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

TESE DE DOUTORADO

ANÁLISE EXTENSIVA DA VIROSFERA E SEUS HOSPEDEIROS: AVANÇANDO NA SISTEMÁTICA, GENÔMICA E TRANSCRIPTÔMICA DE VÍRUS GIGANTES

RODRIGO ARAÚJO LIMA RODRIGUES

BELO HORIZONTE
2018

RODRIGO ARAÚJO LIMA RODRIGUES

ANÁLISE EXTENSIVA DA VIROSFERA E SEUS HOSPEDEIROS: AVANÇANDO NA SISTEMÁTICA, GENÔMICA E TRANSCRIPTÔMICA DE VÍRUS GIGANTES

Tese apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do grau de **Doutor em Microbiologia**.

Orientador: Prof. Jônatas Santos Abrahão

BELO HORIZONTE

2018

043 Rodrigues, Rodrigo Araújo Lima.
Análise extensiva da virosfera e seus hospedeiros: avançando na sistemática, genômica e transcriptômica de vírus gigantes [manuscrito] / Rodrigo Araújo Lima Rodrigues. – 2018.

194 f. : il. ; 29,5 cm.

Orientador: Dr. Jônatas Santos Abrahão.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas.

1. Microbiologia. 2. Vírus Gigantes. 3. Genômica. 4. Transcriptoma. 5. Classificação. I. Abrahão, Jônatas Santos. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 579

Ficha Catalográfica elaborada pela Biblioteca do Instituto de Ciências Biológicas

Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas
Programa de Pós-Graduação Em Microbiologia

D E C L A R A Ç Ã O

DECLARO para os devidos fins que **RODRIGO ARAÚJO LIMA RODRIGUES** cursou o Doutorado em Microbiologia no Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, sob orientação do Prof. Jonatas Santos Abrahão. Defendeu sua Tese intitulada "Análise extensiva da virosfera e seus hospedeiros: Avançando na sistemática, genômica e transcriptômica de vírus gigantes", na data de 17 de dezembro de 2018, diante da banca examinadora composta pelos(as) Drs. Francisco Pereira Lobo (Departamento de Biologia Geral/ICB/UFMG), Betania Paiva Drumond (Departamento de Microbiologia/ICB/UFMG), Pedro Augusto Alves (Instituto René Rachou, Fiocruz Minas), Jaquelline Germano de Oliveira (Instituto René Rachou, Fiocruz Minas) e seu Orientador, tendo sido Aprovado e recebendo a titulação de **DOUTOR EM CIÊNCIAS BIOLÓGICAS: MICROBIOLOGIA**.

Belo Horizonte, 08 de fevereiro de 2019.

Prof. Flávio Guimarães da Fonseca
Coordenador do Programa de Pós-Graduação em Microbiologia
Instituto de Ciências Biológicas
Universidade Federal de Minas Gerais

DEDICATÓRIA

Aos meus pais, Simone e José William, e à minha esposa Michelle,
pelo apoio incondicional desde o início dessa jornada,
sem o qual eu não teria chegado até aqui.

AGRADECIMENTOS

Mais uma etapa se encerra em minha vida. A jornada foi árdua, com muitos desafios e momentos de tensão. Mas o aprendizado foi imensurável e inúmeros foram os momentos de alegria e satisfação. E isso só foi possível graças àqueles que caminharam ao meu lado durante todo esse tempo, aos quais eu presto aqui uma singela homenagem, sabendo que palavras não serão suficientes para expressar toda a minha real gratidão.

A Deus, por me guiar e me confortar nos momentos de difíceis decisões;

Aos meus pais, por estarem sempre presentes e dispostos a me apoiar no que fosse preciso para eu concluir meu sonho. Espero um dia poder retribuir todo o amor, o carinho e a confiança que depositaram em mim. É um privilégio ter pais como vocês, e por isso serei eternamente grato;

A Michelle, por ser a melhor pessoa, companheira, esposa que eu poderia desejar ao meu lado, que me acompanhou em todos os momentos dessa etapa, me incentivando e acreditando em mim nos momentos que eu mais tive dúvidas. Nada do que eu disser ou escrever será suficiente para demonstrar plenamente o quão grato eu sou por ter você ao meu lado;

Ao meu irmão, Flávio, pelas risadas e pela companhia de sempre. Conte comigo sempre que precisar, pois sei que você sempre estará presente para mim;

Aos meus avós, Hugo e Íris (*in memoriam*), e minha tia Silvana, por todo o amor e exemplo de coragem, companheirismo e compaixão, tão escassos nos dias de hoje;

Ao professor Jônatas Abrahão, pela brilhante orientação durante todo o desenvolvimento desse trabalho, pela confiança e por todas as oportunidades que me fizeram (e ainda fazem) crescer profissionalmente. Obrigado por me ensinar que sempre é possível realizar excelentes trabalhos nas mais adversas situações;

Ao professor Bernard La Scola, pela supervisão durante o período sandwich, e aos colegas do IHU pelo aprendizado e companhia nos corredores e laboratórios do IHU;

Aos professores e colegas do Laboratório de Vírus da UFMG, por propiciarem um ambiente agradável de trabalho e um aprendizado tão rico e fundamental para minha formação;

Aos amigos do GEPVIG, pela verdadeira parceria ao longo destes anos. Vocês têm o meu profundo respeito e admiração. Contém sempre comigo;

Aos meus amigos de longa data, por todos os momentos de diversão e por sempre me mostrar que a distância e o tempo não são empecilhos para uma verdadeira amizade;

Aos amigos feitos em Marseille, pela ajuda todas as vezes que foram necessárias (e não foram poucas) e pelos ótimos momentos durante a estadia na França;

Ao Centro de Microscopia da UFMG e seus ótimos profissionais, pelo excelente trabalho prestado que tanto enriqueceram esse trabalho;

Aos membros da banca avaliadora, pelo interesse, disponibilidade e dedicação para colaborar com a minha formação nesta fase final do trabalho;

As agências financiadoras, CNPq, CAPES e FAPEMIG, pelo auxílio financeiro que permitiu a realização desse trabalho;

Por fim, a toda sociedade brasileira, para a qual trabalhamos incansavelmente. Em épocas de crise e ampla disseminação de "*fake news*", o nosso trabalho se torna ainda mais importante, sendo necessário mostrar e fazer entender, diariamente, a importância da ciência. A construção do conhecimento e de uma sociedade melhor passa pelas nossas mãos. É preciso assumir essa responsabilidade e nunca desistir. Eu não desistirei!

Muito obrigado!

“Nada em Biologia faz sentido
exceto à luz da evolução.”

(Theodore Dobzhansky)

RESUMO

Tese de Doutorado

Programa de Pós-Graduação em Microbiologia
Universidade Federal de Minas Gerais

ANÁLISE EXTENSIVA DA VIROSFERA E SEUS HOSPEDEIROS: AVANÇANDO NA SISTEMÁTICA, GENÔMICA E TRANSCRIPTÔMICA DE VÍRUS GIGANTES

RODRIGO ARAÚJO LIMA RODRIGUES
Orientador: Prof. Jônatas Santos Abrahão
Belo Horizonte, Dezembro de 2018

Os vírus são as entidades biológicas mais abundantes e diversas encontradas no planeta, formando uma virosfera. Eles estão associados a organismos de todos os domínios da vida, mas ainda não está claro como essa rede de interação está conectada. No presente trabalho, por meio de meta-análises extensivas, foi possível obter uma visão ampla desta rede, bem como a distribuição de diferentes grupos virais pelo planeta. Os dados revelam uma virosfera antropocêntrica, em que a maioria dos vírus conhecidos está associada ao ser humano e espécies de interesse médico, econômico ou biotecnológico. Além disso, foi possível identificar ao menos 320 espécies virais com representantes que infectam o ser humano. Dentre os vírus não associados a infecções conhecidas no ser humano, os vírus gigantes se destacam pela sua complexidade genômica e estrutural. Muito se têm debatido sobre sua origem, evolução e classificação, sendo necessárias novas análises para aprimorar a sistemática viral. Reconstruções filogenéticas utilizando diferentes estratégias para obtenção de sequências homólogas levam a conclusões divergentes sobre a origem dos mimivírus e tupanvírus, não sendo possível definir este debate baseado em análises de apenas um ou poucos genes. Ao analisar as características dos tupanvírus, uma série de particularidades foi compilada para justificar a criação de um novo grupo taxonômico para classifica-los adequadamente, o gênero "*Tupanvirus*". Ainda, para melhor compreender a biologia e evolução dos vírus gigantes é preciso avançar em análises genômicas e transcriptômicas. Neste sentido, a análise do genoma do Brazilian cedratvirus IHUMI indicou a existência de uma nova linhagem para os cedratvírus. Esse novo isolado possui partículas menores que os demais vírus do grupo, além de apresentar um genoma menor e assintênico. Por fim, análises de dados oriundos do sequenciamento de RNA do marseillevírus confirmaram a existência dos genes previamente preditos e revelou um perfil temporal de transcrição gênica para esses vírus. Este trabalho forneceu pela primeira vez uma visão global da virosfera e seus hospedeiros, além de avançar em diferentes vertentes no estudo dos vírus gigantes, em especial na sistemática, genômica e transcriptômica de diferentes grupos virais.

Palavras-chave: Virosfera, network, vírus gigantes, sistemática, genômica, transcriptômica

ABSTRACT

Doctoral Thesis

Programa de Pós-Graduação em Microbiologia
Universidade Federal de Minas Gerais

EXTENSIVE ANALYSES OF THE VIROSPHERE AND ITS HOSTS: ADVANCING IN THE SYSTEMATICS, GENOMICS AND TRANSCRIPTOMICS OF GIANT VIRUSES

RODRIGO ARAÚJO LIMA RODRIGUES
Advisor: Prof. Jônatas Santos Abrahão
Belo Horizonte, December 2018

Viruses are the most abundant and diverse biological entities on the planet, forming a virosphere. They are associated to organisms from all domains of life, but it is still not clear how this network is connected. Here, by performing an extensive meta-analyses it was possible to obtain a wide view about this network, as well as the distribution of different group of viruses on the planet. The data presented here revealed an anthropocentric virosphere, wherein most of the known viruses are associated to human being and species of medical, economical or biotechnological interest. Moreover, it was possible to identify at least 320 viral species with representatives able to infect humans. Among the viruses that are not associated as etiological agents of infections in humans, giant viruses are noteworthy by their genomic and structural complexity. Much has been debated about their origin, evolution and classification, and new analyzes are needed to improve viral systematics. Phylogenetic reconstructions using different strategies to obtain homologous sequences led to divergent conclusions about the origin of mimiviruses and tupanviruses, and it is not possible to define this debate based on analyzes of only one or a few genes. When reviewing the characteristics of the tupanvirus, a series of particularities was compiled to justify the creation of a new taxonomic group to properly classify them, the genus "*Tupanvirus*". In addition, to better understand the biology and evolution of giant viruses, genomic and transcriptomic analyzes must be advanced. In this sense, the analysis of the genome of the Brazilian cedratvirus IHUMI indicated the existence of a new lineage of cedratvirus. The new isolate has smaller viral particles than the other viruses in the group, besides presenting a shorter and asyntenic genome. Finally, data analysis from the RNA sequencing of marseillevirus confirmed the existence of the genes previously predicted and revealed a temporal profile of gene transcription for these viruses. This work provided, for the first time, a global view of the virosphere and its hosts, besides advancing in different aspects in the study of the giant viruses, especially in the systematics, genomics and transcriptomics of different viral groups.

Key-words: Virosphere, network, giant viruses, systematics, genomics, transcriptomics

SUMÁRIO

1. INTRODUÇÃO

A virologia teve início no final do século XIX com os trabalhos de Adolf Mayer, Dmitry Ivanofsky e Martinus Beijerinck sobre a doença do mosaico do tabaco. Os pesquisadores perceberam que estavam lidando com um agente até então desconhecido pela comunidade acadêmica, o qual mantinha seu caráter infeccioso após filtração e, mesmo diluído, o organismo recuperava sua infectividade ao replicar em tecidos vivos de plantas saudáveis. O patógeno foi denominado "*contagium vivum fluidum*" e posteriormente nomeado como vírus do mosaico do tabaco, o primeiro vírus descrito na literatura (ENQUIST; RACANIELLO, 2013). A partir de então, novos vírus foram sendo constantemente descritos. Em 1901, foi descoberto o vírus da febre amarela – *Yellow fever virus* – por Walter Reed, sendo este o primeiro vírus a ser descrito como causador de doença em humanos. Em 1915, Frederick Twort e Felix d'Herelle descobriram independentemente os primeiros vírus de procariotos, denominados de bacteriófagos, ou apenas fagos (ENQUIST; RACANIELLO, 2013). Porém foi apenas na década de 1950 que o primeiro conceito de vírus foi oficialmente publicado. Em 1957, André Lwoff publicou um trabalho seminal descrevendo pela primeira vez as características para uma entidade ser considerado um vírus, dentre elas: ser um parasita intracelular totalmente dependente da maquinaria biossintética de seu hospedeiro; ser capaz de passar em filtros de 200 nm; e apresentar apenas um tipo de ácido nucléico em sua partícula (LWOFF, 1957). Com o avanço da virologia, foi criado o Comitê Internacional de Taxonomia Viral (ICTV) em 1966 com o objetivo de catalogar e organizar as espécies virais que fossem sendo descritas. Surgiram assim as primeiras regras para a taxonomia viral. Poucos anos depois, David Baltimore propôs uma divisão dos vírus de acordo com as propriedades de seus materiais genéticos, sendo definidos na época seis grupos: I (dsDNA), II (ssDNA), III (dsRNA), IV (ssRNA+), V (ssRNA-), e VI (ssRNA-RT) (BALTIMORE, 1971). Atualmente mais duas categorias foram inseridas, compondo os grupos VII (dsDNA-RT) e VIII (viroides). Tal classificação ainda hoje é bem aceita entre os virologistas.

Com o passar dos anos, diversos vírus foram sendo descritos, isolados de praticamente todos os locais do planeta, e de hospedeiros pertencentes aos três domínios da vida, Eukarya, Bacteria e Archaea, sempre reforçando os critérios estabelecidos na década de 1950 para se reconhecer um vírus. O avanço no campo

na genômica nos últimos anos, em especial da metagenômica (ou ainda, metavirômica), permitiu a identificação de inúmeras sequências virais em diversas regiões do globo, corroborando dados prévios obtidos por microscopia eletrônica que sugeriam a ubiquidade viral e uma quantidade astronômica de vírus na Terra, com cerca de 10^{31} partículas virais em todo o planeta, formando uma virosfera (KRISTENSEN *et al.*, 2010; SUTTLE, 2007). Vários estudos de metavirômica têm sido realizados em diferentes partes do mundo, permitindo a identificação de vários grupos virais em ambientes até então inexplorados, bem como a identificação de possíveis novas espécies virais (PAEZ-ESPINO *et al.*, 2016; SHI *et al.*, 2016). Por outro lado, a grande maioria das sequências identificadas nestes estudos não possui similaridade alguma com sequências depositadas em bancos de dados, sendo denominadas de “matéria escura viral”, as quais constituem o genoma de vírus ainda a serem descobertos (KRISHNAMURTHY; WANG, 2017). A partir de 2003, uma parte deste universo desconhecido começou a ser desvendado com a descoberta dos vírus gigantes associados a amebas de vida livre (COLSON; LA SCOLA; RAOULT, 2017). Estes vírus colocaram em cheque os critérios até então bem estabelecidos do que caracterizaria um vírus, revivendo debates como o fato de serem entidades pouco complexas, totalmente dependentes de seus hospedeiros e se devem ser considerados seres vivos, merecendo uma posição na tradicional árvore da vida (FORTERRE, 2010; RAOULT; FORTERRE, 2008).

Os vírus gigantes fazem parte de um grupo peculiar de vírus denominado de vírus grandes núcleo-citoplasmáticos de DNA (NCLDV, do inglês *nucleocytoplasmic large DNA viruses*). Esse é um grupo hipoteticamente monofilético dentro da virosfera, inicialmente composto pelas famílias *Poxviridae*, *Iridoviridae*, *Asfarviridae* e *Phycodnaviridae* (IYER; ARAVIND; KOONIN, 2001). Com a descoberta dos vírus gigantes de ameba esse grupo expandiu consideravelmente e foi proposta a criação de uma nova ordem viral para classifica-los oficialmente, denominada “*Megavirales*” (ABERGEL; LEGENDRE; CLAVERIE, 2015; COLSON *et al.*, 2013). Os vírus deste grupo apresentam partículas com diferentes formas e tamanhos, podendo atingir dimensões acima de 1.5 µm de comprimento, como os pithovírus, e genomas complexos contendo centenas de genes de origens e funções variadas, muitos inclusive sem função conhecida, chegando a mais de 2000 kb de extensão, como observado para os pandoravírus (Tabela 1) (LEGENDRE *et al.*, 2014; PHILIPPE *et al.*, 2013). Os vírus deste grupo estão associados a diversos hospedeiros, como

mamíferos, répteis, aves, insetos, algas e protozoários, e exploram diferentes estratégias de penetração para estabelecerem ciclos de infecção produtivos, como fusão de membrana, endocitose mediada por receptor, fagocitose, entre outros (Tabela 1) (GHIGO *et al.*, 2008; MOSS, 2016; WANG *et al.*, 2014). Uma vez no interior das células hospedeiras, estes vírus estabelecem fábricas virais em seu citoplasma, região onde pode ocorrer a replicação e transcrição do genoma e morfogênese viral (DE CASTRO; VOLONTÉ; RISCO, 2013). Enquanto para alguns grupos virais seus membros desenvolvem todo o ciclo de multiplicação no citoplasma hospedeiro, como mimivírus e poxvírus, outros dependem do núcleo de seu hospedeiro para completar sua multiplicação, como os phycodnavírus, marseillevírus e iridovírus, o que caracteriza o perfil núcleo-citoplasmático dos NCLDVs (tabela 1).

Tabela 1: Características gerais dos NCLDVs.

Família/grupo viral	Tamanho da partícula (nm)	Morfologia/ simetria	Tamanho do genoma (kb)	Hospedeiro	Sítio de replicação
Phycodnaviridae	190	Icosaédrico	150-400	Algas verdes e cnidários	Citoplasma e núcleo
Poxviridae	220-450 x 140-260	Tijolo ou ovoide	130-380	Insetos, répteis, aves, mamíferos	Citoplasma
Asfarviridae	175-215	Icosaédrico	170-190	Mamíferos	Citoplasma
Ascoviridae	~300 x ~130	Baciliforme ou ovoide	150-190	Insetos	Citoplasma e núcleo
Iridoviridae	185	Icosaédrico	100-220	Insetos e vertebrados de sangue frio	Citoplasma e núcleo
Mimiviridae	300-1200	(Pseudo) Icosaédrico	730-1510	Amoebozoa, Kinetoplastida, *Cafeteria*	Citoplasma
Marseilleviridae	200-250	Icosaédrico	350-390	*Acanthamoeba*	Citoplasma e núcleo
Pandoravirus	~1000 x ~500	Elipsoide	1800-2500	*Acanthamoeba*	Citoplasma e núcleo (?)
Pithovirus	~1600 x ~500	Elipsoide	610-680	*Acanthamoeba*	Citoplasma
Cedratvirus	~1200 x ~400	Elipsoide	575-590	*Acanthamoeba*	Citoplasma
Orpheovirus	~1300 x ~500	Elipsoide	1474	*Vermamoeba*	Citoplasma
Mollivirus	500-600	Esférico	652	*Acanthamoeba*	Citoplasma e núcleo (?)
Faustovirus	240	Icosaédrico	455-490	*Vermamoeba*	Citoplasma
Pacmanvirus	250	Icosaédrico	395	*Acanthamoeba*	Citoplasma
Kaumoebavirus	250	Icosaédrico	351	*Vermamoeba*	Citoplasma

Fonte: Modificado de Yutin *et al.* (2009) e Colson *et al.* (2017); Viralzone.expasy.org.

Dentre os vírus gigantes conhecidos destacam-se os membros da família *Mimiviridae*, o primeiro grupo de vírus gigantes a ser descrito (LA SCOLA *et al.*, 2003). Os mimivírus chamaram a atenção da comunidade científica pela sua estrutura pseudo-icosaédrica peculiar de aproximadamente 700 nm de diâmetro recoberta por uma densa camada de fibrilas glicosiladas e conteúdo genético até então nunca encontrados na virosfera (Figura 1A) (RAOULT *et al.*, 2004; XIAO *et al.*, 2009). Estas fibrilas são importantes para a adesão viral à superfície das células hospedeiras, tendo relevante papel na etapa inicial do ciclo de multiplicação desses vírus (RODRIGUES *et al.*, 2015). Após adesão, os vírus penetram nas células hospedeiras por meio de fagocitose e liberam o genoma por uma face em formato de estrela, denominado *stargate* (GHIGO *et al.*, 2008; ZAUBERMAN *et al.*, 2008). Os mimivírus estabelecem fábricas virais onde ocorre a replicação do genoma e formação de novas partículas virais, as quais são liberadas por meio de lise celular. A morfogênese viral é um processo complexo: membranas do retículo endoplasmático são recrutadas, as quais formarão a membrana lipídica interna; camadas proteicas são agregadas formando o capsídeo; o genoma é incorporado à partícula em formação através de um orifício presente na face oposta ao *stargate*; e finalmente fibrilas são incorporadas às partículas possivelmente ao passarem pela região final da fábrica viral, referida como “área de aquisição de fibrilas” (ANDRADE *et al.*, 2017).

Estes vírus possuem um genoma de aproximadamente 1.2 Mb composto por mais de 1000 genes, dentre os quais estão presentes elementos que conferem aos mimivírus uma maquinaria de glicosilação exclusiva (HÜLSMEIER; HENNET, 2014; PIACENTE *et al.*, 2015), bem como um maquinário transcricional complexo, composto por diversos fatores de transcrição, RNA polimerases, sinais de poliadenilação e promotores, os quais foram associados a expressão de genes precoces durante a multiplicação viral (BYRNE *et al.*, 2009; RAOULT *et al.*, 2004; SUHRE; AUDIC; CLAVERIE, 2005). Além disso, alguns componentes do aparato traducional, como RNA transportadores (tRNA), além de alguns fatores de tradução e aminoacil-tRNA-sintetases (aaRS), enzimas até então exclusivas dos domínios celulares, são difundidas entre os membros dessa família (ABERGEL *et al.*, 2007; JEUDY *et al.*, 2012; RAOULT *et al.*, 2004). A presença destes genes impulsionou debates acirrados sobre a origem e evolução dos mimivírus, onde alguns autores defendem que estes vírus poderiam ter sido originados de organismos mais

complexos (possivelmente compondo um quarto domínio da vida) e evoluíram por meio de redução genômica (BOYER *et al.*, 2010; CLAVERIE; OGATA, 2009; NASIR; KIM; CAETANO-ANOLLES, 2012), enquanto outros pesquisadores defendem uma segunda hipótese, em que os mimivírus tiveram origem a partir de organismos mais simples e evoluíram por meio de ganho de genes por eventos de duplicação gênica e transferência gênica horizontal (TGH) (KOONIN; KRUPOVIC; YUTIN, 2015; MOREIRA; BROCHIER-ARMANET, 2008; YUTIN; WOLF; KOONIN, 2014). Esse debate ainda permanece vigente e merece novas investigações, principalmente utilizando diferentes genes e estratégias variadas para reconstrução da história evolutiva. Além disso, o descobrimento de novos vírus fornecerá peças importantes para auxiliar nessa discussão.

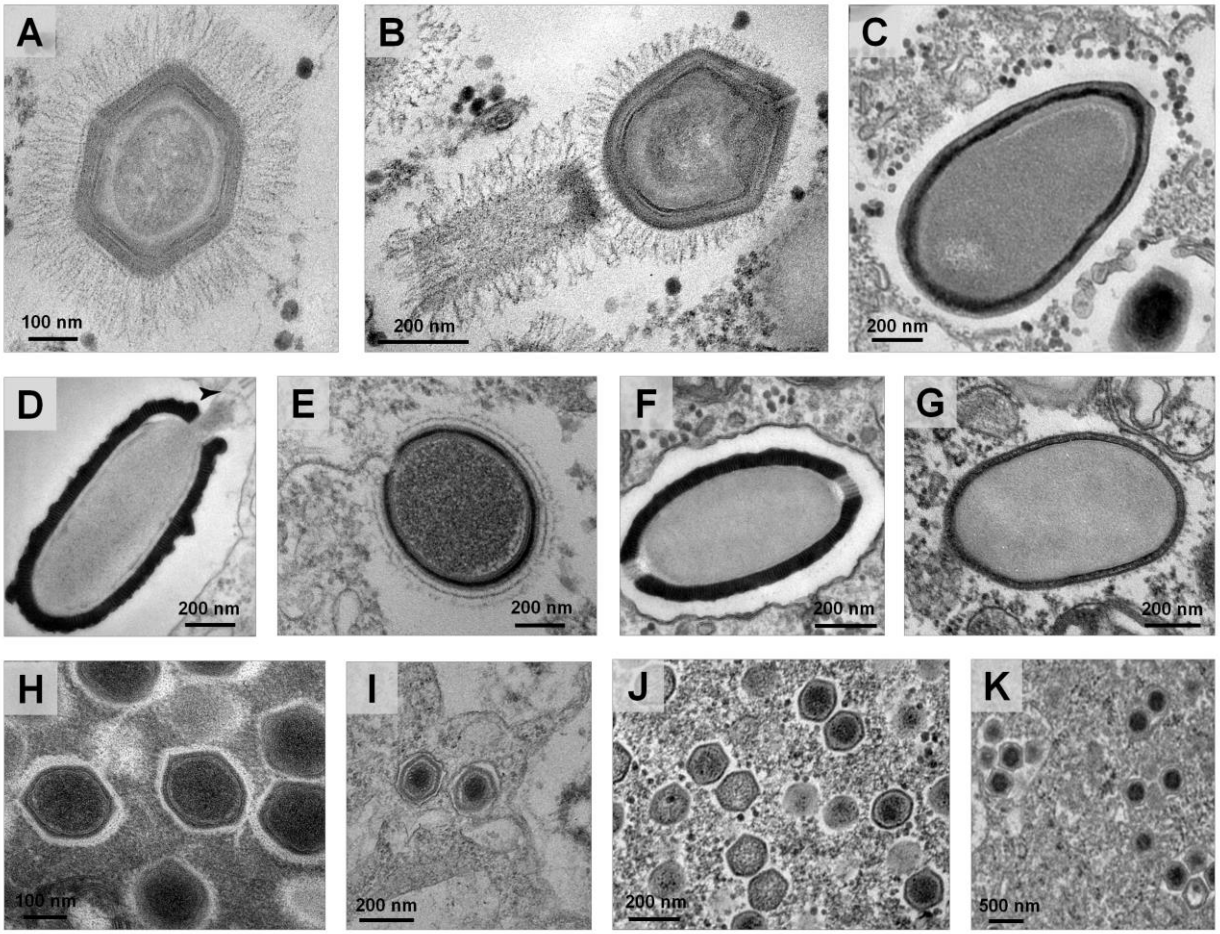


Figura 1: Estrutura dos vírus gigantes conhecidos até o momento. Imagens de microscopia eletrônica de transmissão de diferentes isolados virais, evidenciando as diferenças morfológicas entre os vírus de diferentes grupos. A) Mimivírus; B) Tupanvirus; C) Pandoravírus; D) Pithovírus; E) Mollivírus; F) Cedratvírus; G) Orpheovírus; H) Marseillevírus; I) Pacmanvírus; J) Faustovírus; K) Kaumoebavírus. **Fontes:** Banco de imagens do Gepvig (2018); Andreani *et al*., 2017; Bajrai *et al*., 2016; Legendre *et al*., 2014; Legendre *et al*., 2015.

Nesse sentido, a descrição dos klosneuvírus contribuiu para reforçar a segunda hipótese sobre origem dos vírus gigantes. Estes vírus fazem parte de um novo grupo dentro da família *Mimiviridae*, inicialmente compostos por quatro membros – klosneuvírus, hokovírus, catovírus e indivírus – cujos genomas foram identificados por meio de análises de metagenômica de amostras obtidas de uma estação de tratamento de água residual em Klosterneuburg, Áustria (SCHULZ *et al.*, 2017). Estes vírus ainda não foram isolados, portanto não há informações a respeito de sua biologia. Contudo, seu genoma foi completamente montado e analisado, revelando características inéditas na virosfera. Estes vírus apresentam um vasto arsenal gênico relacionado ao aparato de tradução, em especial o klosneuvírus, o qual apresenta 19 aaRS, e análises filogenéticas destes genes indicaram a ocorrência de múltiplos eventos de TGH, refutando a hipótese de origem viral a partir de um organismo mais complexo (SCHULZ *et al.*, 2017). Este conjunto de genes (aaRS) aumentou ainda mais com a descoberta dos tupanvírus. Estes vírus foram isolados a partir de sedimentos de lagoas salinas do Pantanal e solo oceânico coletados a 3000 metros de profundidade na Bacia de Campos no Brasil (ABRAHÃO *et al.*, 2018). Um vírus foi isolado a partir de cada local, denominados Tupanvirus soda lake (TPV-SL) e Tupanvirus deep ocean (TPV-DO), formando um novo clado dentro da família *Mimiviridae*. Os novos vírus apresentam um capsídeo similar ao observado para os mimivírus, porém possuem uma longa cauda cilíndrica acoplada, resultando em partículas com um tamanho médio de 1.2 µm de extensão, podendo atingir até 2.3 µm devido à plasticidade dessa estrutura (Figura 1B). Diferentemente dos mimivírus, os tupanvírus são capazes de infectar amebas além do gênero *Acanthamoeba* e exibem um fenótipo citotóxico em células hospedeiras e não hospedeiras, o qual está relacionado com um fenômeno de shutdown ribossomal cujo mecanismo ainda permanece desconhecido (ABRAHÃO *et al.*, 2018). Esses vírus possuem o mais completo arsenal de genes relacionados ao processo de tradução já descrito, apresentando diversos fatores de tradução, dezenas de tRNA (até 71 em TPV-DO) e 20 aaRS correspondente a todos os aminoácidos proteinogênicos codificados no código genético padrão. Análises filogenéticas iniciais destas enzimas indicaram que a maioria delas já estaria presentes em um ancestral comum dos vírus gigantes, sugerindo a origem destes vírus a partir de um organismo mais complexo que teria evoluído majoritariamente por perda gênica (ABRAHÃO *et al.*, 2018). Embora muito se tenha avançado com a descoberta destes

isolados no campo dos vírus gigantes, o debate sobre sua origem e evolução ainda é vigente e merece novas investigações. Além disso, os novos vírus descritos ainda não são reconhecidos pelo ICTV e um esforço adicional é necessário para avançar na taxonomia da família *Mimiviridae*.

Além dos mimivírus, outros vírus gigantes de amebas foram descritos nos últimos anos, como pandoravírus, pithovírus, mollivírus, cedratvírus e orpheovírus (Figura 1C-G) (ANDREANI *et al.*, 2016, 2018, LEGENDRE *et al.*, 2014, 2015; PHILIPPE *et al.*, 2013). Dentre estes, os pithovírus e cedratvírus se destacam por apresentarem as maiores partículas virais descritas até o momento. Esses vírus são membros de um novo grupo viral exibindo partículas elipsoides com dimensões acima de 1.0 µm de extensão e genoma circular relativamente pequeno em torno de 600 kb (aproximadamente metade do genoma dos mimivírus e um terço do genoma dos pandoravírus), o qual é liberado através de opérculos presentes nas extremidades das partículas virais (LEGENDRE *et al.*, 2014). O primeiro pithovírus foi isolado em 2014 a partir de amostra de solo congelado siberiano (*permafrost*) datado de 30.000 anos atrás, nomeado Pithovirus sibericum (LEGENDRE *et al.*, 2014). As partículas desse vírus atingem dimensões acima de 1.5 µm de extensão e são menos densas que as partículas dos mimivírus, sendo provável que macromoléculas além do genoma viral estejam presentes em seu interior, embora tal hipótese ainda precise ser verificada (OKAMOTO *et al.*, 2017). Em 2016, um novo isolado foi descrito obtido a partir de amostras de esgoto coletadas na França, denominado Pithovirus massiliensis, e a análise de seu genoma indicou que estes vírus possuem um alto grau de conservação gênica e um perfil de evolução similar ao de procariotos, com uma baixa taxa de substituição de nucleotídeos e mutações por sítio por ano (LEVASSEUR *et al.*, 2016). Neste mesmo ano foram descritos os cedratvírus. Esses novos vírus possuem partículas similares à dos pithovírus, embora um pouco menores (possuem aproximadamente 1.1 µm) e apresentem dois opérculos na partícula viral, enquanto os pithovírus apresentam apenas um (ANDREANI *et al.*, 2016). Os cedratvírus penetram as células hospedeiras por meio de fagocitose e estabelecem grandes fábricas virais onde ocorre a morfogênese viral, a qual envolve estruturas de diferentes formas e uma organização sequencial até a maturação completa das partículas que são posteriormente liberadas por meio de lise celular e possivelmente exocitose (SILVA *et al.*, 2018). O genoma dos cedratvírus apresenta vários genes em comum com os demais NCLDVs e análises

filogenéticas os colocam como grupo irmão dos pithovírus, os quais formariam uma possível nova família "*Pithoviridae*" (ANDREANI *et al.*, 2016; BERTELLI *et al.*, 2017). Por ser um grupo recém-descoberto, pouco ainda se sabe sobre a diversidade e evolução deste grupo. O isolamento de novos vírus e sua caracterização biológica e genômica poderão trazer novas informações para este grupo em expansão.

Além dos vírus gigantes, outros grupos de vírus grandes de DNA foram descritos nos últimos anos infectando espécies do gênero *Acanthamoeba*, como marseillevírus e pacmanvírus, e *Vermamoeba*, como faustovírus e kaumoebavírus (Figura 1H-K) (ANDREANI *et al.*, 2017; BAJRAI *et al.*, 2016; BOYER, M *et al.*, 2009; RETENO *et al.*, 2015). Dentre estes, os marseillevírus são o grupo mais bem caracterizado e compõem a família *Marseilleviridae* (AHERFI *et al.*, 2014; COLSON *et al.*, 2013). Estes vírus foram inicialmente isolados de amostras de água de uma torre de resfriamento de Paris, França, e desde então diversos isolados foram sendo descritos, obtidos de diferentes amostras e regiões do planeta, inclusive no Brasil (AHERFI *et al.*, 2014; BOYER *et al.*, 2009; DORNAS *et al.*, 2016; FABRE *et al.*, 2017; TAKEMURA, 2016; THOMAS *et al.*, 2011). Eles possuem partículas icosaédricas de aproximadamente 250 nm de diâmetro e genoma circular de 370 kb, codificando mais de 400 proteínas. O ciclo de multiplicação destes vírus é rápido, onde novas partículas virais são formadas e liberadas a partir de 8h após infecção (ARANTES *et al.*, 2016; BOYER *et al.*, 2009). Uma vez no interior das células hospedeiras, estes vírus liberam seu genoma, o qual é direcionado ao núcleo celular onde fatores do hospedeiro parecem estar envolvidos em etapas iniciais da transcrição do genoma viral (FABRE *et al.*, 2017). Extensas fábricas virais são formadas onde ocorre a morfogênese viral e a progênie é liberada por lise celular como partículas isoladas ou envolvidas em vesículas membranosas, as quais atingem dimensões micrométricas e podem conter centenas ou mesmo milhares de vírions em seu interior (ARANTES *et al.*, 2016). Estas vesículas infecciosas são importantes também para o início do ciclo, favorecendo a fagocitose dos marseillevírus pelos seus hospedeiros; quando em partículas isoladas, estes vírus exploram a via endocítica para penetração (ARANTES *et al.*, 2016). Com relação ao genoma, grande parte dos genes preditos nos marseillevírus possui função desconhecida e muitos são oriundos de seus hospedeiros e organismos simpátricos presentes em um ambiente intra-amebiano, obtidos por meio de eventos de TGH (BOYER *et al.*, 2009). Mais da metade dos genes é precedido por um motivo

promotor octa-nucleotídico – AAATATTT – cuja importância na expressão gênica foi confirmada experimentalmente, e tal motivo está presente em múltiplas cópias para diferentes genes (OLIVEIRA *et al.*, 2017a). Contudo, não há informações sobre uma possível correlação entre esse promotor e uma expressão gênica temporal como observado para outros NCLDVs, como poxvírus e mimivírus, uma vez que não há informações sobre o perfil transcricional dos marseillevírus (DAVISON; MOSS, 1989; LEGENDRE *et al.*, 2010; OLIVEIRA *et al.*, 2017a). O uso de novas tecnologias para o estudo do transcriptôma, como sequenciamento de RNA, tem sido utilizado para obter importantes dados sobre a biologia de diferentes NCLDVs, e sua aplicação para os marseillevírus poderão trazer novas informações sobre o processo transcricional destes vírus.

2. JUSTIFICATIVA

A virologia avançou consideravelmente desde que foi estabelecida como um ramo da ciência no final do século XIX. Desde a descoberta do vírus do mosaico do tabaco vários outros vírus foram sendo descritos e caracterizados, isolados de diversos ambientes e hospedeiros de todos os domínios da vida. Embora muito se tenha avançado na identificação de novos vírus e nos estudos de suas interações com seus hospedeiros, ainda não se sabe como essa rede de interação está conectada. Além disso, vários estudos de metavirômica têm sido elaborados, identificando diversas sequências virais ao redor do mundo, mas ainda não se tem uma visão clara de como a diversidade viral está distribuída no planeta. Sendo assim, uma análise global da relação entre vírus e seus hospedeiros naturais, permitindo a elucidação da rede de interação entre esses organismos, trará novas informações sobre a virosfera, impulsionando discussões que possibilitarão o avanço da virologia. Dentre os vírus já descritos, os vírus gigantes ganham destaque pela sua complexidade estrutural e genômica. Os mimivírus foram os primeiros a serem descobertos e reascenderam debates sobre a origem e evolução dos vírus, em parte devido à presença de genes envolvidos no processo de síntese proteica até então considerados exclusivos de organismos celulares. Com a descoberta dos tupanvírus, esse conjunto de genes aumentou e impulsionou ainda mais os debates, além de revelarem diferentes características nunca antes vistas na virosfera. A origem e evolução dos vírus gigantes ainda continuam em discussão e novas análises filogenéticas bem como estudos sobre a interação destes organismos com seus hospedeiros poderão trazer novas informações que auxiliarão a elucidar melhor estas questões em aberto. Além disso, estudos de genômica e transcriptômica têm sido amplamente utilizados para desvendar características únicas de diferentes vírus. O avanço dessas áreas aplicadas aos vírus gigantes é de fundamental importância para melhor caracterizar novos isolados e elucidar questões ainda sem respostas para vírus já conhecidos, como os marseillevírus. Finalmente, este conjunto de análises poderá auxiliar no estabelecimento de critérios para aprimorar a classificação dos vírus gigantes, contribuindo assim para o avanço da taxonomia dessas novas e complexas entidades.

3. OBJETIVOS

3.1. Objetivo geral

Elucidar a rede de interação entre a virosfera conhecida e seus hospedeiros e avançar em estudos de sistemática, genômica e transcriptômica dos vírus gigantes.

3.2. Objetivos específicos

- Definir o espectro de hospedeiros naturais dos vírus reconhecidos pelo ICTV;
- Construir um grafo de *network* relacionando vírus e seus hospedeiros naturais;
- Analisar os vírus que afetam os seres humanos e o tropismo destes vírus;
- Analisar a interação entre vírus grandes núcleo-citoplasmáticos de DNA e seus hospedeiros;
- Realizar análises filogenéticas para os genes relacionados ao processo de tradução em mimivírus e tupanvírus;
- Fundamentar os critérios para a criação de um grupo taxonômico para classificar os tupanvírus;
- Realizar análises morfométricas de um cedratvírus isolado no Brasil;
- Montar, anotar e analisar o genoma e a posição filogenética do cedratvírus isolado;
- Realizar o sequenciamento de RNA (RNA-seq) e determinar o perfil transcricional do marseillevírus;
- Avaliar a associação do promotor de marseillevírus com diferentes categorias gênicas.

4. METODOLOGIA, RESULTADOS E DISCUSSÕES PARCIAIS SERÃO APRESENTADOS NA FORMA DE ARTIGOS PUBLICADOS PRECEDIDOS DE UM BREVE RESUMO.

4.1. ARTIGO #1: An anthropocentric view of the virosphere-host relationship

Desde o nascimento da virologia no final do século XIX, diversos vírus foram sendo descritos e isolados de diversas partes do mundo, associados a hospedeiros compreendendo todos os domínios da vida. Contudo, ainda não se tem uma visão clara de como estes vírus estão interligados com seus hospedeiros, nem mesmo como a diversidade viral está distribuída pelo planeta. Por meio de meta-análises extensas, demonstramos neste trabalho que a virosfera conhecida possui um espectro muito restrito de hospedeiros, resultando em uma rede (network) pouco conectada. A análise desta network revelou uma virosfera altamente antropocêntrica, onde a maioria dos vírus conhecidos está associada ao ser humano ou a hospedeiros diretamente relacionados a interesses humanos, sejam econômicos, médicos ou biotecnológicos, independente do grupo de hospedeiros considerados. Além disso, nós demonstramos um contraste entre a ubiquidade de bacteriófagos e a especificidade de alguns grupos virais em determinados ambientes ao redor do planeta baseado em dados de metavirômica, onde discutimos o real avanço da virologia na era da metagenômica. Por fim, revelamos um total de 320 espécies virais com representantes associados com seres humanos, onde a maioria afeta os sistemas tegumentar, respiratório e nervoso, o que nos levou a discutir importantes questões sobre o tropismo destes vírus. Os dados apresentados neste trabalho nos forçam a reconhecer uma verdade inconveniente: após anos de investigação, nós apenas conhecemos uma pequena fração da virosfera, uma vez que temos ignorado tudo ao nosso redor, exceto nós mesmos e aqueles organismos importantes para nós. Isso deixa claro que um grande esforço e uma mudança de perspectiva devem ser realizados para que possamos enxergar um pouco além da ponta do iceberg e de fato avançar nossa compreensão da virosfera.

Este artigo foi publicado no periódico ***Frontiers in Microbiology*** em agosto de 2017.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 August 2017
doi: 10.3389/fmicb.2017.01673

An Anthropocentric View of the Virosphere-Host Relationship

*Rodrigo A. L. Rodrigues, Ana C. dos S. P. Andrade, Paulo V. de M. Boratto, Giliane de S. Trindade, Erna G. Kroon and Jônatas S. Abrahão**

Laboratório de Vírus, Department of Microbiology, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

For over a century, viruses have been known as the most abundant and diverse group of organisms on Earth, forming a virosphere. Based on extensive meta-analyses, we present, for the first time, a wide and complete overview of virus–host network, covering all known viral species. Our data indicate that most of known viral species, regardless of their genomic category, have an intriguingly narrow host range, infecting only 1 or 2 host species. Our data also show that the known virosphere has expanded based on viruses of human interest, related to economical, medical or biotechnological activities. In addition, we provide an overview of the distribution of viruses on different environments on Earth, based on meta-analyses of available metaviromic data, showing the contrasting ubiquity of head-tailed phages against the specificity of some viral groups in certain environments. Finally, we uncovered all human viral species, exploring their diversity and the most affected organic systems. The virus–host network presented here shows an anthropocentric view of the virology. It is therefore clear that a huge effort and change in perspective is necessary to see more than the tip of the iceberg when it comes to virology.

Keywords: virosphere, anthropocentric, virus–host relationship, network, metavirome

OPEN ACCESS

Edited by:
William Michael McShan, University of Oklahoma Health Sciences Center, United States

Reviewed by:
Juliana Felipetto Cargnelutti, Universidade Federal de Santa Maria, Brazil
Jessica Labonté, Texas A&M University at Galveston, United States

****Correspondence:***
Jônatas S. Abrahão
jonatas.abrahao@gmail.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *11 July 2017*
Accepted: *17 August 2017*
Published: *30 August 2017*

Citation:
Rodrigues RAL, Andrade ACSP, Boratto PVM, Trindade GS, Kroon EG and Abrahão JS (2017) An Anthropocentric View of the Virosphere-Host Relationship. Front. Microbiol. 8:1673. doi: 10.3389/fmicb.2017.01673

INTRODUCTION

The virology, as a science field, started at the end of the XIX century with the studies of Adolf Mayer, Dmitry Ivanofsky, and Martinus Beijerinck about tobacco mosaic disease. The investigators noticed that they were dealing with an agent completely unknown to the academic community, which retained its infectious nature even after passing through Chamberland filters (at that time, the most efficient method to retain bacteria). Furthermore, even after being diluted by filtration in a porous membrane, the agent recovered its infectiveness after replication within living tissues of healthy plants. The new pathogen was named "*contagium vivum fluidum*," and only after the advent of *in vitro* plaque assays and electron microscopy it was fully recognized as a virus (Enquist and Racaniello, 2013). Lwoff (1957) published a seminal work in which he established, for the first time, a set of characteristics for an organism to be considered a virus; among them were being an intracellular parasite and completely relying on the biosynthetic machinery of its host, thus being considered a non-living organism. With the advancement of virology, the International Committee on Taxonomy of Viruses (ICTV) was created in the 1960s (originally the International Committee

on Nomenclature of Viruses) with the objective of cataloging and organizing the viruses that were being described in the years to come; it established the first rules for viral taxonomy. A few years later, David Baltimore proposed a strategy to organize the viruses according to the properties of their genetic material, with six groups being defined at that time: I (dsDNA), II (ssDNA), III (dsRNA), IV [ssRNA(+)], V [(ssRNA(−)], and VI (ssRNA-RT) (Baltimore, 1971). In the following years, two additional groups were considered, composing the groups VII (dsDNA-RT) and VIII (viroids). This organization strategy is currently well accepted among virologists.

In the years to come, several viruses were described, being isolated in every corner of the planet from hosts belonging to the three domains of life, i.e., Eukarya, Bacteria, and Archaea. In this context, the virus species concept was created by the ICTV, which is the lowest taxon (group) in a branching hierarchy of viral taxa, defined as a polythetic class of viruses that constitute a replicate lineage and occupy a particular ecological niche (i.e., possess similar biological features) (International Committee on Taxonomy of Viruses - Taxonomy, 2017). These viruses continuously reaffirmed the established criteria raised in the 1950s to recognize an organism as a virus. Only during the last few years this paradigm was broken with the discovery of giant viruses (La Scola et al., 2003; Boyer et al., 2009; Philippe et al., 2013; Legendre et al., 2014). These viruses put the well-established concepts to the test, restoring debates about their complete dependency on their hosts and whether they should be considered living organisms, therefore deserving a place in the metaphorical tree of life (Raoult and Forterre, 2008; Forterre, 2010). Besides, advancements in the field of genomics during the last few years, especially metagenomics (or even metaviromics), have allowed the identification of countless viral sequences in several regions of the globe, supporting previous electron microscopy data which suggested the viral ubiquity and an astronomical number of viruses on Earth, thus forming a virosphere (Suttle, 2005; Kristensen et al., 2010).

Although the identification of new viruses and studies of their interaction with hosts have considerably advanced, we still do not know how this interactive network is truly connected. Moreover, many metaviromic studies have been developed allowing the identification of different viral sequences around the world, but we do not have a clear vision of how the viral diversity is distributed on the planet, or how much we have searched for new viruses. Therefore, a new look into what is currently available and the use of new strategies to explore these data could bring new insights and allow the advancement of the virology field. Through extensive meta-analysis of currently available data, we demonstrate here that the known viruses have a very narrow host range, resulting in a spatially connected network. We found a highly anthropocentric view of the virosphere and demonstrated the existence of some specific viral groups in certain environments on the Earth, leading us to reflect about how far we have progressed in the study of viruses. Finally, we analyzed the diversity of human-associated viruses and the tropism of these viruses. The results presented here show a highly biased virology, confirming that we know only the tip of the iceberg and a lot of work remains to be done so we can have a clearer view of the diversity and ecology of the virosphere.

MATERIALS AND METHODS

Dataset Preparation and Selection Criteria

Virosphere and Hosts

To analyze the host range of the known viruses, only those officially recognized by the International Committee on Taxonomy of Viruses (ICTV) were included in the analysis. The definition of the best dataset to perform this analysis comprises a challenging task. In this context, ICTV proved to be the best option for gathering the largest and most updated dataset of recognized virus species, grouping and reflecting the diversity and circulation of viruses in nature. A list containing all of the virus species was downloaded from ICTV website[1]. A list released on May 26th, 2016 was used. Therefore, new viruses classified by means of metagenomic data, following the new criteria recently approved by the Executive Committee of ICTV (Simmonds et al., 2017), as wells as the reclassification of the family *Bunyaviridae,* were not considered in this analysis. We considered hosts those organisms in which we found consistent and recurrent evidences of the detection of a virus in a given species by means of isolation, serology, and molecular detection. This detection was associated in most cases with clinical manifestation and, in a few cases, in a non-disease context. Organisms used as study models were not considered here. Hosts were associated with each virus at the lowest taxonomic level possible using the Virus–Host Database (Mihara et al., 2016), VIDE database[2], and full research articles related to a given virus. In the latter, only one reference was used to determine the host species, even though more than one study (whenever available) was analyzed to corroborate the reference used. During our research and analyses, we considered (whenever the data were available) different viruses within a virus species and their host-range. Only the viruses in which it was possible to determine the hosts at species or genus taxonomic level were considered for the construction of the network. A total of 4497 nodes were included in the network dataset, classified as virus, animalia, plantae, fungi, protist, bacteria, and archaea, along with 4814 edges directly connecting the nodes, all with weight $(w) = [1]$.

Viral Diversity

To analyze the known viral diversity on the planet, we considered viral groups (families recognized by the ICTV or groups currently unassigned to a proper taxa) identified in diverse metavirome studies performed in the following environments: marine [10], freshwater [7], soil [6], hypersaline [5], thermal springs [4], sewage [4], and polar water [3], in a total of 39 works. The studies were accessed at National Center for Biotechnology Information

[1]https://talk.ictvonline.org/files/master-species-lists/

[2]http://sdb.im.ac.cn/vide/sppindex.htm

(NCBI)[3] using the name of the environments added by virome or metavirome as keywords in the search field. All of the viral groups identified were included in the network analysis, where they were associated with the environments in which they were detected. A total of 103 nodes were included in the network graph, classified according to the analyzed environments and viral order recognized by the ICTV [*Ligamenvirales, Tymovirales, Herpesvirales, Caudovirales, Picornavirales, Mononegavirales, Nidovirales*, and those not classified in order (Unassigned)], and 260 edges indirectly connecting the nodes, with $w = [1]$. To better visualize the viral groups shared between different environments, we created a circular layout image using Circos package (Krzywinski et al., 2009). In addition to the detected viral groups, we computed the type of technology used for nucleic acid sequencing, the type of material analyzed (DNA or RNA), and whether a 200 nm filter was used for sample preparation.

Human Viruses and Viral Tropism

The viruses that affect humans were defined after the association of the hosts of each virus species recognized by the ICTV, as described above. The viruses were associated with the following organic systems, according to the clinical manifestation reported in cases of infection: digestive, integumentary, respiratory, nervous, muscular, skeletal, cardiovascular, urinary, reproductive, lymphatic, immune, endocrine, or none of them, in cases of non-pathogenic viruses, based on clinical manifestation and/or tropism for a particular body tissue. Clinical manifestation and the tropism for each system were defined according to full research articles found at NCBI and using the arboviruses catalog of the Center for Disease Control and Prevention[4]. The viruses were associated with different systems in a bipartite network composed of 333 nodes classified according to the organic systems and viruses, and 497 edges indirectly connecting the nodes, with $w = [1]$. In parallel, we built a unipartite network graph wherein the systems were interconnected according to the viruses that affect different systems simultaneously, in a total of 12 nodes and 42 edges indirectly connecting the nodes, with $w = [1,25]$.

Construction of Networks

The networks presented in this work were built using the program Gephi version 0.9.1 (Bastian et al., 2009). All components of the each graph were listed in a comma-separated values (.csv) spreadsheet, which was imported to the software. Another .csv spreadsheet containing the connections between the components was also imported to generate the raw graph. In all networks, the node diameter is directly proportional to the edge degree. The thickness of the edges is directly proportional to the number of times that a node is connected to another, wherein different weights were assigned to the edges. The layout was generated using algorithms based on force of attraction and repulsion of the nodes (Fruchterman-Reingold followed by ForceAtlas 2), followed by local rearrangement of the nodes for a better visualization of the connections between nodes, without perturbing the general layout of the networks.

[3]https://www.ncbi.nlm.nih.gov/pubmed/

[4]https://wwwn.cdc.gov/arbocat/

RESULTS AND DISCUSSION

The Known Viruses Have a Very Narrow Host Range

The ICTV is the organization responsible for cataloging and classifying viruses into virus species that have been described over time. Historically, this organization has taken into consideration several criteria for a new isolate to be considered a new species, such as the genetic material and the hosts in which it was isolated, as well as any clinical manifestations it may possibly cause (Simmonds et al., 2017). Viral taxonomy covers the levels of order, family (and subfamily in some cases), genus and species, wherein the vast majority of virus species remain outside of a virus order. All of this information is constantly updated by the ICTV, which periodically publishes the Master Species List (MSL). In this work, we evaluated the host range of all known viruses with a virus species officially recognized and published by the ICTV on May 26th, 2016 (MSL#30) [**Supplementary Table S1**]. An extensive search using public databases and indexed publications was performed to define the natural hosts of all of the viruses present in the list (see Materials and Methods). The majority of the viruses present in the MSL#30 (a total of 3704 virus species, henceforward named the known virosphere) comprises group I (dsDNA) and IV [ssRNA(+)] according to Baltimore's classification [35 and 28%, respectively, followed by group II (ssDNA – 17%)], with the remaining groups representing 20% of the known virosphere (**Figure 1A**). It was possible to associate hosts at the species or genus level to 3414 viruses (92.2%), at the family level or higher to 265 viruses (7.15%), and it was not possible to associate any host for only 25 viruses (0.65%), either because the natural hosts for the viruses are not yet known, or due to a complete lack of information in the literature about their host range (**Figure 1B**). For all viral groups, according to Baltimore's classification, the host range is very restricted, with more than 50% of known viruses infecting only one or two host species, reaching up to 75% in some groups, such as those viruses with genomes composed of dsDNA, ssDNA, ssRNA-RT, and viroids (**Figure 1C**). Only the ssRNA(−) viruses seems to possess a slightly broader host range, wherein 42% of the viruses are able to infect more than four host species. Considering the entire known virosphere, 73.3% are associated with only one or two host species; 3.5% with three or four species; 22.5% with more than four species; and only 0.7% have a natural host range which has not been defined (**Figure 1C**). These analyses reveal that, until now, based on the available information we have, viruses have a very narrow host range. This disturbing data must be interpreted carefully. It is likely that several unknown viruses have a broader host-range, which will drastically change the view presented here; however, we might be far from acquire this kind of knowledge since these relationships are likely out of scope of human investigation. Therefore, in light of the research performed so far, we are facing such suspicious data.

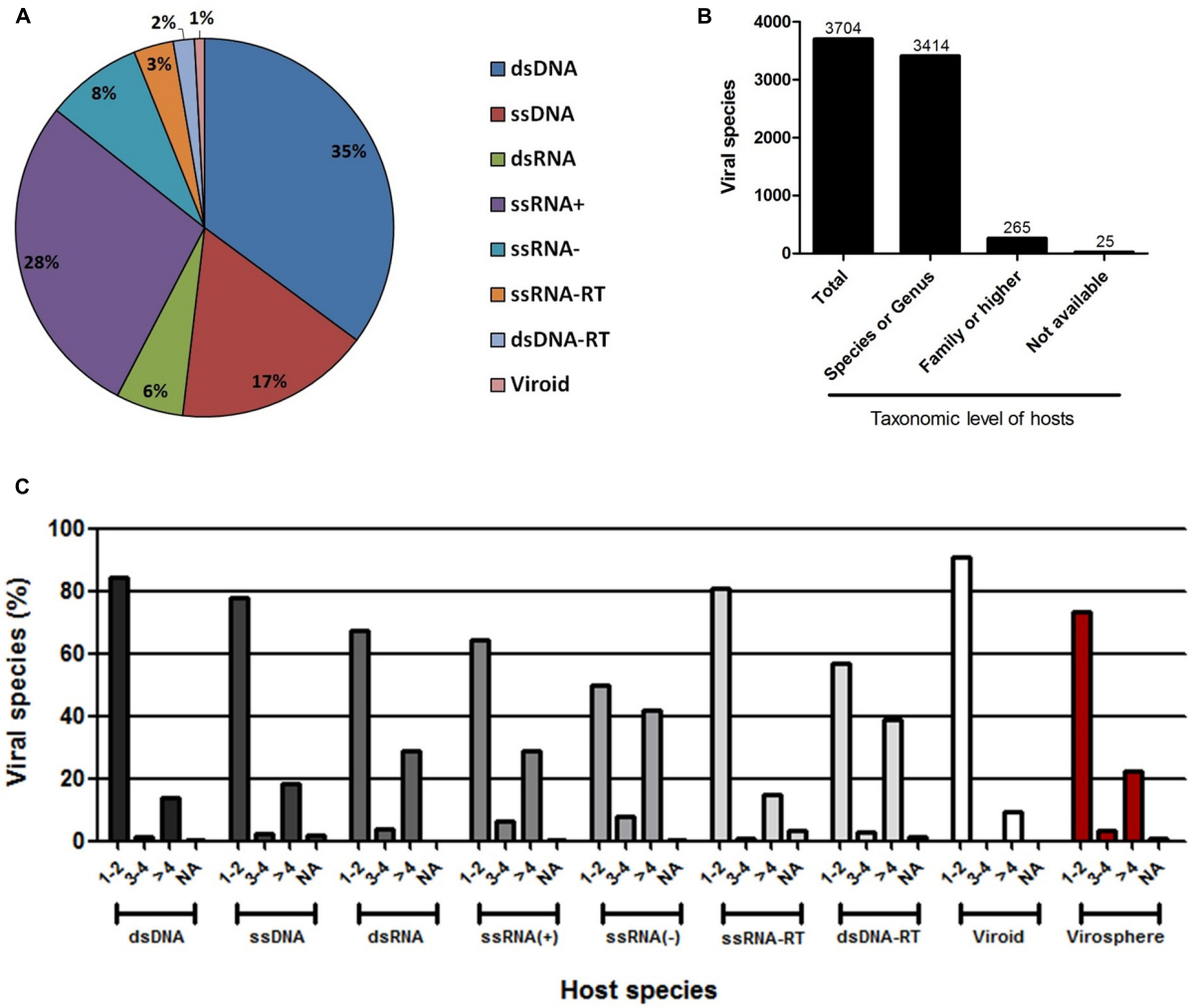


FIGURE 1 | Host range of the known virosphere. **(A)** Pie chart showing the distribution of the viruses recognized by the International Committee on Taxonomy of Viruses (ICTV) according to Baltimore's classification. **(B)** Taxonomic level of the hosts associated to the known viruses. More than 90% of the viruses were associated to hosts at species or genus taxonomic level, which were used in following analysis. **(C)** Amount of host species for viruses according to Baltimore's classification, showing a very narrow host range of the viruses. NA, not available.

An Anthropocentric View of the Known Virosphere

To better represent the interaction between the viruses and the hosts so that we can have a clear vision of how interconnected these organisms are, we built a bipartite network graph composed of 4497 nodes, with 3414 viruses (only viruses associated with hosts at species or genus taxonomic level were included in this analysis) and 1083 hosts (at genus level), all connected by 4814 edges with the same weight (w) = [1]. The hosts were classified according to the major realms and domains of life: Animalia, Plantae, Protist, Fungi, Bacteria, and Archaea (Woese, 2002). We observed a spatially connected network, wherein only a few hosts were associated to a huge amount of viruses, while the majority of the hosts are associated with a few viruses, a reflex of the very narrow host range of the known virosphere (**Figure 2**). Furthermore, the analysis of the network revealed a highly anthropocentric virosphere, in which most viruses are associated with humans or hosts that are directly related to humans by economic, medicinal or biotechnological interests. The vast majority of known viruses are associated with plants (483 genera) or animals (467 genera). These groups are more interconnected than others, even though more than 70% of these hosts possess only one or two associated viruses (**Supplementary Figure S1**). It is noteworthy that some viruses can cross broad host categories, infecting both plants and animals. These viruses are plant pathogens transmitted by arthropod vectors, in which are able to fully replicate and reach the plant host (Dietzgen et al., 2016). Bacteria-infecting viruses (known as bacteriophages or phages) are mainly distributed among the families *Myoviridae*, *Podoviridae*, and *Siphoviridae* (order *Caudovirales*), and are associated with 62 known host genera. This group is spatially connected, reflecting the narrow host range of phages. However, different to animals and plants, almost 40% of known bacteria

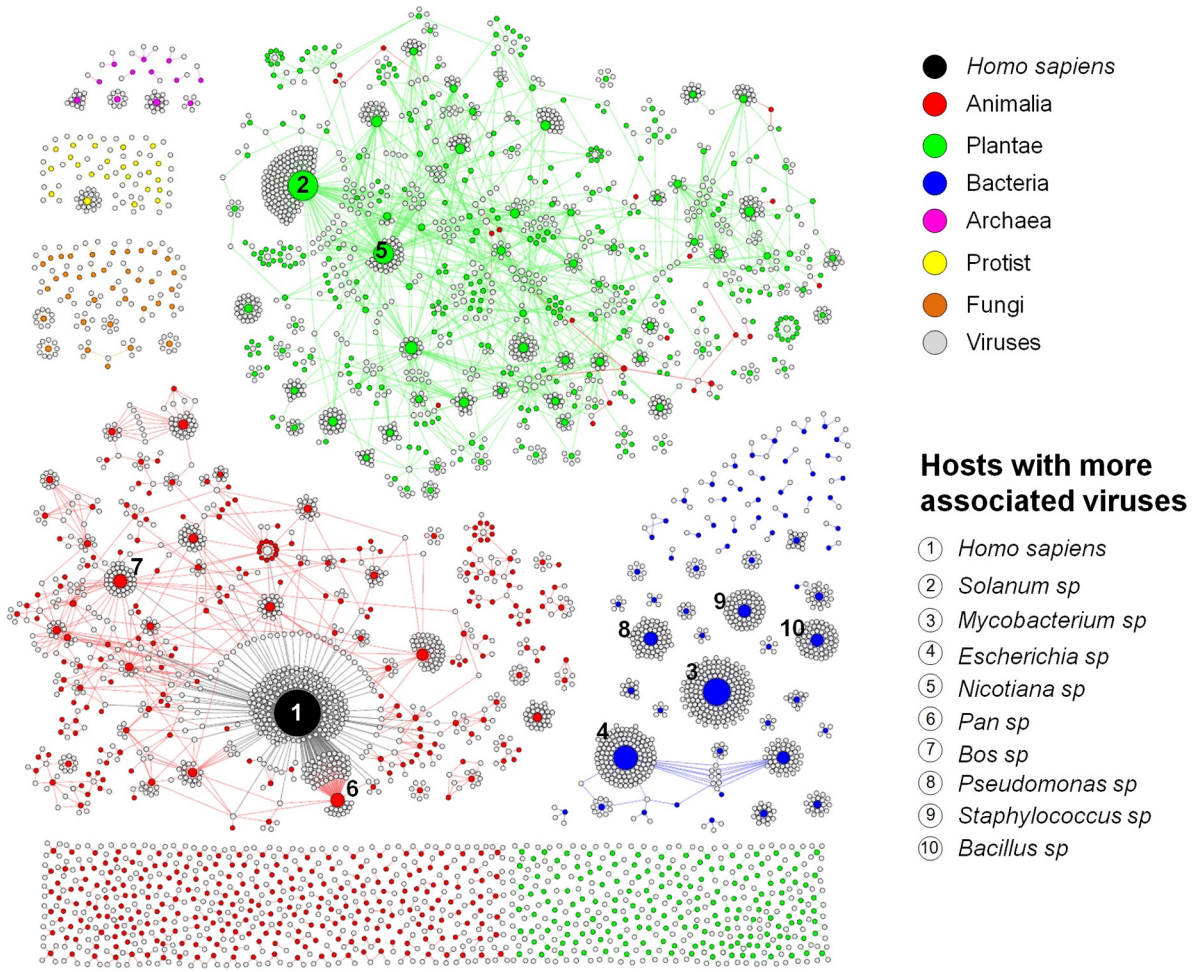


FIGURE 2 | Virus–host interaction network. Bipartite network graph showing a spatially connected network among viruses and hosts, a reflection of our limited knowledge about the viruses and their hosts. Each node represents a virus (gray), or a host genus, classified according to the taxonomic group (colored nodes). The nodes' diameter is proportional to the edge degree. The layout was generated using a force based algorithm followed by manual rearrangement to a better visualization of the connections. A total of 3414 viruses and 1083 hosts (genus level) are represented. The 10 hosts with more associated viruses are depicted.

are infected by more than four viruses. Some bacteria comprised hubs in the network, such as *Mycobacterium* and *Escherichia*, with several associated viruses. Since they are intensively studied due to their medicinal and biotechnological relevance (Korb et al., 2016; Vila et al., 2016), it was expected that a large number of viruses would be identified as parasites of these groups. In fact, a large majority of phage sequences available in GenBank was isolated from a few groups of bacteria associated to human diseases or food processing (Holmfeldt et al., 2013). The knowledge about viruses affecting fungi, protists and archaea is scarce, probably due to the lack of investigation of these groups of viruses and their hosts. These viruses were associated with 36 genera of fungi, 23 protists, and only 12 genera of archaea, reflecting how poorly these microorganisms are studied under the lens of virology.

