

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE BIOCÊNCIAS  
BACHARELADO EM CIÊNCIAS BIOLÓGICAS

FERNANDA ENDERLE

**IDENTIFICAÇÃO MOLECULAR EM MEMBROS DA TRIBO MYRTEAE DC.  
(MYRTACEAE) ATRAVÉS DA TÉCNICA DE *DNA BARCODING***

Porto Alegre

2022

FERNANDA ENDERLE

**IDENTIFICAÇÃO MOLECULAR EM MEMBROS DA TRIBO MYRTEAE DC.  
(MYRTACEAE) ATRAVÉS DA TÉCNICA DE *DNA BARCODING***

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharela em Ciências Biológicas.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Andreia Carina Turchetto Zolet.

Porto Alegre

2022

## CIP - Catalogação na Publicação

Enderle, Fernanda

Identificação molecular em membros da tribo Myrteae DC. (Myrtaceae) através da técnica de DNA barcoding / Fernanda Enderle. -- 2022.

93 f.

Orientador: Andreia Carina Turchetto Zolet.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Bacharelado em Ciências Biológicas, Porto Alegre, BR-RS, 2022.

1. DNA barcoding. 2. Myrteae. 3. Neotrópicos. 4. genética vegetal. 5. bioinformática. I. Turchetto Zolet, Andreia Carina, orient. II. Título.

FERNANDA ENDERLE

**IDENTIFICAÇÃO MOLECULAR EM MEMBROS DA TRIBO MYRTEAE DC.  
(MYRTACEAE) ATRAVÉS DA TÉCNICA DE *DNA BARCODING***

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharela em Ciências Biológicas.

Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Andreia Carina Turchetto Zolet.

**Aprovado em:** 11 de outubro de 2022

BANCA EXAMINADORA:

---

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Andreia Carina Turchetto Zolet (orientadora)  
Instituto de Biociências - Departamento de Genética (UFRGS)

---

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Fernanda Bered  
Instituto de Biociências - Departamento de Genética (UFRGS)

---

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Tatiana Teixeira de Souza Chies  
Instituto de Biociências - Departamento de Botânica (UFRGS)

## AGRADECIMENTOS

Agradeço primeiramente aos meus pais, Ivaldo e Jenilde, por terem oportunizado meu estudo e a concretização desta monografia. Obrigada pelo apoio durante todos esses anos!

Aos meus amigos guaporenses, Gabriel, Lisandra e Patrick, pelos lanches e momentos de descontração em meio ao estresse da rotina acadêmica, com um adendo especial à Lisandra pela amizade de quase duas décadas e por todas as aventuras e desafios que enfrentamos juntas ao decidirmos nos mudar para Porto Alegre.

Aos amigos que fiz no curso de Física Computacional da UFRGS e, posteriormente, na Biologia. Obrigada, Andressa, Caroline, Giovanna, Júlia e Livia, por tornarem as semanas de prova, as saídas de campo e as intermináveis filas do RU mais leves e felizes!

Ao Patric, pela paciência e suporte durante os momentos difíceis e frustrantes, por sempre ter uma palavra amorosa e um abraço tranquilizador. Obrigada por tudo, *dear!*

Agradeço aos servidores, técnicos e docentes da UFRGS, essenciais para o funcionamento da universidade, e à minha orientadora, Andreia, por ter me recebido na iniciação científica, me introduzido à família Myrtaceae e acolhido com carinho as ideias e inseguranças relacionadas ao TCC. Também deixo registrada minha gratidão ao Ossman, que não participou diretamente deste trabalho, mas foi um grande amigo e supervisor ao longo de todas as etapas da iniciação científica.

À Fernanda Antunes e à Thais Vasconcelos, pela troca de ideias durante a elaboração do projeto e pelo compartilhamento de materiais e imagens.

À Fernanda Bered e à Tatiana Chies, que generosamente aceitaram participar da banca avaliadora com entusiasmo e prontidão.

Obrigada!

*We humans look rather different from a tree.  
Without a doubt we perceive the world differently  
than a tree does. But down deep, at the molecular  
heart of life, the trees and we are essentially  
identical.*

**Carl Sagan**

## RESUMO

*DNA barcoding* é uma técnica que visa discriminar indivíduos e delimitar espécies a partir do uso de sequências nucleotídicas curtas, variáveis e universais. Sua aplicação pode ser muito oportuna para a análise de materiais biológicos degradados ou imaturos e espécies crípticas, cujos caracteres morfológicos normalmente não são informativos. Assim, o presente trabalho objetivou testar a performance de marcadores moleculares com propriedades de *barcode* na tribo Myrteae DC. (Myrtaceae), um grupo rico, taxonomicamente complexo e constituído por membros que apresentam importância econômica, ecológica e cultural. Um total de 312 amostras pertencentes a 32 espécies foram selecionadas para a construção de uma biblioteca de referência contendo sequências dos espaçadores internos transcritos do DNA ribossomal (ITS) e de dois marcadores plastidiais (*psbA-trnH* e *matK*). Esses dados foram produzidos principalmente por estudos de filogenia molecular, tendo sido recuperados da plataforma do GenBank. A eficiência discriminatória individual das regiões foi acessada por meio de métodos baseados em distâncias genéticas, incluindo a caracterização da variabilidade a nível intra e interespecífico, a verificação de *barcoding gaps* e a implementação de três testes de similaridade entre as sequências nucleotídicas no programa TAXONDNA. Três metodologias de reconstrução filogenética também foram utilizadas para avaliar a capacidade dos marcadores em formar agrupamentos espécie-específicos, sendo elas o *neighbor-joining*, a máxima verossimilhança e a inferência Bayesiana. Os ITS produziram os melhores resultados, com mais de 50% das espécies apresentando *barcoding gaps* ou sendo reunidas em clados exclusivos. Além disso, de 58% a 81% dos indivíduos foram identificados corretamente com o uso desses espaçadores nos testes de similaridade e nas filogenias, dependendo da estratégia adotada. O marcador *psbA-trnH* foi a segunda região mais polimórfica, mas teve o pior desempenho, possivelmente devido à proporção considerável de divergências interespecíficas nulas. Já o *matK* exibiu taxas de discriminação modestas nas análises executadas com o TAXONDNA, porém insuficientes nos testes de *barcoding gap* e nas reconstruções filogenéticas, onde foi capaz de identificar no máximo 40% das espécies. O desfecho desse trabalho sugere que *barcodes* com origens plastidiais e nucleares diferem quanto à sua performance em Myrteae, sendo que a aplicação da técnica nessa tribo pode ser comprometida por uma série de fatores característicos de linhagens botânicas complexas, ainda que estudos com maior abrangência amostral devam ser realizados. O potencial dos candidatos nucleares, entretanto, pode ser aproveitado como ponto de partida para investigações mais profundas e integrativas dentro desse grupo.

**Palavras-chave:** *DNA barcoding*; Myrteae; Neotrópicos; genética vegetal; bioinformática.

## ABSTRACT

DNA barcoding is a technique that aims to discriminate individuals and delimit species using short, variable and universal nucleotide sequences. Its application can be very opportune for the analysis of degraded or immature biological materials and cryptic species, whose morphological characters are normally uninformative. The present work aimed to test the performance of molecular markers with barcode properties in the tribe Myrteae DC. (Myrtaceae), a rich and taxonomically complex group with economic, ecological and cultural importance. A total of 312 samples belonging to 32 species were selected for the construction of a reference library containing sequences of the nuclear ribosomal internal transcribed spacers (ITS) and two plastid markers (*psbA-trnH* and *matK*). These data were mainly produced by molecular phylogeny studies and retrieved from the GenBank platform. The individual discriminatory efficiency of the regions was accessed by methods based on genetic distances, including the characterization of variability at intra and interspecific levels, verification of barcoding gaps and implementation of three similarity tests between nucleotide sequences in TAXONDNA program. Three phylogenetic reconstruction methodologies were also used to evaluate the ability of markers to form species-specific clusters, namely neighbor-joining, maximum likelihood and Bayesian inference. ITS produced the best results, with more than 50% of the species showing barcoding gaps or being grouped into exclusive clades. In addition, 58% to 81% of individuals were correctly identified using these spacers in similarity tests and phylogenies, depending on the strategy adopted. The *psbA-trnH* marker was the second most polymorphic region but had the worst performance, possibly due to the considerable proportion of null interspecific divergences. On the other hand, *matK* showed modest discrimination rates in the analyses executed with TAXONDNA but insufficient in the barcoding gap tests and phylogenetic reconstructions, where it was able to identify at most 40% of the species. The outcome of this work suggests that barcodes with plastid and nuclear origins differ in terms of performance in Myrteae, and the application of DNA barcoding in this tribe can be compromised by a series of factors typical of complex botanical lineages, although studies with greater sample scope should be carried out. However, the potential of the nuclear candidates can be used as a starting point for deeper and integrative investigations within this group.

**Keywords:** DNA barcoding; Myrteae; Neotropics; vegetal genetics; bioinformatics.



## LISTA DE ILUSTRAÇÕES

Figura 1 - Diversidade de flores e frutos encontrada em representantes da tribo Myrteae.....	18
Figura 2 - Principais etapas envolvidas na geração de um <i>barcode</i> de DNA.....	19
Figura 3 - Representação esquemática de alguns marcadores plastidiais e nucleares testados nas análises de <i>DNA barcoding</i> em plantas. ....	21
Figura 4 - Representação esquemática de um <i>barcoding gap</i> ideal e da versão realista.....	26
Figura 5 - Marcadores moleculares sequenciados em estudos filogenéticos e de <i>DNA barcoding</i> na tribo Myrteae.....	34
Figura 6 - Histogramas com a frequência relativa das distâncias entre pares intra e interespecíficos para os marcadores <i>matK</i> , <i>psbA-trnH</i> e ITS.....	37-38
Figura 7 - Gráficos comparando a distância intraespecífica máxima e a distância interespecífica mínima para os marcadores <i>matK</i> , <i>psbA-trnH</i> e ITS.....	39
Figura 8 - Representação esquemática do possível impacto exercido pelo fluxo gênico entre populações sobre o sucesso discriminatório de um <i>barcode</i> .....	46
Quadro 1 - Relação de alguns estudos sobre <i>DNA barcoding</i> que incluem espécies nativas do Brasil ou da região Neotropical, com os marcadores utilizados, metodologias e considerações dos autores.....	47-49

## LISTA DE TABELAS

Tabela 1 - Principais características e parâmetros de variabilidade encontrados para os três marcadores analisados.....	36
Tabela 2 - Métricas de divergência intra e interespecífica calculadas para os três potenciais <i>barcodes</i> do estudo em termos de unidades de distância.....	36
Tabela 3 - Performance de cada marcador com base nos testes de <i>Best Match</i> , <i>Best Close Match</i> e <i>All Species Barcodes</i> .....	41
Tabela 4 - Performance de cada marcador com base nos testes de <i>Best Close Match</i> e <i>All Species Barcodes</i> considerando dois limites de corte .....	41
Tabela 5 - Número de espécies e indivíduos identificados corretamente através de três metodologias de reconstrução filogenética.....	44

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>12</b>
<b>1.1</b>	<b>Objetivos.....</b>	<b>15</b>
1.1.1	Objetivo geral.....	15
1.1.2	Objetivos específicos.....	15
<b>2</b>	<b>REFERENCIAL TEÓRICO.....</b>	<b>15</b>
<b>2.1</b>	<b>A tribo Myrteae DC. (Myrtaceae) como objeto de estudo.....</b>	<b>16</b>
<b>2.2</b>	<b>Premissas e desafios da técnica de <i>DNA barcoding</i>.....</b>	<b>18</b>
<b>2.3</b>	<b>Principais aplicações em organismos vegetais.....</b>	<b>22</b>
<b>2.4</b>	<b>Estratégias analíticas comumente empregadas.....</b>	<b>24</b>
<b>3</b>	<b>METODOLOGIA.....</b>	<b>28</b>
<b>3.1</b>	<b>Levantamento bibliográfico dos marcadores moleculares.....</b>	<b>28</b>
<b>3.2</b>	<b>Construção da biblioteca de referência.....</b>	<b>29</b>
<b>3.3</b>	<b>Análises computacionais.....</b>	<b>29</b>
3.3.1	Alinhamento das sequências nucleotídicas.....	29
3.3.2	Avaliação da performance dos marcadores através de distâncias genéticas.....	30
3.3.3	Testes de similaridade entre as sequências nucleotídicas.....	31
3.3.4	Avaliação da performance dos marcadores através de filogenias.....	32
<b>4</b>	<b>RESULTADOS E ANÁLISE.....</b>	<b>33</b>
<b>4.1</b>	<b>Marcadores moleculares selecionados.....</b>	<b>33</b>
<b>4.2</b>	<b>Performance dos marcadores.....</b>	<b>35</b>
4.2.1	Caracterização dos alinhamentos, dos parâmetros de variabilidade e das distâncias genéticas.....	35
4.2.2	Detecção dos <i>barcoding gaps</i> .....	38
4.2.3	Taxas de identificação por diferentes testes de similaridade.....	39
4.2.4	Agrupamentos produzidos pelas reconstruções filogenéticas.....	42
<b>5</b>	<b>DISCUSSÃO.....</b>	<b>44</b>
<b>5.1</b>	<b>Diferenças entre regiões plastidiais e nucleares.....</b>	<b>44</b>
<b>5.2</b>	<b>Razões que podem comprometer o uso do <i>DNA barcoding</i> em famílias botânicas.....</b>	<b>50</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS E PERSPECTIVAS.....</b>	<b>54</b>
	<b>REFERÊNCIAS.....</b>	<b>58</b>

APÊNDICE A - RELAÇÃO DOS ARTIGOS SELECIONADOS PARA O LEVANTAMENTO DOS MARCADORES MOLECULARES SEQUENCIADOS NA TRIBO MYRTEAE.....	69
APÊNDICE B - BIBLIOTECA DE REFERÊNCIA DO TRABALHO, INCLUINDO AS ESPÉCIES FILTRADAS DURANTE A CURADORIA E OS CÓDIGOS DE ACESSO DE CADA MARCADOR NO GENBANK.....	73
APÊNDICE C - DISTÂNCIAS INTRAESPECÍFICA MÁXIMA E INTERESPECÍFICA MÍNIMA ENCONTRADAS PARA CADA ESPÉCIE.....	85
APÊNDICE D - TAXAS DE ERRO CUMULATIVO (EM N° DE INDIVÍDUOS) CALCULADAS PARA DIFERENTES LIMITES DE CORTE PELO PACOTE <i>SPIDER</i> NO R.....	90
APÊNDICE E - FILOGENIAS GERADAS POR INFERÊNCIA BAYESIANA.....	91

## 1 INTRODUÇÃO

Os desafios logísticos, financeiros e taxonômicos envolvidos na identificação de novas espécies são fatores que ainda dificultam a quantificação da verdadeira diversidade biológica presente na Terra (CAZZOLLA GATTI *et al.*, 2022). A extensão desse déficit descritivo varia de acordo com os organismos estudados e também depende das realidades socioeconômicas contrastantes entre os países, um panorama que, somado às taxas aceleradas de extinção em muitos grupos, cria uma demanda por ferramentas de monitoramento inovadoras (MAIA *et al.*, 2012). A aplicação dessas ferramentas, por sua vez, é mais urgente em territórios que concentram pontos quentes de biodiversidade global, a exemplo da Mata Atlântica brasileira (BOLSON *et al.*, 2015) e das florestas tropicais da Mesoamérica (LAHAYE *et al.*, 2008), considerando que esses locais ainda são os que apresentam a maior carência de recursos materiais e humanos adequados (MAIA *et al.*, 2012).

Parte desse problema começou a ser solucionada com a implementação de sistemas de diagnóstico em nível genômico. O desenvolvimento de técnicas moleculares como a reação em cadeia da polimerase (PCR) e o sequenciamento de DNA foram imprescindíveis nesse contexto, garantindo que algumas regiões polimórficas do genoma, conhecidas como marcadores moleculares, pudessem ser amplificadas e caracterizadas em diversos taxa relacionados, com destaque inicial para microrganismos e invertebrados de complexa diferenciação morfológica (BROWN; EMBERSON; PATERSON, 1999; BUCKLIN *et al.*, 1999). Assim, o gene mitocondrial que codifica para a subunidade I da enzima citocromo oxidase (COI) passou a ser considerado um marcador central em animais, visto que o mesmo possui em sua sequência nucleotídica variações capazes de distinguir muitas espécies com boa resolução taxonômica, funcionando de modo similar a um código de barras genético (HEBERT *et al.*, 2003). A técnica de isolamento de um fragmento de DNA curto e padronizado, que tem por finalidade identificar um indivíduo através de comparações com sequências previamente disponíveis, popularizou-se, portanto, com a denominação de *DNA barcoding* (HEBERT *et al.*, 2003).

Apesar de bem estabelecido em muitos animais, o emprego do gene COI como marcador é inapropriado para a maior parte das espécies de plantas, considerando que o genoma mitocondrial desses organismos em geral apresenta taxas lentas de substituição (MOWER *et al.*, 2007), além de rearranjos frequentes e transferência horizontal de genes oriundos de parasitas (MOWER *et al.*, 2004). Desse modo, a combinação de duas regiões plastidiais, pertencentes aos genes que codificam para a subunidade maior da enzima

ribulose-1,5-bisfosfato carboxilase/oxigenase (*rbcL*) e para a proteína maturase K (*matK*), foi recomendada como principal alternativa de *barcode* em plantas terrestres, ainda que apresentando um êxito classificatório de apenas 72% (CBOL PLANT WORKING GROUP, 2009).

A despeito dos entraves, os avanços nessa metodologia promoveram impactos significativos em inúmeras áreas do conhecimento, contornando alguns problemas inerentes à identificação baseada apenas em caracteres morfológicos. Essas limitações incluem a presença de plasticidade fenotípica e espécies crípticas, a dependência por indivíduos pertencentes a um determinado sexo ou estágio do ciclo de vida, além da demanda por profissionais habilitados na manipulação de chaves dicotômicas e com expertise em grupos biológicos específicos (HEBERT *et al.*, 2003). Para garantir que a investigação de espécies baseada no DNA seja eficaz, no entanto, é fundamental que existam bibliotecas de referência com boa qualidade e ampla cobertura, compostas por sequências nucleotídicas que estejam associadas a um espécime previamente descrito e depositado como material testemunho (voucher) em uma coleção científica (JONES *et al.*, 2021). A existência de grandes bancos de dados, como o *Barcode of Life Data System* (BOLD) (RATNASINGHAM; HEBERT, 2007) e o GenBank (BENSON *et al.*, 2012) é de grande valia nesse sentido, garantindo que informações genômicas possam ser armazenadas e estejam prontamente acessíveis para o cruzamento de dados em análises posteriores.

O fato de que muitas bibliotecas ainda estão incompletas ou apresentam uma escassez substancial de dados, reitera, portanto, a necessidade de aumento dos esforços futuros de sequenciamento, com prioridade para grupos ricos em espécies, com indivíduos ameaçados e cuja diferenciação pela morfologia seja complicada. Ao considerar a perspectiva botânica e o panorama nacional, esses grupos podem abranger famílias como Lauraceae, Rubiaceae, Fabaceae, Melastomataceae e Myrtaceae (DE LIMA *et al.*, 2018). Muitos indivíduos dessa última destacam-se pela relevância econômica e cultural, incluindo a pimenta-da-jamaica (*Pimenta dioica* (L.) Merr.), a pitanga (*Eugenia uniflora* L.), as jabuticabas (*Plinia* spp.), o cravo-da-índia (*Syzygium aromaticum* (L.) Merr. & L.M.Perry) e os eucaliptos (*Corymbia* spp., *Eucalyptus* spp.), provendo frutos ricos em compostos bactericidas, antioxidantes e antiinflamatórios, condimentos, antissépticos, óleos essenciais, fibras de celulose e madeira, além de serem muitas vezes empregados no paisagismo, na arborização urbana e na recuperação de áreas degradadas (CORADIN; SIMINSKI; REIS, 2011; PROENÇA *et al.*, 2022).

A existência de um grande número de espécies é, entretanto, um aspecto que desafia a delimitação de subgrupos monofiléticos em Myrteae DC., tribo à qual estão subordinados todos os membros brasileiros de Myrtaceae (FLICKINGER *et al.*, 2020; PROENÇA *et al.*, 2022). Os gêneros nativos da região Neotropical, que engloba desde a Flórida e ilhas caribenhas até o cone sul da América do Sul (LUCAS *et al.*, 2007), são frequentemente considerados pesadelos taxonômicos (MCVAUGH, 1968), com algumas espécies apresentando centenas de sinônimos em decorrência de eventos sucessivos de descrição. A uniformidade das estruturas vegetativas e reprodutivas dentro do grupo muitas vezes limita o número de caracteres taxonomicamente informativos (FLICKINGER *et al.*, 2020), ao passo que outras espécies apresentam alta variação fenotípica devido à influência exercida pelas diferentes condições ambientais e geográficas ao longo de suas zonas de distribuição (BÜNGER *et al.*, 2015).

O emprego do *DNA barcoding* em conjunto com abordagens convencionais de identificação pode, portanto, auxiliar na estimativa mais fidedigna da riqueza taxonômica encontrada em Myrtaceae, permitindo que programas adequados de manejo e conservação sejam elaborados. De Lima *et al.* (2018), por exemplo, sugerem essa como uma das famílias prioritárias para a implementação da técnica no estado de São Paulo, considerando aspectos como sua alta diversidade e existência de exemplares endêmicos. Ademais, Myrtaceae figura entre as dez famílias botânicas mais ameaçadas pela degradação e perda de hábitat no território brasileiro, de acordo com levantamentos coordenados pelo Centro Nacional de Conservação da Flora na última década (MARTINELLI; MORAES, 2013).

A delimitação de *barcodes* eficientes também é capaz de gerar aplicações que ultrapassam o campo taxonômico, visto que muitos indivíduos são utilizados como matéria-prima na indústria de alimentos, bebidas, perfumes, cosméticos e fitoterápicos (CORADIN; SIMINSKI; REIS, 2011), garantindo que essa metodologia de diagnóstico venha a ser empregada no controle da qualidade de diferentes classes de produtos. Além disso, espécies como a goiaba (*Psidium guajava* L.) e o araçá (*Psidium cattleianum* Sabine) também apresentam alta tendência invasiva devido à disseminação de suas sementes por aves, o que as tornam potenciais agentes de desequilíbrio natural e econômico em algumas regiões do planeta (WILSON, 2011). Conseqüentemente, sua rastreabilidade em investigações forenses poderá ser beneficiada através da adoção de técnicas moleculares de identificação.

No entanto, a escassez de literatura relacionada a *DNA barcoding* com enfoque em Myrtaceae é uma lacuna de conhecimento que ainda perdura, especialmente no que concerne aos representantes Neotropicais. O sistema de busca taxonômica do BOLD conta com apenas

857 espécies de Myrtaceae registradas com ao menos uma sequência até setembro de 2022 (BOLD, 2022). A maior parte dos trabalhos na área investiga um número bastante limitado de espécies, inseridas principalmente nos gêneros *Syzygium* (ROSLIM, 2019) e *Eucalyptus* (DE LUCA; SIBILIO; GUACCHIO, 2014), que pertencem a outras tribos. Em contrapartida, uma parcela considerável das sequências nucleotídicas disponíveis para espécies de Myrteae nativas das Américas ainda decorre de trabalhos focados em reconstruções filogenéticas. Embora os marcadores sequenciados para essa finalidade não representem registros específicos de *barcodes* de DNA, algumas regiões plastidiais e nucleares comumente empregadas em filogenia molecular têm demonstrado boa performance para a identificação de plantas terrestres, conforme discutido por Chase *et al.* (2005), o que amplia as possibilidades de exploração dos dados publicados no GenBank.

## 1.1 Objetivos

### 1.1.1 Objetivo geral

O presente trabalho almeja investigar a eficiência de marcadores moleculares com potencial para *barcode* em espécies da tribo Myrteae (Myrtaceae), produzindo dados colaborativos para estudos taxonômicos e multidisciplinares.

### 1.1.2 Objetivos específicos

- a) Realizar um levantamento bibliográfico dos marcadores que já foram sequenciados em estudos prévios de *DNA barcoding* e de filogenia molecular em Myrteae.
- b) Construir uma biblioteca de referência a partir das sequências nucleotídicas disponíveis no GenBank para as espécies selecionadas na etapa anterior.
- c) Testar o poder discriminatório dos marcadores através de duas metodologias principais, baseadas na estimativa das distâncias genéticas entre sequências de DNA intra e interespecíficas e em reconstruções filogenéticas.

## 2 REFERENCIAL TEÓRICO

### 2.1 A tribo Myrteae DC. (Myrtaceae) como objeto de estudo



Myrtaceae Juss. é uma família de plantas com fisionomia arbórea ou arbustiva que comporta cerca de 5500 espécies e 142 gêneros (WILSON, 2011). Sua origem remonta ao período Cretáceo, com uma distribuição moderna predominante no hemisfério sul, em territórios derivados da fragmentação do antigo continente de Gondwana (THORNHILL *et al.*, 2015). Atualmente, Myrtaceae é considerada a oitava maior família de angiospermas no mundo e uma das mais ricas em espécies nas regiões Neotropicais (STAGGEMEIER *et al.*, 2015), sendo dividida nas subfamílias Psiloxylodeae, que engloba apenas duas tribos e dois gêneros africanos, e Myrtoideae, constituída pelas 15 tribos restantes (PROENÇA *et al.*, 2022; WILSON *et al.*, 2005). Muitos indivíduos atuam como importantes componentes florísticos e ecológicos em ambientes úmidos da América do Sul, Austrália e Ásia tropical, onde apresentam picos de diversidade (BALBINOTT *et al.*, 2022; PROENÇA *et al.*, 2022; VASCONCELOS *et al.*, 2017). O projeto “Espécies Nativas da Flora Brasileira de Valor Econômico Atual ou Potencial, de Uso Local e Regional - Plantas para o Futuro”, vinculado ao Ministério do Meio Ambiente, também elencou Myrtaceae como um dos grupos botânicos de maior notoriedade na região sul do Brasil, dada sua relevância para uma série de aplicações e pesquisas no presente e futuro (CORADIN; SIMINSKI; REIS, 2011).

Myrteae é a tribo que apresenta o maior número de integrantes, incluindo as 1194 espécies e 29 gêneros de Myrtaceae estimados no território brasileiro (PROENÇA *et al.*, 2022). Sua distribuição abrange os mais variados domínios fitogeográficos, como a Amazônia, Caatinga, Cerrado, Pampa, Pantanal e Mata Atlântica, figurando entre os exemplares lenhosos dominantes nessa última e contribuindo enormemente para a estruturação das formações vegetais nativas (GRESSLER; PIZO; MORELLATO, 2006; PROENÇA *et al.*, 2022). Praticamente toda a diversidade de Myrtaceae encontrada nas regiões Neotropicais é representada por Myrteae, com exceção de uma única espécie da tribo Metrosidereeae (*Metrosideros stipularis*), restrita ao Chile (VASCONCELOS *et al.*, 2017). Ademais, dois terços das espécies já descritas estão concentradas nos gêneros *Eugenia* e *Myrcia*, que possuem cerca de 1000 e 700 espécies, respectivamente (VASCONCELOS *et al.*, 2017).

Além da Mata Atlântica e do Cerrado brasileiros, o Planalto das Guianas e o Caribe constituem os principais centros de diversificação global de Myrteae (MCVAUGH, 1968), embora a América Central continental e as terras baixas da bacia amazônica também contemplem uma biodiversidade significativa (LANDRUM, 1992; VASCONCELOS *et al.*, 2017). Cerca de 450 espécies e 15 gêneros podem ser encontrados em outros continentes, incluindo o sudeste asiático, o nordeste da Austrália, diversas ilhas do Pacífico (como a Nova

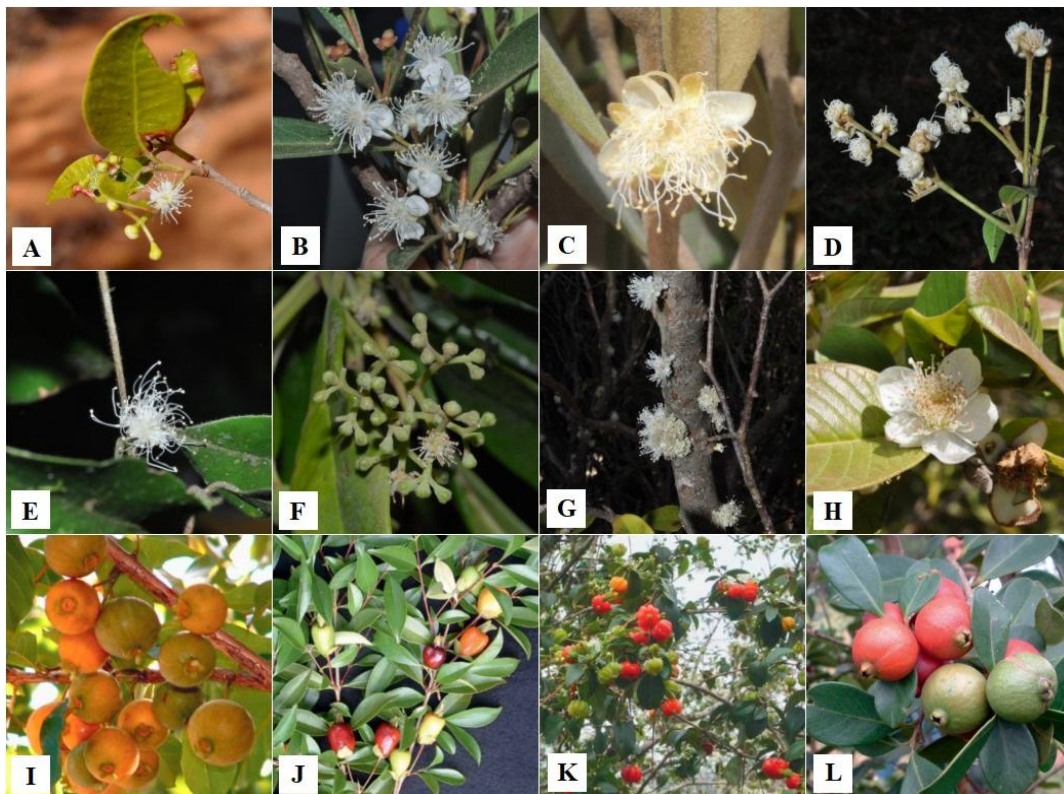
Caledônia e Nova Zelândia) e, em menor proporção, a África, as Ilhas de Madagascar e Maurício, e a Europa, essa última representada por uma única linhagem (VASCONCELOS *et al.*, 2017). Regiões subtropicais, temperadas e com grandes altitudes na América do Sul, como a Patagônia e os Andes, possuem uma riqueza menor de espécies, mas alguns gêneros de Myrteae são endêmicos desses locais (LANDRUM, 1986). As demais tribos ocorrem predominantemente na região Paleotropical e na Austrália (PROENÇA *et al.*, 2022).

