



Universidade Federal de Pernambuco
Centro de Biociências

LUDIMILLA VITÓRIA LEONEL LOPES

**ANÁLISE DE DADOS DE SEQUENCIAMENTO DE
AMOSTRAS DE AR COLETADAS UTILIZANDO DRONES**

Recife
2024

LUDIMILLA VITÓRIA LEONEL LOPES

**ANÁLISE DE DADOS DE SEQUENCIAMENTO DE
AMOSTRAS DE AR COLETADAS UTILIZANDO DRONES**

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação em
Biomedicina da Universidade Federal de
Pernambuco, como pré-requisito à
obtenção do título de Bacharel em
Biomedicina.

Orientador: Drº. Antonio Mauro Rezende
Coorientador: Drº. Helotonio Carvalho

Recife
2024

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Lopes, Ludimilla Vitória Leonel.

Análise de dados de sequenciamento de amostras de ar coletadas utilizando
drones / Ludimilla Vitória Leonel Lopes. - Recife, 2024.

54p. : il., tab.

Orientador(a): Antonio Mauro Rezende

Coorientador(a): Helotonio Carvalho

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de
Pernambuco, Centro de Biociências, Biomedicina, 2024.

Inclui referências.

1. Metagenômica. 2. 16S. 3. Ar. 4. Drone. 5. Sequenciamento. I. Rezende,
Antonio Mauro. (Orientação). II. Carvalho, Helotonio. (Coorientação). IV. Título.

500 CDD (22.ed.)

LUDIMILLA VITÓRIA LEONEL LOPES

ANÁLISE DE DADOS DE SEQUENCIAMENTO DE AMOSTRAS DE AR COLETADAS UTILIZANDO DRONES

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação
em Biomedicina da Universidade
Federal de Pernambuco, como
pré-requisito à obtenção do título de
Bacharel em Biomedicina.

Aprovada em: ____/____/____

BANCA EXAMINADORA

Orientador: Prof. Dr°. Antonio Mauro Rezende
Instituto René Rachou - FIOCRUZ MINAS

Dr°. João Luiz de Lemos Padilha Pitta
Instituto Aggeu Magalhães- FIOCRUZ/ Núcleo de Bioinformática

Dr°. Igor Vasconcelos Rocha
Instituto Aggeu Magalhães - FIOCRUZ/ Departamento de Microbiologia

Dedico este trabalho a minha mãe,
meu pai e meus irmãos, que sempre
acreditaram no meu potencial e
fizeram o possível e o impossível para
que eu chegasse até aqui.

AGRADECIMENTOS

Agradeço ao meu orientador pela oportunidade de ter sido sua aluna de iniciação científica e por toda paciência, atenção e dedicação que me foi concedida ao longo desses anos.

Ao meu coorientador pela atenção e disponibilidade.

A FIOCRUZ/FACEPE e ao CNPq pelo apoio financeiro.

Aos meus colegas de projeto Pedro e André, pela companhia, apoio na realização das atividades e momentos de descontração.

À Crhisllane, João e Ângela por toda ajuda e contribuição com o projeto.

A toda à equipe que compõe o Departamento de Microbiologia do Instituto Aggeu Magalhães e que contribuiu direta ou indiretamente para a realização do projeto.

Agradeço à minha mãe Vitória Regina, ao meu pai Fransuir, e aos meus irmãos Lucas e Laura por todo apoio, incentivo e por nunca perderem a fé em mim. Vocês são a minha força.

À minha madrinha Fabrícia, que sempre torceu por mim.

À minha avó Zilda com quem tenho uma ligação muito forte e que nem mesmo a distância interfere. Te amo, vó.

Agradeço a tia Carla e família por me ajudarem a permanecer em Recife no início da graduação.

À minha avó Alaíde e aos meus avôs Chico e Wilson (in memoriam), que se ainda estivessem nesse plano estariam muito felizes pela minha conquista.

Ao meu namorado Murilo, por toda paciência e cuidado durante os momentos difíceis.

As minhas amigas, com quem eu posso sempre contar.

LOPES, Ludimilla Vitória Leonel. **Análise de dados de sequenciamento de amostras de ar coletadas utilizando drones**. 2024. 54.Trabalho de Conclusão de Curso (Graduação em Biomedicina) - Universidade Federal de Pernambuco, Recife, 2024.

RESUMO

Um microbioma é formado por comunidades de diversas espécies de microrganismos que coexistem e se relacionam. Estudos e a caracterização dessas comunidades microbianas de forma não enviesada podem ser estabelecidos através da metagenômica, que corresponde ao sequenciamento direto dos genomas de um determinado microbioma. A aplicação desta técnica é possível atualmente devido ao aprimoramento e a adesão das tecnologias de sequenciamento de DNA de alto rendimento (HTS). Uma das abordagens conhecidas dessa técnica é o sequenciamento de *amplicon*, que tem como base o sequenciamento de regiões conservadas e hipervariáveis presentes em genes marcadores, e.g. o gene que codifica a subunidade 16S do RNA ribossomal para avaliação de comunidades bacterianas. Em paralelo, as análises sobre a composição de microbiomas ambientais são majoritariamente realizadas em solo ou água, contudo o ar é um compartimento pouco explorado em virtude, principalmente, da dificuldade em realizar coletas. Considerando tal problemática, o presente trabalho apresentou o desenvolvimento de uma nova metodologia de coleta que consistiu em utilizar veículo aéreo não tripulado (VANT), de pequeno porte, popularmente conhecidos como drones, atrelado ao uso de coletores e posterior análise metagenômica utilizando o sequenciamento das regiões V3-V4 do gene 16S rRNA para avaliação das comunidades microbianas encontradas no distrito de Pirituba, Vitória de Santo Antão- PE. Após a coleta, o material foi extraído utilizando o protocolo de Fenol/Clorofórmio e o DNA foi amplificado para as regiões V3-V4 do gene do rRNA 16S através da PCR. Por último, foram construídas as bibliotecas de sequenciamento, que em seguida foram quantificadas com Qubit dsDNA HS e sequenciadas na plataforma Illumina Miseq. O pré-processamento das sequências foi realizado pela ferramenta *fastp*, no qual uma quantidade relevante de bases de baixa qualidade foi eliminada, posteriormente as sequências foram analisadas através do *pipeline* Ampliseq, disponibilizado pelo repositório nf-core, e também pelo pacote *Phyloseq* R. As etapas de análises percorridas pelo *pipeline* foram controle de qualidade de sequências, inferência de ASVs, classificação taxonômica e cálculo das diversidades alfa e beta. Contudo, após a etapa de inferência de ASVs identificou-se uma contaminação considerável dos controles negativos e a partir disso, o pacote *Phyloseq* R foi utilizado para atenuar a contaminação nas outras amostras e assim obter uma classificação taxonômica e cálculos de diversidade mais acurados. Das 2123 ASVs previamente identificadas, restaram 2108 ASVs após a descontaminação. Ao final, apesar das dificuldades de padronização da coleta, extração e sequenciamento, a metodologia empregada permitiu caracterizar a composição microbiana das amostras minimamente.

Palavras-chave: Metagenômica. 16S. Ar. Drone. Sequenciamento.

LOPES, Ludimilla Vitória Leonel. **Análise de dados de sequenciamento de amostras de ar coletadas utilizando drones**. 2024. 54. Trabalho de Conclusão de Curso (Graduação em Biomedicina) - Universidade Federal de Pernambuco, Recife, 2024.

ABSTRACT

A microbiome is made up of communities of different species of microorganisms that coexist and relate to each other. Studies and characterization of these microbial communities in a non-biased way can be established through metagenomics, which corresponds to the direct sequencing of the genomes of a given microbiome. The application of this technique is currently possible due to the improvement and adoption of high-throughput DNA sequencing (HTS) technologies. One of the known approaches to this technique is amplicon sequencing, which is based on the sequencing of conserved and hypervariable regions present in marker genes, e.g. the gene encoding the 16S subunit of ribosomal RNA for the evaluation of bacterial communities. At the same time, analyses of the composition of environmental microbiomes are mostly carried out in soil or water, but air is a compartment that has been little explored, mainly due to the difficulty in collecting samples. Considering this problem, this work presented the development of a new collection methodology that consisted of using a small unmanned aerial vehicle (UAV), popularly known as drones, coupled with the use of collectors and subsequent metagenomic analysis using the sequencing of the V3-V4 regions of the 16S rRNA gene to evaluate the microbial communities found in the district of Pirituba, Vitória de Santo Antão, PE. After collection, the material was extracted using the phenol/chloroform protocol and the DNA was amplified for the V3-V4 regions of the 16S rRNA gene using PCR. Finally, sequencing libraries were constructed and then quantified using Qubit dsDNA HS and sequenced on the Illumina Miseq platform. The sequences were pre-processed using the fastp tool, in which a significant number of low-quality bases were eliminated, after which the sequences were analyzed using the Ampliseq pipeline, available from the nf-core repository, and also using the Phyloseq R package. The analysis steps carried out by the pipeline were sequence quality control, ASV inference, taxonomic classification and calculation of alpha and beta diversity. However, after the ASV inference stage, considerable contamination of the negative controls was identified and the Phyloseq R package was used to mitigate contamination in the other samples and thus obtain a more accurate taxonomic classification and diversity calculations. Of the 2123 ASVs previously identified, 2108 ASVs remained after decontamination. In the end, despite the difficulties in standardizing the collection, extraction and sequencing, the methodology used allowed the microbial composition of the samples to be minimally characterized.

Key words: Metagenomics. 16S. Air. Drone. Sequencing.

LISTA DE ILUSTRAÇÕES

Figura 1 – Diferença entre a estrutura de um pentose de dNTP e ddNTP	14
Figura 2 – Leitura do gel após método de Sanger	15
Figura 3 – Reações enzimáticas do pirosequenciamento	16
Figura 4 – Representação esquemática do sequenciamento por síntese	17
Figura 5 – Ilustração 3D dos coletores	24
Figura 6 – Drone DJI modelo Inspire 2	25
Figura 7 – Trajeto percorrido para coleta de amostras com o coletor de fluxo ativo de ar.	26
Figura 8 – Coletor automatizado com seus elementos expostos (3D)	27
Figura 9 – Bomba de sucção utilizada para coleta das amostras	28
Figura 10 – Coletor manual e automatizado, respectivamente	32
Figura 11 – Localização e alcance de sensores no drone DJI Inspire 2	32
Figura 12 – Eletroforese após PCR confirmando amplificação das regiões V3 e V4 do 16S rRNA para amostras dos coletores.	34
Figura 13 – Eletroforese após PCR confirmando amplificação das regiões V3 e V4 do 16S rRNA para amostras da bomba de sucção.	34
Figura 14 – Eletroforese em gel de agarose 1% das bibliotecas.	35
Figura 15 – Qualidade média das sequências <i>forward</i> (R1) antes e depois do pré-processamento no fastp.	36
Figura 16 – Qualidade média das sequências <i>reverse</i> (R2) antes e depois do pré-processamento no fastp.	37
Figura 17 – Histogramas da frequência de porcentagem de abundância relativa das amostras controles.	42
Figura 18 – Diversidade alfa representada pelos índices de Shannon e Chao1	44
Figura 19 – Abundância relativa de táxons	45
Figura 20 – Diversidade beta representada pelo índice de Bray-curtis.	46

LISTA DE TABELAS

Tabela 1 – Quantificação de DNA das amostras no NanoDrop® 2000.	33
Tabela 2 – Quantidade de leituras que permaneceram após as fases do controle de qualidade do DADA2/ <i>Ampliseq</i> para biblioteca <i>paired-end</i> .	38
Tabela 3 – Quantidade de leituras que permaneceram após as fases do controle de qualidade do DADA2/ <i>Ampliseq</i> para biblioteca single-end.	40

SUMÁRIO

1 INTRODUÇÃO	11
2 REVISÃO BIBLIOGRÁFICA	12
2.1 MICROBIOLOGIA CLÁSSICA	12
2.2 SEQUENCIAMENTO DE PRIMEIRA GERAÇÃO	13
2.3 SEQUENCIAMENTO DE NOVA GERAÇÃO (SEGUNDA GERAÇÃO)	15
2.4 SEQUENCIAMENTO DE TERCEIRA GERAÇÃO	18
2.5 METAGENÔMICA E O AMBIENTE AÉREO	20
3 OBJETIVOS	23
3.1 OBJETIVO GERAL	23
3.2 OBJETIVOS ESPECÍFICOS	23
4 METODOLOGIA	24
4.1 COLETORES DE AR E DRONE	24
4.2 COLETA DAS AMOSTRAS DE AR	26
4.3 EXTRAÇÃO DE DNA	28
4.4 AMPLIFICAÇÃO DAS REGIÕES V3 E V4 DO GENE 16S rRNA.....	29
4.5 SEQUENCIAMENTO DE DNA	29
4.6 ANÁLISE COMPUTACIONAL	30
5 RESULTADOS	31
5.1 COLETORES DE AR E DRONE	31
5.2 EXTRAÇÃO DE DNA E AMPLIFICAÇÃO DAS REGIÕES V3 E V4 DO GENE 16S rRNA	33
5.3 SEQUENCIAMENTO DE DNA	35
5.4 ANÁLISE COMPUTACIONAL	36
6 DISCUSSÃO	47
7 CONCLUSÃO	49
REFERÊNCIAS	50

1 Introdução

A abordagem de metagenômica se caracteriza por utilizar o sequenciamento massivo de ácidos nucleicos para avaliar a microbiota de determinados ambientes e/ou compartimentos biológicos. Essa estratégia é principalmente usada para a prospecção e identificação de microrganismos com o menor viés possível, pois essa caracterização independe de métodos de cultivo normalmente utilizados em estudos sobre a microbiota. O avanço nas tecnologias de sequenciamento impulsionou substancialmente a capacidade de caracterizar amplamente a microbiota tanto no aspecto qualitativo, como quantitativamente. Consequentemente, o conhecimento que temos hoje sobre a diversidade de microrganismos existente em diferentes ambientes, tais como oceano, solo, sedimentos, rios e ambientes extremos expandiu de forma considerável (JIANG et al., 2015). No entanto, a diversidade e o impacto da comunidade microbiana presente no ar associada ao material particulado têm sido pouco estudados utilizando tanto abordagens convencionais de cultivo de microrganismos, quanto por abordagens de metagenômica. Isso se deve principalmente à dificuldade de se obter quantidades suficientes de material para análise devido à baixa densidade do material biológico presente no ar (BEHZAD; GOJOBORI; MINETA, 2015). Dessa forma, em um cenário no qual as limitações relacionadas aos estudos dos microrganismos limitam significativamente o entendimento das comunidades microbianas, surge a necessidade de desenvolver uma metodologia de coleta eficaz e que proporcione o uso adequado da metagenômica no ambiente aéreo.

Devido à importância da comunidade microbiana em vários processos físico-químicos no ambiente aéreo, é de grande relevância avaliarmos a diversidade dessa comunidade frente a diversos fatores tais como níveis de poluição, temperatura, luminosidade e umidade com o objetivo de entendermos como atividades antrópicas estão perturbando a dinâmica deste conjunto de microrganismos. Estas perturbações podem influenciar na disseminação de organismos patogênicos e inclusive no clima local (WOMACK; BOHANNAN; GREEN, 2010). Portanto, entender essa diversidade de forma sistêmica e sem viés com objetivo de definir marcadores biológicos que caracterizam cenários benéficos e maléficos para sociedade podem ter grande impacto na saúde e na tomada de decisões dos gestores locais no sentido de melhorar esses marcadores.

No que tange a poluição do ar e os seus malefícios para a saúde humana,

sabe-se que os efeitos podem ser notados a curto e longo prazo, e que interferem diretamente na mortalidade prematura em todo o mundo (MANISALIDIS et al., 2020). Tendo isso em vista, é de suma importância compreender de forma ampla a dinâmica dos microrganismos no ambiente aéreo levando em consideração os fatores abióticos e a interação destes micróbios em comunidade.

Atualmente, as coletas de materiais do ambiente aéreo são realizadas utilizando grandes equipamentos que são fixados em edifícios, logo são realizadas coletas localizadas, sem alteração do ponto geográfico e sem variações na altitude da coleta. Para material particulado proveniente da poluição atmosférica, normalmente, se utiliza um sistema de sucção com bomba de vácuo que força a passagem do ar por dispositivos conhecidos como impactadores, que selecionam partículas com tamanhos menores que 10 μM , no caso de PM10 ou 2,5 μM , no caso de PM2,5 por exemplo. Este material é coletado em filtros de teflon ou fibra de vidro para posterior análise gravimétrica e/ou química.

