

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

Instituto Aggeu Magalhães

FUNDAÇÃO OSWALDO CRUZ
INSTITUTO AGGEU MAGALHÃES

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOCÊNCIAS E BIOTECNOLOGIA EM
SAÚDE

Pedro Henrique Lopes Ferreira Dantas

**Metagenômica de componentes particulados do ar capturados por drones autônomos em
uma metrópole**

Recife

2023

Pedro Henrique Lopes Ferreira Dantas

Metagenômica de componentes particulados do ar capturados por drones autônomos em uma metrópole

Dissertação de Mestrado Acadêmico em Biociências e Biotecnologia em Saúde apresentada ao Programa de Pós-graduação em Biociências e Biotecnologia em Saúde do Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, como requisito parcial para obtenção do grau de Mestre em Ciências.

Orientador: Dr. Antônio Mauro Rezende

Coorientador: Dr. Helotônio Carvalho

Recife

2023

Título do trabalho em inglês: Metagenomics of air particulate components captured by autonomous drones in a metropolis.

O presente trabalho foi realizado com apoio de Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE) - Código de Financiamento 001.

D192m Dantas, Pedro Henrique Lopes Ferreira.
Metagenômica de componentes particulados do ar capturados por drones autônomos em uma metrópole / Pedro Henrique Lopes Ferreira Dantas. -- 2023.
47 p. : il.color, fotos, mapas.

Orientador: Antonio Mauro Rezende.
Coorientador: Helotonio Carvalho.
Dissertação (Mestrado Acadêmico em Programa de Pós-graduação em Biociências e Biotecnologia em Saúde) - Fundação Oswaldo Cruz, Instituto Aggeu Magalhães, Recife, 2023.
Bibliografia: f. 41-46.

1. Metagenômica. 2. Dispositivos Aéreos não Tripulados. 3. Ar. 4. Microbiologia do Ar. 5. Análise de Sequência de DNA. I. Título.

CDU 579

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da Rede de Bibliotecas da Fiocruz com os dados fornecidos pelo(a) autor(a).

Bibliotecário responsável pela elaboração da ficha catalográfica: Adagilson Batista Bispo da Silva - CRB-1239
Biblioteca Luciana Borges Abrantes dos Santos

Pedro Henrique Lopes Ferreira Dantas

Metagenômica de componentes particulados do ar capturados por drones autônomos em uma metrópole

Dissertação de Mestrado Acadêmico em Biociências e Biotecnologia em Saúde apresentada ao Programa de Pós-graduação em Biociências e Biotecnologia em Saúde do Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, como requisito parcial para obtenção do grau de Mestre em Ciências.

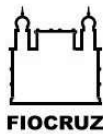
Aprovado em: 28 de julho de 2023

BANCA EXAMINADORA



Digitally signed by Antonio
Mauro
Rezende:07338288682
Date: 2023.10.02 20:55:15
-03'00'

Dr^o Antonio Mauro Rezende
Instituto Aggeu Magalhães/Fundação Oswaldo Cruz



Assinado de forma digital
por Christian Robson de
Souza Reis:02489893408
Dados: 2023.10.02
21:59:59 -03'00'

Dr^o Christian Robson de Souza Reis
Instituto Aggeu Magalhães/Fundação Oswaldo Cruz



Documento assinado digitalmente
CRHISLLANE RAFAELE DOS SANTOS VASCONCE
Data: 02/10/2023 13:29:36-0300
Verifique em <https://validar.iti.gov.br>

Dr^a Chhisllane Rafaela dos Santos Vasconcelos
Symbiomics

AGRADECIMENTOS

Gostaria de agradecer a todos que de alguma forma, diretamente ou indiretamente, contribuíram com este trabalho. Agradeço primeiramente a minha família, meus pais e meus irmãos por todo apoio. Agradeço as amigas que cultivei desde a escola até a universidade e que estão comigo até hoje, eu não sei o que seria sem vocês. Agradeço a Ana Clara por todo o apoio nesses dois anos, entre idas e vindas de Recife a João Pessoa, seu suporte tornou essa jornada mais tranquila. Agradeço a toda a comunidade do Instituto Aggeu Magalhães por terem contribuído para a execução desse trabalho. Agradeço ao meu orientador, Antonio Rezende, pela paciência e pelos ensinamentos. Agradeço ao meu coorientador, Helotonio Carvalho, pelo suporte e contribuições. Agradeço aos amigos André, Ludimilla, João e Ângela pelas boas conversas e por todo o apoio. Agradeço aos amigos do Departamento de Microbiologia, Raylla, Andrei, Darleide e Dourado pela convivência e companheirismo. Agradeço a Lais e Gustavo por toda ajuda no desenvolvimento do trabalho. Agradeço também o suporte e assistência daqueles que conduzem o Departamento de Microbiologia, Christian, Carol; e aos técnicos que sempre estiveram dispostos a ajudar no que fosse: Bruna, Igor, Mateus, Silvana e Claudio. Também agradeço a equipe do NPT, Cássia e Viviane, por toda assistência. Agradeço também aos colegas da Fiocruz Minas Gerais, Rubens e Matheus, que contribuíram para o desenvolvimento do projeto. Agradeço aos colegas do vôlei por proporcionarem os momentos de descontração que foram essenciais no final dessa jornada. Por fim, gostaria de agradecer o Instituto Aggeu Magalhães, a CAPES e a FACEPE pelo financiamento e apoio deste projeto.

RESUMO

DANTAS, Pedro Henrique Lopes Ferreira. Metagenômica de componentes particulados do ar capturados por drones autônomos em uma metrópole. 2023. Dissertação (Mestrado em Biociências e Biotecnologia em Saúde) – Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2023.

A microbiologia estuda a biologia de microrganismos, que engloba organismos dos mais diferentes ramos da árvore da vida. No século XX, parte do conhecimento a respeito das bactérias foi adquirido pelo advento das técnicas de cultivo. Com o surgimento do sequenciamento de nova geração (NGS), metodologias novas foram sendo desenvolvidas a fim de aprofundar o conhecimento que se tinha a respeito desses organismos. A metagenômica é uma técnica de sequenciamento de ácidos nucleicos que permite explorar comunidades microbiológicas em amostras biológicas ou ambientais independentemente de técnicas de cultivo. O ar é um compartimento da biosfera pouco explorado devido a vários fatores, como a dificuldade de coleta, que conseqüentemente leva a deficiência na análise da composição microbiana do ar. Nós propomos uma nova metodologia que proporciona uma coleta mais flexível do ar, por meio de um coletor de ar acoplado a um drone, onde a composição será resolvida por meio da abordagem da metagenômica. Coletores de ar foram desenvolvidos pelo nosso próprio grupo de pesquisa em parceria com a Plataforma Unificada de Modelagem e Manufatura Aditiva (PLUMMA-3D – FIOCRUZ Minas Gerais). Utilizando os coletores acoplados a um drone modelo DJI Inspire 2, foram realizados sobrevoos em diferentes altitudes ao nível do solo e acima do solo. Acoplado aos coletores foram utilizados tubos cônicos possuindo tampão de amostra PBS. As amostras coletadas foram extraídas utilizando protocolo de fenol/clorofórmio, o DNA foi submetido à amplificação através da PCR, e o amplicon foi utilizado na construção das bibliotecas para sequenciamento. As bibliotecas foram quantificadas e o sequenciamento foi realizado em uma plataforma Illumina MiSeq. A análise da qualidade dos dados foi conduzida para remoção das bases ambíguas (Ns) e de baixa qualidade. Posteriormente, foi feita a inferência das variantes de sequenciamento de amplicon (ASVs). Por fim, análises de diversidade alfa e beta foram conduzidas, além de análises para abundância diferencial dos táxons. Os coletores desenvolvidos se mostraram capazes de coletar biomassa a partir do ar e o gel de agarose 1% confirmou a amplificação do nosso fragmento de interesse (rRNA 16S V3/V4). Além disso, as bibliotecas também foram validadas através de gel de agarose 1%. Após o sequenciamento e o tratamento de qualidade, um total de ~3 milhões de leituras foram utilizadas para a etapa inferência de ASVs, com uma média de ~160.000 leituras por amostra. Um total de 1878 ASVs foram preditas ao longo das amostras e tais ASVs foram classificadas em diferentes níveis taxonômicos. Neste trabalho, foi possível estabelecer um fluxo de coleta, processamento e análises de dados para amostras de ar, utilizando drones e a metagenômica para tal. Entretanto, dificuldades nesse tipo de trabalho foram expostas após análises de diversidade alfa e beta e o desenvolvimento de técnicas de processamento de material genético para amostras com baixa biomassa e com maior reprodutibilidade são necessárias. Por fim, espera-se então que este estudo possa contribuir com o conhecimento acerca do microbioma presente no ar e que sirva de estímulo para novos trabalhos desta natureza.

Palavras chaves: metagenômica; drone; ar.

ABSTRACT

DANTAS, Pedro Henrique Lopes Ferreira. Metagenomics of airborne particulate components captured by autonomous drones in a metropolis. 2023. Dissertation (Master in Biosciences and Biotechnology in Health) – Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2023.

Microbiology studies the biology of microorganisms, which encompasses organisms from the most different branches of the tree of life. In the twentieth century, part of the knowledge about bacteria was acquired by the advent of cultivation techniques. With the emergence of next generation sequencing (NGS), new methodologies were being developed in order to deepen the knowledge that was had about these organisms. Metagenomics is a nucleic acid sequencing technique that allows exploring microbiological communities in biological or environmental samples independently of cultivation techniques. The air is a compartment of the biosphere little explored due to several factors, such as the difficulty of collection, which consequently leads to a deficiency in the analysis of the microbial composition of the air. We propose a new methodology that provides a more flexible collection of air, through an air collector coupled to a drone, where the composition will be resolved through the metagenomics approach. Air collectors were developed by our own research group in partnership with the Unified Modeling and Additive Manufacturing Platform (PLUMMA-3D – FIOCRUZ Minas Gerais). Using the collectors coupled to a DJI Inspire 2 model drone, flyovers were performed at different altitudes at ground level and above ground. Connected to the collectors, conical tubes containing PBS sample buffer were used. The collected samples were extracted using a phenol/chloroform protocol, the DNA was subjected to amplification through PCR, and the amplicon was used in the construction of libraries for sequencing. Libraries were quantified and sequencing was performed on an Illumina MiSeq platform. Data quality analysis was conducted to remove ambiguous bases (Ns) and low quality. Subsequently, the inference of amplicon sequencing variants (ASVs) was made. Finally, analyzes of alpha and beta diversity were conducted, in addition to analyzes for differential abundance of taxa. The collectors developed were able to collect biomass from air and the 1% agarose gel confirmed the amplification of our fragment of interest (rRNA 16S V3/V4). In addition, the libraries were also validated using a 1% agarose gel. After sequencing and quality handling, a total of ~3 million reads were used for the ASV inference step, with an average of ~160,000 reads per sample. A total of 1878 ASVs were predicted over the samples and such ASVs were classified into different taxonomic levels. In this work, it was possible to establish a data collection, processing and analysis flow for air samples, using drones and metagenomics to do so. However, difficulties in this type of work were exposed after analyzes of alpha and beta diversity and the development of techniques for processing genetic material for samples with low biomass and with greater reproducibility are necessary. Finally, it is hoped that this study can contribute to knowledge about the microbiome present in the air and that it will serve as a stimulus for further work of this nature.

Keywords: metagenomics; drone; air.

LISTA DE ILUSTRAÇÕES

Figura 1 – Custo por genoma humano sequenciado e sua comparação com a Lei de Moore.	14
Figura 2 – Esquema ilustrativo dos dois tipos de abordagens metagenômicas.	16
Figura 3 – Representações gráficas do gene rRNA 16S.	17
Figura 4 – Modelos de drone comerciais.	21
Figura 5 – Drone modelo DJI Inspire 2 utilizado nas coletas.	22
Figura 6 – Mapas representando o local de coleta e o percurso percorrido pelo drone.	22
Figura 7 – Bomba de sucção utilizada para coleta de amostras de ar.	23
Figura 8 – Modelo esquemático do coletor v2.	26
Figura 9 – Coletores desenvolvidos pelo grupo.	26
Figura 10 – Esquema representativo do funcionamento do coletor.	27
Figura 11 – Resultados das PCRs para amplificação da região de interesse.	29
Figura 12 – Eletroforese em gel de agarose 1% das bibliotecas.	31
Figura 13 - Qualidade média das leituras forward (azul) e reverse (verde).	32
Figura 14 – Qualidade média das leituras forward após trimagem.	33
Figura 15 – Proporção de filós identificados em cada amostra.	34
Figura 16 – Medida de diversidade alfa das amostras a partir do índice de Shannon.	35
Figura 17 – Diversidade beta avaliada entre as amostras utilizando a dissimilaridade de Bray-curtis.	36

LISTA DE ABREVIATURAS E SIGLAS

DNA	Ácido desoxirribonucleico
RNA	Ácido ribonucleico
NGS	Sequenciamento de nova geração
WGS	Sequenciamento de genoma total
rRNA	RNA ribossomal
ITS	Espaçador transcrito interno
ASV	Variante de sequenciamento de <i>amplicon</i>
OTU	Unidade taxonômica operacional
EMP	Projeto Microbioma da Terra
HMP	Projeto Microbioma Humano
PLUMMA	Plataforma Unificada de Modelagem e Manufatura Aditiva
CAD	Desenho assistido por computador
FFF	Fabricação de filamentos fundidos
PETG	Polietileno Tereftalato Glicol
PBS	Tampão fosfato-salino
PCR	Reação em cadeia da polimerase
EDTA	Ácido etilenodiaminotetracético
NaCl	Cloreto de sódio
dNTP	Desoxirribonucleotídeos trifosfato
MgCl ₂	Cloreto de magnésio
DMSO	Dimetilsulfóxido
dsDNA	DNA dupla-fita
HS	Alta sensibilidade

SUMÁRIO

1	INTRODUÇÃO	9
2	OBJETIVOS	11
2.1	OBJETIVO GERAL.....	11
2.2	OBJETIVOS ESPECÍFICOS	11
3	FUNDAMENTAÇÃO TEÓRICA	12
3.1	HISTÓRICO	12
3.2	MÉTODOS TRADICIONAIS DE CULTIVO BACTERIANO	12
3.3	SEQUENCIAMENTO DE NOVA GERAÇÃO.....	13
3.4	METAGENÔMICA	15
3.5	O ESTUDO DO AR A LUZ DA METAGENÔMICA.....	19
4	MATERIAIS E MÉTODOS	21
4.1	DESENVOLVIMENTO DO COLETOR DE AR	21
4.2	COLETA DAS AMOSTRAS DE AR.....	21
4.3	EXTRAÇÃO DE DNA E PCR.....	23
4.4	SEQUENCIAMENTO	24
4.5	ANÁLISES COMPUTACIONAIS.....	25
5	RESULTADOS	25
5.1	DESENVOLVIMENTO DOS COLETORES DE AR PARA OBTENÇÃO DE AMOSTRAS DE MICROBIOMA	25
5.2	AMOSTRAS DE AR COLETADAS APRESENTARAM BOM RENDIMENTO E POTENCIAL PARA SEREM AMPLIFICADAS	28
5.3	SEQUENCIAMENTO FOI CAPAZ DE GERAR QUANTIDADE SUFICIENTE DE DADOS PARA CARACTERIZAÇÃO DA COMPOSIÇÃO MICROBIANA.....	30
5.4	ANÁLISES DE BIOINFORMÁTICA SE MOSTRARAM CAPAZES DE CARACTERIZAR ASPECTOS DO MICROBIOMA AÉREO	32
6	DISCUSSÃO	37
7	CONCLUSÃO	40
	REFERÊNCIAS	41

1 INTRODUÇÃO

A microbiologia é o campo da ciência que estuda a biologia de microrganismos. Atualmente, esse termo é utilizado para englobar organismos dos mais diferentes ramos da árvore da vida: vírus, bactérias, arqueias, fungos e protozoários. O conhecimento a respeito desses organismos cresceu grandemente a partir do século XIX, especialmente devido às contribuições de Pasteur que propôs a teoria da biogênese. Posteriormente, ainda no século XIX, Joseph Lister e Robert Koch contribuíram para a elucidação da relação dos microrganismos com processos patológicos. A partir do século XX a sociedade passou à enxergar a essencialidade dos microrganismos e estes começaram a ser utilizados em diversas atividades econômicas, tais como a utilização na indústria alimentícia, por meio de processos como fermentação, descoberta de novos antibióticos como a penicilina, além da já citada importância para a saúde, não apenas como sendo causadores de doenças mas também atuando, por exemplo, no metabolismo de plantas e animais em relações de simbiose.

