Copyright

by

Bikash Shrestha

2020

The Dissertation Committee for Bikash Shrestha certifies that this is the approved version of the following Dissertation:

Plastid genome evolution and inheritance in *Passiflora*

Committee:

Robert K. Jansen, Supervisor

Lawrence E. Gilbert

David M. Hillis

Craig R. Linder

David H. Herrin

Plastid genome evolution and inheritance in *Passiflora*

by

Bikash Shrestha

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2020

Acknowledgements

Throughout the Ph. D. program I have received tremendous support from many people without which I don't believe this work would be possible.

First of all, I would like to thank my supervisor Dr. Robert Jansen for always being available to help and most importantly, providing considerate guidance and direction in completing the project.

I am very grateful to Dr. Lawrence Gilbert for allowing access to invaluable living *Passiflora* collections and sharing indispensable knowledge about the genus. I am also thankful to my committee members, Dr. David Hillis, Dr. David Herrin and Dr. Craig Linder, for their critical comments and helpful advice in making the project better.

Further, I would like to thank Dr. Mao-Lun Weng and Dr. Jin Zhang for their effort and patience on teaching fundamentals of computational biology during the early stage of the graduate school.

I would also like to express my sincere gratitude to Texas Ecolab for research funds and Graduate School for providing Continuing Fellowship that gave me an opportunity to work independently, which allowed to make substantial progress in my research.

Finally, thanks to my wife Lauren for waiting patiently for me to graduate.

Abstract

Plastid genome evolution and inheritance in *Passiflora*

Bikash Shrestha, PhD

The University of Texas at Austin, 2020

Supervisor: Robert K. Jansen

Plastid genomes (plastomes) of photosynthetic angiosperms are for the most part highly conserved in their organization, mode of inheritance and rates of nucleotide substitution. A small number of distantly related lineages including *Passiflora* share a syndrome of features that deviate from this general pattern, including extensive genomic rearrangements, accelerated rates of nucleotide substitution, biparental inheritance and plastome-genome incompatibility. Plastome evolution studies in *Passiflora* are limited in taxon sampling; hence the phylogenetic extent of the rearrangements is unknown. To gain better understanding in plastome evolution in *Passiflora,* plastomes from 31 taxa and transcriptomes from 6 species were sequenced and assembled. In addition, interspecific crosses within two largest subgenera, *Passiflora* and *Decaloba,* were greatly expanded to understand mode of plastid inheritance in the genus.

Phylogenomic analyses with 68 protein-coding genes generated a fully resolved, strongly supported tree that is congruent with the comprehensive phylogenies based on a few plastid and nuclear loci. Extensive rearrangements were detected including several gene/intron losses, inverted repeat expansion/contraction and inversions, some of which

occurred in parallel. Nucleotide substitution rate analyses of 68 protein-coding genes across the genus showed lineage- and locus-specific acceleration.

Comparative transcriptome analyses identified missing or divergent plastid genes in *Passiflora* that have followed three distinct evolutionary paths: transfer to the nucleus, substitution by the nuclear genes and highly divergent gene that likely remain functional. Plastid-encoded *rps7* was transferred into the intron of a nuclear-encoded plastid-targeted thioredoxin m-type gene, acquiring its plastid transit peptide. Plastid *rpl20* likely experienced a novel substitution by a duplicated, nuclear-encoded mitochondrial-targeted *rpl20* that has a similar gene structure.

Interspecific hybrids in *Passiflora* exhibit diverse modes of plastid inheritance including a clade-specific paternal or maternal pattern along with frequent transmission of biparental plastids. Furthermore, heteroplasmy due to biparental inheritance was restricted to early developmental stage in hybrids and plastid types from either parent were excluded in older plants resulting plastid homogeneity.

These results of unusual plastome dynamics and inheritance identified in *Passiflora* presents the genus as an exciting system to study plastome evolution in angiosperms.

Table of Contents

List of Tables

List of Figures

Chapter One

Plastid genome evolution in *Passiflora*[1]

1.1 Introduction

Plastids are endosymbiotic organelles that contain their own genomes (plastome), which are highly conserved in gene content and structure in photosynthetic angiosperms (Ruhlman and Jansen 2014). Conserved plastome architecture includes a quadripartite structure with two copies of an inverted repeat (IR) that separate large and small single copy regions (LSC and SSC, respectively). The plastome typically contains about 80 protein-coding genes primarily involved in photosynthesis and housekeeping along with 30 tRNA and 4 rRNA genes (Bock, 2007). With the exponential increase in number of published plastome sequences in recent years, a small number of lineages among photosynthetic angiosperms have been identified that depart from this conserved organization and contain highly rearranged plastomes with structural changes, gene and intron losses and rate heterogeneity in protein coding genes (Wicke et al. 2011; Ruhlman and Jansen 2014).

[1] This chapter contains two published manuscripts: (i) Rabah SO, Shrestha B, Hajrah NH, Sabir Mumdooh J, Alharby HF, Sabir Mernan J, Alhebshi AM, Sabir JSM, Gilbert LE, Ruhlman TA, Jansen RK. 2019. *Passiflora* plastome sequencing reveals widespread genomic rearrangements. J. Sys. Evol. 57:1–14. doi: 10.1111/jse.12425 and (ii) Shrestha B, Weng M-L, Theriot EC, Gilbert LE, Ruhlman TA, Krosnick SE, Jansen RK. 2019. Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. Mol. Phylogenet. Evol. 138:53–64. doi: 10.1016/j.ympev.2019.05.030. Bikash Shrestha performed all the experiments, conducted data analyses and wrote the manuscripts.

Structural rearrangements in plastomes are usually caused by inversions and expansion and/or contraction of the IR. Both of these processes shuffle gene order resulting in reduced synteny, and IR expansion/contraction or IR loss has contributed to substantial variation in plastome size (Ruhlman and Jansen 2014). Although large inversions in plastomes have been considered rare and reliable phylogenetic characters (Raubeson and Jansen, 2005), multiple inversions in several lineages including Ranunculaceae (Hoot and Palmer 1994), Campanulaceae (Cosner et al. 2004), Fabaceae (Schwarz et al. 2015) and Geraniaceae (Weng et al. 2014), some of which occurred in parallel, suggest that these changes have limited phylogenetic utility. Major IR expansion is less common but has been documented in several unrelated lineages, such as *Pelargonium* (Chumley et al. 2006; Weng et al. 2017), *Berberis* (Kim and Jansen 1994; Ma et al. 2013), *Trochodendraceae* (Sun et al. 2013), *Plantago* (Zhu et al. 2016) and *Annona* (Blazier et al. 2016a). Although minor IR contractions are common in angiosperms (Goulding et al. 1996; Downie and Jansen 2015), a major contraction excluding a portion of the ribosomal operon from the IR has been documented in *Monsonia* (Guisinger et al. 2011). Similarly, IR loss has been reported in several angiosperm families, including *Fabaceae* (Palmer and Thompson 1981; Downie and Palmer 1992), three genera of Geraniaceae (Guisinger et al. 2011; Blazier et al. 2011, 2016b; Ruhlman et al. 2017), Cactaceae (Sanderson et al. 2015) and Arecaceae (Barrett et al. 2016).

Gene and intron losses and high sequence divergence in protein coding genes are common features in highly rearranged plastomes in angiosperms (Jansen et al. 2007). Common and re-occurring gene loss in angiosperms includes *accD*, encoding the beta

subunit of acetyl-CoA carboxylase complex, *ycf1* and *ycf2*, and several genes that encode ribosomal subunits. Gene losses in plastids are often associated with functional transfer to the nucleus, which has been documented for several genes, such as *accD* in legumes (Magee et al. 2010; Sabir et al. 2014), *rpl32* in Salicaceae (Cusack and Wolfe 2007; Ueda et al. 2007) and *Thalictrum* in Ranunculaceae (Park et al. 2017), and multiple independent transfers of *infA* and *rpl22* in rosids (Millen et al. 2001; Gantt et al. 1991; Jansen et al. 2011). These findings support the hypothesis that endosymbiotic gene transfer of plastid genes to nucleus is a frequent and ongoing process (Timmis et al. 2004). Alternatively, plastid gene loss has been compensated by functional replacement with nuclear genes, such as *rps16* in *Medicago* and some Salicaceae (Ueda et al. 2008), *rpl23* in *Spinacia* (Bubunenko et al. 1994) and *Geranium* (Weng et al. 2016), and *accD* in Poaceae (Konishi et al. 1996) and Geraniaceae (Park et al. 2017). Recent studies have shown unprecedented sequence divergence of two additional genes, *rpoA* (RNA polymerase subunit alpha) and *clpP* (a subunit of ATP-dependent caseinolytic protease) in Geraniaceae (Guisinger et al. 2008; Blazier et al. 2016b) and *clpP* in *Silene* (Erixon and Oxelman 2008; Sloan et al. 2012).

Genomic rearrangements in plastids often correlate with lineage- and/or locus-specific acceleration in nucleotide substitution rates in protein coding genes (Jansen et al. 2007; Weng et al. 2014; Schwarz et al. 2017). Therefore, characterization of substitution rates provides essential insights into the dynamics of plastome evolution. In angiosperms, rates of nucleotide substitution in plastid-encoded protein coding genes are relatively slow and consistent (Wolfe et al. 1987). Comparisons of nucleotide substitution rates among

three cellular compartments showed that the rate in plastids is three times higher than mitochondria but about 2- to 6-fold slower than that of nuclear genomes (Wolfe et al. 1987; Drouin et al. 2008). Within the plastome, IR regions have about 3-fold reduced synonymous substitution rate compared to single copy regions (Wolfe et al. 1987; Perry and Wolfe 2002; Zhu et al. 2016; Schwarz et al. 2017), potentially due to the two-fold available copies for homologous recombination (HR) and gene conversion (Birky and Walsh 1992). However, IR genes in *Pelargonium* are an exception mainly due to accelerated substitution rate in ribosomal subunit (RS) genes, which were incorporated into the IR (Weng et al. 2017). Locus-specific acceleration in RS and plastid-encoded RNA polymerase (PEP) genes has been documented in Geraniaceae (Guisinger et al. 2008; Zhang et al. 2014; Weng et al. 2012). Locus-specific rate acceleration has also been documented in RS genes, *clpP, ycf1* and *ycf2* in the unrelated lineage *Silene* (Erixon and Oxelman 2008; Sloan et al. 2012).

Passiflora is included in a small group of angiosperm taxa with highly rearranged plastomes that exhibit a syndrome of features, including biparental inheritance and plastome-genome incompatibility (Jansen et al. 2007; Greiner et al. 2011). It is the largest genus within Passifloraceae, and includes more than 560 species grouped into the five subgenera *Passiflora, Decaloba, Astrophea, Deidamioides* and *Tetrapathea* (Feuillet and MacDougal 2003; Krosnick et al. 2009; Krosnick et al. 2013). Subgenera *Passiflora* and *Decaloba* are the two largest groups with more than 200 species of vines mainly distributed in Central and South America. The first evidence that *Passiflora* had gene order changes including 10 inversions was an unpublished draft plastome of *P. biflora* (subgenus

Decaloba) (Jansen et al. 2007). The first completed plastome in Passifloraceae was *Passiflora edulis* (subgenus *Passiflor*a), which was found to contain several pseudogenes (*rps7, rpl20, rpl22, accD, ycf1* and *ycf2*) along with three inversions (Cauz-Santos et al. 2017). The remarkable differences in genomic changes in *Passiflora* plastomes based on two species suggests there may be much more variation in the genus. To gain a better understanding of plastome evolution in *Passiflora,* a broader sampling incorporating species across all five subgenera is needed. In this study, we completed *Passiflora* plastome sequences for 29 species from all five subgenera including one species with two accession from different location and a species from the sister genus *Adenia.* With these 31 newly completed plastomes, this study focuses on four questions: i) Does the taxon sampling of plastomes resolve the *Passiflora* phylogeny? ii) Which protein coding genes provide the most phylogenetic signal? iii) What is the extent of genomic rearrangement and its phylogenetic distribution? iv) What are the patterns of nucleotide substitutions in protein coding genes?

1.2 Materials and Methods

Plant materials

Taxon sampling included 13 species from subgenus *Passiflora,* 11 species from subgenus *Decaloba* including a species with two accession from different location, three species from subgenus *Deidamioides,* a species from subgenus *Astrophea,* a silica dried specimen from subgenus *Tetrapathea* and an outgroup *Adenia mannii*. All samples except the species from subgenus *Tetrapathea* were obtained from field-collected populations

grown in Lawrence Gilbert's greenhouses at The University of Texas at Austin. The list of species with their accession numbers, voucher information and original collection location is provided in Table 1.1.

DNA extraction

Newly emerged leaves were collected, flash frozen in liquid nitrogen and stored at -80^{o} C until DNA was isolated. Total genomic DNA was extracted using the method of Doyle and Doyle (1987) with modifications, including the addition of 2% PVP and 2% betamercaptoethanol (Sigma, St. Louis, MO, USA) to the extraction buffer. Organic phase separation using chloroform-isoamyl alcohol was repeated until the aqueous fraction was clear. DNA pellets were resuspended in ~200 μL DNase-free water. After treatment with RNase A (ThermoScientific, Lafayette, CO) samples were re-subjected to phase separation with chloroform-isoamyl alcohol. DNA was recovered by ethanol precipitation, resuspended in DNase-free water and stored at -20^{o}C.

Genome sequencing, assembly and annotation

Passiflora biflora and *P. quadrangularis* plastomes were completed by isolating plastid DNA, shearing DNA, shotgun cloning into plasmids and Sanger sequencing as described in Raubeson et al. (2007). For the remaining samples, library preparation from total genomic DNA and DNA sequencing were carried out at the UT-Austin Genome Sequencing and Analysis Facility on Illumina HiSeq 2500/4000 platforms (Illumina, San Diego, CA) and HiSeq X Ten platform at Beijing Genome Institute (BGI). *De novo*

assembly was performed with Illumina paired-end reads using Velvet v. 1.2.07 (Zerbino and Birney 2008) at the Texas Advanced Computing Center (TACC) using the same approach succesfully utilized across wide diversity of angiosperms (Weng et al. 2014, 2017; Sabir et al. 2014; Schwarz et al. 2015; Rabah et al. 2017). This method involves performing multiple assemblies using different k-mer sizes with coverage cutoffs of 200X, 500X and 1000X to exclude nuclear and mitochondrial contigs. Initial assembly includes five different k-mer sizes ranging from 85 to 93 inclusively, with an increment of 2. Assembled contigs with 1000X coverage for the five kmer parameters were imported into Geneious v. 6.1.8/ Geneious v. 11.0.5 (https://www.geneious.com) and joined using Geneious *de novo* assembly with default settings. To identify putative plastid contigs, the plastome of *Populus trichocarpa,* which is in same order as Passifloraceae, Malpighiales, was used as a reference in Geneious. For three species (*P. nitida, P. pittieri,* and *P. retipetala*) in which the initial k-mer size range was not sufficient to complete the plastome assembly, k-mer sizes up to 119 and coverage over 500X were utilized to complete the assembly in Geneious. Any potential conflicts and misassembles in the genome assemblies, such as single nucleotide polymorphism and imprecise number of nucleotides in repeats, were verified by mapping all Illumina paired-end reads against the contigs using Bowtie2 (Langmead and Salzberg 2012).

Completed genomes were annotated using Dual Organellar GenoMe Annotator (DOGMA; Wyman et al., 2004) and tRNAs were annotated with tRNAscan-SE (Lowe and Eddy, 1997; http://lowelab.ucsc.edu/tRNAscan-SE/). Criteria used to assign protein-coding genes as putatively functional were: i) presence of an open-reading frame (ORF) with a

complete conserved domain as displayed in conserved domain database (CDD, www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), and ii) absence of internal stop codons or if any were present then the sequence should be conserved within the ORF and retain the conserved domain.

PCR validation of IR boundaries

Major IR contractions in *P. menispermifolia* and *P. obovata* and an unusual IR boundary in *P. contracta* were verified using PCR and Sanger sequencing. Primers for the PCR were designed using Primer3 (Untergasser et al. 2012) in Geneious v. 11.0.5 (https://www.geneious.com). Table 1.2 provides primer sequences, target sites and amplification product length.

Plastome rearrangement analysis

Whole genome alignments for *Passiflora, Adenia* and the reference *Populus trichocarpa* were performed using progressiveMauve 2.3.1 (Darling et al. 2010) in Geneious v. 11.0.5 to identify locally collinear blocks (LCBs) shared among species. A copy of inverted repeat (inverted repeat A, IRA) was removed prior the alignment. LCBs generated via progressiveMauve alignment were numbered and strand orientation was assigned (Table 1.3A, B). CREx (common interval rearrangement explore, Bernt et al., 2007) was used to calculate breakpoint (BP) and reversal rearrangement distances (Table 1.3C).

Phylogenetic analysis

Taxon sampling for phylogenetic analysis includes 30 species included in this study along with *Passiflora cincinnata* (NC_037690) available in the NCBI and an unpublished draft plastome of *Passiflora cirrhifolia* for total of 32 *Passiflora* species distributed across five subgenera. *Populus trichocarpa* (NC_009143) and *Adenia mannii* were used as outgroups.

Sixty-eight protein-coding plastid genes (Table 1.4A) shared by all *Passiflora* plastomes and the outgroups *Adenia mannii* and *Populus trichocarpa* were aligned individually using MAFFT (Katoh and Standley,2013) prior to concatenation into single multiple sequence alignment in Geneious. IQ-TREE v. 1.5.2 (Nguyen et al. 2015) was implemented to determine best-fit partition schemes and evolutionary model selection (Table 1.4B). Maximum likelihood (ML) analysis was performed in IQ-TREE to construct phylogenetic trees and branch support was assessed using ultrafast bootstrap and non-parametric bootstrap from 2000 and 100 pseudoreplicates, respectively. The ML tree with bootstrap support values was visualized using FigTree v. 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Commands used in IQ-TREE for phylogenetic analysis are summarized in Table 1.4C.

Phylogenetic informativeness of plastid genes

PhyDesign (http://phydesign.townsend.yale.edu/) was used to estimate the phylogenetic informativeness profile for 68 protein-coding plastid genes using a relative-time ultrametric tree (Lopez-Giraldez and Townsend 2010). The ML tree inferred with the

68 concatenated plastid genes in IQ-TREE was used as an input tree to reconstruct a relative-time ultrametric tree in the *dnamlk* program in PHYLIP (Felsentein 1989). The converted relative-time ultrametric tree and 68 protein-coding plastid genes dataset partitioned by genes were used as input files in PhyDesign to calculate phylogenetic informativeness using the default settings.

Nucleotide substitution rate analyses

Pairwise estimation: Nucleotide sequences of protein coding genes used for the phylogenetic analysis were extracted and translational alignment was carried out for individual genes using MAFFT in Geneious v. 11.0.5. For each gene alignment, pairwise synonymous (*dS*) and non-synonymous (*dN*) substitution rates were calculated between the outgroup *Populus trichocarpa* and all other species using PAML v.4.8 (Yang 2007). The codon frequencies were determined using F3 x 4 model and transitions/transversions were estimated with default settings of initial values 2 and 0.4, respectively. Other parameters in the CODEML control file included cleandata = 0 for treating alignment gap as ambiguous characters, model =0 for a single *dN/dS* value across all branches and runmode = -2 for pairwise comparisons of *dS* and *dN*. Linear mixed-effects models were implemented to estimate statistically significant differences in *dS*, *dN* and *dN/d*S across genes and clades using lme4 package (Bates et al. 2015) in R v3.5.1 (R Core Team 2013). Emmeans package in R v3.5.1 was used for *post-hoc* analysis to compute contrasts for fixed-effect variables in the linear mixed-effects model. The median *dS*, *dN*, and *dN/dS* for each species was visualized using ggplot2 package (Wickham 2016) in R v3.5.1.

Lineage-specific rate analysis: To test whether substitution rates (*dN* and *dS*) and *dN/dS* were significantly different on the branch leading to IR expansion within subgenus *Decaloba*, branch models were used in HyPhy v2.2.4 (Pond et al. 2005) and PAML v.4.8 (Yang, 2007). In HyPhy, the ML tree generated using IQ-TREE (Nguyen et al. 2015) was used as a constraint tree with codon substitution model MG94xHKY85_3x4 to estimate likelihood for two models, global model with single *dN* and *dS*, and branch specific model that allows substitution rates to vary on a specific branch. Likelihood ratio tests (LRTs) were performed in HyPhy (Pond et al. 2005) to detect significant differences in branch specific substitution rates. For the branch model, codon frequencies were determined using F3 x 4 model and transition/transversion and *dN/dS* were estimated setting default initial values of 2 and 0.4, respectively, in CODEML control file using PAML v.4.8 (Yang 2007). The ML tree generated via IQ-TREE (Nguyen et al., 2015) was used as a constraint tree. Two branch models, global-ratio model (model =0) with one *dN/dS* value specified across the entire tree and two-ratio alternative model (model =2) where the branch leading to IR expansion (and all internal branches within it) had one *dN/d*S value and rest of the tree had a different *dN/dS* value. Likelihood ratio tests (LRTs) were conducted to evaluate the model fit. False discovery rate correction was used in R v3.5.1 (R Core Team 2013) to correct for multiple comparisons in estimating significant differences in *dN*, *dS* and *dN/dS*.

1.3 Results

Phylogenetic relationships

The nucleotide alignment of 68 plastid protein-coding genes shared by 34 taxa was 53,065 bp and the optimal phylogenetic tree had a likelihood score of ln(L) = -76476.064 (Figure 1.1A). All nodes except eight had bootstrap support (BS) values of 100%. With *Populus trichocarpa* as the root, *Adenia* was strongly supported as sister to *Passiflora.* Subgenus *Passiflora* was monophyletic whereas subgenus *Deidamioides* was polyphyletic with *P. arbelaezii* sister to subgenus *Astrophea* (BS = 100%)*, P. contracta* and *P. cirrhifolia* sister to subgenera *Tetrapathea* + *Decaloba* (BS = 93%) and *P. obovata* nested within subgenus *Decaloba* (BS =100%)*.* A single species (*P. tetrandra*) from the Old World subgenus *Tetrapathea* was strongly supported (BS=94%) as sister to subgenus *Decaloba.* Within subgenus *Decaloba, P. microstipula* was the earliest diverging lineage and *P. obovata* (subgenus *Deidamioides*) was strongly supported as sister to rest of *Decaloba* (BS = 100%)*.* All internal nodes within *Decaloba* were strongly supported (BS ≥ 98%) and branch lengths in this subgenus were longer compared to other subgenera.

Phylogenetic informativeness

The per-site and net phylogenetic informativeness for 68 protein-coding genes used in phylogenetic analysis were measured using PhyDesign (Figure 1.4, Table S5). A combination of two factors contributed to overall informativeness, frequency of rapidly evolving sites in a gene and gene length. The gene *ycf4* had the highest per-site informativeness, however, *rpoC2* superseded all other protein coding genes in net informativeness because of the considerably longer length of plastid-encoded RNA polymerase (PEP) genes. Likewise, per-site informativeness of RS genes was similar to

PEP genes but their net informativeness was considerably lower than the longer PEP genes. The net informativeness for *clpP* elevated quickly in recent epochs but declined sharply in deeper epochs with the largest standard deviation of informativeness (267) reducing overall informativeness and resolution (Figure 1.4). The slowly evolving genes at the lower end of phylogenetic informativeness were primarily associated with photosynthesis and were shorter in length (~100 bp) (Table 1.5).

Plastome organization

The phylogenetic distribution of all genomic changes in plastomes of *Passiflora* species and *Adenia mannii* were plotted on the ML tree (Figure 1.1A, B, Table 1.6). Among the 31 species of Passifloraceae*, P. arbelaezii* had the largest plastome (170,568 kb) and *P. menispermifolia* had the smallest (133,682 kb). The overall difference of ~37 kb in *Passiflora* plastome size reflected the substantial variation in all three regions: LSC (55 kb – 88 kb), SSC (12 kb – 29 kb) and IR (10 kb – 47 kb). Average GC content in protein coding genes and intergenic regions for *Passiflora* was 38% and 32%, respectively. Gene density (total number of genes per kb) was slightly higher for subgenus *Decaloba* (0.82-0.95) compared to other subgenera (Table 1.6). Overall variation in *Passiflora* plastome size was primarily due to IR expansion or contraction and gene and intron losses.

Gene and intron losses: Based on the criteria used to define presence and absence of a gene (see methods), the distribution of gene and intron losses was plotted on the ML tree (Figure 1.1A). The total number of protein coding genes varied from 69 to 75 in *Passiflora* (Table 1.6). All species in Passifloraceae and *Populus trichocarpa* were missing *rps16*. In

addition, all species in Passifloraceae shared the loss of *rpl22* and the *atpF* intron (Figure 1.1A). All *Passiflora* species except two from subgenus *Deidamioides* (*P. arbelaezii* and *P. cirrhifolia*) and *P. tetrandra* were missing *rpl20.* The clade subgenus *Decaloba* + *P. obovata* shared the loss of *rps7*. The large ribosomal subunit *rpl32* was missing in several lineages, including *P. pittieri,* two species (*P. contracta, P. obovata)* in subgenus *Deidamioides* and seven species (*P. auriculata, P. jatunsachensis, P. rufa, P. filipes, P. misera, P. affinis*, and *P. biflora*) in subgenus *Decaloba* and the outgroup *Populus*. The two largest plastid genes, *ycf1* and *ycf2,* were missing in all species in subgenera *Decaloba* (except *P. microstipula*) and *Passiflora* (except *P. foetida*). Compared to *Populus trichocarpa,* huge variation in pairwise nucleotide sequence identity was detected in the alpha (α) subunit of PEP gene *rpoA* in *Passiflora,* especially in subgenus *Decaloba* (Table 1.7). Pairwise nucleotide identity for *rpoA* between *Populus*, *Adenia* and all the species in *Passiflora* except subgenus *Decaloba* was > 88%. Species in subgenus *Decaloba* had the most divergent *rpoA* sequences, 93.1% in *P. microstipula*, ~72-74% in *P. rufa, P. jatunsachensis* and *P. auriculata*, ~53% in *P. affinis, P. bifora, P. misera, P. lutea* and *P. filipes* and 39% in *P. tenuiloba* and *P. suberosa* (Table 1.7). Pairwise amino acid identity was lowest at ~24% for *P. tenuiloba* and *P. suberosa* compared to *Populus.* A search of the conserved domain database (CDD) for *rpoA* did not predict the functional domains in *P. tenuiloba* and *P. suberosa,* therefore, *rpoA* was considered putatively missing in these two species (Table 1.7).

Both introns in *clpP* were missing in all species in subgenera *Passiflora* and *Decaloba* except *P. microstipula.* Additionally, all species in *Decaloba* except *P.*

microstipula were missing the *rpoC1* intron and the *rpl16* intron was lost in two subgenus *Decaloba* species, *P. lutea* and *P. filipes.*

Inversions: Whole genome alignment of *Passiflora* species with *Populus* as a reference identified 16 locally collinear gene blocks (LCB) (Figure 1.5, Table 1.3A, B). Each LCB was numbered and the distribution of inverted blocks was plotted on the ML tree, allowing for multiple inversions (Figures 1.1B, 1.5). Except for two LCBs, LCB 2 (*trnK-UUU-psbI*) and LCB 15 (*rpoA-ycf1*), all LCBs were inverted in one or more species. All *Passiflora* species and *A. mannii* shared the inversion of LCB 1 (*psbA - trnH-GUG*) indicating an early inversion within the Passifloraceae. *Adenia mannii* had four additional contiguous LCBs (blocks 3-6) inverted, which includes genes from *trnS-GCU - ycf3.* Among *Passiflora* species, *P. foetida* had fewest numbers of inverted LCBs (1, 8-9), and the lowest breakpoint and reversal distances of 4 and 2, respectively. *Passiflora misera* had the largest number of inverted LCBs (1, 3-6, 9-14) and the highest breakpoint and reversal distances of 9 and 6, respectively (Table 1.3C).

IR expansion and contraction: Most *Passiflora* species have experienced minor to major IR expansion or contraction (Figure 1.1C). IR expansion and contraction in *Passiflora* led to substantial variation in IR size (~ 37 kb) ranging from ~10 kb in *P. menispermifolia* to ~47 kb in *P. auriculata* (Table 1.6). The phylogenetic distribution of IR expansion and contraction events on the ML tree revealed that the substantial IR expansion was confined mostly to subgenus *Decaloba,* whereas major IR contraction occurred in two species, *P. menispermifolia* and *P. obovata,* from subgenera *Passiflora* and *Deidamioides,* respectively (Figure 1.1A, C).

The IR of *P. menispermifolia* was substantially reduced by contraction at the IR/SSC boundary resulting in the exclusion all rRNA genes; only four protein coding genes and two tRNAs were retained. A nearly identical contraction in the IR/SSC boundary occurred independently in *P. obovata* resulting an IR similar to *P. menispermifolia.* A major difference in the IR contraction between *P. menispermifolia* and *P. obovata* was the boundary that shifted during the contraction. In *P. obovata,* the boundary at the junction of SSC and IR_B (J_{SB}) was contracted, whereas, in *P. menispermifolia* the boundary at the junction of SSC and IR_A (J_{SA}) was contracted, as a result the SSC genes between these two species appears to be inverted (Table 1.3B). Furthermore, due to pseudogenization of *ycf2* in *P. menispermifolia,* the IR size was much smaller (~10 kb) compared to *P. obovata* (~18 kb). IR contraction in other *Passiflora* species includes two minor contractions at IR/LSC boundary excluding *rps19* and *rpl2* in *P. pittieri* and *rps19, rpl2,* and *rpl23* in *P. microstipula.*

Although major IR expansion in *Passiflora* was restricted to subgenus *Decaloba*, a few instances of minor IR expansion occurred across the genus (Figure 1.1A, C), including the incorporation of *rps15* into the IRs of *P. pittieri, P. arbelaezii* and *P. affinis* and partial copy of *ndhF* in *P. auriculata, P. jatunsachensis, P. rufa, P. suberosa, P. lutea* and *P. filipes*. Except for *P. microstipula,* substantial expansion of IR was mainly at LSC and IR_B (J_{LB}) boundary in all members of subgenus *Decaloba.* Variation in IR expansion in subgenus *Decaloba* resulted in six different J_{LB} boundaries incorporating from seven to twenty-five LSC genes (Figure 1.1C).

Nucleotide substitution rates

Pairwise estimation: Synonymous (*dS*) and nonsynonymous (*dN*) substitution rates in pairwise comparisons between the reference *Populus trichocarpa* and all Passifloraceae were performed for the 68 plastid protein-coding genes used to infer phylogeny (Table 1.8). Variation in *dS*, *dN* and *dN*/*dS* was compared among five clades in *Passiflora* and *Adenia* (Figure 1.3) and significant differences in the substitution rates and *dN/d*S among the clades were calculated (Tables 1.9-1.11).

Median *dS* and *dN* for the majority of plastid genes were consistently low except for species in clade E (Figure 1.3). Multiple comparisons of mean *dS* across the clades showed that several genes including, *atpE, atpH, cemA, ndhF, ndhG, ndhH, ndhI, petD, psbH, psbI, psbM, rbcL, rpl16, rpl23, rps11, rps15, rps2, rps3, rps4, rps8,* and *ycf4*, had significantly elevated *dS* for clade E species (Table 1.9). In clade A, *dS* was elevated for *psaI* in *P. arbelaezii* (0.85) relative to the rest of *Passiflora* (0.14) (Tables 1.8-1.9). Multiple comparisons of means showed that clade E also had elevated *dN* compared to other clades for PEP (*rpoB, ropC1* and *rpoC2*) and RS (*rpl2, rpl14, rpl36, rps2, rps3, rps4, rps11, rps12* and *rps19*) genes (Table 1.10). Clade A and two species in clade E (*P. lutea* and *P. filipes*) had significantly higher *dN* (0.5 and 0.6, respectively) for *ycf4* compared to the other clades (<0.1) (Tables 1.8, 1.10). The average *dS* and *dN* for *clpP* were elevated for all Passifloracae except *dS* for clade C (Table 1.8). Graphical overviews of pairwise *dS* and *dN* estimated for each individual gene in the rate analysis for all *Passiflora* and *Adenia* species are presented in Figures 1.6 and 1.7.

Mean *dN*/*dS* for most plastid genes were consistently low except for *clpP, rpl23, rpl36* and *ycf4* where the ratio was above 1 (Table 1.8). Compared to other plastid genes, the average *dN*/*dS* for *rpl23* (1.7) was significantly higher in clade B and *P. microstipula* of clade E (Tables 1.8, 1.11). The elevated *dN/dS* in subgenus *Passiflora* + *P. microstipula* was due to very low *dS* values compared to other species of *Passiflora.* There was a slight elevation of *dN/dS* for *rpl23* in *A. mannii* (0.99) and *P. arbelaezii* (1.09) (Table 1.8). The average *dN/dS* for *clpP* was elevated in all clades with the highest value in clade C (1.4) followed by clade E (0.96). The *dN/dS* for *ycf4* was significantly elevated in clade A (1.2) compared to rest of *Passiflora*, and two species in clade E, *P. filipes* and *P. lutea*, also had elevated *dN/dS* for *ycf4* at 1.07 and 1.02, respectively (Table 1.8). All four species with elevated *dN/dS* for *ycf4* also had significantly increased *dS* and *dN*. In addition, *P. arbelaezii* and *P. pittieri* had significantly accelerated rates for *psaI, dS* = 0.85 and *dN* = 0.11 for *P. arbelaezii* and *dS* = 0.31 and *dN* = 0.08 for *P. pittieri* compared to rest of *Passiflora* (*dS* < 0.24 and *dN* < 0.037) (Table 1.8). Similarly, three species in clade E, *P. affinis, P. biflora* and *P. misera* had higher *dN*/*dS* for *rpl36* due to elevated *dN*. Table 1.11 provides list of genes with *dN/dS* significantly different among the clades. A graphical overview of *dN*/*dS* values estimated for each individual gene in the rate analysis for all *Passiflora* and *Adenia* species is presented in Figure 1.8.

Lineage-specific rate analysis: We tested branch specific substitution rates and *dN*/*dS* for the branch leading to IR expansion within subgenus *Decaloba* (Figure 1.2). The likelihood ratio test (LRT) detected decreased *dS* but elevated *dN* for several plastid genes on this branch (Table 1.12). Genes with decreased *dS* were primarily associated with

photosystems (*psaA, psaB, psaJ, psbB, psbD, psbZ,)* and the NAD(P)H dehydrogenase complex (*ndhD, ndhE*). Decreased rates were also detected in *dS* for *clpP* and *dN* in *ccsA*. Elevated branch-specific *dN* was primarily restricted to the RS genes *rpl2, rpl23, rps2, rps14,* and *rps15,* plastid-encoded RNA polymerase genes *rpoB, rpoC1* and *rpoC2*, and a few other miscellaneous genes, including *clpP, petD, ycf3* and *ycf4.* The small RS gene *rps11* was elevated in both *dS* and *dN*. The branch specific elevation of *dN* in PEP and RS genes within subgenus *Decaloba* was consistent with increase in mean *dN* for PEP and RS genes in pairwise comparisons (Figure 1.7, Table 1.10).

Substitution rates for the single copy genes incorporated into IR due to IR expansion in subgenus *Decaloba* did not necessarily result in a decreased substitution rate. In the case of ribosomal subunit genes, rates were elevated regardless of the location (Figure 1.9). The LRTs comparing branch-ratio model against global-ratio model detected significantly elevated *dN*/*dS* for the branch leading to IR expansion for five genes, *psaB, rpoB, rpoC1, rpoC2,* and *rps2* (Table 1.12). Except for *psaB,* all PEP genes and *rps2* showed $dN/dS > 1$. A single gene, *rbcL*, had a decrease in *dN*/*dS* on the branch leading to IR expansion.

1.4 Discussion

Phylogenetic relationships

Phylogenetic relationships within *Passiflora* using 32 taxa from all five subgenera and 68 plastid protein-coding genes are well resolved and strongly supported (Figure 1.1A). Subgeneric relationships are congruent with the most recent phylogeny using

comprehensive taxon sampling reconstructed with combined four molecular markers, two each from the plastid and nuclear genomes (Krosnick et al. 2013). Subgenus *Deidamioides* is clearly polyphyletic with species sister to subgenera *Astrophea*, *Tetrapathea* + *Decaloba,* and nested within subgenus *Decaloba*. In contrast, subgenus *Passiflora* is strongly supported as monophyletic. A major incongruence with Krosnick et al. (2013) is the relationship between subgenus *Passiflora* and species from subgenus *Deidamioides* sections *Polyanthea*, *Deidamioides* and *Tetrastylis*. Section *Tetrastylis* includes two species, *P. ovalis* and *P. contracta*, whereas, sections *Polyanthea* and *Deidamioides* are monotypic and include *P. cirrhifolia* and *P. deidamioides*, respectively (Krosnick et al., 2013). In Krosnick et al. (2013), the clade with *P. cirrhifolia* and *P. ovalis* (other member of section *Tetrastylis*) is sister to subgenus *Passiflora* with moderate support (BS = 64%). In contrast, the plastid phylogeny (Figure 1.1A) provides strong support (BS = 93%) for a sister relationship between the *P. cirrhifolia*/*P. contracta* clade and the clade that includes subgenera *Tetrapathea* + *Decaloba*, similar to a previous phylogeny based on two plastid markers (Hansen et al., 2006). Expanded taxon sampling using similar sets of plastid genes resulted in a subgeneric phylogeny (Figure 1.1A) that is congruent with Krosnick et al. (2013). The phylogeny inferred by combining the 10 plastid genes with the highest phylogenetic informativeness (Figure 1.10A) had a nearly identical topology and support values to the 68 plastid genes tree (Figure 1.1A). The only difference is the relationship among three species, *P. edulis, P. quadrangularis* and *P. cincinnata*, in subgenus *Passiflora*. Likewise, the phylogeny constructed by concatenating genes with least phylogenetic informativeness, i.e. similar number of sites to 10 genes with highest

phylogenetic informativeness, generated a topology congruent with the 68 plastid genes tree (Figure 1.10B). Studies have shown that phylogenetic analysis using few taxa but large datasets could lead to systematic error and stronger support for a misleading phylogenetic relationship (Pollock et al. 2002; Hillis et al. 2003; Heath et al. 2008).

Plastome rearrangements

Complete plastomes are now available for 31 *Passiflora* species across all five subgenera, providing an excellent system for examining the evolutionary history of plastome rearrangements in the genus. Three major types of plastomic rearrangements, gene and intron losses, inversions and IR expansion/contraction are widespread in *Passiflora.* The discussion will focus on the phylogenetic distribution of these rearrangements and the phenomena that may underlie the more labile structure of plastomes in the genus.

Gene loss: Compared to the ancestral angiosperm plastome (Bock 2007), *Passiflora* exhibits considerable variation in protein coding gene content (69-75 genes, Table 1.6) primarily due to gene loss, especially RS including *rps7, rps16, rpl20, rpl22* and *rpl32.* Knockout assays have shown that all of these RS genes are essential in tobacco (Rogalski et al. 2008; Fleischmann et al. 2011) except for *rps7,* whose essentiality has yet to be experimentally verified. Losses of ribosomal genes in plastids are often associated with functional transfer to nucleus, such as *rpl32* in Salicaceae (Ueda et al. 2007) and *Thalictrum* (Park et al. 2015)*,* and *rpl22* in multiple lineages of rosids (Gantt et al. 1991; Jansen et al., 2011) or functional replacement by a nuclear copy as in *rps16* in *Medicago*

and some Salicaceae (Ueda et al. 2008). Among the missing ribosomal subunit genes in *Passiflora*, losses of *rps7* and *rpl20* have not yet been reported in any other photosynthetic angiosperm (Ruhlman and Jansen 2014). The phylogenetic distribution of gene loss in *Passiflora* shows multiple events for several genes, including *rpl20, rpl32, ycf1* and *ycf2* (Figure 1.1A). An alternative explanation to a multiple loss and nuclear transfer scenario is that there was a single functional transfer in the ancestor of *Passiflora* but the plastid copies have not yet been lost in all species. Future investigations that search for transferred genes in nuclear transcriptomes are needed to resolve the fate of missing plastid genes. In view of the extensive gene loss in plastomes of *Passiflora* the genus is an excellent system for investigating intracellular gene transfer (IGT).

One of the most interesting putative gene losses in *Passiflora* plastomes is the α subunit of PEP, *rpoA*. All genes encoding PEP subunits are known to be essential since the mutants are photosynthetically defective in tobacco (Serino and Maliga 1998). The only demonstrated case of *rpoA* loss in a land plant occurs in mosses (Sugiura et al. 2003; Goffinet et al. 2005) where there is a documented transfer of the gene to the nucleus in *Physcomitrella*. Earlier studies of Geraniaceae suggested that *rpoA* may have been lost in *Pelargonium* (Palmer et al. 1987; Chumley et al. 2006; Jansen et al. 2007; Guisinger et al. 2008). Blazier et al. (2016a) examined three unrelated lineages of angiosperms (*Annona*, *Passiflora* and *Pelargonium*) with highly divergent *rpoA* sequences and concluded that the gene was likely functional in the plastid in all three groups because the gene sequences contained the three conserved functional domains and were experiencing purifying selection. The authors argued that two factors are responsible for the high divergence, the

labile nature of the *rpoA* gene product and the high level of genomic rearrangements via illegitimate recombination in these lineages. Blazier et al. (2016a) included four species of *Passiflora* in their study with the most divergent species, *P. biflora*, from subgenus *Decaloba*, having 53.6% and 37.4 % nucleotide and amino acid sequence identities compared to *Populus*, respectively. This study includes 11 species from subgenus *Decaloba* with some species having nucleotide identities < 50% (Table 1.7). Among them, the two species *P. tenuiloba* and *P. suberosa* have nucleotide identities of 39%, and amino acid identities of 24.9% and lack all three functional domains. Based on the CDD search and extremely high levels of nucleotide divergence we predict that plastid-encoded *rpoA* is non-functional in these two species. A comparative analysis using transcriptome data for *Passiflora* species with highly divergent *rpoA* is needed to resolve the fate of this gene.

Inversions: This expanded study indicates that inversions are prevalent in *Passiflora* as well as in genus *Adenia* suggesting that the structural rearrangements are not restricted only to the genus *Passiflora* in Passifloraceae (Figure 1.1B). Complete plastomes from the remaining genera of Passifloraceae are needed to reconstruct the ancestral plastome structure in the family and to provide insights into inversions found in *Passiflora.* With our taxon sampling, a synapomorphic inversion of LCB 1 shared by all *Passiflora* and *A. mannii* indicates the structural change occurred prior to origin of the genus *Passiflora* (Figures 1.1B, 1.5, Table 1.3-B). Similarly, LCBs 8 and 9 were inverted multiple times (Figures 1.1B, 1.5). A plausible explanation would be that inversion of LCBs 8-9 occurred at the most basal node including *A. mannii* and probably reversed in some *Passiflora* more recently. This suggests that gene order in *Passiflora* may not be static but is rather dynamic.

Nonetheless, two syntenic gene blocks, LCB 2 and LCB 15, are the only LCBs not inverted in any species indicating possible plastome regions immune to rearrangement in *Passiflora.* Even though inversions are widespread in the genus, subgenus *Decaloba* has experienced the greatest number of inversions, primarily in a few species, *P. misera* and *P. tenuiloba*/*P. suberosa*, with the highest reversal and breakpoint distances, respectively (Table 1.3C).

Despite extensive inversions in Passifloraceae, gene order between the most closely related species is very similar supporting the phylogenetic relationships inferred using plastid protein-coding genes. However, there are a few exceptions to this pattern, including *P. foetida* and *P. obovata* that have a nearly identical gene order but are distantly related (Figure 1.1A). Furthermore, the phylogenetic distribution of inversions (Figures 1.1B, 1.5) indicates considerable homoplasy, suggesting that such events may not be reliable phylogenetic markers in the family. Several previous studies in disparate angiosperm lineages documented homoplasy in plastome inversions (Hoot and Palmer 1994; Cosner et al. 2004; Weng et al. 2014; Schwarz et al. 2015), suggesting that caution should be used in applying these events to infer phylogenetic relationships.

IR expansion and contraction: IR boundaries in some *Passiflora* species have shifted drastically compared to the ancestral angiosperm plastome IR (Zhu et al. 2016) (Figure 1.1A, C). All Passifloraceae share a minor expansion that incorporated a complete copy of *rps19* into the IR. The phylogenetic distribution of IR changes indicates that most other IR modifications occurred recently because they appear in more terminal clades (Figure 1A, C). Among these recent changes, two distinct patterns of IR boundary fluctuations are evident, major IR expansions within subgenus *Decaloba* and severe IR

contraction in two distantly related species, *P. menispermifolia* (subgenus *Passiflora*) and *P. obovata* (subgenus *Deidamioides*). The severe contractions in *P. menispermifolia* and *P. obovata* were independent events that resulted in almost identical gene content that excludes all ribosomal rRNA genes from the IR.

In photosynthetic angiosperms the only example of a reduced IR lacking part of rRNA operon is *Monsonia speciosa* in the Geraniaceae (Guisinger et al. 2011). However, IR loss resulting in only one set of rRNA genes in the plastome has been documented in several lineages, including three genera of Geraniaceae (Guisinger et al. 2011; Blazier et al. 2011, 2016b; Ruhlman et al. 2017), a clade in Fabaceae (Palmer et al. 1987; Lavin et al. 1990; Wojciechowski et al. 2004), and one genus each in Cactaceae (Sanderson et al. 2015) and Arecaceae (Barrett et al. 2016). Major expansion occurred in almost all species in subgenus *Decaloba* except *P. microstipula* (Figure 1.1A, C). A likely scenario for this major expansion involves incorporation of genes up to *rpoA* into IR, followed by multiple incremental expansions integrating genes up to *accD* in *P. jatunsachensis* + *P. rufa*, *psaI* in *P. tenuiloba* + *P. suberosa,* and part of *cemA* in *P. lutea* and *petA* in *P. auriculata,* (Figure 1A, C). The final IR expansion in *P. auriculata* resulted in a similar size of the LSC and IR with nearly equal gene number. To explain the dynamics of IR boundary change in subgenus *Decaloba* precisely, a thorough taxon sampling within the subgenus, specifically complete plastomes from the species in supersections *Disemma, Multiflora, Bryoniodes* and *Decaloba* (see Krosnick et al. 2013 for detailed *Decaloba* phylogeny), will be needed. The major IR expansion and contraction in *Passiflora* is unusual with a similar situation only

known in one other photosynthetic angiosperm family, Geraniaceae (Chumley et al. 2006; Guisinger et al. 2011; Blazier et al. 2011; Ruhlman et al. 2017; Weng et al. 2017).

Geraniaceae includes species with the largest IR in *Pelargonium transvaalense* (88 kb), and species with extremely reduced IR or absent in *Monsonia, Erodium* and *Geranium*. The IR fluctuations in Geraniaceae are more extreme but *Passiflora* is noteworthy because it includes species with both major IR expansion and contraction. The gradual increase in IR size as noted in subgenus *Decaloba* has been reported in monocots but on a smaller scale (Wang et al. 2008). Among eudicots, besides *Passiflora,* the progressive expansion of the IR has been documented in *Pelargonium* (Weng et al. 2017) and perhaps a similar molecular mechanism may be driving the IR evolution in these two unrelated lineages.

Considering plastomes as predominately circular, Goulding et al. (1996) proposed possible mechanisms to account for large and small IR expansions via boundary migration within a single plastome unit. Minor IR expansions were attributed to gene conversion between the two copies of IR. Resolution of the heteroduplex formed when Holliday junction branch migration crosses an IR boundary could have resulted in minor expansions. More extensive expansions of the IR were proposed to arise during double strand break (DSB) repair. A DSB in one IR copy, strand invasion at the other IR copy to template repair, extension beyond the IR boundary and recircularization via illegitimate recombination between polyA tracts could have produced the major IR expansion observed in *Nicotiana acuminata* (Goulding et al. 1996). Double stranded breaks within a circular plastome and subsequent repair has also been postulated for progressive expansion of IR in monocots on a smaller scale (Wang et al. 2008).

However, it is now known that plastomes are mostly linear and/or branched and that most recombination occurs between different copies of the unit genome (Oldenburg and Bendich, 2004; Scharff and Koop, 2006). Thus, a model that invokes intramolecular recombination between IR copies to explain IR expansion/contraction is no longer valid. Strand invasion of internal plastome regions by repeated sequences (i.e. IR sequences) situated at the end of different copies of linear plastomes was proposed to generate plastome single copy region isomers during replication (Oldenburg and Bendich, 2004), a phenomenon that was previously attributed to flip-flop recombination between IRs within a single circular genome (Palmer, 1983). IR expansion and contraction in *Passiflora* was likely caused by recombination between IR copies of different linear unit genomes. Once the IR expansion or contraction occurred, the new boundaries would spread and be maintained by homologous recombination (HR)-mediated gene conversion between plastome copies situated on linear and/or branched structures.

Genes with elevated substitution rates and dN/dS

Nucleotide substitution rate analyses in *Passiflora* reveal that increases in rates are both clade- and locus-specific. Locus-specific increases in substitution rates were detected in *clpP,* where *dS* and *dN* were accelerated in almost all species, except in *dS* for clade C (Figures. 3, 1.6-1.7, Table 1.8). As a consequence, the average *dN/dS* for *clpP* was elevated across the genus and significantly higher in clade C with an average of 1.4 (Figure 1.3). Most species of *Passiflora* have *dN/dS* close to 1 for *clpP* indicating that this gene is evolving neutrally, whereas *dN/dS* is significantly > 1 in two species from subgenus

Deidamioides, *P. cirrhifolia* and *P. contracta*, suggesting lineage-specific positive selection in clade C. Subgenus *Deidamioides* is polyphyletic (Figure 1.1A) and only some species exhibits *dN/dS* significantly > 1. *Passiflora* retains highly divergent *clpP* and both introns have been independently lost twice within the genus. This gene encodes a subunit of ATP-dependent caseinolytic protease (Clp), and is essential gene in plastids (Peltier et al., 2001; Peltier et al., 2004). Significant increases in *clpP* nucleotide substitution rates and *dN*/*dS* have been reported in two other angiosperm lineages, *Oenothera* and *Silene* (Greiner et al. 2008; Erixon and Oxelman 2008; Sloan et al. 2012; Williams et al. 2015).

A comprehensive study of molecular evolution of *clpP* across green plants with a primary focus on *Silene* showed that a divergent *clpP* is still functional and accelerated nucleotide substitutions are correlated with intron loss, plastome structural rearrangements and accelerated substitution rates in nuclear-encoded, plastid-targeted Clp subunits (Williams et al., 2019). In *Passiflora*, significantly elevated *dN*/*dS* for *clpP* occurs in species with intact introns and less rearranged plastomes with lower substitution rates for other plastid protein coding genes. Genus-wide increase in overall substitution rates, loss of introns, and further clade specific increases in substitution rates on the species with intact introns, suggest that different evolutionary forces may be driving *clpP* sequence divergence in *Passiflora*.

In addition to *clpP*, *rpl23* and *ycf4,* have significantly elevated *dN/dS* (> 1) in some clades within *Passiflora* (Figure 1.3). The significantly lower *dS* in subgenus *Passiflora* and *P. microstipula* resulted in elevated *dN*/*dS* in *rpl23*. The significantly higher *dN*/*dS* for *rpl23* without a significant increase in nucleotide substitution rates suggests that the

elevated *dN*/*dS* is consequence of lower *dS* rather than positive selection. Wolf et al. (2009) demonstrated that *dN*/*dS* may not necessarily reflect the selection pressure as the ratio may be heavily influenced by *dS*. The elevated *dN/dS* for *ycf4* in clade A and two species within clade E, *P. lutea* and *P. filipes,* are due to significant increases in both *dN* and *d*S relative to other *Passiflora* indicating relaxation of purifying selection. Accelerated substitution rates in *ycf4* with *dN*/*dS* > 1 has been demonstrated in some legume lineages, presumably due to localized hypermutation (Magee et al. 2010). The concept of mutational hotspot postulated by Magee et al. (2010) in the *ycf4-psaI-accD-rps16* region may also apply to *P. pittieri* and *P. arbelaezii* since both species share features similar to legumes, such as significantly increased substitution rates for *psaI* (Figure 1.3, Table 1.9-1.10), highly divergent *accD* with repetitive elements within the gene, and the loss of *rps16* (Jansen et al. 2007). The accumulation of repeats within *accD* and loss of *rps16* are also shared by other *Passiflora* species with increased substitution rates in several genes including those encoding ribosomal proteins (Figure 1.3). Therefore, it is plausible that localized hypermutation has played a role in plastome evolution in *Passiflora* as well as in legumes.

Accelerated substitution rates in the clade with major IR expansions

Branch lengths in the ML tree are substantially longer in members of clade E with major IR expansions (Figure 1.2). Pairwise comparisons of substitution rates show significantly accelerated *dS* and *dN* for PEP and RS genes in this clade (Figure 1.9, Tables 1.9-1.10). Likelihood ratio tests show deceleration of *dS* for some photosynthetic genes, acceleration of *dN* for genes encoding RS and PEP, and elevated *dN/dS* predominately in

PEP (Table 1.12). Comparisons of substitution rates for genes included in the IR due to IR expansion show that RS genes have higher *dN* regardless of their location in the plastome (Figure 1.9). Lineage-specific and IR independent increase in *dN* in RS genes has been reported in *Pelargonium* (Weng et al. 2017). In terms of increased *dN* for RS and PEP genes, *Passiflora* has the same pattern as *Pelargonium*, where aberrant DNA repair mechanisms have been suggested to drive the increase in substitution rate (Guisinger et al. 2008; Zhang et al. 2015; Blazier et al. 2016b). However, in contrast to *Pelargonium*, several RS genes are missing in *Passiflora* (Figure 1.1A), which suggests either functional transfer of missing RS genes to nucleus or replacement by a nuclear copy. In either scenario, a compensatory substitution in response to selective pressure, allowing the continued interaction of the nuclear and plastid-encoded subunits, may have caused higher nonsynonymous substitutions in other intact plastid-encoded ribosomal genes in *Passiflora.*

Compensatory evolution with increased substitution rates and elevated *dN/dS* in both nuclear-encoded plastid-targeted and plastid-encoded RS genes has been reported in *Silene* and *Pelargonium* (Sloan et al. 2014; Weng et al. 2016). This could also explain the acceleration of *dN* in *rpoB, rpoC1* and *rpoC2* in subgenus *Decaloba*. Because uncertainty in sequence alignment can lead to spurious results in rate analyses, the highly divergent *rpoA* gene in subgenus *Decaloba* was excluded. However, the three remaining PEP genes have elevated *dN/dS* (> 1) for the branches within *Decaloba* indicating positive selection on these subunits, possibly to compensate the high sequence divergence in *rpoA*. Guisinger et al. (2008) proposed that the correlated substitutions in the PEP could be driving the increase in sequence divergence in *Pelargonium* and this may apply to *Passiflora* as well.

1.5 Conclusions

This study shows that *Passiflora*, especially subgenus *Decaloba*, shares the features that are found in other photosynthetic angiosperm lineages with highly rearranged plastomes (see Table 1 in Ruhlman and Jansen 2018). The plastomes of *Passiflora* are analogous to Geraniaceae because both groups have extensive gene and intron losses, numerous gene order changes, substantial variation in IR extent, highly accelerated rates of nucleotide substitutions, biparental inheritance and plastome-genome incompatibility. The astounding resemblance in highly divergent genes, *accD, clpP* and *rpoA*, loss of two largest hypothetical genes, *ycf1* and *ycf2,* and accelerated substitution rates in RS and PEP genes suggests that, with regard to these genes, similar molecular mechanisms may be driving plastome evolution in Geraniaceae and Passifloraceae. *Passiflora* also shares features with other highly rearranged lineages, such as increased substitution rates in *clpP* along with intron losses in *clpP, rpoC1, atpF* and *rpl16* as in *Silene* (Sloan et al., 2012), and a hypermutable region, *ycf4-psaI,* in Fabaceae (Magee et al. 2010). The occurrence of these unusual features in a single genus suggests that multiple molecular mechanisms may be involved in plastome instability in *Passiflora.* A few proteins encoded in nucleus and targeted to the plastid have been identified that play a crucial role in plastome stability, such as Whirlies (Maréchal et al. 2009), MSH1, RecA recombinase and RecG helicase (Odahara et al. 2015a, b). Knockout assays in all of these studies have shown that plastome stability is maintained via suppression of illegitimate or aberrant recombination between dispersed repeats of various sizes and ensuring HR-dependent repair of damaged DNA. It is

conceivable that the impairment of these plastome DNA repair genes may have contributed to extensive rearrangements in *Passiflora*, making this genus an excellent lineage to investigate plastome evolution.

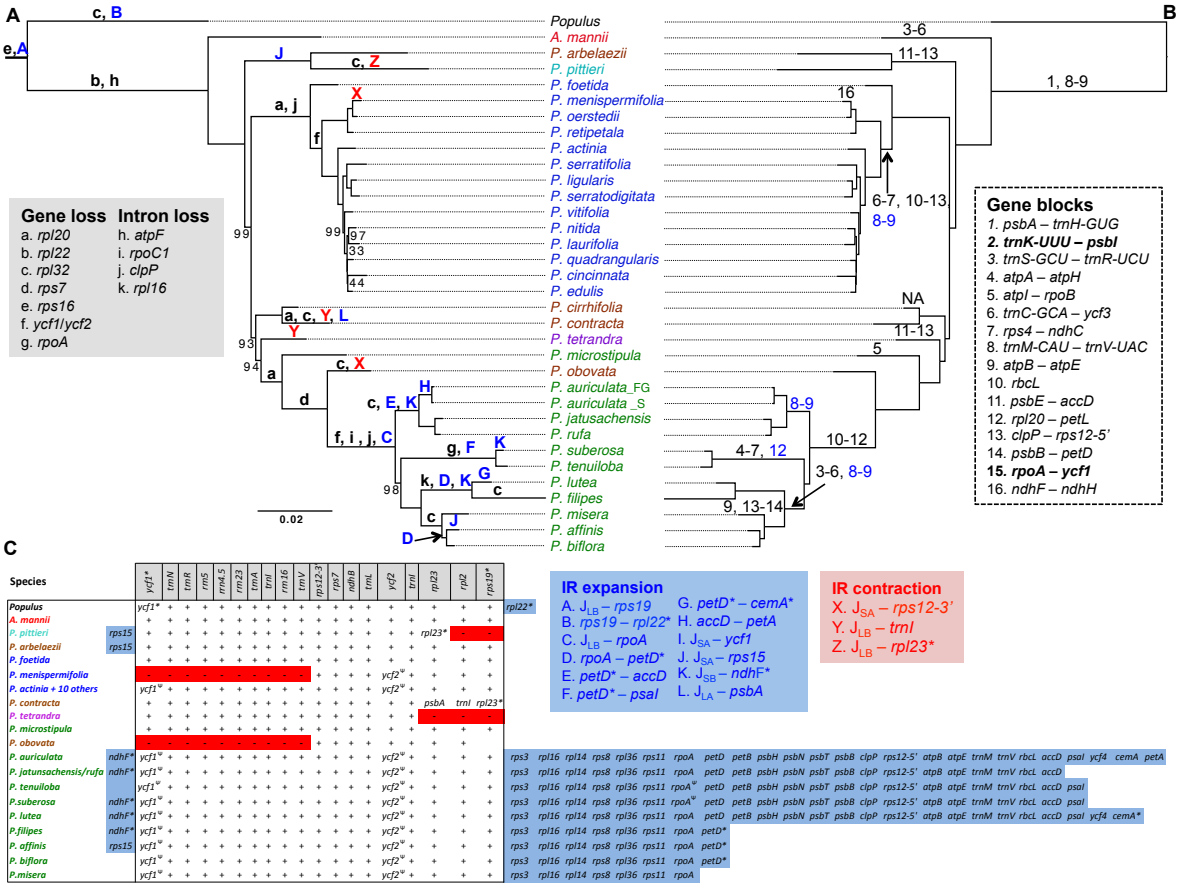


Figure 1.1. Distribution of *Passiflora* plastome rearrangements on maximum likelihood (ML) tree inferred from 68 protein-coding genes using IQ-TREE. (A) Gene/intron losses (black), inverted repeat (IR) expansions (blue) and IR contractions (Red) are plotted on the branches. Except where indicated all bootstrap values were 100%. Horizontal bar indicates the expected number substitutions per site. (B) Inversion events on the ML tree (mirrored from A). Gene blocks within the inversion are numbered. Two gene blocks were not inverted in any species and are highlighted in bold (inset). Gene blocks highlighted in blue represents the reversion events. (C) Genes included in and excluded from the IR by expansion (blue) and contraction (red) of the IR boundary, respectively, for the corresponding species in Passifloraceae. Genes highlighted in grey on the top of the table are within the ancestral IR boundary of angiosperms. Plus (+) and minus (-) represent the presence and absence of the gene, respectively. Genes with asterisks denote the partial copy of gene within the IR. Pseudogenes are indicated by (Ψ). Species names are color-coded to indicate their generic or subgeneric placement: Red (*Adenia*), cyan (*Astrophea*), orange (*Deidamioides*), blue (*Passiflora*), purple (*Tetrapathea*), and green (*Decaloba*). Abbreviations: IR_A- inverted repeat A; IR_B- inverted repeat B; J_{LB} – junction of large single

copy and IR_B; J_{LA} – junction of large single copy and IR_A; J_{SA} – junction of small single copy and IR_A; J_{SB} – junction of small single copy and IR_B.

