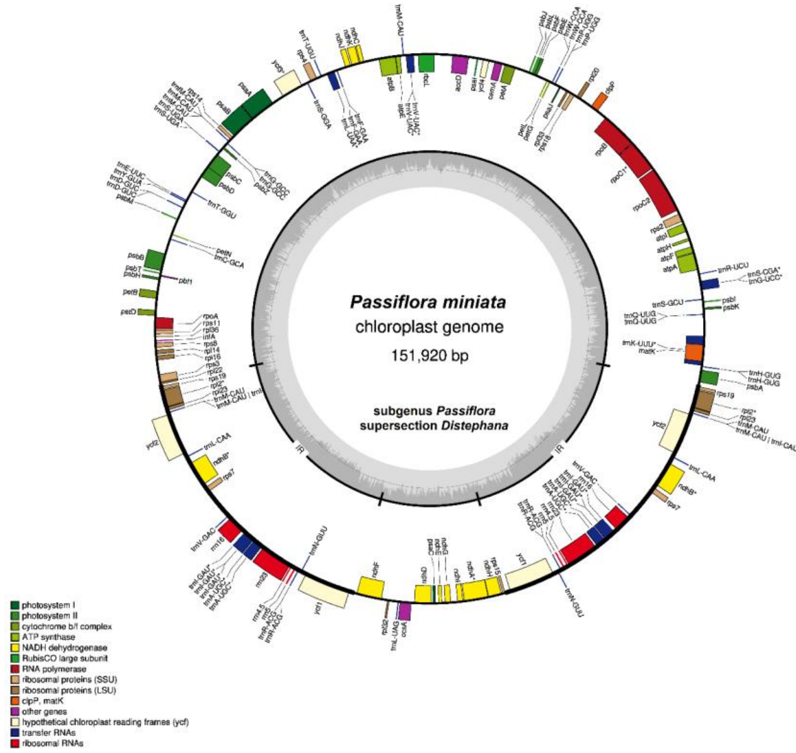
Passiflora deidamioides
chloroplast genome
167,953 bp
subgenus *Deidamioides*
section *Deidamioides*
LSC
IRA
IRB
SSC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs


Passiflora alata
chloroplast genome
147,773 bp
subgenus *Passiflora*
supersection Laurifolia
section Quadrangulares
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

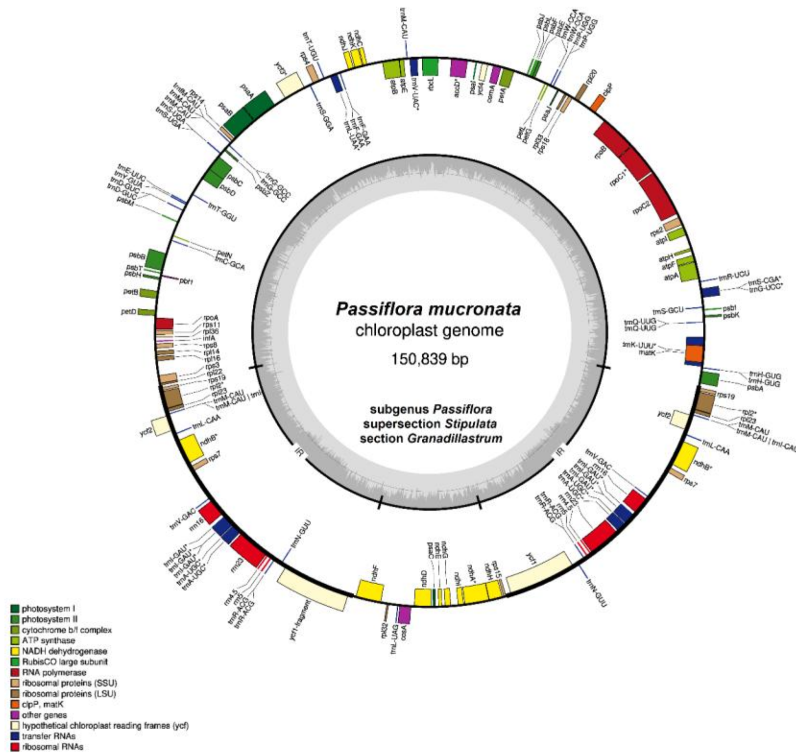

Passiflora cristalina
chloroplast genome
145,054 bp
subgenus *Passiflora*
supersection *Distephana*
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs


Passiflora edmundoi
chloroplast genome
142,737 bp
subgenus Passiflora
supersection Stipulata
section Kermesinae
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

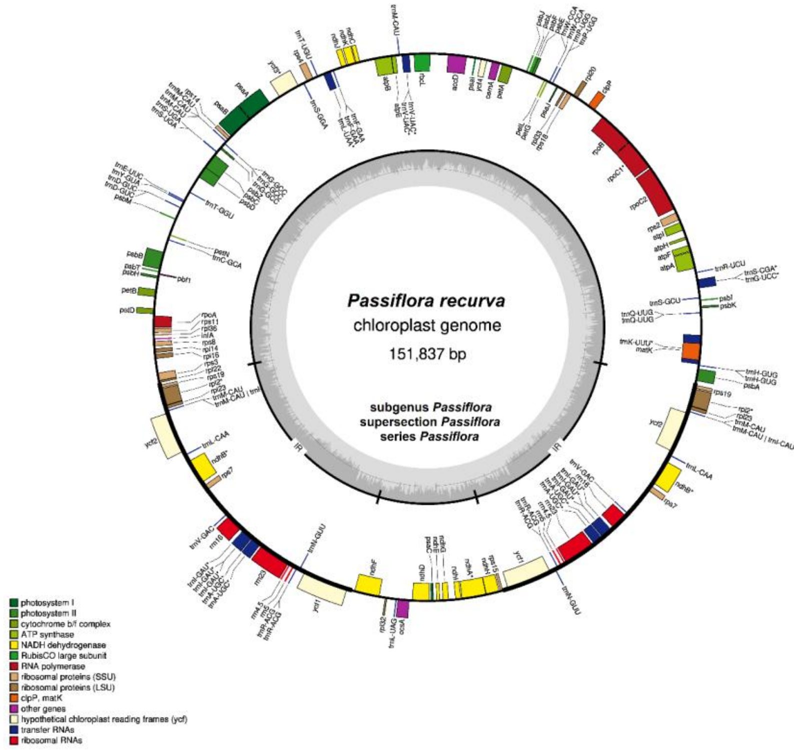

Passiflora loefgrenii
chloroplast genome
146,537 bp
subgenus Passiflora
supersection Stipulata
section Kermesinae
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs



