

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa Interunidades de Pós-graduação em Bioinformática

KELDENN MELO FARIAS MORENO

**INVESTIGAÇÃO DE CASOS DE DOENÇA FEBRIL AGUDA E DOENÇA
NEUROLÓGICA GRAVE COM ETIOLOGIA DESCONHECIDA POR
SEQUENCIAMENTO METAGENÔMICO POR NANOPOROS**

Belo Horizonte

2023

KELDENN MELO FARIAS MORENO

**INVESTIGAÇÃO DE CASOS DE DOENÇA FEBRIL AGUDA E DOENÇA
NEUROLÓGICA GRAVE COM ETIOLOGIA DESCONHECIDA POR
SEQUENCIAMENTO METAGENÔMICO POR NANOPOROS**

Dissertação apresentada ao Programa
Interunidades de Pós-Graduação em
Bioinformática da Universidade Federal de Minas
Gerais como requisito para a obtenção do título de
Mestre em Bioinformática.

Orientadora: Profa. Dra. Marta Giovanetti

Co-orientador: Prof. Dr. Luiz Carlos Júnior Alcântara

Belo Horizonte

2023

043 Moreno, Keldenn Melo Farias.
Investigação de casos de doença febril aguda e doença neurológica grave com etiologia desconhecida por sequenciamento metagenômico por nanoporos [manuscrito] / Keldenn Melo Farias Moreno. – 2023.
61 f. : il. ; 29,5 cm.

Orientador: Marta Giovanetti. Coorientador: Luiz Carlos Júnior Alcântara.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. Doenças do Sistema Nervoso. 3. Sequenciamento por Nanoporos. 4. Metagenômica. 5. Nanoporos. I. Giovanetti, Marta. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 573:004



UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Keldenn Melo Farias Moreno

"INVESTIGAÇÃO DE CASOS DE DOENÇA FEBRIL AGUDA E DOENÇA NEUROLÓGICA GRAVE COM ETIOLOGIA DESCONHECIDA POR SEQUENCIAMENTO METAGENÔMICO"

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Profa Marta Giovanetti - Orientadora
Fundação Oswaldo Cruz

Prof. Luiz Carlos Júnior Alcântara - Coorientador
Fundação Oswaldo Cruz

Prof. Aristóteles Góes Neto
Universidade Federal de Minas Gerais

Prof. Svetoslav Nanev Slavov
Instituto Butantan

Belo Horizonte, 20 de junho de 2024.



Documento assinado eletronicamente por **Aristoteles Goes Neto, Professor do Magistério Superior**, em 24/06/2024, às 13:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Marta Giovanetti, Usuária Externa**, em 24/06/2024, às 13:25, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Svetoslav Nanev Slavov, Usuário Externo**, em 24/06/2024, às 13:56, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luiz Carlos Junior Alcantara, Usuário Externo**, em 15/07/2024, às 12:24, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3322751** e o código CRC **D72F82D2**.

Dedico essa dissertação à memória de Luzia Carneiro Moreno, que partiu deste mundo durante essa jornada e hoje vive em nossos corações. Minha querida avó, muito obrigado por tudo.

AGRADECIMENTOS

Agradeço aos meus pais, Kelin e Waldomiro Moreno, ao meu avô Antônio Farias e a minha avó Maria de Nazare Melo, por embarcaram junto a mim nessa aventura mesmo no início não tendo condições e ainda sim nunca deixaram de me apoiar, acreditaram em mim quando até mesmo eu não acreditava, não me deixaram faltar nada, e sempre estiveram presentes para o que eu mais precisasse, não só nesses últimos dois anos, mas por toda a minha vida. Posso conseguir esse título acadêmico de mestre, mas os meus verdadeiros mestres são vocês. À meu tio Klberwal Farias, por sempre ter me estendido a mão quando precisei e sempre presente quando possível.

Agradeço minha orientadora, Marta Giovanetti, por ter sido uma excelente orientadora, professora, pesquisadora e pessoa de coração de ouro durante todo esse trajeto, obrigado por acreditar em mim, pela paciência que tens comigo, pelo apoio incondicional, pelas críticas construtivas, por todas as oportunidades que me destes para contribuir em projetos, artigos, participação em grandes eventos científicos, e ser um grande exemplo para mim do tipo de pesquisador e pessoa que desejo ser. Acho que não tenho palavras suficientes para descrever quanta sorte tenho de tê-la como orientadora durante todo esse trajeto.

Agradeço ao meu co-orientador Luiz Alcântara, pela oportunidade de poder contribuir em projetos de grande valor para saúde pública como o NAVIO e outros projetos, assim como ao seu grupo de pesquisa em vigilância genômica. Agradeço a Mauricio Lima, Emerson Barbosa, Marcelo Tomé e Vagner Fonseca por suas valiosas contribuições neste trabalho, e em especial agradeço a Hegger Fritsch e Natália Guimarães por terem me acompanhado de perto durante esse trajeto com sua amizade, ensinamentos, confiança, profissionalismo, incentivos e descontrações.

Às agências de fomento, NIH, CAPES, FHEMIG e UFMG, por fornecerem o auxílio necessário para o desenvolvimento deste projeto. À Fundação Ezequiel Dias e ao Hospital Eduardo de Menezes, agradeço pela realização desse projeto e pelo trabalho de excelência que essas instituições e seus servidores realizam.

A todos os amigos que fiz subindo esses morros mineiros, em especial a Lucas Jorge e Emily, muito obrigado pelo seu imprescindível apoio no começo e durante essa jornada, à Mariana Oliveira, pelas sempre leves conversas e aventuras por Beagá, à Lucas Rocha e Tales Melo por terem feito do seu lar de descanso e aconchego meu lar também, e aos meus queridos bioinformatas, em especial Angie Puelles, Vinicius Castro, Maurício Peña, por todos os conselhos, discussões e apoios mútuos; e agradeço a Louella Kasbergen por todo o incentivo e paciência que tiveste comigo ao longo da construção dessa dissertação. Sem a leveza e o apoio de vocês, isso aqui nunca teria sido possível.

Aos meus queridos amigos da terra do verdadeiro açaí, muito obrigado por sempre me incentivarem e grande compreensão nos muitos momentos de ausência, e agradeço a Ramon Araujo, Edivaldo Sousa, Leonardo Sena e Rita Medeiros, exemplares profissionais que foram de vital importância para o começo da minha jornada científica.

Por último, mas não menos importante, gostaria de me agradecer. Quero me agradecer por acreditar em mim, quero agradecer por conseguir fazer o que fiz nos últimos dois anos. Quero me agradecer por ter tido forças de nunca desistir. Quero me agradecer por sempre tentar dar mais do que recebo. Quero me agradecer por sempre tentar fazer o certo. Obrigado.

“Em algum lugar, algo incrível está esperando para ser descoberto”

Carl Sagan

RESUMO

As Doenças Febris Agudas (DFA) e as Doenças Neurológicas Graves (DNG) se representam desafios significativos para o diagnóstico devido à diversidade de agentes infecciosos que podem causá-las e à limitação dos testes diagnósticos convencionais, que frequentemente resultam negativos ou inconclusivos. Diante disso, o sequenciamento metagenômico por nanoporos emergiu como uma ferramenta promissora na identificação de agentes etiológicos de doenças infecciosas desconhecidas, oferecendo estratégias variadas para a análise e a caracterização desses agentes. Este estudo teve como objetivo identificar os agentes etiológicos associados a casos de internação por DFA e DNG com resultado negativo para arboviroses por diagnóstico convencional em um hospital de Belo Horizonte, Minas Gerais. Foram analisadas amostras biológicas de 16 pacientes hospitalizados. Estas amostras foram submetidas a dois protocolos distintos: o sequenciamento shotgun e/ou o protocolo SMART-9N que envolve o enriquecimento de RNA, seguido de sequenciamento por nanoporos. As etapas subsequentes incluíram controle de qualidade, remoção de genes do hospedeiro, montagem de novo do genoma para formação de contigs, duas classificações taxonômicas para minimizar falso-positivos e análise filogenética. Nos casos de DFA, identificamos um espectro de microrganismos associados às relatadas patologias, incluindo *Escherichia coli*, *Streptococcus sp.*, Vírus da Imunodeficiência Humana 1 (Subtipo B), e Pegivírus Humano (Genótipo 2B). Já nos casos de DNG, os microrganismos detectados com prévia associação correlacionada com a patologia identificados incluíram *Escherichia coli*, *Clostridium sp.* e vírus da dengue tipo 2 (Genótipo-II). Este estudo evidencia a eficiência e a rapidez da análise metagenômica, ressaltando sua capacidade de identificar patógenos associados a diversas doenças. Além disso, demonstra a viabilidade de recuperar genomas quase completos de vírus de RNA, sublinhando a relevância desta técnica no aprimoramento da precisão diagnóstica e na personalização de tratamentos.

Palavras-Chave: doença febril aguda; doenças neurológicas severas; sequenciamento por nanoporos; metagenômica.

ABSTRACT

Acute Febrile Illness (AFI) and Severe Neurological Diseases (SND) pose significant diagnostic challenges due to the diversity of infectious agents that can cause them and the limitations of conventional diagnostic tests, which are often negative. In view of this, nanopore metagenomic sequencing has emerged as a promising tool for identifying the etiological agents of unknown infectious diseases, offering a variety of strategies for analyzing and characterizing these agents. The aim of this study was to identify the etiological agents associated with cases of hospitalization for AFI and SND in the Eduardo de Menezes Hospital, Belo Horizonte, Minas Gerais. Biological samples from 16 hospitalized patients were analyzed. These samples were subjected to two different protocols: shotgun sequencing and RNA enrichment using the SMART-9N protocol, followed by nanopore sequencing. Subsequent steps included quality control, removal of host genes, de novo assembly of the genome to form contigs, two taxonomic classifications to minimize false positives and phylogenetic analysis. In cases of AFI, we identified a spectrum of disease-associated microbes including *Escherichia coli*, *Streptococcus sp.*, Human Immunodeficiency Virus 1 (Subtype B), and Human Pegivirus (Genotype 2B). In DNG cases, the pathogens identified included *Escherichia coli*, *Clostridium sp.* and dengue virus type 2 (Genotype-II lineage). This study demonstrates the effectiveness and speed of metagenomic analysis, highlighting its ability to identify pathogens associated with various diseases, the feasibility of recovering almost complete genomes of RNA viruses and the importance of this technique for more accurate diagnoses and personalized treatments.

Keywords: acute febrile illness; severe neurological disorders; nanopore sequencing; metagenomics.

LISTA DE FIGURAS

Figura 1. Workflow da análise metagenômica de DFA e DNG em pacientes hospitalizados.

1) O trabalho começa na coleta de amostras biológicas de pacientes hospitalizados com diagnóstico de DFA ou DNG de etiologia desconhecida; 2) Processamento Laboratorial: Conforme a sintomatologia dos pacientes, uma parte das amostras foi submetida a extração e enriquecimento de RNA usando o protocolo SMART-9N, enquanto outras foram destinadas apenas à extração para posterior análise por sequenciamento shotgun e síntese de cDNA, seguida de sequenciamento genético com tecnologia de nanoporos; 3) Análise de dados: As sequências foram submetidas a um controle de qualidade, remoção de material genético do hospedeiro e geração de contigs através da montagem de novo do genoma e submetidas a três tipos de classificações taxonômicas; 4) Identificação de agentes patogênicos e análise filogenética: Os agentes patogênicos identificados com uma cobertura superior a 80% foram submetidos a um mapeamento e montagem de genoma consenso. A completude destes genomas foi avaliada, culminando na construção de uma árvore filogenética para classificação genética do patógeno.....20

Figura 2. Metaperfil dos organismos presentes nas amostras analisadas por paciente no protocolo Shotgun.....27

Figura 3. Visão geral da proporção de casos por classificação e microrganismos identificados. A) Porcentagem de casos por classificação clínica. B) Proporção de microrganismos associados à DNG nas amostras. C) Microrganismos associados à DNG identificados. D) Proporção de microrganismos associados à DFA nas amostras. E) Microrganismos associados a DFA identificados.....28

Figura 4. Metaperfil dos organismos presentes nas amostras analisadas por paciente no protocolo SMART-9N. As nomenclaturas em vermelho são microrganismos associados a causa de DFA e/ou DNG.....29

Figura 5. Cobertura do genoma dos agentes patogênicos virais identificados. (A) A cobertura do genoma do HIV, (B) A cobertura do genoma do HPgV e (C) A cobertura do genoma de DENV2.....30

Figura 6. Árvore filogenética da análise do genoma do HIV. A amostra do estudo é a amostra destacada agrupada com subtipo B.....31

Figura 7. Árvore filogenética da análise do genoma do HPgV. A amostra do estudo é aquela que se destaca no clado do genótipo 2B.....32

Figura 8. Árvore filogenética da análise do genoma do DENV-2. Comparando os genótipos de DENV-2 detectados no Brasil, a amostra do estudo é a amostra destacada que está agrupada no genótipo II.....33

LISTA DE QUADROS E TABELAS

- Tabela 1.** Perfil clínico e sociodemográfico dos pacientes e descrição do protocolo de análise metagenômica utilizado para cada caso.....25
- Tabela 2.** Dados de sequenciamento dos microrganismos associados a doenças detectados pelo DIAMOND.....30
- Tabela 3.** Código do GenBank e linhagem associada a cada genoma utilizado para a construção das árvores filogenéticas.....38

LISTA DE ABREVIACES

cDNA - Fita complementar de DNA

CZ-ID - Plataforma de anlise metagenmica **Chan-Zuckenberg ID**

DENV-2 - Vrus da Dengue tipo 2, oriundo da nomenclatura viral "**Dengue virus 2**"

DFA - **Doenas Febris Agudas**

DNA - cido desoxirribonucleico, do ingls "**Deoxyribonucleic acid**"

DNG - **Doenas Neurolgicas Graves**

HIV - Vrus da imunodeficincia humana, oriundo da nomenclatura viral **Human Immunodeficiency Virus**"

HEM - **Hospital Eduardo de Menezes**

HPgV - Pegivirus hominis, oriundo da nomenclatura viral "**Human Pegivirus**"

HMM - Algoritmo de Markov Oculto, do ingls "**Hidden Markov Model**"

FUNED - **Fundao Ezequiel Dias**

GTR - Modelo geral de substituio de nucleotdeos reversvel no tempo, do ingls "**General Time Reversible**"

NT-NCBI - Banco de dados Nucleotide Database, pertencente ao National Center for Biotechnology Information

RNA - cido ribonucleico, do ingls "**Ribonucleic acid**"

ONT - **Oxford Nanopore Technologies**

PCR - Reao em Cadeia da Polimerase, do ingls "**Polymerase Chain Reaction**"

SUMÁRIO

| | |
|---|-----------|
| 1. INTRODUÇÃO..... | 15 |
| 1.1 Doenças Febris Agudas e Doenças Neurológicas Severas: Uma breve contextualização..... | 15 |
| 1.2 Metagenômica por nanoporos..... | 16 |
| 1.3 Estratégias de análise bioinformática para identificação de patógenos em dados metagenômicos..... | 17 |
| 1.4 Justificativa..... | 18 |
| 2. OBJETIVOS..... | 19 |
| 2.1 Objetivo Geral..... | 19 |
| 2.2 Objetivos Secundários..... | 19 |
| 3. MATERIAIS E MÉTODOS..... | 20 |
| 3.1 Aspectos éticos..... | 20 |
| 3.2 Seleção amostral..... | 20 |
| 3.3 Extração de RNA e preparação da amostra..... | 21 |
| 3.4 Sequenciamento metagenômico por nanoporos..... | 22 |
| 3.5 Análise bioinformática..... | 22 |
| 3.5.1 Chamada de base e demultiplex..... | 22 |
| 3.5.2 Controle de qualidade..... | 22 |
| 3.5.3 Retirada de genes do hospedeiro..... | 23 |
| 3.5.4 Montagem de novo do genoma..... | 23 |
| 3.5.5 Classificação taxonômica por DIAMOND..... | 23 |
| 3.5.6 Análise confirmatória por classificação taxonômica por Kraken2..... | 23 |
| 3.5.7 Classificação taxonômica por VirSorter2..... | 24 |
| 3.5.8 Análise filogenética..... | 24 |
| 4. RESULTADOS E DISCUSSÃO..... | 25 |
| 4.1 Perfil clínico e sociodemográfico da população estudada..... | 25 |
| 4.2 Análise do sequenciamento Shotgun..... | 26 |
| 4.3 Análise do sequenciamento SMART 9N..... | 27 |
| 5. DISCUSSÃO GERAL..... | 34 |
| 6. CONCLUSÃO..... | 37 |
| REFERÊNCIAS BIBLIOGRÁFICAS..... | 38 |
| APÊNDICES..... | 43 |
| Genomas utilizados para a construção das árvores filogenéticas..... | 43 |
| Artigo publicado..... | 46 |
| Outros artigos publicados..... | 58 |

1. INTRODUÇÃO

1.1 Doenças Febris Agudas e Doenças Neurológicas Severas: Uma breve contextualização

A Doença Febril Aguda (DFA) é uma síndrome clínica caracterizada por febre ($\geq 38,0$ °C) e frequentemente acompanhada por vários sintomas inespecíficos, como dor de cabeça, erupção cutânea, dores musculares e articulares (Kigozi *et al.*, 2023). Essa é uma síndrome que requer frequentemente hospitalização e investigações extensas para determinar a sua etiologia, cuja causa se deve principalmente a agentes patogênicos e infecciosos que variam epidemiologicamente em todo o mundo. Por exemplo, a malária é a causa predominante de DFA em África, enquanto na América Latina e na Ásia, a Dengue e a Leptospirose são as causas mais comuns desta síndrome (Iroh Tam, Obaro and Storch, 2016; Kigozi *et al.*, 2023)

Apesar dos avanços nos diagnósticos médicos, uma proporção significativa de casos de DFA permanece sem solução, com abordagens diagnósticas convencionais que muitas vezes produzem resultados inconclusivos ou negativos. Estudos indicam que quase 50% dos casos de DFA permanecem sem diagnóstico, um número que pode ser maior devido aos frequentes erros de diagnóstico de DFA como doenças arbovirais em ambientes de cuidados primários precários, onde não são realizadas investigações adicionais (Moreira *et al.*, 2018). (Bressan *et al.*, 2023)

Os distúrbios neurológicos graves (DNG) abrangem uma série de condições que afetam o cérebro, a medula espinhal e os nervos. Os sintomas de DNG podem incluir dores de cabeça, disfunção do paladar e do olfato, fraqueza muscular, paralisia, perda de sensibilidade, má coordenação, confusão, convulsões e dor. As DNGs são a segunda principal causa de morte em todo o mundo, com várias causas, incluindo bactérias infecciosas, fungos e particularmente, vírus (Chen, X. *et al.*, 2021; Wouk *et al.*, 2021).

Devido aos sintomas inespecíficos e à ampla gama de infecções associadas, as DNGs e a DFA apresentam desafios diagnósticos significativos e embora exista uma ampla infraestrutura de diagnóstico estabelecido tanto no setor público quanto no privado da área da saúde para investigar essas doenças, muitas vezes a investigação para identificar sua causa pode resultar em achados inconclusivos ou negativos. Por conta disso, muitas vezes é necessário recorrer a técnicas complementares para investigar a origem da patologia.

1.2 Metagenômica por nanoporos

Dentre essas, a metagenômica tem ganhado destaque por ser um método com capacidade revolucionária de fornecer informações abrangentes da composição microbiana em amostras biológicas, por conta de ser uma técnica que tem a capacidade de sequenciar específicas partes de um genoma, conceito conhecido como sequenciamento de amplicon, ou sequenciar randomicamente toda a amplitude de DNA presente em uma amostra biológica, conceito conhecido como sequenciamento Shotgun (Rausch *et al.*, 2019).

Nos últimos tempos, a utilização da metagenômica tem crescido significativamente na identificação de potenciais patógenos emergentes e reemergentes através da análise de sequências de nucleotídeos, e gera dados genômicos valiosos que geralmente não são detectados pelos métodos de diagnóstico tradicionais, com o objetivo principal de descobrir e caracterizar microrganismos associados a patologias como doenças febris ou neurológicas de etiologia desconhecida e melhoram nossa compreensão da biologia das doenças, das interações patógeno-hospedeiro e da epidemiologia (Chiu and Miller, 2019).