Observando especificamente as bactérias, durante o século XX, parte significativa do conhecimento a respeito desses microrganismos foi adquirido pelo advento das técnicas de cultivo, que compreendem basicamente o crescimento de alguma espécie bacteriana em um ambiente controlado utilizando um meio nutritivo para suplementar as necessidades metabólicas do organismo. Essas técnicas de cultivo, no entanto, são limitadas uma vez que poucas espécies de bactérias são passíveis de cultivo, limitando o conhecimento que temos a respeito da diversidade bacteriana existente. Contudo, seguindo os avanços tecnológicos, principalmente com o desenvolvimento das técnicas de sequenciamento em larga escala no início do século XXI, a diversidade do material genético dos mais diversos microrganismos incluindo bactérias passou a ser explorada de forma mais ampla. Com o surgimento das plataformas de sequenciamento de nova geração (NGS), novas metodologias foram sendo criadas a fim de aprofundar o conhecimento que se tinha a respeito do genoma, transcriptoma, proteoma, metaboloma e resistoma desses organismos.

Ainda relacionado ao contexto do avanço tecnológico citado acima, surge a metagenômica, uma técnica de sequenciamento de ácidos nucleicos que permite explorar comunidades microbiológicas independentemente das técnicas de cultivo. Com essa abordagem, amostras de diferentes fontes podem ter o material genético extraído e sequenciado. Diversos estudos se propuseram ao longo dos anos em resolver o microbioma contido em amostras ambientais e em amostras biológicas. Tais estudos buscam relacionar o microbioma ali presente com fenômenos e fatores que determinam esses ambientes.

Dentre as amostras ambientais, destacam-se as amostras de solo e água, entretanto, o ar é um compartimento da biosfera pouco explorado devido a vários fatores, como a dificuldade de coleta, o que conseqüentemente leva a deficiência na análise da composição microbiana do ar. Os métodos tradicionais de coleta de ar são executados por pesadas máquinas de sucção de ar que possuem pouca ou nenhuma mobilidade, limitando a amostragem por questões de logísticas, impossibilitando por exemplo a variação de altitude no ambiente estudado.

Diante da necessidade de se ampliar o conhecimento que possuímos a respeito da composição bacteriana que ocorre na atmosfera, neste trabalho nós propusemos o desenvolvimento de novos métodos que proporcionam uma coleta mais flexível do ar, por meio de um coletor de ar acoplado à um drone. Dessa forma, foi possível obter amostras de diferentes ambientes, altitudes e avaliar a composição e biodiversidade bacteriana presente no ar analisado, composição essa resolvida por meio da abordagem de metagenômica.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Desenvolver e validar um novo método para caracterização da microbiota do ar utilizando aeronaves de pequeno porte não tripuladas (drones) para amostragem em diferentes locais e altitudes.

2.2 OBJETIVOS ESPECÍFICOS

- a) Desenvolver um coletor de ar para coleta de microbioma.
- b) Realizar as coletas de amostras com coletor de ar acoplado a um drone.
- c) Realizar a extração e o sequenciamento de DNA a partir das amostras coletadas.
- d) Analisar os dados obtidos e interpretá-los quanto aos microbiomas identificados.
- e) Comparar os métodos desenvolvidos com os métodos de coleta de ar tradicionais.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 HISTÓRICO

Os microrganismos foram primeiramente observados no século XVII a partir dos microscópios desenvolvidos por Anton Von Levenhook. Denominados inicialmente de “animálculos”, esses organismos foram cogitados como sendo causadores das mais variadas doenças inicialmente por Joseph Lister em 1865, hipótese essa que seria confirmada posteriormente, em 1876, por Robert Koch. Destacam-se também os estudos de Louis Pasteur, que com seu caldo nutritivo comprovou a veracidade da teoria da biogênese, além de contribuir com a descoberta e combate de microrganismos patogênicos de interesse sanitário mundial (PASTEUR, 1882, 1881). Tais estudos pioneiros contribuíram para a comprovação da existência dos microrganismos, bem como na elucidação da relação deles com a saúde humana, a produção de alimentos e outros fenômenos biológicos (FIJAN, 2014; MARCO et al., 2017). Hoje, sabemos que o termo genérico “microrganismo” engloba espécies pertencentes à diversos grupos biológicos como bactérias, arqueias, fungos, protozoários e vírus, e que estes estão presentes nos mais diversos habitats seja de forma livre ou por meio de relações ecológicas intra e interespecíficas (GUSAREVA et al., 2019a).

3.2 MÉTODOS TRADICIONAIS DE CULTIVO BACTERIANO

As técnicas de cultivo foram inicialmente estabelecidas no século XIX, na chamada “Era de Ouro da Microbiologia”, a partir da contribuição de trabalhos de Robert Koch e Louis Pasteur (BLEVINS; BRONZE, 2010; BONNET et al., 2019; URUBURU, 2003). Com objetivo de isolar e estudar aspectos relacionados ao metabolismo e fisiologia dos microrganismos, tais técnicas de cultivo utilizam soluções nutritivas para promover o crescimento microbiano (GILL, 2017; MADIGAN et al., 2018). Entretanto, o estudo dos microrganismos foi particularmente dispendioso ao longo do tempo, devido principalmente à dificuldade de se cultivar espécimes *in vitro*, dificuldade essa que até o presente momento não foi contornada. Estima-se que somente algo próximo de 1% das espécies de bactérias podem ser cultivadas em laboratório (CHAUDHARY; KHULAN; KIM, 2019; NGUYEN et al., 2018; VARTOUKIAN; PALMER; WADE, 2010; WADE, 2002). Essas dificuldades se devem à falta de conhecimento à respeito das condições ideais de crescimento da grande maioria das bactérias, condições essas envolvendo nutrientes e minerais essenciais ao desenvolvimento microbiano (STEWART, 2012). Diante disso, técnicas de sequenciamento, que não dependem necessariamente de cultivo, começaram a ser empregadas a fim de preencher esse espaço sobre o conhecimento

relacionado aos microrganismos que não conseguem ser cultivados a partir de técnicas tradicionais em laboratório (ERCOLINI, 2013; TAYLOR-BROWN et al., 2017).

3.3 SEQUENCIAMENTO DE NOVA GERAÇÃO

No ano de 1977 surge a técnica de sequenciamento por interrompimento da síntese da molécula de DNA conhecida como técnica de sequenciamento de Sanger, uma vez que foi desenvolvida pelo químico Frederik Sanger (SANGER; NICKLEN; COULSON, 1977). Esse foi também o primeiro método de sequenciamento passível de automação. Tal técnica envolve primeiramente o isolamento do DNA de interesse e sua posterior fragmentação. Este DNA fragmentado é submetido à uma reação em cadeia da polimerase (PCR) que possui dentre seus componentes dideoxynucleotídeos (ddATP, ddCTP, ddGTP ou ddTTP), os quais não possuem uma hidroxila na posição 3' da pentose. A PCR é conduzida e ao incorporar um dideoxynucleotídeo, nenhum outro nucleotídeo pode ser adicionado uma vez que a DNA-polimerase não consegue inserir um nucleotídeo sem uma hidroxila livre. A interrupção da amplificação gera fragmentos de DNA de diversos tamanhos que tiveram dideoxynucleotídeos inseridos em diferentes posições. Por fim, o resultado é avaliado utilizando eletroforese em gel, onde as bandas de DNA são organizadas em sequência, sendo possível determinar a base correspondente a cada posição com base no ddNTP que foi inserido no final da sequência amplificada. Isso leva então à construção de uma sequência final do DNA alvo que foi submetido à técnica.

O sequenciamento capilar, como ficou também conhecido o sequenciamento de Sanger, permitiu que o conhecimento a respeito dos microrganismos crescesse significativamente, a medida que os mais diversos genomas iam sendo sequenciados, sendo o primeiro deles o do bacteriófago Φ X174 de *Escherichia coli* (SANGER et al., 1977). Em 1995, nós tivemos o sequenciamento do genoma de *Haemophilus influenzae Rd*, o primeiro genoma de um organismo de vida livre (FLEISCHMANN et al., 1995). Em 1996, o sequenciamento do genoma da levedura *Saccharomyces cerevisiae* (GOFFEAU et al., 1997) foi realizado, e em 1997, o sequenciamento do genoma da bactéria *E. coli* (BLATTNER et al., 1997). Todos esses estudos representaram avanços significativos para a ciência por se tratar de organismos que viriam a se tornar modelos para biologia molecular e genética. Contudo, o método de Sanger possui baixo rendimento quando comparado aos métodos de sequenciamento mais recentes.

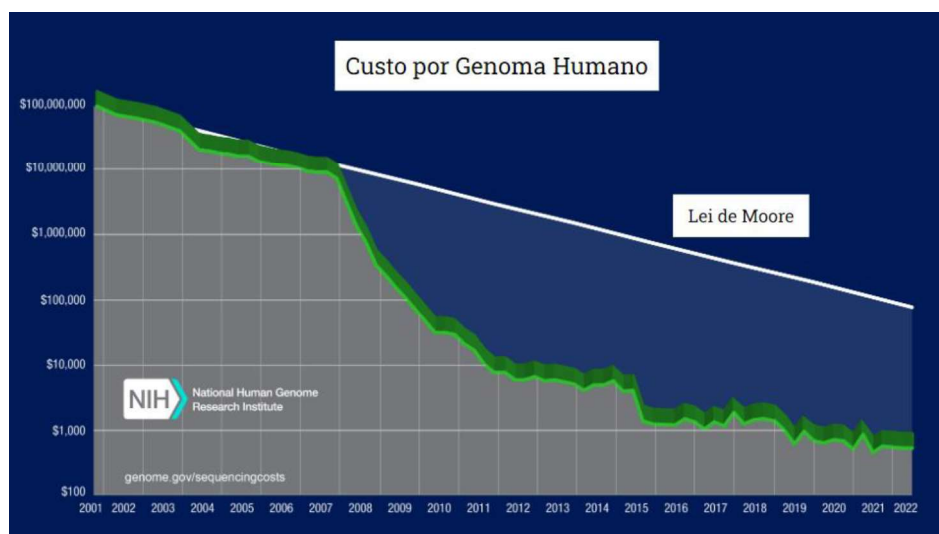
A partir de 2005, o avanço tecnológico e a finalização do projeto Genoma Humano, estimularam o desenvolvimento de metodologias de sequenciamento de larga escala (Next Generation Sequencing ou High Throughput Sequencing). Isso possibilitou aliar novas

abordagens no estudo dos microrganismos, sendo agora possível a determinação das sequências nucleotídicas presentes em uma amostra qualquer a partir da extração do material genético, seja ele DNA ou RNA (CANTALUPO; PIPAS, 2019; DE CASTRO et al., 2018; FEDRIZZI et al., 2017). Tecnologias de segunda geração, como por exemplo as tecnologias utilizadas nas plataformas de sequenciamento das empresas Illumina (MiSeq, NextSeq and NovaSeq) e ThermoFisher Scientific (Ion Torrent) se baseiam na técnica de amplificação clonal e sequenciamento por síntese (HU et al., 2021).

A plataforma Ion Torrent utiliza o princípio da PCR de emulsão, onde diversas esferas microscópicas imersas em uma solução são ligadas à fragmentos de DNA, que são amplificados. Após a amplificação dos fragmentos de interesse, a etapa de sequenciamento realiza a incorporação de nucleotídeos que por sua vez leva à liberação de íons H^+ que são detectados por sensores de pH (GOODWIN; MCPHERSON; MCCOMBIE, 2016). No caso de plataformas Illumina, os fragmentos de DNA são ligados à uma lâmina de vidro (*flow cell*) e posteriormente amplificados, o que resulta na formação de *clusters*. Após a amplificação, o sequenciamento inicia a incorporação dos nucleotídeos, que são distinguidos pelo equipamento a partir da luz emitida por fluoróforos associados à cada dNTP, que por sua vez é capturada por sensores ópticos (MARDIS, 2013).

Dentre as vantagens dessas plataformas estão o custo por genoma sequenciado, que vem caindo cada vez mais no últimos 20 anos (Figura 1), além da alta profundidade de sequenciamento, uma vez que cada nucleotídeo, em uma determinada posição em um genoma alvo, é sequenciado várias vezes, o que garante uma maior acurácia (LEVY; BOONE, 2019).

Figura 1 – Custo por genoma humano sequenciado e sua comparação com a Lei de Moore.



Fonte: adaptado de National Human Genome Research Institute, 2021.

Por outro lado, tecnologias de leituras (fragmentos individuais sequenciados) curtas possuem alto custo computacional, devido à grande quantidade de dados gerados aliado ao alto número de possibilidades no momento do alinhamento das leituras à um genoma de referência; uma outra dificuldade, para o caso de plataformas ThermoFisher Scientific (Ion Torrent), é a resolução de regiões de baixa complexidade (e.g. regiões repetitivas, homopolímeros) pois levam à alta intensidade de sinais químicos/ópticos, o que torna difícil a distinção entre os nucleotídeos incorporados e sua quantidade, podendo gerar inserção/deleção nas leituras sequenciadas (ALSER et al., 2021; HEYDARI et al., 2019). Apesar disso, plataformas de leituras curtas ainda são majoritariamente utilizadas.

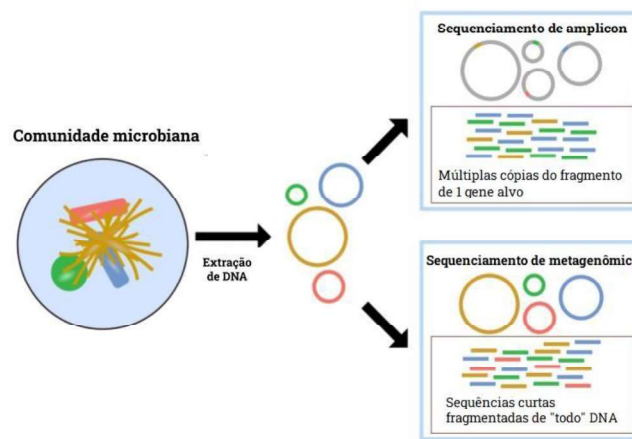
Atualmente, existem novas tecnologias de sequenciamento que utilizam leituras longas, dentre elas, podemos citar as plataformas da empresa Pacific Biosciences (PacBio), cujo primeiro dispositivo foi lançado em 2011. Além disso, existem os sequenciadores portáteis (e.g. MinIon) da empresa Oxford Nanopore Technologies (ONT), que também produzem leituras longas. Os sequenciadores da PacBio (RSII, Sequel) utilizam uma técnica denominada de “Single Molecule Real Time”, onde um fragmento dupla-fita de DNA tem suas extremidades unidas formando uma molécula circular. Esse DNA circular é ligado à uma DNA-polimerase contida no fundo de um poço que por sua vez passa a incorporar nucleotídeos marcados com fluoróforos que emitem sinais ópticos a medida que são adicionados, sendo possível determinar os nucleotídeos que compõem a sequência alvo (RHOADS; AU, 2015). Já os sequenciadores da ONT possuem uma membrana com poros por onde a molécula de DNA passa; essa membrana é submetida à uma diferença de potencial e a medida que os nucleotídeos vão passando pelo poro há uma mudança na condutância, sendo possível determinar qual base foi sequenciada a partir do sinal elétrico captado (LU; GIORDANO; NING, 2016). Sequenciadores da PacBio são capazes de sequenciar fragmentos de 30-50 kb, enquanto que plataformas da ONT pode sequenciar fragmentos que podem chegar a ~1 Mb (SLATKO; GARDNER; AUSUBEL, 2018; VAN DIJK et al., 2018). Utilizando tais tecnologias de leituras longas e curtas, novas metodologias foram desenvolvidas para elucidar o genoma, transcriptoma e proteoma dos mais diversos organismos, dando início ao que ficou conhecido como “Era das Ômicas”. Dentre tais metodologias, destaca-se a metagenômica.

3.4 METAGENÔMICA

A metagenômica se define como o sequenciamento e determinação das sequências nucleotídicas de uma amostra (HANDELSMAN, 2004; RIESENFELD; SCHLOSS; HANDELSMAN, 2004) sem um alvo específico. Há atualmente duas abordagens

metagenômicas que são amplamente usadas e conhecidas: metagenômica de *shotgun* (WGS ou whole genome sequencing) e metagenômica de *amplicon* (Figura 2).