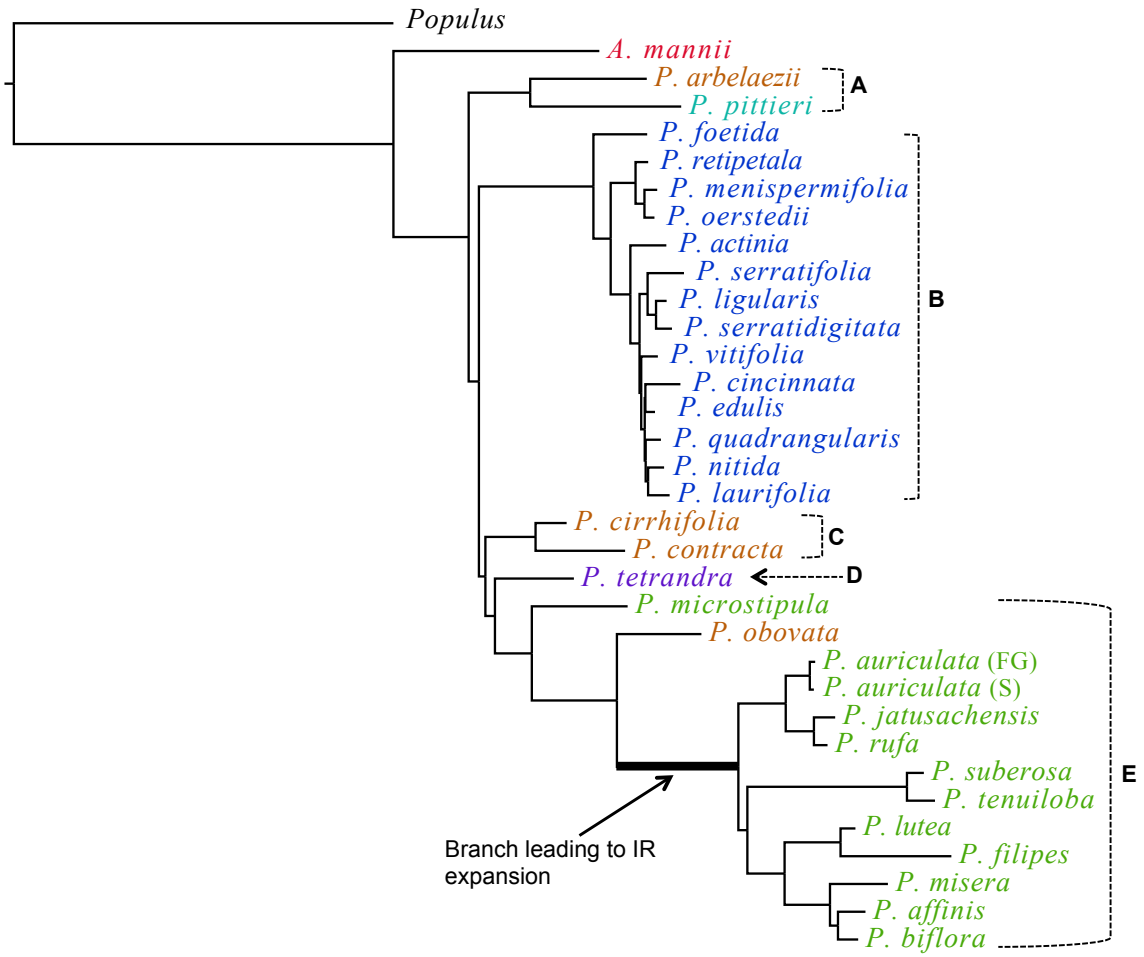


Figure 1.2. ML tree inferred from 68 protein-coding gene using IQ-TREE with five clades (A-E) indicated. The branch leading to IR expansion within clade E has been highlighted. Species names are color-coded to indicate their generic or subgeneric placement: Red (*Adenia*), orange (*Deidamioides*), cyan (*Astrophea*), blue (*Passiflora*), purple (*Tetrapathea*), and green (*Decaloba*).

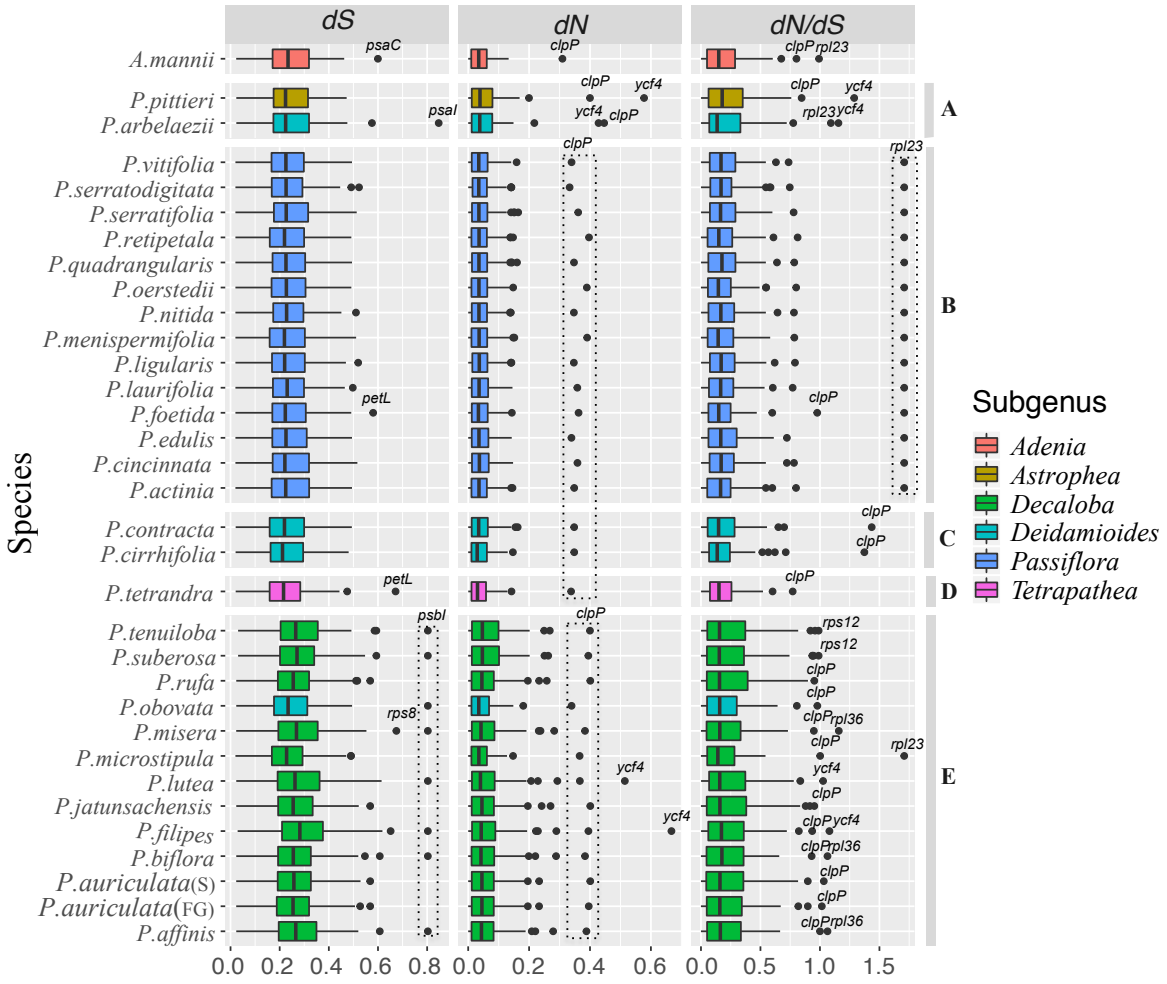


Figure 1.3. Boxplot showing the variation in pairwise *dS* and *dN,* and *dN*/*dS* for the *Adenia* + *Passiflora* species estimated by comparison with *Populus trichocarpa* using PAML (Yang, 2007). Outliers for each category were labeled with gene names. Clades (A-E, Figure 1.2) used to group species to calculate statistical significance are labeled on right-hand side. Boxplots were color-coded to indicate generic or subgeneric placement.

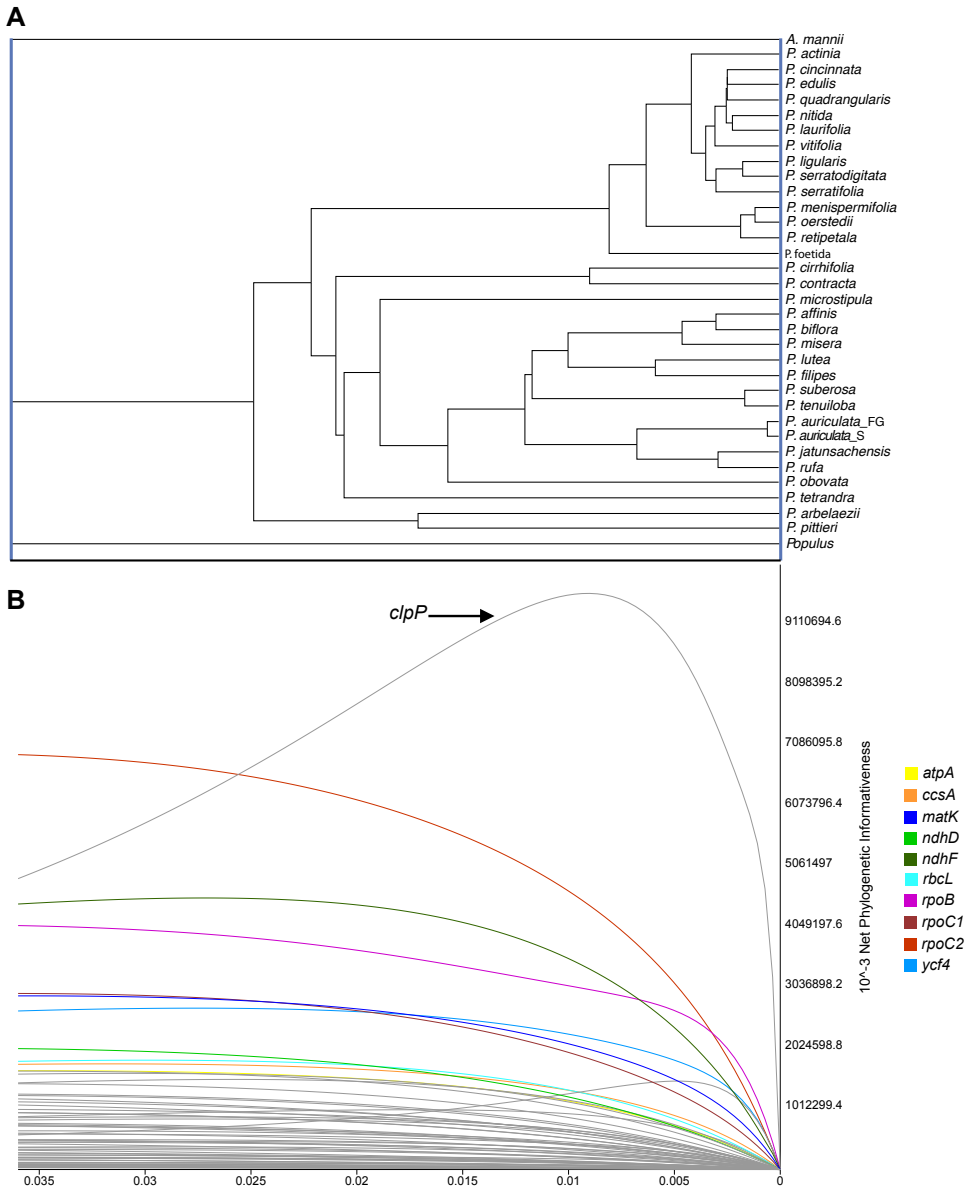


Figure 1.4. *Passiflora* ultrametric tree and phylogenetic informativeness profile estimated in PhyDesign (Lopez-Giraldez and Townsend 2010). (A) The ultramteric tree was generated using *dnamlk* in Phylip (Felsentein 1989). (B) Net phylogenetic informativeness profile for 68 plastid protein-coding genes. Ten genes with the greatest informativeness are color-coded and indicated at the right. X- and Y-axes represent relative-time and net phylogenetic informativeness, respectively.

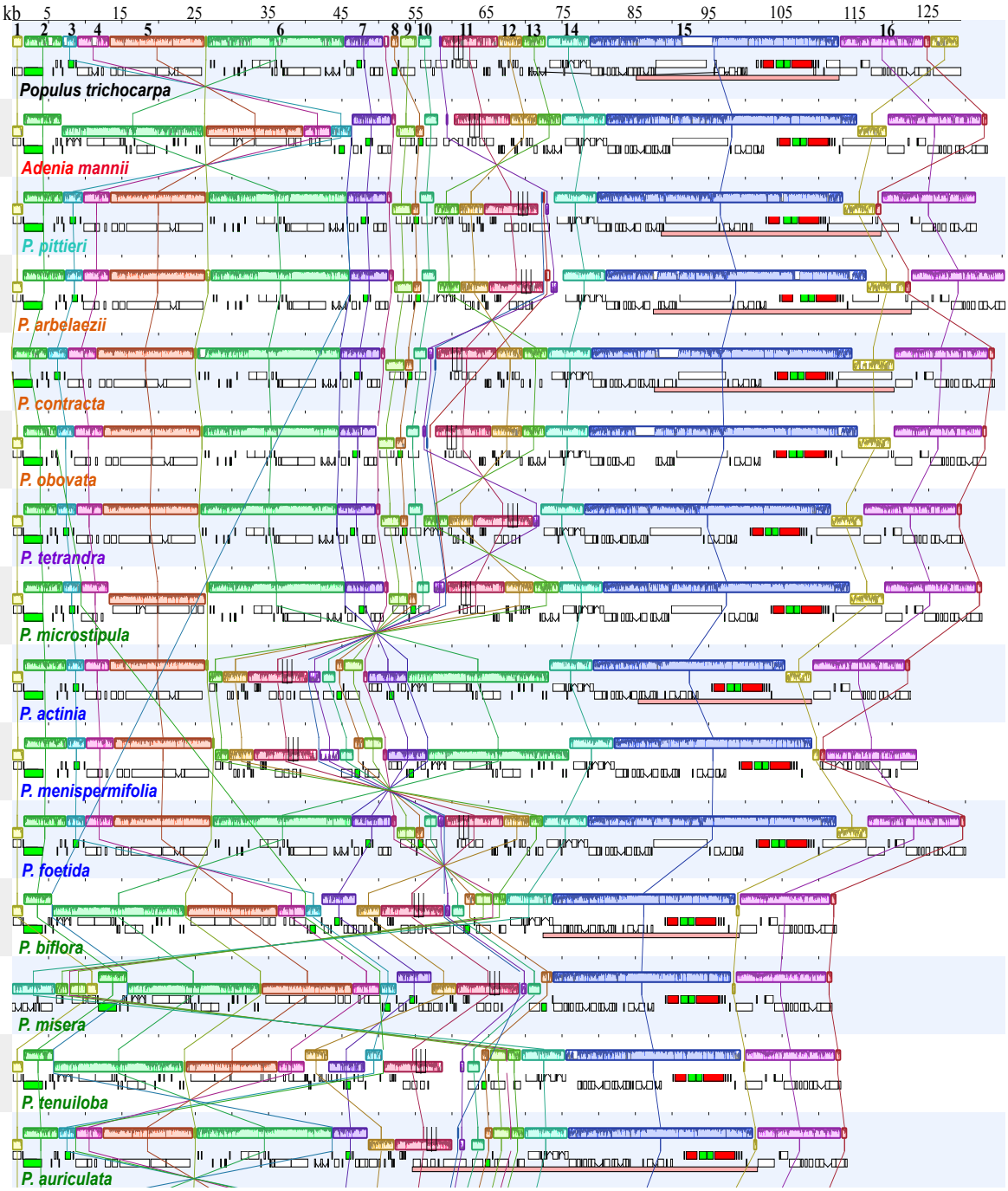


Figure 1.5. Whole genome alignment of *Passiflora* (*P.*) and *Adenia mannii* plastomes compared to *Populus trichocarpa. Populus trichocarpa* (Salicaceae) was used as the reference genome. One copy of the inverted repeat was removed from each species. Species listed in the progressiveMauve (Darling et al. 2010) alignment represent the unique gene order in genera *Passiflora* and *Adenia*. Color-coded Locally Collinear Blocks (LCBs) are

numbered (1-16) and represents syntenic regions. Genes within the numbered LCBs are listed in Table S3-A. Gene orders for all *Passiflora* and *Adenia* species are shown in Table S3-B. Numbers labeled on the top of the x-axis represent plastome coordinates in kilobases (kb). Species names are color-coded to indicate their generic or subgeneric placement; Red (*Adenia*), cyan (*Astrophea*), orange (*Deidamioides*), blue (*Passiflora*), purple (*Tetrapathea*), and green (*Decaloba*).

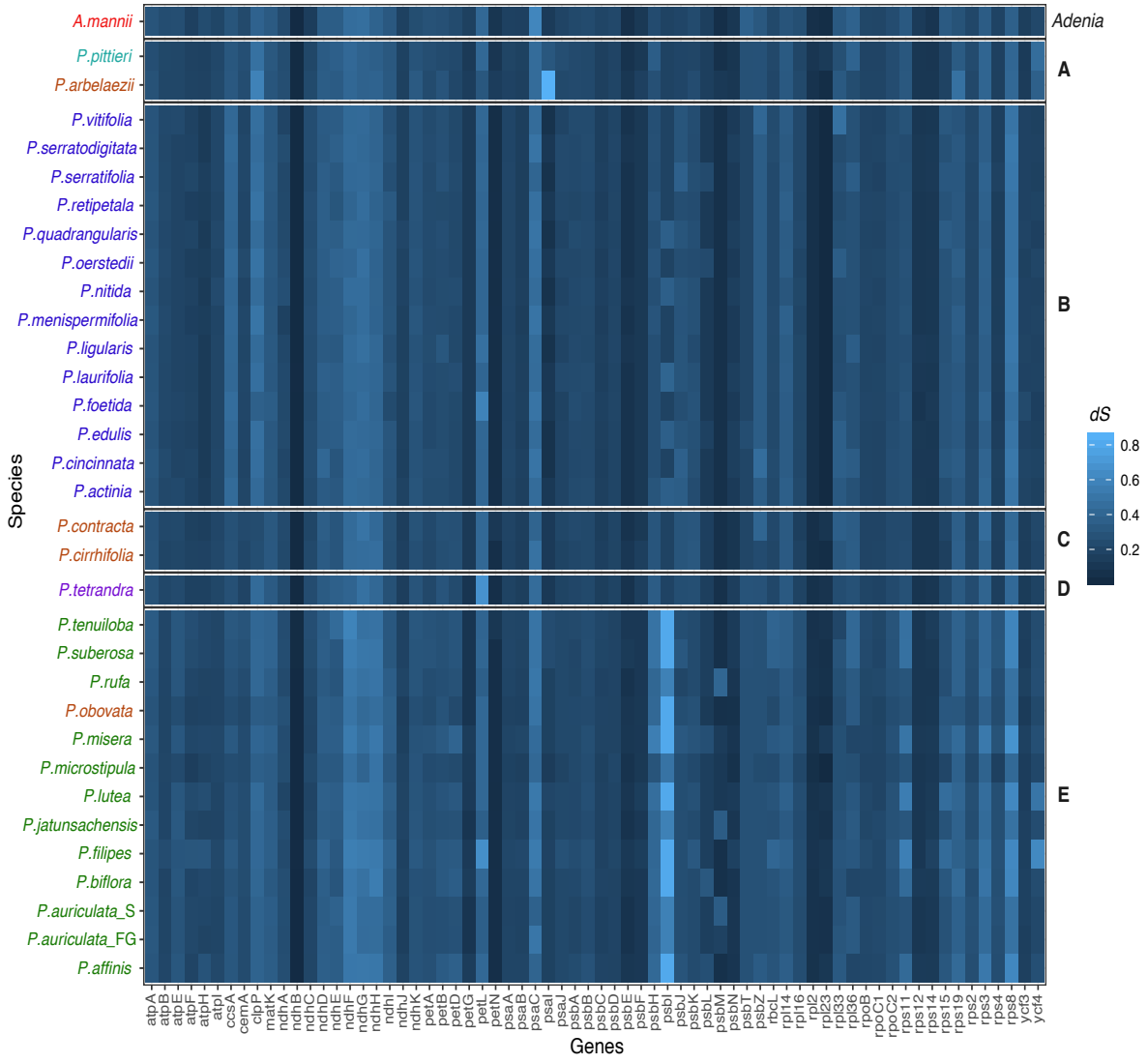


Figure 1.6. Heatmap showing pairwise *dS* for each species in Passifloraceae compared to *Populus trichocarpa* for 68 protein-coding genes. Clades (A-E) of closely related species are labeled on right-hand side based on Figure 1.2. Species names are color-coded to indicate their generic or subgeneric placement; Red (*Adenia*), cyan (*Astrophea*), orange (*Deidamioides*), blue (*Passiflora*), purple (*Tetrapathea*), and green (*Decaloba*).

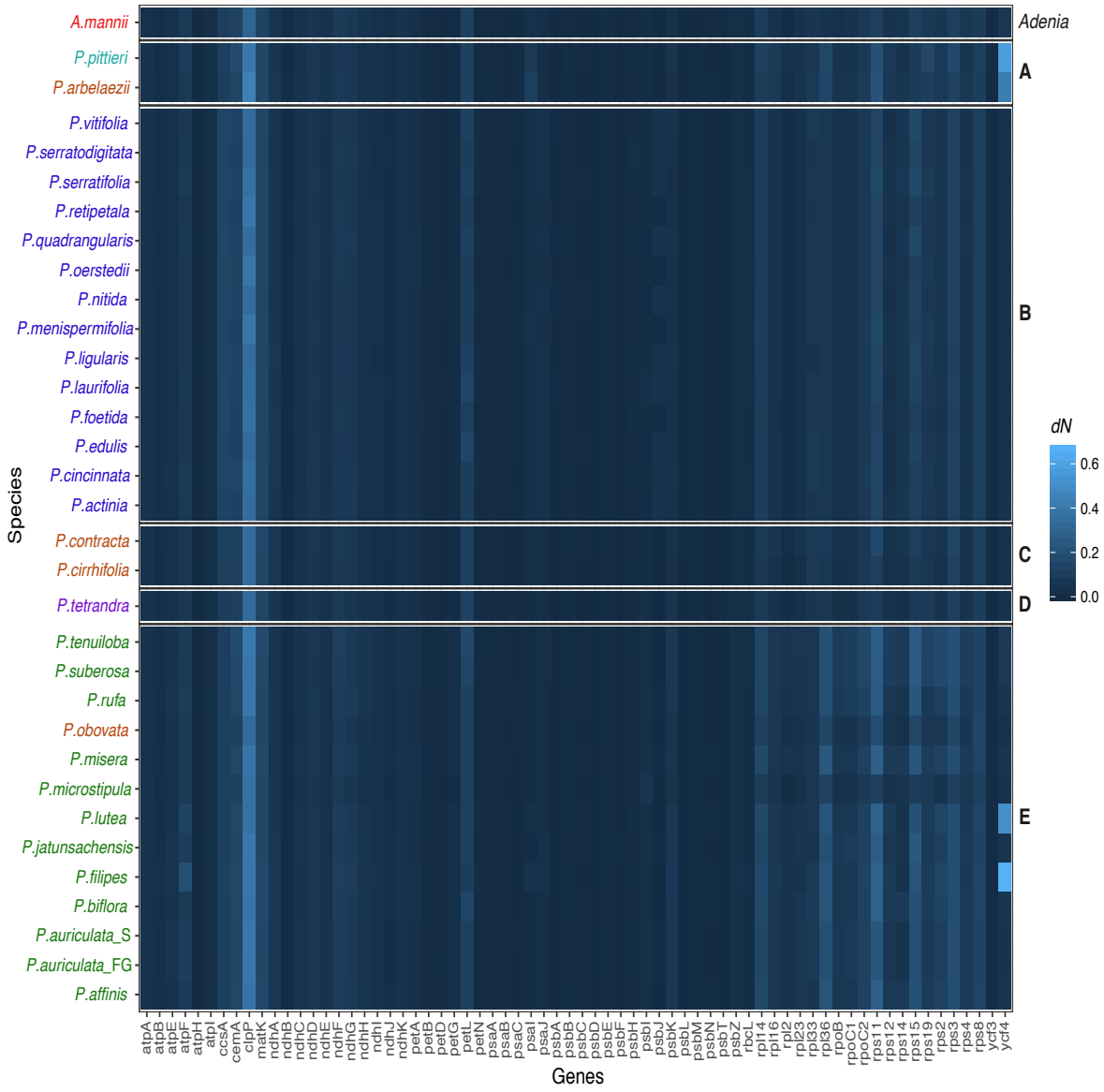


Figure 1.7. Heatmap showing pairwise *dN* for each species in Passifloraceae compared to *Populus trichocarpa* for 68 protein-coding genes. Clades (A-E) of closely related species are labeled on right-hand side based on Figure 1.2. Species names are color-coded to indicate their generic or subgeneric placement; Red (*Adenia*), cyan (*Astrophea*), orange (*Deidamioides*), blue (*Passiflora*), purple (*Tetrapathea*), and green (*Decaloba*).

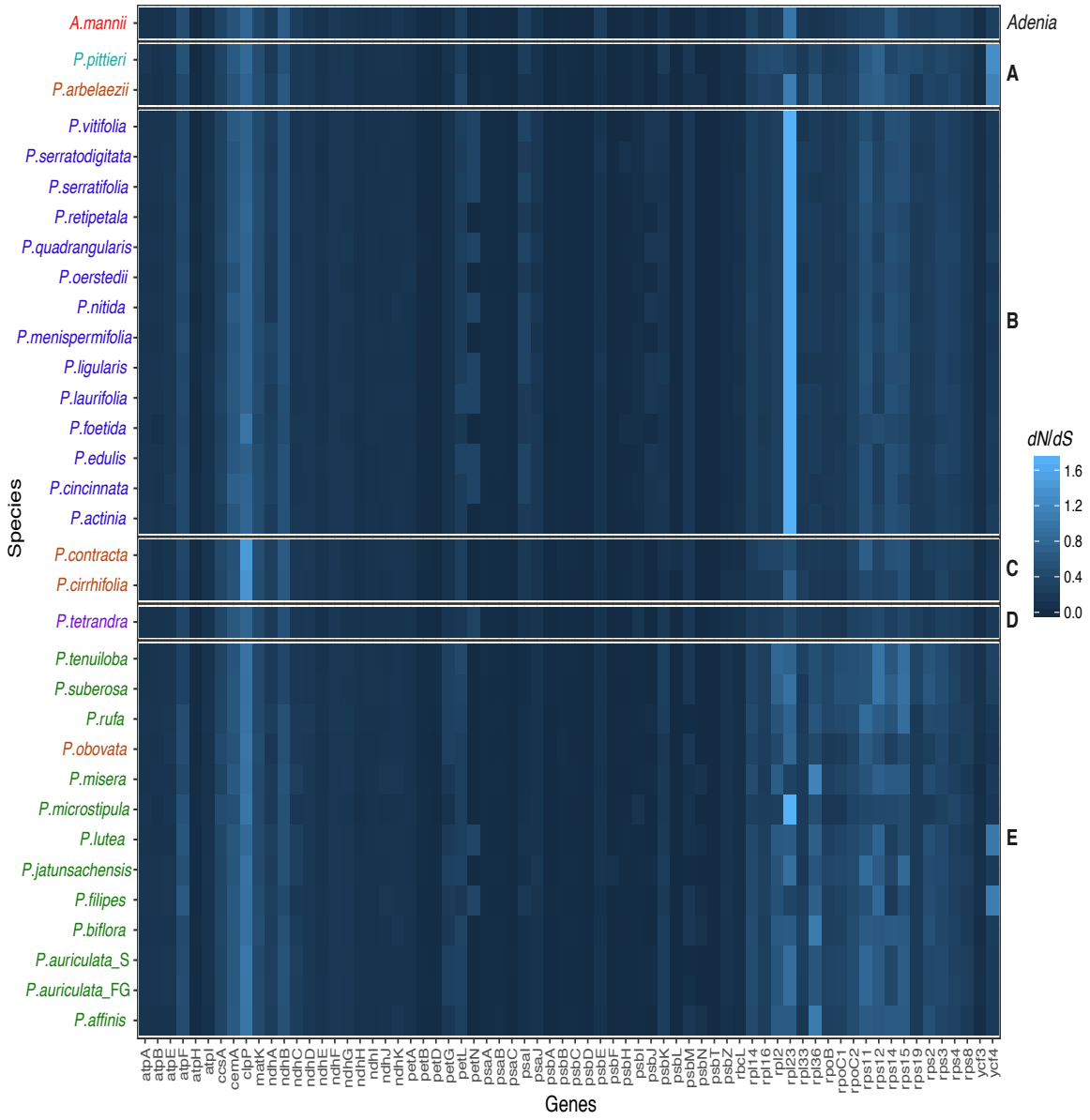


Figure 1.8. Heatmap showing *dN*/*dS* for each species in Passifloraceae compared to *Populus trichocarpa* for 68 protein-coding genes. Clades (A-E) of closely related species are labeled on right-hand side based on Figure 1.2. Species names are color-coded to indicate their generic or subgeneric placement; Red (*Adenia*), cyan (*Astrophea*), orange (*Deidamioides*), blue (*Passiflora*), purple (*Tetrapathea*), and green (*Decaloba*).

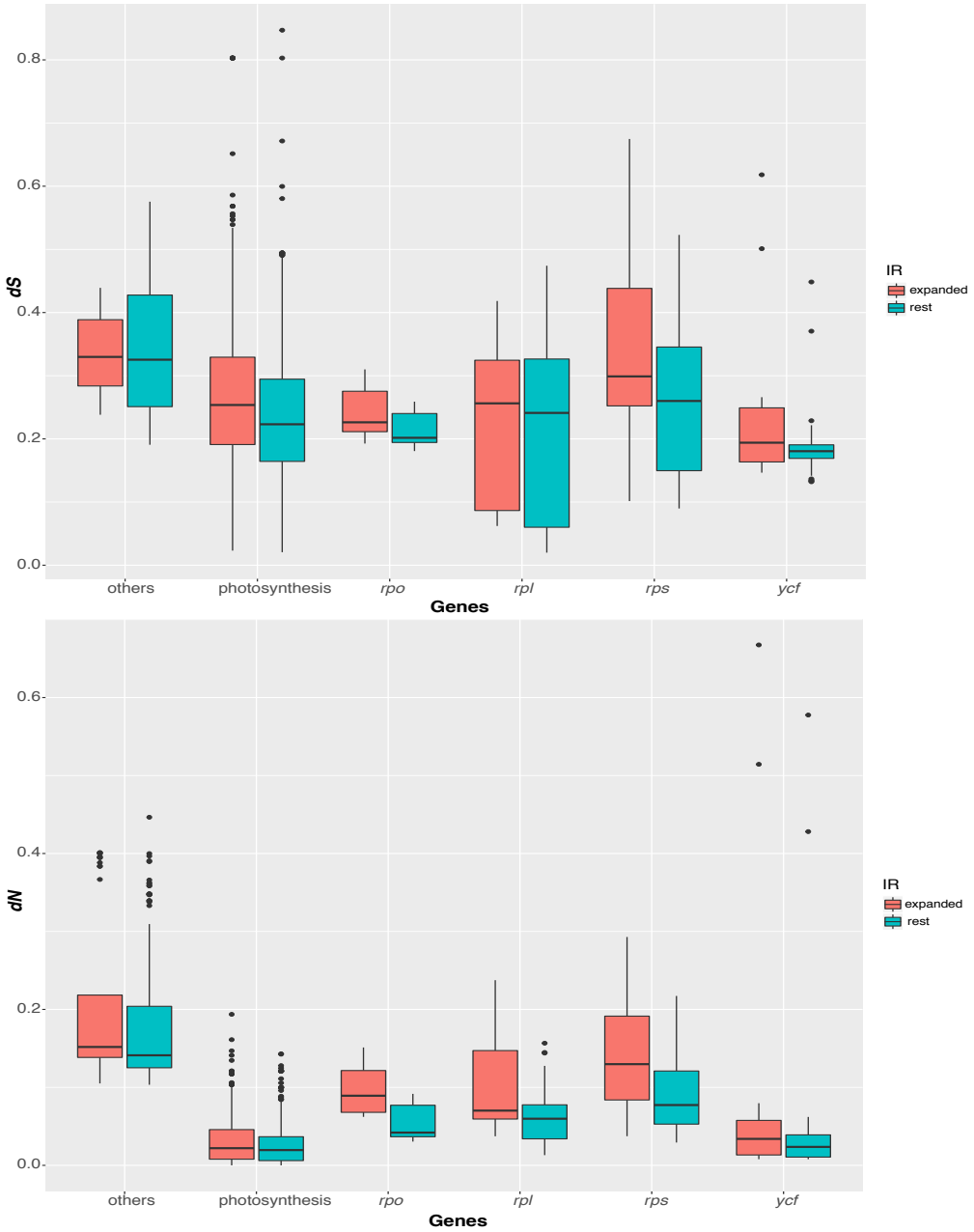


Figure 1.9. Comparison of *dS* and *dN* between genes that are located in the IR or single copy regions categorized by functional group. Each box characterizes interquartile range (IQR, difference between first quartile and third quartile) and the horizontal line through the box represents the median (second quartile). The vertical line extended on each side includes data within 1.5 times IQR. Outliers beyond the vertical line are shown as points.

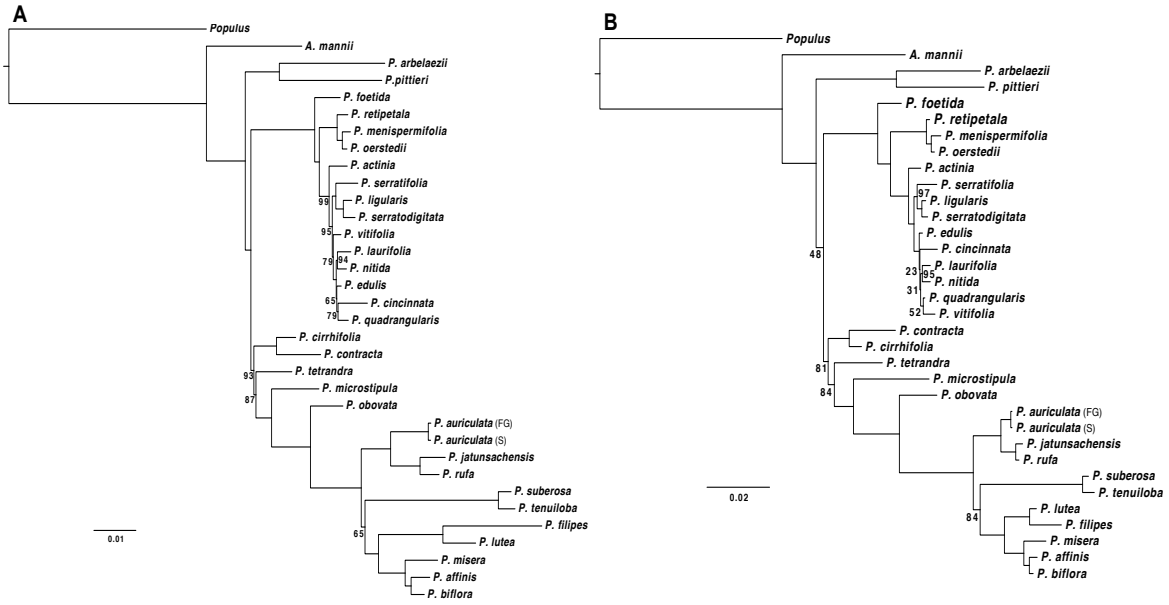


Figure 1.10. Comparison of tree topologies inferred from plastid genes with most and least phylogenetic signal using IQ-TREE. (A) Maximum likelihood (ML) tree generated using 10 genes (20,0001 bp) with most phylogenetic informativeness. The log-likelihood score of the tree is -58845.65. (B) ML tree generated using 48 genes (20,508 bp) with least phylogenetic informativeness and log-likelihood score of -39378.83.
Except where indicated all bootstrap values were 100%. Horizontal bar indicates the expected nucleotide substitutions per site.

Table 1.1. *Passiflora* species with greenhouse accession number (AN), GenBank accession number (GBN) and original collection location. All vouchers are deposited at TEX-LL except for *P. tetrandra*. OS, Ohio State University.

Subgenus	Species	AN	Location	Voucher	GBN
Passiflora	*P. foetida* L.	TX-004	Texas, USA	Shrestha201	MK694932
	P. menispermifolia Kunth.	8039	Corcovado, Costa Rica	Shrestha202	MK694933
	P. actinia Hook.	9031	Not available	Shrestha101	MF807934
	P. edulis Sims.	9408	Cali Valley, Columbia	Shrestha102	MF807938
	P. laurifolia L.	9109	French Guiana	Shrestha103	MF807939
	P. ligularis A. Juss.	9334	Cali Valley, Columbia	Shrestha104	MF807940
	P. nitida Kunth.	8060	Manaus, Brazil	Shrestha105	MF807941
	P. oerstedii Mast.	7005	Selva, Costa Rica	Shrestha106	MF807942
	P. quadrangularis L.	8054	Corcovada, Costa Rica	Shrestha107	MF807944
	P. retipetala Mast.	7007	Arima Pass, Trinidad	Shrestha108	MF807945
	P. serratifolia L.	8058	Belize	Shrestha109	MF807948
	P. serrato-digitata L.	6002	Arima Pass, Trinidad	Shrestha110	MF807946
	P. vitifolia Kunth.	9041	Sirena, Costa Rica	Shrestha111	MF807947
Decaloba	*P. biflora* Lam.	6001	Puerto Viejo, Costa Rica	Shrestha112	MF807937
	P. auriculata Kunth.	9406	French Guiana	Shrestha113	MF807935
	P. auriculata Kunth.	9407	Suriname	Shrestha114	MF807936
	P. affinis Engelm.	TX-003	Texas, USA	Shrestha203	MK694930
	P. filipes Benth.	9493	Texas, USA	Shrestha204	MK694929
	P. jatunsachensis Schwerdtfeger	9402	Ecuador	Shrestha205	MK694920
	P. lutea L.	TX-002	Texas, USA	Shrestha206	MK694922
	P. microstipula Gilbert & MacDougal	9271	Vera Cruz, Mexico	Shrestha207	MK694934
	P. misera Kunth.	9023	J. Turner, Leeds University, England	Shrestha208	MK694928
	P. rufa Feuillet & MacDougal	9086	French Guiana	Shrestha209	MK694924
	P. suberosa L.	9494	Texas, USA	Shrestha210	MK694921
	P. tenuiloba Engelm.	TX-001	Texas, USA	Shrestha211	MK694923
Deidamioides	*P. arbelaezii* L. Uribe	8027	Puerto Viejo, Costa Rica	Shrestha212	MK694926
	P. contracta Vitta.	8071	Espirito de Santo, Brazil	Shrestha213	MK694925
	P. obovata Killip	999	Butterfly World	Shrestha214	MK694931
Tetrapathea	*P. tetrandra* Banks & Sol. ex DC.	N/A	New Zealand	S. Krosnick 266 (OS)	MK694927
Astrophea	*P. pittieri* Mast.	9219	Sirena, Costa Rica	Shrestha115	MF807943
Adenia	*A. mannii* (Mast.) Engl.	7008	Kumba, Cameron	Shrestha215	MK651116

Table 1.2. Oligonucleotide primers used in PCR and Sanger sequencing validation for unusual inverted repeat (IR) boundaries in *Passiflora*. bp, basepairs.

Species	**Location**	**Primer**	**Length (bp)**
P. menispermifolia	*ndhF* IR	5' CGGCGGGATTAACAGCCTTT 3' 5' TCTGTCTCGGTAGGATATACATGT 3'	945
	rrn16 IR	5' CAGGATCGAACTCTCCATCAGA 3' 5' TCTGTCTCGGTAGGATATACATGT 3'	924
P. obovata	*ndhH* *rps15*	5' AGCTCGTAACATTGGTCCCG 3' 5' TGGGGTTACCATTATCCTTTTTGT 3'	937
	rrn16 *rps15*	5' CTTACGCGTTACTCACCCGTC 3' 5' TTTTCGGGGTCTCAAAGGGG 3'	855
P. contracta	*rpl23* *psbA*	5' CCCCAATAACCGAATACTTTTGTC 3' 5' AGACCTGGCTGCTGTTGAAG 3'	1440
	IR *matK*	5' CGTCTTTCAAAAACGGCCTCT 3' 5' CCGAGGGCGAGTTTGGTATT 3'	1505

Table 1.3. (A) Locally collinear blocks (LCBs) and with gene content identified using progressiveMauve (Darling et al. 2010). Asterisks indicate partial copy of the gene. (B) Gene order for *Adenia* and *Passiflora* species compared to *Populus trichocarpa.* Minus sign (-) in front of LCBs indicates inversion. (C) Pairwise comparison of breakpoint and reversal distances estimated using CREx (Bernt et al. 2007). Species names are color-coded to indicate their generic and subgeneric placement (see Figure 1.2): Red (*Adenia*), Cyan (*Astrophea*), Orange (*Deidamioides*), Blue (*Passiflora*), Purple (*Tetrapathea*), and Green (*Decaloba*).

A

LCB	Genes
1	*psbA, trnH-GUG*
2	*trnK-UUU, matK, trnQ-UUG, psbK, psbI*
3	*trnS-GCU*, trnG-UCC, trnR-UCU*
4	*atpA, atpF, atpH*
5	*atpI, rps2, rpoC2, rpoC1, rpoB*
6	*trnC-GCA, petN, psbM, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, psbD, psbC, trnS-UGA, psbZ, trnG-GCC, trnfM-CAU, rps14, psaB, psaA, ycf3, trns-GCU**
7	*rps4, trnT-UGU, trnL-UAA, trnF-GAA, ndhJ, ndhK, ndhC,*
8	*trnM-CAU, trnV-UAC*
9	*atpB, atpE*
10	*rbcL*
11	*psbE, psbF, psbL, psbJ, petA, cemA, ycf4, psaI, accD*
12	*rpl20, rps18, rpl33, psaJ, trnP-UGG, trnW-CCA, petG, petL*
13	*clpP, rps12-5'*
14	*psbB, psbT, psbN, psbH, petB, petD*
15	*rpoA, rpls11, rpl36, rps8, rpl14, rpl16, rps3, rpl22, rps19, rpl2, rpl23, trnI-CAU, ycf2, trnL-CAA, ndhB, rps7, rps12-3', trnV-GAC, 16SrRNA, trnI-GAU, trnA-UGX, 23SrRNA, 4.5SrRNA, 5SrRNA, trnR-ACG, trnN-GUU, ycf1**
16	*ndhF, rpl32, trnL-UAG, ccsA, ndhD, psaC, ndhE, ndhG, ndhI, ndhA, ndhH*

B

Populus trichocarpa	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
A.mannii	-1, 2, -3, -4, -5, -6, 7, -9, -8, 10, 11, 12, 13, 14, 15, 16
P.pittieri/P.arbelaezii	-1, 2, 3, 4, 5, 6, 7, -9, -8, 10, -11, -12, -13, 14, 15, 16
P.foetida	-1, 2, 3, 4, 5, 6, 7, -9, -8, 10, 11, 12, 13, 14, 15, 16
P. menispermifolia	-1, 2, 3, 4, 5, -13, -12, -11, -10, 8, 9, -7, -6, 14, 15, -16
P.actinia + 10 others	-1, 2, 3, 4, 5, -13, -12, -11, -10, 8, 9, -7, -6, 14, 15, 16
P.contracta	-1, 2, 3, 4, 5, 6, 7, -9, -8, 10, 11, 12, 13, 14, 15, 16
P.tetrandra	-1, 2, 3, 4, 5, 6, 7, -9, -8, 10, -13, -12, -11, 14, 15, 16
P.microstipula	-1, 2, 3, 4, -5, 6, 7, -9, -8, 10, 11, 12, 13, 14, 15, 16
P.obovata	-1, 2, 3, 4, 5, 6, 7, -9, -8, 10, 11, 12, 13, 14, 15, 16
P.auriculata/P. jatunsachensis/P. rufa	-1, 2, 3, 4, 5, 6, 7, -12, -11, -10, 8 ,9, 13, 14, 15, 16
P.tenuiloba/P. suberosa	-1, 2, -6, -5, -4, 12, -7, 3, -11, -10, 8, 9, 13, 14, 15, 16
P. lutea/P. filipes/P. biflora/P. affinis	-1, 2, -6, -5, -4, -3, 7, -12, -11, -10, 8, 9,13, 14, 15, 16
P.misera	-14, -13, -9, -1, 2, -6, -5, -4, -3, 7, 12, -11, -10, 8, 15, 16

C

Species	*P. actinia*	*P. menispermifolia*	*P. foetida*	*P. auriculata*	*P. tenuiloba*	*P. lutea*	*P. misera*	*P. microstipula*	*P. pittieri*	*P. contracta*	*P. obovata*	*P. tetrandra*	*A. mannii*	*Populus*
P.actinia + 11 others	0/0													
P. menispermifolia	2/1	0/0												
P. foetida	2/1	4/2	0/0											
P. auriculata/P. jatunsachensis/P. rufa	4/2	6/3	2/1	0/0										
P. tenuiloba/P. suberosa	8/4	10/5	6/3	4/2	0/0									
P. lutea/P. filipes/P. biflora/P. affinis	6/3	8/4	4/2	2/1	2/1	0/0								
P. misera	9/5	10/6	7/4	5/3	5/3	3/2	0/0							
P. microstipula	3/2	5/3	2/1	4/2	8/4	6/3	9/5	0/0						
P. pittieri/P. arbelaezii	6/4	5/3	5/3	7/4	11/6	9/5	11/7	7/4	0/0					
P. contracta	5/3	6/4	3/2	5/3	9/5	7/4	9/6	5/3	7/8	0/0				
P. obovata	5/3	4/2	3/2	5/3	9/5	7/4	9/6	5/3	2/1	5/4	0/0			
P. tetrandra	3/2	5/3	2/1	4/2	8/4	6/3	9/5	4/2	3/2	5/3	5/3	0/0		
A. mannii	4/2	6/3	2/1	4/2	4/2	2/1	5/3	4/2	7/4	5/3	5/3	4/2	0/0	
Populus trichocarpa	6/3	7/4	4/2	5/3	9/5	7/4	9/6	6/3	8/5	6/4	6/4	6/3	6/3	0/0

Table 1.4. Genes, models and commands used in the phylogenetic analyses. (A) Sixty-eight plastid protein-coding genes used for phylogenetic analysis. (B) Best-fit evolutionary model and partition schemes determined in IQ-TREE (C) Commands used in IQ-TREE for partition schemes, tree search and bootstrap support for phylogenetic analysis.

A

Functional Group	Genes
Photosynthesis	*atpA, atpB, atpE, atpF, atpH, atpI, 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*
Ribosomal subunits	*rpl2, rpl14, rpl16, rpl23, rpl33, rpl36, rps2, rps3, rps4, rps8, rps11, rps12, rps14, rps15, rps19*
RNA polymerases	*rpoB, rpoC1, rpoC2*
Others	*clpP, cemA, ccsA, matK, ycf3, ycf4*

B

Data Set	Substitution Model
atpA+atpB+atp+petA+rpoC2+rps2+atpH+psaA+psaB+psbC+psbN+ccsA+ndhF+ndhC+petL+psbI+psbJ+psbK	TVM+F+R3
atpE	TIM2+F+I
atpI+ndhH+ndhJ+psaJ+rps3	TVM+F+G4
cemA+psaI+rps15+ycf4	GTR+F+G4
clpP	GTR+F+I+G4
ndhA+ndhD+ndhE+ndhG	TVM+F+I+G4
ndhB+rpl33+rps8	TVM+F+I
ndhI	GTR+F+R3
ndhK+psbT+rpl16+rpl36	GTR+F+R2
petB+psbB+rbcL+rps12+rps14+rpl2+ycf3	TVM+F+R2
petD+psaC+psbH	TN+F+I
petG+psbZ	TPM2u+F+I
petN+psbE	TPM3u+F
psbA	TN+F+R2
psbD	TPM3u+F+I
psbF+psbL+psbM	K3Pu+F+I
rpl14+rpoB+rpoC1+rps4	TVM+F+R4
rpl23	TIM+F+R2
rps11	TIM3+F+G4
rps19	K3Pu+F+R2

C

Partition schemes	iqtree -s 68genes.phy -st DNA -spp passiflora.parts.nex -nt AUTO -m TESTMERGEONLY -AICc -seed 122755 -pre AICc.models.parts
Tree search	iqtree -s 64genes.phy -st DNA -spp AICc.models.parts.best_scheme.nex -nt 4 -ninit 2000 -ntop 400 -nbest 100 -nstop 200 -bb 2000 -alrt 2000 -lbp 2000 -abayes -pre search
Bootstrap	iqtree -s 68genes.phy -st DNA -spp AICc.models.parts.best_scheme.nex -nt 4 -nstop 200 -b 100 -pre boots

Table 1.5. Phylogenetic informativeness for 68 plastid protein-coding genes. Included are gene length, number of sites for which substitution rates were calculated (#Rates), per-site and net informativeness. Ten genes with the most informativeness are highlighted in bold.

Loci	Length	#Rates	Per site	Net	Standard Deviation
rpoC2	**4414**	**4275**	**0.00223**	**9.52266**	**113.8**
rpoB	**3279**	**3265**	**0.00194**	**6.35037**	**5**
ndhF	**2328**	**2257**	**0.002**	**4.52024**	**137.6**
matK	**1551**	**1530**	**0.00249**	**3.81496**	**4.5**
rpoC1	**2228**	**2170**	**0.00175**	**3.7969**	**3.3**
ndhD	**1500**	**1500**	**0.00179**	**2.69043**	**2.5**
ycf4	**730**	**594**	**0.00407**	**2.41462**	**9.4**
atpA	**1533**	**1533**	**0.00141**	**2.33503**	**3.8**
ccsA	**1011**	**999**	**0.00215**	**2.25122**	**2.9**
rbcL	**1425**	**1425**	**0.00147**	**2.15493**	**2.7**
psaA	2250	2250	0.00104	2.15181	1.5
psaB	2202	2202	0.00102	2.08883	1.5
atpB	1494	1494	0.00132	1.96787	2
ndhA	1101	1089	0.00174	1.89749	2.5
ndhH	1197	1191	0.00155	1.85018	2.6
rps2	780	757	0.00226	1.70732	2.6
cemA	690	690	0.00233	1.60791	2.5
psbB	1524	1524	0.00104	1.58378	1.6
atpF	552	552	0.00285	1.57227	3.2
rps3	741	714	0.00218	1.55899	3.5
petA	960	960	0.00156	1.49336	2
psbC	1419	1419	0.00094	1.33275	1.7
rps11	447	441	0.00275	1.21446	4.5
ndhG	531	531	0.00221	1.17131	2.4
psbA	1059	1059	0.0011	1.16305	2.2
clpP	861	792	0.00146	1.15894	267.7
ndhK	714	708	0.00152	1.07499	1.9
rps4	684	681	0.00157	1.07155	3
rpl2	837	837	0.00125	1.04331	1.7
rps8	411	411	0.00239	0.98355	2.5
atpI	741	741	0.00123	0.91023	1.6
psbD	1059	1059	0.0008	0.8439	1.2
rpl16	477	477	0.00176	0.83898	3.4
ndhI	543	516	0.00151	0.78059	3.4
rpl14	366	366	0.00203	0.74444	3.1
petB	645	645	0.00111	0.71663	2
rps15	307	302	0.00235	0.71105	6.3
atpE	399	399	0.00175	0.69767	1.5
rps19	300	294	0.00222	0.65147	3.1
ndhE	303	303	0.00201	0.61017	1.8
ndhJ	477	474	0.00125	0.59023	1.3
ycf3	504	504	0.00115	0.57918	1.5
ndhC	360	360	0.00138	0.49822	2.5
rps14	300	300	0.00166	0.49803	2.7
petD	495	480	0.00098	0.47238	1.9
ndhB	1530	1530	0.0003	0.45538	0.6
rps12	378	375	0.00118	0.44358	2.1
rpl33	228	222	0.00165	0.36621	3.7
rpl23	429	341	0.00107	0.36475	13.1
psbK	183	183	0.00179	0.32826	2
rpl36	111	111	0.00255	0.28316	2.7
atpH	243	243	0.00104	0.25352	1.8
psbZ	186	186	0.00135	0.25081	1.2
psbH	219	219	0.00109	0.2387	2.1
psaI	117	117	0.002	0.2341	2
psbE	249	249	0.00093	0.23192	0.8
psaC	243	243	0.00092	0.22438	1.9
psbI	108	108	0.00206	0.22233	1.3
petL	93	93	0.00223	0.20784	2.7
psbN	129	129	0.00127	0.16424	1.2
psbJ	120	120	0.00136	0.16279	1.5
psaJ	132	132	0.00097	0.12786	2
petN	87	87	0.00127	0.11008	0.9
petG	111	111	0.00086	0.09597	1.2
psbT	105	105	0.00085	0.08915	2.1
psbL	114	114	0.00076	0.08685	1.2
psbM	102	102	0.0007	0.07093	1.3
psbF	117	117	0.00052	0.06125	1

Table 1.6. Plastome features of *Passiflora* and *Adenia* species. Abbreviations: LSC, large single copy; SSC, small single copy; IR, inverted repeat; FG, French Guiana; S, Suriname; bp, basepair.

Subgenus	Species	Plastome Size (bp)	LSC (bp)	SSC (bp)	IR (bp)	Number of genes	Protein coding genes	tRNA	rRNA	Intergenic region (%)	Protein coding GC (%)	Intergenic GC (%)	Gene density
Passiflora	*P. foetida* L.	162,266	84,635	13,507	32,062	108	74	30	4	30.0	37.4	30.8	0.83
	P. menispermifolia Kunth.	133,682	88,369	24,873	10,220	106	72	30	4	36.8	38.4	30.9	0.86
	P. actinia Hook.	146,255	85,389	13,492	23,687	107	73	30	4	32.2	38.6	30.9	0.85
	P. edulis Sims.	151,286	85,598	13,378	26,155	107	73	30	4	31.5	38.7	30.7	0.82
	P. laurifolia L.	151,422	85,411	13,511	26,250	107	73	30	4	31.5	38.6	30.8	0.82
	P. ligularis A. Juss.	150,827	85,471	13,504	25,926	107	73	30	4	31.4	38.4	30.6	0.82
	P. nitida Kunth.	151,400	85,573	13,479	26,174	107	73	30	4	31.3	38.7	30.6	0.82
	P. oerstedii Mast.	147,073	86,723	13,268	23,541	107	73	30	4	33.2	38.7	30.9	0.84
	P. quadrangularis L.	148,106	85,832	13,494	24,390	107	73	30	4	32.4	38.7	30.6	0.84
	P. retipetala Mast.	146,678	86,154	13,326	23,599	107	73	30	4	33.7	38.7	31.1	0.85
	P. serratifolia L.	143,111	86,196	13,317	21,799	107	73	30	4	33.1	38.5	30.3	0.87
	P. serratodigitata L.	151,509	85,478	13,521	26,255	107	73	30	4	31.9	38.6	30.6	0.82
	P. vitifolia Kunth.	143,845	85,906	13,483	22,228	107	73	30	4	33.6	38.7	30.9	0.86
Decaloba	*P. affinis* Engelm.	139,005	72,281	12,828	26,948	104	70	30	4	31.5	38.2	31	0.93
	P. filipes Benth.	138,086	75,648	13,318	24,560	104	70	30	4	33.4	38.2	30.1	0.92
	P. jatunsachensis Schwerdtfeger	159,860	57,977	13,349	44,267	104	70	30	4	33.0	38.4	30.9	0.90
	P. lutea L.	153,282	55,548	13,252	42,241	105	71	30	4	32.5	38.6	30.9	0.95
	P. microstipula Gilbert & MacDougal	164,672	86,601	13,449	32,311	108	74	30	4	30.6	37.4	31.3	0.82
	P. misera Kunth.	136,455	73,771	13,590	24,547	104	70	30	4	33.2	38.1	31.1	0.93
	P. rufa Feuillet & MacDougal	159,409	58,725	13,292	43,696	104	70	30	4	32.8	38.4	31.5	0.90
	P. suberosa L.	156,807	56,350	13,059	43,699	103	69	30	4	36.3	38.6	32.2	0.91
	P. tenuiloba Engelm.	159,912	58,440	13,226	44,123	103	69	30	4	36.3	38.5	31.7	0.89
	P. biflora Lam.	139,263	72,411	13,290	26,781	105	71	30	4	31.7	38.5	31	0.93
	P. auriculata Kunth. (FG)	161,101	54,593	12,246	47,131	105	71	30	4	30.7	38.9	31.2	0.91
	P. auriculata Kunth. (S)	161,383	54,736	12,229	47,209	105	71	30	4	30.4	38.9	31	0.90
Deidamiodies	*P. arbelaezii* L. Uribe	170,568	87,452	12,764	35,176	109	75	30	4	35.4	38.1	31.4	0.81
	P. contracta Vitta.	166,766	87,597	13,723	32,723	107	73	30	4	32.1	37.3	31.3	0.80
	P. obovata Killip	151,701	84,697	29,586	18,709	107	73	30	4	33.1	37.3	30.4	0.80
Tetrapathea	*P. tetrandra* Banks & Sol. ex DC.	159,223	86,474	13,537	29,606	109	75	30	4	30.0	37.1	30.3	0.84
Astrophea	*P. pittieri* Mast.	161,494	88,539	12,917	30,019	109	75	30	4	30.4	37.3	30.8	0.78
Adenia	*A. mannii* (Mast.) Engl.	165,364	87,176	13,780	32,204	109	75	30	4	30.1	37.3	30.8	0.82

Table 1.7. Variation in pairwise nucleotide and amino acid sequence identity of *rpoA* in Passifloraceae compared to *Populus trichocarpa*. Results of conserved domain database search are presented as yes or no for presence or absence, respectively, of all conserved domains that includes homodimer interface, beta and beta primer interaction sites. Species without any conserved domains in *rpoA* are highlighted in red. bp, basepair; nt, nucleotide; aa, amino acid.

