UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Biológicas

Programa de Pós-Graduação em Microbiologia

Bruna Luiza de Azevedo

O ISOLAMENTO E A CARACTERIZAÇÃO GENÔMICA DE NOVOS ISOLADOS BRASILEIROS DO REINO *BAMFORDVIRAE* EXPANDEM A DIVERSIDADE E A TAXONOMIA DAS FAMÍLIAS *MARSEILLEVIRIDAE* E *SPUTNIVIROVIRIDAE*.

Belo Horizonte

2025

Bruna Luiza de Azevedo

O ISOLAMENTO E A CARACTERIZAÇÃO GENÔMICA DE NOVOS ISOLADOS BRASILEIROS DO REINO *BAMFORDVIRAE* EXPANDEM A DIVERSIDADE E A TAXONOMIA DAS FAMÍLIAS *MARSEILLEVIRIDAE* E *SPUTNIVIROVIRIDAE*.

Tese apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Doutora em Ciências Biológicas: Microbiologia.

Orientador: Prof. Jônatas Santos Abrahão

Coorientador: Prof. Rodrigo Araújo Lima Rodrigues

Belo Horizonte

2025

043 Azevedo, Bruna Luiza de.
O isolamento e a caracterização genômica de novos isolados brasileiros do Reino Bamfordvirae expandem a diversidade e a taxonomia das famílias Marseilleviridae e Sputniviroviridae [manuscrito] / Bruna Luiza de Azevedo. – 2025.
95 f. : il. ; 29,5 cm.

Orientador: Prof. Jônatas Santos Abrahão. Coorientador: Prof. Rodrigo Araújo Lima Rodrigues.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-graduação em Microbiologia.

1. Microbiologia. 2. Vírus gigantes. 3. Genoma Viral. 4. Taxonomia. I. Abrahão, Jônatas Santos. II. Rodrigues, Rodrigo Araújo Lima. III. Universidade Federal e Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 579

Ficha Catalográfica elaborada pela bibliotecária Fabiane C M Reis CRB 6 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PÓS-GRADUAÇÃO EM MICROBIOLOGIA

ATA DE DEFESA DE TESE

ATA DA DEFESA DE TESE DE **BRUNA LUIZA DE AZEVEDO**

Nº REGISTRO: 2021692455

Às 14:00 horas do dia **25 de junho de 2025**, reuniu-se, por via remota, a Comissão Examinadora composta pelos Drs. Giliane de Souza Trindade (Departamento de Microbiologia/ICB/UFMG), Eric Roberto Guimarães Rocha Aguiar (Universidade Estadual de Santa Cruz), Gabriel Magno de Freitas Almeida (UiT Noruega), Juliana Reis Cortines (UFRJ), Prof. Dr. Jônatas Santos Abrahão (Orientador) e o Prof. Dr. Rodrigo Araújo Lima Rodrigues (Coorientador). para julgar o trabalho final **"O isolamento e a caracterização genômica de novos isolados brasileiros do reino bamfordvirae expandem a diversidade e a taxonomia das famílias marseilleviridae e sputnivirovirìdae."** da aluna **Bruna Luiza de Azevedo**, requisito final para a obtenção do Grau de **DOUTORA EM CIÊNCIAS BIOLÓGICAS: MICROBIOLOGIA**. Abrindo a sessão, o Presidente da Comissão, Prof. Dr. Jônatas Santos Abrahão, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para a apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. A candidata foi considerada **APROVADA**. O resultado final foi comunicado publicamente à candidata 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. A candidata tem 60 (sessenta) dias, a partir desta data, para entregar a versão final da tese ao Programa de Pós-graduação em Microbiologia da UFMG e requerer seu diploma.

Belo Horizonte, 25 de junho de 2025

Membros da Banca:

Profa. Dra. Giliane de Souza Trindade

Prof. Dr. Eric Roberto Guimarães Rocha Aguiar

Dr. Gabriel Magno de Freitas Almeida

Prof. Dra. Juliana Reis Cortines

De acordo:

Prof. Dr. Jônatas Santos Abrahão

(Orientador)

Prof. Dr. Rodrigo Araújo Lima Rodrigues

(Coorientador)

Prof. Dr. Caio Tavares Fagundes
(Coordenador do Programa de Pós-graduação
em Microbiologia)

Documento assinado eletronicamente por **Eric Roberto Guimaraes Rocha Aguiar**, **Usuário Externo**, em 27/06/2025, às 15:11, 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 **Juliana Reis Cortines**, **Usuário Externo**, em 27/06/2025, às 20:05, 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 **Gabriel Magno de Freitas Almeida**, **Usuário Externo**, em 30/06/2025, às 03:56, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Caio Tavares Fagundes**, **Coordenador(a)**, em 30/06/2025, às 10:00, 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 **Jonatas Santos Abrahao**, **Professor do Magistério Superior**, em 30/06/2025, às 13:29, 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 **Rodrigo Araújo Lima Rodrigues**, **Professor do Magistério Superior**, em 03/07/2025, às 09:38, 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 **Giliane de Souza Trindade**, **Professora do Magistério Superior**, em 09/07/2025, às 15:21, 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 **4309560** e o código CRC **46A3A8BF**.

Ao meu marido, aos meus pais, meu irmão, e à Nina: "Diante da vastidão do espaço e da imensidade do tempo, é uma alegria para mim partilhar um planeta e uma época com vocês."

– Carl Sagan

AGRADECIMENTOS

Aos meus pais, Rose e Amarildo, por nunca medirem esforços para me dar todo o suporte necessário para que eu fosse bem-sucedida em todos os âmbitos da minha vida. Obrigada por tanta paciência e amor dedicados a mim e ao meu irmão há tantos anos. Sem vocês, eu jamais seria capaz chegar aonde cheguei hoje.

Ao Lucas, meu irmão e companheiro. Você é um ponto de apoio essencial desde o seu nascimento, sempre tão bondoso e amoroso. Obrigada por existir e estar sempre comigo. Você é casa.

Ao Guilherme Prata, o amor da minha vida, meu marido, namorado, companheiro, amigo e tantas outras categorias que ocupa no meu coração. Obrigada por todo o incentivo, pela escuta, por estar presente sempre que eu achei que não ia conseguir e por me deixar sempre com os pés no chão. Sem você, esse trabalho jamais existiria.

À minha filha pet, Nina, que há 13 anos é um suporte emocional. Com a linguagem do seu olhar brilhante e de seu rabinho é capaz de expressar tanto amor, gratidão e ternura, o que sempre são motivos para eu sorrir.

Ao meu orientador, professor Jônatas. Obrigada por um dia ter confiado a mim a responsabilidade de desenvolver projetos tão ricos e de alto nível. Você mudou a minha vida ao me aceitar como aluna. Serei eternamente grata por tudo que aprendi com você ao longo desses oito anos e me sinto extremamente privilegiada por tudo que você me permitiu conquistar como profissional. Você sempre será uma inspiração e exemplo de profissional de altíssimo nível em tudo que faz. Obrigada pela compreensão e paciência diante dos inúmeros obstáculos durante o doutorado.

Ao meu coorientador, professor Rodrigo, por estar sempre disponível para ajudar quando precisei e por ser sempre paciente para compartilhar seu conhecimento. Se consegui desenvolver parte desse trabalho sozinha foi porque um dia você, com muita paciência, me ensinou o caminho.

Às minhas amigas de Biologia, Isabella e Beatriz, minhas parceiras desde o primeiro dia de aula da graduação. 13 anos se passaram e eu tenho muito orgulho de ainda hoje estarmos juntas. Obrigada pela presença e pela escuta. Sem vocês, tudo seria muito mais complicado.

Às minhas amigas do laboratório, Isabella e Talita. Obrigada por todo o apoio desde a iniciação científica até o doutorado. Saber que vocês estavam ali era um porto seguro para enfrentar qualquer desafio que a academia pudesse proporcionar, desde o apoio para fazer um experimento complicado ou para escrever um artigo, até às boas gargalhadas depois de uma palhaçada qualquer durante um lanche ou durante os 15 minutos de UV. Obrigada por deixarem tudo mais leve.

A todos os colegas e amigos do GEPVIG, egressos e atuais. Vocês foram essenciais na minha formação. Ao longo dos anos, foram várias pessoas que me ensinaram de tudo um pouco, engrandecendo a minha jornada acadêmica.

Aos colegas do Laboratório de Vírus. A família Labviridae sempre foi para mim um exemplo de competência e de trabalho em equipe de sucesso. Obrigada pela convivência e por todo o suporte, tanto no âmbito técnico, quanto em um simples lanche na cozinha. Já sinto tanta falta *de tudo e um imenso vazio ao saber que me desvincularei oficialmente do laboratório e esse sentimento é graças ao ambiente saudável que encontrei.*

Aos professores do Laboratório de Vírus. É um privilégio ter tido a oportunidade de compartilhar um local de trabalho e estudo com pesquisadores tão renomados, sábios e excelentes em suas pesquisas.

Ao Centro de microscopia da UFMG, pelo trabalho extremamente bem qualificado, que enriquece as nossas pesquisas com imagens incríveis.

À banca avaliadora pela disponibilidade em colaborar para a melhoria deste trabalho.

Às agências de fomento, CNPq, CAPES e FAPEMIG pelo auxílio financeiro, que garantiu o desenvolvimento deste trabalho.

RESUMO

Os vírus gigantes de amebas representam uma das áreas mais intrigantes da virologia moderna, não apenas por sua morfologia e genoma complexos, mas também por sua relevância evolutiva e taxonômica. Esses vírus estão classificados taxonomicamente no reino *Bamfordvirae,* sendo representados por diferentes famílias virais, como *Marseilleviridae*, *Mimiviridae* e *Sputniviroviridae.* Aqui, descrevemos novos isolados de bamfordvírus obtidos de amostras ambientais brasileiras. Na primeira parte, é descrito o Marseillevirus cajuinensis, primeiro isolado da família *Marseilleviridae* proveniente de água salgada. Este vírus apresenta um genoma com cerca de 380 mil pares de bases, contendo 515 ORFs, entre elas 38 ORFans, e um RNA transportador, um tipo de gene, até então, raramente descrito entre os Marseillevírus. As análises filogenéticas mostraram que o M. cajuinensis é um membro divergente da linhagem A e revelaram uma organização dos Marseillevírus em três grandes grupos, questionando a divisão tradicional em cinco linhagens e sugerindo a necessidade de reorganização taxonômica. Com base nesses dados, foi submetida ao Comitê Internacional de Taxonomia viral (ICTV) uma proposta de inclusão de novos táxons na família *Marseilleviridae*, incluindo a criação de novas espécies e gêneros para ampliar sua diversidade taxonômica. A segunda parte do trabalho aborda o isolamento de um sistema viral tripartite composto por um moumouvirus, um transpoviron e um novo virófago, denominado virófago Pantanal, a partir de amostras da região do Pantanal brasileiro. Embora filogeneticamente agrupado aos *Sputnikvirus*, o virófago Pantanal apresenta características genômicas distintas, incluindo três ORFans e regiões genômicas exclusivas. A baixa identidade de nucleotídeos e de aminoácidos quando comparado com outros Sputnikvírus, somada à filogenia de proteínas conservadas, confirmaram a divergência do virófago e ajudaram a sustentar sua proposta como uma nova espécie dentro do gênero *Sputnikvirus*. Juntos, os resultados revelam a importância dos esforços para prospecção e caracterização de novos vírus gigantes e de seus elementos associados, principalmente em ecossistemas ainda pouco explorados para tal feito, como o litoral e o Pantanal brasileiros. A inclusão desses novos isolados em análises comparativas permite aprofundar o entendimento das relações evolutivas entre vírus conhecidos e destaca a necessidade contínua de revisão e atualização da taxonomia viral, consolidando o papel desses estudos na ampliação do conhecimento básico sobre os vírus do reino *Bamfordvirae*.

Palavras-chave: Marseillevírus cajuinensis, virófago, Pantanal, *Bamfordvirae*, taxonomia

ABSTRACT

Giant amoeba viruses represent one of the most intriguing areas of modern virology, not only because of their complex morphology and genomes, but also because of their evolutionary and taxonomic relevance. These viruses are taxonomically classified in the *Bamfordvirae* kingdom, which is represented by different viral families, such as *Marseilleviridae*, *Mimiviridae* and *Sputniviroviridae*. Here, we describe new isolates of bamfordviruses obtained from Brazilian environmental samples. In the first part of the work, we describe Marseillevirus cajuinensis, the first isolation of a *Marseilleviridae* family virus from saltwater. This virus has a genome with approximately 380 thousand base pairs, containing 515 ORFs, including 38 ORFans, and a transfer RNA (tRNA) which was a gene, util then, rarely described among Marseilleviruses. Phylogenetic analyses showed that M. cajuinensis is a divergent member of lineage A and revealed that Marseilleviruses can be organized into three major groups instead of the classical five, suggesting the need for taxonomic reorganization. Based on these data, a proposal was submitted to the International Committee on Viral Taxonomy (ICTV) for the inclusion of new taxa in the *Marseilleviridae* family, including the creation of new species and genera to expand their taxonomical diversity. The second part of the work addresses the isolation of a tripartite viral system composed of a moumouvirus, a transpoviron and a new virophage, called Pantanal virophage, from samples collected in Pantanal biome region. Although phylogenetically grouped with Sputnikviruses, the Pantanal virophage presents distinct genomic characteristics, including three ORFans and exclusive genomic regions. The low nucleotide and amino acid identity in comparison with other members of the group added to the phylogeny of conserved proteins confirmed their divergence and supported their proposal as a new species within the genus *Sputnikvirus*. Altogether, the results reveal the importance of efforts to prospect and characterize new giant viruses and their associated elements, especially from ecosystems that have not yet been explored for this purpose, such as the Brazilian coast and Pantanal. The inclusion of these new isolates in comparative analyses allows for a deeper understanding of the evolutionary relationships between known viruses and highlights the ongoing need for review and updating of viral taxonomy, consolidating the role of these studies in expanding basic knowledge about viruses in the *Bamfordvirae* kingdom.

Keywords: Marseillevírus cajuinensis, virophage, Pantanal, *Bamfordvirae*, taxonomy

LISTA DE ILUSTRAÇÕES

SUMÁRIO

1 INTRODUÇÃO

1.1 O reino *Bamfordvirae* e seus filos: *Nucleocytoviricota* e *Preplasmiviricota*.

A virosfera é o conjunto dos vírus que habitam o nosso planeta. Ela é vasta e muito diversificada, além de ser composta pelas entidades biológicas mais abundantes do globo. Além de abundantes e ubíquos, os vírus também apresentam uma grande diversidade de formas, tamanhos e genomas. Essa diversidade é tão grande, que ao analisar o conjunto de todos os vírus conhecidos, identifica-se grandes grupos virais que não compartilham uma ancestralidade comum, evidenciando o caráter polifilético dos vírus e dificultando a sua organização taxonômica (KOONIN et al., 2020). Por causa disso, ao longo da história (principalmente na era pré-genômica), os vírus conhecidos foram sendo classificados de acordo com características que não implicam necessariamente em ancestralidade comum e filogenia. Um exemplo é a classificação de Baltimore, que considera a estrutura do ácido nucleico viral e sua estratégia de replicação do genoma para organizá-los (BALTIMORE, 1971). Outro exemplo é a classificação de determinados grupos de vírus de acordo com características fenotípicas, como a forma de disseminação (por exemplo, os “arbovírus”) e a presença de envelope ao redor do capsídeo.

O Comitê Internacional em Taxonomia de Vírus (ICTV - *International Committee on Taxonomy of Viruses*), estabelecido em 1966, tem direcionado esforços para organizar os vírus taxonomicamente, mesmo com toda sua diversidade de grupos. Dessa forma, embora existam outras formas de classificação viral válidas, como a classificação de Baltimore, apenas uma taxonomia viral oficial é considerada, sendo determinada e validada pelos membros do ICTV (SIDDELL et al., 2023). Anteriormente, os vírus eram organizados taxonomicamente dentro de táxons que alcançavam somente até o nível de ordem. No entanto, desde 2016, o comitê executivo do ICTV vinha discutindo sobre a necessidade de reorganização hierárquica dos vírus, permitindo táxons em maior nível hierárquico. Dessa forma, foi feita uma proposta de organização taxonômica global dos vírus incluindo novos táxons, ficando os vírus classificados em: “domínio” (*realm*), reino (*kingdom*), filo, classe, ordem, família, gênero e espécie (ordem da maior hierarquia para a menor hierarquia) (GORBALENYA et al., 2020). É valido mencionar que ficou definido que as atribuições taxonômicas devem ser correspondentes à filogenia dos grupos de vírus, ou seja, os táxons devem ser monofiléticos (GORBALENYA et al., 2020).

Apesar da falta de um ancestral comum entre todos os vírus conhecidos, existem genes amplamente conservados dentro de determinados grupos de vírus, codificando proteínas envolvidas em funções centrais na replicação e morfogênese de partículas. Um exemplo desse tipo de gene é aquele que codifica proteínas principais de capsídeo do tipo "double-jelly-roll" (DJR), amplamente distribuído dentre os vírus de DNA dupla fita (KOONIN et al., 2020). Dessa forma, atualmente, existem 7 "domínios" (*realms*) de vírus (ICTV, 2024). Dentre eles, está o "domínio" *Varidnaviria*, formado por vírus de DNA dupla-fita (DNAdf). O táxon *Varidnaviria* é composto por dois reinos: *Bamfordvirae* e *Abadenavirae* (ICTV, 2024). O reino *Bamfordvirae,* formado com base na presença da proteína de capsídeo com a estrutura de DJR em seus componentes, será o principal táxon estudado neste trabalho. Esse reino é composto pelos filos *Nucleocytoviricota* e *Preplasmiviricota.*

O filo *Nucleocytoviricota* inclui famílias virais que, anteriormente, formavam um grupo não-taxonômico chamado de Vírus Grandes Nucleocitoplasmáticos de DNA (NCLDV – *Nucleo-Cytoplasmic Large DNA Viruses*). Esse é um grupo que compartilha um conjunto de genes ortólogos, que inclui a DNA polimerase tipo B/delta catalítica (IYER et al., 2006; IYER; ARAVIND; KOONIN, 2001). O filo *Nucleocytoviricota* é, até então, dividido em 3 classes: *Megaviricetes* (inclui as famílias *Phycodnaviridae*, *Mimiviridae*, *Ascoviridae*, *Iridoviridae* e *Marseilleviridae*); *Mriyaviricetes* (inclui a família *Yaraviridae) e Pokkesviricetes* (inclui as famílias *Asfarviridae* e *Poxviridae)* (ICTV, 2024). Já o filo *Preplasmiviricota,* inclui*,* dentre outros vírus, os membros da família *Adenoviridae,* os polintonvírus*,* a família *Tectiviridae e* os membros da antiga família Lavidaviridae, responsável por agrupar virófagos, como o mavírus e os sputnikvírus (ICTV, 2024).

1.2 A família *Marseilleviridae*

1.2.1 Diversidade de isolados pertencentes à família

A família *Marseilleviridae,* cujos membros são chamados de Marseillevírus (MsV)*,* foi o segundo grupo de vírus de amebas isolados. A descoberta dos Marseillevírus ocorreu graças ao interesse que a descrição do primeiro vírus gigante de ameba, o mimivírus, causou na comunidade científica, em 2003 (LA SCOLA et al., 2003). Os mimivírus, incluídos na família *Mimiviridae*, chamaram a atenção pelo tamanho da sua partícula e de seu genoma, além dos genes que são capazes de codificar, tendo características absolutamente diferentes do que se conhecia sobre os vírus até então (LA SCOLA et al., 2003). Diversos outros membros da família *Mimiviridae* foram sendo descritos posteriormente, formando o que atualmente é uma família composta por 3 diferentes subfamílias. Dentre essas subfamílias, está a *Megamimivirinae,* que inclui 5 gêneros, dentre os quais se agrupam os primeiros isolados de mimívírus. O gênero *Mimivirus* abriga vírus filogeneticamente relacionados ao primeiro isolado (Acanthamoeba polyphaga mimivirus ou APMV), anteriormente chamados de linhagem A de mimivírus. Já o gênero *Moumouvirus* abriga vírus relacionados à chamada linhagem B de mimivírus. O gênero *Megavirus* é responsável por agrupar os isolados filogeneticamente relacionados à chamada linhagem C de mimivírus (ICTV, 2024).

Graças a descoberta dos primeiros mimivírus, esforços para o isolamento de novos vírus de ameba surgiram, levando ao isolamento do primeiro Marseillevírus, o qual se revelou não só como um outro vírus de ameba, mas também como um componente de uma nova família viral. O primeiro isolado de Marseillevírus, chamado de Marseillevirus T19 ou Marseillevirus marseillevirus foi descoberto em 2007, na cidade de Marseille (França), a partir de amostras de água de uma torre de refrigeração de Paris (BOYER et al., 2009). Durante o isolamento do primeiro Marseillevírus foram observadas grandes partículas icosaédricas, com cerca de 250 nanômetros (nm) (Figura 1). O capsídeo é circundado por pequenas fibrilas (~12 nm) e dentro dele existe um nucleocapsídeo que é cercado por uma membrana interna (BOYER et al., 2009)(Figura 1 B, setas vermelhas).

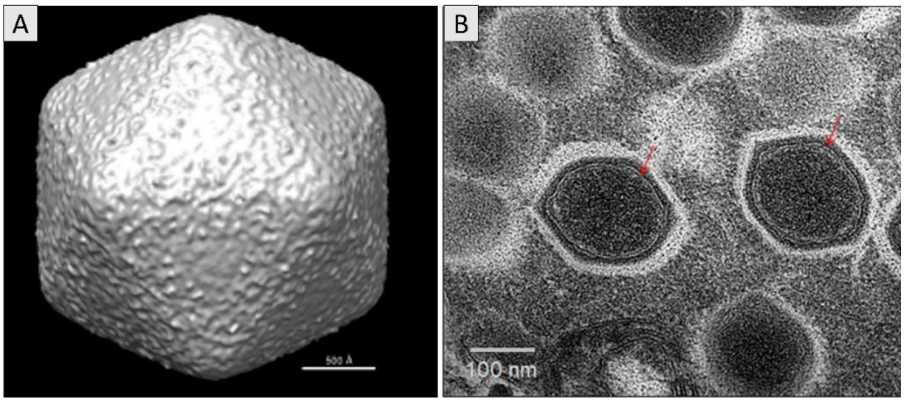
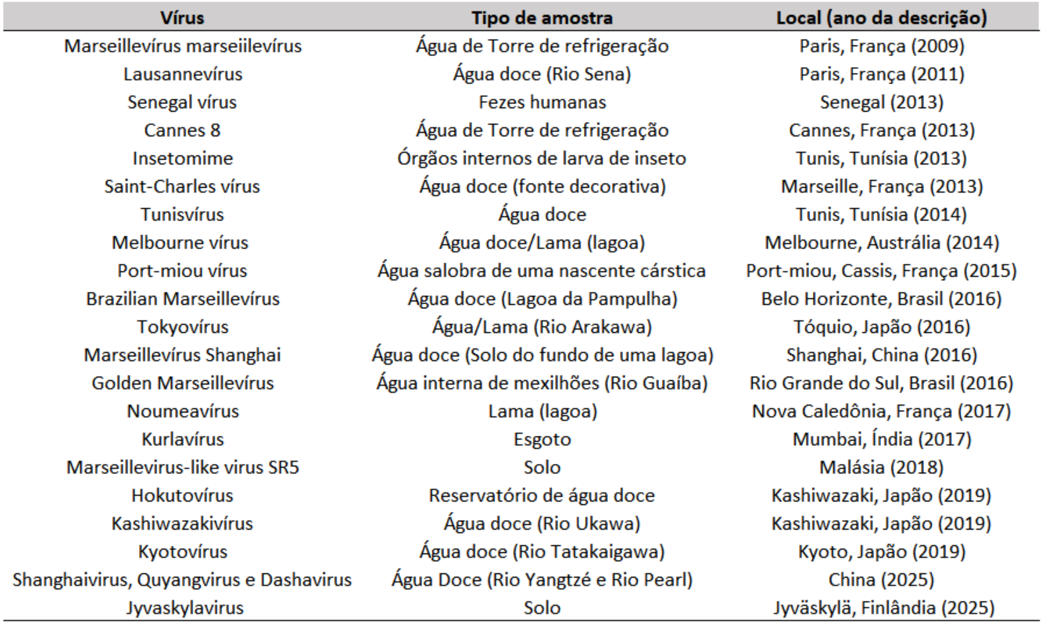


Figura 1 –Partícula viral dos Marseillevírus. (A) Reconstrução em 3D, por crio-microscopia eletrônica, da partícula de Marseillevírus (MsV), mostrando sua morfologia icosaédrica. (B) Imagem de microscopia eletrônica de transmissão das partículas de MsV, mostrando a membrana interna que circunda o nucleocapsídeo (setas vermelhas). Imagens adaptadas de: (ARANTES et al., 2016; BOYER et al., 2009)

O isolamento do Marseillevírus T19 foi o primeiro de diversos outros isolados similares descritos até então. Em 2011, foi descrito o Lausannevírus, segundo membro da família *Marseilleviridae,* obtido de amostras de água do Rio Sena, em Paris (THOMAS et al., 2011). Nos anos seguintes, a busca por novos MsV, principalmente em amostras de água, levou à descrição do Cannes8 virus, isolado da água de uma torre de refrigeração em Cannes, na França (AHERFI et al., 2013). Diversos outros isolados obtidos de amostras relacionadas à água doce foram descritos posteriormente (Tabela 1). Dentre eles, destacam-se o Tunisvírus, o Melbourne vírus, o

Noumeavírus e o Tokyovírus (AHERFI et al., 2014; DOUTRE et al., 2014; FABRE et al., 2017; TAKEMURA, 2016)

Tabela 1 – Lista dos principais Marseillevírus isolados e as respectivas características do seu isolamento.



Vírus	Tipo de amostra	Local (ano da descrição)
Marseillevírus marseiilevírus	Água de Torre de refrigeração	Paris, França (2009)
Lausannevírus	Água doce (Rio Sena)	Paris, França (2011)
Senegal vírus	Fezes humanas	Senegal (2013)
Cannes 8	Água de Torre de refrigeração	Cannes, França (2013)
Insetomime	Órgãos internos de larva de inseto	Tunis, Tunísia (2013)
Saint-Charles vírus	Água doce (fonte decorativa)	Marseille, França (2013)
Tunisvírus	Água doce	Tunis, Tunísia (2014)
Melbourne vírus	Água doce/Lama (lagoa)	Melbourne, Austrália (2014)
Port-miou vírus	Água salobra de uma nascente cárstica	Port-miou, Cassis, França (2015)
Brazilian Marseillevírus	Água doce (Lagoa da Pampulha)	Belo Horizonte, Brasil (2016)
Tokyovírus	Água/Lama (Rio Arakawa)	Tóquio, Japão (2016)
Marseillevírus Shanghai	Água doce (Solo do fundo de uma lagoa)	Shanghai, China (2016)
Golden Marseillevírus	Água interna de mexilhões (Rio Guaíba)	Rio Grande do Sul, Brasil (2016)
Noumeavírus	Lama (lagoa)	Nova Caledônia, França (2017)
Kurlavírus	Esgoto	Mumbai, Índia (2017)
Marseillevirus-like virus SR5	Solo	Malásia (2018)
Hokutovírus	Reservatório de água doce	Kashiwazaki, Japão (2019)
Kashiwazakivírus	Água doce (Rio Ukawa)	Kashiwazaki, Japão (2019)
Kyotovírus	Água doce (Rio Tatakaigawa)	Kyoto, Japão (2019)
Shanghaivirus, Quyangvirus e Dashavirus	Água Doce (Rio Yangtzé e Rio Pearl)	China (2025)
Jyvaskylavirus	Solo	Jyväskylä, Finlândia (2025)

Nota: Dados analisados até junho de 2025.

Posteriormente, outros tipos de amostra começaram a ser testadas para o isolamento de MsV. Em 2015, um isolado de MsV foi obtido a partir da água salobra de uma nascente cárstica e chamado de Port-miou vírus (DOUTRE et al., 2015). Entretanto, alguns anos antes, dois tipos de amostras bem diferentes resultaram no isolamento de dois Marseillevírus. O primeiro é o Senegal vírus, isolado de amostras de fezes humanas de um homem senegalês (COLSON et al., 2013). Já o segundo é o insectomime vírus, obtido a partir dos órgãos internos e trato digestivo de uma larva de inseto da ordem Diptera (BOUGHALMI et al., 2013).

No Brasil, o primeiro Marseillevírus isolado foi o Brazilian Marseillevírus, a partir de amostras de água da Lagoa da Pampulha, Belo Horizonte. Esse vírus foi classificado como um membro fundador de uma nova linhagem de MsV, devido às análises genômicas e filogenéticas realizadas (DORNAS et al., 2016). Pouco tempo depois, outro trabalho brasileiro descreveu o isolamento do Golden Marseillevírus, obtido de amostras de água de dentro de um mexilhão (*Limnoperna fortunei*) do Rio Guaíba, no Rio Grande do Sul. Isso evidenciou ainda mais a diversidade de amostras nas quais os MsV podem ser detectados. Além disso, o Golden Marseillevírus apresentou várias divergências genômicas e filogenéticas, sendo responsável, também, por fundar uma nova linhagem de MsV (DOS SANTOS et al., 2016). Um terceiro isolado de Marseillevírus brasileiro foi descrito em 2018 em amostras de esgoto, em meio a um painel de 68 isolados de vírus gigantes de diferentes grupos (ANDRADE et al., 2018). Dessa forma, os trabalhos que descrevem Marseillevírus brasileiros evidenciam o quanto esses isolados foram capazes de expandir o conhecimento acerca da diversidade, evolução e distribuição de Marseilleviridae. Por causa disso, os trabalhos de prospecção utilizando diferentes amostras mostram-se muito relevantes para garantir que novos vírus sejam, cada vez mais, isolados.

Além de diversos isolados já descritos, existem também estudos de metagenômica que descrevem sequências divergentes de MsV em sedimentos de uma fonte hidrotermal no Oceano Atlântico, localizada em um campo de fontes hidrotermais ativas chamado Castelo de Loki (*Loki's Castle*). Nesse estudo, 5 sequências foram descritas e chamadas de LCMAC 101, LCMAC 102, LCMAC 103, LCMAC 201 e LCMAC 202 (BÄCKSTRÖM et al., 2019). Considerando que a maioria dos Marseillevírus isolados estão relacionados a amostras de água doce (Tabela 1), essas descobertas de metagenômica em ambientes diferentes, como regiões profundas do oceano, expandem o conhecimento sobre os diferentes locais onde MsV podem ser isolados.

1.2.2 Características genômicas dos Marseillevírus

Os Marseillevírus apresentam um genoma composto por uma molécula de DNA de dupla fita de topologia circular. O tamanho das sequências varia em torno de 348 a 404 kpb e o conteúdo de guanina-citosina (G-C) gira em torno de 42,9 a 44,8%. Já o número de genes preditos que foram descritos vai de 386 a 491 (SAHMI-BOUNSIAR et al., 2021). O grupo de sequências detectadas por metagenômica, a partir das amostras da região de *Loki's Castle*, apresenta, curiosamente, tamanhos de genoma, número de genes preditos e conteúdo de G-C maiores do que dos outros Marseillevírus. Por exemplo, o tamanho das sequências pode chegar a 763 kpb, o número de genes preditos a 793, enquanto o conteúdo de G-C alcança 62,3 % (BÄCKSTRÖM et al., 2019; SAHMI-BOUNSIAR et al., 2021)

O genoma dos Marseillevírus é composto por uma grande quantidade de genes que codificam proteínas não caracterizadas, das quais não se sabe a função, assim como acontece com os outros vírus gigantes. Dentre as proteínas com funções preditas, destacam-se as que apresentam homólogos com outros nucleocitovírus, como a DNA polimerase. Além disso, os MsV também codificam um conjunto de genes cujos produtos são proteínas homólogas a histonas, que estão envolvidas na proteção do DNA viral (BRYSON et al., 2022; LIU et al., 2021). Ao contrário do que acontece com outros vírus gigantes, como os mimivírus, os MsV não apresentam uma diversidade de genes envolvidos na tradução de proteínas. Embora seja descrita a codificação de alguns fatores de tradução, não é comum encontrar a descrição de genes que codificam RNAs transportadores e aminoacil-tRNA sintetases (ABRAHÃO et al., 2017; SAHMI-BOUNSIAR et al., 2021). No caso de RNAs transportadores, eles são descritos apenas em Tokyovírus e nas sequências de metagenômica de *Loki's Castle* (BÄCKSTRÖM et al., 2019; TAKEMURA, 2016). Em contrapartida ao que é descrito para a maioria dos MsV, no caso das sequências de *Loki's Castle* são descritos até 26 genes que codificam RNAs transportadores e 15 que codificam aminoacil-tRNA sintetases, valor maior até mesmo do que o encontrado em mimivírus. Isso mostra a divergência dessas sequências em relação aos outros MsV (BÄCKSTRÖM et al., 2019).