Myrteae é definida como um grupo coerente e com valores altos de sustentação dos cladogramas em análises filogenéticas que utilizam dados moleculares plastidiais, nucleares e a combinação de ambos (VASCONCELOS *et al.*, 2017). Em relação à morfologia, seus membros são reconhecidos pelo compartilhamento de alguns caracteres diagnósticos, que incluem (i) presença de folhas simples e opostas, sem estípulas, com pontuações translúcidas, nervura central em geral proeminente, nervuras laterais facilmente detectáveis e margem inteira; (ii) flores tetra ou pentâmeras, bissexuadas, com corola esbranquiçada, por raras vezes com tons de rosa, vermelho ou roxo, dialipétalas, polistêmones, podendo formar diferentes padrões de agregação (inflorescências) e apresentando ovários ínferos que comportam de 2 a 18 lóculos, bi ou multiovulados, com placentação normalmente axial; (iii) frutos carnosos (baga) coroados pelos lobos ou remanescentes do cálice ou ainda por um tipo de cicatriz, contendo de uma a várias sementes com testas que apresentam texturas membranáceas, coriáceas ou ósseas (Figura 1) (LANDRUM; KAWASAKI, 1997; PROENÇA *et al.*, 2022). Esses frutos são consumidos por aves e mamíferos, que atuam como os principais organismos dispersores das sementes, ao passo que as flores são frequentemente visitadas por abelhas, responsáveis pela polinização de boa parte das espécies (GRESSLER; PIZO; MORELLATO, 2006). Tais aspectos justificam, portanto, o emprego da tribo como grupo modelo para estudos de evolução e conservação nos biomas Neotropicais (VASCONCELOS *et al.*, 2017).

Os representantes americanos de Myrteae também foram historicamente divididos em subtribos de acordo com diferentes critérios. A classificação proposta por De Candolle (1828) e Berg (1855–1856, 1857-1859), que circunscreveram as espécies em três grupos principais com base nas características encontradas no embrião, é uma das mais tradicionalmente reconhecidas (LUCAS *et al.*, 2007). O embrião de Myrtinae possui uma conformação circular, espiral ou em C, sendo constituído por dois cotilédones reduzidos e finos e um hipocótilo de mesmo tamanho ou muito mais longo. Já Eugeniinae apresenta embriões com cotilédones espessos e de formato reniforme, por vezes fundidos, além de hipocótilos curtos ou não evidentes. Myrciinae, em contrapartida, é definida por embriões cujos cotilédones são normalmente delgados, membranosos e dobrados, estando envoltos por um hipocótilo de

comprimento similar (LANDRUM; KAWASAKI, 1997; LUCAS *et al.*, 2005). Cada tipo de embrião, entretanto, parece ter surgido várias vezes de forma independente (LUCAS *et al.*, 2007). O caráter artificial desse arranjo fez com que McVaugh (1968) passasse a incluir outras estruturas, como ovário, cálice, inflorescências e testa das sementes, para o estabelecimento de seis grupos informais que representam as principais linhagens evolutivas dos Neotrópicos, com alguns gêneros mantendo-se isolados (MURILLO-A; STUESSY; RUIZ, 2013). Todavia, uma nova proposta de estruturação em nove subtribos foi publicada recentemente, tendo por suporte evidências moleculares e morfológicas (LUCAS *et al.*, 2019). Assim, a alta complexidade taxonômica encontrada em Myrteae pode ser beneficiada por estudos integrativos que adotem abordagens como o *DNA barcoding*.

**Figura 1 - Diversidade de flores e frutos encontrada em representantes da tribo Myrteae.**



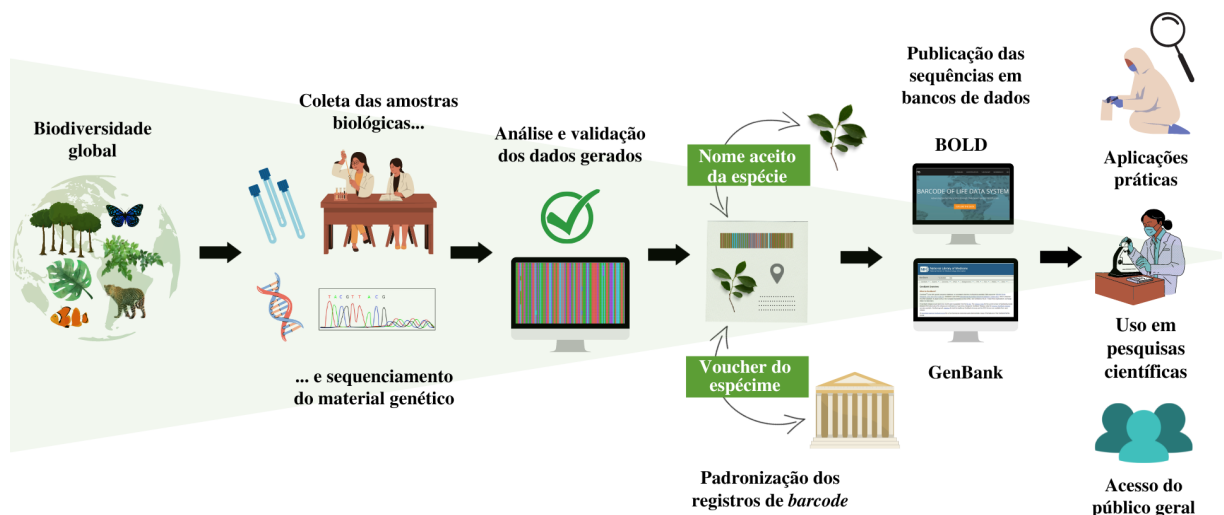
Legenda: A) *Blepharocalyx salicifolius*; B) *Eugenia cf. paracatuana*; C) *Myrceugenia alpigena*; D) *Myrcia splendens*; E) *Myrciaria cf. floribunda*; F) *Pimenta dioica*; G) *Plinia cauliflora*; H) *Psidium guajava*; I) *Campomanesia xanthocarpa*; J) *Eugenia involucrata*; K) *Eugenia uniflora*; L) *Psidium cattleyanum*. Fotos: Thais N. C. Vasconcelos (A-H) e Valdely F. Kinupp (I-L, retiradas de Coradin, Siminski e Reis (2011)).

## 2.2 Premissas e desafios da técnica de *DNA barcoding*

Os resultados promissores relatados pelos estudos pioneiros de *DNA barcoding* em animais (HEBERT *et al.*, 2003, 2004) auxiliaram a impulsionar outras pesquisas na área ao

longo dos últimos 19 anos. Conseqüentemente, grandes iniciativas internacionais com fins colaborativos foram fundadas, objetivando tornar exequível o estabelecimento de um sistema de identificação global baseado no DNA (Figura 2). O Consórcio para o Código de Barras da Vida (CBOL), criado em 2004 e composto por mais de 200 organizações científicas oriundas de 50 países, é um exemplo de projeto dedicado à investigação de novos *barcodes* e geração de seqüências para uma ampla gama de espécies biológicas (CONSORTIUM FOR THE BARCODE OF LIFE, 2012). O mesmo divide-se em equipes de pesquisa com expertise em taxa específicos, a exemplo do grupo de trabalho sobre protistas (PAWLOWSKI *et al.*, 2012), fungos (SEIFERT, 2009) e plantas terrestres (CBOL PLANT WORKING GROUP, 2009). Pode-se citar ainda uma série de alianças locais, como os projetos para *barcoding* da flora encontrada na Área de Conservação de Guanacaste (Costa Rica), das plantas medicinais chinesas e das briófitas do território britânico (HOLLINGSWORTH; GRAHAM; LITTLE, 2011). Nesse sentido, o *International Barcode of Life* (iBOL) também almeja expandir as bibliotecas de DNA disponíveis através do financiamento de programas específicos. O programa BIOSCAN, por exemplo, encontra-se em execução desde 2019 e tem como um de seus objetivos a ampliação da cobertura de *barcodes* para 2 milhões de espécies até o ano de 2028, superando o limite de 500 mil espécies alcançado em 2015 através da iniciativa conhecida como BARCODE 500K (INTERNATIONAL BARCODE OF LIFE, 2022).

Figura 2 - Principais etapas envolvidas na geração de um *barcode* de DNA.



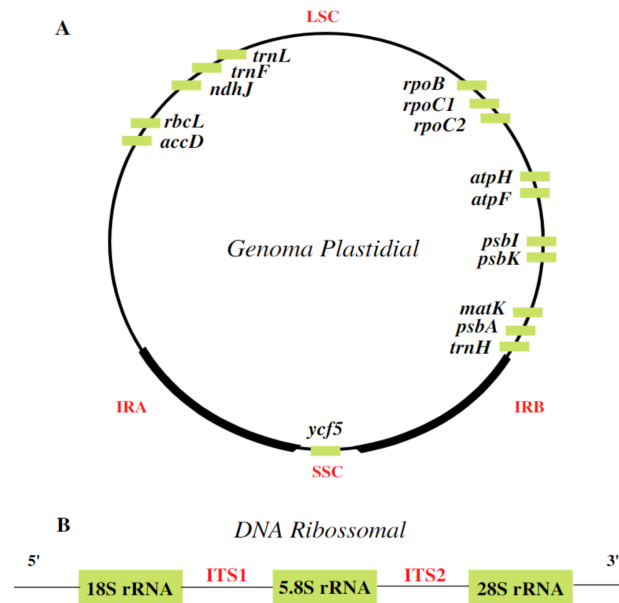
Fonte: elaborada pela autora (2022).

Segundo o CBOL (2009), um *barcode* ideal deve conter regiões flanqueadoras conservadas que permitam a sua recuperação rotineira a partir do uso de um par de iniciadores

(*primers*) universais, e a obtenção de sequências bidirecionais de qualidade não deve ser muito complexa, minimizando a necessidade de intervenções específicas. Além de possuir taxas satisfatórias de amplificação e sequenciamento, esses segmentos de DNA apresentam ainda como característica de interesse a capacidade de distinguir indivíduos de espécies diferentes, exibindo, simultaneamente, pouca divergência entre membros de uma mesma espécie. Desse modo, é esperado que exista uma lacuna separando as variações genéticas a nível intra e interespecífico, conhecida como *barcoding gap*, que possibilite a identificação acurada de um espécime (MEYER; PAULAY, 2005). Propriedades adicionais desejáveis em um *barcode* incluem um tamanho pequeno, visando facilitar o sequenciamento de materiais em diferentes estados de preservação, ausência de heterozigiosidade e facilidade de alinhamento (HOLLINGSWORTH *et al.*, 2009), bem como a habilidade de recuperar o monofiletismo das espécies (HEBERT *et al.*, 2004).

O sucesso da combinação dos genes *rbcL* e *matK*, inicialmente recomendados pelo CBOL para a identificação de espécies vegetais, mostrou-se, no entanto, tipicamente inferior ao do gene COI em animais (HOLLINGSWORTH; GRAHAM; LITTLE; 2011). Considerando esse aspecto, diversos estudos passaram a avaliar e propor marcadores alternativos nos genomas do cloroplasto e do núcleo com propriedades de *barcode*, sendo que o próprio CBOL estimulou a prática de incrementar os *loci* centrais com novas regiões informativas, caso necessário (CBOL PLANT WORKING GROUP, 2009). Dentre os candidatos mais promissores, pode-se citar o espaçador intergênico plastidial *psbA-trnH* e os espaçadores internos transcritos do DNA ribossomal (*internal transcribed spacers* - ITS), que por vezes são amplificados separadamente em dois fragmentos menores, denominados ITS1 e ITS2 (CHEN *et al.*, 2010; CHINA PLANT BOL GROUP, 2011; HOLLINGSWORTH, 2011; KRESS *et al.*, 2005). Demais regiões já testadas incluem genes que codificam para as subunidades da RNA polimerase plastidial (*rpoB*, *rpoC1* e *rpoC2*), para a subunidade  $\beta$  da acetil-CoA carboxilase (*accD*) e para outras proteínas do cloroplasto (*ndhJ* e *ycf5*), além dos espaçadores *psbK-psbI*, *atpF-atpH* e *trnL-trnF* (Figura 3) (VIJAYAN; TSOU, 2010). Um segmento do *locus* codificante 23S rDNA, conhecido como amplicon plastidial universal (UPA), também já foi avaliado, embora sua capacidade de discriminação tenha sido relatada como muito baixa (FAZEKAS *et al.*, 2008). As pesquisas pioneiras nessa área, bem como suas estratégias de amostragem e as recomendações de marcadores encontram-se sumarizadas no trabalho de Hollingsworth *et al.* (2009), ao passo que a revisão conduzida por Vijayan e Tsou (2010) discorre sobre as características básicas das regiões moleculares supracitadas e os principais *primers* utilizados na recuperação das mesmas.

**Figura 3 - Representação esquemática de alguns marcadores (A) plastidiais e (B) nucleares testados nas análises de *DNA barcoding* em plantas.**



Fonte: elaborada pela autora (2022).

Legenda: LSC = região de cópia única longa do genoma do cloroplasto; SSC = região de cópia única curta; IRA e IRB = regiões de repetição invertida A e B, respectivamente.

Embora apresentando atributos altamente desejáveis para um *barcode*, nenhuma dessas regiões enquadra-se integralmente nos critérios básicos de seleção quando avaliada de forma individual. Ao comparar os marcadores centrais (*core barcodes*), nota-se que a recuperação do *matK* acaba sendo desafiadora para alguns clados de samambaias, por exemplo, em função de rearranjos genômicos que tornam as regiões flangeadoras desse gene pouco conservadas, inviabilizando o emprego de *primers* universais (HOLLINGSWORTH; GRAHAM; LITTLE, 2011). Em contrapartida, o marcador *rbcL* é facilmente amplificado e sequenciado para a maioria das plantas terrestres, embora apresente um poder discriminatório considerado baixo (HOLLINGSWORTH, 2011).

O espaçador *psbA-trnH*, por sua vez, consiste no *locus* plastidial mais utilizado após a combinação *rbcL+matK*, tendo bons níveis de amplificação e melhorando a resolução dentro de taxa complexos (HOLLINGSWORTH; GRAHAM; LITTLE, 2011). No entanto, apesar de ser uma das regiões intergênicas com maior variabilidade encontrada em plantas (SHAW *et al.*, 2007), a presença recorrente de microinversões, o tamanho superior a 1000 pares de bases (pb) relatado para algumas espécies de coníferas (HOLLINGSWORTH *et al.*, 2009), a inserção de cópias dos genes/pseudogenes *rps19* e *rpl22* em monocotiledôneas (CHASE *et al.*, 2007), além do comprimento excessivamente curto (< 300 pb) para briófitas e parte das angiospermas (STECH; QUANDT, 2010; VIJAYAN; TSOU, 2010) são fatores que também

limitam sua aplicabilidade, seja pela possibilidade de estimação incorreta das distâncias genéticas ou pelo insucesso em obter sequências bidirecionais adequadas, o que demanda o uso de *primers* internos, táxon-específicos (CBOL PLANT WORKING GROUP, 2009; HOLLINGSWORTH; GRAHAM; LITTLE, 2011). Já os marcadores nucleares ITS, cuja incorporação ao grupo de *barcodes* centrais foi posteriormente recomendada para a identificação de espermatófitas (CHINA PLANT BOL GROUP, 2011), podem ser difíceis em termos de amplificação e sequenciamento devido à existência de estruturas secundárias e cópias divergentes de genes parálogos em um mesmo indivíduo (HOLLINGSWORTH, 2011; KRESS *et al.*, 2005). Somado a isso, algumas sequências fúngicas desses espaçadores são erroneamente atribuídas a espécies vegetais graças a eventos de contaminação ou presença de fungos endofíticos nas amostras biológicas (BUYS *et al.*, 2016; CHEN *et al.*, 2010), ainda que a extensão desses impactos tenha sido relatada como baixa e passível de solução (CHINA PLANT BOL GROUP, 2011; HOLLINGSWORTH, 2011; KRESS *et al.*, 2005).

A demanda por universalidade e taxas máximas de identificação simultaneamente gera, portanto, um conflito que dificulta a escolha de um *locus* global em organismos vegetais (KRESS; ERICKSON, 2007), visto que elevar o nível discriminatório muitas vezes culmina na redução da universalidade e vice-versa (ALVES, 2013). Isso acaba comprometendo a adoção em larga escala da técnica de *DNA barcoding*, de modo que a necessidade de empregar diferentes regiões combinadas para a melhora do poder resolutivo em grupos específicos é um aspecto que vem sendo frisado em diversos trabalhos (CHASE *et al.*, 2007; KRAWCZYK; SZCZECIŃSKA; SAWICKI, 2014; KRESS; ERICKSON, 2007). Entretanto, tal abordagem nem sempre resulta em um incremento significativo da capacidade de identificação, conforme constatado para espécies de Bromeliaceae (MAIA *et al.*, 2012) e Iridaceae (ALVES *et al.*, 2014), por exemplo, ao passo que alguns marcadores únicos já são capazes de diferenciar espécimes com grande sucesso em determinados grupos botânicos (CHEN *et al.*, 2010; LUGON *et al.*, 2021). Cabe observar ainda que há uma carência considerável de dados comparando a performance de todas essas regiões para um conjunto amostral único e taxonomicamente abrangente (HOLLINGSWORTH *et al.*, 2009), o que reforça a importância de novas investigações nesse campo.

### **2.3 Principais aplicações em organismos vegetais**

Os dois objetivos centrais do *DNA barcoding* consistem na identificação de organismos e na delimitação de espécies (COLLINS; CRUICKSHANK, 2013). Algumas

controvérsias e resistências surgiram na comunidade científica desde a proposição dessa abordagem, fundamentadas principalmente na ideia de que a mesma possa mitigar ou substituir a taxonomia tradicional (KRESS *et al.*, 2005), demandar custos elevados de implementação (CAMERON; RUBINOFF; WILL, 2006) e produzir resultados incorretos devido a limitações metodológicas e subamostragens dos taxa, por exemplo (MEYER; PAULAY, 2005; MORITZ; CICERO, 2004). No entanto, os sistemas de identificação baseados em caracteres morfológicos e em dados moleculares não são mutuamente excludentes, visto que empregam estratégias alternativas para alcançar um objetivo comum (DAYRAT, 2005). Um dos métodos pode, portanto, ser mais conclusivo em alguns casos, ao passo que outras situações exigem o uso concomitante de ambos os sistemas para a obtenção de respostas robustas (DAYRAT, 2005). Assim, a integração de procedimentos rápidos, confiáveis e acessíveis no campo da taxonomia é capaz de trazer uma série de benefícios para setores como a economia e a pesquisa básica e aplicada, superando, inclusive, os valores iniciais de investimento nessas técnicas (KRESS, 2017; PIRES; MARINONI, 2010).

O uso do *DNA barcoding* como ferramenta complementar para o aprimoramento de dados taxonômicos e filogenéticos tem auxiliado em aspectos que incluem a discriminação de indivíduos com morfologias convergentes (NAUER *et al.*, 2019) e a criação de bibliotecas contendo sequências de hospedeiros passíveis de infecção por patógenos específicos (BUYS *et al.*, 2016). Além disso, é possível estabelecer inventários de biodiversidade nos mais variados ecossistemas, como florestas tropicais (GONZALEZ *et al.*, 2009), zonas de manguezal (MAO *et al.*, 2021) e vegetações abertas, associadas a afloramentos rochosos (VASCONCELOS *et al.*, 2021). Esses resultados, em conjunto com estudos ecológicos e evolutivos, permitem que planos mais eficazes de monitoramento das espécies, manutenção dos serviços ecossistêmicos e conservação de comunidades vegetais suscetíveis a algum tipo de ameaça sejam executados (BUYS *et al.*, 2016; JONES *et al.*, 2021; MAO *et al.*, 2021).

A literatura sobre o tema abrange ainda inúmeras contribuições notáveis para o campo da botânica forense, especialmente no que concerne à biossegurança e saúde pública, através do controle de espécies invasoras (NIEMANN *et al.*, 2022; XU; LI; JIN, 2018), da fiscalização e autenticação de alimentos e suplementos herbais (LUGON *et al.*, 2021; NAKAMURA *et al.*, 2013), e do combate ao extrativismo e contrabando de plantas de interesse madeireiro, ornamental e medicinal que estejam em risco de extinção (BOLSON *et al.*, 2015; LAHAYE *et al.*, 2008; WILLIAMSON *et al.*, 2016). Ademais, pode-se citar a relevância da técnica na resolução de casos envolvendo envenenamento com espécies tóxicas (NITHANIYAL *et al.*, 2021) e rastreamento de plantações ilegais de *Cannabis sativa* (MELLO *et*



*al.*, 2016). Ribeiro *et al.* (2013), por exemplo, detectaram polimorfismos de nucleotídeo único (SNPs) na sequência do gene *rbcL* para amostras dessa droga oriundas dos Estados Unidos, Reino Unido, China e Rio de Janeiro, sugerindo que tal região pode atuar inclusive como um marcador biogeográfico de procedência, distinguindo os cultivares autorizados dos ilícitos.

Finalmente, a geração de *barcodes* também assegura que diferentes espécies sejam identificadas de forma simultânea e automatizada a partir de amostras ambientais, uma abordagem conhecida como *metabarcoding* (VASCONCELOS *et al.*, 2021). A possibilidade de trabalhar com segmentos pequenos de DNA faz desse método uma alternativa poderosa para análise de material genético degradado, principalmente com a adoção de alguns marcadores específicos, como o loop P6 do intron plastidial *trnL*, que apresenta boas taxas de amplificação (MILLA *et al.*, 2021; YOCCOZ *et al.*, 2012). Sendo assim, estudos envolvendo DNA *metabarcoding* em solo, fezes, pólen e mel, por exemplo, são capazes de realizar estimativas sobre a diversidade e abundância de plantas existentes em diferentes comunidades, incluindo as mudanças de composição observadas ao longo do tempo (YOCCOZ *et al.*, 2012). Essa metodologia viabiliza ainda uma maior compreensão acerca dos padrões de dieta e forrageamento em uma série de organismos que consomem recursos vegetais, como primatas (MALLOTT; GARBER; MALHI, 2018), morcegos (BELL *et al.*, 2021) e abelhas (CANNIZZARO *et al.*, 2022), auxiliando, portanto, em estudos relacionados a teias tróficas e outras redes de interações entre espécies (KRESS, 2017).

## 2.4 Estratégias analíticas comumente empregadas

Hollingsworth *et al.* (2009) definem três estratégias possíveis de amostragem em trabalhos sobre DNA *barcoding*, cada qual com suas potencialidades e limitações. A abordagem conhecida como pares de espécies, por exemplo, envolve a utilização de algumas espécies próximas e muitos gêneros filogeneticamente divergentes. Isso garante uma análise mais fina sobre a capacidade de amplificação e sequenciamento dos *loci* selecionados para uma ampla gama de taxa, ao passo que restringe a avaliação da sua performance discriminatória em nível de espécie, visto que poucos organismos pertencentes a um mesmo gênero são considerados. O trabalho de Kress e Erickson (2007), que aborda 48 gêneros diversificados contendo duas espécies cada, representa esse tipo de metodologia.

A estratégia florística, por outro lado, consiste em amostrar as espécies existentes em uma determinada área geográfica, o que também pode ser útil para testar a universalidade dos marcadores e ilustrar eventuais aplicações práticas do método, como o estudo de Vasconcelos

*et al.* (2021) com a flora presente na Serra dos Carajás (BR). Todavia, a ocorrência de organismos com diferentes graus de parentesco pode superestimar a performance dos *barcodes* nas análises de maior resolução. Assim, é esperado que plantas menos relacionadas apresentem divergências genéticas maiores, responsáveis por favorecer a identificação de cada espécie nesses estudos locais (ALVES, 2013).

Já a técnica baseada em um táxon busca abranger diversas espécies inseridas em um mesmo grupo, como o artigo de Krawczyk, Szczecińska e Sawicki (2014) focado no gênero *Lamium* L. (Lamiaceae). Nesse caso, não há dados suficientes para que conclusões robustas acerca dos níveis de universalidade das regiões estudadas sejam feitas, mas, em contrapartida, a delimitação de cada espécie pode ser investigada com maior precisão. No entanto, esse tipo de trabalho tende a apresentar resultados modestos, visto que as sequências nucleotídicas costumam ser mais homogêneas em grupos intimamente relacionados (ALVES, 2013).

As principais abordagens destinadas a testar a capacidade discriminatória dos marcadores, por sua vez, baseiam-se na estimativa das distâncias genéticas entre as sequências nucleotídicas e na construção de filogenias (MEIER *et al.*, 2006). O algoritmo de busca por alinhamentos básicos locais (BLAST), oferecido pela plataforma do Centro Nacional de Informação Biotecnológica (NCBI), também é frequentemente empregado como um método informal para avaliação dos *barcodes*, objetivando verificar quais sequências disponíveis na base de dados possuem a melhor correspondência (*best hit*) com a amostra investigada e qual a probabilidade do resultado ter sido gerado ao acaso, mensurada através de um parâmetro conhecido como “valor esperado” (*E-value*) (ROSS; MURUGAN; LI, 2008).

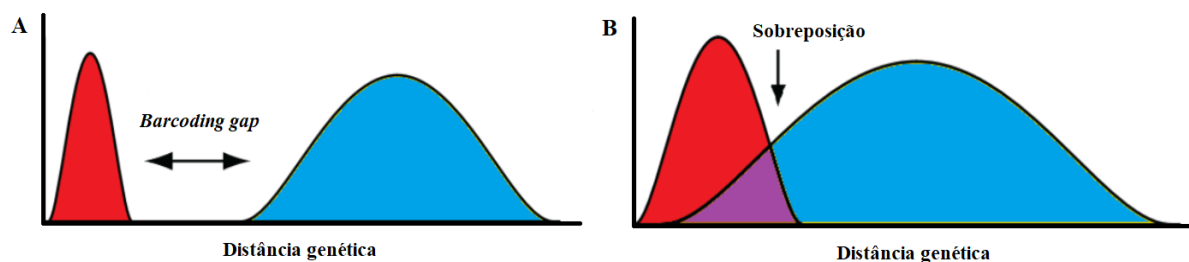
No que concerne às distâncias genéticas, diversas ferramentas e critérios foram propostos a fim de acessar a variabilidade existente a nível intra e interespecífico. O estabelecimento de um limite de corte entre esses dois cenários busca minimizar a chance de nomear erroneamente um espécime desconhecido caso não exista nenhuma sequência coespecífica no banco de referência (COLLINS; CRUICKSHANK, 2013), ainda que sua real eficácia seja debatida, principalmente no que tange a espécies muito próximas (MEYER; PAULAY, 2005). Ademais, é inviável aplicar um valor único para todos os grupos taxonômicos, levando em conta que aspectos como o tempo desde a especiação e as taxas de mutação são variáveis (COLLINS; CRUICKSHANK, 2013; VIRGILIO *et al.*, 2012).

O sistema do BOLD considera até 1% de divergência como parâmetro aceitável para um *match* entre uma sequência indefinida e um *barcode* previamente conhecido na biblioteca (RATNASINGHAM; HEBERT, 2007), enquanto Hebert *et al.* (2003) definiram um limite padronizado de 3% para estudos de *barcode* com lepidópteros. No entanto, o cálculo de

valores otimizados de acordo com os dados disponíveis é mais indicado para reduzir as taxas de falsos-positivos ou falsos-negativos, que podem oscilar consideravelmente quando um limite fixo é adotado (BROWN *et al.*, 2012; MEIER *et al.*, 2006; VIRGILIO *et al.*, 2012). Meier, Zhang e Ali (2008) demonstraram, por exemplo, que entre 7% e 26% de atribuições incorretas podem ser alcançadas ao aplicar os parâmetros arbitrários definidos pelo BOLD em diferentes grupos de metazoários. Os autores também criticam o uso de distâncias genéticas médias para a análise dos *barcoding gaps* e ressaltam a importância de considerar somente as menores distâncias interespecíficas, visto que a dificuldade imposta para identificar uma sequência a nível de espécie dependerá apenas do nível de variação intraespecífica e da diferença existente entre o melhor *match* da biblioteca e a sequência heteroespecífica mais similar. Conseqüentemente, o emprego de valores médios muitas vezes superestima o *gap* real (MEIER; ZHANG; ALI, 2008).

Alguns trabalhos optam por construir histogramas com a distribuição das distâncias genéticas entre cada par de sequências para verificar a existência de *barcoding gaps* globais (Figura 4) (BOLSON *et al.*, 2015; BUYS *et al.*, 2016; LAHAYE *et al.*, 2008). No entanto, Collins e Cruickshank (2013) reiteram que essa abordagem acaba gerando uma interpretação precipitada dos resultados, já que o grau de variabilidade presente dentro de um táxon pode exceder os valores de divergência encontrados entre outras espécies sem comprometer, necessariamente, o sucesso de identificação de um indivíduo, que é definido pela existência de *gaps* locais. A ilustração dessas diferenças através de gráficos de pontos, que comparam a distância intraespecífica máxima e a distância interespecífica mínima para cada espécie, é recomendada, portanto, como uma alternativa de maior acurácia nesse tipo de investigação (LE *et al.*, 2020; LIU *et al.*, 2014; XU; LI; JIN, 2018).

**Figura 4 - (A) Representação esquemática de um *barcoding gap* ideal, onde as distribuições das distâncias genéticas intra e interespecíficas são discretas e sem sobreposição e (B) versão realista, contemplando uma sobreposição significativa desses dados.**



Fonte: Meyer e Paulay, adaptada (2005).

As filogenias, por outro lado, consistem em representações gráficas em forma de árvore que retratam as hipóteses de relações existentes entre organismos ou sequências de nucleotídeos e proteínas ao longo do tempo, de modo que as pontas (descendentes) se conectam aos nós (ancestrais) através dos ramos (CALDART *et al.*, 2016; YANG; RANNALA, 2012). As reconstruções filogenéticas nos trabalhos de *barcode* almejam testar a capacidade dos marcadores em recuperar o monofiletismo das espécies. Sendo assim, a identificação é considerada satisfatória somente quando as sequências formam clados espécie-específicos (BOLSON *et al.*, 2015), ou seja, agrupamentos que incluem todos os membros de uma espécie como descendentes exclusivos de um ancestral comum. Isso costuma ser feito por meio de métodos dependentes de uma matriz de distâncias genéticas, como o *neighbor-joining* (NJ) (SAITOU; NEI, 1987), utilizado com frequência por ser rápido e eficiente para a análise de grandes conjuntos de dados (COLLINS; CRUICKSHANK, 2013; HEBERT *et al.*, 2003).

No entanto, metodologias baseadas em caracteres também podem ser empregadas, a exemplo da máxima verossimilhança (ML) (FELSENSTEIN, 1981) e da inferência Bayesiana (IB) (YANG; RANNALA, 1997), que demandam um maior esforço computacional por compararem todas as sequências do alinhamento considerando um sítio por vez (YANG; RANNALA, 2012). Os procedimentos usuais de avaliação da confiabilidade dessas topologias são a análise de *bootstrap* para os métodos de NJ e ML, e a estimativa da probabilidade posterior no caso da IB (YANG; RANNALA, 2012). A primeira consiste na amostragem aleatória e com reposição dos sítios do alinhamento em uma frequência equivalente ao tamanho das sequências originais, produzindo diversas árvores de *bootstrap* distintas que contemplam o mesmo conjunto de espécies, mas alguns caracteres repetidos e outros ausentes (FELSENSTEIN, 1985). Essas pseudo-réplicas são comparadas com a árvore oficial para verificar quantas vezes determinado agrupamento foi recuperado (CALDART *et al.*, 2016). Já a probabilidade posterior representa a chance de que a filogenia esteja correta a partir dos dados e do modelo evolutivo fornecidos ao algoritmo (YANG; RANNALA, 2012).

