UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FUNGOS, ALGAS E PLANTAS

Victor Soares Santibanez

New genetic markers and gene flow analysis shed light into *Pseudotrimezia* incongruence

Florianópolis
2023

Victor Soares Santibanez

New genetic markers and gene flow analysis shed light into *Pseudotrimezia* incongruence

Dissertação submetida ao Programa de Pós-Graduação em Fungos, Algas e Plantas da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Mestre em Biologia de Fungos, Algas e Plantas.

Orientadora: Dra. Suzana Alcantara
Coorientadora: Dra. Juliana Lovo

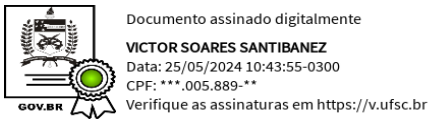
Florianópolis
2023

Santibanez, Victor Soares
New genetic markers and gene flow analysis shed light into Pseudotrimezia incongruence / Victor Soares Santibanez ; orientadora, Suzana de Fátima Alcantara, coorientador, Juliana Lovo Lovo, 2023.
44 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Biologia de Fungos, Algas e Plantas, Florianópolis, 2023.

Inclui referências.

1. Biologia de Fungos, Algas e Plantas. 2. sistemática filogenética, evolução, fluxo gênico. I. Alcantara, Suzana de Fátima. II. Lovo, Juliana Lovo. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Biologia de Fungos, Algas e Plantas. IV. Título.



Victor Soares Santibanez

New genetic markers and gene flow analysis shed light into *Pseudotrimezia* incongruence

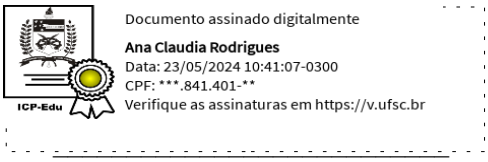
O presente trabalho em nível de Mestrado foi avaliado e aprovado, em vinte de outubro de 2023, pela banca examinadora composta pelos seguintes membros:

Dr. Marcelo Reginato
Universidade Federal do Rio Grande do Sul

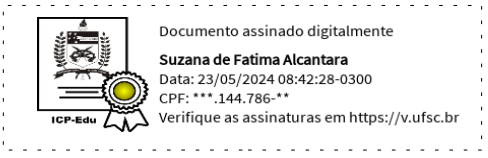
Dr. Alison Gonçalves Nazareno
Universidade Federal de Minas Gerais

Dra. Duane Fernandes de Souza Lima
Universidade Federal de Santa Catarina

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Biologia de Fungos, Algas e Plantas.



Coordenação do Programa de Pós-Graduação



Dra. Suzana Alcantara

Florianópolis, 2023.

AGRADECIMENTOS

Primeiramente, Agradeço a orientadora deste trabalho, Suzana Alcantara. Obrigado pelos 6 anos de parceria, pelas sugestões, contribuições, leituras e releituras que possibilitram esse trabalho. Obrigado também por todos os ensinamentos sobre sistemática, filogenética, evolução e biogeografia que fizeram com que eu fosse capaz de escrever este trabalho. Agradeço também a minha coorientadora, Juliana Lovo. Obrigado pelas leituras e releituras e contribuições que me ajudaram a conhecer melhor o grupo e enriqueceram muito este trabalho.

Agradeço as coordenadoras do curso de pós graduação, Mayara Caddah e posteriormente Fernanda Cordeiro de Oliveira por se mostrarem sempre disponíveis e dispostas a ajudar.

Agradeço também a UFSC e ao Departamento de Pós graduação, em especial ao Kleyton Steinbach, que apesar de não conhecer em pessoa, sempre se mostrou extremamente eficiente e solicito em resolver minhas questões estudantis.

Agradeço também a Capes, a Fapesp e ao Instituto Serrapilheira pelo financiamento que tornou esse trabalho possível.

Finalmente, agradeço a minha amiga Tâmela Madaloz e aos meus alunos por ouvirem as minhas reclamações durante períodos difíceis deste processo, e ao Pipoca, o Bob, a Lupita e a Gertrudes pela companhia que fez com que o percurso fosse mais leve.

RESUMO

Apesar dos avanços da sistemática molecular, alguns clados diversos de radiação recente ainda apresentam desafios quanto à reconstrução das relações de parentesco pela predominância de complexas dinâmicas de fluxo gênico entre espécies distintas, como coalescência e reticulação. Este é o caso do clado *Pseudotrimezia* da tribo Trimezieae (Iridaceae). Este clado é majoritariamente endêmico do *campo rupestre,* uma região de afloramentos rochosos do leste do Brasil, rica em espécies. Estudos prévios demonstraram a presença de incongruência entre marcadores genômicos nucleares e cloroplastidiais neste clado, e foram incapazes de recuperar uma filogenia com alta resolução para o grupo. Neste trabalho, utilizei uma nova matriz de marcadores do tipo RAD para melhorar a resolução da filogenia existente para o grupo e investigar a incongruência encontrada a partir de análises de fluxo gênico, que indicaram índices significativos de reticulação. Esta análise, unida à distribuição geográfica dos espécimes, indicou eventos frequentes de hibridização dentro do grupo.

Palavras-chave: fluxo gênico; Iridaceae, RAD-seq**.**

ABSTRACT

Even considering recent advancements in molecular systematics, some recently diverged clades still pose challenges to phylogenetic reconstruction because of the prevalence of complex gene tree dynamics such as deep coalescence and reticulation. This is the case for the clade *Pseudotrimezia* from the tribe Trimezieae (Iridaceae). This clade is mostly endemic to the *campo rupestre*, a rocky outcrop region of eastern Brazil, which is species rich. Previous studies showed a presence of incongruence between nuclear genomic markers and chloroplast markers in this clade and were unable to recover a phylogeny with high resolution for the group. A new dataset of RAD-seq genetic markers was used to improve the existing phylogeny resolution of the group and probe into the incongruence found through gene flow analysis, which indicated significant signals of reticulation. These analyses, compared to the geographical distribution of the specimens, indicated frequent events of hybridization in the group.

Keywords: gene flow; Iridaceae; RAD-seq.

RESUMO EXPANDIDO

Introdução

Árvores filogenéticas de espécies podem diferir de árvores filogenéticas gênicas. Eventos como introgressão e retenção de polimorfismos ancestrais torna a inferência filogenética difícil para métodos filogenéticos clássicos. Estes eventos evolutivos são comuns em clados que apresentam radiação recente, que é o caso de muitas espécies do Campo Rupestre. As *Pseudotrimezias* são, em sua maioria, endêmicas a esta região e são um exemplo da dificuldade mencionada. *Pseudotrimezia* são em sua maioria endêmicas do sul da Serra do Espinhaço em Chapada Diamantina, Minas Gerais, Brasil. As duas exceções são *P. junciolia* (Klatt) Lovo & A.Gil e *P. pusilla* (Ravenna) A.Gil; com as duas ocorrendo em vegetação de cerrado. A maioria das espécies apresenta flores amarelas simples, com outras ornamentadas também presentes no grupo. Estudos de ontologia floral revelaram que o grupo apresentou repetidos casos de desenvolvimento para características de flor simples a partir de ancestrais ornamentados. Estudos moleculares iniciais revelaram polifiletismo usando os marcadores cloroplastidiais (*trnG* intron, *trnH-psbA* intergenic spacer, *trnK* intron) e nuclear (nrITS) do tipo Sanger. Testes de congruência também revelaram incongruência entre os grupos de marcadores plastidiais e o nuclear. A recircunscrição do gênero *Pseudotrimezia* foi então proposta posteriormente. Apesar disso, a baixa resolução das relações entre as espécies e a incongruência encontrada poderia informar sobre os processos evolutivos atuantes no grupo. Neste caso, congruência se refere à concordância de diferentes árvores gênicas entre si e com a verdadeira história do grupo. Em contraste, incongruência se refere a árvores gênicas e filogenéticas discordantes, ou seja contando histórias diferentes. Esta discordância pode ser causada por hibridização ou pela retenção de polimorfismos ancestrais. A dificuldade em recuperar filogenias com alta sustentação tornou popular o uso de métodos de sequenciamento de baixa representatividade como o RAD-seq, que é um método de sequenciamento de nova geração que utiliza marcadores associados a sítios de restrição presentes por todo o genoma. Este método também é compatível com métodos que testam a variação de frequência de alelos, que é uma forma de testar o fluxo gênico entre dois grupos. A presença, variação e direção do fluxo gênico podem então ser utilizadas para informar hipóteses sobre a história da interação entre grupos de organismos.

Objetivos

Tendo isso, este trabalho teve o objetivo de investigar as incongruências encontradas em estudos anteriores através das estatísticas D, f4 e f-branch. Mais especificamente, diferenciar o efeito da retenção de polimorfismos ancestrais causada pela rápida especiação do grupo do isolamento reprodutivo incipiente que causa a hibridização. Então, comparei os padrões de fluxo gênico com a distribuição geográfica das amostras para verificar se as hipóteses de hibridização são plausíveis dada a distribuição. Além disso, utilizei este novo banco de dados de amostras de sequenciamento RAD para melhorar a resolução da filogenia atual do gênero *Pseudotrimezia*.

Metodologia

A amostragem foi feita tentando incluir o maior número de espécies possíveis. Além disso, repetidas entradas foram incluídas para espécies que não apresentaram monofiletismo recíproco em estudos anteriores. O banco de dados total inclui 31 terminais representando 16 das 24 espécies atualmente aceitas em *Pseudotrimezia,* mais um outgroup do gênero *Trimezia.* Junto com o material de sequenciamento, foram compiladas as localizações geográficas de

cada uma das coletas. As amostras foram obtidas usando a enzima de sítio de restrição Pst1 e submetidas ao protocolo de fita única em um sequenciador Illumina GAIIx. Com o software Ipyrad, os dados crus foram filtrados para baixa qualidade e adaptadores remanescentes do processo de sequenciamento. Em seguida, os reads das amostras foram alinhados entre si, e uma sequência consenso foi gerada para cada terminal. Estas sequências foram por sua vez alinhadas entre os diferentes terminais para gerar um alinhamento final. O próximo passo foi utilizar este alinhamento para gerar uma inferência filogenética bayesiana. O pacote BEAST 2 foi utilizado nesta parte do processo. Em seguida, utilizei o software Dsuite para aplicar o teste de fluxo gênico. Este teste utiliza uma variável D para comparar a frequência de compartilhamento de alelos entre dois organismos, dado um terceiro organismo externo com referência. Desta maneira, sabemos que em caso de incongruência, se a variável D não varia significativamente de zero, a incongruência é gerada pela retenção de polimorfismos ancestrais. Por outro lado, se a variável D varia significativamente de zero, temos evidência de que a incongruência é gerada por hibridização. Outros cálculos subsequentes utilizando a variável D como f4 e f-branch também foram feitos para revelar eventos de fluxo gênico onde é possível identificar a direção da troca de alelos. Finalmente, a distribuição geográfica dos terminais envolvidos nestes eventos de fluxo gênico foram comparadas para verificar a credibilidade da hipótese de hibridização.

Resultados e discussão

O número total de pares de base crus obtidos das 31 amostras foi 8.515.326. A matriz final possui 61,23% de dados faltantes e 45.651 foram alocados. Destes, 9.114 referentes a polimorfismos de nucleotídeo único. A inferência filogenética baseada em 5.400.200 das árvores geradas recuperou alta sustentação, com apenas dois nós com probabilidade posterior inferior a 0,9999 (0.971 and 0.539). Apesar disso, algumas amostras identificadas pela mesma espécie como P. juncifolia não se mostraram monofiléticas. A análise de fluxo gênico gerou 4.060 trios, dos quais, 3,871 (95.34%) mostraram valores de D significativos. O sinal para 12 eventos de fluxo gênico foram identificados, com diferentes níveis de intensidade. A distribuição dos espécimes coletados revelou que a maioria das eventos ocorreu próximo do centro de distribuição das espécies, mas também que espécimes distantes não estão isoladas reprodutivamente, compartilhando alelos com espécimes de diversos locais. Este é o caso de dois espécimes de *P. juncifolia* que apesar de coletados ao mesmo tempo e distantes do centro de distribuição, apresentam diferentes padrões de compartilhamento de alelos. Isso é interessante porque estas amostras diferem uma da outra pela cor das pétalas. Sendo uma amarela e outra laranja. Isso leva a questões sobre a possibilidade de isolamente reprodutivo.

