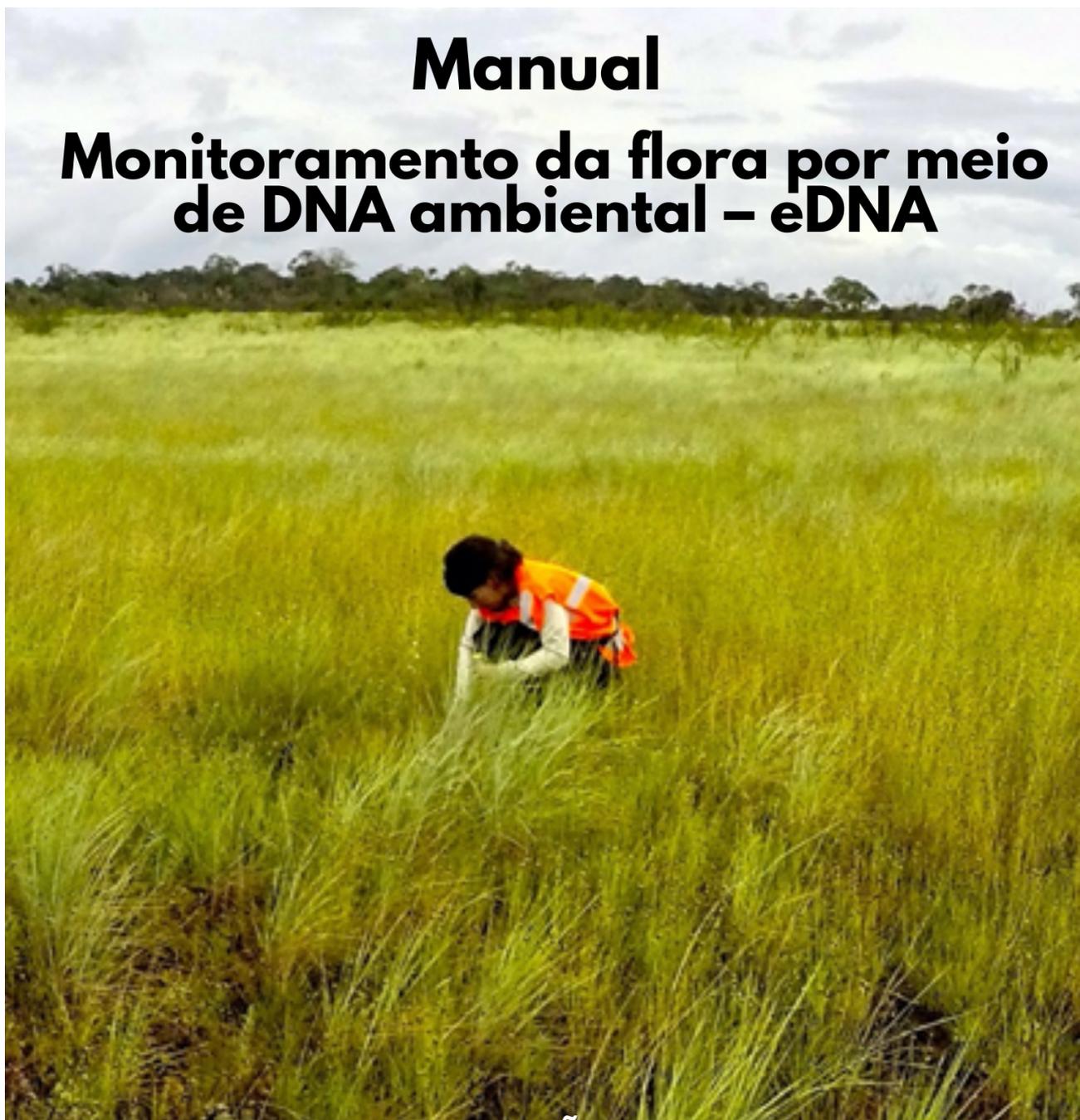




# Manual

## Monitoramento da flora por meio de DNA ambiental – eDNA



Belém / PA  
Março / 2021  
Versão 1



PROD. TEC. ITV DS N023/2020

DOI: 10.29223/PROD.TEC.ITV.DS.N023.2020.Martins

## RELATÓRIO TÉCNICO ITV DS

### MANUAL

## MONITORAMENTO DA FLORA POR MEIO DE DNA AMBIENTAL - eDNA

### RELATÓRIO PARCIAL OU FINAL DO PROJETO

**Vitoria Catarina Cardoso Martins**

**Gisele Lopes Nunes**

**Jamily Lorena**

**Renato Renison Moreira Oliveira**

**José Augusto Pires Bitencourt**

**Eder Pires**

**Mariana Costa Dias**

**Talvâne Glauber Lopes Lima**

**Paulo Henrique de Oliveira Costa**

**Ronnie Alves**

**Santelmo Vasconcelos**

**Guilherme Oliveira**

**Belém / PA  
Março / 2021**

<b>Título:</b> Manual para o monitoramento da flora por meio de DNA ambiental – eDNA	
<b>PROD. TEC. ITV DS-N023/2020</b>	<b>Revisão</b>
<b>Classificação:</b> ( ) Confidencial ( ) Restrita ( ) Uso Interno ( x ) Pública	<b>00</b>

**Informações Confidenciais** - Informações estratégicas para o Instituto e sua Mantenedora. Seu manuseio é restrito a usuários previamente autorizados pelo Gestor da Informação.

**Informações Restritas** - Informação cujo conhecimento, manuseio e controle de acesso devem estar limitados a um grupo restrito de empregados que necessitam utilizá-la para exercer suas atividades profissionais.

**Informações de Uso Interno** - São informações destinadas à utilização interna por empregados e prestadores de serviço

**Informações Públicas** - Informações que podem ser distribuídas ao público externo, o que, usualmente, é feito através dos canais corporativos apropriados

**Citar como:** MARTINS, Vitória Catarina Cardoso; *et al.* **Manual para o monitoramento da flora por meio de DNA ambiental - eDNA.** Belém: ITV, 2021. (Relatório Técnico – N023/2020)

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

M386	<p>Martins, Vitoria Catarina Cardoso.</p> <p>Manual para o monitoramento da flora por meio de DNA ambiental – eDNA / Vitória Catarina Cardoso Martins, Gisele Lopes Nunes, Jamily Lorena, Renato Renison Moreira Oliveira, José Augusto Pires Bitencourt, Eder Pires, Mariana Costa Dias, Talvâne Glauber Lopes Lima, Paulo Henrique de Oliveira Costa, Ronnie Alves, Santelmo Vasconcelos, Guilherme Oliveira -- Belém, 2021.</p> <p>40 p.: il.</p> <p>PROD. TEC. ITV DS-N023/2020 DOI: 10.29223/PROD.TEC.ITV.DS.N023.2020.Martins</p> <p>1. Amazônia. 2. Carajás. 3. Solo. 4. Plantas. 5 DNA <i>metabarcoding</i>. I - Nunes, Gisele Lopes. II- Lorena, Jamily. III - Oliveira, Renato Renison Moreira. IV - Bitencourt, José Augusto Pires. V – Pires, Eder. VI - Dias, Mariana Costa. VII – Lima, Talvâne Glauber Lopes. VIII - Costa, Paulo Henrique de Oliveira. IX – Alves, Ronnie. X. - Vasconcelos, Santelmo. XI - Oliveira, Guilherme. XII - Manual para o monitoramento da flora por meio de DNA ambiental – eDNA</p> <p style="text-align: right;">CDD 23. ed. 581.098115</p>
------	---

## RESUMO EXECUTIVO

A região amazônica é um bioma extremamente vasto e diverso, abrigo de cerca de um quarto das espécies vegetais existentes no planeta. A canga da Serra dos Carajás, no leste da Amazônia, possui um dos maiores depósitos de minérios do planeta e é o lar de uma comunidade de plantas endêmicas e raras. Programas de conservação e monitoramento em áreas megadiversas são muitas vezes dificultados pela falta de conhecimento sobre uma grande parte dos ecossistemas amazônicos, indisponibilidade do profissional taxonomista, dado o grau de especialização necessário para cada grupo taxonômico, lacunas no conhecimento e o tempo para a realização do trabalho, já que a descrição morfológica é ainda uma tarefa demorada e requer várias visitas a uma área de coleta, o que aumenta o tempo e esforço. Devido ao acelerado processo de alterações climáticas e extinção de espécies há a necessidade contínua de desenvolvimento de abordagens aprimoradas de pesquisar a biodiversidade global, e isto inclui aplicar tecnologias baseadas na análise molecular. Nesse sentido, os esforços de especialistas para a construção de um banco de dados genético com códigos de barra de DNA das espécies da Flora de Carajás possibilitaram obter “impressões digitais” das espécies que ali ocorrem. Agora, a partir do banco de dados de referência criado pelo ITV, o ITVBiobase, outras tecnologias moleculares como a de DNA *metabarcoding* pode ser aplicada visando a estratégia de monitoramento das espécies de forma rápida, automatizada e eficiente. A técnica de DNA *metabarcoding* possibilita identificar simultaneamente múltiplas espécies de plantas a partir dos rastros de DNA presente em uma amostra ambiental, como água, sedimento e solo, descartando a necessidade de amostragem direta de espécimes. Mapear a diversidade de plantas em ampla escala e fornecer indicadores da “saúde” do ecossistema dentro de um curto espaço de tempo é de suma importância, tanto para os órgãos regulatórios como para a empresa, uma vez que, levantamentos robustos de biodiversidade são necessários para garantir a proteção das espécies na presença de atividades industriais, incluindo os de mineração. Dessa forma, buscando estratégias rápidas e eficazes, estudos para validação do uso da técnica de DNA *metabarcoding* para monitoramento da biodiversidade da flora em Carajás foram realizados com objetivo de criar um manual padronizado incluindo as etapas desde a coleta em campo até as análises laboratoriais e de bioinformática.

## RESUMO

Os esforços para a construção de um banco de dados genético com códigos de barra de DNA das espécies da Flora de Carajás possibilitaram obter “impressões digitais” das espécies que ali ocorrem. A partir do banco de dados de referência criado pelo ITV, o ITVBiobase, outras tecnologias moleculares como a de DNA *metabarcoding* pode ser aplicada visando o monitoramento das espécies de forma rápida, automatizada e eficiente. A técnica de DNA *metabarcoding* possibilita identificar simultaneamente múltiplas espécies de plantas a partir dos rastros de DNA presente em uma amostra ambiental, descartando a necessidade de amostragem direta de espécimes. Nesse sentido, estudos para validação do uso da técnica de DNA *metabarcoding* para monitoramento da biodiversidade da flora em Carajás, foram realizados com o objetivo de criar um manual padronizado incluindo as etapas desde a coleta em campo até as análises laboratoriais e de bioinformática.

**Palavras-chave:** Amazônia. Carajás. Plantas. DNA Ambiental. Códigos de Barras de DNA.

## ABSTRACT

Efforts to build a genetic database containing DNA barcodes of species of the Carajás flora made it possible to obtain “fingerprints” of local species. From the reference database created by ITV, ITVBiobase, other molecular technologies such as the DNA metabarcoding approach can be applied to monitor species quickly, automatically and efficiently. The DNA metabarcoding technique allows the simultaneous identification of multiple plant species present in the environmental samples through DNA traces buried in the environmental samples (eDNA) without the need to use individual plants as samples. Then, studies to validate the use of the DNA metabarcoding technique for monitoring the biodiversity of the flora of Carajás were carried out to create a standardized manual including the steps from sample collection to laboratory and bioinformatics analysis.