Among the host genera of each group that possess more associated viruses, many are composed of domesticated species such as *Bos* sp., *Sus* sp., and *Gallus* sp. (Animalia; e.g., cattle, swine, and chickens, respectively); *Solanum* sp., *Nicotiana* sp., *Phaseolus* sp., *Capsicum* sp., and *Cucumis* sp. (Plantae; e.g., potato, tobacco, common bean, peppers, and cucumber, respectively); *Chlorella* sp. (Protist); and *Saccharomyces* sp. (Fungi) (**Supplementary Figure S2**). Many species of these groups are employed in farming, such as cattle, pigs and poultry, as well as many grains and legumes consumed worldwide, handling billions of dollars annually (Thornton, 2010; Reganold and Wachter, 2016). In addition, some species of green algae (*Chlorella* sp., *Chlorophyta* phylum) are used as dietary supplementation as sources of vitamins and macronutrients and its efficacy against some human diseases are under constant investigation (Ebrahimi-Mameghani et al., 2016; Panahi et al., 2016). Yeasts of the *Saccharomyces* genus, especially *S. cerevisiae*, are considered domesticated fungi, being used worldwide in the production of alcoholic beverages, also making them economically important (Sicard and Legras, 2011; Gallone et al., 2016). Given the economic relevance of these organisms,

constant efforts are made to reveal parasites that might be considered a threat to them, thus enabling possible strategies of control and prevention to be established. Therefore, it was expected that these groups of hosts had more known viruses.

Other hosts are known due to their medicinal relevance for humans or animals and commercially explored plants, such as *Acanthamoeba* sp. and *Trichomonas* sp. (Protist), both related to severe infections in humans (Siddiqui and Khan, 2012; Menezes et al., 2016); *Heterobasidion* sp., *Cryphonectria* sp., *Rosellinia* sp., and *Ophiostoma* sp. (Fungi), groups of fungi related to diverse plant infections, both domesticated and from native forests, causing severe diseases such as annosum root and chestnut blight (Hillman and Suzuki, 2004; Ďurkovič et al., 2013; Kondo et al., 2013; Vainio and Hantula, 2015); and *Mycobacterium* sp., *Escherichia* sp., *Pseudomonas* sp., *Staphylococcus* sp., and *Bacillus* sp. (Bacteria), all groups of prokaryotes related to life-threatening diseases, such as tuberculosis (Korb et al., 2016), gastrointestinal, respiratory and urinary infections (Langan et al., 2015; Vila et al., 2016), and also used as biological weapons (Goel, 2015). Therefore, it is expected that these species are the target of intense investigation, and the majority of known phages are associated with these bacteria. Finally, some hosts are important in the biotechnology field or used as laboratory study models for molecular biology, such as *Ectocarpus* sp. (Protist) (Lipinska et al., 2016); *Sulfolobus* sp., and *Thermus* sp. (Archaea) (Cava et al., 2009; Zhang et al., 2013) (**Supplementary Figure S2**). Altogether, the data presented here show that in all group of hosts, both eukaryotic and prokaryotic, most of the known viruses are related to hosts that are important for humans in certain aspects. In this way, the virus–host network shows a highly anthropocentric view of the virology performed so far. This biased virology is probably the very reason for our view of a narrow host-range of the known viruses.

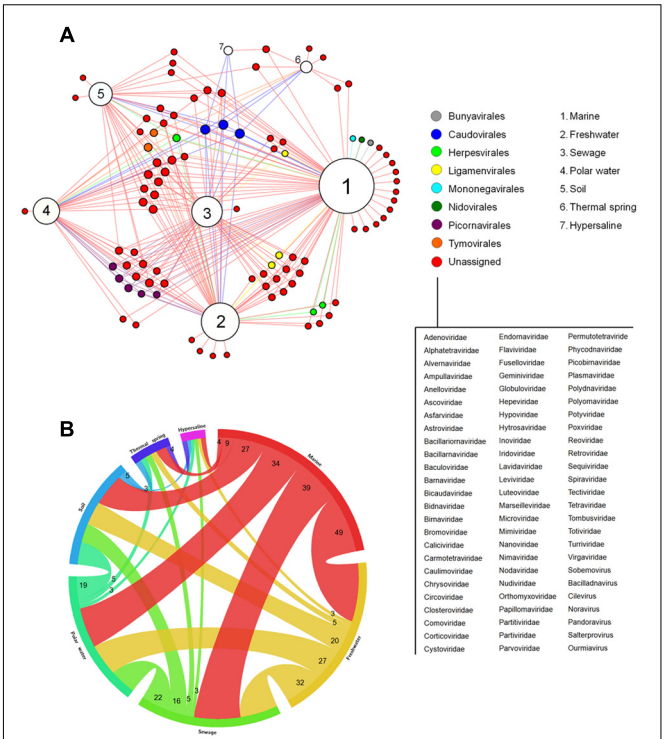


FIGURE 3 | Viral diversity on Earth. **(A)** Network graph showing the viral groups detected in different environments by metaviromic analysis. Each node represents an environment (white) or viral groups (families or known viral genus – colored nodes) classified according to the orders formally recognized by the ICTV. The viruses not currently assigned in any order are listed. The node diameter is proportional to the edge degree. The layout was generated using a force based algorithm followed by manual rearrangement for a better visualization of the connections. A total of 96 viral groups are represented. **(B)** Relationship between the different environments based on the amount of shared viral groups.

Viral Diversity on Earth

Since the discovery of the tobacco mosaic virus at the end of XIX century, many other viruses have been described and biologically characterized in many regions of the planet, thus contributing to the concept of viral ubiquity. With advances in electron microscopy techniques, many studies have been conducted in order to define the abundance and diversity of viruses, coming to an astronomic number, in the order of 10^{31} viral particles on the Earth (Suttle, 2005). However, only with the advent of massive parallel sequencing of nucleic acids and the development of a new research field – metagenomics – it was possible to create a better view of the viral diversity on the planet, reaffirming the viral ubiquity concept (Kristensen et al., 2010).

By analyzing different available metagenomic works, more specifically metaviromic works (analysis of viral nuclei acid sequences in different environments), we built a bipartite network graph connecting the viral groups found within seven distinct environments around the planet: marine, freshwater, polar water, thermal springs, hypersalines, and sewage (**Figure 3A**). A total of 39 works were analyzed (for choice criteria, see Materials and Methods). A total of 96 viral groups (genus or family) were detected in those studies. Different amount of viral groups are shared among the environments, wherein marine shared up to 49 viral groups with other environments, reinforcing the ubiquity of viruses on the planet (**Figure 3B**). Among the viral groups identified, only representatives of the families *Myoviridae*, *Podoviridae*, and *Siphoviridae* (phages belonging to the order *Caudovirales*) were found in all of the searched environments. After the initial studies of metagenomics in marine environments, in which they searched basically for bacteriophages, the hypothesis "Everything is everywhere but environment selects" was applied to these viruses, stating the ubiquity of the phages, even though some groups were specifically found in certain environments (O'Malley, 2008; Thurber, 2009). Our meta-analysis corroborates this hypothesis and goes further, showing that head-tailed phages are found in every location investigated, not only in marine samples. In contrast, the majority of viral groups were found only in two or three environments, and surprisingly, some groups were also restricted to only one environment (**Figure 3A**). The viral diversity is higher in marine environments, wherein 15 groups were exclusive to it. The great diversity of viruses in the oceans is a reflection of the abundance of hosts found there, but also reflects the number of studies

performed, covering all of the oceans and many important seas around the globe, such as the Mediterranean, the Baltic and the Arctic (**Supplementary Table S2**). As expected, extreme environments, such as thermal springs (high temperatures) and hypersalines (high osmolarity), were those with the lowest viral diversity, with only 11 and four viral groups found in each, respectively. The families *Globuloviridae* and *Spiraviridae* were detected exclusively in thermal springs. The viruses of these families infect hyperthermophilic archaea, which are highly abundant in hot springs, thus explaining the exclusivity of those viruses in these environments. No viral group was exclusive to hypersaline environments. Curiously, viruses belonging to the families *Sphaerolipoviridae* and *Pleolipoviridae* (archaea-infecting viruses) have already been isolated and characterized from extreme environments (Luk et al., 2014); however, representatives of these groups were not detected by metaviromic approaches so far.

The absence of some viral groups in certain metaviromic studies might be due to the employed methodology, either in the sequencing platform/method and bioinformatic pipelines, in the type of genetic material that was analyzed (DNA or RNA), or even (and mainly) the procedures employed in the preparation of the samples for sequencing. The vast majority of studies target DNA viruses and use 0.2 μm porous filters during the processing of the collected samples (**Supplementary Table S2**). These strategies restrict the detection of a large part of the viruses (those with RNA genome) and also the giant DNA viruses (Halary et al., 2016), thus making a change in the protocols for the preparation of samples for metaviromic approaches necessary. Nevertheless, it is important to emphasize that the majority of the sequences found in metaviromic studies has no similarities with known sequences available from public databanks. This demonstrates that although the emergence of metagenomic techniques greatly contributed to the discovery of new viruses, even leading the ICTV executive committee to recently approve the use of such information for viral classification (Simmonds et al., 2017), the works on isolation and characterization, both genomically and biologically, should continue and be encouraged. With the association of biological/virological and metaviromic approaches, we might have new insights into the real diversity and distribution of viruses on Earth.

Human-Associated Viruses and Viral Tropism

Since human species is the one with more associated viruses officially recognized by the ICTV among all of the hosts analyzed here, the next step was to turn our attention to these viruses. Until recently, it was thought that about 200 viruses were associated with infections in humans, some with no direct evidence of causing any disease (Woolhouse et al., 2012). Here, we demonstrate that among the known virosphere, 320 virus species are related to human infections (**Supplementary Table S3**). Among them, 146 (45.6%) infect only humans; 116 (36.2%) infect humans and other mammals, some considered important zoonosis, such as rabies (*Rabies lyssavirus*), poxviruses (*Orthopoxvirus*), and hantaviruses (*Hantavirus*) (Shchelkunov, 2013; Jackson, 2016b; Jiang et al., 2017); and 58 (18.2%) are arboviruses (viruses transmitted by arthropods, including mosquitoes, sandflies and ticks) (**Figure 4A**). These viruses are classified within 26 families, wherein *Anelloviridae*, *Bunyaviridae*, and *Papillomaviridae* are the most significant, gathering 44% of the human viruses (**Figure 4B**). These viruses are highly variable, both structurally and genetically, using different replicative strategies. Although all groups of Baltimore's classification possess representatives of human viruses [except for viroids that infect only plants (Steger and Perreault, 2016)], the majority belong to groups I–V, with retroviruses accounting for less than 3% of viruses (**Supplementary Table S3**). Although they are the minority among human viruses, retroviruses were central to the emergence of mammals, thus also to humans, being pivotal components in placenta development (Chuong, 2013). In addition, the human immunodeficiency virus (HIV), the main representative of the group, is one the main life-threatening pathogens, being responsible for immunosuppressive conditions, paving the way to numerous severe secondary infections such as tuberculosis, systemic mycosis, Kaposi sarcoma, among others (Miceli et al., 2011; Godfrey-Faussett and Ayles, 2016; Govindan, 2016).

Many viruses are responsible for severe clinical manifestations, while others are related only to mild symptoms of disease or even asymptomatic infections. To have a better view of the tropism of human viruses and the most affected organic system, we built a network graph associating the viruses with different systems of the human body, according to clinical manifestations related to different viral infections. The viruses that have no direct evidence of causing disease were also included in the analysis. The integumentary, respiratory, and nervous systems were the main affected systems, with 92, 72, and 58 associated viruses, respectively (**Figure 4C**). The integumentary and respiratory systems are the most exposed to infection by different micro-organisms, since they are in direct contact with the environment, thus being expected to be the most affected by viruses. It is noteworthy that many viruses that affect the respiratory tract also affect the muscular system, a reflection of the viruses that cause only flu-like symptoms (**Supplementary Figure S3**). Unlike the two first systems, the nervous system is not directly exposed to the environment, thus making it curious that it is the third most frequently affected system by viruses. Since it is an extremely important and delicate system of the human body, several studies have been conducted to elucidate possible threats for its components, leading to the identification of a considerable range of viruses associated with diseases of the nervous systems. Many of these viruses are associated with severe cases of encephalitis and meningitis, such as herpesviruses (Granerod et al., 2010), lyssaviruses (Jackson, 2016a), and flaviviruses (Daep et al., 2014) (**Supplementary Table S4**), which is why they are target of intense investigation, to better understand the biology of these viruses, thus allowing the development of control mechanisms and possible treatments for diseases. Many of the viruses of the nervous system also affect others, mainly the respiratory and integumentary systems (**Supplementary Figure S3**). In that sense, some viruses are considerable pantropics, affecting

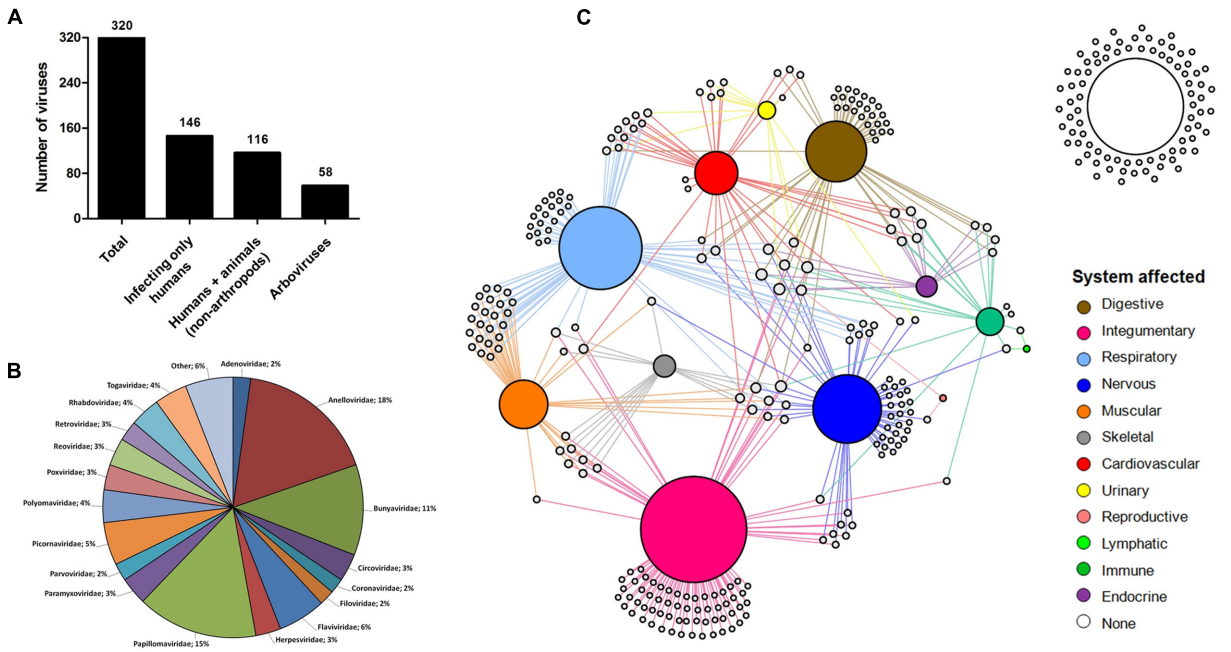


FIGURE 4 | Human viruses and affected systems. **(A)** Human-affecting viruses divided among infecting only humans, infecting humans and other mammals, and arboviruses. **(B)** Pie chart showing the classification of the viruses. A total of 27 groups are represented in the chart. Others: *Deltavirus, Hepadnaviridae, Hepeviridae, Caliciviridae, Picobirnaviridae, Pneumoviridae, Arenaviridae, Orthomyxoviridae, Astroviridae*. **(C)** Network graph showing the viral tropism. Each node represents a virus (white) and an organic system of the human body (colored nodes). The node diameter is proportional to the edge degree. The layout was generated using a force based algorithm followed by manual rearrangement to a better visualization of the connections.

different systems simultaneously, such as ebolavirus, dengue virus and rubella virus, affecting the cardiovascular (hemorrhagic fever), muscular (myalgia), skeletal (arthralgia), and nervous (encephalitis) systems, among others (**Supplementary Table S4**).

The reproductive and lymphatic systems are the least affected by viruses. The first is affected by only two viruses (mumps virus and Rio Bravo virus), responsible for cases of orchitis and oophoritis (Volkova et al., 2012). Although the herpesviruses and papillomaviruses are commonly associated with infections in the reproductive system, where they cause ulcerative lesions and warts in genital regions, we associated these viruses to the integumentary system, since their tropic site of infection is epidermal cells and not specific organs belonging to the reproductive tract. The lymphatic system has also only two associated virus species (*Human gammaherpesvirus 4* and *Primate T-lymphotropic virus 1*), both related to lymphoma cases. Although some viruses trigger lymph node inflammation, these are not considered the tropic site of infection for most viruses, so they are excluded from this analysis. It is possible that other viruses are related to these systems, as well as others included in this network, but further investigations are required. More studies are necessary regarding these systems, thus we can identify the viruses with tropism for these sites. Finally, 83 (26%) viruses analyzed in this work are not connected to any system since they are not related to any known disease so far (**Figure 4C**). The majority of these viruses belong to the family *Anelloviridae* (67.5%), which is mainly composed of the torque teno viruses. These viruses are present in most parts of people, as many metaviromic studies have demonstrated, but there is still no consensus that they carry any kind of loss for our health. As far as we know, they are part of the human virome along with many bacteriophages (Rascovan et al., 2016). Along with the anelloviruses, others have already been detected in human beings by metagenomic approaches, where the association with any disease remains under discussion, such as the giant mimiviruses and marseilleviruses (Popgeorgiev et al., 2013). While there is some evidence linking these viruses with human pathologies, we are still far from ending this debate.

CONCLUSION

It has been more than a century since the discovery of the first viruses. During this time, we have seen great advances in cellular and molecular biology and genetics, which have boosted achievements in the field of virology. Nevertheless, the results presented here show us that, even with great advances, we still know only a tiny fraction of the viral universe, mainly regarding the virus–host interaction. The discovery of giant viruses during the last decade was essential for us to realize how diverse and intriguing the virosphere is, triggering the search for new viruses in hosts completely ignored in the lens of virology. A break of concepts was established after those discoveries, taking us to think again what a virus is and what else is waiting to be discovered. Moreover, the advent of metaviromics had a unique contribution to the expansion of our knowledge about

the virosphere, mainly on the diversity and distribution of these microorganisms, but also with the discovery of new viruses (Alavandi and Poornima, 2012; Shi et al., 2016). However, we are still unable to define the host range of these new viruses with enough accuracy based only on genomic data. In that sense, the improvement of viral isolation techniques is important so that we can look deeper into how these new organisms interact with their hosts and the environment which they inhabit.

The analyses shown here provide a picture of what we know about the entire virosphere and their hosts, and confirm the anthropocentric view of the virology so far. It is likely that the network presented here (**Figure 2**) is largely more interconnected. However, further studies should be performed, especially searching for viruses in hosts that are not of primary human interest, such as environmental fungi and archaea, or even plants and animals that have no added medicinal or economic value. It is an arduous work, but with the improvement of viral isolation techniques and metaviromics, both fundamental tools to this task, it will be possible to continuously add new pieces to fulfill the virus–host network, providing a broader view of the viral universe. In that moment, possibly when science would once again be performed and applied to the understanding of the nature rather than serving the exclusive interests of human beings, we might see beyond just the tip of the iceberg.

AUTHOR CONTRIBUTIONS

RR, AA, and PB prepared the dataset. RR performed the analysis. RR wrote the manuscript. GT, EK, and JA designed the study. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by 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).

ACKNOWLEDGMENTS

We would like to thank our colleagues from Laboratório de Vírus of Universidade Federal de Minas Gerais. JA, GT, and EK are CNPq researchers. JA, EK, RR, and PB are members of a CAPES-COFECUB Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01673/full#supplementary-material

FIGURE S1 | Amount of viruses associated by hosts (at genus level) separated by taxonomic group of the hosts. The total amount of hosts is depicted in the top of each column.

FIGURE S2 | The five hosts with more associated viruses for all six major taxonomic groups, evidencing that most of them is related to human interests. **(A)** Animalia, **(B)** Plantae, **(C)** Protist, **(D)** Fungi, **(E)** Bacteria, **(F)** Archaea. d, domesticated host; i, infection related host; b, biotechnology application host.

FIGURE S3 | Unipartite network graph showing the connections between organic systems according to the viruses that have tropism for more than one system. The nodes' diameter is proportional to the edge degree. The layout was generated using a force based algorithm followed by manual rearrangement to a better visualization of the connections. The thickness of the edges is proportional to the number of viruses that affect the two systems it connects.

TABLE S1 | Viruses and their hosts.

TABLE S2 | Technical information of metaviromic works.

TABLE S3 | Human-infecting viruses and other animals.

TABLE S4 | Tropism of human-infecting viruses and clinical manifestation.

REFERENCES

Alavandi, S. V., and Poornima, M. (2012). Viral metagenomics: a tool for virus discovery and diversity in aquaculture. *Indian J. Virol.* 23, 88–98. doi: 10.1007/s13337-012-0075-2

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol. Rev.* 35, 235–241.

Bastian, M., Heymann, S., and Jacomy, M. (2009). "Gephi: an open source software for exploring and manipulating networks," in *Third International AAAI Conference on Weblogs Social Media*, Paris, 361–362. doi: 10.1136/qshc.2004.010033

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., et al. (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. doi: 10.1073/pnas.0911354106

Cava, F., Hidalgo, A., and Berenguer, J. (2009). *Thermus thermophilus* as biological model. *Extremophiles* 13, 213–231. doi: 10.1007/s00792-009-0226-6

Chuong, E. B. (2013). Retroviruses facilitate the rapid evolution of the mammalian placenta. *Bioessays* 35, 853–861. doi: 10.1002/bies.201300059

Daep, C. A., Muñoz-Jordán, J. L., and Eugenin, E. A. (2014). Flaviviruses, an expanding threat in public health: focus on dengue, West Nile, and Japanese encephalitis virus. *J. Neurovirol.* 20, 539–560. doi: 10.1007/s13365-014-0285-z

Dietzgen, R. G., Mann, K. S., and Johnson, K. N. (2016). Plant virus-insect vector interactions: current and potential future research directions. *Viruses* 8, 1–21. doi: 10.3390/v8110303

Ďurkovič, J., Čaňová, I., Lagana, R., Kučerová, V., Moravčík, M., Priwitzer, T., et al. (2013). Leaf trait dissimilarities between Dutch elm hybrids with a contrasting tolerance to Dutch elm disease. *Ann. Bot.* 111, 215–227. doi: 10.1093/aob/mcs274

Ebrahimi-Mameghani, M., Sadeghi, Z., Abbasalizad Farhangi, M., Vaghef-Mehrabany, E., and Aliashrafi, S. (2016). Glucose homeostasis, insulin resistance and inflammatory biomarkers in patients with non-alcoholic fatty liver disease: beneficial effects of supplementation with microalgae *Chlorella vulgaris*: a double-blind placebo-controlled randomized clinical trial. *Clin. Nutr.* 36, 1001–1006. doi: 10.1016/j.clnu.2016.07.004

Enquist, L. W., and Racaniello, V. R. (2013). "Virology: from contagium fluidum to virome," in *Fields Virology*, eds D. M. Knipe and P. M. Howley (Philadelphia, PA: Lippincott Williams & Wilkins), 1–20.

Forterre, P. (2010). Giant viruses: conflicts in revisiting the virus concept. *Intervirology* 53, 362–378. doi: 10.1159/000312921

Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397–1410.e16. doi: 10.1016/j.cell.2016.08.020

Godfrey-Faussett, P., and Ayles, H. (2016). Why are people living with HIV still dying of tuberculosis? *Lancet* 387, 1141–1143. doi: 10.1016/S0140-6736(16)00699-1

Goel, A. K. (2015). Anthrax: a disease of biowarfare and public health importance. *World J. Clin. Cases* 3, 20–33. doi: 10.12998/wjcc.v3.i1.20

Govindan, B. (2016). Recapitulation of acquired immuno deficiency syndrome associated Kaposi's sarcoma. *Indian J. Sex. Transm. Dis.* 37, 115–122. doi: 10.4103/0253-7184.192120

Granerod, J., Ambrose, H. E., Davies, N. W. S., Clewley, J. P., Walsh, A. L., Morgan, D., et al. (2010). Causes of encephalitis and differences in their clinical presentations in England: a multicentre, population-based prospective study. *Lancet Infect. Dis.* 10, 835–844. doi: 10.1016/S1473-3099(10)70222-X

Halary, S., Temmam, S., Raoult, D., and Desnues, C. (2016). Viral metagenomics: are we missing the giants? *Curr. Opin. Microbiol.* 31, 34–43. doi: 10.1016/j.mib.2016.01.005

Hillman, B. I., and Suzuki, N. (2004). Viruses of the chestnut blight fungus, *Cryphonectria parasitica*. *Adv. Virus Res.* 63, 423–472. doi: 10.1016/S0065-3527(04)63007-7

Holmfeldt, K., Solonenko, N., Shah, M., Corrier, K., Riemann, L., Verberkmoes, N. C., et al. (2013). Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12798–12803. doi: 10.1073/pnas.1305956110

International Committee on Taxonomy of Viruses - Taxonomy (2017). Available at: https://talk. ictvonline. org/taxonomy/w/ictv-taxonomy [accessed July 1, 2017].

Jackson, A. C. (2016a). Diabolical effects of rabies encephalitis. *J. Neurovirol.* 22, 8–13. doi: 10.1007/s13365-015-0351-1

Jackson, A. C. (2016b). Human rabies: a 2016 update. *Curr. Infect. Dis. Rep.* 18, 1–6. doi: 10.1007/s11908-016-0540-y

Jiang, H., Zheng, X., Wang, L., Du, H., Wang, P., and Bai, X. (2017). Hantavirus infection: a global zoonotic challenge. *Virol. Sin.* 32, 32–43. doi: 10.1007/s12250-016-3899-x

Kondo, H., Kanematsu, S., and Suzuki, N. (2013). Viruses of the white root rot fungus, *Rosellinia necatrix*. *Adv. Virus Res.* 86, 177–214. doi: 10.1016/B978-0-12-394315-6.00007-6

Korb, V. C., Chuturgoon, A. A., and Moodley, D. (2016). Mycobacterium tuberculosis: manipulator of protective immunity. *Int. J. Mol. Sci.* 17:131. doi: 10.3390/ijms17030131

Kristensen, D. M., Mushegian, A. R., Dolja, V. V., and Koonin, E. V. (2010). New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* 18, 11–19. doi: 10.1016/j.tim.2009.11.003

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., et al. (2003). A giant virus in amoebae. *Science* 299, 2033. doi: 10.1126/science.1081867

Langan, K. M., Kotsimbos, T., and Peleg, A. Y. (2015). Managing *Pseudomonas aeruginosa* respiratory infections in cystic fibrosis. *Curr. Opin. Infect. Dis.* 28, 547–556. doi: 10.1097/QCO.0000000000000217

Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., et al. (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. doi: 10.1073/pnas.1320670111

Lipinska, A. P., Van Damme, E. J. M., and De Clerck, O. (2016). Molecular evolution of candidate male reproductive genes in the brown algal model Ectocarpus. *BMC Evol. Biol.* 16:5. doi: 10.1186/s12862-015-0577-9

Luk, A. W. S., Williams, T. J., Erdmann, S., Papke, R. T., and Cavicchioli, R. (2014). Viruses of haloarchaea. *Life (Basel)* 4, 681–715. doi: 10.3390/life4040681

Lwoff, A. (1957). The concept of virus the third marjory stephenson memorial lecture. *J. Gen. Microbiol.* 17, 239–253.

Menezes, C. B., Amanda Piccoli Frasson, A. P., and Tasca, T. (2016). Trichomoniasis – are we giving the deserved attention to the most common non-viral sexually transmitted disease worldwide? *Microb. Cell* 3, 404–418. doi: 10.15698/mic2016.09.526

Miceli, M. H., Díaz, J. A., and Lee, S. A. (2011). Emerging opportunistic yeast infections. *Lancet Infect. Dis.* 11, 142–151. doi: 10.1016/S1473-3099(10)70218-8

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., et al. (2016). Linking virus genomes with host taxonomy. *Viruses* 8:66. doi: 10.3390/v8030066

O'Malley, M. A. (2008). 'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography. *Stud. Hist. Philos. Biol. Biomed. Sci.* 39, 314–325. doi: 10.1016/j.shpsc.2008.06.005

Panahi, Y., Darvishi, B., Jowzi, N., Beiraghdar, F., and Sahebkar, A. (2016). *Chlorella vulgaris*: a multifunctional dietary supplement with diverse medicinal properties. *Curr. Pharm. Des.* 22, 164–173. doi: 10.2174/1381612822666151112145226

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., et al. (2013). Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341, 281–286. doi: 10.1126/science.1239181

Popgeorgiev, N., Temmam, S., Raoult, D., and Desnues, C. (2013). Describing the silent human virome with an emphasis on giant viruses. *Intervirology* 56, 395–412. doi: 10.1159/000354561

Raoult, D., and Forterre, P. (2008). Redefining viruses: lessons from Mimivirus. *Nat. Rev. Microbiol.* 6, 315–319. doi: 10.1038/nrmicro1858

Rascovan, N., Duraisamy, R., and Desnues, C. (2016). Metagenomics and the human virome in asymptomatic individuals. *Annu. Rev. Microbiol.* 70, 125–141. doi: 10.1146/annurev-micro-102215-095431

Reganold, J. P., and Wachter, J. M. (2016). Organic agriculture in the twenty-first century. *Nat. Plants* 2, 15221. doi: 10.1038/NPLANTS.2015.221

Shchelkunov, S. N. (2013). An increasing danger of zoonotic orthopoxvirus infections. *PLoS Pathog.* 9:e1003756. doi: 10.1371/journal.ppat.1003756

Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., et al. (2016). Redefining the invertebrate RNA virosphere. *Nature* doi: 10.1038/nature20167 [Epub ahead of print].

Sicard, D., and Legras, J. L. (2011). Bread, beer and wine: yeast domestication in the Saccharomyces sensu stricto complex. *C R Biol.* 334, 229–236. doi: 10.1016/j.crvi.2010.12.016

Siddiqui, R., and Khan, N. A. (2012). Biology and pathogenesis of Acanthamoeba. *Parasit. Vectors* 5:6. doi: 10.1186/1756-3305-5-6

Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., et al. (2017). Consensus statement: virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* 15, 161–168. doi: 10.1038/nrmicro.2016.177

Steger, G., and Perreault, J. P. (2016). Structure and associated biological functions of viroids. *Adv. Virus Res.* 94, 141–172. doi: 10.1016/bs.aivir.2015.11.002

Suttle, C. A. (2005). Viruses in the sea. *Nature* 437, 356–361. doi: 10.1038/nature04160

Thornton, P. K. (2010). Livestock production: recent trends, future prospects. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 365, 2853–2867. doi: 10.1098/rstb.2010.0134

Thurber, R. V. (2009). Current insights into phage biodiversity and biogeography. *Curr. Opin. Microbiol.* 12, 582–587. doi: 10.1016/j.mib.2009.08.008

Vainio, E. J., and Hantula, J. (2015). Taxonomy, biogeography and importance of Heterobasidion viruses. *Virus Res.* 219, 2–10. doi: 10.1016/j.virusres.2015.10.014

Vila, J., Sáez-López, E., Johnson, J. R., Römling, U., Dobrindt, U., Cantón, R., et al. (2016). *Escherichia coli*: an old friend with new tidings. *FEMS Microbiol. Rev.* 40, 437–463. doi: 10.1093/femsre/fuw005

Volkova, E., Tesh, R. B., Monath, T. P., and Vasilakis, N. (2012). Full genomic sequence of the prototype strain (M64) of Rio Bravo virus. *J. Virol.* 86, 4715. doi: 10.1128/JVI.00331-12

Woese, C. R. (2002). On the evolution of cells. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8742–8747. doi: 10.1073/pnas.132266999

Woolhouse, M., Scott, F., Hudson, Z., Howey, R., and Chase-Topping, M. (2012). Human viruses: discovery and emergence. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 2864–2871. doi: 10.1098/rstb.2011.0354

Zhang, C., Krause, D. J., and Whitaker, R. J. (2013). *Sulfolobus islandicus*: a model system for evolutionary genomics. *Biochem. Soc. Trans.* 41, 458–462. doi: 10.1042/BST20120338

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

4.2. ARTIGO #2: Giant among larges – how gigantism impacts giant virus entry into amoebae

Os vírus gigantes fazem parte do grupo dos Vírus Grandes Núcleo-Citoplasmáticos de DNA (NCLDV, ou a recentemente proposta ordem Megavirales). Algumas famílias deste grupo são conhecidas há décadas, como *Poxviridae* e *Phycodnaviridae*, enquanto outras foram descritas há menos tempo, como *Mimiviridae* e *Marseilleviridae*. Junto a estes, outros vírus têm sido descritos, como Pandoravírus, Pithovírus e Mollivírus, embora ainda não estejam oficialmente classificados pelo ICTV. Embora estes vírus sejam bem diferentes estrutural e geneticamente uns dos outros, todos são parasitas de amebas de vida livre, principalmente amebas do gênero *Acanthamoeba*, organismos que surgiram no planeta a mais de um bilhão de anos atrás e, desde então, têm interagido com diversos organismos celulares e virais. Neste artigo de revisão nós discutimos as etapas iniciais da interação dos vírus gigantes e seus hospedeiros, com enfoque nos mimivírus e marseillevírus, e comparamos com os mecanismos de penetração de outros NCLDVs. Ao analisar a interação destes vírus com seus hospedeiros, nós discutimos como as diferentes estratégias de penetração viral exploradas pelos vírus gigantes podem refletir a evolução destes vírus e levantamos a hipótese de que o gigantismo viral teria surgido de maneira independente ao longo da evolução nos variados grupos de vírus gigantes, possibilitando que diferentes vírus explorem nichos distintos. Por fim, revisamos os mecanismos de defesa dos hospedeiros contra a infecção por estes parasitas conhecidos até o momento, ressaltando clássica disputa entre parasitas e hospedeiros, um exemplo típico da teoria da rainha vermelha (ou a mais recentemente discutida, teoria do gato risonho).

Este artigo foi publicado no periódico ***Current Opinion in Microbiolgy*** em junho de 2016.

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Giants among larges: how gigantism impacts giant virus entry into amoebae

Rodrigo Araújo Lima Rodrigues, Jônatas Santos Abrahão, Betânia Paiva Drumond and Erna Geessien Kroon

The proposed order *Megavirales* comprises the nucleocytoplasmic large DNA viruses (NCLDV), infecting a wide range of hosts. Over time, they co-evolved with different host cells, developing various strategies to penetrate them. Mimiviruses and other giant viruses enter cells through phagocytosis, while Marseillevirus and other large viruses explore endocytosis and macropinocytosis. These differing strategies might reflect the evolution of those viruses. Various scenarios have been proposed for the origin and evolution of these viruses, presenting one of the most enigmatic issues to surround these microorganisms. In this context, we believe that giant viruses evolved independently by massive gene/size gain, exploring the phagocytic pathway of entry into amoebas. In response to gigantism, hosts developed mechanisms to evade these parasites.

Address
Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais, Brasil (UFMG), Postal code 486, Belo Horizonte, Minas Gerais, Brazil

Corresponding author: Kroon, Erna Geessien (kroone@icb.ufmg.br)

Current Opinion in Microbiology 2016, **31**:88–93

This review comes from a themed issue on **Environmental microbiology/Megaviromes**

Edited by **Didier Raoult** and **Jonatas Abrahao**

http://dx.doi.org/10.1016/j.mib.2016.03.009

1369-5274/Published by Elsevier Ltd.

Introduction

The recently proposed *Megavirales* order comprises the nucleocytoplasmic large DNA viruses (NCLDVs), the largest and the most complex group of DNA viruses known so far [1]. Some families of this group have been known for some decades, such as *Poxviridae* and *Phycodnaviridae* [2,3]. However, others, such as *Mimiviridae* and *Marseilleviridae* [4,5], have only been discovered in recent years, catching the attention of the entire virology community with the structural and genetic complexity of their members. In the years that followed, many other giant viruses were discovered and characterized, such as pandoraviruses, Pithovirus, Mollivirus, and faustoviruses [6–9]. These new viruses are not yet classified in any viral family, although recent studies have already proven their phylogenetic relationships with other NCLDVs [10–12].

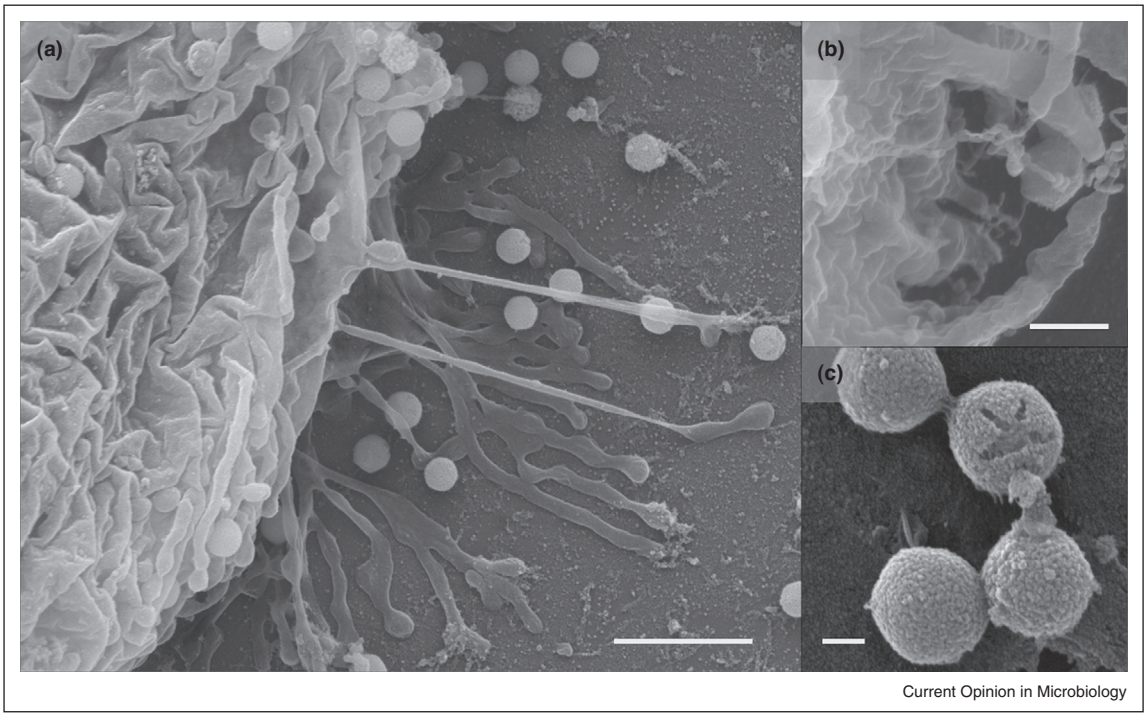
Although each giant virus exhibits its own structural and genetic peculiarities, they possess something in common: the host. These viruses are parasites of free-living amoebae (FLA), most of them from the *Acanthamoeba* genus. Amoebae are ancient unicellular eukaryotes, widely distributed in the environment, presenting two life stages — a vegetative and a resistant one — denominated trophozoite and cyst, respectively [13]. It is estimated that these organisms diverged from the main eukaryotic line of descent approximately 1×10^9 years ago, and since then, have been interacting with several organisms, both eukaryotic and prokaryotic [13]. As a result of these interactions, the amoebae now host many microorganisms, such as bacteria, fungi, and giant viruses [14–16]. Since the discovery of giant viruses, new studies have been developed to better understand the virus–host interaction.

In this brief review, we discuss the initial steps of this interaction, focusing on the attachment and entry of mimiviruses and marseilleviruses, and comparing these mechanisms with those of other NCLDVs. Furthermore, we discuss what the entry strategies explored by the giant viruses can tell us about their evolution, raising new hypotheses concerning their evolutionary history, and how amoebae respond to the presence of their parasites, discussing what we can infer concerning this ancient relationship.

Giant viruses explore different entry strategies

Acanthamoeba polyphaga mimivirus (APMV) was the first giant virus discovered and is the type virus of the *Mimivirus* genus, *Mimiviridae* family [4]. It possesses a pseudo-icosahedral capsid (~500 nm), covered by glycoproteic fibrils (~125 nm), which are immersed in an amorphous matrix of peptidoglycan [17]. Recently, we demonstrated that these proteic structures are important for viral attachment to the surface of the host, interacting with sugars present in the membrane of the amoebae, particularly N-acetylglucosamine and mannose (Figure 1A) [18••]. Following initial attachment to the host cells, the mimiviruses explore the phagocytic pathway as mode of entry, they are one of the first groups of viruses to be described exploring this penetration strategy (Figure 1B) [19••]. An exclusive feature of the mimiviruses is the stargate, a star-shaped structure in one of the vertices of

Figure 1



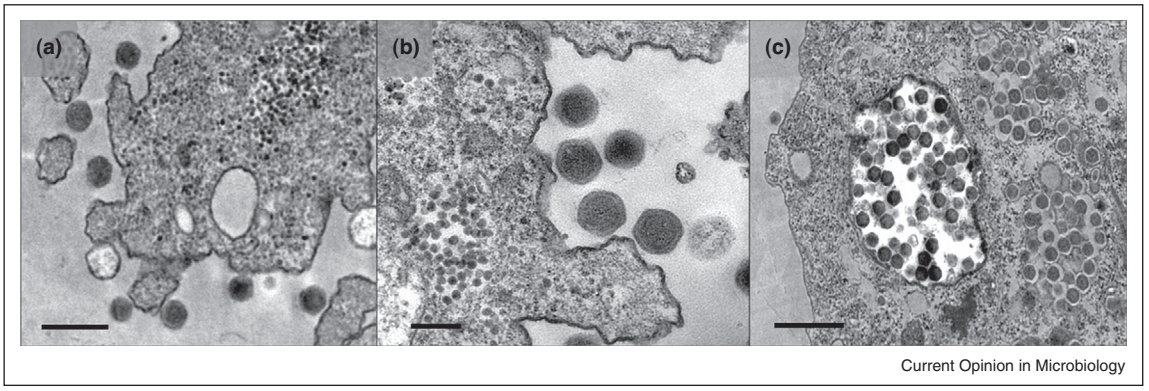
Peculiarities of the mimivirus entry strategy. Scanning electron microscopy images showing **(A)** mimivirus particles attached to the surface of an *Acanthamoeba castellanii* cell, the first step in the virus–host interaction; **(B)** A mimivirus particle being phagocytosed by *A. castellanii* cell (entry strategy); **(C)** stargate portal, the structure responsible for releasing the genome into the amoeba cytoplasm. Scale bar: 2 μm (A); 500 nm (B); 200 nm (C).

the capsid (Figure 1C) [17,20]. Following viral entry, still inside the phagosome, the stargate opens, exposing a lipid inner membrane, which merges with the membrane of the phagosome [20]. In the next step, the viral genome, within a proteic shell (known as the viral seed), is released into the host cytoplasm, establishing an early viral factory, which further develops into a mature viral factory, wherein occurs genome replication and viral morphogenesis [21,22].

The second large amoebal virus discovered was the *Marseillevirus marseillevirus* (MsV). This virus caught much attention due to its genetic content, wherein several horizontal gene transfer (HGT) events were verified to have occurred during the evolution of its genome [5]. The MsV exhibits short fibrils (~12 nm) surrounding the ~220 nm capsid, and likely as for mimiviruses, these structures might act as adhesion factors to the host by interacting with sugars on the surface of the amoebae (Figure 2A) [5]. Pioneering works have indicated that this virus, just like the giant mimivirus, enters host cells through phagocytosis. However, even though the marseilleviruses are large (~220 nm), these viruses do not reach the required size to stimulate phagocytosis (>500 nm) [23]. Surprisingly, data from our group demonstrate that during the replication cycle, the MsV is able to induce the formation and be released by large multimembranous vesicles (>1 μm), which may contain hundreds of viral particles, thus being a structure large enough to stimulate phagocytosis [24]. This process has already been observed for other microorganisms, such as *Legionella pneumophila*, suggesting that this is not specific for viruses, but instead a specificity of some amoebal species [25]. However, it was the first description of such mechanism for a large virus. By evaluating the mechanism of entry of each infectious entity (single viral particle or large vesicle), we noted that when individualized, the MsV particles explored the endocytosis pathway, while in vesicles, the process was carried out through phagocytosis (Figure 2B,C) [24]. After entry, MsV releases the genome into the cell cytoplasm leading to the formation of a large and diffuse viral factory, wherein genome replication and viral morphogenesis occur [5,26•].

Other amoebal-parasite large viruses also explore the phagocytic pathway as an entry mechanism [27••]. It is very likely that pandoraviruses, Pithovirus, and Mollivirus, as well as the mimiviruses, penetrate their hosts only by phagocytosis due to their large size. Meanwhile, for faustoviruses, phagocytosis had already been described as an entry mode [9], but other mechanisms might be involved, as has been observed for marseilleviruses.

Figure 2



Marseilleviruses enter the amoeba cell using different strategies. Transmission electron microscopy images showing **(A)** Marseillevirus particles attached to the surface of an *Acanthamoeba castellanii* cell, possibly due to the interaction between viral nanofibers and host carbohydrates moments before viral uptake; **(B)** Individual Marseillevirus particles being engulfed in a macropinocytosis-like entry mode; **(C)** Marseillevirus particles within a large membranous vesicle after being phagocytosed by an *A. castellanii* cell. Scale bar: 500 nm (A); 200 nm (B); 1 μm (C).

Strategies such as endocytosis or macropinocytosis might be involved in faustoviruses penetration, as observed for asfaviruses [28] and poxviruses [29], the closest phylogenetic groups to faustoviruses [12]. Other large viruses explore those same mechanisms, such as iridoviruses, and possibly ascoviruses, even though the entry mode of the latter has not been fully elucidated [30,31]. Finally, the phycodnaviruses explore a mechanism that diverges from those mentioned above: they degrade the host (algae) cell wall, allowing fusion of the internal lipid membrane with the host plasma membrane and release of the viral genome inside the cell [2]. Therefore, over the course of evolution, several different entry strategies have emerged among the NCLDVs, reinforcing the huge diversity of these complex viruses.

Giant entry mode tells us about evolution

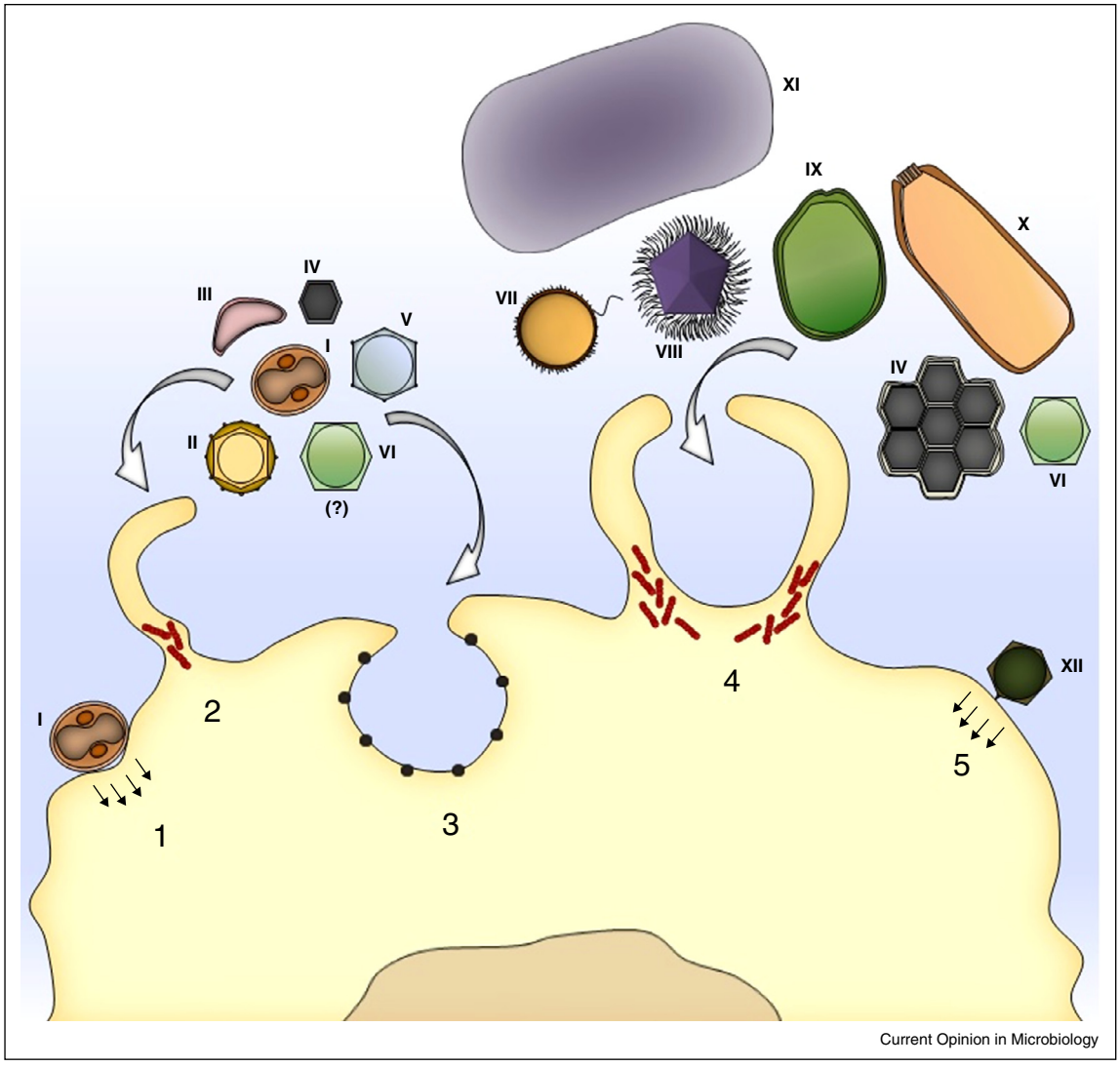
The origin and evolution of the giant viruses have been discussed extensively. Recent studies suggest that they originated from small viruses, more specifically polintoviruses [32•] and their evolution occurred through successive events of gene gain and loss favored by interaction with host and sympatric organisms, following an accordion-like model [33•]. According to this, it is likely that the last common viral ancestor of the NCLDVs was a small virus, which penetrated into the host by receptor-mediated endocytosis and/or macropinocytosis, and other strategies have emerged independently with the advent of gigantism facing new selective pressures. Events of genomic expansion seem to have occurred more than once in the course of giant virus evolution, giving rise to the giant mimiviruses, pandoraviruses, molliviruses, and pithoviruses, since each of them present a more extensive genome than the other NCLDVs, and also very distinct among themselves [27••]. This genetic gain led to an increase in the viral particles size, promoting the entry by phagocytosis. This strategy seems to be positively selected, considering that the hosts of all known giant viruses are amoebae [34]. Since amoebae feed on other microorganisms by phagocytosis [13], it would be beneficial for the viruses to explore this strategy to enter and further replicate within the host, perpetuating themselves over time.

Although advantageous at first sight, penetration by phagocytosis comes with a problem: the release of the genome inside the host cell cytoplasm. With that in mind, the hypothesis of an independent evolution of gigantism among NCLDVs is reinforced. All giant viruses overcame this barrier by presenting structures responsible for releasing the genome, such as the stargate portal (mimiviruses), the ostiole-like portal (pandoraviruses and Pithovirus), and a circular depression (Mollivirus) [27••], even though they are structurally distinct, which indicates an evolutionary convergence of such mechanism within the *Megavirales* order. Moreover, other entry mechanisms have been used by other members of the groups, such as membrane fusion for poxviruses [29], and membrane puncture for phycodnaviruses [2], where the interaction with a different host seems to have been the crucial factor in such selection. Thus, various ancestors of the NCLDVs interacted with different host organisms, resulting in host-specific strategies for penetration (Figure 3).

The host responds to viral infection

While the giant viruses developed strategies to enter and replicate within their hosts, they had also to develop mechanisms to respond to and/or prevent infection by parasites. As mentioned above, amoebae have two life stages. Trophozoites feed on many different microorganisms through phagocytosis, including giant viruses. In this

Figure 3



Nucleocytoplasmic large DNA viruses explore different strategies to enter the host. The poxviruses (I) can merge its external membrane with that of the host, entering through membrane fusion (1). In addition, the poxviruses can explore other actin-dependent pathways, such as macropinocytosis (2) and receptor-mediated endocytosis (3), along with iridoviruses (II), ascoviruses (III), marseilleviruses (IV), asfaviruses (V), and possibly faustoviruses (VI) (not experimentally confirmed). The marseilleviruses can also enter through another actin-dependent pathway, the phagocytosis pathway (4), when inside large membranous vesicles, as well as the mollivirus (VII), mimiviruses (VIII), pandoraviruses (IX), faustoviruses (VI), and Pithovirus (X), alongside with other microorganisms, such as bacteria (XI). The phycodnaviruses (XII) explore a more distinct strategy, forming pores in the host's membrane (5).

stage, the amoebas are susceptible to distinct parasites, but they can differentiate into cysts, spherical cells composed by a double wall, allowing for survival under several harsh conditions, such as a lack of nutrients, osmotic stress, and parasite strikes [13]. The encystment process leads to a remarkable reorganization of the cytoskeleton and a large turnover of cellular components [35]. It has already been demonstrated that some signaling pathways, such as PI3K and MAPK, are involved in this process, along with serine and cysteine proteases [36,37]. Although this process is considered a classical defense strategy of amoebae, some parasites are able to survive inside the cysts, such as *Mycobacterium leprae* and *Legionella pneumophila* [38,39]. The trophozoites are infected by the parasites, and when converted into cysts, these bacteria are not eliminated, remaining viable for long periods. For giant viruses, this seems to be different.

Although they can penetrate and replicate within acanthamoeba during the trophozoite stage, the mimiviruses are not able to infect cysts [40]. In this particular study, the authors demonstrated that when the amoebae

are already encysted, the mimiviruses are incapable of infecting them. However, these viruses can prevent the encystment process by inhibition of the expression of serine protease (an essential step in the process). In this way, mimiviruses evade the classical defense strategy of their natural hosts and ensure viral replication and dispersal in the environment. There are no reports demonstrating whether encystment is effective against other giant viruses, nor even if these parasites have mechanisms to escape the host's defenses, similar to those of the mimiviruses. However, considering that giant viruses such as pandoraviruses, Pithovirus, and Mollivirus, enter through phagocytosis [27••], and that this process does not occur when amoebae are encysted, it is very likely the encystment to be a form of response to them.

This defense strategy also reflects the co-evolution of hosts and giant viruses. Since the emergence of amoebae approximately 1×10^9 years ago, they have coexisted with many different kinds of bacterial and fungal species, establishing different ecological relationships, from symbiosis to parasitism [41]. The same can be extended to the giant viruses. The most ancient giant virus isolated so far belongs to samples dating back 30,000 years [7], but there are no current data pointing to its origin in history. Since mimiviruses are able to attach to other organisms, such as bacteria, fungi, and arthropods, through glycoside interactions [18••], it is reasonable to consider that these viruses and other giants have been interacting with ancestral forms of these organisms [42]. From this perspective, over the course of time the amoebal hosts developed mechanisms to escape their parasites, while the parasites sought different ways to overcome such barriers, thus establishing a classical dispute between parasite and host, a typical example of the Red Queen's theory (or the more recently discussed 'Cheshire Cat' theory [43]).

Future perspectives

Studies regarding the interaction between the giant viruses and their hosts are still in their infancy. Research on the interactions of giant viruses with their hosts might also bring new insights into the enigmatic evolutionary history of these microorganisms. In this context, it seems that the giant viruses originated from small viruses [32•], but other scenarios should not be ruled out, since other studies have suggested that they came from an extinct cellular domain of life [44,45]. Regardless of its origin, gigantism among the NCLDVs seems to have emerged independently, establishing a huge advantage and prompting the viruses to explore new ecological niches. The discovery of new giant viruses will boost our knowledge about the origin, evolution, and ecological roles played by organisms. Thus, many remaining gaps about these viruses will be filled soon, and the puzzle of the biology of them will become increasingly complete and elegant. New studies are coming and outstanding discoveries come with them.

Acknowledgments

Scholarships were provided by the Conselho Nacional de Desenvolvimento Científico e Tecnológico, and Fundação de Amparo à Pesquisa do Estado de Minas Gerais.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Colson P, De Lamballerie, Yutin N, Asgari S, Bigot Y, Bideshi DK, Cheng XW, Federici BA, Van Etten JL, Koonin EV *et al.*: **"Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses**. *Arch Virol* 2012, **158**:2517-2521.

2. Wilson WH, Van Etten JL, Allen MJ: **The Phycodnaviridae: the story of how tiny giants rule the world**. *Curr Top Microbiol Immunol* 2009, **328**:1-42.

3. Moss B: **Poxviridae**. In *Fields Virology*. Edited by Knipe DM, Howley PM. Lippincott Williams & Wilkins; 2013:2129-2159.

4. La Scola B, Audic S, Robert C, Jungang L, de Lambeillerie X, Drancourt M, Birtles R, Claverie JM, Raoult D: **A giant virus in amoebae**. *Science* 2003, **28**:299.

5. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, Robert C, Azza S, Sun S, Rossmann MG *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, **106**:21484-21553.

6. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C *et al.*: **Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes**. *Science* 2013, **341**:281-286.

7. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, Lescot M, Poirot O, Bertaux L, Bruley C *et al.*: **Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology**. *Proc Natl Acad Sci U S A* 2014, **111**:4274-4279.

8. Legendre M, Lartique A, Bertaux L, Jeudy S, Bartoli J, Lescot M, Alempic JM, Ramus C, Bruley C, Labadie K *et al.*: **In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting *Acanthamoeba***. *Proc Natl Acad Sci U S A* 2015, **112**:E5327-E5335.

9. Reteno DG, Benamar S, Khalil JB, Andreani J, Armstrong N, Klose T, Rossmann M, Colson P, Raoult D, La Scola B: **Faustovirus, an asfavirus-related new lineage of giant viruses infecting amoebae**. *J Virol* 2015, **89**:6585-6594.

10. Sharma V, Colson P, Chabrol O, Scheid P, Pontarotti R, Raoult D: **Welcome to pandoraviruses to the 'Fourth TRUC' club**. *Front Microbiol* 2015, **6**:423.

11. Sharma V, Colson P, Chabrol O, Pontarotti R, Raoult D: **Pithovirus sibericum, a new bona fide member of the "Fourth TRUC" club**. *Front Microbiol* 2015, **6**:722.

12. Benamar S, Reteno DG, Bandaly V, Labas N, Raoult D, La Scola B: **Faustoviruses: comparative genomics of new Megavirales family members**. *Front Microbiol* 2016, **7**:3.

13. Siddiqui R, Khan NA: **Biology and pathogenesis of *Acanthamoeba***. *Parasit Vectors* 2012, **5**:6.

14. Schmitz-Esser S, Toenshoff ER, Haider S, Heinz E, Hoenninger VM, Wagner M, Horn M: **Diversity of bacterial endosymbionts of environmental *Acanthamoeba* isolates**. *Appl Environ Microbiol* 2008, **74**:5822-5831.

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

16. Steenbergen JN, Nosanchuk JD, Malliaris SD, Casadevall A: **Interaction of *Blastomyces dermatitidis*, *Sporothrix schenckii*, and *Histoplasma capsulatum* with *Acanthamoeba castellanii***. *Infect Immun* 2004, **72**:3478-3488.

17. Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kotsyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A, Rossmann MG: **Structural studies of the giant mimivirus**. *PLoS Biol* 2009, **7**:e92.

18. •• Rodrigues RA, dos Santos Silva LK, Dornas FP, de Oliveira DB, Magalhães TF, Santos DA, Costa AO, de Macêdo Farias L, Magalhães PP, Bonjardim CA *et al.*: **Mimivirus fibrils are important for viral attachment to the microbial world by a diverse glycoside interaction repertoire**. *J Virol* 2015, **89**:11812-11819.
The first study to demonstrate the initial step of the mimivirus replication cycle, wherein viral fibrils interact with the host, and also with other organisms, through glycoside interactions.

19. •• Ghigo E, Kartenbeck J, Lien P, Pelkmans L, Capo C, Mege JL, Raoult D: **Amoebal pathogen mimivirus infects macrophages through phagocytosis**. *PLoS Pathog* 2008, **4**:e1000087.
An important study that demonstrates the entry strategy explored by mimiviruses (phagocytosis), establishing a new mode of entry among the viruses.

20. Zauberman N, Mutsafi Y, Halevy DB, Shimoni E, Klein E, Xiao C, Sun S, Minsky A: **Distinct DNA exit and packaging portals in the virus *Acanthamoeba polyphaga* mimivirus**. *PLoS Biol* 2008, **6**:e114.

21. Mutsafi Y, Zauberman N, Sabany I, Minsky A: **Vaccinia-like cytoplasmic replication of the giant mimivirus**. *Proc Natl Acad Sci U S A* 2010, **107**:5978-5982.

22. Mutsafi Y, Shimoni E, Shimon A, Minsky A: **Membrane assembly during the infection cycle of the giant Mimivirus**. *PLoS Pathog* 2013, **9**:e1003367.

23. Korn ED, Weisman RA: **Phagocytosis of latex beads by *Acanthamoeba*. II. Electron microscopy studies of the initial events**. *J Cell Biol* 1967, **34**:219-227.

24. Arantes TS, Rodrigues RA, Silva LK, Oliveira GP, de Souza HL, Khalil BJ, Oliveira DB, Torres AA, da Silva LL, Colson P *et al.*: **The large Marseillevirus explores different entry pathways by forming giant infectious vesicles**. *J Virol* 2016. [Epub ahead of print].

25. Berk SG, Ting RS, Turner GW, Ashburn RJ: **Production of respirable vesicles containing live *Legionella pneumophila* cells by two *Acanthamoeba* spp**. *Appl Environ Microbiol* 1998, **64**:279-286.

26. • Aherfi S, La Scola B, Pagnier I, Raoult D, Colson P: **The expanding family Marseilleviridae**. *Virology* 2014, **467**:27-37.
An excellent review of the large marseilleviruses, covering the current knowledge about different biological aspects of these new viruses.

27. •• Albergel C, Legendre M, Claverie JM: **The rapidly expanding universe of giant viruses: mimivirus, pandoravirus, pithovirus and mollivirus**. *FEMS Microbiol Rev* 2015, **36**:779-796.
An excellent review comparing some biological aspects of the giant viruses discovered so far.

28. Sánchez EG, Quintas A, Pérez-Núñez D, Nogal M, Barroso S, Carrascosa ÁL, Revilla Y: **African swine fever virus uses macropinocytosis to enter host cells**. *PLoS Pathog* 2012, **8**:e1002754.

29. Moss B: **Poxvirus cell entry: how many proteins does it take?** *Viruses* 2012, **4**:688-707.

30. Wang S, Huang X, Huang Y, Hao X, Xu H, Cai M, Wang H, Qin Q: **Entry of a novel marine DNA virus, Singapore grouper iridovirus, into host cells occurs via clathrin-mediated endocytosis and macropinocytosis in a pH-dependant manner**. *J Virol* 2014, **88**:13047-13063.

31. Hussain M, Garrad S, Asgari S: **The role of actin filaments in ascovirus replication and pathology**. *Arch Virol* 2009, **154**:1737-1743.

32. • Koonin EV, Krupovic M, Yutin N: **Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses**. *Ann N Y Acad Sci* 2015, **1341**:10-24.
An important work that raises the newest theory about the origin and evolution of the giant viruses, suggesting that they came from polintoviruses.

33. • Filée J: **Genomic comparison of closely related giant viruses supports an accordion-like model of evolution**. *Front Microbiol* 2015, **16**:593.
An extensive analysis of the giant viruses' genome, which supports the most recent theory that the *Megavirales* members evolved by massive gene gain and loss.

34. La Scola B: **Looking at protists as a source of pathogenic viruses**. *Front Microbiol* 2014, **77**:131-135.

35. Lloyd D: **Encystment in *Acanthamoeba*: a review**. *Exp Parasitol* 2014, **145(Suppl.)**:S20-S27.

36. Leitsch D, Kohsler M, Marchetti-Deschmann M, Deutsch A, Allmaier G, Duchêne M, Walochnik J: **Major role for cysteine proteases during the early phase of *Acanthamoeba castellanii* encystment**. *Eukaryot Cell* 2010, **9**:611-618.

37. Moon EK, Chung DI, Hong YC, Kong HH: **Characterization of a serine proteinase mediating encystation of *Acanthamoeba***. *Eukaryot Cell* 2008, **7**:1513-1517.

38. Wheat WH, Casali AL, Thomas V, Spencer JS, Lahiri R, Williams DL, McDonnell GE, Gonzalez-Juarrero M, Brennan PJ, Jackson M: **Long-term survival and virulence of *Mycobacterium leprae* in amoebal cysts**. *PLoS Negl Trop Dis* 2014, **8**:e3405.

39. Kilvington S, Price J: **Survival of *Legionella pneumophila* within cysts of *Acanthamoeba polyphaga* following chlorine exposure**. *J Appl Bacteriol* 1990, **68**:519-525.

40. Boratto P, Albarnaz J, Almeida GM, Botelho L, Fontes AC, Costa AO, Santos DdeA, Bonjardim CA, La Scola B, Kroon EG *et al.*: ***Acanthamoeba* polyphaga mimivirus prevents amoebal encystment-mediating serine proteinase expression and circumvents cell encystment**. *J Virol* 2015, **89**:2962-2965.

41. Scheid P: **Relevance of free-living amoebae as hosts for phylogenetically diverse microorganisms**. *Parasitol Res* 2014, **113**:2407-2414.

42. Nasir A, Kim KM, Caetano-Anolles G: **Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya**. *BMC Evol Biol* 2012, **12**:156.

43. Silva LK, Boratto PV, La Scola B, Bonjardim CA, Abrahao JS: **Acanthamoeba and mimivirus interactions: the role of amoebal encystment and the expansion of the 'Cheshire Cat' theory**. *Curr Opin Microbiol* 2016, **31**:9-15.

44. Boyer M, Madoui MA, Gimenez G, La Scola B, Raoult D: **Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses**. *PLoS ONE* 2010, **5**:e15530.

45. Colson P, Gimenez G, Boyer M, Fournous G, Raoult D: **The giant cafeteria roenbergensis virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of life**. *PLoS ONE* 2011, **6**:e18935.

4.3. ARTIGO #3: The analysis of translation-related gene set boost debates around origin and evolution of mimiviruses

A descoberta dos mimivírus representou um marco para a virologia, uma vez que os mimivírus desafiaram uma série de conceitos bem estabelecidos sobre o que é um vírus. A presença de genes pouco comum entre os vírus, e mesmo alguns nunca antes descritos na virosfera, relacionados ao processo de tradução, incluindo RNAs transportadores, aminoacil-tRNA-sintetases (aaRS), e fatores de síntese protéica, estimularam debates acirrados a respeito da origem e evolução destes vírus. Embora alguns autores defendam que estes elementos são oriundos de eventos de transferência gênica horizontal, sendo os vírus descendentes de organismos mais simples, outros defendem que estes genes já estariam presentes em um ancestral mais complexo, o qual possuiria um arsenal gênico ainda mais completo, e este foi se perdendo ao longo da evolução. Neste artigo, nós compilamos os principais dados disponíveis na literatura referente à maquinaria traducional dos mimivírus e comparamos a distribuição dos elementos de tradução entre os diferentes membros do gênero *Mimivirus* sob uma óptica evolutiva. Realizamos análises de uso preferencial de códons, comparando representantes das três linhagens do gênero *Mimivirus* e *Acanthamoeba castellanii* (hospedeiro), juntamente com análises filogenéticas pelo método de máxima verossimilhança das aaRS presentes nestes vírus. Os resultados obtidos neste trabalho sugerem que é possível que os mimivírus tivessem sido originados de organismos mais complexos, todavia, outros cenários também são discutidos, assim como possíveis problemas de amostragem podem levar a diferentes resultados filogenéticos. Em vista destes dados, discutimos como a presença ou ausência de genes relacionados ao processo de tradução entre os mimivírus levam a importantes ideias que estimulam o debate da origem e da história evolutiva destes vírus.

Este artigo foi publicado no periódico ***PLoS Genetics*** em fevereiro de 2017.

PLOS | GENETICS

REVIEW

The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses

Jônatas Santos Abrahão[1,2☯], Rodrigo Araújo[2☯], Philippe Colson[1], Bernard La Scola[1]*

1 Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE) UM63 CNRS 7278 IRD 198 INSERM U1095, Aix-Marseille Univ., 27 boulevard Jean Moulin, Faculté de Médecine, Marseille, France, **2** Instituto de Ciências Biológicas, Departamento de Microbiologia, Laboratório de Vírus, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

☯ These authors contributed equally to this work.
* bernard.la-scola@univ-amu.fr

Abstract

The giant mimiviruses challenged the well-established concept of viruses, blurring the roots of the tree of life, mainly due to their genetic content. Along with other nucleo-cytoplasmic large DNA viruses, they compose a new proposed order—named *Megavirales*—whose origin and evolution generate heated debate in the scientific community. The presence of an arsenal of genes not widespread in the virosphere related to important steps of the translational process, including transfer RNAs, aminoacyl-tRNA synthetases, and translation factors for peptide synthesis, constitutes an important element of this debate. In this review, we highlight the main findings to date about the translational machinery of the mimiviruses and compare their distribution along the distinct members of the family *Mimiviridae*. Furthermore, we discuss how the presence and/or absence of the translation-related genes among mimiviruses raises important insights to boost the debate on their origin and evolutionary history.

Citation: 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(2): e1006532. doi:10.1371/journal.pgen.1006532

Editor: Jonathan Filée, CNRS, FRANCE

Published: February 16, 2017

Funding: The authors received funding from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), CAPES-COFECUB, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) and the Mediterranean Foundation Infection (FMI). JSA is a CNPq researcher. The funders had no role in the preparation of the article.

Competing interests: The authors have declared that no competing interests exist.

Introduction

In 1957, a time when genetics and cellular biology were in their preliminary stages, André Lwoff proposed a modern concept of viruses based on a set of features that, directly or indirectly, emphasized that viruses are defined by "negative plesiomorphic or apomorphic non-natural characteristics" [1]. In the following years, many research fields evolved—including molecular biology and virology—and several different viruses were discovered, presenting some characteristics which had never been seen before among living organisms. However, even with such progress, most of the features raised by Lwoff have still been well supported if we consider the universe of viral species catalogued by the International Committee on Taxonomy of Viruses (ICTV) [2].

However, the discovery of the giant viruses blurred some of those well-established concepts, surprising the scientific community by their size and genetic content [3]. Although they still fit in some of Lwoff's proposed non-natural features (the same is also true for some intracellular bacteria), giant viruses present an astonishing arsenal of genes not widespread in the

virosphere, some of them related to important steps of the translational process, including transfer RNAs (tRNAs), aminoacyl–tRNA synthetases (aaRS), and translation factors for peptide synthesis [4–6]. Recent studies have shown that some of those genes can be related to the improvement of viral fitness, despite the presence of related genes in their hosts' genomes [7]. The analysis of these intriguing translation-related gene sets has raised interesting theories about the "lifestyle" of giant viruses' ancestors. In this review, we highlight the main features about the diversity, function, and putative origin of mimivirus translation-related genes.

Mimiviridae: A giant virus family with exceptional genetic content

The first giant virus was isolated from a water sample of a cooling tower in Bradford, England, replicating in the protist host *Acanthamoeba polyphaga*. At the time of its isolation, the new microorganism was considered to be a gram-positive bacterium, and its viral nature was only established after transmission electron microscopy analysis, which led to its label *Acanthamoeba polyphaga mimivirus* (APMV) [8]. Since then, new mimivirus-like viruses have been isolated in different parts of the world. These viruses constitute the new family *Mimiviridae*, which was included in the recently proposed order *Megavirales* [9,10].

The mimiviruses present several unusual features, both genetic and structural (for details about the viral structure, see [11]). The genome of these viruses consists of a single linear dsDNA molecule, is A+T rich (reaching up to 1,259 Kb), and may have approximately 1,000 open reading frames with a coding density higher than 90% [4,12]. The mimiviruses are phylogenetically divided into two groups (I and II), the first, which comprises mimiviruses that infect *Acanthamoeba*, being subdivided into three lineages (A, B, and C). Lineage A comprises APMV [8], Mamavirus [13], Samba virus [14], Niemeyer virus [15], and many others. Lineage B is mainly represented by Moumouvirus [16] and lineage C by Megavirus chilensis (MCV) [6], Courdo11 virus [17], and LBA111 virus [18]. Group II comprises smaller *Mimiviridae* members distantly related to APMV and is represented by *Cafeteria roenbergensis* virus (CroV) [5] and some algae viruses, including *Phaeocystis globosa* virus (PgV) [19] and Organic Lake Phycodnaviruses (OLV) [20].

The genome of mimiviruses is impressive not only for its size but also (and mainly) for its genetic content, presenting many genes which have never previously been described for other viruses. The mimiviruses possess many genes codifying DNA repair enzymes and are the first viruses to code for topoisomerase type IA [4]. Moreover, mimiviruses have their own glycosylation apparatus, presenting glycosyltransferases that are involved in the biosynthesis of glycans and post-translational protein modifications [21,22]. Furthermore, and even more impressive, is the presence of genes related to the protein synthesis, such as aaRS, tRNAs, and translation factors, which are present in different amounts in several representatives of the family *Mimiviridae* (Table 1). Other giant viruses, such as Marseillevirus [23], Pandoravirus [24], Faustovirus [25], and Mollivirus [26], also have some of these components, but in much less abundance compared to the mimiviruses.

Aminoacyl-tRNA synthetases among viruses—Breaking barriers

The aaRS are key enzymes in gene translation, during which they catalyse the esterification of a specific amino acid to the 3'-end of its cognate tRNA, forming the aminoacyl-tRNAs [27,28]. There are 20 different aaRS, which are divided into two families named class I and class II [28]. The aaRS are present in a wide variety of different organisms from all domains of life and, until very recently, there were no descriptions of these enzymes in a virus, and they were thus considered trademarks of cellular organisms [29]. However, with the discovery of the mimiviruses, this scenario has changed.

Table 1. Giant viruses' translation-related genes. Representative isolates of each group or family.

Group/Viruses	Aminoacyl-tRNA synthetase	tRNA	Translation Factors
Mimivirus Lineage A			
APMV	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
Mamavirus	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
Lentille	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, eF-TU, eRF1
Hirudovirus	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
SMBV	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
OYTV	ArgRS (2x), CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
KROV	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine	IF4A, IF4E, eF-TU, eRF1
AMAV	CysRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
NYMV	ArgRS, CysRS (2x), MetRS (2x), TyrRS (2x)	Leucine (2x), Histidine, Cysteine	IF4A, IF4E, SUI1, eF-TU, eRF1
Terra2	ArgRS, CysRS, MetRS, TyrRS	Leucine (2x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
Bombay	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
Mimivirus Lineage B			
APMOUV	ArgRS (4x), CysRS, IleRS, MetRS, TyrRS	Leucine, Histidine, Cysteine	IF4E, SUI1, eF-TU, eRF1
Goulette	CysRS, MetRS	Leucine (3x), Histidine, Cysteine	IF4E, SUI1, eF-TU, eRF1
Monve	ArgRS (2x), AsnRS, CysRS, IleRS (2x), MetRS, TyrRS	Leucine, Histidine, Cysteine	IF4A, IF4E (2x), SUI1, eRF1
Mimivirus Lineage C			
MCV	ArgRS, AsnRS, CysRS, IleRS, MetRS, TrpRS, TyrRS	Leucine (2x), Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
Terra1	ArgRS, CysRS, MetRS, TyrRS	Leucine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
LBA111	ArgRS, AsnRS, CysRS, IleRS, MetRS, TrpRS, TyrRS	Leucine (2x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
Courdo7	IleRS, TyrRS	Leucine (3x), Tryptophan	IF4A (2x), IF4E, SUI1, eRF1
Courdo11	ArgRS, AsnRS (2x), CysRS, IleRS, MetRS, TrpRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A (2x), IF4E, SUI1, eRF1
Mimivirus group II			
CroV	IleRS	Leucine (9x), Serine (5x), Tyrosine, Asparagine, Lysine	IF4A, IF4E, SUI1
PgV	-	Leucine (3x), Asparagine (2x), Isoleucine, Arginine, Glutamine	IF4E
OLV	-	Leucine, Isoleucine, Tyrosine, Asparagine, Arginine	IF4E
Other giant viruses			
Marseillevirus	-	-	eIF5, SUI1, EF1α, eRF1
Faustovirus E12	-	-	SUI1
Pandoravirus salinus	TyrRS, TrpRS	Proline, Methionine, Tryptophan	IF4E
Pandoravirus dulcis	TyrRS	Proline	IF4E
Pandoravirus inopinatum	-	Proline	IF4E
Mollivirus sibericum	-	Leucine, Methionine, Tyrosine	IF4E

doi:10.1371/journal.pgen.1006532.t001

A total of four aaRS were found in the APMV genome (Arginyl-RS, Cysteinyl-RS, Methionyl-RS, and Tyrosyl-RS), all of them classified as class I aaRS [4]. In the following years, other mimiviruses were discovered, and the number of mimiviral aaRS expanded. Moumouvirus presents three out of four aaRS described in APMV (ArgRS, CysRS, and TyrRS), plus two others (Asparaginyl-RS [class II] and Isoleucyl-RS [class I]) [16]; and remarkably, MCV presents all the aaRS found in APMV and Moumouvirus, plus another one (Tryptophanyl-RS) [6]. Therefore, to our knowledge, MCV displays the most diverse set of aaRS, i.e., seven different aaRS. It is noteworthy that several paralogs of aaRS can be found in the genome of some mimiviruses. Moumouvirus has four copies of ArgRS and, for this reason, it is the virus with the highest abundance of aaRS (eight); and Niemeyer virus, a new lineage A mimivirus, presents three sets of aaRS duplication (MetRS, CysRS, and TyrRS are duplicated) [15,16]. Regarding CroV (mimivirus group II), only IleRS is present, which is also found in mimiviruses of lineages B and C [5].

The distribution of aaRS follows a clear pattern of diversity in family *Mimiviridae*: viral genomes with a wealthy aaRS gene set (e.g., MCV) contain all types of aaRS found in the rest of the family [6,13,15,16]. This suggests gradual gene loss throughout the evolution of different mimivirus groups (Fig 1A). The evolutionary pressures related to the conservation or loss of aaRS might be linked to specific environmental pressures to which each mimivirus group and/or lineage was submitted after mimiviral species radiation. This scenario would be in accordance with previous phylogenetic and phylogenomic works, which suggested that giant viruses originated from a more complex organism and evolved by genomic reduction [30,31]. However, other studies suggested that some mimivirus aaRS were acquired by horizontal gene transfer (HGT) [32,33]. From this perspective, the mimiviruses should have originated from smaller organisms and evolved mainly by HGT events, being considered "gene pickpockets" [32,33,61] (Fig 2). It is important to highlight, however, that the phylogeny of aaRS is quite complex and sometimes violates the expected pattern of canonical domains of life [29]. The methods of tree construction, alignment, and hits sampling could explain conflicting results observed in different studies (S1–S7 Files). Nevertheless, in Fig 2, we present eight mimivirus aaRS-based trees constructed by maximum likelihood method (very similar results were observed for trees constructed by the neighbor joining method). Considering about 100 of the best hits obtained in GenBank related to each aaRS, we observed that phylogenic reconstructions suggested that all mimiviral genes but TyrRS clustered together as an independent group, with bootstrap values >90 both for viral and cellular taxa in most of the trees (Fig 2). However, both theories regarding mimivirus origin—that mimiviruses either originated from a more complex ancestor or that they originated from a simple ancestor—are plausible, although increasing evidence points to the former hypothesis [4,10,30,31,34].

The analysis of APMV transcriptome revealed that its four aaRS are expressed during the replication cycle of the virus [35]. Furthermore, experimental data have demonstrated that some mimiviruses' aaRS are indeed functional [4,36,37]. Protein structural and functional studies confirmed that APMV, MetRS, and TyrRS act as genuine enzymes. It was shown that TyrRS is a homodimer similar to other class I aaRS described so far, but in contrast to what is found in cellular organisms, the viral TyrRS seems to recognize only two bases in tRNA [37]. Although mimivirus amoebal host encodes aaRS, the conservation of aaRS in the mimiviruses' genomes seems to be associated with an increased viral fitness [38]. The expression of APMV aaRS transcripts can be modulated according to the nutritional status of its host: if mimivirus infection takes place in amoebas cultivated on starvation conditions, a higher mRNA expression of aaRS transcripts is observed. This indicates interplay between nutrient availability sensing of amoeba and the stimulation of the mimivirus aaRS genes as a mechanism related to the circumvention of starvation and maintenance of viral replication in usual levels.

It is intriguing that, even if mimiviruses' aaRS seem to be true proteins and related to the improvement of viral fitness, there is a trend of aaRS repertoire loss in the taxon. An accordion-like evolution model was recently proposed, suggesting such a pathway as the natural history of many mimivirus genes [39]. Indeed, it is possible to see the gain and loss of a given aaRS and tRNAs (e.g., ArgRS and Leu tRNA), but an overview of aaRS and tRNAs through mimiviruses suggests reduction of aaRS and tRNA classes' intra- and inter-lineages (Fig 1A). Considering that the occurrence of sequential HGT events involving mimiviruses' aaRS seems uncertain, this scenario may indicate that the mimiviruses' ancestor presented a more complete set of aaRS. Also, we can hypothesize that the loss of those genes might be a consequence of viral genome reduction and specialization to a given cell environment and to a more restricted host-range lifestyle (*Acanthamoeba*) (Fig 2). It is noteworthy that this gradual gene loss was evidenced and accelerated experimentally, causing the loss of TrpRS (and other genes) in APMV [40]. Another important piece of this puzzle is the lack of correspondence

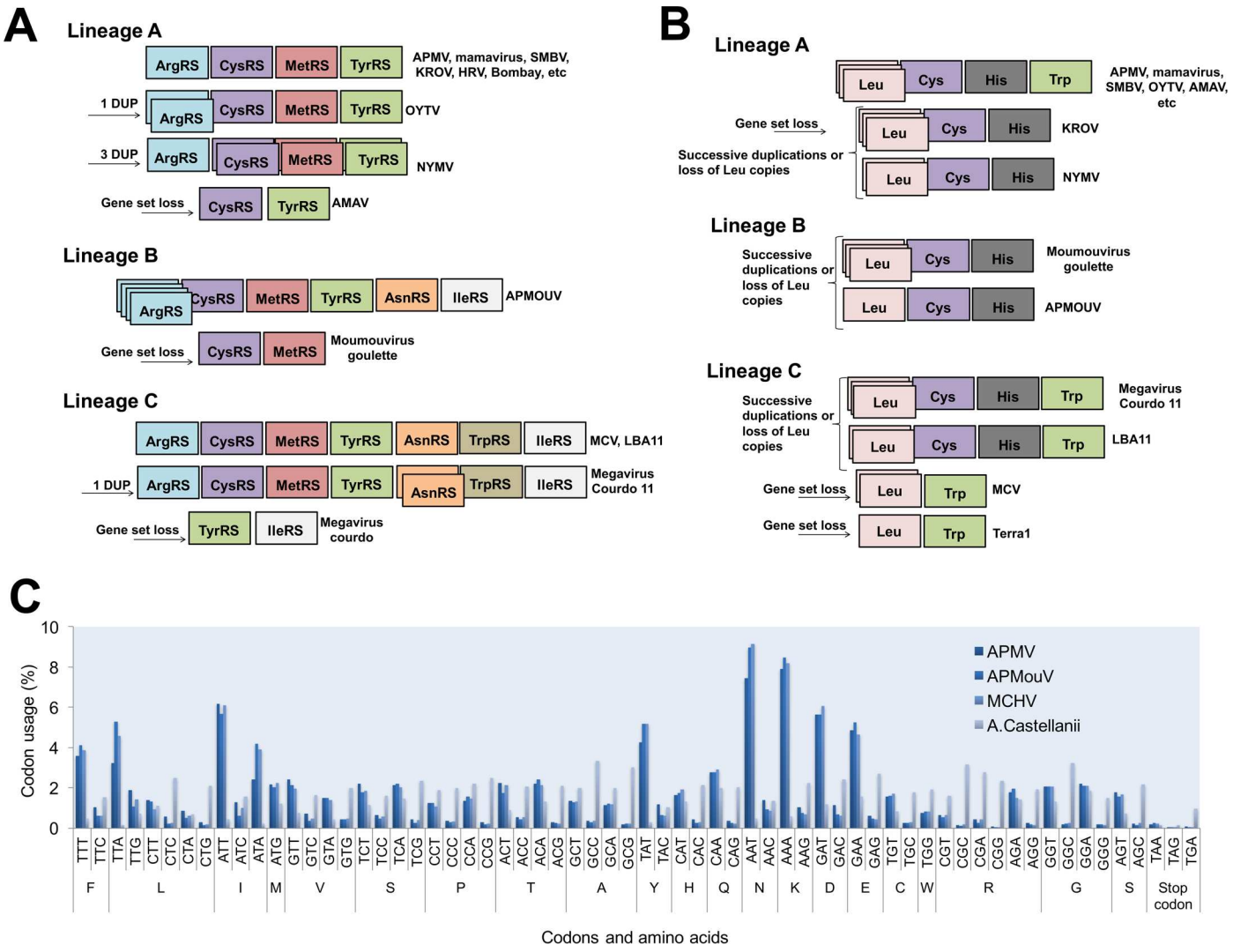


Fig 1. Schematic view of mimiviruses' aaRS (A) and tRNA (B) and codon/amino acid usage (C). The distribution of these genes follows a clear pattern of diversity among *Mimiviridae*: viral genomes with a wealthy aaRS and tRNA gene set (e.g., some lineage C isolates) contain all types of aaRS/tRNAs found in the rest of the family. This indicates a gradual inter- or intra-lineages gene loss throughout the evolution of different mimiviruses. The codon and amino acid usage analysis shows a clear difference between mimiviruses and *Acanthamoeba castellanii* patterns.

doi:10.1371/journal.pgen.1006532.g001

between mimivirus and *Acanthamoeba* codon/amino acid usages [10]. The requirements of mimiviruses for gene translation are quite different than those of *Acanthamoeba*. However, despite the genomes of mimiviruses of lineages A, B, and C presenting important differences and dissimilarities, the codon and amino acid usages are very similar among the members [10] (Fig 1). But, remarkably, the aaRS gene set present in the known mimiviruses does not match with the viral (or host) codon/amino acid usage demands, providing further evidence that a more complex piece of this puzzle has yet to be found.

Transfer RNAs in giant viruses—Expanding the translational apparatus

Similar to aaRS, tRNAs are essential molecules for the process of gene translation, being responsible for transporting an amino acid to a template complementary sequence in the molecule of messenger RNA, where the ribosome will further translate the genetic information [41,42]. The tRNAs are largely diffused among the cellular organisms, being the most abundant type of nucleic acid in the cells and constituting up to 10% of all cellular RNAs [39]. Sequences of tRNAs have already been described in some dsDNA viruses, such as members of the family *Myoviridae* [43], *Herpesviridae* [44], and *Phycodnaviridae* [45]. With the discovery of the mimiviruses, the viral tRNA repertoire increased [4].

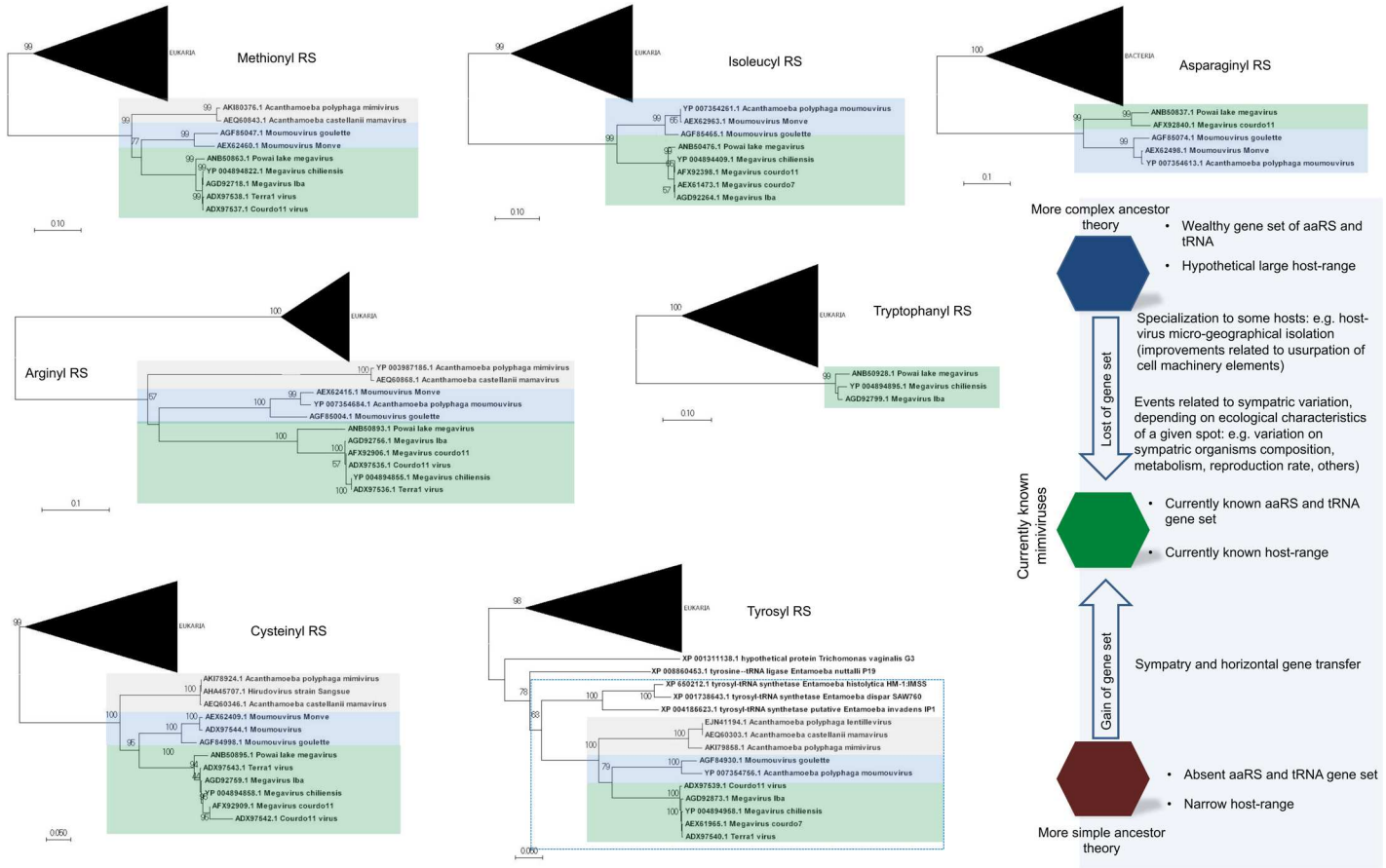


Fig 2. Phylogeny reconstruction of mimiviruses' aaRS. The unrooted trees were generated using MEGA 7 with the maximum likelihood method based on all aaRS found in mimiviruses. The trees were obtained after the alignment of the 100 best hits found in Genbank after BLASTing Megavirus chilensis aaRS predicted aa against all databases. Bacteria or Eukarya taxon, when present, were condensed from the outermost branch presenting bootstrap value >90. Mimiviruses of lineages A (grey), B (blue), and C (green) are highlighted. In all trees but the TyrRS tree, mimiviruses do not cluster inside cellular organism's branches. The bottom–right shows some evolutionary scenarios related to mimivirus evolution that considers their hosts and translation-related genetic data set.

doi:10.1371/journal.pgen.1006532.g002

APMV presents six sequences related to four different tRNAs: leucine (2x TAA and TTG), histidine (CAC), cysteine (TGC), and tryptophan (TGG) [4]. The same tRNAs were found in other mimiviruses of lineage A isolated in Brazil, such as Samba virus, Amazonia virus, Oyster virus, and Kroon virus (except Trp-tRNA) [46], as well as in mimivirus Terra2 [47]. Similar to Kroon virus, Niemeyer virus also has all tRNAs that are found in APMV but Trp-tRNA [15]. It is possible that the evolutionary history of these two viruses differs from that of the other representatives of lineage A, indicating a possible loss of this gene over time. In lineage B, Moumouvirus has sequences related to leucine (TTA), histidine (CAC), and cysteine (TGC), all present in the viruses from lineage A [16]. Considering lineage C, MCV also has three sequences related to tRNA, but only to leucine (TTA and TTG) and tryptophan (TGG) tRNAs [6]. The same tRNAs were identified in the genome of mimivirus Terra1 [47]. By contrast, the mimivirus LBA111 presents all four types of tRNAs found in mimiviruses of lineage A [18]. Similar to viruses from lineage A, it is possible that the ancestor of lineage C viruses had a more complete set of tRNA that was lost during the evolution. If we consider a common ancestor for all group I representatives of the *Mimiviridae* family, the same scenario is reasonable (Fig 1).

In the last years, the family *Mimiviridae* has also expanded the group distantly related to amoeba mimiviruses [48,49]. The analysis of the genome of CroV revealed an even higher range of tRNA-like sequences that had already been identified in their counterparts of group I, with a total of 22 sequences coding for five different tRNAs: leucine (9x TTA), serine (5x TCG), lysine (3x AAA), tyrosine (AAC), and asparagine (AAC), adding some new components in the tRNA set of mimiviruses [5]. Among the algae-infecting mimiviruses, PgV has eight sequences related to tRNA in its genome, coding for leucine (2x TTA e TTG), asparagine (2x AAC), isoleucine (ATA), arginine (AGA), and glutamine (CAA) [19]. OLV presents five tRNAs: leucine (TTG), isoleucine (ATA), tyrosine (TAC), asparagine (AAC), and arginine (AGA) [20] (Table 1). The great diversity of tRNAs coded by mimiviruses of group II is intriguing. Just like for mimiviruses of group I, it is possible that the common ancestor had a more complete set of these molecules that was lost over time. Taking into account the fact that those viruses infect different known hosts (microflagellates and algae), they likely had distinct evolutionary histories and have undergone different selective pressures, which might have contributed to the gain and loss of tRNA genes.

By analyzing the codon/amino acid usage of the mimiviruses and comparing it to the host usage, the hypothesis of multiple events of HGT become even less likely (Fig 2). Among the tRNAs encoded by mimiviruses, leucine (TTA) tRNA is the most common, being present in all of the viruses analyzed so far except for OLV (although it encodes for Leu[TTG]-tRNA), whereas in *Acanthamoeba* sp., it is one of the less frequently occurring tRNAs and is mainly encoded by CTG and CTC [10]. Leu(TAA)-tRNA, which is present in several mimivirus genomes and in multiple copies in some viral genomes, was hypothesized to complement the amoebal tRNA pool and may contribute to accommodating the viral AT-rich codons [10]. The mechanisms of gene expression in the beginning phase of the mimivirus replicative cycle may differ from the mechanisms for gene expression in later phases, and apart from viral RNA transcripts incorporated into mimivirus particles, mimivirus gene expression would first rely primarily on the amoebal machinery and then possibly become increasingly adapted to Mimivirus codon and amino acid usages [10]. In addition, 48% of all mimivirus tRNAs correspond to one of the 10 most frequently used codons in mimiviruses, while 84% of them correspond to one of the 10 least frequently used codons in their hosts [10]. Therefore, such differences suggest that the translational apparatus of the mimiviruses do not come from their currently known natural hosts, which supports the hypothesis that these viruses came from an unknown representative of the fourth TRUC of microbes [30,50,51], although other scenarios cannot be completely ruled out at this point [32,33,52].

Boosting the viral protein synthesis—Translation factors in mimiviruses

The translation of mRNA into proteins involves three major steps: initiation, elongation, and termination. For these steps to occur, some molecules, named translation factors (TFs), are required. Each step demands specific TFs that are essential for the protein synthesis to occur properly, generically known as initiation factors (IFs), elongation factors (EFs), and release factors (RFs) [53]. Each domain of life presents its own TFs that play similar roles during each step of the process [54]. The viruses do not possess such components, which is why they completely rely on the translational apparatus of their hosts. The discovery of the mimiviruses put this rule to the test, suggesting the existence of a possible fourth TRUC of life [7,50] and a "quasi-autonomous" nature of these giant viruses [55].

The genome of APMV presents sequences homologous to five TFs that are related to all three steps of translation: translation initiation factor 4E, translation initiation factor SUI1, translation initiation factor 4A, translation elongation factor eF-TU, and peptide chain release factor eRF1 [4]. The discovery of these genes, along with aaRS, was a milestone in virology since, until then, no virus was known to harbor sequences related to the translational apparatus, prompting an old debate: whether the viruses are living organisms and if they deserve a special place in the tree of life [4,56,57]. The same TFs were found in the genomes of several other mimiviruses of group I, and some of these genes are also present in the smaller mimiviruses, sequences homologous to IFs being found in the genomes of CroV [5], OLV [20], and PgV [19], but EFs and RFs are absent in these viruses (Table 1). The presence of TFs related to the initial step of translation in representatives of the whole family *Mimiviridae* suggests that these viruses have a weaker dependence on their hosts in the beginning of their replication cycle.

Until now, to the best of our knowledge, there have been no experimental studies about mimiviral initiation or elongation factors, which presents a gap in the biology of mimiviruses. However, work by Saini and Fischer (2007), based on 3-D models and an analysis of the conservation of functionally important residues and motifs, demonstrated that it is possible to derive functional attributes for six APMV ORFans, including initiation factor eIF4E [58]. Regarding the peptide chain RF present in mimiviruses, it was initially suggested it was a class-I RF with a sequence homologous to the RF found in eukaryotes and archaea [4]. The genes that encode eRF1 in APMV and MCV present two stop codons inside the coding region, and for the protein to be accurately synthesized, the viruses must be able to change the reading frame (frameshift recoding event), as well as bypass one of the stop codons (readthrough recoding event), which were considered exclusive features of bacteria [59]. Thus, the analysis of mimivirus RFs revealed a new type of TF that had never been seen before in any known group of organism.

This unique mixture of features of RFs from Eukarya, Archaea, and Bacteria found in mimiviruses raised further questions about their origin. The presence of a completely new type of RF makes its origin by successive events of HGT unlikely, thus supporting the hypothesis that the ancestral mimiviruses constituted a fourth TRUC of life [7,50]. If this scenario is true, we might consider that this ancestor presented a complete apparatus that has been gradually lost over successive speciation events. More studies about the TF of giant viruses will bring new insights about this issue, providing valuable clues to the intriguing mystery that is the origin and evolution of *Megavirales*.

What comes next? Unraveling the giant viruses' origin and evolutionary history

Since the discovery of mimiviruses, many theories regarding their origin and evolutionary history have arisen. As soon as the first mimivirus was discovered and its genome analyzed,

authors began to hypothesize that this virus stands within the tree (or rhizome [60]) of life [4]. In the following years, the discovery of new giant viruses increased their known pangenome, which supported the initial theories and opened windows for new ones by suggesting that they originated from a fourth TRUC of life [34,50] and also that they probably coexisted with cellular ancestors, evolving mainly through a genome reductive pattern [30,31]. Nevertheless, these theories were readily contested, with some researchers arguing that giant viruses should not be placed in the tree of life and that they came from other small parasitic elements instead of from an extinct branch of life, which had evolved by several HGT events [32,33,56,61,].

Despite many theories, the origin and evolution of mimiviruses remains a breathtaking mystery. It is not yet possible to exclude any specific scenario of giant viruses' evolution (Fig 2). However, the discovery of new viruses with exceptional genetic content, boosted by the advancement of phylogenomic analysis, provided increasing evidence to support a model wherein the mimivirus ancestor was a more complex organism. Taking this into account, we might speculate that this ancestor possessed a more complete translational-related gene set, which had been constantly losing and gaining genes (mainly through duplication) in accordance with an accordion model of evolution [39]. The *Mimiviridae* ancestor would already be a giant virus, but it would be one with a more independent and generalist lifestyle, able to infect different types of host cells or even interact with them, allowing the gene flow between ancestral lineages. This could have contributed to continuous gene gain and loss over time, shaping the whole viral genome and leading to modern mimiviruses. Considering the new techniques for isolating and discovering giant viruses that are currently implemented [62], we expect to find new mimi- and other giant viruses with genomes that more resemble this ancestor, advancing our understanding of the origin and evolution of this lineage of complex viruses.

Supporting information

S1 File. Arginyl RS alignment.
(MAS)

S2 File. Asparaginyl RS alignment.
(MAS)

S3 File. Cisteinyl RS alignment.
(MAS)

S4 File. Isoleucyl RS alignment.
(MAS)

S5 File. Metionyl RS alignment.
(MAS)

S6 File. Tirosyl RS alignment.
(MAS)

S7 File. Tryptophanyl RS alignment.
(MAS)

References

1. Lwoff A. The concept of virus. J Gen Microbiol. 1957; 17: 239–253. doi: 10.1099/00221287-17-2-239 PMID: 13481308
2. ICTV, 2016 –Last taxonomic release: 2015 release.

3. Raoult D, La Scola B, Birtles R. The Discovery and Characterization of Mimivirus, the Largest Known Virus and Putative Pneumonia Agent. Clin Infect Dis. 2007; 45: 95–102. doi: 10.1086/518608 PMID: 17554709
4. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-Megabase Genome Sequence of Mimivirus. Science. 2004; 306: 1344–1350. doi: 10.1126/science.1101485 PMID: 15486256
5. Fischer MG, Allen MJ, Wilson WH, Suttle CA. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc Natl Acad Sci USA. 2010; 107: 19508–19513. doi: 10.1073/pnas.1007615107 PMID: 20974979
6. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci USA. 2011; 108: 17486–17491. doi: 10.1073/pnas.1110889108 PMID: 21987820
7. Colson P, de Lamballerie X, Fournous G, Raoult D. Reclassification of Giant Viruses Composing a Fourth Domain of Life in the New Order Megavirales. Intervirology. 2012; 55: 321–332. doi: 10.1159/000336562 PMID: 22508375
8. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M. A giant virus in amoebae. Science. 2003; 299: 2033. doi: 10.1126/science.1081867 PMID: 12663918
9. Abrahão JS, Dornas FP, Silva LC, Almeida GM, Boratto PV, Colson P, et al. Acanthamoeba polyphaga mimivirus and other giant viruses: an open field to outstanding discoveries. Virol J. 2014; 11: 120. doi: 10.1186/1743-422X-11-120 PMID: 24976356
10. Colson P, Fournous G, Diene SM, Raoult D. Codon Usage, Amino Acid Usage, Transfer RNA and Amino-Acyl-tRNA Synthetases in Mimiviruses. Intervirology. 2013; 56: 364–375. doi: 10.1159/000354557 PMID: 24157883
11. Klose T, Rossmann MG. Structure of large dsDNA viruses. Biol Chem. 2014; 359: 711–719.
12. Legendre M, Santini S, Rico A, Abergel C, Claverie JM. Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. Virol J. 2011; 8: 99. doi: 10.1186/1743-422X-8-99 PMID: 21375749
13. Colson P, Yutin N, Shabalina SA, Robert C, Fournous G, La Scola B, et al. Viruses with More Than 1,000 Genes: Mamavirus, a New Acanthamoeba polyphaga mimivirus Strain, and Reannotation of Mimivirus Genes. Genome Biol Evol. 2011; 3: 737–742. doi: 10.1093/gbe/evr048 PMID: 21705471
14. Campos RK, Boratto PV, Assis FL, Aguiar ER, Silva LC, Albarnaz JD, et al. Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. Virol J. 2014; 11: 95. doi: 10.1186/1743-422X-11-95 PMID: 24886672
15. Boratto PV, Arantes TS, Silva LC, Assis FL, Kroon EG, La Scola B, et al. Niemeyer Virus: A New Mimivirus Group A Isolate Harboring a Set of Duplicated Aminoacyl-tRNA Synthetase Genes. Front Microbiol. 2015; 6: 1256. doi: 10.3389/fmicb.2015.01256 PMID: 26635738
16. Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, Robert C, et al. Related giant viruses in distant locations and different habitats: *Acanthamoeba polyphaga* moumouvirus represents a third lineage of the *Mimiviridae* that is close to the *Megavirus* lineage. Genome Biol Evol. 2012; 4: 1324–1330. doi: 10.1093/gbe/evs109 PMID: 23221609
17. Saadi H, Pagnier I, Colson P, Cherif JK, Beji M, Boughalmi M, et al. First isolation of Mimivirus in a patient with pneumonia. Clin Infec Dis. 2013; 57: e127–134.
18. Yoosuf N, Pagnier I, Fournous G, Robert C, La Scola B, Raoult D, et al. Complete genome sequence of Courdo11 virus, a member of the family Mimiviridae. Virus Genes. 2014; 48: 218–223. doi: 10.1007/s11262-013-1016-x PMID: 24293219
19. Santini S, Jeudy S, Bartoli J, Poirot O, Lescot M, Abergel C, et al. Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. Proc Natl Acad Sci USA. 2013; 110: 10800–10805. doi: 10.1073/pnas.1303251110 PMID: 23754393
20. Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ, et al. Virophage control of antarctic algal host–virus dynamics. Proc Natl Acad Sci USA. 2011; 108: 6163–6168. doi: 10.1073/pnas.1018221108 PMID: 21444812
21. Piacente F, Marin M, Molinaro A, De Castro C, Seltzer V, Salis A, et al. Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (Viosamine). J Biol Chem. 2012; 287: 3009–3018. doi: 10.1074/jbc.M111.314559 PMID: 22157758
22. Piacente F, Bernardi C, Marin M, Blanc G, Abergel C, Tonetti MG. Characterization of a UDP-N-acetylglucosamine biosynthetic pathway encoded by the giant DNAvirus Mimivirus. Glycobiology. 2014; 24: 51–61. doi: 10.1093/glycob/cwt089 PMID: 24107487

23. 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 USA. 2009; 106: 21848–21853. doi: 10.1073/pnas.0911354106 PMID: 20007369

24. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, et al. Pandoraviruses: amoebae viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science. 2013; 341: 281–286. doi: 10.1126/science.1239181 PMID: 23869018

25. Reteno DG, Benamar S, Khalil JB, Andreani J, Armstrong N, Klose T, et al. Faustovirus, an asfavirus-related new lineage of giant viruses infecting amoebae. J Virol. 2015; 89: 6585–6594. doi: 10.1128/JVI.00115-15 PMID: 25878099

26. Legendre M, Lartigue A, Bertaux L, Jeudy S, Bartoli J, Lescot M, et al. In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. Proc Natl Acad Sci USA. 2015; 112: E5327–5335. doi: 10.1073/pnas.1510795112 PMID: 26351664

27. Ribas de Pouplana L, Schimmel P. Aminoacyl-tRNA synthetases: potential markers of genetic code development. Trends Biochem Sci. 2001; 26: 591–596. PMID: 11590011

28. Li R, Macnamara LM, Leuchter JD, Alexander RW, Cho SS. MD Simulations of tRNA and Aminoacyl-tRNA Synthetases: Dynamics, Folding, Binding, and Allostery. Int J Mol Sci. 2015; 16: 15872–15902. doi: 10.3390/ijms160715872 PMID: 26184179

29. Fournier GP, Andam CP, Alm EJ, Gogarten JP. Molecular Evolution of Aminoacyl tRNA Synthetase Proteins in the Early History of Life. Orig Life Evol Biosph. 2011; 41: 621–632. doi: 10.1007/s11084-011-9261-2 PMID: 22200905

30. Nasir A, Kim KM, Caetano-Anollés G. Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. BMC Evol Biol. 2012; 12: 156. doi: 10.1186/1471-2148-12-156 PMID: 22920653

31. Nasir A, Caetano-Anollés G. A phylogenomic data-driven exploration of viral origins and evolution. Sci Adv. 2015; 1: e1500527. doi: 10.1126/sciadv.1500527 PMID: 26601271

32. Yutin N, Wolf YI, Koonin EV. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology. 2014; 466–467: 38–52. doi: 10.1016/j.virol.2014.06.032 PMID: 25042053

33. Koonin EV, Krupovic M, Yutin N. Evolution of double-strand DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. Ann N Y Acad Sci. 2015; 1341: 10–24. doi: 10.1111/nyas.12728 PMID: 25727355

34. Boyer M, Madoui MA, Gimenez G, La Scola B, Raoult D. Phylogenetic and Phyletic Studies of Informational Genes in Genomes Highlight Existence of a 4th Domain of Life Including Giant Viruses. PLoS ONE. 2010; 5: e15530. doi: 10.1371/journal.pone.0015530 PMID: 21151962

35. Legendre M, Audic S, Poirot O, Hingamp P, Seltzer V, Byrne D, et al. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. Genome Res. 2010; 20: 664–674. doi: 10.1101/gr.102582.109 PMID: 20360389

36. Abergel C, Chenivesse S, Byrne D, Suhre K, Arondel V, Claverie JM. Mimivirus TyrRS: preliminary structural and functional characterization of the first amino-acyl tRNA synthetase found in a virus. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2005; 61(Pt 2): 212–215. doi: 10.1107/S174430910500062X PMID: 16510997

37. Abergel C, Rudinger-Thirion J, Giegé R, Claverie JM. Virus-Encoded Aminoacyl-tRNA Synthetases: Structural and Functional Characterization of Mimivirus TyrRS and MetRS. J Virol. 2007; 81: 12406–12417. doi: 10.1128/JVI.01107-07 PMID: 17855524

38. Silva LC, Almeida GM, Assis FL, Albarnaz JD, Boratto PV, Dornas FP, et al. Modulation of the expression of mimivirus-encoded translation-related genes in response to nutrient availability during *Acanthamoeba castellanii* infection. Front Microbiol. 2015; 6: 539. doi: 10.3389/fmicb.2015.00539 PMID: 26082761

39. Filée J. Genomic comparison of closely related giant viruses supports an accordion-like model of evolution. Front Microbiol. 2015; 6: 593. doi: 10.3389/fmicb.2015.00593 PMID: 26136734

40. Boyer M, Azza S, Barrassi L, Klose T, Campocasso A, Pagnier I, et al. Mimivirus shows dramatic genome reduction after intraamoebal culture. Proc Natl Acad Sci USA. 108: 10296–10301. doi: 10.1073/pnas.1101118108 PMID: 21646533

41. Giegé R. The early history of tRNA recognition by aminoacyl-tRNA synthetases. J Biosci. 2006; 31: 477–488. PMID: 17206068

42. Barciszewska MZ, Perrigue PM, Barciszewski J. tRNA–the golden standard in molecular biology. Mol Biosyst. 2016; 12: 12–17. doi: 10.1039/c5mb00557d PMID: 26549858

43. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. Bacteriophage T4 Genome. Microbiol Mol Biol Rev. 2003; 67: 86–156. doi: 10.1128/MMBR.67.1.86-156.2003 PMID: 12626685

44. Virgin HW 4th, Latreille P, Wamsley P, Hallsworth K, Weck KE, Dal Canto AJ, et al. Complete Sequence and Genomic Analysis of Murine Gammaherpesvirus 68. J Virol. 1997. 71: 5894–5904. PMID: 9223479

45. Van Etten JL, Meints RH. Giant viruses infecting algae. Annu Rev Microbiol. 1999; 53: 447–494. doi: 10.1146/annurev.micro.53.1.447 PMID: 10547698

46. Assis FL, Bajrai L, Abrahao JS, Kroon EG, Dornas FP, Andrade KR, et al. Pan-Genome Analysis of Brazilian Lineage A Amoebal Mimiviruses. Viruses. 2015; 7: 3483–3499. doi: 10.3390/v7072782 PMID: 26131958

47. Yoosuf N, Pagnier I, Fournous G, Robert C, Raoult D, La Scola B, et al. Draft genome sequences of Terra1 and Terra2 viruses, new members of the family Mimiviridae isolated from soil. Virology. 2014; 452–453:125–132. doi: 10.1016/j.virol.2013.12.032 PMID: 24606689

48. Yutin N, Colson P, Raoult D, Koonin EV. Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. Virol J. 2013; 10: 106. doi: 10.1186/1743-422X-10-106 PMID: 23557328

49. Claverie JM. Giant virus in the sea: Extending the realm of Megaviridae to Viridiplantae. Commum Integr Biol. 2013; 6: e25685.

50. Raoult D. TRUC or the need for a new microbial classification. Intervirology. 2013; 56: 349–353. doi: 10.1159/000354269 PMID: 23867259

51. Seligmann H, Raoult D. Unifying view of stem-loop hairpin RNA as origin of current and ancient parasitic and non-parasitic RNAs, including in giant viruses. Curr Opin Microbiol. 2016; 31: 1–8. doi: 10.1016/j.mib.2015.11.004 PMID: 26716728

52. Moreira D, López-García P. Comment on "The 1.2-megabase genome sequence of Mimivirus". Science. 2005; 308: 1114. doi: 10.1126/science.1110820 PMID: 15905382

53. Noble CG, Song H. Structural studies of elongation and release factors. Cell Mol Life Sci. 2008; 65: 1335–1346. doi: 10.1007/s00018-008-7495-6 PMID: 18213444

54. Kyrpides NC, Woese CR. Universally conserved translation initiation factors. Proc Natl Acad Sci USA. 1998; 95: 224–228. PMID: 9419357

55. Claverie JM, Abergel C. Mimivirus: the emerging paradox of quasi-autonomous viruses. Trends Genet. 2010; 26: 431–437. doi: 10.1016/j.tig.2010.07.003 PMID: 20696492

56. Moreira D, Lópes-García P. Ten reasons to exclude viruses from the tree of life. Nat Rev Microbiol. 2009; 7: 306–311. doi: 10.1038/nrmicro2108 PMID: 19270719

57. Raoult D. There is no such thing as a tree of life (and of course viruses are out!). Nat Rev Microbiol. 2009. 7: 615.

58. Saini HK, Fischer D. Structural and functional insights into Mimivirus ORFans. BMC Genomics. 2007; 8: 115. doi: 10.1186/1471-2164-8-115 PMID: 17490476

59. Jeudy S, Abergel C, Claverie JM, Legendre M. Translation in Giant Viruses: A Unique Mixture of Bacterial and Eukaryotic Termination Schemes. PLoS Genet. 2012; 8: e1003122. doi: 10.1371/journal.pgen.1003122 PMID: 23271980

60. Raoult D. The post-Darwinist rhizome of life. Lancet. 2010; 375: 104–105. doi: 10.1016/S0140-6736(09)61958-9 PMID: 20109873

61. Moreira D, Brochier-Armanet C. Giant viruses, giant chimeras: The multiple evolutionary histories of Mimivirus genes. BMC Evol Biol. 2008; 8: 12. doi: 10.1186/1471-2148-8-12 PMID: 18205905

62. Khalil JY, Andreani J, La Scola B. Updating strategies for isolating and discovering giant viruses. Curr Opin Microbiol. 2016; 31: 80–87. doi: 10.1016/j.mib.2016.03.004 PMID: 27039269

4.4. ARTIGO #4: The complex nature of Tupanviruses

Descritos constantemente nos últimos 15 anos, os vírus gigantes são um grupo de entidades com complexidade genômica e estrutural nunca antes observada na virosfera. Sua descoberta levantou questões importantes sobre a diversidade, ecologia e evolução destes vírus. A família *Mimiviridae* foi o primeiro grupo de vírus gigantes de amebas a ser descrito, sendo composta por vírus com diversas características que desafiaram vários conceitos da virologia clássica. Os tupanvírus estão estre os membros mais novos desta família a serem descritos e exibem características estruturais, genéticas e biológicas nunca antes observadas para outros vírus gigantes. A complexidade destes vírus isolados de ambientes extremos no Brasil auxiliou na compreensão da biologia e evolução dos vírus gigantes, mas também levantou importantes questões que precisam ser analisadas. Neste artigo é contada a história por trás do descobrimento de um dos grupos de vírus mais complexos isolados até o momento. Uma ampla revisão da estrutura, ciclo de infecção e perfis de interação com diferentes células é apresentada. Ressaltamos ainda a complexidade gênica encontrada no genoma dos tupanvírus e realizamos novas análises filogenéticas utilizando diferentes estratégias para obtenção de sequências homólogas em bancos de dados para genes de aminoacil-tRNA sintetases e citrato sintase, enzimas chaves nos metabolismos proteico e energético, respectivamente, e que foram detectadas nos genomas destes vírus. Nossos resultados sugerem que o uso da filogenia de um ou poucos genes não é suficiente para encerrar o debate sobre a origem dos vírus gigantes, pois as análises são fortemente influenciadas pela estratégia de obtenção de sequências adotada para reconstruções filogenéticas. Ainda, revisamos a descoberta de sequências ribossomais nos tupanvírus e propomos hipóteses de sua origem entre os membros da família *Mimiviridae*. Finalmente, discutimos como as características exclusivas dos tupanvírus contribuíram para redefinir os limites da virosfera.

Este artigo foi publicado no periódico ***Advances in Virus Research*** em novembro de 2018 (versão online).

The Complex Nature of Tupanviruses

Rodrigo Araújo Lima Rodrigues, Thalita Souza Arantes, Graziele Pereira Oliveira, Ludmila Karen dos Santos Silva, Jônatas Santos Abrahão[1]

Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[1] Corresponding author: *Email address:* jonatas.abrahao@gmail.com (J.S. Abrahão)

Contents

Abstract

The discovery of giant viruses revealed a new level of complexity in the virosphere, raising important questions about the diversity, ecology, and evolution of these viruses. The family *Mimiviridae* was the first group of amoebal giant viruses to be discovered (by Bernard La Scola and Didier Raoult team), containing viruses with structural and genetic features that challenged many concepts of classic virology. The tupanviruses are among the newest members of this family and exhibit structural, biological, and genetic features never previously observed in other giant viruses. The complexity of these viruses has put us one step forward toward the comprehension of giant virus biology and evolution, but also has raised important questions that still need to be addressed. In this chapter, we tell the history behind the discovery of one of the most complex viruses isolated to date, highlighting the unique features exhibited by tupanviruses, and discuss how these giant viruses have contributed to redefining limits for the virosphere.

Advances in Virus Research, Volume xx
ISSN 0065-3527, https://doi.org/10.1016/bs.aivir.2018.09.001

1. DISCOVERY OF NEW GIANT VIRUSES IN UNEXPLORED ENVIRONMENTS

Since the remarkable isolation of *Acanthamoeba polyphaga* mimivirus (APMV) by the team of Bernard La Scola and Didier Raoult in 2003, the research for new giant viruses has significantly increased (La Scola et al., 2003). The use of high-throughput methods and the use of different amoebae species for co-culture approach have allowed the isolation of viruses, such as marseillevirus, pandoravirus, pithovirus, mollivirus, kaumoebavirus and others (Andreani et al., 2016; Bajrai et al., 2016; Boyer et al., 2009; Legendre et al., 2014, 2015; Philippe et al., 2013; Reteno et al., 2015). These viruses are ubiquitous in the environment, often isolated from water, soil, clinical samples, and insects among other sources (Aherfi et al., 2016; Dornas et al., 2015). In addition, metagenomic data have reinforced the ubiquitousness of these viruses, in which giant virus sequences have been found all around the world in places with different levels of biodiversity, and exhibiting a taxa richness that exceeds those of Bacteria and Archaea in the oceans (Dann et al., 2016; Kerepesi and Grolmusz, 2016, 2017; Mihara et al., 2018; Monier et al., 2008). In Brazil, in particular, distinct giant viruses were isolated from different sources, including water, soil, sewage, and others (Andrade et al., 2014, 2018; Assis et al., 2017; Boratto et al., 2015; Dornas et al., 2015, 2016; Dos Santos et al., 2016; dos Santos Silva et al., 2015).

Brazil is one of the most biodiverse countries in the world, and its continental extension allows a large climatic variation, which favors the formation of distinct biomes (Rodrigues et al., 2016). In this context, we highlight a biome known as Soda Lake, an environment that conserves and/or mimics ancient life conditions (extremely high salinity and pH) and is considered one of the most extreme aquatic environments on Earth. Soda lakes are sodium carbonate-dominated environments with varying salinity and high pH values, usually between 9 and 11, but occasionally as high as 13 (Hammer, 1986; Melack, 1981). These ecosystems have been described worldwide, including the East African Rift Valley (Jones et al., 1977), the North and Central Americas (Domagalski et al., 1989), Asia (Ma and Edmunds, 2006), Australia (Hammer, 1986), and Europe (Felföldi et al., 2009). The Brazilian soda lakes are located in the

Pantanal of Mato Grosso do Sul, in a region named Nhecolândia, which has a series of hundreds of soda lakes coexisting with fresh water, in an area of 24,000 km^2. In the last decade, the studies in soda lakes have proliferated and, increasingly, have been the subjects of investigations of the microbial diversity present in these systems (Antony et al., 2013; Sorokin et al., 2015). Another important Brazilian biome is located on the extensive Brazilian marine coast of 3.5 million km^2 and includes ecosystems, such as coral reefs, dunes, mangroves, lagoons, estuaries, and the deep-sea (over 1000 m in depth). These ecosystems play important roles in ecological processes. Although viruses are the most abundant "life forms" in the world's oceans (Suttle, 2005), studies on the presence of these entities have been barely explored in deep-ocean sediments, which could be the source for different and intriguing viruses.