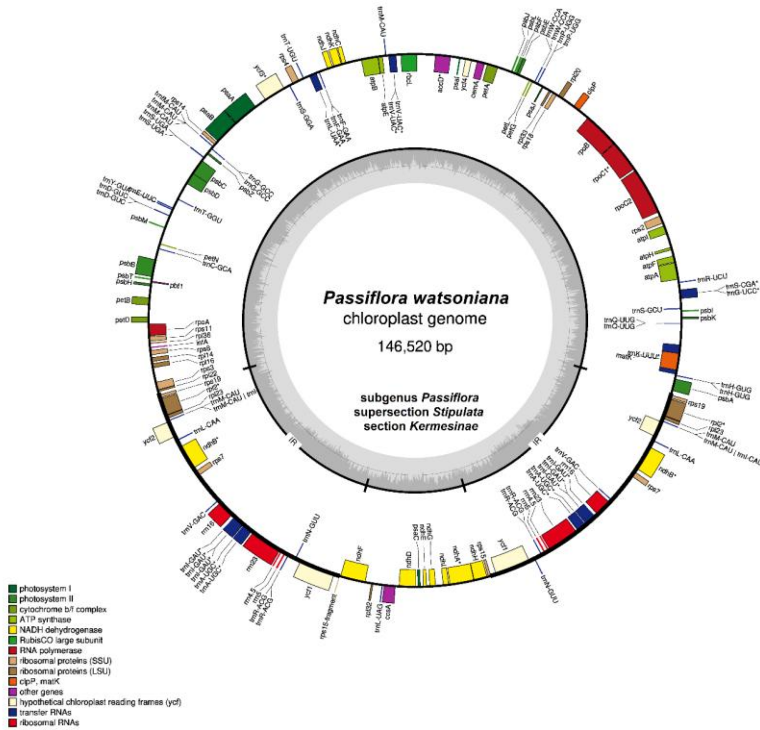
Passiflora miniata
chloroplast genome
151,920 bp
subgenus *Passiflora*
supersection *Distephana*
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs



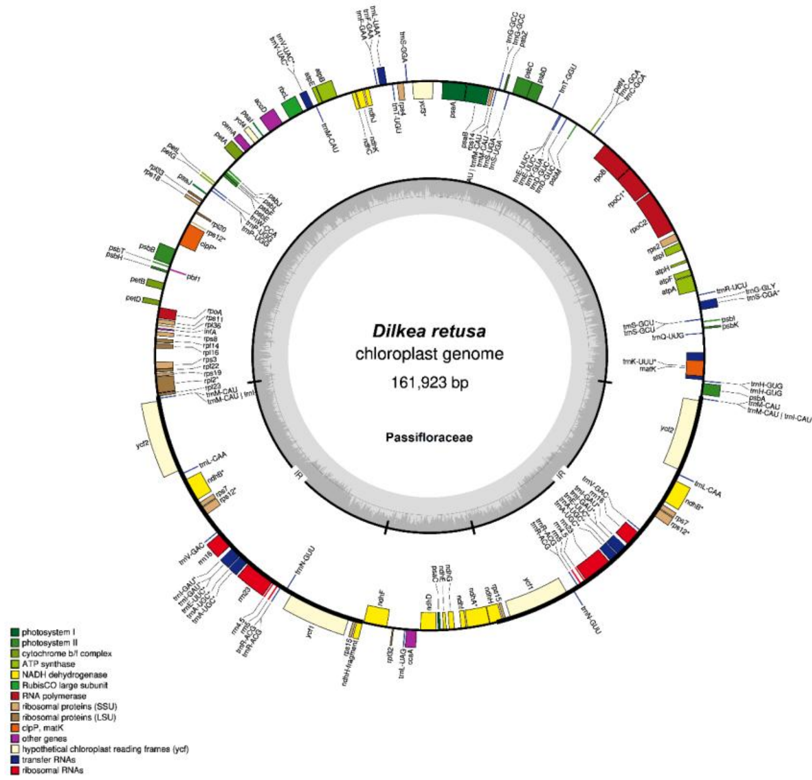
Passiflora mucronata
chloroplast genome
150,839 bp
subgenus *Passiflora*
supersection *Stipulata*
section *Granadillastrum*
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

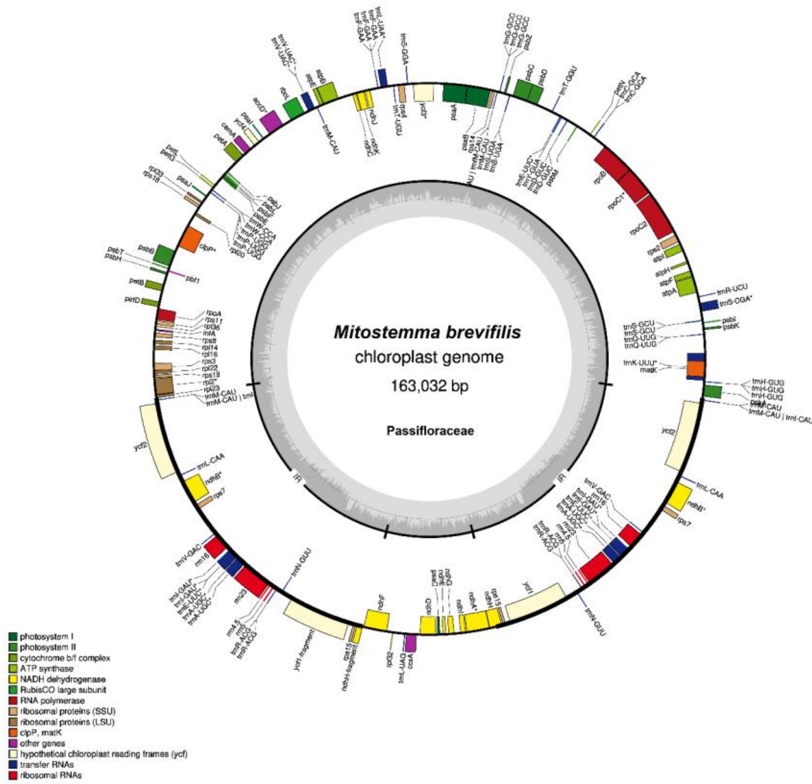


Passiflora recurva
chloroplast genome
151,837 bp
subgenus Passiflora
supersection Passiflora
series Passiflora
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