**Keywords:** Amazon. Carajás. Plants. Environmental DNA. DNA Metabarcoding

## Lista de Figuras

<b>Figura 1-</b> Amostragem em campo (A), saco whirl-pak e espátula (B). .....	13
<b>Figura 2-</b> Fluxograma simplificado com as etapas de extração do DNA total das amostras de solo. ....	16
<b>Figura 3-</b> Amostras de solo (A). Extração do DNA (B). Reação de Amplificação (C). Processamento no termociclador (D). ....	19
<b>Figura 4-</b> Fluxograma simplificado mostrando as etapas de construção de Bibliotecas para o sequenciamento em MiSeq.....	21
<b>Figura 5 -</b> Plataforma Miseq System Illumina. ....	22
<b>Figura 6:</b> Fluxograma simplificado mostrando as etapas de análise do <i>pipeline</i> PIMBA: <i>A Pipeline for MetaBarcoding Analysis</i> . ....	24
<b>Figura 7:</b> Exemplos dos arquivos finais gerados pelas etapas de análise do pipeline PIMBA. Esses são os arquivos que podem ser utilizados em análises estatísticas posteriores. ....	29

## Lista de Tabelas

<b>Tabela 1-</b> <i>Primers</i> testados para amplificação e sequenciamento da região ITS de plantas. 16	
<b>Tabela 2-</b> Lista de reagentes e volumes necessários para a preparação da reação da PCR por amostra. .... 18	
<b>Tabela 3-</b> Configurações para o ciclo de amplificação em termociclador (1ª PCR). .... 18	
<b>Tabela 4-</b> <i>Index</i> PCR (2ª PCR). .... 20	

## LISTA DE SIGLAS E ABREVIATURAS

*Amplicon* – Fragmento de DNA ou RNA oriundos da amplificação ou replicação por PCR.

cpDNA – DNA plastidial

CONAMA – Conselho Nacional do Meio Ambiente

DNA – Ácido Desoxirribonucleico

DNAr – DNA ribossomal

eDNA – *environmental DNA* - DNA ambiental

ITS – *Internal Transcribed Spacer* - Espaçador Interno Transcrito

NGS – *Next Generation Sequencing* - Sequenciamento de próxima geração

nDNA – DNA nuclear

nM – nanomolar

OTUs - *Operational Taxonomic Units* - Unidades Taxonômicas Operacionais

*Primer* - sequência curtas utilizada para dar início a síntese de DNA

PCR - *Polymerase Chain Reaction* - Reação em Cadeia da Polimerase

pmol – picomol

TBE – Tris – Borato - EDTA

UV – Radiação Ultravioleta

μL – microlitro

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>11</b>
<b>2</b>	<b>METODOLOGIA.....</b>	<b>12</b>
2.1	COLETA DE AMOSTRA.....	12
2.2	EXTRAÇÃO DO DNA DE AMOSTRAS DE SOLO .....	14
2.3	AMPLIFICAÇÃO DE MARCADORES GENÉTICOS A PARTIR DO DNA TOTAL EXTRAÍDO DE AMOSTRAS DE SOLO.....	16
<b>2.3.1</b>	<b>Seleção dos <i>Primers</i> .....</b>	<b>16</b>
<b>2.3.2</b>	<b>Reação de amplificação por PCR .....</b>	<b>17</b>
2.4	PREPARAÇÃO DA BIBLIOTECA DE SEQUENCIAMENTO .....	19
2.5	SEQUENCIAMENTO.....	22
2.6	ANÁLISE DE BIOINFORMÁTICA DO MATERIAL SEQUENCIADO .....	23
2.7	PASSO A PASSO CONTENDO TODAS AS ETAPAS E COMANDOS A SEREM UTILIZADOS NO SERVIDOR DO ITV .....	24
<b>2.7.1</b>	<b>Acesso ao Cluster.....</b>	<b>24</b>
<b>2.7.2</b>	<b>Execução do pipeline PIMBA.....</b>	<b>25</b>
<b>3</b>	<b>LISTA DE MATERIAIS.....</b>	<b>30</b>
	<b>REFERÊNCIAS .....</b>	<b>32</b>
	<b>ANEXOS .....</b>	<b>36</b>

## 1 INTRODUÇÃO

A Amazônia é a maior floresta tropical do mundo abrigando uma biodiversidade única. Além disso, também possui um dos maiores depósitos de minérios do planeta, fazendo com que esse ecossistema seja não somente importante do ponto de vista ecológico, mas também econômico. Conhecer a biodiversidade é o passo básico para a sua conservação (GIULIETTI et al., 2019). Porém a realização de inventários florísticos na Amazônia é um desafio devido a sua extensão e ampla diversidade. Taxonomistas qualificados estão cada vez mais escassos, realizar estudos florísticos de ambientes complexos estão se tornando cada mais raros e dispendiosos. Neste sentido, novas tecnologias que visam a caracterização de espécies em larga escala vêm sendo bastante utilizadas, dentre elas, o DNA *metabarcoding* (CRISTESCU, 2014; FAHNER et al., 2016; SHOKRALLA et al. 2016; DEINER, BIK et al., 2017). A técnica de DNA *metabarcoding* se baseia no sequenciamento do DNA ambiental (eDNA) presente em amostras como solos, água, sedimentos ou tecidos vegetais/animais para a obtenção de fragmentos específicos de DNA, o que permite identificar as espécies presentes em um ambiente específico (CORLETT 2017, PORTER e HAJIBABAEI, 2018).

A abordagem de DNA *metabarcoding* depende de bancos de dados de códigos de barra de DNA, que são construídos a partir de sequências de DNA com identificação taxonômica previamente determinadas por especialistas, para identificação das espécies (FAZEKAS et al., 2009; KRESS et al., 2017). A disponibilidade desses bancos referenciados é um pré-requisito para a obtenção de resultados robustos e confiáveis. Recentemente, o ITV criou um banco de códigos de barras de DNA curado por taxonomistas (ITVBioBase), possibilitando a aplicação da técnica de DNA *metabarcoding* em estudos de monitoramento da Flora de Carajás (ZAPPI, 2017).

A geração de código de barras do DNA para plantas é feita a partir da amplificação e sequenciamento de um pequeno fragmento do DNA do genoma do cloroplasto (cpDNA) ou nuclear (nDNA) (HEBERT et al., 2003; SHNEER, 2009). É importante ressaltar que o desenvolvimento de códigos de barras de DNA para plantas não é tão simples quanto para outros eucariotos, como animais e fungos (FAZEKAS et al., 2009; CHEN et al., 2016; HEBERT et al., 2016; HOLLINGSWORTH et al., 2016; VASCONCELOS et al., in prep.). Existe uma dificuldade na padronização das regiões ideais do cpDNA (VASCONCELOS et al., in prep.). Diversos autores relatam taxas de sucesso variáveis usando diferentes marcadores moleculares, como *psbK-psbI*, *rpoB*, *rpoC1*, *atpF-atpH*, *trnH-psbA*, *trnL P6 loop* (FAZEKAS et al., 2008; FAHNER et al., 2016; VASCONCELOS et al., in prep.). A combinação das sequências *rbcL* e *matK* têm sido a mais recomendada como *loci* principal (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 DNA ribossômico 35S, também vem sendo sugerido como uma boa opção para identificação de plantas (MOONHOUSE-GANN et al., 2018; TIMPANO, SCHEIBLE e MEIKLEJOHN, 2020), apesar de algumas restrições evolutivas (CHEN et al., 2010; HOLLINGSWORTH et al., 2011; VASCONCELOS et al., in prep.). A região ITS/ITS2 tem sido uma das principais escolhas de marcadores moleculares, uma vez que existe algumas vantagens metodológicas, como: facilidade na padronização das condições de PCR, tamanho de *amplicon* (~ 450 bp), bons níveis de universalidade, qualidade de sequências geradas, taxa de rendimento/custo eficiente e problemas de amplificação reduzidos quando comparados a outros códigos de barras de DNA de plantas comumente utilizados (CHINA et al., 2011; HOLLINGSWORTH, GRAHAM e LITTLE, 2011; HOLLINGSWORTH et al., 2016; VASCONCELOS et al., in prep.). De acordo com Vasconcelos et al. (in prep.), em uma ampla aplicação da técnica, como no caso da criação do banco de referência de códigos de barras de DNA da flora da Serra dos Carajás, um ambiente extremamente complexo devido a elevada diversidade de plantas, o ITS foi um dos marcadores moleculares mais adequados devido a elevada taxa de sucesso. Para Moorhouse-Gann et al. (2018) e Timpano et al. (2020) o ITS também é o principal marcador utilizado quando pretende-se identificar plantas a partir de amostras de solo.

Dessa forma, o DNA *metabarcoding* é uma poderosa ferramenta, que pode ser aplicada tanto em projetos de monitoramento ambiental como em estudos de conservação da biodiversidade, fornecendo um resultado rápido, robusto e confiável. Levantamentos da biodiversidade, incluindo a florística, são uma demanda legal para mineração (Lei 6.938/81 e Resolução 237/97 do CONAMA), portanto, fornecer alternativas que viabilizem estudos de levantamento da biodiversidade em ampla escala é de suma importância tanto para os órgãos regulatórios como para a empresa, auxiliando a elaboração de planos de conservação e manejo da biodiversidade de forma mais eficiente.

## **2 METODOLOGIA**

### **2.1 COLETA DE AMOSTRA**

A principal preocupação durante a coleta de material para estudos genômicos é com a contaminação da amostra, para isso, todo o procedimento deve ser conduzido utilizando luvas descartáveis (preferencialmente de acetonitrila). Para cada local amostrado, as luvas devem ser trocadas, e caso o coletor toque em uma superfície que não seja o alvo da coleta, troque-as imediatamente.

Para coleta, espátulas de metal limpas e autoclavadas são utilizadas (um número suficiente de espátulas deve ser considerado por coleta) (Figura 1B). Em campo, a assepsia das espátulas deve ser feita utilizando álcool 70% e, quando possível, flambadas ao fogo, tomando sempre cuidado para não provocar danos ao ambiente ou ao coletor (Figura 1A). O solo deve ser armazenado em *containers* estéreis, sendo a recomendação usar sacos de poliuretano de alta densidade (HDPE), selados e esterilizados por radiação gama, como os do tipo whirl-pak (Figura 1B). A identificação deve ser feita previamente e com números únicos, para evitar confusão ou troca de amostras. O número pode ser escrito ou colado no saco com etiquetas próprias para este fim. Caso opte pelas etiquetas, imprimir previamente cópias em triplicata para colar no saco de amostragem, caderno de anotações e uma etiqueta extra para quaisquer eventualidades. Independente da escolha, os números de identificação devem ser anotados/colados no caderno de anotações juntamente com os metadados obtidos na coleta. A coleta deve ser realizada sempre em triplicata.

O transporte das amostras até o laboratório deverá ser feito o mais rápido possível em um *container* termo estável ou isopor. Recomenda-se a temperatura de  $-20^{\circ}\text{C}$ , entretanto, devido as condições de logística e as adversidades do campo, pode-se seguir as recomendações do Guia Nacional de Coleta e Preservação de Amostras(<http://arquivos.ana.gov.br/institucional/sge/CEDOC/Catalogo/2012/GuiaNacionalDeColeta.pdf>), que preconiza o transporte a  $4^{\circ}\text{C}$ . Seguir as devidas recomendações são importantes para manutenção da qualidade da amostra, visando reduzir a degradação do material genético alvo. No campo, uma alternativa é usar sacolas térmicas com gelo químico para o armazenamento da amostra. O uso de gelo deve ser feito com cuidado, devido a possível contaminação das amostras pelo degelo. Ao chegar no laboratório, o material deve imediatamente armazenado em ultra freezer a  $-80^{\circ}\text{C}$ .

**Figura 1** - Amostragem em campo (A), saco whirl-pak e espátula (B).



Fonte: elaborado pelos autores (2021).