Outra característica que se destaca no genoma dos MsV é sua alta taxa de mosaicismo, ou seja, presença de um repertório de genes com várias origens. Além de compartilhar homólogos com os outros NCLDV, uma parte dos genes do Marseillevírus marseillevírus existe também em bactérias e fagos. Outra parte dos genes tem homólogos em amebas e em outros eucariotos (BOYER et al., 2009). A existência de genes com origens múltiplas pode estar ligada com o estilo de vida dos hospedeiros dos MsV, as amebas, uma vez que elas podem ser infectadas por diferentes tipos de organismos ao mesmo tempo. A coexistência de vírus gigantes e outros organismos dentro da mesma ameba, torna essa célula um ambiente favorável para transferência gênica horizontal (TGH), transformando-a no chamado "caldeirão" (*melting pot*) de genes (BOYER et al., 2009).

1.2.2 Taxonomia atual e filogenia da família Marseilleviridae

Como citado anteriormente, os Marseillevírus fazem parte de uma família viral própria, chamada de *Marseilleviridae.* Atualmente, o ICTV classifica os MsV hierarquicamente da seguinte maneira:

Domínio (*realm*): *Varidnaviria*
Reino (*kingdom*): *Bamfordvirae*
Filo: *Nucleocytoviricota*
Classe: *Megaviricetes*
Ordem: *Pimascovirales*
Família: *Marseilleviridae*

Figura 2 – Classificação hierárquica família Marseilleviridae realizada pelo Comitê Internacional em Taxonomia de Vírus (ICTV). Fonte: (ICTV, 2024)

Utilizando análises filogenéticas, geralmente baseadas na proteína DNA polimerase ou na proteína principal do capsídeo (MCP), os Marseillevírus tm sido divididos com base na existência de cinco linhagens filogenéticas, chamadas de A, B, C, D e E (Figura 3). A linhagem A é representada pelo primeiro isolado, o Marseillevírus marseillevírus e vírus filogeneticamente relacionados. Já a linhagem B é representada pelo Lausannevirus, segundo MsV isolado, e outros vírus relacionados. As linhagens A e B, são as que possuem o maior número de isolados descritos. A terceira linhagem (C) é formada por Tunisvírus e Insectomime virus. Já as linhagens D e E, são compostas pelos isolados brasileiros, Brazilian Marseillevírus e Golden Marseillevírus, respectivamente. As sequências de MsV detectadas por metagenômica na região de *Loki's Castle*, quando incluídas nas árvores com os outros MsV, formam dois grupos externos às cinco linhagens (SAHMI-BOUNSIAR et al., 2021).

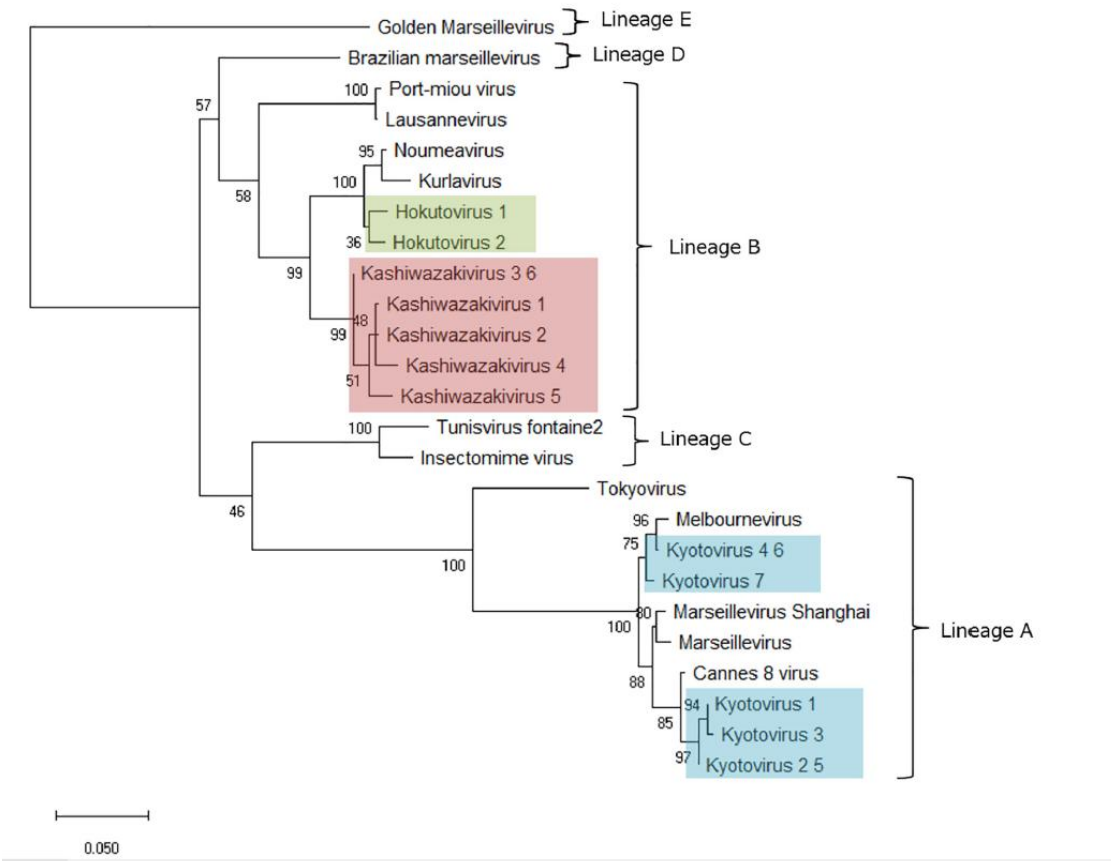


Figura 3 – Árvore filogenética baseada no gene para a proteína principal do capsídeo, indicando as 5 linhagens filogenéticas de Marseillevírus. Árvore sem raiz, construída com o método de máxima verossimilhança. Fonte: (AOKI et al., 2019). Obs.: os vírus destacados na árvore correspondem a sequências obtidas na ocasião do trabalho que publicou essa árvore (AOKI et al., 2019).

Ao considerar a filogenia da família *Marseilleviridae* e compará-la com a taxonomia do grupo proposta oficialmente pelo ICTV, é possível perceber a existência de algumas lacunas, uma vez que estão incluídos apenas membros de três das cinco linhagens aqui descritas. A família é composta por dois diferentes gêneros. O primeiro gênero, *Marseillevirus*, é composto por duas espécies: *Marseillevirus massiliense* and *Marseillevirus senegalense,* ambas representando membros da linhagem A. Já o segundo gênero, *Losannavirus*, também é formado por duas espécies: *Losannavirus lausannense* (vírus relacionados filogeneticamente à linhagem B) e *Losannavirus tunisense* (vírus filogeneticamente relacionados à linhagem C) (ICTV, 2024).

1.3 Os virófagos: família *Sputniviroviridae.*

1.3.1 Descoberta e diversidade dos virófagos

Os virófagos são definidos como vírus de DNAfd, com tamanho de partícula variando entre 50 e 74 nm. Eles são comumente chamados de "vírus que infectam vírus", pois são caracterizados por dependerem não apenas da célula hospedeira, mas também da presença de um vírus gigante, do filo *Nucleocytoviricota*, para completar seu ciclo (Figura 4) (KRUPOVIC; KUHN; FISCHER, 2016; MOUGARI et al., 2019b). Eles foram descritos pela primeira vez em associação com um vírus da família *Mimiviridae* e do gênero *Mimivirus*, chamado de mamavírus, que foi obtido a partir de amostras de água coletadas de uma torre de resfriamento localizada em Paris, França. (LA SCOLA et al., 2008). Os virófagos despertaram o interesse científico por apresentarem uma interação incomum tanto com o mamavírus, quanto com a célula hospedeira.

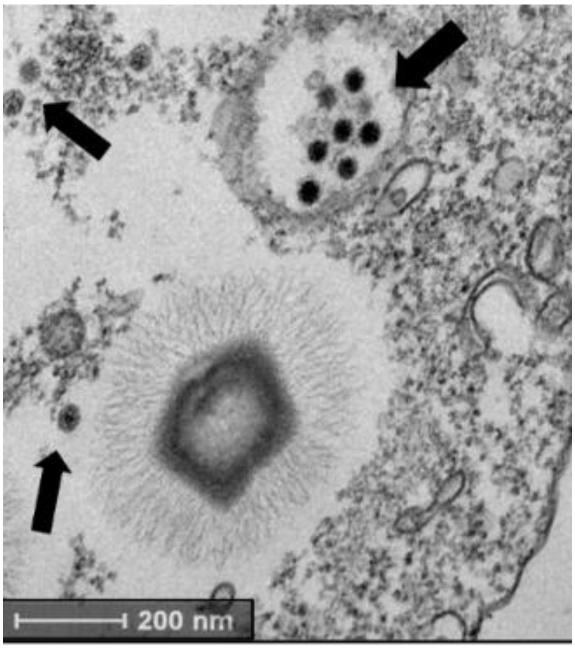


Figura 4 – Partículas de virófago (setas pretas) associadas a um mimivírus dentro de uma ameba hospedeira. A figura evidencia a diferença de tamanho entre as partículas dos tipos de vírus. Fonte: Adaptado de AZEVEDO et al., 2022

Foi observado que os virófagos se multiplicavam dentro da fábrica viral dos mimivírus e que não eram capazes de se multiplicar sozinhos na ameba, ou seja, sem uma coinfecção com o vírus gigante. Em vez de recorrerem diretamente à maquinaria transcricional da célula hospedeira, os virófagos se aproveitam da fábrica viral e dos recursos do vírus gigante para ativar a expressão de seus próprios genes. Essa interferência pode, em algumas situações, comprometer a produção de novos vírus gigantes. Além disso, os estudos também evidenciaram que é possível que os virófagos causem o aumento significativo da formação de partículas de mamavírus defeituosas, o que causa um impacto para a produtividade do vírus gigante (LA SCOLA et al., 2008). Para definir a natureza dessas pequenas partículas recém-descobertas, foi proposto o termo virófago, fazendo referência aos bacteriófagos, que são os vírus que infectam bactérias. O primeiro virófago descrito foi chamado de sputnik (LA SCOLA *et al.*, 2008).

A descoberta do virófago sputnik abriu portas para a posterior descrição de outros isolados de virófagos. Em 2011, o virófago mavírus, foi obtido a partir de águas costeiras do Texas, nos Estados Unidos, associado ao Cafeteria roenbergensis vírus (CroV) que é membro da família *Mimiviridae*, porém de uma subfamília diferente (*Allimimivirinae*) daquela que agrupa os mimivírus “clássicos” (*Megamimivirinae*) (FISCHER; SUTTLE, 2011). Assim como o CroV, os mavírus compõem uma linhagem filogenética mais divergente em relação à linhagem dos virófagos sputnik. Algum tempo depois, outros isolados da linhagem de virófagos sputnik, chamados de sputnik 2 e sputnik 3, foram descobertos na França, a partir de amostras de fluido de lentes de contato e de amostras de solo, respectivamente (DESNUES et al., 2012; GAIA et al., 2013). Posteriormente, outra linhagem de virófagos que infectam *Megamimivirinae* foram descritos e chamados de zamilon, zamilon 2 e zamilon vitis (BEKLIZ et al., 2015; GAIA et al., 2014; JEUDY et al., 2020). Os virófagos zamilon são intrigantes, pois apresentam uma especificidade de hospedeiros. Eles são capazes de infectar os membros do gênero *Megavirus* (antiga linhagem C de mimivírus), mas não são capazes de se multiplicar em coinfecção com os membros do gênero *Mimivirus* (antiga linhagem A de mimivírus). Pesquisas desenvolvidas depois dessa descoberta propuseram a existência de um mecanismo de defesa contra virófagos nos mimivírus gênero *Mimivirus* (LEVASSEUR et al., 2016).

No Brasil, o primeiro virófago, chamado de virófago Rio Negro, foi isolado em associação com o sambavírus, a partir de amostras de água do Rio Negro, na Amazônia (CAMPOS et al., 2014). O segundo isolado de virófago brasileiro, chamado de virófago Guarani, foi obtido a partir de amostras de água da Lagoa da Pampulha, em Belo Horizonte (MOUGARI et al., 2019a). Por fim, foi descrito o Sputnik argentum, que é genotipicamente similar aos outros virófagos sputnik, mas que foi descoberto em uma relação tripla envolvendo o virófago, um mimivirus e um elemento genético móvel chamado de transpoviron, presente no genoma do vírus gigante (AZEVEDO et al., 2022).

É importante ressaltar que a existência de vários estudos com abordagem metagenômica, visando detectar sequências de novos virófagos em diferentes ambientes, tem sido bastante significativa ao contribuir para a expansão da diversidade conhecida de virófagos. São exemplos de sequências completas de virófagos obtidas por metagenômica o "Organic Lake Virophage" (OLV), obtida a partir de amostras de um lago orgânico hipersalino da Antártida e os Yellowstone Lake virophages" (YSLVs), cujas sequências foram detectadas de amostras do lago Yellowstone, nos Estados Unidos (YAU et al., 2011; ZHOU et al., 2013, 2015). Além de genomas completos, algumas sequências incompletas ou genes específicos de virófagos também vêm sendo descritas e detectadas por metagenômica. Essa abordagem metagenômica tem permitido a detecção de virófagos em ambientes muito diversos e, até mesmo, extremos (BELLAS; ANESIO; BARKER, 2015; YUTIN; KAPITONOV; KOONIN, 2015; ZABLOCKI et al., 2014). Em conjunto, esses estudos evidenciam que os virófagos podem estar presentes em diversos ambientes ao redor do mundo e que novos estudos visando o isolamento das partículas nesses ambientes podem ser bem-sucedidos.

1.3.2 Genoma, filogenia e taxonomia dos virófagos sputnik.

Os genomas dos isolados de virófago sputnik apresentam alta similaridade um com o outro, sendo formados por uma molécula de DNA de dupla fita circular. Os genomas apresentam um número de pares de bases que, normalmente, variam em torno de 17 e 18 mil, com uma alta porcentagem de adenina e timina (A+T). Já o número de genes codificadores de proteínas (ORFs) pelos virófagos sputnik varia entre 20 e 22 (MOUGARI et al., 2019b). Os produtos desses genes incluem algumas proteínas conservadas entre os virófagos. As proteínas conservadas estão associadas à morfogênese da partícula dos virófagos e, dentre elas, se inclui a proteína principal do capsídeo (MCP), a proteína secundária do capsídeo (mCP), uma ATPase da superfamília FtsK-HerA relacionada ao empacotamento do DNA para dentro da partícula e uma protease responsável por clivar a MCP durante a maturação do capsídeo (BORN et al., 2018; LA SCOLA et al., 2008; YUTIN; RAOULT; KOONIN, 2013). O par de proteínas de capsídeo MCP-mCP é bastante conservado, e está presente em todos os virófagos, o que evidencia a sua importância para esses organismos (FISCHER, 2021). Além disso, não existem homólogos detectáveis para essas proteínas em outros organismos, além dos virófagos, o que também é indicativo de uma ancestralidade em comum (YUTIN; RAOULT; KOONIN, 2013). É interessante destacar que, assim como em muitos vírus gigantes de ameba, uma parte dos genes dos virófagos sputnik também não apresentaram homólogos nos bancos de dados analisados até então, o que caracteriza esses genes como ORFans

(LA SCOLA et al., 2008). As proteínas conservadas mencionadas acima, em especial a MCP, são frequentemente utilizadas em análises filogenéticas do grupo (Figura 5).

A taxonomia dos virófagos foi recentemente atualizada e aprovada pelo ICTV. Anteriormente, foi proposto que eles fossem classificados em uma única família viral chamada de *Lavidaviridae* (KRUPOVIC; KUHN; FISCHER, 2016), que foi incluída no "domínio" *Varidnaviria,* junto aos outros vírus de DNAdf. Como a proteína principal do capsídeo dos virófagos apresenta a estrutura DJR, o grupo também foi incluído no reino *Bamfordvirae*, formando o filo *Preplasmiviricota,* ao lado dos *Nucleocytoviricota* (KOONIN et al., 2020). Entretanto, apesar de terem sido mantidos no filo *Preplasmiviricota,* a organização interna desse filo foi alterada e, até então, os virófagos isolados (linhagem dos mavírus, linhagem dos sputnik e linhagem dos zamilon) estão incluídos no subfilo *Polisuviricotina* e na classe *Virophaviricetes.*

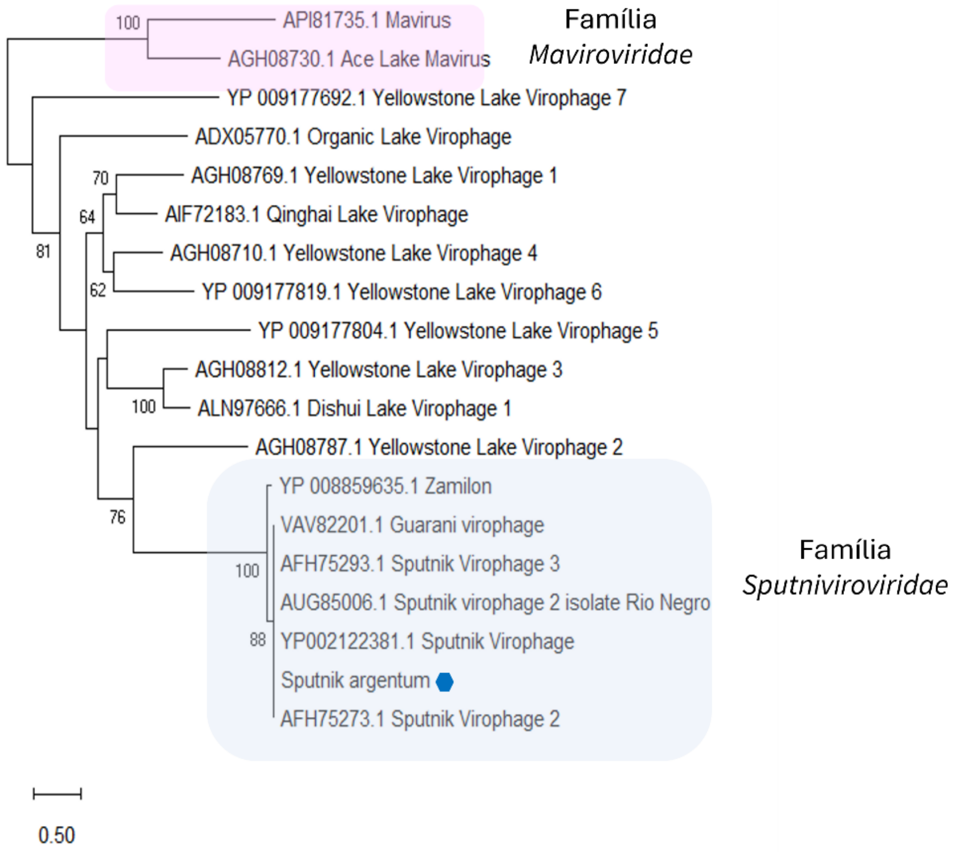


Figura 5 – Árvore filogenética de máxima verossimilhança da proteína principal do capsídeo (MCP) dos virófagos isolados e algumas sequências de metagenômica. Fonte: Adaptado de Azevedo et. al, 2022

A classe *Virophaviricetes* é composta por 4 ordens, dentre elas as ordens *Lavidavirales* e *Mividavirales,* que abrigam os virófagos isolados. A ordem *Lavidavirales* inclui a família *Maviroviridae* e seu único gênero *Mavirus,* contendo a espécie *Mavirus cafeteriae,* correspondente a linhagem dos mavírus. Já a ordem *Mividavirales* é composta pela família *Sputniviroviridae* e seu gênero *Sputnikvirus*, que é formado por duas espécies: *Sputnikvirus mimiviri* (inclui os isolados de virófago sputnik), e *Sputnikvirus zamilonense* (inclui os isolados da linhagem dos zamilon). As outras ordens da classe incluem sequências diversas de virófagos detectados por metagenômica.

Essa atualização compreensiva da taxonomia dos virófagos feita nos últimos anos somada às contantes atualizações na filogenia dos MsV e às lacunas existentes na taxonomia desse grupo demonstram o quanto o conhecimento sobre a evolução desses vírus tem se expandido, graças aos esforços para novos isolamentos e para estudos relacionados. Além disso, reflete-se também a necessidade de uma proposta taxonômica que reflita toda a diversidade conhecida das famílias *Marseilleviridae* e *Sputniviroviridae,* o que contribui para o melhor entendimento da taxonomia e evolução do reino *Bamfordvirae.*

2 JUSTIFICATIVA

Os vírus são as entidades mais abundantes do planeta Terra e apresentam uma ampla distribuição e diversidade. Os vírus gigantes atraíram a atenção da comunidade científica nas últimas décadas, devido a sua singularidade. A descoberta e a caracterização de novos vírus por meio da prospecção de diferentes amostras ambientais, incluindo as de origem brasileira, têm mostrado o seu potencial no isolamento de diferentes grupos de vírus pertencentes ao reino *Barmfordvirae*, incluindo os Marseillevírus e os mimivírus, componentes do filo *Nucleocytoviricota,* e os virófagos, componentes do filo *Preplasmiviricota*.

Os Marseillevírus são vírus de ameba do filo *Nucleocytoviricota* cujos isolados conhecidos são classicamente divididos em 5 linhagens filogenéticas (A, B, C, D e E). O surgimento de novas linhagens foi acontecendo graças aos esforços para isolamento e detecção de novos vírus, incluindo dois isolados brasileiros que são, ainda, os únicos membros pertencentes às linhagens D e E. Os mimivírus e os virófagos passaram pelo mesmo processo de expansão da diversidade de linhagens e do conhecimento sobre sua biologia e evolução, graças ao isolamento e caraterização de novos vírus, além da detecção de sequências por metagenômica.

O potencial dos estudos de prospecção, isolamento e caracterização de novos vírus de amebas é reforçado neste trabalho, pois nele descreveremos novos isolados pertencentes às famílias *Marseilleviridae*, *Mimiviridae* e S*putniviroviridae* e suas respectivas características genômicas e filogenéticas. A caracterização desses novos isolados trouxe novos dados e observações a respeito da filogenia de *Marseilleviridae* e de *Sputniviroviridae*, como a identificação de lacunas a serem preenchidas na taxonomia oficial e a observação de uma necessidade de se estabelecer parâmetros para a classificação dos vírus, além da necessidade de manter uma constante atualização da taxonomia viral. A taxonomia é uma ciência dinâmica e que vai se desenvolvendo na medida em que novos métodos e dados vão surgindo ou sendo melhorados. Com base nisso, esse trabalho se faz importante, pois evidencia que o isolamento de novos vírus de ameba e sua caracterização podem contribuir enriquecendo o conhecimento sobre os aspectos biológicos, genômicos, evolutivos dos vírus do reino *Barfordvirae*, bem como contribuindo com a dinâmica da taxonomia desses vírus.

3 OBJETIVOS

3.1 Objetivo geral

Descrever o isolamento e a caracterização genômica e filogenética de novos isolados de bamfordvírus brasileiros, das famílias *Marseilleviridae* e *Sputniviroviridae,* bem como classificar esses isolados, propondo mudanças na taxonomia oficial dos grupos.

3.2 Objetivo específicos

- Analisar, de maneira geral, as características morfológicas das partículas dos vírus isolados.
- Caracterizar os genomas dos vírus isolados, incluindo a identificação de genes codificadores de proteínas (ORFs).
- Caracterizar as proteínas dos vírus isolados, identificando possíveis funções biológicas.
- Investigar as relações filogenéticas de diferentes proteínas conservadas dentro do grupos aos quais pertencem os vírus isolados.
- Comparar a sintenia entre diferentes genomas virais dos grupos estudados.
- Identificar genes relacionados a RNAs transportadores em diferentes linhagens de Marseillevírus.
- Analisar o pangenoma e o coregenoma dos Marseillevírus.
- Comparar a presença de *Clusters of Orthologous Groups* (COGs ou grupos de proteínas ortólogas) exclusivas e compartilhados entre as diferentes linhagens de Marseillevírus descritas.
- Explorar as relações fenéticas entre diferentes linhagens de Marseillevírus com base na presença e ausência de COGs.
- Avaliar a similaridade genômica e proteica entre vírus isolados e sequências relacionadas, visando estabelecer parâmetros quantitativos que facilitem a sua classificação taxonômica.

4. METODOLOGIA, RESULTADOS E DISCUSSÃO.

A metodologia e os resultados desse trabalho serão apresentados a seguir na forma de artigos científicos já publicados ou em processo de submissão. Os manuscritos serão apresentados em duas partes diferentes e cada um será precedido de um breve resumo, de forma a otimizar a compreensão da conexão entre eles.

4.1- PARTE 1: DESCOBERTA E CARACTERIZAÇÃO DE UM NOVO MEMBRO DA FAMÍLIA *MARSEILLEVIRIDAE* E PROPOSTA DE ATUALIZAÇÃO TAXONÔMICA DO GRUPO.

Nessa primeira parte do trabalho serão descritos o isolamento e a caracterização genômica e filogenética do Marseillevírus cajuinensis, um novo isolado de Marseillevírus brasileiro. Os Marseillevírus são membros da família *Marseilleviridae*, que por sua vez pertence a filo *Nucleocytoviricota* e ao Reino *Bamfordvirae*. A caracterização realizada levou a novas observações e inferências a respeito da evolução e classificação taxonômica desses vírus. Primeiramente, será apresentado o artigo, já publicado, que descreve o isolamento e a caracterização dos vírus, trazendo as análises genômicas e filogenéticas em detalhes. Em segundo lugar, será apresentada a proposta de alteração taxonômica para a família *Marseilleviridae*, que deve ser submetida ao ICTV para que os membros do comitê avaliem as mudanças propostas e as considerem oficialmente.

4.1.1 ARTIGO - The genomic and phylogenetic analysis of Marseillevirus cajuinensis raises questions about the evolution of Marseilleviridae lineages and their taxonomical organization

Esse artigo foi publicado no periódico **Journal of Virology** (DOI: https://doi.org/10.1128/jvi.00513-24)

Resumo: Os Marseillevírus (MsV) são vírus gigantes de ameba pertencentes ao filo *Nucleocytoviricota.* Os MsV já foram isolados de diferentes tipos de amostras, sendo amostras de água doce as mais frequentemente descritas. Esses vírus são classificados em 5 linhagens filogenéticas (A, B, C, D e E), mas a taxonomia oficial da família *Marseilleviridae* ainda não reflete a diversidade de MsV conhecida. Nesse artigo descrevemos o isolamento e a caracterização genômica e filogenética do Marseillevírus cajuinensis, obtido de água salgada do litoral brasileiro. Após o sequenciamento e montagem do genoma, as análises mostraram que o Marseillevírus cajuinensis apresenta uma sequência de 380.653 pares de bases. Esse genoma codifica 515 ORFs, dentre as quais 38 foram consideradas ORFans, pois não foi encontrada similaridade nos bancos de dados. Além disso, o MsV cajuinensis codifica um RNA transportador (RNAt), uma categoria gênica raramente descrita para os Marseillevírus. Nesse contexto, analisamos a presença de RNAt em outros MsV e os resultados revelaram a codificação de RNAt em 5 isolados de Marseillevírus, além do Tokyovírus. As análises filogenéticas sugeriram que o MsV cajuinensis é um membro divergente dentro da linhagem A dos MsV e mostraram também que o ancestral dos Marseillevírus das cinco linhagens se diversificou em três grandes grupos. Outras análises como a distribuição de *Clusters of Orthologous Groups* (COGs) nos MsV, uma análise de distância evolutiva relativa (RED) e análises da média de identidade de aminoácidos e de nucleotídeos entre os diferentes isolados de Marseillevírus reforçaram as observações realizadas, revelando uma nova perspectiva sobre a classificação dos Marseillevírus, tanto com base em suas linhagens filogenéticas, quanto em sua taxonomia oficial. Tomados em conjunto, os resultados deste artigo mostram que a descoberta de novos vírus pode expandir o conhecimento básico sobre famílias virais já conhecidas e pode ajudar a entender cada vez melhor as relações evolutivas existentes entre diferentes isolados. Com uma maior diversidade descoberta, a classificação desses vírus torna-se um desafio e as análises aqui apresentadas podem ajudar a preencher lacunas existentes na taxonomia da família *Marseilleviridae*.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Virology | Full-Length Text

The genomic and phylogenetic analysis of *Marseillevirus cajuinensis* raises questions about the evolution of Marseilleviridae lineages and their taxonomical organization

Bruna Luiza de Azevedo,[1] Victória Fulgêncio Queiroz,[1] Isabella Luiza Martins de Aquino,[1] Talita Bastos Machado,[1] Felipe Lopes de Assis,[1] Erik Reis,[1] João Pessoa Araújo Júnior,[2] Leila Sabrina Ullmann,[2] Philippe Colson,[3,4,5] Gilbert Greub,[6] Frank Aylward,[7,8] Rodrigo Araújo Lima Rodrigues,[1] Jônatas Santos Abrahão[1]

AUTHOR AFFILIATIONS See affiliation list on p. 17.

ABSTRACT Marseilleviruses (MsV) are a group of viruses that compose the Marseilleviridae family within the Nucleocytoviricota phylum. They have been found in different samples, mainly in freshwater. MsV are classically organized into five phylogenetic lineages (A/B/C/D/E), but the current taxonomy does not fully represent all the diversity of the MsV lineages. Here, we describe a novel strain isolated from a Brazilian saltwater sample named Marseillevirus cajuinensis. Based on genomics and phylogenetic analyses, M. cajuinensis exhibits a 380,653-bp genome that encodes 515 open reading frames. Additionally, M. cajuinensis encodes a transfer RNA, a feature that is rarely described for Marseilleviridae. Phylogeny suggests that *M. cajuinensis* forms a divergent branch within the MsV lineage A. Furthermore, our analysis suggests that the common ancestor for the five classical lineages of MsV diversified into three major groups. The organization of MsV into three main groups is reinforced by a comprehensive analysis of clusters of orthologous groups, sequence identities, and evolutionary distances considering several MsV isolates. Taken together, our results highlight the importance of discovering new viruses to expand the knowledge about known viruses that belong to the same lineages or families. This work proposes a new perspective on the *Marseilleviridae* lineages organization that could be helpful to a future update in the taxonomy of the Marseilleviridae family.

IMPORTANCE Marseilleviridae is a family of viruses whose members were mostly isolated from freshwater samples. In this work, we describe the first *Marseillevirus* isolated from saltwater samples, which we called *Marseillevirus cajuinensis*. Most of *M. cajuinensis* genomic features are comparable to other Marseilleviridae members, such as its high number of unknown proteins. On the other hand, *M. cajuinensis* encodes a transfer RNA, which is a gene category involved in protein translation that is rarely described in this viral family. Additionally, our phylogenetic analyses suggested the existence of, at least, three major Marseilleviridae groups. These observations provide a new perspective on Marseilleviridae lineages organization, which will be valuable in future updates to the taxonomy of the family since the current official classification does not capture all the Marseilleviridae known diversity.

KEYWORDS Marseillevirus, Marseilleviridae, lineages, phylogeny, taxonomy, evolution

Some years after the mimivirus discovery (1), the Marseilleviridae family became the second described group of large/giant viruses that are able to infect amoebas. The first Marseilleviridae isolate was obtained from water samples collected in a cooling tower in Paris, France (2). Typically, the members of the Marseilleviridae family exhibit

Editor Kristin N. Parent, Michigan State University, East Lansing, Michigan, USA

Address correspondence to Jônatas Santos Abrahão, jonatas.abrahao@gmail.com.

The authors declare no conflict of interest.

