



Instituto de Biologia do Exército

Divisão de Ensino e Pesquisa

Eduardo Nascimento de Azevedo

**CONSTRUÇÃO DE BANCO DE DADOS LOCAL DA REGIÃO GÊNICA 16S rRNA COM
APLICAÇÃO NA DEFESA BIOLÓGICA**

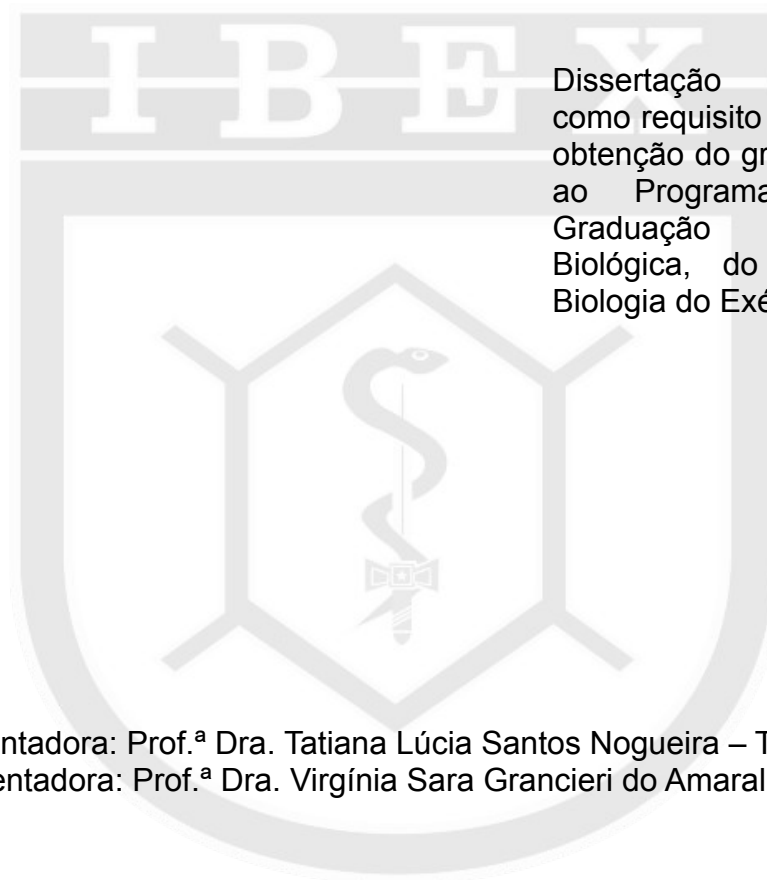
RIO DE JANEIRO

2023

EDUARDO NASCIMENTO DE AZEVEDO – Maj

CONSTRUÇÃO DE BANCO DE DADOS LOCAL DA REGIÃO GÊNICA 16S

rRNA COM APLICAÇÃO NA DEFESA BIOLÓGICA



Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre, ao Programa de Pós-Graduação em Defesa Biológica, do Instituto de Biologia do Exército.

Orientadora: Prof.^a Dra. Tatiana Lúcia Santos Nogueira – Ten Cel
Coorientadora: Prof.^a Dra. Virgínia Sara Grancieri do Amaral – 1º Ten

RIO DE JANEIRO

2023

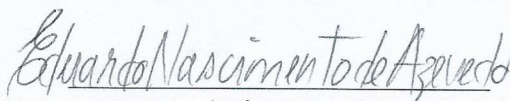
CATALOGAÇÃO NA FONTE
INSTITUTO DE BIOLOGIA DO EXÉRCITO

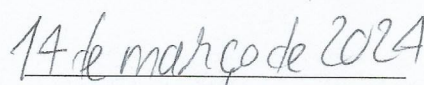
A994 Azevedo, Eduardo Nascimento de.
Construção de Banco de Dados Local da Região Gênica 16S rRNA com
Aplicação na Defesa Biológica. / Eduardo Nascimento de Azevedo. – Rio
de Janeiro, 2023.
77 f.
Orientador (a): Prof.^a Dra. Tatiana Lúcia Santos Nogueira – Ten Cel
Coorientadores: Prof.^a Dra. Virgínia Sara Grancieri do Amaral – 1º
Tenente
Dissertação (Mestrado) – Instituto de Biologia do Exército, Divisão de
Ensino e Pesquisa, Pós-Graduação em Defesa Biológica, 2023.
Referências: f. 68–77.

1. DEFESA BIOLÓGICA. 2. BIOTERRORISMO. 3.
BIOINFORMÁTICA. I. Prof.^a Dra. Tatiana Lúcia Santos Nogueira – Ten
Cel (Orientadora). II. Instituto de Biologia do Exército. III. Título.

CDD 355.232

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial deste trabalho.


Assinatura


Data

EDUARDO NASCIMENTO DE AZEVEDO

EDUARDO NASCIMENTO DE AZEVEDO – Maj

**CONSTRUÇÃO DE BANCO DE DADOS LOCAL DA REGIÃO GÊNICA 16S
rRNA COM APLICAÇÃO NA DEFESA BIOLÓGICA**

Dissertação apresentada
como requisito parcial para a
obtenção do grau de Mestre,
ao Programa de Pós-
Graduação em Defesa
Biológica, do Instituto de
Biologia do Exército.

Data de Aprovação: ____ / ____ / ____

Banca Examinadora:

Prof.^a Dra. Tatiana Lúcia Santos Nogueira – Ten Cel
Instituto de Biologia do Exército – IBEx
Orientadora

Prof. Dr. Marcos Dornelas Ribeiro – Ten Cel
Instituto de Biologia do Exército – IBEx
Membro

Prof. Dr. César Rogério Leal do Amaral
Universidade do Estado do Rio de Janeiro – UERJ
Membro

A vida vem do éter que se
condensa,
Mas o que mais no Cosmos me
entusiasma
É a esfera microscópica do
plasma
Fazer a luz do cérebro que pensa
(Augusto dos Anjos)

RESUMO

Bioterrorismo é o uso deliberado de agentes biológicos, isto é, microrganismos patogênicos, contra populações civis ou militares para promover causas políticas, ideológicas, ou religiosas. O emprego desses meios pode resultar no colapso dos serviços de saúde pública, crise econômica e instabilidade social. As medidas necessárias à contenção dependem da capacidade de detectar e identificar o agente biológico rapidamente. O advento do sequenciamento massivo-paralelo por nanoporos introduziu sequenciadores portáteis com grande capacidade de multiplexação que possibilitam a identificação, independente de culturas microbianas e sem a necessidade de prévio isolamento, de agentes de bioterrorismo em campo ou em laboratórios móveis. O processamento de bioinformática dos dados do sequenciamento se baseia na comparação das *reads* geradas pelo sequenciamento com sequências de alta qualidade (referência) depositadas em bancos de dados públicos, acessíveis pela internet. O emprego de bancos de dados locais, isto é, armazenados em um computador portátil ou na intranet do laboratório sede, permite que os dados sejam analisados sem a necessidade de uma conexão de internet, recurso muitas vezes indisponível em pesquisas de campo ou nas localidades remotas do território nacional, e ainda garante a segurança da informação, em caso de investigações sigilosas, nas quais não seria desejável transmitir dados para terceiros. O objetivo do presente trabalho foi construir um banco de dados local de sequências do gene 16S rRNA de agentes de bioterrorismo (16S rRNA bt) para análises de dados de sequenciamento massivo-paralelo por nanoporos, para fins de identificação bacteriana em ações de defesa biológica, aplicável em laboratório móvel. Para isso foram obtidas sequências de referência dos genes 16S rRNA do banco de dados público 16S RefSeq, do *National Center for Biotechnology Information* (NCBI) das espécies listadas como agentes de bioterrorismo pelo *Centers for Disease Control and Prevention* (CDC), foram sequenciadas amostras das cepas de *Salmonella enterica* ATCC 14028 (agente de bioterrorismo categoria B) e *Proteus vulgaris* ATCC 8427 (cepa controle) por sequenciamento massivo-paralelo por nanoporos, e os dados foram analisados com os bancos de dados locais (16S RefSeq e 16S rRNA bt) e remotamente com o *software* EPI2ME, nos servidores da Nanopore. O banco de dados local 16S rRNA bt com 95 sequências de referência, resultou em 146,9 kB de tamanho, e os alinhamentos realizados com as *reads* geradas a partir das amostras de *S. enterica* e *P. vulgaris* com este banco de dados, duraram $57,02 \pm 0,08$ e $50,99 \pm 0,33$ segundos, respectivamente, enquanto os realizados com o banco de dados 16S RefSeq tiveram duração de $910,47 \pm 5,02$ e $1171,58 \pm 4,00$ segundos, respectivamente. A análise remota com o EPI2ME foi concluída em 06 horas e 04 minutos. Foi possível identificar corretamente *S. enterica* e *P. vulgaris*, tanto pela análise remota, quanto pelas análises *offline* com os bancos de dados locais. Comparado com o banco de dados 16S RefSeq, os alinhamentos com o 16S rRNA bt foram de 16 a 23 vezes mais rápidos. A análise com os bancos de dados locais foi significativamente mais rápida que a realizada no servidor da Nanopore.

Palavras-chave: banco de dados, *Salmonella enterica*, bioinformática, bioterrorismo, biodefesa, sequenciamento massivo-paralelo

ABSTRACT

Bioterrorism is defined as the use of biological agents against civilian population or military personal in order to further political, ideological or religious objectives. Their use may result in public health services being overloaded, economical crisis and social unrest. Containment measures depend upon possessing the capabilities for rapid detection and identification of these agents. The advent of nanopore based massive-parallel sequencing (MPS) introduced portable sequencers devices with great multiplexing capacity, allowing for culture independent identification, with no *a priori* isolation, of bioterrorism agents in the field or in mobile laboratories. The bioinformatics processing of sequencing data is based on the comparison of reads generated by sequencing with the high quality reference sequences deposited in public databases, using internet access. The use of local databases stored in a laptop computer or in the laboratory intranet allows that data analysis is performed without an internet connection, that often is unavailable in field research or remote locations, while simultaneously maintaining the data secure, a desirable feature for sensitive investigations, when submitting data for a third part is precluded. In this work we built a local database with reference sequences of the 16S rRNA gene of bacteria listed by the CDC as bioterrorism agents (16S rRNA bt) for data analysis of nanopore based MPS data as means for bacterial identification in mobile laboratories of biodefense. To accomplish this, reference sequences were obtained from NCBI's 16S RefSeq database, and culture collection strains of *Salmonella enterica* ATCC 14028 (as bioterrorism agent) and *Proteus vulgaris* ATCC 8427 (as control) were sequenced, and sequencing data was analysed using local databases (16S RefSeq and 16S rRNA bt) and remotely using cloud based software EPI2ME from Nanopore. Resulting database contained 95 sequences and amounted for 146,9 kB. Alignment of reads generated from sequencing of *S. enterica* and *P. vulgaris* samples with 16S rRNA bt were made in $57,02 \pm 0,08$ e $50,99 \pm 0,33$ seconds, respectively, while they took, respectively, $910,47 \pm 5,02$ e $1171,58 \pm 4,00$ seconds with 16S RefSeq. Remote analysis with EPI2ME concluded in 06 hours and 04 minutes. All analysis pathways identified correctly the two species. Compared to 16S RefSeq, alignments using 16S rRNA bt were 16 to 23 times faster. The use of local databases significantly sped up sequencing data analysis compared with cloud based option.

Keywords: database, *Salmonella enterica*, bioinformatics, bioterrorism, biodefense, massiv-parallel sequencing

LISTA DE FIGURAS

Figura 1: Processo de descontaminação do <i>Hart Senate Office Building</i> (Washington D.C., EUA), pela <i>Environmental Protection Agency</i> (EPA).....	15
Figura 2: Montagem ilustrativa de uma matéria da revista online Vice, sobre o ataque bioterrorista perpetrado por membros do culto Rajneeshee, no Oregon (EUA), no ano de 1984.....	17
Figura 3: Abertura da carta com esporos de antraz, endereçada ao senador estadunidense Patrick Leahy, no laboratório do forte Detrick (Maryland, EUA).....	17
Figura 4: Representação do gene 16S rRNA e da cobertura dos pares de <i>primers</i> proposta por Fukuda e colaboradores (2016). No desenho, em azul estão representadas as regiões conservadas, e em vermelho, as hipervariáveis.....	27
Figura 5: Gráfico da classificação das <i>reads</i> por espécie.....	51
Figura 6: Histograma da precisão média dos alinhamentos.....	52
Figura 7: Histograma do tamanho das <i>reads</i>	53
Figura 8: Histograma de qualidade das <i>reads</i>	53
Figura 9: Detalhe da árvore filogenética construída com o software Mega. As sequências consenso barcode02 e barcode03 foram agrupadas junto dos clados de <i>S. enterica</i> e <i>P. vulgaris</i> , respectivamente.....	54
Figura 10: Árvore filogenética completa, exibida em formato circular, com as sequências dos genes 16S rRNA de todas as 95 espécies de bactérias reconhecidas com agentes de bioterrorismo, que foram utilizadas na construção do banco de dados 16S rRNA bt. Em verde foram destacadas as sequências do <i>barcode</i> 03 (gerada neste trabalho) e a referência de <i>P. vulgaris</i> (NCBI), e em vermelho as do <i>barcode</i> 02 e referências de <i>S. enterica</i>	56

LISTA DE TABELAS

Tabela 1: Resumo da Classificação de Risco dos Agentes Biológicos usada no Brasil20
Tabela 2: Resumo das Categorias de Agentes de Bioterrorismo, segundo a classificação do CDC.....	.21
Tabela 3: Sequências dos <i>primers</i> utilizados nos estudos do 16S rRNA e respectiva cobertura relativa, no banco de dados <i>Ribosomal Database Project (RDP)</i>28
Tabela 4: Primers alternativos desenhados para modificar a cobertura da reação de amplificação do gene 16S rRNA.....	.28
Tabela 5: Laboratórios civis brasileiros de referência para o diagnóstico dos agentes de bioterrorismo.....	.37
Tabela 6: Relação dos laboratórios públicos de alta contenção, por unidade da federação.....	.38
Tabela 7: Programa do termociclador para a amplificação por PCR do gene 16S rRNA46
Tabela 8: Parâmetros para a corrida de sequenciamento inseridos no software MinKNOW48
Tabela 9: Parâmetros para a análise de dados inseridos no software MinKNOW49
Tabela 10: Sequências que compõem o banco de dados local 16S rRNA de bactérias listadas como agentes de bioterrorismo. Os descritores são o número de acesso NCBI, TaxID, nome científico e tamanho.....	.57
Tabela 11: Comparação entre o tempo necessário para o alinhamento das reads usando o Minimap2 com o banco de dados (16S rRNA bt) e o banco de dados 16S RefSeq do NCBI. Desvio padrão calculado usando-se uma planilha do LibreOffice Calc v.7.0.4.2 para Linux.....	.61

LISTA DE ABREVIÇÕES E SIGLAS

ATCC	Coleção de cultura tipo americana (<i>American Type Culture Collection</i>)
BHI	Infusão cérebro-coração (<i>Brain-Heart Infusion</i>)
CDC	Centro para Prevenção e Controle de Doenças (<i>Centers for Disease Control and Prevention</i>)
DNA	Ácido desoxirribonucleico (<i>DesoxyriboNucleic Acid</i>)
DQBRN	Defesa Química, Biológica, Radiológica e Nuclear
FB	Tampão de descarga (<i>Flush Buffer</i>)
FLT	Guia de descarga (<i>Flush Tether</i>)
IBEx	Instituto de Biologia do Exército
LB	Microesferas de carregamento (<i>Loading Beads</i>)
LFIA	Imunoensaio de fluxo lateral (<i>LateralFlow ImmunoAssay</i>)
MALDI-TOF/MS	Espectrometria de massas com dessorção/ionização laser, assistida por matriz (<i>Matriz-assisted laser desorption/ionization—time-of-flight mass spectrometry</i>)
MPS	Sequenciamento massivo-paralelo (<i>Massive-Parallel Sequencing</i>)
NCBI	<i>National Center for Biotechnology Information</i>
NGS	Sequenciamento de nova geração (<i>Next-Generation Sequencing</i>)
ONT	Oxford Nanopore Technologies
OTU	Unidade Operacional Taxonômica (<i>Operational Taxonomic Unit</i>)
PCR	Reação em cadeia da polimerase (<i>Polymerase Chain Reaction</i>)
rRNA	Ácido ribonucleico ribossomal (<i>ribosomal RiboNucleic Acid</i>)
SQB	Tampão de sequenciamento (<i>Sequencing Buffer</i>)
WHO	Organização Mundial de Saúde (<i>World Health Organization</i>)

SUMÁRIO

INTRODUÇÃO.....	12
1 REFERENCIAL TEÓRICO.....	19
1.1 Classificação dos Microrganismos de Interesse em Bioterrorismo.....	19
1.2 Métodos de Detecção e Identificação de Agentes Microbianos no Campo da Defesa Biológica.....	22
1.2.1 Métodos microbiológicos.....	22
1.2.2 Métodos imunocromatográficos.....	23
1.2.3 Métodos moleculares.....	24
1.3 O gene 16S rRNA e sua Aplicação na Identificação Bacteriana.....	26
1.3.1 Estrutura do gene 16S rRNA.....	26
1.3.2 Classificação taxonômica baseada no gene 16S rRNA.....	29
1.3.3 Ferramentas de bioinformática aplicadas no contexto da identificação microbiana.....	30
1.4 Organizações Nacionais e Internacionais de Enfrentamento aos Agentes Biológicos.....	33
1.5 Atuação do Exército Brasileiro na Defesa Biológica.....	40
2 OBJETIVOS.....	42
2.1 Objetivo Geral.....	42
2.2 Objetivos Específicos.....	42
3 MATERIAIS E MÉTODOS.....	43
3.1 Construção de Bancos de Dados 16S rRNA de Agentes de Bioterrorismo.....	43
3.2 Cultivo Bacteriano e Extração do DNA genômico.....	43
3.3 Quantificação de DNA.....	44
3.4 Preparação da Biblioteca e Corrida de Sequenciamento.....	45
3.5 Análise dos Dados do Sequenciamento.....	49
3.5.1 <i>Basecall</i> , demultiplexação e <i>trimming</i>	49
3.5.2 Mapeamento e alinhamento com o software EPI2ME.....	49
3.5.3 Mapeamento e alinhamento com o banco de dados local.....	50

3.5.4 Construção da árvore filogenética.....	50
3.5.5 Comparação das análises com banco de dados local e remota.....	50
4 RESULTADOS.....	51
4.1 Sequenciamento do Gene 16S rRNA e Análise dos Dados no Sistema ONT.....	51
4.2 Comparação das Análises dos Dados de Sequenciamento.....	54
4.3 Construção da Arvore Filogenética.....	54
4.4 Construção do Banco de Dados 16S rRNA dos Agentes de Bioterrorismo	57
5 DISCUSSÃO.....	62
CONCLUSÃO.....	67
REFERÊNCIAS.....	68

INTRODUÇÃO

Bioterrorismo e Seus Impactos na Sociedade

A pandemia de Covid-19 reacendeu debates e discussões em torno das vulnerabilidades na biossegurança dos sistemas de saúde e colocou um foco no potencial para a transformação de agentes biológicos em armamentos (Tin; Sabeti; Ciottone, 2022). Neste cenário, o conceito de biossegurança ganha destaque, tendo em vista que pode ser definido como um conjunto de ações voltadas para a prevenção, minimização ou eliminação dos riscos inerentes às atividades envolvendo microrganismo (Penna *et al.*, 2020). A pandemia do SARS-Cov-2 demonstrou os efeitos devastadores, da liberação não intencional, de um patógeno respiratório e mutável, na população mundial, provocando extremas morbidade, mortalidade e perdas econômicas (Long; Marzi, 2021).

A recente pandemia, não apenas, expôs deficiências significativas em biocontenção, triagem e vigilância de casos, mobilização de equipamentos de proteção individual (EPI), capacidade de resposta a surtos e contramedidas profiláticas, como, também, a inconsistência na implementação e aderência às intervenções não-farmacológicas, a hesitação e a desigualdade na vacinação global evidenciaram complexidades em liderança de crises e educação pública, que podem ser exploradas no planejamento de ações de bioterrorismo (Tin; Sabeti; Ciottone, 2022).