Testes de *barcode* amparados somente por reconstruções filogenéticas também não são recomendados por alguns autores. Dentre os principais problemas associados a essa prática estão a incapacidade de solucionar sequências posicionadas de forma ambígua (COLLINS; CRUICKSHANK, 2013) e o grau questionável de confiança de algumas topologias (MEIER *et al.*, 2006). Ademais, a existência de espécies parafiléticas em árvores de genes, ou seja, com parte dos indivíduos apresentando alelos mais similares aos de outra espécie do que aos do seu próprio táxon, pode fazer com que algumas sequências nucleotídicas não se agrupem

em um clado espécie-específico, produzindo identificações dúbias ou incorretas (COLLINS; CRUICKSHANK, 2013; FUNK; OMLAND, 2003). Mesmo assim, as filogenias são capazes de fornecer informações relevantes quando utilizadas em conjunto com outros métodos analíticos, especialmente por gerarem dados complementares de cunho evolutivo (VASCONCELOS *et al.*, 2021), o que justifica sua adoção em uma ampla gama de estudos (ALVES *et al.*, 2014; BOLSON *et al.*, 2015; MAIA *et al.*, 2012).

### **3 METODOLOGIA**

#### **3.1 Levantamento bibliográfico dos marcadores moleculares**

Inicialmente, a revisão elaborada por Souza Neto *et al.* (2022) foi empregada como material de suporte para a consulta dos principais trabalhos de inferência filogenética com base em marcadores moleculares na tribo Myrteae. Um levantamento bibliográfico complementar foi conduzido através dos repositórios virtuais Google Scholar (<https://scholar.google.com.br>) e Web of Science (<https://apps.webofknowledge.com/>) entre 13/03/2022 e 17/03/2022, objetivando detectar outros estudos de filogenia molecular e os principais trabalhos sobre *DNA barcoding* em Myrtaceae. Para isso, utilizou-se as terminologias “Myrtaceae” and “phylogen\*” e “Myrtaceae” and “DNA barcod\*”, respectivamente, no campo de busca avançada dessas plataformas. Apenas artigos que contemplassem em suas análises espécies pertencentes à tribo Myrteae foram considerados, resultando na seleção de 33 trabalhos referentes ao período de 2004 a 2022. Desses, 25 (75,8%) geraram sequências nucleotídicas próprias que puderam ser recuperadas através de um código de acesso no banco de dados GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). As principais informações de cada trabalho, como o título, ano de publicação, autores, temática central, marcadores adotados, número total de espécies e número de espécies de Myrteae podem ser consultadas no Apêndice A.

#### **3.2 Construção da biblioteca de referência**

Para o estabelecimento da biblioteca bruta de referência, alguns elementos relevantes foram extraídos de todos os artigos que produziram sequências inéditas e publicadas, visando eliminar coletas redundantes de dados. As informações incluíram: subtribos, gêneros e espécies de Myrteae abordadas, número de indivíduos de cada espécie, localização desses

indivíduos e tipo de marcadores sequenciados. Também foram considerados os vouchers associados a cada exemplar e, finalmente, os códigos de acesso de todas as sequências disponibilizadas no GenBank. Em seguida, apenas as espécies que apresentavam no mínimo 3 indivíduos foram selecionadas para compor a versão definitiva da biblioteca, garantindo que as análises das distâncias genéticas intraespecíficas pudessem ser viabilizadas posteriormente. Do total de 52 espécies obtidas, 6 possuíam distribuição restrita à Nova Zelândia e ilhas asiáticas, tendo sido descartadas com o intuito de abranger apenas taxa nativos dos Neotrópicos. Para as espécies remanescentes, foi realizada uma revisão mais detalhada da nomenclatura atualmente aceita e dos seus sinônimos através de consultas a especialistas e acesso a algumas plataformas online, como o *World Checklist of Selected Plant Families* (<http://wmsp.science.kew.org/qsearch.do>), o Sistema de Informação sobre a Biodiversidade Brasileira ([https://sibbr.gov.br/?lang=pt\\_BR](https://sibbr.gov.br/?lang=pt_BR)) e o Centro Nacional de Conservação da Flora (<http://cncflora.jbrj.gov.br/portal/>).

### 3.3 Análises computacionais

#### 3.3.1 Alinhamento das sequências nucleotídicas

O alinhamento múltiplo das sequências para cada marcador foi executado através do algoritmo MUSCLE (EDGAR, 2004), implementado no software MEGA11 (TAMURA; STECHER; KUMAR, 2021), com as configurações no modo padrão. As sequências passaram por uma curadoria automatizada no programa GBLOCKS v0.91.1 (CASTRESANA, 2000) para eliminar regiões alinhadas de forma ambígua ou com uma grande quantidade de *gaps*, o que otimiza a reprodutibilidade dessa etapa. O tipo de dado foi definido como DNA, enquanto o comprimento mínimo de um bloco de nucleotídeos foi alterado para 3. Metade das posições com *gaps* foram toleradas para evitar reduções drásticas no tamanho dos alinhamentos. Posteriormente, realizou-se uma inspeção visual dos resultados e intervenções manuais foram feitas, com bases ambíguas e *gaps* dispostos em série (dados faltantes) sendo codificados com o caractere “?”. Como muitos programas foram utilizados ao longo do trabalho, decidiu-se atribuir um acrônimo para cada sequência, constituído pela primeira letra do gênero e três primeiras letras do epíteto específico associadas ao número do indivíduo e separadas do código de acesso do GenBank por um subtraço (Ex.: Euni06\_AM489986), almejando evitar problemas de compatibilidade devido ao tipo ou quantidade de caracteres empregados. Após concluir o alinhamento, alguns parâmetros referentes a cada marcador, como a proporção de

*gaps* ou dados faltantes, o número de polimorfismos e sítios parcimoniosamente informativos (PI), bem como a diversidade de nucleotídeos ( $\pi$ ) foram acessados por meio da ferramenta DNASP v6.12.03 (ROZAS *et al.*, 2017) em conjunto com o MEGA.

### 3.3.2 Avaliação da performance dos marcadores através de distâncias genéticas

As distâncias genéticas entre cada par de sequências foram computadas com o emprego do MEGA, que fornece uma matriz com o resultado de todas as combinações possíveis através da opção “*Distance: Compute Pairwise Distances*”. A razão R entre o número de transições e transversões foi previamente verificada, tendo sido maior do que 0,5 para os três marcadores, o que evidencia um viés de substituição nucleotídica e a necessidade de adotar um método de correção para o cálculo dessas divergências. Assim, decidiu-se empregar o modelo evolutivo Kimura 2-parâmetros (KIMURA, 1980), tendo em vista seu amplo uso na literatura sobre *barcodes* (HEBERT *et al.*, 2003; HOLLINGSWORTH *et al.*, 2009; MEIER; ZHANG; ALI, 2008).

As distâncias intra e interespecíficas foram caracterizadas através dos seus valores médios e dos limites mínimos e máximos encontrados para cada marcador (LE *et al.*, 2020; TODESCHINI, 2019; XU; LI; JIN, 2018). Para sequências pertencentes a espécies diferentes, optou-se por analisar as distâncias apenas entre indivíduos congêneros e também dentro de toda a tribo, incluindo, assim, os gêneros representados por uma única espécie na biblioteca. A distribuição das distâncias entre pares, por sua vez, foi plotada em um histograma, com a frequência no eixo y e os valores de divergência representados em intervalos de 0,005 unidades de distância (0,5%) no eixo x.

A avaliação dos *barcoding gaps* envolveu a comparação entre as divergências intraespecífica máxima e interespecífica mínima para cada sequência. Esses parâmetros foram calculados por meio das funções `maxInDist` e `nonConDist` do pacote *Spider* (BROWN *et al.*, 2012), implementado no ambiente de programação R v4.1.3 (R CORE TEAM, 2022), novamente considerando o modelo Kimura 2-parâmetros, além da deleção pareada dos *gaps* (*pairwise deletion*). Uma espécie era classificada como desprovida de descontinuidades quando apresentava ao menos um indivíduo mais próximo ao de outra espécie do que dos seus coespecíficos. A maior variação intraespecífica e a menor distância interespecífica encontradas para cada espécie foram transpostas para um gráfico contendo uma reta de inclinação 1:1, de modo que apenas os pontos dispostos acima dessa reta representam

espécies com *gaps* (KRAWCZYK; SZCZECIŃSKA; SAWICKI, 2014; LE *et al.*, 2020; LIU *et al.*, 2014).

### 3.3.3 Testes de similaridade entre as sequências nucleotídicas

A ferramenta *Species Identifier* v1.9, incluída no programa TAXONDNA (MEIER *et al.*, 2006), foi utilizada para testar a habilidade discriminatória de cada marcador com base em comparações diretas entre as sequências previamente alinhadas. Como o software reconhece o nome das espécies a partir de duas palavras separadas por um espaço, com o primeiro termo iniciado em letra maiúscula, foi necessário renomear as sequências seguindo essa convenção. O modelo Kimura 2-parâmetros foi selecionado para o cálculo das distâncias genéticas na opção *Pairwise Summary*, com o valor mínimo de sobreposição entre as sequências tendo sido alterado para 250 pb. Na estratégia de *Best Match* (BM), o algoritmo inspeciona para cada sequência (*query*) aquela que é a mais próxima no conjunto de dados disponível, definindo como um acerto quando ambas pertencem à mesma espécie. Se sequências provenientes de diferentes espécies tiverem o mesmo nível de correspondência com a *query*, o resultado é classificado como ambíguo, ao passo que o *match* com uma sequência de outra espécie é contabilizado como uma atribuição errada. O *Best Close Match* (BCM) é uma variante mais rigorosa dessa técnica, pois um limite de corte, abaixo do qual 95% das distâncias intraespecíficas são encontradas, é previamente estabelecido. Se nenhuma correspondência for detectada nesse intervalo, a sequência permanece sem identificação. Do contrário, os critérios vigentes no BM são novamente aplicados. Finalmente, o teste de *All Species Barcodes* (ASB) utiliza esse mesmo limite de corte para produzir agrupamentos de sequências ordenadas de acordo com a sua similaridade em relação à *query*. O desfecho é considerado correto caso a sequência a ser comparada seja seguida pelos demais indivíduos da mesma espécie, contanto que existam pelo menos dois deles. Em contrapartida, a presença de apenas um ou alguns desses indivíduos dentro do limite estipulado torna o resultado ambíguo, assim como uma mistura de amostras coespecíficas e heteroespecíficas. Já uma *query* acompanhada por todas as sequências de outra espécie culmina em uma identificação incorreta (MEIER *et al.*, 2006).

A função *threshOpt* do pacote *Spider* foi executada no R como uma estratégia alternativa para calcular o limite de transição entre as distâncias intra e interespecíficas associado à menor taxa de erro cumulativo em cada conjunto de dados. O erro cumulativo corresponde à soma dos falsos positivos (quando o *match* coespecífico não está dentro do



limite de corte delimitado) e falsos negativos (quando existem sequências de outras espécies nesse intervalo, comprometendo a identificação) (BROWN *et al.*, 2012). O script foi configurado para inspecionar distâncias corrigidas pelo modelo Kimura 2-parâmetros e com deleção pareada dos *gaps* entre 0,1% e 5%, calculando os parâmetros de erro a cada 0,1%. Assim, os testes de BCM e ASB foram implementados com duas opções de limite de corte, uma computada pelo próprio TAXONDNA e outra fornecida pela função do R, com o intuito de investigar o impacto gerado por mudanças nesses valores.

### 3.3.4 Avaliação da performance dos marcadores através de filogenias

Os métodos filogenéticos de *neighbor-joining*, máxima verossimilhança e inferência Bayesiana foram empregados para testar a capacidade dos marcadores em formar clados compostos por todas as sequências de uma única espécie. As árvores de NJ foram produzidas no MEGA11 (TAMURA; STECHER; KUMAR, 2021), tendo-se selecionado o modelo de substituição nucleotídica Kimura 2-parâmetros e 1000 replicatas de *bootstrap*. O programa JMODELTEST v2.1.10 (DARRIBA *et al.*, 2012), por sua vez, foi utilizado com o objetivo de computar o modelo evolutivo mais adequado para as filogenias de ML e IB. Optou-se por adotar as configurações padronizadas nessa inspeção, de modo a incluir 88 opções de modelos, combinando parâmetros como a frequências das bases (+F), a proporção de sítios invariáveis (+I) e a taxa de variação entre os sítios (+G). O critério de informação Bayesiana (BIC) foi considerado para a seleção final, tendo definido o modelo TPM1uf+G (KIMURA, 1981) como o mais apropriado para os marcadores plastidiais e o TrNef+I+G (TAMURA; NEI, 1993) para o marcador do genoma nuclear.

As análises de ML foram conduzidas no servidor online do programa IQ-TREE (<http://iqtree.cibiv.univie.ac.at/>) (TRIFINOPOULOS *et al.*, 2016), utilizando os modelos de substituição nucleotídica supracitados e 4 categorias de gama (G). O suporte dos ramos foi verificado através do *bootstrap* padrão, com apenas 100 replicatas devido a limitações computacionais, de modo similar a Bolson *et al.* (2015) e Maia *et al.* (2012). Já o software BEAST v1.10.4 (SUCHARD *et al.*, 2018) foi utilizado para a geração das filogenias por IB, com emprego de um relógio molecular estrito e uma árvore prévia apresentando especiação do tipo Yule (*Yule Process*). Como o modelo de substituição nucleotídica TPM1uf+G não estava disponível no programa, a alternativa HKY+G (HASEGAWA; KISHINO; YANO, 1985) foi adotada como a opção mais similar para os marcadores *matK* e *psbA-trnH*, também com 4 categorias de gama. Duas corridas independentes com 30 milhões de gerações de

cadeias de Markov (MCMC) foram rodadas em cada caso, com os dados sendo registrados no arquivo de saída a cada 3000 gerações. Os resultados foram examinados através do Tracer v1.7.2 (RAMBAUT *et al.*, 2018) para confirmar se os tamanhos de amostra efetiva (ESS) de cada parâmetro eram suficientes ( $ESS > 200$ ). Em seguida, os arquivos contendo todas as árvores geradas foram combinados no *LogCombiner* para que uma árvore consenso de cada marcador pudesse ser finalmente construída através do *TreeAnnotator*, ambas as ferramentas incluídas no pacote do BEAST, com *burn-in* de 3 milhões de estados. Essas árvores e os valores de suporte dos clados foram inspecionados por meio do visualizador gráfico FIGTREE v1.4.4 (RAMBAUT, 2007). A identificação a nível de espécie foi considerada correta quando todos os indivíduos coespecíficos formavam grupos monofiléticos com *bootstrap* de no mínimo 70% (NJ e ML) (BOLSON *et al.*, 2015; GONZALEZ *et al.*, 2009) ou uma probabilidade posterior igual ou superior a 0,95 (IB) (ALVES *et al.*, 2014).

## 4 RESULTADOS E ANÁLISE

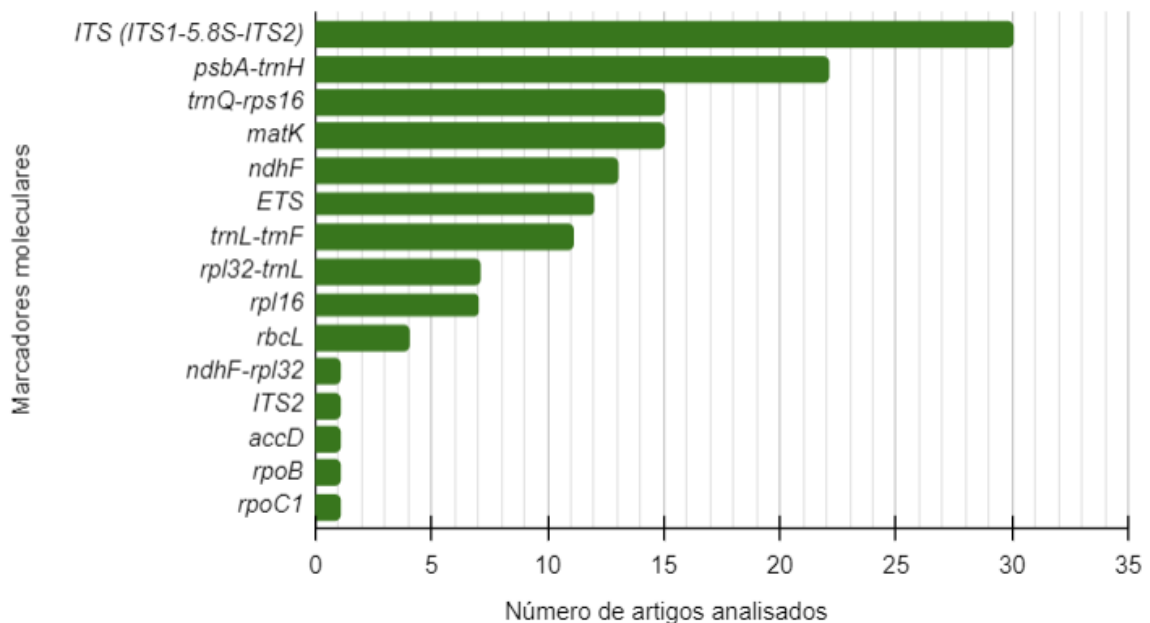
### 4.1 Marcadores moleculares selecionados

Dos 33 trabalhos selecionados no levantamento bibliográfico, 30 (90,9%) apresentaram a temática da filogenia molecular como um de seus assuntos centrais, em contraste com 3 artigos (9,1%) voltados a análises específicas de *DNA barcoding*. Além disso, 15 marcadores pertencentes a quase 600 espécies foram encontrados (Figura 5). Desses, 20% correspondem a espaçadores internos (ITS/ITS2) e externos (*external transcribed spacer - ETS*) de genes ribossomais. Os 80% restantes representam marcadores plastidiais, sendo que 50% são sequências codificantes e 50% envolvem regiões não codificantes (íntrons e espaçadores intergênicos). A região dos ITS lidera o ranking, aparecendo em 90,9% dos trabalhos filtrados, seguida pelos espaçadores plastidiais *psbA-trnH* e *trnQ-rps16*, com 66,7% e 45,5% de prevalência, respectivamente, além do gene *matK*, também presente em 45,5% dos artigos.

De todos os *loci* analisados, 5 destacaram-se por conter no mínimo 3 sequências disponíveis para um número maior de espécies, sendo eles os espaçadores ITS (35 espécies), o *psbA-trnH* (29), o *matK* (12), o ETS (12) e o *trnQ-rps16* (11). Esses marcadores foram submetidos a um alinhamento por meio da ferramenta BLAST com os dados do GenBank para detectar sequências adicionais de Myrtaceae que estivessem depositadas na plataforma. Tal mecanismo, somado à pesquisa ativa no campo de busca do repositório, permitiu que as

lacunas de mais 4 espécies fossem completadas para os ITS, 4 para o *psbA-trnH* e 6 para o *matK*. Informações não verificadas pelo GenBank e sem validação por um voucher ou por um link de acesso aos detalhes do indivíduo no BOLD foram descartadas, assim como espécies com as terminologias de *affinis* (*aff.*) e *conferatum* (*cf.*) e amostras que haviam sido previamente checadas por especialistas e cujo espécime depositado no herbário não correspondia ao táxon indicado. Essa filtragem acaba limitando a quantidade de material acessível, embora um processo criterioso seja importante para aumentar a confiabilidade das sequências utilizadas, especialmente quando todo o trabalho se baseia em dados *in silico*.

**Figura 5 - Marcadores moleculares sequenciados em estudos filogenéticos e de *DNA barcoding* na tribo Myrteae.**



Fonte: elaborado pela autora (2022).

Finalmente, optou-se por selecionar marcadores que contivessem amostras suficientes para o maior número possível de espécies em comum, considerando que uma avaliação comparativa entre os resultados obtidos seria viável apenas se os mesmos marcadores fossem utilizados para todas as espécies. Assim, a curadoria resultou na seleção de duas regiões não codificantes, uma plastidial (*psbA-trnH*) e outra nuclear (ITS), sequenciadas para 29 e 32 espécies de Myrteae pertencentes a 9 e 10 gêneros distintos, respectivamente. Uma margem de 10% de dado faltante foi aceita para o espaçador *psbA-trnH* com o objetivo de aumentar o número de taxa na análise. Desse total, 15 espécies apresentaram ainda uma região codificante completa, representada pelo gene *matK*, de modo que a mesma também foi considerada, ainda

que a amostragem limitada deva refletir em uma interpretação mais cautelosa dos resultados. As sequências do *matK* para a espécie *Eugenia involucrata* foram obtidas do nosso grupo de pesquisa, não estando depositadas no GenBank por fazerem parte de uma tese de doutorado ainda em fase de elaboração (SOUZA NETO *et al.* em preparação). Os gêneros com maior representatividade foram *Eugenia* (8 espécies), seguida por *Myrcia* (7), *Psidium* (6), *Myrceugenia* (3), *Myrcianthes* (2) e *Pimenta* (2), ao passo que *Blepharocalyx*, *Myrciaria*, *Myrrhinium* e *Plinia* apresentaram uma única espécie cada. A biblioteca completa e demais informações complementares podem ser acessadas no Apêndice B.

## 4.2 Performance dos marcadores

### 4.2.1 Caracterização dos alinhamentos, dos parâmetros de variabilidade e das distâncias genéticas

Um total de 312 sequências nucleotídicas foram selecionadas para as análises de *barcode*, sendo 129 pertencentes à região dos espaçadores ITS (ITS1-5.8S-ITS2), 111 ao espaçador *psbA-trnH* e 72 ao gene *matK*. Em muitos casos, foi inevitável utilizar indivíduos diferentes para cada marcador em razão da dependência aos dados disponíveis no GenBank. Ainda que isso configure uma limitação metodológica, todos os espécimes que passaram pela curadoria manual foram aceitos, incluindo aqueles classificados como variedades, considerando a necessidade de contemplar múltiplas amostras coespecíficas no trabalho.

Com relação à facilidade de alinhamento, os ITS e o *matK* não impuseram grandes dificuldades, minimizando a demanda por edições. No caso desse último, praticamente não houveram *gaps* internos, embora algumas sequências de *Myrceugenia ovata* tenham sido mais curtas do que as demais (263 pb), o que aumentou o número de *gaps* nas terminações. O espaçador *psbA-trnH*, por outro lado, foi mais desafiador nesse sentido pela existência de regiões contendo rearranjos genômicos e sítios com frequências consideráveis de dados faltantes (> 20%), que acabaram sendo removidos da análise. O número significativo de eventos de inserção, deleção, microssatélites e repetições invertidas é um aspecto frisado em outros trabalhos e torna o comprimento desse marcador altamente variável, podendo comprometer o alinhamento correto das sequências em alguns casos (KRESS; ERICKSON, 2007; LAHAYE *et al.*, 2008; XU; LI; JIN, 2018).

O conjunto de dados para os ITS forneceu a maior proporção de sítios polimórficos (38,5%), seguido pelo *psbA-trnH* (30,5%) e *matK* (11,3%). O mesmo padrão foi observado

para os sítios parcimoniosamente informativos, ou seja, aqueles que apresentam mais de um tipo de nucleotídeo, com ao menos dois deles ocorrendo em uma frequência mínima de duas sequências (TAMURA; STECHER; KUMAR, 2021), e para a diversidade nucleotídica  $\pi$ , um índice que representa o número médio de diferenças nucleotídicas entre duas sequências por sítio (ROZAS *et al.*, 2017) (Tabela 1).

**Tabela 1 - Principais características e parâmetros de variabilidade encontrados para os três marcadores analisados.**

Barcode	Nº de amostras (Nº de espécies)	Tamanho do alinhamento (pb)	Nº de sítios com <i>gaps</i> ou dados faltantes	Nº de sítios polimórficos (%)	Nº de sítios PI (%)	Diversidade $\pi$ de nucleotídeos
ITS	129 (32)	592	185	228 (38,5)	175 (29,6)	0,03190
<i>psbA-trnH</i>	111 (29)	416	198	127 (30,5)	88 (21,2)	0,02722
<i>matK</i>	72 (15)	531	344	60 (11,3)	38 (7,2)	0,01821

Fonte: elaborada pela autora (2022).

Em termos de descrição geral das distâncias genéticas, o gene *matK* foi o que apresentou a menor média intraespecífica, enquanto o locus *psbA-trnH* demonstrou um grau intermediário de variação e os espaçadores ITS exibiram o resultado mais elevado. O mesmo padrão também foi observado em relação às divergências entre espécies, cujas médias sempre excederam a das distâncias intraespecíficas. Pode-se ainda considerar que as divergências encontradas dentro da tribo tendem a apresentar valores mais elevados do que as distâncias calculadas apenas entre indivíduos do mesmo gênero, algo esperado em função da maior proximidade evolutiva das sequências nesse último caso (Tabela 2).

**Tabela 2 - Métricas de divergência intra e interespecífica calculadas para os três potenciais barcodes do estudo em termos de unidades de distância.**

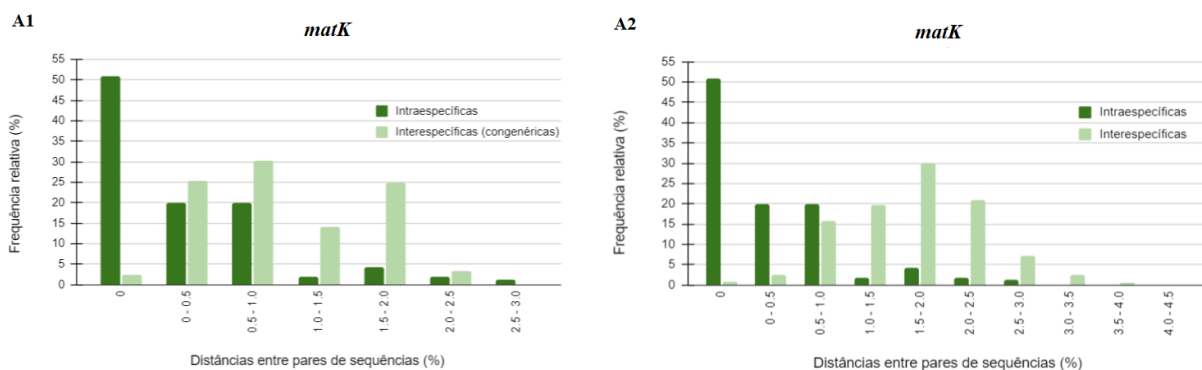
Barcode	Distância intraespecífica média (extensão)	Distância interespecífica média	
		Dentro dos gêneros (extensão)	Dentro da tribo (extensão)
ITS	0,01082 (0 - 0,05503)	0,03978 (0 - 0,08531)	0,05782 (0 - 0,11820)
<i>psbA-trnH</i>	0,00674 (0 - 0,06871)	0,02273 (0 - 0,06165)	0,04071 (0 - 0,10824)
<i>matK</i>	0,00352 (0 - 0,02546)	0,00993 (0 - 0,02329)	0,01669 (0 - 0,04372)

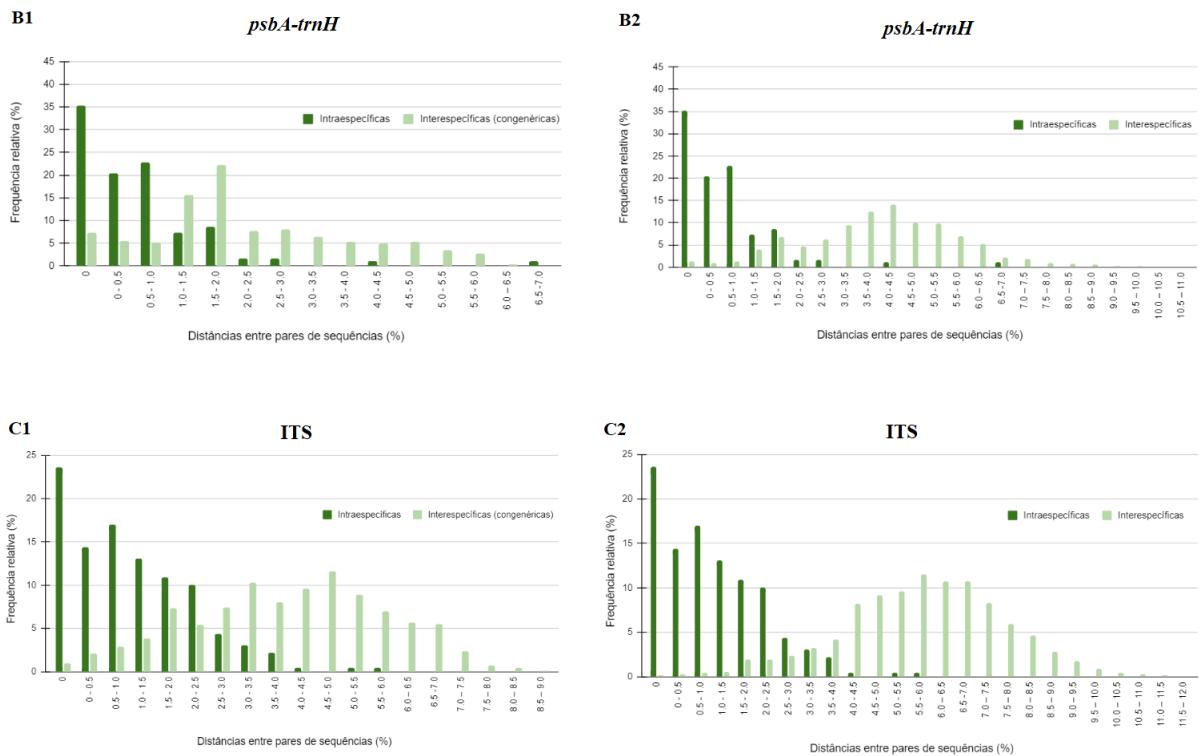
Fonte: elaborada pela autora (2022).

A distribuição das divergências intraespecíficas estende-se de 0 a 2,5% para o *matK*. Por outro lado, o intervalo observado para os ITS é maior, abrangendo entre 0 e 5,5%, de modo similar ao *psbA-trnH*, onde uma proporção reduzida de amostras possui quase 7% de divergência. No entanto, as distâncias com valor nulo (0%) são as mais frequentes para todos os marcadores, sendo que no caso do *matK* as mesmas correspondem a mais da metade dos dados (Figura 6). É interessante sinalizar, portanto, que a distribuição das variações intraespecíficas não segue uma curva do tipo normal ou gaussiana, de modo que a análise das medianas (*matK* = 0; *psbA-trnH* = 0,003 e ITS = 0,007) poderia ser mais informativa do que a das médias nesse caso, já que essa medida de tendência central possui uma menor sensibilidade aos valores extremos.

Quando as divergências entre espécies congêneras são analisadas, a amplitude de distribuição varia de 0 a 2,3% para o *matK* e 0 a 6,2% para o *psbA-trnH*, sobrepondo-se totalmente aos dados intraespecíficos encontrados para esses marcadores, enquanto a distribuição dos ITS é um pouco mais ampla (0 - 8,5%) (Figura 6-A1, B1, C1). Além disso, o *psbA-trnH* é o marcador com maior frequência de divergências interespecíficas nulas, que alcança 7,3% (FIGURA 6-B1). Ao considerar todas as espécies na análise, incluindo os gêneros com um único representante na biblioteca de referência, nota-se que o intervalo de distribuição das distâncias interespecíficas aumenta, variando de 0 a 4,4% para o *matK*, 0 a 10,8% para o *psbA-trnH* e 0 a 11,8% para os ITS, e que distâncias maiores passam a ser mais frequentes (Figura 6-A2, B2, C2). Os ITS também apresentam a menor sobreposição global entre as variações a nível intra e interespecífico, pois quase metade das distâncias entre pares de indivíduos de espécies diferentes ultrapassa o intervalo de distribuição observado para distâncias entre indivíduos coespecíficos (Figura 6-C2).