Considerações finais

Esta investigação demonstrou que a hibridização frequente no grupo é responsável pelos padrões de compartilhamento de alelos que geraram incongruência entre os marcadores nucleares e cloroplastidiais. Essa dinâmica de fluxo gênico extenso pode indicar como as características pedomórficas se desenvolveram repetidamente no grupo, através do compartilhamento do genótipo que leva ao desenvolvimento deste fenótipo. A pressão de seleção de polinizadores também pode ter sido um fator na seleção destas características. Além disso, a falta de fluxo gênico entre populações paripátricas que diferem somente em cor, parece ser um fator crucial na evolução do gênero, e estudos posteriores devem investigar estas diferenças.

Palavras-chave: fluxo gênico; Iridaceae, RAD-seq**.**

LISTA DE ABREVIATURAS E SIGLAS

ILS - Incomplete lineage sorting

RAD - Restriction site associated DNA

ILD - Incongruence Length Difference test

GTR - General Time Reversible

pp – posterior probability

SUMÁRIO

1. INTRODUÇÃO GERAL

Dentro das árvores filogenéticas de espécies, estão inseridas árvores filogenéticas gênicas, que em vez de contar a história do conjunto de espécimes amostrados, conta a história de diversificação de um gene ou locus na linhagem. Eventos como introgressão e retenção de polimorfismos ancestrais tornam a inferência filogenética particularmente difícil para métodos filogenéticos clássicos, pois criam linhagens gênicas que podem diferir umas das outras e da verdadeira história evolutiva dos organismos (Hibbins et al., 2023). Estes eventos são comuns em clados apresentando radiação recente. Áreas com especiação *in situ* contendo várias linhagens endêmicas geralmente incluem vários destes clados. Este é o caso do Campo Rupestre, que é uma região rica em espécies, localizada no leste do Brasil. Essa região é composta de afloramentos rochosos e é habitat de muitas espécies endêmicas que mostraram alta diversificação nos últimos 5 milhões de anos (Vasconcelos et al., 2020).

Esta dificuldade em gerar filogenias robustas de clados recentemente divergentes pode ser vista em *Pseudotrimezia* (sensu Lovo et al., 2018), que é, em sua maioria, uma linhagem endêmica ao campo rupestre. Estudos moleculares iniciais demonstraram que o grupo previamente chamado *Pseudotrimezia* não era monofilético, e testes demonstraram incongruência entre marcadores cloroplastidiais (*trnG* intron, *trnH-psbA* intergenic spacer, *trnK* intron) e nuclear (nrITS). Estes incluíram o teste de incongruência por diferença de comprimento (ILD, Farris et al., 1994) e o teste Templeton (Templeton, 1983) (Lovo et al., 2012). Apesar da incongruência, foi possível indentificar quatro grupos distintos com alto suporte, nomeados informalmente Barretoi, Martinicensis, Fluminensis e Violacea. Em 2018, com a adição de dois novos marcadores plastidiais (*rpoC1* intron, *rps16* intron), Lovo et al recircunscreveram o clado Barretoi como o novo gênero *Pseudotrimezia* (ver figura complementar 1). A ausência de isolamento reprodutivo e ocorrência de hibridização e introgressão são alguns dos mecanismos que pode gerar desafios para a recuperação de uma árvore completamente bifurcante. Além disso, a diferenciação rápida e recente do grupo pode ser outra razão, refletindo espécies incipientes ainda suscetíveis à retenção de polimorfismos ancestrais (Vasconcelos et al., 2020; Lovo et al. 2021). A baixa resolução das relações entre espécies pode ser um problema do ponto de vista taxonômico, dificultando a recuperação de árvores completamente resolvidas, mas pode também informar sobre os processos evolutivos atuantes no clado.

De fato, o desafio da reconstrução filogenética é distinguir sinais de homologia mediante o ruído da homoplasia (Avise & Robinson, 2008). Congruência neste caso se refere a concordância entre árvores gênicas de diferentes locais do genoma entre si e com a história do grupo. Na presença de congruência, é possível explicar de forma parcimoniosa que as divergências desta árvore têm uma origem singular compartilhada por todos os descendentes do grupo. Em contraste, a presença de incongruência significa que diferentes árvores gênicas estão contando uma história evolutiva diferente umas das outras e/ou da verdadeira história de parentesco entre as linhagens. Esta discordância de árvores gênicas pode ser causada por hibridização ou pela presença de polimorfismos ancestrais. Este último fenômeno também é chamado de coalescência profunda, baseado na teoria de coalescência de Hudson (1990), que permite pesar a perda de caracteres contra a falha de amostragem (i.e. estimar se a amostragem é adequada e representa a variação de alelos na população). Neste caso, a incongruência é identificada porque mais de um alelo é retido após a diferenciação do grupo. Esta discordância torna-se mais proeminente quando a espessura estimada do ramo (representando o tamanho populacional) aproxima-se do comprimento do ramo (representado pela idade das linhagens) em uma determinada árvore. O outro fator que pode causar incongruência entre árvores gênicas é reticulação ou hibridização, que inclui especiação por hibridização e introgressão (Karimi et al., 2020). A especiação por hibridização acontece quando indivíduos híbridos compõem uma nova linhagem distinta de ambas linhagens parentais. Introgressão, por outro lado, é quando material genético de uma linhagem é introduzido em uma linhagem distinta através de hibridização e o híbrido gerado cruza posteriormente com uma ou ambas linhagens parentais (= retrocruzamento), sem criar uma nova população (Harrison & Larson, 2014).

Vários trabalhos foram desenvolvidos em busca de marcadores apropriados para inferir filogenias em plantas (Dong, 2012). A utilização de marcadores Sanger frequentemente só é útil para um grupo limitado de organismos e o uso de vários marcadores pode ser necessário para recuperar uma filogenia com alta resolução (Defilippis & Moore, 2000). Estas limitações, somadas com os desafios impostos por dinâmicas de distribuição de alelos como coalescência profunda e reticulação, fizeram com que métodos que recuperam múltiplos loci distribuídos pelo genoma se popularizassem. Estes métodos geram bancos de dados de loci com representação reduzida, contendo em torno de 100 pares de base, e incluem RAD-seq (DNA associado a sítios de restrição) e GBS (genotipagem através do sequenciamento) (Elworth et al., 2019). Estes métodos se mostraram eficazes em resolver relações filogenéticas

obscurecidas por processos recentes e contínuos de divergências. Isso se dá porque ao contrário de métodos clássicos que testam regiões individuais que podem não ter acumulado sinal filogenético significativo, este erro local é reduzido em métodos que amostram todo o genoma, que são capazes de recuperar um número maior de loci variáveis (Maddison & Knowles, 2006). Inicialmente, dúvidas foram levantadas sobre a capacidade deste tipo de marcador em gerar filogenias informativas, dada a grande quantidade de dados faltantes nas matrizes geradas, e sobre a robustez de métodos correntes de análise em gerar filogenias com alta resolução (Eaton et al., 2017). Desde então, métodos para análise destes dados foram desenvolvidos e testados com bons resultados, tanto para bancos de dados reais quanto para dados simulados (Rubin et al., 2012; Eaton & Overcast, 2020). Atualmente, o campo encontra-se mais maduro e estudos filogenéticos utilizando estes dados têm se tornado comuns (Acha et al., 2021; Paetzold et al., 2019; Donkpegan et al., 2020; Piwczyński et al., 2023; Su et al., 2023; Sitam et al., 2023).

A popularização destes métodos de sequenciamento acontece consecutivamente com o desenvolvimento de estratégias de investigação das dinâmicas de espécie previamente mencionadas. Testes de evidência de fluxo gênico baseados na estística D de Patterson (2012) foram desenvolvidos para testar variação em frequência de alelos entre populações, mas os mesmos princípios podem e são aplicados para testar os padrões de frequência de alelos entre espécies. A estatística D parte do pressuposto de que a taxa de variação de alelos é constante em uma população isolada reprodutivamente (D=0). Deste modo, é possível inferir se a variação de alelos é significativamente diferente da hipótese nula, indicando reticulação. Além disso, com o uso de outras estimativas, como f4 e fbranch statistics (Malinsky et al., 2018), é possível inferir o momento e a direção do compartilhamento de alelos, permitindo apontar eventos específicos de compartilhamento de alelo entre espécimes (Malinsky et al., 2021).

Com base nisso, este trabalho foi desenvolvido como dissertação de mestrado do autor e apresentado na forma de artigo científico, formatado de acordo com as normas do periódico *Molecular Phylogenetics and Evolution*, para aprofundar a investigação de Lovo et al. (2018) sobre a incongruência encontrada no grupo *Pseudotrimezia*. Um novo banco de dados de sequenciamento RAD foi gerado incluindo a maioria das espécies do grupo e uma nova filogenia foi inferida. Análises de fluxo gênico foram feitas para testar se a principal causa da incongruência observada pode ser atribuída a fluxo gênico ou à coalescência profunda.

Eventos pontuais de reticulação foram investigados em relação a sua distribuição geográfica para gerar hipóteses sobre a natureza dos eventos encontrados.

2. ARTIGO CIENTÍFICO

NEW GENETIC MARKERS AND GENE FLOW ANALYSIS SHED LIGHT INTO *PSEUDOTRIMEZIA* INCONGRUENCE

2.1 INTRODUCTION

The inference of phylogenetic relationships within lineages undergoing recent speciation are made particularly difficult by ongoing species dynamics such as incomplete lineage sorting and introgression. Those present a challenge to classic phylogenetic methods because of species-gene trees discordance (Hibbins et al., 2023). Areas with high rates of *in situ* speciation and several endemic lineages usually include several examples of such lineages. This is the case for the *campo rupestre*, which is a species rich region in eastern Brazil (Vasconcelos et al., 2020). It is a habitat composed of rocky outcrops and home to many endemic species, which have shown faster diversification rates in the last 5 Myr. (Vasconcelos et al., 2020). These complications in inferring the detailed phylogeny of recently diverged clades can be seen in the genus *Pseudotrimezia* (*sensu* Lovo et al., 2018), a clade within the tribe Trimezieae (Iridaceae), which is mostly endemic to the *campo rupestre*. A previous study showed evidence for incongruence between phylogenetic inferences from nuclear and chloroplast markers based on Incongruence Length Difference test (ILD, Farris et al., 1994) and Templeton test (Templeton, 1983) (see Lovo et al., 2012). Regardless of this incongruence, the group has been consistently recovered as monophyletic, being recircumscribed as *Pseudotrimezia* genus by Lovo et al. (2018). The unsolved tree topology and the short branches of the new circumscribed *Pseudotrimezia* (Lovo et al., 2012; 2018) might result from the fast and recent differentiation of the group, reflecting incipient species still prone to incomplete lineage sorting (Vasconcelos et al., 2020; Lovo et al. 2021). In addition, lack of reproductive isolation and the occurrence of hybridization and introgression among endemic and co-occurring species may also pose a challenge to recovering a totally bifurcating phylogenetic tree (Lovo et al., 2018). As the authors themselves pointed out, the limited species relationships resolution may be a problem from a taxonomic standpoint, but may also provide important insights into the evolutionary process that shaped its diversity.

The challenge of phylogenetic reconstruction is to distinguish the phylogenetic noise of homoplasy from the signal of homology (Avise & Robinson, 2008). Congruent trait patterns can be parsimoniously explained as having a single origin in some ancestral taxon

shared by all descendants. In contrast, incongruent trait patterns may be produced by discordant gene trees and ancient polymorphisms. This latter phenomenon is called deep coalescence, after Hudson's (1990) coalescent theory, which helps to weigh trait loss against failure to sample. This discordance becomes more prominent when the estimated width of the branches (representing population size) approaches the length of the branches (i.e., lineages ages) of a given tree (Maddison, 1997). The term hemiplasy was coined to describe this incomplete lineage sorting (ILS) or deep coalescence, that can lead to gene trees and organismal trees discordance (Avise & Terrence., 2008). Moreover, gene tree incongruence in plants are also largely caused by the widespread phenomenon of reticulation, which includes hybrid speciation and introgression (Karimi et al, 2020). Hybrid speciation happens when a new species is composed by hybrid individuals forming a separate distinct lineage from either of its parental populations. Introgression, on the other hand, involves the integration of genetic material into the genome of a population through interbreeding and backcrossing without creating a new population (Harrison & Larson, 2014).