Apesar da Convenção para a Proibição de Armas Bacteriológicas (Biológicas) e Toxínicas (CPAB) em vigor, ratificada por 185 países (Biological Weapons Convention – UNODA, [s. d.]), proibindo o desenvolvimento, produção, estocagem, aquisição, conservação e transferência dessas armas (Rambauske; Cardoso; Navarro, 2014), estados párias e grupos terroristas, dificilmente, sentir-se-ão constrangidos por acordos internacionais, e é justificada a preocupação com o potencial do bioterrorismo para causar doença, morte e pânico desproporcionais aos meios empregados (Green *et al.*, 2019).

O bioterrorismo é definido como a liberação intencional de uma toxina ou microrganismo patogênico para causar morte ou dano a uma população, criação

animal ou plantação, tendo motivação ideológica (Jansen *et al.*, 2014). É importante fazer a distinção entre o bioterrorismo e outros atos de agressão, perpetrados com o emprego de agentes biológicos. A expressão guerra biológica refere-se a uma ação que envolve entes estatais, na qual os agentes biológicos são usados como armas, em um cenário de guerra (Oliveira *et al.*, 2020). Por fim, há o biocrime que é uma ação motivada por vingança ou ganho pessoal, contra uma pessoa ou um pequeno grupo de pessoas, usando agentes biológicos (Jansen *et al.*, 2014).

As percepções globais da biodefesa e bioproteção nos meios científicos e clínicos foi permanentemente alterada pelo ataque com esporos de antraz, nos Estados Unidos, em 2001 (Artenstein, 2017). Biodefesa ou defesa biológica compreende o conjunto de medidas desenvolvidas em resposta às ameaças de bioterrorismo ou de surtos epidêmicos de ocorrência natural, que inclui a vigilância, o diagnóstico, o tratamento, e a prevenção (Artenstein, 2017). Bioproteção aborda os princípios de proteção, controle e responsabilidades sobre materiais biológicos de alto risco, buscando evitar o acesso não autorizado, o extravio ou a liberação intencional (Brasil, 2019).

As armas biológicas oferecem vantagens significativas aos terroristas: são relativamente mais baratas que armamentos convencionais ou nucleares; podem ser lançadas de forma furtiva, devido ao período variável de latência clínica, facilitando a fuga dos perpetradores; o seu alcance pode ser amplificado pela transmissão pessoa a pessoa; provocam ansiedade e pânico na população, que podem ser desproporcionais aos seus efeitos materiais, em alguns casos (Artenstein, 2017).

Eventos de bioterrorismo são considerados de baixa probabilidade, mas de alto impacto (Rathjen; Shahbodaghi, 2021). Essas ações têm como objetivos produzir pânico, grande número de vítimas e interrupção social. O banco de dados *Global Terrorism Database* lista 37 ataques de bioterrorismo entre os anos de 1981 a 2018, em todo o mundo. (Global Terrorism Database, [s. d.]).

No caso das cartas contendo esporos de antraz, endereçadas a políticos e aos veículos de imprensa estadunidenses nas semanas seguintes aos ataques ao *World Trade Center*, além dos 22 casos da doença, com cinco fatalidades, pelo

menos 10.000 pessoas foram submetidas ao tratamento profilático com antibióticos, por 60 dias, e a descontaminação das instalações afetadas teve custos estimados em US\$ 320 milhões (Long; Marzi, 2021). Entre as instalações contaminadas havia três grandes centros de processamento e distribuição do *United States Postal Service* (USPS), a sede da Suprema Corte, escritórios do Senado e da Câmara de Representantes, dos EUA (Schmitt; Zacchia, 2012). O *Hart Senate Office Building* (Figura 1) permaneceu em quarentena por 96 dias, até a completa descontaminação. Em algumas semanas, o FBI recebeu 29.000 denúncias sobre pacotes suspeitos, e interceptou 7.000 cartas, supostamente contendo antraz, enquanto o USPS recebeu 19.800 denúncias de antraz (Jacobs, 2004). O orçamento dos EUA para a biodefesa, que em 2001 havia sido de US\$ 700 milhões, aumentou para US\$ 4 bilhões, em 2002, chegando a um teto de US\$ 8 bilhões, em 2005 (Long; Marzi, 2021).

A possibilidade do uso dual do conhecimento empregado em pesquisas legítimas para ações de bioterrorismo fez com que aumentasse a regulação para a pesquisa das doenças. O impacto dessas regulações mais rígidas, em termos de oportunidades perdidas de cooperação internacional, intercâmbio de patógenos e compartilhamento de novos agentes, não pode ser mensurado e é frequentemente ignorado (Green *et al.*, 2019).

A história do bioterrorismo e da guerra biológica está intrinsecamente ligada ao progresso científico e à pesquisa das doenças (Carus, 2017), muito antes da teoria dos germes, materiais contaminados eram lançados no território inimigo, evoluindo para a produção de culturas de microrganismos em larga escala no alvorecer da moderna microbiologia e, por fim, culminando com a produção de armamentos a partir de patógenos após a 1ª Guerra Mundial (Long; Marzi, 2021).



Figura 1: Processo de descontaminação do *Hart Senate Office Building* (Washington D.C., EUA), pela *Environmental Protection Agency* (EPA).
Fonte: GAO, 2003

Contexto Histórico do Bioterrorismo

Muitas são as dificuldades impostas ao estudo histórico do bioterrorismo e da guerra biológica: a ausência de dados científicos confiáveis; a dificuldade de se interpretar os documentos disponíveis; e a interpretação equivocada dos fatos históricos, sob uma perspectiva contemporânea (Barras; Greub, 2014).

O emprego de microrganismos patogênicos como arma não é um fenômeno novo: evidências arqueológicas sugerem o uso de pontas de flechas contaminadas com bactérias capazes de causar o tétano na pré-história. Relatos medievais descrevem corpos de vítimas da peste, catapultados sobre os muros de cidades

sitiadas, e correspondências coloniais narram a distribuição de cobertores usados por doentes de varíola aos nativos americanos (Carus, 2017).

Com o advento da microbiologia e a descoberta da relação causal entre microrganismos e doença, abriu-se a possibilidade de selecionar e produzir esses agentes, e durante as 1ª e 2ª Guerras Mundiais, vários países envolvidos mantiveram programas de desenvolvimento de armas biológicas até que a CPAB entrasse em vigor, em 1975. A Convenção, intitulada “*The Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction*” ou na forma curta *The Biological Weapons Convention* (BWC), é constituída de 15 artigos e foi o primeiro tratado de desarmamento multilateral a banir uma categoria inteira de armas de destruição em massa (Biological Weapons Convention – UNODA, [s. d.]).

A Convenção foi negociada na Conferência do Comitê de Desarmamento, em Genebra, Suíça, passando a vigorar a partir de 26 de março de 1975, suplementando a proibição do uso de armas biológicas, prevista nos Protocolos de Genebra, de 1925 (Biological Weapons Convention – UNODA, [s. d.]). A cada cinco anos, aproximadamente, são promovidos encontros entre os estados partícipes para rever a operação da Convenção. Desde 1980, nove Conferências de Revisão foram realizadas, enquanto no período entre Conferências, os estados partícipes promoveram diversas atividades e iniciativas visando o fortalecimento da efetividade e o aprimoramento da implementação da Convenção (Biological Weapons Convention – UNODA, [s. d.]).

Em 1984, integrantes de uma seita originada na cidade de *Poona*, na Índia, e liderado por *Bhagwan Shree* (Figura 2), que haviam estabelecido uma comunidade na vizinhança da pequena cidade de *Antelope* (*Oregon*, EUA), contaminaram os bufês de salada de vários estabelecimentos, no condado *The Dalles*, com *Salmonella*, para interferir nas eleições locais, vitimando 751 pessoas, sem fatalidades. Quase 20 anos depois, nas semanas que se seguiram aos ataques terroristas de 11 de setembro de 2001, cartas contendo esporos de *Bacillus anthracis* (Figura 3) foram enviadas para empresas de mídia, jornalistas e políticos, vitimando 22 pessoas, com cinco fatalidades (Oliveira *et al.*, 2020).



Figura 2: Montagem ilustrativa de uma matéria da revista online Vice, sobre o ataque bioterrorista perpetrado por membros do culto Rajneeshee, no Oregon (EUA), no ano de 1984.
Fonte: Sen, 2018

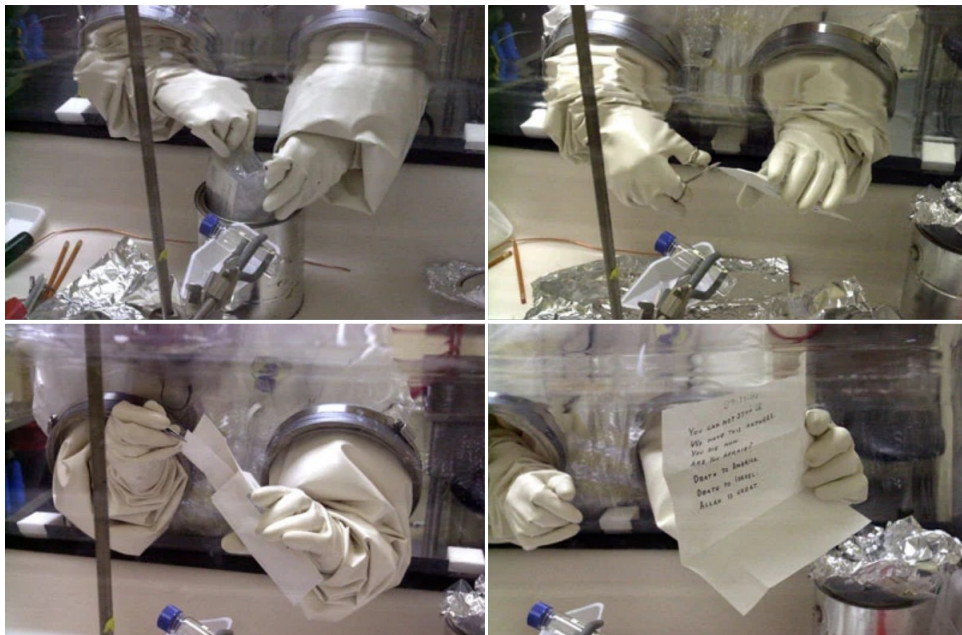


Figura 3: Abertura da carta com esporos de antraz, endereçada ao senador estadunidense Patrick Leahy, no laboratório do forte Detrick (Maryland, EUA).
Fonte: THE LEAHY LETTER, [s. d.]

Ainda que não se trate de um incidente de bioterrorismo, como os citados anteriormente, o vazamento acidental de esporos de antraz, ocorrido em 1979, na cidade de Sverdlovsk (atualmente Yekaterinburg, Rússia), um dos maiores surtos da história de antraz inalatório humano, com 66 mortes, foi considerado uma clara evidência de violação da CPAB (Sahl *et al.*, 2016). O incidente, provocado pela falha dos filtros de segurança, durante uma manutenção rotineira, resultou na liberação de uma nuvem de esporos, espalhados pelo vento, adoeceu pessoas, distantes 4 km, e animais, distantes 50 km, da origem. Inicialmente foi negado pelas autoridades soviéticas (Sahl *et al.*, 2016), e somente em 1992, após o fim da União Soviética, o presidente Bóris Yeltsin admitiu que se tratava de um laboratório militar (Oliveira *et al.*, 2020).

O trabalho com microrganismos patogênicos é uma atividade de risco e, somente com uma avaliação correta desses riscos é possível desenvolver pesquisas dentro de condições adequadas de segurança. Para orientar a implementação de medidas de biossegurança adequadas, foram elaboradas classificações de riscos, segundo a infectividade e virulência dos agentes biológicos (Penna *et al.*, 2020).

Justificativa

A análise dos dados é a etapa final em qualquer processo baseado em sequenciamento de DNA. Em geral, essas análises consistem em fazer o alinhamento das sequências, obtidas de um sequenciador, com aquelas depositadas em um banco de dados público, transferindo-se os arquivos locais para os servidores que hospedam os bancos de dados. Essa abordagem tem como inconvenientes o dispêndio de parte do tempo de análise com o *upload* de grandes volumes de dados por conexões de internet, que estão sujeitas à instabilidade, além de comprometer a segurança de informações sensíveis. Neste contexto, o presente trabalho se faz extremamente relevante, tanto no âmbito da biodefesa, quanto da cibersegurança, onde a rápida resposta e a proteção da informação precisam

caminhar em conjunto, além de representarem pontos cruciais na tomada de decisão diante de um ataque bioterrorista.

1 REFERENCIAL TEÓRICO

1.1 Classificação dos Microrganismos de Interesse em Bioterrorismo

No Brasil, a Comissão de Biossegurança em Saúde (CBS), do Ministério da Saúde, é o órgão responsável por instituir, designar e coordenar uma Comissão de Especialistas que revisa e atualiza, a cada dois anos, a Classificação de Risco de Agentes Biológicos (Tabela 1). Esta classificação divide os agentes em quatro classes de risco: a Classe de Risco 1 engloba os microrganismos que não causam doenças em pessoas saudáveis; a Classe de Risco 2, aqueles capazes de infectar pessoas ou animais, mas com baixo potencial para propagação na comunidade e disseminação no meio ambiente; a Classe de Risco 3, os microrganismos com capacidade de transmissão respiratória, que causam doenças potencialmente letais em humanos e animais, e transmissão de pessoa a pessoa, para os quais há medidas profiláticas e tratamento eficazes; e a Classe de Risco 4, os agentes que causam doenças graves, no homem ou nos animais, com alta transmissibilidade, para os quais não são conhecidas medidas profiláticas ou terapêuticas eficazes (Brasil, 2023). A classificação brasileira é bastante similar àquelas usadas pela Organização Mundial da Saúde (WHO, 2020) e pelo *National Institute of Health* (NIH, 2019).

Tabela 1: Resumo da Classificação de Risco dos Agentes Biológicos usada no Brasil Adaptado de: Brasil, 2023

Classificação	Avaliação de riscos	Características	Exemplos
Classe de risco 1	Baixo risco individual e para a comunidade	Não causam doenças no ser humano ou nos animais adultos saudios	<i>Lactobacillus spp.</i> e <i>Bacillus subtilis</i>

Classificação	Avaliação de riscos	Características	Exemplos
Classe de risco 2	Moderado risco individual e limitado risco para a comunidade	Provocam infecções no ser humano ou nos animais, cujo potencial de propagação na comunidade e de disseminação no meio ambiente é limitado, e para os quais existem medidas profiláticas e terapêuticas conhecidas eficazes	<i>Schistosoma mansoni</i> e vírus da rubéola
Classe de risco 3	Alto risco individual e moderado risco para a comunidade	Capacidade de transmissão, em especial por via respiratória, e que causam doenças potencialmente letais em humanos ou animais, e para as quais existem, usualmente, medidas profiláticas e terapêuticas. Os agentes biológicos representam risco se disseminados na comunidade e no meio ambiente, podendo se propagar de pessoa a pessoa.	<i>Bacillus anthracis</i> e Vírus da Imunodeficiência Humana (HIV)
Classe de risco 4	Alto risco individual e para a comunidade	Grande poder de transmissibilidade, em especial por via respiratória, ou de transmissão desconhecida. Até o momento, não há nenhuma medida profilática ou terapêutica eficaz contra infecções ocasionadas por esses agentes biológicos. Eles causam doenças de alta gravidade em humanos e animais, tendo grande capacidade de disseminação na comunidade e no meio ambiente	Vírus ebola e vírus da varíola

Outra classificação bem estabelecida e muito adotada é a do *Centers for Disease Control and Prevention* (CDC), dos EUA (CDC, 2019). Nesta foi estabelecida uma classificação somente para os agentes de bioterrorismo, na forma de três categorias (Tabela 2): as bactérias *Bacillus anthracis*, *Yersinia pestis*, *Francisella tularensis*, os vírus da varíola, Ebóla, Marburg, Lassa e Machupo, e a toxina de *Clostridium botulinum*, na categoria A, como os da mais alta prioridade; as bactérias dos gêneros *Brucella*, *Salmonella* e *Shigella*, das espécies *Burkholderia mallei*, *Burkholderia pseudomallei*, *Chlamydia psittaci*, *Coxiella burnetii*, *Rickettsia prowazekii* e *Vibrio cholerae*, do subtipo *Escherichia coli* O157:H7, o parasita *Cryptosporidium parvum*, os vírus das encefalites equinas ocidental, oriental e venezuelana, e as toxinas do *Clostridium perfringens*, ricina e enterotoxina B estafilocócica, na categoria B, como os de segunda maior prioridade, e uma

categoria C de doenças infecciosas emergentes como os Nipah vírus e hantavírus (CDC, 2019).

Tabela 2: Resumo das Categorias de Agentes de Bioterrorismo, segundo a classificação do CDC. Adaptado de: CDC, 2019

Categorias	Definição	Características	Exemplos
Categoria A	Agentes biológicos com alta prioridade, que representam um risco à segurança nacional	Podem ser facilmente disseminados ou transmitidos de pessoa a pessoa; altas taxas de mortalidade e alto potencial de impacto sobre a saúde pública; Podem provocar pânico e disrupção social; demandam capacidades especiais de preparação, na saúde pública	<i>Bacillus anthracis</i> Ebola
Categoria B	Segunda mais alta prioridade	Moderada facilidade de disseminação; baixas taxas de mortalidade, e morbidade moderada; Demandam aprimoramento das capacidades diagnósticas e de vigilância	<i>Rickettsia prowazekii</i> <i>Vibrio cholerae</i>
Categoria C	Terceira mais alta prioridade, inclui os patógenos emergentes que possam ter sido modificados para disseminação em massa	Grande disponibilidade; Facilidade de produção e disseminação; Potencial de altas morbidade, mortalidade e impacto na saúde pública	Nipah virus hantavirus

O *National Institute of Allergy and Infectious Diseases* (NIAID), uma divisão do *National Institutes of Health* (NIH), utiliza uma classificação baseada nas mesmas regras daquela do CDC, porém, mais abrangente, em número de organismos (NIAID, 2018). Ainda outra classificação em uso é a *Biological Select Agents and Toxins* (BSATs), uma publicação conjunta do *US Department of Health and Human Services* (DHHS) e do *US Department of Agriculture* (USDA), dividida em três seções: uma primeira lista de agentes que afetam seres humanos, uma segunda lista de agentes que afetam a agricultura e uma terceira lista de agentes que afetam ambos (SELECT AGENTS AND TOXINS LIST, 2023).

1.2 Métodos de Detecção e Identificação de Agentes Microbianos no Campo da Defesa Biológica

1.2.1 Métodos microbiológicos

Por volta de 1673, o holandês Anton van Leeuwenhoek fez a primeira observação de uma bactéria, em um microscópio, inventado por ele (Bonnet *et al.*, 2020). Quase dois séculos mais tarde, foi o francês Louis Pasteur quem descobriu o meio de cultura (Bonnet *et al.*, 2020). Apenas duas décadas haviam se passado, e o médico alemão Robert Koch, utilizou, pela primeira vez, o ágar para produzir meios de cultura sólidos, e assim obter culturas puras de bactérias (Bonnet *et al.*, 2020). Em 1887, Julius Richard Petri foi quem introduziu as “caixas circulares” que, não apenas, permitiam a observação das culturas, como, também, as protegiam da contaminação (Bonnet *et al.*, 2020). Esses foram alguns dos avanços fundamentais para o desenvolvimento da microbiologia, no século XIX, culminando em sua ampla utilização para diagnóstico e identificação de microrganismos.

Os métodos tradicionais de identificação de microrganismos, baseados em cultura, são considerados o padrão ouro (Žukowska, 2021). Esses métodos compreendem a coloração de Gram, o crescimento de cultura bacteriana em meios artificiais, e a identificação baseada no perfil bioquímico de cada espécie (Žukowska, 2021). A maior vantagem dos métodos microbiológicos convencionais está na confiabilidade, por serem técnicas bem validadas (Žukowska, 2021).

Esses métodos têm limitações claras pela relativa baixa sensibilidade e seletividade para os microrganismos nutricionalmente demandantes ou aqueles com limitada atividade bioquímica, e são aplicáveis apenas aos microrganismos com condições de crescimento conhecidas (Žukowska, 2021). Existe, ainda, a dificuldade de resolver espécies estreitamente relacionadas, como, por exemplo, *Bacillus anthracis* e *Bacillus cereus*, porque foram identificados isolados de *B. cereus* não-hemolíticos, não-móveis, sensíveis à penicilina e à lise por fago gama, bem como, isolados de *B. anthracis* hemolíticos, resistentes à penicilina e ao fago gama (Zasada, 2020).