Os metadados são informações extras e referentes a cada amostra. A coleta dessas informações é uma etapa obrigatória que tem como objetivo organizar e identificar os parâmetros de amostragem. Em campo, os metadados obrigatórios são basicamente: coordenadas geográficas, código da amostra, data da coleta e coletor. Outros fatores podem ser considerados, isso vai depender do tipo do material a ser coletado. Tomar o devido cuidado para anotar os metadados ao código associado à amostra.

Recomenda-se uma preparação prévia de etiquetas adesivas para identificação das amostras a serem coletadas e para o caderno de coleta. No caderno de coleta são coladas as etiquetas associadas e os metadados associados. Fotos do local de coleta e dos espécimes são também desejáveis, sendo que a mesma câmera utilizada para as fotos pode ser utilizada para bater fotos das anotações associadas, assim facilita a conexão das imagens aos dados da amostra. A planilha com os metadados deve ser digitalizada preferencialmente no mesmo dia da coleta.

## 2.2 EXTRAÇÃO DO DNA DE AMOSTRAS DE SOLO

A etapa de extração de DNA deve ser feita em laboratório, devidamente limpo e equipado. O kit sugerido para extração de DNA de solos é o Kit DNeasy PowerSoil Kit (Anexo A) (QIAGEN, Hilden, Germany). Segundo a experiência do Grupo de Genômica Ambiental do ITV com extração de DNA de micro-organismos a partir de amostras de solo, os resultados gerados pelo kit acima são bastante satisfatórios, reduzindo a ação de inibidores naturais do solo que podem causar prejuízos a qualidade do DNA a ser extraído.

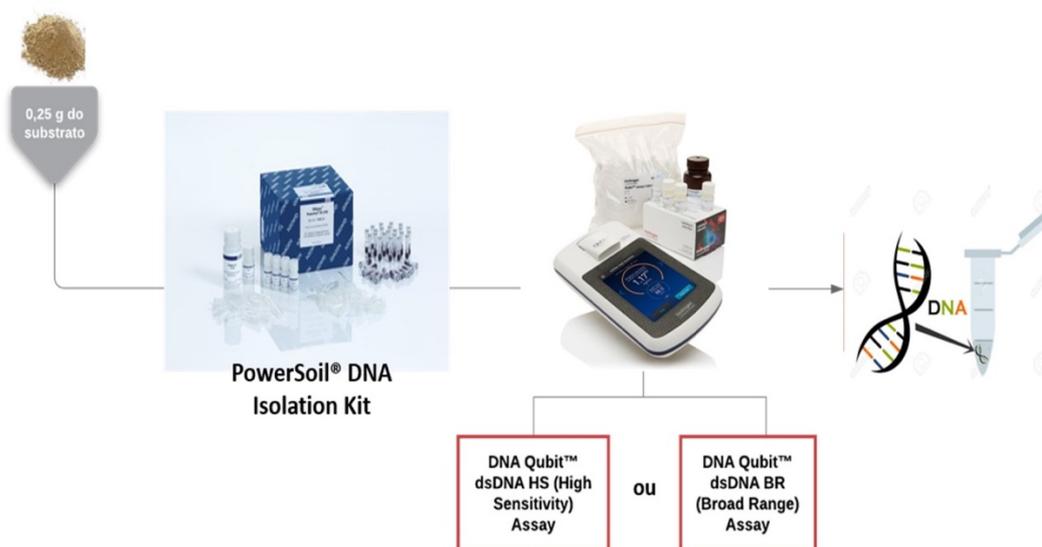
Para evitar a contaminação, todo o processo de extração deve ser executado em capela de fluxo laminar, previamente limpa com hipoclorito de sódio a 2% e etanol 70%, e posteriormente, esterilizada com luz UV. Além disso, controles negativos devem ser preparados a cada nova extração. Os controles negativos são feitos utilizando água no lugar no DNA, seguindo os mesmos procedimentos e kits utilizados para as amostras de DNA. Recomenda-se fazer extrações de DNA em triplicatas, mas isso depende da questão a ser respondida e da necessidade do projeto. A extração do DNA ambiental (eDNA) deverá ser feita usando o Kit DNeasy PowerSoil (QIAGEN, Hilden, Germany), conforme as instruções abaixo:

1. Identificar os tubos com o nome das amostras;
2. Pesar 0,25 g de solo de cada amostra e adicionar nos tubos identificados;
3. Adicionar 60 µL de solução C1 e manualmente invertê-los por 10 vezes;

4. Vortexar por 10 min à 3000 rpm. Recomenda-se adaptar o mix para vortexar horizontalmente;
5. Posteriormente, os tubos devem ser centrifugados a 14.000 rpm por 30 seg. Transferir um volume de 500 µL do sobrenadante para um novo microtubo, proporcionado pelo kit;
6. Adicionar 250 µL de solução C2, agitar no vortex (Mix Mate) por 5 seg e imediatamente incubar a 4°C por 5 min;
7. Centrifugar por 1 min a 14.000 rpm e transferir 600 µL de sobrenadante para um novo microtubo proporcionado pelo kit;
8. Adicionar 200 µL de solução C3 e agitar no vortex (Mix Mate) por 5 seg e incubar novamente a 4°C por 5 min. Centrifugar os microtubos a 14.000 rpm por um min e transferir 750 µL de sobrenadante para um novo microtubo. Posteriormente, adicionar 1200 µL da solução C4 e agitar no vortex (Mix Mate) por 5 seg;
9. Em seguida, adicionar 675 µL da solução em um novo microtubo contendo uma coluna de sílica acoplada e centrifugar por 1 min a 14.000 rpm. Esse passo é repetido por três vezes. Logo após, adicionar 500 µL de solução C5 e centrifugar por 30 seg a 14.000 rpm.
10. Centrifugar o microtubo vazio novamente por 30 seg a 14.000 rpm. Em seguida, transferir a coluna de sílica para um novo microtubo e adicionar 100 µL de solução C6 (10 mM Tris- HCl pH 8.5) fornecido pelo próprio kit no centro da coluna para fazer a eluição do DNA. A solução C6 não contém EDTA Centrifugar por 30 seg a 14.000 rpm.
11. Descartar a coluna de sílica e armazenar o tubo contendo o DNA em uma temperatura de -20°C a - 80°C, para garantir um maior tempo de viabilidade para as análises. O produto da extração deve ser colocado em microtubos Eppendorf Safe-Lock 1,5 µL.

Ao final, quantificar o DNA no fluorômetro Qubit® 3.0 (Thermo Fisher Scientific) com o kit Qubit™ dsDNA HS (High Sensitivity) Assay (Thermo Fisher Scientific) (Anexo B, Anexo C) caso a concentração de DNA esteja entre 10 pg/µL à 100 ng/µL. Caso a concentração esteja entre 100 pg/µL à 1,000 ng/µL utilizar o kit Qubit™ dsDNA BR (Broad Range) Assay (Thermo Fisher Scientific) (Anexo D) (Figura 2).

**Figura 2** - Fluxograma simplificado com as etapas de extração do DNA total das amostras de solo.



**Fonte:** elaborado pelos autores<sup>1</sup> (2021).

## 2.3 AMPLIFICAÇÃO DE MARCADORES GENÉTICOS A PARTIR DO DNA TOTAL EXTRAÍDO DE AMOSTRAS DE SOLO

### 2.3.1 Seleção dos *Primers*

Para a adequação do protocolo de sequenciamento nas plataformas NGS (Ion Torrent e Illumina MiSeq), diferentes pares de *primers* foram testados para a amplificação da região ITS (Tabela 1). Após a análise de eficiência de diferentes combinações de *primers*, o par ITS2F-ITS3R (CHEN et al., 2010; CHIOU et al., 2007) apresentou melhores resultados de amplificação para as amostras da canga.

**Tabela 1** - *Primers* testados para amplificação e sequenciamento da região ITS de plantas.

Região	Primer	Sequência (5'-3')	Autor
ITS2	ITS2F	ATGCGATACTTGGTGTGAAT	Chen et al. (2010)
ITS2	ITS3R	GACGCTTCTCCAGACTACAAT	Chiou et al. (2007)
ITS	ITS1	TCCGTAGGTGAACCTGCGG	White et al. (1990)
ITS	ITS1F	CTTGATCATTAGAGGAAGTAA	Gardes e Bruns (1993)
ITS	ITS4	TCCTCCGCTTATTGATATGC	White et al. (1990)
ITS	ITS4-R	CAGGAAACAGCTATGACTCCTCCGCTTAT TGATATGC	Gardes e Bruns (1993)

**Fonte:** elaborado pelos autores (2021).

Os *primers* podem ser sintetizados pela empresa IDT (Integrated DNA technologies) em escala de síntese padrão.

<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/>).

- Preparo da solução de estoque (*Primer*)

Antes de preparar a solução estoque deve-se centrifugar o tubo do *primer* liofilizado por 30 seg. Para a solução estoque, a quantidade de número de mols é multiplicada por 10. O valor da multiplicação é a quantidade em microlitros de água ultrapura que precisa ser adicionada ao *primer* liofilizado. Recomenda-se vortexar por 30 seg a 1 min para ter certeza que o *primer* está bem dissolvido.

- Preparo da solução de uso (*Primer*)

Para preparar a solução que vai ser adicionada na reação em cadeia da polimerase (PCR) é necessário diluir para 10 micromolar. Para tanto, utiliza-se 10  $\mu\text{L}$  da solução estoque em 90  $\mu\text{L}$  de água ultrapura.

### **2.3.2 Reação de amplificação por PCR**

As reações de amplificação devem ser realizadas em triplicata para cada amostra e controles negativos devem ser utilizados toda vez que for realizado um conjunto de reação em cadeia da polimerase (PCR) (Figura 3). Outra recomendação importante é não utilizar a pipeta multicanal para pipetar o DNA, pipetar sempre as amostras separadamente.

A preparação da reação da PCR (MIX), mistura que contém todos os reagentes necessários à síntese de novas cópias de DNA, deve ser feita de acordo com a

Tabela 2. O volume dos reagentes deve ser multiplicado pelo número de amostras +1 durante o preparo do MIX. Com base nos testes realizados no ITV é recomendado não utilizar o 5x Green GoTaq™ Flexi Buffer do kit Promega GoTaq™ G2 Flexi DNA Polymerase. Esse reagente vem com o tampão de carregamento e permite análise direta em gel após a reação de PCR. Para evitar a contaminação do estoque, separar em alíquotas o reagente dimethyl sulphoxide (DMSO), adquirido pela empresa HIMEDIA, em tubos de eppendorf. O reagente TBT-PAR (solução 5x) é preparado no laboratório conforme o protocolo do Anexo E.