A lot of work has been done in finding genetic markers appropriate for inferring phylogenies in plants (Dong, 2012). They need to be both variable enough to inform that there is in fact a cladistic difference accumulated in the genetic regions analysed at the same time, as well as do not vary so much as being prone to high rates of homoplasy preventing to track homologies along the object taxa. Apart from that, the choice must also take into consideration the nature of the information, which may depend on the biological question and the practical criteria that refers to lab equipment and resources available (Chenuil, 2006). Although most phylogenetic inferences so far used Sanger markers for solving relationships, they are often only useful for a limited group of organisms and may require several markers to achieve inferences with high resolution (Defilippis & Moore, 2000). Because of the challenges posed by gene coalescence and reticulation, genome wide multi-locus methods of reduced representation have become popular for phylogenetic history inference. Those reduced representation data sets can be created by methods such as restriction site associated DNA sequencing (RAD-seq) markers and genotyping by sequencing (GBS), which retrieve loci as roughly 100 base pairs with palindromic restriction enzyme recognition sites (Elworth et al., 2019). These high-throughput sequencing (HTS) technologies, also called NGS (e.g., next-generation sequencing), introduced more generalistic genetic markers which allow for sampling a larger proportion of the genome, resulting in a larger amount of specimen data sampling and coverage. Initially, RAD-seq markers were used to explore intra specific

relationships between organisms and Quantitative Trait Locus analysis to identify phenotypic-genotypic variation within populations, since it was able to find variations spread through several regions of the genetic material (Emerson et al., 2010; Baxter et al., 2011; Pfender et al., 2011). This characteristic also showed potential to solve another type of question, the one of the relationships among different species (Eaton & Ree, 2013; Herrera & Shank, 2015). The benefits in comparison to the Sanger markers is that its generalistic nature could explore several different regions of the genome, without the need for the trial and error process of finding specific markers for each species group, even without an existing reference genome (Eaton and Ree, 2013). That makes it a much more cost effective and time saving strategy.

RAD-seq markers could be especially useful in cases of recent or continuing divergence among lineages, since one single region encompassed by a Sanger marker may not have accumulated enough phylogenetic signal at the lowest taxonomic levels for an accurate inference (Maddison & Knowles, 2006). Also, by sampling the whole genome, local gene topology bias is reduced. Initially, the very generalistic aspect of RAD-sequencing was seen as a problem, since it was not possible to be completely sure of what regions generated the short reads resulting. This created doubts and following investigations about the efficacy of current assembly and phylogenetic inference methods of accounting for this data (Eaton et al., 2017). Since then, novel assembly methods were developed to unite those several short reads into reliable alignments with techniques that weighted and filtered those reads by quality, clustered them and reached consensus based on the relationships among reads, taxa or even reference genomes (Eaton & Overcast, 2020). Alongside those, it was also necessary to develop phylogenetic inference techniques that accounted for the large proportion of missing data and were able to deal with the much larger matrixes generated using reasonable computing resources (Escudero et al., 2020). As of the time of this research, the field has matured and the efficacy of this kind of analysis has been tested with success for artificially created data sets (Rubin et al., 2012), as also applied to empirical phylogenetic investigations (Acha et al., 2021; Paetzold et al., 2019; Donkpegan et al., 2020; Piwczyński et al., 2023; Su et al., 2023; Sitam et al., 2023).

Beyond allowing for the recovery of phylogenies with high support, the choice of RAD markers is also compatible with model-based clustering methods of testing for allele frequency variation such as STRUCTURE (Pritchard et al., 2000) and ADMIXTURE (Alexander et al., 2009). These were initially designed to infer admixture among populations, but the same principles can be applied to test hybridization and introgression among closely

related species. This is true as long as common population assumptions hold (Malinsky, 2021). In this sense, D statistics and its related f4-ratio statistics are commonly used to assess evidence of gene flow (Patterson, 2012). D statistics or ABBA BABA test is a parsimony like method that uses allele frequency to find parsimony informative sites supporting a different phylogeny from the species tree. In practice, the calculation of the D statistic is done by using a group of four taxa and an established phylogeny to detect gene flow between two ingroups that are not the closest-related species. Because of some shortcomings of D statistics such as being a poor estimator for the amount of and direction of introgression, f4-ratio statistics have been designed to investigate signatures of introgression in genomic windows along chromosomes (Malinsky et al., 2021). To help disentangle f4-ratio results and assign gene flow evidence to specific branches, the f-branch metric was introduced (Malinsky et al., 2018). This can be used to directly test gene flow hypotheses.

Here, we deepen the investigation on *Pseudotrimezia* (sensu Lovo et al. 2018) by expanding Lovo et al. (2012, 2018) sampling through the use of RAD markers and use the resulting phylogeny to directly probe the reasons of gene trees incongruence found by Lovo using D, f4 and f-branch statistics. In particular, we intend to disentangle the contribution of incomplete lineage sorting, due to the recent and fast speciation, and incipient reproductive isolation, causing introgression, in the incongruence found among nuclear and chloroplast genomes. Then, we compare the patterns of gene flow inferred with the geographical distribution of the samples to assess whether the assumptions of introgression are plausible given their distributions. Moreover, other aspects of the species biology (pollination systems, phenology) were also considered in order to infer the most plausible evolutionary scenario.

2.2. MATERIAL AND METHODS

2.2.1 STUDY GROUP

The tribe Trimeziae is endemic to the Neotropics and is currently composed of five genera: *Deluciris* A.Gil & Lovo, *Pseudiris* Chukr & A.Gil, *Pseudotrimezia* R.C.Foster, *Trimezia* Salisb. ex Herb. and *Neomarica* Sprague. The tribe Trimezieae as a whole is a great example of a clade that highly benefited from molecular analysis, having its relationships clarified by studies that used Sanger markers to uncover false assumptions of homology based on morphology alone. This group was strongly supported as monophyletic in a study by Lovo et al. (2012), although these authors also found that its historically recognized genera were not all monophyletic. These included *Pseudotrimezia* and *Trimezia*. Lovo et al. (2012) identified four distinct clades within the tribe and informally named them Barretoi, Martinicensis, Fluminensis, and Violacea. The group named Barretoi needed special attention for its low internal support and for including members of *Pseudotrimezia* and *Trimezia*, being described as an especially difficult one to solve. In Lovo et al. (2018), the authors performed further analyses with a higher number of Sanger markers from nuclear and chloroplast genomes of Trimezieae. These analyses corroborated the revision of the generic organization of this tribe. The clade Barretoi showed consistently low internal support in both chloroplast (trnG intron, trnH-psbA intergenic spacer, trnK intron) and nrITS datasets, even after the addition of two chloroplast markers (rpoC1 intron, rps16 intron) (Lovo et at., 2018). However, it remained been recovered as monophyletic, and the authors proposed it should now constitute a recircumscribed monophyletic *Pseudotrimezia* (*sensu* Lovo et al. 2018). Therefore, for the purposes of this study, we follow the revised classification of tribe Trimezieae proposed by Lovo et al. 2018 (see supplementary material 2 and 3).

Pseudotrimezia species are almost entirely endemic to the *campos rupestres* in Diamantina Plateau District, Southern Espinhaço Province (Colli-Silva et al., 2019) with two exceptions: *P. juncifolia* (Klatt) Lovo & A.Gil and *P. pusilla* (Ravenna) A.Gil, with both occurring in Cerrado vegetation. *Pseudotrimezia juncifolia* is more widely distributed and can be found in central-western, southern and southeastern Brazil, while *P. pusilla* is endemic to Cerrado in Goias state, central Brazil. This latter is not represented in our sampling. The tribe Trimezieae is often characterized by conspicuous, mostly colorful flowers, displaying sepals and petals with contrasting color patterns and glandular trichomes, which are common traits

in Iridaceae (Hamlin et al., 2017). In *Pseudotrimezia*, although some species display ornamented iridaceous flowers, most species show contrasting plain yellow flowers (Lovo et al., 2021). In a study into the floral ontogeny of Trimezieae, lineages within *Pseudotrimezia* were shown to have developed from ornamented to plain flowers several times. The research debated the hypothesis that heterochronic events allowed some *Pseudotrimezia* lineages to maintain a paedomorphic state in the mature flower. It also showed that this repeated evolution to simple plain yellow flower leads to an increase in the specialization regarding their pollination system, which was supported by their probe into flower visitors and their respective behaviour (Lovo et al., 2021).

2.2.2 SAMPLING

Sampling was focused on the genus *Pseudotrimezia*, aiming to include as many species as possible. Moreover, we also included multiple individuals from species that failed in presenting reciprocal monophyly in previous investigations (e.g., *P. cipoana*, *P. diamantinensis*, *P. juncifolia*, *P. cathartica*, Lovo et al., 2012, 2018). We included a total of 31 accessions, representing 16 from the 24 species currently accepted in *Pseudotrimezia* (66,6%) and one outgroup from the genus *Trimezia* (Table 1).

TABLE 1. Sampled specimens. *P.* = *Pseudotrimezia, T.* = *Trimezia, aff.* = affinis, sp= unidentified species, collect ID= field collecting number, Alt. = elevation (meters), S=south, W=west.

Species	Author	collect. ID	Alt.	Latitude	Longitude
P. barretoi	R.C.Foster	Mello-Silva 2680	1417	S18º 20' 35"	W 43º 40' 51"
P. cathartica	(Klatt) Ravenna	Lovo 269	1325	S18º 14' 5.9"	W 43º 41' 5.8"
P. cathartica	(Klatt) Ravenna	Lovo 433	1169	S18º 9' 24.2"	W 43º 43' 45.5"
P. cipoana	Ravenna	Lovo 75	1216	S19º 17' 25.7"	W 43º 33' 47.04"
P. cipoana	Ravenna	Lovo 246	1083	S18º 28' 38.4"	W 43º 27' 28.6"
P. cipoana	Ravenna	Lovo 374	1228	S18º 27' 19.1"	W 43º 26' 56.4"
P. cipoana	Ravenna	Mello-Silva 2743	1332	S16º 51' 21"	W 43º 2' 0"
P. diamantinensis	Ravenna	Lovo173	1047	S18º 17' 36.8"	W 43º 44' 8.01"
P. diamantinensis	Ravenna	Mello-Silva 2674	1192	S18º 34' 19"	W 43º 35' 33"
P. elegans	Ravenna	Lovo144	1515	S18º 24' 3.6"	W 43º 19' 13.2"
P. fulva	Ravenna	Lovo117	1089	S18º 11' 32.2"	W 43º 37' 37"
P. fulva	Ravenna	Lovo376	1170	S18º 9' 24.9"	W 43º 43' 36.1"
P. gracilis	Chukr	Lovo139	1063	S19º 21' 52.8"	W 43º 34' 28.9"
P. juncifolia	(Klatt) Lovo & A.Gil	Lovo36	1176	S18º 51' 34"	W 43º 45' 27"
P. juncifolia	(Klatt) Lovo & A.Gil	Lovo216	1302	S19º 13' 8.4"	W 43º 29' 52.3"
P. juncifolia	(Klatt) Lovo & A.Gil	Lovo 424	1270	S18º 25' 47.6"	W 43º 40' 45.5"
P. juncifolia	(Klatt) Lovo & A.Gil	Mello-Silva 3443	1180	S20º 17' 52.3"	W 46º 31' 23.5"
P. juncifolia	(Klatt) Lovo & A.Gil	Mello-Silva 3444	1218	S20º 17' 46.2"	W 46º 31' 14.7"
P. laevis	Ravenna	Lovo 123	1220	S18º 17' 41.9"	W 43º 50' 44.5"
P. nana	Lovo & Mello-Silva	Lovo 418	1130	S18º 36' 1.9"	W 43º 52' 55.5"
P. pauloi	Chukr	Lovo 152	1097	S18º 11' 0.1"	W 43º 37' 4.1"
P. planifolia	Ravenna	Lovo 147	1316	S18º 17' 33.8"	W 43º 44' 6.8"
P. recurvata	Ravenna	Lovo 171	1431	S18º 15' 32.4"	W 43º 42' 43.9"
P. recurvata	Ravenna	Lovo 375	1228	S18º 27' 19.1"	W 43º 26' 56.4"
P. striata	Lovo & Mello-Silva	Lovo 297	1119	S18º 18' 36.1"	W 43º 55' 6.7"
P. sublateralis	Ravenna	Lovo 189	1455	S18º 16' 9"	W 43º 42' 16.1"
P. truncata	(Ravenna) Lovo & A.Gil	Lovo 254	1538	S18º 24' 2.8"	W 43º 19' 11.3"
Pseudotrimezia sp1		Lovo 280	1047	S18º 9' 19.1"	W 43º 41' 39.9"
Pseudotrimezia sp2		Lovo 360	1184	S18º 34' 21.5"	W 43º 35' 32"
Pseudotrimezia sp3		Lovo 373	1228	S18º 27' 19.1"	W 43º 26' 56.4"
T. aff. chimantensis	Steyerm	Lovo 409	1603	S13º 31' 23.7"	W 41º 57' 29.1"