Received 19 March 2024
Accepted 19 April 2024
Published 16 May 2024

icosahedral particles with an average diameter of approximately 250 nm. The discovery of the first Marseillevirus (MsV) paved the way to describe several new isolates. Amoebae are likely marseillevirus hosts in the environment, and *Acanthamoeba* was the species used at the laboratory to isolate these viruses. Most of them were obtained from freshwater samples, including Lausannevirus, Tunisvirus, Melbournevirus, Cannes 8 virus, Noumeavirus, Brazilian marseillevirus, and Tokyovirus (3–9). Samples from other environments have been tested leading to the discovery of new Marseilleviridae isolates. Golden marseillevirus and Insectomime virus, for example, were isolated from golden mussels and internal organs of insect larvae, respectively (10, 11). In addition to several isolates already described, metagenomic studies have reported five divergent Marseilleviridae sequences detected in sediments from a hydrothermal vent (Loki's Castle), in the Atlantic Ocean (12). These discoveries in different types of environments, such as deep ocean regions, have expanded our understanding about the ecology of these viruses and locations where they can be isolated.

Marseilleviridae genomes are composed by circular double-stranded DNA molecules that range in size from 348 to 404 kbp (13). The G-C content ranges from 42.9% to 44.8%, and the number of predicted genes varies from 386 to 491 (13). Most part of Marseillevirus genomes encodes uncharacterized proteins. This is a general characteristic for giant viruses. Besides, Marseilleviruses do not have a diversity of genes involved in protein translation as observed in other giant viruses, such as those of Mimiviridae family (14). Although genes encoding translation factors have been described, it is not common to find in Marseilleviruses genome genes that encode transfer RNAs (tRNA) and aminoacyl-tRNA synthetases (13, 15). Considering tRNAs, they are currently described only in Tokyovirus and in the Loki's Castle metagenomic sequences (8, 12). As other Nucleocytoviricota phylum viruses, Marseillevirus genomes present a high mosaicism, which means that they have sets of genes of multiple origins (2).

The International Committee on Taxonomy of Viruses (ICTV) classifies all members of the Marseilleviridae family within the Nucleocytoviricota phylum, Pimascorivales order (16, 17). Currently, Marseilleviridae is composed by one genus (*Marseillevirus*), in which two species are included: *Marseillevirus marseillevirus* and *Senegalvirus marseillevirus*. There are two other species in the family, which are not included in any genus, the species *Lausannevirus* and *Tunisvirus* (16). However, a new organization and nomenclature for *Marseillevirus*-related taxa was recently proposed and is under approval for official publishing by ICTV. Phylogenetic analyses based on DNA polymerase proteins revealed the existence of five different Marseillevirus phylogenetic lineages, named A, B, C, D, and E. Lineage A is represented by the first isolate, *Marseillevirus marseillevirus* (2). Lineage B is represented by Lausannevirus, the second Marseilleviridae isolated (3). Lineages A and B have the largest number of isolates. The lineage C is represented by Tunisvirus and Insectomime virus (5, 17), while lineages D and E are composed of the Brazilian isolates, Brazilian marseillevirus and Golden marseillevirus, respectively (9, 11). Taking that into account, the known diversity of Marseilleviridae members is not being completely represented by the official taxonomy of the group. For example, the representatives of lineages D and E are not considered yet.

Almost 15 years after the discovery of the first *Marseillevirus*, it becomes apparent that the isolation of new viruses is important to elucidate the evolutionary history of this family. In this study, we report the discovery of *Marseillevirus cajuinensis*, a new isolate obtained from a saltwater sample from the Northeast coast of Brazil. This discovery paved the way for a genomic and phylogenetic characterization that suggested the existence of at least three major consistent Marseilleviridae groups. Analysis involving Clusters of Orthologous Groups (COGs), relative evolutionary distance (RED), and average amino acid/nucleotide identity (AAI and ANI) reinforces this three-group organization and helps to establish parameters for a future taxonomic organization of Marseilleviridae into three genera and different species. These findings provide a new perspective on Marseilleviridae lineages, which will be valuable in evolutive studies and future updates to the taxonomy of the family.

RESULTS

Marseillevirus cajuinensis: a new amoebae-infecting virus isolated from saltwater

From a saltwater sample collected in the Northeast coast of Brazil, we identified a new viral isolate able to infect *Acanthamoeba castellanii*. The transmission electronic microscopy (TEM) images containing infected amoebae showed icosahedral particles presenting shape similar to members of Marseilleviridae (Fig. 1A). With TEM analysis, it was also possible to observe some general morphological aspects from the isolate replication cycle. For example, we observed viral factories (VF) with an electron lucent aspect that occupied a large part of the cell cytoplasm (Fig. 1B, highlighted in pink). The images showed several viral particles in different maturation stages inside the VF (Fig. 1C, black arrow). Amorphic structures were also observed inside VFs (Fig. 1B through D), some of them horseshoe shaped (Fig. 1C and D, yellow arrows), resembling the crescent precursors of viral particles found in poxvirus and mimivirus viral factories (18, 19). Also, we observed giant vesicles harboring several viral particles at the end of the replication cycle. One of these giant vesicles reached a size of 5 µm (Fig. 1E, red arrow).

Besides TEM, the viral sample was analyzed by genome sequencing for a more accurate identification of the virus. Next-generation sequencing generated 225,574 reads that were assembled into a 166× depth single scaffold containing 380,653 bp. When compared with the National Center for Biotechnology Information (NCBI) database, the genome sequence best matched with a member of the Marseilleviridae family, presenting 79% of average nucleotide identity with Tokyovirus. Since we could confirm that the isolate is a Marseilleviridae, we named it *Marseillevirus cajuinensis*.

Marseillevirus cajuinensis genome

The circular double-stranded DNA molecule that composes *M. cajuinensis* genome has a G-C content of 45.24%. A total of 515 open reading frames (ORFs) were predicted that encode proteins with sizes ranging from 50 to 1,520 amino acids. GenBank sequence database searches suggested that 40 of the 515 proteins encoded by *M. cajuinensis* genome have functions related to DNA replication, recombination, and repair (Fig. 2A). This category includes a chaperone, a DNA topoisomerase, different nucleases, helicases, histones, and the DNA polymerase protein, which is commonly used as a marker for phylogeny. Other remarkable categories include signal transduction regulation and miscellaneous (Fig. 2A). The former primarily comprises serine/threonine protein kinases, while the latter consists of proteins whose function cannot be reliably predicted due to the presence of non-specific domains and/or repeats, such as ankyrin repeat-containing proteins and zinc finger proteins. *M. cajuinensis* genome also encodes for proteins involved in different metabolic processes, such as lipases and proteases. Furthermore, the major capsid protein (MCP) and the A32-like packaging ATPase are important proteins found in the virion structure and morphogenesis category.

More than half (59%) of the 515 *M. cajuinensis* proteins were classified as uncharacterized (Fig. 2A). Another 38 proteins were considered as ORFans since they had no hits with any other sequence in the used databases (Fig. 2A). Since several Marseilleviridae isolates have been described, it could be considered a high number of ORFans for a new isolate. However, by analyzing the ORFans' sizes, we observed that most of them (27/38) correspond to short polypeptides ranging from 50 to 100 amino acids (Fig. 2B). The possible high number of ORFans and their short sizes was intriguing. To confirm these results, a new gene prediction was performed using a different program. Thus, Prodigal was used instead of GeneMarkS. After this new analysis, 493 proteins (bigger than 50 amino acids) were predicted by Prodigal, 22 less than the 515 that were predicted by GeneMarkS. Interestingly, the number of predicted ORFans decreased considerably. Instead of 38, the new prediction returned only 9 ORFans. When analyzing the size of these newly predicted ORFans, it was possible to note that the number of ORFans bigger than 150 amino acids was increased, while the number of ORFans that were composed

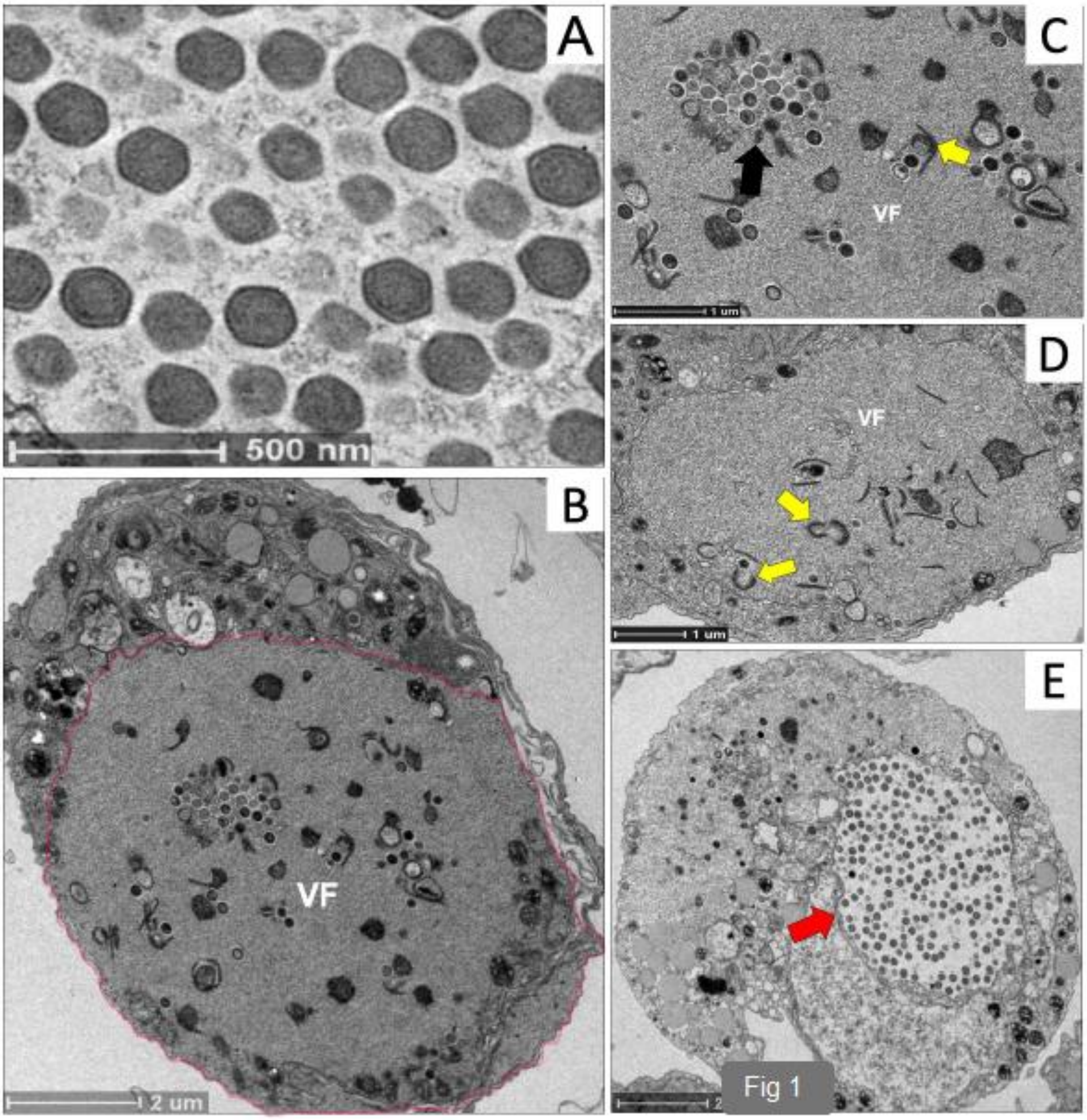


FIG 1 Transmission electronic microscopy images showing morphological characteristics from the new isolate inside *Acanthamoeba castellanii* cells. (A) Several icosahedral Marseillevirus-like particles. (B) *Acanthamoeba castellanii* cell containing a viral factory occupying a large portion of the cell. Viral factory boundaries are highlighted in pink. (C) A portion of Fig. 1B seen closer, showing a VF containing particles in different maturation stages (black arrow) and horseshoe-shaped structures (yellow arrow). (D) Another amoeba cell containing a VF filled with amorphic and horseshoe-shaped structures (yellow arrows). (E) Giant vesicle (red arrow) harboring several viral particles in a final stage of the cycle.

by less than 100 amino acids was considerably low (Fig. 2B). Here, predictions generated by GeneMarkS were selected to perform all the analysis in this work because this tool was the one the most consistently used for Marseilleviruses in previous works (2, 4–6, 9, 10, 17, 20).

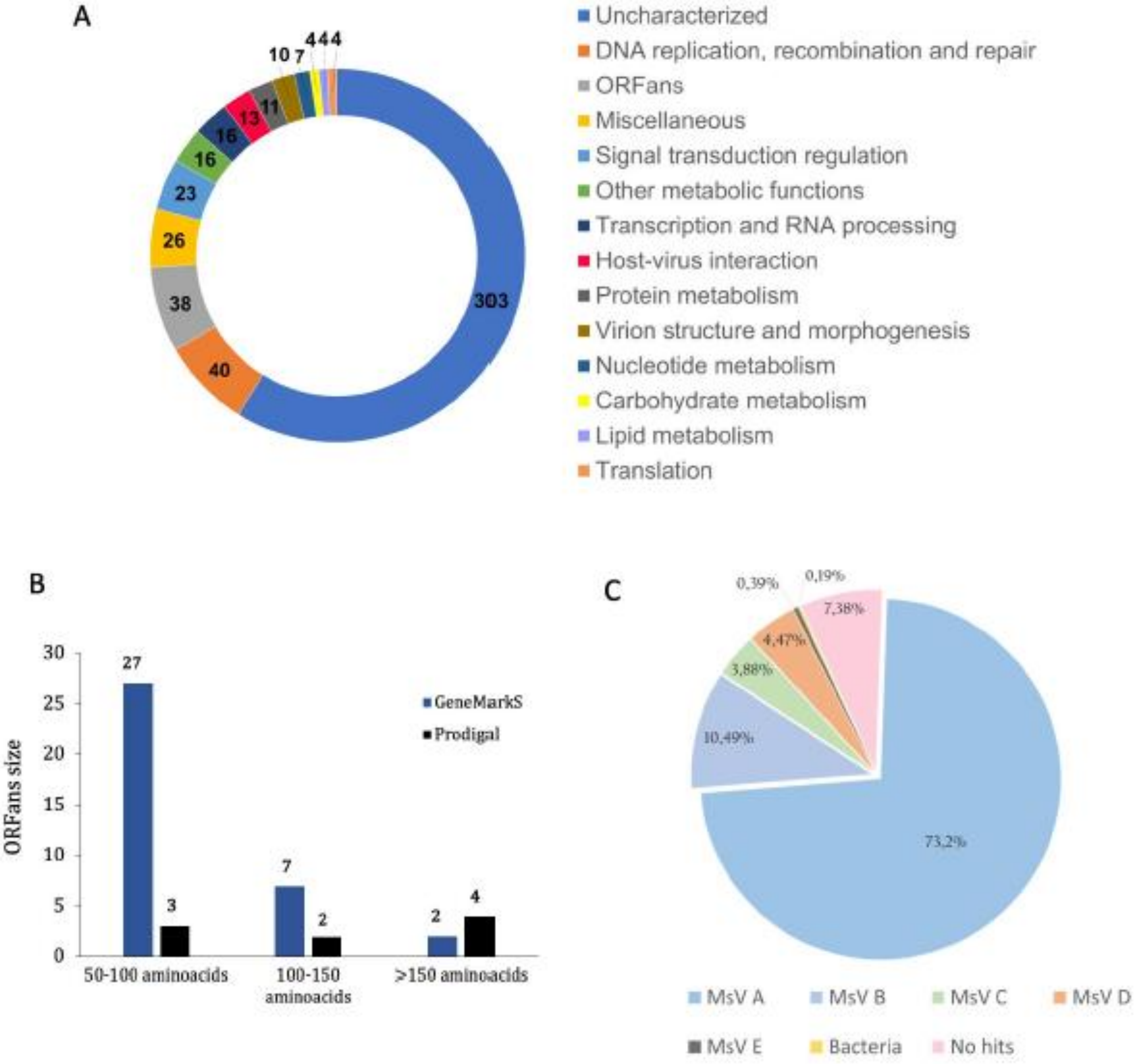


FIG 2 Genomic analysis of the *Marseillevirus cajuinensis* genome. (A) Number of *M. cajuinensis* proteins according to the function predicted during genome annotation. $n = 515$. (B) Comparison between the number of *M. cajuinensis* ORFans and their sizes when using different gene prediction tools. (C) Percentage of *Marseillevirus cajuinensis* proteins that presented best hit (BLASTp) with the five different Marseilleviridae lineages currently described (A–E). $n = 515$.

To know which organisms had their sequences considered as best hits with *M. cajuinensis* proteins, we analyzed hit by hit the results obtained in the BLASTp searches . Thus, for more than half (73.2%) of *M. cajuinensis* proteins, best hits from GenBank database, were Marseillevirus lineage A sequences (MsV A) (Fig. 2C). This suggests that *M. cajuinensis* could present a closer evolutionary relationship with lineage A, which can be confirmed by phylogeny. Also, this best hit analysis showed one *M. cajuinensis* protein (P86) whose best hit was a sequence encoded by a bacterium (Fig. 2C). P86 was the only protein for which no best hits were sequences from other members of the family Marseilleviridae.

Genome synteny analysis of different members of Marseilleviridae lineages highlighted the relatedness between *M. cajuinensis* and previously described lineage A members.

It is possible to observe that the organization of the *M. cajuinensis* genome blocks, separated and colored according to similarity, is more like that of the representative member of lineage A than of viruses from other lineages (Fig. 3). In this analysis, it is also possible to observe that in all the genomes analyzed, there is a region that appears to be more conserved, being approximately the last third of genome lengths (Fig. 3). Such conserved region in Marseilleviridae genomes was already described before and called a "core region" (21). More detailed synteny analysis containing different Marseilleviruses isolates from the five lineages can be found in Supplementary Figure 1 at https://www.giantviruses.com/sup-material-of-papers/sup-material-the-genomic-and-phylogenetic-analysis-of-marseillevirus-cajuinensis-raises-questions-about-the-evolution-of-marseilleviridae-lineages-and-their-taxonomical-organization.

Marseillevirus cajuinensis' translation-related genes and detection of tRNAs in different Marseilleviruses

The genomic analysis showed that *M. cajuinensis* encodes four different translation factors. No aminoacyl-tRNA-synthetase genes were found. Additionally, a search for transfer RNA sequences in *M. cajuinensis* genome was performed. To perform this search, two different programs (Aragorn and tRNAscanSE) were used. No tRNA sequence was found in *M. cajuinensis* genome by using tRNAscanSE. However, Aragorn was able to detect 1 tRNA sequence (tRNA-Gln-CTG) that has a 1,379-nucleotide intron.

Considering all the known giant/large amoeba-infecting viruses, tRNA encoding is not commonly described for Marseilleviridae isolates neither for other families from Pimascovirales. Otherwise, in groups phylogenetically related to the Pimascovirales order, such as cedratviruses and orpheoviruses, tRNAs were already described (22). Because of the difference in results between the two programs used to predict the tRNA, we carried out a search for tRNA in genomes from the five classical Marseilleviridae lineages that had complete sequences available in GenBank (March 2023). For this, we used both Aragorn and tRNAscanSE. The Aragorn program allows changing its parame-

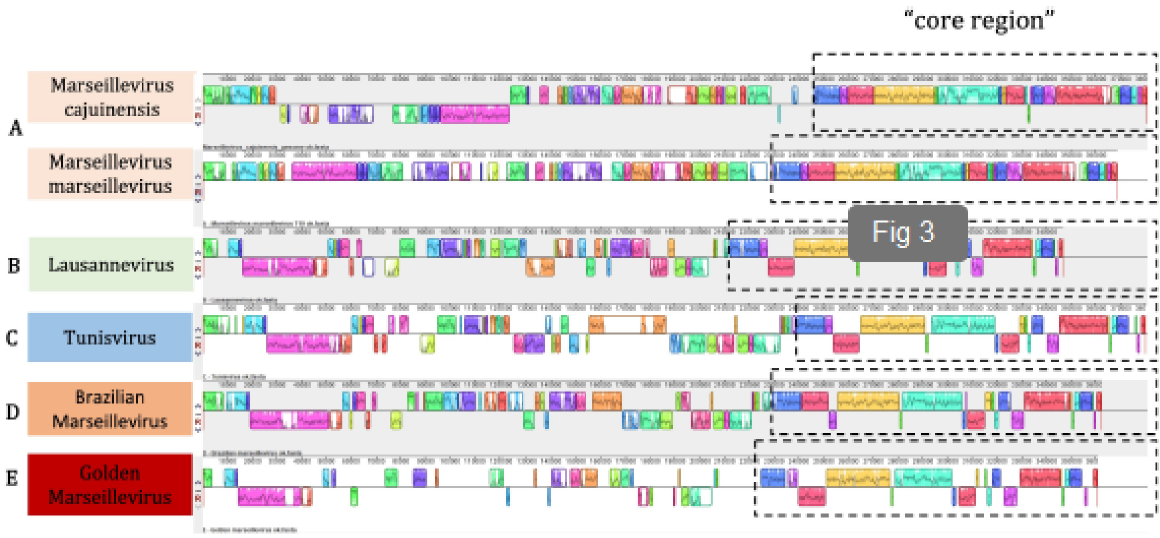


FIG 3 Genome synteny analysis of Marseilleviridae isolates representing the five currently described lineages and *Marseillevirus cajuinensis*. Each line represents the sequence of a different virus, which is identified in the legend on the left. The letters A, B, C, D, and E indicate the respective phylogenetic lineages of each analyzed virus. Blocks of the same color indicate similar regions between sequences. The areas without any colored blocks represent regions exclusive to that virus, that is, which do not show similarity with other viruses used in the analysis. Note: As they have a circular topology, the sequences were adjusted to start from the MCP aiming to facilitate interpretation of this figure. *Marseillevirus marseillevirus* was used as the reference genome.

ters to consider or not the presence of introns, and we performed the analysis in both conditions.

Thus, we detected tRNA sequences in Tokyovirus, *Marseillevirus marseillevirus*, Melbournevirus, Insectomime virus, Tunisvirus, and Golden marseillevirus using the Aragorn program with parameter allowing introns detection. When introns detection was not considered, it was only possible to detect tRNA in Tokyovirus, as the two sequences encoded by this virus do not have introns. Using tRNAscanSE, it was possible to detect tRNA only in the Tokyovirus sequence. Interestingly, tRNAscanSE detected three tRNA sequences in Tokyovirus, while Aragorn detected only two (see Supplementary Figure 2 at https://www.giantviruses.com/sup-material-of-papers/sup-material-the-genomic-and-phylogenetic-analysis-of-marseillevirus-cajuinensis-raises-questions-about-the-evolution-of-marseilleviridae-lineages-and-their-taxonomical-organization). Considering all tRNA sequences detected in different Marseillevirus, only two of the Tokyovirus tRNA were already described (8). Albeit little described in the Marseilleviruses of the five previously reported lineages, tRNAs have already been described in the sequences detected by metagenomics from samples from Loki's Castle. Although genomes assembled from metagenomes should be considered with caution, one of these sequences, called LCMAC202, was reported to encode 26 types of tRNA (12).

Phylogeny of different Nucleocytoviricota conserved proteins raises questions about Marseilleviridae lineages organization

To better elucidate the evolutionary relationship of *M. cajuinensis* with other Marseilleviridae members, phylogenetic analyses were performed using protein sequences that are considered conserved in Nucleocytoviricota. This set of conserved proteins includes the DNA polymerase, the A32-like packaging ATPase, and the late transcription factor VLT3 like, which were used as markers to construct phylogenetic trees (Fig. 4). It is noteworthy that the topology within lineages or genera varies not only among Marseilleviruses but also among other amoebal viruses, depending on the gene analyzed. Virus evolution is modular, with each gene subject to various nuances of a multitude of selective pressures.

Analyzing all phylogenies, it was possible to observe that *M. cajuinensis* groups together with sequences from Marseilleviruses of lineage A but represents a more divergent branch within this group (Fig. 4). Similar results were described for Tokyovirus within lineage A in previous works (8). It is noteworthy that *M. cajuinensis* and Tokyovirus cluster together in separate branch in the VLT3-like tree but not in all constructed trees, including the concatenated one. It is important to note that the divergence of *M. cajuinensis* within lineage A is comparable with the divergence that separates two different lineages (C and D). It raises questions about which criteria should be used to define what can be considered a new lineage or a new genus within the Marseilleviridae family. For example, by comparing lineage C and D branches, Brazilian marseillevirus is currently considered as a different lineage. Based on this, *M. cajuinensis*, and even Tokyovirus, could also be considered new lineages.

In addition to this questioning, it is also possible to observe the clear divergence of the common ancestor of the five classical Marseilleviridae lineages into three major groups: one that groups the current members of lineage A, corresponding to the current genus *Marseillevirus*, another that groups members of the current lineages B-C-D, and finally Golden marseillevirus (lineage E) in a third group. Noteworthy, the lineage E is closer from B-C-D branch than from the *Marseillevirus* genus (lineage A) but still presents a high divergence within its clade. Indeed, this same topology can be observed in a concatenated phylogenetic tree based on the three conserved sequences former analyzed individually (Fig. 4D).

Marseillevirus cajuinensis expands the pangenome of Marseilleviridae isolates

To understand the impact of the *M. cajuinensis* isolation on the pangenome and core genome of Marseilleviridae isolates, we searched for Clusters of Orthologous Groups of

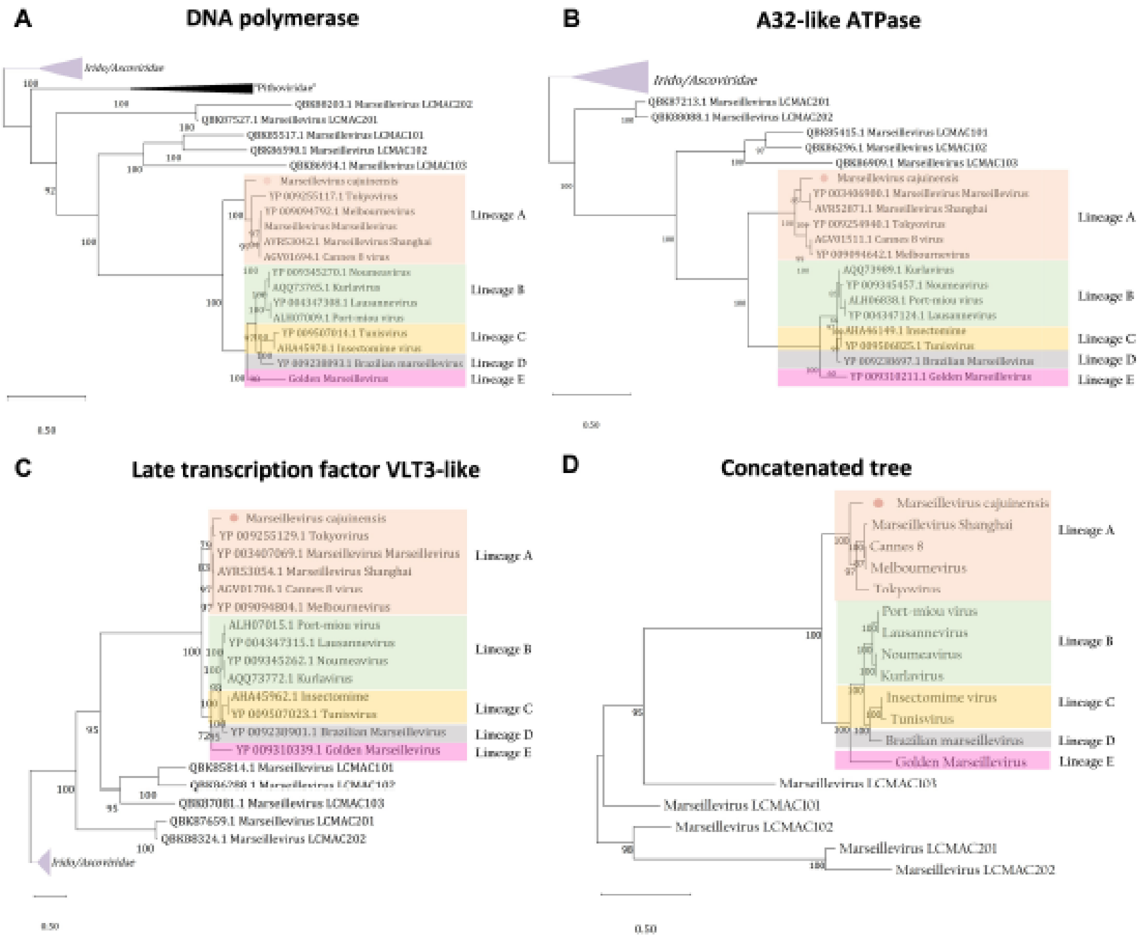


FIG 4 Marseilleviridae phylogeny using different conserved protein sequences. (A) Phylogeny based on DNA polymerase sequences. (B) Phylogenetic tree based on the A32-like packaging ATPase sequences. (C) Phylogeny based on the amino acid sequence of the VLT3-like late transcription factor sequences. (D) Concatenated sequence tree based on DNA polymerase, A32-like packaging ATPase, and VLT3-like late transcription factor sequences. *Marseillevirus cajuinensis* sequence is labeled in the tree with a pink disk. The trees were built using the maximum likelihood method, with statistical support based on 1,000 replicates (bootstrap). The best model, selected by IQtree (ModelFinder), for the trees was VT + F + I + G4 for (A), LG + G4 for (B), LG + I + G4 for (C), and VT + F + I + G4 for (D). The trees shown in A, B, and C were rooted on Iridoviridae branch as an outgroup. Concatenated tree was rooted at the midpoint. The tree scale bars represent the number of amino acid substitutions per site.

proteins shared between complete genome sequences of the isolates available in GenBank. Thus, it was possible to analyze the pangenome and core genome of isolated members of the Marseilleviridae family after including *M. cajuinensis* (Fig. 5) as well as the sharing of COGs between each lineage. It was observed that the pangenome of Marseilleviridae isolates increases from 598 to 1,626 when 13 new isolates are added in the analysis (Fig. 5). The first inserted sequence was Golden marseillevirus because it is the most divergent member of the five lineages. It was expected that with each new discovery of a different virus, the number of total COGs (pangenome) increases. This shows that the Marseilleviridae pangenome is still expanding and that the discovery of new additional viruses is warranted and important, as it will consequently lead to the discovery of new genes.

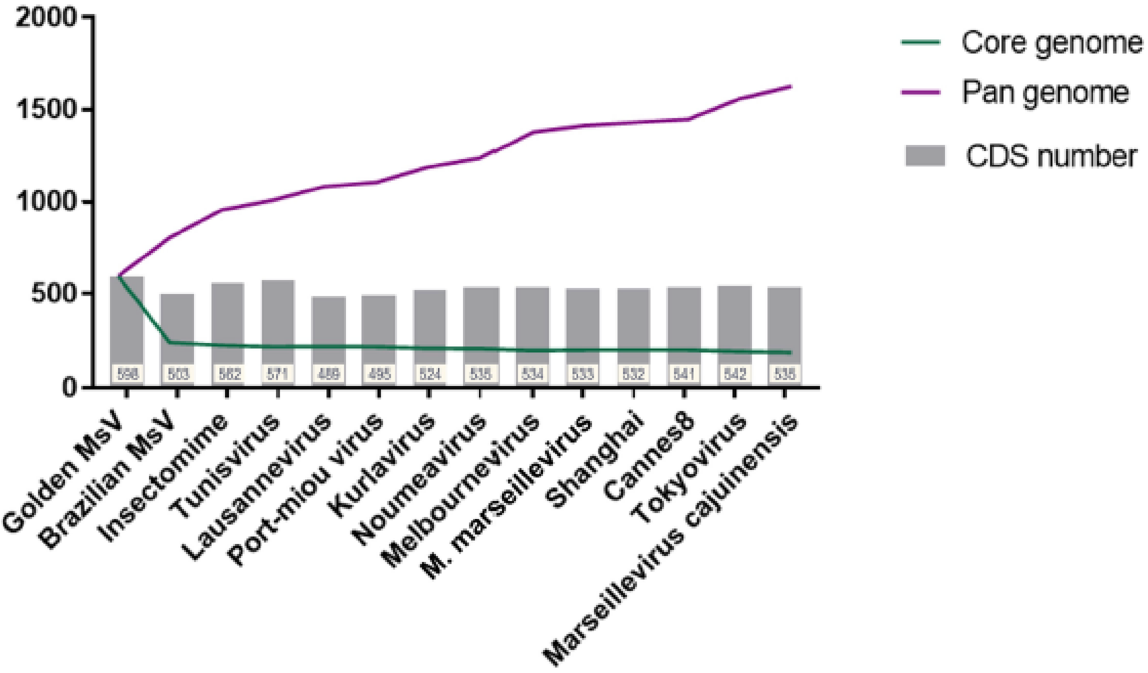


FIG 5 Analysis of the pangenome and core genome of Marseilleviridae isolates, including *Marseillevirus cajuinensis*. The curves indicate the variation in the number of Clusters of Orthologous Groups as new sequences are inserted in analysis. White boxes in the bottom of gray bars indicate the number of coding sequences (CDS) for each virus. Note: The number of CDS is based on the new gene prediction performed exclusively to this analysis.