Figura 2 – Esquema ilustrativo dos dois tipos de abordagens metagenômicas. A representação traz uma comunidade microbioana que passa pelo processo de extração de DNA, e a partir disso, duas abordagens podem ser aplicadas: a abordagem do sequenciamento de *amplicon*, onde um gene alvo de todos os organismos presentes é sequenciado; ou a abordagem de sequenciamento de metagenômica, onde espera-se sequenciar o genoma total de todos os organismos presentes.



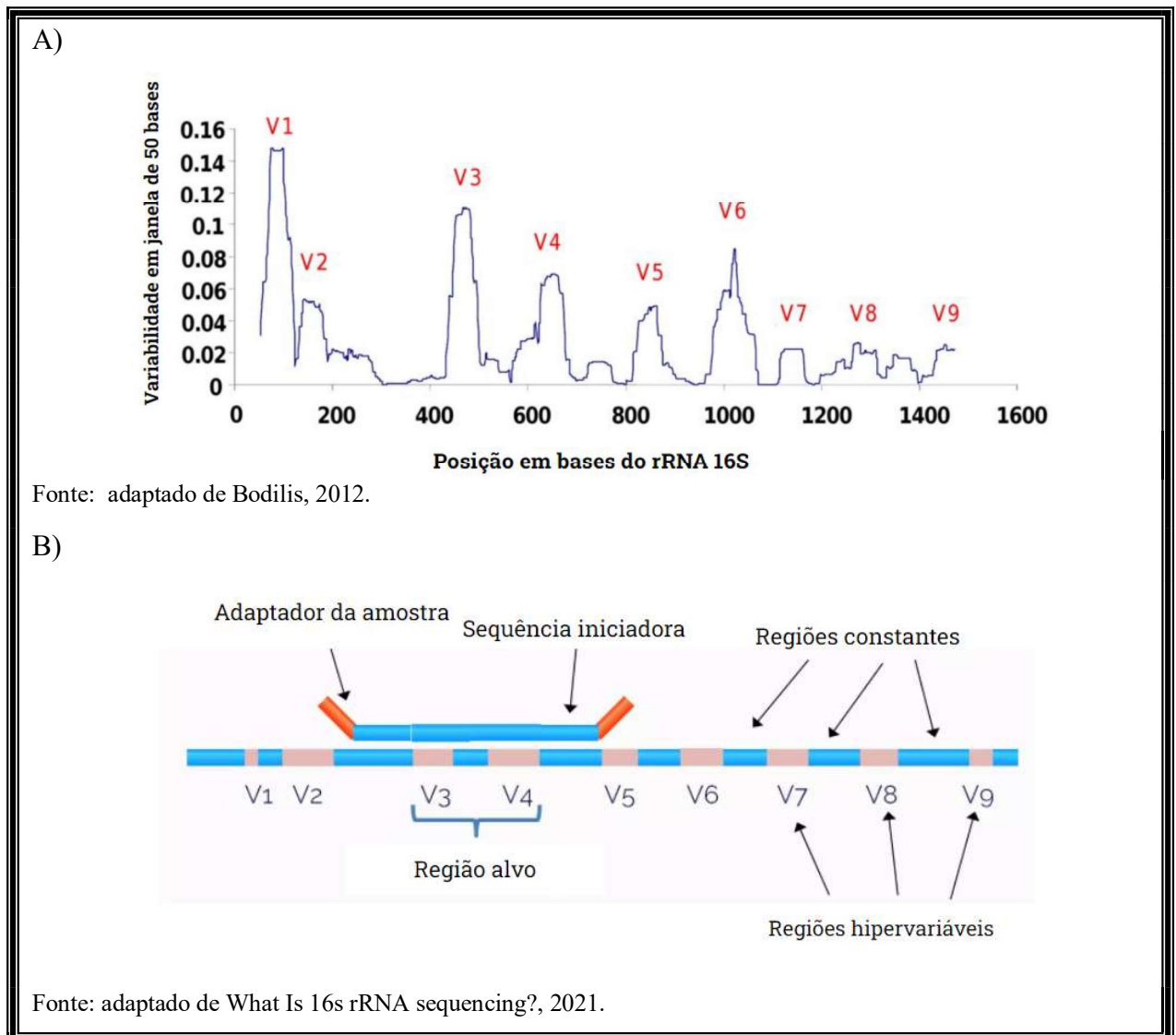
Fonte: adaptado de Lee, 2023.

No caso de WGS, o sequenciamento é feito para o DNA total da amostra, na tentativa de amostrar os genomas presentes (CHEN; PACHTER, 2005). Já a metagenômica de *amplicon* consiste no sequenciamento em larga escala de uma sequência específica (marcador taxonômico) em uma amostra visando principalmente a classificação taxonômica dos microrganismos ali contidos (COTTRELL et al., 2005; OLSEN et al., 1986). Para esse propósito são usados genes amplamente distribuídos em seus respectivos grupos de organismos, e que ainda possuem certa variabilidade nas sequências que permite seu uso como marcador para classificação taxonômica; são exemplos o rRNA (RNA ribossomal) 16S no caso de procariotos e rRNA 18S e ITS (Internal Transcribed Spacer) para eucariotos (GAO et al., 2021).

Como mencionado acima, o marcador utilizado para bactérias e arqueias é o gene 16S, um dos genes que codifica para uma parte da subunidade menor (30S) do ribossomo procarioto (WAGNER; GASSEN, 1976). Devido a sua baixa taxa evolutiva, o gene 16S é altamente conservado entre os procariotos, contudo, esse gene possui o que são conhecidas como regiões hipervariáveis (CHAKRAVORTY et al., 2007; PEI et al., 2010). Nove regiões hipervariáveis no gene 16S (V1-V9) são conhecidas (Figura 3A), sendo as regiões V3-V4 (Figura 3B) usualmente utilizadas para fins de metagenômica de *amplicon* uma vez que elas apresentam

maior heterogeneidade se comparadas às outras regiões do 16S e são mais bem estabelecidas (ATA et al., 2019; GARCÍA-LÓPEZ et al., 2020). Apesar de tal hipervariabilidade permitir a distinção em níveis taxonômicos mais altos, como classe e família, a conservação do gene rRNA 16S torna a classificação à nível de espécie um desafio (PLUMMER et al., 2015).

Figura 3 – Representações gráficas do gene rRNA 16S. A) Heterogeneidade das regiões hipervariáveis do gene rRNA 16S. No eixo x é apresentado a posição das bases do rRNA 16S, o qual possui ~1600 pares de base (pb). No eixo y é mostrado a variabilidade das regiões a cada 50 nucleotídeos. B) Gene 16S rRNA e a posição de anelamento dos primers. Destacado em azul estão as regiões conservadas do rRNA 16S, regiões essas onde os primers se anelam. Em rosa estão destacadas as regiões hipervariáveis.



Após o estabelecimento de protocolos para estudar e acessar a diversidade bacteriana, estudos foram feitos para determinar a melhor forma de analisar os dados produzidos a partir do sequenciamento de *amplicon*, e duas abordagens se destacaram ao longo dos anos: a Unidade

Taxonômica Operacional (OTU) e a Variante de Sequência de *Amplicon* (ASV). As OTUs utilizam o princípio da similaridade, onde sequências que apresentam um limiar de similaridade, normalmente 97%, são agrupadas em um mesmo grupo. Uma sequência representativa desse cluster então é classificada taxonomicamente e tal classificação é atribuída a todas as outras sequências desse mesmo grupo (NGUYEN et al., 2016). Diversos *pipelines* utilizam essa abordagem e dentre eles temos o MOTHUR e UPARSE (EDGAR, 2013; SCHLOSS et al., 2009). A abordagem de ASV por sua vez leva em consideração a frequência de cada sequência exata aliada à um modelo de erro para estabelecer se tal sequência é um atributo biológico real ou fruto de erros de sequenciamento. Dessa forma, cada ASV recebe um valor de confiança (p-valor) que é utilizado para filtrar os resultados.

A abordagem por ASV tem ganhado cada vez mais destaque, substituindo metodologias já consolidadas como OTU, tendo como principais motivos o melhor custo computacional, acurácia e reprodutibilidade (CALLAHAN; MCMURDIE; HOLMES, 2017; PRODAN et al., 2020; SIERRA et al., 2020). Um dos principais *pipelines* que empregam tal abordagem é o DADA2 (CALLAHAN et al., 2016).

Além das diferentes abordagens para predição de ASVs e OTUs, métricas de estatística ecológica começaram a ser empregadas na caracterização de comunidade microbiológicas, como no caso de estudos de metagenômica. Algumas métricas usualmente empregadas são os índices de diversidade alfa e beta. Os índices de diversidade alfa avaliam a estrutura de uma comunidade microbiológica de acordo com sua riqueza, uniformidade, ou ambos (WILLIS, 2019), como geralmente ocorre quando se utiliza o índice de Shannon (SHANNON; WEAVER, 1998), uma das métricas de diversidade alfa mais utilizadas em trabalhos dessa natureza (REESE; DUNN, 2018). Por sua vez, os índices de diversidade beta se propõem a comparar duas amostras entre si, avaliando a similaridade ou dissimilaridade (ANDERMANN et al., 2022), e para isso, métricas como a dissimilaridade de Bray-curtis e UniFrac são amplamente utilizadas (BRAY; CURTIS, 1957; LOZUPONE et al., 2011).

Utilizando as abordagens citadas acima, estudos foram conduzidos avaliando a composição microbiana de amostras ambientais, especialmente solo e água (LI et al., 2020; SALEEM et al., 2018). Além da composição microbiana, esses estudos também buscam relacionar o perfil microbiano e o ambiente em que se encontram (NEAL et al., 2020; OH et al., 2019; REDDY; DUBEY, 2021). Tais resultados representam o espaço ocupado pelos microrganismos, e os mesmos estando presente nos mais variados locais, é de se esperar que tais organismos estejam intimamente relacionados com diversos processos biológicos, inclusive em nosso próprio corpo. Como exemplos podemos citar a influência da microbiota fecal sobre

a saúde (JIANG et al., 2015); microbiota associada à metabolização de medicamentos (JAVDAN et al., 2020) e microrganismos associados à proteção contra vírus e bactérias patogênicas (LI et al., 2019). Além de tais estudos, diversos consórcios ao redor do mundo lançaram esforços para caracterizar os mais diversos microbiomas, dentre os quais pode-se destacar o Earth Microbiome Project (EMP) (THOMPSON et al., 2017) e o Human Microbiome Project (HMP) (THE INTEGRATIVE HMP (IHMP) RESEARCH NETWORK CONSORTIUM, 2019). Apesar dos mais diversos esforços e grandes avanços nesse campo, um importante compartimento da biosfera vem sendo pouco explorado: o ar.

3.5 O ESTUDO DO AR A LUZ DA METAGENÔMICA

Pesquisadores tem dividido a biosfera em várias camadas e sistemas fechados de organismos, a fim de que seja proporcionada uma melhor maneira de entender a dinâmica desses sistemas. Além disso, sabemos que todos os componentes da biosfera atuam de forma interconectada, influenciando um ao outro. Isso pode ser melhor observado em diferentes estudos que mostram, por exemplo, a dispersão de areia da África para a América do Sul, e como essa areia fertiliza o solo amazônico (NOGUEIRA et al., 2021; PROSPERO; GLACCUM; NEES, 1981), e mais especificamente no caso da dispersão de organismos fora de sua região de origem a partir de uma longa viagem seguindo fluxos de ar como no caso dos mosquitos que se dispersam até 300 km (HUESTIS et al., 2019).

Apesar dos avanços no entendimento da relação entre os organismos vivos e a matéria inorgânica do planeta, ainda se tem pouco conhecimento a respeito da relação dos microrganismos com o ar e como estes são dispersados. O ar é ubíquo, ou seja, está presente em toda parte, compondo a biosfera, possuindo também uma vasta microbiota associada, porém, os estudos que tratam dessas comunidades de microrganismos subestimam a diversidade desses organismos, pois tais estudos são baseados apenas em alguma espécie cultiváveis (FANG et al., 2007; LIGHTHART, 2006; MANCINELLI; SHULLS, 1978).

Poucos estudos de fato tentaram caracterizar os microrganismos dispersos no ar, além de fungos e vírus patogênicos, e usando uma metodologia menos tendenciosa, como o sequenciamento de alta geração (AMATO et al., 2019; AN; SIN; DUBOW, 2015; MARON et al., 2005). Um desses estudos caracterizou a microbiota associada à nuvens de regiões montanhosas e foi observado uma grande diversidade de bactéria, arqueia e fungos normalmente encontrados em ambientes aquáticos, mostrando a atuação de tais microrganismos no desempenho de seus papéis como fagocitose e expressando vias de sinalização que os protegem das hostilidade do ambiente. Um estudo mais recente foi o primeiro a caracterizar

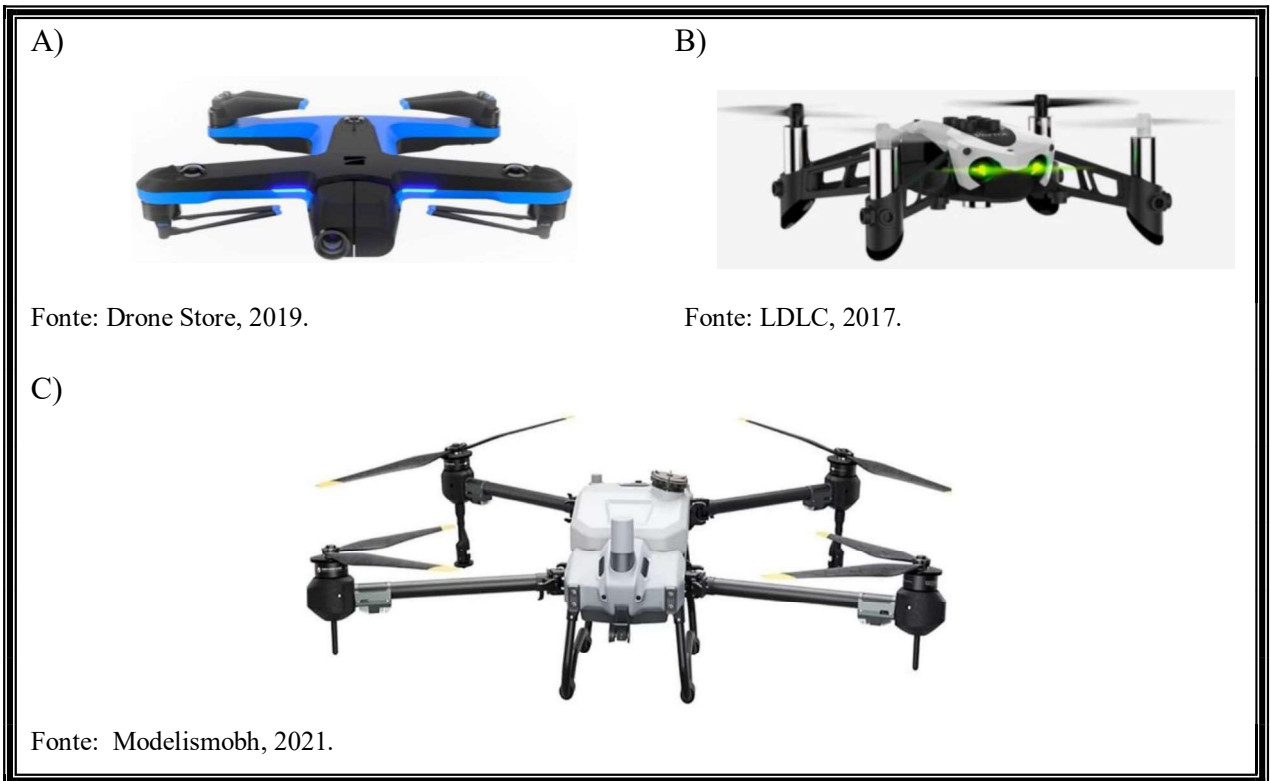
consistentemente a modulação do microbioma do ar na troposfera próxima ao solo, considerando amostragens diárias e sazonais por um longo período na cidade de Singapura.

Grandes descobertas foram encontradas a partir desse estudo, como a alta riqueza de espécies associadas ao ar, riqueza esta comparável com ambientes como solo e o mar, apesar dos eucariotos serem mais presentes. Além disso, há uma variação diária significativa da biomassa do microbioma e uma abundância e diversidade constantes e variáveis ao longo do ano de vários táxons. No entanto, uma das descobertas mais marcantes é que a maioria do DNA sequenciado (> 80%) não pôde ser identificada em nenhum nível taxonômico, sugerindo que a maior parte do microbioma do ar é composta por microrganismos não caracterizados (GUSAREVA et al., 2019b), mostrando a importância da nossa proposta.