2.2.3 DNA EXTRACTION AND SEQUENCING

The DNA extractions were made with the CTAB method from silica gel dried leaf tissue (Doyle and Doyle, 1990) in the Pritzker Laboratory in Chicago. Quantity and concentration of DNA were assessed using the Qubit and Quant-it assay system (Invitrogen), with a target concentration of 20-40 ng/µl. The resulting DNA was normalized to a concentration of 30ng/µl and sent to the Floragenex company (www.floragenex.com) for the NGS process. The method utilized was the RAD-seq of the single end type, where several

samples are read on the same plate, and the sequences were obtained through the clivage of the DNA samples with the restriction site enzyme Pst1 (Walder et al., 1984), anchored to specific barcodes to each sample, and submitted to a multiplex run in a single strand of the Illumina GAIIx sequencer for 75 cycles (generating single-end sequences – hereforth referred to as reads). Output data was in an Illumina FASTQ format, which was processed for downstream analyses.

2.2.4 ASSEMBLY AND ALIGNMENT

The raw sequence reads were demultiplexed and de novo assembled, since there are no reference genomes for species of Trimezieae. Data demultiplex and for low quality reads, adapters filtering was performed in ipyrad v. 0.9.77 (Eaton, 2014). Assembly was also performed in ipyrad v. 0.9.77 (Eaton, 2014) was used to demultiplex the data and filter for low quality reads, adapters and primers. In the next step ipyrad clustered the reads within samples and then identified and removed duplicate reads by using the ipyrad implementation of MUSCLE (Edgar, 2004). Then, the sequencing error rate was calculated based on this within-sample alignment. At this point, a single consensus sequence was kept per sample.

Finally, homologous loci were aligned across samples in a MUSCLE alignment (Edgar, 2004). Default parameters were kept for minimum depth for statistical base calling (6 reads) and minimum depth for majority-rule base calling (6 reads). The maximum number of low quality bases per read was set to five and clustering threshold similarity to 85%. Filtering for primers/adapters was set to its strictest setting (2) and zero barcodes mismatches were set to ensure reads were assigned to their proper specimen barcode. The maximum number of alleles allowed per site in the consensus sequences was set to two. The minimum number of samples allowed per locus was set to four out of the total thirty-one. The maximum number of uncalled bases (N) and heterozygotes in consensus sequences was left at default (0.05). The maximum allowed percentage of SNPs per locus was set at 20% and the maximum number of indels per locus was set to eight, as is default and recommended in the software documentation (Eaton and Overcast, 2020).

2.2.5 PHYLOGENETIC INFERENCE

We used the software package Beast2 (Bayesian Evolutionary Analysis by Sampling Trees 2 – Bouckaert et al, 2019) to recover the phylogenetic relationship among *Pseudotrimezia* species based on a Bayesian inference. We used a relaxed clock model as implemented on Beast2 because assuming that the rate of mutation is constant for the group is unrealistic, and the relaxed clock model attempts to account for this variation (Lepage et al., 2007). The nucleotide substitution model was the GTR+I+Gamma, given it was the model available with the least amount of assumptions and has been shown to be common in such analyses which is robust to over-parametrization (Fabreti and Höhna, 2022). Three separate runs with chain length of 20 million generations were carried out, sampled every 1000 trees, with the initial 10% of each run removed as burn in. The results were compiled using the BEAST package Logcombiner and analysed using the BEAST package Tracer (Rambaut et al., 2018). A single tree was summarized to represent the posterior distribution generated with the BEAST2 package TreeAnnotator (Rambaut et al., 2018).

2.2.6 TESTING FOR INGRESSION USING D, F AND F-BRANCH STATISTICS

D statistic or ABBA BABA test is a parsimony like method that uses allele frequency to find parsimony informative sites supporting a different phylogeny from the species tree (Patterson, 2012). In practice, the calculation of the D ratio is done by using groups of four taxa, three specimen samples from the group being studied (P1, P2 and P3) and an established Outgroup (O). Thousands of tree topologies are inferred for each portion of the dataset information of these groups. The name ABBA BABA origins from this variation in local topology. This variation is quantified and compared to the average expected allele variation along the whole tree. In the null hypothesis, D is zero; the phylogenetic signal for gene flow is stronger as this variable diverges from zero (Zheng and Janke, 2018). A significant D value indicates two non-sister species/samples share more alleles to each other than expected when compared to phylogenetically closely related samples (same or sister species). In incongruent datasets where the D value deviation from zero is statistically significant, the primary reason for this incongruence is hybridism and introgression. In contrast, when the D value does not deviate from zero, the primary reason for incongruence is ILS (Aardema et al., 2023).

The software dtrios from the Dsuite package (Malinsky et al., 2021) was used to create trios of specimen to be compared against the outgroup. Its input was the tree topology and the resulting assembly with read depth information. It uses a jackknife method to estimate the

standard error deviation for each trio and compare it to a normal distribution (null hypothesis). From it, D, Z-score, p and F4-ratio values are calculated. The Z-score is the D variable divided by its standard error and its significance value is p (Zheng & Janke, 2018). This significance was used to rule out f4 values for trios were p value was > 0.01 as is standard. This is done because f4 ratio calculations for trios nearly or equaly related can produce large but insignificant values even in the absence of gene flow (Malinsky et al., 2021). D values were evaluated in order to rule out ILS as the primary cause of gene tree incongruence.

Results from the previous step, plus a species list with a designated outgroup were input into the fbranch software from the same package and f-branch ratios were calculated. This is done because the number of possible gene flow donor-recipient combinations increase rapidly with the number of species sampled. The fbranch metric is thus used to assign gene flow to specific nodes and branches of a given tree. Later, the software dtools from the same package was used to plot the f-branch ratios into a heatmap, where darker red colored squares represent gene flow events between the specimens or nodes and each terminal specimen (Malinsky et al., 2021).

2.3. RESULTS

2.3.1 SEQUENCE ASSEMBLY AND ALIGNMENT

The total number of base pairs obtained from all raw reads among the 31 samples (i.e., specimens) was 8,515,326. The highest number and lowest number or raw reads (755,909-53,580) coincided with the highest and lowest numbers of loci recovered for each sample (2,473-411). The size of the final matrix was 31,319,569 and 61.23% of those were missing sites. Read alignment matrix resulted in 3,768 columns in length. Out of the 45,651 base pairs allocated to those loci, 33,389 were variable and 9,114 contained parsimony informative data (Supplementary table 1). The mean sequencing error rate was 0.203%, ranging from 0.182% to 0.248%. The mean heterozygosis error rate was 0.558%, ranging from 1.114% to 0.440%. There were 405 loci represented by the minimum amount of samples (4), 5 loci represented in 26 of the samples and no loci was present in all samples. The number of base pairs recovered in the assembly for each terminal is listed in table 2.

TABLE 2. Total number of nucleotides recovered by RAD-sequencing before processing (raw reads), followed by the total number of nucleotides allocated in the final assembly for each terminal (loci in assembly). *P. = Pseudotrimezia, T. = Trimezia*.

Specimen	raw reads	loci in assembly
P. barretoi MS2680	112592	1097
P. cathartica Lovo269	306092	1922
P. cathartica Lovo433	178126	1585
P. cipoana Lovo75	66579	525
P. cipoana Lovo246	144651	1265
P. cipoana Lovo374	481131	2152
P. cipoana MS2743	262529	1786
P. diamantinensis Lovo173	211478	1618
P. diamantinensis MS2674	237367	716
P. elegans Lovo144	124595	931
P. fulva Lovo117	113692	1051
P. fulva Lovo376	276787	1658
P. gracilis Lovo139	414306	2141
P. juncifolia Lovo36	214251	1555
P. juncifolia Lovo216	130599	782
P. juncifolia Lovo424	755909	2473
P. juncifolia MS3443	353964	2119
P. juncifolia MS3444	142741	1201
P. laevis Lovo123	110827	1057
P. nana Lovo418	512237	1985
P. pauloi Lovo152	99416	863
P. planifolia Lovo147	240164	1847
P. recurvata Lovo171	53580	411
P. recurvata Lovo375	595755	2214
P. striata Lovo297	241883	1742
P. sublateralis Lovo189	98182	984
P. truncata Lovo254	130428	1152
Pseudotrimezia sp1 Lovo280	533780	1885
Pseudotrimezia sp2 Lovo360	712155	2353
Pseudotrimezia sp3 Lovo373	209799	1598
T. aff. chimantensis Lovo409	449731	987

2.3.2 PHYLOGENETIC INFERENCE

The final Bayesian inference was based on the distribution of posterior probabilities over the 5,400,200 trees remaining after burn in removal. Nodal posterior probabilities where high, e.g., only two nodes with a posterior probability (pp) under 0.999 (0.971 and 0.539),

recovering a summarized tree with high overall support. Effective sample size for pp and likelihood were 399 and 1817 respectively.

