UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

Wylerson Guimarães Nogueira

VIGILÂNCIA METAGENÔMICA DE CORPOS D'ÁGUA NA AMAZÔNIA: uma abordagem de Saúde Única

Belo Horizonte
2025

Wylerson Guimarães Nogueira

VIGILÂNCIA METAGENÔMICA DE CORPOS D'ÁGUA NA AMAZÔNIA: uma abordagem de Saúde Única

Tese de doutorado apresentada ao Programa Interunidades de Pós-graduação em Bioinformática da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do título de Doutor em Bioinformática.

Orientador: Prof. Dr. Rommel Thiago Jucá Ramos

Belo Horizonte

2025

043 Nogueira, Wylerson Guimarães.
Vigilância metagenômica de corpos d'água na Amazônia: uma abordagem de Saúde Única [manuscrito] / Wylerson Guimarães Nogueira. – 2025.
155 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Rommel Thiago Jucá Ramos.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. Ecossistema Amazônico. 3. Água Doce. 4. Metagenômica. 5. Resistência Microbiana a Medicamentos. I. Ramos, Rommel Thiago Jucá. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 573:004

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis – CRB 6 - 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS

ATA

INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

DEFESA DE TESE
Wylerson Guimarães Nogueira

Às nove horas do dia **23 de maio de 2025**, reuniu-se, por videoconferência através do aplicativo Zoom, a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"Vigilância metagenômica de corpos d'água na Amazônia: uma abordagem de Saúde Única"**, requisito para obtenção do grau de Doutor em **Bioinformática.** Abrindo a sessão, o Presidente da Comissão, **Dr. Rommel Thiago Juca Ramos**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Professor(a) /Pesquisador(a)	Instituição	Indicação
Dr. Rommel Thiago Juca Ramos - Orientador	Universidade Federal do Pará	**Aprovado**
Dr. Alexandre Soares Rosado	King Abdullah University of Science and Technology	**Aprovado**
Dr. Aristóteles Góes Neto	Universidade Federal de Minas Gerais	**Aprovado**
Dr. Siomar de Castro Soares	Universidade Federal do Triângulo Mineiro	**Aprovado**
Dr. Ulisses de Padua Pereira	Universidade Estadual de Londrina	**Aprovado**

Pelas indicações, o candidato foi considerado: **Aprovado**

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 23 de maio de 2025.

Documento assinado eletronicamente por **Siomar de Castro Soares**, **Usuário Externo**, em 23/05/2025, às 12:25, 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 **Aristoteles Goes Neto**, **Professor do Magistério Superior**, em 23/05/2025, às 13:41, 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 **Ulisses de Padua Pereira**, **Usuário Externo**, em 23/05/2025, às 13:42, 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 **Alexandre Soares Rosado**, **Usuário Externo**, em 24/05/2025, às 04:39, 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 **Rommel Thiago Jucá Ramos**, **Usuário Externo**, em 27/05/2025, às 16:09, 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 **4220896** e o código CRC **8C2B812C**.

Referência: Processo nº 23072.230754/2025-19 SEI nº 4220896

Dedico este trabalho à memória de todos os bichinhos com que eu cresci na floresta nativa amazônica – os bichos-preguiça, macaquinhos e tucanos – que em função do desmatamento ilegal e urbanização irregular não tem hoje mais um lugar onde morar.

AGRADECIMENTOS

Nenhuma jornada se completa sozinho. E nesta longa jornada que me trouxe até aqui, eu fui agraciado com a honra e a bênção de ser acompanhado por tanta gente incrível… pessoas pelas quais eu não poderia deixar de expressar a minha eterna gratidão.

Primeiramente, eu gostaria de agradecer às parceiras – Diego, Luccas e Juan, as mais mais da vida acadêmica, perigosas que só elas. Uma veada sozinha não faz verão e, mesmo eu não gostando de sol, eu não sei o que seria de mim sem o apoio delas, principalmente nos momentos de maior insegurança e incerteza. Obrigado por serem garotas, serem mulheres e, acima de tudo, irmãs para toda a vida.

Não poderia deixar de agradecer a meus amigos que, ainda que fisicamente distantes, se fazem sempre presentes – Mateus Almeida, Raoni Kempfer, Vinícius Carvalho, Carolina Ramos, Bianca ‘Nhoim’ Ferreira, Milena ‘Cyrus’ Fontenele, Yuri ‘Yurinha’ Aguiar e Tarsis ‘Tarsila’ Menezes. Por todas as *watch parties*, presepadas, aventuras narradas e compartilhadas em horas e mais horas de chamadas de vídeo, jogatinas online, memes trocados, e inúmeros podcasts em formato de mensagem de voz: a minha eterna gratidão.

Em especial, gostaria de agradecer também a Roney Goulart Sampaio, que me adotou e também foi adotado como parte da minha *chosen family*, minha irmã a quem recorro sempre que preciso de conselhos de vida, consultando sempre a sua sabedoria. Obrigado, amigo, por todo o apoio, carinho, cuidado e consideração. Amo a Sra. Além disso, gratidão imensa por me trazer Lucas e Greg para adicionar na minha vida, por quem tenho um carinho enorme e amo muito também.

Como não poderia deixar de ser, gostaria de agradecer aos órgãos de fomento CNPQ e CAPES, pelo aporte financeiro que possibilitaram a manutenção da minha jornada no doutorado e a realização de meu intercâmbio sanduíche no exterior. Inclusive, gostaria de expressar minha mais humilde e sincera gratidão ao Prof. Alexandre Rosado, por me aceitar, acolher e aportar em seu laboratório durante a minha estadia em KAUST. Essa experiência transformadora e divisora de águas na minha vida não seria possível sem que me desses essa oportunidade. Muitíssimo obrigado por me dar essa chance de me provar pessoal e profissionalmente, Prof., eu vou colher os frutos dessa vivência pelo resto da vida e isso não seria possível sem a sua colaboração.

Estendo também os meus votos de gratidão a todo time do MEGB Lab, ao Niketan Patel e Ping Xu pelo suporte, à Lila, Sharifah e Ayman por todas as trocas científicas e, em especial, à Júnia Schultz, por toda a amizade, carinho e acolhimento em todos momentos. Para além do laboratório, gostaria de agradecer a todos os amigos que fiz durante a minha jornada em KAUST e que tornaram a experiência ainda mais gratificante – Bárbara Ribeiro, Raquel Lubambo, Pedro Cardoso, Phillipe Rosado, Erika Santoro, Israel Filho, Helena Villela, Paula García, Axel Ortiz, Rodrigo Jimenez, Sebastian 'Jezusito' Celis, Violeta Velez, Yazeed 'Zez' Faden, Ekaterina Toshcheva, Isabel Diercks, Alexandra Cardenas, Luis Limo, Gustavo Valentim, e tantas outras pessoas que ficarão na memória e no coração. Em especial, aos meus colegas de apartamento, Moinak, Ayan, Yuji, agradeço por todas as trocas, convívio e conversas, mas acima de tudo agradeço imensamente ao Anderson, que foi meu porto seguro, parceiro e colega de trabalho e aventuras desde o instante em que pisei em KAUST. Muitíssimo obrigado por tudo, amigo.

A special shoutout to my boys, Nicholas Kuntis, Stefanos Georgiadis, Biel 'El Cabrón' Garau, and Teddy Sealy – I love you, guys. Thank you so much for all the craziness, joy, and unhinged good times, I'll never forget you.

Gostaria de agradecer a todos os amigos do CVBioinfo 2023, Aline, Angie, Camila, Giovana, Luana, Henry, Rafael, Stephane e Talita, foi uma honra servir com vocês. Também ao Diego Mariano, Pamella Miranda, Joicy Xavier, Izabela Mamede e Catarina, grandes amigos que foram super importantes nessa trajetória de pesquisa e extensão. Thiago 'Thiaguinho' Sousa, Daniel Viana, Eula Neves e Arun Jaiswal por todos os momentos. Ao Felipe Caixeta, João Paulo de Almeida, Douglas, Pedro, Mateus, Gaybriel, Mário e aos demais colegas da UFMG, eu agradeço imensamente por toda a parceria nessa jornada.

Ao Prof. Vasco, agradeço imensamente por todo o apoio e incentivo durante o meu doutorado. E, também, agradeço enormemente à Profa. Glória, que esteve sempre ao meu lado, tanto como professora e coordenadora do curso de verão da bioinfo mas, também, como conselheira e mãezona durante toda a minha jornada de pós-graduação. Agradeço também a Tiago e Sheila da Secretaria, sempre dispostos a ajudar em todos os momentos, carregando nas costas a excelência do PPG, e por quem tenho um enorme carinho.

Agradeço ao time do LBGM - UFPA, à todos aqueles que eu já cheguei a conhecer pessoalmente ou não. Sem o esforço de toda a equipe local e de seus colaboradores no Pará,

esse projeto não teria sido possível. Sendo assim, sou muitíssimo grato a todos os envolvidos. Agradeço também à ex-aluna da equipe, Bruna Gois, grande amiga do laboratório para a vida.

Aos amigos do Kung Fu, agradeço demais ao Mestre Vânio, a Fernando, Léo, Graci e Henrique, meu amado amigo. Agradeço também a minha família biológica – a minha irmã, por todo o auxílio e amizade, e por me emprestar seu namorado para ler as cartas de tarot para mim; a minha mãe, por todo o amor e pela fé; e ao meu pai, por torcer por mim à distância.

E, por fim, mas extremamente importante, eu não poderia deixar de agradecer a pessoa sem a qual nada disso teria sido possível – meu orientador, Prof. Rommel Ramos. Eu acredito que nesse período tão desafiador, para muito além do aspecto intelectual do trabalho, mesmo a pessoa mais saudável e segura de si se vê afligida pela incerteza, dúvida e questionamento intrínsecos à ciência – se vendo abalada em algum momento pelo quão difícil realmente se é vivenciar um doutorado. Nesse contexto, é inquestionável a diferença que se faz ter ao seu lado não somente um mentor, mas um líder visionário que acredita no trabalho que conduz e entende que esse trabalho é feito por pessoas e, por fim, para pessoas, para dar um retorno científico para a sociedade. Com esse senso de propósito, o Prof. Rommel nos lidera rumo a metas – quando não, extremamente ambiciosas – mas sempre de um jeito muito humano. Não consigo contar a quantidade de vezes em que busquei ao Prof., inconscientemente buscando um reforço negativo do meu trabalho que eu nunca creio ser o suficiente e, ao invés de corroborar minhas expectativas, na orientação do Prof. eu encontrei acolhimento, parceria e direção real, que mantinham meus pés no chão, sonhando mas sempre caminhando, seguindo em frente. Nunca me faltou norte, apoio e compreensão. Através da minha experiência com o Prof. Rommel, eu pude não somente abraçar oportunidades que eu nunca imaginei, ampliar meus conhecimentos, me capacitar profissionalmente e concretizar sonhos que eu nem sabia serem possíveis, mas pude também reeducar diversas noções pessoais que eu nem imaginava estarem ali e carregar comigo tais aprendizados como lições de vida. Se hoje eu espero muito mais da vida, almejo um futuro em que acredito e, acima de tudo, acredito muito mais em mim mesmo é graças às oportunidades de aprendizado cedidas pelo Prof.. Eu serei grato para sempre por todos esses bons anos em que trabalhamos juntos, desde a monografia até a tese. Muitíssimo obrigado por toda a confiança, fé e apoio durante todos esses anos, Prof.. Eterna gratidão.

"*What is the bravest thing you've ever said?*
asked the boy.
'*Help,*' said the horse.
'*Asking for help isn't giving up,*' said the horse.
'*It's refusing to give up.*"

(Charlie Mackesy
– *The Boy, the Mole, the Fox and the Horse*, 2019)

RESUMO

Atualmente, a resistência antimicrobiana representa uma das principais ameaças globais à saúde pública, especialmente em regiões assoladas por desigualdades estruturais e pela degradação ambiental. Na Amazônia brasileira, fatores como a urbanização desenfreada, o desmatamento e a falta de saneamento básico favorecem o surgimento e a disseminação de doenças de veiculação hídrica e de elementos genéticos de resistência, sendo de grande importância para saúde única. Nesse cenário, a vigilância genômica de ambientes aquáticos se destaca como uma abordagem essencial para a compreensão de riscos ecológicos e sanitários emergentes. Esta tese propõe uma abordagem integrada de metagenômica ambiental voltada à vigilância de corpos d'água na região amazônica. A partir da análise de dados metagenômicos de amostras coletadas no Estado do Pará, comunidades microbianas foram caracterizadas, genes de resistência a antimicrobianos (ARGs) foram identificados e suas possíveis associações com variáveis ambientais e antrópicas foram avaliadas. Além disso, o trabalho inclui uma investigação sobre a diversidade viral em ambientes amazônicos de água doce, bem como uma revisão crítica de ferramentas computacionais para a descoberta computacional de fagos, contribuindo para a consolidação de bases metodológicas em metagenômica viral. Os resultados alcançados reforçam a importância da metagenômica como instrumento de monitoramento ambiental e vigilância em saúde pública, ampliando a compreensão sobre os impactos da ação humana na microbiota aquática da Amazônia, sob a perspectiva conceitual de Saúde Única.

Palavras-chave: Amazônia. Água doce. Metagenômica. Resistência antimicrobiana. Resistoma. Virosfera. Saúde Única.

ABSTRACT

Antimicrobial resistance is currently one of the greatest global threats to public health, especially in regions burdened by structural inequality and environmental degradation. In the Brazilian Amazon, factors such as unregulated urban expansion, deforestation, and the lack of basic sanitation contribute to the emergence and spread of waterborne diseases and genetic elements associated with resistance. In this context, genomic surveillance of aquatic environments stands out as an essential approach to understanding emerging ecological and sanitary risks. This thesis proposes an integrated environmental metagenomics approach applied to the surveillance of freshwater ecosystems in the Amazon region. Based on metagenomic data from water samples collected in the State of Pará, microbial communities were characterized, antimicrobial resistance genes (ARGs) were identified, and their possible associations with environmental and anthropogenic variables were evaluated. Additionally, the work includes an investigation of viral diversity in Amazonian freshwater ecosystems, as well as a critical review of computational tools for in silico phage discovery, contributing to the consolidation of methodological frameworks in viral metagenomics. The results reinforce the importance of metagenomics as a tool for environmental monitoring and public health surveillance, expanding our understanding of the impacts of human activity on Amazonian aquatic microbiota under the conceptual framework of One Health.

Key-words: Amazon. Freshwater. Metagenomics. Antimicrobial resistance. Resistome. Virome. One Health.

ESTRUTURA DO MANUSCRITO

Esse manuscrito referente à Tese de Doutorado está organizado em seis seções, um apêndice e onze anexos:

(I). Referencial Teórico, Justificativa e Objetivos;

(II). Artigo I - Viroma do Lago Bolonha;

(III). Artigo II - Descoberta *In Silico* de Fagos de Água Doce;

(IV). Artigo III - Resistoma de Corpos Hídricos do Pará;

(V). Conclusões e Perspectivas;

(VI). Referências Bibliográficas.

O Apêndice A contém o acesso para o material de registros fotográficos e audiovisuais coletados por drone durante as coletas de amostras de água doce realizadas neste estudo.

O Anexo A está constituído de um Prêmio de Mérito em Divulgação Científica, na categoria de *Best Paper Award*, referente ao capítulo de livro “Imunobioinformática para leigos”, publicado pela Revista BIOINFO.

Os Anexos B e C estão constituídos de artigos publicados em periódicos científicos indexados na condição de primeiro autor ou de primeira autoria compartilhada.

Os Anexos D à G estão constituídos de artigos publicados em periódicos científicos indexados na condição de co-autor.

Os Anexos H e I estão constituídos de dois livros publicados como co-organizador, disponíveis na forma de e-book publicado junto à Editora Alfahelix.

O Anexo J está constituído de oito capítulos de livros publicados – um capítulo como único autor; um capítulo como primeiro autor; e seis capítulos como co-autor.

E, por fim, o Anexo K contém o Currículo Lattes dos últimos quatro anos relatando as atividades realizadas durante os oito semestres de doutorado.

LISTA DE FIGURAS

Figura 1 – Da origem astrológica à saúde humana: a jornada da água 16

Figura 2 – A Guerra Invisível: resistência a antimicrobianos 23

Figura 3 – A Trama Oculta: ecologia microbiana revelada pela metagenômica 27

LISTA DE ABREVIATURAS E SIGLAS

ARGs	Genes de resistência a antibióticos
ONU	Organização das Nações Unidas
OMS	Organização Nacional da Saúde
WHO	*World Health Organization*
AMR	Resistência antimicrobiana
DNA	Ácido desoxirribonucleico
EUA	Estados Unidos da América
CDC	*Centers for Disease Control and Prevention*
MRSA	*Staphylococcus aureus* resistente à meticilina
MGEs	Elementos genéticos móveis
HIV	Vírus da imunodeficiência humana
AIDS	Síndrome da imunodeficiência adquirida
GAP-AMR	Plano de Ação Global em Resistência Antimicrobiana
GLASS	Sistema Global de Vigilância de Uso e Resistência Antimicrobiana
RNA	Ácido ribonucleico
rRNA	RNA ribossômico
PCR	Reação em cadeia da polimerase
NGS	Next Generation Sequencing
OTU	Unidade taxonômica
bBSI	Infecções bacterianas da corrente sanguínea
IPCC	Intergovernmental Panel on Climate Change
IDH	Índice de Desenvolvimento Humano
IBGE	Instituto Brasileiro de Geografia e Estatística
ISSN	International Standard Serial Number
PA	State of Pará, Brazil

SUMÁRIO

1. INTRODUÇÃO GERAL **15**
1.1. ÁGUA – FONTE DA VIDA E DE DOENÇAS 15
1.2. RESISTÊNCIA A ANTIMICROBIANOS 21
1.3. METAGENÔMICA – O QUE TRANSCENDE A GENÔMICA 26
2. JUSTIFICATIVA **31**
3. OBJETIVOS **33**
3.1. OBJETIVO GERAL 33
3.2. OBJETIVOS ESPECÍFICOS 33
4. ARTIGO I – VIROMA DO LAGO BOLONHA **34**
5. ARTIGO II – DESCOBERTA IN SILICO DE FAGOS DE ÁGUA DOCE **45**
6. ARTIGO III – RESISTOMA DE CORPOS HÍDRICOS DO PARÁ **56**
7. CONSIDERAÇÕES FINAIS **123**
8. PERSPECTIVAS **124**
9. REFERÊNCIAS BIBLIOGRÁFICAS **125**
10. APÊNDICE **136**
APÊNDICE A – Dados coletados por drone – Registros fotográficos e audiovisuais 137
11. ANEXOS **138**
ANEXO A – Premiação – Prêmio de Mérito em Divulgação Científica na Categoria Best Paper Award para o capítulo de livro “Imunobioinformática para leigos” 139
ANEXO B – Artigo Publicado – Computational identification of putative common genomic drug and vaccine targets in Mycoplasma genitalium 140
ANEXO C – Artigo Publicado – From In-Person to the Online World: Insights Into Organizing Events in Bioinformatics 141
ANEXO D – Artigo Publicado – Genomic and phenotypic insight into antimicrobial resistance of Pseudomonas fluorescens from King George Island, Antarctica 142
ANEXO E – Artigo Publicado – Microbiome analyses of the Uraim River in the Amazon and georeferencing analyses to establish correlation with anthropogenic impacts of land use 143
ANEXO F – Artigo Publicado – In silico approach to identify microsatellites candidate biomarkers to differentiate the biovar of Corynebacterium pseudotuberculosis genomes 144
ANEXO G – Artigo Publicado – Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil 145
ANEXO H – Livro Publicado – BIOINFO #02 - Revista Brasileira de Bioinformática e Biologia Computacional 146
ANEXO I – Livro Publicado – BIOINFO #03 - Revista Brasileira de Bioinformática e Biologia Computacional 147
ANEXO J – Capítulos de Livro Publicados 148
ANEXO K – Currículo Lattes – 01/2021 à 04/2025 149

1. INTRODUÇÃO GERAL

1.1. ÁGUA – FONTE DA VIDA E DE DOENÇAS

O Sistema Solar herdou sua água de grãos interestelares cobertos de gelo na nuvem de poeira da qual o Sol e os planetas se formaram há 4,6 bilhões de anos (ZOLENSKY *et al.*, 1999). Sob as crostas geladas de luas como Encélado e Europa, ou na geologia de Marte, podemos encontrar vestígios de água ao longo de toda a sua vizinhança celeste (HAND *et al.*, 2020). Dos oceanos a um copo de água gelada em um dia quente, a história cosmológica por trás desse líquido transparente carrega consigo as bases para a evolução da vida na Terra.

No espaço, as moléculas de água formam-se no espaço interestelar por reações químicas que se agregam de dois modos principais: na forma de objetos gelados como cometas, longe do Sol onde as temperaturas são baixas; ou, na forma de minerais hidratados, quando a água reage com materiais rochosos sob temperaturas mais altas (MORBIDELLI *et al.*, 2000). De acordo com a ciência planetária, segundo as teorias mais aceitas, acredita-se que o planeta Terra tenha herdado a sua água há pelo menos 4,3 bilhões de anos pelo choque e colisão com objetos gelados, como asteroides e cometas, que possam ter sido então assimilados pelo crescente manto da Terra (RAYMOND; IZIDORO, 2017), embora esta não seja a única explicação possível para este mundo aquático. Outras hipóteses sugerem que a maior parte da água da Terra foi assimilada ainda durante o período protoplanetário e veio à tona com o tempo, com estimativas de que podem existir até 6 vezes mais água dentro do manto rochoso terrestre do que na superfície terrestre (GU *et al.*, 2022; PIANI *et al.*, 2020).

Atualmente, apesar do mistério que perdura sob a sua origem, cerca de 71% da superfície da Terra é coberta por água, o chamado "planeta azul". Dessa água, 97% é salgada e está nos mares e oceanos, 2,2% estão na forma de gelo e permafrost e apenas 0,008% dela está disponível na forma de água doce no estado líquido na superfície do planeta (GLEICK, 1993). Sendo considerada um pré-requisito para que a vida surja e evolua, acredita-se que o surgimento da vida na Terra se deu em água líquida que, como solvente, permite a organização de biopolímeros em estruturas tridimensionais e atua como parceira química da maioria das vias bioquímicas que conhecemos (BRACK, 1993). Sendo assim, a água é um dos ingredientes fundamentais para a vida na Terra e grande parte dos organismos são compostos principalmente por ela.

Figura 01 – Da origem astrológica à saúde humana: a jornada da água. A água, elemento essencial à vida, tem uma origem ancestral ligada a processos geológicos e cósmicos. Desde os impactos de cometas na formação da Terra até sua presença em aquíferos e rios tropicais, a água percorre um ciclo que conecta ecossistemas e seres humanos. Sua qualidade e disponibilidade influenciam diretamente a saúde pública, tornando-se um elo entre o ambiente e as doenças de veiculação hídrica. Fonte: ilustração gerada por inteligência artificial, em 22 de abril de 2025.

Inclusive, o próprio corpo humano é composto em mais de 70% por água, exigindo uma ingestão regular de água para que o organismo se mantenha hidratado, transporte nutrientes até as células e elimine toxinas do corpo através do suor e da urina, desempenhando suas funções vitais (YAMADA *et al.*, 2022). De fato, um ser humano adulto em média sobrevive por apenas 3 dias sem água (POPKIN; D'ANCI; ROSENBERG, 2010). Apesar disto, mais 2,2 bilhões de pessoas em todo o mundo ainda carecem de uma gestão segura de água potável, incluindo 785 milhões sem acesso básico à água potável para beber (UNITED NATIONS, 2020). Os serviços de água, saneamento e higiene são uma das maneiras mais baratas, fáceis e eficazes de prevenir a propagação de doenças, como o coronavírus, ainda assim cerca de 3 bilhões de pessoas em todo o mundo não tem a capacidade de lavar as mãos com segurança em casa, e mesmo serviços de instalações de assistência médica ainda carecem de abastecimento básico de água (UNITED NATIONS, 2020).

A Agenda 2030 para o Desenvolvimento Sustentável da Organização das Nações Unidas (ONU) dedica a sua meta nº 06 para água limpa e saneamento, a fim de "garantir o acesso à água e saneamento para todos, a necessidade humana mais básica para a saúde e o bem-estar" (UNITED NATIONS, 2022a). Enquanto a demanda por água aumenta – devido ao rápido crescimento populacional, urbanização e aumento das necessidades de água da agricultura, indústria e setores de energia – após décadas de uso indevido, má gestão, extração excessiva de águas subterrâneas e contaminação de fontes de água doce, o estresse dos recursos hídricos atinge um nível crítico (UNITED NATIONS, 2020). Além disso, desafios crescentes relacionados à degradação dos ecossistemas relacionados à água, escassez causada pela mudança climática, subinvestimento em água e saneamento, e cooperação insuficiente em águas transfronteiriças atuam como uma grande ameaça a todos os países (UNITED NATIONS, 2020).

A Agenda 2030 da ONU também aponta, na sua meta nº 15 que trata sobre a vida terrestre, sobre a importância de "proteger, restaurar e promover o uso sustentável dos ecossistemas terrestres, gerir de forma sustentável as florestas, combater a desertificação, mitigar e reverter a degradação dos solos e deter a perda da biodiversidade" (UNITED NATIONS, 2022b). De acordo com o mais recente Relatório dos Objetivos de Desenvolvimento Sustentável (UNITED NATIONS, 2020), nos últimos 300 anos, mais de 85% das zonas úmidas do planeta foram perdidas e degradadas, principalmente por meio de drenagem e conversão de terras, levando 81% das espécies dependentes desses ambientes à beira da extinção de 1970 até hoje, com um número crescente de espécies definitivamente extintas.

Conforme descrito pela Organização Mundial da Saúde (OMS), melhorias no saneamento e higiene da água, gestão de águas residuais em todos os setores e maiores esforços em vigilância e pesquisa são componentes fundamentais para o atingimento dessas metas (WHO, 2020). Sem uma melhor infraestrutura e gestão, milhões de pessoas morrem todos os anos de doenças relacionadas com a água, como a diarreia, e perdas irreparáveis na biodiversidade e na resiliência do ecossistema minam a prosperidade e os esforços para um futuro mais sustentável. Além disso, ameaças à saúde pública relacionadas à condições precárias de higiene, transmissibilidade por água e à resistência antimicrobiana têm levado a um número cada vez maior de doenças infecciosas emergentes (VELAZQUEZ-MEZA *et al.*, 2022).

Os impactos das ações antrópicas no meio ambiente, incluindo o manejo da terra e da vida animal, doméstica ou selvagem, e das mudanças climáticas podem levar a contaminação da água e diminuir a capacidade dos ecossistemas e zonas úmidas de filtrar a água naturalmente, promovendo a transmissão e disseminação de doenças (SHAHEEN, 2022). O entendimento de que a saúde humana, a saúde animal e a saúde ambiental estão intrinsecamente inter-relacionadas é um componente-chave para garantir a qualidade de vida e bem-estar de toda a humanidade. Esse conceito holístico é chamado de *One Health*, ou Saúde Única, e seu conceito é definido como "o esforço colaborativo de múltiplas disciplinas – trabalhando local, nacional e globalmente – para alcançar a saúde ideal para pessoas, animais e nosso meio ambiente" (ADISASMITO *et al.*, 2022). Ao levar em consideração todos os três aspectos da saúde, podem ser geradas soluções que não apenas abordem os problemas de saúde de um grupo específico, mas também mitiguem a origem desses problemas.

As abordagens em *One Health* voltadas ao ambiente aquático possibilitam a identificação de vias críticas de exposição à patógenos, o desenvolvimento de sistemas de vigilância baseados em água para detecção precoce de agentes causadores de doenças em nível populacional, e a implementação de abordagens de intervenção para bloquear as vias críticas de exposição (O'BRIEN; XAGORARAKI, 2019). Quanto à patógenos virais, o transporte de vírus no ambiente facilitado devido aos seus pequenos tamanhos em comparação com outros patógenos, a sua resistência à desinfecção e a capacidade de sobreviver por períodos prolongados em água e sólidos tornam provável a transmissão de vírus do ambiente para hospedeiros suscetiveis (XAGORARAKI; YIN; SVAMBAYEV, 2014). Além disso, a sua baixa dose infecciosa necessária para causar uma nova infecção, a incapacidade de serem tratados por antibióticos e sua propensão à mutação adaptativa tornam vírus dispersos pela água no ambiente em uma potencial ameaça de saúde pública. Além disso, por não se replicarem fora de suas células hospedeiras, a detecção de vírus em amostras ambientais pode indicar contaminação direta de excrementos relativos à população humana ou animal (XAGORARAKI; YIN; SVAMBAYEV, 2014).

No que tange à patógenos bacterianos, a ameaça imposta pelo tratamento inadequado das águas, contaminação de corpos hídricos e ausência de saneamento básico na interface ambiente, humano e animal é ainda mais drástica por fomentar as condições ideais para a promoção da resistência antimicrobiana (AMR). A resistência aos antibióticos pode surgir tanto de mutações no genoma pré-existente de uma bactéria quanto da incorporação de DNA

estranho. No que diz respeito à absorção de novos fatores de resistência, a imensa diversidade do microbioma ambiental presente na água, no solo e em outros ambientes com nichos ecológicos altamente variáveis fornecem um *pool* genético incomparável, fornecendo numerosos genes que potencialmente podem ser adquiridos e usados por patógenos para neutralizar o efeito de antibióticos (LARSSON; FLACH, 2022). Atualmente, os EUA registram mais de 2,8 milhões de infecções resistentes a antimicrobianos ao ano e mais de 35.000 pessoas morrem como resultado dessas infecções multirresistentes (CDC, 2019).

A água, portanto, se posiciona tanto como um bem essencial à vida quanto como um potencial agente disseminador de doenças. Um exemplo emblemático da importância da água e de sua dualidade para a saúde pública remonta a 1854, quando John Snow conduziu o primeiro estudo epidemiológico até então documentado ao rastrear a origem de um surto de cólera até uma bomba de água contaminada em Londres, considerado como um marco fundador da epidemiologia moderna (TULCHINSKY, 2018).

Como parte da estratégia para combater possíveis novas infecções e dispersão de elementos de resistência a antimicrobianos no meio ambiente é fundamental que seja realizado o monitoramento de corpos hídricos que sirvam como reservatórios de água para o consumo humano e animal, a vigilância genômica de efluentes e águas residuais, e a caracterização do microbioma desses ambientes aquáticos, saudáveis ou contaminados (DAVIS *et al.*, 2023; KO; CHNG; NAGARAJAN, 2022).

Nesse intuito, estudos recentes mapearam genes de resistência a antibióticos (ARGs) por abordagens em metagenômica em estações de tratamento de esgoto em Hong Kong (YANG *et al.*, 2014), em efluentes hospitalares (MARKKANEN *et al.*, 2023; TALAT *et al.*, 2023), em águas costeiras no Reino Unido (LEONARD *et al.*, 2018), e por outras abordagens nos Lagos Geneva e Maggiore (ECKERT *et al.*, 2018), na bacia do Rio Platte Sul (PRUDEN; ARABI; STORTEBOOM, 2012) e em águas residuais urbanas em dez países europeus (HUIJBERS; LARSSON; FLACH, 2020). Outros estudos realizaram a caracterização metagenômica geral de amostras de água dos Lagos Tanganyika (TRAN *et al.*, 2021), do Lago Han (MOON *et al.*, 2020), de lagos na Nova Zelândia (BIESSY *et al.*, 2022), do Lago Bolonha (ALVES *et al.*, 2020) e de outros lagos da bacia amazônica (TOYAMA *et al.*, 2016). Adicionalmente, estudos de identificação de ARGs na microbiota intestinal humana de indivíduos residentes na China, Espanha, Dinamarca e Estados Unidos utilizando estratégias de metagenômica também foram realizados (KARKMAN; PÄRNÄNEN; LARSSON, 2019),

o que auxilia no entendimento da dinâmica de contaminação na interface entre humanos e o ambiente e no estabelecimento de correlações entre a dispersão de elementos de resistência e as suas rotas de distribuição geográficas entre as populações.

Da ancestralidade cosmológica da água para origem da vida até os dias atuais, a ciência ainda não conhece plenamente a infinitude da nossa origem interestelar e, tampouco, a diversidade do universo microscópico dos ambientes aquáticos, responsável pela evolução e manutenção da vida na Terra. Avanços em pesquisa empregando abordagens como a metagenômica de amostras ambientais de água, de ambientes saudáveis e impactados, são essenciais para auxiliar na compreensão da ecologia desses berços hídricos de doenças emergentes, e no entendimento de seus mecanismos de evolução e de manutenção desses microbiomas, a fim de desenvolver sistemas de vigilância, de avaliação de risco e de planos de resposta e controle contra futuros surtos e epidemias que ameacem o equilíbrio em *One Health*.

1.2. RESISTÊNCIA A ANTIMICROBIANOS

"Resistência" é um termo abrangente adotado por diversas áreas do conhecimento. Quando avaliamos diversos tópicos e disciplinas, as suas distintas definições demonstram que o que estudiosos reconhecem como resistência nas humanidades e na natureza pode ser observado a partir de um amplo espectro de comportamentos, configurações e fenômenos naturais (HOLLANDER; EINWOHNER, 2004).

Em uma das citações mais conhecidas de Michel Foucault, ele afirma que "onde há poder, há resistência" (VERA; FOUCAULT, 1979). E, nas ciências da vida, quando esse poder se manifesta na forma de uma pressão seletiva pela exposição a substâncias e agentes antimicrobianos, ela é chamada de "resistência antimicrobiana" (LARSSON, D. G.JOAKIM; FLACH, 2022). A resistência antimicrobiana (AMR) é definida pela *Nature Portfolio* como a capacidade de microrganismos, como bactérias, fungos ou protozoários, de se multiplicarem apesar da exposição a substâncias antimicrobianas naturalmente disponíveis ou projetadas para inibir seu crescimento (LARSSON, D. G.J. *et al.*, 2023).

Na natureza, a resistência existe naturalmente, onde várias espécies de microrganismos desenvolveram a capacidade de tolerar moléculas nocivas para si (DCOSTA *et al.*, 2011). Registros de cavernas isoladas, núcleos de permafrost e outros ambientes e espécimes que foram preservados da contaminação bacteriana antropogênica ajudam a revelar informações sobre os mecanismos de resistência prevalentes durante a era pré-antibiótica (BHULLAR *et al.*, 2012; LUGLI *et al.*, 2017; PERRY; WAGLECHNER; WRIGHT, 2016), onde a competição interminável por recursos entre os microorganismos, incluindo a produção natural de metabólitos secundários semelhantes a antibióticos atuais, promovia o surgimento de agentes de resistência a antimicrobianos (ALLEN *et al.*, 2010). Inclusive, foi com base na observação desse fenômeno natural em 1928, notando a inibição do crescimento bacteriano em uma placa de petri contaminada com fungo do gênero *Penicillium*, que o Sir Alexander Fleming descobriu o primeiro antibiótico da era moderna - a penicilina (FLEMING, 1929; TAN; TATSUMURA, 2015).

Ao longo da história, o manejo de infecções microbianas já foi documentado previamente no antigo Egito, Grécia e China (SENGUPTA; CHATTOPADHYAY; GROSSART, 2013). Contudo, a revolução na medicina moderna aconteceu com a descoberta da penicilina por Sir Alexander Fleming e com o desenvolvimento do seu método de

purificação pelos farmacêuticos Ernst Boris Chain e Howard Walter Florey, o que permitiu a síntese e distribuição comercial da penicilina para a população, dando início a chamada Era dos Antibióticos (AMINOV, 2010).

Desde então, os antibióticos não apenas transformaram a medicina moderna e salvaram milhões de vidas, mas também desempenharam um papel fundamental na obtenção de grandes avanços na prática clínica e cirúrgica (GOULD; BAL, 2013). Os antibióticos preveniram e trataram com sucesso infecções comuns a pacientes que recebem tratamentos de quimioterapia, que têm doenças crônicas, como diabetes, doenças renais, artrite reumatoide, ou que passam por cirurgias complexas, como transplantes de órgãos, traumatológicas ou cirurgias cardíacas (ROSSOLINI *et al.*, 2014). Alterando o resultado de infecções bacterianas, os antibióticos também ajudaram a prolongar a expectativa de vida para pessoas nos Estados Unidos em mais de 24 anos, de 1920 à 2015, apresentando efeitos benéficos semelhantes em todo o mundo (PIDDOCK, 2012; VENTOLA, 2015). Nos países em desenvolvimento, onde o saneamento é precário, os antibióticos também ajudam a combater a morbidade e a mortalidade causadas por infecções transmitidas por alimentos e outras causas relacionadas à pobreza (ROSSOLINI *et al.*, 2014).

Após as primeiras prescrições de antibióticos para o tratamento de infecções graves na década de 1940, apesar do sucesso da penicilina no controle de infecções bacterianas entre soldados da Segunda Guerra Mundial, na década seguinte a resistência à penicilina já havia se tornado um problema clínico substancial (SPELLBERG; GILBERT, 2014). Em resposta, novos antibióticos foram investigados, desenvolvidos e implantados, os beta-lactâmicos. Contudo, ainda na mesma década, os primeiros casos de cepas de *Staphylococcus aureus* resistentes à meticilina (MRSA) foram notificados no Reino Unido e nos Estados Unidos (SENGUPTA; CHATTOPADHYAY; GROSSART, 2013), já sinalizando o desafio de saúde global que a resistência aos antibióticos viria a se tornar.

A introdução de antibióticos como agentes clínicos mudou radicalmente as pré-condições para a evolução de microrganismos e disseminação da resistência a antimicrobianos, fornecendo pressões de seleção sem precedentes, especialmente em membros da microbiota de humanos e animais domésticos, e presentes em ambientes fortemente poluídos com antibióticos (ALCOCK *et al.*, 2020). Essa pressão de seleção imposta pelos antibióticos promove a evolução das bactérias para escapar dos seus efeitos, o que pode se dar através de vários mecanismos diferentes.

Figura 02 – A Guerra Invisível: resistência a antimicrobianos. A resistência antimicrobiana configura-se como uma ameaça invisível que emerge na interface entre saúde humana, animal e ambiental. Dentro de corpos d'água, bactérias resistentes se proliferam e compartilham genes de resistência por meio de elementos genéticos móveis, como plasmídeos e fagos. Nesta ilustração, um campo simbólico de batalha revela bactérias protegidas por escudos, elementos de transferência gênica e os impactos seletivos do uso inadequado de antimicrobianos. Ao fundo, o ambiente natural, os animais e instituições de saúde ressaltam a importância da abordagem integrada de Saúde Única. Fonte: ilustração gerada por inteligência artificial, em 22 de abril de 2025.

Enquanto algumas bactérias são capazes de neutralizar um antibiótico alterando seu componente para torná-lo ineficaz, outras podem exportar os antibióticos para fora das bactérias, e, algumas podem modificar sua estrutura externa ou receptores para que os antibióticos não possam se ligar a eles (PAUL; DAS, 2022). Esses mecanismos podem fazer com que algumas bactérias sobrevivam ao uso de um antibiótico específico e, posteriormente, esta resistência pode ser transmitida a outras bactérias à medida que se multiplicam.

Os mecanismos de resistência aos antibióticos podem ser categorizados em quatro grupos:

- resistência intrínseca – onde as bactérias podem sobreviver a um antibiótico devido à resistência intrínseca através da evolução, estrutura ou componentes (i.e., um antibiótico que afeta o mecanismo de construção da parede das bactérias não pode afetar as bactérias que não possuem parede celular);

- resistência adquirida – onde a capacidade de resistir à atividade de um determinado agente antimicrobiano ao qual era previamente suscetível é adquirida, seja esta por meio de uma nova mutação genética ou assimilando o DNA de uma bactéria que já é resistente;
- mudança genética – o DNA da bactéria pode mudar e alterar a produção de proteínas, levando a diferentes componentes e receptores bacterianos que tornam a bactéria não reconhecível ao antibiótico;
- transferência de DNA – onde as bactérias adquirem material genético externo contendo genes de resistência por meio de uma transferência horizontal de genes com a ajuda de elementos genéticos móveis (MGEs), como plasmídeos, transposons ou fagos, o que geralmente ocorre por transformação (através da incorporação de DNA nu), por transdução (através de fagos), ou por conjugação (por contato direto) (HABBOUSH; GUZMAN, 2022).

Os impactos de eventos de mobilização e transferência horizontal de genes de resistência a antibióticos (ARGs) para espécies bacterianas podem ser drásticos quando se tratam de agentes patogênicos. De acordo com relatório de 2013, casos de infecções por MRSA mataram mais americanos ao ano do que HIV/AIDS, doença de Parkinson, enfisema pulmonar e homicídio juntos (GOLKAR; BAGASRA; GENE PACE, 2014; GROSS, 2013). Atualmente, epidemias globais de *Staphylococcus aureus*, *Candida spp.*, *Neisseria gonorrhoeae* e *Enterococcus* multirresistentes representam as maiores ameaças (ROSSOLINI *et al.*, 2014; CDC, 2019), paralelamente a disseminação global de resistência entre patógenos respiratórios comuns, incluindo *Streptococcus pneumoniae* e *Mycobacterium tuberculosis*, que também é epidêmica (ROSSOLINI *et al.*, 2014).

A resistência antimicrobiana causa 1,3 milhões de mortes e contribui para mais 5 milhões de mortes anualmente e, até 2030, a AMR poderia forçar mais de 24 milhões de pessoas à pobreza extrema (WHO, 2025). Enquanto isso, testes diagnósticos novos e acessíveis, ferramentas de predição de doenças, vacinas e alternativas não antimicrobianas continuam indisponíveis. Para combater essa ameaça, a Organização Mundial da Saúde (OMS) formulou o plano de ação global em resistência antimicrobiana (GAP-AMR) que foi adotado por todos os países em 2015 por meio de decisões na Assembleia Mundial da Saúde (WHO, 2017). O GAP-AMR estabelece responsabilidades para as nações e seus parceiros nacionais e internacionais envolvidos na resposta global à AMR. E, desde então, alguns

progressos foram feitos, como a inauguração do Sistema Global de Vigilância de Uso e Resistência Antimicrobiana (GLASS), o primeiro sistema de vigilância global a incorporar dados nacionais oficiais sobre resistência antimicrobiana e consumo de antimicrobianos em humanos, mais recentemente atualizado em 2022 (ACHMAD ALI FIKRI, SYAMSUL ARIFIN, 2022).

Apesar dessas iniciativas, os avanços em pesquisa científica, de forma multidisciplinar e multissetorial, continuam sendo fundamentais para embasar as intervenções, soluções e abordagens mais adequadas no combate à resistência a antimicrobianos. Como bactérias e genes frequentemente cruzam ambientes e fronteiras de espécies, é fundamental entender e reconhecer as conexões entre a microbiota humana, animal e ambiental para gerenciar esse desafio de saúde pública global (BUSCHHARDT *et al.*, 2021; MACKENZIE et al., 2019).

O envolvimento de formuladores de políticas, chefes de governo, ministros e cientistas por todo o espectro da *One Health* (saúde humana, animal e ambiental) é primordial na busca de resoluções para o desafio da resistência antimicrobiana (LARSSON, D. G. J. *et al.*, 2023). O fortalecimento dos vínculos entre pesquisa, política e prática clínica, humana e veterinária, é essencial para gerar soluções sustentáveis e uma agenda de pesquisa mais ampla para atender às necessidades do combate de patógenos multirresistentes em todos os setores.

1.3. METAGENÔMICA – O QUE TRANSCENDE A GENÔMICA

No sentido moderno da nomenclatura, o prefixo *meta-* faz alusão ao auto-referencial, transmitindo a ideia de "mais elevado do que, transcendente, abrangente, lidando com os assuntos mais fundamentais de" (HARSANYI, 1973). Sendo assim, a *metagenômica* poderia ser entendida como a genômica da genômica, ou aquilo que a transcende.

Sob a influência dos estudos de Carl Woese, figura proeminente da ecologia e evolução microbiana, a visão da ciência sobre a diversidade biológica mudou fundamentalmente com o advento da utilização da molécula de RNA ribossômico (rRNA) 16S como um marcador filogenético (WOESE, 1987). Com estes trabalhos, Carl Woese adicionou '*archaea*' à árvore da vida como um novo domínio (WOESE; FOX, 1977) e preparou o gene 16S rRNA para ser usado posteriormente como uma ferramenta em estudos de sequenciamento de amplicon de alto rendimento para explorar a ecologia microbiana. As ideias apresentadas na revisão do autor, intitulada "*Bacterial Evolution*", trouxeram perspectivas que impactam ainda hoje a maneira como a metagenômica e a microbiologia são percebidas. Uma destas premissas destaca como a pesquisa em microbiologia tem sido fundamentalmente impulsionada pela tecnologia. Em particular, como a nossa capacidade de sequenciar genes mudou drasticamente nossa compreensão da diversidade e da evolução das proteínas (WOESE, 1987).

Superando algumas das limitações de estudos dependentes de cultivo microbiológico convencional, os estudos de sequenciamento do gene 16S rRNA expandiram nossa compreensão da diversidade e ecologia microbiana, catalisados pelo melhoramento de primers de PCR e pelo sequenciamento de Sanger, mais economicamente viável (ESCOBAR-ZEPEDA; DE LEÓN; SANCHEZ-FLORES, 2015). Contudo, ainda faltava contexto genômico para que o sequenciamento de amplicon conseguisse extrapolar capacidades metabólicas e fenótipos dentre representantes cultivados em laboratório e novos filotipos recém-descobertos. Do ponto de vista ecológico, como ilustrado por Carl Woese (em tradução livre), "no extremo, as trocas de genes entre espécies podem ser tão desenfreadas, tão amplamente disseminadas, que uma bactéria única não teria isoladamente uma história por direito próprio; seria uma quimera evolutiva, cada uma com sua própria história" (WOESE, 1987). E, em vista dessa complexidade, e na tentativa de explorar o material genético da microflora do solo, a pesquisadora Jo Handelsman cunhou o termo "metagenômica",

inferindo que "ir além do genoma" nesse contexto seria abarcar para além de um organismo o "o genoma coletivo" de um microambiente (HANDELSMAN *et al.*, 1998).

Deste modo, a metagenômica, também conhecida como ecogenômica, genômica comunitária ou genômica ambiental, é o estudo de todo o material genético recuperado diretamente de amostras ambientais, sendo considerado uma ciência emergente que explora várias abordagens genômicas para caracterizar as comunidades microbianas nas amostras ambientais e desdobrar os genomas de microrganismos não cultivados (RIESENFELD *et al.*, 2004; SCHMEISSER *et al.*, 2007; UHLIK *et al.*, 2013). Estima-se que as bactérias, ainda que invisíveis a olho nu, totalizam mais de 70 gigatoneladas de carbono, representando cerca de 15% da biomassa total da Terra, uma fração superada apenas pelas plantas (BAR-ON; PHILLIPS; MILO, 2018).

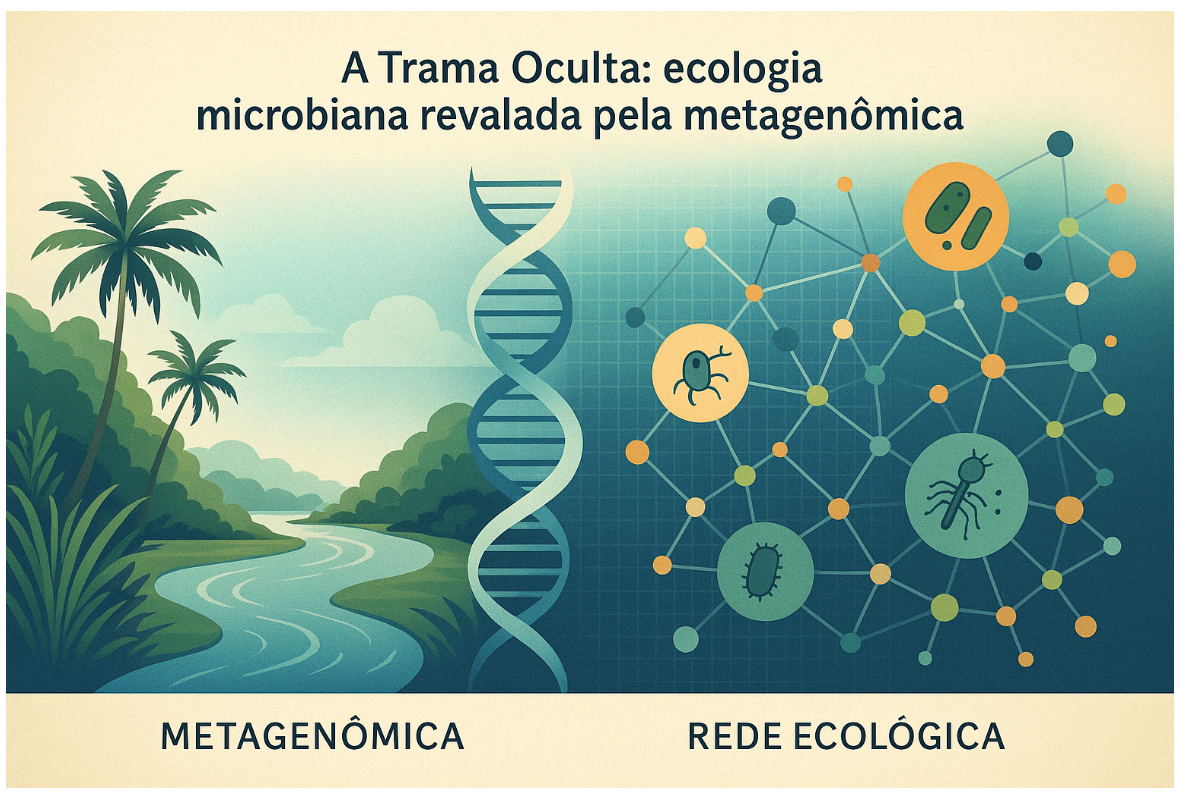


Figura 03 – A Trama Oculta: ecologia microbiana revelada pela metagenômica. A metagenômica permite investigar comunidades microbianas diretamente em seus ambientes, revelando interações ecológicas e funções genéticas que antes permaneciam invisíveis aos métodos clássicos. Nesta ilustração, diferentes microrganismos compõem uma rede ecológica complexa, onde o microbioma é representado como um sistema interconectado de espécies, genes e processos funcionais. A abordagem metagenômica transforma dados genéticos em conhecimento ecológico, ampliando a compreensão sobre a dinâmica microbiana em ecossistemas naturais. Fonte: ilustração gerada por inteligência artificial, em 22 de abril de 2025.

O conceito de metagenômica introduzido por Handelsman *et al.* envolve a extração do metagenoma (DNA genômico total de uma determinada amostra ambiental) para sua posterior fragmentação, clonagem, transformação e triagem de biblioteca construída (HANDELSMAN *et al.*, 1998). Posteriormente, com o advento das tecnologias de sequenciamento NGS de alto rendimento, uma nova abordagem metagenômica surgiu, uma triagem baseada em sequência. Essa abordagem foi demonstrada pela primeira vez pelo sequenciamento do genoma ambiental do Mar dos Sargaços, que revelou a vasta diversidade filogenética e metabólica das suas comunidades microbianas (VENTER *et al.*, 2004; YUN; RYU, 2005).

No contexto atual, as abordagens metagenômicas mais empregadas diferenciam-se por adotar duas estratégias distintas: sequenciamento direcionado, também conhecido como sequenciamento por amplicon ou *targeted sequencing*; ou por sequenciamento de shotgun. Na metagenômica de sequenciamento direcionado, a diversidade de um único gene é investigada em um ambiente, sendo mais frequentemente empregada para investigar tanto a diversidade filogenética quanto a abundância relativa de um determinado gene em uma amostra (TECHTMANN; HAZEN, 2016). Enquanto que na metagenômica por shotgun, o conteúdo genômico total de uma comunidade ambiental é investigado por meio do sequenciamento genômico.

Para realizar metagenômica direcionada, o DNA ambiental é extraído e o gene de interesse é amplificado por PCR usando primers projetados para amplificar uma maior diversidade de sequências para o gene de interesse. Por sua vez, esses genes amplificados são sequenciados usando tecnologia NGS, 16S rRNA *amplicon sequencing*, o que resulta em milhares de leituras de rRNA de subunidades pequenas por amostra e permite a sondagem de centenas de amostras simultaneamente (GUPTA *et al.*, 2019). Durante a análise dos dados, as sequências geradas podem ser agrupadas por sequências próximas em unidades taxonômicas (OTU), a fim de evitar que possíveis erros oriundos do sequenciamento de amplicon influenciem nas análises (MORGAN; HUTTENHOWER, 2012). A metagenômica direcionada captura a diversidade de um único gene de interesse, mas é limitada pela universalidade dos primers de PCR escolhidos para a análise (PARADA *et al.*, 2016).

Na metagenômica por shotgun, o DNA total da amostra ambiental é extraído e fragmentado para preparar as suas bibliotecas de sequenciamento, que são posteriormente sequenciadas para determinar o conteúdo genômico total da amostra (QUINCE *et al.*, 2017). Através desta técnica, o potencial funcional de uma comunidade microbiana pode ser

identificado e estudado, mas uma importante limitação é a profundidade do seu sequenciamento e o uso de réplicas para dar robustez ao estudo. É importante que trabalhos que adotem essa abordagem garantam uma boa cobertura de todo o conteúdo genômico ali presente para que se torne possível uma análise abrangente do microbioma estudado (DELMONT; SIMONET; VOGEL, 2012).

Apesar das suas limitações, a metagenômica é uma abordagem rápida, eficiente e altamente versátil que supera as restrições das técnicas moleculares convencionais. Com a diminuição dos preços de sequenciamento, a popularidade das tecnologias de NGS, e com o desenvolvimento de diversas ferramentas computacionais e avanço da bioinformática, a aplicação de técnicas de metagenômica tornou-se mais ampla e escalonável, impulsionando o seu emprego em diversas áreas (GHOSH; MEHTA; KHAN, 2018). Desde a caracterização de microrganismos ambientais e prospecção de novos genes e enzimas com potencial biotecnológico (NWACHUKWU; BABALOLA, 2022) à vigilância genômica em saúde pública (KO; CHNG; NAGARAJAN, 2022), as abordagens metagenômicas são aplicadas de formas cada vez mais abrangentes.

A metagenômica tem uma ampla gama de aplicações, desde amostras clínicas à ambientais, voltada desde a segurança alimentar, estudos forenses, análises de resíduos industriais e à identificação de patógenos. Por exemplo, a metagenômica pode detectar vírus humanos e animais em amostras de alface, o que pode identificar a contaminação viral das folhas verdes no campo (AW; WENGERT; ROSE, 2016). Essas abordagens ajudam a determinar a fonte de contaminação microbiana e viral e podem ser usadas para identificar as etapas críticas em que essa contaminação ocorre e implementar processos aprimorados para garantir a qualidade e a segurança dos alimentos (DOYLE; O'TOOLE; COTTER, 2017).

Na pesquisa industrial, a metagenômica funcional tem sido utilizada para a identificação de diversos biocatalisadores disponíveis no mercado (FERNÁNDEZ-ARROJO *et al.*, 2010). Abordagens de metagenômica, metatranscriptômica e metaproteômica aprimoram a descoberta de enzimas de interesse biotecnológico e podem ser adotadas para rastrear com eficiência enzimas altamente ativas (MADHAVAN *et al.*, 2017). Além disso, a metagenômica também pode ser utilizada na biorremediação, ajudando na identificação de diferentes microrganismos, amplamente disseminados, que habitem ambientes poluídos e possuam a capacidade de degradar poluentes tóxicos (DEVARAPALLI; KUMAVATH, 2015).

Ademais, a metagenômica também está presente em investigações médicas e forenses para resolver desafios englobando os campos da medicina, agricultura e ecologia, em vários âmbitos. No desenvolvimento de diagnósticos, abordagens metagenômicas têm sido usadas para identificar novos biomarcadores bacterianos não invasivos a partir de amostras fecais para diagnóstico de câncer colorretal (LIANG *et al.*, 2017), e para detectar infecções bacterianas da corrente sanguínea por 16S rRNA (bBSI) (DECUYPERE *et al.*, 2016). No âmbito da vigilância genômica de patógenos, a metagenômica viral tem cumprido um papel fundamental na identificação da causa de novas doenças epidêmicas (MOKILI *et al.*, 2012) e, também, no monitoramento de surtos de patógenos reemergentes (SHAW *et al.*, 2020).

Em suas abordagens e métodos, a metagenômica contorna a incapacidade de cultivar a maioria dos microrganismos, enquanto reconhece a necessidade do desenvolvimento de métodos computacionais que maximizem a compreensão da composição genética e de atividades de comunidades microbianas tão complexas. Retornando à sua etimologia, *meta-* como "transcendente", a metagenômica é tanto um conjunto de técnicas de pesquisa quanto um campo de pesquisa por si só, transcendendo o organismo individual para focar nos genes da comunidade e em como estes genes podem influenciar nas atividades uns dos outros, ao servir às suas funções coletivas (NATIONAL RESEARCH COUNCIL, 2007). Nas próximas décadas, com a ajuda da metagenômica, entenderemos cada vez mais as comunidades microbianas e o conjunto de comunidades que formam a nossa biosfera, tal como um sistema holístico de sistemas, que tanto incluem os seres humanos quanto dos quais intrinsecamente depende a sobrevivência humana.

2. JUSTIFICATIVA

A Organização Mundial da Saúde (OMS) declarou em 2020 que a resistência antimicrobiana (AMR) é uma das 10 principais ameaças à saúde pública e ao desenvolvimento global no século XXI (WHO, 2025). A ausência de saneamento básico e o uso indevido de antibióticos, tanto na medicina humana quanto na agroindústria, têm favorecido o surgimento e a disseminação de patógenos multirresistentes em diversos contextos ambientais (MURRAY, 2020; RUIZ, 2021). Ao mesmo tempo, vírus presentes na água, como os fagos, também desempenham papeis centrais na dinâmica ecológica desses ambientes, podendo tanto regular populações bacterianas quanto atuar como vetores de genes de resistência (MOON *et al.*, 2020).

Essas interações complexas entre comunidades microbianas e virais em ecossistemas aquáticos têm ganhado destaque em estratégias de vigilância ambiental e epidemiológica (DE MANDAL *et al.*, 2021). Principalmente, à medida que as mudanças climáticas induzidas pelas ações humanas causam perturbações perigosas na natureza, afetando a vida de bilhões de pessoas em todo o mundo, onde comunidades e ecossistemas menos capazes de lidar com elas sofrem cada vez mais e estão sujeitos a riscos de eventos naturais catastróficos (IPCC, 2022). Em especial, a região Norte do Brasil apresenta um dos piores índices de saneamento público do país: segundo o IBGE, mais de 80% dos domicílios paraenses não estão conectados à rede de tratamento de esgoto e 51.1% não têm sequer abastecimento pela rede geral de distribuição de água. Além disso, o Estado do Pará possui o quinto menor Índice de Desenvolvimento Humano (IDH) do Brasil, 0.690 (IBGE, 2025). Essa fragilidade estrutural, somada à crescente pressão antrópica sobre os ecossistemas da Amazônia, representa um cenário de risco significativo para o surgimento e a reemergência de doenças infecciosas associadas à água (LEVY *et al.*, 2018).