Subgenus	Species	nt length (bp)	nt identity (%)	aa identity (%)	Conserved domains
Populus	*P. trichocarpa*	1020	100.0	100.0	Yes
Adenia	*Adenia mannii*	1026	92.5	88.0	Yes
Passiflora	*P. menispermifolia*	1002	91.0	86.7	Yes
	P. foetida	1020	93.5	89.7	Yes
	P. actinia	1020	93.5	89.7	Yes
	P. edulis	1020	93.0	88.8	Yes
	P. laurifolia	1020	93.3	89.1	Yes
	P. ligularis	1020	93.0	88.2	Yes
	P. nitida	1020	93.2	89.1	Yes
	P. oerstedii	1020	93.1	89.1	Yes
	P. quadrangularis	1020	93.3	89.1	Yes
	P. retipetala	1020	93.3	89.4	Yes
	P. serratifolia	1020	93.3	89.4	Yes
	P. serratodigitata	1020	93.3	88.8	Yes
	P. vitifolia	1020	93.4	89.1	Yes
	P. cincinnata	1020	93.1	88.5	Yes
Astrophea	*P. pittieri*	1020	93.3	89.1	Yes
Decaloba	*P. microstipula*	1020	93.1	89.4	Yes
	P. auriculata	945	73.8	59.8	Yes
	P. jatunsachensis	1047	72.7	57.3	Yes
	P. rufa	966	74.1	58.9	Yes
	P. lutea	1098	53.1	38.2	Yes
	P. filipes	1098	53.3	37.9	Yes
	P. misera	1119	55.1	35.8	Yes
	P. affinis	1026	53.7	37.2	Yes
	P. biflora	1026	53.7	36.9	Yes
	P. suberosa	**1041**	**39.0**	**24.9**	**No**
	P. tenuiloba	**1041**	**39.0**	**24.9**	**No**
Deidamioides	*P. arbelaezii*	1035	91.1	86.0	Yes
	P. cirrhifolia	1020	93.8	90.6	Yes
	P. contracta	1062	88.3	82.7	Yes
	P. obovata	1056	89.1	84.3	Yes
Tetrapathea	*P. tetrandra*	1020	94.1	91.2	Yes

Table 1.8-1.11 are provided in the link**, https://github.com/bshrestha0/Substitution-rates-analyses-results** as well as in supplemental files.

Table 1.12. Summary of branch specific log-likelihood test (LRT) results for substitution rates and *dN*/*dS* using PAML (Yang 2007) and HyPhy (Pond et al. 2005), respectively. Genes with branch specific accelerated and decelerated rates and ratio are red and blue, respectively. False discovery rate (FDR) was used to adjust p-values for the multiple comparisons. IR, inverted repeats.

Genes	Rate/ Ratio	Null model		Alternative model			LRT statistic	p-value	FDR p-value
		rate/ratio	lnL	Other Passifloraceae	IR expansion clade	lnL			
ccsA	*dN*	0.0146	-4475.66	0.0156	0.0088	-4471.222	8.88	0.0029	0.0187
clpP	*dS*	0.0576	-9378.77	0.0616	0.0358	-9375.2	7.15	0.0075	0.0393
clpP	*dN*	0.0878	-9378.77	0.0832	0.1125	-9370.691	16.17	0.0001	0.0007
ndhD	*dS*	0.0329	-5528.69	0.0355	0.0189	-5523.009	11.37	0.0007	0.0073
ndhE	*dS*	0.0280	-932.36	0.0326	0.0035	-927.832	9.05	0.0026	0.0187
petD	*dN*	0.0020	-1278.63	0.0012	0.0066	-1273.389	10.49	0.0012	0.0109
psaA	*dS*	0.0195	-5318.58	0.0210	0.0113	-5313.624	9.91	0.0016	0.0124
psaB	*dS*	0.0203	-5076.73	0.0218	0.0119	-5072.292	8.88	0.0029	0.0187
psaJ	*dS*	0.0268	-373.85	0.0317	0.0000	-369.849	8.00	0.0047	0.0276
psbB	*dS*	0.0228	-3782.60	0.0246	0.0126	-3778.349	8.50	0.0036	0.0220
psbD	*dS*	0.0164	-2200.02	0.0183	0.0059	-2194.843	10.35	0.0013	0.0110
psbZ	*dS*	0.0261	-414.43	0.0310	0.0000	-410.647	7.57	0.0060	0.0337
rbcL	*dS*	0.0323	-5206.67	0.0269	0.0629	-5185.984	41.37	0.0000	0.0000
rpl2	*dN*	0.0049	-2208.13	0.0039	0.0100	-2200.446	15.37	0.0001	0.0009
rpl23	*dN*	0.0176	-1851.84	0.0112	0.0523	-1820.728	62.22	0.0000	0.0000
rpoB	*dN*	0.0112	-12377.35	0.0077	0.0307	-12229.67	295.37	0.0000	0.0000
rpoC1	*dN*	0.0148	-8859.53	0.0103	0.0400	-8745.25	228.57	0.0000	0.0000
rpoC2	*dN*	0.0159	-18865.92	0.0133	0.0309	-18784.1	163.62	0.0000	0.0000
rps2	*dN*	0.0154	-3260.20	0.0113	0.0396	-3222.841	74.71	0.0000	0.0000
rps11	*dS*	0.0328	-2904.57	0.0238	0.0823	-2889.119	30.90	0.0000	0.0000
rps11	*dN*	0.0286	-2904.57	0.0247	0.0498	-2893.371	22.39	0.0000	0.0000
rps14	*dN*	0.0143	-1106.43	0.0119	0.0278	-1101.343	10.18	0.0014	0.0114
rps15	*dN*	0.0391	-2008.23	0.0324	0.0784	-1995.574	25.32	0.0000	0.0000
ycf3	*dN*	0.0200	-1171.18	0.0014	0.0059	-1167.535	7.29	0.0069	0.0377
ycf4	*dN*	0.0426	-5191.60	0.0384	0.0665	-5179.547	24.11	0.0000	0.0000
psaB	*dN*/*dS*	0.0512	-4839.89	0.0347	0.0970	-4835.22	9.34	0.0022	0.0254
rbcL	*dN*/*dS*	0.1339	-4921.58	0.2116	0.0547	-4901.70	39.77	0.0000	0.0000
rpoB	*dN*/*dS*	0.5169	-11643.34	0.3119	1.1024	-11599.32	88.05	0.0000	0.0000
rpoC1	*dN*/*dS*	0.6113	-8312.63	0.3624	1.2813	-8281.68	61.91	0.0000	0.0000
rpoC2	*dN*/*dS*	0.6084	-17767.49	0.4700	1.0695	-17740.25	54.49	0.0000	0.0000
rps2	*dN*/*dS*	0.6490	-3023.86	0.4676	1.1640	-3018.12	11.48	0.0007	0.0096

Chapter Two

Evolutionary fate of missing or divergent plastid genes in *Passiflora*[2]

2.1. Introduction

The origin of plastids is attributed to primary endosymbiosis in which a eukaryote engulfed a cyanobacterium that initially retained its genome. Subsequent relocation of genes to the host nucleus resulted in a highly reduced endosymbiont or plastid genome (plastome) (Timmis et al. 2004). Accordingly, the genome size (~1.4 to 9.1 Mb) and number of protein coding genes (~1000 to 8000) of cyanobacteria (Larsson et al. 2011) are substantially larger than land plant plastomes (~100 to 200 kb, ~120 to 130 genes; Raubeson and Jansen 2005; Bock 2007). Most land plant plastomes have a highly conserved quadripartite structure that contains protein coding genes involved in photosynthesis or gene expression along with ~30 transfer RNA and 4 ribosomal RNA genes (Bock 2007).

DNA transfer from the plastid to the nucleus is an ongoing process (Martin 2002; Huang et al. 2003; Stegemann et al. 2003). Studies of plastid DNA transfer in angiosperms have shown size variation from small fragments <100 bp to several kb (Matsuo et al. 2005; Yoshida et al. 2014) to entire plastomes in *Oryza sativa* (Matsuo et al. 2005) and *Populus*

[2] This chapter contains published manuscript: Shrestha B, Gilbert LE, Ruhlman TA, Jansen RK. 2020. Rampant nuclear transfer and substitutions of plastid genes in *Passiflora*. Genome Biol. Evol. https://doi.org/10.1093/gbe/evaa123. Bikash Shrestha performed all the experiments, conducted data analyses and wrote the manuscript.

trichocarpa (Salicaceae; Huang et al. 2017). Together these findings suggest that plastid DNA transfers to nucleus are not uncommon. Despite frequent DNA transfers to nucleus, only a few functional plastid gene transfers have been confirmed. Functional transfers require the acquisition of elements for nuclear expression along with targeting peptides (Bruce 2000), which are essential for plastid localization. Mechanisms for the acquisition of N-terminal signal sequences are better understood for mitochondrial genes transferred to the nucleus (Adams and Palmer 2003). A common acquisition mechanism is insertion of an organelle gene into a duplicate copy of a pre-existing nuclear-encoded organelle-targeted gene mediated by exon shuffling, which has been documented for the mitochondrial gene *rps11* in *Oryza* (Kadowaki et al. 1996). Similarly, transfer of plastid *rpl32* involved the gain of a transit peptide by integration into a duplicated copy of nuclear-encoded plastid-targeted Cu-Zn superoxide dismutase in *Populus* (Ueda et al. 2007). Although transit peptides for most transferred plastid genes have been identified, little is known about their origin.

At least four plastid genes are known to have functional transfers to nucleus in the land plants. Among these, *rpoA*, which encodes the α-subunit of the plastid RNA polymerase (PEP), was transferred in mosses (Sugiura et al. 2003; Goffinet et al. 2005). Within angiosperms, *infA*, which encodes translation initiation factor IF-1, has undergone multiple independent transfers to the nucleus (Millen et al. 2001). Similarly, at least two independent transfers of *rpl22*, in Fabaceae and Fagaceae, have been reported (Gantt et al. 1991; Jansen et al. 2011) and third putative transfer in *Passiflora* was suggested (Jansen et al. 2011). Likewise, independent transfers of *rpl32* have been reported in Rhizophoraceae,

Salicaceae and Ranunculaceae (Cusack and Wolfe 2007; Ueda et al. 2007; Park et al. 2015). An alternative to the functional transfer of plastid genes to the nucleus is replacement of function by a nuclear gene, such as the substitution of nuclear-encoded mitochondrial *rps16* gene in *Medicago truncatula* and *Populus alba* plastomes (Ueda et al. 2008), *accD* in grasses (Konishi et al. 1996) and *rpl23* in spinach and *Geranium* (Bubunenko et al. 1994; Weng et al. 2016).

Evolutionary studies based on 31 sequenced *Passiflora* plastomes reported the loss of several essential genes that encode large or small ribosomal subunits (*rpl20, rpl22, rpl32, rps7* and *rps16*) as well as the two largest plastid genes, *ycf1* and *ycf2* (Cauz-Santos et al. 2017; Rabah et al. 2019; Shrestha et al. 2019). The function of the proteins encoded by the latter two genes has been long debated but recent findings suggested that YCF1 is an essential component of the primary translocon complex of the plastid inner envelope membrane (Kikuchi et al. 2013) and YCF2 is a component of the associated ATPase motor protein (Kikuchi et al. 2018). Patterns of gene loss or pseudogenization in *Passiflora* plastomes are quite unusual. All species have lost *rpl22* and *rps16* completely. However, the phylogenetic distribution of gene losses for *rpl20, rpl32*, *rps7, ycf1* and *ycf2* suggested multiple independent losses within the genus (Shrestha et al. 2019). The pattern for two genes, *rpl20* and *rps7,* is highly variable with some species having only remnants of the gene, while others contain pseudogenes with premature stop codon(s) or complete sequences with conserved domains (Rabah et al. 2019; Shrestha et al. 2019). In addition, a highly divergent *rpoA* was reported to be non-functional due to very low sequence identity and the lack of conserved domains, although an earlier study that included only four species

of *Passiflora* (Blazier et al. 2016a) suggested that this gene may still be functional. To understand the evolutionary fate of missing plastid genes in *Passiflora*, transcriptome data were gathered for at least one species from each of the four subgenera *Passiflora, Decaloba, Astrophea* and *Deidamioides.* The results indicate that plastomes in the genus have followed a diverse trajectory represented by extensive gene transfers and/or substitutions with several novel events among angiosperms.

2.2. Materials and Methods

Plant material and RNA isolation

Plant sampling for RNA isolation included six species from *Passiflora* (*P.*), *P. pittieri, P. contracta* and *P. oerstedii* from the three subgenera *Astrophea, Deidamioides* and *Passiflora*, respectively, and three species, *P. tenuiloba, P. auriculata* and *P. biflora,* from subgenus *Decaloba.* Young leaves were flash frozen in liquid nitrogen from field-collected populations grown in greenhouses at The University of Texas at Austin. Total RNA isolation was carried out using RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Denaturing gel electrophoresis and NanoDrop® (ND-1000, ThermoScientific) were used for qualitative and quantitative assessment of RNA.

Transcriptome sequencing and assembly

Library preparation and transcriptome sequencing was performed at Beijing Genomics Institute on BGISEQ-500 platform or at UT-Austin Genome Sequencing and Analysis Facility on Illumina HiSeq 4000 platform (Illumina, San Diego, CA). Ribosomal

RNA (rRNA) was removed using Ribo-Zero rRNA Removal Kit (Epicentre Biotechnologies, WI, USA) prior to sequencing.

Quality assessment of RNA reads was carried out using FastQC v.0.11.5 (Andrews 2010) prior to and after removal of rRNA. SortMeRNA v.2.1b (Kopylova et al. 2012) was employed for removal of rRNA by mapping against eight available rRNA databases (bacteria, archaea and eukarya). Transcriptome data for *P. biflora* contained low quality reads so a wrapper tool, Trim Galore v.0.4.4 (https://github.com/FelixKrueger/TrimGalore, last accessed June 5, 2019), was used to trim low quality reads. All transcriptome data were assembled *de novo* using Trinity v.2.8.4 (Grabherr et al. 2011). Three different methods were used to characterize the quality of transcriptome assembly: (i) read representation was assessed by mapping reads against the assembled transcriptome using Bowtie 2 v.3.4 (Langmead and Salzberg 2012); (ii) contig N50 was calculated; and (iii) completeness of the assembly was estimated by mapping against the single-copy orthologs database using Benchmarking Universal Single-copy Orthologs (BUSCO) (Waterhouse et al. 2018). Eudicots OrthoDB (*odb10*) was selected within BUSCO trinity-assembled transcript mapping. All computational analyses for transcriptome assembly including quality assessments were carried out at the Texas Advanced Computing Center (http://www.tacc.utexas.edu) at the University of Texas at Austin. The clean RNA reads for all six *Passiflora* species included in this study can be accessed via https://www.ncbi.nlm.nih.gov/sra/PRJNA634675.

Identification of genes

Two approaches were employed to identify genes of interest: (i) transcriptome data was aligned with UniProt protein database followed by functional annotation of aligned transcripts; and (ii) a protein database was created to identify genes of interest from a list of reference species and used as a query to map against the assembled transcriptome data. Both approaches are described in detail below.

Functional annotation of assembled transcripts: Prior to annotation, assembled transcripts were aligned to the Protein Knowledgebase (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz, last accessed, August 25, 2019). BLASTx was employed to align transcripts against the UniProt BLAST database with an e-value of 1 e^{-4}. The results of BLASTx were processed to extract coding sequences using scripts available at https://github.com/z0on/annotatingTranscriptomes (last accesses, September 5, 2019). Coding sequences aligned with the UniProt database were extracted using script "CDS_extractor_v2.pl" and subsequently filtered to extract the single best hit by removing isoforms for each gene using the script "fasta2BH.pl". The output generated a file that contained protein coding sequence in Multi-FASTA format, which was used for functional annotation on online server eggNOG-mapper v.4.5.1 (Huerta-Cepas et al. 2016) under default parameter settings.

Mining orthologous genes using reference species protein database: Plastid-encoded proteins sequences were obtained from the completed plastomes of reference species available at NCBI including *Arabidopsis* (*A.*) *thaliana* (NC_000932.1), *Nicotiana tabacum* (NC_001879.2), *Vitis vinifera* (NC_007957.1), *Salix purpurea* (KP019639.1) and *Populus*

trichocarpa (NC_009143.1). Genes of interest were extracted, translated and aligned using MUSCLE (Edgar 2004) in Geneious v.11.0.5 (https://www.geneious.com, last accessed, January 28, 2018). Similarly, for the nuclear-encoded proteins in *Arabidopsis thaliana* and *Populus trichocarpa*, sequences were downloaded from The Arabidopsis Information Resource (TAIR, https://www.arabidopsis.org/, last accessed, February 10, 2020) and Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html#, last accessed, February 10, 2020), respectively. Orthologous genes in *Passiflora* were identified using tBLASTn with the reference Multi-FASTA protein sequences against transcriptome database with parameters "tblastn –evalue $1e^{-3}$ –outfmt 7 –max_target_seqs 1 –out tblastn.out –num_threads 12". Open reading frames (ORFs) were identified using Geneious and web BLAST (BLASTn and BLASTp, https://blast.ncbi.nlm.nih.gov/Blast.cgi, last accessed, March 5, 2020) was used to identify similar sequences in the NCBI database.

Prediction of transit peptides

Three online software programs, TargetP-2.0 (Armenteros et al. 2019; http://www.cbs.dtu.dk/services/TargetP/, last accessed March 5, 2020), LOCALIZER (Sperschneider et al. 2017; http://localizer.csiro.au/, last accessed, March 5, 2020) and Predotar (Small et al. 2004; https://urgi.versailles.inra.fr/predotar/, last accessed, March 5, 2020), were used to predict putative transit peptides for nuclear transferred genes. The ORFs identified in the transcript of interest were translated in Geneious and used for the prediction under default settings.

Phylogenetic analysis of nuclear-encoded rpl20

Phylogenetic relationships among nuclear-encoded *rpl20* sequences in *Passiflora* were inferred by maximum likelihood (ML) using IQ-TREE v.1.5.2 (Nguyen et al. 2015). The translated amino acid alignment for the analysis included RPL20-1 and RPL20-2 from the six *Passiflora* species, nuclear-encoded mitochondrial targeted RPL20 from *A. thaliana* (AT1G16740.1) and *Populus trichocarpa* (XM_006383341) and plastid-encoded RPL20 from *A. thaliana* (NP_051082) and *Populus trichocarpa* (ABO36728.1). The alignment also included 50S ribosomal protein L20 (RPLT) from two bacterial species, *Microcystis aeruginosa* (AP009552) and *Rickettsia prowazekii* (NZ_CP014865), which share an endosymbiotic ancestry with plastids and mitochondria, respectively and a thermophilic bacterium, *Thermotoga caldifontis* (NZ_AP014509), was used as an outgroup. Amino acid sequences were aligned using MUSCLE in Geneious. IQ-TREE v.1.5.2 (Nguyen et al. 2015) was used for evolutionary model selection, ML analyses and assessment of branch support by non-parametric bootstrapping using 100 pseudoreplicates.

Evolutionary rate analysis

Pairwise and branch-specific substitution rate analyses were performed for *rpoA* using PAML v.4.8 (Yang 2007). The nucleotide sequence alignment for *rpoA* included 11 species from subgenus *Decaloba*, two from subgenus *Deidamioides*, a single species each from subgenera *Passiflora* and *Astrophea*, and a species of *Adenia* as an outgroup (Table 2.1). Translational alignment was carried out using MAFFT (Katoh and Standley 2013) in Geneious. For both pairwise and branch-specific analyses, codon frequencies were

estimated using F3 x 4 model and transition/transversion ratio and omega (*dN/dS*) were estimated with default setting of 2 and 0.4, respectively. Parameters for pairwise estimation in the CODEML control file included runmode = -2, model = 0 and cleandata = 0 for treating alignment gaps as ambiguous data. Branch-specific synonymous (*dS*) and non-synonymous (*dN*) rates and *dN*/*dS* ratio were calculated using free-ratio model, where each branch was allowed to have its own *dN/dS* value, and global-ratio model with single *dN/dS* value for the entire tree. Parameters for the free-ratio model included model =1, runmode = 0 and a maximum likelihood (ML) tree generated using 68 plastid genes (Shrestha et al. 2019) was used as the constraint tree. Similarly, parameters for the global-ratio model included model= 0 and runmode = 0 with the 68 plastid genes ML tree used as the constraint tree. For the branches with *dN/dS* ratio >1, a two-ratio model (model =2) was used, where the branch with *dN/dS* > 1 was allowed a different *dN/dS* value from rest of the tree and likelihood ratio tests (LRTs) were performed to verify the significant differences. False discovery rate correction was used in R v3.5.1 (R core Team 2013) to correct for the multiple comparisons in estimating significant differences in *dN/dS.*

Validation of intron in the nuclear *rps7* and *rpl20*

Introns in nuclear-encoded *rps7* and *rpl20* were validated with PCR amplification. Genomic DNAs were isolated for six *Passiflora* species using NucleoSpin® plant II DNA extraction kit (MACHEREY-NAGEL, Düren, Germany). The nuclear transcripts of *rps7* and *rpl20* were aligned to design primers with Primer3 (Untergasser et al. 2012) in Geneious. The primers used to amplify the target regions in nuclear *rps7* and *rpl20* are

provided in Table 2.2. Products were amplified with TaKaRa PrimeSTAR® GXL DNA polymerase (TaKaRa Bio, Shiga, Japan) with the following parameters: 1 min at 98°C, followed by 32 cycles of 10 s at 98°C, 15 s at 60°C, and 1 min or 2 min at 68°C, and final extension of 5 min at 68°C. The intron sequences were determined with a combination of Sanger sequencing and mapping of high-throughput DNA reads available for *P. pittieri, P. contracta, P. oerstedii, P. tenuiloba* and *P. auriculata* (Rabah et al. 2019; Shrestha et al. 2019) with Bowtie 2 v.3.4 (Langmead and Salzberg 2012) and the Geneious mapper in Geneious.

2.3. Results

Transcriptome assembly and assessment

Transcriptome sequencing and assembly were carried out for six *Passiflora* species. Transcriptome data contained from 0.23% to 10.69% rRNA reads (Table 2.3), which were removed prior to assembly. *Passiflora biflora* transcriptome read quality was relatively poor compared to other species, hence, reads were trimmed to improve the quality prior to assembly resulting in read length variation ranging from 70 bp to 151 bp for this species. Quality assessment by mapping clean paired-end reads to the assembled transcriptome showed high read support values (> 98%) for all species except *P. biflora* (87.6%). Total number of assembled bases was slightly higher for *P. biflora* compared to other species, whereas mean contig length and N50 were similar for all species. The completeness of the transcriptome assembly was >92% for the 2121 single-copy orthologs searched. Comparison of basic evaluation metrics for the transcriptome assembly, such as total

assembled bases, mean contig length and N50 statistics based on single longest isoform per gene are shown in Table 2.3.

Fate of the missing plastid genes

A brief summary of the results of transcriptome analyses is provided Table 2.4. More detailed results for the each gene assessed in this study are described below. Sequence identity for each gene was compared among *Passiflora* species and with reference species. For each comparison nucleotide (nt) identities are reported first, followed by amino acid (aa) identities.

rps7

Nuclear transcripts for *rps7* that included predicted transit peptides were identified in all six species of *Passiflora* with nt and aa sequence identities >77% (Table 2.5). Two included species, *P. pittieri* and *P. contracta,* had intact *rps7* in their plastomes (Rabah et al. 2019; Shrestha et al. 2019). In these two species, the plastid and nuclear-encoded *rps7* had pairwise nt and aa identities of ~73% and ~62%. Nuclear-encoded RPS7 in *Passiflora* was ~218 aa long, which was ~60 amino acids longer than plastid-encoded RPS7 in *Arabidopsis thaliana* (155 aa) (Figure 2.1A).

TargetP and LOCALIZER predicted nuclear RPS7 was targeted to the plastid with high probabilities (Pr: 0.99-1.0) but the length of predicted transit peptides varied depending on the software (Table 2.6). Predotar also predicted the localization of nuclear RPS7 to the plastid but the probability varied from 0.73 to 0.99 among the species. TargetP

predicted 60 aa transit peptide that shared 76.3% identity in all *Passiflora* species, whereas, LOCALIZER predicted species-specific transit peptides of various lengths (Table 2.6). BLAST searches (BLASTn and BLASTp) against NCBI performed to identify the source of the transit peptide for the nuclear-encoded *rps7* did not find any significant match but the protein search identified a 37 - 45% match with Thioredoxin m-type 3 protein (TRX-m3) of *Populus alba* (TKS05236.1).

A tBLASTN search with the *Populus alba* TRX-m3 sequence along with eight isoforms of TRX-m from *Populus trichocarpa* (Chibani et al. 2009) as queries returned several isoforms of *trx-m* in each of the *Passiflora* species examined, with isoform 3 (*trx-m3*) as the best match. The *trx-m3* transcripts in *Passiflora* were ~513 nt (171 aa) long and had nt and aa identities of 90.4% and 87.1%. The *Passiflora* TRX-m3 consensus sequence shared 95.9% and 67.6% aa identity with *Populus alba* and *Populus trichocarpa* TRX-m3, respectively. Between TRX-m3 and nuclear-encoded RPS7 in *Passiflora*, nt and aa identities were ~60% and ~38% (Figure 2.1C). However, the transit peptides (60 aa) between TRX-m3 and nuclear RPS7 sequence had 100% pairwise identity in each *Passiflora* species (Figure 2.1C).

To confirm the transfer of plastid *rps7* into the intron of the nuclear gene *trx-m3*, the gene was amplified with two PCR reactions that shared a forward primer on the targeting sequence but employed unique reverse primers (Figure 2.2A). Amplification with the reverse primer in *rps7* produced a band of ~500 bp, whereas, the reverse primer in *trx-m3* amplified a band of ~2000 bp (Figure 2.2B). The presence of two introns, an ~400 bp intron that separated *rps7* from the targeting sequence and second intron of ~850 bp that separated

the *trx-m3* exon from *rps7* was verified with PCR and Sanger sequencing (Figure 2.2A-C). Mapping of Illumina DNA reads for the five *Passiflora* species *P. pittieri, P. contracta, P. oerstedii, P. tenuiloba* and *P. auriculata* also validated the presence of introns. Accession numbers for transcripts and genes associated with *rps7*, *trx-m3,* and chimeric *rps7-trx-m3* are provided in Table 2.7.

To gain insight into the timing of *rps7* nuclear transfer, the *Passiflora* nuclear RPS7 and TRX-m3 protein sequences were used as queries to identify transcripts in two Salicaceae genera, *Salix purpurea* in ONEKP project (db.cngb.org/onekp/) and *Populus trichocarpa* at NCBI. The tBLASTn search identified nuclear transcripts of *rps7* and *trx-m3* in both Salicaceae species (Table 2.7). The translated aa sequences for RPS7 and TRX-m3 of *S. purpurea* had overall pairwise identities of 35.1% and 76% for the transit peptide. Similarly, in *Populus trichocarpa* the aa pairwise identities between RPS7 and TRX-m3 was 38% for entire alignment and 82.5% for the transit peptide (Figure 2.3).

rpl22

Nuclear *rpl22* transcripts were identified in all *Passiflora* species and varied in length from 621 bp - 645 bp and had nt and aa identities >80% (Table 2.5). *Passiflora* nuclear *rpl22* had nt and aa sequence identities > 76% with *Arabidopsis thaliana* plastid *rpl22* (Table 2.5). Compared to the length of *Arabidopsis* plastid RPL22 protein, *Passiflora* RPL22 was 46-54 aa longer (Figure 2.4A). All three prediction software programs predicted N-terminal sequence in nuclear *rpl22* as a plastid transit peptide with high probabilities but with discordance in the length between the programs. Predicted transit

peptide lengths and probabilities are provided in Table 2.6. Due to the variation in predicted length, it was not possible to define the precise extent of the transit peptide. The alignment of *Passiflora* nuclear RPL22 with the *Arabidopsis* plastid RPL22 contained an overhang of 83-89 aa in the N-terminal region (Figure 2.4A). The overhang has 85.8% nt and 77.8% aa identities across *Passiflora* species and likely represents a transit peptide.

To examine the source of the transit peptide, *Passiflora* nuclear RPL22 sequences were aligned with nuclear RPL22 from three Fabaceae (*Pisum sativum, Medicago sativa* and *Glycine max*) and two Fagaceae (*Quercus rubra* and *Castanea mollisssima*). The alignment of the transit peptide (89 aa) had < 20% identity, whereas the remaining sequence had 60-70% identity and the entire alignment has ~48% aa identity (results not shown). BLAST searches against NCBI for nuclear *rpl22* resulted in a 70% nt and 45% aa match with a 164 aa organelle RNA recognition motif domain-containing protein 1 in *Populus trichocarpa* (ORRM1). Using *Populus* ORRM1 as a query, a 152 aa RNA binding protein in *Arabidopsis thaliana* (AT4G20030) with 53.2% identity was identified. tBLASTn searches of the *Passiflora* transcriptomes with the *Populus* and *Arabidopsis* ORRM sequences as queries identified putative ORRM transcripts that contained a RNA recognition motif and shared 76.6% aa identity. The alignment of *Populus trichocarpa* ORRM, the putative *Passiflora* ORRM and the *Passiflora* nuclear RPL22 showed that RPL22 contains a fragmented ORRM sequence within the RPL22 overhang sequence in the N-terminal region (Figure 2.4B). The fragmented ORRM sequence in *Passiflora* RPL22 shared 46.7% and 50% aa identity with the *Populus* and *Passiflora* ORRM, respectively.

Accession numbers for the sequence of *rpl22* transcripts and ORMM genes are provided in Table 2.8.

rpl32

Nuclear *rpl32* transcripts were identified in all *Passiflora* species that were substantially longer compared to plastid *rpl32* in *Arabidopsis thaliana* (~828 bp vs. 159 bp). *Passiflora* nuclear *rpl32* had nt and aa identities >88% (Table 2.5). Two examined *Passiflora* species, *P. oerstedii* and *P. tenuiloba*, had intact *rpl32* in their plastomes with conserved domains (Shrestha et al. 2019). Compared to the identified nuclear *rpl32,* plastid *rpl32* had nt and aa pairwise identities of 79.5 % and 75.9% in *P. oerstedii,* whereas, plastid *rpl32* in *P. tenuiloba* had the pairwise identities of 72.8% and 57.1%. All three software programs predicted that nuclear RPL32 in *Passiflora* was targeted to the plastid with high probabilities. TargetP predicted ~75 aa sequence at N-terminal region as a transit peptide for all *Passiflora* species but LOCALIZER predicted variable transit peptide lengths among species (Table 2.6). BLAST searches against NCBI for the source of transit peptide identified a significant match with a chloroplast targeted copper/zinc superoxide dismutase gene (cp *rpl32*) from several Malpighiales species. Plastid-targeted *rpl32* (cp *rpl32*) and cp *sod-1* for *Populus alba* (Ueda et al. 2007) were downloaded, translated and aligned with nuclear RPL32 in *Passiflora.* Copies of nuclear RPL32 in *Passiflora* were longer (~275 aa) than *Populus* cp RPL32 (183 aa) due to retention of additional pt *sod-1* exons in *Passiflora* (Figure 2.5). The entire alignment had ~50% aa identity but the identity increased to ~64%

for the transit peptide and ~73% for the RPL32 sequence at the C-terminus. Accession numbers for the *Passiflora* nuclear *rpl32* transcripts are provided in Table 2.9.

rpl20

Transcriptome mining for nuclear *rpl20* identified two distinct nuclear-encoded *rpl20* sequences (*rpl20-1* and *rpl20-2*) in the six species of *Passiflora.* Alignment of the 12 transcripts had nt and aa identities of 43.7% and 68%, respectively but identity within each transcript type was much higher. The nt and aa identities for *rpl20-1* were >92%, whereas *rpl20-2* were >76% (Table 2.5). The *rpl20-1* transcripts were slightly longer than *rpl20-2* (~375 bp vs. ~360 bp). Compared to *Arabidopsis thaliana* plastid *rpl20, Passiflora rpl20-1* and *rpl20-2* had nt and aa identities <40% (Table 2.5). Nuclear-encoded mitochondrial targeted *rpl20* from *Arabidopsis thaliana* (Bonen and Calixte 2005) was used to identify orthologs in *Populus trichocarpa* in NCBI. *Populus trichocarpa* mitochondrial *rpl20* is located on chromosome 17 (NC_037301.1) and shared 87% aa identity with *Arabidopsis.* The mitochondrial *rpl20* in *Populus* had a substantially longer intron compared to *Arabidopsis* (1657 bp vs. 797 bp). Compared to the *Arabidopsis* mitochondrial *rpl20, Passiflora rpl20-1* had nt and aa identities >76% but slightly lower for *rpl20-2* (<56%, table 3). Likewise, compared to *Populus* mitochondrial *rpl20,* nt and aa identities were higher (~84% and ~92%) for *rpl20-1* but lower for *rpl20-2* (~64% and ~48%). TargetP failed to predict subcellular targeting sequences for RPL20-1 and RPL20-2 and LOCALIZER predicted plastid transit peptides for RPL20-2 for the three species, *P. oerstedii*, *P, auriculata* and *P. biflora*, respectively but failed to predict targeting sequence

for RPL20-1. In contrast, Predotar strongly predicted localization of RPL20-1 to mitochondria and RPL20-2 to plastids in all six *Passiflora* species (Table 2.6). Phylogenetic analysis of nuclear RPL20 strongly supported the placement of *Passiflora* RPL20-1 in a clade with nuclear-encoded mitochondrial-targeted RPL20 of *Arabidopsis thaliana* and *Populus trichocarpa* (Figure 2.6). The *Passiflora* RPL20-2 formed a clade sister to RPL20-1 and together as a clade sister to the α-proteobacterium species (Figure 2.6). BLAST searches against NBCI for *rpl20-1* resulted in ~80% nt and ~90% aa matches to 50S ribosomal protein L20 for several angiosperm lineages including two families of Malpighiales, Euphorbiaceae and Salicaceae. Conserved domain searches of RPL20-1 predicted binding sites for 23S rRNA, and RPL13 and RPL21 proteins (Figure 2.7A). BLAST searches for *rpl20-2* generated similar results to *rpl20-1* but with slightly lower sequence identities, ~73% nt and 55-65% aa identities with 50S ribosomal protein L20 and binding sites for RPL13 and RPL21 (Figure 2.7B).

BLAST searches for *rpl20-2* also matched a *Passiflora edulis* BAC clone Pe84M23 (AC278199.1) with high identity (82-95%). Mapping of nuclear *rpl20-2* against the *P. edulis* BAC clone identified an ORF of 685 bp with a putative intron of 313 bp (Figure 2.7C). Transcriptome assembly has been completed for *P. edulis* in ONEKP (db.cngb.org/onekp/). A tBLASTn search using *Arabidopsis thaliana* plastid RPL20 as a query identified a *P. edulis rpl20* transcript of 372 bp in the ONEKP database. The transcript was 99.5% nt and 100% aa identical to a coding domain in the ORF of *P. edulis* BAC clone that has an intron. The intron in *rpl20-1* and *rpl20-2* was validated with PCR and Sanger sequencing (Figure 2.7D). The amplicon for *rpl20-1* was 1800-2000 bp,

whereas the amplicons were much smaller (700-900 bp) for *rpl20-2* and had ~50% nt identity (Figure 2.7D). Intron size varied from 1643 bp in *P. oerstedii* to 2066 bp in *P. contracta* for *rpl20-1* and 324 bp in *P. oerstedii* to 573 bp in *P. contracta* for *rpl20-2* and all introns contained splice sites GT at the 5'end and AG at the 3' end. A BLASTn search for the *rpl20-1* intron against NCBI produced a 92% match with an unpublished nuclear sequence of *P. edulis* (MUZT01065614.1) that contained the intron and second exon for *rpl20-1* gene but lacks the first exon. Among the three species in subgenus *Decaloba*, Sanger sequencing for intron validation was carried out only for *P. auriculata*. Accession numbers for the *Passiflora rpl20* transcripts and genes are provided in Table 2.10

rps16

Two isoforms of the nuclear *rps16* transcript (*rps16-1* and *rps16-2*) were identified in all *Passiflora* species. *Passiflora rps16-1* had nt and aa identities >88% and *rps16-2* had identities >64% (Table 2.5). Additionally, two non-identical copies of *rps16-1* transcripts (*rps16-1a* and *rps16-1b)* were identified in *P. oerstedii* that had pairwise nt and aa identities of 83.7% and 84.6%. Mapping of transcriptome reads to copies of *rps16-1* in *P. oerstedii* provided support for both copies but the number of reads mapped varied substantially (1309 reads for *rps16-1a* vs. 446 reads for *rps16-1b*). Nuclear-encoded mitochondrial targeted *rps16* in *Populus alba* (Ueda et al. 2008)*, rps16-1* and *rps16-2*, were downloaded from NCBI and aligned with *Passiflora rps16* transcripts. The *Passiflora rps16-1* alignment, including the *Populus rps16-1*, had nt and aa identities >86%, whereas, the *rps16-2* had nt and aa identities >62% (Table 2.5). The N-terminal organelle signal

sequence (90 aa) of the *Populus* RPS16 was compared with the *Passiflora* RPS16 proteins. *Passiflora* RPS16-1 shares 95.5% aa identity with the *Populus* RPS16-1 and *Passiflora* RPS16-2 shares 70% aa identity with the *Populus* RPS16-2. Accession numbers for the *Passiflora* nuclear-encoded *rps16* transcripts and references are provided in the Table 2.11.

rpoA

No *rpoA* nuclear transcripts were detected in any *Passiflora* species. Searches for sigma factor genes (*sig*), nuclear-encoded components of plastid-encoded RNA polymerase (PEP), resulted in identification of transcripts of six sigma factors (*sig1*-*sig6*). The total number of *sig* genes and the copy number of the individual *sig* genes varied across species (Table 2.12). All six *sig* genes known in *Arabidopsis thaliana* (Chi et al. 2015) were identified in four *Passiflora* species *P. contracta, P. auriculata, P. tenuiloba* and *P. biflora* but *sig3* was not located in *P. pittieri* and *sig4* was not identified in *P. pittieri* and *P. oerstedii.* Despite high nt identity (>90%) of *P. tenuiloba sig3* with other *Decaloba* species, the *sig3* transcript in *P. tenuiloba* contained frame shift deletions and the ORF is present as two fragments. Mapping of transcriptome reads to the *P. tenuiloba sig3* transcript validated the frame shift deletions.

Pairwise estimations of synonymous (*dS*) and nonsynonymous (*dN*) substitutions and the *dN/dS* ratio for *rpoA* were substantially higher for all species in subgenus *Decaloba* except *P. microstipula* (Table 2.13). *Decaloba* species included in rate analyses belonged to four supersections, *Pterosperma* (*P. microstipula*), *Auriculata* (*P. auriculata*, *P. jatunsachensis* and *P. rufa*), *Cieca* (*P. tenuiloba* and *P. suberosa*) and *Decaloba* (*P. biflora,*

P. affinis and *P. misera*). The species from supersection *Cieca* had the most divergent *rpoA* with the highest *dS* and *dN* values of ~2.4 and ~0.97, respectively. The *dS* and *dN* values were also higher for species in supersection *Decaloba* but the *dN/dS* values were < 1. Only the species in supersection *Auriculata* had *dN/dS* > 1 due to slight increases in *dN* compared to *dS*. Branch-specific *dN/dS* values were estimated and plotted on the constraint tree (Figure 2.8). The branches with *dN/dS* >1 due to *dS* value close to zero were fixed to a value of 0.731, which was estimated using global-ratio model. All together five branches (one leading to *P. contracta* and four within subgenus *Decaloba*) have *dN/dS* >1 due to larger *dN* and *dS* value not close to zero. Likelihood ratio tests (LRTs) identified three branches with *dN/dS* >1 within subgenus *Decaloba* that were significantly different, including the branch leading to subgenus *Decaloba* excluding *P. microstipula*, the branch leading to supersection *Auriculata* and the branch leading to *P. misera* (Table 2.14 and Figure 2.8).

ycf1/ycf2

No nuclear transcripts of *ycf1* and *ycf2* were identified in any *Passiflora* species. The TIC214 protein, encoded by *ycf1*, along with three nuclear-encoded proteins, TIC20, TIC56 and TIC100, form the 1-MD (megadalton) protein translocon of the plastid inner envelope (TIC) (Kikuchi et al. 2013). Similarly, *ycf2* encodes a subunit of the 2-MD AAA-ATPase complex, a protein motor that contains six nuclear components, FTSHI1, FTSHI2, FTSHI4, FTSHI5, FTSH12 and NAD-malate dehydrogenase (Kikuchi et al. 2018). *Arabidopsis thaliana* 1-MD TIC complex proteins were used to query the assembled

transcripts of *Passiflora.* Transcripts for all 1-MD TIC components including all *tic20* isoforms were identified in *P. pittieri* and *P. contracta.* In contrast, in subgenera *Passiflora* and *Decaloba* only transcripts for *tic20* isoforms (except isoform I) were detected: II, IV and V in *P. tenuiloba, P. auriculata* and *P. biflora*; and IV and V in *P. oerstedii* (Table 2.15). Transcripts identified for *tic100* and *tic56* were substantially shorter with fragmented ORFs that contained multiple stop codons. To assess whether *tic100, tic56*, and *tic20-I* transcripts were missing in subgenera *Passiflora* and *Decaloba*, RNA reads were mapped and a tBLASTn search was performed using the sequences identified in other *Passiflora* species as queries. No transcripts with complete ORFs for *tic100, tic56* and *tic20-I* were identified in subgenera *Passiflora* and *Decaloba.*

Components of the 2-MD motor protein complex in *Passiflora* were investigated using the *Arabidopsis thaliana* 2-MD complex components as a query. All six nuclear-encoded components of the 2-MD complex were identified in *P. pittieri* and *P. contracta* including two isoforms of pdNAD-MDH that had pairwise aa identities of 91.5% for type 1 and 94.2% for type 2 (Table 2.16). In *P. oerstedii,* a transcript with a complete ORF was identified only for pdNAD-MDH of 2-MD protein complex in addition to fragmented transcripts lacking ORFs for *fstHi4* and *fstHi5* but no transcripts for *fstHi1*, *fstHi2* and *fstH12* were identified. However, transcripts for several other plastid FTSH/FTSHI proteins not known to be associated with the 2-MD motor complex were found in *P. oerstedii*. In subgenus *Decaloba* transcripts for *ftsHi4*, *ftsH12* and *pdNAD-MDH* of the 2-MD protein complex were identified in all species and an additional isoform of *pdNAD-MDH* only in *P. auriculata* but no transcripts for the remaining components were found. Similar to *P.*

oerstedii, transcripts for other plastid FTSH/FTSHI proteins not known to be associated with the 2-MD protein complex were identified in subgenus *Decaloba* as well (Table 2.16).

2.4. Discussion

Missing or divergent plastid genes in *Passiflora* have followed three distinct evolutionary paths: transfer to the nucleus, substitution by the nuclear genes and highly divergent gene that likely remain functional. Demonstrating that a gene synthesizes a protein that is subsequently targeted to the plastid constitutes another step necessary to validate the functionality of nuclear transfers. Hence, identification of nuclear transcripts that contain subcellular localization sequences with transcriptomic analysis suggests only that the gene has potential to be targeted to the plastids. Therefore, in the discussion the term 'nuclear transfer of plastid genes' in *Passiflora* indicates that these are putative functional transfers.

Comparative analyses of *Passiflora* indicate that three plastid genes (*rps7, rpl22* and *rpl32*) were transferred to nucleus, four (*rpl20, rps16, ycf1* and *ycf2*) were substituted by nuclear genes and the highly divergent *rpoA* remains functional in plastids (Figure 2.9). Transfers of *rpl22, rpl32* and substitution of *rps16* are known in several other angiosperm lineages (e.g., Gantt et al. 1991; Ueda et al. 2007; Ueda et al. 2008; Jansen et al. 2011; Park et al. 2015) therefore, discussion of these three genes is provided in supplementary text S1 (Supplementary Material online). The discussion will focus on the novel findings regarding the evolutionary fate of *rps7, rpl20, rpoA, ycf1* and *ycf2* in *Passiflora*, most of which have not been reported in angiosperms.

Transfer of plastid *rps7* to the nucleus

Plastid *rps7* encodes a component of the small subunit (30S) of the 70S ribosome. Bacterial *rps7* is essential for cell survival (Shoji et al. 2011) and in green algae, RPS7 plays important role in translation initiation in the plastid (Fargo et al. 2001). *Passiflora* plastid-encoded *rps7* presents an interesting evolutionary scenario since subgenus *Passiflora* species have an internal stop codon, whereas the gene is lost in *P. obovata* (subgenus *Deidamioides*) and subgenus *Decaloba* species except *P. microstipula* (Cauz-Santos et al. 2017; Rabah et al. 2019; Shrestha et al. 2019). In contrast, a complete sequence of *rps7* with conserved domains is present in species of polyphyletic subgenus *Deidamioides* and two species examined in subgenera *Astrophea* and *Tetrapathea* (Shrestha et al. 2019). Nuclear *rps7* with high sequence identity to *Arabidopsis thaliana* plastid *rps7* is present in transcriptomes of all six species of *Passiflora* examined, including *P. pittieri* and *P. contracta*, which also have an intact *rps7* in their plastomes (Figure 2.1). This suggests that *rps7* transferred to the nucleus early in the evolution of *Passiflora* and that the plastid-encoded *rps7* is differentially degraded across the genus. A single nuclear transfer of *rps7* is also supported by the presence of predicted transit peptide that has high sequence identity.

The transit peptide for nuclear *rps7* is identical to the transit peptide for nuclear-encoded plastid targeted thioredoxin M-type protein isoform 3 (TRX-m3) in each *Passiflora* species (Figure 2.1C). This could be due to the transfer of plastid *rps7* into the intron of nuclear *trx-m3*, which is co-transcribed but alternatively spliced resulting into two

gene products with same transit peptide. PCR amplification and Sanger sequencing as well as Illumina read mapping confirmed the insertion of plastid *rps7* into the intron (Figure 2.2). The insertion split the intron in two, forming a chimeric gene that encodes RPS7 as well as TRX-m3. The identification of functional transfer of a plastid gene into the intron of the nuclear-encoded plastid-targeted gene has not been reported among angiosperms. A similar example is the mitochondrial gene *rps14* that was transferred into the intron of the nuclear-encoded mitochondrial-targeted succinate dehydrogenase gene *sdh2*, which is processed by alternative splicing in maize and rice (Figueroa et al. 1999; Kubo et al., 1999).

Two previous studies reported that *rps7* has been pseudogenized at least four times in Salicaceae and suggested that the gene may have been transferred to the nucleus (Huang et al. 2017; Zhang et al. 2018) but neither examined nuclear data to support this hypothesis. Nuclear transcripts of *rps7* are present in Salicaceae species, *Salix purpurea* and *Populus trichocarpa,* and both contain transit peptides derived from nuclear *trx-m3* gene, suggesting that the transfer occurred prior to the divergence of Passifloraceae and Salicaceae (Figure 2.9). However, the transit peptides of RPS7 and TRX-m3 are not identical, as they are as in *Passiflora* species (Figure 2.3), suggesting that the nuclear *rps7* and *trx-m3* transcripts in Salicaceae may be derived from two separate nuclear loci. After the gene transfer, Salicaceae species may have experienced further evolutionary change that caused divergence of the targeting sequences in *rps7* and *trx-m3,* possibly due to gene duplication. There is evidence of whole-genome duplication (WGD) within Salicaceae, specifically prior to the divergence of *Salix* and *Populus* (Soltis et al. 2009; Qiao et al. 2019). If the transfer of plastid *rps7* into the nuclear *trx-m3* intron occurred prior the divergence of

Salicaceae and Passifloraceae, the WGD in Salicaceae would have duplicated the chimeric *rps7-trx-m3* gene and the duplicated copies could accumulate mutations independently. The duplicated *rps7-trx-m3* copies could generate *rps7* and *trx-m3* transcripts separately in Salicaceae, whereas in *Passiflora* a single *rps7-trx-m3* may be alternatively spliced to produce *rps7* and *trx-m3* transcripts. A thorough examination of Salicaceae is needed to understand the variation of the plastid-targeting sequences of nuclear-encoded *rps7-trx-m3* and *trx-m3* genes. Furthermore, denser taxon sampling of Malpighiales would elucidate the precise timing of plastid *rps7* transfer to the nucleus.

Thioredoxins are ubiquitous proteins that reduce disulfide bonds by thiol-disulfide interchange of reacting proteins and regulate redox environment (Schurmann and Jacquot 2000). Plant genomes harbor six classes of thioredoxin genes (*trx*- f, h, m, o, x and y) of prokaryotic and eukaryotic origin, of which many localize to the organelles (Gelhaye et al. 2005). Among the *trx-m* isoforms in *Arabidopsis,* the divergent isoform *trx-m3* plays a role in redox regulation of callose (a polysaccharide) deposition and regulates the permeability of plasmodesmata and symplastic transport (Benitez-Alfonso et al. 2009). Passifloraceae and Salicaceae are the only angiosperm families that have some species with plastid *rps7* either missing or pseudogenized. Since there are nuclear copies with transit peptides in both families it is likely that this gene will eventually be lost entirely from the plastome.

Substitution of plastid *rps16* by nuclear encoded *rps16*

Several angiosperm lineages have lost *rps16* from their plastomes, including Fabaceae (Saski et al. 2005; Schwarz et al. 2015), Salicaceae (Okumura et al. 2006),

Brassicaceae (Roy et al. 2010) and Passifloraceae (Jansen et al. 2007; Rabah et al. 2019; Shrestha et al. 2019). Ueda et al. (2008) experimentally validated that plastid *rps16* was substituted by dual targeted nuclear-encoded mitochondrial *rps16*. They suggested that the functional transfer of mitochondrial *rps16* to the nucleus occurred prior the divergence of monocots and eudicots, subsequently duplicated and most lineages have retained the duplicated copies. In *Medicago truncatula* and *Populus alba*, the dual targeted nuclear-encoded mitochondrial *rps16* has set the stage for the loss of the plastid copy. In *Passiflora rps16* transcripts are present with high sequence similarity to duplicated nuclear-encoded mitochondrial *rps16* in *Populus alba*, suggesting a similar scenario of substitution and subsequent loss of the plastid copy.

Substitution of plastid *rpl20* by putatively duplicated nuclear-encoded mitochondrial *rpl20*

Two distinct nuclear transcripts that contain RPL20 conserved domains and belong to 50S ribosomal protein family L20 were identified in all *Passiflora* species examined (Figure 2.7A-B). Phylogenetic analysis using amino acid sequences placed RPL20 in *Passiflora* into two clades, RPL20-1 and RPL20-2, and RPL20-1 was nested within a clade that includes nuclear-encoded, mitochondrial-targeted RPL20 (Figure 2.6). For *Passiflora* RPL20, only Predotar strongly predicted RPL20-1 is targeted to the mitochondrion and RPL20-2 to the plastid, whereas TargetP predicted "other" and LOCALIZER predicted plastid for RPL20-2 in three of the six species (Table 1.6). These results suggest

localization of RPL20-1and RPL20-2 in *Passiflora* to mitochondria and plastids, respectively but experimental validation is needed to confirm the target location.

Two alternative pathways are proposed for the origin of *rpl20-2* in the nucleus. In one scenario, nuclear-encoded, mitochondria-targeted *rpl20*, which is present across land plants (Bonen and Calixte 2005), was duplicated in the ancestor of *Passiflora* and the duplicate copy gained a plastid transit peptide (Figure 2.10A). Substantial deletion in the intron of *rpl20-2* as well as substitutions in the coding region would account for sequence divergence and intron length variation. Similarity in the gene structure of mitochondrial *rpl20-1* and plastid-targeted *rpl20-2* and the phylogenetic position of RPL20-2 sister to RPL20-1 indicates that the *rpl20-2* may have originated from a duplicated copy of nuclear-encoded mitochondrial *rpl20.* This scenario is analogous to the evolution of *rps13* but occurs in the opposite direction. A gene of plastid origin was transferred to nucleus, subsequently duplicated and the duplicate copy was targeted to mitochondria resulting in functional replacement of mitochondrial RPS13 (Adams et al. 2002). Alternatively, plastid *rpl20* was transferred to the nucleus in the ancestor of *Passiflora* and gained an intron as well as a plastid transit peptide (Figure 2.10B). Intron gains in organelle genes transferred to the nucleus are common and attributed to signal sequence acquisition via exon shuffling (Gantt et al. 1991; Wischmann and Schuster 1995; Adams and Palmer 2003; Ueda et al. 2007). Another plausible explanation for intron gain in *rpl20-2* is *de novo* insertion of an intron or intron gain via homing, a process in which an intron is transferred from an intron-containing allele to intron-less allele that is mediated by sequence homology (Lambowitz

and Belfort 1993). In *Passiflora,* the intron from the nuclear-encoded mitochondrial *rpl20* (*rpl20-1*) could act as source given the high sequence identity between the two *rpl20* genes.

The loss of plastid *rpl20* has not been reported previously for angiosperms. Only a few *Passiflora* species*, P. arbelaezii* and *P. cirrhifolia* from the polyphyletic subgenus *Deidamioides* and *P. tetrandra* from an Old World subgenus *Tetrapathea*, have intact *rpl20* in their plastomes. In contrast, subgenus *Decaloba* entirely lacks *rpl20*, subgenus *Passiflora* species have *rpl20* with multiple stop codons and *Astrophea* species have a single stop codon in the gene (Cauz-Santos et al. 2017; Rabah et al. 2019; Shrestha et al. 2019). The nuclear-encoded *rpl20-2* has likely substituted the role of plastid *rpl20* in *Passiflora* resulting in loss or pseudogenization of this gene in the plastome. It is probable that species of *Passiflora* with intact *rpl20* in their plastomes will eventually lose this gene.

Transfer of plastid *rpl22* to the nucleus

Plastid *rpl22* encodes an essential ribosomal protein of the large subunit (50S) (Fleischmann et al. 2011). Early evidence of *rpl22* transfer to nucleus in angiosperms was detected in Fabacecae (Gantt et al. 1991), and a second independent transfer in two species of Fagaceae was later reported (Jansen et al. 2011). Both studies characterized a nuclear copy of *rpl22* with two exons separated by an intron, in which first exon encodes a plastid transit peptide. Jansen et al. (2011) suggested a third transfer of plastid *rpl22* to the nucleus in *Passiflora* but no transcriptome data were available. The present study provides further evidence of a putative functional transfer of plastid *rpl22* to the nucleus in *Passiflora.* Nuclear *rpl22* transcripts identified in *Passiflora* share high sequence identity with plastid

rpl22 of *Arabidopsis thaliana* (Figure 2.4A). Although the length of the predicted transit peptide could not be confirmed, high sequence similarity for the N-terminal overhang suggests a single transfer of plastid *rpl22* to the nucleus in the ancestor of the genus. Furthermore, *Adenia mannii*, the sister genus to *Passiflora*, lacks plastid *rpl22* (Shrestha et al. 2019). Thus, the nuclear transfer of *rpl22* may have occurred early in the divergence of Passifloraceae. Further study incorporating species from other genera in Passifloraceae as well as other lineages in Malpighiales are needed to clarify the evolutionary timing of the *rpl22* transfer.

The origin of the *rpl22* transit peptide in Fabaceae and Fagaceae is unknown and hypothesized to be independent based on low sequence similarity (Jansen et al. 2011). Although nuclear RPL22 in *Passiflora* shares high sequence similarity with Fabaceae and Fagaceae, the identity for the transit peptide is < 20% suggesting that the nuclear transfer in *Passiflora* was independent. *Passiflora* RPL22 contains a fragmented RNA recognition motif (RRM) in the N-terminal overhang sequence (Figure 2.4B). The RRM shares high sequence similarity with a *Populus trichocarpa* and *Passiflora* plastid-targeted organelle RNA recognition motif containing protein (ORMM). These proteins are essential *trans* factors required for the post-transcriptional C-to-U RNA editing of organelle transcripts (Sun et al. 2013; Shi et al. 2016). Detection of an RRM in the *Passiflora* nuclear RPL22 may indicate the presence of a chimeric gene comprising *rpl22* of plastid origin and a fragment of duplicated ORRM encoding the N-terminal plastid transit peptide. Under this scenario, the ORRM portion would lack functional constraint allowing degradation of the RRM sequence. Several nuclear proteins containing an ORRM are known and at least 23

plastid ORRMs with RRM either at the N-terminus, C-terminus or embedded in the protein have been reported in *Arabidopsis thaliana* (Ruwe et al. 2011). It is plausible that a partial ORMM in the nuclear RPL22 in *Passiflora* was derived from a member of the ORRM gene family, however partial RRM deletion will make it difficult to discern the precise origin. The fragmented ORRM of *Passiflora* RPL22 shares high sequence similarity with both *Populus and Arabidopsis* ORRM proteins.

Transfer of plastid *rpl32* to the nucleus

Plastid *rpl32* has been lost in several Salicaceae and there is evidence for gene transfer to the nucleus (Ueda et al. 2007). Cusack and Wolfe (2007) proposed that following the transfer of plastid *rpl32* to the nucleus, functionalization was achieved with the formation of a chimeric gene that included the plastid transit peptide of Cu-Zn superoxide dismutase (cp *sod-1)*. Ueda et al. (2007) experimentally verified the functional transfer of plastid *rpl32* and showed that the chimeric cp SOD-RPL32 localizes to plastids in *Populus*. *Passiflora* transcriptome searches identified pt *sod-rpl32* transcripts that share high sequence identity to RPL32 and the transit peptide as well as overall structure of the *Populus* sequence (Figure 2.5). This suggests that the transfer of plastid *rpl32* to nucleus in *Passiflora* is similar to *Populus* and other Salicaceae species. A major difference is that pt SOD-RPL32 in *Passiflora* is ~ 90 aa longer relative to *Populus.* The chimeric pt *sod-rpl32* in *Populus* has six exons and five introns and pt *sod*-1 has eight exons and seven introns (Ueda et al. 2007). In contrast, *Passiflora* pt *sod-rpl32* contains all seven introns and eight exons with the last exon being substituted by plastid *rpl32*, similar to the situation in

Bruguiera in the Rhizophoraceae, another family of Malpighiales (Cusack and Wolfe 2007). The presence of chimeric pt *sod-rpl32* in the three Malpighiales families Salicaceae, Rhizophoraceae and Passifloraceae suggests a single nuclear transfer of plastid *rpl32* prior to divergence of the order (APG IV, 2016). Although the transfer of plastid *rpl32* occurred early in the divergence of Malpighiales, not all species in the order have lost plastid *rpl32* and many have retained copies with complete conserved domains (Shrestha et al. 2019). An independent transfer of plastid *rpl32* has been also reported in Ranunculaceae (Park et al. 2015) but has low similarity in transit peptide sequence and gene structure compared to Malpighiales. Thus, at least two independent transfers of plastid *rpl32* have occurred during the evolution of angiosperms.

Highly divergent *rpoA* is likely functional

The plastomes of photosynthetic plants contain four genes (*rpoA, rpoB, rpoC1* and *rpoC2*) encoding subunits of the plastid RNA polymerase (PEP, Serino and Maliga 1998). In *Passiflora*, the α-subunit (*rpoA*) is highly divergent compared to *Populus trichocarpa.* A previous study (Blazier et al. 2016a) reported that *P. biflora rpoA* has only 37.4% aa identity with *Populus* but the authors concluded that the gene is likely functional because it has conserved domains and is under purifying selection. Recently more divergent copies of *rpoA* were identified in two species in subgenus *Decaloba* (*P. tenuiloba* and *P. suberosa*) that have pairwise aa identity < 25% compared to *Populus* and lack conserved domains. For these reasons, *rpoA* was suggested to be a pseudogene in these species (Shrestha et al. 2019). No *rpoA* nuclear transcripts were detected in transcriptomes, which suggests that

there has not been a nuclear transfer of *rpoA.* The PEP holoenzyme comprises both the plastid-encoded subunits as well as nuclear-encoded sigma factors required for promoter recognition and initiation of transcription (Tiller and Link 1995). The *Arabidopsis thaliana* genome encodes six sigma factor genes (*sig1-sig6*) that have specific as well as overlapping functions (Chi et al. 2015). Transcripts for almost all *sig* genes are present in *Passiflora,* including the two species with most divergent *rpoA*, *P. tenuiloba* and *P. suberosa.* The presence of sigma factors and all other plastid-encoded PEP components and lack of nuclear *rpoA* transcripts suggests that the PEP is likely functional in *Passiflora.* Similar lines of evidence, lack of *rpoA* in the nuclear transcriptome, identification of all nuclear-encoded sigma factor genes and evolutionary rate comparisons, were used to argue for the functionality of highly divergent *rpoA* in *Pelargonium* species (Zhang et al. 2013; Blazier et al. 2016a).