A aplicação da metagenômica na detecção de agentes infecciosos, especialmente aqueles indescritíveis aos métodos diagnósticos convencionais, teve um aumento acentuado nos últimos anos, aproveitando uma variedade de tecnologias de sequenciamento de próxima geração (NGS). Entre estes, a plataforma de sequenciamento MinION é cada vez mais favorecida devido à sua capacidade de gerar rapidamente volumes substanciais de dados. Sua capacidade de sequenciamento de longas sequências genéticas (também chamadas de *long reads*) abordam os desafios associados a regiões repetitivas ambíguas, melhorando a contiguidade genômica (Ciuffreda, Rodríguez-Pérez and Flores, 2021; Deng *et al.*, 2022). A portabilidade e a relação custo-benefício desta tecnologia reforçam ainda mais sua adoção na pesquisa genômica, e à medida que o panorama epidemiológico das doenças infecciosas continua a mudar, a relevância de metodologias metagenômicas sofisticadas torna-se mais pronunciada (Latorre-Pérez *et al.*, 2020).

Através dessa plataforma, é possível realizar análises de forma rápida e integrar essa tecnologia com uma metodologia de análise eficaz e ágil para a identificação dos microrganismos presentes, não apenas acelerando a recuperação do paciente afetado, mas também aprimorando substancialmente o sistema de vigilância genômica.

1.3 Estratégias de análise bioinformática para identificação de patógenos em dados metagenômicos

Para analisar a presença de um microorganismo, há diferentes meios análises dos dados brutos de sequenciamento, alguns estudos utilizam ferramentas de classificação taxonômica diretamente com as reads sequenciadas, enquanto outros realizam um mapeamento prévio das reads, utilizando o genoma de referência do hospedeiro, antes de proceder com a classificação taxonômica. No entanto, a metodologia mais recomendada atualmente é que, após o processo de controle de qualidade, seja feito o processo de montagem das sequências (Breitwieser, Lu and Salzberg, 2019) (Ghurye, Cepeda-Espinoza and Pop, 2016).

Em análises metagenômicas para investigar agentes etiológicos desconhecidos, não é possível ter um organismo de referência para realizar a montagem, e por conta disso, necessita-se realizar montagens *de novo* das sequências. Para long-reads, montadores de contigs utilizam principalmente o algoritmo de consenso de sobreposição, como o Canu (Koren *et al.*, 2017), Flye (Kolmogorov *et al.*, 2019) e NECAT (Chen *et al.*, 2021), que têm a capacidade de gerar contigs de maior qualidade e confiabilidade.

Após a montagem dos genomas, é possível realizar o agrupamento dos contigs, seja de acordo com um genoma de referência ou seguindo um padrão para a geração de genomas conhecidos ou desconhecidos, em um processo conhecido como "binning" (Nissen *et al.*, 2021). No entanto, este procedimento pode resultar em múltiplos contigs, o que pode complicar o processamento do algoritmo e aumentar o risco de produzir resultados falsos positivos, e por esse motivo, algumas estratégias de classificação taxonômica aconselham evitar essa etapa (Nayfach, *et al.*, 2021). Como é possível realizar os processos de classificação sem a necessidade do "binning", este último procedimento é considerado opcional.

Finalizado o processo de montagem de genomas e opcional uso do binning, é possível realizar o processo de classificação taxonômica. Para a identificação de genes presentes, os classificadores mais utilizados atualmente são o Kraken2 e o DIAMOND. O Kraken2 funciona através do algoritmo de k-mers, em que faz uma pesquisa comparativa das sequências com um banco de dados que contém todos os k-mer de cada genoma com o identificador de espécie (Wood, Lu and Langmead, 2019). Além do Kraken2, nos últimos tempos tem crescido o uso do DIAMOND, que em vez de utilizar as sequências diretamente, traduz essas sequências e conduz a pesquisa comparativa por homologia de proteínas (Buchfink, Reuter and Drost, 2021).

Além dessas ferramentas, outras estratégias estão sendo empregadas para a identificação de microrganismos, especialmente no que diz respeito aos patógenos virais. Novas metodologias estão sendo desenvolvidas, com destaque para aquelas que se baseiam principalmente no Algoritmo de Markov Oculto (HMM, do inglês "Hidden Markov Model") (Oliveira and Gruber, 2021), um modelo matemático utilizado para o reconhecimento de padrões, e no contexto das análises genéticas, é empregado para identificar e classificar as leituras de DNA com relação aos microrganismos aos quais pertencem (Nayfach, Páez-Espino, *et al.*, 2021). Por meio deste algoritmo, foram desenvolvidas diversas ferramentas e pipelines de análise voltadas para a identificação de genomas virais, tais como o VIRIFY, VirSorter e VirFinder, entre outros (Nooij *et al.*, 2018).

Essas ferramentas podem ser integradas com outras oriundas do HMM, permitindo a estimativa da completude e integridade de fragmentos de genoma, a remoção de regiões hospedeiras de provírus integrados, como no CheckV (Nayfach, Camargo, *et al.*, 2021), além de possibilitar a anotação e classificação funcional de vírus, sejam eles descritos ou inéditos. Isso é feito por meio de ferramentas como o Distilled and Refined Annotation of Metabolism (DRAM) (Guo *et al.*, 2021) (Shaffer *et al.*, 2020).

1.4 Justificativa

A implementação de diversas estratégias na investigação metagenômica desempenha um papel crucial. Isso se deve à importância de realizar comparações abrangentes, analisar e mitigar falsos positivos, além de aprimorar as técnicas já existentes. A necessidade de compreender cada vez mais a dinâmica das doenças infecciosas e obter respostas que contribuam para a vigilância eficaz de patógenos é uma razão fundamental para testar, definir e implementar os melhores métodos de análise. Esses métodos são essenciais para a identificação e caracterização precisa de microrganismos associados a patologias. Portanto, a busca por estratégias aprimoradas na investigação metagenômica se justifica plenamente, dado o impacto significativo que isso pode ter no campo da saúde e na nossa compreensão das doenças infecciosas.

2. OBJETIVOS

2.1 Objetivo Geral

Identificar os agentes etiológicos associados a casos de internação por Doença Febril Aguda e Doenças Neurológicas Graves com resultado negativo ou inconclusivo por diagnóstico convencional em um hospital de Belo Horizonte, Minas Gerais.

2.2 Objetivos Secundários

- Detectar potenciais microrganismos associados a origem do quadro clínico de Doença Febril Aguda e Doenças Neurológicas Graves circulantes na cidade de Belo Horizonte
- Implementar uma estratégia de identificação de microrganismo ágil e eficaz para a geração de respostas rápidas e confiáveis para a solicitação médica.
- Mapear e gerar genomas de microrganismos importantes para a saúde pública, para fortalecer a vigilância genômica do Estado de Minas Gerais

3. MATERIAIS E MÉTODOS

3.1 Aspectos éticos

Este estudo foi aprovado pelo Comitê de Ética e Pesquisa da Universidade Federal de Minas Gerais, sob o registro CAAE: 90249218.6.1001.5248

3.2 Seleção amostral

O Hospital Eduardo de Menezes (HEM), em Belo Horizonte, Brasil, desempenha várias funções críticas na prestação de cuidados de saúde, incluindo a realização de investigações etiológicas de pacientes hospitalizados com diversas patologias, entre essas a Doença Febril Aguda ou a Doença Neurológica Grave e em colaboração com a Fundação Ezequiel Dias (FUNED), o hospital procura elucidar as possíveis associações entre infecções com diagnósticos inconclusivos e infecções microbianas através de investigações sorológicas e de biologia molecular abrangentes destinadas a detectar arbovírus, hepatite, rickettsiose, hantavíroses, leptospirose, febre maculosa e outros agentes patogênicos, expandindo cada vez mais as metodologias de análise. No entanto, em 2022, as amostras de dezesseis pacientes hospitalizados resultaram em resultados inconclusivos e/ou negativos, e como abordagem alternativa para identificar os agentes etiológicos, amostras biológicas desses pacientes foram submetidas a uma análise metagenômica (Figura 1).

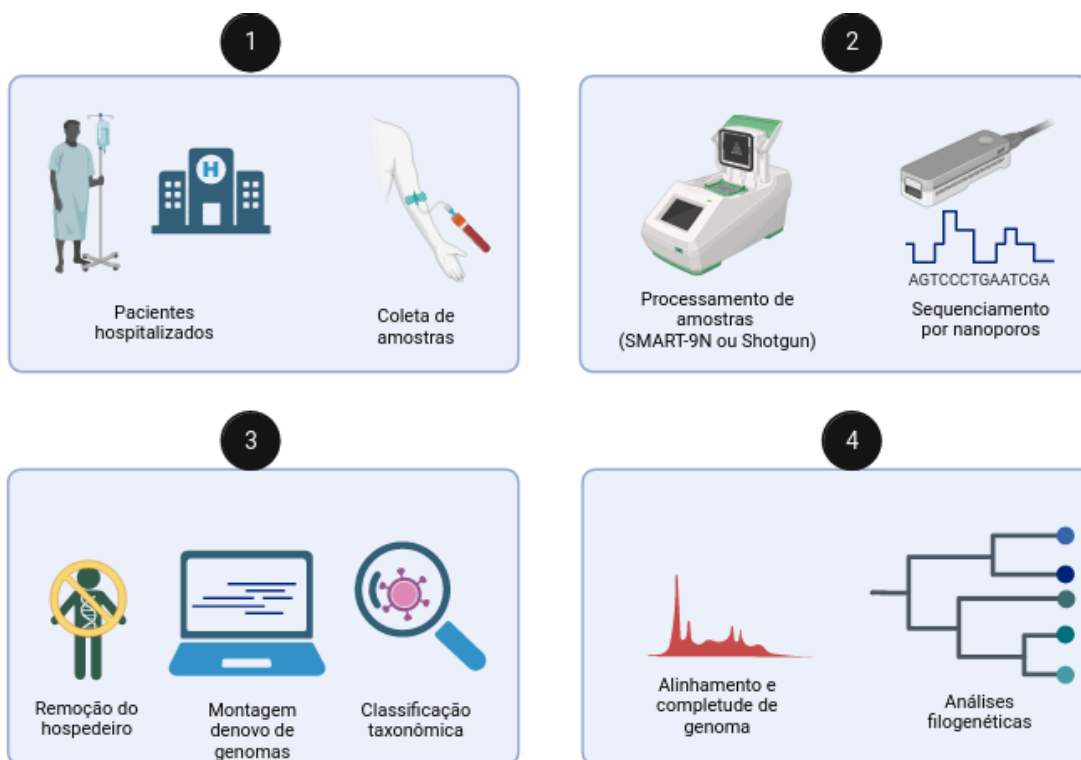


Figura 1. Workflow da análise metagenômica de DFA e DNG em pacientes hospitalizados. 1) O trabalho começa na coleta de amostras biológicas de pacientes hospitalizados com diagnóstico de DFA ou DNG de etiologia desconhecida; 2) Processamento Laboratorial: Dependendo da sintomatologia, parte das amostras selecionadas para metagenômica foram submetidas a extração e enriquecimento usando o protocolo SMART-9N e outra somente extração para análise shotgun, síntese de cDNA, seguido de sequenciamento genético empregando tecnologia de nanoporos; 3) Análise de dados: As sequências foram submetidas a um controle de qualidade, remoção de material genético do hospedeiro e geração de contigs através da montagem de novo do genoma e submetidos a três tipos de classificações taxonômicas; 4) Identificação de agentes patogênicos e análise filogenética: Os agentes patogênicos identificados com uma cobertura superior a 80% foram submetidos a um mapeamento e montagem de genoma consenso. A completude destes genomas foi avaliada, culminando na construção de uma árvore filogenética para classificação genética do patógeno.

3.3 Extração de RNA e preparação da amostra

Com base nos sintomas, classificação clínica e input de material genético extraído da amostra biológica, as amostras biológicas foram submetidas a dois tipos de sequenciamento: Algumas foram submetidas a uma análise shotgun, que consistiu na extração de ácido ribonucleico (RNA), síntese de DNA complementar (cDNA) e preparação de biblioteca, e outras foram submetidas ao protocolo SMART-9N de tratamento metagenômico (Claro *et al.*, 2021), utilizando um modelo SMART-Seq com três primers aleatórios compatível com os adaptadores da Oxford Nanopore Technologies (ONT). Este protocolo é particularmente eficaz para melhorar a detecção e sequenciamento de genomas com foco em vírus de RNA, porém também permite a identificação e sequenciamento de vírus de DNA, bactérias, fungos, eucariotas e archaea. Das amostras submetidas a esse protocolo, o RNA foi extraído das amostras utilizando o High Pure RNA Isolation Kit (Hoffmann-La Roche, Suíça), concentrado utilizando o RNA Clean & Concentrator-5 (Zymo Research, EUA) e feita a reação de síntese de cDNA seguida de PCR utilizando o protocolo SMART-9N. Para esse processo, foram utilizados dois iniciadores durante a síntese de cDNA: O primeiro iniciador 9N é composto de uma sequência de 9 nucleotídeos randômicos que se ligam ao RNA. Esse iniciador possui ainda uma região complementar a um outro iniciador que será utilizado na PCR. O segundo iniciador utilizado na síntese de cDNA é o TSO, que tem como função se ligar à extremidade final do cDNA recém-sintetizado e adicionar uma tag RLB, que também será complementar ao primer da etapa de PCR. Em seguida foi feita a PCR utilizando um primer que tem como alvo as sequências adicionadas pelos iniciadores na etapa de cDNA. Além das amostras, um controle negativo foi adicionado.

3.4 Sequenciamento metagenômico por nanoporos

Para a preparação da biblioteca, houve primeiro a purificação das amostras submetidas a PCR do protocolo SMART-9N usando as beads magnéticas AMPure XP (Beckman Coulter, EUA). Após essa purificação, todas as amostras foram submetidas a normalização da concentração de DNA, outra purificação por beads magnéticas seguido de preparação da cauda poliar e endprep usando o Ultra II End Prep Kit (New England Biolabs, EUA), seguido de mais uma purificação, ligação de barcode usando o PCR Barcoding Expansion (Oxford Nanopore Technologies, Reino Unido), outra etapa de purificação e ligação do adaptador com o protocolo de expansão Adapter Mix II (Oxford Nanopore Technologies, Reino Unido). A concentração de material genômico em cada etapa foi quantificada utilizando o Qubit dsDNA Quantification Assay Kit (Life Technologies, EUA). As bibliotecas genômicas preparadas foram carregadas em flowcells Oxford Nanopore MinION R9.4.1 (um para cada tipo de sequenciamento), preparada de acordo com as instruções do fabricante usando o Flow Cell Priming Kit SQK-112 (Oxford Nanopore Technologies, Reino Unido) e sequenciadas no dispositivo MinION Mk1B da ONT usando o software MinKNOW.

3.5 Análise bioinformática

Para a análise bioinformática foi realizada uma série de etapas: Conversão do sinal elétrico em bases, demultiplexação das amostras, avaliação inicial da qualidade das sequências geradas (reads), mapeamento e retirada de reads de seres humanos, montagem *denovo* das sequências restantes, classificação taxonômica dos contigs gerados, análise de completude, mapeamento e montagem das sequências-consensos dos microrganismos identificados e análise filogenética.

3.5.1 Chamada de base e demultiplex

A execução inicial do sequenciamento gerou dados brutos no formato FAST5, capturando sinais elétricos. Esses arquivos foram processados para chamada de base, convertendo-os para o formato FASTQ, e demultiplexados utilizando o software Guppy, desenvolvido pela ONT (Wick, Judd and Holt, 2019).

3.5.2 Controle de qualidade

Após a chamada de base e demultiplexação, o controle de qualidade foi feito no software fastp (Chen *et al.*, 2018) para retirar reads de baixa qualidade ou tamanho

insuficiente para utilizar no processo de montagem, colocando phred score mínimo de qualidade 9, tamanho mínimo de 100 pares de base, ativação do filtro de sequências de baixa qualidade e permanência de duplicação. O comando utilizado para a realização do controle de qualidade foi o seguinte: *fastp -i sample_x.fastq --qualified_quality_phred 9 --length_required 100 --low_complexity_filter --complexity_threshold 30 --dont_eval_duplication --failed_out --html fastp.html -o sample_xx_quality_filtered.fastq*

3.5.3 Retirada de genes do hospedeiro

A retirada de genes humanos foi realizada através do software minimap2 comparando com as reads que atenderam os critérios de qualidade com o banco de dados de genoma humano do software (Li, 2018). O comando utilizado foi *minimap2 -t 210 -ax splice minimap_human_db sample_xx_quality_filtered.fastq -o sampleXX.humanfiltered.sam --split-prefix temp_name*

3.5.4 Montagem *de novo* do genoma

A partir das sequências que atenderam aos critérios de qualidade e sem genes de hospedeiro, os arquivos FASTQ foram processados pelo software de montagem para long-reads Flye (Kolmogorov *et al.*, 2019), empregando o parâmetro específico para análise metagenômica metaFlye (Kolmogorov *et al.*, 2020). Esse processo resultou na produção de contigs que foram posteriormente salvos em um arquivo no formato FASTA. O comando utilizado para montagem foi: *flye --nano sampleXX_humanfiltered.fastq.gz --out-dir /sampleXX_assembled --threads 10 --meta*

3.5.5 Classificação taxonômica por DIAMOND

Por conta do grande uso computacional e levando em conta a praticidade do uso de ferramentas em nuvem, para a análise no software DIAMOND foi utilizado o pipeline de análise metagenômica em nuvem de código aberto fornecida pela plataforma Chan-Zuckerberg IDseq (CZ-ID) (Kalantar *et al.*, 2020), cruzando os contigs gerados com a base de dados NCBI Nucleotide (NT).

3.5.6 Análise confirmatória por classificação taxonômica por Kraken2

Como forma de confirmar, a classificação taxonômica dos contigs montados, também foi feita utilizando o software Kraken2 de acordo com as instruções fornecidas na

documentação (Wood, Lu and Langmead, 2019). Os parâmetros utilizados incluíram a especificação do formato de entrada e saída, a alocação de 10 threads de processamento e a inclusão do parâmetro "--use-mpa-style" para gerar um perfil metagenômico normalizado em termos de dados percentuais, que estava associado aos contigs. Para conduzir essa classificação, empregamos o banco de dados mais recente disponível, denominado "standard-db-kraken2". O comando utilizado para classificação taxonômica no Kraken2 foi:

```
kraken2 assembly.fasta --threads 20 --db standard_db_kraken2 --output classification-kraken2-sampleXX --report --use-mpa-style.
```

3.5.7 Classificação taxonômica por VirSorter2

Para investigar a presença de genes virais utilizando HMM, os contigs resultantes da montagem foram submetidos ao pipeline de análise genômica denominado "VirSorter2," que é especializado na identificação de genes virais (Guo *et al.*, 2021). O arquivo FASTA gerado por meio desse pipeline, que continha os prováveis genes virais, foi posteriormente submetido ao software CheckV para uma análise mais detalhada desses genes (Nayfach, Camargo, *et al.*, 2021). Para a execução dessas análises, utilizamos os parâmetros padrão conforme descritos na documentação correspondente e utilizamos o banco de dados mais atualizado disponível.

3.5.8 Análise filogenética

Os agentes patogênicos com uma cobertura do genoma superior a 80% foram submetidos à montagem do genoma utilizando a ferramenta Genome Detective (Vilsker *et al.*, 2019). Os genomas de consenso resultantes foram depois avaliados quanto à sua qualidade e integralidade utilizando o software CheckV, através do comando *checkv end_to_end final-viral-combined.fa checkv -t 10 -d checkv-db-v1.5*. Estas sequências genômicas foram alinhadas com genomas de referência extraídos da base de dados do NCBI utilizando o MAFFT com parâmetros predefinidos (Katoh *et al.*, 2002). A curadoria manual foi efetuada utilizando o Aliview para eliminar artefatos biológicos (Larsson, 2014). As árvores filogenéticas foram construídas utilizando a estimativa de Máxima Verossimilhança (ML) efetuada pelo IQ-TREE2 (Minh *et al.*, 2020), empregando o modelo geral de substituição de nucleotídeos reversível no tempo (GTR) e apoiado por 1.000 réplicas bootstrap. A visualização final das árvores filogenéticas foi feita pelo software FigTree (Rambaut, 2018). Os conjuntos de dados utilizados para construir estas árvores filogenéticas estão detalhados na seção de apêndice 7.1.

4. RESULTADOS E DISCUSSÃO

4.1 Perfil clínico e sociodemográfico da população estudada

Neste estudo, amostras biológicas de 16 pacientes hospitalizados no Hospital Eduardo de Menezes por DFA e DNG não tiveram resultados conclusivos nas investigações laboratoriais e por conta disso, foram submetidos a análise metagenômica. Desses, dois não tiveram resultados processados por conta da ausência de informações clínicas e socio epidemiológicas acerca dos pacientes.

