



MESTRADO PROFISSIONAL  
USO SUSTENTÁVEL DE RECURSOS NATURAIS EM REGIÕES  
TROPICAIS

VITÓRIA CATARINA CARDOSO MARTINS

**MONITORAMENTO DA BIODIVERSIDADE DA FLORA DE  
CANGA, SERRA DOS CARAJÁS, PARÁ, ATRAVÉS DE DNA  
METABARCODING**

Belém / PA

2021

**VITÓRIA CATARINA CARDOSO MARTINS**

**MONITORAMENTO DA BIODIVERSIDADE DA FLORA DE  
CANGA, SERRA DOS CARAJÁS, PARÁ, ATRAVÉS DE DNA  
METABARCODING**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciências Ambientais, do Programa de Mestrado Profissional em Uso Sustentável de Recursos Naturais em Regiões Tropicais, do Instituto Tecnológico Vale Desenvolvimento Sustentável (ITV DS).

Orientador: Santelmo Selmo de Vasconcelos Jr., Dr  
Coorientador: Guilherme Côrrea de Oliveira, Dr.

**Belém / PA**

**2021**

### Dados Internacionais de Catalogação na Publicação (CIP)

M386 Martins, Vitória Catarina Cardoso  
Monitoramento da biodiversidade da flora de Canga, Serra dos Carajás  
- Pará, através de DNA *Metabarcoding* / Vitória Catarina Cardoso Martins. -  
Belém: ITV, 2021.  
96 f.: il.

Dissertação (Mestrado Profissional em Uso Sustentável de Recursos  
Naturais em Regiões Tropicais) - Instituto Tecnológico Vale, 2021  
Orientação: Dr. Santelmo Selmo de Vasconcelos Júnior.  
Coorientação: Dr. Guilherme Corrêa de Oliveira.

1. Genômica - Plantas. 2. Genômica - DNA *Metabarcoding*. 3. Solos -  
Amazônia. 4. Carajás, Serra dos (PA. I. Vasconcelos Júnior, Santelmo  
Selmo de. II. Oliveira, Guilherme Côrrea de. III. Título.

CDD 23. ed. 622.752098115

Bibliotecária responsável: Nisa Gonçalves / CRB 2 – 525

VITÓRIA CATARINA CARDOSO MARTINS

**MONITORAMENTO DA BIODIVERSIDADE DA FLORA DE  
CANGA, SERRA DOS CARAJÁS, PARÁ, ATRAVÉS DE DNA  
METABARCODING**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciências Ambientais, do Programa de Mestrado Profissional em Uso Sustentável de Recursos Naturais em Regiões Tropicais, do Instituto Tecnológico Vale Desenvolvimento Sustentável (ITV DS).

Data de aprovação:

Banca examinadora:

---

Santelmo Selmo de Vasconcelos Júnior  
Orientador – Dr. – Instituto Tecnológico Vale (ITV)

---

Markus Gastauer  
Membro interno – Dr. – Instituto Tecnológico Vale (ITV)

---

Adriano Reis Lucheta  
Membro externo – Dr. - Instituto SENAI de Inovação em Tecnologias Minerais

## DEDICATÓRIA

Dedico este trabalho a minha mãe, Ângela Suely Cardoso Martins, e ao meu avô, José Maria Martins, meus instrutores de vida, que contribuíram para a busca dos meus sonhos, fortaleceram a minha fé e foram minhas motivações para seguir e trilhar, diante de todas as adversidades, essa longa jornada.

## **AGRADECIMENTOS**

Agradeço a Deus, que através da minha fé sempre me auxiliou quando precisei, especialmente nos períodos mais difíceis da minha vida, inclusive no de pandemia, concedendo-me saúde, perseverança e forças para chegar ao final e me fazendo ir além das adversidades encontradas no caminho.

Ao Instituto Tecnológico Vale, no qual foi possível a realização de um dos sonhos, o Mestrado, com todas as condições que me proporcionaram dias de aprendizagem muito ricos.

A todos os docentes do Programa de Pós-Graduação em Uso Sustentável de Recursos Naturais em Regiões Tropicais do ITV, foram eles que me deram recursos e ferramentas para evoluir um pouco mais todos os dias.

Sou grata e honrada por todos os pesquisadores e bolsistas do Grupo de Genômica Ambiental, pelos ensinamentos que colhi durante os dois anos de Mestrado e pela certeza da contribuição árdua desses profissionais para mudanças significativas e inspirações para a minha vivência e crescimento profissional. Também agradeço ao técnico Manoel Lopes por todo auxílio no laboratório geral, seu apoio foi fundamental.

Aos membros da banca examinadora, que tão gentilmente aceitaram participar e colaborar com esta dissertação, contribuindo para a conclusão de mais essa etapa da minha vida.

Ao meu orientador, Dr. Santelmo Selmo de Vasconcelos, por toda ajuda neste último período, que foi crucial, pela confiança, orientação, competência, profissionalismo e ensinamentos valorosos no mundo das plantas, que me ajudaram a construir e consolidar muitas ideias que estão ilustrados neste trabalho e que foram imprescindíveis no sucesso dessa pesquisa. Obrigada por tudo, de coração!

Ao meu coorientador, Dr. Guilherme Côrrea de Oliveira, pela confiança, ensinamentos, competência, profissionalismo e apoio na realização desse projeto ao longo desses dois anos. Obrigada por acreditar no meu potencial e pelo voto de confiança de que eu seria capaz de concluir esse projeto, mostrando-me como é importante a dedicação e trabalho árduo para a concretização dos meus objetivos. Sinceramente, obrigada por tudo!

À minha mãe e aos meus avós, especialmente ao meu avô, por serem o meu pilar, obrigada por sempre acreditarem no meu potencial, por terem, apesar das dificuldades, me proporcionado orientações e apoio de toda a natureza, investindo na minha formação moral, intelectual, profissional desde o nascimento até a etapa que ora concluo. Sem eles não seria possível estar aqui hoje, vocês têm a minha eterna gratidão, jamais os esquecerei.

A Ariane Castro e José Augusto Bitencourt por me auxiliarem integralmente na coleta das amostras e no campo realizado na Floresta Nacional de Carajás. Obrigada pela amizade, apoio, companheirismo, ensinamentos e risadas. Tenho certeza que sem a ajuda de vocês esta pesquisa não seria possível. Com certeza, esses momentos se tornarão inesquecíveis!

À Amanda Manuelly, Jamily Lorena, Keyvilla Aguiar, Michelle Molina, Anderson Feitosa e Paulo Henrique, minha eterna gratidão a vocês, pela parceria, cumplicidade, apoio, ensinamentos, risadas, amizade, companheirismo, carinho, força e fé, para tudo que tivemos que enfrentar durante esses anos, almejando sempre muitas vitórias. Vocês me ajudaram a relaxar nos momentos difíceis e me colocaram para cima quando mais precisei. Vocês ficarão eternamente no meu coração. Alan Erik, obrigada por tirar minhas dúvidas de estatística no R. Você vai longe com sua competência! Aos meus colegas de laboratório, pela colaboração na pesquisa científica. Aos meus colegas da turma do mestrado, pela amizade nesses dois anos.

E, por fim, agradeço todas as pessoas que, de alguma forma, direta ou indiretamente, acreditaram em mim e foram essenciais para que alcançasse este objetivo com o qual sempre sonhei. Obrigada pela atenção, compreensão e carinho. Essa vitória é de todos nós!

## RESUMO

Os esforços de especialistas para a construção de um banco genético com códigos de barra de DNA para a identificação da Flora de Carajás geraram uma biblioteca de referência das espécies que ali ocorrem. Os dados estão disponíveis no banco de dados criado pelo ITV, o ITVBioBase e em repositórios públicos. Agora, outras tecnologias como a técnica de DNA *metabarcoding*, podem ser aplicadas. Neste sentido, a partir de traços de DNA no solo, é possível a identificação simultânea de múltiplas espécies de plantas com base no sequenciamento em massa de código de barras de DNA de uma amostra ambiental. Contudo, por se tratar de uma abordagem ainda nova, o DNA *metabarcoding* apresenta limitações metodológicas tanto no campo quanto no laboratório e exige esforços para estabelecer protocolos padronizados. Este estudo objetivou avaliar o método de DNA *metabarcoding* como uma ferramenta de monitoramento da flora da canga da Serra dos Carajás, no sudoeste do Pará, sendo estudadas três parcelas na região. Após a extração das amostras de DNA do solo, a região ITS2 foi amplificada e sequenciada na plataforma Illumina MiSeq. As análises de dados foram feitas usando o *pipeline* PIMBA, que permite a personalização do banco de dados de referência para etapa de classificação taxonômica. Adicionalmente, análises estatísticas foram realizadas para estimar as abundâncias relativas, diversidades  $\alpha$  e  $\beta$ , e a existência de dependência espacial. As sequências geradas foram agrupadas em unidades taxonômicas operacionais, sendo utilizado o banco de referências genéticas de plantas de Carajás para a determinação taxonômica. A diversidade de plantas detectada por eDNA incluiu 37 espécies de 33 gêneros e 19 famílias, identificando um número maior de espécies e gêneros que o método morfológico tradicional. Ainda, a análise baseada em DNA *metabarcoding* evidenciou 13 espécies comuns às listadas no inventário florístico das mesmas áreas de estudo, além de outras sete espécies que não haviam sido registradas pelo método tradicional. O fato do DNA *metabarcoding* ter permitido observar uma concordância com os dados do levantamento florístico aponta para uma robustez da abordagem molecular para ser utilizada em estudos de monitoramento da diversidade vegetal. Adicionalmente, com o desenvolvimento dos estudos de validação da metodologia tanto no campo quanto no laboratório foi criado um manual com o objetivo de facilitar a implementação da técnica de DNA *metabarcoding* em futuros levantamentos de espécies de plantas da canga.

**Palavras-chave:** Amazônia. Carajás. DNA *metabarcoding*. Plantas. Solo.

## ABSTRACT

Efforts to build a referenced DNA barcode genetic database to identify the flora of Carajás made it possible to obtain “fingerprints” of the species that occur there. From the reference database created by ITV, ITVBiobase, other approaches such as DNA *metabarcoding* can be applied to monitor species and environments. The DNA *metabarcoding* approach can simultaneously identify multiple plant species present in environmental samples through DNA traces in the soil matrix based on DNA barcodes mass sequencing from an environmental sample. However, as a new approach, DNA *metabarcoding* has methodological limitations both in the field and in the laboratory and requires efforts to establish standardized protocols based on environmental samples. The goal of this study was to evaluate the DNA *metabarcoding* method as a tool for monitoring the flora of Canga da Serra dos Carajás, in southwestern Pará, in three different plots. After extracting the DNA samples from the soil, the ITS2 region was amplified and sequenced on the Illumina Miseq platform. Data analysis was performed using the PIMBA pipeline, which allows for the customization of the reference database for the taxonomic classification stage. Statistical analyses estimated relative abundances,  $\alpha$  and  $\beta$  diversity, and the existence of spatial dependence. Sequences were grouped into operational taxonomic units (OTUs) using the Carajás plant genetic reference bank for taxonomic determination. The detected plant diversity by using eDNA included 37 species from 33 genera and 19 families, identifying a greater number of species and genera than the traditional morphological method. Furthermore, the analysis based on DNA *metabarcoding* showed 13 species common to those listed in the floristic inventory of the same study areas and seven other species that had not been registered by the traditional morphological method. The DNA *metabarcoding* results obtained agreed with the traditional morphological method. These results point out to the viability of using the molecular approach to determine and monitor plant diversity. As a result of validation studies of the methodology both in the field and in the laboratory, a manual was created with the aim of facilitating the implementation of the DNA *metabarcoding* approach in future routine environmental studies.

**Keywords:** Amazon. Carajás. DNA *metabarcoding*. Plants. Soil.

## LISTA DE FIGURAS

<b>Figura 1 -</b>	Localização da região da Serra dos Carajás (laranja), limites da FLONA Carajás (verde) e identificação dos principais platôs de canga (vermelho)	18
<b>Figura 2 -</b>	Diversos ambientes (fitofisionomias) associados às cangas da Serra dos Carajás: canga arbustiva (A), campo de Vellozia (B), campo rupestre arbustivo (C) e capão florestal (D).	29
<b>Figura 3 -</b>	Espécies de plantas vasculares consideradas endêmicas da Serra dos Carajás.	31
<b>Figura 4 -</b>	Região ITS	33
<b>Figura 5 -</b>	Códigos de barra de DNA, uma ferramenta molecular para identificação automatizada da Flora das Cangas de Carajás	35
<b>Figura 6 -</b>	Número de estudos usando as palavras 'DNA ambiental' ou 'eDNA' para o período entre 1 de janeiro de 2008 e 31 de dezembro de 2019 recuperado de uma pesquisa bibliográfica.	37
<b>Figura 7 -</b>	Número de estudos a partir de grupo de organismo e ecossistema diferentes usando as palavras 'DNA ambiental' ou 'eDNA' para o período entre 1 de janeiro de 2008 e 31 de dezembro de 2019, recuperado de uma pesquisa bibliográfica.	37
<b>Figura 8 -</b>	Fluxo de trabalho DNA metabarcoding	39
<b>Figura 9 -</b>	Custo do sequenciamento do genoma.	40
<b>Figura 10 -</b>	(A) Exemplo de um ponto de amostragem em campo. (B) Saco estéril (Whirl-pak) e espátula usados na coleta.	42
<b>Figura 11 -</b>	Localização das parcelas inventariadas pela empresa de consultoria, na Canga; Projeto Vale S11D, Inventários florísticos, Serra Sul, Canaã dos Carajás, Pará.	42
<b>Figura 12 -</b>	Plots selecionados dentro da área do Projeto Vale S11D, Serra Sul, Canaã dos Carajás, Pará.	43
<b>Figura 13 -</b>	Desenho experimental da amostragem em plots de canga.	43
<b>Figura 14 -</b>	Fluxograma simplificado mostrando as etapas de extração do DNA total das amostras de solo.	44
<b>Figura 15 -</b>	Fluxograma simplificado mostrando as etapas de construção de bibliotecas para o sequenciamento em MiSeq	47
<b>Figura 16 -</b>	Fluxograma simplificado mostrando as etapas de análise do pipeline PIMBA: A Pipeline for MetaBarcoding Analysis.	49
<b>Figura 17 -</b>	Curva de rarefação do sequenciamento por amostra.	52
<b>Figura 18 -</b>	Número de espécies, gêneros e famílias detectadas pelo método tradicional (verde e azul) e pelo método de DNA ambiental (amarelo). Amplo Total refere-se a espécies, gêneros e famílias identificados em todos os plots da S11D indicados na Figura 8, e Amplo Plots são espécies, gêneros e famílias identificados	55

somente nos plots analisados por eDNA, conforme indicado na Tabela 2

- Figura 19 -** Diagrama de Venn indicando a sobreposição das famílias, gêneros e espécies observadas nos três diferentes plots 56
- Figura 20 -** Diagrama de Venn indicando a sobreposição das espécies observadas nos diferentes plots pela abordagem tradicional (A) e pela abordagem de DNA metabarcoding (B). 56
- Figura 21 -** Diagramas de Venn indicando o número de espécies em comum encontradas pela identificação morfológica tradicional e pela abordagem de DNA metabarcoding por parcela nos plots PCG-121 (A), PCG-168 (B) e PCG-235 (C). 57
- Figura 22 -** Frequência por família amostrada em cada um dos três plots estudados (PCG-121, PCG-168 e PCG-235), indicando as frequências por amostra (A) e por parcela (B). 60
- Figura 23 -** Frequência por gênero amostrada em cada um dos três plots estudados (PCG-121, PCG-168 e PCG-235), indicando as frequências por amostra (A) e por parcela (B). 61
- Figura 24 -** Frequência por espécie amostrada em cada um dos três plots estudados. As frequências por amostra e por parcela estão representadas em A e B, respectivamente. 65
- Figura 25 -** Análise de diversidade Alfa baseada nos índices de riqueza (Chao1) e diversidade de espécies (Shannon), considerando presença/ausência nas diferentes parcelas localizadas na S11D. Letras diferentes no mesmo boxplot indicam uma diferença significativa ( $p < 0,05$ ), de acordo com o teste post-hoc de Tukey HSD. 69
- Figura 26 -** Análise de diversidade Alfa baseada nos índices de riqueza (A) (Chao1) e diversidade de espécies (B) (Shannon), considerando o método tradicional e eDNA em diferentes parcelas localizadas na S11D. 69
- Figura 27 -** Análise de diversidade Alfa baseada nos índices de riqueza (A) (Chao1) e diversidade de espécies (B) (Shannon), considerando diferentes amostras localizadas na S11D. 70
- Figura 28 -** Análise das Coordenadas Principais (PCoA) da diversidade beta dos locais de amostragem (distância de Jaccard). As elipses representam 95% dos indivíduos dos grupos indicados na análise. 71
- Figura 29 -** Distâncias geográficas e sua relação com a composição (Índice de Sorensen) das comunidades de plantas dentro dos três plots de canga avaliados. 72
- Figura 30 -** Distâncias geográficas e sua relação com a composição (Índice de Sorensen) das comunidades de plantas entre os três plots de canga avaliados. A distância geográfica foi logaritmizada. 73

<b>Figura 31 -</b>	Eletoforese em gel de agarose 1% dos DNA's provenientes de amostras de solo da parcela PCG-235, PCG-168 e PCG-121 extraídos a partir do Kit Power Soil.	88
<b>Figura 32 -</b>	Corrida no Tapestation System 4200 mostrando o perfil dos tamanhos dos amplicons	89
<b>Figura 33 -</b>	Frequência de identificação das famílias de plantas detectadas via DNA metabarcoding por amostra.	91
<b>Figura 34 -</b>	Frequência de identificação das famílias de plantas detectadas via DNA metabarcoding e método tradicional em distintas parcelas	91
<b>Figura 35 -</b>	Frequência de identificação das espécies de plantas detectadas via DNA metabarcoding e método tradicional em distintas parcelas localizadas na Serra Sul em Canaã dos Carajás.	92

## LISTA DE TABELAS

<b>Tabela 1 -</b>	Coordenadas geográficas, Leste e Norte, das três parcelas estudadas na região do projeto S11D, em Canaã dos Carajás-PA. Analisados – blocos estudados e relatados neste documento.	42
<b>Tabela 2 -</b>	Lista de espécies encontradas pela consultoria nas áreas selecionadas em S11D.	81
<b>Tabela 3 -</b>	Metadados da parcela 168	82
<b>Tabela 4 -</b>	Metadados da parcela 121	83
<b>Tabela 5 -</b>	Metadados da parcela 235.	84
<b>Tabela 6 -</b>	Informações das concentrações de DNA da parcela 235	85
<b>Tabela 7 -</b>	Informações das concentrações de DNA da parcela 121	86
<b>Tabela 8 -</b>	Informações das concentrações de DNA da parcela 168	87
<b>Tabela 9 -</b>	Número de sequências e bases produzidas e que passaram pelo controle de qualidade do sequenciamento.	90
<b>Tabela 10 -</b>	Número de OTU e frequência de ocorrência das espécies por parcela.	93
<b>Tabela 11 -</b>	Valores de diversidade Alfa baseada nos índices de riqueza (Chao1) e diversidade de espécies (Shannon), considerando presença/ausência nas diferentes parcelas localizadas na S11D.	94
<b>Tabela 12 -</b>	Valores de diversidade Alfa baseada nos índices de riqueza (A) (Chao1) e diversidade de espécies (B) (Shannon) considerando presença/ausência no método tradicional e eDNA em diferentes parcelas localizadas na S11D.	95

## LISTA DE SIGLAS E ABREVIATURAS

- BOLD – *Barcoding of Life Database* - Banco de dados de códigos de barras da vida
- CBOL – *Consortium for Barcoding of Life* - Consórcio para o Código de Barras da Vida
- CTAB - *Cetyl trimethylammonium bromide* - brometo de cetil trimetil amônio
- DMSO – Di metil sulfoxido
- DNTP - Desoxirribonucleotídeos Fosfatados
- eDNA – *environmental DNA* - DNA ambiental
- FCC – Flora das Cangas de Carajás
- FLONA – Floresta Nacional
- GSPC - *Global Strategy for Plant Conservation* - Estratégia Global para a Conservação de Plantas
- ITV-DS – Instituto Tecnológico Vale Desenvolvimento Sustentável
- ITS – *Internal Transcribed spacer* - Espaçador Interno Transcrito
- OTU – *Operational Taxonomic Unit* - Unidade Taxonômica Operacional
- ONU – Organização das Nações Unidas
- NCBI - *National Center for Biotechnology Information* - Centro Nacional de Informações sobre Biotecnologia
- NGS – *Next Generation Sequencing* - Sequenciamento de próxima geração
- PIMBA- *Pipeline for MetaBarcoding Analysis*
- PCR – *Polymerase Chain Reaction* - Reação em Cadeia da Polimerase

## SUMÁRIO

<b>ORGANIZAÇÃO DA DISSERTAÇÃO</b> .....	<b>12</b>
<b>1 INTRODUÇÃO</b> .....	<b>12</b>
<b>2 OBJETIVOS</b> .....	<b>16</b>
2.1 OBJETIVO GERAL.....	16
2.2 OBJETIVOS ESPECÍFICOS .....	16
<b>3 REFERENCIAL TEORICO OU FUNDAMENTAÇÃO CIENTÍFICA</b> .....	<b>17</b>
3.1 SERRA DOS CARAJÁS .....	17
3.2 CANGA DE CARAJÁS .....	18
3.3 PROJETO FLORA DAS CANGAS DA SERRA DOS CARAJÁS.....	20
3.3.1 Banco de dados de código de barras de DNA .....	23
3.4 DNA METABARCODING .....	26
<b>4 MATERIAIS E MÉTODOS</b> .....	<b>32</b>
4.1 COLETA DAS AMOSTRAS.....	32
4.2 EXTRAÇÃO DE DNA .....	35
4.3 SELEÇÃO DE PRIMERS .....	36
4.4 PROCEDIMENTO DE AMPLIFICAÇÃO DO DNA.....	37
4.5 PREPARAÇÃO DA BIBLIOTECA E SEQUENCIAMENTO NGS .....	37
4.6 ANÁLISES BIOINFORMÁTICAS.....	39
4.7 ANÁLISES ESTATÍSTICAS .....	41
<b>5 RESULTADOS E DISCUSSÃO</b> .....	<b>43</b>
5.1 LEVANTAMENTO DA FLORA POR MEIO DE TAXONOMIA CLÁSSICA .....	43
5.2 LEVANTAMENTO DA FLORA VIA DNA METABARCODING .....	43
5.2.1 Amostras extraídas.....	43
5.2.2 Sequenciamento NGS e pré-processamento .....	43
5.2.3 Atribuição taxonômica .....	44
5.2.4 Índices de diversidade alfa e beta .....	61
5.3 RECOMENDAÇÕES PARA FUTURAS PESQUISAS.....	67
<b>6 CONCLUSÃO</b> .....	<b>68</b>
<b>REFERÊNCIAS</b> .....	<b>69</b>
<b>APÊNDICES</b> .....	<b>80</b>

## **ORGANIZAÇÃO DA DISSERTAÇÃO**

O presente trabalho encontra-se organizado sob a forma de relatório técnico, como disposto no Regulamento Interno do Programa de Pós-Graduação em Uso Sustentável de Recursos Naturais em Regiões Tropicais do ITV DS.

### **RELATÓRIO TÉCNICO**

#### **Monitoramento da Biodiversidade da Flora de Canga, Serra dos Carajás, Pará, através de DNA metabarcoding**

##### **Sob autoria de**

Vitória Catarina Cardoso Martins, Gisele Lopes Nunes, Renato Renison Moreira Oliveira, Raíssa Lorena Silva da Silva, Ronnie Alves, Santelmo Vasconcelos e Guilherme Oliveira,

PROD.TEC.ITV.DS. / N003/2021  
DOI 10.29223/PROD.TEC.ITV.DS.2021.03.Martins

**Belém / PA**

**Fevereiro / 2021**

# 1 INTRODUÇÃO

A região amazônica contém uma enorme biodiversidade, a qual é distribuída em uma variedade de ecossistemas com características únicas (ZAPPI et al., 2015), destacando-se a Serra dos Carajás, que é caracterizada pela presença isolada de altos platôs de afloramento de rochas ferruginosas e relevo acidentado (MOTA et al., 2018). A região de Carajás abriga um tipo de vegetação sobre as rochas ferruginosas denominada 'canga' (SECCO e MESQUITA, 1983; VIANA et al., 2016; SOUZA-FILHO et al., 2019). A canga apresenta uma vegetação aberta, predominantemente herbáceo-arbustiva, com a presença de várias espécies endêmicas e adaptadas ao ambiente peculiar, principalmente às condições edáficas restritivas, tanto fisicamente quanto quimicamente (SKIRYCZ et al., 2014).

Recentemente, o inventariamento das espécies que compõem a flora da Serra dos Carajás foi objeto do projeto Flora das Cangas de Carajás, ou FCC, que iniciou no ano de 2015 e teve como um dos seus objetivos disponibilizar informações taxonômicas sobre a totalidade da cobertura vegetal da região. Participaram do projeto 145 especialistas de mais de 30 instituições, culminando em diversas publicações incluídas em quatro volumes especiais do periódico *Rodriguésia* – a revista científica do Jardim Botânico do Rio de Janeiro (VIANA et al., 2016, ZAPPI, 2017).

Nas expedições realizadas durante o projeto FCC, as plantas foram coletadas, e os espécimes foram enviados para herbários públicos, principalmente para a coleção botânica do Museu Paraense Emílio Goeldi (MPEG), em Belém, Pará. As análises taxonômicas tradicionais foram realizadas para a validação de cada um dos espécimes depositados, assim como foram coletados tecidos para a caracterização genética das plantas (MOTA et al., 2018).

Segundo Zappi (2017), foi o extenso trabalho de campo dos taxonomistas no início que propiciou a formação de um banco de dados genéticos com códigos de barra de DNA referenciado para identificação da flora. O código de barras de DNA (*DNA barcode*) é uma tecnologia que possibilita a identificação de espécies a partir da sequência de um pequeno fragmento do genoma organelar ou nuclear (HEBERT et al., 2003; SHNEER, 2009) e mesmo quase 20 anos após sua formalização, continua em plena expansão com mais de 6.000 artigos publicados até 2018 (DORMOTT et al., 2018).

A construção de uma biblioteca de DNA referenciada com as informações genéticas disponíveis sobre a flora de Carajás (OLIVEIRA et al., 2019; VASCONCELOS et al., 2021) vem possibilitando de maneira inovadora a identificação automatizada das espécies de plantas da região. Analogamente, é como obter “impressões digitais” das espécies que ali ocorrem (ZAPPI, 2017). A partir da biblioteca de referência, outras abordagens podem ser utilizadas, por exemplo, a partir de traços de DNA no solo é possível analisar as plantas em conjunto, obtendo um sequenciamento em massa e sem a necessidade de amostragens diretas de espécimes (FLØJGAARD et al., 2019).

Dessa forma, foi imprescindível o trabalho prévio dos taxonomistas no caminho da descrição de espécies, pois foi o que viabilizou a descrição da base de referência, somando-se aos esforços para a geração dos códigos de barra de DNA. De agora em diante, o taxonomista pode se ocupar de áreas mais problemáticas de identificação e descrição da diversidade vegetal ao invés de executar atividades de repetição como o acompanhamento da biodiversidade de áreas já bem caracterizadas (RIEDEL et al., 2013).

Em Carajás, por exemplo, os métodos tradicionais para identificar plantas geralmente têm algumas limitações. São usualmente necessárias várias visitas para coletar a maioria dos espécimes na fase fértil (em flor, fruto ou esporos) para que possam ser identificados até o nível de espécie (ZAPPI, 2017). Existem também plantas com ciclos de vida que duram apenas alguns meses, e depois desaparecem completamente durante a época seca, complicações na identificação de espécies taxonomicamente próximas e morfologicamente parecidas ou espécies crípticas (ZAPPI, 2017). Para resolver estas questões são necessários especialistas para cada grupo taxonômico, equipamentos especializados, empréstimo de matérias de coleções ou tempos de observação específicos, o que aumenta o tempo e esforço (DEINER et al., 2017).

Estes fatos influenciam diretamente no planejamento das ações de conservação em regiões que serão afetadas por atividades industriais, como a mineração, cuja eficácia depende da capacidade de mapear a biodiversidade e fornecer indicadores da “saúde” do ecossistema dentro de um curto espaço de tempo (JARDIM e MARTINS, 2018). Levantamentos da biodiversidade, incluindo a florística, são uma demanda legal para empreendimentos industriais, incluindo os de mineração

(Lei 6.938/81 e Resolução 237/97 do CONAMA). Os levantamentos visam gerar o conhecimento necessário para que o empreendimento impacte minimamente o meio ambiente. O desconhecimento sobre o ambiente e a falta de cobertura dos levantamentos de biodiversidade dificultam os *stakeholders* a tomarem decisões substanciadas e em tempo hábil (CORADIN; TAFURI; DUARTE, 2012). O resultado são incertezas no desenvolvimento dos projetos industriais e sobre a capacidade das decisões direcionadas à conservação do meio ambiente.

O DNA *metabarcoding* vem sendo considerado um método rápido e eficiente para o levantamento e monitoramento da biodiversidade, pois permite a identificação simultânea de múltiplas espécies com base no sequenciamento em massa de código de barras de DNA de uma amostra ambiental (DEINER et al., 2017; RUPPERT et al., 2019). Em diversas regiões do planeta, o DNA *metabarcoding* já é utilizado com o objetivo de esclarecer questões ambientais e ecológicas (FLØJGAARD et al., 2019; LITTLEFAIR et al., 2018; MATESANZ et al., 2019). Para a identificação taxonômica, os códigos de barras obtidos são comparados com base em um banco de dados de referência de sequência de DNA gerado anteriormente, a partir de espécies determinadas por um especialista (HEBERT et al., 2003; TABERLET et al., 2012a). Assim, o uso dessa ferramenta molecular possibilita de forma não invasiva a detecção de rastros de DNA das plantas em amostras ambientais (água, solo), contribuindo com rapidez e objetividade no levantamento de espécies, mesmo nos ecossistemas mais remotos ou novos, dinamizando as pesquisas e inventários de biodiversidade em termos de custos e cronogramas (COWART et al., 2015; FAHNER et al., 2016).

O solo é uma fonte potencial para a obtenção de amostras ambientais que possibilita a eficiente identificação de táxons que deixam traços de DNA no substrato, como, por exemplo, as comunidades vegetais. A abordagem tem sido usada com sucesso para rastrear traços de DNA e recriar a composição da biota atual ou passada de uma região (ALSOS et al., 2018; EDWARDS et al., 2018; FAHNER et al., 2016; YOCCOZ et al., 2012). Além disso, sabe-se que no cenário onde a taxa global de extinção de espécies é maior do que a taxa de descrição, torna-se cada vez mais difícil o monitoramento da biodiversidade apenas com o uso de métodos tradicionais (THOMSEN e WILLERSLEV, 2015).

Logo, inserido no contexto dos Objetivos de Sustentabilidade da Vale e o 15º Objetivo de Sustentabilidade da ONU, a Vida na Terra, o DNA *metabarcoding* oferece

uma poderosa ferramenta molecular que permite a captura de sinais da biodiversidade local sem a necessidade de amostragem direta de espécimes. Isso possibilita eficiência nos esforços de coleta, com menor exposição de pessoas ao trabalho de campo, digitalização da informação possibilitando facilmente a reanálise dos resultados, e com alto potencial de automatização dos procedimentos sem causar distúrbios nos ecossistemas (RUPPERT et al., 2019).