**Tabela 2** - Lista de reagentes e volumes necessários para a preparação da reação da PCR por amostra.

Reagentes	$\mu\text{L}$
5x Buffer	2,5
25 mM $\text{MgCl}_2$	1,2
2 mM dNTP	1
TBT	2,5
DMSO	1
ITS 2F 10pmol ( <i>primer</i> F)	0,25
ITS 3R 10pmol ( <i>primer</i> R)	0,25
Taq	0,1
DNA	2
$\text{H}_2\text{O}$ ultrapura	1,7
Volume final (por amostra)	12,5

**Fonte:** elaborado pelos autores (2021).

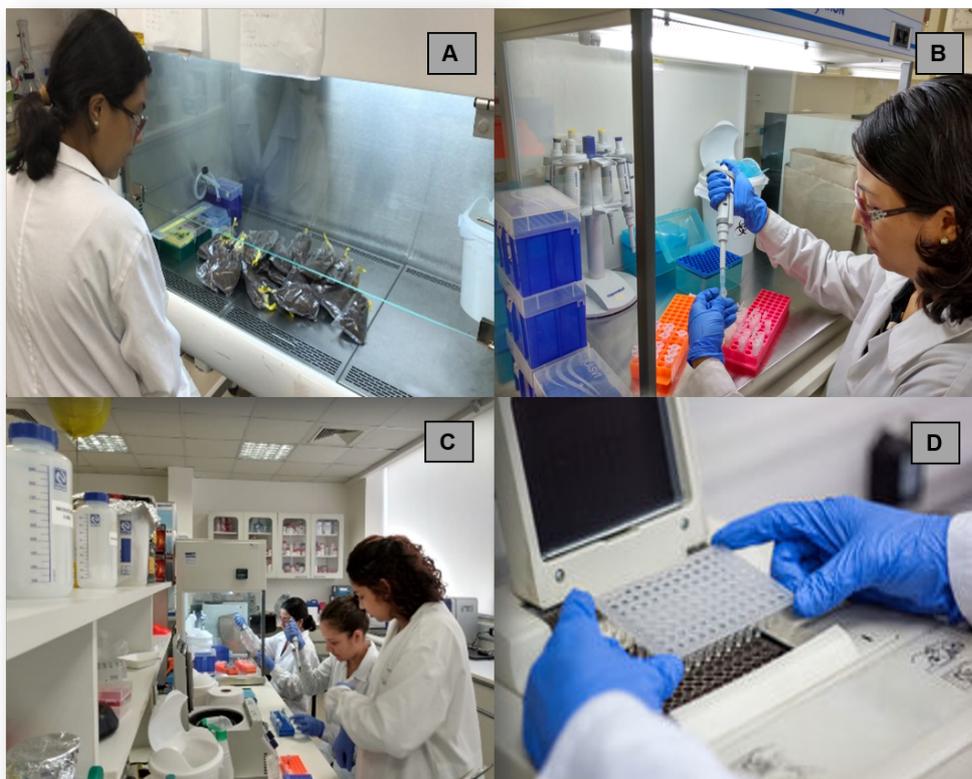
Posteriormente, as condições de amplificação devem ser programadas no termociclador. Para esse conjunto de *primers* as condições ideais são as seguintes: desnaturação inicial a  $94^\circ\text{C}$  por 3 min; seguida por 30 ciclos de  $94^\circ\text{C}$  por 1 min (desnaturação do DNA),  $54^\circ\text{C}$  por 1 min (anelamento dos *primers*) e 1 min à  $72^\circ\text{C}$  (extensão da fita de DNA); e extensão final por 7 min a  $72^\circ\text{C}$  (Tabela 3).

**Tabela 3** - Configurações para o ciclo de amplificação em termociclador (1ª PCR).

Desnaturação inicial	30 ciclos			Extensão final
	Desnaturação	Anelamento	Extensão	
3 min	1 min	1min	1min	7 min
$94^\circ\text{C}$	$94^\circ\text{C}$	$54^\circ\text{C}$	$72^\circ\text{C}$	$72^\circ\text{C}$

**Fonte:** elaborado pelos autores (2021).

**Figura 3** - Amostras de solo (A). Extração do DNA (B). Reação de Amplificação (C). Processamento no termociclador (D).



Fonte: elaborado pelos autores (2021).

## 2.4 PREPARAÇÃO DA BIBLIOTECA DE SEQUENCIAMENTO

Esta etapa é responsável pela inclusão de adaptadores de sequências únicas Illumina para o sequenciamento dos fragmentos de DNA amplificados. Recomendamos que a construção das bibliotecas seja realizada de acordo com o protocolo 16S Metagenomic Sequencing Library Preparation da Illumina (Illumina) (Anexo F).

Inicialmente, todas as triplicatas devem ser unidas antes da primeira purificação. Os procedimentos da primeira purificação do DNA devem ser feitos a partir do produto da PCR (*amplicons*) utilizando o kit de esferas magnéticas Agencourt Ampure XP (Bechman Coulter), seguindo as instruções do protocolo anteriormente mencionado. Para avaliar a qualidade dos *amplicons*, as amostras devem ser submetidas à eletroforese em gel de agarose 1%. Para preparo do gel, 1g de agarose é adicionado a 100 mL de TBE 1X e a solução dissolvida em micro-ondas por 2 min. Posteriormente, o corante fluorescente Syber Safe Thermo Fisher Scientific (10  $\mu$ L para cada 100 mL de gel de agarose) é adicionado ao gel e a mistura adicionada na cuba de eletroforese para solidificação (aproximadamente 30 min). Para preparo da amostra, 5  $\mu$ L de produto de PCR, 1  $\mu$ L de DNA-ladder 50 pb (Invitrogen<sup>TM</sup>) e 1,0  $\mu$ L de corante azul de bromofenol devem ser misturados. Assim que o gel estiver solidificado, inserir

o tampão TBE 1X ou 0,5X na cuba de eletroforese juntamente com o gel e aplicar as amostras nos poços com uma ajuda de uma pipeta. Iniciar a corrida utilizando uma corrente elétrica de 90 V por 40 min. Para uma verificação mais precisa dos tamanhos dos *amplicons*, utiliza-se o Bioanalyzer Agilent Technology 2100 (PANARO et al. 2000), de acordo com instruções do kit Agilent DNA 1000 (Agilent Technologies).

Após a verificação dos *amplicons*, adaptadores de sequências únicas (*indexes*) devem ser adicionados a cada amostra através da etapa de *Index* utilizando o kit Nextera XT Library prep (Illumina, San Diego, CA, USA), conforme instruções do protocolo (16S Metagenomic Sequencing Library Preparation da Illumina). Os *indexes* utilizados na montagem da biblioteca não devem ser repetidos, uma vez que eles são utilizados para identificar as amostras na etapa de bioinformática. Em seguida, uma nova PCR é feita utilizando 5 µL de DNA, 5µL de Nextera XT *index Primer*1, 5 µL de Nextera XT *index Primer* 2, 25,0 µL de enzima 2x Kapa Hifi HotStart Ready mix e 10 µL de água ultrapura, para um volume final de 50 µL. A etapa de *Index* PCR deve ser feita de acordo com as condições descritas na Tabela 4.

**Tabela 4 - *Index* PCR (2ª PCR).**

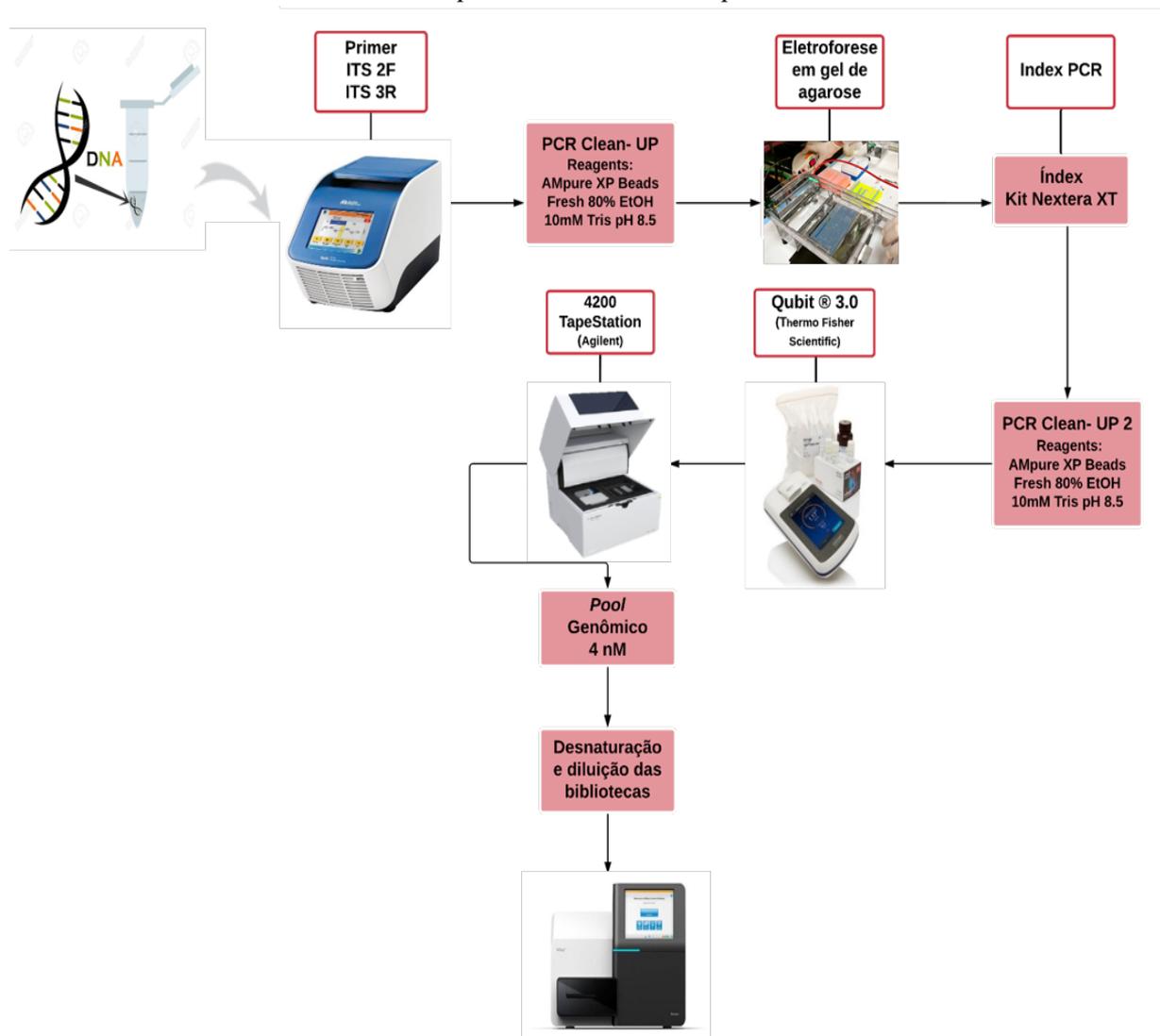
Index PCR (8 ciclos)				
3 min	30 s	30 s	30 s	5 min
95°C	95°C	55°C	72°C	72°C

**Fonte:** elaborado pelos autores (2021).

O produto da *Index* PCR deve ser purificado novamente utilizando o kit de esferas magnéticas Agencourt Ampure XP (Bechman Coulter), conforme instruções do protocolo 16S (Illumina). As bibliotecas devem ser quantificadas no fluorômetro Qubit® 3.0 (Thermo Fisher Scientific) e o tamanho do *amplicon* verificado no TapeStation System 4200 (Agilent). Para a escolha do kit Agilent TapeStation DNA ScreenTape e Reagentes (d1000, d5000 ou High Sensitive) deve-se primeiro verificar o tamanho do fragmento para depois seguir o protocolo do fabricante. O *pool* genômico (produto do *Index* PCR purificado) deve ser padronizado para concentração de 4 nM utilizando a tabela *conversion calculator* (solicitar a tabela a Eder Pires, pelo email [eder.pires@itv.org](mailto:eder.pires@itv.org)), que deve ser preenchida de acordo com a quantidade de DNA quantificado e o tamanho do *amplicon* obtido. Este processo equaliza a quantidade de DNA em cada biblioteca e certifica uma representação igualitária entre as amostras. Logo após, o *pool* genômico é agrupado em um único tubo de acordo com o protocolo 16S Metagenomic Sequencing Library Preparation da Illumina (Illumina). Se necessário, ao *pool* de bibliotecas

poderá ser adicionado PhiX, uma biblioteca padrão da Illumina utilizada para aumentar a diversidade dos fragmentos e permitir um bom sequenciamento dos *amplicons* (ILLUMINA, 2016). A corrida de sequenciamento deve ser realizada utilizando o kit de corrida MiSeq V3 600 ciclos (Figura 4).

**Figura 4** - Fluxograma simplificado mostrando as etapas de construção de Bibliotecas para o sequenciamento em MiSeq.



Fonte: elaborado pelos autores<sup>2</sup> (2021).

<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/>).