On the other hand, the number of COGs shared by all the analyzed viruses (core genome) is 182 (Fig. 5). The graph analysis suggests that the core genome of Marseilleviridae isolates appears to have reached a plateau, suggesting that the gene content essential for the existence of these viruses is already relatively well defined, although the functions of most of these genes still need to be deciphered.

A detailed analysis of COGs shared between different Marseilleviruses reinforces the organization of Marseilleviridae in three major groups

In addition to pangenome and coregenome analyses, our data indicate the number of COGs that are shared between the members of the five Marseilleviridae classical lineages (Fig. 6). This analysis shows that *M. cajuinensis* (Fig. 6A, VI, red arrow) has 65 singletons, that is, clusters of proteins that are found only in its sequence. Among the members of lineage A (Fig. 6A, red disks), *M. cajuinensis* (VI, red arrow) and Tokyovirus (I) are those that have the greatest numbers of singletons. This reinforces the assumption that both viruses are the most divergent members of the lineage, corroborating the phylogeny results mentioned above. Altogether, the members of lineage A analyzed here share a total of 370 COGs that are unique to their lineage (Fig. 6A, red disks).

Among the other Marseilleviridae lineages analyzed, in lineage B (Fig. 6A, blue disks), it is possible to observe that all the four lineage members share 137 COGs that are absent in other lineages. Similarly, the lineage C members (Fig. 6A, purple disks) share 133 COGs that are exclusive to their lineage. On the other hand, the only known member of lineage D, Brazilian marseillevirus, has 52 singletons (Fig. 6A, gray arrow). This virus is described to compose its own lineage; however, in this analysis, it presents a smaller number of COGs than *M. cajuinensis* and Tokyovirus which are considered members of a same phylogenetic lineage (lineage A). Golden marseillevirus, the only known representative member of lineage E, has 294 singletons (Fig. 6A, orange arrow). The comparison of the

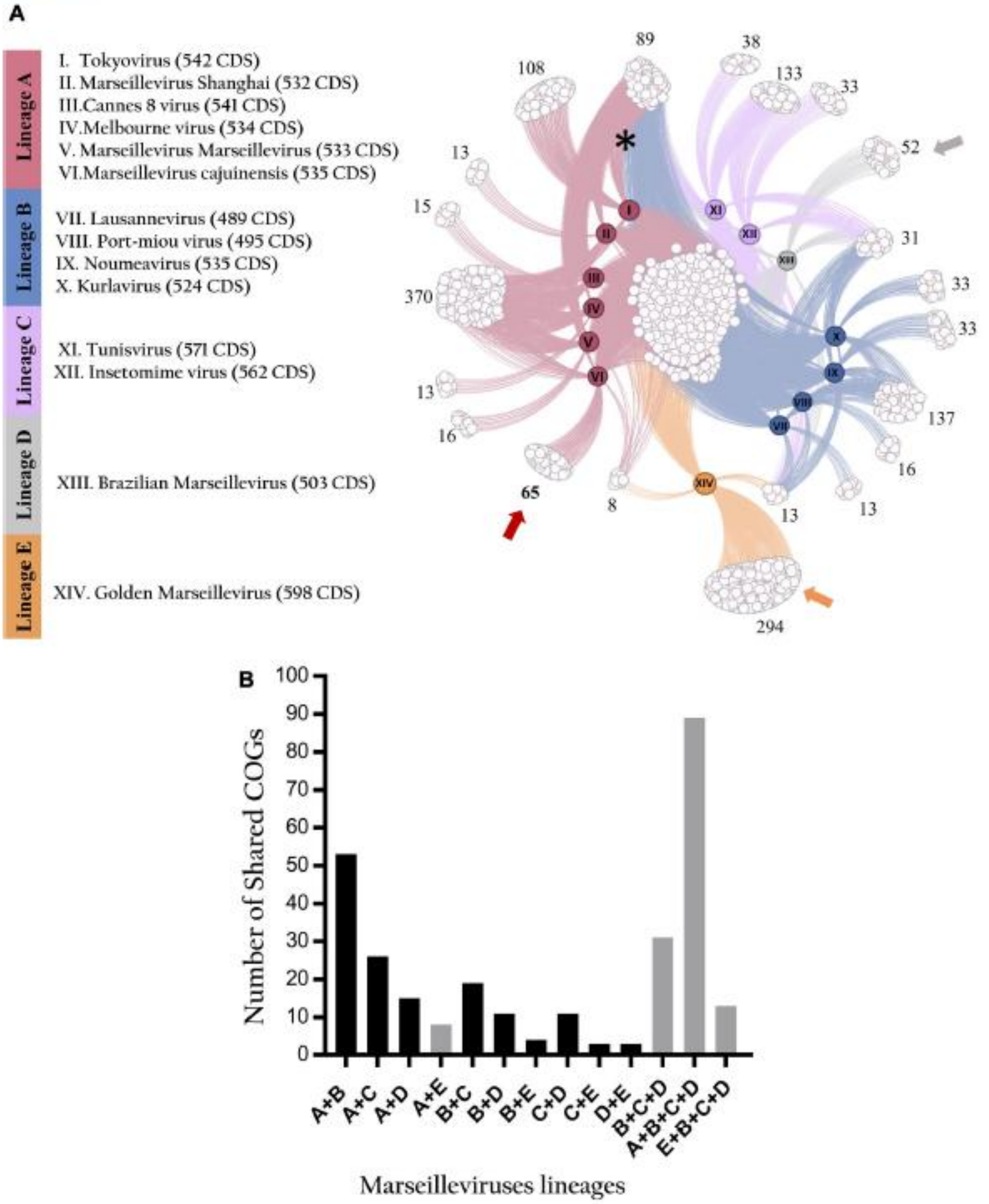


FIG 6 The sharing of Clusters of Orthologous Groups among different viruses belonging to the five classical lineages of Marseilleviruses. (A) Network showing the distribution of COGs among viruses. White circles represent COGs. The colored circles represent the analyzed viruses, separated by color, according to their respective phylogenetic lineages. Roman numerals individually identify each virus analyzed, while Arabic numbers indicate the number of COGs contained

(Continued on next page)

FIG 6 (Continued)

In each cluster. Note: The number of proteins obtained though GeneMarkS is indicated in parentheses for each virus. *Marseillevirus cajuinensis* is highlighted by a red arrow, Golden marseillevirus is highlighted by an orange arrow, Brazilian marseillevirus is highlighted by a gray arrow, and the sharing of COGs between lineages A, B, C, and D is highlighted by an asterisk. (B) Graph detailing the number of exclusive COGs shared between combinations of different classical Marseilleviridae lineages. The black bars correspond to combinations of lineages that are not represented in Fig. 6A, while the gray bars correspond to combinations already represented in Fig. 6A. Note: Some of the lineage's combinations were not represented in Fig. 6A because some of them often overlap themselves in the network, making analysis hard.

number of exclusive shared COGs between the different lineages is detailed in Fig. 6B. In this graph, it is possible to analyze the sharing of COGs between different lineages in a simpler way. Lineages A and E (A + E), for example, share only eight COGs. This number decreases when comparing the other lineages with lineage E. Lineages B and E (B + E) share only four COGs, while lineages C and E (C + E), and D and E (D + E) share only COGs. Together, lineages A, B, C, and D (A + B + C + D) share 89 COGs that are not found in lineage E (Fig. 6A, asterisk). Thus, these data reinforce the divergence of lineage E among Marseilleviridae and support its assignment in a distinct group of the family.

Using the data obtained in COGs analysis described above, a hierarchical clustering phenophyletic tree was constructed based on the presence and absence of COGs in the analyzed sequences (Fig. 7). In this figure, it was possible to observe a topology very similar to what was observed in the phylogenetic trees described in this work. Representatives of lineage A are organized into a group (group I) that represents the current genus *Marseillevirus* and apart from representatives of lineages B, C, D (group II), and E (group III) (Fig. 7).

Within lineage A, it was possible to observe the organization of the viruses into two subgroups, one was composed of *M. cajuinensis* and Tokyovirus, and the other was composed of the other lineage A viruses (Fig. 7). The same happens in group II since there are three subgroups composed by each member of lineages B, C, and D

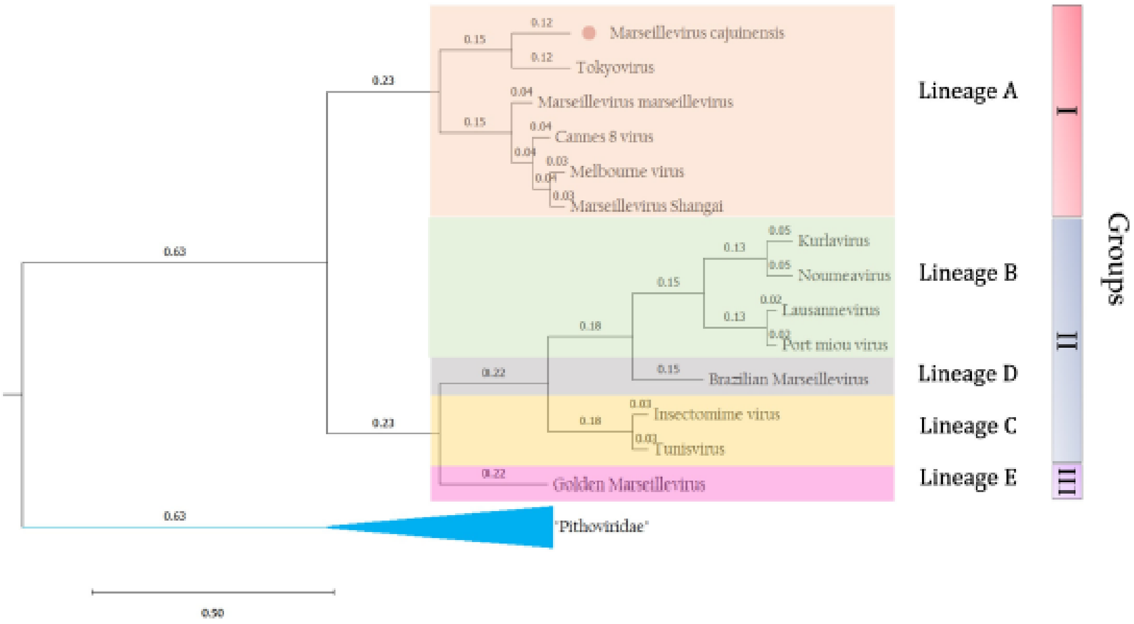


FIG 7 Hierarchical clustering tree considering the presence and absence of Clusters of Orthologous Groups in different viruses of Marseilleviridae family. *Marseillevirus cajuinensis* is labeled in the tree with a pink circle. Scale bar represents arbitrary values that express the evolutionary distance based on presence-absence of COGs.

;

(Fig. 7). These results reinforce what was observed in phylogeny and genomic analyses. Also, they reinforce the questions raised about the classical organization of Marseilleviridae viruses in five lineages. For example, if the tree branches (lineages B, C, and D) that compose group II are represented by viruses originally classified into three different lineages, this could justify that *Marseillevirus cajuinensis* and Tokyovirus could be classified in its own lineages. On the other hand, the different branches that compose the classical Marseilleviridae lineages could also be considered as members of major groups.

Relative evolutionary distance and average amino acid/nucleotide identity analyses suggest the organization of Marseilleviridae isolates in three putative genera

To quantify all these observations and make them more consistent, an RED analysis was performed. Thus, we considered three groups for this analysis based on the three major clades observed in DNA polymerase phylogeny: group I (lineage A), group II (lineages B, C, D), and group III (lineage E) (Fig. 8A). DNA polymerase phylogeny was selected because it is the main phylogenetic marker for Nucleocytoviricota, but the same topology was observed for all phylogenies (Fig. 4) and for the hierarchical clustering tree as well (Fig. 7).

RED values varies from 0 to 1, and threshold values for different taxonomic levels in Nucleocytoviricota were defined previously (23). The previously reported RED values for genus ranged from 0.69 to 0.995 (23). The present analysis showed that group I (lineage A) had a RED value of 0.86, while group II (lineages B, C, and D) had a RED value of 0.83 (Fig. 8A). These numbers are consistent with values expected for the genus level (23). Because group III (lineage E) is composed by a single genome (Golden marseillevirus), RED analysis cannot be performed. However, when Golden marseillevirus sequence is included in group II (lineages B, C, and D), the RED value decreases to 0.78. Although this value could still classify group II as a genus while including Golden marseillevirus, the lower RED value makes group II less consistent when comparing with group I. Thus, excluding lineage E from group II and considering it as a separate group were most consistent here.

Additionally, sequences of the Marseillevirus isolates previously analyzed in this work were submitted to an average nucleotide identity analysis. The ANI analysis delineated three main groups of Marseilleviruses, considering an ANI cutoff >75% (Fig. 8B). This value corroborates with our phylogenetic analyses and hierarchical clustering of COGs, revealing the existence of three distinct groups in Marseilleviridae, possibly corresponding to three distinct genera. Within groups, we can use pairwise ANI >95% to define viral species, as used for members of Imitervirales (24). In this case, we can define eight viral species, among which three belong to group I—one of them corresponding to the *Marseillevirus cajuinensis*; four belong to group II; and only one, consisting of the Golden marseillevirus isolate, belongs to group III. Also, an average amino acid identity analysis was performed. The same three groups could be observed in the bidirectional AAI analysis, with AAI >65% (Fig. 8B). Moreover, in the AAI estimation, we clearly saw eight putative viral species, with AAI >95%.

DISCUSSION

In this work, we describe the isolation of *M. cajuinensis* from a saltwater sample from the Northeast coast of Brazil. Although members of the family Marseilleviridae have already been isolated from samples of different sources (2, 10, 11, 17, 25), most of them were obtained from freshwater or mud from rivers and lakes (2–4, 6–8). To our knowledge, this is the first time that a Marseilleviridae member has been isolated from ocean water, although they have already been detected in this type of environment through metagenomic analyses (12).

TEM images revealed that the replication cycle of *M. cajuinensis* shares similar characteristics with other members of the Marseilleviridae family, such as the presence of

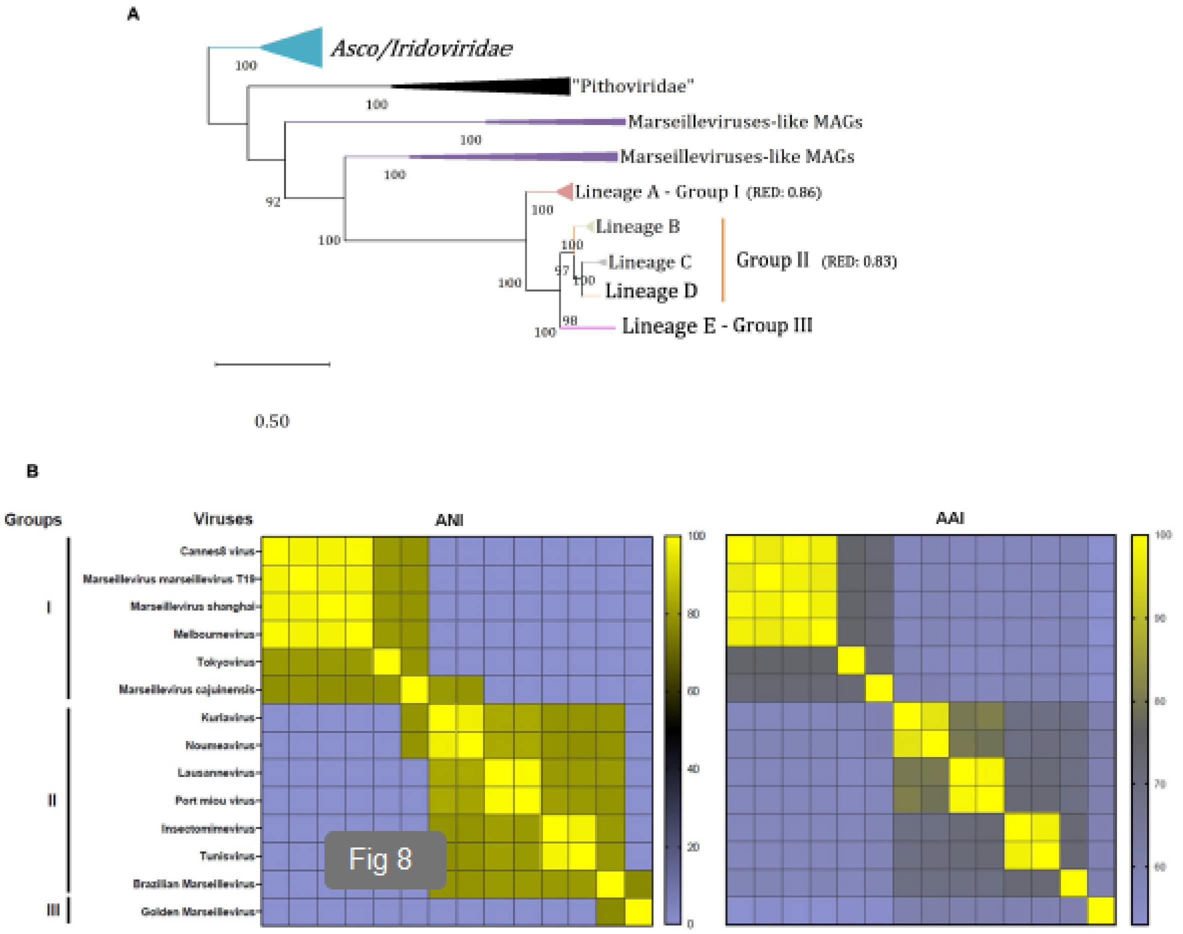


FIG 8 (A) DNA polymerase phylogenetic tree illustrating the three-group proposal for Marseilleviridae. RED values are indicated for groups I and II, and the numbers are consistent with the genus level for taxonomy. Group III does not have a RED value because there is only one virus in lineage E. The tree was built using the maximum likelihood method, with statistical support based on 1,000 replicates (bootstrap). The best model, selected by IQtree (ModelFinder), for the tree was VT + F + I + G4. The tree was rooted on Iridoviridae branch as an outgroup. The tree scale bars represent the number of amino acid substitutions per site. Note: MAGs, metagenome-assembled genomes. (B) Average nucleotide identity and average amino acid identity analysis of Marseilleviridae. Fourteen Marseilleviruses are grouped based on a similarity matrix composed by ANI (left heatmap) and AAI (right heatmap). The three viral groups are indicated. ANI <75% was set to zero. ANI values ranged from 77 to 100, and AAI values ranged from 53 to 100.

giant vesicles by the end of the cycle (26). These vesicles are important structures for their replication cycles, as the viruses can be released from the cell inside these vesicles, referred to as "expelled vesicles" (27, 28). This mechanism is important to ensure greater effectiveness in the subsequent entry of particles into another amoeba to start new cycle, corresponding to a sort of Trojan horse strategy. Alternatively, such structure may be related to the exocytosis of the viral progeny (26) and their increased resistance to harsh environments when being outside amoebae, waiting for new hosts. The presence of amorphic and horseshoe-like structures inside the viral factory (Fig. 1C and D) requires further investigations, and more detailed analyses of the replication cycle are needed to infer a biological function to these structures. It is known that in the initial stages of Marseilleviridae isolates viral factory formation, endosomal membranes are recruited. These recruited membranes are involved in the formation of the internal membranes that compose viral capsids (26). Therefore, it can be hypothesized that *M. cajuinensis* VF structures have been recruited from some cellular component to perform a function that is not yet known.

After analyzing the genome of *M. cajuinensis*, we observed that its genome size, GC content, and number of genes are compatible with that described for other members of the Marseilleviridae family (2, 3, 5, 9, 11, 13). The functional categorization of proteins suggests that more than half of the *M. cajuinensis* genome encode proteins with unknown functions. This is a very common characteristic among giant amoeba viruses and reinforces the need for new studies that aim the elucidation of the unknown functions of these proteins (1, 29, 30). Also, we detected 38 ORFans in a first gene prediction, but most of them have sizes that range from 50 to 100 amino acids. This puts in doubt whether these sequences are real proteins or whether they are artifacts due to the main gene prediction protocol used in this study. When using a new gene prediction tool, the number of ORFans decreased to 9. This considerable difference in results might be due to the lack in updates of prediction tools that currently are mostly indicated for prokaryotes, eukaryotes, and for viruses that have smaller and fewer complex genomes. The number of ORFs was also affected when different parameters and tools for prediction are utilized. In addition, we detected a gene that encodes a tRNA, which is not commonly described in members of the family Marseilleviridae (13, 14). By analyzing sequences from other Marseilleviridae members using different programs and parameters, we detected tRNAs in five Marseilleviruses that, to our knowledge, have not been previously described as encoding such sequences. The absence of tRNA detection might be attributed to the use of a single algorithm in most previous analyses. Our current analysis shows a difference in the sensitivity of tRNA detection, with Aragorn being more sensitive than the tRNAscanSE algorithm tested in parallel. This is likely due to the different tRNA search models and parameters used by each tool (31, 32). For example, the Aragorn does not depend on the taxonomic lineage specification as parameter to achieve maximum search sensitivity, whereas the tRNAscan-SE does (31). Thus, current gene prediction algorithms lack updates that consider the singularities of giant viruses. This highlights the importance to systematically use more than one algorithm for ORFs and tRNA prediction, as they can complement each other and can stimulate deeper investigations.

Comparisons between *M. cajuinensis* and members of the *Marseillevirus* genus, conducted through genomic and phylogenic analyses, showed that they are all grouped within the lineage A. Despite this, *M. cajuinensis* and Tokyovirus form a divergent branch within this lineage. Phylogenetic analyses also showed that the common ancestor of the five classical Marseilleviridae lineages is in fact diversified into three main branches, which we refer to as group I (lineage A), group II (lineages B, C, and D), and group III (lineage E). After analyzing parameters such as RED, AAI, and ANI, it is possible to suggest that these three groups could potentially be considered as three genera. Taxonomically, group I currently corresponds to *Marseillevirus* genus, and group II corresponds to the recently proposed *Losanna* genus (that includes the *L. lausannense* and the *L. tunisiense* species). Group III and some members of the other proposed groups (e.g. Brazilian marseillevirus) remain to be officially assigned taxonomically. Also, AAI and ANI analyses suggested the organization of Marseilleviridae isolates in eight species (see Supplementary Figure 3 at https://www.giantviruses.com/sup-material-of-papers/sup-material-the-genomic-and-phylogenetic-analysis-of-marseillevirus-cajuinensis-raises-questions-about-the-evolution-of-marseilleviridae-lineages-and-their-taxonomical-organization), according to the percentage of sequence identity between each other. This could be helpful to classify the isolates that are still not considered officially.

The analyses of COGs shared between different lineages highlight conserved and variable COGs in each lineage. An in-depth analysis to understand the function of each protein belonging to the clusters was not performed here, and most part of these proteins might not have a known function. However, it is possible to hypothesize that conserved COGs might represent important genes to the lineage, possibly inherited from their common ancestor. Conversely, the proteins belonging to clusters that vary among lineages might have different origins. The presence/absence of COGs analysis complemented the phylogeny and reinforced both the greater divergence of *Marseillevirus cajuinensis* within lineage A and the organization of Marseilleviridae classical lineages into three groups. However, is it worth mentioning that defining the number of Marseilleviridae lineages/genera is a big challenge and might be treacherous as it depends on the methods and the viruses considered in the analysis. Taking that into account, it is clear that there is a need to continue efforts to obtain new isolates as the Marseilleviridae pangenome is still open. The raised questions about the number of lineages within the Marseilleviridae family reflect the impact of new virus discovery on taxonomists' perspectives, as these new strains add new information to the analyses, sometimes leading to different tree topologies. Such new topologies call for updating phylogenetic organization and taxonomy to ensure that genus and species taxonomic levels better reflect the reality of diversity in a given taxon, rather than being biased due to sampling mostly specific habitats, such as freshwater. For this taxonomic update to happen, it is necessary to first establish the parameters that are needed to classify these viruses into new species or genus. Thus, this work represents a contribution to shape future updates in Marseilleviridae taxonomy.

MATERIALS AND METHODS

Viral isolation, multiplication, purification, and titration

The isolate was obtained through the collection of saltwater samples at Cajueiro da Praia city, located in Piaui state (Northeast coast of Brazil). The protocol was based on the inoculation of the collected samples on 96-well plates containing *Acanthamoeba castellanii* cells (33). The inoculated wells that presented cytopathic effects (i.e., rounding cells and cellular lysis) had their content collected and analyzed through transmission electron microscopy. After confirming the isolation, the virus was inoculated at a multiplicity of infection (MOI) of 0.01 in cell culture flasks containing 1.4×10^7 *Acanthamoeba castellanii* cells and 35 mL of peptone-yeast extract-glucose (PYG) medium, supplemented with penicillin (100 U/mL; Cellofarm, Brazil), streptomycin (100 µg/mL; Sigma-Aldrich, USA), and amphotericin B (0.25 µg/mL; Cultilab, Brazil). The cells were incubated at 32°C. Non-infected cells maintained in the same conditions were used as control. When viral-induced cytopathic effects were observed, the flask's content was collected. This content was filtered through 0.45 µm pores, and then it was ultracentrifuged (36,000 x g) in a 25% sucrose cushion for 2 hours. The pellet containing purified viral particles was homogenized in 300 µL of phosphate-buffered saline (PBS 1×). All the viral titers were obtained and calculated using the end-point method (34).

Transmission electron microscopy

To analyze the morphology of isolated viral particles, the samples were prepared for TEM. First, 7×10^6 *A. castellanii* cells, cultured in 25 mL of PYG medium, were inoculated with the virus at an MOI of 0.01. Once cytopathic effects were observed, we performed two consecutive washes with 0.1 M sodium phosphate buffer, and we subsequently fixed the cells for 2 hours at room temperature under rotation in an orbital mixer. The fixation solution consisted of 0.1 M sodium phosphate buffer and 2.5% glutaraldehyde. Following this initial fixation step, the cells underwent a secondary fixation with 2% osmium tetroxide before being embedded in Epon resin. This resin allowed an ultramicrotomy, and the 60-nm-thick sections were then examined using a transmission electron microscope (Spirit Biotwin FEI-120 kV) at the Center of Microscopy of the Federal University of Minas Gerais (CM-UFMG).

Sequencing, assembly, and annotation

The purified virus was sequenced using an Illumina MiSeq instrument with a paired-end library using the Illumina DNA Prep Kit (Illumina Inc., San Diego, CA, USA). The FastQC program was used for quality control of the obtained reads, and the per base sequence quality was considered satisfactory (phred >28). The reads were trimmed using the Trimmomatic tool (35). Genome *de novo* assembly was performed using Spades 3.12 program with default parameters (36, 37). The obtained scaffold was compared with sequences from the NCBI database, using BLASTn (database: nr/nt; expect threshold: 10^{-3}). Open reading frames were predicted with the GeneMarkS tool and Prodigal (38, 39), considering only proteins that were bigger than 50 amino acids. Additionally, tRNA-coding sequences (CDSs) were predicted using ARAGORN (parameters – type: tRNA; allow intron: yes and no (alternately); topology: circular; strand- both) and tRNAscanSE (parameters - Sequence source: general tRNA model; Search mode: default, Genetic Code for tRNA Isotype Prediction: universal) (31, 32). ORFs were annotated using BLASTp (expect threshold: 10^{-3}) against the NCBI non-redundant protein sequence (nr) database aiming to search for similar sequences in this database. The functional categorization of predicted proteins was carried out based on the Nucleo-Cytoplasmic Virus Orthologous Groups (40, 41).

Synteny and phylogenetic analysis

To perform synteny analyses, genome sequences of different MsV isolates were obtained from the NCBI GenBank database. Only genomes from isolated Marseilleviruses (excluding those built from metagenomic data) and that were complete and available in GenBank (March 2023) were selected. As they have a circular topology, the sequences were manually curated to start from the major capsid protein aiming to facilitate image interpretation. After curating the sequences, synteny analysis was performed using the MAUVE program, with default parameters (42). The following genome sequences were retrieved from GenBank and then analyzed: Tokyovirus (NC_030230.1); *Marseillevirus marseillevirus* (GU071086.1); Cannes 8 virus (KF261120.1); Marseillevirus Shanghai (MG827395.1); Melbournevirus (KM275475.1); Kurlavirus (KY073338.1); Lausannevirus (HQ113105.1); Noumeavirus (KX066233.1); Port-miou virus (KT428292.1); Insectomime (HG428764.1); Tunisvirus (KF483846.1); Brazilian MsV (KT752522.1); Golden MsV (KT835053.1).

Phylogenetic trees were constructed using the IQtree software (version 1.6.12) with 1,000 bootstrap replicates as branch support (43). To prepare the data sets for alignment, a search for similar sequences was performed using the NCBI non-redundant protein sequences (nr) database and BLASTp with an expected threshold of 10^{-3}. Sequence alignment was performed using the MUSCLE algorithm (44). The best-fit substitution models were determined using the ModelFinder algorithm within IQtree (45). Finally, the resulting phylogenetic trees were visualized and edited using MEGA X software (46).

Relative evolutionary distance analyses were performed using phylogeny constructed according to parameters mentioned above. RED values were calculated using the R package "castor" (47), and the thresholds for taxonomic levels were defined as described in the Results section based in a previous work (23).

Pangenome and COGs analysis

All complete MsV sequences that were previously obtained for synteny analyses from GenBank were also subjected to a new gene prediction using GeneMarkS (38). The amino acid sequences of each predicted CDS were analyzed using the ProteinOrtho software (parameters - selfblast, identity: 30%, coverage: 50%, and *e*-value of 10^{-5}) (48). The output files generated by ProteinOrtho were used to analyze the pangenome and core genome of isolated MsV. Also, output files of orthologous proteins were used to compare the number of Clusters of Orthologous Groups that are shared between the studied viruses and to construct a hierarchical clustering based on the presence and absence of COGs in different MsV. To analyze the sharing of COGs between different MsV, a network representation was constructed using the Gephi 0.10 software. For this, data obtained in ProteinOrtho analysis were used to create spreadsheets containing the "nodes" (viruses and COGs) and the "edges" (presence of COGs in each virus). Network representation was built using an algorithm based on attraction and repulsion forces (Force Atlas). To perform COGs presence and absence analysis, a binary file was generated, and a phenetic tree was created in the MultiExperiment Viewer program, version 4.9.0, using the hierarchical clustering algorithm and the Pearson correlation as distance metric (49).

Average nucleotide and amino acid identities

Whole-genome average nucleotide identity analysis was performed using FastANI (50) implemented on Galaxy Server (https://usegalaxy.eu/), on the complete genomes of 14 Marseilleviruses obtained inform the NCBI GenBank database. ANI <75% was considered 0. Average amino acid identity was calculated using reciprocal best hits (two-way AAI) between two Marseilleviruses' protein genomic data sets, considering an identity cutoff of 20%. AAI was estimated using the AAI calculator (http://enve-omics.ce.gatech.edu/aai/).

ACKNOWLEDGMENTS

We would like to thank all colleagues from Grupo de Estudo e Prospecção de Vírus Gigantes (GEPVIG) and from Laboratório de Vírus of Universidade Federal de Minas Gerais (UFMG). Also, we thank Centro de Microscopia of UFMG for the contribution of microscopy images.

We thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), and FAPEMIG (Fundação de Amparo à Pesquisa do estado de Minas Gerais) for research funding.

This research is registered at SISGEN and SISBIO.

J.S.A., J.P.A.J., and R.A.L.R. are CNPq researchers.