Uma vez que a localização geográfica e a altitude influenciam de forma direta em diversas variáveis abióticas do ambiente (luz, radiação, concentração de partículas, etc.), é esperado que exista uma variação na comunidade bacteriana em função dessas variáveis. Isto posto, o uso de drones para a coleta dos microrganismos possibilitará um levantamento mais representativo da comunidade microbiológica, uma vez que, a maneira atual de coleta com equipamentos fixos não leva em conta variações em fatores abióticos importantes para a diversidade microbiológica. Dessa forma, o presente projeto propõe uma validação dos materiais e métodos utilizados e análise dos dados que serão produzidos.

2 REVISÃO BIBLIOGRÁFICA

2.1 MICROBIOLOGIA CLÁSSICA

Os estudos dos microrganismos tiveram início no século XVII, quando Antonie van Leeuwenhoek desenvolveu o primeiro microscópio e descreveu pela primeira vez a visualização de bactérias, fungos e protozoários de amostras coletadas da água, lama e placa dentária. Posteriormente, Robert Koch também se destacou com o produção de técnicas de fixação e coloração para visualização no microscópio,

além do uso de placas de Petri, onde conseguiu isolar e cultivar microrganismos em meios de cultura para associá-los a doenças humanas como antraz e cólera. A partir desses marcos, estudos voltados para a identificação de microrganismos causadores de doenças em humanos se intensificaram. (ESCOBAR-ZEPEDA; VERA-PONCE DE LEÓN; SANCHEZ-FLORES, 2015).

Sergei Winogradsky, considerado o pai da microbiologia moderna, desenvolveu técnicas de enriquecimento de meio e deu início ao estudo da ecologia microbiana, compreendendo a importância do estudo das populações *in situ* ou o mais próximo possível das condições naturais. A partir dessa abordagem, surgiu uma nova perspectiva a respeito do estudo dos microrganismos, na qual os objetos de estudos não necessariamente trazem riscos à saúde humana, mas como seres que coexistem em um determinado ambiente e desempenham papéis que por muitas vezes são benéficos para os humanos (BERG et al., 2020).

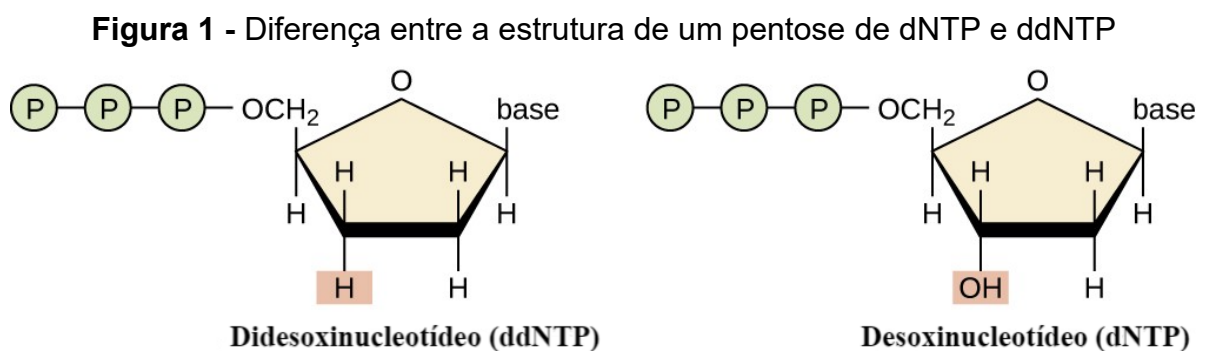
Outra informação que surge nesse cenário é a respeito do cultivo das espécies, visto que, as descobertas de Winogradsky demonstraram que nem todos os microrganismos são suscetíveis ao cultivo em laboratório, problemática que persiste até a atualidade. Sabe-se que a quantidade de microrganismos cultiváveis em laboratório é pequena, uma vez que, alguns organismos precisam de nutrientes e condições desconhecidas e que estão presentes em seu habitat natural ou só crescem em ambientes absolutamente inóspitos, logo, na maioria das vezes o ambiente natural não é reproduzível em laboratório (CHAUDHARY; KHULAN; KIM, 2019).

2.2 SEQUENCIAMENTO DE PRIMEIRA GERAÇÃO

Por muito tempo, os avanços a respeito da ecologia microbiana permaneceram limitados, dado que as técnicas disponíveis ainda se restringiam a compreensão aos níveis morfológico e bioquímico. Nesse contexto, Carl Woese estabeleceu o RNA ribossomal como uma molécula passível de determinar o histórico filogenético dos organismos (WOESE; FOX, 1977), contribuição fundamental para o desenvolvimento de técnicas moleculares para identificação de microrganismos.

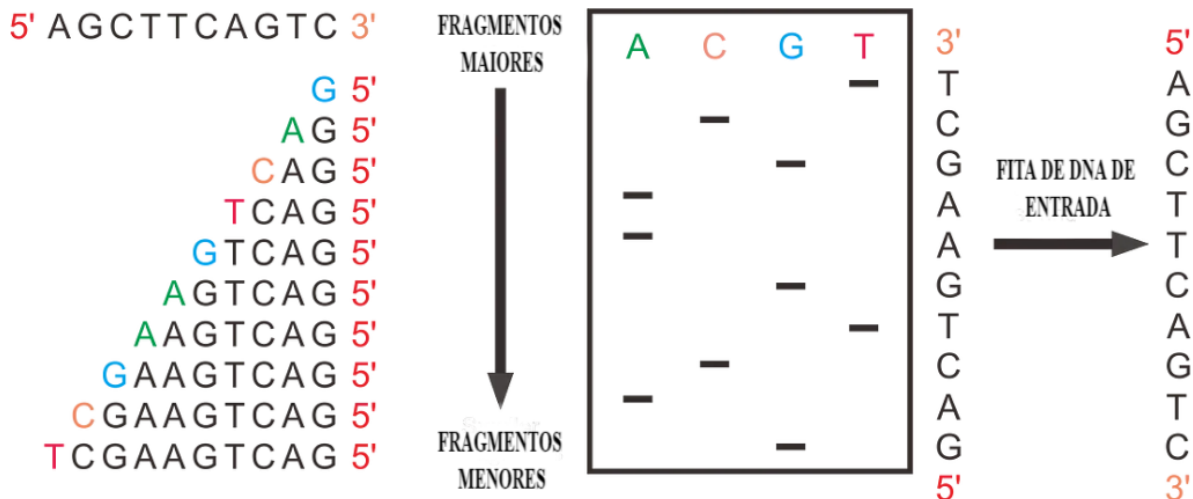
Paralelamente, técnicas de sequenciamento de DNA estavam sendo desenvolvidas, com destaque para o método de terminação de cadeia criado por

Coulson e Sanger (SANGER; NICKLEN; COULSON, 1977), além do método de clivagem química do DNA desenvolvido por Allan Maxam e Walter Gilbert. No entanto, a técnica de Maxam-Gilbert perdeu destaque pela complexidade e uso de reagentes tóxicos. Dessa forma, o padrão ouro do sequenciamento de primeira geração é o método de Sanger, que tem como protagonistas os didesoxinucleotídeos (ddATP, ddGTP, ddCTP e ddTTP) radiomarcados, que não possuem um grupo hidroxila (OH) ligado ao carbono 3' da pentose (Figura 1), característica que impede a adição de um novo nucleotídeo na cadeia de DNA.



Fonte: Adaptado de "Visualizing and Characterizing DNA, RNA, and Protein | Microbiology", ([s.d.])

A técnica utiliza a enzima DNA polimerase, iniciadores e desoxinucleotídeos (dATP, dGTP, dCTP e dTTP), os iniciadores se ligam as fitas simples dos fragmentos de DNA e a polimerase inicia a produção da nova fita inserindo os dNTPs, a síntese será interrompida quando, ao acaso, um ddNTP for inserido na fita, garantindo assim a produção de cadeias de diversos tamanhos. Posteriormente, as moléculas são separadas de acordo com o tamanho por eletroforese em gel, e o resultado da eletroforese indica a sequência de DNA da amostra, dado que, de acordo com a posição das bandas e pelo ddNTP representado é possível saber o posicionamento de cada nucleotídeo na sequência (Figura 2). Através desta técnica foi possível sequenciar o primeiro genoma completo: o do vírus que infecta a *Escherichia Coli* (bacteriófago PhiX174). Tendo em vista essa primeira conquista e as várias outras que a sucederam, o sequenciamento de Sanger é reconhecido pelas contribuições à comunidade científica, e apesar das inúmeras inovações em sequenciamentos, ainda é uma técnica utilizada quando não há necessidade de alto rendimento (SLATKO; GARDNER; AUSUBEL, 2018).

Figura 2 - Leitura do gel após método de Sanger

Fonte: Adaptado de KREMER, 2019.

O método de Sanger foi automatizado e passou por diversas melhorias, sendo uma delas a substituição de radiomarcadores por corantes fluorescentes, tais aperfeiçoamentos resultaram em um crescimento exponencial dos dados gerados, motivando assim a criação de bancos de dados genômicos (e.g. Genbank). Além disso, esse método contribuiu com o sequenciamento do genoma humano e serviu como estímulo para a criação das técnicas de sequenciamento das próximas gerações. (SLATKO; GARDNER; AUSUBEL, 2018).

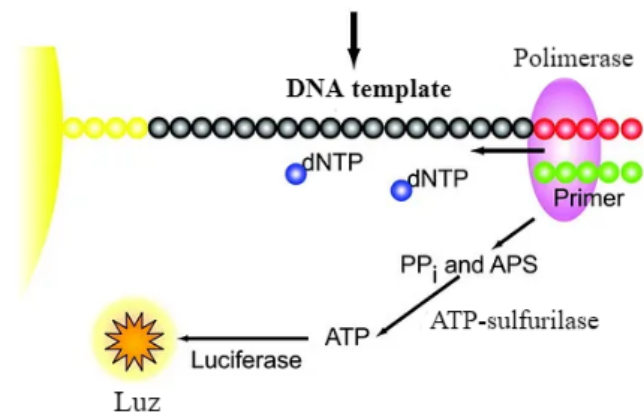
2.3 SEQUENCIAMENTO DE NOVA GERAÇÃO (SEGUNDA GERAÇÃO)

Considerando os inúmeros e contínuos progressos alcançados através do sequenciamento de primeira geração, bem como suas limitações, novas tecnologias surgem para possibilitar um sequenciamento rápido, preciso e de baixo custo. Neste cenário, no início do século XXI, novas tecnologias de sequenciamento emergem, sendo elas denominadas de Sequenciamento de Nova Geração (NGS). Essas tecnologias apresentavam como principal vantagem a capacidade de produzir dados em larga escala, podendo chegar a mais de 1 bilhão de leituras curtas por execução. Tamanha agilidade só é possível devido a paralelização e miniaturização dos processos (METZKER, 2009).

A primeira tecnologia de sequenciamento de segunda geração que obteve

êxito comercial foi o Pirosequenciamento Roche 454, produzido pela *Pyrosequencing AB* - futuramente adquirida pela Roche. Embora a tecnologia tenha sido descontinuada, na época do lançamento revolucionou o mercado como o primeiro sequenciamento de alto rendimento (GIANI et al., 2020). Além disso, apresenta uma metodologia incomum que consiste em uma reação enzimática na qual, durante a replicação do DNA, a cada nucleotídeo incorporado pela DNA polimerase, um pirofosfato é liberado, e este utilizado como substrato da enzima sulfúrilase, que produz uma molécula de ATP. Esta molécula é então degradada pela luciferase, gerando um sinal luminoso (Figura 3). O resultado do sequenciamento é produzido em ciclos e em cada ciclo uma base diferente é incorporada e emite sua luminescência (KULSKI, 2016).

Figura 3 - Reações enzimáticas do pirosequenciamento



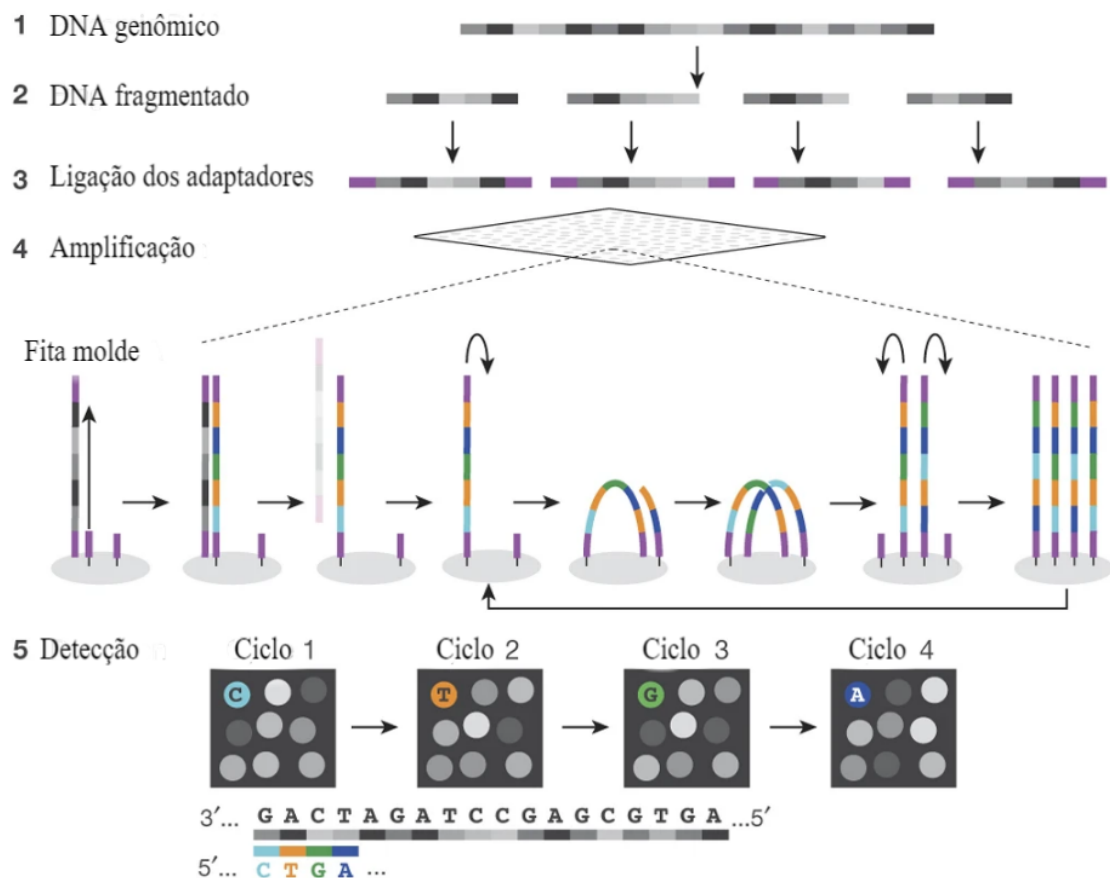
Fonte: Adaptado de KREMER, 2019.

O sequenciamento da Roche 454 foi capaz de produzir leituras com comprimento entre 300-500 nucleotídeos, sendo inferior apenas ao sequenciamento de Sanger que suportava entre 800 e 1000 nucleotídeos, entretanto, a quantidade de reagentes utilizados encareceu a técnica (AHMADIAN; EHN; HOBBER, 2006). Nesse contexto, a Illumina conquistou o mercado com suas plataformas (HiSeq e MiSeq), as quais utilizam o método de sequenciamento por síntese com terminação cíclica reversível (Figura 4) e oferecem maior capacidade de produção de dados por um menor custo.

O fluxo de trabalho dos sequenciadores Illumina inicia com a preparação da biblioteca, onde o DNA é fragmentado e os adaptadores (sequências conhecidas) são inseridos. Em seguida, a amplificação e o sequenciamento acontecem em uma

placa de vidro chamada de *flow cell* que armazena diversos iniciadores complementares aos adaptadores adicionados aos fragmentos de DNA da biblioteca. Dessa forma, os fragmentos ligam-se aos seus complementares na placa e são amplificados em um processo conhecido por amplificação por pontes gerando diversas cópias do mesmo fragmento (*clusters*), o objetivo desta etapa é aumentar a intensidade do sinal luminoso durante o sequenciamento. Após a formação dos *clusters*, a DNA polimerase adiciona o dNTP terminador reversível apropriado, e a cada base incorporada pela DNA polimerase uma fluorescência é emitida, detectada por um sensor e armazenada. Essa etapa ocorre em ciclos e paralelamente em todos os clusters, a cada finalização de ciclo ocorre uma lavagem para que os fluoróforos e terminadores sejam removidos da sequência, permitindo assim a incorporação de outro nucleotídeo. (SHENDURE et al., 2017).