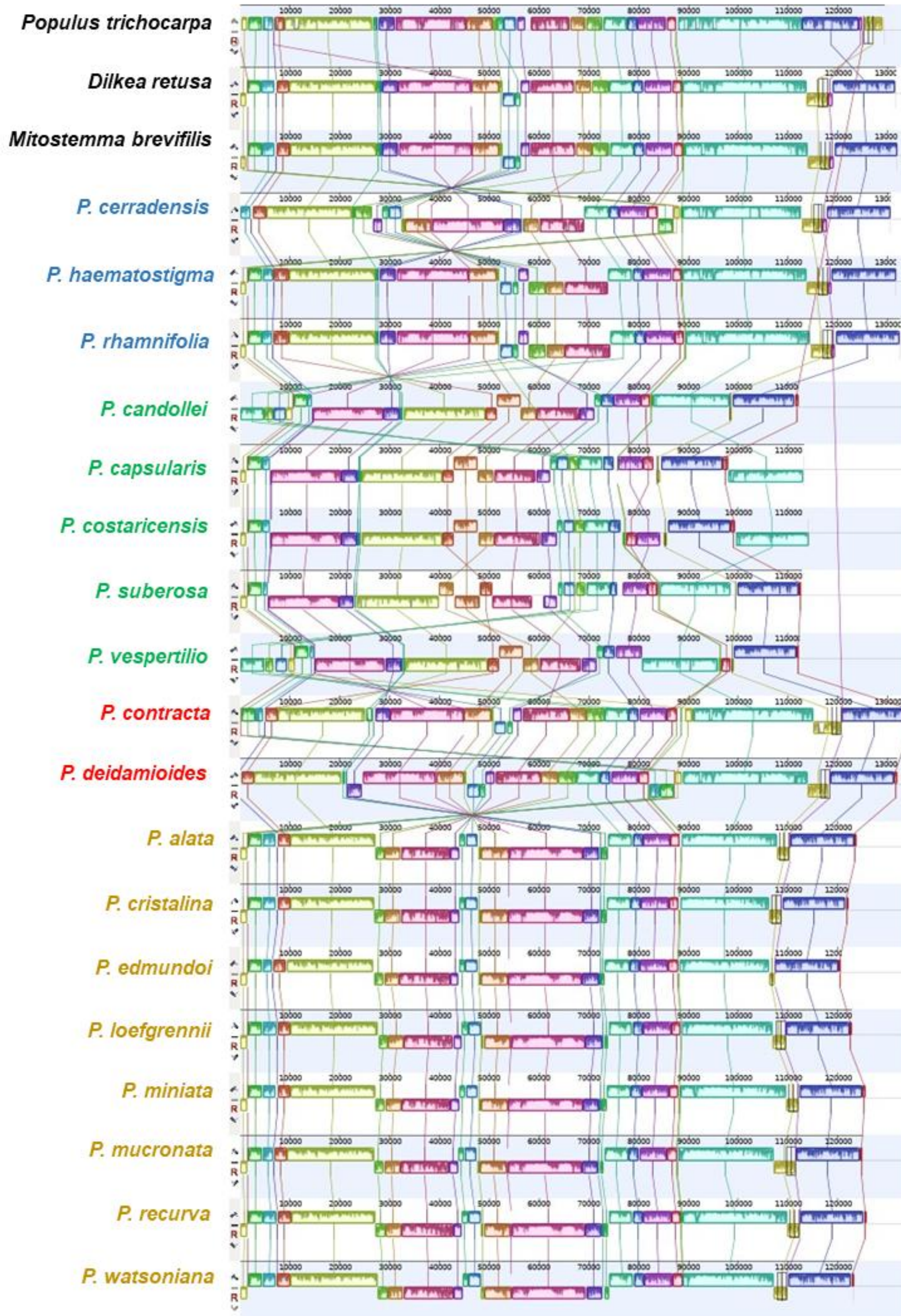


Passiflora watsoniana
chloroplast genome
146,520 bp
subgenus Passiflora
supersection Stipulata
section Kermesinae
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs


Dilkea retusa
chloroplast genome
161,923 bp
Passifloraceae
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs


Mitostemma brevifilis
chloroplast genome
163,032 bp
Passifloraceae
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

APPENDIX H. Comparative analysis of chloroplast genomes from 18 *Passiflora* species in addition to the Passifloraceae, *Dilkea retusa* and *Mitostemma brevifilis*. *Populus trichocarpa* (Salicaceae) was used as a reference.



APPENDIX I. List of species used in the plastid phylogenomics and nuclear phylogenetic studies.

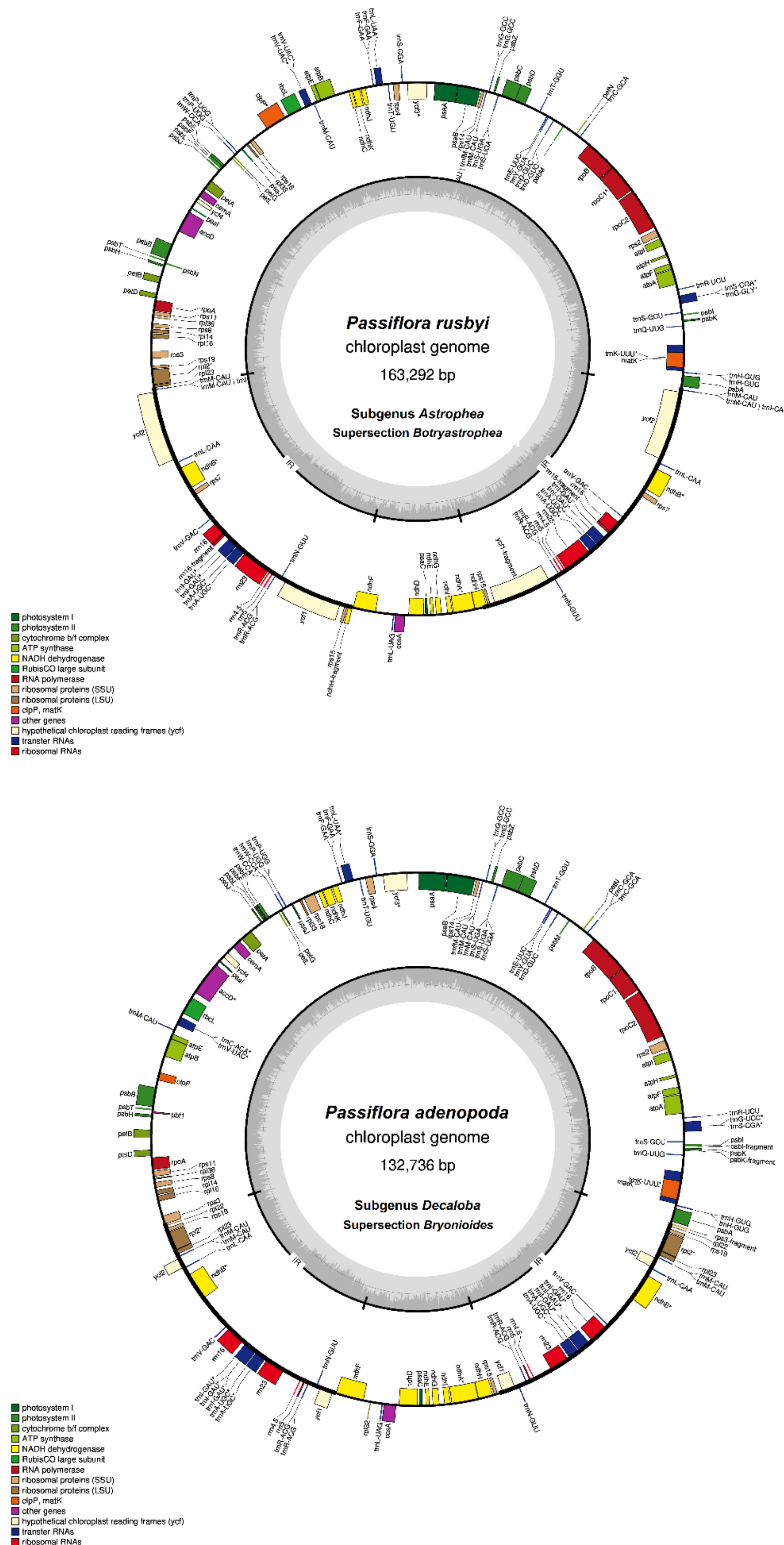
Subgenus	Species	Data	Plastid Genes	Whole Plastid	18S/26S
Astrophea	*Passiflora cerradensis*	This study	X		X
	Passiflora haematostigma	This study	X		X
	Passiflora pittieri	NC_038125.1	X		
	Passiflora rhamnifolia	This study	X		X
	Passiflora rusbyi	This study	X		X
Decaloba	*Passiflora adenopoda*	This study	X		X
	Passiflora affinis	NC_043823.1	X		
	Passiflora auriculata	NC_038119.1	X		
	Passiflora biflora	NC_038120.1	X		
	Passiflora candollei	This study	X		X
	Passiflora capsularis	This study	X		X
	Passiflora costaricensis	This study	X		X
	Passiflora filipes	NC_043822.1	X		
	Passiflora intricata	This study	X		X
	Passiflora jatunsachensis	NC_043813.1	X		
	Passiflora lutea	NC_043815.1	X		
	Passiflora microstipula	NC_043827.1	X		
	Passiflora misera	NC_043821.1	X		
	Passiflora organensis	This study	X		X
	Passiflora rufa	NC_043817.1	X		
	Passiflora suberosa	This study	X		X
	Passiflora tenuiloba	NC_043816.1	X		
	Passiflora vespertilio	This study	X		X
	Passiflora xiikzozdz	This study	X		X
Deidamioides	*Passiflora arbelaezii*	NC_043819.1	X		
	Passiflora contracta	This study	X		X
	Passiflora deidamioides	This study	X		X
	Passiflora discophora	This study	X		X
	Passiflora obovata	NC_043824.1	X		
	Passiflora ovalis	This study	X		X
Tetraphatea	*Passiflora tetrandra*	This study	X		X
Passiflora	*Passiflora actinia*	NC_038118.1	X	X	
	Passiflora alata	This study	X	X	X
	Passiflora chaparensis	This study	X	X	X
	Passiflora cincinnata	NC_037690.1	X	X	
	Passiflora cristalina	This study	X	X	X
	Passiflora edmundoi	This study	X	X	X
	Passiflora edulis	NC_034285.1	X	X	
	Passiflora foetida	NC_043825.1	X	X	
	Passiflora garckei	This study	X	X	X
	Passiflora laurifolia	NC_038121.1	X	X	
	Passiflora ligularis	NC_038122.1	X	X	
	Passiflora loefgrenii	This study	X	X	X
	Passiflora	NC_043826.1	X	X	