Apesar de tais descobertas fantásticas, grandes lacunas ainda estão presentes, principalmente por causa do desafio de amostrar a baixa biomassa de microrganismos no ar. Associado a isso, há também a dificuldade relacionada com os equipamentos costumeiramente utilizados para esse tipo de coleta. Bombas pesadas e fixas são necessárias para coletar vários metros cúbicos de ar que normalmente passam por um filtro para reter microrganismos. Existem outros métodos empregados que não utilizam filtros (HARNPICHARNCHAI et al., 2023; PARKER et al., 2020; TAN et al., 2022), mas todos dependem de máquinas pesadas para sucção de ar. Tais condições de amostragem limitam substancialmente os locais de coleta e, portanto, a variabilidade das camadas de ar que podem ser amostradas. Diante disso, se faz necessário o desenvolvimento de novas metodologias que empreguem o uso de equipamentos leves e versáteis capazes de amostrar microrganismos de uma maneira mais fácil.

Drones (Figura 4) são dispositivos voadores muitas vezes autônomos, e que possuem tal versatilidade, sendo capazes de voar até 5.000 metros de altura acima do nível do mar e em velocidade constante.

Figura 4 – Modelos de drone comerciais. A) Skydio 2. B) Parrot Anafi. C) DJI Agras.



Seu uso em projetos científicos foi inicialmente proposto para análises geoestatísticas, mas também tem sido utilizados para outras análises, como na caracterização de novas espécies e conservação (KOH; WICH, 2012; NEZAMI et al., 2020), além de auxiliar na amostragem do microbioma de pequenos cetáceos (CENTELLEGHE et al., 2020). Diante disso, propomos desenvolver um método que permita coletar e investigar, a partir da metagenômica, o microbioma do ar em diferentes altitudes utilizando drones e comparar estes métodos com as técnicas tradicionais de coleta de ar.

4 MATERIAIS E MÉTODOS

4.1 DESENVOLVIMENTO DO COLETOR DE AR

O coletor foi desenvolvido pelo nosso grupo de pesquisa em parceria com a Plataforma Unificada de Modelagem e Manufatura Aditiva (PLUMMA-3D – FIOCRUZ Minas Gerais). Duas versões do coletor foram elaboradas (v1 e v2). Ambas as versões foram modeladas utilizando o programa CAD e impressos em impressora 3D através do método de Fabricação com Filamento Fundido (FFF), utilizando como matéria prima filamentos de Polietileno Tereftalato Glicol (PETG).

4.2 COLETA DAS AMOSTRAS DE AR

Utilizando o coletor v1 acoplado à um drone modelo Inspire 2 da fabricante DJI (Figura 5), foram sobrevoadas duas altitudes acima do solo (20m e 50m), e as amostras foram coletadas em duplicatas em função da autonomia de voo da aeronave.

Figura 5 – Drone modelo DJI Inspire 2 utilizado nas coletas.

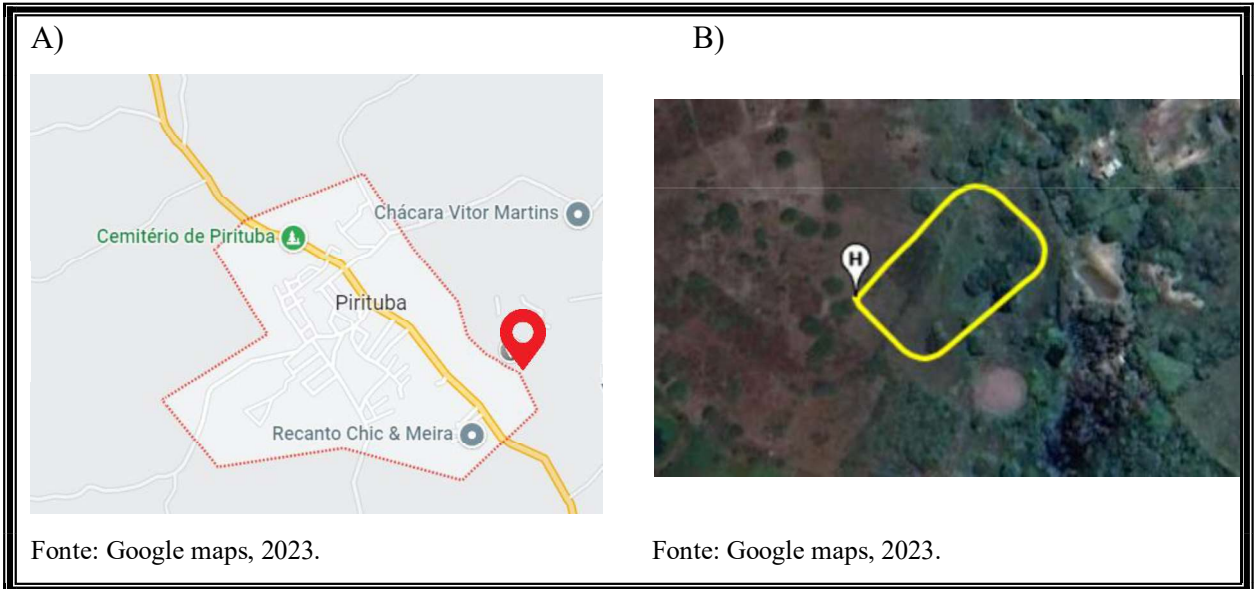


Fonte: DJI, 2016.

Os vôos tiveram duração média de 3m30s e uma velocidade de ~25km/h. Para o coletor v1, a fim de garantir a remoção dos contaminantes coletados durante a subida e descida do drone até a altura de interesse, uma vez que o coletor v1 possui um duto de sucção de ar sempre aberto, duas amostras de *background* para 20 e 50 metros foram coletadas verticalmente. As coletas foram realizadas no distrito de Pirituba, Vitória de Santo Antão-PE (-8.117425, -35.327279) (Figura 6).

O coletor v2, quando testado, foi visto que obstruía os sensores de altitude do drone, e por essa razão, por questões de segurança, foi avaliado estando apenas acoplado à um veículo (carro) em movimento na altura do solo e em diferentes velocidades (40, 50 e 60 km/h). O tempo médio de coleta foi de 2min30s. Por fim, acoplado aos coletores foram utilizados tubos cônicos, possuindo 20ml de tampão de amostra PBS 1X previamente autoclavado.

Figura 6 – Mapas representando o local de coleta e o percurso percorrido pelo drone. A) Mapa esquemático do distrito de Pirituba - Vitória de Santo Antão - PE e o local da coleta apontado com marcador vermelho. B) Mapa representando o trajeto realizado pelo drone durante as coletas. Foram percorridos no total 380 metros a partir do ponto de decolagem (H).



A fim de comparar os métodos propostos neste trabalho com métodos tradicionais de coleta de ar, coletas foram feitas utilizando uma bomba de sucção de ar modelo Ambient Fine Dust Sampler IPM-FDS-M2.5/10 (Figura 7). As coletas foram realizadas em duplicata, durante 1h cada, no teto do departamento de Biofísica da Universidade Federal de Pernambuco (UFPE). Para captura do material particulado, foram utilizados filtros de fibra de vidro e para o branco da coleta, foram usados filtros não expostos.

Figura 7 – Bomba de sucção utilizada para coleta de amostras de ar.



Fonte: Instrumex, 2019.

4.3 EXTRAÇÃO DO DNA E PCR

O DNA coletado foi extraído utilizando protocolo adaptado de acordo com (FERGUSON et al., 2019). Para as amostras coletadas através da bomba de sucção de ar, os filtros foram imersos em tubos cônicos contendo PBS 1x e vortexados posteriormente. Os filtros foram então retirados e as etapas subsequentes foram realizadas de forma igual para as amostras de filtro e para as amostras coletadas utilizando os coletores v1 e v2. Os tubos cônicos contendo as amostras foram então centrifugados a 1.792g por 45 minutos a 4°C. O sobrenadante foi

descartado por inversão e o sedimento foi ressuspensão em um tampão de lise Dodecil Sulfato de Sódio (SDS) 1% [10mM Tris-HCl pH 8, 25mM Na₂EDTA pH 8, 100mM NaCl]. A solução obtida foi incubada em banho maria à 70°C por 45 minutos e resfriada em gelo. Após isso, 500 µL de fenol/clorofórmio/alcool isoamílico (25:24:1) foi acrescentado. A solução foi vortexada 2x por 30 segundos, e então centrifugada por 5 minutos à 4°C e 16.000g. A fase aquosa (500 µL) foi transferida para um microtubo de 2,5mL e isopropanol foi adicionado (500 µL). A solução foi incubada por 12h a temperatura ambiente e após isso foi centrifugada por 30 minutos à 4°C e 16.000g. O sobrenadante foi então descartado por inversão e o sedimento lavado com etanol 70% gelado (500 µL). Por fim, a solução foi mais uma vez centrifugada por 5 minutos à 4°C e 16.000g e o sobrenadante descartado por inversão. O sedimento foi incubado a temperatura ambiente para secar e após isso foi ressuspensão em 25 µL de água ultrapura. Após a extração, o DNA foi avaliado quanto a concentração e garantia de qualidade por meio do NanoDrop® 2000 (Thermo Scientific). A reação em cadeia da polimerase (PCR) foi conduzida para cada amostra utilizando os iniciadores fornecidos pela Illumina (Part #15044223 Rev. B, Illumina, CA, USA) para as regiões V3/V4 do 16S rRNA (iniciador *forward* S-D-Bact-0341-b-S-17 5'- CCTACGGGNGGCWGCAG e iniciador *reverse* S-D-Bact-0785-a-A-21 5'- GACTACHVGGGTATCTAATCC) (KLINDWORTH et al., 2013). O programa utilizado no termociclador consiste em 98°C por 30 segundos, seguido de 30 ciclos de 98°C por 10 segundos, 55°C por 30 segundos e 72°C por 1 minuto e 30 segundos, finalizando com 72°C por 10 minutos à 4°C (Eppendorf Mastercycler EP Gradient S). Os componentes utilizados na PCR (volume final 50 µL por amostra) contem 5 µL de DNA template, 5x tampão phusion HF, 10 mM de dNTP, 10 µM de cada iniciador, 50 mM de MgCl₂, DMSO 100% e 1U de Phusion High-fidelity DNA polymerase (New England Biolabs). Para o controle positivo da PCR, foi utilizado o DNA genômico de *Yersinia pestis* cedido para o estudo pelo Serviço de Referência em Peste (SRP). A solução de componentes utilizados na PCR, sem DNA molde, foi utilizado como branco. O resultado da PCR foi avaliado através de eletroforese em gel de agarose 1% corado com SYBR™ Green. O produto da PCR foi então purificado utilizando beads magnéticas (Agencourt AMPure XP, Beckman Coulter™).

4.4 SEQUENCIAMENTO

Após as etapas de extração, avaliação, amplificação e purificação do DNA, as bibliotecas de sequenciamento foram construídas utilizando o kit Nextera XT Index Kit (Illumina, San Diego, CA, EUA), seguindo as instruções do fabricante. As bibliotecas foram purificadas com beads magnéticas (Agencourt AMPure XP, Beckman Coulter™) e quantificadas utilizando o

kit Qubit™ dsDNA HS. O sequenciamento foi realizado com o kit v3 600 ciclos para bibliotecas do tipo “pair-end” e no sequenciador Illumina MiSeq.

4.5 ANÁLISES COMPUTACIONAIS

A análise exploratória da qualidade dos dados foi conduzida inicialmente com o programa fastqc (v0.11.9) (ANDREWS S., 2010). O programa fastp (v. 0.23.2) então foi utilizado para o tratamento das bases ambíguas (Ns) presente nas leituras (<https://github.com/OpenGene/fastp>) (CHEN et al., 2018). Posteriormente, devido à baixa qualidade das leituras *reverse*, análises tipo *single-end* utilizando as leituras *forward* foram conduzidas e a trimagem das leituras foi realizada, de acordo com a qualidade phred de 20, utilizando o pacote DADA2 (v1.20.0) (CALLAHAN et al., 2016) no ambiente de programação R (v3.6.3). Também através do DADA2, o modelo de erro foi estabelecido, as duplicatas foram tratadas e a etapa de inferência das ASVs foi configurada para “*pooled*”. Por fim, procedeu-se a remoção das quimeras também através do pacote DADA2. As ASVs foram então classificadas taxonomicamente utilizando o banco de dados SILVA (v138.) (QUAST et al., 2013) e a função IDTaxa do pacote DECIPHER (v.2.12.0) (MURALI; BHARGAVA; WRIGHT, 2018). As análises subsequentes foram realizadas em um computador local, utilizando o ambiente de programação R (v4.3.0), devido à limitação da versão v.3.6.3 para gerenciar pacotes mais atuais e que foram necessários para etapas de filtragem e análises de diversidade.

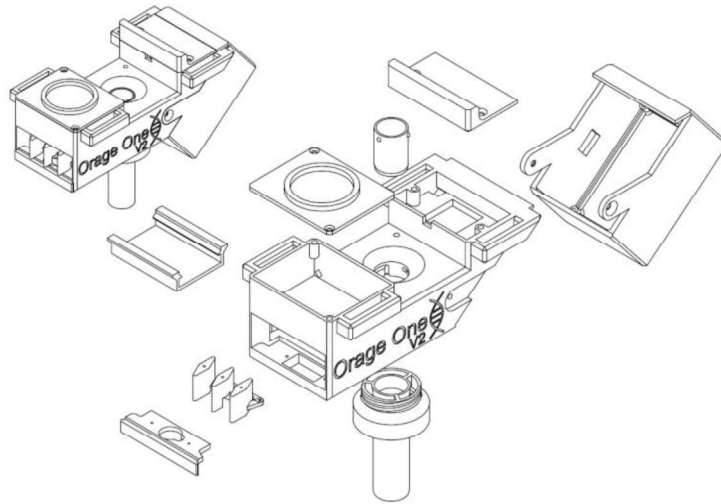
As amostras de controle (foram utilizados para filtragem das demais amostras através do pacote SCRuB (v0.0.1) (AUSTIN et al., 2023). As tabelas de contagem das ASVs e da classificação taxonômica foram compiladas, junto com os metadados, em um objeto utilizando o pacote phyloseq (v1.44.0) (MCMURDIE; HOLMES, 2013). Utilizando o objeto phyloseq, análises dos índices de diversidade alfa e beta foram conduzidas para avaliar a diversidade intra e entre as amostras, respectivamente. A diversidade alfa foi calculada utilizando o índice de Shannon. A diversidade beta por sua vez foi calculada utilizando a dissimilaridade de Bray-curtis e as análises estatísticas feitas utilizando o índice de similaridade (ANOSIM).

5 RESULTADOS

5.1 DESENVOLVIMENTO DOS COLETORES DE AR PARA OBTENÇÃO DE AMOSTRAS DE MICROBIOMA

Os coletores desenvolvidos foram avaliados quanto a sua capacidade de coletar a biomassa associada ao material particulado suspenso no ar (Figura 8).

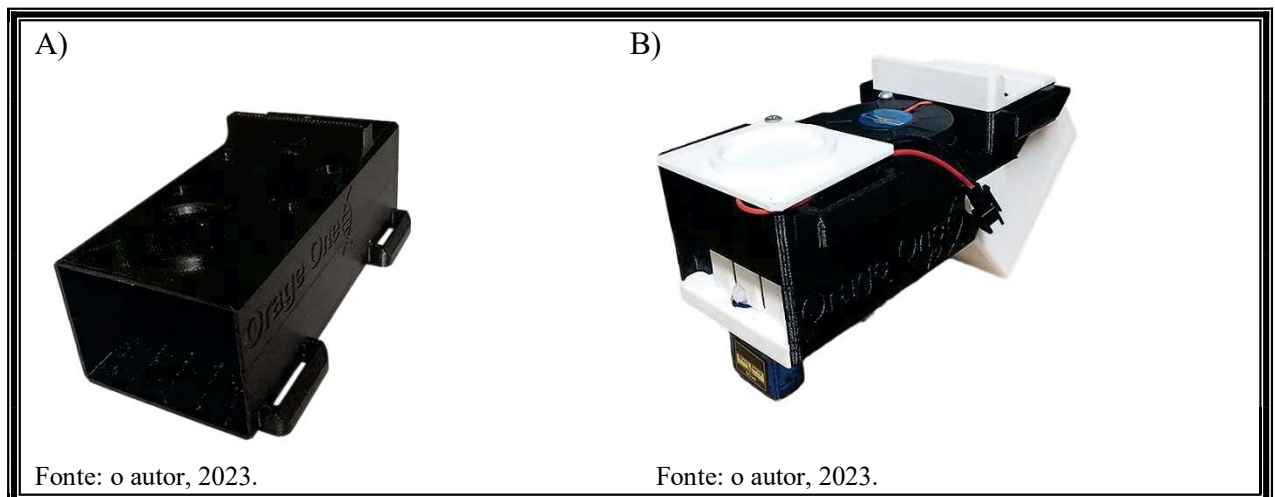
Figura 8 – Modelo esquemático do coletor v2. O modelo do coletor v2 se assemelha ao coletor v1, com a exceção de que o coletor v2 possui aberturas no duto que são controladas eletronicamente, além de possuir um resfriador na parte superior que serve para sucção de ar através do duto.