A Amazônia abriga uma enorme diversidade de microrganismos, muitos dos quais ainda desconhecidos, e encontra-se sob constante ameaça devido ao desmatamento, à urbanização desordenada, à poluição e às mudanças climáticas (FERNANDES *et al.*, 2018; FRANCO FILHO *et al.*, 2019; ELLWANGER *et al.*, 2020). A abundância microbiana e viral desses ambientes, associada à degradação ambiental e à vulnerabilidade social, pode favorecer a emergência de novos agentes infecciosos e o agravamento da crise global da resistência

antimicrobiana, que já é diretamente responsável por mais de 1,27 milhão de mortes ao ano mundialmente (MURRAY *et al.*, 2022).

Nesse cenário, a aplicação de ferramentas de metagenômica voltadas ao meio ambiente desponta como uma abordagem essencial para a caracterização desses riscos (ALVES *et al.*, 2020). Ao permitir a investigação simultânea de comunidades microbianas, virais e elementos genéticos associados à resistência antimicrobiana em amostras ambientais, a metagenômica amplia nossa capacidade de compreender a ecologia microbiana e viral em ambientes tropicais e altamente biodiversos, tais como a Amazônia.

Para além da caracterização do resistoma de corpos d'água utilizados para o abastecimento populacional, investigações paralelas sobre a virosfera deste bioma e sobre metodologias computacionais voltadas à descoberta de fagos contribuem para a formação de um panorama mais amplo e abrangente em vigilância genômica e avaliação de risco da região. Dessa forma, este trabalho se desenvolve sob uma perspectiva integrada e interdisciplinar, alinhada aos princípios de Saúde Única (*One Health*), buscando compreender como fatores ambientais, microbianos e sociais se inter-relacionam na dinâmica de saúde ambiental da Amazônia.

3. OBJETIVOS

3.1. OBJETIVO GERAL

- Investigar por meio de abordagens metagenômicas a diversidade microbiana, viral e o resistoma de corpos d’água em ambientes amazônicos, visando compreender suas implicações ecológicas e sanitárias sob a perspectiva de Saúde Única.

3.2. OBJETIVOS ESPECÍFICOS

- Coletar amostras de água doce de corpos hídricos do Estado do Pará, em triplicata, em diferentes localidades amazônicas;
- Caracterizar a diversidade viral presente nas amostras oriundas de estudos prévios;
- Explorar a aplicação de metodologias *in silico* para identificação de vírus bacteriófagos em ambientes aquáticos;
- Realizar sequenciamento de alto rendimento do metagenoma de amostras ambientais coletadas;
- Caracterizar a diversidade microbiana presente nas amostras sequenciadas;
- Identificar e classificar os genes de resistência a antimicrobianos (ARGs) presentes nas amostras de sequenciamento;
- Analisar os impactos de ações antrópicas e ambientais com base nos registros obtidos para os pontos de coleta;
- Correlacionar os dados ecológicos e metagenômicos para identificar padrões relacionados ao resistoma e potenciais zonas de risco associados à resistência antimicrobiana amazônica.

4. ARTIGO I – VIROMA DO LAGO BOLONHA

O primeiro artigo que compõe esta tese, indexado em periódico científico, é intitulado "Viral Metagenomics Reveals Widely Diverse Viral Community of Freshwater Amazonian Lake", e foi publicado junto à revista *Frontiers in Public Health* (ISSN: 2296-2565), sob o *Research Topic* "Tackling the Presence of Pathogens in Water and Wastewater", em abril de 2022, fator de impacto: 3.0 (2024).

Este estudo (https://doi.org/10.3389/fpubh.2022.869886) apresenta uma análise metagenômica da diversidade viral presente no Lago Bolonha, importante reservatório de água doce localizado na região metropolitana de Belém (PA), com foco especial em vírus bacteriófagos e cianofágicos. Através de abordagens de sequenciamento de nova geração (NGS), o trabalho revelou a presença de uma comunidade viral ampla e funcionalmente diversa, composta majoritariamente por vírus com potencial ecológico relevante. Os resultados indicam que os fagos podem desempenhar papeis fundamentais na regulação de populações bacterianas, no controle da proliferação de cianobactérias e até mesmo na disseminação de genes de resistência antimicrobiana.

Além de preencher uma lacuna significativa no conhecimento sobre a virosfera amazônica, este artigo contribui para o entendimento da ecologia viral em ambientes aquáticos tropicais, com implicações importantes para a saúde pública, a segurança hídrica e a vigilância ambiental.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fpubh.2022.869886

Viral Metagenomics Reveals Widely Diverse Viral Community of Freshwater Amazonian Lake

Wylerson Guimarães Nogueira[1†], Bruna Verônica Azevedo Gois[2†], Kenny da Costa Pinheiro[2], Andressa de Oliveira Aragão[3], Ana Lídia Cavalcante Queiroz[2], Artur Luiz da Silva[3], Adriana Carneiro Folador[3] and Rommel Thiago Jucá Ramos[2]*

[1] Department of Biochemistry and Immunology, Federal University of Minas Gerais, Belo Horizonte, Brazil, [2] Institute of Biological Sciences, Federal University of Pará, Belém, Brazil, [3] Laboratory of Genomic and Bioinformatics, Center of Genomics and System Biology, Federal University of Pará, Belém, Brazil

OPEN ACCESS

Edited by:
Maria Ines Zanoli Sato, Companhia Ambiental do Estado de São Paulo (CETESB), Brazil

Reviewed by:
Niti B. Jadeja, Ashoka Trust for Research in Ecology and the Environment (ATREE), India
Mahmuda Yasmin, University of Dhaka, Bangladesh

***Correspondence:**
Rommel Thiago Jucá Ramos
rommelthiago@gmail.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Environmental health and Exposome, a section of the journal Frontiers in Public Health

Received: 05 February 2022
Accepted: 24 March 2022
Published: 25 April 2022

Citation:
Nogueira WG, Gois BVA, Pinheiro KdC, Aragão AdO, Queiroz ALC, Silva ALd, Folador AC and Ramos RTJ (2022) Viral Metagenomics Reveals Widely Diverse Viral Community of Freshwater Amazonian Lake. Front. Public Health 10:869886. doi: 10.3389/fpubh.2022.869886

Despite the importance of understanding the ecology of freshwater viruses, there are not many studies on the subject compared to marine viruses. The microbiological interactions in these environments are still poorly known, especially between bacteriophages and their host bacteria and between cyanophages and cyanobacteria. Lake Bologna, Belém, capital of the Brazilian State of Pará, is a water source that supplies the city and its metropolitan region. However, it remains unexplored regarding the contents of its virome and viral diversity composition. Therefore, this work aims to explore the taxonomic diversity of DNA viruses in this lake, especially bacteriophages and cyanophages, since they can act as transducers of resistance genes and reporters of water quality for human consumption. We used metagenomic sequencing data generated by previous studies. We analyzed it at the taxonomic level using the tools Kraken2, Bracken, and Pavian; later, the data was assembled using Genome Detective, which performs the assembly of viruses. The results observed here suggest the existence of a widely diverse viral community and established microbial phage-regulated dynamics in Lake Bolonha. This work is the first ever to describe the virome of Lake Bolonha using a metagenomic approach based on high-throughput sequencing, as it contributes to the understanding of water-related public health concerns regarding the spreading of antibiotic resistance genes and population control of native bacteria and cyanobacteria.

Keywords: Amazon, Lake Bolonha, freshwater, viruses, bacteriophages, cyanophages, virome, metagenomics

INTRODUCTION

Amazonia is a 10 million years old unrivaled nest of biodiversity that reigns over South America; from bird-eating spiders to emperor tamarins to pink river dolphins, biologists find a new species every other day (1). Its ecosystems are essential for biodiversity preservation, climate regulation, energy production, and food and water security. The Amazon has a vital role in controlling zoonotic diseases and vector-borne and water-borne infections (2, 3).

Despite its importance, policies, laws, agreements, funds, and practical actions focused on Amazon protection have been weakened in Brazil, encouraging deforestation and culminating in losing about 20% of the original Amazon forest cover in Brazil by 2019 (4). The association between anthropogenic action in the Amazon rainforest, eutrophication of its water bodies, climate change,

and alterations in vector dynamics, human migration, genetic changes in pathogens, and the poor social and environmental conditions in many Latin-American countries serve as an opportunity for the emergence and re-emergence of human infectious diseases in Brazil and other Amazonian countries (5).

Amazonian fauna hosts a vast diversity of well-known pathogens and many other potential new or even unknown pathogens (6–11). This abundance of microorganisms indicates that the emergence of new infections from the forest is a constant threat to human health, particularly favorable to water-borne diseases due to anthropogenic activities nearby freshwater sources, such as ponds, rivers, basins, and lakes (12–14).

Lakes close to urban areas are increasingly changing their ecosystem as human population expansion occurs and commercial, recreational, and residential uses increase (15). The eutrophication process offers particular conditions for the replication of viruses, as environments of this type seem to provide high viral activation and hypothetically control host abundance, respiration, and production (16).

Regarding lake environments, little is known about the ecology of the freshwater virus when compared to marine viruses (17). Most are bacteriophages or human and other animal viruses, but plant viruses are also identified (18). Bacteriophages can play an essential function in the aquatic ecosystem as they can contribute to the acquisition and spread of antibiotic resistance genes (ARGs) (19). Some studies have shown ARGs-carrying phages to be abundant in many environments, especially those impacted by anthropogenic activities (20–24), which demonstrates that these types of viruses are relevant to local microbial ecology (20).

Like bacteriophages, another essential group of DNA viruses is cyanophages, which infects cyanobacteria and have a similar morphology (25). Being abundant in both fresh and saltwater, they play an essential role in modulating cyanobacterial populations and preserving water quality (25, 26). Also, cyanophages are abundant in aquatic environments and play a fundamental role in flowering dynamics, including growth regulation and photosynthesis of cyanobacteria (27). However, unlike bacteriophages, it has many genera of possible hosts; therefore, freshwater cyanophages can be classified according to the taxonomy of their host organisms (28).

Given the need to study the diversity of different environments, techniques have been developed, such as viral metagenomics, also known as virome. This technique allows the study of various viruses from environmental samples (29). In this way, metagenomics and next-generation sequencing (NGS) have demonstrated considerable genetic complexity and inter-species and intra-species interaction by exploring viral populations both in aquatic environments and within the human microbiome (30, 31). However, despite increasing studies using the technique, there are still significant gaps in the virome databases. It has been estimated that 1,031 viral particles are infecting bacterial populations. However, <2,200 double-chain DNA virus (dsDNA) and retrovirus genomes are deposited at the National Center for Biotechnology Information (NCBI), compared to more than 45,000 bacterial genomes (32).

In order to understand the relationships that may exist between human actions and the emergence of new diseases from water sources in the Amazon, it is crucial to comprehend the role and dynamics displayed by the present viruses inside the local community. Therefore, our objective was to identify and describe the diversity of DNA viruses through viral metagenomics analyses in Lake Bolonha, especially those that have bacteriophagic and cyanophagic behavior, thus contributing to the future handling of water-borne diseases associated with resistance to antimicrobials of public health concern.

MATERIALS AND METHODS

Sample Collection

All sequencing data employed in this study were generated previously by Alves et al. (33). The water samples were collected in January of 2017 at Lake Bolonha, Belém-PA, at three different points, namely: P1 (S 01°25.530″ W 048°26.043″), upstream of the Water Treatment Plant uptake; P2 (S 01°25.530″ W 048°26.018″) in the morning-glory spillway that supplies other water treatment substations, and P3 (S 01°24.992″ W 048°25.785″) in the channel connecting the lakes Água Preta and Bolonha (**Figure 1**). The assessment of water quality, DNA extraction, and Total Community DNA (TC-DNA) metagenomics sequencing by the Ion Proton^{TM} platform are described in Alves et al. (33).

Metagenomic Dataset

The raw metagenomic data used in this work is currently available at SRA/NCBI database under the access numbers SRR8893560 (P1), SRR8893561 (P2), and SRR8893559 (P3) (**Table 1**). This data can also be found at

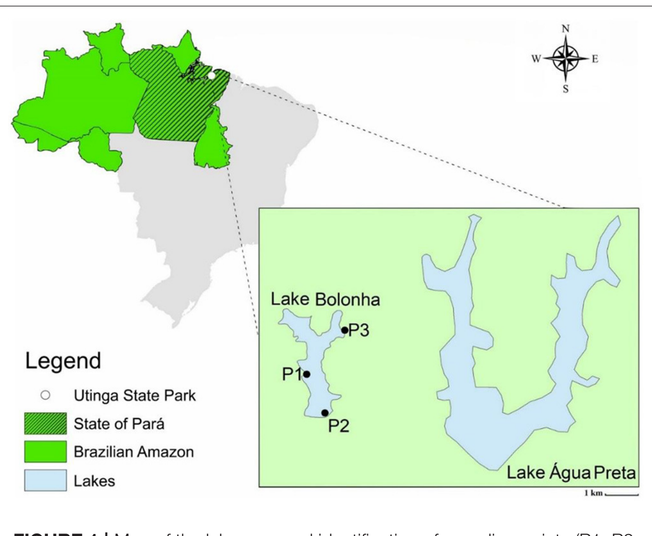


FIGURE 1 | Map of the lake area and identification of sampling points (P1, P2, and P3). Alves et al. (33).

Abbreviations: ARGs, Antibiotic resistance genes; NGS, Next-Generation Sequencing; dsDNA, double-chain DNA; NCBI, National Center for Biotechnology Information.

TABLE 1 | General information about the metagenomic dataset used in this work.

SRA accession number	Sample code	Number of spots	Number of bases	Size	Study	Sequencing technology
SRR8893560	P1	16,671,734	2.3 Gb	1.7 Gb	Freshwater Metagenome	Ion Proton^{TM}
SRR8893561	P2	16,278,151	2.2 Gb	1.7 Gb	Freshwater Metagenome	Ion Proton^{TM}
SRR8893559	P3	12,236,522	1.7 Gb	1.3 Gb	Freshwater Metagenome	Ion Proton^{TM}

TABLE 2 | List of samples (P1, P2, and P3) and their respective number of reads obtained after the sequencing, after quality control, the number of reads associated with viruses, and the number of reads used for coverage depth.

Sample code	Raw data	After control quality	Reads associated with viruses	Average depth of assembly
P1	16,671,734 reads	8,410,372 reads	202,318 reads	5,359 reads
P2	16,278,151 reads	8,339,898 reads	44,877 reads	4,067 reads
P3	12,236,522 reads	6,299,043 reads	131,370 reads	2,217 reads

www.ncbi.nlm.nih.gov/sra/PRJNA506429. It is important to note that the data employed here was generated through high-throughput metagenomic sequencing of environmental samples using the Total Community DNA approach. However, we only included in our analyses the viral portion of the metagenomic dataset for this work being the virome our aimed subject of study.

Taxonomy Analysis

The raw data was used to perform the taxonomic analysis through the Kraken2 tool (34) with the parameter "–download-library viral" to download the complete viral sequences of RefSeq and classify the reads regards its taxonomy.

Assessment of Viral Diversity

The output generated by Kraken2 was submitted to the tool Bracken (35), using abundance and diversity to generate more accurate estimations on the viruses genus and species levels. The input parameters were "${CLASSIFICATION_LVL} = 'S' (Species)" and "input data = kraken2 output (report)." All other parameters were set as default. Later, the results were displayed with the Pavian tool (36), which allows comparing the taxonomic classifications obtained by Kraken2 and Bracken and presenting abundance estimations in several samples. Veen diagrams were generated using the web-based tool InteractiVenn (37).

Viral Metagenome Assembly

The online tool Genome Detective (38) was used in default parameters to assemble the sequencing data and classify the contigs formed into their respective taxa, identified using metaSPAdes software for single-end reads (39).

RESULTS

Virome Assembly

The viral portion of the raw metagenomic data was assembled to prepare the data for the classification of their respective taxon per each of the freshwater samples (**Table 2**). The species with the highest percentage of coverage after assembly were Cyanophage KBS-S-2A (P1) with 25.37% and 69.66% identity, *Cladosporium fulvum* T-1 virus (P2) with 23.14% coverage and 56.90% identity, and Cyanophage KBS-S-2A (P3) with 20.65% coverage and 71.18% identity.

Viral Diversity of Lake Bolonha

Taxonomic analysis performed by Kraken2 revealed that there might be more than 3.500 distinct species of viruses in Lake Bolonha. A Sankey diagram is represented summarizing taxonomic diversity at samples P1, P2, and P3, respectively, in **Figure 2**. The complete data from the taxonomic analysis comparison between each sample site is reported as **Supplementary Tables S1–S3**.

The 20 species with a higher reads count in all freshwater samples are represented in **Figures 3–5**. The number of reads found for each virus on this top selection was compared with their respective number of reads on the other two samples from this study. Also, a Venn diagram showcasing the overlap between these top 20 most represented viruses for samples P1, P2, and P3, is displayed in **Figure 6**. A similar diagram for the complete dataset of the identified viruses is reported as **Supplementary Figure S1**.

The sample P1 presented an overall abundance of *Synechococcus* phage, a cyanophage (**Figure 3**). We obtained the highest reads for a single virus in sample P2 for *Choristoneura fumiferana granulovirus* (**Figure 4**). The sample P2 also presented a much more significant amount of the genus Pandoravirus and Mimivirus compared to other collection sites. *Haemophilus* phage HP1 was only observed on sample P3 (**Figure 5**).

DISCUSSION

Aquatic ecological environments have a broad range of viruses that are crucial in controlling bacterial communities and regulating biogeochemical cycles (40, 41). While most of the literature on water virome has concentrated on viruses in marine waters (42, 43), previous research indicated that freshwater harbored specific viral communities distinct from other aquatic environments (44). Other studies have described the characterization of freshwater viromes in ballast water (45), sewage (46, 47), lakes (48, 49), river estuaries, where marine

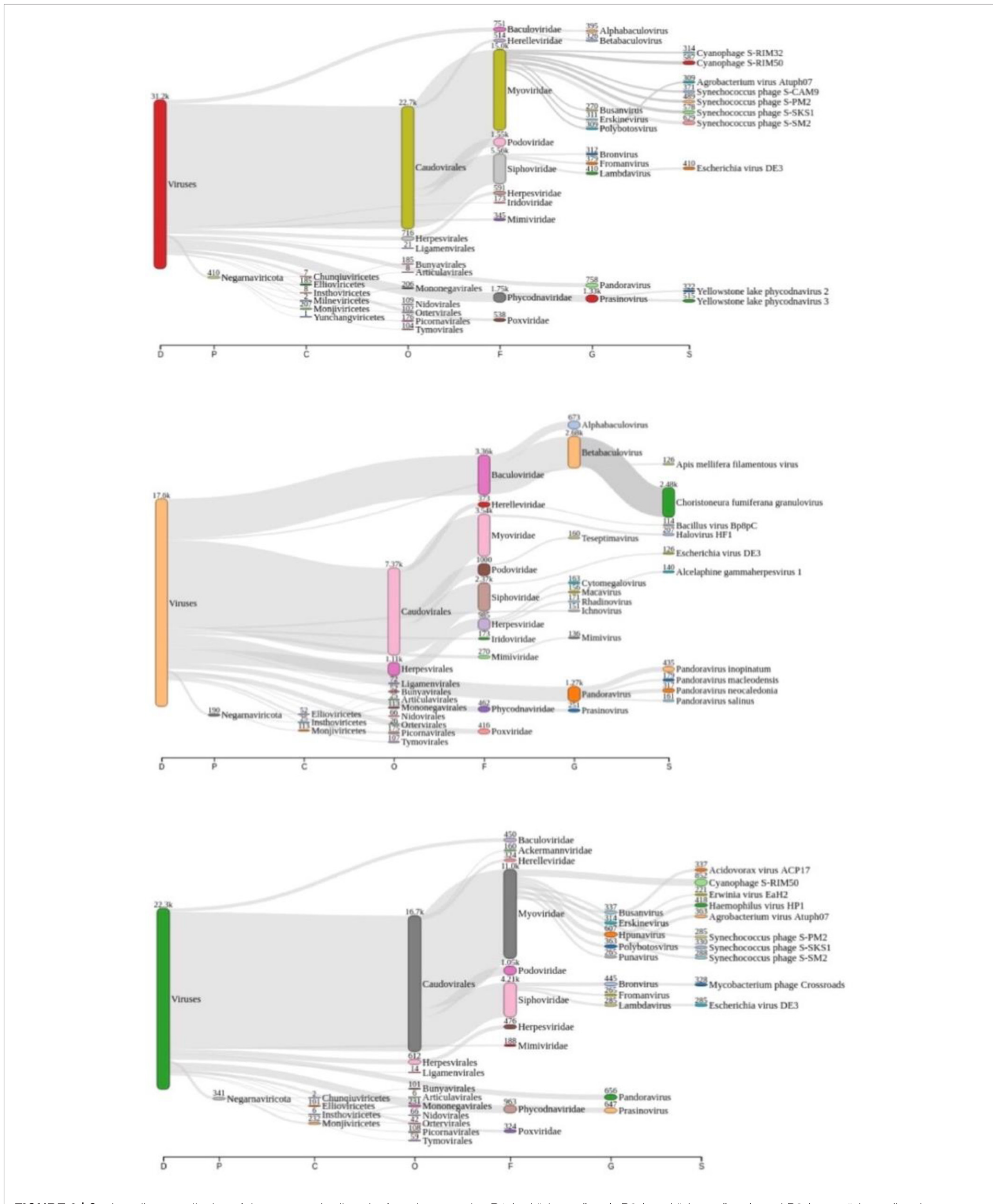


FIGURE 2 | Sankey diagram display of the taxonomic diversity found at samples P1 (red "viruses" tag), P2 (coral "viruses" tag), and P3 (green "viruses" tag), respectively.

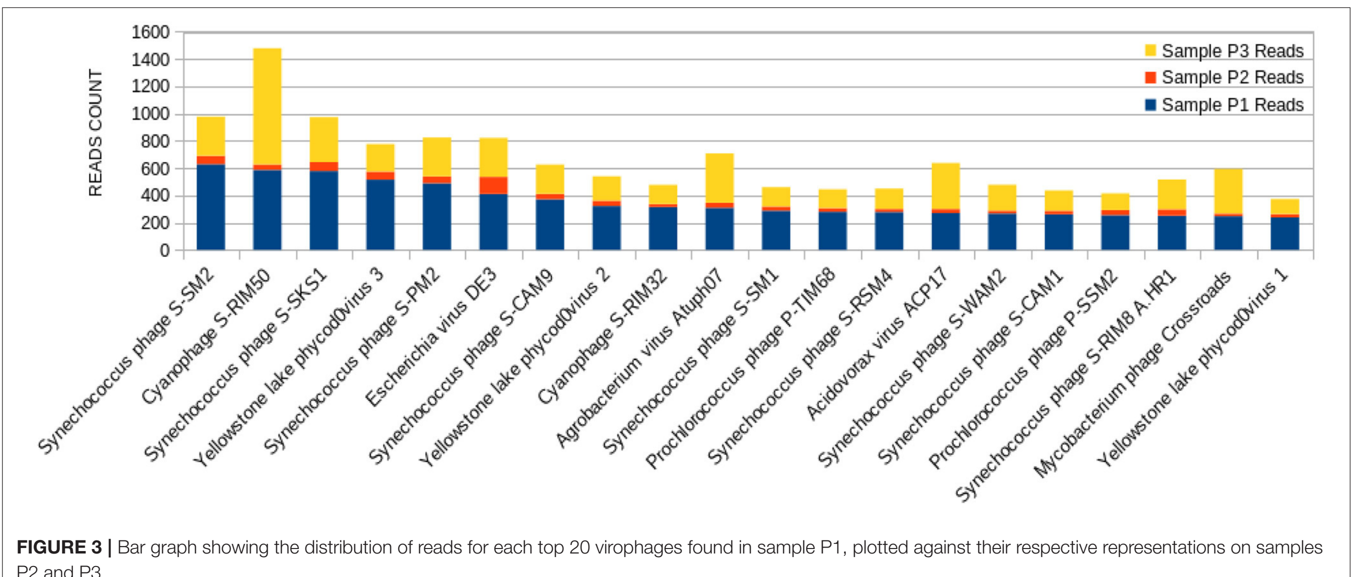


FIGURE 3 | Bar graph showing the distribution of reads for each top 20 virophages found in sample P1, plotted against their respective representations on samples P2 and P3.

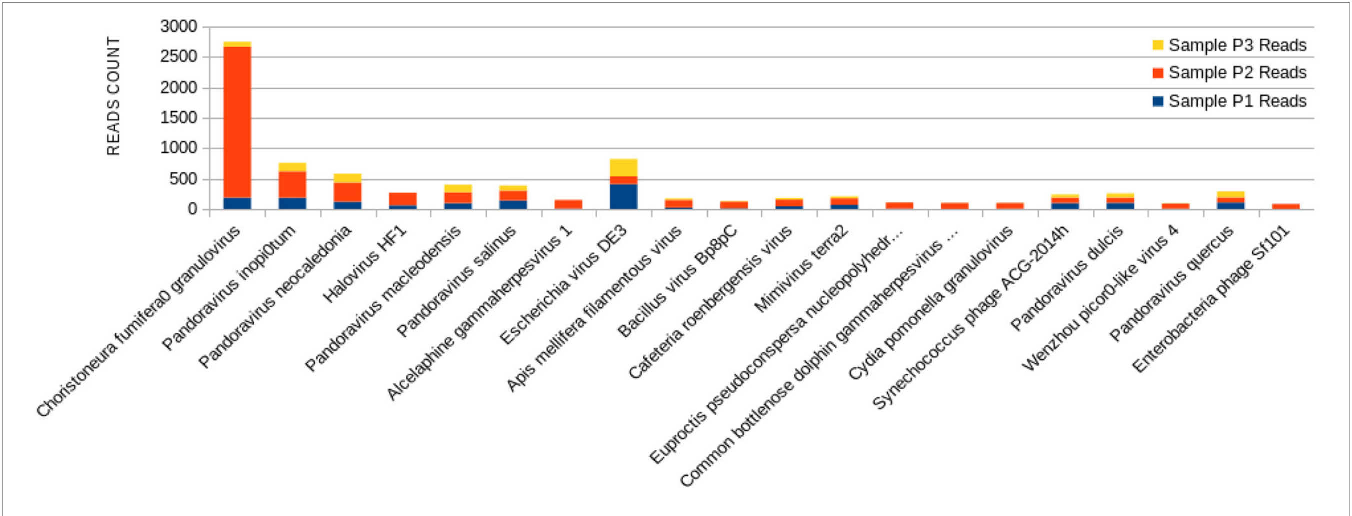


FIGURE 4 | Bar graph showing the distribution of reads for each top 20 virophages found in sample P2, plotted against their respective representations on samples P1 and P3.

and freshwater mix (50–52), and long rivers (53). In the present study, we identified the diversity of the viral community for three different metagenomic samples from the Lake Bolonha, obtained by the previous work of Alves et al. (33), to acquire a vivid understanding of this Amazonian lake water system.

According to Alves et al. (33), the Amazonian vegetation on its shore characterizes Lake Bolonha and the propagation of large plants under its surface, resulting in eutrophication. Also, it presents increased phosphorus and total nitrogen values in physical-chemical analysis and a high fecal coliform rate (33). In this study, we noted the significant presence of cyanophages, mainly at P1 and P3 (**Figure 2**). This finding contributes as an essential indicator of the interference of these phages in the environment, known for their ability to perform photosynthesis by consuming oxygen and their potential for binding to nitrogen and producing toxins (54). Nitrogen or phosphorus supplies and reduced growth rate and biomass may naturally limit freshwater ecosystems, including those involving cyanobacteria accumulation (55).

When it comes to viral abundance, taxonomic analysis reveals that a plethora of more than 3.500 distinct viruses is present at freshwater Brazilian Lake Bolonha, an area of environmental preservation. It is possible to observe that the phages with a higher abundance of reads in P1 and P3 have a lower distribution among their P2 (**Figures 3–5**), which can be explained by the proximity between sampling sites P1 and P3. The alpha diversity observed in the raw sequencing data of **Table 1** of Alves et al. (33) also shows significant sample abundance in P1 and P3, which could also be related to this observation.

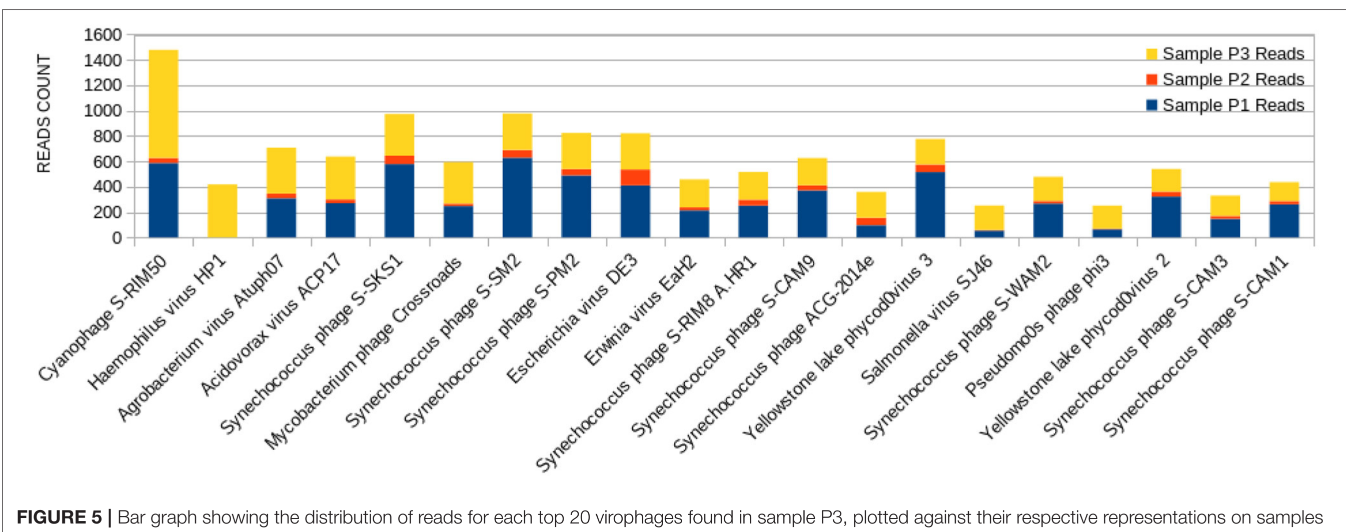


FIGURE 5 | Bar graph showing the distribution of reads for each top 20 virophages found in sample P3, plotted against their respective representations on samples P1 and P2.

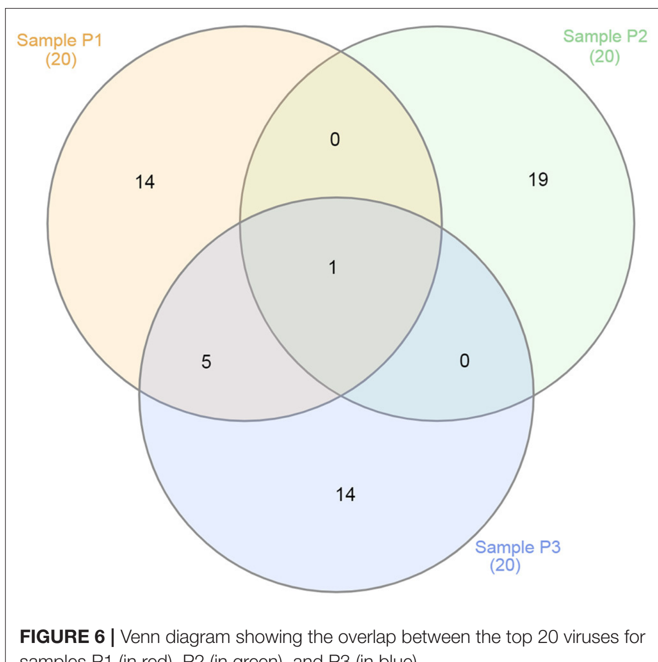


FIGURE 6 | Venn diagram showing the overlap between the top 20 viruses for samples P1 (in red), P2 (in green), and P3 (in blue).

The diversity and abundance imbalance of essential viruses, such as bacteriophages and cyanophages, can cause significant changes in the aquatic ecosystem. Some of these phages can mediate the transduction of resistance genes between bacteria, which can provide evolutionary advantages to microorganisms and affect the water quality (33). In addition, understanding this viral community's diversity could help prevent the spread of antimicrobial resistance elements and circumvent possible future multi-resistant pathogen epidemics.

Although all samples came from the same lake and environment, the overlap of viruses best represented at each sampling site showed a vast distinction when considering its top 20 viruses (**Figure 6**). For instance, only Escherichia phage DE3 was at the top 20 viruses for all three samples. At the top, only shared by samples P1 and P3 were five viruses: *Mycobacterium* phage Crossroads; Yellowstone lake phycodnavirus 3; *Agrobacterium* phage Atuph07; Yellowstone lake phycodnavirus 2; and Acidovorax virus ACP17. Moreover, sample P2 had no virus shared exclusively with sample P1 or P3 at its top 20. That observation shows the diversity of viral entities present in a single environment, such as Lake Bolonha. It highlights choosing different and representative sites when studying an environmental microbiome.

The sample P1 presented an abundance of *Synechococcus* phage (**Figure 3**). This phage frequently infects cyanobacteria of the *Synechococcus* genus in diurnal patterns of infection due to the photosynthetic activities of its host (56). This infection affects their population dynamics by killing part of this cyanobacteria population daily, estimated between 0.005 and 30% per day (27, 57, 58). In addition, it has been described that such cyanophages play a vital role in the diversity and evolution of their host cyanobacteria (18, 32, 54, 59–62).

Amongst the abundant species, the cyanophage S-RIM50 was found in P1, which could infect cyanobacteria of the genus *Synechococcus* (63). The abundance of cyanophage S-RIM50 has been reported in both fresh and seawater (26). Cyanophages such as the cyanophage S-RIM50 and *Synechococcus* phage are abundant in freshwater environments. They have been isolated from various freshwater reserves, including lakes, ponds, streams, and sewage points (27, 56, 58). They have an essential contribution to maintaining the cyanobacterial community and the preservation of water quality (25).

Synechococcus phages, present in both samples P1 (**Figure 3**) and P3 (**Figure 5**), can be associated with health problems such as multiple sclerosis. This phage expresses proteins containing consensus peptide stretches that are highly homologous to the products of 16 autoantigens related to multiple sclerosis susceptibility genes (64). Other viruses associated with multiple sclerosis, such as the Epstein-Barr virus, have also shown

this behavior, and the bacteriophage *Synechococcus* has been identified as a new relevant contributor to this phenomenon. Its cyanobacterial host prefers a temperate climate, indicating that the ecology of this cyanophage is consistent with the overall distribution and epidemiology of multiple sclerosis (63).

Yellowstone Lake phycodnaviruses, a double-stranded DNA virus that infects algae, and *Escherichia* phage DE3 on the sample P1 were also observed (**Figure 3**). A more significant number of reads associated with *Shigella* phage SfIV was observed compared to samples P2 and P3 (**Figure 3**). It is important to mention a relationship with the possible environmental presence of the *Shigella* bacterial host, responsible for causing intestinal infection followed or not by fever, colic, and diarrhea with blood and mucus (65). This observation demonstrates the importance of this study in characterizing the environmental conditions as a possible source of information for public health.

The P2 collection point presented many *Choristoneura fumiferana granulovirus* (**Figure 4**), part of the *Baculoviridae* family (66). *Halovirus* HF1, responsible for infecting members of the *Halobacteriaceae* family (67), was also identified (**Figure 4**). Interestingly, a small number of reads (25 reads) associated with the species *Diplodia scrobiculate* RNA virus 1 is only present in P2. As a preferable host, this phage has the endophytic fungus *Diplodia scrobiculate*, which primarily affects the genus *Pinus* spp. among other conifers (68), which would be odd to find on a water sample from a tropical locality.

A considerable amount of *Bacillus* phage Bp8pC has also been observed on sample P2 (**Figure 4**), which hosts the bacteria *Bacillus thuringiensis* and *Bacillus pumilus*. Both are of economic importance because they are used in agriculture as pest control bringing little harm to humans (69). The presence of the genus *Bacillus* in lake water may indicate the contamination of the water environment by different types of residues coming from the watershed to the lake (70).

Exclusively on P2, a much more significant amount of the genus Pandoravirus and Mimivirus was observed compared to the other collection points (**Figure 4**). Previous studies suggest a potential role of Mimivirus in respiratory pathology displayed during seroconversion in patients with pulmonary pneumonia. In addition, positive serology for Mimivirus is associated with increased duration of mechanical ventilation supported breathing and intensive care unit in patients with ventilator-associated pneumonia (71). Both genera are constituted of giant viruses and have Amoeba as their typical host (72–74). Pandoravirus has a size of about 1 micron and may resemble some types of bacteria. Their genome contains more than 100 distinct genes and can be twice as large as the Mimivirus genome, besides the fact that their genome is quite different compared to other known organisms (75).

Haemophilus virus HP1, a bacteriophage that infects the *Haemophilus influenzae* bacterium (76), was only observed at sample P3 (**Figure 5**). Its sampling site is located at the starting point of the channel connecting both Lakes Bolonha and Água Preta. It is crucial to note that what occurs at this site may in the future influence the environment of nearby Lake Água Preta.

Overall, these results denote the presence of a diverse viral community and suggest the existence of established regulation dynamics in the local microbial environment of Lake Bolonha, highly influenced by the bacteriophages and cyanophages that inhabit the location. The dispersion of those biological entities along the water distribution channels using Lake Bolonha as a water source and general eutrophic activity might contribute to the spread of minor genetic elements like ARGs and future unbalance the microenvironment of close by freshwater sources, such as Lake Água Preta.

CONCLUSIONS

Given the importance of Lake Bolonha as a source of drinking water supply for the metropolitan region of Belém, the elucidation of the viral diversity from this environment is relevant to provide a better understanding of how its exploration can affect it. The results observed in this work indicate a widely diverse viral community, especially bacteriophages and cyanophages. These findings also suggest the existence of established micro-environmental dynamics in Lake Bolonha, possibly regulated by such phage entities. The dispersion of those viral beings bare similarity along the course of the lake, apparently more related the deeper they are into the lake (P1, P3) and the further away they are from the water evacuation sites to other treatment substations (P2). This study is the first-ever work to describe the virome of Lake Bolonha and, as such, contributes to the understanding of water-related public health concerns regarding the spreading of antibiotic resistance genes and population control of native bacteria and cyanobacteria.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/; https://www.ncbi.nlm.nih.gov/sra/PRJNA506429.

AUTHOR CONTRIBUTIONS

BG, KP, and RR designed the study. WN, BG, and KP compiled and curated the data, performed bioinformatic analysis, and interpreted the results. AF and RR supervised and administered the project and provided funding. WN and BG wrote the original draft and manuscript with input from KP, AA, AQ, AF, and RR. All authors critically reviewed the manuscript and approved the final version.

FUNDING

We thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–CAPES and PROPESP/UFPA (Pró-Reitoria de Pesquisa e Pós-Graduação / Universidade Federal do Pará), the funding agencies FAPESPA (Fundação Amazônia de Amparo à Estudos e Pesquisas), and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) for financial support on this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.869886/full#supplementary-material

Supplementary Figure 1 | Venn diagram representation of overlapping viruses for Samples P1, P2, and P3 complete datasets.

Supplementary Table 1 | Full report table of identified viruses from taxonomic analysis performed by Kraken2 for Lake Bolonha's metagenome Sample P1.

Supplementary Table 2 | Full report table of identified viruses from taxonomic analysis performed by Kraken2 for Lake Bolonha's metagenome Sample P2.

Supplementary Table 3 | Full report table of identified viruses from taxonomic analysis performed by Kraken2 for Lake Bolonha's metagenome Sample P3.

REFERENCES

1. Sandy M. *Why Is the Amazon Rain Forest Disappearing?* Time (2019).
2. Baker JCA, Spracklen DV. Climate benefits of intact amazon forests and the biophysical consequences of disturbance, front. *For Glob Chang.* (2019) 2:47. doi: 10.3389/ffgc.2019.00047
3. Metzger JP, Bustamante MMC, Ferreira J, Fernandes GW, Librán-Embid F, Pillar VD, et al. Why brazil needs its legal reserves, perspect. *Ecol Conserv.* (2019) 17:91–103. doi: 10.1016/j.pecon.2019.07.002
4. *INPE PRODES — Coordenação-Geral de Observação da Terra, Inst. Nac. Pesqui. Espac.* (2020). p. 1. Available online at: http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes (accessed January 27, 2022).
5. Ellwanger JH, Kulmann-Leal B, Kaminski VL, Valverde-Villegas JM, VEIGA ABGDA, Spilki FR, et al. Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. *An Acad Bras Cienc.* (2020) 92:20191375. doi: 10.1590/0001-3765202020191375
6. de Cássio Veloso de Barros B, Chagas EN, Bezerra LW, Ribeiro LG, Barboza Duarte JW, Pereira D, et al. Rotavirus A in wild and domestic animals from areas with environmental degradation in the Brazilian Amazon. *PLoS ONE.* (2018) 13:e0209005. doi: 10.1371/journal.pone.0209005
7. Silva MBda, Portela JM, Li W, Jackson M, Gonzalez-Juarrero M, Hidalgo AS, et al. Evidence of zoonotic leprosy in Pará, Brazilian Amazon, and risks associated with human contact or consumption of armadillos. *PLoS Negl Trop Dis.* (2018) 12:e0006532. doi: 10.1371/journal.pntd.0006532
8. Farikoski IO, Medeiros LS, Carvalho YK, Ashford DA, Figueiredo EES, Fernandes DVGS, et al. The urban and rural capybaras (Hydrochoerus hydrochaeris) as reservoir of Salmonella in the western Amazon, Brazil. *Pesqui Vet Bras.* (2019) 39:66–9. doi: 10.1590/1678-5150-pvb-5761
9. Franco Filho LC, Barata RR, Cardoso JF, de V. Massafra JM, da S. Lemos P, Casseb LMN, et al. Metagenomic analysis of samples from three bat species collected in the amazon rain forest. *Microbiol Resour Announc.* (2019) 8:e01422–18. doi: 10.1128/MRA.01422-18
10. Fernandes J, Guterres A, De Oliveira RC, Chamberlain J, Lewandowski K, Teixeira BR, et al. Xapuri virus, a novel mammarenavirus: Natural reassortment and increased diversity between New World viruses. *Emerg Microbes Infect.* (2018) 7:120. doi: 10.1038/s41426-018-0119-9
11. Medeiros DBA, Vasconcelos PFC. Is the brazilian diverse environment is a crib for the emergence and maintenance of exotic arboviruses? *An Acad Bras Cienc.* (2019) 91:e20190407. doi: 10.1590/0001-3765201920190407
12. Confalonieri UEC, Margonari C, Quintão AF. Environmental change and the dynamics of parasitic diseases in the Amazon. *Acta Trop.* (2014) 129:33–41. doi: 10.1016/j.actatropica.2013.09.013
13. Fonseca PAM, Hacon Sde S, Reis VL, Costa D, Brown IF. O uso de dados de satélite para estudar a relação entre chuva e doenças diarreicas em uma bacia na amazônia sul-ocidental. *Cienc Saude Coletiva.* (2016) 21:731–42. doi: 10.1590/1413-81232015213.20162015
14. Levy K, Smith SM, Carlton EJ. Climate change impacts on waterborne diseases: moving toward designing interventions. *Curr Environ Heal Reports.* (2018) 5:272–82. doi: 10.1007/s40572-018-0199-7
15. Putt AE, MacIsaac EA, Herunter HE, Cooper AB, Selbie DT. Eutrophication forcings on a peri-urban lake ecosystem: context for integrated watershed to airshed management. *PLoS ONE.* (2019) 14:e0219241. doi: 10.1371/journal.pone.0219241
16. Cook A, Rushton S, Allan JA, Baxter A. An evaluation of techniques to control problem bird species on landfill sites. *Environ Manage.* (2008) 41:834–43. doi: 10.1007/s00267-008-9077-7
17. Palermo CN, Fulthorpe RR, Saati R, Short SM. Metagenomic analysis of virus diversity and relative abundance in a eutrophic freshwater harbour. *Viruses.* (2019) 11:792. doi: 10.3390/v11090792
18. Mohiuddin M, Schellhorn HE. Spatial and temporal dynamics of virus occurrence in two freshwater lakes captured through metagenomic analysis. *Front Microbiol.* (2015) 6:960. doi: 10.3389/fmicb.2015.00960
19. Koskella B, Brockhurst MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev.* (2014) 38:916–31. doi: 10.1111/1574-6976.12072
20. Calero-Cáceres W, Melgarejo A, Colomer-Lluch M, Stoll C, Lucena F, Jofre J, et al. Sludge as a potential important source of antibiotic resistance genes in both the bacterial and bacteriophage fractions. *Environ Sci Technol.* (2014) 48:7602–11. doi: 10.1021/es501851s
21. Colomer-Lluch M, Calero-Cáceres W, Jebri S, Hmaied F, Muniesa M, Jofre J. Antibiotic resistance genes in bacterial and bacteriophage fractions of Tunisian and Spanish wastewaters as markers to compare the antibiotic resistance patterns in each population. *Environ Int.* (2014) 73:167–75. doi: 10.1016/j.envint.2014.07.003
22. Lekunberri I, Villagrasa M, Balcázar JL, Borrego CM. Contribution of bacteriophage and plasmid DNA to the mobilization of antibiotic resistance genes in a river receiving treated wastewater discharges. *Sci Total Environ.* (2017) 601–2:206–9. doi: 10.1016/j.scitotenv.2017.05.174
23. Marti E, Variatza E, Balcázar JL. Bacteriophages as a reservoir of extended-spectrum β-lactamase and fluoroquinolone resistance genes in the environment. *Clin Microbiol Infect.* (2014) 20:0456–9. doi: 10.1111/1469-0691.12446
24. Subirats J, Sànchez-Melsió A, Borrego CM, Balcázar JL, Simonet P. Metagenomic analysis reveals that bacteriophages are reservoirs of antibiotic resistance genes. *Int J Antimicrob Agents.* (2016) 48:163–7. doi: 10.1016/j.ijantimicag.2016.04.028
25. Xia H, Li T, Deng F, Hu Z. Freshwater cyanophages. *Virol Sin.* (2013) 28:253–9. doi: 10.1007/s12250-013-3370-1
26. Shane MS. Distribution of blue-green algal viruses in various types of natural waters. *Water Res.* (1971) 5:711–6. doi: 10.1016/0043-1354(71)90093-5
27. Jiang X, Ha C, Lee S, Kwon J, Cho H, Gorham T, et al. Characterization of cyanophages in lake erie: Interaction mechanisms and structural damage of toxic cyanobacteria. *Toxins.* (2019) 11:444. doi: 10.3390/toxins11080444
28. Singh P. Structural, functional and molecular basis of cyanophage-cyanobacterial interactions and its significance. *AFRICAN J Biotechnol.* (2012) 11:2591–608. doi: 10.5897/AJB10.790
29. Wommack KE, Nasko DJ, Chopyk J, Sakowski EG. Counts and sequences, observations that continue to change our understanding of viruses in nature. *J Microbiol.* (2015) 53:181–92. doi: 10.1007/s12275-015-5068-6
30. Aggarwala V, Liang G, Bushman FD. Viral communities of the human gut: Metagenomic analysis of composition and dynamics. *Mob DNA.* (2017) 8:1–10. doi: 10.1186/s13100-017-0095-y
31. Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, et al. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat Commun.* (2017) 8:1–12. doi: 10.1038/s41467-016-0009-6
32. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. *Nature.* (2016) 536:425–30. doi: 10.1038/nature19094
33. Alves J, Dias L, Mateus J, Marques J, Graças D, Ramos R, et al. Resistome in lake bolonha, brazilian amazon: identification of genes related to resistance to broad-spectrum antibiotics. *Front Microbiol.* (2020) 11:67. doi: 10.3389/fmicb.2020.00067

34. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. (2019) 20:257. doi: 10.1186/s13059-019-1891-0
35. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci.* (2017) 2017:e104. doi: 10.7717/peerj-cs.104
36. Breitwieser FP, Salzberg SL. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics*. (2020) 36:1303–4. doi: 10.1093/bioinformatics/btz715
37. Heberle H, Meirelles VG, Silva FRda, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*. (2015) 16:1–7. doi: 10.1186/s12859-015-0611-3
38. Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, et al. Genome detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics*. (2019) 35:871–3. doi: 10.1093/bioinformatics/bty695
39. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.* (2017) 27:824–34. doi: 10.1101/gr.213959.116
40. Fischer UR, Velimirov B. High control of bacterial production by viruses in a eutrophic oxbow lake. *Aquat Microb Ecol.* (2002) 27:1–12. doi: 10.3354/ame027001
41. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat Rev Microbiol.* (2014) 12:519–28. doi: 10.1038/nrmicro3289
42. Suttle CA. Environmental microbiology: viral diversity on the global stage. *Nat Microbiol*. (2016) 1:16205. doi: 10.1038/nmicrobiol.2016.205
43. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell.* (2019) 177:1109–23.e14. doi: 10.1016/j.cell.2019.03.040
44. Roux S, Enault F, Robin A, Ravet V, Personnic S, Theil S, et al. Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS ONE*. (2012) 7:e33641. doi: 10.1371/journal.pone.0033641
45. Kim Y, Aw TG, Teal TK, Rose JB. Metagenomic Investigation of viral communities in ballast water. *Environ Sci Technol*. (2015) 49:8396–407. doi: 10.1021/acs.est.5b01633
46. Fernandez-Cassi X, Timoneda N, Martínez-Puchol S, Rusiñol M, Rodriguez-Manzano J, Figuerola N, et al. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. *Sci Total Environ.* (2018) 618:870–80. doi: 10.1016/j.scitotenv.2017.08.249
47. Martínez-Puchol S, Rusiñol M, Fernández-Cassi X, Timoneda N, Itarte M, Andrés C, et al. Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. *Sci Total Environ*. (2020) 713:136604. doi: 10.1016/j.scitotenv.2020.136604
48. López-Bueno, J. Tamames, D. Velázquez, A. Moya, A. Quesada, A. Alcamí. High diversity of the viral community from an Antarctic lake. *Science.* (2009) 326:858–61. doi: 10.1126/science.1179287
49. Ge X, Wu Y, Wang M, Wang J, Wu L, Yang X, et al. Viral metagenomics analysis of planktonic viruses in East Lake, Wuhan, China. *Virol Sin.* (2013) 28:280–90. doi: 10.1007/s12250-013-3365-y
50. Cai L, Zhang R, He Y, Feng X, Jiao N. Metagenomic analysis of Virioplankton of the subtropical Jiulong river estuary, China. *Viruses*. (2016) 8:35. doi: 10.3390/v8020035
51. Wolf YI, Silas S, Wang Y, Wu S, Bocek M, Kazlauskas D, et al. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat Microbiol.* (2020) 5:1262–70. doi: 10.1038/s41564-020-0755-4
52. Wu S, Zhou L, Zhou Y, Wang H, Xiao J, Yan S, et al. Diverse and unique viruses discovered in the surface water of the East China Sea. *BMC Genomics*. (2020) 21:441. doi: 10.1186/s12864-020-06861-y
53. Lu J, Yang S, Zhang X, Tang X, Zhang J, Wang X, et al. Metagenomic analysis of viral community in the yangtze river expands known eukaryotic and prokaryotic virus diversity in freshwater. *Virol Sin*. (2022) 37:60–9. doi: 10.1016/j.virs.2022.01.003
54. Jaskulska. J. Mankiewicz-Boczek Cyanophages specific to cyanobacteria from the genus Microcystis. *Ecohydrol Hydrobiol.* (2020) 20:83–90. doi: 10.1016/j.ecohyd.2019.06.001
55. Cheng K, Frenken T, Brussaard CPD, Van De Waal DB. Cyanophage propagation in the freshwater cyanobacterium phormidiumis constrained by phosphorus limitation and enhanced by elevated pCO2. *Front Microbiol.* (2019) 10:617. doi: 10.3389/fmicb.2019.00617
56. Liu R, Liu Y, Chen Y, Zhan Y, Zeng Q. Cyanobacterial viruses exhibit diurnal rhythms during infection. *Proc Natl Acad Sci U S A*. (2019) 116:14077–82. doi: 10.1073/pnas.1819689116
57. Bragg L, Tyson GW. Metagenomics using next-generation sequencing. *Methods Mol Biol*. (2014) 1096:183–201. doi: 10.1007/978-1-62703-712-9_15
58. Toyama D, Kishi LT, Santos-Júnior CD, Soares-Costa A, Souza de Oliveira TC, de Miranda FP, et al. Metagenomics analysis of microorganisms in freshwater lakes of the Amazon Basin. *Genome Announc.* (2016) 4:1440–56. doi: 10.1128/genomeA.01440-16
59. Handelsman J. Metagenomics: application of genomics to uncultured microorganisms, microbiol. *Mol Biol Rev*. (2004) 68:669–85. doi: 10.1128/MMBR.68.4.669-685.2004
60. Lorenz P, Eck J. Metagenomics and industrial applications. *Nat Rev Microbiol.* (2005) 3:510–6. doi: 10.1038/nrmicro1161
61. Tundisi JG. *Água no século 21: enfrentando a escassez* (2003).
62. Fedida A, Lindell D. Two synechococcus genes, two different effects on cyanophage infection. *Viruses*. (2017) 9:136. doi: 10.3390/v9060136
63. Stoddard LI, Martiny JBH, Marston MF. Selection and characterization of cyanophage resistance in marine synechococcus strains. *Appl Environ Microbiol*. (2007) 73:5516–22. doi: 10.1128/AEM.00356-07
64. Carter CJ. Epstein-Barr and other viral mimicry of autoantigens, myelin and vitamin D-related proteins and of EIF2B, the cause of vanishing white matter disease: massive mimicry of multiple sclerosis relevant proteins by the Synechococcus phage. *Immunopharmacol Immunotoxicol*. (2012) 34:21–35. doi: 10.3109/08923973.2011.572262
65. Martinez-Becerra FJ, Chen X, Dickenson NE, Choudhari SP, Harrison K, Clements JD, et al. Characterization of a novel fusion protein from IpaB and IpaD of Shigella spp and its potential as a Pan-Shigella vaccine. *Infect Immun.* (2013) 81:4470–7. doi: 10.1128/IAI.00859-13
66. Clem RJ, Passarelli AL. Baculoviruses: sophisticated pathogens of insects. *PLoS Pathog.* (2013) 9:e1003729. doi: 10.1371/journal.ppat.1003729
67. Ventosa. *Halophilic Microorganisms*. Berlin; Heidelberg: Springer (2004). doi: 10.1007/978-3-662-07656-9
68. Bihon W, Slippers B, Burgess T, Wingfield MJ, Wingfield BD. Diplodia scrobiculata found in the southern hemisphere. *For Pathol.* (2011) 41:175–81. doi: 10.1111/j.1439-0329.2010.00649.x
69. Stotzky G. Persistence and biological activity in soil of insecticidal proteins from bacillus thuringiensis and of bacterial DNA bound on clays and humic acids. *J Environ Qual.* (2000) 29:691–705. doi: 10.2134/jeq2000.00472425002900030003x
70. Shafi S, Kamili AN, Shah MA, Bandh SA, Dar R. Dynamics of bacterial class Bacilli in the deepest valley lake of Kashmir-the Manasbal Lake. *Microb Pathog.* (2017) 104:78–83. doi: 10.1016/j.micpath.2017.01.018
71. Vanspauwen MJ, Franssen FME, Raoult D, Wouters EFM, Bruggeman CA, Linssen CFM. Infections with mimivirus in patients with chronic obstructive pulmonary disease. *Respir Med.* (2012) 106:1690–4. doi: 10.1016/j.rmed.2012.08.019
72. Legendre M, Fabre E, Poirot O, Jeudy S, Lartigue A, Alempic JM, et al. Diversity and evolution of the emerging pandoraviridae family. *Nat Commun.* (2018) 9:1–12. doi: 10.1038/s41467-017-02088-w
73. La Scola B, Audic S, Robert C, Jungang L, De Lamballerie X, Drancourt M, et al. A giant virus in amoebae. *Science.* (2003) 299:2033. doi: 10.1126/science.1081867
74. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of mimivirus. *Science*. (2004) 306:1344–50. doi: 10.1126/science.1101485
75. Abergel C, Legendre M, Claverie JM. The rapidly expanding universe of giant viruses: mimivirus, pandoravirus, pithovirus and mollivirus. *FEMS Microbiol Rev*. (2015) 39:779–96. doi: 10.1093/femsre/fuv037

76. Hakimi JM, Scocca JJ. Purification and characterization of the integrase from the Haemophilus influenzae bacteriophage HP1; identification of a four-stranded intermediate and the order of strand exchange. *Mol Microbiol*. (1996) 21:147–58. doi: 10.1046/j.1365-2958.1996.6311351.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Nogueira, Gois, Pinheiro, Aragão, Queiroz, Silva, Folador and Ramos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

5. ARTIGO II – DESCOBERTA *IN SILICO* DE FAGOS DE ÁGUA DOCE

O segundo artigo desta tese, indexado em periódico científico, é intitulado “Tools and methodology to *in silico* phage discovery in freshwater environments”, publicado junto à revista *Frontiers in Microbiology* (ISSN: 1664-302X), sob a seção de *Phage Biology*, em maio de 2024, fator de impacto: 4.076 (2024). Este artigo (https://doi.org/10.3389/fmicb.2024.1390726) foi publicado em condição de primeira autoria compartilhada, juntamente com os M.Sc. Carlos W. D. Dantas e M.Sc. David T. Martins.

Esta publicação apresenta uma revisão crítica sobre as principais ferramentas computacionais aplicadas à identificação de vírus bacteriófagos em metagenomas ambientais, com ênfase em ambientes de água doce. O artigo aborda o atual panorama metodológico da descoberta de fagos *in silico*, destacando as principais estratégias disponíveis para detecção, anotação, análise de qualidade, taxonomia, predição de hospedeiros e estudos de diversidade viral. Além disso, discute os desafios técnicos ainda enfrentados na área, como a ausência de padronização de pipelines e a limitação de dados sobre interações vírus-hospedeiro.

Este trabalho contribui com a consolidação de uma base metodológica sólida para estudos virais metagenômicos, oferecendo suporte técnico e conceitual para análises de vigilância ambiental de vírus, como as conduzidas no capítulo anterior desta tese.

Frontiers in Microbiology

TYPE Mini Review
PUBLISHED 31 May 2024
DOI 10.3389/fmicb.2024.1390726

Check for updates

OPEN ACCESS

EDITED BY
Marcin Łoś,
University of Gdansk, Poland

REVIEWED BY
Przemyslaw Decewicz,
University of Warsaw, Poland

*CORRESPONDENCE
Rommel Thiago Jucá Ramos
rommelthiago@gmail.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 23 February 2024
ACCEPTED 16 May 2024
PUBLISHED 31 May 2024

CITATION
Dantas CWD, Martins DT, Nogueira WG, Alegria OVC and Ramos RTJ (2024) Tools and methodology to *in silico* phage discovery in freshwater environments.
Front. Microbiol. 15:1390726.
doi: 10.3389/fmicb.2024.1390726

COPYRIGHT
© 2024 Dantas, Martins, Nogueira, Alegria and Ramos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Tools and methodology to *in silico* phage discovery in freshwater environments

Carlos Willian Dias Dantas[1,2,3†], David Tavares Martins[2,3†], Wylerson Guimarães Nogueira[1†], Oscar Victor Cardenas Alegria[2,3] and Rommel Thiago Jucá Ramos[2,3*]

[1]Department of Biochemistry and Immunology, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, [2]Laboratory of Simulation and Computational Biology — SIMBIC, High Performance Computing Center — CCAD, Federal University of Pará, Belém, Pará, Brazil, [3]Laboratory of Bioinformatics and Genomics of Microorganisms, Institute of Biological Sciences, Federal University of Pará, Belém, Pará, Brazil

Freshwater availability is essential, and its maintenance has become an enormous challenge. Due to population growth and climate changes, freshwater sources are becoming scarce, imposing the need for strategies for its reuse. Currently, the constant discharge of waste into water bodies from human activities leads to the dissemination of pathogenic bacteria, negatively impacting water quality from the source to the infrastructure required for treatment, such as the accumulation of biofilms. Current water treatment methods cannot keep pace with bacterial evolution, which increasingly exhibits a profile of multidrug resistance to antibiotics. Furthermore, using more powerful disinfectants may affect the balance of aquatic ecosystems. Therefore, there is a need to explore sustainable ways to control the spreading of pathogenic bacteria. Bacteriophages can infect bacteria and archaea, hijacking their host machinery to favor their replication. They are widely abundant globally and provide a biological alternative to bacterial treatment with antibiotics. In contrast to common disinfectants and antibiotics, bacteriophages are highly specific, minimizing adverse effects on aquatic microbial communities and offering a lower cost–benefit ratio in production compared to antibiotics. However, due to the difficulty involving cultivating and identifying environmental bacteriophages, alternative approaches using NGS metagenomics in combination with some bioinformatic tools can help identify new bacteriophages that can be useful as an alternative treatment against resistant bacteria. In this review, we discuss advances in exploring the virome of freshwater, as well as current applications of bacteriophages in freshwater treatment, along with current challenges and future perspectives.