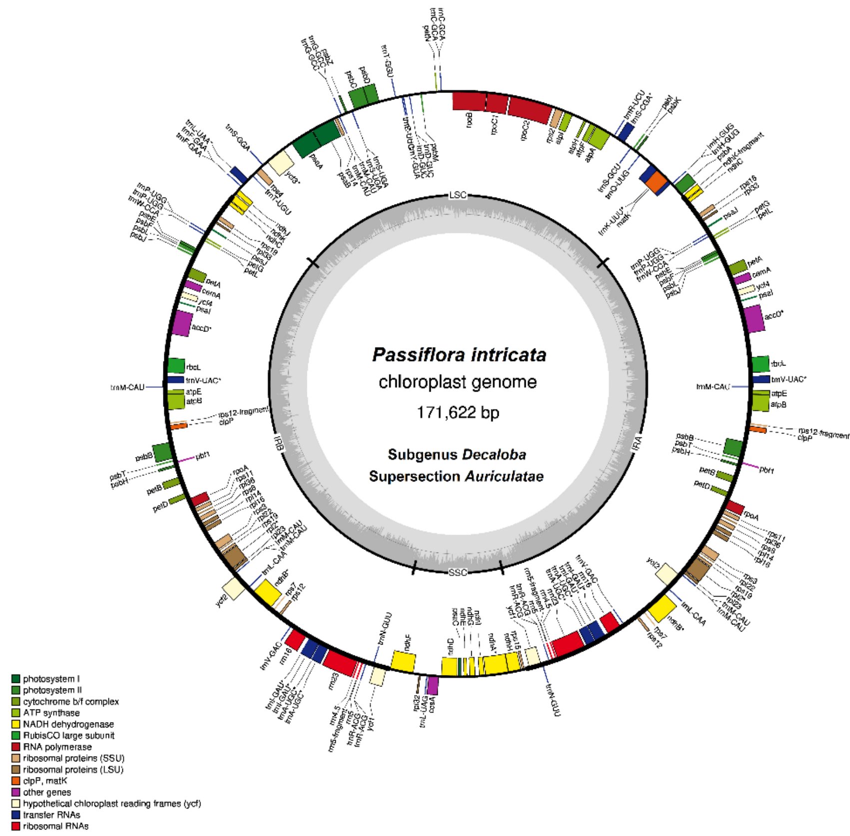
	menispermifolia				
	Passiflora miniata	This study	X	X	X
	Passiflora mucronata	This study	X	X	X
	Passiflora nitida	NC_038123.1	X	X	
	Passiflora oerstedii	NC_038124.1	X	X	
	Passiflora palenquensis	This study	X	X	X
	Passiflora phoenicea	This study	X	X	X
	Passiflora popenovii	This study	X	X	X
	Passiflora quadrangularis	NC_038126.1	X	X	
	Passiflora racemosa	This study	X	X	
	Passiflora recurva	This study	X	X	X
	Passiflora retipetala	NC_038188.1	X	X	
	Passiflora serratifolia	NC_038129.1	X	X	
	Passiflora serratodigitata	NC_038127.1	X	X	
	Passiflora vitifolia	NC_038128.1	X	X	
	Passiflora watsoniana	This study	X	X	X
	Adenia mannii	NC_043791.1	X		
	Dilkea retusa	This study	X		X
	Mitostemma brevifilis	This study	X		X
	Populus trihocarpa	NC_009143.1	X		

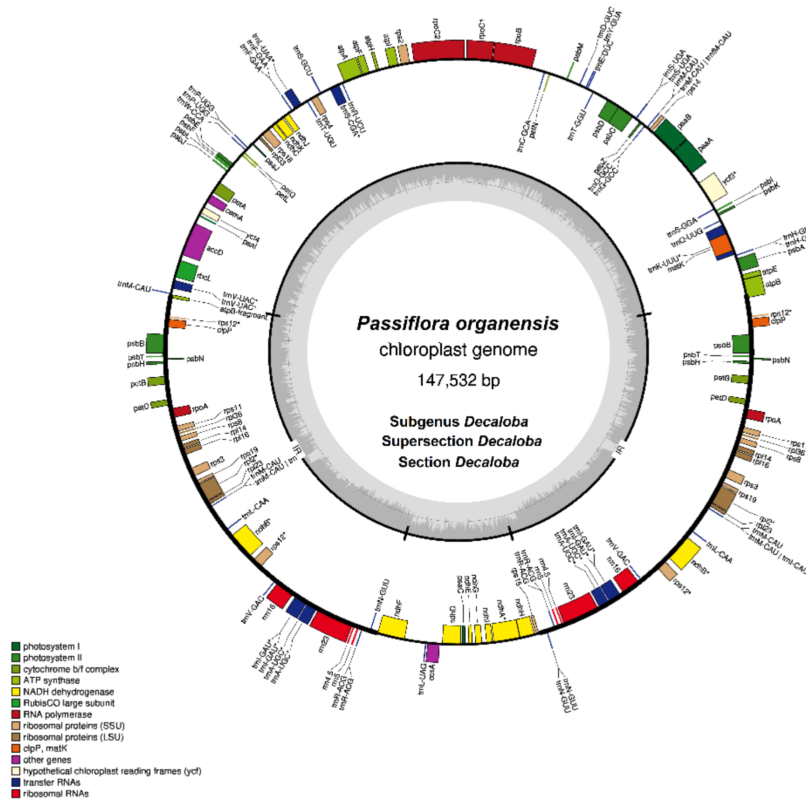
APPENDIX J. PacBio sequencing results.