Figura 4 - Representação esquemática do sequenciamento por síntese



Fonte: Adaptado de SHENDURE et al. (2017).

Ainda no âmbito do sequenciamento de segunda geração, a Life Technologies desenvolveu o sequenciador Ion Torrent, sendo o primeiro a não produzir

fluorescência em sua reação. Esta tecnologia, assim como a da Illumina, tem como princípio o sequenciamento por síntese, contudo, a abordagem utilizada é a adição de nucleotídeo único (SNA), logo, o procedimento consiste em realizar uma PCR em emulsão para amplificar os fragmentos de DNA, em seguida, as beads com as cópias dos fragmentos de DNA são mergulhadas separadamente nos poços do microchip onde a reação acontece. Neste momento, os nucleotídeos fluem através dos poços e a cada incorporação realizada pela DNA polimerase, íons H^+ são liberados, alterando o pH da solução. Os sensores abaixo do poço medem a mudança no pH e convertem em voltagem, indicando qual nucleotídeo foi incorporado. (GOODWIN; MCPHERSON; MCCOMBIE, 2016).

Embora o sequenciamento Illumina também contenha falhas (e.g a diminuição da precisão em função do aumento da leitura), ainda apresenta uma taxa de erro inferior a 0,5%, enquanto as plataformas Ion Torrent apresentam uma taxa de erro superior a 1% (LOMAN et al., 2012). Isto posto, dentre as tecnologias de segunda geração, as plataformas Illumina mantém a superioridade no mercado em razão, principalmente, da alta qualidade dos dados e pela diversidade de plataformas apresentadas que podem ser escolhidas de acordo com a aplicação (GOODWIN; MCPHERSON; MCCOMBIE, 2016).

2.4 SEQUENCIAMENTO DE TERCEIRA GERAÇÃO

É visível o progresso das metodologias da segunda geração em relação à primeira, visto que, as problemáticas de custo alto por base sequenciada e tempo longo de sequenciamento em relação a quantidade de leituras geradas por ensaio foram mitigadas, todavia, outras questões surgiram em função da necessidade de amplificação do DNA, na qual o número de cópias de uma molécula é aumentado para que a intensidade do sinal luminoso ou elétrico seja percebido, pois, além de ser uma etapa a mais de processamento das amostras, a PCR tende a aumentar a taxa de erro do sequenciamento (KEOHAVONG; THILLY, 1989).

Nessa conjuntura, a tecnologia SMRT (single-molecule real-time sequencing) da empresa Pacific Biosciences (PacBio) foi criada e baseia-se no sequenciamento em tempo real de uma única molécula, além disso, é capaz de gerar leitura longas e consequentemente suprir algumas das limitações da geração anterior de sequenciadores. O processo de sequenciamento SMRT inicia com o preparo da

amostra, na qual as extremidades de uma dupla-fita de DNA são ligadas formando uma molécula circular com uma polimerase modificada que é fixada no fundo de poços de observação em nanoescala (Zero Mode Waveguides- ZMWs). Então, à medida em que a polimerase incorpora nucleotídeos modificados com fluorescência específica, os sensores detectam e gravam o tempo e intensidade da fluorescência. É o aumento do tempo e da intensidade de fluorescência de determinado nucleotídeo que indica a incorporação deste na sequência (VOELKERDING; DAMES; DURTSCHI, 2009).

Além da PacBio, a Oxford Nanopore Technologies (ONT) também fornece uma tecnologia de sequenciamento SMRT, porém utiliza princípios de eletroquímica e dispensa o uso de nucleotídeos marcados por fluorescência, uma vez que, a identificação da sequência é realizada através da medição de um sinal elétrico correspondente a grupos de nucleotídeos. O sequenciamento acontece em um poro protéico com condutividade elétrica e a medida em que a fita de DNA transpassa o poro, há um distúrbio na corrente iônica diretamente ligada a morfologia do grupo de nucleotídeos que foi translocado, logo, esse grupo de nucleotídeos será determinado de acordo com o distúrbio elétrico captado (KREMKOW; LEE, 2014).

Nesse contexto, as tecnologias da terceira geração superaram a geração anterior no que concerne à possibilidade de produzir leituras longas, além disso, tecnologias como a PacBio Hi-Fi podem ser equiparadas às plataformas Illumina quanto a acurácia, dado que, a fidelidade média das leituras identificadas é de 99,8% com um comprimento médio de aproximadamente 13,5 quilobases (kb), semelhante a acurácia do Illumina NovaSeq (99,9% para leituras de 2×151 pb) e HiSeq 2500 (99,5% para leituras de 2×250 pb) (WENGER et al., 2019). Entretanto, todas as plataformas de sequenciamento apresentam vantagens e desvantagens, logo, é importante entender as metodologias e sequenciadores para assim escolher o que melhor se adequa a finalidade.

Portanto, a evolução das metodologias de sequenciamento trouxeram consequências positivas para as mais diversas áreas de pesquisa genômica, dado que, as tecnologias estão cada vez mais acessíveis e com maior qualidade e capacidade. Transcriptômica, epigenética, proteômica e metagenômica são alguns dos campos beneficiados pelo sequenciamento de nova geração.

2.5 METAGENÔMICA E O AMBIENTE AÉREO

Os avanços das tecnologias de sequenciamento ampliaram as possibilidades dos estudos relacionados à ecologia microbiana, uma vez que não há mais a necessidade de cultivo para a identificação de microrganismos. Nesse contexto, surge a metagenômica, uma técnica que inicialmente tratava-se do sequenciamento direto dos genomas de toda uma comunidade de organismos heterogêneos presentes em dado ambiente, contudo, atualmente a metagenômica dispõe de duas abordagens: *shotgun*, que corresponde ao conceito apresentado a princípio, no qual todos os genomas presentes na amostra são sequenciados; e *amplicon*, na qual o sequenciamento é realizado apenas para determinado alvos utilizados como marcadores moleculares (OULAS et al., 2015).

A metagenômica de *amplicon* dispõe de regiões alvo conhecidas com resolução suficiente para identificar grupos de organismos. Geralmente, o gene que codifica a subunidade 16S do rRNA é utilizado para identificação de bactéria e archaea, o gene que codifica a subunidade 18S do rRNA e a região ITS dos genes para caracterização de eucariotos. Esta conduta é mais utilizada com o objetivo de classificar taxonomicamente e analisar a diversidade dos táxons alvo presentes nas amostras (RAUSCH et al., 2019).

Aprofundando sobre a abordagem utilizando marcadores específicos, as etapas posteriores ao sequenciamento consistem em: pré-processamento das leituras e controle de qualidade, formação de Unidades Taxonômicas Operacionais (*Operational Taxonomic Units* - OTU) ou Variantes de Sequências de Amplicon (*Amplicon Sequence Variants* - ASVs), classificação taxonômica e análises de diversidade. Na atualidade, existem fluxos computacionais de trabalhos conhecidos como *pipelines* (e.g. *Ampliseq/nf-core*) que implementam diversas ferramentas de análises de dados microbianos e assim permitem que as etapas citadas anteriormente sejam realizadas em um único ambiente, garantindo reprodutibilidade nas análises.

No que concerne ao pré-processamento de leituras e controle de qualidade, Fastp pode ser citado como exemplo de ferramenta que executa com eficiência as atribuições de remoção de adaptadores, de leituras curtas e de leituras com baixa qualidade, excesso de bases ambíguas (Ns). Quanto ao agrupamento taxonômico, existem duas modalidades: OTU, que reúne sequências com no mínimo 97% de

similaridade em um mesmo grupo, e ASV, que agrupa as sequências que diferem em no máximo um nucleotídeo. QIIME 2 é um exemplo de ferramenta que utiliza OTUs em seu agrupamento taxonômico, enquanto DADA2 utiliza ASVs, que atualmente é a modalidade de primeira escolha, dado que, apresenta maior sensibilidade, cobertura e reprodutibilidade (CALLAHAN; MCMURDIE; HOLMES, 2017). Posteriormente, para a classificação taxonômica, os principais bancos de dados utilizados são o Silva e Greengenes. O QIIME 2 também realiza classificação taxonômica, e além disso, análise de diversidade alfa e beta, sendo a primeira equivalente à riqueza de espécies dentro da amostra e tem os índices Shannon e Chao como métricas. A diversidade beta, por sua vez, calcula o quão semelhantes as amostras são entre si, e possui a distância de Bray-Curtis e UniFrac como principais métricas.

Em contrapartida, a metagenômica *shotgun* permite que as sequências sejam classificadas a nível de espécies e cepas, viabilizando uma caracterização e exploração mais profunda acerca da diversidade, abundância e funcionalidade do ambiente estudado. Entretanto, a metagenômica de *amplicon* costuma ser mais acessível no âmbito de custos e condições para análise computacional (PÉREZ-COBAS; GOMEZ-VALERO; BUCHRIESER, 2020).

A metagenômica detém uma ampla gama de aplicabilidades, incluindo a área médica com a identificação de patógenos intestinais, descrição funcional da microbiota intestinal, identificação de bactérias resistentes a antibióticos e caracterização de infecções. Na área ambiental, a metagenômica tem sido utilizada para descrever a microbiota dos ecossistemas, na qual destaca-se os estudos relacionados ao solo em virtude das inúmeras descobertas de interesse médico, como por exemplo, os antibióticos Turbomicina que foram isolados a partir de DNA extraído do solo (GILLESPIE et al., 2002).

Portanto, o compartimento ambiental menos explorado pelos estudos metagenômicos é o ar, embora haja comprovação de que os organismos presentes neste ambiente desempenham funções que interferem em processos físico-químicos da atmosfera (DELORT et al., 2010; MORRIS et al., 2011), existem fatores que dificultam sua exploração. Nesse contexto, a abordagem utilizada primordialmente para estudo deste microbioma foi o meio de cultura, e a partir desta técnica um estudo realizado em Pequim foi capaz de relacionar o aumento da concentração de microrganismos com o aumento das atividades antrópicas (FANG et al., 2007). No

entanto, supõe-se que o ambiente aéreo apresenta uma diversidade taxonômica tão abundante quanto o solo, mas, apenas 1% dessa riqueza pode ser representada a partir do cultivo (FRANZETTI et al., 2010), sendo assim, ainda que progressos tenham sido atingidos, a técnica mostrou-se limitada no que se refere a identificação da riqueza do ambiente aéreo em sua totalidade.

Nessa conjuntura, o uso das técnicas de sequenciamento para caracterização do microbioma aéreo expandiu-se e uma nova metodologia de amostragem passou a ser utilizada, na qual bombas com um sistema de sucção a vácuo realizam a coleta do ar que passa por um filtro onde os microrganismos ficam retidos. O uso desta abordagem foi responsável por diversos progressos na metagenômica aérea como o estudo da propagação de partículas pequenas contendo microrganismo patogênicos em evento de poeira africana no Mediterrâneo Oriental (POLYMENAKOU et al., 2008).

Um extenso trabalho realizado com 795 metagenomas coletados durante 1 ano em Cingapura também fez uso da metodologia de amostragem através de um sistema de sucção com filtro, e seus resultados comprovaram que a riqueza e diversidade de microrganismos presentes no ar inquestionavelmente pode ser equiparada a outros microbiomas (e.g solo), além disso, foi capaz de identificar a temperatura como principal fator resultante em mudanças na comunidade microbiana. No entanto, a informação mais significativa relatada é a de que todos os resultados foram pautados em aproximadamente 20% do material sequenciado (GUSAREVA et al., 2019), logo, boa parte dos microrganismos sequenciados no ar ainda não possuem identificação, aspecto que demonstra a importância do presente trabalho, dado que, ainda há muito para ser investigado a respeito da comunidade microbiana do compartimento aéreo.

Todavia, as metodologias tradicionais de coleta utilizando bomba de sucção a vácuo, restringem as possibilidades de amostragem a uma mesma localização, com mesma altitude e fatores abióticos, uma vez que, as bombas são fixas. Nesse sentido, atentando-se aos obstáculos das metodologias tradicionais, o presente projeto busca validar uma metodologia de coleta alternativa utilizando aeronaves não tripuladas (drones), de pequeno porte, atrelada ao uso de coletores para posterior análise metagenômica.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Validação de fluxo de trabalho para estudo microbiológico do ar a partir de amostras coletadas com um drone.

3.2 OBJETIVOS ESPECÍFICOS

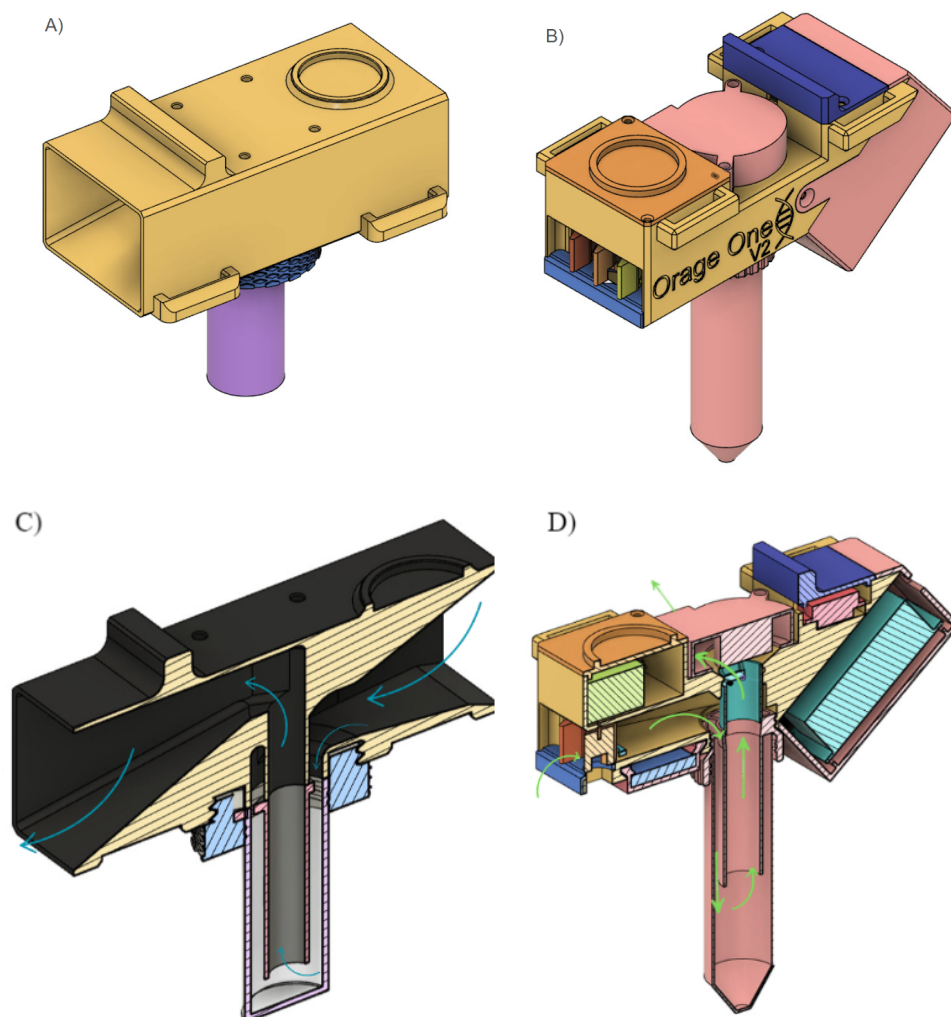
1. Aprimorar o protocolo de coleta utilizando o protótipo do coletor de amostras aéreas.
2. Implementar um procedimento de processamento e preparação das amostras coletadas.
3. Realizar análise da composição taxonômica e diversidade dos microrganismos presentes nas amostras.

4 METODOLOGIA

4.1 COLETORES DE AR E DRONE

Em parceria com a Plataforma Unificada de Modelagem e Manufatura Aditiva (PLUMMA-3D – FIOCRUZ Minas Gerais), foram produzidas duas versões de coletores de ar para serem acoplados ao drone: uma versão com o fluxo de ar passivo e outra com o fluxo de ar ativo (Figura 5). A segunda versão reúne os seguintes componentes: o Arduino Nano que é uma placa microcontroladora responsável por regular a abertura e o fechamento do duto de sucção de ar, um barômetro BMP280, um cooler radial de 35mm, um leitor de cartão SD e uma bateria lipo de 11,1V (450mah ou 850mah). Ambos foram modelados no software CAD utilizando a plataforma integrada de softwares Autodesk Fusion 360, a impressão foi realizada por uma impressora 3D com a tecnologia Fused Deposition Modeling (FDM) e o material escolhido para uso foi o Polietileno tereftalato glicol (PETG).