KEYWORDS

freshwater, phage, bacteriophage, cyanophage, virome

1 Introduction

Freshwater is an indispensable resource for maintaining life on Earth and has been consistently impacted by the increasing anthropogenic influence. Urban and rural expansion around water bodies, coupled with waste disposal from hospitals, water treatment systems, industry, agriculture, and residences, contribute to rivers and lakes becoming hotspots for the

proliferation of pathogenic microorganisms (Reddy et al., 2022). Water disinfection methods have become limited due to the growing demand for water reuse and the inefficiency of a significant portion of antibiotics against the spread of antibiotic-resistant bacteria (Mathieu et al., 2019). Therefore, there is a pressing need to explore natural compounds to control multidrug-resistant bacteria, such as bacteriophages.

Bacteriophages (or phages) are the most abundant biological entities globally. They were first described in the early 1900s, and by now, we know they are widespread in the environment, with estimates of ~10^{31} phages present in the biosphere (Twort, 1915; Rohwer and Edwards, 2002). Phages act as natural predators of bacteria and archaea, and exploit host machinery favoring their own replication (Dion et al., 2020). Phages may interact with bacterial or archaeal hosts by transferring genes that might be ecologically relevant, thus favoring the host genetic fitness through horizontal gene transfer (HGT) (Touchon et al., 2017; De Mandal et al., 2021). When associated with their hosts as prophages, phages may introduce auxiliary metabolism genes that potentially enhance host adaptability (Luo et al., 2022). The initial discovery that phages were highly abundant in aquatic samples (Bergh et al., 1989) laid the groundwork for the eventual determination of their pivotal impact on the ecosystem.

The paramount significance of phages arises from the viral shunt phenomenon, wherein organic matter is recycled through the lysis of host cells, driving global-scale biogeochemical cycles (Breitbart et al., 2018). Bacteriophages represent an ecological alternative to the use of antibiotics, with a lower cost–benefit ratio of production, and exhibit high specificity to their hosts, minimizing dysbiosis (Romero-Calle et al., 2019). They have been employed for at least a century in controlling bacterial infections in humans (Rohde et al., 2018), and have recently been advocated for applications in freshwater environments (Naknaen et al., 2021; Ben Saad et al., 2022; Hu et al., 2023).

Phages can be classified into three groups: (1) virulent bacteriophages that solely undergo the lytic cycle, leading to the lysis of the host cell; (2) temperate phages that can suffer lysogenic cycles, remaining dormant within the host cell (prophages) but can be induced to switch to the lytic or chronic cycle; and (3) filamentous phages: go through a chronic cycle in which viral replication occurs without host cell lysis (Chevallereau et al., 2022; Zhang et al., 2022). Lytic phages are the most desirable due to their cell lysis capability and lower risk of horizontal gene transfer.

Classical studies of phages relied on isolation and culture methods for their identification (Hyman, 2019). Currently, with the advancement of culture-independent methodologies such as metagenomics, databases are increasingly enriched with viral data, enabling a more comprehensive understanding at the taxonomic level and potential interactions of phages with their hosts (Santiago-Rodriguez and Hollister, 2023) showing that bioinformatics tools for mining viral data can be a powerful aid in discovering bacteriophages.

This review discusses the identification of phages in freshwater environments, the primary *in silico* tools used for phage data exploration, and types of phage applications in freshwater. We also discuss the possible challenges and future possibilities for the field.

2 Identification of phages in freshwater

The metaviromics field (phage metagenomics) essentially is a shotgun metagenomic approach focused on studying the genomes of viral populations from the environment (Hurwitz and Sullivan, 2013; Coutinho et al., 2017; Moon and Cho, 2021), and due to the importance of freshwater bodies as sources of drinking water, recreation, and commerce, more recent studies have dedicated their efforts to freshwater systems (Bruder et al., 2016). Since water chemistry and hydrological factors can contribute to a dynamic environment on a microbial level, likely to be reflected in the indigenous phage populations, the exploration of metagenomic data sampled from freshwater sources from different biomes and places in the world is bound to reveal a plethora of yet unknown and undocumented species of phages (Hayes et al., 2017; Alanazi et al., 2022).

Previous studies have explored how nutrient availability, seasonality, temperature, and human activity influence freshwater viral communities (Bruder et al., 2016). By example, the study of Mohiuddin and Schellhorn (2015) observed that geographic location does not appear to have had a major impact on viral abundance and diversity for two freshwater lakes of the lower Great Lakes region, Lake Ontario and Lake Erie, since the virome composition of both lakes were found to be similar. However, temporal variation in taxonomic composition was observed for both lakes after a year apart sampling.

Another interesting relationship against phage diversity are the possibly related effects of anthropogenic actions on the microbial environment. The study of Green et al., (2015) of the Virginian Lake Matoaka found viral species richness and diversity to be negatively correlated with the level of human activity at the sampling sites, observing the highest levels of diversity and species richness at the main body of the lake, the area least affected by human activity. Another study, conducted by Fancello et al. (2013), observed that the most anthropogenically influenced out of four perennial ponds of the Mauritanian Sahara presented the lowest amount of viral diversity, and higher abundance of heterotrophic microorganisms and human pathogens.

Freshwater viral metagenomics studies also can assist in tackling significant threats to global health, such as the spread of antibiotic resistance. Not only antibiotic resistance genes (ARGs) can spread across different bacterial populations through horizontal gene transfer mediated by bacteriophages, but bacteriophage-carried ARGs are especially threatening due to their prolonged persistence in the environment, fast replication rates, and ability to infect diverse hosts (Brown-Jaque et al., 2015). Moon et al. (2020) explored ARGs recovered from urban surface water viral metagenome data, revealing novel phage-borne antibiotic resistance genes that were also found in bacterial metagenomes, indicating that they were harbored by actively infecting phages. These results suggest that those environmental bacteriophages could act as reservoirs of unknown ARGs that could be widely disseminated via virus-host interactions and illustrate the potential of the viral metagenomics for the discovery of phages involved in spreading antimicrobial resistance on the environment.

In addition, freshwater metagenomic data can also be used to study the viral ecology in the context of other organisms. Chen et al. (2019) investigated and revealed a worldwide distribution of distinct phage genotypes that may infect Fonsibacter, one the most abundant bacterioplankton in freshwater ecosystems, suggesting their substantial role in shaping indigenous microbial communities and potentially significant influence on biogeochemical cycling.

3 *In silico* phage mining with bioinformatic tools

Due to the advances in sequencing technologies and in viral databases, we selected some of the currently most used tools developed to analyze the viral community on metaviromic data. A classic virome analysis pipeline include tools for (i) assembly, (ii) viral sequence prediction, (iii) quality check, (iv) annotation, (v) taxonomy classification, (vi) phage-host prediction tools and (vii) viral microdiversity analysis (Table 1), some being also present in general metagenomic studies (steps i, iv and v). They are essential to understand the diversity of viruses and know their function in the environment, and can be used to identify new uncultivated viral genomes (UViGs) (Green et al., 2015; Moon and Cho, 2021; Naknaen et al., 2021).

In 2017, Roux et al. (2017), identified IDBA-UD (Peng et al., 2012), Megahit (Li et al., 2016), and MetaSpades (Nurk et al., 2017) as the best available options for assembly of viral contigs from short reads. Later on Sutton et al. (2019) analyzed a set of simulated, mocked, and human gut virome with 16 assemblers and identified MetaSpades as the most efficient. However, it showed less effectiveness in reconstructing microdiversity, being more useful to study the mutation rates of the virome. Additionally, although not present in the previous study for being later published, MetaViralSpades (Antipov et al., 2020), a variation of MetaSpades (Nurk et al., 2017), outperformed it in an analysis of 18 real virome data sets, where the contig completeness was superior in 12 cases (Antipov et al., 2020).

After the assembly, a viral sequence prediction analysis can be applied to filter out phages' host sequences from the metagenomic data. There are three main approaches (Andrade-Martínez et al., 2022) which includes tools that uses protein homology searches to databases: VirSorter (Roux et al., 2015), Prophet (Reis-Cunha et al., 2019), PHASTEST (Wishart et al., 2023), MetaPhinder (Jurtz et al., 2016); machine learning based tools that employs reference-free viral genomic features detection: VirFinder (Ren et al., 2017), DeepVirFinder (Ren et al., 2020), PPR-Meta (Fang et al., 2019), PhaMers (Deaton et al., 2019); and hybrid tools that employ machine learning classification reference based or reference independent: VirSorter2 (Nurk et al., 2017; Guo et al., 2021), ViralVerify (Nurk et al., 2017), geNomad (Camargo et al., 2023), Marvel (Amgarten et al., 2018), and VIBRANT (Kieft et al., 2020) (which can do the steps of identify viral sequences, annotation, and determine genome quality and completeness) (Table 1). Each methodology will have its limitations, and, for machine learning, is related to how updated is the training dataset, the alignment-based tools may also be limited by how updated are the datasets and the difficulty to handle large data. The best approach would be a combination of results from tools that utilize different methodologies for phage sequence prediction (Andrade-Martínez et al., 2022).

Contigs obtained from short-read metagenomic sequencing are normally segmented and it might have misleading information, making it difficult to perform further analysis. To help with this issue, the use of tools such as CheckV (Nayfach et al., 2021), ViralComplete (Antipov et al., 2020), or VIBRANT (Kieft et al., 2020), that identify the completeness and possible host contamination on viral genomes is essential, but yet, still need improvements due to be dependable of the database of virus and the tools used (Green et al., 2015; Sutton et al., 2019). In terms of annotation, some of the most known tools to predict ORFs (Open Reading Frames) are prodigal (Hyatt et al., 2010), Glimmer (Delcher et al., 2007) and GeneMarks (Besemer et al., 2001; Andrade-Martínez et al., 2022), but there are other more specific tools for virus annotations such as VIBRANT (Kieft et al., 2020), viral-Eggnog-mapper (Cantalapiedra et al., 2021), DRAM-v (Shaffer et al., 2020), and PHANOTATE (Mcnair et al., 2019; Table 1). They are suitable for viral annotation and can be applied in manual curation of possible viral false positive results taking into account characteristics such as number of viral and cellular genes hits, bitscores, absence of viral hallmark genes, and presence of plasmid genes (Guo et al., 2021).

For taxonomic classification, it is currently a challenge to find tools that can classify viral sequences under the latest ICTV taxonomy framework, given the high variability, lack of universally conserved genes, and unknown regarding viruses. Kraken 2 (Wood et al., 2019), is a powerful tool for virus taxonomy and identification, and a study performed by Ho et al. (2023) detected a high F1 score of 0.86 in the correct detection of sequences of a viral mock community of characterized viruses. However, it has limited homology to the reference used, so it's a good option for the identification of known viruses, and when discovery of new viruses is considered, the use of Kraken 2 combined with other tools is advised (Ho et al., 2023). Among the tools that do taxonomy analysis MMSeqs 2 (Steinegger and Söding, 2017), and CAT (von Meijenfeldt et al., 2019), perform protein homology searches to own databases, VContact 2 (Bin Jang et al., 2019), who employs clustering of viral contigs based on shared genes, PhaGCN (Shang et al., 2021), a deep learning classifier based on gene-sharing networks, and VirusTaxo (Raju et al., 2022), that uses a k-mer enrichment database approach (Table 1). All of these tools have customizable databases or the option to retrain their machine-learning models with the latest ICTV taxonomy, which is essential since the ICTV taxonomy is frequently changing (Zhu et al., 2022).

Considering that freshwater environments are expected to have a considerable percentage of new uncultivated viral genomes (UViGs), if a researcher needs to identify its possible host, it is necessary to perform a phage-host prediction. Current methods include mainly similar oligonucleotide frequency (ONF) analysis (VirHostMatcher) (Ahlgren et al., 2017), k-mer similarity (PHIST) (Zielezinski et al., 2022), CRISPR spacer alignment (Dion et al., 2021), and machine learning algorithms (RAFaH) (Coutinho et al., 2021). For researchers new to metavirome analysis it might helpful to use a software that computes the results of other tools such as iPHoP (Roux et al., 2023), which computes the results of six tools utilizing different methodologies and summarizes the putative taxonomy of phage hosts in a table.

The high volume of data produced by the metagenomic studies stimulated the development of tools to simplify the analysis of metagenomic data that also can be applied to metaviromic datasets. Among them, packages such as Phyloseq (McMurdie and Holmes, 2013), MicrobiomeAnalyst (Dhariwal et al., 2017), Animalcules (Zhao et al., 2021), and Microeco (Liu et al., 2021) are some of the most known integrated R packages available (Wen et al., 2023) and offer great set of graphics to support analysis of environmental viruses and their role through metagenomics.

4 Applications of phages in freshwater

Safe drinking water is a high demand limited resource that gains more attention in research as water resources get scarcer worldwide, and multi-resistant water-borne pathogens and overall pollution grows as an even bigger threat to society over the years (Mathieu et al.,

10.3389/fmicb.2024.1390726

TABLE 1 Tools available for metagenomic analysis of data for viral identification from environmental data.

Tool name	Metavirome analysis type	Description	Input type	Accessibility—web or tandalone	Citation
IDBA-UD	Assembly	Assembly of fastq reads	Processed fastq files	Standalone	Peng et al. (2012)
Megahit	Assembly	Assembly of fastq reads	Processed fastq files	Standalone	Li et al. (2016)
MetaSpades	Assembly	Assembly of fastq reads	Processed fastq files	Standalone	Nurk et al. (2017)
MetaViralSpades	Assembly	Assembly of fastq reads	Processed fastq files	Standalone	Antipov et al. (2020)
VirSorter	Viral sequence prediction	Predicts viral regions using probabilistic similarity models and referenced-based protein homology searches	Contigs in FASTA	Standalone	Roux et al. (2015)
Prophet	Viral sequence prediction	Predicts viral regions based on similarity searches against own database	Contigs in FASTA	Standalone	Reis-Cunha et al. (2019)
PHASTEST	Viral sequence prediction	Predicts viral regions based on similarity searches against own database	Contigs in Genbank or FASTA	Standalone or Web	Wishart et al. (2023)
MetaPhinder	Viral sequence prediction	Identifies phage sequences in assembled contigs by integrating BLAST matches to several phage genomes in a database.	Contigs in FASTA	Standalone	Jurtz et al. (2016)
VirFinder	Viral sequence prediction	Machine learning method for identification of viral contigs based on K-mer distribution	Contigs in FASTA	Standalone	Ren et al. (2017)
DeepVirFinder	Viral sequence prediction	Uses convolutional neural networks to learn viral genomic signatures and predict if a sequence is viral	Contigs in FASTA	Standalone	Ren et al. (2020)
PPR-Meta	Viral sequence prediction	Utilizes neural network structure (CNN) to predict viral sequences	Contigs in FASTA	Standalone	Fang et al. (2019)
PhaMers	Viral sequence prediction	Utilizes a machine learning model based on k-mer frequencies to identify viral sequences	Contigs in FASTA	Standalone	Deaton et al. (2019)
VirSorter2	Viral sequence prediction	Predicts viral regions based on HMM alignment to database	Contigs in FASTA	Standalone	Guo et al. (2021)
ViralVerify	Viral sequence prediction	Analyses the gene content of a contig through a Naive Bayesian classifier and classifies it as viral/bacterial/uncertain	Contigs in FASTA	Standalone	Antipov et al. (2020)
geNomad	Viral sequence prediction	Uses an hybrid approach with gene content and a deep neural network to identify sequences of plasmids and viruses	Contigs in FASTA	Standalone	Camargo et al. (2023)
Marvel	Viral sequence prediction	Identify viral bins based on a Random Forest machine learning approach	Bins in FASTA	Standalone	Amgarten et al. (2018)
CheckV	Quality check	Estimation of viral completeness by viral sequence comparison to database of complete viral genomes	Contigs in FASTA	Standalone	Nayfach et al. (2021)
ViralComplete	Quality check	Estimate viral completeness using the Naive Bayesian Classifier to compute the similarity of a sequence to known virus from the RefSeq database	Contigs in FASTA	Standalone	Antipov et al. (2020)
DRAM-v	Annotation	Protein-similarity-based pipeline specific for viral functional and metabolic profiling	Contigs in FASTA plus a affi table generated by VirSorter2	Standalone	Shaffer et al. (2020)
PHANOTATE	Annotation	A gene calling annotation tool that treats a phage genome as a network of paths, being ORFs treated as favorable, and overlaps and gaps less favorable, but still possible. These paths are represented as a weighted network of connections graph to find the optimal path.	Contigs in FASTA	Standalone	Mcnair et al. (2019)

(Continued)

TABLE 1 (Continued)

Tool name	Metavirome analysis type	Description	Input type	Accessibility—web or tandalone	Citation
VIBRANT	Annotation	Hybrid pipeline that uses protein similarity and machine learning approach to annotate viral sequences	Contigs in FASTA	Standalone	Kieft et al. (2020)
viral-EggNog-Mapper	Annotation	Pipeline that uses orthology assignment approach to annotate eukaryotic or prokaryotic organisms from genome or metagenome samples	Contigs in FASTA	Standalone and Web	Cantalapiedra et al. (2021)
Kraken 2	Read or contig taxonomic classification	Taxonomic classification of microbiome fastq reads or contigs based on k-mer alignment	Fastq or Fasta files	Standalone	Wood et al. (2019)
VContact 2	Contig taxonomic classification	Utilizes whole genome gene-sharing profiles integrating distance-based hierarchical clustering to generate confidence scores for virus classification	FASTA protein file plus a Gene-to-genome mapping table	Standalone	Bin Jang et al. (2019)
MMSeqs 2	Read or contig taxonomic classification	A protein-search-based taxonomy assignment tool that uses a weighted method to assign taxonomic labels	Fastq or Fasta files	Standalone	Steinegger and Söding (2017)
CAT	Contig taxonomic classification	DIAMOND BLASTP homology for contig taxonomic classification	Contigs in FASTA	Standalone	von Meijenfeldt et al. (2019)
PhaGCN	Contig taxonomic classification	Machine-Learning -Based tools that combines the DNA sequence features and protein sequence similarity to assign taxonomic labels	Contigs in FASTA	Standalone	Shang et al. (2021)
VirusTaxo	Contig taxonomic classification	Taxonomic classification tool that provides a framework to organize the population of viruses from metagenomic data.	Contigs in FASTA	Standalone	Raju et al. (2022)
PHIST	Phage-host prediction	Infer virus host relationships based on the number of k-mers shared between their sequences	FASTA file with host sequences and a FASTA file with viral sequences	Standalone	Zielezinski et al. (2022)
CrisprOpenDB pipeline	Phage-host prediction	Predicts virus-hosts relationships by searching for CRISPR spacer matches and uses several criteria to create predictions	Contigs in FASTA	Standalone and Web	Dion et al. (2021)
RaFah	Phage-host prediction	Utilizes Random Forest model to assign phage-host interaction independent from databases	Contigs in FASTA	Standalone	Coutinho et al. (2021)
VirHostMatcher	Phage-host prediction	Assign virus-host relations based on oligonucleotide frequency similarity	FASTA file with host sequences, FASTA file with viral sequences and Host taxonomy table	Standalone	Ahlgren et al. (2017)
IPHoP	Phage-host prediction	A pipeline that combined multiple tools and methods for phage-host prediction	Contigs in FASTA	Standalone	Roux et al. (2023)
Phyloseq	Diversity analysis	Generate microbial diversity statistics	Virus count table, viral taxonomy table, metadata table	Standalone	McMurdie and Holmes (2013)
MicrobiomeAnalyst	Diversity analysis	Generate microbial diversity statistics	Virus count table, viral taxonomy table, metadata table	Standalone and Web	Dhariwal et al. (2017)
Animalcules	Diversity analysis	Generate microbial diversity statistics	Virus count table, viral taxonomy table, metadata table	Standalone	Zhao et al. (2021)

(Continued)

10.3389/fmicb.2024.1390726

TABLE 1 (Continued)

Tool name	Metavirome analysis type	Description	Input type	Accessibility—web or tandalone	Citation
Microeco	Diversity analysis	Generate microbial diversity statistics	Virus count table, viral taxonomy table, metadata table	Standalone	Liu et al. (2021)

2019). Approximately one-ninth of the global population reportedly lacks access to safe drinking water (Jassim et al., 2016). Given the capacity of phages to infect bacterial hosts, they have recently been used as novel tools in water pollution control, to monitor and treat fresh and wastewater (Ji et al., 2021).

4.1 Bacteriophages as pollution indicators in water

There have been a few applied methods using phages to evaluate water quality as properties indicators to monitor pathogenic bacteria in wastewater. Immobilized phages have been used on an electrode surface as biorecognition elements, through a technique known as electrochemical impedance spectroscopy (EIS), to detect bacteria, such as *E. coli*, *Staphylococcus aureus*, and *Pseudomonas aeruginosa* (Yue et al., 2017; Zhou et al., 2017; Richter et al., 2018). Phages have also been employed as capture elements by other alternative combinations with nanoparticles for bacterial pathogen detection (Richter et al., 2018), and as biomechanisms to assess membrane performance and monitor membrane integrity in water treatment facilities (McMinn et al., 2017; Wu et al., 2017; Dias et al., 2018).

A specific group of bacteriophages named crAssphages have been proposed as potential universal human feces viral indicators in water bodies (Farkas et al., 2019; Mafumo et al., 2023). CrAssphages were described by Dutilh et al. (2014) as the most abundant phages in the human gut virome. Further studies identified that crAssphages are highly specific and abundant to human feces (Sabar et al., 2022), highly prevalent in sewage samples (Stachler et al., 2017), and maintain correlation to the presence of human enteric viruses in water (Jennings et al., 2020). Given the previous characteristics, crAssphages have been preconized in favor of currently used fecal indicator bacteria (FIB), which poorly explain viral pathogen dynamics in water and have low host specificity, making difficult the identification of the source of contamination (Ward et al., 2020; Toribio-Avedillo et al., 2021; Mafumo et al., 2023). CrAssphage applicability has been evaluated in several countries (Crank et al., 2020; Ward et al., 2020; Sangkaew et al., 2021; Nam et al., 2022) and shows promising possibilities for human fecal contamination detection in freshwater.

4.2 Bacteriophages in water treatment

Another challenge that greatly affects the operation of wastewater treatment systems is the formation of flocs and sludge bulking by filamentous microorganisms that proliferate excessively that form thick, viscous foams (Aracic et al., 2015). The study conducted by Petrovski et al. (2011a,b) showed how phages that can lyse multiple host bacteria can circumvent the stability of foams. Additionally, Liu et al. (2015) performed tests in a simulated aeration tank system using isolated Gordonia phages, achieving significant reduction in the sludge sedimentation volume. However, all these methods are still experimental as current research still focuses on evaluating and monitoring the behavior of potential phage candidates on wastewater treatment systems (Reisoglu and Aydin, 2023).

Other lines of research have employed phages as low-cost biological control agents to treat specific pathogenic bacteria in sewage. Studies reported the successful inhibition and lysis of

drug-resistant *A. baumannii* (Lin et al., 2010), waterborne disease-causing *Vibrio cholerae* (Wei et al., 2011), and dysentery-causing *Shigella* (Jun et al., 2016) through the combination of different phages in co-culture essays. Also, some studies act on the biological control of cyanobacteria, harmful prokaryotes often causing water blooms on green or red tides and producing cyanotoxins, which endanger the surrounding wildlife, aquatic farming animals, threaten human health and can cause tremendous economic losses (Jassim and Limoges, 2013). The strategy in some of those studies is to isolate and employ cyanophages that effectively reduce phycobilisome proteins and destroy the thylakoid structure of cyanobacteria (Gao et al., 2012; Yoshida-Takashima et al., 2012).

However, for both cases, some problems still emerge in the practical application of phage-based biological control, with the emergence of host-resistant mutants, the reduction of cyanophage infectivity caused by sunlight irradiation, and the feasibility of multiple-host approaches are still challenges to be overcome. Nonetheless, phage-based technology also has the advantage of reducing the use of chemical reagents, thus reinforcing the appeal of such strategies and interest in their future development (Mathieu et al., 2019; Ji et al., 2021).

5 Current limitations and perspectives

The study of viral sequences in environmental samples is challenging due to the low representativity or fragmentation of DNA in short sequencing data, the high error rate and the large amount of DNA necessary for long-read sequencing (Warwick-Dugdale et al., 2019). As technology advances, improved read length and sequencing quality have partially addressed this issue. This progress has also opened up the opportunity to implement hybrid approaches for sequencing, combining short and long reads that might allow better environmental virus detection, characterization, and understanding of the microdiversity of virus populations (Warwick-Dugdale et al., 2019; Pratama et al., 2021; Andrade-Martínez et al., 2022).

For the identification of phages, common tools employ distinct methods, such as sequence composition, sequence similarity, and machine learning approaches (Titus Brown and Irber, 2016; Fang et al., 2019; Kieft et al., 2020), but there is no standardization for these techniques. Currently, each method yields slightly different results, and phage identification still relies heavily on trial and error usage of software packages. It is crucial that a golden standard be established to ensure the robustness of methodologies and techniques, thereby enhancing the replicability and reliability of phage identification.

An alternative for an assembly-free, culture-independent study of phages is the analysis of the whole genome of phages by using long-read sequencing technologies, like Oxford nanopore or PacBio technologies (Warwick-Dugdale et al., 2019; Zaragoza-Solas et al., 2022). The advantages of this approach are avoiding over-fragmentation of sequencing data and adopting portable sequencing technologies, allowing the researcher to identify phages from natural sources *in situ* (Warwick-Dugdale et al., 2019). This opportunity leads to the study of phages directly from their natural environment, allowing for the identification of phages and the analysis of the samples in real time, which is a significant and desirable feature for the genomic surveillance field (Lisotto et al., 2021).

Most virus databases are derived from uncultivated viral genomes (UViGs) representing >95% of public databases (Roux et al., 2019), leading to another significant problem: most of the phage-host interactions are obtained solely from *in silico* predictions of the study of metagenomes. This lack of lab-studied observations implies the absence of a clear understanding of host-phage dynamics in nature (Coclet and Roux, 2021). In addition to avoiding the intrinsic wet lab biases, such as the identification of false positive or negative viruses due to contamination, the increase of biases related to the process of sample collection, storage, genetic material extraction, purification, and sequencing (Cantalupo and Pipas, 2019). However, the lack of this holistic vision might affect the build of future databases and the scientific interpretations from related results, so it is vital to keep these current limitations presented by bioinformatics tools in mind and apply different combinations of analysis to confirm the identity of phages coming from metagenomic data (Roux et al., 2013, 2019).

Author contributions

CD: Writing – original draft. DM: Writing – original draft. WN: Writing – original draft. OA: Formal analysis, Supervision, Writing – review & editing. RR: Conceptualization, Formal analysis, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for this article's research, authorship, and publication. The authors would like to thank the funding agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and National Council for Scientific and Technological Development (CNPq), the support by the CNPq project #312316/2022-4, the Secretary of State for Science, Technology, and Professional and Technological Education (SECTET), the Dean's Office for Research and Graduate Studies/Federal University of Pará–PROPESP/UFPA (PAPQ), and the partnership SECTEC/UFPA/FADESP for the financial support on this work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1390726/full#supplementary-material

References

Ahlgren, N. A., Ren, J., Lu, Y. Y., Fuhrman, J. A., and Sun, F. (2017). Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res.* 45, 39–53. doi: 10.1093/nar/gkw1002

Alanazi, F., Nour, I., Hanif, A., Al-Ashkar, I., Aljowaie, R. M., and Eifan, S. (2022). Novel findings in context of molecular diversity and abundance of bacteriophages in wastewater environments of Riyadh, Saudi Arabia. *PLoS One* 17:e0273343. doi: 10.1371/journal.pone.0273343

Amgarten, D., Braga, L. P. P., da Silva, A. M., and Setubal, J. C. (2018). MARVEL, a tool for prediction of bacteriophage sequences in metagenomic bins. *Front. Genet.* 9:304. doi: 10.3389/fgene.2018.00304

Andrade-Martínez, J. S., Camelo Valera, L. C., Chica Cárdenas, L. A., Forero-Junco, L., López-Leal, G., Moreno-Gallego, J. L., et al. (2022). Computational tools for the analysis of uncultivated phage genomes. *Microbiol. Mol. Biol. Rev.* 86:e0000421. doi: 10.1128/mmbr.00004-21

Antipov, D., Raiko, M., Lapidus, A., and Pevzner, P. A. (2020). Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinformatics* 36, 4126–4129. doi: 10.1093/bioinformatics/btaa490

Aracic, S., Manna, S., Petrovski, S., Wiltshire, J. L., Mann, G., and Franks, A. E. (2015). Innovative biological approaches for monitoring and improving water quality. *Front. Microbiol.* 6:826. doi: 10.3389/fmicb.2015.00826

Ben Saad, M., Ben Said, M., Bousselmi, L., and Ghrabi, A. (2022). Use of bacteriophage to inactivate pathogenic bacteria from wastewater. *J. Environ. Sci. Health A* 57, 111–116. doi: 10.1080/10934529.2022.2036551

Bergh, Ø., KY, B. Ø., Bratbak, G., and Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature* 340, 467–468. doi: 10.1038/340467a0

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618. doi: 10.1093/nar/29.12.2607

Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* 37, 632–639. doi: 10.1038/s41587-019-0100-8

Breitbart, M., Bonnain, C., Malki, K., and Sawaya, N. A. (2018). Phage puppet masters of the marine microbial realm. *Nat. Microbiol.*. Nature Publishing Group 3, 754–766. doi: 10.1038/s41564-018-0166-y

Brown-Jaque, M., Calero-Cáceres, W., and Muniesa, M. (2015). Transfer of antibiotic-resistance genes via phage-related mobile elements. *Plasmid* 79, 1–7. doi: 10.1016/j.plasmid.2015.01.001

Bruder, K., Malki, K., Cooper, A., Sible, E., Shapiro, J. W., Watkins, S. C., et al. (2016). Freshwater metaviromics and bacteriophages: a current assessment of the state of the art in relation to bioinformatic challenges. *Evol. Bioinforma.* 12, 25–33. doi: 10.4137/EBO.S38549

Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., et al. (2023). Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* doi: 10.1038/s41587-023-01953-y

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293

Cantalupo, P. G., and Pipas, J. M. (2019). Detecting viral sequences in NGS data. *Curr. Opin. Virol.* 39, 41–48. doi: 10.1016/j.coviro.2019.07.010

Chen, L. X., Zhao, Y., McMahon, K. D., Mori, J. F., Jessen, G. L., Nelson, T. C., et al. (2019). Wide distribution of phage that infect freshwater SAR11 Bacteria. *mSystems.* 4:e00410-19. doi: 10.1128/mSystems.00410-19

Chevallereau, A., Pons, B. J., van Houte, S., and Westra, E. R. (2022). Interactions between bacterial and phage communities in natural environments. *Nat Rev Microbiol* 20, 49–62. doi: 10.1038/s41579-021-00602-y

Coclet, C., and Roux, S. (2021). Global overview and major challenges of host prediction methods for uncultivated phages. *Curr. Opin. Virol.* 49, 117–126. doi: 10.1016/j.coviro.2021.05.003

Coutinho, F. H., Silveira, C. B., Gregoracci, G. B., Thompson, C. C., Edwards, R. A., Brussaard, C. P. D., et al. (2017). Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat. Commun.* 8:15955. doi: 10.1038/ncomms15955

Coutinho, F. H., Zaragoza-Solas, A., López-Pérez, M., Barylski, J., Zielezinski, A., Dutilh, B. E., et al. (2021). RaFAH: host prediction for viruses of Bacteria and Archaea based on protein content. *Patterns.* 2:100274. doi: 10.1016/j.patter.2021.100274

Crank, K., Li, X., North, D., Ferraro, G. B., Iaconelli, M., Mancini, P., et al. (2020). CrAssphage abundance and correlation with molecular viral markers in Italian wastewater. *Water Res.* 184:116161. doi: 10.1016/j.watres.2020.116161

De Mandal, S., Panda, A. K., Kumar, N. S., Bisht, S. S., and Jin, F. (2021). *Metagenomics and microbial ecology*. Boca Raton: CRC Press.

Deaton, J., Yu, F. B., and Quake, S. R. (2019). Mini-metagenomics and nucleotide composition aid the identification and host Association of Novel Bacteriophage Sequences. *Adv Biosyst* 3:e1900108. doi: 10.1002/adbi.201900108

Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics* 23, 673–679. doi: 10.1093/bioinformatics/btm009

Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., and Xia, J. (2017). MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res.* 45, W180–W188. doi: 10.1093/nar/gkx295

Dias, E., Ebdon, J., and Taylor, H. (2018). The application of bacteriophages as novel indicators of viral pathogens in wastewater treatment systems. *Water Res.* 129, 172–179. doi: 10.1016/j.watres.2017.11.022

Dion, M. B., Oechslin, F., and Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nat. Rev. Microbiol.* 18, 125–138. doi: 10.1038/s41579-019-0311-5

Dion, M. B., Plante, P. L., Zufferey, E., Shah, S. A., Corbeil, J., and Moineau, S. (2021). Streamlining CRISPR spacer-based bacterial host predictions to decipher the viral dark matter. *Nucleic Acids Res.* 49, 3127–3138. doi: 10.1093/nar/gkab133

Dutilh, B. E., Cassman, N., McNair, K., Sanchez, S. E., Silva, G. G. Z., Boling, L., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* 5:4498. doi: 10.1038/ncomms5498

Fancello, L., Trape, S., Robert, C., Boyer, M., Popgeorgiev, N., Raoult, D., et al. (2013). Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME J.* 7, 359–369. doi: 10.1038/ismej.2012.101

Fang, Z., Tan, J., Wu, S., Li, M., Xu, C., Xie, Z., et al. (2019). PPR-meta: a tool for identifying phages and plasmids from metagenomic fragments using deep learning. *Gigascience* 8:giz066. doi: 10.1093/gigascience/giz066

Farkas, K., Adriaenssens, E. M., Walker, D. I., McDonald, J. E., Malham, S. K., and Jones, D. L. (2019). Critical evaluation of CrAssphage as a molecular marker for human-derived wastewater contamination in the aquatic environment. *Food Environ Virol* 11, 113–119. doi: 10.1007/s12560-019-09369-1

Gao, E. B., Gui, J. F., and Zhang, Q. Y. (2012). A novel Cyanophage with a cyanobacterial nonbleaching protein a gene in the genome. *J. Virol.* 86, 236–245. doi: 10.1128/JVI.06282-11

Green, J., Rahman, F., Saxton, M., and Williamson, K. (2015). Metagenomic assessment of viral diversity in Lake Matoaka, a temperate, eutrophic freshwater lake in southeastern Virginia, USA. *Aquat. Microb. Ecol.* 75, 117–128. doi: 10.3354/ame01752

Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., et al. (2021). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9:37. doi: 10.1186/s40168-020-00990-y

Guo, J., Vik, D., Pratama, A. A., Roux, S., and Sullivan, M.. *Viral sequence identification SOP with VirSorter2. protocols.io*. (2021). Available at: https://www.protocols.io/view/viral-sequence-identification-sop-with-virsorter2-5qpvoyqebg4o/v3 (Accessed Jan 5, 2024).

Hayes, S., Mahony, J., Nauta, A., and van Sinderen, D. (2017). Metagenomic approaches to assess bacteriophages in various environmental niches. *Viruses* 9:127. doi: 10.3390/v9060127

Ho, S. F. S., Wheeler, N. E., Millard, A. D., and van Schaik, W. (2023). Gauge your phage: benchmarking of bacteriophage identification tools in metagenomic sequencing data. *Microbiome* 11:84. doi: 10.1186/s40168-023-01533-x

Hu, M., Xing, B., Yang, M., Han, R., Pan, H., Guo, H., et al. (2023). Characterization of a novel genus of jumbo phages and their application in wastewater treatment. *iScience* 26:106947. doi: 10.1016/j.isci.2023.106947

Hurwitz, B. L., and Sullivan, M. B. (2013). The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS One* 8:e57355. doi: 10.1371/journal.pone.0057355

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Hyman, P. (2019). Phages for phage therapy: isolation, characterization, and host range breadth. *Pharmaceuticals* 12:35. doi: 10.3390/ph12010035

Jassim, S. A. A., and Limoges, R. G. (2013). Impact of external forces on cyanophage–host interactions in aquatic ecosystems. *World J. Microbiol. Biotechnol.* 29, 1751–1762. doi: 10.1007/s11274-013-1358-5

Jassim, S. A. A., Limoges, R. G., and El-Cheikh, H. (2016). Bacteriophage biocontrol in wastewater treatment. *World J. Microbiol. Biotechnol.* 32:70. doi: 10.1007/s11274-016-2028-1

Jennings, W. C., Gálvez-Arango, E., Prieto, A. L., and Boehm, A. B. (2020). CrAssphage for fecal source tracking in Chile: covariation with norovirus, HF183, and bacterial indicators. *Water Res X* 9:100071. doi: 10.1016/j.wroa.2020.100071

Ji, M., Liu, Z., Sun, K., Li, Z., Fan, X., and Li, Q. (2021). Bacteriophages in water pollution control: advantages and limitations. *Front. Environ. Sci. Eng.* 15:84. doi: 10.1007/s11783-020-1378-y

Jun, J. W., Giri, S. S., Kim, H. J., Yun, S. K., Chi, C., Chai, J. Y., et al. (2016). Bacteriophage application to control the contaminated water with Shigella. *Sci. Rep.* 6:22636. doi: 10.1038/srep22636

Jurtz, V. I., Villarroel, J., Lund, O., Voldby Larsen, M., and Nielsen, M. (2016). MetaPhinder—identifying bacteriophage sequences in metagenomic data sets. *PLoS One* 11:e0163111. doi: 10.1371/journal.pone.0163111

Kieft, K., Zhou, Z., and Anantharaman, K. (2020). VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 8:90. doi: 10.1186/s40168-020-00867-0

Li, D., Luo, R., Liu, C. M., Leung, C. M., Ting, H. F., Sadakane, K., et al. (2016). MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102, 3–11. doi: 10.1016/j.ymeth.2016.02.020

Lin, N. T., Chiou, P. Y., Chang, K. C., Chen, L. K., and Lai, M. J. (2010). Isolation and characterization of ϕAB2: a novel bacteriophage of *Acinetobacter baumannii*. *Res. Microbiol.* 161, 308–314. doi: 10.1016/j.resmic.2010.03.007

Lisotto, P., Raangs, E. C., Couto, N., Rosema, S., Lokate, M., Zhou, X., et al. (2021). Long-read sequencing-based in silico phage typing of vancomycin-resistant *Enterococcus faecium*. *BMC Genomics* 22:758. doi: 10.1186/s12864-021-08080-5

Liu, C., Cui, Y., Li, X., and Yao, M. (2021). Microeco: an R package for data mining in microbial community ecology. *FEMS Microbiol. Ecol.* 97:fiaa255. doi: 10.1093/femsec/fiaa255

Liu, M., Gill, J. J., Young, R., and Summer, E. J. (2015). Bacteriophages of wastewater foaming-associated filamentous Gordonia reduce host levels in raw activated sludge. *Sci. Rep.* 5:13754. doi: 10.1038/srep13754

Luo, X. Q., Wang, P., Li, J. L., Ahmad, M., Duan, L., Yin, L. Z., et al. (2022). Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. *Microbiome* 10:190. doi: 10.1186/s40168-022-01384-y

Mafumo, N., Bezuidt, O. K. I., le Roux, W., and Makhalanyane, T. P. (2023). CrAssphage may be viable markers of contamination in pristine and Contaminated River water. *mSystems* 8:e0128222. doi: 10.1128/msystems.01282-22

Mathieu, J., Yu, P., Zuo, P., Da Silva, M. L. B., and Alvarez, P. J. J. (2019). Going viral: emerging opportunities for phage-based bacterial control in water treatment and reuse. *Acc. Chem. Res.* 52, 849–857. doi: 10.1021/acs.accounts.8b00576

McMinn, B. R., Ashbolt, N. J., and Korajkic, A. (2017). Bacteriophages as indicators of faecal pollution and enteric virus removal. *Lett. Appl. Microbiol.* 65, 11–26. doi: 10.1111/lam.12736

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217

Mcnair, K., Zhou, C., Dinsdale, E. A., Souza, B., and Edwards, R. A. (2019). PHANOTATE: a novel approach to gene identification in phage genomes. *Bioinformatics* 35, 4537–4542. doi: 10.1093/bioinformatics/btz265

Mohiuddin, M., and Schellhorn, H. E. (2015). Spatial and temporal dynamics of virus occurrence in two freshwater lakes captured through metagenomic analysis. *Front. Microbiol.* 6:960. doi: 10.3389/fmicb.2015.00960

Moon, K., and Cho, J. C. (2021). Metaviromics coupled with phage-host identification to open the viral 'black box.'. *J. Microbiol.* 59, 311–323. doi: 10.1007/s12275-021-1016-9

Moon, K., Jeon, J. H., Kang, I., Park, K. S., Lee, K., Cha, C. J., et al. (2020). Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. *Microbiome* 8:75. doi: 10.1186/s40168-020-00863-4

Naknaen, A., Suttinun, O., Surachat, K., Khan, E., and Pomwised, R. (2021). A novel jumbo phage PhiMa05 inhibits harmful Microcystis sp. *Front. Microbiol.* 12:660351. doi: 10.3389/fmicb.2021.660351

Nam, S. J., Hu, W. S., and Koo, O. K. (2022). Evaluation of crAssphage as a human-specific microbial source-tracking marker in the Republic of Korea. *Environ. Monit. Assess.* 194:367. doi: 10.1007/s10661-022-09918-5

Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., and Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* 39, 578–585. doi: 10.1038/s41587-020-00774-7

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116

Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174

Petrovski, S., Seviour, R. J., and Tillett, D. (2011a). Prevention of Gordonia and Nocardia stabilized foam formation by using bacteriophage GTE7. *Appl. Environ. Microbiol.* 77, 7864–7867. doi: 10.1128/AEM.05692-11

Petrovski, S., Seviour, R. J., and Tillett, D. (2011b). Characterization of the genome of the polyvalent lytic bacteriophage GTE2, which has potential for biocontrol of Gordonia-, Rhodococcus-, and Nocardia-stabilized foams in activated sludge plants. *Appl. Environ. Microbiol.* 77, 3923–3929. doi: 10.1128/AEM.00025-11

Pratama, A. A., Bolduc, B., Zayed, A. A., Zhong, Z. P., Guo, J., Vik, D. R., et al. (2021). Expanding standards in viromics: in silico evaluation of dsDNA viral genome identification, classification, and auxiliary metabolic gene curation. *PeerJ* 9:e11447. doi: 10.7717/peerj.11447

Raju, R. S., Al Nahid, A., Chondrow Dev, P., and Islam, R. (2022). VirusTaxo: taxonomic classification of viruses from the genome sequence using k-mer enrichment. *Genomics* 114:110414. doi: 10.1016/j.ygeno.2022.110414

Reddy, S., Kaur, K., Barathe, P., Shriram, V., Govarthanan, M., and Kumar, V. (2022). Antimicrobial resistance in urban river ecosystems. *Microbiol. Res.* 263:127135. doi: 10.1016/j.micres.2022.127135

Reis-Cunha, J. L., Bartholomeu, D. C., Manson, A. L., Earl, A. M., and Cerqueira, G. C. (2019). ProphET, prophage estimation tool: a stand-alone prophage sequence prediction tool with self-updating reference database. *PLoS One* 14:e0223364. doi: 10.1371/journal.pone.0223364

Reisoglu, Ş., and Aydin, S. (2023). Bacteriophages as a promising approach for the biocontrol of antibiotic resistant pathogens and the reconstruction of microbial interaction networks in wastewater treatment systems: a review. *Sci. Total Environ.* 890:164291. doi: 10.1016/j.scitotenv.2023.164291

Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., and Sun, F. (2017). VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 5:69. doi: 10.1186/s40168-017-0283-5

Ren, J., Song, K., Deng, C., Ahlgren, N. A., Fuhrman, J. A., Li, Y., et al. (2020). Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* 8, 64–77. doi: 10.1007/s40484-019-0187-4

Richter, Ł., Janczuk-Richter, M., Niedziółka-Jönsson, J., Paczesny, J., and Hołyst, R. (2018). Recent advances in bacteriophage-based methods for bacteria detection. *Drug Discov. Today* 23, 448–455. doi: 10.1016/j.drudis.2017.11.007

Rohde, C., Resch, G., Pirnay, J. P., Blasdel, B., Debarbieux, L., Gelman, D., et al. (2018). Expert opinion on three phage therapy related topics: bacterial phage resistance, phage training and prophages in bacterial production strains. *Viruses* 10:178. doi: 10.3390/v10040178

Rohwer, F., and Edwards, R. (2002). The phage proteomic tree: a genome-based taxonomy for phage. *J. Bacteriol.* 184, 4529–4535. doi: 10.1128/JB.184.16.4529-4535.2002

Romero-Calle, D., Guimarães Benevides, R., Góes-Neto, A., and Billington, C. (2019). Bacteriophages as alternatives to antibiotics in clinical care. *Antibiotics* 8:138. doi: 10.3390/antibiotics8030138

Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., et al. (2019). Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* 37, 29–37. doi: 10.1038/nbt.4306

Roux, S., Camargo, A. P., Coutinho, F. H., Dabdoub, S. M., Dutilh, B. E., Nayfach, S., et al. (2023). iPHoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol.* 21:e3002083. doi: 10.1371/journal.pbio.3002083

Roux, S., Emerson, J. B., Eloe-Fadrosh, E. A., and Sullivan, M. B. (2017). Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5:e3817. doi: 10.7717/peerj.3817

Roux, S., Enault, F., Hurwitz, B. L., and Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. doi: 10.7717/peerj.985

Roux, S., Krupovic, M., Debroas, D., Forterre, P., and Enault, F. (2013). Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biol.* 3:130160. doi: 10.1098/rsob.130160

Sabar, M. A., Honda, R., and Haramoto, E. (2022). CrAssphage as an indicator of human-fecal contamination in water environment and virus reduction in wastewater treatment. *Water Res.*. Elsevier Ltd 221:118827. doi: 10.1016/j.watres.2022.118827

Sangkaew, W., Kongprajug, A., Chyerochana, N., Ahmed, W., Rattanakul, S., Denpetkul, T., et al. (2021). Performance of viral and bacterial genetic markers for sewage pollution tracking in tropical Thailand. *Water Res.* 190:116706. doi: 10.1016/j.watres.2020.116706

Santiago-Rodriguez, T. M., and Hollister, E. B. (2023). Viral metagenomics as a tool to track sources of fecal contamination: a one health approach. *Viruses*. MDPI 15:236. doi: 10.3390/v15010236

Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., et al. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* 48, 8883–8900. doi: 10.1093/nar/gkaa621

Shang, J., Jiang, J., and Sun, Y. (2021). Bacteriophage classification for assembled contigs using graph convolutional network. *Bioinformatics* 37, i25–i33. doi: 10.1093/bioinformatics/btab293

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., and Shanks, O. C. (2017). Quantitative CrAssphage PCR assays for human fecal pollution measurement. *Environ. Sci. Technol.* 51, 9146–9154. doi: 10.1021/acs.est.7b02703

Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026–1028. doi: 10.1038/nbt.3988

Sutton, T. D. S., Clooney, A. G., Ryan, F. J., Ross, R. P., and Hill, C. (2019). Choice of assembly software has a critical impact on virome characterisation. *Microbiome* 7:12. doi: 10.1186/s40168-019-0626-5

Titus Brown, C., and Irber, L. (2016). Sourmash: a library for MinHash sketching of DNA. *J. Open Source Softw.* 1:27. doi: 10.21105/joss.00027

Toribio-Avedillo, D., Blanch, A. R., Muniesa, M., and Rodríguez-Rubio, L. (2021). Bacteriophages as fecal pollution indicators. *Viruses* 13:1089. doi: 10.3390/v13061089

Touchon, M., Moura de Sousa, J. A., and Rocha, E. P. (2017). Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Curr. Opin. Microbiol.* 38, 66–73. doi: 10.1016/j.mib.2017.04.010

Twort, F. W. (1915). An investigation on the nature of ultra-microscopic viruses. *Lancet* 186, 1241–1243. doi: 10.1016/S0140-6736(01)20383-3

von Meijenfeldt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H., and Dutilh, B. E. (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol.* 20:217. doi: 10.1186/s13059-019-1817-x

Ward, L. M., Ghaju Shrestha, R., Tandukar, S., Sherchand, J. B., Haramoto, E., and Sherchan, S. P. (2020). Evaluation of CrAssphage marker for tracking fecal contamination in river water in Nepal. *Water Air Soil Pollut.* 231:282. doi: 10.1007/s11270-020-04648-1

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A. C., Allen, M. J., et al. (2019). Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* 7:e6800. doi: 10.7717/peerj.6800

Wei, Y., Kirby, A., and Levin, B. R. (2011). The population and evolutionary dynamics of *vibrio cholerae* and its bacteriophage: conditions for maintaining phage-limited communities. *Am. Nat.* 178, 715–725. doi: 10.1086/662677

Wen, T., Niu, G., Chen, T., Shen, Q., Yuan, J., and Liu, Y. X. (2023). The best practice for microbiome analysis using R. *Protein Cell* 14, 713–725. doi: 10.1093/procel/pwad024

Wishart, D. S., Han, S., Saha, S., Oler, E., Peters, H., Grant, J. R., et al. (2023). PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res.* 51, W443–W450. doi: 10.1093/nar/gkad382

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with kraken 2. *Genome Biol.* 20:257. doi: 10.1186/s13059-019-1891-0

Wu, B., Wang, R., and Fane, A. G. (2017). The roles of bacteriophages in membrane-based water and wastewater treatment processes: a review. *Water Res.* 110, 120–132. doi: 10.1016/j.watres.2016.12.004

Yoshida-Takashima, Y., Yoshida, M., Ogata, H., Nagasaki, K., Hiroishi, S., and Yoshida, T. (2012). Cyanophage infection in the bloom-forming Cyanobacteria <i>Microcystis aeruginosa</i> in surface freshwater. *Microbes Environ.* 27, 350–355. doi: 10.1264/jsme2.ME12037

Yue, H., He, Y., Fan, E., Wang, L., Lu, S., and Fu, Z. (2017). Label-free electrochemiluminescent biosensor for rapid and sensitive detection of *pseudomonas aeruginosa* using phage as highly specific recognition agent. *Biosens. Bioelectron.* 94, 429–432. doi: 10.1016/j.bios.2017.03.033

Zaragoza-Solas, A., Haro-Moreno, J. M., Rodriguez-Valera, F., and López-Pérez, M. (2022). Long-read metagenomics improves the recovery of viral diversity from complex natural marine samples. *mSystems* 7:e0019222:202228. doi: 10.1128/msystems.00192-22

Zhang, M., Zhang, T., Yu, M., Chen, Y. L., and Jin, M. (2022). The life cycle transitions of temperate phages: regulating factors and potential ecological implications. *Viruses* 14, 14:1904:1904. doi: 10.3390/v14122818

Zhao, Y., Federico, A., Faits, T., Manimaran, S., Segrè, D., Monti, S., et al. (2021). Animalcules: interactive microbiome analytics and visualization in R. *Microbiome* 9:76. doi: 10.1186/s40168-021-01013-0

Zhou, Y., Marar, A., Kner, P., and Ramasamy, R. P. (2017). Charge-directed immobilization of bacteriophage on nanostructured electrode for whole-cell electrochemical biosensors. *Anal. Chem.* 89, 5734–5741. doi: 10.1021/acs.analchem.6b03751

Zhu, Y., Shang, J., Peng, C., and Sun, Y. (2022). Phage family classification under Caudoviricetes: a review of current tools using the latest ICTV classification framework. *Front. Microbiol.* 13:1032186. doi: 10.3389/fmicb.2022.1032186

Zielezinski, A., Deorowicz, S., and Gudyś, A. (2022). PHIST: fast and accurate prediction of prokaryotic hosts from metagenomic viral sequences. *Bioinformatics* 38, 1447–1449. doi: 10.1093/bioinformatics/btab837

6. ARTIGO III – RESISTOMA DE CORPOS HÍDRICOS DO PARÁ

O manuscrito completo para o terceiro e principal artigo desta tese, intitulado "Mapping the Resistome: Metagenomic Surveillance Reveals AMR Hotspots and Ecological Risk in Amazonian Freshwaters", referente às análises realizadas neste trabalho, foi submetido a periódico científico indexado, à revista *iMeta* (ISSN: 2770-5986), em maio de 2025, fator de impacto: 23.7 (2024).

Este estudo representa o núcleo central desta pesquisa de doutorado, apresentando uma caracterização abrangente do resistoma de corpos d'água localizados em quatro municípios do Estado do Pará. Utilizando abordagens de sequenciamento metagenômico de alta profundidade, o trabalho descreve a composição microbiana, bem como a abundância e diversidade de genes de resistência a antimicrobianos (ARGs) detectados nas amostras. Além disso, propõe um escore de ameaça de risco baseado em parâmetros funcionais dos ARGs identificados e explora associações espaciais entre os níveis de desmatamento, indicadores de impacto antrópico e a distribuição do resistoma nas regiões analisadas.

Este artigo fornece uma contribuição inédita ao mapeamento genômico da resistência antimicrobiana em ambientes amazônicos, evidenciando zonas de risco ecológico e sanitário. Seus resultados reforçam a importância do monitoramento ambiental contínuo e posicionam a metagenômica como uma ferramenta essencial para ações de vigilância em saúde pública sob a perspectiva de Saúde Única.

Mapping the Resistome: Metagenomic Surveillance Reveals AMR Hotspots and Ecological Risk in Amazonian Freshwaters

Wylerson Guimarães Nogueira[1,2], David Tavares Martins[3,4,5], Carlos Willian Dias Dantas[3,4,5], Victor Benedito Costa Ferreira[3,4,5], Oscar Victor Cardenas-Alegria[3,4,5], Edian Franklin Franco de Los Santos[6], Paulo Rogenes Monteiro Pontes[7], Rosane Barbosa Lopes Cavalcante[7], Júnia Schultz[2], Vasco Ariston de Carvalho Azevedo[8], Alexandre Soares Rosado[2], Rommel Thiago Jucá Ramos[3,4,5*]

[1]Department of Biochemistry and Immunology, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[2]Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[3]Center of High Performance Computer and Artificial Intelligence, Federal University of Pará, Belém, Pará, Brazil
[4]Laboratory of Bioinformatics and Genomics of Microorganisms, Federal University of Pará-UFPA, Belém, Pará, Brazil
[5]Laboratory of Simulation and Computational Biology (SIMBIC), Federal University of Pará, Belém, Pará, Brazil
[6]Universidade Tecnológica de Santiago, Santo Domingo, Republic Dominican.
[7]Vale Institute of Technology – Sustainable Development, Belém, Pará, Brazil
[8]Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil

*** Correspondence:**
Rommel Thiago Jucá Ramos
rommelthiago@gmail.com

Keywords: Metagenomics, Antimicrobial Resistance (AMR), Resistome, Freshwater Microbiomes, Amazon Basin, Waterborne Pathogens, One Health.

Abstract

Background: Antimicrobial resistance (AMR) is a global health crisis increasingly linked to environmental reservoirs, yet the spatial dynamics of resistance in biodiverse and understudied ecosystems remain poorly characterized.

Methods: This study employed high-throughput metagenomic sequencing to investigate microbial communities and antimicrobial resistance profiles across 20 freshwater sites in four municipalities of the Brazilian Amazon - Altamira, Bragança, Marabá, and Paragominas. Samples were collected in triplicate in five different sites in each municipality. Ecological and demographic metadata were incorporated into a multidimensional analysis encompassing microbial diversity, resistance gene abundance, and a risk-based ranking of antimicrobial resistance genes (ARGs).

Results: Taxonomic classification revealed diverse and spatially structured microbial communities, with both clinically relevant pathogens and ecologically significant taxa identified across sites. ARGs were detected in all samples, including those from remote or low-impact areas. Bragança and Marabá emerged as critical AMR hotspots, exhibiting high ARG loads and elevated threat scores, particularly at sites with dense urbanization or severe deforestation. In contrast, low-resistance profiles were observed at semi-pristine locations in Altamira. A novel ARG Threat Score and 3D landscape visualization framework was developed to integrate abundance, risk, and ecological gradients, allowing the identification of sites with disproportionate resistance potential.

Conclusions: This study provides the first metagenomic risk landscape of AMR across Amazonian freshwater systems, revealing that ARGs are both widespread and ecologically contextual. By framing resistance through spatial and ecological lenses, this work establishes a flexible strategy for environmental AMR surveillance and highlights the Amazon as a critical frontier for One Health-focused resistance monitoring.

Resistome of Amazonian Freshwaters

1. Introduction

Microbial resistance long precedes modern medicine. Before the first antibiotics were discovered, microbes were already engaged in an ancient arms race, competing, adapting, and evolving molecular strategies to suppress one another [1]. In this context, antimicrobial resistance (AMR) is not just a clinical issue but a fundamental aspect of microbial ecology. What transforms this biological phenomenon into a global health crisis is not its existence but its acceleration, driven by human activity, environmental disruption, and the breakdown of ecological balance [2].

The Amazon Basin is home to Earth's most diverse freshwater ecosystems, a reservoir of microbial novelty, complexity, and ecological balance [3]. Its vast mosaic of rivers, floodplains, and dense vegetation harbors biodiversity and supports microbial networks, which are essential to environmental resilience [4]. However, this balance is increasingly at risk. Decades of deforestation, agro-industrial expansion, hydroelectric projects, and unregulated urban growth have dramatically reshaped these landscapes, altering their physical contours and the microbial communities they sustain [5]. AMR may no longer follow its natural evolutionary rhythms in such shifting environments. Instead, it risks being amplified by ecological disruption and maintained by human-driven selective pressure [6].

While AMR has long been framed through a clinical lens, it is now increasingly understood as a complex, ecological phenomenon shaped by the intersections of environmental degradation, social inequality, and pathogen evolution [7]. Surveillance efforts have begun to extend beyond hospitals and wastewater plants to encompass rivers, lakes, and soils [8], [9], [10]. Nevertheless, environmental AMR remains significantly under-characterized in tropical biodiversity regions, particularly in the Amazon, where microbial ecosystems are both rich in untapped genetic potential and vulnerable to anthropogenic stress.