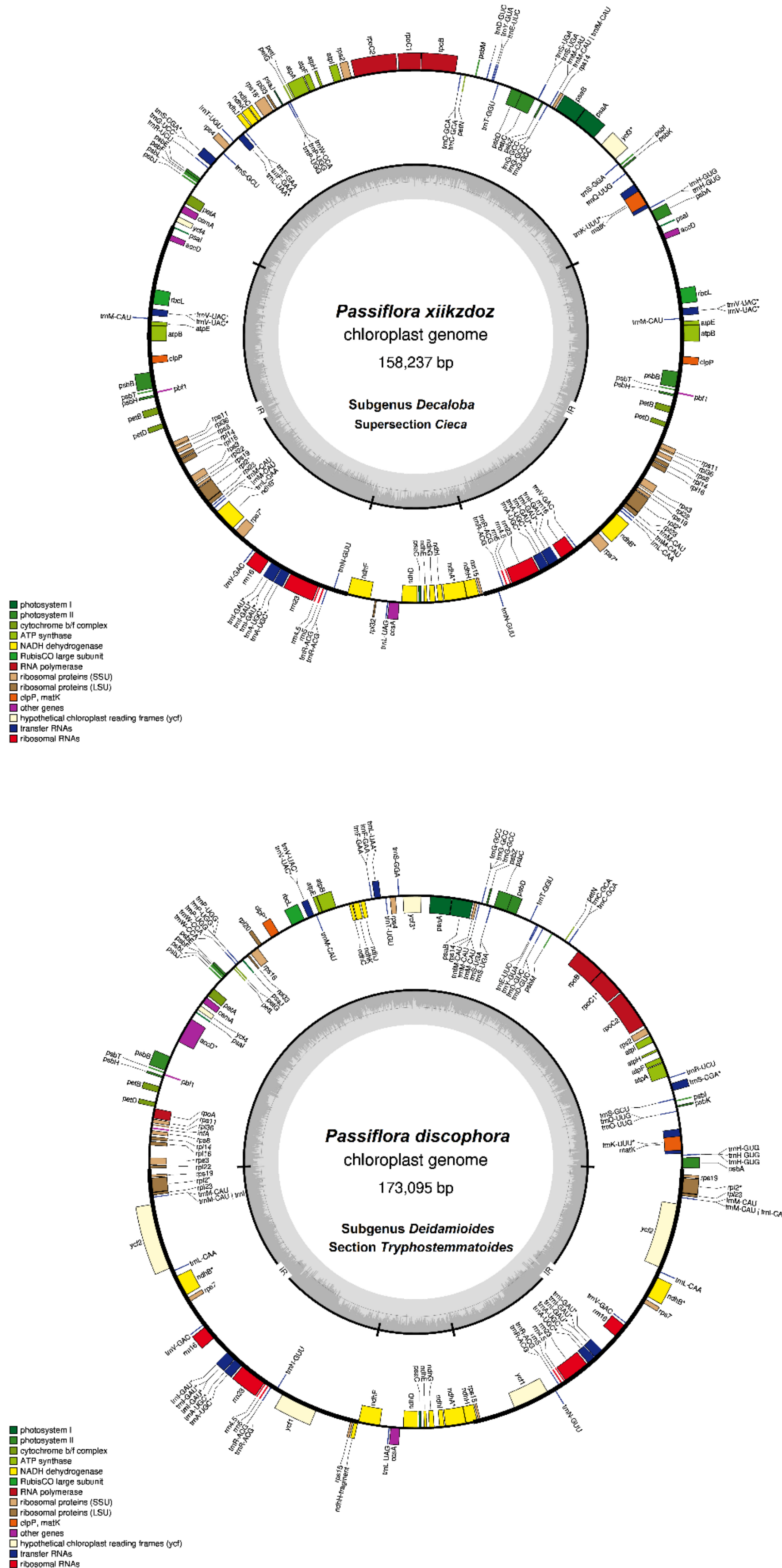
Subgenus	Species	Data (Gb)	Number of reads	Number of contigs	N50	Max. Contig size (bp)
Astrophea	*Passiflora rusbyi*	1,03	166,773	6,233	9,909	92,786
Decaloba	*Passiflora adenopoda*	0,13	27,93	4,907	9,294	124,092
	Passiflora intricata	3,30	623,544	5,451	8,233	154,278
	Passiflora xiikzozdz	0,46	84,887	5,265	8,187	44,588
Deidamioides	*Passiflora discophora*	2,05	62,393	6,932	11,293	109,214
	Passiflora ovalis	0,50	90,496	5,606	9,741	73,360
Passiflora	*Passiflora chaparensis*	1,2	724,271	6,385	8,233	127,466
	Passiflora garckei	3,50	547,863	6,415	9,271	77,447
	Passiflora palenquensis	3,07	472,178	6,522	8,580	115,869
	Passiflora phoenicea	1,2	250,990	5,715	8,577	93,946
	Passiflora popenovii	1,2	479,303	6,132	8,421	139,036
	Passiflora racemosa	0,54	15,888	3,407	9,856	127,276
Tetrapathea	*Passiflora tetrandra*	0,42	62,393	6,974	8,329	72,454

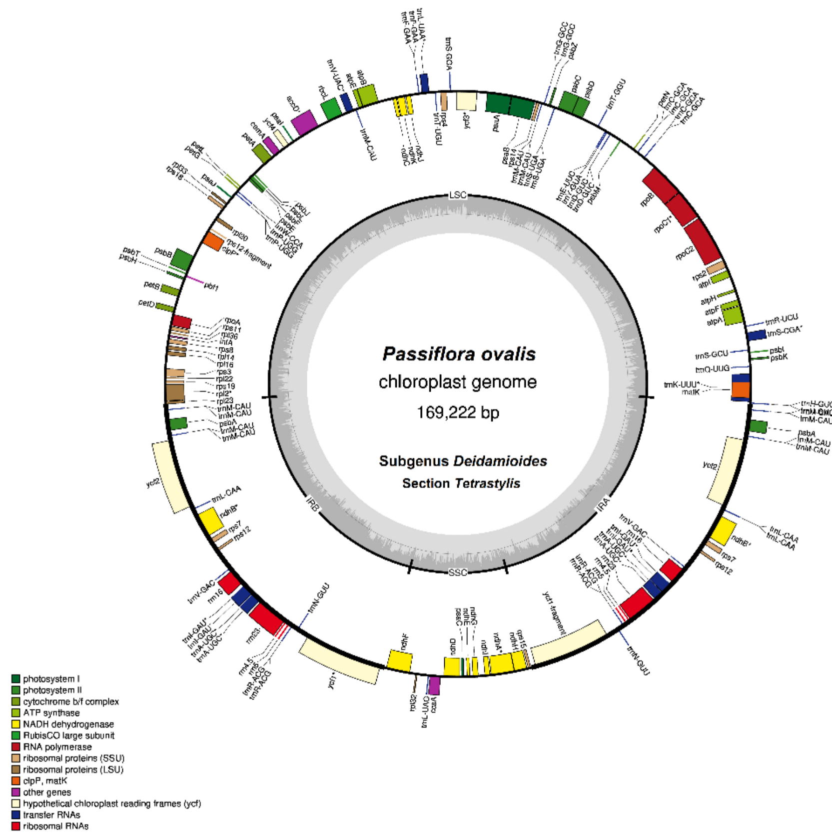
APPENDIX K. Chloroplast genome maps of 14 *Passiflora* species assembled with PacBio reads.