Figura 5- Ilustração 3D dos coletores



Fonte: Autora, 2022. A) Primeira versão- Fluxo de ar passivo; B) Segunda versão- Fluxo de ar ativo; C) Demonstração de passagem do fluxo de ar na primeira versão do coletor. D) Demonstração de passagem do fluxo de ar na segunda versão do coletor.

O drone utilizado nas coletas é do fabricante DJI modelo Inspire 2 (Figura 6). Para as amostragens com o coletor em que o fluxo de ar é passivo o uso do drone é indispensável, uma vez que este coletor, por não ter um mecanismo ativo de sucção de ar, depende exclusivamente do arraste do drone para gerar o fluxo de ar necessário durante o processo de coleta. Entretanto, o coletor automatizado não depende do arraste do drone, visto que, através do seus componentes e tecnologia, possui a capacidade de sucção, gerando assim um fluxo de ar dentro do tubo acoplado.

Figura 6 - Drone DJI modelo Inspire 2

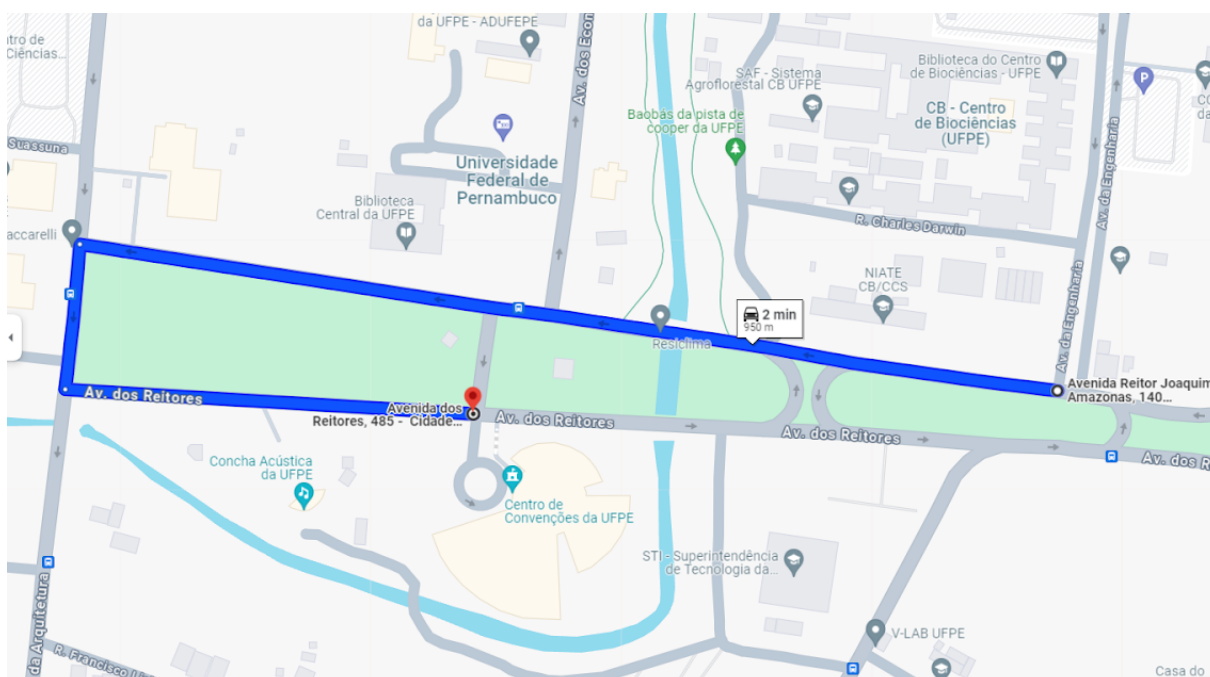


Fonte: DJI, 2016

4.2 COLETA DAS AMOSTRAS DE AR

Não foi possível utilizar a segunda versão do coletor acoplado ao drone, pois sua estrutura bloqueia parcialmente os sensores de altura do drone. Dessa forma, por razões de segurança, a coleta com o coletor de fluxo ativo de ar foi realizada acoplado a um automóvel. A coleta teve uma duração média de aproximadamente 2 minutos e 30 segundos, em diferentes velocidades: 40, 50 e 60 km/h. O trajeto percorrido pelo veículo para cada amostra coletada iniciou na Avenida Reitor Joaquim Amazonas, 140, Cidade Universitária, e foi finalizado na Avenida dos Reitores, 485, Cidade Universitária. (Figura 7).

Figura 7 - Trajeto percorrido para coletar amostras com o coletor de fluxo ativo de ar.

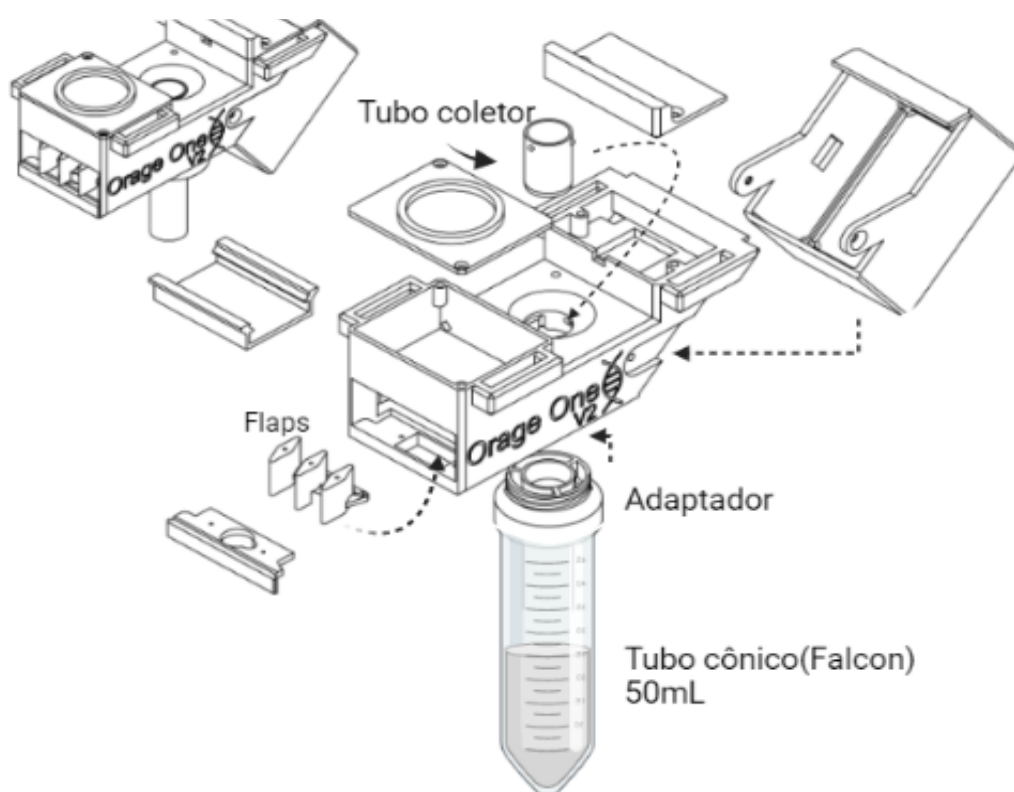


Fonte: Google Maps, 2024

No que concerne ao coletor de fluxo passivo de ar, a coleta ocorreu no distrito de Pirituba, Vitória de Santo Antão-PE onde foram coletadas duplicatas nas alturas de 20 e 50 metros no mesmo ponto, predeterminado pela função Waypoints do drone DJI Inspire 2 que permite o planejamento e pré-programação de um percurso de voo específico. Considerando que o coletor manual permanece com o duto de ar

aberto durante a subida e descida do drone, foram coletadas também amostras de *background* (movimentos de subida e descida com o objetivo de, no momento da análise, excluir possíveis contaminantes das amostras de interesse). Em todas as coletas, tubos cônicos de 50mL contendo 20 mL de tampão PBS 1X com pH 7,4 autoclavado foram utilizados (Figura 8).

Figura 8 - Coletor automatizado com seus elementos expostos (3D)



Fonte: Autora, 2022

Para fins de comparação, avaliação da eficácia e viabilidade da nova abordagem em relação aos métodos convencionais, foi empregado também um método tradicional de coleta, utilizando uma bomba de sucção ar de ar modelo Ambient Fine Dust Sampler IPM-FDS-M2.5/10 (Figura 9). As amostras foram coletadas em fibras de vidro, ao longo de 1h, e ocorreram no teto externo do departamento de Biofísica da Universidade Federal de Pernambuco(UFPE), local onde o dispositivo foi fixado.

Figura 9 - Bomba de sucção utilizada para coleta das amostras



Fonte: Instrumex, 2019.

4.3 EXTRAÇÃO DE DNA

Durante a coleta, o ar entrou em contato com o tampão presente no tubo cônico acoplado ao coletor, momento em que os microrganismos ficaram retidos para posterior extração do DNA. O DNA genômico do tampão foi extraído seguindo o protocolo de extração de Fenol/Clorofórmio (FERGUSON et al., 2019). Portanto, as amostras foram centrifugadas no tubo cônico a 4000 rpm, por 45 minutos e com temperatura de 4°C, após isso, o sobrenadante foi descartado, e o pellet preservado ressuspensionado em 500 µL do tampão de extração SDS Buffer, após homogeneizar vinte vezes. Posteriormente, esse volume foi transferido para microtubo de 1,5mL e colocado em banho-maria a 70° C por 45 min. Após o resfriamento do material, acrescentou-se a solução de extração do DNA contendo fenol tamponado, clorofórmio e álcool isoamílico (25:24:1). Em seguida, o conteúdo foi homogeneizado duas vezes por 30 segundos, e centrifugado a 16.000g, por 5 minutos e a 4 °C. O procedimento anterior removeu proteínas e contaminantes da solução aquosa do DNA, criando assim três camadas: aquosa, proteínas e fenol. Sendo assim, transferimos apenas a fase aquosa (superior) para outro tubo, no qual adicionou-se Isopropanol e foi incubado durante 12h à temperatura ambiente. Nesta etapa, 500 µL da fase aquosa foi transferido para um microtubo de 2,5mL, onde 500 µL de isopropanol foi adicionado. Após a incubação, o volume obtido foi novamente levado à centrífuga por 30 minutos, a 16.000g e 4 °C. Posteriormente, o sobrenadante foi descartado por inversão e o pellet lavado com 500 µL de etanol 70% gelado. Para finalizar, o material foi centrifugado, a 16.000g por 5 minutos e com a temperatura de 4°C, assim o sobrenadante foi descartado por inversão, e aguardamos o pellet secar

para ressuspender em 25 µL de água ultrapura. Vale ressaltar, que toda a manipulação das amostras foi realizada no fluxo laminar localizado na sala de bactéria do NB2 do departamento de microbiologia do Instituto Aggeu Magalhães. Após a extração, as amostras foram submetidas à análise do espectrofotômetro Nanodrop 2000 Thermo Scientific, para quantificação de DNA e verificação dos parâmetros de qualidade e pureza.

4.4 AMPLIFICAÇÃO DAS REGIÕES V3 E V4 DO GENE 16S rRNA

Após a extração, uma PCR foi realizada com o DNA extraído para avaliação da presença de microrganismos nas amostras. Esta amplificação teve como alvos as regiões V3 e V4 do gene que codifica a subunidade 16s do RNA ribossomal. Para esta reação foram utilizados os iniciadores *Forward* e *Reverse* sugeridos pela empresa Illumina (https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). O kit utilizado na reação foi o da Phusion High-fidelity DNA polymerase (New England Biolabs) e os reagentes utilizados possuíam as seguintes concentrações: Tampão Phusion 5x, DNTP a 10mM, iniciadores *Forward* e *Reverse* a 10 µM, MgCl₂ a 50mM e DMSO a 100%; para a produção do master mix adicionamos 5 µl de Tampão Phusion, 1,0 µl de DNTP, 1,0µl do iniciador *Forward*, 1,0 µl do primer *Reverse*, 1,0 µl de MgCl₂, 0,5 µl de Taq Phusion, 1,5 µl de DMSO, 34 µl de água ultrapura e 5µL de DNA, compondo um volume final de 50µL. O termociclador utilizado foi o T100 Bio-Rad, com o programa que consiste em 98°C por 30s, seguido de 30 ciclos de 98°C por 10s, 55°C por 30s e 72°C por 1min30s, finalizando com 72°C por 10 min à 4°C. O controle negativo da reação foram apenas os reagentes sem o DNA, enquanto para o controle positivo foi utilizado o DNA genômico de *Yersinia pestis*, gentilmente cedido pela Coleção de *Yersinia pestis* FIOCRUZ/CYP do Instituto Aggeu Magalhães. O resultado da PCR foi analisado através de uma eletroforese com gel de agarose a 1% corado com *SYBR Green*.

4.5 SEQUENCIAMENTO DE DNA

Após a PCR, as bibliotecas de sequenciamento foram preparadas utilizando o

Nextera XT Index Kit (Illumina, Inc), seguindo instruções do fabricante. A purificação das bibliotecas foi realizada com as esferas magnéticas AMPure™ XP (Agencourt, Beckman-Coulter, USA) e a quantificação concluída com o kit Qubit™ dsDNA HS. Por fim, o sequenciamento foi realizado com o kit v3 de 600 ciclos para bibliotecas do tipo *pair-end*, utilizando a plataforma Illumina MiSeq com objetivo de obter alta profundidade e cobertura de sequenciamento.

4.6 ANÁLISE COMPUTACIONAL

Na primeira etapa da análise dos dados, as leituras do sequenciamento foram submetidas ao pré-processamento pela ferramenta *fastp* (v. 0.23.4), que executa o controle de qualidade, correção de bases, filtragem de sequências e remoção de adaptadores (CHEN et al., 2018). Durante esta etapa foram removidas sequências com tamanho inferior a 15 pb, sequências com mais de 5 representações dúbias de bases (Ns), 30 pb das sequências *forward* na extremidade 5' e 30 pb das sequências *reverse* na extremidade 3'. O resultado do pré-processamento foi compilado em um relatório através da ferramenta MultiQC (v. 1.15) (EWELS et al., 2016).

Subsequentemente, a análise dos dados foi realizada através do fluxo computacional automatizado (*pipeline*) Ampliseq disponibilizado pelo repositório *nf-core* (<https://nf-co.re/>). O repositório *nf-core* utiliza o Nextflow como estruturador do seu fluxo de trabalho. Usar o Nextflow é um diferencial para qualquer *pipeline*, pois, oferece vantagens que facilitam a implementação de ferramentas e manipulação dos dados, além de contribuir com a validação do método científico. Padronização, portabilidade e reprodutibilidade são algumas das características positivas implementadas pelo Nextflow, e com base nessas vantagens o Ampliseq foi o *pipeline* escolhido.

As etapas de análises percorridas pelo *pipeline* foram controle de qualidade de sequências, inferência de ASVs, classificação taxonômica e cálculo das diversidades alfa e beta. Inicialmente, o controle de qualidade no *pipeline* foi realizado pela ferramenta DADA2 (CALLAHAN et al., 2016), onde as sequências *forward* e *reverse* foram truncadas em 200pb e 120pb, respectivamente, e 3 foi o limite máximo de erros esperados determinado. Em seguida, DADA2 também realizou a inferência de ASVs no modo "pooled", prezando pela sensibilidade do

cálculo, independentemente da demanda computacional requisitada e classificação taxonômica, utilizando o banco de dados SILVA(v. 138) (QUAST et al., 2013). Por fim, com QIIME 2 foram produzidos as tabelas e gráficos que expressam os índices de alfa e beta diversidade.

No entanto, durante a análise dos resultados gerados pelo *Ampliseq* verificou-se uma eliminação excessiva de leituras após o *merged* realizado pelo DADA2 e uma contaminação considerável nas amostras controles (PBS, Mix e Mix2). Buscando manter uma quantidade suficiente de leituras para análise, as sequências passaram a ser tratadas como *single-end* e foram truncadas em 200pb. A partir da tabela de abundância de ASVs geradas pelo DADA2, foram produzidos no R (v.4.3.3) histogramas da abundância relativa das amostras contaminadas. Com base nos histogramas, foi possível determinar a distribuição das ASVs e estipular uma abundância máxima permitida nos controles para assim realizar a descontaminação das amostras.