A Tabela 1 resume os dados epidemiológicos dos pacientes incluídos, mostrando distribuição etária variando de 14 a 61 anos. A distribuição por gênero foi igual, com 50% de participantes do sexo feminino e 50% do sexo masculino. Todos os pacientes eram residentes de Minas Gerais, predominantemente de Belo Horizonte, embora alguns fossem de cidades próximas, como Ouro Preto ou Ribeirão das Neves. Clinicamente, os pacientes internados apresentavam uma variedade de sintomas que necessitavam de sua internação. Aqueles com DFS mais comumente relataram febre, mialgia, artralgia e dores de cabeça. Em contraste, os casos de SND foram associados principalmente a sintomas como dores de cabeça, convulsões e potencial neurosífilis. Um total de 18 amostras foram analisadas desses 16 pacientes, com dois indivíduos (pacientes 7 e 11) contribuindo com duas amostras cada. Os tipos de amostras incluíram cinco amostras de sangue total, duas de líquido cefalorraquidiano, cinco de soro e duas amostras de sangue total coletadas em um tubo de conservante de RNA (Tempus).

De acordo com os sintomas relatados, onze pacientes tiveram suas amostras submetidas ao protocolo SMART-9N e quatro foram submetidos ao protocolo Shotgun conforme descrito na Tabela 1. Os resultados da análise com o protocolo SMART-9N já foram publicados em artigo científico, inserido na seção de apêndices da atual dissertação.

Tabela 1. Perfil clínico e sociodemográfico dos pacientes e o protocolo de análise metagenômica utilizado para cada caso

| Paciente | Gênero | Idade | Cidade | Estado | Classificação clínica | Sintomas | Protocolo metagenômico | Tipo de amostra |
|----------|-----------|-------|----------------|--------------|-----------------------|---|------------------------|-----------------|
| 1 | Masculino | 53 | Belo Horizonte | Minas Gerais | DNG | Dores fortes de cabeça | SMART-9N | Soro |
| 2 | Feminino | 38 | Belo Horizonte | Minas Gerais | DFA | Febre, mialgia, artralgia, fadiga, dores de cabeça e lesões na pele | SMART-9N | Tempus |
| 3 | Masculino | 38 | Belo | Minas | DNG | Dores de cabeça, | SMART-9N / | Sangue |

| | | | Horizonte | Gerais | | convulsões e sintomas neurológicos não-específicos | Shotgun | total |
|----|-----------|----|--------------------|--------------|-----|---|--------------------|--------------|
| 4 | Masculino | 53 | Ouro Preto | Minas Gerais | DNG | Quadro sugestivo de neurosífilis e meningite bacteriana | SMART-9N / Shotgun | Sangue total |
| 5 | Feminino | 43 | Belo Horizonte | Minas Gerais | DFA | Mialgia, artralgia, fadiga e petéquias na pele | SMART-9N | Soro |
| 6 | Masculino | 34 | Belo Horizonte | Minas Gerais | DFA | Febre, mialgia, dores de cabeça e petéquias na pele | SMART-9N | Sangue total |
| 7 | Feminino | 27 | Ribeirão das Neves | Minas Gerais | DFA | Febre, Tosse, mialgia, fadiga e petéquias na pele | SMART-9N | Tempus, Soro |
| 8 | Feminino | 14 | Belo Horizonte | Minas Gerais | DFA | Febre, mialgia e dores de cabeça | SMART-9N | Sangue total |
| 9 | Masculino | 37 | Belo Horizonte | Minas Gerais | DFA | Febre | SMART-9N | Soro |
| 10 | Masculino | 61 | Belo Horizonte | Minas Gerais | DFA | Febre | SMART-9N | Soro |
| 11 | Feminino | 32 | Belo Horizonte | Minas Gerais | DFA | Febre e mielorradiculopatia subaguda | SMART-9N | Liquor (2) |
| 12 | Feminino | 32 | Belo Horizonte | Minas Gerais | DFA | Febre, mialgia, lesões cutâneas, imunossupressão e hepatite aguda | SMART-9N | Sangue total |
| 13 | SI | SI | SI | SI | DFA | SI | Não realizado | - |
| 14 | SI | SI | SI | SI | DFA | SI | Não realizado | - |
| 15 | Masculino | 57 | Belo Horizonte | Minas Gerais | DFA | Quadro sugestivo de leptospirose | Shotgun | Soro |
| 16 | Masculino | 49 | Belo Horizonte | Minas Gerais | DFA | Paquimeningite com acometimento encefálico e medular extenso. Lesões expansivas intramedulares, mesenfálica e temporal. | Shotgun | Soro |

SI: Sem informação

4.2 Análise do sequenciamento Shotgun

A partir do sequenciamento Shotgun não foi possível detectar a presença de nenhum microrganismo. Para verificar o que houve, a análise dos dados foi refeita sem a remoção dos genes humanos e o resultados deram de fato que 100% das sequências geradas no sequenciamento Shotgun foram de seres humanos (Figura 2), incluindo em amostras que posteriormente na análise do sequenciamento SMART-9N foram detectados genomas quase completos de notórios microrganismos associados a origem dos quadros clínico de DNG e DFA.

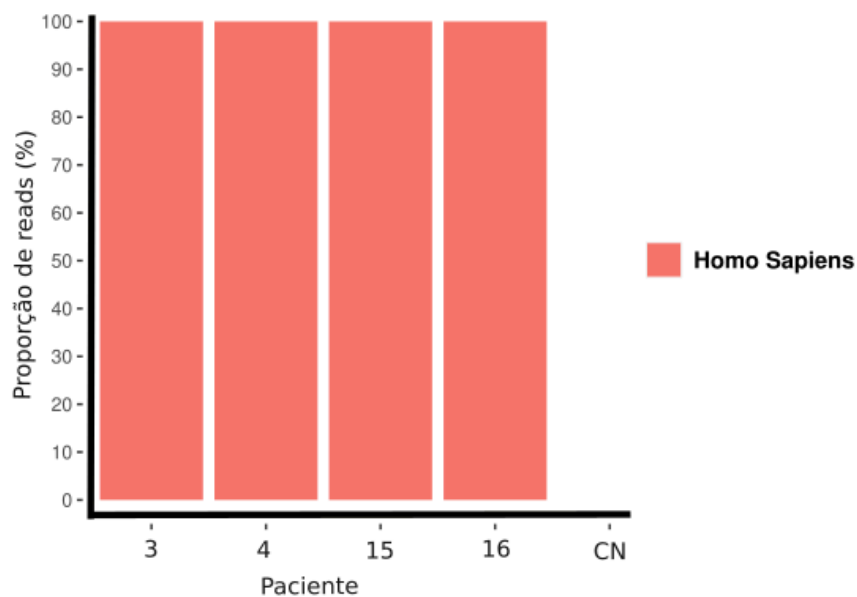


Figura 2. Metaperfil dos organismos presentes nas amostras analisadas por paciente no protocolo Shotgun

4.3 Análise do sequenciamento SMART 9N

Dentre os pacientes cujas amostras foram submetidas ao protocolo SMART-9N, 25% foram classificados clinicamente como casos de DNG (n=3), enquanto 75% foram classificados como casos de DFA (n=9) (Figura 3A). Na análise pelo DIAMOND, entre os casos de DNS, foram identificados nove microrganismos, sendo que três deles (34%) estavam associados à DNS: *Escherichia coli* (33%), *Clostridium sp.* (33%) e vírus da dengue tipo 2 (DENV-2) (33%) (Figura 3B e 3C).

Relativamente aos casos de DFA, foram identificados um total de 21 microrganismos, sendo que oito (34%) eram agentes microbianos associados a sintomas de DFA (Figura 3D). A detecção de bactérias foi predominante, sendo *Escherichia coli* (76%) a mais prevalente, seguida de *Streptococcus sp.* (8%). Além disso, foram identificados vírus em dois casos, incluindo o Vírus da Imunodeficiência Humana 1 (HIV-1) (8%) e o Pegivirus Hominis (HPgV) (8%) (Figura 3E).

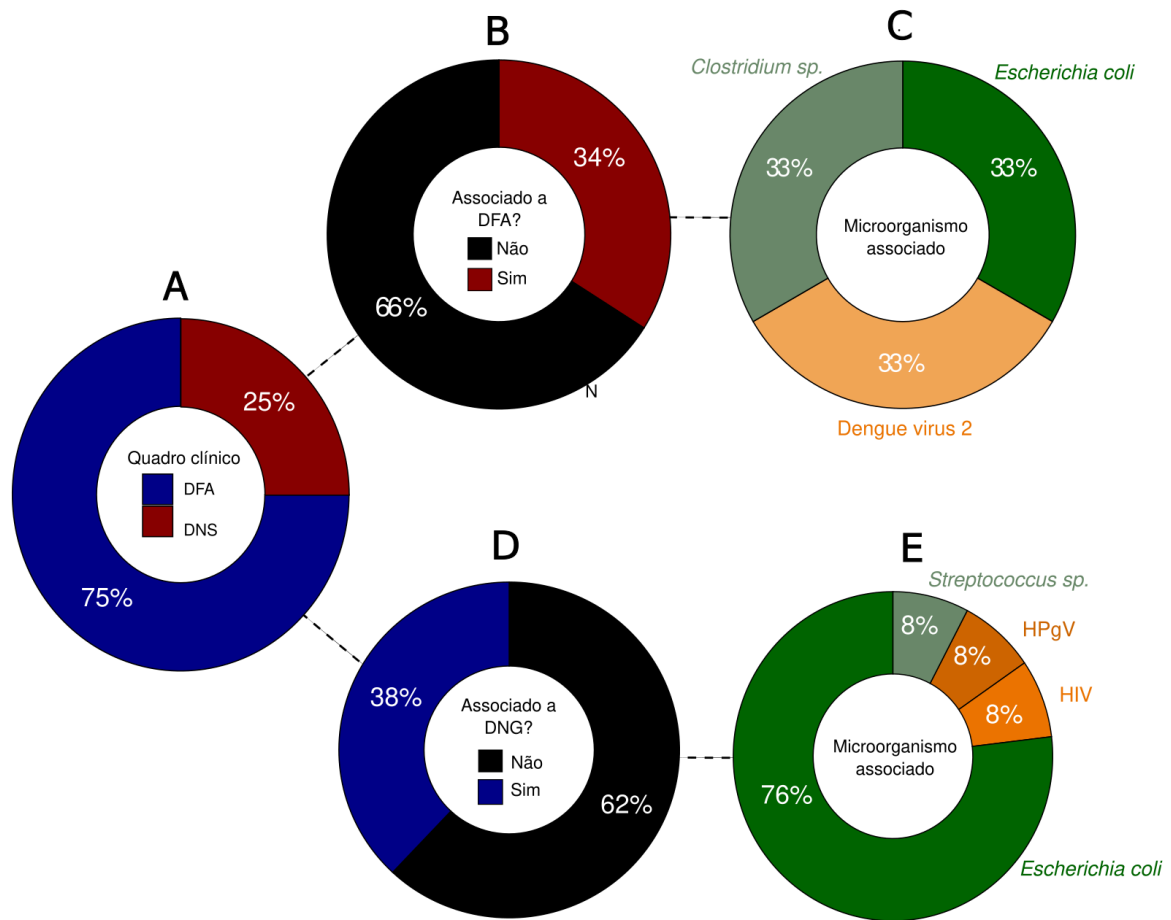


Figura 3. Visão geral da proporção de casos por classificação e microrganismos identificados. A) Porcentagem de casos por classificação clínica. B) Proporção de microrganismos associados à DNG nas amostras. C) Microrganismos associados à DNG identificados. D) Proporção de microrganismos associados à DFA nas amostras. E) Microrganismos associados a DFA identificados.

Entre as bactérias e fungos detectados, a *Escherichia coli* foi identificada em todos os tipos de amostras coletadas e foi associada a uma variedade de sintomas, incluindo febre, mialgia, artralgia, fadiga, dores de cabeça, lesões cutâneas, petéquias e sinais não específicos de neurosífilis e meningite. Embora a *E. coli* seja frequentemente considerada parte do microbioma normal, também pode atuar como um agente patogénico, particularmente quando é o único agente patogénico detectado ou dependendo do tipo de amostra em que é encontrada.

Foram identificadas espécies de *Clostridium* numa amostra de soro de um paciente internado com DNG que apresentava sintomas graves de cefaléia. Dada a variedade de doenças causadas por diferentes espécies de *Clostridium*, a sua deteção não pode ser negligenciada. Uma situação semelhante ocorreu com as espécies de *Streptococcus*, que foram encontradas numa amostra de sangue total de um paciente com DFS que apresentava

sintomas de febre, mialgia, cefaléia e petéquias. No controle negativo da sequenciamento, não foram detectados microrganismos nos contigs, o que evidencia a ausência de contaminação cruzada. Os microrganismos associados a essas doenças foram confirmados na análise por Kraken2 (Figura 4).

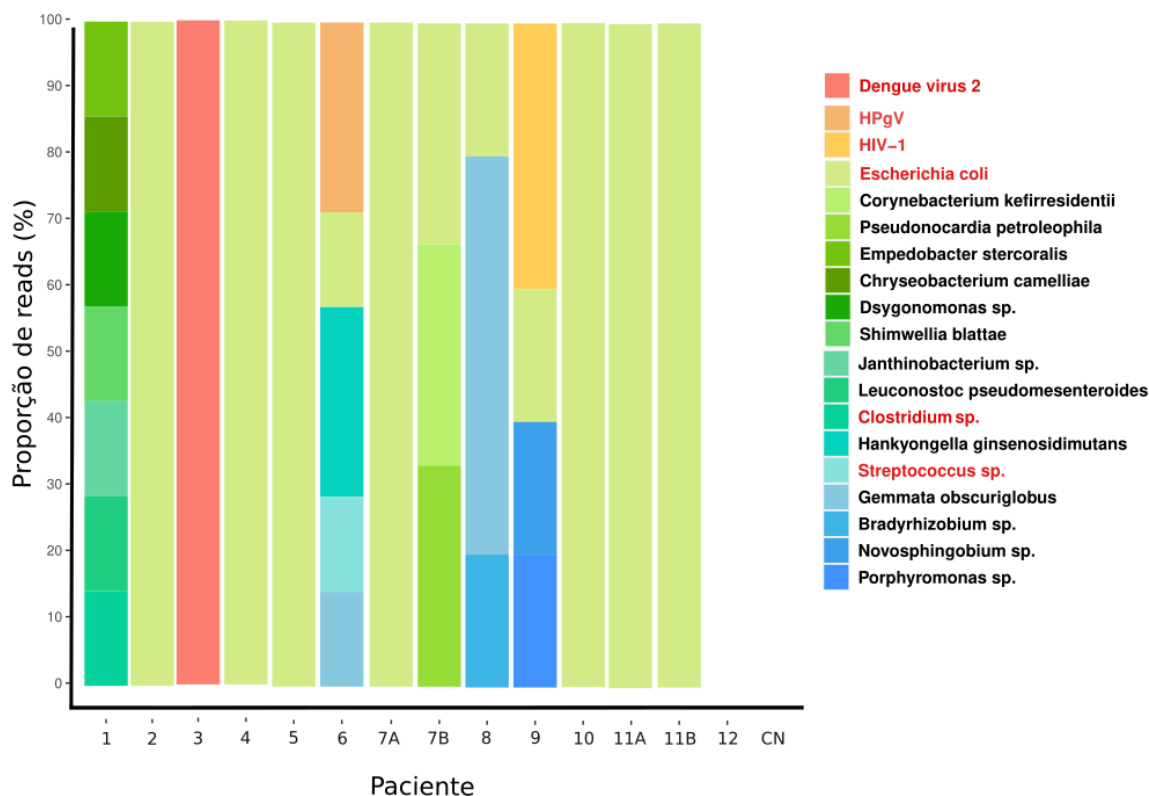


Figura 4. Metaperfil dos organismos presentes nas amostras analisadas por paciente no protocolo SMART-9N. As nomenclaturas em vermelho são microrganismos associados a causa de DFA e/ou DNG.

Em relação à análise de predição por HMM, identificamos diversos genes virais nos contigs por meio do VirSorter2. Contudo, ao verificar esses genes identificados pelo VirSorter2 utilizando o CheckV, nenhum deles foi confirmado como um gene viral, mesmo em amostras onde vírus amplamente catalogados foram detectados pelo Kraken2 e DIAMOND.

As bactérias associadas à doença apresentaram uma cobertura e profundidade muito baixas (Tabela 2), o que impediu uma análise mais aprofundada. Em contrapartida, os vírus detectados foram sequenciados com elevada qualidade, permitindo a montagem de genomas quase completos de HIV, HPgV e DENV-2, e isto permitiu a construção de uma árvore filogenética para identificar as linhagens virais associadas a estas infecções (Figura 5).

Tabela 2. Dados de sequenciamento dos microrganismos associados a doenças detectados pelo DIAMOND.

| Paciente | Tipo de amostra | Patógeno | Reads | Contigs | Cobertura | Profundidade |
|----------|-------------------|--------------------------|-------|---------|-----------|--------------|
| 1 | Soro | <i>Clostridium sp.</i> | 6554 | 1 | 0,8% | 0,25x |
| 2 | Tempus | <i>E. Coli</i> | 63 | 1 | 0,1% | 0,005x |
| 3 | Sangue total | DENV-2 | 72078 | 3 | 99,8% | 1145x |
| 4 | Sangue total | <i>E. coli</i> | 6939 | 1 | 0,1% | 0,61x |
| 5 | Soro | <i>E. coli</i> | 10968 | 1 | 0,3% | 1x |
| | | HPgV | 511 | 2 | 86,2% | 41x |
| 6 | Sangue total | <i>E. coli</i> | 4635 | 1 | 0,4% | 0,18x |
| | | <i>Streptococcus sp.</i> | 1347 | 1 | 0,1% | 0,17x |
| 7 | Tempus | <i>E. coli</i> | 173 | 1 | 0,7% | 0,012x |
| | Soro | <i>E. coli</i> | 9889 | 1 | 0,1% | 0,95x |
| 8 | Sangue total | <i>E. coli</i> | 5280 | 1 | 0,1% | 0,41x |
| 9 | Soro | HIV-1 | 52420 | 2 | 92% | 1805x |
| | | <i>E. coli</i> | 13778 | 1 | 0,2% | 1,2x |
| 10 | Soro | <i>E. coli</i> | 22622 | 1 | 0,1% | 1,6x |
| 11 | Liquor 1 | <i>E. coli</i> | 2084 | 1 | 0,1% | 0,19x |
| | Liquor 2 | <i>E. coli</i> | 1703 | 1 | 0,2% | 0,11x |
| 12 | Sangue total | - | - | - | - | - |
| - | Controle negativo | <i>E.coli</i> | 27 | 0 | - | - |

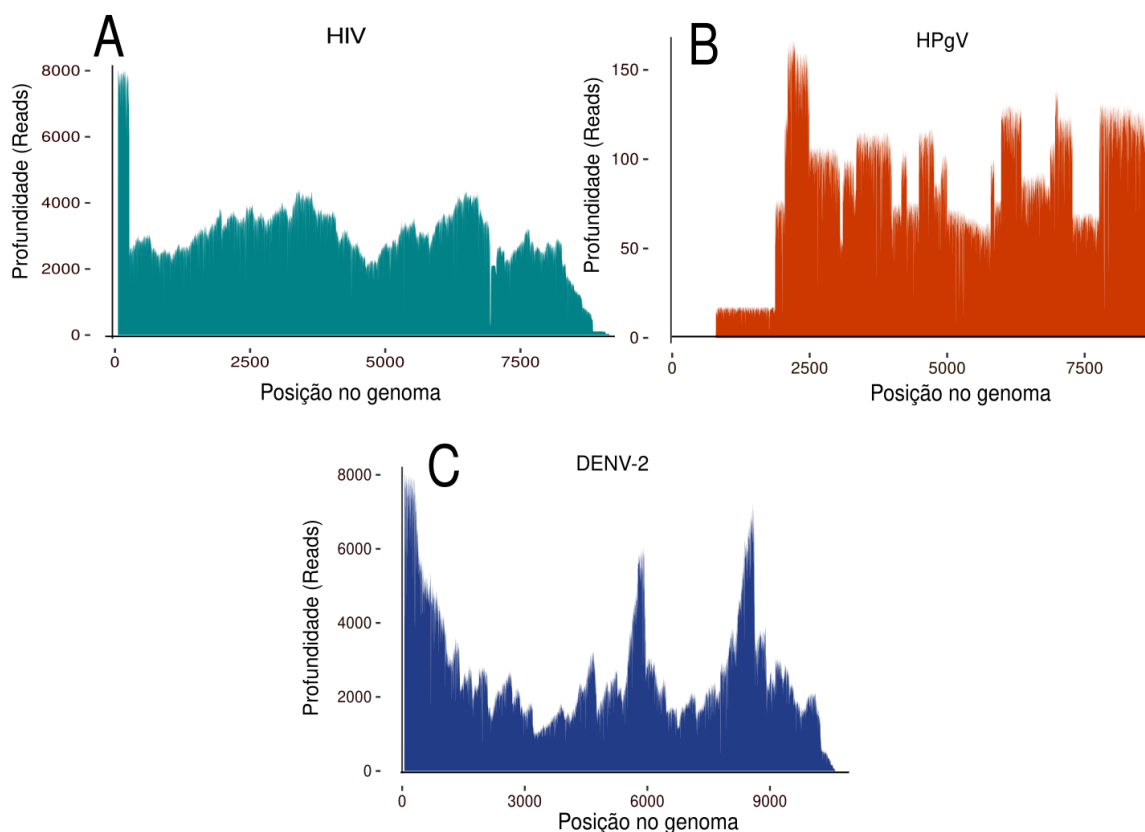


Figura 5. Cobertura do genoma dos agentes patogênicos virais identificados. (A) A cobertura do genoma do HIV, (B) A cobertura do genoma do HPgV e (C) A cobertura do genoma de DENV2.