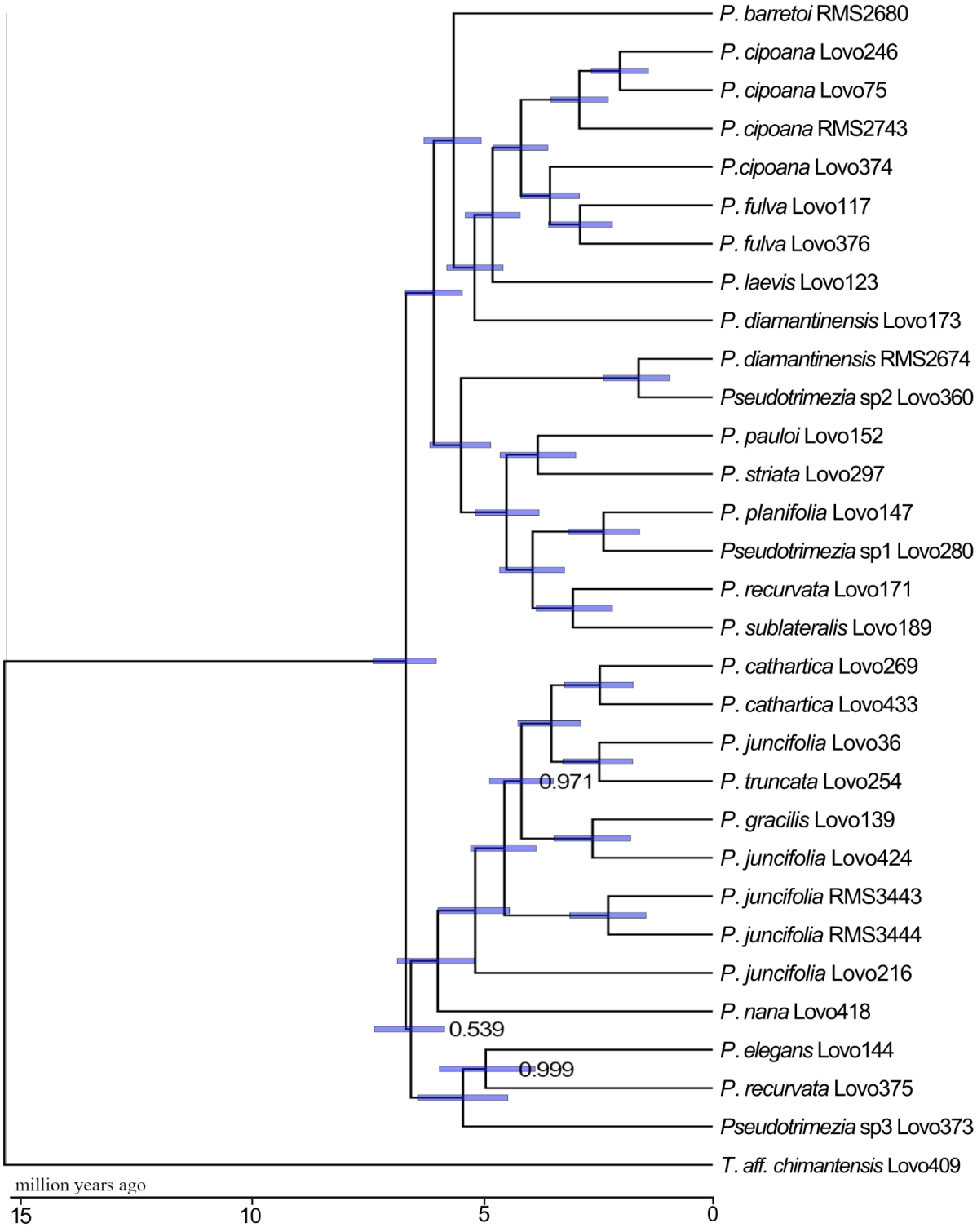


FIGURE 1. Phylogenetic relationships in *Pseudotrimezia* based on RAD-seq molecular data. The tree displays the highest pp recovered from the Bayesian distribution of 5,400,200 trees after burn in. Node labels indicate the specific pp at the nodes, when differing from 1 - unlabelled nodes were recovered with pp = 1. Blue bars indicate the margin of error for the divergence of each node.

2.3.3 GENE FLOW TEST

A total of 4,060 trios were created, each containing 1,668 variant base pairs, and their D and f4-ratio values were calculated. Out of those, 3,871 (95.34%) showed D values higher than 0.01 and 325 trios (8%) were discarded because there wasn't enough informative variation. A significant Z-score of 3 or higher was found for 95 trios (2.3%). Four trios resulted in a D value of 1 (Z-score=infinite), representing complete allele sharing. D value variation along the tree was plotted for each pair which showed allele deviation from its trio and outgroup.

The resulting Z-score, p-value and f4-ratio were summarized into a fbranch matrix containing 12 occurrences of significant gene flow between specimens ($P < 0.01$). This fbranch matrix was then plotted into a heatmap where each square represents a possible gene flow event. Squares were significant fbranch values were found are colored. Darker colors represent a higher fbranch value.

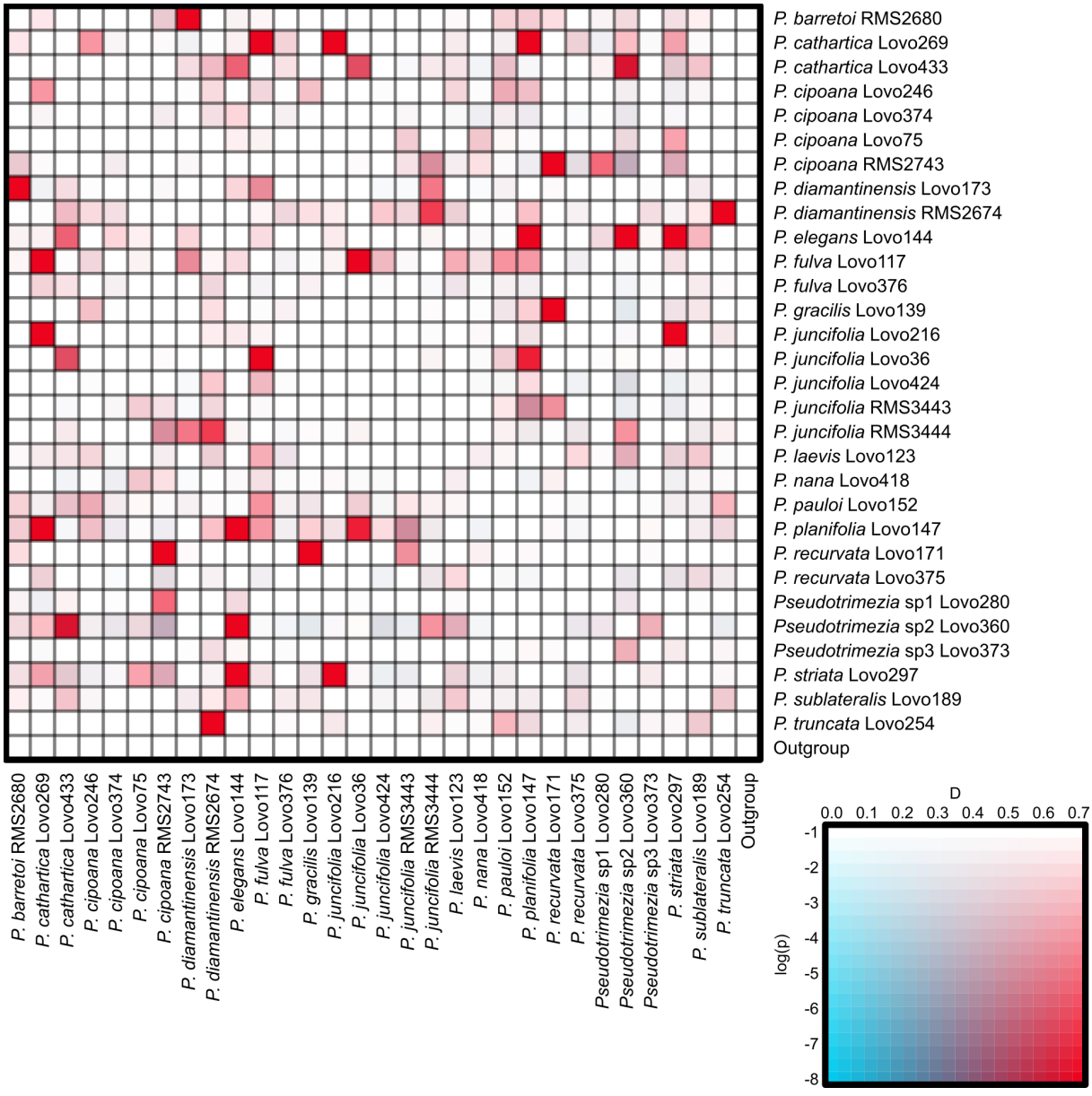


FIGURE 2. Heatmap for species P2 and P3, in which the color in each cell corresponds to the most significant D-statistic found between these two species, across all possible species in P1.

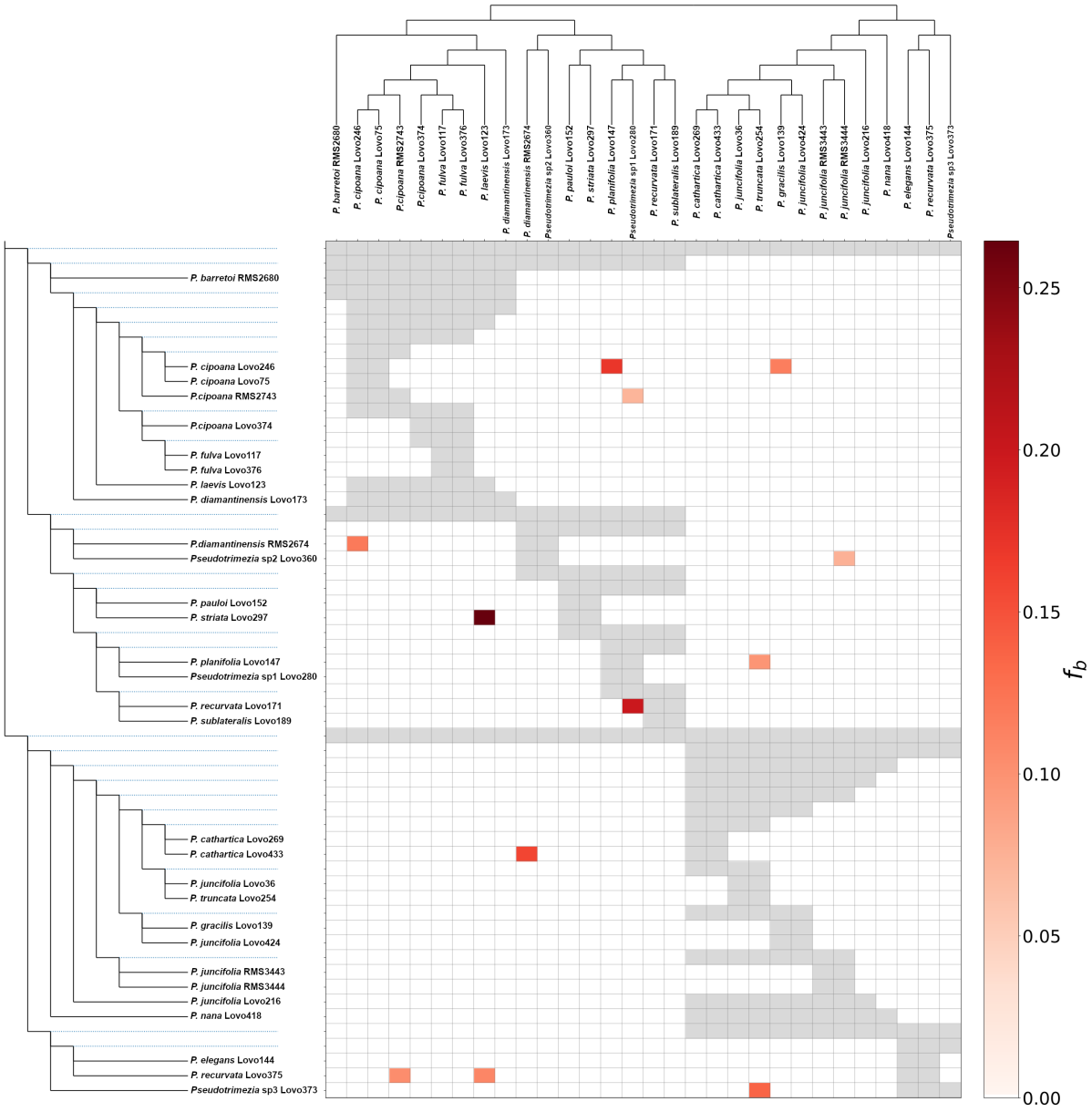


FIGURE 3. F-branch results. The values in the matrix refer to excess allele sharing (fb statistics) between the tree branches and nodes in the y axis and the specimen on the x axis. Dotted lines in the phylogeny represent ancestral lineage, and non-dotted lines represent extant lineages. The colored squares indicate significant results ($P < 0.01$). Gray cells are inestimable relationships.

TABLE 3. Statistically significant gene flow events characterized by donor, recipient and f-branch values daviation from zero, Minimum value being 0 representing no signal of gene flow and 1 total allele sharing).