Fonte: PLUMMA-3D, 2022.

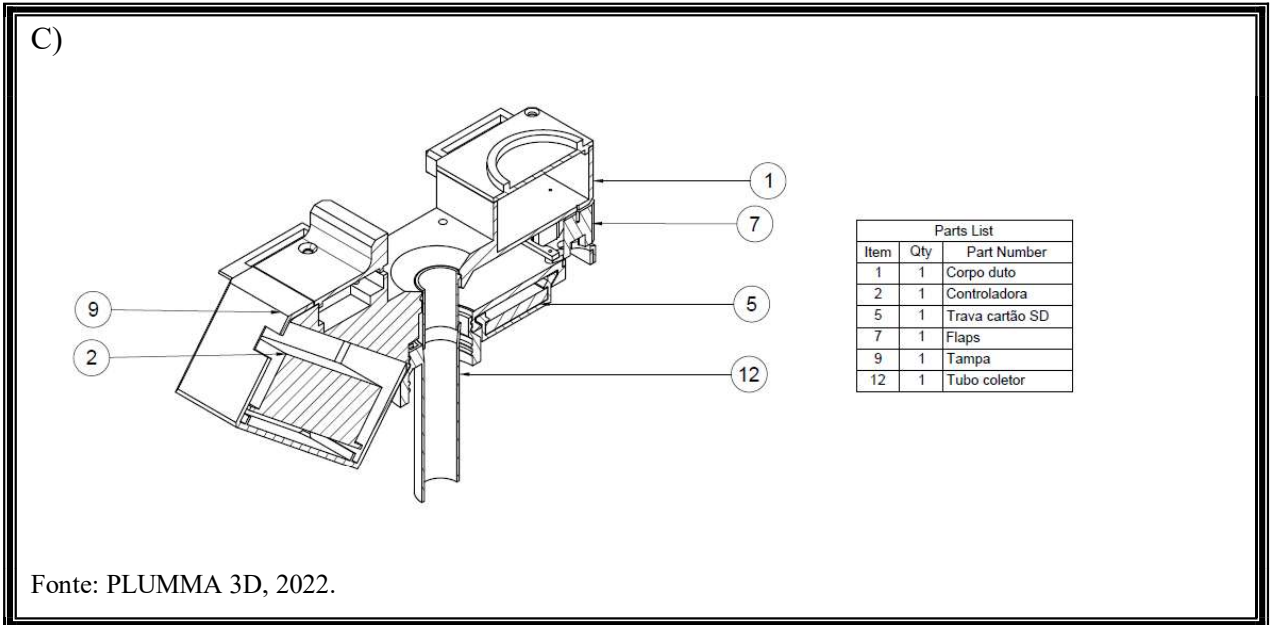
A primeira versão (v1) (Figura 9A) possui um mecanismo de sucção passiva, dependendo apenas do arraste do drone, enquanto a segunda versão (v2) (Figura 9B) possui um mecanismo de sucção ativa que é realizado por um resfriador radial de 35mm. Além disso, o coletor v1 não possui controle da entrada e saída de ar, estando o duto de entrada de ar sempre aberto; já o coletor v2 possui um controlador Arduino Nano, que regula a abertura e o fechamento do duto de sucção de ar, além de possuir também um barômetro BMP280 que é capaz de medir a pressão atmosférica e a partir disso calcular a altura a partir do solo (Figura 9C).

Figura 9 – Coletores desenvolvidos pelo grupo. A) Coletor v1. B) Coletor v2. C) Modelo em corte transversal do coletor v2, onde é indicado seus principais componentes.



Fonte: o autor, 2023.

Fonte: o autor, 2023.



A medição da pressão atmosférica permite que o coletor v2 possa ser configurado para abrir na altura de interesse de quem vai utilizá-lo. Para o coletor v2, as informações referentes à altura, abertura do duto e funcionamento do resfriador são registradas em um arquivo de texto gravado em um cartão de memória que vai acoplado ao coletor (Tabela 1).

Tabela 1 – Modelo de registro gravado em formato de texto no cartão de memória acoplado ao coletor v2. A primeira, terceira e quarta coluna registram os dados de forma binária. A primeira coluna indica se o coletor de fato está coletando (0 = não; 1 = sim). A segunda coluna indica a altura registrada a cada segundo. A terceira coluna indica se o flap do duto está aberto (0 = não; 1 = sim). Por fim, a quarta coluna indica se o resfriador está ligado (0 = não; 1 = sim).

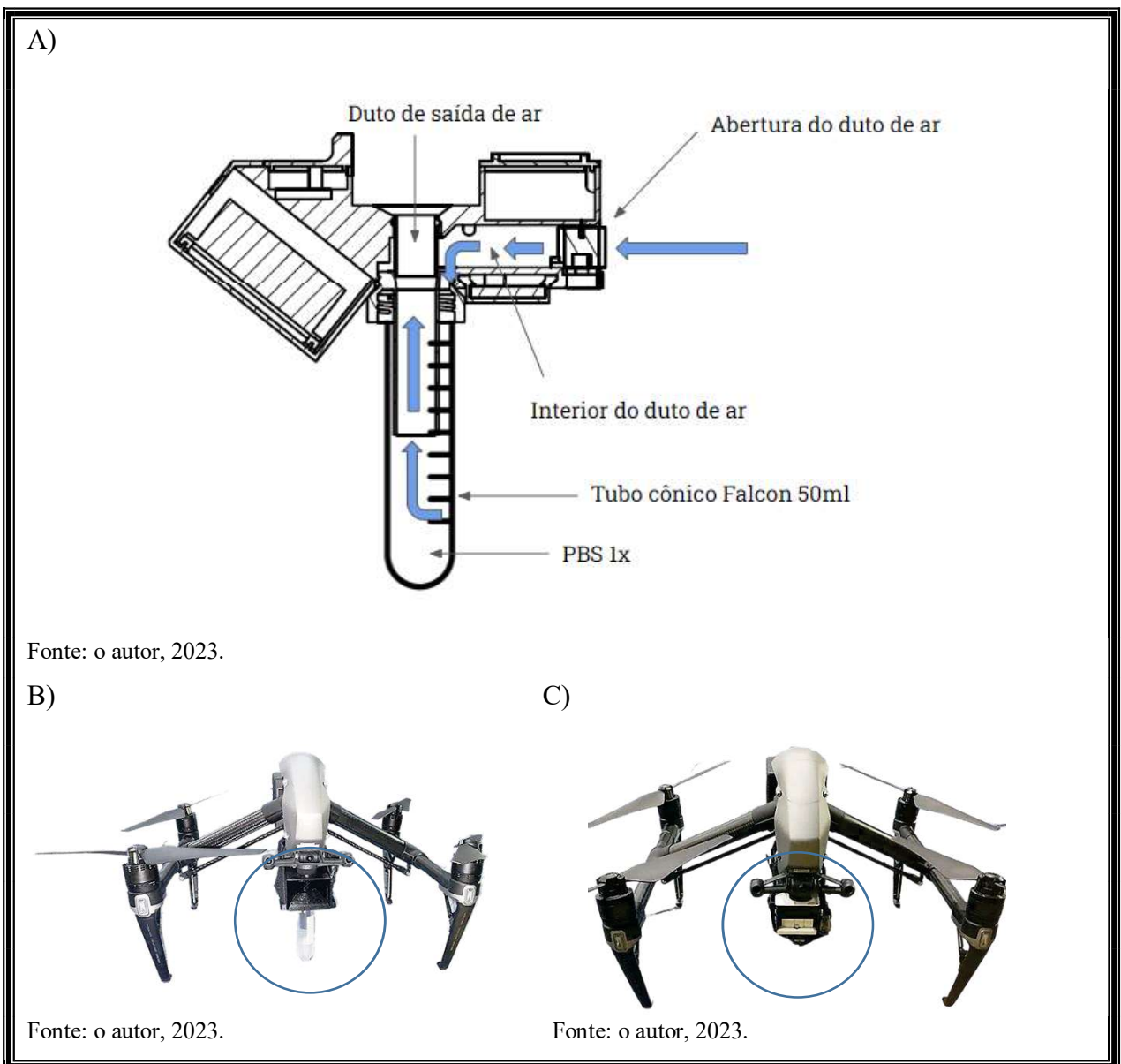
Coleta	Altura	Flap	Resfriador
0	18	0	0
0	19	0	0
1	20	1	1

Fonte: o autor, 2023.

Para além das diferenças nos componentes, o fluxo de ar funciona de forma semelhante para ambos os coletores (Figura 10^a-C).

Figura 10 – Esquema representativo do funcionamento do coletor. A) Esquema representando o funcionamento do fluxo de ar dentro dos coletores. O ar (seta azul) entra pela abertura do duto de ar, passa através do interior do coletor e desce por uma passagem entre o duto de saída de ar e a parede do tubo cônico. O ar então entra em contato com a solução PBS 1x contida no tubo cônico e então sobe para sair pelo duto de saída de ar, onde também é localizado o resfriador

radial responsável pela sucção de ar. B) Coletor v1 e seu acoplamento ao drone modelo DJI Inspire 2. C) Coletor v2 e seu acoplamento ao drone modelo DJI Inspire 2.



5.2 AMOSTRAS DE AR COLETADAS APRESENTARAM BOM RENDIMENTO E POTENCIAL PARA SEREM AMPLIFICADAS

O material extraído foi avaliado quanto a sua concentração e qualidade utilizando o NanoDrop® 2000 (Thermo Scientific). Os resultados estão expressos na tabela abaixo (Tabela 2).

Tabela 2 – Qualidade e concentração do DNA extraído.

Amostra	Concentração (ng/ μ l)	260/280
Controle – Extração (PBS)	11,6	1,44

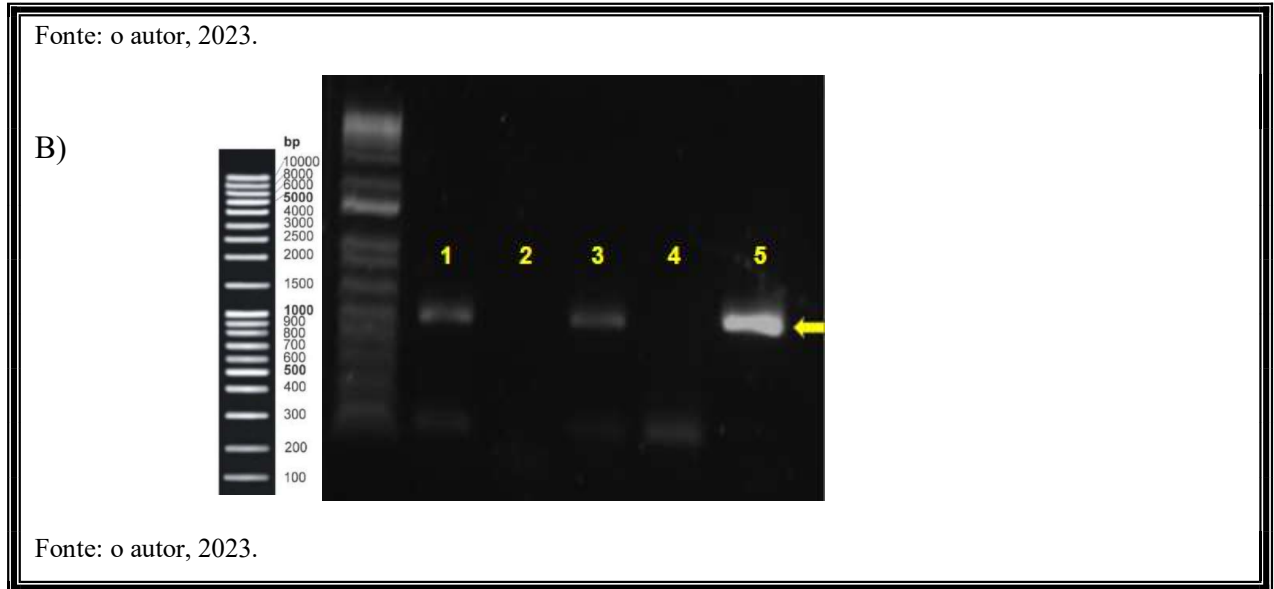
20 metros – background	28,1	1,46
Coletor 1 – 20 metros	7,4	1,35
Coletor 1 – 20 metros	24,9	1,41
50 metros – background	47,9	1,43
Coletor 1 – 50 metros	11,1	1,28
Coletor 1 – 50 metros	26,5	1,39
Coletor 2 – 40km/h	36,5	1,46
Coletor 2 – 40km/h	12,0	1,57
Coletor 2 – 50km/h	30,0	1,46
Coletor 2 – 60km/h	8,4	1,31
Bomba de sucção – Branco	49,3	1,47
Bomba de sucção – filtro	80,5	1,47
Bomba de sucção – filtro	57,2	1,49

Fonte: O autor, 2023.

Após a extração, prosseguiu-se para a amplificação do material utilizando os iniciadores universais para região V3/V4 do rRNA 16S fornecidos pela Illumina (Part #15044223 ver. B, Illumina, CA, USA). O resultado pode ser visualizado em gel de agarose 1% (Figura 11 A-B).

Figura 11 – Resultados das PCRs para amplificação da região de interesse. A) Gel de agarose 1% da região rRNA 16S V3/V4 amplificada para amostras de ar obtidas com os coletores. 1) controle extração (PBS); 2) controle PCR (Yersinia); 3) branco PCR (Mix); 4) 20 metros – background; 5 e 6) Coletor-1 – 20 metros; 7) 50 metros – background; 8 e 9) Coletor-1 – 50 metros; 10 e 11) Coletor 2 – 40km/h; 12) Coletor 2 – 50km/h; 13) Coletor 2 – 60km/h. B) Gel de agarose 1% da região rRNA 16S V3/V4 amplificada para amostras de ar obtidas com a bomba de sucção. 1) Bomba de sucção – Branco; 2 e 3) Bomba de sucção – filtro; 4) branco PCR (Mix); 5) controle PCR (Yersinia).





Através dos resultados da eletroforese, a amplificação do fragmento do V3/V4 do rRNA 16S para nossas amostras foi confirmada, com o tamanho esperado de ~500 pb, tanto para as amostras obtidas com ambas as versões dos coletores (v1 e v2) quanto para as amostras obtidas através da bomba de sucção.

5.3 SEQUENCIAMENTO FOI CAPAZ DE GERAR QUANTIDADE SUFICIENTE DE DADOS PARA CARACTERIZAÇÃO DA COMPOSIÇÃO MICROBIANA

Após a purificação do produto da PCR, prosseguiu-se para o preparo da biblioteca com a ligação dos adaptadores aos fragmentos amplificados na etapa anterior. Foi utilizado o kit Nextera XT Index Kit (Illumina, San Diego, CA, EUA) de acordo com o protocolo Illumina. A biblioteca foi então quantificada utilizando o kit Qubit™ dsDNA HS (Tabela 3) e o tamanho foi validado através de eletroforese em gel de agarose 1% (Figura 12).

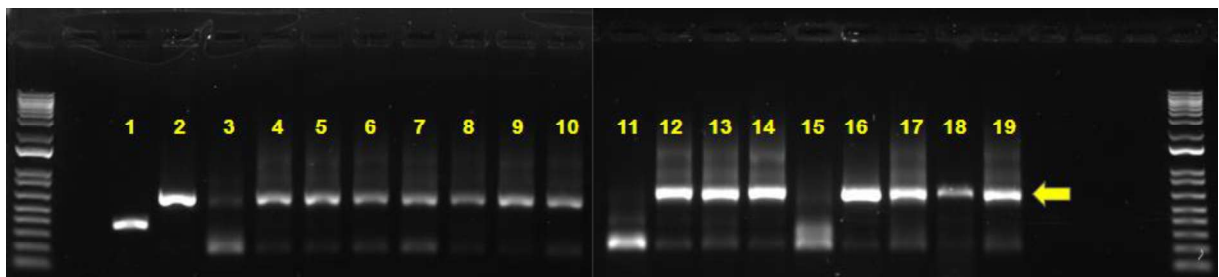
Tabela 3 – Concentração das bibliotecas.