A importância destes resultados é realçada pela identificação de subtipos virais específicos, como demonstrado pela confirmação do HIV numa amostra de soro de um

paciente diagnosticado com DFA. Este vírus foi classificado filogeneticamente como HIV tipo 1, subtipo B (Figura 6), elucidando assim o subtipo viral presente no paciente.

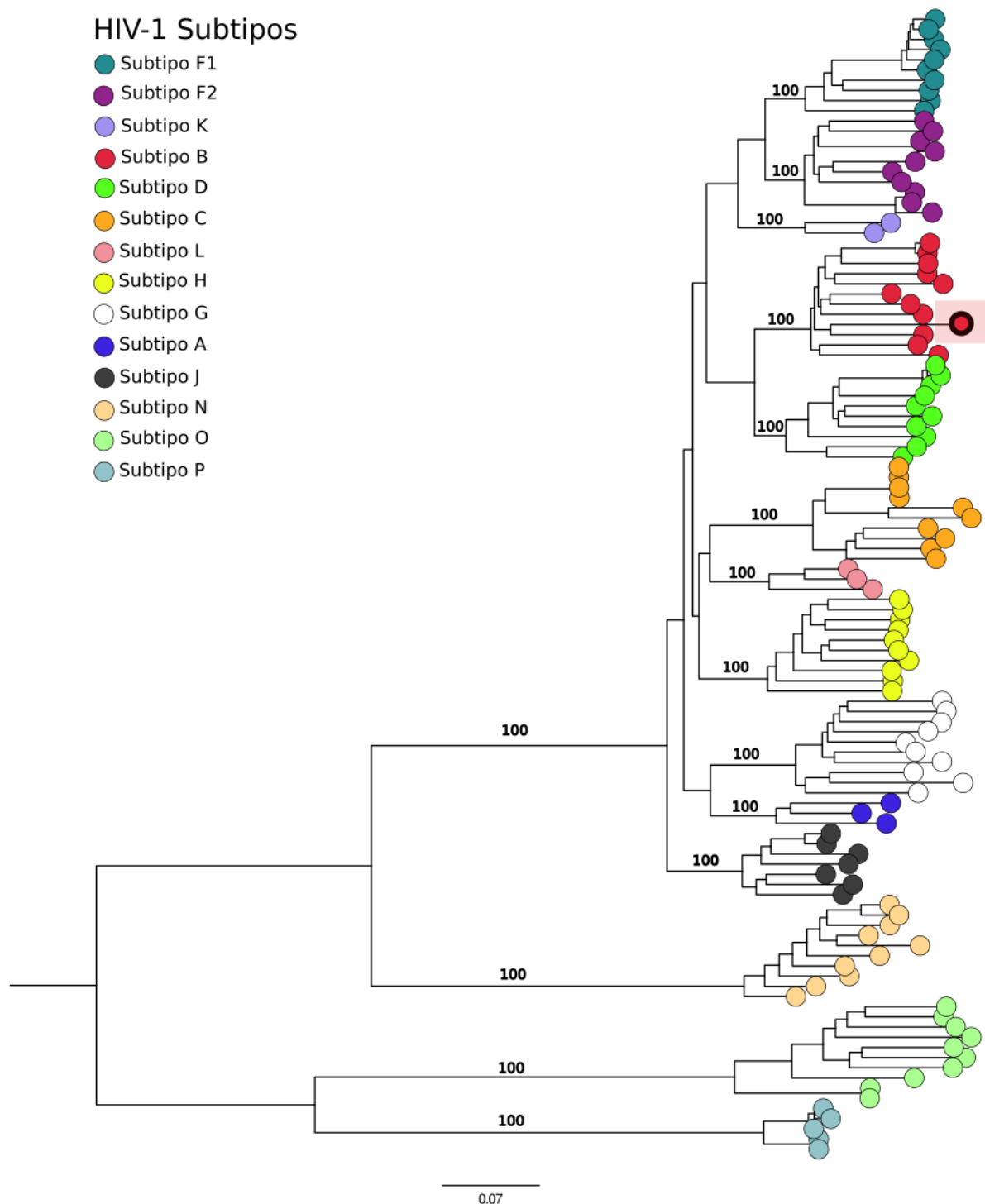


Figura 6. Árvore filogenética da análise do genoma do HIV. A amostra do estudo é a amostra destacada agrupada com subtipo B.

A amostra positiva para HPgV foi obtida a partir de uma amostra de sangue total, que também foi positiva para *Escherichia coli* e *Streptococcus sp.* Esta amostra estava associada a

sintomas típicos de DFS, incluindo febre, mialgia, dores de cabeça e petéquias na pele. A análise do genoma da amostra revelou um genoma de HPgV 86% completo. A análise filogenética subsequente classificou este vírus como pertencente ao genótipo 2B do HPgV-1 (Figura 7).

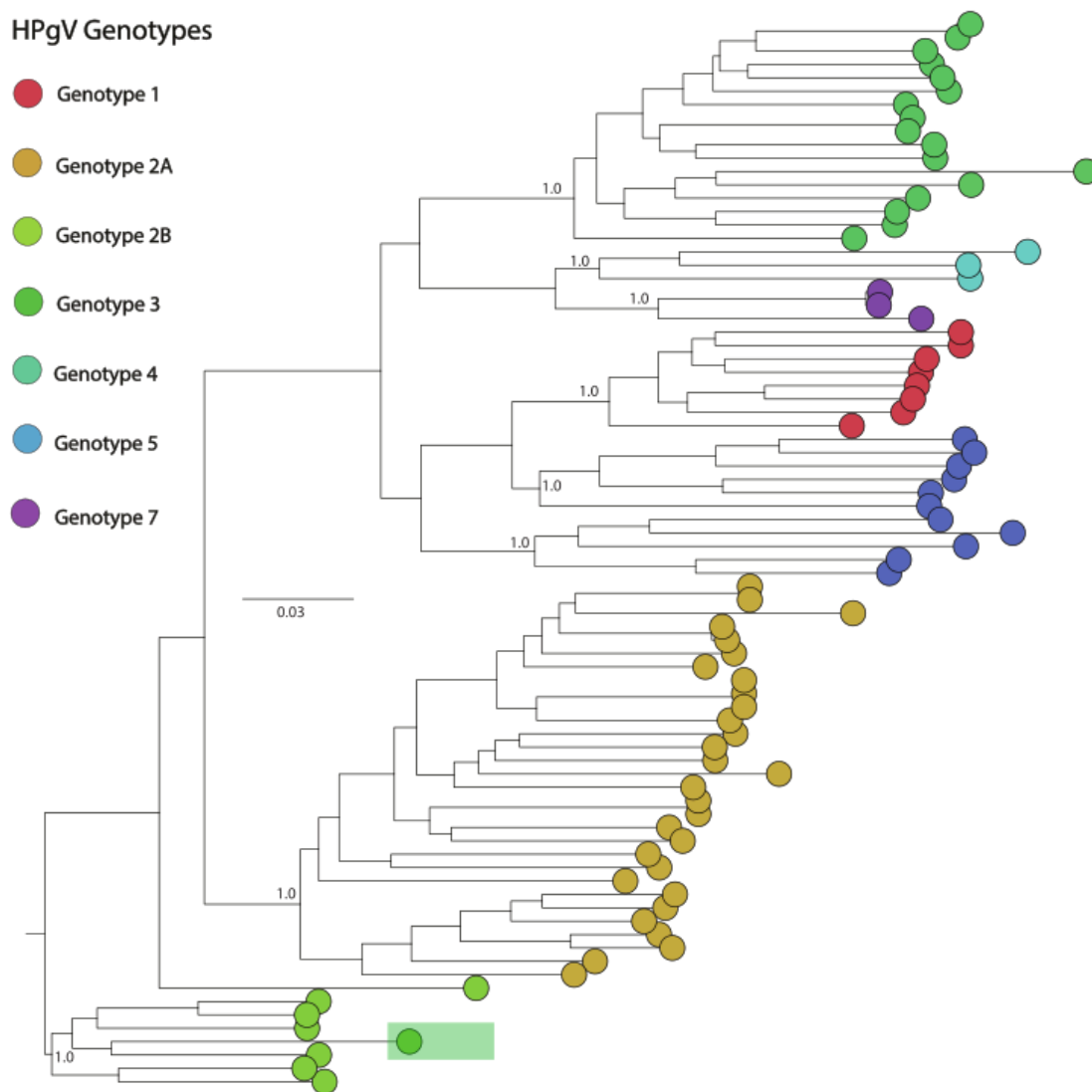


Figura 7. Árvore filogenética da análise do genoma do HPgV. A amostra do estudo é aquela que se destaca no clado do genótipo 2B.

Do caso positivo de DENV-2, foi recuperado um genoma com 99% de cobertura (Figura 5C), que foi classificado como pertencente ao genótipo II, também conhecido como linhagem cosmopolita (Figura 8). Este vírus foi detectado numa amostra de sangue total de um paciente diagnosticado com Doença Neurológica Grave (DNS), que apresentava sintomas como dores de cabeça, crises convulsivas e outros sintomas neurológicos inespecíficos.

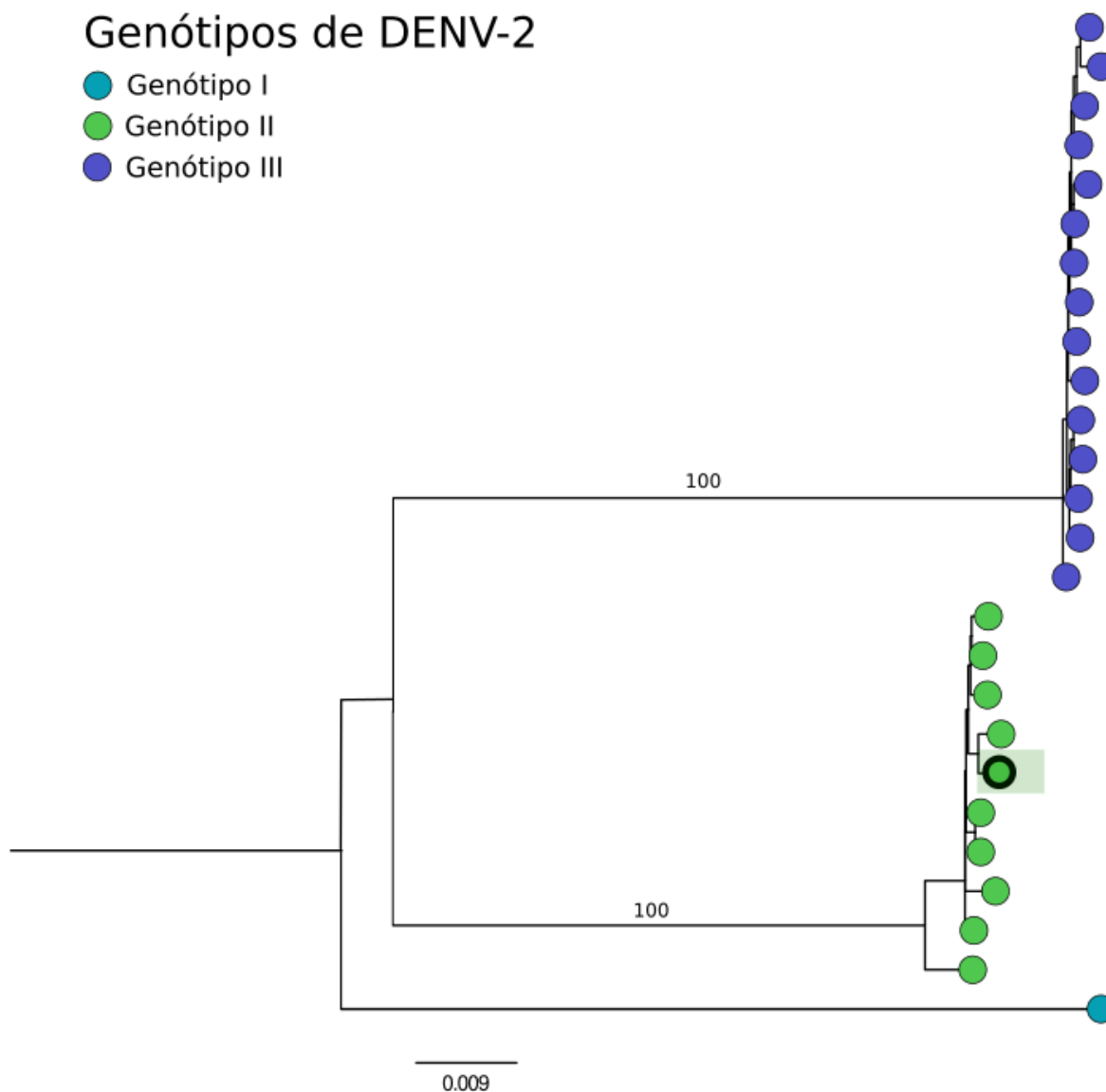


Figura 8. Árvore filogenética da análise do genoma do DENV2. Comparando os genótipos de DENV2 detectados no Brasil, a amostra do estudo é a amostra destacada que está agrupada no genótipo II.

Em relação à análise de predição por HMM, identificamos diversos genes virais nos contigs por meio do VirSorter2. Contudo, ao verificar esses genes identificados pelo VirSorter2 utilizando o CheckV, nenhum deles foi confirmado como um gene viral, mesmo em amostras onde vírus amplamente catalogados foram detectados pelo Kraken2 e DIAMOND.

5. DISCUSSÃO GERAL

A DFA e a DNG de origem desconhecida constituem um grande desafio para a medicina, e conhecer a etiologia dessas doenças é imprescindível para proporcionar um tratamento específico e bem sucedido aos pacientes, e para auxiliar nessa elucidação, a metagenômica possui um grande potencial de utilidade (Chiu and Miller, 2019) .

No estudo atual, a análise metagenômica shotgun não foi capaz de detectar microrganismos, porém a abordagem SMART-9N utilizada foi capaz de identificar diversos microrganismos que podem explicar a origem dos sintomas e a necessidade de hospitalização dos pacientes. Dentre esses, a bactéria *Escherichia coli* foi o patógeno mais detectado, e ainda que esta seja uma bactéria comum da microbiota humana, certas condições e linhagens descritas como patogênicas deste microrganismo pode levar a uma série de doenças, incluindo entre elas síndromes febris e problemas neurológicos (Kaper, Nataro and Mobley, 2004; Schuppner *et al.*, 2016; Soedarmono *et al.*, 2022). Nesse estudo, a presença de *E.coli* em amostras como swab dificilmente podem estar associada a doença do paciente, porém a detecção deste microrganismo em amostras de líquido, sangue total e/ou soro demanda uma grande atenção devido ao fato de que a presença de *E.coli* na corrente sanguínea pode acionar uma grande resposta do sistema imunológico, resultando em uma grande gama de sintomas e até mesmo levar a quadro de sepse (Daga *et al.*, 2019), portanto a detecção destes pode explicar o quadro clínico e a necessidade de hospitalização dos pacientes afetados neste estudo. Além de *E.coli*, *Streptococcus* sp. outra bactéria que pode estar associada a DFS também foi identificada. O gênero *Streptococcus* inclui mais de 35 espécies associadas a infecções, incluindo *Streptococcus pyogenes* e *Streptococcus pneumoniae*, ambos ligados à DFS (Brouwer *et al.*, 2023; Hercik *et al.*, 2017; Krzyściak *et al.*, 2013), porém, com a qualidade da sequência detectada em nosso estudo, não conseguimos inferir se foi uma espécie associada a patologias ou não.

Das amostras positivas virais, foi possível recuperar genomas quase completos, permitindo uma análise mais aprofundada, como linhagens e genótipos conhecidos, que são informações cruciais para adaptar estratégias de tratamento adequadas e obter informações sobre as características epidemiológicas do vírus na população. Esta descoberta contribui para a nossa compreensão mais ampla das dinâmicas de transmissão, auxilia no desenvolvimento de intervenções específicas e informa as iniciativas de saúde pública destinadas a controlar as propagações virais.

HIV é um agente patogênico humano bem conhecido cuja infecção leva a quadro de imunodeficiência, tornando os indivíduos mais suscetíveis a outras infecções (Bekker *et al.*,

2023), e no caso deste estudo, *E. coli* foi encontrada na mesma amostra de sangue que o HIV e é uma causa comum de infecções da corrente sanguínea em pacientes soropositivos (Taramasso *et al.*, 2019), o que pode explicar o quadro clínico. A análise filogenética classificou-a com sucesso como HIV tipo 1 subtipo B, que é o subtipo mais prevalente de HIV-1 em todo o mundo, inclusive em estudos locais feitos no estado de Minas Gerais (Da Silva, Aleixo and Tupinambás, 2022; Junqueira and Almeida, 2016). Posteriormente, esse caso de HIV também foi identificado por meio de triagem molecular na FUNED, e o paciente relatou um status positivo, mas optou por não se submeter a tratamento adicional em 2020.

Todos os microrganismos citados previamente são popularmente e previamente já relatados como microrganismos associados à DFA, porém o HPgV continua a ser um vírus com patogenicidade controversa. Enquanto alguns estudos não o associam a quaisquer doenças infecciosas humanas e o consideram um vírus comensal, outros associam as infecções por HPgV a doenças como linfoma ou encefalite cerebral (Yu *et al.*, 2022). No nosso estudo, a amostra positiva de HPgV-1 estava co-infectada com *Streptococcus*, o que sugere que o HPgV pode não ser a principal causa de infecção por DFA; no entanto, a associação do HPgV com sintomas semelhantes aos dos arbovírus, tais como febre, mialgia, cefaleias e petéquias na pele, realça a necessidade de mais investigação para esclarecer esta correlação. Essa é a primeira amostra de HPgV-1 sequenciada no Estado de Minas Gerais, e o genótipo encontrado e o genótipo mais prevalentes em outras regiões do país, em análises feitas em pacientes portadores de hepatites e transplantados ([Savassi-Ribas et al. 2020](#); [da Silva et al. 2023](#)).

Para além dos casos associados a DFS e a *Escherichia coli*, dois outros microrganismos foram associados à DNG: *Clostridium sp.* e DENV-2. Embora não tenha sido possível identificar a espécie específica de *Clostridium* nesta análise, a presença de uma bactéria do gênero *Clostridium*, que se sabe incluir várias espécies associadas a sintomas neurológicos graves (por exemplo, *Clostridium perfringens*, *difficile*, *tetani* e *botulinum*), suscita preocupações em termos de saúde pública (Cai, Kumar and Singh, 2021; Finsterer and Hess, 2007). No entanto, a detecção de microrganismos de espécies não-inferíveis deve-se à qualidade da sequência e, para resolver este problema, podem ser testadas várias outras abordagens metagenômicas de nanoporos para melhorar ainda mais a detecção de agentes patogênicos.

O DENV, particularmente os sorotipos 2 e 3, é um vírus notoriamente conhecido por causar complicações neurológicas, incluindo convulsões, dores de cabeça e outros sintomas neurológicos, como no caso deste estudo (Kulkarni, Pujari and Gupta, 2021; Trivedi and

Chakravarty, 2022). Este DENV-2 positivo foi classificado como linhagem do genótipo II, também conhecida como linhagem cosmopolita. Essa linhagem tem se disseminado recentemente no Brasil, com relatos de circulação em diversas regiões (Amorim *et al.*, 2023; Giovanetti *et al.*, 2022; Souza *et al.*, 2023). Embora a disseminação dessa linhagem em Minas Gerais careça de pesquisas extensas, os achados deste estudo sugerem que o genótipo II do DENV-2 foi inicialmente introduzido na região em abril de 2022, o que é consistente com o fato de que esse genótipo foi o mais prevalente em Minas Gerais no período subsequente (Secretaria de Saúde do Estado de Minas Gerais, 2023).