## 2.5 SEQUENCIAMENTO

O sequenciamento deve ser realizado na plataforma MiSeq System da Illumina (2 x 300bp) com capacidade de leitura de até 600 pb (Figura 5). A geração de dados varia de 300 Mb a 15 Gb, tempo de máquina 4 - 40 horas (sequenciamento e análise) e tempo de preparação da amostra dura 2 dias, considerando uma jornada de 8 horas.

**Figura 5** - Plataforma MiSeq System Illumina.



**Fonte:** MiSeq System Guide (2019).

Passo a passo da execução do sequenciamento acordo com o protocolo Guia do Sistema MiSeq (Anexo G):

- 1° Passo: Preparar o cartucho de reagente pré-carregado para uso.
- 2° Passo: Desnaturar e diluir as bibliotecas.
- 3° Passo: Colocar a mistura da biblioteca no cartucho de reagente no reservatório designado.
- 4° Passo: Na interface do *software*, selecione *Sequence* (Sequenciamento) para iniciar as etapas de configuração da execução ou conecte-se ao BaseSpace ou ao BaseSpace Onsite.
- 5° Passo: Limpar e secar cuidadosamente a lâmina de fluxo. Carregue a lâmina de fluxo.
- 6° Passo: Carregar o frasco de PR2 e verificar se o frasco de resíduos está vazio. Carregue o cartucho de reagente.
- 7° Passo: Revisar os parâmetros da execução e os resultados da verificação pré-execução. Selecionar *Start Run* (Iniciar execução).
- 8° Passo: Monitorar a execução pela interface do MCS ou por outro computador usando o *Sequencing Analysis Viewer* (SAV).

9º Passo: Realizar uma limpeza após a execução.

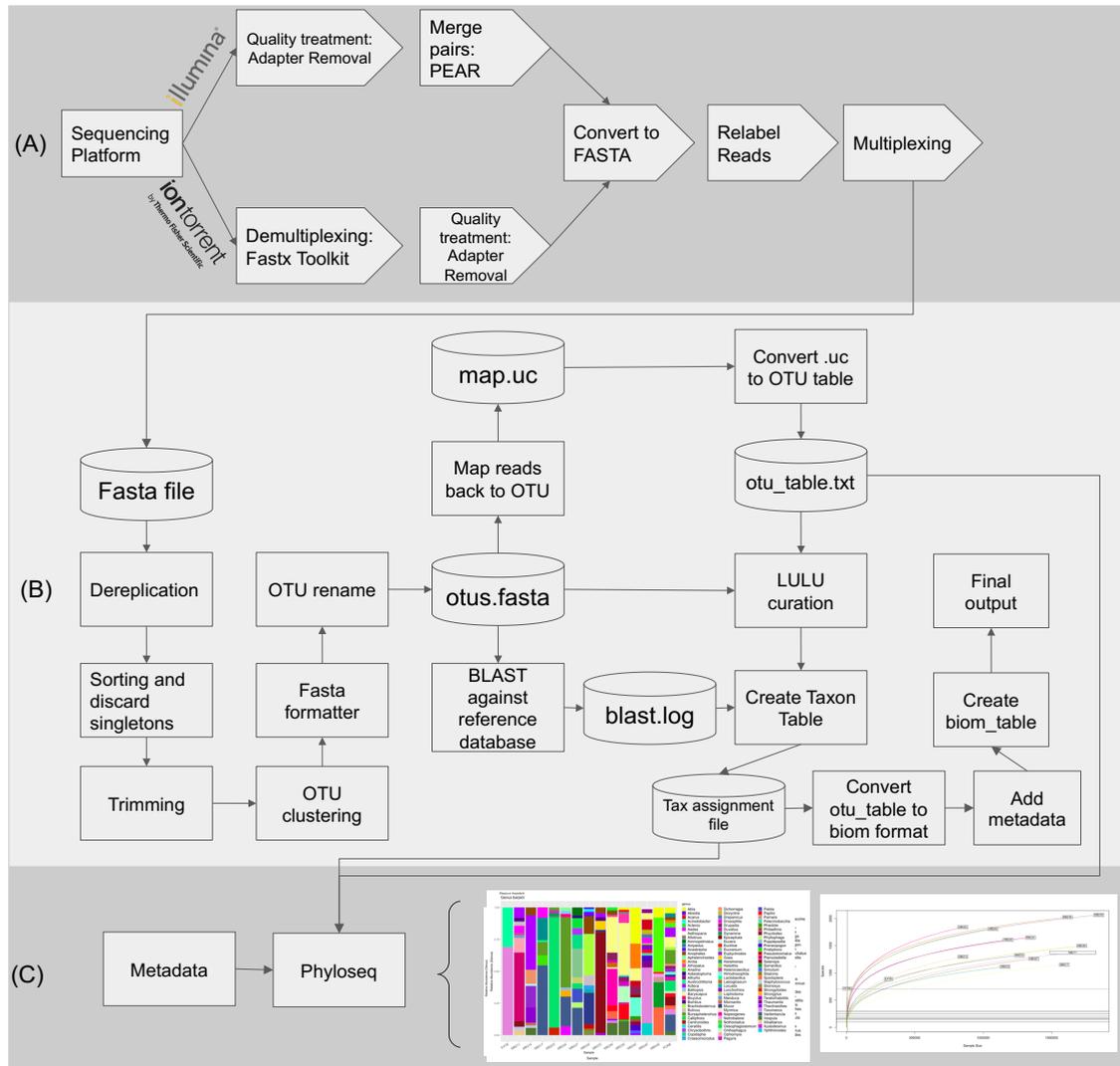
## 2.6 ANÁLISE DE BIOINFORMÁTICA DO MATERIAL SEQUENCIADO

As sequências brutas geradas pelo sequenciamento devem ser submetidas ao PIMBA (*A Pipeline for MetaBarcoding Analysis*) (OLIVEIRA et al., 2021) um *pipeline* voltado para análise de dados de DNA *metabarcoding* que permite a utilização de bancos de dados taxonômicos distintos. A escolha do banco de dados é uma etapa muito importante e depende do alvo amostral (plantas, animais, microrganismos, etc.) e marcador utilizado (ITS, *rbcL*, COI, 16S DNAr etc.). O PIMBA foi desenvolvido pelo grupo de bioinformática do Instituto Tecnológico Vale – Desenvolvimento Sustentável para análise de metabarcodes de plantas e tem como base o *pipeline* QIIME (*Quantitative Insights Into Microbial Ecology*) (CAPORASO et al., 2010, BOYLEN et al., 2019), um *pipeline* referência desenvolvido para estudos de microbioma. Este *pipeline* é composto por várias etapas (tratamento e qualidade das sequências, classificação taxonômica, reconstrução filogenética, análise de diversidade) que foram adaptadas para a análise de DNA *metabarcoding* de plantas (Figura 6). O download do *pipeline* e manual de uso (Anexo H) está disponível no GitHub (<https://github.com/reinator/pimba>). O PIMBA funciona nos sistemas operacionais: LINUX e macOS.

Resumidamente, o PIMBA realiza a etapa de trimagem (remoção de adaptadores) e filtragem por qualidade (Phred >20) utilizando a ferramenta Adapter Removal (SCHUBERT, LINDGREEN & ORLANDO, 2016). Posteriormente, somente sequências *forward* e *reverse* de alta qualidade serão montadas através do montador Pear (ZHANG et al., 2014). Após a montagem, ocorre a etapa de derreplicação para a remoção das sequências duplicadas e em seguida a exclusão de sequências com cópia única. Com intuito de melhorar a qualidade dos dados, as sequências menores que 100 pb são descartadas (isso depende do tamanho do marcador molecular utilizado). Em seguida, as sequências são agrupadas em Unidades Taxonômicas Operacionais (OTUs) utilizando o VSEARCH (ROGNES et al., 2016). O agrupamento das sequências OTUs é usada para classificar grupos de indivíduos com similaridade > 97% (essa percentagem depende do grupo de organismo a ser trabalhado ou marcador molecular utilizado). Finalmente as OTUs são comparadas com as sequências disponíveis em bancos de dados públicos como o NCBI (<https://www.ncbi.nlm.nih.gov>), ou bancos específicos como no caso do ITVBioBase, um banco de dados que contém sequências de plantas (*matK*, *rbcL*, *rpoB*, and *rpoC1*, *atpF-atpH*, *psbK-psbI* and *trnH-psbA* e ITS2), invertebrados (COI), morcegos (COI) e peixes (12S e 16S DNAr) que ocorrem em Carajás.

Esse banco foi devidamente construído e curado por especialistas com foco na área de atuação da Vale.

**Figura 6:** Fluxograma simplificado mostrando as etapas de análise do *pipeline* PIMBA: *A Pipeline for MetaBarcoding Analysis.*



Fonte: Oliveira *et al.* (2021).

## 2.7 PASSO A PASSO CONTENDO TODAS AS ETAPAS E COMANDOS A SEREM UTILIZADOS NO SERVIDOR DO ITV

### 2.7.1 Acesso ao Cluster.

O Acesso ao cluster do ITV acontece via *ssh* utilizando o terminal disponível em sistema Linux, macOS ou Windows Power Shell (Windows 10). Todos os programas a serem utilizados já estão instalados nos servidores do ITV e devidamente atualizados. A liberação do acesso ao cluster deve ser solicitada a Guilherme Oliveira pelo email

[guilherme.oliveira@itv.org](mailto:guilherme.oliveira@itv.org). Uma vez liberado, um **usuário** e **senha** serão disponibilizados e estes serão utilizados a seguir

1. Entrar no terminal

```
ssh <usuário>@uv3000
```

2. Criar uma pasta com seu projeto dentro da sua pasta principal

```
mkdir <nome_projeto>
```

3. Entrar na pasta criada

```
cd <nome_projeto>
```

4. Copiar para sua pasta <nome\_projeto> os arquivos “pbs” que estão dentro da pasta **pimba\_prepare** (/illumina ou /ionTorrent), **pimba\_run** e **pimba\_plot** que estão alocados no endereço /bio/pimba\_metabarcoding/docker/examples/

```
cp /bio/pimba_metabarcoding/docker/examples/pimba_prepare/illumina/job_prepare.pbs .
```

```
cp /bio/pimba_metabarcoding/docker/examples/pimba_run/job_pimba.pbs .
```

```
cp /bio/pimba_metabarcoding/docker/examples/pimba_run/databases.txt .
```

```
cp bio/pimba_metabarcoding/docker/examples/pimba_plot/job_pimbaplot.pbs
```

```
cp /bio/pimba_metabarcoding/examples/pimba_prepareData/illumina/adapters.txt
```

OBS:(configurar o arquivo adapters.txt de acordo com seus marcadores)

5. Criar uma pasta chamada rawdata para alocar os dados brutos

```
mkdir rawdata
```

```
chmod 777 rawdata (dar permissão de acesso a pasta)
```

OBS: Os dados do sequenciamento que serão copiados para a pasta rawdata devem ser solicitados a Gisele Nunes ([gisele.nunes@itv.org](mailto:gisele.nunes@itv.org)) ou Eder Pires ([eder.pires@itv.org](mailto:eder.pires@itv.org)).

## 2.7.2 Execução do pipeline PIMBA

Assim que os dados do sequenciamento estiverem disponíveis na pasta rawdata execute os comandos dos 3 passos abaixo. Lembrar que para executar o comando, o usuário deve estar dentro da pasta <nome\_projeto>.