O tempo requerido para a identificação de microrganismos por métodos convencionais, que pode levar, pelo menos, de dois a cinco dias, o trabalho intensivo (preparação de meios, diluição, plaqueamento, contagem, isolamento e caracterização), as distorções quantitativas, pelo enriquecimento, e a completa omissão das espécies não cultiváveis, são outras limitações reconhecidas para essas técnicas. (Franco-Duarte *et al.*, 2019).

1.2.2 Métodos imunocromatográficos

Os testes rápidos, baseados nos Imunoensaios de Fluxo Lateral (*Lateral Flow ImmunoAssay*, LFIA), também conhecidos como imunocromatografia, alcançaram enorme sucesso como ferramentas auxiliares do diagnóstico, inclusive na detecção e identificação de agentes de bioterrorismo, graças ao baixo custo, facilidade, rapidez e portabilidade, combinados com as altas sensibilidade e especificidade característicos dos métodos imunológicos (Calabria *et al.*, 2021). São capazes de detectar alvos mesmo em matrizes biológicas complexas, sem tratamento prévio da amostra (Calabria *et al.*, 2021).

Um teste comercial do tipo LFIA para a detecção do antígeno capsular de *B. anthracis* em amostras de plasma, soro ou sangue humano, não precisa de refrigeração, requer mínimo volume de amostra (10 a 20 µL) e apresenta um resultado visual em 15 minutos. Foram encontrados níveis de sensibilidade e especificidade de 98% e 96%, respectivamente, em uma avaliação feita em um surto de antraz em búfalos e hipopótamos, na Namíbia (Kolton *et al.*, 2019). A imunocromatografia tem, como limitações, a sensibilidade, reprodutibilidade, multiplexação, reatividade cruzada e baixa integração com sistemas eletrônicos (Calabria *et al.*, 2021).

1.2.3 Métodos moleculares

Já, entre as técnicas moleculares, a dessorção/ionização a laser auxiliada por matriz—espectrometria de massas por tempo de voo (*Matrix-assisted laser desorption/ionization—time-of-flight mass spectrometry*, MALDI-TOF) é um dos métodos mais usados para a identificação de microrganismos, no ambiente clínico, por ser rápido, conveniente, preciso e custo-efetivo (Tsuchida; Umemura; Nakayama, 2020). A técnica consiste em obter um espectro, baseado na composição dos peptídios presentes em cada amostra. A abundância relativa de cada peptídeo é plotada em função da sua relação massa/carga (m/z), calculada com base no tempo de voo após a ionização da amostra. Por fim, a identificação é realizada comparando os espectros obtidos com os espectros presentes no banco de dados de referência, fornecido pelo fabricante (Chen *et al.*, 2021).

A identificação de bactérias MALDI-TOF/MS é dependente de cultura e isolamento prévios, e há dificuldades para resolver espécies estreitamente relacionadas, como *Shigella* e *E. coli* (Solntceva; Kostrzewa; Larrouy-Maumus, 2021). Uma análise de 160 isolados de bactérias do grupo *Bacillus cereus*, incluindo 57 isolados de *B. anthracis*, usando um sistema MALDI-TOF/MS comercial, reportou a totalidade dos isolados como *B. anthracis* (Manzulli *et al.*, 2021).

Outra técnica molecular, possivelmente a mais importante técnica de diagnóstico, é a Reação em Cadeia da Polimerase (*Polymerase Chain Reaction*, PCR), descrita em 1983, por Kary Banks Mullis, ganhador do Prêmio Nobel de Química de 1993. A PCR possibilita a multiplicação de fragmentos de DNA em algumas horas, permitindo realizar pesquisas genéticas em quantidades mínimas de material biológico (Soroka; Wasowicz; Rymaszewska, 2021).

A técnica consiste na repetição de ciclos de desnaturação do DNA, anelamento de *primers*, flanqueando as extremidades da região de DNA alvo, extensão desses *primers* empregando a polimerase do *Thermophilus aquaticus* (Taq polimerase), e a desnaturação do DNA sintetizado. A reação é controlada por um termociclador que eleva e abaixa a temperatura da reação, sucessivamente, em intervalos de tempo programados (Mullis *et al.*, 1986; Zhu *et al.*, 2020).

Em um estudo, foi demonstrada a detecção simultânea de *B. anthracis*, *Y. pestis*, *F. Tularensis*, e *Brucella spp.*, usando a técnica de *Multiplex Oligonucleotide Ligation-PCR* (MOL-PCR), num ensaio com duração de quatro horas, sem falsos positivos, falsos negativos ou reações cruzadas entre as bactérias, com limites de detecção inferiores a 1 ng/μL (Jelinkova *et al.*, 2021).

Para a detecção de microrganismos patogênicos, o sequenciamento de DNA é a técnica que possibilita multiplexação ilimitada, e nenhum conhecimento prévio do agente etiológico, em contraste com a PCR (Minogue *et al.*, 2019). Duas técnicas pioneiras de sequenciamento de DNA foram desenvolvidas, independentemente, na década de 70, por Frederick Sanger (Sanger; Nicklen; Coulson, 1977), e Walter Gilbert (Maxam; Gilbert, 1977), que, por esses trabalhos, foram laureados com o Prêmio Nobel de Química de 1980 (Walker, 2014). A Primeira Geração de sequenciadores empregava o método Sanger, com algumas evoluções, como o uso de fluorescência, para identificar os nucleotídeos, e eletroforese capilar, para separar os fragmentos de DNA (Shendure *et al.*, 2017).

Em meados dos anos 2000, surgia a Segunda Geração de tecnologias de sequenciamento de DNA: o pirosequenciamento, o sequenciamento por ligação e o sequenciamento por terminadores de cadeia reversíveis (Minogue *et al.*, 2019). Também conhecidas, coletivamente, como Sequenciamento Massivo-Paralelo (*Massive-Parallel Sequencing*, MPS) ou Sequenciamento de Nova-Geração (*Next-Generation Sequencing*, NGS), essas tecnologias se caracterizam pelo altíssimo desempenho, podendo gerar bilhões de *reads* curtas (sequências de fragmentos de DNA, com tamanhos de 50 a 250 bases), totalizando terabases de sequências, em dois dias de trabalho de um único sequenciador, ao custo de alguns milhares de dólares (Shendure *et al.*, 2017).

Uma Terceira de Geração de sequenciadores surgiria a partir de 2010, capazes de produzir *reads* longas (maiores que 1.000 bases), podendo chegar até megabases de comprimento (Linde *et al.*, 2023), introduzindo duas tecnologias de sequenciamento de única molécula, em tempo real (Shendure *et al.*, 2017).

A detecção rápida de *Y. pestis* (6,5 h) e *B. anthracis* (8,5 h), com a investigação de marcadores de resistência a antimicrobianos, engenharia genética,

e variantes estruturais, utilizando o sequenciamento por nanoporos, foi demonstrada (Gargis *et al.*, 2019). Metodologias de identificação de microrganismos, empregando o MPS, incluem o sequenciamento do gene 16S rRNA, que tem como vantagens o sequenciamento da mesma região gênica repetidamente, aumentando a profundidade do sequenciamento e a confiabilidade dos resultados, e o uso da amplificação por PCR possibilita a análise de amostras com mínimas quantidades de DNA (Dias *et al.*, 2020).

1.3 O gene 16S rRNA e sua Aplicação na Identificação Bacteriana

Na década de 1980, a classificação sistemática dos seres vivos, foi revolucionada pela noção de que “as técnicas de sequenciamento molecular permitiam a medição direta das relações genealógicas” (Fox *et al.*, 1980). Entretanto, as metodologias bem estabelecidas para filogenia dos eucariotos, como a análise comparativa das sequências do citocromo c oxidase, não eram adequadas ao estudo das bactérias (Fox *et al.*, 1980).

A molécula escolhida para explorar a filogenia bacteriana, o RNA ribossômico (rRNA), tem distribuição universal, constância de função, e sua sequência muda muito lentamente. A taxa de divergência estimada do 16S rRNA é de 1% cada 50 milhões de anos (Ochman; Elwyn; Moran, 1999). Desde a década de 1970, as sequências de 16S rRNA tem sido usadas para explorar a filogenia dos procariotos. Nesses estudos, as relações entre os organismos passaram a ser quantificadas em termos de um coeficiente de associação, calculado para cada par de organismos. O coeficiente de associação está relacionado ao número de diferenças de nucleotídeos entre as sequências do 16S rRNA (Fox *et al.*, 1980).

1.3.1 Estrutura do gene 16S rRNA

O tamanho médio do gene 16S rRNA é de aproximadamente 1.500 nucleotídeos (1,5 kb), considerando que alguns organismos podem ter sequências mais longas ou mais curtas. Ele codifica o RNA que é parte da subunidade 30S do

ribossomo procarioto. O 16S rRNA tem diversas funções e é fundamental para a síntese de proteínas. Todo microrganismo possui, pelo menos, uma cópia do gene 16S, e o gene, além de ser altamente conservado, tem uma evolução lenta. Por essas razões, é o gene alvo mais amplamente usado em estudos filogenéticos de arqueias e bactérias (Church *et al.*, 2020).

A classificação taxonômica baseada no gene 16S rRNA tem se tornado uma ferramenta comumente empregada na identificação bacteriana por três razões: a primeira se deve ao tamanho relativamente pequeno do gene; a segunda é que as regiões conservadas do gene são comuns entre a maioria das bactérias, e já existem *primers* universais para as regiões conservadas; a terceira é que o número de sequências registradas em bases de dados públicas tem crescido substancialmente (Fukuda *et al.*, 2016).

Na sua estrutura, altamente conservada, estão interespaçadas 9 regiões hipervariáveis (Figura 4). As regiões conservadas (Figura 4) podem ser usadas para a ligação de *primers*, para a captura de um grande número de espécies de bactérias, e, possivelmente, arqueias, enquanto as hipervariáveis permitem a discriminação dessas espécies (Abellan-Schneyder *et al.*, 2021).

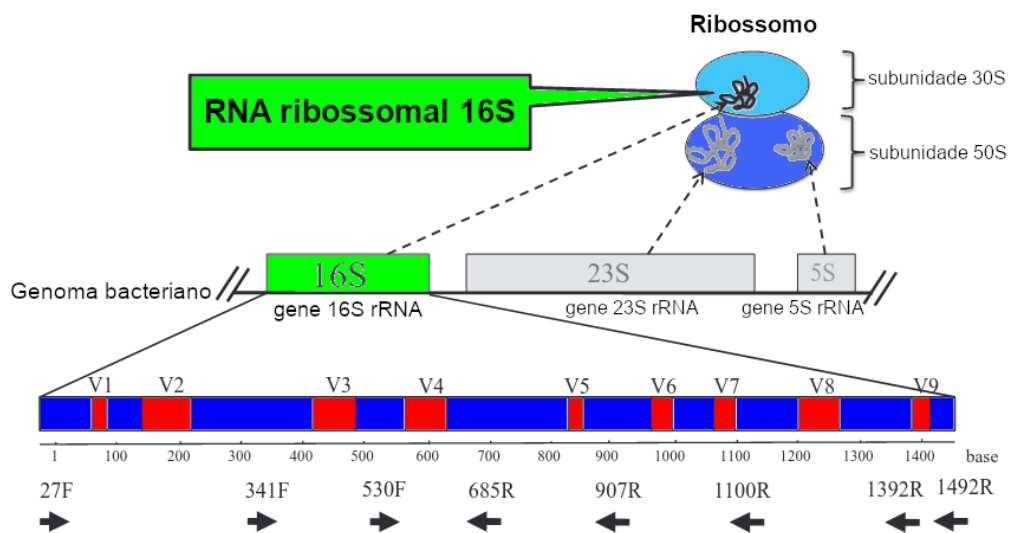


Figura 4: Representação do gene 16S rRNA e da cobertura dos pares de *primers* proposta por Fukuda e colaboradores (2016). No desenho, em azul estão representadas as regiões conservadas, e em vermelho, as hipervariáveis. Adaptado de: Fukuda *et al.*, 2016

As sequências dos *primers* comumente usados nos estudos do 16S rRNA são apresentadas na Tabela 3. Os *primers* 27F, que se estende da posição 8 a 27, no 16S rRNA de *E. coli*, e 1492R, que se estende da posição 1.492 a 1.507, conforme as posições no rRNA de *Escherichia coli*, permitem amplificar o 16S rRNA na quase totalidade de sua extensão (Frank *et al.*, 2008; Waechter *et al.*, 2023).

Tabela 3: Sequências dos primers utilizados nos estudos do 16S rRNA e respectiva cobertura relativa, no banco de dados Ribosomal Database Project (RDP) Adaptado de: Fukuda *et al.*, 2016

Nome	Sequência (5' para 3')	Nº de <i>matches</i> / Sequências	Cobertura (%)
27F	AGAGTTTGATCMTGGCTCAG	1846/9752	18.9
109R	ACGYGTTACKCACCCGT	6741/9752	69.1
341F	CCTACGGGAGGCAGCAG	9261/9752	94.9
530F	GTGCCAGCMGCCGCGG	9576/9752	98.2
685R	TCTRCGCATTYACCGCTAC	5976/9752	61.3
907R	CCGTCAATTCMTTTRAGTTT	9047/9752	92.8
1100R	GGGTTGCGCTCGTTG	8424/9752	86.4
1392R	ACGGGCGGTGTGTRC	9237/9752	94.7
1492R	TACGGYTACCTTGTTACGACTT	3687/9752	37.8

Há, ainda, a possibilidade de se desenhar *primers* para atender aos objetivos específicos de cada estudo. Foram descritas versões degeneradas do *primer* 27F, apresentadas na Tabela 4, anotadas usando o código IUPAC (*International Union of Pure and Applied Chemistry*) para especificação ambígua de nucleotídeos (Johnson, 2010), como também *primers* para posições específicas do gene 16S para os gêneros *Lactobacillus* e *Gardnerella* (Frank *et al.*, 2008).

Tabela 4: Primers alternativos desenhados para modificar a cobertura da reação de amplificação do gene 16S rRNA

Adaptado de: Frank *et al.*, 2008

Primer/ Região gênica	Nome/ Versão	Sequência (5' para 3')	Código IUPAC/ (Equivalência)	Organismos de interesse
27F	27f-CM	AGAGTTTGATCMTGGCTCAG	M (A ou C)	-
	27f-YM	AGAGTTTGATYMTGGCTCAG	Y (C ou T)	-
	27f-Bif	AGGGTTCGATTCTGGCTCAG	-	<i>Bifidobacteriaceae</i>
	27f-Bor	AGAGTTTGATCCTGGCTTAG	-	<i>Borrelia</i>
	27f-Chl	AGAATTTGATCTTGGTTCAG	-	<i>Chlamydiales</i>
16S, posições 200 a 400	LactoF	TGGAAACAGRTGCTAATACCG	R (A ou G)	<i>Lactobacillus</i>
	LactoR	GYCCATTGTGGAAGATTCCC	Y (C ou T)	
16S, posições 180 a 360	GardnF	GACTGAGATACGGCCAGAC	-	<i>Gardnerella</i>
	GardnR	ATTCGAAAGGTACTCACC	-	

1.3.2 Classificação taxonômica baseada no gene 16S rRNA

A partir das sequências do gene 16S rRNA de alta qualidade, obtidas por sequenciamento massivo-paralelo, é possível fazer a comparação com sequências de referência depositadas em um banco de dados, e determinar a classificação taxonômica, usando algoritmos, como o *Basic Local Alignment Search Tools* (BLAST) (Altschul *et al.*, 1990). A resolução do nível hierárquico da classificação taxonômica depende dos resultados do sequenciamento e, tipicamente, os limites de 80%, 85%, 90%, 92%, 94% e 97% são atribuídos, respectivamente, aos níveis de filo, classe, ordem, família, gênero e espécie (Fukuda *et al.*, 2016).

Dada a predominância das tecnologias de sequenciamento de *short reads*, isto é, sequências produzidas a partir de fragmentos curtos de DNA, geralmente entre 100 e 500 pares de bases, a maior parte dos estudos baseados na região 16S rRNA foi realizada com sequências de uma ou mais regiões do gene (V1-V2, V1-V3, V3-V5, V4, V6, e V6-V9) (Johnson *et al.*, 2019). A capacidade de resolução taxonômica a partir das sub-regiões do 16S rRNA foi sempre inferior em comparação com a realizada empregando todo o gene (Johnson *et al.*, 2019).

Limitações importantes desta metodologia foram demonstradas em diversos estudos de composição de comunidades de bactérias. Diferenças no desenho dos

primers também resultaram em avaliações diferentes da diversidade microbiana em comunidades (Frank *et al.*, 2008; Waechter *et al.*, 2023) e foi demonstrado que diferentes métodos de extração de DNA, vieses na amplificação por PCR, a escolha da região do gene 16S rRNA, o conjunto de *primers* empregados, a base de dados de referência, e o *pipeline* de análise, influenciam a avaliação da composição das comunidades bacterianas (Abellan-Schneyder *et al.*, 2021; Johnson *et al.*, 2019; Pollock *et al.*, 2018).

1.3.3 Ferramentas de bioinformática aplicadas no contexto da identificação microbiana

Aplicações de bioinformática para genômica e metagenômica foram surgindo, acompanhando o desenvolvimento e a popularização dos sequenciadores de DNA. Com o advento do sequenciamento massivo-paralelo, o número de sequências biológicas depositadas em bases de dados públicas cresceu exponencialmente. Um banco de dados é um conjunto de informações armazenadas, que correspondem a uma representação de um aspecto do mundo real (Matioli, 2021).

O sequenciamento massivo-paralelo, também chamado de sequenciamento de nova geração, é uma tecnologia de sequenciamento de alta capacidade de produção de dados (*high-throughput*), capaz de processar múltiplas reações de sequenciamento, em paralelo (Bruijns; Tiggelaar; Gardeniers, 2018). Para lidar com o enorme volume de dados produzido pelos modernos sequenciadores, a comunidade de bioinformatas tem produzido novos bancos de dados, ferramentas de montagem e alinhamento de *reads*, e *pipelines* de análise. *Pipeline* é um conjunto de algoritmos de bioinformática, executados em sequência pré-determinada, para processar os dados produzidos por sequenciamento de nova geração (Roy *et al.*, 2018).

O *National Center for Biotechnology Information* (NCBI, *Bethesda, Maryland, EUA*) é, possivelmente, a base de dados biológicos mais acessada e conhecida. Foi criado em 04 de novembro de 1988, como uma divisão da *National Library of Medicine* (NLM) do *National Institutes of Health* (NIH) (Our Mission - NCBI, [s. d.]).

As sequências do *16S RefSeq* é uma coleção curada, criada com os objetivos de “prover ferramentas para melhor analisar e validar os dados de rRNA, manter informações taxonômicas completas e atualizadas sobre arqueias e bactérias, aprimorar a anotação de genes, possibilitar a identificação baseada em sequências, e possibilitar a extração de subconjuntos de dados específicos” (16S RefSeq records processing and curation, [s. d.]). Atualmente, a base de dados 16S rRNA do NCBI conta com mais de 27 mil registros. O NCBI disponibiliza, ainda, o pacote de ferramentas de bioinformática chamado BLAST+ que, entre outras funcionalidades, permite a execução do BLAST em um computador local.

O *Greengenes* é outra base de dados destinada à classificação taxonômica de arqueias e bactérias, baseada no gene 16S rRNA (McDonald *et al.*, 2012). O banco de dados foi criado para resolver o problema de subclassificação das sequências depositadas no NCBI (McDonald *et al.*, 2012). A taxonomia do *Greengenes* está baseada na construção *de novo* de uma árvore taxonômica, ou seja, do princípio, a partir de mais de 400 mil sequências, filtradas por qualidade, calculada com o uso da ferramenta *FastTree* (McDonald *et al.*, 2012). Atualmente, a versão 13.5 do banco de dados do *Greengenes* possui mais de 1 milhão de sequências, de mais de 90 mil organismos (greengenes.secondgenome.com, [s. d.]).

A base de dados *Ribosomal Database Project* (RDP-II) também tem uma proposta de oferecer sequências anotadas, ferramentas de alinhamento e serviços de análises de sequências (Cole *et al.*, 2005). O serviço era hospedado pela *Michigan State University* (MSU, Lansing, Michigan, EUA), mas o *site* do RDP não está mais *online* e, atualmente a versão 18 do banco de dados, contendo 20 mil sequências de 16S rRNA de bactérias, e 600 de arqueias, pode ser acessada através do *site* da ferramenta *mothur* (mothur website, [s. d.]).