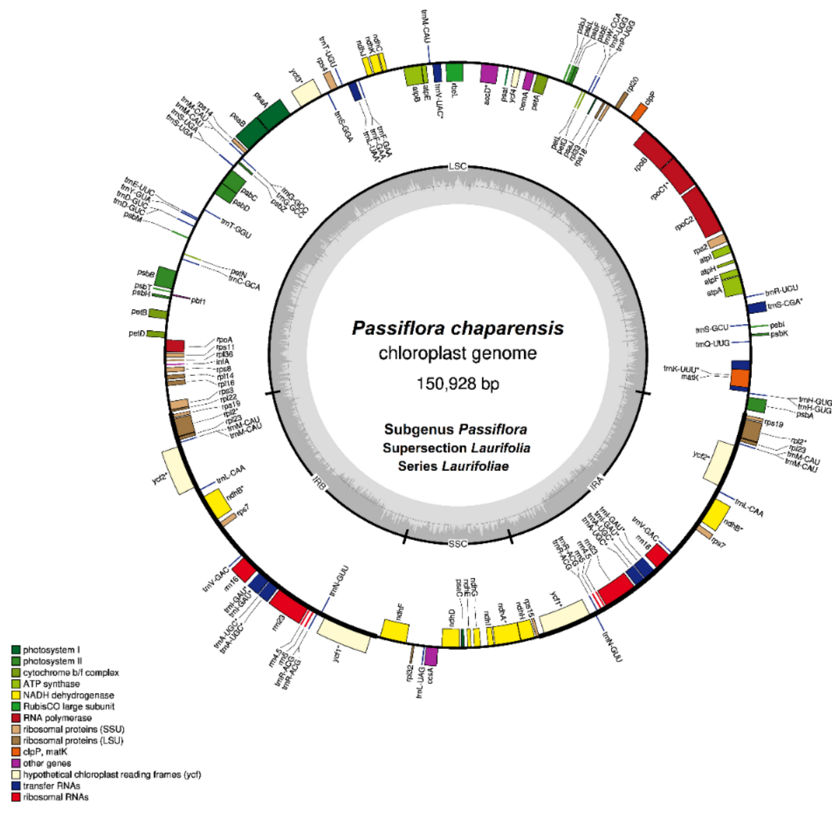

Passiflora intricata
chloroplast genome
171,622 bp
Subgenus Decaloba
Supersection Auriculatae
LSC
SSC
IRA
IRB
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs


Passiflora organensis
chloroplast genome
147,532 bp
Subgenus Decaloba
Supersection Decaloba
Section Decaloba
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

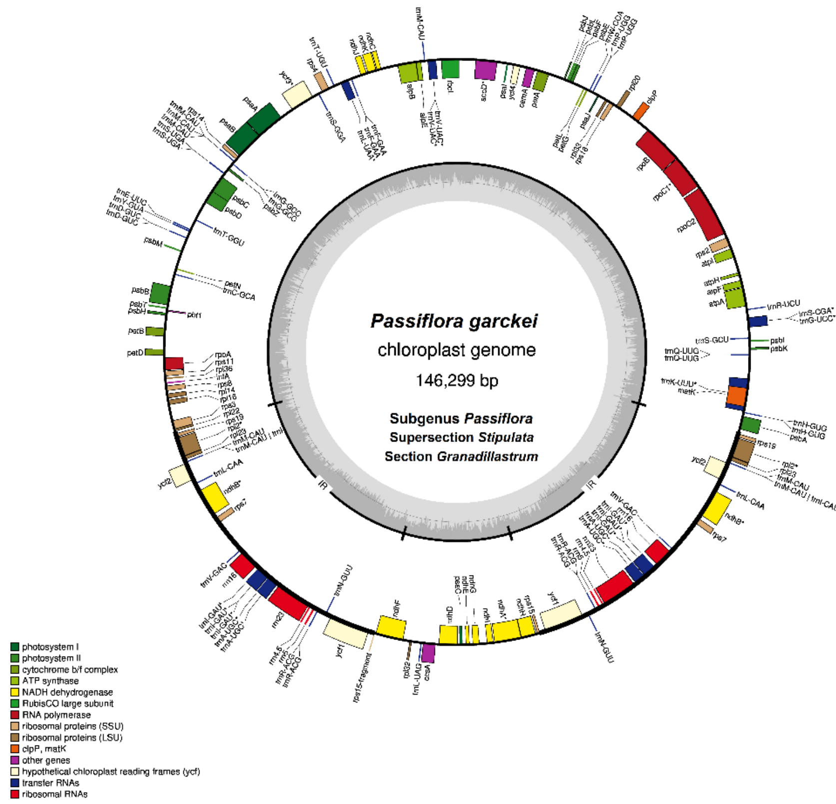

Passiflora xiikzdoz
chloroplast genome
158,237 bp
Subgenus Decaloba
Supersection Cieca
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs
Passiflora discophora
chloroplast genome
173,095 bp
Subgenus Deidamioides
Section Tryphostemmatoides
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