Despite the increase of giant virus isolation in the last decades, only a few were isolated from extreme environments, i.e., pithovirus, mollivirus and mimivirus (Andrade et al., 2018; Legendre et al., 2014, 2015). Taking this into account, our group, in collaboration with the Bernard La Scola and Didier Raoult group (Aix Marseille University), conducted prospective studies in extreme environments found in Brazil, which led to the discovery of a new kind of virus: the tupanviruses (Abrahão et al., 2018). Initially, soil samples of the Pantanal soda lake region were inoculated with cultures of *Acanthamoeba castellanii*, a standard organism used for isolation of giant viruses, which exhibited a cytopathic effect characterized by rounding and amoebae aggregation. The observation of the sample by electron microscopy revealed the presence of a virus with an extremely distinct morphology, presenting a long tail coupled to the capsid. We named this strain Tupanvirus soda lake (TPV-SL) as a tribute to the South American Guarani Indigenous tribes, for whom Tupan—or Tupã—is the God of Thunder, an important mythological figure. Driven by this discovery we performed a prospective study in deep ocean sediments collected at 3000 m depth in the region of Campos dos Goytacazes, in the Brazilian Atlantic Ocean. Surprisingly, we isolated another new strain of tupanvirus, which was named Tupanvirus deep ocean (TPV-DO). Although these viruses were found in extreme environments, it is possible that tupanvirus relatives are present in other different places. However, it is intriguing that no tupanvirus DNA-dependent RNA polymerase sequence was found in ocean metagenomics data, while sequences from other *Mimiviridae* relatives are even more abun-

dant when compared to bacterial and archaeal organisms (Mihara et al., 2018). These viruses present biological and genomic characteristics that distinguish them from all other known viruses and considerably expanded our knowledge of the virosphere. In the next sections we describe the remarkable features of tupanviruses and discuss how this discovery contributes to the advance of mimivirology.

2. MORPHOLOGY AND REPLICATION CYCLE OF TUPANVIRUSES

The discovery of giant viruses has evidenced new structures and placed the viral world in a comparable dimension to cellular organisms, with viruses ranging up to 1.5 μm in size (*Pithovirus sibericum*), dimensions that exceed some bacterial cells (Legendre et al., 2014). The huge particle size exhibited by the giant viruses of amoebae is one of the reasons their discovery was only made in 2003 (over a century after the discovery of the first viruses), with mimiviruses exhibiting a peculiar pseudo-icosahedral particle (Xiao et al., 2009). The tupanviruses particle is composed of a capsid, similar to that of mimiviruses, with a size of ~ 450 nm in diameter, a vertex modified in a starfish shape containing the star-gate portal (Zauberman et al., 2008), and fibrils covering the majority of the structure. These fibrils seem to be slightly shorter and thicker, compared to the fibrils of mimiviruses, but could have an important role in the attachment of viral particles in the host cells, in an analogous way as observed for mimivirus (Rodrigues et al., 2015). In addition, a lipid membrane was observed in the inner part of the capsid and, as in other giant viruses, it is associated with the beginning of the replication cycle of tupanviruses. Despite these similarities, tupanvirus presents a long cylindrical tail attached to the capsid, a unique structure among giant viruses. This structure is ~ 550 nm in extension and ~ 450 nm in diameter, including fibrils, and is coupled to the capsid base in the opposite region of the star gate (Abrahão et al., 2018). The nature of the content of the tail is still uncertain. Although they do not appear to be tightly bound, mechanical attempts and different enzymatic treatments were not able to separate the tail from the capsid of the tupanviruses (Abrahão et al., 2018). The average length of a complete virion is 1.2 μm, although some particles can reach lengths of up to 2.3 μm due to the variation in the

tail's size, which makes them one of the longest viral particles described to date.

The replication cycle of tupanviruses was studied by transmission electron microscopy in *A. castellanii* and *Vermamoeba vermiformis* and occurs in a similar way in both cells (Fig. 1). Initially the viral particles attach to the cell surface and are internalized by phagocytosis, and subsequently observed within phagosomes in the cytoplasm (1–2 h). The inner lipid membrane of the capsid fuses with the phagosome membrane, releasing the viral genome through the star-gate portal. Interestingly, the opening of the star gate can be preceded by invagination of the phagosome membrane into the tail, causing the release of the tail's content into the cytoplasm of the amoebas. A typical phase of eclipse occurs, followed by the formation of an electron-dense viral factory in the host cytoplasm. The viral morphogenesis begins to be observed after 7–12 h of infection. During morphogenesis, it was possible to observe the presence of several lamellar structures in the inner part of the viral factory preceding the capsid formation. These lamellar structures have different sizes and begin in a similar way to the crescents described for mimiviruses and marseilleviruses (Andrade et al., 2017; Arantes et al., 2016). Analysis of viral morphogenesis revealed that the tail is attached to the capsid only after the formation and closure of the capsid. In later times (16–24 h), the cytoplasm of the amoeba is filled with mature viral particles, followed by cell lysis and virus release. It is noteworthy that the tupanvirus cycle is considerably longer compared to the cycle of other giant viruses (e.g., mimivirus and marseilleviruses), and reflects a higher complexity of these new viruses in comparison to previously known viruses.

3. A GIANT VIRUS WITH A BROAD HOST RANGE

The majority of giant viruses are often called amoeba-infecting viruses, although the term "amoebae" encompasses one of the most diverse groups of organisms in nature comprising the Amoebozoa taxa, wherein genera such as *Acanthamoeba* and *Vermamoeba* belong to different classes, Tubulinea and Discosea, respectively (Khan, 2006; Schilde and Schaap, 2013). Species of the *Acanthamoeba* genus are the hosts for most giant viruses including pandoraviruses, pacmanvirus, mollivirus, cedratvirus, marseillevirus, and most isolates of mimivirus (Aherfi et al., 2014; Andreani et al., 2016, 2017; Bajrai et al., 2016;

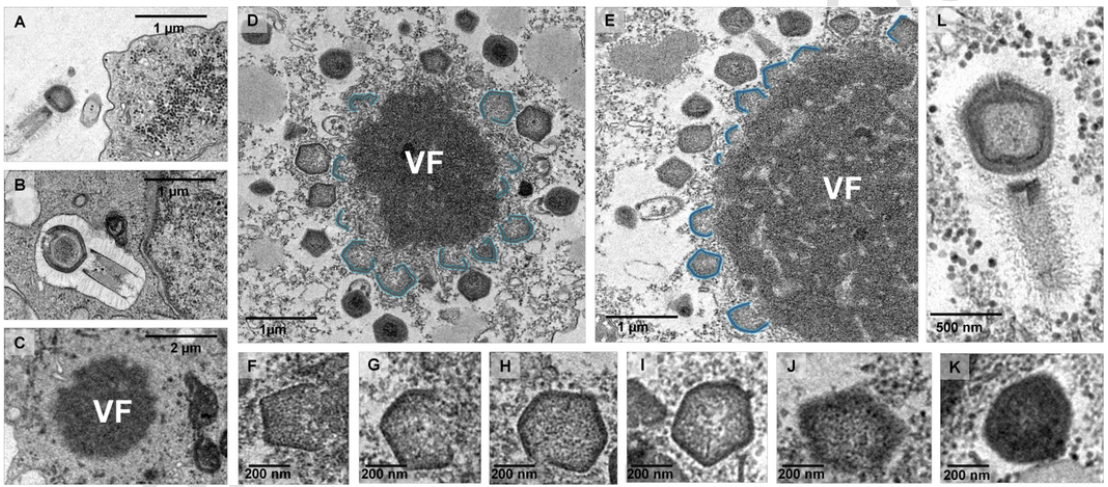


Fig. 1. Replication cycle of tupanviruses by transmission electron microscopy analyses. (A) Tupanvirus particle phagocytosis; (B) viral particle inside of a phagosome; (C) early viral factory; (D, E) mature viral factories (viral particles in different morphogenesis steps are highlighted in blue); (F–K) growing lamellar structures with different sizes in the viral factory, demonstrating the viral crescent-like structures until the fully formation of the capsid; (L) complete viral particle.

Legendre et al., 2014, 2015; Philippe et al., 2013; Reteno et al., 2015). On the other hand, faustovirus, kaumoebavirus, and orpheovirus are phylogenetically related groups of giant viruses that can infect *V. vermiformis*, another species of amoebae (Andreani et al., 2018; Bajrai et al., 2016; Reteno et al., 2015). Curiously, distinct from other giant viruses, tupanviruses are able to replicate in different types of hosts, exhibiting four different profiles of interaction (Fig. 2) (Abrahão et al., 2018). Despite the large evolutionary distance between many members of Amoebozoa, as well as physiological differences, tupanviruses can produc-

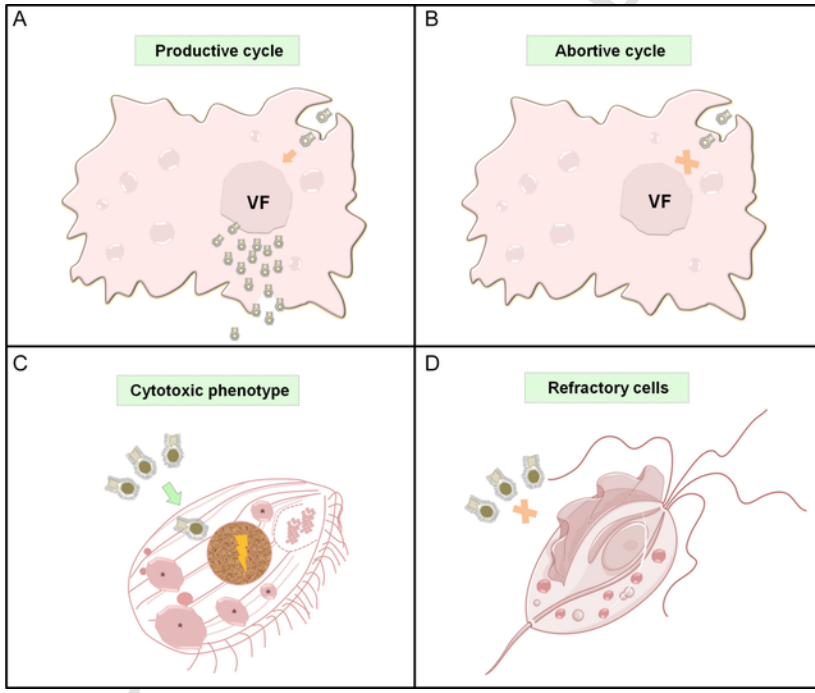


Fig. 2. Interaction profiles of tupanviruses. (A) Productive cycle: The productive cycle involves the virus entry, morphogenesis, and formation of a new viral progeny. A productive infection is observed in *A. castellanii*, *A. polyphaga*, *A. sp E4*, *A. griffini*, *V. vermiformis*, *Dictyostelium discoideum*, and *Willartia magna*; (B) abortive cycle: In the abortive cycle the virus enters the cell, the genome is replicated but no viral progeny are generated. This profile is observed in *A. michelline* and *A. royreba*; (C) cytotoxic phenotype: The particles are phagocytized but there is neither genome replication nor viral progeny generation and the non-host cells exhibit a cytotoxic phenotype involving vacuole formation and nuclear degradation. This profile is observed in *Tetrahymena* sp.; (D) refractory cells: This profile was observed with *Trichomonas tenax*, where there is a refractory interaction, i.e., the virus is not able to enter the cells. VF: viral factory; *vacuoles formation.

tively infect *A. castellanii*, *A. polyphaga*, *A. sp E4*, *A. griffini*, *V. vermiformis*, *Dictyostelium discoideum*, and *Willartia magna*, showing the broadest host spectrum described among the giant viruses (Abrahão et al., 2018). In addition, tupanvirus displays an abortive cycle in *A. michelline* and *A. royreba*, exhibiting evident cytopathic effect and genome replication, but with no particle formation. Moreover, it is not able to replicate within *Tetrahymena* sp., but viral particles were phagocytized by the protist and were able to release their capsid and tail content in the host cytoplasm. Furthermore, some alterations in this protozoa cell were also observed during tupanvirus infection, associated with a previously undescribed cytotoxic profile. Finally, tupanviruses were not able to replicate in more distant protozoa, such as *Trichomonas* sp., thus constituting refractory cells and the fourth replication profile (Abrahão et al., 2018).

The majority of giant viruses are specialized in acanthamoeba species, amoebas that have been found in diverse environments such as air, soil, water, and animal bodies, and suggested as one of the most ubiquitous protozoans in nature (Khan, 2006). This provides viruses with a wider access to reach unlimited susceptible hosts and environments. In contrast, tupanviruses were found in extreme environments with high salinity and pH (soda lake) and under massive pressure in ocean sediments collected at a depth of 3000 m, harsh conditions to survive for most known life forms (Abrahão et al., 2018). In addition, hypersaline lakes have a lower biodiversity and a less dense population of surviving organisms compared to other environments (Sergeev et al., 2002). The inhospitable characteristics of these environments may be directly correlated with the generalist host range profile presented by tupanviruses. In these environments, certain host populations would be limited by the available conditions and resources, and the virus would have a lower number of hosts, hindering the encounter between viral particle and host cell. Therefore, a generalist strategy of broad host range could provide tupanviruses with a greater possibility of finding suitable hosts, even under extreme conditions. The description of other tupanviruses in different environments could provide valuable insights into the ecology of these complex viruses and the answers for some of the hypothesis raised here.

4. TOXICITY OF TUPANVIRUSES AND RIBOSOMAL SHUTDOWN: AN UNSOLVED MYSTERY

Tupanviruses present remarkable structural and genomic characteristics, as well as the broadest host range among the giant viruses (Abrahão et al., 2018). During the study of tupanvirus hosts, it was observed that tupanvirus triggered a cytotoxic profile in *Tetrahymena hyperangularis*, a non-host organism. *Tetrahymena* is a genus composed of ciliated, ovoid/pyriform protozoa, which members exhibit high motility and rate of phagocytosis, a process used for feeding similar to other protozoa (Grønlien et al., 2002). The cytotoxic profile was also observed in host cells (*A. castellanii*) at a high multiplicity of infection (MOI), a phenotype not observed for APMV (Abrahão et al., 2018). Associated with this toxic profile, the shutting down of host ribosomal RNA (rRNA) abundance was observed, a phenomenon not previously detected in the field of giant viruses. Different shutdown profiles were observed between *T. hyperangularis* and *A. castellanii*. The effect was clearly observed after 4 days of infection in *T. hyperangularis*, whereas in *A. castellanii*, it was observed at around 9 h of infection. In addition, in *T. hyperangularis*, an interesting loss of motility, increase of vacuolization, presence of a large amount of extracellular vesicles, and a decrease in the phagocytosis rate, indicated by the reduced ingestion of tupanvirus' particles, were detected (Abrahão et al., 2018).

Crucial for the discovery of the shutdown phenomenon in host cells were the failed attempts to perform viral RNA sequencing (RNA-seq), since the quantification of RNA after viral replication at high MOI was recurrently lower than expected. In addition, electrophoretic analyses demonstrated an unexpected absence of ribosomal subunits 18S and 28S, which suggested the occurrence of ribosomal degradation. From this moment forward, different analyses were carried out to identify what would lead to the reduction of host rRNA. The first hypothesis was related to the process of ribophagy, an autophagy process responsible for the degradation of ribosomes in prolonged periods of nutrient deprivation (Kraft et al., 2008). The autophagy is a canonical mechanism of macromolecule and cell component degradation, which has some typical markers such as formation of double-membrane vesicles, acidification of autophagosomes, and the participation of a group of proteins coded by

genes referred to as *Atg* (autophagy-related genes) (Choi et al., 2018; Kraft et al., 2008). These genes are widely spread among cellular organisms, and genes homologous to the Atg-3, Atg-8, and Atg-13 were identified in *A. castellanii* (Moon et al., 2015). Initial assays using the pharmacological inhibitors chloroquine and bafilomycin A were not able to prevent rRNA shutdown, suggesting that both lysosomal acidification and autophagosome/lysosome fusion are non-obligatory processes for the occurrence of ribosomal degradation induced by tupanvirus infection. This is the first evidence that a classical ribophagy process might not be responsible for RNA shutdown (Abrahão et al., 2018; Jha et al., 2014; Moon et al., 2015; Redmann et al., 2017; Solitro and MacKeigan, 2016). Another indication for the absence of ribophagy came from gene silencing assays, where siRNA targeting Atg8-2 in *A. castellanii* cells infected with tupanvirus did not prevent ribosomal shutdown (Abrahão et al., 2018). In addition, assays to evaluate the occurrence of acidification of cell compartments revealed an intense acidification of the entire cytoplasm of *A. castellanii* infected by tupanvirus (not only cell compartments, as expected in the autophagy process) concomitantly with rRNA degradation (Choi et al., 2018). Furthermore, the presence of double-membrane vesicles containing ribosomes was rarely observed in tupanvirus-infected amoebae, although single-membrane vesicles were frequently noticed in electron microscopy analyses, thus suggesting that the canonical ribophagy process was not responsible for the ribosomal degradation observed during tupanvirus infection (Abrahão et al., 2018).

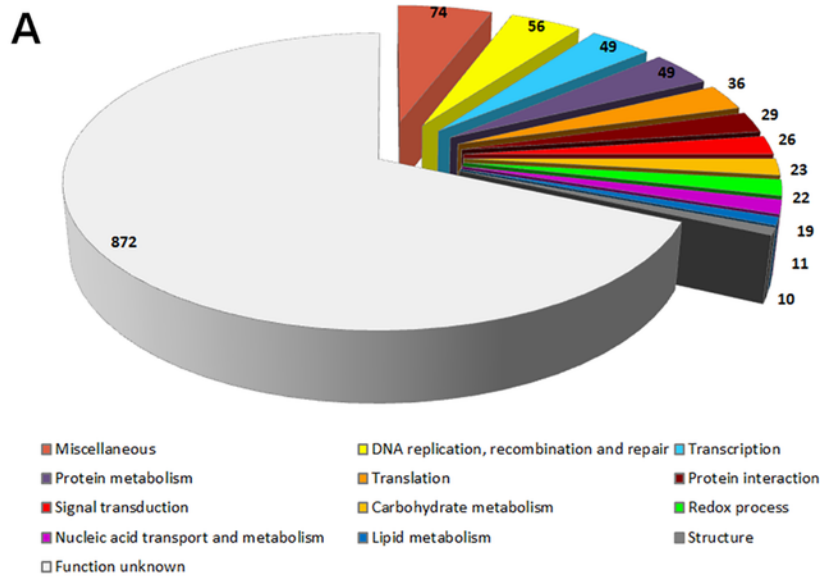
Interestingly, tupanvirus infection also induced nuclear/nucleolar degradation in *A. castellanii* and *T. hyperangularis* cells. The nucleolus is the main cellular component for ribosome biogenesis and its degradation by tupanvirus infection could be a pivotal factor for the shutting down of ribosomes (Nerurkar et al., 2015). It has also been demonstrated that the toxicity profile and rRNA degradation are independent of tupanvirus replication, since the shutdown occurred in the presence of UV inactivated particles, but not with particles inactivated by heat, thus suggesting that this phenomenon could be induced by the presence of a viral factor, most likely a protein element (Abrahão et al., 2018). It is possible that this unknown factor can be carried by the viral particle and released into the cytoplasm of the protozoan. The hypothesis that the formation of vesicles containing ribosomes and nuclear/nucleolar degradation might be related to the occurrence of rRNA shutdown is to date the best at-

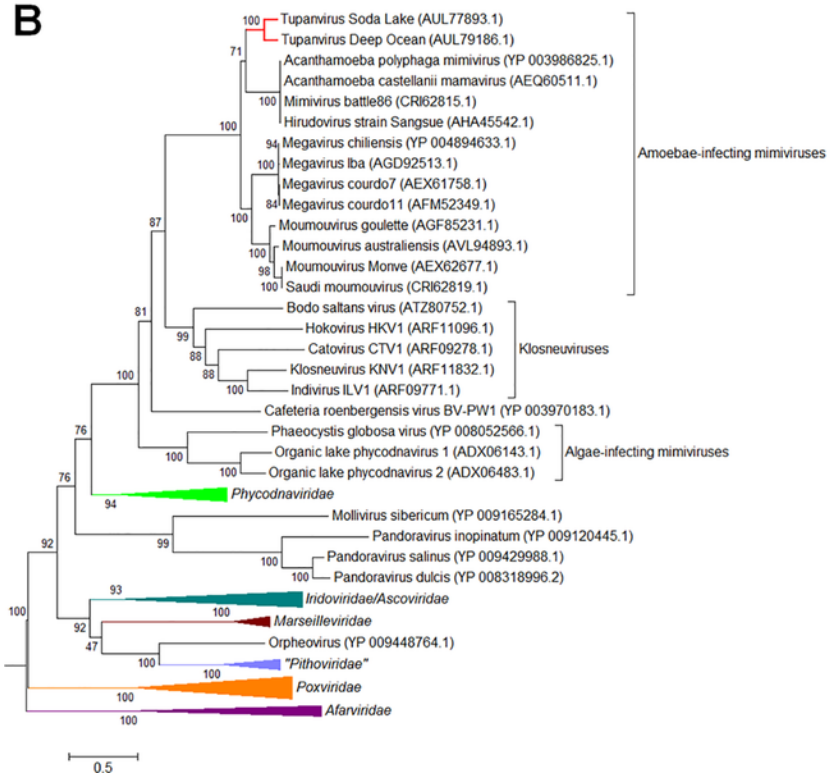
tempts to explain this phenomenon. It is also possible that tupanviruses induce ribosomal modifications that favor the translation of its own proteins, similar to what was observed in poxviruses (Jha et al., 2017). Nevertheless, there is still much to investigate in order to fully understand this peculiar history, possibly revealing a totally new process of RNA degradation.

5. THE COMPLEX GENOMIC AND PROTEOMIC SET OF TUPANVIRUSES

TPV-SL and TPV-DO each have a linear double-stranded DNA genome of 1,439,508 and 1,516,267 bp, respectively (Genbank accession number KY523104 and MF405918), the largest genomes described for members of the family *Mimiviridae*, and smaller only than the genomes of pandoraviruses (Philippe et al., 2013) and orpheovirus IHUMI-LCC2 (Andreani et al., 2018) (considering TPV-SL). The genome of tupanviruses has a coding density of ~ 88% with a total of 1276 (TPV-SL) and 1359 (TPV-DO) open reading frames (ORFs) distributed roughly equally between the two strands, with 407 and 378 ORFans (ORFs with no hits in databases), respectively. The genome is A/T-rich (~ 72%), similar to other amoebae-infecting mimiviruses, which reflect the preferential use of codons formed by A/T triplets, especially asparagine (AAT), lysine (AAA), and isoleucine (ATT) (Abrahão et al., 2018). Moreover, as reported for other mimiviruses, the profile of amino acid and codon usage of tupanviruses are substantially different from *Acanthamoeba* spp. and suggest the absence of large scale events of horizontal transfer with these host organisms in recent evolutionary history. The "AAAATTGA" promoter motif is associated with several tupanviruses genes, in a similar frequency as observed for other members of the family *Mimiviridae* (Suhre et al., 2005), which is associated with the control of early expressed genes. The true profile of gene expression of tupanviruses is still unknown, but it is expected that it is similar to that observed for APMV, where three well-defined categories of genes were described (Legendre et al., 2010).

While most of the genes encoded by tupanviruses have no known function, others are associated with different functional groups (Fig. 3A). These genes have different origins, as demonstrated by analysis of the rhizome of tupanviruses, in which 50% exhibit best hits with genes found


A
74
56
49
49
36
29
26
23
22
19
11
10
872
Miscellaneous
DNA replication, recombination and repair
Transcription
Protein metabolism
Translation
Protein interaction
Signal transduction
Carbohydrate metabolism
Redox process
Nucleic acid transport and metabolism
Lipid metabolism
Structure
Function unknown


B
Tupanvirus Soda Lake (AUL77893.1)
Tupanvirus Deep Ocean (AUL79186.1)
Acanthamoeba polyphaga mimivirus (YP 003986825.1)
Acanthamoeba castellanii mamavirus (AEQ60511.1)
Mimivirus battle86 (CRI62815.1)
Hirudovirus strain Sangsue (AHA45542.1)
Megavirus chiliensis (YP 004894633.1)
Megavirus Iba (AGD92513.1)
Megavirus courdo7 (AEX61758.1)
Megavirus courdo11 (AFM52349.1)
Moumouvirus goulette (AGF85231.1)
Moumouvirus australiensis (AVL94893.1)
Moumouvirus Monve (AEX62677.1)
Saudi moumouvirus (CRI62819.1)
Amoebae-infecting mimiviruses
Bodo saltans virus (ATZ80752.1)
Hokovirus HKV1 (ARF11096.1)
Catovirus CTV1 (ARF09278.1)
Klosneuvirus KNV1 (ARF11832.1)
Indivirus ILV1 (ARF09771.1)
Klosneuviruses
Cafeteria roenbergensis virus BV-PW1 (YP 003970183.1)
Phaeocystis globosa virus (YP 008052566.1)
Organic lake phycodnavirus 1 (ADX06143.1)
Organic lake phycodnavirus 2 (ADX06483.1)
Algae-infecting mimiviruses
Phycodnaviridae
Mollivirus sibericum (YP 009165284.1)
Pandoravirus inopinatum (YP 009120445.1)
Pandoravirus salinus (YP 009429988.1)
Pandoravirus dulcis (YP 008318996.2)
Iridoviridae/Ascoviridae
Marseilleviridae
Orpheovirus (YP 009448764.1)
"Pithoviridae"
Poxviridae
Afarviridae
0.5

in amoebae-infecting mimiviruses and klosneuviruses (Abrahão et al., 2018; Schulz et al., 2017). Many nucleocytoplasmic viruses orthologous group of genes (NCVOGs) are found in tupanviruses [e.g., Major Capsid Protein (NCVOG0022) and DNA polymerase B family (NCVOG0038)], and phylogenetic analyses based on those genes put tupanviruses as a sister-group of amoebae-infecting mimiviruses among different groups of giant viruses (Fig. 3B). In addition, pan-genome analysis comprising the family *Mimiviridae* showed the presence of many genes in common with these viruses, although a total of 775 tupanvirus genes are absent in other genomes of known mimiviruses (Abrahão et al., 2018). Most of these core genes are usually found in the central region of the genome of the giant viruses. Interestingly, genome synteny analysis comprising different amoebae-infecting mimiviruses (tupanvirus included) revealed that a central part of the genomes is more highly conserved, while the ends seemed to be more susceptible to variation (Fig. 4). These ends likely contain genes related to viral fitness in a competition scenario with other microorganisms, as suggested by the loss of genes in these regions observed when APMV was cultured in allopatric conditions for over a year (Boyer et al., 2011).

The tupanviruses have many genes related to the metabolism of nucleic acids, some of them involved in the process of replication and DNA repair, such as DNA polymerase, ATP-dependent DNA ligase, DNA helicase, DNA topoisomerase, and DNA repair proteins (including MutS mismatch repair and uracil-DNA glycosylase). This suggests a relative independence from the hosts, regarding the metabolic pathways involving nucleic acids, and was also observed for other large DNA viruses (Abrahão et al., 2018; Boyer et al., 2009; Raoult et al., 2004). In addition, the tupanviruses encode several nucleases, including homing endonucleases, which compose the genomic mobilome, described for giant viruses along with provirophages and transpovirons, although these other

Fig. 3. Gene content and phylogenetic position of tupanvirus. (A) Functional assignment of TPV-SL genome content based on BLASTp analysis. Numbers indicate the amount of genes from each functional class, which are designated by different colors indicated below the pie-chart; (B) phylogenetic tree based on DNA polymerase B amino acid sequences of nucleocytoplasmic large DNA viruses (NCLDVs). The tree was constructed using MEGA version 6.0, applying the maximum likelihood method and the Jones-Taylor-Thornton (JTT) model of evolution with 1000 bootstrap replicates (Tamura et al., 2013). The tupanviruses are clustered along with other amoebae-infecting mimiviruses, highlighted in red. The scale bar indicates the rate of evolution.

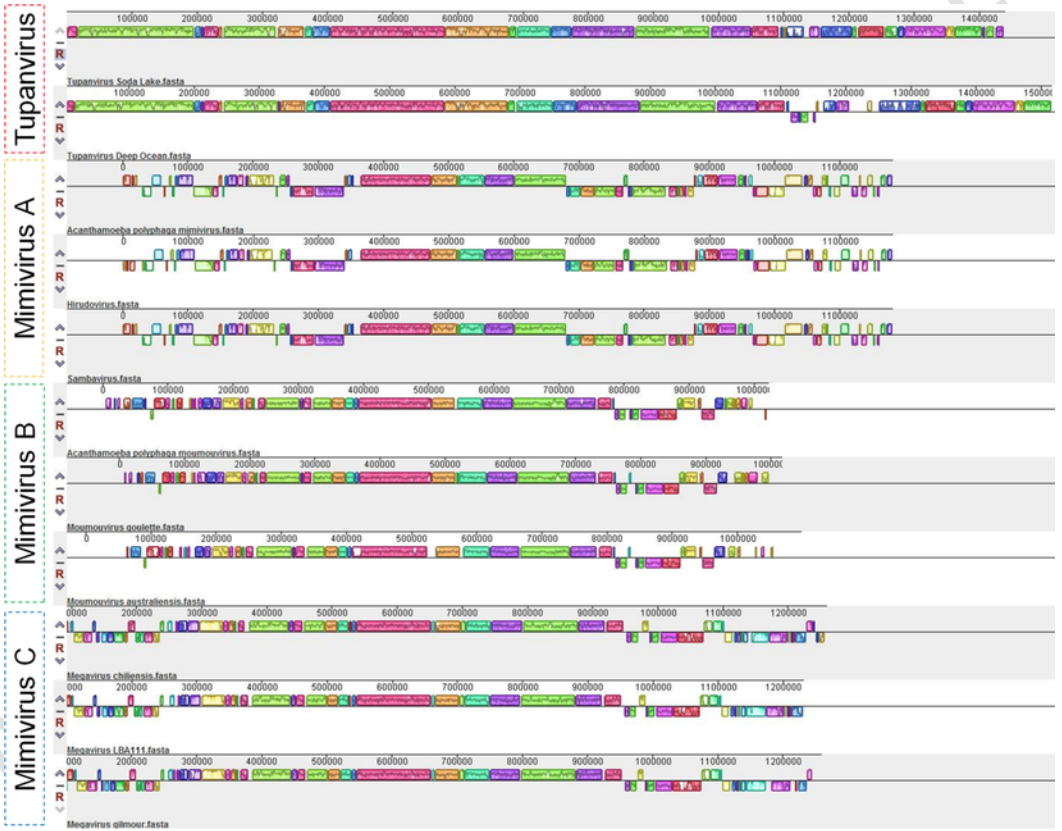

Tupanvirus
Mimivirus A
Mimivirus B
Mimivirus C
Tupanvirus Soda Lake.fasta
Tupanvirus Deep Ocean.fasta
Acanthamoeba polyphaga mimivirus.fasta
Hirudovirus.fasta
Sambavirus.fasta
Acanthamoeba polyphaga moumouvirus.fasta
Moumouvirus goulette.fasta
Moumouvirus australiensis.fasta
Megavirus chiliensis.fasta
Megavirus LBA111.fasta
Megavirus gilmour.fasta

Fig. 4. Mimivirus (including tupanviruses) genome synteny analysis. Schematic genome alignment diagram obtained using the Mauve software package. The analysis was performed using the genome of different amoebae-infecting mimiviruses, named Tupanvirus soda lake (KY523104.1); Tupanvirus deep ocean (MF405918.1); *Acanthamoeba polyphaga* mimivirus (NC 014649.1); *Acanthamoeba castellanii* mamavirus (JF801956.1); Sambavirus (KF959826.2); *Acanthamoeba polyphaga* moumouvirus (NC 020104.1); Moumouvirus goulette (KC008572.1); Moumouvirus australiensis (MG807320.1); Megavirus chiliensis (NC 016072.1); Megavirus LBA111 (JX885207.1); and Megavirus gilmour (MG602507). The blocks illustrated above the *x* axis are in the positive strand (forward sense), while blocks below the *x* axis are in the negative strand (reverse sense).

elements have not been associated with the genome of tupanviruses (Desnues et al., 2012). Furthermore, they have genes involved in the transcription process, including at least five subunits of DNA-dependent RNA polymerase (subunits 1, 2, 5, 6, and 9) and enzymes important for the processing of mRNA, such as mRNA capping enzyme and poly-A polymerase. The transcriptional machinery is also accompanied by the presence of transcription factors involved in the initiation (TFIIB) and elongation (TFIIS), in addition to VLTF3, a gene involved in late transcription present in many other giant viruses. Taking this into account, the synthesis of mRNA in tupanviruses is similar to that observed for APMV, wherein the presence of the "AAAATTGA" promoter motif, along with transcription factors and polymerases, is responsible for the production of viral transcripts in the host cytoplasm, with the process being terminated by hairpin-like structures in the 3′-end region that act as polyadenylation signals (Byrne et al., 2009). Finally, the transcripts are likely translated into proteins with a remarkable contribution of the translation-related elements encoded by tupanviruses, which present an unparalleled gene arsenal in the virosphere related to the process of protein synthesis, along with cell components (see next topic).

Different from other viruses, the members of the family *Mimiviridae* have their own glycosylation machinery capable of the synthesis of complex sugars, including viosamine, a rare glycan usually found only in a few bacterial species (Piacente et al., 2012, 2014, 2015). The tupanviruses also encode proteins related to the synthesis of carbohydrates such as UDP-glucose 4-epimerase and UDP-*N*-acetylglucosamine pyrophosphorylase, in addition to many glycosyltransferases, which can be involved in different post-translational modifications. Moreover, genes encoding proteins involved in the metabolism of lipids and proteins, such as phospholipases, lipases, and ubiquitins, as well as several genes related to signal transduction and host-interaction, especially serine/threonine kinases and ankyrin repeat-containing proteins have been detected. There are still few studies about giant virus-host interactions, especially at the molecular level, but it is possible that these viruses exhibit an important modulation of the host machinery due to their genomic complexity, as was observed for other large DNA viruses, such as poxviruses, which have a broad gene arsenal to counter the antiviral defense of their hosts (Alzhanova and Fruh, 2010; Bidgood and Mercer, 2015).

Genes encoding proteins related to viral structure were also identified in tupanviruses, such as major core protein and major capsid protein (MCP), which are usually found in NCLDVs, the latter considered a hallmark gene among giant viruses (Yutin et al., 2009). Among the genes encoding proteins related to the fibrils of mimiviruses (Fibril Associated Proteins—FAPs) (Sobhy et al., 2015), only the genes L725 and L829 of APMV have orthologs in the genome of tupanviruses (coverage > 85%, identity > 50%). Curiously, homologous genes to the R135 gene of APMV are not found in tupanviruses, which is implicated as the main component of the mimiviruses' fibrils and displays an important function in viral entry (Klose et al., 2015). It is possible that the tupanvirus fibrils have a distinct constitution compared to mimiviruses, which would explain the morphological difference.

Structural proteins were detected in purified viral particles of TPV-SL by proteomic approaches with a total of 127 proteins detected within the viral particle (Abrahão et al., 2018). Over half of them have no predicted function (67/127 = 52.8%), including 11 ORFans. In addition to structural components, genes involved in different metabolic processes have been detected in the TPV-SL particles, including DNA polymerase, triacylglycerol lipase, and protein phosphatases, among others. Even more impressive was the presence of several components of the transcription machinery (a total of 13 genes, including RNA polymerase), indicating that the synthesis of viral mRNA has a very early onset and is independent of the cell machinery, unlike what has been observed for members of the family *Marseilleviridae* (Fabre et al., 2017). The large amount and variety of proteins constituting the viral particle are not uncommon among the giant viruses (Fabre et al., 2017; Fischer et al., 2014; Fridmann-Sirkis et al., 2016; Legendre et al., 2015). This reflects their huge genomic complexity.

6. TRANSLATION MACHINERY OF TUPANVIRUSES—AN UNEXPECTED COMPLEXITY

Protein synthesis and metabolism are key processes of life as we know it, and were the features historically used to separate the living from the inanimate world or, more specifically, cellular organisms from viruses (Lwoff, 1957; Nierhaus and Wilson, 2006). Protein translation is a complex process that requires several elements: ribosomal RNA and

proteins, responsible for the formation of ribosomes; transfer RNAs (tRNAs) responsible for adding an amino acid to the cognate sequence in the mRNA molecule (Giegé, 2006); aminoacyl-tRNA synthetases (aaRS), which perform the esterification reaction between an amino acid and the 3′-end of its cognate tRNA (Ribas de Pouplana and Schimmel, 2001); and many other proteins, collectively named translation factors, which assist in the whole process of the synthesis of a polypeptide chain. These elements have been conceived as a barrier between the cellular and viral worlds, since they are abundantly found in organisms belonging to the three domains of life (Eukarya, Bacteria, Archaea) and are absent in viruses, with the exception of a few tRNAs found in some groups of viruses, such as phycodnavirus, herpesvirus, and myovirus (Miller et al., 2003; Van Etten and Meints, 1999; Virgin et al., 1997). However, this scenario has been put challenged with the discovery of mimiviruses and other giant viruses of amoebae, since their genome codes for different components of the translational apparatus, thus promoting a hot debate about the origin and evolution of viruses (Abrahão et al., 2017; Claverie and Ogata, 2009; Filée, 2013; Marcelino et al., 2017; Moreira and López-garcía, 2009; Nasir et al., 2012; Schulz et al., 2017).

Until very recently, Cafeteria roenbergensis virus had the largest quantity and diversity of viral proteins involved in translation, including an isoleucyl-tRNA synthetase and 22 tRNAs (Fischer et al., 2010). This number was largely exceeded by the discovery of the klosneuvirus, whose genome encodes 25 tRNAs and 19 aaRS, as well as some translation factors and tRNA modifying enzymes (Schulz et al., 2017). Now, the tupanviruses expand the repertoire of translational components found in the virosphere even more, exceeding even some cellular organisms (Abrahão et al., 2018). With a total of 67 (TPV-SL) and 70 (TPV-DO) genes encoding tRNAs, these viruses have tRNA with anticodons associated with 46 and 47 codons respectively, representing 21 amino acids, including pyrrolysine. The profile of tRNAs present in tupanviruses is very different from that observed for the other domains of life (Novoa et al., 2012), exhibiting some tRNAs found only in eukaryotes (Fig. 5A). It is not possible to unequivocally identify the origin of these viral genes, although they seem to have multiple origins. To define the evolutionary history of tRNAs is a difficult task due to the susceptibility to horizontal transfer (Novoa et al., 2012). In addition, the tupanviruses have 20 aaRS related to all amino acids present in the genetic code, thus allowing the

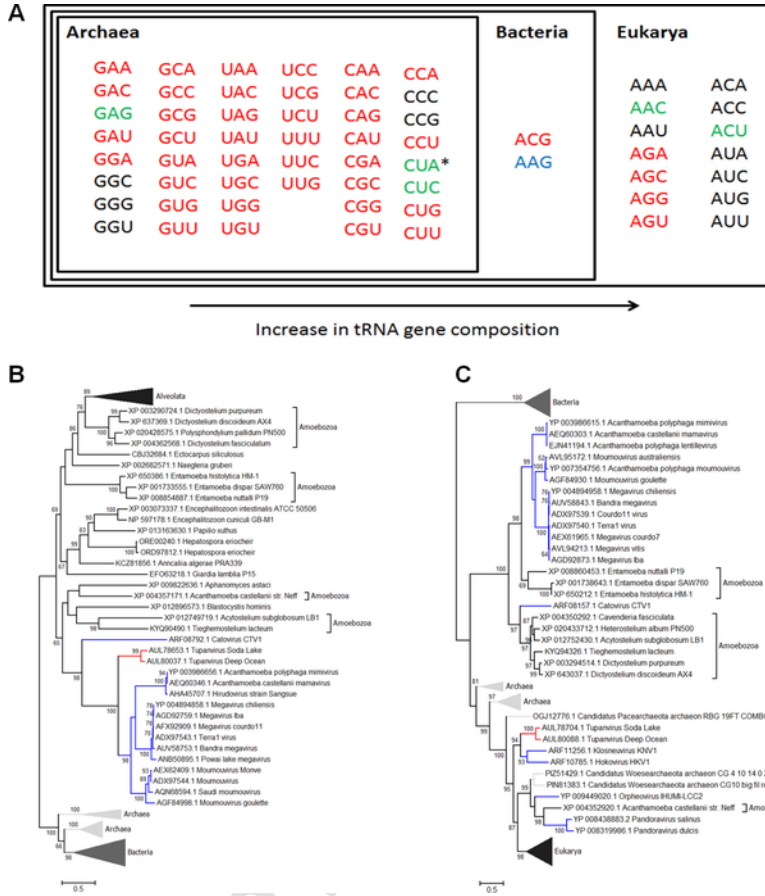


Fig. 5. Translational components of tupanviruses. (A) Diagram showing the increase in tRNA population complexity in the domains of life (data obtained from Novoa et al., 2012), showing a miscellaneous set of tRNA in tupanviruses. Each tRNA is designated by its anticodon sequence. The colored sequences stand for the tRNA coded for tupanviruses (red: both viruses; blue: only TPV-SL; green: only TPV-DO). The anticodon labeled with an asterisk (CUA) corresponds to a tRNA for pyrrolysine, a rare amino acid found only in some methanogenic archaea, bacteria, and TPV-DO. (B) Phylogenetic reconstruction based on amino acid sequences of cysteinyl-tRNA synthetase; (C) phylogenetic reconstruction based on amino acid sequences of tyrosinyl-tRNA synthetase. Both phylogenetic trees were built considering sequences from organisms of all domains of life and viruses. The alignments were performed by using Muscle software (Edgar, 2004). Phylogenetic reconstructions were performed by using FastTree software, with maximum likelihood method, Jones-Taylor-Thornton (JTT) model for amino acid substitution and 1000 bootstrap replicates. The trees were visualized with MEGA7 software (Kumar et al., 2016). In both trees, the tupanviruses are highlighted in red and other viruses in blue.

Bootstrap values lower than 50 were suppressed. The scale bars indicate the rate of evolution.

synthesis of all necessary tRNAs for the translation process. Similar to tRNAs, attempting to define the origin of these enzymes is not trivial. Since the discovery of APMV, an exciting debate about the origin and evolution of giant viruses has revolved around these genes, with two main hypotheses: (1) origin from a simpler organism and increase of complexity by gene gain events (Moreira and Brochier-Armanet, 2008; Yutin et al., 2014); (2) origin from a more complex organism that has lost genes throughout evolution, resulting in the current giant viruses (Arslan et al., 2011; Boyer et al., 2010). An attempt to set this debate based on aaRS genes might not be the best alternative, since using different sequence sampling approaches and distinct methods for phylogenetic reconstruction, canonical groups can be placed in different clades and viruses can form external groups to the cellular organisms or cluster with them, either giving the idea of an independent origin, or origin by lateral gene transfer events.

Initial phylogenetic analyses using the aaRS of tupanviruses as bait for retrieval of only the best hits and some amoebozoa organisms in public databases placed the viruses as external groups, thus suggesting an independent origin from most of aaRS genes in tupanviruses, related to cellular genes (Abrahão et al., 2018). New analyses using TPV-SL aaRS sequences against specific taxa in databases [Eukarya (taxid: 2759); Amoebozoa (taxid: 554915); Archaea (taxid: 2157); Firmicutes (taxid: 1239); Proteobacteria (taxid: 1224); Viruses (taxid: 10239)] and retrieving up to 30 random sequences for each taxa (when available, and 10 for amoebozoa), we observed an alternative topology of the trees, with viruses clustered within cellular branches, either in a monophyletic group or dispersed in the tree, suggestive of horizontal gene transfer events (Fig. 5B and C; Supplementary Fig. 1 in the online version at https://doi.org/10.1016/bs.aivir.2018.09.001). It is noteworthy that canonical taxonomic groups are frequently separated in the trees, independent of the gene used. This pattern has already been observed in phylogenetic reconstructions performed by different groups (Schulz et al., 2017). Methodological issues related to analyses using aaRS have been and still are debated. Therefore, in light of the available data and methods the origin of these genes remains a mystery.

The tupanviruses also encode many translation factors, including eight proteins involved in the initiation process [IF2α, IF2β, IF2γ, IF4a, IF4e (2 copies in TPV-SL), IF5a (2 copies in TPV-DO) and SUI1], an initiation/elongation factor (GTP-binding elongation/initiation), an elongation factor (EF-2), and a release factor (eRF1). Some of these factors are also present in other mimiviruses, mainly in the klosneuviruses, where a considerable diversity of these factors has been identified (Schulz et al., 2017). Moreover, the tupanviruses have additional genes related to the maturation and stabilization of tRNAs (tRNA nucleotidyltransferase, tRNA guanylyltransferase, cytidine deaminase, RNA methyl transferase) and modification of ribosomal proteins (ribosomal-protein-alanine *N*-acetyltransferase, FtsJ-like methyltransferase). With > 100 virus-encoded elements related to translation ranging from tRNAs to proteins, tupanviruses specify the most complex protein synthesis machinery ever observed in the virosphere. The purpose of such a vast gene pool related to protein translation observed for giant viruses is still unclear.

7. RIBOSOMAL GENE COMPONENTS IN TUPANVIRUSES: WHAT IS GOING ON?

After the discovery and genomic characterization of tupanviruses, new questions about the presence of other components that could be involved with the translation process were raised. In parallel with viral genome characterization, experiments were conducted in order to analyze the transcriptome of TPV-SL; however, due to the shutdown of ribosomal RNA, these analyses were thwarted by the generation of just a small quantity of viral reads (Abrahão et al., 2018). Curiously, an in-depth evaluation of these reads revealed the presence of two sequences in the genome of TPV-SL, named copy 1 and copy 2, which presented a high similarity with intronic regions of the 18S gene, more specifically with self-splicing group I introns of different organisms, such as fungi, bacteria, amoeba and algae (Abrahão et al., 2018). Surprisingly, the 18S rRNA intronic regions appear to be widespread in the three lineages of the family *Mimiviridae*, with lineages A and B presenting just one copy of the sequences, and lineage C and TPV-DO presenting two copies (Abrahão et al., 2018). The analyses involving the localization of these self-splicing group I intron sequences in the viral genomes showed that the single copies of lineages A and B and the copy 1 of lineage C were

located in an intronic region of the DNA-directed RNA polymerase subunit 1 gene, next to a group I intron endonuclease (Fig. 6A). In contrast, the copy 2 of lineage C and both copies of tupanviruses are located (ac-

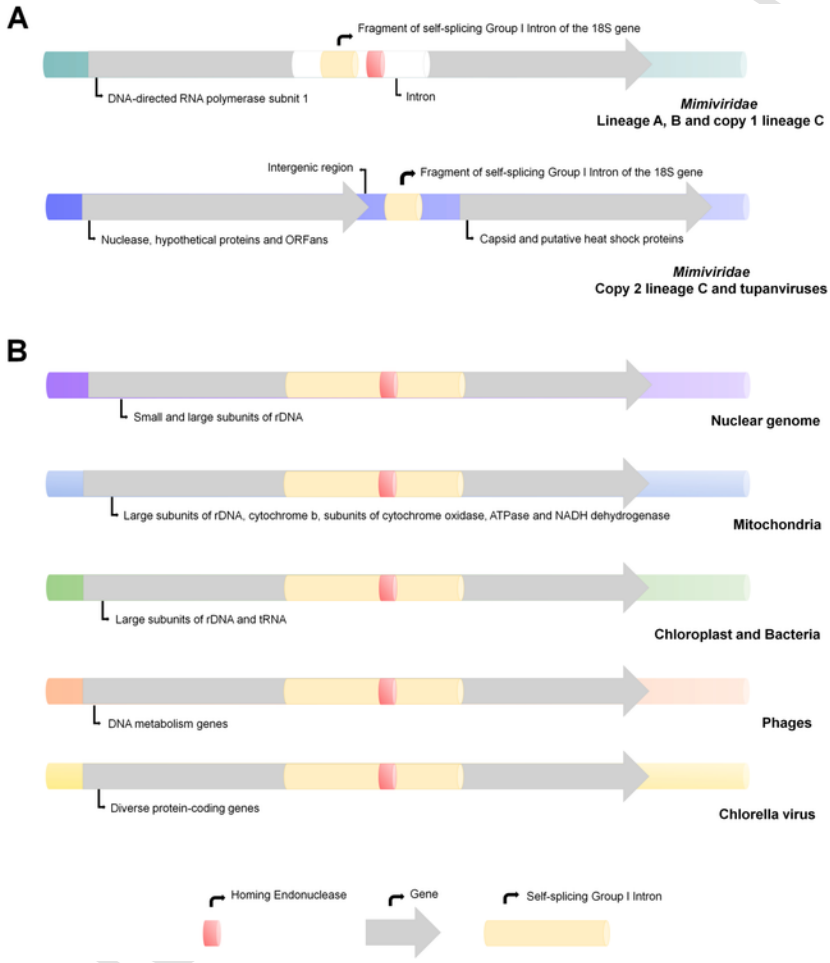


Fig. 6. Occurrence of self-splicing group I in different organisms. (A) Fragments of self-splicing group I introns of 18S genes in intronic or intergenic regions of the mimiviruses genome. (B) Self-splicing group I introns are widespread in a variety of genes in different organisms. Nuclear group I introns are exclusively located in both large and small rRNA genes of several eukaryotic organisms. Mitochondrial introns are located in large rRNA and in genes encoding components of the electron transport system of several eukaryotic organisms. Chloroplast and bacteria 18S introns are found in large rRNA and also in tRNA genes. Bacteriophage 18S introns are located in different genes involved mainly in DNA metabolism. Chlorella viral 18S introns were found in protein-coding genes, such as transcriptional elongation factor TFIIS, URF genes, and major capsid protein.

cording to prediction) in an intergenic region (Fig. 6A). Despite both copies of TPV-SL being located in intergenic regions, experiments involving fluorescence in situ hybridization (FISH) and qPCR were able to confirm the expression of these copies during the viral replicative cycle (Abrahão et al., 2018).

Phylogenetic analyses of these sequences present in the family *Mimiviridae* suggested that copies 1 and 2 of TPV-SL have separate origins and seem to be related to the single copies of lineage A and B, to copy 1 of lineage C and to the fungi mitochondrial 18S rRNA intronic region. Furthermore, copy 2 of lineage C is more related to the fungi 18S rRNA intronic region and to 18S-like sequences present in some Chlorella viruses, which were all phylogenetically related (Abrahão et al., 2018). The presence of more than one self-splicing group I intron in the genome of viruses belonging to the family *Phycodnaviridae* has been previously reported and these sequences also exhibit a phylogenetic relationship with introns found in the rRNA genes of a huge variety of organisms, such as algae, yeasts, fungi and protozoa (Nishida et al., 1998; Yamada et al., 1994). The occurrence of this group of viral introns does not seem to be limited to the NCLDVs, since an abundance of self-splicing introns in the genome of bacteriophages has already been described, showing their presence in multiple genes and also in single genes with many introns (Bonocora and Shub, 2004; Landthaler and Shub, 1999; Sandegren and Sjöberg, 2007).

Group I introns are a distinct class of self-splicing ribozymes, found widespread in a variety of organisms, genes, and genomes throughout all the domains of life, and occur less frequently in viruses compared to cellular organisms (Haugen et al., 2005; Hedberg and Johansen, 2013). In eukaryotes, the nuclear group I introns are exclusively located in both large and small rRNA genes, while mitochondrial introns are located not only in large rRNA but also in genes encoding components of the electron transport system (Fig. 6B). In contrast, chloroplast introns are found in large rRNA and also in tRNA genes (Fig. 6B) (Lambowitz and Belfort, 1993; Yamada et al., 1994). Bacteriophage introns are located in different genes involved mainly in DNA metabolism, whereas bacterial introns are found primarily in tRNA genes, although some studies have also reported the presence of them in 23S rRNA genes (Fig. 6B) (Lambowitz and Belfort, 1993; Nesbo and Doolittle, 2003; Raghavan et al., 2008; Yamada et al., 1994). Similar to bacteriophage introns,

Chlorella virus introns are found in protein-coding genes, such as transcriptional elongation factor TFIIS, unassigned reading frame (URF) genes and major capsid protein (Fig. 6B) (Nishida et al., 1998; Yamada et al., 1994). Curiously, an intergenic position has not been previously described for this group of self-splicing introns in any other organism, other than tupanviruses and members of lineage C of the family *Mimiviridae* (Abrahão et al., 2018). Many group I introns contain ORFs encoding endonucleases that could have mobility-promoting activities enabling these DNA elements to move within and between genomes, suggesting that the presence of a group I intron endonuclease next to the 18S-like intronic copies of some mimiviruses could be involved with the movement and genesis of those introns in the viral genome (Edgell et al., 2011; Haugen et al., 2005; Hedberg and Johansen, 2013; Lambowitz and Belfort, 1993). Moreover, endonucleases could promote intron mobility using a transposition mechanism known as "homing," introducing double-strand breaks (DSBs) and initiating intron mobility via a DSB-repair process (Edgell et al., 2011). Furthermore, homing endonucleases are site-specific but also sequence-tolerant to DNA, often supporting multiple sequence changes within their recognition site, thus allowing homing into new sites (Belfort and Bonocora, 2014).

Several lines of evidence also suggest that many of these mobile elements have undergone lateral transfer events into different species and genes, resulting in their sporadic and highly biased distribution (Belfort and Bonocora, 2014; Biniszkiewicz et al., 1994; Haugen et al., 2005; Lambowitz and Belfort, 1993). Previously groups have already reported the presence of introns in intracellular bacteria that were putatively acquired from their eukaryotic hosts (Nesbo and Doolittle, 2003). Furthermore, viruses are considered as possible vectors of horizontal transfer. Sequence analyses of phage T4 introns, for example, showed that these elements shared ancestry with bacterial introns, leading some authors to speculate that phages with a broad host range might be vectors for shuttling introns between various organisms (Bernstein and Bernstein, 1989; Haugen et al., 2005; Lambowitz and Belfort, 1993).

Taking into account these points, and based on the model of intron evolution of Goddard and Burt (1999) that involved gain and loss, we propose an evolutionary scenario for the origin and distribution of the 18S-like intronic sequences in the *Mimiviridae* (Goddard and Burt, 1999; Haugen et al., 2005). As previously described, amoebae are considered a

"melting pot" for microbial genetic diversity, facilitating genetic exchanges between bacteria, yeasts and viruses that cohabit in the same host (Boyer et al., 2009; Wang and Wu, 2017). In this intracellular environment we hypothesize that group I self-splicing introns located in genes of ribosomal subunits of amoeba, bacteria, and/or fungi, and containing an ORF encoding an endonuclease, could act as an intron donor (Fig. 7A). Further, the expressed endonuclease could recognize a homing site in the mimivirus genome and the donor intron could be introduced by a DSB repair mechanism (Fig. 7B). Once the intron becomes fixed in the viral population, the endonuclease has no longer any active biological function, and because of that it could accumulate mutations and eventually be inactivated or lost, as observed for tupanviruses, copy 2 of lineage C and also in many Chlorella viruses (Fig. 7C,D). Unable to spread, the intron could also be destined to be lost or degenerate overtime (Fig. 7D and E) (Haugen et al., 2005). Since more than one copy is observed in the genome of mimiviruses and phycodnaviruses, we also hypothesize that the acquisition of intron donors happened independently, as shown by phylogenetic analyses, and more than once during the evolution of these viruses.

Although self-splicing introns are commonly characterized as selfish parasites, some authors indicated that the dynamics of these elements could be involved with environmental stressors like oxidizing agents, reactive oxygen and nitrogen species, starvation, temperature, osmolarity, and DNA damage (Belfort, 2017; Edgell et al., 2011). However, the real functions of these 18S rRNA intronic region sequences in tupanviruses and other mimiviruses require further clarification. Furthermore, no exonic region of 18S rRNA was found in the genomes of tupanviruses or any previously described mimivirus, although an apparent colocalization between these 18S rRNA intronic region sequences and *Acanthamoeba* ribosomal RNA was observed by FISH, raising questions about the possibility of interaction between host ribosomes and viral copies of 18S intron-like sequences (Abrahão et al., 2018).

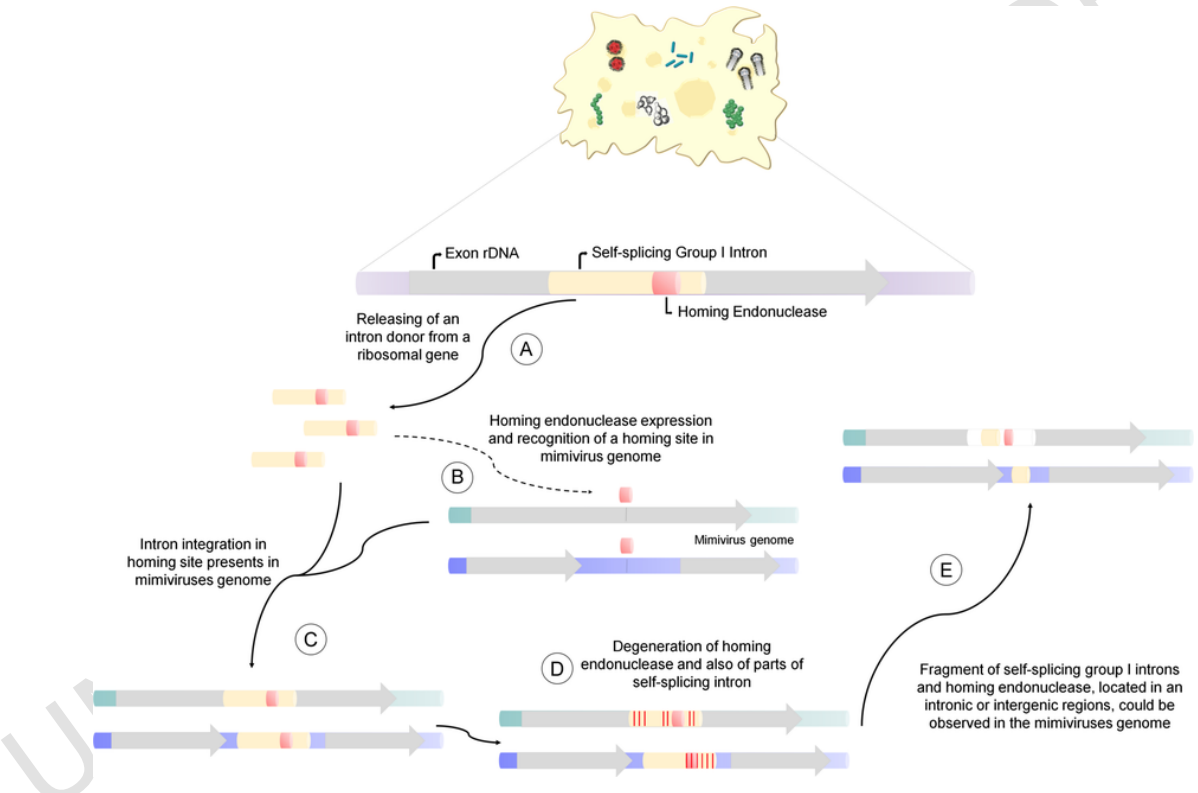

Exon rDNA
Self-splicing Group I Intron
Homing Endonuclease
Releasing of an intron donor from a ribosomal gene
A
Homing endonuclease expression and recognition of a homing site in mimivirus genome
B
Mimivirus genome
Intron integration in homing site presents in mimiviruses genome
C
D
Degeneration of homing endonuclease and also of parts of self-splicing intron
E
Fragment of self-splicing group I introns and homing endonuclease, located in an intronic or intergenic regions, could be observed in the mimiviruses genome

Fig. 7. Possible evolutionary scenario concerning the rise and distribution of the 18S-like intronic sequences in some members belonging to the family *Mimiviridae.* Giant viruses infect amoebae that are considered a genetic "melting pot." In this intracellular environmental, group I self-splicing introns located in genes of ribosomal subunits of amoeba, bacteria, and/or fungi, and containing an ORF encoding a homing endonuclease, act as intron donors (A). Expressed endonuclease recognizes a homing site in mimivirus genome (B) and the donor intron is introduced in mimivirus genome (C). Homing endonuclease accumulates mutations being inactivated or lost (D). Intron is also destined to be lost or degenerated overtime and fragments of self-splicing group I introns and homing endonuclease, located in intronic or intergenic regions, are observed in the mimivirus genome (E).

8. BONA FIDE VIRUSES: TUPANVIRUSES ARE STILL DEPENDENT ON TRANSLATIONAL AND ENERGY-PRODUCING APPARATUS OF THE HOST

Giant viruses have been breaking some criteria traditionally used to define the viral nature of an organism, e.g., the fact that they are not retained in 0.22 μm porous filters and they exhibit both RNA and DNA within their virions (Lwoff, 1957). Moreover, the presence of many genes related to protein synthesis has suggested that those viruses have a relative independence from the hosts, being "quasi-autonomous" viruses (Claverie and Abergel, 2010). In this way, the tupanviruses would be the height of complexity in the virosphere described so far. However, despite exhibiting several components of the translational apparatus, the tupanviruses have no ribosomal exonic RNA/proteins, making them dependent on the translational machinery of their hosts to synthesize their proteins, which (together with other characteristics) classified them as bona fide viruses. The ribosomal proteins are pivotal in the protein synthesis process, since they are key components in the ribosome structure, without which the translation of the mRNA does not occur (Chang et al., 2015).

Another key point that places giant viruses into the viral domain is the lack of genes related to the metabolic pathways of ATP production. Their absence makes the viruses dependent on their hosts to generate the energy needed for their development. Recently, a new algae-infecting mimivirus was isolated in Hawaii (USA), which displays two genes involved in the fermentation process (anaerobic strategy for energy production), pyruvate formate-lyase and pyruvate formase-lyase activating enzyme, never detected before in a virus (Schvarcz and Steward, 2018). These genes seem to have originated by means of horizontal gene transfer from its host. It is speculated that this virus can use those proteins to ensure energy production in conditions of low oxygen concentration that would maintain the fermentation pathway of the host fully active, although such a hypothesis has yet to be tested (Schvarcz and Steward, 2018).

Surprisingly, the tupanviruses have a gene coding for citrate synthase, an enzyme involved in the citric acid cycle, the central part of the metabolic pathway for ATP production. The Krebs cycle is a complex process

involving a series of organic acids and enzymes; the citrate synthase is responsible for the first step of the process, converting acetyl-CoA and oxaloacetate to citrate (Fig. 8A). There are no homologs of this gene in any other known virus. Phylogenetic analyses using the 100 best hits strategy suggested an independent origin of this gene in tupanviruses, but when sequences of different groups of organisms were considered, it is likely that tupanvirus acquired this gene by horizontal gene transfer from sympatric bacteria (Fig. 8B,C). Similar to what was observed for aaRS genes, the use of different strategies to reconstruct phylogeny provides contradictory results, and therefore it is difficult to affirm the most probable origin of such genes. Furthermore, canonical taxonomic groups are separated in these analyses. The specific impact of this gene on the life cycle of tupanviruses is still uncertain, but it is undoubtedly intriguing that these viruses have genes related to energy metabolism. Nevertheless, this enzyme constitutes only one piece of the metabolic pathway, leaving the tupanviruses still dependent on their hosts for ATP synthesis.