O banco de dados SILVA (floresta, em latim) é um banco de dados de sequências das subunidades ribossomais pequenas (*small subunit*, SSU) e subunidades ribossomais grandes (*large subunit*, LSU) (Quast *et al.*, 2013). A versão mais recente 138.1, acumula mais de 9 milhões de sequências com alinhamentos superiores a 50% de identidade e qualidade acima de 40, a partir de sequências

com escores de qualidades de, pelo menos, 30. Há, também, o subconjunto de sequências SSU Ref, com apenas sequências maiores que 1.200 bp, de bactérias, e 900 bp, de arqueias, e identidade maior que 70% ou qualidade maior que 50, com mais de 2 milhões de sequências, e ainda o SSU Ref NR 99, somente com sequências de 99% de identidade, com 500 mil sequências (Silva, [s. d.]).

A ferramenta Arb (de *arbor*, árvore, em latim), foi uma iniciativa interdisciplinar de bioinformática da *Technische Universität München* (Munique, Alemanha) (Ludwig *et al.*, 2004). O conceito por trás do Arb estabeleceu como tarefas prioritárias a “manutenção de uma base de dados secundária, estruturada e integrativa, combinando dados estruturais primários processados e quaisquer dados adicionais associados às sequências”, e “uma abrangente seleção de ferramentas de *software*, integradas entre si e com a base de dados central, todas controladas por uma interface gráfica de usuário comum” (Ludwig *et al.*, 2004). Arb integra ferramenta de busca, alinhamento de sequências, editor de sequências, gerador de árvore filogenética, e desenho e avaliação de sondas (Ludwig *et al.*, 2004). Desde de 2014, o Arb passou a ser mantido pelo *Department of Molecular Ecology*, no *Max Planck Institute for Marine Microbiology* (Bremen, Alemanha) (ARB: Project, [s. d.]).

O aumento de escala e de complexidade das pesquisas de centenas para centenas de milhares de sequências criou uma demanda por algoritmos mais rápidos (Schloss *et al.*, 2009). O *mothur* foi desenvolvido, implementando algoritmos usados nas ferramentas *DOTUR*, *SONS*, *TreeClimber*, *LIBSHUFF*, *[-LIBSHUFF]* e *UniFrac*, incorporando, ainda, calculadoras para quantificar parâmetros ecológicos chave, para medidas de diversidade α e β ; ferramentas de visualização para diagramas Venn, mapas de calor e dendogramas; funções para triagem de coleções de sequências por qualidade; alinhador de sequências; e calculadora de distância de sequências (Schloss *et al.*, 2009). Possui a flexibilidade para invocar os comandos individuais ou como arquivos de lote (Schloss, 2020; Schloss *et al.*, 2009).

Outra ferramenta muito utilizada, o *Quantitative Insights into Microbial Ecology* (QIIME) consiste em um *software* de código fonte aberto desenvolvido para lidar com dados brutos de sequenciamento, até a interpretação e depósito em base

de dados, e suporta um grande número de análises e visualizações de comunidades microbianas (Caporaso *et al.*, 2010). A interação com o usuário é feita por meio de interface gráfica (Caporaso *et al.*, 2010). O QIIME foi implementado de forma altamente modular, possibilitando o uso de componentes alternativos para seleção de unidades taxonômicas operacionais (*Operational Taxonomic Units*, OTU), alinhamento de sequências, construção de árvores filogenéticas, e avaliação de diversidade, baseada em *taxon*, intra e inter-amostras (Caporaso *et al.*, 2010).

Em geral, esses bancos de dados tanto permitem que os usuários depositem sequências produzidas por eles, o que garante a expansão continuada dos acervos de sequências, como também copiem as sequências depositadas por outros usuários, o que possibilita fazer o uso local das mesmas, usando o *software* e *hardware* disponíveis no próprio laboratório (The Sequence Read Archive (SRA), [s. d.]).

1.4 Organizações Nacionais e Internacionais de Enfrentamento aos Agentes Biológicos

Não existe uma única organização internacional ou nacional dedicada ao enfrentamento dos agentes biológicos, e diversas organizações compartilham competências concorrentes, com algum grau de sobreposição. Assim a resposta aos incidentes com agentes biológicos é organizada de forma diferente, em cada país. Mesmo as informações sobre o funcionamento e a organização dessas instituições não seguem um padrão, e só oferecem uma visão geral da sua operacionalização. Foram relacionadas as informações sobre diferentes organismos, planos, estratégias e iniciativas, nacionais ou multilaterais, que, na sua constituição, preveem, entre outras atribuições, o enfrentamento aos surtos epidêmicos e às ações de bioterrorismos (Canada, 2018; International Health Regulations (2005) – Third edition, [s. d.]; National Biodefense Strategy, [s. d.]; Outi, 2022).

A Organização Mundial da Saúde, por meio do Anexo I das Regulações Internacionais de Saúde (*International Health Regulations*), atualizadas em 2005,

estabeleceu requisitos mínimos de vigilância, notificação, verificação, detecção, implementação de controles preliminares, assistência, comunicação, análise de amostras, investigação, operação e manutenção de um plano nacional de emergência em saúde pública, com “capacidade para detectar eventos envolvendo doenças e mortes em números acima do esperado, para determinado tempo e local, em todo o território” (International Health Regulations (2005) – Third edition, [s. d.]). Esse instrumento obriga que os estados-membros fortaleçam e mantenham “capacidades de avaliar, notificar e relatar eventos”, comunicando à OMS, em até 24 horas, quaisquer eventos que possam constituir “emergência de saúde pública internacional”, ocorridos em seus territórios (International Health Regulations (2005) – Third edition, [s. d.]).

Na União Europeia, os estados-membros adotaram o *Health Security Programme* que tem como uma de suas prioridades “implementar um mecanismo de troca de informações, consultas e coordenação para a cuidar de questões de saúde relativas a ataques nos quais agentes químicos ou biológicos foram ou possam ter sido usados” (Outi, 2022). A coordenação entre os estados-membros, as agências como o *European Centre for Disease Prevention and Control* (ECDC) e *European Food Safety Authority* (EFSA), e organizações internacionais, como a OMS, é feita pela *Health Emergency Operations Facility* (HEOF) (Outi, 2022). As questões relacionadas com agentes químicos, biológicos e nucleares também integram essa estrutura, por meio do *Health Security Committee* (HSC) (Outi, 2022).

A Comissão Europeia conta ainda com ferramentas como o *Medical Intelligence System* (MedISys), desenvolvida pela *Joint Research Centre* (JRC), para detectar e monitorar potenciais ameaças à saúde pública, inclusive Químico, Biológico, Radiológico e Nuclear (QBRN), através do acompanhamento da mídia nos níveis local, nacional e internacional, de forma totalmente automatizada (Steinberger *et al.*, 2008), e o RAS-BICHAT, um sistema de alerta rápido, baseado em web, para a troca de informações sobre ameaças à saúde pública pela liberação intencional de agentes QBRN.

Nos EUA, o *National Biodefense Strategy and Implementation Plan for Countering Biological Threats, Enhancing Pandemic Preparedness, and Achieving*

Global Health Security é o documento que explica como o governo gerencia as ações destinadas a avaliar, prevenir, preparar, responder e recuperar-se das ameaças biológicas, coordenando os esforços de biodefesa com as entidades estaduais, locais, territoriais e tribais, parceiros internacionais, indústrias, academia, organizações não-governamentais, e o setor privado (National Biodefense Strategy, [s. d.]). Essa estratégia define como missão do governo federal “salvar vidas, reduzir o sofrimento humano e animal, proteger propriedades e o meio ambiente, controlar a disseminação das doenças, apoiar os esforços das comunidades para superar os impactos físicos, emocionais, ambientais e econômicos do incidente, e identificar a causa e fonte do incidente” (National Biodefense Strategy, [s. d.]). O desenvolvimento dessa estratégia foi realizado sobre os seguintes pressupostos: as ameaças biológicas são persistentes; as ameaças biológicas originam-se de múltiplas fontes; doenças infecciosas não conhecem fronteiras; incidentes biológicos afetam a infraestrutura crítica e as cadeias de suprimento; a cooperação multilateral e multissetorial é crítica para uma biodefesa efetiva; a abordagem da saúde única (visão integrada da saúde humana, saúde animal e meio ambiente) reduz a ocorrência e o impacto dos incidentes biológicos; ciência e tecnologia continuarão a se desenvolver globalmente (National Biodefense Strategy, [s. d.]). Os cinco objetivos estabelecidos para a estratégia foram: viabilizar a consciência de risco e a detecção para a tomada de decisão informada em biodefesa; assegurar as capacidades para a prevenção dos incidentes biológicos; assegurar preparação em biodefesa para reduzir os impactos dos incidentes biológicos; responder rapidamente para limitar os impactos dos incidentes biológicos; e facilitar a recuperação para restaurar a comunidade, a economia, e o ambiente, após o incidente biológico (National Biodefense Strategy, [s. d.]).

O Canadá envolveu os diversos níveis de governo na elaboração do *Chemical, Biological, Radiological, Nuclear and Explosives Resilience Strategy for Canada*, numa abordagem que se baseou em quatro componentes de gerenciamento de emergências, a prevenção, o preparo, a resposta e a recuperação, e, também, em cinco objetivos estratégicos chave, a liderança, o gerenciamento de riscos, o planejamento baseado em capacidades, a

interoperabilidade e efetividade das equipes, e a gestão do conhecimento e informação (Canada, 2018). A estratégia foi desenhada para funcionar em conjunto com outros mecanismos e jurisdições existentes como o *Federal Emergency Response Plan* (FERP), o *National Emergency Response System* (NERS), o *Federal Nuclear Emergency Plan* (FNEP), o *National Counter-Terrorism Plan*, e, ainda, os planos de resposta dos níveis provinciais, territoriais e municipais (Canada, 2018).

No Brasil, o Sistema Nacional de Laboratórios de Saúde Pública (SISLAB), instituído em 2004, é formado por um “conjunto de redes nacionais de laboratórios, organizadas em sub-redes, por agravos ou programas, de forma hierarquizada por grau de complexidade das atividades relacionadas à vigilância em saúde – compreendendo a vigilância epidemiológica e vigilância em saúde ambiental, vigilância sanitária e assistência médica”. Dentro desse sistema, as unidades laboratoriais foram classificadas em sete níveis de complexidade: Centros Colaboradores (CC), Laboratórios de Referência Nacional (LRN), Laboratórios de Referência Regional (LRR), Laboratórios de Referência Estadual (LRE), Laboratórios de Referência Municipal (LRM), Laboratórios Locais (LL) e Laboratórios de Fronteira (LF) (Brasil, 2019).

O Guia para Diagnóstico Laboratorial em Saúde Pública relaciona os laboratórios de referência (Tabela 5) para diversos agentes biológicos de interesse da biodefesa (Brasil, 2022):

Tabela 5: Laboratórios civis brasileiros de referência para o diagnóstico dos agentes de bioterrorismo. Adaptado de: Brasil, 2022

AGENTE BIOLÓGICO	CENTRO COLABORADOR	LABORATÓRIO DE REFERÊNCIA NACIONAL	LABORATÓRIOS DE REFERÊNCIA REGIONAIS
Toxina botulínica	Não há	Instituto Adolfo Lutz – IAL/São Paulo	Não há
<i>Brucella</i>	Instituto Adolfo Lutz – IAL/São Paulo	Não há	Não há
Antraz	Não há	Instituto Oswaldo Cruz – Fiocruz/RJ	Instituto Evandro Chagas – IEC/PA
<i>V. colera</i>	Não há	Fundação Oswaldo Cruz – Fiocruz/RJ	Fundação Ezequiel Dias – Funed/MG, LACEN/PE, Instituto Adolfo Lutz – IAL/São Paulo, LACEN/DF, Instituto Evandro Chagas – IEC/PA
Ebola	Não há	Instituto Evandro Chagas – IEC/PA	Não há
<i>Salmonella, E. coli O157:H7, Shigella</i>	Não há	Fundação Oswaldo Cruz – Fiocruz/RJ	Fundação Ezequiel Dias – Funed/MG, LACEN/PE, Instituto Adolfo Lutz – IAL/São Paulo, LACEN/DF, Instituto Evandro Chagas – IEC/PA
Lassa	Não há	Instituto Evandro Chagas – IEC/PA	
<i>Rickettsia prowazekii</i>	Não há	Fundação Ezequiel Dias – Funed/MG	Instituto Adolfo Lutz – IAL/São Paulo
Hantavírus	Instituto Carlos Chagas – Fiocruz/Paraná	Instituto Adolfo Lutz – IAL/São Paulo	Fundação Oswaldo Cruz – Fiocruz/RJ, Instituto Evandro Chagas – IEC/PA
Toxina estafilocócica	Fundação Ezequiel Dias – Funed/MG	Não há	Não há
<i>Y. pestis</i>	Não há	Centro de Pesquisas Aggeu (CPqAM)	Não há
<i>F. tularensis</i>	Centro de Pesquisas Aggeu (CPqAM)	Não há	Não há

Além dos laboratórios vinculados ao Ministério da Saúde, o Ministério da Ciência, Tecnologia e Inovação relaciona 17 laboratórios com capacidade NB-3 (Tabela 6), funcionando em Universidades (Plataforma Nacional de Infraestrutura de Pesquisa, [s. d.]):

Tabela 6: Relação dos laboratórios públicos de alta contenção, por unidade da federação Adaptado de: PLATAFORMA NACIONAL DE INFRAESTRUTURA DE PESQUISA, [s. d.]

UF	Órgão Gestor	Depto. Responsável	Finalidade
MG	Universidade Federal de Lavras (UFLA)	Biotério Central	<ul style="list-style-type: none"> – pesquisas sobre doenças que afetam humanos, animais e até mesmo plantas; – pesquisas avançadas com patógenos.
ES	Universidade Federal do Espírito Santo (UFES)	Não informa	<ul style="list-style-type: none"> – pesquisas relacionadas a biovigilância, contenção, isolamento e caracterização de microrganismos de alto risco biológico.
AM	Instituto Nacional de Pesquisas da Amazônia (Inpa)	Laboratório Temático Biotério Central (LTBC)	<ul style="list-style-type: none"> – etapas pré-clínicas do desenvolvimento de um fármaco, com testes in vitro (células) e in vivo (animais), antes de seguir para os testes clínicos em humanos; – avaliar as substâncias isoladas e os extratos (vegetais, fúngicos e animais) sobre microrganismos altamente patogênicos, incluindo o vírus SARS-CoV-2 e outros multirresistentes hospitalares
MG	Universidade Federal de Minas Gerais (UFMG)	Instituto de Ciências Biológicas (ICB)	<ul style="list-style-type: none"> – estudos com microrganismos potencialmente ameaçadores para a sociedade, como o vírus Sars-CoV-2, causador da covid-19
RS	Universidade Federal do Rio Grande do Sul (UFRGS)	Instituto de Ciências Básicas da Saúde (ICBS)	<ul style="list-style-type: none"> – cultivo do SARS-CoV-2; – estudos com <i>Mycobacterium tuberculosis</i>.
RS	Universidade Federal de Santa Maria (UFSM)	Não informa	<ul style="list-style-type: none"> – pesquisa com amostras vivas do novo coronavírus – pesquisa de microrganismos que causam doenças potencialmente letais, incluindo vírus que provocam doenças como a Aids, febre amarela, gripe A, dengue, Zika e Chikungunya e bactérias como o bacilo de Koch
PR	Universidade Estadual de Maringá (UEM)	Central de Tecnologia em Saúde (CTS)	<ul style="list-style-type: none"> – Análise e cultivo celular – Manipulação de agentes infecciosos de nível 3 (NB3) – Cultura de agentes infecciosos de nível 3 (NB3) – Experimentação animal com agentes infecciosos de nível 3 (NB3) – Ensaios in vitro com agentes infecciosos de nível 3 (NB3)

UF	Órgão Gestor	Depto. Responsável	Finalidade
PA	Instituto Evandro Chagas (IEC)	Seção de Arbovirologia e Febres Hemorrágicas	– pesquisas e na vigilância de diversos patógenos de importância em saúde pública.
CE	Universidade Federal do Ceará (UFC)	Núcleo de Pesquisa e Desenvolvimento de Medicamentos (NPDM)	– desenvolvimento de todas as etapas pré-clínicas de pesquisa com patógenos.
RS	Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)	Centro de Pesquisas em Biologia Molecular e Funcional (CPBMF)	<ul style="list-style-type: none"> – Cultivo em meio de cultura sólido e líquido de <i>Mycobacterium tuberculosis</i> – Extração de ácidos nucleicos (DNA e RNA) e proteínas de <i>M. tuberculosis</i> – Ensaio de determinação de Concentração Inibitória Mínima (CIM) em <i>M. tuberculosis</i> – Ensaio de nocaute e silenciamento gênicos em <i>M. tuberculosis</i> – Curvas de crescimento e curvas de morte (time-kill curves) em <i>M. tuberculosis</i> – Ensaio de infecção de células eucarióticas com <i>M. tuberculosis</i> – Ensaio de infecção de células eucarióticas com o vírus SARS -CoV-2 – Ensaio de efeito citopático com SARS-CoV-2 – Ensaio de quantificação de carga viral com SARS-CoV-2
SP	Universidade de São Paulo (USP)	Departamento de Análises Clínicas e Toxicológicas	– pesquisa com isolados de Micobactérias de pacientes com infecção pelo bacilo que apresentava resistência a tratamento
PE	Universidade Federal de Pernambuco (UFPE)	Núcleo de Pesquisa em Inovação Terapêutica Suely Galdino (Nupit SG)	<ul style="list-style-type: none"> – desenvolvimento de estudos com micro-organismos causadores de doenças como a covid-19; – estudos com outros vírus emergentes e reemergentes e patógenos como hantavírus, tuberculose, HIV, além dos vírus da hepatite, febre amarela e dengue.
RJ	Universidade Federal do Rio de Janeiro (UFRJ)	Centro de Pesquisas de Medicina de Precisão (CPMF)	– apoiar projetos de pesquisa que necessitem de equipamentos e/ou infraestrutura para experimentação e manipulação de organismos altamente patogênicos
BA	Fiocruz Bahia	Instituto Gonçalo Moniz (IGM)	<ul style="list-style-type: none"> – pesquisas sobre a tuberculose; – estudos com os vírus HIV e HTLV.
SP	Universidade de São Paulo (USP)	Centro de Pesquisa em Virologia da Faculdade de Medicina de Ribeirão Preto	<ul style="list-style-type: none"> – manipulação e armazenamento de vírus patogênicos; – manipulação e cultura de células infectadas com vírus

UF	Órgão Gestor	Depto. Responsável	Finalidade
MG	Universidade Federal de Juiz de Fora (UFJF)	Não informa	– desenvolvimento de projetos com patógenos que requerem nível 3 de biossegurança, como SARs-CoV-2, Mycobacterium entre outros.
SP	Universidade de São Paulo (USP)	Departamento de Parasitologia do ICB	– pesquisa de patógenos humanos de risco biológico da classe 3, como protozoários, vírus e <i>Rickettsia rickettsii</i>

1.5 Atuação do Exército Brasileiro na Defesa Biológica

Entre 2007 e 2016, o Brasil sediou oito eventos internacionais promovidos por várias entidades como, por exemplo, o Comitê Olímpico Internacional (COI) e as Organizações das Nações Unidas (ONU), chamados, coletivamente, como os Grandes Eventos, porque envolveram diversos países, grande interesse midiático, participação de grande público, etc (Brasil, 2018). O Exército Brasileiro (EB) atuou na segurança desses eventos juntamente às outras Forças Armadas e também aos órgãos de segurança pública (Brasil, 2018). Entre outras atribuições, coube ao EB a Defesa Química, Biológica, Radiológica e Nuclear (DQBRN), realizando varreduras, monitoramento e operando postos de descontaminação (Brasil, 2018).

Nesse período, houve intensa atividade envolvendo os componentes do Sistema de Defesa Química, Biológica, Radiológica e Nuclear do Exército (SisDQBRNEx), que é o sistema que estrutura a capacitação, a preparação e a resposta do EB nos incidentes QBRN (Brasil, 2012), com investimentos que resultaram na aquisição de diversos equipamentos para as unidades militares integrantes do sistema, entre as quais o Instituto de Biologia do Exército (IBEx), órgão de assessoria científica do sistema, na área biológica (Brasil, 2018).