**Figura 6 - Histogramas com a frequência relativa das distâncias entre pares intra e interespecíficos para os marcadores (A) *matK*, (B) *psbA-trnH* e (C) ITS. As variações entre espécies foram consideradas a nível de gênero (A1, B1 e C1) e de tribo (A2, B2 e C2).**



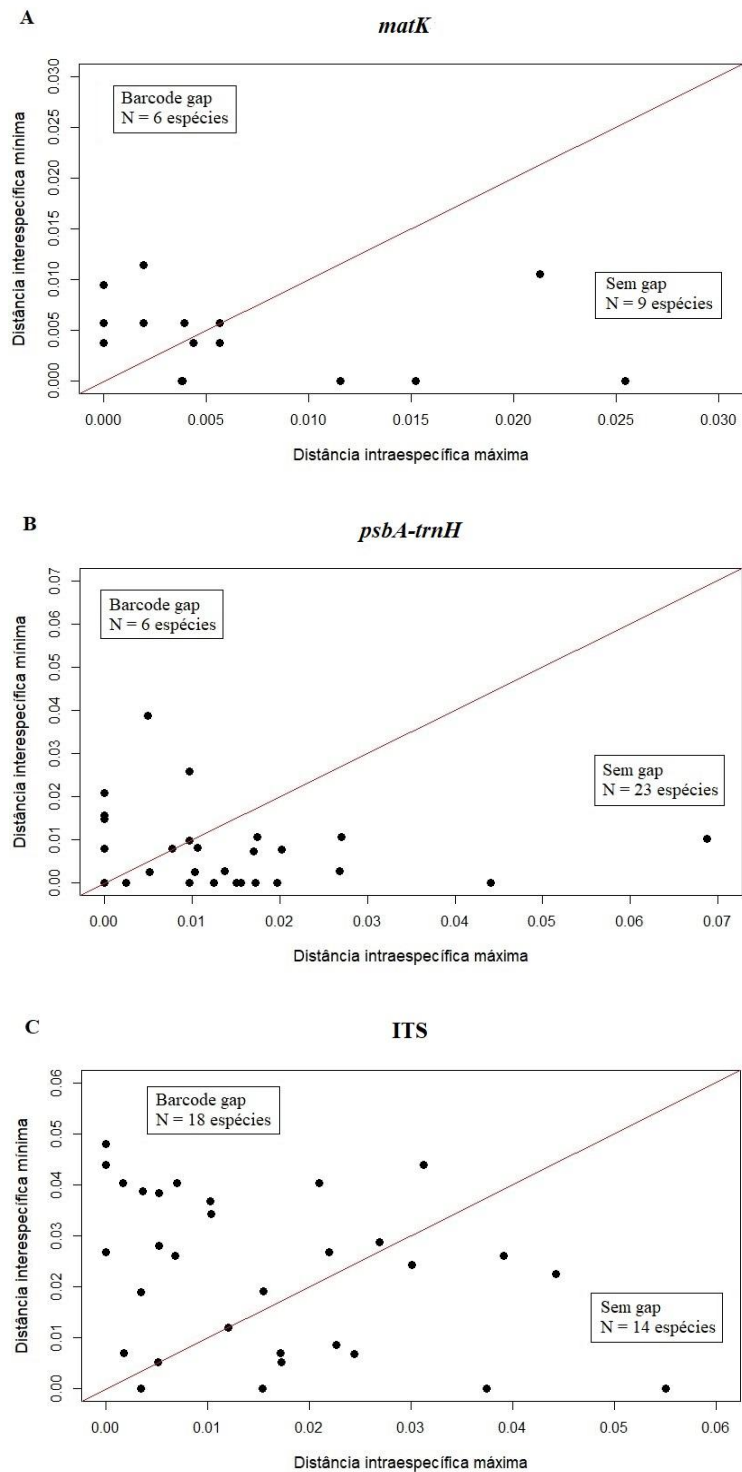


Fonte: elaborado pela autora (2022).

#### 4.2.2 Detecção dos *barcoding gaps*

Os gráficos comparando as divergências intraespecífica máxima e interespecífica mínima calculadas pelo modelo Kimura 2-parâmetros para cada espécie indicam uma baixa recorrência de *barcoding gaps* nos três marcadores analisados. O pior desempenho foi observado para o *psbA-trnH*, onde apenas 20,7% das espécies apresentaram lacunas, estando dispostas acima da reta com inclinação 1:1 (Figura 7-B). Esse índice sobe para 40% no que concerne ao *matK* (Figura 7-A) e 56,3% para os ITS (Figura 7-C), de modo que apenas nesse último caso foram constatadas descontinuidades satisfatórias entre os dois parâmetros para mais da metade das espécies investigadas. Além disso, 11 espécies (cerca de 38%) compartilharam sequências idênticas do *psbA-trnH* com ao menos uma outra espécie. Valores nulos de divergência interespecífica também foram encontrados em 5 espécies para o *matK* (33,3%) e ITS (15,6%). No Apêndice C, é possível consultar esses dados com maiores detalhes.

**Figura 7 - Gráficos comparando a distância intraespecífica máxima e a distância interespecífica mínima para os marcadores (A) *matK*, (B) *psbA-trnH* e (C) ITS.**



Fonte: elaborado pela autora (2022).

#### 4.2.3 Taxas de identificação por diferentes testes de similaridade

Os testes implementados no programa TAXONDNA apontam para índices de sucesso que variam de baixos a medianos, dependendo do nível de rigor adotado durante a



comparação das sequências (Tabela 3). Como todas as espécies foram contempladas com pelo menos três indivíduos distintos, a possibilidade de uma identificação incorreta pela ausência de *barcodes* coespecíficos na biblioteca é nula.

A região dos ITS apresentou os resultados mais promissores, com uma performance superior à dos demais marcadores nas três análises, atingindo 81,39% de amostras identificadas corretamente no teste de *Best Match*, 80,62% no *Best Close Match* e 64,34% no *All Species Barcodes*. O segundo melhor desempenho foi observado para o *matK*, com 72,22% de identificações satisfatórias para os dois primeiros testes e 52,77% na análise de ASB. O espaçador *psbA-trnH*, no entanto, atingiu níveis mais modestos de acertos, de modo que apenas 62,16% dos indivíduos foram atribuídos às suas espécies correspondentes no BM e BCM, e menos da metade das sequências (45,04%) foi corretamente discriminada no ASB. A proporção de amostras classificadas de modo equivocado também é superior no caso do *psbA-trnH*, variando de 0,9% a 12,61% (Tabela 3). Além disso, cabe destacar que a diminuição substancial do número de acertos na abordagem de ASB para todos os marcadores reflete o fato de que algumas sequências divergem mais do que outras dentro de um mesmo táxon. Assim, a identificação por meio desse critério já fica comprometida quando um único indivíduo de outra espécie é mais similar à *query* do que a amostra coespecífica mais distante, justificando a quantidade elevada de resultados ambíguos.

Os testes de BCM e ASB executados com os valores de corte obtidos no R foram mais conservativos (Tabela 4). Como os limites otimizados entre as divergências intra e interespecíficas nesse caso acabaram sendo bem inferiores àqueles calculados pela estratégia do TAXONDNA (Apêndice D), era esperado que a frequência de resultados ambíguos ou incorretos fosse reduzida, mas, em compensação, a capacidade dos métodos em identificar corretamente as sequências também diminuiu, enquanto a proporção de indivíduos que permaneceram indefinidos atingiu taxas de 14,72% a 18,91%. Além disso, é possível constatar que índices elevados de erro cumulativo permaneceram mesmo para os melhores limites, sendo de 37,5% (27 indivíduos) para o *matK*, 54,9% (61 indivíduos) para o *psbA-trnH* e 37,2% (48 indivíduos) para os ITS (Apêndice D). Problemas similares também foram relatados para animais, onde Meyer e Paulay (2005) alcançaram uma taxa mínima de 33% de erro total ao considerar espécies de gastrópodes marinhos circunscritas de forma tradicional.

**Tabela 3 - Performance de cada marcador com base nos testes de *Best Match*, *Best Close Match* e *All Species Barcodes*.**

Barcode	N	<i>Best Match</i> , N (%)			<i>Best Close Match</i> , N (%)				<i>All Species Barcodes</i> , N (%)				
		Correto	Ambíguo	Incorreto	Correto	Ambíguo	Incorreto	Sem match	Correto	Ambíguo	Incorreto	Sem match	Limite (%)
ITS	129	105 (81,39)	14 (10,85)	10 (7,75)	104 (80,62)	14 (10,85)	9 (6,97)	2 (1,55)	83 (64,34)	43 (33,33)	1 (0,77)	2 (1,55)	3,48
<i>psbA-trnH</i>	111	69 (62,16)	28 (25,22)	14 (12,61)	69 (62,16)	28 (25,22)	12 (10,81)	2 (1,80)	50 (45,04)	58 (52,25)	1 (0,90)	2 (1,8)	2,47
<i>matK</i>	72	52 (72,22)	20 (27,77)	0 (0,0)	52 (72,22)	20 (27,77)	0 (0,0)	0 (0,0)	38 (52,77)	34 (47,22)	0 (0,0)	0 (0,0)	1,71

Fonte: elaborada pela autora (2022).

**Tabela 4 - Performance de cada marcador com base nos testes de *Best Close Match* e *All Species Barcodes* considerando dois limites de corte.**

Barcode	N	<i>Best Close Match</i> , N (%)				<i>All Species Barcodes</i> , N (%)				
		Correto	Ambíguo	Incorreto	Sem match	Correto	Ambíguo	Incorreto	Sem match	Limite (%)
ITS	129	104 (80,62)	14 (10,85)	9 (6,97)	2 (1,55)	83 (64,34)	43 (33,33)	1 (0,77)	2 (1,55)	3,48 <sup>a</sup>
		89 (68,99)	14 (10,85)	7 (5,42)	19 (14,72)	75 (58,13)	34 (26,35)	1 (0,77)	19 (14,72)	1,10 <sup>b</sup>
<i>psbA-trnH</i>	111	69 (62,16)	28 (25,22)	12 (10,81)	2 (1,80)	50 (45,04)	58 (52,25)	1 (0,90)	2 (1,80)	2,47 <sup>a</sup>
		56 (50,45)	26 (23,42)	8 (7,2)	21 (18,91)	42 (37,83)	47 (42,34)	1 (0,90)	21 (18,91)	0,60 <sup>b</sup>
<i>matK</i>	72	52 (72,22)	20 (27,77)	0 (0,0)	0 (0,0)	38 (52,77)	34 (47,22)	0 (0,0)	0 (0,0)	1,71 <sup>a</sup>
		45 (62,50)	16 (22,22)	0 (0,0)	11 (15,27)	34 (47,22)	27 (37,50)	0 (0,0)	11 (15,27)	0,20 <sup>b</sup>

Fonte: elaborado pela autora (2022).

Legenda: <sup>a</sup> = limites calculados pelo TAXONDNA (MEIER *et al.*, 2006); <sup>b</sup> = limites calculados pela função threshOpt no R (BROWN *et al.*, 2012).

Esse desfecho é corroborado pela existência de sobreposição considerável entre as distribuições das distâncias genéticas nos histogramas da Figura 6, sugerindo que não há um valor de corte universal, capaz de segregar a maior parte das sequências coespecíficas daquelas oriundas de diferentes espécies, ao menos no contexto de grupos biológicos altamente diversificados e que ainda carecem de mais estudos (MEYER; PAULAY, 2005). Ao avaliar 117 espécies de dípteros representadas por múltiplas sequências no GenBank, Meier *et al.* (2006) já haviam constatado que as distâncias entre trios de indivíduos coespecíficos não costumam ser equilaterais, fazendo com que a tentativa de adotar um valor de corte fixo produza agrupamentos que não são consistentes com boa parte das espécies atualmente aceitas, o que demandaria a redescritção de diversos taxa (MEIER *et al.*, 2006).

#### 4.2.4 Agrupamentos produzidos pelas reconstruções filogenéticas

A performance dos marcadores via reconstruções filogenéticas foi avaliada a partir do número de espécies e também de indivíduos/sequências identificados corretamente (Tabela 5). As topologias geradas através dos métodos de *neighbor-joining*, máxima verossimilhança e inferência Bayesiana apresentaram alguns detalhes distintos, mas, em geral, revelaram os mesmos agrupamentos espécie-específicos, apesar dos valores de suporte divergirem em determinadas situações. Além disso, os padrões de ramificação não demonstraram ser capazes de diferenciar as variedades dos demais indivíduos coespecíficos nos casos em que essa categoria taxonômica foi contemplada, conforme pode ser consultado no Apêndice E.

Os resultados para o *matK* e para os ITS foram muito próximos nas três abordagens. No que concerne ao *matK*, a maior diferença observada foi o fato de *Psidium cattleianum* não aparecer como um grupo monofilético na análise de NJ, enquanto as sequências de *Blepharocalyx salicifolius* foram reunidas em um clado somente nessa mesma metodologia. No entanto, ambas as espécies não receberam valores adequados de suporte quando agrupadas (*bootstrap* < 70% ou probabilidade posterior (PP) < 0,95), não tendo sido contabilizadas, portanto, como identificações satisfatórias. *Pimenta racemosa* formou um clado exclusivo, porém com suporte insuficiente em todas as topologias, enquanto alguns indivíduos de *Myrcia amazonica* não foram agrupados na análise de ML e a espécie não recebeu suporte nos demais casos. Já *Myrciaria floribunda* apresentou monofilétismo através dos três métodos, mas foi considerada significativa apenas na ML e IB.

Em relação aos ITS, é possível observar uma espécie correta a menos na análise de ML em detrimento das demais devido às sequências de *Myrceugenia alpigena*, cujo clado foi

desconsiderado por exibir um valor de *bootstrap* de apenas 67%. Outras diferenças de topologia encontradas incluem o fato de *Psidium laruotteanum* não ter sido recuperado na árvore de NJ e *B. salicifolius* não aparecer como um grupo monofilético somente na ML, embora essas espécies não tenham alcançado valores consideráveis de suporte em nenhum caso. Já o marcador *psbA-trnH* foi o que obteve os resultados mais variáveis de acordo com a metodologia. 10 espécies foram recuperadas em comum, embora o clado composto pelas sequências de *M. amazonica* não tenha recebido suporte suficiente. Por outro lado, o agrupamento de *Eugenia involucrata* foi considerado significativo apenas na ML e IB, enquanto um valor de suporte alto foi atribuído a *E. brevistyla* e *Pimenta dioica* somente na IB (PP > 0,99). A espécie *Myrceugenia ovata* também foi corretamente discriminada na IB, mas não recebeu suporte elevado na ML e nem chegou a ser recuperada como um clado na topologia de NJ. Além disso, *Myrrhinium atropurpureum* apareceu como um agrupamento espécie-específico exclusivamente na análise de IB, porém com um valor baixo de suporte (PP = 0,84).

Esses resultados estão sujeitos a algumas alterações de acordo com os critérios adotados, como a calibração dos parâmetros em cada programa e a estipulação de limites adequados de *bootstrap* e probabilidade posterior. Nota-se, no entanto, que a performance de cada marcador segue um padrão similar àquele encontrado nos testes baseados em distâncias genéticas, com o *psbA-trnH* apresentando os índices mais insatisfatórios de espécies identificadas corretamente, que variam de 20,69% para a árvore de NJ até 34,48% no caso da IB. O gene *matK* ocupa uma posição intermediária, de modo que 40% das taxa chegaram a formar clados nas topologias de ML e IB. Todavia, é válido ressaltar que um número menor de espécies foi utilizado para essa região. *Myrcia citrifolia*, *M. guianensis*, *M. rufipes* e *M. variabilis*, que não chegaram a manifestar *barcoding gaps* (Apêndice C) ou formar agrupamentos monofiléticos (Apêndice E) nos demais marcadores, por exemplo, são espécies que não foram contempladas para o *matK*, de modo que a performance desse gene poderia estar sendo sub ou superestimada, dependendo do comportamento de tais grupos. Já os ITS demonstraram novamente o desfecho mais promissor, onde 65,63% a 68,75% das espécies agregaram-se de forma exclusiva.

Ao observar a tabela 5, é possível constatar que a inferência Bayesiana destacou-se apenas no caso do *psbA-trnH*. Tal método de reconstrução filogenética pode, portanto, ter sido capaz de lidar melhor com ambiguidades ou divergências recentes dentro desse marcador, ao passo que as resoluções dos ITS e do *matK* mantiveram-se aproximadamente constantes, independentemente do tipo de filogenia adotada. Ainda assim, é interessante observar que o

número de indivíduos identificados corretamente é mais comedido nos métodos amparados em filogenias se comparado às análises de BM e BCM, principalmente para as regiões plastidiais (Tabela 3). Isso já foi relatado em outros trabalhos (ALVES, 2013; GIUDICELLI; MÄDER; DE FREITAS, 2015) e pode estar associado ao fato do teste de BCM ser capaz de realizar identificações corretas mesmo sob presença de parafiletismo (COLLINS; CRUICKSHANK, 2013). Ademais, o fato de duas amostras de *Eugenia monticola* (Emon07 e Emon08), sequenciadas pela equipe do iBOL somente para o *matK*, estarem inseridas no mesmo clado de *Myrcia splendens* em todas as topologias geradas com esse marcador é inesperado. De acordo com as informações do BOLD, esses indivíduos e alguns espécimes de *M. splendens* foram coletados no mesmo local, de modo que a possibilidade de uma inserção incorreta dos dados no sistema ou um erro de identificação não deve ser descartada.

**Tabela 5 - Número de espécies e indivíduos identificados corretamente através de três metodologias de reconstrução filogenética.**

Barcode	Nº de espécies (Nº de indivíduos)	Nº de espécies identificadas corretamente (%)			Nº de indivíduos identificados corretamente (%)		
		NJ	ML	IB	NJ	ML	IB
ITS	32 (129)	22 (68,75)	21 (65,63)	22 (68,75)	91 (70,54)	88 (68,22)	91 (70,54)
<i>psbA-trnH</i>	29 (111)	6 (20,69)	7 (24,14)	10 (34,48)	26 (23,42)	30 (27,03)	39 (35,14)
<i>matK</i>	15 (72)	5 (33,33)	6 (40)	6 (40)	27 (37,50)	30 (41,67)	30 (41,67)

Fonte: elaborada pela autora (2022).

Legenda: NJ = *neighbor-joining*; ML = máxima verossimilhança; IB = inferência Bayesiana.

## 5 DISCUSSÃO

### 5.1 Diferenças entre regiões plastidiais e nucleares

Os estudos dedicados à avaliação da técnica de *DNA barcoding* em Myrtaceae ainda são muito escassos, conforme constatado por meio do levantamento bibliográfico realizado. Mesmo assim, algumas pesquisas têm sido fundamentais para o incremento de novas sequências nucleotídicas nos repositórios digitais, possibilitando uma análise mais refinada da eficácia de diferentes marcadores nessa família. Buys *et al.* (2016), por exemplo, construíram uma base de dados contendo informações relativas a mais de 100 espécies presentes na Nova

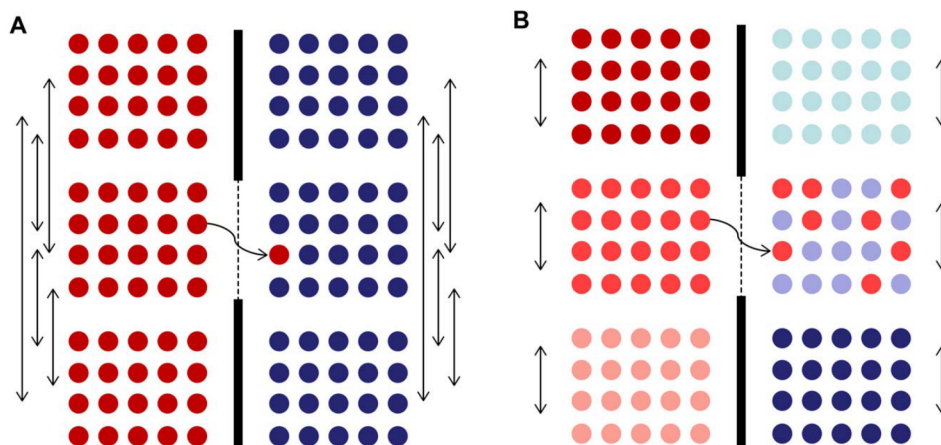
Zelândia, tendo comparado a performance dos marcadores ITS, ETS e *matK* e verificado que o primeiro apresentou as melhores taxas individuais de sucesso na identificação das amostras através das metodologias de BLAST1 e *Nearest Distance* (equivalente ao *Best Match*). Já Wati *et al.* (2022) recomendaram a região dos ITS como *barcode* para a delimitação de espécies de *Syzygium* na ilha de Sumatra (Indonésia) a partir de análises de *barcoding gap*, BLAST e inferência Bayesiana, sugerindo que a combinação *rbcL+matK* seja empregada de forma suplementar nesse gênero. Em ambos os casos, no entanto, o número de espécies de Myrteae sequenciadas foi muito baixo (13 e 3, respectivamente). Parra-O e Bohórquez-Osório (2016) também utilizaram as regiões do *matK*, *rbcL* e ITS para confirmar o gênero de uma nova espécie colombiana de Myrtaceae com caracteres florais incomuns (*Myrcianthes roncesvallensis*), integrando dados morfológicos e moleculares em uma aplicação prática.

Os diversos testes implementados no presente trabalho sugerem que os marcadores *matK*, *psbA-trnH* e ITS diferem quanto à capacidade de discriminar indivíduos a nível de espécie na tribo Myrteae, com o *locus* nuclear apresentando os resultados mais promissores. A baixa performance dos *barcodes* plastidiais já foi constatada em outros gêneros de plantas mono e eudicotiledôneas com taxonomias complexas, incluindo *Sisyrinchium* (Iridaceae) (ALVES *et al.*, 2014) e *Passiflora* (Passifloraceae) (GIUDICELLI; MÄDER; DE FREITAS, 2015). Os ITS ou o ITS2, por sua vez, são considerados alguns dos marcadores mais úteis em estudos filogenéticos devido à presença simultânea em diferentes grupos de eucariotos e à baixa restrição funcional, o que favorece uma suscetibilidade a rápidas mudanças evolutivas (VIJAYAN; TSOU, 2010). Seu poder discriminatório normalmente superior ao dos *loci* plastidiais em categorias taxonômicas mais específicas é um aspecto bem estabelecido na sistemática molecular (HOLLINGSWORTH; GRAHAM; LITTLE, 2011), o que fez desses espaçadores regiões recomendadas como *barcodes* para famílias botânicas como Asteraceae (GAO *et al.*, 2010), Rosaceae (PANG *et al.*, 2011) e Rutaceae (LUO *et al.*, 2010).

Uma das hipóteses para explicar tais discrepâncias pode residir nas próprias particularidades de dispersão das regiões plastidiais e nucleares. Em angiospermas, as variantes plastidiais tendem a ser adquiridas do parental feminino, em um sistema de herança não-mendeliano, ainda que uma potencial transmissão biparental tenha sido relatada em cerca de 20% dos gêneros (ZHANG; SODMERGEN, 2010). Para boa parte das espécies, o pólen percorre distâncias muito mais amplas quando comparado ao deslocamento relativamente restrito das sementes, sendo que o material genético nuclear está presente em ambas as estruturas, mas o plastidial apenas nessa última (HOLLINGSWORTH; GRAHAM; LITTLE, 2011). Assim, marcadores com herança materna normalmente apresentam uma diferenciação

a nível populacional consideravelmente maior do que regiões provenientes do núcleo devido ao fluxo gênico reduzido, o que aumenta a permeabilidade das espécies à introgressão de alelos heteroespecíficos em decorrência da menor competição local contra essas variantes (HOLLINGSWORTH; GRAHAM; LITTLE, 2011). Como consequência, tais grupos acabam sendo menos propícios a exibir *barcodes* característicos (Figura 8). Regiões do genoma plastidial costumam ser utilizadas em estudos de filogeografia justamente pela sua habilidade de reconstrução de eventos históricos e padrões demográficos em diferentes populações. O trabalho de Turchetto-Zolet *et al.* (2016) com *Eugenia uniflora* (Myrtaceae), por exemplo, empregou espaçadores intergênicos do cloroplasto para evidenciar a ocorrência de duas linhagens divergentes principais dentro da espécie, uma no norte e outra no sul do Domínio da Mata Atlântica, de modo que a baixa taxa de migração observada entre essas populações provavelmente está atrelada à presença de barreiras fitogeográficas que impedem a dispersão zoocórica das sementes (TURCHETTO-ZOLET *et al.*, 2016). A limitação de fluxo gênico é, portanto, um dos possíveis fatores associados ao pior desempenho do *psbA-trnH* como *barcode* nos testes realizados com as espécies de Myrteae.

**Figura 8 - Representação esquemática do possível impacto exercido pelo fluxo gênico entre populações sobre o sucesso discriminatório de um *barcode*. Em A, as três populações da espécie vermelha estão conectadas por uma intensa troca de material genético (indicada pelas setas verticais alongadas), permanecendo relativamente isoladas da espécie azul, e vice-versa. Em B, o fluxo gênico entre populações está comprometido (setas verticais curtas), tornando-as mais diferenciadas entre si e suscetíveis à introgressão de alelos de outra espécie (seta inclinada), o que reduz o poder da técnica de *DNA barcoding*.**



Fonte: Hollingsworth, Graham e Little (2011).

Alguns trabalhos sobre *DNA barcoding* realizados com espécies que ocorrem no território brasileiro ou na região Neotropical podem ser consultados no Quadro 1. Os resultados que servem como embasamento para as recomendações em cada táxon por vezes são gerados através de abordagens analíticas distintas, fazendo com que esses detalhes

metodológicos acabem interferindo em uma comparação direta dos dados. Mesmo assim, é interessante notar que os ITS são indicados de forma recorrente quando testados, seja como marcadores individuais ou como candidatos potencialmente informativos para uso em conjunto com regiões suplementares. Gonzalez *et al.* (2009) não preconizam esses espaçadores nucleares como principais alternativas de *barcode* devido ao índice limitado de sequenciamento obtido para espécies tropicais da floresta Amazônica (41%), embora admitam que sua alta variabilidade pode ser muito oportuna na delimitação de famílias que são particularmente complexas e desafiadoras para as regiões plastidiais, a exemplo de Sapotaceae.

Vivas *et al.* (2014) demonstraram que os ITS realmente apresentam uma boa capacidade discriminatória em espécies de Sapotaceae na Mata Atlântica brasileira, superando os testes com combinações de diferentes marcadores. De maneira similar, Bolson *et al.* (2015) encontraram resultados positivos para esses espaçadores ao analisar espécies madeireiras ameaçadas no estado do Paraná (RS), incluindo alguns representantes do gênero *Ocotea* (Lauraceae), outro grupo diverso e de difícil taxonomia. Os autores, no entanto, ressaltam que as taxas relativamente baixas de sucesso durante o sequenciamento (78%) consistem na única inconveniência associada a essa região. Índices ainda menores de universalidade (39%) foram encontrados para os ITS em membros sul americanos da tribo Tigridieae (Iridaceae) (ALVES, 2013), demonstrando que a recuperação de regiões do núcleo não é tão simples em alguns casos devido à presença de poliploidia, elementos de transposição, heterozigiosidade e mutações nos sítios de *priming*, por exemplo (HOLLINGSWORTH; GRAHAM; LITTLE, 2011; SPOONER, 2009). Ainda assim, um estudo com maior abrangência em Lauraceae, que incorporou a prospecção de marcadores táxon-específicos do cloroplasto, também apontou os ITS como candidatos com a melhor resolução para a família (38,9% a 66,7%), sugerindo que novas opções nucleares sejam investigadas em detrimento das plastidiais (LIU *et al.*, 2022).