No Brasil, a falta de conhecimento sobre determinadas plantas ou animais dificulta o alcance dos objetivos da ONU, e segundo Giulietti et al. (2019), é preciso ter domínio sobre o ambiente para melhor planejar as ações e ajudar na conservação, especialmente tendo em vista os cenários de mudanças climáticas previstos para a região (GIULIETTI et al., 2016; LEVINE et al., 2016; MIRANDA et al., 2019). Ter conhecimentos sobre a composição florística de ambientes naturais é de suma importância, tendo os empreendedores grande responsabilidade sobre a proteção da biodiversidade nas áreas em que atuam (GIULIETTI et al., 2016).

Portanto, são essenciais os estudos sobre o emprego de novas tecnologias com foco em ambientes diversos e pouco explorados, já que adotar estratégias como esta, representam o maior passo dado na direção do fechamento de lacunas no conhecimento taxonômico (RIEDEL et al., 2013). Adicionalmente, é importante termos disponíveis ferramentas que gerem respostas rápidas. As partes interessadas necessitam de respostas ágeis e que permitam análises robustas e integradas ao mundo digital, em sintonia com a Indústria 4.0 (CORADIN; TAFURI; DUARTE, 2012).

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

- Avaliar o método de DNA *metabarcoding* como uma ferramenta de monitoramento da flora da canga da Serra dos Carajás, Pará.

### 2.2 OBJETIVOS ESPECÍFICOS

- Determinar as espécies de plantas da canga através de DNA *metabarcoding*;
- Determinar parâmetros ótimos para o esforço de amostragem e esforço de sequenciamento para o biomonitoramento de plantas com o uso de *metabarcoding*;
- Comparar o método de DNA *metabarcoding* com o método padrão realizado pelos consultores na mesma área de estudo;
- Desenvolver um manual para DNA *metabarcoding* de plantas (documento apresentado separadamente).

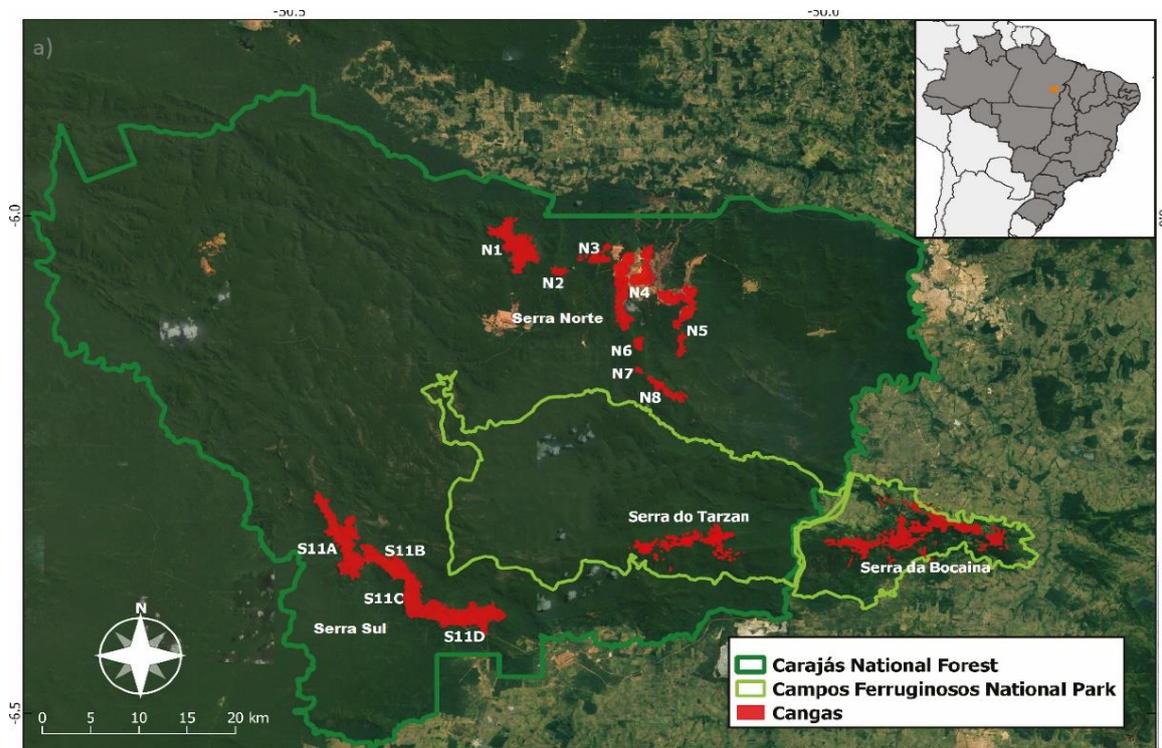
### **3 REFERENCIAL TEORICO OU FUNDAMENTAÇÃO CIENTÍFICA**

#### **3.1 SERRA DOS CARAJÁS**

O Brasil possui uma das maiores jazidas de ferro de alto teor do planeta, localizado no sul do Pará, a Serra dos Carajás, a qual destaca-se por possuir uma vasta área de 144 km<sup>2</sup>, áreas de proteção ambiental, áreas de povos indígenas e reservas naturais, como a Floresta Nacional (FLONA) de Carajás (SOUZA-FILHO et al., 2019). Esta, criada em fevereiro de 1998, resguarda parte da floresta amazônica e inclui as áreas abertas de cangas que possuem uma diversidade florística singular (VIANA et al., 2016).

De acordo com Corrêa (2011), a região de Carajás se encontrava isolada e inserida no cenário mais desconhecido do território brasileiro até a década de 1970. Porém, com o início do programa de prospecção geológica de Carajás, por volta dos anos de 1966 e 1967, foram descobertas suas grandes e valiosas jazidas de minério, iniciando-se a investigação de viabilidade econômica e atraindo várias empresas de mineração (SILVEIRA, 2016). Através dos estudos de Rezende e Barbosa (1972), a Serra dos Carajás foi setorizada em quatro conjuntos principais de serras (Figura 1), tais como Serra Norte, Serra Sul, Serra Leste e Serra de São Félix. Nas áreas de canga das Serras do Tarzan e da Bocaina, foi decretada em junho de 2017 a criação do Parque Nacional dos Campos Ferruginosos (PNCF), de maneira a propiciar estabilidade ambiental na região (MOTA et al., 2015).

**Figura 1.** Localização da região da Serra dos Carajás (laranja), limites da FLONA Carajás (verde) e identificação dos principais platôs de canga (vermelho).



Fonte: Maia et al. (2021).

### 3.2 CANGA DE CARAJÁS

As áreas de canga são ecossistemas singulares e se referem a um depósito de hematita superficial que ocorrem majoritariamente em Minas Gerais (MG), Carajás (PA), Caetité (BA) e Morraria de Urucum (MS) (CARMO et al., 2012). A canga abriga uma flora especializada, com a presença de espécies endêmicas e adaptadas ao ambiente peculiar, como: *Paspalum carajasensis*, *Isoetes cangae* e *Ipomea cavalcantei*, por exemplo (GIULIETTI et al., 2019; ZAPPI, 2017). Essas áreas possuem estações úmidas e secas extremas, altas concentrações de elementos ou metais potencialmente tóxicos, solo superaquecido e impermeável, baixa retenção de água e composição geoquímica particular (SAHOO et al., 2017; SCHETTINI et al., 2018).

A canga abriga uma mistura de fitofisionomias associadas às rochas ferruginosas, como canga arbustiva, campo de *Vellozia*, campo rupestre arbustivo e capão florestal (Figura 2), que possuem variações na composição florística em relação a abundância e dominância de espécies, solos ricos em matéria orgânica e são barreiras naturais contra o fogo natural ou antrópico (JACOBI et al., 2007). Os campos

gramíneos e savanas são onde o substrato rochoso está mais fragmentado; as florestas com vegetação higrófila, são associadas a riachos e lagoas, perenes ou sazonais; e os capões florestais são áreas com maior acúmulo de matéria orgânica sobre o substrato ferruginosos, proporcionando condições para as matas se desenvolvem altas, bem estruturadas e ricas em variedades de espécies (SKIRYCYZ et al., 2014; VIANA et al., 2016).

Adicionalmente, a flora das cangas é marcada por espécies herbáceas anuais e efêmeras durante o período chuvoso e uma tendência com diferentes graus de decíduidade do componente arbustivo e arbóreo, algumas com grande riqueza de trepadeiras e cipós no período mais seco (MOTA et al., 2015; VIANA et al., 2016). O clima na região de Carajás é caracterizado pela sazonalidade, as estações seca e chuvosa são bem distintas. Geralmente no período de novembro a maio, a precipitação total anual varia de 1.800 a 2.300 mm no período chuvoso e 350 mm durante a estação seca, apresentando pequenas variações entre as serras (MORAES et al., 2005; SAHOO et al., 2017). Segundo Corrêa (2011) e Skiryucz et al. (2014), o relevo e as condições climáticas selecionaram espécies de plantas capazes de lidar com condições ambientais de grande amplitude térmica e de chuva.

**Figura 2** - Diversos ambientes (fitofisionomias) associados às cangas da Serra dos Carajás: canga arbustiva (A), campo de *Vellozia* (B), campo rupestre arbustivo (C) e capão florestal (D).



**Fonte:** Autora, 2021.

### 3.3 PROJETO FLORA DAS CANGAS DA SERRA DOS CARAJÁS

Desde a descoberta das Serras de Carajás, em 1967, diversos estudos vêm sendo feitos com a flora da região, especialmente para atender licenciamentos para a mineração. Até recentemente, a lista da Flora de Carajás que abrangia o maior número de espécies havia sido preparada em 1991, por Manoela Ferreira e Silva e colaboradores, incluindo 58 famílias, 145 gêneros e 232 espécies, sendo que 72 estão identificadas apenas até o nível de gênero (ZAPPI, 2017). Dessa forma, a elaboração do projeto da flora das cangas foi iniciada no período de janeiro de 2015, seguindo até dezembro de 2017. O projeto contou com um conjunto de 145 especialistas, que através de um convênio entre Museu Paraense Emilio Goeldi (MPEG) e Instituto Tecnológico Vale (ITV), colaboraram com tratamento de diversas famílias (MOTA et al., 2018; OLIVEIRA DA SILVA e ILKIU-BORGES, 2018; SALINO e ALMEIDA, 2018).

Foram registradas espécies nativas e/ou naturalizadas na área, briófitas, licófitas, samambaias, gimnospermas e angiospermas, perfazendo um total de aproximadamente 1.100 espécies para vegetação associada a canga, representando 15% das espécies conhecidas do estado do Pará (MOTA et al., 2018; VIANA et al., 2016; ZAPPI, 2017).

No total, foram compilados mais de 15.000 espécimes de plantas de Carajás, incluindo as áreas de cangas e florestas, registrando 164 monografias de famílias, incluindo 89 espécies briófitas, 173 de samambaias e licófitas e 831 de plantas com sementes (MOTA e GIULIETTI, 2016; MOTA et al., 2018; SALINO e ALMEIDA, 2018; VIANA et al., 2016). As famílias com maior número de espécies registradas para a canga foram: Poaceae (87), Fabaceae (65) e Rubiaceae (48). Ao longo desses anos, ao menos quatro novos gêneros foram descritos, entre eles *Brasilianthus*, *Monogereion*, *Parapiquieria* e *Carajasia* (Figura 3) além de mais de 24 espécies recentemente descritas (GIULIETTI et al., 2019; MOTA et al., 2018).

Os dados reunidos foram depositados em herbários, principalmente no Museu Paraense Emilio Goeldi (MG), na Universidade Federal de Minas Gerais (BHCB) e na Vale, em Parauapebas (HCJS), locais que concentram grande volume de material oriundo das cangas de Carajás e servem como base para a taxonomia das plantas da região (MOTA et al., 2018; VIANA et al., 2016).

O projeto Flora das Cangas de Carajás (FCC) formou uma base de dados florística única que compõem a área delimitada da Serra dos Carajás, o que auxilia na relação entre biodiversidade e os negócios do setor mineral. O FCC alavancou a pesquisa botânica das espécies, gerando quatro fascículos na revista *Rodriguésia*, do Jardim Botânico do Rio de Janeiro, entre 2016 e 2017, sendo possível consultar informações sobre taxonomia e morfologia das espécies, distribuição geográfica, ilustração de estruturas, bibliografia, diferenciação de espécies próximas, época de floração e frutificação e seus usos (MOTA et al., 2018; VIANA et al., 2016; ZAPPI, 2017). Isto representou uma contribuição imprescindível para o conhecimento da Floresta Amazônica, que apesar da importância, ainda é o bioma menos conhecido do país, como também para a ampliação da versão eletrônica da Flora do Brasil 2020, que está sendo construída por mais de 700 botânicos, visando atender à meta 1 da Estratégia Global de Conservação de Plantas (GSPC) (FILARDI et al., 2018; ZAPPI, 2017).

**Figura 3** - Espécies de plantas vasculares consideradas endêmicas da Serra dos Carajás. *a.* *Monogereion carajensis*; *b.* *Parapiqueria cavalcantei*; *c.* *Ipomoea cavalcantei*; *d.* *Sinningia minima*; *e.* *Brasilianthus carajasensis*; *f.* *Paspalum carajasense*; *g.* *Carajasia cangae*; *h.* *Perama carajensis*; *i.* *Blechnum areolatum*; *j.* *Anemopaegma carajasensis*; *k.* *Isoetes cangae*; *l.* *Isoetes serracarajensis*



Fonte: adaptado de Viana et al. (2016).

### 3.3.1 Banco de dados de código de barras de DNA

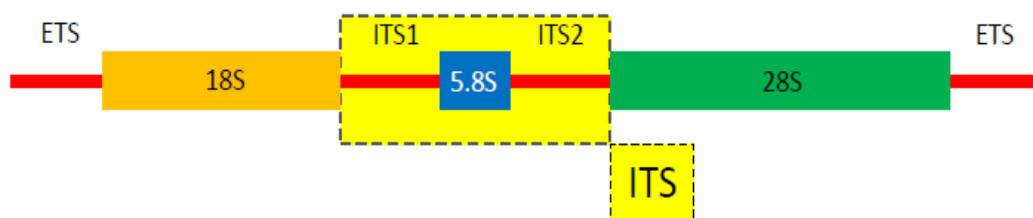
A base de conhecimento criada para as plantas foi estabelecida a partir das expedições feitas pelos taxonomistas durante o período do estudo. Nas formações de canga, foi coletado material fértil (com flor e fruto) de todas as espécies da área usando saquinhos com sílica gel ou tubos fechados contendo gel de CTAB, utilizados para a formação de um banco de DNA do ITV (VIANA et al., 2016). O banco de DNA é destinado para a elaboração do código de barras de DNA das plantas das cangas de Carajás, auxiliando nos processos de identificação de espécimes, dados de biodiversidade, entre outras aplicações (HEBERT et al., 2003; SHNEER, 2009; VIANA et al., 2016; ZAPPI, 2017) (Figura 4).

Assim, ao ter uma planta identificada por um botânico, por meio de métodos tradicionais e associando a ela um código de barras de DNA, é incorporado nesse código todo o conhecimento taxonômico, que agora pode ser trabalhado no mundo digital (Figura 4). Nesse contexto, os dados presentes na Flora representam a base para trabalhos que utilizem essas espécies. Logo, com o material obtido, é possível incrementar o acervo já disponível, comparar com a bibliografia especializada, coleções e base de dados, e armazenar tecidos vegetais para estudos anatômicos e moleculares, visando a identificação computacional das plantas (VIANA et al., 2016; ZAPPI, 2017).

A utilização de códigos de barra de DNA é complementar à descrição morfológica realizada por taxonomistas e pode ajudar na identificação molecular de grupos, como as plantas (KRESS et al., 2017). Ressalta-se que a região do DNA escolhida que funcionará como código de barra depende de cada organismo em particular, no entanto é fato que o desenvolvimento para plantas não é tão simples quanto para outros eucariotos, como animais e fungos (FAZEKAS et al., 2009; HEBERT et al., 2016; HOLLINGSWORTH et al., 2016; VASCONCELOS et al., 2021). Existe uma dificuldade em padronizar quais regiões do cpDNA funcionarão como código de barra de DNA confiáveis para plantas (VASCONCELOS et al., 2021). Diversos autores, relatam taxas de sucesso variáveis usando diferentes marcadores para plantas, como *psbK-psbI*, *rpoB*, *rpoC1*, *atpF-atpH*, *trnH-psbA*, *trnL* P6 loop (FAZEKAS et al., 2008; FAHNER et al., 2016; VASCONCELOS et al., 2021) e até a combinação das sequências *rbcL* e *matK* tem sido recomendada como principais loci de código de barras (CBOL et al., 2009; KRESS et al., 2017). Além dos marcadores

organelares, algumas regiões do genoma nuclear, como os espaçadores transcritos internos (ITS1 e ITS2) do gene 35S rRNA, também produzem códigos de barras de DNA adequados para plantas (CHEN et al., 2010; HOLLINGSWORTH et al., 2011; VASCONCELOS et al., 2021). Nesse contexto, para o levantamento de várias espécies de plantas de Carajás, o ITS2 tem sido uma das principais escolhas de marcadores moleculares (Figura 4), uma vez que existe várias vantagens metodológicas do uso deste código de barras de DNA, como a facilidade de padronizar as condições de PCR e um tamanho de *amplicon* menor (~ 450 bp) em comparação com outros códigos de barras de DNA de plantas frequentemente usados (CHEN et al., 2010; RICHARDSON et al., 2015; GOUS et al., 2019; VASCONCELOS et al., 2021).

**Figura 4 - Região ITS**  
35S rDNA



**Fonte:** Relatório de Atividades ITV, 2019.

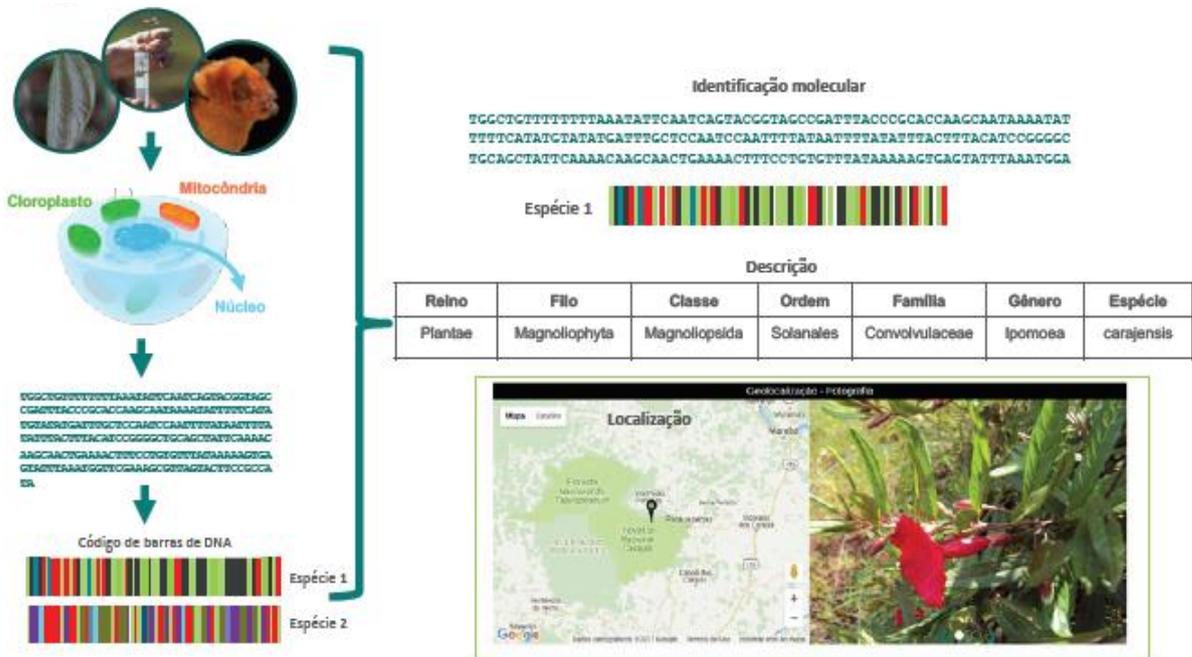
Os esforços de biólogos e taxonomistas para a construção de um banco genético com códigos de barra de DNA para a identificação da Flora de Carajás possibilitaram obter uma biblioteca de referência das espécies que ali ocorrem (OLIVEIRA et al., 2019). Os dados da fauna e flora são disponíveis no banco de dados criado pelo ITV, o chamado ITVBioBase, onde é possível acessar os códigos de barra de DNA, associados a espécimes caracterizados por um taxonomista, dados de coleta realizados pelos especialistas, localização das amostras e registros fotográficos, entre outras informações (Figura 4) (ITV, 2019). Com base na lista de plantas vasculares para a canga relatadas por Mota et al. (2018), Salino e Almeida (2018) e Giulietti et al. (2019), a quantidade de código de barras de DNA depositados no ITVBioBase cobrem aproximadamente 36,40% da diversidade de espécies da canga (380 das 1044 spp descritas) (VASCONCELOS et al., 2021). É importante notar que, apesar do fato de

que a maioria, das espécies de plantas da canga ainda não possuam códigos de barras de DNA válido, o banco de dados genético ITVBiobase caracteriza-se como o mais extenso e atualizado esforço para a produção de código de barras de DNA para a Amazônia brasileira (VASCONCELOS et al., 2021).

A ideia de padronizar o uso de dados genéticos para auxiliar taxonomia já vem mobilizando a comunidade científica há algum tempo, tendo sido lançada uma grande iniciativa para agregar os grupos que utilizam os códigos de barras de DNA para a descrição de espécies, o Consórcio Barcode of Life (CBOL 2010). O CBOL criou um sistema para armazenar, gerenciar e publicar os dados online, sendo que, houve um crescimento do ano de 2011, onde havia 1.005.505 sequências (com 70.532 spp. descritas) para o ano de 2020, com cerca de 8,449,641 sequências (com 314,669 spp. descritas) (CBOL, 2010; DEWALT, 2011). No entanto, o CBOL abrange apenas cerca de 13.1% das angiospermas (66,669) e 20,8% das gimnospermas (785), até o momento, sendo uma proporção baixa da biodiversidade global, principalmente no caso das plantas com flores (DORMONTT et al., 2018)

Obviamente, a disponibilidade de sequências de determinados marcadores, como *rbcl* e *ITS2* em repositórios públicos (BOLD) e locais (ITVBiobase), além de ser essencial para um inventário eficaz da diversidade de plantas, também contribui para o fortalecimento da base de dados online do Consórcio Barcode of Life (CBOL 2010), uma vez que, para angiospermas, *ITS2* tem sido uma das regiões de código de barras mais usadas até agora, cobrindo 26,7% do ca. 340.000 sequências disponíveis na base de dados BOLD (até 12 de fevereiro de 2020), atrás apenas de *rbcl* e *matK*, respectivamente com 35,8% 31,6% (VASCONCELOS et al., 2021). Agora, outras tecnologias como a abordagem de DNA *metabarcoding* podem ser aplicadas visando o monitoramento das espécies de forma rápida, automatizada e eficiente.

**Figura 5** - Códigos de barra de DNA, uma ferramenta molecular para identificação automatizada da Flora das Cangas de Carajás.



Fonte: Relatório de Atividades ITV, 2019.

### 3.4 DNA METABARCODING

Segundo Deiner et al. (2017) e Porter e Hajibabaei (2018), as principais abordagens para estudar a biodiversidade e realizar levantamento de riqueza e abundância de espécies de plantas têm sido realizadas através dos métodos convencionais de observação direta e coleta de espécimes. Porém, estas abordagens apresentam limitações, já que a descrição morfológica de uma determinada planta para a construção de inventários florísticos ainda é uma tarefa demorada e que requer várias visitas a uma área de coleta, aumentando a exposição. Conseqüentemente, os consultores e taxonomistas não conseguem fazer uma rápida identificação de espécies, independentemente da existência de uma base de referência (DEINER et al., 2017). De fato, a descrição de caracteres morfológicos para as pesquisas requer a atenção de taxonomistas experientes, normalmente especializados na análise morfológica de algumas poucas famílias ou gêneros. A indisponibilidade do profissional taxonomista, dado o grau de especialização necessário, o tempo necessário para a realização do trabalho e lacunas no conhecimento, entre outros

aspectos, é amplamente reconhecido como impedimento taxonômico (*taxonomic impediment*) (RIEDEL et al. 2013).

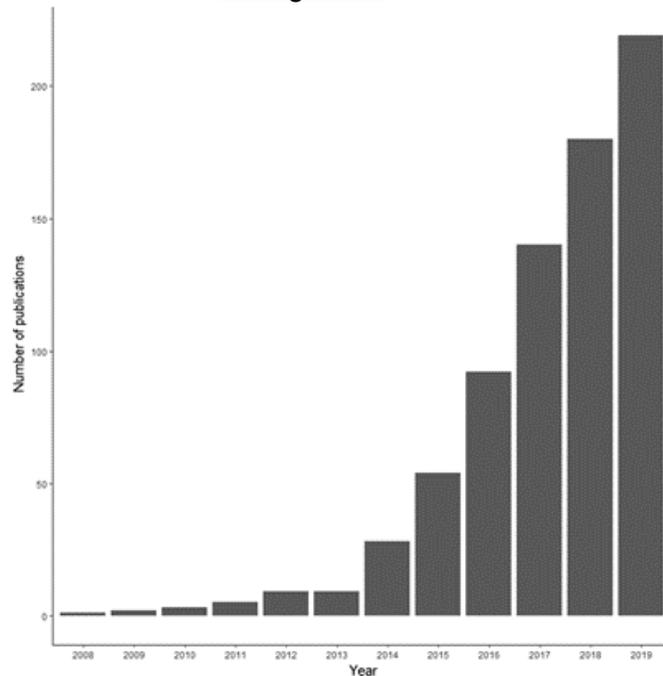
Em países megadiversos como o Brasil, a caracterização da diversidade biológica torna-se um desafio ainda maior. Devido ao acelerado processo de alterações climáticas e extinção de espécies (LEVINE et al., 2016; MIRANDA et al., 2019), Deiner et al. (2017) e Hebert et al. (2003) ressaltam que há a necessidade contínua de desenvolvimento de abordagens aprimoradas de pesquisar a biodiversidade global, e isto inclui aplicar uma tecnologia baseada na análise molecular.

Estudos recentes identificaram que o DNA isolado dos solos pode ser utilizado para caracterizar a biodiversidade de minhocas (PASUN et al., 2015), invertebrados (MCGEE e EATON, 2015) e espécies de vertebrados (ANDERSEN et al., 2012) com sucesso. O DNA encontrado no solo também inclui DNA de tecidos vegetais ativos e dormentes, sementes, pólen e detritos de plantas, e pode potencialmente revelar a diversidade total de plantas de um local (HIIESALU et al., 2011; YOCCOZ et al., 2012;). Nesse sentido, o DNA *metabarcoding* é o método usado para verificar quais espécies estão presentes na amostra ambiental e sua taxonomia, sem a necessidade de amostragem direta do espécime vegetal (BARSOUM et al., 2018; COMPSON et al., 2020).

A técnica de DNA *metabarcoding* é uma aplicação emergente de ferramentas moleculares para a descrição da biodiversidade, contando com mais de 200 artigos publicados no ano de 2019 (Figura 6) (DORMONTT et al., 2018; KINGSLY e CORLETT, 2020). O potencial para as aplicações de DNA *metabarcoding* pode fornecer identificação confiável e com bom custo-benefício, com a perspectiva de determinar rapidamente a composição de espécies de qualquer amostra, seja para fins de monitoramento ambiental (KIMMERLING et al., 2018), análise de dieta (BUGLIONE et al., 2018), biossegurança (BORRELL et al., 2017) e detecção de comércio ilegal (DE BOER et al., 2017). Porém, de todos os grupos taxonômicos aos quais DNA *metabarcoding* foram aplicados, a identificação de plantas ainda é pouco representada na literatura (Figura 7). Logo, mostra-se necessário a continuidade desses estudos, uma vez que o DNA *metabarcoding* surge como uma forma para a identificação taxonômica de grupos muito diversificados, já que as plantas poderão

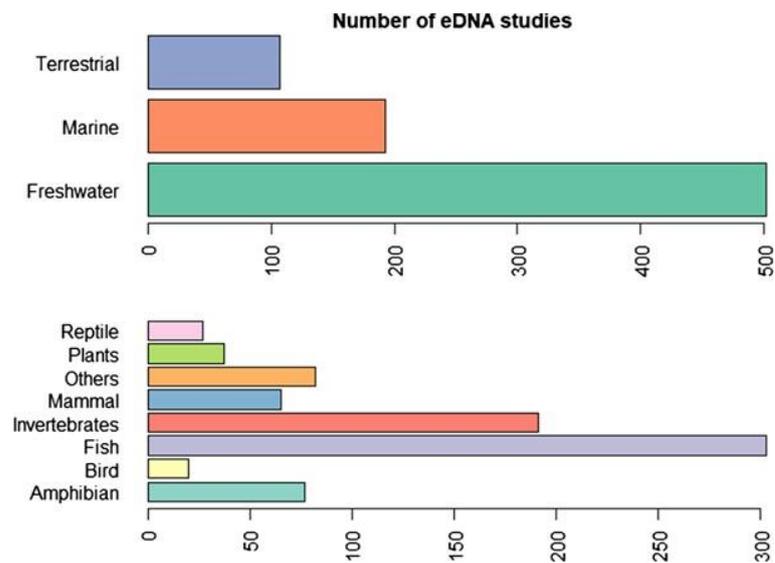
ser analisadas em grupo através do sequenciamento em massa da amostra ambiental (COMPSON et al., 2020; KINGSLEY e CORLETT, 2020).

**Figura 6** - Número de estudos usando as palavras 'DNA ambiental' ou 'eDNA' para o período entre 1 de janeiro de 2008 e 31 de dezembro de 2019 recuperado de uma pesquisa bibliográfica.



Fonte: KINGSLEY e CORLETT (2020)

**Figura 7** - Número de estudos a partir de grupo de organismo e ecossistema diferentes usando as palavras 'DNA ambiental' ou 'eDNA' para o período entre 1 de janeiro de 2008 e 31 de dezembro de 2019, recuperado de uma pesquisa bibliográfica.



Fonte: KINGSLEY e CORLETT (2020).

O DNA extraído de amostras coletadas no ambiente (ar, fezes, solo e água) é conhecido como DNA ambiental (eDNA – *environmental DNA*). Uma única amostra de DNA ambiental, a exemplo do solo/sedimento poderá conter traços de DNA de diferentes organismos (Figura 8), possibilitando detectar a presença de diversas espécies (TEDERSOO et al., 2016).