9. FUTURE PERSPECTIVES

The study of giant viruses raised questions regarding several paradigms of classic virology including the origin and evolution of viruses, revealing an exceptional diversity of viruses, unimaginable until very recently. The discovery of tupanviruses represents a breakthrough in virology, bringing new and intriguing challenges. The majority of viral genes and proteins (~ 70%) have no assigned function, many are completely novel (ORFans), and their study and characterization could bring valuable scientific discoveries. The complex structure of tupanviruses is equally intriguing. The hitherto unsuccessful separation of the tail from the capsid for a better characterization of each virion component is a current challenge. Moreover, the genomic and structural complexity of these viruses provides the basis for biotechnological approaches applying genetic engineering that may generate new tools for different uses in basic and applied sciences. Considering the continuous efforts in the search for and characterization of viruses from unexplored environments, our quest for unraveling the virosphere is bound to yield more exciting surprises impacting on virology and other scientific disciplines.

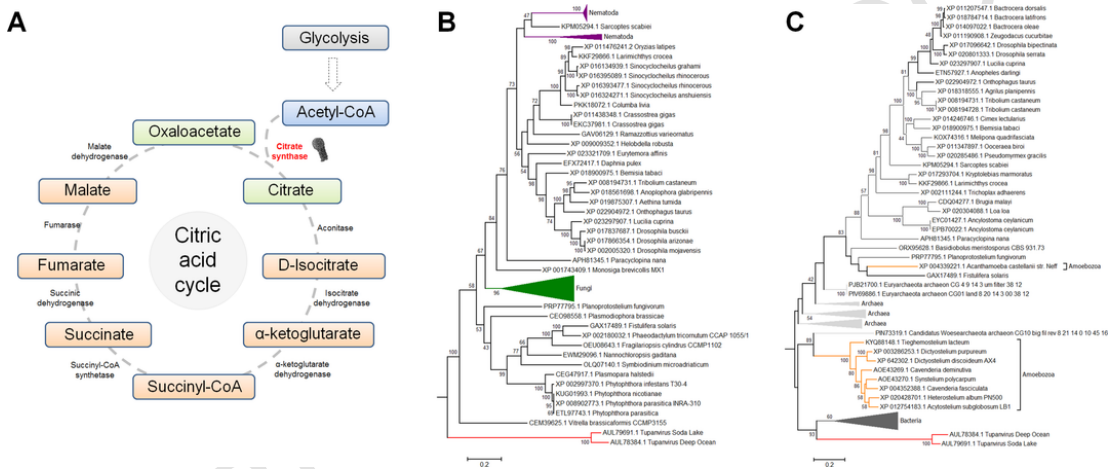


Fig. 8. Citrate synthase in tupanviruses. (A) Schematic representation of the citric acid cycle, showing the first step of the cycle (in green) and the required enzyme citrate synthase (in red), which is found in the genomes of tupanviruses; (B) phylogenetic reconstruction based on amino acid sequences of citrate synthase using the 100 best hits of TPV-SL sequence against NCBI nr database; (C) phylogenetic reconstruction based on amino acid sequences of citrate synthase using sequences from organisms of all domains of life and viruses. The alignments were performed with Muscle software (Edgar, 2004). Phylogenetic reconstructions were performed with MEGA6 software, with maximum likelihood method, Jones-Taylor-Thornton (JTT) model for amino acid substitution, and 1000 bootstrap replicates (Tamura et al., 2013). Bootstrap values lower than 40 were suppressed. The scale bars indicate the rate of evolution.

ACKNOWLEDGMENTS

We thank colleagues from Laboratorio de Virus and from IHU-Aix Marseille University. We also thank CAPES, CNPq, FAPEMIG for scholarships. J.S.A. is a CNPq researcher.

REFERENCES

Abrahão, J.S., 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.

Abrahão, J., Silva, L., Silva, L.S., Khalil, J.Y.B., Rodrigues, R., Arantes, T., Assis, F., Boratto, P., Andrade, M., Kroon, E.G., Ribeiro, B., Bergier, I., Seligmann, H., Ghigo, E., Colson, P., Levasseur, A., Kroemer, G., Raoult, D., La Scola, B., 2018. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat. Commun. 9, 749.

Aherfi, S., La Scola, B., Pagnier, I., Raoult, D., Colson, P., 2014. The expanding family Marseilleviridae. Virology 466–467, 27–37.

Aherfi, S., Colson, P., La Scola, B., Raoult, D., 2016. Giant viruses of amoebas: an update. Front. Microbiol. 7, 1–14.

Alzhanova, D., Fruh, K., 2010. Modulation of the host immune response by cowpox virus. Microbes Infect. 12, 900–909.

Andrade, K.R., Boratto, P.P.V.M., Rodrigues, F.P., Silva, L.C.F., Dornas, F.P., Pilotto, M.R., La Scola, B., Almeida, G.M.F., Kroon, E.G., Abrah??o, J.S., 2014. Oysters as hot spots for mimivirus isolation. Arch. Virol. 160, 477–482.

Andrade, A.C.d.S.P., Rodrigues, R.A.L., Oliveira, G.P., Andrade, K.R., Bonjardim, C.A., La Scola, B., Kroon, E.G., Abrahão, J.S., 2017. Filling gaps about mimivirus entry, uncoating and morphogenesis. J. Virol. 91, e01335–17.

Andrade, A.C.D.S.P., Arantes, T.S., Rodrigues, R.A.L., Machado, T.B., Dornas, F.P., Landell, M.F., Furst, C., Borges, L.G.A., Dutra, L.A.L., Almeida, G., Trindade, G.D.S., Bergier, I., Abrahão, W., Borges, I.A., Cortines, J.R., De Oliveira, D.B., Kroon, E.G., Abrahão, J.S., 2018. Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. Virol. J. 15, 1–10.

Andreani, J., Aherfi, S., Khalil, J.Y.B., Di Pinto, F., Bitam, I., Raoult, D., Colson, P., La Scola, B., 2016. Cedratvirus, a double-cork structured giant virus, is a distant relative of pithoviruses. Viruses 8, 1–11.

Andreani, J., Khalil, J.Y.B., Sevvana, M., Benamar, S., Di Pinto, F., Bitam, I., Colson, P., Klose, T., Rossmann, M.G., Raoult, D., La Scola, B., 2017. Pacmanvirus, a new giant icosahedral virus at the crossroads between asfarviridae and faustoviruses. J. Virol. 91, e00212–e00217.

Andreani, J., Khalil, J.Y.B., Baptiste, E., Hasni, I., Michelle, C., Raoult, D., Levasseur, A., La Scola, B., 2018. Orpheovirus IHUMI-LCC2: a new virus among the giant viruses. Front. Microbiol. 8, 1–11.

Antony, C.P., Kumaresan, D., Hunger, S., Drake, H.L., Murrell, J.C., Shouche, Y.S., 2013. Microbiology of Lonar Lake and other soda lakes. ISME J. 7, 468–476.

Arantes, T.S., Rodrigues, R.A.L., dos Santos Silva, L.K., Oliveira, G.P., de Souza, H.L., Khalil, J.Y.B., de Oliveira, D.B., Torres, A.A., da Silva, L.L., Colson, P., Kroon, E.G., da Fonseca, F.G., Bonjardim, C.A., La Scola, B., Abrahão, J.S., 2016. The

large marseillevirus explores different entry pathways by forming giant infectious vesicles. J. Virol. 90, 5246–5255.

Arslan, D., Legendre, M., Seltzer, V., Abergel, C., Claverie, J.-M., 2011. Distant mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc. Natl. Acad. Sci. U.S.A. 108, 17486–17491.

Assis, F.L., Franco-Luiz, A.P.M., dos Santos, R.N., Campos, F.S., Dornas, F.P., Borato, P.V.M., Franco, A.C., Abrahao, J.S., Colson, P., Scola, B.L., 2017. Genome characterization of the first mimiviruses of lineage C isolated in Brazil. Front. Microbiol. 8, 1–11.

Bajrai, L.H., Benamar, S., Azhar, E.I., Robert, C., Levasseur, A., Raoult, D., La Scola, B., 2016. Kaumoebavirus, a new virus that clusters with Faustoviruses and Asfarviridae. Viruses 8, 278.

Belfort, M., 2017. Mobile self-splicing introns and inteins as environmental sensors. Curr. Opin. Microbiol. 38, 51–58.

Belfort, M., Bonocora, R.P., 2014. Homing endonucleases: from genetic anomalies to programmable genomic clippers. Methods Mol. Biol. 1123, 1–26.

Bernstein, H., Bernstein, C., 1989. Bacteriophage T4 genetic homologies with bacteria and eucaryotes. J. Bacteriol. 171, 2265–2270.

Bidgood, S.R., Mercer, J., 2015. Cloak and dagger: alternative immune evasion and modulation strategies of poxviruses. Viruses 7, 4800–4825.

Biniszkiewicz, D., Cesnaviciene, E., Shub, D.A., 1994. Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. EMBO J. 13, 4629–4635.

Bonocora, R.P., Shub, D.A., 2004. A self-splicing group I intron in DNA polymerase genes of T7-like bacteriophages. J. Bacteriol. 186, 8153–8155.

Boratto, P.V.M., Arantes, T.S., Silva, L.C.F., Assis, F.L., Kroon, E.G., La Scola, B., Abrahão, J.S., 2015. Niemeyer virus: a new mimivirus group A isolate harboring a set of duplicated aminoacyl-tRNA synthetase genes. Front. Microbiol. 6, 1–11.

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., Robert, C., Azza, S., Sun, S., Rossmann, M.G., Suzan-Monti, M., La Scola, B., Koonin, E.V., 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.

Boyer, M., Madoui, M.A., Gimenez, G., La Scola, B., Raoult, D., 2010. Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4th domain of life including giant viruses. PLoS One 5, e15530.

Boyer, M., Azza, S., Barrassi, L., Klose, T., Campocasso, A., Pagnier, I., Fournous, G., Borg, A., Robert, C., Zhang, X., Desnues, C., Henrissat, B., Rossmann, M.G., La Scola, B., Raoult, D., 2011. Mimivirus shows dramatic genome reduction after intraamoebal culture. Proc. Natl. Acad. Sci. U.S.A. 108, 10296–10301.

Byrne, D., Grzela, R., Lartigue, A., Audic, S., Chenivesse, S., Encinas, S., Claverie, J.M., Abergel, C., 2009. The polyadenylation site of Mimivirus transcripts obeys a stringent "hairpin rule". Genome Res. 19, 1233–1242.

Chang, K.C., Wen, J.D., Yang, L.W., 2015. Functional importance of mobile ribosomal proteins. Biomed. Res. Int. 2015, 539238.

Choi, Y., Bowman, J.W., Jung, J.U., 2018. Autophagy during viral infection—a double-edged sword. Nat. Rev. Microbiol. 16, 341–354.

Claverie, J.M., Abergel, C., 2010. Mimivirus: the emerging paradox of quasi-autonomous viruses. Trends Genet. 26, 431–437.

Claverie, J.-M., Ogata, H., 2009. Ten good reasons not to exclude giruses from the evolutionary picture. Nat. Rev. Microbiol. 7, 615, author reply 615.

Dann, L.M., Rosales, S., McKerral, J., Paterson, J.S., Smith, R.J., Jeffries, T.C., Oliver, R.L., Mitchell, J.G., 2016. Marine and giant viruses as indicators of a marine microbial community in a riverine system. Microbiologyopen 5, 1071–1084.

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

Domagalski, J.L., Orem, W.H., Eugster, H.P., 1989. Organic geochemistry and brine composition in Great Salt, Mono, and Walker Lakes. Geochim. Cosmochim. Acta 53, 2857–2872.

Dornas, F.P., Khalil, J.Y.B., Pagnier, I., Raoult, D., Abrahão, J., La Scola, B., 2015. Isolation of new Brazilian giant viruses from environmental samples using a panel of protozoa. Front. Microbiol. 6, 1–9.

Dornas, F.P., Assis, F.L., Aherfi, S., Arantes, T., Abrahão, J.S., Colson, P., La Scola, B., 2016. A Brazilian marseillevirus is the founding member of a lineage in family Marseilleviridae. Viruses 8, 76.

Dos Santos, R.N., Campos, F.S., Medeiros De Albuquerque, N.R., Finoketti, F., Côrrea, R.A., Cano-Ortiz, L., Assis, F.L., Arantes, T.S., Roehe, P.M., Franco, A.C., 2016. A new marseillevirus isolated in Southern Brazil from Limnoperna fortunei. Sci. Rep. 6, 35237.

dos Santos Silva, L.K., Arantes, T.S., Andrade, K.R., Lima Rodrigues, R.A., Miranda Boratto, P.V., de Freitas Almeida, G.M., Kroon, E.G., Scola, B.L., Clemente, W.T., Santos Abrahão, J., 2015. High positivity of mimivirus in inanimate surfaces of a hospital respiratory-isolation facility, Brazil. J. Clin. Virol. 66, 62–65.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Edgell, D.R., Chalamcharla, V.R., Belfort, M., 2011. Learning to live together: mutualism between self-splicing introns and their hosts. BMC Biol. 9, 22.

Fabre, E., Jeudy, S., Legendre, M., Trauchessec, M., Claverie, J., Abergel, C., 2017. Noumeavirus replication relies on a transient remote control of the host nucleus. Nat. Commun. 8, 15087.

Felföldi, T., Somogyi, B., Márialigeti, K., Vörös, L., 2009. Characterization of photoautotrophic picoplankton assemblages in turbid, alkaline lakes of the Carpathian Basin (Central Europe). J. Limnol. 68, 385–395.

Filée, J., 2013. Route of NCLDV evolution: the genomic accordion. Curr. Opin. Virol. 3, 595–599.

Fischer, M.G., Allen, M.J., Wilson, W.H., Suttle, C.A., 2010. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc. Natl. Acad. Sci. U.S.A. 107, 19508–19513.

Fischer, M.G., Kelly, I., Foster, L.J., Suttle, C.A., 2014. The virion of Cafeteria roenbergensis virus (CroV) contains a complex suite of proteins for transcription and DNA repair. Virology 466–467, 82–94.

Fridmann-Sirkis, Y., Milrot, E., Mutsafi, Y., Ben-Dor, S., Levin, Y., Savidor, A., Kartvelishvily, E., Minsky, A., 2016. Efficiency in complexity: composition and dynamic nature of mimivirus replication factories. J. Virol. 90, 01319-16.

Giegé, R., 2006. The early history of tRNA recognition by aminoacyl-tRNA synthetases. J. Biosci. 477–488.

Goddard, M.R., Burt, A., 1999. Recurrent invasion and extinction of a selfish gene. Proc. Natl. Acad. Sci. U.S.A. 96, 13880–13885.

Grønlien, H.K., Berg, T., Løvlie, A.M., 2002. In the polymorphic ciliate Tetrahymena vorax, the non-selective phagocytosis seen in microstomes changes to a highly selective process in macrostomes. J. Exp. Biol. 205, 2089–2097.

Hammer, U., 1986. Saline Lake Ecosystems of the World. Springer, Dordrecht.

Haugen, P., Simon, D.M., Bhattacharya, D., 2005. The natural history of group I introns. Trends Genet. 21, 111–119.

Hedberg, A., Johansen, S.D., 2013. Nuclear group I introns in self-splicing and beyond. Mob. DNA 4, 17.

Jha, B.K., Jung, H.J., Seo, I., Kim, H.A., Suh, S.I., Suh, M.H., Baek, W.K., 2014. Chloroquine has a cytotoxic effect on acanthamoeba encystation through modulation of autophagy. Antimicrob. Agents Chemother. 58, 6235–6241.

Jha, S., Rollins, M.G., Fuchs, G., Procter, D.J., Hall, E.A., Cozzolino, K., Sarnow, P., Savas, J.N., Walsh, D., 2017. Trans-kingdom mimicry underlies ribosome customization by a poxvirus kinase. Nature 546, 651–655.

Jones, B.F., Eugster, H.P., Rettig, S.L., 1977. Hydrochemistry of the Lake Magadi basin, Kenya. Geochim. Cosmochim. Acta 41, 53–72.

Kerepesi, C., Grolmusz, V., 2016. Giant viruses of the Kutch Desert. Arch. Virol. 161, 721–724.

Kerepesi, C., Grolmusz, V., 2017. The "Giant Virus Finder" discovers an abundance of giant viruses in the Antarctic dry valleys. Arch. Virol. 162, 1671–1676.

Khan, N.A., 2006. Acanthamoeba: biology and increasing importance in human health. FEMS Microbiol. Rev. 30, 564–595.

Klose, T., Herbst, D.A., Zhu, H., Max, J.P., Kenttämaa, H.I., Rossmann, M.G., 2015. A mimivirus enzyme that participates in Viral entry. Structure 23, 1058–1065.

Kraft, C., Deplazes, A., Sohrmann, M., Peter, M., 2008. Mature ribosomes are selectively degraded upon starvation by an autophagy pathway requiring the Ubp3p/Bre5p ubiquitin protease. Nat. Cell Biol. 10, 602–610.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Lambowitz, A.M., Belfort, M., 1993. Introns as mobile genetic elements. Annu. Rev. Biochem. 62, 587–622.

Landthaler, M., Shub, D.A., 1999. Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. Proc. Natl. Acad. Sci. U.S.A. 96, 7005–7010.

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.

Legendre, M., Audic, S., Poirot, O., Hingamp, P., Seltzer, V., Byrne, D., Lartigue, A., Lescot, M., Bernadac, A., Poulain, J., Abergel, C., Claverie, J.M., 2010. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. Genome Res. 20, 664–674.

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, 201320670.

Legendre, M., Lartigue, A., Bertaux, L., Jeudy, S., Bartoli, J., Lescot, M., Alempic, J.-M., Ramus, C., Bruley, C., Labadie, K., Shmakova, L., Rivkina, E., Couté, Y., Abergel, C., Claverie, J.-M., 2015. In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. Proc. Natl. Acad. Sci. U.S.A. 112, 5327–5335.

Lwoff, A., 1957. The concept of virus the third Marjory Stephenson memorial lecture. J. Gen. Microbiol. 17, 239–253.

Ma, J., Edmunds, W.M., 2006. Groundwater and lake evolution in the Badain Jaran Desert ecosystem, Inner Mongolia. Hydrgeol. J. 14, 1231–1243.

Marcelino, V.M., Espinola, M.V.P.C., Serrano-Solis, V., Farias, S.T., 2017. Evolution of the genus Mimivirus based on translation protein homology and its implication in the tree of life. Genet. Mol. Res. 16.

Melack, J.M., 1981. 7. Photosynthetic activity of phytoplankton in tropical African soda lakes. Hydrobiologia 81–82, 71–85.

Mihara, T., Koyano, H., Hingamp, P., Grimsley, N., Goto, S., Ogata, H., 2018. Taxon richness of "Megaviridae" exceeds those of Bacteria and Archaea in the Ocean. Microbes Environ. 33 (2), 162–171.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., Ruger, W., 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67, 86–156.

Monier, A., Claverie, J.-M., Ogata, H., 2008. Taxonomic distribution of large DNA viruses in the sea. Genome Biol. 9, R106.

Moon, E.K., Kim, S.H., Hong, Y., Chung, D.I., Goo, Y.K., Kong, H.H., 2015. Autophagy inhibitors as a potential antiamoebic treatment for Acanthamoeba keratitis. Antimicrob. Agents Chemother. 59, 4020–4025.

Moreira, D., Brochier-Armanet, C., 2008. Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. BMC Evol. Biol. 8, 12.

Moreira, D., López-garcía, P., 2009. Ten reasons to exclude viruses from the tree of life. Nat. Rev. Microbiol. 7, 306–311.

Nasir, A., Kim, K.M., Caetano-Anolles, G., 2012. Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. BMC Evol. Biol. 12, 156.

Nerurkar, P., Altvater, M., Gerhardy, S., Schütz, S., Fischer, U., Weirich, C., Panse, V.G., 2015. Eukaryotic ribosome assembly and nuclear export. In: International Review of Cell and Molecular Biology. pp. 107–140.

Nesbo, C.L., Doolittle, W.F., 2003. Active self-splicing group I introns in 23S rRNA genes of hyperthermophilic bacteria, derived from introns in eukaryotic organelles. Proc. Natl. Acad. Sci. U.S.A. 100, 10806–10811.

Nierhaus, K.H., Wilson, D.N., 2006. Protein Synthesis and Ribosome Structure, Protein Synthesis and Ribosome Structure.

Nishida, K., Suzuki, S., Kimura, Y., Nomura, N., Fujie, M., Yamada, T., 1998. Group I introns found in Chlorella viruses: biological implications. Virology 242, 319–326.

Novoa, E.M., Pavon-Eternod, M., Pan, T., Ribas De Pouplana, L., 2012. A role for tRNA modifications in genome structure and codon usage. Cell 149, 202–213.

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C., Garin, J., Claverie, J.-M., Abergel, C., 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341, 281–286.

Piacente, F., Marin, M., Molinaro, A., De Castro, C., Seltzer, V., Salis, A., Damonte, G., Bernardi, C., Claverie, J.M., Abergel, C., Tonetti, M., 2012. Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (viosamine). J. Biol. Chem. 287, 3009–3018.

Piacente, F., Bernardi, C., Marin, M., Blanc, G., Abergel, C., Tonetti, M.G., 2014. Characterization of a UDP-N-acetylglucosamine biosynthetic pathway encoded by the giant DNA virus Mimivirus. Glycobiology 24, 51–61.

Piacente, F., Gaglianone, M., Laugieri, M.E., Tonetti, M.G., 2015. The autonomous glycosylation of large dna viruses. Int. J. Mol. Sci. 16, 29315–29328.

Raghavan, R., Hicks, L.D., Minnick, M.F., 2008. Toxic introns and parasitic intein in Coxiella burnetii: legacies of a promiscuous past. J. Bacteriol. 190, 5934–5943.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., Scola, B.L., Suzan, M., Claverie, J.-M., 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344–1350.

Redmann, M., Benavides, G.A., Berryhill, T.F., Wani, W.Y., Ouyang, X., Johnson, M.S., Ravi, S., Barnes, S., Darley-Usmar, V.M., Zhang, J., 2017. Inhibition of autophagy with bafilomycin and chloroquine decreases mitochondrial quality and bioenergetic function in primary neurons. Redox Biol. 11, 73–81.

Reteno, D.G., Benamar, S., Khalil, J.B., Andreani, J., Armstrong, N., Klose, T., Rossmann, M., Colson, P., Raoult, D., La Scola, B., 2015. Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. J. Virol. 89, 6585–6594.

Ribas de Pouplana, L., Schimmel, P., 2001. Aminoacyl-tRNA synthetases: potential markers of genetic code development. Trends Biochem. Sci. 26, 591–596.

Rodrigues, R.A.L., dos Santos Silva, L.K., Dornas, F.P., de Oliveira, D.B., Magalhães, T.F.F., Santos, D.A., Costa, A.O., de Macêdo Farias, L., Magalhães, P.P., Bonjardim, C.A., Kroon, E.G., La Scola, B., Cortines, J.R., Abrahão, J.S., 2015. Mimivirus fibrils are important for viral attachment to the microbial world by a diverse glycoside interaction repertoire. J. Virol. 89, 11812–11819.

Rodrigues, T., Belmok, A., Catão, E., Kyaw, C.M., 2016. Archaea in natural and impacted Brazilian environments. Archaea 2016, 1259608.

Sandegren, L., Sjöberg, B.M., 2007. Self-splicing of the bacteriophage T4 group I introns requires efficient translation of the pre-mRNA in vivo and correlates with the growth state of the infected bacterium. J. Bacteriol. 189, 980–990.

Schilde, C., Schaap, P., 2013. The Amoebozoa. Methods Mol. Biol. 983, 1–15.

Schulz, F., Yutin, N., Ivanova, N.N., Ortega, D.R., Lee, T.K., Vierheilig, J., Daims, H., Horn, M., Wagner, M., Jensen, G.J., Kyrpides, N.C., Koonin, E.V., Woyke, T., 2017. Giant viruses with an expanded complement of translation system components. Science 356, 82–85.

Schvarcz, C.R., Steward, G.F., 2018. A giant virus infecting green algae encodes key fermentation genes. Virology 518, 423–433.

Sergeev, V.N., Gerasimenko, L.M., Zavarzin, G.A., 2002. The proterozoic history and present state of cyanobacteria. Microbiology 71, 623–637.

Sobhy, H., La Scola, B., Pagnier, I., Raoult, D., Colson, P., 2015. Identification of giant Mimivirus protein functions using RNA interference. Front. Microbiol. 6, 1–11.

Solitro, A.R., MacKeigan, J.P., 2016. Leaving the lysosome behind: novel developments in autophagy inhibition. Future Med. Chem. 8, 73–86.

Sorokin, D.Y., Banciu, H.L., Muyzer, G., 2015. Functional microbiology of soda lakes. Curr. Opin. Microbiol. 25, 88–96.

Suhre, K., Audic, S., Claverie, J.-M., 2005. Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. Proc. Natl. Acad. Sci. U.S.A. 102, 14689–14693.

Suttle, C.A., 2005. Viruses in the sea. Nature 437, 356–361.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Van Etten, J.L., Meints, R.H., 1999. Giant viruses infecting algae. Annu. Rev. Microbiol. 53, 447–494.

Virgin, H.W., Latreille, P., Wamsley, P., Hallsworth, K., Weck, K.E., Dal Canto, A.J., Speck, S.H., 1997. Complete sequence and genomic analysis of murine gammaherpesvirus 68. J. Virol. 71, 5894–5904.

Wang, Z., Wu, M., 2017. Comparative genomic analysis of acanthamoeba endosymbionts highlights the role of amoebae as a "melting pot" shaping the Rickettsiales evolution. Genome Biol. Evol. 9, 3214–3224.

Xiao, C., Kuznetso, Y.G., Sun, S., Hafenstein, S.L., Kostyuchenko, V.A., Chipman, P.R., Suzan-Monti, M., Raoult, D., McPherson, A., Rossmann, M.G., 2009. Structural studies of the giant Mimivirus. PLoS Biol. 7, 0958–0966.

Yamada, T., Tamura, K., Aimi, T., Songsri, P., 1994. Self-splicing group I introns in eukaryotic viruses. Nucleic Acids Res. 22, 2532–2537.

Yutin, N., Wolf, Y.I., Raoult, D., Koonin, E.V., 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol. J. 6, 223.

Yutin, N., Wolf, Y.I., Koonin, E.V., 2014. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology 466–467, 38–52.

Zauberman, N., Mutsafi, Y., Halevy, D.B., Shimoni, E., Klein, E., Xiao, C., Sun, S., Minsky, A., 2008. Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. PLoS Biol. 6, 1104–1114.

Keywords: Tupanvirus; giant virus; *Mimiviridae*; Host range; Genetic complexity; Translational apparatus; 18S-like intron; virus evolution

Initial phylogenetic analyses using the aaRS of tupanviruses as bait for retrieval of only the best hits and some amoebozoa organisms in public databases placed the viruses as external groups, thus suggesting an independent origin from most of aaRS genes in tupanviruses, related to cellular genes (Abrahão et al., 2018). New analyses using TPV-SL aaRS se-

quences against specific taxa in databases [Eukarya (taxid: 2759); Amoebozoa (taxid: 554915); Archaea (taxid: 2157); Firmicutes (taxid: 1239); Proteobacteria (taxid: 1224); Viruses (taxid: 10239)] and retrieving up to 30 random sequences for each taxa (when available, and 10 for amoebozoa), we observed an alternative topology of the trees, with viruses clustered within cellular branches, either in a monophyletic group or dispersed in the tree, suggestive of horizontal gene transfer events (Fig. 5B and C; Supplementary Fig. 1). It is noteworthy that canonical taxonomic groups are frequently separated in the trees, independent of the gene used. This pattern has already been observed in phylogenetic reconstructions performed by different groups (Schulz et al., 2017). Methodological issues related to analyses using aaRS have been and still are debated. Therefore, in light of the available data and methods the origin of these genes remains a mystery.

4.5. ARTIGO #5: “*Tupanvirus*”, a new genus in the family *Mimiviridae*

Os tupanvírus são um novo grupo de vírus gigantes recentemente isolados de amostras coletadas em ambientes extremos no Brasil. A descoberta destes vírus chamou a atenção da comunidade científica devido a sua estrutura peculiar e conteúdo gênico único apresentando o mais completo arsenal de genes relacionados ao processo de síntese proteica identificado na virosfera. Reconstruções filogenéticas colocaram os tupanvírus como grupo irmão de mimivírus, sendo assim parte da família *Mimiviridae*. Contudo, estes vírus apresentam várias características que os distinguem dos demais mimivírus. Neste trabalho nós compilamos uma série de propriedades dos tupanvírus que contrastam com aquelas observadas para outros membros da família *Mimiviridae*, o que justificaria a criação de um novo grupo taxonômico para classifica-los adequadamente. Os tupanvírus apresentam um capsídeo similar ao observado para os mimivírus, porém acompanhado de uma longa cauda cilíndrica, resultado em partículas com tamanho médio de 1,2 μm. Eles são capazes de infectar amebas de diferentes grupos taxonômicos, além de causarem um efeito tóxico tanto em células permissivas quanto não hospedeiras, um fenótipo associado a um processo de shutdown ribossomal. Além disso, o efeito citopático observado é distinto daquele induzido por outros vírus gigantes. O genoma dos tupanvírus apresentam diversos genes exclusivos, incluindo 30% de ORFans, além de uma organização gênica consideravelmente diferentes de outros membros da família *Mimiviridae*. Ainda, realizamos análises filogenéticas utilizando diferentes genes conservados para os vírus gigantes e os resultados corroboram dados anteriores, colocando os tupanvírus como um grupo a parte do gênero *Mimivirus*. Diante disso propusemos a criação do novo gênero “*Tupanvirus*” contendo duas novas espécies, *Tupanvirus soda lake* e *Tupanvirus deep ocean*. Em paralelo, a proposta oficial foi encaminhada ao líder do grupo de estudos responsável pela família taxonomia da família *Mimiviridae* dentro do Comitê Internacional de Taxonomia Viral (ICTV) e se encontra em análise pelo comitê executivo da organização.

Este artigo foi publicado no periódico ***Archives of Virology*** em outubro de 2018.

Archives of Virology
https://doi.org/10.1007/s00705-018-4067-4

VIROLOGY DIVISION NEWS

"*Tupanvirus*", a new genus in the family *Mimiviridae*

Rodrigo Araújo Lima Rodrigues[1,2] · Said Mougari[1] · Phillipe Colson[1] · Bernard La Scola[1] · Jônatas Santos Abrahão[2]

Received: 6 July 2018 / Accepted: 30 September 2018

Abstract
The genus "*Tupanvirus*" is a new proposed taxon to be included in the family *Mimiviridae*. The two known tupanvirus isolates were isolated from soda lake and oceanic sediments samples collected in Brazil and were named "tupanvirus soda lake" and "tupanvirus deep ocean", respectively. These viruses exhibit similarities to amoeba-infecting mimiviruses, but there are also several differences that place them in a separate group within the family *Mimiviridae*. Their virions have a mean size of 1.2 µm, which include a mimivirus-like capsid and a large cylindrical tail, both covered by fibrils. The linear double-stranded DNA genomes of up to 1,516,267 base pairs encode over 1,200 genes, among which ~30% have no homologs in any database, including in other mimivirus genomes. Compared to other mimiviruses, tupanviruses exhibit a broader host range and cause a cytotoxic effect in host and non-host organisms, a phenotype that is not observed for other mimiviruses. Remarkably, these viruses possess the most complete gene set related to the protein synthesis process, including 20 aminoacyl-tRNA synthetases, 67-70 tRNAs, many translation factors, and genes involved in maturation and modification of tRNA and mRNA, among others. Moreover, diverse phylogenomic analyses put tupanviruses in a distinct group within the family *Mimiviridae*. In light of the set of different features observed for these giant viruses, we propose establishment of a new genus to allow proper classification of two known tupanviruses and possibly many more similar viruses yet to be characterized.

Introduction

The discovery of Acanthamoeba polyphaga mimivirus (APMV) in 2003 paved the way for new and outstanding discoveries in the virology field, expanding our knowledge about diversity, evolution and complexity of viruses [1]. Given to its distinct morphological and genetic features, a new viral family, named *Mimiviridae*, was created to accommodate this virus [2]. This family currently includes two genera, the genus *Mimivirus*, in which the only officially recognized species is typified by APMV, and the genus *Cafeteriavirus*, with a single species, *Cafeteria roenbergensis virus*, whose members are distantly related to APMV and infect marine flagellates [3].

Dozens of other mimiviruses infecting free-living amoebae have been described over the past decade in different parts of the world and from distinct environments/hosts, and these viruses exhibit biological, structural and genomic characteristics similar to those observed for APMV [4–11]. In addition, some viruses that are known to infect unicellular algae are proposed members of the family *Mimiviridae*, since they are phylogenetically closer to amoebae-infecting mimiviruses than to other algae-infecting viruses, which are usually classified in the family *Phycodnaviridae* [12]. Furthermore, a new virus infecting a kinetoplastid protozoan, named "Bodo saltans virus" [13], was recently proposed to be part of a new group of complex mimiviruses, along with four other viruses whose genomes were assembled from environmental metagenomes but that were not isolated, and these were proposed to comprise a subfamily named "*Klosneuvirinae*" within the family *Mimiviridae* [14].

Handling Editor: Sead Sabanadzovic.

Rodrigo Araújo Lima Rodrigues and Said Mougari contributed equally to this work.

The taxonomic changes proposed here have not been endorsed by the ICTV Executive Committee. This article is related to an ongoing taxonomic proposal, submitted to the ICTV but not yet accepted at the time of submission.

✉ Jônatas Santos Abrahão
jonatas.abrahao@gmail.com

[1] Aix-Marseille Université, IRD, APHM, MEPHI, IHU-Méditerranée Infection, Marseille, France

[2] Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Published online: 05 October 2018

Recently, we described the discovery of two new members of the family *Mimiviridae*, the tupanviruses, which have a set of features that strongly differ from the mimiviruses described to date [15]. These viruses were isolated from soda lake water and oceanic sediment samples collected in Brazil and were named in tribute to Tupan – or Tupã – (God of Thunder) which is an important mythological figure to the South American Guarani indigenous tribes. The new and intriguing characteristics observed for the tupanviruses led us to propose the creation of a new taxon in the family *Mimiviridae*, the genus "*Tupanvirus*".

Morphological properties

The tupanviruses have a capsid similar in size to those observed for other mimiviruses (~450 nm), which exhibit a 'stargate' structure [16] and is covered by fibrils. However, unlike other viruses, the tupanviruses have a large cylindrical tail attached to the base of the capsid of ~550 nm in length and ~450 nm in diameter (including fibrils) (Fig. 1A and B). The mean size of tupanvirus' virions is 1.2 µm, which allows them to be visualized easily using optical microscopy, although particles up to 2.3 µm have been observed due to a high degree of plasticity in size of the tail [15]. A lipid membrane is present in the capsid, similar to other giant viruses [17, 18], which is associated with fusion with the

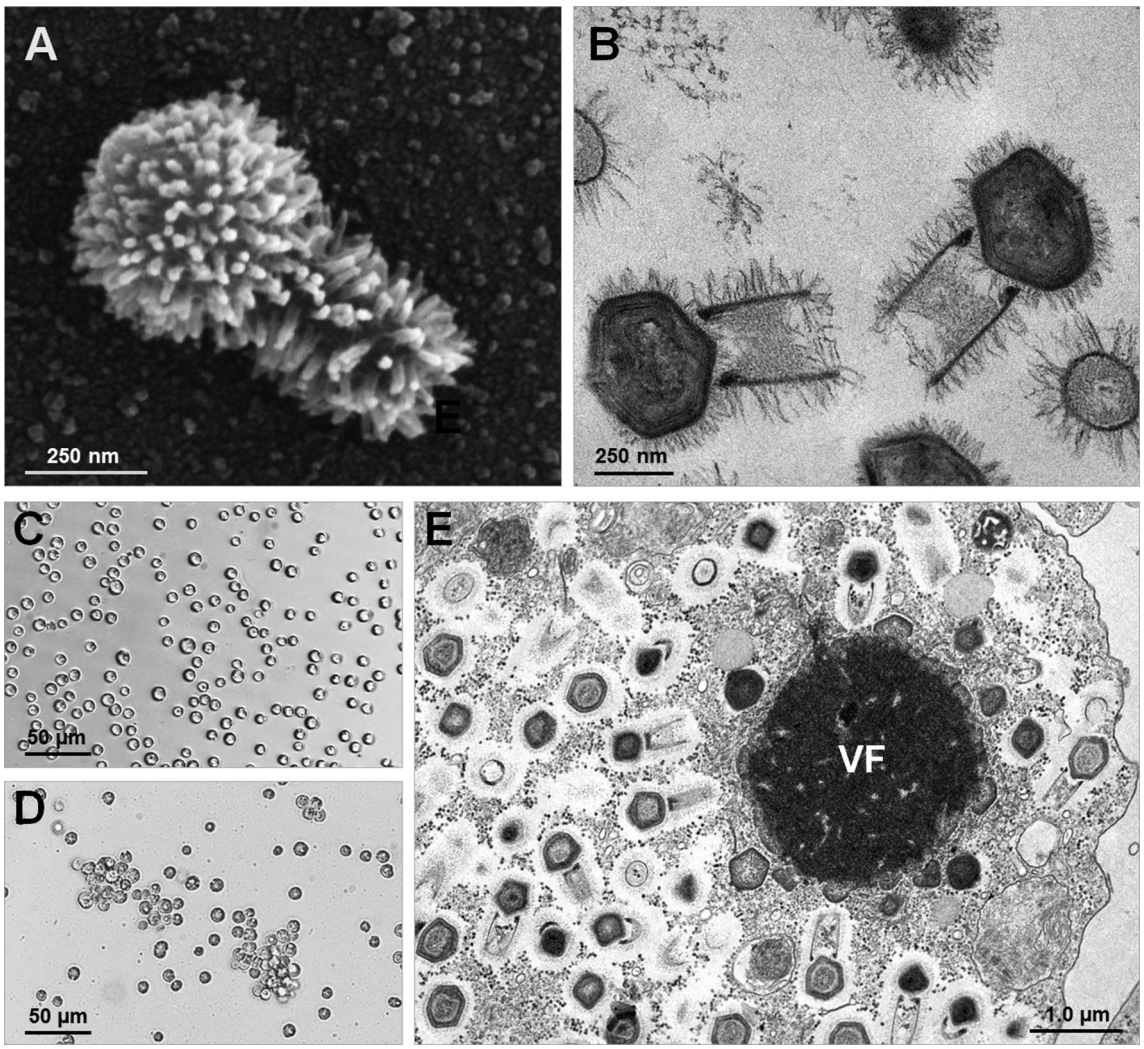


Fig. 1 Morphological and replication properties of tupanviruses. (A) Scanning electron microscopy of a tupanvirus particle. (B) Transmission electron microscopy (TEM) of tupanvirus particles, showing a mimivirus-like capsid and a cylindrical tail attached at its base. Optical microscopy of cytopathic effect of mimivirus (C) and tupanvirus (D). (E) TEM image of an *A. castellanii* cell infected with a tupanvirus, representing a late stage of the replication cycle, showing the viral factory (VF) and mature viruses around this structure

phagosome membrane and release of the capsid contents. The tail is less electron-dense than the capsid and its contents are still unknown (Fig. 1B). Nevertheless, this material seems to be released into the host cytoplasm after invagination of the phagosome membrane inside the tail [15].

Prevalence and host range

The distribution of tupanviruses in nature is uncertain. The two viruses were isolated from separate samples and had different characteristics. The first one was isolated from water samples collected from Brazilian soda lakes, which have high salinity and pH (~9.0), and it was named "tupanvirus soda lake" (TPV-SL), while the second one, named "tupanvirus deep ocean" (TPV-DO), was isolated from ocean sediment samples collected at a depth of 3000 m in the Brazilian Atlantic Ocean (Table 1). Interestingly, a mimivirus-like endoparasite with a thin tail attached to the capsid was isolated from plane tree bark samples collected in Germany using *Saccamoeba* sp. as bait, and was named KSL5x [19]. Since no genomic information is available about KSL5x, it is not possible to confirm that it is another tupanvirus, although they are morphologically similar. New analysis should be conducted in this regard, which could provide valuable insights about the ecology and evolution of these viruses.

The natural hosts of tupanviruses are also unknown. Unlike other amoeba-infecting mimiviruses, the tupanviruses are able to infect and establish a productive cycle *in vitro* in different species of amoebae of the genus *Acanthamoeba*, *Vermamoeba vermiformis*, *Dictyostelium discoideum* and *Willartia magna* [15]. Furthermore, a cytotoxic phenotype in non-host cells (*Tetrahymena hyperangularis*, RAW247, and THP-1 cells) was observed, as well as for *A. castellanii* at high multiplicity of infection, something that is not observed for APMV. This toxic profile seems to be related to a shutdown of ribosomal RNA in infected cells, while the autophagy/ribophagy canonical pathways are not involved in the process, and thus the mechanism of this phenotype is still unknown [15].

Properties in culture

The tupanviruses cause a cytopathic effect in *A. castellanii* cells that is typical of mimiviruses, characterized by loss of motility, rounding, and further cell lysis. However, unlike what is usually observed for mimiviruses, the tupanviruses caused cell aggregation, forming clusters of round cells previous to the lysis step (Fig. 1C and D). At the ultrastructural level, the tupanviruses display a replicative cycle similar to that of other mimiviruses [15, 20]. Viral particles enter into host cells by phagocytosis 1 h postinfection. After opening

Table 1 Description of primary features of the tupanviruses

Virus	Source	Country/region	Year of isolation	Genome GenBank accession no.	Genome topology	Genome size (bp)	Number of genes	Number of translation-related genes[a]	Number of tRNAs	Number of ORFans
TPV-SL	Soda lake	Brazil (Pantanal)	2014	KY523104.1	Linear	1,439,508	1276	61	67	375
TPV-DO	Ocean sediment	Brazil (Campos dos Goytacazes)	2016	MF405918.1	Linear	1,516,267	1359	64	70	378

[a]Including aminoacyl-tRNA synthetases, tRNA-modifying proteins, mRNA and ribosomal protein maturation

of the 'stargate' and fusion of capsid inner membrane to the phagosome membrane, the genome is released to the host cytoplasm, where a viral factory is established (Fig. 1E). In this region, genome replication and viral morphogenesis takes place; the capsid is built first and then the tail is attached. Mature virions are released by cell lysis in the last step of the viral cycle.

Genomic and proteomic features

The genomes of tupanviruses are linear double-stranded DNA molecules with ~28% G+C. The TPV-SL genome is 1,439,508 bp in length and contains 1276 genes, while the TPV-DO has a genome of 1,516,267 bp that contains 1359 genes (Table 1). Around 30% of the predicted genes for the tupanviruses are completely new, named ORFans (open reading frames with no similarities in the databases). Among the ORFs with known function, up to 50% exhibit best hits in other viruses, especially members of the family *Mimiviridae*, while the other genes have best hits in cellular organisms, mainly eukaryotes and bacteria [15]. The structure and genome architecture are well conserved between the two tupanviruses, except for the terminal region of the genome of TPV-SL, which appears to be translocated and inverted in the genome of TPV-DO. In contrast, these genomes exhibit a strongly different synteny when compared to the genomes of the prototype viruses of the genera *Mimivirus* and *Cafeteriavirus* (Fig. 2). Moreover, the tupanviruses have the promoter motif AAAATTGA and a codon and amino acid usage frequency similar to that reported for other mimiviruses [3, 15, 21].

The tupanviruses have the most complete arsenal of genes related to the protein translation process among the members of the virosphere. Surprisingly, these viruses encode 67-70 tRNAs, 20 aminoacyl-tRNA synthetases (aaRS), and 41-44 other proteins involved in the translation process, such as translation factors (initiation, elongation and release), and factors related to maturation of tRNA and mRNA, among others (Table 1). The amount and diversity of these genes far exceed what was previously known for other giant viruses, even including the klosneuviruses, which possess up to 19 aaRS [14], and Bodo Saltans virus. The origin of these genes is still a matter of a hot debate in the scientific community, and a consensus has not yet been reached on whether these genes came from several events of lateral gene transfer from hosts and/or sympatric organisms or were already contain in the ancestors of mimiviruses [22–24]. Furthermore, the tupanviruses present two copies of an intronic region of the 18S rRNA gene, which are highly expressed during the viral replication cycle, but their function remains to be elucidated [15]. It is noteworthy that no exonic region of the 18S rRNA gene was found in tupanviruses or other viruses, nor do they possess genes related to the energy synthesis process [15].

Proteomic analysis of TPV-SL particles revealed the presence of 127 proteins, of which 67 have no known function, and 11 are encoded by ORFans. No aaRS or translation factors were detected in the viral particles. Among the

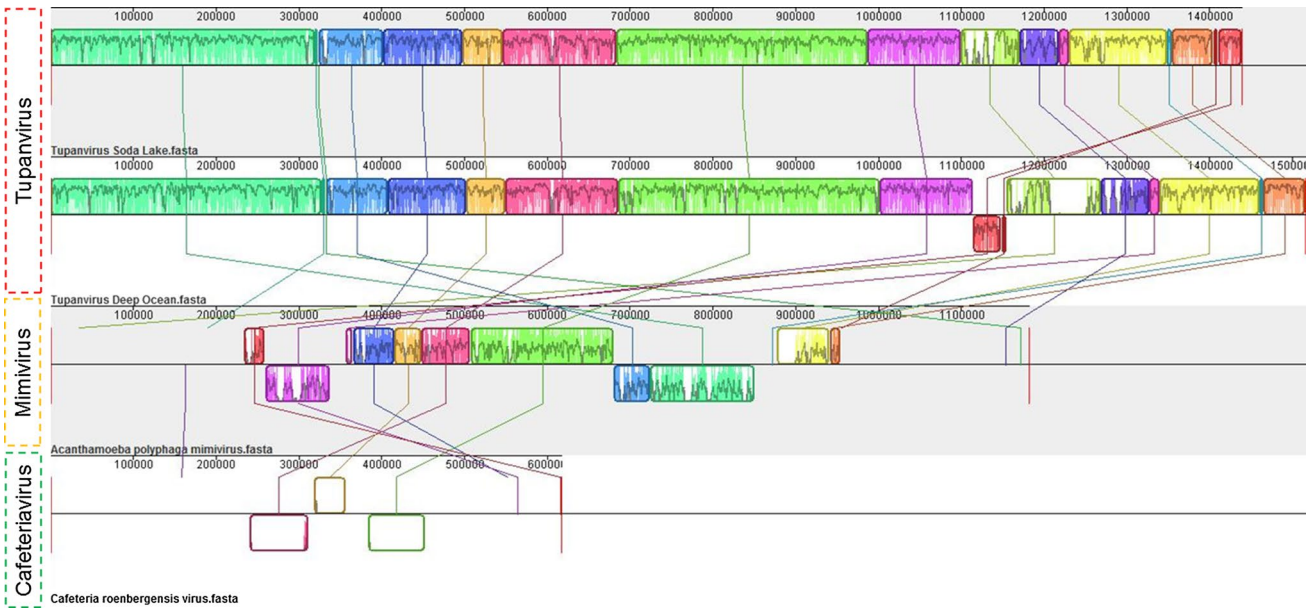


Fig. 2 Genome alignment of tupanviruses, APMV (NC_014649.1), and CroV (NC_014637.1). The figure shows genome architecture and synteny of prototype viruses of the genera of the family *Mimiviridae*. The boxes highlight the collinear blocks of similarity between the genomes. The graphs within the boxes represent the average similarity of each region, with a range of values and the mean value darkened. The schematic genome alignment diagram was produced using the MAUVE software package [26]

proteins with a known function, several structural components were observed, such as the major capsid protein and the core protein, as well as enzymes involved in nucleic acid metabolism, such as a DNA polymerase, a DNA-dependent RNA polymerase, a RNA helicase, and an mRNA capping enzyme. It is notable that fewer than 50% of the proteins found in TPV-SL particles are shared with APMV or CroV [15].

Phylogenomics

Phylogenetic analysis based on different core genes of giant viruses, named DNA polymerase B family, major capsid protein, D5 primase helicase and D6/D11 helicase (both individual and concatenated) put the tupanviruses within the family *Mimiviridae* (Fig. 3). Notably, when the phylogenetic reconstruction is performed using the D6/D11 helicase gene, the tupanviruses are placed as a sister group of amoebae-infecting mimiviruses (Fig. 3D). A comparable topology is observed when the analysis is carried out by hierarchical clustering based on a presence-absence matrix of clusters of orthologous genes shared by the different mimiviruses [15]. Moreover, our analyses corroborate previous studies that indicate the existence of two other groups of mimiviruses, referred to as "extended-family" *Mimiviridae*, comprising the algae-infecting mimiviruses and the klosneuviruses [12, 13]. These analyses highlight the differences between tupanviruses and other mimiviruses and the need to create a new taxonomic group to properly classify these new viruses.

Conclusion

The tupanviruses represent a new group of viruses isolated from extreme environments in Brazil. Although they share many features with mimiviruses of amoebae, the tupanviruses have a set of distinctive features that place them in a different group within the family *Mimiviridae*, the proposed genus "*Tupanvirus*". This new taxon would initially include

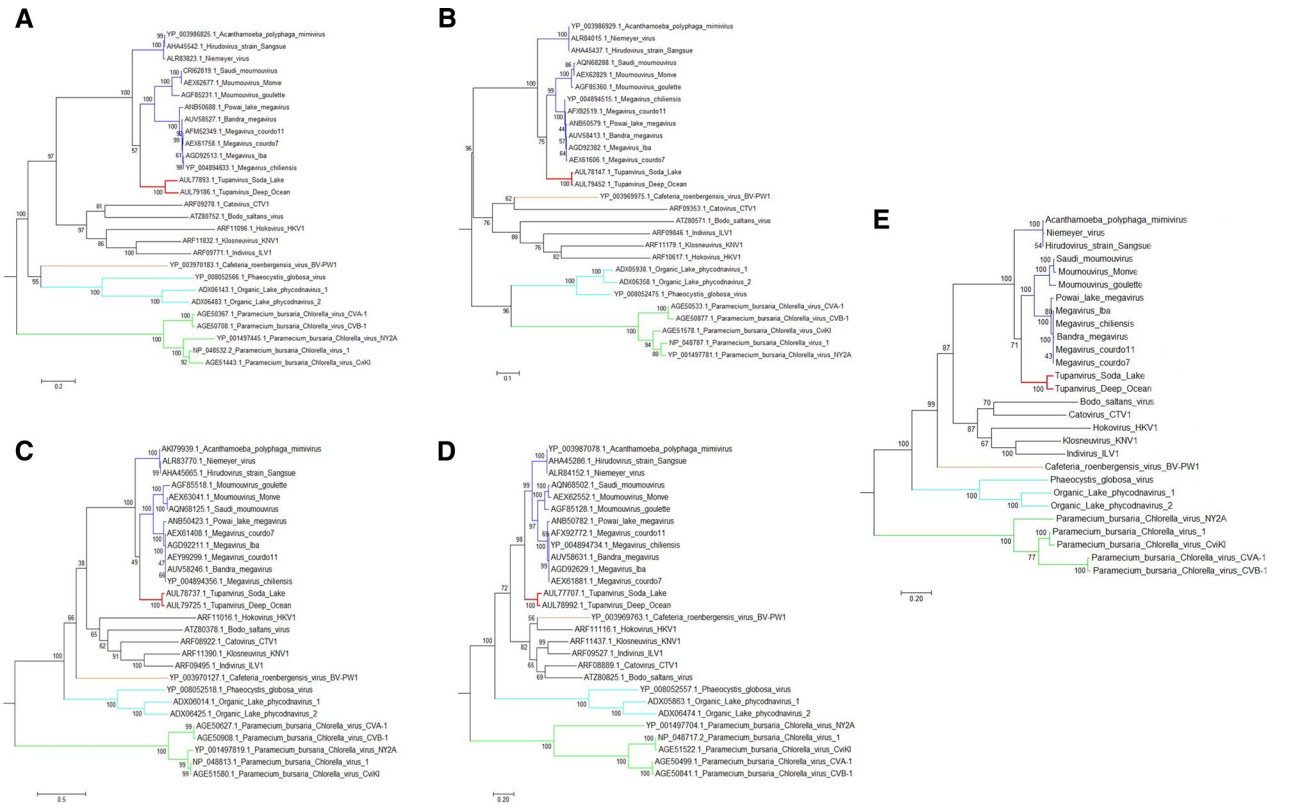


Fig. 3 Phylogenetic reconstruction using amino acid sequences of core genes of the nucleo-cytoplasmic large DNA viruses. The analyses were performed using different representatives of the family *Mimiviridae* and phycodnaviruses as an external group, based on (A) DNA polymerase B family, (B) major capsid protein, (C) D5 primase helicase, (D) D6/D11 helicase, and (E) the genes concatenated. The alignments were built using Muscle software [27]. The tree was constructed using MEGA version 6.0, applying the maximum-likelihood method and JTT model of evolution with 1000 bootstrap replicates [28]. Colors indicate different viral groups: blue, mimivirus lineage A, B, and C; red, tupanviruses; orange, CroV; black, klosneuviruses; cyan, extended *Mimiviridae*; green, *Phycodnaviridae*. Scale bars indicate rate of evolution

two species, "*Tupanvirus soda lake*" and "*Tupanvirus deep ocean*". With the advancement of the isolation and characterization techniques of giant viruses, we can expect even more new viruses to be revealed in the near future, therefore expanding our knowledge about the virosphere.

The *Mimiviridae* taxonomy is an open field that requires further attention. Many mimiviruses have been described during the last few years, comprising three different clades named lineage A, B, and C, represented by mimivirus, moumouvirus, and megavirus, respectively [4–6]. These viruses exhibit considerable differences at the genomic level that might constitute different genera, thus splitting the genus *Mimivirus*. Furthermore, the klosneuviruses could represent another genus within the family, alongside with the now proposed "*Tupanvirus*", which would be part of a subfamily, as previously suggested for the amoebae-infecting mimiviruses [25]. Additional efforts should be made to reorganize the taxonomy of the whole family *Mimiviridae*, which will strongly contribute to the systematics of mimiviruses.

Acknowledgements We thank colleagues from Aix Marseille University and from Universidade Federal de Minas Gerais, and the Center of Microscopy of UFMG for technical assistance. We also thank CAPES, CNPq, FAPEMIG and Mediterranée Infection for scholarships and grants. JSA is a CNPq researcher. The authors are members of a CAPES-COFECUB project. The authors declare no conflict of interest.

Compliance with ethical standards

Ethical statement We declare that this work complies with the ethical rules applicable for this journal.

References

1. Colson P, La Scola B, Levasseur A et al (2017) Mimivirus: leading the way in the discovery of giant viruses of amoebae. Nat Rev Microbiol 15:243–254. https://doi.org/10.1038/nrmicro.2016.197
2. Raoult D, Audic S, Robert C et al (2004) The 1.2-Megabase genome sequence of Mimivirus. Science 306(80):1344–1350. https://doi.org/10.1126/science.1101485
3. Fischer MG, Allen MJ, Wilson WH, Suttle CA (2010) Giant virus with a remarkable complement of genes infects marine zooplankton. Proc Natl Acad Sci 107:19508–19513. https://doi.org/10.1073/pnas.1007615107
4. Colson P, Yutinf N, Shabalina SA et al (2011) Viruses with more than 1,000 genes: Mamavirus, a new *Acanthamoeba* polyphaga mimivirus strain, and reannotation of mimivirus genes. Genome Biol Evol 3:737–742. https://doi.org/10.1093/gbe/evr048
5. Arslan D, Legendre M, Seltzer V et al (2011) Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci 108:17486–17491. https://doi.org/10.1073/pnas.1110889108
6. Yoosuf N, Yutin N, Colson P et al (2012) Related giant viruses in distant locations and different habitats: *Acanthamoeba* polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the Megavirus lineage. Genome Biol Evol 4:1324–1330. https://doi.org/10.1093/gbe/evs109
7. Dornas FP, Khalil JYB, Pagnier I et al (2015) Isolation of new Brazilian giant viruses from environmental samples using a panel of protozoa. Front Microbiol 6:1–9. https://doi.org/10.3389/fmicb.2015.01086
8. Saadi H, Reteno DGI, Colson P et al (2013) Shan virus: a new mimivirus isolated from the stool of a tunisian patient with pneumonia. Intervirology 56:424–429. https://doi.org/10.1159/000354564
9. Campos RK, Boratto PV, Assis FL et al (2014) Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. Virol J 11:95. https://doi.org/10.1186/1743-422X-11-95
10. Andrade KR, Boratto PPVM, Rodrigues FP et al (2014) Oysters as hot spots for mimivirus isolation. Arch Virol 160:477–482. https://doi.org/10.1007/s00705-014-2257-2
11. Andrade ACDSP, Arantes TS, Rodrigues RAL et al (2018) Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. Virol J 15:1–10. https://doi.org/10.1186/s12985-018-0930-x
12. Santini S, Jeudy S, Bartoli J et al (2013) Genome of *Phaeocystis* globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. Proc Natl Acad Sci USA 110:10800–10805. https://doi.org/10.1073/pnas.1303251110
13. Deeg CM, Chow C-ET, Suttle CA (2018) The kinetoplastid-infecting *Bodo saltans* virus (BsV), a window into the most abundant giant viruses in the sea. Elife. https://doi.org/10.7554/elife.33014
14. Schulz F, Yutin N, Ivanova NN et al (2017) Giant viruses with an expanded complement of translation system components. Science 356(80):82–85. https://doi.org/10.1126/science.aal4657
15. Abrahão J, Silva L, Silva LS et al (2018) Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat Commun 9:749. https://doi.org/10.1038/s41467-018-03168-1
16. Zauberman N, Mutsafi Y, Halevy DB et al (2008) Distinct DNA exit and packaging portals in the virus *Acanthamoeba* polyphaga mimivirus. PLoS Biol 6:1104–1114. https://doi.org/10.1371/journal.pbio.0060114
17. Mutsafi Y, Shimoni E, Shimon A, Minsky A (2013) Membrane assembly during the infection cycle of the giant mimivirus. PLoS Pathog 9:100. https://doi.org/10.1371/journal.ppat.1003367
18. Mutsafi Y, Fridmann-Sirkis Y, Milrot E et al (2014) Infection cycles of large DNA viruses: emerging themes and underlying questions. Virology 466–467:3–14. https://doi.org/10.1016/j.virol.2014.05.037
19. Michel R, Loch S, Muller KD et al (2017) Reisolation of mimivirus-like giant viruses by using *Saccamoeba* sp. as bait resulted in loss of virophages. Endocytobiosis Cell Res 28:50–53
20. Andrade ACDSP, Rodrigues RAL, Oliveira GP et al (2017) Filling gaps about mimivirus entry, uncoating and morphogenesis. J Virol. https://doi.org/10.1128/jvi.01335-17
21. Suhre K, Audic S, Claverie J-M (2005) Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. Proc Natl Acad Sci USA 102:14689–14693. https://doi.org/10.1073/pnas.0506465102
22. 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
23. Colson P, De Lamballerie X, Fournous G, Raoult D (2012) Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. Intervirology 55:321–332. https://doi.org/10.1159/000336562
24. Moreira D, López-García P (2015) Evolution of viruses and cells: do we need a fourth domain of life to explain the origin of eukaryotes? Philos Trans R Soc Lond B Biol Sci 370:20140327. https://doi.org/10.1098/rstb.2014.0327

25. Gallot-Lavallée L, Blanc G, Claverie J-M (2017) Comparative genomics of *Chrysochromulina* ericina virus and other microalga-infecting large DNA viruses highlights their intricate evolutionary relationship with the established mimiviridae family. J Virol 91:JVI-00230. https://doi.org/10.1128/jvi.00230-17
26. 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
27. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinf 5:113. https://doi.org/10.1186/1471-2105-5-113
28. Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197

4.6. ARTIGO #6: Morphologic and genomic analyses of new isolates reveal a second lineage of Cedratviruses

Vírus gigantes têm sido isolados e caracterizados a partir de amostras distintas coletadas em todo o planeta. Estes vírus apresentam formas variadas e genomas extensos, compostos por centenas de genes, muitos dos quais não possuem função conhecida. Em 2016, foi descoberto um novo grupo viral denominado “Cedratvirus” composto atualmente por apenas dois isolados, sendo membros da possível nova família “*Pithoviridae*”. Neste trabalho nós apresentamos o isolamento e caracterização biológica e genômica de dois novos cedratvírus, isolados de amostras coletadas na França e no Brasil. Os vírus foram isolados em células de *Acanthamoeba castellanii*, purificados e submetidos a análises de microscopia eletrônica de transmissão para caracterização morfológica e do ciclo. Em paralelo foi realizado o sequenciamento completo do genoma utilizado as plataformas MiSeq e MinIon para posterior caracterização genômica. Ambos os vírus apresentam partículas ovoides com opérculos nas duas extremidades, uma característica típica dos cedratvírus. O vírus brasileiro, denominado Brazilian Cedratvirus IHUMI, apresenta partículas com 910 nm e possui um genoma de 460.038 bp, dimensões menores que o observado para os demais cedratvírus conhecidos. Análises de regressão linear indicam que os membros da família “*Pithoviridae*” estão no limite de predição da lei alométrica envolvendo o volume das partículas virais e o tamanho do genoma. Além disso, esse vírus possui um genoma completamente assintênico e uma diferença considerável na similaridade de aminoácidos de genes ortólogos (73% em relação aos demais vírus). Análises de pan-genoma envolvendo os quatro cedratvírus revelaram um aumento do pan-genoma e uma redução do arsenal gênico comum aos isolados ao adicionar o vírus brasileiro. Por fim, análises filogenéticas utilizando diferentes genes colocaram o vírus brasileiro em um ramo separado dos demais cedratvírus. O conjunto de dados obtidos nos levou a propor a existência de uma segunda linhagem para este novo grupo de vírus e fornece novas informações sobre a biodiversidade destes vírus gigantes.