Os comandos executados no cluster ITV devem ser via PBS chamado pela função *qsub*, na fila 2 (-q) e arquivo contendo os scripts desejados (*arquivo.pbs*)

### **Passo 1: Prepare seus dados (pimba\_prepare.sh)**

A primeira etapa executa os *scripts* de tratamento de qualidade como controle de qualidade e *merge* das sequências caso seja *paired-end*.

```
qsub -q op2 job_prepare.pbs
```

Para ver se o *job* está rodando ou se já finalizou utilize o comando abaixo:

```
qstat -a
```

OBS: caso queira editar os parâmetros de tratamento de qualidade deve-se editar o arquivo:

```
job_prepare.pbs.
```

Abra o arquivo no terminal usando o comando vim. Para editar apertar *i*, para sair e salvar apertar *Esc*, digitar *:wq* e dar *Enter*. Para sair sem salvar apertar *Esc*, digitar *:q* e dar *Enter*

```
vim job_prepare.pbs
```

O comando padrão é o seguinte:

```
/bio/pimba_metabarcoding/docker/pimba_prepare.sh      illumina      <rawdata_dir>  
<output_reads> <num_threads> <adapters.txt>
```

Parâmetros:

<rawdata\_dir> = caminho dos arquivos R1 e R2 *reads*;

<output\_reads> = nome da pasta dos resultados;

<num\_threads> = número de *threads*;

<adapters.txt> = arquivo contendo os adaptadores usados no sequenciamento.

### **EXEMPLO:**

```
/bio/pimba_metabarcoding/docker/pimba_prepare.sh illumina rawdata_illumina/ AllSamples  
24 adapters.txt
```

### **Passo 2: Análise de clusterização (OTU) e atribuição taxonômica (pimba\_run.sh)**

Etapa de agrupamento das OTUs e atribuição taxonômica. O arquivo *fasta* gerado na etapa anterior é mandatório na etapa 2. Verificar se o arquivo foi gerado.

```
qsub -q op2 job_pimba.pbs
```

Para ver se o *job* está rodando ou se já finalizou utilize o comando abaixo:

```
qstat -a
```

OBS: caso queira editar os parâmetros de similaridade para agrupamento das OTUs, nome da pasta contendo os resultados ou bancos de dados a ser utilizado para classificação taxonômica das sequências, deve-se editar o arquivo:

**job\_pimba.pbs**

Abra o arquivo no terminal usando o comando vim. Para editar apertar *i*, para sair e salvar apertar *Esc*, digitar *:wq* e dar *Enter*. Para sair sem salvar apertar *Esc*, digitar *:q* e dar *Enter*

**vim job\_pimba.pbs**

O comando padrão é o seguinte:

```
/bio/pimba_metabarcoding/docker/pimba_run.sh -i <input_reads> -o <output_dir> -s <otu_similarity> -a <assign_similarity> -c <coverage> -l <otu_length> -h <hits_per_subject> -g <marker_gene> -t <num_threads> -e <E-value> -d <databases.txt>
```

Parâmetros:

-i <input_reads> =	arquivo FASTA contendo as <i>reads</i> geradas pelo <i>pimba_prepare.sh</i>
-o <output_dir> =	diretório com os resultados
-s <otu_similarity> =	porcentagem de similaridade utilizada para clusterização das OTUs. <i>Default</i> é 0.97
-a <assign_similarity> =	porcentagem de similaridade utilizada para atribuição taxonômica. <i>Default</i> é 0.9
-c <coverage> =	cobertura mínima para alinhamento. <i>Default</i> é 0.9
-l <otu_length> =	comprimento da trimagem das reads. Se 0, <i>reads</i> são trimadas
-h <hits_per_subject> =	Se 1, escolhe o melhor <i>best hit</i> . If > 1, escolhe pela maioria. <i>Default</i> é 1
-g <marker_gene> =	gene marcador e database. Pode ser: (16S-SILVA, 16S-GREENGENES, 16S-RDP, 16S-NCBI, ITS-FUNGI-NCBI, ITS-FUNGI-UNITE, ITS-PLANTS-NCBI, ITS-PLANTS-BIOBASE, COI-NCBI, 12S-BIOBASE).
-t <num_threads> =	número de <i>threads</i> usado para o <i>blast</i> . <i>Default</i> é 1
-e <e-value> =	usado para <i>blast</i> . <i>Default</i> é 10
-d <databases_file.txt> =	Arquivo contendo o caminho da base de dados

### EXEMPLO:

```
/bio/pimba_metabarcoding/docker/pimba_run.sh -i AllSamples.fasta -o ITS-BioBase -s 0.97 -a 0.7 -c 0.8 -l 100 -h 1 -g ITS-PLANTS-BIOBASE -t 24 -e 0.1 -d databases.txt
```

### **Passo 3: Plot os resultados (pimba\_plot.sh)**

Quando terminar o *pimba\_run.sh*, você será capaz de gerar alguns gráficos básicos para seus resultados, como PCoA, curvas de rarefação, gráficos de diversidade alfa e beta. Tudo o

que você precisa são dois arquivos que o `pimba_run.sh` irá gerar (`otu_table.txt` e `tax_assignment`) e um arquivo de metadados (tabela em formato csv) que você terá que fornecer (`metadata.csv`).

```
qsub -q op2 job_pimbaplot.pbs
```

Para ver se o *job* está rodando e se já finalizou utilize o comando abaixo:

```
qstat -a
```

OBS: Caso queira editar, abra o arquivo no terminal usando o comando `vim`. Para editar apertar *i*, para sair e salvar apertar *Esc*, digitar `:wq` e dar *Enter*. Para sair sem salvar apertar *Esc*, digitar `:q` e dar *Enter*

```
vim job_pimbaplot.pbs
```

O comando padrão é o seguinte:

```
/bio/pimba_metabarcoding/docker/pimba_plot.sh -t <otu_table> -a <tax_assignment> -m <metadata> -g <group_by>
```

Parâmetros:

<code>-t &lt;otu_table&gt; =</code>	tabela OTU gerada por <code>pimba_run</code>
<code>-a &lt;tax_assignment&gt; =</code>	arquivo de atribuição taxonômica gerado por <code>pimba_run</code>
<code>-m &lt;metadata&gt; =</code>	arquivo CSV com colunas "sample_id" (primeira coluna contendo os IDs das amostras) e outros atributos relacionados a cada amostra (local, estação do ano etc.)
<code>-g &lt;groupby&gt; =</code>	indica a coluna que deseja agrupar os resultados. Por exemplo, local. Se não quiser agrupar os resultados, não o especifique.

### **EXEMPLO:**

```
/bio/pimba_metabarcoding/docker/pimba -t AllSamples_otu_table.txt -a AllSamples_otus_tax_assignments.txt -m mapping_file.csv -g local
```

Uma maior exploração dos dados pode ser feita utilizando o software R e o pacote `phyloseq` (<https://joey711.github.io/phyloseq/index.html>), usando como entrada as matrizes de contagem de OTUs (`OTU_table`) e de taxonomia (`tax_assignment`).

OBS: Caso for correr o PIMBA em um computador pessoal ou outro servidor, o manual de uso e instalação está disponível em <https://github.com/reinator/pimba>.

**Figura 7** - Exemplos dos arquivos finais gerados pelas etapas de análise do pipeline PIMBA. Esses são os arquivos que podem ser utilizados em análises estatísticas posteriores.

Example of OTU table (otu\_table.txt) generated by pimba\_run.sh:

OTUID	EXTB	MB211	MB214	MB217	MB223	MB224	MB227	MB230	MB233	MB236	MB239	MB242	MB245	MB248	PCRB
OTU955365479	74371	116	24	52	236	98	147	19	45	1039	2355	60	16	3	20
OTU812100800	155496	225	32	88	86	152	525	17	147	329	126	308	73	9	35
OTU789743117	54074	21	15	22	20	48	191	8	11	329	14	2	71	4	14
OTU645601205	8938	4	2	7	3	5	27	2	4	10	6	1	8	0	1
OTU212911189	2037	2	0	0	0	0	2	0	0	1	0	1	0	0	0
OTU964442511	140	76	82	227	64	128	291	135	94	373557	265	167	957	113	76
OTU533441192	164	133	62	265895	97	67	246	43	55	444	106	31	124	51	92
OTU103992458	724	0	2	1	0	1	2	0	6	1	1	9	0	0	0
OTU524853145	530	389	212	799483	347	3301	722	142	183	1150	284	118	377	165	236
OTU844428705	310	310	180	390	320	253	350	518	214	353599	610124	64796	39526	1131	212
OTU293268522	270	37	28	92	1007	120	121096	48	14	51	33	71	71	76	18
OTU884692304	253	0	0	1	0	0	3	0	0	0	0	0	0	0	0
OTU986112541	434	0	0	0	0	1	0	0	0	2	0	0	0	0	0
OTU176656043	97	0	0	0	0	0	2	0	0	0	0	0	0	0	0
OTU689340459	3	0	0	0	0	2	1	0	296	0	1	3	0	1	0
OTU886162264	411	2	0	3	2	1	0	0	1	0	1	2	3	0	1
OTU771688223	499	0	0	0	1	0	1	0	2	1	1	0	0	0	0
OTU484938833	8	10	6	4	14	3	1	11940	272	3	4	5	2	9	0
OTU46757698	1	0	0	0	0	0	0	262	5	243	0	0	0	0	0

Example of OTU Tax assignment (tax\_assignment.txt):

OTU598595363	k	Metazoa	p	Arthropoda	c	Insecta	o	Lepidoptera	f	Eriocraniidae	g	Eriocrania	s	sp.ER06	0.96154
OTU766111137	k	Metazoa	p	Arthropoda	c	Insecta	o	Diptera	f	Calliphoridae	g	Huascaromusca	s	sp.1MATM-2015	0.94574
OTU679530351	k	Metazoa	p	Arthropoda	c	Insecta	o	Lepidoptera	f	Nymphalidae	g	Hypothyris	s	moebiusmoebiusi	0.96154
OTU120497499	k	Metazoa	p	Arthropoda	c	Insecta	o	Diptera	f	Culicidae	g	Anopheles	s	koliensis	0.84733
OTU277420622	k	Metazoa	p	Arthropoda	c	Insecta	o	Lepidoptera	f	Nymphalidae	g	Ideopsis	s	gaura	0.95385
OTU374560010	k	Metazoa	p	Arthropoda	c	Arachnida	o	Scorpiones	f	Buthidae	g	Isometrus	s	maculatus	0.87692
OTU8660394	k	Metazoa	p	Arthropoda	c	Insecta	o	Diptera	f	Empididae	g	Rhaphomyia	s	sp.1BKC-2015	0.96899
OTU261821968	k	Metazoa	p	Arthropoda	c	Insecta	o	Diptera	f	Syrphidae	g	Toxomerus	s	apegiensis	0.95349
OTU302289343	k	Metazoa	p	Arthropoda	c	Insecta	o	Lepidoptera	f	Lycaenidae	g	Allotinus	s	davidis	0.84733
OTU412412883	k	Metazoa	p	Arthropoda	c	Insecta	o	Diptera	f	Drosophilidae	g	Zygothrica	s	ptilialis	0.94615
OTU285620827	k	p	c	Florideophyceae	o	Gigartinales	f	Peyssonneliaceae	g	Riquetophycus	s	sp.HSY-2014a	0.91473		
OTU288350188	k	Metazoa	p	Arthropoda	c	Insecta	o	Lepidoptera	f	Nymphalidae	g	Hypomenitis	s	alphesiboea	0.96124
OTU425378105	k	Metazoa	p	Nematoda	c	Chromadorea	o	Tylenchida	f	Aphelenchoididae	g	Bursaphelenchus	s	paraluxuriosae	0.89552
OTU733572620	k	Metazoa	p	Arthropoda	c	Insecta	o	Diptera	f	Syrphidae	g	Toxomerus	s	difficilis	0.96923
OTU240332622	k	Metazoa	p	Arthropoda	c	Insecta	o	Diptera	f	Culicidae	g	Anopheles	s	nuneztovari	0.93846

Example of Metadata file (metadata.csv):

SampleID	SampleName	BarcodeSequence	LinkerPrimer_Description
EXTB	EXTB	CGTACTAG+CGCTTAGA	CTGTCTCTIATACATCT, Branco
PCRB	PCRB	CGTACTAG+TCTTAGCC	CTGTCTCTIATACATCT, Branco
MB211	MB211	AGGCAGAA+TCTTAGCC	CTGTCTCTIATACATCT, Cav1
MB214	MB214	TCCTGAGC+TCTTAGCC	CTGTCTCTIATACATCT, Cav1
MB217	MB217	GGACTCCT+TCTTAGCC	CTGTCTCTIATACATCT, Cav1
MB223	MB223	TAGGCATG+TCTTAGCC	CTGTCTCTIATACATCT, Cav2
MB224	MB224	CTCTCTAC+TCTTAGCC	CTGTCTCTIATACATCT, Cav2
MB227	MB227	CAGAGAGG+TCTTAGCC	CTGTCTCTIATACATCT, Cav2
MB230	MB230	GCTACGCT+TCTTAGCC	CTGTCTCTIATACATCT, Cav3
MB233	MB233	CGAGGCTG+TCTTAGCC	CTGTCTCTIATACATCT, Cav3
MB236	MB236	AAGAGGCA+TCTTAGCC	CTGTCTCTIATACATCT, Cav3
MB239	MB239	GTAGAGGA+TCTTAGCC	CTGTCTCTIATACATCT, Cav4
MB242	MB242	TAAGGCGA+ATAGAGAG	CTGTCTCTIATACATCT, Cav4
MB245	MB245	CGTACTAG+ATAGAGAG	CTGTCTCTIATACATCT, Cav4
MB248	MB248	AGGCAGAA+ATAGAGAG	CTGTCTCTIATACATCT, Cav4

Fonte: Oliveira *et al.* (2021).