Amostra	Concentração (ng/ µl)
Controle PCR (Yersinia)	33,6
Branco PCR (Mix)	4,77
Controle extração (PBS)	25,5
Coletor 2 – 60km/h	25,1
Coletor 2 – 50km/h	23,7
Coletor 2 – 40km/h	18,8

Coletor 2 – 40km/h	19,4
20 metros – background	24,6
Coletor 1 – 20 metros	20,1
Coletor 1 – 20 metros	0,188
50 metros – background	24,5
Coletor 1 – 50 metros	20,1
Coletor 1 – 50 metros	23,1
Branco PCR (Mix)	3,85
Controle PCR (Yersinia)	28,8
Bomba de sucção – Branco	21,1
Bomba de sucção – filtro	0
Bomba de sucção – filtro	20,6

Fonte: o autor, 2023.

Figura 12 – Eletroforese em gel de agarose 1% das bibliotecas. 1) Controle Qpcr; 2) Controle PCR (Yersinia); 3) Branco PCR (Mix); 4) Controle extração (PBS); 5) Coletor 2 – 60km/h; 6) Coletor 2 – 50km/h; 7) Coletor 2 – 40km/h; 8) Coletor 2 – 40km/h; 9) 20 metros – background; 10) Coletor 1 – 20 metros; 11) Coletor 1 – 20 metros; 12) 50 metros – background; 13) Coletor 1 – 50 metros; 14) Coletor 1 – 50 metros; 15) Branco PCR (Mix); 16) Controle PCR (Yersinia); 17) Bomba de sucção - Branco; 18) Bomba de sucção - filtro; 19) Bomba de sucção - filtro.



Fonte: o autor, 2023.

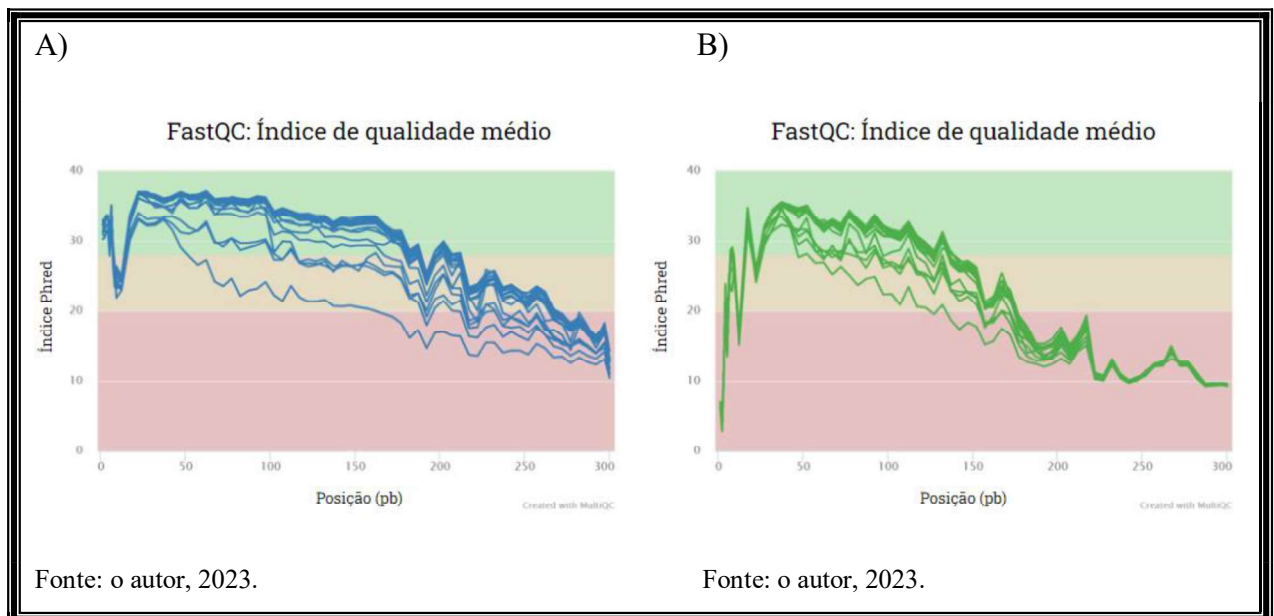
A partir da análise do gel, foi possível confirmar a ligação dos adaptadores nas sequências presentes nas amostras, que por fim apresentaram um tamanho aproximado de ~600 pb, assim como estabelecido no protocolo Illumina (Part #15044223 Rev. B, Illumina, CA,

USA). A partir disso, as bibliotecas foram então diluídas, misturadas e desnaturadas para posterior sequenciamento no sequenciados Illumina MiSeq.

5.4 ANÁLISES DE BIOINFORMÁTICA SE MOSTRARAM CAPAZES DE CARACTERIZAR ASPECTOS DO MICROBIOMA AÉREO

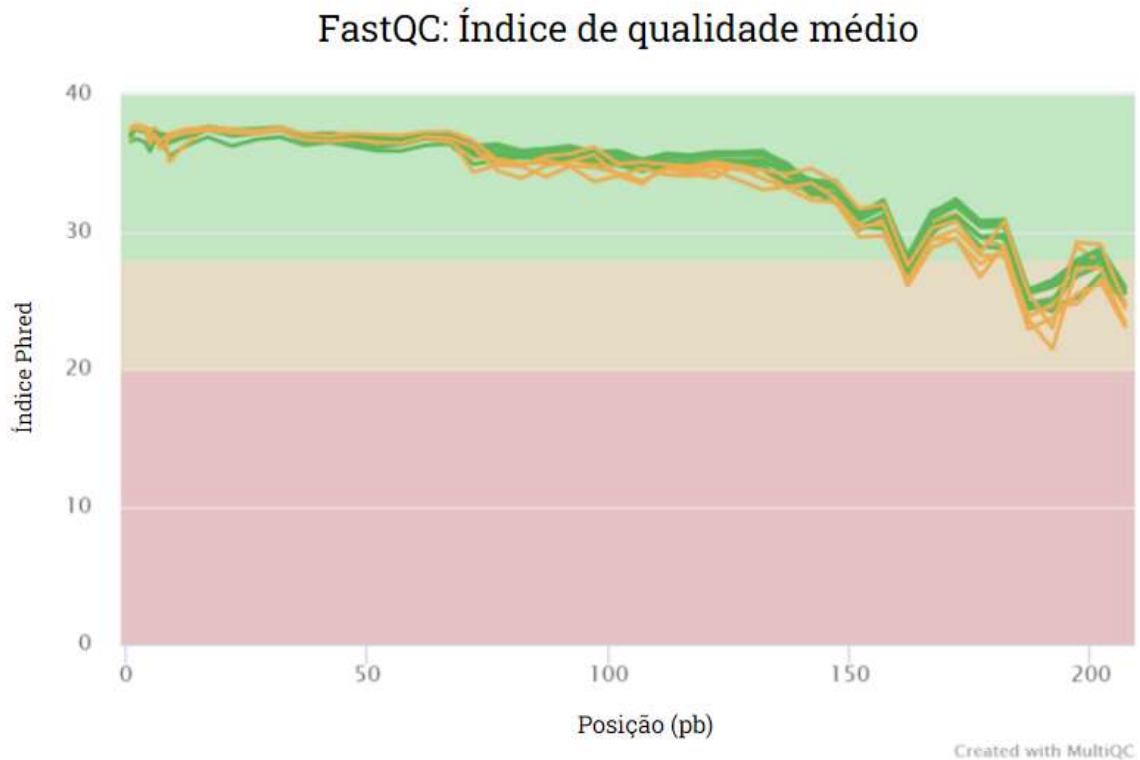
Um total de 20.141.040 de leituras foram convertidas em arquivos fastq que serviram de input para as análises computacionais subsequentes. Devido à baixa qualidade das leituras *reverse*, estas foram excluídas das análises posteriores (Figura 13).

Figura 13 - Qualidade média das leituras A) *forward* e B) *reverse*. No eixo y está descrito a qualidade Phred e no eixo x está descrito a posição em pares de base das sequências.



As leituras *forward* foram então filtradas e trimadas, através da ferramenta DADA2, em 209 pb, considerando uma qualidade Phred de 20 (Figura 14).

Figura 14 – Qualidade média das leituras *forward* após trimagem. No eixo y está descrito a qualidade Phred e no eixo x está descrito a posição em pares de base das sequências.



Fonte: o autor, 2023.

Por fim, após a remoção de quimeras, um total de 3.006.009 milhões de leituras foram utilizadas para a etapa inferência de ASVs, com uma média de ~160.000 leituras por amostra (Tabela 4). Um total de 1878 ASVs foram preditas ao longo das amostras e tais ASVs foram classificadas em diferentes níveis taxonômicos (e.g. filo) (Figura 15). Posteriormente as contagens das amostras de controle da extração (PBS) e de background para 20 e 50 metros foram utilizadas como filtros e subtraídas das demais amostras, portanto não estão presentes nas análises de diversidade.

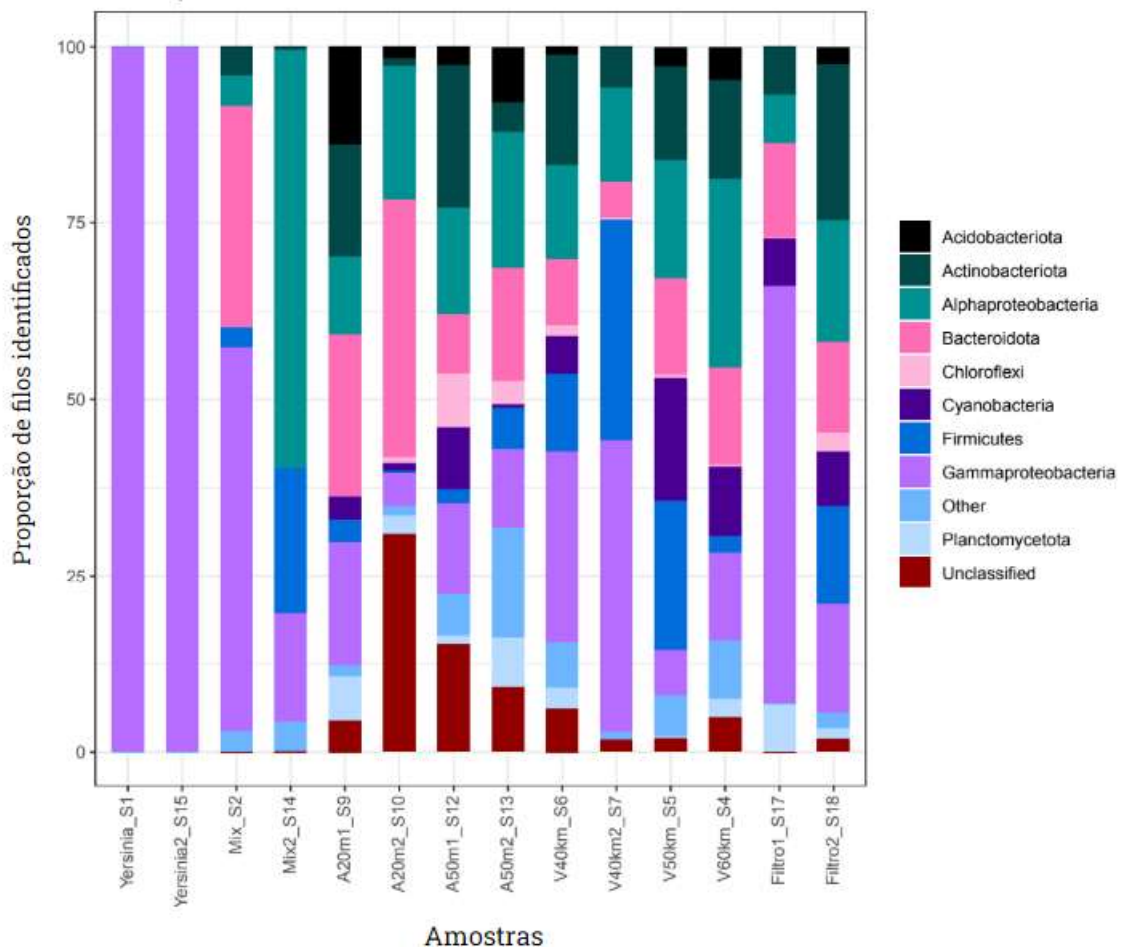
Tabela 4 – Quantidade de leituras finais por amostra após filtragem, tratamento de qualidade e remoção de quimeras.

Amostra	Entrada	Filtrado	Não quimeras	% de leituras retidas
Controle PCR (Yersinia)	212997	202990	202762	95.2
Branco PCR (Mix)	79422	74165	68468	86.2
Controle extração (PBS)	269639	248204	230546	85.5
20 metros - background	256186	238632	219671	85.7
Coletor 1 - 20 metros	275351	261074	250733	91.1
Coletor 1 - 20 metros	9350	2502	2419	25.9

50 metros - background	251473	235163	217635	86.5
Coletor 1 - 50 metros	285598	261056	244320	85.5
Coletor 1 - 50 metros	244143	230455	210698	86.3
Coletor 2 – 40km/h	317194	287669	273893	86.3
Coletor 2 – 40km/h	315048	279526	269285	85.5
Coletor 2 – 50km/h	240379	224148	208908	86.9
Coletor 2 – 60km/h	323133	296957	273742	84.7
Controle PCR (Yersinia)	123847	115486	115426	93.2
Branco PCR (Mix)	9936	9168	9045	91.0
Bomba de sucção - Branco	39692	36705	34430	86.7
Bomba de sucção - filtro	197	178	167	84.8
Bomba de sucção - filtro	197698	187400	173861	87.9

Fonte: o autor, 2023.

Figura 15 – Proporção de filamentos identificados em cada amostra. No eixo y temos a proporção de cópias do gene do rRNA 16S de acordo com os filamentos expostos na relação no canto direito. No eixo x temos as amostras (Yersinia e Yersinia2 = controle positivo da PCR; Mix e Mix2 = branco da PCR; A20m1 e A20m2 = duplicatas coletadas com o drone e coletor v1 à 20m de altura do solo; A50m1 e A50m2 = duplicatas coletadas com o drone e coletor v1 à 50m de altura do solo; V40km e V40km2 = amostras coletadas com coletor v2 acoplado à um carro em movimento em velocidade de 40km/h; V50km = amostras coletadas com coletor v2 acoplado à um carro em movimento em velocidade de 50km/h; V60km = amostras coletadas com coletor v2 acoplado à um carro em movimento em velocidade de 60km/h; Filtro1 e Filtro2 = amostras coletadas em duplicada utilizando filtros e bomba de sucção).