Allele donor	Allele recipient	f-branch value
P. striata Lovo297	*P. laevis* Lovo123	0.2644
P. recurvata Lovo171	*Pseudotrimezia* sp1 Lovo280	0.1983
P. cipoana Lovo246	*P. planifolia* Lovo147	0.1698
P. cathartica Lovo433	*P. diamantinensis* Mello-Silva2674	0.1580
Pseudotrimezia sp3 Lovo373	*P. truncata* Lovo254	0.1366
P. diamantinensis Mello-Silva2674	*P. cipoana* Lovo246	0.1190
P. cipoana Lovo246	*P. gracilis* Lovo139	0.1166
P. recurvata Lovo375	*P. laevis* Lovo123	0.1100
P. recurvata Lovo375	*P. cipoana* Mello-Silva2743	0.1035
P. planifolia Lovo147	*P. truncata* Lovo254	0.0975
Pseudotrimezia sp2 Lovo360	*P. juncifolia* Mello-Silva3444	0.0752
P. cipoana Mello-Silva2743	*Pseudotrimezia* sp1 Lovo280	0.0718

2.4 DISCUSSION

Our results greatly improve the phylogenetic relationships among *Pseudotrimezia* species. More importantly, it indicates that the previous incongruencies between nuclear and plastidial genomes are mainly caused by hybridization. The previous phylogenetic hypotheses available for *Pseudotrimezia* were based on separated trees recovered from both plastid and nrITS matrices (Lovo et al., 2018). In both instances, some nodes posterior probabilities were high, but there were recalcitrant polytomies, especially in early divergent nodes (see supplementary material 2 and 3). Our inference however, recovered a pp of 1 for most nodes except 3. One node showed 0.971 pp, and includes two specimen of *P. juncifolia* (Lovo 36 and Lovo 424) which failed to group with each other and with *P. juncifolia* (Mello-Silva 3443 and Mello-Silva 3444). Out of those *P. juncifolia* specimen, only one (Mello-Silva 3444) was

involved in the gene flow events inferred, being an allele donor to *Pseudotrimezia* sp2. Another early divergent node, which included the 0.971 pp node, showed low support (0.539). Within this lineage, 4 instances of allele reception and other 4 instances of allele donation where found between terminal species. Also, high D values were present in this group, as they were across the dataset. This may evidence that, although RAD datasets may yield higher pp for incongruent datasets, species dynamics such as reticulation still allude phylogenetic reconstruction when not taken into account.

Our positive results are mirrored by other studies based on RAD markers in plant species. Acha et al. (2021) worked with a larger dataset of 206 species of *Passiflora*, using RAD-2b markers and Maximum likelihood inference. Their final alignment included 424,008 base pairs and missing information varied from 62 to 92% among samples, reflecting similar size and amount of missing data of the final dataset recovered here, i.e., considering their sampling is one order of magnitude larger than ours. The study recovered overall lower support than recovered for *Pseudotrimezia*, with many nodes below 50%, but other well defined branches with high support. Efforts to filter out sites with missing data recovered lower support than when inference included the full dataset. Regardless of a low support for several nodes, the approach allowed the recovery of the largest most well resolved phylogeny of Neotropical vines to date (Acha et al. 2021). Another similar study included 101 samples (41 species + outgroup) of Hawaiian *Melicope* (Rutaceae) (Paetzold et al., 2019). The authors compiled five distinct datasets by varying thresholds for the minimum number of samples per locus, recovering 36,622 loci and 385,871 SNPs across the final alignment in the main analysis. These numbers are much higher than ours, considering our dataset included a third of the number of entries. Their inference method involved a concatenated approach including Bayesian, Maximum likelihood and SVD quartets, which revealed incongruences among datasets and analyses, as well as clades with low overall support. D statistics test showed positive signal for ancient introgression among clades, corroborating the hypothesis of two events of colonization of the Marquesas Islands by *Melicope*, and being used as basis for a new taxonomic treatment (Paetzold et al., 2019). In addition, Rad sequencing data on 38 assessments of *Prunus* (Rosaceae) was also used successfully to probe into its species dynamics. Like in this study, they were able to infer a phylogeny with high support and found significant evidence of introgression, but differently from us, they also found faint evidence for ancient introgression (Su et al., 2023).

Most specimen sampled in this study concentrate in the southernmost part of the Espinhaço Mountains, mostly on the east portion of the mountain range. When the geographical distribution of the specimens is compared to the level of gene flow found between the paired samples (e.g., Fig. 3), some specific patterns are worth noting. For instance, out of four samples for the species *P. cipoana*, the only one showing 2 significant instances of allele reception is *P. cipoana* Lovo 246, which is in the center of the species' distribution (Fig. 4). The strongest signal of introgression in *P. cipoana* Lovo 246 refers to the reception from the closer located specimen *P. planifolia* and the weakest to the further located *P. gracilis.* The northernmost specimen of *P. cipoana* (Mello-Silva 2743) was identified as an allele recipient of *Pseudotrimezia* sp1 (Lovo 280), which is nearer to it than most other specimens of *P. cipoana* (Fig. 4). *P. diamantinensis* Mello-Silva 2674 received alleles from *P. cipoana* Lovo 246 (Fig. 3), which are both around the center of the specimen distribution (Fig. 4). The same is evident for the following recipient-donor pairs: *Pseudotrimezia* sp3 Lovo 373 < *P. truncata* Lovo 254, *P. cathartica* Lovo 433 < *P. diamantinensis* Mello-Silva 2674, *P. planifolia* Lovo 147 < *P. truncata* Lovo 254, and *P. striata* Lovo 297 < *P. laevis* Lovo 123, which was the event detected with the strongest signal (Fig. 3). *P. recurvata* Lovo 375 also showed signal of allele reception from two specimen, being *P. laevis* Lovo 123, which is in the center of distribution, and *P. cipoana* Mello-Silva 2743, which is further north. *Pseudotrimezia* sp2 Lovo 360 received alleles from *P. juncifolia* Mello-Silva 3444, which was sampled at the southwestern limit of distribution, outside the Espinhaço mountain range (Fig. 4). Interestingly, the same signal was not found between the other instance of *P. juncifolia* (Mello-Silva 3443) collected in the same location as *P. juncifolia* Mello-Silva 3444, indicating that the allele being shared by one, is not shared by the other.

The case of the two *Pseudotrimezia juncifolia* specimens from the limit of our geographical sampled range is noteworthy because the populations are located side by side in the Serra da Canastra National Park, mirroring a parapatric distribution with no detectable barriers in between. The only reason why both specimens were sampled as representing distinct populations sources is the conspicuous color difference, with *P. juncifolia* Mello-Silva 3443 population displaying deep orange to red tepals; while *P. juncifolia* Mello-Silva 3444 displays totally yellow tepals. There were no visible individuals displaying intermediate color patterns between the deep orange and the clear yellow characterizing the contiguous populations (R. Mello-Silva & S. Alcantara, fieldwork observation). A closer look at the Dstatistics heatmap (Fig. 2) reveals that both specimens did not display signals of reciprocal

gene flow. Moreover, both instances display significant and distinct levels of gene flow (e.g., D value) with other specimens representing distinct species in the clade. It indicates that, even though those are the most distant specimens from the genus distribution center, they are far from being isolated reproductively from the other species. In particular, *P. juncifolia* Mello-Silva 3444 shows a low (but significant) gene flow signal with *P. juncifolia* Lovo 36, which is located in the central portion of the Espinhaço range and also displays vibrant yellow flowers (as all the other populations of this species except *P. juncifolia* Mello-Silva 3443, J. Lovo, personal communication). Moreover, the absence of detectable gene flow between the two parapatric *P. juncifolia* with orange/red versus yellow flowers might be suggestive of a reproductive isolation due to pollinator shift caused by change in flower colour (see Schemske and Bradshaw, 1999). Such changes have been shown to occur by the substitution of a single allele substitution in a large effect gene in *Mimulus* (Bradshaw and Schemske, 2003). In that case, however, additional morphological changes were also associated with the flowers of distinct colours, with distinct species (*M. lewisii* and *M. cardinalis*) displaying distinct pollinator guilds (bumblebees and hummingbirds, respectively; Bradshaw and Schemske, 2003). *P. juncifolia* Mello-Silva 3443 and *P. juncifolia* Mello-Silva 3444 did not display other morphological differences but flower colors, and further investigations should focus on the elucidation whether the two populations are indeed representatives of *P. juncifolia* (Klatt) Lovo & A.Gil. An alternative scenario might indicate the population identified as *P. juncifolia* Mello-Silva 3443 is reproductively, ecologically and/or physiologically distinct from *P. juncifolia* and then it should be described as a new species. Such parapatric pattern of speciation is well known for plants occurring in mining or heavy metals contaminated soils, with reproductive reinforcement taking place subsequentially to the evolution of soil specialization (e.g., *Agrotis*, Gregory and Bradshaw, 1965). As mentioned earlier, no evident barriers, including soil/habitat differences were detect in the field during the samples of the populations. In this sense, a more detailed ecological investigation is needed to identify the putative drivers of the differing flower colours between *P. juncifolia* Mello-Silva 3443 and *P. juncifolia* Mello-Silva 3444 populations.

The clade that includes *P. elegans* Lovo 144, *Pseudotrimezia* sp. Lovo 373 and *P. recurvata* Lovo 375 seems to be influenced by a geographic factor. *P. elegans* is a microendemic species from the Itambé peak. It is one of the few species of *Pseudotrimezia* with simple plain flower which is easily recognizable due the caespitose habit. The terminal entry *Pseudotrimezia* sp. Lovo 373 is from a region not far from *P. elegans* Lovo 144 and *P.*

recurvata Lovo 375 wasn't expected to be in that clade since it's also present in another region, where the voucher represents its most common occurrence (Lovo 2009). The species *P. sublateralis* is also easy to identify and is appears as closely related to the specimen *P. recurvata* Lovo 171. The placement of *P. recurvata* specimens in two distinct clades is an evidence of the labile morphology of those groups. Species delimitation within this lineage is quite difficult and relies almost entirely on leaf morphology, which in turn is also variable in *P. recurvata* (Lovo 2009). This latter is one of the most polimorphic regarding vegetative characters. In this sense, the two specimens identified as *P. recurvata* Ravenna may in turn not represent the same species. The divergent placement may be caused by the fact that both specimens show significant signal of allele sharing with distinct species.

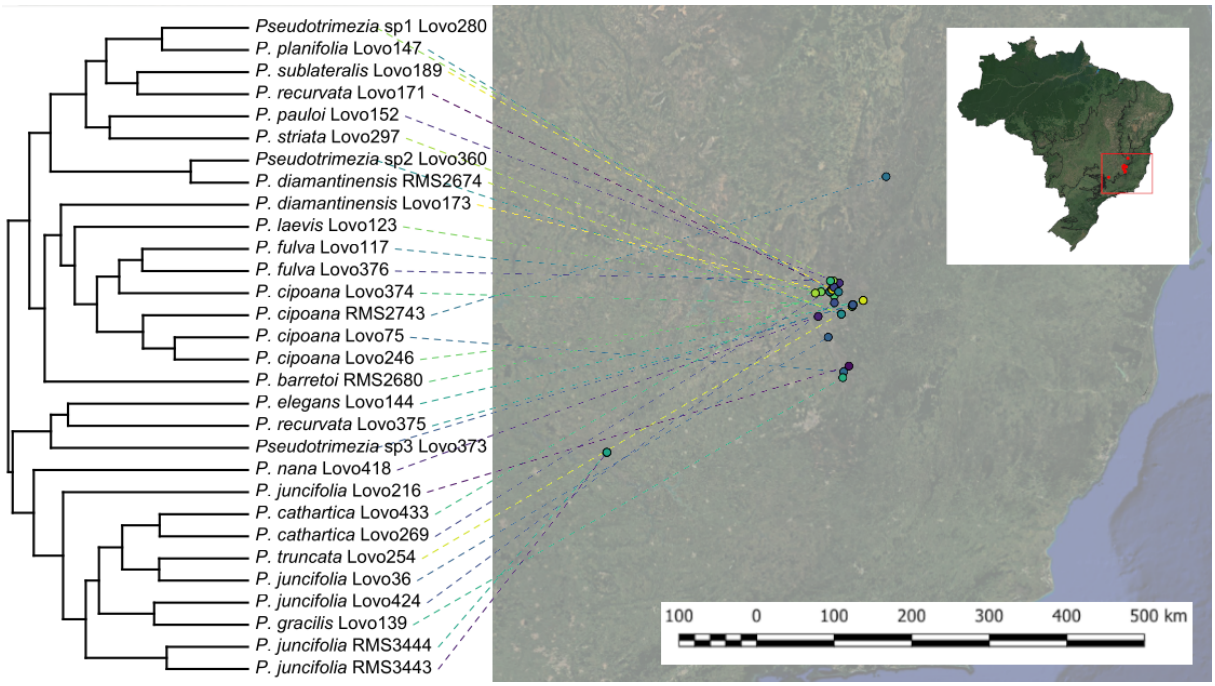


Figure 4. Phylogenetic relationships plotted over specimens geographical distribution. Different colors represent different terminals.