### 3 LISTA DE MATERIAIS

Ítem	Descrição	Fabricante	Núm. Parte	Site web
1	Luvas de nitrila sem pó	Merck	Z672327-50EA	<a href="https://www.sigmaaldrich.com/catalog/product/aldrich/z672327?lang=pt&amp;region=BR">https://www.sigmaaldrich.com/catalog/product/aldrich/z672327?lang=pt&amp;region=BR</a>
2	Whirl-Pak	Sigma	WPB01546WA-500EA	<a href="https://www.sigmaaldrich.com/catalog/product/sigma/wpb01546wa?lang=pt&amp;region=BR&amp;cm_sp=Insite-_-prodRecCold_xviews-_-prodRecCold10-1">https://www.sigmaaldrich.com/catalog/product/sigma/wpb01546wa?lang=pt&amp;region=BR&amp;cm_sp=Insite-_-prodRecCold_xviews-_-prodRecCold10-1</a>
3	Gelo artificial reutilizável	Americanas	94055821	<a href="https://www.americanas.com.br/produto/94055821/kit-com-10-unidades-gelo-artificial-reutilizavel-cliogel-rigido-500-ml?DCSext.recom=RR_item_page.rr1-CategorySiloedViewCP&amp;dcsext.recom=RR_item_page.rr1-CategorySiloedViewCP&amp;nm_origem=rec_item_page.rr1-CategorySiloedViewCP&amp;nm_ranking_rec=5">https://www.americanas.com.br/produto/94055821/kit-com-10-unidades-gelo-artificial-reutilizavel-cliogel-rigido-500-ml?DCSext.recom=RR_item_page.rr1-CategorySiloedViewCP&amp;dcsext.recom=RR_item_page.rr1-CategorySiloedViewCP&amp;nm_origem=rec_item_page.rr1-CategorySiloedViewCP&amp;nm_ranking_rec=5</a>
4	Sacola térmica	Americanas	93144353	<a href="https://www.americanas.com.br/produto/93144353/sacola-termica-poliester-yins-cinza?pfm_carac=sacola%20%C3%A9rmica&amp;pfm_index=4&amp;pfm_page=search&amp;pfm_pos=grid&amp;pfm_type=search_page">https://www.americanas.com.br/produto/93144353/sacola-termica-poliester-yins-cinza?pfm_carac=sacola%20%C3%A9rmica&amp;pfm_index=4&amp;pfm_page=search&amp;pfm_pos=grid&amp;pfm_type=search_page</a>
5	Qubit™ 4 Fluorometer	Thermo Fisher	Q33226	<a href="https://www.thermofisher.com/order/catalog/product/Q33226?SID=srch-srp-Q33226#/Q33226?SID=srch-srp-Q33226">https://www.thermofisher.com/order/catalog/product/Q33226?SID=srch-srp-Q33226#/Q33226?SID=srch-srp-Q33226</a>
6	Qubit™ dsDNA HS Assay Kit	Thermo Fisher	Q32854	<a href="https://www.thermofisher.com/order/catalog/product/Q32854#/Q32854">https://www.thermofisher.com/order/catalog/product/Q32854#/Q32854</a>
7	Qubit™ dsDNA BR Assay Kit	Thermo Fisher	Q32850	<a href="https://www.thermofisher.com/order/catalog/product/Q32850?SID=srch-srp-Q32850#/Q32850?SID=srch-srp-Q32850">https://www.thermofisher.com/order/catalog/product/Q32850?SID=srch-srp-Q32850#/Q32850?SID=srch-srp-Q32850</a>
8	Eppendorf Safe-Lock Tubes 1,5 ml	Eppendorf	0030120086	<a href="https://online-shop.eppendorf.com.br/BR-pt/Ponteiras-tubos-e-placas-44512/Tubos-44515/Eppendorf-Safe-Lock-Tubes-PF-8863.html">https://online-shop.eppendorf.com.br/BR-pt/Ponteiras-tubos-e-placas-44512/Tubos-44515/Eppendorf-Safe-Lock-Tubes-PF-8863.html</a>
9	Veriti 96-Well Thermal Cycler1	Thermo Fisher	4375786	<a href="https://www.thermofisher.com/order/catalog/product/4375786#/4375786">https://www.thermofisher.com/order/catalog/product/4375786#/4375786</a>
10	Ponteiras com barreiras para PCR	Thermo Fisher	AM12650, AM12660, AM12635, AM12648	<a href="https://www.thermofisher.com/br/en/home/life-science/dna-rna-purification-analysis/rna-extraction/rna-extraction-products/nuclease-free-tubes-tips-and-buffers/nuclease-free-tubes-and-tips.html?gclid=CjwKCAiA_f3uBRAmEiwAzPuaM2QKXJO0XfyKOWifG68Rpm0KDA2G1k1K6evZmKUIQvmD-iTtyq47IRoC_cIQAvD_BwE&amp;s_kwid=AL!3652!3!350493651280!b!!g!!%2Bbarrier%20%2Btips&amp;ef_id=CjwKCAiA_f3uBRAmEiwAzPuaM2QKXJO0XfyKOWifG68Rpm0KDA2G1k1K6evZmKUIQvmD-iTtyq47IRoC_cIQAvD_BwE:G:s&amp;s_kwid=AL!3652!3!350493651280!b!!g!!%2Bbarrier%20%2Btips">https://www.thermofisher.com/br/en/home/life-science/dna-rna-purification-analysis/rna-extraction/rna-extraction-products/nuclease-free-tubes-tips-and-buffers/nuclease-free-tubes-and-tips.html?gclid=CjwKCAiA_f3uBRAmEiwAzPuaM2QKXJO0XfyKOWifG68Rpm0KDA2G1k1K6evZmKUIQvmD-iTtyq47IRoC_cIQAvD_BwE&amp;s_kwid=AL!3652!3!350493651280!b!!g!!%2Bbarrier%20%2Btips&amp;ef_id=CjwKCAiA_f3uBRAmEiwAzPuaM2QKXJO0XfyKOWifG68Rpm0KDA2G1k1K6evZmKUIQvmD-iTtyq47IRoC_cIQAvD_BwE:G:s&amp;s_kwid=AL!3652!3!350493651280!b!!g!!%2Bbarrier%20%2Btips</a>
11	Pipetadores automáticos	Eppendorf	3120000020, 3120000011	<a href="https://online-shop.eppendorf.com.br/BR-pt/Pipetas-e-Dispensadores-44563/Pipetagem-manual--dispensacao-44564/Eppendorf-Research-plus-PF-534798.html">https://online-shop.eppendorf.com.br/BR-pt/Pipetas-e-Dispensadores-44563/Pipetagem-manual--dispensacao-44564/Eppendorf-Research-plus-PF-534798.html</a>
12	96-well 0.2 ml skirtless PCR plates	Fisher	14-222-321	<a href="https://www.fishersci.com/shop/products/axxygen-96-well-semi-skirted-pcr-microplates-2/p-4104932">https://www.fishersci.com/shop/products/axxygen-96-well-semi-skirted-pcr-microplates-2/p-4104932</a>
13	Agencourt AMPure XP 60 ml kit	Beckman Coulter Genomics	A63881	<a href="https://www.mybeckman.com.br/reagents/genomic/cleanup-and-size-selection/pcr/a63881">https://www.mybeckman.com.br/reagents/genomic/cleanup-and-size-selection/pcr/a63881</a>
14	Álcool absolute para biologia molecular	Sigma	E7023	<a href="https://www.sigmaaldrich.com/catalog/product/sigald/e7023?lang=pt&amp;region=BR&amp;gclid=CjwKCAiA_f3uBRAmEiwAzPuaMyDQvnYQ4JXhsrF6Z3BqTGkQwPhQNqleHpfNN7n4vGZ9PpKL83wsxoCVLEQAvD_BwE">https://www.sigmaaldrich.com/catalog/product/sigald/e7023?lang=pt&amp;region=BR&amp;gclid=CjwKCAiA_f3uBRAmEiwAzPuaMyDQvnYQ4JXhsrF6Z3BqTGkQwPhQNqleHpfNN7n4vGZ9PpKL83wsxoCVLEQAvD_BwE</a>
15	Microseal 'A' adhesive seals	Bio-Rad	MSA-5001	<a href="https://www.bio-rad.com/pt-br/sku/msa5001-microseal-pcr-plate-pcr-tube-sealing-film?ID=msa5001">https://www.bio-rad.com/pt-br/sku/msa5001-microseal-pcr-plate-pcr-tube-sealing-film?ID=msa5001</a>
16	Microseal 'B' adhesive seals	Bio-Rad	MSB-1001	<a href="https://www.bio-rad.com/pt-br/sku/msb1001edu-microseal-b-adhesive-seals?ID=msb1001edu">https://www.bio-rad.com/pt-br/sku/msb1001edu-microseal-b-adhesive-seals?ID=msb1001edu</a>