To address this knowledge gap, the present study applies a metagenomic approach to explore the structure of microbial communities and the distribution of antimicrobial resistance genes (ARGs) across freshwater sites in four municipalities of the Brazilian Amazon. The resistome, defined as the collection of all ARGs in a microbial community, is examined in spatial context by integrating environmental metadata, taxonomic profiles, and risk-based ARG classification. This work provides a foundation for understanding how resistance is embedded within dynamic freshwater ecosystems, informing future efforts in environmental AMR surveillance and resilience-based One Health strategies.

2. Results

2.1. *High-Quality Metagenomic Data from a Complex Amazonian Landscape*

The samples were initially collected in triplicate from 20 sites distributed across four municipalities in the Brazilian state of Pará, each representing distinct aquatic environments influenced by varying degrees of anthropogenic activity. DNA was extracted following suitable protocols for environmental samples, and libraries were constructed from samples that met quality and concentration thresholds.

Shotgun metagenomic sequencing produced an average of 71.8 million reads per sample (range: 59.7–91.6 million), with a high rate of quality retention across all libraries. After quality control (QC) filtering, host DNA removal, and duplicate read filtering, a majority of reads were retained for downstream analyses, with final usable read counts averaging 62.0 million reads per sample (range: 52.2–79.1 million) and a mean QC retention rate of ~90.9% (Table 1). The final dataset comprises high-quality environmental DNA libraries suitable for metagenomic profiling and comparative analysis.

Table 1. Summary of sequencing output and quality control metrics for all environmental DNA samples. Summary of sequencing and preprocessing metrics for the 58 environmental water samples analyzed in this study. Values include total reads, quality control retention, and remaining reads after QC filtering, host DNA removal, and duplicate read removal. Metadata columns indicate sample type, nucleotide type, collection date, and location. The ‘Sample type’ and ‘Nucleotide type’ for all of the samples was Environmental and DNA, respectively.

Each sample was annotated with environmental and contextual metadata, including geographic location, collection date, sample type, key physicochemical parameters (e.g., pH, temperature, and pressure), and municipality-level socioeconomic indicators. These variables were used as covariates in subsequent analyses. The metagenomic metadata adopted in this study is detailed at the sample, sampling site, and city levels in Supplementary Tables 1–3, respectively.

2.2. *Microbial Communities Mirror Environmental Gradients Across the Amazon*

The taxonomic composition of microbial communities across Amazonian freshwater samples revealed high diversity both within and between cities. Figure 1A introduces the sampling landscape, showing the geographic distribution of the 20 sites and proportional bubble charts representing the relative abundance and richness of bacterial families per

municipality. Taxonomic classification of the samples, at family, genus, and species levels, enabled further examination of microbial profiles.

A global Sankey diagram (Figure 1B) summarizes the aggregated taxonomic distribution, highlighting the dominance of the domain *Bacteria*, particularly across the phyla *Pseudomonadota*, *Actinomycetota*, *Cyanobacteriota*, and *Bacteroidota*. This reflects the study's focus on prokaryotic diversity, as preprocessing steps removed low-quality and human-associated reads to improve ARG profiling resolution. Summary data, including microbial composition and filtering statistics, are in Supplementary Table 4.

Figure 1. Geographic distribution and taxonomic profiles of Amazonian freshwater microbiomes. (A) Regional map of the 20 freshwater sampling sites across four municipalities in the state of Pará, Brazil. Colored polygons indicate municipal boundaries; bubble charts display bacterial family-level abundance per city, scaled by relative abundance and sequencing depth. (B) Sankey diagram showing the cumulative taxonomic classification of all 58 metagenomic samples across hierarchical levels from domain (D) to species (S). The majority of classified reads belong to the domain *Bacteria*, spanning diverse phyla and several clinically relevant genera. Flows are scaled by read abundance. (C) Attribute column charts of bacterial communities stratified by percentage of regional deforestation. Left (I): relative abundance by family; Right (II): absolute abundance. Bottom panel lists the top-ranked taxonomic lineages by cumulative sequence count. These patterns illustrate the influence of environmental disturbance on microbial structure and dominance.

Several bacterial families with known environmental and clinical relevance were detected, including *Comamonadaceae*, *Enterobacteriaceae*, and *Burkholderiaceae*, alongside notable genera such as *Acinetobacter*, *Klebsiella*, *Aeromonas*, *Pseudomonas*, and *Escherichia*. Cyanobacterial genera, especially *Synechococcus* and *Cyanobium*, were also detected across multiple sites. The classification flow highlights the diversity captured through metagenomic profiling. City-level Sankey diagrams (Supplementary Figure 1, panels A–D) show differences in dominant lineages, including the relative abundance of *Enterobacterales*, *Burkholderiales*, and *Synechococcales*. These diagrams emphasize site-specific microbial structures that reflect the environmental and anthropogenic heterogeneity of the region.

To explore further environmental influences, Figure 1C presents dominant bacterial families stratified by regional deforestation. *Comamonadaceae* and *Burkholderiaceae* remained consistently abundant, suggesting adaptability to varied conditions. In contrast, families like *Pseudomonadaceae* and *Aeromonadaceae* were more prominent in highly deforested areas, suggesting potential associations with anthropogenic disturbance. Less

disturbed regions exhibited more even and diverse profiles, lacking dominance by any single lineage. Supplementary Figure 2 provides additional stratifications based on physicochemical and socioeconomic variables, offering deeper ecological context to these taxonomic trends.

2.3. *Local Structure Within a Diverse System: Richness, Beta Diversity, and Community Separation*

Alpha diversity indices revealed consistently high values across Amazonian freshwater samples (Table 2). Shannon diversity ranged from 7.91 to 10.59 (mean = 9.92), with low variation, indicating rich and stable microbial communities. Simpson evenness scores were uniformly high (mean = 0.9999), suggesting balanced communities without dominance by any single taxon. These trends were consistent across all municipalities, underscoring the ecological complexity of Amazonian microbial assemblages.

Statistical comparisons across cities using the Kruskal-Wallis test revealed significant differences in microbial diversity metrics, including Shannon diversity ($p < 0.001$), Simpson evenness ($p < 0.001$), and the metagenomic diversity (MD) index ($p < 0.001$), indicating that microbial community diversity varied significantly among municipalities despite overall high diversity across sites. These municipality-level bacterial diversity patterns are shown in Supplementary Figure 3, and the full statistical results, including Dunn's post hoc comparisons, are provided in Supplementary Table 5.

Table 2. Summary of alpha diversity metrics per sampling site. Median values and standard deviations of alpha diversity indices were calculated from triplicate samples at each of the 20 freshwater sampling sites. Metrics include total predicted gene content, protein richness, Shannon diversity, Simpson evenness, protein dissimilarity, and metagenomic diversity (MD). These indices capture taxonomic and functional diversity and assess microbial community richness and ecological balance across sampling locations.

In addition to taxonomic diversity, functional potential was assessed through the total number of predicted genes and protein richness. Although these metrics showed greater variability across sampling sites, as expected in environmental metagenomes, this heterogeneity reflects the natural variation in microbial community complexity and metabolic capacity among distinct aquatic environments. The metagenomic diversity (MD) index further supported this interpretation. MD values, ranging from 0.76 to 0.98, revealed substantial variation in the functional dissimilarity of protein-encoding genes across sites. Sites with higher MD scores, such as BP1 (Bragança) and MP1 (Marabá), exhibited a

broader array of non-redundant proteins, suggesting greater ecological flexibility and functional breadth. In contrast, sites like PP1, PP2, and PP3 (Paragominas) showed lower MD values, indicating greater functional redundancy and possibly reflecting stronger environmental filtering. These observations align with the total gene and protein richness patterns and suggest that distinct local pressures, including hydrological conditions, pollution gradients, or human influence, may shape communities in specific locations.

At the city-level, Altamira and Bragança showed greater variability in MD and functional indices, suggesting more heterogeneous aquatic environments. Conversely, Paragominas and Marabá displayed more uniform diversity metrics, possibly indicating shared environmental drivers or anthropogenic pressures. These findings highlight the ecological richness and spatial variability of microbial life in Amazonian freshwaters. Full metrics are provided in Supplementary Table 6.

Beta diversity analyses revealed distinct community structures across cities and sampling sites. These compositional differences were further supported by ANOSIM analysis ($R = 0.36$, $p = 0.001$), confirming significant but partially overlapping microbial community structuring among municipalities. Principal Coordinates Analysis (PCoA; Figure 2), based on Bray-Curtis dissimilarity, showed tight clustering in Bragança and Marabá, indicating consistent community composition, while Altamira and Paragominas exhibited more dispersed profiles, suggesting greater intra-city variability.

Figure 2. Principal Coordinates Analysis (PCoA) of microbial community composition across all samples. PCoA plot based on Bray-Curtis dissimilarity matrix of all 58 samples, colored by municipality. Axis 1 and 2 explain 23.96% and 22.57% of the variation, respectively. Ellipses represent 95% confidence intervals for each city. The plot illustrates both city-level clustering and intra-city variation in microbial community composition.

Compositional differences were statistically supported by PERMANOVA ($R^2 = 0.407$, $F = 12.361$, $p < 0.001$), confirming that sampling location significantly influenced community structure (Supplementary Table 7). Beta dispersion analyses (Supplementary Figure 4) revealed higher internal variability in Altamira and Paragominas, while Bragança and Marabá showed more compact clustering. These patterns were confirmed by PERMDISP ($F = 3.51$, $p = 0.025$), indicating significant multivariate dispersion differences (Supplementary Table 7).

Together, alpha and beta diversity analyses revealed that Amazonian microbial communities are both functionally rich and spatially structured. High within-site diversity

and clear between-city differentiation reflect localized ecological dynamics and provide a solid baseline for assessing the distribution and drivers of antimicrobial resistance.

2.4. *Pathogens Detected Across the Region — But They Do Not Explain AMR Alone*

The taxonomic profiling of potentially pathogenic bacteria revealed distinct patterns of occurrence and abundance across sites and municipalities (Figure 3). A heatmap of the ten most abundant pathogenic species per sample (reads per million, RPM) highlights microbial putative hotspots, particularly in Bragança (BP4, BP5), Marabá (MP1, MP3), and Paragominas (PP2, PP3, PP5). In contrast, Altamira samples consistently exhibited lower pathogen-associated abundance.

Figure 3. Heatmap of bacterial species with known pathogenic potential across metagenomic samples. Heatmap displays the ten most abundant pathogenic species per sample (reads per million, RPM), clustered by taxon similarity. Several clinically relevant taxa – including *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, *Enterobacter cloacae*, and *Stenotrophomonas maltophilia* – were detected, particularly in samples from Marabá, Bragança, and Paragominas. White cells represent taxa not detected in the corresponding sample. Filtering thresholds applied for inclusion were: nucleotide rPM ≥ 10, protein rPM ≥ 1, and alignment length ≥ 50 bp.

Several well-known clinically relevant species were detected, including core members of the ESKAPE group, *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, *Enterobacter cloacae*, and *Klebsiella pneumoniae*, as well as *Stenotrophomonas maltophilia*, a multidrug-resistant opportunist frequently associated with hospital settings. Notably, *K. pneumoniae* appeared enriched in BP4, BP5, and MP3. Other notable findings included *Burkholderia cepacia* and *Ralstonia pickettii*, opportunists that appeared across all cities. Waterborne pathogens such as *Aeromonas hydrophila* and *A. veronii* were also broadly distributed, especially in Bragança and Paragominas. Environmental taxa with pathogenic potential, including non-tuberculous mycobacteria (NTM) *Mycobacterium chelonae* and *M. intracellulare*, *Nocardia spp.*, and *Legionella pneumophila*, were also present.

Detecting this broad and clinically significant pathogenic spectrum, alongside marked differences in their distribution across sites, supports the hypothesis of localized anthropogenic influence on microbial profiles. However, as non-pathogenic or environmentally dominant bacteria may also carry antimicrobial resistance genes (ARGs), a

more comprehensive examination of the total AMR burden across samples was conducted to assess resistance potential beyond known pathogens better.

Heatmaps of log-transformed ARG genes and drug classes abundance (Supplementary Figures 5 and 6) showed widespread detection of clinically relevant genes (e.g., *tet*, *sul*, *qac*, *van*) and resistance classes (e.g., glycopeptides, fluoroquinolones, cephalosporins, aminoglycosides, disinfectants), with notable enrichment in Bragança. These findings emphasize the importance of assessing AMR risk beyond known pathogens, capturing broader environmental reservoirs of resistance.

2.5. *Antimicrobial Resistance Genes Are Widespread, Even in Seemingly Pristine Sites*

The composite analysis of antimicrobial resistance burden revealed pronounced differences in ARG distribution, prevalence, and load across cities and sampling sites (Figure 4). The prevalence heatmap (Panel A) showed that Bragança and Marabá had the broadest AMR class repertoires, with over 10 classes detected in more than 60% of samples. Meanwhile, bacitracin and multidrug resistance classes were consistently present across all cities, suggesting a potential baseline resistome characteristic of Amazonian aquatic ecosystems.

Figure 4. Composite analysis of antimicrobial resistance (AMR) burden across sampling sites and cities. (A) Heatmap of AMR class prevalence across the four cities expressed as a percentage of samples in which each class was detected. (B) Relative abundance of ARG types per sampling site, grouped by city. (C) Boxplot of normalized ARG copies per estimated number of cells (log_{10} scale), grouped by city. (D) Radar plots show the mean ARG load per sample within each city. Together, these panels reveal localized ARG richness and load patterns, with Bragança and Paragominas emerging as notable putative hotspots for AMR diversity and intensity.

At the site level, Bragança's BP4 and BP5 exhibited high ARG diversity, while Paragominas sites showed elevated loads dominated by β-lactam and sulfonamide classes. BP4 stood out as the clearest compositional hotspot. MP3 in Marabá also displayed a high ARG burden, surpassing many Paragominas sites, indicating that localized enrichment can occur even within moderate-profile cities. These trends were corroborated by absolute ARG abundance (Supplementary Figure 7), where BP4, BP5, and MP3 were the most heavily burdened. ARG alpha diversity patterns across municipalities are further detailed in Supplementary Figure 8. Differences in ARG composition among municipalities were also explored using NMDS ordination, which revealed partial clustering with substantial overlap

among cities (Supplementary Figure 9), consistent with the moderate but significant separation detected by ANOSIM (Supplementary Table 8).

Boxplots of normalized ARG copies (Panel C) revealed uniformly high burden in Paragominas, while Bragança showed more variability, from low to extreme outliers. These patterns suggest contrasting anthropogenic pressures: chronic or diffuse in Paragominas versus episodic or point-source in Bragança. Radar plots (Panel D) consolidated these profiles, highlighting Bragança's outlier status, Paragominas's elevated baseline, and Altamira's uniformly low burden.

Differences in ARG load among municipalities were statistically evaluated using the Kruskal-Wallis test, which revealed significant variation across cities ($\chi^2 = 19.86$, $p < 0.001$). Post hoc Dunn's tests indicated that Bragança exhibited significantly higher ARG loads compared to Altamira ($p = 0.001$) and Marabá ($p < 0.001$), supporting the identification of Bragança as a major resistance hotspot in the region. These city-level ARG load distributions are shown in Supplementary Figure 10, and the corresponding statistical analyses are summarized in Supplementary Table 8.

Several sites, including Altamira's AP1 to AP5 and Marabá's MP1 and MP2, showed consistently low ARG load. While not necessarily pristine, these areas may experience minimal anthropogenic influence or benefit from hydrological conditions that dilute resistance accumulation. As such, they serve as valuable ecological reference points. Altogether, these results underscore the importance of analyzing both aggregate and site-level AMR trends to identify key putative hotspots and potential baselines.

2.6. *High-Risk Resistance Hotspots Are Concentrated in Bragança and Marabá*

To further characterize the public health relevance of the identified resistance genes, ARGs were classified into four hierarchical risk ranks based on their association with human hosts, mobility potential, and clinical relevance. Rank I includes mobile ARGs already detected in pathogens. Rank II covers mobile ARGs not yet found in pathogens but with potential clinical relevance. Ranks III and IV include genes with limited mobility or no known association with human hosts. A composite figure (Figure 5) displays total ARG abundance per site (top) and the proportional distribution of ARG risk ranks (bottom). The Threat Score, indicated by color gradient, was calculated by weighting the relative abundance of Rank I and II ARGs, capturing clinically relevant resistance potential.

Figure 5. ARG Risk Ranking Across Sampling Sites. ARG Risk profiling across Amazonian freshwater sampling sites. (Top) Bubble plot representing the total ARG abundance per sampling site, color-scaled by Threat Score log10 (ThreatScore+ε), where ε=1e−6. Bubble size corresponds to total ARG abundance normalized by estimated cell count. (Bottom) Stacked bar chart showing the proportional distribution of ARGs per site by risk rank, from Rank I (highest risk: mobile and detected in pathogens) to Rank IV (lowest risk: not human-associated). Threat Scores were calculated by summing the relative abundance of Rank I and Rank II ARGs per sample, weighted by total ARG abundance: Threat Score = $(RA_{(Rank\ I)} + RA_{(Rank\ II)}) \times$ Total ARGs. This approach emphasizes clinically relevant resistance potential while retaining quantitative resolution across sites.

Distinct patterns were observed across sampling sites. The concentration of high-risk ARGs in specific locations was consistent with the observed differences in overall ARG burden, reinforcing the role of localized environmental pressures in shaping resistance risk landscapes. Bragança's BP4 site demonstrated the highest combined ARG load and Threat Score, followed by BP5 and MP3 in Marabá, which also exhibited elevated risk indicators. In contrast, sites such as AP1, AP2 (Altamira), and MP2 (Marabá) presented low overall ARG abundance and negligible contributions from high-risk ARGs, supporting their classification as potentially pristine AMR reference points. While several Paragominas sites displayed moderate to elevated risk profiles, none exceeded the threat level observed for MP3, reinforcing the importance of site-specific analysis within broader city trends. All calculated values used in the risk ranking, including total abundance, ARG risk rank distributions, and Threat Scores, are provided in Supplementary Table 9.

To visualize the intersection of microbial resistance burden and environmental context, three-dimensional surface plots were generated (Figure 6). These landscapes integrate total ARG abundance and Threat Scores with (i) the inverted Metagenomic Diversity (MD) index, where higher values reflect lower diversity and presumed more significant ecological disruption, and (ii) regional deforestation percentages at the municipality level. In both models, BP4, BP5, and MP3 consistently formed prominent topographical peaks, highlighting their convergence of high ARG abundance, elevated risk classification, and environmental stressors. Consistent with this pattern, correlation analyses showed that ARG load was positively associated with population density and deforestation, and negatively associated with dissolved oxygen (Supplementary Figure 11). This multi-layered visualization underscores the importance of ecological and anthropogenic variables in shaping the spatial distribution of AMR potential across Amazonian freshwater systems.

Resistome of Amazonian Freshwaters

Figure 6. 3D ARG Landscapes Across Amazonian Freshwater Samples. Three-dimensional surface models integrating ARG abundance and Threat Score with ecological disturbance proxies. (Left) Landscape shaped by the inverted Metagenomic Diversity (MD) index, where higher values indicate lower within-sample diversity and presumed ecological imbalance. (Right) Landscape shaped by municipality-level deforestation percentage, obtained from publicly available environmental datasets. For both models, the z-axis represents ARG abundance log10(ARGs+1), and color gradients reflect Threat Score magnitude log10(ThreatScore+ε). Peaks correspond to sampling sites with elevated resistance potential under combined biological and anthropogenic pressure.

Altogether, our findings reveal a multifaceted picture of microbial and resistance dynamics. While diversity remained high across sites, compositional and functional patterns were spatially structured by environmental pressures. ARG abundance, risk, and distribution varied locally, identifying both resistance putative hotspots and low-impact baselines. The integration of risk classification, ARG burden, and ecological proxies offers a powerful framework to pinpoint critical areas for AMR emergence, and underscores the critical need to incorporate environmental AMR surveillance into broader public health and One Health frameworks.

3. Discussion

The Amazon Basin harbors one of Earth's richest ecosystems, yet its microbial diversity remains underexplored [11]. Recent metagenomic efforts have begun to map microbial life across Amazonian soils [12], sediments [13], and waters [14], underscoring high local heterogeneity. Our findings reinforce this pattern, revealing diverse and ecologically structured microbial communities shaped by environmental gradients.

Environmental factors such as deforestation and other environmental factors were associated with microbial composition in this study. Sites exhibiting high levels of forest loss tended to host distinct taxonomic profiles compared to more preserved regions, supporting the hypothesis that ecological filtering plays a strong role in shaping microbial community dynamics in these systems. While these associations do not imply causation, they align with emerging literature suggesting that anthropogenic disturbance can influence not only species presence but microbial functionality and resilience as well [15], [16]. In parallel, ARG load was positively correlated with deforestation and population density in this study, reinforcing the link between anthropogenic disturbance and resistance burden observed across sampling sites (Supplementary Figure 11). In this context, our data offer a timely addition to ongoing efforts to characterize the microbial complexity of the Amazon,

particularly as environmental degradation and climate instability continue to accelerate across the region [17].

Beyond ecological diversity, taxonomic analyses revealed the presence of several bacterial species with known pathogenic potential across all 58 freshwater metagenomes, including members of the high-priority ESKAPE group such as *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, and *Enterobacter cloacae*. Detecting these taxa, especially in surface water bodies used for fishing, recreation, or local water supplies, signals potential exposure risks to nearby communities, many of which are geographically isolated and face well-documented challenges in accessing public health infrastructure [18]. In such contexts, even pathogens that remain largely treatable can pose substantial health threats due to delayed or insufficient clinical response and limited availability of appropriate antimicrobials [19].

Beyond human health risks, the taxonomic profiles also included environmental pathogens and opportunists that play dual ecological and clinical roles. Remarkably, cyanobacterial genera such as *Synechococcus* and *Cyanobium* were prevalent across sites. While not typically associated with infectious disease, these taxa are known for their capacity to form harmful algal blooms and produce cyanotoxins, particularly in tropical freshwater systems [20], [21]. The co-occurrence of toxin-producing organisms and potentially pathogenic bacteria reinforces the multifaceted threat posed by altered microbial communities, where human health, aquatic biodiversity, and ecosystem function are simultaneously at risk [22].

While many pathogens identified remain treatable with current therapeutic strategies, several belong to genera with well-documented histories of acquiring antimicrobial resistance. Species such as *Escherichia coli*, *Pseudomonas putida*, *Enterobacter hormaechei*, and *Aeromonas caviae*, all widely detected, have been implicated in the environmental dissemination of ARGs, particularly in aquatic systems impacted by high organic load or industrial discharge [23], [24], [25], [26]. These opportunistic pathogens are known for their ability to integrate and disseminate resistance determinants through mobile genetic elements, including plasmids and integrons [27]. Their widespread detection in Amazonian freshwater reinforces concerns over their potential role in shaping local resistance profiles, particularly where surveillance and mitigation are limited.

Resistance is an ancient microbial trait, shaped by evolutionary arms races long before clinical antibiotics. In aquatic systems, low-abundance background ARGs likely reflect these natural dynamics. However, environmental stressors such as pollution and

deforestation can increase selective pressure, disrupt microbial communities, and accelerate ARG mobilization into clinically relevant contexts [28]. Thus, detecting ARGs in semi-pristine sites is not surprising. These genes may represent ancestral resistomes maintained by natural selection. Yet, their comparison to urbanized or agro-industrial zones, where ARG abundance and mobility are higher, underscores how human pressures amplify risk. Sites like AP1, AP2 (Altamira), and MP2 (Marabá) offer ecological baselines against which such enrichment can be evaluated.

Altamira's AP1–AP2 sites, located in less disturbed areas of the Xingu River basin, exhibited low ARG load and minimal high-risk ARGs. MP2, situated upstream from Marabá's urban core, showed similar trends. These patterns likely result from limited direct pollution and intact riparian vegetation buffers that reduce selective pressure. In contrast, Bragança's BP4 site emerged as a resistance epicenter, exhibiting the highest ARG load and Threat Score, driven by abundant Rank I ARGs. MP3 in Marabá and multiple Paragominas sites also revealed high ARG burdens and risk signatures, despite differing in population density and land-use context. While Bragança's hotspots may be linked to untreated sewage and dense settlement, Paragominas's elevated ARG levels likely stem from widespread agro-industrial activity and deforestation.

These observations suggest a dual pathway for environmental AMR amplification: concentrated urban discharge and diffuse agricultural runoff. Both converge on microbial communities, increasing the likelihood of ARG acquisition and dissemination. Integrating deforestation maps, population data, and ARG metrics provides a multidimensional lens for tracking this phenomenon. Interestingly, despite similar deforestation rates, Altamira did not present the same ARG risk levels. This reinforces the idea that AMR emergence reflects more than just land cover, it is shaped by interactions among microbial load, ecological disruption, and exposure to selective agents like antibiotics, metals, and waste.

Ultimately, ARGs were detected in all sites, reaffirming their widespread ecological presence. However, high-risk ARGs clustered in specific locations with clear anthropogenic signatures. This gradient, from natural reservoirs to amplified threats, frames AMR not just as a clinical crisis but as an ecological indicator of system imbalance.

Recognizing this duality is crucial. Identifying both putative hotspots and baselines allows for better risk stratification and early intervention. Environmental AMR surveillance must adopt such contextualized frameworks, especially in biodiverse and rapidly changing regions like the Amazon. Our study contributes to this effort by mapping resistance not just in terms of presence but ecological relevance.

Resistome of Amazonian Freshwaters

To translate these patterns into a surveillance-oriented framework, we introduced a composite ARG Threat Score that integrates resistance gene abundance with established risk ranking criteria. Rather than relying solely on total ARG counts, which may mask differences in clinical relevance, this metric emphasizes genes classified as higher risk due to their association with mobility and occurrence in pathogenic hosts. By weighting the relative abundance of Rank I and Rank II ARGs, those most closely linked to clinically relevant resistance, the Threat Score highlights sites where environmentally detected resistomes may carry greater potential for downstream public health impact.

Importantly, this metric should be interpreted as a comparative ecological indicator rather than a direct measure of clinical risk. Its primary function is to facilitate spatial prioritization in environmental AMR surveillance, particularly in regions where systematic monitoring resources remain limited. Although future studies integrating longitudinal sampling and functional validation may further refine this framework, the approach presented here provides a pragmatic strategy for identifying locations where resistance reservoirs warrant closer ecological and epidemiological attention. While these findings provide new insights into the ecological distribution and risk potential of antimicrobial resistance in Amazonian freshwater systems, several limitations should be considered when interpreting the results.

Limitations and Scope of the Study

Despite providing a comprehensive metagenomic assessment of antimicrobial resistance across multiple Amazonian freshwater systems, several limitations inherent to environmental surveillance studies should be considered when interpreting the results.

First, the spatial coverage of the sampling design represents only a fraction of the ecological and hydrological diversity of the Amazon basin. Although the study included twenty sites across four municipalities spanning gradients of urbanization and deforestation, the Amazon encompasses vast and heterogeneous freshwater systems influenced by diverse land-use patterns, hydrological regimes, and anthropogenic pressures. As such, the patterns described here should be interpreted as a regional snapshot rather than a basin-wide characterization of resistome dynamics. Nevertheless, the multi-site design and triplicate sampling strategy provide a valuable baseline for understanding antimicrobial resistance distribution within previously underexplored freshwater environments.

Second, sampling was conducted within a single hydrological window (September to November), which may not capture seasonal fluctuations in microbial community structure

or antimicrobial resistance gene (ARG) dynamics. Amazonian aquatic ecosystems undergo pronounced seasonal changes driven by rainfall cycles, river discharge, and sediment transport, all of which can influence microbial composition and gene distribution. Longitudinal monitoring across wet and dry seasons would therefore be valuable for evaluating temporal stability and seasonal drivers of resistome patterns.

Third, the ecological associations identified in this study, including relationships between ARG abundance, microbial diversity metrics, and environmental gradients such as deforestation, should be interpreted as correlations rather than direct causal relationships. While spatial patterns suggest links between anthropogenic disturbance and elevated resistance burdens, the observational nature of the dataset does not allow causal attribution of specific drivers. Future work integrating controlled environmental measurements, longitudinal sampling, and mechanistic modeling would help clarify these relationships.

Fourth, the metagenomic approach employed here enables sensitive detection of ARG presence across microbial communities but does not directly measure gene expression or functional activity. The detection of ARG sequences therefore reflects the genetic potential for resistance rather than confirmed phenotypic resistance or active gene expression. Complementary approaches such as metatranscriptomics, functional assays, or culture-based validation could further refine the interpretation of resistance activity within these ecosystems.

Finally, the ARG Threat Score proposed in this study should be interpreted as a comparative ecological risk proxy rather than a direct measure of clinical risk. The metric was designed to integrate ARG abundance with established risk classifications to highlight sites with elevated resistance potential. While this framework facilitates spatial prioritization of monitoring efforts, its predictive capacity would benefit from future validation using additional environmental datasets and epidemiological context.

Taken together, these limitations reflect the inherent complexity of environmental resistome surveillance but do not diminish the value of the dataset presented. Instead, they highlight the importance of expanding systematic monitoring efforts across diverse freshwater systems and emphasize the need for integrative approaches linking environmental, ecological, and public health perspectives within a One Health framework.

4. Conclusion

This study offers a novel framework for contextualizing antimicrobial resistance (AMR) within the ecological fabric of Amazonian freshwater systems. Through

high-resolution metagenomic surveillance coupled with environmental and spatial indicators, we reveal the widespread presence of resistance genes and the emergence of localized putative hotspots shaped by deforestation, urbanization, and disrupted ecological balance. By modeling resistance across ecological gradients, we demonstrate that AMR is a clinical challenge and a complex ecological signal influenced by land use and anthropogenic pressures. This approach provides a scalable strategy for environmental AMR monitoring, particularly vital for biodiverse regions under pressure, and reinforces the importance of integrative, One Health surveillance to inform public health and policy decisions.

5. Methods

5.1. *Study Area, Sample Collection, and Metadata Acquisition*

Surface freshwater samples were collected from four municipalities in the Brazilian state of Pará, Altamira, Bragança, Marabá, and Paragominas, between September and November 2022. Five sampling sites were selected within each municipality to represent gradients of anthropogenic influence. At each site, three biological replicates were collected at a depth of approximately 20 cm, about 1 meter from the riverbank or channel center, totaling 60 samples across 20 sites.

For each replicate, 5 L of surface water were collected using a Van Dorn bottle, stored on ice, and transported immediately to the laboratory for processing. Samples were filtered using a negative pressure vacuum system with 47 mm nitrocellulose membranes (0.22 µm pore size). Membranes were transferred to 50 mL Falcon tubes containing TSE buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA) and stored at -20°C until DNA extraction.

In situ physicochemical parameters were recorded at the time of sampling using a multiparameter water analyzer (HI 9829, HANNA Instruments), including pH, temperature, pressure, electrical conductivity, resistivity, total dissolved solids (TDS), salinity, oxidation-reduction potential (ORP), and dissolved oxygen.

Municipality-level socioeconomic indicators (including population density, GDP per capita, and infant mortality) were obtained from the Brazilian Institute of Geography and Statistics (IBGE) Cities and States Portal [29]. Deforestation levels were quantified through geospatial analysis following the methodology of Cavalcante *et al.* [30], using MapBiomas Brazil land-use and land-cover data (Collection 6, year 2020). The deforested area was

calculated as the proportion of non-natural land cover associated with each sampling site within the corresponding hydrographic basin.

Geographic coordinates for all sampling locations are provided in Supplementary Methods Table S1 (Supplementary Methods 1.1). The combined dataset, including physicochemical measurements, socioeconomic indicators, and spatial environmental context, was integrated into downstream multivariate analyses.

5.2. *DNA Extraction, Library Preparation, and Metagenomic Sequencing*

Following field filtration, filter membranes were transported to the central laboratory and incubated at room temperature under constant agitation (250 RPM) to release retained particles. The resulting suspension was centrifuged at 13,200 RPM for 10 minutes, and DNA was extracted from the resulting pellet using the DNeasy PowerSoil Pro Kit (QIAGEN) according to the manufacturer's protocol. DNA concentration was measured using a Qubit fluorometer, and purity was assessed using a NanoDrop spectrophotometer. DNA integrity was verified by agarose gel electrophoresis. Samples meeting quality thresholds ($\geq$50 ng/µL and 260/280 ratio between 1.4 and 2.0) were retained for downstream analysis ($n = 58$). Metagenomic libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina) and sequenced on the Illumina NextSeq 550 platform using the High-Output kit, generating 150 bp paired-end reads. Raw reads were delivered in FASTQ format for subsequent preprocessing and analysis.

5.3. *Quality Control and Preprocessing of Sequencing Data*

Raw sequencing reads were processed using the CZ ID metagenomics (mNGS) pipeline (Illumina mNGS Pipeline v8.3). This pipeline performs host filtering, adapter trimming, quality filtering, low-complexity read removal, and human read depletion, being fully described by Kalantar *et al.* [31]. The resulting high-quality non-host reads were retained for all downstream taxonomic and functional analyses. Detailed pipeline parameters and reference alignment steps are described in Supplementary Methods 1.2.

5.4. *Taxonomic Classification and Microbial Diversity Analysis*

Taxonomic classification was performed using Kraken2 and Bracken on the Galaxy Europe platform [32], [33], [34], aligning paired-end reads against the RefSeq PlusPF database (version 2024-01-12) with a confidence score threshold of 0.1. Species-level

abundance estimates were visualized using Pavian [35] and Phinch [36], incorporating environmental metadata for comparative analysis. Alpha diversity metrics, including species richness, Shannon index, Simpson index, and metagenomic diversity, were calculated using the MD.py script, which estimates diversity based on clustering of predicted protein sequences [37]. Assemblies and protein predictions required for MD index calculation were generated using MEGAHIT [38] and Prodigal [39], respectively. Beta diversity was evaluated using Bray-Curtis dissimilarity, with ordination performed using principal coordinates analysis (PCoA). Group differences were tested using PERMANOVA and PERMDISP implemented in the vegan package (v2.6-10) [40]. Additional workflow details are provided in Supplementary Methods 1.3.

5.5. *Detection and Quantification of Antimicrobial Resistance Genes (ARGs)*

Antimicrobial resistance genes (ARGs) were characterized using two complementary analytical approaches. First, ARG detection was performed using the CZ ID AMR pipeline (v1.4.2), which identifies resistance genes across both quality-filtered reads and assembled contigs using the Resistance Gene Identifier (RGI v6.0.0) and the CARD (v3.2.6) and WildCARD (v4.0.0) databases [41]. These results were used to generate ARG and drug-class heatmaps and pathogen association profiles. In parallel, quantitative resistome profiling was conducted using ARGs-OAP (v3.2.4), which estimates ARG abundance normalized by microbial cell counts based on alignments to the SARG database (v3.0) [42]. Microbial cell numbers were inferred from 16S rRNA hypervariable region alignments, with copy-number correction performed using the Greengenes and CopyRighter databases [43], [44]. Normalized ARG copy numbers were used to generate abundance-based visualizations including stacked bar plots, boxplots, radar charts, and prevalence heatmaps. The use of two independent pipelines enabled cross-validation of ARG detection and improved robustness of resistome characterization. Full details of the analytical pipeline and normalization procedures are provided in Supplementary Methods 1.4.

5.6. *ARG Risk Ranking and Threat Score Calculation*

To evaluate the potential public health relevance of detected ARGs, resistance genes were classified using ARG_ranker (v3.7.2) [45]. This framework prioritizes ARGs according to mobility potential, occurrence in human-associated environments, and presence in known pathogens. Based on these criteria, ARGs are assigned to four ranks.

Rank I includes high-risk ARGs that are both mobile and present in human pathogens. Rank II includes genes with strong potential for horizontal transfer and clinical emergence. Rank III comprises mobile ARGs not enriched in human-associated environments, while Rank IV includes non-mobile ARGs with limited clinical relevance. A Threat Score was calculated for each sample to summarize the potential clinical risk of the resistome. This metric was used to identify sites with elevated concentrations of clinically relevant and mobile ARGs. This score combines the relative abundance of Rank I and Rank II ARGs, those considered most relevant to public health, and multiplies this value by the total normalized ARG abundance. The resulting values were log-transformed to generate a continuous risk metric for each sample. The full formula and implementation details are provided in Supplementary Methods 1.5.

5.7. *Spatial Modeling and 3D ARG Landscapes*

Three-dimensional ARG landscape models were constructed to visualize resistance trends across environmental gradients. These models integrate ARG abundance and Threat Score values across two ecological axes: the Metagenomic Diversity Index (MD index) and the percentage of regional deforestation surrounding each sampling site. To improve visualization, MD index values were inverted so that lower microbial diversity appears as higher values on the y-axis, aligning visually with ecological disturbance. This transformation allows resistance putative hotspots to appear as elevated surface features in the landscape model. Full modeling procedures, including interpolation methods and code references, are described in Supplementary Methods 1.6.

5.8. *Statistical Analysis and Visualization Framework*

All statistical analyses were performed in R (v4.4.1) using packages including vegan (v2.6-10) and the tidyverse ecosystem. Ordination analyses of microbial community composition and ARG profiles were conducted using principal coordinates analysis (PCoA) and non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity matrices. Microbial alpha diversity metrics were calculated using the *MD.py* workflow based on protein-coding gene clustering, which generates multiple diversity indices including species richness, Shannon diversity, Simpson evenness, and metagenomic diversity, the MD index [37]. For statistical comparisons and visualizations, Shannon diversity indices derived from taxonomic abundance tables were used. ARG alpha diversity

was calculated using Shannon indices derived from normalized ARG abundance matrices. Differences in ARG abundance among municipalities were evaluated using the Kruskal-Wallis test, followed by Dunn's post hoc pairwise comparisons where significant differences were detected. Associations between ARG abundance, microbial diversity, and environmental variables, including deforestation levels, were evaluated using Spearman rank correlation analyses. AMR heatmaps were generated using the Make_AMR_Heatmap.ipynb workflow from the CZ ID GitHub repository (https://github.com/chanzuckerberg/czid-amr-heatmap), using the platform's Combined AMR Report as input. Additional visualizations, including taxonomic bar plots, radar charts, boxplots, and ARG bubble plots, were generated using custom R scripts and visualization libraries including *ggplot2*, *viridis*, *fmsb*, and *pheatmap*. Sampling maps and spatial layouts were produced using QGIS (v3.40), incorporating manually collected GPS coordinates and ARG abundance data to generate composite spatial visualizations. More details are provided in Supplementary Methods.

6. Declarations

Author Contributions:

WGN performed all analyses, generated visualizations, and wrote both the draft and final versions of the manuscript. WGN, DTM, CWDD, VBCF, OVCA, EFFLS, PRMP, and RBLC contributed to data generation. JS, VACA, ASR, and RTJR provided supervision. RTJR also coordinated the project. All authors reviewed the final manuscript and approved the submitted version.

Acknowledgments:

The authors would like to thank the funding agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), including support through the CAPES/PRINT program (Edital nº 41/2017) during the first author's doctoral internship at King Abdullah University of Science and Technology (KAUST); Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA); the National Council for Scientific and Technological Development (CNPq), project #312316/2022–4; the Secretary of State for Science, Technology, and Professional and Technological Education (SECTET). The authors also acknowledge the SECTET/UFPA/FADESP partnership for their financial support of this work. This work was also produced within the scope of the National Institute of Science and Technology IAmazônia project. We thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for approving the proposal under n° 409001/2024-4.

Conflict Of Interest Statement:

The authors declare no conflicts of interest.

Ethics Statement:

Not applicable.

Data Availability Statement:

All the sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under accession number PRJNA1122411 (https://www.ncbi.nlm.nih.gov/sra/).

8. References

[1] R. Agudo and M. P. Reche, "Revealing antibiotic resistance's ancient roots: insights from pristine ecosystems," *Front Microbiol*, vol. 15, p. 1445155, Oct. 2024, doi: 10.3389/FMICB.2024.1445155/BIBTEX.

[2] Samreen, I. Ahmad, H. A. Malak, and H. H. Abulreesh, "Environmental antimicrobial resistance and its drivers: a potential threat to public health," *J Glob Antimicrob Resist*, vol. 27, pp. 101–111, Dec. 2021, doi: 10.1016/J.JGAR.2021.08.001.

[3] P. ; et al. Val, "Amazon Assessment Report 2021," *United Nations Sustainable Development Solutions Network, New York, USA*, 2021, doi: 10.55161/POFE6241.

[4] J. C. A. Baker and D. V. Spracklen, "Climate Benefits of Intact Amazon Forests and the Biophysical Consequences of Disturbance," *Frontiers in Forests and Global Change*, vol. 2, Aug. 2019, doi: 10.3389/FFGC.2019.00047.

[5] B. M. Flores *et al.*, "Critical transitions in the Amazon forest system," *Nature 2024 626:7999*, vol. 626, no. 7999, pp. 555–564, Feb. 2024, doi: 10.1038/s41586-023-06970-0.

[6] B. Aslam *et al.*, "AMR and Sustainable Development Goals: at a crossroads," *Global Health*, vol. 20, no. 1, pp. 1–21, Dec. 2024, doi: 10.1186/S12992-024-01046-8/FIGURES/13.

[7] WHO – World Health Organization, "Antimicrobial resistance: accelerating national and global responses WHO strategic and operational priorities to address drug-resistant bacterial infections in the human health sector, 2025-2035," 2024, Accessed: Apr. 15, 2025. [Online]. Available: https://www.woah.org/app/uploads/2021/03/en-amr-strategy-2022-final-

[8] L. M. Clarke, J. W. O'Brien, A. K. Murray, W. H. Gaze, and K. V. Thomas, "A review of wastewater-based epidemiology for antimicrobial resistance surveillance," *J Environ Expo Assess 2024;3:7.*, vol. 3, no. 1, p. N/A-N/A, Feb. 2024, doi: 10.20517/JEEA.2023.29.

[9] J. Bengtsson-Palme *et al.*, "Towards monitoring of antimicrobial resistance in the environment: For what reasons, how to implement it, and what are the data needs?," *Environ Int*, vol. 178, p. 108089, Aug. 2023, doi: 10.1016/J.ENVINT.2023.108089.

[10] U. N. E. Programme, "Bracing for Superbugs: Strengthening Environmental Action in the One Health Response to Antimicrobial Resistance," *Bracing for Superbugs*, Feb. 2023, doi: 10.18356/9789210025799.

Resistome of Amazonian Freshwaters

[11] A. C. F. M. de Oliveira *et al.*, “A metabologenomics approach reveals the unexplored biosynthetic potential of bacteria isolated from an Amazon Conservation Unit,” *Microbiol Spectr*, vol. 13, no. 1, Jan. 2025, doi: 10.1128/SPECTRUM.00996-24.

[12] K. B. do Carmo *et al.*, “Assessment of soil bacterial communities in integrated crop production systems within the Amazon Biome, Brazil: a comparative study,” *Braz J Microbiol*, vol. 55, no. 3, Sep. 2024, doi: 10.1007/S42770-024-01352-8.

[13] C. C. Thompson *et al.*, “Diversity of Microbiomes Across a 13,000-Year-Old Amazon Sediment,” *Microb Ecol*, vol. 86, no. 3, pp. 2202–2209, Oct. 2023, doi: 10.1007/S00248-023-02202-0.

[14] W. G. Nogueira *et al.*, “Viral Metagenomics Reveals Widely Diverse Viral Community of Freshwater Amazonian Lake,” *Front Public Health*, vol. 10, Apr. 2022, doi: 10.3389/FPUBH.2022.869886.

[15] R. E. Beattie, A. Bandla, S. Swarup, and K. R. Hristova, “Freshwater Sediment Microbial Communities Are Not Resilient to Disturbance From Agricultural Land Runoff,” *Front Microbiol*, vol. 11, p. 539921, Oct. 2020, doi: 10.3389/FMICB.2020.539921/BIBTEX.

[16] X. Zhu *et al.*, “Effects of different types of anthropogenic disturbances and natural wetlands on water quality and microbial communities in a typical black-odor river,” *Ecol Indic*, vol. 136, p. 108613, Mar. 2022, doi: 10.1016/J.ECOLIND.2022.108613.

[17] J. H. Ellwanger *et al.*, “Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health,” *An Acad Bras Cienc*, vol. 92, no. 1, p. 20191375, 2020, doi: 10.1590/0001-3765202020191375.

[18] L. Garnelo, R. C. P. Parente, M. L. R. Puchiarelli, P. C. Correia, M. V. Torres, and F. J. Herkrath, “Barriers to access and organization of primary health care services for rural riverside populations in the Amazon,” *Int J Equity Health*, vol. 19, no. 1, pp. 1–14, Jul. 2020, doi: 10.1186/S12939-020-01171-X/FIGURES/1.

[19] M. C. R. Fausto, L. Giovanella, J. G. Lima, L. M. da Silva Cabral, and H. Seidl, “Primary Health Care sustainability in rural remote territories at the fluvial Amazon: organization, strategies, and challenges,” *Cien Saude Colet*, vol. 27, no. 4, pp. 1605–1618, 2022, doi: 10.1590/1413-81232022274.01112021.

[20] K. Y. H. Gin *et al.*, “Novel cyanotoxin-producing Synechococcus in tropical lakes,” *Water Res*, vol. 192, Mar. 2021, doi: 10.1016/J.WATRES.2021.116828.

[21] Z. Y. Sim, K. C. Goh, Y. He, and K. Y. H. Gin, “Present and future potential role of toxin-producing Synechococcus in the tropical region,” *Science of the Total Environment*, vol. 896, Oct. 2023, doi: 10.1016/j.scitotenv.2023.165230.

[22] J. S. Metcalf and G. A. Codd, “Co-Occurrence of Cyanobacteria and Cyanotoxins with Other Environmental Health Hazards: Impacts and Implications,” *Toxins 2020, Vol. 12, Page 629*, vol. 12, no. 10, p. 629, Oct. 2020, doi: 10.3390/TOXINS12100629.

[23] S. Zhang *et al.*, “Dissemination of antibiotic resistance genes (ARGs) via integrons in Escherichia coli: A risk to human health,” *Environ Pollut*, vol. 266, no. Pt 2, Nov. 2020, doi: 10.1016/J.ENVPOL.2020.115260.

[24] T. T. Mallick *et al.*, “Molecular and genomic investigation unveils Pseudomonas putida as an emerging multidrug-resistant pathogen linked to bovine clinical mastitis,” *Microb Pathog*, vol. 203, p. 107461, Jun. 2025, doi: 10.1016/J.MICPATH.2025.107461.

[25] A. Ghazawi *et al.*, “Genomic Study of High-Risk Clones of Enterobacter hormaechei Collected from Tertiary Hospitals in the United Arab Emirates,” *Antibiotics*, vol. 13, no. 7, p. 592, Jul. 2024, doi: 10.3390/ANTIBIOTICS13070592.

[26] T. Xu, J. Song, J. Liu, L. Huang, Z. Li, and K. Zhou, "First report of multidrug-resistant carbapenemase-producing Aeromonas caviae co-harboring mcr-3.43 and mcr-7.2," *Microbiol Spectr*, vol. 12, no. 5, pp. e03685-23, May 2024, doi: 10.1128/SPECTRUM.03685-23.

[27] P. Chaturvedi *et al.*, "Dissemination of antibiotic resistance genes, mobile genetic elements, and efflux genes in anthropogenically impacted riverine environments," *Chemosphere*, vol. 273, p. 129693, Jun. 2021, doi: 10.1016/J.CHEMOSPHERE.2021.129693.

[28] P. Sivalingam *et al.*, "Anthropogenic pollution may enhance natural transformation in water, favouring the spread of antibiotic resistance genes," *J Hazard Mater*, vol. 475, p. 134885, Aug. 2024, doi: 10.1016/J.JHAZMAT.2024.134885.

[29] IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, "Pará | Cidades e Estados | IBGE." Accessed: Apr. 15, 2025. [Online]. Available: https://www.ibge.gov.br/cidades-e-estados/pa.html

[30] R. B. L. Cavalcante, A. Fleischmann, P. R. M. Pontes, P. Rógenes, and M. Pontes, "Looking upstream: analyzing the protection of the drainage area of Amazon rivers," Nov. 2023, doi: 10.31223/X51M4F.

[31] K. L. Kalantar *et al.*, "IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring," *Gigascience*, vol. 9, no. 10, pp. 1–14, Oct. 2020, doi: 10.1093/GIGASCIENCE/GIAA111.

[32] D. E. Wood, J. Lu, and B. Langmead, "Improved metagenomic analysis with Kraken 2," *bioRxiv*, p. 762302, Sep. 2019, doi: 10.1101/762302.

[33] L. A. L. Abueg *et al.*, "The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update," *Nucleic Acids Res*, vol. 52, no. W1, pp. W83–W94, Jul. 2024, doi: 10.1093/NAR/GKAE410.

[34] J. Lu, F. P. Breitwieser, P. Thielen, and S. L. Salzberg, "Bracken: Estimating species abundance in metagenomics data," *PeerJ Comput Sci*, vol. 2017, no. 1, p. e104, Jan. 2017, doi: 10.7717/PEERJ-CS.104/SUPP-5.

[35] F. P. Breitwieser and S. L. Salzberg, "Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification," *Bioinformatics*, vol. 36, no. 4, pp. 1303–1304, Feb. 2020, doi: 10.1093/BIOINFORMATICS/BTZ715.

[36] H. M. Bik and P. I. Inc., "Phinch: An interactive, exploratory data visualization framework for –Omic datasets," *bioRxiv*, p. 009944, Oct. 2014, doi: 10.1101/009944.

[37] D. R. Finn, "A metagenomic alpha-diversity index for microbial functional biodiversity," *FEMS Microbiol Ecol*, vol. 100, no. 3, Feb. 2024, doi: 10.1093/FEMSEC/FIAE019.

[38] D. Li, C. M. Liu, R. Luo, K. Sadakane, and T. W. Lam, "MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph," *Bioinformatics*, vol. 31, no. 10, pp. 1674–1676, May 2015, doi: 10.1093/BIOINFORMATICS/BTV033.

[39] D. Hyatt, G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser, "Prodigal: Prokaryotic gene recognition and translation initiation site identification," *BMC Bioinformatics*, vol. 11, no. 1, pp. 1–11, Mar. 2010, doi: 10.1186/1471-2105-11-119/TABLES/5.

[40] Jari Oksanen; et al., "Community Ecology Package [R package vegan version 2.6-10]," Jan. 2025, doi: 10.32614/CRAN.PACKAGE.VEGAN.

[41] B. P. Alcock *et al.*, "CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database," *Nucleic Acids Res*, vol. 48, no. D1, p. D517, Jan. 2019, doi: 10.1093/NAR/GKZ935.

[42] X. Yin, X. Zheng, L. Li, A. N. Zhang, X. T. Jiang, and T. Zhang, "ARGs-OAP v3.0: Antibiotic-Resistance Gene Database Curation and Analysis Pipeline Optimization," *Engineering*, vol. 27, pp. 234–241, Aug. 2023, doi: 10.1016/J.ENG.2022.10.011.

[43] D. McDonald *et al.*, "Greengenes2 unifies microbial data in a single reference tree," *Nature Biotechnology 2023 42:5*, vol. 42, no. 5, pp. 715–718, Jul. 2023, doi: 10.1038/s41587-023-01845-1.

[44] F. E. Angly, P. G. Dennis, A. Skarshewski, I. Vanwonterghem, P. Hugenholtz, and G. W. Tyson, "CopyRighter: A rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction," *Microbiome*, vol. 2, no. 1, pp. 1–13, Apr. 2014, doi: 10.1186/2049-2618-2-11/FIGURES/6.

[45] A. N. Zhang *et al.*, "An omics-based framework for assessing the health risk of antimicrobial resistance genes," *Nature Communications 2021 12:1*, vol. 12, no. 1, pp. 1–11, Aug. 2021, doi: 10.1038/s41467-021-25096-3.

Mapping the Resistome: Metagenomic Surveillance Reveals AMR Hotspots and Ecological Risk in Amazonian Freshwaters

Wylerson Guimarães Nogueira[1,2], David Tavares Martins[3,4,5], Carlos Willian Dias Dantas[3,4,5], Victor Benedito Costa Ferreira[3,4,5], Oscar Victor Cardenas-Alegria[3,4,5], Edian Franklin Franco de Los Santos[6], Paulo Rogenes Monteiro Pontes[7], Rosane Barbosa Lopes Cavalcante[7], Júnia Schultz[2], Vasco Ariston de Carvalho Azevedo[8], Alexandre Soares Rosado[2], Rommel Thiago Jucá Ramos[3,4,5*]

[1]Department of Biochemistry and Immunology, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[2]Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[3]Center of High Performance Computer and Artificial Intelligence, Federal University of Pará, Belém, Pará, Brazil
[4]Laboratory of Bioinformatics and Genomics of Microorganisms, Federal University of Pará-UFPA, Belém, Pará, Brazil
[5]Laboratory of Simulation and Computational Biology (SIMBIC), Federal University of Pará, Belém, Pará, Brazil
[6]Universidade Tecnológica de Santiago, Santo Domingo, Republic Dominican.
[7]Vale Institute of Technology – Sustainable Development, Belém, Pará, Brazil
[8]Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil

*** Correspondence:**
Rommel Thiago Jucá Ramos
rommelthiago@gmail.com

Keywords: Metagenomics, Antimicrobial Resistance (AMR), Resistome, Freshwater Microbiomes, Amazon Basin, Waterborne Pathogens, One Health.

Figures & Tables

2. Results

2.1. *High-Quality Metagenomic Data from a Complex Amazonian Landscape*

Sample name	Total reads	% Passed Quality Control	QC filtering		Host Filtering Step		Duplicate read removal		Compression ratio	Collection date	Collection location
			Reads remaining	% Reads Remaining	Reads remaining	% Reads Remaining	Reads remaining	% Reads Remaining			
AP1 A	64.198.626	89,21	57.270.084	89,21	57.270.076	89,21	54.389.928	84,72	1.05	2022-11	Altamira
AP1 B	78.681.368	88,58	69.697.756	88,58	69.697.744	88,58	65.831.676	83,67	1.06	2022-11	Altamira
AP1 C	80.471.742	88,34	71.089.598	88,34	71.089.582	88,34	67.015.492	83,28	1.06	2022-11	Altamira
AP2 A	71.502.404	89,78	64.194.120	89,78	64.194.114	89,78	60.844.716	85,09	1.05	2022-11	Altamira
AP2 B	68.388.590	88,68	60.648.834	88,68	60.648.828	88,68	57.661.168	84,31	1.05	2022-11	Altamira
AP2 C	87.342.288	89,38	78.065.982	89,38	78.065.966	89,38	73.130.646	83,73	1.07	2022-11	Altamira
AP3 A	69.350.748	92,20	63.941.940	92,20	63.941.940	92,20	60.901.832	87,82	1.05	2022-11	Altamira
AP3 B	72.664.952	91,97	66.828.946	91,97	66.828.940	91,97	63.540.948	87,44	1.05	2022-11	Altamira
AP3 C	78.653.682	91,96	72.331.526	91,96	72.331.524	91,96	68.445.006	87,02	1.06	2022-11	Altamira
AP4 A	78.212.402	92,34	72.223.282	92,34	72.223.278	92,34	68.571.612	87,67	1.05	2022-11	Altamira
AP4 B	73.994.760	91,91	68.010.354	91,91	68.010.350	91,91	64.302.448	86,90	1.06	2022-11	Altamira
AP4 C	62.940.342	91,32	57.474.550	91,32	57.474.550	91,32	54.849.728	87,15	1.05	2022-11	Altamira
AP5 A	69.068.632	90,14	62.254.466	90,13	62.254.462	90,13	59.276.324	85,82	1.05	2022-11	Altamira
AP5 B	70.558.070	90,77	64.047.956	90,77	64.047.952	90,77	60.834.698	86,22	1.05	2022-11	Altamira
AP5 C	71.076.230	90,82	64.548.440	90,82	64.548.438	90,82	61.505.660	86,53	1.05	2022-11	Altamira
BP1 A	90.713.876	89,34	81.042.960	89,34	81.042.954	89,34	77.206.530	85,11	1.05	2022-10	Bragança
BP2 A	75.738.794	90,30	68.394.534	90,30	68.394.530	90,30	64.656.220	85,37	1.06	2022-10	Bragança
BP2 B	74.073.108	90,98	67.391.214	90,98	67.391.208	90,98	64.187.154	86,65	1.05	2022-10	Bragança
BP2 C	69.314.484	91,12	63.159.168	91,12	63.159.168	91,12	60.054.982	86,64	1.05	2022-10	Bragança
BP3 A	65.481.786	91,94	60.206.398	91,94	60.206.394	91,94	57.339.498	87,57	1.05	2022-10	Bragança
BP3 B	79.938.780	91,60	73.222.402	91,60	73.222.398	91,60	69.556.394	87,01	1.05	2022-10	Bragança
BP3 C	75.346.710	91,89	69.236.156	91,89	69.236.152	91,89	65.536.566	86,98	1.06	2022-10	Bragança
BP4 A	71.428.774	91,82	65.583.144	91,82	65.583.144	91,82	62.290.844	87,21	1.05	2022-10	Bragança
BP4 B	68.611.014	92,47	63.445.582	92,47	63.445.582	92,47	60.532.266	88,23	1.05	2022-10	Bragança
BP4 C	78.480.142	91,58	71.872.536	91,58	71.872.534	91,58	67.710.506	86,28	1.06	2022-10	Bragança
BP5 A	65.630.128	91,32	59.935.884	91,32	59.935.880	91,32	57.147.438	87,08	1.05	2022-10	Bragança
BP5 B	62.228.644	91,15	56.721.652	91,15	56.721.652	91,15	54.231.188	87,15	1.05	2022-10	Bragança
BP5 C	70.917.012	90,81	64.400.650	90,81	64.400.650	90,81	61.066.076	86,11	1.05	2022-10	Bragança
MP1 A	91.615.886	90,89	83.267.604	90,89	83.267.604	90,89	79.056.638	86,29	1.05	2022-11	Marabá
MP1 B	70.608.562	90,02	63.559.328	90,02	63.559.326	90,02	60.878.314	86,22	1.04	2022-11	Marabá
MP1 C	74.851.014	90,84	67.996.666	90,84	67.996.662	90,84	64.823.914	86,60	1.05	2022-11	Marabá
MP2 A	74.560.210	90,90	67.772.806	90,90	67.772.806	90,90	64.528.036	86,54	1.05	2022-11	Marabá
MP2 B	75.436.092	89,74	67.698.750	89,74	67.698.750	89,74	64.771.694	85,86	1.04	2022-11	Marabá
MP2 C	74.001.388	91,63	67.806.596	91,63	67.806.594	91,63	64.222.092	86,78	1.06	2022-11	Marabá
MP3 A	72.355.746	90,97	65.821.374	90,97	65.821.370	90,97	62.246.604	86,03	1.06	2022-11	Marabá
MP3 B	67.060.528	91,03	61.043.854	91,03	61.043.854	91,03	58.122.364	86,67	1.05	2022-11	Marabá
MP3 C	62.274.932	91,37	56.899.958	91,37	56.899.958	91,37	54.052.108	86,80	1.05	2022-11	Marabá
MP4 A	60.186.754	90,95	54.736.724	90,94	54.736.724	90,94	52.505.064	87,24	1.04	2022-11	Marabá
MP4 B	65.123.244	89,77	58.461.592	89,77	58.461.590	89,77	55.597.634	85,37	1.05	2022-11	Marabá
MP4 C	61.596.106	91,03	56.069.896	91,03	56.069.892	91,03	53.755.628	87,27	1.04	2022-11	Marabá
MP5 A	67.507.898	91,76	61.947.286	91,76	61.947.286	91,76	58.916.494	87,27	1.05	2022-11	Marabá
MP5 B	78.869.912	91,68	72.308.694	91,68	72.308.694	91,68	68.041.480	86,27	1.06	2022-11	Marabá

MP5 C	59.747.138	91,22	54.501.368	91,22	54.501.368	91,22	52.161.768	87,30	1.04	2022-11	Marabá
PP1 A	74.980.716	89,44	67.060.776	89,44	67.060.774	89,44	63.744.572	85,01	1.05	2022-10	Paragominas
PP1 B	68.322.002	91,76	62.693.752	91,76	62.693.752	91,76	59.481.370	87,06	1.05	2022-10	Paragominas
PP1 C	73.749.982	91,00	67.114.896	91,00	67.114.890	91,00	63.770.654	86,47	1.05	2022-10	Paragominas
PP2 A	74.205.958	91,82	68.137.688	91,82	68.137.680	91,82	64.475.680	86,89	1.06	2022-10	Paragominas
PP2 B	61.296.256	93,08	57.053.214	93,08	57.053.210	93,08	54.153.078	88,35	1.05	2022-10	Paragominas
PP2 C	65.297.132	90,16	58.871.110	90,16	58.871.108	90,16	56.232.716	86,12	1.05	2022-10	Paragominas
PP3 A	71.259.250	91,41	65.134.094	91,40	65.134.088	91,40	61.817.930	86,75	1.05	2022-10	Paragominas
PP3 B	59.921.434	91,68	54.937.114	91,68	54.937.106	91,68	52.631.588	87,83	1.04	2022-10	Paragominas
PP3 C	66.066.316	90,86	60.029.992	90,86	60.029.990	90,86	57.190.686	86,57	1.05	2022-10	Paragominas
PP4 A	71.142.442	91,01	64.744.604	91,01	64.744.602	91,01	61.603.880	86,59	1.05	2022-10	Paragominas
PP4 B	63.385.400	89,93	57.004.922	89,93	57.004.918	89,93	54.586.588	86,12	1.04	2022-10	Paragominas
PP4 C	89.490.902	91,59	81.965.662	91,59	81.965.658	91,59	76.851.902	85,88	1.07	2022-10	Paragominas
PP5 A	82.174.602	90,69	74.521.492	90,69	74.521.478	90,69	70.377.238	85,64	1.06	2022-10	Paragominas
PP5 B	67.160.506	89,66	60.215.758	89,66	60.215.758	89,66	57.705.872	85,92	1.04	2022-10	Paragominas
PP5 C	75.592.326	90,08	68.094.230	90,08	68.094.226	90,08	64.707.924	85,60	1.05	2022-10	Paragominas

Table 1. Summary of sequencing output and quality control metrics for all environmental DNA samples. Summary of sequencing and preprocessing metrics for the 58 environmental water samples analyzed in this study. Values include total reads, quality control retention, and remaining reads after QC filtering, host DNA removal, and duplicate read removal. Metadata columns indicate sample type, nucleotide type, collection date, and location. The ‘Sample type’ and ‘Nucleotide type’ for all of the samples was Environmental and DNA, respectively.