O Centro de Estudos em Biodefesa (Sousa *et al.*, 2019), do IBEx, integra, numa única edificação, dois laboratórios de segurança biológica, sendo um de Nível de Biossegurança 2 e o outro NB-3, concebidos para funcionar de maneira dual, atendendo tanto às demandas de diagnóstico clínico, como àquelas de pesquisa em biodefesa. A seção NB-2 está equipada com MALDI-TOF/MS, extrator automatizado de DNA, estufa bacteriológica, Cabines de Segurança Biológica (CSB) tipo 2 e

ultrafreezer, enquanto a seção NB-3 é equipada com uma CSB tipo 3 (*glovebox*), CSB tipo 2, ultrafreezer e estrutura para cultura de células.

A diretriz de funcionamento do IBEx, aprovada no ano de 2021, consolidou o papel multidimensional do Instituto, abrangendo as diversas atividades que este desenvolve, a começar pela assistência à saúde, como serviço de apoio ao diagnóstico, juntamente ao desenvolvimento do ensino e pesquisa, com foco na biodefesa, participando das atividades do SisDQBRNEx, na dimensão biológica. A diretriz prescreve, ainda, “o acompanhamento da evolução tecnológica”, consoante com a tradição de pioneirismo do Instituto, na incorporação de novas tecnologias.

2 OBJETIVOS

2.1 Objetivo Geral

Construir um banco de dados local de sequências do gene 16S rRNA de agentes de bioterrorismo para análises de dados de sequenciamento massivo-paralelo por nanoporos, para fins de identificação bacteriana em ações de defesa biológica, aplicável em laboratório móvel.

2.2 Objetivos Específicos

- Construir um banco de dados local de sequências do gene 16S rRNA de agentes de bioterrorismo, a partir de sequências depositadas em bancos de dados públicos;
- Realizar o sequenciamento do gene 16S rRNA da cepa de *Salmonella enterica* ATCC 14028 no sistema Mk1B (ONT), com o emprego do pacote de *softwares* próprio da ONT;
- Analisar os dados de sequenciamento massivo-paralelo usando o banco de dados local de 16S rRNA do agente de bioterrorismo construído no presente trabalho; e
- Comparar os resultados da análise dos dados gerados com o banco de dados local com a análise remota realizada com o *software* da Nanopore.

3 MATERIAIS E MÉTODOS

3.1 Construção de Bancos de Dados 16S rRNA de Agentes de Bioterrorismo

Para a construção do banco de dados local 16S rRNA de agentes de bioterrorismo, referido como 16S rRNA bt, foram utilizadas sequências do banco de dados 16S RefSeq (16S RefSeq, [s. d.]), parte do projeto *NCBI RefSeq Targeted Loci* (NCBI RefSeq Targeted Loci Project, [s. d.]), do NCBI, obtidas do servidor https://ftp.ncbi.nlm.nih.gov/blast/db/16S_ribosomal_RNA, atualizado em 21/10/2023.

O arquivo do 16S RefSeq contém 26.910 sequências de 16S rRNA, cada uma das espécies é referenciada por um número identificador de taxonomia (TaxID), que são as mesmas usadas para acessar a ferramenta *blastn* (Nucleotide BLAST, [s. d.]), e ocupa 198,6 MB (39,44 MB, compactado). O arquivo foi descompactado, e as sequências foram extraídas, filtradas e convertidas para o formato FASTA, usando-se as ferramentas do *software BLAST+* v2.15.0, conforme o manual (BLAST, 2008), e uma lista dos TaxID correspondentes às espécies de bactérias listadas como agentes de bioterrorismo (CDC, 2019), obtidos no banco de dados *Taxonomy* (Federhen, 2012), do NCBI, a partir dos nomes científicos. Uma sequência de *Proteus vulgaris* foi acrescentada ao banco de dados, porque este microrganismo foi selecionado para servir como controle, representando um agente bacteriano não pertencente ao grupo de interesse para bioterrorismo.

3.2 Cultivo Bacteriano e Extração do DNA genômico

Para o experimento, foram selecionadas cepas da coleção de culturas, disponíveis no Instituto de Biologia do Exército (IBEx). A cepa de *Salmonella enterica* ATCC 14028, Plastilabor (Rio de Janeiro, Rio de Janeiro, BR), foi escolhida como representante de agente de bioterrorismo categoria B (CDC, 2019). Como controle, representando um agente bacteriano não pertencente às categorias de interesse em bioterrorismo, foi utilizada a cepa de *P. vulgaris* ATCC 8427, Plastilabor. As cepas liofilizadas foram semeadas em caldo infusão cérebro-coração (*brain-heart infusion*, BHI), Laborclin (Pinhais, Paraná, BR) e incubadas em estufa à

36° C por 24 horas. Após o crescimento no caldo, foram transferidas com alça bacteriológica para placas de meio cisteína-lactato eletrólito deficiente (CLED), por esgotamento, e incubadas por 24 horas na estufa à 36° C. Colônias isoladas de cada bactéria foram transferidas, com alça bacteriológica, para novos tubos de caldo BHI, e incubadas em estufa, à 36° C, por 6 horas, previamente à extração.

A extração por *beads* magnéticas foi feita no extrator automático EXTRACTA STATION 9600, da Locus (Cotia, São Paulo, BR), nas instalações do laboratório NB-2 do IBEx. Foi utilizado o protocolo MPTA-PU16, empregando o kit Extracta DNA e RNA de Patógenos, da Locus. Pipetou-se 100 µl do caldo BHI de cada cepa, em um poço diferente da placa do kit.

3.3 Quantificação de DNA

O kit 16S *Barcoding* 1-24, da *Oxford Nanopore Technologies* (ONT – Oxford, UK), selecionado para a preparação da biblioteca, requer, pelo menos, 10 ng de DNA de entrada, de cada amostra. Quantidades de DNA muito acima ou abaixo desse valor podem comprometer a preparação da biblioteca. Para assegurar que a quantidade e a qualidade do DNA estivessem dentro dos parâmetros do kit, após a extração, as amostras foram quantificadas por fluorimetria, utilizando o fluorímetro *Quantus*, com o kit *QuantiFluor dsDNA System*, ambos da Promega (*Madison, Wisconsin*, EUA). Foram preparadas as seguintes soluções: solução 1X TE *Buffer* (1 mL de 20X TE *Buffer* diluído com 19 mL de água sem nuclease; solução de trabalho (1 µl de *QuantiFluor dsDNA Dye* diluído com 399 µl de 1X TE *Buffer*); branco (200 µl de solução de trabalho); solução padrão de DNA (2 µl de solução padrão de DNA 100 ng/µl diluídos em 200 µl de solução de trabalho); amostra (1 µl de amostra diluído em 200 µl de solução de trabalho). Cada solução preparada foi agitada com *vortex* por 2 segundos, e coberta com papel laminado para proteger da luz. Cada amostra preparada foi incubada à temperatura ambiente por 5 minutos. Antes de iniciar as medições, o protocolo dsDNA foi selecionado, seguido da calibração com a leitura do branco e do padrão de DNA, e o volume de amostra adicionado foi informado.

3.4 Preparação da Biblioteca e Corrida de Sequenciamento

A preparação da biblioteca foi realizada com o kit 16S Barcoding 1-24, que inclui 24 *barcodes* para multiplexação de amostras. Os *barcodes* de números 2 e 3, selecionados para o experimento, foram destacados do cartão com 24 *barcodes*, descongelados à temperatura ambiente, e deixados sobre um *cooler*, para manter sob refrigeração.

A solução *LongAmp Hot Start Taq 2X Master Mix*, da *New England BioLabs* (*Ipswich, Massachusetts*, EUA) foi descongelada à temperatura ambiente. Com uma micropipeta, a solução foi adicionada ao próprio tubo, repetidamente, para misturar. O tubo com a solução foi deixado sobre o *cooler*, para manter sob refrigeração.

De cada amostra foram transferidos 10 ng do DNA extraído para tubos de PCR e adicionados 10µl de água livre de nuclease. Os tubos foram agitados manualmente e depois centrifugados (*spin down*). Em um segundo tubo de PCR de 0,2 mL, para cada amostra foram adicionados 5 µl de água livre de nuclease, mais 25 µl da solução *LongAmp Hot Start Taq 2X Master Mix*, e transferidos 10 µl da solução de DNA do primeiro tubo. Com o auxílio de uma micropipeta para misturar, pipetando repetidamente, no próprio tubo. Depois da mistura, os tubos foram centrifugados (*spin down*).

Os lacres dos *barcodes* foram perfurados, usando-se para cada um uma ponteira nova. A solução foi pipetada no próprio tubo por 10 vezes, para misturar. Depois de homogeneizada, foram adicionados 10 µl de solução dos *barcodes* 2 e 3, respectivamente, para os tubos de *S. enterica* e *P. vulgaris*. Depois da adição dos *barcodes*, o conteúdo de cada tubo foi pipetada por 10 vezes no próprio tubo, para misturar. Por fim, seguiu-se a etapa de amplificação no termociclador VERITI 96 Thermal Cycler, da Applied Biosystems (*Waltham, Massachusetts*, EUA), programado como no Tabela 7.

Tabela 7: Programa do termociclador para a amplificação por PCR do gene 16S rRNA
Fonte: o autor

Estágio	Temperatura	Tempo	Ciclos
Desnaturação inicial	95° C	1 min	1
Desnaturação	95° C	20s	25
Anelamento	55° C	30s	25
Extensão	65° C	2min	25
Extensão final	65° C	5min	1
Fim	4° C	∞	-

Concluída a amplificação, as amostras foram transferidas para tubos de 1,5 mL para proceder a purificação dos produtos de PCR, por *beads* magnéticas. Para a purificação foi usado o reagente comercial *AMPure XP Reagent for PCR Purification*, da *Beckman Coulter Life Sciences (Brea, California, EUA)*. Primeiro o reagente foi descongelado à temperatura ambiente. Depois de descongelado, o frasco da *AMPure* foi colocado em agitador do tipo *vortex*, para ressuspender o reagente. Imediatamente após o *vortex* foram transferidos, com pipeta automática, 30 µl da suspensão para cada tubo de amostra. O reagente e a amostra foram misturados pipetando-se repetidamente no próprio tubo de reação. Depois de misturados, os tubos foram incubados em agitador rotativo por 5 minutos, à temperatura ambiente.

Foram preparados 500 µl de solução de etanol a 70% em água livre de nuclease. Os tubos de amostra com a suspensão de *beads* magnéticas foram centrifugados e colocados em uma estante magnética. Após a separação das *beads*, foi usada uma pipeta para aspirar o líquido sobrenadante, sem perturbar as *beads* aglomeradas na lateral do tubo. O líquido aspirado foi descartado. Com o tubo ainda na estante magnética, foi usada uma micropipeta para fazer a lavagem das *beads* com 200 µl do etanol a 70%. O etanol usado na lavagem das *beads* foi aspirado e desprezado. A lavagem das *beads* com etanol 70% foi repetida. Os tubos foram centrifugados e recolocados na estante magnética. O resíduo de etanol foi

aspirado e descartado, e os tubos foram deixados abertos por 30 segundos para evaporar o etanol restante.

Os tubos foram retirados da estante magnética e os agregados de partículas magnéticas foram ressuspensos com 10 µl de tampão Tris-HCl 10 mM com NaCl 50 mM, e incubados à temperatura ambiente, por 2 minutos. Os tubos foram recolocados na estante magnética até agregar as partículas magnéticas e clarificar o eluato. Os 10 µl de eluato foram transferidos com pipeta para novos tubos de 1,5 mL. Os agregados de partículas magnéticas foram desprezados.

Com o intuito de buscar uma representatividade homogênea de todas as amostras durante o sequenciamento e diminuir vieses, os volumes das bibliotecas foram ajustados proporcionalmente à concentração de DNA, antes de serem incorporados ao *pool*. Para isso foi realizada a quantificação da biblioteca por fluorimetria com o fluorímetro Quantus, da Promega, seguindo o mesmo protocolo usado anteriormente, no item 3.3.

Para o gene 16S rRNA a quantidade total de DNA do *pool* deve estar na faixa entre 50 e 100 ng, conforme recomendado pela ONT no manual do kit de sequenciamento. Foram adicionados volumes de 0,6 µl e 0,4 µl, respectivamente, das amostras de *S. enterica* e *P. vulgaris* em um tubo de 1,5 mL.

Por fim, segue-se a etapa de adição dos adaptadores, incluídos no kit 16S *Barcoding* Kit 1-24, tendo sido adicionado 1 µl de *Rapid Adapter* (RAP) ao tubo com *pool* de amostras, homogeneizado por agitação manual e incubado à temperatura ambiente, por 5 minutos.

Os reagentes do *Sequencing Buffer* (SQB), *Loading Beads* (LB), *Flush Tether* (FLT) e *Flush Buffer* (FB) foram descongelados à temperatura ambiente. Depois de descongelados, os tubos foram agitados em *vortex* e centrifugados (*spin down*). Com uma micropipeta, foram transferidos 30 µl do FLT para o tubo do FB, e agitado com *vortex*, para preparar o *priming mix* da célula de fluxo.

Uma célula de fluxo R9.4.1 foi inserida no MinION, e este foi ligado ao computador, para verificar a quantidade de poros disponíveis. A checagem da célula de fluxo foi feita com o *software* MinKNOW v.5.1.8, da ONT, seguindo-se as instruções na tela.

A *priming port* foi aberta, girando-se a cobertura no sentido horário. Para remover bolhas de ar da célula de fluxo, foi aspirado um volume de 20 a 30 µl do tampão estoque, girando o embolo da pipeta até que fosse possível visualizar o tampão na parte inferior da ponteira. Foram transferidos, cuidadosamente, 800 µl do *priming mix* para a célula de fluxo pela *priming port*, evitando-se a formação de bolhas de ar. Após a transferência do *priming mix*, a célula de fluxo foi mantida em repouso, por 5 minutos. A mistura para carregamento da biblioteca foi preparada, em um novo tubo, no qual foram adicionados 34 µl de SQB, 25,5 µl de LB (imediatamente após agitação com *vortex*), 4,5 µl de água sem nuclease, e 11 µl do *pool* de biblioteca de DNA. A cobertura da porta *spot on* foi aberta para expor a entrada. A seguir, foram adicionados 200 µl de *priming mix* à *priming port*. A mistura de carregamento foi gentilmente homogeneizada por pipetagem, repetidamente, no próprio tubo. Foram adicionados 75 µl da mistura de carregamento, gota a gota, pela porta *spot on*. A cobertura da porta *spot on* foi recolocada, a *priming port* foi fechada, e o MinION foi fechado.

A corrida de sequenciamento foi iniciada com o *software* MinKNOW v.5.1.8, usando-se um computador tipo *notebook* modelo Positivo XR9430 (processador Intel Core i7 4510U de 3,10GHz, memória 8GB), usando os parâmetros presentes no Tabela 8.

Tabela 8: Parâmetros para a corrida de sequenciamento inseridos no software MinKNOW
Fonte: o autor

Qualidade e <i>Barcoding</i>	
Mínimo escore de qualidade	10
<i>Barcoding</i>	Informação gerada pelo MinKNOW
Filtro comprimento mínimo das <i>reads</i>	1200
Filtro comprimento máximo das <i>reads</i>	1800

3.5 Análise dos Dados do Sequenciamento

3.5.1 *Basecall*, demultiplexação e *trimming*

Os arquivos *fast5*, com os dados brutos do sequenciamento provenientes da corrida, foram processados usando-se o *software* Guppy GPU v6.3.2 para Linux, da ONT, em um computador com processador Ryzen 5 5600x, de 3,7 GHz, da AMD, com 16 GB de memória RAM, e uma GPU GeForce RTX 3070, com 8 GB de memória de vídeo, rodando o sistema operacional Debian 11, versão 64-Bits. O *basecall* foi realizado com o modelo “*Super Accurate*” (SUP), com as opções de *trimming* dos *barcodes* e adaptadores selecionadas.

3.5.2 Mapeamento e alinhamento com o software EPI2ME

A análise dos dados foi feita utilizando-se o serviço de computação em nuvem usando o *software* EPI2ME Desktop Agent v.3.5.7, da ONT. Após a etapa de *basecall*, demultiplexação e *trimming*, os arquivos resultantes *fastq* foram transferidos, usando-se o mesmo *notebook* da etapa de sequenciamento. As análises foram feitas com o *workflow* Fastq 16S, revisão 20220107, que executa o alinhamento local, empregando o BLAST (versão não divulgada pela ONT), da NCBI (*Bethesda, Maryland, EUA*), com os parâmetros do Tabela 9.

Tabela 9: Parâmetros para a análise de dados inseridos no software MinKNOW
Fonte: o autor

16S Microbial	
Filtro Blast <i>E-value</i>	0,01
Cobertura mínima	30%
Identidade mínima	70%
Máximo número de sequências alvo	3

3.5.3 Mapeamento e alinhamento com o banco de dados local

Os arquivos no formato FASTQ obtidos da etapa de *basecall* dos *barcodes* 02 e 03 foram alinhados com as sequências de referência do banco de dados 16S rRNA de agentes de bioterrorismo, usando-se o *software* Minimap2 v2.26 r1175 (Li, 2018), em um computador com processador Ryzen 9 5900x, de 3,7 GHz, da AMD, com 128 GB de memória RAM, com o sistema operacional Linux Mint Cinnamon, versão 64-Bits. Os melhores alinhamentos foram filtrados usando-se o *software* Samtools v.1.18 (Danecek *et al.*, 2021), que também foi usado para gerar as sequências consenso.

3.5.4 Construção da árvore filogenética

Foi construída uma árvore filogenética pelo método *Neighbor-Joining*, usando-se o *software* Mega v.10.2.6 (Kumar *et al.*, 2018), conforme as instruções (Newman; Duffus; Lee, 2016). As sequências consenso obtidas na análise com o bando de dados 16S rRNA de agentes de bioterrorismo e as sequências do banco de dados construído foram alinhadas, por alinhamento global com o algoritmo Clustal W (Thompson; Higgins; Gibson, 1994). Como teste de suporte foram calculadas 1000 repetições do *bootstrap*, e para a substituição nucleotídica foi empregado o algoritmo Jukes-Cantor.

3.5.5 Comparação das análises com banco de dados local e remota

Foram medidos os tempos real e de CPU para realizar, em triplicata, os alinhamentos com o *software* Minimap2 v2.26 r1175 (Li, 2018) para as sequências dos *barcodes* 02 e 03, usando-se como bancos de dados de referência o 16S RefSeq e o banco de dados 16S rRNA bt, construído para este trabalho.

4 RESULTADOS

4.1 Sequenciamento do Gene 16S rRNA e Análise dos Dados no Sistema ONT

A extração automatizada das amostras, obtidas de culturas puras de cepas de *S. enterica* ATCC 14028 e *P. vulgaris* ATCC 8427 resultou em quantidades de DNA genômico de 16 ng/ μ L e 23 ng/ μ L, respectivamente. Após a normalização e corrida de sequenciamento foram geradas 544.553 *reads*, totalizando 773,6 Megabases. As *reads* foram analisadas com o *workflow* Fastq 16S, no *software* EPI2ME, por serviço de computação em nuvem. Do total de *reads* analisadas, 462.885, correspondendo a 85% de todas as *reads* geradas, foram classificadas com sucesso pelo *workflow* proposto. O restante das *reads* não foi analisado porque foram rejeitadas pelos filtros de qualidade mínima (> 10) e de tamanho máximo e mínimo (> 1.200 e < 1.800 bp). Em relação a identidade encontrada, 190.710 (41,2% das *reads* classificadas), alinharam-se com o genoma de referência de *S. enterica*, e 233.125 *reads*, (50,4%) com *P. vulgaris*. Outras 39.050 *reads* (8,4%), alinharam-se com outras espécies (Figura 5). Todas as espécies classificadas com abundância inferior a 1% foram desconsideradas.

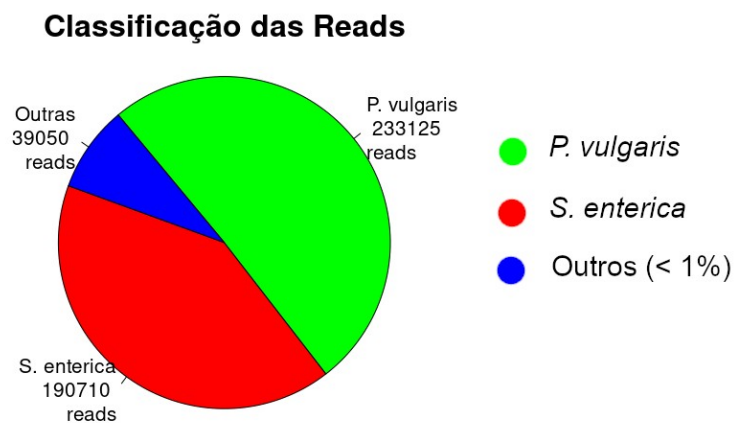


Figura 5: Gráfico da classificação das *reads* por espécie
Fonte: o autor

Em ambas as espécies, que tiveram abundância acima de 1% do número de *reads*, a média da precisão dos alinhamentos com os respectivos genomas de referência (identidade) foi de 93%. A Figura 6 apresenta o histograma da distribuição das *reads* por percentual de precisão.