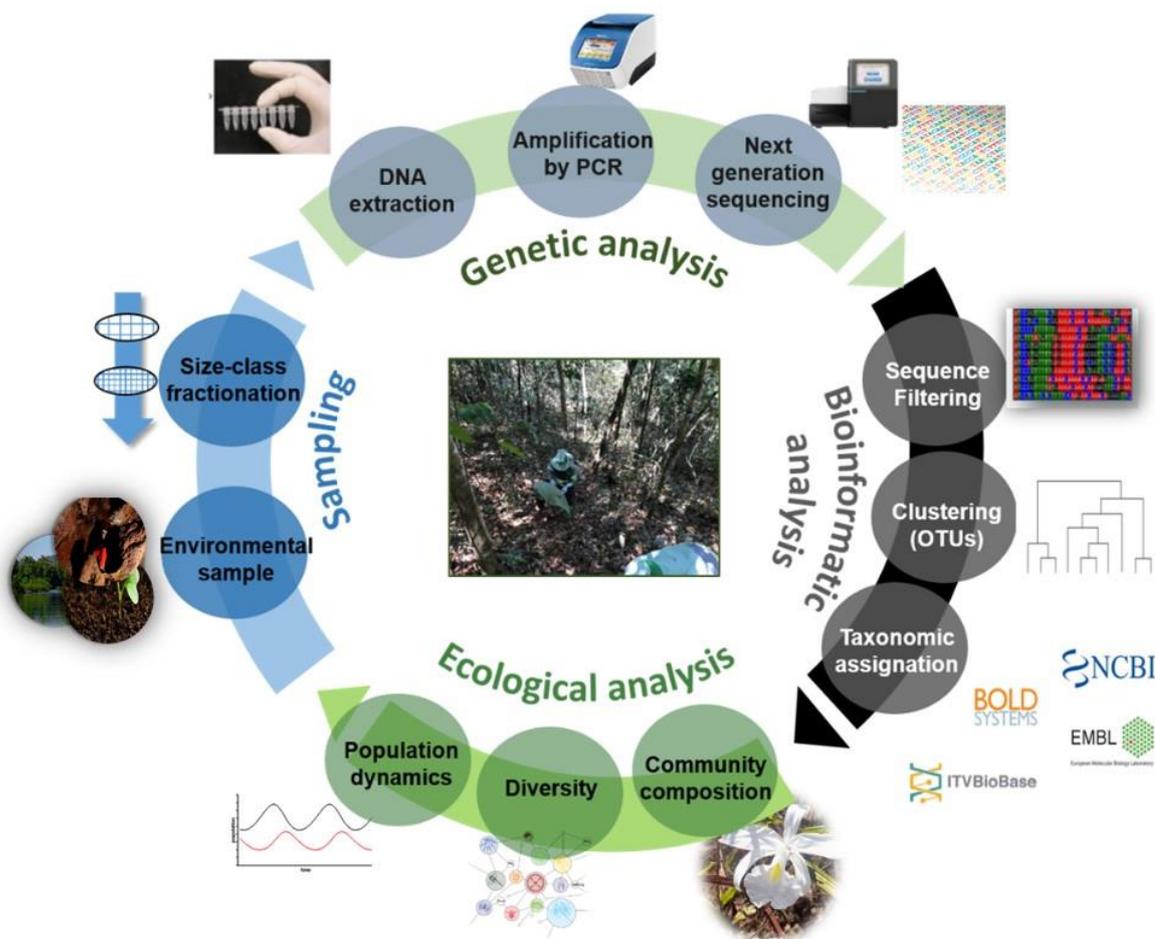
Para a distinção das espécies, usa-se de marcadores moleculares curtos dos genomas dos organismos alvo do estudo. O eDNA total é extraído e posteriormente amplificado por meio de uma reação em cadeia da polimerase (PCR – *Polymerase Chain Reaction*). A região do DNA escolhida depende de cada grupo taxonômico, e é importante selecionar uma região conservada para o uso de iniciadores (*primers*) universais. O locus alvo deve apresentar um tamanho adequado para facilitar a amplificação e sequenciamento, além de conter variabilidade genética idealmente suficiente para permitir que os organismos sejam identificados a nível de espécie (BARSOUM et al., 2015).

A sequência gerada é comparada às sequências que estão depositadas no banco de referências genéticas para a determinação taxonômica. A construção desses bancos referenciados é um pré-requisito para a obtenção de resultados confiáveis de forma robusta, porém, espécies anteriormente não descritas por um código de barra, ou seja, sequências sem correspondências nos bancos de dados, também podem ser detectadas e a elas será atribuída uma unidade taxonômica operacional (OTU – *Operational Taxonomic Unit*) (BARSOUM et al., 2018).

Segundo Kelly (2016), as maneiras pelas quais o DNA *metabarcoding* pode complementar e estender pesquisas convencionais são promissoras, e com o desenvolvimento/aperfeiçoamento da técnica é possível verificar quais são os desafios no campo e no laboratório. Diversos estudos como Cheng et al. (2016), Zun et al. (2017), Zhao et al. (2018), Moonhouse-Gann et al. (2018) e Timpano, Scheible e Meiklejohn (2020), relatam que tiveram melhores resultados em análises de códigos de barra de plantas usando o espaçador interno transcrito (ITS – *Internal Transcribed Spacer*), em particular a região ITS2, do DNA ribossômico 35S do genoma nuclear. Isso se deve principalmente aos altos níveis de divergência interespecífica, e, portanto, alta resolução taxonômica, que são maiores do que em códigos de barra plastidiais curtos (CHEN et al., 2010). Além disso, a região ITS/ITS2 apresenta bons

níveis de universalidade, qualidade de sequências geradas, boa taxa de rendimento/custo e menos problemas de amplificação do que outros marcadores comumente utilizados (CHINA PLANT BOL GROUP et al., 2011; HOLLINGSWORTH, GRAHAM e LITTLE, 2011; HOLLINGSWORTH et al., 2016).

**Figura 8** - Fluxo de trabalho DNA *metabarcoding*.

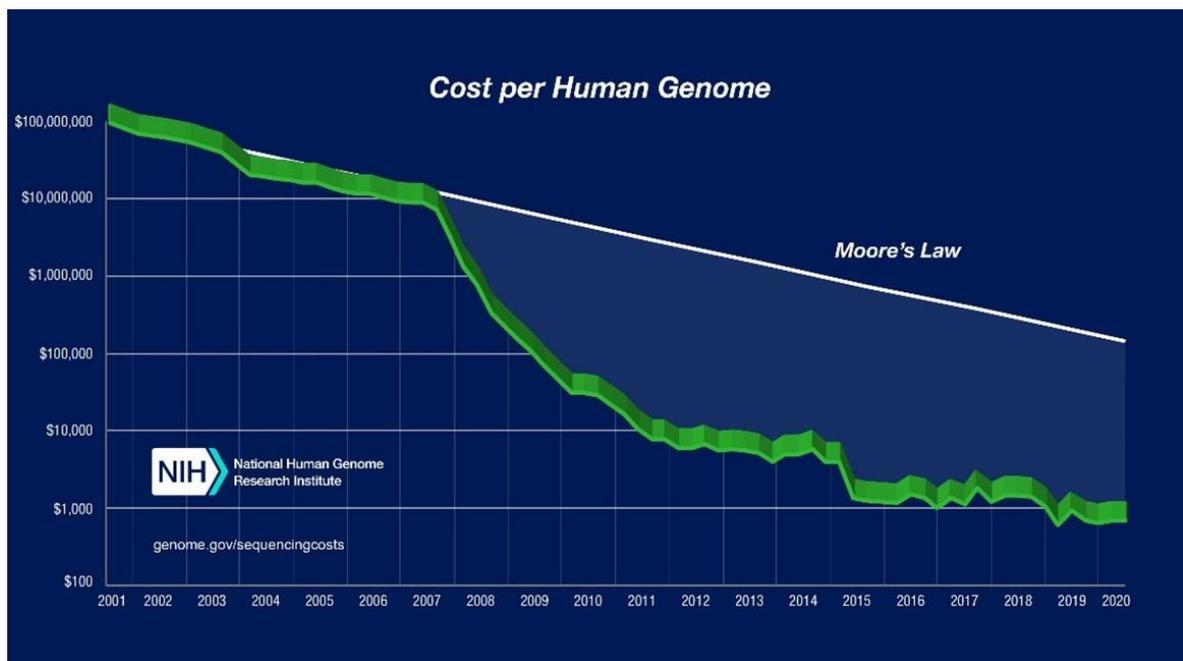


**Fonte:** R. Siano & S. Arnaud-Haond, com adaptações, 2016.

O desenvolvimento de métodos moleculares, tanto para a geração dos dados, quanto para as análises destes, permitiu que um grande número de leituras de sequências de DNA de alta qualidade seja obtido de forma relativamente barata, a partir de amostras de DNA ambiental em uma única execução de sequenciamento (COMPSON et al., 2020; YOUNG et al., 2017). Ressalta-se que, a popularização das tecnologias para sequenciamento massivo de alto rendimento (High Throughput Sequencing), durante os últimos treze anos, revolucionou o campo da pesquisa genômica, promovendo o sequenciamento de genoma com baixo custo e em um

período de tempo muito curto (PARK e KIM, 2016; DEINER et al., 2017). Por exemplo, o custo de sequenciamento de um genoma de tamanho humano (3.000 Mb) em 2020 é cerca de US \$ 1.000, sendo que o genoma de US \$ 100 pode estar à vista (Figura 9). Isso é quase uma redução de 10.000 vezes no custo de sequenciamento de um genoma humano desde 2004 (STEPHENS et al., 2015; PARK e KIM,2016).

**Figura 9** - Custo do sequenciamento do genoma.



**Fonte:** <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>

Espera-se que nos próximos anos as plataformas de sequenciamento de terceira geração (de leitura longa), por exemplo Nanopore e PacBio sequencing, aumentem a capacidade e a velocidade de sequenciamento, enquanto reduzem os custos de maneira muito significativa (KARST et al., 2021). Nota-se, que as novas tecnologias de sequenciamento do DNA geram uma grande quantidade de informações que só é possível ser analisada por computadores de alto desempenho (COMPSON et al., 2020; KINGSLY e CORLETT, 2020), e o *know-how* para lidar, processar e extrair padrões dos dados, leva a aumentar os custos para a interpretação do genoma, exigindo diferentes níveis de treinamento em bioinformática e novas abordagens, como a Inteligência Artificial (LOGARES et al., 2012; YOUNG et al., 2017; GERHARD e GUNSCH., 2019). A questão importante é que esses campos

estão agora se tornando estreitamente conectados e interdependentes, e o avanço de um requer o avanço de outros (LOGARES et al., 2012).

Ou seja, ocorreu uma mudança de ciência descritiva para uma ciência quantitativa e com base em grandes volumes de dados (EDMUNDS et al., 2013). Entretanto, a produção massiva de dados requer *pipelines* computacionais capazes de lidar com grande volume de informação. Nessa circunstância, a bioinformática tem contribuído significativamente para a geração de *pipelines* e apresentação dos dados obtidos, além da investigação de novas e complexas perguntas sobre a biologia dos organismos (RUPPERT et al., 2019).

Portanto, são necessárias pesquisas contínuas em diferentes ambientes para entender quais são os desafios para uma detecção de espécies robusta a partir de eDNA, bem como, a dinâmica espaço-temporal complexa dos vários tipos de amostras ambientais, sobre os quais atualmente sabemos muito pouco (DEINER et al., 2017). Com o desenvolvimento da aplicação da abordagem de DNA *metabarcoding* para as cangas da Serra de Carajás, será possível levar, de uma maneira analítica, o conhecimento taxonômico existente sobre a região para o mundo digital, ajudando a criar novas formas de identificar e monitorar espécies, possibilitando uma comparação entre os dados existente e estudos futuros, e conseqüentemente a implementação da técnica em larga escala para o monitoramento da biodiversidade.

## **4 MATERIAIS E MÉTODOS**

### **4.1 COLETA DAS AMOSTRAS**

As amostras de solo foram coletadas nos dias 20, 21 e 22 de agosto de 2019, numa profundidade rasa (<50 cm), no período de seca da região, em três parcelas de 20x5 m na área do Projeto Vale S11D, Serra Sul, Canaã dos Carajás, Pará. Os *plots* amostrados foram os mesmos utilizados para o levantamento tradicional da flora pela empresa de consultoria (Inventário florístico e florestal em 422,89 hectares de vegetação da Serra Sul, projeto Vale S11D, Canaã dos Carajás, Pará, Amplo Consultoria, maio de 2016) (Figura 11). Os detalhes sobre os blocos de amostragem estão nas Figuras 12 e 13. Fotos do local de coleta também foram obtidas (Figura 10A) e informações adicionais sobre os metadados estão no Apêndice B (Tabela 3, Tabela 4 e Tabela 5). Um total de 20 amostras de solo foram coletadas por parcela, totalizando 60 amostras. Contudo, devido à dificuldade de acesso à área de coleta,

uma amostra da parcela 168 (PCG-168) foi perdida. A coleta foi realizada utilizando sacos estéreis (*Whirl-pak*), espátulas esterilizadas, luvas estéreis e álcool 70% (Figura 10B). O pacote com as amostras foi etiquetado para a alocação de um ID único (Tabela 1). As amostras de solo foram armazenadas em isopor contendo gelo químico congelado previamente *overnight* a  $-20^{\circ}\text{C}$ . Ao término do dia, as amostras de solo foram congeladas a  $-20^{\circ}\text{C}$  e transferidas para o laboratório do ITV, em Belém, e armazenadas a  $-80^{\circ}\text{C}$  até a extração de DNA.

**Figura 10.** (A) Exemplo de um ponto de amostragem em campo. (B) Saco estéril (*Whirl-pak*) e espátula usados na coleta.



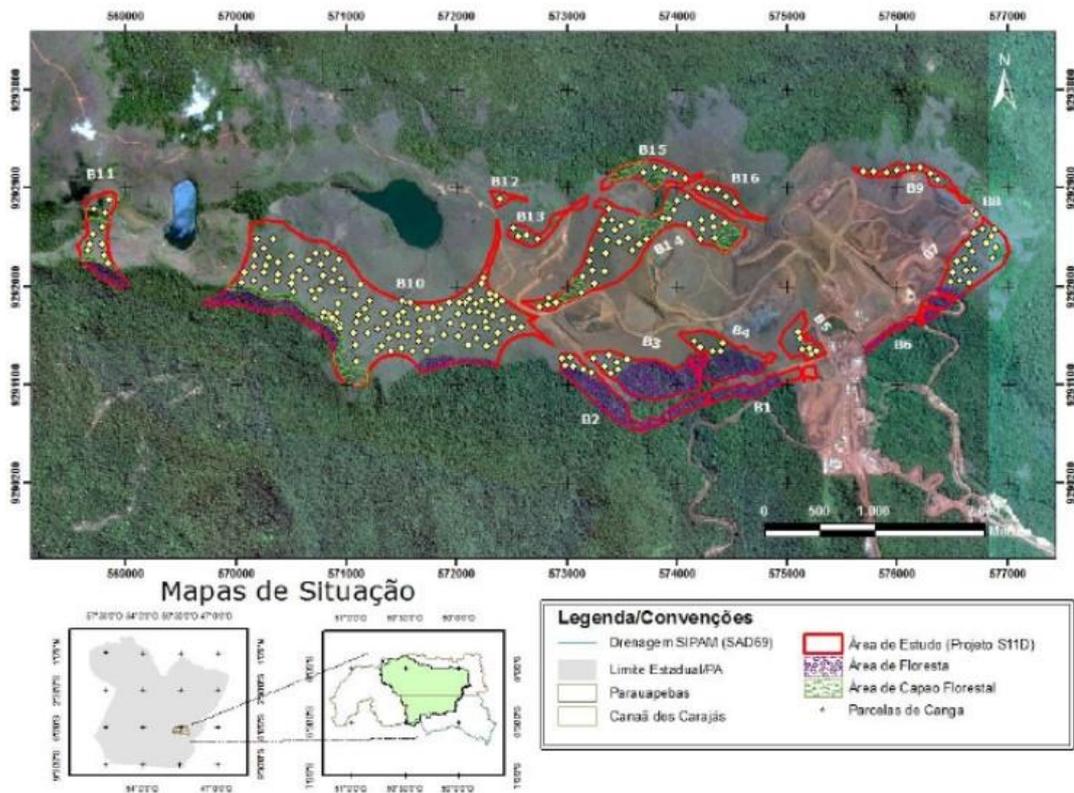
A



B

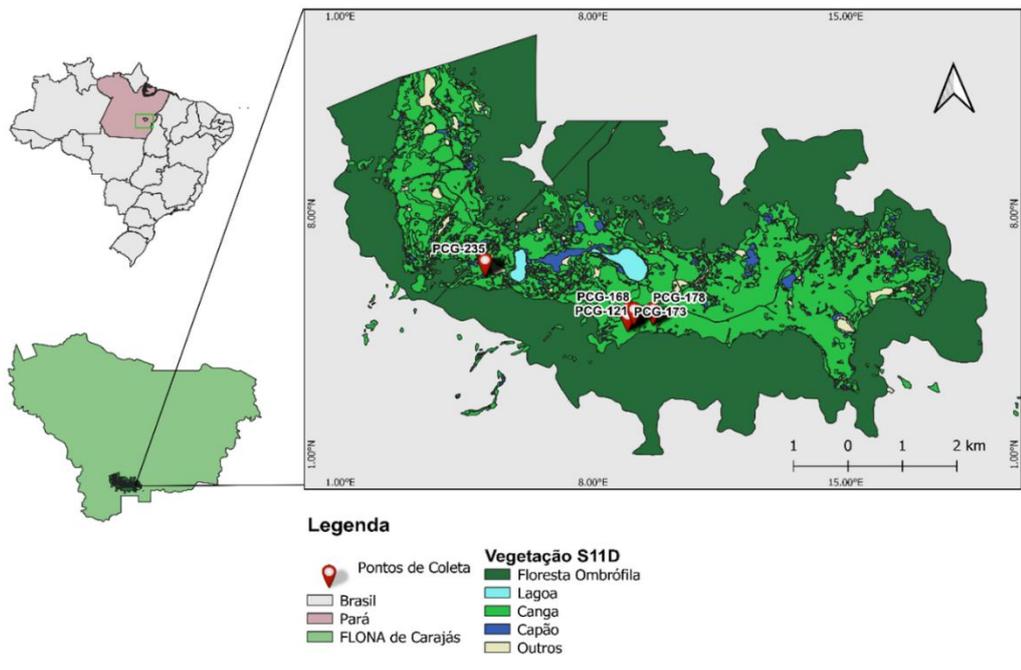
Fonte: Autora, 2021.

**Figura 11.** Localização das parcelas inventariadas pela empresa de consultoria, na Canga; Projeto Vale S11D, Inventários florísticos, Serra Sul, Canaã dos Carajás, Pará.



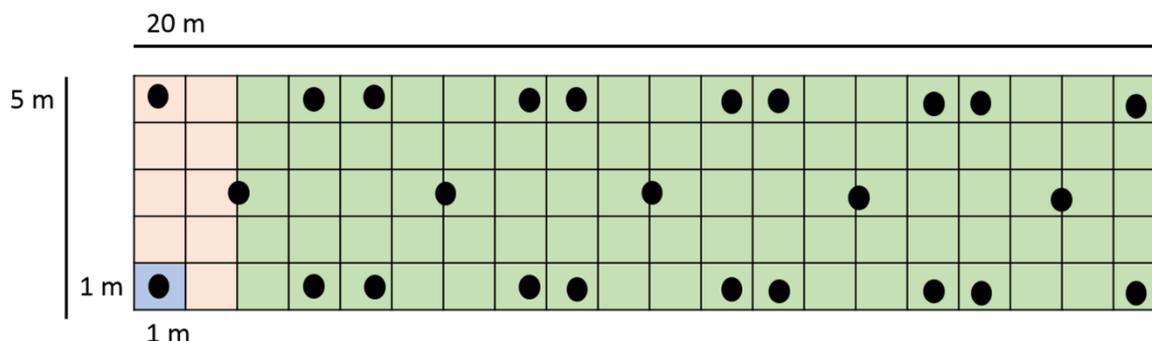
Fonte: Consultoria, 2016.

**Figura 12.** Plots selecionados dentro da área do Projeto Vale S11D, Serra Sul, Canaã dos Carajás, Pará.



Fonte: Autora, 2021.

**Figura 13.** Desenho experimental da amostragem em plots de canga.



**Fonte:** Gastauer (2019).

**Tabela 1.** Coordenadas geográficas, Leste e Norte, das três parcelas estudadas na região do projeto S11D, em Canaã dos Carajás-PA. Analisados – blocos estudados e relatados neste documento.

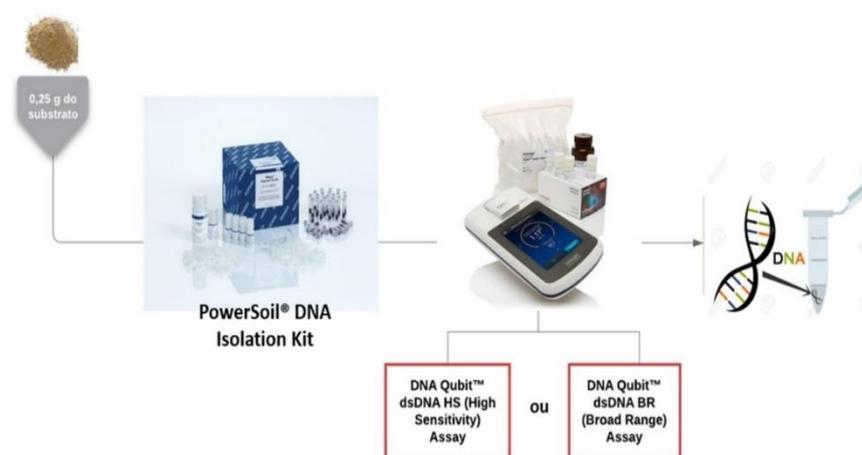
Bloco	Parcela	Tipologia	Zona	Leste	Norte	Amostras
B10	PCG-121	Canga	22M	571466,470	9291729,971	ITV21580-ITV21599
B10	PCG-168	Canga	22M	571502,316	9291865,336	ITV21541-ITV21560
B11	PCG-235	Canga	22M	568845,310	9292791,845	ITV21620-ITV21639

**Fonte:** Autora, 2021.

#### 4.2 EXTRAÇÃO DE DNA

A extração de DNA foi realizada a partir de 0,25 g de solo, em triplicata, utilizando o kit DNeasy PowerSoil (QIAGEN, Hilden, Germany), seguindo as instruções do fabricante. Controles negativos de extração foram utilizados para cada procedimento de extração de DNA realizado. No final, as amostras foram eluídas em 100 µL do tampão do kit e quantificadas no fluorômetro Qubit® 3.0 (Thermo Fisher Scientific) com o kit Qubit™ dsDNA HS Assay (Thermo Fisher Scientific). As informações das concentrações de DNA para cada uma das amostras estão descritas no Apêndice B (Tabelas 6-8). Após o isolamento e a quantificação, os produtos da extração foram armazenados à -20 °C, sendo posteriormente utilizados na etapa de amplificação do DNA e construção das bibliotecas Illumina.

**Figura 14.** Fluxograma simplificado mostrando as etapas de extração do DNA total das amostras de solo.



**Fonte:** Compilação da autora<sup>1</sup>, 2021.

#### 4.3 SELEÇÃO DE PRIMERS

Pares de *primers* direcionados para a amplificação das regiões de ITS foram testados para a adequação do protocolo de sequenciamento nas plataformas NGS (Ion Torrent e Illumina MiSeq). De acordo com estudos de Moonhouse-Gann et al. (2018) e Timpano, Scheible e Meiklejohn (2020), o espaçador interno transcrito (ITS) do rDNA nuclear é o principal marcador usado em análises ambientais de código de barras de plantas a partir de solo. Após a verificação de eficiência de diferentes combinações de primers, o par ITS2F (sequência 5'-3': ATGCGATACTTGGTGTGAAT) e ITS3R (sequência 5'-3': GACGCTTCTCCAGACTACAAT) (CHEN et al., 2010) apresentou melhores resultados de amplificação, considerando a diversidade de grupos taxonômicos na canga, padronização e automação razoável dos protocolos para processamento e análise das amostras de eDNA do presente trabalho e foi selecionado para as etapas posteriores de amplificação e preparo das bibliotecas para sequenciamento na plataforma Illumina MiSeq.

<sup>1</sup> Montagem a partir de imagens coletadas nos sites da Qiagen (<https://www.qiagen.com/>), ThermoFisher Scientific (<https://www.thermofisher.com/>) e via Malgorzata Slusarczyk (<https://br.123rf.com/>).

#### 4.4 PROCEDIMENTO DE AMPLIFICAÇÃO DO DNA

As reações de amplificação foram realizadas em triplicatas independentes para cada amostra, incluindo controles negativos de amplificação. O tamanho do fragmento esperado estava em torno de 350 - 450 bp. As reações de PCR foram realizadas conforme descrito a seguir: 2,5 µL de tampão 5X, 1,2 µL de MgCl<sub>2</sub> 25 mM, 1,0 µL de dNTP 2 mM, 1,0 µL de DMSO, 0,25 µL de cada *primer* forward e reverse a 10 pmol, 0,1 U de Taq (Promega), 2 µL de DNA, 2,5 µL de TBT 5X (SAMARAKOON et al., 2013) e 1,7 µL de água ultrapura ou Nuclease- Free Water (ThermoFisher Scientific), para um volume final de 12,5 µL. As PCRs foram realizadas em termocicladores Veriti 96-Well Thermal Cycler (Applied Biosystems), com as seguintes condições: desnaturação inicial a 94°C por 3 min, seguida por 30 ciclos de amplificação com 1 min a 94 °C, 1 min a 54 °C, em seguida 1 min a 72 °C e a extensão final por 7 min a 72 °C.

#### 4.5 PREPARAÇÃO DA BIBLIOTECA E SEQUENCIAMENTO NGS

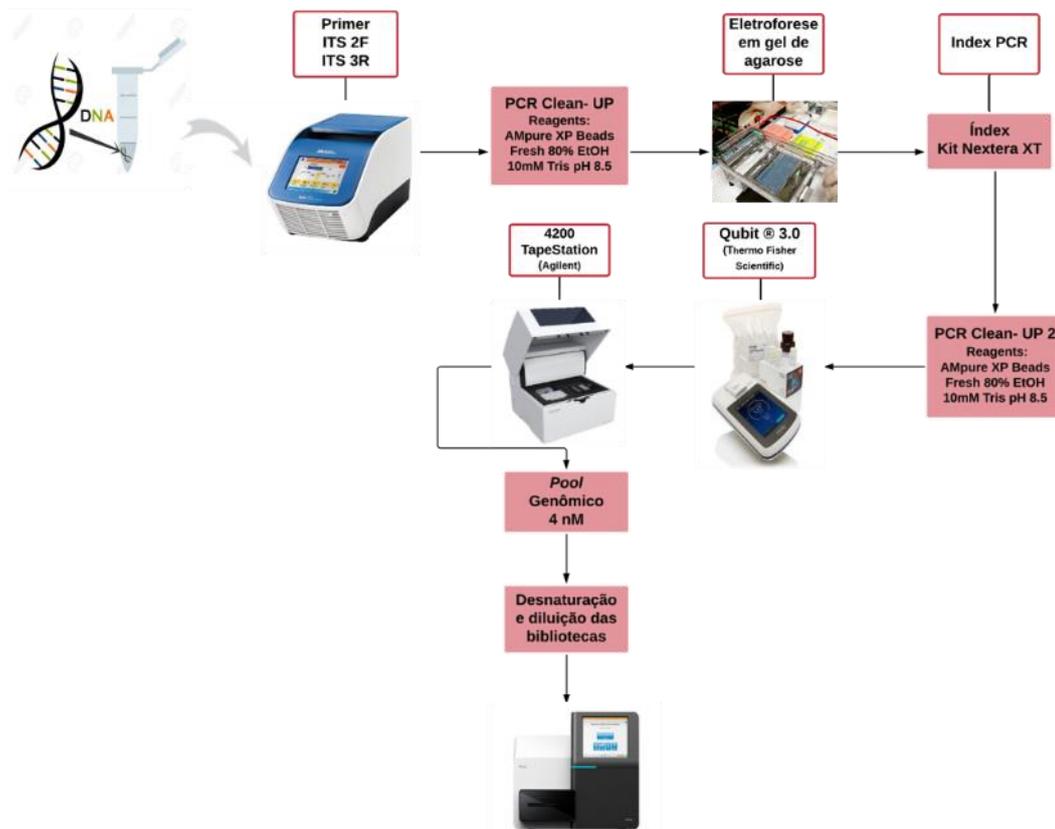
As bibliotecas foram construídas de acordo com o protocolo 16S Metagenomic Sequencing Library Preparation da Illumina (Illumina, San Diego, CA, USA). Todas as triplicatas de PCR foram misturadas antes dos procedimentos de purificação. A primeira purificação do DNA amplificado foi realizada com o kit de esferas magnéticas Agencourt Ampure XP beads (Bechman Coulter, Inc., Brea, EUA), seguindo o recomendado no protocolo. Para avaliar a qualidade dos amplicons, as amostras foram submetidas à eletroforese em gel de agarose 1%, corados com corante fluorescente Syber Safe Thermo Fisher Scientific (10 µL para cada 100 mL de gel de agarose). Foram utilizados 5 µL de produto de PCR, 1 µL de DNA-ladder 50 pb (Invitrogen™) e 1,0 µL de corante azul de bromofenol. A eletroforese foi realizada em tampão TBE 1X, por 40 min a 90 V. Para uma verificação mais precisa dos tamanhos dos amplicons, foram realizadas eletroforeses em capilares com um Bioanalyzer Agilent Technology 2100 (Agilent Technologies), com o kit Agilent DNA 1000 (Agilent Technologies).

Posteriormente, os índices (adaptadores de sequências únicas) do kit Nextera XT Library prep (Illumina, San Diego, CA, USA) foram adicionados a cada amostra conforme instruções do protocolo 16S. Uma nova PCR com os índices foi realizada utilizando 5 µL de DNA, 5 µL de Nextera XT index Primer1, 5 µL de Nextera XT index

Primer2, 25 µL de enzima 2X Kapa Hifi HotStart Ready Mix e 10 µL de água ultrapura ou Nuclease- Free Water (ThermoFisher Scientific), para um volume final de 50 µL. A etapa de Index PCR foi feita com os seguintes passos: desnaturação inicial a 95 °C por 3 min, seguida por oito ciclos de amplificação por 30 s a 95°C, 30 s a 55 °C, 30 s a 72 °C e a extensão final por 5 min a 72 °C.

O DNA foi purificado novamente conforme descrito previamente e as bibliotecas foram quantificadas no fluorômetro Qubit 3.0 (Thermo Fisher Scientific) com o kit Qubit™ dsDNA HS Assay (High Sensitivity) (Thermo Fisher Scientific). A verificação dos tamanhos dos amplicons foi realizada em um Tapestation System 4200 (Agilent Technologies), utilizando o kit DS1000 (Agilent Technologies). Posteriormente, o pool das bibliotecas foi padronizado para concentração de 4 nM, adicionando 20% do padrão PhiX. A corrida de sequenciamento foi realizada utilizando o kit de corrida MiSeq V3 600 ciclos (Illumina) em um sequenciador Illumina MiSeq System do Laboratório de Genética Humana e Médica no Instituto de Ciências Biológicas da Universidade Federal do Pará (Belém, Pará).

**Figura 15.** Fluxograma simplificado mostrando as etapas de construção de bibliotecas para o sequenciamento em MiSeq



Fonte: Compilação da autora<sup>2</sup>, 2021.