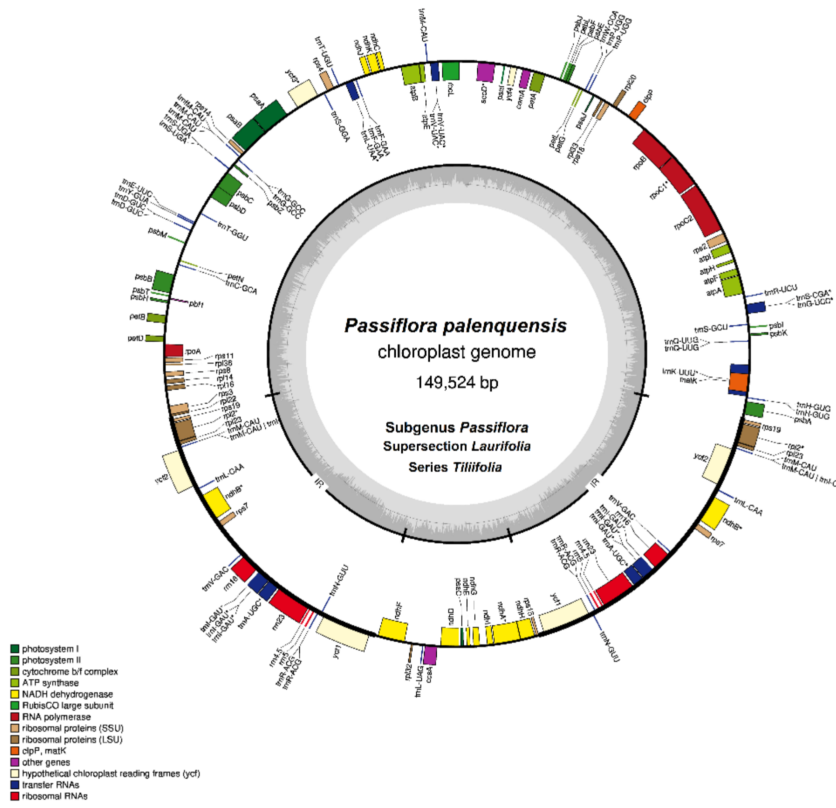


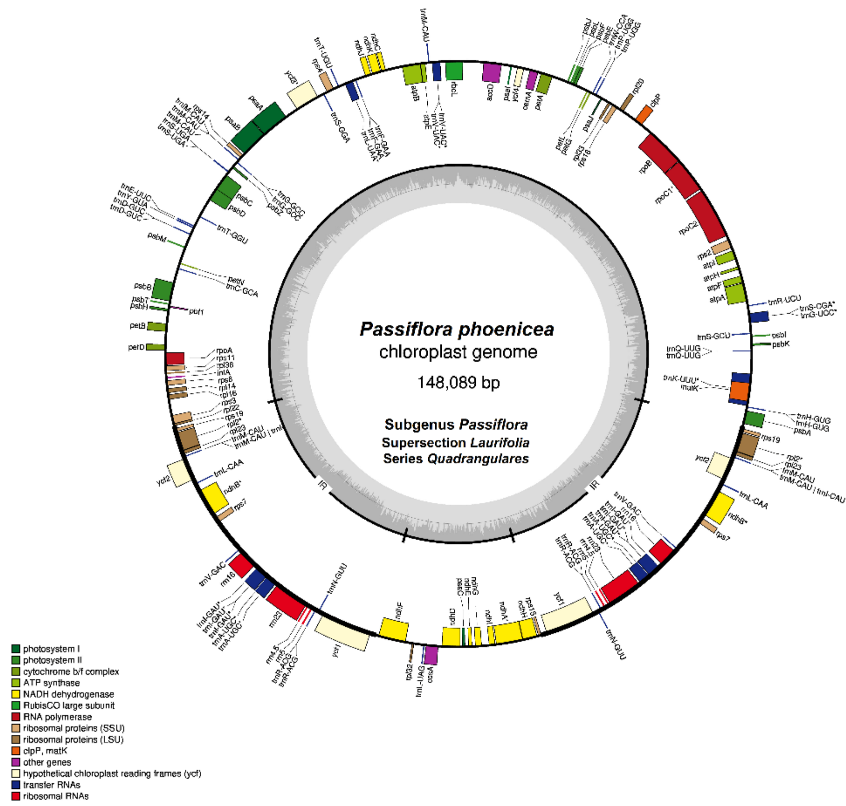
Passiflora ovalis
chloroplast genome
169,222 bp
Subgenus Deidamioides
Section Tetrastylis
LSC
SSC
IRA
IRB
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

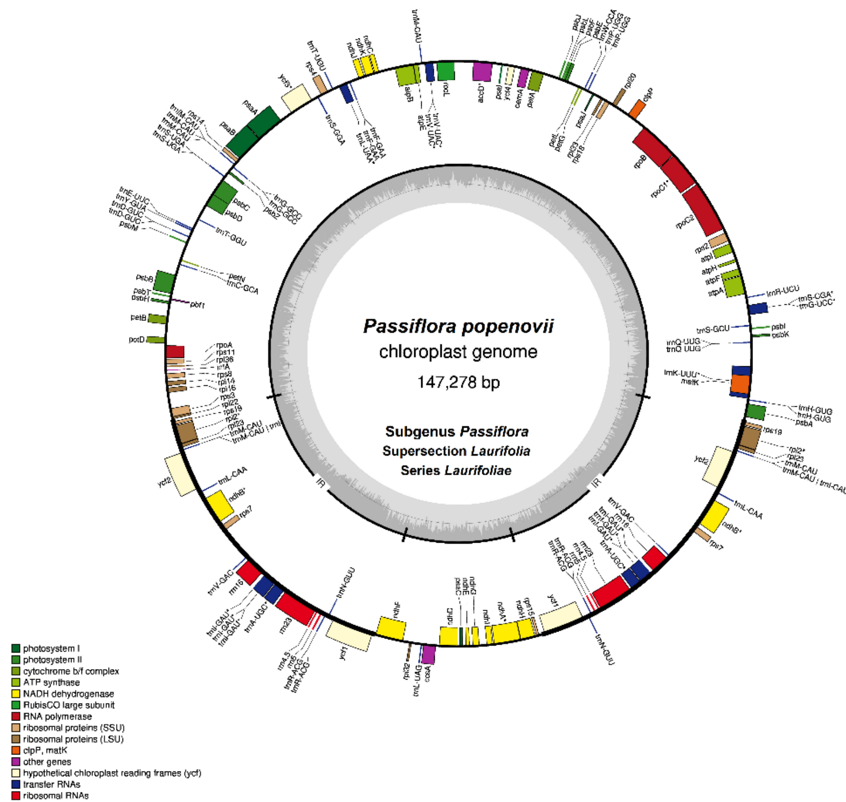


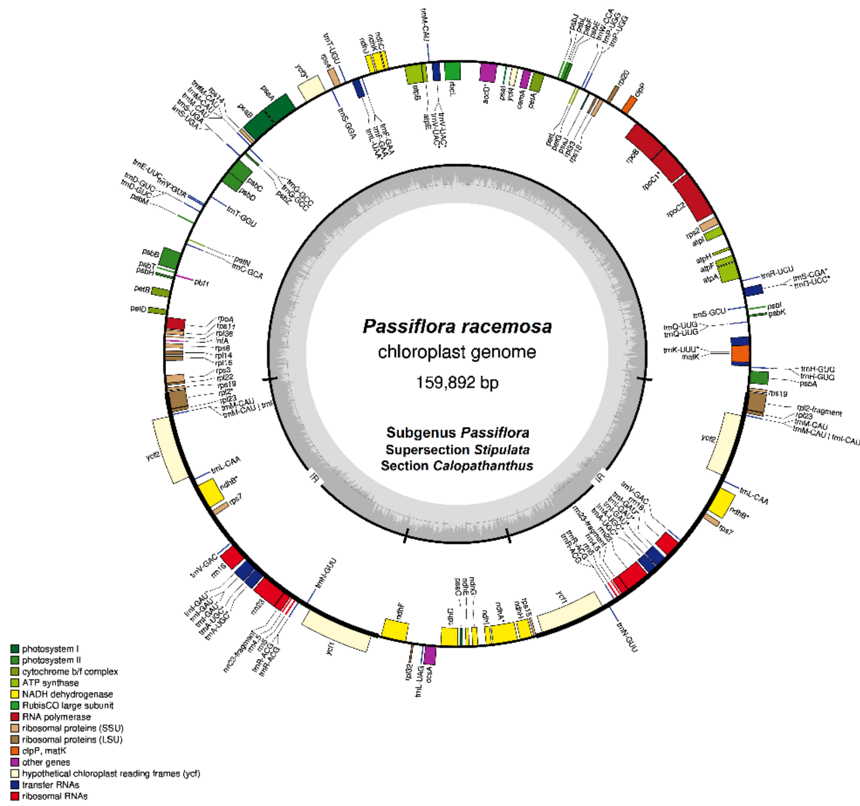
Passiflora chaparensis
chloroplast genome
150,928 bp
Subgenus Passiflora
Supersection Laurifolia
Series Laurifoliae
LSC
SSC
IRA
IRB
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