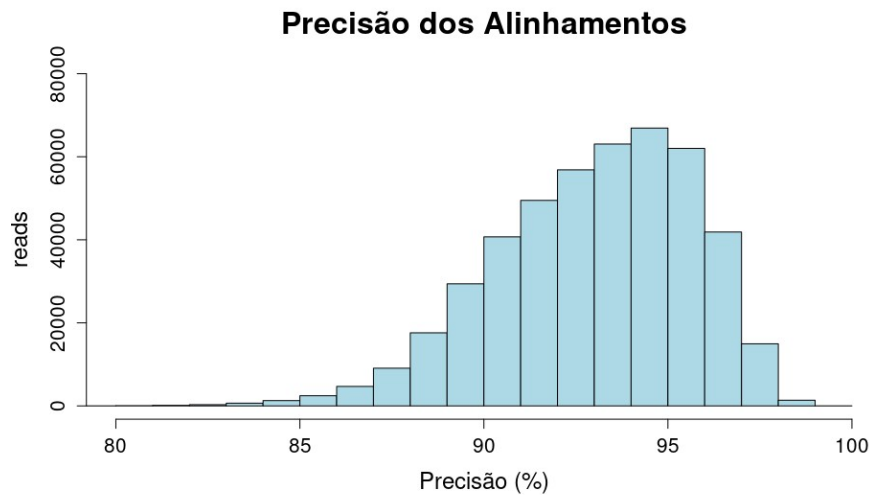


Figura 6: Histograma da precisão média dos alinhamentos
Fonte: o autor

As 23.303 *reads*, com tamanhos menores que 1.200 bp ou maiores que 1.800 bp, foram excluídas da análise pelo filtro estabelecido nos parâmetros da corrida de sequenciamento. O tamanho médio das *reads* foi de 1.420 bp. A distribuição das *reads* por tamanho pode ser vista na Figura 7.

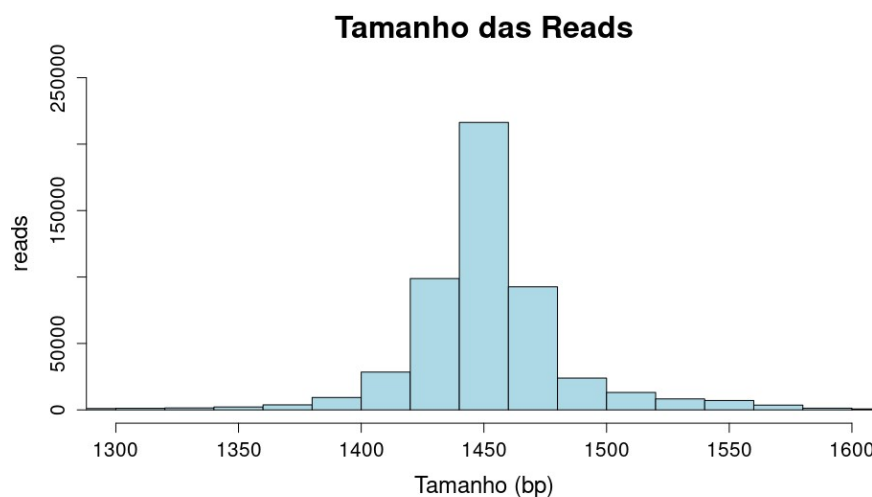


Figura 7: Histograma do tamanho das *reads*
 Fonte: o autor

Um total de 55.357 *reads*, com escore médio de qualidade inferior a 10 (escala *Phred*), foram excluídas da análise, pelo filtro estabelecido nos parâmetros da corrida de sequenciamento. A média do escore de qualidade das *reads* foi de 12,16. A distribuição das *reads* segundo o escore de qualidade pode ser vista na Figura 8.

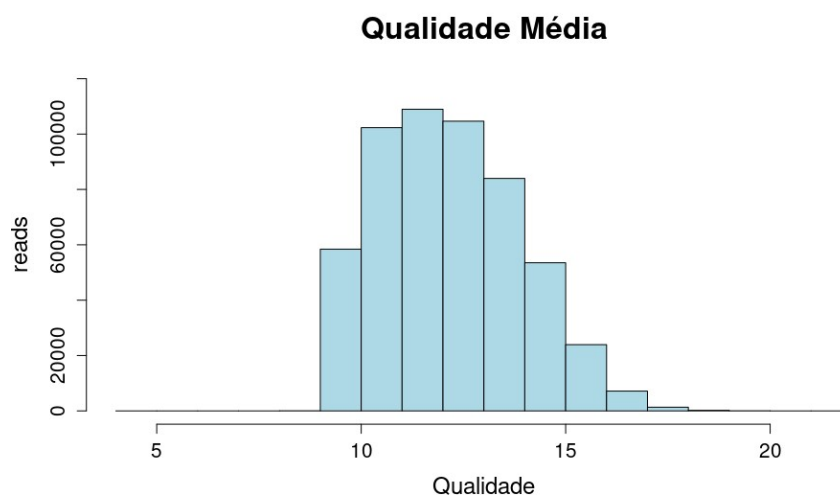


Figura 8: Histograma de qualidade das *reads*.
 Fonte: o autor

4.2 Comparação das Análises dos Dados de Sequenciamento

O tempo total da análise remota realizada com o software EPI2ME de 14 horas e 04 minutos, para um total de 1,8 Gigabases, foi proporcionalmente ajustado para o volume dados correspondente aos *barcodes* 02 e 03, de 773,6 Megabases, que equivalem a 6 horas e 02 minutos, considerando uma taxa de processamento constante.

Para comparação entre o desempenho do alinhamento dos dados do sequenciamento entre os bancos de dados 16S rRNA bt e 16S RefSeq, foram executadas três repetições com o *software* Minimap2 v2.26 r1175.

4.3 Construção da Árvore Filogenética

A árvore filogenética construída demonstrou a relação (distância) das sequências consenso, obtidas a partir dos dados experimentais, com as sequências de referência depositadas no banco de dados do NCBI. Os *barcodes* agrupados com os respectivos clados podem ser vistos em detalhe na Figura 9.

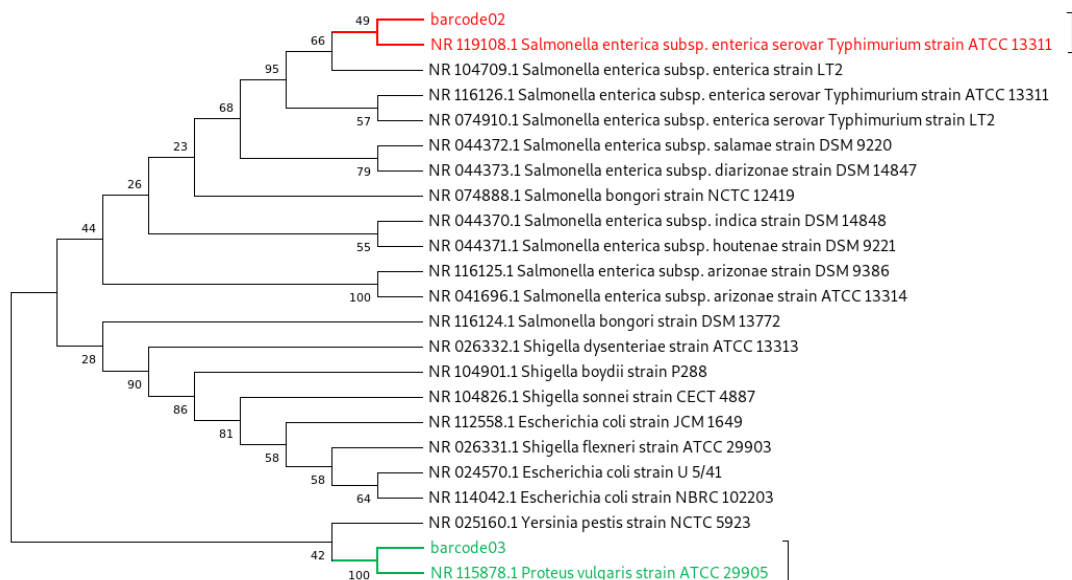


Figura 9: Detalhe da árvore filogenética construída com o software Mega. As sequências consenso barcode02 e barcode03 foram agrupadas junto dos clados de *S. enterica* e *P. vulgaris*, respectivamente.

Fonte: o autor

A sequência consenso obtida para o *barcode* 02 mostrou-se filogeneticamente semelhante às sequências da cepa de *S. enterica* subespécie *enterica* LT2 e a da cepa *S. enterica* subespécie *enterica* sorovariante *Typhimurium* ATCC 13311. Já a sequência do *barcode* 03 mostrou-se filogeneticamente semelhante à sequência da cepa *P. vulgaris* ATCC 29905.

A árvore filogenética completa, exibida em formato circular, com as sequências dos genes 16S rRNA de todas as 95 espécies de bactérias reconhecidas com agentes de bioterrorismo, que foram utilizadas na construção do banco de dados 16S rRNA bt, mostrando as relações entre elas, pode ser vista na Figura 10.

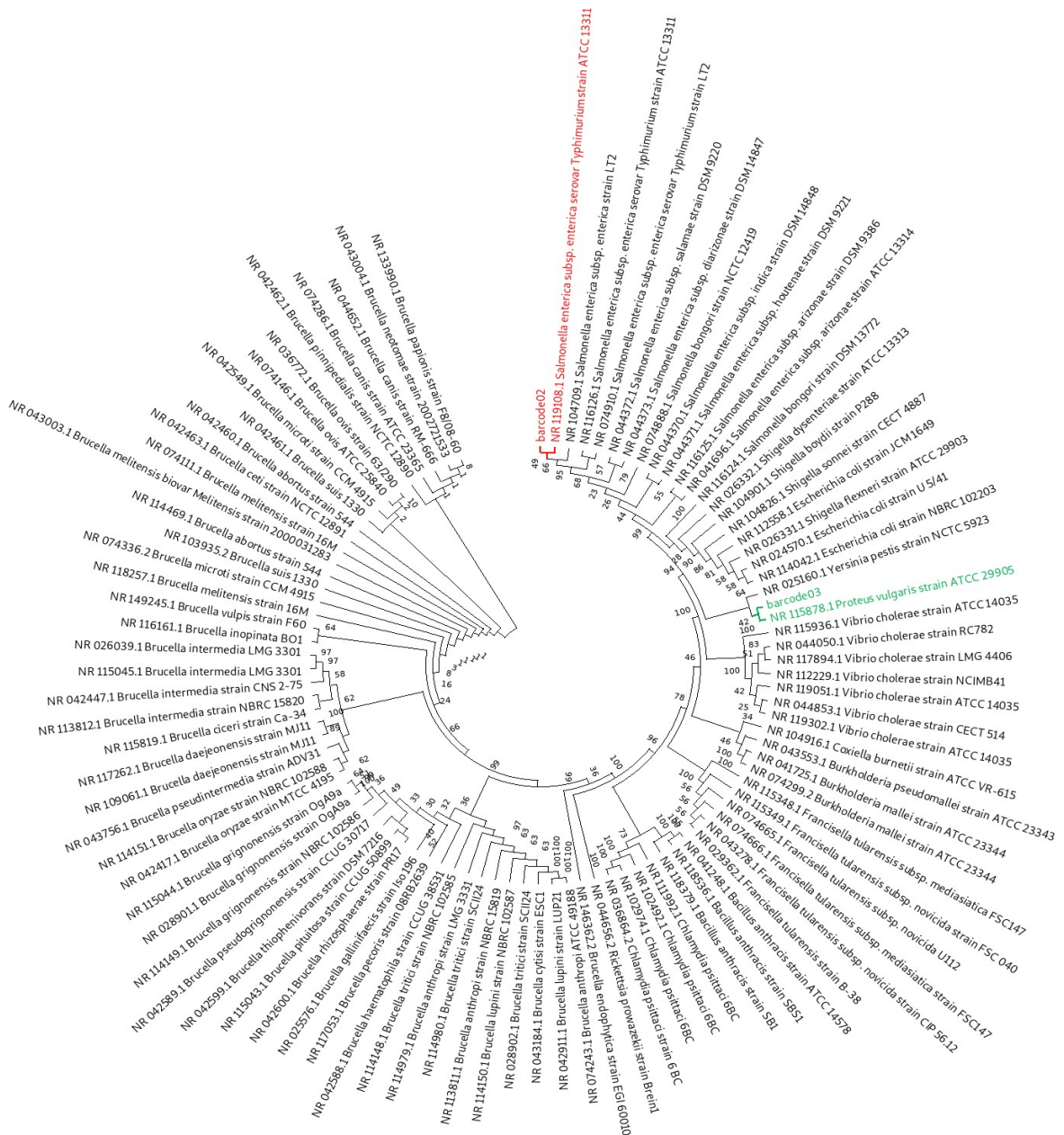


Figura 10: Árvore filogenética completa, exibida em formato circular, com as sequências dos genes 16S rRNA de todas as 95 espécies de bactérias reconhecidas com agentes de bioterrorismo, que foram utilizadas na construção do banco de dados 16S rRNA bt. Em verde foram destacadas as sequências do *barcode* 03 (gerada neste trabalho) e a referência de *P. vulgaris* (NCBI), e em vermelho as do *barcode* 02 e referências de *S. enterica*.
Fonte: o autor

4.4 Construção do Banco de Dados 16S rRNA dos Agentes de Bioterrorismo

Para a construção do banco de dados 16S rRNA bt (“bt”, de bioterrorismo), as sequências de DNA dos genes 16S rRNA das bactérias classificadas pelo CDC como agentes de bioterrorismo foram selecionadas a partir dos respectivos TaxID. As sequências foram extraídas do banco de dados global 16S RefSeq, do NCBI. A sequência de *P. vulgaris* também foi selecionada por ser uma espécie não classificada como agente de bioterrorismo, servindo, portanto, como um controle. Do total de 26.910 sequências foram selecionadas 95 para compor o banco de dados local, com tamanho de 146,9 kB.

As sequências usadas na construção do banco de dados 16S rRNA bt foram listadas no Tabela 10, com os respectivos números de acesso, TaxID e tamanho.

Tabela 10: Sequências que compõem o banco de dados local 16S rRNA de bactérias listadas como agentes de bioterrorismo. Os descritores são o número de acesso NCBI, TaxID, nome científico e tamanho.

Fonte: o autor

Número de acesso NCBI	TaxID	Nome científico	Tamanho (bp)
NR_118379.1	1392	<i>Bacillus anthracis</i>	1311
NR_041248.1	1392	<i>Bacillus anthracis</i>	1306
NR_118536.1	1392	<i>Bacillus anthracis</i>	1311
NR_042460.1	235	<i>Brucella abortus</i>	1430
NR_114469.1	235	<i>Brucella abortus</i>	1412
NR_114979.1	529	<i>Brucella anthropi</i>	1388
NR_113811.1	529	<i>Brucella anthropi</i>	1408
NR_074243.1	439375	<i>Brucella anthropi</i> ATCC 49188	1476
NR_074286.1	36855	<i>Brucella canis</i>	1485
NR_044652.1	36855	<i>Brucella canis</i>	1418
NR_042463.1	120577	<i>Brucella ceti</i>	1430
NR_115819.1	391287	<i>Brucella ciceri</i>	1492
NR_043184.1	407152	<i>Brucella cytisi</i>	1476
NR_109061.1	659015	<i>Brucella daejeonensis</i>	1389
NR_117262.1	659015	<i>Brucella daejeonensis</i>	1431

Número de acesso NCBI	TaxID	Nome científico	Tamanho (bp)
NR_146362.2	1963359	<i>Brucella endophytica</i>	1485
NR_025576.1	215590	<i>Brucella gallinifaecis</i>	1453
NR_028901.1	94627	<i>Brucella grignonensis</i>	1472
NR_115044.1	94627	<i>Brucella grignonensis</i>	1387
NR_114149.1	94627	<i>Brucella grignonensis</i>	1407
NR_042588.1	419474	<i>Brucella haematophila</i>	1388
NR_116161.1	470735	<i>Brucella inopinata BO1</i>	1412
NR_042447.1	94625	<i>Brucella intermedia</i>	1390
NR_113812.1	94625	<i>Brucella intermedia</i>	1410
NR_115045.1	641118	<i>Brucella intermedia LMG 3301</i>	1390
NR_026039.1	641118	<i>Brucella intermedia LMG 3301</i>	1454
NR_114150.1	255457	<i>Brucella lupini</i>	1408
NR_042911.1	255457	<i>Brucella lupini</i>	1477
NR_074111.1	29459	<i>Brucella melitensis</i>	1432
NR_118257.1	29459	<i>Brucella melitensis</i>	1400
NR_043003.1	274058	<i>Brucella melitensis biovar Melitensis</i>	1412
NR_042549.1	444163	<i>Brucella microti</i>	1390
NR_074336.2	444163	<i>Brucella microti</i>	1485
NR_043004.1	29460	<i>Brucella neotomae</i>	1412
NR_042417.1	335286	<i>Brucella oryzae</i>	1477
NR_114151.1	335286	<i>Brucella oryzae</i>	1412
NR_036772.1	236	<i>Brucella ovis</i>	1412
NR_074146.1	444178	<i>Brucella ovis ATCC 25840</i>	1454
NR_133990.1	1462582	<i>Brucella papionis</i>	1407
NR_117053.1	867683	<i>Brucella pecoris</i>	1386
NR_042462.1	120576	<i>Brucella pinnipedialis</i>	1430
NR_115043.1	571256	<i>Brucella pituitosa</i>	1429
NR_043756.1	370111	<i>Brucella pseudintermedia</i>	1396
NR_042589.1	419475	<i>Brucella pseudogrignonensis</i>	1387
NR_042600.1	571254	<i>Brucella rhizosphaerae</i>	1387
NR_042461.1	204722	<i>Brucella suis 1330</i>	1430
NR_103935.2	204722	<i>Brucella suis 1330</i>	1485
NR_042599.1	571255	<i>Brucella thiophenivorans</i>	1387
NR_114980.1	94626	<i>Brucella tritici</i>	1388

Número de acesso NCBI	TaxID	Nome científico	Tamanho (bp)
NR_114148.1	94626	<i>Brucella tritici</i>	1408
NR_028902.1	94626	<i>Brucella tritici</i>	1398
NR_149245.1	981386	<i>Brucella vulpis</i>	1425
NR_041725.1	13373	<i>Burkholderia mallei</i>	1488
NR_074299.2	13373	<i>Burkholderia mallei</i>	1528
NR_043553.1	28450	<i>Burkholderia pseudomallei</i>	1488
NR_036864.2	83554	<i>Chlamydia psittaci</i>	1552
NR_111992.1	331636	<i>Chlamydia psittaci</i> 6BC	1507
NR_102492.1	331636	<i>Chlamydia psittaci</i> 6BC	1538
NR_102974.1	331636	<i>Chlamydia psittaci</i> 6BC	1538
NR_104916.1	777	<i>Coxiella burnetii</i>	1465
NR_024570.1	562	<i>Escherichia coli</i>	1450
NR_112558.1	562	<i>Escherichia coli</i>	1110
NR_114042.1	562	<i>Escherichia coli</i>	1467
NR_029362.1	263	<i>Francisella tularensis</i>	1521
NR_074666.1	135248	<i>Francisella tularensis</i> subsp. <i>mediasiatica</i>	1524
NR_115348.1	441952	<i>Francisella tularensis</i> subsp. <i>mediasiatica</i> FSC147	1450
NR_043278.1	264	<i>Francisella tularensis</i> subsp. <i>novicida</i>	1520
NR_115349.1	264	<i>Francisella tularensis</i> subsp. <i>novicida</i>	1450
NR_074665.1	401614	<i>Francisella tularensis</i> subsp. <i>novicida</i> U112	1524
NR_115878.1	585	<i>Proteus vulgaris</i>	1478
NR_044656.2	782	<i>Rickettsia prowazekii</i>	1498
NR_116124.1	54736	<i>Salmonella bongori</i>	1443
NR_074888.1	54736	<i>Salmonella bongori</i>	1542
NR_116125.1	59203	<i>Salmonella enterica</i> subsp. <i>arizonae</i>	1438
NR_041696.1	59203	<i>Salmonella enterica</i> subsp. <i>arizonae</i>	1491
NR_044373.1	59204	<i>Salmonella enterica</i> subsp. <i>diarizonae</i>	1368
NR_104709.1	59201	<i>Salmonella enterica</i> subsp. <i>enterica</i>	1467
NR_116126.1	90371	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i>	1262
NR_074910.1	90371	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i>	1544
NR_119108.1	90371	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i>	1539
NR_044371.1	59205	<i>Salmonella enterica</i> subsp. <i>houtenae</i>	1437
NR_044370.1	59207	<i>Salmonella enterica</i> subsp. <i>indica</i>	1414
NR_044372.1	59202	<i>Salmonella enterica</i> subsp. <i>salamae</i>	1448

Número de acesso NCBI	TaxID	Nome científico	Tamanho (bp)
NR_104901.1	621	<i>Shigella boydii</i>	1515
NR_026332.1	622	<i>Shigella dysenteriae</i>	1487
NR_026331.1	623	<i>Shigella flexneri</i>	1488
NR_104826.1	624	<i>Shigella sonnei</i>	1530
NR_044050.1	666	<i>Vibrio cholerae</i>	1465
NR_115936.1	666	<i>Vibrio cholerae</i>	1465
NR_117894.1	666	<i>Vibrio cholerae</i>	1461
NR_112229.1	666	<i>Vibrio cholerae</i>	1465
NR_044853.1	666	<i>Vibrio cholerae</i>	1538
NR_119051.1	666	<i>Vibrio cholerae</i>	1463
NR_119302.1	666	<i>Vibrio cholerae</i>	1452
NR_025160.1	632	<i>Yersinia pestis</i>	1461

O banco de dados 16S rRNA bt foi utilizado como referência na etapa de alinhamento das *reads* produzidas no sequenciamento das amostras das culturas bacterianas.