Este artigo foi publicado no periódico ***Journal of Virology*** em junho de 2018.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

Morphologic and Genomic Analyses of New Isolates Reveal a Second Lineage of Cedratviruses

Rodrigo Araújo Lima Rodrigues,[a,b] Julien Andreani,[a] Ana Cláudia dos Santos Pereira Andrade,[b] Talita Bastos Machado,[b] Souhila Abdi,[a] Anthony Levasseur,[a] Jônatas Santos Abrahão,[b] Bernard La Scola[a]

[a]Aix Marseille Université, IRD, APHM, MEPHI, IHU-Méditerranée Infection, Marseille, France

[b]Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

ABSTRACT Giant viruses have been isolated and characterized in different environments, expanding our knowledge about the biology of these unique microorganisms. In the last 2 years, a new group was discovered, the cedratviruses, currently composed of only two isolates and members of a putative new family, "Pithoviridae," along with previously known pithoviruses. Here we report the isolation and biological and genomic characterization of two novel cedratviruses isolated from samples collected in France and Brazil. Both viruses were isolated using *Acanthamoeba castellanii* as a host cell and exhibit ovoid particles with corks at either extremity of the particle. Curiously, the Brazilian cedratvirus is ~20% smaller and presents a shorter genome of 460,038 bp, coding for fewer proteins than other cedratviruses. In addition, it has a completely asyntenic genome and presents a lower amino acid identity of orthologous genes (~73%). Pangenome analysis comprising the four cedratviruses revealed an increase in the pangenome concomitant with a decrease in the core genome with the addition of the two novel viruses. Finally, phylogenetic analyses clustered the Brazilian virus in a separate branch within the group of cedratviruses, while the French isolate is closer to the previously reported *Cedratvirus lausannensis*. Taking all together, we propose the existence of a second lineage of this emerging viral genus and provide new insights into the biodiversity and ubiquity of these giant viruses.

IMPORTANCE Various giant viruses have been described in recent years, revealing a unique part of the virosphere. A new group among the giant viruses has recently been described, the cedratviruses, which is currently composed of only two isolates. In this paper, we describe two novel cedratviruses isolated from French and Brazilian samples. Biological and genomic analyses showed viruses with different particle sizes, genome lengths, and architecture, revealing the existence of a second lineage of this new group of giant viruses. Our results provide new insights into the biodiversity of cedratviruses and highlight the importance of ongoing efforts to prospect for and characterize new giant viruses.

KEYWORDS *Cedratvirus*, giant virus, NCLDV, new lineage, virion volume, genome length, pangenome

Viruses are the most abundant microorganisms in the biosphere and present the greatest genetic and morphological diversity among the known biological organisms (1, 2). Different groups of viruses have specific characteristics that define them as a group, such as capsid structure (e.g., icosahedral and helical) and type of genome (e.g., double-stranded DNA [dsDNA] and single-stranded RNA negative sense [ssRNA−]), which implicate differences in the life cycles and evolution of these viruses. Moreover, remarkable differences can be seen in the virion volumes and genome

Received 6 March 2018 **Accepted** 13 April 2018

Accepted manuscript posted online 25 April 2018

Citation Rodrigues RAL, Andreani J, Andrade ACSP, Machado TB, Abdi S, Levasseur A, Abrahão JS, La Scola B. 2018. Morphologic and genomic analyses of new isolates reveal a second lineage of cedratviruses. J Virol 92: e00372-18. https://doi.org/10.1128/JVI.00372-18.

Editor Rozanne M. Sandri-Goldin, University of California, Irvine

Address correspondence to Bernard La Scola, bernard.la-scola@univ-amu.fr.

R.A.L.R. and J.A. contributed equally to this work.

lengths of viruses, exhibiting a difference of 4 orders of magnitude in the former and ranging from 1.2 kbp to 2,500 kbp in the latter (3). Despite these differences, the virion sizes and genome lengths of viruses display an allometric relationship independent of the type of capsid or genetic material, the ebolaviruses (ssRNA−) being the only exception to this scaling law described to date (3). This relationship has also been observed for some giant viruses such as mimivirus and pandoravirus (both dsDNA).

The giant viruses were discovered at the beginning of the last decade with the description of mimiviruses, revealing a new world within the virosphere (4, 5). These viruses replicate in free-living amoebas of the *Acanthamoeba* genus, although other protists have been associated with giant viruses, such as *Cafeteria roenbergensis* (6) and *Bodo saltans* (7). The discovery of viruses with gigantic particles (>500 nm) and the presence of genes considered hallmarks of the cellular world (e.g., those encoding aminoacyl-tRNA synthetases and translation factors) broke many paradigms of classical virology and reopened a hot debate about the origin and evolution of viruses (8–13). In subsequent years, other giant viruses have been isolated and characterized (14, 15), thus expanding the group of nucleocytoplasmic large DNA viruses (NCLDVs)—the proposed "Megavirales" order (16).

Among these new viruses, the pithoviruses drew attention due to their huge size (~1.5 μm) and relatively "small" genomes (~610 kbp) (17, 18), which suggest that those viruses do not fit the scaling law observed for other viruses, although no analysis has been performed in this regard to date. The same case would appear to apply to the cedratviruses, a new group of recently described viruses, with only two members characterized thus far, *Cedratvirus A11* (19), and *Cedratvirus lausannensis* (20). These viruses have an ovoid particle of about 1.0 μm and possess a circular dsDNA genome of ~590 kbp. These viruses are related to the pithoviruses but have two corks in the viral particle, one at either extremity, rather than the single one displayed by pithoviruses. Recently, it has been reported that some pithovirus particles can have complex alternative shapes and sometimes have two corks, as observed for cedratviruses (21). These viruses replicate in *Acanthamoeba* sp., entering the cells by phagocytosis. The genome is released through the cork, and an eclipse phase is established, followed by the formation of an electron-lucent viral factory, where a complex morphogenesis occurs (19, 20, 22). It is possible that there is a nuclear phase during the replication of cedratviruses, since the host nucleus remains intact during the viral cycle, although further investigation into this aspect needs to be performed (22). After ~12 h of infection, mature viral particles are released by cell lysis. The real extent of the diversity of cedratviruses is currently unknown, and the discovery of new members of this group could bring valuable information about it.

Here we describe the isolation and biological and genomic analyses of two new cedratviruses, one from samples collected in France and a second from samples collected in Brazil, which have morphological and genomic features distinct from those of the previously known cedratviruses, suggesting the existence of a second lineage among this new group of viruses. The discovery of new cedratviruses in different parts of the world reflects their ubiquity and high diversity and improves our knowledge about these viruses, thus reinforcing the importance of continuing to prospect for and biologically/genomically characterize the giant viruses.

RESULTS

New cedratviruses with different virion sizes and genome lengths. In the search for a better understanding of the diversity of giant viruses in different parts of the world, we isolated two new cedratviruses, named *Cedratvirus Zaza IHUMI* and *Brazilian Cedratvirus IHUMI*. Transmission electron microscopy (TEM) analyses revealed viruses with ovoid particles and with corks at either extremity of the particle (Fig. 1A to C; see Fig. S1 posted at http://www.mediterranee-infection.com/article.php?laref=983&titer=morphological-and-genomic-analyses-of-new-isolates-reveal-a-second-lineage-of-cedratviruses), a singular feature of cedratviruses (19, 20). Unlike pithoviruses, cedratviruses usually have two corks, although some alternative shapes with a single cork can

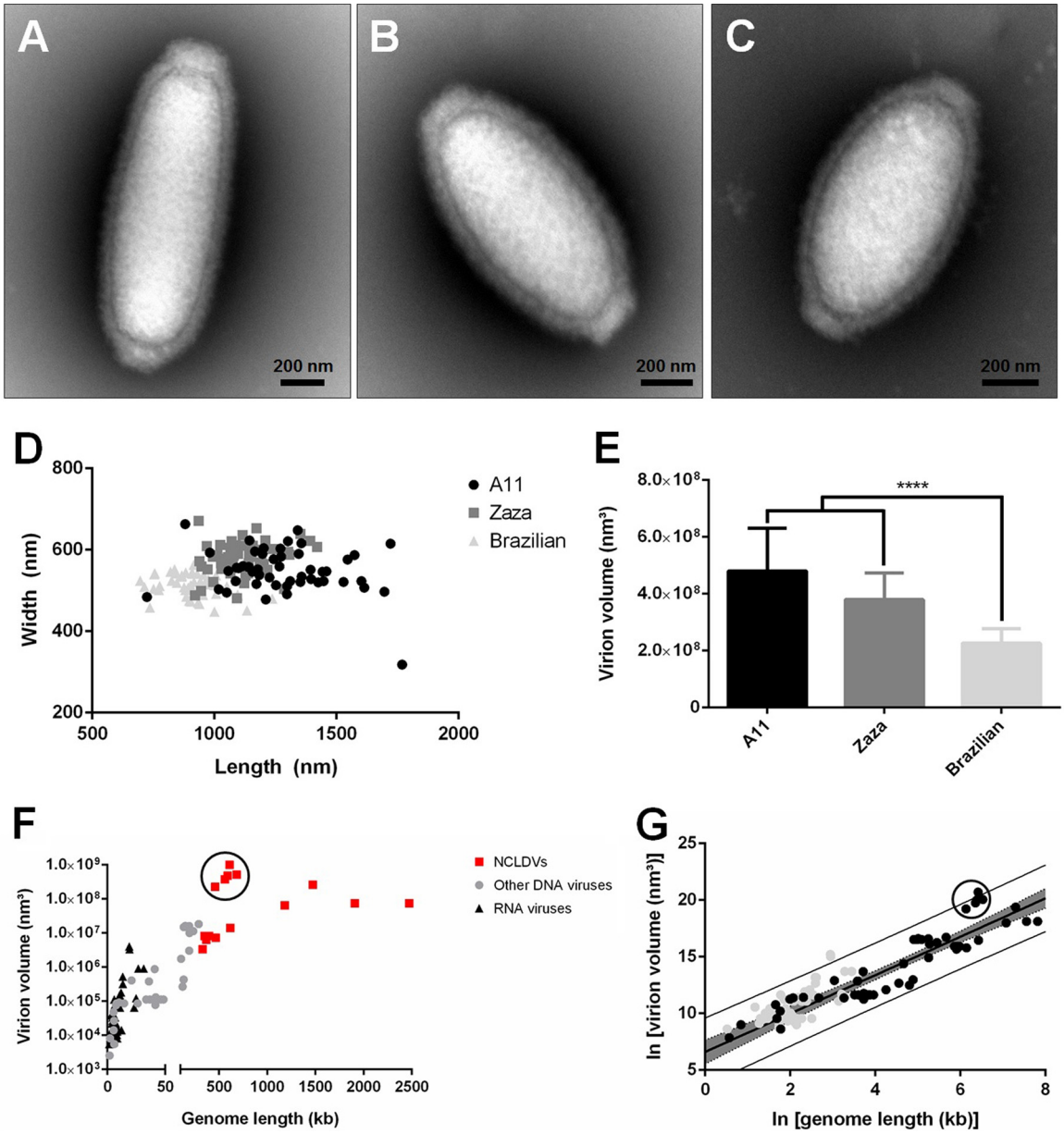


FIG 1 Morphology and volume analysis of new cedratviruses. (A to C) Negative-staining images exhibiting the characteristic ovoid shape and the presence of two corks in the particles of cedratvirus A11, cedratvirus Zaza IHUMI, and Brazilian cedratvirus IHUMI, respectively. Scale bars are indicated on each panel. (D) Length and width of 50 particles of each cedratvirus. Each point represents a single particle analyzed by using ImageJ software. (E) Volumes of different cedratviruses based on the analyses of 50 individual particles, indicating that Brazilian cedratvirus IHUMI has a significantly smaller volume than the other viruses. (F) Relationship between genome length and virion volume for different DNA and RNA viruses. Black circles highlight the pithoviruses and cedratviruses. (G) Relationship between genome length and virion volume for different viruses. The solid black line marks the linear regression between ln-ln-transformed data. The gray area represents the 95% confidence interval for the linear regression line. Black circles highlight the pithoviruses and cedratviruses. The outer gray lines represent the 95% prediction interval, within which we expect 95% of virion volume to lie for a given genome size. ****, $P < 0.0001$ (ANOVA).

be discerned, such as in pithoviruses. Moreover, as recently described for pithoviruses (21), we also observed membranous structures in empty particles in some negative-staining images (see Fig. S2 posted at the URL mentioned above). Cedratvirus Zaza IHUMI particles have a mean size of 1,110 nm (range, 921 to 1,420 nm) in length and 580 nm (range, 481 to 671 nm) in width, values closer to those observed for cedratvirus A11 (1,280 nm by 550 nm), while the Brazilian cedratvirus IHUMI particle is smaller, with the particle displaying a mean size of 910 nm (696 to 1,237 nm) in length and 510 nm

TABLE 1 Main genomic characteristics of known cedratviruses[a]

Virus	Mean particle length × width (nm)	Genome size (bp)	GC content (%)	No. of predicted proteins	Coding density (%)
Cedratvirus A11	1,280 × 550	589,068	42.6	574	78.5
Cedratvirus lausannensis	~1,000 × 500	575,161	42.8	643	83
Cedratvirus Zaza IHUMI	1,110 × 580	560,887	42.7	636	84.3
Brazilian cedratvirus IHUMI	910 × 510	460,038	42.9	533	84.3

[a]All these viruses showed ovoid, double-cork morphology, and none had tRNA.

(448 to 563 nm) in width (Fig. 1D). The difference in the particle size reflects the difference in virion volume, in that the Brazilian Cedratvirus IHUMI has the smallest volume (2.26×10^8 nm^3) among the cedratviruses analyzed, significantly smaller than cedratvirus A11 (4.8×10^8 nm^3) and cedratvirus Zaza IHUMI (3.79×10^8 nm^3) ($P < 0.0001$) (Fig. 1E). Despite these physical differences, the replication cycle of the Brazilian isolate is similar to those previously observed for other cedratviruses, exhibiting the same infection profile (see Fig. S1 posted at the URL mentioned above). The virus enters the host through phagocytosis and releases the capsid content into the cytoplasm, establishing an eclipse phase 2 h after infection. A viral factory is formed in the cytoplasm, where morphogenesis occurs, and mature virions are released 12 h after infection.

In addition to the size of the particles, the new cedratviruses have different genome lengths. Cedratvirus Zaza IHUMI has a genome of 560,887 bp coding for 636 proteins, while the Brazilian cedratvirus IHUMI has a genome of 460,038 bp coding for 533 proteins. Despite the remarkable difference in the length and numbers of predicted open reading frames (ORFs) in the genomes of the new viruses, both exhibit a circular dsDNA genome with the same coding density (84.3%) and very similar G+C contents, 42.7% and 42.9% for cedratvirus Zaza IHUMI and Brazilian cedratvirus IHUMI, respectively (Table 1). It is noteworthy that the Brazilian isolate is the smallest cedratvirus described to date and also has the smallest genome among representatives of this new group of viruses.

Cedratviruses and pithoviruses: exceptions to the allometric scaling law? The fact that the new cedratviruses exhibit different genome lengths led us to analyze the relationship between the genome length and volume size of different NCLDVs, in order to check whether they lie within the prediction interval and are thus in line with the allometric scaling law, as observed for other groups of viruses (3). We calculated the volume of 15 different viruses, including mimiviruses, marseilleviruses, pithoviruses, cedratviruses, faustovirus, kaumoebavirus, pacmanvirus, phycodnavirus, and iridovirus. The volume was calculated by considering the dimensions of viral particles resulting from previous studies using cryo-EM or negative-staining methods (see Table S1 posted at http://www.mediterranee-infection.com/article.php?laref=983&titer=morphological-and-genomic-analyses-of-new-isolates-reveal-a-second-lineage-of-cedratviruses), with the exception of the cedratviruses, for which the volume used in the analysis was the mean volume obtained from the analyses of 50 viral particles using the negative-staining method. Data concerning all other viruses were obtained from previous studies (3).

The volume of the viruses varied by 5 orders of magnitude, with porcine circovirus 1 displaying the smallest volume of the viruses under consideration (2.5×10^3 nm^3), and pithovirus sibericum presenting the largest volume (9.9×10^8 nm^3). Regarding genome length, this ranged from 1.76 kbp (porcine circovirus 1) to 2,474 kbp (pandoravirus salinus) (Fig. 1F). Considering only the volumes of the NCLDVs calculated in this study, volumes ranged from 3.31×10^6 nm^3 (chilo iridescent virus) to 9.9×10^8 nm^3 (pithovirus sibericum) (see Table S1 posted at http://www.mediterranee-infection.com/article.php?laref=983&titer=morphological-and-genomic-analyses-of-new-isolates-reveal-a-second-lineage-of-cedratviruses).

Plotting the new data on NCLDVs alongside other viruses on a log-log scale, the linear relationship is maintained ($P < 0.0001$, $R^2 = 0.83$, slope = 1.58), with values even more stringent than those previously reported (3). The vast majority of viruses fall within the 95% prediction interval, indicating that almost all viruses follow the allo-

metric scaling law for volume size and genome length, i.e., the larger the volume size of a viral particle, the longer the genome enclosure by the virus (Fig. 1G). Curiously, cedratviruses and pithoviruses are at the limit of the prediction level, with pithovirus sibericum actually outside the interval. The same was observed when considering only dsDNA viruses (data not shown). This suggests that the putative "Pithoviridae" family could be the first dsDNA group of viruses that does not conform to the allometric scaling law, along with ebolaviruses (ssRNA−). It is notable that although cedratviruses and pithoviruses appear to be exceptions to this scaling law, this appears to be true only when comparing group of viruses, since a virus with a larger volume (e.g., cedratvirus A11) has a longer genome than does a virus displaying a smaller volume, as verified for the Brazilian cedratvirus IHUMI.

Genome comparison of new cedratviruses. The cedratvirus Zaza IHUMI genome exhibit 636 genes, of which 313 (49.2%) code for proteins with no known function (hypothetical proteins). Of these, three had no hits in all searched databases and were considered ORFans (proteins that were longer than 100 amino acids and with no hits in any database). Regarding Brazilian cedratvirus IHUMI, 269 of its 533 predicted genes (50.5%) have no known function and 11 are considered to be ORFans. Among the ORFs with known functions, the presence of genes related to the metabolism of nucleic acids (e.g., those coding for DNA polymerase, DNA-dependent RNA polymerase, helicases, nucleases, DNA repair proteins) and transcription process (e.g., TFIIB initiation factor, TFIIS elongation factor, viral transcription late factor 3) was observed. Moreover, we identified 76 ankyrin repeat-containing-domain proteins in the cedratvirus Zaza IHUMI genome, while only 42 were observed in the Brazilian cedratvirus IHUMI genome. No tRNA or aminoacyl-tRNA synthetases were detected in the genomes of the new viruses. Regarding the nucleocytoplasmic virus orthologous group (NCVOG) core genes, we found some conserved genes also present in some other NCLDVs, e.g., those encoding a divergent major capsid protein (NCVOG0022), D5 helicase-primase (NCVOG0023), DNA topoisomerase II (NCVOG0037), ribonucleotide reductase (NCVOG0276 and NCVOG1353), and an mRNA capping enzyme (NCVOG1117) similar to that observed for other cedratviruses (19, 20).

Although the gene content does not exhibit significant differences at first glance, the genome organization of the Brazilian isolate is completely different from that observed for other cedratviruses, being totally asyntenic (Fig. 2). The synteny analysis revealed the presence of conserved and aligned blocks between cedratvirus Zaza IHUMI, cedratvirus A11, and cedratvirus lausannensis, while the same blocks are organized in a different orientation in the Brazilian cedratvirus IHUMI genome. Compared to the genome of the other viruses, the genome of the Brazilian isolate exhibits many inversions and rearrangements of blocks throughout its entire length. Such differences in the genomic architecture among similar viruses are observed among different lineages of mimiviruses (23) and marseilleviruses (24), which led us to consider the existence of a second lineage of cedratviruses, with Brazilian cedratvirus IHUMI being its first member.

In addition, the Brazilian cedratvirus IHUMI amino acid sequences showed lower identity than other cedratviruses (Fig. 3). The orthologous genes of the Brazilian isolate have a mean identity of 73.48% compared to cedratvirus A11, 73.6% compared to cedratvirus lausannensis, and 73.56% compared to cedratvirus Zaza IHUMI (Fig. 3A to C). In contrast, when we compared the orthologous genes from other cedratviruses to one another, we observed an amino acid identity higher than 90%, reaching 95.76% between cedratvirus lausannensis and the new isolate, cedratvirus Zaza IHUMI (Fig. 3D to F). Therefore, not only is the genomic architecture between the Brazilian isolate and the other viruses different, but also amino acid identity is considerably different, reinforcing the existence of a new lineage among the group of cedratviruses.

Pangenome and phylogenetic analyses of cedratviruses. The pangenome analysis of the cedratviruses isolated thus far revealed an increase in the pangenome content with the addition of a gene repertoire by way of the new viruses described in

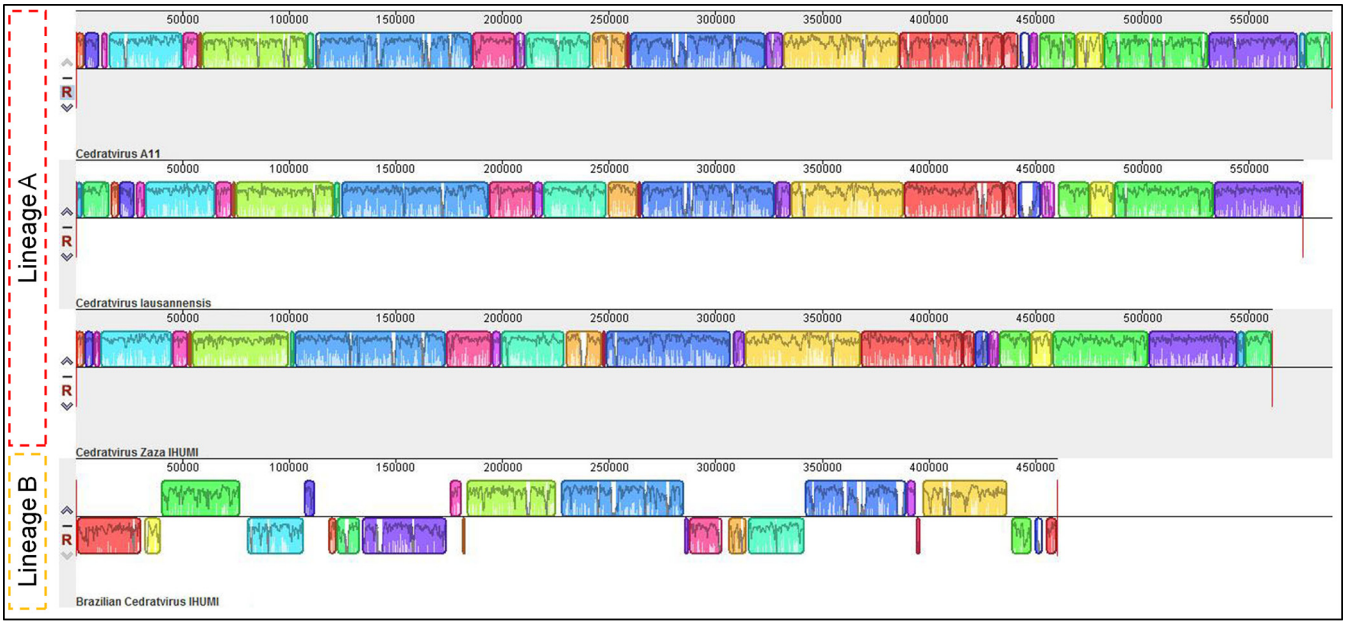


FIG 2 Genome synteny analysis. Schematic genome alignment diagram obtained using the Mauve software package. The analysis was performed using the genome of cedratvirus A11 (NC_032108.1) and cedratvirus lausannensis (LT907979.1), besides the genome sequences of the new isolates. The blocks illustrated above the *x* axis are in the positive strand (forward sense), while blocks below the *x* axis are in the negative strand (reverse sense).

this study. A total of 2,386 proteins were grouped into 821 clusters of orthologous genes (COGs) (Fig. 4), including 613 COGs comprising genes for at least two proteins of different cedratviruses. The size of the pangenome content displayed a continuous increase with the addition of the two new viruses, including an increase of 61 new COGs with the addition of the Brazilian cedratvirus IHUMI, even though this virus presented a genome coding for fewer proteins than the other viruses. Furthermore, it is the virus that presents the greatest number of unique COGs (numbering 72), while the others present only 59 (cedratvirus lausannensis), 47 (cedratvirus Zaza IHUMI), and 30 (cedratvirus A11) unique COGs (see Fig. S3 posted at http://www.mediterranee-infection.com/article.php?laref=983&titer=morphological-and-genomic-analyses-of-new-isolates-reveal-a-second-lineage-of-cedratviruses). However, the most remarkable point is the existence of a break in the curve of the core genome content when genes of the Brazilian isolate are added (−102), reaching a total of 386 COGs for this proposed viral genus (Fig. 4). These data suggest that different lineages of cedratviruses could contribute to a slight increase in the pangenome and could share a reduced core gene set.

To better understand the evolutionary relationship between the new cedratviruses and other members of the proposed Megavirales order, we performed phylogenetic analyses based on different NCLDV genes (NCVOGs) including those coding for the family B DNA polymerase (NCVOG0038) (Fig. 5), the major capsid protein (NCVOG0022), the DNA-dependent RNA polymerase subunit 1 (NCVOG0274), and the VV-A18 helicase (NCVOG0508) (see Fig. S4 posted at http://www.mediterranee-infection.com/article.php?laref=983&titer=morphological-and-genomic-analyses-of-new-isolates-reveal-a-second-lineage-of-cedratviruses). Moreover, we performed additional phylogenetic analyses using the D6/D11 helicase (NCVOG0031), DNA repair exonuclease (NCVOG0308), Flap endonuclease (NCVOG1060), and ATP-dependent DNA ligase (NCVOG0034), focusing on the cedratviruses and closer viral groups, i.e., marseilleviruses and irido/ascoviruses (see Fig. S5 posted at the URL mentioned above). Phylogenetic trees recurrently clustered the new isolates alongside previously described cedratviruses, pithoviruses, and orpheovirus. Furthermore, all trees based on the core genes showed the cedratvirus Zaza IHUMI as being closer to cedratvirus lausannensis and cedratvirus A11 and the Brazilian isolate being in a branch distant

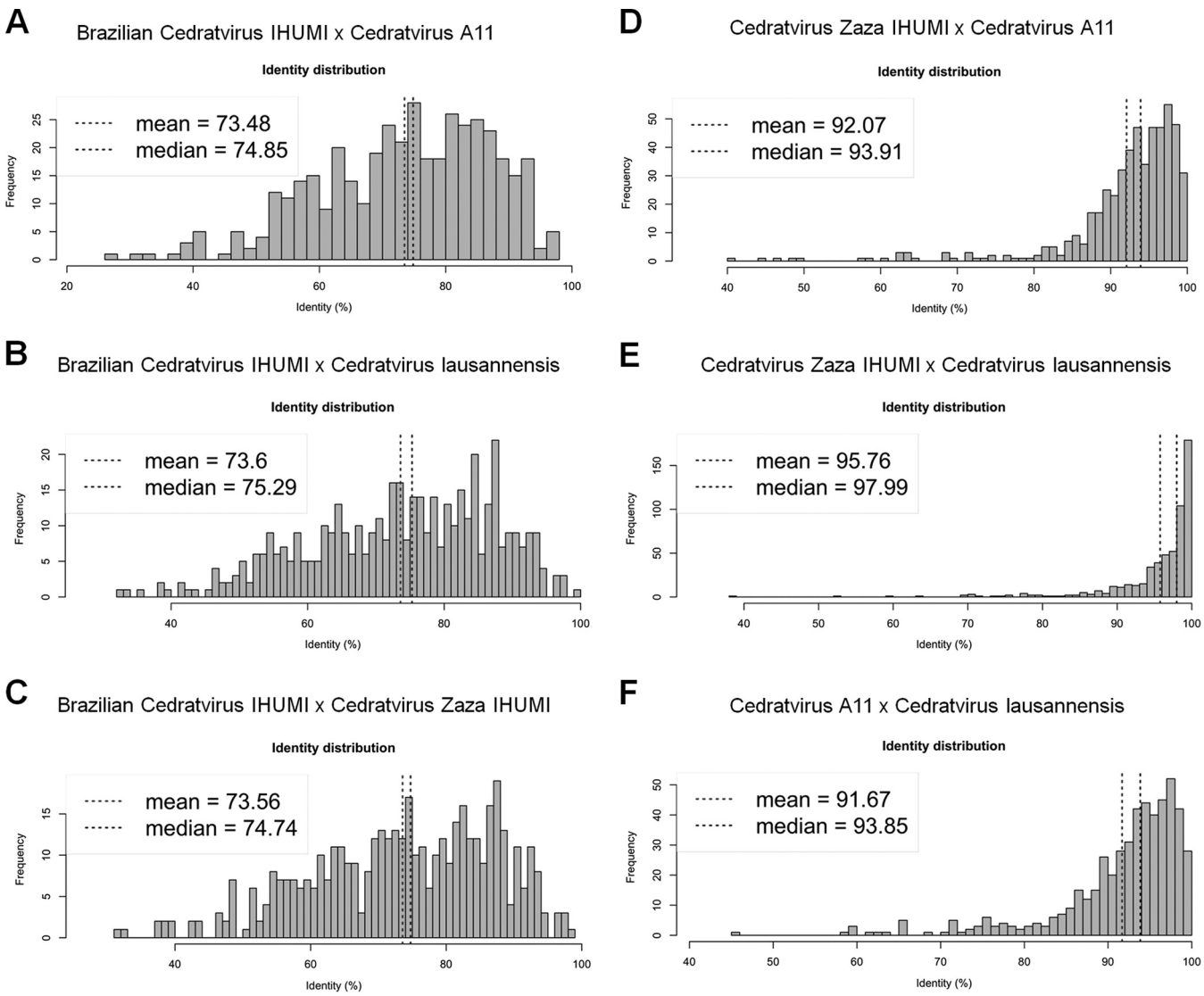


FIG 3 Average amino acid identity. In this analysis, estimates were reached using both best hits (one-way AAI) and reciprocal best hits (two-way AAI) between two data sets of proteins from different cedratviruses. Plots A to C demonstrate the amino acid comparisons between Brazilian cedratvirus IHUMI and other cedratviruses; plots D and E compare cedratvirus Zaza IHUMI and previously known cedratviruses; plot F compares cedratvirus A11 and cedratvirus lausannensis.

from the other cedratviruses with a bootstrap value of >90, corroborating the existence of a new lineage among cedratviruses. Finally, the putative "Pithoviridae" family is clustered along with marseilleviruses or irido/ascoviruses depending of the gene used, the tree topology not being always congruent. An in-depth phylogenetic analysis must be performed to better establish the phylogenetic relationship among these groups of giant viruses.

DISCUSSION

The isolation of new giant viruses associated with biological and genomic analyses has significantly contributed to broadening our understanding of the diversity, ecology, and evolution of this complex group of viruses. The discovery of pithoviruses (17, 18) and cedratviruses (19, 20) drew particular attention, since these viruses exhibit very large particles constraining relatively short genomes, forming a putative novel viral family among the group of NCLDVs. In this study, we describe the isolation and the biological and genomic analyses of two new members of this group, providing new insights into the biodiversity and evolution of these viruses.

The analyses performed in this study revealed two new viruses with significant

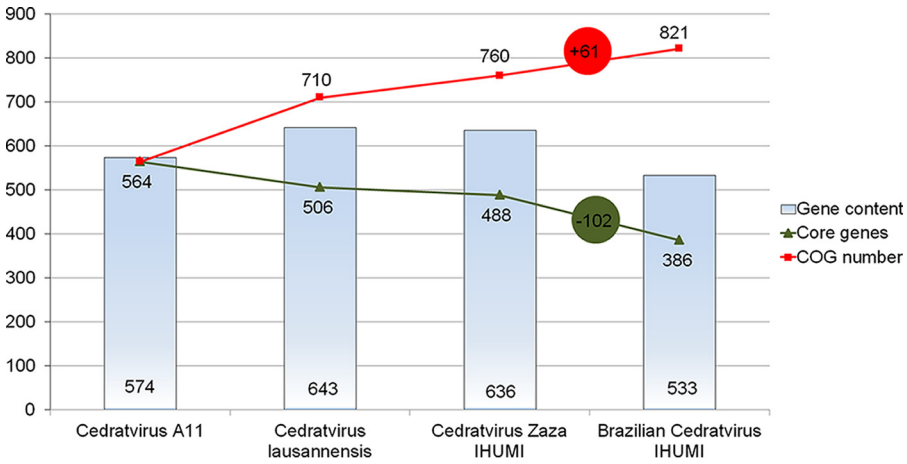


FIG 4 Pangenome (red line) and core genome (green line) sizes of cedratviruses. Numbers at the base of the column refer to the number of genes carried by each virus strain. Numbers at line nodes represent the cumulative COG numbers after the inclusion of a new genome. Numbers in (red and green) circles demonstrate the variation of COGs after the inclusion of the Brazilian cedratvirus IHUMI (proposed new lineage).

structural differences, both physical and genomic. Although the cedratvirus Zaza IHUMI exhibits a particle size and genome length similar to those of other cedratviruses that have been described, analysis of the Brazilian cedratvirus IHUMI revealed a virus with particles that were smaller (~20%) than those of the other viruses of the same group and a considerably smaller genome. By analyzing the relationship between the virion volume and the genome length of viruses, including those from different groups of giant viruses, we noticed that the majority of viruses fall into the allometric scaling law and, curiously, the pithoviruses and cedratviruses are at the limits of the prediction interval. This suggests that these viruses might be exceptions to this scaling law. Since we considered only data from comparable imaging methods (i.e., cryo-EM and negative staining [25, 26]) to calculate the volumes of giant viruses, only a few viruses were analyzed. It is possible that with new, forthcoming structural data on viruses, particularly on giant viruses, it may be discovered that the pithoviruses and cedratviruses definitively fall outside the prediction interval. Indeed, when the virion size data for other amoebal giant viruses (e.g., mimiviruses and marseilleviruses) from TEM images were considered in our analysis, the members of the putative "Pithoviridae" did not fit with the allometric scaling law (data not shown). It is notable that, along with ebolaviruses (*Filoviridae* family), the members of the putative "Pithoviridae" family are the only known viruses that display a massive particle but a "small" genome. Such features raise important questions about what is inside these viral particles. A recent study comparing the internal density of pithoviruses' and mimiviruses' particles demonstrated that the former viruses have three-quarters of the internal density of the latter, suggesting that the pithoviruses may carry macromolecules other than nucleic acids inside the particles (21). The same would appear to be the case for the cedratviruses, but further studies are needed to define exactly which macromolecules could be carried by those viruses.

The fact that the Brazilian isolate has a smaller genome is equally curious. Similar to other cedratviruses, this new virus exhibits only a few repeat zones throughout the genome (data not shown), and we identified the presence of genes also present in other cedratviruses, such as those coding for polymerases, helicases, nucleases, etc. Among the genes with known function, we noticed differences mainly in the quantity of those coding for proteins containing repeat domains, especially coding for ankyrin repeat motifs, as Brazilian cedratvirus IHMU (a virus with a smaller genome) has fewer genes of these category than do other cedratviruses. This is in

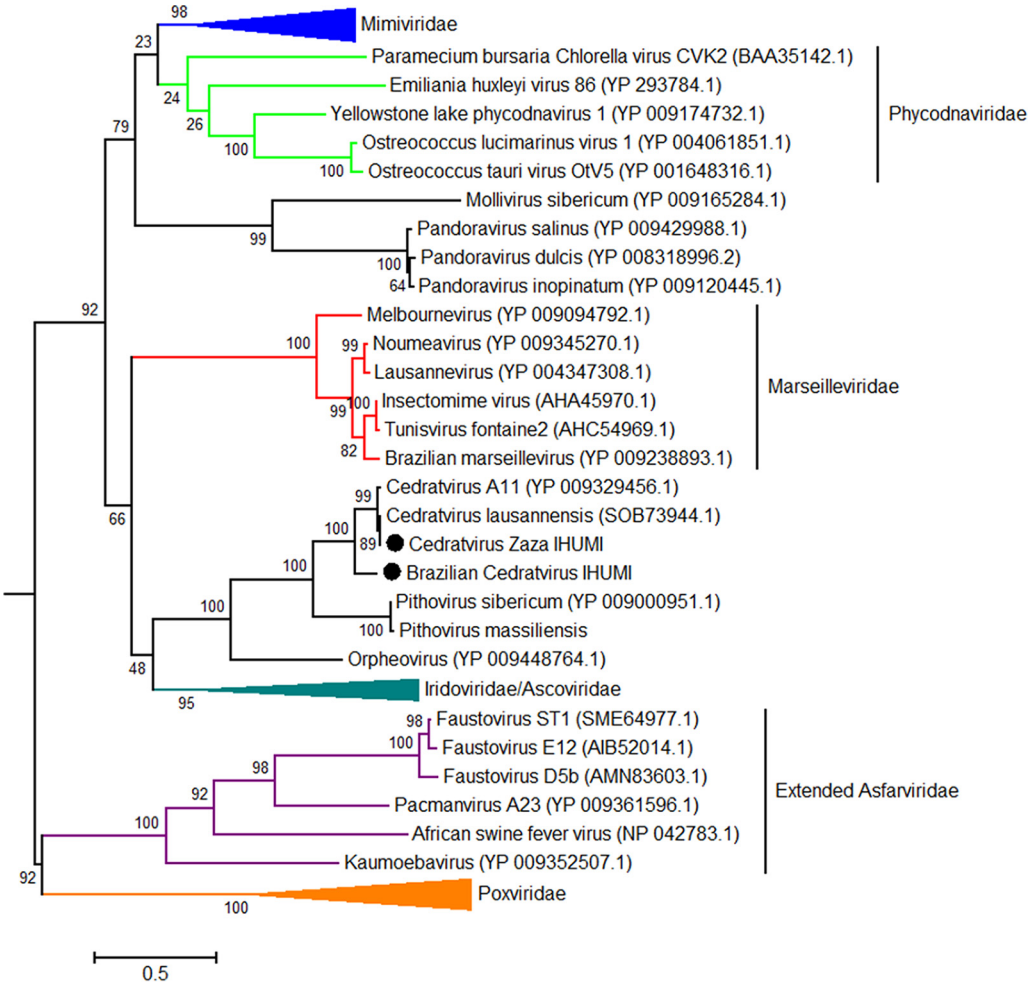


FIG 5 Phylogenetic tree based on DNA polymerase B amino acid sequences of nucleocytoplasmic large DNA viruses (NCLDVs). The tree was constructed using MEGA version 6.0, applying the maximum likelihood method and the JTT model of evolution with 1,000 bootstrap replicates. Colors indicate viral families: blue was used for *Mimiviridae*; green for *Phycodnaviridae*, red for *Marseilleviridae*, navy blue for *Iridoviridae/Ascoviridae*, purple for extended *Asfarviridae*, and orange for *Poxviridae*. The new cedratviruses are highlighted with black circles. The scale bar indicates the rate of evolution.

accordance with the recent analysis conducted by Shukla and colleagues, wherein they demonstrated that in different groups of giant viruses infecting amoebae, the quantity of this class of genes is proportional to the length of the genome (27), which has also been observed for some intracellular bacteria (28). These analyses also seem to apply to viruses within the same group, such as the cedratviruses described here. Taking this into account, it is possible that the Brazilian cedratvirus IHUMI underwent different selective pressures, thus contributing to a different evolutionary history. This would be in accordance with our proposed new lineage within the cedratviruses. Such a proposal is supported by the observation of a completely different genomic architecture between the Brazilian cedratvirus IHUMI and the other viruses, in addition to a significant difference in the amino acid identity of orthologous genes, similar to that observed for members of the *Mimiviridae* and *Marseilleviridae* families (23, 24). In addition, this virus has more exclusive COGs and contributes to an increase in the pangenome with 61 new COGs and, even more strikingly, with the reduction of the core genome by 102 COGs. Finally, phylogenetic analyses based on different core genes of giant viruses clearly clustered the Brazilian isolate in a separate branch from other cedratviruses, therefore reinforcing the existence of a lineage "B" among this new group of viruses.

It is still premature to dive deep into the evolutionary history of cedratviruses, but it is possible that they underwent an accordion-like model of evolution as observed for other giant viruses (29), although new analysis must be performed to confirm this hypothesis. In any case, it is clear that this new, expanding group of viruses deserves attention, and new structural and evolutionary analyses could help to solve some unanswered questions around them.

MATERIALS AND METHODS

Virus isolation, production, and purification. Two novel cedratviruses were isolated by the coculture method as previously described (30), one from an *Alpova* sp. (*Basidiomycota*, *Paxillaceae* family) homogenate in sterile distilled water collected near Toulon, France, and the other from a water sample supplemented with bio-floc, collected in Belo Horizonte, Brazil. The former virus was isolated at the Institut-Hospitalo-Universitaire (IHU) Méditerrannée Infection at Marseille, France, and was named *Cedratvirus Zaza IHUMI*, while the second was isolated in the Laboratürio de Vírus of UFMG at Belo Horizonte, Brazil. The Brazilian isolate was then sent to IHU for further production, genome sequencing, and analysis and was given the name of *Brazilian Cedratvirus IHUMI*. For multiplication of the viruses, *Acanthamoeba castellanii* (strain Neff [ATCC 30010]) was cultured in a 150-cm^2 cell culture flask with 50 ml of a peptone-yeast extract-glucose (PYG) medium at 28°C. When the flasks contained a fresh monolayer of *A. castellanii*, the PYG medium was replaced by starvation medium (TS). The amoebas were then infected with the isolated virus, and the flasks were kept at 30°C for 72 h. The cell lysates were then collected and centrifuged at 400 $\times$ *g* for 10 min to remove amoeba debris. The supernatants were then centrifuged at 6,500 $\times$ *g* for 15 min at 4°C, and the pellets were suspended in Page's amoeba saline (PAS) solution. This process was repeated twice. The pellets were suspended in 3 ml of phosphate-buffered saline (PBS) solution, added to a sucrose cushion (50%), and centrifuged at 16,000 $\times$ *g* for 15 min at 4°C. The final pellets were suspended in PAS solution.

Characterization of the replicative cycle. In order to study possible differences in the replicative cycle of Brazilian cedratvirus IHUMI, ultrathin sections of infected amoebas were evaluated under TEM and a comparative one-step growth curve was performed. For the microscopy analysis, *A. castellanii* cells were infected with Brazilian cedratvirus IHUMI at a multiplicity of infection (MOI) of 0.01 for 24 h in PYG medium (asynchronous cycle). The cells were then collected and fixed with 2.5% glutaraldehyde in a 0.1 M sodium phosphate buffer for 1 h at room temperature. The amoebas were postfixed with 2% osmium tetroxide and embedded in Epon resin. Ultrathin sections were then analyzed under transmission electron microscopy (Spirit Biotwin FEI, 120 kV). For the one-step growth curves, *A. castellanii* cells were infected with different cedratviruses at an MOI of 10 in TS medium in 24-well microplates. After 30 min of adsorption, the inoculum was removed, and fresh medium was added. The cell and supernatant were collected at different time points and further titrated using the endpoint method (31). The experiment was performed in duplicate.

Virus particle morphometry and volume calculation. For particle morphometry, negative staining was performed on the fixed supernatant from coculture. A total of 5 μl was deposited onto the glow-discharged grid for 20 min at room temperature. The dried grid was contrasted with a small drop of 1% ammonium molybdate for 5 s, and the grid was then observed using a Tecnai G20 electron microscope (FEI, Germany) operated at 200 kV. At least 50 particles of each virus were analyzed using ImageJ software (32). For the volume calculation of cedratvirus particles, we employed the formula for spheroid particles as previously described for ovoid viruses (3), $V = 4/3 \times \pi a^2 c$, where V is the volume of the viral particle, a is the equatorial radius of the spheroid, c is the distance from the center to the pole along the symmetry axis, and π is a constant. The data used for the volume calculation of other NCLDVs were obtained from previous publications, considering data only from cryo-electron microscopy or transmission electron microscopy negative-staining data (see Table S1 posted at http://www.mediterranee-infection.com/article.php?laref=983&titer=morphological-and-genomic-analyses-of-new-isolates-reveal-a-second-lineage-of-cedratviruses). For icosahedral viruses, we used the formula for spherical particles, $V = 4/3 \times \pi r^3$, employing a strategy described elsewhere (3), where r is one-half of the diameter of the virus capsid. For other viruses, we kept the volume data previously calculated by Cui and colleagues (3).

Statistical analysis. We used analysis of variance (ANOVA) to compare the virion volumes of different cedratviruses and linear regression between the natural logarithm of genome length and the natural logarithm of virion volume to test whether the allometric relationship previously described for other viruses (3) also applied to giant viruses of amoebas, which had not previously been evaluated. The statistical analyses were performed by using GraphPad Prism 6.0.

DNA extraction and genome sequencing and assembly. The genomes of the new cedratviruses were extracted using the automated EZ1 virus minikit v.2 (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. DNA was quantified using a Qubit assay with the high-sensitivity kit (Life Technologies, Carlsbad, CA, USA) to 30.3 ng/μl (cedratvirus Zaza IHUMI) and 16 ng/μl (Brazilian cedratvirus IHUMI). DNA was sequenced using MiSeq Technology (Illumina Inc., San Diego, CA, USA) with the paired-end application. DNA was barcoded in order to be mixed with other projects for the Nextera XT DNA sample prep kit (Illumina).

To prepare the paired-end library, dilution was performed to yield 1.0 ng of each genome as input. The "tagmentation" step fragmented and tagged the DNA. Limited-cycle PCR amplification (12 cycles) then completed the tag adapters and introduced dual-index barcodes. The library profile was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies Inc., Santa Clara, CA, USA) with a DNA high-

sensitivity LabChip, and the fragment size was estimated to 1.5 kb. After purification on AMPure XP beads (Beckman Coulter Inc., Fullerton, CA, USA), the libraries were then normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled for sequencing on the MiSeq. Automated cluster generation and paired-end sequencing with dual index reads were performed in a single 39-hour run in 2× 250 bp. A total of 2.8 Gb of information was obtained from a 277,000/mm² cluster density in the first run with a cluster passing quality control filters of 98.2% (5,333,000 passed filtered clusters). Within this run, the index representation for cedratvirus Zaza IHUMI was determined to be 2.18%. The 149,880 paired-end reads were trimmed and filtered according to the read qualities. Additionally, a total of 7.5 Gb of information was obtained from an 802,000/mm² cluster density in the second run with a cluster passing quality control filters of 96.4% (14,444,000 clusters). Within this run, the index representation for Brazilian cedratvirus IHUMI was determined to be 8.75%. The 1,264,356 paired-end reads were filtered according to the read qualities.

In addition, a run was performed through the MinIon Oxford Nanopore for the Brazilian isolate. The Oxford Nanopore approach was performed on 1D genomic DNA sequencing for the MinIon device using the SQK-LSK108 kit. A library was constructed from 1.5 μg genomic DNA without fragmentation and end repair. Adapters were ligated to both ends of the genomic DNA. After purification on AMPure XP beads (Beckman Coulter Inc., Fullerton, CA, USA), the library was quantified using a Qubit assay with the high-sensitivity kit (Life Technologies, Carlsbad, CA, USA). An amount of 136.8 ng, adapted and tethered as a library, was loaded on the flow cell via the SpotON port. A total of 1,110 active pores were detected for sequencing, and the WIMP workflow was chosen for live bioinformatic analysis. After 4 h and 40 min of run time, the EPI2ME software led to 6,299 classified reads of the Brazilian cedratvirus IHUMI of the 98,601 analyzed reads, with an average length of 2.6 kb.

The sequence reads were assembled *de novo* using the CLC Genomics Workbench v7.5 (http://www.clcbio.com/blog/clc-genomics-workbench-7-5/) for the cedratvirus Zaza IHUMI and hybridSPAdes (33) for the Brazilian cedratvirus IHUMI.

Study of genome organization and genome annotation. Open reading frames were predicted by GeneMarkS (34), and the draft genomes were deposited in NCBI. The tRNA genes were searched using the tRNAscan-SE and ARAGORN software (35, 36). Predicted proteins of fewer than 50 amino acids in length were discarded. A BLASTp search against the NCBI nonredundant (nr) protein sequence database was performed on 5 January 2018. Homology was considered significant if the E value was lower than 1×10^{-3}. A BLASTp search was also computed with the same parameters against the clusters of orthologous groups (COGs) of proteins of the nucleocytoplasmic large DNA virus (known as NCVOGs) (37). In addition, we searched for conserved domains and putative functions of predicted proteins using the online InterProScan software, version 66.0 (https://www.ebi.ac.uk/interpro/search/sequence-search). In addition, predicted ORFs ranging from 50 to 99 amino acids were submitted for tridimensional folding analyses using Phyre2 (38). Proteins ranging from 50 to 99 amino acids in length were discarded if they exhibited no hits either against the BLASTp or against the InterProScan searches or if they exhibited abnormal folding as modeled by Phyre2. Those proteins that were longer than 100 amino acids and with no hits in any database were kept and referred to as ORFans. Finally, the genome annotation was manually revised and curated.

Comparative genomic and pangenome analysis. The genome synteny among cedratviruses was checked using the Mauve program (39) with default parameters. The Proteinortho tool (40) was used to identify orthologous gene sequences based on the reciprocal best hit shared by cedratviruses using an amino acid sequence identity of 30% and sequence coverage of 60% as thresholds. The average amino acid identity (AAI) calculator tool (41) was used to compare identity between orthologous genes from cedratvirus isolates. To estimate the size of the pangenome, their predicted proteins were clustered using the Proteinortho tool (40), applying the same criteria as those given above. We also described pangenome and core genes size variation by stepwise inclusion of each new virus annotation in the pairwise comparisons of the gene contents of the available cedratvirus genome sequences.

Phylogenetic analysis. Phylogenetic reconstructions were based on individual alignment of distinct genes, namely, those encoding the DNA polymerase B family, the DNA-dependent RNA polymerase subunit 1, the VV-A18 helicase, the major capsid protein, the D6/D11 helicase, the Flap endonuclease, the ATP-dependent DNA ligase, and the DNA repair exonuclease. Amino acid sequences were aligned using the Muscle software (42). Phylogenetic trees were built using the MEGA6 software (43), the Jones-Taylor-Thornton (JTT) model for amino acid substitution, and the maximum likelihood method with 1,000 bootstrap replicates.

Accession number(s). Sequences for the draft genomes were deposited in NCBI under the accession numbers LT994651 (Brazilian cedratvirus IHUMI) and LT994652 (cedratvirus Zaza IHUMI).

ACKNOWLEDGMENTS

We thank our colleagues at the Laboratório de Vírus and Center of Microscopy of UFMG for their excellent technical support. Thanks also go to Enora Tomei and Caroline Michele for genome sequencing and to Emeline Baptiste for genome submission.

This work was funded by CNPq, FAPEMIG, CAPES-COFECUB. J.S.A. is a CNPq researcher. This work was also supported by the French Government under the "Investissements d'avenir" (Investments for the Future) program managed by the Agence Nationale de la Recherche (ANR, French National Research Agency) (reference: Méditerranée Infection 10-IAHU-03).

R.A.L.R. and J.A. performed genomic analyses and wrote the paper; R.A.L.R. performed morphological and phylogenetic analyses; J.A., S.A., A.C.S.P.A., and T.B.M. performed virus isolation and initial biological characterization; A.L., J.S.A., and B.L.S. conceived the study and revised the manuscript.

REFERENCES

1. Suttle CA. 2005. Viruses in the sea. Nature 437:356–361. https://doi.org/10.1038/nature04160.
2. Hulo C, De Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, Le Mercier P. 2011. ViralZone: a knowledge resource to understand virus diversity. Nucleic Acids Res 39(Database issue):D576–D582. https://doi.org/10.1093/nar/gkq901.
3. Cui J, Schlub TE, Holmes EC. 2014. An allometric relationship between the genome length and virion volume of viruses. J Virol 88:6403–6410. https://doi.org/10.1128/JVI.00362-14.
4. 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. https://doi.org/10.1126/science.1081867.
5. Colson P, La Scola B, Levasseur A, Caetano-Anollés G, Raoult D. 2017. Mimivirus: leading the way in the discovery of giant viruses of amoebae. Nat Rev Microbiol 15:243–254. https://doi.org/10.1038/nrmicro.2016.197.
6. Fischer MG, Allen MJ, Wilson WH, Suttle CA. 2010. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc Natl Acad Sci U S A 107:19508–19513. https://doi.org/10.1073/pnas.1007615107.
7. Deeg CM, Chow C-ET, Suttle CA. 2018. The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea. Elife 7:e33014. https://doi.org/10.7554/eLife.33014.
8. Raoult D, Forterre P. 2008. Redefining viruses: lessons from Mimivirus. Nat Rev Microbiol 6:315–319. https://doi.org/10.1038/nrmicro1858.
9. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie J-M. 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306:1344–1350. https://doi.org/10.1126/science.1101485.
10. Moreira D, López-García P. 2009. Ten reasons to exclude viruses from the tree of life. Nat Rev Microbiol 7:306–311. https://doi.org/10.1038/nrmicro2108.
11. Raoult D. 2009. There is no such thing as a tree of life (and of course viruses are out!). Nat Rev Microbiol 7:615. https://doi.org/10.1038/nrmicro2108-c6.
12. Koonin EV, Krupovic M, Yutin N. 2015. Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. Ann N Y Acad Sci 1341:10–24. https://doi.org/10.1111/nyas.12728.
13. Nasir A, Caetano-Anollés G. 2015. A phylogenomic data-driven exploration of viral origins and evolution. Sci Adv 1:e1500527. https://doi.org/10.1126/sciadv.1500527.
14. Abergel C, Legendre M, Claverie JM. 2015. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. FEMS Microbiol Rev 39:779–796. https://doi.org/10.1093/femsre/fuv037.
15. Colson P, La Scola B, Raoult D. 2017. Giant viruses of amoebae: a journey through innovative research and paradigm changes. Annu Rev Virol 4:61–85. https://doi.org/10.1146/annurev-virology-101416-041816.
16. Colson P, De Lamballerie X, Fournous G, Raoult D. 2012. Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. Intervirology 55:321–332. https://doi.org/10.1159/000336562.
17. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, Lescot M, Poirot O, Bertaux L, Bruley C, Coute 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.
18. Levasseur A, Andreani J, Delerce J, Bou Khalil J, Robert C, La Scola B, Raoult D. 2016. Comparison of a modern and fossil Pithovirus reveals its genetic conservation and evolution. Genome Biol Evol 8:2333–2339. https://doi.org/10.1093/gbe/evw153.
19. Andreani J, Aherfi S, Khalil JYB, Di Pinto F, Bitam I, Raoult D, Colson P, La Scola B. 2016. Cedratvirus, a double-cork structured giant virus, is a distant relative of pithoviruses. Viruses 8:1–11. https://doi.org/10.3390/v8110300.
20. Bertelli C, Mueller L, Thomas V, Pillonel T, Jacquier N, Greub G. 2017. Cedratvirus lausannensis—digging into Pithoviridae diversity. Environ Microbiol 19:4022–4034. https://doi.org/10.1111/1462-2920.13813.
21. Okamoto K, Miyazaki N, Song C, Maia FRNC, Reddy HKN, Abergel C, Claverie JM, Hajdu J, Svenda M, Murata K. 2017. Structural variability and complexity of the giant Pithovirus sibericum particle revealed by high-voltage electron cryo-tomography and energy-filtered electron cryo-microscopy. Sci Rep 7:13291. https://doi.org/10.1038/s41598-017-13390-4.
22. Silva LKDS, Andrade ACDSP, Dornas FP, Rodrigues RAL, Arantes T, Kroon EG, Bonjardim CA, Abrahaõ JS. 2018. Cedratvirus getuliensis replication cycle: an in-depth morphological analysis. Sci Rep 8:4000. https://doi.org/10.1038/s41598-018-22398-3.
23. Assis FL, Franco-Luiz APM, dos Santos RN, Campos FS, Dornas FP, Borato PVM, Franco AC, Abrahao JS, Colson P, La Scola B. 2017. Genome characterization of the first mimiviruses of lineage C isolated in Brazil. Front Microbiol 8:2562. https://doi.org/10.3389/fmicb.2017.02562.
24. 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(3):76. https://doi.org/10.3390/v8030076.
25. Hoenger A, Aebi U. 1996. 3-D reconstructions from ice-embedded and negatively stained biomacromolecular assemblies: a critical comparison. J Struct Biol 117:99–116. https://doi.org/10.1006/jsbi.1996.0075.
26. De Carlo S, Harris JR. 2011. Negative staining and cryo-negative staining of macromolecules and viruses for TEM. Micron 42:117–31. https://doi.org/10.1016/j.micron.2010.06.003.
27. Shukla A, Chatterjee A, Kondabagil K. 2018. The number of genes encoding repeat domain-containing proteins positively correlates with genome size in amoebal giant viruses. Virus Evol 4:1–11. https://doi.org/10.1093/ve/vex039.
28. Moliner C, Fournier PE, Raoult D. 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS Microbiol Rev 34:281–94. https://doi.org/10.1111/j.1574-6976.2009.00209.x.
29. Filée J. 2015. Genomic comparison of closely related giant viruses supports an accordion-like model of evolution. Front Microbiol 6:593. https://doi.org/10.3389/fmicb.2015.00593.
30. Khalil JYB, Robert S, Reteno DG, Andreani J, Raoult D, La Scola B. 2016. High-throughput isolation of giant viruses in liquid medium using automated flow cytometry and fluorescence staining. Front Microbiol 7:26. https://doi.org/10.3389/fmicb.2016.00026.
31. Reed LJ, Muench H. 1938. A simple method of estimating fifty per cent endpoints. Am J Hyg 27:493–497.
32. Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675. https://doi.org/10.1038/nmeth.2089.
33. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. HybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics 32:1009–1015. https://doi.org/10.1093/bioinformatics/btv688.
34. Besemer J. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29:2607–2618.
35. Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33(Web Server issue):W686–W689.
36. 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.
37. Yutin N, Wolf YI, Raoult D, Koonin EV. 2009. Eukaryotic large nucleocytoplasmic 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.
38. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053.
39. 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.

40. 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.
41. Rodriguez-R LM, Konstantinidis KT. 2014. Bypassing cultivation to identify bacterial species. Microbe Mag 9:111–118. https://doi.org/10.1128/microbe.9.111.1.
42. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113.
43. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197.

4.7. ARTIGO #7: The analysis of Marseillevirus transcriptome reveals a temporal profile of gene expression

A família *Marseilleviridae* agrupa vírus grandes de DNA pertencentes à proposta ordem “Megavirales”. Estes vírus infectam amebas de vida livre do gênero *Acanthamoeba*, possuem partículas icosaédricas de aproximadamente 250 nm de diâmetro. O sequenciamento do genoma viral revelou uma molécula circular com alto grau de mosaicismo, codificando centenas de genes de origens distintas. Mais da metade dos genes codificados pelos marseillevírus estão acompanhados de um motivo promotor octamérico – AAATATTT – cuja importância da regulação foi biologicamente demonstrada. Entretanto, ainda não há estudos sobre o perfil transcricional destes vírus e nem mesmo se os genes preditos são de fato expressos durante o ciclo de multiplicação viral. Neste trabalho nós realizamos o sequenciamento de RNA (RNA-seq) para caracterizar o perfil de transcrição dos marseillevírus. Células de *A. castellanii* foram infectadas com Marseillevirus marseillevirus, o vírus protótipo da família, e mantidas por diferentes tempos entre 0h e 12h. Após o sequenciamento do RNA nós obtivemos mais de oito milhões de sequências (75-bp) que foram mapeadas contra todos os 457 genes inicialmente preditos. As sequências foram heterogeneamente mapeadas contra os genes, e diferentes análises de agrupamento indicaram a existência de três categorias temporais de expressão gênica: precoce, intermediária e tardia. Este perfil foi validado por ensaios de RT-qPCR utilizando diferentes genes como alvos. Genes de diferentes grupos funcionais são expressos em momentos distintos ao longo do ciclo de infecção. Além disso, nós observamos que o motivo promotor está homogeneamente distribuído ao longo do genoma viral e não está diretamente relacionado com a regulação de nenhuma classe temporal ou funcional específica. Os dados obtidos sugerem que a maquinaria transcricional dos marseillevírus é mais complexa do que inicialmente imaginada e envolve fatores ainda não identificados. Novos estudos são necessários para avançar e obter informações sobre a transcrição destes vírus gigantes.

Este artigo está em fase final de análise e redação e será submetido para publicação após as considerações e sugestões da banca.

Title: The analysis of Marseillevirus transcriptome reveals a temporal profile of gene expression

Running title: Transcriptomic analysis of Marseillevirus

Authors: Rodrigo Araújo Lima Rodrigues[1,2]; Amina Cherif Louazani[2]; Philippe Colson[2,3]; Bernard La Scola[2,3]; Jônatas Santos Abrahão[1*]

Affiliation:

[1]Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[2] Microbes, Evolution, Phylogeny and Infection (MEΦI), Aix-Marseille Université UM63, IRD 198, AP-HM, Marseille, France

[3] Institut Hospitalo-Universitaire (IHU) - Méditerranée Infection, Marseille, France

*Corresponding author

e-mail: jonatas.abrahao@gmail.com

Abstract

Marseilleviruses comprise a family of large double-stranded DNA viruses belonging to the proposed order “Megavirales”. These viruses have a circular genome with a high degree of mosaicism, coding hundreds of genes of distinct origins. Over a half of their genes are associated to an AT-rich putative promoter motif, which was demonstrated to be important for gene regulation. However, the transcriptional profile of marseilleviruses is currently unknown, constituting an important gap to be filled for a better comprehension upon the biology of these viruses. Here we used RNA sequencing technology to get a general view of the transcriptional landscape of marseilleviruses. We generated over eight million 75-bp-long reads robustly mapped to all 457 genes initially predicted for Marseillevirus isolate T19, the prototype strain of the family. These reads were heterogeneously mapped to the genes according to the replicative cycle time point from which they were obtained, and different clustering analyses indicated the existence of three main temporal categories of gene expression: early, intermediate and late. This temporal profile of expression was validated by RT-qPCR assays targeting several genes. Genes belonging to different functional groups exhibited distinct expression levels throughout the infection cycle. Moreover, we observed that the previously predicted AAATATTT promoter motif was not directly related to any of the temporal classes of genes, suggesting that a different machinery regulates viral transcription. This work provides an overview of the transcriptional landscape of marseilleviruses.

Importance

Marseilleviruses, discovered in 2009, have complex genomes containing hundreds of genes from different origins, over a half of them being associated to an AT-rich putative promoter motif. Despite family *Marseilleviridae* has expanded with dozens isolates from several continents and classified in at least five putative different lineages, no transcriptional profile is available for one of its representatives. In this work we used a RNA sequencing approach to reveal a temporal dynamic of gene expression for the prototype isolate of genus *Marseillevirus*. Our data indicate a differential temporal expression of genes from different functional categories throughout the replicative cycle, with some genes expressed at least for 12h. These findings expand our knowledge about the biology of these giant viruses.