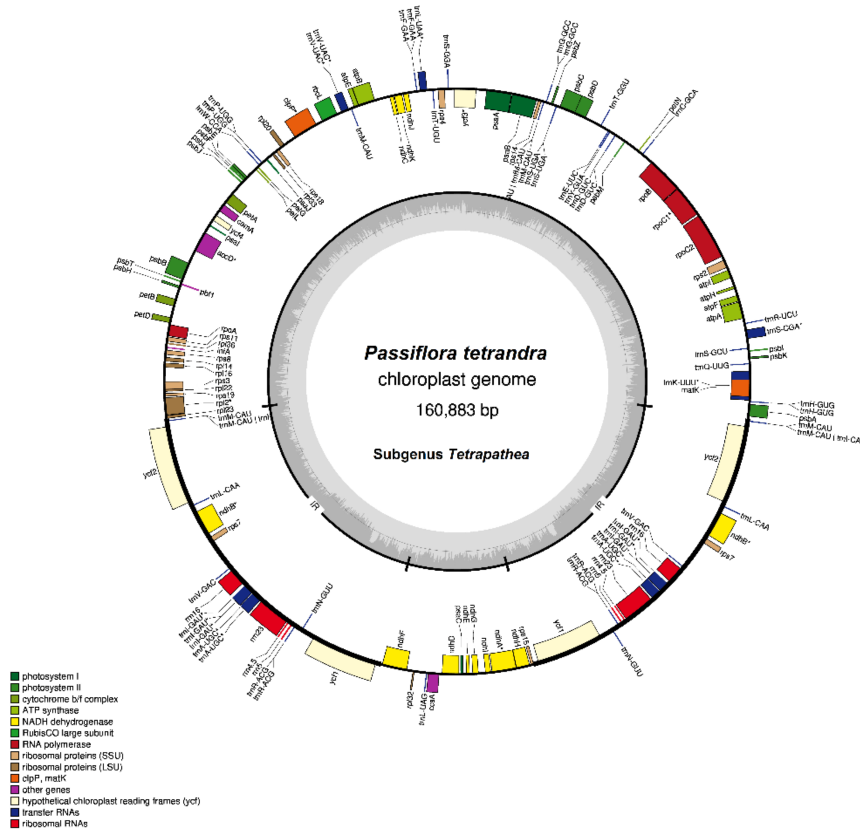

Passiflora garckei
chloroplast genome
146,299 bp
Subgenus *Passiflora*
Supersection *Stipulata*
Section *Granadillastrum*
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs


Passiflora palenquensis
chloroplast genome
149,524 bp
Subgenus *Passiflora*
Supersection *Laurifolia*
Series *Tiliifolia*
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

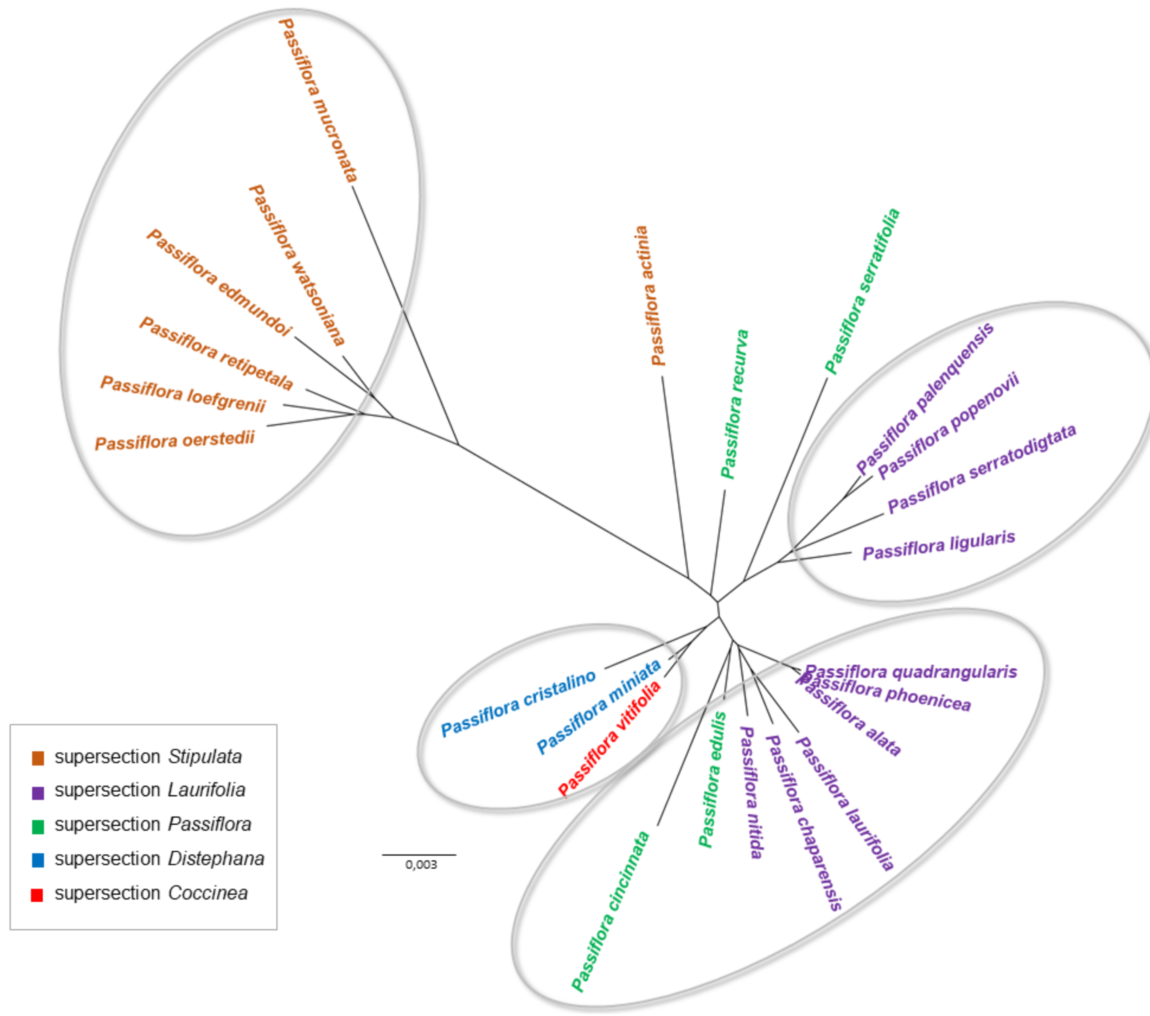

Passiflora phoenicea
chloroplast genome
148,089 bp
Subgenus Passiflora
Supersection Laurifolia
Series Quadrangulares
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs


Passiflora popenovii
chloroplast genome
147,278 bp
Subgenus Passiflora
Supersection Laurifolia
Series Laurifoliae
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs


Passiflora racemosa
chloroplast genome
159,892 bp
Subgenus Passiflora
Supersection Stipulata
Section Calopathanthus
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs


Passiflora tetrandra
chloroplast genome
160,883 bp
Subgenus Tetrapathea
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

APPENDIX L. Phylogenetic tree obtained using Bayesian inference and whole chloroplast genome sequences from the subgenus *Passiflora*. The colors of the taxons indicate the supersections' classification.



APPENDIX M. Phylogenetic tree of *Passiflora* genus obtained with Maximum Likelihood analysis. A) Phylogenetic tree with the complete sequences of nuclear 18S/26S genes; B) Phylogenetic tree with 68 chloroplast protein-coding genes. * indicates BS < 50.