Para as sequências correspondentes aos *barcodes* 02 e 03, foram executadas três repetições do alinhamento com o banco de dados construído neste estudo (16S rRNA bt) e também com o banco de dados 16S RefSeq do NCBI, e anotados o tempo total dispendido (tempo real) e o tempo total de processamento (tempo de CPU). Os resultados obtidos foram registrados na Tabela 11.

Tabela 11: Comparação entre o tempo necessário para o alinhamento das *reads* usando o Minimap2 com o banco de dados (16S rRNA bt) e o banco de dados 16S RefSeq do NCBI. Desvio padrão calculado usando-se uma planilha do LibreOffice Calc v.7.0.4.2 para Linux
 Fonte: o autor

Banco de dados	Barcode 02				Barcode 03			
	16S rRNA bt		16S RefSeq		16S rRNA bt		16S RefSeq	
	Real	CPU	Real	CPU	Real	CPU	Real	CPU
1	57,11	1127,69	905,37	19752,87	50,85	983,61	1167,00	26146,62
2	57,21	1134,18	915,39	19974,14	50,75	992,01	1174,40	26313,00
3	57,28	1141,20	910,64	19865,44	51,37	1000,63	1173,34	26303,70
Média	57,20	1134,36	910,47	19864,15	50,99	992,08	1171,58	26254,44
Desvio padrão	0,08	6,75	5,02	110,64	0,33	8,51	4,00	93,49

Usando-se o banco de dados 16S rRNA bt para os alinhamentos das *reads* dos *barcodes* 02 e 03 foram necessários, em média, $57,20 \pm 0,08$ segundos e $50,99 \pm 0,33$ segundos, respectivamente, enquanto os alinhamentos dos mesmos *barcodes*, quando realizados com o banco de dados 16S RefSeq, do NCBI, levaram, respectivamente, $910,47 \pm 5,02$ segundos (aproximadamente 15 minutos e 11 segundos) e $1171,58 \pm 4,00$ segundos (aproximadamente 19 minutos e 30 segundos). O tamanho reduzido do banco de dados 16S rRNA possibilitou a redução do tempo de execução dos alinhamentos em aproximadamente 16 vezes e 23 vezes.

O banco de dados 16S rRNA bt foi armazenado no *notebook* do Centro de Estudos de Biodefesa, modelo Alienware m15 R7 e no servidor do IBEx, modelo Dell PowerEdge R750.

5 DISCUSSÃO

No presente trabalho, a bactéria *S. entérica* foi selecionada, sendo a amostra obtida a partir de uma cultura da cepa ATCC 14028. Todas as espécies do gênero *Salmonella* são considerados agentes de bioterrorismo da categoria B, segundo a classificação do CDC. O banco de dados chamado 16S rRNA bt foi construído a partir de sequências de genes 16S rRNA depositadas no banco de dados 16S RefSeq, do NCBI. O tempo de análise dos dados de sequenciamento massivo-paralelo de *S. enterica* ATCC 14028 usando o banco de dados local 16S rRNA bt foi comparado com a análise usando o banco de dados 16S RefSeq completo, e com a análise pelo serviço de computação em nuvem com o software EPI2ME, da Nanopore.

Na construção do banco de dados local 16S rRNA bt, foram usadas somente 95 sequências de genes 16S rRNA dos agentes de bioterrorismo, assim compreendidas as espécies de bactérias listadas nas categorias A e B da classificação do CDC, que foram previamente curadas para o 16S RefSeq, do NCBI. Originalmente, mais de 26 mil sequências de genes 16S rRNA de bactérias e arqueias foram agregados ao 16S RefSeq, na versão mais recente, até o momento da escrita deste texto. O tamanho reduzido do banco de dados 16S rRNA bt proporcionou uma redução do tempo médio dispendido para o alinhamento das *reads* da ordem de 20 vezes, de $910,47 \pm 5,02$ segundos para $57,02 \pm 0,08$ segundos, e de $1171,58 \pm 4,00$ segundos para $50,99 \pm 0,33$ segundos, respectivamente para os *barcodes* 02 e 03. Como referência, o tempo total da análise com o *software* EPI2ME foi de 21.763,95 segundos (aproximadamente 6h e 04 minutos). A análise remota não permite estimar separadamente o tempo efetivo de processamento dos dados e o tempo de transmissão dos mesmos pela conexão de internet, para que fosse possível fazer uma comparação direta com o processamento local. A Nanopore não tornou público o código fonte do *pipeline* do EPI2ME, mas a documentação deixa evidente que a análise se baseia no uso do BLAST (Workflow: 16S, [s. d.]). Segundo o guia de utilização da interface de programação de aplicativo do BLAST (*Application Programming Interface*, API), que é a camada de *software*

que gerencia a interação entre dois ou mais programas de computador, o NCBI esclarece que “para assegurar a disponibilidade do serviço poderá limitar as buscas de usuários de alta demanda”, que “usuários que submeterem de mais de 100 buscas em um período de 24 horas, serão direcionados para filas mais lentas e, em casos extremos, bloqueados”, e recomenda expressamente que a versão local (Stand-alone BLAST+) seja utilizada quando for necessário um grande número de buscas (Overview — BLASTHelp documentation, [s. d.]). Ainda que a rapidez da análise possa ser do interesse dos usuários dessas ferramentas de análise remota, elas não foram desenvolvidas expressamente com esse objetivo. Como esperado, as análises locais mostraram-se mais rápidas que as remotas.

A construção de bancos de dados 16S rRNA específicos para determinados nichos ecológicos como, por exemplo, o DAIRYdb, de microbiomas associados à produção de laticínios (Meola *et al.*, 2019), o FreshTrain, de comunidades microbianas de ecossistemas aquíferos (Rohwer *et al.*, 2018), o HITdb, de microbiota intestinal humana (Ritari *et al.*, 2015), também foi baseada em sequências depositadas em bancos de dados públicos.

Entre as possíveis estratégias de sequenciamento, optou-se por fazer o sequenciamento alvo-específico do gene 16S rRNA, que codifica um componente da subunidade 30S do ribossomo. As *long reads*, obtidas pelo sequenciamento por nanoporos, permitem que se sequencie toda a extensão do gene 16S rRNA, o que tem como vantagem a maior resolução taxonômica e filogenética para a identificação bacteriana, porque todas as regiões do gene são analisadas (Santos *et al.*, 2020).

A corrida de sequenciamento produziu um total de 773,6 Megabases de dados, correspondendo a 544.553 *reads* de sequenciamento das amostras com os *barcodes* 02 e 03 (média de 272.000, por amostra), quantidade de dados que se mostrou suficiente para a identificação bacteriana pretendida. Na literatura, foram reportados volumes de dados de 1.695.436 (média de 66.000, por amostra), para a análise de amostras de solos (Stevens *et al.*, 2023), e 4,6 Gigabases, equivalentes a 3.168.160 de *reads* (média de 33.000, por amostra), para o estudo de composição da microbiota intestinal de crianças de 3 meses de idade (Angell *et al.*, 2020). O

sequenciamento massivo-paralelo ainda é uma tecnologia relativamente recente e ainda não foram estabelecidos padrões, em termos de número mínimo de *reads* que constitua uma identificação positiva (Wołkowicz, 2018).

O percentual de *reads* aproveitadas após a filtragem por escore de qualidade (> 10) e por tamanho (> 1.200 e < 1.800 bp) foi de 85%. A finalidade dos filtros é remover *reads* de baixa qualidade e muito curtas, antes da análise. Os parâmetros de filtragem revelaram-se apropriados à análise pretendida, porque a maioria das *reads* foi preservada para análise. Aplicando-se o filtro de escore de qualidade > 7, combinado com filtro de tamanhos entre 1.300 e 1.950, em um estudo de diversidade da flora intestinal humana, verificou-se aproveitamentos de 71,8 a 88,2% do total de *reads* (Matsuo *et al.*, 2021).

As *reads* tiveram um tamanho médio das de 1.420 pares de bases (bp), após a remoção dos adaptadores (*trimming*), realizada pelo *software* Guppy GPU v6.3.2, durante a etapa do *basecall*. Esta extensão seria suficiente para abarcar as nove regiões hipervariáveis do gene 16S rRNA, tomando-se por base as posições do gene de *E. coli*, com 1.398 bp entre o início de V1, na posição 68, e o fim de V9, na posição 1.466 (Church *et al.*, 2020). Foram encontrados, na literatura, tamanhos médios de 1.473 bp, em estudo de diversidade da flora intestinal humana (Matsuo *et al.*, 2021), de 1.450 bp, do monitoramento ambiental da Estação Espacial Internacional (*International Space Station*, ISS) (Burton *et al.*, 2020), e de 1.137 bp, em estudo da flora intestinal de crianças (Angell *et al.*, 2020).

Nas duas amostras, a precisão média da *reads* observada para 462.885 *reads* classificadas, 190.710 como *S. enterica*, e 233.125 como *P. vulgaris*, foi de 93,3%. Para *S. enterica* e *P. vulgaris* respectivamente, foram obtidas 7.279 e 9.301 *reads* com precisão média dos alinhamentos superior a 97%. Em amostras de água de poço, foram identificadas as espécies *Aeromans hydrophila* e *Legionella pneumophila*, com precisões médias que variaram entre 82,3 e 93,2 % (Mimura *et al.*, 2023), enquanto as espécies *Enterococcus hirae* e *Enterococcus faecalis* foram identificadas com 96,4% e 97,5% de precisão, em experimento de sensibilidade a antimicrobianos (Kawai *et al.*, 2022). Foi demonstrada a resolução taxonômica em nível de espécie mesmo com percentuais de identidades de 85%, devido ao maior

poder de discriminação possível com *long reads* (Cha *et al.*, 2023). A média de *reads* por amostra foi de 230.000. Em uma análise do perfil microbiano da flora intestinal de neonatos, os autores concluíram que a diferença de diversidade, em todos os níveis taxonômicos, com 100.000 *reads* por amostra e 30.000 *reads* por amostra era inferior a 0,1% (Cha *et al.*, 2023).

O escore médio de qualidade das *reads* foi de 12,16 na escala *Phred*, o que equivale a 94% de precisão do *basecall* ou, aproximadamente, uma base errada a cada 15 bases. Escores na faixa de 8,2 até 10,1 foram reportados na análise da microbiota intestinal de roedores (Jahan *et al.*, 2021). O uso de um nível de corte mais elevado (> 10) pode explicar a obtenção de um escore de qualidade mais elevado neste trabalho.

A cobertura média por base resultante foi de 544,8x, calculada com base no total de bases sequenciadas e tamanho médio das *reads*. Como as amostras eram provenientes de culturas puras, não foi possível avaliar se a profundidade de cobertura seria suficiente para diferenciar espécies muito proximamente relacionadas como, por exemplo, *B. anthracis* e *Bacillus cereus* (Braun *et al.*, 2022), e algumas espécies de *Salmonella* e *Shigella* (Kawai *et al.*, 2022).

Analisando-se filogeneticamente o relacionamento das sequências consenso das *reads* dos *barcodes* 02 e 03 com as sequências usadas na construção do banco de dados 16S rRNA bt, observou-se que a sequência referente ao *barcode* 02 foi agrupada no mesmo clado das subespécies de *S. enterica*, enquanto a sequência do *barcode* 03 foi agrupada com a única sequência de referência de *P. vulgaris* ATCC 29905 contida no banco de dados 16S rRNA bt (Figura 10).

O presente trabalho não explorou a identificação de agentes de bioterrorismo em amostras reais, isto é, amostras simuladas em exercícios de campo, em condições análogas àquelas de uma liberação intencional ou acidental de um agente de bioterrorismo. Em uma amostra simulada estariam presentes diversos contaminantes ambientais (solo, matéria orgânica, etc), inclusive inibidores de PCR, além de toda uma comunidade microbiana de fundo.

Também não foi possível determinar a possível interferência que alteração do protocolo utilizado na preparação das amostras nos resultados obtidos, porque as

bactérias usadas na preparação das amostras não haviam sido sequenciadas previamente e não se tratavam de cepas com sequências curadas conhecidas. Até o momento da elaboração do banco de dados não havia sequências de referência das cepas utilizadas depositadas no 16S RefSeq.

Como o estudo foi iniciado a partir das cepas cultivadas em meio de cultura, não foi feita a avaliação de possíveis influências das diferentes matrizes, técnicas de amostragem, métodos de extração, purificação e conservação do DNA, pares de *primers* e *pipelines* de análise nos resultados da análise, como já foi demonstrado (Abellan-Schneyder *et al.*, 2021; Pollock *et al.*, 2018).

Mais estudos são necessários para estabelecer padrões de confiabilidade e reprodutibilidade necessários às hipóteses previstas para o emprego em defesa biológica. Ainda não foram estabelecidos padrões mínimos para a identificação de microrganismos por sequenciamento de DNA para os parâmetros experimentais de bancada, como quantidade de biomassa, formas de preservação, método ideal de extração e forma de preparação de biblioteca, nem para o processamento de bioinformática posterior, como número, tamanho e cobertura das *reads*.

A identificação simultânea de agentes de bioterrorismo a partir de amostras com pequena abundância relativa com a microbiota de fundo e o desenvolvimento de um *pipeline* dedicado para a identificação desses agentes seriam desdobramentos importantes para a continuidade deste trabalho.

A Nanopore tem mantido um padrão de desenvolvimento acelerado e lançado atualizações periódicas de sua química de sequenciamento, que chegou à versão 14, e seria interessante determinar se as novas químicas de sequenciamento conseguiram aumentar a confiabilidade do sequenciamento por nanoporos.

CONCLUSÃO

Neste trabalho foi demonstrada a construção de um banco de dados local (16S rRNA bt), para servir como referência na identificação de bactérias consideradas potenciais agentes de bioterrorismo, baseado no sequenciamento massivo-paralelo do gene 16S rRNA, com tecnologia de nanoporos. Foi verificado que a utilização de um banco de dados formado exclusivamente com sequências dos microrganismos de interesse possibilitou realizar a análise dos dados de sequenciamento *offline*, sem depender de conexão com a internet, e a redução do tempo necessário para o alinhamento das *reads* de mais de 20 minutos para menos de 1 minuto.

As duas amostras, provenientes de culturas bacterianas das cepas *S. enterica* ATCC 14028 e *P. vulgaris* ATCC 8427, foram corretamente identificadas, até o nível taxonômico de espécie, após sequenciamento massivo-paralelo por nanoporos, seguido da análise de dados com o *software* EPI2ME. A análise, feita de maneira remota nos servidores da Nanopore, foi concluída em 06 horas e 04 minutos.

As amostras também foram corretamente identificadas, quando os mesmos dados de sequenciamento foram analisados localmente, utilizando os bancos de dados 16S rRNA bt, construído para este trabalho, e o 16S RefSeq, do NCBI.

O tempo dedicado ao alinhamento das *reads* com o 16S rRNA bt foi, respectivamente, de $57,02 \pm 0,08$ segundos e $50,99 \pm 0,33$ segundos, e $910,47 \pm 5,02$ segundos e de $1171,58 \pm 4,00$ segundos, com o 16S RefSeq. Comparado com o banco de dados 16S RefSeq, os alinhamentos com o 16S rRNA bt foram de 16 a 23 vezes mais rápidos.

Com bases nos resultados obtidos, foi possível reduzir o tempo gasto na identificação de agentes de bioterrorismo com a utilização dos bancos de dados locais, além de assegurar a manutenção da custódia restrita sobre os dados do experimento.

REFERÊNCIAS

16S REFSEQ RECORDS PROCESSING AND CURATION. [S. l.], [s. d.]. Disponível em: https://www.ncbi.nlm.nih.gov/refseq/targetedloci/16S_process/. Acesso em: 15 out. 2023.

ABELLAN-SCHNEYDER, Isabel *et al.* Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. **mSphere**, [s. l.], v. 6, n. 1, p. e01202-20, 2021.

ALTSCHUL, Stephen F. *et al.* Basic local alignment search tool. **Journal of Molecular Biology**, [s. l.], v. 215, n. 3, p. 403–410, 1990.

ANGELL, Inga Leena *et al.* De novo species identification using 16S rRNA gene nanopore sequencing. **PeerJ**, [s. l.], v. 8, p. e10029, 2020.

ARB: PROJECT. [S. l.], [s. d.]. Disponível em: <http://www.arb-home.de/home.html>. Acesso em: 17 nov. 2023.

ARTENSTEIN, Andrew W. 75 - Bioterrorism and Biodefense. *In*: COHEN, Jonathan; POWDERLY, William G.; OPAL, Steven M. (org.). **Infectious Diseases (Fourth Edition)**. [S. l.]: Elsevier, 2017. p. 670-679.e1. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780702062858000757>. Acesso em: 29 out. 2023.

BARRAS, V.; GREUB, G. History of biological warfare and bioterrorism. **Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases**, [s. l.], v. 20, n. 6, p. 497–502, 2014.

BIOLOGICAL WEAPONS CONVENTION – UNODA. *In*: [s. d.]. Disponível em: <https://disarmament.unoda.org/biological-weapons/>. Acesso em: 12 nov. 2023.

BLAST® COMMAND LINE APPLICATIONS USER MANUAL. [S. l.]: National Center for Biotechnology Information (US), 2008.

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**, [s. l.], v. 34, p. 100622, 2020.

BRASIL. **A Participação do Exército na Segurança dos Grandes Eventos - o Legado**. [S. l.: s. n.], 2018.

BRASIL. **Classificação de risco dos agentes biológicos**. Brasília, DF: Ministério da Saúde, 2023.

BRASIL. Construindo a política nacional de biossegurança e bioproteção: ações estratégicas da saúde. *In*: CONSTRUINDO A POLÍTICA NACIONAL DE BIOSSEGURANÇA E BIOPROTEÇÃO: AÇÕES ESTRATÉGICAS DA SAÚDE. [S. l.:

s. n.], 2019. p. 152–152. Disponível em:
https://bvsms.saude.gov.br/bvs/publicacoes/construindo_politica_nacional_biosseguranca_bioprotecao.pdf. Acesso em: 29 out. 2023.

BRASIL. **Guia para diagnóstico laboratorial em saúde pública : orientações para o sistema nacional de laboratórios de saúde pública**. Brasília, DF: Ministério da Saúde, 2022.

BRASIL. **PORTARIA Nº 204 EME, DE 14 DE DEZEMBRO DE 2012**. [S. l.], 2012. Disponível em:
http://www.sgex.eb.mil.br/sg8/006_outras_publicacoes/01_diretrizes/04_estado-maior_do_exercito/port_n_204_eme_14dez2012.html. Acesso em: 1 dez. 2023.