Visando contornar as limitações encontradas no *Ampliseq* durante o processamento das leituras, foi utilizado o pacote *Phyloseq* R (MCMURDIE; HOLMES, 2013) para realizar descontaminação e as análises de diversidade alfa e beta. Todavia, para fazer uso do pacote *Phyloseq* (v.1.46.0) faz-se necessário criar uma objeto composto por uma tabela de metadados, de classificação taxonômica e de abundância de ASVs, dessa forma, esses arquivos foram aproveitados da análise realizada anteriormente no *Ampliseq*. Após a compilação, o objetivo do *phyloseq* foi processado novamente para excluir as ASVs indesejadas e executar as análises de diversidade alfa, utilizando os índices de Shannon e Chao como métricas. Além disso, foi produzido um gráfico de abundância de táxons e calculada a diversidade beta, utilizando a dissimilaridade de Bray-Curtis.

5 RESULTADOS

5.1 COLETORES DE AR E DRONE

Para ambos coletores, com fluxo de ar passivo e ativo (Figura 10), limitações e problemas foram identificados. A segunda versão do coletor possui capacidade de sucção ativa, logo, não precisaria do arraste do drone ou de um veículo para gerar o fluxo de ar dentro do tubo cônico, entretanto, nos testes realizados com o coletor

automatizado estático não foi possível detectar a presença de DNA após a extração. Dessa forma, constatamos que o fluxo de ar gerado apenas pela tecnologia do coletor mostrou-se insuficiente para coleta de material particulado do ar. Além disso, a estrutura física deste coletor impediu o seu uso acoplado ao drone, uma vez que, havia um bloqueio parcial dos sensores de altitude localizados na parte posterior do drone (Figura 11). Considerando esses problemas, o coletor foi utilizado fixado a um veículo e a partir daí obtivemos resultados positivos de DNA após a extração.

Figura 10 - Coletor manual e automatizado, respectivamente.



Fonte: Autora, 2024

Figura 11 - Localização e alcance de sensores no drone DJI Inspire 2.



Fonte: Adaptado de “Inspire 2 - DJI”, [s.d.]. No canto superior direito o coletor automatizado está acoplado ao drone e bloqueando alcance dos sensores.

5.2 EXTRAÇÃO DE DNA E AMPLIFICAÇÃO DAS REGIÕES V3 E V4 DO GENE 16S rRNA

Após a coleta, as amostras foram extraídas e em seguida quantificadas com NanoDrop® 2000 (Thermo Scientific). A razão 260/280 da maioria das amostras ficou abaixo do valor de 1,8 (Tabela 1), contudo, a leitura do Nanodrop possui interferências quando se trata de amostras com baixa biomassa, dessa forma, as amostras foram conduzidas a PCR.

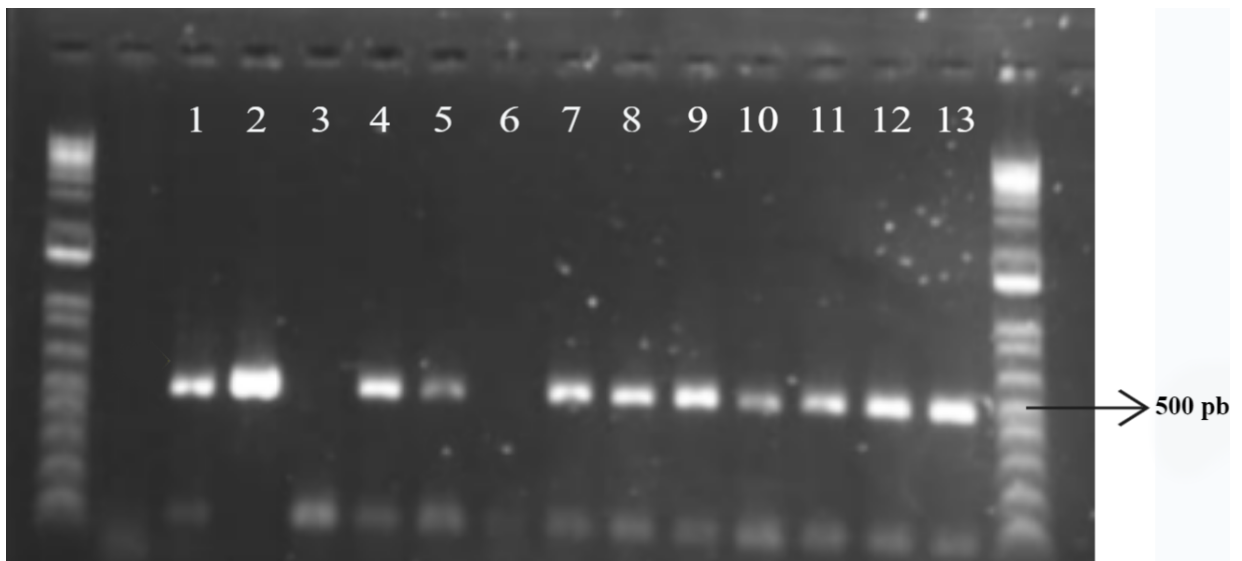
Tabela 1 - Quantificação de DNA das amostras no NanoDrop® 2000

Amostra	Concentração (ng/ µl)	260/280
Controle negativo – Extração (PBS)	11,6	1,44
20 metros – background	28,1	1,46
Coletor Manual – 20 metros	7,4	1,35
Coletor Manual – 20 metros	24,9	1,41
50 metros – background	47,9	1,43
Coletor Manual – 50 metros	11,1	1,28
Coletor Manual – 50 metros	26,5	1,39
Coletor Automatizado – 40km/h	36,5	1,46
Coletor Automatizado – 40km/h	12,0	1,57
Coletor Automatizado – 50km/h	30,0	1,46
Coletor Automatizado – 60km/h	8,4	1,31
Bomba de sucção – Controle negativo	49,3	1,47
Bomba de sucção – filtro	80,5	1,47
Bomba de sucção – filtro	57,2	1,49

Fonte: Autora, 2024

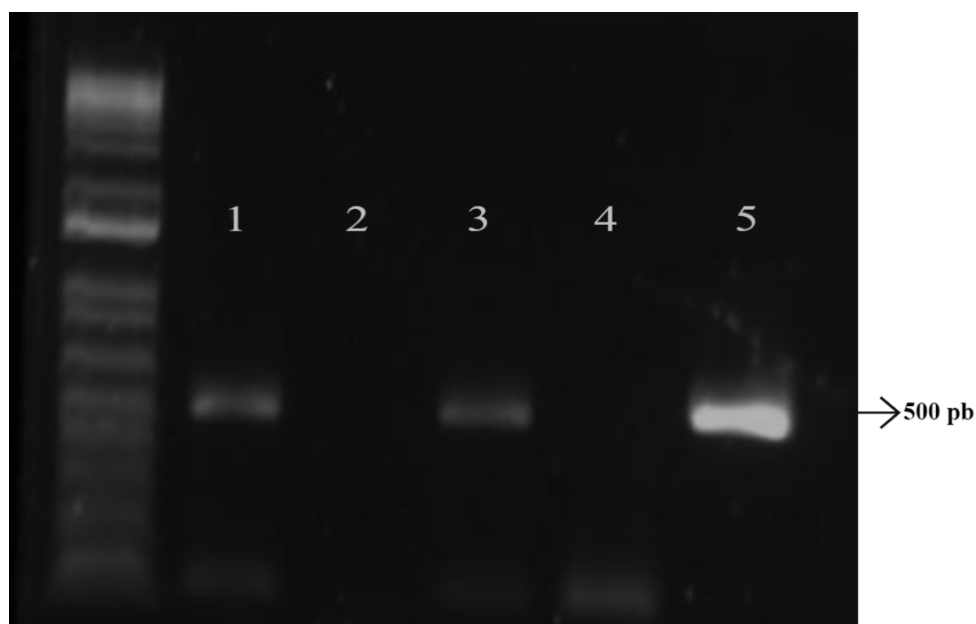
Para amplificação das amostras, a PCR que teve como alvo as regiões V3 e V4 do gene 16S rRNA foi realizada e seu produto foi avaliado através de uma eletroforese em gel de agarose a 1%, na qual confirmou-se a amplificação do fragmento de DNA das regiões definidas (Figura 12 e 13).

Figura 12 - Eletroforese após PCR confirmando amplificação das regiões V3 e V4 do 16S rRNA para amostras dos coletores.



Fonte: Autora, 2024. 1) Controle negativo extração (PBS); 2) Controle positivo PCR (*Yersinia*); 3) Controle negativo PCR (Mix); 4) 20 metros – background; 5 e 6) Coletor manual – 20 metros; 7) 50 metros – background; 8 e 9) Coletor manual – 50 metros; 10 e 11) Coletor automatizado – 40km/h; 12) Coletor automatizado – 50km/h; 13) Coletor automatizado – 60km/h

Figura 13 - Eletroforese após PCR confirmando amplificação das regiões V3 e V4 do 16S rRNA para amostras da bomba de sucção.

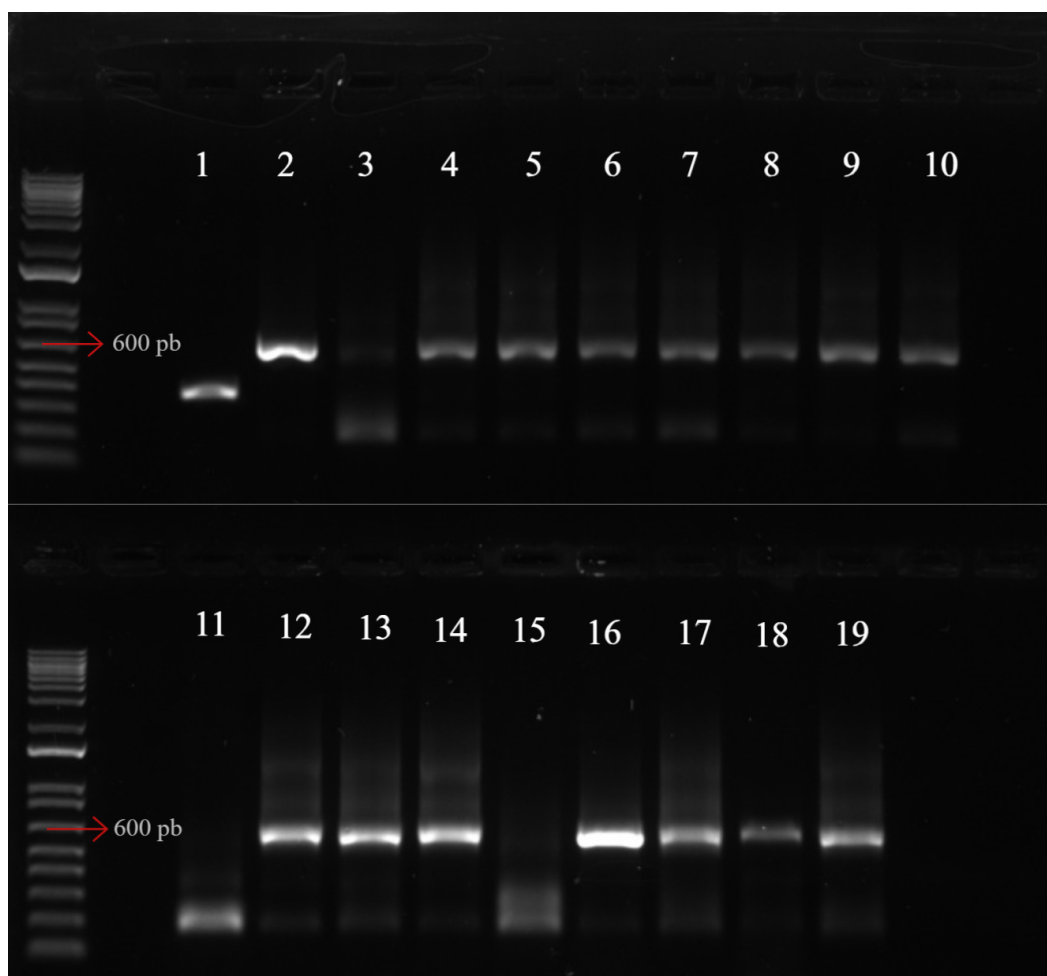


Fonte: Autora, 2024. 1) Bomba de sucção – Controle negativo extração; 2 e 3) Bomba de sucção – filtro; 4) Controle negativo PCR (Mix2); 5) controle PCR (*Yersinia*).

5.3 SEQUENCIAMENTO DE DNA

Para realização do sequenciamento, o produto da PCR foi purificado com esferas magnéticas e em seguida os fragmentos de DNA foram conectados aos adaptadores para construção das bibliotecas. Para quantificar as bibliotecas, o método escolhido foi o fluorimétrico, com o kit Qubit™ dsDNA HS. A quantificação mostrou-se satisfatória e a eletroforese em gel de agarose a 1% sinalizou que a união com os adaptadores foi bem sucedida e os fragmentos possuíam em torno de 600 pb (Figura 14). Por fim, o preparo das bibliotecas foi finalizado conforme instruções do fabricante Illumina e o sequenciamento foi realizado na plataforma Illumina Miseq.

Figura 14 – Eletroforese em gel de agarose 1% das bibliotecas.

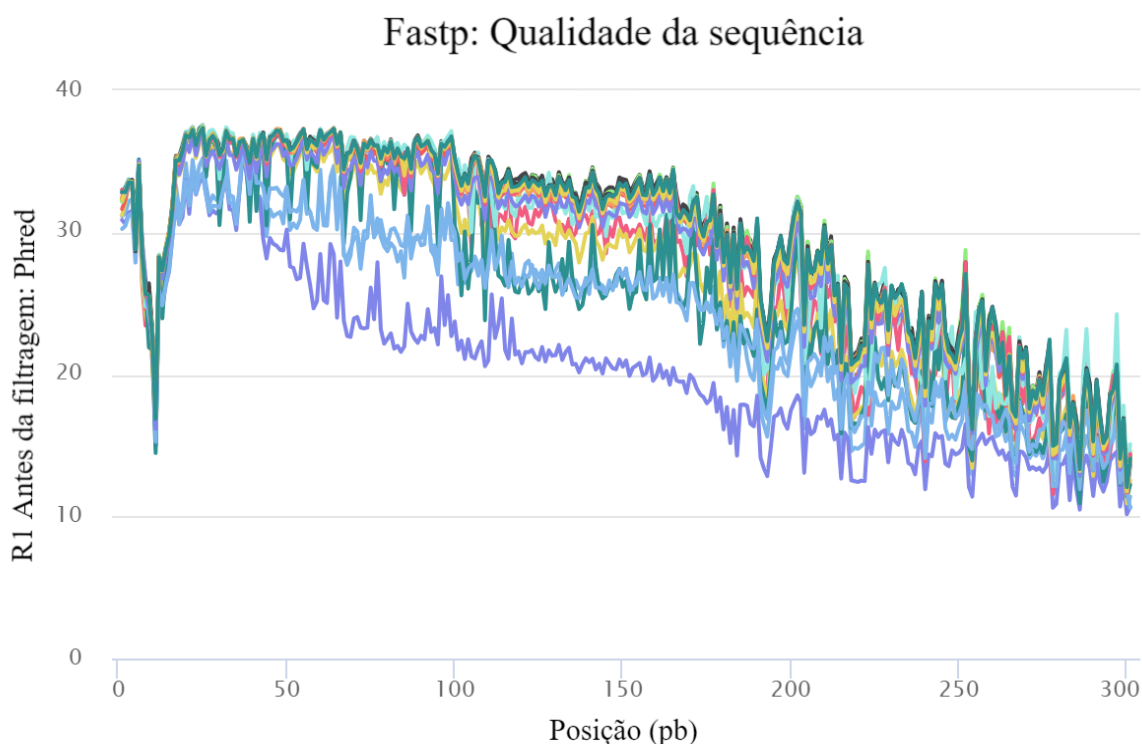


Fonte: Autora, 2024. 1) Controle QuantPCR; 2) Controle positivo PCR (*Yersinia*); 3) Controle negativo PCR (Mix); 4) Controle negativo extração (PBS); 5) Coletor automatizado – 60km/h; 6) Coletor automatizado – 50km/h; 7) Coletor automatizado – 40km/h; 8) Coletor automatizado – 40km/h; 9) 20 metros – background; 10) Coletor manual – 20 metros; 11) Coletor manual – 20 metros; 12) 50 metros – background; 13) Coletor manual – 50 metros; 14) Coletor manual – 50 metros; 15) Controle negativo PCR (Mix2); 16) Controle positivo PCR (*Yersinia*); 17) Bomba de sucção - Controle negativo; 18) Bomba de sucção - filtro; 19) Bomba de sucção - filtro.