#### 4.6 ANÁLISES BIOINFORMÁTICAS

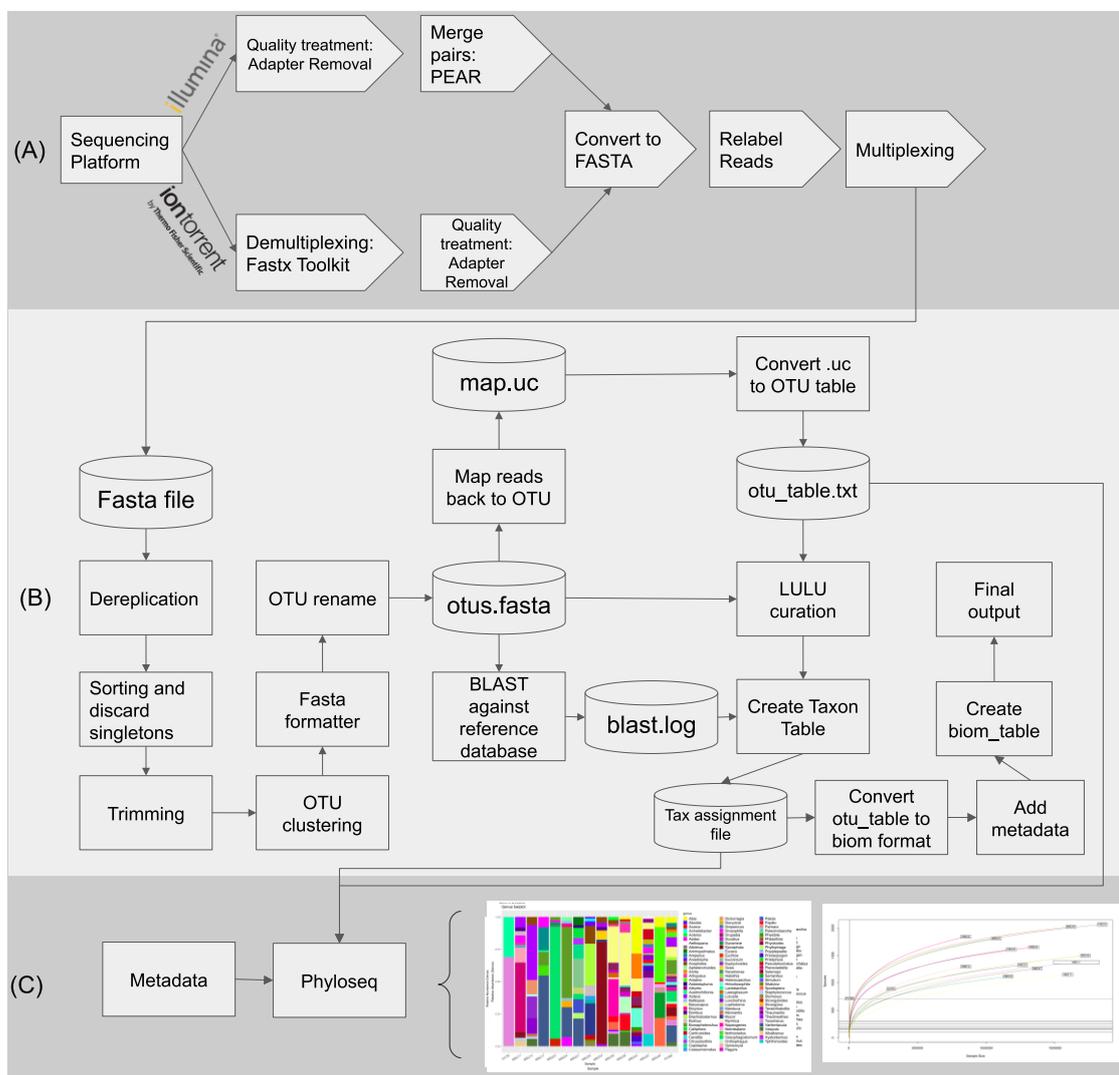
As sequências brutas geradas foram submetidas ao *pipeline* PIMBA (OLIVEIRA et al., 2021.) para análise de dados de DNA *metabarcoding*, o qual permite a personalização do banco de dados de referência para etapa de classificação taxonômica. Para análises de DNAr, o PIMBA (Figura 16) utiliza o *pipeline* QIIME (CAPORASO et al., 2010) como base. Primeiramente, a etapa de “trimagem” (retirada de contaminações) e filtragem por qualidade (Phred >20) foi realizado com a ferramenta PRINSEQ (CANTU, SANDURAL e EDWARDS, 2019). Posteriormente, somente as sequências *forward* e *reverse* de alta qualidade foram montadas através do montador Pear (ZHANG et al., 2014). Após a montagem, ocorreu a etapa de

<sup>2</sup> Montagem a partir de imagens coletadas nos sites da Qiagen (<https://www.qiagen.com/>), ThermoFisher Scientific (<https://www.thermofisher.com/>), Agilent (<https://www.agilent.com/>), Illumina (<https://www.illumina.com/>), Biosystems (<http://www.biosystems.com.br/>) e via Malgorzata Slusarczyk (<https://br.123rf.com/>).

derreplicação, na qual foram removidas todas as duplicadas e em seguida a abundância das sequências foi aferida, descartando aquelas com contagens únicas. Como controle adicional de qualidade, todas as sequências menores que 100 pb foram descartadas. As bibliotecas obtidas foram normalizadas.

Testes foram realizados para verificar a melhor porcentagem de similaridade para o agrupamento das OTUs e para a determinação taxonômica, de acordo com os parâmetros utilizados no QIIME e considerando as listas de levantamentos taxonômicos tradicionais para as áreas. Os parâmetros que apresentaram os melhores resultados estão descritos a seguir: as sequências foram agrupadas em Unidades Taxonômicas Operacionais (OTUs) utilizando o algoritmo USEARCH 7 (<https://www.drive5.com/usearch/>), com base na similaridade de 97% (otu\_similarity); A classificação taxonômica das OTUs foi realizada com base em sequências acima de 95% de identidade (assign\_similarity); e a cobertura mínima para utilização das sequências foi de 70%. Houve a filtragem das sequências de fungos. As OTUs sem taxonomia atribuída (unassigned) não foram consideradas para as análises posteriores (ALBERDI et al., 2018; LEAR et al., 2018). A atribuição taxonômica das OTUs foi realizada com o banco de referências genéticas de plantas de Carajás disponível no ITVBiobase, sem cortes adicionais no comprimento das sequências obtidas (otu\_lenght zero) e observando os 10 melhores resultados da busca BLAST. O banco de dados do NCBI também foi utilizado para comparar os táxons com a lista de plantas registradas para as áreas de cangas de Carajás

**Figura 16.** Fluxograma simplificado mostrando as etapas de análise do pipeline PIMBA: A Pipeline for *MetaBarcoding Analysis*.



Fonte: Oliveira et al., 2021.

#### 4.7 ANÁLISES ESTATÍSTICAS

Todas as análises estatísticas foram realizadas no programa R v.3.6.3 (R Core Team 2016) através dos pacotes phyloseq v.1.30.0 (MCMURDIE e HOLMES, 2013), vegan v.2.5-6 (OKSANEN et al., 2016) e ggplot2 v.3.3.2 (WICKHAM, 2009). Para testar a saturação da geração de reads por amostra, foi construída a curva de rarefação utilizando a função *pimba plot* através do PIMBA. As abundâncias relativas foram transformadas em dados de presença-ausência devido à grande quantidade de variáveis que influenciam a abundância de fragmentos observados em amostras de DNA ambiental sequenciado (KINGSLY e CORLETT, 2020). Com a matriz de dados de presença/ausência atribuídas a nível de família, gênero e espécies, foi calculado o

número de vezes que cada amostra foi identificada a nível de espécie, e então calculadas as frequências relativas das parcelas. A composição taxonômica foi representada utilizando o `plot_bar` gerado através pacote `phyloseq`. Utilizando os pacotes `heatmap3` v.1.3.7, `RColorBrewer` v1.1-2 e `phyloseq`, foi realizado um *heatmap* com a matriz de dados de presença/ausência para verificar a frequência de ocorrência das espécies e famílias por parcela. Os diagramas de Venn foram calculados utilizando os pacotes `VennDiagram` v.1.6.20 e `ggplot2`.

Com o objetivo de avaliar a riqueza de espécies, foi calculada a diversidade alfa para as parcelas e amostras. A análise de diversidade alfa foi realizada utilizando a função `plot_richness` do pacote `phyloseq`, adotando os índices de Chao e Shannon. Em seguida, foi realizado um teste ANOVA unilateral para identificar diferenças nas médias ( $p < 0,005$ ), bem como um teste post-hoc de Tukey HSD após verificação de normalidade e homogeneidade de variância. A diversidade beta foi estimada com o pacote `phyloseq`, a qual foi posteriormente avaliada por uma análise de coordenadas principais (PCoA), com o índice de Jaccard. Para verificar se havia diferenças significativas entre as parcelas foi realizado um teste de análise de variância multivariada permutacional (PERMANOVA), usando as funções `adonis`, `betadisper` e `permutest` do pacote `vegan`, seguindo com um teste post-hoc de Tukey (HSD). Para verificar a existência de dependência espacial na comunidade de plantas, foi gerada uma matriz de distância da composição de espécies pelo índice Sorensen, por meio da função `vegdist` do pacote `vegan`. O teste de Mantel foi realizado para avaliar a correlação entre as duas matrizes de dissimilaridade, usando o comando `mantel.test` do pacote `Ade4` v.1.7-13 (DRAY et al., 2007), com base na correlação de Spearman e 999 permutações das matrizes.

## 5 RESULTADOS E DISCUSSÃO

### 5.1 LEVANTAMENTO DA FLORA POR MEIO DE TAXONOMIA CLÁSSICA

O Inventário florístico e florestal de vegetação da Serra Sul submetido pela Consultores à empresa Vale foi obtido através da Gerência de Meio Ambiente Corredor Norte. O relatório enviado data de 2016 e o Apêndice A (Tabela 2) lista as espécies encontradas pela consultoria nos mesmos pontos de amostragem utilizados para o levantamento da flora por DNA *metabarcoding*.

### 5.2 LEVANTAMENTO DA FLORA VIA DNA *METABARCODING*

#### 5.2.1 Amostras extraídas

Para a avaliação da eficiência da abordagem de DNA *metabarcoding* com eDNA de solos da canga, três parcelas coletadas foram selecionadas: PCG-235, PCG-168 e PCG-121, totalizando 59 amostras. As informações das concentrações de DNA estão no Apêndice C. Todos os cinco controles negativos de extração apresentaram concentração de DNA abaixo do detectável, conforme descrito no Apêndice C (Tabelas 6-8). Os perfis das bibliotecas geradas com os tamanhos dos amplicons podem ser visualizados no Apêndice D e E.

#### 5.2.2 Sequenciamento NGS e pré-processamento

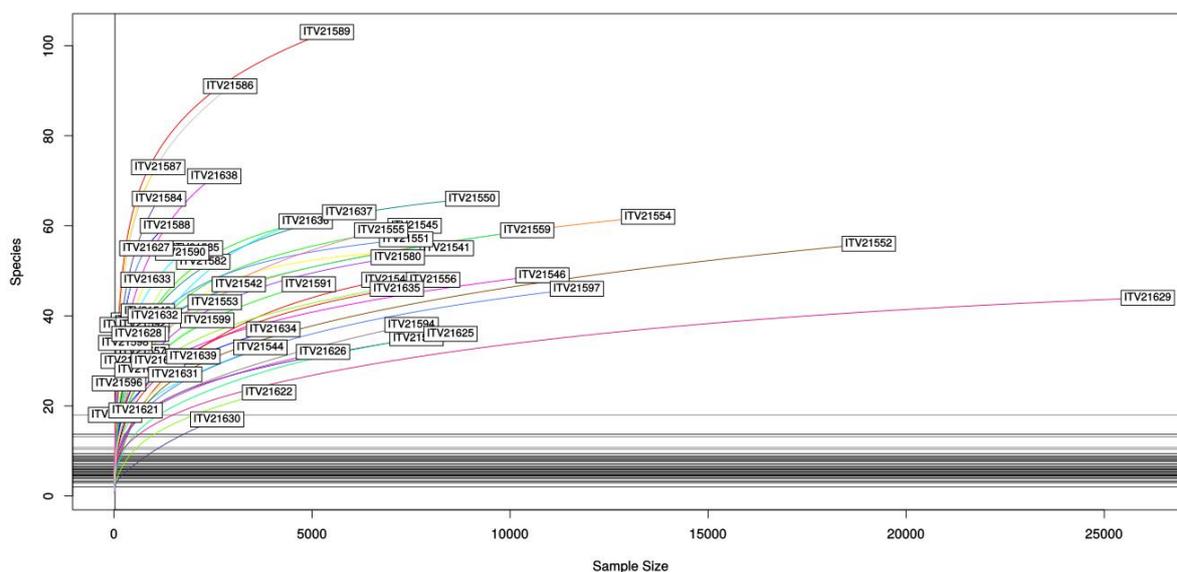
O volume bruto de dados produzidos e após o tratamento de qualidade, considerando *reads* e pares de base, estão indicados no Apêndice F. As bibliotecas foram normalizadas. Em média, foram gerados 270.284 reads e 79.749.562 pb de dados brutos por amostra, restando 232.350 reads e 139.108.021 pb após o tratamento de qualidade. No entanto, a curva de rarefação obtida (Figura 17) evidencia a necessidade de rodadas adicionais de sequenciamento para parte das amostras testadas.

Estudos anteriores de DNA *metabarcoding* utilizando o marcador ITS2 para caracterização de plantas a partir de amostras do solo, indicam um número maior de rodadas de sequenciamento sendo realizadas (FLØJGAARD et al., 2019), gerando, em média, 1.000.000 à 10.000.000 reads (TIMPANO, SCHEIBLE e MEIKLEJOHN, 2020; URUMARUDAPPA et al., 2020), tendo um número maior de sequências alvo recuperadas por amostra (~10.000 sequências / amostra) (TIMPANO, SCHEIBLE e MEIKLEJOHN, 2020; FAHNER et al., 2016; FLØJGAARD et al., 2019). Todavia, a profundidade necessária para localidades menos diversas ou *primers* mais

específicos é provavelmente menor, havendo a necessidade de avaliação prévia com uma amostragem mais reduzida (GREY et al., 2018).

Timpano, Scheible e Meiklejohn (2020) relataram que diferenças principais na metodologia nesses estudos podem explicar o aumento da recuperação de sequências, por exemplo: 1) o solo foi armazenado a  $-80^{\circ}\text{C}$  antes da subamostragem para preservar o material biológico (FAHNER et al., 2016), 2) múltiplas extrações de DNA foram concluídas por amostra (com o DNA resultante agrupado para aumentar a quantidade inicial para PCR (FAHNER et al., 2016), 3) uma única extração utilizando uma grande quantidade de solo como entrada (FLØJGAARD et al., 2019).

**Figura 17.** Curva de rarefação do sequenciamento por amostra.



Fonte: Autora, 2021.

### 5.2.3 Atribuição taxonômica

Com relação ao teste de configuração de porcentagens utilizadas no PIMBA, com base na identificação morfológica prévia como parâmetro de calibração, observou-se que o uso de similaridades mínimas de 97% para o agrupamento em OTUs, 95% para identidade taxonômica e 70% para cobertura de alinhamento demonstrou melhorar a qualidade de atribuições taxonômicas dos conjuntos de dados obtidos. Estudos ressaltam que o limite de similaridade de 97% para o agrupamento em OTUs é, de fato, o valor padrão para a maioria dos algoritmos de agrupamento (ALBERDI et al., 2018), com base na dissimilaridade média mínima de 3% entre

espécies quando utilizados os principais marcadores alvo de *metabarcoding* (HEBERT et al., 2003). Notou-se que com o aumento no limite da identidade taxonômica para 97%, algumas espécies, como *Eugenia puniceifolia* e *Ichthyothere terminalis*, não puderam ser encontradas, enquanto que 98%, as perdas na detecção são ainda maiores. Por outro lado, com a diminuição do valor da identidade para 90%, aparecem espécies que não ocorrem na canga, ou mesmo na área de S11D, como *Tibouchina* sp. e *Rhabdodendron amazonicum*. Isso está relacionado com falhas no banco de referências, pois o banco a ser utilizado deve estar apenas com as espécies de canga, da Serra Sul, S11D e etc. Ainda, com o aumento da cobertura de alinhamento para 90%, por exemplo, há um decréscimo considerável na identificação de espécies que tipicamente ocorrem nas cangas, como *Mimosa acutistipula* var. *ferrea*, *Monogereion carajensis*, *Erythroxylum nelson-rosae* e *Vellozia glochidea*.

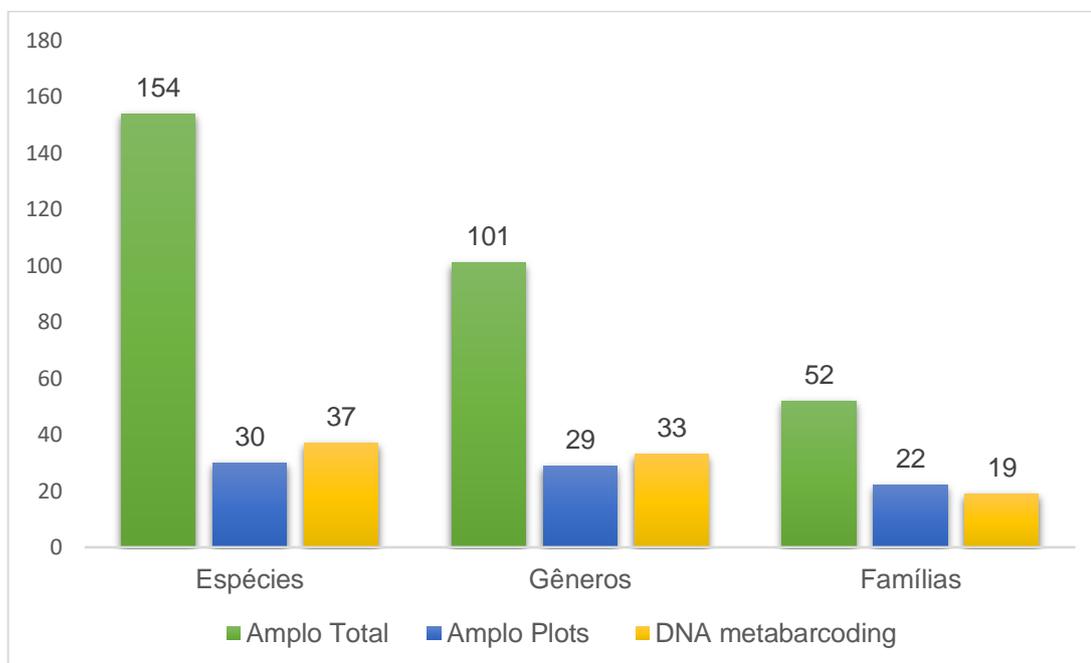
Cada marcador utilizado pode identificar um grupo diferente de organismos, dependendo da variabilidade da região e do comprimento obtido do *amplicon*, dentre outros parâmetros (ZAFEIROPOULOS et al., 2020). Embora seja difundido que a otimização de parâmetros deve melhorar a precisão ao fazer atribuições taxonômicas (DEINER et al., 2017; RUPPERT et al., 2019; COMPSON et al., 2020 e KINGSLEY e CORLETT, 2020), muitos estudos ainda empregam apenas parâmetros padronizados previamente publicados (BOGGS et al., 2019; FAHNER et al., 2016; PORNON et al., 2016), potencialmente propagando erros. Ajustes de parâmetros podem desempenhar um papel crucial na precisão das análises de DNA *metabarcoding*, evitando atribuições taxonômicas errôneas. Dessa forma, ao calibrar os parâmetros do *pipeline* aqui empregado, várias configurações puderam ser testadas para verificar os melhores ajustes para as porcentagens de similaridade para o agrupamento das OTUs, classificação taxonômica e cobertura do alinhamento, sendo possível, no futuro, implantar dentro do *pipeline* parâmetros distintos para as espécies.

No total, 565 OTUs foram identificadas, com similaridade de 97%, mas apenas 63 delas tiveram correspondências com sequências de plantas depositadas no ITVBiobase. É importante ressaltar que parte das 565 OTUs que não tiveram as sequências identificadas podem estar relacionadas com organismos que não estão depositados no ITVBiobase, como fungos, cianobactérias, dentre outros. A diversidade de plantas detectada por eDNA inclui 37 espécies de 33 gêneros e 19 famílias. A Figura 18 mostra o total relativo de espécies, gêneros e famílias

identificadas pelo método tradicional e pelo método de eDNA. Pôde-se observar que o DNA *metabarcoding* foi capaz de identificar um maior número de espécies e gêneros em relação ao método tradicional. Porém, o número de famílias foi menor, possivelmente devido ao fato de algumas famílias terem pouca representatividade de referências genéticas no ITVbiobase. Enquanto algumas famílias como Fabaceae e Poaceae tendem a não apresentar problemas de incompatibilidade de primers, sendo facilmente detectadas em estudos de eDNA (TIMPANO, SCHEIBLE e MEIKLEJOHN, 2020; GOUS et al., 2019), alguns estudos relatam variação de sucesso de amplificação e sequenciamento para espécimes de Asteraceae (ALSOS et al., 2018; PARDUCCI et al., 2013; BOESSENKOOL et al., 2014; HIIESALU et al., 2011). Ainda, sabe-se que nas cangas algumas espécies de plantas são difíceis de obter uma quantidade satisfatória de tecido para extrair DNA (VASCONCELOS et al., 2021), bem como, alguns táxons são raros com ocorrência dispersa na vegetação (GIULIETTI et al., 2019), além de haver problemas metodológicos de amplificação e sequenciamento da região alvo para algumas famílias relativamente abundantes da região, como Melastomataceae e Myrtaceae, por exemplo (VASCONCELOS et al., 2021).

Uma das explicações para essa diferença de sucesso de detecção pode estar relacionada à alta porcentagem de táxons com incompatibilidade de bases para os *primers* utilizados (SØNSTENØ et al., 2010). Para evitar qualquer viés metodológico relacionado à variação de reações de PCR, sugere-se que para aumentar a representatividade de famílias utilizando a técnica de DNA *metabarcoding*, a solução seria empregar primers específicos para famílias reconhecidamente problemáticas (KRESS, 2017; LIMA et al., 2018; VASCONCELOS et al., 2021). Para amostras com DNA altamente degradado, como frequentemente é o caso de eDNA, a resolução taxonômica em estudos de DNA *metabarcoding* pode ser potencialmente aumentada com uma combinação de vários marcadores moleculares, ou abordagens de sequenciamento genômico completo de baixa cobertura (COISSAC et al., 2016; PARDUCCI et al., 2017; LAMB et al., 2016).

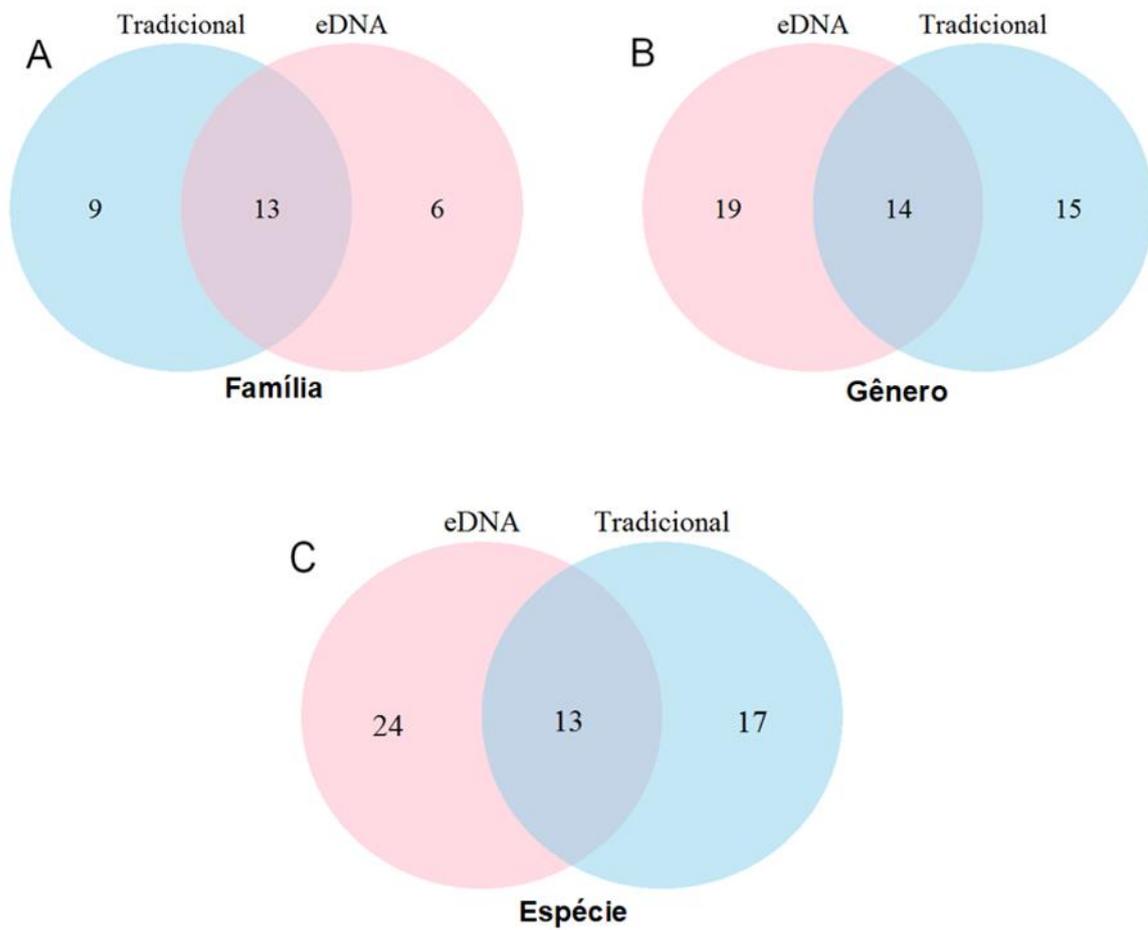
**Figura 18.** Número de espécies, gêneros e famílias detectadas pelo método tradicional (verde e azul) e pelo método de DNA ambiental (amarelo). Amplo Total refere-se a espécies, gêneros e famílias identificados em todos os plots da S11D indicados na Figura 8, e Amplo Plots são espécies, gêneros e famílias identificados somente nos plots analisados por eDNA, conforme indicado na Tabela 2.



**Fonte:** Autora, 2021.

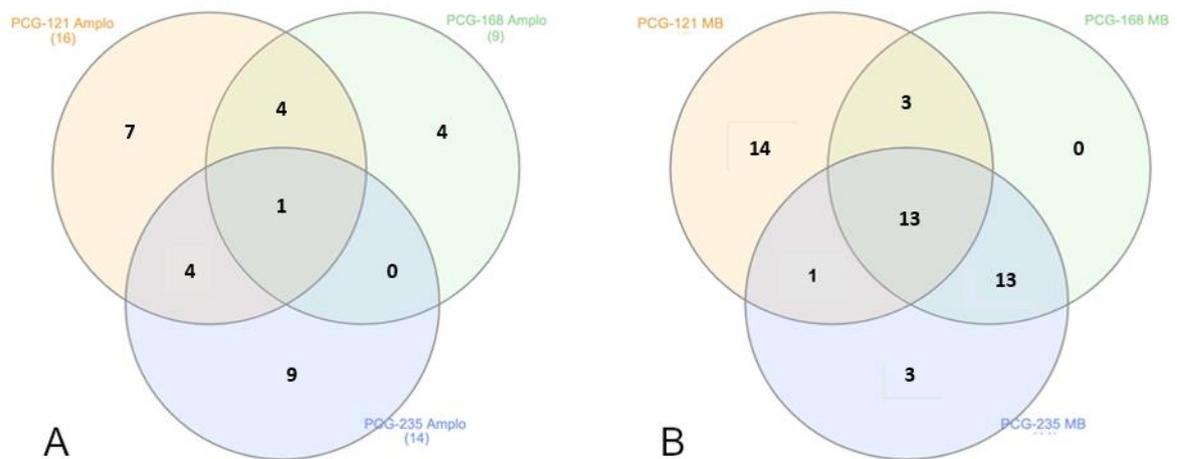
O levantamento taxonômico tradicional e a abordagem de DNA *metabarcoding* apresentaram alguma concordância entre as identificações de famílias, gêneros e espécies (Figura 19), havendo uma pequena sobreposição entre as espécies encontradas por parcela (Figura 21). No total, 67 espécies de plantas foram detectadas considerando ambos os métodos, ou seja, 30 espécies pelo tradicional e 37 espécies por eDNA, (Figura 19), e 13 espécies foram detectadas por ambos (23,6%). Ambos os métodos detectaram gêneros (14) comuns em proporções mais altas do que as detecções em nível de família (13) e espécie (13) (Figura 19). Das 30 espécies de plantas detectadas apenas pelo método tradicional, 19 espécies (63%) foram encontradas no banco de referências ITVBiobase.

**Figura 19:** Diagrama de Venn indicando a sobreposição das famílias, gêneros e espécies observadas nos três diferentes plots



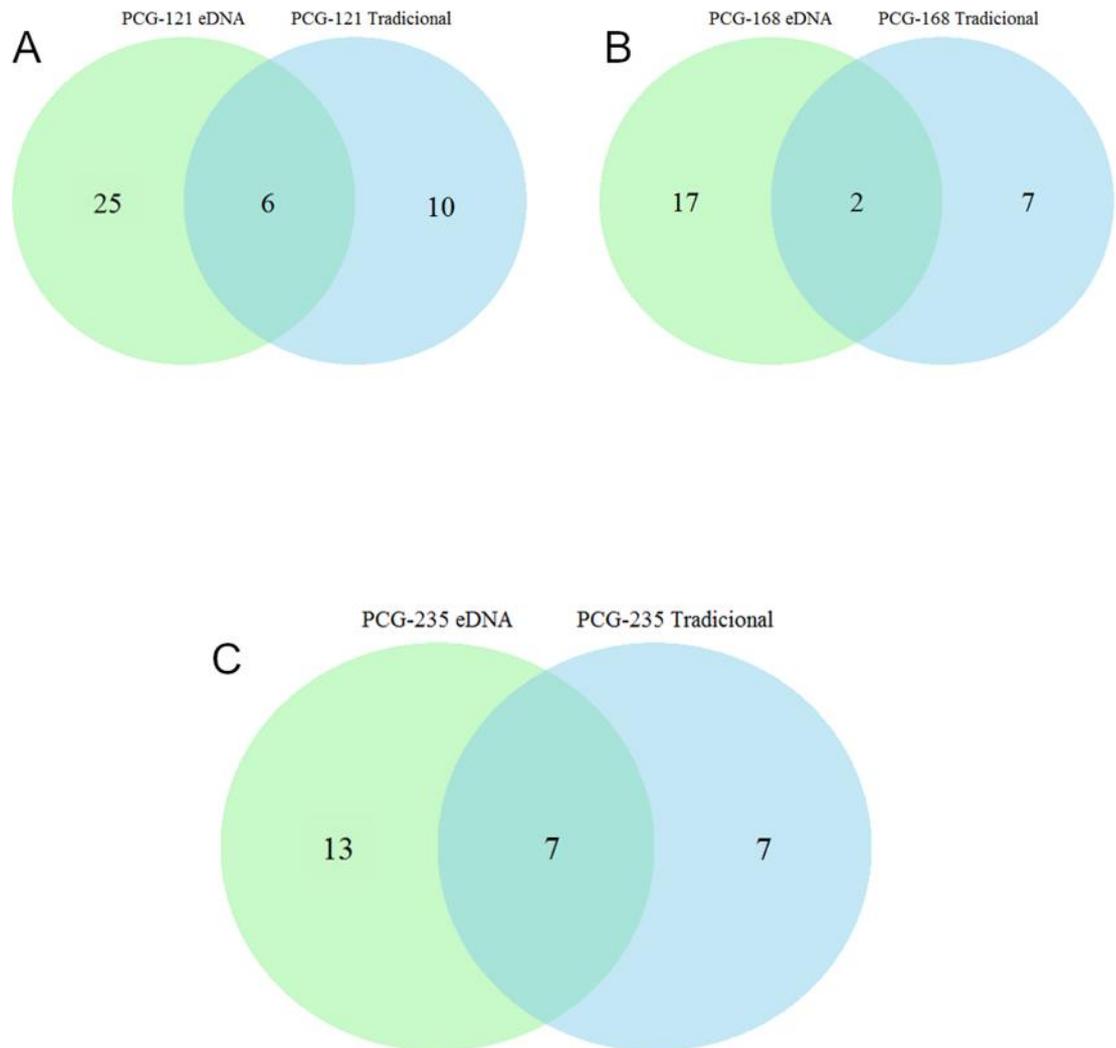
Fonte: Autora, 2021

**Figura 20:** Diagrama de Venn indicando a sobreposição das espécies observadas nos diferentes plots pela abordagem tradicional (A) e pela abordagem de DNA *metabarcoding* (B).