Highly divergent *rpoA* in *Passiflora* is confined to subgenus *Decaloba*. Within *Decaloba,* species in supersection *Cieca* are most divergent with substantially higher *dS* and *dN* values compared to species in supersections *Decaloba* and *Auriculata* (Table 2.13). However, $dN/dS < 1$ indicates that the gene is under purifying selection in supersection *Cieca.* In contrast, $dN/dS > 1$ for species in supersection *Auriculata* suggests that positive selection may have contributed to divergence of *rpoA* in this clade. Subgenus *Decaloba* includes clades that have experienced different evolutionary pressures resulting in a divergent *rpoA.* Branch-specific rate analyses further indicate changes in selection pressure for *rpoA* over time within *Decaloba* (Figure 2.8). Significantly higher *dN/dS* for *rpoA* during the early divergence of subgenus *Decaloba* corresponds with $dN/dS > 1$ for the other

three PEP genes *rpoB, rpoC1* and *rpoC2* (Shrestha et al. 2019). This suggests that during the early divergence of subgenus *Decaloba* all components of PEP experienced positive selection resulting in divergent *rpo* genes.

Plastid *rpoA* is an essential subunit of the PEP (Serino and Maliga 1998) and its functional transfer to nucleus has been reported only in mosses (Sugiura et al. 2003; Goffinet et al. 2005). Beside *Passiflora,* highly divergent *rpoA* has been reported in three unrelated angiosperm lineages, *Annona, Berberis* and *Pelargonium* (Blazier et al. 2016a). These authors proposed two potential factors causing divergence of *rpoA*, the labile nature of the gene product and high level of genomic rearrangements via illegitimate recombination. Genomic rearrangements in subgenus *Decaloba* are widespread but divergent *rpoA* is specifically found in supersection *Cieca*. In agreement with Blazier et al. (2016a) the location of *rpoA* in the plastome may have also influenced the divergence of the gene. Except for *P. lutea, rpoA* in supersection *Decaloba* is located at the boundary of the inverted repeat (IR) (Shrestha et al. 2019). Subgenus *Decaloba* has experienced several IR expansions and *rpoA* is located in the region of IR boundary changes. However, the most divergent *rpoA* in *P. tenuiloba* and *P. suberosa* is currently located in the middle of the IR.

Loss of the two largest plastid genes in *Passiflora*

The phylogenetic distribution of plastid gene loss in *Passiflora* showed that almost all species in subgenera *Passiflora* and *Decaloba* lack *ycf1* and *ycf2,* and that these losses were independent (Figure 2.9). Experiments with *ycf1* and *ycf2* in *Nicotiana tabacum*

demonstrated that the gene products are essential for cell survival (Drescher et al. 2001) and recent proteomic studies have provided crucial insight into the function of these two genes. Kikuchi et al. (2013) proposed that *ycf1* encodes the TIC214 protein, an essential component of the plastid inner membrane protein translocon (TIC). Along with plastid-encoded TIC214, three other essential nuclear-encoded proteins, TIC20, TIC100 and TIC56, form a 1-megadalton (MD) complex (photosynthetic-type TIC) that facilitates the transfer of proteins across the inner plastid membrane (Kikuchi et al. 2009, 2013). Among the components of the TIC complex, TIC20 isoform I (TIC20-I) is considered the core protein that functions as the protein-conducting channel (Kikuchi et al. 2009). Similarly, *ycf2* encodes a component of the 2-MD AAA-ATPase complex, a motor protein that generates ATP required for inner membrane translocation (Kikuchi et al. 2018). The 2-MD protein complex also includes five nuclear-encoded FTSH proteases, FTSHI1, FTSHI2, FTSHI4, FTSHI5 and FTSH12 and plastid NAD-malate dehydrogenase (pdNAD-MDH) (Kikuchi et al. 2018). These authors verified that the 2-MD motor protein complex physically coordinates with the 1-MD TIC complex to facilitate plastid import. Filamentation temperature sensitive protein H (FTSH) in the 2-MD complex is a membrane bound ATP-dependent metalloprotease with diverse biological roles. *ftsH* was originally identified in bacteria as a single copy gene but four different *ftsH* protease genes have been identified in cyanobacteria and 17 in *Arabidopsis* (Sokolenko et al. 2002; Wagner et al. 2012). All 17 *ftsH* proteases in plants are either targeted to mitochondria or plastids, five of which are inactive isoforms (FTSHI 1-5) of unknown function as they lack the zinc-binding motif required for proteolytic activity (Sokolenko et al. 2002; Wagner et al. 2012). Kikuchi

et al. (2018) have shown that nuclear-encoded proteins, FTSHI1, FTSHI2, FTSHI4, FTSHI5 and FTSH12 and plastid-encoded YCF2 are associated with translocation of protein in plastids but did not find any association between FTSHI3 and plastid targeted proteins.

Nuclear transcripts for *ycf1* and *ycf2* were not detected in *Passiflora*, suggesting that the transfer of these genes to nucleus is unlikely. To assess whether the two largest plastid genes are lacking entirely other components associated with the *ycf1* and *ycf2* gene products were evaluated. Transcripts for all other components were identified, including members of the 1-MD TIC complex (*tic100*, *tic56* and *tic20-I*) as well as 2-MD AAA ATPase protein motor complex (*ftsHi1*, *ftsHi2*, *ftsHi4*, *ftsHi5*, *ftsH12* and *pdNAD-MDH*) in *P. pittieri* and *P. contracta*, both of which contain intact *ycf1* and *ycf2* in their plastomes. However, for the species that lack *ycf1* and *ycf2*, no other components of 1-MD complex and only some components of 2-MD complex were identified (Tables 2.15-2.16). The independent loss of both *ycf1* and *ycf2* in the genus and the lack of transcripts for the components associated with the 1-MD and 2-MD complexes in *Passiflora* supports the suggestion of Kikuchi et al. (2018) that these two complexes are functionally coordinated. A paralog of *tic20*, *tic20-IV*, is known to partially compensate for the role of *tic20-I* in knockout assays (Kasmati et al. 2011; Kikuchi et al. 2013) suggesting that TIC20-IV may be involved in an alternative import pathway (Nakai 2015a, 2015b). The *tic20-IV* paralog is present in all the *Passiflora* species that lack *tic20-I* and other 1-MD TIC components indicating TIC20-IV may have substituted for *ycf1* in *Passiflora*.

Passiflora species that lack *ycf2* are also missing transcripts for all/most FTSH/FTSHI proteins of the 2-MD protein complex. FTSHI3 is the only inactive isomer found in all *Passiflora* species examined including those with intact *ycf2* in their plastomes, supporting the hypothesis that its expression is independent of *ycf2* expression (Kikuchi et al. 2018). In addition, several other plastid FTSH proteases are present that are not known to be associated with 2-MD protein complex in *Passiflora* species that lack *ycf2* (Table 2.16). Perhaps, these plastid FTSH proteases have substituted the role of YCF2 in delivering the energy required for protein translocation, acting as an alternative to the 2-MD motor protein complex in *ycf2* lacking species. A comparative study including lineages with and without *ycf1* and *ycf2* in their plastomes may improve the understanding of protein import mechanisms and identify factors associated with the process. Since *Passiflora* includes numerous species with or without *ycf1* and *ycf2*, it is an ideal system to investigate alternative TIC and motor protein complexes required for plastid protein import.

2.5. Conclusion

In addition to substitution of plastid functions by nuclear encoded proteins, *Passiflora* also exhibits several cases of plastid ribosomal genes transferred to the nucleus providing evidence for ongoing endosymbiotic gene transfer. Some of these evolutionary events occurred early, during the divergence of the order Malpighiales, while others are restricted to the Passifloraceae (Figure 2.9). Examples of nuclear transfer of plastid genes in *Passiflora* include *rpl22*, which has been transferred independently in multiple angiosperm lineages, as well as the unprecedented transfer of *rps7*. The adoption of a preexisting transit

peptide by *rps7* is similar to the gain of a transit peptide by another plastid gene in *Passiflora*, *rpl32,* however the underlying mechanisms are likely different. Nuclear transfers of *rps7* and *rpl32* can provide essential insights into the processes behind ongoing endosymbiotic transfer of plastid genes to nucleus, which is limited for the plastid genes. In addition, the likely substitution of plastid *rpl20* by nuclear-encoded *rpl20* provides an example of recent gene substitution resulting from gene duplication, an ancient evolutionary process for ribosomal genes (Adams et al. 2002; Ueda et al. 2008). The substitution of two missing plastid genes, *ycf1* and *ycf2* by nuclear counterparts in *Passiflora* requires further investigation. Together, evidence for common and novel gene transfers or substitutions indicates multiple underlying mechanisms have mediated the loss of essential plastid genes in *Passiflora*. It is possible that the genus may have experienced a high frequency of plastid DNA transfer to the nucleus and estimates of plastid DNA content in the nucleus would enhance the understanding of cytonuclear interactions in *Passiflora*. In addition to gene loss, *Passiflora* plastomes also have experienced extensive structural rearrangements making it an excellent system to study cytonuclear coevolution.

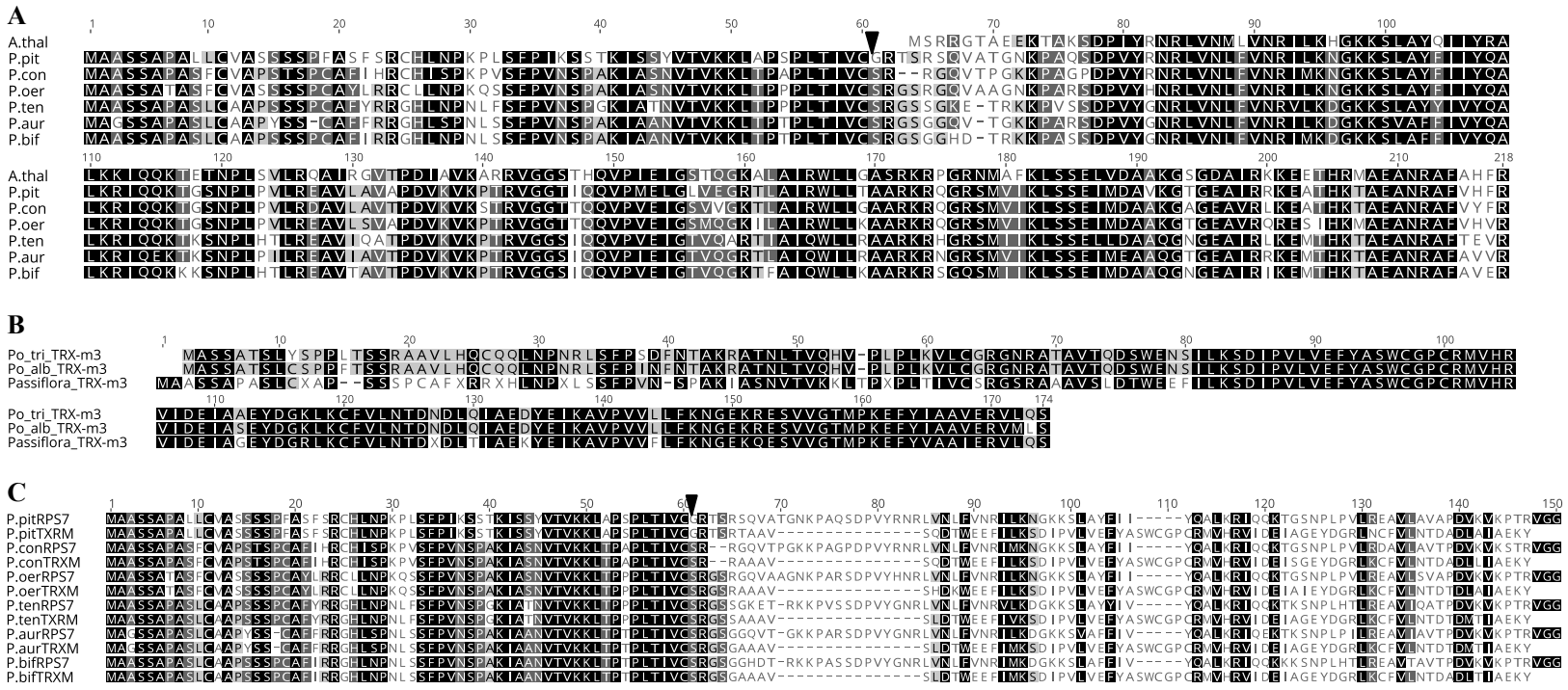
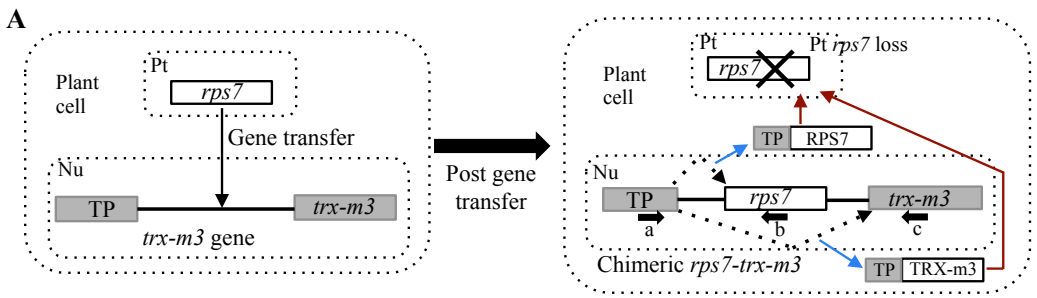
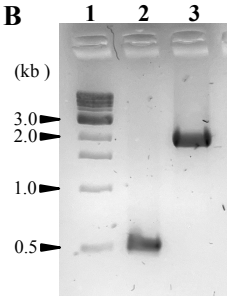


Figure 2.1. Amino acid alignments of *Passiflora* RPS7 and TRX-m3. (A) *Arabidopsis thaliana* plastid RPS7 amino acid (aa) alignment with nuclear RPS7 in six species of *Passiflora* (*P.*). (B) Comparison of TRX-m3 in *Populus trichocarpa* (Po_tri) and *Populus alba* (Po_alb) with the consensus TRX-m3 sequence for six *Passiflora* species. (C) Alignment of nuclear RPS7 against TRX-m3 among six *Passiflora* species with only the first 150 aa of sequence alignment shown. The aa identity for the transit peptide between RPS7 and TRX-m3 for each species is 100%. Black triangles denote transit peptide cleavage site predicted by TargetP. Abbreviations, P.pit, *P. pittieri;* P.con*, P. contracta*; P.oer, *P. oerstedii*; P.ten, *P. tenuiloba*; P.aur, *P. auriculata*; P.bif, *P. biflora*.





C

```
ATGGCTGCTT CTTCCGCTCC TGCTTTGCTT TGTGTTGCTT CTTCCTCTTC CCCATTCGCA TCCTTCAGCC GCTGCCACCT CAACCCTAAA CCGCTTTCAT TTCCTATCAA GTCTTCCACC AAAATATCAT CATATGTGAC  140
Localization sequence

TGTCAAGAAG CTCGCGCCCT CTCCTCTTAC GATTGTCTGC GGGCGTACCA GCCGAAGTAA TAGAGCTAAA ACTTTAAACC TCTGATGTTT TGTTGTTTGA TTTTTTGTTG ATTTACACTC ACACTCACAG GTCACGTCTG  280
Localization sequence
                                                                  5' splice site

TTTGCTGTAG CTTTTTTCAC GTGGATTTGT TATGTTCATA ATTGTGAGAA AGCTACAGTC TTCATAAACT TTAGCTCCGG TTTCGGACGA ATTCGTGGAT AGTTGTTAGT GCTTTATTTC GTTCATTGTG AACTATATGT  420

TGTTTGGTAA CAGCGTCAAG ATAGGAATTA ATTTCATTTG AAGAATCAAA TATTCTGTCT TTTCTATTGT GATCCTGAAA CGTTATACGC TGGCTTTTTC TGAAAGTTGC ATTCTTATGG ATCACATTTT TGGAGAAAAC  560

                                                                  3' splice site
TGATTTTATT GACCGATTTG AAGTAAAACT GATTCTTGGA TCCCTTTTAT CCTCATCTTA AGGTCAAGTTG CTACAGGGAA TAAACCTGCC CAATCTGATC CCGTTTATCG TAATCGGTTA GTTAACTTGT TTGTCAACCG  700
                                                                                                                                                rps7

TATCCTGAAA AATGGGAAGA AATCTCTGGC TTATTTCATT ATCTATCAAG CCTTAAAAAG GATCCAACAA AAGACAGGAT CCAATCCCTT ACCTGTTTTG CGCGAGGCAG TTCTTGCAGT GGCTCCTGAT GTGAAAGTGA  840
rps7

AACCAACTCG TGTGGGTGGA ACTATTCAGC AAGTGCCTAT GGAATTAGGG CTTGTGGAAG GAAGAACACT TGCCATTCGC TGGCTATTAA CGGCTGCTCG AAAACGTCAA GGCCGAAGTA TGGTTATCAA ATTAAGTTCT  980
rps7

GAGATAATGG ATGCTGTCAA AGGAACTGGT GAAGCCATAC GACGAAAGGA AGCCACACAC AAAACTGCTG AGGCGAACAG AGCTTTTGTA CACTTTCGTT AATCCGTGGA TAAGATGTAG ACACATGGAC TCTTGGAGCC 1120
rps7

AAACAACTTG GTCCGGAAGA TTGTGTAGAG GTTATGACTA TGAGATATAA TGTTAGGAAA CAGAAAACAA GAAACCCCTC TTTGTACTTC AATTAAATGT GCTGTATTGT TATTTCATCC CTAATTTGGA AACTTATGTA 1260

GAACGTGGTT TGGACTTCCT TTTGGTTGTG TTTCTGCATT GTATCATGAC ATGATTCAAT TGTCATGTAC GTTCTTAATT GATGTATTAT CTTCTGGGTT CTGGTCCTCT GATTGTCTCA TTAATCCCCT CTTGATGAAA 1400

AATTTCATGC TTCTTTACAT AGTTTTGTCA TCTCTGGAAT GCTTGCAAAT AGTTTCATCT AGGATAATGT TGTGGAGGAT GTGCATGGTG AAACTGAGAT AAGAGATGAA AGTTGTTGGA GATCGTAAGA AAAGGGTTTT 1540

TTTTTAGTGA CTCTTCTTGC TAGATGTGAT GCAGTAAAGA TTAGATACTT ATGCCTTGAG CTAATGGTGA TTTATCCTAG CACTTTTGAG CAGGTGCAAT TTTATTTATA TATGGTAGGC ATATGTTATG CATTCCAAGA 1680

CAGTTGCATT TTTTATAATC AAATTTTGAA ATTAGGTGAC GTTGTACAGG GTCTTTCAGA ATTGCTCAAA ACTGTGAAAG TTGAATTGGT TTGCAGAAAT CTTCAAATTT AGTTGGGAAC TTCGATTTTT TTCCTTCAAC 1820

TCCCACTTGG AAAGATGGTT TCAAGTTCAA TGTTGCTATT GCTTTATGCT TTCTCAAATC AAAGTTTTCA CTGACACGTT GGCATGAGAT CTATAAGCGT AATAGTTGGC AAGTTTGTTA ACCACATCAT TTTGATATGT 1960
3' splice site
TCAGCTGCAG CTGTTTCTCA GGACACTTGG GAGGAGTTTA TTCTGAAGAG TGATATCCCT GTCCTTGTTG AATTTTATGC GAGCTGGTGT GGTCCTTGTA GGATGGTTCA TCGAGTTATT GACGAGATTG CTGGAGAGTA 2100
Trx-m3

TGACGGAAGA CTTAATTGCT TTGTGCTAAA CACAGATGCG GACTTGGCAA TTGCAGAGAA ATATGAGATT AAGGCTGTAC CAGTTGTCTG TCTATTCAAG AACGGAGAGA AGCAAGAGTC TGTGGTTGGT ACCATGCCAA 2240
Trx-m3

AGGAATTCTA TGTTGCTGCC ATTGAGAGGG TTTTGCAGTC CTAA                                                                                                          2284
Trx-m3
```

Figure 2.2. Integration of plastid *rps7* into the intron of nuclear-encoded thioredoxin gene in *Passiflora*. (A) Schematic diagram (not to scale) depicts the insertion of plastid *rps7* into the intron of thioredoxin (*trx-m3*) that contains transit peptide (TP) known for plastid localization. Grey boxes indicate the exons of the *trx-m3* gene and the black line in between indicates the intron. The first exon of *trx-m3* gene contains TP. White box represents the plastid *rps7.* Alternative splicing is shown in dotted arrows. Blue and red arrows represent the gene product of alternative splicing and localization of the product to the plastid, respectively. Arrows (a, b and c) below the chimeric *rps7-trx-m3* indicate the location annealing sites of primers designed to amplify the gene product. The figure is not drawn to scale. (B) PCR amplifications of the chimeric *rps7-trx-m3* in *P. pittieri* with the primers designed in figure (A). Lane 1, 1 kb DNA ladder from NEB; Lane 2, PCR product with primer set a and b; and lane 3, PCR product with primer set a and c as indicated in (A). (C) *Passiflora pittieri* chimeric *rps7-trx-m3* as a representation for all other *Passiflora* species. The three exons of the gene are annotated in yellow. Intron 5' and 3' splice sites are boxed in grey. Abbreviations, Nu, nucleus; Pt, plastid, Mt, mitochondrion.

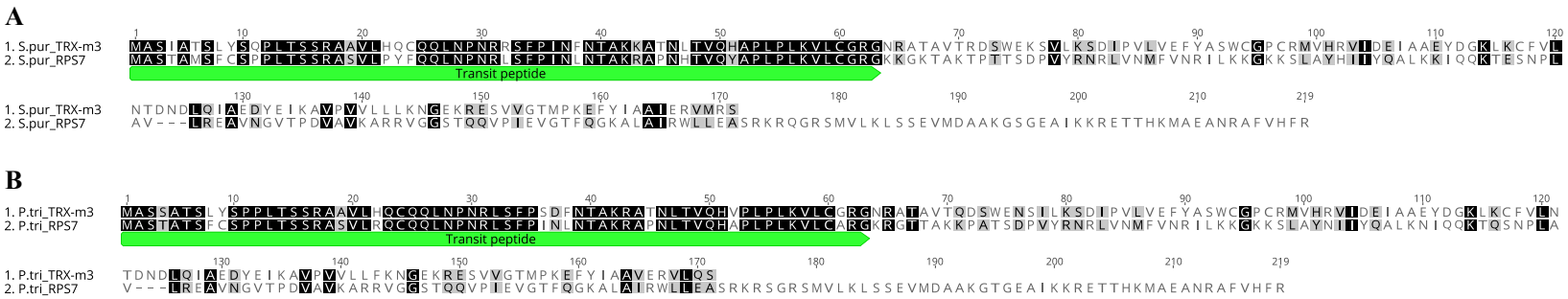


Figure 2.3. Amino acid alignments of nuclear RPS7 and TRX-m3 in two Salicaceae species. (A) *Salix purpurea* (S.pur) nuclear RPS7 and TRX-m3. (B) *Populus trichocarpa* (P.tri) nuclear RPS7 and TRX-m3. Predicted TRX-m3 transit peptide is labeled in green.

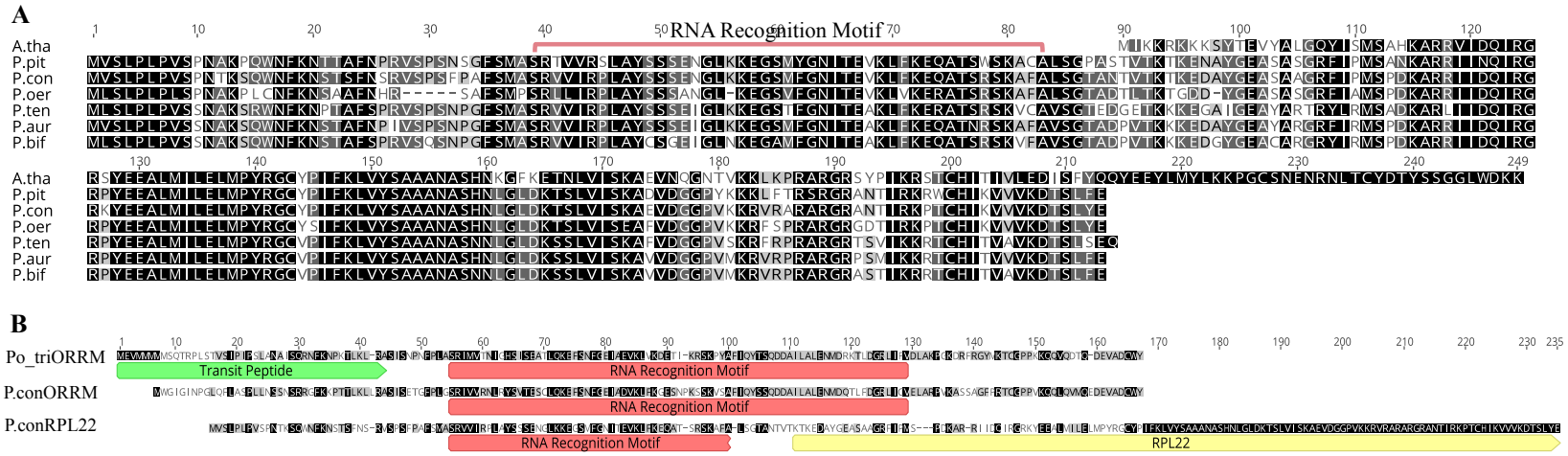


Figure 2.4. Amino acid alignments of *Passiflora* nuclear RPL22. (A) Amino acid (aa) comparison of *Passiflora* nuclear RPL22 with plastid RPL22 in *Arabidopsis thaliana*. (B) Comparison of the organelle RNA recognition motif protein (ORRM) sequence among *Populus trichocarpa* (Po_triORRM), *Passiflora contracta* (P.conORRM) and *Passiflora contracta* RPL22 (P.conRPL22). *Passiflora contracta* RPL22 is used to represent RPL22 identified in all *Passiflora* species. The predicted transit peptide of the *Populus* ORMM along with the ORRM and RPL22 sequences are labeled. Abbreviations, A.thal, *Arabidopsis thalianai*; P.pit, *Passiflora pittieri; P.con, P. contracta*; P.oer, *P. oerstedii*; P.ten, *P. tenuiloba*; P.aur, *P. auriculata*; P.bif, *P. biflora*.

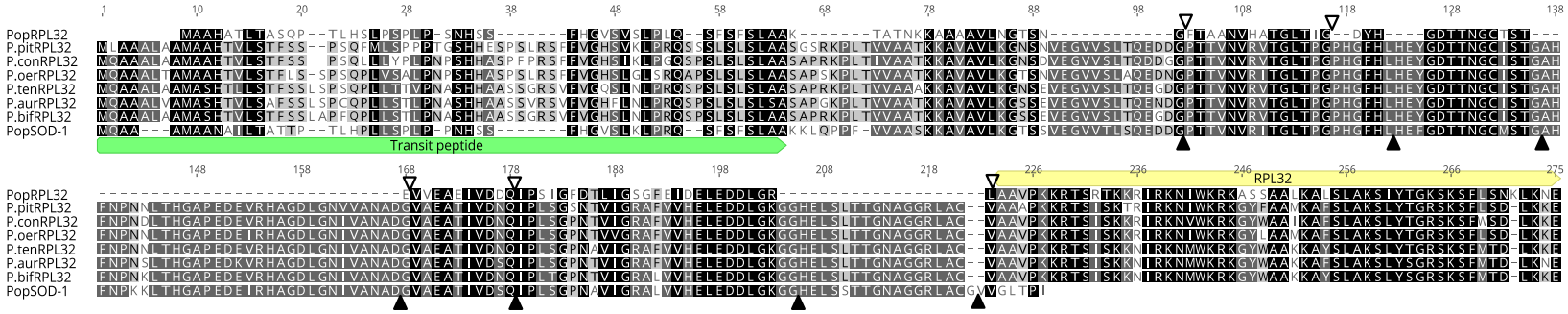


Figure 2.5. Amino acid alignment of nuclear RPL32 in *Passiflora* and *Populus.* The alignment includes *Populus* nuclear SOD-1 and *Passiflora* nuclear RPL32 for comparison. Open and filled triangles represent position of introns (cleaved from the transcript) described by Ueda et al. (2007) for plastid-targeted RPL32 and SOD-1 in *Populus alba,* respectively. The transit peptide of SOD-1 and the RPL32 domain of *Populus* polypeptides are annotated in green and yellow, respectively. Abbreviations, Pop, *Populus alba;* P.pit, *Passiflora pittieri;* P.con*, P. contracta*; P.oer, *P. oerstedii*; P.ten, *P. tenuiloba*; P.aur, *P. auriculata*; P.bif, *P. biflora*.

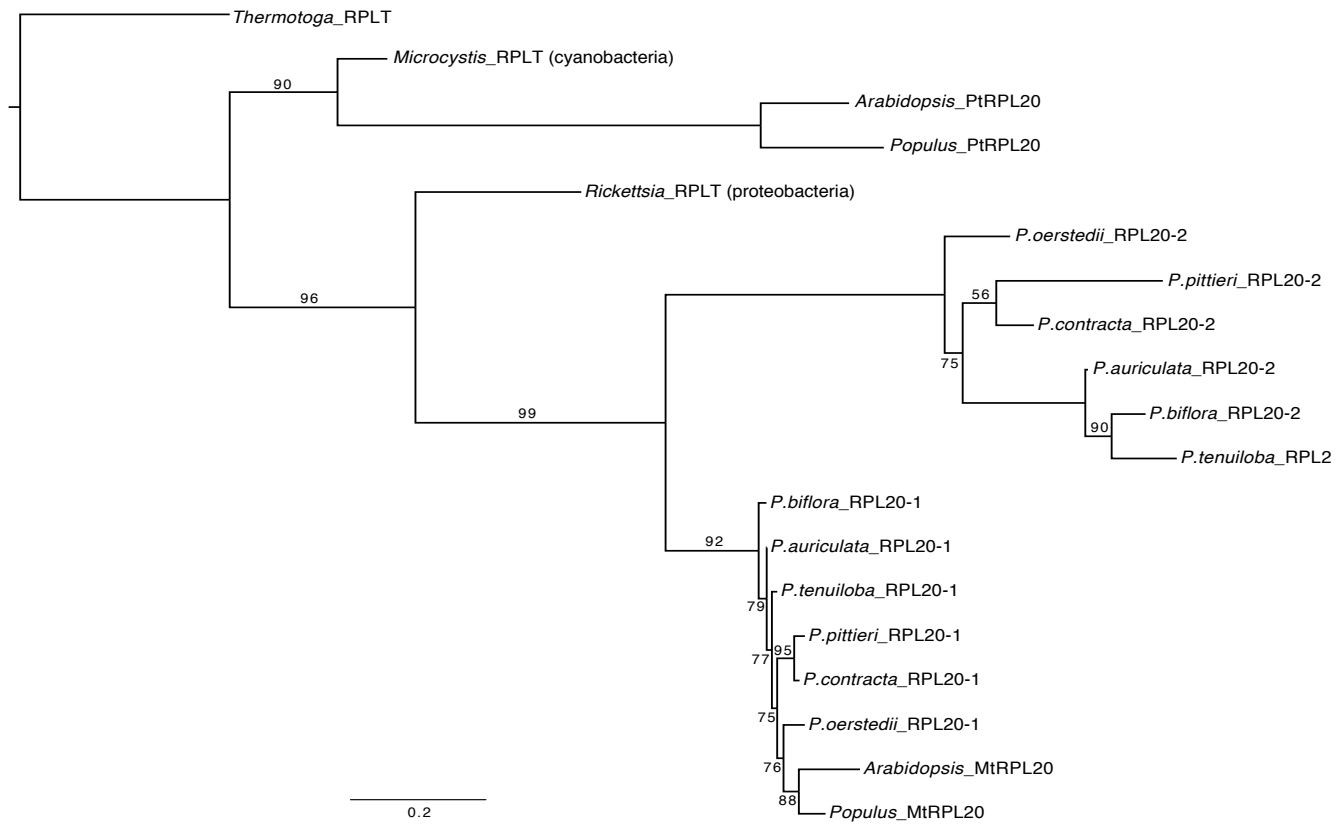


Figure 2.6. Phylogeny of *Passiflora* RPL20. The maximum likelihood phylogeny (-ln = -2289.199) was inferred using amino acid (aa) sequences of *Passiflora* (*P.*), *Arabidopsis* and *Populus* nuclear-encoded mitochondria-targeted RPL20 along with plastid-encoded (pt), and bacterial 50S ribosomal protein L20 (RPLT). Sequences from three bacterial species, *Thermotoga caldifontis*, *Microcystis aeruginosa* and *Rickettsia prowazekii* were included to infer phylogenetic position of nuclear-encoded RPL20 in *Passiflora.* Bootstrap values less than 100% are plotted above or below the branches. Horizontal bar indicates the expected aa substitutions per site.

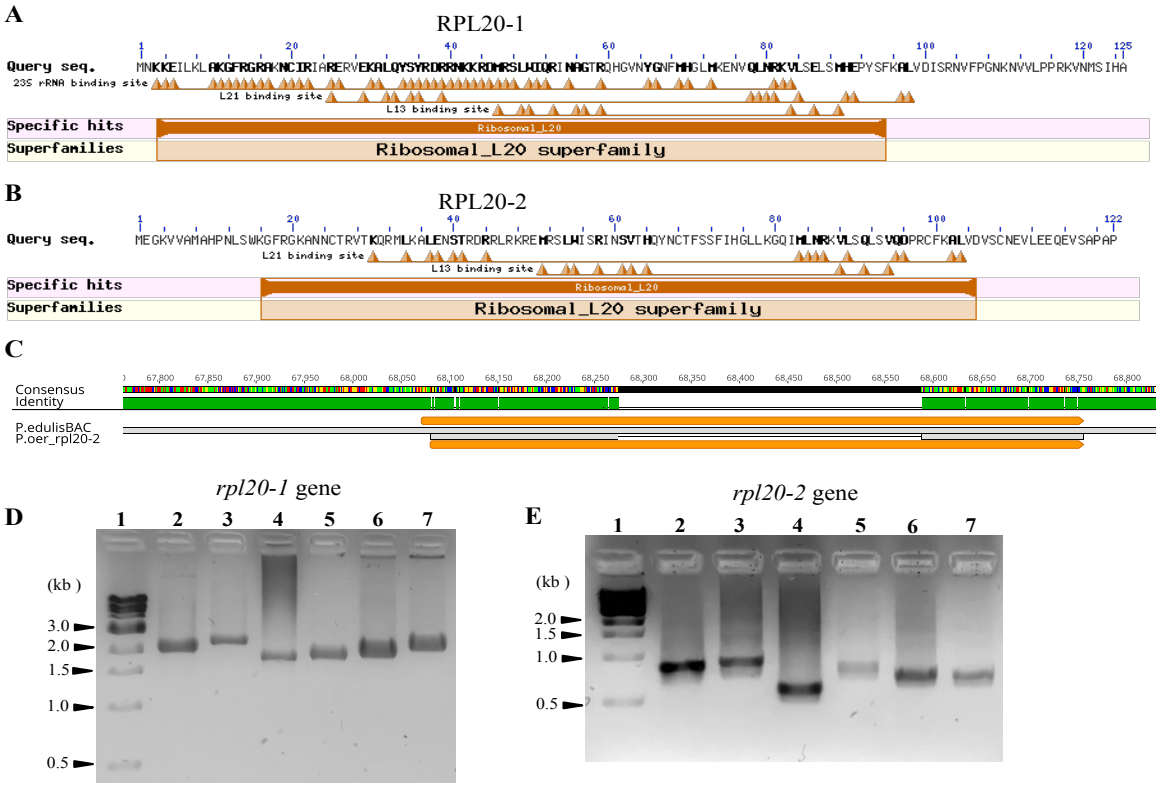


Figure 2.7. Nuclear-encoded RPL20 isoforms in *Passiflora*. The NCBI Conserved Domain (CD) Database was used for CD prediction. (A) Putative mitochondrial RPL20 (RPL20-1) in *Passiflora* containing RNA binding site as well as binding sites for other ribosomal subunits. (B) Putative plastid RPL20 (RPL20-2) in *Passiflora* with predicted binding sites for ribosomal subunits. (C) Mapping of *P. oerstedii rpl20-2* transcript against the *P. edulis* BAC clone Pe84M23 indicates the presence of an intron. (D-E) PCR amplifications to verify intron presence in *rpl20-1* and *rpl20-2* genes. Lane 1, 1 kb DNA ladder from NEB; lane 2, *Passiflora pittieri;* lane 3, *P. contracta*; lane 4, *P. oerstedii*; lane 5, *P. tenuiloba*; lane 6, *P. auriculata*; and lane 7, *P. biflora.*

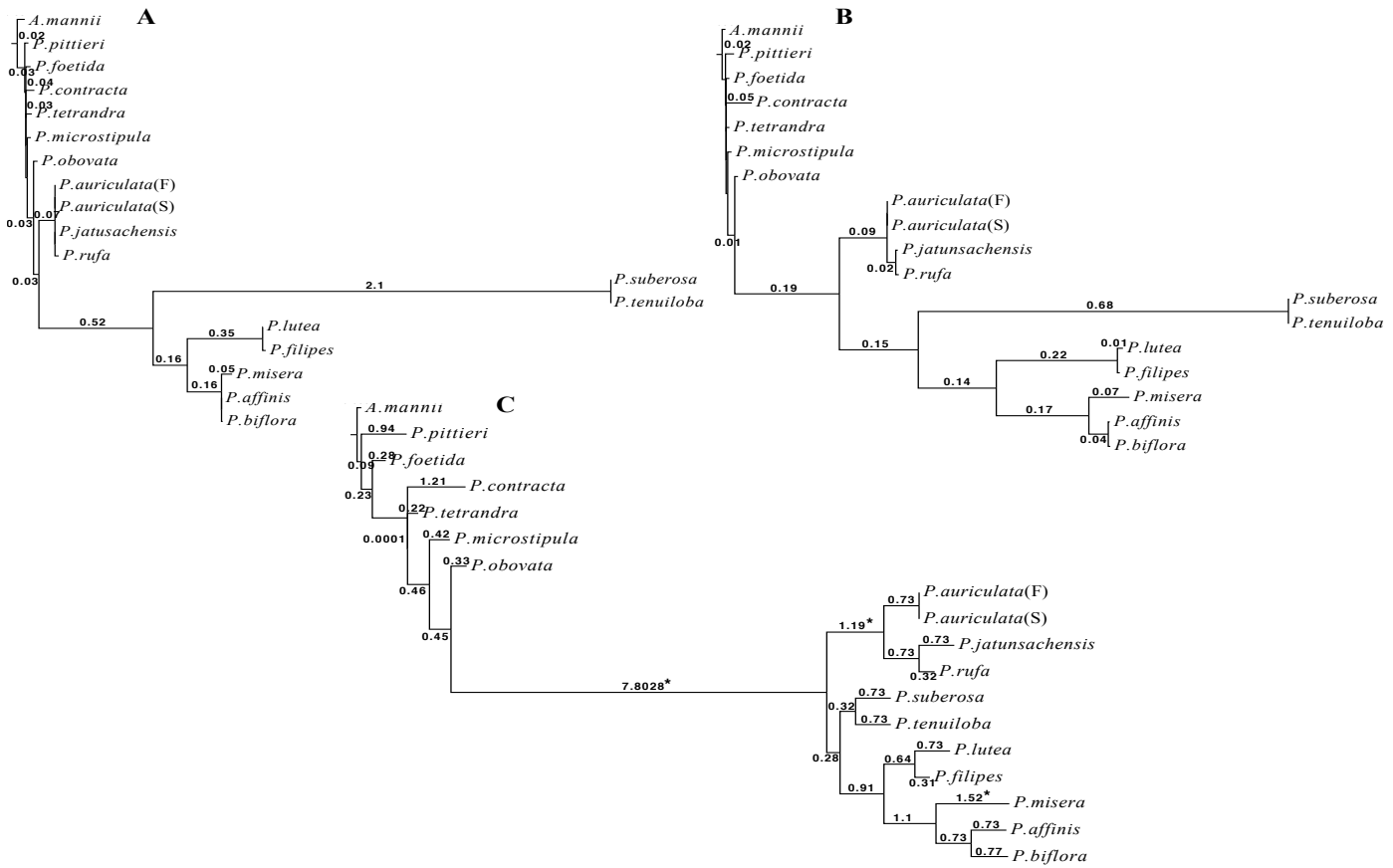


Figure 2.8. Substitution rates and *dN/dS* for *Passiflora rpoA* plotted on maximum likelihood constraint tree generated using plastid protein-coding genes (Shrestha et al. 2019). (A) Synonymous (*dS*) rate > 0.2. (B) Nonsynonymous (*dN*) rate > 0.1. (C) *dN/dS* was calculated using free-ratio model. Asterisks denote the branches with significantly higher *dN/dS* ($p<0.05$) evaluated by likelihood ratio test (LRT). Branches with *dN/dS* >1 due to *dS* closer to 0 were fixed at 0.731, a value generated with a global-ratio model. Branch lengths in all figures are proportional to *dS, dN* and *dN/dS* values.

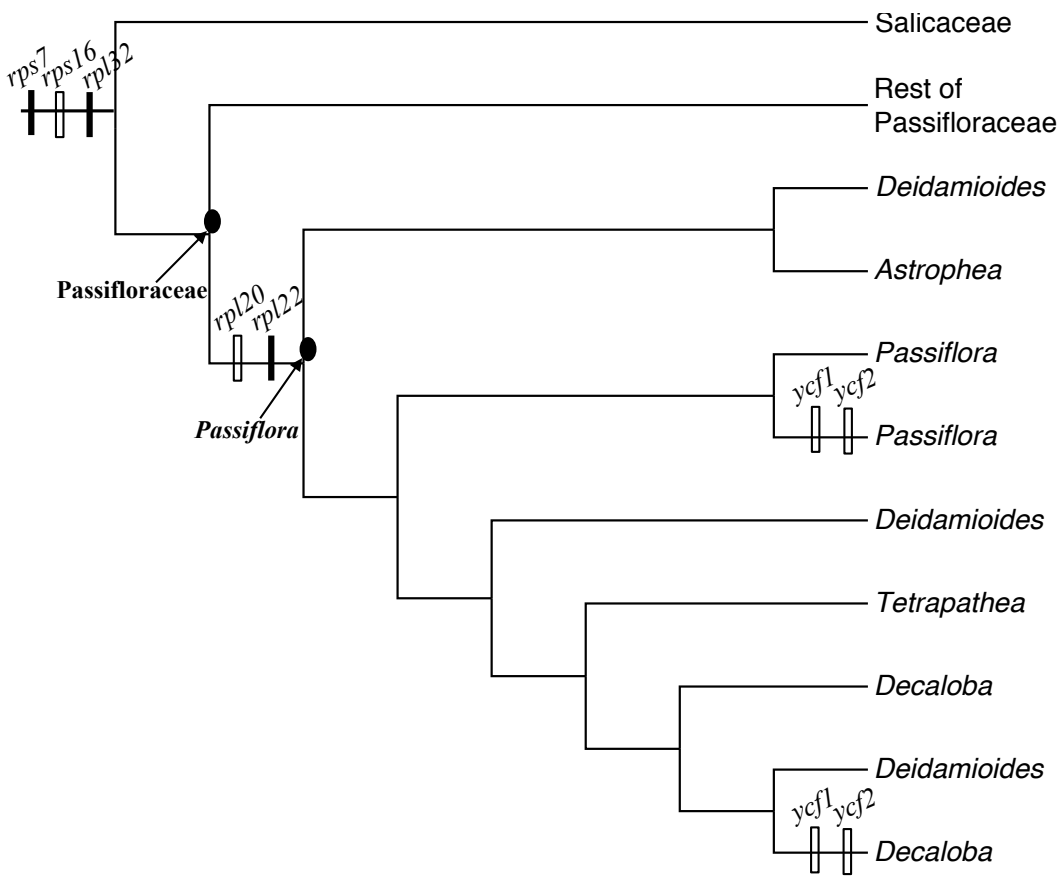


Figure 2.9. Phylogenetic distribution of nuclear transfer or substitution of plastid genes in *Passiflora.* The cladogram depicts the subgeneric relationships within *Passiflora* based on Shrestha et al. (2019) with Salicaceae as an outgroup. Distribution of plastid gene transfers to the nucleus (solid bar) and substitutions by nuclear genes (open bar) are plotted on the tree.

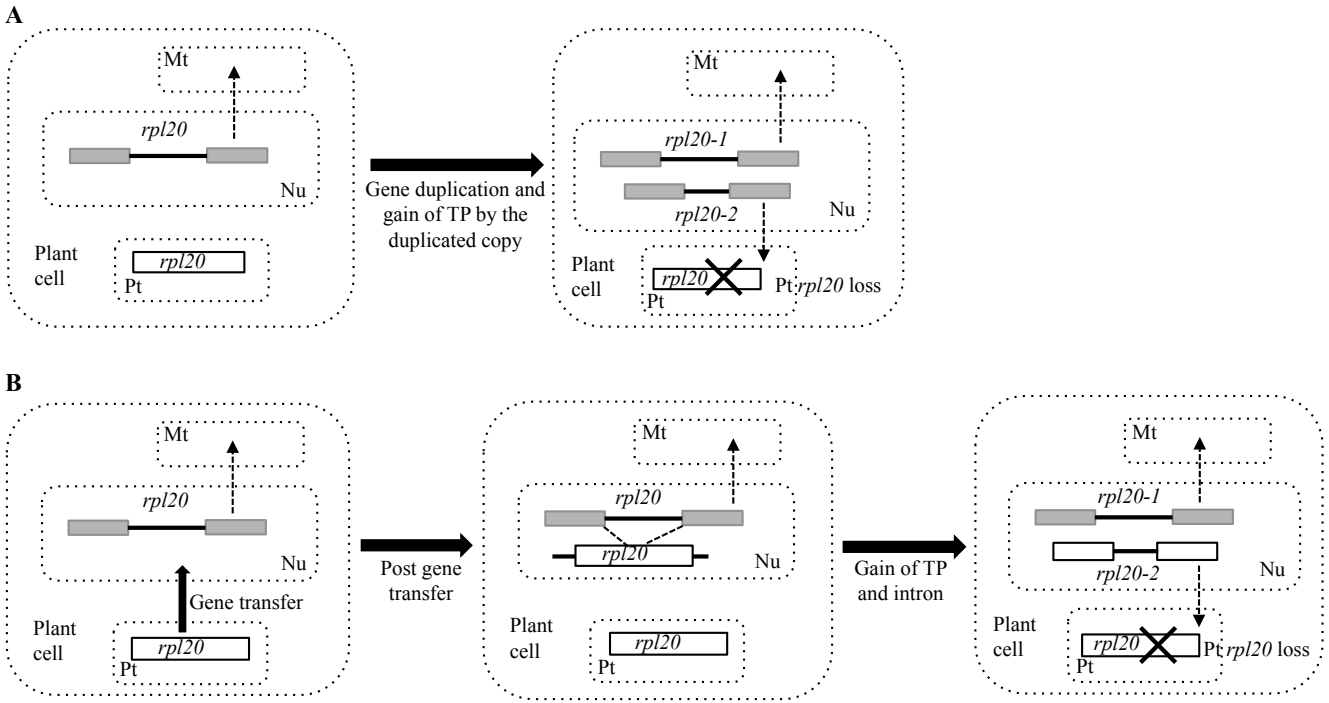


Figure 2.10. Schematic representation of two alternative scenarios for the origin of the nuclear-encoded plastid-targeted *rpl20* gene in *Passiflora.* (A) Duplication of nuclear-encoded mitochondrial *rpl20* followed by gain of a plastid-targeted transit peptide (TP) by the duplicated copy, followed by the loss of *rpl20* from the plastome. (B) Transfer of plastid *rpl20* to nucleus that includes acquisition of a TP and an intron. A possible scenario for intron gain could be intron transfer from nuclear-encoded mitochondrial *rpl20* due to sequence homology with nuclear-transferred *rpl20*, which is shown with dotted lines. Gain of a TP by nuclear-transferred plastid *rpl20* facilitates plastid localization of its product. Grey and white boxes represent exons for the nuclear and plastid genes, respectively. Black lines between the exons represent introns. Dotted lines with arrowheads indicate proteins that are targeted either to mitochondria or plastids. Major evolutionary events are shown in thick black arrows and descriptions are provided. The figure is not drawn to scale. Abbreviations, Nu, nucleus, Mt, mitochondrion, Pt, Plastid.

Table 2.1. List of species included for nucleotide substitution rate analyses of *rpoA*.

Subgenus	**Species**
Astrophea	*Passiflora pittieri*
Passiflora	*Passiflora foetida*
Deidamioide	*Passiflora contracta*
	Passiflora obovata
Tetrapathea	*Passiflora tetrandra*
Decaloba	*Passiflora microstipula*
	Passiflora auriculata
	Passiflora jatunsachensis
	Passiflora rufa
	Passiflora suberosa
	Passiflora tenuiloba
	Passiflora lutea
	Passiflora filipes
	Passiflora misera
	Passiflora affinis
	Passiflora biflora
Adenia	*Adenia mannii*

Table 2.2. Oligonucleotide primers used in PCR amplification and Sanger sequencing. Asterisks denote that the species was excluded from Sanger sequencing.

Gene	Primer	Direction	Species
rps7-trx-m3	5' ACTGTCAAGAAGCTCACGCCC 3'	Forward	*P. pittieri, P. contracta, P. oerstedii, P. tenuiloba*, P. auriculata, P. biflora**
	5' AAACGGGATCAGATCTGGCAG 3'	Reverse	
	5' ACTGTCAAGAAGCTCACGCCC 3'	Forward	*P. pittieri, P. contracta, P. oerstedii, P. tenuiloba, P. auriculata, P. biflora**
	5' CTACAAGGACCACACCAGCTC 3'	Reverse	
rpl20-1	5' GAGATCGACGCAACAAGAAGC 3'	Forward	*P. pittieri, P. contracta, P. oerstedii, P. tenuiloba, P. auriculata, P. biflora**
	5' GGCTCATGCATTGAAAGCTCC 3'	Reverse	
rpl20-2	5' GGCTCACCCAAACCTGTCAT 3'	Forward	*P. pittieri, P. contracta*
	5' CAACCAAGGCCTTGAAGCAG 3'	Reverse	
	5' GGATAGCCAAGCAGAGGACG 3'	Forward	*P. oerstedii, P. tenuiloba, P. auriculata, P. biflora**
	5' CAACCAAGGCCTTGAAGCAG 3'	Reverse	

Table 2.3. Transcriptome assembly statistics for the *Passiflora* (*P.*) species. Asterisks denote the statistic is based on single longest isoform per gene.

Species	Total reads	Read length (bp)	GC (%)	Reads after rRNA removal	rRNA (%)	Bowtie read mapping (%)	Total assembled bases*	Mean contig length*	N50*	BUSCO alignment (%)
P. pittieri	67,826,440	100	47	62,306,774	8.05	98.6	45,638,979	1024	1937	94.2
P. contracta	71,417,224	100	46	63,888,012	10.44	98.21	47,793,015	1048	2014	94.4
P. oerstedii	74,046,768	100	46	72,726,108	1.69	98.42	48,667,883	1108	2157	94.2
P. tenuiloba	67,265,786	100	44	66,971,196	0.39	98.6	48,862,335	1150	2265	92.2
P. auriculata	62,271,730	100	45	62,093,554	0.23	98.31	47,326,508	1045	2050	94.3
P. biflora	64,638,446	70-151	44	51,194,260	10.69	87.59	58,708,174	1171	2209	95.8

Table 2.4. A brief summary of results on fate of missing or divergent plastid genes with transcriptome analyses.

Gene	Description	Gene status in *Passiflora* plastome	Transcritpome results
rpl20	ribosomal protein L20, 50S subunit	Missing in subgenera *Passiflora* and *Decaloba. P. pittieri* contains premature stop codon. Present in all species in *Deidamioides* and *P. tetrandra.*	Putatively substituted by a duplicated nuclear-encoded mitochondrial *rpl20.*
rpl22	ribosomal protein L22, 50S subunit	Missing in all *Passiflora* species including *Adenia Mannii,* species from a genus sister to *Passiflora*.	Functional transfer to a nuclear gene containg RNA recognition motif.
rpl32	ribosomal protein L32, 50S subunit	Missing in *P. pittieri* (*Astrophea*)*, P. contracta* + *P. obovata* (*Deidamioides*), *P. jatunsachensis* + *P. rufa* + *P. auriculata* + *P. filipes* + *P. misera* + *P. affinis* + *P. biflora* (*Decaloba*).	Transfer to the duplicated copy of nuclear chloroplastic Cu-Zn dismutatse gene.
rps7	ribosomal protein S7, 30S subunit	Missing in *P. obovata (Deidamioides)* and in subgenus *Decaloba* except *P. microstipula.*	Functional transfer into the intron of nuclear Thioredoxin (m type) gene.
rps16	ribosomal protein S16, 30S subunit	Missing in all *Passiflora* species including outgroup genus *Populus.*	Substituted by dual targeted nuclear-encoded mitochondrial gene.
ycf1/ycf2	Components of TIC and motor protein complexes	Missing in subgenera *Decaloba* and *Passiflora* expect in species *P. microstipula* and *P. foetida.*	No nuclear transcript identified. Potentially substituted by alternative TIC and motor protein complexes.
rpoA	RNA polymerase subunit alpha	Highly divergent gene in subgenus *Decaloba.*	No nuclear transcript identified. The divergent plastid *rpoA* remain potentially functional.

Table 2.5. Nucleotide (nt) and amino acid (aa) identities for the transcripts identified in transcriptome analyses. For each transcript, comparisons were made among six *Passiflora* species and against genes from the reference species *Arabidopsis thaliana* or *Populus* (those marked with asterisk) species (see text). Abbreviations, Pt- Plastid. Mt- Mitochondrion.

Transcripts	**Among *Passiflora* species**		***Passiflora* vs. reference**			
			pt gene		**mt gene**	
	nt (%)	**aa (%)**	**nt (%)**	**aa (%)**	**nt (%)**	**aa (%)**
rps7	87	77.8	77.2	74.6	-	-
rps16-1*	88.8	94.7	-	-	86.7	93.7
rps16-2*	71	64.8	-	-	67.5	62.6
rpl20-1	92.2	95.6	34	30	76	86
rpl20-2	84.9	76.1	38	32	56	47
rpl22	86.8	80.3	82.8	76.9	-	-
rpl32	90.8	88.7	-	-	-	-

Table 2.6. Subcellular localization of nuclear-encoded *Passiflora* proteins predicted based on three prediction softwares, TargetP, LOCALIZER and Predotar. Prediction by TargetP is based on likelihood, whereas, the values provided by LOCALIZER and Predotar are probabilities. Abbreviations, Nu- Nucleus, Mt- Mitochondrion, Pt- Plastid, TL- Thylakoid Lumen, ER- Endoplasmic reticulum, TP- Transit Peptide, NA- Not available.

Species	Protein	TargetP							LOCALIZER				Predotar				
		Signal peptide	Mt transfer peptide	Pt tranfer peptide	TL transfer peptide	Other	Cleavage site	Prediction	Pt	Mt	Nu	TP length	Mt	Pt	ER	Elsewhere	Prediction
P. pittieri	RPS7	0.0001	0.0001	0.9987	0.0007	0.0005	60-61	Pt	1.0	-	-	41	0.02	0.96	0.03	0.04	Pt
P. contracta		0	0	0.9986	0.0007	0.0007	60-61	Pt	0.998	-	-	64	0.01	0.91	0.01	0.09	Pt
P. oerstedii		0	0	0.9988	0.0004	0.0008	60-61	Pt	1.0	-	-	51	0.01	0.94	0.01	0.05	Pt
P. tenuiloba		0	0	0.9996	0.0001	0.0004	60-61	Pt	0.997	-	-	68	0.01	0.8	0.01	0.2	Pt
P. auriculata		0	0	0.9945	0.0002	0.0053	59-60	Pt	0.996	-	-	52	0.01	0.73	0.03	0.26	Pt
P. biflora		0	0	0.9994	0.0002	0.0004	60-61	Pt	0.997	-	-	63	0.01	0.88	0.01	0.12	Pt
P. pittieri	RPL22	0	0.0005	0.9048	0.0082	0.0864	45-46	Pt	0.983	-	-	41	0.02	0.97	0.01	0.03	Pt
P. contracta		0.0001	0.001	0.8695	0.0095	0.1199	48-49	Pt	0.993	-	-	30	0.02	0.96	0.01	0.04	Pt
P. oerstedii		0	0.0434	0.8843	0.0241	0.0482	43-44	Pt	0.999	-	-	41	0.06	0.92	0.02	0.08	Pt
P. tenuiloba		0	0.0171	0.9078	0.001	0.0741	82-83	Pt	0.997	-	-	32	0.17	0.81	0.01	0.16	Pt
P. auriculata		0.0002	0.0057	0.8562	0.0004	0.1375	48-49	Pt	0.997	-	-	23	0.02	0.96	0.01	0.04	Pt
P. biflora		0.0001	0.0035	0.9073	0.0015	0.0877	49-50	Pt	0.999	-	-	41	0.04	0.91	0.01	0.09	Pt
P. pittieri	RPL32	0.0038	0.0009	0.9935	0.0017	0.0001	75-76	Pt	0.998	-	-	71	0.06	0.65	0.32	0.22	Pt
P. contracta		0.001	0.0003	0.9975	0.0009	0.0002	75-76	Pt	0.984	-	-	71	0.01	0.62	0.3	0.26	Pt
P. oerstedii		0.0037	0.0006	0.9836	0.0119	0.0002	75-76	Pt	0.998	-	-	44	0.02	0.94	0.16	0.05	Pt
P. tenuiloba		0.0003	0.0001	0.9975	0.002	0.0001	77-78	Pt	0.998	-	-	61	0.01	0.76	0.19	0.19	Pt
P. auriculata		0.0018	0.0018	0.9946	0.0016	0.0002	77-78	Pt	0.996	-	-	82	0.01	0.53	0.44	0.26	Pt
P. biflora		0	0.0001	0.9992	0.0006	0.0001	77-78	Pt	0.998	-	-	73	0.01	0.78	0.15	0.18	Pt
P. pittieri	RPL20-1	0.0001	0.0017	0	0	0.9982	NA	Other	-	-	-	NA	0.81	0	0	0.19	Mt
P. contracta		0.0001	0.0014	0	0	0.9985	NA	Other	-	-	-	NA	0.81	0	0	0.19	Mt
P. oerstedii		0.0001	0.0015	0	0	0.9985	NA	Other	-	-	-	NA	0.81	0	0	0.19	Mt
P. tenuiloba		0.0001	0.0013	0	0	0.9987	NA	Other	-	-	-	NA	0.81	0	0	0.19	Mt
P. auriculata		0.0001	0.0013	0	0	0.9987	NA	Other	-	-	-	NA	0.81	0	0	0.19	Mt
P. biflora		0.0001	0.0017	0	0	0.9982	NA	Other	-	-	-	NA	0.81	0	0	0.19	Mt
P. pittieri	RPL20-2	0.0001	0.0014	0.0275	0.0037	0.9673	NA	Other	-	-	-	NA	0.02	0.85	0.02	0.14	Pt
P. contracta		0.0001	0.0009	0.007	0.0265	0.9653	NA	Other	-	-	-	NA	0.02	0.8	0.01	0.2	Pt
P. oerstedii		0	0.0017	0.0884	0.0152	0.8947	NA	Other	0.905	-	-	72	0.02	0.81	0	0.18	Pt
P. tenuiloba		0	0.0001	0.0115	0.0114	0.977	NA	Other	-	-	-	NA	0.01	0.62	0	0.38	Pt
P. auriculata		0	0.0001	0.0153	0.0132	0.9713	NA	Other	0.903	-	-	43	0.01	0.64	0.01	0.35	Pt
P. biflora		0	0.0002	0.0236	0.0087	0.9675	NA	Other	0.951	-	-	38	0.01	0.6	0	0.4	Pt

Table 2.7. GenBank accession number (GBN) for *rps7* and *trx-m3* transcripts and the chimeric *rps7-trx-3* gene.