2.2. *Microbial Communities Mirror Environmental Gradients Across the Amazon*

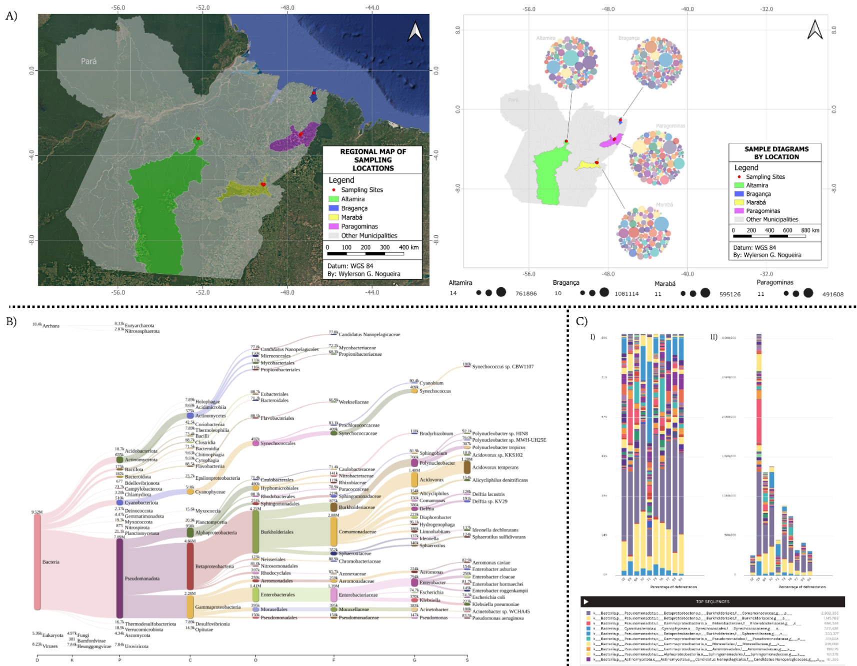


Figure 1. Geographic distribution and taxonomic profiles of Amazonian freshwater microbiomes. (A) Regional map of the 20 freshwater sampling sites across four municipalities in the state of Pará, Brazil. Colored polygons indicate municipal boundaries;

bubble charts display bacterial family-level abundance per city, scaled by relative abundance and sequencing depth. (B) Sankey diagram showing the cumulative taxonomic classification of all 58 metagenomic samples across hierarchical levels from domain (D) to species (S). The majority of classified reads belong to the domain *Bacteria*, spanning diverse phyla and several clinically relevant genera. Flows are scaled by read abundance. (C) Attribute column charts of bacterial communities stratified by percentage of regional deforestation. Left (I): relative abundance by family; Right (II): absolute abundance. Bottom panel lists the top-ranked taxonomic lineages by cumulative sequence count. These patterns illustrate the influence of environmental disturbance on microbial structure and dominance.

2.3. *Local Structure Within a Diverse System: Richness, Beta-Diversity, and Community Separation*

	Indexes						Standard Deviation (SD)					
Site	Total Genes	Protein Richness	Shannon Diversity	Simpson Evenness	Log10 Protein Dissimilarity	Metagenomic Diversity	Total Genes	Protein Richness	Shannon Diversity	Simpson Evenness	Log10 Protein Dissimilarity	Metagenomic Diversity
AP1	52550	42754	10,563	0,9999690	4,643	0,836	5,01E+03	3,38E+03	7,52E-02	2,07E-06	3,72E-02	1,57E-02
AP2	51344	43288	10,591	0,9999706	4,648	0,865	3,57E+03	2,63E+03	5,70E-02	1,46E-06	2,70E-02	7,14E-03
AP3	43840	33269	10,281	0,9999566	4,539	0,789	3,25E+03	1,88E+03	5,00E-02	2,13E-06	2,56E-02	1,26E-02
AP4	36289	28742	10,160	0,9999534	4,473	0,805	5,49E+03	3,56E+03	1,02E-01	3,80E-06	5,04E-02	1,78E-02
AP5	41335	34488	10,335	0,9999601	4,552	0,863	3,75E+03	2,08E+03	5,24E-02	1,90E-06	2,87E-02	3,12E-02
BP1	2811	2748	7,910	0,9996276	3,441	0,982	---	---	---	---	---	---
BP2	9682	8938	9,065	0,9998777	3,957	0,934	2,58E+03	2,00E+03	1,81E-01	1,67E-05	8,71E-02	2,48E-02
BP3	15069	12455	9,339	0,9998920	4,107	0,851	2,14E+03	1,44E+03	1,01E-01	6,84E-06	5,22E-02	2,77E-02
BP4	42500	34831	10,363	0,9999624	4,555	0,845	9,39E+03	7,27E+03	2,47E-01	1,24E-05	1,11E-01	1,20E-02
BP5	16912	14661	9,529	0,9999192	4,174	0,884	1,67E+03	1,28E+03	8,91E-02	7,23E-06	4,13E-02	9,71E-03
MP1	7822	7505	8,905	0,9998599	3,879	0,967	3,28E+02	2,98E+02	3,70E-02	4,50E-06	1,69E-02	1,89E-03
MP2	31334	26792	10,120	0,9999539	4,438	0,876	2,02E+03	1,42E+03	4,90E-02	1,95E-06	2,41E-02	9,12E-03
MP3	30902	26253	10,098	0,9999529	4,430	0,880	2,98E+03	2,46E+03	8,94E-02	3,88E-06	4,04E-02	7,61E-03
MP4	28333	24743	10,050	0,9999511	4,403	0,892	1,68E+03	1,42E+03	5,58E-02	2,35E-06	2,54E-02	2,32E-03
MP5	32101	27327	10,140	0,9999551	4,447	0,875	6,71E+02	5,22E+02	1,94E-02	9,66E-07	8,46E-03	2,67E-03
PP1	41918	31342	10,213	0,9999534	4,515	0,781	2,24E+04	1,54E+04	7,79E-01	5,32E-05	3,80E-01	1,09E-01
PP2	48795	35498	10,322	0,9999572	4,570	0,764	7,60E+03	4,37E+03	1,09E-01	3,01E-06	5,92E-02	3,10E-02
PP3	47039	34157	10,287	0,9999568	4,553	0,767	4,57E+03	2,55E+03	6,44E-02	2,10E-06	3,54E-02	1,95E-02
PP4	42193	31977	10,240	0,9999550	4,524	0,791	7,33E+03	4,04E+03	1,05E-01	2,79E-06	5,95E-02	3,56E-02
PP5	27864	22942	9,954	0,9999452	4,374	0,850	3,59E+03	2,25E+03	7,93E-02	3,22E-06	4,15E-02	1,90E-02

Table 2. Summary of alpha diversity metrics per sampling site. Median values and standard deviations of alpha diversity indices were calculated from triplicate samples at each of the 20 freshwater sampling sites. Metrics include total predicted gene content, protein richness, Shannon diversity, Simpson evenness, protein dissimilarity, and metagenomic diversity (MD). These indices capture taxonomic and functional diversity and assess microbial community richness and ecological balance across sampling locations.

Resistome of Amazonian Freshwaters

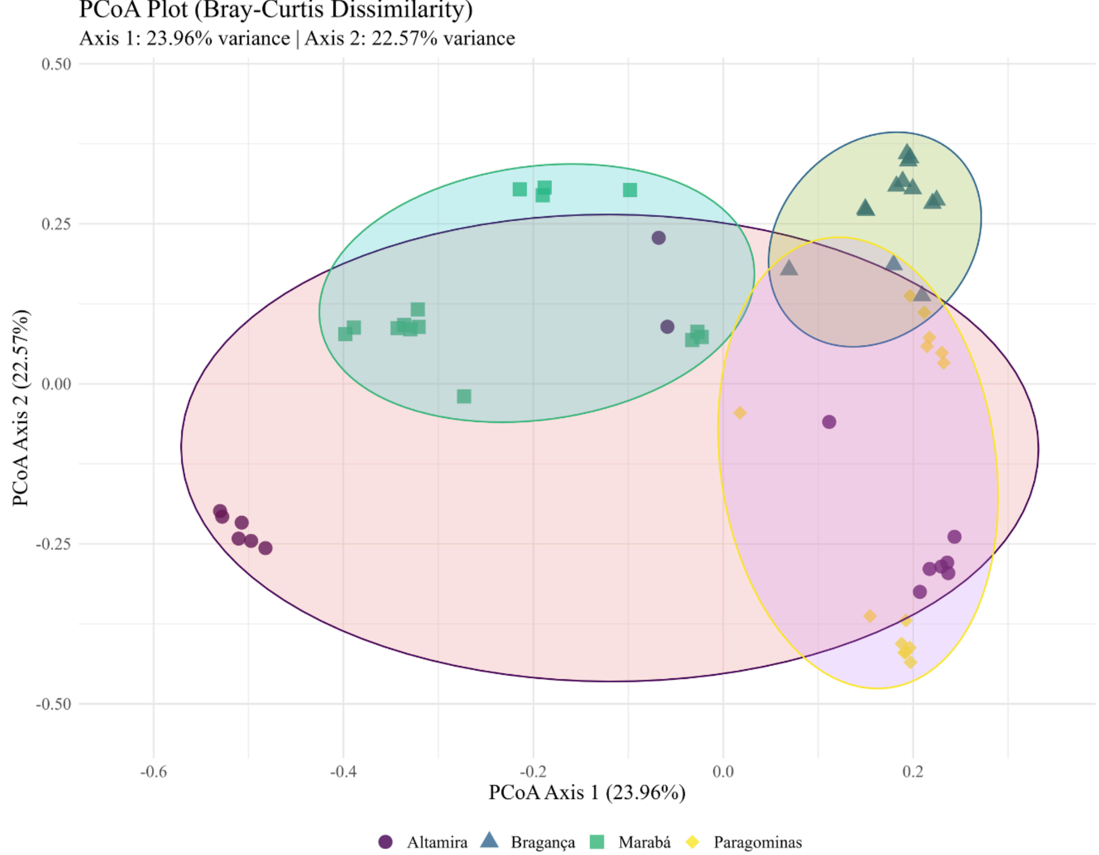


Figure 2. Principal Coordinates Analysis (PCoA) of microbial community composition across all samples. PCoA plot based on Bray-Curtis dissimilarity matrix of all 58 samples, colored by municipality. Axis 1 and 2 explain 23.96% and 22.57% of the variation, respectively. Ellipses represent 95% confidence intervals for each city. The plot illustrates both city-level clustering and intra-city variation in microbial community composition.

2.4. *Pathogens Detected Across the Region — But They Don’t Explain AMR Alone*

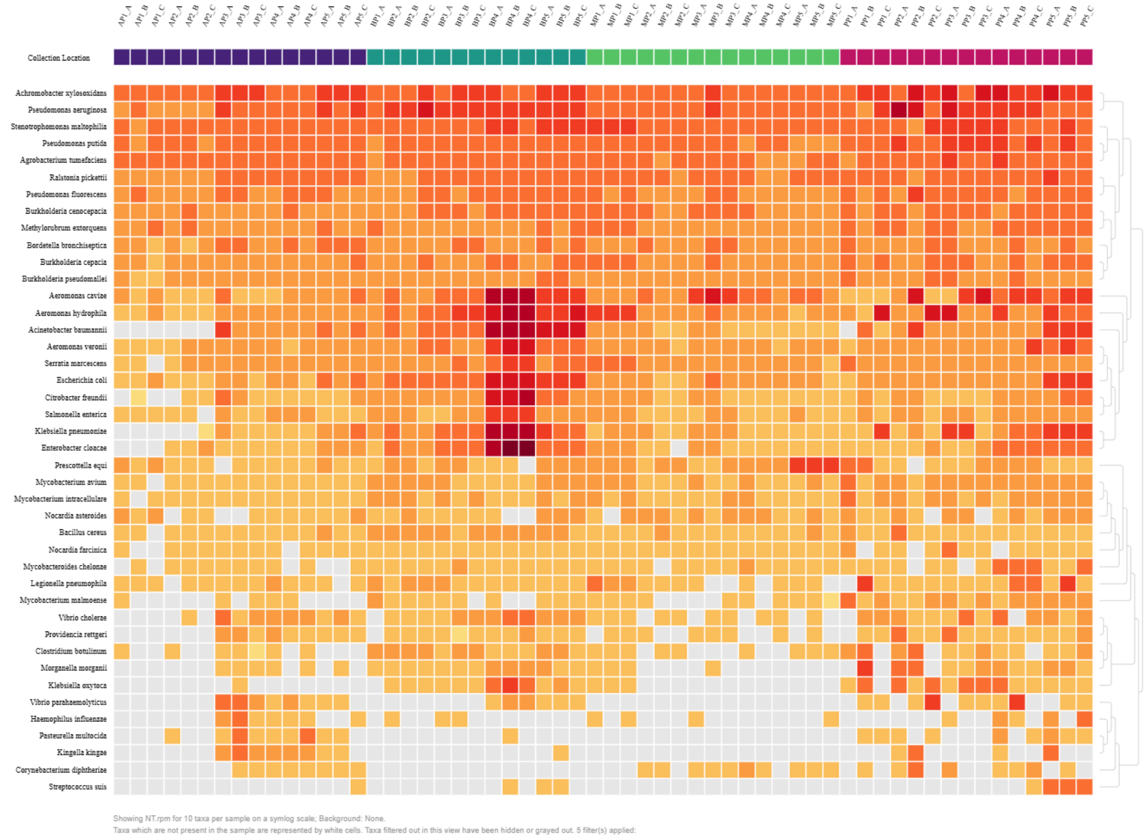


Figure 3. Heatmap of bacterial species with known pathogenic potential across metagenomic samples. Heatmap displays the ten most abundant pathogenic species per sample (reads per million, RPM), clustered by taxon similarity. Several clinically relevant taxa – including *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, *Enterobacter cloacae*, and *Stenotrophomonas maltophilia* – were detected, particularly in samples from Marabá, Bragança, and Paragominas. White cells represent taxa not detected in the corresponding sample. Filtering thresholds applied for inclusion were: nucleotide rPM ≥ 10, protein rPM ≥ 1, and alignment length ≥ 50 bp.

2.5. *Antimicrobial Resistance Genes Are Widespread, Even in Seemingly Pristine Sites*

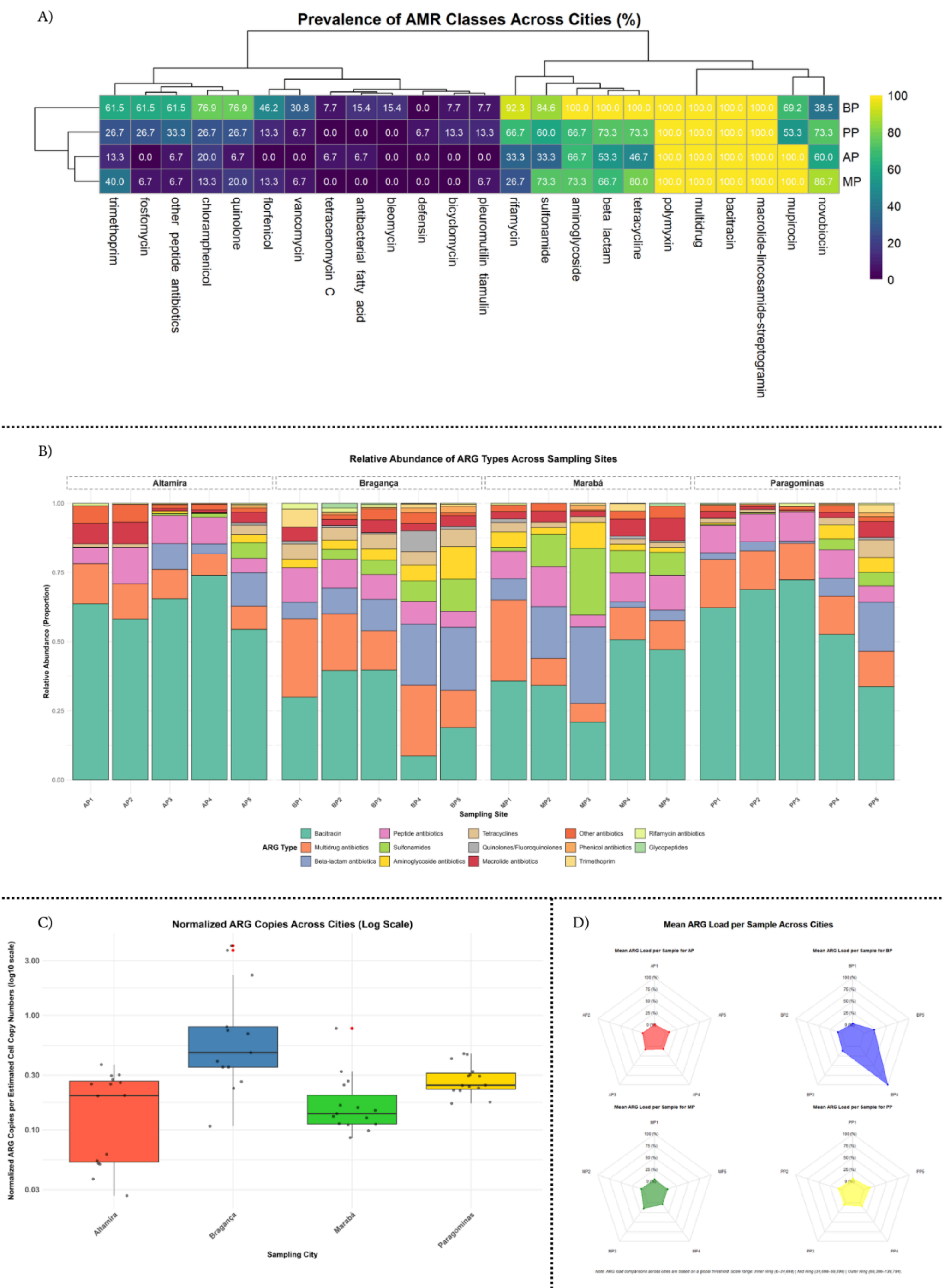


Figure 4. Composite analysis of antimicrobial resistance (AMR) burden across sampling sites and cities. (A) Heatmap of AMR class prevalence across the four cities expressed as a percentage of samples in which each class was detected. (B) Relative abundance of ARG types per sampling site, grouped by city. (C) Boxplot of normalized ARG copies per estimated number of cells (log_{10} scale), grouped by city. (D) Radar plots show the mean ARG load per sample within each city. Together, these panels reveal localized ARG richness and load patterns, with Bragança and Paragominas emerging as notable putative hotspots for AMR diversity and intensity.

2.6. *High-Risk Resistance Hotspots Are Concentrated in Bragança and Marabá*

Figure 5. ARG Risk Ranking Across Sampling Sites. ARG Risk profiling across Amazonian freshwater sampling sites. (Top) Bubble plot representing the total ARG abundance per sampling site, color-scaled by Threat Score log10 (ThreatScore+ε), where ε=1e−6. Bubble size corresponds to total ARG abundance normalized by estimated cell count. (Bottom) Stacked bar chart showing the proportional distribution of ARGs per site by risk rank, from Rank I (highest risk: mobile and detected in pathogens) to Rank IV (lowest risk: not human-associated). Threat Scores were calculated by summing the relative abundance of Rank I and Rank II ARGs per sample, weighted by total ARG abundance: Threat Score = $(RA_{(Rank\ I)} + RA_{(Rank\ II)}) \times$ Total ARGs. This approach emphasizes clinically relevant resistance potential while retaining quantitative resolution across sites.

Resistome of Amazonian Freshwaters

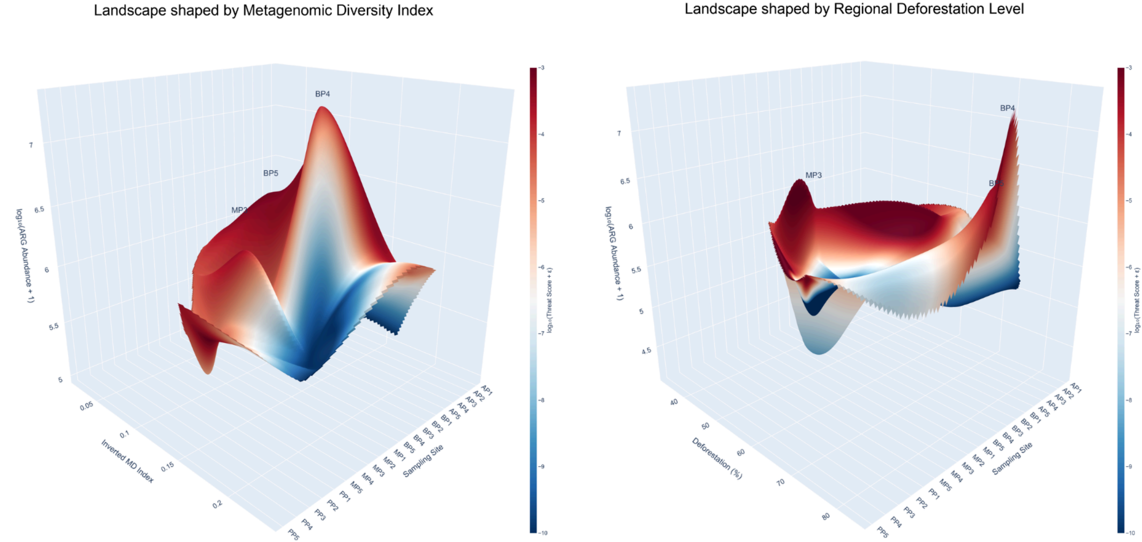


Figure 6. 3D ARG Landscapes Across Amazonian Freshwater Samples. Three-dimensional surface models integrating ARG abundance and Threat Score with ecological disturbance proxies. (Left) Landscape shaped by the inverted Metagenomic Diversity (MD) index, where higher values indicate lower within-sample diversity and presumed ecological imbalance. (Right) Landscape shaped by municipality-level deforestation percentage, obtained from publicly available environmental datasets. For both models, the z-axis represents ARG abundance log10(ARGs+1), and color gradients reflect Threat Score magnitude log10(ThreatScore+ε). Peaks correspond to sampling sites with elevated resistance potential under combined biological and anthropogenic pressure.

Mapping the Resistome: Metagenomic Surveillance Reveals AMR Hotspots and Ecological Risk in Amazonian Freshwaters

Wylerson Guimarães Nogueira[1,2], David Tavares Martins[3,4,5], Carlos Willian Dias Dantas[3,4,5], Victor Benedito Costa Ferreira[3,4,5], Oscar Victor Cardenas-Alegria[3,4,5], Edian Franklin Franco de Los Santos[6], Paulo Rogenes Monteiro Pontes[7], Rosane Barbosa Lopes Cavalcante[7], Júnia Schultz[2], Vasco Ariston de Carvalho Azevedo[8], Alexandre Soares Rosado[2], Rommel Thiago Jucá Ramos[3,4,5*]

[1]Department of Biochemistry and Immunology, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[2]Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[3]Center of High Performance Computer and Artificial Intelligence, Federal University of Pará, Belém, Pará, Brazil
[4]Laboratory of Bioinformatics and Genomics of Microorganisms, Federal University of Pará-UFPA, Belém, Pará, Brazil
[5]Laboratory of Simulation and Computational Biology (SIMBIC), Federal University of Pará, Belém, Pará, Brazil
[6]Universidade Tecnológica de Santiago, Santo Domingo, Republic Dominican.
[7]Vale Institute of Technology – Sustainable Development, Belém, Pará, Brazil
[8]Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil

*** Correspondence:**
Rommel Thiago Jucá Ramos
rommelthiago@gmail.com

Keywords: Metagenomics, Antimicrobial Resistance (AMR), Resistome, Freshwater Microbiomes, Amazon Basin, Waterborne Pathogens, One Health.

Supplementary Materials

2. Results

2.1. *High-Quality Metagenomic Data from a Complex Amazonian Landscape*

	pH	pH/mV	Temperature (ºC)	Pressure (mbar)	Resistivity (MΩ.cm)	EC (µS/cm)	Absolute EC (µS/cm) *	TDS (ppm)	Salinity (ppm)	ORP (mV)	DO%	DO (ppm)
	ALTAMIRA											
Palt 1-A	7,34	-2,4	30,46	993,9	0,037	27	30	19	0,01	-221,9	92,6	6,81
Palt 1-B	6,9	21,6	30,36	993,9	0,0398	25	28	18	0,01	-196,3	98,3	7,24
Palt 1-C	6,76	29,7	29,5	994	0,0385	26	28	18	0,01	-175,8	98,2	7,34
Palt 2-A	5,29	114,6	29,52	1005,2	0,0361	28	30	19	0,01	-128,5	71,1	5,38
Palt 2-B	5,35	111,2	29,53	1009,4	0,0362	28	30	19	0,01	-125,6	67,3	5,11
Palt 2-C	5,37	109,8	29,51	1009,9	0,0363	28	30	19	0,01	-120,1	71,8	5,46
Palt 3-A	6,55	41,7	26,76	1001,5	0,0442	23	23	16	0,01	-167	54,3	4,29
Palt 3-B	5,55	99,1	26,58	1002,3	0,0595	17	17	12	0,01	-142,2	61,9	4,91
Palt 3-C	5,38	108,3	26,58	1003,2	0,062	16	17	11	0,01	-134,9	54,6	4,34
Palt 4-A	5,29	114,1	28,66	1001,5	0,0284	35	38	25	0,01	-107,2	59,4	4,54
Palt 4-B	5,05	128,1	27,92	1001,5	0,0365	27	29	19	0,01	-102,5	64,3	4,98
Palt 4-C	5,63	95,9	32,63	1002	0,0103	97	111	68	0,04	-126,3	68,2	4,87
Palt 5-A	6,71	32,9	32,63	1003,5	0,0439	23	26	16	0,01	-96,4	98	7,01
Palt 5-B	5,91	79	31,11	1003,5	0,0418	24	27	17	0,01	-99,7	95,2	6,99
Palt 5-C	5,83	83,9	31,09	1004,2	0,0361	28	31	19	0,01	-95,6	81,2	5,96
	BRAGANÇA											
Pbra 1-A	3,98	186,2	27,03	1010,8	0,082	12	13	8	0	-147,5	97,3	7,73
Pbra 1-B	4,21	173,1	27,01	1010,7	0,077	13	13	9	0	-188,1	35,2	2,8
Pbra 1-C	3,95	187,9	26,94	1010,8	0,081	12	13	9	0	-185,9	37,2	2,96
Pbra 2-A	4,97	129,3	27,04	1013,2	0,0261	38	40	27	0,02	-175,1	53,1	4,22
Pbra 2-B	4,77	141,2	27,04	1013,4	0,0246	41	42	29	0,02	-171	50,3	4
Pbra 2-C	4,73	143,4	27,06	1013,5	0,0244	41	43	29	0,02	-175	52,9	4,21
Pbra 3-A	5,25	113,6	27,96	1015,1	0	0	0	0	0	-150,6	88,2	6,92
Pbra 3-B	5,39	105,4	27,29	1015,2	0,0103	97	102	68	0,04	-167,9	64,6	5,13
Pbra 3-C	5,41	104,5	27,27	1015,2	0,0102	98	102	69	0,04	-189,6	50	3,97
Pbra 4-A	5,36	107,1	27,03	1013,6	0,0084	119	123	83	0,05	-170,4	22,8	1,82
Pbra 4-B	5,27	112,5	27,02	1013,7	0,0084	120	124	84	0,06	-166,4	22	1,75
Pbra 4-C	5,28	111,5	26,07	1013,9	0,0166	60	62	42	0,03	-180	24,6	2
Pbra 5-A	6	70,5	26,63	1010,4	0,008	126	130	88	0,06	-221,2	16,5	1,32
Pbra 5-B	5,98	71,3	26,63	1010,5	0,008	126	130	88	0,06	-219,2	19,2	1,54
Pbra 5-C	6,41	47	26,63	1011,2	0,008	125	128	87	0,06	-222,8	18,9	1,51
	MARABÁ											
Pmar 1-A	6,22	60,3	27,2	1002,6	0,0252	40	41	28	0,02	-114,7	93,2	7,32
Pmar 1-B	6,04	70,8	27,03	1002,6	0,0253	39	41	28	0,02	-108,5	86,6	6,82
Pmar 1-C	7,3	-0,3	27,05	1002,6	0,0249	40	42	28	0,02	-120,4	86,1	6,78

Pmar 2-A	6,59	39,7	31,05	999,3	0,0115	87	97	61	0,04	85,4	87,7	6,41
Pmar 2-B	6,96	18,3	31,06	999,8	0,0115	87	97	61	0,04	55,1	87,3	6,39
Pmar 2-C	7,05	12,8	31,08	1000,2	0,0115	87	97	61	0,04	-131,2	88,4	6,47
Pmar 3-A	6,84	25,2	31,25	1000,4	0,0115	87	98	61	0,04	-139,1	86,8	6,34
Pmar 3-B	6,83	25,9	31,3	1000,4	0,0114	88	98	61	0,04	-141,2	85,3	6,22
Pmar 3-C	6,85	24,3	30,27	1000,4	0,0112	89	98	62	0,04	-139,3	91,8	6,82
Pmar 4-A	7,49	-11	31,08	1002,9	0,0115	87	97	61	0,04	-128,8	90,7	6,66
Pmar 4-B	7,35	-3,4	31,66	1003	0,011	91	102	63	0,04	-131,9	92,7	6,74
Pmar 4-C	6,98	16,9	31,57	1003,1	0,0112	89	100	62	0,04	-137,6	92,8	6,76
Pmar 5-A	7,52	-12,1	29,71	1002,8	0,0206	49	53	34	0,02	-159,2	84,8	6,37
Pmar 5-B	6,99	16,6	30,46	1003,2	0,0113	89	98	62	0,04	-132,3	90,2	6,7
Pmar 5-C	6,9	21,3	30,47	1003,3	0,0114	88	97	61	0,04	-138,5	87,8	6,51
	PARAGOMINAS											
Ppar 1-A	4,65	148,8	28,61	1000,4	0,0387	26	28	18	0,01	-145,3	78,5	6
Ppar 1-B	4,43	161	28,59	1000,5	0,0353	28	30	20	0,01	-143	75,6	5,48
Ppar 1-C	4,22	173,2	28,59	1000,5	0,0351	28	30	20	0,01	-138	73,5	5,62
Ppar 2-A	3,3	225,7	27,96	1004,3	0,0364	27	29	19	0,01	-22	66,6	5,17
Ppar 2-B	3,77	199	28,12	1004,4	0,0467	21	23	15	0,01	-72,6	70,8	5,48
Ppar 2-C	3,57	210,5	27,97	1006	0,0418	24	25	17	0,01	-81,1	76,4	5,94
Ppar 3-A	4,37	164,2	27,71	1000,8	0,036	28	29	19	0,01	-82,8	73,8	5,73
Ppar 3-B	4,16	176,4	27,72	1001,2	0,0336	30	31	21	0,01	-78,9	73,5	5,71
Ppar 3-C	3,99	185,9	27,7	1001,4	0,034	29	31	31	0,01	-83,2	68,6	5,33
Ppar 4-A	3,71	202,9	28,43	1003,1	0,0408	25	26	17	0,01	-131,8	69,4	5,34
Ppar 4-B	3,71	202,6	28,54	1003,1	0,0367	27	29	19	0,01	-145,2	82,7	6,34
Ppar 4-C	3,66	205,5	28,54	1003,3	0,0349	29	31	20	0,01	-132,5	78	5,99
Ppar 5-A	4,74	143,2	28,13	1002,3	0,0314	32	34	22	0,01	-162,4	57	4,4
Ppar 5-B	4,64	148,8	28,13	1002,6	0,0311	32	34	23	0,01	-154,3	56,5	4,37
Ppar 5-C	4,5	157	28,12	1002,8	0,0313	32	34	22	0,01	-157,5	57,8	4,47

Supplementary Table 1. Physicochemical parameters measured per sample at the time of collection. Field measurements collected for each of the original 60 water samples across 20 sites, including pH, temperature, pressure, conductivity, resistivity, total dissolved solids (TDS), salinity, dissolved oxygen (DO), and oxidation-reduction potential (ORP). The values of pH, temperature, and pressure were used as environmental covariates in downstream metagenomic analyses.

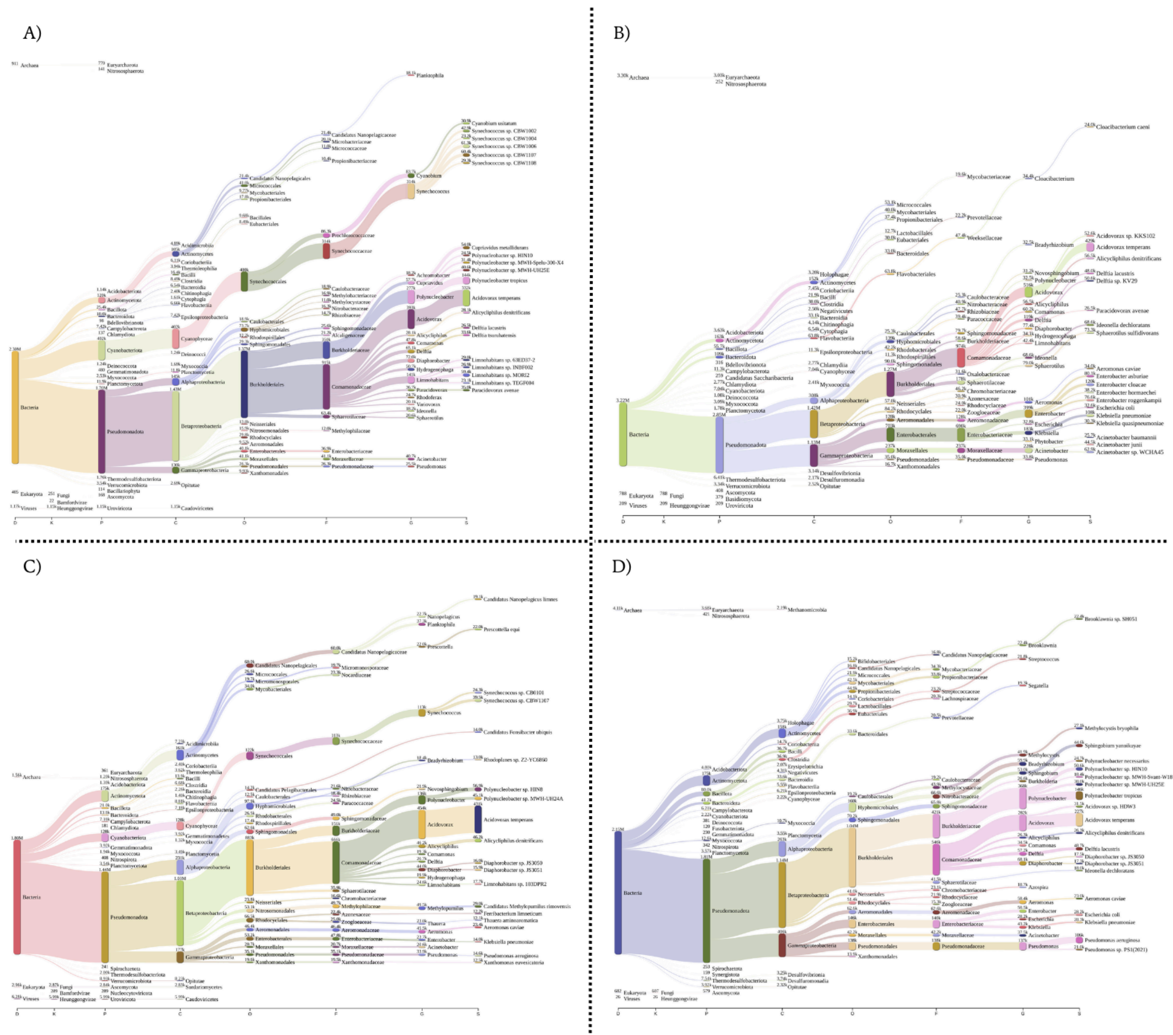
	Altamira					Bragança					Marabá					Paragominas				
Sampling Site	AP1	AP2	AP3	AP4	AP5	BP1	BP2	BP3	BP4	BP5	MP1	MP2	MP3	MP4	MP5	PP1	PP2	PP3	PP4	PP5
Population density (pop/km²)	0	701	190	360	449	0	1434	2135	3728	3886	6.68	13.02	14.44	15.83	16.06	1	4	31	61	184
Percentage of deforestation	73	76	64	64.7	64.9	71.4	76	77.2	83.2	84	32.9	53.51	53.47	53.46	53.46	70	71.1	70.3	70.8	73

Supplementary Table 2. Socio environmental indicators per sampling site. Metadata per sampling site indicating estimated population density (pop/km²) and percentage of deforestation for each of the 20 locations across the four municipalities. Values were used as contextual variables for interpreting antimicrobial resistance trends and were included in relevant multivariate analyses.

Municipality	Land Area (km2) [2023]	Estimated Population (nº of people) [2024]	Population Density (hab/km2) [2022]	IDHM (Municipal Human Development Index) [2010]	Infant Mortality - deaths per thousand live births [2022]	GDP per capita (R$) [2021]
Altamira	159.533.306	136.982	0,79	0,665	13,58	26.595,80
Bragança	2.124.735	131.679	57,93	0,600	11,59	10.679,00
Marabá	15.127.872	288.513	17,62	0,668	14,75	47.010,21
Paragominas	19.342.565	112.843	5,46	0,645	16,18	36.952,54

Supplementary Table 3. Socioeconomic indicators per city. Municipality-level statistics for Altamira, Bragança, Marabá, and Paragominas. Data includes land area, estimated population, population density, Human Development Index (IDHM), infant mortality rate, and GDP per capita. These indicators were used to contextualize each region's sampling environment.

2.2. *Microbial Communities Mirror Environmental Gradients Across the Amazon*

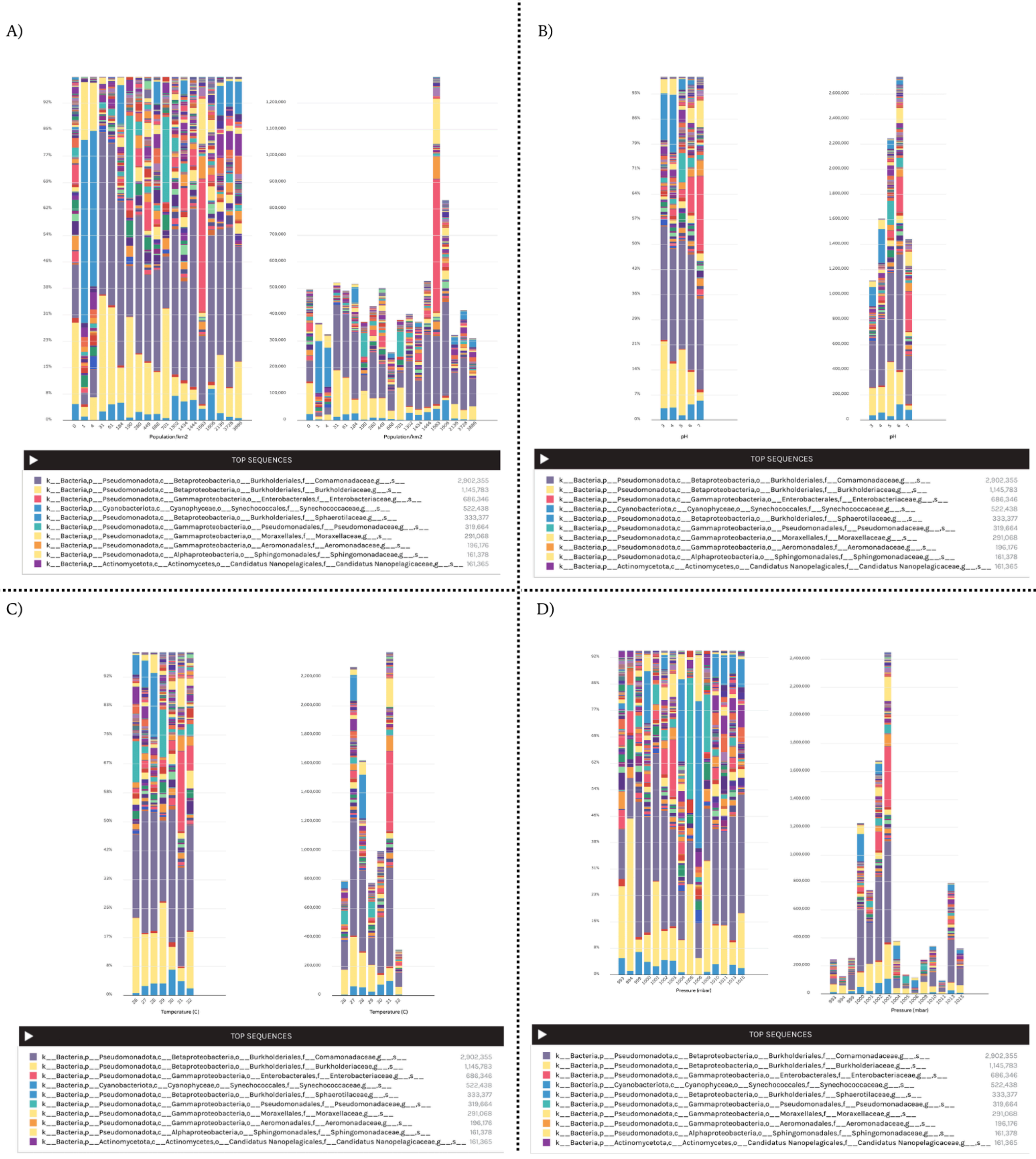


Supplementary Figure 1. City-level taxonomic profiles of microbial communities. Sankey diagrams representing aggregated taxonomic composition per city: (A) Altamira, (B) Bragança, (C) Marabá, and (D) Paragominas. Each diagram follows the taxonomic hierarchy from domain to species, highlighting differences in microbial community structure across municipalities. These visualizations reveal city-specific trends in the relative abundance of dominant bacterial lineages.

Name	Number of raw reads	Classified reads	Chordate reads	Artificial reads	Unclassified reads	Microbial reads	Bacterial reads	Viral reads	Fungal reads	Protozoan reads
AP1 A	1050953	11.9	0.045	0	88.1	11.9	11.7	0.00971	0.004	0.00019
AP1 B	1056233	12.3	0.0618	0	87.7	12.2	12	0.00795	0.00275	0.000568
AP1 C	1058200	12.5	0.0711	0	87.5	12.4	12.2	0.00652	0.00369	0.0000945
AP2 A	1052714	10.6	0.0421	0	89.4	10.6	10.4	0.0095	0.00285	0.000475
AP2 B	1048823	10.7	0.0696	0	89.3	10.6	10.4	0.00734	0.00381	0.000191
AP2 C	1065779	11.4	0.0509	0	88.6	11.3	11.1	0.00722	0.003	0.0000938
AP3 A	1049110	18.5	0.00458	0	81.5	18.5	18.4	0.00315	0.000953	0
AP3 B	1051505	15.8	0.00466	0	84.2	15.8	15.7	0.00285	0.00152	0.00019
AP3 C	1056466	17	0.00521	0	83	16.9	16.8	0.00256	0.00123	0
AP4 A	1052501	14.6	0.00637	0	85.4	14.5	14.4	0.00428	0.002	0.00019
AP4 B	1057341	17.1	0.00558	0	82.9	17.1	16.9	0.00255	0.00142	0
AP4 C	1047167	17.1	0.00535	0	82.9	17.1	17	0.00392	0.00229	0.000286
AP5 A	1049174	16.7	0.0114	0	83.3	16.7	16.5	0.0338	0.00324	0.000286
AP5 B	1051615	17.9	0.0152	0	82.1	17.9	17.7	0.0294	0.00181	0.0000951
AP5 C	1048188	16.5	0.018	0	83.5	16.4	16.3	0.0356	0.00296	0.0000954
BP1 A	1049526	7.22	0.00553	0	92.8	7.2	7.06	0.002	0.00562	0.000191
BP2 A	1057056	13.5	0.00719	0	86.5	13.5	13.3	0.0115	0.00587	0.000189
BP2 B	1049859	12.3	0.00514	0	87.7	12.3	12.1	0.00667	0.00705	0.000286
BP2 C	1051190	14.6	0.00675	0	85.4	14.6	14.5	0.00666	0.0059	0.00019
BP3 A	1049741	14.7	0.00762	0	85.3	14.7	14.5	0.0143	0.00638	0.0000953
BP3 B	1052174	18.8	0.00494	0	81.2	18.8	18.6	0.00922	0.0038	0.00038
BP3 C	1056354	19	0.00322	0	81	19	18.9	0.0126	0.00369	0.000284
BP4 A	1052606	35.4	0.00228	0	64.6	35.3	35.2	0.00922	0.00732	0
BP4 B	1047789	44.9	0.00353	0	55.1	44.8	44.6	0.0084	0.0111	0.000191
BP4 C	1061074	45.5	0.00254	0	54.5	45.4	45.3	0.00933	0.00933	0
BP5 A	1048640	26.9	0.00486	0	73.1	26.9	26.7	0.0154	0.00782	0.000286
BP5 B	1045250	27.4	0.00258	0	72.6	27.4	27.2	0.0161	0.00765	0.000383
BP5 C	1054161	27.9	0.00465	0	72.1	27.9	27.7	0.0124	0.00683	0.00019
MP1 A	1052564	12.6	0.00542	0	87.4	12.5	12.3	0.0144	0.087	0.000475
MP1 B	1043173	12.4	0.00489	0	87.6	12.4	12.1	0.0143	0.0869	0.0000959
MP1 C	1048576	12.6	0.00525	0	87.4	12.6	12.3	0.0144	0.0873	0.000191
MP2 A	1049420	12.6	0.0131	0	87.4	12.6	12.4	0.0359	0.00172	0.0000953
MP2 B	1044540	10.2	0.0109	0	89.8	10.1	9.95	0.0435	0.00124	0
MP2 C	1055074	9.41	0.0164	0	90.6	9.38	9.19	0.0435	0.00152	0.0000948
MP3 A	1056961	12.9	0.0115	0	87.1	12.9	12.7	0.0358	0.00189	0.000189
MP3 B	1049359	18.2	0.00972	0	81.8	18.2	18	0.0331	0.00114	0.000191
MP3 C	1052094	12.3	0.00855	0	87.7	12.3	12.1	0.0385	0.00181	0.00019
MP4 A	1041886	11	0.0116	0	89	11	10.8	0.0479	0.00154	0.000288

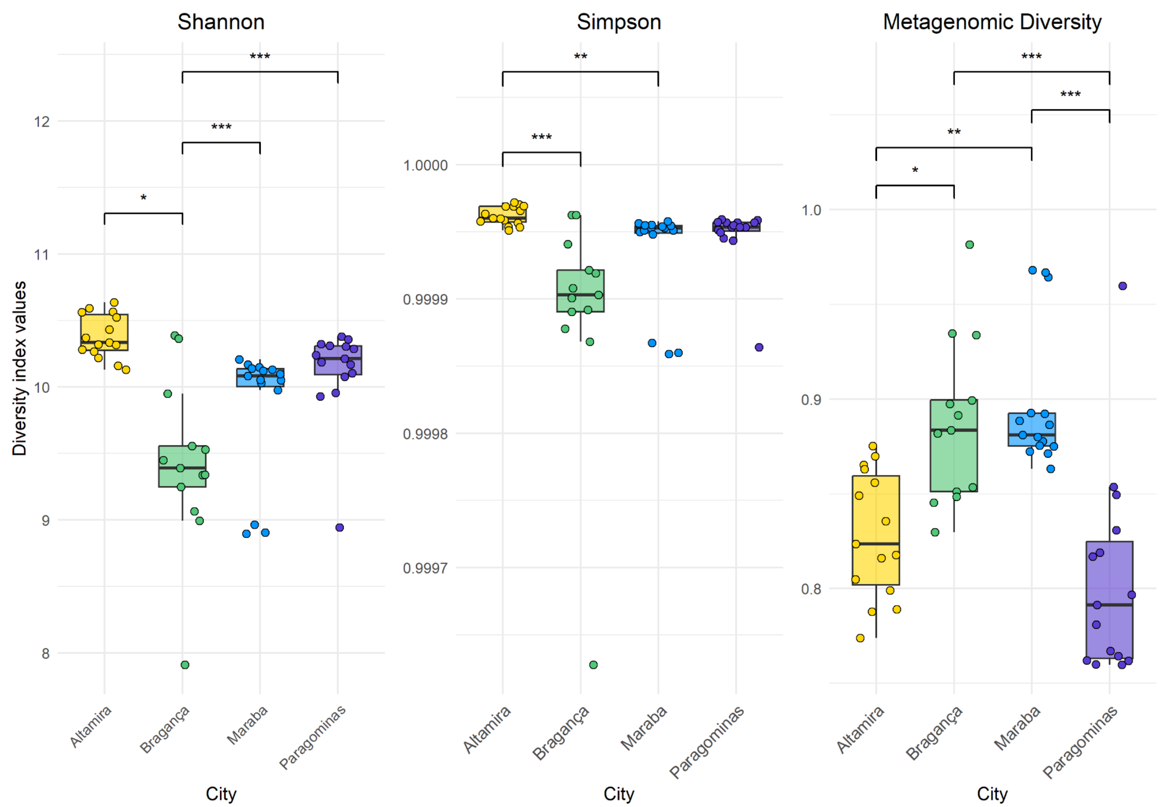
MP4 B	1050487	10.5	0.0137	0	89.5	10.4	10.2	0.0593	0.00181	0.000381
MP4 C	1042508	11.3	0.00835	0	88.7	11.3	11.1	0.0531	0.00144	0.0000959
MP5 A	1051602	9.57	0.00704	0	90.4	9.54	9.35	0.0686	0.00105	0
MP5 B	1062669	9.47	0.00772	0	90.5	9.43	9.23	0.0707	0.00245	0.000376
MP5 C	1044504	9.47	0.00651	0	90.5	9.44	9.24	0.0709	0.00182	0
PP1 A	1051568	13	0.00514	0	87	13	12.7	0.000285	0.0228	0.0000951
PP1 B	1053096	11.7	0.00921	0	88.3	11.7	11.6	0.00256	0.0038	0.00038
PP1 C	1051255	12.6	0.0097	0	87.4	12.6	12.4	0.00124	0.00209	0.000285
PP2 A	1056449	13.6	0.00861	0	86.4	13.6	13.4	0.00151	0.00312	0.0000947
PP2 B	1053226	12.4	0.00484	0	87.6	12.4	12.3	0.00142	0.00389	0
PP2 C	1046375	12.3	0.00812	0	87.7	12.2	12.1	0.000956	0.00296	0.0000956
PP3 A	1052437	12.6	0.00846	0	87.4	12.6	12.4	0.00095	0.00428	0.000475
PP3 B	1043489	12	0.00757	0	88	12	11.8	0.00134	0.00307	0.000575
PP3 C	1048775	13.1	0.00772	0	86.9	13.1	12.9	0.000763	0.00257	0.000191
PP4 A	1050387	14.4	0.0059	0	85.6	14.4	14.2	0.000762	0.00295	0
PP4 B	1043550	14.3	0.00728	0	85.7	14.2	14	0.00163	0.00498	0
PP4 C	1065829	14.7	0.00741	0	85.3	14.7	14.5	0.0015	0.00281	0
PP5 A	1057702	16.9	0.013	0	83.1	16.8	16.7	0.00539	0.00369	0.000567
PP5 B	1042939	16.9	0.0105	0	83.1	16.8	16.6	0.00508	0.0047	0.000384
PP5 C	1051360	16.7	0.0112	0	83.3	16.7	16.5	0.00561	0.00732	0.00019

Supplementary Table 4. Taxonomic distribution and classification summary of metagenomic reads per sample. Overview of Kraken2 and Bracken-based taxonomic assignments across the 58 metagenomic samples. The table includes the number of raw reads, proportions of classified and unclassified reads, and the relative abundance of host, microbial, bacterial, viral, fungal, and protozoan reads. These data provide supporting detail for the taxonomic composition analyses presented in the main text.



Supplementary Figure 2. Taxonomic distribution of microbial communities stratified by environmental and socioeconomic variables. Attribute column charts showing family-level taxonomic composition across samples, grouped by: (A) population density, (B) pH, (C) temperature, and (D) pressure. These metadata groupings were used to explore environmental trends in microbial diversity. The ten most abundant taxa across all samples are listed below each panel.

2.3. *Local Structure Within a Diverse System: Richness, Beta-Diversity, and Community Separation*



Supplementary Figure 3. Bacterial diversity patterns across municipalities based on taxonomic and metagenomic diversity metrics. Boxplots showing bacterial diversity patterns across the four sampled municipalities (Altamira, Bragança, Marabá, and Paragominas) based on three complementary metrics: Shannon diversity (left), Simpson evenness (center), and Metagenomic Diversity Index (MD) (right). Each point represents one sampling site. Shannon diversity reflects taxonomic richness and evenness, Simpson values indicate community evenness, and MD captures protein-clustering-based metagenomic diversity. Altamira exhibited comparatively higher Shannon diversity, whereas Bragança showed lower Shannon values and intermediate MD values. Paragominas presented the lowest MD values, while Marabá showed relatively high MD values overall. Differences among municipalities were assessed using the Kruskal-Wallis test, followed by Dunn's post hoc pairwise comparisons. Significant pairwise differences are indicated in the figure, and full statistical results are provided in Supplementary Table S9.

Diversity metric	Test	Statistic	df	p-value	Interpretation
Shannon diversity	Kruskal–Wallis	χ^2 = 27.600	3	4.406e-06	Significant differences among municipalities
Simpson evenness	Kruskal–Wallis	χ^2 = 24.591	3	1.88e-05	Significant differences among municipalities
Metagenomic Diversity Index (MD)	Kruskal–Wallis	χ^2 = 31.559	3	6.484e-07	Significant differences among municipalities

Pairwise post hoc comparisons for Shannon diversity

Comparison	Mean rank difference	Adjusted p-value	Significance
Bragança vs Altamira	18.810256	0.0197	*
Marabá vs Altamira	-9.866667	0.6574	ns
Paragominas vs Altamira	-1.666667	1.0000	ns
Marabá vs Bragança	-28.676923	4.4e-05	***
Paragominas vs Bragança	-20.476923	0.0082	**
Paragominas vs Marabá	8.200000	1.0000	ns

Pairwise post hoc comparisons for Simpson evenness

Comparison	Mean rank difference	Adjusted p-value	Significance
Bragança vs Altamira	-30.143590	1.5e-05	***
Marabá vs Altamira	-22.266667	0.0018	**
Paragominas vs Altamira	-15.666667	0.0664	.
Marabá vs Bragança	7.876923	1.0000	ns
Paragominas vs Bragança	14.476923	0.1420	ns
Paragominas vs Marabá	6.600000	1.0000	ns

Pairwise post hoc comparisons for Metagenomic Diversity Index

Comparison	Mean rank difference	Adjusted p-value	Significance
Bragança vs Altamira	18.866667	0.01917	*
Marabá vs Altamira	22.733333	0.00136	**
Paragominas vs Altamira	-6.733333	1.00000	ns
Marabá vs Bragança	3.866667	1.00000	ns
Paragominas vs Bragança	-25.600000	0.00038	***
Paragominas vs Marabá	-29.466667	1.1e-05	***

Supplementary Table 5. Additional statistical analyses for bacterial diversity metrics across municipalities. Differences in microbial diversity among municipalities were assessed using Kruskal-Wallis tests followed by Dunn's post hoc multiple comparisons with Bonferroni correction. Diversity metrics included Shannon diversity, Simpson evenness, and the Metagenomic Diversity Index (MD), all derived from taxonomic or protein-clustering-based diversity estimates described in the Supplementary Methods. Significance codes were defined as follows: *** for $p \leq 0.001$; ** for $0.001 < p \leq 0.01$; * for $0.01 < p \leq 0.05$; . for $0.05 < p \leq 0.1$; and *ns* for $p > 0.1$.