Fonte: Autora, 2021

**Figura 21:** Diagramas de Venn indicando o número de espécies em comum encontradas pela identificação morfológica tradicional e pela abordagem de DNA metabarcoding por parcela nos plots PCG-121 (A), PCG-168 (B) e PCG-235 (C).



**Fonte:** Autora, 2021.

Algumas famílias como Fabaceae, Poaceae, Convolvulaceae, Myrtaceae e Euphorbiaceae foram observadas como dominantes em todos os plots, apresentando proporções variáveis (Figura 22), assim como observado nos dados de frequência de ocorrência (Apêndice G). Fabaceae foi a família com maior riqueza de espécies, sendo representada por sete gêneros e oito espécies. Asteraceae, Poaceae e Rubiaceae apresentaram a segunda maior riqueza, sendo três gêneros e três espécies para cada. Em seguida, Myrtaceae aparece com a terceira maior riqueza, com dois gêneros e três espécies. Para mais da metade das famílias (14), foram identificadas apenas uma ou duas espécies (Tabela 10). Na identificação morfológica tradicional, representantes de Fabaceae foram observados em todos os plots, sendo representada por três espécies, assim como Malpighiaceae (Apêndice A- Tabela 2). Poaceae, Vochysiaceae, Araceae e Myrtaceae vieram em seguida, com duas espécies cada. Apesar de serem de uma mesma fitofisionomia (canga arbustiva), na Figura 21B, pode-se observar a variação na riqueza dos plots e na representatividade de cada família identificada por meio do eDNA, mostrando uma clara diferença entre eles, tal como reportado no método convencional. É importante frisar que o trabalho dos taxonomistas em campo geralmente apresenta algumas limitações. A identificação de plantas é feita com base em características morfológicas externas ou internas, um espécime de cada vez, e exige várias visitas para coletar a maioria das espécies na fase fértil. Ainda, a maioria das plantas apresentam ciclos de vida sazonais, que duram apenas alguns meses, e depois desaparecem completamente do ambiente, o que dificulta o processo, além de casos de complicações na taxonomia de famílias próximas e morfológicamente parecidas ou espécies crípticas (ZAPPI, 2017), o que pode ocasionar erros.

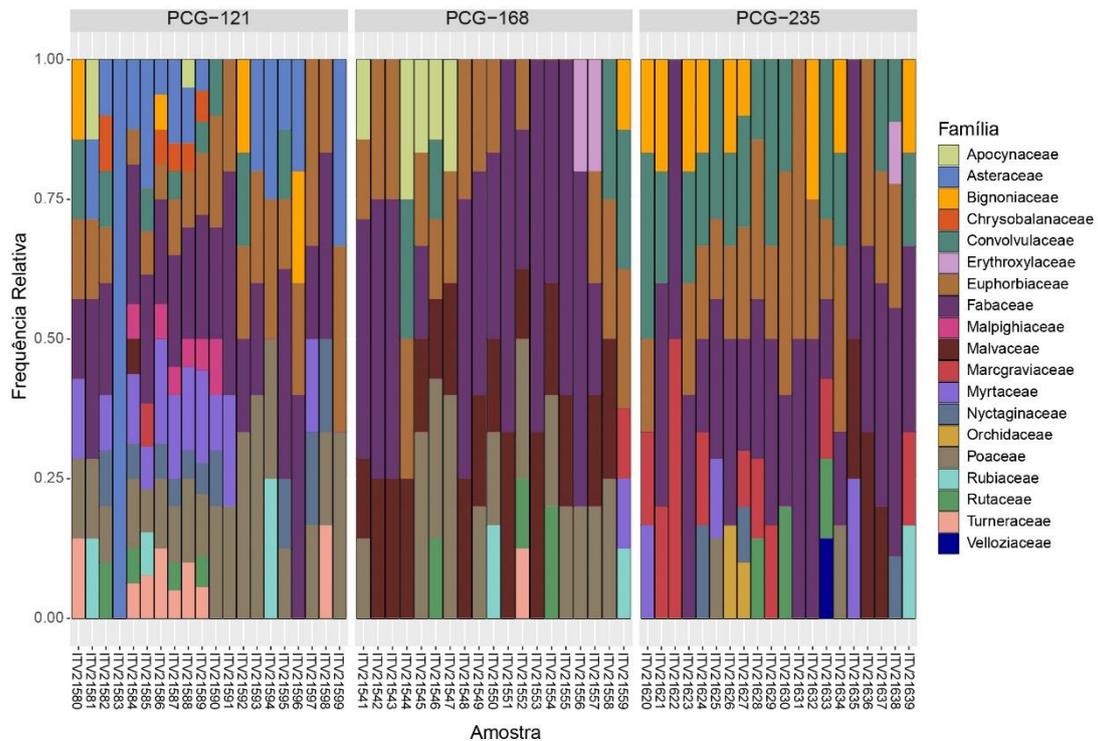
A distribuição das famílias foi bastante distinta entre os plots (Figura 22A), com Poaceae sendo observada com maior frequência no plot PCG-121, e Marcgraviaceae mais representada no plot PCG-235, por exemplo (Figura 22B). Adicionalmente, espécimes de Erythroxylaceae, de ocorrência mais limitada, foram identificados no plot PCG-235 (Figura 22B), confirmando o levantamento florístico tradicional para o mesmo plot (Apêndice A -Tabela 2). Entretanto, foi observada uma divergência para a detecção da família Asteraceae, onde a identificação baseada em morfologia a detectou no plot PCG -168 (Apêndice A -Tabela 2) e o DNA *metabarcoding* no plot PCG-121 (Figura 22B). Outro resultado interessante obtido por eDNA incluiu a

detecção de famílias que são da área de canga e S11D, mas que não foram encontradas pela amostragem tradicional nas áreas selecionadas, como Apocynaceae (PCG-121 e PCG-235), Malvaceae (todos os plots), Nyctaginaceae (PCG-121 e PCG-235) e Turneraceae (PCG-121 e PCG-168). A presença de algumas famílias de plantas ameaçadas de Carajás também pôde ser observada, destacando-se, Erythroxylaceae presentes nas amostras ITV21557, ITV21558 e ITV21638. Destaca-se também a identificação de algumas famílias em poucas amostras, como no caso da espécie de Orchidaceae encontrada apenas nas amostras ITV21626 e ITV21627, e Velloziaceae, observada apenas na amostra ITV21634. Adicionalmente, ambos os métodos de identificação detectaram espécimes das famílias Asteraceae, Bignoniaceae, Convolvulaceae, Erythroxylaceae, Euphorbiaceae, Fabaceae, Malpighiaceae, Marcgraviaceae, Myrtaceae, Orchidaceae, Poaceae, Rubiaceae e Velloziaceae. Ainda, as diferenças observadas dentro de um mesmo plot indicam a possibilidade de detecção de variabilidade mesmo em pequenas distâncias geográficas (Figura 22), uma vez que as comunidades são dinâmicas e podem apresentar bolsões, evidenciando microhabitats que abrigam plantas próprias (GASTAUER et al., 2021).

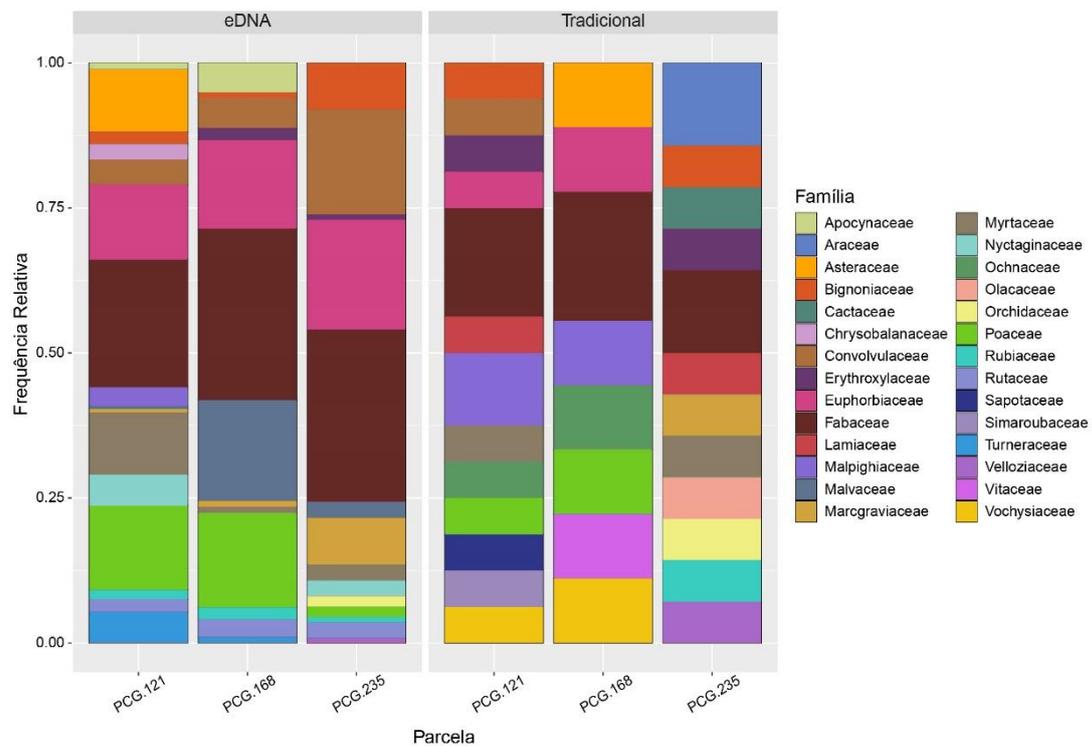
Com relação aos gêneros, pôde-se observar que *Alchornea*, *Dioclea*, *Neea* e *Tunera* estavam mais representados em PCG-121, enquanto *Anemopagegma*, *Bauhinia*, *Croton*, *Ipomoea* e *Norantea* foram observados mais vezes em PCG-235, e *Abrus*, *Chamaecrista*, *Melochia* e *Paspalum* com distribuição bastante distinta, sendo mais abundantes em PCG-168 (Figura 23B). Também se destaca a identificação de alguns gêneros em poucas amostras, como *Erythroxylum* (ITV21556, ITV21557, ITV21638), *Sobralia* (ITV21625, ITV21626) e *Vellozia* (ITV21633) (Figura 23A).

**Figura 22.** Frequência por família amostrada em cada um dos três plots estudados (PCG-121, PCG-168 e PCG-235), indicando as frequências por amostra (A) e por parcela (B).

**A**



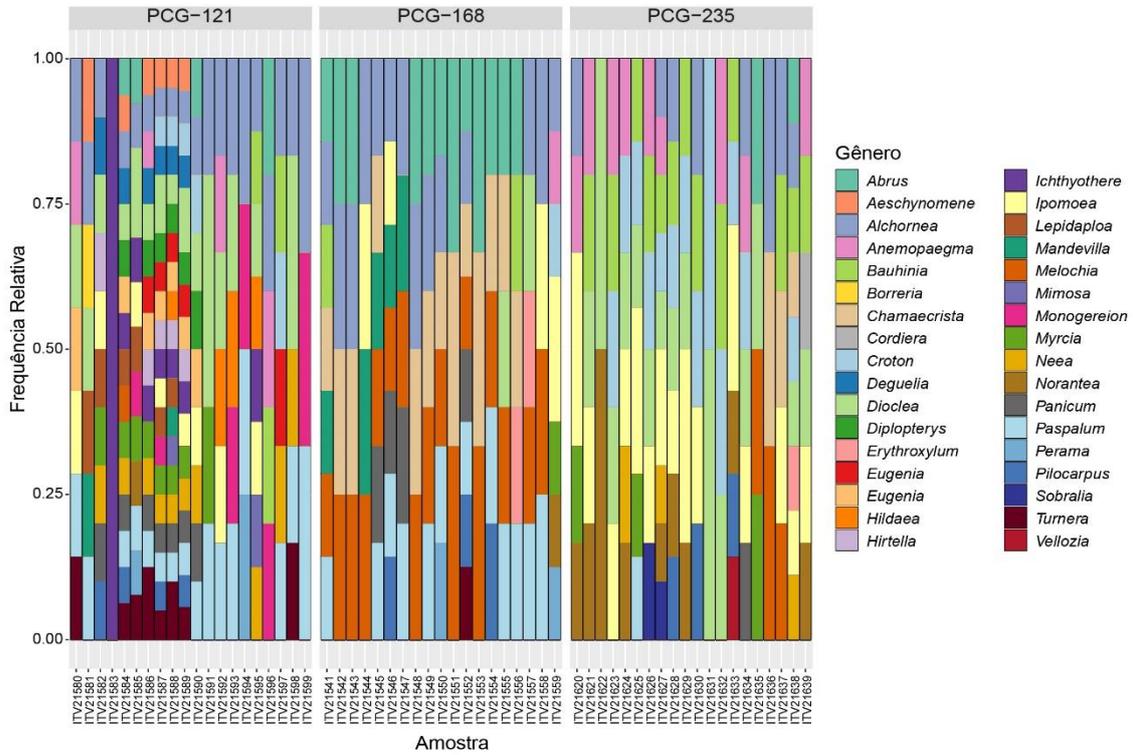
**B**



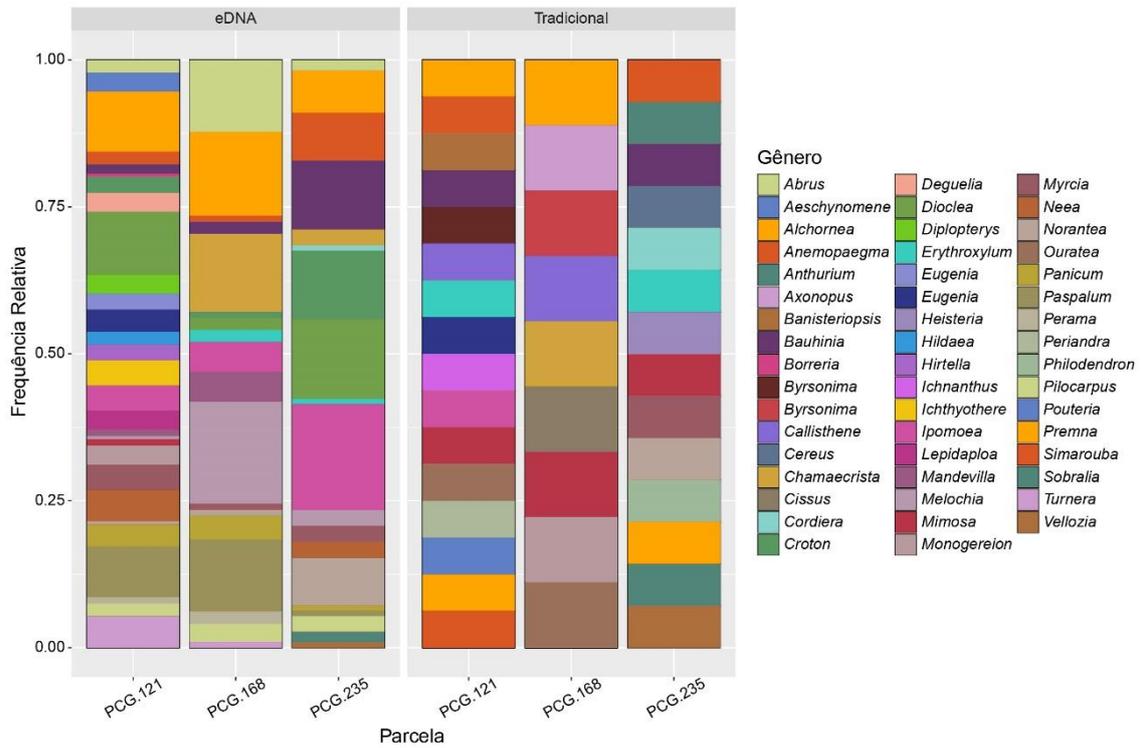
Fonte: Autora, 2021.

**Figura 23.** Frequência por gênero amostrada em cada um dos três plots estudados (PCG-121, PCG-168 e PCG-235), indicando as frequências por amostra (A) e por parcela (B).

**A**



**B**



Fonte: Autora, 2021.

Em termos de diversidade de espécies, é interessante apontar que a parcela PCG-121 claramente apresentou uma diversidade superior às parcelas PCG-168 e PCG-235 (Figura 24), corroborando com o perfil taxonômico do inventário florístico (Apêndice A- Tabela 2). Curiosamente, a parcela de canga arbustiva PCG-121 apresenta uma proximidade com à beira da estrada e com áreas de capão florestal e floresta, o que pode ter influenciado na maior heterogeneidade das espécies encontradas, já que existe a possibilidade de ocorrer transporte de matéria orgânica e traços de DNA de espécies de plantas do capão florestal/floresta para canga por lixiviação. É interessante notar o efeito da conectividade que às áreas com matas bem estruturadas, acúmulo de matéria orgânica, disponibilidade de nutrientes e maior riqueza de plantas podem ter com as porções de canga e vice-versa. Estudos indicam que diferenças ambientais entre as fisionomias, a disponibilidade de habitat na paisagem e a heterogeneidade da paisagem influenciam concomitantemente na diversidade de plantas da canga amazônica (GASTAUER et al., 2021). Observa-se também que, dentro e entre os plots amostrados, a diversidade de táxons foi relativamente alta, mesmo com uma cobertura de menos de um terço (29,31%) das espécies de canga com códigos de barras ITS2 (VASCONCELOS et al., 2021). Segundo Gastauer et al. (2021), as fisionomias possuem heterogeneidade abiótica e podem apresentar filtros ambientais que segregam a comunidade de plantas da canga em pools de espécies específicas, causando diferenças composicionais.

A abordagem baseada em DNA *metabarcoding* conseguiu identificar 13 espécies em comum às listadas no inventário florístico das mesmas áreas de estudo: *Alchornea discolor* (Euphorbiaceae), *Anemopaegma carajasense* (Bignoniaceae), *Bauhinia pulchella* (Fabaceae), *Chamaecrista desvauxii* (Fabaceae), *Cordia myrciifolia* (Rubiaceae), *Erythroxylum nelson-rosae* (Erythroxylaceae), *Eugenia flavescens* (Myrtaceae), *Ipomoea marabaensis* (Convolvulaceae), *Mimosa acutistipula* var. *ferrea* (Fabaceae), *Monogereion carajensis* (Asteraceae), *Norantea guianensis* (Marcgraviaceae), *Sobralia liliastrum* (Orchidaceae) e *Vellozia glauca* (Velloziaceae). Dentre essas, *Monogereion carajensis* e *Erythroxylum nelson-rosae* são espécies enquadradas como ameaçadas de extinção nas listas da IUCN, do MMA (Portaria 443/2014) e da SEMA/PA (Resolução 054/2007). Vale ressaltar que a alta resolução taxonômica obtida só foi possível devido à disponibilidade de uma biblioteca de

referências genéticas previamente disponibilizada para as espécies de plantas dentro da área de estudo em questão (VASCONCELOS et al., 2021). Quando o banco de dados do NCBI foi utilizado para determinação taxonômica e comparação com a base de dados do ITVBiobase, poucos táxons disponíveis eram compatíveis com a lista de plantas registradas para as áreas de cangas de Carajás. Uma das principais críticas para estudos de detecção indireta de plantas está relacionada à falta de sequências de referência disponíveis em bancos de dados de referência, como o GenBank (HOLLINGSWORTH, 2011). Além disso, muitas vezes pode haver pouca confiabilidade em dados disponíveis que não foram derivados de espécimes bem curados (DORMONTT et al., 2018). Ainda existe pouca informação sobre os níveis de certeza e comprovação de identificações taxonômicas realizadas em relação a vouchers dos espécimes amostrados e qualidade de dados brutos de sequenciamento, podendo haver riscos de atribuições taxonômicas incorretas (SMITH et al., 2016; BALAKIREV et al., 2017). O banco de dados do BOLDSYSTEMS (RATNASINGHAM e HEBERT, 2007) conta atualmente com sequências de DNA para aproximadamente 20% das plantas terrestres descritas. Em comparação ao GenBank, o BOLDSYSTEMS exige a disponibilização de metadados mais completos e espécimes devidamente verificados antes que os dados de sequência possam ser aceitos (DORMONTT et al., 2018). No entanto, devido a esses requisitos mais rígidos, o número de espécies incluídas nessa base de dados é muito menor do que o representado em outros repositórios menos rigorosos. Dessa forma, reforça-se que a geração e utilização de bases de dados bem curadas manualmente e direcionadas para as áreas de estudos se fazem essenciais para a correta aplicação das abordagens de levantamento de espécies por meio de DNA *metabarcoding*.

Outro ponto interessante é que apesar de *Mimosa acutistipula* var. *ferrea* ter sido identificada em todos os plots pelo método convencional (Apêndice A - Tabela 2), a espécie foi detectada por eDNA apenas em PCG-121 (Figura 24B). Os dados de frequência de identificação das espécies para eDNA (Apêndice H), sugerem que a maioria das espécies não são exclusivas de uma área, ficando claro que algumas são comuns a todos os pontos enquanto outras são específicas. Dessa forma, pode-se ressaltar a necessidade de ampliação das abordagens de levantamentos florísticos e da aplicação de métodos de eDNA para entender essa heterogeneidade. Tanto a variação de PCR versus agregação espacial de eDNA quanto a impossibilidade de

utilização de dados de frequência de espécies com os dados do levantamento tradicional, com espécies listadas por plot e não por ponto amostral, são fatores a serem considerados.

Onze espécies previamente listadas e não detectadas por DNA *metabarcoding*, não estavam presentes no banco de referências de ITS2: *Anthurium lindmanianum* (Araceae), *Axonopus purpusii* (Poaceae), *Byrsonima spicata* (Malpighiaceae), *Byrsonima crassifolia* (Malpighiaceae), *Callisthene microphylla* (Vochysiaceae), *Cereus jamacaru* (Cactaceae), *Ichnanthus* sp. (Poaceae), *Myrcia cuprea* (Myrtaceae), *Periandra mediterranea* var. *mediterrânea* (Fabaceae), *Philodendron quinquenervium* (Araceae) e *Premna serratifolia* (Lamiaceae). Tais resultados reforçam a importância de continuar a expandir o banco de dados e ter disponível as sequências de DNA *barcodes* nos repositórios de referência, para um inventário eficaz da diversidade de plantas (LIMA et al., 2018; VASCONCELOS et al., 2021). Adicionalmente, seis espécies do levantamento florístico com DNA *barcodes* disponíveis não foram detectadas por eDNA: *Banisteriopsis stellaris* (Malpighiaceae), *Cissus erosa* (Vitaceae), *Heisteria ovata* (Olacaceae), *Ouratea castaneifolia* (Ochnaceae), *Pouteria ramiflora* (Sapotaceae) e *Simarouba amara* (Simaroubaceae) (Figura 24B; Apêndice A - Tabela 2).

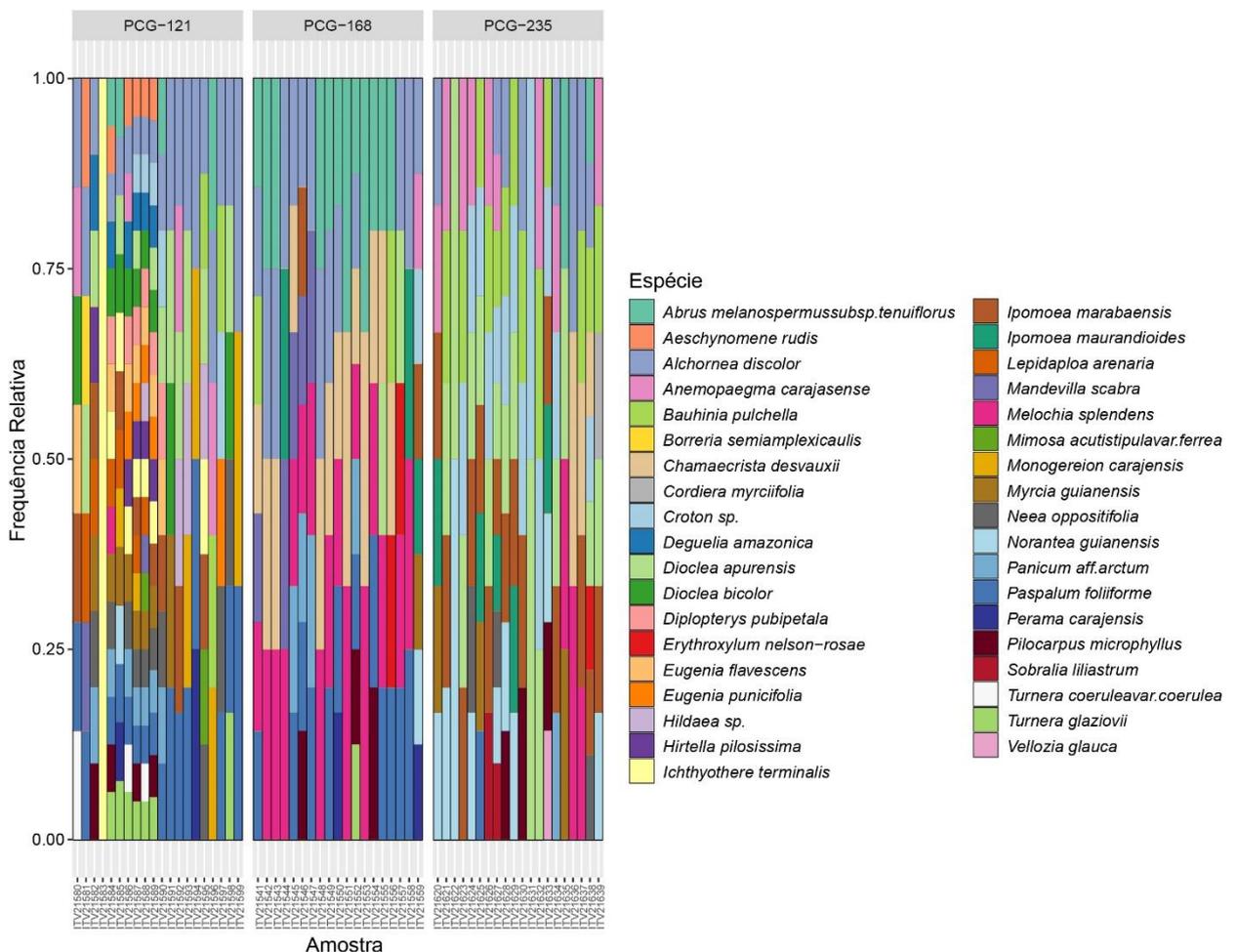
Um dos problemas mais comuns de usar *primers* universais em uma amostragem taxonômica muito diversa em termos de famílias e espécies estreitamente relacionadas é a possível necessidade de utilização de mais de um marcador molecular, quando nem todos os táxons de plantas podem ser detectados com uma única região genômica (FAHNER et al., 2016; HOLLINGSWORTH, GRAHAM e LITTLE, 2011; CBOL et al., 2009). Vários autores vêm relatando dificuldades na utilização de DNA *barcodes* baseados em ITS de espécies tropicais (GONZALEZ et al., 2009; BOLSON et al., 2015). Por exemplo, Gonzalez et al. (2009), obtiveram resultados de sequenciamento problemáticos para ITS, com apenas 41% das amostras de árvores da Amazônia sendo codificadas com sucesso pelos autores. Segundo Moonhouse-Gann et al. (2018) tal problema pode ser associado a diversos fatores, como baixa taxa de sequenciamento do locus, múltiplas cópias de parálogos em um genoma individual e pseudogenes, além da alta probabilidade de contaminação fúngica endofítica, dada a semelhança entre os sítios de ancoragem dos *primers* universais de plantas e fungos (CHENG et al., 2016; HOLLINGSWORTH,

2011). Em amostras ambientais, como solo, os fungos são abundantes e desempenham um papel central na decomposição (TIMPANO, SCHEIBLE e MEIKLEJOHN, 2020).

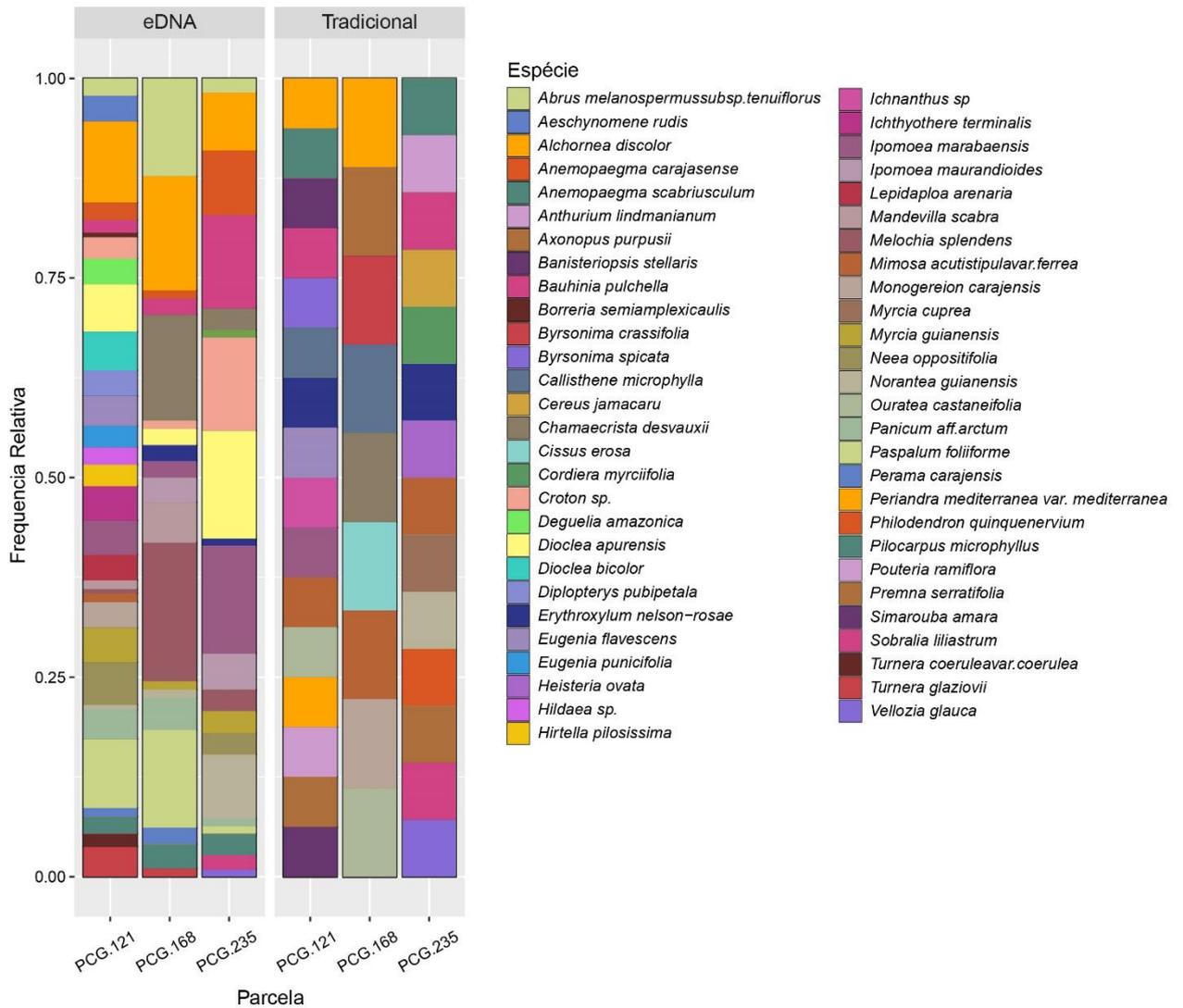
Além disso, dependendo do grupo de plantas, à exemplo da família Araceae, em que há porções ricas em repetições curtas de DNA na região (VASCONCELOS et al., 2018), o uso apenas da região nuclear do ITS como DNA *barcoding* pode não ser suficiente para detectar espécies por ferramentas de eDNA, havendo a necessidade da inclusão de outros marcadores moleculares (BOLSON et al., 2015). Por outro lado, a região ITS2 menor tem sido indicada como uma das melhores regiões para o código de barras de plantas, apresentando um alto índice de sucesso de sequenciamento para amostras de DNA de qualidade inferior (CHEN et al., 2010; KUZMINA et al., 2012; RAMALHO et al., 2018).