.

Species	Sequence	GBN
P. pittieri	*rps7* transcript	MT259499
	trx-m3 transcript	MT259504
	rps7-trx-m3 gene	MT259512
P. contracta	*rps7* transcript	MT259496
	trx-m3 transcript	MT259507
	rps7-trx-m3 gene	MT259513
P. oerstedii	*rps7* transcript	MT259501
	trx-m3 transcript	MT259505
	rps7-trx-m3 gene	MT259511
P. tenuiloba	*rps7* transcript	MT259502
	trx-m3 transcript	MT259506
	rps7-trx-m3 gene	MT259510
P. auriculata	*rps7* transcript	MT259497
	trx-m3 transcript	MT259509
	rps7-trx-m3 gene	MT259514
P. biflora	*rps7* transcript	MT259498
	trx-m3 transcript	MT259508
Populus trichocarpa	*rps7* transcript	XM_006384342
	trx-m3 transcript	XM_002313085
Salix purpurea	*rps7* transcript	MT259500
	trx-m3 transcript	MT259503
Populus alba	TRX-m3 protein	TKS05236.1
Arabidopsis thaliana	TRX-m3 protein	AT2G15570.2

Table 2.8. GenBank accession number (GBN) for nuclear *rpl22* , ORMM, and ORRM1 transcripts.

Species	Transcripts	GBN
P. pittieri	*rpl22*	MT259540
	ORRM	MT259549
	ORRM1	MT259549
P. contracta	*rpl22*	MT259537
	ORRM	MT259544
	ORRM1	MT259552
P. oerstedii	*rpl22*	MT259538
	ORRM	MT259546
	ORRM1	MT259551
P. tenuiloba	*rpl22*	MT259542
	ORRM	MT259543
	ORRM1	MT259550
P. auriculata	*rpl22*	MT259539
	ORRM	MT259548
	ORRM1	MT259554
P. biflora	*rpl22*	MT259541
	ORRM	MT259545
	ORRM1	MT259553
Populus trichocarpa	ORRM	XM_024596424; XP_024452192.1
	ORRM1	XM_024609853
Arabidopsis thaliana	ORRM	AT4G20030
	ORRM1	AT3G20930

Table 2.9. GenBank accession number (GBN) for the *rpl32* transcripts identified in *Passiflora* with references.

Species	Sequences	GBN
P. pittieri	*rpl32* transcript	MT259557
P. contracta	*rpl32* transcript	MT259558
P. oerstedii	*rpl32* transcript	MT259555
P. tenuiloba	*rpl32* transcript	MT259560
P. auriculata	*rpl32* transcript	MT259559
P. biflora	*rpl32* transcript	MT259556
Populus alba	cp *rpl32*	AB302219
Populus alba	cp *sod-1*	AB302220

Table 2.10. GenBank accession number (GBN) for *rpl20* transcripts and genes with references.

Species	Sequence	GBN
P. pittieri	*rpl20-1* transcript	MT259515
	rpl20-1 gene	MT259522
	rpl20-2 transcript	MT259526
	rpl20-2 gene	MT259533
P. contracta	*rpl20-1* transcript	MT259517
	rpl20-1 gene	MT259524
	rpl20-2 transcript	MT259530
	rpl20-2 gene	MT259535
P. oerstedii	*rpl20-1* transcript	MT259518
	rpl20-1 gene	MT259523
	rpl20-2 transcript	MT259528
	rpl20-2 gene	MT259534
P. tenuiloba	*rpl20-1* transcript	MT259516
	rpl20-1 gene	MT259521
	rpl20-2 transcript	MT259527
	rpl20-2 gene	MT259532
P. auriculata	*rpl20-1* transcript	MT259520
	rpl20-1 gene	MT259525
	rpl20-2 transcript	MT259531
	rpl20-2 gene	MT259536
P. biflora	*rpl20-1* transcript	MT259519
	rpl20-2 transcript	MT259529
Populus trichocarpa	nuclear-encoded mitochondrial *rpl20*	XM_006383341
	Plastid *rpl20* gene	ABO36728.1
Arabidopsis thaliana	nuclear-encoded mitochondrial *rpl20*	AT1G16740
	Plastid *rpl20* gene	NP_051082

Table 2.11. GenBank accession number (GBN) for the nuclear *rps16* transcripts in *Passiflora* with references.

Species	Transcripts	GBN
P. pittieri	*rps16-1*	MT259561
	rps16-2	MT259563
P. contracta	*rps16-1*	MT259568
	rps16-2	MT259569
P. oerstedii	*rps16-1a*	MT259567
	rps16-1b	MT259566
	rps16-2	MT259564
P. tenuiloba	*rps16-1*	MT259565
	rps16-2	MT259562
P. auriculata	*rps16-1*	MT259573
	rps16-2	MT259572
P. biflora	*rps16-1*	MT259571
	rps16-2	MT259570
Populus alba	*rps16-1*	AB365529
	rps16-2	AB365530

Table 2.12. GenBank accession number (GBN) for the sigma factor transcripts identified in *Passiflora* with references. Asterisk represents the sequence lacks complete open reading frame.

Species	Transcripts	GBN
P. pittieri	*sig1*	MT145371
	sig2-A	MT145375
	sig2-B	MT145373
	sig5	MT145374
	sig6	MT145372
P. contracta	*sig1*	MT145346
	sig2	MT145344
	sig3	MT145342
	sig4	MT145345
	sig5	MT145341
	sig6	MT145343
P. oerstedii	*sig1*	MT145356
	sig2	MT145358
	sig3	MT145357
	sig5	MT145354
	sig6	MT145355
P. tenuiloba	*sig1*	MT145352
	sig2-A	MT145353
	sig2-B	MT145351
	*sig3**	MT145350
	sig4	MT145348
	sig5	MT145347
	sig6	MT145349
P. auriculata	*sig1*	MT145364
	sig2	MT145360
	*sig3**	MT145363
	sig4	MT145359
	sig5	MT145362
	sig6	MT145361
P. biflora	*sig1*	MT145368
	sig2	MT145370
	sig3	MT145369
	sig4	MT145366
	sig5	MT145367
	sig6	MT145365
Arabidopsis thaliana	*sig1*	AT1G64860.1
	sig2	AT1G08540.1
	sig3	AT3G53920.1
	sig4	AT5G13730.1
	sig5	AT5G24120.1
	sig6	AT2G36990.1

Table 2.13. Synonymous (*dS*), nonsynonymous (*dN*) and *dN/dS* ratio calculated for *rpoA* in *Passiflora* along with *rpoA* nucleotide (nt) and amino acid (aa) identity compared to *Adenia mannii*. Pairwise estimation of nucleotide substitution rate for *rpoA* in *Passiflora* carried out using PAML v.4.8 (Yang, 2007). Highly divergent species are highlighted in bold.

Subgenus	Species	*dS*	*dN*	*dN/dS*	nt identity (%)	aa identity (%)
Astrophea	*P. pittieri*	0.0985	0.0229	0.2325	95.7	93.3
Passiflora	*P. foetida*	0.0938	0.0198	0.2111	96.2	94.7
Deidamioides	*P. contracta*	0.0996	0.0516	0.5181	90	84.2
	P. obovata	0.1177	0.0255	0.2167	92.8	90.9
Tetrapathea	*P. tetrandra*	0.0978	0.0168	0.1718	96.4	95
Decaloba	*P. microstipula*	0.1028	0.0183	0.1780	96.1	94.4
	P. auriculata	0.1913	0.2128	1.1124	74.7	59.3
	P. jatunsachensis	0.1811	0.227	1.2535	74.2	57.3
	P. rufa	0.1907	0.2272	1.1914	75.3	59.3
	P. suberosa	**2.3714**	**0.9829**	**0.4145**	**38.2**	**22**
	P. tenuiloba	**2.4554**	**0.9735**	**0.3965**	**38.1**	**22**
	P. lutea	1.0814	0.6173	0.5708	50.7	33.9
	P. filipes	1.0783	0.6133	0.5688	50.8	33.9
	P. misera	0.9084	0.5508	0.6063	51.5	33.6
	P. affinis	0.8509	0.5421	0.6371	53.7	35.9
	P. biflora	0.8535	0.5417	0.6347	53.7	35.6

Table 2.14 Branch-specific loglikehood ratio tests (LRTs) for *rpoA* estimated using branch model in PAML v.4.8 (Yang, 2007). False discovery rate (FDR) was used to adjust p-values for the multiple comparisons

Branch	Null model		Alternative model		LRT	p-value	FDR corrected p-value
	lnL	*dN/dS*	lnL	*dN/dS*			
Branch leading to supersections *Auriculata*, *Cieca* and *Decaloba*	-7132.7183	0.7316	-7120.9901	7.8028	23.456	1.28E-06	5.112E-06
Branch leading to supersection *Auriculata*	-7132.7183	0.7316	-7128.8344	1.1859	7.768	0.00532	0.0071
Branch leading to *P. misera*	-7132.7183	0.7316	-7128.1736	1.5226	9.089	0.00257	0.0051

Table 2.15. GenBank accession number (GBN) for the transcripts associated with 1 megadalton (MD) TIC complex identified in *Passiflora* with reference**.** The sequences highlighted in bold are part of the TIC complex and rest are not known to be associated with the 1-MD TIC. Asterisk denotes the sequence is partial or lacks complete open reading frame.

Species	Transcritps	GBN
P. pittieri	***tic100***	MT145393
	tic56	MT145395
	tic20-I	MT145397
	tic20-II	MT145394
	tic20-IV	MT145396
	tic20-V	MT145392
P. contracta	***tic100***	MT145399
	tic56	MT145402
	tic20-I	MT145403
	tic20-II	MT145401
	tic20-IV	MT145400
	tic20-V	MT145398
P. oerstedii	*tic20-IV*	MT145390
	tic20-V	MT145391
	tic20-I*	MT145388
	tic100*	MT145387
	tic56*	MT145389
P. tenuiloba	*tic20-II*	MT145381
	tic20-IV	MT145380
	tic20-V	MT145382
P. auriculata	*tic20-II*	MT145386
	tic20-IV	MT145383
	tic20-V	MT145385
	tic100*	MT145384
P. biflora	*tic20-II*	MT145376
	tic20-IV	MT145378
	tic20-V	MT145379
	tic100*	MT145377
Arabidopsis thaliana	*tic100*	AT5G22640
	tic56	AT5G01590
	tic20-I	AT1G04940
	tic20-II	AT2G47840
	tic20-IV	AT4G03320
	tic20-V	AT5G55710

Table 2.16. GenBank accession number (GBN) for the transcripts associated with 2 megadalton (MD) AAA-ATPase protein motor complex identified in *Passiflora* with reference. The sequences highlighted in bold are part of the 2-MD complex and rest are not known to be associated with the 2-MD complex. Asterisk denotes the sequence is partial or lack complete open reading frame.

Species	Sequences	GBN
P. pittieri	***ftsHi1***	MT145454
	ftsHi2	MT145453
	ftsHi4	MT145450
	ftsHi5	MT145452
	ftsH12	MT145455
	pdNAD-MDH-1	MT145448
	pdNAD-MDH-2	MT145449
	ftsHi3	MT145451
P. contracta	***ftsHi1***	MT145446
	ftsHi2	MT145445
	ftsHi4	MT145440
	ftsHi5	MT145444
	ftsH12	MT145447
	pdNAD-MDH-1	MT145441
	pdNAD-MDH-2	MT145443
	ftsHi3	MT145442
P. oerstedii	***pdNAD-MDH***	MT145431
	ftsHi4*	MT145434
	ftsHi5*	MT145432
	ftsH	MT145439
	ftsH2	MT145438
	ftsH6	MT145437
	ftsH7/9	MT145436
	ftsH11	MT145433
	ftsHi3	MT145435
P. tenuiloba	***ftsHi4***	MT145424
	ftsH12	MT145423
	pdNAD-MDH	MT145427
	ftsH	MT145429
	ftsH2	MT145430
	ftsH7/9	MT145428
	ftsH11	MT145426
	ftsHi3	MT145425
P. auriculata	***ftsHi4***	MT145414
	ftsH12	MT145417
	pdNAD-MDH-1	MT145415
	pdNAD-MDH-2	MT145416
	ftsH	MT145421
	ftsH2	MT145420
	ftsH6	MT145419
	ftsH7/9	MT145422
	ftsH11	MT145418
	ftsHi3	MT145413
P. biflora	***ftsHi4***	MT145407
	ftsH12	MT145408
	pdNAD-MDH	MT145404
	ftsH	MT145412
	ftsH2	MT145411
	ftsH6	MT145410
	ftsH7/9	MT145409
	ftsH11	MT145405
	ftsHi3	MT145406
Arabidopsis thaliana	***ftsHi1***	AT4G23940
	ftsHi2	AT3G16290
	ftsHi4	AT5G64580
	ftsHi5	AT3G04340
	ftsH12	AT1G79560
	pdNAD-MDH-1	AT3G47520
	ftsHi3	AT3G02450

Chapter Three

Modes of plastid inheritance in *Passiflora*[3]

3.1 Introduction

Plastids in seed plants contain their own genome (plastome) that differs from the nuclear genome mainly in their function, prokaryotic origin and non-Mendelian mode of inheritance. Plastid inheritance is usually uniparental and shares an important feature, vegetative segregation, a stochastic process where plastids of different genotypes segregate (sort-out) when the cell divides, usually early in plant development (Kirk and Tilney-Bassett 1978; Birky 1994, 2001). Due to non-Mendelian characteristics, uniparental inheritance and vegetative segregation, plastomes are generally considered to be genetically homogenous despite the presence of many plastids per plant cell with each containing numerous copies of the unit genome in each nucleoid (Greiner et al. 2019). Within seed plants, it is generally accepted that plastid inheritance is maternal among angiosperms, whereas paternal inheritance is common among gymnosperms with exceptions in both groups (Kirk and Tilney-Bassett 1978; Szmidt et al. 1987; Corriveau and Coleman 1988; Neale and Sederoff 1989; Zhang et al. 2003). Cytological screenings of pollen for plastid DNA (ptDNA) of nearly 300 angiosperm species reported that approximately 80% and

[3] This chapter contains a manuscript currently in preparation, "Clade-specific plastid inheritance patterns including frequent biparental transmission in *Passiflora* interspecific crosses" by Shrestha B, Gilbert LE, Ruhlman TA, and Jansen RK to be submittted to the journal Theoretical and Applied Genetics. Bikash Shrestha performed all the experiments, conducted data analyses and wrote the manuscript.

20% have the potential for maternal plastid and biparental transmission, respectively (Corriveau and Coleman 1988; Zhang et al. 2003). The bias toward maternal inheritance is attributed to the distribution of plastids during pollen development. During mitotic divisions of a microspore, sperm cells fail to receive any plastids or receive fewer plastids that largely contribute to strictly maternal plastid transmission (Hagemann 1979; Hagemann and Schröder 1989). Even when ptDNA is not completely excluded in sperm cells, other mechanisms are known to prevent paternal plastid transmission during and post fertilization, consequently enforcing maternal plastid transmission (Sears 1980; Hagemann and Schröder 1989; Mogensen 1996).

The prevalence of maternal inheritance in angiosperms is considered to be driven by constraints to avoid intracellular conflicts and the spread of selfish elements associated with paternal plastids (Reboud and Zeyl 1994; Zhang and Sodmergen 2010; Greiner et al 2014). However, uniparental inheritance is prone to the accumulation of deleterious mutations over evolutionary time. To overcome this, biparental inheritance has evolved to rescue defective plastids and alleviate mutational load (Zhang and Sodmergen 2010; Greiner et al. 2014; Barnard-Kubow et al. 2017). In the past few decades, plastid inheritance studies have expanded in many plant lineages and evidence of paternal leakage and biparental inheritance has been reported in lineages that were thought to exhibit purely maternal inheritance (Azhagiri and Maliga 2007; McCauley et al. 2007; Thyssen et al. 2012), which may support the hypothesis that paternal transmission helps to prevent the perils of strictly maternal inheritance but empirical evidence to support this is lacking. Nonetheless, it certainly indicates that the underlying mechanisms of strict maternal transmission are not

completely effective. Since the initial report of biparental plastid inheritance in *Pelargonium* (Baur 1909) based on the variegated phenotype, additional evidence has been described in a few other lineages including *Oenothera* (Chiu et al. 1988), *Medicago* (Smith et al. 1986; Smith 1989; Matsushima et al. 2008), *Turnera* (Shore et al. 1994; Shore and Triassi 1998), *Zantedeschia* (Brown et al. 2005), *Passiflora* (Hansen et al. 2007) and *Campanulastrum* (Barnard-Kubow et al. 2017). Two lineages, *Pelargonium* and *Oenothera*, have been extensively studied and variations of plastid transmission patterns are under strong influence of nuclear or plastid genomes (Kirk and Tilney-Bassett 1978; Tilney-Bassett and Birky 1981; Chiu et al. 1988). With an increasing number of progeny being assayed in inheritance studies, low frequency paternal transmission has been detected in lineages that were considered to display strictly maternal inheritance (Medgyesy et al 1986; Ruf et al. 2007; Azhagiri and Maliga 2007; Ellis et al 2008). Low frequency paternal inheritance in angiosperms is frequently observed within interspecific crosses, thus the correlation of paternal transmission in crosses involving divergent parents has been postulated, which is potentially due to the failure of mechanisms to prevent paternal inheritance (Cruzan et al. 1993; Rebound and Zeyl 1994; Yang et al. 2000; Hansen et al. 2007). Similarly, two lineages within rosids, *Medicago* and *Turnera*, exhibit rare predominant paternal inheritance even in intraspecific crosses, a mode that is commonly observed among gymnosperms (Schumann and Hancock 1989; Smith 1989; Matsushima et al. 2008; Shore et al 1994; Shore and Triassi 1998). These two lineages inherit plastids predominantly from paternal parents occasionally with biparental and maternal inheritance that strikingly contradicts with general patterns documented among angiosperms.

Since uniparental, specifically maternal plastid inheritance is prevalent among angiosperms, ptDNA has been used in understanding evolutionary relationships among plant lineages and in chloroplast genetic engineering to improve agronomic traits (Jansen and Ruhlman 2012; Daniell et al. 2016). The utility of ptDNA has been critical in inferring phylogenetic relationships with numerous applications in plant research including breeding, conservation (Daniell et al. 2016) and biogeography (Ronquist and Sanmartin 2011). However, exceptions in inheritance patterns that include biparental inheritance and occasional paternal leakage could be problematic for phylogenetic inference. When genetic variation exists in the plastomes between parents, paternal leakage or biparental plastid transmission could result in heteroplasmy, a condition with presence of different plastome types in a cell or in an individual. In such cases, if the plastome variation is substantial, heteroplasmy could introduce complications of paralogy resulting in different phylogenetic histories (Wolfe and Randle 2004). The implications of heteroplasmy in phylogeny reconstruction have been documented in several lineages, including *Hyobanche* (Wolfe and Randle 2001, 2004) and *Passiflora* (Hansen et al. 2006). In these cases, conflicting phylogenetic relationships were inferred based on the sampling of heterologous loci. It should be noted that several other processes other than plastid inheritance, such as gene duplication and transfer, could also contribute to heteroplasmy (Wolfe and Randle 2004; Ramsey and Mandel 2019). Nonetheless, recognition of the patterns of plastid transmission could help to achieve correct conclusions when using ptDNA to estimate phylogenetic relationships (Wolfe and Randle 2004; Gonçalves et al. 2020).

Plastid inheritance studies in *Passiflora* have reported uniparental (paternal or maternal) and biparental inheritance. *Passiflora* is the largest genus in Passifloraceae and includes about 560 species grouped into five subgenera, *Astrophea, Decaloba, Deidamioides*, *Passiflora* and *Tetrapathea,* with subgenera *Passiflora* and *Decaloba* each containing more than 200 species (Feuillet and MacDougal 2003; Krosnick et al. 2009; Muschner et al. 2012; Krosnick et al. 2013). The cytological study by Corriveau and Coleman (1988) included a commonly cultivated species *P. edulis* (subgenus *Passiflora*) that exhibited the potential for biparental plastid transmission. Based on three interspecific artificial crosses and one natural hybrid, Muschner et al. (2006) found that hybrids within subgenus *Passiflora* inherited paternal plastids whereas those in subgenus *Decaloba* inherited maternal plastids suggesting distinct inheritance patterns between the subgenera. Hansen et al. (2007) expanded the number of interspecific hybrids from subgenus *Passiflora* and included an intraspecific hybrid from subgenera *Passiflora* and *Decaloba.* The authors found all interspecific hybrids in subgenus *Passiflora* inherited paternal plastids whereas intraspecific hybrids inherited maternal plastids suggesting variation in the inheritance pattern was due to a difference in taxonomic level of the cross. In addition, the increase in number of progeny examined did not affect the observed paternal transmission in interspecific crosses but occasional biparental inheritance was detected in intraspecific hybrids (Hansen et al. 2007). Biparental inheritance in *Passiflora* was also reported in interspecific hybrids from subgenus *Passiflora* in crosses involving *Passiflora menispermifolia,* which displayed a hybrid-bleaching phenotype resulting in hybrid lethality due to plastome-genome incompatibility (Mráček 2005). The remarkable

difference in inheritance patterns observed in *Passiflora* suggests that multiple mechanisms may be involved in plastid inheritance. However, due to limited number of crosses and progeny examined the extent of variation in inheritance patterns is still uncertain. Considering the species diversity in the genus, a comprehensive sampling would improve the understanding of the diversity of modes of plastid inheritance in *Passiflora.* In the present study, plastid inheritance was examined with an increased number of interspecific crosses in subgenera *Passiflora* and *Decaloba* and a single interspecific cross from subgenus *Astrophea.* Polymerase chain reaction (PCR) amplification of ptDNA followed by restriction endonuclease (RE) digestion of amplicons was used to detect the plastid types in the hybrids. Plastid inheritance was examined in embryos of *Passiflora* hybrids together with sampling tissues at the different stages of plant development to understand inheritance as well as retention of plastid types at different phases of plant life cycle.

3.2 Materials and Methods

Passiflora hybrids and seed germination

Passiflora hybrids were generated from field-collected populations grown in greenhouses at The University of Texas at Austin. Interspecific hybrids were generated from self-incompatible *Passiflora* clones whereas self-compatible species were excluded from the study. Crosses were made by brushing pollen from male parents onto the stigma of female parents. The anthers of the female parents were removed prior to pollen maturation. Reciprocal crosses were made when possible. Two approaches were used for seed germination. First, the seed coat was mechanically nicked and seeds were soaked in water

(25°C) for 24 hours and placed between moist paper towels inside a plastic bag in dark until seed germination was observed. The germinated seedlings were transferred in Pro-Mix® growing medium and moved to the greenhouse. Second, seeds were soaked in water (25°C) without scarification for 24 hours and directly transferred to Pro-Mix® growing medium in the greenhouse.

Direct PCR amplification using crude embryo extract

Passiflora hybrid seeds were soaked in water for 24 hours and embryos were excised by removing the seed coat and endosperm under a dissecting microscope. The excised embryos were rinsed in sterile water and stored at -80°C until further use. Stored embryos were resuspended in 88 µl of 50 mM NaOH and homogenized for 1 min in presence of five to six 1 mm glass beads (BioSpec, Inc.) using Mini-BeadBeater-96 (Glen Mills, Inc.). The disrupted embryos were incubated at 95°C for 10 min and neutralized by adding 12 µl of 1 M Tris-HCl (pH 8.0). The crude extract was mixed using a vortex, briefly centrifuged and 5 µl of supernatant was used for PCR.

The PCR amplification of crude embryo extract was carried out using Seed-Direct™ PCR mix (D300, Lamda Biotech, Inc). The PCR reaction contained 5 µl of crude extract, 10 µl of 2X Seed-Direct™ PCR mix, 1 µl of forward and reverse primers and 7 µl of H_20. The PCR program included 5 min at 95°C, followed by 34 cycles of 30 sec at 95°C, 30 sec at 52-56°C, and 1 min at 72°C and final extension of 5 min at 72°C. The annealing temperature of the PCR cycle was adjusted according to the melting temperature of the different primer sets.

DNA isolation and PCR amplification

Total genomic DNA was isolated from young leaves of parents, cotyledons and leaves of hybrids seedlings and also from hybrid embryos that failed the direct PCR amplification. The tissues were flash frozen in liquid nitrogen and homogenized using Geno Grinder ^{TM} (SPEX, Metuchen, NJ) in presence of metallic beads at 1200 rpm for 15 sec. Total genomic DNA was isolated using NucleoSpin® plant II DNA extraction kit (MACHEREY-NAGEL, Düren, Germany). Except for the embryos, qualitative and quantitative assessments of isolated DNA for other tissues were carried out using agarose gel electrophoresis and NanoDrop® (ND-1000, Thermo Scientific).

Isolated DNA was used for PCR amplification in a final reaction volume of 25 µl that contained approximately 200 ng of DNA, 10 µl of FailSafe^{TM} PCR PreMix D (FSP995D, Lucigen), 1 µl of forward and reverse primers, 0.5 µl of Taq polymerase and 5-10 µl of H_20. The PCR program included 5 min at 95°C, followed by 34 cycles of 30 sec at 95°C, 30 sec at 52-56°C, and 1 min at 72°C and final extension of 5 min at 72°C. The annealing temperature of the PCR cycle was adjusted according to the melting temperature of the different primer sets.

Identification of target regions to assess plastid type

Complete plastomes available at the NCBI for 31 *Passiflora* species (Rabah et al. 2019; Shrestha et al. 2019) were used to identify regions that contain high nucleotide variability. The variable regions were amplified and Sanger sequencing was carried out to

determine the sequence for species that lack complete plastome sequences. Primer design for the PCR amplification and Sanger sequencing was carried out with Primer3 (Untergasser et al. 2012) in Geneious v. 11.0.5 (https://www.geneious.com). The newly obtained sequences were aligned with available *Passiflora* plastome data using MAFFT (Katoh and Standley 2013) in Geneious v.11.0.5. The ptDNA target regions were selected based on presence of polymorphic restriction sites between the parents. The target regions for the parents were amplified followed by restriction digestion and visualized in agarose gels and subsequently the process was repeated for hybrids. Banding patterns observed for parents prior and post digestion were used as references to assess plastid types in hybrids. Restriction digestions of amplified ptDNA were performed for variable time periods at the optimal temperature recommended by the restriction enzyme (RE) suppliers (New England Biolabs and Thermo Scientific). Restriction digestions of amplicons were carried out directly without prior purification of the PCR products and visualized in 1-2% agarose gels stained with RedSafe™ (iNtRON Biotechnology). One kb DNA ladder (N3232L, New England Biolabs, Inc.) was used as a marker to estimate fragment sizes for the PCR products prior and post digestion.

3.3 Results

Crosses were considered successful only when the pollen transfer from male parent onto the stigma of female parent produced a fruit that contained seeds with embryos (Figure 3.1). With this criterion, the number of successful crosses was substantially reduced compared to the number of crosses attempted. Forty-four crosses from the three subgenera

Passiflora, *Decaloba* and *Astrophea* were examined. Most crosses (37) were confined to subgenus *Passiflora* and included four reciprocal crosses (Table 3.1). Seven successful crosses were examined in subgenus *Decaloba*, two of which involved reciprocal crosses. In subgenus *Astrophea* only a single interspecific cross was successful. Multiple attempts to generate hybrids between species in subgenus *Deidamioides* were unsuccessful. Several ptDNA target regions were selected for PCR and RE analysis with the *rpl32-trnL* region being the most useful for assessing plastid types. Detailed information on *Passiflora* species used to generate hybrids, direction of crosses, ptDNA target regions, accession numbers for sequences, primers used to amplify target regions, restriction enzymes and vouchers are provided in Tables 3.2-3.5.

Plastid inheritance in hybrid embryos

Substantial differences in the embryo size were noted between hybrids in different subgenera such that the embryos in subgenus *Decaloba* were generally much smaller compared to embryos in subgenera *Passiflora* and *Astrophea* (Figure 3.1). Direct PCR using crude extract from embryos successfully amplified ptDNA target regions for most of the hybrids except three, *P. nephrodes* (♀) x *P. oerstedii* (♂), *P. lancetillensis* (♀) x *P. microstipula* (♂) and *P. microstipula* (♀) x *P. lancetillensis* (♂). Therefore, DNA was isolated from embryos prior PCR amplification, however *P. nephrodes* (♀) x *P. oerstedii* (♂) failed even after DNA isolation, thus the cross was excluded from the analyses. Plastid assessment in the embryos exhibited uniparental (paternal or maternal) and biparental modes of inheritance in subgenus *Passiflora* (Figures 3.2-3.3). Plastid inheritance can be

grouped as purely paternal, maternal, biparental or a combination of any of the three patterns. For the 37 interspecific crosses examined, hybrids from 24 crosses displayed only paternal inheritance and the remaining crosses had maternal and/or biparental inheritance (Table 3.1). When total embryos in subgenus *Passiflora* were taken in account, 280 embryos inherited paternal plastids, 45 embryos had maternal inheritance and 47 embryos had biparental inheritance. A general inheritance pattern was observed with predominant paternal inheritance and occasional maternal and biparental inheritance. In some cases maternal or biparental inheritance was the only observed inheritance mode. Solely maternal transmission was found in two hybrids, *P. menispermifolia* (9224, ♀) x *P. oerstedii* (♂) and *P. nephrodes* (♀**)** x *P. choconiana* (♂); however, their reciprocal crosses had exclusively paternal inheritance (Table 3.1). This indicated that *P. menispermifolia* and *P. nephrodes* plastids were inherited in their hybrids regardless of the direction of the crosses. Solely biparental plastid inheritance was restricted two crosses, *P. oerstedii* (♀) x *P. menispermifolia* (8039, ♂) and *P. hastifolia* (♀) x *P. menispermifolia* (8039, ♂). In addition, predominant biparental inheritance was also noted in other hybrids with paternal plastid inheritance (Table 3.1).

Based on six interspecific crosses in subgenus *Decaloba,* maternal inheritance was predominant in most hybrids with frequent biparental and rare paternal inheritance (Table 3.1). A total of 31 embryos exhibited maternal inheritance, 26 embryos had biparental inheritance and only two embryos had paternal inheritance. Two hybrids, *P. organensis* (♀) x *P. biflora* (♂) and *P. microstipula* (♀) x *P. lancetillensis* (♂), exhibited solely maternal inheritance in all the embryos examined while other crosses had predominantly maternal

with biparental and paternal inheritance less common (Table 3.1). In addition, a few interspecific hybrids in subgenus *Decaloba* displayed predominantly biparental inheritance. A single interspecific hybrid, *P. sphaerocarpa* (♀) x *P. pittieri* (♂) from subgenus *Astrophea*, had predominantly paternal inheritance with a single embryo with maternal inheritance. Examples of paternal, maternal and biparental inheritance detected in the hybrid embryos are shown in Figure 3.4, including hybrids of *P. menispermifolia* (♀) x *P. miersii* (♂) representing subgenus *Passiflora* and *P. rufa* (♀) x *P. auriculata* (♂) from subgenus *Decaloba.* Gel images of restriction digestion for all parent species and hybrid embryos are shown in Figures 3.2 and 3.3, respectively.

Plastid inheritance in seedlings and older plants

When possible, plastid inheritance was assessed using different tissues from seedlings including cotyledon, first leaf and subsequent leaves, to identify plastid type at the early stages of plant development. When plastid inheritance was biparental in the early developmental stage, if feasible, further assessment was continued using subsequent emerging leaves. For the hybrids with uniparental plastid inheritance detected in seedlings, no further assessments were carried out at older stages. For some hybrids, tissues were not collected at the early stages and plastid inheritance was assessed using leaves from year old hybrids, which we refer here as leaves from older plants.

Seed germination for hybrids was substantially lower compared to the number of successful crosses generated, and as a consequence the total number of interspecific crosses were limited to 10 hybrids in subgenus *Passiflora* and six in subgenus *Decaloba* including

a reciprocal cross (Table 3.6). Within subgenus *Passiflora* paternal inheritance was detected in all the progeny of the hybrids assessed (Figure 3.5). The hybrids that displayed paternal inheritance in the cotyledons also had paternal inheritance in the embryos. The hybrids of *P. menispermifolia* (8039, ♀) x *P. hastifolia* (♂), *P. nephrodes* (♀) x *P. sprucei* (♂) and *P. kermesina* (♀) x *P. miersii* (♂), which had maternal and biparental inheritance in embryos, displayed solely paternal inheritance in the seedlings.

Hybrids in subgenus *Decaloba* displayed all three modes of plastid inheritance in their seedlings with most hybrids displaying predominantly maternal inheritance with occasional biparental inheritance and rarely paternal inheritance. Solely maternal inheritance was detected in all the seedlings for two hybrids, *P. organensis* (♀) x *P. biflora* (♂) and *P. microstipula* (♀) x *P. lancetillensis* (♂), which was consistent with solely maternal inheritance detected in the embryos (Figure 3.5). Three hybrids, *P. rufa* (♀) x *P. auriculata* (♂), *P. lancetillensis* (♀) x *P. microstipula* (♂) and *P. misera* (9023, ♀) x *P. misera* (9335, ♂) displayed uniparental (paternal or maternal) or biparental inheritance in the seedlings, which was also present in the embryos. Biparental inheritance detected in the hybrids was limited to seedlings, mainly during early plant development (Table 3.6; Figure 3.6). Biparental inheritance was detected in the cotyledons for all three seedlings analyzed for the hybrid *P. lancetillensis* (♀) x *P. microstipula* (♂). However, among the three seedlings examined only two seedlings retained plastids from both parents in the first leaf and only maternal plastids were detected in the first leaf of the third seedling (Figure 3.6B). All three seedlings died and plastid assessment in the later developmental stages was not possible. Therefore, plastid inheritance was assessed using the leaves of one-year-old *P.*

lancetillensis (♀) x *P. microstipula* (♂) hybrids in four individuals, which represent plastids at the later developmental stage. In the older *P. lancetillensis* (♀) x *P. microstipula* (♂) hybrids, maternal plastids were detected in three individuals and paternal in one individual (Figure 3.6C).

A thorough assessment of plastid inheritance at different developmental stages was carried out for the hybrid *P. misera* (9023, ♀) x *P. misera* (9335, ♂) (Figures 3.6E-H). Among the seven seedlings examined five inherited plastids biparentally and two inherited maternal plastids in their cotyledons. When the first leaf of these seedlings was examined, maternal plastids were detected in four and biparental inheritance was present in three (Figure 3.6F-G). Among the three seedlings with biparental inheritance (seedlings 4, 6 and 7) seedling 4 showed the paternal plastid as the dominant type compared to remaining two seedlings. Subsequent examination of plastid type in the second and third leaves of seedling 4 showed retention of only the paternal plastid type. In contrast, similar assessment of plastid types using leaves from different developmental stages in the seedling 6 displayed retention of maternal plastid type (Figure 3.7H). Plastid types from both parents were detected up to second leaf in the seedling 6 but only the maternal plastid type was detected in leaf 3 and in older leaves (Figure 3.7H). Similar observations, presence of plastids from both parents up to second leaf and retention of maternal plastid afterwards, were also noted in seedling 7.

Hybrids that displayed incompatibility phenotypes were primarily restricted to subgenus *Decaloba* with a single case observed in subgenus *Passiflora.* The hybrid *P. menispermifolia* (8039, ♀) x *P. P. hastifolia* (♂) from subgenus *Passiflora* exhibited a

dwarf phenotype with disrupted growth, severely reduced leaf size and altered leaf morphology (Figure 3.8A). Within subgenus *Decaloba*, all progeny of *P. organensis* (♀) x P. *biflora* (♂) had normal green cotyledons but displayed white or whitish-green leaves that died within a few of weeks (Figure 3.7A-B). Hybrid lethality was also observed in two hybrids, *P. rufa* (♀) x *P. auriculata* (♂) and *P. rufa* (♀) x *P. jatunsachensis* (♂). Both hybrids produced albino seedlings that senesced without producing leaves or even cotyledons in some progeny (Figure 3.7C-D). A variegated phenotype was observed for *P. microstipula* (♀) x *P. lancetillensis* (♂) only in the cotyledons (Figure 3.7E) and their reciprocal cross, *P. lancetillensis* (♀) x *P. microstipula* (♂), displayed variegation in the cotyledons as well as in leaves in which white sectors gradually disappeared as the plant developed (Figure 3.7F). Similarly, hybrid variegation was also observed in *P. misera* (9023, ♀) x *P. misera* (9335, ♂), mostly in the cotyledons and occasionally in the first leaf (Figure 3.7G-H).

3.4 Discussion

The presence of ptDNA in generative or sperm cells indicates potential transmission of paternal plastids; hence, staining of ptDNA in pollen grains has been used to examine plastid inheritance in a large group of angiosperms (Corriveau and Coleman 1988; Zhang and Sodmergen 2003). However, this method does not confirm that paternal plastids are transmitted into a zygote, as multiple mechanisms behind paternal plastid exclusions occur during gametogenesis, fertilization and post fertilization (Sears 1980; Hagemann and Schröder 1989; Birky 1994, 2001; Mogensen 1996; Nagata 2010). Low frequency

transmissions of paternal plastids are also susceptible to random drift reducing heteroplasmy during embryo development (Tilney-Bassett and Birky 1989; Birky 1994). Similarly, biparental plastid inheritance during early plant development is known to segregate or sort-out and resulting in homogenous plastomes (Birky 1994, 2001; Matsushima et al. 2008). This suggests that examination of plastid types in older plants may only detect retained plastids and does not reflect how plastids are initially transmitted. Therefore, plastid inheritance in *Passiflora* hybrids was investigated by assaying different tissues, such as embryos, cotyledons and leaves to capture inheritance and retention of plastids at the different stages of plant development. To assess the plastid types in hybrids, RE digestion of PCR amplicons was performed, a method utilized in previous studies of other angiosperms (Cruzan et al., 1993; Yang et al. 2000; Trusty et al. 2007). It is possible that trace amounts of ptDNA due to extremely low transmission may fail to amplify or amplify with lower yield resulting in an incorrect characterization of the mode as uniparental. Nonetheless, the method will only underestimate the actual cases of biparental plastid inheritance. Therefore, the number of cases of biparental inheritance reported for *Passiflora* hybrids in this study may be underestimated.

Clade-specific mode of inheritance

The results of plastid inheritance in *Passiflora* hybrids can be roughly summarized into four categories; (i) predominant paternal inheritance in subgenus *Passiflora*, (ii) predominant maternal inheritance in subgenus *Decaloba*, (iii) solely maternal inheritance in few distinct hybrids in subgenus *Passiflora* and (iv) predominant biparental inheritance in

some hybrids in subgenera *Passiflora* and *Decaloba* (Tables 3.1 and 3.6). The dichotomy of paternal inheritance in subgenus *Passiflora* and maternal inheritance in subgenus *Decaloba* is in agreement with the previous observation based on few interspecific hybrids (Muschner et al. 2006). Paternal transmissions are commonly observed within interspecific crosses or crosses that include divergent parents within a population, which is attributed to failure in the mechanism to prevent paternal leakage (Cruzan et al. 1993; Reboud and Zeyl 1994; Yang et al. 2000; Xu 2005; Nagata 2010). Paternal inheritance in subgenus *Passiflora* also supports the correlation between paternal inheritance and interspecific crosses; however, maternal inheritance observed for the interspecific hybrids in subgenus *Decaloba* contradicts this phenomenon. Hansen et al. (2007) suggested that there is a dichotomy of paternal and maternal inheritance in *Passiflora* between inter- and intraspecific crosses, respectively. The correlation of maternal inheritance with intraspecific crosses was based on only two hybrids from subgenera *Passiflora* and *Decaloba*, *P. pseudo-oerstedii* x *P. oerstedii* and *P. costaricensis* from two geographical locations, respectively (Hansen et al. 2007). *Passiflora pseudo-oerstedii* is now considered as the distinct species *P. dispar* (Ulmer and McDougal 2004); therefore, *P. pseudo-oerstedii* x *P. oerstedii* is not an intraspecific hybrid. Thus only a single intraspecific cross from subgenus *Decaloba* has been performed and showed maternal plastid inheritance. Hence, the conclusion that maternal inheritance is correlated with intraspecific crosses in *Passiflora* is tenuous since it is based on a single observation. In the present study, maternal inheritance is the predominant mode of plastid inheritance even in interspecific crosses in subgenus *Decaloba.* Therefore, regardless of inter- or intraspecific crosses, maternal inheritance is

likely to be the dominant mode of inheritance in this subgenus. Hybrids between *P. misera* (9023) and *P. misera* (9335) in this study are unlikely to represent intraspecific crosses as the two *P. misera* are very distinct in morphology of flowers, leaves and stems (Figure 3.8J-L). *Passiflora misera* is known for its conspicuously flattened stems lacking showy flowers (Vanderplank 1996) but the species is considered to be highly variable and it was suggested that multiple species may be represented under the name *P. misera* (Boender 2019). Among the two *P. misera* examined in the present study, *P. misera* (9023) has notably flattened stems with flowers similar to commonly described *P. misera*, whereas *P. misera* (9335) has square stems and morphologically distinct flowers (Figure 3.8J-L). This suggests that these two accessions should be treated as distinct species. Therefore, hybrids from the crosses between *P. misera* (9023) and *P. misera* (9335) should not be considered intraspecific.

Paternal inheritance is considered an anomaly among angiosperms but cases have been reported in a limited number of lineages including *Actinidia, Turnera* and *Medicago* (Schumann and Hancock 1989; Shore and Triassi 1998; Chat et al. 1999). It is noteworthy that the pattern of plastid inheritance observed in subgenus *Passiflora* is analogous to intraspecific crosses in *Turnera* (Shore et al. 1994; Shore and Triassi 1998), which is now classified as the subfamily Turneroideae in Passifloraceae (APG 2016). Interspecific crosses in subgenus *Passiflora* and *Turnera* display predominant paternal plastid transmission with occasional maternal or biparental inheritance. It is possible that intraspecific crosses in subgenus *Passiflora* also follow paternally biased plastid inheritance and thus represent one of the few angiosperm lineages with predominant paternal

inheritance. This pattern could be tested by performing more intraspecific crosses in subgenus *Passiflora.* The similarity in the inheritance patterns between two subfamilies of Passifloraceae suggests that paternal inheritance may be the norm in these closely related taxa regardless of whether the cross is within or between species. Within *Passiflora* there may have been a shift towards predominantly maternal inheritance specifically in subgenus *Decaloba.* A single interspecific hybrid from subgenus *Astrophea* also shows predominantly paternal inheritance similar to species in subgenus *Passiflora.* However, the result should be cautiously interpreted as a general pattern of inheritance for subgenus *Astrophea* because of the small sample size and the nature of the cross. Since the present study lacks intraspecific crosses in *Passiflora*, results based on interspecific crosses will not be sufficient to explain the variation noted in plastid inheritance patterns. The hypothesis of paternal and maternal inheritance associated with inter- and intraspecific crosses, respectively, in *Passiflora* requires evidence from additional intraspecific crosses. Interspecific crosses suggest that species in subgenus *Passiflora* exhibit predominant paternal plastid inheritance whereas species in subgenus *Decaloba* predominantly inherit maternal plastids.

Occurrence of biparental plastid inheritance

Two separate studies presented molecular evidence of biparental inheritance in *Passiflora;* first, in an interspecific cross between *P. menispermifolia* and *P. oerstedii* and their reciprocal cross from subgenus *Passiflora* (Mráček 2005) and, second, in an intraspecific cross of *P. costaricensis* from two geographical locations (Hansen et al. 2007).

The F1 hybrids of *P. menispermifolia* and *P. oerstedii* displayed a bleached phenotype with white and green sectors in leaves containing plastomes of *P. menispermifolia* and *P. oerstedii*, respectively, and non-differentiated plastids of *P. menispermifolia* indicated incompatibility with F1 nuclear genotype (Mráček 2005). In contrast, Hansen et al. (2007) did not report biparental inheritance but instead found strictly paternal inheritance in hybrids including *P. menispermifolia* and *P. oerstedii* as parents and in backcrosses. It is noteworthy that *P. oerstedii* illustrated in Mráček (2005) has deeply trilobed leaves that differ from commonly described *P. oerstedii* with oblong-ovate simple leaves included in Hansen et al. (2007). Ulmer and MacDougal (2004) suggested two species, *P. oerstedii* var. *choconiana* (also known as *P. choconiana*) and *P. purpusii* with trilobed leaves, are erroneously identified under the name *P. oerstedii*. Therefore, the difference in plastid inheritance noted in these two studies is likely due to incorrect identification of the parent species. Crosses in the present study with *P. menispermifolia* and *P. oerstedii* as parents, the results depended on varieties being used. The crosses between *P. menispermifolia* (9224) and *P. oerstedii* inherited plastids from *P. menispermifolia* (9224) regardless of the direction. However, the cross including a different accession of *P. menispermifolia* (8039) displayed biparental plastid inheritance in all hybrid embryos (Table 3.1; Figure 3.3). Although nothing is known about the nuclear genomes of *P. menispermifolia* and *P. oerstedii,* it may be noteworthy to consider plastome size variation between the species. The *P. menispermifolia* plastome is substantially smaller (14 kilobases) than *P. oerstedii* and the size difference may play a role in efficient plastome replication resulting in an advantage over the *P. oerstedii* plastome during plastid transmission. It has been proposed

that speed of plastid multiplication plays a role in transmission of plastids in *Oenothera* (Chiu et al. 1988).

In the present study, heteroplasmy due to biparental inheritance was detected specifically in the embryos and seedlings of *Passiflora* hybrids but not in older plants, which is a clear distinction from previous findings. During plant development biparentally inherited plastids in *Passiflora* hybrids segregated and excluded the plastid types from either parent resulting in homogeneity. This observation is very similar to the segregation of biparentally inherited plastids in *Medicago* (Matsushima et al. 2008). The vegetative segregation in *Passiflora* hybrids is complete in the F1 generation similar to most *Medicago* hybrids. However, in some *Medicago* hybrids segregation is not complete until the F2 generation. Sorting-out of plastid types observed in *Passiflora* and *Medicago* supports the prediction that the completion of vegetative segregation occurs within a generation (Birky 2001). In fact, homogeneity of plastid types in *Passiflora* is attained in the first leaf for most hybrids but in a few cases the process did not occur until the second or third leaves indicating the rate of segregation differs among hybrid individuals (Figure 3.6F and H). Out of 11 progenies examined from two hybrids, *P. lancetillensis* (♀) x *P. microstipula* (♂) and *P. misera* (9023, ♀) x *P. misera* (9335, ♂), only two hybrids from each cross retain paternal plastid in the older plants (Figure 3.6C and F) and the remaining ended with maternal plastid types. Vegetative segregation strongly favoring maternal plastid type was also reported in *Medicago* (Matsushima et al. 2008). Plastomes are referred as relaxed genomes that are under stochastic replication and division suggesting that homogeneity toward the maternal plastid could result from a stochastic process (Birky

1983, 1994, 2001). The number of maternal plastids or plastomes per plastid may be substantially higher due to the stochastic replication and division favoring maternal plastid over paternal in *Passiflora* hybrids. It is also possible that the preference of maternal plastid over paternal type could be due to selection. The limited number of crosses and progeny in the present study does not allow a conclusion regarding whether the preference of maternal parent is due to selection or a consequence of a random process. Nonetheless, vegetative segregation of biparentally inherited plastids as observed in *Passiflora* seedlings mandates that the modes of plastid inheritance need to be examined early during plant development and the assessment using older plants may not accurately reflect inheritance patterns.

Hybrid incompatibility phenotypes

Hybrid variegation in *Passiflora* seedlings was observed for three crosses, *P. microstipula* (♀) x *P. lancetillensis* (♂), *P. lancetillensis* (♀) x *P. microstipula* (♂) and *P. misera* (9023, ♀) x *P. misera* (9335, ♂) (Figure 3.7E-H). Variegation in the cotyledons and leaves of *P. misera* (9023, ♀) x *P. misera* (9335, ♂) often correlates with segregation of biparentally inherited plastids (Figures 3.6F-H; 3.7G-H). Excising tissues from different sectors and assessing plastid types in each sector could confirm if variegation is due to plastid type. Variegation was observed in the cotyledons of *P. microstipula* (♀) x *P. lancetillensis* (♂) but heteroplasmy was not detected in the leaves (Figure 3.7E). Similarly, the reciprocal cross *P. lancetillensis* (♀) x *P. microstipula* (♂) displayed variegation in cotyledons as well as leaves and the white sectors gradually disappeared as the plants developed (Figures 3.7F; 3.8F-H). Heteroplasmy was detected in this hybrid only in the

cotyledons but not in the leaves (Seedling 3, Figure 3.6B). This indicates that in hybrids with *P. lancetillensis* and *P. microstipula* as parents variegation is unlikely to be associated with the segregation of plastids but more likely to be mediated by another mechanism under nuclear influence (Kirk and Tilney-Bassett 1978; Rodermel 2002). Many species from subgenus *Decaloba* are known for conspicuous leaf variegation in younger and/or older plants (Ulmer and MacDougal 2004; Krosnick et al. 2013). Thus additional evidence is needed to confirm hybrid variegation is associated with vegetative segregation in the genus.

Passiflora hybrids that displayed incompatibility phenotypes were mostly within subgenus *Decaloba.* Hybrid incompatibility resulted in hybrid inviability for three hybrids, *P. organensis* (♀) x P. *biflora* (♂), *P. rufa* (♀) x *P. jatunsachensis* (♂) and *P. rufa* (♀) x *P. auriculata* (♂) (Figures 3.7A-D; 3.8B-E). Two hybrids with *P. rufa* inherited uniparental (maternal or paternal) or biparental plastids, and regardless of the mode of plastid inheritance all progeny displayed an albino phenotype and failed to produce leaves. The white/pale phenotypes in hybrids are generally associated with lower chlorophyll content with retarded thylakoid development as a consequence of reduced photosynthesis (Greiner et al. 2011). Hybrid inviability observed in these *Passiflora* hybrids is most likely under the influence of the nuclear genome as the phenotype was consistent regardless of the plastid inheritance pattern. A similar phenotype with albino/whitish-green leaves was observed in *P. organensis* (♀) x P. *biflora* (♂) hybrids that solely inherited maternal plastids in all progeny, although the hybrids produced green, healthy cotyledons. This suggests that the incompatibility is associated with leaf development, perhaps due to failure of proper assembly of photosynthetic components, possibly due to incompatible interaction between

plastid and nuclear genomes. It has been suggested that postzygotic reproductive barriers can be generally predicted from the observed phenotype (Rieseberg and Blackman 2010). Post-zygotic barriers that leads to hybrid inviability usually result when co-adapted loci accumulate mutations independently in their lineages and subsequent reunion of the incompatible alleles in hybrids results in a negative interaction as explained by Dobzhansky-Muller model (reviewed in Bomblies and Weigel 2007; Rieseberg and Blackman 2010). This model has been further extended for co-adapted loci in two cellular compartments, nucleus and plastid, which are disrupted in hybrids leading to plastome-genome incompatibility (Greiner et al. 2011). The availability of complete plastomes for *Passiflora* hybrids that displayed the incompatibility phenotype could be used to identify underlying plastome regulatory elements associated with the hybrid incompatibility.

3.5 Conclusions

Clade-specific plastid inheritance along with frequent occurrence of biparental inheritance makes *Passiflora* as an intriguing system to study the underlying evolutionary mechanisms associated with plastid inheritance. The diverse inheritance patterns within a single genus suggest multiple evolutionary forces may be in play. Biparental inheritance is thought to be advantageous for rescuing defective plastids and slowing down the accumulation of mutations through strictly uniparental inheritance in angiosperms (Zhang and Sodmergen 2010; Greiner et al. 2014). This may explain the low frequency of biparental inheritance and paternal/maternal leakage detected in *Passiflora* but is not adequate to explain predominant paternal inheritance in subgenus *Passiflora.* It is unlikely

that the failure of mechanisms to prevent paternal plastid leakage entirely contributed to predominant paternal inheritance in *Passiflora*. Detection of paternally biased plastid transmission also observed in closely related subfamily Turneroideae suggests that paternal inheritance may be the ancestral condition in these lineages. Further study with conventional microscopy comparing pollen development in species from subgenera *Passiflora* and *Decaloba* could provide essential insights in the mechanism of plastid inheritance in *Passiflora*.

Self-incompatibility is common in *Passiflora* and many species outcross easily and viable F1 hybrids that can backcross with either parent (Ulmer and MacDougal 2004; Hansen et al. 2007) and also with other species generating viable new hybrids (BS and LEG, personal observation). Surprisingly, only a few natural hybrids have been reported (Killip 1938; Lorenz-Lemke et al. 2005). The ability to hybridize without difficulty coupled with diverse plastid inheritance patterns could be problematic in resolving phylogenetic relationships. In fact, a conflicting relationship was identified in *Passiflora* phylogenetic studies using plastome loci due to the heteroplasmy, suggesting that researchers should be cautious in interpreting phylogenetic relationship in the genus using plastid loci (Hansen et al. 2006). In the present study, although heteroplasmy was restricted to seedlings and vegetative segregation led to retention of either maternal or paternal plastid types in later stages of development, this phenomenon could still contribute to conflicting phylogenetic relationship in *Passiflora* as suggested by Hansen et al. (2007). It is notable that cytonuclear gene transfer in *Passiflora* is widespread and duplicated plastid loci co-exist in two cellular compartments under different evolutionary constraints (Shrestha et al. 2020).

As a consequence, this may also introduce conflicts associated with paralogy while inferring phylogenetic relationships, thus, one should be cautious using plastid loci for phylogeny in *Passiflora.*

Multiple modes of plastid inheritance in *Passiflora* could also play a role in plastome evolution in the genus. *Passiflora* plastomes are highly rearranged with extensive structural variation and highly accelerated substitution rates in certain genes and clades (Rabah et al. 2019; Shrestha et al. 2019; Cauz-Santos et al. 2020). Although recombination between plastomes from different plastids is unknown, empirical evidence of plastome recombination has been documented in protoplast fusion experiments (Medgyesy et al. 1986; Thanh and Medgyesy 1989). Evidence of plastome structural rearrangements correlated with biparental inheritance has been recognized in angiosperms (Jansen and Ruhlman 2012) and isogamous algae (Choi et al. 2020). Therefore, it will be interesting to examine whether the plastid inheritance pattern has had any role in the unusual rearrangements in *Passiflora* plastomes.

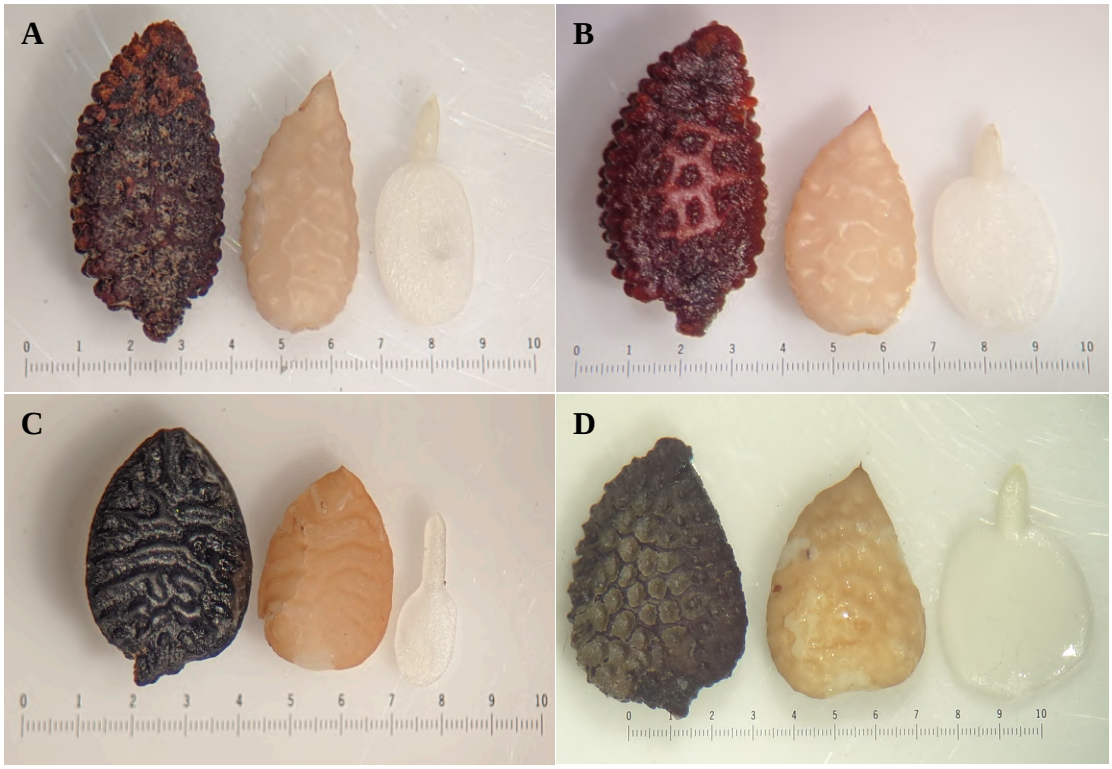


Figure 3.1. *Passiflora* hybrid seeds under dissecting microscope at 10X magnification. (A) *P. menispermifolia* (8039, ♀) x *P. menispermifolia* (Sirena, ♂) progenies for the comparison with interspecific hybrid in B that includes same maternal parent. (B) Interspecific hybrid *P. menispermifolia* (8039, ♀) x *P. miersii* (♂) from subgenus *Passiflora*. (C). Interspecific hybrid *P. rufa* (♀) x *P. auriculata* (♂) from subgenus *Decaloba*. (D) Interspecific hybrid *P. sphaerocarpa* (♀) x *P. pittieri* (♂) from subgenus *Astrophea*. Left to right, seeds soaked in water for 24 hours, seed without seed coat and embryo only.