Sample	Total Genes	Protein Richness	Shannon Diversity	Simpson Evenness	Log10 Protein Dissimilarity	Metagenomic Diversity
AP1 A	44786	37099	10.432754741667383	0.9999655959172447	4.58011320295247	0.8491236305743052
AP1 B	52550	42754	10.562845913980526	0.9999692974318184	4.642603872658739	0.8356629588614156
AP1 C	54154	43129	10.56316882094871	0.9999690348961885	4.646281707396293	0.8178076462772609
AP2 A	47486	40255	10.523724087009468	0.999968943258015	4.615984895685423	0.8697998080980835
AP2 B	51344	43288	10.590854161607556	0.9999705559862189	4.647661116213829	0.8653093898554852

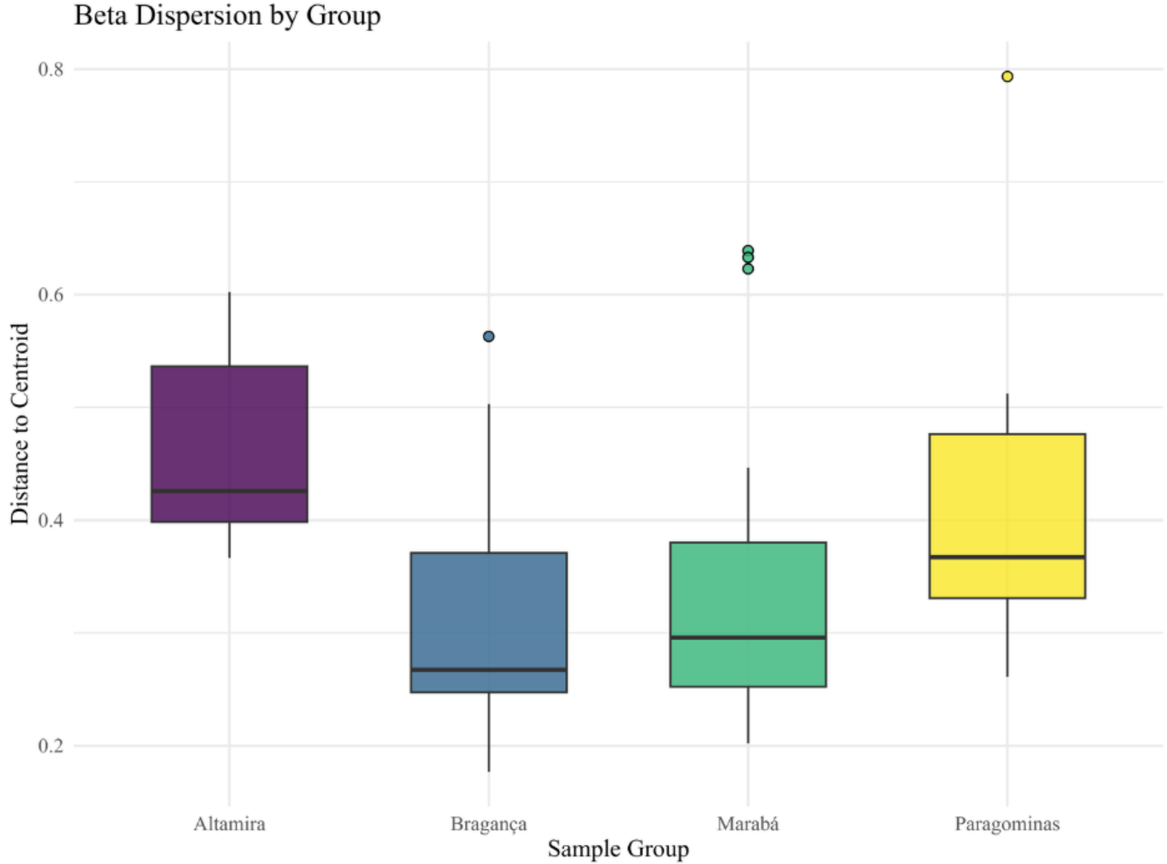
AP2 C	54620	45491	10.636995078897247	0.9999718608564644	4.669733540193413	0.8558188431051984
AP3 A	46852	34802	10.316333154795942	0.9999577412751732	4.559391948915431	0.7738625218636086
AP3 B	40361	31063	10.217752286297426	0.9999536278201355	4.508475236214712	0.7989384418896752
AP3 C	43840	33269	10.281405334921796	0.9999566115049852	4.5390373322481175	0.7891631302183815
AP4 A	36104	28742	10.15966527753337	0.999953377956298	4.473206334378197	0.8234761281378723
AP4 B	45700	34599	10.319976742523513	0.9999585760053583	4.556347462982843	0.7878277297312632
AP4 C	36289	28162	10.130104320351984	0.9999511657851777	4.465530120146044	0.804924449868606
AP5 A	41335	34695	10.370124997397651	0.9999632472187611	4.552326349368531	0.8629952541274342
AP5 B	43656	34488	10.33454121049456	0.9999598461672153	4.551799631334842	0.8161231376184815
AP5 C	36328	30998	10.266972249152595	0.9999600604225152	4.502323171532359	0.8751483587273664
BP1 A	2811	2748	7.910061690731113	0.9996275953854388	3.4407825779972474	0.9815710304755129
BP2 A	9682	8938	9.06519095173031	0.9998777415741199	3.9566190921682507	0.9346607682641327
BP2 B	9021	8318	8.99245894291067	0.9998682116461882	3.925539452876859	0.9338664431142151
BP2 C	13786	12046	9.336356689375586	0.9999006764057834	4.089469271565824	0.89131455269883
BP3 A	12485	11002	9.247835289356683	0.9998919755698876	4.049446587676022	0.897548738772241
BP3 B	15069	12455	9.338512056673494	0.9998904734851901	4.106661691760443	0.848354287141058
BP3 C	16736	13889	9.449496226692368	0.9999030072671627	4.153618880805524	0.8510737852947262
BP4 A	27614	22910	9.947849277392208	0.9999409563372902	4.372306290090941	0.8534478124936985
BP4 B	42500	34831	10.363320716025434	0.9999624075181801	4.555441095639959	0.8453803825261497
BP4 C	44968	36083	10.38671975478204	0.9999623747506936	4.571854402460457	0.8297568190336597
BP5 A	14295	12648	9.390112541580352	0.9999080146414777	4.109161441587427	0.899450529706196
BP5 B	16912	14661	9.52947969026583	0.9999191813732005	4.174434462797338	0.8835670114486192
BP5 C	17391	15036	9.556060713342326	0.999921561164535	4.185708442043032	0.8818280049295704
MP1 A	7731	7433	8.896726829798556	0.99985899147741	3.874134221788001	0.968051692320959
MP1 B	7822	7505	8.905376512963814	0.9998598670182138	3.878643868681859	0.9667764373379115
MP1 C	8339	7982	8.964751472690512	0.9998671931372881	3.9053448638712416	0.9643415737298636
MP2 A	28995	24957	10.05308203378628	0.9999511260088453	4.407315117156028	0.8810326518977373
MP2 B	33007	27756	10.14831079508169	0.9999548615362188	4.4547333836566505	0.8632316008998658
MP2 C	31334	26792	10.120470878431387	0.9999539464265661	4.438299330088371	0.8755551179704567
MP3 A	28062	24318	10.030465493368023	0.9999500055263278	4.395775209101452	0.8864547853480024
MP3 B	34024	29196	10.207688494193684	0.999957684330925	4.476322850895161	0.8801110651872783
MP3 C	30902	26253	10.09846779755264	0.9999529172573609	4.430151334442803	0.8712940459309853
MP4 A	29455	25598	10.083256378529851	0.999952920925407	4.4178159621047834	0.8884989829926198
MP4 B	28333	24743	10.049811939928611	0.9999511213681451	4.40279865141726	0.8922901314436141
MP4 C	26157	22834	9.974283283393454	0.9999482626036288	4.368294340433401	0.8927017950368077
MP5 A	32101	27327	10.139931661064523	0.9999551159671836	4.447270927234168	0.8724737426645779
MP5 B	31440	26966	10.128897218261457	0.9999545898407886	4.44088884732715	0.877821694817587
MP5 C	32781	27994	10.1666598484392	0.999956462321991	4.457656918960643	0.8750537316076303
PP1 A	8203	7806	8.941798042144185	0.9998641265180557	3.8960982790142062	0.9596780629038157
PP1 B	50730	36886	10.357284437176155	0.9999588052758477	4.587154529696266	0.7618854282541728
PP1 C	41918	31342	10.21282661803508	0.999953418535572	4.5150238251940165	0.7809500766502436
PP2 A	48795	35498	10.322091380073243	0.999957247254283	4.5703341606524805	0.7620067849085229

PP2 B	51387	37555	10.379690710624285	0.9999593932137949	4.594123642287907	0.7643114719775246
PP2 C	37119	29170	10.167990060708762	0.9999534549366363	4.481735787204464	0.8168447452065304
PP3 A	39340	30072	10.185514064665973	0.9999533015248196	4.49618694586335	0.7968059900685652
PP3 B	47440	34745	10.304850527345168	0.9999570527770042	4.560992466582099	0.767092587667692
PP3 C	47039	34157	10.28697904571646	0.9999568292418671	4.553298402925969	0.7600467782156045
PP4 A	48351	35045	10.310012548675793	0.9999569687222615	4.565096536260872	0.7597856304266037
PP4 B	33747	27038	10.103551326448095	0.9999514604623342	4.447703228396746	0.8307456136967195
PP4 C	42193	31977	10.239967703105428	0.9999549594024784	4.5236003394360145	0.791334196869204
PP5 A	33630	26518	10.076809888029025	0.9999496493303696	4.439975661788938	0.8189337967521955
PP5 B	27864	22942	9.954053858825358	0.9999451674574594	4.374252907450317	0.8495901057261119
PP5 C	27028	22363	9.928588076648946	0.9999434073145035	4.363075977668268	0.853615057723795

Supplementary Table 6. Alpha diversity metrics for all individual samples. Complete set of alpha diversity results calculated per sample (n = 58), including total predicted genes, protein richness, Shannon diversity, Simpson evenness, protein dissimilarity, and metagenomic diversity (MD). These values reflect the taxonomic and functional diversity captured by each metagenome and serve as the basis for site-level summary statistics presented in Table 3.

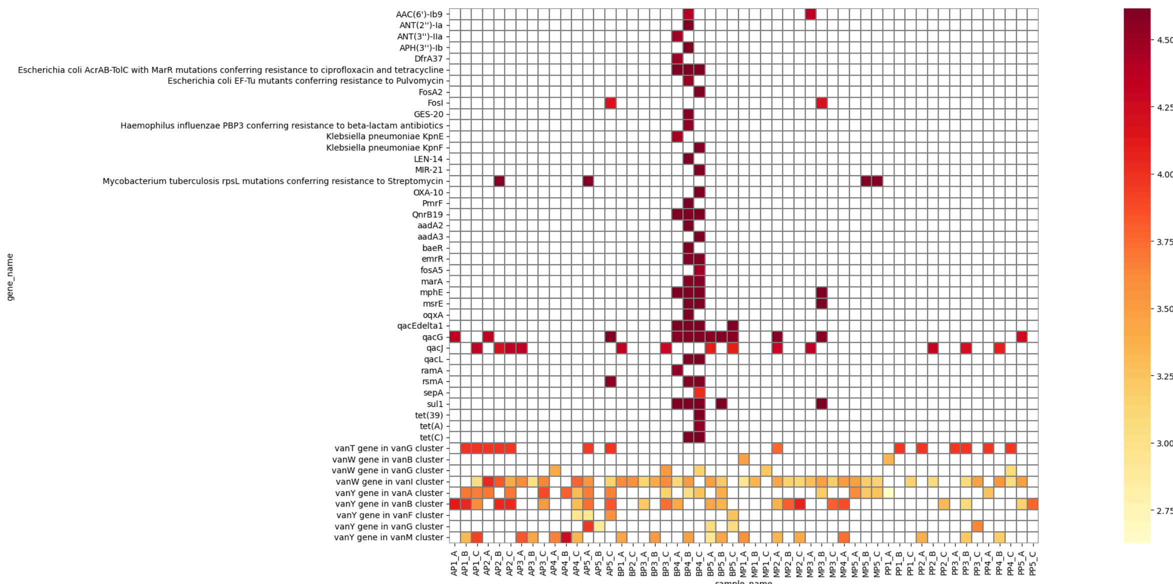
Test	Source	Df	Sum of Squares	Mean Squares	F-value	R^2	N.Perm	p-value	Significance
PERMANOVA	Model	3	6.7397	---	12.361	0.40714	999	0.001	***
	Residual	54	9.814	---	---	0.59286	---	---	
	Total	57	16.5537	---	---	1.0	---	---	
PERMDISP	Groups	3	0.17098	0.056994	3.5098	---	999	0.025	*
	Residual	54	0.87689	0.016239	---	---	---	---	

Supplementary Table 7. PERMANOVA and PERMDISP results for beta diversity validation. Summary of statistical validation tests for beta diversity results. PERMANOVA was performed using the adonis2 function, and PERMDISP using the permutest function, both from the vegan package in R. Significance codes: '***' for $p \leq 0.001$; '**' for $0.001 < p \leq 0.01$; '*' for $0.01 < p \leq 0.05$; '.' for $0.05 < p \leq 0.1$; no symbol for $p > 0.1$.

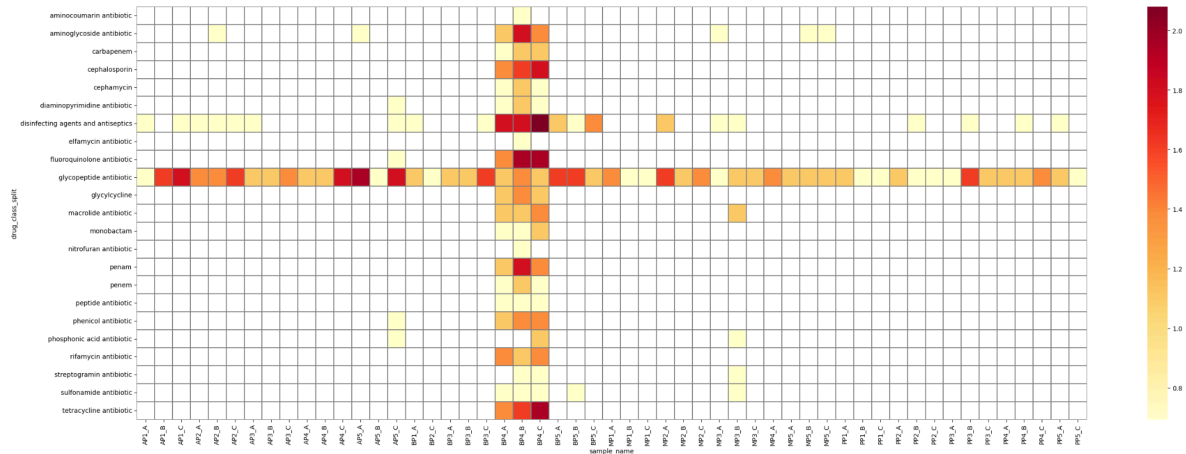


Supplementary Figure 4. Beta dispersion analyses of microbial communities by municipality. Boxplot showing beta dispersion (distance to centroid) for each city group. These results reflect within-group variation in community composition. Altamira and Paragominas display greater internal variability compared to Bragança and Marabá.

2.4. *Pathogens Detected Across the Region — But They Don't Explain AMR Alone*

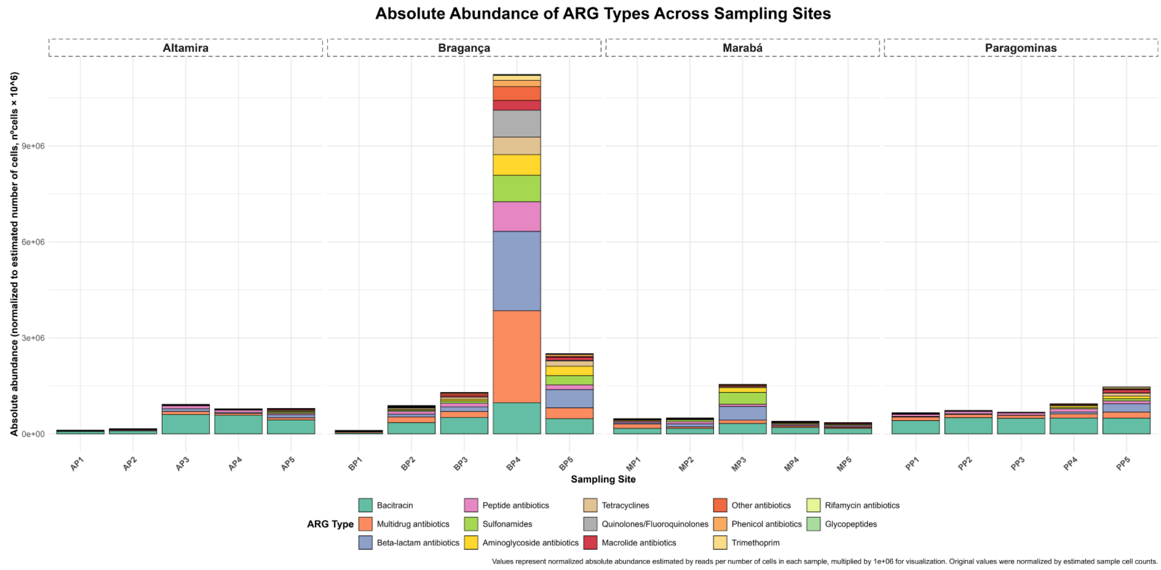


Supplementary Figure 5. Heatmap of antimicrobial resistance genes across all metagenomic samples. Log-transformed abundance ($\log_{10}$(RPM + 1)) of detected antimicrobial resistance genes (ARGs) across 58 samples. ARGs are displayed on the y-axis and samples on the x-axis, clustered by hierarchical similarity. Clinically relevant resistance determinants, including *tet*, *sul*, *qac*, and *van* gene families, were detected at variable levels across cities, with Bragança and Paragominas exhibiting higher ARG signal intensity. Only ARGs meeting filtering thresholds (nucleotide rPM ≥ 10, protein rPM ≥ 1, alignment length ≥ 50 bp) are displayed.

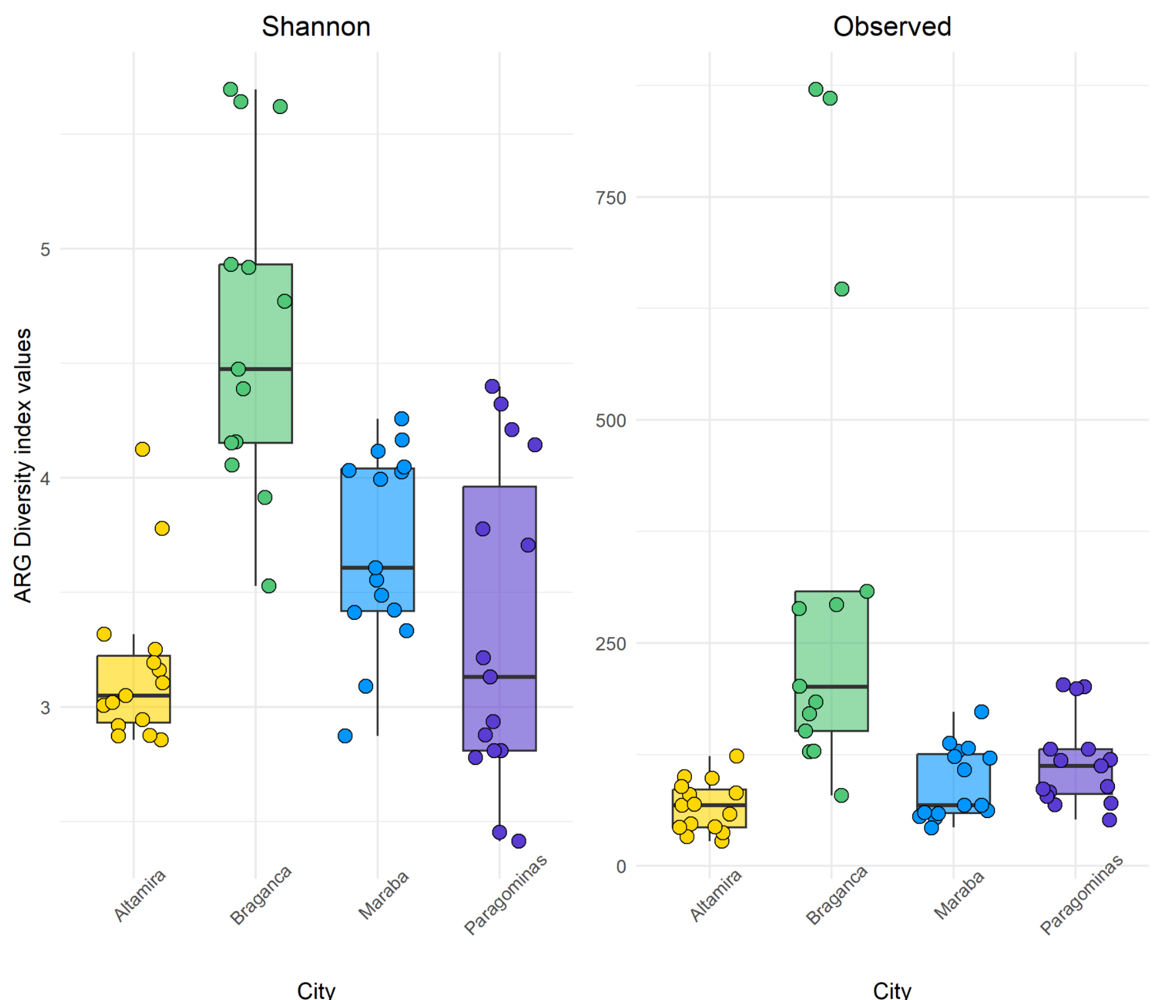


Supplementary Figure 6. Heatmap of antimicrobial resistance drug classes across all metagenomic samples. Log-transformed abundance ($\log_{10}$(RPM + 1)) of antimicrobial resistance drug classes, inferred from gene annotations, across 58 samples. Bragança and Paragominas showed enriched profiles for resistance to critical antibiotics such as glycopeptides, fluoroquinolones, and disinfectants. Only drug classes associated with genes meeting filtering thresholds (nucleotide rPM ≥ 10, protein rPM ≥ 1, alignment length ≥ 50 bp) are displayed.

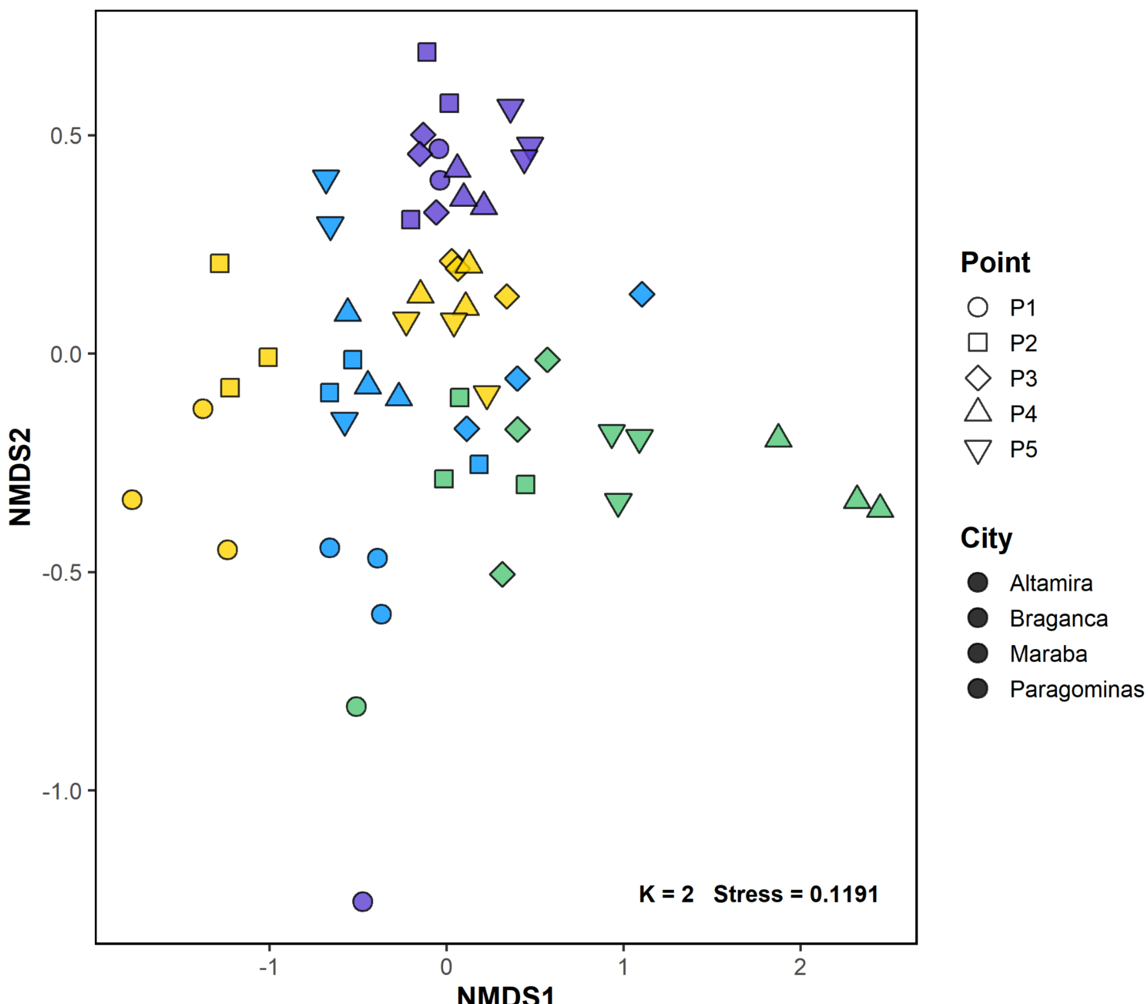
2.5. *Antimicrobial Resistance Genes Are Widespread, Even in Seemingly Pristine Sites*



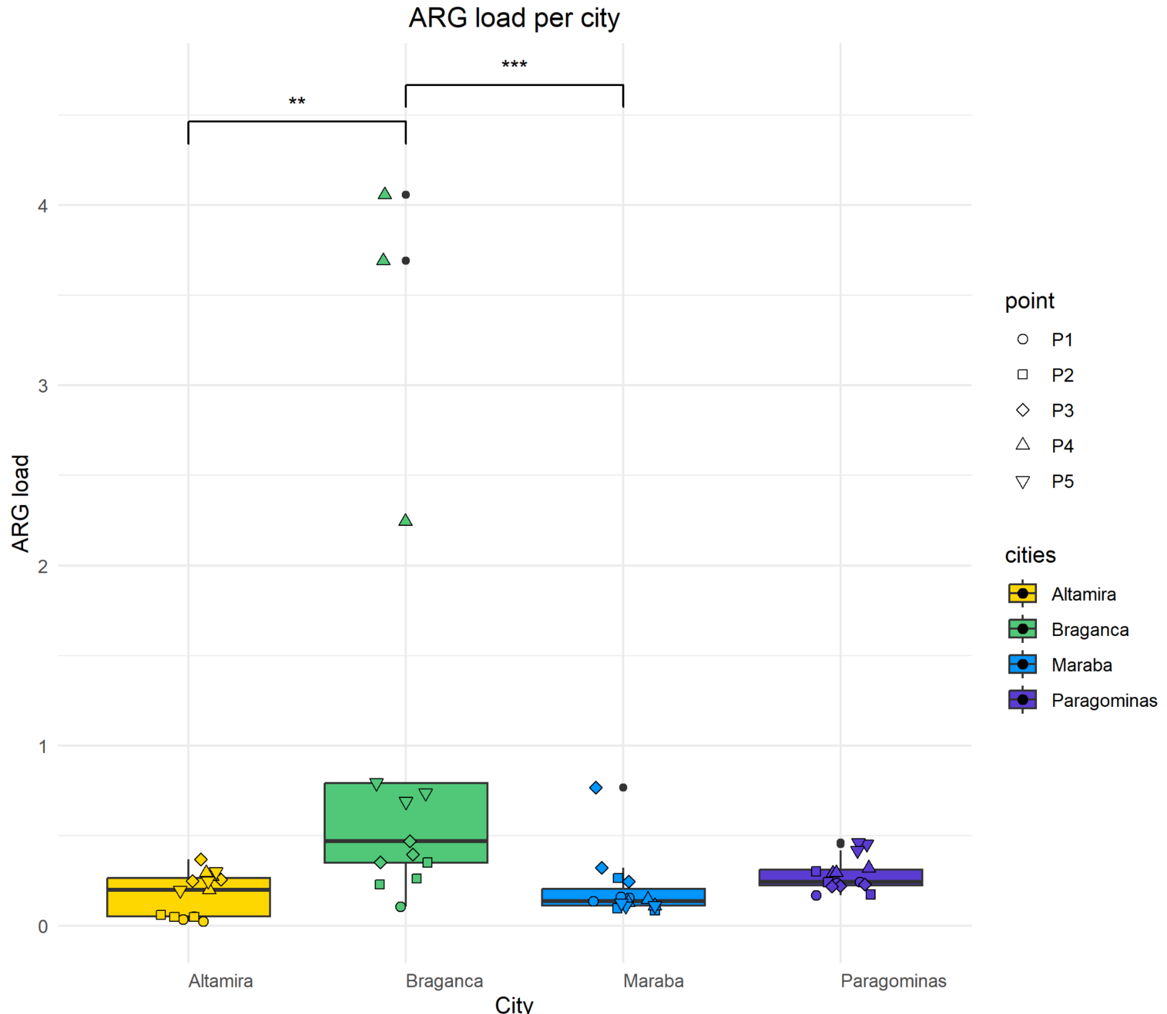
Supplementary Figure 7. Absolute abundance of antimicrobial resistance gene (ARG) types across sampling sites. Stacked bar plots represent normalized absolute ARG abundance per sampling site, calculated as ARG copies per million estimated cells (rescaled by 10^6). Each color indicates a distinct ARG type. Bragança's BP4 and BP5, and Marabá's MP3 sampling sites exhibit the highest burden across nearly all ARG categories, underscoring localized amplification of resistance potential.



Supplementary Figure 8. Alpha diversity of antimicrobial resistance gene profiles across municipalities. Boxplots showing alpha diversity patterns of antimicrobial resistance gene (ARG) profiles across the four sampled municipalities: Altamira, Bragança, Marabá, and Paragominas. Two complementary metrics are shown: Shannon diversity (left), reflecting ARG profile diversity and evenness, and Observed richness (right), indicating the number of ARG features detected per sample. Each point represents one metagenomic sample, and colors indicate municipality. Bragança showed the highest ARG alpha diversity and observed richness overall, whereas Altamira displayed lower and less variable values. Differences in ARG alpha diversity among municipalities were evaluated using the Kruskal-Wallis test and were not statistically significant overall (Supplementary Table S8).



Supplementary Figure 9. Non-metric multidimensional scaling (NMDS) ordination of ARG composition across sampling sites. Non-metric multidimensional scaling (NMDS) ordination based on Bray-Curtis dissimilarity of normalized antimicrobial resistance gene (ARG) profiles across the 58 metagenomic samples. Colors represent municipalities (Altamira, Bragança, Marabá, and Paragominas), and point shapes indicate sampling sites within each city (P1–P5). The ordination was generated in two dimensions (K = 2) with a final stress value of 0.1191, indicating an acceptable representation of dissimilarity relationships in reduced space. Samples from different municipalities showed partial clustering with substantial overlap, consistent with moderate but significant differences in resistome composition among cities as supported by ANOSIM analysis (R = 0.3618, p = 0.001; Supplementary Table S8).



Supplementary Figure 10. Differences in normalized ARG load across municipalities. Boxplots showing the distribution of normalized antimicrobial resistance gene (ARG) load across the four sampled municipalities: Altamira, Bragança, Marabá, and Paragominas. Each point represents one sampling site, and point shapes indicate site identity (P1–P5) within each city. ARG load values represent normalized abundance estimates derived from metagenomic profiles. Bragança showed the highest overall ARG load and the greatest variability, including several high-burden sites, whereas Altamira and Marabá displayed lower and more homogeneous values. Differences among municipalities were evaluated using the Kruskal-Wallis test, followed by Dunn's post hoc pairwise comparisons. Significant differences were detected between Altamira and Bragança and between Bragança and Marabá. Detailed statistical results are provided in Supplementary Table S8.

Analysis	Response variable	Test	Statistic	df	p-value	Interpretation
ARG load among municipalities	Normalized ARG abundance	Kruskal-Wallis	$\chi^2 = 19.862$	3	0.0001813	Significant differences among municipalities
ARG alpha diversity among municipalities	Shannon diversity of normalized ARG profiles	Kruskal-Wallis	—	3	> 0.05	No significant differences among municipalities
ARG beta diversity among municipalities	Bray-Curtis dissimilarity of ARG composition	ANOSIM	R = 0.3618	—	0.001	Significant but moderate separation among municipalities

Pairwise post hoc comparisons for ARG load

Comparison	Mean rank difference	Adjusted p-value	Significance
Bragança vs Altamira	23.67692	0.00129	**
Marabá vs Altamira	-1.20000	1.00000	ns
Paragominas vs Altamira	12.00000	0.30986	ns
Marabá vs Bragança	-24.87692	0.00061	***
Paragominas vs Bragança	-11.67692	0.40818	ns
Paragominas vs Marabá	13.20000	0.19379	ns

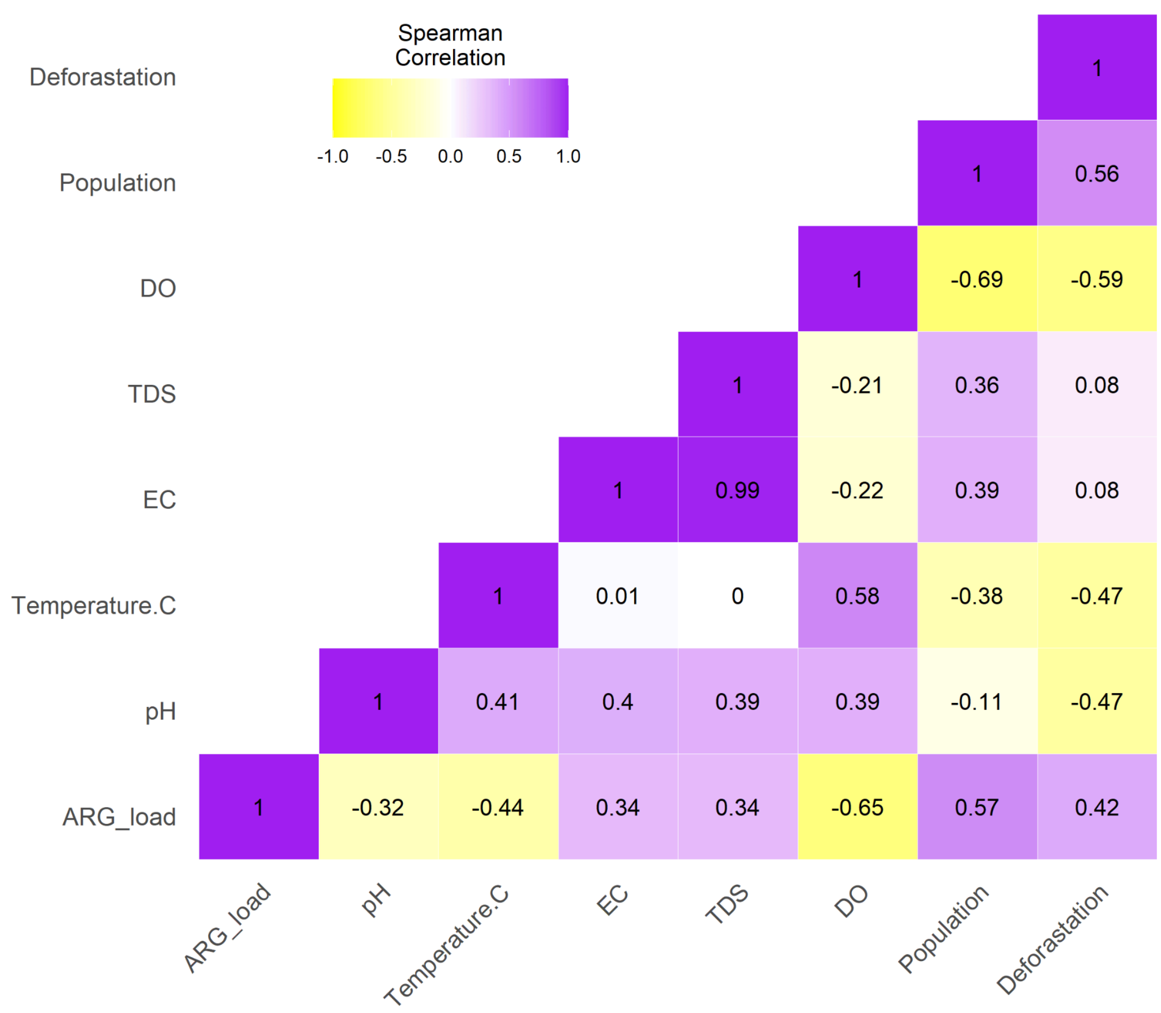
Supplementary Table 8. Additional statistical analyses for antimicrobial resistance gene (ARG) load and ARG diversity across municipalities. Differences in normalized ARG abundance among municipalities were assessed using the Kruskal-Wallis test, followed by Dunn's post hoc multiple comparisons with Bonferroni correction. ARG alpha diversity was evaluated using Shannon diversity indices calculated from normalized ARG abundance matrices. Differences in ARG composition among municipalities were assessed using ANOSIM based on Bray-Curtis dissimilarity matrices with 999 permutations. The ANOSIM result was significant, although the low R statistic indicates substantial overlap in ARG composition among cities. Significance codes were defined as follows: *** for $p \leq 0.001$; ** for $0.001 < p \leq 0.01$; * for $0.01 < p \leq 0.05$; . for $0.05 < p \leq 0.1$; and *ns* for $p > 0.1$.

2.6. *High-Risk Resistance Hotspots Are Concentrated in Bragança and Marabá*

	Percentage of ARG Abundance Load per Risk Rank								Standard Deviation (SD)					
Site	Rank I	Rank II	Rank III	Rank IV	Total abundance	Weighted Risk	**Threat Score**	**Log Threat Score**	Rank I	Rank II	Rank III	Rank IV	Total abundance	Weighted Risk
AP1	0,000	0,000	0,000	1,000	0,00057	4,00	0,00E+00	-10,00	0,00E+00	0,00E+00	6,35E-03	5,77E-03	1,85E-04	6,34E-03
AP2	0,000	0,000	0,000	1,000	0,00078	4,00	0,00E+00	-10,00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	8,53E-05	0,00E+00
AP3	0,005	0,000	0,008	0,985	0,00200	3,97775	9,60E-06	-5,017724	3,91E-03	4,79E-02	6,47E-03	5,51E-02	2,47E-04	1,11E-01
AP4	0,000	0,004	0,013	0,980	0,00210	3,97225	9,14E-06	-5,039287	0,00E+00	6,08E-03	8,08E-03	1,04E-02	2,08E-04	1,75E-02
AP5	0,012	0,033	0,057	0,905	0,00185	3,85436	8,23E-05	-4,084468	1,97E-02	2,08E-02	3,59E-02	7,29E-02	2,52E-04	1,27E-01
BP1	0,000	0,000	0,078	0,925	0,00190	3,92269	0,00E+00	-10,00	---	---	---	---	---	---
BP2	0,018	0,015	0,032	0,935	0,00230	3,88400	7,59E-05	-4,119758	1,12E-02	9,54E-03	6,05E-03	3,00E-02	2,02E-04	5,84E-02
BP3	0,022	0,042	0,075	0,840	0,00250	3,73127	1,61E-04	-3,792500	6,33E-03	1,95E-02	3,18E-02	4,54E-02	2,02E-04	7,08E-02
BP4	0,046	0,062	0,200	0,690	0,00635	3,53912	6,79E-04	-3,167842	7,26E-03	4,86E-03	0,00E+00	1,53E-02	7,09E-04	3,29E-02
BP5	0,060	0,080	0,140	0,715	0,00260	3,51122	3,64E-04	-3,438898	1,06E-02	1,32E-02	1,04E-02	2,93E-02	2,02E-04	5,98E-02
MP1	0,016	0,014	0,017	0,955	0,00180	3,90446	5,40E-05	-4,267605	1,22E-02	4,35E-03	8,70E-03	2,89E-03	1,26E-04	2,64E-02
MP2	0,000	0,023	0,004	0,980	0,00135	3,95414	3,11E-05	-4,507937	2,08E-02	3,84E-02	7,97E-02	1,37E-01	6,79E-04	2,16E-01
MP3	0,060	0,078	0,145	0,730	0,00285	3,55800	3,93E-04	-3,405276	1,28E-02	5,00E-02	5,36E-02	1,11E-01	1,05E-03	1,80E-01
MP4	0,000	0,016	0,042	0,915	0,00155	3,86920	2,40E-05	-4,619335	2,37E-02	2,26E-02	3,86E-02	8,66E-03	1,61E-04	4,13E-02
MP5	0,021	0,019	0,019	0,915	0,00155	3,86613	6,12E-05	-4,213070	6,25E-03	3,37E-02	2,66E-02	2,75E-02	7,64E-05	6,58E-02
PP1	0,000	0,000	0,004	0,995	0,00330	3,99171	0,00E+00	-10,00	0,00E+00	1,41E-03	1,51E-02	1,61E-02	1,05E-03	1,38E-02

PP2	0,000	0,000	0,006	0,995	0,00365	3,99450	0,00E+00	-10,00	0,00E+00	0,00E+00	1,09E-02	1,04E-02	7,85E-04	1,09E-02
PP3	0,000	0,000	0,009	0,990	0,00305	3,99059	0,00E+00	-10,00	0,00E+00	0,00E+00	3,40E-03	2,89E-03	2,65E-04	3,39E-03
PP4	0,023	0,015	0,042	0,920	0,00320	3,84858	1,22E-04	-3,915066	1,27E-02	5,35E-03	1,15E-02	7,64E-03	2,84E-04	2,27E-02
PP5	0,043	0,056	0,140	0,765	0,00320	3,62052	3,17E-04	-3,499215	1,73E-02	1,01E-02	2,81E-02	5,75E-02	3,06E-04	9,73E-02

Supplementary Table 9. Summary of ARG Threat Score Calculations and Ranking Metrics. Table detailing the total normalized abundance of antimicrobial resistance genes (ARGs), their proportional distribution across risk ranks (Rank I–IV), calculated Threat Scores, and log-transformed values log10(ThreatScore+ε) for each sample. Risk ranks were assigned according to established criteria based on ARG mobility, clinical detection, and host association. All values used to generate Figure 5 are provided for transparency and reproducibility. The Threat Scores were calculated by summing the relative abundance of Rank I and Rank II ARGs per sample, weighted by total ARG abundance: Threat Score = $(RA_{(Rank\ I)} + RA_{(Rank\ II)}) \times$ Total ARGs.



Supplementary Figure 11. Spearman correlation matrix between ARG load and environmental variables across sampling sites. Heatmap showing Spearman rank correlation coefficients (ρ) among normalized antimicrobial resistance gene (ARG) load and environmental variables measured or compiled for each sampling site, including pH,

temperature, electrical conductivity (EC), total dissolved solids (TDS), dissolved oxygen (DO), population density, and deforestation percentage. Correlation coefficients are displayed within each cell and range from -1 to 1, with purple tones indicating positive associations and yellow tones indicating negative associations. ARG load showed moderate positive correlations with population density ($\rho = 0.57$) and deforestation ($\rho = 0.42$), and a moderate negative correlation with dissolved oxygen ($\rho = -0.65$). These patterns support the association between resistome burden and anthropogenic or ecologically disturbed freshwater environments.

Mapping the Resistome: Metagenomic Surveillance Reveals AMR Hotspots and Ecological Risk in Amazonian Freshwaters

Wylerson Guimarães Nogueira[1,2], David Tavares Martins[3,4,5], Carlos Willian Dias Dantas[3,4,5], Victor Benedito Costa Ferreira[3,4,5], Oscar Victor Cardenas-Alegria[3,4,5], Edian Franklin Franco de Los Santos[6], Paulo Rogenes Monteiro Pontes[7], Rosane Barbosa Lopes Cavalcante[7], Júnia Schultz[2], Vasco Ariston de Carvalho Azevedo[8], Alexandre Soares Rosado[2], Rommel Thiago Jucá Ramos[3,4,5*]

[1]Department of Biochemistry and Immunology, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[2]Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[3]Center of High Performance Computer and Artificial Intelligence, Federal University of Pará, Belém, Pará, Brazil
[4]Laboratory of Bioinformatics and Genomics of Microorganisms, Federal University of Pará-UFPA, Belém, Pará, Brazil
[5]Laboratory of Simulation and Computational Biology (SIMBIC), Federal University of Pará, Belém, Pará, Brazil
[6]Universidade Tecnológica de Santiago, Santo Domingo, Republic Dominican.
[7]Vale Institute of Technology – Sustainable Development, Belém, Pará, Brazil
[8]Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil

*** Correspondence:**
Rommel Thiago Jucá Ramos
rommelthiago@gmail.com

Keywords: Metagenomics, Antimicrobial Resistance (AMR), Resistome, Freshwater Microbiomes, Amazon Basin, Waterborne Pathogens, One Health.

Supplementary Methods

Resistome of Amazonian Freshwaters

1. Supplementary Methods

1.1. *Study Area, Sample Collection, and Metadata Acquisition*

Surface freshwater samples were collected from four municipalities in the Brazilian state of Pará — Altamira, Bragança, Marabá, and Paragominas — between September and November 2022. Five sampling sites were selected within each municipality based on varying degrees of anthropogenic influence. At each site, three biological replicates were collected at a depth of 20 cm, approximately 1 meter from the riverbank or channel center, totaling 60 samples across 20 sites. For each replicate, 5 liters of surface water were collected using a Van Dorn bottle, stored on ice, and immediately transported to the laboratory for processing.

Samples were filtered using a negative pressure vacuum system with 47 mm nitrocellulose membranes (0.22 µm pore size). Membranes were transferred to 50 mL Falcon tubes containing TSE buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA) and stored at −20°C until DNA extraction. In situ, physicochemical parameters were recorded at the time of collection using a multiparameter water analyzer (HI 9829, HANNA Instruments), including pH, temperature (°C), pressure (mbar), electrical conductivity (EC and ECabsolute), resistivity, total dissolved solids (TDS), salinity, oxidation-reduction potential (ORP), and dissolved oxygen (DO%, DO ppm).

Municipality-level socioeconomic indicators, such as population density, GDP per capita, and infant mortality, were obtained from the Brazilian Institute of Geography and Statistics (IBGE) Cities and States Portal (https://www.ibge.gov.br/cidades-e-estados/) [1]. Deforestation levels were determined through geospatial analysis following the methodology of Cavalcante et al. [2], using MapBiomas Brazil land-use and land-cover data (Collection 6, year 2020). The deforested area was calculated as the total proportion of non-natural land cover associated with each sampling site within the hydrographical basin.

To support spatial analyses and ecological contextualization of each sampling site, geographic coordinates (latitude and longitude) were recorded at the time of sampling using a handheld GPS device. Sampling sites were selected to represent a gradient of anthropogenic influence across the municipalities of Altamira, Bragança, Marabá, and Paragominas in Pará, Brazil. Sites were named using a city code (e.g., AP1 = Altamira Point 1), and each site was sampled in biological triplicate. Coordinates listed in Supplementary Methods Table 01 correspond to the exact central point at which triplicate samples were collected.

The combined metadata, including physicochemical parameters, socioeconomic indicators, and spatial ecological context, was integrated into downstream multivariate analyses.

	ALTAMIRA			
	Latitude	Longitude	Elevation	Altitude
Palt 1	-3,20889	-52,24922	106	128,27
Palt 2	-3,20216	-52,23962667	109,2	131,47
Palt 3	-3,2009152	-52,2161569	83,4	105,67
Palt 4	-3,2003535	-52,2093526	79,9	102,17
Palt 5	-3,2020395	-52,2023505	111,7	133,97

	BRAGANÇA			
	Latitude	Longitude	Elevation	Altitude
Pbra 1	-1,068325	-46,791195	12,3	37,69
Pbra 2	-1,0565863	-46,7767607	0,4	25,79
Pbra 3	-1,0537596	-46,771585	-151,6	9,79
Pbra 4	-1,0508683	-46,7651933	-19	6,39
Pbra 5	-1,0481952	-46,7598743	-20,9	4,49

	MARABÁ			
	Latitude	Longitude	Elevation	Altitude
Pmar 1	-5,34653667	-49,20456833	79,6	104,09
Pmar 2	-5,402505	-49,13356167	72,8	97,29
Pmar 3	-5,3660059	-49,0935374	75,1	99,59
Pmar 4	-5,35686833	-49,12218167	65	89,49
Pmar 5	-5,35506333	-49,14269833	65,8	90,29

	PARAGOMINAS			
	Latitude	Longitude	Elevation	Altitude
Ppar 1	-3,05792	-47,41727	77,3	103,34
Ppar 2	-3,0068714	-47,3813726	55	81,04
Ppar 3	-2,9989622	-47,371593	71,7	97,78
Ppar 4	-2,9824708	-47,3605382	50,2	76,28
Ppar 5	-2,9650469	-47,3375389	73,1	99,18

Supplementary Methods Table 01. Geographic coordinates of freshwater sampling sites. Table listing the georeferenced coordinates (latitude and longitude) of the 20 freshwater sampling sites across the municipalities of Altamira, Bragança, Marabá, and Paragominas, Pará, Brazil. Site codes correspond to labels used throughout the manuscript.

1.2. *DNA Extraction, Library Preparation, and Metagenomic Sequencing*

Upon transfer to the central laboratory following field filtration at local facilities, the filter membranes were incubated at room temperature under constant agitation (250 RPM) to facilitate the release of retained particles. The resulting suspension was centrifuged at 13,200 RPM for 10 minutes in Eppendorf tubes. DNA was extracted from the resulting pellet using the DNeasy PowerSoil Pro Kit (QIAGEN), following the manufacturer's protocol. DNA concentration was quantified using a Qubit fluorometer (Thermo Fisher Scientific), and purity indices were assessed with a NanoDrop spectrophotometer (Thermo Fisher Scientific). General DNA integrity was verified by agarose gel electrophoresis. Samples meeting minimum quality thresholds (concentration ≥50 ng/μL and a 260/280 ratio between 1.4 and 2.0) were retained for downstream analysis (n = 58). Metagenomic libraries were constructed using the Nextera XT DNA Library Preparation Kit (Illumina)

according to the manufacturer's protocol. Sequencing was performed on the Illumina NextSeq 550 platform using the High-Output kit, generating 150 bp paired-end reads. Raw reads were delivered in FASTQ format for subsequent preprocessing and analysis.

1.3. *Quality Control and Preprocessing of Sequencing Data*

All raw paired-end metagenomic reads were processed using the CZ ID metagenomics pipeline (Illumina mNGS Pipeline v8.3) [3], a cloud-based bioinformatics platform designed for microbial identification and host-read depletion. The following preprocessing steps were automatically performed as part of the CZ ID host filtering and quality control workflow:

- Adapter trimming and low-quality read removal were performed using fastp, with default parameters;
- Low-complexity filtering was applied using a modified version of the SDUST algorithm, removing sequences with >40% low-complexity regions;
- Short-read filtering removed reads shorter than 35 bp and those containing >15 undetermined bases (Ns);
- Host read depletion included sequential alignment of reads to reference genomes using Bowtie2 (for initial host and ERCC alignment) and HISAT2 (for secondary human depletion), based on the GRCh38 reference;
- Read deduplication was performed using the czid-dedup module, retaining a single representative from any group of reads found to be 100% identical over the first 70 bp;
- ERCC spike-in removal was conducted by aligning to ERCC sequences using Bowtie2;
- Subsampling was performed to a maximum of 2 million read pairs per sample for normalization across analyses.

All resulting non-host, quality-filtered reads were retained for downstream analyses, including taxonomic classification, AMR gene profiling, and ecological modeling. Full documentation of the pipeline is available at: https://czid.org.

1.4. *Taxonomic Classification and Microbial Diversity Analysis*

Taxonomic classification of the filtered metagenomic reads was performed using Kraken2 (Galaxy Version 2.1.3+galaxy1) via the Galaxy Europe platform [4], [5]. Input was provided as a paired-end read collection in FASTQ format, and classification was performed against the Prebuilt RefSeq PlusPF database (Version: 2024-01-12), which

includes bacteria, archaea, viruses, protozoa, and fungi. A confidence score threshold of 0.1 was applied, meaning a minimum of 10% of the k-mers in a read had to match a given taxon. Refined species-level abundance estimates were generated using Bracken (Galaxy Version 3.1+galaxy0), selecting the corresponding PlusPF database (K-mer = 35, Read Length = 100 bp) [6]. Results were visualized using Pavian (Galaxy Version 1.0) and Phinch (Galaxy Version 0.1) [7], [8], with the latter integrating sample metadata to generate bubble charts, bar charts, and attribute column plots.

To evaluate beta diversity, Krakentools (Galaxy Version 1.2+galaxy2) was used to generate Bray-Curtis dissimilarity matrices from the Bracken output files [9]. These matrices were used for Principal Coordinates Analysis (PCoA), performed in R using the vegan package (v2.6-10) [10]. Significance of community differences between groups was assessed using PERMANOVA (adonis2) and PERMDISP (permutest), with 999 permutations for both. Alpha diversity indices were calculated using the MD.py script developed by Damien Finn (available at: https://github.com/DamienFinn/MD/blob/main/MD.py) [11]. This tool computes multiple diversity metrics — including species richness, Shannon, Simpson Evenness, protein-coding gene richness, protein dissimilarity, and overall metagenomic diversity — based on MMSeqs2 clustering output. To generate the necessary inputs for MD.py, metagenomic assemblies were constructed for each sample using MEGAHIT (Galaxy Version 1.2.9+galaxy2) [12]. Assembly was performed using the k-mer strategy: min = 21, max = 91, step = 12. Following assembly, Prodigal was used to predict open reading frames and generate .faa protein sequence files for each sample [13]. These were then processed using MMSeqs2 and submitted to MD.py to calculate the diversity metrics.

Triplicates were aggregated using medians per sampling site for alpha diversity comparisons, while individual samples (n = 58) were retained for beta diversity analysis to preserve site-level variability.

1.5. *Detection and Quantification of Antimicrobial Resistance Genes (ARGs)*

Detection of antimicrobial resistance genes (ARGs) was performed using two independent platforms to enable both sensitive detection and accurate quantification. First, ARG identification was carried out using the CZ ID AMR pipeline (v1.4.2), which applies the Resistance Gene Identifier (RGI) (v6.0.0) to both read and contig-level inputs [14]. Alignments were performed against the Comprehensive Antibiotic Resistance Database

(CARD v3.2.6) and the WildCARD (v4.0.0) module. WildCARD supplements the canonical CARD database with computationally predicted resistomes and variant data, increasing sensitivity to uncharacterized or emerging ARGs. These results were used to generate heatmaps of ARG presence and resistance drug classes, as well as species-level profiles of pathogenic bacteria.

To assess ARG abundance and ecological burden, the same samples were also analyzed using ARGs-OAP (v3.2.4), a pipeline specifically designed for large-scale ARG annotation and quantification [15]. The platform utilizes the SARG database (release v3.0), which integrates curated ARG sequences from both CARD and the archived ARDB. The database supports hierarchical classification into ARG type, subtype, and reference sequence, and annotation output is provided at all three levels.

Abundance normalization was performed by estimating microbial cell numbers based on 16S rRNA gene profiles. To this end, ARGs-OAP constructs a hypervariable region (HVR) database using Greengenes, and aligns reads via USEARCH to extract 16S-like sequences. These are taxonomically classified to estimate the community structure, and lineage-specific copy number correction is applied using the CopyRighter database. This method outputs ARG copy numbers normalized by estimated cell number (i.e., ARGs per million cells), accounting for both taxonomic composition and gene copy variation [16], [17]. Normalized abundance tables were used to generate all site-level visualizations: ARG-type prevalence heatmaps, stacked bar plots (relative and absolute), boxplots, and radar charts. This dual-pipeline strategy allowed for robust detection of ARG diversity (via CZ ID + CARD/WildCARD) and comprehensive quantification of ARG burden (via ARGs-OAP + SARG).

To further characterize resistome structure across municipalities, alpha diversity of ARG profiles was calculated from normalized ARG abundance matrices using Shannon diversity and observed richness metrics. Beta diversity of ARG composition was assessed using Bray-Curtis dissimilarity, and ordination was performed using non-metric multidimensional scaling (NMDS). Differences in ARG load among municipalities were evaluated using the Kruskal-Wallis test followed by Dunn's post hoc multiple comparisons, while differences in ARG composition were tested using analysis of similarities (ANOSIM). These analyses supported the supplementary visualizations and statistical summaries presented in Supplementary Figures S7–S10 and Supplementary Table S8.

1.6. *ARG Risk Ranking and Threat Score Calculation*

Risk classification of antimicrobial resistance genes (ARGs) was performed using ARG_ranker (v3.7.2), a tool that evaluates ARGs according to their potential to contribute to the spread of clinically relevant resistance [18]. The classification is based on three criteria: gene mobility, enrichment in human-associated environments, and presence in known pathogenic hosts. Based on these criteria, ARGs are assigned to one of four risk categories: Rank I represents high-risk ARGs already detected in human pathogens and associated with mobile genetic elements; Rank II includes ARGs that are mobile and enriched in human-associated environments but have not yet been reported in clinical pathogens; Rank III consists of mobile ARGs not enriched in human or pathogenic contexts; Rank IV includes non-mobile ARGs with limited relevance to human health.

To quantify site-level AMR risk, a composite Threat Score was calculated for each sample. Only Rank I and II ARGs — those deemed high-risk under the ARG_ranker framework — were included in the score computation. These ranks were summed and then weighted by the total normalized ARG abundance for that sample. The resulting metric captures both the clinical risk and quantitative burden of ARGs present. The formula used was:

$$Threat\ Score\ = log_{10}[(Abundance_Rank\ I\ +\ Abundance_Rank\ II) \times Total\ ARG\ Abundance\ +\ \varepsilon]$$

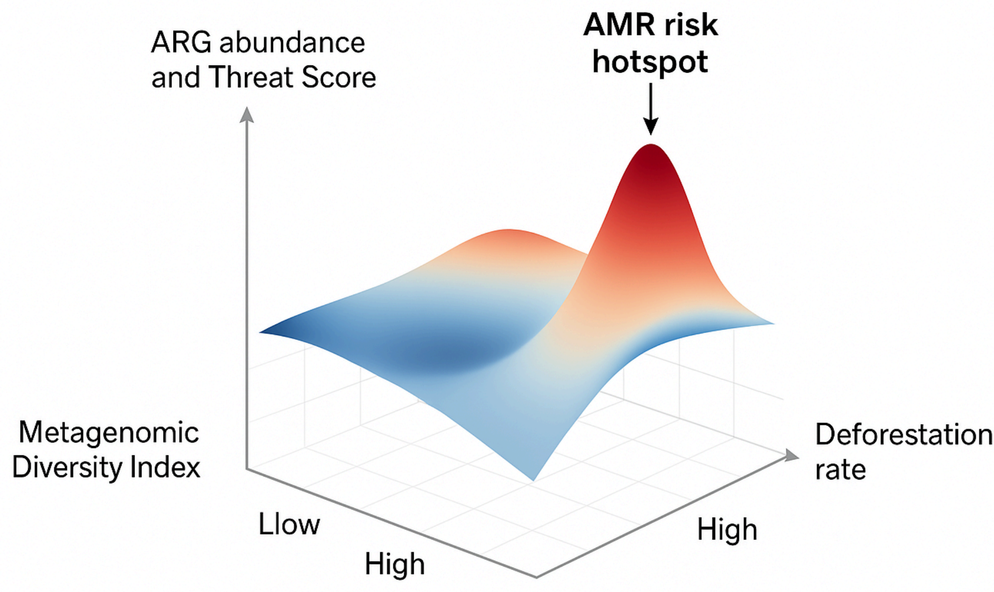
In this equation, Abundance_Rank I and Abundance_Rank II are the normalized abundance values (ARG copies per million cells) of high-risk ARGs in the sample; Total ARG Abundance is the overall ARG load, also normalized; ($\varepsilon = 1\times10^{-6}$) is a small constant to prevent undefined log values. This score was used to rank sampling sites by AMR threat and to build the three-dimensional resistance landscapes presented in Section 2.6.