BRAUN, Peter *et al.* In-Depth Analysis of Bacillus anthracis 16S rRNA Genes and Transcripts Reveals Intra- and Intergenomic Diversity and Facilitates Anthrax Detection. **mSystems**, [s. l.], v. 7, n. 1, p. e01361-21, 2022.

BRUIJNS, Brigitte; TIGGELAAR, Roald; GARDENIERS, Han. Massively parallel sequencing techniques for forensics: A review. **ELECTROPHORESIS**, [s. l.], v. 39, n. 21, p. 2642–2654, 2018.

BURTON, Aaron S. *et al.* Off Earth Identification of Bacterial Populations Using 16S rDNA Nanopore Sequencing. **Genes**, [s. l.], v. 11, n. 1, p. 76, 2020.

CALABRIA, Donato *et al.* Recent Advancements in Enzyme-Based Lateral Flow Immunoassays. **Sensors**, [s. l.], v. 21, n. 10, p. 3358, 2021.

CANADA, Public Safety. **Chemical, Biological, Radiological, Nuclear and Explosives Resilience Strategy for Canada**. [S. l.], 2018. Disponível em:
<https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/rsln-c-strtg/index-en.aspx>. Acesso em: 18 out. 2023.

CAPORASO, J. Gregory *et al.* QIIME allows analysis of high-throughput community sequencing data. **Nature Methods**, [s. l.], v. 7, n. 5, p. 335–336, 2010.

CARUS, W Seth. A Short History of Biological Warfare: From Pre-History to the 21st Century. [s. l.], 2017.

CDC | BIOTERRORISM AGENTS/DISEASES (BY CATEGORY) | EMERGENCY PREPAREDNESS & RESPONSE. [S. l.], 2019. Disponível em:
<https://emergency.cdc.gov/agent/agentlist-category.asp>. Acesso em: 24 jun. 2023.

CHA, Teahyen *et al.* Gut microbiome profiling of neonates using Nanopore MinION and Illumina MiSeq sequencing. **Frontiers in Microbiology**, [s. l.], v. 14, 2023. Disponível em: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1148466>. Acesso em: 16 ago. 2023.

CHEN, Xin-Fei *et al.* Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry (MALDI-TOF MS) Analysis for the Identification of Pathogenic Microorganisms: A Review. **Microorganisms**, [s. l.], v. 9, n. 7, p. 1536, 2021.

CHURCH, Deirdre L. *et al.* Performance and Application of 16S rRNA Gene Cycle Sequencing for Routine Identification of Bacteria in the Clinical Microbiology Laboratory. **Clinical Microbiology Reviews**, [s. l.], v. 33, n. 4, p. e00053-19, 2020.

COLE, J. R. *et al.* The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. **Nucleic Acids Research**, [s. l.], v. 33, n. suppl_1, p. D294–D296, 2005.

DANECEK, Petr *et al.* Twelve years of SAMtools and BCFtools. **GigaScience**, [s. l.], v. 10, n. 2, p. giab008, 2021.

DIAS, Victor H. G. *et al.* Evaluation of 16S rRNA Hypervariable Regions for Bioweapon Species Detection by Massively Parallel Sequencing. **International Journal of Microbiology**, [s. l.], v. 2020, p. e8865520, 2020.

FEDERHEN, Scott. The NCBI Taxonomy database. **Nucleic Acids Research**, [s. l.], v. 40, n. D1, p. D136–D143, 2012.

FOX, G. E. *et al.* The Phylogeny of Prokaryotes. **Science**, [s. l.], v. 209, n. 4455, p. 457–463, 1980.

FRANCO-DUARTE, Ricardo *et al.* Advances in Chemical and Biological Methods to Identify Microorganisms—From Past to Present. **Microorganisms**, [s. l.], v. 7, n. 5, p. 130, 2019.

FRANK, Jeremy A. *et al.* Critical Evaluation of Two Primers Commonly Used for Amplification of Bacterial 16S rRNA Genes. **Applied and Environmental Microbiology**, [s. l.], v. 74, n. 8, p. 2461–2470, 2008.

FUKUDA, Kazumasa *et al.* Molecular Approaches to Studying Microbial Communities: Targeting the 16S Ribosomal RNA Gene. **Journal of UOEH**, [s. l.], v. 38, n. 3, p. 223–232, 2016.

GAO, US. Capitol Hill Anthrax Incident—EPA’s Cleanup Was Successful; Opportunities Exist to Enhance Contract Oversight. *In*: C. O. F. REPORT TO THE CHAIRMAN, US SENATE. [S. l.: s. n.], 2003.

GARGIS, Amy S. *et al.* Rapid Detection of Genetic Engineering, Structural Variation, and Antimicrobial Resistance Markers in Bacterial Biothreat Pathogens by Nanopore Sequencing. **Scientific Reports**, [s. l.], v. 9, n. 1, p. 13501, 2019.

GLOBAL TERRORISM DATABASE. [S. l.], [s. d.]. Disponível em: <http://apps.start.umd.edu/gtd/>. Acesso em: 13 out. 2023.

GREEN, Manfred S. *et al.* Confronting the threat of bioterrorism: realities, challenges, and defensive strategies. **The Lancet Infectious Diseases**, [s. l.], v. 19, n. 1, p. e2–e13, 2019.

GREENGENES.SECONDGENOME.COM. [S. l.], [s. d.]. Disponível em: <https://greengenes.secondgenome.com/>. Acesso em: 17 nov. 2023.

INTERNATIONAL HEALTH REGULATIONS (2005) – THIRD EDITION. [S. l.], [s. d.]. Disponível em: <https://www.who.int/publications-detail-redirect/9789241580496>. Acesso em: 14 out. 2023.

JACOBS, Michael K. The history of biologic warfare and bioterrorism. **Dermatologic Clinics**, [s. l.], v. 22, n. 3, p. 231–246, 2004.

JAHAN, Nusrat A. *et al.* Nanopore-Based Surveillance of Zoonotic Bacterial Pathogens in Farm-Dwelling Peridomestic Rodents. **Pathogens**, [s. l.], v. 10, n. 9, p. 1183, 2021.

JANSEN, H. J. *et al.* Biological warfare, bioterrorism, and biocrime. **Clinical Microbiology and Infection**, [s. l.], v. 20, n. 6, p. 488–496, 2014.

JELINKOVA, Pavlina *et al.* Development and Inter-Laboratory Validation of Diagnostics Panel for Detection of Biothreat Bacteria Based on MOL-PCR Assay. **Microorganisms**, [s. l.], v. 9, n. 1, p. 38, 2021.

JOHNSON, Andrew D. An extended IUPAC nomenclature code for polymorphic nucleic acids. **Bioinformatics**, [s. l.], v. 26, n. 10, p. 1386–1389, 2010.

JOHNSON, Jethro S. *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. **Nature Communications**, [s. l.], v. 10, n. 1, p. 5029, 2019.

KAWAI, Yuto *et al.* Development of an efficient antimicrobial susceptibility testing method with species identification by Nanopore sequencing of 16S rRNA amplicons. **PLOS ONE**, [s. l.], v. 17, n. 2, p. e0262912, 2022.

KOLTON, C.B. *et al.* Detection of *Bacillus anthracis* in animal tissues using InBios active anthrax detect rapid test lateral flow immunoassay. **Letters in Applied Microbiology**, [s. l.], v. 68, n. 6, p. 480–484, 2019.

KUMAR, Sudhir *et al.* MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. **Molecular Biology and Evolution**, [s. l.], v. 35, n. 6, p. 1547–1549, 2018.

LI, Heng. Minimap2: pairwise alignment for nucleotide sequences. **Bioinformatics**, [s. l.], v. 34, n. 18, p. 3094–3100, 2018.

LINDE, Jörg *et al.* Comparison of Illumina and Oxford Nanopore Technology for genome analysis of *Francisella tularensis*, *Bacillus anthracis*, and *Brucella suis*. **BMC Genomics**, [s. l.], v. 24, n. 1, p. 258, 2023.

LONG, Carrie M.; MARZI, Andrea. Biodefense Research Two Decades Later: Worth the Investment?. **The Lancet. Infectious diseases**, [s. l.], v. 21, n. 8, p. e222–e233, 2021.

LUDWIG, Wolfgang *et al.* ARB: a software environment for sequence data. **Nucleic Acids Research**, [s. l.], v. 32, n. 4, p. 1363–1371, 2004.

MANZULLI, Viviana *et al.* Discrimination of *Bacillus cereus* Group Members by MALDI-TOF Mass Spectrometry. **Microorganisms**, [s. l.], v. 9, n. 6, p. 1202, 2021.

MATIOLI, Sergio Russo. **Introdução à Bioinformática**. Campinas, SP: Fundação de Desenvolvimento da Unicamp - Funcamp, 2021.

MATSUO, Yoshiyuki *et al.* Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. **BMC Microbiology**, [s. l.], v. 21, n. 1, p. 35, 2021.

MAXAM, A M; GILBERT, W. A new method for sequencing DNA. **Proceedings of the National Academy of Sciences**, [s. l.], v. 74, n. 2, p. 560–564, 1977.

MCDONALD, Daniel *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. **The ISME Journal**, [s. l.], v. 6, n. 3, p. 610–618, 2012.

MEOLA, Marco *et al.* DAIRYdb: a manually curated reference database for improved taxonomy annotation of 16S rRNA gene sequences from dairy products. **BMC Genomics**, [s. l.], v. 20, n. 1, p. 560, 2019.

METZKER, Michael L. Sequencing technologies — the next generation. **Nature Reviews Genetics**, [s. l.], v. 11, n. 1, p. 31–46, 2010.

MIMURA, Mayumi *et al.* Microbial Water Quality Assessment of Private Wells Using 16S rRNA Gene Amplicon Sequencing with a Nanopore Sequencer. **Biological and Pharmaceutical Bulletin**, [s. l.], v. 46, n. 2, p. 263–271, 2023.

MINOGUE, Timothy D. *et al.* Next-Generation Sequencing for Biodefense: Biothreat Detection, Forensics, and the Clinic. **Clinical Chemistry**, [s. l.], v. 65, n. 3, p. 383–392, 2019.

MOTHUR WEBSITE. [S. l.], [s. d.]. Disponível em: <https://mothur.org>. Acesso em: 17 nov. 2023.

MULLIS, K. *et al.* Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. **Cold Spring Harbor Symposia on Quantitative Biology**, [s. l.], v.

51, p. 263–273, 1986.

NATIONAL BIODEFENSE STRATEGY. [S. l.], [s. d.]. Disponível em: <https://aspr.hhs.gov:443/biodefense/Pages/default.aspx>. Acesso em: 18 out. 2023.

NATIONAL INSTITUTES OF HEALTH. NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules (NIH Guidelines) - April 2019. [s. l.], 2019.

NCBI REFSEQ TARGETED LOCI PROJECT. [S. l.], [s. d.]. Disponível em: <https://www.ncbi.nlm.nih.gov/refseq/targetedloci/>. Acesso em: 15 out. 2023.

NEWMAN, Lucas; DUFFUS, Amanda L. J.; LEE, Cathy. Using the Free Program MEGA to Build Phylogenetic Trees from Molecular Data. **The American Biology Teacher**, [s. l.], v. 78, n. 7, p. 608–612, 2016.

NIAID EMERGING INFECTIOUS DISEASES/PATHOGENS | NIH: NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES. [S. l.], 2018. Disponível em: <https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens>. Acesso em: 21 nov. 2023.

NUCLEOTIDE BLAST: SEARCH NUCLEOTIDE DATABASES USING A NUCLEOTIDE QUERY. [S. l.], [s. d.]. Disponível em: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome. Acesso em: 29 nov. 2023.

OCHMAN, Howard; ELWYN, Susannah; MORAN, Nancy A. Calibrating bacterial evolution. **Proceedings of the National Academy of Sciences**, [s. l.], v. 96, n. 22, p. 12638–12643, 1999.

OLIVEIRA, Manuela *et al.* Biowarfare, bioterrorism and biocrime: A historical overview on microbial harmful applications. **Forensic Science International**, [s. l.], v. 314, p. 110366, 2020.

OUR MISSION - NCBI. [S. l.], [s. d.]. Disponível em: <https://www.ncbi.nlm.nih.gov/home/about/mission/>. Acesso em: 17 nov. 2023.

OUTI, Karvonen. Assessing public health preparedness and response in the EU. **the European Union**, [s. l.], 2022.

OVERVIEW — BLASTHELP DOCUMENTATION. [S. l.], [s. d.]. Disponível em: <https://blast.ncbi.nlm.nih.gov/doc/blast-help/developerinfo.html>. Acesso em: 17 dez. 2023.

PENNA, P. M. M. *et al.* BIOSSEGURANÇA: UMA REVISÃO. **Arquivos do Instituto Biológico**, [s. l.], v. 77, p. 555–565, 2020.

PLATAFORMA NACIONAL DE INFRAESTRUTURA DE PESQUISA. [S. l.], [s. d.]. Disponível em: <https://pnipe.mctic.gov.br/>. Acesso em: 17 out. 2023.

POLLOCK, Jolinda *et al.* The Madness of Microbiome: Attempting To Find Consensus “Best Practice” for 16S Microbiome Studies. **Applied and Environmental Microbiology**, [s. l.], v. 84, n. 7, p. e02627-17, 2018.

QUAST, Christian *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. **Nucleic Acids Research**, [s. l.], v. 41, n. D1, p. D590–D596, 2013.

RAMBAUSKE, Dora; CARDOSO, Telma Abdalla de Oliveira; NAVARRO, Marli Brito Moreira de Albuquerque. Bioterrorismo, riscos biológicos e as medidas de biossegurança aplicáveis ao Brasil. **Physis: Revista de Saúde Coletiva**, [s. l.], v. 24, p. 1181–1205, 2014.

RATHJEN, Nicholas A.; SHAHBODAGHI, S. David. Bioterrorism. **American Family Physician**, [s. l.], v. 104, n. 4, p. 376–385, 2021.

RITARI, Jarmo *et al.* Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database. **BMC Genomics**, [s. l.], v. 16, n. 1, p. 1056, 2015.

ROHWER, Robin R. *et al.* TaxAss: Leveraging a Custom Freshwater Database Achieves Fine-Scale Taxonomic Resolution. **mSphere**, [s. l.], v. 3, n. 5, p. 10.1128/msphere.00327-18, 2018.

ROY, Somak *et al.* Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. **The Journal of Molecular Diagnostics**, [s. l.], v. 20, n. 1, p. 4–27, 2018.

SAHL, Jason W. *et al.* A Bacillus anthracis Genome Sequence from the Sverdlovsk 1979 Autopsy Specimens. **mBio**, [s. l.], v. 7, n. 5, p. e01501-16, 2016.

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

SANTOS, Andres *et al.* Computational methods for 16S metabarcoding studies using Nanopore sequencing data. **Computational and Structural Biotechnology Journal**, [s. l.], v. 18, p. 296–305, 2020.

SCHLOSS, Patrick D. *et al.* Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. **Applied and Environmental Microbiology**, [s. l.], v. 75, n. 23, p. 7537–7541, 2009.

SCHLOSS, Patrick D. Reintroducing mothur: 10 Years Later. **Applied and Environmental Microbiology**, [s. l.], v. 86, n. 2, p. e02343-19, 2020.

SCHMITT, Ketra; ZACCHIA, Nicholas A. Total decontamination cost of the anthrax letter attacks. **Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science**, [s. l.], v. 10, n. 1, p. 98–107, 2012.

SELECT AGENTS AND TOXINS LIST | FEDERAL SELECT AGENT PROGRAM. [S. l.], 2023. Disponível em: <https://www.selectagents.gov/sat/list.htm>. Acesso em: 21 nov. 2023.

SEN, Mayukh. How a Cult Used Salad Bars to Orchestrate the Worst Bioterror Attack in US History. *In*: VICE. 15 mar. 2018. Disponível em: <https://www.vice.com/en/article/kzp4n9/wild-wild-country-netflix-salad-bar-bioterror-attack>. Acesso em: 13 nov. 2023.

SHENDURE, Jay *et al.* DNA sequencing at 40: past, present and future. **Nature**, [s. l.], v. 550, n. 7676, p. 345–353, 2017.

SILVA. [S. l.], [s. d.]. Disponível em: <https://www.arb-silva.de/>. Acesso em: 16 nov. 2023.

SOLNTCEVA, Vera; KOSTRZEWA, Markus; LARROUY-MAUMUS, Gerald. Detection of Species-Specific Lipids by Routine MALDI TOF Mass Spectrometry to Unlock the Challenges of Microbial Identification and Antimicrobial Susceptibility Testing. **Frontiers in Cellular and Infection Microbiology**, [s. l.], v. 10, 2021. Disponível em: <https://www.frontiersin.org/articles/10.3389/fcimb.2020.621452>. Acesso em: 2 set. 2023.

SOROKA, Marianna; WASOWICZ, Barbara; RYMASZEWSKA, Anna. Loop-Mediated Isothermal Amplification (LAMP): The Better Sibling of PCR?. **Cells**, [s. l.], v. 10, n. 8, p. 1931, 2021.

SOUSA, Roberto B. *et al.* A New Method for Extraction and Analysis of Ricin Samples through MALDI-TOF-MS/MS. **Toxins**, [s. l.], v. 11, n. 4, p. 201, 2019.

STEINBERGER, Ralf *et al.* **MedSys: A Multilingual Media Monitoring Tool for Medical Intelligence and Early Warning**. [S. l.], 2008. Disponível em: <https://publications.jrc.ec.europa.eu/repository/handle/JRC45523>. Acesso em: 14 out. 2023.

STEVENS, Bo Maxwell *et al.* Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. **Scientific Reports**, [s. l.], v. 13, n. 1, p. 9323, 2023.

THE LEAHY LETTER. [S. l.], [s. d.]. Disponível em: <https://www.fbi.gov/about-us/history/famous-cases/anthrax-amerithrax/the-leahy->

letter. Acesso em: 13 nov. 2023.

THE SEQUENCE READ ARCHIVE (SRA). [S. l.], [s. d.]. Disponível em: <https://www.ncbi.nlm.nih.gov/sra/docs/>. Acesso em: 1 dez. 2023.

THOMPSON, Julie D.; HIGGINS, Desmond G.; GIBSON, Toby J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, [s. l.], v. 22, n. 22, p. 4673–4680, 1994.

TIN, Derrick; SABETI, Pardis; CIOTTONE, Gregory R. Bioterrorism: An analysis of biological agents used in terrorist events. **The American Journal of Emergency Medicine**, [s. l.], v. 54, p. 117–121, 2022.

TSUCHIDA, Sachio; UMEMURA, Hiroshi; NAKAYAMA, Tomohiro. Current Status of Matrix-Assisted Laser Desorption/Ionization–Time-of-Flight Mass Spectrometry (MALDI-TOF MS) in Clinical Diagnostic Microbiology. **Molecules**, [s. l.], v. 25, n. 20, p. 4775, 2020.

WAECHTER, Christian *et al.* Comparative analysis of full-length 16s ribosomal RNA genome sequencing in human fecal samples using primer sets with different degrees of degeneracy. **Frontiers in Genetics**, [s. l.], v. 14, 2023. Disponível em: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1213829>. Acesso em: 16 ago. 2023.

WALKER, John. Frederick Sanger (1918–2013). **Nature**, [s. l.], v. 505, n. 7481, p. 27–27, 2014.

WOŁKOWICZ, Tomasz. The utility and perspectives of NGS-based methods in BSL-3 and BSL-4 laboratory – sequencing and analysis strategies. **Briefings in Functional Genomics**, [s. l.], v. 17, n. 6, p. 471–476, 2018.

WORKFLOW: 16S. [S. l.], [s. d.]. Disponível em: <https://epi2me.nanoporetech.com/workflow-docs-7?version=2022.01.07>. Acesso em: 17 dez. 2023.

WORLD HEALTH ORGANIZATION. **Laboratory biosafety manual, 4th edition: Risk Assessment**. [S. l.], 2020. Disponível em: <https://www.who.int/publications-detail-redirect/9789240011458>. Acesso em: 13 out. 2023.

ZASADA, Aleksandra A. Detection and Identification of Bacillus anthracis: From Conventional to Molecular Microbiology Methods. **Microorganisms**, [s. l.], v. 8, n. 1, p. 125, 2020.

ZHU, Hanliang *et al.* PCR past, present and future. **BioTechniques**, [s. l.], v. 69, n. 4, p. 317–325, 2020.

ŻUKOWSKA, Magdalena Elwira. Advanced methods of bacteriological identification in a clinical microbiology laboratory. **Journal of Pre-Clinical and Clinical Research**, [s. l.], v. 15, n. 2, p. 68–72, 2021.