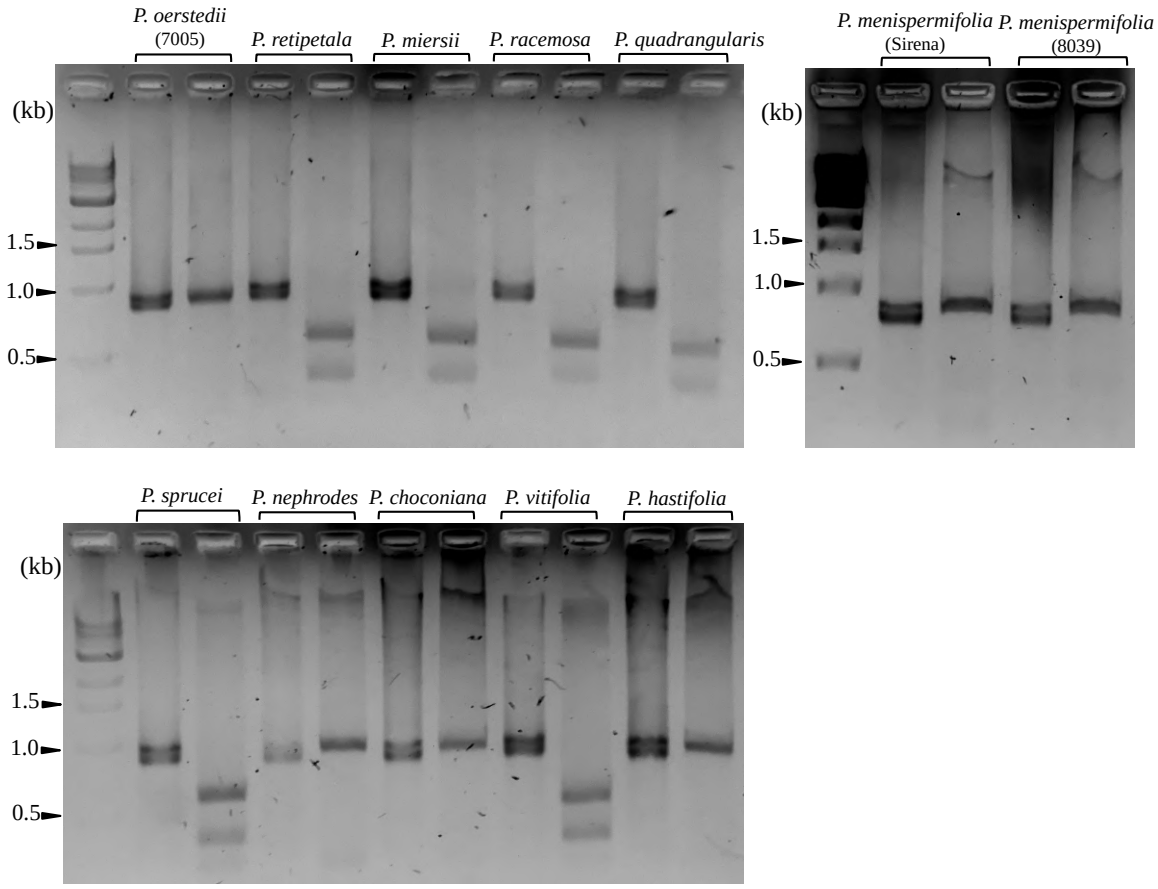


Figure 3.2.1. Agarose gels of PCR amplified *rpl32-trnL* region following digestion with NdeI endonuclease in *Passiflora* parents. For each species, PCR amplicons of *rpl32-trnL* are on the left lane and products after NdeI digestion are on the right lane. The restriction digestion was carried out at 37°C for 12 hrs. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe™ (INtRON Biotechnology).

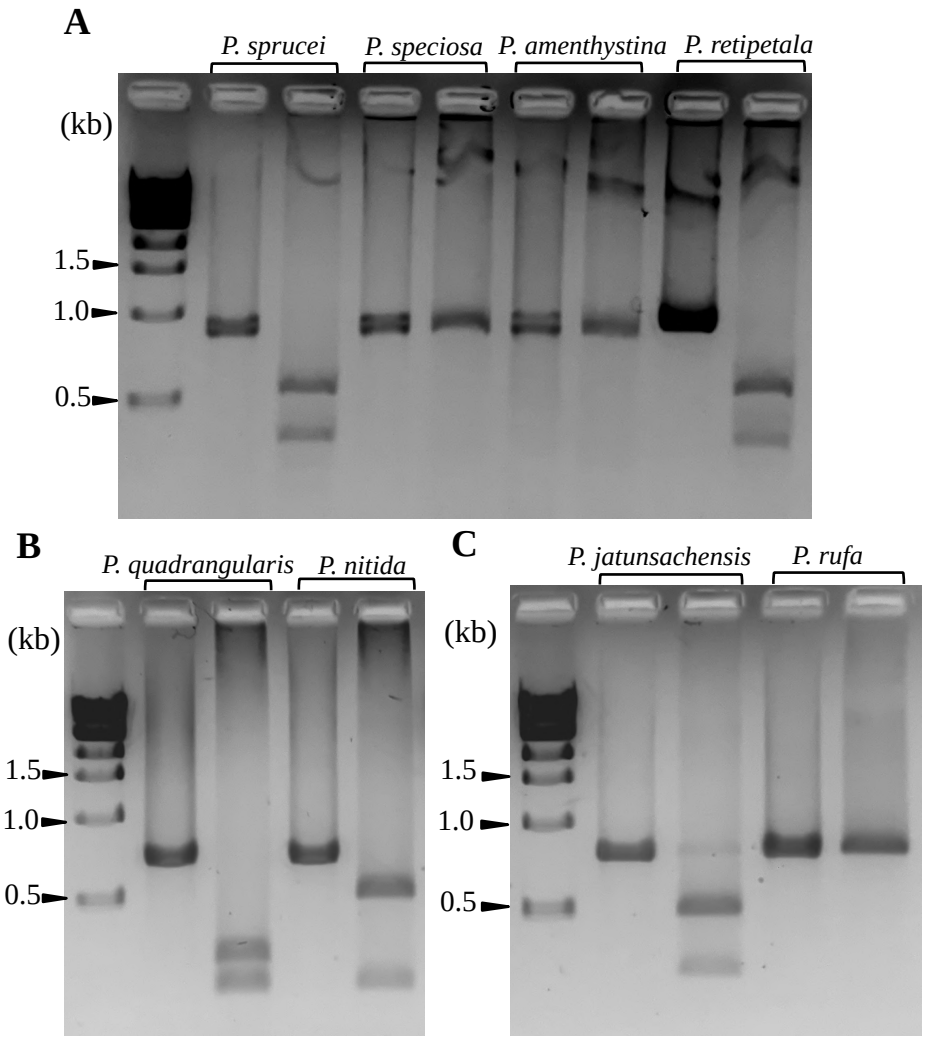


Figure 3.2.2. Agarose gels of PCR amplified various target regions (A, *rpl32-trnL*; B, *atpF-atpH*; C, *ycf4-psaI*) following digestion with Bsp119I endonuclease in *Passiflora* parents. For each species, PCR amplicons of the target regions are on the left lane and products after Bsp119I digestion are on the right lane. The restriction digestion was carried out at 37°C for 8 hrs. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe™ (INtRON Biotechnology).

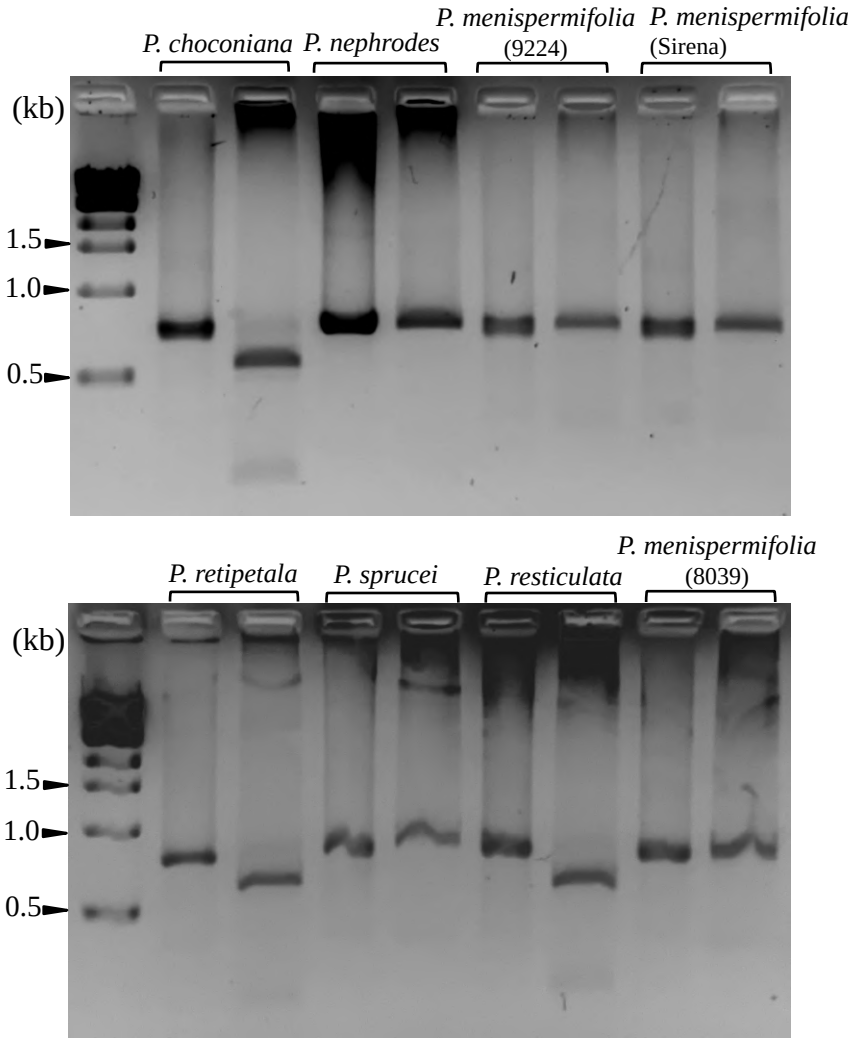


Figure 3.2.3. Agarose gels of PCR amplified intergenic-*trnL-1* region following digestion with EcoRI endonuclease in *Passiflora* parents. For each species, PCR amplicons of intergenic-*trnL-1* are on the left lane and products after EcoRI digestion are on the right lane. The restriction digestion was carried out at 37°C for 12 hrs. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe^{TM} (INtRON Biotechnology).

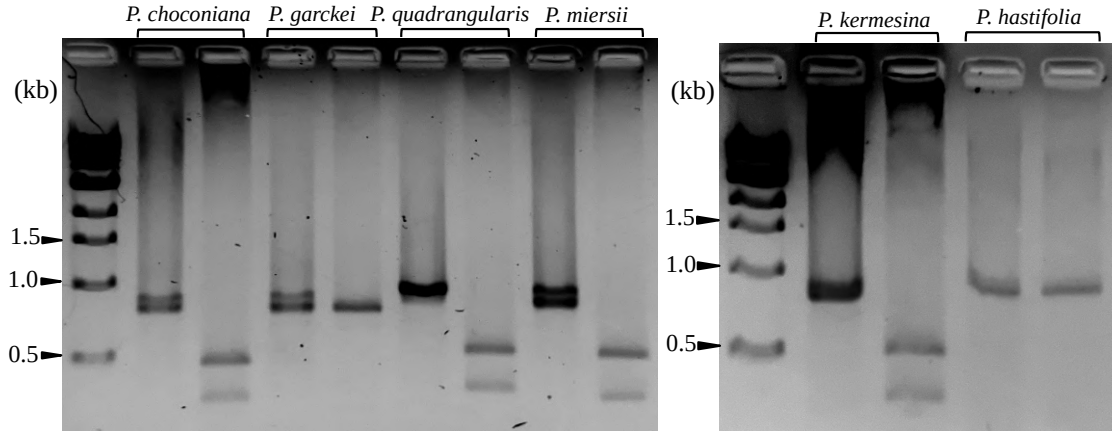


Figure 3.2.4. Agarose gels of PCR amplified *rpl32-trnL* region following digestion with HindIII endonuclease in *Passiflora* parents. For each species, PCR amplicons of *rpl32-trnL* are on the left lane and products after HindIII digestion are on the right lane. The restriction digestion was carried out at 37°C for 12 hrs. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe™ (INtRON Biotechnology).

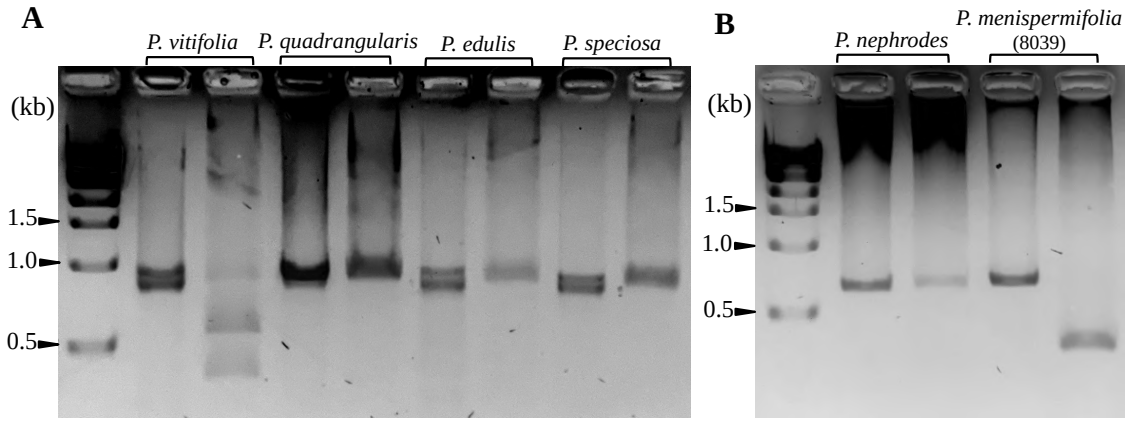


Figure 3.2.5. Agarose gels of PCR amplified target regions (A, *rpl32-trnL*; B, *accD-rbcL*) following digestion with SwaI endonuclease in *Passiflora* parents. For each species, PCR amplicons are on the left lane and products after SwaI digestion are on the right lane. The restriction digestion was carried out at 25°C for 8 hrs. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe™ (INtRON Biotechnology).

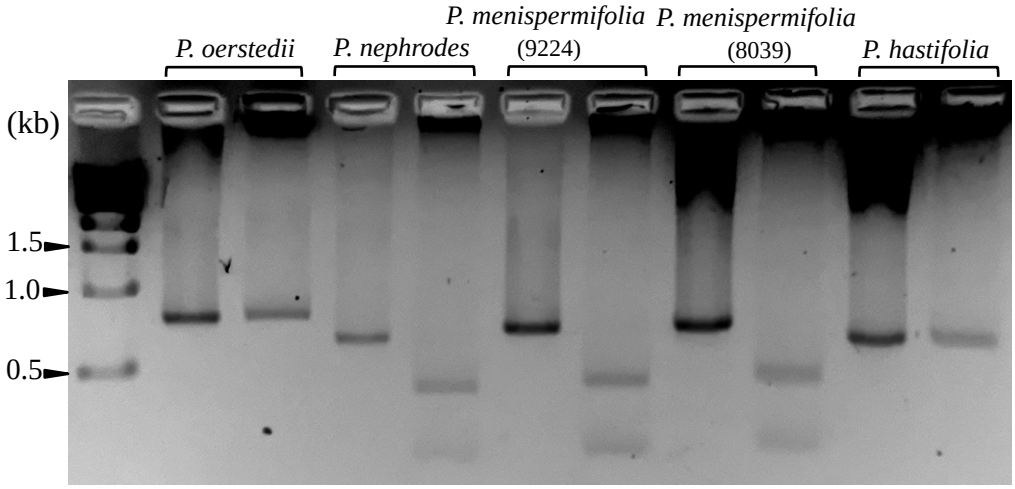


Figure 3.2.6. Agarose gels of PCR amplified *accD-rbcL* intergenic region following digestion with BsmI endonuclease in *Passiflora* parents. For each species, PCR amplicons are on the left lane and products after BsmI digestion are on the right lane. The restriction digestion was carried out at 65°C for 2:30 hrs. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe^{TM} (INtRON Biotechnology).

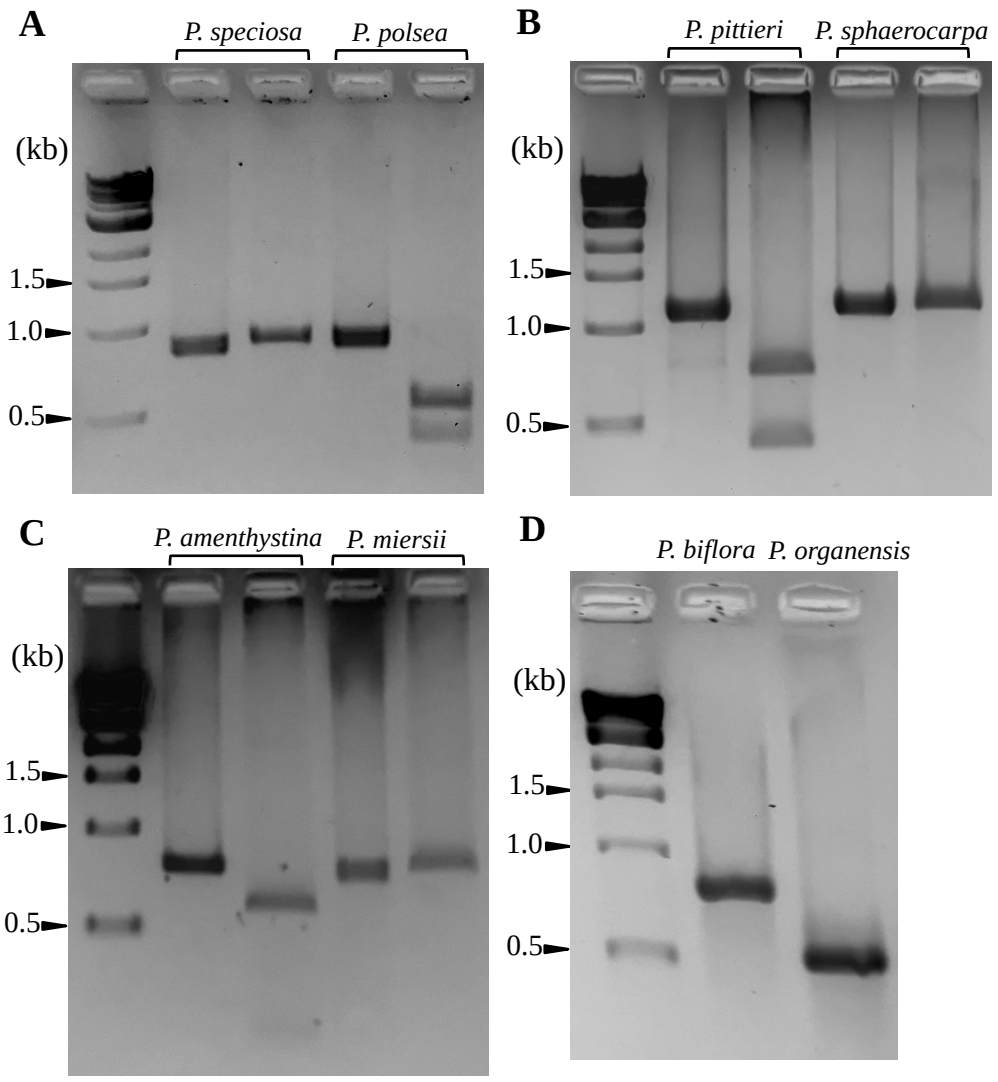
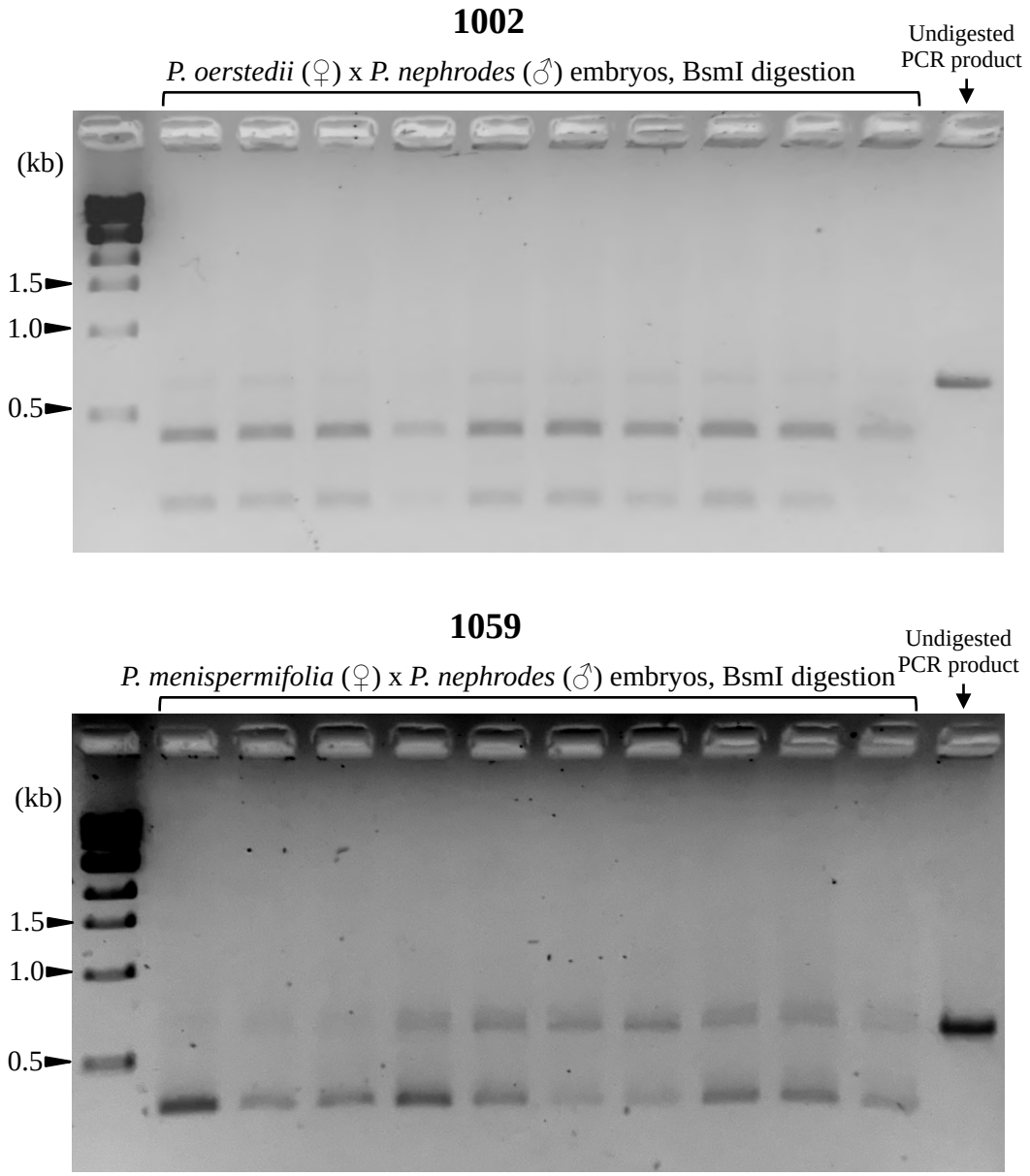
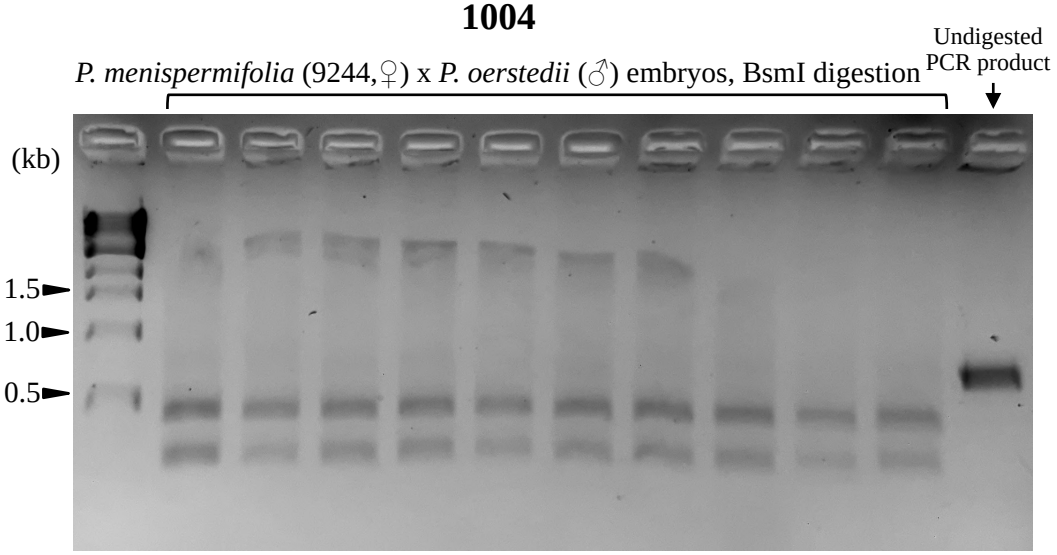


Figure 3.2.7. Agarose gels of PCR amplified various target regions following restriction digestion in *Passiflora* parents. (A) The intergenic region *rpl32-trnL* was amplified and digested with endonuclease AflII. (B) The intergenic region *rpl32-trnL* was amplified and digested with endonuclease Eco32I. (C) The intergenic region *psbK-trnS* was amplified and digested with endonuclease BsaI. (D) The intergenic region *ycf4-psaI* was amplified and plastid type was assessed based on the size difference of amplicons. For each species (figures A-C), PCR amplicons are on the left lane and products after restriction digestion are on the right. The restriction digestion was carried out at 37°C for 8 hrs. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe^{TM} (INtRON Biotechnology).

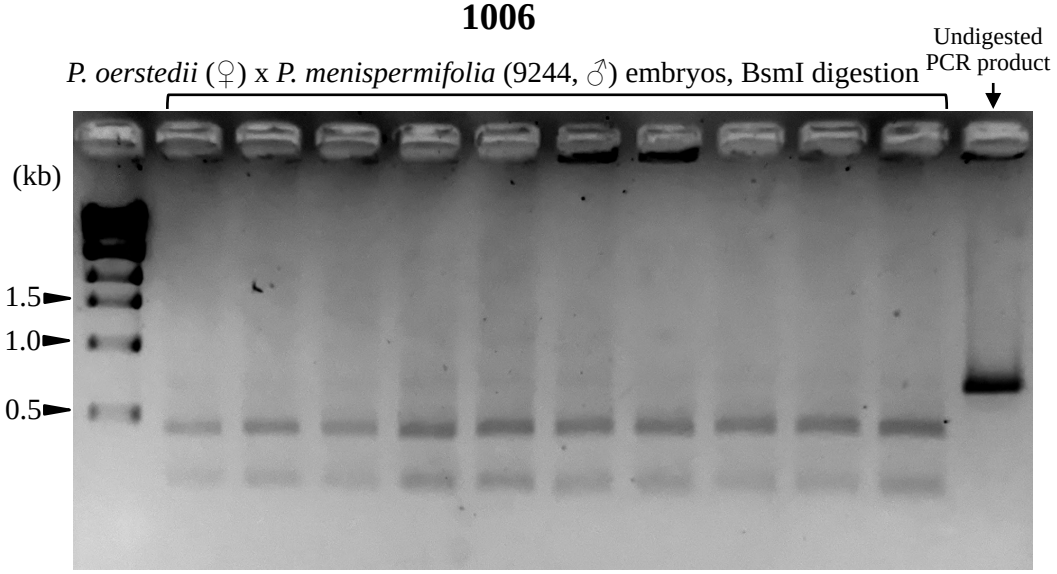
Figure 3.3. Agarose gels of PCR amplified ptDNA regions followed by restriction digestion in *Passiflora* hybrid embryos. Detail information regarding to ptDNA target region for amplification and restriction enzymes used for digestion is provided in Table 3.4. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe™ (INtRON Biotechnology).



1004


Undigested
PCR product
P. menispermifolia (9244,♀) x P. oerstedii (♂) embryos, BsmI digestion
(kb)
1.5
1.0
0.5

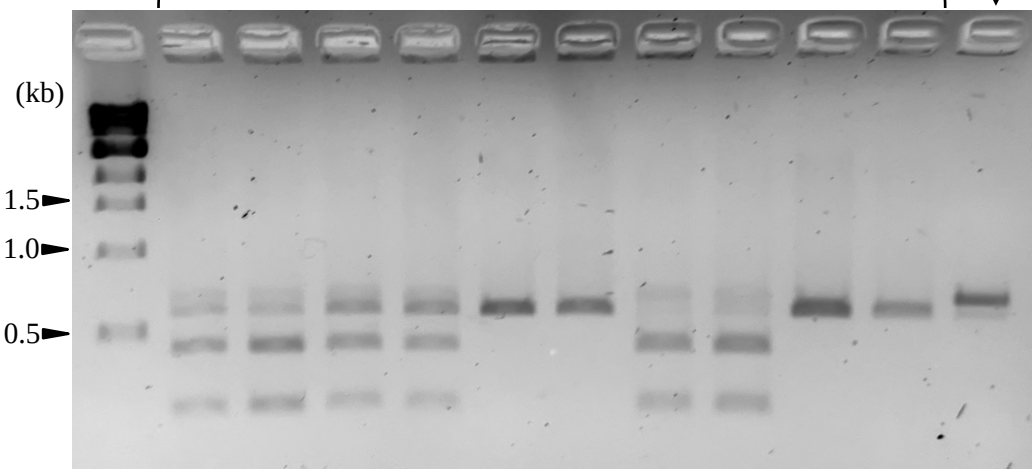
1006


Undigested
PCR product
P. oerstedii (♀) x P. menispermifolia (9244, ♂) embryos, BsmI digestion
(kb)
1.5
1.0
0.5

1028

P. menispermifolia (8039, ♀) x *P. hastifolia* (♂) embryos, BsmI digestion

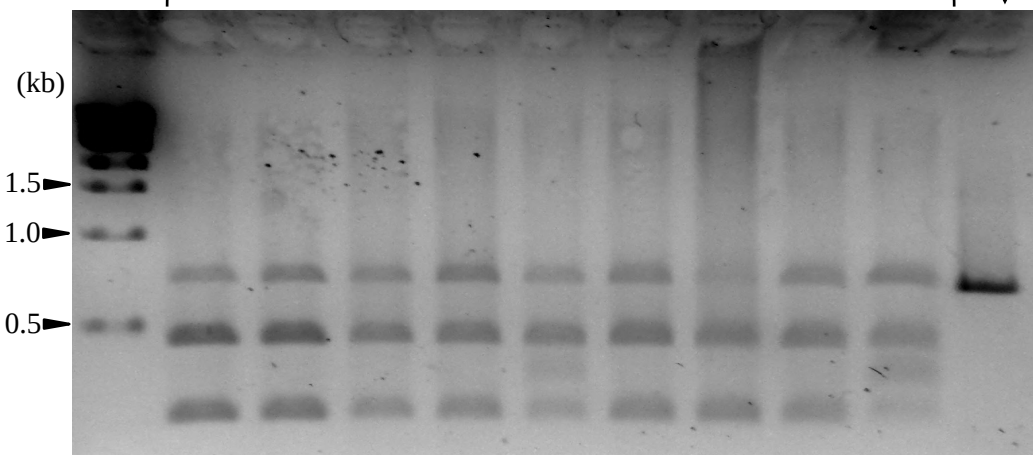
Undigested PCR product



1030

P. hastifolia (♀) x *P. menispermifolia* (8039, ♂) embryos, BsmI digestion

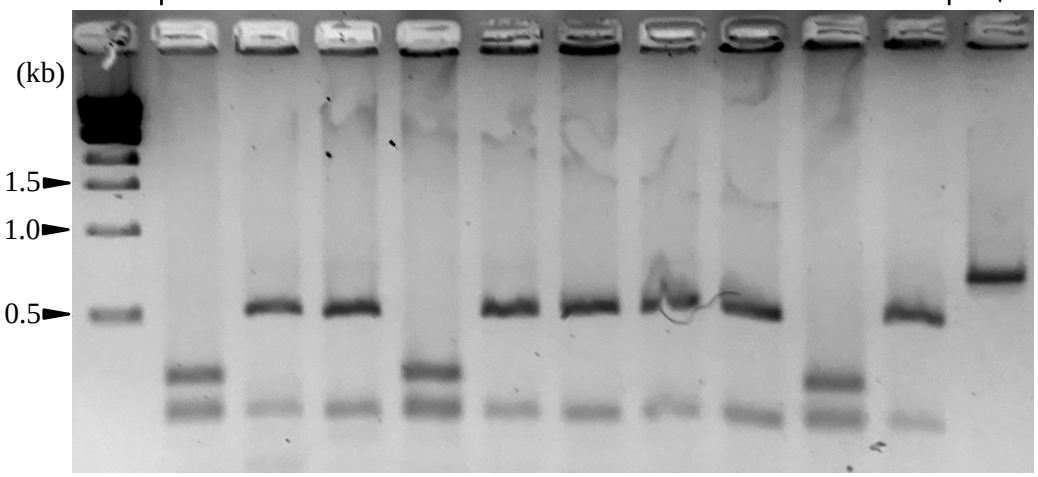
Undigested PCR product



1080

P. nitida (♀) x *P. quadrangularis* (♂) embryos, Bsp119I digestion

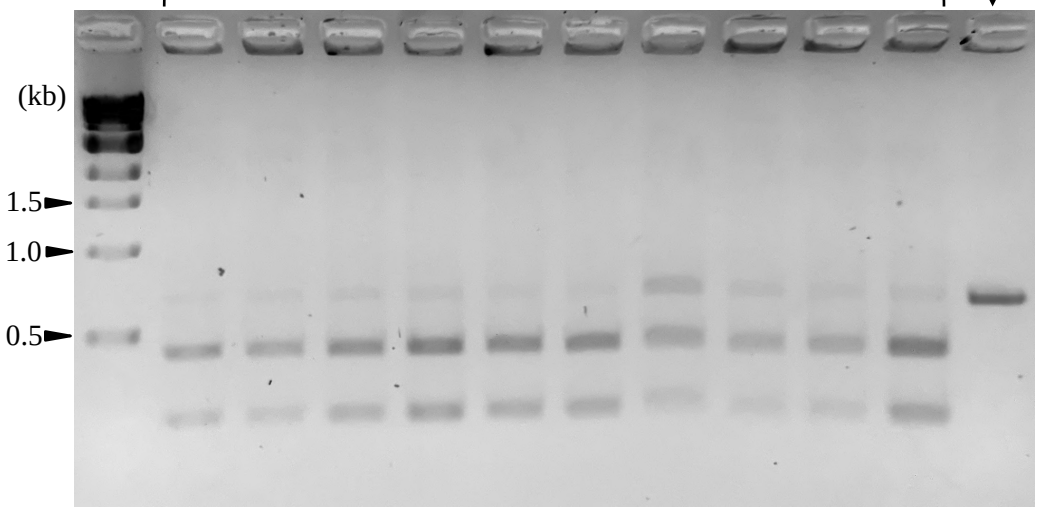
Undigested PCR product



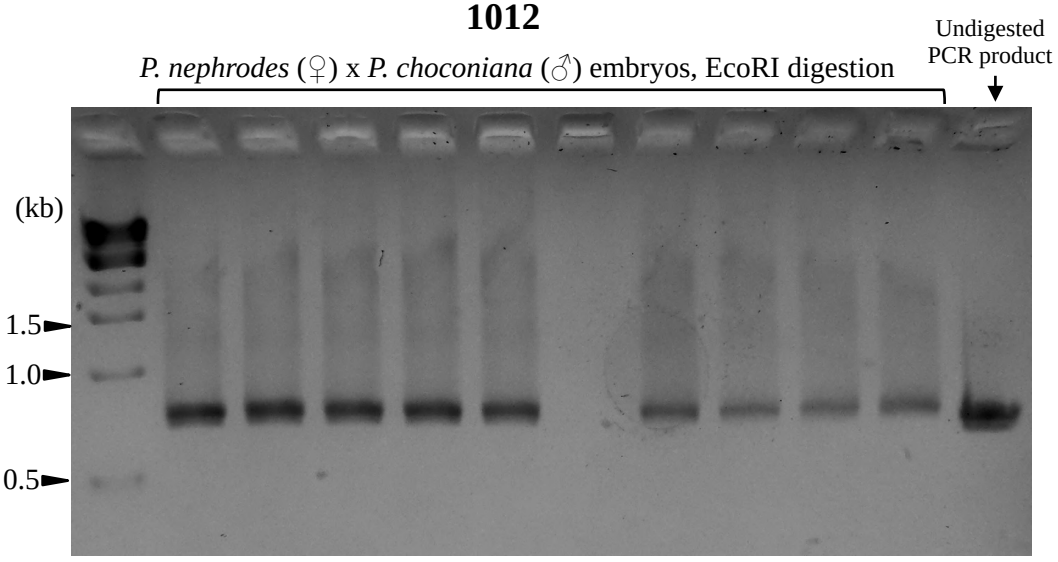
1031

P. oerstedii (♀) x *P. menispermifolia* (8039, ♂) embryos, BsmI digestion

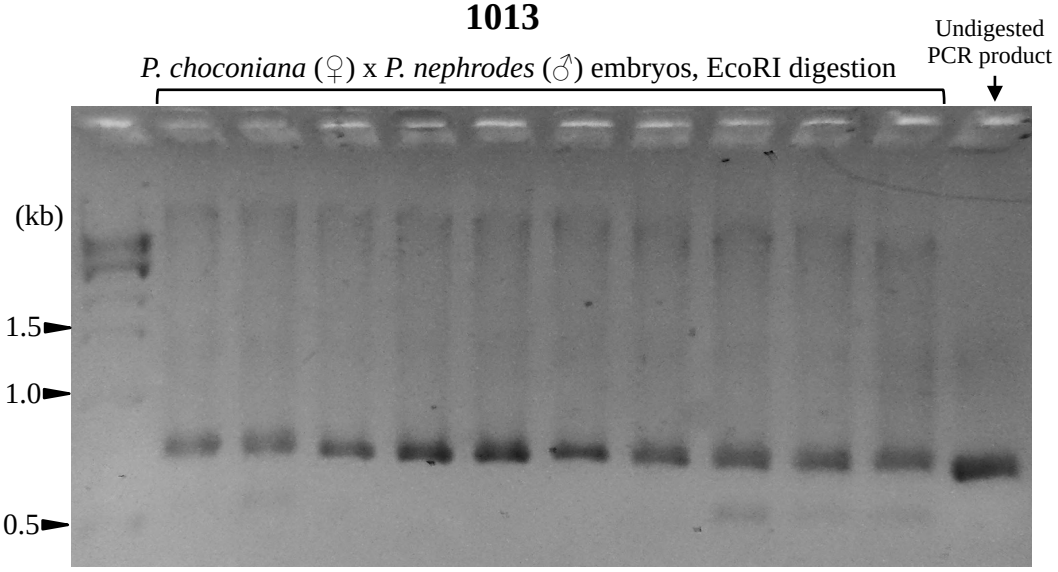
Undigested PCR product



1012


Undigested
PCR product
P. nephrodes (♀) x P. choconiana (♂) embryos, EcoRI digestion
(kb)
1.5
1.0
0.5

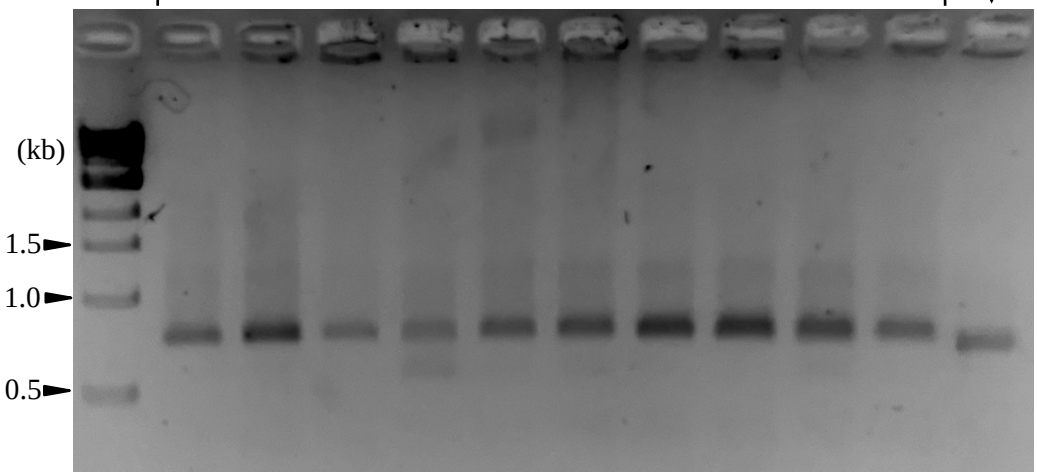
1013


Undigested
PCR product
P. choconiana (♀) x P. nephrodes (♂) embryos, EcoRI digestion
(kb)
1.5
1.0
0.5

1008

P. choconiana (♀) x *P. menispermifolia* (9224, ♂) embryos, EcoRI digestion

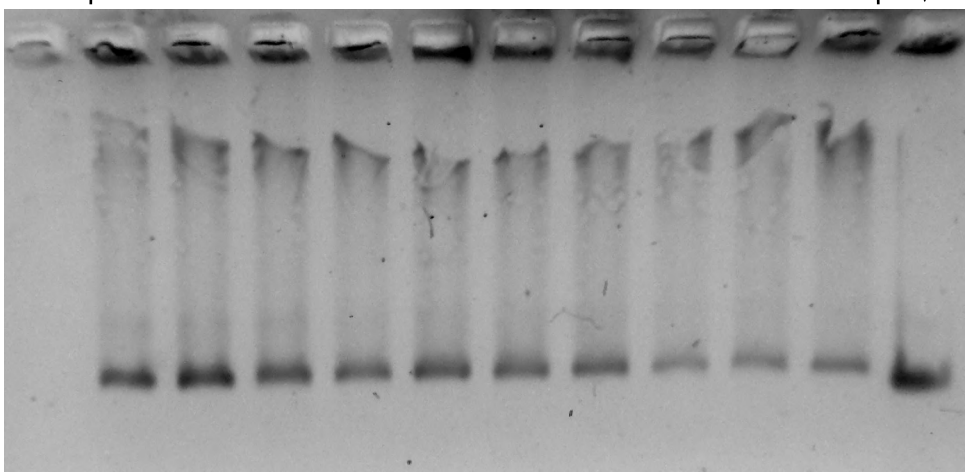
Undigested PCR product



1033

P. choconiana (♀) x *P. menispermifolia* (Sirena, ♂) embryos, EcoRI digestion

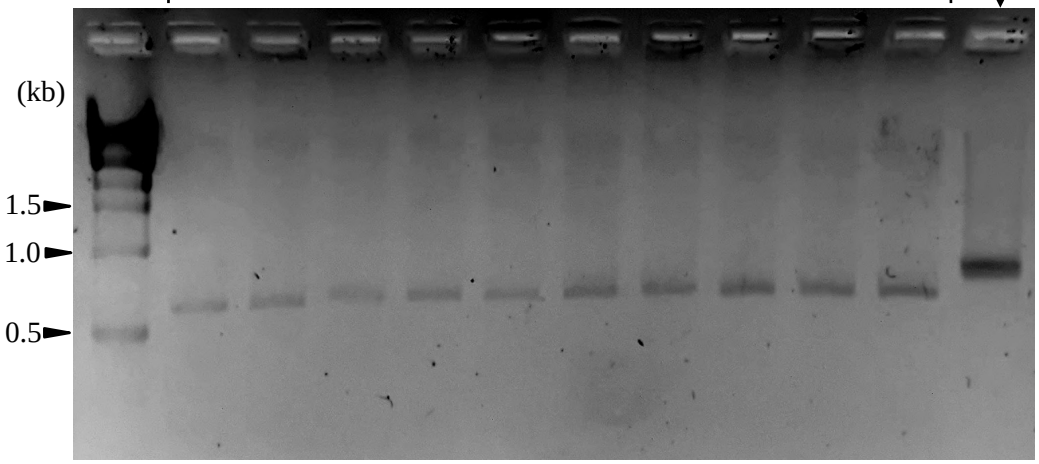
Undigested PCR product



1046

P. sprucei (♀) x *P. retipetala* (♂) embryos, EcoRI digestion

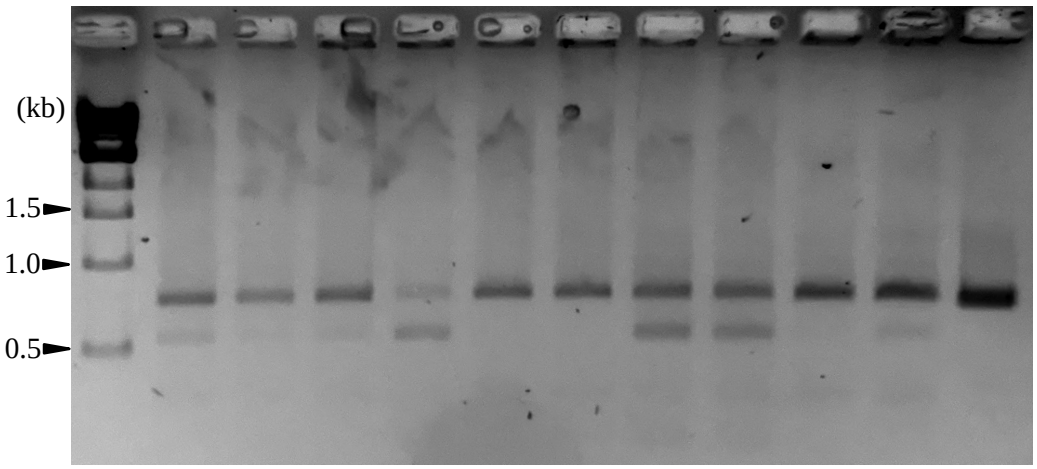
Undigested PCR product



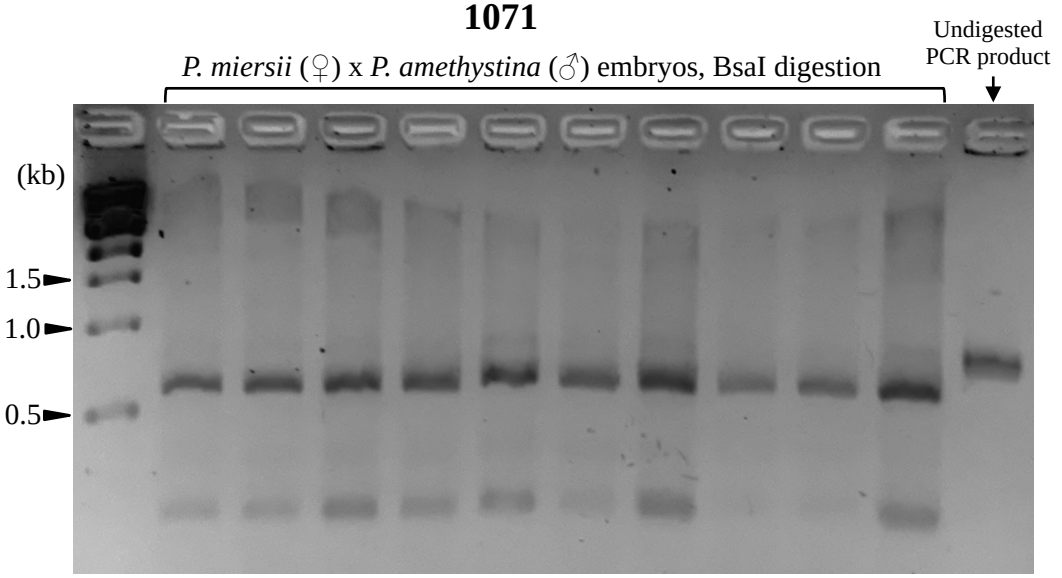
1068

P. menispermifolia (8039, ♀) x *P. resticulata* (♂) embryos, EcoRI digestion

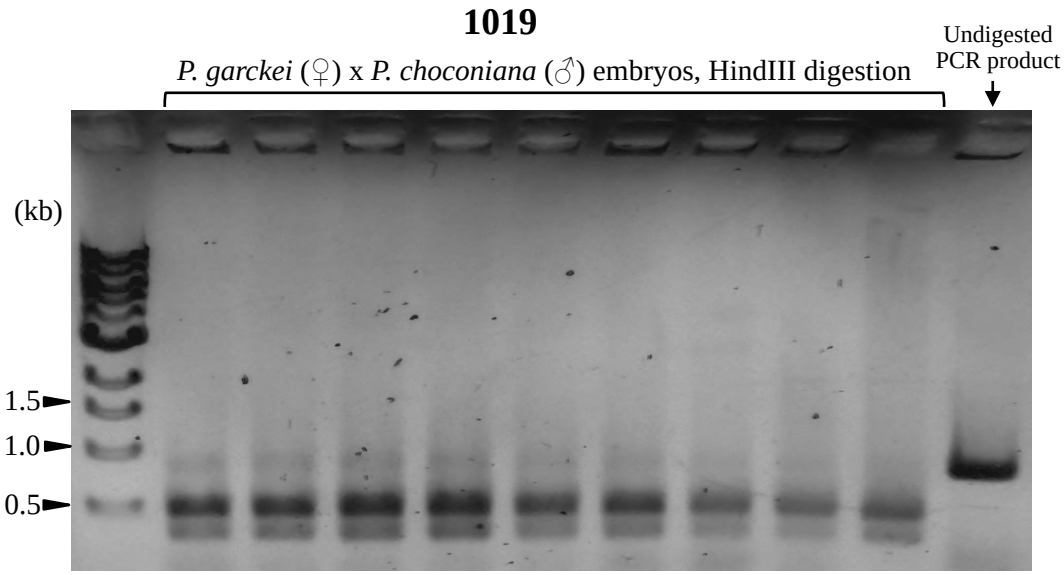
Undigested PCR product



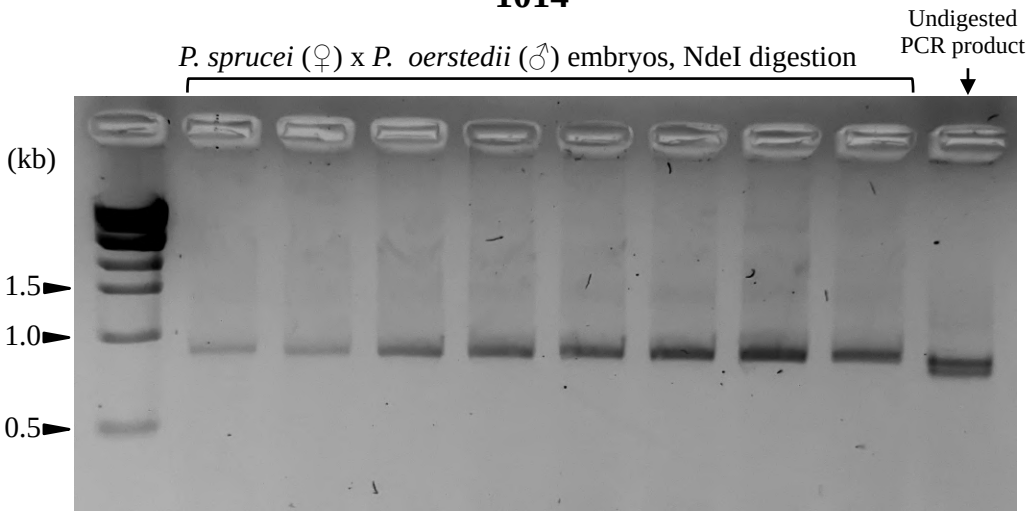
1071


Undigested
PCR product
P. miersii (♀) x P. amethystina (♂) embryos, BsaI digestion
(kb)
1.5
1.0
0.5

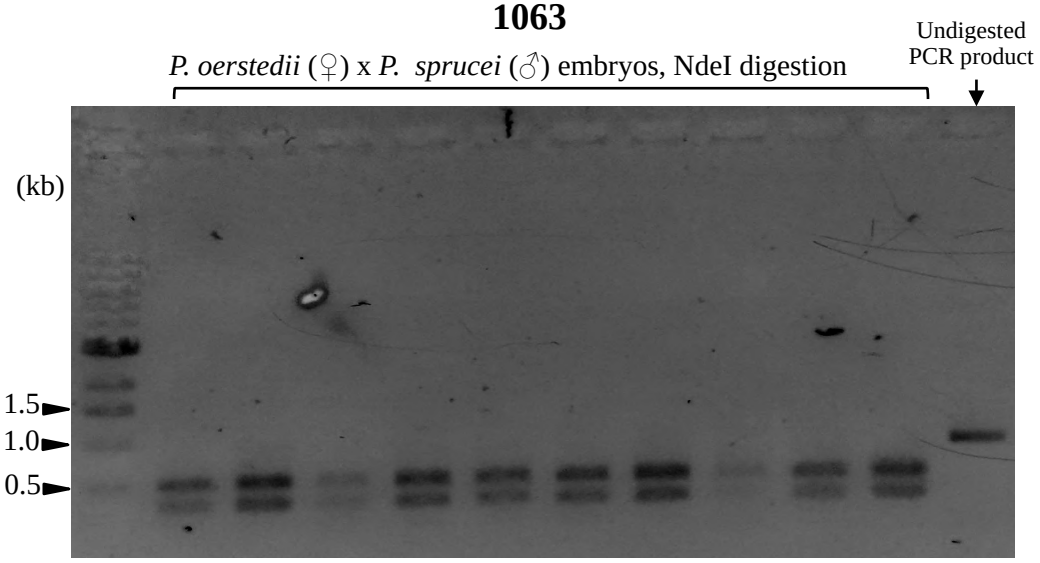
1019


Undigested
PCR product
P. garckei (♀) x P. choconiana (♂) embryos, HindIII digestion
(kb)
1.5
1.0
0.5

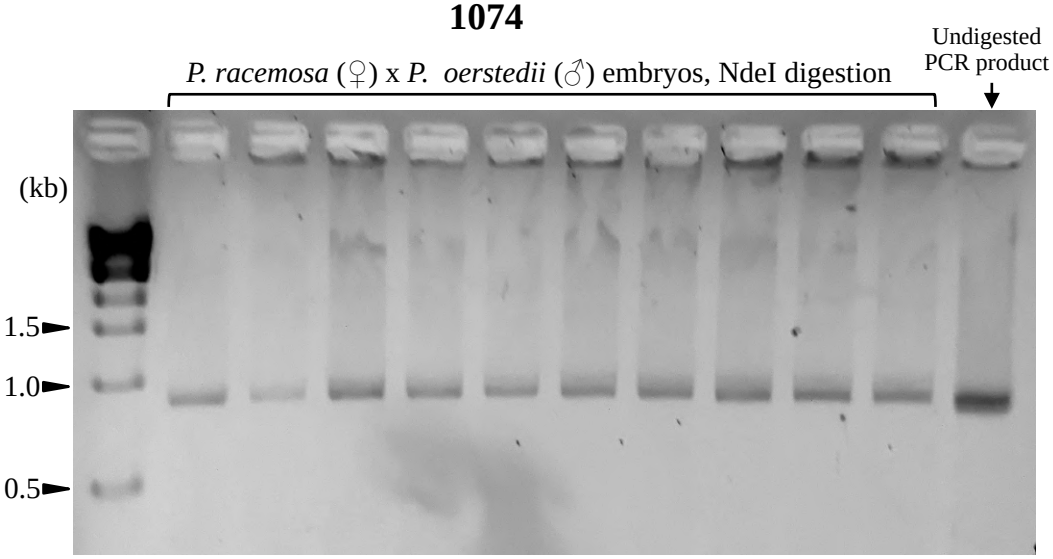
1014


Undigested
PCR product
P. sprucei (♀) x P. oerstedii (♂) embryos, NdeI digestion
(kb)
1.5
1.0
0.5

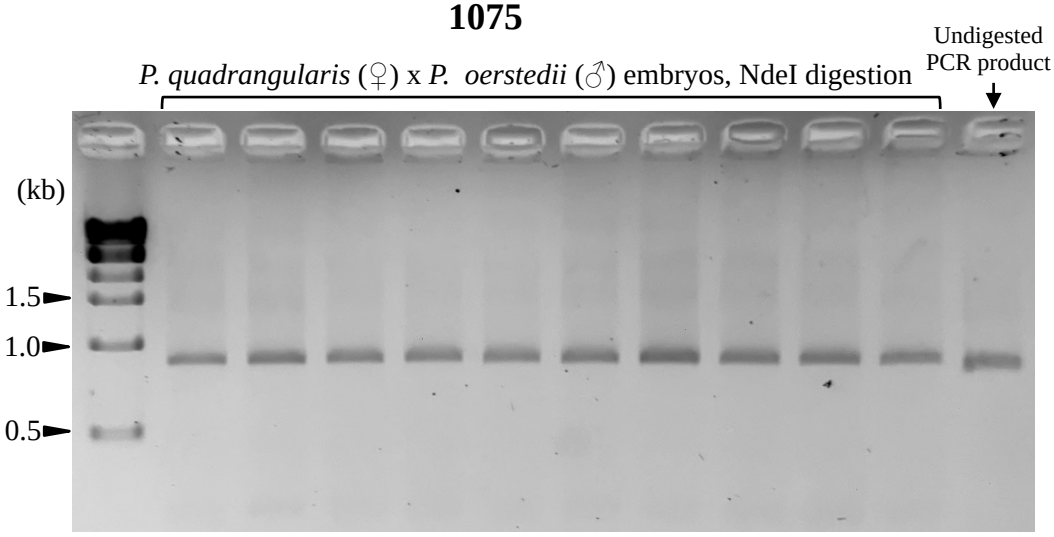
1063


Undigested
PCR product
P. oerstedii (♀) x P. sprucei (♂) embryos, NdeI digestion
(kb)
1.5
1.0
0.5

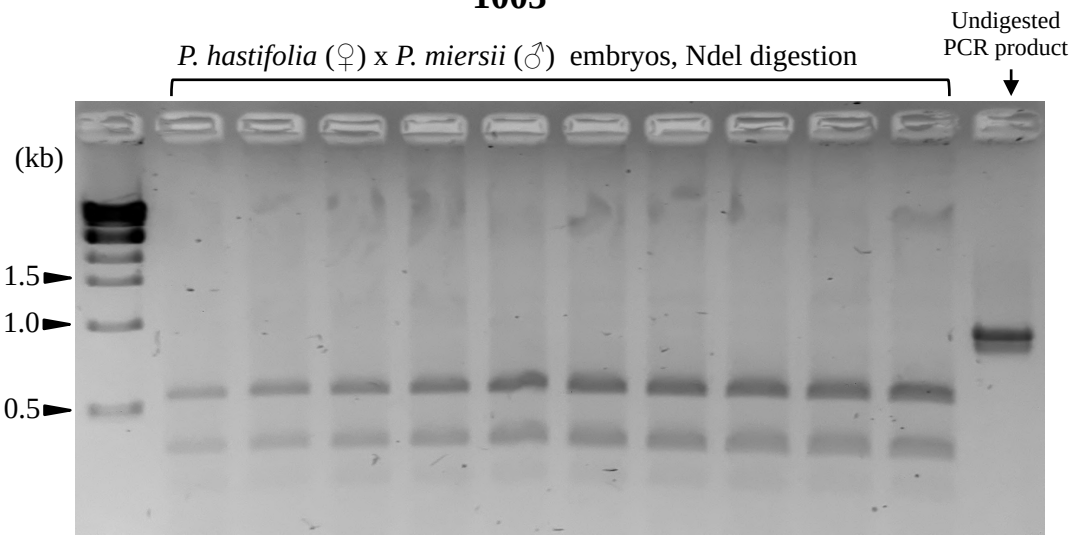
1074


Undigested
PCR product
P. racemosa (♀) x *P. oerstedii* (♂) embryos, NdeI digestion
(kb)
1.5
1.0
0.5