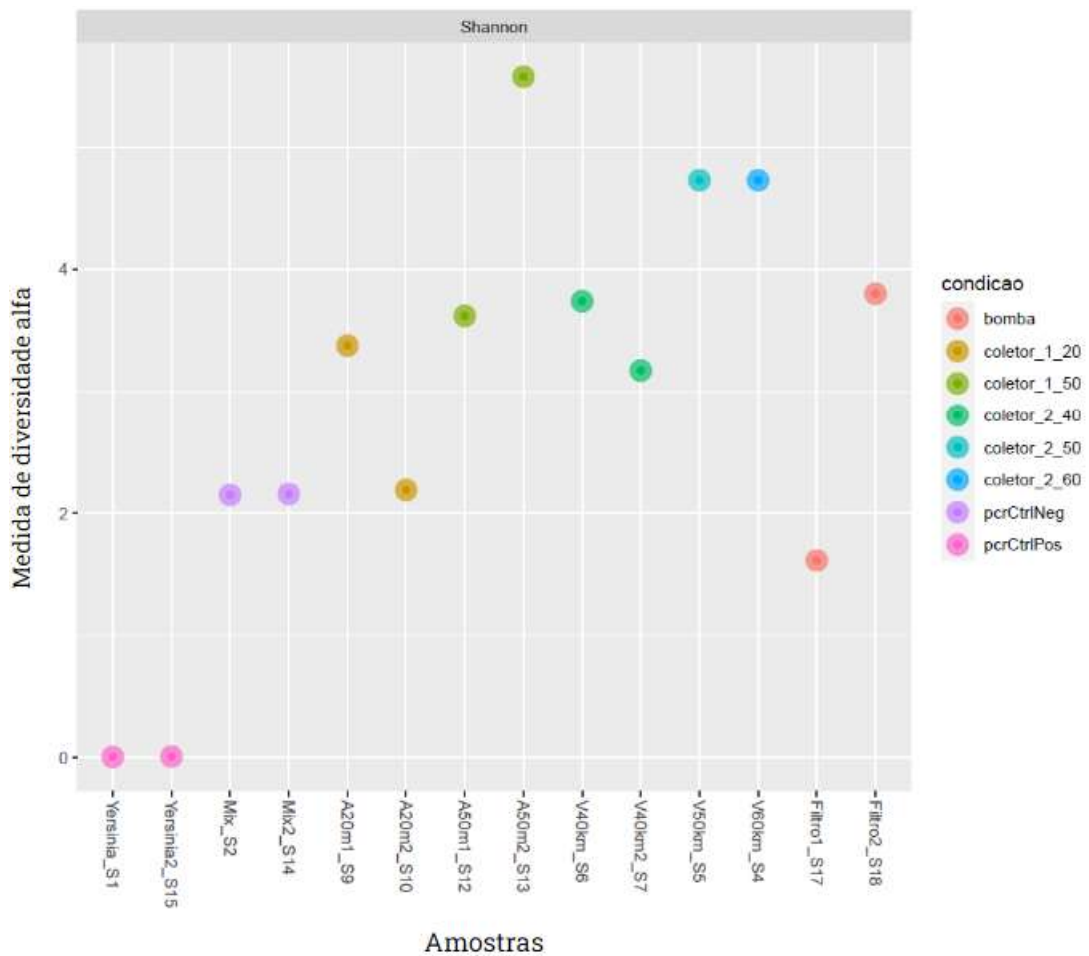


Fonte: o autor, 2023.

Além da abundância de táxons presentes nas amostras, a diversidade alfa foi calculada para cada amostra a partir do índice de Shannon, que estabelece a riqueza presente em cada amostra (Figura 16). No gráfico abaixo é possível observar uma tendência, onde amostras coletadas em altitudes maiores (50 metros) e em maior velocidade (50 e 60 km/h) apresentam uma maior riqueza. Também é possível observar variabilidade na riqueza entre amostras coletadas em uma mesma condição.

Figura 16 – Medida de diversidade alfa das amostras a partir do índice de Shannon. No eixo y é exibido a medida de diversidade alfa e no eixo x as amostras (Yersinia e Yersinia2 = controle positivo da PCR; Mix e Mix2 = branco da PCR; A20m1 e A20m2 = duplicatas coletadas com o drone e coletor v1 à 20m de altura do solo; A50m1 e A50m2 = duplicatas coletadas com o drone e coletor v1 à 50m de altura do solo; V40km e V40km2 = amostras coletadas com coletor v2 acoplado à um carro em movimento em velocidade de 40km/h; V50km = amostras coletadas com coletor v2 acoplado à um carro em movimento em velocidade de 50km/h; V60km = amostras coletadas com coletor v2 acoplado à um carro em movimento em velocidade de

60km/h; Filtro1 e Filtro2 = amostras coletadas em duplicada utilizando filtros e bomba de sucção).

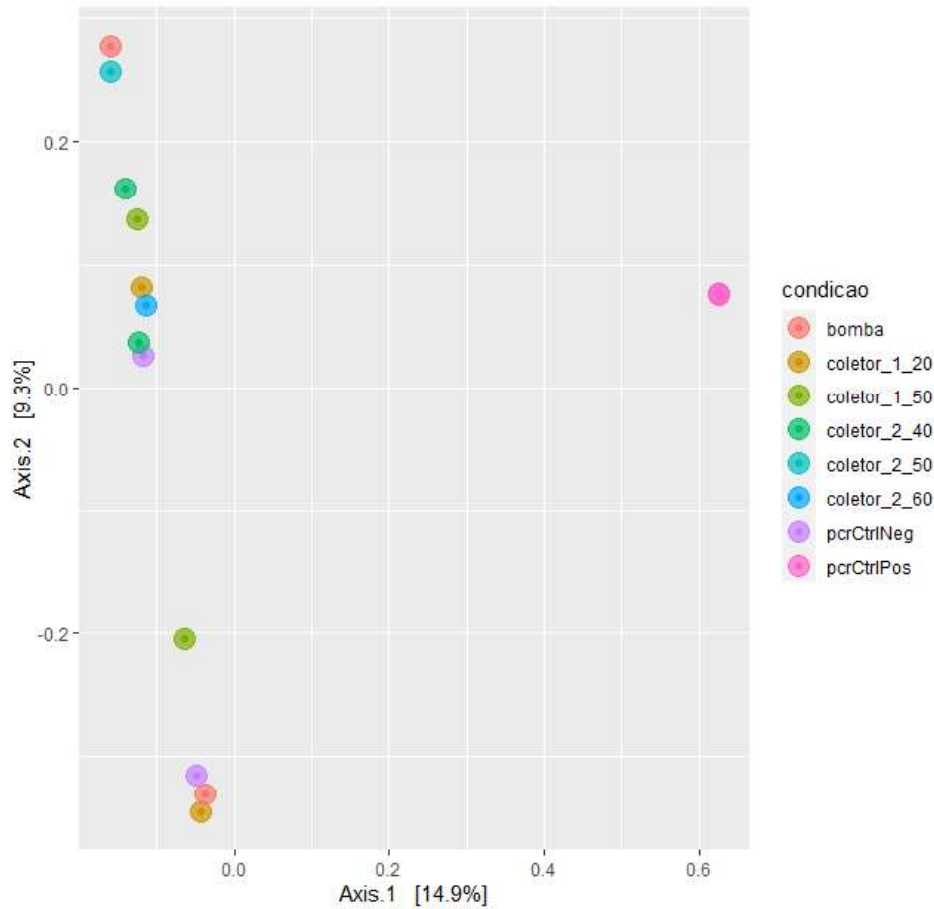


Fonte: o autor, 2023.

Também foi calculada a diversidade beta que avalia a similaridade entre as amostras. Tal cálculo foi feito utilizando a dissimilaridade de Bray-curtis (Figura 17). As análises de similaridade (ANOSIM) indicaram que os grupos apresentam mais semelhanças do que diferenças entre si (ANOSIM statistic R: 0.3059). Por fim, foi observado que não houve diferença significativa entre a similaridade das amostras (p-valor = 0.0665).

Figura 17 – Diversidade beta avaliada entre as amostras utilizando a dissimilaridade de Bray-curtis. Os pontos representam as condições (pcrCtrlPos = controle positivo da PCR; pcrCtrlNeg = branco da PCR; coletor_1_20 = duplicatas coletadas com o drone e coletor v1 à 20m de altura do solo; coletor_1_50 = duplicatas coletadas com o drone e coletor v1 à 50m de altura do solo; coletor_2_40 = amostras coletadas com coletor v2 acoplado à um carro em movimento em velocidade de 40km/h; coletor_2_50 = amostras coletadas com coletor v2 acoplado à um carro em movimento em velocidade de 50km/h; coletor_2_60 = amostras coletadas com coletor v2

acoplado à um carro em movimento em velocidade de 60km/h; bomba = amostras coletadas em duplicada utilizando filtros e bomba de sucção).



Fonte: o autor, 2023.

6 DISCUSSÃO

Neste trabalho, propomos uma nova forma de se amostrar o microbioma aéreo e analisá-lo através da técnica de metagenômica por *amplicons*. Como mencionado anteriormente, há uma grande dificuldade de se estudar o ambiente aéreo devido à escassez de métodos eficazes de coleta e tratamento das amostras de ar. A coleta com drones surgiu como uma possibilidade pois tais equipamentos permitem uma versatilidade na amostragem de diferentes ambientes e altitudes. Os drones vem se tornando cada vez mais utilizados para uma ampla gama de finalidade devido a versatilidade mencionada acima, finalidades estas como, por exemplo, coleta de compostos orgânicos presentes no ar e para mensuração da biomassa de plantas em ambientes áridos (MCCANN; KEITH; KINGSFORD, 2022; PUSFITASARI et al., 2022). Explorando mais ainda a possibilidade do uso de drones, estes começaram a ser empregados na coleta de amostras ambientais como a água, para caracterização da comunidade microbológica ali presente (HORRICKS et al., 2022). Apesar de tais inovações, ainda não se tem na literatura

o registro do uso de drones para coleta de amostras de ar, juntamente com o emprego de técnicas de sequenciamento de alto rendimento como a metagenômica para a caracterização do microbioma aéreo. Dessa forma, este trabalho se propõe a estabelecer uma nova alternativa para a coleta de amostras de ar a fim de que a comunidade microbiológica associada ao ambiente aéreo possa ser efetivamente amostrada e caracterizada.

Além do desenvolvimento de métodos de coleta, tais como o emprego do drone e o desenvolvimento de um coletor capaz de capturar amostras de ar, o resultado depende amplamente do tratamento que as amostras irão receber posteriormente. Por possuir uma baixa densidade de biomassa, é particularmente difícil realizar uma extração de material genético, a partir do ar, com um bom rendimento e qualidade. Protocolos que utilizam kits de extração, geralmente a partir de colunas de sílica, possuem baixo rendimento, retendo boa parte do material genético, enquanto que métodos como fenol/clorofórmio possuem um melhor rendimento, ao menos em amostras de tecido (SHUTTLEWORTH; OOSTHUIZEN, 2022). Por outro lado, o emprego de fenol/clorofórmio traz desvantagens como a alta variação na reprodução dos resultados e a presença de contaminantes fenólicos que, por exemplo, podem inibir a PCR, além de conter reagentes químicos altamente tóxicos (SCHRADER et al., 2012; TONI et al., 2018).

Esses fatores poderiam explicar a variação na amplificação em algumas amostras do nosso experimento. Essas variações se refletiram nas análises posteriores, com o baixo número de leituras sequenciadas em algumas amostras (e.g. Bomba de sucção – filtro 1). Tal diferença na quantidade de leituras prejudica diretamente a avaliação da composição microbiana de uma amostra, e usualmente tais situações são contornadas com técnicas de normalização do tamanho das bibliotecas, a fim de evitar que uma amostra sequencie mais do que outra.

A rarefação é uma técnica que envolve a seleção aleatória de uma quantidade de leituras a fim de que se iguale os dados entre as amostras. Essa quantidade é definida baseada na menor amostra sequenciada e envolve o descarte dos dados que excederam ou que ficaram abaixo do limiar estabelecido (WILLIS, 2019). Entretanto, a fim de contornar o viés da quantidade de leituras sequenciadas, tal técnicas insere um outro viés, pois grande parte dos dados descartados são de leituras raras, e que impactam algumas métricas de diversidade que são altamente dependente dessas leituras raras, como o índice de Shannon (KONOPÍŃSKI, 2020).

Essas ponderações foram muito bem exploradas por (MCMURDIE; HOLMES, 2014) que sugerem outras técnicas para normalização das amostras (e.g. transformação estabilizadora de variância, DESeq2). Por essa razão, decidimos não realizar a rarefação nos nossos dados, pois isso nos forneceria uma cenário mais aproximado da comunidade bacteriana presente no

ar. Apesar disso, foram observadas diferenças relevantes entre amostras coletadas na mesma condição, tanto para diversidade alfa quanto para diversidade beta, e podemos supor que a razão para tal pode ser a alta variação no método de extração, que impactou análises subsequentes como a PCR e conseqüentemente o sequenciamento, levando a um desbalanceamento na quantidade de leituras.

Contaminações em estudos de metagenômica são recorrentes e podem levar a uma falsa interpretação dos dados gerados (DAVIS et al., 2018), porém a hipótese de que tais contaminantes, sejam eles advindos de reagentes ou de outras fontes, são a fonte da variação observada entre amostras de mesmo tratamento pode ser afastada uma vez que as leituras classificadas na amostra de controle de extração (PBS) foram subtraídas das demais amostras, e ainda assim, uma ampla variação pôde ser observada. Dessa forma é imperativo o estabelecimento de protocolos de extração que se proponham não apenas a extrair material genético de amostras de baixa biomassa, mas também que possuam um alto grau de reprodutibilidade, a fim de mitigar o viés causado por etapas de processamento das amostras.

Para além das dificuldades citadas acima, os dados taxonômicos corroboram alguns achados sobre a diversidade bacteriana encontrada no ambiente aéreo, como a ampla diversidade de filos bacterianos identificados nas amostras além da alta taxa de leituras não classificadas em qualquer nível taxonômico (GUSAREVA et al., 2019b), característica observada em amostras coletada à 20 metros de altitude, onde ~30% de leituras não foram classificadas em qualquer nível taxonômico. Tal achado lança luz sobre a lacuna no conhecimento da comunidade microbiana presente no ar, lacuna essa que pode ser contornada a partir de estudos mais aprofundados, incluindo outras técnicas de sequenciamento, como sequenciamento de “*shotgun*”, que permitiria não apenas a resolução, em partes, da identificação de tais táxons, mas também a dinâmica funcional que ocorre no ambiente aéreo.

7 CONCLUSÃO

- A partir dos dados obtidos nesse trabalho, ficou evidente a necessidade do estabelecimento de técnicas de processamento de amostras genéticas que permitam uma maior reprodutibilidade associada à um maior rendimento em amostras de baixa biomassa.
- Os protocolos desenvolvidos e executados nesse trabalho se mostraram capazes de mostrar, ainda que minimamente, as características da comunidade bacteriana no ar, especialmente sua composição taxonômica.
- Espera-se então que este estudo possa contribuir com o conhecimento acerca do microbioma presente no ar e que sirva de estímulo para o desenvolvimento e aprimoramento de novas técnicas para o estudo de um componente da biosfera tão amplo e tão pouco explorado como o ar.

REFERÊNCIAS

- ALSER, M. et al. Technology dictates algorithms: recent developments in read alignment. **Genome Biology**, v. 22, n. 1, p. 249, 26 ago. 2021.
- AMATO, P. et al. Metatranscriptomic exploration of microbial functioning in clouds. **Scientific Reports**, v. 9, n. 1, p. 4383, 13 mar. 2019.
- AN, S.; SIN, H. H.; DUBOW, M. S. Modification of atmospheric sand-associated bacterial communities during Asian sandstorms in China and South Korea. **Heredity**, v. 114, n. 5, p. 460–467, 2015.
- ANDERMANN, T. et al. Estimating Alpha, Beta, and Gamma Diversity Through Deep Learning. **Frontiers in Plant Science**, v. 13, p. 839407, 19 abr. 2022.
- ANDREWS S. **FastQC A Quality Control tool for High Throughput Sequence Data**. Disponível em: <<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>>. Acesso em: 29 jun. 2023.
- ATA, B. et al. The Endobiota Study: Comparison of Vaginal, Cervical and Gut Microbiota Between Women with Stage 3/4 Endometriosis and Healthy Controls. **Scientific Reports**, v. 9, n. 1, p. 2204, dez. 2019.
- AUSTIN, G. I. et al. Contamination source modeling with SCRuB improves cancer phenotype prediction from microbiome data. **Nature Biotechnology**, p. 1–9, 16 mar. 2023.
- BLATTNER, F. R. et al. The complete genome sequence of Escherichia coli K-12. **Science (New York, N.Y.)**, v. 277, n. 5331, p. 1453–1462, 5 set. 1997.
- BLEVINS, S. M.; BRONZE, M. S. Robert Koch and the ‘golden age’ of bacteriology. **International Journal of Infectious Diseases**, v. 14, n. 9, p. e744–e751, 1 set. 2010.
- BODILIS, Josselin et al. Variable copy number, intra-genomic heterogeneities and lateral transfers of the 16S rRNA gene in Pseudomonas. **PloS one**, v. 7, n. 4, p. e35647, 2012.
- BONNET, M. et al. Bacterial culture through selective and non-selective conditions: the evolution of culture media in clinical microbiology. **New Microbes and New Infections**, v. 34, p. 100622, 30 nov. 2019.
- BRAY, J. R.; CURTIS, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. **Ecological Monographs**, v. 27, n. 4, p. 325–349, out. 1957.
- 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, dez. 2017.
- CANTALUPO, P. G.; PIPAS, J. M. Detecting viral sequences in NGS data. **Current Opinion in Virology**, v. 39, p. 41–48, 2019.
- CENTELLEGHE, C. et al. The use of Unmanned Aerial Vehicles (UAVs) to sample the blow microbiome of small cetaceans. **PLoS ONE**, v. 15, n. 7 July, p. 1–14, 2020.
- CHAKRAVORTY, S. et al. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. **Journal of Microbiological Methods**, v. 69, n. 2, p. 330–339, maio 2007.
- 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.
- CHEN, K.; PACHTER, L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. **PLoS computational biology**, v. 1, n. 2, p. 106–112, jul. 2005.
- 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.
- COTTRELL, M. T. et al. Bacterial diversity of metagenomic and PCR libraries from the Delaware River. **Environmental Microbiology**, v. 7, n. 12, p. 1883–1895, dez. 2005.