5.4 ANÁLISE COMPUTACIONAL

O sequenciamento gerou 20.141.040 leituras agrupadas em arquivos fastq utilizados nas análises posteriores. O pré-processamento das sequências foi realizado com *fastp*, e demonstrou que o início das leituras *forward* e *reverse* apresentam uma qualidade muito baixa, tendo isso em vista, foram removidos 30 pb da porção inicial de R1 e R2 (Figura 15 e 16), além disso, sequências com tamanho inferior a 15 pb ou mais de 5 representações dúbias de bases (Ns) também foram eliminadas.

Figura 15 – Qualidade média das sequências *forward* (R1) antes e depois do pré-processamento no *fastp*.



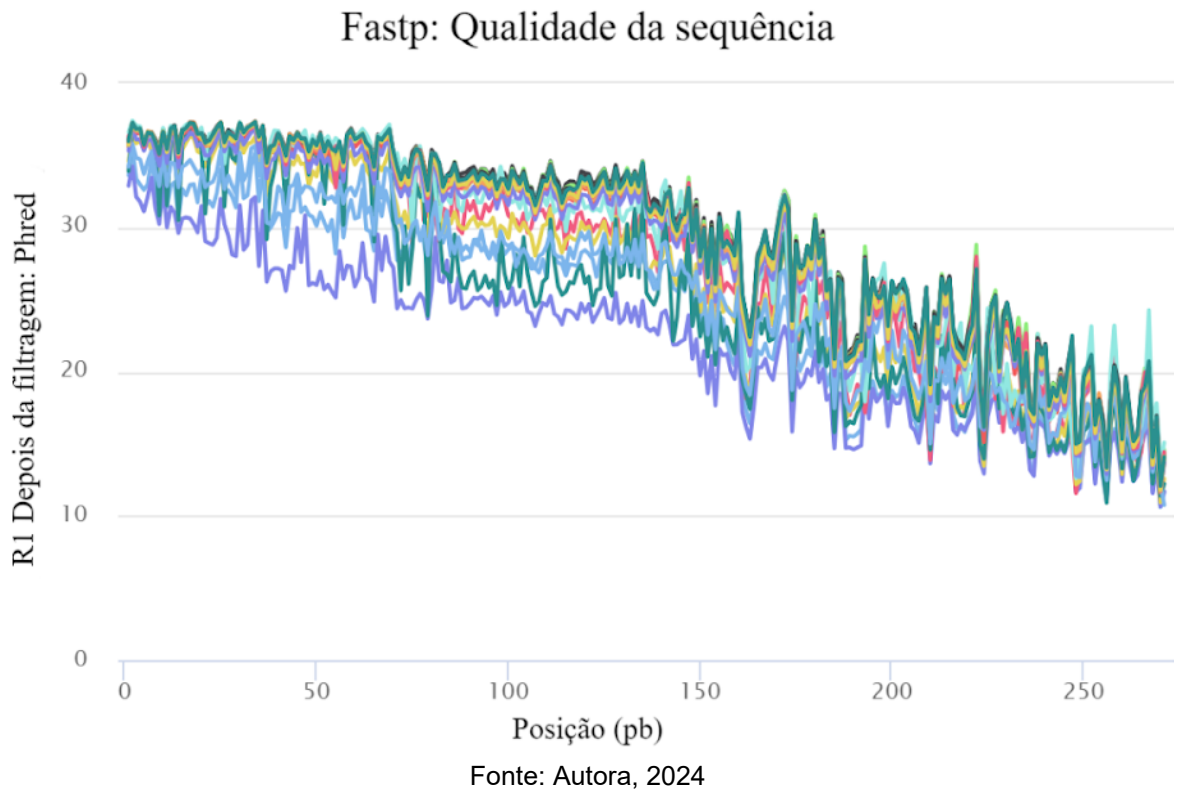
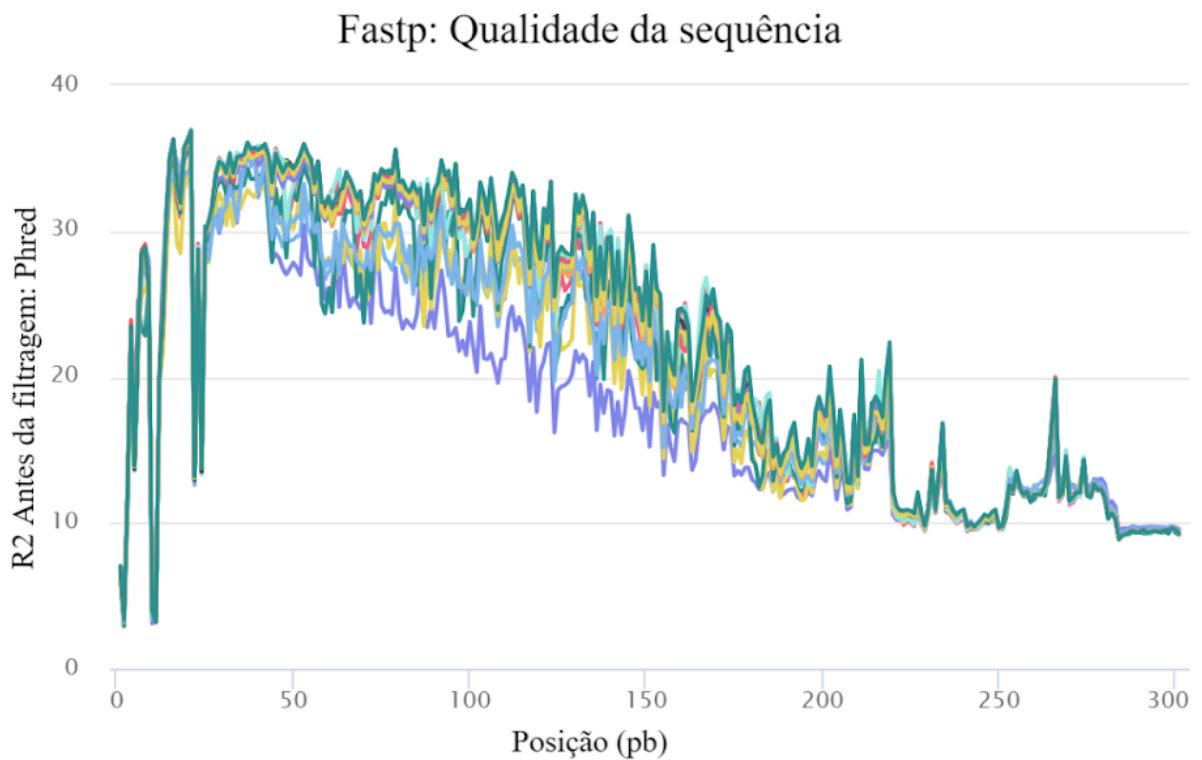
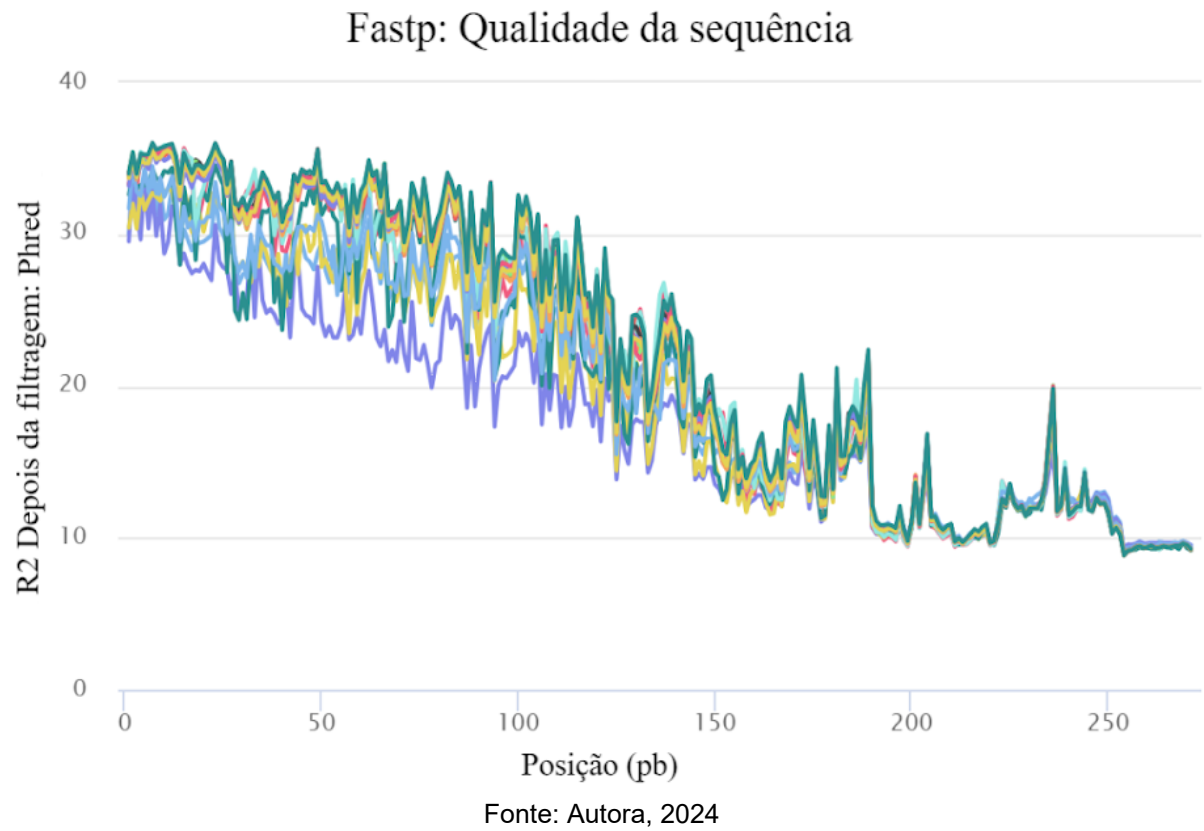


Figura 16 – Qualidade média das sequências *reverse* (R2) antes e depois do pré-processamento no fastp.





Após o pré-processamento, 6.057.295 leituras foram utilizadas como dado de entrada para o pipeline Ampliseq. Nesse momento, as sequências forward e reverse foram truncadas em 200pb e 120pb, respectivamente. O limite máximo de erros determinado foi 3, e a inferência de ASVs foi realizada no modo “pooled”. No entanto, ao observar o output do DADA2 (Tabela estatística), constatou-se que a abordagem deveria ser modificada, dado que, durante a etapa de união (merged) entre as leituras forward e reverse, muitas sequências foram perdidas (Tabela 2). Além disso, a quantidade de leituras iniciais presentes nas amostras Mix, Mix2 e PBS revelou uma contaminação a ser reduzida.

Tabela 2 – Quantidade de leituras que permaneceram após as fases do controle de qualidade do DADA2/Ampliseq para biblioteca *paired-end*.

Amostras	Entrada	Filtrado	Denoised F	Denoised R	Merged	Não quimeras
20 metros – BG	440342	310428	309495	309981	248	247

Coletor M. – 20 metros	362211	250978	250795	250907	103	103
Coletor M. – 20 metros	32989	3715	3681	3689	822	822
50 metros – BG	388878	292984	292372	292675	538	535
Coletor M. – 50 metros	458467	330176	329792	329981	673	670
Coletor M. – 50 metros	418168	304153	302847	303546	302	301
Bomba de sucção – filtro	695	257	257	257	0	0
Bomba de sucção – filtro	330632	229141	228721	228874	334	331
Bomba de sucção – Branco	180056	83600	83475	83490	231	229
Branco PCR (Mix)	271230	87590	87540	87533	268	268
Branco PCR (Mix2)	97801	11476	11465	11462	119	119
Controle – Extração (PBS)	433934	314581	313996	314300	617	616
Coletor A. – 40km/h	533793	338350	337912	338101	615	615
Coletor A.– 40km/h	523514	358040	357781	357870	482	480
Coletor A. – 50km/h	425412	292523	291884	292134	185	184
Coletor A. – 60km/h	560695	390199	389409	389795	1088	1086
Controle PCR (Yersinia)	362211	250978	250795	250907	103	103
Controle PCR (Yersinia)	236267	146576	146542	146521	65	65

Fonte: Autora, 2024. BG= background; Coletor M.=Coletor Manual; Coletor A. = Coletor Automatizado. As colunas representam = Amostras- Nome das amostras utilizadas; Entrada:

Quantidade de leituras que deram entrada no DADA2; Filtrado: Quantidade de leituras que restaram após truncamento e filtragem de acordo com a quantidade de erros esperados determinada; Denoised F: Leituras Forward que restaram após correção de erros de sequenciamento baseado em modelos de erro do DADA2; Denoised R: Leituras Reverse que restaram após correção de erros de sequenciamento baseado em modelos de erro do DADA2; Merged: Leituras que foram capazes de mesclar Forward com Reverse; Não Quimeras: Leituras que restaram após remoção de quimeras.

As sequências *forward* e *reverse* de cada amostra foram então reunidas em um só arquivo e passaram a ser tratadas como *single-end*. A ferramenta *fastp* foi novamente utilizada e eliminou-se 30pb da porção 5' e 30pb da porção 3' das leituras, entretanto os outros parâmetros permaneceram iguais. Em seguida, o produto deste processamento retornou ao *Ampliseq*, e o único parâmetro acrescentado foi o - - *single-end*, além do truncamento das sequências em 200pb. Essa abordagem mostrou-se positiva com relação a quantidade de leituras que passaram pela filtragem do DADA2 para as próximas etapas (Tabela 3). O DADA2 recebeu 12.093.270 sequências de *input*, das quais 4.057.405 foram aprovadas na filtragem de qualidade e seguiram para inferência de ASVs, onde 2123 foram identificadas.

Tabela 3 – Quantidade de leituras que permaneceram após as fases do controle de qualidade do DADA2/*Ampliseq* para biblioteca single-end.

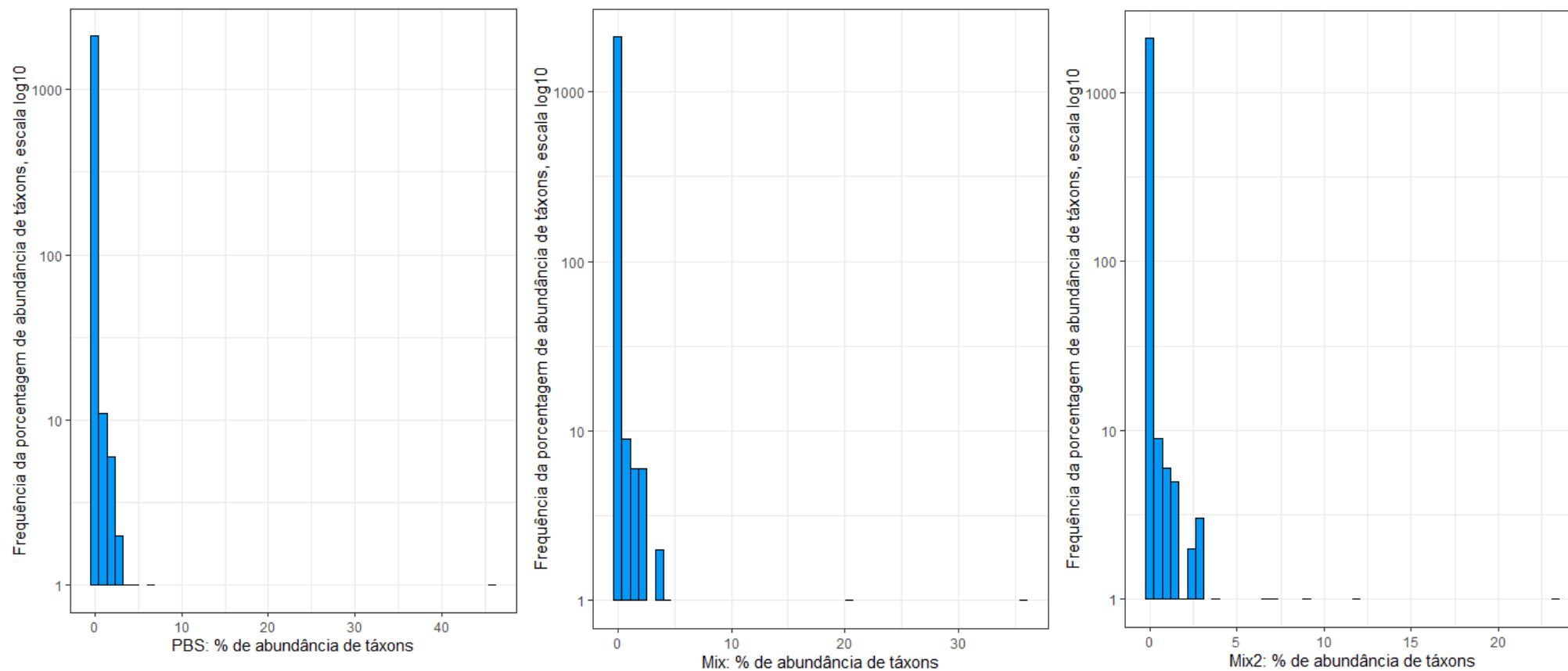
Amostras	Entrada	Filtrado	Denoised	Não quimeras
20 metros – BG	877186	331424	329788	302447
Coletor M. – 20 metros	800361	322996	322060	307166
Coletor M. – 20 metros	58364	5269	5152	4765
50 metros – BG	776491	313473	312268	288573
Coletor M. – 50 metros	913760	354691	354101	328662
Coletor M. – 50 metros	835073	326024	323436	296839

Bomba de sucção – filtro1	1397	300	300	270
Bomba de sucção – filtro2	659225	243289	242635	224655
Bomba de sucção – Branco	358353	95726	95573	88466
Branco PCR (Mix)	469993	100420	100272	98733
Branco PCR (Mix2)	235075	15227	15107	14338
Controle – Extração (PBS)	863509	336631	335615	309799
Coletor A. – 40km/h	1046814	366928	366029	344961
Coletor A.– 40km/h	1041607	381635	380911	363597
Coletor A. – 50km/h	840335	311113	309781	287321
Coletor A. – 60km/h	1118595	416057	414425	379599
Controle PCR (<i>Yersinia</i>)	724515	265509	265465	264965
Controle PCR (<i>Yersinia</i>)	472617	152516	152463	152249

Fonte: Autora, 2024. BG= background; Coletor M.=Coletor Manual; Coletor A. = Coletor Automatizado. As colunas representam = Amostras- Nome das amostras utilizadas; Entrada: Quantidade de leituras que deram entrada no DADA2; Filtrado: Quantidade de leituras que restaram após truncamento e filtragem de acordo com a quantidade de erros esperados determinada; Denoised: Leituras que restaram após correção de erros de sequenciamento baseado em modelos de erro do DADA2; Não Quimeras: Leituras que restaram após remoção de quimeras.