Key words: Marseillevirus, giant virus, transcriptomics, RNA-seq, gene expression, promoter motif

Introduction

The family *Marseilleviridae* is a recently established taxon encompassing marseilleviruses, a peculiar group of giant viruses isolated by co-culturing on *Acanthamoeba* spp., mostly from water samples (1, 2). This viral family has been expanding over the last years, with isolates from different regions of the planet, such as Senegal, Tunisia, India, Japan, Australia, Brazil, and New Caledonia (3–9). Along with other giant viruses isolated on amoebae, the marseilleviruses are members of the proposed order “Megavirales”, which comprises the nucleocytoplasmic large DNA viruses (NCLDVs) (10). These viruses replicate in free-living amoebae of genus *Acanthamoeba*. The replication cycle begins with phagocytosis of multiple viral particles or membranous vesicles containing up to thousands of particles within 1h after infection (11). Once within the host cells, the marseilleviruses establish a large viral factory around 3 to 4h post-infection, where genome replication (assisted by host nuclear proteins) and morphogenesis occur, and the viral progeny is released by cell lysis or wrapped inside giant infectious vesicles 8h after infection (9, 11).

Marseillevirus marseillevirus T19 (MRSV) was the first virus to be described in this group, thus being the prototype of the genus *Marseillevirus,* family *Marseilleviridae*. It has an icosahedral capsid of 250 nm in diameter and a circular double-strand DNA genome of 368 kb containing 457 open reading frames (ORFs) predicted to encode proteins, the large majority of which has no known function (1). The genomes of marseilleviruses exhibit a high degree of mosaicism, with genes having different origins (1). Over a half of these genes are associated to a promoter motif – AAATATTT – which has been shown to be important to drive gene expression *in vitro* and are present as single or multiple copies in intergenic regions (12). AT-rich

promoter motifs have been predicted for other amoebal giant viruses, such as mimiviruses, faustoviruses and kaumoebavirus, which seem to be associated to the temporal regulation of the expression of viral genes throughout the infection cycle (13, 14). Currently there is no information about the transcriptional profile of marseilleviruses and even if the predicted genes are all expressed during viral replication in its amoebae host, which constitutes an important gap for a better comprehension upon the biology of these viruses.

In this work we aimed to decipher the transcriptomic pattern of the marseillevirus prototype isolate. Data provided by RNA sequencing (RNA-seq) confirmed the existence of genes previously predicted only by bioinformatics tools, and indicated the existence of three main temporal classes of genes, which were validated by RT-qPCR. We verified that genes belonging to distinct functional groups exhibit different expression levels throughout the infection cycle. Furthermore, we noticed that the promoter motif previously described is not related to any specific temporal class. This work finally provides an overview of the transcriptional profile of marseilleviruses, expanding our knowledge about the biology of these viruses.

Results

Predicted genes of marseillevirus are validated by RNA-seq data

In this study we explored a ribosomal depletion approach for library preparation of the RNA to be sequenced. *Acanthamoeba castellanii* strain Neff were infected with MRSV, cells were collected at different times, and the reads obtained after sequencing were mapped to the viral genome for each dataset corresponding to

the infection times (T = 0h, 1h, 2h, 4h, 5h, 6h, 8h, 10h, and 12h). All 457 initially predicted genes of MRSV (1) were validated by at least 10 reads superposed to corresponding ORFs (considering raw data from all datasets). The total number of normalized reads in transcripts per million reads (TPM) exhibits a highly heterogeneous distribution among the genes (min = 32; max = 318,169; median = 7,036), indicating a large difference in the expression level of these genes (Fig. S1). Indeed, an initial analysis of the 20 most expressed genes suggests a differential gene expression throughout the viral replication cycle, wherein some genes have more cognate reads in earlier periods (e.g. MAR_ORF412 and MAR_ORF014) and others have more cognate reads in later moments (e.g. MAR_ORF300 and MAR_ORF342) (Table 1). Among these most expressed genes, only four have predicted function, one of them being the major capsid protein gene. The fact that the majority of the most expressed genes of MRSV have no known function is intriguing and deserves to be further studied, since they are highly relevant genes for the virus and such studies might provide new insights upon the viral biology and its relationship with its host.

Table 1: Most expressed genes. Top 20 most expressed annotated genes. The total number of normalized reads counts is presented for each gene as well as the read count at each time point of infection.

Gene	Presence of promoter	Expression pattern	Predicted function	0h	1h	2h	4h	5h	6h	8h	10h	12h	Total
MAR_ORF300	N	L	-	1349	31	69	36587	49038	50252	55730	67099	58015	318169
MAR_ORF390	Y	L	multiple zinc ribbon protein	1776	9869	27205	51234	49629	43289	39411	45095	50282	317790
MAR_ORF317	N	L	-	1354	170	40	36306	39576	39271	54230	35973	68786	275705
MAR_ORF342	Y	L	major capsid protein	609	48	26	61325	45875	40048	66725	23909	30871	269436
MAR_ORF384	N	L	-	899	134	23	38560	35602	42222	50020	54510	34821	256790
MAR_ORF370	N	L	-	1317	317	948	17200	28825	27160	39832	89992	36077	241669
MAR_ORF219	Y	L	-	793	106	12744	84839	47325	22149	20213	17578	19583	225329
MAR_ORF029	N	I	-	1843	24117	60761	27563	19259	17693	24533	14184	22469	212422
MAR_ORF305	Y	L	-	379	32	28	35828	39222	43383	28499	15090	25099	187560
MAR_ORF413	Y	L	histone H3	520	44	66	16092	35578	41292	29851	22200	25556	171199
MAR_ORF109	Y	L	-	926	260	24676	29570	17876	9340	5808	5875	10734	105065
MAR_ORF421	N	L	-	210	62	1986	17949	17338	21450	21452	9317	15270	105033
MAR_ORF412	N	E	-	54414	40209	5914	197	101	196	130	202	192	101555
MAR_ORF026	Y	L	-	402	179	173	6830	15641	19574	15741	14843	27536	100920
MAR_ORF193	Y	L	-	385	74	2351	13831	16722	14791	16327	15747	17850	98078
MAR_ORF021	Y	L	zinc finger protein	720	77	16	8775	16500	19814	19344	15478	17237	97962
MAR_ORF389	Y	E	-	27733	43921	21364	514	409	534	396	744	438	96053
MAR_ORF250	N	L	-	202	227	94	15714	20837	21501	16581	9799	10734	95689
MAR_ORF014	N	E	-	66374	22286	2559	98	84	17	59	53	62	91592
MAR_ORF147	N	E	-	44432	36392	2414	141	94	117	85	77	58	83810

Transcriptional profile of Marseillevirus genes

In order to confirm the existence of a temporal profile of gene expression for marseilleviruses, we initially performed a traditional hierarchical cluster analysis, which allows the clustering of genes into groups that share a similar expression profile. This analysis indicated the existence of three main groups, comprising (i) genes for which expression levels are high between 0h and 1h of infection, remaining slightly elevated until 2h and then decreasing; (ii) genes that are expressed between 1h and 2h of infection, with low expression at 0h and also during later moments of the cycle; and finally, (iii) genes with high expression levels after 4h of infection and remaining detected for at least 12h post infection (Figure 1A). This initial analysis demonstrates that not only all MRSV genes are expressed, but that such expression is remarkably fast, wherein all genes are expressed in just 4h after infection, many of them remaining active until the step of viral progeny release. A second cluster analysis, this time using the *k*-means strategy, where *k* was set to 3, corroborated the presence of three categories of genes, which we considered as the traditional temporal categories of “early”, “intermediate” and “late” viral genes, as observed in the hierarchical clustering analysis (Figure 1B).

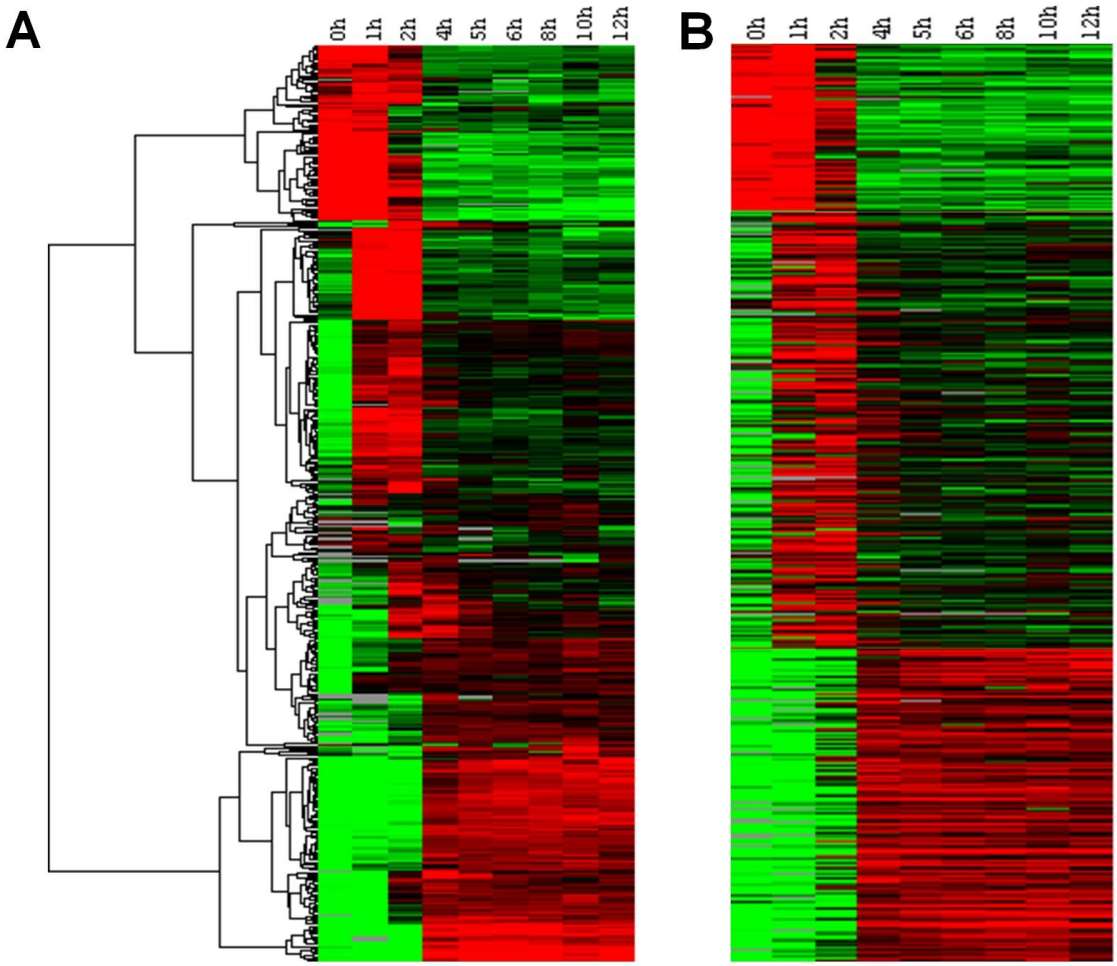


Figure 1: Marseillevirus gene expression classes. (A) Heat map of Marseillevirus gene expression profiles. Rows correspond to all 457 genes and columns to the nine infection time points. Expression profiles are clustered using hierarchical clustering. A dendrogram of the clustering is shown on the left; (B) Heat map of the same expression profiles partitioned into three main classes, ‘‘early’’ (top), ‘‘intermediate’’ (center), and ‘‘late’’ (bottom), by k-means clustering algorithm. Expression levels are displayed from green (low expression) to red (high expression). Gray lines indicate absence of data from the gene at a particular infection time point. Both clustering methods were performed by applying Euclidean distance as similarity metric.

Since data obtained from high throughput sequencing can contain some biases, which could lead us to an erroneous interpretation about the existence of a temporal profile of gene expression for marseilleviruses, we also evaluated the presence of the three categories by means of RT-qPCR, a more sensitive technique and the golden standard method in gene expression analyses (15, 16). Using six different genes, we confirmed the existence of three different classes of gene expression throughout the replication cycle of marseilleviruses (Figure S1). Early

genes exhibit a high expression level as soon as at 0h and have an activation peak 1h post infection, then having a considerably reduced expression after this period. Intermediate genes have an increase in the expression level at 1h and reach the highest level at 2h post infection. Differently from the early genes, the expression level of these genes seems to decrease more slowly after 4h of infection, which suggests maintenance of mRNA in the host cell. Lastly, the late genes have an expression peak at 4h post infection, and the expression is maintained over the remaining steps of the viral replication. Moreover, the genes belonging to different temporal classes are homogeneously distributed along the viral genome. Thus, structural clusters of these genes in the genome related to the transcriptional profile of the virus could not be observed (Figure 2).

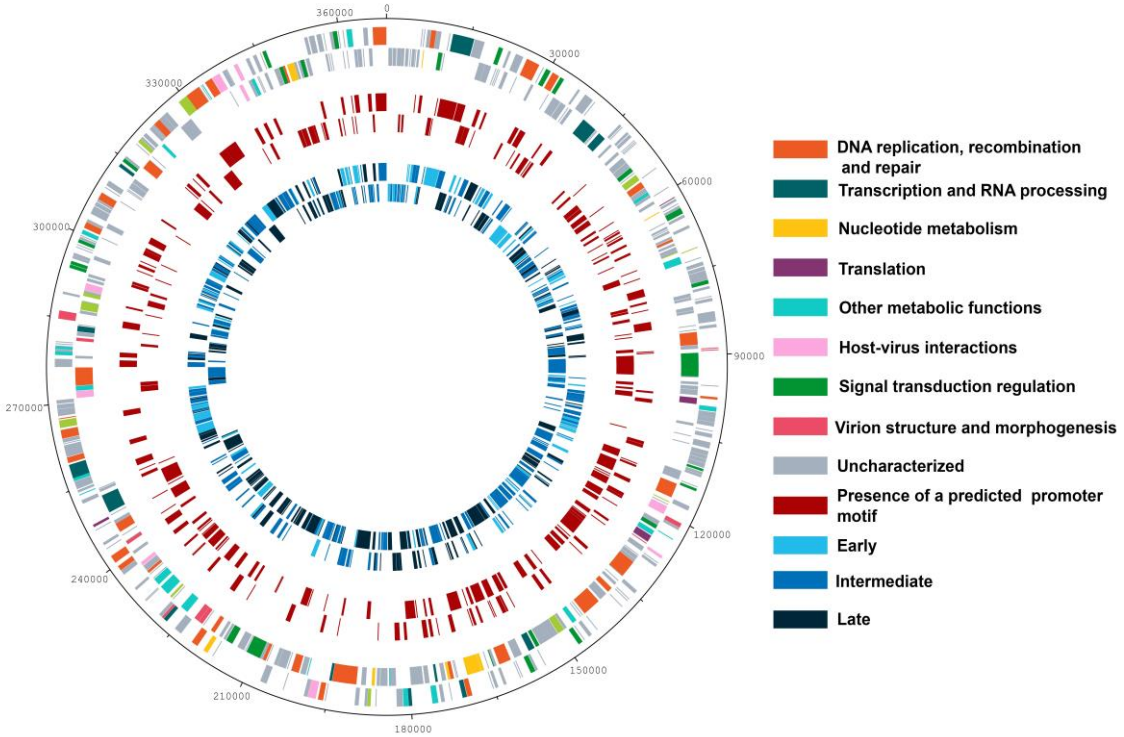


Figure 2: Genome map of Marseillevirus highlighting the different categories of genes. Rings starting from outer to innermost correspond to (i) genome coordinates in kilobases; (ii) predicted

protein-coding genes oriented forward or reverse on DNA strands with different colors corresponding to distinct functional gene categories; (iii) genes followed by the predicted AAATATTT promoter motif; (iv) distribution of genes from different temporal classes, named early, intermediate, and late. Color legend is provided in the right side of the figure.

Functional categories of genes throughout time of infection

The NCLDVs share an intrinsic evolutionary history being considered for some authors a monophyletic group of viruses, which have many homologous genes divided in different functional groups, the so-called nucleo-cytoplasmic virus orthologous groups (NCVOGs) (17). Based on this functional classification, 316 genes (69.1%) of MRSV have uncharacterized function, and the remaining genes are divided into nine categories, among which 36 genes related to DNA replication, repair and recombination, 25 genes related to signal transduction regulation, and others such as those related to nucleotide metabolism (8), transcription (13), translation (4), and viral structure and morphogenesis (10) (Figure 3A). Genes are not clustered in the genome based on this functional classification (Figure 2). These groups comprise genes with differential expression levels throughout the replication cycle, although some categories (e.g. viral structure and morphogenesis) have genes that are only expressed during later moments, as expected (Figure 3B).

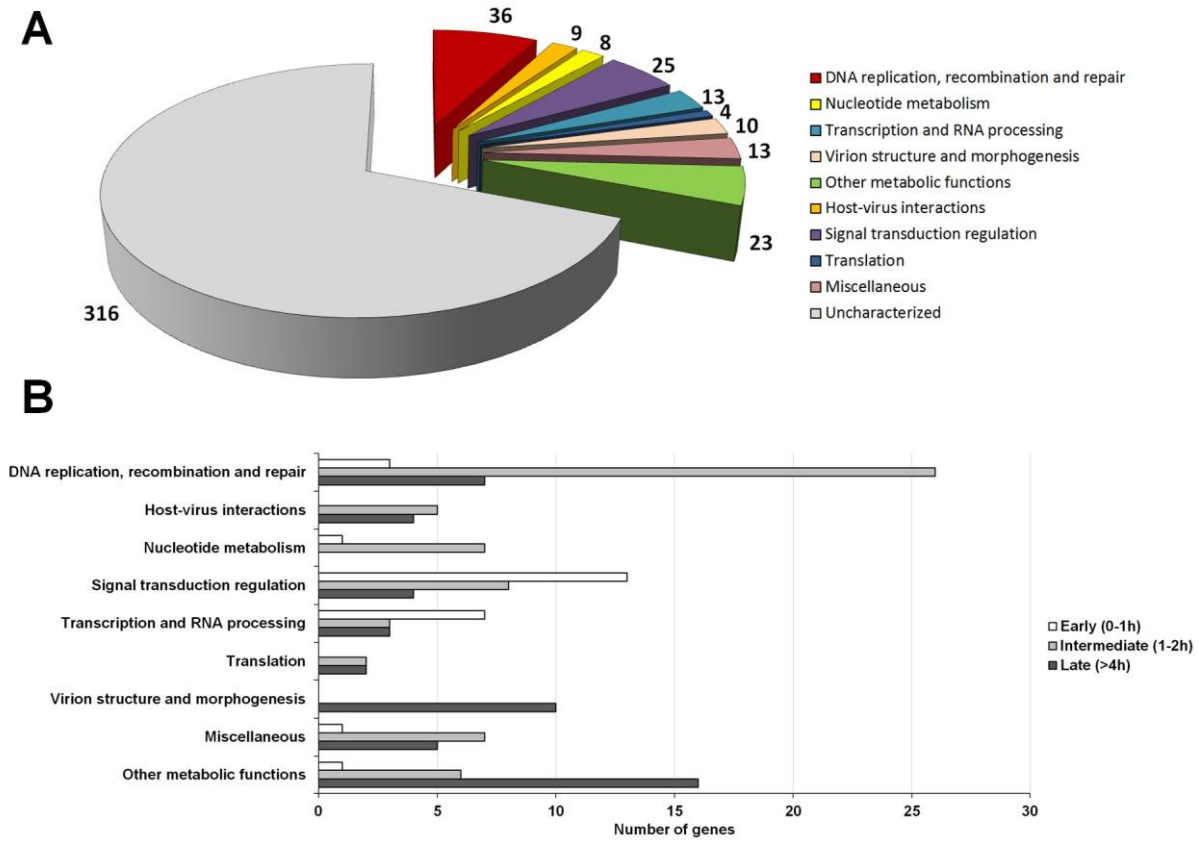


Figure 3: Functional category analysis and correlation to the temporal expression profile of viral genes. (A) Pie-chart representing the amount of marseillevirus genes distributed into different functional categories according to the NCVOG clusters; (B) Distribution of genes from different functional categories into temporal classes of gene expression. Only genes with known functions were included in the graph.

Among the genes related to DNA replication, repair and recombination, the majority are intermediate and late genes, although some genes are expressed as soon as the virus enters the host cells, such as alkylated DNA repair protein and XRN1 5'-3' exonuclease genes (Figure 3B). DNA polymerase B and DNA topoisomerase II genes are expressed 2h post infection, suggesting that the virus exploits the cell machinery to express the genes that will further allow to synthesize its own genetic material. Moreover, some genes as D6/D11 helicase and D5-primase-helicase exhibit late activation peak (at 4h post infection) when the viral factory is fully established, which can be used and possibly incorporated into the viral

particle, as previously observed for other D6-helicases and DNA polymerases (1, 9). It is worthy to note that restriction enzymes coded by MRSV, which are important proteins in virus-host interaction, are expressed 2h post infection, suggesting their participation in moments prior to the establishment of the viral factory. The viral-encoded histones exhibit a late expression peak, indicating their direct role in final steps of the synthesis of viral nucleic acid, as previously suggested (1, 18). With respect to the genes involved in nucleotide metabolism, the great majority is expressed at intermediate moments of the cycle, with peak expression within 2h after infection, such as the ribonucleotide reductase and thymidine kinase genes (Figure 3B).

MRSV has a large set of genes involved in signal transduction regulation, especially serine/threonine (ST) kinases family, which indicates that the virus has a great potential to manipulate the host response against infection. Most of these genes (13/25 = 52%) are readily expressed (T = 0h), since these transcripts have not been detected inside viral particles (1) (Figure 3B). This fast expression profile suggests that the virus interfere with the host response as soon as the virus and the hosts are in contact, which could facilitate the establishment of a productive infectious cycle. Furthermore, some genes are expressed lately (after 4h), indicating that such regulation of host response is maintained throughout the entire replication cycle. We could not observe a trend regarding the expression level of genes belonging to large paralogous families, such as ST kinases or Membrane Occupation and Recognition Nexus (MORN) repeat-containing proteins, since some genes have relatively high expression levels (e.g. MAR_ORF191 (ST-kinase) and MAR_ORF361 (MORN repeat-containing protein), the 53rd and 68th most expressed genes respectively), while others have very low expression levels (e.g. MAR_ORF352 (ST-

kinase) and MAR_ORF366 (MORN repeat-containing protein), the 36th and 6th less expressed genes, respectively).

Among the genes involved in the transcription process are transcription factors such as transcription factor initiation factor IIB (TFIIB) and early transcription factor (eTF), viral RNA polymerase, as well as enzymes involved in the processing of RNA, such as RNA methyltransferase and mRNA capping enzyme. Some of these enzymes are expressed early as the RNA polymerase, indicating a fast onset of viral transcripts synthesis (Figure 3B). Previous data have indicated that viral transcription may begin by using cell nuclear proteins, since marseilleviruses do not carry the RNA polymerase within their virions (9). Our data show that genes responsible for the synthesis of the transcripts, such as RNA polymerases, are rapidly expressed, thus suggesting that the viral enzymes also have an important role in the transcription process, which might act along with and/or just after cell proteins initiate the transcription of viral genes. Curiously, most of transcription factors are expressed only after 2h of infection. It is noteworthy that some of these factors have been detected in the MRSV particle by proteomic analysis (1), which in association with our data suggests that these proteins might be used in the beginning of the cycle.

Differently from other giant viruses, such as mimiviruses, klosneuviruses and tupanviruses (19–21), the marseillevirus have a restricted gene arsenal related to the translation process, consisting in only four translation factors (Figure 3). These genes are expressed from 2h post infection and two of them only after 4h, indicating their role in the protein synthesis possibly inside the fully matured viral factory. Moreover, genes related to viral morphogenesis are also expressed in late moments, which was totally expected since they are essential genes involved in the formation of the viral particle. These include membrane component and major capsid protein genes (the

latter being one of the most expressed gene throughout the cycle – Table 1), besides the core gene encoding A32-like packaging ATPase, which is related to the packaging of the viral genome into the virion (22, 23) (Figure 3B). Finally, with respect to the genes involved in other metabolic functions, such as those encoding lipases, proteases and proteins involved in redox reactions, the majority (16/23 = 69.5%) have a late expression. This is for instance the case for class 3 lipases, peptidoglycan peptidase and thioredoxin, indicating their role in the final steps of replicative cycle, probably in the regulation of host response to the viral infection and production of new progeny.

AT-rich promoter motif is not associated to a specific class of genes

Over a half of MRSV genes are associated to an AT-rich promoter motif – AAATATTT – which impact in gene expression was observed (12). At the time of the description of the motif there was no information about the temporal expression profile as evidenced in this work, thus an attempt to associate the promoter to any class of genes could not be made, as observed for other Megavirales representatives, such as mimiviruses and poxviruses (13, 24, 25). The genes that are associated to the promoter are homogeneously distributed along the genome (Figure 2). Our data show that of the 457 MRSV-encoded genes, 83 (18%) are expressed early between 0-1h, 218 (48%) are intermediate genes with an expression peak at 1-2h, and 156 (34%) are late genes, with peak of expression from 4h of infection (Figure 4A). However, the promoter motif does not seem to be associated to any specific temporal class of genes, wherein it is present in 43.4% of early genes, 49.1% of intermediate genes and 58.3% of late genes (Figure 4B). Similarly, it was not

possible to establish any specific association between the promoter motif to different functional class of genes (Figure 4C). In all classes are found genes that are not followed by the promoter motif at a rate ranging from 37.5% to 77.8%, the classes of nucleotide metabolism and virus-host interaction being those composed by the smallest and highest numbers of genes associated with the promoter motif, respectively (Figure 4C). Among the 20 most expressed genes (~5% of all MRSV genes), half of them are followed by the promoter motif and the majority (15 genes) are expressed at late times during the replication cycle (Table 1). Thus, although the promoter motif is important in gene regulation (12), this promoter is not associated with any specific class of MRSV genes, which suggests that the transcriptional regulation of marseillevirus has different mechanisms than the already proposed promoter.

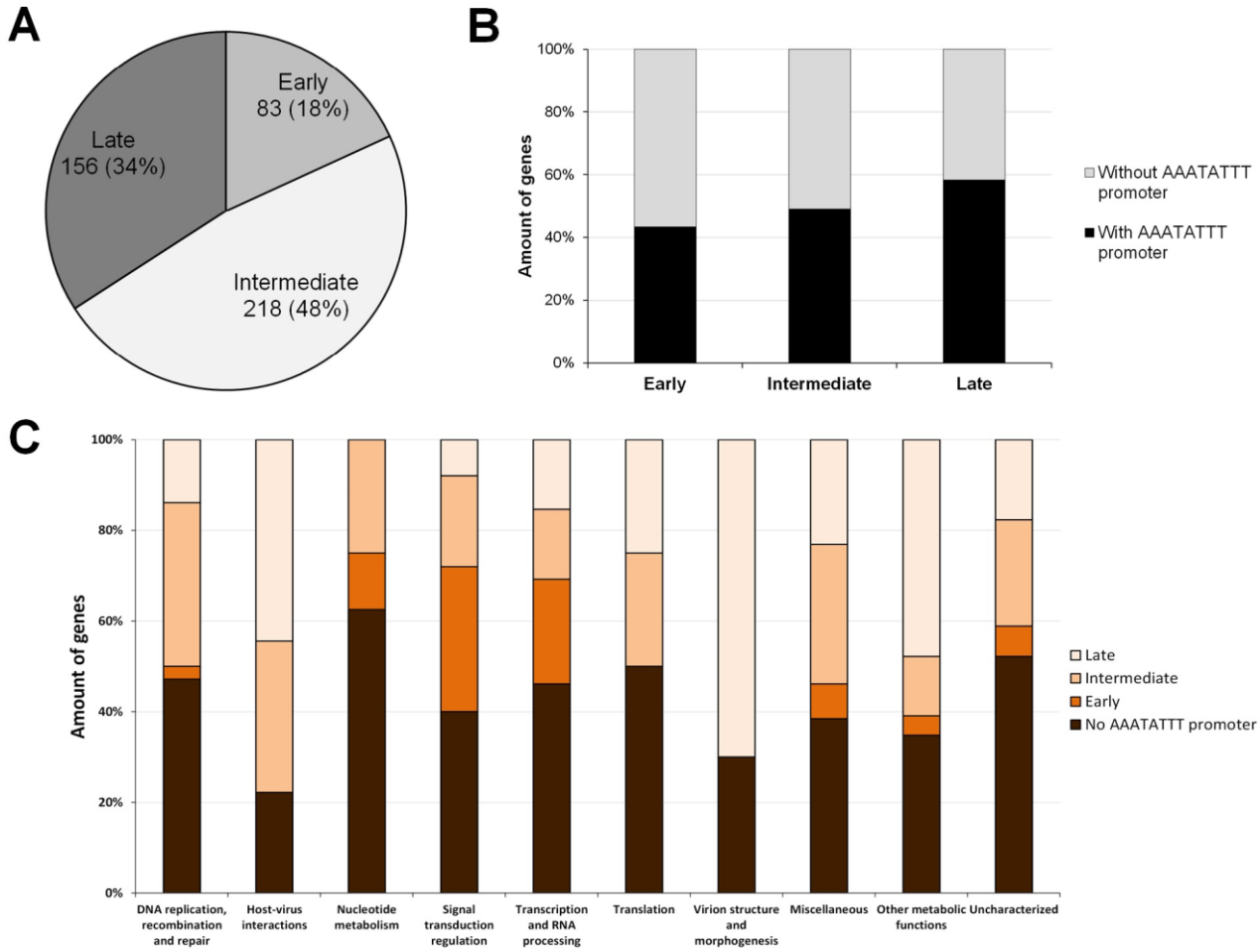


Figure 4: Promoter motif association to temporal classes of genes. (A) Pie-chart demonstrating the 457 viral genes divided into the three temporal categories, early, intermediate and late; (B) Association of genes from different temporal classes of expression to the presence/absence of AAATATTT promoter motif; (C) Promoter motif association to genes from different functional categories and temporal classes of expression.

Discussion

To elucidate the transcriptional landscape of marseilleviruses, we initially explored the RNA-seq approach, a technique that has been widely diffused and in constant improvement in this age of high throughput sequencing (26, 27). To avoid a massive sequencing of rRNA that is a major component of the cellular RNA pool, we used the rRNA depletion strategy, which was never previously used in studies involving giant viruses and amoebae. The method was efficient with respect to its

purpose and revealed important features about marseillevirus transcription profile. In view of the high coverage of RNA reads mapped to the viral genome observed in this work, it is possible that these regions of the transcripts are very extensive, turning the intergenic regions shorter than expected, a characteristic also observed in phycodnaviruses, another member group of NCLDVs (28). The temporal expression profile observed for Marseillevirus T19 – genes divided into early, intermediate and late categories – is similar to that observed for other giant DNA viruses (29). This profile was validated by RT-qPCR data, a more sensitive technique compared to the high throughput sequencing. Our data confirm the existence of all 457 genes initially predicted for MRSV with a minimum of 10 reads of mapped RNA and show an incredibly rapid gene expression where all genes are expressed within 4 h after viral infection. Considering that the virus factory established by these viruses in the host cytoplasm is recurrently observed from 4h of infection (9, 11), most of the viral genes are still expressed outside this structure, or the viral transcripts are still present in the host cytoplasm. Interestingly the late viral genes are expressed from 4h but remain active for at least 12h after infection. It is possible that the maintenance of the expression of these genes is performed inside the viral factory, where the mRNAs would be less subjected to degradation by the cellular machinery.

Genes belonging to different functional classes have peaks of expression at varying moments, many of them exhibiting an expected pattern, such as translation and viral morphogenesis genes that are mostly expressed later during the cycle. However, some classes have genes that are expressed at different times, such as genes involved in the DNA replication, repair, and recombination process. This profile shows that these genes are important throughout the entire viral cycle, some being used at the beginning of the cycle and others being used lately, whose products can

be packaged into the viral particle, as already verified for the viral DNA polymerase (1). In addition, we observed that the MRSV genes are not organized in the genome according to their functions, not even regarding their temporal expression profile. This reinforces the idea of genomic mosaicism previously proposed for these viruses (1).

It is worth to note that the majority of the MRSV genes with highest level of expression have no known function (hypothetical proteins). Overall, a substantial part of the most abundant part of the marseillevirus transcriptome represents a “dark transcriptome”. It indicates that the most abundant part of the MRSV transcripts encode for putative proteins that remain to be structurally and functionally studied. A similar pattern is also observed for all giant viruses that have been studied in the lens of transcriptomics so far, for instance mimivirus, mollivirus and pithovirus (13, 30, 31). This shows that a substantial part of the gene content of these entities differs considerably from that of other living organisms, thus representing new expressed elements.

Similar to other NCLDVs, marseilleviruses have a conserved AT-rich promoter motif, which is associated with over 50% of the genes (12). Our data show that this promoter is not associated with any specific temporal class of genes since it accompanies between 40% and 60% of the early, intermediate and late genes. This contrasts with what was observed for mimivirus and poxvirus, in which an AT-rich promoter motif is mostly associated with early genes (13, 24). The presence of an AT-rich promoter in a GC-richer genome (45%) than for the case of the mimivirus genomes, and its non-correlation with any temporal class, suggest the existence of other regulatory mechanisms yet to be discovered for marseillevirus. It is possible that other viral promoters (probably degenerated regions) exist, which could be associated with some specific temporal class, as observed for the phycodnaviruses

(32). However, further studies to accurately define the 5 'and 3'-UTR of the viral transcripts would be required to confirm this hypothesis. Differing from other strategies, the single molecule real time sequencing technology would be useful for such goal (33). Indeed, it applies specific tags at both extremities of the mRNA, thus allowing defining with great accuracy the complete structure of the transcript, i.e. 5' and 3' untranslated regions (UTR), what the ribosome depletion strategy does not allow. Further studies employing this strategy could reveal new and interesting features of marseillevirus transcription complex.

Altogether our work provides new information about the biology of a newly discovered group of viruses, suggesting the existence of a much more complex transcriptional machinery than originally thought for these viruses. Advances in RNA sequencing and *in silico* analysis technologies and analytical tools may provide important insights into the molecular mechanisms exhibited by marseilleviruses, especially in the context of interactions with their host and sympatric organisms within amoebae.

Materials and methods

Cell culture and infection by Marseillevirus

For viral production, *Acanthamoeba castellanii* strain Neff (ATCC 30010) cells cultivated in T150 cm^2 cell culture flasks containing peptone-yeast extract-glucose (PYG) medium were infected with Marseillevirus T19 (MRSV) and kept at 30°C for up to seven days. After complete cell lysis, the material was collected and passed through 0.45 µm membranous filter (Millipore, USA) to retain cell debris. The virus

was titrated by end-point method, the titer being expressed as tissue culture infectious doses 50% ($TCID_{50}$) (34), and kept at refrigeration until further use.

For transcriptomic analysis, a total of nine infection points was performed, that is 0h, 1h, 2h, 4h, 5h, 6h, 8h, 10h and 12h, according to the replication profile observed for the virus during one-step growth curve assays (data not shown). To do so, T150 cm² cell culture flasks containing each $2x10^7$ cells of *A. castellanii* kept in 25 mL of PYG medium were infected with MRSV at a multiplicity of infection (MOI) of 100. Cells were kept at 30°C for 30 minutes and then the supernatant was removed and the cell monolayer was washed twice with Page's Amoeba Saline buffer to remove excess virus. Fresh PYG medium was added in each flask which was kept at 30°C and the cells collected at each determined time point. The time point of 0h corresponds to 30 minutes after infection (adsorption period). The flask content was collected and centrifuged for 10 minutes at 1,000 *g*, the supernatant was discarded and the cell pellet resuspended in RNA*later* stabilization reagent (QIAgen, France) and kept at -20°C until further use.

RNA extraction and ribosomal depletion

RNA extraction was performed using RNeasy Midi Kit (cat no: 75144 QIAgen, France), following the manufacturer's protocol. Briefly, cells were centrifuged for 10 minutes at 3,000 *g* to remove the RNA*later* reagent and then resuspended in 4 mL of RLT buffer for cell lysis. An on-column DNA digestion was performed. The total RNA was eluted with two successive addition of ~200 µL of RNase free water. Then, two successive digestions with Turbo DNase (Invitrogen, USA) were performed for each RNA sample to completely eliminate DNA contamination. Each treatment was

performed by incubating the samples and enzyme for 30 minutes at 37°C. Finally, enzyme inactivation reagent was added and incubated at room temperature for five minutes. The samples were centrifuged at 10,000 g for 1.5 min, collected and quantified by NanoDrop spectrophotometer. The absence of contaminating DNA was checked by qPCR (data not shown).

For mRNA enrichment, crucial step before RNA sequencing, the ribosomal RNA (rRNA) depletion strategy was chosen and performed by using Ribo-Zero rRNA removal Kit (Bacteria) (Illumina, France), following the manufacturer's protocol. Briefly, a total of 2.5 µg of each RNA sample was hybridized with probes for rRNA, being incubated at 68°C for 10 minutes, and then associated to magnetic beads for rRNA removal (incubation at 50°C for 5 minutes). Then, RNA was purified by ethanol precipitation, resuspended in specific buffer for sequencing library preparation and the ribosomal depletion was checked by using Agilent 2100 Bioanalyzer with a RNA 6000 Pico Chip (data not shown). The depleted RNAs were used for construction of sequencing library.

cDNA production and sequencing

cDNA production and library construction for sequencing was performed by using TruSeq stranded total RNA kit (Illumina, France), following the manufacturer's protocol. Briefly, the first strand cDNA synthesis was performed with SuperScript II enzyme. After second strand synthesis, the sequences were adenylated at the 3' end, adaptors were linked and PCR amplification was performed to obtain the library. Each library profile was visualized on a DNA 1000 Bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA) to read the optimum size in bp and the final

concentration library was measured in nmol/l. The libraries were normalized at 2 nM and pooled for sequencing using MiSeq Reagent kit V3 in 150 cycles.

Read mapping, counting and normalization

Reads generated from each RNA-seq dataset were uploaded to the Galaxy web platform, and we used the public server at usegalaxy.org to analyze the data (35). Reads were mapped on the Marseillevirus T19 genome (GenBank accession number NC_013756.1) by using the HISAT2 software with default parameters and considering a maximum intron length of 5000 (36). Mapping results were analyzed using the HTseq-count software, with the union mode (37). Only unique mapped reads were taken into account for further analyses. We were able to assign at least 10 reads to each Marseillevirus gene (Supplementary Fig. 2). For expression analysis, the raw data was normalized considering gene length and sequencing coverage by means of Transcripts Per Million reads (TPM) (26). Normalized read counts ranged from 32 to 318,169.

Gene expression cluster analyses

To reveal the transcriptional pattern during Marseillevirus infection, we clustered gene transcription profiles using hierarchical and k-means clustering methods. We first log-transformed the normalized read count profiles and centered this data by the mean. Both cluster analyses were performed with Cluster 3.0 program (38) using Euclidean distance as similarity metric and setting the number of cluster to 3. Data visualization was done using Java TreeView program (39).

485

486 Expression profile validation by real-time quantitative PCR

487

488 For validating the expression profile observed with the RNA-seq analysis, we
489 performed reverse transcription real-time polymerase chain reaction (RT-qPCR)
490 using different genes (Supplementary table 1). Specific primers for each gene were
491 designed using the primer-blast tool at National Center for Biotechnology Information
492 platform (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Viral infection and RNA
493 extraction was performed as previously described. The RT-qPCR was performed in a
494 one-step reaction by using QuantiTect SYBR Green RT-PCR Kit (QIAgen, France)
495 following the manufacturer's recommendations. The assays were performed in a
496 BioRad Real-Time PCR Detection System (BioRad) using the following thermal
497 conditions for all genes: 30 min at 50°C for reverse transcription step, followed by 15
498 min at 95ºC and 40 amplification cycles of 15 s at 94°C, 30 s at 60°C and 30 s at
499 72°C. The values were expressed as arbitrary units (delta-Ct). The experiment was
500 performed twice in duplicate.

501

502 Acknowledgments

503

504 We thank Catherine Robert for excellent technical support in library
505 preparation and sequencing. This work was supported by the French Government
506 under the "*Investissements d'avenir*" program managed by the *Agence Nationale de*
507 *la Recherche (ANR)*, [reference: *Méditerranée-Infection* 10-IAHU-03], by Région
508 Provence-Alpes-Côte d'Azur and European funding FEDER PRIMI. R.A.L.R.

scholarship was provided by CNPq and CAPES. R.A.L.R., P.C., B.L.S. and J.S.A. are members of a CAPES-COFECUB project. J.S.A. is a CNPq researcher.

References

1. 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 E V, 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.
2. Colson P, Pagnier I, Yoosuf N, Fournous G, La Scola B, Raoult D. 2013. "Marseilleviridae", a new family of giant viruses infecting amoebae. Arch Virol 158:915–920.
3. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, Trape JF, Koonin E V., La Scola B, Raoult D. 2012. Microbial culturomics: Paradigm shift in the human gut microbiome study. Clin Microbiol Infect 18:1185–1193.
4. 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.
5. 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.
6. Takemura M. 2016. Morphological and Taxonomic Properties of Tokyovirus, the First Marseilleviridae Member Isolated from Japan. Microbes Environ

31:442–448.

7. Doutre G, Philippe N, Abergel C, Claverie JM. 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.

8. 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.

9. Fabre E, Jeudy S, Legendre M, Trauchessec M, Claverie J, Abergel C. 2017. Noumeavirus replication relies on a transient remote control of the host nucleus. Nat Commun 1–57.

10. Colson P, De Lamballerie X, Yutin N, Asgari S, Bigot Y, Bideshi DK, Cheng XW, Federici BA, Van Etten JL, Koonin E V., La Scola B, Raoult D. 2013. “Megavirales”, a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Arch Virol 158:2517–2521.

11. 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.

12. Oliveira GP, Lima MT, Arantes TS, Assis FL, Rodrigues RAL, da Fonseca FG, Bonjardim CA, Kroon EG, Colson P, La Scola B, Abrahão JS. 2017. The Investigation of Promoter Sequences of Marseilleviruses Highlights a Remarkable Abundance of the AAATATTT Motif in Intergenic Regions. J Virol 91:e01088-17.

13. Legendre M, Audic S, Poirot O, Hingamp P, Seltzer V, Byrne D, Lartigue A,

559 Lescot M, Bernadac A, Poulain J, Abergel C, Claverie JM. 2010. mRNA deep
560 sequencing reveals 75 new genes and a complex transcriptional landscape in
561 Mimivirus. Genome Res 20:664–674.
562 14. Oliveira GP, de Aquino ILM, Luiz APMF, Abrahão JS. 2018. Putative promoter
563 motif analyses reinforce the evolutionary relationships among faustoviruses,
564 kaumoebavirus, and asfarvirus. Front Microbiol 9.
565 15. Bustin SA. 2000. Absolute quantification of mrna using real-time reverse
566 transcription polymerase chain reaction assays. J Mol Endocrinol.
567 16. Bustin SA, Benes V, Nolan T, Pfaffl MW. 2005. Quantitative real-time RT-PCR
568 - A perspective. J Mol Endocrinol.
569 17. Yutin N, Wolf YI, Raoult D, Koonin E V. 2009. Eukaryotic large nucleo-
570 cytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of
571 viral genome evolution. Virol J 6.
572 18. Thomas V, Bertelli C, Collyn F, Casson N, Telenti A, Goesmann A, Croxatto A,
573 Greub G. 2011. Lausannevirus, a giant amoebal virus encoding histone
574 doublets. Environ Microbiol 13:1454–1466.
575 19. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B,
576 Suzan M, Claverie J-M. 2004. The 1.2-Megabase genome sequence of
577 Mimivirus. Science (80-) 306:1344–1350.
578 20. Schulz F, Yutin N, Ivanova NN, Ortega DR, Lee TK, Vierheilig J, Daims H,
579 Horn M, Wagner M, Jensen GJ, Kyrpides NC, Koonin E V., Woyke T. 2017.
580 Giant viruses with an expanded complement of translation system components.
581 Science (80-) 356:82–85.
582 21. Abrahão J, Silva L, Silva LS, Khalil JYB, Rodrigues R, Arantes T, Assis F,
583 Boratto P, Andrade M, Kroon EG, Ribeiro B, Bergier I, Seligmann H, Ghigo E,

584 Colson P, Levasseur A, Kroemer G, Raoult D, La Scola B. 2018. Tailed giant
585 Tupanvirus possesses the most complete translational apparatus of the known
586 virosphere. Nat Commun 9:749.
587 22. Koonin E V., Senkevich TG, Chernos VI. 1993. Gene A32 product of vaccinia
588 virus may be an ATPase involved in viral DNA packaging as indicated by
589 sequence comparisons with other putative viral ATPases. Virus Genes 7:89–
590 94.
591 23. Chelikani V, Ranjan T, Kondabagil K. 2014. Revisiting the genome packaging
592 in viruses with lessons from the “Giants.” Virology 466–467:15–26.
593 24. Davison AJ, Moss B. 1989. Structure of vaccinia virus early promoters. J Mol
594 Biol 210:749–769.
595 25. Knutson B a, Liu X, Oh J, Broyles SS. 2006. Vaccinia virus intermediate and
596 late promoter elements are targeted by the TATA-binding protein. J Virol
597 80:6784–6793.
598 26. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson
599 A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. 2016. A survey
600 of best practices for RNA-seq data analysis. Genome Biol.
601 27. Todd E V., Black MA, Gemmell NJ. 2016. The power and promise of RNA-seq
602 in ecology and evolution. Mol Ecol 25:1224–1241.
603 28. Blanc G, Mozar M, Agarkova I V., Gurnon JR, Yanai-Balser G, Rowe JM, Xia
604 Y, Riethoven JJ, Dunigan DD, Van Etten JL. 2014. Deep RNA sequencing
605 reveals hidden features and dynamics of early gene transcription in
606 Paramecium bursaria chlorella virus 1. PLoS One 9.
607 29. Oliveira G, Andrade A, Rodrigues R, Arantes T, Boratto P, Silva L, Dornas F,
608 Trindade G, Drumond B, La Scola B, Kroon E, Abrahão J. 2017. Promoter

Motifs in NCLDVs: An Evolutionary Perspective. Viruses 9:16.

30. Legendre M, Lartigue A, Bertaux L, Jeudy S, Bartoli J, Lescot M, Alempic J-M, Ramus C, Bruley C, Labadie K, Shmakova L, Rivkina E, Couté Y, Abergel C, Claverie J-M. 2015. In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. Pnas 112:5327–5335.

31. 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 111:201320670.

32. Fitzgerald LA, Boucher PT, Yanai-Balser GM, Suhre K, Graves M V., Van Etten JL. 2008. Putative gene promoter sequences in the chlorella viruses. Virology 380:388–393.

33. Fish R, Zhang S, Bostick M, Chang C, Gandlur S. 2014. An Improved cDNA Library Generation Protocol for Transcriptome Analysis from a Single Cell. J Biomol Tech 25(Suppl).

34. Reed LJ, Muench H. 1938. A simple method of estimating fifty percent endpoints. Am J Hyg.

35. Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Ech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544.

36. Kim D, Langmead B, Salzberg SL. 2015. HISAT: A fast spliced aligner with low memory requirements. Nat Methods 12:357–360.

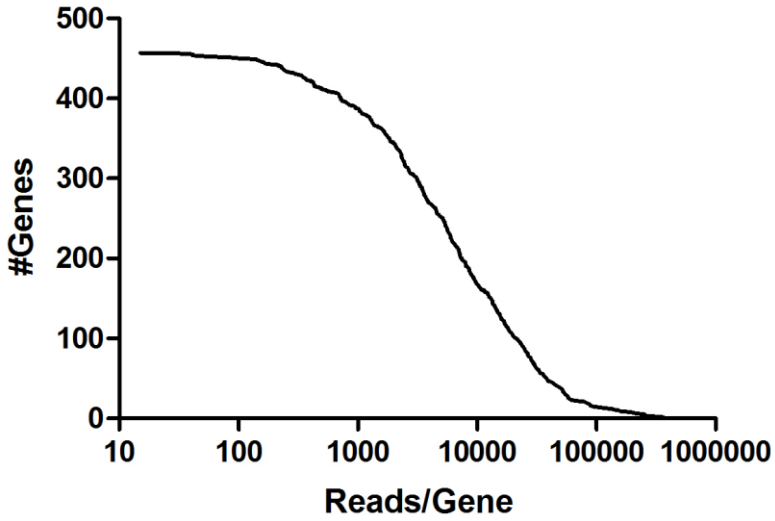
634 37. Anders S, Pyl PT, Huber W. 2015. HTSeq-A Python framework to work with
635 high-throughput sequencing data. Bioinformatics 31:166–169.
636 38. de Hoon MJL, Imoto S, Nolan J, Miyano S. 2004. Open source clustering
637 software. Bioinformatics 20:1453–1454.
638 39. Saldanha AJ. 2004. Java Treeview - Extensible visualization of microarray
639 data. Bioinformatics 20:3246–3248.

640

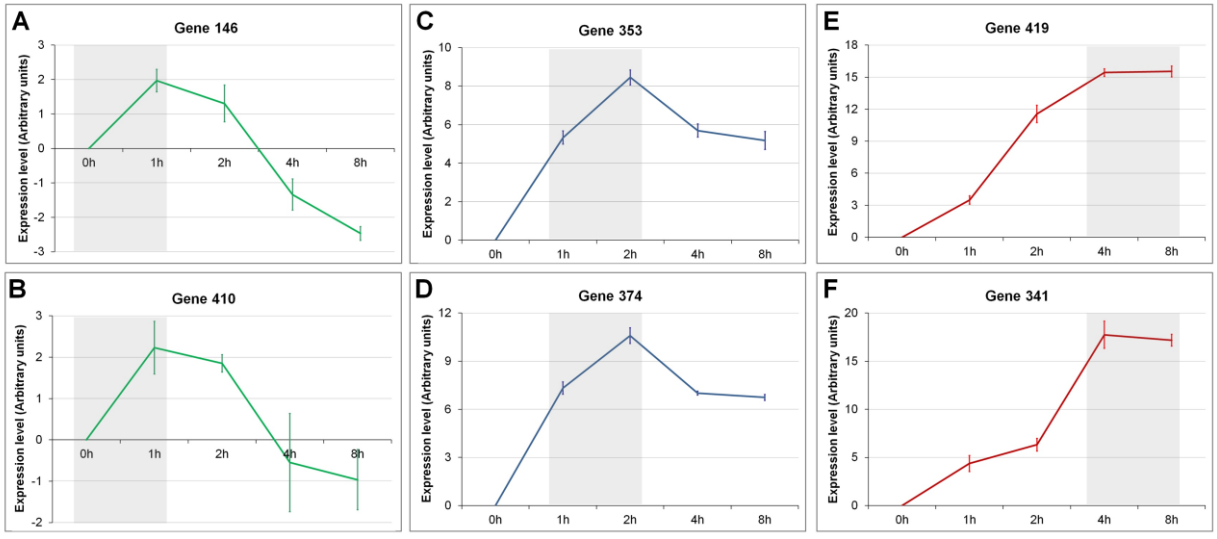
641 **Supplementary material**

642 **Supplementary table 1:** Genes and primers used in RT-qPCR reactions.

Gene	Forward primer (5'-3')	Reverse primer (5'-3')	Tm (ºC)
MAR_ORF147	agc cat tgg aag agc aga gg	tgc ctt ccc ata tcg ttc gc	60
MAR_ORF342	ggt ttt gtg gac ttg gcg ac	cct ttt tga tgg cac gca ca	60
MAR_ORF354	caa cac aac cgg acc aga ga	ggg gct cgc ttt ttc ttt cc	60
MAR_ORF375	ggg gag gtc agt tgt tct cg	cca gag cag tct cgg tat cc	60
MAR_ORF412	cct gtg aga tgg cag gag tc	cct tcc tct tcc ccg tca aa	60
MAR_ORF421	gca aga gcc cca aaa aga gg	ccg tgt tgt acg gaa tca gc	60



Supplementary figure 1: Number of genes *vs* number of total cognate reads (cumulative). Cumulative distribution of total cognate reads to the genes of Marseillevirus. Before normalization, the number of reads per gene ranged from 15 to 372,331.



Supplementary figure 2: Validation of RNA-seq data by RT-qPCR assays. Molecular assays were performed to validate the temporal profile for the gene expression of marseillevirus by using different genes. Gray boxes highlight the peak of activation of the genes, indicating the (A,B) 'Early' genes; (C,D) 'Intermediate' genes; and (E,F) 'Late' genes. The expression levels are depicted as arbitrary units, calculated using ΔCt method. All assays were performed twice in duplicate. Error bars: standard deviation.

5. DISCUSSÃO

Desde seu surgimento no final do século XIX, a virologia vem avançando consideravelmente, com diversos vírus já descritos capazes de interagir com organismos dos três domínios da vida. Os vírus formam uma extensa rede de interação com seus hospedeiros, formando uma virosfera, e até o momento não se sabia o grau de conexão desse sistema. Dentre os vírus descritos nos últimos anos, os vírus gigantes de amebas surpreendeu a comunidade científica pelas suas características únicas, tanto biológicas quanto genômicas, que os distinguem dos demais vírus (ANDREANI *et al.*, 2016, 2018, LEGENDRE *et al.*, 2014, 2015; PHILIPPE *et al.*, 2013; RAOULT *et al.*, 2004). Apesar do crescente progresso observado nessa área de estudos, ainda existem importantes lacunas a serem preenchidas, especialmente no que tange a evolução e a diversidade destes vírus. Neste contexto, o presente trabalho apresenta uma visão ampla da rede de interação entre os vírus conhecidos e seus hospedeiros e fornece uma série de dados que contribuem para a construção do conhecimento em múltiplas vertentes no campo dos vírus gigantes, como a sistemática, a genômica e a transcriptômica de diferentes grupos virais.

A rede de interação entre os vírus e seus hospedeiros apresentada nesse trabalho foi baseada na lista de vírus oficialmente reconhecidos pelo ICTV publicada em 26 de maio de 2016 (MSL#30, *Master Species List*). A criação do ICTV foi um marco na história da virologia, pois permitiu, a partir de então, classificar e sistematizar o conhecimento adquirido sobre os vírus, estabelecendo os critérios para a taxonomia viral. O comitê publica periodicamente uma MSL atualizada contendo todos os táxons conhecidos, os quais são propostos por diferentes pesquisadores e possivelmente aprovados pelo comitê executivo da organização. A lista utilizada nesse trabalho apresentava um total de 3704 espécies virais, distribuídas em 112 famílias e 07 ordens. Após uma cuidadosa análise para designar os hospedeiros naturais dos vírus conhecido à época, foi possível a construção de um grafo representando a rede de interação (daqui em diante referido como *network*) entre esses organismos, resultando em uma visão ampla da relação entre a virosfera conhecida e seus hospedeiros. O grafo foi construído considerando os hospedeiros em nível taxonômico de gênero, pois não foi possível defini-los com precisão em nível de espécie. Os dados revelaram que os vírus apresentam um

espectro muito restrito de hospedeiros, em que mais de 75% dos vírus conhecidos é capaz de infectar apenas uma ou duas espécies. Esse valor chega a mais de 80% considerando alguns grupos seguindo a classificação de Baltimore, como vírus que apresentam genoma de fita dupla de DNA (grupo I) e fita simples de RNA capazes de realizar transcrição reversa (grupo VI). Em decorrência disso, o que observamos foi um *network* pouco conectado com a maioria dos vírus conhecidos infectando plantas e animais, especialmente o ser humano e espécies de interesse clínico, econômico ou biotecnológico, revelando uma virosfera altamente antropocêntrica.

Um total de 483 gêneros de plantas e 467 de animais foram identificados contendo espécies hospedeiras de vírus conhecidos. Estes grupos estão mais interconectados que outros, embora mais de 70% dos hospedeiros possui apenas uma ou duas espécies de vírus com representantes capazes de infectá-los. É notável que alguns poucos vírus sejam capazes de cruzar a barreira de hospedeiros, infectando tanto animais quanto plantas. Estes vírus são parasitas de plantas transmitidos por vetores artrópodes, onde são capazes de se multiplicar e alcançar os hospedeiros vegetais (DIETZGEN; MANN; JOHNSON, 2016). Dentre os vírus que infectam bactérias (bacteriófagos), a maioria está associada a espécies do gênero *Mycobacterium* e *Escherichia*. Visto que algumas espécies destes grupos são intensamente estudadas devido ao seu interesse médico e biotecnológico, tal característica já era esperada (KORB; CHUTURGOON; MOODLEY, 2016; VILA *et al.*, 2016). Alguns grupos possuem poucos vírus descritos, como arqueias, fungos e protistas, o que indica uma forte necessidade de explorar esta área da virosfera. Ainda assim, os hospedeiros que mais possuem vírus conhecidos são aqueles de interesse humano, com destaque para os gêneros *Saccharomyces* (fungo) e *Chlorella* (protista). O primeiro agrupa leveduras de forte interesse econômico e biotecnológico, como *S. cerevisiae*, um micro-organismo considerado domesticado, utilizado mundialmente na produção de bebidas alcóolicas, resultando em enorme geração de renda global (GALLONE *et al.*, 2016; SICARD; LEGRAS, 2011). O segundo grupo comporta as algas verdes, as quais são utilizadas como suplementação nutricional por serem fontes de vitaminas e micronutrientes e sua eficácia contra doenças humanas está em constante investigação (EBRAHIMI-MAMEGHANI *et al.*, 2017; PANAHI *et al.*, 2016). Por fim, dentre as espécies com mais vírus associados estão animais dos gêneros *Bos*, *Sus* e *Gallus*, e plantas dos gêneros *Solanum*, *Nicotiana* e *Phaseolus*. Nestes grupos encontram-se muitas

espécies exploradas na agropecuária, como bovinos, porcos, galinhas e diversas leguminosas consumidas em todo o mundo, resultando em uma economia de bilhões de dólares anualmente (REGANOLD; WACHTER, 2016; THORNTON, 2010). O espectro restrito de hospedeiros observado é muito provavelmente resultante de uma virologia fortemente enviesada pelos interesses humanos e uma mudança de perspectiva é necessária para avançarmos nosso conhecimento sobre a virosfera.

Uma ferramenta moderna fundamental para a virologia é a metagenômica. Diversos trabalhos envolvendo análises de sequências de origem viral nos mais distintos locais e hospedeiros têm sido realizados, contribuindo para a expansão da virosfera (ZHANG; SHI; HOLMES, 2018). Ao analisar diversos trabalhos de metavirômica realizados em diferentes regiões do planeta, foi possível identificar grupos virais distribuídos por todo o globo, em especial os bacteriófagos da ordem *Caudovirales*. Em contrapartida, alguns grupos foram restritos a um ou dois ambientes, em particular locais considerados extremos, como fontes termais e zonas polares. Esses dados corroboram a hipótese proposta por O'malley (2008) a qual afirma que os vírus estão em todos os lugares, mas o ambiente os selecionam (O'MALLEY, 2008). A distribuição dos vírus no planeta corrobora o conceito de virosfera e demonstra que os diferentes ambientes estão conectados pelos vírus que ali circulam. Um estudo recente envolvendo análises baseadas em mais de 5.0 Tb de dados de metagenômica oriundos de 3042 locais espalhados pelo planeta corroboram os resultados aqui apresentados, contribuindo para avançar na compreensão da ecologia viral a nível global (PAEZ-ESPINO *et al.*, 2016). Estudos de metagenômica e metatranscriptômica têm sido realizados para identificação de novos vírus, tanto dispersos no ambiente quanto associados a hospedeiros diversos, como insetos, crustáceos, anfíbios, entre outros, o que vem contribuindo para ampliar a virosfera conhecida (REMNANT *et al.*, 2017; RUSSO *et al.*, 2018; SHI *et al.*, 2016; ZHANG; SHI; HOLMES, 2018). Recentemente o ICTV passou a considerar dados provenientes de metavirômica para criação de novos táxons e estabelecimento de critérios para taxonomia viral (SIMMONDS *et al.*, 2017). Com isso, a última divulgação da lista pelo ICTV (MSL#33), publicada em 19 de outubro de 2018, apresenta um total de 4958 espécies virais, distribuídas em 143 famílias e 14 ordens, além de estabelecer os táxons "Classe", "Subfilo" e "Filo" para a taxonomia viral, um avanço considerável em relação à MSL#30 utilizada para as

análises do presente trabalho (ICTV, 2018). É possível que o espectro de hospedeiros destes vírus seja igualmente restrito ao dos vírus descritos até 2016, embora novas investigações sejam necessárias para confirmar tal hipótese. Embora a metagenômica esteja auxiliando muito o avanço da virologia, é preciso associá-la à virologia clássica para que se tenha uma compreensão mais ampla sobre os vírus, especialmente no que se refere às características biológicas dessas entidades.

Considerando que *Homo sapiens* é a espécie que mais possui vírus associados, um dos objetivos desta tese foi determinar quais seriam estes vírus e qual o tropismo destes parasitas. Foi possível identificar ao menos 320 espécies de vírus que possuem isolados capazes de infectar o ser humano. Grande parte destes são parasitas restritos à espécie humana (146/320 = 45,6%) e a maioria dos vírus causa algum tipo de patologia nos mais variados sítios, sendo os sistemas tegumentar, respiratório e nervoso os mais afetados, com um total de 92, 72 e 58 vírus associados respectivamente. Visto que os sistemas tegumentar e respiratório são os mais expostos ao ambiente, era esperado que fossem os mais afetados pelos vírus. Em contrapartida, o sistema nervoso está entre os mais protegidos e é curioso o fato de ser o terceiro mais afetado. Por se tratar de um sistema extremamente delicado e importante para o corpo humano, diversos estudos são realizados para identificar possíveis ameaças para seus componentes, o que leva consequentemente à descoberta de vários patógenos. Muitos destes vírus estão associados a quadros graves de meningite e encefalite, como herpesvírus, lyssavírus e flavivírus, estando em sob constante investigação sob a óptica epidemiológica e farmacológica (DAEP; MUNOZ-JORDAN; EUGENIN, 2014; GRANEROD *et al.*, 2010; JACKSON, 2016). Já os sistemas linfático e reprodutor foram os que apresentaram menos vírus associados. É provável que existam outros vírus que infectem os componentes desses sistemas, mas são necessários novos estudos que permitam sua identificação. Por fim, 83 vírus (26%) analisados neste trabalho não estão associados a nenhuma doença. A maioria destes é representada pelos torque teno vírus (família *Anelloviridae*), os quais tem sido apontados como parte do viroma humano não patogênico, juntamente com vários bacteriófagos (RASCOVAN; DURAISAMY; DESNUES, 2016). É notável que estes vírus já tenham sido identificados em diversas pessoas por meio de análises de metagenômica, bem como alguns vírus gigantes, como mimivírus e marseillevírus, mas sua possível

patogenia ainda está em debate e foge do propósito desta tese (COLSON *et al.*, 2013).