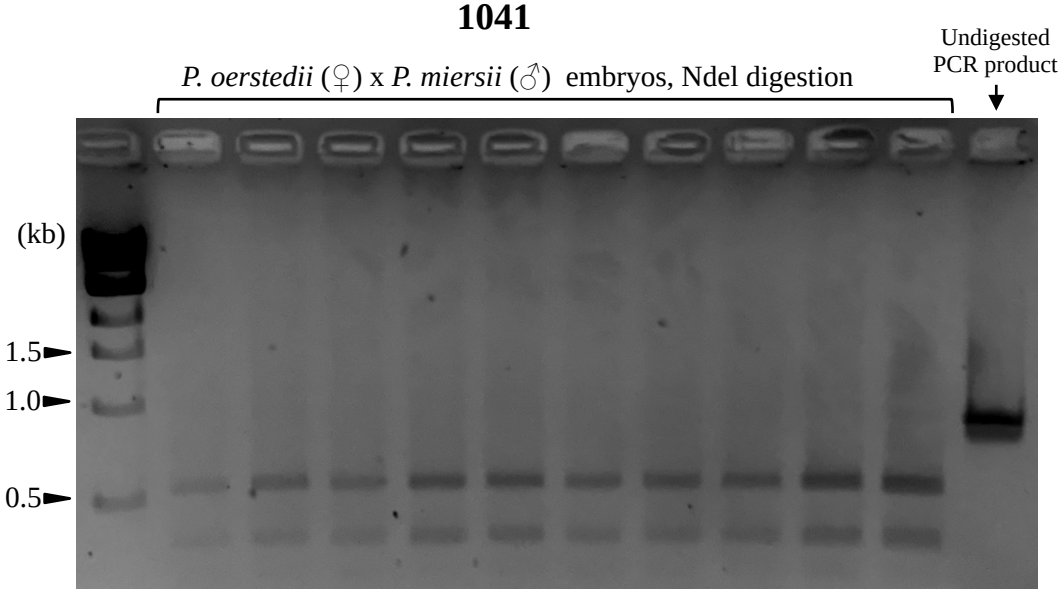
1075


Undigested
PCR product
P. quadrangularis (♀) x *P. oerstedii* (♂) embryos, NdeI digestion
(kb)
1.5
1.0
0.5

1003


Undigested
PCR product
P. hastifolia (♀) x P. miersii (♂) embryos, Ndel digestion
(kb)
1.5
1.0
0.5

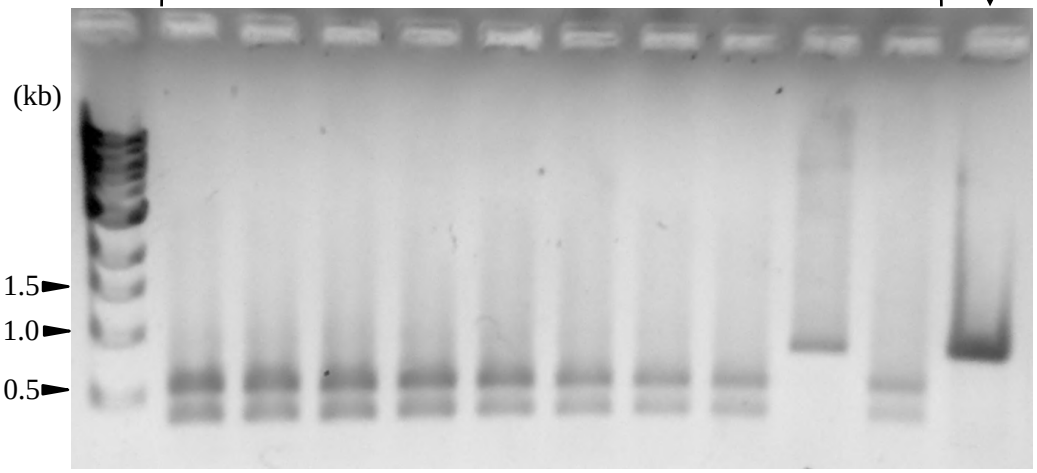
1041


Undigested
PCR product
P. oerstedii (♀) x P. miersii (♂) embryos, Ndel digestion
(kb)
1.5
1.0
0.5

1061

P. nephrodes (♀) x *P. sprucei* (♂) embryos, NdeI digestion

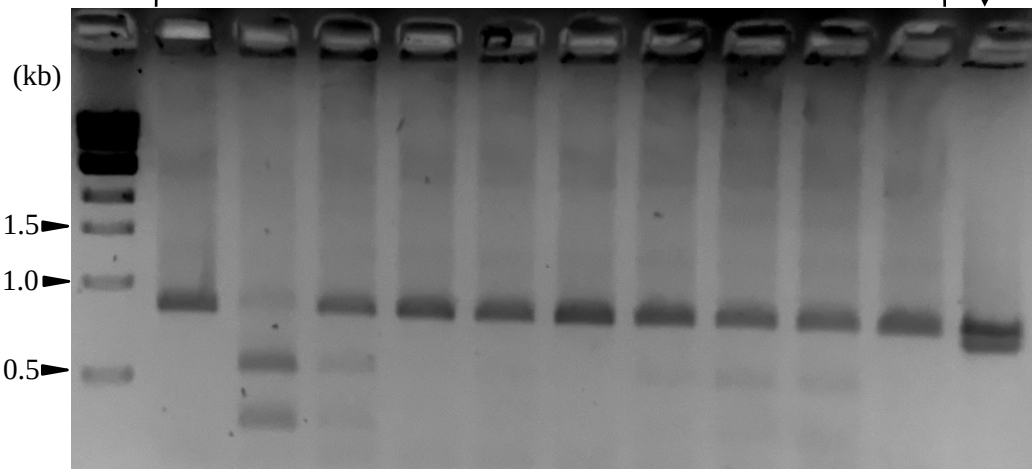
Undigested PCR product



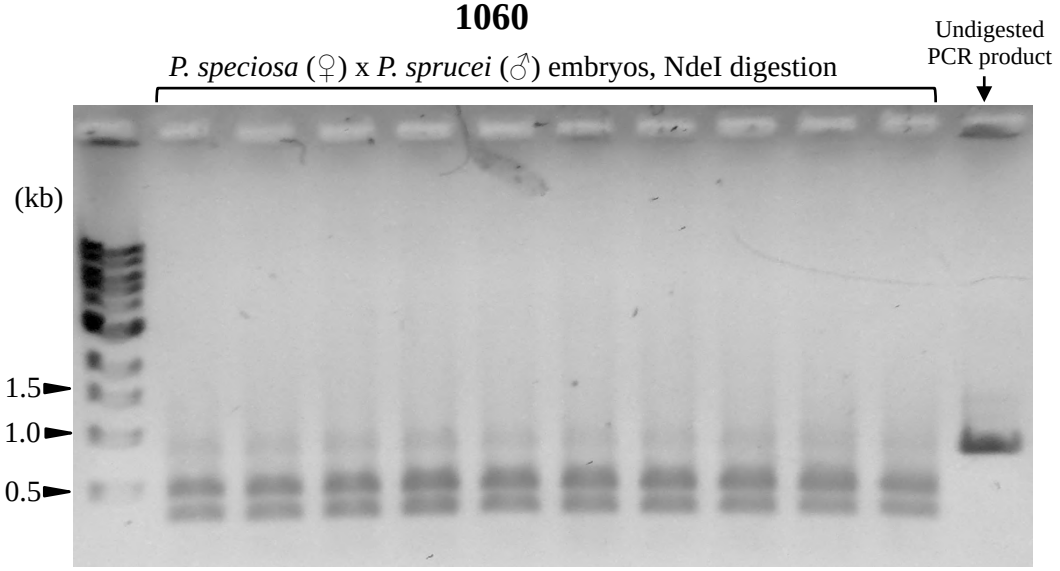
1062

P. choconiana (♀) x *P. sprucei* (♂) embryos, NdeI digestion

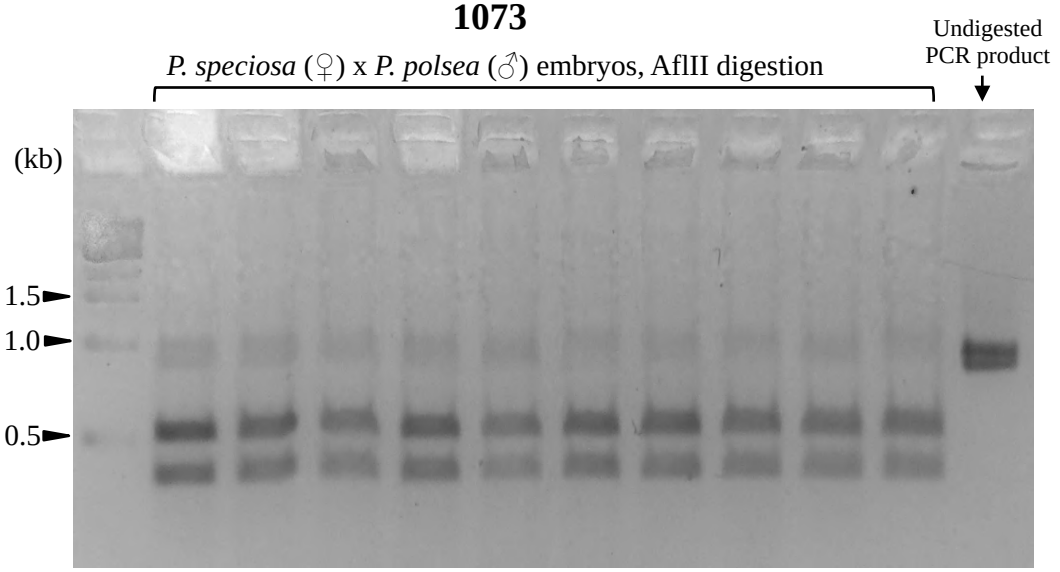
Undigested PCR product



1060


Undigested
PCR product
P. speciosa (♀) x P. sprucei (♂) embryos, NdeI digestion
(kb)
1.5
1.0
0.5

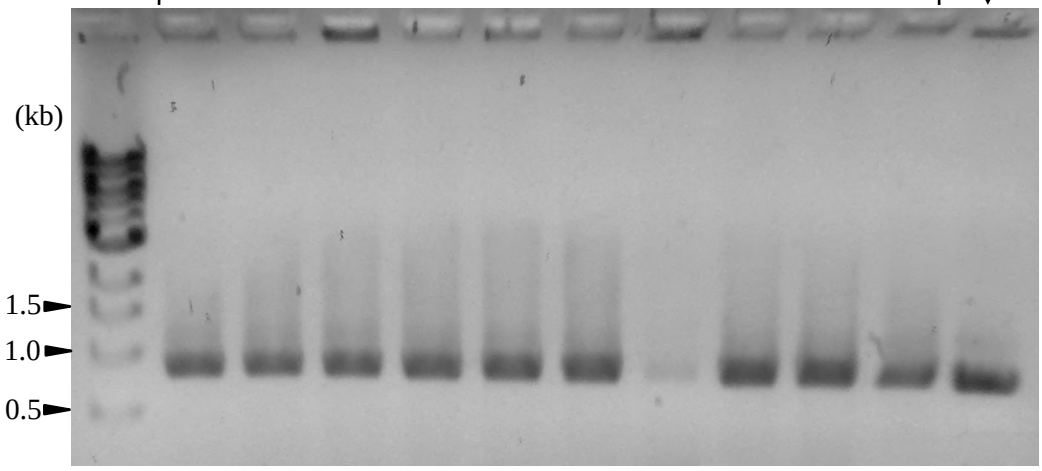
1073


Undigested
PCR product
P. speciosa (♀) x P. polsea (♂) embryos, AflII digestion
(kb)
1.5
1.0
0.5

1032

P. vitifolia (♀) x *P. hastifolia* (♂) embryos, NdeI digestion

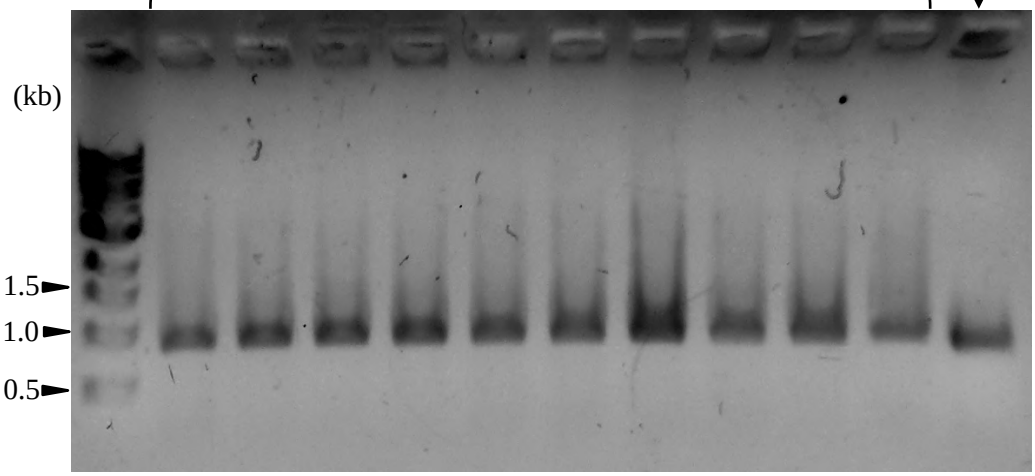
Undigested PCR product



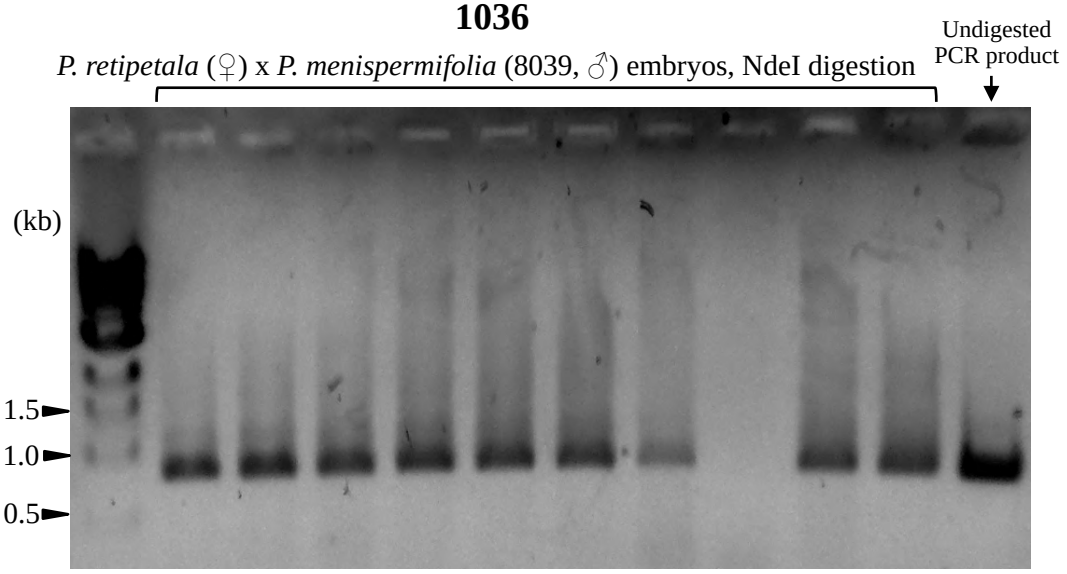
1053

P. vitifolia (♀) x *P. speciosa* (♂) embryos, SwaI digestion

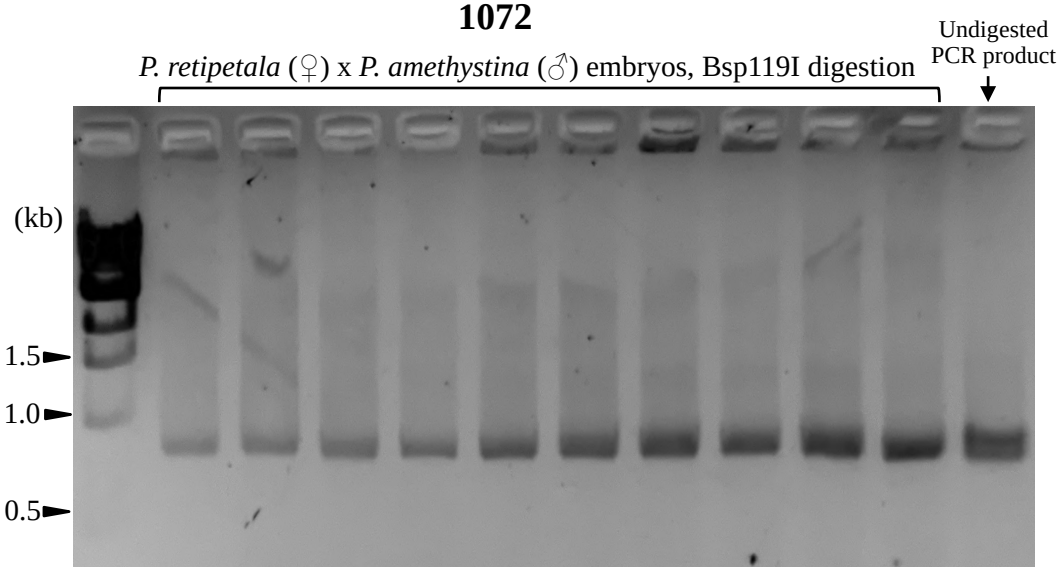
Undigested PCR product



1036


Undigested
PCR product
P. retipetala (♀) x *P. menispermifolia* (8039, ♂) embryos, NdeI digestion
(kb)
1.5
1.0
0.5

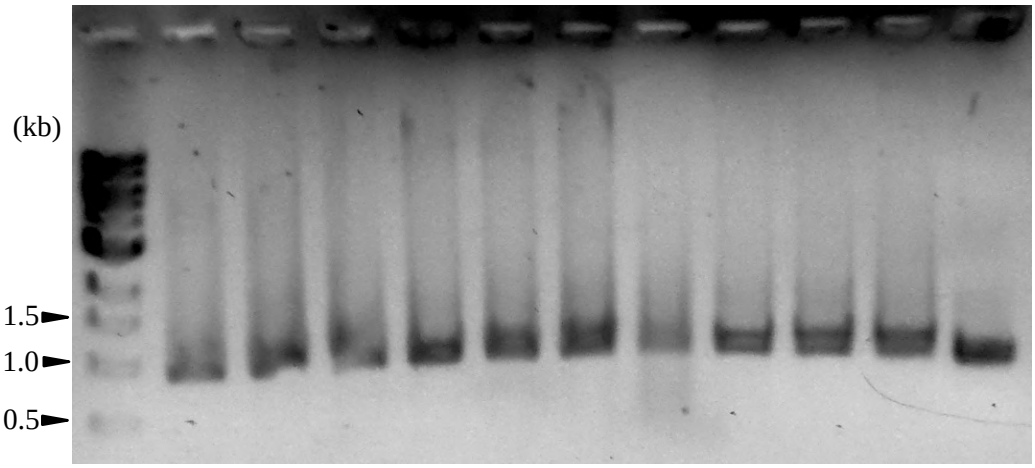
1072


Undigested
PCR product
P. retipetala (♀) x *P. amethystina* (♂) embryos, Bsp119I digestion
(kb)
1.5
1.0
0.5

1077

P. vitifolia (♀) x *P. edulis* (♂) embryos, SwaI digestion

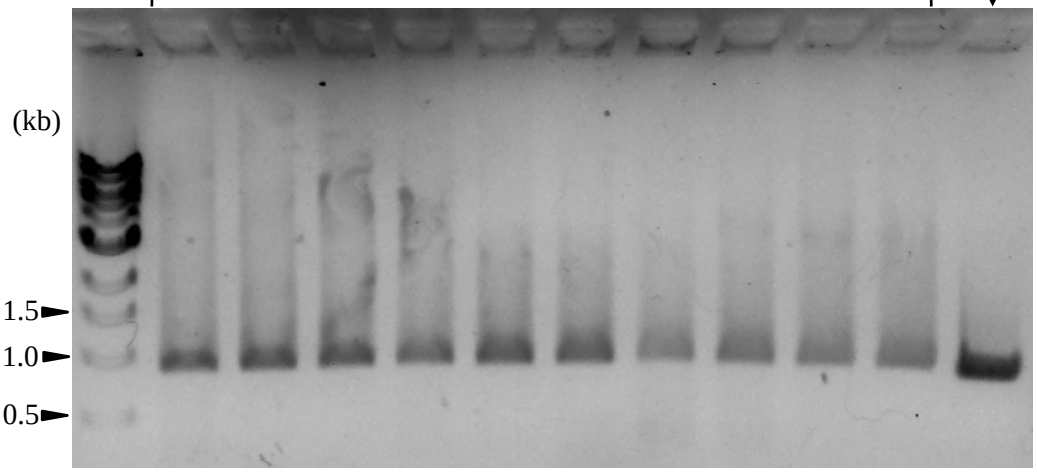
Undigested PCR product



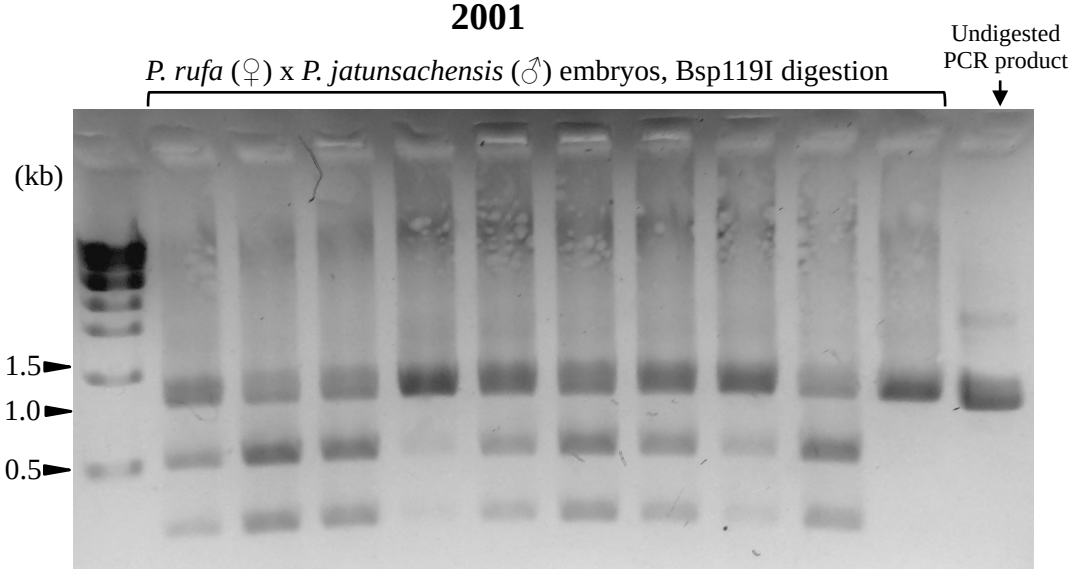
1078

P. vitifolia (♀) x *P. quadrangularis* (♂) embryos, SwaI digestion

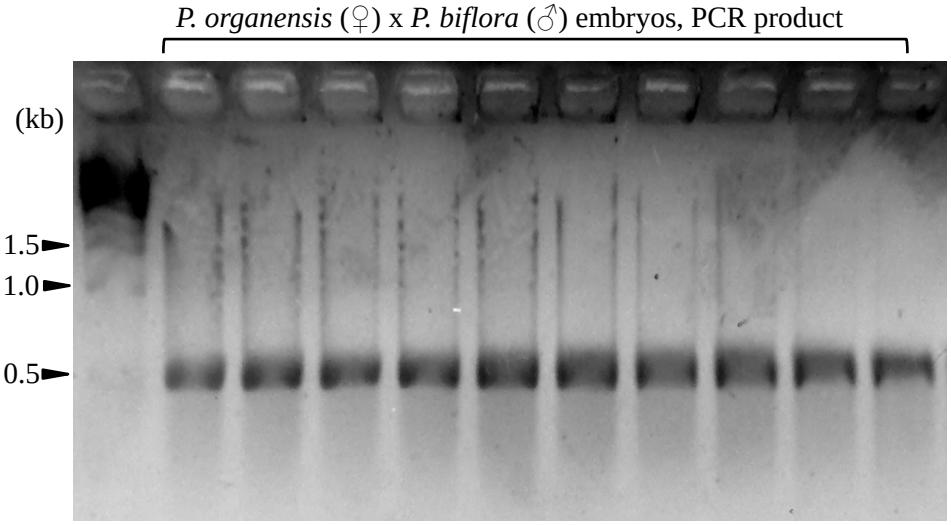
Undigested PCR product



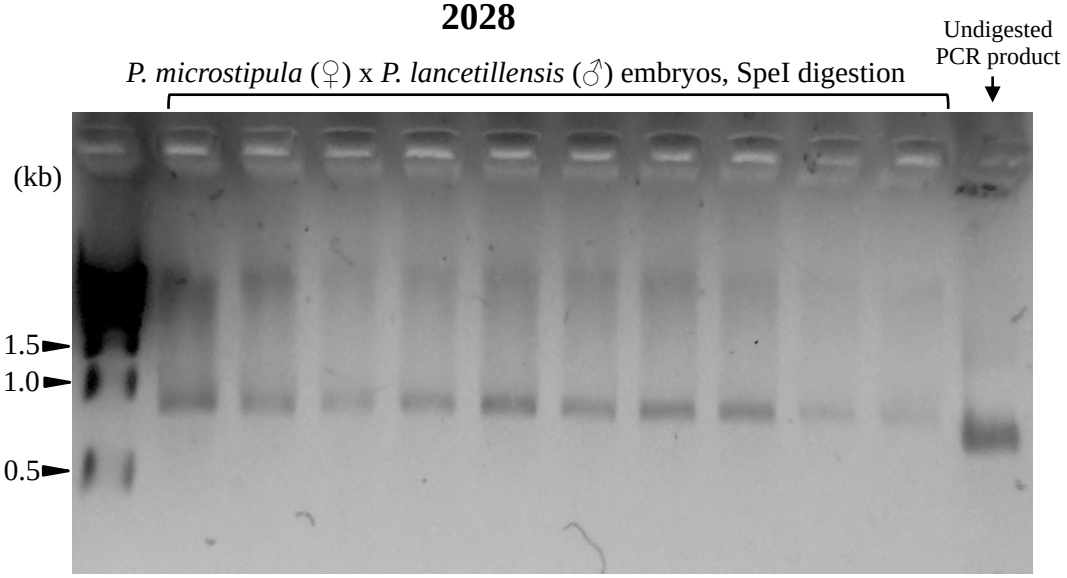
2001


Undigested
PCR product
P. rufa (♀) x P. jatunsachensis (♂) embryos, Bsp119I digestion
(kb)
1.5
1.0
0.5

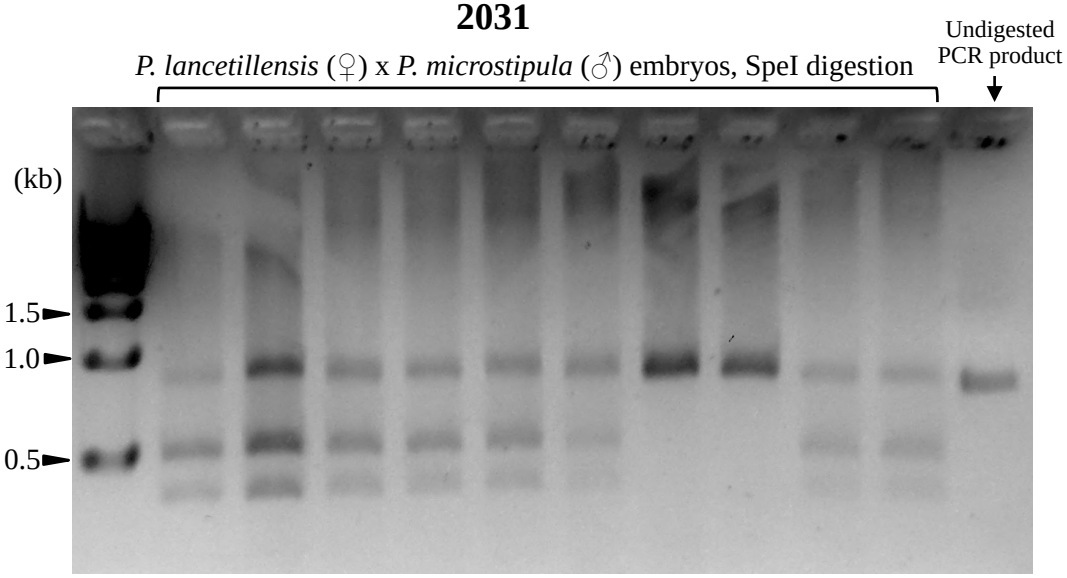
2014


P. organensis (♀) x P. biflora (♂) embryos, PCR product
(kb)
1.5
1.0
0.5

2028


Undigested
PCR product
P. microstipula (♀) x P. lancetillensis (♂) embryos, SpeI digestion
(kb)
1.5
1.0
0.5

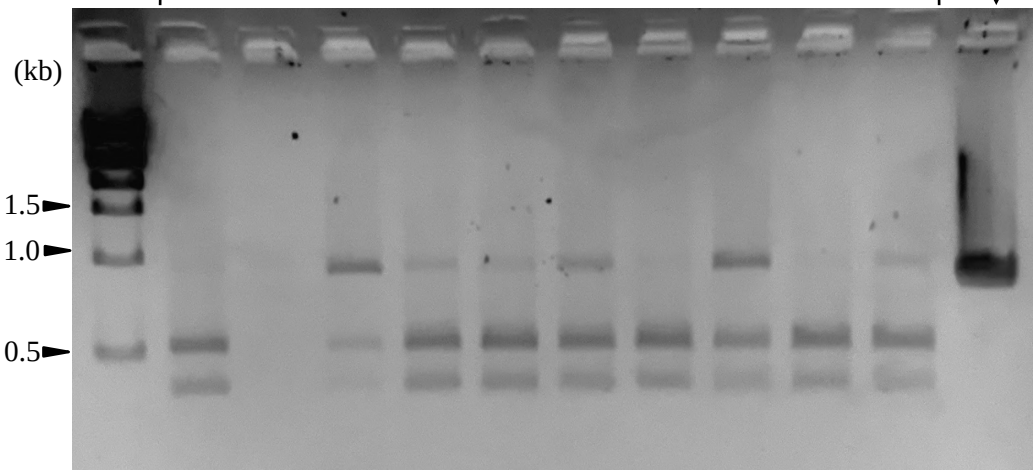
2031


Undigested
PCR product
P. lancetillensis (♀) x P. microstipula (♂) embryos, SpeI digestion
(kb)
1.5
1.0
0.5

2027

P. misera (9023,♀) x *P. misera* (9335,♂) embryos, EarI digestion

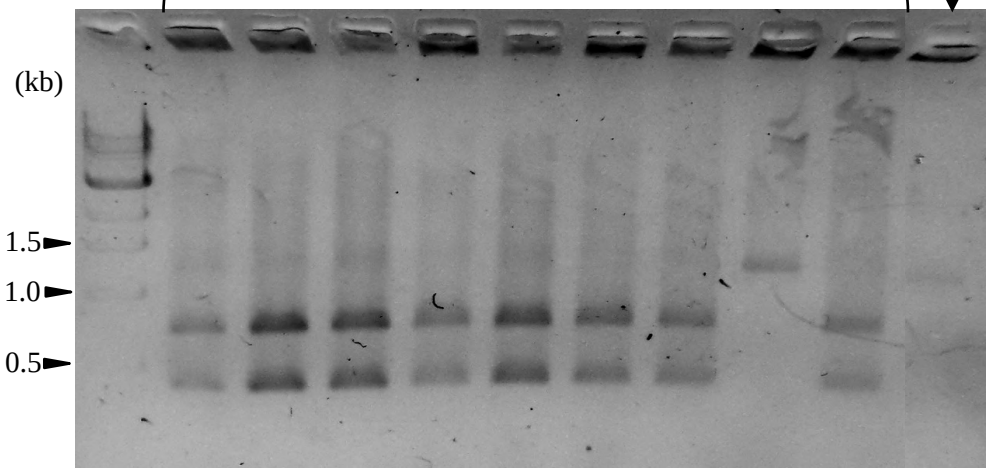
Undigested PCR product



3002

P. sphaerocarpa (♀) x *P. pittieri* (♂) embryos, Eco32I digestion

Undigested PCR product



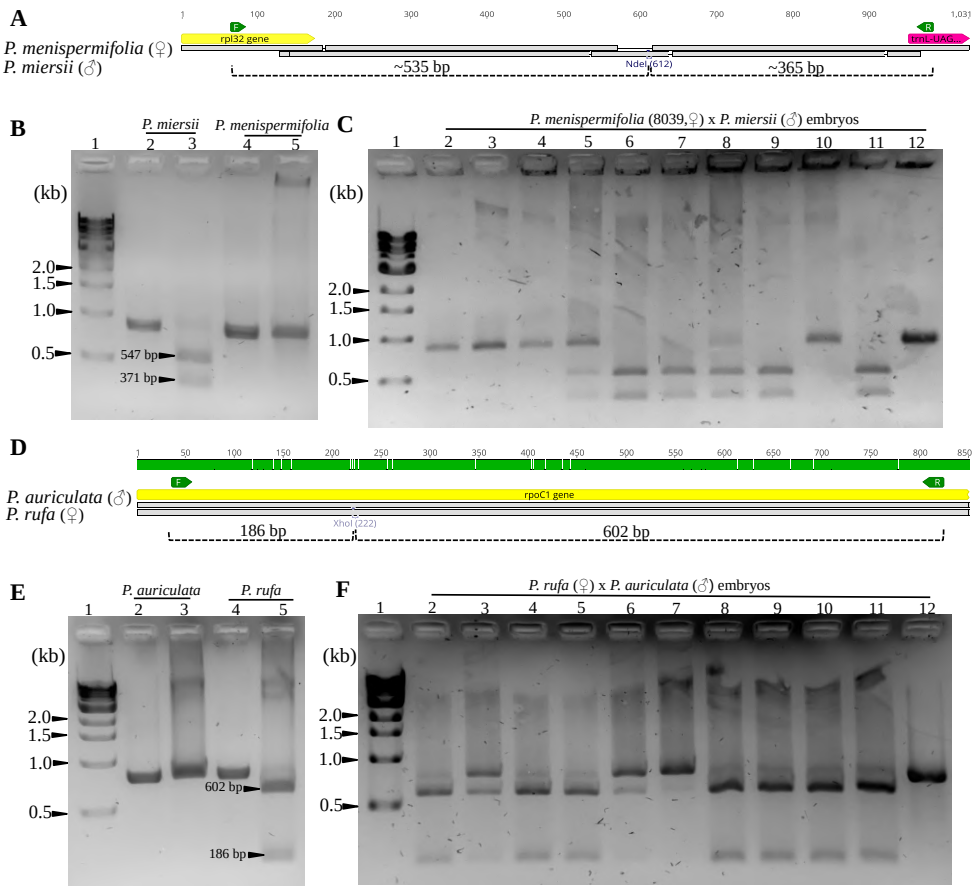


Figure 3.4. Agarose gels of PCR amplified target regions following digestion with restriction endonucleases in *Passiflora* hybrid embryos. (A) The intergenic region (*rpl32-trnL*) amplified from *P. menispermifolia* (♀) and *P. miersii* (♂) and their hybrid (1044) with estimated fragment sizes after NdeI digestion. (B) Lanes 2 and 4 are amplicons of *rpl32-trnL* for *P. miersii* and *P. menispermifolia* and lanes 3 and 5 are products after digestion with NdeI. (C). Lanes 2-11 are *Nde*I digestion products of amplified *rpl32-trnL* for 10 embryos of hybrid 1044. Lane 12 is undigested PCR product for embryo 1 of hybrid 1044 as a reference. (D) The partial coding region within *rpoC1* amplified from *P. rufa* (♀) and *P. auriculata* (♂) and their hybrid progeny (2005) with estimated fragment sizes after XhoI digestion. (E) Lanes 2 and 4 are amplicons of *rpoC1* for *P. auriculata* and *P. rufa* and lanes 3 and 5 are products after digestion with XhoI. (F) Lanes 2-11 are XhoI digestion products of amplified *rpoC1* for 10 embryos of 2005. Lane 12 is undigested PCR product for embryo 1 of hybrid 2005 as a reference. Lane 1 in all gel images (B-C, E-F) contains one kb DNA ladder (NEB). DNA bands were separated in 1.5% agarose gel stained with RedSafe™ (INtRON Biotechnology).

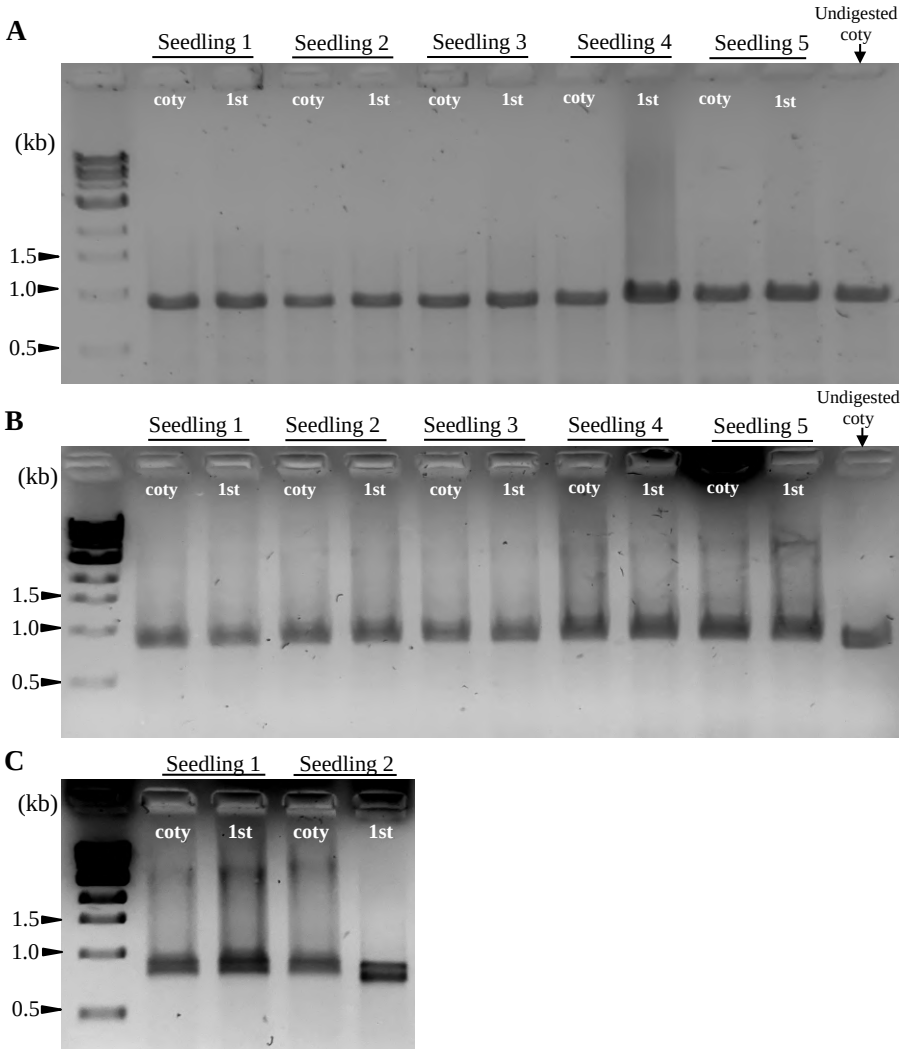


Figure 3.5.1. Agarose gels of PCR amplified *rpl32-trnL* region following digestion with SwaI endonuclease in *Passiflora* hybrid seedlings. (A) 1053 seedlings, *P. vitifolia* (♀) x *P. speciosa* (♂). (B) 1077 seedlings, *P. vitifolia* (♀) x *P. edulis* (♂). (C) 1078 seedlings, *P. vitifolia* (♀) x *P. quadrangularis* (♂). For each seedling, the digested PCR products are shown for cotyledon (coty) and the first leaf (1^{st}). Last lanes in figures A-B contain undigested PCR product of cotyledon (coty) as a reference. The restriction digestion was carried out at 37°C for 9 hrs. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe^{TM} (INtRON Biotechnology).

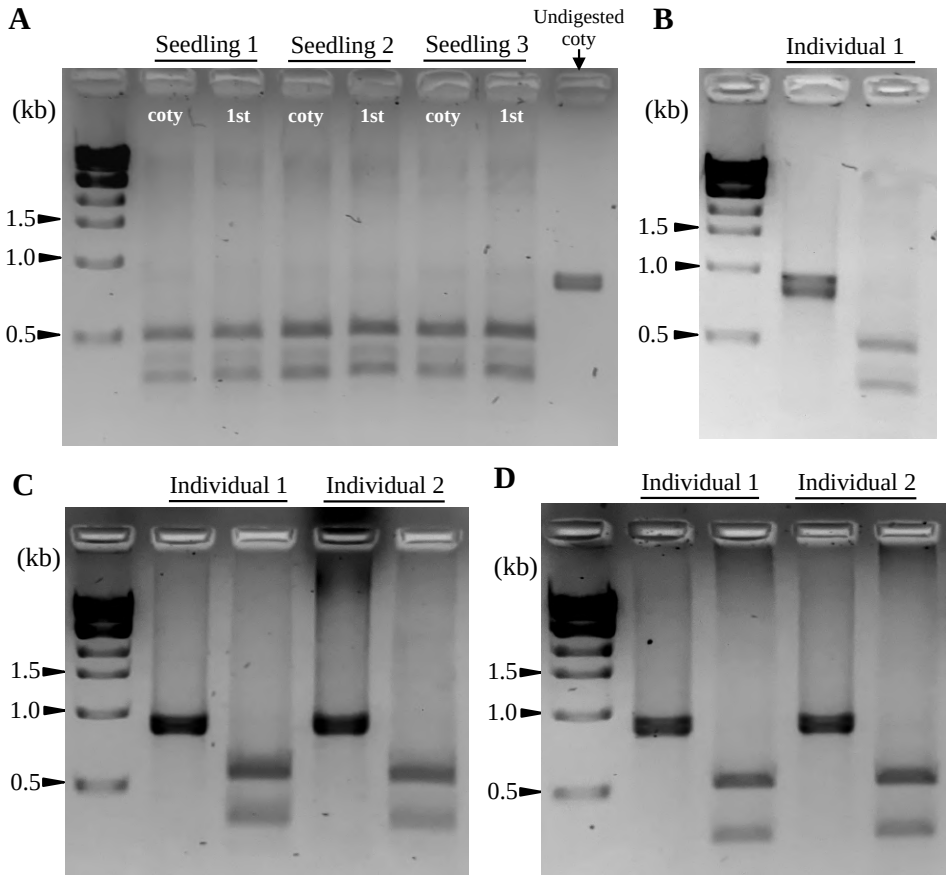


Figure 3.5.2. Agarose gels of PCR amplified *rpl32-trnL* region following digestion with NdeI or HindIII endonuclease in *Passiflora* hybrids (A) 1060 seedlings, *P. speciosa* (♀) x *P. sprucei* (♂). The digested PCR products are shown for cotyledon (coty) and the first leaf (1^{st}). The last lane contains undigested PCR product of cotyledon. (B) 1061 leaf, *P. nephrodes* (♀) x *P. sprucei* (♂). (C) 1063 leaves, *P. choconiana* (♀) x *P. sprucei* (♂). The restriction digestion was carried out at 37°C for 12 hrs for NdeI and 9 hrs for HindIII. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe™ (INtRON Biotechnology).

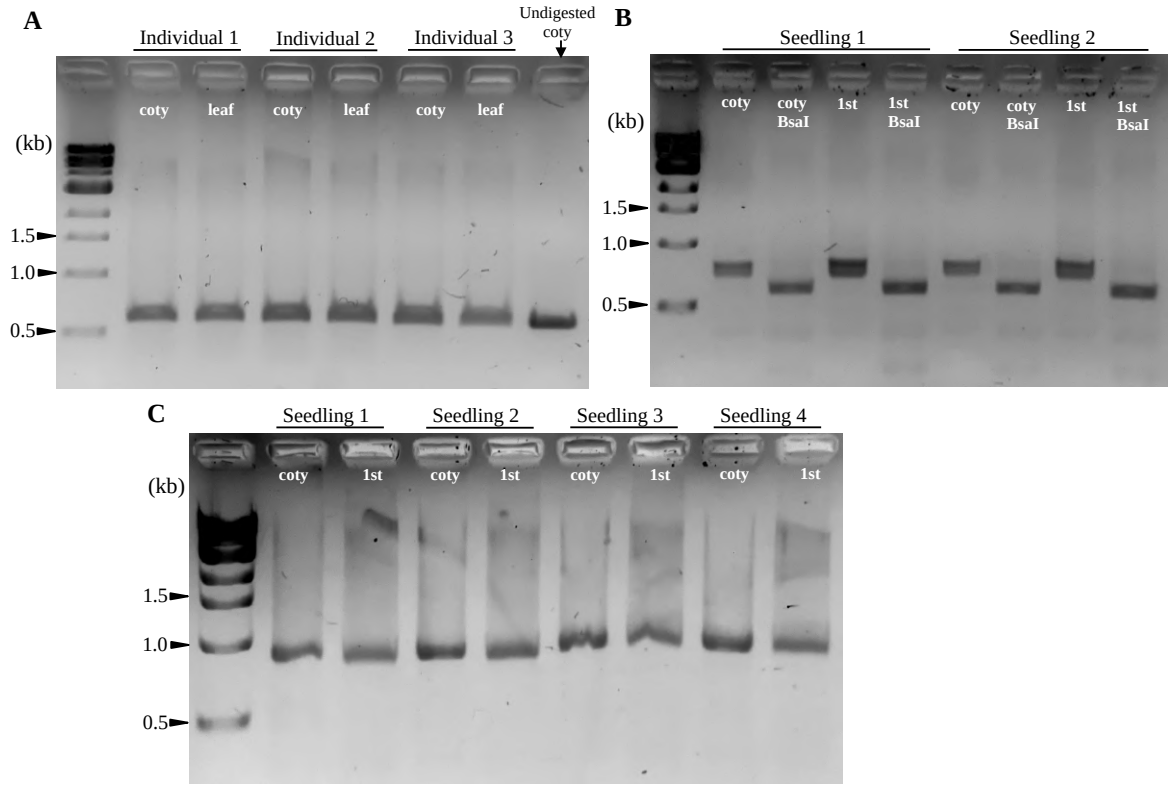


Figure 3.5.3. Agarose gels of PCR amplified various target regions following digestion with endonuclease in *Passiflora* hybrids. (A) 1028 cotyledons and leaves from 3 months old hybrids, *P. menispermifolia* (8039,♀) x *P. hastifolia* (♂). Target region *accD-rbcL* amplified and digested with BsmI at 65°C for 9 hrs. (B) 1071 seedlings, *P. miersii* (♀) x *P. amenthystina* (♂). Target region *psbK-trnS* amplified and digested with BsaI at 37°C for 12 hrs. (C) 1072 seedlings, *P. retipetala* (♀) x *P. amenthystina* (♂). Target region *rpl32-trnL* amplified and digested with NdeI at 37°C for 12 hrs. In figures A and C, the digested PCR products are shown for cotyledon (coty) and the first leaf (1^{st}), where as in figure B, products after BsaI digestion (coty BsaI; 1st BsaI) were loaded subsequently after the amplicons of *psbK-trnS* for cotyledons and 1^{st} leaf (coty, 1^{st}) for each seedlings. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe^{TM} (INtRON Biotechnology).

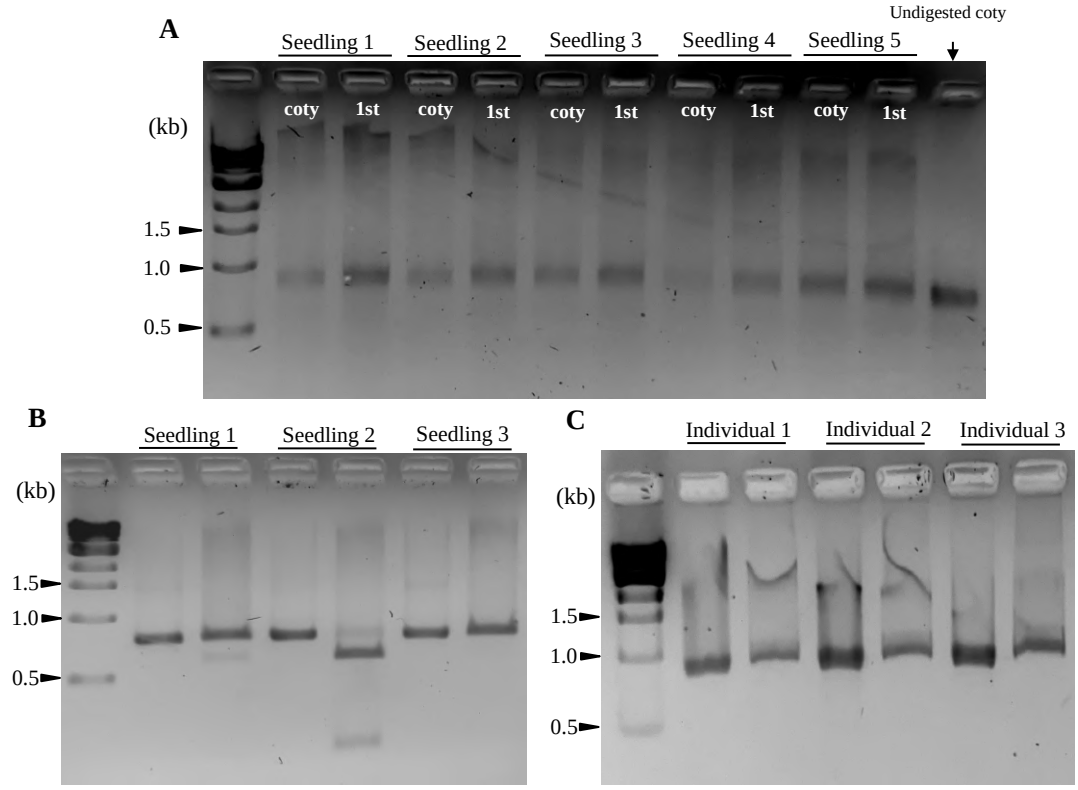


Figure 3.5.4. Agarose gels of PCR amplified various target regions following digestion with endonuclease in *Passiflora* hybrids. (A) 2028 seedlings, *P. microstipula* (♀) x *P. lancetillensis* (♂). The intergenic region *psbA-matK* amplified and digested with SpeI at 37°C for 12 hrs. The digested PCR products are shown for cotyledon (coty) and the first leaf (1st). The last lane contains undigested PCR product of cotyledon. (B) 2005 seedlings, *P. rufa* (♀) x *P. auriculata* (♂). The partial coding region within *rpoC1* was amplified and digested with XhoI at 37°C for 9 hrs. (C) 2018 leaves from a year old hybrids, *P. misera* (9335, ♀) x *P. misera* (9023, ♂). The partial coding region within *rpoB* amplified was amplified and digested with EarI at 37°C for 9 hours. In the figures B-C, amplicons of the target region are one the left lane and the digested products are on the right for each seedlings/individuals. Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe^{TM} (INtRON Biotechnology).

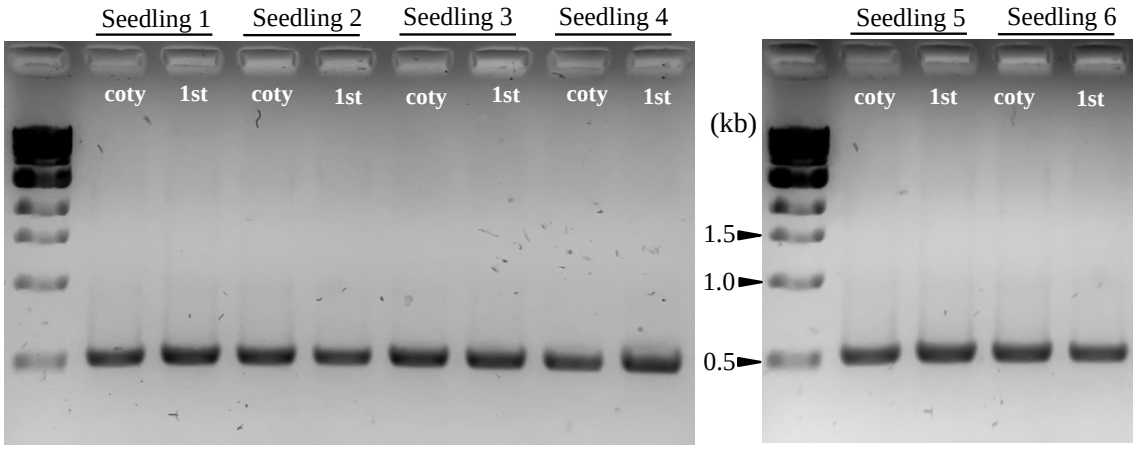


Figure 3.5.5. Agarose gels of PCR amplified *ycf4-psaI* region in seedlings of *Passiflora* hybrid *P. organensis* (♀) x *P. biflora* (♂). In the figures, the amplicons of *ycf4-psaI* are shown for cotyledon (coty) and the first leaf (1^{st}) for the six seedlings. Plastid type was assessed based on the size difference of the amplicons between P. *biflora* (800 bp) and *P. organensis* (~500 bp). Lane 1 in all gel images contains one kb DNA ladder (N323L, New England Biolabs, Inc). DNA bands were separated in 2% agarose gel and visualized with RedSafe^{TM} (INtRON Biotechnology).

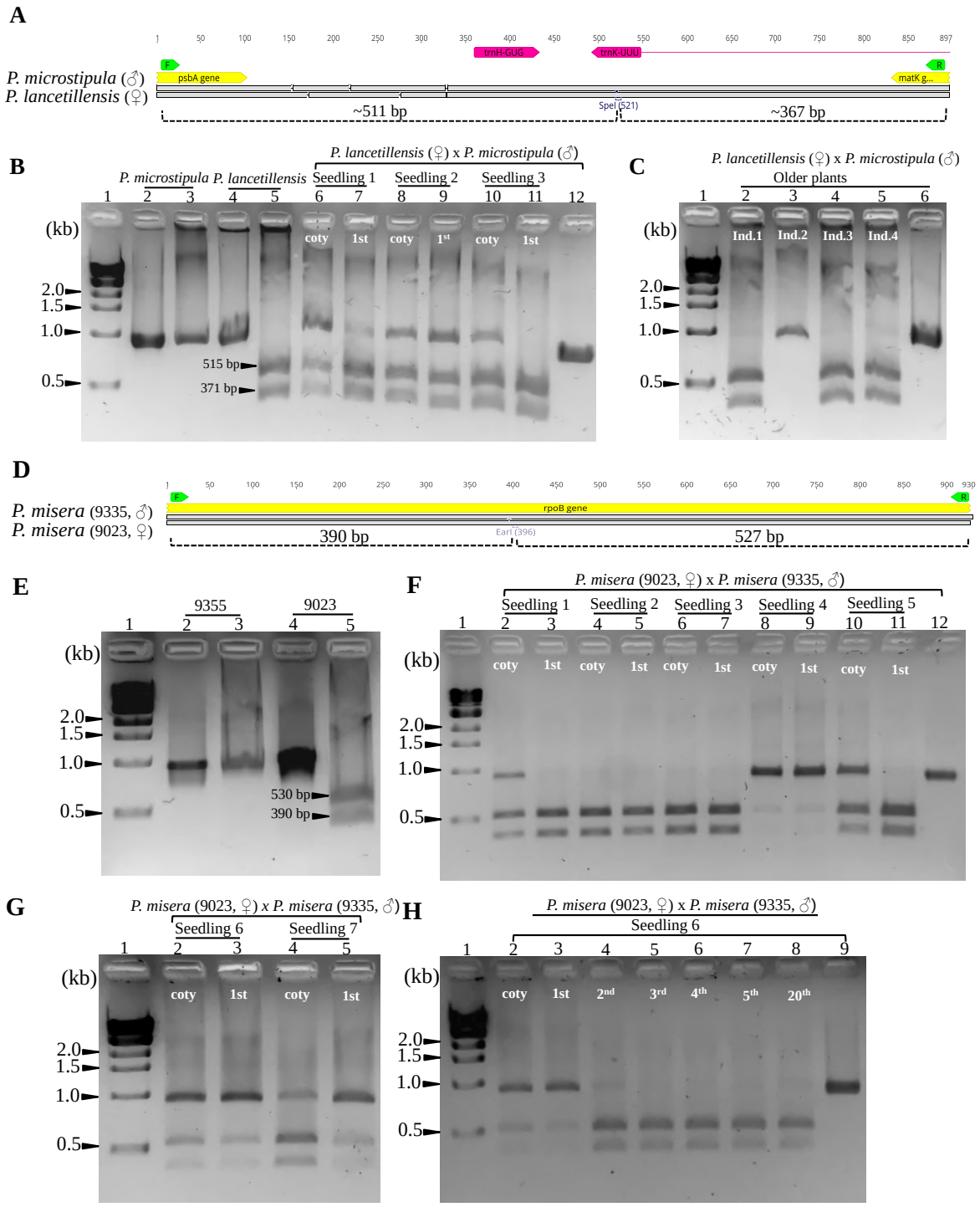


Figure 3.6. Agarose gels of PCR amplified target regions following digestion with restriction endonucleases in *Passiflora* hybrid seedlings and older leaves. (A) The intergenic region (*psbA-matK*) amplified from *P. lancetillensis* (♀) and *P. microstipula* (♂)

and their hybrids (2031) with estimated fragment sizes after SpeI digestion. (B) Lanes 2 and 4 are amplicons of *psbA-matK* from *P. microstipula* and *P. lancetillensis* and lanes 3 and 5 are products after SpeI digestion. Lanes 6-11 are SpeI digested PCR products amplified from hybrid seedlings (2031). For each seedling, the digested PCR products are shown for cotyledon (coty) and the first leaf (1^{st}). Lane 12, undigested PCR product of 2031 seedling 1 (cotyledon) as a reference. (C) Lanes 2-5, SpeI digestion products for four older hybrid (2031) individuals (Ind.), which were not included in the seedlings analyses. Lane 6, undigested PCR product for 2031 older individual 1 (Ind.1). (D) The partial coding region within *rpoB* amplified from *P. misera* 9023 (♀) and *P. misera* 9355 (♂) and their hybrid progeny (2027) with estimated fragment sizes after EarI digestion. (E) Lanes 2 and 4 are amplicons of the *rpoB* fragment for *P. misera* 9355 and *P. misera* 9023 and lanes 3 and 5 are products after EarI digestion. (F) Lanes 2-11, EarI digested PCR product for five hybrid seedlings (2027) of *P. misera* 9023 (♀) x *P. misera* 9355 (♂). For each seedling, the digested products are shown for cotyledon (coty) and the first leaf (1^{st}`). Lane 12, undigested PCR product amplified from 2027 seedling 1 (cotyledon). (G) Lanes 2-5, EarI digested PCR products for two additional hybrid seedlings of 2027, in the same order as F. (H) Lanes 2-8, EarI digested PCR products for seedling 6 of 2027 at different developmental stages, coty- cotyledons, 1^{st} - first leaf, 2^{nd} – second leaf, 3^{rd} – third leaf, 4^{th} – fourth leaf, 5^{th} – fifth leaf, 20^{th} – 20^{th} leaf. Lane 9, undigested PCR product of 2027 seedling 6 (20^{th} leaf) as a reference. Lane 1 in all gel images (B-C, E-H) contains one kb DNA ladder (NEB). DNA bands were separated in 2% agarose gel and visualized with RedSafe™ (INtRON Biotechnology).