No geral, esta análise metagenômica de amostras de pacientes que apresentam sintomas clínicos de DFA e DNG forneceu informações valiosas sobre os diversos agentes patogênicos associados a estas doenças. O processo de análise utilizado identificou uma série de microrganismos, incluindo bactérias, fungos e, principalmente, vírus, para elucidar seu papel potencial nessas doenças, além de diversos microrganismos comuns da microbiota ou não relacionados a patologias. Dentre as bactérias, a *Escherichia coli* foi o agente patogênico predominante nos casos de AFI, assim como HIV e *Streptococcus* foram identificados como contribuintes para os sintomas de DFS. Nos casos de DNS, a nossa análise revelou agentes patogênicos como o *Clostridium sp.* e o DENV-2, realçando os diversos fatores etiológicos associados aos sintomas neurológicos graves.

6. CONCLUSÃO

Em conclusão, a nossa análise metagenômica produziu conhecimentos significativos sobre os agentes patogênicos associados à DFA e à DNG. Esta abordagem abrangente delineou um espectro de bactérias, fungos e vírus, esclarecendo os seus potenciais papéis na patogênese destas condições. A identificação acurada desses microrganismos potencialmente patogênicos, juntamente com a caracterização de seus perfis genéticos e o seu papel nos cenários de co-infecção, amplia significativamente nossa capacidade de desenvolver estratégias terapêuticas precisas e direcionadas para o tratamento adequado. Além disso, este estudo destaca a importância crítica de investigação contínua sobre a patogenicidade de microrganismos ainda pouco estudados, como o HPgV. Os resultados obtidos não apenas contribuem para uma compreensão mais aprofundada da dinâmica das doenças infecciosas, mas também promovem a aprimoração das metodologias metagenômicas em aplicações de diagnóstico clínico e saúde pública, estabelecendo fundamentos sólidos para futuras investigações exploratórias e aplicadas neste campo.

REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, M. T. *et al.* Emergence of a New Strain of DENV-2 in South America: Introduction of the Cosmopolitan Genotype through the Brazilian-Peruvian Border. **Tropical medicine and infectious disease**, v. 8, n. 6, 17 Jun. 2023.

BEKKER, L.-G. *et al.* HIV infection. **Nature reviews. Disease primers**, v. 9, n. 1, p. 42, 17 Aug. 2023.

BREITWIESER, F. P.; LU, J.; SALZBERG, S. L. A review of methods and databases for metagenomic classification and assembly. **Briefings in Bioinformatics**, v. 20, n. 4, p. -1125–1136, 19 Jul. 2019.

BRESSAN, C. DA S. *et al.* Challenges of acute febrile illness diagnosis in a national infectious diseases center in Rio de Janeiro: 16-year experience of syndromic surveillance. **PLoS Neglected Tropical Diseases**, v. 17, n. 4, p. e0011232, 3 Apr. 2023.

BROUWER, S. *et al.* Pathogenesis, epidemiology and control of Group A Streptococcus infection. **Nature Reviews Microbiology**, v. 21, n. 7, p. 431–447, Jul. 2023.

BUCHFINK, B.; REUTER, K.; DROST, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. **Nature Methods**, v. 18, n. 4, p. 366–368, 7 Apr. 2021.

CAI, S.; KUMAR, R.; SINGH, B. R. Clostridial neurotoxins: structure, function and implications to other bacterial toxins. **Microorganisms**, v. 9, n. 11, 23 Oct. 2021.

CHEN, S. *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. **Bioinformatics**, v. 34, n. 17, p. i884–i890, 1 Sep. 2018.

CHEN, X. *et al.* A systematic review of neurological symptoms and complications of COVID-19. **Journal of Neurology**, v. 268, n. 2, p. 392–402, Feb. 2021.

CHEN, Y. *et al.* Efficient assembly of nanopore reads via highly accurate and intact error correction. **Nature Communications**, v. 12, n. 1, p. 60, 4 Jan. 2021.

CHIU, C. Y.; MILLER, S. A. Clinical metagenomics. **Nature Reviews. Genetics**, v. 20, n. 6, p. 341–355, Jun. 2019.

CIUFFREDA, L.; RODRÍGUEZ-PÉREZ, H.; FLORES, C. Nanopore sequencing and its application to the study of microbial communities. **Computational and structural biotechnology journal**, v. 19, p. 1497–1511, 7 Mar. 2021.

CLARO, I. M. *et al.* Rapid viral metagenomics using SMART-9N amplification and nanopore sequencing. **Wellcome Open Research**, v. 6, p. 241, 2021.

DA SILVA, G. J.; ALEIXO, A. W.; TUPINAMBÁS, U. Prevalence of HIV-1 transmitted drug resistance and its impact on the effectiveness of antiretroviral therapy - Minas Gerais state, Brazil / Prevalência de resistência transmitida do HIV-1 e seu impacto na efetividade da terapia antirretroviral – Minas Gerais, Brasil. **Brazilian Journal of Health Review**, v. 5, n. 1, p. 1044–1059, 17 Jan. 2022.

DAGA, A. P. *et al.* Escherichia coli Bloodstream Infections in Patients at a University Hospital: Virulence Factors and Clinical Characteristics. **Frontiers in cellular and infection microbiology**, v. 9, p. 191, 6 Jun. 2019.

DENG, Q. *et al.* Nanopore-based metagenomic sequencing for the rapid and precise detection of pathogens among immunocompromised cancer patients with suspected infections. **Frontiers in cellular and infection microbiology**, v. 12, p. 943859, 20 Sep. 2022.

FINSTERER, J.; HESS, B. Neuromuscular and central nervous system manifestations of Clostridium perfringens infections. **Infection**, v. 35, n. 6, p. 396–405, Dec. 2007.

GHURYE, J. S.; CEPEDA-ESPINOZA, V.; POP, M. Metagenomic assembly: overview, challenges and applications. **The Yale journal of biology and medicine**, v. 89, n. 3, p. 353–362, 30 Sep. 2016.

GIOVANETTI, M. *et al.* Emergence of dengue virus serotype 2 cosmopolitan genotype, Brazil. **Emerging Infectious Diseases**, v. 28, n. 8, p. 1725–1727, Aug. 2022.

GUO, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. **Microbiome**, v. 9, n. 1, p. 37, 1 Feb. 2021.

HERCIK, C. *et al.* A diagnostic and epidemiologic investigation of acute febrile illness (AFI) in Kilombero, Tanzania. **Plos One**, v. 12, n. 12, p. e0189712, 29 Dec. 2017.

IROH TAM, P.-Y.; OBARO, S. K.; STORCH, G. Challenges in the Etiology and Diagnosis of Acute Febrile Illness in Children in Low- and Middle-Income Countries. **Journal of the Pediatric Infectious Diseases Society**, v. 5, n. 2, p. 190–205, Jun. 2016.

JUNQUEIRA, D. M.; ALMEIDA, S. E. DE M. HIV-1 subtype B: Traces of a pandemic. **Virology**, v. 495, p. 173–184, Aug. 2016.

KALANTAR, K. L. *et al.* IDseq-An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. **GigaScience**, v. 9, n. 10, 15 Oct. 2020.

KAPER, J. B.; NATARO, J. P.; MOBLEY, H. L. Pathogenic Escherichia coli. **Nature Reviews Microbiology**, v. 2, n. 2, p. 123–140, Feb. 2004.

KATOH, K. *et al.* MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, v. 30, n. 14, p. 3059–3066, 15 Jul. 2002.

KIGOZI, B. K. *et al.* Investigating the etiology of acute febrile illness: a prospective clinic-based study in Uganda. **BMC Infectious Diseases**, v. 23, n. 1, p. 411, 16 Jun. 2023.

KOLMOGOROV, M. *et al.* Assembly of long, error-prone reads using repeat graphs. **Nature Biotechnology**, v. 37, n. 5, p. 540–546, May 2019.

KOLMOGOROV, M. *et al.* metaFlye: scalable long-read metagenome assembly using repeat graphs. **Nature Methods**, v. 17, n. 11, p. 1103–1110, Nov. 2020.

KOREN, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. **Genome Research**, v. 27, n. 5, p. 722–736, May 2017.

KRZYŚCIAK, W. *et al.* The pathogenicity of the Streptococcus genus. **European Journal of Clinical Microbiology & Infectious Diseases**, v. 32, n. 11, p. 1361–1376, Nov. 2013.

KULKARNI, R.; PUJARI, S.; GUPTA, D. Neurological manifestations of dengue fever. **Annals of Indian Academy of Neurology**, v. 24, n. 5, p. 693–702, 28 May 2021.

LARSSON, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. **Bioinformatics**, v. 30, n. 22, p. 3276–3278, 15 Nov. 2014.

LATORRE-PÉREZ, A. *et al.* Assembly methods for nanopore-based metagenomic sequencing: a comparative study. **Scientific Reports**, v. 10, n. 1, p. 13588, 12 Aug. 2020.

LI, H. Minimap2: pairwise alignment for nucleotide sequences. **Bioinformatics**, v. 34, n. 18, p. 3094–3100, 15 Sep. 2018.

MINH, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. **Molecular Biology and Evolution**, v. 37, n. 5, p. 1530–1534, 1 May 2020.

MOREIRA, J. *et al.* Epidemiology of acute febrile illness in Latin America. **Clinical Microbiology and Infection**, v. 24, n. 8, p. 827–835, Aug. 2018.

NAYFACH, S.; CAMARGO, A. P.; *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. **Nature Biotechnology**, v. 39, n. 5, p. 578–585, May 2021.

NAYFACH, S.; PÁEZ-ESPINO, D.; *et al.* Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. **Nature Microbiology**, v. 6, n. 7, p. 960–970, Jul. 2021.

NISSEN, J. N. *et al.* Improved metagenome binning and assembly using deep variational autoencoders. **Nature Biotechnology**, v. 39, n. 5, p. 555–560, May 2021.

NOOIJ, S. *et al.* Overview of virus metagenomic classification methods and their biological applications. **Frontiers in Microbiology**, v. 9, p. 749, 23 Apr. 2018.

OLIVEIRA, L. S.; GRUBER, A. Rational design of profile hidden markov models for viral classification and discovery. *In*: HELDER I., N. (Ed.). . **Bioinformatics**. Brisbane (AU): Exon Publications, 2021.

RAMBAUT, A. FigTree v1.4.4. Disponível em: <http://tree.bio.ed.ac.uk/software/figtree/>. Acesso em 28-12-2023

RAUSCH, P. *et al.* Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. **Microbiome**, v. 7, n. 1, p. 133, 14 Sep. 2019.

SCHUPPNER, R. *et al.* Neurological Sequelae in Adults After E coli O104: H4 Infection-Induced Hemolytic-Uremic Syndrome. **Medicine**, v. 95, n. 6, p. e2337, Feb. 2016.

SECRETARIA DE SAUDE DO ESTADO DE MINAS GERAIS. **Epidemia de Arboviroses**. Disponível em: <https://www.saude.mg.gov.br/component/gmg/story/18774-epidemia-arboviroses>>. Acesso em: 11 apr. 2024.

SHAFFER, M. *et al.* DRAM for distilling microbial metabolism to automate the curation of microbiome function. **Nucleic Acids Research**, v. 48, n. 16, p. 8883–8900, 18 Sep. 2020.

SOEDARMONO, P. *et al.* The characteristics of bacteremia among patients with acute febrile illness requiring hospitalization in Indonesia. **Plos One**, v. 17, n. 9, p. e0273414, 8 Sep. 2022.

SOUZA, U. J. B. DE *et al.* Circulation of Dengue Virus Serotype 1 Genotype V and Dengue Virus Serotype 2 Genotype III in Tocantins State, Northern Brazil, 2021-2022. **Viruses**, v. 15, n. 11, 24 Oct. 2023.

TARAMASSO, L. *et al.* Bloodstream infections in patients living with HIV in the modern cART era. **Scientific Reports**, v. 9, n. 1, p. 5418, 1 Apr. 2019.

TRIVEDI, S.; CHAKRAVARTY, A. Neurological complications of dengue fever. **Current Neurology and Neuroscience Reports**, v. 22, n. 8, p. 515–529, Aug. 2022.

VILSKER, M. *et al.* Genome Detective: an automated system for virus identification from high-throughput sequencing data. **Bioinformatics**, v. 35, n. 5, p. 871–873, 1 Mar. 2019.

WICK, R. R.; JUDD, L. M.; HOLT, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. **Genome Biology**, v. 20, n. 1, p. 129, 24 Jun. 2019.

WOOD, D. E.; LU, J.; LANGMEAD, B. Improved metagenomic analysis with Kraken 2. **Genome Biology**, v. 20, n. 1, p. 257, 28 Nov. 2019.

WOUK, J. *et al.* Viral infections and their relationship to neurological disorders. **Archives of Virology**, v. 166, n. 3, p. 733–753, Mar. 2021.

YU, Y. *et al.* Review of human pegivirus: Prevalence, transmission, pathogenesis, and clinical implication. **Virulence**, v. 13, n. 1, p. 324–341, Dec. 2022.

APÊNDICES

Genomas utilizados para a construção das árvores filogenéticas

Tabela 3. Código do GenBank e a linhagem de cada genoma utilizado para a construção das árvores filogenéticas.

| HIV | | HPgV | | DENV-2 | |
|-------------|----------|------------|-------------|------------|--------------|
| Código | Linhagem | Código | Linhagem | Código | Linhagem |
| KT276260.F1 | F1 | AB003290.1 | Genotype 3 | MT929639.1 | Genotype_III |
| KT276257.F1 | F1 | AB003289.1 | Genotype 2A | MT929620.1 | Genotype_III |
| KY989957.F1 | F1 | AB003291.1 | Genotype 6 | MT929624.1 | Genotype_III |
| KY514084.F1 | F1 | AB003292.1 | Genotype 4 | MT929615.1 | Genotype_III |
| KY989955.F1 | F1 | AB003293.1 | Genotype 3 | MT929635.1 | Genotype_III |
| KY514085.F1 | F1 | AB008335.1 | Genotype 3 | MT929610.1 | Genotype_III |
| KT427774.F1 | F1 | AB008336.1 | Genotype 6 | MT929630.1 | Genotype_III |
| KT427814.F1 | F1 | AB008342.1 | Genotype 3 | MT929631.1 | Genotype_III |
| MK041565.F1 | F1 | AB013500.1 | Genotype 1 | MT929638.1 | Genotype_III |
| MK177827.F1 | F1 | AB018667.1 | Genotype 4 | MT929640.1 | Genotype_III |
| KU749420.F2 | F2 | AB021287.1 | Genotype 4 | MT929629.1 | Genotype_III |
| MN153485.F2 | F2 | AF006500.1 | Genotype 3 | MT929586.1 | Genotype_III |
| KU749422.F2 | F2 | AF031827.1 | Genotype 2A | MT929643.1 | Genotype_III |
| MT349417.F2 | F2 | AF031828.1 | Genotype 2A | MT929628.1 | Genotype_III |
| JX140673.F2 | F2 | AF031829.1 | Genotype 2A | MT929633.1 | Genotype_III |
| MH705144.F2 | F2 | AF081782.1 | Genotype 2A | MN589883.1 | Genotype_III |
| MN153483.F2 | F2 | AF104403.1 | Genotype 2A | HQ012537.1 | Genotype_III |
| JX140672.F2 | F2 | AF121950.1 | Genotype 2A | MN589884.1 | Genotype_III |
| KX228809.F2 | F2 | AF309966.1 | Genotype 2A | MN589880.1 | Genotype_III |
| MK086131.F2 | F2 | AX338086.1 | Genotype 2A | JX669487.1 | Genotype_III |
| AJ249235.K | K | AY196904.1 | Genotype 2A | JX669488.1 | Genotype_III |
| AJ249239.K | K | AY949771.1 | Genotype 5 | OR389017.1 | Genotype_III |
| MG571988.B | B | D87255.1 | Genotype 2A | OK605761.1 | Genotype_III |
| MG571987.B | B | D87262.1 | Genotype 3 | HQ012535.1 | Genotype_III |
| MG571991.B | B | D87263.1 | Genotype 3 | MN589882.1 | Genotype_III |
| MG572010.B | B | D87709.1 | Genotype 3 | HQ012538.1 | Genotype_III |
| MG571982.B | B | D87710.1 | Genotype 3 | OR389306.1 | Genotype_III |
| MG571981.B | B | D87711.1 | Genotype 3 | HQ012534.1 | Genotype_III |
| MG571999.B | B | D87712.1 | Genotype 3 | OR389305.1 | Genotype_III |
| sample | sample | D87713.1 | Genotype 3 | HQ012533.1 | Genotype_III |
| MG571983.B | B | D87714.1 | Genotype 3 | OP599720.1 | Genotype_II |
| MG572002.B | B | D87715.1 | Genotype 3 | OP599699.1 | Genotype_II |
| MK041586.B | B | D90600.1 | Genotype 2A | OR025669.1 | Genotype_II |
| OQ686013.D | D | D90601.1 | Genotype 3 | OR025670.1 | Genotype_II |
| OQ686012.D | D | HQ331171.1 | Genotype 4 | OR138998.1 | Genotype_II |
| OQ686008.D | D | HQ331192.1 | Genotype 7 | OP599710.1 | Genotype_II |
| MW006081.D | D | HQ331209.1 | Genotype 7 | OM744143.1 | Genotype_II |
| MF109710.D | D | HQ331210.1 | Genotype 7 | Sample_45 | Sample |
| MF109689.D | D | HQ331233.1 | Genotype 7 | OQ622206.1 | Genotype_II |
| MF109713.D | D | HQ331234.1 | Genotype 7 | OQ511274.1 | Genotype_II |

| | | | | | |
|------------|---|-------------|-------------|------------|-------------|
| MF109532.D | D | HQ331235.1 | Genotype 7 | OP599661.1 | Genotype_II |
| KJ787684.D | D | JN127373.1 | Genotype 2B | OP599708.1 | Genotype_II |
| KJ787683.D | D | KC618398.1 | Genotype 5 | OP599709.1 | Genotype_II |
| MW262775.C | C | KC618399.1 | Genotype 5 | OP599724.1 | Genotype_II |
| MW262771.C | C | KC618400.1 | Genotype 5 | OP599733.1 | Genotype_II |
| MW262772.C | C | KC618401.1 | Genotype 5 | OP599746.1 | Genotype_II |
| MW262774.C | C | KP710598.1 | Genotype 1 | OP599729.1 | Genotype_II |
| OR736070.C | C | KP710599.1 | Genotype 1 | OP599726.1 | Genotype_II |
| OR736069.C | C | KP710600.1 | Genotype 5 | OP599727.1 | Genotype_II |
| MK041564.C | C | KP710601.1 | Genotype 1 | OP599742.1 | Genotype_II |
| MK041561.C | C | KP710602.1 | Genotype 1 | OP599747.1 | Genotype_II |
| MK041563.C | C | KP710603.1 | Genotype 5 | JQ686088.2 | Genotype_I |
| MK041550.C | C | KP710604.1 | Genotype 1 | | |
| AF286236.L | L | KP710605.1 | Genotype 5 | | |
| AF457101.L | L | KP710606.1 | Genotype 5 | | |
| MN271384.L | L | MH053115.1 | Genotype 2A | | |
| AF190127.H | H | MH053119.1 | Genotype 2A | | |
| KY392777.H | H | MH053120.1 | Genotype 2A | | |
| FJ711703.H | H | MH053121.1 | Genotype 2A | | |
| KU168279.H | H | MH179063.1 | Genotype 2A | | |
| AF005496.H | H | MK234885.1 | Genotype 2B | | |
| AF190128.H | H | MK684252.1 | Genotype 2A | | |
| KY392778.H | H | MN215897.1 | Genotype 2A | | |
| KY392779.H | H | MN215898.1 | Genotype 2A | | |
| MN486046.H | H | MN215900.1 | Genotype 2A | | |
| KU168273.H | H | MN215901.1 | Genotype 2B | | |
| KT276261.G | G | MN215902.1 | Genotype 2A | | |
| MK254639.G | G | MN215903.1 | Genotype 2A | | |
| OQ121873.G | G | MN215906.1 | Genotype 2B | | |
| KY275364.G | G | MN215907.1 | Genotype 2A | | |
| MF109598.G | G | MN215908.1 | Genotype 2A | | |
| MF109568.G | G | MN215909.1 | Genotype 2B | | |
| MK254637.G | G | MN215910.1 | Genotype 2B | | |
| OQ121858.G | G | MN215911.1 | Genotype 2B | | |
| OQ121909.G | G | MN551061.1 | Genotype 1 | | |
| KU685592.G | G | MN551062.1 | Genotype 5 | | |
| JQ403028.A | A | MN551063.1 | Genotype 2A | | |
| MH705157.A | A | MN551064.1 | Genotype 1 | | |
| DQ396400.A | A | MN551065.1 | Genotype 2A | | |
| AF082394.J | J | MW526240.1 | Genotype 2B | | |
| AF082395.J | J | NC_001710.1 | Genotype 2A | | |
| KU168280.J | J | U36380.1 | Genotype 5 | | |
| GU237072.J | J | U44402.1 | Genotype 2A | | |
| KU310620.J | J | U45966.1 | Genotype 2A | | |
| EF614151.J | J | U63715.1 | Genotype 2B | | |
| KY392776.J | J | U75356.1 | Genotype 2A | | |

| | | | |
|------------|---|------------------------|-------------|
| DQ017382.N | N | U94695.1 | Genotype 2A |
| DQ017383.N | N | sample_35_HEM_2 022 | Sample |
| AY532635.N | N | | |
| AJ271370.N | N | | |
| MF767262.N | N | | |
| GQ324958.N | N | | |
| AJ006022.N | N | | |
| JN572926.N | N | | |
| GQ324959.N | N | | |
| KY498771.N | N | | |
| KU168283.O | O | | |
| MH705149.O | O | | |
| MH705148.O | O | | |
| JX245015.O | O | | |
| MH705150.O | O | | |
| JN571034.O | O | | |
| JX245014.O | O | | |
| KU168292.O | O | | |
| KU168298.O | O | | |
| KY112585.O | O | | |
| GU111555.P | P | | |
| KY953207.P | P | | |
| GQ328744.P | P | | |
| HQ179987.P | P | | |
| JA446000.P | P | | |

Artigo publicado

Os resultados da análise com o protocolo SMART-9N já foram publicados em artigo científico. Segue o artigo abaixo.