**Figura 24:** Frequência por espécie amostrada em cada um dos três plots estudados. As frequências por amostra e por parcela estão representadas em A e B, respectivamente.

**A**



## B



Fonte: Autora, 2021.

Também é importante indicar que algumas espécies listadas no levantamento florístico (Apêndice A - Tabela 2) foram identificadas por DNA *metabarcoding* em localidades diferentes do que foi originalmente indicado, como no caso de *Monogereion carajensis* e *Ipomoea marabaensis*, por exemplo. Outro resultado interessante é que o método morfológico tinha identificado a espécie *Anemopaegma scabriusculum* na parcela PCG-121 e PCG-235 (Figura 24B), porém estudos recentes identificaram, descreveram e ilustraram uma espécie nova, *Anemopaegma carajasense*, para as cangas de Carajás. Até então *Anemopaegma carajasense* era identificada como *Anemopaegma scabriusculum*. *Anemopaegma carajasense* foi identificada por DNA *metabarcoding* (Figura 24B), e é uma espécie endêmica da Serra

dos Carajás, ocorrendo na S11D e em áreas de canga, e assemelha-se a duas outras espécies do gênero, *Anemopaegma goyazense* e *Anemopaegma scabriusculum*. Entretanto, *A. carajasense* possui diferenças em vários caracteres vegetativos, apresentando a morfologia floral do tipo “*Anemopaegma*” (LOHMANN, FIRETTI E GOMES, 2018). Isto evidencia a importância do uso de ferramentas moleculares no monitoramento ambiental, uma vez que, o relatório florístico enviado pelos consultores data de 2016, e as comunidades florísticas são dinâmicas, logo existe a iminência de espécies novas por vir entre esse período de tempo. Sem dúvidas, o uso da tecnologia molecular será um grande avanço para os taxonomistas, principalmente, pelo fato que muitas vezes as determinações taxonômicas são realizadas por análises comparativas, e podem ser tendenciosas ou até mesmo gerar níveis de incertezas, por isso as informações contidas no DNA serão mais precisas e valiosas.

Ainda comparando os métodos morfológico e molecular, espécies diferentes de um mesmo gênero puderam ser detectadas em alguns casos, como *Myrcia guianensis* e *Myrcia cuprea*. (Figura 24). *Myrcia guianensis* é uma espécie que ocorre nas áreas de canga da S11A, com floração praticamente o ano todo (TRINDADE, ROSÁRIO E SANTOS, 2018) e foi identificada nas amostras de eDNA, em vez de *Myrcia cuprea* (Apêndice A - Tabela 2). Contudo, *Myrcia cuprea* não está presente no banco de referências de ITS2 e possivelmente, essa detecção de espécies próximas em um mesmo gênero ocorreu devido problemas de resolução filogenética causadas pela alta similaridade entre sequências de táxons proximamente relacionados (VASCONCELOS et al., 2021).

Entretanto, também não se pode descartar uma contaminação da amostra de solo com DNA de plantas provenientes de outras áreas, como por transporte de pólen e/ou sementes por água ou vento, por exemplo (SILVA et al., 2018). Além disso, a visualização de partes de tecidos vegetais presentes no local no momento da coleta das amostras de solo é geralmente difícil, aumentando a probabilidade de contaminação do solo com DNA de espécies de outras áreas. KINGSLEY e CORLETT (2020) demonstram que o poder do DNA *metabarcoding* está relacionado a quanto tempo o DNA permanece no local sem perda considerável por degradação, considerando que fatores como temperaturas altas e pH do solo podem impactar enormemente na qualidade do DNA e, conseqüentemente, na identificação de uma espécie em uma dada área. Fragmentos de DNA mais longos provavelmente

representam o DNA mais recente e permanecem detectáveis por dias após a presença da espécie (TABERLET et al. 2012). Assim, o comprimento do fragmento amplificado pode alterar a resolução temporal da composição de espécies visualizadas por meio de amostras de solo. Destaca-se a detecção pelo eDNA de espécies que são da área de canga e S11D, mas que não foram encontrados pela amostragem tradicional em todos os plots de canga, como: *Aeschynomene rudis* (Fabaceae), *Ichthyothere terminalis* (Asteraceae), *Ipomoea maurandioides* (Convolvulaceae), *Lepidaploa arenaria* (Asteraceae), *Mandevilla scabra* (Apocynaceae), *Melochia splendens* (Malvaceae) e *Neea oppositifolia* (Nyctaginaceae).

A resolução taxonômica, ou seja, a taxa de eDNA a ser identificada em nível de espécie, depende do nível de representatividade da flora local na biblioteca de referência disponível (ALSOS et al., 2018; KOLTER e GEMEINHOLZER, 2020), levando em conta as características específicas da região, como a complexidade do tipo de vegetação (BIRKS et al., 2016), bem como, a variação morfológica e/ou genética exibida por diferentes grupos taxonômicos (ALSOS et al., 2018). Considerando que apenas onze espécies do presente estudo que foram listadas em levantamentos taxonômicos prévios estavam ausentes no banco de sequências de DNA de referência do local, o DNA *metabarcoding* pode ser considerado como bastante promissor para ser utilizado no monitoramento ambiental das áreas de canga da Serra dos Carajás.

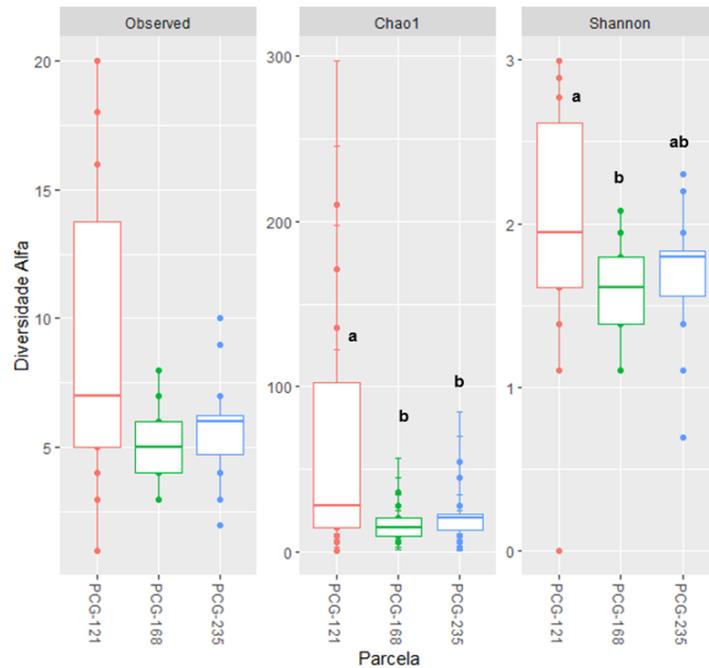
Adicionalmente, o entendimento da distribuição das espécies e os processos que moldam a variação espacial e a composição da comunidade são cruciais para a aplicação de estratégias de monitoramento ambiental e garantir a conservação da biodiversidade, especialmente para espécies endêmicas e ameaçadas. Tais ações também requerem métodos que permitam a detecção rápida e robusta da biodiversidade em diferentes escalas espaciais (ZINGER et al., 2019; BUSH et al., 2020). Dessa forma, foi demonstrado que a robustez do método de DNA *metabarcoding* para monitoramento de espécies de plantas na Serra dos Carajás é fortemente influenciada pela otimização dos parâmetros escolhidos. Além disso, aqui foi demonstrado que um *benchmarking* robusto de DNA *metabarcoding* para a identificação de espécies de plantas da canga depende de diversos fatores metodológicos, como a disponibilidade de um banco abrangente de referências genéticas, *primers* e condições de amplificação apropriados (KRESS, 2017; ZEPEDA

MENDOZA, SICHERITZ-PONTÉN e GILBERT, 2015; COMPSON et al., 2020; KINGSLEY e CORLETT, 2020).

#### **5.2.4 Índices de diversidade alfa e beta**

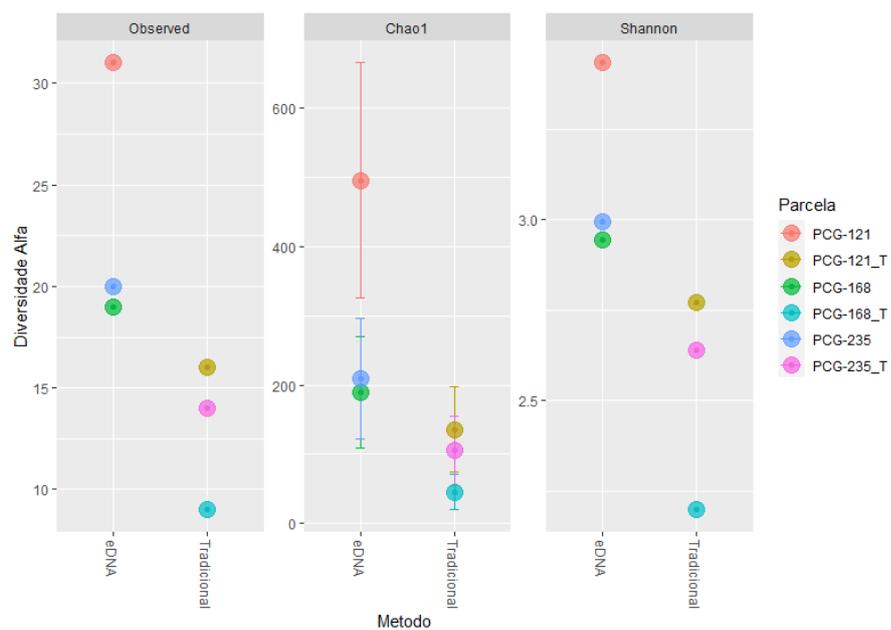
O cálculo do índice de diversidade alfa indicou claramente que PCG-121 possui níveis de riqueza (Chao1, ANOVA,  $F = 7,977$ ;  $p\text{-value} = 0,000894$ ) e diversidade (Shannon, ANOVA,  $F = 3,786$ ;  $p\text{-value} = 0,0287$ ) de OTUs significativamente maiores dentre as parcelas amostradas (Figura 25), em concordância com os o indicado pelo método de identificação morfológica tradicional (Kruskal-Wallis,  $p\text{-value} = 0,049$ , Figura 26). É importante retomar que o fato desta parcela ter uma proximidade maior com áreas de capão florestal e floresta, pode ter uma relação com os resultados encontrados. Para o índice de Shannon, embora não houve uma diferença significativa nos valores de diversidade entre PCG-235 e PCG-168 e entre PCG-235 e PCG-121, foi observada uma diferença significativa entre PCG-121 e PCG-168 (Figura 25). Entre amostras, a riqueza e a diversidade também variaram, apesar de serem bastante homogêneas no total (Kruskal-Wallis,  $p\text{-value} = 0,4$ ) (Figura 27) (Apêndice I).

**Figura 25:** Análise de diversidade Alfa baseada nos índices de riqueza (Chao1) e diversidade de espécies (Shannon), considerando presença/ausência nas diferentes parcelas localizadas na S11D. Letras diferentes no mesmo boxplot indicam uma diferença significativa ( $p < 0,05$ ), de acordo com o teste post-hoc de Tukey HSD.



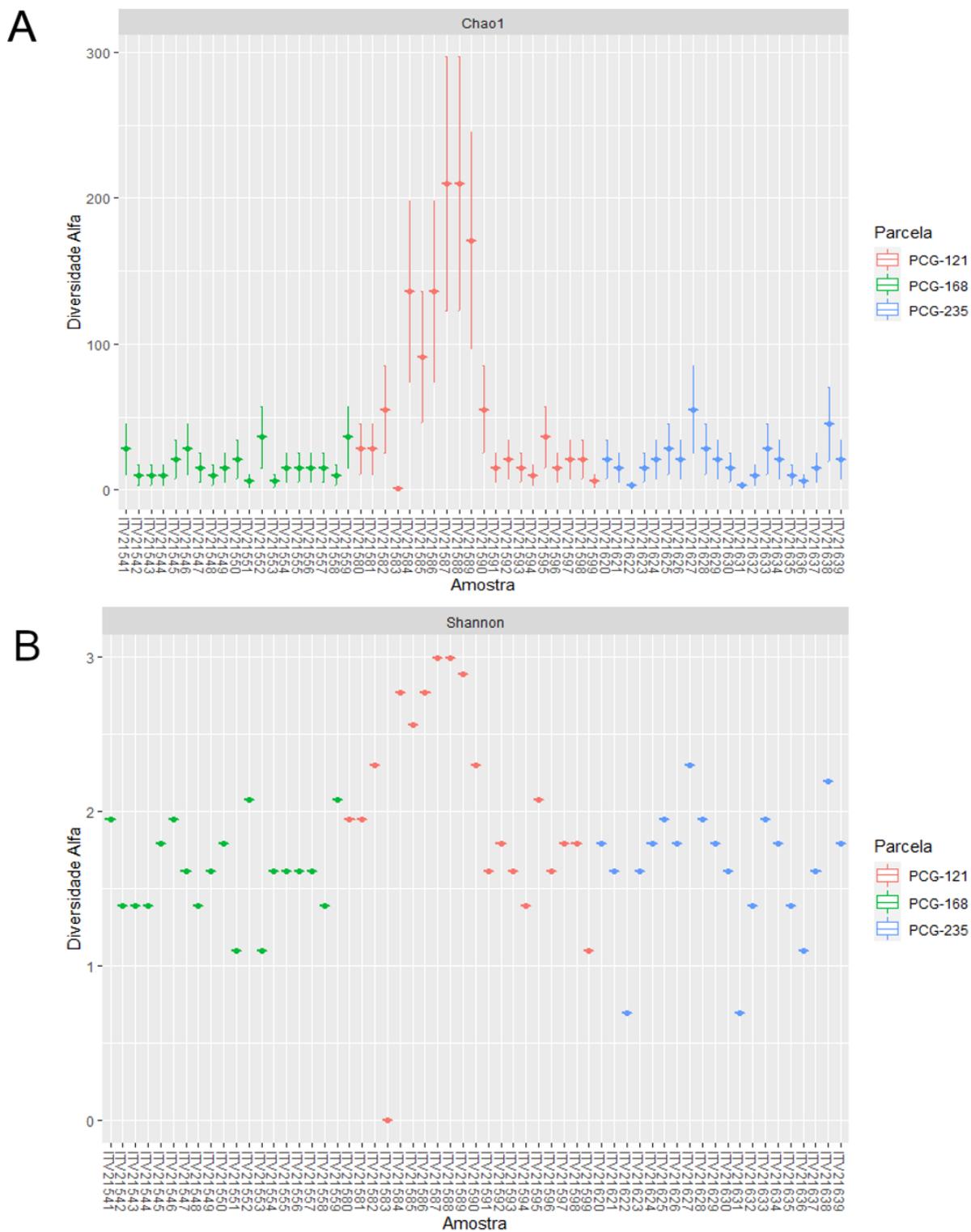
Fonte: Autora, 2021.

**Figura 26:** Análise de diversidade Alfa baseada nos índices de riqueza (A) (Chao1) e diversidade de espécies (B) (Shannon), considerando o método tradicional e eDNA em diferentes parcelas localizadas na S11D.



Fonte: Autora, 2021.

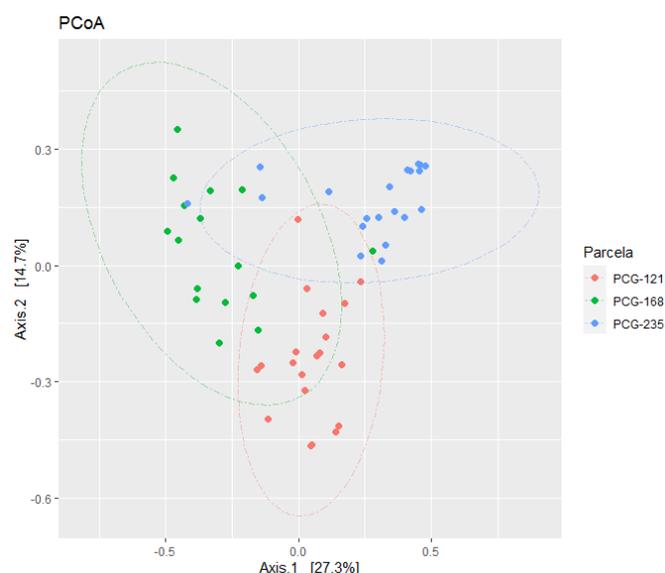
**Figura 27:** Análise de diversidade Alfa baseada nos índices de riqueza (A) (Chao1) e diversidade de espécies (B) (Shannon), considerando diferentes amostras localizadas na S11D.



Fonte: Autora, 2021.

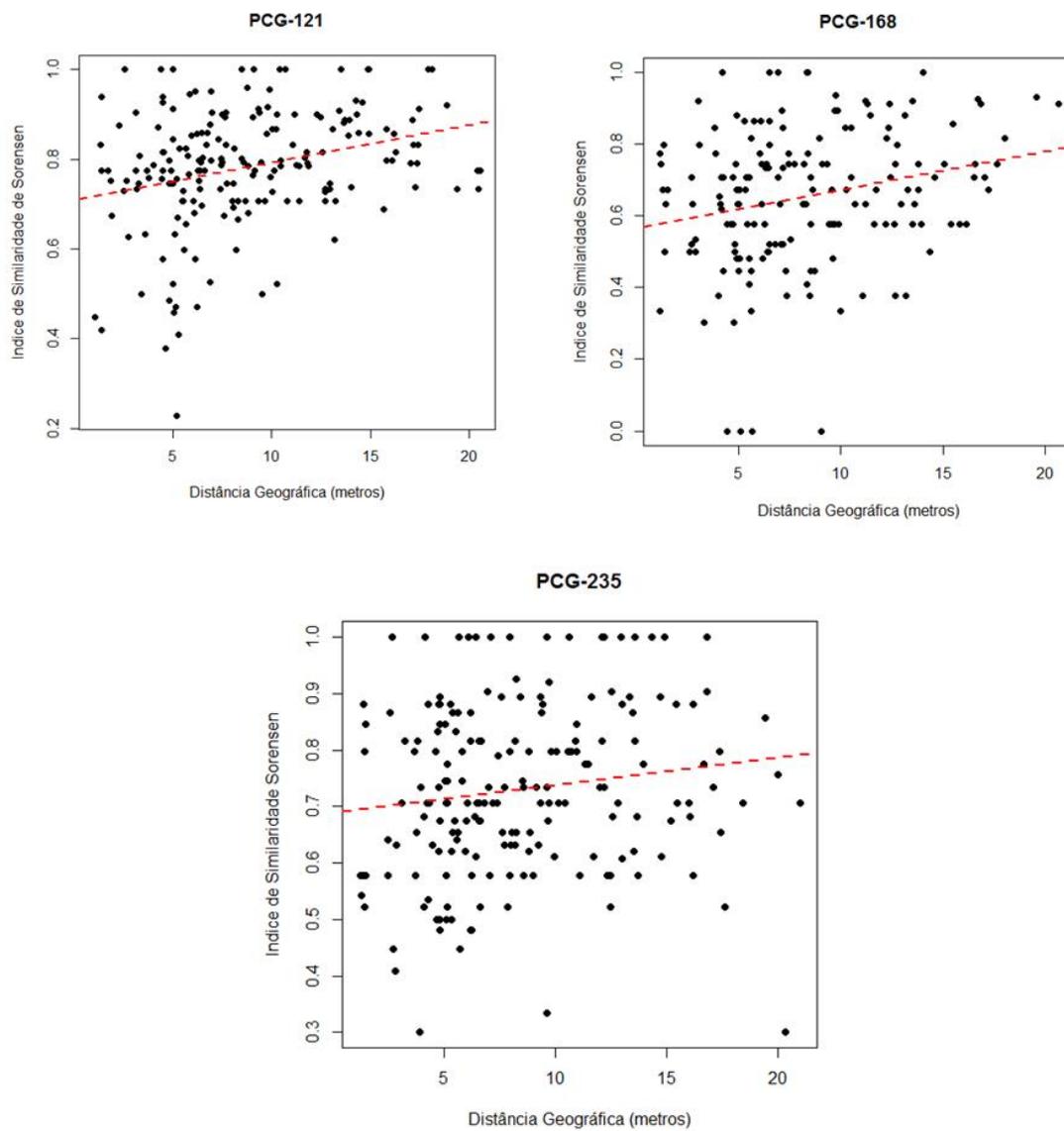
Algumas diferenças significativas foram observadas na composição taxonômica de plantas detectadas, conforme indicado pelos resultados de PERMANOVA ( $R^2 = 0,28$ ,  $p$ -value =  $1e-04$ ). Os padrões de diversidade beta para o conjunto de dados de DNA *metabarcoding* mostraram que as parcelas PCG-168 e PCG-121 são diferentes significativamente entre si (Tukey HSD,  $p = 0,03$ ). Mesmo sendo uma parcela localizada numa área geograficamente mais distante das demais, a análise apontou que não há diferenças significativas entre PCG-235 e as demais parcelas (Tukey HSD,  $p = 0,4$ ). É interessante notar que houve um maior compartilhamento de táxons entre as parcelas PCG-168 e PCG-235 do que com a PCG-121 (Figura 28). Adicionalmente, o teste de Mantel (Figura 29), indicou uma fraca correlação entre composição de espécies e distância geográfica dentro de cada parcela (PCG-121:  $r = 0,26$  e  $p < 0,005$ ; PCG-235:  $r = 0,17$ ;  $p < 0,04$ ; PCG-168:  $r = 0,22$ ;  $p < 0,009$ ). Já entre as parcelas houve correlação significativa ( $r = 0,4202$  e  $p < 0,005$ ), indicando que a similaridade de espécies varia com a distância geográfica (Figura 30). Em relação ao desenho amostral, há necessidade de mais estudos para que compreendamos precisamente quais os eventos que determinam o que torna a comunidade de plantas mais parecidas ou não nos pontos de coleta nas cangas da Serra dos Carajás ao serem identificadas por DNA *metabarcoding*.

**Figura 28:** Análise das Coordenadas Principais (PCoA) da diversidade beta dos locais de amostragem (distância de Jaccard). As elipses representam 95% dos indivíduos dos grupos indicados na análise.



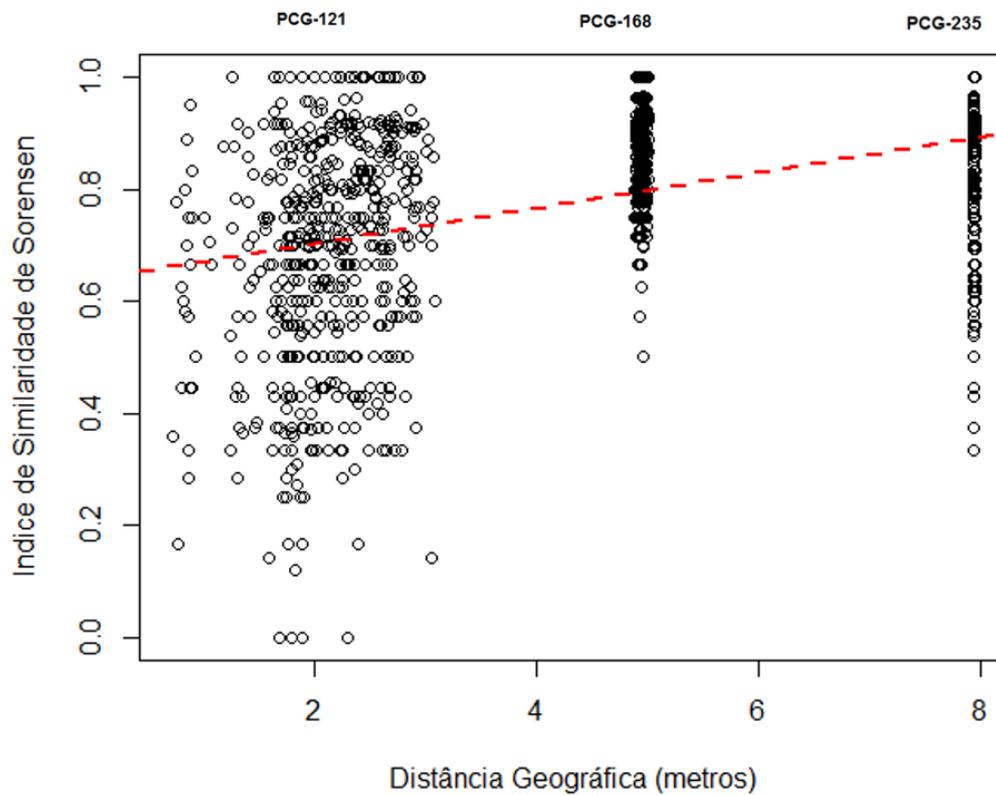
**Fonte:** Autora, 2021

**Figura 29:** Distâncias geográficas e sua relação com a composição (Índice de Sorensen) das comunidades de plantas dentro dos três plots de canga avaliados.



Fonte: Autora, 2021.

**Figura 30:** Distâncias geográficas e sua relação com a composição (Índice de Sorensen) das comunidades de plantas entre os três plots de canga avaliados. A distância geográfica foi logaritimizada.



Fonte: Autora, 2021.

### 5.3 RECOMENDAÇÕES PARA FUTURAS PESQUISAS

- Combinação ou uso de outros marcadores moleculares capazes de detectar mais táxons de plantas;
- Rodadas adicionais de sequenciamento para aumentar a profundidade de *reads* por amostra e saber quanto é preciso sequenciar para estimar a riqueza de espécies e a dissimilaridade da comunidade;
- Ampliação da base de dados referências genéticas para as plantas locais;
- Analisar os efeitos dos esforços de amostragem nos padrões de biodiversidade, para saber quantas amostras de eDNA é preciso por plot para capturar toda a biodiversidade de espécies e estimar a dissimilaridade da comunidade;
- Analisar o esforço amostral para saber se as amostras de um mesmo plot parecem mais umas com as outras do que as de outro plot;
- Continuar os estudos de expansão da amostragem para novas fitofisionomias.

## 6 CONCLUSÃO

Utilizando uma abordagem pioneira para o Brasil, baseada em eDNA proveniente de amostras de solo para avaliar a utilidade do DNA *metabarcoding* como uma ferramenta de monitoramento da flora da Canga da Serra dos Carajás, os resultados aqui obtidos foram muito promissores. Observou-se uma diversidade taxonômica relativamente alta dentro e entre os plots amostrados, pois apesar de serem de uma mesma fitofisionomia (canga arbustiva), existem diferenças entre a ocorrência e a densidade de plantas nos locais, já que fitofisionomias com as mesmas características podem apresentar pequenos pools que abrigam plantas específicas. O número de espécies identificadas por DNA *metabarcoding* foi maior do que foi observado pelo método tradicional nos mesmos pontos de amostragem, embora com um número menor de famílias. Ainda, foi possível identificar alguns táxons de plantas raras e ameaçadas de extinção, além de espécies que são da área de canga e S11D, mas que não foram encontrados pela amostragem tradicional. Importante ressaltar, que a identificação das plantas em níveis taxonômicos mais refinados, como famílias, gêneros e espécies só foi possível devido ao banco de dados de referências genéticas de plantas de Carajás disponível no ITVBiobase. Além disso, encontrar a melhor porcentagem de agrupamento em OTUs, identidade taxonômica e cobertura de alinhamento no *pipeline* de bioinformática empregado, desempenhou uma etapa importante na obtenção de resultados satisfatórios. Para o conjunto de dados de DNA *metabarcoding*, o índice de diversidade alfa mostrou maiores níveis de riqueza e diversidade na parcela 121, como também foi indicado pelo método de identificação morfológica. Os padrões de diversidade beta mostraram que as parcelas PCG-168 e PCG-121 são diferentes significativamente entre si e que não há diferenças significativas entre PCG-235 e as demais parcelas. A tecnologia de DNA *metabarcoding* pode ser uma ferramenta poderosa e complementar aos esforços na rotina de monitoramento da flora da canga, uma vez que, possibilita menor exposição de pessoas ao campo, diminui o tempo e o risco de trabalho. Sendo assim, os resultados aqui obtidos apontam que a validade da técnica de DNA *metabarcoding* com ITS2 para monitoramento de espécies de plantas na Serra dos Carajás foi demonstrada com sucesso.

## REFERÊNCIAS

- ALBERDI, A. et al. Scrutinizing key steps for reliable metabarcoding of environmental samples. **Methods in Ecology and Evolution**. v. 9, n. 1, p. 134-147, 2018.
- ALSOS, I. G. et al. Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. **PLoS ONE**, v. 13, n. 4, p. e0195403. 2018.
- ANDERSEN, K. et al. Metabarcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. **Molecular Ecology**, v. 21, n. 8, p. 1966-1979, 2012.
- BALAKIREV, E. S. et al. Complete mitochondrial genomes of the Cherskii's sculpin *Cottus cherskii* and Siberian taimen *Hucho taimen* reveal GenBank entry errors: incorrect species identification and recombinant mitochondrial genome. **Evolutionary Bioinformatics**. v. 13, n. 1176934317726783, 2017.
- BARSOUM, N. et al. Using DNA barcoding and metabarcoding to detect species and improve forest biodiversity monitoring. **Forest Research**, s/v, s/n, p. 1-14, 2018. Disponível em: <<https://www.forestresearch.gov.uk/research/using-dna-barcoding-and-metabarcoding-to-detect-species-and-improve-forest-biodiversity-monitoring/>> Acesso em: 13 de setembro de 2020.
- BARSOUM, N. et al. Demonstrating a rapid biodiversity assessment methodology to facilitate the management of multiple forest ecosystem services. **Forestry Commission Ecosystem Services Report**. Forestry Commission, s/v, s/n, p. 1-95, 2015.
- BIRKS, H. J. B. et al. Does pollen-assemblage richness reflect floristic richness? A review of recent developments and future challenges. **Review of Palaeobotany and Palynology**. v. 228, s/n, p. 1-25, 2016.
- BOESSENKOOL, S. et al. Use of ancient sedimentary DNA as a novel conservation tool for high-altitude tropical biodiversity. **Conservation Biology**. v. 28, n. 2, p. 446-455, 2014.
- BOGGS, L. M. et al. Single fragment or bulk soil DNA metabarcoding: which is better for characterizing biological taxa found in surface soils for sample separation? **Genes**, v. 10, n. 6, p. 431, 2019.
- 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**. v. 10, n. 12, p. e0143049, 2015.
- BORRELL, Y. J. et al. DNA in a bottle-Rapid metabarcoding survey for early alerts of invasive species in ports. **PLoS ONE**. v. 12, n. 9, p. e0183347, 2017.
- BUGLIONE, M. et al. A pilot study on the application of DNA metabarcoding for non-invasive diet analysis in the Italian hare. **Mammalian Biology**, v. 88, p. 31-32, 2018.

BUSH, A. et al. DNA metabarcoding reveals metacommunity dynamics in a threatened boreal wetland wilderness. **Proceedings of the National Academy of Sciences of the United States of America**, v. 117, p. 8539-8545, 2020.