**Quadro 1 - Relação de alguns estudos sobre *DNA barcoding* que incluem espécies nativas do Brasil ou da região Neotropical, com os marcadores utilizados, metodologias e considerações dos autores.**

(continua)

Grupo botânico	Marcadores analisados	Principais estratégias de análise discriminatória	Recomendações	Referência
86 espécies de angiospermas para todos os <i>loci</i> + 1084 espécies de Orchidaceae restritas ao <i>matK</i>	<i>rbcL</i> , <i>matK</i> , <i>psbA-trnH</i> , <i>ycf5</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>accD</i> , <i>ndhJ</i>	Avaliação dos <i>barcoding gaps</i> via histogramas; reconstruções filogenéticas pelos métodos de MP, ML, NJ, UPGMA e IB; análises de coalescência (agrupamentos genéticos)	<b><i>matK</i></b> , com o <b><i>psbA-trnH</i></b> sendo um marcador alternativo ou complementar	Lahaye <i>et al.</i> (2008)



**Quadro 1 - Relação de alguns estudos sobre *DNA barcoding* que incluem espécies nativas do Brasil ou da região Neotropical, com os marcadores utilizados, metodologias e considerações dos autores.**

(continuação)

Grupo botânico	Marcadores analisados	Principais estratégias de análise discriminatória	Recomendações	Referência
Myristicaceae: <i>Compsoeura</i> (8 espécies), <i>Iryanthera</i> (1) e <i>Virola</i> (1)	UPA, <i>rpoB</i> , <i>rpoC1</i> , <i>accD</i> , <i>rbcL</i> , <i>matK</i> , <i>psbA-trnH</i>	Escalonamento multidimensional não métrico; histogramas com as distâncias intra e interespecíficas sem correção; TAXONDNA (BM e BCM)	<b><i>matK</i> + <i>psbA-trnH</i></b>	Newmaster <i>et al.</i> (2008)
54 famílias de angiospermas da Amazônia	<i>rbcL</i> , <i>rpoC1</i> , <i>rpoB</i> , <i>matK</i> , <i>ycf5</i> , <i>trnL</i> , <i>psbA-trnH</i> , ITS	TAXONDNA (BCM); ferramenta <i>cluster</i> do TAXONDNA e método de <i>blastclust</i> para agrupamento das sequências em MOTUs; reconstruções filogenéticas (ML e NJ)	<b><i>psbA-trnH</i></b> , embora nenhum marcador plastidial tenha alcançado taxas de sucesso superiores a 70%	Gonzalez <i>et al.</i> (2009)
Fabaceae ( <i>Inga</i> ): 26 espécies + 2 famílias com enfoque fora dos Neotrópicos (Aytoniaceae e Araucariaceae)	<i>rpoC1</i> , <i>rpoB</i> , <i>rbcL</i> , <i>matK</i> , <i>psbA-trnH</i> , <i>atpF-atpH</i> , <i>psbK-psbI</i>	Gráficos comparando os níveis de divergência intragenérica para cada par de <i>loci</i> ; histogramas com a distribuição das distâncias genéticas (K2P); filogenias de MP, NJ e UPGMA	<b><i>rpoC1</i> + <i>rbcL</i> + <i>matK</i> ou <i>rbcL</i> + <i>matK</i> + <i>psbA-trnH</i> (<i>Inga</i>)</b> = discriminação de até 69% das espécies	Hollingsworth <i>et al.</i> (2009)
Bromeliaceae: 46 espécies	<i>rbcL</i> , <i>matK</i> , <i>psbA-trnH</i>	Teste de similaridade (BLAST); distribuição das divergências interespecíficas (K2P); análises filogenéticas (MP, ML e NJ)	<b><i>rbcL</i> + <i>matK</i> + <i>psbA-trnH</i></b> apresentaram o melhor resultado, mas insatisfatório (44,44% das espécies)	Maia <i>et al.</i> (2012)
Iridaceae: 7 gêneros e 17 espécies da tribo Tigridieae	ITS, <i>rbcL</i> , <i>matK</i> , <i>psbA-trnH</i>	Histogramas com as divergências intra e interespecíficas (K2P); TAXONDNA (BM e BCM); filogenias (MP, NJ, UPGMA)	<b><i>matK</i></b> (ITS tiveram baixa universalidade: 39,06%)	Alves (2013)
Iridaceae: 88 espécies de <i>Sisyrinchium</i> + 10 espécies de outros gêneros	ITS, <i>matK</i> , <i>psbA-trnH</i>	Avaliação dos <i>barcoding gaps</i> via distribuição das distâncias intra e interespecíficas (K2P); TAXONDNA (BM e BCM); filogenias (MP, NJ, UPGMA e IB)	<b>ITS como ponto inicial</b> , com 30,61% a 38,78% das espécies identificadas	Alves <i>et al.</i> (2014)
Sapotaceae: 26 espécies	<i>rbcL</i> , <i>matK</i> , <i>psbA-trnH</i> , ITS	Distribuição relativa das distâncias intra e interespecíficas (K2P); TAXONDNA (BM e BCM); filogenia de NJ	<b>ITS</b>	Vivas <i>et al.</i> (2014)
30 espécies e 17 famílias de angiospermas do Paraná	<i>rbcL</i> , <i>matK</i> , <i>psbA-trnH</i> , ITS	Histogramas com as frequências das distâncias genéticas sem correção; TAXONDNA (BCM); abordagens filogenéticas pelos métodos de MP, ML e NJ	<b>ITS</b> , com o <b><i>psbA-trnH</i></b> como <i>locus</i> adicional	Bolson <i>et al.</i> (2015)

**Quadro 1 - Relação de alguns estudos sobre *DNA barcoding* que incluem espécies nativas do Brasil ou da região Neotropical, com os marcadores utilizados, metodologias e considerações dos autores.**  
(conclusão)

Grupo botânico	Marcadores analisados	Principais estratégias de análise discriminatória	Recomendações	Referência
<i>Passiflora</i> (Passifloraceae): 222 espécies	<i>rbcL</i> , <i>matK</i> , <i>psbA-trnH</i> , <i>trnL</i> , ITS1, ITS2, ITS1+2	Frequência relativa das distâncias intra e interespecíficas (K2P); TAXONDNA (BM e BCM); filogenias (ML, NJ e IB)	<b>ITS1 como ponto inicial</b> (> 60% de acertos via BM e BCM)	Giudicelli, Mäder e de Freitas (2015)
Arecaceae: 4 espécies coletadas no estudo + várias espécies de 5 subfamílias de Palmae (GenBank)	<i>rps16-trnK</i> , <i>rpl32-trnL</i> , <i>trnL-trnF</i> , ITS2	Cálculo das divergências intra e interespecíficas médias (K2P) e da distância do <i>Nearest Neighbor</i> para cada espécie no BOLD; construção de árvores de NJ	<b><i>rpl32-trnL</i> ou concatenações dos três marcadores plastidiais</b>	Todeschini (2019)
<i>Piper</i> (Piperaceae): <i>P. gaudichaudianum</i> , <i>P. malacophyllum</i> e <i>P. regnellii</i>	ITS2, <i>psbA-trnH</i> , <i>rbcL</i> + testagem de 3 primers para amplificar microssatélites	Avaliação dos <i>barcoding gaps</i> através de histogramas das distâncias genéticas sem correção; teste de identificação via BLASTn; análises filogenéticas por MP e UPGMA	<b>ITS2</b>	Brandão <i>et al.</i> (2020)
<i>Euterpe</i> (Arecaceae): <i>E. edulis</i> , <i>E. oleracea</i> e <i>E. precatória</i>	<i>rbcL</i> , <i>matK</i> , <i>psbA-trnH</i> , <i>ycf1b</i> , <i>trnL</i> , <i>trnL-trnF</i> , <i>psbK-psbI</i> , ETS, ITS2	Inspeção do número de sítios polimórficos dentro de cada espécie e entre espécies diferentes; reconstrução filogenética pelo método de IB	<b><i>psbK-I</i></b>	Lugon <i>et al.</i> (2021)
538 espécies e 115 famílias de plantas vasculares na Serra dos Carajás (PA)	<i>matK</i> , <i>rbcL</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>atpF-atpH</i> , <i>psbK-psbI</i> , <i>psbA-trnH</i> , ITS2	Análises de BLASTn; árvores de ML	<b><i>rbcL</i> + ITS2</b>	Vasconcelos <i>et al.</i> (2021)
Myrtaceae: 10 gêneros e 32 espécies da tribo Myrteae	<i>matK</i> , <i>psbA-trnH</i> , ITS	Verificação dos <i>barcoding gaps</i> a partir das distâncias intraespecífica máxima e interespecífica mínima (K2P) para cada espécie; TAXONDNA (BM, BCM e ASB); filogenias (ML, NJ e IB)	<b>ITS como ponto inicial</b>	Este trabalho (2022)

Fonte: elaborado pela autora (2022).

Legenda: ASB = *All Species Barcodes*; BC = *Best Match*; BCM = *Best Close Match*; BOLD = *Barcode of Life Data System*; IB = inferência Bayesiana; K2P = Kimura 2-parâmetros; ML = máxima verossimilhança; MOTU = unidade taxonômica operacional molecular (*molecular operational taxonomic unit*); MP = máxima parcimônia; NJ = *neighbor-joining*; UPGMA = método de agrupamento não ponderado com médias aritméticas (*unweighted pair group method with arithmetic mean*).

Na contramão dessas estatísticas, o *matK* demonstra ser um marcador filogenético e um *barcode* mais eficiente para *Carex* seção *Phyllostachyae* (Cyperaceae) do que os ITS,

possivelmente em razão desses últimos estarem evoluindo tão rápido a ponto de criar homoplasias em nível infraseccional (STARR; NACZI; CHOUINARD, 2009). Mesmo que as taxas de espécies resolvidas nesse gênero com o uso do gene plastidial não sejam elevadas (até 60%), Starr, Naczi e Chouinard (2009) propõem que esse *locus* faça parte do sistema de *barcodes* universais para o grupo e seja sequenciado em amostras com origens previamente conhecidas, o que pode auxiliar na melhora da performance da técnica. Regiões do cloroplasto também já apresentaram resultados promissores em alguns levantamentos florísticos, incluindo análises com orquídeas da Mesoamérica (Orchidaceae), onde o *matK* foi capaz de revelar espécies crípticas quando empregado individualmente (LAHAYE *et al.*, 2008). Por fim, a combinação desse *locus* com o *psbA-trnH* revelou-se útil para a delimitação de espécies de *Compsoeura* (Myristicaceae) coletadas em diferentes países da América do Sul e Central, considerando que a identificação morfológica desses organismos depende fortemente da presença de estruturas reprodutivas (NEWMASER *et al.*, 2008). É importante destacar, no entanto, que esses estudos costumam envolver amostragens a nível regional ou um conjunto relativamente limitado de espécies-irmãs. Assim, a relevância dos marcadores plastidiais não deve ser ignorada, embora sua eficácia varie de acordo com os taxa inspecionados (ALVES, 2013) e possa diminuir consideravelmente quando muitos membros próximos são incluídos (STARR; NACZI; CHOUINARD, 2009).

## 5.2 Razões que podem comprometer o uso do *DNA barcoding* em famílias botânicas

Inúmeros fatores são capazes de prejudicar a estratégia de identificação a nível de espécie baseada em *barcodes* de DNA, sendo um deles a própria história evolutiva dos grupos taxonômicos. O ritmo de especiação acaba interferindo no acúmulo de diferenças espécie-específicas, fazendo com que um sistema de diagnóstico genético se mostre desfavorável em plantas que apresentam taxas muito lentas de substituição e, conseqüentemente, baixa variabilidade nas suas sequências nucleotídicas, como algumas gimnospermas do gênero *Araucaria* (Araucariaceae) (HOLLINGSWORTH *et al.*, 2009). A evolução molecular nos maiores clados de angiospermas também é mais vagarosa em árvores e arbustos quando comparada a linhagens herbáceas, sugerindo que espécies lenhosas, caracterizadas pela existência de gerações que perduram por longos períodos, possam ser particularmente problemáticas (GONZALEZ *et al.*, 2009; SMITH; DONOGHUE, 2008).

As taxas de diversificação na região Neotropical, todavia, costumam ser altas, o que gera eventos frequentes de rápida radiação (RICHARDSON *et al.*, 2001). Dificuldades para a

implementação do *DNA barcoding* são igualmente verificadas nessas situações, com relatos de insucesso na família Bromeliaceae, cujas subfamílias começaram a divergir a apenas 19 milhões de anos (MAIA *et al.*, 2012), no gênero *Inga* (Fabaceae), com especiações predominando nos últimos 10 milhões de anos (HOLLINGSWORTH *et al.*, 2009; RICHARDSON *et al.*, 2001), e na seção *Solanum* sect. *Petota* (Solanaceae) (SPOONER, 2009). Não é surpreendente, portanto, que desafios similares sejam encontrados em pesquisas aplicadas aos membros Neotropicais da tribo Myrteae (Myrtaceae), levando em conta que um período de radiação relativamente recente é estimado para esse grupo (Oligoceno - Mioceno) e sua grande riqueza de espécies provavelmente está atrelada a taxas aceleradas de diversificação em algumas linhagens (BIFFIN *et al.*, 2010; LIMA *et al.*, 2021). Nesse caso, o emprego exclusivo de marcadores moleculares que evoluem de forma mais lenta, como sequências codificadoras de proteínas, pode fornecer informações insuficientes para a resolução de filogenias (BALBINOTT *et al.*, 2022) e também para uso como *barcode* em plantas. As análises conduzidas nesse trabalho indicam que o gene *matK* de fato possui o menor número de sítios polimórficos parcimoniosamente informativos dentre os marcadores selecionados, mesmo sendo conhecido como um dos *loci* plastidiais codificantes de mais rápida evolução (VIJAYAN; TSOU, 2010).

Evidências de parafiletismo em árvores de genes também aparentam ser mais comuns nas plantas do que em animais, o que acaba fomentando um debate acerca da existência de fronteiras genéticas ambíguas entre espécies vegetais (FAZEKAS *et al.*, 2008, 2009). Esse é um dos problemas associados ao uso do *bootstrap* como uma métrica de validação de *barcodes* em filogenias, visto que o suporte de um grupo monofilético é mensurado a partir dos dados disponibilizados, mas desconsidera a existência de outras sequências com monofiletismo possivelmente interrompido, principalmente nos casos em que não foi feita uma amostragem minuciosa de indivíduos e espécies (FAZEKAS *et al.*, 2008). Além disso, ramos mais curtos com espécies-irmãs que divergiram recentemente costumam apresentar valores baixos de suporte, mesmo que em alguns casos seja possível diferenciar esses grupos por mutações únicas (COLLINS; CRUICKSHANK, 2013).

O parafiletismo tem múltiplas causas, podendo, por exemplo, ser decorrente de um evento conhecido como classificação de linhagem incompleta (*incomplete lineage sorting*) ou coalescência profunda, onde os conjuntos alélicos (haplótipos) herdados em populações com tamanho efetivo grande e que sofreram especiação rápida e recente costumam produzir árvores de genes incompatíveis com as árvores de espécies devido à retenção de polimorfismos ancestrais, afetando a habilidade discriminatória dos *barcodes* (FAZEKAS *et*

*al.*, 2009; FUNK; OMLAND, 2003; MADDISON, 1997). Nesse sentido, espécies que se originaram por auto ou aloploidia também estão sujeitas, ao menos inicialmente, a compartilhar haplótipos plastidiais com seus respectivos progenitores (HOLLINGSWORTH; GRAHAM; LITTLE, 2011). Ademais, Fazekas *et al.* (2009) já constataram que gêneros botânicos com evidências de hibridação natural possuem um grau menor de descontinuidade genética, o que torna esse parâmetro um preditor significativo da magnitude dos *barcoding gaps* nesses grupos.

A dificuldade em identificar os membros Neotropicais de Myrtaceae provavelmente está associada à hibridação atuando em conjunto com a poliploidia para gerar novos caracteres intermediários entre espécies muito próximas, de modo que a diferenciação a nível cromossômico acaba isolando esses indivíduos geneticamente e favorecendo o processo de especiação (DA COSTA; FORNI-MARTINS, 2006a, 2006b). Séries poliploides são frequentes na subfamília Myrtoideae, tendo sido reportadas em cerca de 75% das espécies de *Psidium* examinadas em estudos citogenéticos, além de representantes de *Eugenia*, *Gomidesia* e *Myrcia* (DA COSTA, 2004). Esses dados são, portanto, um indicativo da importância da poliploidia para a evolução da tribo Myrteae (DA COSTA; FORNI-MARTINS, 2006a). Somado a isso, a alta prevalência de metabólitos secundários em espécies Neotropicais também complexifica procedimentos como a extração de DNA, fundamental para as etapas subsequentes das análises de *barcode* (GONZALEZ *et al.*, 2009).

A adesão dos taxonomistas a diferentes conceitos de espécie ao longo do tempo é outro elemento que pode ter contribuído para as atuais incongruências encontradas em alguns grupos botânicos (SPOONER, 2009). Limites excessivamente inclusivos (*lumping*) ou restritivos (*splitting*) são estabelecidos para a delimitação de uma espécie dependendo dos critérios considerados, como morfologia, ecologia ou grau de fluxo gênico e isolamento reprodutivo, produzindo taxa que nem sempre correspondem adequadamente às variações existentes a nível genético (FAZEKAS *et al.*, 2009). Além disso, o uso de um ou poucos marcadores por vezes é incapaz de diferenciar grupos taxonomicamente complexos, cujas fronteiras entre espécies não são claramente definidas devido a fatores como presença de ecótipos e microespécies originadas por apomixia (HOLLINGSWORTH; GRAHAM; LITTLE, 2011). *Myrcia splendens* é uma das espécies da flora Neotropical mais desafiadora nesse aspecto, pois abrange cerca de 170 sinônimas e uma série de morfotipos distribuídos em uma ampla área geográfica, que inclui a Mata Atlântica, Amazônia e Cerrado brasileiros, além da Guiana Francesa, Costa Rica e República Dominicana (LIMA DOS SANTOS *et al.*, 2021). O não-monofiletismo desse complexo foi confirmado por Lima dos Santos *et al.*

(2021), com múltiplos acessos emergindo em clados independentes de acordo com os biomas de procedência. Tal instabilidade taxonômica estende-se também a outras espécies com altos níveis de plasticidade morfológica e grande abrangência territorial, como *Myrcia amazonica*, *M. guianensis* e *M. multiflora*, que apresentam 29, 158 e 41 sinônimos, respectivamente (WCSP, 2022), e cujos morfotipos podem vir a ser reconhecidos como espécies separadas no futuro (LIMA *et al.*, 2021).

Para além de algumas espécies, toda a família Myrtaceae tem sido submetida a uma série de revisões nas últimas décadas devido às dificuldades impostas para a circunscrição de seus membros (BALBINOTT *et al.*, 2022). Diversos tratamentos taxonômicos regionais e trabalhos filogenéticos baseados no uso de marcadores nucleares não codificantes, bem como de genes e espaçadores intergênicos plastidiais, têm contribuído para o esclarecimento de algumas relações dentro da família (BALBINOTT *et al.*, 2022; PROENÇA *et al.*, 2022). A classificação moderna ao nível hierárquico de tribo, por exemplo, está fundamentada em dados provenientes do gene *matK* em conjunto com o íntron *trnK* (WILSON *et al.*, 2005), ao passo que estudos moleculares focados em grupos específicos, inseridos principalmente em *Eugenia* s.l. (MAZINE *et al.*, 2014) e *Myrcia* s.l. (LUCAS *et al.*, 2011), foram conduzidos no intuito de refinar aspectos evolutivos e biogeográficos em Myrteae, incluindo a atualização e o reposicionamento de alguns taxa.

Portanto, mesmo considerando a necessidade de informações provenientes dos vouchers, é esperado que muitos espécimes atualmente depositados em herbários com nomenclatura errônea ou ultrapassada não sejam reclassificados (MEIER *et al.*, 2006), e que identificações conflituosas possam surgir quando diversos grupos trabalham com o mesmo táxon (COLLINS; CRUICKSHANK, 2013). Ao realizar uma breve inspeção na base de dados do BOLD, é possível encontrar uma série de gêneros que atualmente não constam na lista de nomes oficiais em Myrtaceae, a exemplo de *Marlierea* (19 espécimes), *Gomidesia* (6) e *Hexachlamys* (8), além de algumas delimitações genéricas (*Campomanesia* sp., *Eugenia* sp.) e sinonímias, como *Blepharocalyx tweediei* e *B. salicifolius* (WCSP, 2022). Qualquer banco de armazenamento de sequências nucleotídicas, incluindo o GenBank, será suscetível a esse tipo de adversidade. Assim, no contexto do *DNA barcoding*, a falta de padronização ocasionada pelo envio de sequências com múltiplos nomes para a mesma espécie ou classificadas incorretamente também acaba gerando um efeito negativo importante. Nesse caso, o estabelecimento de políticas mais específicas para a submissão de dados associados a espécimes com características atípicas, como altas divergências intraespecíficas (MEIER *et al.*, 2006), somada a um sistema que elenca o grau de expertise e esforço dedicados a uma

determinada identificação (COLLINS; CRUICKSHANK, 2013) são aspectos que podem amenizar parte do problema.

## 6 CONSIDERAÇÕES FINAIS E PERSPECTIVAS

O presente trabalho objetivou contribuir para a geração de conhecimento acerca da técnica de *DNA barcoding* na tribo Myrteae (Myrtaceae), algo até então pouco explorado na literatura, a partir de metodologias baseadas no cálculo das distâncias genéticas intra e interespecíficas e em filogenias. No entanto, algumas limitações importantes comprometeram a geração de resultados mais robustos sobre a eficácia dos marcadores testados.

Cerca de 1% das 2500 espécies e 20% dos 51 gêneros previstos para Myrteae (VASCONCELOS *et al.*, 2017) foram contemplados na construção da biblioteca de referência, o que subestima consideravelmente a riqueza existente no grupo como um todo e dentro de cada gênero, alguns representados por uma única espécie nesse levantamento. As zonas de distribuição dos taxa selecionados também apresentaram grande variação de amplitude, podendo ser restritas, a exemplo de *Psidium macahense*, endêmico da Mata Atlântica brasileira (TULER *et al.*, 2019), ou englobar diversos domínios fitogeográficos, como no caso de *Eugenia uniflora* (TURCHETTO-ZOLET *et al.*, 2016) e *Myrcia splendens* (LIMA DOS SANTOS *et al.*, 2021). Desse modo, nem sempre foi possível abranger indivíduos coletados em múltiplas localidades, minimizando a verdadeira diversidade genética associada a cada um dos marcadores. Ademais, boa parte dos trabalhos filogenéticos, utilizados nessa monografia como a principal fonte de recuperação dos dados avaliados, empregam poucas sequências nucleotídicas de uma mesma espécie, geralmente uma ou duas, o que atua como um empecilho adicional para a obtenção de quantidades ideais de amostras intraespecíficas. A dificuldade em encontrar espécies com os mesmos *loci* sequenciados na plataforma do GenBank também compromete a realização de estudos comparativos, embora não impossibilite que as informações moleculares atualmente disponíveis sejam exploradas para a delimitação de novas hipóteses.

Todos esses problemas já foram relatados na literatura como fatores importantes para a geração de vieses (LIU *et al.*, 2022; MEIER *et al.*, 2006; MEYER; PAULAY, 2005; MORITZ; CICERO, 2004). Espera-se, portanto, que os índices de adequabilidade aqui mensurados diminuam à medida que o esforço amostral se intensifique, seja pela inserção de indivíduos coespecíficos com uma representatividade geográfica mais fidedigna ou pela incorporação de novos grupos, incluindo membros intimamente relacionados, à biblioteca.

Apesar de todas as limitações, os dados obtidos a partir desse trabalho forneceram um panorama condizente com outras avaliações implementadas em grupos Neotropicais. Alguns dos *loci* comumente sequenciados para estudos filogenéticos em Myrteae coincidiram com os marcadores clássicos de DNA *barcoding* em plantas, possibilitando que o poder discriminatório dessas regiões fosse verificado através de diferentes abordagens analíticas. Nesse sentido, os espaçadores nucleares ITS demonstraram a melhor performance individual quando comparados a dois marcadores plastidiais selecionados, um não codificante (*psbA-trnH*) e outro codificante (*matK*), apresentando o maior grau de polimorfismo genético intra e interespecífico. Mais da metade das espécies exibiram *barcoding gaps* para essa região ou formaram agrupamentos monofiléticos nos três métodos de reconstrução filogenética utilizados. Além disso, de 58% a 81% dos indivíduos foram identificados corretamente nos testes de similaridade, dependendo da estratégia adotada (nível de rigor e limite de corte). Os marcadores plastidiais, em contrapartida, expressaram resultados menos satisfatórios, principalmente pela proporção considerável de divergências interespecíficas nulas e pela incapacidade de reunir a maior parte das espécies em clados exclusivos. Tais resultados sugerem que o *locus* nuclear pode servir como um ponto inicial para investigações mais profundas na tribo, visto que seu desempenho é o mais promissor, apesar de não ser o ideal.

Conforme previamente discutido, os desafios impostos pelos atributos evolutivos, genéticos e bioquímicos encontrados em famílias botânicas costumam ser mais pronunciados do que em animais, especialmente nos grupos caracterizados por uma alta diversidade biológica, o que acaba comprometendo as taxas de identificação via *DNA barcoding*. Embora Lahaye *et al.* (2008) tenham sugerido que cerca de 90% das espécies de plantas serão identificadas através de *barcodes* universais, as estimativas discretas adotadas por estudos subsequentes parecem ser mais coerentes. Nesse cenário, os *loci* plastidiais não chegarão a ultrapassar 70% de atribuições adequadas, sendo que o melhor desfecho possível em alguns casos será apenas a indicação de um grupo de espécies compatível com o material analisado (FAZEKAS *et al.*, 2008; HOLLINGSWORTH *et al.*, 2009). A inclusão de marcadores complementares que pertencem a um mesmo grupo de ligação gênica (como o genoma plastidial) nem sempre superará esse impasse, visto que aumentos sutis de resolução tendem a ser esperados, muitas vezes não compensando os esforços de sequenciamento (FAZEKAS *et al.*, 2008, 2009). Assim, a sondagem de genes nucleares apropriados para as principais ordens e famílias botânicas complexas, como Myrtaceae, poderá ser uma estratégia de incremento da variabilidade dos *barcodes* nesses grupos a curto ou médio prazo (FAZEKAS *et al.*, 2009). Analogamente, o desenvolvimento de protocolos de amplificação e sequenciamento



específicos para alguns taxa problemáticos se tornará inevitável (VASCONCELOS *et al.*, 2021), mesmo que isso desobedeça algumas das premissas ideais estipuladas para a seleção de um *barcode*.

O uso de regiões com performances modestas, todavia, já pode ser suficiente para análises de indivíduos imaturos ou fragmentados dentro de uma área geográfica restrita, sem a coocorrência de muitos taxa aparentados (HOLLINGSWORTH; GRAHAM; LITTLE, 2011). Discriminações a nível de gênero ou até mesmo família também acabam sendo de grande utilidade, principalmente na detecção de alguns tipos de adulterantes e na orientação de usuários desprovidos de qualquer conhecimento prévio sobre o organismo investigado (HOLLINGSWORTH; GRAHAM; LITTLE, 2011). Sendo assim, o emprego de um método molecular ajuda a restringir o espectro de identidades que podem ser atribuídas a um espécime, sendo posteriormente suplementado com consultas a especialistas ou buscas melhor direcionadas em chaves dicotômicas, quando viável. A taxonomia e sistemática tradicionais, portanto, continuarão exercendo um papel importante e insubstituível, embora a elaboração de mais estudos sobre *DNA barcoding* garantirá que a integração de dados genéticos e morfológicos seja mais eficiente e pormenorizada.

Como desdobramento desse trabalho, espera-se que novas sequências nucleotídicas possam ser geradas em laboratório, visando completar algumas lacunas da biblioteca de referência e ampliar o número de espécies abordadas, especialmente com representações mais significativas dentro de cada gênero. Esse procedimento auxiliará na condução de análises sobre o perfil de sequenciamento e amplificação de cada *barcode* em Myrteae, algo que não pôde ser aqui averiguado, além de permitir que diferentes marcadores sejam sequenciados e concatenados em um mesmo indivíduo. Assim, combinações dos ITS com outras regiões genômicas poderão ser testadas com o intuito de verificar se a habilidade discriminatória desses espaçadores é passível de melhorias em uma abordagem baseada em múltiplos *loci*.

Finalmente, a adoção de protocolos com o máximo de padronização possível em conjunto com inovações como os sequenciamentos de nova geração e ferramentas robustas de bioinformática determinará os rumos futuros do *DNA barcoding*. Análises de alta resolução por fusão (*High Resolution Melting* - HRM) acopladas ao emprego dos *barcodes* clássicos já estão garantindo progressos substanciais, visto que essa tecnologia, conhecida como Bar-HRM, permite autenticar com alta resolução plantas medicinais comercializadas em um estado menos processado (YU *et al.*, 2021). Por outro lado, a estratégia de *genome skimming*, que envolve o sequenciamento de fragmentos curtos de DNA com baixa cobertura, pode ser utilizada na reconstrução de plastomas e inspeção de polimorfismos de nucleotídeo único

espécie-específicos (MASCARELLO *et al.*, 2021). Tal processo auxiliará, por exemplo, no delineamento de regiões altamente variáveis cujo potencial discriminatório supera o dos *loci* convencionais, ao menos em alvos particulares, como madeiras tropicais cobiçadas (MASCARELLO *et al.*, 2021). Nesse sentido, o genoma completo do cloroplasto ou algumas sequências mais longas de DNA ribossomal (~6000 pb) também estão sendo testadas para discernir linhagens abaixo do nível de espécie, como subespécies, variedades e até mesmo indivíduos (KANE *et al.*, 2012). Consequentemente, a metodologia de *ultra-barcoding* (UBC) e outras tecnologias emergentes certamente trarão perspectivas ambiciosas ou promissoras para o campo da taxonomia vegetal, sobretudo pela agregação de novos dados informativos nas análises moleculares.

## REFERÊNCIAS

- ALVES, T. L. S. Applying DNA barcoding to South-American Tigridaeae (Iridaceae): restricted variation limits identification in related species. *In*: ALVES, T. L. S. **Aplicação de DNA Barcoding para identificação de espécies pertencentes às tribos Sisyrinchieae e Tigridaeae (Iridaceae)**. 2013. Tese (Doutorado em Botânica) - Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.
- ALVES, T. L. S. *et al.* Species discrimination in *Sisyrinchium* (Iridaceae): Assessment of DNA barcodes in a taxonomically challenging genus. **Molecular Ecology Resources**, [s. l.], v. 14, n. 2, p. 324–335, 2014.
- BALBINOTT, N. *et al.* Perspectives in Myrtaceae evolution from plastomes and nuclear phylogenies. **Genetics and Molecular Biology**, [Ribeirão Preto], v. 45, n. 1, p. 20210191, 2022.
- BELL, K. L. *et al.* Optimisation of a pollen DNA metabarcoding method for diet analysis of flying-foxes (*Pteropus* spp.). **Australian Journal of Zoology**, [Canberra], v. 68, n. 6, p. 273–284, 2021.
- BENSON, D. A. *et al.* GenBank. **Nucleic Acids Research**, [Oxford], v. 41, n. D1, p. D36-D42, 2012.
- BERG, O. Revisio Myrtacearum Americae. *In*: VON SCHLECHTENDAL, D. F. L. **Linnaea**, v. 27. [S.l.: s.n.], 1855–1856.
- BERG, O. Myrtaceae. *In*: VON MARTIUS, C. F. P. (ed.). **Flora Brasiliensis**, v. 14. Leipzig: Frid. Fleischer, 1857–1859.
- BIFFIN, E. *et al.* Evolution of exceptional species richness among lineages of fleshy-fruited Myrtaceae. **Annals of Botany**, [Oxford], v. 106, n. 1, p. 79-93, 2010.
- BOLD. **Barcode of Life Data System v4**. 2022. Disponível em: <https://www.boldsystems.org/>. Acesso em: 03 de set. de 2022.
- BOLSON, M. *et al.* ITS and trnH-psbA as Efficient DNA Barcodes to Identify Threatened Commercial Woody Angiosperms from Southern Brazilian Atlantic Rainforests. **PLoS ONE**, [San Francisco], v. 10, n. 12, p. e0143049, 2015.
- BRANDÃO, A. P. M. E. *et al.* Evaluation of DNA markers for molecular identification of three *Piper* species from Brazilian Atlantic Rainforest. **PLoS ONE**, [San Francisco], v. 15, n. 10, p. 1–17, 2020.
- BROWN, B.; EMBERSON, R. M.; PATERSON, A. M. Mitochondrial COI and II provide useful markers for wiseana (Lepidoptera: Hepialidae) species identification. **Bulletin of Entomological Research**, [Cambridge], v. 89, n. 4, p. 287–293, 1999.
- BROWN, S. D. J. *et al.* Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. **Molecular Ecology Resources**, [s. l.], v. 12, n. 3, p. 562-565, 2012.

- BUCKLIN, A. *et al.* Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR. *In: ZEHR, J. P.; VOYTEK, M. A. (ed.). Molecular Ecology of Aquatic Communities*. 1. ed. Dordrecht: Springer, 1999.
- BÜNGER, M. O. *et al.* Resolving species delimitations in the *Eugenia involucrata* Group (*Eugenia* sect. *Phyllocalyx* - Myrtaceae) with morphometric analysis. **Systematic Botany**, [St. Louis], v. 40, n. 4, p. 995-1002, 2015.
- BUYS, M. H. *et al.* Preparing for the invasion: efficacy of DNA barcoding to discern the host range of myrtle rust (*Puccinia psidii*) among species of Myrtaceae. **Forestry**, [s. l.], v. 89, n. 3, p. 263–270, 2016.
- CALDART, E. T. *et al.* Análise filogenética: conceitos básicos e suas utilizações como ferramenta para virologia e epidemiologia molecular. **Acta Scientiae Veterinariae**, Porto Alegre, v. 44, n. 1392, p. 1-20, 2016.
- CAMERON, S.; RUBINOFF, D.; WILL, K. Who will actually use DNA barcoding and what will it cost?. **Systematic Biology**, Oxford, v. 55, n. 5, p. 844-847, 2006.
- CANNIZZARO, C. *et al.* Forest landscapes increase diversity of honeybee diets in the tropics. **Forest Ecology and Management**, [s. l.], v. 504, p. 119869, 2022.
- CASTRESANA, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. **Molecular Biology and Evolution**, [Oxford], v. 17, n. 4, p. 540-552, 2000.
- CAZZOLLA GATTI, R. *et al.* The number of tree species on Earth. **Proceedings of the National Academy of Sciences**, [Washington], v. 119, n. 6, p. 1–11, 2022.
- CBOL PLANT WORKING GROUP. A DNA barcode for land plants. **Proceedings of the National Academy of Sciences**, [Washington], v. 106, n. 31, p. 12794–12797, 2009.
- CHASE, M. W. *et al.* Land plants and DNA barcodes: short-term and long-term goals. **Philosophical Transactions of the Royal Society B: Biological Sciences**, [London], v. 360, n. 1462, p. 1889-1895, 2005.
- CHASE, M. W. *et al.* A proposal for a standardised protocol to barcode all land plants. **Taxon**, [s. l.], v. 56, n. 2, p. 295–299, 2007.
- CHEN, S. *et al.* Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. **PLoS ONE**, [San Francisco], v. 5, n. 1, p. e8613, 2010.
- CHINA PLANT BOL GROUP. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. **Proceedings of the National Academy of Sciences**, [Washington], v. 108, n. 49, p. 19641–19646, 2011.
- COLLINS, R. A.; CRUICKSHANK, R. H. The seven deadly sins of DNA barcoding. **Molecular Ecology Resources**, [s. l.], v. 13, n. 6, p. 969–975, 2013.