Article

Exploring Microorganisms Associated to Acute Febrile Illness and Severe Neurological Disorders of Unknown Origin: A Nanopore Metagenomics Approach

Keldenn Melo Farias Moreno^{1,†}, Virgínia Antunes de Andrade^{2,†}, Felipe Campos de Melo Iani³, Vagner Fonseca⁴, Maurício Teixeira Lima^{1,5}, Emerson de Castro Barbosa⁵, Luiz Marcelo Ribeiro Tomé⁵, Natália Rocha Guimarães⁵, Hegger Machado Fritsch^{1,6}, Talita Adelino³, Tatiana Oliveira Fereguetti², Maíra Cardoso Aspahan², Tereza Gamarano Barros², Luiz Carlos Junior Alcantara⁵ and Marta Giovanetti^{7,8,*}

- ¹ Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte 31270-901, Brazil; keldennfmoreno@gmail.com (K.M.F.M.); maurili15@hotmail.com (M.T.L.); hegger.fritsch@gmail.com (H.M.F)
 - ² Eduardo de Menezes Hospital, Belo Horizonte 30622-020, Brazil; virginiaantunesdeandrade@gmail.com (V.A.d.A.); tataniofereguetti@gmail.com (T.O.F.); maiaaspahan@gmail.com (M.C.A.); teresa.barros@fhemig.mg.gov.br (T.G.B)
 - ³ Central Public Health Laboratory of the State of Minas Gerais, Ezequiel Dias Foundation, Belo Horizonte 30510-010, Brazil; felipe.iani@funed.mg.gov.br (F.C.d.M.I.); talitadelino@gmail.com (T.A.)
 - ⁴ Department of Exact and Earth Sciences, University of the State of Bahia, Salvador 41150-000, Brazil; vagner.fonseca@gmail.com
 - ⁵ René Rachou Institute, Oswaldo Cruz Foundation, Belo Horizonte 30190-002, Brazil; emersoncb7@gmail.com (E.d.C.B.); luiz.tome@funed.mg.gov.br (L.M.R.T.); natyroguiaman@yahoo.com.br (N.R.G.); luiz.alcantara@fiocruz.br (L.C.J.A.)
 - ⁶ Morphogenesis and Antigenicity of HIV and Hepatitis Viruses, University of Tours, 37032 Tours, France
 - ⁷ Department of Sciences and Technologies for Sustainable Development and One Health, Università Campus Bio-Medico di Roma, 00128 Rome, Italy
 - ⁸ Oswaldo Cruz Foundation, Rio de Janeiro 21040-900, Brazil
- * Correspondence: giovanetti.marta@gmail.com
 † These authors contributed equally to this work.



Citation: Moreno, K.M.F.; de Andrade, V.A.; de Melo Iani, F.C.; Fonseca, V.; Lima, M.T.; de Castro Barbosa, E.; Tomé, L.M.R.; Guimarães, N.R.; Fritsch, H.M.; Adelino, T.; et al. Exploring Microorganisms Associated to Acute Febrile Illness and Severe Neurological Disorders of Unknown Origin: A Nanopore Metagenomics Approach. *Genes* **2024**, *15*, 922. <https://doi.org/10.3390/genes15070922>

Academic Editor: Claudia Ricci

Received: 18 June 2024

Revised: 10 July 2024

Accepted: 12 July 2024

Published: 15 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Acute febrile illness (AFI) and severe neurological disorders (SNDs) often present diagnostic challenges due to their potential origins from a wide range of infectious agents. Nanopore metagenomics is emerging as a powerful tool for identifying the microorganisms potentially responsible for these undiagnosed clinical cases. In this study, we aim to shed light on the etiological agents underlying AFI and SND cases that conventional diagnostic methods have not been able to fully elucidate. Our approach involved analyzing samples from fourteen hospitalized patients using a comprehensive nanopore metagenomic approach. This process included RNA extraction and enrichment using the SMART-9N protocol, followed by nanopore sequencing. Subsequent steps involved quality control, host DNA/cDNA removal, de novo genome assembly, and taxonomic classification. Our findings in AFI cases revealed a spectrum of disease-associated microbes, including *Escherichia coli*, *Streptococcus* sp., Human Immunodeficiency Virus 1 (Subtype B), and Human Pegivirus. Similarly, SND cases revealed the presence of pathogens such as *Escherichia coli*, *Clostridium* sp., and Dengue virus type 2 (Genotype-II lineage). This study employed a metagenomic analysis method, demonstrating its efficiency and adaptability in pathogen identification. Our investigation successfully identified pathogens likely associated with AFI and SNDs, underscoring the feasibility of retrieving near-complete genomes from RNA viruses. These findings offer promising prospects for advancing our understanding and control of infectious diseases, by facilitating detailed genomic analysis which is critical for developing targeted interventions and therapeutic strategies.

Keywords: acute febrile illness; severe neurological disorders; nanopore sequencing; metagenomics

1. Introduction

Acute febrile illness (AFI) is a clinical syndrome characterized by fever (≥ 38.0 °C) and often accompanied by various non-specific symptoms such as headache, rash, and muscle and joint pains. This illness frequently necessitates hospitalization and extensive investigations to determine its cause, which is primarily due to infectious pathogens that vary epidemiologically around the world. For instance, malaria is the predominant cause of AFI in Africa, whereas in Latin America and Asia, dengue and leptospirosis are more common causes of this syndrome [1,2].

Despite advancements in medical diagnostics, a significant proportion of AFI cases remain unresolved. Conventional diagnostic approaches often yield inconclusive or negative results. Studies indicate that nearly 50% of AFI cases remain undiagnosed, a figure that may be higher due to the frequent misdiagnosis of AFI as arboviral disease in primary care settings, where no further investigations are conducted [3,4].

Severe neurological disorders (SNDs) encompass a range of conditions that impact the brain, spinal cord, and nerves. Symptoms of SNDs may include headaches, taste and smell dysfunction, muscle weakness, paralysis, a loss of sensation, poor coordination, confusion, seizures, and pain. SNDs are the second leading cause of death globally, with various causes including infectious bacteria, fungi, and particularly viruses [5,6].

Due to the non-specific symptoms and the wide range of associated infections, SNDs and AFI present significant diagnostic challenges. In many cases, supplementary methodologies are necessary to determine the etiological origins of these complex diseases. Among these methodologies, metagenomics has emerged as a revolutionary approach. It provides a comprehensive analysis of the microbial composition within biological samples. This technique facilitates the identification of potential emerging and re-emerging pathogens through nucleotide sequence analysis and generates valuable genomic data. These insights enhance our understanding of disease biology, host–pathogen interactions, and epidemiology [7].

The application of metagenomics in the detection of infectious agents, especially those elusive to conventional diagnostic methods, has seen a marked increase in recent years [8,9]. Metagenomic analysis leverages a variety of next-generation sequencing (NGS) technologies. Among these, the MinION sequencing platform is increasingly favored due to its capacity to rapidly generate substantial data volumes. Its long-read sequencing capabilities address the challenges associated with ambiguous repetitive regions, enhancing genomic contiguity. The portability and cost-effectiveness of this technology further bolster its adoption in genomic research [10]. As the epidemiological landscape of infectious diseases continues to shift, the relevance of sophisticated metagenomic methodologies becomes more pronounced. These technologies are crucial in elucidating the intricate interactions between pathogens and their hosts, aiming to mitigate the public health burden of diseases such as AFI and SND. They are instrumental in safeguarding public health in an era characterized by the frequent emergence and re-emergence of infectious diseases.

In light of these capabilities, this study sought to employ the nanopore sequencing platform to identify and characterize microorganisms linked to febrile or neurological disorders of unknown etiology. The investigation focused on a cohort of hospitalized patients with undiagnosed cases of AFI or SND, demonstrating the utility of metagenomics in resolving complex diagnostic challenges in clinical settings.

2. Materials and Methods

2.1. Clinical Sampling

The Eduardo de Menezes Hospital in Belo Horizonte, Brazil, plays a crucial role in healthcare by conducting etiological investigations for patients hospitalized with conditions such as acute febrile illness or severe neurological disorders. The hospital ensures the highest biosafety standards during the collection and identification of samples, following medical requests. Collaborating with the Ezequiel Dias Foundation (FUNED), the hospital aims to uncover possible links between infections with inconclusive diagnoses

and microbial infections. This is carried out through extensive serological and molecular biology investigations targeting arboviruses, hepatitis, rickettsiosis, hantavirus, leptospirosis, and other pathogens. Samples are collected based on clinical information and medical requests and may include serum, whole blood, swabs, or cerebrospinal fluid. If an RNA virus is suspected, blood is also collected in a Tempus tube, which contains chemicals that preserve RNA for a longer period. Despite these efforts, in 2022, fourteen patients had inconclusive results even after serological and molecular biology investigations. As an alternative approach to identify the etiological agents, these patient samples were subjected to metagenomic analysis (Figure 1).

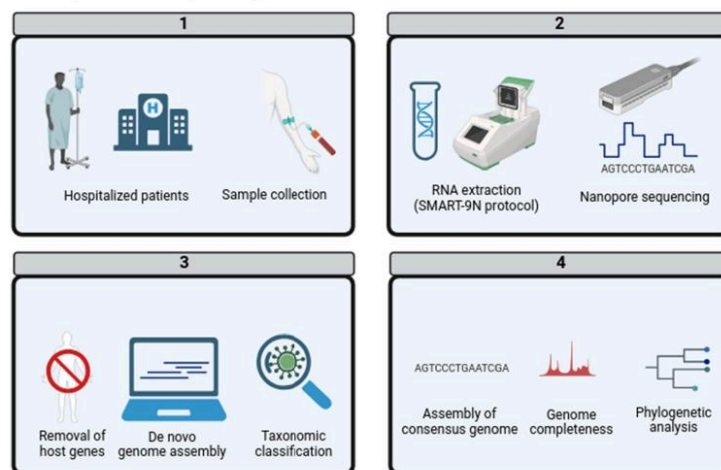


Figure 1. Metagenomic workflow analysis of AFI and SND in hospitalized patients. (1) Sample collection: Collection of biological samples from hospitalized patients diagnosed with acute febrile illness [AFI] or severe neurological disorder [SND] of unidentified origin. (2) Laboratory Processing: Samples underwent RNA extraction using the SMART-9N protocol, followed by metagenomic sequencing employing nanopore technology. (3) Data Analysis: Sequences were subjected to quality control, the removal of host genetic material, and the generation of contigs through de novo genome assembly. These contigs were then classified taxonomically. (4) Pathogen Identification and Further Analysis: Identified pathogens with coverage exceeding 80% were used to assemble a consensus genome. The completeness of these genomes was assessed, culminating in the construction of a phylogenetic tree.

2.2. RNA Extraction and Nanopore Sequencing

RNA was extracted from a range of 200 to 500 μ L of biological samples using the High Pure RNA Isolation Kit (Hoffmann–La Roche, Basel, Switzerland) and concentrated using the RNA Clean & Concentrator-5 (Zymo Research, Irvine, CA, USA). The samples then underwent the SMART-9N PCR protocol [11], utilizing a three-random-primed SMART-Seq model compatible with Oxford Nanopore Technologies (ONT) adapters. This protocol is particularly effective in enhancing the detection and sequencing of pathogen genomes, including RNA viruses, and facilitates the identification and sequencing of DNA viruses, bacteria, fungi, eukaryotes, and archaea. Library preparation involved multiple steps: PCR purification using AMPure XP beads (Beckman Coulter, Indianapolis, IN, USA), the normalization of DNA concentration, purification, end repair, and dA-tailing using the Ultra II End Prep Kit (New England Biolabs, Ipswich, MA, USA). This was followed by further purification, native barcode ligation using the PCR Barcoding Expansion (Oxford Nanopore Technologies, Oxford, UK), and adapter ligation with the Adapter Mix II Expansion protocol (Oxford Nanopore Technologies, UK).

The genomic material at each step was quantified using the Qubit dsDNA Quantification Assay Kit (Life Technologies, Carlsbad, CA, USA). Following barcoding, the samples were pooled and loaded into an Oxford Nanopore MinION R9.4.1 flow cell, primed according to the manufacturer's instructions using the Flow Cell Priming Kit (Oxford Nanopore Technologies, UK). Sequencing was conducted on the ONT MinION Mk1B device using MinKNOW software.

The initial sequencing run generated raw data in FAST5 format, capturing electrical signals. These files were processed for base calling, converting them to FASTQ format, and demultiplexed using Guppy software, developed by ONT [12].

2.3. De Novo Genome Assembly and Taxonomic Classification

The resulting FASTQ files were analyzed using the open-source metagenomic cloud-pipeline provided by the Chan-Zuckerberg IDseq platform [13]. This analysis involved a series of steps: initial quality assessment using Fastp software [14], followed by the removal of human host sequences using minimap2 [15]. Subsequently, the remaining reads were de novo-assembled using Flye software, specifically employing the metaFlye parameter tailored for metagenomic analysis [16]. The resulting contigs and unaligned reads were then classified using the DIAMOND software [17] and cross-referenced with the NCBI Nucleotide database (NT). The complete list of detected microorganisms is provided in Supplementary Table S1.

2.4. Phylogenetic Analysis

Pathogens with genome coverage exceeding 80% were subjected to genome assembly using the Genome Detective online tool [18]. The resulting consensus genomes were then assessed for quality and completeness using CheckV software [19]. These genomic sequences were aligned to reference genomes retrieved from the NCBI database using MAFFT with default parameters [20]. Manual curation was performed using Aliview [21] to eliminate biological artifacts. Phylogenetic trees were constructed using Maximum Likelihood (ML) estimation performed by IQ-TREE2, employing the general time reversible (GTR) model of nucleotide substitution and supported by 1000 bootstrap replicates, with the command `iqtree -s archive.fasta -m GTR+F+R5 -bb 1000`. The final visualization of the phylogenetic trees was facilitated using FigTree software. The datasets used to build the phylogenetic trees are in Supplementary Table S2.

3. Results

In this study, nine hospitalized patients were diagnosed with acute febrile illness (AFI), while three exhibited symptoms indicative of severe neurological disease (SND). Notably, two AFI patients lacked sufficient epidemiological data, resulting in their exclusion from further analysis. Table 1 summarizes the epidemiological data of the included patients, showing an age distribution ranging from 14 to 61 years. The gender distribution was equal, with 50% female and 50% male participants. All patients were residents of Minas Gerais, predominantly from Belo Horizonte, although some were from nearby cities such as Ouro Preto or Ribeirão das Neves.

Clinically, the hospitalized patients presented a variety of symptoms that necessitated their admission. Those with AFI commonly reported fever, myalgia, arthralgia, and headaches. In contrast, SND cases were primarily associated with symptoms like headaches, seizures, and potential neurosyphilis. A total of 14 samples were analyzed from these 12 patients, with 2 individuals (patients 7 and 11) contributing 2 samples each. The types of samples included five whole blood, two cerebrospinal fluid, five serum, and two RNA Tempus blood samples.

Table 1. Epidemiological information from the patients analyzed in this study.

| Patient | Gender | Age | City | State | Clinical Classification | Symptoms | Sample |
|---------|--------|-----|--------------------|--------------|-------------------------|--|----------------------|
| 1 | Male | 53 | Belo Horizonte | Minas Gerais | SND | Headaches | Serum |
| 2 | Female | 38 | Belo Horizonte | Minas Gerais | AFI | Fever, myalgia, arthralgia, fatigue, headache, and skin lesions (rash) | Tempus |
| 3 | Male | 38 | Belo Horizonte | Minas Gerais | SND | Headache, convulsive seizure, and neurological symptoms | Whole blood |
| 4 | Male | 53 | Ouro Preto | Minas Gerais | SND | Suggestive of neurosyphilis and bacterial meningitis | Whole blood |
| 5 | Female | 43 | Belo Horizonte | Minas Gerais | AFI | Myalgia, arthralgia, fatigue, and petechiae | Serum |
| 6 | Male | 34 | Belo Horizonte | Minas Gerais | AFI | Fever, myalgia, headache, and petechiae | Whole blood |
| 7 | Female | 27 | Ribeirão das Neves | Minas Gerais | AFI | Fever, cough, myalgia, fatigue, and petechiae | Tempus, serum |
| 8 | Female | 14 | Belo Horizonte | Minas Gerais | AFI | Fever, myalgia, and headache | Whole blood |
| 9 | Male | 37 | Belo Horizonte | Minas Gerais | AFI | Fever | Serum |
| 10 | Male | 61 | Betim | Minas Gerais | AFI | Fever | Serum |
| 11 | Female | 23 | Belo Horizonte | Minas Gerais | AFI | Fever and subacute myeloradiculopathy | Cerebrospinal fluids |
| 12 | Female | 32 | Belo Horizonte | Minas Gerais | AFI | Fever, myalgia, skin lesions, immunosuppression, and acute hepatitis | Whole blood |

Among the patients, 25% were clinically classified as SND cases ($n = 3$), while 75% were classified as AFI cases ($n = 9$) (Figure 2A). Among the SND cases, nine microorganisms were identified, with three pathogens (34%) associated with SND: *Escherichia coli* (33%), *Clostridium* sp. (33%), and Dengue virus type 2 (33%) (Figure 2B,C).

Regarding the AFI cases, a total of 21 microorganisms were identified, with 8 (38%) of them being pathogens associated with AFI symptoms (Figure 2D). Bacterial detection was predominant, with *Escherichia coli* (76%) being the most prevalent, followed by *Streptococcus* sp. (8%). Furthermore, viruses were identified in two cases, including Human Immunodeficiency Virus 1 (HIV-1) (8%), and Human pegivirus (HPgV) (8%) (Figure 2E).

Among the bacteria and fungi detected, *Escherichia coli* was identified in all types of collected samples and was associated with a variety of symptoms including fever, myalgia, arthralgia, fatigue, headaches, skin lesions, petechiae, and non-specific signs of neurosyphilis and meningitis. While *E. coli* is often considered part of the normal microbiome, it can also act as a pathogen, particularly when it is the sole pathogen detected or depending on the type of sample in which it is found.

Clostridium species were identified in a serum sample from a hospitalized patient with SND presenting severe headache symptoms. Given the range of diseases caused by different species of *Clostridium*, their detection cannot be overlooked. A similar situation occurred with *Streptococcus* species, which were found in a whole blood sample from an AFI patient exhibiting symptoms of fever, myalgia, headaches, and petechiae. In the sequencing control, no microorganisms were detected in the contigs, which highlights the absence of cross-contamination.

The disease-associated bacterial detections had a diverse range of read counting, with some bacterial detections presenting a high number of reads. However, all exhibited very low coverage and depth, which hindered further analysis (Table 2). In contrast, the viruses detected were sequenced with a high number of reads and great quality, allowing the assembly of near-complete genomes of HIV, HPgV, and Dengue virus type 2 (DENV-2)

with the current protocol. This enabled the construction of a phylogenetic tree to identify viral lineages associated with these infections (Figure 3).