P.C., G.G., and J.S.A. are contributors of the ICTV Marseillevirus study group (2016-2023).

AUTHOR AFFILIATIONS

[1]Laboratório de Vírus, Departamento de Microbiologia, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil
[2]Laboratório de Virologia, Departamento de Microbiologia e Imunologia, Instituto de Biotecnologia, Universidade Estadual Paulista (Unesp), Alameda das Tecomarias s/n, Chácara Capão Bonito, Botucatu, Brazil
[3]IHU Méditerranée Infection, Marseille, France
[4]Microbes Evolution Phylogeny and Infections (MEPHI), Aix-Marseille University, Marseille, France
[5]Assistance Publique-Hôpitaux de Marseille (AP-HM), Marseille, France
[6]Centre for Research on Intracellular Bacteria and Giant Viruses, Institute of Microbiology, University Hospital Centre and University of Lausanne, Lausanne, Switzerland
[7]Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia, USA
[8]Center for Emerging, Zoonotic, and Arthropod-Borne Infectious Disease Virginia Tech, Blacksburg, Virginia, USA

AUTHOR ORCIDs

Gilbert Greub http://orcid.org/0000-0001-9529-3317
Frank Aylward http://orcid.org/0000-0002-1279-4050
Rodrigo Araújo Lima Rodrigues http://orcid.org/0000-0001-7148-4012
Jônatas Santos Abrahão http://orcid.org/0000-0001-9420-1791

DATA AVAILABILITY

The *M. cajuinensis* genome sequence is available in GenBank under accession number OR991738.

REFERENCES

1. Scola BL, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raoult D. 2003. A giant virus in amoebae. Science 299:2033–2033. https://doi.org/10.1126/science.1081867
2. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, Robert C, Azza S, Sun S, Rossmann MG, Suzan-Monti M, La Scola B, Koonin EV, Raoult D. 2009. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc Natl Acad Sci U S A 106:21848–21853. https://doi.org/10.1073/pnas.0911354106
3. Thomas V, Bertelli C, Collyn F, Casson N, Telenti A, Goesmann A, Croxatto A, Greub G. 2011. Lausannevirus, a giant amoebal virus encoding histone doublets. Environ Microbiol 13:1454–1466. https://doi.org/10.1111/j.1462-2920.2011.02446.x
4. Aherfi S, Pagnier I, Fournous G, Raoult D, La Scola B, Colson P. 2013. Complete genome sequence of Cannes 8 virus, a new member of the proposed family "Marseilleviridae". Virus Genes 47:550–555. https://doi.org/10.1007/s11262-013-0965-4
5. Aherfi S, Boughalmi M, Pagnier I, Fournous G, La Scola B, Raoult D, Colson P. 2014. Complete genome sequence of Tunisvirus, a new member of the proposed family *Marseilleviridae*. Arch Virol 159:2349–2358. https://doi.org/10.1007/s00705-014-2023-5
6. Doutre G, Philippe N, Abergel C, Claverie J-M. 2014. Genome analysis of the first *Marseilleviridae* representative from Australia indicates that most of its genes contribute to virus fitness. J Virol 88:14340–14349. https://doi.org/10.1128/JVI.02414-14
7. Fabre E, Jeudy S, Santini S, Legendre M, Trauchessec M, Couté Y, Claverie J-M, Abergel C. 2017. Noumeavirus replication relies on a transient remote control of the host nucleus. Nat Commun 8:15087. https://doi.org/10.1038/ncomms15087
8. Takemura M. 2016. Morphological and taxonomic properties of *Tokyovirus*, the first *Marseilleviridae* member isolated from Japan. Microbes Environ 31:442–448. https://doi.org/10.1264/jsme2.ME16107
9. Dornas FP, Assis FL, Aherfi S, Arantes T, Abrahão JS, Colson P, La Scola B. 2016. A Brazilian marseillevirus is the founding member of a lineage in family *Marseilleviridae*. Viruses 8:76. https://doi.org/10.3390/v8030076
10. Colson P, Fancello L, Gimenez G, Armougom F, Desnues C, Fournous G, Yoosuf N, Million M, La Scola B, Raoult D. 2013. Evidence of the megavirome in humans. J Clin Virol 57:191–200. https://doi.org/10.1016/j.jcv.2013.03.018
11. Dos Santos RN, Campos FS, Medeiros de Albuquerque NR, Finoketti F, Côrrea RA, Cano-Ortiz L, Assis FL, Arantes TS, Roehe PM, Franco AC. 2016. A new marseillevirus isolated in Southern Brazil from *Limnoperna fortunei*. Sci Rep 6:35237. https://doi.org/10.1038/srep35237
12. Bäckström D, Yutin N, Jørgensen SL, Dharamshi J, Homa F, Zaremba-Niedwiedzka K, Spang A, Wolf YI, Koonin EV, Ettema TJG. 2019. Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. mBio 10:e02497-18. https://doi.org/10.1128/mBio.02497-18
13. Sahmi-Bounsiar D, Rolland C, Aherfi S, Boudjemaa H, Levasseur A, La Scola B, Colson P. 2021. Marseilleviruses: an update in 2021. Front Microbiol 12:648731. https://doi.org/10.3389/fmicb.2021.648731
14. Rodrigues RAL, da Silva LCF, Abrahão JS. 2020. Translating the language of giants: translation-related genes as a major contribution of giant viruses to the virosphere. Arch Virol 165:1267–1278. https://doi.org/10.1007/s00705-020-04626-2
15. Abrahão JS, Araújo R, Colson P, La Scola B. 2017. The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. PLoS Genet 13:e1006532. https://doi.org/10.1371/journal.pgen.1006532
16. Current ICTV taxonomy release. ICTV. Available from: https://ictv.global/taxonomy. Retrieved 01 Oct 2022.
17. Boughalmi M, Pagnier I, Aherfi S, Colson P, Raoult D, La Scola B. 2013. First isolation of a Marseillevirus in the Diptera Syrphidae Eristalis tenax. Intervirology 56:386–394. https://doi.org/10.1159/000354560
18. Andrade A, Rodrigues RAL, Oliveira GP, Andrade KR, Bonjardim CA, La Scola B, Kroon EG, Abrahão JS. 2017. Filling knowledge gaps for mimivirus entry, uncoating, and morphogenesis. J Virol 91:e01335-17. https://doi.org/10.1128/JVI.01335-17
19. Maruri-Avidal L, Weisberg AS, Moss B. 2013. Association of the vaccinia virus A11 protein with the endoplasmic reticulum and crescent precursors of immature virions. J Virol 87:10195–10206. https://doi.org/10.1128/JVI.01601-13
20. Chatterjee A, Kondabagil K. 2017. Complete genome sequence of Kurlavirus, a novel member of the family *Marseilleviridae* isolated in Mumbai, India. Arch Virol 162:3243–3245. https://doi.org/10.1007/s00705-017-3469-z
21. Blanca L, Christo-Foroux E, Rigou S, Legendre M. 2020. Comparative analysis of the circular and highly asymmetrical *Marseilleviridae* genomes. Viruses 12:1270. https://doi.org/10.3390/v12111270
22. Queiroz VF, Carvalho J, de Souza FG, Lima MT, Santos JD, Rocha KLS, de Oliveira DB, Araújo JP, Ullmann LS, Rodrigues RAL, Abrahão JS. 2023. Analysis of the genomic features and evolutionary history of pithovirus-like isolates reveals two major divergent groups of viruses. J Virol 97:e0041123. https://doi.org/10.1128/jvi.00411-23
23. Aylward FO, Moniruzzaman M, Ha AD, Koonin EV. 2021. A phylogenomic framework for charting the diversity and evolution of giant viruses. PLOS Biol 19:e3001430. https://doi.org/10.1371/journal.pbio.3001430
24. Aylward FO, Abrahão JS, Brussaard CPD, Fischer MG, Moniruzzaman M, Ogata H, Suttle CA. 2023. Taxonomic update for giant viruses in the order Imitervirales (phylum *Nucleocytoviricota*). Arch Virol 168:283. https://doi.org/10.1007/s00705-023-05906-3
25. Doutre G, Arfib B, Rochette P, Claverie J-M, Bonin P, Abergel C. 2015. Complete genome sequence of a new member of the *Marseilleviridae* recovered from the brackish submarine spring in the Cassis Port-Miou Calanque, France. Genome Announc 3:e01148-15. https://doi.org/10.1128/genomeA.01148-15
26. Arantes TS, Rodrigues RAL, Dos Santos Silva LK, Oliveira GP, de Souza HL, Khalil JYB, de Oliveira DB, Torres AA, da Silva LL, Colson P, Kroon EG, da Fonseca FG, Bonjardim CA, La Scola B, Abrahão JS. 2016. The large Marseillevirus explores different entry pathways by forming giant infectious vesicles. J Virol 90:5246–5255. https://doi.org/10.1128/JVI.00177-16
27. Greub G, Raoult D. 2004. Microorganisms resistant to free-living amoebae. Clin Microbiol Rev 17:413–433. https://doi.org/10.1128/CMR.17.2.413-433.2004
28. Greub G, Raoult D. 2002. Crescent bodies of *Parachlamydia acanthamoeba* and its life cycle within *Acanthamoeba polyphaga*: an electron micrograph study. Appl Environ Microbiol 68:3076–3084. https://doi.org/10.1128/AEM.68.6.3076-3084.2002
29. Boratto PVM, Oliveira GP, Machado TB, Andrade A, Baudoin J-P, Klose T, Schulz F, Azza S, Decloquement P, Chabrière E, Colson P, Levasseur A, La Scola B, Abrahão JS. 2020. Yaravirus: a novel 80-nm virus infecting *Acanthamoeba castellanii*. Proc Natl Acad Sci U S A 117:16579–16586. https://doi.org/10.1073/pnas.2001637117
30. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, Lescot M, Poirot O, Bertaux L, Bruley C, Couté Y, Rivkina E, Abergel C, Claverie J-M. 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. Proc Natl Acad Sci U S A 111:4274–4279. https://doi.org/10.1073/pnas.1320670111
31. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152
32. Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1

33. Machado TB, de Aquino ILM, Abrahão JS. 2022. Isolation of giant viruses of *Acanthamoeba castellanii*. Curr Protoc 2:e455. https://doi.org/10.1002/cpz1.455

34. Reed LJ, Muench H. 1938. A simple method of estimating fifty per cent endpoints. Am J Epidemiol 27:493–497. https://doi.org/10.1093/oxfordjournals.aje.a118408

35. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

36. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

37. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using spades de novo assembler. Curr Protoc Bioinformatics 70:e102. https://doi.org/10.1002/cpbi.102

38. Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451–W454. https://doi.org/10.1093/nar/gki487

39. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119

40. Yutin N, Wolf YI, Raoult D, Koonin EV. 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol J 6:223. https://doi.org/10.1186/1743-422X-6-223

41. Rodrigues RAL, Queiroz VF, Ghosh J, Dunigan DD, Van Etten JL. 2022. Functional genomic analyses reveal an open pan-genome for the chloroviruses and a potential for genetic innovation in new isolates. J Virol 96:e0136721. https://doi.org/10.1128/JVI.01367-21

42. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704

43. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300

44. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

45. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285

46. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

47. Louca S, Doebeli M. 2018. Efficient comparative phylogenetics on large trees. Bioinformatics 34:1053–1055. https://doi.org/10.1093/bioinformatics/btx701

48. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics 12:124. https://doi.org/10.1186/1471-2105-12-124

49. Howe E, Holton K, Nair S, Schlauch D, Sinha R, Quackenbush J. 2010. MeV: MultiExperiment viewer, p 267–277. In Ochs MF, Casagrande JT, Davuluri RV (ed), Biomedical informatics for cancer research. Springer US, Boston, MA.

50. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9

4.1.2 PROPOSTA TAXONÔMICA – Criação de um novo gênero e inclusão de novas espécies dentro dos gêneros atuais da família Marseilleviridae (ordem Pimascovirales, classe Megaviricetes, filo Nucleocytoviricota)

Essa é uma proposta taxonômica submetida ao Comitê Internacional de Taxonomia de Vírus (ICTV) como parte do protocolo para que as mudanças taxonômicas sejam avaliadas e consideradas oficialmente. Posteriormente, essa proposta será publicada como um artigo.

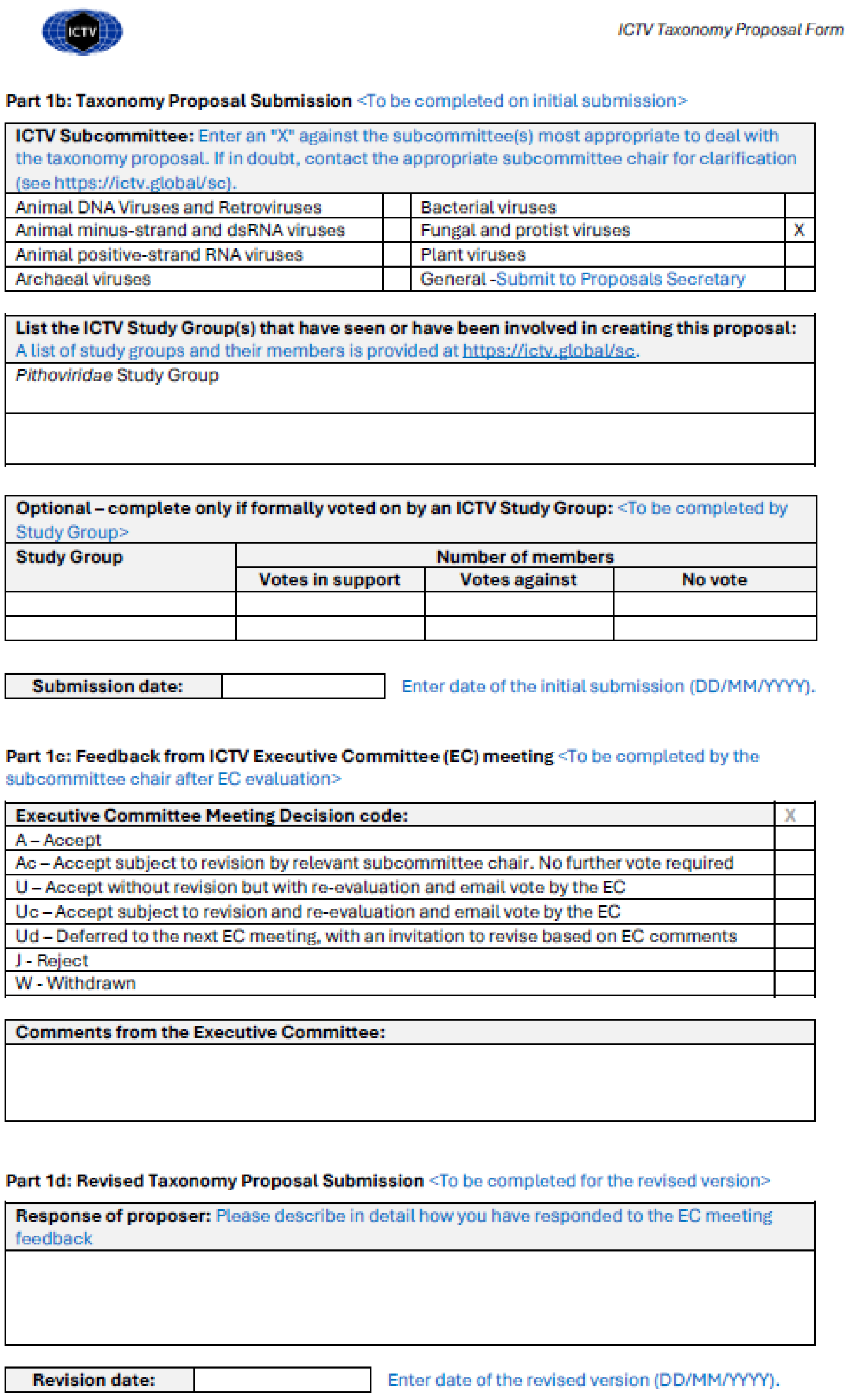
Resumo: As análises realizadas durante a descrição do Marseillevirus cajuinensis abriram as portas para uma reorganização da taxonomia oficial da família *Marseilleviridae*. Para tal é necessário que uma proposta de alteração taxonômica seja feita oficialmente ao ICTV, para que os membros do comitê analisem tal proposta. As mudanças foram sugeridas, pois a taxonomia atual da família *Marseilleviridae*, apesar de recentemente atualizada, ainda não abrange boa parte da diversidade conhecida de linhagens virais da família. Atualmente, a família é composta por dois gêneros. O primeiro gênero, *Marseillevirus*, é composto por duas espécies: *Marseillevirus massiliense* e *Marseillevirus senegalense*. Os membros desse gênero são filogeneticamente relacionados a linhagem A de Marseillevirus. Tradicionalmente, os Marseillevirus isolados vinham sendo classificados não oficialmente em 5 linhagens filogenéticas chamadas de A, B, C, D e E. O segundo gênero atual da família é o *Losannavirus*, que também é composto por duas espécies: *Losannavirus lausannense e Losannavirus tunisense,* que representam as linhagens B e C, respectivamente. Dessa forma, as linhagens D e E não estão representadas na taxonomia atual e outros isolados poderiam ser incluídos nos táxons já existentes. Com base na filogenia da DNA polimerase e nas análises de média de identidade de nucleotídeo e aminoácidos, foi proposta a criação de duas novas espécies dentro do gênero *Marseillevirus*: *Marseillevirus cajuinensis* e *Marseillevirus tokyoensis*. Ambas são compostas por isolados filogeneticamente divergentes dentro da linhagem A de Marseillevírus. Quanto ao gênero *Losannavirus*, foi proposto que ele agrupasse vírus que estejam filogeneticamente associados tanto às linhagens B e C, quanto à linhagem D. Assim, propusemos que duas novas espécies fosse incluídas nesse gênero: *Losannavirus noumeaensis,* que é representada por um membro divergente da linhagem B, e *Losannavirus brasiliensis,* representada pelo isolado brasileiro Brazilian Marseillevirus, fundador da linhagem D. Por fim, para agrupar a linhagem E, foi proposta a criação de um terceiro gênero para a família, chamado *Goldenvírus*, contendo uma única espécies que representa o único isolado dessa linhagem, o brasileiro Golden Marseillevirus.

The International Committee on Taxonomy of Viruses
Taxonomy Proposal Form, 2025

Part 1a: Details of taxonomy proposals

Title:	Create 1 new genus and include new species within the current genera from *Marseilleviridae* family (order *Pimascovirales*, class *Megaviricetes*, phylum *Nucleocytoviricota*)
Code assigned:	<to be assigned by ICTV officers>

Author(s), affiliation and email address(es): One author per row - add additional rows as necessary				
Given name (+middle initial(s))	**Surname**	**Affiliation** Include and limit to Department, Institution, City, Country	**Email address** One only	**Corr. author(s)** Mark with an X
Bruna	Azevedo	Microbiology Department, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil	azvdobruna@gmail.com	
Gabriel	Almeida	The Norwegian College of Fishery Science, The Arctic University of Norway, Trom so, Norway	gabriel.d.almeida@uit.no	
Julien	Andriani	Centre Hospitalier Universitaire de Grenoble, La Tronche, France	JAndreani@chu-grenoble.fr	
Frank	Aylward	Department of Biological Sciences, Virginia Tech, Blacksburg, USA	faylward@vt.edu	
Jean-Michel	Claverie	School of Medicine, Aix-Marseille University, Marseille, France	claverie@igs.cnrs-mrs.fr	
Philippe	Colson	IHU Méditerranée Infection, Aix-Marseille University, Marseille, France	philippe.colson@univ-amu.fr	
Gilbert	Greub	Institute of Microbiology, University of Lausanne, Lausanne, Switzerland	gilbert.greub@chuv.ch	
Victória	Queiroz	Department of Biological Sciences, Virginia Tech, Blacksburg, USA	victoriafq18@gmail.com	
Rodrigo	Rodrigues	Microbiology Department, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil	rodriguesral07@gmail.com	
Masaharu	Takemura	Institute of Arts and Sciences, Tokyo University of Science, Tokyo, Japan	giantvirus@rs.tus.ac.jp	
Jônatas	Abrahão	Microbiology Department, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil	jonatas.abrahao@gmail.com	X



ICTV Taxonomy Proposal Form 2025 v.2

Part 1b: Taxonomy Proposal Submission <To be completed on initial submission>

ICTV Subcommittee: Enter an "X" against the subcommittee(s) most appropriate to deal with the taxonomy proposal. If in doubt, contact the appropriate subcommittee chair for clarification (see https://ictv.global/sc).			
Animal DNA Viruses and Retroviruses		Bacterial viruses	
Animal minus-strand and dsRNA viruses		Fungal and protist viruses	X
Animal positive-strand RNA viruses		Plant viruses	
Archaeal viruses		General -Submit to Proposals Secretary	

List the ICTV Study Group(s) that have seen or have been involved in creating this proposal: A list of study groups and their members is provided at https://ictv.global/sc.
Pithoviridae Study Group

Optional – complete only if formally voted on by an ICTV Study Group: <To be completed by Study Group>			
Study Group	**Number of members**		
	Votes in support	**Votes against**	**No vote**

Submission date:	

Enter date of the initial submission (DD/MM/YYYY).

Part 1c: Feedback from ICTV Executive Committee (EC) meeting <To be completed by the subcommittee chair after EC evaluation>

Executive Committee Meeting Decision code:	X
A – Accept	
Ac – Accept subject to revision by relevant subcommittee chair. No further vote required	
U – Accept without revision but with re-evaluation and email vote by the EC	
Uc – Accept subject to revision and re-evaluation and email vote by the EC	
Ud – Deferred to the next EC meeting, with an invitation to revise based on EC comments	
J - Reject	
W - Withdrawn	

Comments from the Executive Committee:

Part 1d: Revised Taxonomy Proposal Submission <To be completed for the revised version>

Response of proposer: Please describe in detail how you have responded to the EC meeting feedback

Revision date:	

Enter date of the revised version (DD/MM/YYYY).

Part 3: TAXONOMIC PROPOSAL

Please use this section for proposing taxonomic changes. The changes MUST also be presented in the accompanying form, TP_Template_Excel_module_2025.xlsx, available at https://ictv.global/taxonomy/templates.

Taxonomic changes proposed: Enter an "X" against one or more actions			
Establish new taxon	X	Split taxon	
Abolish taxon		Merge taxon	
Move taxon		Promote taxon	
Rename taxon		Demote taxon	
Move and rename			

Etymology (origin) of proposed taxonomic names: Briefly describe the origin and rationale for each proposed taxonomic name in the proposal. A longer explanation can be provided in the main section of the taxonomy proposal if required. Use one row for each taxonomic name; additional rows can be inserted where needed. See examples for existing names in https://ictv.global/taxonomy/etymology.

Taxon name	**Etymology of the term**
Marseillevirus cajuinensis	"Marseillevirus" from the genus and "cajuinensis" in reference to a popular drink (cajuina) from the region where the virus was isolated: Piaui, Brazil.
Marseillevirus tokyoensis	"Marseillevirus" from the genus and "tokyoensis" from Tokyo, the Japanese city where "tokyovirus" was isolated.
Losannavirus brasiliensis	"Losannavirus" from the genus and "brasiliensis" from Brazil, country where 'Brazilian marseillevirus' where isolated. It was the first marseillevirus isolated in Brazil.
Losannavirus noumeaensis	"Losannavirus" from the genus and "noumeaensis" from the place where noumeavirus was isolated: Noumea Airport (New Caledonia, France)
Goldenvirus	"Goldenvirus" is derived from 'golden mussels' that was the type of sample where golden marseillevirus was isolated.
Goldenvirus brasilensis	"Goldenvirus" form the genus and "brasiliensis" from Brazil, country where golden marseillevirus was isolated.

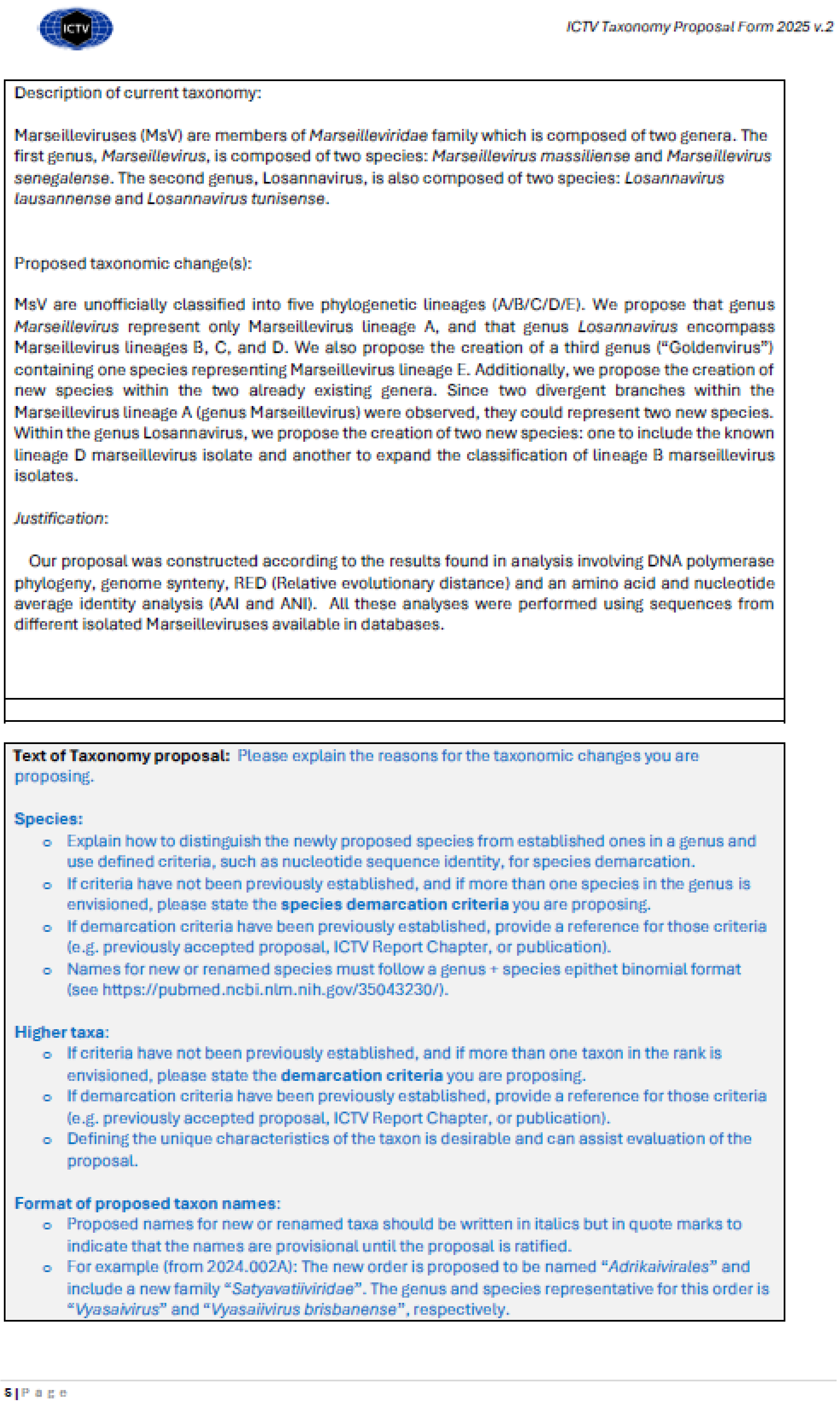
Permission for use of names derived from a living person: If you propose taxon name(s) which are derived from the name of a living person or persons, you must attach documents to verify that permission has been obtained from those persons and that they agree to the form in which their name or an element of it is to be used. These documents should be listed in the Accompanying Files table below. The documents will not be posted online but will be retained by ICTV as evidence of authorization.

Taxon name	**Full name of person from whom the name is derived**	**Attached** X

Abstract of Taxonomy Proposal: Please provide a structured summary of your taxonomic proposal using the headings provided (maximum 250 words). The abstract should provide sufficient context to be understandable without reference to the full text of the proposal as it will be used as written for the automatic population of public documents such as taxonomy summary publications.

Taxonomic rank(s) affected:

The marseilleviruses belong to the family *Marseilleviridae*, a taxon found within the order *Pimascovirales*, phylum *Nucleocytoviricota*, kingdom *Bamfordvirae*, and realm *Varidnaviria*.



ICTV Taxonomy Proposal Form 2025 v.2

Description of current taxonomy:

Marseilleviruses (MsV) are members of *Marseilleviridae* family which is composed of two genera. The first genus, *Marseillevirus*, is composed of two species: *Marseillevirus massiliense* and *Marseillevirus senegalense*. The second genus, Losannavirus, is also composed of two species: *Losannavirus lausannense* and *Losannavirus tunisense*.

Proposed taxonomic change(s):

MsV are unofficially classified into five phylogenetic lineages (A/B/C/D/E). We propose that genus *Marseillevirus* represent only Marseillevirus lineage A, and that genus *Losannavirus* encompass Marseillevirus lineages B, C, and D. We also propose the creation of a third genus ("Goldenvirus") containing one species representing Marseillevirus lineage E. Additionally, we propose the creation of new species within the two already existing genera. Since two divergent branches within the Marseillevirus lineage A (genus Marseillevirus) were observed, they could represent two new species. Within the genus Losannavirus, we propose the creation of two new species: one to include the known lineage D marseillevirus isolate and another to expand the classification of lineage B marseillevirus isolates.

Justification:

Our proposal was constructed according to the results found in analysis involving DNA polymerase phylogeny, genome synteny, RED (Relative evolutionary distance) and an amino acid and nucleotide average identity analysis (AAI and ANI). All these analyses were performed using sequences from different isolated Marseilleviruses available in databases.

Text of Taxonomy proposal: Please explain the reasons for the taxonomic changes you are proposing.

Species:
- Explain how to distinguish the newly proposed species from established ones in a genus and use defined criteria, such as nucleotide sequence identity, for species demarcation.
- If criteria have not been previously established, and if more than one species in the genus is envisioned, please state the **species demarcation criteria** you are proposing.
- If demarcation criteria have been previously established, provide a reference for those criteria (e.g. previously accepted proposal, ICTV Report Chapter, or publication).
- Names for new or renamed species must follow a genus + species epithet binomial format (see https://pubmed.ncbi.nlm.nih.gov/35043230/).

Higher taxa:
- If criteria have not been previously established, and if more than one taxon in the rank is envisioned, please state the **demarcation criteria** you are proposing.
- If demarcation criteria have been previously established, provide a reference for those criteria (e.g. previously accepted proposal, ICTV Report Chapter, or publication).
- Defining the unique characteristics of the taxon is desirable and can assist evaluation of the proposal.

Format of proposed taxon names:
- Proposed names for new or renamed taxa should be written in italics but in quote marks to indicate that the names are provisional until the proposal is ratified.
- For example (from 2024.002A): The new order is proposed to be named "*Adrikaivirales*" and include a new family "*Satyavatiiviridae*". The genus and species representative for this order is "*Vyasaivirus*" and "*Vyasaiivirus brisbanense*", respectively.

5 | P a g e

- Proposed taxonomic names are formatted in italics in the Excel module, but they do not require quote marks.

Availability of genome sequences:

- Proposals for the establishment of new taxa require that annotated and coding complete genome sequences of the species exemplars are available from an INSDC database (eg. GenBank).
- Short read archive (SRA) records (unassembled sequences) are not acceptable.

Taxonomic rank(s) affected:

Marseilleviruses (MsV) are amoeba-infecting viruses found in several different types of samples, such as freshwater, saltwater and animals (1–3). These viruses are members of the family *Marseilleviridae* that is classified within the *Pimascovirales* order, *Nucleocytoviricota* phylum, *Bamfordvirae* kingdom and *Varidnaviria* realm (4).