High levels of gene flow favours phenotypic homogeneity across the distribution ranges due to the shared alleles, especially in cases where gene flow is multidirectional. Then, once putative advantageous alleles are widespread, they may be selected upon in different locations. That may allow for the clade with highly supported species, but with high gene flow and absence of phenotypical differentiation (Crispo, 2008). The positive effect of gene flow in counteracting inbreeding depression is also significant for microendemic species with small populations, as are some *Pseudotrimezia* (Lovo et al., 2018). This may have helped the

survival and later diversification of the group by replenishing the genetic variation on which evolution depends (Garant et al, 2007). Previous studies on other clades from the *campo rupestre* do not show the same the results for gene flow dynamics among and within species. *Vellozia auriculata* (Velloziaceae) is endemic from the *campo rupestre* and displays low gene flow within populations and high local population persistence (Fiorini, 2019). Another study into bromeliads in the region also revealed that *Pitcairnia flammea,* which belongs to a recently diverged group, also shows low levels of gene flow among its populations and with other *Pitcairnia* (Mota et al., 2020). Moreover, results from the undergrad thesis of this author also approaching RAD-seq data for the genus *Barbacenia* (Velloziaceae), also did not indicate gene flow events despite the difficult taxonomic delimitation among the species and similar highly endemic distribution in this area (Santibañez, 2018). This might indicate that different processes drive the diversification and radiation of those groups in this species rich region.

High D values across the majority of trios generated for *Pseudotrimezia* is strong indication that introgression is a major source for the incongruence found, considering the absence of those are evidence of ILS in the presence of incongruence (Aardema et al., 2023). F-branch results did not show events where alleles from one specimen where shared among all specimens of an unrelated node, which would be a sign of ongoing introgressive hybridization (Harrison and Larson, 2014). Even though, there seems to be a geographical component of proximity influencing the gene flow events at first glance, since most events recovered were between individuals in close proximity. It is worth noting that significant D signals were present within specimen across all the spatial distribution sampled, which indicates that gene flow is ubiquitous in the group. This high gene flow may be the reason for the lack of evidence for ancient introgression events inherited by all descendants of a node., since high gene flow constrains adaptive divergence by homogenizing the genome of interbreeding groups (Garant et al., 2007).

Our further perspective is considering the effect of spatial distribution over the events will require a formal test such as the Mantel test to quantify the spatial processes driving population structure (Diniz-Filho et al, 2013). Also, further studies focusing on a genomic/developmental investigation is necessary to identify what mechanisms are responsible for the complex flower variation in order to ascertain if they were indeed acquired through gene flow.

2.5 CONCLUSION

Our analyses provide robust evidence for allele sharing due to the widespread occurrence of hybridization in the genus *Pseudotrimezia*, although with few identified events of introgressive hybridization. Additionally, we show significant evidence that gene flow was the main reason for the difficulty in retrieving high support phylogenies in previous studies (Lovo et al, 2012; Lovo et al, 2018). These dynamics of extensive gene flow among species may shed light on how paedomorphic flower characteristics developed repeatedly throughtout the group, which would be by sharing the genotype that allowed for this morphological phenotype. The selective pressure of pollination dynamics described by Lovo et al. (2021), may have then caused the selection of those traits. As also suggested by the absence of gene flow between parapatric populations of *P. juncifolia* that differ only on floral colours, the pressures of pollinators' interactions affecting reproductive isolation (or lack of thereof) seems to be crucial for the evolution of the genus as a whole.

3. REFERENCES

Aardema, M. L., Schmidt, K. L., & Amato, G. (2023). Patterns of cytonuclear discordance and divergence between subspecies of the scarlet macaw (Ara macao) in Central America. Genetica, 10.1007/s10709-023-00193-x. Advance online publication. https://doi.org/10.1007/s10709-023-00193-x

Acha, S., Linan, A., MacDougal, J., & Edwards, C. (2021). The evolutionary history of vines in a neotropical biodiversity hotspot: Phylogenomics and biogeography of a large passion flower clade (Passiflora section Decaloba). Molecular phylogenetics and evolution, 164, 107260. https://doi.org/10.1016/j.ympev.2021.107260

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome research, 19(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Avise, J. C., & Robinson, T. J. (2008). Hemiplasy: a new term in the lexicon of phylogenetics. Systematic biology, 57(3), 503–507. https://doi.org/10.1080/10635150802164587

Baxter SW, Davey JW, Johnston JS, et al. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. PLoS One, 6(4), e19315. Published 2011 Apr 26. doi:10.1371/journal.pone.0019315

Bouckaert, Remco et al. “BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis.” PLoS computational biology vol. 15, 4 e1006650. 8 Apr. 2019, doi:10.1371/journal.pcbi.1006650

Bradshaw, H. D., & Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. Nature, 426(6963), 176-8. doi: 10.1038/nature02106.

Chenuil, A. (2006). Choosing the right molecular genetic markers for studying biodiversity: from molecular evolution to practical aspects [published correction appears in Genetica. 2007 Sep;131(1):107. Anne, Chenuil [corrected to Chenuil, Anne]]. Genetica, 127(1-3), 101-120. doi: 10.1007/s10709-005-2485-1.

Colli-Silva, M., Vasconcelos, T. N. C., & Pirani, J. R. (2019). Outstanding plant endemism levels strongly support the recognition of campo rupestre provinces in mountaintops of eastern South America. J Biogeogr, 46, 1723-1733. https://doi.org/10.1111/jbi.13585.

Crispo, E. (2008). Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation, and gene flow. Journal of Evolutionary Biology, 21, 1460-1469. https://doi.org/10.1111/j.1420-9101.2008.01592.x.

DeFilippis, V. R., & Moore, W. S. (2000). Resolution of phylogenetic relationships among recently evolved species as a function of the amount of DNA sequence: an empirical study based on woodpeckers (Aves: Picidae). Mol Phylogenet Evol, 16(1), 143-160. doi: 10.1006/mpev.2000.0780.

Diniz-Filho, J. A., Soares, T. N., Lima, J. S., Dobrovolski, R., Landeiro, V. L., de Campos Telles, M. P., Rangel, T. F., & Bini, L. M. (2013). Mantel test in population genetics. Genetics and molecular biology, 36(4), 475-485. https://doi.org/10.1590/S1415-47572013000400002.

Dong, W., Liu, J., Yu, J., Wang, L., & Zhou, S. (2012). Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One, 7(4), e35071. doi: 10.1371/journal.pone.0035071.

Donkpegan ASL, Doucet JL, Hardy OJ, Heuertz M, Piñeiro R. Miocene Diversification in the Savannahs Precedes Tetraploid Rainforest Radiation in the African Tree Genus Afzelia (Detarioideae, Fabaceae). Front Plant Sci. 2020;11:798. Publicado em: 17 jun. 2020. doi:10.3389/fpls.2020.00798.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13.

Eaton, Deren A R, and Isaac Overcast. “ipyrad: Interactive assembly and analysis of RADseq datasets.” Bioinformatics (Oxford, England) vol. 36,8 (2020): 2592-2594. doi:10.1093/bioinformatics/btz966.

Eaton DA, Ree RH. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). Syst Biol. 2013;62(5):689-706. doi:10.1093/sysbio/syt032.

Eaton DAR, Spriggs EL, Park B, Donoghue MJ. Misconceptions on Missing Data in RAD-seq Phylogenetics with a Deep-scale Example from Flowering Plants. Syst Biol. 2017;66(3):399-412. doi:10.1093/sysbio/syw092.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792–1797. https://doi.org/10.1093/nar/gkh340.

Elworth, R.A.L., Ogilvie, H.A., Zhu, J., Nakhleh, L. (2019). Advances in Computational Methods for Phylogenetic Networks in the Presence of Hybridization. In: Warnow, T. (eds) Bioinformatics and Phylogenetics. Computational Biology, vol 29. Springer, Cham. https://doi.org/10.1007/978-3-030-10837-3_13.

Emerson KJ, Merz CR, Catchen JM, et al. Resolving postglacial phylogeography using high-throughput sequencing. Proc Natl Acad Sci U S A. 2010;107(37):16196-16200. doi:10.1073/pnas.1006538107.

Donkpegan ASL, Doucet JL, Hardy OJ, Heuertz M, Piñeiro R. Miocene Diversification in the Savannahs Precedes Tetraploid Rainforest Radiation in the African Tree Genus Afzelia (Detarioideae, Fabaceae). Front Plant Sci. 2020;11:798. Publicado em: 17 jun. 2020. doi:10.3389/fpls.2020.00798.

Escudero M, Nieto Feliner G, Pokorny L, Spalink D, Viruel J. Editorial: Phylogenomic Approaches to Deal With Particularly Challenging Plant Lineages. Front Plant Sci. 2020;11:591762. Publicado em: 24 nov. 2020. doi:10.3389/fpls.2020.591762.

Farris JS, Ka¨llersjo¨ M, Kluge AG, Bult C. 1994. Testing significance of congruence. Cladistics 10: 315–319.

Fiorini CF, Miranda MD, Silva-Pereira V, Barbosa AR, Oliveira U, Kamino LH, Mota NF, Viana PL, Borba EL. The phylogeography of Vellozia auriculata (Velloziaceae) supports low zygotic gene flow and local population persistence in the campo rupestre, a Neotropical OCBIL. Botanical Journal of the Linnean Society. 2019;191(3):381–398. doi:10.1093/botlinnean/boz051.

Fabreti LG, Höhna S. Bayesian inference of phylogeny is robust to substitution model over-parameterization. bioRxiv. 2022.02.17.480861. doi: https://doi.org/10.1101/2022.02.17.480861.

Garant D, Forde SE, Hendry AP. The multifarious effects of dispersal and gene flow on contemporary adaptation. Functional Ecology. 2007;21:434-443. doi:10.1111/j.1365-2435.2006.01228.x.

Gregory RPG, Bradshaw AD. Heavy metal tolerance in populations of Agrostis tenuis Sibth. and other grasses. New Phytol. 1965;64:131–143.

Hamlin JAP, Simmonds TJ, Arnold ML. Niche conservatism for ecological preference in the Louisiana Iris species complex. Biological Journal of the Linnean Society. 2017;120:144–154.

Harrison RG, Larson EL. Hybridization, introgression, and the nature of species boundaries. The Journal of heredity. 2014;105(Suppl 1):795-809. doi:10.1093/jhered/esu033.

Herrera S, Shank TM. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Molecular phylogenetics and evolution. 2016;100:70-79. doi:10.1016/j.ympev.2016.03.010.

Hibbins MS, et al. Phylogenomic comparative methods: Accurate evolutionary inferences in the presence of gene tree discordance. Proceedings of the National Academy of Sciences of the United States of America. 2023;120(22):e2220389120. doi:10.1073/pnas.2220389120.

Hudson RR, Kaplan NL. The coalescent process and background selection. Philosophical transactions of the Royal Society of London. Series B, Biological sciences. 1995;349(1327):19-23. doi:10.1098/rstb.1995.0086.

Karimi N, et al. Reticulate Evolution Helps Explain Apparent Homoplasy in Floral Biology and Pollination in Baobabs (Adansonia; Bombacoideae; Malvaceae). Systematic biology. 2020;69(3):462-478. doi:10.1093/sysbio/syz073.

Lepage T, et al. A general comparison of relaxed molecular clock models. Molecular biology and evolution. 2007;24(12):2669-2680. doi:10.1093/molbev/msm193.

Lovo J. Filogenia e Revisão de Pseudotrimezia (Iridaceae). São Paulo, 2009. [Dissertação].