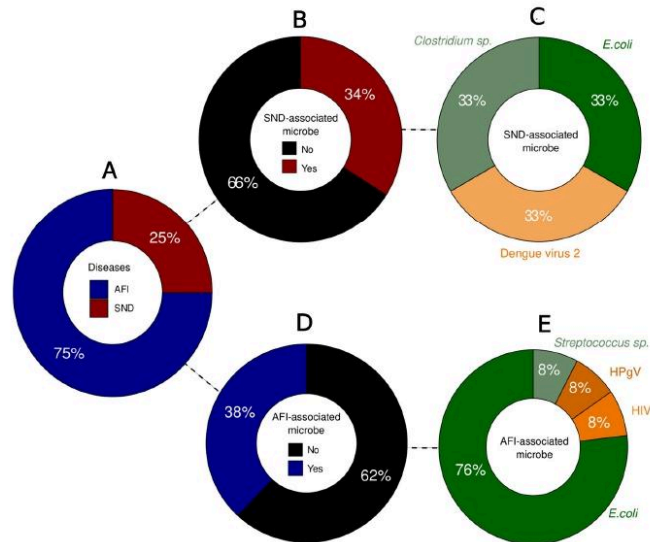


Figure 2. Overview of the proportion of cases per classification and microorganisms identified. (A) Percentage of cases per clinical classification. (B) Proportion of SND-associated microorganisms in the samples. (C) SND-associated microorganisms identified: *Clostridium* sp. (light green), *E. coli* (dark green) and DENV-2 (light orange). (D) Proportion of AFI-associated microorganisms in the samples. (E) AFI-associated microorganisms identified: *E. coli* (dark green), *Streptococcus* sp. (light green), HIV (orange) and HPgV (dark orange).

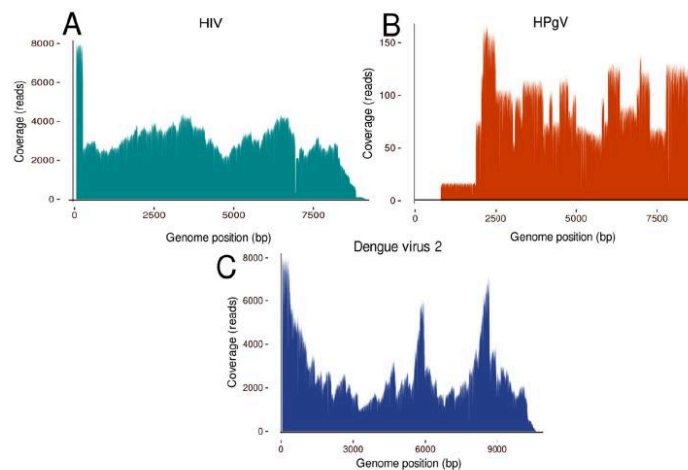
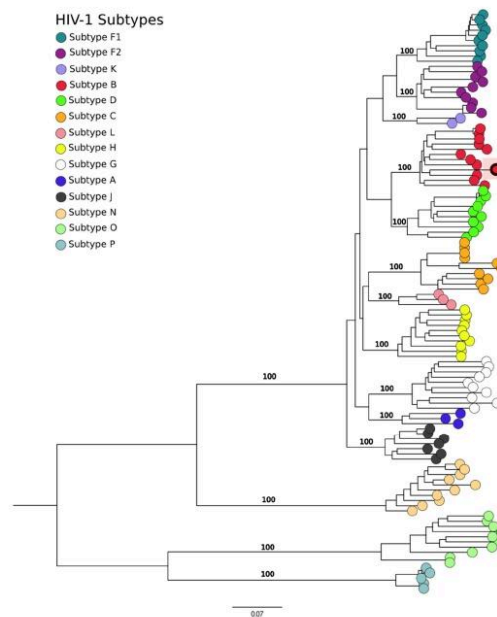


Figure 3. Genome coverage of the viral pathogens identified. (A) The genome coverage of the HIV genome, (B) the genome coverage of the HPgV genome, and (C) the genome coverage of the Dengue virus 2 genome.

Table 2. Sequencing data from the detected disease-associated microorganisms.

| Patient | Sample Type | Pathogen | Reads | Contigs | Coverage | Depth |
|---------|-----------------------|--------------------------|--------|---------|----------|--------|
| 1 | Serum | <i>Clostridium</i> sp. | 6554 | 1 | 0.8% | 0.25× |
| 2 | Tempus | <i>E. coli</i> | 63 | 1 | 0.1% | 0.005× |
| 3 | Whole blood | DENV-2 | 72,078 | 3 | 99.8% | 1145× |
| 4 | Whole blood | <i>E. coli</i> | 6939 | 1 | 0.1% | 0.61× |
| 5 | Serum | <i>E. coli</i> | 10,968 | 1 | 0.3% | 1× |
| 6 | Whole blood | HPgV | 511 | 2 | 86.2% | 41× |
| | | <i>E. coli</i> | 4635 | 1 | 0.4% | 0.18× |
| 7 | Tempus | <i>Streptococcus</i> sp. | 1347 | 1 | 0.1% | 0.17× |
| | | <i>E. coli</i> | 173 | 1 | 0.7% | 0.012× |
| 8 | Serum | <i>E. coli</i> | 9889 | 1 | 0.1% | 0.95× |
| | | <i>E. coli</i> | 5280 | 1 | 0.1% | 0.41× |
| 9 | Whole blood | HIV-1 | 52,420 | 2 | 92% | 1805× |
| 10 | Serum | <i>E. coli</i> | 13,778 | 1 | 0.2% | 1.2× |
| | | <i>E. coli</i> | 22,622 | 1 | 0.1% | 1.6× |
| 11 | Cerebrospinal fluid 1 | <i>E. coli</i> | 2084 | 1 | 0.1% | 0.19× |
| 11 | Cerebrospinal fluid 2 | <i>E. coli</i> | 1703 | 1 | 0.2% | 0.11× |
| 12 | Whole blood | - | - | - | - | - |
| 13 | Negative control | <i>E. coli</i> | 27 | 0 | - | - |

The significance of these findings is highlighted by the identification of specific viral subtypes, as demonstrated by the confirmation of HIV in a serum sample from a patient diagnosed with acute febrile illness (AFI). This virus was phylogenetically classified as HIV type 1, subtype B (Figure 4), thereby elucidating the viral subtype present in the patient. This classification enhances our understanding of the pathogenic landscape within the AFI patient cohort.

**Figure 4.** Phylogenetic tree of the HIV genome analysis. The sample from the study is the highlighted sample grouped with subtype B.

The HPgV-positive sample was obtained from a whole blood specimen, which also tested positive for *Escherichia coli* and *Streptococcus* species. This sample was associated with symptoms typical of AFI, including fever, myalgia, headaches, and petechiae on the skin. The genome analysis of the sample yielded an 86% complete genome of HPgV. Subsequent phylogenetic analysis classified this virus as belonging to the human strain HPgV-1 (Figures 3C and 5), providing insights into the viral strains affecting this patient population.

From the DENV-2 positive case, a genome was recovered with 99% coverage (Figure 4, which was classified as belonging to genotype II, also known as the cosmopolitan lineage (Figure 6). This virus was detected in a whole blood sample from a patient diagnosed with a severe neurological disorder (SND), presenting with symptoms including headaches, convulsive seizures, and other nonspecific neurological symptoms.

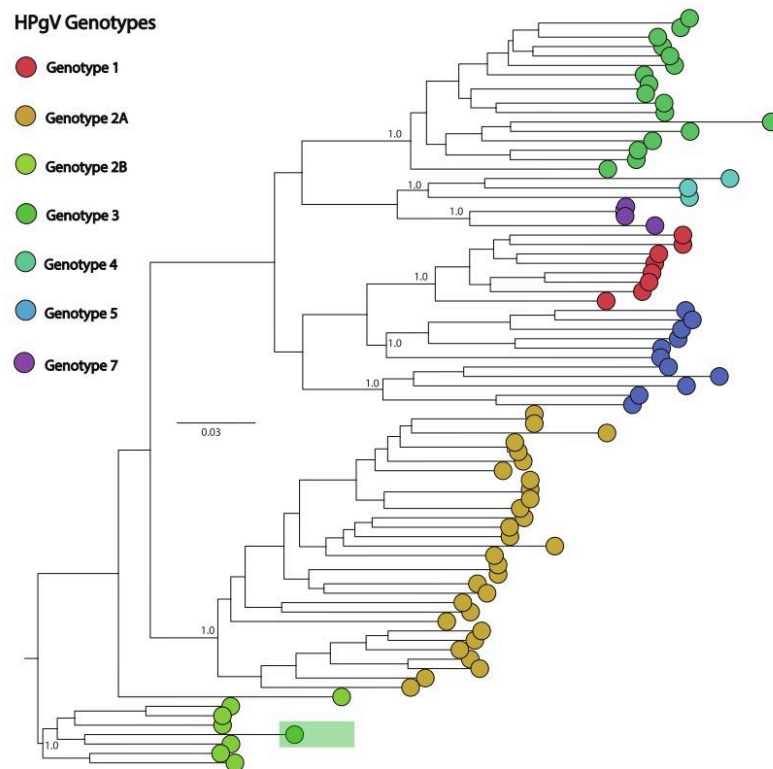


Figure 5. Phylogenetic tree of the HPgV genome analysis. The study's sample is the one that is highlighted in the human clade.

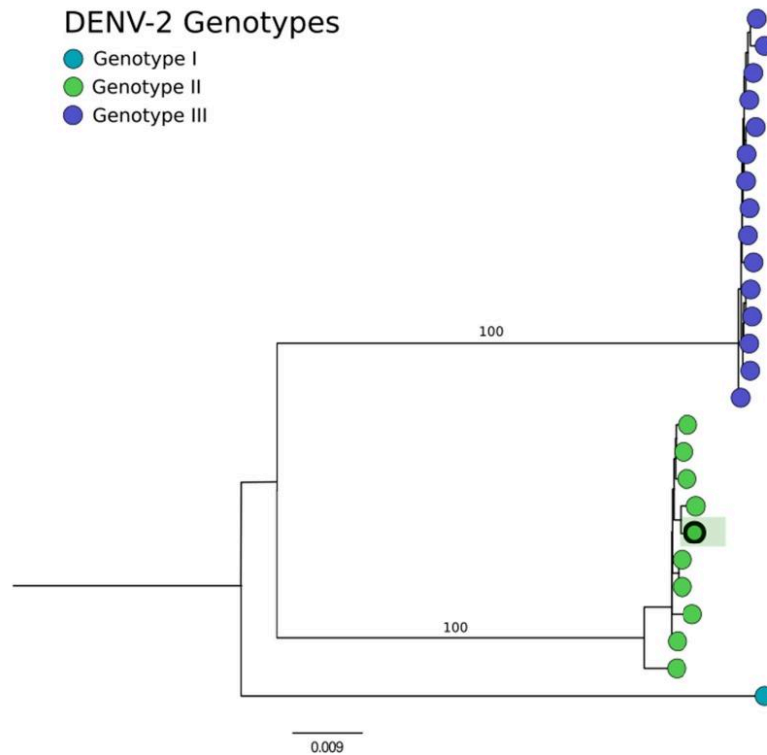


Figure 6. Phylogenetic tree of the DENV2 genome analysis. Comparing the detected genotypes of DENV2 in Brazil, the sample from the study is the highlighted sample that is grouped in genotype II.

4. Discussion

The utilization of metagenomic analysis to elucidate the etiology of AFI and SNDs with unknown origins presents a significant advancement in clinical medicine. The ability to identify a specific etiology can greatly enhance the efficacy of treatments for hospitalized patients. In our research, the application of metagenomic techniques facilitated the identification of diverse microorganisms. This detection capability potentially elucidates the etiological factors underlying the symptoms and resultant hospitalizations of the patients included in the study. Notably, *Escherichia coli* was the most commonly identified pathogen. While *E. coli* is a typical constituent of the human microbiome, under certain conditions, pathogenic strains can lead to various diseases, including acute febrile and neurological disorders [22–26]. The detection of *E. coli* in critical samples such as cerebrospinal fluid, whole blood, and serum necessitates heightened attention due to its potential to trigger systemic inflammatory responses, potentially leading to septicemia [27]. Additionally, our metagenomic analysis recovered nearly complete genomes of several viruses, facilitating further analyses to determine their lineages and genotypes. Such information is crucial for developing targeted treatment strategies and understanding the epidemiological characteristics of these viruses within the population. For example, HIV was identified and classified phylogenetically as HIV type 1, subtype B, which is prevalent both globally and in the Minas Gerais state [28,29]. This subtype was also detected in a patient who had previously reported being HIV-positive but had elected not to pursue treatment in 2020. This case

underscores the utility of metagenomic analysis in identifying secondary infections that could explain the patient's hospitalization for AFI.

The study also highlighted the controversial pathogenicity of HPgV, detected in a sample co-infected with *Streptococcus*. While some studies do not associate HPgV with any specific human diseases, others suggest potential links to conditions such as lymphoma or brain encephalitis [30]. This ambiguity points to the need for further research to clarify the clinical significance of HPgV, particularly when found in conjunction with typical arbovirus-like symptoms such as fever, myalgia, headache, and petechiae. Our findings also identified pathogens associated with SNDs, including *Clostridium* sp. and DENV-2. The presence of the *Clostridium* genus, known for its association with severe neurological symptoms, raises public health concerns, although the specific species could not be determined [31,32]. Moreover, DENV-2 is a well-known neurological pathogen [33,34], which genetically belonged to the genotype II cosmopolitan lineage, recently reported to be spreading in Brazil and also specifically in Minas Gerais State [35–37]. This lineage was first detected in the region in April 2022 and has since been prevalent, highlighting the benefits of metagenomic analysis in tracking the spread and evolution of pathogens.

5. Conclusions

In conclusion, our metagenomic analysis has yielded significant insights into the pathogens associated with AFI and SND. This comprehensive approach has delineated a spectrum of bacteria, fungi, and viruses, illuminating their potential roles in the pathogenesis of these conditions. The precise identification of these pathogens, along with their genetic profiles and involvement in co-infection scenarios, substantially augments our capacity to formulate timely and specifically targeted therapeutic strategies. Furthermore, this study underscores the critical need for continued research into the pathogenicity of relatively unexplored microbes such as HPgV. The outcomes of this research not only contribute to a deeper understanding of infectious disease dynamics but also enhance the implementation of metagenomic methodologies in clinical diagnostics and public health, thereby laying the groundwork for future exploratory and applied research in this field.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes15070922/s1>, Table S1: Complete List of Detected Microorganisms; Table S2: Datasets Used for Phylogenetic Tree Construction.

Author Contributions: Conceptualization, L.C.J.A. and M.G.; methodology, K.M.F.M., V.F., L.M.R.T., N.R.G., M.T.L., E.d.C.B., H.M.F., T.A., F.C.d.M.I., T.O.F., M.C.A., T.G.B., V.A.d.A., L.C.J.A. and M.G.; formal analysis, K.M.F.M.; investigation, K.M.F.M., M.T.L., T.A., E.d.C.B. and H.M.F.; data curation, K.M.F.M., V.F. and H.M.F.; writing—original draft preparation, K.M.F.M., L.M.R.T., and M.G.; writing—review and editing, K.M.F.M., V.F., L.M.R.T., N.R.G., M.T.L. and M.G.; visualization, K.M.F.M. and M.G.; supervision, M.G., V.F. and L.C.J.A.; project administration, L.C.J.A., M.G., V.F., F.C.d.M.I. and V.A.d.A.; funding acquisition, L.C.J.A. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This study was in part financed by the Coordination for the Improvement of Higher Education Personnel, Brazil (CAPES). This study was supported by the National Institutes of Health USA grant U01 AI151698 for the United World Arbovirus Research Network [UWARN] and the CRP-ICGEB RESEARCH GRANT 2020 Project CRP/BRA20-03, Contract CRP/20/03. M. Giovanetti's funding is provided by PON "Ricerca e Innovazione" 2014–2020.

Institutional Review Board Statement: This study received approval from the Ethics and Research Committee of the Federal University of Minas Gerais, under the registration CAAE: 90249218.6.1001.5248.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available in FigShare at <https://figshare.com/account/home#/projects/201114> accessed on 17 June 2024.

Acknowledgments: The authors would like to acknowledge the Global Consortium to Identify and Control Epidemics—CLIMADE. We are grateful for the support from all funding institutions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Mbidde, E.K.; Lutwama, J.J.; Perniciaro, J.L.; Nicholson, W.L.; Bower, W.A.; Bwogi, J.; Blaney, D.D. Investigating the etiology of acute febrile illness: A prospective clinic-based study in Uganda. *BMC Infect. Dis.* **2023**, *23*, 411. [CrossRef] [PubMed] [PubMed Central]
- Tam, P.Y.I.; Obaro, S.K.; Storch, G. Challenges in the Etiology and Diagnosis of Acute Febrile Illness in Children in Low- and Middle-Income Countries. *J. Pediatr. Infect. Dis. Soc.* **2016**, *5*, 190–205. [CrossRef] [PubMed] [PubMed Central]
- Moreira, J.; Bressan, C.S.; Brasil, P.; Siqueira, A.M. Epidemiology of acute febrile illness in Latin America. *Clin. Microbiol. Infect.* **2018**, *24*, 827–835. [CrossRef] [PubMed] [PubMed Central]
- Bressan, C.D.S.; Teixeira, M.L.B.; Gouvêa, M.I.F.D.S.; de Pina-Costa, A.; Santos, H.F.P.; Calvet, G.A.; Lupi, O.; Siqueira, A.M.; Valls-de-Souza, R.; Valim, C.; et al. Challenges of acute febrile illness diagnosis in a national infectious diseases center in Rio de Janeiro: 16-year experience of syndromic surveillance. *PLoS Negl. Trop. Dis.* **2023**, *17*, e0011232. [CrossRef] [PubMed] [PubMed Central]
- Chen, X.; Laurent, S.; Onur, O.A.; Kleineberg, N.N.; Fink, G.R.; Schweitzer, F.; Warnke, C. A systematic review of neurological symptoms and complications of COVID-19. *J. Neurol.* **2021**, *268*, 392–402. [CrossRef] [PubMed] [PubMed Central]
- Wouk, J.; Rechenchoski, D.Z.; Rodrigues, B.C.D.; Ribelato, E.V.; Faccin-Galhardi, L.C. Viral infections and their relationship to neurological disorders. *Arch. Virol.* **2021**, *166*, 733–753. [CrossRef] [PubMed]
- Chiu, C.Y.; Miller, S.A. Clinical metagenomics. *Nat. Rev. Genet.* **2019**, *20*, 341–355. [CrossRef] [PubMed] [PubMed Central]
- Ciuffreda, L.; Rodríguez-Pérez, H.; Flores, C. Nanopore sequencing and its application to the study of microbial communities. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1497–1511. [CrossRef] [PubMed] [PubMed Central]
- Deng, Q.; Cao, Y.; Wan, X.; Wang, B.; Sun, A.; Wang, H.; Wang, Y.; Wang, H.; Gu, H. Nanopore-based metagenomic sequencing for the rapid and precise detection of pathogens among immunocompromised cancer patients with suspected infections. *Front. Cell Infect. Microbiol.* **2022**, *12*, 943859. [CrossRef] [PubMed] [PubMed Central]
- Latorre-Pérez, A.; Villalba-Bermell, P.; Pascual, J.; Vilanova, C. Assembly methods for nanopore-based metagenomic sequencing: A comparative study. *Sci. Rep.* **2020**, *10*, 13588. [CrossRef] [PubMed] [PubMed Central]
- Claro, I.M.; Ramundo, M.S.; Coletti, T.M.; da Silva, C.A.M.; Valença, I.N.; Candido, D.S.; Sales, F.C.S.; Manuli, E.R.; de Jesus, J.G.; de Paula, A.; et al. Rapid viral metagenomics using SMART-9N amplification and nanopore sequencing. *Wellcome Open Res.* **2023**, *6*, 241. [CrossRef] [PubMed] [PubMed Central]
- Wick, R.R.; Judd, L.M.; Holt, K.E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **2019**, *20*, 129. [CrossRef] [PubMed] [PubMed Central]
- Kalantar, K.L.; Carvalho, T.; de Bourcy, C.F.A.; Dimitrov, B.; Dingle, G.; Egger, R.; Han, J.; Holmes, O.B.; Juan, Y.F.; King, R.; et al. IDseq-An open-source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *GigaScience* **2020**, *9*, g1aa111. [CrossRef] [PubMed] [PubMed Central]
- Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef] [PubMed] [PubMed Central]
- Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef] [PubMed] [PubMed Central]
- Kolmogorov, M.; Bickhart, D.M.; Behsaz, B.; Gurevich, A.; Rayko, M.; Shin, S.B.; Kuhn, K.; Yuan, J.; Polevikov, E.; Smith, T.P.L.; et al. metaFlye: Scalable long-read metagenome assembly using repeat graphs. *Nat. Methods* **2020**, *17*, 1103–1110. [CrossRef] [PubMed] [PubMed Central]
- Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [CrossRef] [PubMed]
- Vilsker, M.; Moosa, Y.; Nooij, S.; Fonseca, V.; Ghysens, Y.; Dumon, K.; Pauwels, R.; Alcantara, L.C.; Vanden Eynden, E.; Vandamme, A.M.; et al. Genome Detective: An automated system for virus identification from high-throughput sequencing data. *Bioinformatics* **2019**, *35*, 871–873. [CrossRef] [PubMed] [PubMed Central]
- Nayfach, S.; Camargo, A.P.; Schulz, F.; Eloie-Fadrosch, E.; Roux, S.; Kyrpides, N.C. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **2021**, *39*, 578–585. [CrossRef] [PubMed] [PubMed Central]
- Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef] [PubMed] [PubMed Central]
- Larsson, A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **2014**, *30*, 3276–3278. [CrossRef] [PubMed] [PubMed Central]
- Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534, Erratum in *Mol. Biol. Evol.* **2020**, *37*, 2461. [CrossRef] [PubMed] [PubMed Central]
- Rambaut, A. FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. 2010. Available online: <http://tree.bio.ed.ac.uk/software/figtree/> (accessed on 17 June 2024).
- Kaper, J.B.; Nataro, J.P.; Mobley, H.L. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2004**, *2*, 123–140. [CrossRef] [PubMed]
- Schuppner, R.; Maehlmann, J.; Dirks, M.; Worthmann, H.; Tryc, A.B.; Sandorski, K.; Bahlmann, E.; Kielstein, J.T.; Giesemann, A.M.; Lanfermann, H.; et al. Neurological Sequelae in Adults After *E. coli* O104: H4 Infection-Induced Hemolytic-Uremic Syndrome. *Medicine* **2016**, *95*, e2337. [CrossRef] [PubMed] [PubMed Central]