Description of current taxonomy:

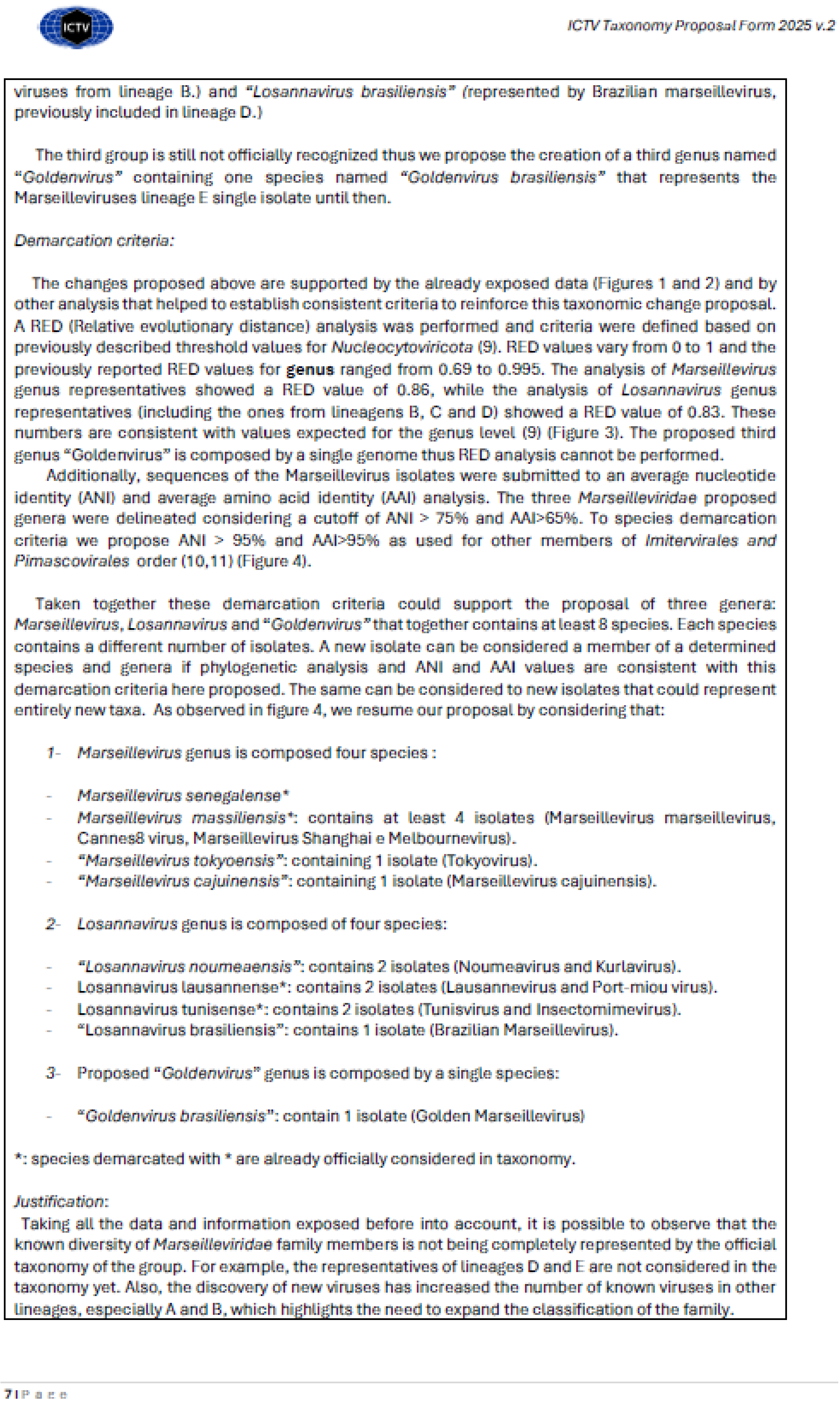
Phylogenetic analyses based on DNA polymerase proteins revealed the existence of five different Marseilleviruses phylogenetic lineages, named A, B, C, D, and E. Lineage A is represented by the first isolate, Marseillevirus marseillevirus and related viruses (1). Lineage B is represented by the second isolate Lausannevirus and related viruses (5). The lineage C is represented by Tunisvirus and Insectomime virus (2,6), while lineages D and E are composed of the Brazilian isolates, Brazilian marseillevirus and Golden marseillevirus, respectively (7,8). About the taxonomic ranks, currently, *the Marseilleviridae* family is composed of two genera. The first genus, *Marseillevirus*, is composed of two species: *Marseillevirus massiliense* and *Marseillevirus senegalense*. *Marseillevirus* genus members are related to lineage A marseilleviruses. The second genus, Losannavirus, is also composed of two species: *Losannavirus lausannense* (lineage B related) and *Losannavirus tunisense* (lineage C related).

Proposed taxonomic change(s):

We performed a genomic and phylogenetic analysis comparing Marseilleviruses from the five phylogenetic lineages that were isolated and had sequences available in databases (GenBank). The DNA polymerase phylogeny suggested the existence of three major consistent Marseilleviruses groups instead of five lineages, as they were classically organized (Figure 1). A genome synteny analysis (Figure 2) reinforces this three-group organization. Together the phylogeny and the genome synteny can help to establish the parameters for a taxonomic organization of *Marseilleviridae* into three genera that together include at least 8 different species.

The first group is already officially classified as the *Marseillevirus* genus and includes two species, as mentioned before. We propose the creation of two new species in this genus that together with the other existing species could represent the Marseilleviruses related to lineage A. Based on DNA polymerase phylogeny (Figure 1) it is possible to observe that within the former lineage A there are at least three divergent branches, one including M. cajuinensis, other including Tokyovirus and a third with several other lineage A isolates, including the one that represents the already existing species *Marseillevirus massiliense* (represented by Marseillevirus marseillevirus). Therefore, we propose the creation of two new species within the Marseillevirus genus named "*Marsellevirus cajuinensis*" (to include M. cajuinensis isolate) and "*Marseillevirus tokyoensis*" (represented by Tokyovirus).

The second group is already officially classified as the *Losannavirus* genus and already includes two species, as mentioned before. We propose that this genus could represent all the Marseilleviruses lineages B, C and D. These three lineages are all included in the same group derived from a more recent common ancestor (Figure 1) and this proposal corroborates with the current classification that already includes lineages B and C in the same genus. Also, we propose that two new species could be included in *Losannavirus* genus named "*Losannavirus noumeaensis*" (represented by Noumeavirus and related



viruses from lineage B.) and *"Losannavirus brasiliensis"* (represented by Brazilian marseillevirus, previously included in lineage D.)

The third group is still not officially recognized thus we propose the creation of a third genus named *"Goldenvirus"* containing one species named *"Goldenvirus brasiliensis"* that represents the Marseilleviruses lineage E single isolate until then.

Demarcation criteria:

The changes proposed above are supported by the already exposed data (Figures 1 and 2) and by other analysis that helped to establish consistent criteria to reinforce this taxonomic change proposal. A RED (Relative evolutionary distance) analysis was performed and criteria were defined based on previously described threshold values for *Nucleocytoviricota* (9). RED values vary from 0 to 1 and the previously reported RED values for **genus** ranged from 0.69 to 0.995. The analysis of *Marseillevirus* genus representatives showed a RED value of 0.86, while the analysis of *Losannavirus* genus representatives (including the ones from lineagens B, C and D) showed a RED value of 0.83. These numbers are consistent with values expected for the genus level (9) (Figure 3). The proposed third genus "Goldenvirus" is composed by a single genome thus RED analysis cannot be performed.

Additionally, sequences of the Marseillevirus isolates were submitted to an average nucleotide identity (ANI) and average amino acid identity (AAI) analysis. The three *Marseilleviridae* proposed genera were delineated considering a cutoff of ANI > 75% and AAI>65%. To species demarcation criteria we propose ANI > 95% and AAI>95% as used for other members of *Imitervirales and Pimascovirales* order (10,11) (Figure 4).

Taken together these demarcation criteria could support the proposal of three genera: *Marseillevirus*, *Losannavirus* and *"Goldenvirus"* that together contains at least 8 species. Each species contains a different number of isolates. A new isolate can be considered a member of a determined species and genera if phylogenetic analysis and ANI and AAI values are consistent with this demarcation criteria here proposed. The same can be considered to new isolates that could represent entirely new taxa. As observed in figure 4, we resume our proposal by considering that:

1- *Marseillevirus* genus is composed four species :

- *Marseillevirus senegalense**
- *Marseillevirus massiliensis**: contains at least 4 isolates (Marseillevirus marseillevirus, Cannes8 virus, Marseillevirus Shanghai e Melbournevirus).
- *"Marseillevirus tokyoensis"*: containing 1 isolate (Tokyovirus).
- *"Marseillevirus cajuinensis"*: containing 1 isolate (Marseillevirus cajuinensis).

2- *Losannavirus* genus is composed of four species:

- *"Losannavirus noumeaensis"*: contains 2 isolates (Noumeavirus and Kurlavirus).
- Losannavirus lausannense*: contains 2 isolates (Lausannevirus and Port-miou virus).
- Losannavirus tunisense*: contains 2 isolates (Tunisvirus and Insectomimevirus).
- "Losannavirus brasiliensis": contains 1 isolate (Brazilian Marseillevirus).

3- Proposed "*Goldenvirus*" genus is composed by a single species:

- *"Goldenvirus brasiliensis"*: contain 1 isolate (Golden Marseillevirus)

*: species demarcated with * are already officially considered in taxonomy.

Justification:

Taking all the data and information exposed before into account, it is possible to observe that the known diversity of *Marseilleviridae* family members is not being completely represented by the official taxonomy of the group. For example, the representatives of lineages D and E are not considered in the taxonomy yet. Also, the discovery of new viruses has increased the number of known viruses in other lineages, especially A and B, which highlights the need to expand the classification of the family.

ng

References:

1. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc Natl Acad Sci U S A. 2009 Dec 22;106(51):21848–53.

2. Boughalmi M, Pagnier I, Aherfi S, Colson P, Raoult D, La Scola B. First isolation of a Marseillevirus in the Diptera Syrphidae Eristalis tenax. Intervirology. 2013;56(6):386–94.

3. The genomic and phylogenetic analysis of Marseillevirus cajuinensis raises questions about the evolution of Marseilleviridae lineages and their taxonomical organization - PubMed [Internet]. [cited 2025 Apr 18]. Available from: https://pubmed.ncbi.nlm.nih.gov/38752754/

4. Current ICTV Taxonomy Release | ICTV [Internet]. [cited 2022 Oct 1]. Available from: https://ictv.global/taxonomy

5. Thomas V, Bertelli C, Collyn F, Casson N, Telenti A, Goesmann A, et al. Lausannevirus, a giant amoebal virus encoding histone doublets. Environmental Microbiology. 2011;13(6):1454–66.

6. Aherfi S, Boughalmi M, Pagnier I, Fournous G, La Scola B, Raoult D, et al. Complete genome sequence of Tunisvirus, a new member of the proposed family Marseilleviridae. Arch Virol. 2014 Sep;159(9):2349–58.

7. Dornas F, Assis F, Aherfi S, Arantes T, Abrahão J, Colson P, et al. A Brazilian Marseillevirus Is the Founding Member of a Lineage in Family Marseilleviridae. Viruses. 2016 Mar 10;8(3):76.

8. Dos Santos RN, Campos FS, Medeiros de Albuquerque NR, Finoketti F, Côrrea RA, Cano-Ortiz L, et al. A new marseillevirus isolated in Southern Brazil from Limnoperna fortunei. Sci Rep. 2016 Oct 14;6:35237.

9. Aylward FO, Moniruzzaman M, Ha AD, Koonin EV. A phylogenomic framework for charting the diversity and evolution of giant viruses. PLOS Biology. 2021 Oct 27;19(10):e3001430.

10. Queiroz VF, Rodrigues RAL, Abrahão JS. A taxonomic proposal for cedratviruses, orpheoviruses, and pithoviruses. Arch Virol. 2024 Jun 1;169(6):132.

11. Aylward FO, Abrahão JS, Brussaard CPD, Fischer MG, Moniruzzaman M, Ogata H, et al. Taxonomic update for giant viruses in the order Imitervirales (phylum Nucleocytoviricota). Arch Virol. 2023 Oct 31;168(11):283.

Accompanying files: List the names of supplementary files associated with the taxonomy proposal and description of their contents; add additional rows if required. Accompanying files may include letters of permission for use of personal names for taxa, data tables, distance matrices, or larger figures such as phylogenetic trees *etc.* Acceptable file formats: docx, xlsx, PDF.

Filename	Description of contents

Tables, Figures: The use of Figures and Tables as supporting evidence is strongly recommended (note that using material from publications will require permission from the copyright holder if the publication is not open access).

For phylogenetic analysis, please provide a tree based on aligned nucleotide or translated amino acid sequences with branch lengths proportional to genetic distance. Trees should be generated using an appropriate algorithm (*eg.* neighbor-joining, maximum likelihood, or maximum clade credibility); the use of UPGMA is discouraged. Provide evidence of the reliability of the branching (*eg.* by boot-strapping or posterior probabilities). Where possible, trees should be rooted using an outgroup sequence.

Include INSDC (*eg.* GenBank) nucleotide accession numbers in the sequence labels. Annotate trees with existing and proposed taxa.

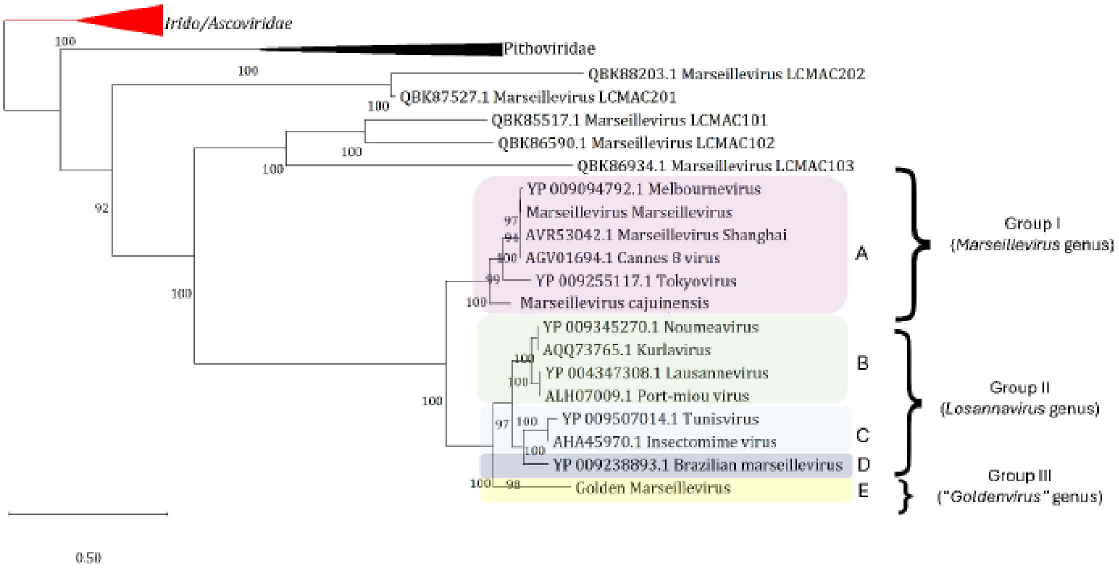


Figure 1 –Phylogenetic tree based on DNA polymerase amino acid sequences of Marseilleviridae isolates available at GenBank. The letters A, B, C, D and E represent the classical five Marseilleviruses phylogenetic lineages while “group I”, “group II” and “group III” represent the proposed organization of Marseilleviruses in three major groups that could represent three genera (indicate in parenthesis). The tree was built using the maximum likelihood method, with statistical support based on 1,000 replicates (bootstrap). The best model, selected by IQtree (ModelFinder), for the tree was VT + F + I + G4. The tree is rooted on *Iridoviridae* branch as an outgroup. The tree scale bars represent the number of amino acid substitutions per site.

Figure 2 – Genome synteny analysis with representatives of the five phylogenetic lineages of Marseilleviruses, labeled by the letters from A to E. Each line represents the sequence of a different virus, which is identified in the legend on the colored boxes at the left. The organization of the sequences in the three major groups (I, II and III) proposed as genera in *Marseilleviridae* family is indicated on the left. Blocks of the same color indicate similar regions between sequences. The areas without any colored blocks represent regions exclusive to that virus, that is, which do not show similarity with other viruses used in the analysis. Note: As Marseilleviruses genomes have circular topology, the sequences were adjusted to start from the MCP aiming to facilitate interpretation of this figure. Marseillevirus marseillevirus was used as the reference genome.

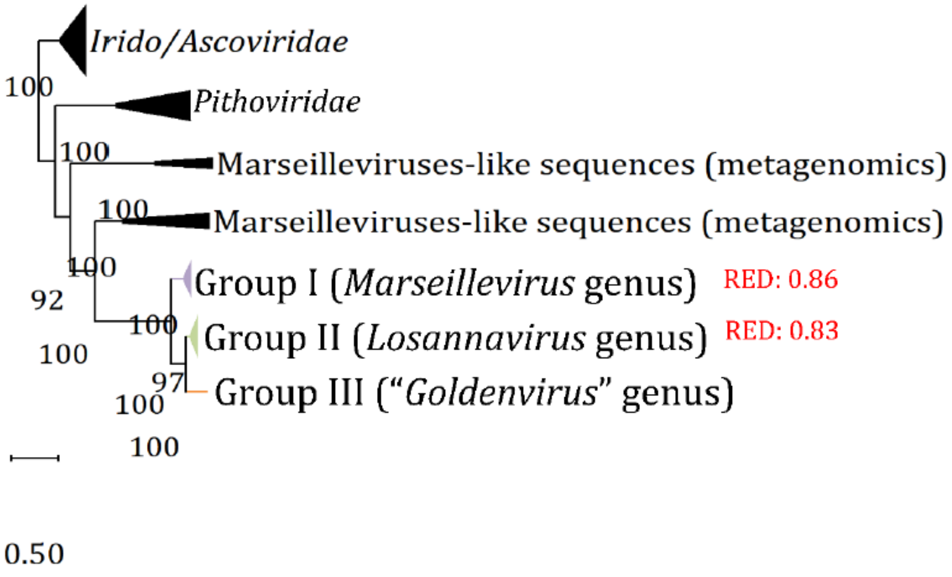


Figure 3 – DNA polymerase phylogenetic tree illustrating the three-genera proposal for Marseilleviridae family. RED (Relative Evolutionary Distance) values are indicated for groups I and II, and the numbers are consistent with the genus level for taxonomy. Group III does not have a RED value because there is only one virus in lineage E. The sequences used in this tree are the same used in figure 1. The tree was built using the maximum likelihood method, with statistical support based on 1,000 replicates (bootstrap). The best model, selected by IQtree (ModelFinder), for the tree was VT + F + I + G4. The tree was rooted on Iridoviridae branch as an outgroup. The tree scale bars represent the number of amino acid substitutions per site.

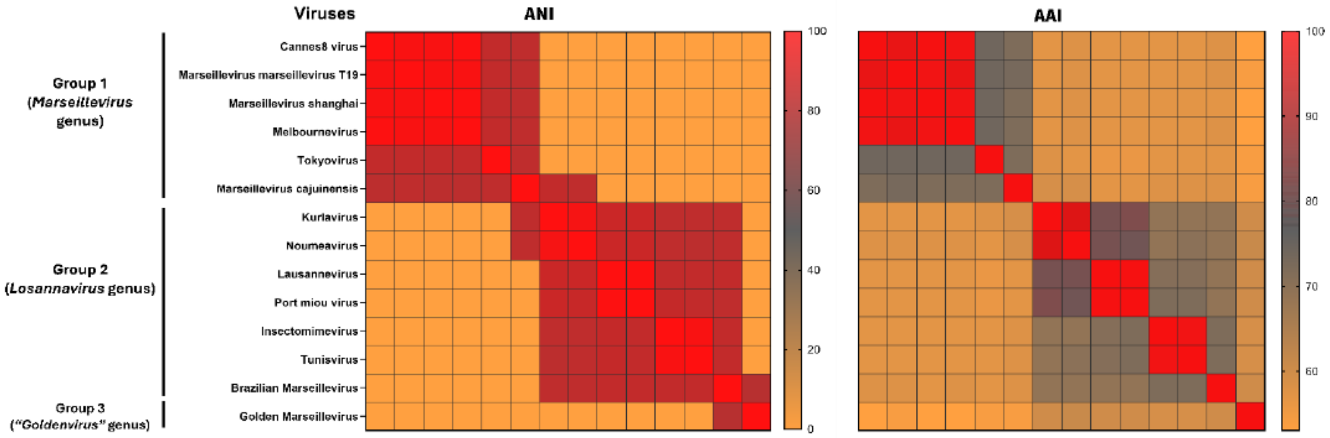


Figure 4 - Average nucleotide identity (ANI) and average amino acid identity (AAI) analysis of *Marseilleviridae* family isolates. Fourteen Marseilleviruses are grouped based on a similarity matrix composed by ANI (left heatmap) and AAI (right heatmap). The three viral groups and their corresponding genus are indicated. ANI <75% was set to zero. ANI values ranged from 77 to 100, and AAI values ranged from 53 to 100.

4. 2 – PARTE 2: ISOLAMENTO E CARACTERIZAÇÃO DE UM NOVO ISOLADO DE VIRÓFAGO PROPOSTO COMO NOVA ESPÉCIE DO GÊNERO *SPUTNIKVIRUS.*

A parte dois deste trabalho descreve o isolamento e caracterização genômica e filogenética do virófago Pantanal, associado a um moumouvírus e a um transpoviron. O virófago aqui descrito faz parte da família *Sputniviroviridae,* que está incluída no filo *Preplasmiviricota,* fazendo parte, portanto, do reino *Bamfordvirae*.

4.2.1 MANUSCRITO: Pantanal Virophage: a divergent lineage of virophage associated with a moumouvirus and a transpoviron suggests an expansion of Sputnikvirus genus

Este manuscrito está em processo de submissão ao periódico **Journal of Virology**.
Obs.: As figuras e suas legendas estão listadas em sequência, ao final do manuscrito, após a descrição dos resultados.

Resumo: Este artigo descreve a descoberta e caracterização de um novo virófago, chamado de virófago Pantanal (*Pantanal virophage)*, isolado a partir de amostras do bioma Pantanal, no Brasil. O vírus foi obtido em associação com um *moumouvirus* (denominado *Moumouvirus pantanalensis*) e um transpoviron, formando um sistema tripartite que pela primeira vez foi obtido de uma mesma amostra para a linhagem dos moumouvírus. A presença simultânea desses três elementos sugere uma complexa rede de interações entre vírus gigantes, virófagos e elementos genéticos móveis. A caracterização morfológica, genômica e filogenética teve foco no virófago, devido as suas características mais divergentes. O virófago possui um genoma que codifica com 21 ORFs, incluindo genes conservados e três ORFans, ou seja, genes cuja função ainda não é conhecida. A análise filogenética posiciona o virófago Pantanal como um ramo divergente dentro do gênero *Sputnikvirus*. A análise de média de identidade de nucleotídeos e de aminoácidos confirmou essa distinção, com valores abaixo do limite de 95% usados como parâmetro para demarcar de espécies virais, indicando que o isolado se trata de uma nova espécie de Sputnikvirus. Este trabalho amplia o conhecimento sobre a diversidade e evolução de virófagos, contribuindo para o avanço e atualização da taxonomia viral e destaca a importância de investigar ecossistemas como o bioma pantanal que ainda são pouco explorados como fonte de novos vírus gigantes e de seus elementos associados.

Pantanal Virophage: a divergent lineage of virophage associated with a moumouvirus and a transpoviron suggests an expansion of *Sputnikvirus* genus

Bruna Luiza de Azevedo[a], Nidia Esther Colquehuanca Arias [a], Talita Bastos Machado[a], João Pessoa Araújo Júnior[b], Leila Sabrina Ullmann[b], Luiz Carlos Junior Alcantara[c] Rodrigo Araújo Lima Rodrigues[a] , Jônatas Santos Abrahão[a] #.

Author information:

[a] Laboratório de Vírus, Departamento de Microbiologia, Universidade Federal de Minas Gerais (UFMG). Belo Horizonte, Minas Gerais, 31270-901, Brazil.

[b] Laboratório de Virologia, Departamento de Microbiologia e Imunologia, Instituto de Biotecnologia, Universidade Estadual Paulista (Unesp), Alameda das Tecomarias s/n, Chácara Capão Bonito, Botucatu, SP 18607-440, Brazil.

[C] Instituto Rene Rachou, Fundação Oswaldo Cruz and Expanded Navigation for Intensive and Optimized Surveillance (NAVIO) Network, Belo Horizonte, 30190-002, Brazil

Running Head: Pantanal virophage expands Sputnikvirus diversity

#Address correspondence to Jônatas S. Abrahão, jonatas.abrahao@gmail.com

Abstract:

Amoeba-infecting giant viruses, such as mimiviruses, challenged paradigms of virology since their discovery. Their extensive and complex genomes and particle structures can interact with virophages and transpovirons revealing unique aspects of the viral ecology and evolution. Here, we describe the isolation of Pantanal virophage, a new virophage obtained from Brazilian pantanal biome. The virophage was found in association with Moumouvirus pantanalensis and a transpoviron sequence. Sequencing and genomic analysis revealed that Pantanal virophage presents a 17,964 bp genome with 21 coding sequences, including three ORFAns that code for proteins with completely unknown functions. Phylogeny showed that Pantanal virophage composes a divergent branch within virophages from *Sputnikvirus* genus. Average nucleotide and amino acid identity analysis comparing different virophage sequences indicated that Pantanal virophage could represent a new *Sputnikvirus* species. These findings shed light into the giant virus-virophage-transpoviron triplet diversity, evolution and ecology, contributing to a future update in virophages taxonomy.

Importance:

Studies on prospecting, isolation and characterization of new amoeba viruses are important to provide new information about biology, diversity, evolution, ecology and taxonomy of these viruses. This work reinforces this importance since we describe Pantanal virophage, a new species of Sputnikvirus found in association with a moumouvirus and a transpoviron. The characterization of Pantanal virophage provided new data and observations regarding the phylogeny and taxonomy of *Sputnikvirus* genus evidencing the need for constant updates in taxonomic classification. Taxonomy is a dynamic science that develops as new methods and data emerge or are improved. Thus, this work shows that the efforts for isolation of new amoeba viruses and their characterization can contribute to enriching the knowledge about taxonomy and evolutionary dynamics of these viruses and of their parasitic-associated elements.

Keywords: virophage, Pantanal, moumouviruses, transpoviron, species, taxonomy, Sputnikvirus

Introduction

Giant viruses of amoeba have surprised the scientific community since the description of their first representative, the mimivirus, in 2003 (1). With particles that exceed the dimensions (750 nm) of some bacterial cells and double-stranded DNA genomes with more than 1.2 million base pairs, mimiviruses represent a milestone in the history of modern virology and they opened the door to the discovery of new amoeba-infecting viruses. In the last years, it was revealed the great diversity of mimiviruses. Currently, *Mimiviridae* family is composed of different subfamilies among them the *Megamimivirinae*, which includes genera such as *Mimivirus*, *Moumouvirus*, *Megavirus* and *Tupanvirus* (2). Considering the high structural and genomic complexity of megamimiviruses, subsequent studies have shown that they have atypical relationships with other organisms, including the possibility of being parasitized by satellite viruses, known as virophages (3). These entities parasitize the viral factories of mimiviruses and are only able to form their progeny in the amoeba cytoplasm during co-infection with mimiviruses.

Different virophages have been described in recent years, capable of parasitizing not only mimiviruses but also megaviruses, moumouviruses, tupanviruses, and Cafeteria roenbergensis virus (4–7) . The impact of virophage parasitism on the morphogenesis of giant viruses is variable. Some virophages strongly affect the formation of their associated giant virus progeny, while others appear not to interfere with their development (3, 4, 8, 9). The interaction between virophages and their viral hosts is complex and can also be related to specific genetic mobile elements known as transpovirons (10–12). Transpovirons are episomal and linear sequences with around 7kb that code their own proteins. It is proposed that transpovirons use the DNA replication machinery of the giant virus and that they can use the virophage to disseminate and to increase their gene expression without affecting the productivity and the gene expression patterns from the virophage or from the giant virus (11, 13).

Besides the virophage isolates, thousands of virophage sequences have been detected through metagenomics, including high-quality metagenome-assembled genomes (MAGs), revealing great diversity (14–18). As a result, a recent taxonomic proposal submitted to the ICTV established a restructuring in the taxonomy and nomenclature of virophage-related taxa. In this context, species related to Sputnik- and Zamilon-type virophages were allocated to the family *Sputniviroviridae*, genus *Sputnikvirus*, with two species: *Sputnikvirus mimiviri* and *Sputnikvirus zamilonense*, respectively (2). These two virus groups, along with mavirusеs, are linked to the initial studies on virophages. Because some of their representatives have been isolated, they have enabled fundamental studies into the nature of the virophage–giant virus relationship and allowed for structural analyses of the particles.

Here, we present the discovery of a new virophage of the genus *Sputnikvirus*, named Pantanal virophage. This virus was isolated in association with a moumouvirus and a transpoviron, suggesting a complex interplay among multiple entities within the same host. The Pantanal virophage exhibits intermediate genomic features that position it between previously described virophages of the Sputnik and Zamilon types. These findings not only highlight the continuum of diversity within the *Sputnikvirus* genus but also provide further evidence of the evolutionary plasticity and ecological versatility of virophages in amoebal environments.

Results

A moumouvirus-virophage-transpoviron tripartite association obtained from Brazilian Pantanal samples.

From the samples collected from Pantanal biome, in Brazil, we isolated an amoeba-infecting giant virus, using *Acanthamoeba castellanii* as host platform. The isolation was confirmed through transmission electronic microscopy (TEM) images that showed the presence of mimivirus-like particles (Figure 1 A). The isolated virus presents replication cycle characteristics corresponding to those observed for other mimiviruses, such as the viral factory morphology (Figure 1 B). Also, we observed the presence of smaller particles that were suggestive to be virophage particles (Figure 1 C, black arrows). These smaller particles were often found inside vesicles in the final stages of the cycle. In Figure 1 C, it is possible to observe a mimivirus-like defective particle (red arrow) inside the same vesicle where virophage particles are observed.

After sequencing and genome assembly, scaffolds representing the giant virus genome were obtained. Compared to the database (GenBank), these scaffolds had as 'best hit' the Moumouvirus maliensis genome (>90% of nucleotide identity). This indicates that the isolate can be grouped into the *Mimiviridae* family, the *Megamimivirinae* subfamily and to the *Moumouvirus* genus. Then, the new isolate was called Moumouvirus pantanalensis. An additional parameter that corroborates with the genomic identification of the isolate as a moumouvirus is its pattern of fibril organization in the particle. Moumouviruses are known to have fewer fibrils that are not uniformly distributed (19), such as is observed in Figure 1 A.

Also, a scaffold containing 6,711 nt were identified and it had 99,40% of nucleotide identity with a transpoviron associated with moumouviruses. We analyzed the transpoviron sequence performing genes prediction and functional annotation. The moumouvirus pantanalensis-associated transpoviron presented 8 ORFs that code for proteins that vary from 24 to 1,028 amino acids (Figure 2 A). ORFs 1 to 4, code hypothetical proteins whose functions were not defined yet. ORF 7 codes for the biggest protein (1,028 amino acids) which is predicted to be a DNA helicase. The DNA helicase is a conserved protein among the known transpovirons and can be used to phylogeny analysis. In transpovirons phylogeny, the known DNA helicase protein sequences form three different lineages that correspond to the giant viral lineage to which they are related (*Mimivirus* genus, *Moumouvirus* genus or *Megavirus* genus). (Figure 2 B).

Another scaffold containing 17,964 nucleotides (nt) were obtained after the sequencing and it presented a high copy number, indicated by the sequencing coverage of 1,926. This sequence presents 86,12% of nucleotide identity with a sputnik virophage, confirming the isolation of a virophage in the sample, which we named Pantanal Virophage.

Pantanal virophage composes a divergent branch within the sputnik virophages phylogeny

In contrast with the moumouvirus and the transpoviron obtained sequences, the Pantanal virophage scaffolds presented a lower nucleotide identity (86,12%) with its best hit from GenBank. Thus, we decided to focus on the genomic and phylogenetic characterization of the isolated virophage. Pantanal virophage genome sequence presents a G-C content of 28,45% and 21 ORFs which code for proteins that vary from

62 to 775 amino acids in size. By searching for similar sequences in databases we were able to identify conserved proteins among the virophages such as the major capsid protein (MCP), the minor virion protein (mCP), and the DNA packaging ATPase (Figure 3 A). On the other hand, 3 ORFs code for proteins to which there are no similar sequences in databases, the so called ORFans. Thus, Pantanal virophage are predicted to encoded 3 completely new proteins, which represents 14% of its proteome (Figure 3 A and B).