Observando a Tabela 3, é possível perceber que as amostras Mix, Mix2 e PBS (Controles negativos) apresentam uma quantidade expressiva de leituras, sinalizando uma contaminação. Dessa forma, a tabela de abundância de ASVs gerada pelo DADA2 foi direcionada ao R (v.4.3.3) e histogramas da abundância relativa das amostras controle (Mix, Mix2 e PBS) foram construídos (Figura 17).

Figura 17– Histogramas da frequência de porcentagem de abundância relativa das amostras controles: PBS, Mix e Mix2.



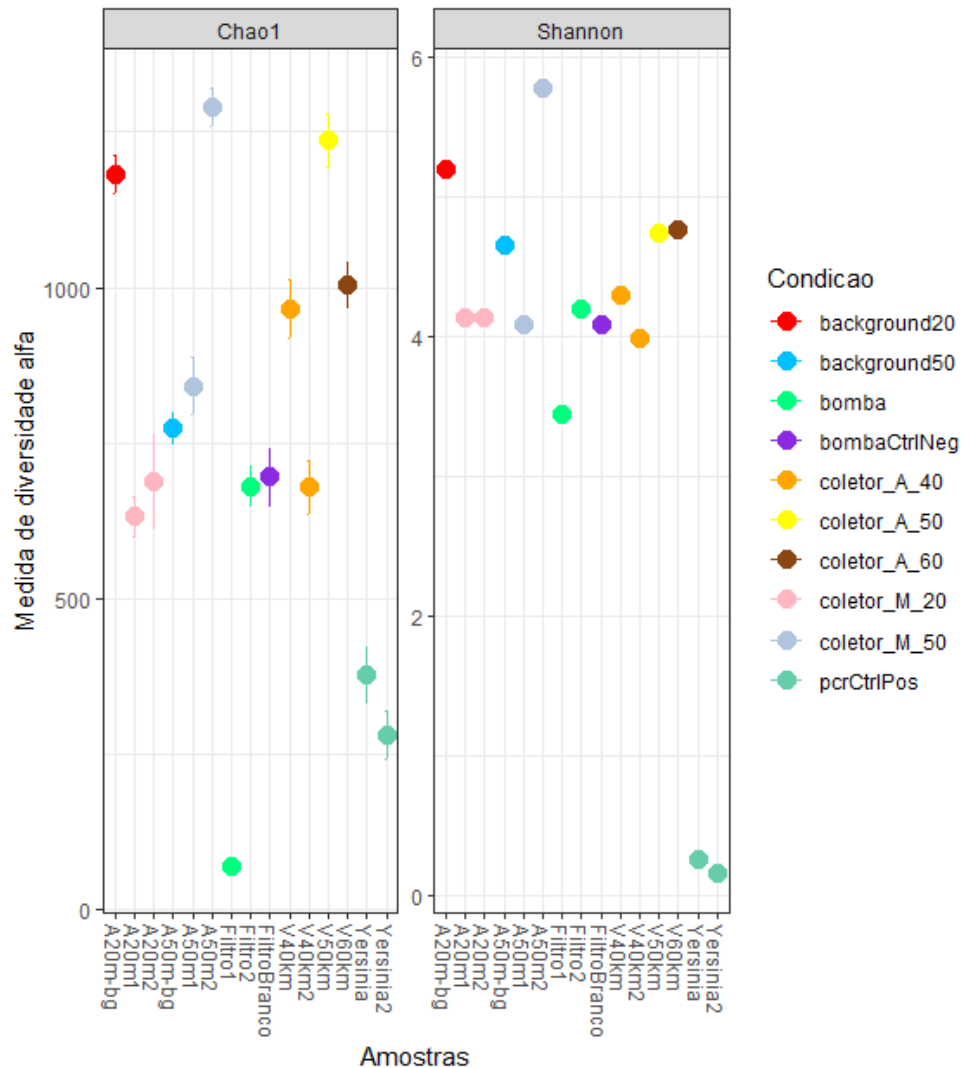
Fonte: Autora, 2024

A partir dos histogramas ficou evidente que a frequência de ASVs com abundância elevada é baixa, tornando factível eliminar as de porcentagem elevada, visto que não são maioria, pois, a maioria das ASVs possui de 0 a 2% de abundância de táxons. Sendo assim, decidiu-se que ASVs das amostras controle com porcentagem de abundância maior que 3% deveriam ser eliminadas de todas as amostras.

Subsequentemente, a descontaminação foi realizada com o pacote *Phyloseq* R, e para produzir o objeto *phyloseq*, as tabelas de metadados, de classificação taxonômica e de abundância de ASVs foram importadas do *Ampliseq* e compiladas. Utilizando o objeto *phyloseq*, as ASVs foram eliminadas de todas as amostras seguindo as condições determinadas anteriormente pela análise dos histogramas, desse modo, 15 ASVs foram removidas permanecendo 2108 para análise de diversidade alfa e beta, além disso, as amostras Mix, Mix1 e PBS que foram utilizadas para a remoção da contaminação, não fizeram parte das análises posteriores.

As métricas utilizadas para calcular a diversidade alfa foram o índice de Shannon e Chao1 (Figura 18), ambos indicam a riqueza de cada amostra, contudo, Shannon também representa a uniformidade dessas espécies dentro de cada amostras, enquanto o índice de Chao1 destaca a presença de espécies representadas por um ou dois indivíduos na amostra (espécies raras). Com base nesses gráficos, foi possível perceber que a riqueza das amostras de background 20 e 50 metros (Subida e descida até a altura indicada) apresentam diversidade semelhante às amostras de 20 e 50 metros, o que pode ser explicado pela estrutura do coletor manual, onde o duto pelo qual o fluxo de ar passa sempre está aberto, problemática que foi resolvida no coletor automatizado. Além disso, as amostras coletadas em altitude (50 metros) e velocidade maior (50 e 60 km/h) tem propensão a uma riqueza maior. Presumivelmente, a riqueza sofre mais influência da altitude e velocidade do que do método de coleta, pois as amostras coletadas nos dois coletores apresentaram resultados semelhantes.

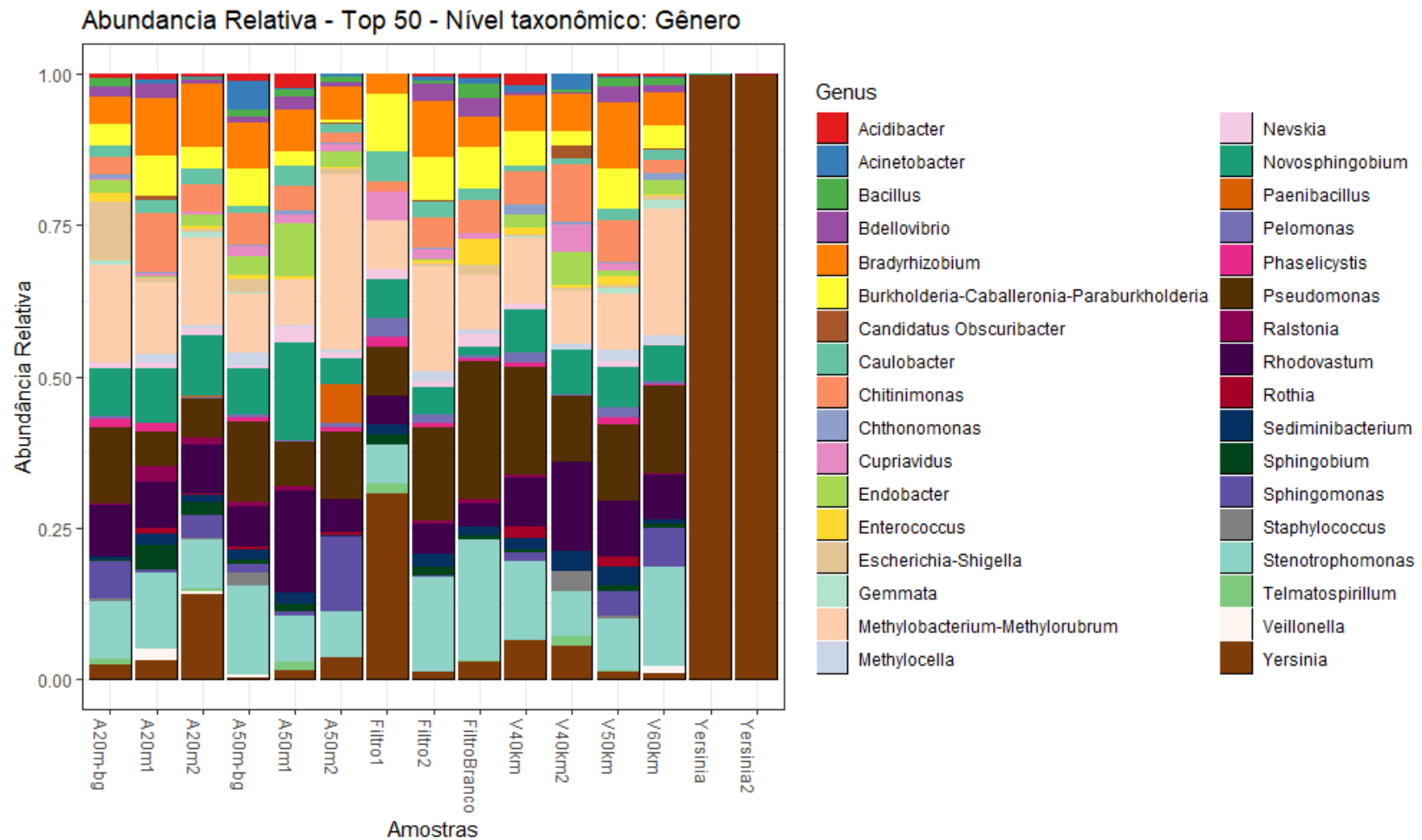
Figura 18 – Diversidade alfa representada pelos índices de Shannon e Chao1.



Fonte: Autora, 2024. No eixo x as amostras estão representadas (A20m-bg= amostra coletada com o drone e coletor manual à 20m de altura do solo (subida e descida); A20m1 e A20m2 = duplicatas coletadas com o drone e coletor manual à 20m de altura do solo; A50m-bg= amostra coletada com o drone e coletor manual à 50m de altura do solo (subida e descida); A50m1 e A50m2 = duplicatas coletadas com o drone e coletor manual à 50m de altura do solo; Filtro 1 e 2 = amostras coletadas em duplicata utilizando filtros e bomba de sucção; FiltroBranco= Controle negativo do filtro; V40km e V40km2 = amostras coletadas com coletor automatizado acoplado à um carro em movimento em velocidade de 40km/h; V50km = amostra coletada com coletor automatizado acoplado à um carro em movimento em velocidade de 50km/h; V60km = amostra coletada com coletor automatizado acoplado à um carro em movimento em velocidade de 60km/h; *Yersinia* e *Yersinia2* = controle positivo da PCR.)

O gráfico de abundância relativa de táxons é um recurso visual importante para compreender a composição e a diversidade das comunidades microbianas nas amostras, e assim fornecer percepções sobre as interações e a ecologia microbiana. No entanto, devido à limitação de resolução no sequenciamento, as amostras apresentadas não permitiram a identificação de padrões distintos (Figura 19). No máximo, foi possível observar uma semelhança geral entre os táxons presentes em todas as amostras.

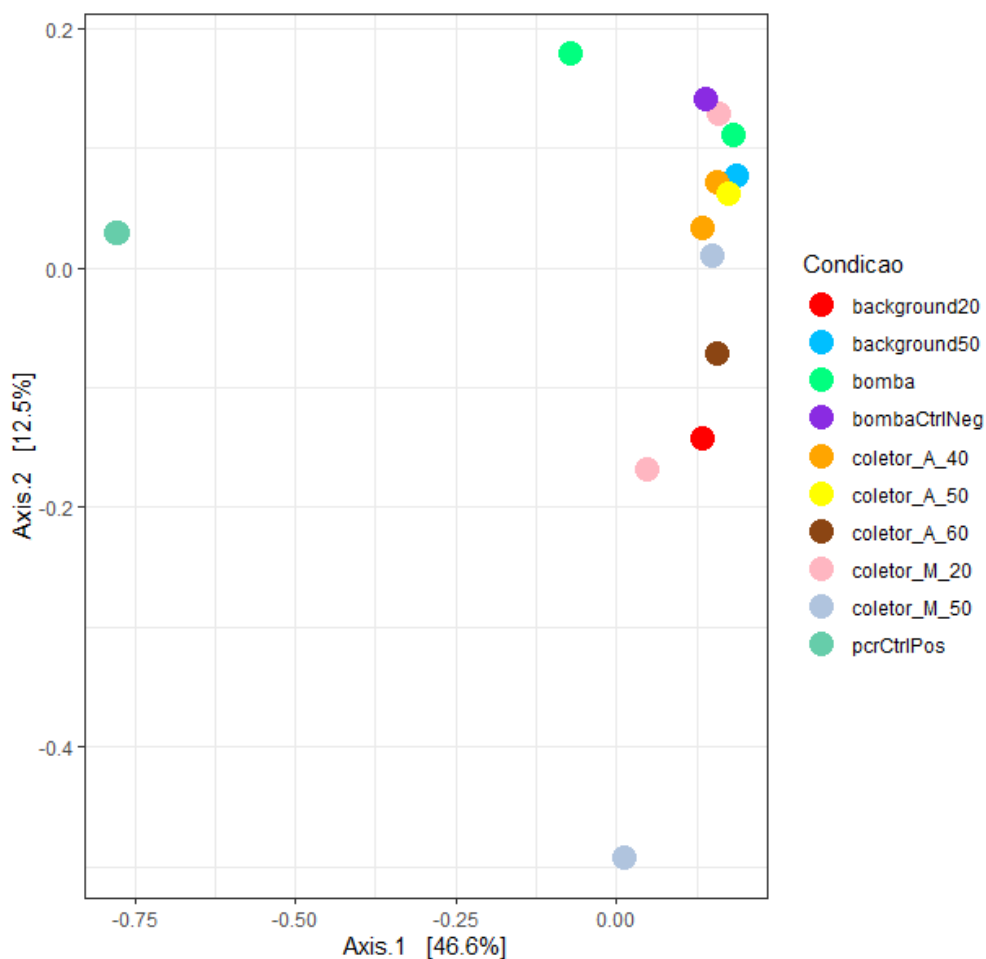
Figura 19– Abundância relativa de táxons



Fonte: Autora, 2024. No eixo x as amostras estão representadas (A20m-bg= amostra coletada com o drone e coletor manual à 20m de altura do solo (subida e descida); A20m1 e A20m2 = duplicatas coletadas com o drone e coletor manual à 20m de altura do solo; A50m-bg= amostra coletada com o drone e coletor manual à 50m de altura do solo (subida e descida); A50m1 e A50m2 = duplicatas coletadas com o drone e coletor manual à 50m de altura do solo; Filtro 1 e 2 = amostras coletadas em duplicata utilizando filtros e bomba de sucção; FiltroBranco= Controle negativo do filtro; V40km e V40km2 = amostras coletadas com coletor automatizado acoplado à um carro em movimento em velocidade de 40km/h; V50km = amostra coletada com coletor automatizado acoplado à um carro em movimento em velocidade de 50km/h; V60km = amostra coletada com coletor automatizado acoplado à um carro em movimento em velocidade de 60km/h; Yersinia e Yersinia2 = controle positivo da PCR.)