1.7. *Spatial Modeling and 3D ARG Landscapes*

Three-dimensional ARG landscape models were constructed to contextualize antimicrobial resistance trends within environmental gradients. To model the spatial distribution of antimicrobial resistance in relation to ecological factors, two variables were selected to represent environmental gradients: the regional deforestation percentage and the Metagenomic Diversity Index (MD index). Deforestation data were extracted from the

MapBiomas v6 dataset [19], while MD values were computed based on protein-level dissimilarity metrics as described in Supplementary Methods 1.4.

To improve the interpretability of the resulting plots, MD index values were inverted (i.e., subtracted from the maximum observed value). Mathematically, this means $inverted\ MD\ index\ =\ 1 - MD\ index$. This inversion allowed areas of lower microbial diversity, which may reflect anthropogenic disturbance or selective pressure, to be visually emphasized as zones of ecological risk. In this format, ARG abundance and risk hotspots appeared as raised features on the 3D surface, aligning better with intuitive visual logic and aiding reader interpretation. To facilitate reader interpretation of the ARG landscape modeling approach, a conceptual diagram was created to illustrate how ecological variables and AMR indicators were mapped onto three-dimensional visualizations at Supplementary Methods Figure 01. This diagram serves solely as a conceptual aid and does not reflect the exact spatial data or analytical results used in the study.



Supplementary Methods Figure 01. Conceptual framework for interpreting 3D ARG landscape models. This diagram illustrates how antimicrobial resistance data were integrated into three-dimensional visualizations by mapping sample-level ARG abundance (z-axis) against ecological variables such as metagenomic diversity (x-axis) and deforestation rate (y-axis). The landscape highlights how resistance burden can rise in response to ecological stressors. This schematic is intended as a conceptual aid and does not represent the actual modeling or interpolation used in the analytical visualizations presented in the main manuscript.

Each landscape was generated using Python (v3.9), with data points structured as follows:

- X-axis: Sampling site index, alphabetically ordered (e.g., AP1 to PP5);
- Y-axis: Ecological gradient (inverted MD index or deforestation percentage);
- Z-axis: Log_{10}-transformed ARG abundance (log_{10}(ARGs + 1));
- Color: Log_{10}-transformed ARG Threat Score, mapped using a diverging color scale.

Surfaces were interpolated using scipy.interpolate.griddata, and visualized with plot_surface from matplotlib.pyplot. Annotated labels were manually added to highlight sites with extreme values or clustering patterns.

1.8. *Statistical Analysis and Visualization Framework*

Statistical analyses and visualizations were conducted using a multi-platform approach to support transparency, reproducibility, and high-quality figure generation. Sampling site maps were generated using the free and open-source software QGIS (v3.40) by georeferencing GPS coordinates collected in the field. These base maps were later combined with visual outputs to create the composite figure used in the main manuscript for sampling location context.

Taxonomic classification visualizations, including Sankey diagrams, bubble charts, bar plots, and attribute column charts, were produced using Pavian and Phinch via Galaxy Europe [7], [8]. Pavian facilitated inspection of hierarchical taxonomic structure, while Phinch allowed for exploration of sample-level attributes and microbial composition. Beta diversity visualizations were created in RStudio (v4.4.1) using the vegan package (v2.6-10) [10]. Bray-Curtis dissimilarity matrices were computed based on Bracken-derived species-level abundances. Ordination was carried out via PCoA, and statistical validation was done through PERMANOVA (*adonis2*) and PERMDISP (*permutest*), both with 999 permutations.

Additional statistical analyses were performed in R (v4.4.1) to support the revised comparison of resistome and microbial diversity patterns across municipalities. For ARG data, alpha diversity was summarized using Shannon diversity and observed richness calculated from normalized ARG abundance matrices. Differences in ARG load among municipalities were assessed using the Kruskal-Wallis test, followed by Dunn's post hoc pairwise comparisons with Bonferroni correction. Differences in ARG beta diversity were

evaluated using ANOSIM based on Bray-Curtis dissimilarity matrices, and NMDS was used for ordination and visualization of compositional relationships among samples.

For bacterial diversity, municipality-level comparisons were performed using Kruskal-Wallis tests followed by Dunn's post hoc pairwise comparisons for Shannon diversity, Simpson evenness, and the Metagenomic Diversity Index (MD). Correlations between ARG load and environmental variables, including pH, temperature, electrical conductivity, total dissolved solids, dissolved oxygen, population density, and deforestation, were assessed using Spearman rank correlation. The complete outputs of these analyses are provided in Supplementary Tables S8 and S9, and the corresponding visualizations are shown in Supplementary Figures S7–S11.

AMR heatmaps were created using the Make_AMR_Heatmap.ipynb script from the CZ ID GitHub repository (https://github.com/chanzuckerberg/czid-amr-heatmap), which takes the platform's Combined AMR Report as input and allows filter thresholds to be applied for read alignment and abundance. ARG burden figures, including stacked bar plots, boxplots, and radar plots, were generated using the following R packages: *ggplot2*, *fmsb*, *dplyr*, *tidyr*, *pheatmap*, *patchwork*, *viridis*, *readr*, *gridExtra*, *grid*, *png*, *scales*, *plotly*, and *tidyverse*. These were used for data wrangling, multi-panel layout, and graphical export.

3D ARG landscape plots were created in Python (v3.9) using *numpy*, *pandas*, *scipy*, and *matplotlib*. For each sampling site, the z-axis represented log-transformed ARG abundance; the x-axis corresponded to site codes (ordered alphabetically); and the y-axis represented either regional deforestation (%) or the inverted Metagenomic Diversity (MD) index. Color gradients were mapped to the $\log_{10}$-transformed Threat Score. Surfaces were generated with plot_surface, and interpolation was handled by scipy.interpolate.griddata.

This framework ensured rigorous processing and high-quality visual reporting of the microbial and resistance-related data presented in this study.

2. References

[1] IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, "Pará | Cidades e Estados | IBGE." Accessed: Apr. 15, 2025. [Online]. Available: https://www.ibge.gov.br/cidades-e-estados/pa.html

[2] R. B. L. Cavalcante, A. Fleischmann, P. R. M. Pontes, P. Rógenes, and M. Pontes, "Looking upstream: analyzing the protection of the drainage area of Amazon rivers," Nov. 2023, doi: 10.31223/X51M4F.

[3] K. L. Kalantar *et al.*, "IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring," *Gigascience*, vol. 9, no. 10, pp. 1–14, Oct. 2020, doi: 10.1093/GIGASCIENCE/GIAA111.

[4] D. E. Wood, J. Lu, and B. Langmead, "Improved metagenomic analysis with Kraken 2," *bioRxiv*, p. 762302, Sep. 2019, doi: 10.1101/762302.

[5] L. A. L. Abueg *et al.*, "The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update," *Nucleic Acids Res*, vol. 52, no. W1, pp. W83–W94, Jul. 2024, doi: 10.1093/NAR/GKAE410.

[6] J. Lu, F. P. Breitwieser, P. Thielen, and S. L. Salzberg, "Bracken: Estimating species abundance in metagenomics data," *PeerJ Comput Sci*, vol. 2017, no. 1, p. e104, Jan. 2017, doi: 10.7717/PEERJ-CS.104/SUPP-5.

[7] F. P. Breitwieser and S. L. Salzberg, "Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification," *Bioinformatics*, vol. 36, no. 4, pp. 1303–1304, Feb. 2020, doi: 10.1093/BIOINFORMATICS/BTZ715.

[8] H. M. Bik and P. I. Inc., "Phinch: An interactive, exploratory data visualization framework for –Omic datasets," *bioRxiv*, p. 009944, Oct. 2014, doi: 10.1101/009944.

[9] J. Lu *et al.*, "Metagenome analysis using the Kraken software suite," *Nature Protocols 2022 17:12*, vol. 17, no. 12, pp. 2815–2839, Sep. 2022, doi: 10.1038/s41596-022-00738-y.

[10] Jari Oksanen; et al., "Community Ecology Package [R package vegan version 2.6-10]," Jan. 2025, doi: 10.32614/CRAN.PACKAGE.VEGAN.

[11] D. R. Finn, "A metagenomic alpha-diversity index for microbial functional biodiversity," *FEMS Microbiol Ecol*, vol. 100, no. 3, Feb. 2024, doi: 10.1093/FEMSEC/FIAE019.

[12] D. Li, C. M. Liu, R. Luo, K. Sadakane, and T. W. Lam, "MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph," *Bioinformatics*, vol. 31, no. 10, pp. 1674–1676, May 2015, doi: 10.1093/BIOINFORMATICS/BTV033.

[13] D. Hyatt, G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser, "Prodigal: Prokaryotic gene recognition and translation initiation site identification," *BMC Bioinformatics*, vol. 11, no. 1, pp. 1–11, Mar. 2010, doi: 10.1186/1471-2105-11-119/TABLES/5.

[14] B. P. Alcock *et al.*, "CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database," *Nucleic Acids Res*, vol. 48, no. D1, p. D517, Jan. 2019, doi: 10.1093/NAR/GKZ935.

[15] X. Yin, X. Zheng, L. Li, A. N. Zhang, X. T. Jiang, and T. Zhang, "ARGs-OAP v3.0: Antibiotic-Resistance Gene Database Curation and Analysis Pipeline Optimization," *Engineering*, vol. 27, pp. 234–241, Aug. 2023, doi: 10.1016/J.ENG.2022.10.011.

[16] F. E. Angly, P. G. Dennis, A. Skarshewski, I. Vanwonterghem, P. Hugenholtz, and G. W. Tyson, "CopyRighter: A rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction," *Microbiome*, vol. 2, no. 1, pp. 1–13, Apr. 2014, doi: 10.1186/2049-2618-2-11/FIGURES/6.

[17] D. McDonald *et al.*, "Greengenes2 unifies microbial data in a single reference tree," *Nature Biotechnology 2023 42:5*, vol. 42, no. 5, pp. 715–718, Jul. 2023, doi: 10.1038/s41587-023-01845-1.

[18] A. N. Zhang *et al.*, "An omics-based framework for assessing the health risk of antimicrobial resistance genes," *Nature Communications 2021 12:1*, vol. 12, no. 1, pp. 1–11, Aug. 2021, doi: 10.1038/s41467-021-25096-3.

[19] R. B. L. Cavalcante, A. Fleischmann, P. R. M. Pontes, P. Rógenes, and M. Pontes, "Looking upstream: analyzing the protection of the drainage area of Amazon rivers," Nov. 2023, doi: 10.31223/X51M4F.

7. CONSIDERAÇÕES FINAIS

Esta tese teve como objetivo central investigar, por meio de abordagens metagenômicas, a diversidade microbiana, viral e o resistoma de corpos d'água na região amazônica, com ênfase em sua relevância ecológica e sanitária no contexto de Saúde Única. Os três estudos que a compõem, cada um com escopos e metodologias específicas, representam diferentes dimensões dessa problemática, articulando descobertas complementares sobre a dinâmica microbiana e viral de ambientes aquáticos tropicais.

O primeiro artigo trouxe uma contribuição inédita ao mapear, pela primeira vez, a diversidade viral do Lago Bolonha, revelando uma virosfera rica, funcionalmente diversa e potencialmente impactante em termos ecológicos e de saúde pública. Já o segundo artigo apresentou uma revisão crítica e atualizada sobre ferramentas computacionais para a descoberta *in silico* de vírus bacteriófagos e cianofágos em ambientes aquáticos, consolidando uma base metodológica essencial para futuras análises metagenômicas virais. Por fim, o terceiro e principal artigo forneceu uma caracterização abrangente do resistoma em diferentes municípios paraenses, destacando zonas de risco ecológico e sanitário associadas à presença de genes de resistência a antimicrobianos em águas superficiais de uso humano.

Ao integrar os dados de diversidade microbiana, presença de ARGs e informações ecológicas contextuais, esta tese reforça a importância de abordagens genômicas aplicadas à vigilância ambiental. Seus resultados contribuem não apenas para o avanço científico em ecologia microbiana e resistência antimicrobiana, mas também oferecem subsídios relevantes para a formulação de políticas públicas voltadas ao monitoramento da qualidade da água e à prevenção de riscos sanitários em regiões altamente vulneráveis, como a Amazônia, uma das regiões mais carentes do Brasil com os menores índices de acesso a saneamento básico.

Assim, este trabalho reafirma o potencial da metagenômica ambiental como ferramenta estratégica para a promoção da saúde pública, da conservação ecológica e da compreensão integrada entre ambiente, microrganismos e sociedade — pilares fundamentais do conceito de Saúde Única.

8. PERSPECTIVAS

O avanço desta pesquisa abre caminhos promissores para novas investigações sobre os riscos ecológicos e sanitários associados à resistência antimicrobiana em ambientes aquáticos amazônicos. Um dos principais passos futuros previstos para este estudo é a caracterização dos mecanismos de disseminação dos genes de resistência detectados, a serem analisados nos metagenomas obtidos, com foco na identificação e análise de elementos genéticos móveis (MGEs). A integração dessas informações permitirá uma compreensão mais robusta sobre o potencial de transferência horizontal desses genes de resistência e seus impactos ambientais.

Outro desdobramento importante consiste na análise metagenômica do microbioma intestinal de peixes nativos das áreas de coleta. No futuro, a análise metagenômica do microbioma de peixes poderá revelar interações entre o resistoma ambiental e a fauna aquática, e fornecer evidências importantes sobre a transferência e risco de incorporação de genes de resistência a antimicrobianos (ARGs) ao longo da cadeia alimentar, ampliando a abordagem ecológica e de Saúde Única deste estudo.

Paralelamente, a expansão da amostragem para novas localidades e municípios do Estado do Pará já está em andamento, permitindo ampliar a cobertura espacial da vigilância metagenômica proposta. Tomando como base este projeto, outros projetos já estão em execução em mais duas frentes relacionadas ao monitoramento de rios na Amazônia no Estado do Pará – um em região ocupada pela comunidade índigena Tembé; e outro na região costeira de Bragança, onde há presença de mangues, em áreas que futuramente irão receber influência da exploração de petróleo na Foz do Amazonas. Pretende-se também intensificar a coleta em pontos já analisados, com o objetivo de realizar estudos temporais e comparativos entre estações chuvosas e de baixa pluviosidade, visando investigar a influência de variáveis sazonais na dinâmica do resistoma.

Por fim, destaca-se a importância de que os dados gerados por este trabalho sirvam como subsídios para ações intersetoriais entre pesquisa, vigilância sanitária e políticas públicas. A comunicação efetiva entre ciência, setores de saúde pública e gestão ambiental é essencial para a transformação de conhecimento em estratégia, especialmente em contextos de alta vulnerabilidade, como no caso da região amazônica.

9. REFERÊNCIAS BIBLIOGRÁFICAS

ACHMAD ALI FIKRI, SYAMSUL ARIFIN, M. Fuad Fahruddin. *Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report 2022*. [S.l: s.n.], 2022. v. 2.

ADISASMITO, Wiku B. *et al.* One Health: A new definition for a sustainable and healthy future. *PLOS Pathogens*, v. 18, n. 6, p. e1010537, 1 jun. 2022.

ALCOCK, Brian P. *et al.* CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, v. 48, n. D1, p. D517–D525, 8 jan. 2020.

ALLEN, Heather K. *et al. Call of the wild: Antibiotic resistance genes in natural environments*. *Nature Reviews Microbiology*. [S.l.]: Nat Rev Microbiol. , abr. 2010

ALVES, Jorianne *et al.* Resistome in Lake Bolonha, Brazilian Amazon: Identification of Genes Related to Resistance to Broad-Spectrum Antibiotics. *Frontiers in Microbiology*, v. 11, 4 fev. 2020.

AMGARTEN, Deyvid *et al.* MARVEL, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins. *Frontiers in genetics*, v. 9, n. AUG, 7 ago. 2018.

AMINOV, Rustam I. A brief history of the antibiotic era: Lessons learned and challenges for the future. *Frontiers in Microbiology*, v. 1, n. DEC, 2010.

ANDREWS, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. . [S.l: s.n.]. , 2015

ARAUJO, Fabricio Almeida *et al.* GO FEAT: A rapid web-based functional annotation tool for genomic and transcriptomic data. *Scientific Reports*, v. 8, n. 1, p. 1794, 1 dez. 2018.

AW, Tiong Gim; WENGERT, Samantha; ROSE, Joan B. Metagenomic analysis of viruses associated with field-grown and retail lettuce identifies human and animal viruses. *International Journal of Food Microbiology*, v. 223, p. 50–56, 16 abr. 2016.

BAR-ON, Yinon M.; PHILLIPS, Rob; MILO, Ron. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America*, v. 115, n. 25, p. 6506–6511, 19 jun. 2018.

BHULLAR, Kirandeep *et al.* Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS ONE*, v. 7, n. 4, 2012.

BIESSY, Laura *et al.* Metagenomic insights to the functional potential of sediment microbial communities in freshwater lakes. *Metabarcoding and Metagenomics*, v. 6, p. 59–74, 2022.

BOLGER, Anthony M.; LOHSE, Marc; USADEL, Bjoern. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, v. 30, n. 15, p. 2114–2120, 8 ago. 2014.

BRACK, André. Liquid water and the origin of life. *Origins of Life and Evolution of the Biosphere*, v. 23, n. 1, p. 3–10, fev. 1993.

BREITWIESER, Florian P.; SALZBERG, Steven L. Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics*, v. 36, n. 4, p. 1303–1304, 2020.

BUSCHHARDT, Tasja *et al.* A one health glossary to support communication and information exchange between the human health, animal health and food safety sectors. *One Health*, v. 13, 1 dez. 2021.

DAVIS, Benjamin C. *et al. Recommendations for the use of metagenomics for routine monitoring of antibiotic resistance in wastewater and impacted aquatic environments*. *Critical Reviews in Environmental Science and Technology*. [S.l.]: Taylor & Francis. , 2023

DCOSTA, Vanessa M. *et al. Antibiotic resistance is ancient*. *Nature*. [S.l.]: Nature. , 22 set. 2011

DE MANDAL, S. et al. Metagenomics and Microbial Ecology. Metagenomics and Microbial Ecology, 15 nov. 2021.

DE NIES, Laura *et al.* PathoFact: a pipeline for the prediction of virulence factors and antimicrobial resistance genes in metagenomic data. *Microbiome*, v. 9, n. 1, p. 1–14, 1 dez. 2021.

DECUYPERE, Saskia *et al.* Diagnosis of Bacterial Bloodstream Infections: A 16S Metagenomics Approach. *PLoS Neglected Tropical Diseases*, v. 10, n. 2, p. e0004470, 29 fev. 2016.

DELMONT, Tom O.; SIMONET, Pascal; VOGEL, Timothy M. *Describing microbial communities and performing global comparisons in the omic era*. *ISME Journal*. [S.l.]: Nature Publishing Group. , 21 jun. 2012

DEMŠAR, Janez *et al.* Orange: Data mining toolbox in python. *Journal of Machine Learning Research*, v. 14, p. 2349–2353, 2013.

DEVARAPALLI, Pratap; KUMAVATH, Ranjith N. Metagenomics — A Technological Drift in Bioremediation. *Advances in Bioremediation of Wastewater and Polluted Soil*. [S.l.]: IntechOpen, 2015. .

DHARIWAL, Achal *et al.* MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Research*, v. 45, n. W1, p. W180–W188, 3 jul. 2017.

DOYLE, Conor J.; O'TOOLE, Paul W.; COTTER, Paul D. *Metagenome-based surveillance and diagnostic approaches to studying the microbial ecology of food production and processing environments*. *Environmental Microbiology*. [S.l.]: John Wiley & Sons, Ltd. , 1 nov. 2017

ECKERT, Ester M. *et al.* Persistence of antibiotic resistance genes in large subalpine lakes: the role of anthropogenic pollution and ecological interactions. *Hydrobiologia*, v. 824, n. 1, p. 93–108, 4 jan. 2018.

ELLWANGER, Joel Henrique *et al.* Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. *Anais da Academia Brasileira de Ciencias*, v. 92, n. 1, p. 20191375, 17 abr. 2020.

ESCOBAR-ZEPEDA, Alejandra; DE LEÓN, Arturo Vera Ponce; SANCHEZ-FLORES, Alejandro. *The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics*. *Frontiers in Genetics*. [S.l.]: Frontiers Media SA. , 2015

FERNANDES, Jorlan *et al.* Xapuri virus, a novel mammarenavirus: Natural reassortment and increased diversity between New World viruses. *Emerging Microbes and Infections*, v. 7, n. 1, 1 dez. 2018.

FERNÁNDEZ-ARROJO, Lucía *et al. Metagenomic era for biocatalyst identification*. *Current Opinion in Biotechnology*. [S.l.]: Elsevier Current Trends. , 1 dez. 2010

FLEMING, Alexander. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *British journal of experimental pathology*, v. 10, n. 3, p. 226, 1929.

FRANCO FILHO, Luciano Chaves *et al.* Metagenomic Analysis of Samples from Three Bat Species Collected in the Amazon Rain Forest. *Microbiology Resource Announcements*, v. 8, n. 2, 10 jan. 2019.

GHOSH, Arpita; MEHTA, Aditya; KHAN, Asif M. Metagenomic analysis and its applications. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*. [S.l.]: Academic Press, 2018. v. 1–3. p. 184–193.

GLEICK, Peter. *Water in Crisis: Chapter 2*. Disponível em: <https://www.academia.edu/902661/Water_in_Crisis_Chapter_2_Oxford_University_Press_1993>. Acesso em: 11 abr. 2023.

GOLKAR, Zhabiz; BAGASRA, Omar; GENE PACE, Donald. *Bacteriophage therapy: A potential solution for the antibiotic resistance crisis*. *Journal of Infection in Developing Countries*. [S.l.]: J Infect Dev Ctries. , fev. 2014

GOULD, Ian M.; BAL, Abhijit M. New antibiotic agents in the pipeline and how hey can help overcome microbial resistance. *Virulence*, v. 4, n. 2, p. 185–191, 15 fev. 2013.

GROSS, Michael. Antibiotics in crisis. *Current Biology*, v. 23, n. 24, 16 dez. 2013.

GU, Tingting *et al.* Hydrous peridotitic fragments of Earth's mantle 660 km discontinuity sampled by a diamond. *Nature Geoscience*, v. 15, n. 11, p. 950–954, 26 set. 2022.

GUPTA, Shashank *et al.* Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. *Communications Biology*, v. 2, n. 1, p. 1–7, 5 ago. 2019.

HABBOUSH, Yacob; GUZMAN, Nilmarie. *Antibiotic Resistance*. [S.l.]: StatPearls Publishing, 2022.

HAND, K. P. *et al. On the Habitability and Future Exploration of Ocean Worlds*. *Space Science Reviews*. [S.l.]: Springer. , 1 ago. 2020

HANDELSMAN, Jo *et al.* Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry and Biology*, v. 5, n. 10, 1998.

HARSANYI, John C. Paradoxes of Rationality: Theory of Metagames and Political Behavior. By Nigel Howard. (Cambridge, Mass.: M.I.T. Press, 1971. Pp. 248. $12.95.). *American Political Science Review*, v. 67, n. 2, p. 599–600, jun. 1973.

HAWKEY, Jane *et al.* ISMapper: Identifying transposase insertion sites in bacterial genomes from short read sequence data. *BMC Genomics*, v. 16, n. 1, 3 set. 2015.

HOLLANDER, Jocelyn A.; EINWOHNER, Rachel L. *Conceptualizing resistance*. *Sociological Forum*. [S.l: s.n.]. , 2004

HUA, Xiaoting *et al.* BacAnt: A Combination Annotation Server for Bacterial DNA Sequences to Identify Antibiotic Resistance Genes, Integrons, and Transposable Elements. *Frontiers in Microbiology*, v. 12, p. 2015, 23 jul. 2021.

HUGHES, Donna M.; MLADJENOVIĆ, Lepa; MRŠEVIĆ, Zorica. Feminist Resistance in Serbia. *European Journal of Women's Studies*, v. 2, n. 4, p. 509–532, 1 nov. 1995.

HUIJBERS, Patricia M.C.; LARSSON, D. G.Joakim; FLACH, Carl Fredrik. Surveillance of antibiotic resistant Escherichia coli in human populations through urban wastewater in ten European countries. *Environmental Pollution*, v. 261, 1 jun. 2020.

IBGE. Censo Demográfico 2022: Características dos domicílios - Resultados do universo. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes &id=73106>. Acesso em: 21 abr. 2025.

IPCC. *Climate Change 2022: Impacts, Adaptation and Vulnerability*. [S.l: s.n.], 2022.

KANG, Dongwan D. *et al.* MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, v. 2019, n. 7, 2019.

KARKMAN, Antti; PÄRNÄNEN, Katariina; LARSSON, D. G.Joakim. Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nature Communications*, v. 10, n. 1, 1 dez. 2019.

KEEGAN, Kevin P.; GLASS, Elizabeth M.; MEYER, Folker. MG-RAST, a metagenomics service for analysis of microbial community structure and function. *Methods in Molecular Biology*. [S.l.]: Methods Mol Biol, 2016. v. 1399. p. 207–233.

KIEFT, Kristopher; ANANTHARAMAN, Karthik. Deciphering Active Prophages from Metagenomes. *mSystems*, v. 7, n. 2, 26 abr. 2022.

KO, Karrie K.K.; CHNG, Kern Rei; NAGARAJAN, Niranjan. Metagenomics-enabled microbial surveillance. *Nature Microbiology*, v. 7, n. 4, p. 486–496, 1 abr. 2022.

LARSSON, D. G.J. *et al. AMR, One Health and the environment*. *Nature Microbiology*. [S.l.]: Nature Publishing Group. , 30 mar. 2023

LARSSON, D. G.Joakim; FLACH, Carl Fredrik. *Antibiotic resistance in the environment*. *Nature Reviews Microbiology*. [S.l.]: Nature Publishing Group. , 4 nov. 2022

LEONARD, Anne F.C. *et al.* Exposure to and colonisation by antibiotic-resistant E. coli in UK coastal water users: Environmental surveillance, exposure assessment, and epidemiological study (Beach Bum Survey). *Environment International*, v. 114, p. 326–333, 1 maio 2018.

LEVY, Karen; SMITH, Shanon M.; CARLTON, Elizabeth J. *Climate Change Impacts on Waterborne Diseases: Moving Toward Designing Interventions*. *Current environmental health reports*. [S.l.]: NIH Public Access. , 1 jun. 2018

LI, Dinghua *et al.* MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, v. 31, n. 10, p. 1674–1676, 15 maio 2015.

LIANG, Qiaoyi *et al.* Fecal bacteria act as novel biomarkers for noninvasive diagnosis of colorectal cancer. *Clinical Cancer Research*, v. 23, n. 8, p. 2061–2070, 15 abr. 2017.

LIU, Qiaoliang *et al.* VFM: Identification of Bacteriophages from Metagenomic Bins and Contigs Based on Features Related to Gene and Genome Composition. *IEEE Access*, v. 7, p. 177529–177538, 2019.

LU, Jennifer *et al.* Bracken: Estimating species abundance in metagenomics data. *PeerJ Computer Science*, v. 2017, n. 1, p. e104, 2 jan. 2017.

LUGLI, Gabriele Andrea *et al.* Ancient bacteria of the ötzi's microbiome: A genomic tale from the Copper Age. *Microbiome*, v. 5, n. 1, 2017.

MACKENZIE, John S.; JEGGO, Martyn. *The one health approach-why is it so important? Tropical Medicine and Infectious Disease*. [S.l.]: Multidisciplinary Digital Publishing Institute (MDPI). , 31 maio 2019

MADHAVAN, Aravind *et al.* Metagenome Analysis: a Powerful Tool for Enzyme Bioprospecting. *Applied Biochemistry and Biotechnology*, v. 183, n. 2, p. 636–651, 16 ago. 2017.

MARKKANEN, Melina A. *et al.* Metagenomic Analysis of the Abundance and Composition of Antibiotic Resistance Genes in Hospital Wastewater in Benin, Burkina Faso, and Finland. *mSphere*, v. 8, n. 1, 21 fev. 2023.

MIKHEENKO, Alla; SAVELIEV, Vladislav; GUREVICH, Alexey. MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics*, v. 32, n. 7, p. 1088–1090, 1 abr. 2016.

MOGHISSI, Haideh. Away from home: Iranian women displacement cultural resistance and change. *Journal of Comparative Family Studies*, v. 30, n. 2, p. 207–217, 15 mar. 1999.

MOKILI, John L.; ROHWER, Forest; DUTILH, Bas E. *Metagenomics and future perspectives in virus discovery*. *Current Opinion in Virology*. [S.l.]: Elsevier. , 1 fev. 2012

MOON, Kira *et al.* Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. *Microbiome*, v. 8, n. 1, p. 1–15, 1 jun. 2020.

MORBIDELLI, A. *et al.* Source regions and timescales for the delivery of water to the Earth. *Meteoritics and Planetary Science*, v. 35, n. 6, p. 1309–1320, 1 nov. 2000.

MORGAN, Xochitl C.; HUTTENHOWER, Curtis. Chapter 12: Human Microbiome Analysis. *PLoS Computational Biology*, v. 8, n. 12, p. e1002808, 2012.

MURRAY, Aimee K. The Novel Coronavirus COVID-19 Outbreak: Global Implications for Antimicrobial Resistance. *Frontiers in Microbiology*, v. 11, p. 1020, 13 maio 2020.

MURRAY, C. J. et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet, v. 399, n. 10325, p. 629–655, 12 fev. 2022.

NWACHUKWU, Blessing Chidinma; BABALOLA, Olubukola Oluranti. *Metagenomics: A Tool for Exploring Key Microbiome With the Potentials for Improving Sustainable Agriculture*. *Frontiers in Sustainable Food Systems*. [S.l.]: Frontiers Media S.A. , 17 jun. 2022

O'BRIEN, Evan; XAGORARAKI, Irene. *A water-focused one-health approach for early detection and prevention of viral outbreaks*. *One Health*. [S.l.]: Elsevier. , 1 jun. 2019

OHM, Georg Simon. *The Galvanic Circuit Investigated Mathematically*. [S.l: s.n.], 1827.

PAL, Chandan *et al.* BacMet: Antibacterial biocide and metal resistance genes database. *Nucleic Acids Research*, v. 42, n. D1, p. D737, 1 jan. 2014.

PARADA, Alma E.; NEEDHAM, David M.; FUHRMAN, Jed A. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, v. 18, n. 5, p. 1403–1414, 1 maio 2016.

PARKS, Donovan H. *et al.* CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, v. 25, n. 7, p. 1043–1055, 1 jul. 2015.

PAUL, Deepjyoti; DAS, Bhabatosh. Gut microbiome in the emergence of antibiotic-resistant bacterial pathogens. *Progress in Molecular Biology and Translational Science*. [S.l.]: Academic Press, 2022. v. 192. p. 1–31.

PERRY, Julie; WAGLECHNER, Nicholas; WRIGHT, Gerard. The prehistory of antibiotic resistance. *Cold Spring Harbor Perspectives in Medicine*, v. 6, n. 6, 1 jun. 2016.

PIANI, Laurette *et al.* Earth's water may have been inherited from material similar to enstatite chondrite meteorites. *Science*, v. 369, n. 6507, p. 1110–1113, 28 ago. 2020.

PICKERING, Sharon. Women, The home and resistance in Northern Ireland. *Women and Criminal Justice*, v. 11, n. 3, p. 49–82, 18 jul. 2000.

PIDDOCK, Laura J.V. *The crisis of no new antibiotics-what is the way forward? The Lancet Infectious Diseases*. [S.l.]: Lancet Infect Dis. , mar. 2012

POPKIN, Barry M.; D'ANCI, Kristen E.; ROSENBERG, Irwin H. *Water, hydration, and health*. *Nutrition Reviews*. [S.l.]: Nutr Rev. , 2010

PRUDEN, Amy; ARABI, Mazdak; STORTEBOOM, Heather N. Correlation between upstream human activities and riverine antibiotic resistance genes. *Environmental Science and Technology*, v. 46, n. 21, p. 11541–11549, 6 nov. 2012.

QUINCE, Christopher *et al. Shotgun metagenomics, from sampling to analysis*. *Nature Biotechnology*. [S.l.]: Nature Publishing Group. , 12 set. 2017

RAYMOND, Sean N.; IZIDORO, Andre. Origin of water in the inner Solar System: Planetesimals scattered inward during Jupiter and Saturn's rapid gas accretion. *Icarus*, v. 297, p. 134–148, 15 nov. 2017.

RIESENFELD, Christian S.; SCHLOSS, Patrick D.; HANDELSMAN, Jo. *Metagenomics: Genomic analysis of microbial communities*. *Annual Review of Genetics*. [S.l.]: Annual Reviews. , 29 nov. 2004

ROBERTSON, James; NASH, John H.E. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microbial genomics*, v. 4, n. 8, 1 ago. 2018.

ROSSOLINI, Gian Maria *et al. Update on the antibiotic resistance crisis*. *Current Opinion in Pharmacology*. [S.l.]: Curr Opin Pharmacol. , 2014

RUIZ, J. *Enhanced antibiotic resistance as a collateral COVID-19 pandemic effect? Journal of Hospital Infection*. [S.l.]: W.B. Saunders Ltd. , 1 jan. 2021

SCHMEISSER, Christel; STEELE, Helen; STREIT, Wolfgang R. *Metagenomics, biotechnology with non-culturable microbes*. *Applied Microbiology and Biotechnology*. [S.l.]: Springer. , 30 jul. 2007

SEEMANN, Torsten. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, v. 30, n. 14, p. 2068–2069, 15 jul. 2014.

SENGUPTA, Saswati; CHATTOPADHYAY, Madhab K.; GROSSART, Hans Peter. *The multifaceted roles of antibiotics and antibiotic resistance in nature*. *Frontiers in Microbiology*. [S.l.]: Front Microbiol. , 2013

SHAHEEN, Mohamed N.F. *The concept of one health applied to the problem of zoonotic diseases*. *Reviews in Medical Virology*. [S.l.]: Rev Med Virol. , 1 jul. 2022

SHAW, Alexander G. *et al.* Rapid and sensitive direct detection and identification of poliovirus from stool and environmental surveillance samples by use of nanopore sequencing. *Journal of Clinical Microbiology*, v. 58, n. 9, 1 set. 2020.

SILVA, N. Ku'e! Hawaiian women's resistance to the annexation. *Social Process in Hawai'i*, v. 38, p. 4–1, 1997.

SOARES, Siomar C. *et al.* GIPSy: Genomic island prediction software. *Journal of Biotechnology*, v. 232, p. 2–11, 20 ago. 2016.

SPELLBERG, Brad; GILBERT, David N. The future of antibiotics and resistance: A tribute to a career of leadership by John Bartlett. *Clinical Infectious Diseases*, v. 59, n. Suppl 2, p. S71–S75, 2014.

TALAT, Absar *et al.* Metagenomic Insight into Microbiome and Antibiotic Resistance Genes of High Clinical Concern in Urban and Rural Hospital Wastewater of Northern India Origin: a Major Reservoir of Antimicrobial Resistance. *Microbiology Spectrum*, 14 fev. 2023.

TAN, Siang Yong; TATSUMURA, Yvonne. *Alexander Fleming (1881–1955): Discoverer of penicillin*. *Singapore Medical Journal*. [S.l.]: Wolters Kluwer -- Medknow Publications. , 1 jul. 2015

TECHTMANN, Stephen M.; HAZEN, Terry C. *Metagenomic applications in environmental monitoring and bioremediation*. *Journal of Industrial Microbiology and Biotechnology*. [S.l.]: J Ind Microbiol Biotechnol. , 1 out. 2016

TOYAMA, Danyelle *et al.* Metagenomics analysis of microorganisms in freshwater lakes of the Amazon Basin. *Genome Announcements*, v. 4, n. 6, p. 1440–1456, 2016.

TRAN, Patricia Q. *et al.* Depth-discrete metagenomics reveals the roles of microbes in biogeochemical cycling in the tropical freshwater Lake Tanganyika. *ISME Journal*, v. 15, n. 7, p. 1971–1986, 9 fev. 2021.

TULCHINSKY, Theodore H. John Snow, Cholera, the Broad Street Pump; Waterborne Diseases Then and Now. *Case Studies in Public Health*, p. 77, 2018.

UHLIK, Ondrej *et al. Stable isotope probing in the metagenomics era: A bridge towards improved bioremediation*. *Biotechnology Advances*. [S.l.]: Elsevier. , 1 mar. 2013

UNITED NATIONS. *Sustainable Development Goals Report – United Nations Sustainable Development*. *United Nations Publications,*. [S.l: s.n.]. , 2020

UNITED NATIONS (UN). *Forests, desertification and biodiversity - United Nations Sustainable Development*. Disponível em: <https://www.un.org/sustainabledevelopment/biodiversity/>. Acesso em: 11 abr. 2023a.

UNITED NATIONS (UN). *Water and Sanitation - United Nations Sustainable Development*. Disponível em: <https://www.un.org/sustainabledevelopment/water-and-sanitation/>. Acesso em: 11 abr. 2023b.

US DEPARTMENT OF HEALTH AND HUMAN SERVICES (CENTERS FOR DISEASE CONTROL AND PREVENTION). Antibiotic Resistance Threats in The United States 2019. *Cdc*, v. 10, n. 1, 2019.

US NATIONAL RESEARCH COUNCIL, Committee on Metagenomics: Challenges and; FUNCTIONAL APPLICATIONS. Why Metagenomics? *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*. [S.l.]: National Academies Press (US), 2007. .

VELAZQUEZ-MEZA, Maria Elena *et al. Antimicrobial resistance: One Health approach*. *Veterinary World*. [S.l.]: Veterinary World. , 1 mar. 2022

VENTER, J. Craig *et al.* Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science*, v. 304, n. 5667, p. 66–74, 2 abr. 2004.

VENTOLA, C. Lee. The antibiotic resistance crisis: causes and threats. *P & T journal*, v. 40, n. 4, p. 277–83, 2015.

VERA, Hernan; FOUCAULT, Michel. The History of Sexuality, Volume 1: An Introduction. *Contemporary Sociology*, v. 8, n. 4, p. 589, 1979.

WHO. Global Action Plan on Antimicrobial Resistance. *World Health Organization*, v. 10, n. 9, p. 1–28, 2017.

WHO. *Technical Brief on Water, Sanitation, Hygiene and Wastewater Management to Prevent Infections and Reduce the Spread of Antimicrobial Resistance*. [S.l: s.n.], 2020.

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

WOESE, C R. Bacterial evolution. *Microbiological Reviews*, v. 51, n. 2, p. 221–271, jun. 1987.

WOOD, Derrick E.; LU, Jennifer; LANGMEAD, Ben. Improved metagenomic analysis with Kraken 2. *Genome Biology*, v. 20, n. 1, 28 nov. 2019.

WORD HEALTH ORGANIZATION. *Antimicrobial resistance*. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>. Acesso em: 22 abril 2025.

XAGORARAKI, Irene; YIN, Ziqiang; SVAMBAYEV, Zhassulan. Fate of Viruses in Water Systems. *Journal of Environmental Engineering*, v. 140, n. 7, p. 04014020, 9 maio 2014.

YAMADA, Yosuke *et al.* Variation in human water turnover associated with environmental and lifestyle factors. *Science*, v. 378, n. 6622, p. 909–915, 11 nov. 2022.

YANG, Ying *et al.* Fate of antibiotic resistance genes in sewage treatment plant revealed by metagenomic approach. *Water Research*, v. 62, p. 97–106, 1 out. 2014.

YUN, Jiae; RYU, Sangryeol. *Screening for novel enzymes from metagenome and SIGEX, as a way to improve it*. *Microbial Cell Factories*. [S.l.]: BioMed Central. , 25 mar. 2005.

ZOLENSKY, Michael E. *et al.* Asteroidal water within fluid inclusion-bearing halite in an H5 chondrite, Monahans (1998). *Science*, v. 285, n. 5432, p. 1377–1379, 27 ago. 1999.

10. APÊNDICE

Nesta seção estão os dados suplementares coletados ao longo do período de coleta e planejamento amostral deste estudo. Segue como:

- APÊNDICE A – Dados coletados por drone: Registros fotográficos e audiovisuais.

APÊNDICE A – Dados coletados por drone – Registros fotográficos e audiovisuais

Neste repositório online, os registros fotográficos e audiovisuais capturados pelo Drone DJI Inspire 1 estão disponíveis para a visualização. O conteúdo pode ser acessado através da leitura do QR Code acima ou através do link: https://drive.google.com/drive/folders/1OWg6DFRUGyqiibqe54zhJkT48OyzT7MO?usp=share_link.

11. ANEXOS

Nesta seção estão descritos os demais artigos e atividades nos quais eu participei como primeiro autor, co-autor, organizador ou co-organizador, incluindo produções bibliográficas, participações em eventos, organizações de eventos, apresentações de palestras e minicursos ministrados. Seguem como:

- ANEXO A – Premiação: Prêmio de Mérito em Divulgação Científica na Categoria *Best Paper Award* para o capítulo de livro "Imunobioinformática para leigos";
- ANEXO B – Artigo Publicado: Computational identification of putative common genomic drug and vaccine targets in *Mycoplasma genitalium*;
- ANEXO C – Artigo Publicado: From In-Person to the Online World: Insights Into Organizing Events in Bioinformatics;
- ANEXO D – Artigo Publicado: Genomic and phenotypic insight into antimicrobial resistance of *Pseudomonas fluorescens* from King George Island, Antarctica;
- ANEXO E – Artigo Publicado: Microbiome analyses of the Uraim River in the Amazon and georeferencing analyses to establish correlation with anthropogenic impacts of land use;
- ANEXO F – Artigo Publicado: *In silico* approach to identify microsatellites candidate biomarkers to differentiate the biovar of *Corynebacterium pseudotuberculosis* genomes;
- ANEXO G – Artigo Publicado: Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil;
- ANEXO H – Livro Publicado: BIOINFO #02 - Revista Brasileira de Bioinformática e Biologia Computacional;
- ANEXO I – Livro Publicado: BIOINFO #03 - Revista Brasileira de Bioinformática e Biologia Computacional;
- ANEXO J – Capítulos de Livro Publicados: 2 capítulos em "Reverse Vaccinology", 4 capítulos em "O ESTADO DA ARTE NAS PESQUISAS EM VACINOLOGIA" e 2 capítulos em "BIOINFO – Revista Brasileira de Bioinformática. Edição #01";
- ANEXO K – Currículo Lattes: Atividades realizadas de 01/2021 à 04/2025.

ANEXO A – Premiação – Prêmio de Mérito em Divulgação Científica na Categoria *Best Paper Award* para o capítulo de livro "Imunobioinformática para leigos"

O Prêmio BIOINFO de Mérito em Divulgação Científica – 2023 é uma iniciativa do comitê editorial da Revista BIOINFO para reconhecer autores de artigos publicados na revista, como reconhecimento pelas suas contribuições para a disseminação da bioinformática e da biologia computacional no Brasil. No ano de 2023, o prêmio contemplou três principais categorias: Artigo mais citado; Artigo mais lido; *Best paper award*. A lista completa de vencedores do Prêmio BIOINFO 2023 está disponível em https://bioinfo.com.br/premio-bioinfo-2023.

ANEXO B – Artigo Publicado – Computational identification of putative common genomic drug and vaccine targets in *Mycoplasma genitalium*

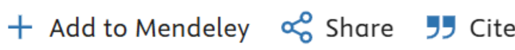
Genomics
Volume 113, Issue 4, July 2021, Pages 2730-2743

Original Article

Computational identification of putative common genomic drug and vaccine targets in *Mycoplasma genitalium*

Wylerson G. Nogueira [a 1], Arun Kumar Jaiswal [a c 1], Sandeep Tiwari [a], Rommel T.J. Ramos [b], Preetam Ghosh [d], Debmalya Barh [e], Vasco Azevedo [a], Siomar C. Soares [c]

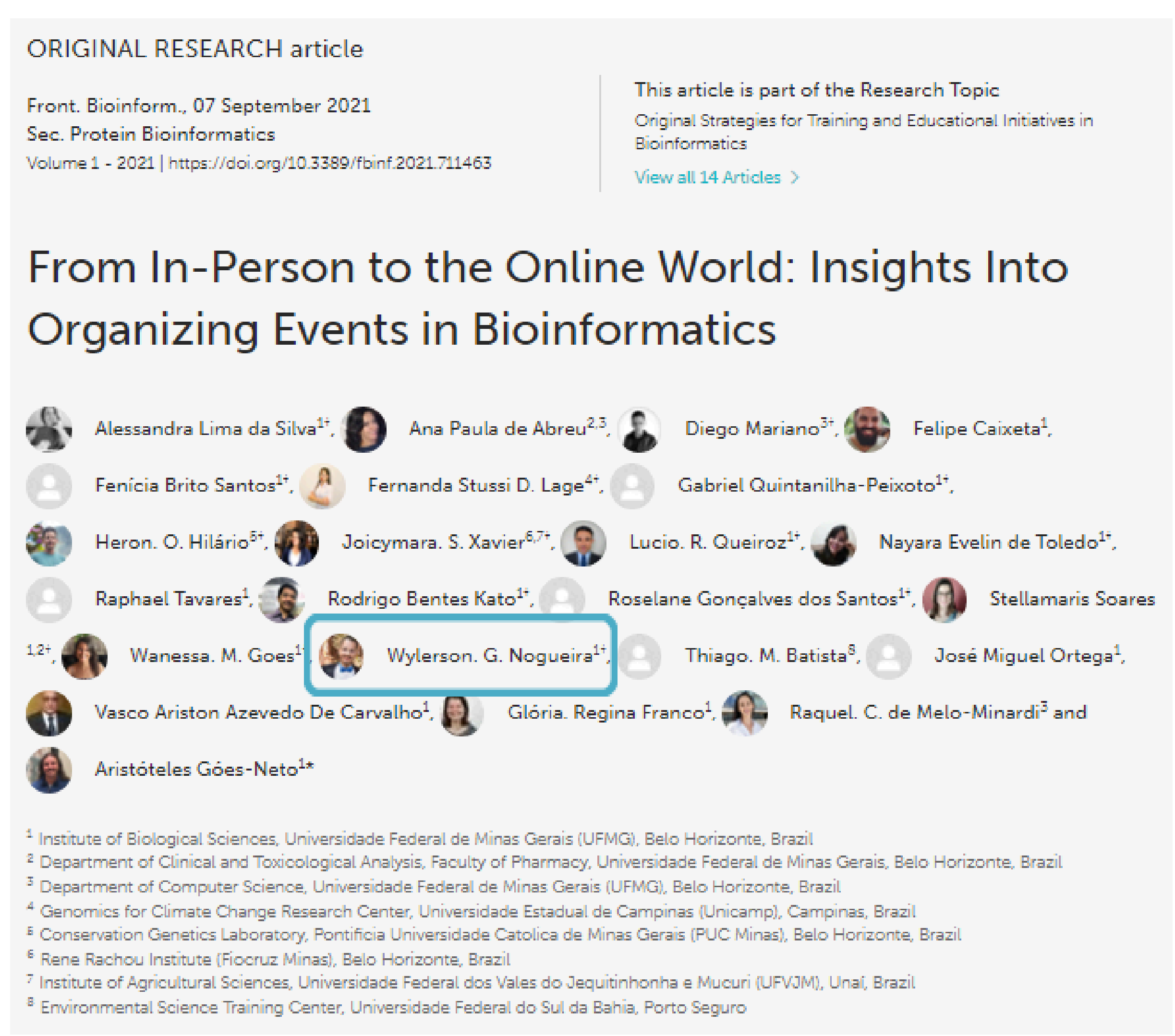
Show more



https://doi.org/10.1016/j.ygeno.2021.06.011 Get rights and content

Neste artigo (https://doi.org/10.1016/j.ygeno.2021.06.011) de primeira autoria, eu participei da análise formal, investigação, redação do rascunho original, revisão e edição, e na visualização dos resultados alcançados. Este artigo é oriundo do legado de trabalhos desenvolvidos ao longo do mestrado e foi publicado no periódico científico *Genomics*.

ANEXO C – Artigo Publicado – From In-Person to the Online World: Insights Into Organizing Events in Bioinformatics



ORIGINAL RESEARCH article

Front. Bioinform., 07 September 2021
Sec. Protein Bioinformatics
Volume 1 - 2021 | https://doi.org/10.3389/fbinf.2021.711463

This article is part of the Research Topic
Original Strategies for Training and Educational Initiatives in Bioinformatics
View all 14 Articles

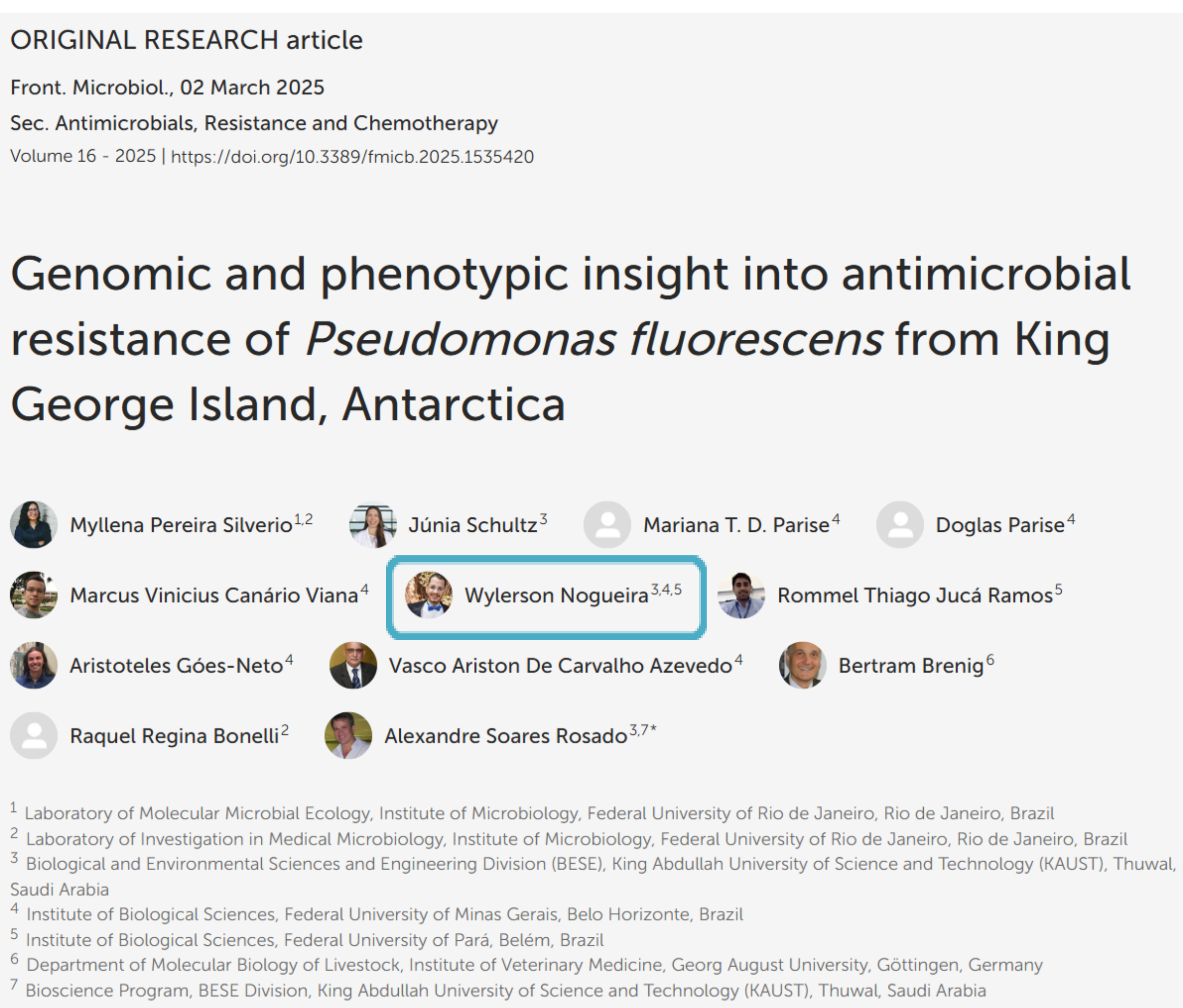
From In-Person to the Online World: Insights Into Organizing Events in Bioinformatics

Alessandra Lima da Silva[1†], Ana Paula de Abreu[2,3], Diego Mariano[3†], Felipe Caixeta[1], Fenícia Brito Santos[1†], Fernanda Stussi D. Lage[4†], Gabriel Quintanilha-Peixoto[1†], Heron. O. Hilário[5†], Joicymara. S. Xavier[6,7†], Lucio. R. Queiroz[1†], Nayara Evelin de Toledo[1†], Raphael Tavares[1], Rodrigo Bentes Kato[1†], Roselane Gonçalves dos Santos[1†], Stellamaris Soares[1,2†], Wanessa. M. Goes[1], Wylerson. G. Nogueira[1†], Thiago. M. Batista[8], José Miguel Ortega[1], Vasco Ariston Azevedo De Carvalho[1], Glória. Regina Franco[1], Raquel. C. de Melo-Minardi[3] and Aristóteles Góes-Neto[1*]

[1] Institute of Biological Sciences, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil
[2] Department of Clinical and Toxicological Analysis, Faculty of Pharmacy, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[3] Department of Computer Science, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil
[4] Genomics for Climate Change Research Center, Universidade Estadual de Campinas (Unicamp), Campinas, Brazil
[5] Conservation Genetics Laboratory, Pontificia Universidade Catolica de Minas Gerais (PUC Minas), Belo Horizonte, Brazil
[6] Rene Rachou Institute (Fiocruz Minas), Belo Horizonte, Brazil
[7] Institute of Agricultural Sciences, Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Unaí, Brazil
[8] Environmental Science Training Center, Universidade Federal do Sul da Bahia, Porto Seguro

Neste artigo (https://doi.org/10.3389/fbinf.2021.711463) de primeira autoria compartilhada, eu participei na definição de escopo metodológico, análise formal, investigação, redação do rascunho original, revisão e edição, e na visualização dos resultados alcançados. Este artigo documenta a experiência de realização das edições presenciais e uma edição online do Curso de Verão em Bioinformática da Universidade Federal de Minas Gerais, publicado no período científico *Frontiers in Bioinformatics,* sob o *Research Topic* “Original Strategies for Training and Educational Initiatives in Bioinformatics”.

ANEXO D – Artigo Publicado – Genomic and phenotypic insight into antimicrobial resistance of *Pseudomonas fluorescens* from King George Island, Antarctica



ORIGINAL RESEARCH article
Front. Microbiol., 02 March 2025
Sec. Antimicrobials, Resistance and Chemotherapy
Volume 16 - 2025 | https://doi.org/10.3389/fmicb.2025.1535420

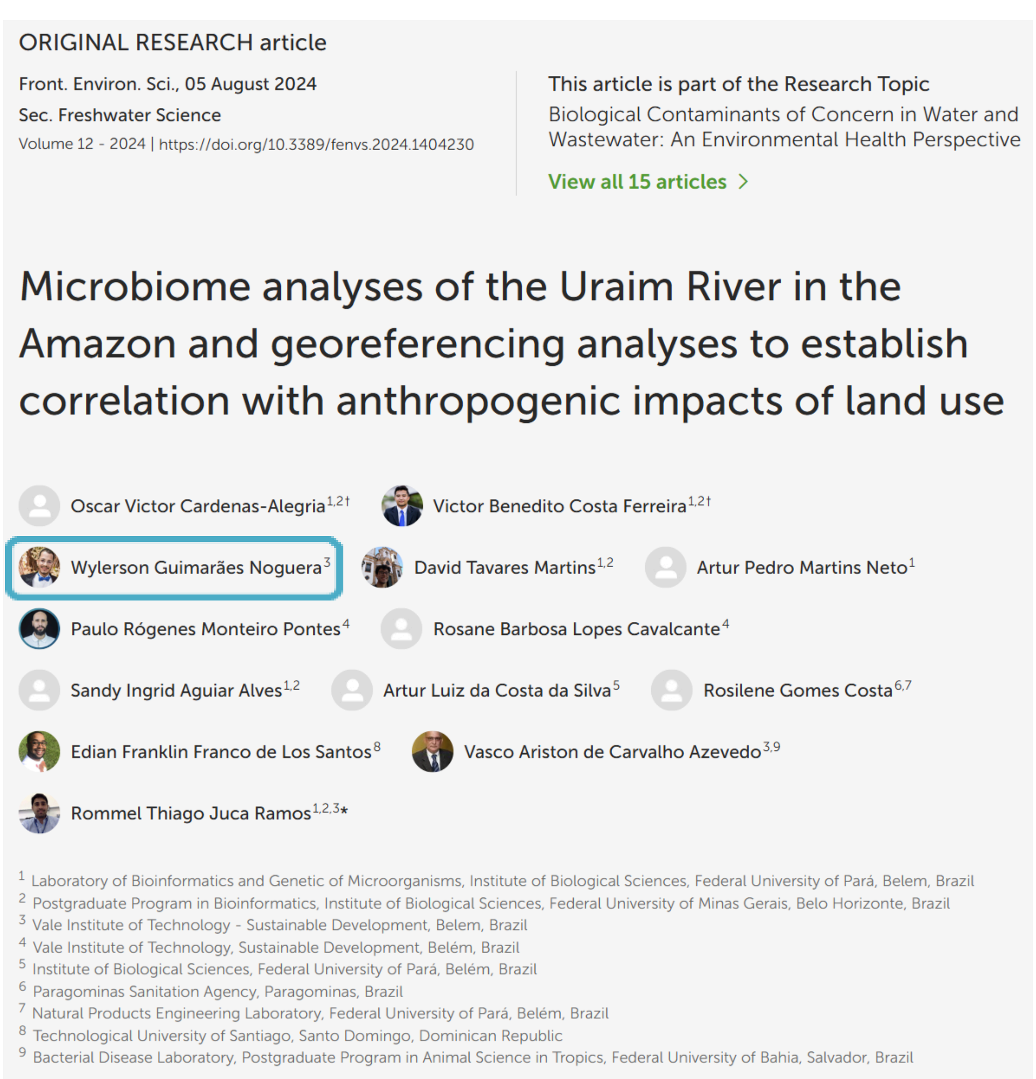
Genomic and phenotypic insight into antimicrobial resistance of *Pseudomonas fluorescens* from King George Island, Antarctica

Myllena Pereira Silverio[1,2] Júnia Schultz[3] Mariana T. D. Parise[4] Doglas Parise[4] Marcus Vinicius Canário Viana[4] Wylerson Nogueira[3,4,5] Rommel Thiago Jucá Ramos[5] Aristoteles Góes-Neto[4] Vasco Ariston De Carvalho Azevedo[4] Bertram Brenig[6] Raquel Regina Bonelli[2] Alexandre Soares Rosado[3,7]*

[1] Laboratory of Molecular Microbial Ecology, Institute of Microbiology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
[2] Laboratory of Investigation in Medical Microbiology, Institute of Microbiology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
[3] Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[4] Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil
[5] Institute of Biological Sciences, Federal University of Pará, Belém, Brazil
[6] Department of Molecular Biology of Livestock, Institute of Veterinary Medicine, Georg August University, Göttingen, Germany
[7] Bioscience Program, BESE Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Neste artigo (https://doi.org/10.3389/fmicb.2025.1535420), como co-autor, eu participei na análise parcial de dados, revisão e na visualização de alguns dos resultados alcançados. A participação neste artigo é oriunda da colaboração proveniente do intercâmbio de doutorado sanduíche em KAUST, sob a supervisão do Prof. Dr. Alexandre Rosado. Este artigo foi publicado no periódico científico *Frontiers in Microbiology*, sob a seção de *Antimicrobials, Resistance and Chemotherapy*.