CONSORTIUM FOR THE BARCODE OF LIFE. **Barcode of Life: Identifying Species with DNA barcoding**. 2012. Disponível em: <https://web.archive.org/web/20120626065130/http://www.barcodeoflife.org/>. Acesso em: 04 de set. de 2022.

CORADIN, L.; SIMINSKI, A.; REIS, A. **Espécies nativas da flora brasileira de valor econômico atual ou potencial: plantas para o futuro – Região Sul**. Brasília: MMA, 2011.

DA COSTA, I. R. **Estudos cromossômicos em espécies de Myrtaceae Juss. no sudeste do Brasil**. 2004. Dissertação (Mestrado em Biologia Vegetal) - Instituto de Biologia, Universidade Estadual de Campinas, Campinas, 2004.

DA COSTA, I. R.; FORNI-MARTINS, E. R. Chromosome studies in species of *Eugenia*, *Myrciaria* and *Plinia* (Myrtaceae) from south-eastern Brazil. **Australian Journal of Botany**, [Cambera], v. 54, n. 4, p. 409-415, 2006a.

DA COSTA, I. R.; FORNI-MARTINS, E. R. Chromosome studies in Brazilian species of *Campomanesia* Ruiz & Pávon and *Psidium* L. (Myrtaceae Juss.). **Caryologia**, [s. l.], v. 59, n. 1, p. 7-13, 2006b.

DARRIBA, D. *et al.* jModelTest 2: more models, new heuristics and parallel computing. **Nature Methods**, [s. l.], v. 9, n. 8, p. 772, 2012.

DAYRAT, B. Towards integrative taxonomy. **Biological Journal of the Linnean Society**, [London], v. 85, n. 3, p. 407-417, 2005.

DE CANDOLLE, A. P. Myrtaceae. *In*: DE CANDOLLE, A. P. **Prodromus Systematis Naturalis Regni Vegetabilis**, v. 3. [S.l.: s.n.], 1828.

DE LIMA, R. A. F. *et al.* Can plant DNA barcoding be implemented in species-rich tropical regions? A perspective from São Paulo State, Brazil. **Genetics and Molecular Biology**, [Ribeirão Preto], v. 41, n. 3, p. 661–670, 2018.

DE LUCA, A.; SIBILIO, G.; GUACCHIO, E. DEL. DNA barcoding analysis of *Eucalyptus camaldulensis* Dehnh. (Myrtaceae). **Delpinoa**, [s. l.], v. 56, 2014.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, [Oxford], v. 32, n. 5, p. 1792-1797, 2004.

FAZEKAS, A. J. *et al.* Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. **PloS ONE**, [San Francisco], v. 3, n. 7, p. e2802, 2008.

FAZEKAS, A. J. *et al.* Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? **Molecular Ecology Resources**, [s. l.], v. 9, n. s1, p. 130-139, 2009.

FELSENSTEIN, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. **Journal of Molecular Evolution**, [s. l.], v. 17, n. 6, p. 368-376, 1981.

FELSENSTEIN, J. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, [s. l.] v. 39, n. 4, p. 783-791, 1985.

- FLICKINGER, J. A. *et al.* A phylogenetic survey of Myrtaceae in the Greater Antilles with nomenclatural changes for some endemic species. **Taxon**, [s. l.], v. 69, n. 3, p. 448–480, 2020.
- FUNK, D. J.; OMLAND, K. E. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. **Annual Review of Ecology, Evolution and Systematics**, [s. l.], v. 34, p. 397–423, 2003.
- GAO, T. *et al.* Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. **BMC Evolutionary Biology**, [s. l.], v. 10, n. 324, p. 1–7, 2010.
- GIUDICELLI, G. C.; MÄDER, G.; DE FREITAS, L. B. Efficiency of ITS sequences for DNA barcoding in *Passiflora* (Passifloraceae). **International Journal of Molecular Sciences**, [s. l.], v. 16, n. 4, p. 7289–7303, 2015.
- GONZALEZ, M. A. *et al.* Identification of Amazonian Trees with DNA Barcodes. **PLoS ONE**, [San Francisco], v. 4, n. 10, p. e7483, 2009.
- GRESSLER, E.; PIZO, M. A.; MORELLATO, L. P. C. Polinização e dispersão de sementes em Myrtaceae do Brasil. **Brazilian Journal of Botany**, São Paulo, v. 29, n. 4, p. 509–530, 2006.
- HASEGAWA, M.; KISHINO, H.; YANO, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. **Journal of Molecular Evolution**, [s. l.], v. 22, p. 160–174, 1985.
- HEBERT, P. D. N. *et al.* Biological identifications through DNA barcodes. **Proceedings of the Royal Society B: Biological Sciences**, [London], v. 270, n. 1512, p. 313–321, 2003.
- HEBERT, P. D. N. *et al.* Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. **Proceedings of the National Academy of Sciences**, [Washington], v. 101, n. 41, p. 14812–14817, 2004.
- HOLLINGSWORTH, M. L. *et al.* Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. **Molecular Ecology Resources**, [s. l.], v. 9, n. 2, p. 439–457, 2009.
- HOLLINGSWORTH, P. M. Refining the DNA barcode for land plants. **Proceedings of the National Academy of Sciences**, [Washington], v. 108, n. 49, p. 19451–19452, 2011.
- HOLLINGSWORTH, P. M.; GRAHAM, S. W.; LITTLE, D. P. Choosing and using a plant DNA barcode. **PLoS ONE**, [San Francisco], v. 6, n. 5, p. e19254, 2011.
- INTERNATIONAL BARCODE OF LIFE. **IBOL Programs**. 2022. Disponível em: <https://ibol.org/programs/program-overview/>. Acesso em: 04 de set. de 2022.
- JONES, L. *et al.* Barcode UK: A complete DNA barcoding resource for the flowering plants and conifers of the United Kingdom. **Molecular Ecology Resources**, [s. l.], v. 21, n. 6, p. 2050–2062, 2021.

- KANE, N. *et al.* Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. **American Journal of Botany**, [St. Louis], v. 99, n. 2, p. 320-329, 2012.
- KIMURA, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. **Journal of Molecular Evolution**, [*s. l.*], v. 16, n. 2, p. 111-120, 1980.
- KIMURA, M. Estimation of evolutionary distances between homologous nucleotide sequences. **Proceedings of the National Academy of Sciences**, [Washington], v. 78, n. 1, p. 454-458, 1981.
- KRAWCZYK, K.; SZCZECIŃSKA, M.; SAWICKI, J. Evaluation of 11 single-locus and seven multilocus DNA barcodes in *Lamium* L. (Lamiaceae). **Molecular Ecology Resources**, [*s. l.*], v. 14, n. 2, p. 272–285, 2014.
- KRESS, W. J. Plant DNA barcodes: Applications today and in the future. **Journal of Systematics and Evolution**, [*s. l.*], v. 55, n. 4, p. 291–307, 2017.
- KRESS, W. J.; ERICKSON, D. L. A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcL* Gene Complements the Non-Coding *trnH-psbA* Spacer Region. **Plos ONE**, [San Francisco], v. 2, n. 6, p. e508, 2007.
- KRESS, W. J. *et al.* Use of DNA barcodes to identify flowering plants. **Proceedings of the National Academy of Sciences**, [Washington], v. 102, n. 23, p. 8369–8374, 2005.
- LAHAYE, R. *et al.* DNA barcoding the floras of biodiversity hotspots. **Proceedings of the National Academy of Sciences**, [Washington], v. 105, n. 8, p. 2923–2928, 2008.
- LANDRUM, L. R. *Campomanesia, Pimenta, Blepharocalyx, Legrandia, Acca, Myrrhinium, and Luma* (Myrtaceae). **Flora Neotropica**, New York, v. 45, p. 1–178, 1986.
- LANDRUM, L. R. *Mosiera* (Myrtaceae) in Mexico and Mesoamerica. **Novon**, [St. Louis], v. 2, n. 1, p. 26–29, 1992.
- LANDRUM, L. R.; KAWASAKI, M. L. The genera of Myrtaceae in Brazil: an illustrated synoptic treatment and identification keys. **Brittonia**, New York, v. 49, n. 4, p. 508-536, 1997.
- LE, D. T. *et al.* The utility of DNA barcodes to confirm the identification of palm collections in botanical gardens. **Plos ONE**, [San Francisco], v. 15, n. 7, p. e0235569, 2020.
- LIMA, D. F. *et al.* Phylogeny and biogeography of *Myrcia* sect. *Aguava* (Myrtaceae, Myrteae) based on phylogenomic and Sanger data provide evidence for a Cerrado origin and geographically structured clades. **Molecular Phylogenetics and Evolution**, [*s. l.*], v. 157, p. 107043, 2021.
- LIMA DOS SANTOS, L. *et al.* Phylogenetic and biogeographic analysis in *Myrcia* sect. *Myrcia* (*Myrcia* s.l., Myrtaceae) with focus on highly polyphyletic *Myrcia splendens*. **International Journal of Plant Sciences**, Chicago, v. 182, n. 9, p. 778–792, 2021.
- LIU, J. *et al.* Identification of species in the angiosperm family Apiaceae using DNA barcodes. **Molecular Ecology Resources**, [*s. l.*], v. 14, n. 6, p. 1231–1238, 2014.

- LIU, Z-F. *et al.* Do taxon-specific DNA barcodes improve species discrimination relative to universal barcodes in Lauraceae? **Botanical Journal of the Linnean Society**, [London], v. 199, p. 741–753, 2022.
- LUCAS, E. J. *et al.* Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae (Myrtales). **Taxon**, [*s. l.*], v. 56, n. 4, p. 1105–1128, 2007.
- LUCAS, E. J. *et al.* Phylogenetics, morphology and evolution of the large genus *Myrcia* s.l. (Myrtaceae). **International Journal of Plant Sciences**, Chicago, v. 172, n. 7, p. 915–934, 2011.
- LUCAS, E. J. *et al.* A New Subtribal Classification of Tribe Myrteae (Myrtaceae). **Systematic Botany**, [St. Louis], v. 44, n. 3, p. 560–569, 2019.
- LUGON, M. D. *et al.* Is Your Açaí Really from Amazon? Using DNA Barcoding to Authenticate Commercial Products. **Food Analytical Methods**, [*s. l.*], v. 14, p. 1559–1566, 2021.
- LUO, K. *et al.* Assessment of candidate plant DNA barcodes using the Rutaceae family. **Science China Life Sciences**, [*s. l.*], v. 53, n. 6, p. 701-708, 2010.
- MADDISON, W. P. Gene trees in species trees. **Systematic Biology**, Oxford, v. 46, n. 3, p. 523-536, 1997.
- MAIA, V. H. *et al.* DNA Barcoding Bromeliaceae: Achievements and Pitfalls. **PLoS ONE**, [San Francisco], v. 7, n. 1, p. e29877, 2012.
- MALLOTT, E. K.; GARBER, P. A.; MALHI, R. S. trnL outperforms rbcL as a DNA metabarcoding marker when compared with the observed plant component of the diet of wild white-faced capuchins (*Cebus capucinus*, Primates). **PLoS ONE**, [San Francisco], v. 13, n. 6, p. e0199556, 2018.
- MAO, X. *et al.* Establishing community-wide DNA barcode references for conserving mangrove forests in China. **BMC Plant Biology**, [*s. l.*], v. 21, n. 1, p. 1–14, 2021.
- MARTINELLI, G.; MORAES, M. A. (org.). **Livro vermelho da flora do Brasil**. Rio de Janeiro: Andrea Jakobsson - Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, 2013.
- MASCARELLO, M. *et al.* Genome skimming reveals novel plastid markers for the molecular identification of illegally logged African timber species. **PloS ONE**, [San Francisco], v. 16, n. 6, p. e0251655, 2021.
- MAZINE, F. *et al.* A preliminary phylogenetic analysis of *Eugenia* (Myrtaceae: Myrteae), with a focus on Neotropical species. **Kew Bulletin**, Kew, v. 69, n. 2, p. 1-14, 2014.
- MCVAUGH, R. The genera of American Myrtaceae: an interim report. **Taxon**, [*s. l.*], v. 17, n. 4, p. 354-418, 1968.
- MEIER, R. *et al.* DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. **Systematic Biology**, Oxford, v. 55, n. 5, p. 715-728, 2006.



MEIER, R.; ZHANG, G.; ALI, F. The Use of Mean Instead of Smallest Interspecific Distances Exaggerates the Size of the “Barcoding Gap” and Leads to Misidentification. **Systematic Biology**, Oxford, v. 57, n. 5, p. 809–813, 2008.

MELLO, I. C. T. *et al.* A segment of *rbcL* gene as a potential tool for forensic discrimination of *Cannabis sativa* seized at Rio de Janeiro, Brazil. **International Journal of Legal Medicine**, [s. l.], v. 130, n. 2, p. 353-356, 2016.

MEYER, C. P.; PAULAY, G. DNA Barcoding: Error Rates Based on Comprehensive Sampling. **PLoS Biology**, [San Francisco], v. 3, n. 12, p. e422, 2005.

MILLA, L. *et al.* Pollen DNA metabarcoding identifies regional provenance and high plant diversity in Australian honey. **Ecology and Evolution**, [s. l.], v. 11, n. 13, p. 8683–8698, 2021.

MORITZ, C.; CICERO, C. DNA barcoding: promise and pitfalls. **PLoS Biology**, [San Francisco], v. 2, n. 10, p. e354, 2004.

MOWER, J. P. *et al.* Gene transfer from parasitic to host plants. **Nature**, [United Kingdom], v. 432, n. 7014, p. 165-166, 2004.

MOWER, J. P. *et al.* Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. **BMC Evolutionary Biology**, [s. l.], v. 7, n. 1, p. 1–14, 2007.

MURILLO-A, J.; STUESSY, T. F.; RUIZ, E. Phylogenetic relationships among *Myrceugenia*, *Blepharocalyx* and *Luma* (Myrtaceae) based on paired-sites models and the secondary structures of ITS and ETS sequences. **Plant Systematics and Evolution**, [s. l.], v. 299, n. 4, p. 713-729, 2013.

NAKAMURA, S. S. *et al.* Autenticação molecular de *Maytenus* sp por PCR-RFLP. **Semina: Ciências Agrárias**, Londrina, v. 34, n. 2, p. 627–634, 2013.

NAUER, F. *et al.* A taxonomic review of the genus *Hypnea* (Gigartinales, Rhodophyta) in Brazil based on DNA barcode and morphology. **Brazilian Journal of Botany**, São Paulo, v. 42, p. 561–574, 2019.

NEWMASER, S. G. *et al.* Testing candidate plant barcode regions in the Myristicaceae. **Molecular Ecology Resources**, [s. l.], v. 8, n. 3, p. 480–490, 2008.

NIEMANN, H. J. *et al.* Using a DNA barcoding approach to facilitate biosecurity: Identifying invasive alien macrophytes traded within the South African aquarium and pond plant industry. **South African Journal of Botany**, [s. l.], v. 144, p. 364–376, 2022.

NITHANIYAL, S. *et al.* Forensic application of DNA barcoding in the identification of commonly occurring poisonous plants. **Journal of Forensic and Legal Medicine**, [s. l.], v. 78, p. 102126, 2021.

PANG, X. *et al.* Applying plant DNA barcodes for Rosaceae species identification. **Cladistics**, [Washington], v. 27, n. 2, p. 165-170, 2011.

PARRA-O, C.; BOHÓRQUEZ-OSORIO, A. F. Effectiveness of DNA barcoding markers in the description of a new and unusual calyprate species of *Myrcianthes* (Myrtaceae). **Phytotaxa**, [Auckland], v. 284, n. 3, p. 203–210, 2016.

PAWLOWSKI, J. *et al.* CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. **PLoS Biology**, [San Francisco], v. 10, n. 11, p. e1001419, 2012.

PIRES, A. C.; MARINONI, L. DNA barcoding and traditional taxonomy unified through Integrative Taxonomy: a view that challenges the debate questioning both methodologies. **Biota Neotropica**, São Paulo, v. 10, n. 2, p. 339-346, 2010.

PROENÇA, C. E. B. *et al.* Myrtaceae Juss. *In: Flora e Funga do Brasil*. Jardim Botânico do Rio de Janeiro. [2022?]. Disponível em: <https://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB171>. Acesso em: 03 de set. de 2022.

R CORE TEAM. **The R Project for Statistical Computing**. [2022?]. Disponível em: <https://www.r-project.org>. Acesso em: 04 de set. de 2022.

RAMBAUT, A. **Figtree, a graphical viewer of phylogenetic trees**. 2007. Disponível em: <http://tree.bio.ed.ac.uk/software/figtree>. Acesso em: 04 de set. de 2022.

RAMBAUT, A. *et al.* Posterior summarization in Bayesian phylogenetics using Tracer 1.7. **Systematic Biology**, Oxford, v. 67, n. 5, p. 901-904, 2018.

RATNASINGHAM, S; HEBERT, P. D. N. BOLD: The Barcode of Life Data System. **Molecular Ecology Notes**, [s. l.], v. 7, n. 3, p. 355-364, 2007.

RIBEIRO, A. S. D. *et al.* O gene *rbcL* como barcode para identificação forense de *Cannabis sativa*. **Saúde, Ética & Justiça**, São Paulo, v. 18, p. 67-71, 2013.

RICHARDSON, J. E. *et al.* Rapid diversification of a species-rich genus of neotropical rain forest trees. **Science**, [s. l.], v. 293, n. 5538, p. 2242-2245, 2001.

ROSLIM, D. I. Analysis of *matK*, *rbcL* and *trnL-trnF* intergenic spacer sequences on Durik-Durik (*Syzygium* sp). *In: Journal of Physics: Conference Series*. Bristol: IOP Publishing, 2019.

ROSS, H. A.; MURUGAN, S.; LI, W. L. S. Testing the reliability of genetic methods of species identification via simulation. **Systematic Biology**, Oxford, v. 57, n. 2, p. 216-230, 2008.

ROZAS, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. **Molecular Biology and Evolution**, Oxford, v. 34, n. 12, p. 3299-3302, 2017.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, Oxford, v. 4, n. 4, p. 406-425, 1987.

SEIFERT, K. A. Progress towards DNA barcoding of fungi. **Molecular Ecology Resources**, [s. l.], v. 9, n. s1, p. 83-89, 2009.

SHAW J. *et al.* Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. **American Journal of Botany**, [St. Louis], v. 94, n. 3, p. 275–288, 2007.

SMITH, S. A.; DONOGHUE, M. J. Rates of molecular evolution are linked to life history in flowering plants. **Science**, [s. l.], v. 322, n. 5898, p. 86-89, 2008.

SOUZA NETO, J. D. *et al.* Advances and perspectives on the evolutionary history and diversification of Neotropical Myrteae (Myrtaceae). **Botanical Journal of the Linnean Society**, [London], v. 199, n. 1, p. 173-195, 2022.

SPOONER, D. M. DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. **American Journal of Botany**, [St. Louis], v. 96, n. 6, p. 1177–1189, 2009.

STAGGEMEIER, V. G. *et al.* Clade-specific responses regulate phenological patterns in Neotropical Myrtaceae. **Perspectives in Plant Ecology, Evolution and Systematics**, [s. l.], v. 17, n. 6, p. 476-490, 2015.

STARR, J. R.; NACZI, R. F. C.; CHOUINARD, B. N. Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). **Molecular Ecology Resources**, [s. l.], v. 9, n. Suppl. 1, p. 151–163, 2009.

STECH, M.; QUANDT, D. 20,000 species and five key markers: the status of molecular bryophyte phylogenetics. **Phytotaxa**, [Auckland], v. 9, n. 1, p. 196-228, 2010.

SUCHARD, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. **Virus Evolution**, Oxford, v. 4, n. 1, p. vey016, 2018.

TAMURA, K.; NEI, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. **Molecular Biology and Evolution**, Oxford, v. 10, n. 3, p. 512-526, 1993.

TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. **Molecular Biology and Evolution**, Oxford, v. 38, n. 7, p. 3022–3027, 2021.

THORNHILL, A. H. *et al.* Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. **Molecular Phylogenetics and Evolution**, [s. l.], v. 93, p. 29–43, 2015.

TODESCHINI, C. C. **Uso de DNA barcode para identificação de espécies de palmito como ferramenta para a genética forense**. 2019. Dissertação (Mestrado em Genética e Biologia Molecular) - Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019.

TRIFINOPOULOS, J. *et al.* W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. **Nucleic Acids Research**, Oxford, v. 44, n. W1, p. W232-W235, 2016.

TULER, A. C. *et al.* Re-establishment of *Psidium macahense* (Myrtaceae, Myrteae), an endemic species from the Brazilian Atlantic forest. **Phytotaxa**, [Auckland], v. 397, n. 1, p. 34–44, 2019.

TURCHETTO-ZOLET, A. C. *et al.* Phylogeography and ecological niche modelling in *Eugenia uniflora* (Myrtaceae) suggest distinct vegetational responses to climate change between the southern and the northern Atlantic Forest. **Botanical Journal of the Linnean Society**, [London], v. 182, n. 3, p. 670-688, 2016.

VASCONCELOS, S. *et al.* Unraveling the plant diversity of the Amazonian canga through DNA barcoding. **Ecology and Evolution**, [s. l.], v. 11, n. 19, p. 13348–13362, 2021.

VASCONCELOS, T. N. C. *et al.* Myrteae phylogeny, calibration, biogeography and diversification patterns: Increased understanding in the most species rich tribe of Myrtaceae. **Molecular Phylogenetics and Evolution**, [s. l.], v. 109, p. 113–137, 2017.

VIJAYAN, K.; TSOU, C. H. DNA barcoding in plants: Taxonomy in a new perspective. **Current Science**, [Bengaluru], v. 99, n. 11, p. 1530–1541, 2010.

VIRGILIO, M. *et al.* Identifying insects with incomplete DNA barcode libraries, African Fruit Flies (Diptera: Tephritidae) as a test case. **PLoS ONE**, [San Francisco], v. 7, n. 2, p. e31581, 2012.

VIVAS, C. V. *et al.* DNA barcoding in Atlantic Forest plants: What is the best marker for Sapotaceae species identification? **Genetics and Molecular Biology**, [Ribeirão Preto], v. 37, n. 4, p. 662–670, 2014.

WATI, R. *et al.* Filling gaps of reference DNA barcodes in *Syzygium* from rainforest fragments in Sumatra. **Tree Genetics & Genomes**, [s. l.], v. 18, n. 6, p. 1–15, 2022.

WCSP. **World Checklist of Selected Plant Families**. Facilitated by the Royal Botanic Gardens, Kew. [2022?]. Disponível em: <http://wvsp.science.kew.org/qsearch.do>. Acesso em: 05 de set. de 2022.

WILLIAMSON, J. *et al.* Exposing the illegal trade in cycad species (Cycadophyta: *Encephalartos*) at two traditional medicine markets in South Africa using DNA barcoding. **Genome**, [Ottawa], v. 59, n. 9, p. 771-781, 2016.

WILSON, P. G. Myrtaceae. *In*: KUBITZKI, K. (ed.). **The Families and Genera of Vascular Plants: Flowering Plants - Eudicots**, v. 10. Berlin: Springer, 2011.

WILSON, P. G. *et al.* Relationships within Myrtaceae sensu lato based on a matK phylogeny. **Plant Systematics and Evolution**, [s. l.], v. 251, n. 1, p. 3–19, 2005.

XU, S. Z.; LI, Z. Y.; JIN, X. H. DNA barcoding of invasive plants in China: A resource for identifying invasive plants. **Molecular Ecology Resources**, [s. l.], v. 18, n. 1, p. 128–136, 2018.

YANG, Z.; RANNALA, B. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. **Molecular Biology and Evolution**, Oxford, v. 14, n. 7, p. 717-724, 1997.

YANG, Z.; RANNALA, B. Molecular phylogenetics: principles and practice. **Nature Reviews Genetics**, [s. l.], v. 13, n. 5, p. 303–314, 2012.

YOCCOZ, N. G. *et al.* DNA from soil mirrors plant taxonomic and growth form diversity. **Molecular Ecology**, [s. l.], v. 21, n. 15, p. 3647–3655, 2012.

YU, J. *et al.* Progress in the use of DNA barcodes in the identification and classification of medicinal plants. **Ecotoxicology and Environmental Safety**, [s. l.], v. 208, p. 111691, 2021.

ZHANG, Q.; SODMERGEN. Why does biparental plastid inheritance revive in angiosperms?. **Journal of Plant Research**, [s. l.], v. 123, n. 2, p. 201-206, 2010.

**APÊNDICE A - RELAÇÃO DOS ARTIGOS SELECIONADOS PARA O LEVANTAMENTO DOS MARCADORES MOLECULARES SEQUENCIADOS NA TRIBO MYRTEAE**

(continua)

<b>Autores</b>	<b>Ano</b>	<b>Título</b>	<b>Temática Central</b>	<b>Marcadores</b>	<b>Nº Total de Espécies</b>	<b>Nº de Espécies de Myrteae</b>	<b>Sequências Novas e Publicadas</b>
Sytsma <i>et al.</i>	2004	Clades, clocks, and continents: Historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the Southern Hemisphere	Biogeografia e Filogenia	<i>matK, rbcL, ndhF</i>	152	20	Sim
Lucas <i>et al.</i>	2005	Phylogenetic patterns in the fleshy-fruited Myrtaceae – preliminary molecular evidence	Filogenia	ITS, <i>psbA-trnH</i>	64	54	Não
Wilson <i>et al.</i>	2005	Relationships within Myrtaceae sensu lato based on a <i>matK</i> phylogeny	Filogenia	<i>matK</i>	85	22	Sim
Lucas <i>et al.</i>	2007	Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae (Myrtales)	Biogeografia e Filogenia	ITS, ETS, <i>psbA-trnH, matK</i>	88	75	Sim
Biffin <i>et al.</i>	2010	Evolution of exceptional species richness among lineages of fleshy-fruited Myrtaceae	Filogenia	ITS, <i>ndhF, matK</i>	104	21	Sim
Lucas <i>et al.</i>	2011	Phylogenetics, Morphology, and Evolution of the Large Genus <i>Myrcia</i> s.l. (Myrtaceae)	Evolução, Filogenia e Morfologia	ITS, ETS, <i>psbA-trnH, trnL-trnF, matK</i>	74	74	Sim
Murillo-A <i>et al.</i>	2012	Phylogenetic relationships in <i>Myrceugenia</i> (Myrtaceae) based on plastid and nuclear DNA sequences	Filogenia	ITS, ETS, <i>matK, rpl32-trnL, trnQ-rps16, rpl16</i>	40	40	Sim
da Cruz <i>et al.</i>	2013	Phylogenetic analysis of the genus <i>Hexachlamys</i> (Myrtaceae) based on plastid and nuclear DNA sequences and their taxonomic implications	Filogenia	ITS2, <i>psbA-trnH, accD, rpoB, rpoC1</i>	31	31	Sim

<b>Autores</b>	<b>Ano</b>	<b>Título</b>	<b>Temática Central</b>	<b>Marcadores</b>	<b>Nº Total de Espécies</b>	<b>Nº de Espécies de Myrteae</b>	<b>Sequências Novas e Publicadas</b>
Murillo-A <i>et al.</i>	2013	Phylogenetic relationships among <i>Myrceugenia</i> , <i>Blepharocalyx</i> , and <i>Luma</i> (Myrtaceae) based on paired-sites models and the secondary structures of ITS and ETS sequences	Filogenia	ITS, ETS	93	91	Não
Mazine <i>et al.</i>	2014	A preliminary phylogenetic analysis of <i>Eugenia</i> (Myrtaceae: Myrteae), with a focus on Neotropical species	Filogenia	ITS, ETS, <i>psbA-trnH</i>	68	65	Sim
Staggemeier <i>et al.</i>	2015	Clade-specific responses regulate phenological patterns in Neotropical Myrtaceae	Fenologia e Filogenia	ITS, ETS, <i>psbA-trnH</i> , <i>ndhF</i> , <i>trnL-trnF</i> , <i>trnQ-rps16</i> , <i>matK</i> , <i>rbcL</i>	58	58	Não
Staggemeier <i>et al.</i>	2015	Phylogenetic analysis in <i>Myrcia</i> section <i>Aulomyrcia</i> and inferences on plant diversity in the Atlantic rainforest	Filogenia	ITS, <i>psbA-trnH</i> , <i>ndhF</i> , <i>trnL-trnF</i> , <i>trnQ-rps16</i>	85	85	Sim
Thornhill <i>et al.</i>	2015	Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny	Biogeografia e Filogenia	ITS, <i>ndhF</i> , <i>matK</i>	202	38	Sim
Bunger <i>et al.</i>	2016	The evolutionary history of <i>Eugenia</i> sect. <i>Phyllocalyx</i> (Myrtaceae) corroborates historically stable areas in the southern Atlantic forests	Filogenia e Modelagem de Nicho Ecológico	ITS, <i>psbA-trnH</i> , <i>rpl16</i> , <i>rpl32-trnL</i> , <i>trnQ-rps16</i>	42	42	Sim
Buys <i>et al.</i>	2016	Preparing for the invasion: efficacy of DNA barcoding to discern the host range of myrtle rust ( <i>Puccinia psidii</i> ) among species of Myrtaceae	DNA barcode	ITS, ETS, <i>matK</i>	114	13	Sim
Murillo-A <i>et al.</i>	2016	Explaining disjunct distributions in the flora of southern South America: evolutionary history and biogeography of <i>Myrceugenia</i> (Myrtaceae)	Biogeografia e Filogenia	ITS, ETS, <i>matK</i> , <i>rpl32-trnL</i> , <i>trnQ-rps16</i> , <i>rpl16</i>	89	87	Não
Parra-o e Bohórquez-Osorio	2016	Effectiveness of DNA barcoding markers in the description of a new and unusual calyprate species of <i>Myrcianthes</i> (Myrtaceae)	DNA barcode	ITS, <i>rbcL</i> , <i>matK</i>	1	1	Sim