BRASIL. Congresso. Planalto. Lei no 6.938, de 31 de agosto de 1981 e suas alterações (Leis nos 7.804, de 18 de julho de 1989, e 8.028, de 12 de abril de 1990) - Dispõe sobre a Política Nacional do Meio Ambiente, seus fins e mecanismos de formulação e aplicação. Disponível em< [www.planalto.gov.br](http://www.planalto.gov.br)>. Acesso em: 20 de agosto de 2020.

BRASIL. Ministério do Meio Ambiente. CONAMA. Resolução do CONAMA no 237 de 19 de dezembro de 1997 – Dispõe sobre os procedimentos e critérios utilizados no licenciamento ambiental. Disponível em:< [www.planalto.gov.br](http://www.planalto.gov.br)>. Acesso em: 20 de agosto de 2020.

CBOL, P. W. G. et al. A DNA barcode for land plants. **Proceedings of the National Academy of Sciences**. v. 106, n. 31, p. 12794-12797, 2009.

CBOL (CONSORTIUM FOR BARCODE OF LIFE). Consortium for Barcode of Life. 2010. Disponível em: <http://barcoding.si.edu/>). Acesso em: 15 set. 2020.

CANTU, V.A; SADURAL, J; EDWARDS, R. PRINSEQ++, a multi-threaded tool for fast and efficient quality control and preprocessing of sequencing datasets. **PeerJ Preprints**. v. 7, s/n, p. 1-3, 2019. Disponível em: <https://peerj.com/preprints/27553/>. Acesso em: 18 ago. 2020.

CAPORASO, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. **Nature Methods**, v. 7, p. 335-336, 2010.

CARMO, F. F. et al. Cangas: ilhas de ferro estratégicas para a conservação. **Ciência Hoje**, v. 295, n. 4, p. 48-50, 2012.

COISSAC, E. et al. From barcodes to genomes: extending the concept of DNA barcoding. **Molecular Ecology**. v. 25, n. 7, p. 1423-1428, 2016.

COMPSON, G. Z., et al. Metabarcoding From Microbes to Mammals: Comprehensive Bioassessment on a Global Scale. **Frontiers in Ecology and Evolution**. v.8, online, p. 379, 2020.

CORADIN, L.; TAFURI, A.; DUARTE, L. **Diretrizes para o engajamento do setor empresarial com a biodiversidade: para a promoção de ações voluntárias por entidades empresariais, visando a conservação e o uso sustentável da biodiversidade**. Brasília, DF: Ministério do Meio Ambiente, 2012. 164 p.

CORRÊA, G. R. **Pedogênese em platôs de Canga Ferrífera e Basaltos na Serra dos Carajás – PA**. 2011. 213f. Tese (Doutorado em Solos e Nutrição de Plantas). Universidade Federal de Viçosa, Minas Gerais, 2011.

COWART, D. A. et al. Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and Molecular Surveys of Seagrass Communities. **PLoS ONE**, v. 10, n. 2, p. e0117562, 2015.

CHINA PLANT BOL GROUP et al. Comparative analysis of a large dataset indicates that ITS should be incorporated into the core barcode for seed plants. **Proceedings of the National Academy of Sciences of the United States of America**. v. 108, n. 49, p. 19641-19646, 2011.

CHEN, S. et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. **PLoS ONE**, v. 5, n. 1, p. e8613, 2010.

CHENG, T. et al. Barcoding the kingdom Plantae: New PCR primers for ITS regions of plants with improved universality and specificity. **Molecular Ecology Resources**, v. 16, n. 1, p. 138-149, 2016.

DALSTRA H.; GUEDES S. Giant hydrothermal hematite deposits with Mg-Fe metasomatism: a comparison of the Carajás, Hamersley, and other iron ores. **Economic Geology**, v. 99, n. 8, p. 1793-1800, 2004.

DEINER, K. et al. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. **Molecular Ecology**, v. 26, n. 21, p. 5872-5895, 2017.

DEINER, K. et al. Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. **Methods in Ecology and Evolution**, v. 8, n. 12, p. 1888-1898, 2017.

DE BOER, H. J. et al. DNA metabarcoding of orchid-derived products reveals widespread illegal orchid trade. **Proceeding of the Royal Society B-Biological Sciences**, v. 284, n. 1863, p. 1-9, 2017.

DEWALT, E. R. DNA barcoding: a taxonomic point of view. **Journal of North American Benthological Society**, v. 30, n.1, p – 174 -181, 2011.

DORMONTT, E. E. et al. Advancing DNA Barcoding and Metabarcoding Applications for Plants Requires Systematic Analysis of Herbarium Collections - An Australian Perspective. **Frontiers in Ecology and Evolution**, v. 6, n. 134, p. 1-12, 2018.

DRAY, S.; DUFOUR, A. -B. The ade4 Package: Implementing the Duality Diagram for Ecologists. **Journal of Statistical Software**, v. 22, n. 4, p. 1-20, 2007.

EDWARDS, M. E. et al. Metabarcoding of modern soil DNA gives a highly local vegetation signal in Svalbard tundra. **The Holocene**, v. 28, n. 12, p. 2006-2016, 2018.

EDMUNDS, S. C. et al. Biodiversity research in the “big data” era: GigaScience and Pensoft work together to publish the most data-rich species description. **GigaScience**, v. 2, n. 1, p. 1-3 2013.

- FAHNER, N. A. et al. Large-Scale Monitoring of Plants through Environmental DNA Metabarcoding of Soil: Recovery, Resolution, and Annotation of Four DNA Markers. **PLoS ONE**, v. 11, n. 6, p. e0157505, 2016.
- FAZEKAS, A. J., et al. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. **PLoS ONE**, v.3, 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**, v.9, p.130-139, 2009.
- FILARDI, F. B. et al. Brazilian Flora 2020: Innovation and collaboration to meet Target 1 of the Global Strategy for Plant Conservation (GSPC). **Rodriguésia**, v. 69, n. 4, p. 1513-1527, 2018.
- FLØJGAARD, C. et al. Predicting provenance of forensic soil samples: Linking soil to ecological habitats by metabarcoding and supervised classification. **PLoS ONE**, v. 14, n. 7, p. e0202844, 2019.
- GASTAUER, M., et al. A Landscape heterogeneity and habitat amount drive plant diversity in Amazonian canga ecosystems. **Landscape Ecology**. v. 36, p.393–406, 2021.
- GERHARD, A, W; GUNSCH. Metabarcoding and machine learning analysis of environmental DNA in ballast water arriving to hub ports. **Environment International**. V. 124, online, p.312 – 319, 2019.
- GIULIETTI, A. M. et al. Eriocaulaceae in the Brazilian Amazon and the use of species distribution modelling in its conservation. **Rodriguésia**, v. 67, n. 4, p. 905-916, 2016.
- GIULIETTI, A. M; NOGUEIRA, M. G. C. Flora das cangas da Serra dos Carajás, Pará, Brasil: Nyctaginaceae. **Rodriguésia**, v. 68, n. 3, p.1045-1051, 2017.
- GIULIETTI, A. et al. Edaphic endemism in the Amazon: vascular plants of the canga of Carajás, Brazil. **The Botanical Review**, v. 85, p. 357-383, 2019.
- GONZALEZ, M. A, et al. Identification of Amazonian trees with DNA barcodes. **PLOS ONE**. v.4, n. 10, p. e7483, 2009.
- GOUS, A. et al. Plant-pollinator interactions over time: Pollen metabarcoding from bees in a historic collection. **Evolutionary Applications**, v.12, n. 2, p. 187-197, 2019.
- GREY, E.K. *et al.* Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. **Scientific Reports**. v.8, n.8843, 2018.
- HEBERT, P. D. N. et al. Biological identifications through DNA barcodes. **Proceedings of the Royal Society of London Series B. Biological sciences**, v. 270, n. 1512, p.313-321, 2003.

- HEBERT, P. D. N. et al. From writing to reading the encyclopedia of life. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v.371, p. 20150321. 2016
- HIIESALU, I. et al. Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. **Molecular Ecology**, v. 21, n. 8, p. 2004-2016, 2011.
- HOLLINGSWORTH, P.M. et al. Telling plant species apart with DNA: from barcodes to genomes. **Philosophical Transactions of the Royal Society B**, v. 371, n. 1702, 2016.
- HOLLINGSWORTH, P. M.; GRAHAM, S. W.; LITTLE, D. P. Choosing and using a plant DNA barcode. **PLoS ONE**, v. 6, n. 5, p. e19254, 2011.
- HOLLINGSWORTH, P. M. Refining the DNA barcode for land plants. **Proceeding of the National Academy of Sciences of the United States of America**, v. 108, n. 49, p. 19451-19452, 2011.
- ITV. **Relatório de atividades do ITV**. Instituto Tecnológico Vale (ITV). Belém. 128 p. 2019. Disponível em: <<http://www.itv.org/wp-content/uploads/2020/01/Relat%C3%B3rio-de-Atividades-ITV-Vers%C3%A3o-final.pdf>>. Acesso em: 5 de setembro de 2020.
- JACOBI, C. M. et al. Plant communities on ironstone outcrops - a diverse and endangered Brazilian ecosystem. **Biodiversity and Conservation**. v. 16, p. 2185-2200, 2007.
- JARDIM, M. A. G.; MARTINS, M. B. **Reflexões em Biologia da Conservação**. 1. Ed. Belém, PA: Museu Paraense Emílio Goeldi, 2018. 186 p.
- KARST, S.M., *et al.* High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. **Nature Methods**, v.18, online, p. 165–169 2021.
- KELLY, R. P. Making environmental DNA count. **Molecular Ecology Resources**, v. 16, n. 1, p. 10-12, 2016.
- KIMMERLING, N. et al. Quantitative species-level ecology of reef fish larvae via metabarcoding. **Nature Ecology and Evolution**, v. 2, p. 306-316, 2018.
- KINGSLEY C. B.; CORLETT, R. T. Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. **Biodiversity and Conservation**. v. 29, p- 2089-2121, 2020.
- KOLTER, A.; GEMEINHOLZER, B. Plant DNA barcoding necessitates marker-specific efforts to establish more comprehensive reference databases. **Genome**, s/v, s/n, p. 1-34, 2020. Disponível em: <https://cdnsiencepub.com/doi/abs/10.1139/gen-2019-0198?mobileUi=0&#pill-metrics-citations>. Acesso em: 11 set. 2020.

KUZMINA, M. L. et al. Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library. **BMC Ecology**, v. 12, n. 25, 2012.

KRESS, J. W. Plant DNA barcodes: Applications today and in the future. **Journal of Systematics and Evolution**. v. 55, n. 4, p.291-307, 2017.

LAMB, E. G. et al. A high-throughput belowground plant diversity assay using next-generation sequencing of the trnL intron. **Plant and Soil**, v. 404, p. 361-372, 2016.

LEAR, G. et al. Methods for extraction, storage, amplification and sequencing of DNA from environmental samples. **New Zealand Journal of Ecology**, v. 42, n. 1, p. 10-50, 2018.

LEVINE, N. M., et al. Ecosystem heterogeneity determines the ecological resilience of the Amazon to climate change. **Proceedings of the National Academy of Sciences of the United States of America**, v.113, s/n, p. 793-797, 2016.

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

LITTLEFAIR, J. E et al. DNA metabarcoding reveals changes in the contents of carnivorous plants along an elevation gradient. **Molecular Ecology Resources**, v. 28, n. 2, p. 281-292, 2018.

LOGARES, R., et al. Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. **Journal Microbiology Methods**. v. 91, n.1, p. 106-113, 2012.

LOHMANN, L. G; FIRETTI, F; GOMES, B. M. Flora das cangas da Serra dos Carajás, Pará, Brasil: Bignoniaceae. **Rodriguésia**, v. 69, n. 3, p. 1063-1079, 2018.

MAIA, U. M. et al. Forest matrix fosters high similarity in bee composition occurring on isolated outcrops within Amazon biome. **Environmental Entomology**, n. nva115, 2020.

MATESANZ, S et al. Estimating belowground plant abundance with DNA metabarcoding. **Molecular Ecology Resources**, v. 19, n. 5, p. 1265-1277, 2019.

MIRANDA, L. S., et al. Climate change impact on ecosystem functions provided by birds in southeastern Amazonia. **PLoS ONE**, v. 14, s/n, p. e0215229, 2019.

MOORHOUSE-GANN, R. J. et al. New universal ITS2 primers for high-resolution herbivory analyses using DNA metabarcoding in both tropical and temperate zones. **Scientific Reports**, v. 8, n. 8542, p. 1-15, 2018.

MORAES, B. C. et al. Variação espacial e temporal da precipitação no estado do Pará. **Acta Amazonica**, v. 35, n. 2, p. 207-214, 2005.

MOTA, N. F. de O; GIULIETTI, A. M. Flora das cangas da Serra dos Carajás, Pará, Brasil: Gnetaceae. **Rodriguésia**, Rio de Janeiro, v. 67, n. 5, p. 1191-1194, 2016.

MOTA, N. F. O. et al. **Vegetação sobre Sistemas Ferruginosos da Serra dos Carajás**. In: CARMO, F.F., KAMINO, L.H.Y. (orgs.). **Geossistemas ferruginosos do Brasil: áreas prioritárias para conservação da diversidade geológica e biológica, patrimônio cultural e serviços ambientais**. 1. ed. Belo Horizonte: 3i, 551 p. v. 1. 2015.

MOTA, N. F. O et al. Cangas da Amazônia: a vegetação única de Carajás evidenciada pela lista de fanerógamas. **Rodriguésia**, v. 69, n. 3, p. 1435-1488, 2018.

MCGEE, K. M.; EATON, W. D. A comparison of the wet and dryseason DNA-based soil invertebrate community characteristics in large patches of the bromeliad *Bromelia pinguin* in a primary forest in Costa Rica. **Applied Soil Ecology**, v. 87, p. 99-107, 2015.

MCMURDIE, P. J; HOLMES, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. **PLoS ONE**, v. 8, n. 4, p. e61217, 2013.

OLIVEIRA-DA-SILVA, F. R.; ILKIU-BORGES, A. L. Briófitas (Bryophyta e Marchantiophyta) das cangas da Serra dos Carajás, Pará, Brasil. **Rodriguésia**, v. 69, n. 3, p. 1405-1416, 2018.

OLIVEIRA, G. et al. DNA barcoding and genomics in the megadiverse Amazon altitude fields. **iBOL Barcode Bulletin**, v. 9, n. 5498, 2019.

OLIVEIRA, R. et al. PIMBA: a Pipeline for MetaBarcoding Analysis. **bioRxiv**, 2021.

OKSANEN, J. et al. Vegan: community ecology package. **R package version**, 2016.  
PARDUCCI, L. et al. Transley Reviews: Ancient plant DNA from lake sediments. **New Phytologist**, v. 213, n. 3, p.924-942, 2017.

PARDUCCI, L. et al. Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. **Molecular Ecology**, v. 22, n. 13, p. 3511-3524, 2013.

PARK, S.T; KIM, J. Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. **International Neurology Journal**. v. 20, online, p. 76 – 83, 2016.

PANSU, J. et al. Landscape-scale distribution patterns of earthworms inferred from soil DNA. **Soil Biology and Biochemistry**, v. 83, p. 100-105, 2015.

PEREIRA, J. B. S. et al. Two new species of *Isoetes* (Isoetaceae) from northern Brazil. **Phytotaxa**, v. 272, n. 2, p. 141-148, 2016.

PORNON, A. et al. Using metabarcoding to reveal and quantify plant-pollinator interactions. **Scientific Reports**, v. 6, n. 27282, 2016.

PORTER, T. M.; HAJIBABAEI, M. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. **Molecular Ecology**, v. 27, n. 2, p. 313-338, 2018.

RAMALHO, A. J. et al. Blind testing: DNA barcoding sheds light upon the identity of plant fragments as a subsidy for cave conservation. **Frontiers in Plant Science**, v. 9, n. 1052, 2018.

RATNASINGHAM, S.; HEBERT, P. D. N. BOLD: The Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). **Molecular Ecology Notes**, v. 7, n. 3, p. 355-364, 2007.

RESENDE, N. P.; BARBOSA, A. L. M. **Relatório de pesquisa de minério de ferro, Distrito Ferrífero da Serra dos Carajás, estado do Pará**. AMZA. Belém. 1248p. 1972.

RIEDEL, A. et al. Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. **Frontiers in Zoology**, v. 10, n. 15, p. 1-9, 2013.

RICHARDSON, R. T., et al. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. **Applications in Plant Sciences**, v. 3, p. apps.1400066, 2015.

RIZZINI, C. T. Tratado de fitogeografia do Brasil. Vol. 2. **Aspectos ecológicos**. Hucitec/Edusp, São Paulo. 374 p., 1979.

RUPPERT, K.; KLINE, R.; RAHMAN, M. D. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. **Global Ecology and Conservation**, v. 17, p. 55-63, 2019.

SAHOO, P. K. et al. Limnological characteristics and planktonic diversity of five tropical upland lakes from Brazilian Amazon. **Annales de Limnologie – International Journal of Limnology**, v. 53, p. 467-483, 2017.

SALINO, A; ARRUDA, A. J; ALMEIDA, T. E; Ferns and lycophytes from Serra dos Carajás, an Eastern Amazonian mountain range. **Rodriguésia**, v. 69, n. 3, p. 1417-1434, 2018.

SAMARAKOON, T.; WANG, S. Y.; ALFORD, M. H. Enhancing PCR amplification of DNA from recalcitrant plant specimens using a trehalose-based additive. **Applications in Plant Sciences**, v. 1, n.1, apps.1200236, 2013.

SECCO, R. S.; MESQUITA, A. L. Notas sobre a vegetação de canga da Serra Norte - I. **Boletim do Museu Paraense Emílio Goeldi**, v. 59, s/n, p. 1-13, 1983.

SILVA, C. E. P; WATANABE, M.; BRITO, R. M.; GIANNINI, T. C. **Sistemas de Polinização nas cangas de Carajás - Relatório final do projeto Biodiversidade e Mineração**. Instituto Tecnológico Vale (ITV), Belém. 2018. 33 p.

SILVA, M. A. **Arranjos político-institucionais: a criação de novos municípios, novas estruturas de poder e as lideranças locais - a divisão territorial de Marabá na década de 1980**. 2006. 188 f. Tese (Doutorado em Desenvolvimento Sustentável do Trópico Úmido) – Núcleo de Altos Estudos Amazônicos, Universidade Federal do Pará. Belém, 2006.

SILVEIRA, F. A. O. et al. Ecology and evolution of plant diversity in the endangered campo rupestre: a neglected conservation priority. **Plant and Soil**, v. 403, p. 129-152, 2016.

SOUZA-FILHO, P. W. M. et al. Mapping and quantification of ferruginous outcrop savannas in the Brazilian Amazon: a challenge for biodiversity conservation. **PLoS ONE**, v. 14, n. 1, p. e0211095, 2019.

SCHETTINI, A. T. et al. Exploring Al, Mn and Fe phytoextraction in 27 ferruginous rocky outcrops plant species. **Flora**, v. 238, p. 175-182, 2018.

SHNEER, V. S. DNA barcoding is a new approach in comparative genomics of plants. **Genetika**, v. 45, n. 11, p. 1436-1448, 2009.

SKIRYCYZ, A. et al. Canga biodiversity , a matter of mining. **Frontiers in Plant Science**, v. 5, n. 653, 2014.

SMITH, B. E. et al. From GenBank to GBIF: phylogeny-based predictive niche modeling tests accuracy of taxonomic identifications in large occurrence data repositories. **PLoS ONE**, v. 11, n. 3, p. e0151232, 2016.

STEPHENS, Z. D. et al. Big Data: Astronomical or Genomical? **PLoS Biology**, v. 13, n. 7, p. e1002195, 2015.

SØNSTEBØ, J. H. et al. Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. **Molecular Ecology Resources**, v. 10, n. 6, p. 1009-1018, 2010.

TARBELET P. et al. Soil sampling and isolation of extra-cellular DNA from large amount of starting material suitable for metabarcoding studies. **Molecular Ecology**, v. 21, n. 8, p. 1816-1820, 2012.

TEDERSOO, L. et al. Tree diversity and species identity effects on soil fungi, protists and animals are context-dependent. **The International Society for Microbial Ecology Journal**, v. 10, n. 2, p. 346-362, 2016.

TIMPANO, E. K.; SCHEIBLE, M. K. R.; MEIKLEJOHN, K. A. Optimization of the second internal transcribed spacer (ITS2) for characterizing land plants from soil. **PLoS ONE**, v. 15, n. 4, p. e0231436, 2020.

THOMSEN, P. F.; WILLERSLEV, E. Environmental DNA e an emerging tool in conservation for monitoring past and present biodiversity. **Biology Conservation**, v. 183, p. 4-18, 2015.

- TRINDADE, J. R.; ROSARIO, A. S.; SANTOS, J. U. M. Flora das cangas da Serra dos Carajás, Pará, Brasil: Myrtaceae. **Rodriguésia**, v. 69, n. 3, p. 1259-1277, 2018.
- URUMARUDAPPA, S. K. J. et al. DNA metabarcoding to unravel plant species composition in selected herbal medicines on the National List of Essential Medicines (NLEM) of Thailand. **Scientific Reports**. v.10, n.18259, 2020.
- VALENTINI, A. et al. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. **Molecular Ecology**, v. 25, n. 4, p. 929-942, 2016.
- VASCONCELOS, S. et al. New insights on the phylogenetic relationships among the traditional *Philodendron* subgenera and the other groups of the *Homalomena* clade (Araceae). **Molecular Phylogenetics and Evolution**, v. 127, p. 168-178, 2018.
- VASCONCELOS, S. et al. Unravelling the plant diversity of the Amazonian *canga* through DNA barcoding. **Authorea**. 2021.
- VERNOOY, R. et al. Barcoding life to conserve biological diversity: beyond the taxonomic imperative. **PLoS Biology**, v. 8, n. 7, p. e1000417, 2010.
- VIANA, P. L. et al. Flora das cangas da Serra dos Carajás, Pará, Brasil: história, área de estudos e metodologia. **Rodriguésia**, Rio de Janeiro, v. 67, n. 5, p. 1107-1124, 2016.
- ZAFEIROPOULOS H, et al. Supporting data for “PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S rRNA, ITS, and COI marker genes”. **GigaScience Database**. v. 9, n. 3, 2020.
- ZAPPI, D. et al. Growing knowledge: an overview of seed plant diversity in Brazil. **Rodriguésia**, Rio de Janeiro, v. 66, n. 4, p. 1085-1113, 2015.
- ZAPPI, D. C. Paisagens e plantas de Carajás/Landscapes and plants of Carajás. **Instituto Tecnológico Vale (ITV)**, Belém. 248 p. 2017.
- ZEPEDA MENDOZA, M. L.; SICHERITZ-PONTÉN, T.; GILBERT, M. T. P. Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. **Briefings in Bioinformatics**. v. 16, n. 5, p-745-758, 2015.
- ZINGER, L., et al. DNA metabarcoding – need for robust experimental designs to draw sound ecological conclusions. **Molecular Ecology**, v. 28, p.1857-1862, 2019.
- ZHAO, L. L. et al. Internal transcribed spacer 2 (ITS2) barcodes: A useful tool for identifying Chinese *Zanthoxylum*. **Applications in Plant Sciences**, v. 6, n. 6, p. 1-8, 2018.
- ZHANG, J. et al. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. **Bioinformatics**, v. 30, n. 5, p. 614-620. 2014.

ZHU, R. W. et al. Establishment of the most comprehensive ITS2 barcode database to date of the traditional medicinal plant *Rhodiola* (Crassulaceae). **Scientific Reports**, v. 7, n. 10051, 2017.

YOUNG, J. M., AUSTIN J. J., WEYRICH, L. S. Soil DNA metabarcoding and high-throughput sequencing as a forensic tool: considerations, potential limitations and recommendations. **FEMS Microbiology Ecology**, v. 93, n. 2, p. 207, 2017.

YOCCOZ, N. et al. DNA from soil mirrors plant taxonomic and growth form diversity. **Molecular Ecology**, v. 21, n. 15, p. 3647-3655, 2012.

WICKHAM, H. ggplot2: Elegant Graphics for Data Analysis. **Springer-Verlag**, 2009.

## **APÊNDICES**

**APÊNDICE A** - Levantamento da flora por meio de taxonomia clássica.

**Tabela 2.** Lista de espécies encontradas pela consultoria nas áreas selecionadas em S11D.

Família	Espécie	Canga		
		PCG -168	PCG- 121	PCG -235
Araceae	<i>Anthurium lindmanianum</i>	-	-	Sim
Araceae	<i>Philodendron quinquenervium</i>	-	-	Sim
Asteraceae	<i>Monogereion carajensis</i>	Sim	-	-
Bignoniaceae	<i>Anemopaegma scabriusculum</i>	-	Sim	Sim
Cactaceae	<i>Cereus jamacarum</i>	-	-	Sim
Convolvulaceae	<i>Ipomoea marabensis</i>	-	Sim	-
Erythroxylaceae	<i>Erythroxylum nelson-rosae</i>	-	Sim	Sim
Euphorbiaceae	<i>Alchornea discolor</i>	Sim	Sim	-
Fabaceae	<i>Bauhinia pulchella</i>	-	Sim	Sim
Fabaceae	<i>Mimosa acutistipula var. ferrea</i>	Sim	Sim	Sim
Fabaceae	<i>Periandra mediterranea var. mediterranea</i>	-	Sim	-
Lamiaceae	<i>Premna serratifolia</i>	-	Sim	Sim
Malpighiaceae	<i>Banisteriopsis stellaris</i>	-	Sim	-
Malpighiaceae	<i>Byrsonima crassifolia</i>	Sim	-	-
Malpighiaceae	<i>Byrsonima spicata</i>	-	Sim	-
Marcgraviaceae	<i>Norantea guianensis</i>	-	-	Sim
Myrtaceae	<i>Eugenia flavescens</i>	-	Sim	-
Myrtaceae	<i>Myrcia cuprea</i>	-	-	Sim
Ochnaceae	<i>Ouratea castaneifolia</i>	Sim	Sim	-
Olacaceae	<i>Heisteria ovata</i>	-	-	Sim
Orchidaceae	<i>Sobralia liliastrum</i>	-	-	Sim
Poaceae	<i>Axonopus purpusii</i>	Sim	-	-
Poaceae	<i>Ichnanthus sp</i>	-	Sim	-
Rubiaceae	<i>Cordia myrciifolia</i>	-	-	Sim
Sapotaceae	<i>Pouteria ramiflora</i>	-	Sim	-
Simaroubaceae	<i>Simarouba amara</i>	-	Sim	-
Velloziaceae	<i>Vellozia glochidea</i>	-	-	Sim
Vitaceae	<i>Cissus erosa</i>	Sim	-	-
Vochysiaceae	<i>Callisthene microphylla</i>	Sim	Sim	-
Vochysiaceae	<i>Chamaecrista desvauxii</i>	Sim	-	-

**Fonte:** Consultoria (2016).

## APÊNDICE B – Metadados das parcelas

**Tabela 3:** Metadados da parcela 168.

Bloco	Tipologia	Coletor *	Parcela	Data da Coleta	Zona Leste	Zona Norte	codeITV
B10	Canga	V, J, A	PCG-168	20.08.19	22M 571502,32	9291865,3	ITV21541
							ITV21542
							ITV21543
B10	Canga	V, J, A	PCG-168	20.08.19	22M 571502,32	9291865,3	ITV21544
							ITV21545
							ITV21546
B10	Canga	V, J, A	PCG-168	20.08.19	22M 571502,32	9291865,3	ITV21547
							ITV21548
							ITV21549
B10	Canga	V, J, A	PCG-168	20.08.19	22M 571502,32	9291865,3	ITV21550
							ITV21551
							ITV21552
B10	Canga	V, J, A	PCG-168	20.08.19	22M 571502,32	9291865,3	ITV21553
							ITV21554
							ITV21555
B10	Canga	V, J, A	PCG-168	20.08.19	22M 571502,32	9291865,3	ITV21556
							ITV21557
							ITV21558
B10	Canga	V, J, A	PCG-168	20.08.19	22M 571502,32	9291865,3	ITV21559

\*Vitória Catarina Martins (V), José Augusto Bittencourt (J), Ariane Flexa de Castro (A).

**Fonte:** Autora, 2021.

**Tabela 4:** Metadados da parcela 121.

Bloco	Tipologia	Coletor *	Parcela	Data da Coleta	Zona		codeITV
					Leste	North	
B10	Canga	V, J, A	PCG-121	22.08.19	22M 571466,47	9291730	ITV21580
							ITV21581
							ITV21582
B10	Canga	V, J, A	PCG-121	22.08.19	22M 571466,47	9291730	ITV21583
							ITV21584
							ITV21585
B10	Canga	V, J, A	PCG-121	22.08.19	22M 571466,47	9291730	ITV21586
							ITV21587
							ITV21588
B10	Canga	V, J, A	PCG-121	22.08.19	22M 571466,47	9291730	ITV21589
							ITV21590
							ITV21591
B10	Canga	V, J, A	PCG-121	22.08.19	22M 571466,47	9291730	ITV21592
							ITV21593
							ITV21594
B10	Canga	V, J, A	PCG-121	22.08.19	22M 571466,47	9291730	ITV21595
							ITV21596
							ITV21597
B10	Canga	V, J, A	PCG-121	22.08.19	22M 571466,47	9291730	ITV21598
							ITV21599

\*Vitória Catarina Martins (V), José Augusto Bittencourt (J), Ariane Flexa de Castro (A).

Fonte: Autora, 2021.

**Tabela 5:** Metadados da parcela 235.

Bloco	Tipologia	Coletor *	Parcela	Data da Coleta	Zona Leste	Zona North	codeITV
B11	Canga	V, J, A	PCG-235	22.08.19	22M 568845,31	9292791,8	ITV21620
							ITV21621
							ITV21622
							ITV21623
B11	Canga	V, J, A	PCG-235	22.08.19	22M 568845,31	9292791,8	ITV21624
							ITV21625
							ITV21626
B11	Canga	V, J, A	PCG-235	22.08.19	22M 568845,31	9292791,8	ITV21627
							ITV21628
B11	Canga	V, J, A	PCG-235	22.08.19	22M 568845,31	9292791,8	ITV21629
							ITV21630
							ITV21631
B11	Canga	V, J, A	PCG-235	22.08.19	22M 568845,31	9292791,8	ITV21632
							ITV21633
B11	Canga	V, J, A	PCG-235	22.08.19	22M 568845,31	9292791,8	ITV21634
							ITV21635
							ITV21636
B11	Canga	V, J, A	PCG-235	22.08.19	22M 568845,31	9292791,8	ITV21637
							ITV21638
B11	Canga	V, J, A	PCG-235	22.08.19	22M 568845,31	9292791,8	ITV21639

\*Vitória Catarina Martins (V), José Augusto Bittencourt (J), Ariane Flexa de Castro (A).