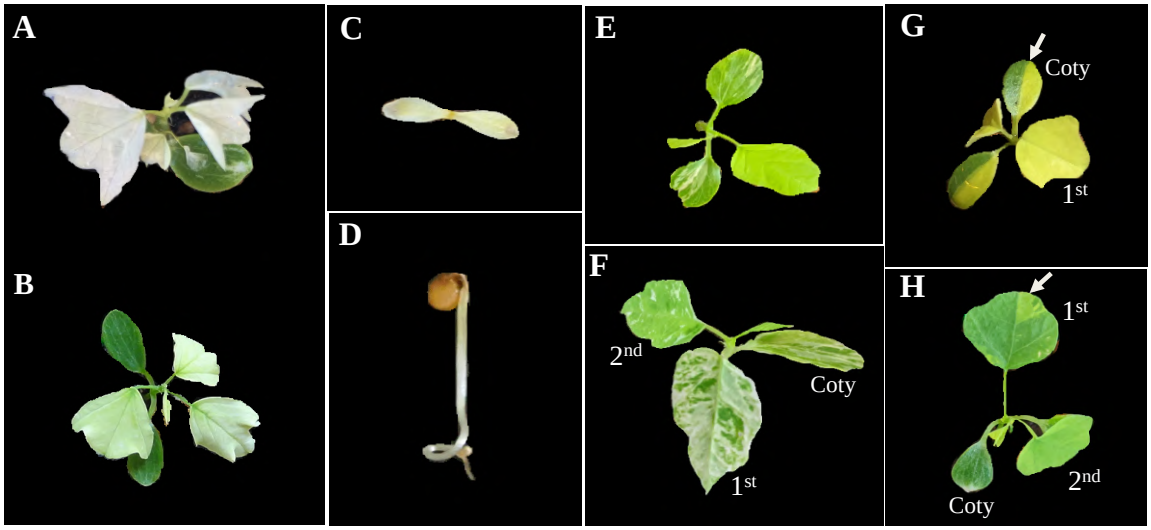


Figure 3.7. Seedlings of *Passiflora* hybrids. (A-B) *P. organensis* (♀) x *P. biflora* (♂) hybrids. The hybrids displayed green cotyledons with white (A) or greenish white (B) leaves and all hybrids died within a few weeks. (C-D) *P. rufa* (♀) x *P. auriculata* (♂) hybrids. The hybrids produce white/pale cotyledons (C) or hypocotyl (D) and die off before developing leaves. (E) *P. microstipula* (♀) x *P. lancetillensis* (♂) hybrid. Variegation was observed in cotyledons. (F) *P. lancetillensis* (♀) x *P. microstipula* (♂). Variegation was observed in cotyledons as well as few early leaves in the seedling. (G-H) *P. misera* (9023, ♀) x *P. misera* (9335, ♂) hybrids. Variegation was noted in cotyledons and occasionally in first leaf in some seedlings. The observed variegation phenotype often coordinates with vegetative segregation of biparentally inherited plastids detected through molecular analysis.

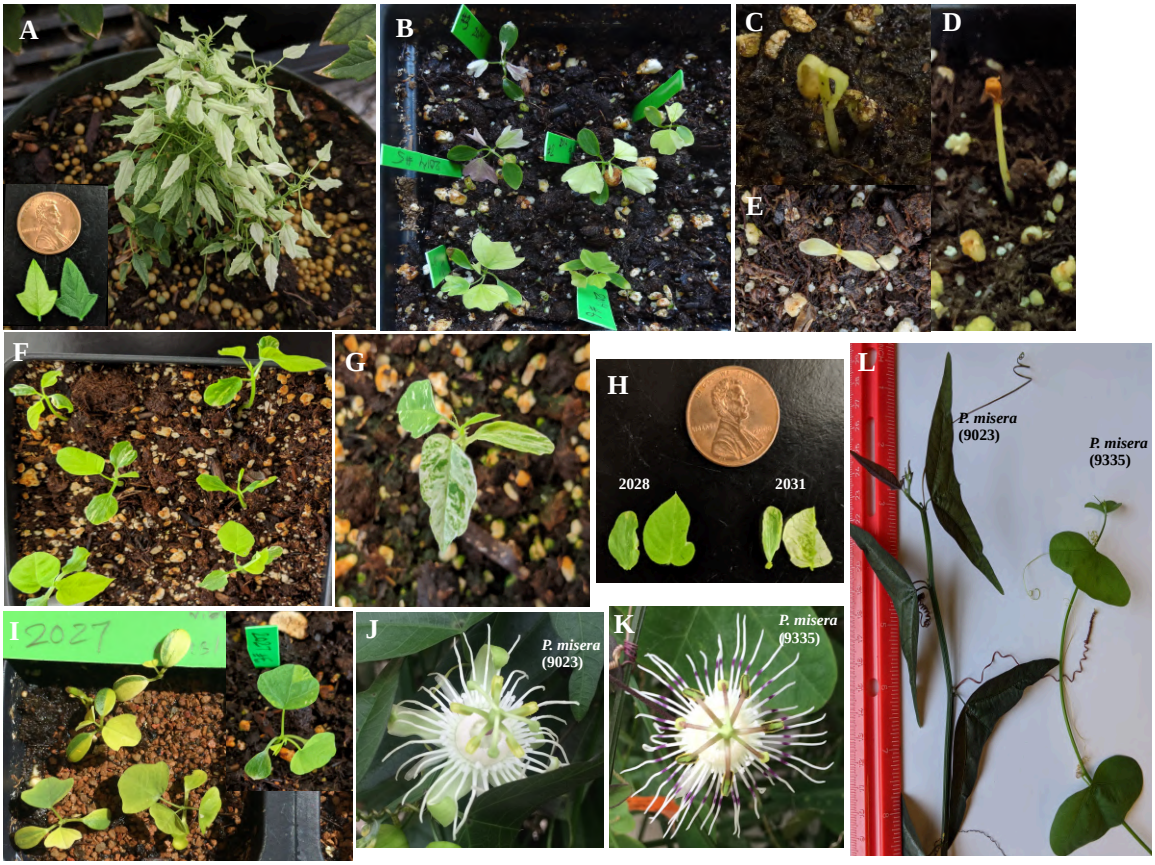


Figure 3.8. *Passiflora* hybrid seedlings and their parents. (A) 1028, *P. menispermifolia* (8039, ♀) x *P. hastifolia* (♂), six months old seedling. The inset shows the comparison of leaves size for a year old plant. (B) 2014, *P. organensis* (♀) x *P. biflora* (♂) hybrids. (C-E) 2005, *P. rufa* (♀) x *P. auriculata* (♂) hybrids. (F) 2028, *P. microstipula* (♀) x *P. lancetillensis* (♂) hybrids. (G) *P. lancetillensis* (♀) x *P. microstipula* (♂) hybrids. (H) Comparison of cotyledons and first leaf between 2028 and 2031. (I) 2027, *P. misera* (9023, ♀) x *P. misera* (9335, ♂) hybrids. The inset shows variegation in the first leaf. (J) *P. misera* (9023) flower. (K) *P. misera* (9335) flower. (L) Comparison of plant morphology between *P. misera* (9023) and *P. misera* (9335).

Table 3.1. Plastid inheritance in embryos of *Passiflora* hybrids. Hybrids highlighted in grey indicates reciprocal crosses. Abbreviations, P- Paternal, M- Maternal, B - Biparental.

Subgenus	Hybrids (♀ x ♂)	Accession	Inheritance
Passiflora	*P. retipetala* x *P. oerstedii*	1015	18P
	P. racemosa x *P. oerstedii*	1074	10P
	P. quadrangularis x *P. oerstedii*	1075	11P
	P. sprucei x *P. oerstedii*	1014	8P
	P. oerstedii x *P. sprucei*	1063	10P
	P. nephrodes x *P. sprucei*	1061	9P/1M
	P. choconiana x *P. sprucei*	1062	5M/5B
	P. speciosa x *P. sprucei*	1060	10P
	P. hastifolia x *P. miersii*	1003	10P
	P. oerstedii x *P. miersii*	1041	10P
	P. menspermifolia (8039) x *P. miersii*	1044	4P/4M/2B
	P. garckei x *P. choconiana*	1019	10P
	P. quadrangularis x *P. garckei*	1054	1P
	P. oerstedii x *P. nephrodes*	1002	10P
	P. menspermifolia (9224) x *P. oerstedii*	1004	10M
	P. oerstedii x *P. menspermifolia* (9244)	1006	10P
	P. oerstedii x *P. menspermifolia* (8039)	1031	10B
	P. vitifolia x *P. hastifolia*	1032	10P
	P. vitifolia x *P. quadrangularis*	1078	10P
	P. vitifolia x *P. edulis*	1077	11P
	P. vitifolia x *P. speciosa*	1053	12P
	P. menspermifolia (8039) x *P. hastifolia*	1028	4P/6B
	P. hastifolia x *P. menspermifolia* (8039)	1030	10B
	P. retipetala x *P. amethystina*	1072	10P
	P. retipetala x *P. menspermifolia* (Sirena)	1036	9P
	P. nephrodes x *P. choconiana*	1012	9M
	P. choconiana x *P. nephrodes*	1013	10P
	P. menspermifolia (8039) x *P. nephrodes*	1059	3M/7B
	P. choconiana x *P. menspermifolia* (9244)	1008	9P/1B
	P. choconiana x *P. menspermifolia* (Sirena)	1033	11P
	P. kermesina x *P. miersii*	1040	9P/1M
	P. kermesina x *P. hastifolia*	1023	10P
	P. sprucei x *P. retipetala*	1046	10P
	P. menspermifolia (8039) x *P. resticulata*	1068	3M/7B
	P. nitida x *P. quadrangularis*	1080	3P/7M
	P. speciosa x *P. polsea*	1073	10P
	P. miersii x *P. amethystina*	1071	10P
Decaloba	*P. rufra* x *P. jatunsachensis*	2001	1M/9B
	P. rufra x *P. auriculata*	2005	7M/3B
	P. organensis x *P. biflora*	2014	10M
	P. misera (9023) x *P. misera* (9335)	2027	3M/6B
	P. microstipula x *P. lancetillensis*	2028	10M
	P. lancetillensis x *P. microstipula*	2031	2P/8B
Astrophea	*P. sphaerocarpa* x *P. pittieri*	3002	8P/1M

Table 3.2. *Passiflora* species included in this study. Abrreviation, AN- Accession numbe; GBN- GenBank Number; NA- Not Available.

Subgenus	Species	AN	Location	Voucher	GBN
Passiflora	*P. oerstedii* Mast.	7005	Selva, Costa Rica	Shrestha106	MF807942
	P. choconiana (S. Wats.) Killip	8011	Altantida, Honduras	NA	NA
	P. nephrodes Mast.	9260		NA	NA
	P. menispermifolia Kunth. (Smooth leaves).	8039	Corcovado, Costa Rica	Shrestha202	MK694933
	P. menispermifolia Kunth. (Sirena)		Sirena, Costa Rica	NA	NA
	P. menispermifolia Kunth.	9224	Corcovado, Costa Rica	NA	NA
	P. hastifolia Killip		Bolivia	NA	NA
	P. retipetala Mast.	7007	Arima Pass, Trinidad	Shrestha108	MF807945
	P. quadrangularis L.	8054	Corcovada, Costa Rica	Shrestha107	MF807944
	P. nitida Kunth.	8060	Manaus, Brazil	Shrestha105	MF807941
	P. sprucei Mast.	9410		NA	NA
	P. resticulata Mast.and Andre.	9234		NA	NA
	P. amethystina Mikan	8061	Rio de Janeiro, Brazil	NA	NA
	P. miersii Mast.	9255	Sao Paulo, Brazil	NA	NA
	P. polsea			NA	NA
	P. speciosa Gardner	9274		NA	NA
	P. kermesina Link and Otto			NA	NA
	P. garckei Mast.	9080	French Guiana	NA	NA
	P. racemosa Brot.	9453		NA	NA
	P. vitifolia Kunth.	9041	Sirena, Costa Rica	Shrestha111	MF807947
	P. edulis Sims.	9408	Cali Valley, Columbia	Shrestha102	MF807938
Decaloba	*P. jatunsachensis* Schwerdtfeger	9402	Ecuador	Shrestha205	MK694920
	P. rufa Feuillet & MacDougal	9086	French Guiana	Shrestha209	MK694924
	P. auriculata Kunth.	8028	Costa Rica	NA	NA
	P. biflora Lam.	6001	Puerto Viejo, Costa Rica	Shrestha112	MF807937
	P. organensis Gardn.	9450		NA	NA
	P. misera Kunth.	9335	Mato Grosso do Sul, Brazil	NA	NA
	P. misera Kunth. (flattened stem)	9023	J. Turner, Leeds University, England	Shrestha208	MK694928
	P. microstipula Gilbert & MacDougal	9271	Vera Cruz, Mexico	Shrestha207	MK694934
	P. lancetillensis MacDougal and Meerman			NA	NA
Astrophea	*P. pittieri* Mast.	9219	Sirena, Costa Rica	Shrestha115	MF807943
	P. sphaerocarpa Tr .and Planch	9263		NA	NA

Table 3.3. List of *Passiflora* included in the study with the information regarding the ptDNA target regions, restriction enzymes used, amplicon size prior and post digestion. Abbreviation, AN- Accession number; bp- base pair. Reciprocal crosses are highlighted in grey. Asterik denotes the crosses performed by L. E. Gilbert. All other crosses were performed by B. Shrestha.

Subgenus	Hybrids (♀ x ♂)	AN	ptDNA target region	Restriction Enzyme (REase)	Parent with REase site	PCR product size	PCR product size post REase digestion
Passiflora	*P. oerstedii* x *P. nephrodes*	1002	*accD - rbcL*	BsmI (NEB)	*P. nephrodes*	~700 bp	~251 bp/447 bp
	P. menspermifolia (9224) x *P. oerstedii*	1004	*accD - rbcL*		*P. menispermifolia* (9224)	~700 bp	~251 bp/447 bp
	P. oerstedii x *P. menspermifolia* (9244)	1006	*accD - rbcL*			~700 bp	251 bp/447 bp
	P. oerstedii x *P. menspermifolia* (8039)	1031	*accD - rbcL*			~700 bp	251 bp/447 bp
	P. menspermifolia (8039) x *P. hastifolia*	1028	*accD - rbcL*		*P. menispermifolia* (8039)	~700 bp	251 bp/447 bp
	P. hastifolia x *P. menspermifolia* (8039)	1030	*accD - rbcL*			~700 bp	251 bp/447 bp
	P. menspermifolia (8039) x *P. nephrodes*	1059	*accD - rbcL*	SwaI (NEB)		700 - 800 bp	352 bp/346 bp
	P. nitida x *P. quadrangularis*	1080*	*atpF - atpH*	Bsp119I (Thermo Scientific)	*P. quadrangularis* (twice) vs. *P. nitida* (once)	~700 bp	286 bp/221 bp/199 bp vs. 501bp/199 bp
	P. nephrodes x P. choconiana	1012	intergenic - *trnL*-1	EcoRI (NEB)	*P. choconiana*	700-750 bp	~540 bp/200 bp
	P. choconiana x P. nephrodes	1013					
	P. choconiana x *P. menspermifolia* (9244)	1008					
	P. choconiana x *P. menspermifolia* (Sirena)	1033					
	P. sprucei x *P. retipetala*	1046	*psbK - trnS*	BsaI (NEB)	*P. retipetala*	700-750 bp	593 bp/191 bp
	P. menspermifolia (8039) x *P. resticulata*	1068			*P. resticulata*	700-750 bp	~593 bp/191 bp
	P. miersii x *P. amethystina*	1071			*P. amethystina*	~800 bp	~200 bp/600 bp
	P. retipetala x *P. oerstedii*	1015			*P. retipetala*	~890 bp	535 bp/365 bp
	P. sprucei x *P. oerstedii*	1014			*P. sprucei*	~890 bp	~535 bp/365 bp
	P. racemosa x *P. oerstedii*	1074			*P. racemosa*	~890 bp	~535 bp/365 bp
	P. quadrangularis x *P. oerstedii*	1075		NdeI (NEB)	*P. quadrangularis*	~890 bp	522 bp/372 bp
	P. hastifolia x *P. miersii*	1003					
	P. oerstedii x *P. miersii*	1041			*P. miersii*	~890 bp	~535 bp/365 bp
	P. menspermifolia (8039) x *P. miersii*	1044					
	P. garckei x *P. choconiana*	1019		HindIII (Thermo Scientific)	*P. choconiana*	~890 bp	~358 bp/525 bp
	P. quadrangularis x *P. garckei*	1054			*P. quadrangularis*	~890 bp	358 bp/536 bp
	P. nephrodes x *P. sprucei*	1061					
	P. oerstedii x *P. sprucei*	1063	*rpl32 - trnL*	NdeI (NEB)	*P. sprucei*	~890 bp	~535 bp/365 bp
	P. choconiana x *P. sprucei*	1062					
	P. speciosa x *P. sprucei*	1060					
	P. vitifolia x *P. hastifolia*	1032			*P. vitifolia*	~890 bp	521 bp/373 bp
	P. retipetala x *P. menspermifolia* (Sirena)	1036			*P. retipetala*	~890 bp	535 bp/365bp
	P. retipetala x *P. amethystina*	1072		Bsp119I (Thermo Scientific)		~890 bp	535 bp/365 bp
	P. vitifolia x *P. quadrangularis*	1078					
	P. vitifolia x *P. edulis*	1077		SwaI (NEB)	*P. vitifolia*		
	P. vitifolia x *P. speciosa*	1053				~890 bp	346 bp/548 bp
	P. kermesina x *P. miersii*	1040		HindIII (Thermo Scientific)	*P. miersii*	~890 bp	~ 356 bp/532 bp
	P. kermesina x *P. hastifolia*	1023			*P. hastifolia*	~890 bp	~ 356 bp/532 bp
	P. speciosa x *P. polsea*	1073		AflII (NEB)	*P. polsea*	~890 bp	~385 bp/510 bp
Decaloba	*P. rufra* x *P. jatunsachensis*	2001	*ycf4 - psaI*	Bsp119I (Thermo Scientific)	*P. jatunsachensis*	~800 bp	~ 296 bp/504 bp
	P. rufra x *P. auriculata*	2005	*rpoC1*	XhoI (Thermo Scientific)	*P. rufa*	~790 bp	186 bp/602 bp
	P. organensis x *P. biflora*	2014	*ycf4 - psaI*	Size difference		*P. biflora* (802 bp) / *P. organensis* (~500 bp)	
	P. misera (9335) *x P. misera* (9023)	2018	*rpoB*	EarI (NEB)	*P. misera* (9335)	~920 bp	390 bp/527 bp
	P. misera (9023) x *P. misera* (9335)	2027					
	P. microstipula x *P. lancetillensis*	2028*	*psbA - matK*	SpeI (NEB)	*P. lancetillensis*	~880 bp	~511 bp/367 bp
	P. lancetillensis x *P. microstipula*	2031*					
Astrophea	*P. sphaerocarpa* x *P. pittieri*	3002	*rpl32 - trnL*	Eco32I (NEB)	*P. pittieri*	~1130 bp	421 bp/712 bp

Table 3.4. Oligonucleotide primers used for PCR amplification and Sanger sequencing of the target regions. Abbreviation, bp- base pair.

Target region	Direction	Sequence	Product length
rpl32 - trnL	Forward Reverse	5' GAAAAGGATCTTGGGCAGCG 3' 5' GAGCAGCGTGTCTACCGATTTC 3'	~890 bp
intergenic - *trnL*-1	Forward Reverse	5' TTCCCCATCGACCCACTTGTC 3' 5' GACGTTATGCCGCTACTCGG 3'	700-750 bp
accD - rbcL	Forward Reverse	5' ACTGCAACAAGGCGTCCATA 3' 5' GCTGCTGCGTGTGAAGTATG 3'	700-800 bp
atpF - atpH	Forward Reverse	5' GGATGGCCAGTGAACCAAGA 3' 5' GCGGAGGGAAAAATACGAGG 3'	~700 bp
psbK - trnS	Forward Reverse	5' GGCTGAGTGGACTAAAGCGTC 3' 5' GGCTGAGTGGACTAAAGCGTC 3'	~800 bp
ycf4 - psaI	Forward Reverse	5' ACCACTACCCACATTCCACG 3' 5' GGCTTAGTCTTTCCAGCAATTG 3'	802 bp for *P. biflora* and ~500 bp for *P. organensis*
psbA - matK	Forward Reverse	5' CCGTGCTAACCTTGGTATGGA 3' 5' TTATACCGAGGGCGAGTTTGG 3'	~880 bp
rpoB	Forward Reverse	5' AGGTTCATTGATCAGGGCTT 3' 5' CGAACCAGAGCCAATCCGAA 3'	~920 bp
rpoC1	Forward Reverse	5' TGGATCAGTTTCGCCTCAACA 3' 5' TCGGGAGGAAGAACGGGTAAT 3'	~790 bp

Table 3.5. GenBank accession numbers (GBN) for the target region sequences generated with Sanger sequencing for *Passiflora* (*P.*) species and their hybrids.

Subgenus	Species	Target region	GBN
Passiflora	*P. oerstedii*	*rpl32-trnL*	MF807942
	P. retipetala	*rpl32-trnL*	
	P. sprucei	*rpl32-trnL*	
	P. racemosa	*rpl32-trnL*	
	P. quadrangularis	*rpl32-trnL*	
	P. hastifolia	*rpl32-trnL*	
	P. miersii	*rpl32-trnL*	
	P. menispermifolia (8039)	*rpl32-trnL*	
	P. menispermifolia (Sirena)	*rpl32-trnL*	
	P. menispermifolia (9244)	*rpl32-trnL*	
	P. garckei	*rpl32-trnL*	
	P. choconiana	*rpl32-trnL*	
	P. nephrodes	*rpl32-trnL*	
	P. speciosa	*rpl32-trnL*	
	P. vitifolia	*rpl32-trnL*	
	P. amethystina	*rpl32-trnL*	
	P. edulis	*rpl32-trnL*	
	P. kermesina	*rpl32-trnL*	
	P. polsea	*rpl32-trnL*	
	P. resticulata	*rpl32-trnL*	
	P. oerstedii	*accD-rbcL*	
	P. nephrodes	*accD-rbcL*	
	P. menispermifolia (9244)	*accD-rbcL*	
	P. menispermifolia (8039)	*accD-rbcL*	
	P. hastifolia	*accD-rbcL*	
	P. organensis	*ycf4-psaI*	
Decaloba	*P. misera* (9335)	*rpoB*	
	P. lancetillensis	*psbA-matK*	
Astrophea	*P. sphaerocarpa*	*rpl32-trnL*	

Table 3.6. Plastid inheritance in seedling and/or mature plant of *Passiflora* (*P.*) hybrids. Abbreviations, P- Paternal, M- Maternal, B – Biparental, NA – Not available. Asterisk denotes the hybrid seedling that occasionally produced cotyledons but not leaves and senesced.

Subgenus	Hybrids (♀ x ♂)	Accession	Inheritance			
			Embryo	Seedlings		Mature Plant
				Cotyledons	1st leaf	Leaf
Passiflora	*P. vitifolia* x *P. speciosa*	1053	12P	5P	5P	NA
	P. vitifolia x *P. edulis*	1077	11P	5P	5P	NA
	P. retipetala x *P. amenthystina*	1072	10P	4P	4P	NA
	P. miersii x *P. amenthystina*	1071	10P	2P	2P	NA
	P. vitifolia x *P. quadrangularis*	1078	10P	2P	2P	NA
	P. menspermifolia (8039) x *P. hastifolia*	1028	4P/2M/4B	3P	NA	3P
	P. speciosa x *P. sprucei*	1060	10P	3P	3P	NA
	P. nephrodes x *P. sprucei*	1061	9P/1M	1P	1P	NA
	P. oerstedii x *P. sprucei*	1063	10P	NA	NA	2P
	P. kermesina x *P. miersii*	1040	9P/1M	NA	NA	2P
Decaloba	*P. organensis* x *P. biflora*	2014	10M	6M	6M	NA
	P. rufra x *P. auriculata**	2005	1P/7M/2B	1M/1P/1B	NA	NA
	P. microstipula x *P. lancetillensis*	2028	10M	5M	5M	NA
	P. lancetillensis x *P. microstipula*	2031	2P/8B	3B	2B/1M	1M
	P. lancetillensis x *P. microstipula*	2031	2P/8B	NA	NA	3M/1P
	P. misera (9335) x *P. misera* (9023)	2018	NA	NA	NA	2M
	P. misera (9023) x *P. misera* (9335)	2027	3M/6B	2M/5B	4M/3B	6M/1P

References

Adams, K.L., Daley, D.O., Whelan, J., Palmer, J.D., 2002. Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. Plant cell. 14,931-943.

Adams, K.L., Palmer, J.D., 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol. Phylogenet. Evol. 29, 380–395.

Allen, J.F., 2018. Translating photosynthesis. Nature Plants 4, 199–200.

Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., Nielsen, H., 2019. Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance 2.

Ankenbrand, M. J., Pfaff, S., Terhoeven, N., Qureischi, M., Gündel, M., L Weiss, C., Hackl, T., Förster, F., 2018. ChloroExtractor: extraction and assembly of the chloroplast genome from whole genome shotgun data. J. Open Source Softw. 3, 464.

Azhagiri, A.K., Maliga, P., 2007. Exceptional paternal inheritance of plastids in Arabidopsis suggests that low-frequency leakage of plastids via pollen may be universal in plants. Plant J. 52, 817–823.

Barnard-Kubow, K.B., McCoy, M.A., Galloway, L.F., 2017. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. New Phytol. 213, 1466–1476.

Barrett, C.F., Baker, W.J., Comer, J.R., Conran, J.G., Lahmeyer, S.C., Leebens-Mack, J.H., Li, J., Lim, G.S., Mayfield-Jones, D.R., Perez, L., 2016. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytol. 209, 855–870.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Baur, E., 1909. Wesen und die Erbli- chkeitsverha..ltnisse der “Varietates albomarginatae hort.” von Pelargonium zonale. Zeitschrift für Induktive Abstammungsund Vererbungslehre 1: 330–351.

Benitez-Alfonso, Y., Cilia, M., Roman, A.S., Thomas, C., Maule, A., Hearn, S., Jackson, D., 2009. Control of Arabidopsis meristem development by thioredoxin-dependent regulation of intercellular transport. Proc. Natl. Acad. Sci.106, 3615–3620.

Bernt, M., Merkle, D., Ramsch, K., Fritzsch, G., Perseke, M., Bernhard, D., Schlegel, M., Stadler, P.F., Middendorf, M., 2007. CREx: inferring genomic rearrangements based on common intervals. Bioinformatics 23, 2957–2958.

Birky, C., 1983. Relaxed cellular controls and organelle heredity. Science 222, 468–475.

Birky, C.W., Walsh, J.B., 1992. Biased gene conversion, copy number, and apparent mutation rate differences within chloroplast and bacterial genomes. Genetics 130, 677–683.

Birky, C.W., 1994. Relaxed and Stringent Genomes: Why Cytoplasmic Genes Don't Obey Mendel's Laws. J Hered. 85, 355–365.

Birky, C.W., 2001. The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models. Annu. Rev. Genet. 35, 125–148.

Blazier, J.C., Guisinger, M.M., Jansen, R.K., 2011. Recent loss of plastid-encoded ndh genes within *Erodium* (Geraniaceae). Plant Mol. Biol. 76, 263–272.

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

Blazier, J. C., Jansen, R.K., Mower, J.P., Govindu, M., Zhang, J., Weng, M.-L., Ruhlman, T.A., 2016b. Variable presence of the inverted repeat and plastome stability in *Erodium*. Ann. Bot. 117, 1209–1220.

Bock, R., 2007. Structure, function, and inheritance of plastid genomes, in: Cell and Molecular Biology of Plastids. Springer, pp. 29–63.

Boender R., 2019. Passionflowers: A Pictorial Guide.

Bomblies, K., Weigel, D., 2007. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. Nat. Rev. Genet. 8, 382–393.

Bonen, L., Calixte, S., 2006. Comparative Analysis of Bacterial-Origin Genes for Plant Mitochondrial Ribosomal Proteins. Mol. Biol. Evol. 23, 701–712.

Brown, F.S., Snijder, R.C., van Tuyl, J.M., 2004. Biparental Plastid Inheritance in *Zantedeschia albomaculata* (Araceae). Acta Hortic. 673, 463-468.

Bruce, B.D., 2000. Chloroplast transit peptides: structure, function and evolution. Trends Cell Biol. 10, 440–447.

Bubunenko, M.G., Schmidt, J., Subramanian, A.R., 1994. Protein substitution in chloroplast ribosome evolution: a eukaryotic cytosolic protein has replaced its organelle homologue (L23) in spinach. J. Mol. Biol. 240, 28–41.

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., Berges, H., 2017. The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: Structural organization and phylogenomic studies in Malpighiales. Front. Plant Sci. 8, 334.

Cauz-Santos, L.A., da Costa, Z.P., Callot, C., Cauet, S., Zucchi, M.I., Bergès, H., van den Berg, C., Vieira, M.L.C., 2020. A repertory of rearrangements and the loss of an inverted repeat region in Passiflora chloroplast genomes. Genome Biol. Evol.

Chat, J., Chalak, L., Petit, R.J., 1999. Strict paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in intraspecific crosses of kiwifruit: Theor. Appl. Genet. 99, 314–322.

Chi, W., He, B., Mao, J., Jiang, J., Zhang, L., 2015. Plastid sigma factors: Their individual functions and regulation in transcription. Biochim. Biophy. Acta. Bioenerg. 1847, 770–778.

Chibani, K., Wingsle, G., Jacquot, J.-P., Gelhaye, E., Rouhier, N., 2009. Comparative Genomic Study of the Thioredoxin Family in Photosynthetic Organisms with Emphasis on Populus trichocarpa. Mol. Plant. 2, 308–322.

Chiu, W.-L., Stubbe, W., Sears, B.B., 1988. Plastid inheritance in Oenothera: organelle genome modifies the extent of biparental plastid transmission. Curr. Genet. 13, 181–189.

Choi, J.W., Graf, L., Peters, A.F., Cock, J.M., Nishitsuji, K., Arimoto, A., Shoguchi, E., Nagasato, C., Choi, C.G., Yoon, H.S., 2020. Organelle inheritance and genome architecture variation in isogamous brown algae. Sci.Rep. 10, 2048.

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

Collin, V., Issakidis-Bourguet, E., Marchand, C., Hirasawa, M., Lancelin, J.-M., Knaff, D.B., Miginiac-Maslow, M., 2003. The Arabidopsis Plastidial Thioredoxins NEW FUNCTIONS AND NEW INSIGHTS INTO SPECIFICITY. J. Biol. Chem. 278, 23747–23752.

Corriveau, J.L., Coleman, A.W., 1988. Rapid Screening Method to Detect Potential Biparental Inheritance of Plastid DNA and Results for Over 200 Angiosperm Species. Am. J. Bot. 75, 1443–1458.

Cosner, M.E., Raubeson, L.A., Jansen, R.K., 2004. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. BMC Evol. Biol. 4, 27.

Cruzan, M.B., Arnold, M.L., Carney, S.E., Wollenberg, K.R., 1993. cpDNA Inheritance in Interspecific Crosses and Evolutionary Inference in Louisiana Irises. Am. J. Bot. 80, 344–350.

Cusack, B.P., Wolfe, K.H., 2007. When gene marriages don’t work out: divorce by subfunctionalization. Trends Genet. 23, 270–272.

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

Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. PloS ONE 5, e11147.

Downie, S.R., Jansen, R.K., 2015. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid

transfer of DNA, and identification of highly divergent noncoding regions. Syst. Bot. 40, 336–351.

Downie, S.R., Palmer, J.D., 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny, in: Molecular Systematics of Plants. Springer, pp. 14–35.

Drouin, G., Daoud, H., Xia, J., 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol. Phylogenet. Evol. 49, 827–831.

Drescher, A., Ruf, S., Calsa, T., Carrer, H., Bock, R., 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. Plant J. 22, 97–104.

Eberhard, W.G., 1980. Evolutionary Consequences of Intracellular Organelle Competition. Q. Rev. Biol. 55, 231–249.

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

Ellis, J.R., Bentley, K.E., McCauley, D.E., 2008. Detection of rare paternal chloroplast inheritance in controlled crosses of the endangered sunflower Helianthus verticillatus. Heredity 100, 574–580.

Erixon, P., Oxelman, B., 2008. Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast clpP1 gene. PLoS ONE 3, e1386.

Fargo, D.C., Boynton, J.E., Gillham, N.W., n.d. Chloroplast Ribosomal Protein S7 of Chlamydomonas Binds to Chloroplast mRNA Leader Sequences and May Be Involved in Translation Initiation 13.

Felsenstein, J., 1989. PHYLIP – Phylogeny Inference Package (Version 3.2). Cladistics 5, 164-166.

Feuillet, C., MacDougal, J.M., 2003. A new infrageneric classification of *Passiflora* L. Passifloraceae). Passiflora 13, 34–38.

Figueroa, P., Gómez, I., Holuigue, L., Araya, A., Jordana, X., 1999. Transfer of rps14 from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron–sulphur subunit of succinate dehydrogenase and expression by alternative splicing. Plant J.18, 601–609.

Fleischmann, T.T., Scharff, L.B., Alkatib, S., Hasdorf, S., Schöttler, M.A., Bock, R., 2011. Nonessential plastid-encoded ribosomal proteins in tobacco: A developmental role for plastid translation and implications for reductive genome evolution. Plant Cell 23, 3137-3155.

Gantt, J.S., Baldauf, S.L., Calie, P.J., Weeden, N.F., Palmer, J.D., 1991. Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J. 10, 3073–3078.

Gelhaye, E., Rouhier, N., Navrot, N., Jacquot, J.P., 2005. The plant thioredoxin system. Cell. Mol. Life Sci. 62, 24–35.

Goffinet, B., Wickett, N.J., Shaw, A.J., Cox, C.J., 2005. Phylogenetic significance of the *rpoA* loss in the chloroplast genome of mosses. Taxon. 54, 353–360.

Gonçalves, D. J., Jansen, R. K., Ruhlman, T. A., Mandel, J. R., 2020. Under the rug: Abandoning persistent misconceptions that obfuscate organelle evolution. Mol. Phylogenet. Evol. 106903

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

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 29, 644–652.

Greiner, S., Wang, X., Rauwolf, U., Silber, M.V., Mayer, K., Meurer, J., Haberer, G., Herrmann, R.G., 2008. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. Nucleic Acids Res. 36, 2366–2378.

Greiner, S., Rauwolf, U.W.E., Meurer, J., Herrmann, R.G., 2011. The role of plastids in plant speciation. Mol. Ecol. 20, 671–691.

Greiner, S., Sobanski, J., Bock, R., 2015. Why are most organelle genomes transmitted maternally? BioEssays 37, 80–94.

Greiner, S., Golczyk, H., Malinova, I., Pellizzer, T., Bock, R., Börner, T., Herrmann, R.G., 2020. Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. Plant J. 102, 730–746.

Guisinger, M.M., Kuehl, J.V., Boore, J.L., Jansen, R.K., 2008. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. Proc. Natl. Acad. Sci. 105, 18424–18429.

Guisinger, M.M., Kuehl, J.V., Boore, J.L., Jansen, R.K., 2010. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. Mol. Biol. Evol. 28, 583–600.

Hagemann, R., 1979. Genetics and molecular biology of plastids of higher plants : plastids in male gametes, mutation induction, nitroso - urea, plastid mutants, thylakoid proteins, plastid ribosome deficiency, pollen, chloroplast, photosynthesis, mitochondria

Hagemann, R., Schröder, M-B., 1989. The cytological basis of the plastid inheritance in angiosperms. Protoplasma 152:57–64.

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.

Heath, T.A., Hedtke, S.M., Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46, 239–257.

Hillis, D.M., Pollock, D.D., McGuire, J.A., Zwickl, D.J., 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. 52, 124.

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

Huang, C.Y., Ayliffe, M.A., Timmis, J.N., 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature. 422, 72–76.

Huang, Y., Wang, J., Yang, Y., Fan, C., Chen, J., 2017. Phylogenomic Analysis and Dynamic Evolution of Chloroplast Genomes in Salicaceae. Front. Plant Sci. 8, 1050.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., Jensen, L.J., von Mering, C., Bork, P., 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44, D286–D293.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., et al., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci. 104, 19369–19374.

Jansen, R.K., Saski, C., Lee, S.-B., Hansen, A.K., Daniell, H., 2010. 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.

Jansen, R.K., Ruhlman, T.A., 2012. Plastid Genomes of Seed Plants, in: Bock, R., Knoop, V. (Eds.), Genomics of Chloroplasts and Mitochondria, Advances in Photosynthesis and Respiration. Springer Netherlands, Dordrecht, pp. 103–126.

Kadowaki, K., Kubo, N., Ozawa, K., Hirai, A., 1996. Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing targeting signals. EMBO J. 15, 6652–6661.

Kasmati, A.R., Töpel, M., Patel, R., Murtaza, G., Jarvis, P., 2011. Molecular and genetic analyses of Tic20 homologues in Arabidopsis thaliana chloroplasts. Plant J. 66, 877–889.

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

Kikuchi, S., Asakura, Y., Imai, M., Nakahira, Y., Kotani, Y., Hashiguchi, Y., Nakai, Y., Takafuji, K., Bédard, J., Hirabayashi-Ishioka, Y., Mori, H., Shiina, T., Nakai, M., 2018. A Ycf2-FtsHi Heteromeric AAA-ATPase Complex Is Required for Chloroplast Protein Import. Plant Cell. 30, 2677–2703.

Kikuchi, S., Bédard, J., Hirano, M., Hirabayashi, Y., Oishi, M., Imai, M., Takase, M., Ide, T., Nakai, M., 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. Science. 339, 571–574.

Kikuchi, S., Oishi, M., Hirabayashi, Y., Lee, D.W., Hwang, I., Nakai, M., 2009. A 1-Megadalton Translocation Complex Containing Tic20 and Tic21 Mediates Chloroplast Protein Import at the Inner Envelope Membrane. Plant Cell. 21, 1781–1797.

Killip, E.P., 1938. The American species of Passifloraceae. The American species of Passifloraceae.

Kirk, J.T.O., Tilney-Bassett, R.A.E., 1978. The Plastids, Their Chemistry, Structure, Growth, and Inheritance. Elsevier/North Holland Biomedical Press.

Kim, Y.-D., Jansen, R.K., 1994. Characterization and phylogenetic distribution of a chloroplast DNA rearrangement in the Berberidaceae. Plant Syst. Evol. 193, 107–114.

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

Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 28, 3211–3217.

Krosnick, S.E., Ford, A.J., Freudenstein, J.V., 2009. Taxonomic revision of *Passiflora* subgenus *Tetrapathea* including the monotypic genera *Hollrungia* and *Tetrapathea* (Passifloraceae), and a new species of *Passiflora*. Syst. Bot. 34, 375–385.

Krosnick, S.E., Porter-Utley, K.E., MacDougal, J.M., Jørgensen, P.M., McDade, L.A., 2013. New insights into the evolution of *Passiflora* subgenus *Decaloba* (Passifloraceae): phylogenetic relationships and morphological synapomorphies. Syst. Bot. 38, 692–713.

Kubo, N., Harada, K., Hirai, A., Kadowaki, K., 1999. A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: Common use of the same mitochondrial targeting signal for different proteins. Proc. Natl. Acad. Sci. 96, 9207–9211.

Lambowitz, A.M., Belfort, M., 1993. Introns as Mobile Genetic Elements. Annu. Rev. Biochem. 62, 587–622.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9, 357–359.

Larsson, J., Nylander, J.A., Bergman, B., 2011. Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. BMC Evol. Biol. 11, 187.

Lavin, M., Herendeen, P.S., Wojciechowski, M.F., 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Syst. Biol. 54, 575–594.

López-Giráldez, F., Townsend, J.P., 2011. PhyDesign: An online application for profiling phylogenetic informativeness. BMC Evol. Biol. 11, 152.

Lorenz-Lemke, A.P., Muschner, V.C., Bonatto, S.L., Cervi, A.C., Salzano, F.M., Freitas, L.B., 2005. Phylogeographic Inferences Concerning Evolution of Brazilian *Passiflora actinia* and *P. elegans* (Passifloraceae) Based on ITS (nrDNA) Variation. Ann. Bot. 95, 799–806.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955-964.

Ma, J., Yang, B., Zhu, W., Sun, L., Tian, J., Wang, X., 2013. The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. Gene 528, 120–131.

Magee, A.M., Aspinall, S., Rice, D.W., Cusack, B.P., Sémon, M., Perry, A.S., Stefanović, S., Milbourne, D., Barth, S., Palmer, J.D., 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res. 20, 1700-1710.

Maréchal, A., Parent, J.-S., Véronneau-Lafortune, F., Joyeux, A., Lang, B.F., Brisson, N., 2009. Whirly proteins maintain plastid genome stability in *Arabidopsis*. Proc. Natl. Acad. Sci. 106, 14693–14698.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., Penny, D., 2002. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. 99, 12246–12251.

Matsuo, M., Ito, Y., Yamauchi, R., Obokata, J., 2005. The Rice Nuclear Genome Continuously Integrates, Shuffles, and Eliminates the Chloroplast Genome to Cause Chloroplast–Nuclear DNA Flux. Plant Cell 17, 665–675.

Matsushima, R., Hu, Y., Toyoda, K., Sodmergen, Sakamoto, W., 2008. The Model Plant Medicago truncatula Exhibits Biparental Plastid Inheritance. Plant Cell Physiol. 49, 81–91.

McCauley, D.E., Sundby, A.K., Bailey, M.F., Welch, M.E., 2007. Inheritance of chloroplast DNA is not strictly maternal in Silene vulgaris (Caryophyllaceae): evidence from experimental crosses and natural populations. Am. J. Bot. 94, 1333–1337.

Medgyesy, P., Páy, A., Márton, L., 1986. Transmission of paternal chloroplasts in Nicotiana. Molec. Gen. Genet. 204, 195–198.

Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Lao, N.T., Heggie, L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C., Morden, C.W., 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13, 645–658.

Mogensen, H.L., 1996. INVITED SPECIAL PAPER: The hows and whys of cytoplasmic inheritance in seed plants. Am. J. Bot. 83, 383–404.

Mracek, J., n.d. Investigation of interspecific genome-plastome incompatibility in *Oenothera* and *Passiflora* 118.

Munhoz, C.F., Costa, Z.P., Cauz-Santos, L.A., Reátegui, A.C.E., Rodde, N., Cauet, S., Dornelas, M.C., Leroy, P., Varani, A. de M., Bergès, H., Vieira, M.L.C., 2018a. A

gene-rich fraction analysis of the Passiflora edulis genome reveals highly conserved microsyntenic regions with two related Malpighiales species. Sci. Rep. 8, 13024.

Muschner, V.C., Lorenz-Lemke, A.P., Vecchia, M., Bonatto, S.L., Salzano, F.M., Freitas, L.B., 2006. Differential organellar inheritance in *Passiflora's* (Passifloraceae) subgenera. Genetica 128, 449–453.

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.

Nagata, N., 2010. Mechanisms for independent cytoplasmic inheritance of mitochondria and plastids in angiosperms. J. Plant Res. 123, 193–199.

Nakai, M., 2015a. YCF1: A Green TIC: Response to the de Vries et al. Commentary. Plant Cell 27, 1834–1838.

Nakai, M., 2015b. The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. Biochim Biophy Acta Bioenerg. 1847, 957–967.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2014. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

Odahara, M., Inouye, T., Nishimura, Y., Sekine, Y., 2015a. RECA plays a dual role in the maintenance of chloroplast genome stability in *Physcomitrella patens*. Plant J. 84, 516–526.

Odahara, M., Masuda, Y., Sato, M., Wakazaki, M., Harada, C., Toyooka, K., Sekine, Y., 2015b. RECG maintains plastid and mitochondrial genome stability by suppressing extensive recombination between short dispersed repeats. PLoS Genet. 11, e1005080.

Okumura, S., Sawada, M., Park, Y.W., Hayashi, T., Shimamura, M., Takase, H., Tomizawa, K.-I., 2006. Transformation of poplar (Populus alba) plastids and expression of foreign proteins in tree chloroplasts. Transgenic Res 15, 637–646.

Oldenburg, D.J., Bendich, A.J., 2004. Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. J. Mol. Biol. 335, 953–970.

Palmer, J.D., Nugent, J.M., Herbon, L.A., 1987. Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. Proc. Natl. Acad. Sci. USA 84, 769–773.

Palmer, J.D., Thompson, W.F., 1981. Rearrangements in the chloroplast genomes of mung bean and pea. Proc. Natl. Acad. Sci. USA 78, 5533–5537.

Palmer, J.D. 1983. Chloroplast DNA exits in two orientations. Nature 301: 92-93.

Park, S., Jansen, R.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 Biol. 15, 40.

Park, S., Ruhlman, T.A., Weng, M.-L., Hajrah, N.H., Sabir, J.S., Jansen, R.K., 2017. Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of *Geranium*. Genome Biol. Evol. 9, 1766–1780.

Peltier, J.-B., Ytterberg, J., Liberles, D.A., Roepstorff, P., van Wijk, K.J., 2001. Identification of a 350 kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of *Arabidopsis thaliana*. J. Biol. Chem. 19, 16318-16327.

Peltier, J.-B., Ripoll, D.R., Friso, G., Rudella, A., Cai, Y., Ytterberg, J., Giacomelli, L., Pillardy, J., van Wijk, K.J., 2004. Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. J. Biol. Chem. 279, 4768–4781.

Perry, A.S., Wolfe, K.H., 2002. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. J. Mol. Evol. 55, 501–508.

Pollock, D.D., Zwickl, D.J., McGuire, J.A., Hillis, D.M., 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51, 664.

Pond, S.L.K., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies, in: Statistical Methods in Evolution. Springer, pp. 125–181.

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., Paterson, A.H., 2019. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biol 20, 38.

Rabah, S.O., Shrestha, B., Hajrah, N.H., Sabir, Mu. J., Alharby, H.F., Sabir, Me. J., Alhebshi, A.M., Sabir, J.S., Gilbert, L.E., Ruhlman, T.A., Jansen, R.K. 2019. *Passiflora* plastome sequencing reveals widespread genomic rearrangements. J. Syst. Evol. 57, 1–14.

Ramsey, A.J., Mandel, J.R., 2019. When One Genome Is Not Enough: Organellar Heteroplasmy in Plants, in: Annual Plant Reviews Online. 619–658.

Raubeson, L.A., Jansen, R.K. 2005. Chloroplast genomes of plants. Plant diversity and evolution: genotypic and phenotypic variation in higher plants. CABI Publishing. London.

Reboud, X., Zeyl, C., 1994. Organelle inheritance in plants. Heredity 72, 132–140.

Rieseberg, L.H., Blackman, B.K., 2010. Speciation genes in plants. Ann. Bot. 106, 439–455.

Rodermel, S., 2002. Arabidopsis Variegation Mutants. Arabidopsis Book/American Society of Plant Biologists, 1.

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

Ronquist, F., Sanmartín, I., 2011. Phylogenetic Methods in Biogeography. Annu. Rev. Ecol. Evol. Syst. 42, 441–464.

Roy, S., Ueda, M., Kadowaki, K., Tsutsumi, N., 2010. Different status of the gene for ribosomal protein S16 in the chloroplast genome during evolution of the genus Arabidopsis and closely related species. Genes Genet. Syst. 85, 319–326.

Ruf, S., Karcher, D., Bock, R., 2007. Determining the transgene containment level provided by chloroplast transformation. Proc. Natl. Acad. Sci. 104, 6998–7002.

Ruhlman, T.A., Jansen, R.K., 2014. The Plastid Genomes of Flowering Plants, in: Maliga, P. (Ed.), Chloroplast Biotechnology: Methods and Protocols, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 3–38.

Ruhlman, T.A., Zhang, J., Blazier, J.C., Sabir, J.S., Jansen, R.K., 2017. Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. Am. J. Bot. 104, 559–572.

Ruhlman, T.A., Jansen, R.K., 2018. Aberration or analogy? The atypical plastomes of Geraniaceae. In Adv. Bot. Res. 85, 223–262.

Ruwe, H., Kupsch, C., Teubner, M., Schmitz-Linneweber, C., 2011. The RNA-recognition motif in chloroplasts. Journal of Plant Physiology, Regulation of plant primary metabolism 168, 1361–1371.

Sabir, J., Schwarz, E., Ellison, N., Zhang, J., Baeshen, N.A., Mutwakil, M., Jansen, R., Ruhlman, T., 2014. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. Plant Biotechnol. J. 12, 743–754.

Sanderson, M.J., Copetti, D., Búrquez, A., Bustamante, E., Charboneau, J.L., Eguiarte, L.E., Kumar, S., Lee, H.O., Lee, J., McMahon, M., 2015. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the ndh gene suite and inverted repeat. Am. J. Bot. 102, 1115–1127.

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

Scharff, L.B., Koop, H.-U., 2006. Linear molecules of tobacco ptDNA end at known replication origins and additional loci. Plant Mol. Biol. 62, 611–621.

Schumann, C.M., Hancock, J.F., 1989. Paternal inheritance of plastids in Medicago sativa. Theoret. Appl. Genetics 78, 863–866.

Schürmann, P., Jacquot, J.-P., 2000. Plant Thioredoxin Systems Revisited. Annual Review of Plant Physiology and Plant Mol.Biol. 51, 371–400.

Schwarz, E.N., Ruhlman, T.A., Sabir, J.S., Hajrah, N.H., Alharbi, N.S., Al-Malki, A.L., Bailey, C.D., Jansen, R.K., 2015. Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids. J. Syst. Evol. 53, 458–468.

Schwarz, E.N., Ruhlman, T.A., Weng, M.-L., Khiyami, M.A., Sabir, J.S., Hajarah, N.H., Alharbi, N.S., Rabah, S.O., Jansen, R.K., 2017. Plastome-wide nucleotide substitution rates reveal accelerated rates in Papilionoideae and correlations with genome features across legume subfamilies. J. Mol. Evol. 84, 187–203.

Sears, B.B., 1980. Elimination of plastids during spermatogenesis and fertilization in the plant kingdom. Plasmid 4, 233–255.

Serino, G., Maliga, P., 1998. RNA polymerase subunits encoded by the plastid rpo genes are not shared with the nucleus-encoded plastid enzyme. Plant Physiol. 117, 1165–1170.

Shi, X., Bentolila, S., Hanson, M.R., 2016. Organelle RNA recognition motif-containing (ORRM) proteins are plastid and mitochondrial editing factors in Arabidopsis. Plant Signaling & Behavior 11, e1167299.

Shoji, S., Dambacher, C.M., Shajani, Z., Williamson, J.R., Schultz, P.G., 2011. Systematic chromosomal deletion of bacterial ribosomal protein genes. J. Mol. Biol. 413, 751–761.

Shore, J.S., McQueen, K.L., Little, S.H., 1994. Inheritance of Plastid DNA in the *Turnera ulmifolia* Complex (Turneraceae). Am. J. Bot. 81, 1636–1639.

Shore, J.S., Triassi, M., 1998. Paternally biased cpDNA inheritance in *Turnera ulmifolia* (Turneraceae). Am. J. Bot. 85, 328–332.

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. Phylogent. Evol. 138, 53–64.

Shrestha, B., Gilbert, L.E., Ruhlman, T.A., Jansen, R.K., 2020. Rampant nuclear transfer and substitutions of plastid genes in Passiflora. Genome Biol. Evol.

Sloan, D.B., Alverson, A.J., Wu, M., Palmer, J.D., Taylor, D.R., 2012. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. Genome Biol. Evol. 4, 294–306.

Sloan, D.B., Triant, D.A., Wu, M., Taylor, D.R., 2013. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Mol. Biol. Evol. 31, 673–682.

Small, I., Peeters, N., Legeai, F., Lurin, C., 2004. Predotar: A tool for rapidly screening proteomes forN-terminal targeting sequences. PROTEOMICS 4, 1581–1590.

Smith, S.E., 1989. Influence of Parental Genotype on Plastid Inheritance in *Medicago sativa*. J Hered. 80:214–217.

Smith, S.E., Bingham, E.T., Fulton, R.W., 1986. Transmission of chlorophyll deficiencies in *Medicago sativa*. J. Hered. 77:35–38.

Sokolenko, A., Pojidaeva, E., Zinchenko, V., Panichkin, V., Glaser, V., Herrmann, R., Shestakov, S., 2002. The gene complement for proteolysis in the cyanobacterium Synechocystis sp. PCC 6803 and Arabidopsis thaliana chloroplasts. Curr. Genet. 41, 291–310.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., de Pamphilis, C.W., Wall, P.K., Soltis, P.S., 2009. Polyploidy and angiosperm diversification. Am. J. Bot. 96, 336–348.

Sperschneider, J., Catanzariti, A.-M., DeBoer, K., Petre, B., Gardiner, D.M., Singh, K.B., Dodds, P.N., Taylor, J.M., 2017. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci. Rep. 7, 44598.

Stegemann, S., Hartmann, S., Ruf, S., Bock, R., 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. Proc. Natl. Acad. Sci. 100, 8828–8833.

Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, C., Sugita, M., 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: Evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. Nucleic Acids Res. 31, 5324–5331.

Sun, T., Germain, A., Giloteaux, L., Hammani, K., Barkan, A., Hanson, M.R., Bentolila, S., 2013. An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. Proc. Natl. Acad. Sci. 110, E1169–E1178.

Sun, Y., Moore, M.J., Meng, A., Soltis, P.S., Soltis, D.E., Li, J., Wang, H., 2013. Complete plastid genome sequencing of Trochodendraceae reveals a significant expansion of the inverted repeat and suggests a Paleogene divergence between the two extant species. PLoS ONE 8, e60429.

Szmidt, A.E., Aldén, T., Hällgren, J-E., 1987. Paternal inheritance of chloroplast DNA in *Larix*. Plant Mol. Biol. 9:59–64.

Team, R.C., 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Thanh, N.D., Medgyesy, P., 1989. Limited chloroplast gene transfer via recombination overcomes plastomegenome incompatibility between Nicotiana tabacum and Solanum tuberosum. Plant Mol. Biol. 12, 87–93.

The Angiosperm Phylogeny Group, 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. 181, 1–20.

Thyssen, G., Svab, Z., Maliga, P., 2012. Exceptional inheritance of plastids via pollen in Nicotiana sylvestris with no detectable paternal mitochondrial DNA in the progeny. Plant J. 72, 84–88.

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. 45, W6–W11.

Tilney-Bassett, R.A.E., Birky, C.W., 1981. The mechanism of the mixed inheritance of chloroplast genes in *Pelargonium*: Evidence from gene frequency distributions among the progeny of crosses. Theoret. Appl. Genetics 60, 43–53.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y., Martin, W., 2004. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5, 123.

Trusty, J.L., Johnson, K.J., Lockaby, B.G., Goertzen, L.R., 2007. Bi-Parental Cytoplasmic DNA Inheritance in Wisteria (Fabaceae): Evidence from a Natural Experiment. Plant Cell Physiol 48, 662–665.

Ueda, M., Fujimoto, M., Arimura, S., Murata, J., Tsutsumi, N., Kadowaki, K., 2007. Loss of the *rpl32* gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*. Gene 402, 51–56.

Ueda, M., Nishikawa, T., Fujimoto, M., Takanashi, H., Arimura, S., Tsutsumi, N., Kadowaki, K., 2008. Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. Mol. Biol. Evol. 25, 1566–1575.

Ulmer, T., MacDougal, J.M., 2004. *Passiflora*: passionflowers of the world. Timber Press (OR).

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3—new capabilities and interfaces. Nucleic Acids Res. 40, e115–e115.

Vanderplank, J., 1996. Passion Flowers. The MIT Press Cambridge, Massachusetts. 2nd ed.

Wagner, R., Aigner, H., Funk, C., 2012. FtsH proteases located in the plant chloroplast. Physiol. Plant. 145, 203–214.

Wang, P., Liu, J., Liu, B., Feng, D., Da, Q., Wang, P., Shu, S., Su, J., Zhang, Y., Wang, J., Wang, H.-B., 2013. Evidence for a Role of Chloroplastic m-Type Thioredoxins in the Biogenesis of Photosystem II in Arabidopsis. Plant Physiology 163, 1710–1728.

Wang, R.-J., Cheng, C.-L., Chang, C.-C., Wu, C.-L., Su, T.-M., Chaw, S.-M., 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol. Biol. 8, 36.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M., 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol. Biol. Evol. 35, 543–548.

Weng, M.-L., Ruhlman, T.A., Gibby. M., Jansen, R.K. 2012. Phylogeny, rate variation and genome size evolution of *Pelargonium* (Geraniaceae). Mol. Phylogenet. Evol. 64. 654-670.

Weng, M.-L., Blazier, J.C., Govindu, M., Jansen, R.K., 2013. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome

rearrangements, repeats, and nucleotide substitution rates. Mol. Biol. Evol. 31, 645–659.

Weng, M.-L., Ruhlman, T.A., Jansen, R.K., 2016. Plastid–nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. Genome Biol. Evol. 8, 1824–1838.

Weng, M.-L., Ruhlman, T.A., Jansen, R.K., 2017. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. New Phytol. 214, 842–851.

Wolfe, A.D., Randle, C.P., 2004. Recombination, Heteroplasmy, Haplotype Polymorphism, and Paralogy in Plastid Genes: Implications for Plant Molecular Systematics. Sys. Bot. 29, 1011–1020.

Wolfe, A.D., Randle, C.P., 2001. Relationships within and among Species of the Holoparasitic Genus Hyobanche (Orobanchaceae) Inferred from ISSR Banding Patterns and Nucleotide Sequences. Sys. Bot. 26, 120–130.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. Springer.

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

Williams, A.M., Friso, G., van Wijk, K.J., Sloan, D.B., 2019. Extreme variation in rates of evolution in the plastid Clp protease complex. Plant J.

Williams, A.V., Boykin, L.M., Howell, K.A., Nevill, P.G., Small, I., 2015. The complete sequence of the *Acacia ligulata* chloroplast genome reveals a highly divergent clpP1 gene. PLoS ONE 10, e0125768.

Wischmann, C., Schuster, W., 1995. Transfer of rps10 from the mitochondrion to the nucleus in Arabidopsis thaliana: evidence for RNA-mediated transfer and exon shuffling at the integration site. FEBS Lett. 374, 152–156.

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. 91, 1846–1862.

Wolf, J.B., Künstner, A., Nam, K., Jakobsson, M., Ellegren, H., 2009. Nonlinear dynamics of nonsynonymous (*dN*) and synonymous (*dS*) substitution rates affects inference of selection. Genome Biol. Evol. 1, 308–319.

Wolfe, K.H., Li, W.-H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84, 9054–9058.

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

Xu, J., 2005. The inheritance of organelle genes and genomes: patterns and mechanisms. Genome 48, 951–958.

Xu, Y.-Z., Arrieta-Montiel, M.P., Virdi, K.S., de Paula, W.B., Widhalm, J.R., Basset, G.J., Davila, J.I., Elthon, T.E., Elowsky, C.G., Sato, S.J., 2011. MutS HOMOLOG1 is a nucleoid protein that alters mitochondrial and plastid properties and plant response to high light. Plant Cell 23, 3428-3441.

Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Yang, T.W., Yang, Y.A., Xiong, Z., 2000. Paternal inheritance of chloroplast DNA in interspecific hybrids in the genus *Larrea* (Zygophyllaceae). American Journal of Botany 87, 1452–1458.

Yoshida, T., Furihata, H.Y., Kawabe, A., 2014. Patterns of Genomic Integration of Nuclear Chloroplast DNA Fragments in Plant Species. DNA Res. 21, 127–140.

Zerbino, D., Birney, E., 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18, 821-829.

Zhang, J., Ruhlman, T.A., Mower, J.P., Jansen, R.K., 2013. Comparative analyses of two Geraniaceae transcriptomes using next-generation sequencing. *BMC Plant Biol*. 13:228.

Zhang, J., Ruhlman, T.A., Sabir, J., Blazier, J.C., Jansen, R.K., 2015. Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. Plant Cell 27, 563-573.

Zhang, Q., Liu, Y., Sodmergen, 2003. Examination of the Cytoplasmic DNA in Male Reproductive Cells to Determine the Potential for Cytoplasmic Inheritance in 295 Angiosperm Species. Plant Cell Physiol. 44, 941–951.

Zhang, Q., Sodmergen, 2010. Why does biparental plastid inheritance revive in angiosperms? J. Plant Res. 123, 201–206.

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

Vita

Bikash Shrestha was born in Bhaktapur, Nepal. He received Bachelor degree in Biotechnology from Kathmandu University, Nepal, in 2007 and moved to USA to pursue Master's degree in Biology at the University of Nebraska at Omaha in 2009. After completion of Master's program, he worked in a synthetic biology lab at Integrated DNA technologies Inc. from 2012-2014. In August 2014, he embarked on the graduate school at the University of Texas at Austin to pursue Ph.D. in plant biology.

Email: b.shrestha@utexas.edu

This manuscript was typed by Bikash Shrestha.