When searching for similar sequences in GenBank database, it is possible to observe that the Pantanal virophage proteins matched with sequences from different virophages. The most part (34%, n=7) of Pantanal virophage proteins showed a bigger identity with sputnik virophage 1. Four proteins (19%, n= 4) had as 'best hit' a Guarani virophage protein sequence, while three proteins (14%, n=3) presented a bigger amino acid identity with Zamilon virophage proteins. The remaining proteins had as best hits giant virus sequences (moumouvirus and mimivirus) (Figure 3 B).

To better understand the relationship between Pantanal virophage and the other virophage we performed phylogenetic analysis. MCP phylogeny revealed that Pantanal virophage is included within the sputnik and zamilon virophages clade (Figure 4). Also, it is possible to observe that Pantanal Virophage forms a divergent branch within the virophages belonging to *Sputnikvirus* genus. The same topology is observed when analyzing the other conserved proteins (Supplementary Figures 1 and 2).

Additionally, a genome synteny was performed, comparing complete genomic sequences from different virophages from Sputnikvirus genus with Pantanal virophage (Figure 5). In this analysis it is possible to observe that the Pantanal virophage presents at least four genomic regions that do not show similarity with the other viruses used in the analysis (Figure 5, red arrows). This data corroborates with phylogenies that show Pantanal virophage as a divergent sequence in this group.

Nucleotide and Amino Acid Identity analysis reveal Pantanal virophage as a new virophage species

The sequences of *Sputnikvirus* genus isolates were submitted to Average Nucleotide Identity (ANI) and Average Amino Acid Identity (AAI) analysis. In both ANI and AAI analysis it is possible to delineate three major groups whitin *Sputnikvirus* genus group, based on a cutoff of ANI and AAI >95% (Figure 6). This ANI cutoff was previously considered as a parameter to define viral species in *Nucleocytoviricota* phylum (20, 21). The first group represents *Sputnikvirus mimiviri* species and its isolates with ANI values that vary from 98% to 100% and AAI values ranging from 97,29% to 100% (Figure 6). Guarani virophage is the sequence with a smaller ANI and AAI numbers when compared with the other isolates from *Sputnik mimiviri* species. The second group represents *Sputnikvirus zamilonense* species. The results show ANI <83% and AAI < 62% corroborating with its classification as a different species (Figure 6). The third group corresponds to Pantanal virophage sequence that when compared to the other sequences presented ANI values ranging from 80 to 84% and AAI values ranging from 61,81% to 70,68%. These data corroborate with our genomic and phylogenetic analyses, revealing Pantanal virophage as new species whitin the *Sputnikvirus* genus.

Discussion

The findings and analysis described here represent the isolation of a tripartite system involving Moumouvirus pantanalensis, its associated transpoviron, and the Pantanal virophage. This kind of relationship was already described before for *Mimivirus* genus (10, 12) and *Megavirus* genus (4, 11) members. For *Moumouvirus* genus it was described the detection of associated transpovirons sequences, but not simultaneously with virophages (10, 11). Also, it was described that zamilon virophage is able to infect moumouviruses, but they were not isolated from the same sample (4). This work describes, to our knowledge, the first report of simultaneous isolation of this tripartite system from the same sample involving moumouviruses. Also, it is the first report of this interaction in Pantanal biome which adds new information to the ecology and diversity of giant viruses and virophages.

The isolation of Moumouvirus pantanalensis was confirmed both morphologically and genomically. The particle presents fewer fibrils with a non-uniform distribution corroborating with the fibril pattern described for moumouviruses previously (19). Genomic comparisons further support this classification, with >90% nucleotide identity, placing the virus within the Moumouvirus group. The identification of a closely related transpoviron, sharing >99% identity with known Moumouvirus-associated transpovirons and the phylogeny based on the conserved helicase gene of the transpoviron reinforce the established three-lineage model corresponding to *Mimivirus*, *Moumouvirus*, and *Megavirus* genera, supporting the hypothesis that transpovirons are co-evolving with their associated viral lineages (11, 12).

The major contribution of this study is the characterization of the Pantanal virophage. The description of three ORFans may reflect adaptation to a unique ecological niche or host interaction pattern, considering the different location where it was isolated and the tripartite system with which the virophage is related. These unknown proteins could play key roles in the biology of the virophage which could warrant the need for an in-depth functional characterization. Also, the presence of protein-coding genes showing highest similarity to multiple virophage (including Sputnik, Guarani, and Zamilon) and giant viruses lineages reflects the mosaicism of Pantanal virophage genome, as it is observed for virophages in general (3). This kind of genetic mosaicism can be associated to horizontal gene transfer (HGT) possibly facilitated by co-infection scenarios due to the amoebas phagocytic behavior or/and to the presence of genetic mobile elements like the transpoviron (10, 22–24).

Although it clusters within the *Sputnikvirus* genus, Pantanal virophage exhibits several features that distinguish it from previously described members. The relatively low nucleotide identity (86.12%) with its closest relative, along with divergent phylogenetic placement and unique genomic regions lacking synteny with other Sputnikvirus genomes, suggest that Pantanal virophage constitutes a new group of virophages. These findings are reinforced by ANI and AAI analyses, which show that Pantanal virophage falls below the 95% identity threshold previously used to demarcate viral species (20, 21), suggesting that it can be classified as a new species whitin the *Sputnikvirus* genus.

Taking together, our results extend the known biodiversity of virophages and their associated elements and underscore the importance of further exploring underrepresented ecosystems for novel giant viruses' discovery. The identification of Pantanal virophage as a new species within *Sputnikvirus* genus not only expands the known phylogenetic and taxonomic diversity of this group but also raises important questions about virophage-host coevolution, adaptation, and ecological impact in the ecosystems.

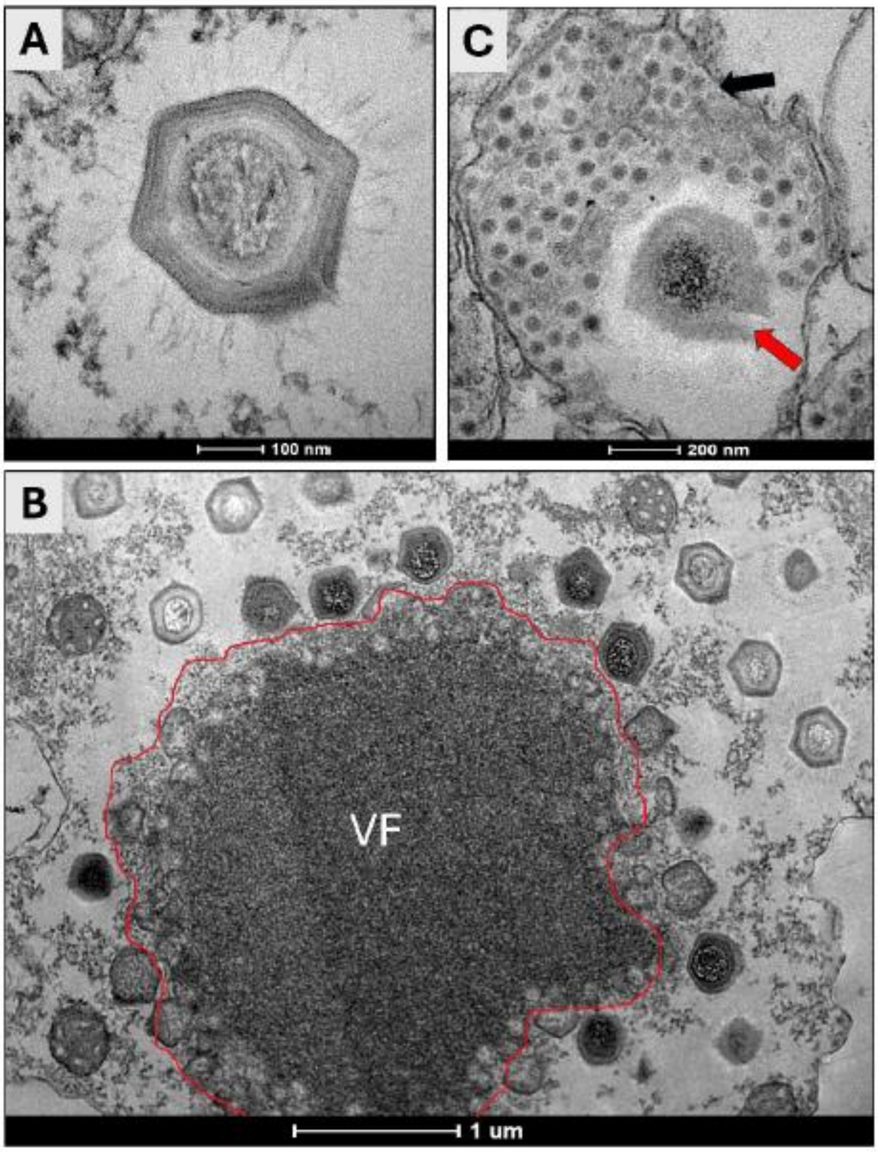


Figure 1 – Morphological aspects of the viral isolates analyzed through transmission electronic microscopy (TEM) images. (**A**) A mimivirus-like with a fibril distribution pattern that corresponds to a moumouvirus particle. (**B**) Virophage-like particles (black arrows) inside a vesicle. A moumouvirus defective particle is also observed inside the same vesicle (red arrow). (**C**) The isolated moumouvirus viral factory (VF), highlighted in red, surrounded by newly assembled particles.

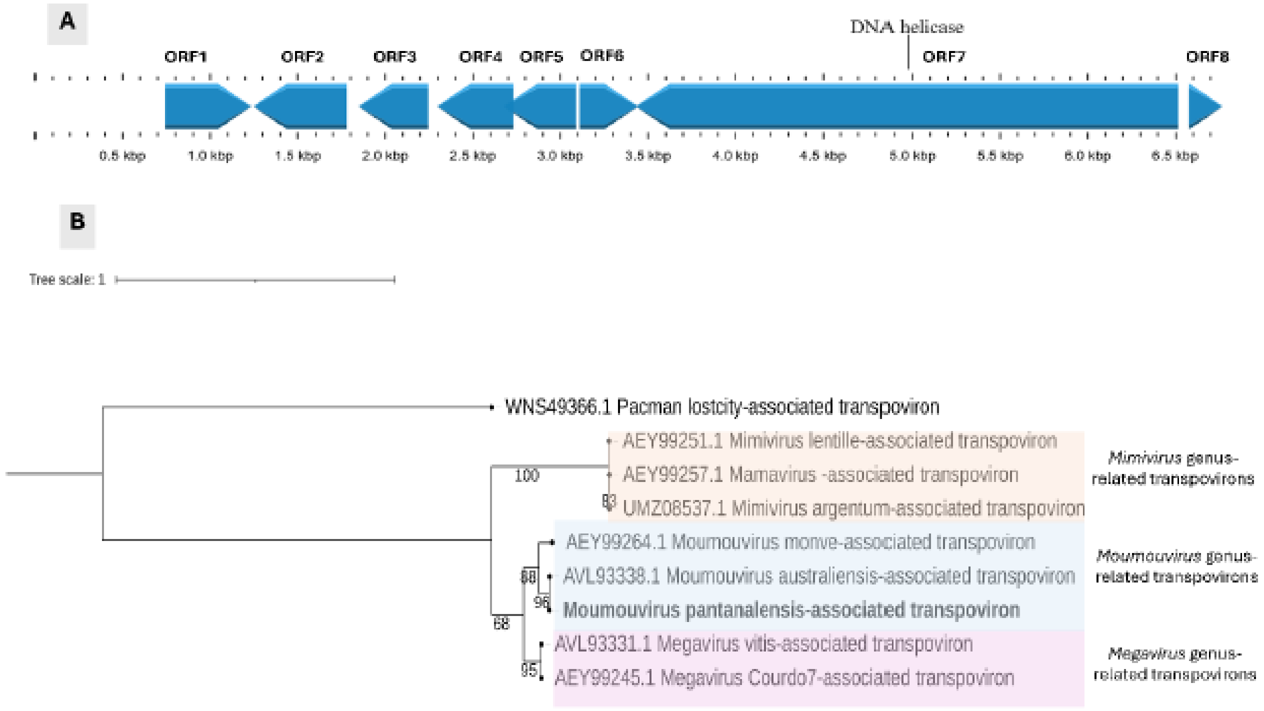


Figure 2 – Genomic and phylogenetic characteristics of the Moumouvirus pantanalensis-associated transpoviron. **(A)** A genome map showing its linear topology and the 8 encoded ORFs. ORF 7 represents DNA helicase protein, which is conserved among the known transpovirons. **(B)** Maximum likelihood phylogenetic tree using the DNA helicase amino acid sequence from known transpovirons sequences. The Megamimiviruses-associated transpovirons are highlighted by different colors and the respective genera are indicated. Moumouvirus pantanalensis-associated transpoviron are highlighted in bold. The tree was built with statistical support based on 1000 replicates (bootstrap). The tree was rooted on the pacman lostcity-associated transpoviron branch as an outgroup. The best model, selected by IQtree (ModelFinder), for this tree was WAG+F+G4. The tree scale bar indicates the number of amino acid substitutions per site.

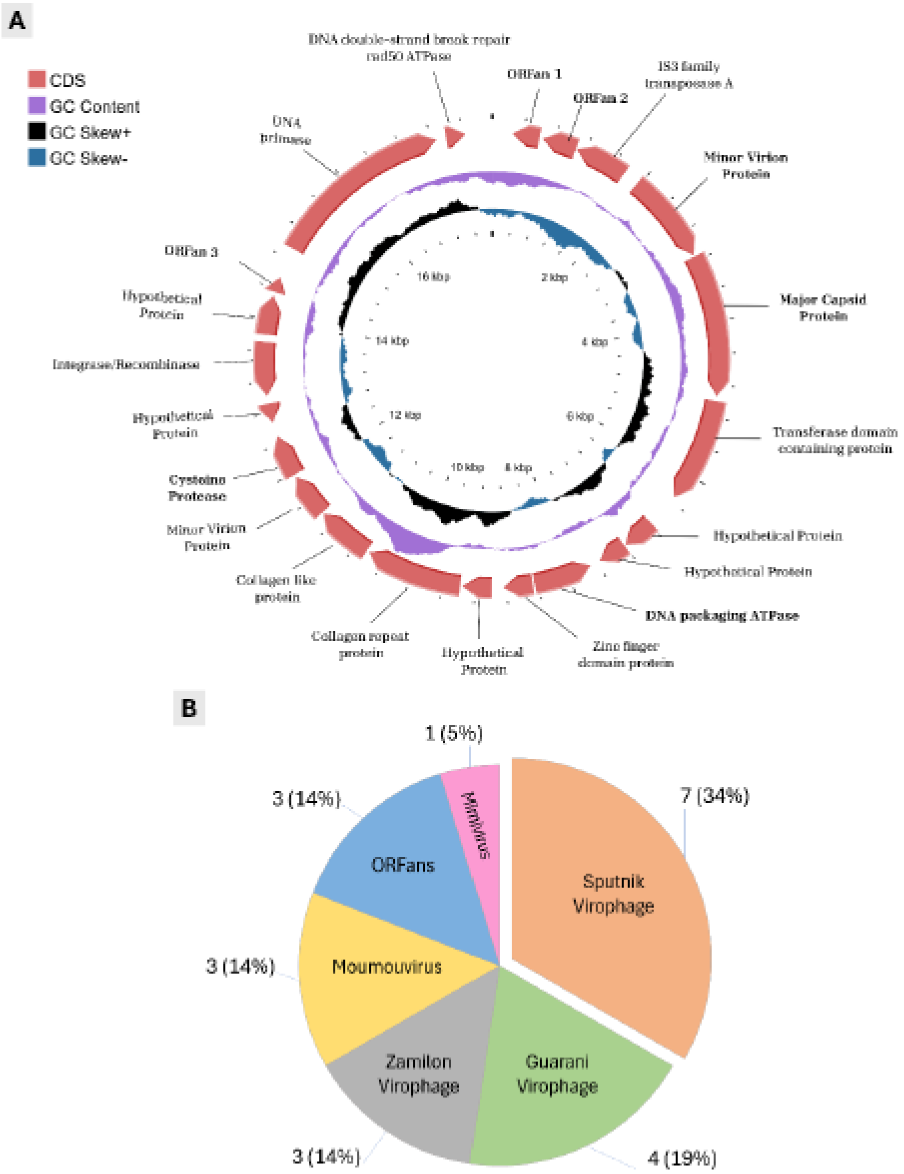


Figure 3– Genomic characterization of Pantanal virophage. (**A**) Genome map showing the circular topology of the Pantanal virophage genome and the distribution of Coding Sequences (CDSs) throughout the genome. The CDSs, the G-C content and the G-C skew are illustrated by different ring colors, as indicated in the color legend at top left. (**B**) Proportion of Pantanal virophage proteins and their corresponding best hits among the different virophage and giant virus sequences available in GenBank database.

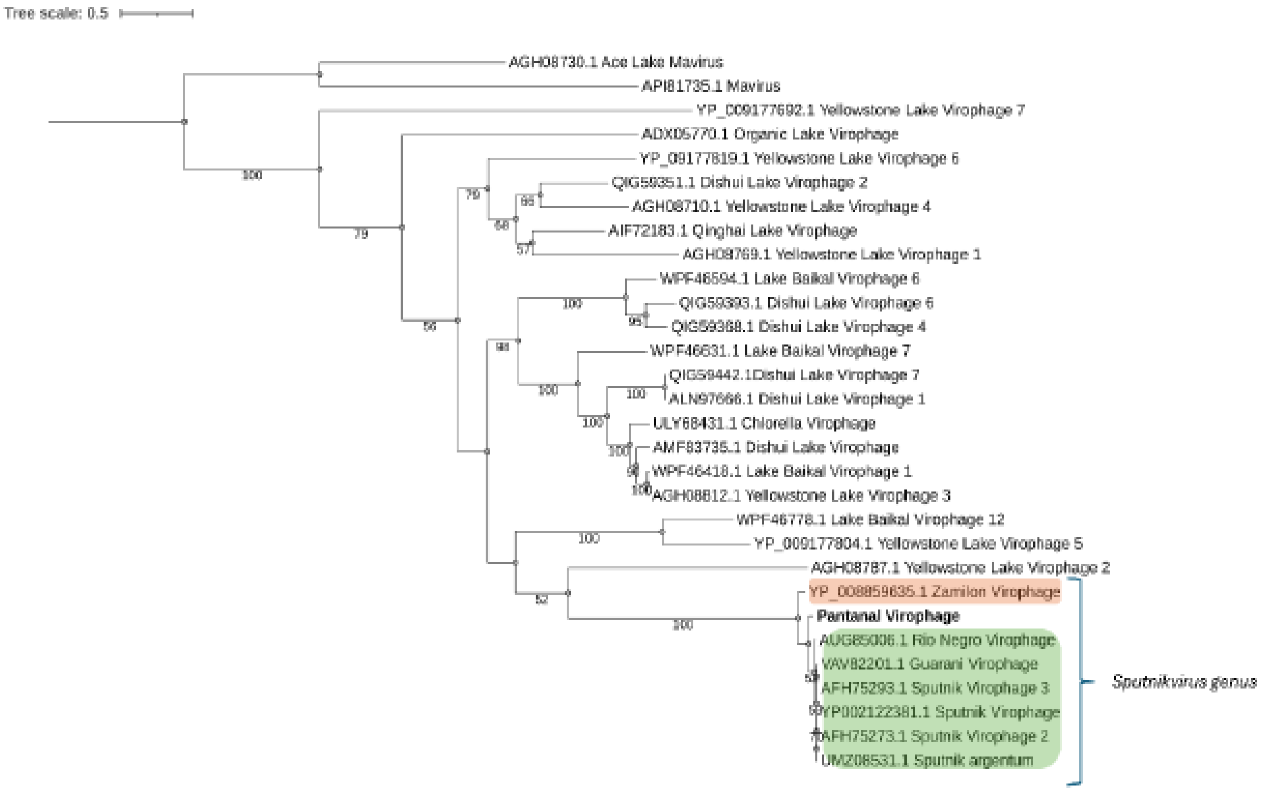


Figure 4 – Major Capsid Protein (MCP) phylogeny based on amino acid sequences from different virophages. Pantanal virophage is labeled in bold. Sequences labeled in green are classified whitin the current *Sputnikvirus mimiviri* species while the sequence labeled in orange belongs to the current *Sputnikvirus zamilonense* species. This maximum likelihood phylogenetic tree was built with statistical support based on 1000 replicates (bootstrap) and was rooted at the midpoint. The best model, selected by IQtree (ModelFinder), for the tree was rtREV+F+R3. The scale bar represents the number of amino acid substitutions per site.

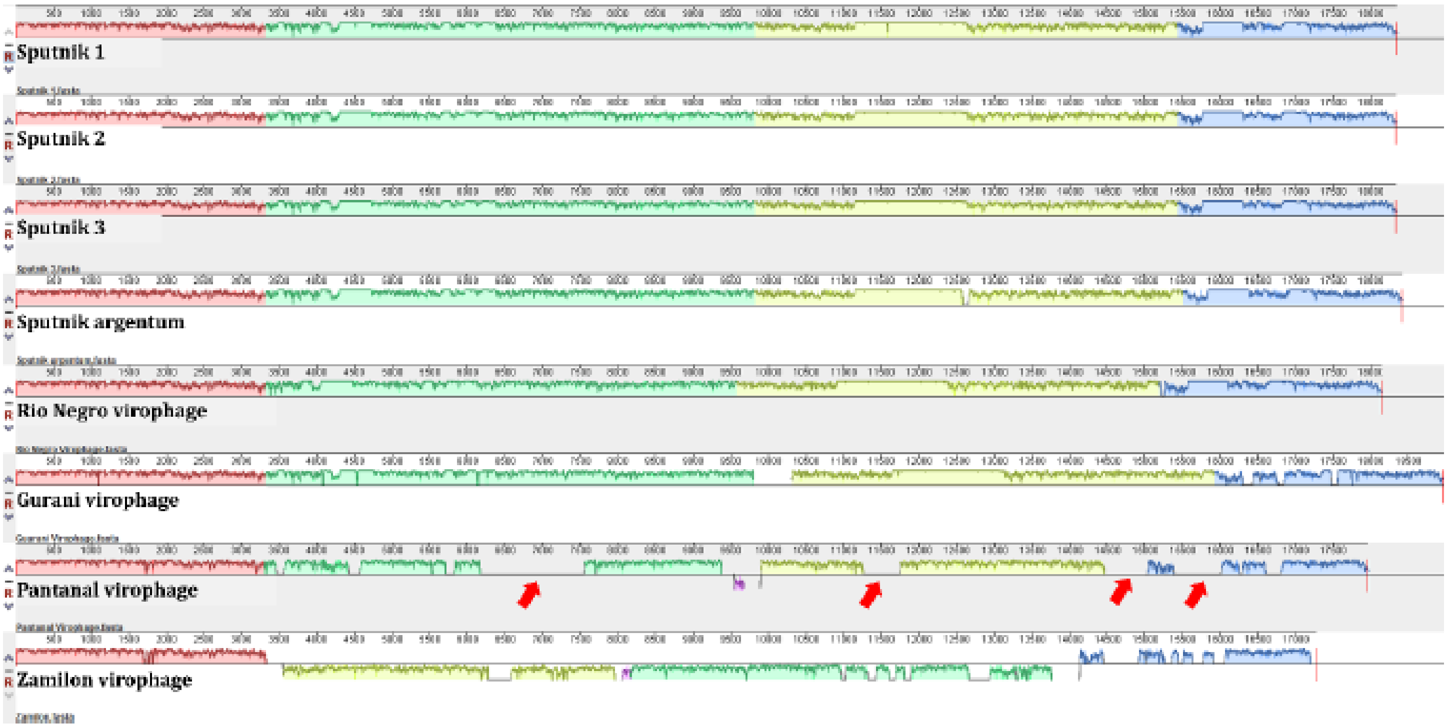


Figure 5 – Genomic synteny based on complete sequences from different *Sputnikvirus* virophages. Each line represents a different virophage sequence. The sequences identification is found below each line. In this analysis the blocks of the same color indicate similar regions between sequences. The areas without any colored blocks represent regions exclusive to that virus, that is, which do not show similarity with the other viruses used in the analysis. Some of the exclusive regions from Pantanal virophage are indicated by red arrows. Note: All the sequences were adjusted to start from the main capsid protein (MCP) aiming to facilitate interpretation of this figure, because they have a circular topology.

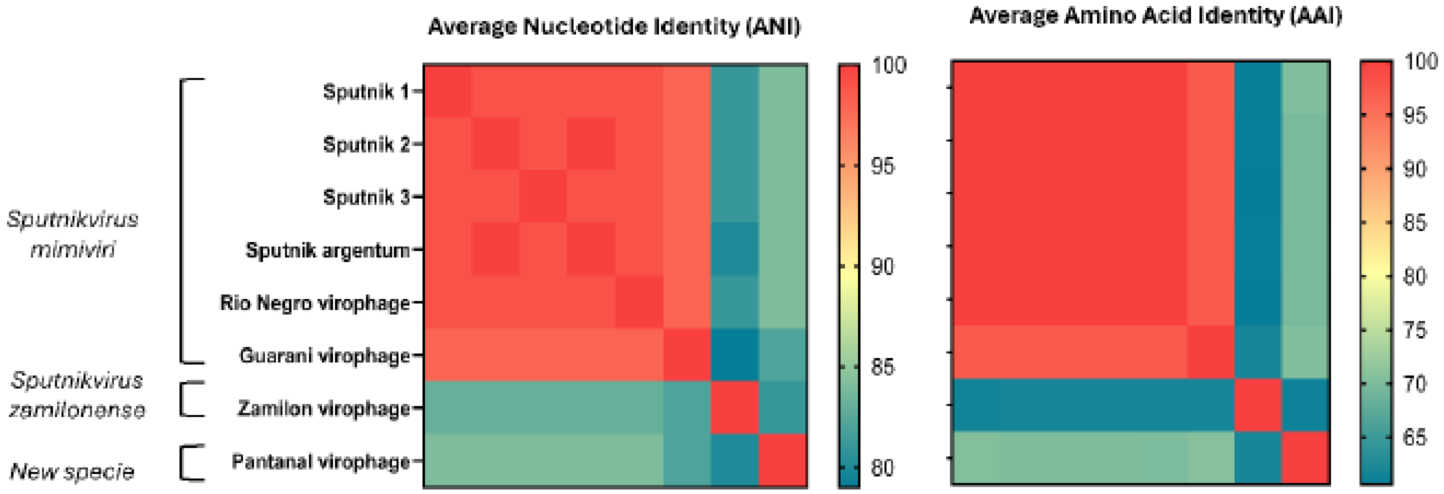
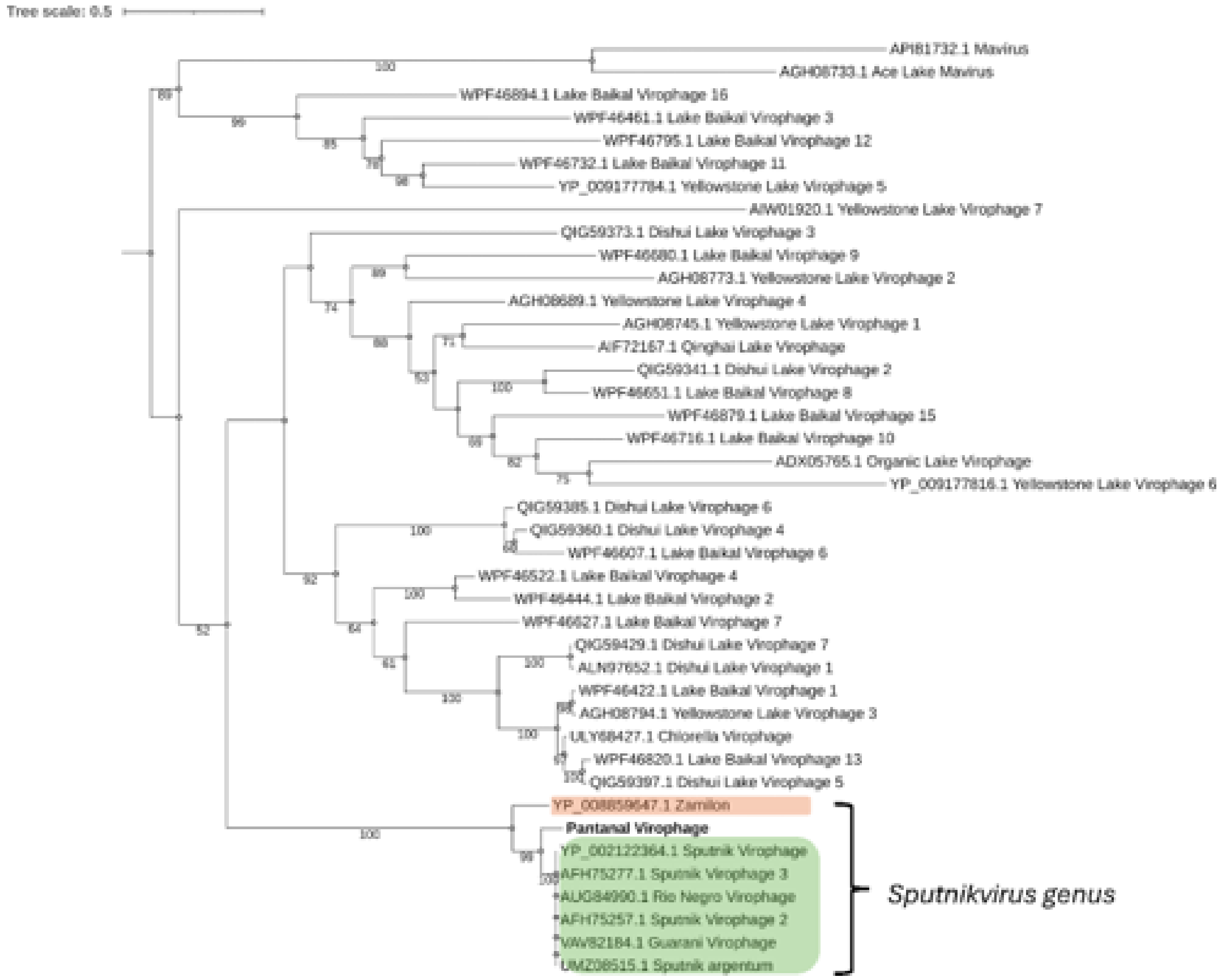
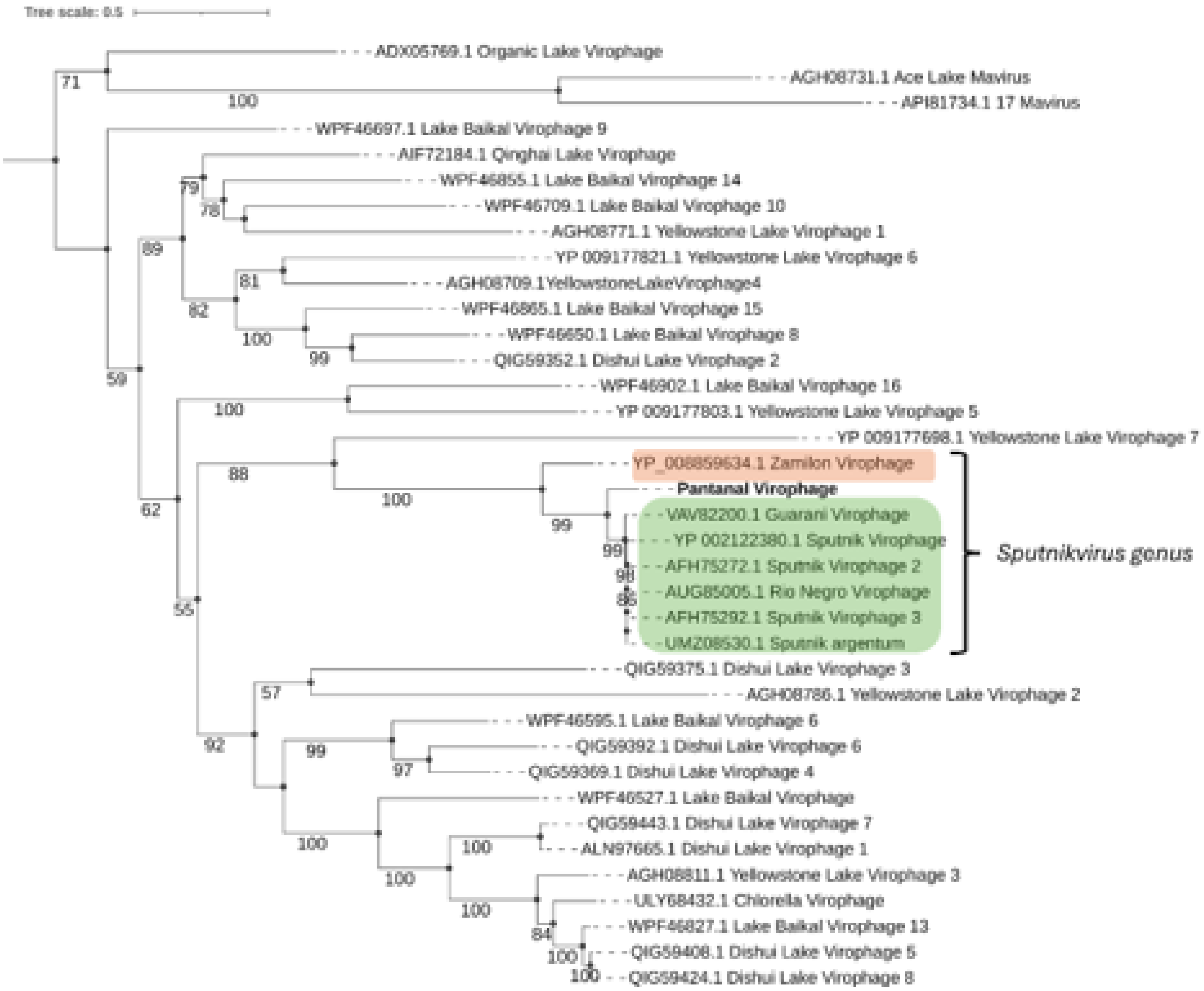


Figure 6 – Average nucleotide identity (ANI) and Average Amino Acid Identity (AAI) analysis comparing different virophage sequences from *Sputnikvirus* genus, including Pantanal virophage. The analysis is based on a similarity matrix composed by ANI and AAI values presented on left and right heatmaps, respectively. The two current species from the genus (*Sputnikvirus mimiviri* and *Sputnikvirus zamilonense*) are indicated. Pantanal virophage is indicated as a new species. ANI values range from 80 to 100% while AAI values range from 60 to 100%.