**Fonte:** Autora, 2021.

## APÊNDICE C – Quantificação de DNA

**Tabela 6:** Informações das concentrações de DNA da parcela 235

Parcela	Sample	Triplicata	PowerSoil	Sample	Triplicata	PowerSoil
			ng/µl			ng/µl
PCG-235	ITV21620	20.1	13,4	ITV21630	30.1	12,6
		20.2	13,4		30.2	14,9
		20.3	12,4		30.3	20,4
		21.1	17,14		31.1	7,24
PCG-235	ITV21621	21.2	14,6	ITV21631	31.2	9,32
		21.3	21,2		31.3	6,12
		22.1	14,8		32.1	19
PCG-235	ITV21622	22.2	13,8	ITV21632	32.2	18,4
		23.3	17,8		32.3	25
		23.1	9,38		33.1	8,16
PCG-235	ITV21623	23.2	11,3	ITV21633	33.2	13,5
		23.3	10,9		33.3	18,7
		24.1	16,3		34.1	11,3
PCG-235	ITV21624	24.2	13	ITV21634	34.2	12,1
		24.3	19		34.3	9,6
		25.1	11,9		35.1	8,86
PCG-235	ITV21625	25.2	10,5	ITV21635	35.2	32,2
		25.3	10,4		35.3	35,4
		26.1	25		36.1	39,2
PCG-235	ITV21626	26.2	9,18	ITV21636	36.2	62
		26.3	23,4		36.3	19,9
		27.1	9,88		37.1	28,8
PCG-235	ITV21627	27.2	8,02	ITV21637	37.2	31,4
		27.3	9,08		37.3	40,4
		28.1	14,7		38.1	7,04
PCG-235	ITV21628	28.2	19	ITV21638	38.2	7,18
		28.3	13		38.3	9,04
		29.1	11,1		39.1	16
PCG-235	ITV21629	29.2	9,3	ITV21639	39.2	14,3
		29.3	16,5		39.3	15,3
PCG-235	Controle	1	Too Low	Volume Final		100 µl

Fonte: Autora, 2021.

**Tabela 7:** Informações das concentrações de DNA da parcela 121.

Parcela	Sample	Triplicata	PowerSoil	Sample	Triplicata	PowerSoil
			ng/ $\mu$ l			ng/ $\mu$ l
PCG-121	ITV21580	80.1	34,6	ITV21590	90.1	27,8
		80.2	20,2		90.2	31
		80.3	21,4		90.3	27
PCG-121	ITV21581	81.1	17,5	ITV21591	91.1	52,2
		81.2	16,1		91.2	43,2
		81.3	22,2		91.3	54
PCG-121	ITV21582	82.1	12,2	ITV21592	92.1	49,2
		82.2	18		92.2	36
		82.3	19,2		92.3	44
PCG-121	ITV21583	83.1	24	ITV21593	93.1	34,2
		83.2	27,4		93.2	30,6
		83.3	17,5		93.3	35,6
PCG-121	ITV21584	84.1	30	ITV21594	94.1	48,6
		84.2	41,6		94.2	52,4
		84.3	31,6		94.3	46
PCG-121	ITV21585	85.1	25,8	ITV21595	95.1	39,6
		85.2	40		95.2	42,2
		85.3	25,6		95.3	27,8
PCG-121	ITV21586	86.1	28,2	ITV21596	96.1	22
		86.2	25		96.2	37,4
		86.3	24,8		96.3	17,3
PCG-121	ITV21587	87.1	28,4	ITV21597	97.1	57,4
		87.2	34,8		97.2	39,4
		87.3	40,6		97.3	45,8
PCG-121	ITV21588	88.1	21,6	ITV21598	98.1	36,2
		88.2	32,2		98.2	28,6
		88.3	27,2		98.3	25,6
PCG-121	ITV21589	89.1	36,8	ITV21599	99.1	25,8
		89.2	18,1		99.2	23,8
		89.3	18,9		99.3	15,9
PCG-121	Controle	2	Too Low	Volume Final		100 $\mu$ l

Fonte: Autora, 2021.

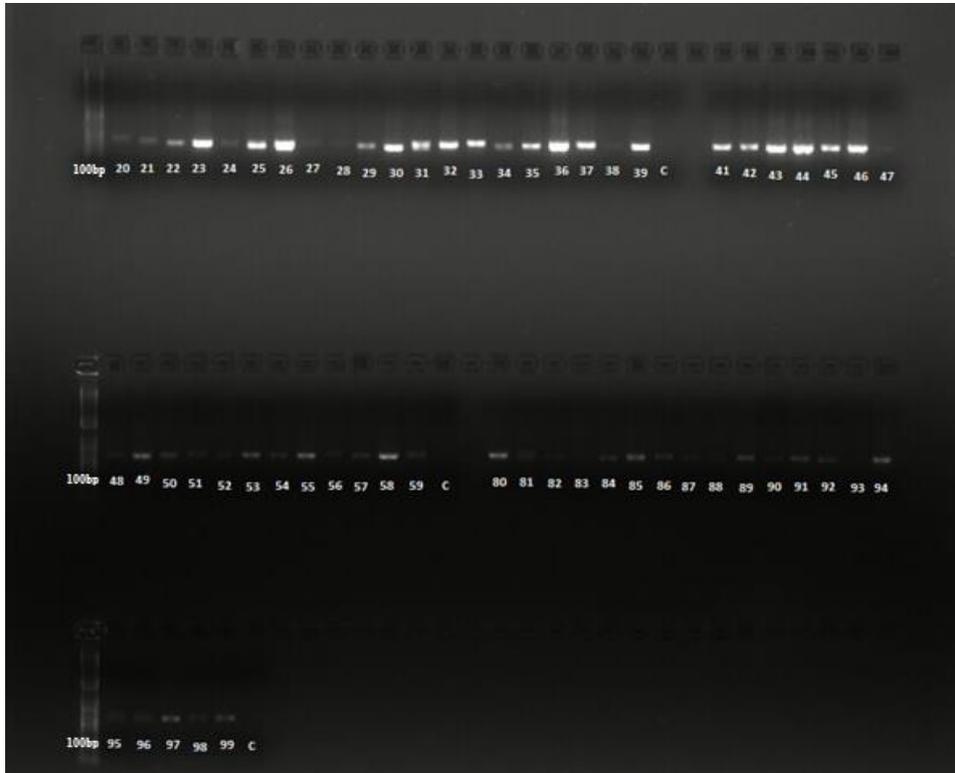
**Tabela 8:** Informações das concentrações de DNA da parcela 168.

Parcela	Sample	Triplicata	PowerSoil	Sample	Triplicata	PowerSoil
			ng/ $\mu$ l			ng/ $\mu$ l
PCG-168	ITV21541	41.1	25,8	ITV21551	51.1	57,4
		41.2	31,6		51.2	61,8
		41.3	19,2		51.3	77,2
PCG-168	ITV21542	42.1	27,6	ITV21552	52.1	22,8
		42.2	40,8		52.2	15,1
		42.3	30,2		52.3	19,1
PCG-168	ITV21543	43.1	40,2	ITV21553	53.1	29,2
		43.2	33		53.2	38,4
		43.3	52,8		53.3	24,8
PCG-168	ITV21544	44.1	16	ITV21554	54.1	49,4
		44.2	24,4		54.2	40
		44.3	34,2		54.3	33
PCG-168	ITV21545	45.1	38,6	ITV21555	55.1	39,4
		45.2	17,6		55.2	27,6
		45.3	20,8		55.3	46
PCG-168	ITV21546	46.1	26,4	ITV21556	56.1	39,8
		46.2	44		56.2	37,8
		46.3	17,4		56.3	33,2
PCG-168	ITV21547	47.1	12,1	ITV21557	57.1	23
		47.2	12,1		57.2	10,3
		47.3	17,9		57.3	11,4
PCG-168	ITV21548	48.1	13	ITV21558	58.1	35
		48.2	21,2		58.2	22,6
		48.3	16,4		58.3	16,8
PCG-168	ITV21549	49.1	35	ITV21559	59.1	15,2
		49.2	40,2		59.2	15,2
		49.3	25,2		59.3	19
PCG-168	ITV21550	50.1	61,6	ITV21560	60.1	X
		50.2	35,2		60.2	X
		50.3	52		60.3	X
PCG-168	Controle	2	TL	Volume Final		100 $\mu$ l

Fonte: Autora, 2021.

## APÊNDICE D – Amplificação DNA

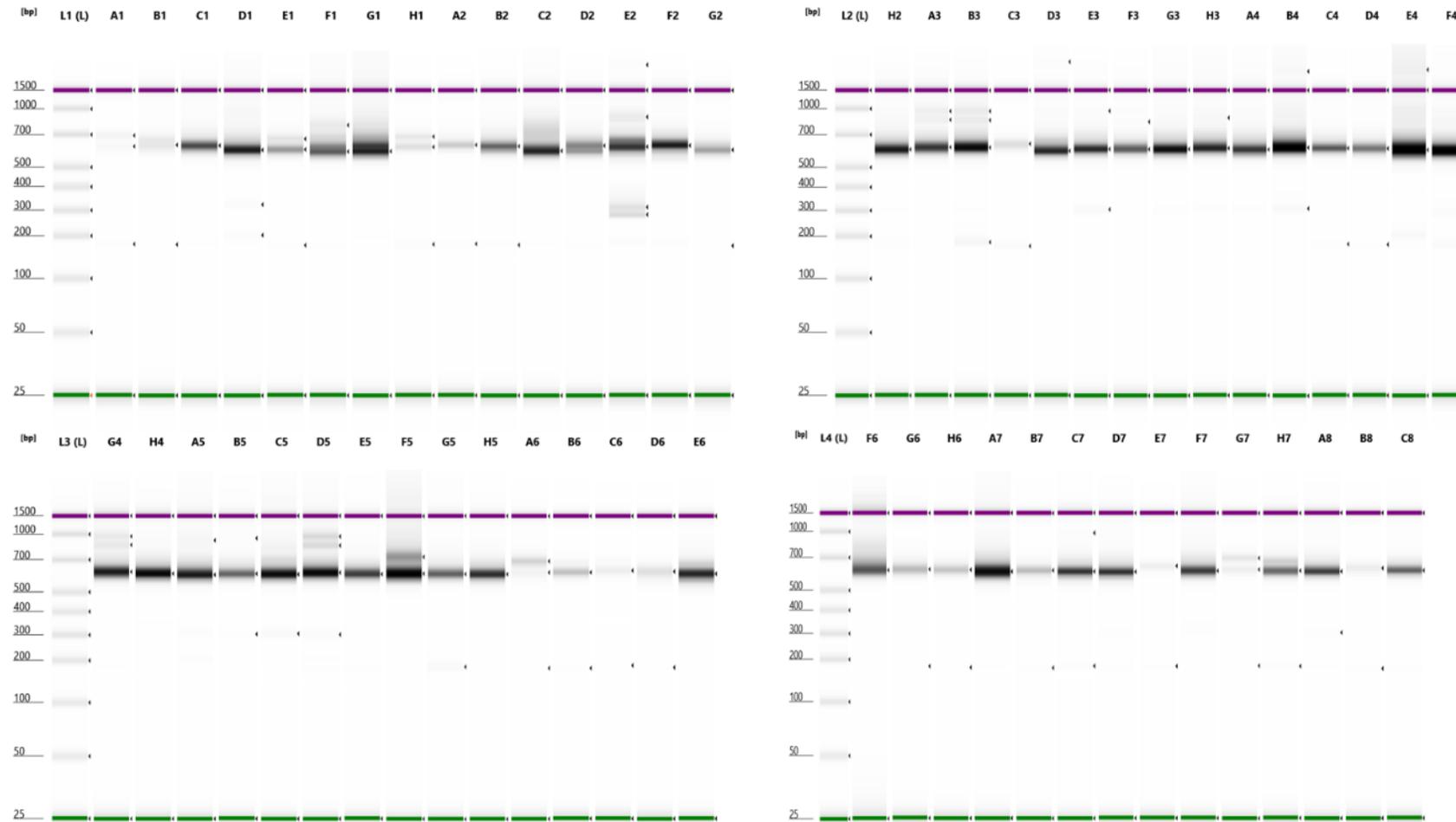
**Figura 31:** Eletroforese em gel de agarose 1% dos DNA's provenientes de amostras de solo da parcela PCG-235, PCG-168 e PCG-121 extraídos a partir do Kit Power Soil.



Fonte: Autora, 2021.

## APÊNDICE E – Tamanho dos amplicons

Figura 32: Corrida no TapeStation System 4200 mostrando o perfil dos tamanhos dos amplicons



Fonte: Autora, 2021.

## APÊNDICE F – Profundidade do Sequenciamento

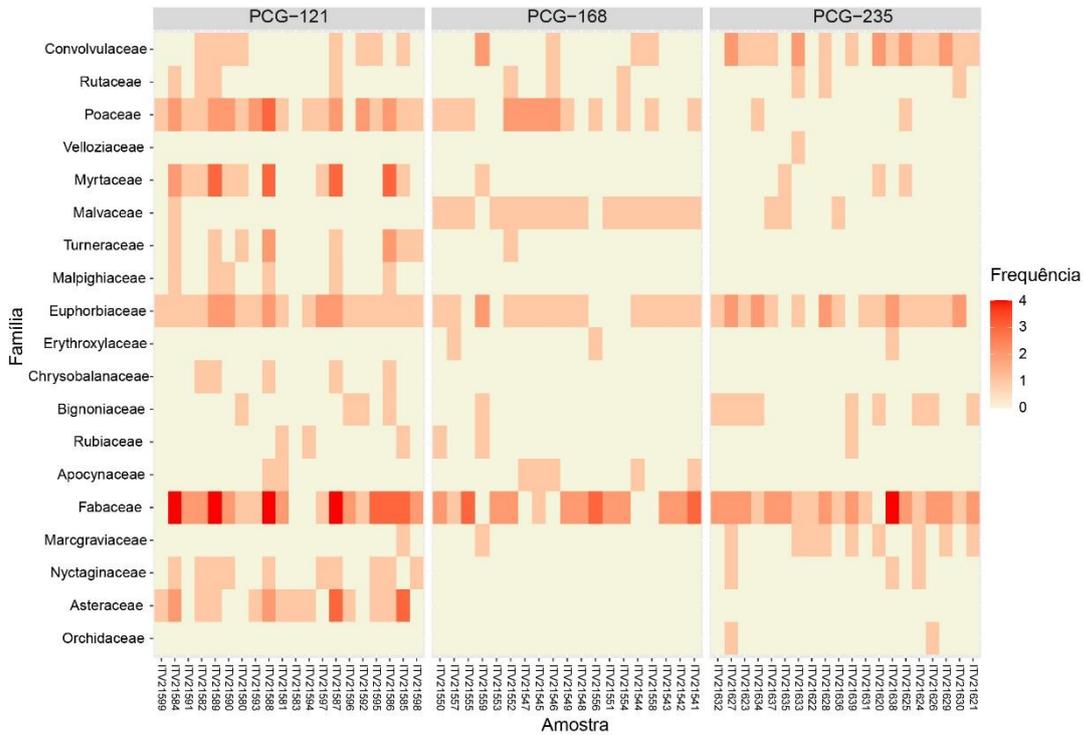
**Tabela 9:** Número de sequências e bases produzidas e que passaram pelo controle de qualidade do sequenciamento.

Parcela	Amostra	Input sequences	Input_bases	Good_sequences (pairs)	Good_bases (pairs)
PCG-168	ITV21541	333151	99601657	303270	181665054
	ITV21542	379346	113532878	337516	201965142
	ITV21543	324209	97191638	290295	173924274
	ITV21544	267568	80114639	243386	145842465
	ITV21545	323142	96166708	291468	174524823
	ITV21546	862698	258072372	790517	473669703
	ITV21547	57579	16910459	49865	29883923
	ITV21548	85242	25241101	75456	45154274
	ITV21549	256544	76812425	215765	128944720
	ITV21550	786160	231938878	667772	399426836
	ITV21551	681475	203175478	600130	359154574
	ITV21552	723454	215563341	645913	386930892
	ITV21553	418660	125086959	369861	221301997
	ITV21554	761460	226573118	677890	405928069
	ITV21555	463706	138749968	409476	245253207
	ITV21556	404132	120585016	360290	215820892
	ITV21557	82362	24428121	67696	40477557
	ITV21558	619939	185469396	558479	334583496
	ITV21559	376292	105792940	314918	188342676
PCG-121	ITV21580	354926	106140200	312994	187449197
	ITV21581	42735	12402029	35143	20984553
	ITV21582	127555	36542001	102312	61063171
	ITV21583	10416	3067228	6034	3597956
	ITV21584	58813	17010931	47776	28521568
	ITV21585	218796	65035870	193592	115886373
	ITV21586	275101	81695686	237097	141813548
	ITV21587	100564	29692319	82386	49279225
	ITV21588	81959	23643695	68129	40772496
	ITV21589	662132	197694098	529624	316905725
	ITV21590	118220	33989611	98266	58782370
	ITV21591	207618	61091914	183589	109943446
	ITV21592	81917	24229718	70590	42258021
	ITV21593	75461	18588416	48053	28619774
	ITV21594	422513	126478384	383665	229927105
	ITV21595	69258	20023644	60480	36188427
	ITV21596	40672	11912635	34257	20491369
	ITV21597	555330	165270956	500409	299841974
	ITV21598	47305	12138976	32609	19485811
ITV21599	151516	44349758	131634	78768632	
PCG-235	ITV21620	67805	15972152	41269	24573717
	ITV21621	33614	9755674	25866	15496524
	ITV21622	119352	35364716	101145	60630921
	ITV21623	153654	45638571	136371	81651611
	ITV21624	91892	26701898	68779	41144059
	ITV21625	249053	73648171	190992	114516283
	ITV21626	214882	64208886	182140	109189595
	ITV21627	130341	33318705	89766	53539255
	ITV21628	129768	35795759	96624	57848796
	ITV21629	521619	152713642	452674	271333336
	ITV21630	174563	52084035	132679	79537336
	ITV21631	174709	51373758	125866	75407554
	ITV21632	213156	62537129	171376	102689139
	ITV21633	213515	62521642	150488	90264195
	ITV21634	327971	96417105	260908	156085263
	ITV21635	300677	89104107	268101	160614490
	ITV21636	216358	64397126	197717	118074357
	ITV21637	347986	100813311	293418	175591990
	ITV21638	110180	27486461	74129	44227289
ITV21639	245757	73366139	219758	131582218	

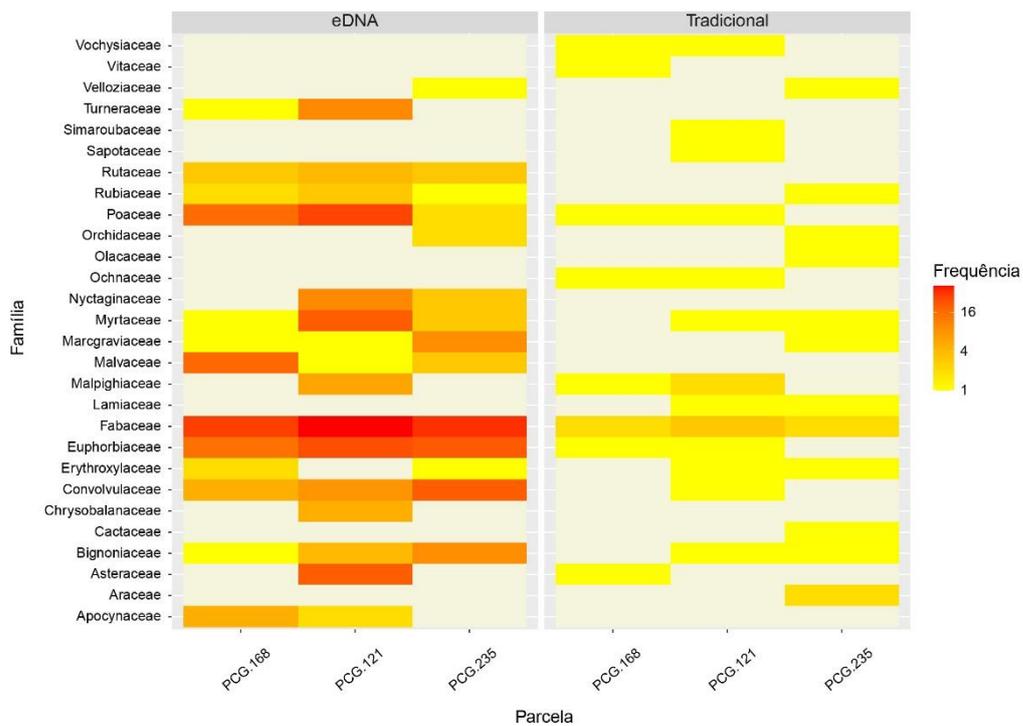
Fonte: Autora, 2021.

## APÊNDICE G – Frequência de ocorrência das famílias

**Figura 33:** Frequência de identificação das famílias de plantas detectadas via DNA metabarcoding por amostra.



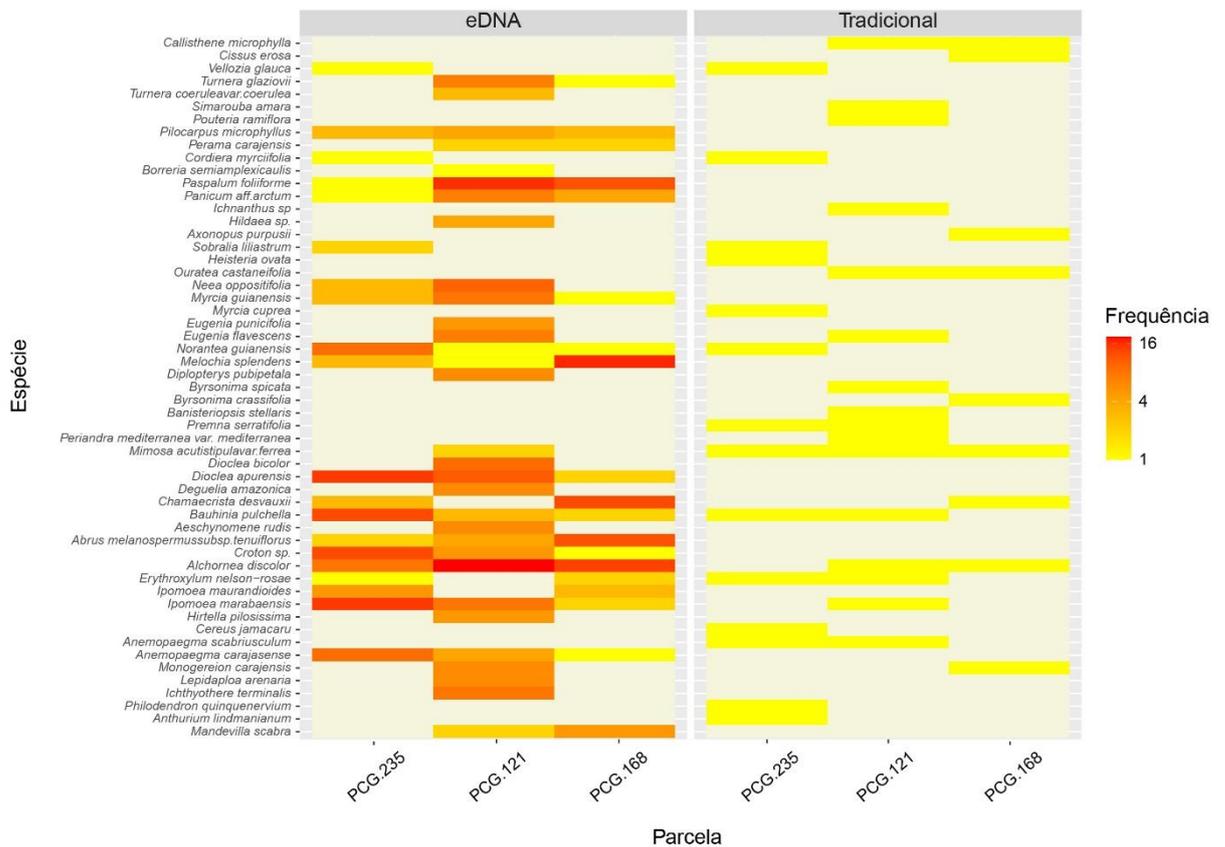
**Figura 34:** Frequência de identificação das famílias de plantas detectadas via DNA metabarcoding e método tradicional em distintas parcelas.



Fonte: Autora, 2021.

## APÊNDICE H – Frequência de ocorrência das espécies

**Figura 35:** Frequência de identificação das espécies de plantas detectadas via DNA metabarcoding e método tradicional em distintas parcelas localizadas na Serra Sul em Canaã dos Carajás.



Fonte: Autora, 2021.

Tabela 10: Número de OTU e frequência de ocorrência das espécies por parcela.

Família	Parcela	Espécie	PCG-168		PCG-121		PCG-235	
			Nº de OTU's com a presença da espécie	Frequência de ocorrência de OTU's da Espécie %	Nº de OTU's com a presença da espécie	Frequência de ocorrência de OTU's da Espécie %	Nº de OTU's com a presença da espécie	Frequência de ocorrência de OTU's da Espécie %
Apocynaceae		<i>Mandevilla scabra</i>	5	5,10	2	1,08	0	0,00
Asteraceae		<i>Ichthyothere terminalis</i>	0	0,00	8	4,30	0	0,00
Asteraceae		<i>Lepidaploa arenaria</i>	0	0,00	6	3,23	0	0,00
Asteraceae		<i>Monogereion carajensis</i>	0	0,00	6	3,23	0	0,00
Bignoniaceae		<i>Anemopaegma carajasense</i>	1	1,02	4	2,15	9	8,11
Chrysobalanaceae		<i>Hirtella pilosissima</i>	0	0,00	5	2,69	0	0,00
Convolvulaceae		<i>Ipomoea marabaensis</i>	2	2,04	8	4,30	15	13,51
Convolvulaceae		<i>Ipomoea maurandioides</i>	3	3,06	0	0,00	5	4,50
Erythroxylaceae		<i>Erythroxylum nelson-rosae</i>	2	2,04	0	0,00	1	0,90
Euphorbiaceae		<i>Alchornea discolor</i>	14	14,29	19	10,22	8	7,21
Euphorbiaceae		<i>Croton sp.</i>	1	1,02	5	2,69	13	11,71
Fabaceae		<i>Abrus melanospermussubsp.tenuiflorus</i>	12	12,24	4	2,15	2	1,80
Fabaceae		<i>Aeschynomene rudis</i>	0	0,00	6	3,23	0	0,00
Fabaceae		<i>Bauhinia pulchella</i>	2	2,04	3	1,61	13	11,71
Fabaceae		<i>Chamaecrista desvauxii</i>	13	13,27	0	0,00	3	2,70
Fabaceae		<i>Deguelia amazonica</i>	0	0,00	6	3,23	0	0,00
Fabaceae		<i>Dioclea apurensis</i>	2	2,04	11	5,91	15	13,51
Fabaceae		<i>Dioclea bicolor</i>	0	0,00	9	4,84	0	0,00
Fabaceae		<i>Mimosa acutistipularvar.ferrea</i>	0	0,00	2	1,08	0	0,00
Malpighiaceae		<i>Diplopterys pubipetala</i>	0	0,00	6	3,23	0	0,00
Malvaceae		<i>Melochia splendens</i>	17	17,35	1	0,54	3	2,70
Marcgraviaceae		<i>Norantea guianensis</i>	1	1,02	1	0,54	9	8,11
Myrtaceae		<i>Eugenia flavescens</i>	0	0,00	7	3,76	0	0,00
Myrtaceae		<i>Eugenia punicifolia</i>	0	0,00	5	2,69	0	0,00
Myrtaceae		<i>Myrcia guianensis</i>	1	1,02	8	4,30	3	2,70
Nyctaginaceae		<i>Neea oppositifolia</i>	0	0,00	10	5,38	3	2,70
Orchidaceae		<i>Sobralia liliastrum</i>	0	0,00	0	0,00	2	1,80
Poaceae		<i>Hiladaea sp.</i>	0	0,00	4	2,15	0	0,00
Poaceae		<i>Panicum aff.arctum</i>	4	4,08	7	3,76	1	0,90
Poaceae		<i>Paspalum foliiforme</i>	12	12,24	16	8,60	1	0,90
Rubiaceae		<i>Borreria semiamplexicaulis</i>	0	0,00	1	0,54	0	0,00
Rubiaceae		<i>Cordia myrciifolia</i>	0	0,00	0	0,00	1	0,90
Rubiaceae		<i>Perama carajensis</i>	2	2,04	2	1,08	0	0,00
Rutaceae		<i>Pilocarpus microphyllus</i>	3	3,06	4	2,15	3	2,70
Turneraceae		<i>Turnera coeruleavar.coerulea</i>	0	0,00	3	1,61	0	0,00
Turneraceae		<i>Turnera glaziovii</i>	1	1,02	7	3,76	0	0,00
Velloziaceae		<i>Vellozia glauca</i>	0	0,00	0	0,00	1	0,90
<b>Total</b>		<b>Total</b>	<b>98</b>		<b>186</b>		<b>111</b>	

Fonte: Autora, 2021.

## APÊNDICE I – Riqueza e diversidade de espécies

**Tabela 11:** Valores de diversidade Alfa baseada nos índices de riqueza (Chao1) e diversidade de espécies (Shannon), considerando presença/ausência nas diferentes parcelas localizadas na S11D.

Parcela	Amostra	Observed	Chao1	Shannon
PCG-168	ITV21541	7	28	1.9459101
	ITV21542	4	10	1.3862944
	ITV21543	4	10	1.3862944
	ITV21544	4	10	1.3862944
	ITV21545	6	21	1.7917595
	ITV21546	7	28	1.9459101
	ITV21547	5	15	1.6094379
	ITV21548	4	10	1.3862944
	ITV21549	5	15	1.6094379
	ITV21550	6	21	1.7917595
	ITV21551	3	6	1.0986123
	ITV21552	8	36	2.0794415
	ITV21553	3	6	1.0986123
	ITV21554	5	15	1.6094379
	ITV21555	5	15	1.6094379
	ITV21556	5	15	1.6094379
	ITV21557	5	15	1.6094379
	ITV21558	4	10	1.3862944
	ITV21559	8	36	2.0794415
	PCG-121	ITV21580	7	28
ITV21581		7	28	1.9459101
ITV21582		10	55	2.3025851
ITV21583		1	1	0.0000000
ITV21584		16	136	2.7725887
ITV21585		13	91	2.5649494
ITV21586		16	136	2.7725887
ITV21587		20	210	2.9957323
ITV21588		20	210	2.9957323
ITV21589		18	171	2.8903718
ITV21590		10	55	2.3025851
ITV21591		5	15	1.6094379
ITV21592		6	21	1.7917595
ITV21593		5	15	1.6094379
ITV21594		4	10	1.3862944
ITV21595		8	36	2.0794415
ITV21596		5	15	1.6094379
ITV21597		6	21	1.7917595
ITV21598		6	21	1.7917595
ITV21599		3	6	1.0986123
PCG-235	ITV21620	6	21	1.7917595
	ITV21621	5	15	1.6094379
	ITV21622	2	3	0.6931472
	ITV21623	5	15	1.6094379
	ITV21624	6	21	1.7917595
	ITV21625	7	28	1.9459101
	ITV21626	6	21	1.7917595
	ITV21627	10	55	2.3025851
	ITV21628	7	28	1.9459101
	ITV21629	6	21	1.7917595
	ITV21630	5	15	1.6094379
	ITV21631	2	3	0.6931472
	ITV21632	4	10	1.3862944
	ITV21633	7	28	1.9459101
	ITV21634	6	21	1.7917595
	ITV21635	4	10	1.3862944
	ITV21636	3	6	1.0986123
	ITV21637	5	15	1.6094379
	ITV21638	9	45	2.1972246
	ITV21639	6	21	1.7917595

Fonte: Autora, 2021.

**Tabela 12:** Valores de diversidade Alfa baseada nos índices de riqueza (A) (Chao1) e diversidade de espécies (B) (Shannon) considerando presença/ausência no método tradicional e eDNA em diferentes parcelas localizadas na S11D.

<b>Método</b>	<b>Parcela</b>	<b>Observed</b>	<b>Chao1</b>	<b>Shannon</b>
eDNA	PCG-168	19	190	2.944439
	PCG-121	31	496	3.433987
	PCG-235	20	210	2.995732
Tradicional	PCG-168	9	45	2.197225
	PCG-121	16	136	2.772589
	PCG-235	14	105	2.639057

**Fonte:** Autora, 2021