<b>Autores</b>	<b>Ano</b>	<b>Título</b>	<b>Temática Central</b>	<b>Marcadores</b>	<b>Nº Total de Espécies</b>	<b>Nº de Espécies de Myrteae</b>	<b>Sequências Novas e Publicadas</b>
Santos <i>et al.</i>	2016	Phylogeny, morphology and circumscription of <i>Myrcia</i> sect. <i>Sympodiomyrcia</i> ( <i>Myrcia</i> s.l., Myrtaceae)	Filogenia e Morfologia	ITS, <i>psbA-trnH</i> , <i>ndhF</i> , <i>trnL-trnF</i> , <i>trnQ-rps16</i>	94	94	Sim
Wilson <i>et al.</i>	2016	Phylogenetic Relationships in <i>Calyptanthes</i> (Myrtaceae) with Particular Emphasis on its Monophyly Relative to <i>Myrcia</i> s. l.	Filogenia	ITS, <i>psbA-trnH</i> , <i>trnL-trnF</i> , <i>trnQ-rps16</i>	74	74	Sim
Santos <i>et al.</i>	2017	Biogeographical patterns of <i>Myrcia</i> s.l. (Myrtaceae) and their correlation with geological and climatic history in the Neotropics	Biogeografia e Filogenia	ITS, <i>psbA-trnH</i> , <i>ndhF</i> , <i>trnL-trnF</i> , <i>trnQ-rps16</i>	173	173	Não
Vasconcelos <i>et al.</i>	2017	Links between parallel evolution and systematic complexity in angiosperms—A case study of floral development in <i>Myrcia</i> s.l. (Myrtaceae)	Filogenia e Ontogenia	ITS, <i>psbA-trnH</i> , <i>ndhF</i> , <i>trnL-trnF</i> , <i>trnQ-rps16</i>	69	69	Não
Vasconcelos <i>et al.</i>	2017	Myrteae phylogeny, calibration, biogeography and diversification patterns: Increased understanding in the most species rich tribe of Myrtaceae	Biogeografia e Filogenia	ITS, <i>psbA-trnH</i> , <i>matK</i> , <i>ndhF</i> , <i>trnL-trnF</i> , <i>trnQ-rps16</i> , <i>rpl16</i> , <i>rpl32-trnL</i>	114	99	Sim
Mazine <i>et al.</i>	2018	Phylogeny and biogeography of the hyper-diverse genus <i>Eugenia</i> (Myrtaceae: Myrteae), with emphasis on <i>E.</i> sect. <i>Umbellatae</i> , the most unmanageable clade	Biogeografia e Filogenia	ITS, <i>psbA-trnH</i> , <i>rpl16</i> , <i>rpl32-trnL</i> , <i>trnQ-rps16</i>	100	100	Sim
Nadra <i>et al.</i>	2018	Evolution of pollination by frugivorous birds in Neotropical Myrtaceae	Filogenia e Morfologia	ITS, ETS, <i>matK</i> , <i>psbA-trnH</i>	85	76	Sim
Amorim <i>et al.</i>	2019	Advanced understanding of phylogenetic relationships, morphological evolution and biogeographic history of the mega-diverse plant genus <i>Myrcia</i> and its relatives (Myrtaceae: Myrteae)	Biogeografia e Filogenia	ITS, ETS, <i>psbA-trnH</i> , <i>ndhF</i> , <i>matK</i> , <i>rpl16</i> , <i>trnQ-rps16</i> , <i>rpl32-trnL</i> , <i>trnL-trnF</i>	253	251	Sim
Giaretta <i>et al.</i>	2019	Calyx (con)fusion in a hyper-diverse genus: Parallel evolution of unusual flower patterns in <i>Eugenia</i> (Myrtaceae)	Filogenia e Ontogenia	ITS, <i>psbA-trnH</i> , <i>rpl16</i> , <i>rpl32-trnL</i> , <i>trnQ-rps16</i>	79	79	Sim



<b>Autores</b>	<b>Ano</b>	<b>Título</b>	<b>Temática Central</b>	<b>Marcadores</b>	<b>Nº Total de Espécies</b>	<b>Nº de Espécies de Myrteae</b>	<b>Sequências Novas e Publicadas</b>
Giaretta <i>et al.</i>	2019	Phylogenetic Placement of New Species with Fused Calyx Reveals Homoplastic Character in <i>Eugenia</i> (Myrtaceae)	Filogenia	ITS, ETS, <i>psbA-trnH</i>	54	54	Sim
Flickinger <i>et al.</i>	2020	A phylogenetic survey of Myrtaceae in the Greater Antilles with nomenclatural changes for some endemic species	Biogeografia e Filogenia	ITS, <i>psbA-trnH</i> , <i>trnL-trnF</i> , <i>ndhF-rpl32</i>	160	154	Sim
Lima <i>et al.</i>	2021	Phylogeny and biogeography of <i>Myrcia</i> sect. <i>Aguava</i> (Myrtaceae, Myrteae) based on phylogenomic and Sanger data provide evidence for a Cerrado origin and geographically structured clades	Biogeografia e Filogenia	ITS, <i>psbA-trnH</i> , <i>ndhF</i> , <i>trnL-trnF</i> , <i>trnQ-rps16</i>	58	58	Sim
Lima dos Santos <i>et al.</i>	2021	Phylogenetic and biogeographic analysis in <i>Myrcia</i> sect. <i>Myrcia</i> ( <i>Myrcia</i> S.L., Myrtaceae) with focus on highly polyphyletic <i>Myrcia splendens</i>	Biogeografia e Filogenia	ITS, <i>psbA-trnH</i> , <i>ndhF</i> , <i>trnL-trnF</i> , <i>trnQ-rps16</i>	127	127	Sim *
Uehbe De Oliveira <i>et al.</i>	2021	Understanding molecular relationships in <i>Campomanesia</i> Ruiz & Pav. (Myrtaceae): emphasizing the <i>C. xanthocarpa</i> complex based on multiple accessions	Filogenia	ITS	21	21	Sim **
Proença <i>et al.</i>	2022	Diversity, phylogeny and evolution of the rapidly evolving genus <i>Psidium</i> L. (Myrtaceae, Myrteae)	Filogenia	ITS, ETS, <i>psbA-trnH</i> , <i>ndhF</i>	49	48	Sim
Wati <i>et al.</i>	2022	Filling gaps of reference DNA barcodes in <i>Syzygium</i> from rainforest fragments in Sumatra	DNA barcode	ITS, <i>rbcL</i> , <i>matK</i>	27	3	Sim

\* Novas sequências foram geradas, mas não foi possível rastreá-las no GenBank (os autores forneceram apenas o código das amostras de DNA depositadas no Jardim Botânico Real de Kew - RBG Kew).

\*\* As sequências ficaram disponíveis no GenBank somente após a execução deste trabalho, de modo que as respectivas amostras não foram contabilizadas nas análises.

**APÊNDICE B - BIBLIOTECA DE REFERÊNCIA DO TRABALHO, INCLUINDO AS ESPÉCIES FILTRADAS DURANTE A CURADORIA E OS CÓDIGOS DE ACESSO DE CADA MARCADOR NO GENBANK**

(continua)

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Bsal01	Blepharocalycinae	<i>Blepharocalyx</i>	<i>Blepharocalyx salicifolius</i> (Kunth) O. Berg	Uruguai	Landrum 11232 (ASU)	JN661035		JN660935
Bsal02	Blepharocalycinae	<i>Blepharocalyx</i>	<i>Blepharocalyx salicifolius</i> (Kunth) O. Berg	Argentina	Negritto 927 (CONC)	JN661034		JN660936
Bsal03	Blepharocalycinae	<i>Blepharocalyx</i>	<i>Blepharocalyx salicifolius</i> (Kunth) O. Berg	Brasil	E. Lucas 78 (K)	AM489979	AM489815	AM234084
Bsal04	Blepharocalycinae	<i>Blepharocalyx</i>	<i>Blepharocalyx salicifolius</i> (Kunth) O. Berg	-	Katinas 200 (LPAG)	AY521531		
Bsal05	Blepharocalycinae	<i>Blepharocalyx</i>	<i>Blepharocalyx salicifolius</i> (Kunth) O. Berg	-	Kew DNA bank code 46211		MF954275	MF954018
Bsal06	Blepharocalycinae	<i>Blepharocalyx</i>	<i>Blepharocalyx salicifolius</i> (Kunth) O. Berg	-	P. S. De-Carvalho 579 (UB)		MH835450	
Ebre01	Eugeniinae	<i>Eugenia</i>	<i>Eugenia brevistyla</i> D. Legrand	Brasil	Mazine 993 (ESA, K)		KJ469663	KJ187614
Ebre02	Eugeniinae	<i>Eugenia</i>	<i>Eugenia brevistyla</i> D. Legrand	Brasil	A. Giaretta 1493 (K, SPF)		MN887368	MN296371
Ebre03	Eugeniinae	<i>Eugenia</i>	<i>Eugenia brevistyla</i> D. Legrand	Brasil, RS, Torres	ICN<BRA>:167434		JQ033338	
Ebre04	Eugeniinae	<i>Eugenia</i>	<i>Eugenia brevistyla</i> D. Legrand	Brasil, SP, Sete Barras	ESA: 109420			MG707977
Eflo01	Eugeniinae	<i>Eugenia</i>	<i>Eugenia florida</i> DC.	Guiana Francesa	Lucas 106 (K)		AM489830	AM234090
Eflo02	Eugeniinae	<i>Eugenia</i>	<i>Eugenia florida</i> DC.	Brasil, RS, Porto Alegre	ICN<BRA>:167429		JQ033340	
Eflo03	Eugeniinae	<i>Eugenia</i>	<i>Eugenia florida</i> DC.	Brasil, PR, Cerro Azul	Mazine 965 (ESA, K, UEC)		KJ469671	KJ187622
Eflo04	Eugeniinae	<i>Eugenia</i>	<i>Eugenia florida</i> DC.	Brasil, SP, Gália	ESA: 110057			MG708056

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Einv01	Eugeniinae	<i>Eugenia</i>	<i>Eugenia involucrata</i> DC.	Brasil, SP	ICN<BRA>: 167417		JQ033333	
Einv02	Eugeniinae	<i>Eugenia</i>	<i>Eugenia involucrata</i> DC.	Brasil, MG, Uberlândia	Bunger 551 (BHCB)		KX789294	KX789281
Einv03	Eugeniinae	<i>Eugenia</i>	<i>Eugenia involucrata</i> DC.	Brasil, DF	T. Vasconcelos 256 (RB)		MF954293	MF954035
Einv04	Eugeniinae	<i>Eugenia</i>	<i>Eugenia involucrata</i> DC.	Brasil	Mazine 1018 (ESA, K)		KJ469665	KJ187616
Einv05	Eugeniinae	<i>Eugenia</i>	<i>Eugenia involucrata</i> DC.		-		Indisponível	
Einv06	Eugeniinae	<i>Eugenia</i>	<i>Eugenia involucrata</i> DC.		-		Indisponível	
Einv07	Eugeniinae	<i>Eugenia</i>	<i>Eugenia involucrata</i> DC.		-		Indisponível	
Einv08	Eugeniinae	<i>Eugenia</i>	<i>Eugenia involucrata</i> DC.		-		Indisponível	
Emon01	Eugeniinae	<i>Eugenia</i>	<i>Eugenia monticola</i> (Sw.) DC.	República Dominicana	Lucas 1117 (K)		MH446064	MH446007
Emon02	Eugeniinae	<i>Eugenia</i>	<i>Eugenia monticola</i> (Sw.) DC.	República Dominicana	T. Vasconcelos 566 (K)		MF954295	MF954037
Emon03	Eugeniinae	<i>Eugenia</i>	<i>Eugenia monticola</i> (Sw.) DC.	República Dominicana (cultivada)	Flickinger 2015-DR-27 (FTG)		MN295112	MN295295
Emon04	Eugeniinae	<i>Eugenia</i>	<i>Eugenia monticola</i> (Sw.) DC.	Porto Rico	R.A. Muscarella 104 (US)	KJ012591	KJ426730	
Emon05	Eugeniinae	<i>Eugenia</i>	<i>Eugenia monticola</i> (Sw.) DC.	Costa Rica	BioBot11708	JQ588484		
Emon06	Eugeniinae	<i>Eugenia</i>	<i>Eugenia monticola</i> (Sw.) DC.	Costa Rica	BioBot11707	JQ588483		
Emon07	Eugeniinae	<i>Eugenia</i>	<i>Eugenia monticola</i> (Sw.) DC.	Costa Rica	BioBot10796	JQ588482		
Emon08	Eugeniinae	<i>Eugenia</i>	<i>Eugenia monticola</i> (Sw.) DC.	Costa Rica	BioBot10795	JQ588481		
Epar01	Eugeniinae	<i>Eugenia</i>	<i>Eugenia paracatuana</i> O. Berg	Brasil, GO	Faria 2928 (UB)		MH446057	MH446000
Epar02	Eugeniinae	<i>Eugenia</i>	<i>Eugenia paracatuana</i> O. Berg	Brasil, SP	Silva 64 (SORO)		MH446068	MH446011
Epar03	Eugeniinae	<i>Eugenia</i>	<i>Eugenia paracatuana</i> O. Berg	Brasil, GO	P.O. Rosa 1399		MF954297	MF954039

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Epet01	Eugeniinae	<i>Eugenia</i>	<i>Eugenia petaloidea</i> Giaretta & B.S. Amorim	Brasil	B.S. Amorim 1765 [3] (UFP)		MH341169	MH340543
Epet02	Eugeniinae	<i>Eugenia</i>	<i>Eugenia petaloidea</i> Giaretta & B.S. Amorim	Brasil	B.S. Amorim 1765 [4] (UFP)		MH341170	MH340544
Epet03	Eugeniinae	<i>Eugenia</i>	<i>Eugenia petaloidea</i> Giaretta & B.S. Amorim	Brasil	B.S. Amorim 1765 [23] (UFP)		MH341167	MH340542
Epet04	Eugeniinae	<i>Eugenia</i>	<i>Eugenia petaloidea</i> Giaretta & B.S. Amorim	Brasil	B.S. Amorim 1765 [24] (UFP)		MH341168	
Epet05	Eugeniinae	<i>Eugenia</i>	<i>Eugenia petaloidea</i> Giaretta & B.S. Amorim	Brasil	B.S. Amorim 1765 [106] (UFP)		MH341166	MH340541
Esti01	Eugeniinae	<i>Eugenia</i>	<i>Eugenia stipitata</i> McVaugh	Guiana Francesa	Holst 8872 (CAY, SEL)		KJ469694	KJ187645
Esti02	Eugeniinae	<i>Eugenia</i>	<i>Eugenia stipitata</i> McVaugh	Singapura (cultivada)	Vasconcelos 677 (K)		MF954301	MF954043
Esti03	Eugeniinae	<i>Eugenia</i>	<i>Eugenia stipitata</i> McVaugh	EUA, Flórida (cultivada)	Flickinger 90 (FTG)		MN295129	MN295312
Euni01	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	Nova Zelândia	NZFRI 29277	KM065110		KM064791
Euni02	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	Nova Zelândia	AK258847	KM065113		
Euni03	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	Nova Zelândia	AK241422	KM065173		KM064994
Euni04	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	Nova Zelândia	NZFRI 29293	KM065203		
Euni05	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	Nova Zelândia	AK319813			KM064828
Euni06	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	Reino Unido (cultivada)	Lucas 207 (K)	AM489986	AM489828	AM234088
Euni07	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	Brasil, RS, Palmares	ICN<BRA>: 167404		JQ033327	
Euni08	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	EUA, Flórida	Flickinger 2015-FL-1 (FTG)		MN295133	MN295316

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Euni09	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	EUA, Flórida	J.R. Abbott 23855 (FLAS)	GU135006	GU135338	
Euni10	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	Brasil, SP, Gália	ESA: 109949	MG718959		MG708059
Euni11	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	EUA, Flórida	Trotta 950534 (FLAS)		MH621900	
Euni12	Eugeniinae	<i>Eugenia</i>	<i>Eugenia uniflora</i> L.	-	PGW 1335 (NSW)	AF368207.3*		
Malp01	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia alpigena</i> (DC.) Landrum	Brasil, MG	E. Lucas 167 (K)	JN660991		JN660892
Malp02	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia alpigena</i> var. <i>longifolia</i>	Brasil	R. Harley 26218 (ASU)	JN660992		JN660893
Malp03	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia alpigena</i> var. <i>fuliginea</i>	Brasil	G. Hatschbach 59697 (ASU)	JN660990		JN660891
Mmyr01	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia myrcioides</i> (Cambess.) O.Berg	Brasil	E. Lucas 503 (K)	JN661014		JN660915
Mmyr02	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia myrcioides</i> (Cambess.) O.Berg	Brasil	E. Lucas 82 (K)	AM490000		AM234097
Mmyr03	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia myrcioides</i> var. <i>acrophylla</i>	Brasil	O. Ribas 229 (ASU)	JN661012		JN660913
Mmyr04	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia myrcioides</i> (Cambess.) O.Berg	Brasil, SP, Sete Barras	ESA: 109387	MG718712		MG708014
Mova01	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia ovata</i> (Hook. & Arn.) O.Berg	Escócia (cultivada)	RBGE 1998-2353C (E)		AM489852	AM234096
Mova02	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia ovata</i> var. <i>acutata</i>	Brasil	F. Chagas 1979 (ASU)	JN661017		JN660918
Mova03	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia ovata</i> var. <i>nannophylla</i>	Chile	M. Mihoc 5162 (CONC)	JN661019.2		JN660920

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Mova04	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia ovata</i> var. <i>ovata</i>	Chile	F. Gardner 19 (CONC)	JN661021		JN660922
Mova05	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia ovata</i> var. <i>regnelliana</i>	Brasil	V. Souza 10621 (ASU)	JN661020		JN660921
Mova06	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia ovata</i> var. <i>regnelliana</i>	Brasil	J. Silva 18 (ASU)	JN661037.2		
Mova07	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia ovata</i> var. <i>ovata</i>	-	H. Retamales 11801 (EIF)		KX522595	
Mova08	Luminae	<i>Myrceugenia</i>	<i>Myrceugenia ovata</i> var. <i>nannophylla</i>	-	Crawford & Baeza 157851 (CONC)		KX522594	
Mama01	Myrciinae	<i>Myrcia</i>	<i>Myrcia amazonica</i> DC.	Brasil, PR	E. Lucas 189 (K)	JN091304	JN091403	JN091212
Mama02	Myrciinae	<i>Myrcia</i>	<i>Myrcia amazonica</i> DC.	Brasil, SP	E. Lucas 59 (K)	JN091305	JN091404	JN091213
Mama03	Myrciinae	<i>Myrcia</i>	<i>Myrcia amazonica</i> DC.	Guiana Francesa	Prevost 4751 (K)	JN091306	JN091405	JN091214
Mama04	Myrciinae	<i>Myrcia</i>	<i>Myrcia amazonica</i> DC.	Brasil, PR	E. Lucas 130 (K)		JN091406	JN091215
Mama05	Myrciinae	<i>Myrcia</i>	<i>Myrcia amazonica</i> DC.	Brasil, SP, Cananéia	ESA: 109274	MG719057		MG708083
Mcit01	Myrciinae	<i>Myrcia</i>	<i>Myrcia citrifolia</i> (Aubl.) Urb.	Ilhas Virgens Britânicas	Pollard 1193 (K)		JN091410	JN091219
Mcit02	Myrciinae	<i>Myrcia</i>	<i>Myrcia citrifolia</i> (Aubl.) Urb.	República Dominicana, Monseñor Nouel	Flickinger 2015-DR-26 (FTG)		MN295156	MN295339
Mcit03	Myrciinae	<i>Myrcia</i>	<i>Myrcia citrifolia</i> (Aubl.) Urb.	República Dominicana	A.C. Araujo 1791 (K)		MW023354	MW023968
Mgui01	Myrciinae	<i>Myrcia</i>	<i>Myrcia guianensis</i> (Aubl.) DC.	Brasil	Lucas 141 (K)		JN091415	JN091224
Mgui02	Myrciinae	<i>Myrcia</i>	<i>Myrcia guianensis</i> (Aubl.) DC.	Brasil	Harley 50307 (K)		JN091416	JN091225
Mgui03	Myrciinae	<i>Myrcia</i>	<i>Myrcia guianensis</i> (Aubl.) DC.	Brasil, MG	D.F. Lima 468 (UPCB)		MW023352	MW023966

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Mgui04	Myrciinae	<i>Myrcia</i>	<i>Myrcia guianensis</i> (Aubl.) DC.	Brasil, MG	D.F. Lima 517 (UPCB)	MW023348	MW023962	
Mgui05	Myrciinae	<i>Myrcia</i>	<i>Myrcia guianensis</i> (Aubl.) DC.	Brasil, GO	D.F. Lima 530 (UPCB)	MW023357	MW023971	
Mgui06	Myrciinae	<i>Myrcia</i>	<i>Myrcia guianensis</i> (Aubl.) DC.	Brasil, BA	D.F. Lima 512 (UPCB)	MW023370	MW023984	
Mmul01	Myrciinae	<i>Myrcia</i>	<i>Myrcia multiflora</i> (Lam.) DC.	Brasil	Lucas 65 (K)	AM489858	AM234117	
Mmul02	Myrciinae	<i>Myrcia</i>	<i>Myrcia multiflora</i> (Lam.) DC.	Brasil	V.G. Staggemeier 422 (UB)	KP722286	KP722379	
Mmul03	Myrciinae	<i>Myrcia</i>	<i>Myrcia multiflora</i> (Lam.) DC.	Brasil	V.G. Staggemeier 863 (UB, K)	KP722293	KP722386	
Mmul04	Myrciinae	<i>Myrcia</i>	<i>Myrcia multiflora</i> (Lam.) DC.	Brasil	V.G. Staggemeier 867 (UB)	KP722294	KP722387	
Mruf01	Myrciinae	<i>Myrcia</i>	<i>Myrcia rufipes</i> DC.	Brasil	Lucas 280 (K)	JN091430	JN091239	
Mruf02	Myrciinae	<i>Myrcia</i>	<i>Myrcia rufipes</i> DC.	Brasil, MG	D.F. Lima 548 (UPCB)	MW023372	MW023986	
Mruf03	Myrciinae	<i>Myrcia</i>	<i>Myrcia rufipes</i> DC.	Brasil, MG	D.F. Lima 534 (UPCB)	MW023358	MW023972	
Mruf04	Myrciinae	<i>Myrcia</i>	<i>Myrcia rufipes</i> DC.	Brasil, MG	D.F. Lima 523 (UPCB)	MW023345	MW023959	
Mspl01	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	República Dominicana	T. Vasconcelos 587 (K)	MF954395	MF954059	
Mspl02	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Guiana Francesa	Prevost 4716 (K)	AM490006	AM489865	AM234124
Mspl03	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Brasil	Lucas 73 (K)	AM489863	AM234122	
Mspl04	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Brasil	Lucas 165 (K)	JN091433	JN091242	
Mspl05	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Bolívia	Wood 15435 (K)	JN091434	JN091243	
Mspl06	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Brasil, MG, São Gonçalo do Rio Abaixo	T. Vasconcelos 477 (K)	MW023377*	MW023991	

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Mspl07	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Costa Rica	BioBot10549	JQ588499		
Mspl08	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Costa Rica	BioBot10546	JQ588498		
Mspl09	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Costa Rica	BioBot10383	JQ588496		
Mspl10	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Costa Rica	BioBot06044	JQ588492		
Mspl11	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Costa Rica	BioBot06110	JQ588494		
Mspl12	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Costa Rica	BioBot06045	JQ588493		
Mspl13	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Costa Rica	BioBot01485	JQ588489		
Mspl14	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Costa Rica	BioBot01486	JQ588490		
Mspl15	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Costa Rica	BioBot01487	JQ588491		
Mspl16	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Brasil	Blum 10-098 (UPCB)		KF421072	KF420979
Mspl17	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Brasil	Blum 11-003 (UPCB)		KF421073	KF420978
Mspl18	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Brasil	Blum 11-024 (UPCB)		KF421074	KF420977
Mspl19	Myrciinae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.	Porto Rico	N.Swensen 177 (US)	HM446718*		
Mvar01	Myrciinae	<i>Myrcia</i>	<i>Myrcia variabilis</i> DC.	Brasil	Lucas 277 (K)		JN091439	JN091248
Mvar02	Myrciinae	<i>Myrcia</i>	<i>Myrcia variabilis</i> DC.	Brasil, MG	D.F. Lima 365 (UPCB)		MW023347	MW023961
Mvar03	Myrciinae	<i>Myrcia</i>	<i>Myrcia variabilis</i> DC.	Brasil, MG	D.F. Lima 544 (UPCB)		MW023356	MW023970
Mfra01	Eugeniinae	<i>Myrcianthes</i>	<i>Myrcianthes fragrans</i> (Sw.) McVaugh	-	Holst 8862 (SEL)		KJ469705	KJ187655
Mfra02	Eugeniinae	<i>Myrcianthes</i>	<i>Myrcianthes fragrans</i> (Sw.) McVaugh	Ilhas Turcas e Caicos	Hamilton 552 (FTG, K)		MH446046	MH445989
Mfra03	Eugeniinae	<i>Myrcianthes</i>	<i>Myrcianthes fragrans</i> (Sw.) McVaugh	EUA, Flórida (cultivada)	Flickinger 2015-FTG-2 (FTG)		MN295154	MN295337
Mfra04	Eugeniinae	<i>Myrcianthes</i>	<i>Myrcianthes fragrans</i> (Sw.) McVaugh	EUA, Flórida	Trotta 950247 (FLAS)		MH621999	



Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Mpun01	Eugeniinae	<i>Myrcianthes</i>	<i>Myrcianthes pungens</i> (O.Berg) D.Legrand	Argentina	Tressens et al 5481 (K)	AM489867	AM234099	
Mpun02	Eugeniinae	<i>Myrcianthes</i>	<i>Myrcianthes pungens</i> (O.Berg) D.Legrand	Brasil	Forster 1013 (ESA)	KJ469706	KJ187656	
Mpun03	Eugeniinae	<i>Myrcianthes</i>	<i>Myrcianthes pungens</i> (O.Berg) D.Legrand	Brasil, DF	Faria 2759 (UB)	MH446059	MH446002	
Mpun04	Eugeniinae	<i>Myrcianthes</i>	<i>Myrcianthes pungens</i> (O.Berg) D.Legrand	-	Rodriguez S/N b (JUA)		MH716440	
Mflo01	Pliniinae	<i>Myrciaria</i>	<i>Myrciaria floribunda</i> (H.West ex Willd.) O.Berg	Brasil	Vasconcelos 388	MF954398	MF954062	
Mflo02	Pliniinae	<i>Myrciaria</i>	<i>Myrciaria floribunda</i> (H.West ex Willd.) O.Berg	Porto Rico, Quebradillas	Flickinger 87 (FTG)	MN295163	MN295346	
Mflo03	Pliniinae	<i>Myrciaria</i>	<i>Myrciaria floribunda</i> (H.West ex Willd.) O.Berg	Porto Rico	118738361 (US)	KJ426839		
Mflo04	Pliniinae	<i>Myrciaria</i>	<i>Myrciaria floribunda</i> (H.West ex Willd.) O.Berg	Brasil, SP, Gália	ESA: 110085	MG718987		MG708060
Mflo05	Pliniinae	<i>Myrciaria</i>	<i>Myrciaria floribunda</i> (H.West ex Willd.) O.Berg	Brasil, SP, Assis	ESA: 91793	MG718859		MG708048
Mflo06	Pliniinae	<i>Myrciaria</i>	<i>Myrciaria floribunda</i> (H.West ex Willd.) O.Berg	Brasil, SP, Sete Barras	ESA: 109405	MG718720		MG708022
Matr01	Pimentinae	<i>Myrrhinium</i>	<i>Myrrhinium atropurpureum</i> Schott	Brasil, RJ	Costa 594 (EAC, UB)	MF954400	MF954064	
Matr02	Pimentinae	<i>Myrrhinium</i>	<i>Myrrhinium atropurpureum</i> var. octandrum Benth.	-	SI:MGN 2	MH713602	MH716441	
Matr03	Pimentinae	<i>Myrrhinium</i>	<i>Myrrhinium atropurpureum</i> var. octandrum Benth.	-	SI:MGN 30	MH713603	MH716442	

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Pdio01	Pimentinae	<i>Pimenta</i>	<i>Pimenta dioica</i> (L.) Merr.	Nova Zelândia	NZFRI 29338	KM065289		KM064833
Pdio02	Pimentinae	<i>Pimenta</i>	<i>Pimenta dioica</i> (L.) Merr.	Nova Zelândia	NZFRI 29276	KM065298		KM065057
Pdio03	Pimentinae	<i>Pimenta</i>	<i>Pimenta dioica</i> (L.) Merr.	Reino Unido (cultivada)	Lucas 212 (K)	AM490011	AM489874	AM234081
Pdio04	Pimentinae	<i>Pimenta</i>	<i>Pimenta dioica</i> (L.) Merr.	EUA, Flórida (cultivada)	Flickinger 2016-FTG-3 (FTG)		MN295165	MN295348
Pdio05	Pimentinae	<i>Pimenta</i>	<i>Pimenta dioica</i> (L.) Merr.	EUA, Missouri (cultivada)	Carlsen 3255 (MO)	OL690114	OL312147	
Prac01	Pimentinae	<i>Pimenta</i>	<i>Pimenta racemosa</i> (Mill.) J.W.Moore	-	Holst 8866 (K)	AM490012	AM489875	AM234082
Prac02	Pimentinae	<i>Pimenta</i>	<i>Pimenta racemosa</i> (Mill.) J.W.Moore	-	UNSW 22372	AY521545		
Prac03	Pimentinae	<i>Pimenta</i>	<i>Pimenta racemosa</i> var. <i>grisea</i>	Porto Rico, Maricao	Flickinger 72 (FTG)		MN295167	MN295350
Prac04	Pimentinae	<i>Pimenta</i>	<i>Pimenta racemosa</i> var. <i>ozua</i>	República Dominicana (cultivada)	Flickinger 2015-DR-30 (FTG)		MN295169	MN295352
Prac05	Pimentinae	<i>Pimenta</i>	<i>Pimenta racemosa</i> var. <i>racemosa</i>	EUA, Flórida (cultivada)	J. Flickinger 89 (FTG)		MN295171	MN295354
Prac06	Pimentinae	<i>Pimenta</i>	<i>Pimenta racemosa</i> (Mill.) J.W.Moore	Porto Rico	R.A. Muscarella 113 (US)	KJ012715		
Pcau01	Pliniinae	<i>Plinia</i>	<i>Plinia cauliflora</i> (Mart.) Kausel	Reino Unido (cultivada)	E. Lucas 210 (K)	AM490007		AM234093
Pcau02	Pliniinae	<i>Plinia</i>	<i>Plinia cauliflora</i> (Mart.) Kausel	Nova Zelândia	NZFRI 29257	KM065264		KM064982
Pcau03	Pliniinae	<i>Plinia</i>	<i>Plinia cauliflora</i> (Mart.) Kausel	Nova Zelândia	NZFRI 29271	KM065319		KM064870
Pcau04	Pliniinae	<i>Plinia</i>	<i>Plinia cauliflora</i> (Mart.) Kausel	EUA, Missouri (cultivada)	Carlsen 3084 (MO)	OL690122		
Paus01	Pimentinae	<i>Psidium</i>	<i>Psidium australe</i> var. <i>argenteum</i>	Brasil, MG, Diamantina	J.E.Q. Faria 6689 (UB)		MZ682332	MZ710576