Para o cálculo da diversidade beta a métrica utilizada foi a dissimilaridade de Bray-Curtis, que indica o quão semelhantes ou diferentes as amostras são entre si. Nessa perspectiva, o gráfico indica que existem mais semelhanças do que diferenças entre as amostras (Figura 20).

Figura 20– Diversidade beta representada pelo índice de Bray-curtis.



Fonte: Autora, 2024. Os pontos representam as condições (background20 = amostra coletada com o drone e coletor manual à 20m de altura do solo (subida e descida); background50 = amostra coletada com o drone e coletor manual à 50m de altura do solo (subida e descida); bomba = amostras coletadas em duplicatas utilizando filtros e bomba de sucção; bombaCtrlNeg = controle negativo da bomba; coletor_A_40 = amostras coletadas com coletor automatizado acoplado à um carro em movimento com velocidade de 40km/h; coletor_A_50 = amostras coletadas com coletor automatizado acoplado à um carro em movimento com velocidade de 50km/h; coletor_A_60 = amostras coletadas com coletor automatizado acoplado à um carro em movimento com velocidade de 60km/h; coletor_M_20 = duplicatas coletadas com o drone e coletor manual à 20m de altura do solo; coletor_M_50 = duplicatas coletadas com o drone e coletor manual à 50m de altura do solo; pcrCtrlPos = controle positivo da PCR.)

6 DISCUSSÃO

Com o avanço tecnológico das metodologias de sequenciamento, juntamente com a diminuição do custo para executá-las, houve uma aplicação crescente da abordagem conhecida por metagenômica para estudar os microrganismos, como eles interagem entre si e quais os papéis funcionais que eles desempenham no microbioma em que vivem. Entretanto, a maioria dos estudos de metagenômica se concentram em amostras provenientes do solo ou da água (AMATO et al., 2005). Portanto, o ambiente aéreo não tem sido o foco dos trabalhos de metagenômica, isso se deve principalmente pelas dificuldades que são apresentadas para o estudo desse compartimento do nosso planeta, tais como, a baixa biomassa de microrganismos presentes no ar, o que dificulta a recuperação eficaz do material genético, o que consequentemente prejudica o desenvolvimento de metodologias robustas para análises do microbioma aéreo.

Diante disso dessa lacuna, o presente trabalho buscou validar uma metodologia alternativa de coleta de material do ar. O uso do drone com um coletor acoplado permite que o material genético seja coletado do ambiente aéreo, proporcionando coletas em altitudes e localização diferentes. No entanto, aqui não foi possível correlacionar a composição de organismos das amostras com a localização ou o método de coleta, apenas verificou-se uma tendência do aumento da riqueza da amostra com o aumento da altitude ou velocidade da coleta (e.g amostras de 50m, 50km/h e 60/h). Portanto, os coletores se mostraram parcialmente capazes de realizar uma amostragem representativa do microbioma aéreo, o que

sugere a necessidade de aprimoramento desses protótipos.

Além do método de amostragem, a etapa de extração do DNA é crucial para o sucesso de um ensaio de sequenciamento, garantido que dados de qualidade sejam produzidos, possibilitando uma análise computacional robusta. Tendo em vista isso, para extração do DNA, o protocolo utilizado foi o de fenol/clorofórmio, que é o mais indicado para amostras com baixa biomassa, todavia, é uma técnica com diversas etapas manuais e que utiliza reagentes tóxicos, passíveis de contaminação, o que pode refletir na etapa de amplificação do DNA (DANTAS, 2023). Sendo assim, o material extraído das coletas foi passível de amplificação, entretanto, é possível perceber na tabela de quantificação de DNA no Nanodrop (Tabela 1), que as amostras não apresentam a pureza ideal e possuem uma baixa concentração de DNA, fatores determinantes nas fases seguintes.

Para realizar a análise computacional, inicialmente iríamos utilizar somente o pipeline *Ampliseq/nf-core* para maior otimização do tempo de análise e reprodutibilidade dos resultados, todavia, as leituras de sequenciamento não apresentaram um bom rendimento após o sequenciamento, e as amostras de controle apresentaram contaminação, dessa forma, as análises passaram a ser executadas no R, evidenciando que apesar de muito útil, a utilização de pipelines dificilmente se aplica a leituras com baixa qualidade, dado que, são sequências que precisam de protocolos de manipulação singulares e o uso de pipelines não permite esse manuseio.

Nessa conjuntura, apesar das consequências negativas experimentadas pela baixa qualidade do sequenciamento, que pode ser explicada pela dificuldade em extrair de maneira eficaz o DNA de amostras com baixa biomassa, é possível afirmar que a metodologia aplicada possui potencial de caracterizar o microbioma aéreo, contudo aprimoramentos são necessários para aumento do fluxo de ar, consequentemente, coleta de maior quantidade de biomassa. Além disso, é urgente a elaboração/aplicação de técnicas de extração de DNA com alto rendimento e menos etapas de manipulação da amostra voltadas para amostras com baixa densidade de microrganismos, visto que, essa é a principal causa de sequenciamentos com baixo rendimento e com contaminação, o que afeta diretamente as análises.

7 CONCLUSÃO

- Os coletores mostraram potencial para capturar o material particulado do ambiente aéreo, contudo o sequenciamento das amostras coletadas foi insatisfatório para que os dados gerados determinassem a composição microbiana.
- O trabalho demonstrou a carência de técnicas de extração de DNA eficiente e com alto rendimento para o processamento de amostras com baixa biomassa.
- Os resultados obtidos demonstram que o pipeline *Ampliseq* não consegue gerenciar informações de sequenciamento de baixa qualidade.
- O coletor de ar com fluxo ativo precisa de aprimoramentos para que uma nova versão permita o seu uso no drone com maior eficiência de coleta.

REFERÊNCIAS

- AHMADIAN, A.; EHN, M.; HOBER, S. Pyrosequencing: History, biochemistry and future. *Clinica Chimica Acta*, v. 363, p. 83-94, 2006.
- AMATO, P. et al. Microbial population in cloud water at the Puy de Dôme: Implications for the chemistry of clouds. *Atmospheric Environment*, v. 39, n. 22, p. 4143–4153, jul. 2005.
- BARBA, M.; CZOSNEK, H.; HADIDI, A. Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. *Viruses*, v. 6, n. 1, p. 106–136, 6 jan. 2014.
- BEHZAD, H.; GOJOBORI, T.; MINETA, K. Challenges and Opportunities of Airborne Metagenomics. *Genome Biology and Evolution*, v. 7, n. 5, p. 1216–1226, 6 maio 2015.
- BERG, G. et al. Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, v. 8, n. 1, 30 jun. 2020.
- BOLYEN, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, v. 37, n. 8, p. 852–857, 24 jul. 2019.
- CALLAHAN, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, v. 13, n. 7, p. 581–583, jul. 2016.
- CALLAHAN, B. J.; MCMURDIE, P. J.; HOLMES, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, v. 11, n. 12, p. 2639–2643, 21 jul. 2017.
- CAMERON, E. S. et al. Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities. *Scientific Reports*, v. 11, n. 1, p. 22302, 16 nov. 2021.
- CARNEIRO, M. O. et al. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics*, v. 13, n. 1, p. 375, 2012.
- CHAUDHARY, D. K.; KHULAN, A.; KIM, J. Development of a novel cultivation technique for uncultured soil bacteria. *Scientific Reports*, v. 9, n. 1, p. 6666, 30 abr. 2019.
- CLARKE, J. et al. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology*, v. 4, n. 4, p. 265–270, 22 fev. 2009.
- CHEN, S. et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)*, v. 34, n. 17, p. i884–i890, 1 set. 2018.
- DANTAS, Pedro Henrique Lopes Ferreira. Metagenômica de componentes particulados do ar capturados por drones autônomos em uma metrópole. Orientador:

Antonio Mauro Rezende. 2023. 47p. Dissertação (Mestrado Acadêmico em Programa de Pós-graduação em Biociências e Biotecnologia em Saúde) - Fundação Oswaldo Cruz, Instituto Aggeu Magalhães, Recife, 2023.

DELORT, A.-M. et al. A short overview of the microbial population in clouds: Potential roles in atmospheric chemistry and nucleation processes. *Atmospheric Research*, v. 98, n. 2, p. 249–260, 1 nov. 2010.

ESCOBAR-ZEPEDA, A.; VERA-PONCE DE LEÓN, A.; SANCHEZ-FLORES, A. The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Frontiers in Genetics*, v. 6, 17 dez. 2015.

EWELS, P. et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, v. 32, n. 19, p. 3047–3048, 16 jun. 2016.

FANG, Z. et al. Culturable Airborne Bacteria in Outdoor Environments in Beijing, China. *Microbial Ecology*, v. 54, n. 3, p. 487–496, 17 fev. 2007.

FATIMA, T.; EBERTZ, A. A journey through the history of DNA sequencing. Disponível em: <https://the-dna-universe.com/2020/11/02/a-journey-through-the-history-of-dna-sequencing/>.

FRANZETTI, A. et al. Seasonal variability of bacteria in fine and coarse urban air particulate matter. *Applied Microbiology and Biotechnology*, v. 90, n. 2, p. 745–753, 24 dez. 2010.

FERGUSON, R. M. W. et al. Bioaerosol biomonitoring: Sampling optimization for molecular microbial ecology. *Molecular Ecology Resources*, v. 19, n. 3, p. 672–690, maio. 2019.

GIANI, A. M. et al. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, v. 18, p. 9–19, 2020.

GILLESPIE, D. E. et al. Isolation of Antibiotics Turbomycin A and B from a Metagenomic Library of Soil Microbial DNA. *Applied and Environmental Microbiology*, v. 68, n. 9, p. 4301–4306, 1 set. 2002.

GOODWIN, S.; MCPHERSON, J. D.; MCCOMBIE, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nature reviews. Genetics*, v. 17, n. 6, p. 333–51, 2016.

GUSAREVA, E. S. et al. Microbial communities in the tropical air ecosystem follow a precise diel cycle. *Proceedings of the National Academy of Sciences*, v. 116, n. 46, p. 23299–23308, 28 out. 2019.

HEATHER, J. M.; CHAIN, B. The Sequence of sequencers: the History of Sequencing DNA. *Genomics*, v. 107, n. 1, p. 1–8, jan. 2016.

HUGENHOLTZ, P.; TYSON, G. W. Metagenomics. *Nature*, v. 455, n. 7212, p.

481–483, set. 2008.

ILLUMINA. 16S Metagenomic Sequencing Library Preparation: preparing 16s ribosomal rna gene amplicons for the illumina miseq system. San Diego: Illumina, . 28 p. 2013. Disponível em: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf. Acesso em: 25 Fev. 2024.

Inspire 2 - DJI. Disponível em: <<https://www.dji.com/br/inspire-2>>. Acesso em: 22 mar. 2024.

INSTRUMEX. Dual Channel Dust Sampler: Dual Channel Dust Sampler-(PM10 & PM2.5) manufacturer, supplier, and exporter, Mumbai-India.. 2019. Imagem. Disponível em: <https://instrumexindia.com/dual-channel-dust-sampler.html>. Acesso em: 25 Fev. 2024.

JIANG, H. et al. Altered fecal microbiota composition in patients with major depressive disorder. *Brain, Behavior, and Immunity*, v. 48, n. 48, p. 186–194, ago. 2015.

KEOHAVONG, P.; THILLY, W. G. Fidelity of DNA polymerases in DNA amplification. *Proceedings of the National Academy of Sciences of the United States of America*, v. 86, n. 23, p. 9253–9257, 1 dez. 1989.

KREMER, F. S. Entendendo o NGS (parte 1): plataformas de sequenciamento. Disponível em: <<https://medium.com/omixdata/entendendo-o-ngs-parte-i-plataformas-de-sequenciamento-66017d6d38d3>>. Acesso em: 20 mar. 2024.

KREMKOW, B. G.; LEE, K. H. Sequencing technologies for animal cell culture research. *Biotechnology Letters*, v. 37, n. 1, p. 55–65, 12 set. 2014.

KULSKI, J. K. Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications. [s.l.] IntechOpen, 2016.

LOMAN, N. J. et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, v. 30, n. 5, p. 434–439, 22 abr. 2012.

MANISALIDIS, I. et al. Environmental and health impacts of air pollution: A review. *Frontiers in Public Health*, v. 8, n. 14, p. 1–13, 20 fev. 2020.

MCFALL-NGAI, M. Are biologists in “future shock”? Symbiosis integrates biology across domains. *Nature Reviews Microbiology*, v. 6, n. 10, p. 789–792, out. 2008.

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, 22 abr. 2013.

METZKER, M. L. Sequencing technologies — the next generation. *Nature Reviews Genetics*, v. 11, n. 1, p. 31–46, 8 dez. 2009.

MORRIS, C. E. et al. Microbiology and atmospheric processes: research challenges concerning the impact of airborne micro-organisms on the atmosphere and climate. *Biogeosciences*, v. 8, n. 1, p. 17–25, 3 jan. 2011.

OULAS, A. et al. Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies. *Bioinformatics and Biology Insights*, v. 9, p. 75–88, 5 maio 2015.

PÉREZ-COBAS, A. E.; GOMEZ-VALERO, L.; BUCHRIESER, C. Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microbial Genomics*, v. 6, n. 8, 1 ago. 2020.

QUINCE, C. et al. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, v. 35, n. 9, p. 833–844, set. 2017.

RAUSCH, P. et al. Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome*, v. 7, n. 1, 14 set. 2019.

REGALADO, J. et al. Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe–microbe interaction networks in plant leaves. *The ISME Journal*, v. 14, n. 8, p. 2116–2130, 13 maio 2020.

ROSS, M. G. et al. Characterizing and measuring bias in sequence data. *Genome Biology*, v. 14, n. 5, p. R51, 2013.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, v. 74, n. 12, p. 5463–5467, 1 dez. 1977.

SATAM, H. et al. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology*, v. 12, n. 7, p. 997, 1 jul. 2023.

SCHECHTER, Matthew. The history of metagenomics: An incomplete summary. Disponível em: <<https://merenlab.org/2020/07/27/history-of-metagenomics/>>. Acesso em: 25 Fev. 2024.

SHENDURE, J. et al. DNA sequencing at 40: past, present and future. *Nature*, v. 550, n. 7676, p. 345–353, out. 2017.

SLATKO, B. E.; GARDNER, A. F.; AUSUBEL, F. M. Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, v. 122, n. 1, p. e59, abr. 2018.

SUN, Y. et al. Plastisphere microbiome: Methodology, diversity, and functionality. *iMeta*, 31 mar. 2023.

Visualizing and Characterizing DNA, RNA, and Protein | Microbiology. Disponível em: <<https://courses.lumenlearning.com/suny-mcc-microbiology/chapter/visualizing-and-characterizing-dna-rna-and-protein/>>. Acesso em: 19 mar. 2024.

VOELKERDING, K. V.; DAMES, S. A.; DURTSCHI, J. D. Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry*, v. 55, n. 4, p. 641–58, 2009.

WENGER, A. M. et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, v. 37, n. 10, p. 1155–1162, 12 ago. 2019.

WILLIS, A. D. Rarefaction, Alpha Diversity, and Statistics. *Frontiers in Microbiology*, v. 10, 23 out. 2019.

WOESE, C. R.; FOX, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, v. 74, n. 11, p. 5088–5090, 1 nov. 1977.

WOMACK, A. M.; BOHANNAN, B. J. M.; GREEN, J. L. Biodiversity and biogeography of the atmosphere. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 365, n. 1558, p. 3645–3653, 2010.

ZHANG, L. et al. Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers in Microbiology*, v. 12, 17 dez. 2021.