- DAVIS, N. M. et al. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. **Microbiome**, v. 6, n. 1, p. 226, 17 dez. 2018.
- DE CASTRO, O. et al. Sanger and next generation sequencing in the characterisation of arbuscular mycorrhizal fungi (AMF) in *Pancreaticum maritimum* L. (Amaryllidaceae), a representative plant species of Mediterranean sand dunes. **Planta**, v. 248, n. 6, p. 1443–1453, 2018.
- DJI. Inspire 2. 2016. Imagem. Disponível em: <https://www.dji.com/be/inspire-2>. Acesso em: 26 jun. 2023.
- DRONE STORE. SKYDIO 2. 2019. Imagem. Disponível em: <https://www.drone-store.fr/articles/skydio2>. Acesso em: 24 maio 2023.
- EDGAR, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. **Nature Methods**, v. 10, n. 10, p. 996–998, out. 2013.
- ERCOLINI, D. High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. **Applied and Environmental Microbiology**, v. 79, n. 10, p. 3148–3155, maio 2013.
- FANG, Z. et al. Culturable airborne bacteria in outdoor environments in Beijing, China. **Microbial Ecology**, v. 54, n. 3, p. 487–496, 2007.
- FEDRIZZI, T. et al. Genomic characterization of Nontuberculous Mycobacteria. **Scientific Reports**, v. 7, n. February, p. 1–14, 2017.
- 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.
- FIJAN, S. Microorganisms with claimed probiotic properties: An overview of recent literature. **International Journal of Environmental Research and Public Health**, v. 11, n. 5, p. 4745–4767, 2014.
- FLEISCHMANN, R. D. et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. **Science (New York, N.Y.)**, v. 269, n. 5223, p. 496–512, 28 jul. 1995.
- GAO, B. et al. An Introduction to Next Generation Sequencing Bioinformatic Analysis in Gut Microbiome Studies. **Biomolecules**, v. 11, n. 4, p. 530, 2 abr. 2021.
- GARCÍA-LÓPEZ, R. et al. Doing More with Less: A Comparison of 16S Hypervariable Regions in Search of Defining the Shrimp Microbiota. **Microorganisms**, v. 8, n. 1, p. 134, 17 jan. 2020.
- GILL, A. The Importance of Bacterial Culture to Food Microbiology in the Age of Genomics. **Frontiers in Microbiology**, v. 8, 2017.
- GOFFEAU, A. et al. The Yeast Genome Directory. **Nature**, v. 387, n. S6632, p. 5–5, 29 maio 1997.
- 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. 333–351, jun. 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 of the United States of America**, v. 116, n. 46, p. 23299–23308, 2019a.
- 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, 12 nov. 2019b.
- HANDELSMAN, J. Metagenomics: Application of Genomics to Uncultured Microorganisms. **Microbiology and Molecular Biology Reviews**, v. 68, n. 4, p. 669–685, dez. 2004.
- HARNPICHARNCHAI, P. et al. AirDNA sampler: An efficient and simple device enabling high-yield, high-quality airborne environment DNA for metagenomic applications. **PLoS One**, v. 18, n. 6, p. e0287567, 2023.

- HEYDARI, M. et al. Illumina error correction near highly repetitive DNA regions improves de novo genome assembly. **BMC Bioinformatics**, v. 20, p. 298, 3 jun. 2019.
- HORRICKS, R. A. et al. Comparison of drone and vessel-based collection of microbiological water samples in marine environments. **Environmental Monitoring and Assessment**, v. 194, n. 6, p. 439, 2022.
- HU, T. et al. Next-generation sequencing technologies: An overview. **Human Immunology**, v. 82, n. 11, p. 801–811, nov. 2021.
- HUESTIS, D. L. et al. Windborne long-distance migration of malaria mosquitoes in the Sahel. **Nature**, v. 574, n. 7778, p. 404–408, 2019.
- 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: 3 jul. 2023.
- JAVDAN, B. et al. Personalized Mapping of Drug Metabolism by the Human Gut Microbiome. **Cell**, v. 181, n. 7, p. 1661- 1679.e22, 2020.
- JIANG, H. et al. Altered fecal microbiota composition in patients with major depressive disorder. **Brain, Behavior, and Immunity**, v. 48, n. April, p. 186–194, 2015.
- KLINDWORTH, A. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. **Nucleic Acids Research**, v. 41, n. 1, p. e1, 7 jan. 2013.
- KOH, L. P.; WICH, S. A. Dawn of drone ecology: Low-cost autonomous aerial vehicles for conservation. **Tropical Conservation Science**, v. 5, n. 2, p. 121–132, 2012.
- KONOPIŃSKI, M. K. Shannon diversity index: a call to replace the original Shannon’s formula with unbiased estimator in the population genetics studies. **PeerJ**, v. 8, p. e9391, 29 jun. 2020.
- LEE, Michael. Amplicon and metagenomics overview. 2023. Imagem. Disponível em: https://astrobiomike.github.io/misc/amplicon_and_metagen. Acesso em: 21 jun. 2022.
- LEVY, S. E.; BOONE, B. E. Next-Generation Sequencing Strategies. **Cold Spring Harbor Perspectives in Medicine**, v. 9, n. 7, p. a025791, jul. 2019.
- LI, N. et al. The commensal microbiota and viral infection: A comprehensive review. **Frontiers in Immunology**, v. 10, n. JULY, p. 1–16, 2019.
- LI, R. et al. Metagenomic Analysis Exploring Taxonomic and Functional Diversity of Soil Microbial Communities in Sugarcane Fields Applied with Organic Fertilizer. **BioMed Research International**, v. 2020, p. 1–11, 21 out. 2020.
- LIGHTHART, B. The ecology of bacteria in the alfresco atmosphere. **FEMS Microbiology Ecology**, v. 23, n. 4, p. 263–274, 17 jan. 2006.
- LDLC. Parrot MiniDrone Mambo. 2017. Imagem. Disponível em: <https://www.ldlc.com/fr-be/fiche/PB00215403.html>. Acesso em: 24 maio 2023.
- LOZUPONE, C. et al. UniFrac: an effective distance metric for microbial community comparison. **The ISME Journal**, v. 5, n. 2, p. 169–172, fev. 2011.
- LU, H.; GIORDANO, F.; NING, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. **Genomics, Proteomics & Bioinformatics**, v. 14, n. 5, p. 265–279, out. 2016.
- MADIGAN, M. T. et al. **Brock biology of microorganisms**. Fifteenth edition ed. NY, NY: Pearson, 2018.
- MANCINELLI, R. L.; SHULLS, W. A. Airborne bacteria in an urban environment. **Applied and Environmental Microbiology**, v. 35, n. 6, p. 1095–1101, 1978.
- MARCO, M. L. et al. Health benefits of fermented foods: microbiota and beyond. **Current Opinion in Biotechnology**, v. 44, p. 94–102, 2017.
- MARDIS, E. R. Next-Generation Sequencing Platforms. **Annual Review of Analytical Chemistry**, v. 6, n. 1, p. 287–303, 2013.

- MARON, P. A. et al. Assessing genetic structure and diversity of airborne bacterial communities by DNA fingerprinting and 16S rDNA clone library. **Atmospheric Environment**, v. 39, n. 20, p. 3687–3695, 2005.
- MCCANN, J. A.; KEITH, D. A.; KINGSFORD, R. T. Measuring plant biomass remotely using drones in arid landscapes. **Ecology and Evolution**, v. 12, n. 5, p. e8891, 13 maio 2022.
- 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.
- MCMURDIE, P. J.; HOLMES, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. **PLOS Computational Biology**, v. 10, n. 4, p. e1003531, 3 abr. 2014.
- MODELISMOBH. Drone DJI Agras T20P. 2021. Imagem. Disponível em: <https://www.modelismobh.com.br/drone-dji-agras-t20p-kit-basico>. Acesso em: 24 maio 2023.
- MURALI, A.; BHARGAVA, A.; WRIGHT, E. S. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. **Microbiome**, v. 6, n. 1, p. 140, dez. 2018.
- NATIONAL HUMAN GENOME RESEARCH INSTITUTE (USA). **The Cost of Sequencing a Human Genome**. 1 nov. 2021. Imagem. Disponível em: <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>. Acesso em: 15 ago. 2023.
- NEAL, A. L. et al. Soil as an extended composite phenotype of the microbial metagenome. **Scientific Reports**, v. 10, n. 1, p. 10649, 30 jun. 2020.
- NEZAMI, S. et al. Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks. **Remote Sensing**, v. 12, n. 7, 2020.
- NGUYEN, N.-P. et al. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. **npj Biofilms and Microbiomes**, v. 2, n. 1, p. 1–8, 20 abr. 2016.
- NGUYEN, T. M. et al. Effective Soil Extraction Method for Cultivating Previously Uncultured Soil Bacteria. **Applied and Environmental Microbiology**, v. 84, n. 24, p. e01145-18, 30 nov. 2018.
- NOGUEIRA, J. et al. Dust arriving in the Amazon basin over the past 7,500 years came from diverse sources. **Communications Earth & Environment**, v. 2, n. 1, p. 1–11, 4 jan. 2021.
- OH, H. N. et al. Antarctic tundra soil metagenome as useful natural resources of cold-active lignocellulolytic enzymes. **Journal of Microbiology (Seoul, Korea)**, v. 57, n. 10, p. 865–873, out. 2019.
- OLSEN, G. J. et al. Microbial ecology and evolution: a ribosomal RNA approach. **Annual Review of Microbiology**, v. 40, p. 337–365, 1986.
- PARKER, A. et al. Review of Field Sampling Technologies for Characterizing Bioaerosols in Compact Spaces. 2020.
- PASTEUR, L. Remarks on Anthracic Vaccination as a Prophylactic of Splenic Fever. **BMJ**, v. 1, n. 1110, p. 489–489, 8 abr. 1882.
- PASTEUR, M. An Address on Vaccination in Relation to Chicken Cholera and Splenic Fever. **BMJ**, v. 2, n. 1076, p. 283–284, 13 ago. 1881.
- PEI, A. Y. et al. Diversity of 16S rRNA Genes within Individual Prokaryotic Genomes. **Applied and Environmental Microbiology**, v. 76, n. 12, p. 3886–3897, jun. 2010.
- PLUMMER, E. et al. A Comparison of Three Bioinformatics Pipelines for the Analysis of Preterm Gut Microbiota using 16S rRNA Gene Sequencing Data. **Journal of Proteomics & Bioinformatics**, v. 8, n. 12, 2015.
- PRODAN, A. et al. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. **PLOS ONE**, v. 15, n. 1, p. e0227434, 16 jan. 2020.
- PROSPERO, J. M.; GLACCUM, R. A.; NEES, R. T. Atmospheric transport of soil dust from Africa to South America. **Nature**, v. 289, n. 5798, p. 570–572, fev. 1981.

- PUSFITASARI, E. D. et al. Aerial drone furnished with miniaturized versatile air sampling systems for selective collection of nitrogen containing compounds in boreal forest. **Science of The Total Environment**, v. 808, p. 152011, 20 fev. 2022.
- QUAST, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. **Nucleic Acids Research**, v. 41, n. D1, p. D590–D596, 1 jan. 2013.
- REDDY, B.; DUBEY, S. K. Exploring the allochthonous pollution influence on bacterial community and co-occurrence dynamics of River Ganga water through 16S rRNA-tagged amplicon metagenome. **Environmental Science and Pollution Research International**, v. 28, n. 21, p. 26990–27005, jun. 2021.
- REESE, A. T.; DUNN, R. R. Drivers of Microbiome Biodiversity: A Review of General Rules, Feces, and Ignorance. **mBio**, v. 9, n. 4, p. 10.1128/mbio.01294-18, 31 jul. 2018.
- RHOADS, A.; AU, K. F. PacBio Sequencing and Its Applications. **Genomics, Proteomics & Bioinformatics**, v. 13, n. 5, p. 278–289, out. 2015.
- RIESENFELD, C. S.; SCHLOSS, P. D.; HANDELSMAN, J. Metagenomics: genomic analysis of microbial communities. **Annual Review of Genetics**, v. 38, p. 525–552, 2004.
- SALEEM, F. et al. Metagenomic Characterization of Bacterial Communities in Drinking Water Supply System of a Mega City. **Microbial Ecology**, v. 76, n. 4, p. 899–910, nov. 2018.
- SANGER, F. et al. Nucleotide sequence of bacteriophage phi X174 DNA. **Nature**, v. 265, n. 5596, p. 687–695, 24 fev. 1977.
- SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences of the United States of America**, v. 74, n. 12, p. 5463–5467, dez. 1977.
- SCHLOSS, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. **Applied and Environmental Microbiology**, v. 75, n. 23, p. 7537–7541, dez. 2009.
- SCHRADER, C. et al. PCR inhibitors – occurrence, properties and removal. **Journal of Applied Microbiology**, v. 113, n. 5, p. 1014–1026, 2012.
- SHANNON, C. E.; WEAVER, W. **The mathematical theory of communication**. 21. print ed. Urbana: Univ. of Illinois Press, 1998.
- SHUTTLEWORTH, L.; OOSTHUIZEN, C. J. Comparing DNA yield from fish scales following different extraction protocols. **Scientific Reports**, v. 12, p. 2836, 18 fev. 2022.
- SIERRA, M. A. et al. The Influences of Bioinformatics Tools and Reference Databases in Analyzing the Human Oral Microbial Community. **Genes**, v. 11, n. 8, p. 878, 3 ago. 2020.
- 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.
- STEWART, E. J. Growing Unculturable Bacteria. **Journal of Bacteriology**, v. 194, n. 16, p. 4151–4160, 27 jul. 2012.
- TAN, K. S. et al. Detection of hospital environmental contamination during SARS-CoV-2 Omicron predominance using a highly sensitive air sampling device. **Frontiers in Public Health**, v. 10, p. 1067575, 2022.
- TAYLOR-BROWN, A. et al. Culture-independent metagenomics supports discovery of uncultivable bacteria within the genus Chlamydia. **Scientific Reports**, v. 7, p. 10661, 6 set. 2017.
- THE INTEGRATIVE HMP (IHMP) RESEARCH NETWORK CONSORTIUM. The Integrative Human Microbiome Project. **Nature**, v. 569, n. 7758, p. 641–648, maio 2019.
- THOMPSON, L. R. et al. A communal catalogue reveals Earth’s multiscale microbial diversity. **Nature**, v. 551, n. 7681, p. 457–463, 23 nov. 2017.

- TONI, L. S. et al. Optimization of phenol-chloroform RNA extraction. **MethodsX**, v. 5, p. 599–608, 30 maio 2018.
- URUBURU, F. History and services of culture collections. **International Microbiology**, v. 6, n. 2, p. 101–103, 1 jun. 2003.
- VAN DIJK, E. L. et al. The Third Revolution in Sequencing Technology. **Trends in genetics: TIG**, v. 34, n. 9, p. 666–681, set. 2018.
- VARTOUKIAN, S. R.; PALMER, R. M.; WADE, W. G. Strategies for culture of ‘unculturable’ bacteria: Culturing the unculturable. **FEMS Microbiology Letters**, p. no-no, 27 abr. 2010.
- WADE, W. Unculturable bacteria—the uncharacterized organisms that cause oral infections. v. 95, p. 3, 2002.
- WAGNER, R.; GASSEN, H. G. Identification of a 16S rna sequence located in the decoding site of 30S ribosomes. **FEBS letters**, v. 67, n. 3, p. 312–315, 1 set. 1976.
- WHAT Is 16s rRNA sequencing?. [S. l.]: CD Genomics, 2021. Disponível em: https://www.youtube.com/watch?v=_IpZOqFP9iY. Acesso em: 17 fev. 2023.
- WILLIS, A. D. Rarefaction, Alpha Diversity, and Statistics. **Frontiers in Microbiology**, v. 10, p. 2407, 23 out. 2019.