Ítem	Descrição	Fabricante	Núm. Parte	Site web
17	MiSeq Reagent Kit v3 (600 cycle)	Illumina	MS-102-3003	<a href="https://www.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/miseq-reagent-kit-v3.html">https://www.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/miseq-reagent-kit-v3.html</a>
18	Nextera XT Index Kit	Illumina	FC-131-1001	<a href="https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-xt-dna.html">https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-xt-dna.html</a>
19	PhiX Control Kit v3	Illumina	FC-110-3001	<a href="https://www.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/phix-control-v3.html">https://www.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/phix-control-v3.html</a>
20	Tris-HCl 10 mM, pH 8.5	Sigma	10812846001	<a href="https://www.sigmaaldrich.com/catalog/search?term=10+mM+Tris-HCl&amp;interface=All&amp;N=0&amp;mode=partialmax&amp;lang=pt&amp;region=global&amp;focus=product">https://www.sigmaaldrich.com/catalog/search?term=10+mM+Tris-HCl&amp;interface=All&amp;N=0&amp;mode=partialmax&amp;lang=pt&amp;region=global&amp;focus=product</a>
21	Magnetic stand-96	Thermo Fisher	AM10027	<a href="https://www.thermofisher.com/order/catalog/product/AM10027?SID=srch-srp-AM10027#/AM10027?SID=srch-srp-AM10027">https://www.thermofisher.com/order/catalog/product/AM10027?SID=srch-srp-AM10027#/AM10027?SID=srch-srp-AM10027</a>
22	Centrífuga de microplaca	Eppendorf	5804 R	<a href="https://online-shop.eppendorf.com.br/BR-pt/Centrifugacao-44533/Centrifugas-44534/Centrifuge-5804-5804R-PF-240993.html">https://online-shop.eppendorf.com.br/BR-pt/Centrifugacao-44533/Centrifugas-44534/Centrifuge-5804-5804R-PF-240993.html</a>
23	TruSeq Index Plate Fixture Kit	Illumina	FC-130-1005	<a href="https://www.illumina.com/products/by-type/ivd-products/truseq-custom-amplicon-dx.html">https://www.illumina.com/products/by-type/ivd-products/truseq-custom-amplicon-dx.html</a>
24	High Speed Micro Plate Shaker	VWR	13500-890	<a href="https://us.vwr.com/store/product/4779955/vwr-signaturetm-high-speed-microplate-shaker">https://us.vwr.com/store/product/4779955/vwr-signaturetm-high-speed-microplate-shaker</a>
25	Espátulas de metal	Sigma	Z283274-1EA	<a href="https://www.sigmaaldrich.com/catalog/product/aldrich/z283274?lang=pt&amp;region=BR">https://www.sigmaaldrich.com/catalog/product/aldrich/z283274?lang=pt&amp;region=BR</a>
26	Direct Thermal Cryo-Tags	Sigma		<a href="https://www.sigmaaldrich.com/catalog/substance/directthermalcryotags1234598765?lang=pt&amp;region=BR">https://www.sigmaaldrich.com/catalog/substance/directthermalcryotags1234598765?lang=pt&amp;region=BR</a>
27	DNeasy PowerSoil Kit Print	Qiagen	12888-100	<a href="https://www.qiagen.com/de/products/dneasy-powersoil-kit/#orderinginformation">https://www.qiagen.com/de/products/dneasy-powersoil-kit/#orderinginformation</a>
28	Capela de fluxo laminar	SPLabor	Linha100	<a href="http://www.splabor.com.br/produto/capela-de-fluxo-laminar-vertical-dimensoes-internas-1890x652x652-mm-linha-100/">http://www.splabor.com.br/produto/capela-de-fluxo-laminar-vertical-dimensoes-internas-1890x652x652-mm-linha-100/</a>
29	Agarose	Thermo Fisher	16500100	<a href="https://www.thermofisher.com/br/en/home/life-science/dna-rna-purification-analysis/nucleic-acid-gel-electrophoresis/dna-electrophoresis/agarose-gel-electrophoresis/routine-dna-gels.html?open=pure#pure">https://www.thermofisher.com/br/en/home/life-science/dna-rna-purification-analysis/nucleic-acid-gel-electrophoresis/dna-electrophoresis/agarose-gel-electrophoresis/routine-dna-gels.html?open=pure#pure</a>
30	Kit de eletroforese horizontal	BioRad	1655044EDU	<a href="https://www.bio-rad.com/pt-br/sku/1655044edu-wide-mini-sub-cell-gt-with-powerpac-mini-power-supply?ID=1655044EDU">https://www.bio-rad.com/pt-br/sku/1655044edu-wide-mini-sub-cell-gt-with-powerpac-mini-power-supply?ID=1655044EDU</a>
31	Bioanalyzer Agilent Technology 2100	Agilent	G2939BA	<a href="https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-instrument/2100-bioanalyzer-instrument-228250">https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-instrument/2100-bioanalyzer-instrument-228250</a>
32	DNA 1000 Kit	Agilent	5067-1504	<a href="https://www.agilent.com/store/productDetail.jsp?catalogId=5067-1504">https://www.agilent.com/store/productDetail.jsp?catalogId=5067-1504</a>
33	4200 TapeStation System	Agilent	G2991AA	<a href="https://www.agilent.com/en/product/automated-electrophoresis/tapestation-systems/tapestation-instruments/4200-tapestation-system-228263">https://www.agilent.com/en/product/automated-electrophoresis/tapestation-systems/tapestation-instruments/4200-tapestation-system-228263</a>

## REFERÊNCIAS

BOLYEN, E. J. R. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. **Nature Biotechnology**, v. 37, s/n, p. 852-857, 2019. Disponível em: <https://www.nature.com/articles/s41587-019-0209-9>. Acesso em: 04 set. 2020.

BRASIL. Congresso. Planalto. **Lei n. 6.938, de 31 de agosto de 1981 e suas alterações (Leis n. 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 ago. 2020

BRASIL. Ministério do Meio Ambiente. Conselho Nacional do Meio Ambiente. **Resolução do CONAMA n. 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 ago. 2020.

CAPORASO, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. **Nature Methods**. v.7, s/n, p. 335-336, 2010. Disponível em: <https://www.nature.com/articles/nmeth.f.303>. Acesso em: 04 de setembro de 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

CORLETT, R. T. "A Bigger Toolbox: Biotechnology in Biodiversity Conservation." **Trends Biotechnology**, v. 35, n. 1, p. 55-65, 2017. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0167779916300919>. Acesso em: 04 set. 2020.

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

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.

CHEN, 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

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**, v. 108, n. 49, p. 19641-19646, 2011. Disponível em: <https://www.pnas.org/content/pnas/108/49/19641.full.pdf>. Acesso em: 07 ago. 2020.

CHIOU, S-J. *et al.* Authentication of medicinal herbs using PCR-amplified ITS2 with specific primers. **Planta medica**, v. 73, s/n, p. 1421-1426, 2007.

CRISTESCU, M. E. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. **Trends Ecology Evolution**, v. 29, n. 10, p. 566-571, 2014.

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. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/mec.14350>. Acesso em: 07 ago. 2020.

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, online, 2016. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0157505>. Acesso em: 18 ago. 2020.

FAZEKAS, A. J., *et al.* Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. **PLoS ONE**, v. 3, n. 7, online, 2008. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0002802>. Acesso em: 18 ago. 2020

FAZEKAS, A. J. *et al.* "Are plant species inherently harder to discriminate than animal species using DNA barcoding markers?" **Molecular Ecology Resources**, v. 9, s/n, p. 130-139, 2009. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/21564972/>. Acesso em: 04 set. 2020.

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, online, 2019. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0202844>. Acesso em: 23 ago. 2020.

GARDES, M.; BRUNS, T. D. ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. **Molecular Ecology**, v. 2, n. 2, p. 113-118, 1993.

GIULIETTI, A. *et al.* Edaphic Endemism in the Amazon: Vascular Plants of the canga of Carajás, Brazil. **The Botanical Review**, v. 85, s/n, p. 357-383, 2019. Disponível em: <https://link.springer.com/article/10.1007/s12229-019-09214-x#citeas>. Acesso em: 27 ago. 2020.

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.; HOLLIINGSWORTH, P. M.; HAJIBABAEI, M. From writing to reading the encyclopedia of life. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 371, n. 1702, online, 2016. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rstb.2015.0321>. Acesso em: 18 ago. 2020

HOLLINGSWORTH, P. M.; GRAHAM, S. W.; LITTLE, D. P. Choosing and using a plant DNA barcode. **PLoS One**, v. 6, n. 5, online, 2011. Disponível em:

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019254>. Acesso em: 30 ago. 2020.

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.

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, online, 2016. Disponível em: [https://royalsocietypublishing.org/doi/10.1098/rstb.2015.0338?url\\_ver=Z39.88-2003&rfr\\_id=ori:rid:crossref.org&rfr\\_dat=cr\\_pub%20%20pubmed](https://royalsocietypublishing.org/doi/10.1098/rstb.2015.0338?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed). Acesso em: 29 ago. 2020.

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. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/jse.12254>. Acesso em: 12 set. 2020.

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. Disponível em: <https://www.nature.com/articles/s41598-018-26648-2#citeas>. Acesso em: 17 ago. 2020.

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

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. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/mec.14478>. Acesso em: 16 jul. 2020.

ROGNES, T. *et al.* "VSEARCH: a versatile open source tool for metagenomics." **PeerJ**, v. 4, p. e2584. 2016.

SCHUBERT, M.; LINDGREEN, S.; ORLANDO, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. **BMC Res Notes**, v. 9, n. 88. 2016.

SHNEER, V. S. DNA barcoding is a new approach in comparative genomics of plants. **Genetika**, v. 45, n. 11, p. 1436-1448. 2009. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/20058792/>. Acesso em: 23 set. 2020

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, online, 2020. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231436>. Acesso em: 09 jun. 2020

VASCONCELOS, S. *et al.* Unravelling the plant diversity of the Amazonian *canga* through DNA barcoding. In prep.

WHITE, T. J. *et al.* Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. **PCR Protocols: A Guide to Methods and Applications**, v. 18, s/n, p. 315-322, 1990.

ZAPPI, D. C. **Paisagens e plantas de Carajás** = Landscapes and plants of Carajás. Belém: ITV, 2017.

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. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6025816/pdf/APS3-6-e01157.pdf>. Acesso em: 18 ago. 2020.

ZHANG, J. *et al.* PEAR: a fast and accurate Illumina Paired-End reAd mergeR. **Bioinformatics**, v. 30, n. 5, p. 614-620. 2014. Disponível em: <https://academic.oup.com/bioinformatics/article/30/5/614/247231>. Acesso em: 18 ago. 2020.

## ANEXOS

### A - Kit de extração de DNA DNeasy PowerSoil

O manual pode ser obtido neste link:

<https://www.qiagen.com/th/resources/download.aspx?id=91cf8513-a8ec-4f45-921e-8938c3a5490c&lang=en>

### B - Qubit™ 4 Fluorometer;

O manual do Qubit™ 4 Fluorometer pode ser obtido neste link:

<https://www.thermofisher.com/order/catalog/product/Q33226?SID=srch-srp-Q33226#/Q33226?SID=srch-srp-Q33226>

### C - Qubit® dsDNA HS Assay Kit;

O manual do Qubit® dsDNA HS Assay Kits pode ser obtido neste link:

<https://www.thermofisher.com/order/catalog/product/Q33226?SID=srch-srp-Q33226#/Q33226?SID=srch-srp-Q33226>

### D - Qubit® dsDNA BR Assay Kit

O manual do Qubit® dsDNA BR Assay Kits pode ser obtido neste link:

<https://www.thermofisher.com/order/catalog/product/Q32853#/Q32853>

### E - Preparação da solução TBT – PAR (Solução 5X)

Concentrações finais:

Trealose 750 mM (catalog no. T9531, Sigma – Aldrich)

BSA (não – acetilado) 1 mg/mL

Tween-20 1%

Tris –HCl 8.5 mM pH 8.0

Passos:

1. Preparar 1ml de uma solução de BSA a 20 mg/mL. Dissolver 20 mg de BSA sólido em 800 µL de água ultrapura e então ajustar o volume final para 1mL;
2. Preparar 8,5 de tampão Tris 10 mM adicionando 85 µL de Tris-HCl 1 M (pH 8.0) a 8.415 mL de água ultrapura em um tubo Falcon de 15 mL e misturar bem;
3. Em um tubo Falcon de 50 mL, dissolver 2.835 g de trealose (no. de catálogo T9531, Sigma - Aldrich) em 6 mL;
4. Adicionar 500 µL da solução de BSA preparada na etapa 1 e 100 µL de Tween – 20 à solução de trealose preparada na etapa 3 e misturar bem.

### **F -16S Metagenomic Sequencing Library Preparation**

O manual do procedimento para a construção das bibliotecas pode ser obtido neste link:

[https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry\\_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf)

### **G - MiSeq System Guide**

O manual do MiSeq System pode ser obtido neste link:

[https://support.illumina.com/sequencing/sequencing\\_instruments/miseq/documentation.html](https://support.illumina.com/sequencing/sequencing_instruments/miseq/documentation.html)

### **H - Guia de instalação e uso do PIMBA**

<https://github.com/reinator/pimba> .