26. Soedarmono, P.; Diana, A.; Tauran, P.; Lokida, D.; Aman, A.T.; Alisjahbana, B.; Arlinda, D.; Tjitra, E.; Kosasih, H.; Merati, K.T.P.; et al. The characteristics of bacteremia among patients with acute febrile illness requiring hospitalization in Indonesia. *PLoS ONE* **2022**, *17*, e0273414. [CrossRef] [PubMed] [PubMed Central]
27. Daga, A.P.; Koga, V.L.; Soncini, J.G.M.; de Matos, C.M.; Perugini, M.R.E.; Pelisson, M.; Kobayashi, R.K.T.; Vespero, E.C. *Escherichia coli* Bloodstream Infections in Patients at a University Hospital: Virulence Factors and Clinical Characteristics. *Front. Cell Infect. Microbiol.* **2019**, *9*, 191. [CrossRef] [PubMed] [PubMed Central]
28. Junqueira, D.M.; Almeida, S.E. HIV-1 subtype B: Traces of a pandemic. *Virology* **2016**, *495*, 173–184. [CrossRef] [PubMed]
29. Da Silva, G.J.; Aleixo, A.W.; Tupinambás, U. Prevalence of HIV-1 transmitted drug resistance and its impact on the effectiveness of antiretroviral therapy—Minas Gerais state, Brazil. *Braz. J. Health Rev.* **2022**, *5*, 1044–1059. [CrossRef]
30. Yu, Y.; Wan, Z.; Wang, J.H.; Yang, X.; Zhang, C. Review of human pegivirus: Prevalence, transmission, pathogenesis, and clinical implication. *Virulence* **2022**, *13*, 324–341. [CrossRef] [PubMed] [PubMed Central]
31. Finsterer, J.; Hess, B. Neuromuscular and central nervous system manifestations of *Clostridium perfringens* infections. *Infection* **2007**, *35*, 396–405. [CrossRef] [PubMed]
32. Cai, S.; Kumar, R.; Singh, B.R. Clostridial Neurotoxins: Structure, Function and Implications to Other Bacterial Toxins. *Microorganisms* **2021**, *9*, 2206. [CrossRef] [PubMed] [PubMed Central]
33. Trivedi, S.; Chakravarty, A. Neurological Complications of Dengue Fever. *Curr. Neurol. Neurosci. Rep.* **2022**, *22*, 515–529. [CrossRef] [PubMed] [PubMed Central]
34. Kulkarni, R.; Pujari, S.; Gupta, D. Neurological Manifestations of Dengue Fever. *Ann. Indian. Acad. Neurol.* **2021**, *24*, 693–702. [CrossRef] [PubMed] [PubMed Central]
35. Giovanetti, M.; Pereira, L.A.; Santiago, G.A.; Fonseca, V.; Mendoza, M.P.G.; de Oliveira, C.; de Moraes, L.; Xavier, J.; Tosta, S.; Fristch, H.; et al. Emergence of Dengue Virus Serotype 2 Cosmopolitan Genotype, Brazil. *Emerg. Infect. Dis.* **2022**, *2*, 1725–1727. [CrossRef] [PubMed] [PubMed Central]
36. Amorim, M.T.; Hernández, L.H.A.; Naveca, F.G.; Essashika Prazeres, I.T.; Wanzeller, A.L.M.; Silva, E.V.P.D.; Casseb, L.M.N.; Silva, F.S.D.; da Silva, S.P.; Nunes, B.T.D.; et al. Emergence of a New Strain of DENV-2 in South America: Introduction of the Cosmopolitan Genotype through the Brazilian-Peruvian Border. *Trop. Med. Infect. Dis.* **2023**, *8*, 325. [CrossRef] [PubMed] [PubMed Central]
37. Arboviruses Epidemic. Available online: <https://www.saude.mg.gov.br/component/gmg/story/18774-epidemia-arbovirozes> (accessed on 12 April 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Outros artigos publicados

Além dos resultados da atual dissertação, outros artigos foram publicados com contribuições do discente.

O artigo intitulado “Phylogenetic Reconstructions Reveal the Circulation of a Novel Dengue Virus-1V Clade and the Persistence of a Dengue Virus-2 III Genotype in Northeast Brazil” buscou através do sequenciamento genético fazer análises filodinâmicas e epidemiológicas para revelar um novo clado de genótipo V de DENV-1 e a persistência de circulação do genótipo III do DENV-2 no Nordeste Brasileiro.



Article

Phylogenetic Reconstructions Reveal the Circulation of a Novel Dengue Virus-1V Clade and the Persistence of a Dengue Virus-2 III Genotype in Northeast Brazil

Hegger Fritsch ^{1,†}, Keldenn Moreno ^{1,†}, Italo Andrade Barbosa Lima ^{2,†}, Cleiton Silva Santos ², Bernardo Gratalval Gouveia Costa ², Breno Lima de Almeida ², Ronald Alves dos Santos ², Marcos Vinicius Lima de Oliveira Francisco ², Maria Paula Souza Sampaio ², Maricelia Maia de Lima ^{3,4}, Felicidade Mota Pereira ⁵, Vagner Fonseca ⁶, Stephane Tosta ¹, Joilson Xavier ¹, Carla de Oliveira ⁷, Talita Adelino ⁸, Arabela Leal e Silva de Mello ⁴, Tiago Gräf ⁹, Luiz Carlos Junior Alcantara ^{10,*}, Marta Giovanetti ^{10,11} and Isadora Cristina de Siqueira ^{2,*}

- ¹ Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Avenida Presidente Antônio Carlos, 6627, Belo Horizonte 31270-901, MG, Brazil
 - ² Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Rua Waldemar Falcão, 121, Salvador 40296-710, BA, Brazil
 - ³ Secretaria Municipal de Saúde de Feira de Santana, Avenida João Durval Carneiro, s/n, Feira de Santana 44027-010, BA, Brazil
 - ⁴ Departamento de Saúde, Universidade Estadual de Feira de Santana, Avenida Transnordestina, s/n, Feira de Santana 44036-900, BA, Brazil
 - ⁵ Laboratório Central de Saúde Pública Prof Gonçalo Moniz, Rua Waldemar Falcão, 123, Salvador 40295-010, BA, Brazil
 - ⁶ Organização Pan-Americana de Saúde/Organização Mundial de Saúde, Setor das Embaixadas, Lote 19, Avenida das Nações, Brasília 70-800400, SP, Brazil
 - ⁷ Laboratório de Flavivirus, Instituto Oswaldo Cruz/Fundação Oswaldo Cruz, Avenida Brasil, 4365, Rio de Janeiro 21040-900, RJ, Brazil
 - ⁸ Laboratório Central de Saúde Pública do Estado de Minas Gerais, Fundação Ezequiel Dias, Rua Conde Pereira Carneiro, 80, Belo Horizonte 30510-010, MG, Brazil
 - ⁹ Laboratório de Virologia Molecular, Instituto Carlos Chagas/Fiocruz-PR, Avenida Professor Algacyr Munhoz Mader, 3775, Curitiba 81310-020, PA, Brazil
 - ¹⁰ Instituto Rene Rachou, Fundação Oswaldo Cruz, Avenida Augusto de Lima, 1715, Belo Horizonte 30190-002, MG, Brazil
 - ¹¹ Sciences and Technologies for Sustainable Development and One Health, University of Campus Bio-Medico, Via Álvaro del Portillo, 21, 00128 Rome, Italy
- * Correspondence: luiz.alcantara@fiocruz.br (L.C.J.A.); isadora.siqueira@fiocruz.br (I.C.d.S.)
 † These authors contributed equally to this work.



Citation: Fritsch, H.; Moreno, K.; Lima, I.A.B.; Santos, C.S.; Costa, B.G.G.; de Almeida, B.L.; dos Santos, R.A.; Francisco, M.V.L.d.O.; Sampaio, M.P.S.; de Lima, M.M.; et al. Phylogenetic Reconstructions Reveal the Circulation of a Novel Dengue Virus-1V Clade and the Persistence of a Dengue Virus-2 III Genotype in Northeast Brazil. *Viruses* **2023**, *15*, 1073. <https://doi.org/10.3390/v15051073>

Academic Editor: Erich Mackow

Received: 6 April 2023

Revised: 26 April 2023

Accepted: 26 April 2023

Published: 28 April 2023

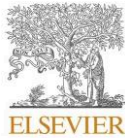


Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Dengue fever is among the most significant public health concerns in Brazil. To date, the highest number of Dengue notifications in the Americas has been reported in Brazil, with cases accounting for a total number of 3,418,796 reported cases as of mid-December 2022. Furthermore, the northeastern region of Brazil registered the second-highest incidence of Dengue fever in 2022. Due to the alarming epidemiological scenario, in this study, we used a combination of portable whole-genome sequencing, phylodynamic, and epidemiological analyses to reveal a novel DENV-1 genotype V clade and the persistence of DENV-2 genotype III in the region. We further report the presence of non-synonymous mutations associated with non-structural domains, especially the NS2A (non-structural protein 2A), as well as describe synonymous mutations in envelope and membrane proteins, distributed differently between clades. However, the absence of clinical data at the time of collection and notification, as well as the impossibility of monitoring patients in order to observe worsening or death, restricts our possibility of correlating mutational findings with possible clinical prognoses. Together, these results reinforce the crucial role of genomic surveillance to follow the evolution of circulating DENV strains and understand their spread across the region through inter-regional importation events, likely mediated by human mobility, and also the possible impacts on public health and outbreak management.

O artigo intitulado “Global SARS-CoV-2 genomic surveillance: What we have learned (so far)” buscou analisar o panorama global da circulação e evolução genômica do SARS-CoV-2.

Infection, Genetics and Evolution 108 (2023) 105405



Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid



Global SARS-CoV-2 genomic surveillance: What we have learned (so far)

Stephane Tosta^a, Keldenn Moreno^a, Gabriel Schuab^{b,c}, Vagner Fonseca^{d,**}, Fátima María Cardozo Segovia^e, Simone Kashima^f, Maria Carolina Elias^g, Sandra Coccuzzo Sampaio^g, Massimo Ciccozzi^h, Luiz Carlos Junior Alcantara^{a,c}, Svetoslav Nanev Slavov^{f,g}, José Lourençoⁱ, Eleonora Cella^{j,**}, Marta Giovanetti^{a,c,k,*}

^a Interunit Postgraduate Program in Bioinformatics, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

^b Federal University of Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil

^c Laboratório de Flavivirus, Instituto Oswaldo Cruz, Rio de Janeiro, Rio de Janeiro, Brazil

^d Organização Pan-Americana da Saúde/Organização Mundial da Saúde, Brasília, Distrito Federal, Brazil

^e Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción, San Lorenzo, Paraguay

^f Blood Center of Ribeirão Preto, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil

^g Butantan Institute, São Paulo, Brazil

^h Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Italy

ⁱ BioISI (Biosystems and Integrative Sciences Institute), Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal

^j Burnett School of Biomedical Sciences, University of Central Florida, Orlando, FL 32827, USA

^k Department of Science and Technology for Humans and the Environment, University of Campus Bio-Medico di Roma, Rome, Italy

ARTICLE INFO

Keywords:
SARS-CoV-2
Viral evolution
Genomic surveillance
Epidemic-pandemic preparedness

ABSTRACT

The COVID-19 pandemic has brought significant challenges for genomic surveillance strategies in public health systems worldwide. During the past thirty-four months, many countries faced several epidemic waves of SARS-CoV-2 infections, driven mainly by the emergence and spread of novel variants. In that line, genomic surveillance has been a crucial toolkit to study the real-time SARS-CoV-2 evolution, for the assessment and optimization of novel diagnostic assays, and to improve the efficacy of existing vaccines. During the pandemic, the identification of emerging lineages carrying lineage-specific mutations (particularly those in the Receptor Binding domain) showed how these mutations might significantly impact viral transmissibility, protection from reinfection and vaccination. So far, an unprecedented number of SARS-CoV-2 viral genomes has been released in public databases (i.e., GISAID, and NCBI), achieving 14 million genome sequences available as of early-November 2022. In the present review, we summarise the global landscape of SARS-CoV-2 during the first thirty-four months of viral circulation and evolution. It demonstrates the urgency and importance of sustained investment in genomic surveillance strategies to timely identify the emergence of any potential viral pathogen or associated variants, which in turn is key to epidemic and pandemic preparedness.

1. SARS-CoV-2: an emerging threat of international concern

In late December 2019, The World Health Organization (WHO) office in China was informed about a cluster of novel cases of pneumonia of unknown aetiology detected in the city of Wuhan, Hubei province (Huang et al., 2020; Wu et al., 2020) (Fig. 1). Shortly afterwards, a new type of coronavirus, later named SARS-CoV-2, was isolated and identified by the Chinese authorities and its genetic sequence was shared with the international community on 10 January 2020 (Zhou et al., 2020)

(Fig. 1). The coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 was classified as a pandemic on 11 of March 2020 (Cucinotta and Vanelli, 2020) by The World Health Organization (WHO). Up to early November 2022, more than 627 million confirmed cases of SARS-CoV-2 with exceeding 6.5 million associated deaths have been registered worldwide (WHO, 2022) (Fig. 1). As the pandemic advanced, a variety of methods were applied to understanding viral biology and identifying viral factors that influenced the evolution and global spread of this emerging threat.

* Corresponding author at: Interunit Postgraduate Program in Bioinformatics, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.

** Corresponding authors.

E-mail addresses: vagner.fonseca@gmail.com (V. Fonseca), eleonora.cella@yahoo.it (E. Cella), giovanetti.marta@gmail.com (M. Giovanetti).

<https://doi.org/10.1016/j.meegid.2023.105405>

Received 13 December 2022; Received in revised form 13 January 2023; Accepted 17 January 2023

Available online 18 January 2023

1567-1348/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

O artigo intitulado “Exploring the urban arbovirus landscape in Rio de Janeiro, Brazil: transmission dynamics and patterns of disease spread” buscou analisar a dinâmica de transmissão de arbovírus no Rio de Janeiro, identificando padrões de propagação da doença.

Exploring the urban arbovirus landscape in Rio de Janeiro, Brazil: transmission dynamics and patterns of disease spread



Gabriel Schuab,^{a,b,p} Stephane Tosta,^{c,p} Keldenn Moreno,^{c,p} Vagner Fonseca,^{d,p} Luciane Amorim Santos,^e Svetoslav Nanev Slavov,^{f,g} Simone Kashima,^f Massimo Ciccozzi,^h José Lourenço,^{i,j} Eleonora Cella,^k Carla de Oliveira,^b Andréa Cany Cavalcanti,^l Luiz Carlos Junior Alcantara,^m Fernanda de Bruycker-Nogueira,^b Ana Maria Bispo de Filippis,^{b,*} and Marta Giovanetti^{n,o,**}

^aUniversidade Federal do Rio de Janeiro, Duque de Caxias, Rio de Janeiro, Brazil

^bLaboratório de Arbovírus e Vírus Hemorrágicos, Instituto Oswaldo Cruz, Rio de Janeiro, Rio de Janeiro, Brazil

^cPrograma Interunidades de Pós-Graduação em Bioinformática, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

^dDepartment of Exact and Earth Sciences, University of the State of Bahia, Salvador, Brazil

^eBahiana School of Medicine and Public Health, Salvador, Bahia, Brazil

^fBlood Center of Ribeirão Preto, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil

^gButantan Institute, São Paulo, Brazil

^hUnit of Medical Statistics and Molecular Epidemiology, Università Campus Bio-Medico di Roma, Rome, Italy

ⁱUniversidade Católica Portuguesa, Faculdade de Medicina, Biomedical Research Center, Lisboa, Portugal

^jClimate Amplified Diseases and Epidemics (CLIMADE), Portugal, Europe

^kBurnett School of Biomedical Sciences, University of Central Florida, Orlando, FL, 32827, USA

^lCentral Laboratory of Health of the Rio de Janeiro State, Brazil

^mInstituto René Rachou, Fundação Oswaldo Cruz, Minas Gerais, Brazil

ⁿDepartment of Science and Technology for Humans and the Environment, University of Campus Bio-Medico di Roma, Rome, Italy

^oOswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

Summary

Background This study focuses on urban arboviruses, specifically dengue (DENV), chikungunya (CHIKV), and Zika (ZIKV), which pose a significant public health challenge in Rio de Janeiro state, Southeast Brazil. In our research, we highlight critical findings on the transmission dynamics of these arboviruses in Rio de Janeiro, identifying distinct patterns of disease spread.

Methods By combining genomic data with case reports from the Brazilian Ministry of Health, we have analysed the phylogenetics, prevalence and spatial distribution of these endemic viruses within the state.

Findings Our results revealed sustained DENV transmission primarily in the northern part of the state, a significant ZIKV epidemic in 2016 affecting all mesoregions, and two major CHIKV outbreaks in 2018 and 2019, predominantly impacting the northern and southern areas. Our analysis suggests an inverse relationship between arboviral case incidence and urban density, with less populous regions experiencing higher transmission rates, potentially attributed to a complex interplay of factors such as the efficacy of vector control measures, environmental conditions, local immunity levels, and human mobility. Furthermore, our investigation unveiled distinct age and gender trends among affected individuals. Notably, dengue cases were predominantly observed in young adults aged 32, while chikungunya cases were more prevalent among individuals over 41. In contrast, cases of ZIKV were concentrated around the 33-year age group. Intriguingly, females accounted for nearly 60% of the cases, suggesting a potential gender-based difference in infection rates.

Interpretation Our findings underscore the complexity of arbovirus transmission and the need for interventions tailored to different geographical mesoregions. Enhanced surveillance and genomic sequencing will be essential for a deeper, more nuanced understanding of regional arbovirus dynamics. Identifying potential blind spots within the state will be pivotal for developing and implementing more effective public health strategies, specifically designed to address the unique challenges posed by these viruses throughout the state.

*Corresponding author.

**Corresponding author. Department of Science and Technology for Humans and the Environment, University of Campus Bio-Medico di Roma, Rome, Italy.

E-mail addresses: ana.bispo@ioc.fiocruz.br (A.M. Bispo de Filippis), m.giovanetti@unicampus.it (M. Giovanetti).

^pDenote equal contribution.

Translation: For the Portuguese translation of the abstract see Supplementary Materials section.

The Lancet Regional Health - Americas
2024;35: 100786

Published Online xxx
<https://doi.org/10.1016/j.lana.2024.100786>