ANEXO E – Artigo Publicado – Microbiome analyses of the Uraim River in the Amazon and georeferencing analyses to establish correlation with anthropogenic impacts of land use



ORIGINAL RESEARCH article

Front. Environ. Sci., 05 August 2024
Sec. Freshwater Science
Volume 12 - 2024 | https://doi.org/10.3389/fenvs.2024.1404230

This article is part of the Research Topic
Biological Contaminants of Concern in Water and Wastewater: An Environmental Health Perspective

View all 15 articles

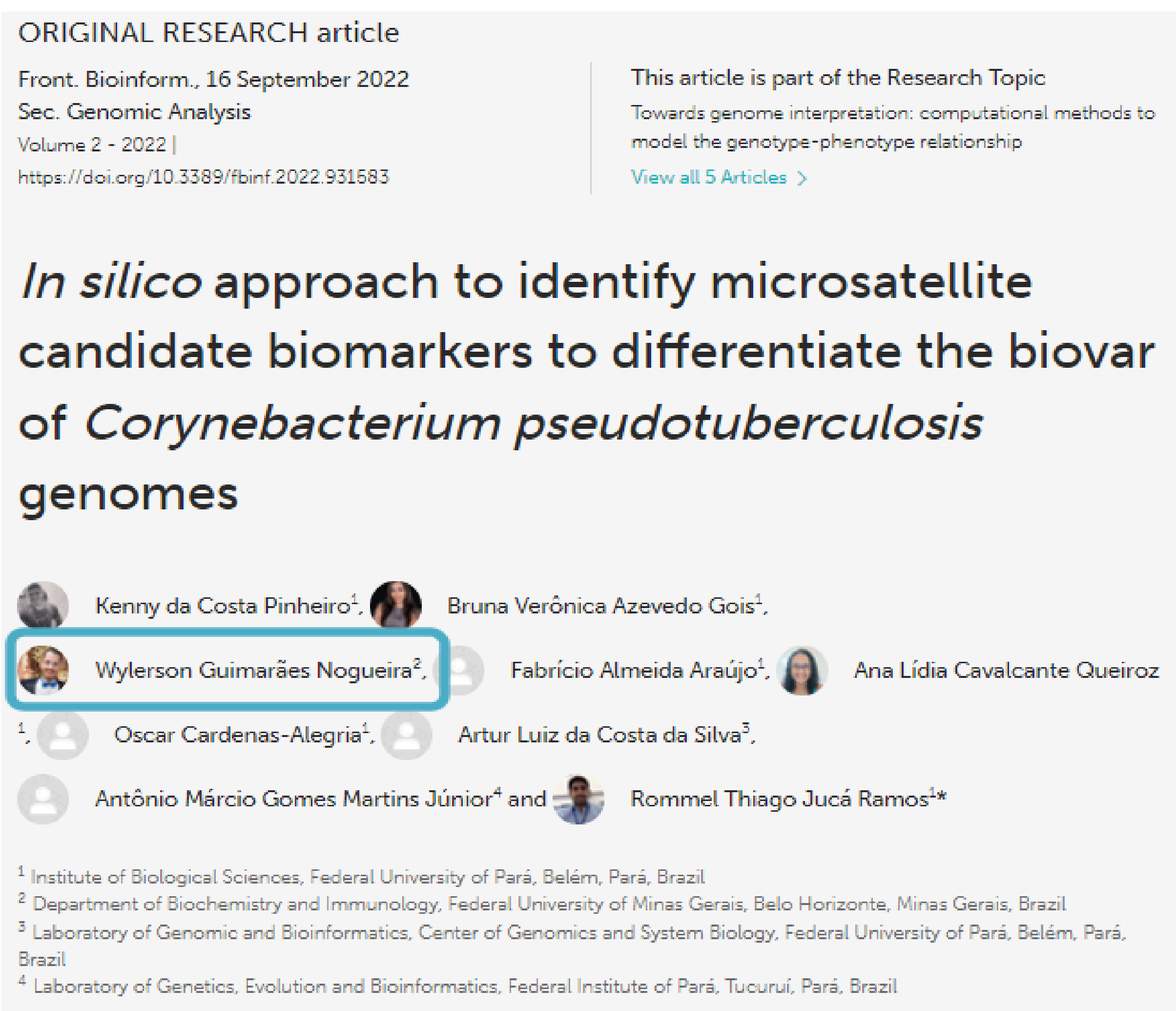
Microbiome analyses of the Uraim River in the Amazon and georeferencing analyses to establish correlation with anthropogenic impacts of land use

Oscar Victor Cardenas-Alegria[1,2†] Victor Benedito Costa Ferreira[1,2†]
Wylerson Guimarães Noguera[3] David Tavares Martins[1,2] Artur Pedro Martins Neto[1]
Paulo Rógenes Monteiro Pontes[4] Rosane Barbosa Lopes Cavalcante[4]
Sandy Ingrid Aguiar Alves[1,2] Artur Luiz da Costa da Silva[5] Rosilene Gomes Costa[6,7]
Edian Franklin Franco de Los Santos[8] Vasco Ariston de Carvalho Azevedo[3,9]
Rommel Thiago Juca Ramos[1,2,3*]

[1] Laboratory of Bioinformatics and Genetic of Microorganisms, Institute of Biological Sciences, Federal University of Pará, Belem, Brazil
[2] Postgraduate Program in Bioinformatics, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil
[3] Vale Institute of Technology - Sustainable Development, Belem, Brazil
[4] Vale Institute of Technology, Sustainable Development, Belém, Brazil
[5] Institute of Biological Sciences, Federal University of Pará, Belém, Brazil
[6] Paragominas Sanitation Agency, Paragominas, Brazil
[7] Natural Products Engineering Laboratory, Federal University of Pará, Belém, Brazil
[8] Technological University of Santiago, Santo Domingo, Dominican Republic
[9] Bacterial Disease Laboratory, Postgraduate Program in Animal Science in Tropics, Federal University of Bahia, Salvador, Brazil

Neste artigo (https://doi.org/10.3389/fenvs.2024.1404230), eu participei como co-autor neste trabalho que deu origem à publicação no periódico científico *Frontiers in Environmental Science*, sob a seção de *Freshwater Science*, como parte do *Research Topic* ”Biological Contaminants of Concern in Water and Wastewater: An Environmental Health Perspective”.

ANEXO F – Artigo Publicado – *In silico* approach to identify microsatellites candidate biomarkers to differentiate the biovar of *Corynebacterium pseudotuberculosis* genomes



ORIGINAL RESEARCH article

Front. Bioinform., 16 September 2022
Sec. Genomic Analysis
Volume 2 - 2022 |
https://doi.org/10.3389/fbinf.2022.931583

This article is part of the Research Topic
Towards genome interpretation: computational methods to model the genotype-phenotype relationship
View all 5 Articles

In silico approach to identify microsatellite candidate biomarkers to differentiate the biovar of *Corynebacterium pseudotuberculosis* genomes

Kenny da Costa Pinheiro[1], Bruna Verônica Azevedo Gois[1], Wylerson Guimarães Nogueira[2], Fabrício Almeida Araújo[1], Ana Lídia Cavalcante Queiroz[1], Oscar Cardenas-Alegria[1], Artur Luiz da Costa da Silva[3], Antônio Márcio Gomes Martins Júnior[4] and Rommel Thiago Jucá Ramos[1]*

[1] Institute of Biological Sciences, Federal University of Pará, Belém, Pará, Brazil
[2] Department of Biochemistry and Immunology, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[3] Laboratory of Genomic and Bioinformatics, Center of Genomics and System Biology, Federal University of Pará, Belém, Pará, Brazil
[4] Laboratory of Genetics, Evolution and Bioinformatics, Federal Institute of Pará, Tucuruí, Pará, Brazil

Neste artigo (https://doi.org/10.3389/fbinf.2022.931583) de co-autoria, eu participei na definição de escopo metodológico, análise formal, investigação, redação do rascunho original, revisão e edição, e na visualização dos resultados alcançados. Este artigo foi publicado no período científico *Frontiers in Bioinformatics,* sob o *Research Topic* ”Towards genome interpretation: computational methods to model the genotype-phenotype relationship”.

ANEXO G – Artigo Publicado – Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil

nature > nature communications > articles > article

Article | Open Access | Published: 16 April 2021

Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil

Talita Émile Ribeiro Adelino, Marta Giovanetti, Vagner Fonseca, Joilson Xavier, Álvaro Salgado de Abreu, Valdinete Alves do Nascimento, Luiz Henrique Ferraz Demarchi, Marluce Aparecida Assunção Oliveira, Vinicius Lemes da Silva, Arabela Leal e. Silva de Mello, Gabriel Muricy Cunha, Roselene Hans Santos, Elaine Cristina de Oliveira, Jorge Antônio Chamon Júnior, Felipe Campos de Melo Iani, Ana Maria Bispo de Filippis, André Luiz de Abreu, Ronaldo de Jesus, Carlos Frederico Campelo de Albuquerque, Jairo Mendez Rico, Rodrigo Fabiano do Carmo Said, Joscélio Aguiar Silva, Noely Fabiana Oliveira de Moura, Priscila Leite, Latin American Genomic Surveillance Arboviral Network, ... Luiz Carlos Junior Alcantara

Latin American Genomic Surveillance Arboviral Network ×

Erenilde Marques de Cerqueira, Tiago Graf, Walter Ramalho, Wildo Navegantes, Renato Barbosa Reis, Clara Guerra Duarte, Maira Alves Pereira, Paulo Eduardo de Souza da Silva, Raoni Almeida de Souza, Alex Pauvolid-Corrêa, Anne Aline Pereira de Paiva, Hegger Machado Fritsch, Maria Angélica Mares-Guia, Maria Celeste Torres, Maurício Teixeira Lima, Patrícia Sequeira, William de Almeida Marques, Jorlan Fernandes de Jesus, Felipe Gomes Naveca, Alessandra Lima Silva, Anne Cybelle Pinto, Arun Kumar Jaiswal, Élisson Nogueira Lopes, Francielly Morais Rodrigues da Costa, Gabriel Quintanilha-Peixoto, Gilson Carlos Soares, Paula Luize Camargos Fonseca, Renan Pedra de Souza, Rodrigo Bentes Kato, Rodrigo Profeta Silveira Santos, Sandeep Tiwari, Wylerson Guimarães Nogueira, Beatriz Senra Álvares da Silva Santos, Bruna Lopes Bueno, Isadora

Borba, Alix Sandra Mazzetto, Francisco de Assis Araújo Aguiar, Irenio da Silva Gomes, Jayra Juliana Paiva Alves Abrantes, Luiz Takao Watanabe, Marta Ferreira da Silva Rego, Vanessa Brandão Nardy, Shirlei Ferreira de Aguiar, Fabiana Cristina Pereira dos Santos, Alice Louize Nunes Queiroz, Bruno Tardelli Diniz Nunes, Lívia Carício Martins, Márcio Roberto Teixeira Nunes, Flávia Cristina da Silva Salles, Ingra Morales Claro, Jaqueline Goes de Jesus, Darlan da Silva Cândido, Cintia Marcela Fabbri, Claudia González, Lisseth Saéz, María Chen-Germán, Jaime Lagos Barrera, José Ernesto Ramírez-González, Josefina Campos, Noelia Morel Faller, Marta Eugenia Víquez Villalobos, Roberto Kaslin, Silvia Paola Salgado Cisneros, Flávia Figueira Aburjaile, Carolina Dourado Amaral, Danielle Bandeira Costa de Sousa Freire, Laura Nogueira Cruz, Daniel Mattos, Leandro Ferreira Lopes Landeira, Mariane Talon de Menezes, Ieda Maria Orioli, Ariane Coelho Ferraz, Daiane Teixeira de Oliveira, Alexandre Barbosa Reis, Renata Guerra de Sá Cota, Rafael dos Santos Bezerra, Melissa Barreto Falcão & Rodrigo Dias de Oliveira Carvalho

Neste artigo (https://doi.org/10.1038/s41467-021-22607-0), eu participei como parte do *Latin American Genomic Surveillance Arboviral Network*, consórcio de participantes do evento "Nanopore-based genome sequencing technology for temporal investigation and epidemiology of dengue outbreak: training, research, surveillance, and scientific dissemination", que deu origem à publicação no periódico científico *Nature Communications*.

ANEXO H – Livro Publicado – BIOINFO #02 - Revista Brasileira de Bioinformática e Biologia Computacional

bioinfo

Revista Brasileira de Bioinformática e Biologia Computacional

Alfahelix

BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional

ISBN: 978-65-992753-5-7 | doi: 10.51780/978-65-992753-5-7

www.bioinfo.com.br

2ª edição - Jul. 2022

Organização & Revisão

Luana Luiza Bastos

Doutoranda do programa de pós-graduação em Bioinformática da UFMG

Angie Atoche Puelles

Mestranda do programa de pós-graduação em Bioinformática da UFMG

Vivian Paixão

Mestranda do programa de pós-graduação em Bioinformática da UFMG

Giovanna Fiorini

Mestranda do programa de pós-graduação em Bioinformática da UFMG

Eduardo Utsch

Mestrando do programa de pós-graduação em Bioinformática da UFMG

Wylerson Guimarães Nogueira

Doutorando do programa de pós-graduação em Bioinformática da UFMG

Tatiane Senna Bialves

Doutoranda do Programa de pós-graduação em Ciências Fisiológicas da FURG

Dra. Lucianna Helene Santos

Pesquisadora do Institut Pasteur de Montevideo

Ana Carolina Silva Bulla

Mestranda em Biologia Computacional e Sistemas no Instituto Oswaldo Cruz

Flavia Sanches

Mestranda do Programa de Pós-Graduação Multicêntrico em Bioquímica e Biologia Molecular na UFBA

Débora Soares Brandão

Doutoranda do programa de pós-graduação em Biologia Vegetal da UFMG

Isaac Farias Cansanção

Professor Adjunto, Universidade Federal do Vale do São Francisco, UNIVASF

Sávio Costa

Parque de Ciência e Tecnologia Guamá, Laboratório de Engenharia Biológica, Belém, PA, Brazil

Frederico Chaves Carvalho

Doutorando do programa de pós-graduação em Ciência da Computação da UFMG

Dr. Diego Mariano

Residente pós-doutoral | Departamento de Ciência da Computação (UFMG) | Editor-in-chief Alfahelix

Ficha catalográfica

Sandro Alex Batista CRB6/2433

Bibliotecário

Publicação

Editora Alfahelix, CNPJ: 37.524.984/0001-10

Lagoa Santa, MG, Brasil

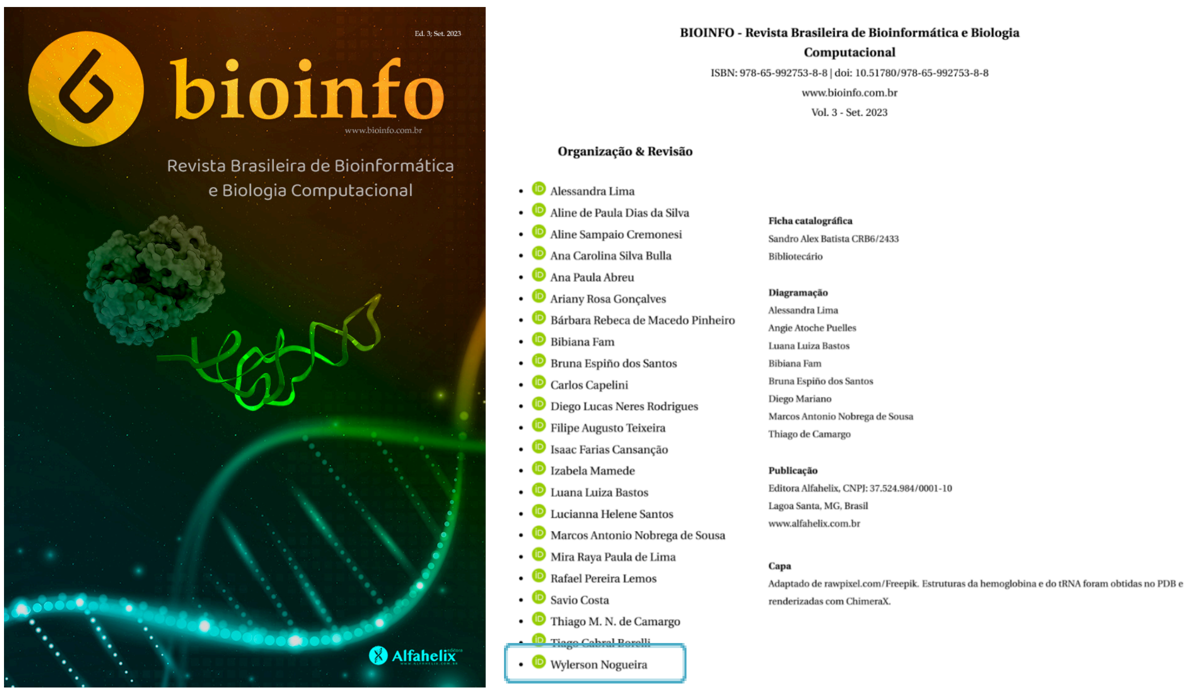
www.alfahelix.com.br

Capa

Adaptado de rawpixel.com/Freepik e NASA (https://webb.nasa.gov/). Estruturas da hemoglobina e do tRNA foram obtidas no PDB e renderizadas com ChimeraX.

Neste livro, intitulado “BIOINFO #02 - Revista Brasileira de Bioinformática e Biologia Computacional” (ISBN: 978-65-992753-5-7 | doi.org/10.51780/978-65-992753-5-7), eu participei da organização e revisão do manuscrito. Este e-book é fruto da 2ª edição de produções bibliográficas de BIOINFO, uma plataforma para divulgação científica em bioinformática, que tem como objetivo fornecer material de estudo de fácil compreensão referente a área em português e fornecer a alunos de pós-graduação a possibilidade de compartilhar seus conhecimentos através de artigos e tutoriais, bem como exercitar sua habilidade de comunicação científica.

ANEXO I – Livro Publicado – BIOINFO #03 - Revista Brasileira de Bioinformática e Biologia Computacional



BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional

ISBN: 978-65-992753-8-8 | doi: 10.51780/978-65-992753-8-8

www.bioinfo.com.br

Vol. 3 - Set. 2023

Organização & Revisão

- Alessandra Lima
- Aline de Paula Dias da Silva
- Aline Sampaio Cremonesi
- Ana Carolina Silva Bulla
- Ana Paula Abreu
- Ariany Rosa Gonçalves
- Bárbara Rebeca de Macedo Pinheiro
- Bibiana Fam
- Bruna Espiño dos Santos
- Carlos Capelini
- Diego Lucas Neres Rodrigues
- Filipe Augusto Teixeira
- Isaac Farias Cansanção
- Izabela Mamede
- Luana Luiza Bastos
- Lucianna Helene Santos
- Marcos Antonio Nobrega de Sousa
- Mira Raya Paula de Lima
- Rafael Pereira Lemos
- Savio Costa
- Thiago M. N. de Camargo
- Tiago Cabral Borelli
- Wylerson Nogueira

Ficha catalográfica
Sandro Alex Batista CRB6/2433
Bibliotecário

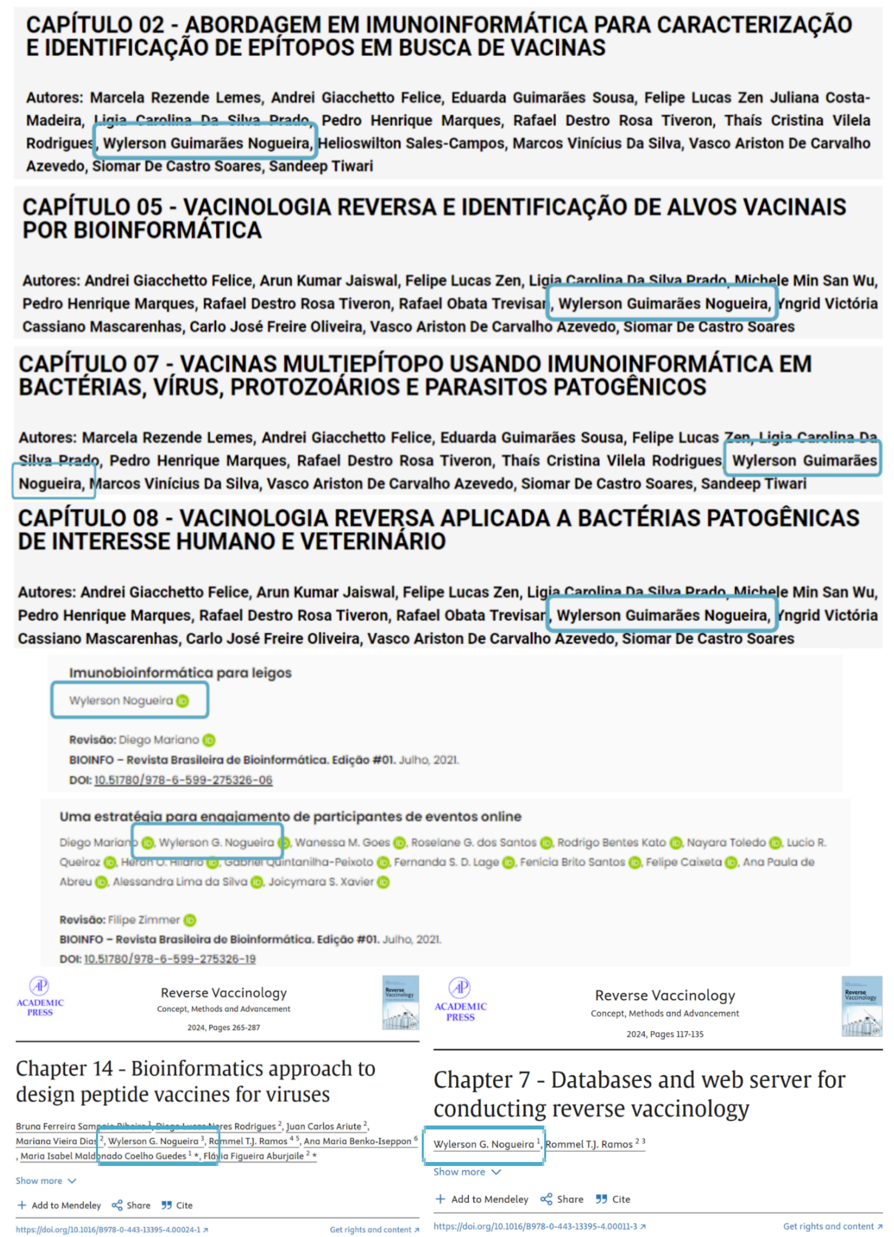
Diagramação
Alessandra Lima
Angie Atoche Puelles
Luana Luiza Bastos
Bibiana Fam
Bruna Espiño dos Santos
Diego Mariano
Marcos Antonio Nobrega de Sousa
Thiago de Camargo

Publicação
Editora Alfahelix, CNPJ: 37.524.984/0001-10
Lagoa Santa, MG, Brasil
www.alfahelix.com.br

Capa
Adaptado de rawpixel.com/Freepik. Estruturas da hemoglobina e do tRNA foram obtidas no PDB e renderizadas com ChimeraX.

Neste livro, intitulado "BIOINFO #03 - Revista Brasileira de Bioinformática e Biologia Computacional" (ISBN: 978-65-992753-8-8 | doi.org/10.51780/978-65-992753-8-8), eu participei da organização e revisão do manuscrito. Este e-book é fruto da 3ª edição de produções bibliográficas de BIOINFO, uma plataforma para divulgação científica em bioinformática, que tem como objetivo fornecer material de estudo de fácil compreensão referente a área em português e fornecer a alunos de pós-graduação a possibilidade de compartilhar seus conhecimentos através de artigos e tutoriais, bem como exercitar sua habilidade de comunicação científica.

ANEXO J – Capítulos de Livro Publicados



CAPÍTULO 02 - ABORDAGEM EM IMUNOINFORMÁTICA PARA CARACTERIZAÇÃO E IDENTIFICAÇÃO DE EPÍTOPOS EM BUSCA DE VACINAS

Autores: Marcela Rezende Lemes, Andrei Giacchetto Felice, Eduarda Guimarães Sousa, Felipe Lucas Zen Juliana Costa-Madeira, Ligia Carolina Da Silva Prado, Pedro Henrique Marques, Rafael Destro Rosa Tiveron, Thaís Cristina Vilela Rodrigues, Wylerson Guimarães Nogueira, Helioswilton Sales-Campos, Marcos Vinícius Da Silva, Vasco Ariston De Carvalho Azevedo, Siomar De Castro Soares, Sandeep Tiwari

CAPÍTULO 05 - VACINOLOGIA REVERSA E IDENTIFICAÇÃO DE ALVOS VACINAIS POR BIOINFORMÁTICA

Autores: Andrei Giacchetto Felice, Arun Kumar Jaiswal, Felipe Lucas Zen, Ligia Carolina Da Silva Prado, Michele Min San Wu, Pedro Henrique Marques, Rafael Destro Rosa Tiveron, Rafael Obata Trevisan, Wylerson Guimarães Nogueira, Yngrid Victória Cassiano Mascarenhas, Carlo José Freire Oliveira, Vasco Ariston De Carvalho Azevedo, Siomar De Castro Soares

CAPÍTULO 07 - VACINAS MULTIEPÍTOPO USANDO IMUNOINFORMÁTICA EM BACTÉRIAS, VÍRUS, PROTOZOÁRIOS E PARASITOS PATOGÊNICOS

Autores: Marcela Rezende Lemes, Andrei Giacchetto Felice, Eduarda Guimarães Sousa, Felipe Lucas Zen, Ligia Carolina Da Silva Prado, Pedro Henrique Marques, Rafael Destro Rosa Tiveron, Thaís Cristina Vilela Rodrigues, Wylerson Guimarães Nogueira, Marcos Vinícius Da Silva, Vasco Ariston De Carvalho Azevedo, Siomar De Castro Soares, Sandeep Tiwari

CAPÍTULO 08 - VACINOLOGIA REVERSA APLICADA A BACTÉRIAS PATOGÊNICAS DE INTERESSE HUMANO E VETERINÁRIO

Autores: Andrei Giacchetto Felice, Arun Kumar Jaiswal, Felipe Lucas Zen, Ligia Carolina Da Silva Prado, Michele Min San Wu, Pedro Henrique Marques, Rafael Destro Rosa Tiveron, Rafael Obata Trevisan, Wylerson Guimarães Nogueira, Yngrid Victória Cassiano Mascarenhas, Carlo José Freire Oliveira, Vasco Ariston De Carvalho Azevedo, Siomar De Castro Soares

Imunobioinformática para leigos

Wylerson Nogueira

Revisão: Diego Mariano
BIOINFO – Revista Brasileira de Bioinformática. Edição #01. Julho, 2021.
DOI: 10.51780/978-6-599-275326-06

Uma estratégia para engajamento de participantes de eventos online

Diego Mariano, Wylerson G. Nogueira, Wanessa M. Goes, Roselane G. dos Santos, Rodrigo Bentes Kato, Nayara Toledo, Lucio R. Queiroz, Heron O. Hilário, Gabriel Quintanilha-Peixoto, Fernanda S. D. Lage, Fenícia Brito Santos, Felipe Caixeta, Ana Paula de Abreu, Alessandra Lima da Silva, Joicymara S. Xavier

Revisão: Filipe Zimmer
BIOINFO – Revista Brasileira de Bioinformática. Edição #01. Julho, 2021.
DOI: 10.51780/978-6-599-275326-19

ACADEMIC PRESS
Reverse Vaccinology
Concept, Methods and Advancement
2024, Pages 265-287

Chapter 14 - Bioinformatics approach to design peptide vaccines for viruses

Bruna Ferreira Sampaio Ribeiro [1], Diego Lucas Neres Rodrigues [2], Juan Carlos Ariute [2], Mariana Vieira Dias [2], Wylerson G. Nogueira [3], Rommel T.J. Ramos [4 5], Ana Maria Benko-Iseppon [6], Maria Isabel Maldonado Coelho Guedes [1] *, Flávia Figueira Aburjaile [2] *

Show more
Add to Mendeley Share Cite
https://doi.org/10.1016/B978-0-443-13395-4.00024-1 Get rights and content

ACADEMIC PRESS
Reverse Vaccinology
Concept, Methods and Advancement
2024, Pages 117-135

Chapter 7 - Databases and web server for conducting reverse vaccinology

Wylerson G. Nogueira [1], Rommel T.J. Ramos [2 3]

Show more
Add to Mendeley Share Cite
https://doi.org/10.1016/B978-0-443-13395-4.00011-3 Get rights and content

Neste período, participei da publicação de 2 capítulos no livro “Reverse Vaccinology”, como primeiro autor e como co-autor (doi.org/10.1016/B978-0-443-13395-4.00011-3 | doi.org/10.1016/B978-0-443-13395-4.00024-1); 4 capítulos no livro “O ESTADO DA ARTE NAS PESQUISAS EM VACINOLOGIA”, todos como co-autor (doi.org/10.53924/vac1.02 | doi.org/10.53924/vac1.05 | doi.org/10.53924/vac1.07 | doi.org/10.53924/vac1.08); e, 2 capítulos no livro “BIOINFO – Revista Brasileira de Bioinformática. Edição #01”, como autor único e como co-autor (doi.org/10.51780/978-6-599-275326-06 | doi.org/10.51780/978-6-599-275326-19).

ANEXO K – Currículo Lattes – 01/2021 à 04/2025

Wylerson Guimarães Nogueira

Endereço para acessar este CV: **https://lattes.cnpq.br/7773763584173970**

Última atualização do currículo em 20/04/2025

Doutorado no Programa Interunidades de Pós-Graduação em Bioinformática (CAPES 7) da Universidade Federal de Minas Gerais (UFMG), com mestrado em Bioinformática pela UFMG e bacharelado em Biotecnologia pela Universidade Federal do Pará (UFPA). Wylerson realizou pesquisas em imunobioinformática, com foco em bactérias zoonóticas e patógenos sexualmente transmissíveis, além de One Health, em colaboração com a RECOM Rede de Ciências Ômicas. Sua pesquisa atual aborda a ameaça à saúde pública representada pela resistência aos antimicrobianos, com o objetivo de mapear a disseminação da resistência antimicrobiana nos corpos d'água da Amazônia, um hotspot de biodiversidade com inúmeros potenciais patógenos emergentes. Recentemente, juntou-se ao Laboratório de Ecogenômica Microbiana e Biotecnologia (MEGBLab) na King Abdullah University of Science and Technology (KAUST) como estudante visitante sob a supervisão do Prof. Alexandre Rosado. **(Texto informado pelo autor)**

Identificação

Nome Wylerson Guimarães Nogueira

Lattes ID 7773763584173970

Orcid ID https://orcid.org/0000-0001-9910-0690

Nome em citações bibliográficas NOGUEIRA, W. G.;NOGUEIRA, WYLERSON;NOGUEIRA, WYLERSON G.;NOGUEIRA, WYLERSON. G.;NOGUEIRA, WYLERSON GUIMARÃES;NOGUERA, WYLERSON GUIMARÃES

Prêmios e títulos

2023 Prêmio BIOINFO de Mérito em Divulgação Científica - Best Paper Award, Revista BIOINFO

2019 Biotecnologias para Remediação e Monitoramento Ambiental: Menção Honrosa., Centro Brasileiro-Argentino de Biotecnologia - CBAB

2006 Olimpíada Brasileira de Astronomia e Astronáutica, Medalha de Bronze, Agência Espacial Brasileira (AEB)

Formação acadêmica/titulação

2021 Doutorado em Bioinformática.
Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil
Orientador: Rommel Thiago Jucá Ramos

2018 - 2021 Mestrado em Bioinformática.
Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil
Título: Identificação computacional de alvos de drogas e candidatos a vacinas em Mycoplasma genitalium, Ano de obtenção: 2021
Orientador: Rommel Thiago Jucá Ramos
Bolsista do(a): Fundação de Amparo à Pesquisa do Estado de Minas Gerais, FAPEMIG, Brasil.
Palavras-chave: Bioinformatics, Reverse Vaccinology, Immunobioinformatics.
Grande área: Ciências Biológicas / Área: Biotecnologia
Grande área: Ciências Biológicas / Área: Biotecnologia

2011 - 2018 Graduação em Biotecnologia.
Universidade Federal do Pará, UFPA, Belem, Brasil
Título: Predição in silico de alvos antigênicos em pan-exoproteoma de Corynebacterium pseudotuberculosis
Orientador: Rommel Thiago Jucá Ramos
Bolsista do(a): Fundação Amazônia Paraense de Amparo à Pesquisa, FAPESPA, Brasil.

2008 - 2010 Ensino Médio (2o grau) .
Sistema Elite de Ensino, ELITE, Brasil

2024 - 2024 Aperfeiçoamento em CAPES-PRINT - 88887.911664/2023-00.
King Abdullah University of Science and Technology, KAUST, Jeddah, Arábia Saudita
Título: Metagenomic characterization of water bodies in the Amazon to monitor the spread of antimicrobial resistance in the environment
Orientador: Alexandre Soares Rosado
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil.

Formação complementar

2023 - 2023 Curso de curta duração em CADDE Workshop on Portable Metagenomics for Pathogen Surveillance. (Carga horária: 40h).
Instituto de Medicina Tropical - Universidade de São Paulo, IMT, Brasil

2022 - 2022 Curso de curta duração em Vacin. Reversa: Clonagem, Expressão e Avaliação de Antígenos Recombinantes. (Carga horária: 40h).
Centro Brasileiro-Argentino de Biotecnologia, CBAB, Brasil

2021 - 2021 Curso de curta duração em Curso de Preparação e Resposta às Emergências em Saúde Pública. (Carga horária: 41h).
Organizacão Pan-Americana da Saude/Organizacão Mundial da Saude, OPAS/OMS, Brasília, Brasil

2019 - 2019 Curso de curta duração em Biotecnologias para Remediação e Monitoramento Ambiental. (Carga horária: 48h).
Centro Brasileiro-Argentino de Biotecnologia, CBAB, Brasil

2019 - 2019 Curso de curta duração em Curso OPAS: Tecnologia de sequenciamento genético baseada em nanoporos. (Carga horária: 96h).
Organizacão Pan-Americana da Saude/Organizacão Mundial da Saude, OPAS/OMS, Brasília, Brasil

2019 - 2019 Curso de curta duração em Mining Microbial and Viral Genomes and Metagenomes for Biotechnological App. (Carga horária: 10h).
DOE's Joint Genome Institute, JGI, Estados Unidos

2012 - 2012 Extensão universitária em II Curso de Verão em Bioinformática Estrutural da UFMG. (Carga horária: 40h).
Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil

Atuação profissional

1Sociedade Brasileira de Genética - SBG

2023 - 2023 Vínculo: Monitor , Enquadramento funcional: Monitor , Carga horária: 24, Regime: Sociedade Brasileira de Genética Parcial
Outras informações:
Monitoria no evento internacional GENÉTICA 2023 68th Brazilian Congress of Genetics

2Universidade Federal de Minas Gerais - UFMG

2023 - 2023 Vínculo: Monitor , Enquadramento funcional: Monitor , Carga horária: 19, Regime: Universidade Federal de Minas Gerais Parcial
Outras informações:
Monitor do 2 Congresso Internacional de Ciência, Biodiversidade e Sustentabilidade.

3King Abdullah University of Science and Technology - KAUST

2024 - 2024 Vínculo: Academic Mentor , Enquadramento funcional: KAUST Space2102 Program - Academic Mentor , Carga horária: 40, Regime: King Abdullah University of Science and Technology Integral

2024 - 2024 Vínculo: Assistant Mentor , Enquadramento funcional: SRSI 2024 - Assistant Mentor , Carga horária: 40, Regime: King Abdullah University of Science and Technology Integral

Producão

Produção bibliográfica

Artigos completos publicados em periódicos

1. doi> SILVERIO, MYLLENA PEREIRA; SCHULTZ, JÚNIA; PARISE, MARIANA T. D.; PARISE, DOGLAS; VIANA, MARCUS VINICIUS CANÁRIO; **NOGUEIRA, WYLERSON**; RAMOS, ROMMEL THIAGO JUCÁ; GÓES-NETO, ARISTOTELES; AZEVEDO, VASCO ARISTON DE CARVALHO; BRENIG, BERTRAM; BONELLI, RAQUEL REGINA; ROSADO, ALEXANDRE SOARES. Genomic and phenotypic insight into antimicrobial resistance of Pseudomonas fluorescens from King George Island, Antarctica. Frontiers in Microbiology. JCR, v.16, p.01 - 16, 2025.

2. doi> CARDENAS-ALEGRIA, OSCAR VICTOR; FERREIRA, VICTOR BENEDITO COSTA; **NOGUERA, WYLERSON GUIMARÃES**; MARTINS, DAVID TAVARES; MARTINS NETO, ARTUR PEDRO; MONTEIRO PONTES, PAULO RÓGENES; LOPES CAVALCANTE, ROSANE BARBOSA; AGUIAR ALVES, SANDY INGRID; LUIZ DA COSTA DA SILVA, ARTUR; GOMES COSTA, ROSILENE; FRANCO DE LOS SANTOS, EDIAN FRANKLIN; AZEVEDO, VASCO ARISTON DE CARVALHO; RAMOS, ROMMEL THIAGO JUCA. Microbiome analyses of the Uraim River in the Amazon and georeferencing analyses to establish correlation with anthropogenic impacts of land use. FRONTIERS IN ENVIRONMENTAL SCIENCE. JCR, v.12, p.01 - 12, 2024.

3. doi> DANTAS, CARLOS WILLIAN DIAS; MARTINS, DAVID TAVARES; **NOGUEIRA, WYLERSON GUIMARÃES**; ALEGRIA, OSCAR VICTOR CARDENAS; RAMOS, ROMMEL THIAGO JUCÁ. Tools and methodology to in silico phage discovery in freshwater environments. Frontiers in Microbiology. JCR, v.15, p.01 - 10, 2024. **Citações:** WEB OF SCIENCE ™ 1 | SCOPUS 1

4. doi> PINHEIRO, K. C.; GOIS, B. V. A.; NOGUEIRA, W. G.; ARAUJO, F. A.; CAVALCANTE, A. L. Q.; ALEGRIA, O. V. C.; SILVA, A. L. C.; MARTINS JUNIOR, A.; RAMOS, R. T. J.. In silico approach to identify microsatellites candidate biomarkers to differentiate the biovar of Corynebacterium pseudotuberculosis genomes. Frontiers in Bioinformatics. v.1, p.1 - 15, 2022. **Citações:** WEB OF SCIENCE ™ 4 | SCOPUS 4

5. doi> **NOGUEIRA, WYLERSON GUIMARÃES**; GOIS, BRUNA VERÔNICA AZEVEDO; PINHEIRO, KENNY DA COSTA; ARAGÃO, ANDRESSA DE OLIVEIRA; QUEIROZ, ANA LÍDIA CAVALCANTE; SILVA, ARTUR LUIZ DA; FOLADOR, ADRIANA CARNEIRO; RAMOS, ROMMEL THIAGO JUCÁ. Viral Metagenomics Reveals Widely Diverse Viral Community of Freshwater Amazonian Lake. FRONTIERS IN PUBLIC HEALTH. JCR, v.10, p.11 - 10, 2022. **Citações:** WEB OF SCIENCE ™ 7 | SCOPUS 6

6. doi> **NOGUEIRA, WYLERSON G.**; JAISWAL, ARUN KUMAR; TIWARI, SANDEEP; RAMOS, ROMMEL T.J.; GHOSH, PREETAM; BARH, DEBMALYA; AZEVEDO, VASCO; SOARES, SIOMAR C.. Computational identification of putative common genomic drug and vaccine targets in Mycoplasma genitalium. GENOMICS. JCR, v.113, p.2730 - 2743, 2021. **Citações:** WEB OF SCIENCE ™ 12 | SCOPUS 12

7. doi> ADELINO, TALITA ÉMILE RIBEIRO; GIOVANETTI, MARTA; FONSECA, VAGNER; XAVIER, JOILSON; DE ABREU, ÁLVARO SALGADO; DO NASCIMENTO, VALDINETE ALVES; DEMARCHI, LUIZ HENRIQUE FERRAZ; OLIVEIRA, MARLUCE APARECIDA ASSUNÇÃO; DA SILVA, VINÍCIUS LEMES; DE MELLO, ARABELA LEAL E. SILVA; CUNHA, GABRIEL MURICY; SANTOS, ROSELENE HANS; DE OLIVEIRA, ELAINE CRISTINA; JÚNIOR, JORGE ANTÔNIO CHAMON; DE MELO IANI, FELIPE CAMPOS; DE FILIPPIS, ANA MARIA BISPO; DE ABREU, ANDRÉ LUIZ; DE JESUS, RONALDO; DE ALBUQUERQUE, CARLOS FREDERICO CAMPELO; RICO, JAIRO MENDEZ; DO CARMO SAID, RODRIGO FABIANO; SILVA, JOSCÉLIO AGUIAR; DE MOURA, NOELY FABIANA OLIVEIRA; LEITE, PRISCILA; FRUTUOSO, LÍVIA CARLA VINHAL; *et al* . Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil. Nature Communications. JCR, v.12, p.01 - 12, 2021. **Citações:** WEB OF SCIENCE ™ 35 | SCOPUS 33

8. doi> DA SILVA, ALESSANDRA LIMA; ABREU, ANA PAULA DE; MARIANO, DIEGO; CAIXETA, FELIPE; SANTOS, FENÍCIA BRITO; LAGE, FERNANDA STUSSI D.; QUINTANILHA-PEIXOTO, GABRIEL; HILÁRIO, HERON. O.; XAVIER, JOICYMARA. S.; QUEIROZ, LUCIO. R.; DE TOLEDO, NAYARA EVELIN; TAVARES, RAPHAEL; KATO, RODRIGO BENTES; DOS SANTOS, ROSELANE GONÇALVES; SOARES, STELLAMARIS; GOES, WANESSA. M.; **NOGUEIRA, WYLERSON. G.**; BATISTA, THIAGO. M.; ORTEGA, JOSÉ MIGUEL; DE CARVALHO, VASCO ARISTON AZEVEDO; FRANCO, GLÓRIA. REGINA; MELO-MINARDI, RAQUEL. C. DE; GÓES-NETO, ARISTÓTELES. From In-Person to the Online World: Insights Into Organizing Events in Bioinformatics. Frontiers in Bioinformatics. v.1, p.1 - 10, 2021. **Citações:** WEB OF SCIENCE ™ 1 | SCOPUS 1

Livros publicados

1. MARIANO, DIEGO; Da Silva, A. P. D.; Cremonesi, A. S.; BULLA, Ana Carolina Silva; ABREU, ANA PAULA DE; Gonçalves, A. R.; Pinheiro, B. R. M.; Fam, B.; Dos Santos, B. E.; Capelini, C.; Rodrigues, D. L. N.; Teixeira, F. A.; CANSANÇÃO, Isaac Farias; MAMEDE, I.; BASTOS, Luana Luiza; SANTOS, Lucianna Helene; De Sousa, M. A. N.; De Lima, M. R. P.; Lemos, R. P.; COSTA, Savio; CAMARGO, T. M. N.; BORELLI, T. C.; **NOGUEIRA, W. G.**; PUELLES, Angie Atoche; OLIVEIRA, E. S.; BAPTISTA, G.; MOREIRA, J. C.; SCHEKIERA, L. G. A.; PAIXÃO, Vivian; LIMA, A.. BIOINFO #03 - Revista Brasileira de Bioinformática e Biologia Computacional, ed.3. Lagoa Santa - MG: Alfahelix Publicações, 2023

2. BASTOS, Luana Luiza; PUELLES, Angie Atoche; PAIXÃO, Vivian; FIORINI, Giovanna; UTSCH, Eduardo; **NOGUEIRA, W. G.**; BIALVES, Tatiane Senna; SANTOS, Lucianna Helene; BULLA, Ana Carolina Silva; SANCHES, Flavia; BRANDÃO, Débora Soares; CANSANÇÃO, Isaac Farias; COSTA, Savio; CARVALHO, Frederico Chaves; MARIANO, DIEGO. BIOINFO #02 - Revista Brasileira de Bioinformática e Biologia Computacional, ed.2. Lagoa Santa - MG: Editora Alfahelix, 2022, v.2., p.167.

Capítulos de livros publicados

1. Ribeiro, Bruna Ferreira Sampaio; Rodrigues, Diego Lucas Neres; Ariute, Juan Carlos; Dias, Mariana Vieira; **NOGUEIRA, WYLERSON G.**; RAMOS, ROMMEL T.J.; Benko-Iseppon, Ana Maria; Guedes, Maria Isabel Maldonado Coelho; ABURJAILE, FLÁVIA FIGUEIRA. Bioinformatics approach to design peptide vaccines for viruses In: Reverse Vaccinology, ed.01. : Elsevier, 2024, p. 265 - 287.

2. **NOGUEIRA, WYLERSON G.**; RAMOS, ROMMEL T.J.. Databases and web server for conducting reverse vaccinology In: Reverse Vaccinology, ed.01. : Elsevier, 2024, p. 117 - 135.

3. Lemes, M. R.; Felice, A. G.; Sousa, E. G.; Zen, F. L.; Costa-Madeira, J.; Prado, L. C. S.; Marques, P. H.; Tiveron, R. D. R.; Rodrigues, T. C. V.; **NOGUEIRA, W. G.**; Sales-Campos, H.; da Silva, M. V.; AZEVEDO, V. A. C.; SOARES, S. C.; TIWARI, S.. Abordagem Em Imunoinformática Para Caracterização E Identificação De Epítopos Em Busca De Vacinas In: O ESTADO DA ARTE NAS PESQUISAS EM VACINOLOGIA, ed.1. João Pessoa - PB: Editora Creative, 2022, p. 17 - 40.

4. Lemes, M. R.; Felice, A. G.; Sousa, E. G.; Zen, F. L.; Prado, L. C. S.; Marques, P. H.; Tiveron, R. D. R.; Rodrigues, T. C. V.; **NOGUEIRA, W. G.**; da Silva, M. V.; AZEVEDO, V. A. C.; SOARES, S. C.; TIWARI, S.. Vacinas Multiepítopo Usando Imunoinformática Em Bactérias, Vírus, Protozoários E Parasitos Patogênicos In: O ESTADO DA ARTE NAS PESQUISAS EM VACINOLOGIA, ed.1. João Pessoa - PB: Editora Creative, 2022, p. 121 - 136.

5. Felice, A. G.; JAISWAL, A. K.; Zen, F. L.; Prado, L. C. S.; Wu, M. M. S.; Marques, P. H.; Tiveron, R. D. R.; Trevisan, R. O.; **NOGUEIRA, W. G.**; Mascarenhas, Y. V. C.; Oliveira, C. J. F.; AZEVEDO, V. A. C.; SOARES, S. C.. Vacinologia Reversa Aplicada A Bactérias Patogênicas De Interesse Humano E Veterinário In: O ESTADO DA ARTE NAS PESQUISAS EM VACINOLOGIA, ed.1. João Pessoa - PB: Editora Creative, 2022, p. 137 - 159.

6. Felice, A. G.; JAISWAL, A. K.; Zen, F. L.; Prado, L. C. S.; Wu, M. M. S.; Marques, P. H.; Tiveron, R. D. R.; Trevisan, R. O.; **NOGUEIRA, W. G.**; Mascarenhas, Y. V. C.; Oliveira, C. J. F.; AZEVEDO, V. A. C.; SOARES, S. C.. Vacinologia Reversa E Identificação De Alvos Vacinais Por Bioinformática In: O ESTADO DA ARTE NAS PESQUISAS EM VACINOLOGIA, ed.1. João Pessoa - PB: Editora Creative, 2022, p. 77 - 101.

7. **NOGUEIRA, WYLERSON**. Imunobioinformática para leigos In: BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional, ed.1. : Alfahelix, 2021, v.1, p. 57 - 76.

8. MARIANO, DIEGO; **NOGUEIRA, WYLERSON G.**; Goes, Wanessa M.; dos Santos, Roselane G.; KATO, RODRIGO BENTES; Toledo, Nayara; Queiroz, Lucio R.; Hilário, Heron O.; QUINTANILHA-PEIXOTO, GABRIEL; Lage, Fernanda S. D.; SANTOS, FENÍCIA BRITO; CAIXETA, FELIPE; de Abreu, Ana Paula; DA SILVA, ALESSANDRA LIMA; Xavier, Joicymara S.. Uma estratégia para engajamento de participantes de eventos online In: BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional, ed.1. : Alfahelix, 2021, v.1, p. 281 - 291.

Trabalhos publicados em anais de eventos (resumo)

1. PUELLES, Angie Atoche; SANTOS, G. A. C.; PAIXÃO, Vivian; **NOGUEIRA, W. G.**; GOMES, L. G. R.; MELO-MINARDI, RAQUEL. C. DE. An engineered chimeric protein against Neisseria gonorrhoeae as a new multi-epitope vaccine candidate to prevent the sexually transmitted infections of Gonorrhea In: GENÉTICA 2023 - 68 Brazilian Congress of Genetics, 2023, Ouro Preto - MG. **Livro de Resumos do Congresso de Genética 2023**. Editora SBG, 2023, v.01,

2. **NOGUEIRA, W. G.**; JAISWAL, A. K.; GOMES, L. G. R.; SOARES, S. C.; AZEVEDO, V. A. C.; RAMOS, R. T. J.; TIWARI, S.. An immunoinformatics-based designed multi-epitope vaccine against the sexually transmitted bacteria Mycoplasma genitalium In: GENÉTICA 2023 - 68 Brazilian Congress of Genetics, 2023, Ouro Preto - MG. **Livro de Resumos do Congresso de Genética 2023**. Editora SBG, 2023, v.01,

3. MELO-MINARDI, RAQUEL. C. DE; BASTOS, Luana Luiza; PAIXÃO, Vivian; BULLA, Ana Carolina Silva; PUELLES, Angie Atoche; MAIA, G. C. F.; COSTA, Savio; ABREU, A. P.; CANSANÇÃO, Isaac Farias; **NOGUEIRA, W. G.**; Teixeira, F. A.; BAPTISTA, G.; OLIVEIRA, E. S.; SOUSA, T. J.; BORELLI, T. C.; BIALVES, Tatiane Senna; SANCHES, F. S.; MOREIRA, E. U. M.; CARVALHO, Frederico Chaves; XAVIER, J. S.; SANTOS, L. H. S.; MARIANO, D. C. B.. BIOINFO: The Brazilian Journal of Bioinformatics and Computational Biology In: X-meeting / BSB 2023, 2023, Curitiba - PR. **X-Meeting presentations**. 2023, v.01,

4. DA SILVA, ALESSANDRA LIMA; MAIA, G. C. F.; MAMEDE, I.; COSTA, Savio; BAPTISTA, G.; CAMARGO, T. M. N.; FAM, B. S. O.; **NOGUEIRA, W. G.**; MELO-MINARDI, RAQUEL. C. DE; MARIANO, D. C. B.. BIOINFO: the first Brazilian journal of scientific dissemination of bioinformatics and computational biology in Portuguese In: GENÉTICA 2023 - 68 Brazilian Congress of Genetics, 2023, Ouro Preto - MG. **Livro de Resumos do Congresso de Genética 2023**. Editora SBG, 2023, v.01,

5. MAIA, G. C. F.; PAIXÃO, Vivian; GOMES, L. G. R.; **NOGUEIRA, W. G.**; RAMOS, R. T. J.; MELO-MINARDI, RAQUEL. C. DE. In silico prediction of antigenic targets of Mesomycoplasma hyopneumoniae: an immunoinformatics approach In: GENÉTICA 2023 - 68 Brazilian Congress of Genetics, 2023, Ouro Preto - MG. **Livro de Resumos do Congresso de Genética 2023**. Editora SBG, 2023, v.01,

Apresentação de trabalho e palestra

1. PUELLES, Angie Atoche; SANTOS, G. A. C.; PAIXÃO, Vivian; **NOGUEIRA, W. G.**; GOMES, L. G. R.; MELO-MINARDI, RAQUEL. C. DE. An engineered chimeric protein against neisseria gonorrhoeae as a new multi-epitope vaccine candidate to prevent the sexually transmitted infections of gonorrhea, 2023. (Congresso,Apresentação de Trabalho)

2. **NOGUEIRA, W. G.**; JAISWAL, A. K.; GOMES, L. G. R.; SOARES, S. C.; AZEVEDO, V. A. C.; RAMOS, R. T. J.; TIWARI, S.. An immunoinformatics-based designed multi-epitope vaccine against the sexually transmitted bacteria Mycoplasma genitalium, 2023. (Congresso,Apresentação de Trabalho)

3. MELO-MINARDI, RAQUEL. C. DE; BASTOS, Luana Luiza; PAIXÃO, Vivian; BULLA, Ana Carolina Silva; PUELLES, Angie Atoche; MAIA, G. C. F.; COSTA, Savio; ABREU, A. P.; CANSANÇÃO, Isaac Farias; **NOGUEIRA, W. G.**; Teixeira, F. A.; BAPTISTA, G.; OLIVEIRA, E. S.; SOUSA, T. J.; BORELLI, T. C.; BIALVES, Tatiane Senna; SANCHES, F. S.; MOREIRA, E. U. M.; CARVALHO, Frederico Chaves; XAVIER, J. S.; SANTOS, L. H. S.; MARIANO, D. C. B.. BIOINFO: The Brazilian Journal of Bioinformatics and Computational Biology, 2023. (Congresso,Apresentação de Trabalho)

4. DA SILVA, ALESSANDRA LIMA; MAIA, G. C. F.; MAMEDE, I.; COSTA, Savio; BAPTISTA, G.; CAMARGO, T. M. N.; FAM, B. S. O.; **NOGUEIRA, W. G.**; MELO-MINARDI, RAQUEL. C. DE; MARIANO, D. C. B.. BIOINFO: the first Brazilian journal of scientific dissemination of bioinformatics and computational biology in Portuguese, 2023. (Congresso,Apresentação de Trabalho)

5. MAIA, G. C. F.; PAIXÃO, Vivian; GOMES, L. G. R.; **NOGUEIRA, W. G.**; RAMOS, R. T. J.; MELO-MINARDI, RAQUEL. C. DE. In silico prediction of antigenic targets of mesomycoplasma hyopneumoniae: an immunoinformatics approach, 2023. (Congresso,Apresentação de Trabalho)

6. **NOGUEIRA, W. G.**. Expectativa x Realidade, 2022. (Seminário,Apresentação de Trabalho)

7. **NOGUEIRA, W. G.**. Imunobioinformática aplicada ao combate de doenças veterinárias de interesse comercial, 2021. (Conferência ou palestra,Apresentação de Trabalho)

Eventos

Eventos

Participação em eventos

1. Apresentação de Poster / Painel no(a) **68th Brazilian Congress of Genetics**, 2023. (Congresso) An immunoinformatics-based designed multi-epitope vaccine against the sexually transmitted bacteria Mycoplasma genitalium.

2. **7th International Symposium on Immunobiologicals**, 2023. (Simpósio) Easy ISI 3 - Introdução à aplicação da bioinformática em imunobiológicos (curso virtual).

3. Apresentação de Poster / Painel no(a) **CADDE Workshop on Portable Metagenomics for Pathogen Surveillance**, 2023. (Outra) Viral Metagenomics Reveals Widely Diverse Viral Community of Freshwater Amazonian Lake.

4. **Simpósio e Diplomação dos Membros Afiliados da ABC Região SP 2023-2027**, 2023. (Simpósio) .

5. Apresentação de Poster / Painel no(a) **X-meeting / BSB 2023**, 2023. (Congresso) BIOINFO: The Brazilian Journal of Bioinformatics and Computational Biology.

6. **AULA MAGNA DA PÓS-GRADUAÇÃO DO ICB 2022 - UFMG**, 2022. (Outra) .

7. Apresentação Oral no(a) **CENTRO LATINO-AMERICANO DE BIOTECNOLOGIA - VI CURSO DE VACINOLOGIA REVERSA: CLONAGEM, EXPRESSÃO E AVALIAÇÃO DE ANTÍGENOS RECOMBINANTES**, 2022. (Outra) Expressão Recombinante da Proteína P42 como Antígeno contra a Pneumonia Enzoótica Suína.

8. **REUNIÃO MAGNA DA ABC 2022 - O FUTURO É AGORA**, 2022. (Seminário) .

9. **WEBINAR - MANEJO CLÍNICO DAS ARBOVIROSES - DENGUE, CHIKUNGUNYA, ZIKA E FEBRE AMARELA**, 2022. (Outra) .

10. **AULA MAGNA DA PÓS GRADUAÇÃO DO ICB 2021 - UFMG**, 2021. (Outra) .

11. **Ciclo de Palestras da 2ª Liga Brasileira de Bioinformática**, 2021. (Outra) .

12. **SIMPÓSIOS CIENTÍFICOS | ED.02 • MICROBIOLOGIA E IMUNOLOGIA | CIÊNCIAS SOCIAIS**, 2021. (Outra) .

13. **WEBINÁRIOS DA ABC – ED. 43 | LIBERDADE E RESPONSABILIDADE SOCIAL DA CIÊNCIA**, 2021. (Outra) .
14. **WEBINÁRIOS DA ABC – ED. 45 | COMUNICAÇÃO CIENTÍFICA: COMO FALAR MAIS ALTO DO QUE A DESINFORMAÇÃO?**, 2021. (Outra) .
15. **WEBINÁRIOS DA ABC – ED. 48 | COMPLICAÇÕES PÓS-COVID: ALERTA AO SISTEMA DE SAÚDE**, 2021. (Outra) .
16. **XMeeting XPerience 2021**, 2021. (Congresso) .

Organização de evento

1. **NOGUEIRA, W. G.**; MAMEDE, I.; GOMES, L. G. R.. Curso de Férias em Bioinformática da UFMG - 2025, 2025. (Outro, Organização de evento)
2. **NOGUEIRA, W. G.**. Curso de Inverno em Bioinformática 2023 da UFMG, 2023. (Outro, Organização de evento)
3. **NOGUEIRA, W. G.**. Curso de Verão em Bioinformática da UFMG - Edição de Inverno - 2022, 2022. (Outro, Organização de evento)
4. NOGUEIRA, W. G.; RAMOS, R. T. J.; HAMOY, I. G.. Bioinformática na Estrada 2.0 - Evento Online, 2021. (Outro, Organização de evento)

Totais de produção

Produção bibliográfica	
Artigos completos publicados em periódico	9
Livros publicados	1
Livros publicados	1
Capítulos de livros publicados	9
Trabalhos publicados em anais de eventos	12
Apresentações de trabalhos (Conferência ou palestra)	2
Apresentações de trabalhos (Congresso)	14
Apresentações de trabalhos (Seminário)	4
Apresentações de trabalhos (Simpósio)	1

Produção técnica	
Curso de curta duração ministrado (extensão)	4

Eventos	
Participações em eventos (congresso)	10
Participações em eventos (seminário)	4
Participações em eventos (simpósio)	7
Participações em eventos (oficina)	1
Participações em eventos (encontro)	3
Participações em eventos (outra)	14
Organização de evento (congresso)	3
Organização de evento (outro)	6

Página gerada pelo sistema Currículo Lattes em 21/04/2025 às 19:16:27.