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Paus02	Pimentinae	<i>Psidium</i>	<i>Psidium australe</i> var. <i>argenteum</i>	Brasil, MG Santa Cruz de Minas	I.R. Costa 540 (UB)		MZ682346	
Paus03	Pimentinae	<i>Psidium</i>	<i>Psidium australe</i> Cambess.	Brasil, SP, Assis	I.R. Costa 649 (EAC)		MZ682333	MZ710578
Paus04	Pimentinae	<i>Psidium</i>	<i>Psidium australe</i> Cambess.	Brasil, MG, Diamantina	C.E.B. Proença 5376 (UB)			MZ710577
Paus05	Pimentinae	<i>Psidium</i>	<i>Psidium australe</i> var. <i>suffruticosum</i>	Brasil, MG, Santa Cruz de Minas	I.R. Costa 538 (UB)			MZ710603
Pcat01	Pimentinae	<i>Psidium</i>	<i>Psidium cattleianum</i> Sabine	Nova Zelândia	NZFRI 29007	KM065073		KM065002
Pcat02	Pimentinae	<i>Psidium</i>	<i>Psidium cattleianum</i> Sabine	Nova Zelândia	NZFRI 29059	KM065132		KM064972
Pcat03	Pimentinae	<i>Psidium</i>	<i>Psidium cattleianum</i> Sabine	Reino Unido (cultivada)	Lucas 213 (K)	AM490014	AM489878	AM234080
Pcat04	Pimentinae	<i>Psidium</i>	<i>Psidium cattleianum</i> Sabine	Brasil, SP, Campinas (cultivada)	I.R. Costa 486 (UEC)		MZ682337	MZ710583
Pcat05	Pimentinae	<i>Psidium</i>	<i>Psidium cattleianum</i> Sabine	Brasil, SP, Campinas (cultivada)	I.R. Costa 819b (UB)		MZ682345	MZ710604
Pcat06	Pimentinae	<i>Psidium</i>	<i>Psidium cattleianum</i> Sabine	EUA, Flórida (cultivada)	Flickinger 2015-FL-02 (FTG)		MN295176	MN295359
Pcat07	Pimentinae	<i>Psidium</i>	<i>Psidium cattleianum</i> Sabine	EUA, Flórida	J.R. Abbott 24905 (FLAS)	GU135031	GU135360*	
Pcat08	Pimentinae	<i>Psidium</i>	<i>Psidium cattleianum</i> Sabine	Brasil, RJ	I.R. Costa 606 (UB)			MZ710584
Pcat09	Pimentinae	<i>Psidium</i>	<i>Psidium cattleianum</i> Sabine	Nova Zelândia	NZFRI 26525			KM064916
Pcat10	Pimentinae	<i>Psidium</i>	<i>Psidium cattleianum</i> Sabine	Brasil, RJ, Nova Friburgo	Tuler 553 (RB)		MK310546	
Pgua01	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	Brasil, SP, Campinas (cultivada)	I.R. Costa 476 (UEC)		MK041282	MZ710589
Pgua02	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	África do Sul	ASU199194 (PRU)			AY487283
Pgua03	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	África do Sul, Limpopo	LTM026 (JRAU)	JQ024987		

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						<i>matK</i>	<i>psbA-trnH</i>	ITS
Pgua04	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	EUA, Flórida (cultivada)	Flickinger 2015-FTG-8 (FTG)		MN295177	MN295360
Pgua05	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	EUA, Flórida	M.S. Frank 0541 (FLAS)	GU135087	GU135421	
Pgua06	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	Costa Rica	BioBot10556	JQ588512		
Pgua07	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	Costa Rica	BioBot05032	JQ588510		
Pgua08	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	Costa Rica	BioBot05031	JQ588509		
Pgua09	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	Costa Rica	BioBot10555	JQ588511		
Pgua10	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	Costa Rica	BioBot05030	JQ588508		
Pgua11	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	EUA, Missouri (cultivada)	Carlsen 3227 (MO)	OL690131	OL312278	
Pgua12	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	Nicarágua	NPL0233 (HULE)		HG963647	
Pgua13	Pimentinae	<i>Psidium</i>	<i>Psidium guajava</i> L.	EUA, Havaí, Maui	BISH 687876			AY864898*
Plar01	Pimentinae	<i>Psidium</i>	<i>Psidium laruotteanum</i> Cambess.	Brasil, BA	J.E.Q. Faria 2362		MF954277	
Plar02	Pimentinae	<i>Psidium</i>	<i>Psidium laruotteanum</i> Cambess.	Brasil, DF	I.R. Costa 624 (UB)			MZ710591
Plar03	Pimentinae	<i>Psidium</i>	<i>Psidium laruotteanum</i> Cambess.	Brasil, DF	M. Kuhlmann 34 (UB)		MZ682341	MZ710592
Plar04	Pimentinae	<i>Psidium</i>	<i>Psidium laruotteanum</i> Cambess.	Brasil, DF	C.E.B. Proença 3069 (UB)		MZ682342	MZ710593
Pmac01	Pimentinae	<i>Psidium</i>	<i>Psidium macahense</i> O.Berg	Brasil, ES, Linhares	J.E.Q. Faria 2519 (UB)		MZ682343	MZ710594
Pmac02	Pimentinae	<i>Psidium</i>	<i>Psidium macahense</i> O.Berg	Brasil, ES, Linhares	Tuler 532 (RB)		MK310534	MK313861
Pmac03	Pimentinae	<i>Psidium</i>	<i>Psidium macahense</i> O.Berg	Brasil, ES, Linhares	Tuler 455 (RB)		MK310535	MK313862
Pmac04	Pimentinae	<i>Psidium</i>	<i>Psidium macahense</i> O.Berg	Brasil, ES, Linhares	Tuler 457 (RB)		MK310536	MK313863
Pruf01	Pimentinae	<i>Psidium</i>	<i>Psidium rufum</i> Mart. ex DC.	Brasil, DF (cultivada)	C.E.B. Proença & M.C. Silva Junior 3230 (UB)		MK041285	MZ710599

Acrônimo	Subtribo	Gênero	Espécie (nome aceito)**	Localização	Voucher/Herbário ***	Códigos de acesso no GenBank		
						matK	psbA-trnH	ITS
Pruf02	Pimentinae	<i>Psidium</i>	<i>Psidium rufum</i> Mart. ex DC.	Brasil, RJ, Nova Friburgo	Tuler 536 (RB)		MK310540	MK313867
Pruf03	Pimentinae	<i>Psidium</i>	<i>Psidium rufum</i> Mart. ex DC.	Brasil, RJ, Nova Friburgo	Tuler 538 (RB)		MK310541	MK313868

\* Sequências desconsideradas nas análises por serem muito curtas ou apresentarem diversas bases ambíguas.

\*\* Sinonímias adotadas:

*Blepharocalyx salicifolius* (Kunth) O. Berg = *Blepharocalyx tweediei* (Hook. et Arn.) O. Berg;

*Eugenia brevistyla* D. Legrand = *Eugenia neoaustralis* Sobral;

*Eugenia involucrata* DC. = *Eugenia calycina* Cambess.;

*Myrcia guianensis* (Aubl.) DC. = *Myrcia angustifolia* (O.Berg) Nied. = *Myrcia decrescens* (O.Berg) Mattos = *Myrcia rotundifolia* (O.Berg) Kiaersk;

*Myrcia splendens* (Sw.) DC. = *Myrcia fallax* (DC.) O. Berg = *Myrcia rostrata* DC.;

*Myrcia variabilis* DC. = *Myrcia amethystina* (O. Berg) Kiaersk. = *Myrcia dealbata* DC.;

*Plinia cauliflora* (Mart.) Kausel = *Myrciaria cauliflora* (Mart.) O.Berg.

\*\*\* Abreviaturas: AK - Museu Memorial da Guerra de Auckland; ASU - Herbário da Universidade Estadual do Arizona; BHCb - Herbário da Universidade Federal de Minas Gerais; CAY - Herbário da Instituição de Pesquisa para o Desenvolvimento (IRD) da Guiana Francesa; CONC - Herbário da Universidade de Concepción, Chile; E - Herbário do Jardim Botânico Real de Edimburgo, Escócia; EAC - Herbário da Universidade Federal do Ceará; EIF - Herbário da Universidade do Chile; ESA - Herbário da Escola Superior de Agricultura Luiz de Queiroz; FLAS - Herbário da Universidade da Flórida; FTG - Herbário do *Fairchild Tropical Botanic Garden*; HULE - Herbário da Universidade Nacional Autónoma de León, Nicarágua; ICN - Herbário da Universidade Federal do Rio Grande do Sul; JRAU - Herbário da Universidade de Joanesburgo, África do Sul; JUA - Herbário da Universidade Nacional de Jujuy, Argentina; K - Jardim Botânico Real de Kew, Reino Unido; LPAG - Herbário da Universidade Nacional de La Plata, Argentina; MO - Herbário do Jardim Botânico de Missouri, EUA; NSW - Herbário Nacional de New South Wales; NZFRI - Herbário do Instituto de Pesquisa Florestal da Nova Zelândia; PRU - Herbário da Universidade de Pretoria, África do Sul; RB - Herbário do Jardim Botânico do Rio de Janeiro; SEL - Herbário do *Marie Selby Botanical Gardens*; SI - Instituto de Botânica Darwinion; SORO - Herbário da Universidade Federal de São Carlos, campus Sorocaba; SPF - Herbário do Instituto de Biociências da Universidade de São Paulo; UB - Herbário da Universidade de Brasília; UEC - Herbário da Universidade Estadual de Campinas; UFP - Herbário da Universidade Federal de Pernambuco; UNSW - Herbário da Universidade de New South Wales; UPCB - Herbário da Universidade Federal do Paraná; US - Herbário da *Smithsonian Institution*.

**APÊNDICE C - DISTÂNCIAS INTRAESPECÍFICA MÁXIMA E INTERESPECÍFICA MÍNIMA ENCONTRADAS PARA CADA ESPÉCIE**

(continua)

<b>Marcador</b>	<b>Espécie</b>	<b>Maior distância intraespecífica</b>	<b>Menor distância interespecífica</b>	<b>Espécie(s) mais próxima(s)</b>
<i>matK</i>	<i>Blepharocalyx salicifolius</i>	0,021264	0,010518	<i>Eugenia monticola</i> e <i>Myrcia splendens</i>
	<i>Eugenia involucrata</i>	0,000000	0,009481	<i>Eugenia monticola</i>
	<i>Eugenia monticola</i>	0,025457	0,000000	<i>Myrcia splendens</i>
	<i>Eugenia uniflora</i>	0,001919	0,011429	<i>Eugenia monticola</i> e <i>Myrcia splendens</i>
	<i>Myrceugenia alpigena</i>	0,003827	0,000000	<i>Myrceugenia myrcioides</i> e <i>Myrceugenia ovata</i>
	<i>Myrceugenia myrcioides</i>	0,003791	0,000000	<i>Myrceugenia alpigena</i> e <i>Myrceugenia ovata</i>
	<i>Myrceugenia ovata</i>	0,011550	0,000000	<i>Myrceugenia alpigena</i> e <i>Myrceugenia myrcioides</i>
	<i>Myrcia amazonica</i>	0,004380	0,003776	<i>Eugenia monticola</i> e <i>Myrcia splendens</i>
	<i>Myrcia splendens</i>	0,015221	0,000000	<i>Eugenia monticola</i>
	<i>Myrciaria floribunda</i>	0,005671	0,005674	<i>Eugenia monticola</i> e <i>Myrcia splendens</i>
	<i>Pimenta dioica</i>	0,005671	0,003777	<i>Pimenta racemosa</i>
	<i>Pimenta racemosa</i>	0,000000	0,003777	<i>Pimenta dioica</i>
	<i>Plinia cauliflora</i>	0,000000	0,005674	<i>Eugenia monticola</i>
	<i>Psidium cattleyanum</i>	0,001919	0,005674	<i>Eugenia monticola</i> e <i>Psidium guajava</i>
	<i>Psidium guajava</i>	0,003937	0,005674	<i>Psidium cattleyanum</i>

<b>Marcador</b>	<b>Espécie</b>	<b>Maior distância intraespecífica</b>	<b>Menor distância interespecífica</b>	<b>Espécie(s) mais próxima(s)</b>
<i>psbA-trnH</i>	<i>Blepharocalyx salicifolius</i>	0,068711	0,010170	<i>Myrceugenia ovata</i>
	<i>Eugenia brevistyla</i>	0,017382	0,010688	<i>Eugenia involucrata</i>
	<i>Eugenia florida</i>	0,044063	0,000000	<i>Eugenia paracatuana</i>
	<i>Eugenia involucrata</i>	0,027060	0,010688	<i>Eugenia brevistyla</i>
	<i>Eugenia monticola</i>	0,000000	0,020808	<i>Eugenia petaloidea</i>
	<i>Eugenia paracatuana</i>	0,015031	0,000000	<i>Eugenia florida</i>
	<i>Eugenia petaloidea</i>	0,000000	0,015631	<i>Eugenia involucrata</i>
	<i>Eugenia stipitata</i>	0,004890	0,038622	<i>Eugenia brevistyla</i>
	<i>Eugenia uniflora</i>	0,009678	0,025822	<i>Eugenia petaloidea</i>
	<i>Myrceugenia ovata</i>	0,010601	0,008136	<i>Pimenta dioica</i>
	<i>Myrcia amazonica</i>	0,017021	0,007251	<i>Myrcia guianensis</i>
	<i>Myrcia citrifolia</i>	0,000000	0,000000	<i>Myrcia guianensis, Myrcia rufipes, Myrcia variabilis</i>
	<i>Myrcia guianensis</i>	0,009680	0,000000	<i>Myrcia citrifolia, Myrcia rufipes, Myrcia variabilis</i>
	<i>Myrcia multiflora</i>	0,020222	0,007658	<i>Myrcia rufipes</i>
	<i>Myrcia rufipes</i>	0,002449	0,000000	<i>Myrcia citrifolia, Myrcia guianensis, Myrcia variabilis</i>
	<i>Myrcia splendens</i>	0,013750	0,002651	<i>Myrciaria floribunda</i>
	<i>Myrcia variabilis</i>	0,002408	0,000000	<i>Myrcia citrifolia, Myrcia guianensis, Myrcia rufipes</i>

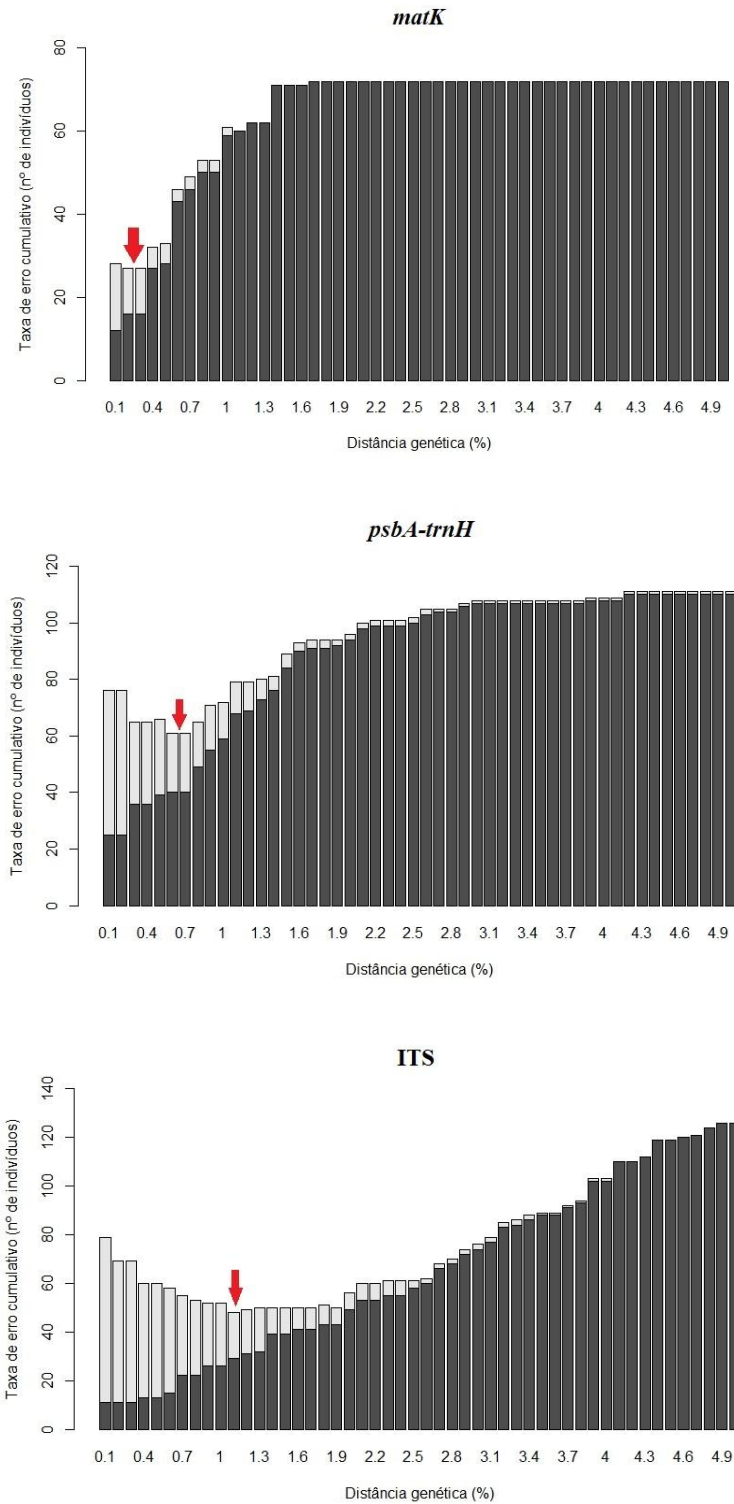
Marcador	Espécie	Maior distância intraespecífica	Menor distância interespecífica	Espécie(s) mais próxima(s)
<i>psbA-trnH</i>	<i>Myrcianthes fragrans</i>	0,005133	0,002562	<i>Myrcianthes pungens</i>
	<i>Myrcianthes pungens</i>	0,010301	0,002562	<i>Myrcianthes fragrans</i>
	<i>Myrciaria floribunda</i>	0,026871	0,002651	<i>Myrcia splendens</i>
	<i>Myrrhinium atropurpureum</i>	0,009702	0,009709	<i>Psidium cattleyanum, Psidium macahense, Psidium rufum</i>
	<i>Pimenta dioica</i>	0,000000	0,007916	<i>Pimenta racemosa</i>
	<i>Pimenta racemosa</i>	0,007712	0,007916	<i>Pimenta dioica</i>
	<i>Psidium australe</i>	0,019724	0,000000	<i>Psidium cattleyanum, Psidium laruotteanum, Psidium macahense, Psidium rufum</i>
	<i>Psidium cattleyanum</i>	0,017229	0,000000	<i>Psidium australe, Psidium laruotteanum, Psidium macahense, Psidium rufum</i>
	<i>Psidium guajava</i>	0,000000	0,014783	<i>Myrrhinium atropurpureum</i>
	<i>Psidium laruotteanum</i>	0,012480	0,000000	<i>Psidium australe, Psidium cattleyanum, Psidium macahense, Psidium rufum</i>
	<i>Psidium macahense</i>	0,000000	0,000000	<i>Psidium australe, Psidium cattleyanum, Psidium laruotteanum, Psidium rufum</i>
<i>Psidium rufum</i>	0,015565	0,000000	<i>Psidium australe, Psidium cattleyanum, Psidium laruotteanum, Psidium macahense</i>	
ITS	<i>Blepharocalyx salicifolius</i>	0,044218	0,022543	<i>Myrcia guianensis, Myrcia rufipes, Myrcia variabilis</i>
	<i>Eugenia brevistyla</i>	0,006933	0,040339	<i>Eugenia involucrata</i>



Marcador	Espécie	Maior distância intraespecífica	Menor distância interespecífica	Espécie(s) mais próxima(s)
ITS	<i>Eugenia florida</i>	0,055029	0,000000	<i>Eugenia paracatuana</i>
	<i>Eugenia involucrata</i>	0,005155	0,038453	<i>Eugenia petaloidea</i>
	<i>Eugenia monticola</i>	0,000000	0,026685	<i>Eugenia petaloidea</i>
	<i>Eugenia paracatuana</i>	0,037460	0,000000	<i>Eugenia florida</i>
	<i>Eugenia petaloidea</i>	0,021942	0,026685	<i>Eugenia monticola</i>
	<i>Eugenia stipitata</i>	0,000000	0,047966	<i>Myrcia rufipes</i>
	<i>Eugenia uniflora</i>	0,010241	0,036809	<i>Eugenia petaloidea</i>
	<i>Myrceugenia alpigena</i>	0,005130	0,005119	<i>Myrceugenia myrcioides</i>
	<i>Myrceugenia myrcioides</i>	0,017275	0,005119	<i>Myrceugenia alpigena</i>
	<i>Myrceugenia ovata</i>	0,024361	0,006840	<i>Myrceugenia myrcioides</i>
	<i>Myrcia amazonica</i>	0,026881	0,028809	<i>Myrcia guianensis, Myrcia variabilis</i>
	<i>Myrcia citrifolia</i>	0,022661	0,008621	<i>Myrcia guianensis, Myrcia rufipes, Myrcia variabilis</i>
	<i>Myrcia guianensis</i>	0,015414	0,000000	<i>Myrcia rufipes, Myrcia variabilis</i>
	<i>Myrcia multiflora</i>	0,012051	0,011996	<i>Myrcia variabilis</i>
	<i>Myrcia rufipes</i>	0,003419	0,000000	<i>Myrcia guianensis, Myrcia variabilis</i>
	<i>Myrcia splendens</i>	0,030103	0,024345	<i>Myrcia variabilis</i>
	<i>Myrcia variabilis</i>	0,003390	0,000000	<i>Myrcia guianensis, Myrcia rufipes</i>

Marcador	Espécie	Maior distância intraespecífica	Menor distância interespecífica	Espécie(s) mais próxima(s)
ITS	<i>Myrcianthes fragrans</i>	0,003581	0,038660	<i>Myrcia guianensis</i>
	<i>Myrcianthes pungens</i>	0,010347	0,034299	<i>Myrceugenia ovata</i>
	<i>Myrciaria floribunda</i>	0,039068	0,025991	<i>Myrcia rufipes</i>
	<i>Myrrhinium atropurpureum</i>	0,005193	0,028064	<i>Psidium cattleyanum</i>
	<i>Pimenta dioica</i>	0,000000	0,043966	<i>Pimenta racemosa</i>
	<i>Pimenta racemosa</i>	0,031259	0,043966	<i>Pimenta dioica</i>
	<i>Plinia cauliflora</i>	0,006811	0,026011	<i>Myrcia guianensis, Myrcia variabilis</i>
	<i>Psidium australe</i>	0,001695	0,040239	<i>Psidium macahense</i>
	<i>Psidium cattleyanum</i>	0,015467	0,019168	<i>Psidium laruotteanum</i>
	<i>Psidium guajava</i>	0,020929	0,040286	<i>Psidium australe</i>
	<i>Psidium laruotteanum</i>	0,017160	0,006923	<i>Psidium rufum</i>
	<i>Psidium macahense</i>	0,003396	0,019009	<i>Psidium rufum</i>
	<i>Psidium rufum</i>	0,001703	0,006923	<i>Psidium laruotteanum</i>

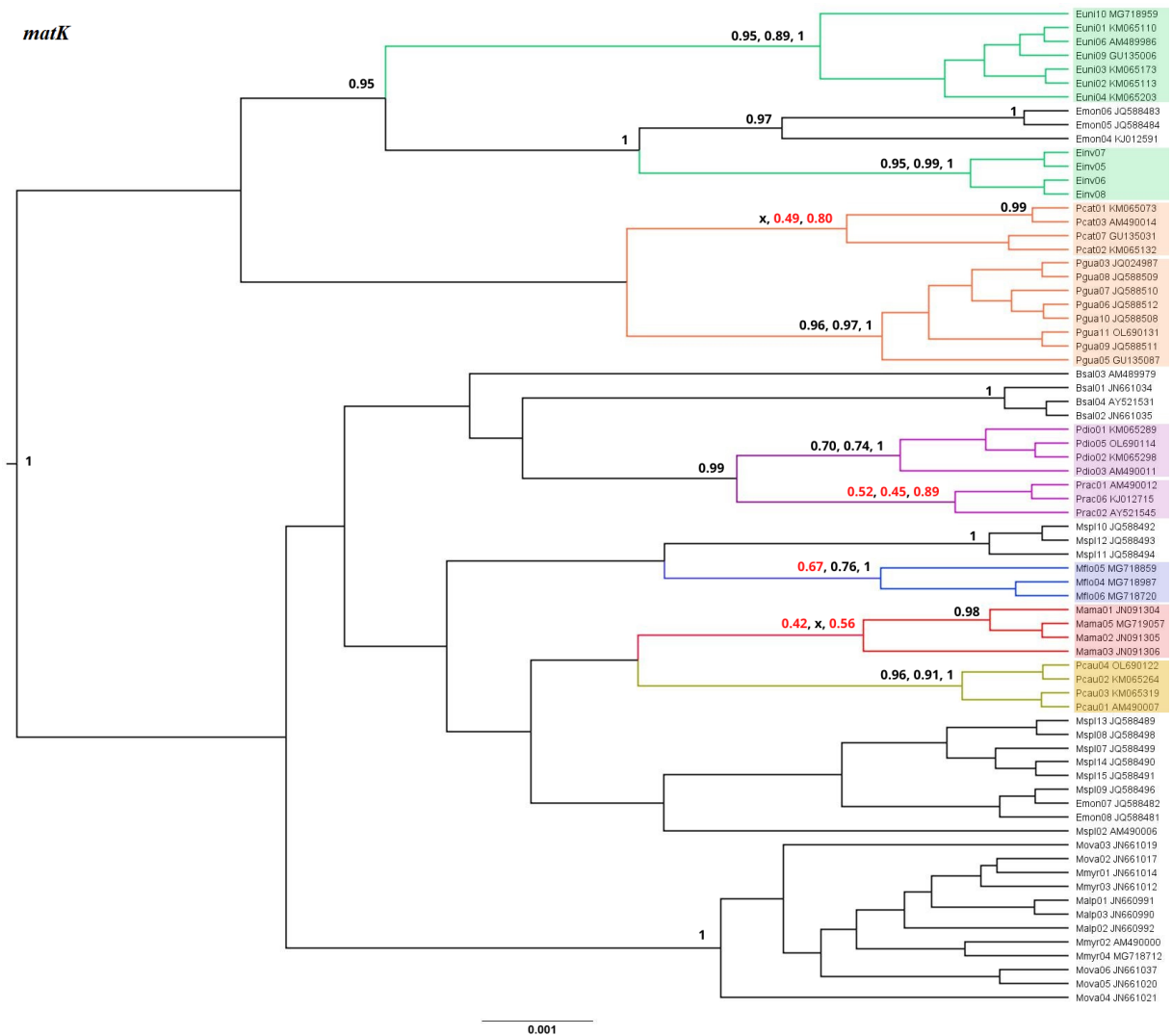
**APÊNDICE D - TAXAS DE ERRO CUMULATIVO (EM Nº DE INDIVÍDUOS)  
CALCULADAS PARA DIFERENTES LIMITES DE CORTE PELO PACOTE *SPIDER*  
NO R\***



\* Nota: limites de corte em intervalos de 0,001 unidades de distância genética (0,1%). As setas em vermelho indicam os valores de limite otimizados, enquanto as barras em cinza claro representam identificações com erros do tipo falso positivo (quando há ausência de um correspondente coespecífico dentro do limite). Já as barras em cinza escuro simbolizam identificações com erros do tipo falso negativo (onde sequências pertencentes a outras espécies se encontram dentro do limite).

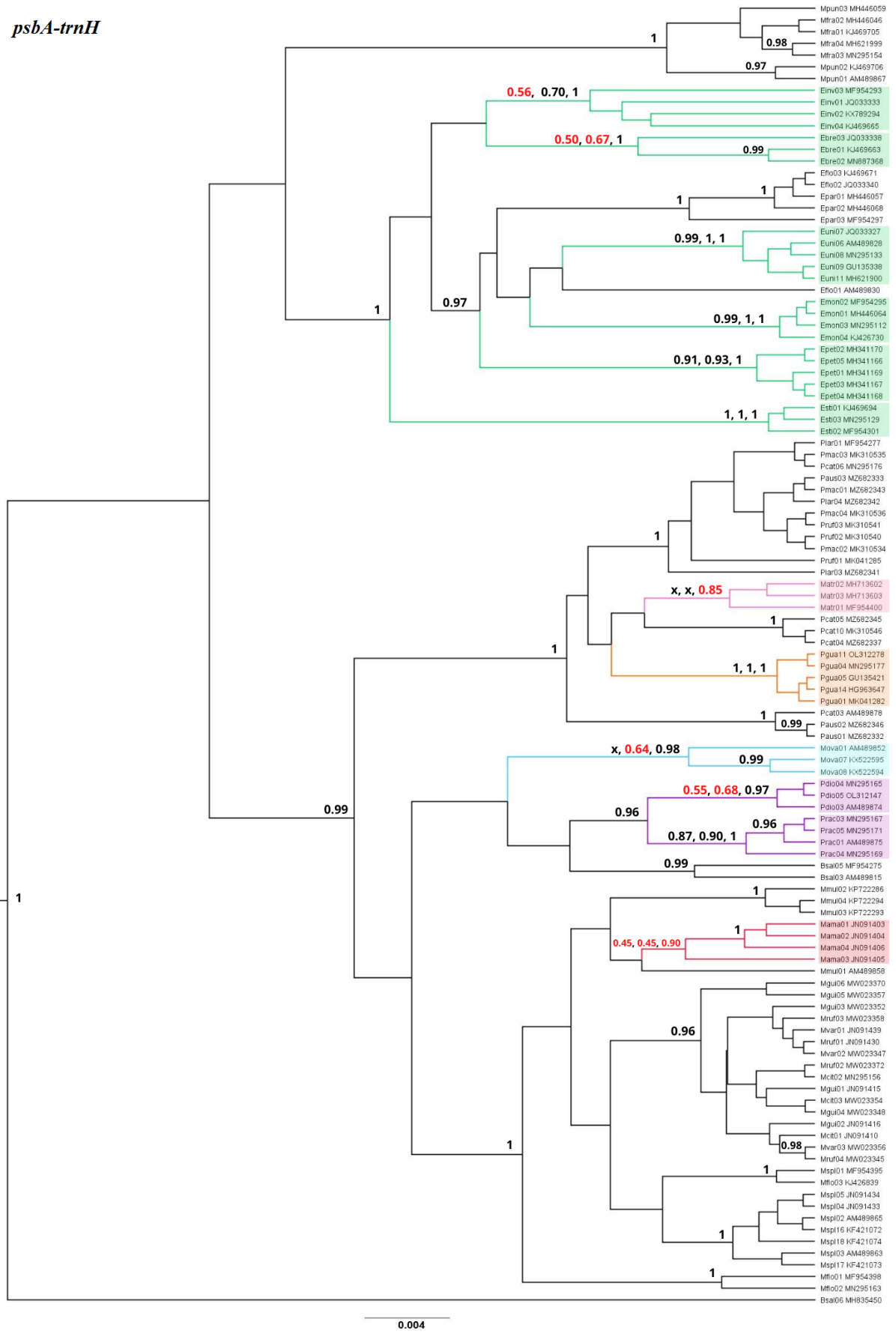
## APÊNDICE E - FILOGENIAS GERADAS POR INFERÊNCIA BAYESIANA\*

(continua)



\* Nota: os clados coloridos denotam espécies que tiveram todos os seus indivíduos agrupados de forma exclusiva, e cada cor está associada a um gênero distinto. Três valores de suporte para cada um desses clados encontram-se dispostos sobre os ramos, simbolizando, respectivamente, o *bootstrap* do *neighbor-joining* e da máxima verossimilhança, além da probabilidade posterior (PP). Valores em vermelho representam suportes insatisfatórios (*bootstrap* < 0,70 e PP < 0,95), enquanto o “x” indica que o agrupamento não foi recuperado no método de reconstrução filogenética correspondente. Valores de PP iguais ou superiores a 0,95 também foram incluídos nos demais ramos da filogenia. Para maiores esclarecimentos acerca dos acrônimos utilizados nas pontas da árvore, consulte o Apêndice B.

*psbA-trnH*



ITS