Lovo J, Winkworth RC, Gil ASB, Amaral MCE, Bittrich V, Mello-Silva R. A revised genus-level taxonomy for Trimezieae (Iridaceae) based on expanded molecular and morphological analyses. Taxon. 2018;67(3):503-520. doi:10.12705/673.4.

Lovo J, Winkworth RC, Mello-Silva R. New insights into Trimezieae (Iridaceae) phylogeny: what do molecular data tell us? Ann Bot. 2012;110(3):689-702. doi:10.1093/aob/mcs127.

Maddison WP. Gene trees in species trees. Systematic Biology. 1997;46:523–536.

Maddison WP, Knowles LL. Inferring phylogeny despite incomplete lineage sorting. Syst Biol. 2006;55(1):21-30. doi:10.1080/10635150500354928.

Malinsky M, et al. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat Ecol Evol. 2018;2(12):1940-1955. doi:10.1038/s41559-018-0717-x.

Malinsky M, Matschiner M, Svardal H. Dsuite - Fast D-statistics and related admixture evidence from VCF files. Mol Ecol Resour. 2021;21(2):584-595. doi:10.1111/1755-0998.13265.

Mota MR, Pinheiro F, Leal BSS, et al. From micro- to macroevolution: insights from a Neotropical bromeliad with high population genetic structure adapted to rock outcrops. Heredity. 2020;125(4):353–370. doi:10.1038/s41437-020-0342-8.

Paetzold C, Wood KR, Eaton DAR, Wagner WL, Appelhans MS. Phylogeny of Hawaiian Melicope (Rutaceae): RAD-seq Resolves Species Relationships and Reveals Ancient Introgression. Front Plant Sci. 2019;10:1074. doi:10.3389/fpls.2019.01074.

Patterson N, Moorjani P, Luo Y, et al. Ancient admixture in human history. Genetics. 2012;192(3):1065-1093. doi:10.1534/genetics.112.145037.

Pfender WF, Saha MC, Johnson EA, Slabaugh MB. Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in Lolium perenne. Theor Appl Genet. 2011;122(8):1467-1480. doi:10.1007/s00122-011-1546-3.

Piwczyński M, et al. Unraveling phylogenetic relationships and species boundaries in the arid adapted Gerbillus rodents (Muridae: Gerbillinae) by RAD-seq data. Molecular phylogenetics and evolution. 2023;107913. doi:10.1016/j.ympev.2023.107913.

Pritchard JK, et al. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945-959. doi:10.1093/genetics/155.2.945.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 2018;67(5):901-904. doi:10.1093/sysbio/syy032.

Rubin BE, Ree RH, Moreau CS. Inferring phylogenies from RAD sequence data. PLoS One. 2012;7(4):e33394. doi:10.1371/journal.pone.0033394.

SANTIBANEZ, Victor. Avaliação de marcadores de sequenciamento RAD para inferência filogenética de máxima verossimilhança no gênero Barbacenia (Velloziaceae). 2020. 59 p. Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina,

Centro de Ciências Biológicas, Graduação em Ciências Biológicas, Florianópolis, 2020. Orientadora: Suzana Alcantara.

Schemske DW, Bradshaw HD Jr. Pollinator preference and the evolution of floral traits in monkeyflowers (Mimulus). Proc Natl Acad Sci U S A. 1999;96(21):11910-11915. doi:10.1073/pnas.96.21.11910.

Sitam FT, Salgado-Lynn M, Denel A, Panjang E, McEwing R, Lightson A, Ogden R, Maruji NA, Yahya NK, Ngau C, Mohd Kulaimi NA, Ithnin H, Rovie-Ryan J, Abu Bakar MS, Ewart KM. Phylogeography of the Sunda pangolin, Manis javanica: Implications for taxonomy, conservation management and wildlife forensics. Ecol Evol. 2023;13(8):e10373. doi:10.1002/ece3.10373.

Su N, et al. Molecular phylogeny and inflorescence evolution of Prunus (Rosaceae) based on RAD-seq and genome skimming analyses. Plant diversity. 2023;45(4):397-408. doi:10.1016/j.pld.2023.03.013.

Templeton A. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. Evolution. 1983;37:221–244.

Vasconcelos TNC, Alcantara S, Andrino CO, et al. Fast diversification through a mosaic of evolutionary histories characterizes the endemic flora of ancient Neotropical mountains. Proc Biol Sci. 2020;287(1923):20192933. doi:10.1098/rspb.2019.2933.

Walder RY, Walder JA, Donelson JE. The organization and complete nucleotide sequence of the PstI restriction-modification system. J Biol Chem. 1984;259(12):8015-8026.

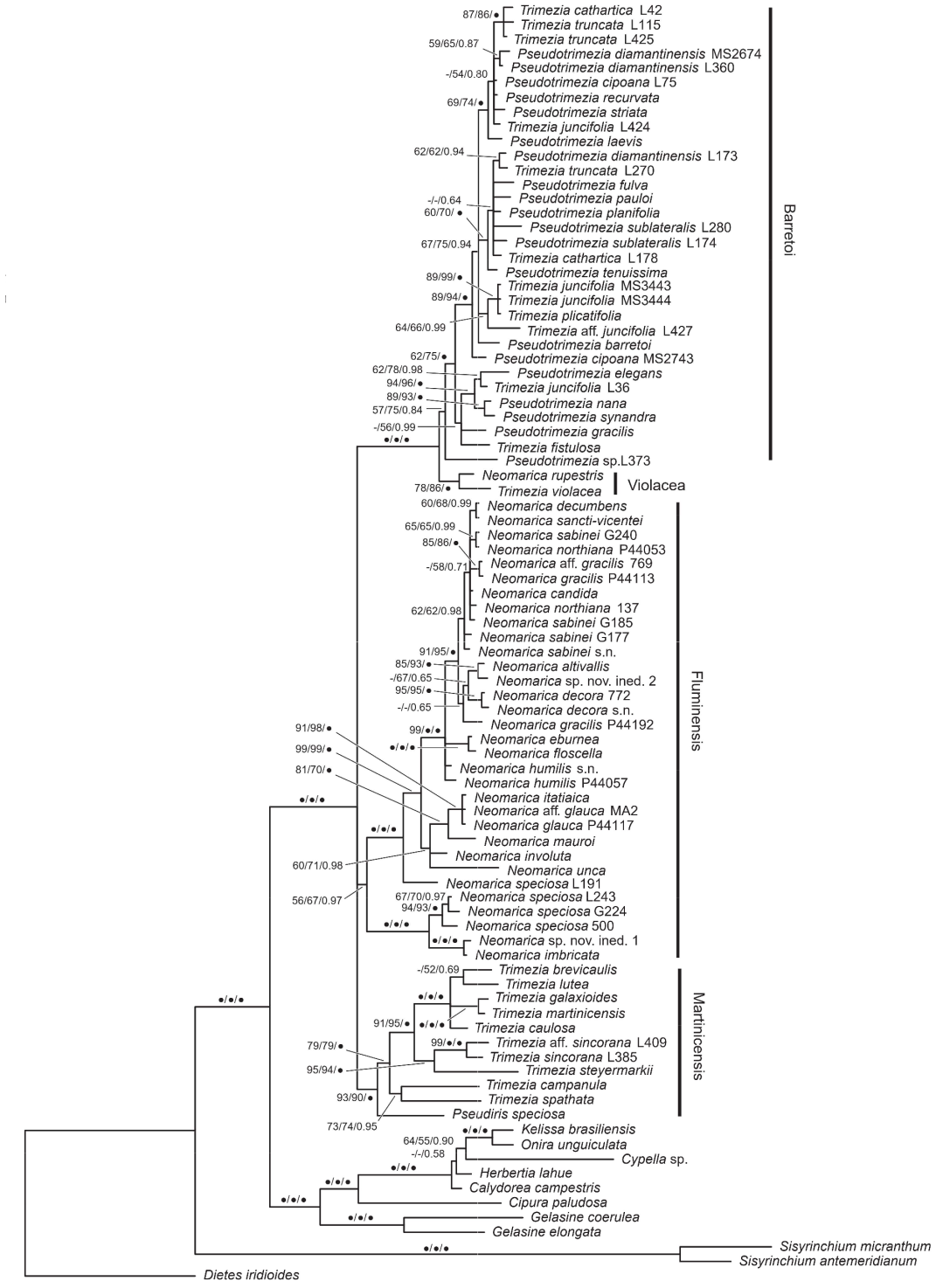
Zheng Y, Janke A. Gene flow analysis method, the D-statistic, is robust in a wide parameter space. BMC Bioinformatics. 2018;19:10. doi:10.1186/s12859-017-2002-4.

4. SUPPLEMENTARY MATERIAL

Supplementary material 1. Title.

Specimen	reads raw	reads passed filter	clusters total	clusters hidepth	hetero est	error est	reads consens	loci in assembly
P. barretoi Mello-Silva2680	112592	111808	3082	1940	0.005407	0.001955	1861	1103
P. cathartica Lovo269	306092	289521	6891	4497	0.006058	0.002073	4283	1926
P. cathartica Lovo433	178126	176904	3487	2509	0.006159	0.001961	2393	1580
P. cipoana Lovo246	144651	143407	2969	2039	0.004996	0.002073	1951	1265
P. cipoana Lovo374	481131	474622	5917	3960	0.009271	0.00214	3656	2141
P. cipoana Lovo75	66579	66019	1998	1208	0.001548	0.001912	1193	521
P. cipoana Mello-Silva2743	262529	260146	4381	2962	0.005165	0.002045	2826	1786
P. diamantinensis Lovo173	211478	210163	4413	2970	0.007605	0.001997	2810	1628
P. diamantinensis Mello-Silva2674	237367	233219	6608	3757	0.001909	0.002024	3688	713
P. elegans Lovo144	124595	123590	3831	2329	0.003101	0.001881	2262	942
P. fulva Lovo117	113692	112845	3255	1937	0.005893	0.002019	1846	1051
P. fulva Lovo376	276787	273126	5070	3322	0.006486	0.002133	3156	1658
P. gracilis Lovo139	414306	409916	6944	4475	0.006041	0.002135	4266	2145
P. juncifolia Lovo216	130599	129152	3902	2250	0.004318	0.002061	2181	781
P. juncifolia Lovo36	214251	211963	4311	2857	0.005913	0.002135	2732	1564
P. juncifolia Lovo424	755909	745546	7373	4766	0.008546	0.002106	4446	2477
P. juncifolia Mello-Silva3443	353964	350608	5187	3620	0.007771	0.002073	3402	2113
P. juncifolia Mello-Silva3444	142741	141777	3403	2183	0.005402	0.002016	2083	1208
P. laevis Lovo123	110827	109979	3134	1948	0.005082	0.001995	1869	1058
P. nana Lovo418	512237	506331	11681	7319	0.004623	0.002105	7060	1984
P. pauloi Lovo152	99416	98280	3123	1861	0.003165	0.002024	1813	861
P. planifolia Lovo147	240164	238070	4042	2918	0.007169	0.001905	2762	1845
P. recurvata Lovo171	53580	52946	1821	1007	0.0031	0.00204	980	401
P. recurvata Lovo375	595755	589222	9432	6005	0.005351	0.00206	5753	2210
P. striata Lovo297	241883	238806	4198	2770	0.004599	0.00205	2677	1752
P. sublateralis Lovo189	98182	97489	2809	1761	0.004505	0.001865	1702	988
P. truncata Lovo254	130428	128929	3117	2026	0.0044	0.002003	1952	1151
Pseudotrimezia sp1 Lovo280	533780	529556	11340	7774	0.007327	0.002481	7324	1889
Pseudotrimezia sp2 Lovo360	712155	705233	6905	4662	0.011143	0.002106	4195	2348
Pseudotrimezia sp3 Lovo373	209799	208567	3688	2604	0.005505	0.00182	2501	1579
T. aff. chimantensis Lovo409	449731	446715	6078	4347	0.005451	0.001884	4190	983

Supplementary material 2. Trimezieae filogeny recovered with the chloroplast markers trnG intron, trnH-psbA intergenic spacer, trnK intron, rpoC1 intron, and rps16 intron by Lovo et al. (2018).



Supplementary material 3. Trimezieae filogeny recovered with nrITS markers by Lovo et al. (2018).