Supplementary figure 1 – Maximum likelihood phylogeny based on DNA packaging ATPase amino acid sequences from virophages. Pantanal virophage is labeled in bold among *Sputnikvirus* genus sequences. Sequences highlighted in green are classified whitin the current *Sputnikvirus mimiviri* species while the sequence highlighted in orange belongs to the current *Sputnikvirus zamilonense* species. The tree was built with statistical support based on 1000 replicates (bootstrap) and was rooted at the midpoint. The best model, selected by IQtree (ModelFinder), for the tree was rtREV+F+I+G4. The tree scale bar represents the number of amino acid substitutions per site.



Supplementary figure 2 – Phylogenetic tree based on minor virion protein (mCP) sequences from different virophages. Pantanal virophage is labeled in bold among *Sputnikvirus* genus sequences. Sequences highlighted in green are classified whitin the current *Sputnikvirus mimiviri* species while the sequence highlighted in orange belongs to the current *Sputnikvirus zamilonense* species. This maximum likelihood phylogenetic tree was built with statistical support based on 1000 replicates (bootstrap). The branches were rooted at the midpoint. The best model, selected by IQtree (ModelFinder), for the tree was Blosum62+F+I+G4. The tree scale bar represents the number of amino acid substitutions per site.

Materials and Methods

Viral isolation

The sample was collected from sewage samples obtained during an expedition to Porto Murtinho city (21°41'36"S 57°53'25"W), Mato Grosso do Sul state, Brazil. This city is located at the southern end of the Pantanal biome, which is a South American ecosystem known for its wetlands and great biodiversity. The giant virus and the virophage were isolated following a well-stablished protocol based on the inoculation of the collected samples on 96-well plates containing *Acanthamoeba castellanii* cells (25). Non-infected wells were reserved in the same conditions to be used as the experiment control. The inoculated wells were often observed aiming to search for cytopathic effects (CPE), such as rounding cells and cellular lysis. When the CPE was observed the well content was collected and analyzed through transmission electron microscopy (TEM) aiming to confirm the isolation.

Viral production and titration

To conduct genomic analysis, we needed to produce a great quantity of the isolated virus. Then, cell culture flasks filled with 1.4 × 10^7 *Acanthamoeba castellanii* cells and 35 mL of peptone-yeast extract-glucose (PYG) medium, supplemented with penicillin (100 U/mL; Cellofarm, Brazil), streptomycin (100 µg/mL; Sigma-Aldrich, USA), and amphotericin B (0.25 µg/mL; Cultilab, Brazil) was inoculated with the isolates considering at a multiplicity of infection (MOI) of 0.01 for the giant virus as performed before (26). The viral titration were obtained and calculated using the end-point method (27). The infected cells were incubated at 30°C. Non-infected cells maintained in the same conditions were used as control. When viral-induced CPE were observed, the flask's content was collected and then it was ultracentrifuged (36,000 x g) in tubes containing a 22% sucrose cushion for 30 minutes. The purified viral particles were stored at -20°C in microtubes containing 300 µL of phosphate-buffered saline (PBS 1×).

Transmission electron microscopy

The samples containing the isolated viruses were processed to transmission electron microscopy (TEM) analysis. A total of 7×10^6 *Acanthamoeba castellanii* cells, grown in a cell culture flask containing 25 mL of PYG medium, were infected with the virus at a multiplicity of infection (MOI) of 0.01. After the observation of CPEs the flask's contet was washed twice using 0.1 M sodium phosphate buffer. After, the sample was fixed in a solution containing glutaraldehyde (2.5%) and sodium phosphate buffer (0.1 M) for 2 hours. Post-fixation methodology was performed by using 2% osmium tetroxide followed by the inclusion of the sample in Epon resin to allow the sectioning. A Spirit Biotwin FEI transmission electron microscope (120 kV) was used to analyze the samples at the Microscopy Center of the Federal University of Minas Gerais (CM-UFMG).

Genome sequencing, assembly and annotation

The virus-virophage-transpoviron sequences were obtained using an Illumina MiSeq instrument with a single-end library using the Illumina DNA Prep Kit (Illumina Inc., San Diego, CA, USA). The obtained reads were submitted to FastQC program to a sequence quality control. The cutoff for satisfactory sequence quality was phred values that were bigger than 28. The reads were trimmed using the Trimmomatic tool (28). The sequences were de novo assembled through Spades 3.12 program with default parameters (29, 30). After assembling, the scaffolds were compared with the NCBI database, using BLASTn (database: nr/nt; expect threshold: 10^{-3}). FASTQC, Trimmomatic and Spades were all used in this work through the Galaxy server (31). Virophage and transpoviron genes prediction were performed using the Prodigal software (32) with the meta mode parameter. The functional annotation of the proteins was obtained through BLASTp (expect threshold: 10^{-3}) analysis against the NCBI non-redundant protein sequence (nr) database.

Genomic synteny

The comparison between different virophage genomic sequences was performed by making a genome synteny analysis using the MAUVE program, with its default parameters (33). The sequences used in this analysis were obtained from the NCBI GenBank database available until February 2025: Sputnik 1 (NC_011132.1); Sputnik 2 (JN603369.1); Sputnik 3 (JN603370.1); Rio Negro virophage (MG676470.1); Sputnik argentum (OL770071.1); Guarani virophage (LS999520.1); Zamilon virophage (NC_022990.1). A manual curation was performed aiming to standardize the sequences to start with the major capsid protein since virophages have genomes with circular topology.

Phylogenetic analysis

The sequences for alignment and phylogenetic analysis were obtained through a search for similar sequences against the NCBI non-redundant protein sequences (nr) database using BLASTp with an expected threshold of 10^{-3}. The data sets were aligned with the MUSCLE algorithm using the MEGA X software (34, 35). Maximum likelihood phylogenetic trees were obtained using the IQtree software (version 1.6.12) with 1,000 bootstrap replicates as branch support (36). The best-fit substitution models were determined using the ModelFinder algorithm within IQtree (37). The resulting phylogenetic trees were visualized and edited using iTOL (38).

Average nucleotide and amino acids' identity analysis.

The average nucleotide identity (ANI) analysis was performed using whole-genome sequences obtained from NCBI database as described before for genomic synteny. These genomic sequences were submitted to FastANI (39) by using the Galaxy Server (31). Average amino acid identity (AAI) was calculated using the AAI calculator available at http://enve-omics.ce.gatech.edu/aai/. It analyzed reciprocal best hits (two-way AAI) between two virophages protein genomic data sets. A minimum identity cutoff of 20% was considered for AAI. The ANI and AAI obtained data was plotted in heatmaps developed using GraphPad Prism 10.5.0 software.

Acknowledgments: We are grateful to all members of the Giant Virus Study and Prospecting Group (GEPVIG) and the Virus Laboratory at the Federal University of Minas Gerais (UFMG) for their valuable support. We also acknowledge the Microscopy Center at UFMG for providing microscopy images. This research was financially supported by CNPq (National Council for Scientific and Technological Development), CAPES (Coordination for the Improvement of Higher Education Personnel), and FAPEMIG (Research Support Foundation of the State of Minas Gerais).

The authors declare no conflicts of interest.

References

1. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raoult D. 2003. A giant virus in amoebae. Science 299:2033.

2. ICTV. 2024. Current ICTV Taxonomy Release | ICTV. Current ICTV Taxonomy Release. https://ictv.global/taxonomy. Retrieved 2 May 2025.

3. La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, Merchat M, Suzan-Monti M, Forterre P, Koonin E, Raoult D. 2008. The virophage as a unique parasite of the giant mimivirus. Nature 455:100–104.

4. Gaia M, Benamar S, Boughalmi M, Pagnier I, Croce O, Colson P, Raoult D, La Scola B. 2014. Zamilon, a novel virophage with Mimiviridae host specificity. PLoS One 9:e94923.

5. Fischer MG, Hackl T. 2016. Host genome integration and giant virus-induced reactivation of the virophage mavirus. Nature 540:288–291.

6. Gaia M, Pagnier I, Campocasso A, Fournous G, Raoult D, La Scola B. 2013. Broad spectrum of mimiviridae virophage allows its isolation using a mimivirus reporter. PLoS One 8:e61912.

7. Mougari S, Chelkha N, Sahmi-Bounsiar D, Di Pinto F, Colson P, Abrahao J, La Scola B. 2020. A virophage cross-species infection through mutant selection represses giant virus propagation, promoting host cell survival. Commun Biol 3:248.

8. Duponchel S, Fischer MG. 2019. Viva lavidaviruses! Five features of virophages that parasitize giant DNA viruses. PLoS Pathog 15:e1007592.

9. Mougari S, Sahmi-Bounsiar D, Levasseur A, Colson P, La Scola B. 2019. Virophages of Giant Viruses: An Update at Eleven. Viruses 11:733.

10. Desnues C, La Scola B, Yutin N, Fournous G, Robert C, Azza S, Jardot P, Monteil S, Campocasso A, Koonin EV, Raoult D. 2012. Provirophages and transpovirons as the diverse mobilome of giant viruses. Proc Natl Acad Sci U S A 109:18078–18083.

11. Jeudy S, Bertaux L, Alempic J-M, Lartigue A, Legendre M, Belmudes L, Santini S, Philippe N, Beucher L, Biondi EG, Juul S, Turner DJ, Couté Y, Claverie J-M, Abergel C. 2020. Exploration of the propagation of transpovirons within Mimiviridae reveals a unique example of commensalism in the viral world. ISME J 14:727–739.

12. Azevedo BL de, Júnior JPA, Ullmann LS, Rodrigues RAL, Abrahão JS. 2022. The Discovery of a New Mimivirus Isolate in Association with Virophage-Transpoviron Elements in Brazil Highlights the Main Genomic and Evolutionary Features of This Tripartite System. Viruses 14:206.

13. Bessenay A, Bisio H, Belmudes L, Couté Y, Bertaux L, Claverie J-M, Abergel C, Jeudy S, Legendre M. 2024. Complex transcriptional regulations of a hyperparasitic quadripartite system in giant viruses infecting protists. Nat Commun 15:8608.

14. Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ, Andrews-Pfannkoch C, Lewis M, Hoffman JM, Gibson JA, Cavicchioli R. 2011. Virophage

control of antarctic algal host-virus dynamics. Proc Natl Acad Sci U S A 108:6163–6168.

15. Zhou J, Zhang W, Yan S, Xiao J, Zhang Y, Li B, Pan Y, Wang Y. 2013. Diversity of virophages in metagenomic data sets. J Virol 87:4225–4236.

16. Zhou J, Sun D, Childers A, McDermott TR, Wang Y, Liles MR. 2015. Three novel virophage genomes discovered from Yellowstone Lake metagenomes. J Virol 89:1278–1285.

17. Gong C, Zhang W, Zhou X, Wang H, Sun G, Xiao J, Pan Y, Yan S, Wang Y. 2016. Novel Virophages Discovered in a Freshwater Lake in China. Front Microbiol 7:5.

18. Oh S, Yoo D, Liu W-T. 2016. Metagenomics Reveals a Novel Virophage Population in a Tibetan Mountain Lake. Microbes Environ 31:173–177.

19. de Aquino ILM, Serafim MSM, Machado TB, Azevedo BL, Cunha DES, Ullmann LS, Araújo JP, Abrahão JS. 2023. Diversity of Surface Fibril Patterns in Mimivirus Isolates. J Virol 97:e0182422.

20. Aylward FO, Abrahão JS, Brussaard CPD, Fischer MG, Moniruzzaman M, Ogata H, Suttle CA. 2023. Taxonomic update for giant viruses in the order Imitervirales (phylum Nucleocytoviricota). Arch Virol 168:283.

21. The genomic and phylogenetic analysis of Marseillevirus cajuinensis raises questions about the evolution of Marseilleviridae lineages and their taxonomical organization - PubMed. https://pubmed.ncbi.nlm.nih.gov/38752754/. Retrieved 18 April 2025.

22. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, Robert C, Azza S, Sun S, Rossmann MG, Suzan-Monti M, La Scola B, Koonin EV, Raoult D. 2009. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc Natl Acad Sci U S A 106:21848–21853.

23. Raoult D, Boyer M. 2010. Amoebae as genitors and reservoirs of giant viruses. Intervirology 53:321–329.

24. Fischer MG, Suttle CA. 2011. A virophage at the origin of large DNA transposons. Science 332:231–234.

25. Machado TB, de Aquino ILM, Abrahão JS. 2022. Isolation of Giant Viruses of Acanthamoeba castellanii. Curr Protoc 2:e455.

26. Abrahão JS, Boratto P, Dornas FP, Silva LC, Campos RK, Almeida GMF, Kroon EG, La Scola B. 2014. Growing a giant: evaluation of the virological parameters for mimivirus production. J Virol Methods 207:6–11.

27. Reed LJ, Muench H. 1938. A simple method of estimating fifty per cent endpoints. American Journal of Epidemiology 27:493–497.

28. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

29. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly

algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.

30. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics 70:e102.

31. Galaxy Community. 2024. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. Nucleic Acids Res 52:W83–W94.

32. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

33. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. 7. Genome Res 14:1394–1403.

34. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

35. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547–1549.

36. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274.

37. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589.

38. Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128.

39. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114.

5 CONCLUSÕES

A primeira parte do trabalho traz um artigo onde está descrito o primeiro isolado de Marseillevirus de água salgada, o qual apresenta um genoma com elementos raramente descritos para esse grupo de vírus, os RNAs transportadores. Além disso, a filogenia revelou uma divergência do isolado dentro dos Marseillevírus de linhagem A e uma organização dos Marseillevírus em três grandes grupos ao invés de cinco linhagens, o que implica em uma reorganização em níveis taxonômicos como gênero e espécie. Ainda na primeira parte, uma proposta de reorganização taxonômica oficial foi construída e submetida ao comitê de taxonomia viral (ICTV), destacando ainda mais a importância da descoberta do Marseillevírus cajuinensis para a virologia.

A segunda parte deste trabalho é composta por um artigo que descreve o isolamento de um moumouvírus envolvido em uma relação tripla com um virófago e um transpoviron, a partir de uma amostra do Pantanal brasileiro. Embora esteja agrupo junto aos Sputnikvirus, o virófago pantanal apresenta características genômicas e filogenéticas que o distinguem dos virófagos já descritos. Ele apresenta 3 ORFans, cujas funções ainda são completamente desconhecidas. A sintenia dos genomas de diferentes Sputnikvirus, bem como a filogenia das principais proteínas conservadas entre os virófagos revelou a divergência do novo isolado em relação aos Sputnikvirus já descritos. Por fim, dados de identidade de aminoácidos e nucleotídeos reforçam essa divergência e ajudam a classificar o virófago pantanal como uma nova espécie de Sputnikvirus, expandindo esse gênero viral.

Em conjunto os resultados desse trabalho mostram que a adição de novos vírus nas análises genômicas e filogenéticas ajuda a entender cada vez melhor as relações evolutivas existentes entre os diferentes isolados e que com uma maior diversidade descoberta, se faz necessária uma constante atualização taxonômica. A taxonomia é uma ciência que demanda constantes atualizações e melhorias, na medida em que novas informações e métodos são descritos. Isso evidencia que o isolamento de novos vírus e sua caracterização podem contribuir para preencher lacunas que existam no conhecimento sobre seus aspectos biológicos, genômicos, evolutivos e taxonômicos, enriquecendo, portanto, o conhecimento básico acerca dos vírus do reino *Barmfordvirae*.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRAHÃO, J. S. et al. The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. **PLoS genetics**, v. 13, n. 2, p. e1006532, fev. 2017.

AHERFI, S. et al. Complete genome sequence of Cannes 8 virus, a new member of the proposed family "Marseilleviridae". **Virus Genes**, v. 47, n. 3, p. 550–555, dez. 2013.

AHERFI, S. et al. Complete genome sequence of Tunisvirus, a new member of the proposed family Marseilleviridae. **Archives of Virology**, v. 159, n. 9, p. 2349–2358, set. 2014.

ANDRADE, A. C. D. S. P. et al. Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. **Virology Journal**, v. 15, n. 1, p. 22, 24 jan. 2018.

AOKI, K. et al. Fifteen Marseilleviruses Newly Isolated From Three Water Samples in Japan Reveal Local Diversity of Marseilleviridae. **Frontiers in Microbiology**, v. 10, p. 1152, 2019.

ARANTES, T. S. et al. The Large Marseillevirus Explores Different Entry Pathways by Forming Giant Infectious Vesicles. **Journal of Virology**, v. 90, n. 11, p. 5246–5255, 12 maio 2016.

AZEVEDO, B. L. DE et al. The Discovery of a New Mimivirus Isolate in Association with Virophage-Transpoviron Elements in Brazil Highlights the Main Genomic and Evolutionary Features of This Tripartite System. **Viruses**, v. 14, n. 2, p. 206, 21 jan. 2022.

BÄCKSTRÖM, D. et al. Virus Genomes from Deep Sea Sediments Expand the Ocean Megavirome and Support Independent Origins of Viral Gigantism. **mBio**, v. 10, n. 2, p. e02497-18, 5 mar. 2019.

BALTIMORE, D. Expression of animal virus genomes. **Bacteriological Reviews**, v. 35, n. 3, p. 235–241, set. 1971.

BEKLIZ, M. et al. A New Zamilon-like Virophage Partial Genome Assembled from a Bioreactor Metagenome. **Frontiers in Microbiology**, v. 6, p. 1308, 2015.

BELLAS, C. M.; ANESIO, A. M.; BARKER, G. Analysis of virus genomes from glacial environments reveals novel virus groups with unusual host interactions. **Frontiers in Microbiology**, v. 6, p. 656, 2015.

BLANCA, L. et al. Comparative Analysis of the Circular and Highly Asymmetrical Marseilleviridae Genomes. **Viruses**, v. 12, n. 11, p. 1270, 7 nov. 2020.

BORN, D. et al. Capsid protein structure, self-assembly, and processing reveal morphogenesis of the marine virophage mavirus. **Proceedings of the National Academy of Sciences of the United States of America**, v. 115, n. 28, p. 7332–7337, 10 jul. 2018.

BOUGHALMI, M. et al. First isolation of a Marseillevirus in the Diptera Syrphidae Eristalis tenax. **Intervirology**, v. 56, n. 6, p. 386–394, 2013.

BOYER, M. et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 51, p. 21848–21853, 22 dez. 2009.

BRYSON, T. D. et al. A giant virus genome is densely packaged by stable nucleosomes within virions. **Molecular Cell**, v. 82, n. 23, p. 4458- 4470.e5, 1 dez. 2022.

CAMPOS, R. K. et al. Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. **Virology Journal**, v. 11, n. 1, p. 95, dez. 2014.

COLSON, P. et al. Evidence of the megavirome in humans. **Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology**, v. 57, n. 3, p. 191–200, jul. 2013.

DESNUES, C. et al. Provirophages and transpovirons as the diverse mobilome of giant viruses. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 44, p. 18078–18083, 30 out. 2012.

DESNUES, C.; RAOULT, D. Virophages question the existence of satellites. **Nature Reviews. Microbiology**, v. 10, n. 3, p. 234; author reply 234, 16 fev. 2012.

DORNAS, F. et al. A Brazilian Marseillevirus Is the Founding Member of a Lineage in Family Marseilleviridae. **Viruses**, v. 8, n. 3, p. 76, 10 mar. 2016.

DOS SANTOS, R. N. et al. A new marseillevirus isolated in Southern Brazil from Limnoperna fortunei. **Scientific Reports**, v. 6, n. 1, p. 35237, dez. 2016.

DOUTRE, G. et al. Genome analysis of the first Marseilleviridae representative from Australia indicates that most of its genes contribute to virus fitness. **Journal of Virology**, v. 88, n. 24, p. 14340–14349, dez. 2014.

DOUTRE, G. et al. Complete Genome Sequence of a New Member of the Marseilleviridae Recovered from the Brackish Submarine Spring in the Cassis Port-Miou Calanque, France. **Genome Announcements**, v. 3, n. 6, p. e01148-15, 25 nov. 2015.

FABRE, E. et al. Noumeavirus replication relies on a transient remote control of the host nucleus. **Nature Communications**, v. 8, p. 15087, 21 abr. 2017.

FISCHER, M. G. The Virophage Family Lavidaviridae. **Current Issues in Molecular Biology**, v. 40, p. 1–24, 2021.

FISCHER, M. G.; SUTTLE, C. A. A virophage at the origin of large DNA transposons. **Science (New York, N.Y.)**, v. 332, n. 6026, p. 231–234, 8 abr. 2011.

GAIA, M. et al. Broad spectrum of mimiviridae virophage allows its isolation using a mimivirus reporter. **PloS One**, v. 8, n. 4, p. e61912, 2013.

GAIA, M. et al. Zamilon, a novel virophage with Mimiviridae host specificity. **PloS One**, v. 9, n. 4, p. e94923, 2014.

GORBALENYA, A. E. et al. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. **Nature Microbiology**, v. 5, n. 5, p. 668–674, maio 2020.

ICTV. **Current ICTV Taxonomy Release | ICTV**. Disponível em: <https://ictv.global/taxonomy>. Acesso em: 2 maio. 2025.

IYER, L. M. et al. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. **Virus Research**, v. 117, n. 1, p. 156–184, abr. 2006.

IYER, L. M.; ARAVIND, L.; KOONIN, E. V. Common origin of four diverse families of large eukaryotic DNA viruses. **Journal of Virology**, v. 75, n. 23, p. 11720–11734, dez. 2001.

JEUDY, S. et al. Exploration of the propagation of transpovirons within Mimiviridae reveals a unique example of commensalism in the viral world. **The ISME journal**, v. 14, n. 3, p. 727–739, mar. 2020.

KOONIN, E. V. et al. Global Organization and Proposed Megataxonomy of the Virus World. **Microbiology and molecular biology reviews: MMBR**, v. 84, n. 2, p. e00061-19, 20 maio 2020.

KORNBERG, R. D.; THOMAS, J. O. Chromatin structure; oligomers of the histones. **Science (New York, N.Y.)**, v. 184, n. 4139, p. 865–868, 24 maio 1974.

KRUPOVIC, M.; KUHN, J. H.; FISCHER, M. G. A classification system for virophages and satellite viruses. **Archives of Virology**, v. 161, n. 1, p. 233–247, jan. 2016.

LA SCOLA, B. et al. A giant virus in amoebae. **Science (New York, N.Y.)**, v. 299, n. 5615, p. 2033, 28 mar. 2003.

LA SCOLA, B. et al. The virophage as a unique parasite of the giant mimivirus. **Nature**, v. 455, n. 7209, p. 100–104, 4 set. 2008.

LEVASSEUR, A. et al. MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. **Nature**, v. 531, n. 7593, p. 249–252, 10 mar. 2016.

LIU, Y. et al. Virus-encoded histone doublets are essential and form nucleosome-like structures. **Cell**, v. 184, n. 16, p. 4237- 4250.e19, 5 ago. 2021.

MOUGARI, S. et al. Guarani Virophage, a New Sputnik-Like Isolate From a Brazilian Lake. **Frontiers in Microbiology**, v. 10, p. 1003, 2019a.

MOUGARI, S. et al. Virophages of Giant Viruses: An Update at Eleven. **Viruses**, v. 11, n. 8, p. 733, 8 ago. 2019b.

OLIVEIRA, G. P. et al. The Investigation of Promoter Sequences of Marseilleviruses Highlights a Remarkable Abundance of the AAATATTT Motif in Intergenic Regions. **Journal of Virology**, v. 91, n. 21, p. e01088-17, 1 nov. 2017.

RODRIGUES, R. A. L. et al. Analysis of a Marseillevirus Transcriptome Reveals Temporal Gene Expression Profile and Host Transcriptional Shift. **Frontiers in Microbiology**, v. 11, p. 651, 2020.

SAHMI-BOUNSIAR, D. et al. Marseilleviruses: An Update in 2021. **Frontiers in Microbiology**, v. 12, p. 648731, 2021.

SIDDELL, S. G. et al. Virus taxonomy and the role of the International Committee on Taxonomy of Viruses (ICTV). **The Journal of General Virology**, v. 104, n. 5, p. 001840, 4 maio 2023.

TAKEMURA, M. Morphological and Taxonomic Properties of Tokyovirus, the First Marseilleviridae Member Isolated from Japan. **Microbes and Environments**, v. 31, n. 4, p. 442–448, dez. 2016.

THOMAS, V. et al. Lausannevirus, a giant amoebal virus encoding histone doublets. **Environmental Microbiology**, v. 13, n. 6, p. 1454–1466, 2011.

YAU, S. et al. Virophage control of antarctic algal host-virus dynamics. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 15, p. 6163–6168, 12 abr. 2011.

YUTIN, N.; KAPITONOV, V. V.; KOONIN, E. V. A new family of hybrid virophages from an animal gut metagenome. **Biology Direct**, v. 10, p. 19, 25 abr. 2015.

YUTIN, N.; RAOULT, D.; KOONIN, E. V. Virophages, polintons, and transpovirons: a complex evolutionary network of diverse selfish genetic elements with different reproduction strategies. **Virology Journal**, v. 10, p. 158, 23 maio 2013.

ZABLOCKI, O. et al. High-level diversity of tailed phages, eukaryote-associated viruses, and virophage-like elements in the metaviromes of antarctic soils. **Applied and Environmental Microbiology**, v. 80, n. 22, p. 6888–6897, nov. 2014.

ZHOU, J. et al. Diversity of virophages in metagenomic data sets. **Journal of Virology**, v. 87, n. 8, p. 4225–4236, abr. 2013.

ZHOU, J. et al. Three novel virophage genomes discovered from Yellowstone Lake metagenomes. **Journal of Virology**, v. 89, n. 2, p. 1278–1285, 15 jan. 2015.

PRODUÇÕES E ATIVIDADES REALIZADAS DURANTE O DOUTORADO

- Artigos científicos publicados durante o doutorado.

- Artigos como primeira autora:

- ✓ Primeira autora do artigo intitulado: "*The genomic and phylogenetic analysis of Marseillevirus cajuinensis raises questions about the evolution of Marseilleviridae lineages and their taxonomical organization,* publicado no periódico **Journal of Virology,** 2024.
    - DOI: https://doi.org/10.1128/jvi.00513-24
    - **Obs.: Essa publicação contempla os resultados do Capítulo 1 desta tese.**

- ✓ Primeira autora do artigo intitulado: *"The Discovery of a New Mimivirus Isolate in Association with Virophage-Transpoviron Elements in Brazil Highlights the Main Genomic and Evolutionary Features of This Tripartite System",* publicado no periódico **Viruses**, 2022. DOI: https://doi.org/10.3390/v14020206

- Artigos como coautora, em colaboração com colegas:

- ✓ Coautora do artigo intitulado: *"A Brief History of Giant Viruses' Studies in Brazilian Biomes."*, publicado no periódico **Viruses**, 2022. DOI: https://doi.org/10.3390/v14020191

- ✓ Coautora do artigo intitulado: *"Giant Viruses as a Source of Novel Enzymes for Biotechnological Application"*, publicado no periódico **Pathogens**, 2022. DOI: https://doi.org/10.3390/pathogens11121453

- ✓ Coautora do artigo intitulado: *"Diversity of Surface Fibril Patterns in Mimivirus Isolates"*, publicado no periódico **Journal of Virology**, 2023. DOI: https://doi.org/10.1128/jvi.01824-22

- ✓ Coautora do artigo intitulado: *"Gene duplication as a major force driving the genome expansion in some giant viruses"*, publicado no periódico **Journal of Virology,** 2023. DOI: **https://doi.org/10.1128/jvi.01309-23**

- ✓ Coautora do artigo intitulado: *" A long-term prospecting study on giant viruses in terrestrial and marine Brazilian biomes",* publicado no periódico **Virology Journal**, 2024. DOI: https://doi.org/10.1186/s12985-024-02404-z

- ✓ Coautora do artigo intitulado: *"Investigations into the Diversity and Distribution of tRNA and Phylogenetics of Translation Factors in Amoebozoa-Infecting Nucleocytoviricota"*, publicado no periódico **Viruses**, 2025. DOI: https://doi.org/10.3390/v17030328

- ✓ Coautora do artigo intitulado: *"The final cut: how giant viruses of protists are released from their hosts' cells",* publicado no períodico **Archives of Virology**, 2025. DOI: https://doi.org/10.1007/s00705-025-06261-1

- ✓ Coautora do artigo intitulado: *"Genomic and structural insights into*
    - *Jyvaskylavirus, the first giant virus isolated from Finland",* publicado no periódico **eLife**, 2025. DOI: https://doi.org/10.7554/eLife.103492

- Eventos científicos

- ✓ Participação no XXXII Congresso Brasileiro de Virologia (Sociedade Brasileira de Virologia) – 2021 - Resumo publicado em anais: *"Mobilome analysis of a new Brazilian mimivirus isolate reveals a provirophage fully integrated in its genome"*

- ✓ Participação no VIII Simpósio de Microbiologia da UFMG – CONECTA SIM, 2021. – Resumo: "Caracterização do genoma de um isolado de mimivírus e seus elementos genéticos associados"

- ✓ Participação e apresentação de poster no IX Simpósio de Microbiologia da UFMG, 2022. (Belo Horizonte, Minas Gerais) – Poster: *"A descoberta de um novo isolado de mimivírus em associação com um virófago e um transpoviron no Brasil destaca as principais características genômicas e evolutivas desse sistema tripartite."*
    - **Obs.: esse trabalho foi destacado por Relevância Acadêmica no evento.**

- ✓ Participação e apresentação de poster no X Simpósio de Microbiologia da UFMG, 2023. (Belo Horizonte, Minas Gerais) – Poster: "*As análises genômicas e filogenéticas de Marseillevírus cajuinensis levantam questões sobre a evolução das linhagens de Marseilleviridae e sua organização taxonômica."*

- Outras atividades

- ✓ **Curso: "**Introdução à estrutura, evolução e análise por NGS de genomas virais". (Sociedade Brasileira de Virologia, 2021)
- ✓ Participação na equipe de diagnóstico de COVID-19, durante a pandemia da doença, no Laboratório de Vírus (ICB/UFMG), no ano de 2021.
- ✓ Trabalho como professora de Biologia nos Ensinos Fundamental e Médio, desde 2022, com carga horária fixa de 30-40h semanais.