Dentre os diferentes grupos de hospedeiros analisados neste trabalho, os protistas estão entre aqueles que possuem menos vírus oficialmente classificados pelo ICTV. Destes fazem parte os vírus gigantes cuja descoberta foi um marco para a virologia, visto o rompimento de uma série de paradigmas bem aceitos até então (CLAVERIE; ABERGEL, 2016; FORTERRE, 2017). Duas características principais distinguem estes dos demais vírus: 1) possuem partículas com tamanhos acima do usualmente observado para vírus, chegando a dimensões micrométricas; 2) apresentam genoma extenso e complexo, contendo genes nunca antes observados na virosfera, como aminoacil-tRNA sintetases (aaRS), enzimas fundamentais no processo de tradução proteica (RIBAS DE POUPLANA; SCHIMMEL, 2001). Estas características desencadearam debates intensos sobre a origem do gigantismo viral e como se deu a história evolutiva destes vírus.

Os NCLDVs apresentam diversos formatos e tamanhos (tabela 1), e consequentemente diferentes estratégias para penetrar as células hospedeiras. Os vírus grandes (mas não gigantes) exploram distintas vias de penetração que independem do tamanho, como fusão de membranas (*Poxviridae*), endocitose mediada por receptores (*Iridoviridae*) e até mesmo a formação de poros na membrana celular do hospedeiro (*Phycodnaviridae*) (MOSS, 2016; WANG *et al.*, 2014; WILSON; VAN ETTEN; ALLEN, 2009). Já os vírus gigantes exploram a via fagocítica, a qual é desencadeada primariamente devido ao tamanho da partícula aderida ao fagócito (> 500 nm) (KORN; WEISMAN, 1967). A revisão dos mecanismos de penetração de cada grupo viral da proposta ordem "*Megavirales*" nos levou a propor uma hipótese em que o gigantismo viral teria emergido independentemente em cada grupo de vírus gigantes. O ancestral dos NCLDVs seria um vírus pequeno e adquiriu complexidade ao longo da evolução por meio de ganho e perda de genes, como defendido por diferentes autores (FILÉE, 2013; YUTIN; WOLF; KOONIN, 2014). O saldo positivo no ganho gênico levaria a um aumento do tamanho das partículas, promovendo a penetração por meio da fagocitose. Esta estratégia parece ter sido selecionada positivamente, visto que os hospedeiros conhecidos dos vírus gigantes são todos organismos fagotróficos, principalmente amebas de vida livre.

O debate sobre a origem dos vírus gigantes tem envolvido principalmente os componentes do aparato traducional observado nesses vírus. Neste aspecto, os membros da família *Mimiviridae* exibem a maior quantidade e diversidade destes elementos, em especial tRNAs e aaRS (ABRAHÃO *et al.*, 2018; ARSLAN *et al.*, 2011; SCHULZ *et al.*, 2017). Análises iniciais sugeriram que estes genes já estariam presentes nos ancestrais dos vírus gigantes e que este seria um organismo mais complexo, que teria coexistido com outros organismos celulares e evoluído por redução genômica, fazendo parte de um elusivo quarto domínio da vida (BOYER *et al.*, 2010; LEGENDRE *et al.*, 2012; NASIR; KIM; CAETANO-ANOLLES, 2012). Entretanto, análises filogenômicas têm indicado um cenário alternativo, em que estes genes teriam sido transferidos para os vírus gigantes a partir de hospedeiros e organismos simpátricos, sendo o ancestral um vírus menor e menos complexo cujo genoma expandiu por meio de um balanço positivo entre perda e ganho de genes (FILÉE, 2015; SCHULZ *et al.*, 2017). Ao analisar a distribuição dos genes de tRNA e aaRS nos membros do gênero *Mimivirus*, observamos que os vírus que possuem maior quantidade e diversidade gênica contém todos os genes presentes no resto do grupo, sugerindo uma perda gradual destes genes ao longo da evolução. Análises filogenéticas utilizando o método de máxima verossimilhança dos sete genes de aaRS presentes nos membros do gênero *Mimivirus* indicam que esses genes já estariam presentes no ancestral desses vírus gigantes, visto que eles se apresentam como grupo externo aos organismos eucariotos, resultados similares aos previamente descritos (BOYER *et al*., 2010). Para estas análises foram utilizados os 100 melhores *hits* após a busca em bancos de dados pelo programa Blastp utilizando os genes presentes em Megavirus chiliensis. É importante ressaltar que os métodos de reconstrução filogenética e principalmente as estratégias de obtenção de sequências para as análises influenciam fortemente o resultado das análises, o que explicaria os resultados discordantes em diversos trabalhos, levando a hipóteses de cenários divergentes para a origem e evolução dos vírus gigantes. Análises de uso preferencial de códons e aminoácidos dão suporte à hipótese de que os genes dos vírus gigantes não teriam sido oriundos por eventos de TGH, uma vez que o perfil exibido pelos vírus é muito diferente daquele observado para a célula hospedeira *Acanthamoeba castellanii*. Contudo, métodos para detecção destes eventos baseados em uso de códons tendem a gerar resultados controversos, não sendo os mais adequados para inferir a ocorrência da

transferência horizontal de genes (FRIEDMAN; ELY, 2012; KOSKI; MORTON; GOLDING, 2001). É possível que, caso a transferência gênica tenha ocorrido muito remotamente, o acúmulo de mutações ao longo da evolução resultaria em uma alteração no perfil de uso de códons e aminoácidos, não sendo, portanto, uma característica suficientemente adequada para encerrar o debate.

Com a descoberta dos klosneuvírus e tupanvírus, o arsenal de aaRS viral aumentou drasticamente. A presença de 20 aaRS nestes vírus foi um forte indício inicial de que a hipótese de redução genômica ganharia força. Análises filogenéticas baseadas nos genes presentes em klosneuvírus, entretanto, indicaram que apenas alguns destes genes estariam presentes em um ancestral viral, e a maioria teria sido transferido de organismos celulares (SCHULZ *et al.*, 2017). Por outro lado, análises iniciais com os genes de tupanvírus sugeriram um cenário oposto, uma vez que os vírus ficaram como grupo externo a organismos celulares em muitas árvores filogenéticas, levando à conclusão de que os genes já estariam presentes no ancestral dos vírus (ABRAHÃO *et al.*, 2018). Neste cenário de contradições e diante das características exclusivas observadas nos tupanvírus, decidimos realizar novas análises com estratégias alternativas para reconstrução filogenética utilizando genes de aaRS e revisitar a história da descoberta destes vírus que levou a ampliação da complexidade na virosfera. Ao buscar sequências homólogas em bancos de dados considerando grupos taxonômicos específicos (a saber, Eukarya, Archaea, Firmicutes, Protobacteria, Amoebozoa e Virus) para reconstrução filogenética em detrimento da estratégia de melhores *hits*, observamos árvores com topologia muito diferente do que havia sido previamente obtido. As novas análises levam à conclusão de que os tupanvírus e outros vírus gigantes teriam obtido a maioria dos genes de aaRS por meio de eventos de TGH. Além disso, análises filogenéticas utilizando o gene da citrato sintase, uma enzima chave no processo de geração de energia (necessária para o primeiro passo do ciclo do ácido cítrico e a primeira vez encontrada em um vírus), aplicando ora a estratégia de melhores *hits*, ora a busca por sequências homólogas em diferentes grupos taxonômicos, levaram a conclusões divergentes: no primeiro caso, o gene parecia estar presente em um ancestral dos vírus, diferente dos organismos celulares; no segundo, o gene teria sido obtido por meio de transferência a partir de bactérias. Esses dados demonstram que a história evolutiva dos vírus gigantes é muito complexa, e análises utilizando poucos genes não serão suficientes para definir questões importantes sobre origem e evolução dos

mesmos. A utilização da estratégia de busca de sequências homólogas em múltiplos grupos taxonômicos parece ser mais coerente, visto que a reconstrução filogenética é feita considerando uma ampla variedade de organismos, o que pode indicar com mais fidedignidade a ocorrência ou não de TGH entre os vírus e outros organismos. Neste sentido, é provável que os vírus gigantes tenham adquirido a maioria dos genes envolvidos no processo de tradução a partir de hospedeiros e organismos simpátricos, o que corroboraria a hipótese de origem a partir de um vírus mais simples cujo genoma foi se tornando mais complexo ao longo da evolução (YUTIN; WOLF; KOONIN, 2014).

Embora sejam filogeneticamente próximos dos mimivírus, os tupanvírus apresentam uma série de características que os distinguem dos demais gigantes, como uma partícula apresentando uma cauda cilíndrica acoplada ao capsídeo, recobertas por fibrilas mais curtas que as observadas para os mimivírus; efeito citopático caracterizado por agregação celular e posterior lise; diferentes perfis de interação com células hospedeiras, sendo capazes inclusive de infectar células e não multiplicar, mas causar um efeito tóxico e shutdown de rRNA. O fenômeno de shutdown foi descrito pela primeira vez em um vírus gigante, parece ser exclusivo dos tupanvírus, e seu mecanismo ainda não é conhecido, embora pareça estar envolvido com componentes presentes na partícula viral (ABRAHÃO *et al.*, 2018). Além disso, os tupanvírus apresentam diversos genes não observados nos mimivírus, incluindo inúmeras ORFans (genes sem similaridade com nenhuma sequência depositada em bancos de dados), e uma organização genômica distinta dos membros do gênero *Mimivirus* e *Cafeteriavirus*. Por fim, análises filogenéticas baseadas em diferentes genes conservados para os vírus gigantes colocam os tupanvírus em um ramo diferente dos mimivírus. Isso é reforçado pelo agrupamento hierárquico dos vírus baseado na presença e ausência de genes, o qual coloca os tupanvírus como um grupo-irmão dos mimivírus (ABRAHÃO *et al.*, 2018). Diante das claras diferenças entre tupanvírus e mimivírus, tanto biológicas quanto genômicas, foi realizada a proposta oficial ao ICTV para a criação de um novo grupo taxonômico para classificar adequadamente os novos vírus, o gênero *Tupanvirus*. A taxonomia dos vírus gigantes ainda é precária, muito devido à negligência dos grupos de pesquisa que descrevem novos vírus para com este ramo da ciência. É preciso estabelecer critérios e propor oficialmente para o ICTV a criação de novos táxons para que tenhamos maior clareza na taxonomia destes vírus.

O isolamento e a caracterização de novos vírus gigantes são objetivos centrais do nosso grupo de pesquisa. Diversos vírus já foram descritos oriundos de amostras coletadas em todas as regiões do Brasil, compreendendo distintos grupos virais, como mimivírus, marseillevírus, pandoravírus e cedratvírus (ANDRADE *et al.*, 2018; CAMPOS *et al.*, 2014; DORNAS *et al.*, 2016). Análises iniciais por microscopia eletrônica de transmissão (MET) e varredura de um novo cedratvírus isolado a partir de amostras de água acrescida de bioflocos coletadas em Belo Horizonte sugeriram um vírus menor que os demais já descritos para o grupo. Ao realizarmos a medição do tamanho e calcularmos o volume deste vírus baseado em dezenas de imagens oriundas de MET por contraste negativo, observamos que o novo isolado é significativamente menor que os demais cedratvírus. Além disso, o genoma possui aproximadamente 100.000 pares de bases a menos que o genoma dos demais vírus. Ao analisar a relação entre o volume das partículas virais e o tamanho do genoma, observamos que os pithovírus e cedratvírus estão no limite do intervalo de predição da escala alométrica, podendo ser exceções à regra, a qual prevê que vírus com menor volume tende a ter um genoma pequeno (CUI; SCHLUB; HOLMES, 2014). Visto que consideramos apenas dados oriundos de métodos de imagens comparáveis para calcular o volume dos vírus gigantes, como MET por contraste negativo e microscopia eletrônica criogênica, apenas alguns vírus foram analisados (DE CARLO; HARRIS, 2011; HOENGER; AEBI, 1996). É possível que com o surgimento de novos dados sobre a estrutura viral, em particular dos vírus gigantes, os cedratvírus e pithovírus fiquem realmente fora da regra de escala alométrica.

Os genes preditos para o novo vírus, denominado Brazilian cedratvirus IHUMI [em homenagem ao país de origem (Brasil) e o local onde a caracterização viral foi realizada (Institut Hopitalo-Universitaire Méditerranée Infection, França)], são similares aos descritos para outros cedratvírus. Entre os genes com função conhecida, observamos diferenças principalmente na quantidade daqueles codificando proteínas que contém domínios repetidos, com destaque para os genes com motivos repetidos de anquirina. Um estudo recente conduzido por Shukla e colaboradores (2018) demonstrou que, entre os vírus gigantes, a quantidade de genes desta classe é proporcional ao tamanho do genoma, o que também foi observado para bactérias intracelulares (SHUKLA; CHATTERJEE; KONDABAGIL, 2018). É possível que o mesmo padrão seja válido para os cedratvírus. Análises de identidade de aminoácidos e sintenia revelaram diferenças consideráveis entre o

vírus brasileiro e os demais cedratvírus, levando a hipótese de uma segunda linhagem para o grupo viral. Avaliação do pan-genoma reforça a hipótese, uma vez que ao adicionar o novo vírus foi constatado considerável aumento do pan-genoma e redução do genoma central, similar ao observado para mimivírus e marseillevírus (ASSIS *et al.*, 2017; DORNAS *et al.*, 2016). Junto a isso, análises filogenéticas utilizando diferentes genes conservados para os vírus gigantes colocaram o Brazilian cedratvirus IHUMI em um ramo diferente dos demais vírus, corroborando esta hipótese, o que nos levou a propor a criação da linhagem 'B' para os cedratvírus. Ainda não há dados suficientes para explorar adequadamente a história evolutiva dos cedratvírus, mas é possível que eles tenham um vírus menor como ancestral e evoluíram seguindo o modelo de acordeão descrito para outros vírus gigantes (FILÉE, 2013; YUTIN; WOLF; KOONIN, 2014). O isolamento e caracterização de novos vírus deste grupo emergente poderão trazer importantes informações a esse respeito.

A genômica tem contribuído fortemente para melhorar nossa compreensão sobre a diversidade e evolução dos vírus gigantes. Porém, para melhor compreender a biologia destes vírus, é preciso avançar em outras áreas da ciência e, nesse sentido, o uso da transcriptômica vem auxiliando a identificar peculiaridades dos NCLDVs (BLANC *et al.*, 2014; JIA *et al.*, 2017; LEGENDRE *et al.*, 2010). Neste trabalho realizamos o sequenciamento de RNA (RNA-seq) de células de *A. castellanii* infectadas com Marseillevirus marseillevirus, o vírus protótipo do gênero *Marseillevirus*, para obter uma visão geral do perfil de transcrição destes vírus. Todos os 457 genes inicialmente preditos para este vírus tiveram *reads* cognatas com cobertura suficiente para afirmar que são expressos ao longo do ciclo de multiplicação viral (> 10 *reads* por gene). O perfil temporal de expressão observado para os marseillevírus é similar ao descrito para outros NCLDVs, exibindo genes precoces, intermediários e tardios (OLIVEIRA *et al.*, 2017b). A expressão gênica dos marseillevírus é rápida, com todos os genes já expressos a partir de 4h, sendo que os genes tardios são expressos em altos níveis por pelo menos até 12h após infecção. É provável que a manutenção da expressão destes genes seja realizada no interior da fábrica viral, onde os mRNA estariam menos sujeitos a degradação pela maquinaria celular. Contudo, a transcrição de genes precoces parece ocorrer fora das fábricas virais, uma vez que estas são observadas no citoplasma das células hospedeiras recorrentemente a partir de 4h de infecção

(ARANTES *et al.*, 2016). Genes com funções diferentes possuem pico de expressão em momentos variados. Como esperado, genes relacionados aos processos de tradução e morfogênese viral são expressos em momentos tardios do ciclo, enquanto genes de replicação de DNA e transcrição são majoritariamente expressos em momentos mais iniciais do ciclo (em até 2h). Os genes de marseillevírus não estão organizados no genoma de acordo com suas funções nem mesmo em relação ao perfil temporal de expressão gênica. Tal característica reforça a ideia de mosaicismo genômico descrito para os marseillevírus, em que genes com funções e perfis de expressão diferentes teriam sido incorporados ao genoma viral em diferentes momentos ao longo da evolução, a partir de organismos distintos (BOYER *et al.*, 2009).

Assim como observado para outros NCLDVs, os marseillevírus possuem um motivo promotor conservado rico em AT, o qual está associado com mais de 50% dos genes (OLIVEIRA. *et al.*, 2017a). Não foi possível estabelecer uma relação direta entre a presença desse promotor com nenhuma classe funcional ou temporal de genes, semelhante ao observado para os mimivírus e poxvírus (DAVISON; MOSS, 1999; LEGENDRE *et al.*, 2010). Estes dados sugerem que a maquinaria para regulação transcricional dos marseillevírus é complexa e merece novas investigações. É possível que outros promotores virais (provavelmente sequências degeneradas) existam, as quais poderiam estar associadas à regulação específica de genes de alguma classe temporal, assim como descrito para os phycodnavírus (FITZGERALD *et al.*, 2008). Entretanto, estudos adicionais que tenham como objetivo definir as regiões terminais (5’ e 3’) dos transcritos virais devem ser realizados para confirmar esta hipótese. Ao optar pela estratégia de depleção do rRNA não foi possível determinar tais regiões. Embora a estratégia tenha sido eficiente para o propósito geral do estudo, ela não permite definir com acurácia as regiões terminais dos transcritos, diferentemente de outras estratégias, como a tecnologia SMRT (Single Molecule Real Time sequencing) (FISH *et al.*, 2014). Estudos futuros poderão avançar mais na transcriptômica destes vírus e revelar novas características que nos permitirão compreender melhor os mecanismos moleculares, especialmente no contexto de interação com seus hospedeiros e organismos simpátricos.

6. CONCLUSÕES

- Em geral, os vírus possuem um espectro de hospedeiros restrito;
- Conhecemos apenas uma pequena parte dos vírus existentes no planeta;
- A virosfera conhecida é enviesada pelo interesse humano: “Virosfera antropocêntrica”;
- Ambientes marinhos e hipersalinos apresentam a maior e a menor diversidade viral, respectivamente;
- Metavirômica associada à virologia clássica são fundamentais para ampliar nosso conhecimento sobre a virosfera;
- Pelo menos 320 espécies virais possuem membros capazes de infectar o ser humano, afetando principalmente os sistemas tegumentar, respiratório e nervoso;
- Grande parte dos vírus que afetam humanos não é patogênica;
- O mecanismo de penetração dos diferentes vírus gigantes sugere que o gigantismo viral pode ter evoluído independentemente em cada grupo;
- A distribuição dos elementos de tradução na família *Mimiviridae* sugere que estes vírus teriam um ancestral mais complexo e genes foram sendo perdidos ao longo da evolução;
- Diferentes estratégias para obtenção de sequências em bancos de dados afetam profundamente o resultado das árvores filogenéticas, levando a interpretações divergentes;
- Tupanvírus apresentam várias características distintas em relação aos demais membros da família *Mimiviridae* que justifica a criação de um novo grupo taxonômico para classifica-los adequadamente;

- Novos critérios para classificação de membros da família *Mimiviridae* são necessários para uma taxonomia eficiente;

- Novo cedratvírus brasileiro é menor do que os isolados da França e Argélia;

- Cedratvírus e pithovírus estão no limite de predição da escala alométrica entre volume da partícula e tamanho do genoma, podendo ser exceções à regra;

- Brazilian cedratvirus IHUMI possui um genoma menor e assintênico em relação aos demais cedratvirus;

- Os cedratvírus apresentam um pan-genoma aberto e a inclusão do isolado brasileiro reduz consideravelmente o genoma central;

- O isolado brasileiro é o primeiro membro de uma nova linhagem dos gigantes cedratvírus;

- Marseillevirus exibe um perfil temporal de transcrição gênica;

- Ensaios de RT-qPCR corroboraram dados de RNA-seq e confirmaram o perfil temporal de expressão dos genes de marseillevírus;

- Expressão de alguns genes é mantida pelo menos até 12h após a infecção;

- Genes de diferentes classes temporais estão homogeneamente distribuídos ao longo do genoma;

- Não há correlação entre a presença do promotor predito com as diferentes categorias temporais ou funcionais de genes;

- A maquinaria transcricional dos marseillevírus é mais complexa do que inicialmente imaginada, contando com fatores ainda não identificados.

7. CONSIDERAÇÕES FINAIS

- Após anos de investigação, nós conhecemos apenas uma pequena fração da virosfera, pois temos ignorado tudo exceto nós mesmos e os organismos relevantes para nós;

- A origem, evolução e taxonomia dos vírus gigantes estão em um campo aberto de pesquisa, e discussões acerca de metodologias e critérios são fundamentais para avançar nesse campo;

- Análises morfológicas e genômicas são essenciais para caracterizar um novo vírus e podem revelar aspectos inéditos da virosfera;

- A transcriptômica é uma área em constante evolução e sua aplicação é necessária para investigar o complexo maquinário transcricional dos vírus gigantes.

8. REFERÊNCIAS BIBLIOGRÁFICAS

ABERGEL, C. *et al.* Virus-Encoded Aminoacyl-tRNA Synthetases: Structural and Functional Characterization of Mimivirus TyrRS and MetRS. *Journal of Virology*, v. 81, n. 22, p. 12406–12417, 2007.

ABERGEL, C.; LEGENDRE, M.; CLAVERIE, J. M. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiology Reviews*, v. 39, n. 6, p. 779–796, 2015.

ABRAHÃO, J. *et al.* Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nature Communications*, v. 9, n. 1, p. 749, 2018.

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, 2014.

AHERFI, S. *et al.* The expanding family Marseilleviridae. *Virology*, v. 466–467, p. 27–37, 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. 1–10, 2018.

ANDRADE, A. C. DOS S. P. *et al.* Filling Knowledge Gaps for Mimivirus Entry, Uncoating, and Morphogenesis. *Journal of Virology*, v. 91, n. 22, p. e01335-17, 2017.

ANDREANI, J. *et al.* Cedratvirus, a double-cork structured giant virus, is a distant relative of pithoviruses. *Viruses*, v. 8, n. 11, p. 1–11, 2016.

ANDREANI, J. *et al.* Orpheovirus IHUMI-LCC2: A new virus among the giant viruses. *Frontiers in Microbiology*, v. 8, n. January, p. 1–11, 2018.

ANDREANI, J. *et al.* Pacmanvirus, a New Giant Icosahedral Virus at the Crossroads

between Asfarviridae and Faustoviruses. *Journal of Virology*, v. 91, n. 14, p. e00212-17, 2017.

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, 2016.

ARSLAN, D. *et al.* Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proceedings of the National Academy of Sciences*, v. 108, n. 42, p. 17486–17491, 2011.

ASSIS, F. L. *et al.* Genome characterization of the first mimiviruses of lineage C isolated in Brazil. *Frontiers in Microbiology*, v. 8, n. DEC, p. 1–11, 2017.

BAJRAI, L. H. *et al.* Kaumoebavirus, a new virus that clusters with Faustoviruses and Asfarviridae. *Viruses*, v. 8, n. 11, 2016.

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

BERTELLI, C. *et al.* Cedratvirus lausannensis – digging into Pithoviridae diversity. *Environmental Microbiology*, v. 19, n. 10, p. 4022–4034, 2017.

BLANC, G. *et al.* Deep RNA sequencing reveals hidden features and dynamics of early gene transcription in Paramecium bursaria chlorella virus 1. *PLoS ONE*, v. 9, n. 3, 2014.

BOYER, M. *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*, v. 106, n. 51, p. 21848–21853, 2009.

BOYER, M. *et al.* Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4th domain of life including giant viruses. *PLoS ONE*, v. 5, n. 12, 2010.

BYRNE, D. *et al.* The polyadenylation site of Mimivirus transcripts obeys a stringent “hairpin rule”. *Genome Research*, v. 19, n. 7, p. 1233–1242, 2009.

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, 2014.

CLAVERIE, J. M.; OGATA, H. Ten good reasons not to exclude giruses from the evolutionary picture. *Nature Reviews. Microbiology*, v. 7, n. 8, p. 615; author reply 615, 2009.

CLAVERIE, J. M.; ABERGEL, C. Giant viruses: The difficult breaking of multiple epistemological barriers. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, v. 59, p. 89–99, 2016.

COLSON, P. *et al.* Evidence of the megavirome in humans. *Journal of Clinical Virology*, v. 57, n. 3, p. 191–200, 2013.

COLSON, P. *et al.* “Marseilleviridae”, a new family of giant viruses infecting amoebae. *Archives of Virology*, v. 158, n. 4, p. 915–920, 2013.

COLSON, P. *et al.* “Megavirales”, a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Archives of Virology*, v. 158, n. 12, p. 2517–2521, 2013.

COLSON, P.; LA SCOLA, B.; RAOULT, D. Giant viruses of amoebae: A journey through innovative research and paradigm changes. *Annual Review of Virology*, v. 4, n. 1, p. annurev-virology-101416-041816, 2017.

CUI, J.; SCHLUB, T. E.; HOLMES, E. C. An allometric relationship between the genome length and virion volume of viruses. *Journal of Virology*, v. 88, n. 11, p. 6403–6410, 2014.

DAEP, C. A.; MUNOZ-JORDAN, J. L.; EUGENIN, E. A. Flaviviruses, an expanding threat in public health: focus on dengue, West Nile, and Japanese encephalitis virus. *Journal of Neurovirology*, v. 20, n. 6, p. 539–560, 2014.

DAVISON, A. J.; MOSS, B. Structure of vaccinia virus late promoters. *Journal of Molecular Biology*, v. 210, n. 4, p. 771–784, 1989.

DE CARLO, S.; HARRIS, J. R. Negative staining and cryo-negative staining of macromolecules and viruses for TEM. *Micron*. [S.l: s.n.]. , 2011

DE CASTRO, I. F.; VOLONTÉ, L.; RISCO, C. Virus factories: Biogenesis and structural design. *Cellular Microbiology*. [S.l: s.n.]. , 2013

DIETZGEN, R. G.; MANN, K. S.; JOHNSON, K. N. Plant virus-insect vector interactions: Current and potential future research directions. *Viruses*. [S.l: s.n.]. , 2016

DORNAS, F. P. *et al.* A Brazilian marseillevirus is the founding member of a lineage in family marseilleviridae. *Viruses*, v. 8, n. 3, 2016.

EBRAHIMI-MAMEGHANI, M. *et al.* Glucose homeostasis, insulin resistance and inflammatory biomarkers in patients with non-alcoholic fatty liver disease: Beneficial effects of supplementation with microalgae Chlorella vulgaris: A double-blind placebo-controlled randomized clinical trial. *Clinical Nutrition*, v. 36, n. 4, p. 1001–1006, 2017.

ENQUIST, L. W.; RACANIELLO, V. R. Virology: From Contagium Fluidum to Virome. In: *Fields Virology*. 6th. ed. Philadelphia: Lippincott Williams & Wilkins, 2013. p. 1–20.

FABRE, E. *et al.* Noumeavirus replication relies on a transient remote control of the host nucleus. *Nature Communications*, p. 1–57, 2017.

FILÉE, J. Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution. *Frontiers in Microbiology*, v. 6, n. MAY, p. 1–13,

2015.

FILÉE, J. Route of NCLDV evolution: The genomic accordion. *Current Opinion in Virology*, v. 3, n. 5, p. 595–599, 2013.

FISH, R. *et al.* An Improved cDNA Library Generation Protocol for Transcriptome Analysis from a Single Cell. *J Biomol Tech*, v. 25(Suppl), n. S16, 2014.

FITZGERALD, L. A. *et al.* Putative gene promoter sequences in the chlorella viruses. *Virology*, v. 380, n. 2, p. 388–393, 2008.

FORTERRE, P. Giant viruses: Conflicts in revisiting the virus concept. *Intervirology*, v. 53, n. 5, p. 362–378, 2010.

FORTERRE, P. Viruses in the 21st Century: From the Curiosity-Driven Discovery of Giant Viruses to New Concepts and Definition of Life. *Clinical Infectious Diseases*, v. 65, n. Suppl 1, p. S74–S79, 2017.

FRIEDMAN, R.; ELY, B. Codon usage methods for horizontal gene transfer detection generate an abundance of false positive and false negative results. *Current Microbiology*, v. 65, n. 5, p. 639–642, 2012.

GALLONE, B. *et al.* Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. *Cell*, v. 166, n. 6, p. 1397–1410.e16, 2016.

GHIGO, E. *et al.* Ameobal pathogen mimivirus infects macrophages through phagocytosis. *PLoS Pathogens*, v. 4, n. 6, 2008.

GRANEROD, J. *et al.* Causes of encephalitis and differences in their clinical presentations in England: A multicentre, population-based prospective study. *The Lancet Infectious Diseases*, v. 10, n. 12, p. 835–844, 2010.

HOENGER, A.; AEBI, U. 3-D reconstructions from ice-embedded and negatively stained biomacromolecular assemblies: A critical comparison. *Journal of Structural*

Biology, v. 117, n. 2, p. 99–116, 1996.

HÜLSMEIER, A. J.; HENNET, T. O-Linked glycosylation in Acanthamoeba polyphaga mimivirus. *Glycobiology*, v. 24, n. 8, p. 703–714, 2014.

ICTV. *International Committee in Taxonomy of Viruses*. Disponível em: <https://talk.ictvonline.org/>. Acesso em: 7 nov. 2018.

IYER, L. M.; ARAVIND, L.; KOONIN, E. V. Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses. *Journal of Virology*, v. 75, n. 23, p. 11720–11734, 2001.

JACKSON, A. C. Human Rabies: a 2016 Update. *Current Infectious Disease Reports*, v. 18, n. 11, p. 1–6, 2016.

JEUDY, S. *et al.* Translation in Giant Viruses: A Unique Mixture of Bacterial and Eukaryotic Termination Schemes. *PLoS Genetics*, v. 8, n. 12, 2012.

JIA, H. *et al.* Transcriptome analysis of sheep oral mucosa response to Orf virus infection. *PLoS ONE*, v. 12, n. 10, p. 1–14, 2017.

KOONIN, E. V.; KRUPOVIC, M.; YUTIN, N. Evolution of double-stranded DNA viruses of eukaryotes: From bacteriophages to transposons to giant viruses. *Annals of the New York Academy of Sciences*, v. 1341, n. 1, p. 10–24, 2015.

KORB, V. C.; CHUTURGOON, A. A.; MOODLEY, D. *Mycobacterium tuberculosis:* Manipulator of protective immunity. *International Journal of Molecular Sciences*. [S.l: s.n.]. , 2016

KORN, E. D.; WEISMAN, R. A. Phagocytosis of latex beads by Acanthamoeba. II. Electron microscopic study of the initial events. *Journal of Cell Biology*, v. 34, n. 1, p. 219–227, 1967.

KOSKI, L. B.; MORTON, R. A.; GOLDING, G. B. Codon bias and base composition

are poor indicators of horizontally transferred genes. *Molecular Biology and Evolution*, v. 18, n. 3, p. 404–412, 2001.

KRISHNAMURTHY, S. R.; WANG, D. Origins and Challenges of Viral Dark Matter. *Virus Research*, 2017.

KRISTENSEN, D. M. *et al.* New dimensions of the virus world discovered through metagenomics. *Trends in Microbiology*, v. 18, n. 1, p. 11–19, 2010.

LA SCOLA, B. *et al.* A giant virus in amoebae. *Science*, v. 299, n. 5615, p. 2033, 2003.

LEGENDRE, M. *et al.* Genomics of megavirus and the elusive fourth domain of life. *Communicative and Integrative Biology,* v. 5, n. 1, p. 102–106, 2012.

LEGENDRE, M. *et al.* In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. *Proceedings of the National Academy of Sciences*, v. 112, n. 38, p. 5327–5335, 2015.

LEGENDRE, M. *et al.* mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Research*, v. 20, n. 5, p. 664–674, 2010.

LEGENDRE, M. *et al.* Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proceedings of the National Academy of Sciences*, v. 111, n. 11, p. 201320670, 2014.

LEVASSEUR, A. *et al.* Comparison of a modern and fossil *Pithovirus* reveals its genetic conservation and evolution. *Genome Biology and Evolution*, p. 153, 2016.

LWOFF, A. The Concept of Virus The Third Marjory Stephenson Memorial Lecture. *Journal of General Microbiology*, v. 17, n. 17, p. 239–253, 1957.

MOREIRA, D.; BROCHIER-ARMANET, C. Giant viruses, giant chimeras: the multiple

evolutionary histories of Mimivirus genes. *BMC Evol Biol*, v. 8, p. 12, 2008.

MOSS, B. Membrane fusion during poxvirus entry. *Seminars in Cell and Developmental Biology*, v. 60, p. 89–96, 2016.

NASIR, A.; KIM, K. M.; CAETANO-ANOLLES, G. Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. *BMC Evolutionary Biology*, v. 12, n. 1, p. 156, 2012.

O'MALLEY, M. A. "Everything is everywhere: but the environment selects": ubiquitous distribution and ecological determinism in microbial biogeography. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, v. 39, n. 3, p. 314–325, 2008.

OKAMOTO, K. *et al.* Structural variability and complexity of the giant Pithovirus sibericum particle revealed by high-voltage electron cryo-Tomography and energy-filtered electron cryo-microscopy. *Scientific Reports*, v. 7, n. 1, p. 1–12, 2017.

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, 2017a.

OLIVEIRA, G. *et al.* Promoter Motifs in NCLDVs: An Evolutionary Perspective. *Viruses*, v. 9, n. 1, p. 16, 2017b.

PAEZ-ESPINO, D. *et al.* Uncovering Earth's virome. *Nature*, v. 536, n. 7617, p. 425–430, 2016.

PANAHI, Y. *et al.* Chlorella vulgaris: A Multifunctional Dietary Supplement with Diverse Medicinal Properties. *Current pharmaceutical design*, v. 22, n. 2, p. 164–73, 2016.

PHILIPPE, N. *et al.* Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science*, v. 341, n. 6143, p. 281–6, 2013.

PIACENTE, F. *et al.* The autonomous glycosylation of large dna viruses. *International Journal of Molecular Sciences*, v. 16, n. 12, p. 29315–29328, 2015.

RAOULT, D. *et al.* The 1.2-Megabase genome sequence of Mimivirus. *Science*, v. 306, n. 2004, p. 1344–1350, 2004.

RAOULT, D.; FORTERRE, P. Redefining viruses: lessons from Mimivirus. *Nature Reviews Microbiology.*, v. 6, n. 1740–1534 (Electronic), p. 315–319, 2008.

RASCOVAN, N.; DURAISAMY, R.; DESNUES, C. Metagenomics and the Human Virome in Asymptomatic Individuals. *Annual Reviews in Microbiology*, v. 70, p. 125–141, 2016.

REGANOLD, J. P.; WACHTER, J. M. Organic agriculture in the twenty-first century. *Nature Plants*, v. 2, n. February, p. 15221, 2016.

REMNANT, E. J. *et al.* A Diverse Range of Novel RNA Viruses in Geographically Distinct Honey Bee Populations. *Journal of Virology*, p. JVI.00158-17, 2017.

RETENO, D. G. *et al.* Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *Journal of virology*, v. 89, n. 13, p. 6585–94, 2015.

RIBAS DE POUPLANA, L.; SCHIMMEL, P. Aminoacyl-tRNA synthetases: Potential markers of genetic code development. *Trends in Biochemical Sciences*. [S.l: s.n.]. , 2001

RODRIGUES, R. A. L. *et al.* Mimivirus Fibrils Are Important for Viral Attachment to the Microbial World by a Diverse Glycoside Interaction Repertoire. *Journal of Virology*, v. 89, n. 23, p. 11812–11819, 2015.

RUSSO, A. G. *et al.* Viral Discovery in the Invasive Australian Cane Toad (Rhinella marina) Using Metatranscriptomic and Genomic Approaches. *Journal of Virology*, v. 92, n. 17, p. 1–18, 2018.

SCHULZ, F. *et al.* Giant viruses with an expanded complement of translation system components. *Science*, v. 356, n. 6333, p. 82–85, 2017.

SHI, M. *et al.* Redefining the invertebrate RNA virosphere. *Nature*, p. 1–12, 2016.

SHUKLA, A.; CHATTERJEE, A.; KONDABAGIL, K. The number of genes encoding repeat domain-containing proteins positively correlates with genome size in amoebal giant viruses. *Virus Evolution*, v. 4, n. 1, p. 1–11, 2018.

SICARD, D.; LEGRAS, J. L. Bread, beer and wine: Yeast domestication in the Saccharomyces sensu stricto complex. *Comptes Rendus - Biologies*. [S.l: s.n.]. , 2011

SILVA, L. K. D. S. *et al.* Cedratvirus getuliensis replication cycle: An in-depth morphological analysis. *Scientific Reports*, v. 8, n. 1, p. 1–11, 2018.

SIMMONDS, P. *et al.* Consensus statement: Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*, v. 15, n. 3, p. 161–168, 2017.

SUHRE, K.; AUDIC, S.; CLAVERIE, J.-M. J.-M. Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, v. 102, n. 41, p. 14689–14693, 2005.

SUTTLE, C. A. Marine viruses--major players in the global ecosystem. *Nature Reviews Microbiology*, v. 5, n. 10, p. 801–812, 2007.

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, 2016.

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

THORNTON, P. K. Livestock production: Recent trends, future prospects. *Philosophical Transactions of the Royal Society B: Biological Sciences*. [S.l: s.n.]. , 2010

VILA, J. *et al. Escherichia coli:* An old friend with new tidings. *FEMS Microbiology Reviews*. [S.l: s.n.]. , 2016

WANG, S. *et al.* Entry of a novel marine DNA virus, Singapore grouper iridovirus, into host cells occurs via clathrin-mediated endocytosis and macropinocytosis in a pH-dependent manner. *Journal of virology*, v. 88, n. 22, p. 13047–63, 2014.

WILSON, W. H.; VAN ETTEN, J. L.; ALLEN, M. J. The Phycodnaviridae: The story of how tiny giants rule the world. *Current Topics in Microbiology and Immunology*, v. 328, p. 1–42, 2009.

XIAO, C. *et al.* Structural studies of the giant Mimivirus. *PLoS Biology*, v. 7, n. 4, p. 0958–0966, 2009.

YUTIN, N. *et al.* Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virology journal*, v. 6, n. 3, p. 223, 2009.

YUTIN, N.; WOLF, Y. I.; KOONIN, E. V. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology*, v. 466–467, p. 38–52, 2014.

ZAUBERMAN, N. *et al.* Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. *PLoS Biology*, v. 6, n. 5, p. 1104–1114, 2008.

ZHANG, Y. Z.; SHI, M.; HOLMES, E. C. Using Metagenomics to Characterize an Expanding Virosphere. *Cell*, v. 172, n. 6, p. 1168–1172, 2018.

9. TRABALHOS PUBLICADOS DURANTE O DOUTORADO

9.1. Artigos científicos publicados em periódicos indexados

1. **Rodrigues RAL**, dos Santos Silva LK, Dornas FP, de Oliveira DB, Magalhães TF, Santos DA, Costa AO, de Macêdo Farias L, Magalhães PP, Bonjardim CA, Kroon EG, La Scola B, Cortines JR, Abrahão JS. Mimivirus fibrils are important for viral attachment to microbial world by a diverse glycoside interaction repertoire. ***Journal of Virology***, 89(23): 11812-9, 2015.

2. **Rodrigues RAL**, Abrahão JS, Drumond BP, Kroon EG. Giants among larges: how gigantism impacts giant virus entry into amoebae. ***Current Opinion in Microbiology***, 31:88-93, 2016.

3. Arantes TS, **Rodrigues RAL**, Dos Santos Silva LK, Oliveira GP, de Souza HL, Khalil JY, de Oliveira DB, Torres AA, da Silva LL, Colson P, Kroon EG, da Fonseca FG, Bonjardim CA, La Scola B, Abrahão JS. The large marseillevirus explores different entry pathways by forming giant infectious vesicles. ***Journal of Virology***, 90(11): 5246-55; 2016.

4. Abrahão JS*, **Araújo R***, Colson P, La Scola B. The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. ***PLoS Genetics***, 13(2): e1006532, 2017.
   *Estes autores contribuíram igualmente para o trabalho.

5. Oliveira GP, Andrade AC, **Rodrigues RAL**, Arantes TS, Boratto PV, Silva LK, Dornas FP, Trindade GS, Drumond BP, La Scola B, Kroon EG, Abrahão JS. Promoter Motifs in NCLDVs: An Evolutionary Perspective. ***Viruses***, 9(1): 16, 2017.

6. **Rodrigues RAL**, Andrade ACDSP, Boratto PVM, Trindade GS, Kroon EG, Abrahão JS. An Anthropocentric View of the Virosphere-Host Relationship. ***Frontiers in Microbiology***, 8: 1673, 2017.

7. Oliveira GP, Lima MT, Arantes TS, Assis FL, **Rodrigues RAL**, da Fonseca FG, Bonjardim CA, Kroon EG, Colson P, La Scola B, Abrahão JS. The Investigation of Promoter Sequences of Marseilleviruses Highlights a Remarkable Abundance of the AAATATTT Motif in Intergenic Regions. ***Journal of Virology***, 91(21): e01088-17, 2017.

8. Andrade ACDSP*, **Rodrigues RAL***, Oliveira GP*, Andrade KR, Bonjardim CA, La Scola B, Kroon EG, Abrahão JS. Filling Knowledge Gaps for Mimivirus Entry, Uncoating, and Morphogenesis. ***Journal of Virology***, 91(22): e01335-17, 2017.
*Estes autores contribuíram igualmente para o trabalho.

9. Silva FC, Rodrigues VG, Duarte LP, Lula IS, Sinisterra RD, Vieira-Filho SA, **Rodrigues RAL**, Kroon EG, Oliveira PL, Farias LM, Magalhães PP, Silva GDF. Antidiarrheal activity of extracts from Maytenus gonoclada and inhibition of Dengue virus by lupeol. ***Anais da Academia Brasileira de Ciências***, 89(3): 1555-1564, 2017.

10. Oliveira GP*, **Rodrigues RAL***, Lima MT*, Drumond BP, Abrahão JS. Poxvirus Host Range Genes and Virus-Host Spectrum: A Critical Review. ***Viruses***, 9(11): e331, 2017.
*Estes autores contribuíram igualmente para o trabalho.

11. Boratto PVM, Dornas FP, da Silva LCF, **Rodrigues RAL**, Oliveira GP, Cortines JR, Drumond BP, Abrahão JS. Analyses of the Kroon Virus Major Capsid Gene and Its Transcript Highlight a Distinct Pattern of Gene Evolution and Splicing among Mimiviruses. ***Journal of Virology***, 92(2): e01782-17.

12. Andrade ACDSP, Arantes TS, **Rodrigues RAL**, Machado TB, Dornas FP, Landell MF, Furst C, Borges LGA, Dutra LAL, Almeida G, Trindade GS, Bergier I, Abrahão W, Borges IA, Cortines JR, de Oliveira DB, Kroon EG, Abrahão JS. Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. ***Virology Journal***, 15(1): 22, 2018.

13. Abrahão J, Silva L, Silva LS, Khalil JYB, **Rodrigues R**, Arantes T, Assis F, Boratto P, Andrade M, Kroon EG, Ribeiro B, Bergier I, Seligmann H, Ghigo E, Colson P, Levasseur A, Kroemer G, Raoult D, La Scola B. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. ***Nature communications***, 9(1): 749, 2018.

14. Silva LKDS, Andrade ACDSP, Dornas FP, **Rodrigues RAL**, Arantes T, Kroon EG, Bonjardim CA, Abrahão JS. Cedratvirus getuliensis replication cycle: an in-depth morphological analysis. ***Scientific Reports***, 8(1): 4000, 2018.

15. **Rodrigues RAL**, Andreani J, Andrade ACDSP, Machado TB, Abdi S, Levasseur A, Abrahão JS, La Scola B. Morphologic and Genomic Analyses of New Isolates Reveal a Second Lineage of Cedratviruses. ***Journal of Virology***, 92(13): e00372-18, 2018.

16. **Rodrigues RAL**, Mougari S, Colson P, La Scola B, Abrahão JS. "Tupanvirus", a new genus in the family Mimiviridae. ***Archives of Virology***, epub ahead of print, 2018.

17. **Rodrigues RAL**, Arantes TS, Oliveira GP, Silva LKS, Abrahão JS. The complex nature of Tupanviruses. ***Advances in Virus Research***, in press, corrected proof, 2018.

9.2 Artigos científicos aceitos para publicação

1. Andrade ACSP, Boratto PVM, **Rodrigues RAL**, Bastos TM, Azevedo BL, Dornas FP, Oliveira DB, Drumond BP, Kroon EG, Abrahão JS. New isolates of pandoraviruses: contribution to the study of replication cycle steps. ***Journal of Virology***, aceito para publicação em novembro de 2018.

9.3. Artigos científicos em avaliação para publicação

1. **Rodrigues RAL**, Cherif-Louazani A, Colson P, La Scola B, Abrahão JS. The analysis of Marseillevirus transcriptome reveals a temporal profile of gene expression. Manuscrito em fase final de análise e redação para submissão posterior às considerações da banca avaliadora.

2. Boratto PVM, Andrade ACSP, **Rodrigues RAL**, La Scola B, Abrahão JS. The multiple origins of proteins present in tupanvirus particles. Manuscrito em análise, submetido ao periódico ***Current Opinion in Virology*** em novembro de 2018.

9.4. Texto publicado em jornais ou revistas

- **Rodrigues RAL**, Abrahão JS. Redefining life. ***The Biologist***, 63(2): 12-15, 2016. Royal Society of Biology, United Kingdom.

10. ATIVIDADES DESENVOLVIDAS DURANTE O DOUTORADO

10.1. Estágio doutoral no exterior

- Aix-Marseille Université, Institut Hopitalo-Universitaire Méditerranée Infection, Marseille, França. Estágio doutoral sob a supervisão do Dr. Bernard La Scola. Período: Agosto/2017 a Julho/2018. Bolsista pelo programa Capes-Cofecub.

10.2. Participação de eventos científicos

- XXVI Brazilian Congress of Virology and X Mercosur Meeting of Virology. Período: 11 a 14 de outubro de 2015. Florianópolis, SC, Brasil.

- II Simpósio de Microbiologia da UFMG – Microbiologia translacional: do ambiente natural às aplicações biotecnológicas. Período: 05 e 06 de outubro de 2015. Belo Horizonte, MG, Brasil.

- III Simpósio de Microbiologia da UFMG – Doenças microbianas emergentes. Período: 05 e 06 de setembro de 2016. Belo Horizonte, MG, Brasil.

- V Simpósio de Microbiologia da UFMG – Desafios atuais no enfrentamento de doenças microbianas. Período: 05 e 06 de setembro de 2018. Belo Horizonte, MG, Brasil.

10.3. Organização de eventos científicos

- II Simpósio de Microbiologia da UFMG – Microbiologia translacional: do ambiente natural às aplicações biotecnológicas. Período: 05 e 06 de outubro de 2015. Belo Horizonte, MG, Brasil.

10.4. Participação em banca de conclusão de curso

- **Rodrigues, RAL**; Oliveira, DB; Ferreira, PCP. Participação em banca de Lucas Moreira Botelho. Níveis de mRNA de genes da família 2'5'OAS em pacientes com esclerose sistêmica. 2015. Trabalho de Conclusão de Curso (Graduação em Ciências Biológicas) - Universidade Federal de Minas Gerais.

10.5. Disciplinas cursadas e aproveitamento

Período	Disciplina	Turma	Nota	Conceito	Créditos
2015/2	Dispensa de crédito*	DCR GER000	-	-	27
2015/2	Seminários	DIP MIC808 1	-	-	02
2015/2	Tópicos especiais em microbiologia: metagenômica	DIP MIC836 B	95	A	03
2015/2	Micro-organismos patogênicos	DIP MIC853 A	90	A	03
2016/2	Tópicos especiais em genética e evolução II: DNAs repetitivos: organização, função e evolução	DIP BIG847 A	98	A	02
2017/1	Tópicos especiais em microbiologia: Isolamento e identificação de micro-organismos: bases para a taxonomia	DIP MIC836 A	100	A	03

*Disciplinas concluídas durante o Mestrado em Microbiologia pela UFMG: Virologia básica, vírus com importância na saúde humana, técnicas básicas em virologia, bacteriologia de anaeróbios, bacteriologia de aeróbios, biologia molecular de micro-organismos, biossegurança e bioética, biologia de leveduras, imunidade inata a infecções microbianas, antimicrobianos, treinamento didático em microbiologia.

ANEXOS – Outros trabalhos produzidos durante o doutorado

ANEXOS

Larger than life

Rodrigo Rodrigues and Jônatas Abrahão look at so-called giant viruses – and how their unique features have forced us to reconsider whether or not viruses are living organisms

Viruses were discovered towards the end of the 19th century, thanks to the pioneering studies of Louis Pasteur, and Dmitri Iwanowsky and Martinus Beijerinck's separate work on tobacco mosaic disease. They were first defined as nanoscopic filterable agents with a maximum size of about 200nm.

In 1957, André Lwoff published a paper[1] on the basic features for an organism to be considered a virus: infectious; potentially pathogenic; but presenting only one kind of nucleic acid, DNA or RNA; unable to grow and replicate by binary fission, and lacking its own metabolic machinery.

Viruses have also traditionally been considered a group of polyphyletic microorganisms (having many origins) and kept out of the tree of life. Are all these definitions still valid to virologists and biologists today? A discovery in 2003 changed most of them.

The year was a milestone for virology. A new virus, called *Acanthamoeba polyphaga mimivirus*, was the first of many pathogens to be discovered that are now known as giant viruses.

The history of giant viruses began in 1992, in Bradford, England. Researchers were working on a pneumonia outbreak and in their search for pathogenic microorganisms they found an organism similar to a Gram-positive coccus, isolated in a culture of the amoeba *Acanthamoeba polyphaga* in a water sample from a hospital cooling tower. The pathogen, named *Bradfordcoccus*, was impossible to label with classical techniques for bacterial identification.

After several attempts to identify the new microorganism, the original sample was stored away until the early 2000s, when it was taken to the University of Marseille, France. There, a team led by Dr Bernard La Scola and Dr Didier Raoult resumed the investigation.

They decided to look directly at it using transmission electron microscopy. To the entire scientific community's surprise, *Bradfordcoccus* was actually an enormous virus, with very unusual structures never seen before in the virosphere.[2]

It was renamed Acanthamoeba polyphaga mimivirus (APMV), a reference to the host in which it was first isolated and the fact it was able to mimic a microorganism. For its unusual features to be properly accommodated, a new viral family was created, named Mimiviridae. The APMV presented some features that suggested it was related to other large viruses, generically named as nucleo-cytoplasmic large DNA viruses

To the entire scientific community's surprise, it was actually an enormous virus with structures never seen before

An illustration showing a giant virus (blue) being infected by smaller virophages

(NCLDV, a proposed order named Megavirales). This group comprises the families Poxviridae (including Variola virus), Iridoviridae (insect-infecting viruses) and Phycodnaviridae (algae-infecting viruses), among others, as well as some newly discovered but unclassified giant viruses, such as *Pithovirus*, *Pandoravirus* and *Mollivirus*.

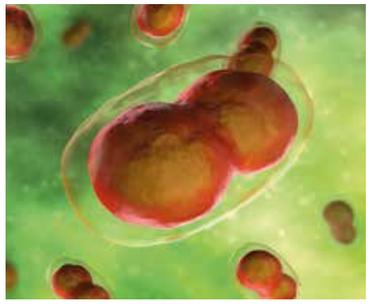
After analysis of the genome of APMV, this hypothesis was confirmed, grouping it definitively with the other large viruses. In addition, the discovery of other mimiviruses in many parts of the world (North and South America, Europe and Africa so far), legitimised the existence of the Mimiviridae family.

So what makes this virus so different? At about 750nm in diameter, it was far bigger than any other known viruses at the time of its discovery, and even bigger than some bacteria such as *Mycoplasma* and *Rickettsia*. Its capsid is pseudo-icosahedral (not the classic icosahedron), due to a star-shaped structure in one of the capsid vertex, named stargate, which is responsible for the genome release into the cytoplasm of a host cell.

Inside the capsid there is a lipid membrane, surrounding a proteic core that conceals the genome. Also, the capsid is covered by a dense layer of glycoproteic fibrils of approximately 125nm in length. The only region that lacks fibrils is the stargate, and one can easily see this in a scanning electron microscopy image. These fibrils are immersed in a peptidoglycan matrix, which is why the mimivirus was considered positive in the Gram staining.

Considering the virus's size and the presence of peptidoglycan, it is not surprising that those researchers in 1992 thought they were dealing with a bacterium.

Thanks to the fibrils, the mimiviruses are capable of attaching to several environmental organisms, such as amoebae, Gram-positive bacteria, fungi and arthropods.[3]



***Variola* is in the order Megavirales, which shares some features with mimiviruses**

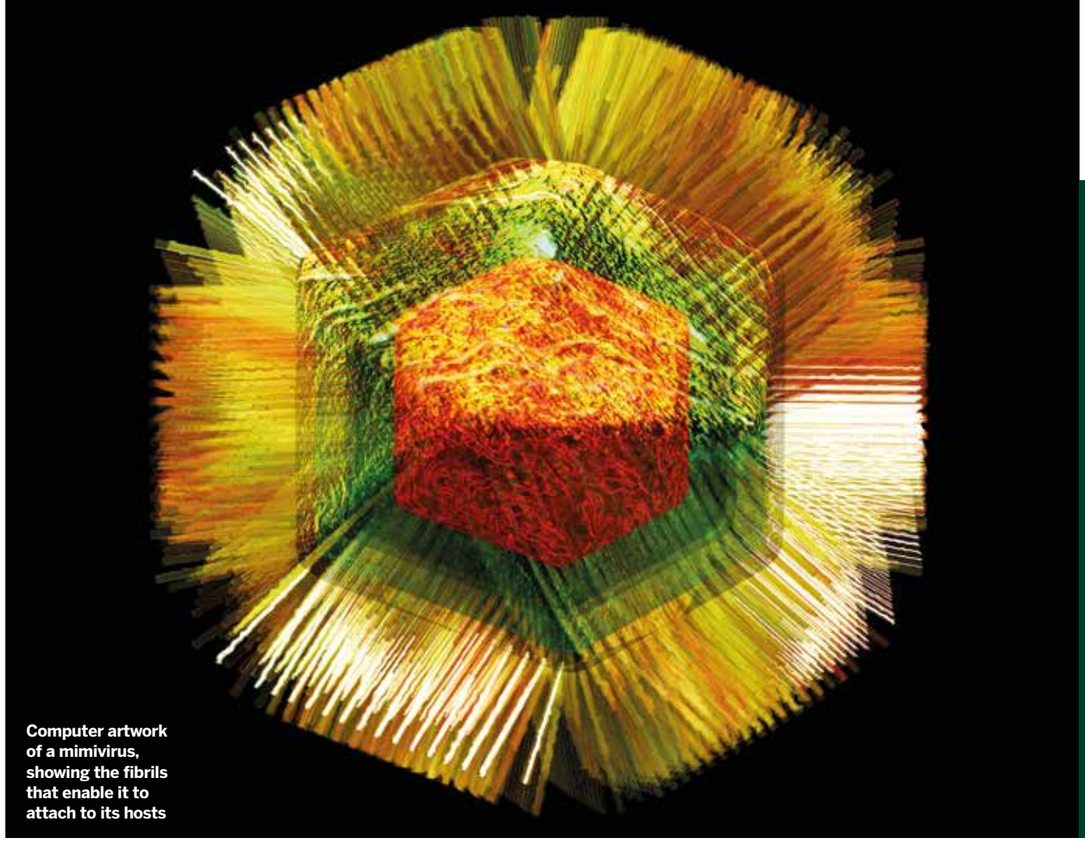
Coloured transmission electron micrograph of virions of the tobacco mosaic virus

Their ability to attach to arthropods such as crustaceans and mosquitoes might account for the ubiquity of these viruses, as these animals could disperse the giant viruses to many places on the planet, and also come into contact with humans.

The differences are not restricted to the viruses' structure. The genome of APMV is also distinct and complex. Its linear double-strand DNA molecule is 1.2 million base pairs – larger than the bacterium *Mycoplasma pneumonia* and most viruses, whose genomes are normally measured in thousands of base pairs.

The APMV genome's content is impressive too. Among its almost 1,000 genes, genes related to protein translation apparatus (normally associated only with cellular organisms) were found. There were also enzymes related to DNA repair (e.g. endonucleases), RNA modification (e.g. tRNA-methyltransferases) and carbohydrate metabolism (e.g. glycosyltransferases).[4] Some transfer RNA (tRNA) was also identified within the viral particle.

The presence of classical cell genes in the mimivirus's genome dramatically changed scientists' view about viruses, reopening an old debate: are viruses living organisms?



Computer artwork of a mimivirus, showing the fibrils that enable it to attach to its hosts

RUSSELL KIGHTLEY/SCIENCE PHOTO LIBRARY

Are giant viruses a threat to people?

- It has been shown that mimiviruses can enter and replicate in human macrophages and in certain blood cells, where they interfere with the innate immune system.[8] There is evidence that they infect humans (such as the detection of mimiviruses and anti-mimivirus antibodies in patients), but there is still no consensus.[9,10]
- It remains a mystery whether they caused the pneumonia outbreak that occurred when they were first isolated in Bradford in 1992.

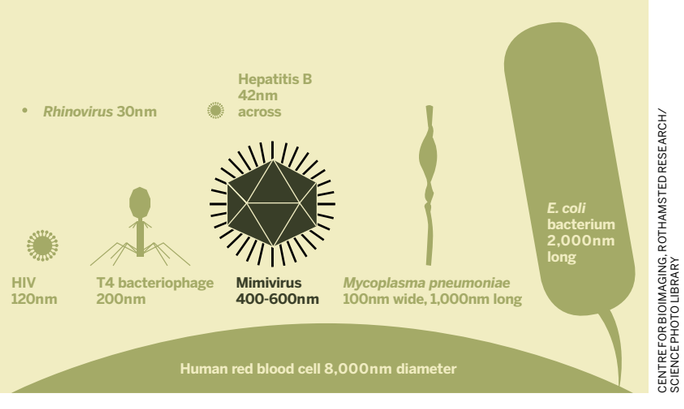
***Pandoravirus* has one of the largest genomes of any virus**

A QUESTION OF SCALE

Mimivirus in perspective

How it compares to other organisms

The mimivirus is 400–600nm across and has 1.2 million base pairs. Compare this to the hepatitis B virus, which is a mere 42nm across, with 3,200 base pairs. One of the largest phage viruses, T4 bacteriophage, at 200nm is around half the size of the mimivirus and has a fraction of the number of base pairs (69,000). The mimivirus is around the same size as the *Staphylococcus aureus* bacterium.



CENTRE FOR BIOIMAGING, ROTHAMSTED RESEARCH/ SCIENCE PHOTO LIBRARY

Before the discovery of the mimiviruses, a consensus considered viruses to be 'just' small, obligate, intracellular microorganisms, completely dependent on cellular machinery, with few reasons to group them along with cellular or living organisms. However, the huge size and presence of classical cell genes within mimivirus's genome put this consensus to the test. For the first time, it was possible to phylogenetically cluster the viruses along with the other domains of life, Eukarya, Bacteria and Archaea in a metaphorical tree of life, leading to the proposal of a fourth domain of life, which includes the viruses from Megavirales.[5]

Moreover, with genes related to translational apparatus, the mimiviruses possess a partial independence from the host. Since they can also be parasitised by other viruses (virophages), it reinforced the idea that viruses could be regarded as living organisms, and so should be represented in the tree of life. However, in science, new proposals are not accepted easily.

Some scientists are against the creation of a fourth domain of life for a number of reasons. Even though the mimiviruses have characteristics that partially exclude them from the classical definition of viruses proposed by André Lwoff, and have a relative independence from the host, they are still inert out of a cell. In addition, the viruses do not have a single ancestral lineage. They are polyphyletic organisms, and some of their genes come from horizontal gene transfer; it is impossible to establish a reliable phylogenetic relationship based on these genes.[6] Others believe what should be considered is the virus while inside the cell host[7] ('the virocell'), which forms a viral factory and is able to replicate and evolve, both essential characteristics of life.

Its genome contains classical 'cell' genes related to protein translation

The debate will rumble on, but the entire classification system could change in the coming years. It is already feasible that all the organisms on the planet will be divided into two groups, ribosome encoding organisms (i.e. cellular organisms, or REOs) and capsid encoding organisms (viral organisms, or CEOs), leaving the idea of domains behind. Some evolutionists still claim that the term 'domain' should not be extinguished, but used exclusively for REO, and the term 'lineage' should be for CEO members. Only time will tell if these ideas succeed.

There is still a lot to discover. Whether viruses will be considered alive or not is uncertain, but no one can deny the role they play in the evolution of life.

Jônatas Abrahão is professor of virology at Federal University of Minas Gerais, Brazil. His research interests focus on relationships among large/giant viruses and their hosts.

Rodrigo Rodrigues is a PhD student, Dept of Microbiology, Federal University of Minas Gerais, Brazil. His research interests focus on the structural, ecological and evolutionary aspects of giant viruses.

References
1) Lwoff, A. The concept of virus. *J Gen Microbiol* **17**(2), 239–253 (1957).
2) La Scola, B. *et al.* A giant virus in amoebae. *Science* **299**(5615), 2033 (2003).
3) Rodrigues, R. A. *et al.* Mimivirus fibrils are important for viral attachment to the microbial world by a diverse glycoside interaction repertoire. *J Virol* **89**(23), 11812–11819 (2015).
4) Raoult, D. *et al.* The 1.2-megabase genome sequence of Mimivirus. *Science* **306**(5700), 1344–1350 (2004).
5) Boyer, M. *et al.* Phylogenetic and phyletic studies of informational genes in genome highlight existence of a 4th domain of life including giant viruses. *PLoS One* **5**(12), e15530 (2010).
6) Moreira, D., López-García, P. Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol* **7**(4), 306–311 (2009).
7) Claverie, J. M. & Ogata, H. Ten good reasons not to exclude giruses from the evolutionary picture. *Nat Rev Microbiol* **7**(8), 615 (2009).
8) Silva, L. C. *et al.* A resourceful giant: APMV is able to interfere with the human type I interferon system. *Microbes and Infect* **16**(3), 187–195 (2014).
9) La Scola, B. *et al.* Mimivirus in pneumonia patients. *Emerg Infect Dis* **11**(3), 449–452 (2005).
10) Saadi, H. *et al.* First isolation of Mimivirus in a patient with pneumonia. *Clin Infect Dis* **54**(7), e127–134 (2013).

JVI
Journals.ASM.org

Mimivirus Fibrils Are Important for Viral Attachment to the Microbial World by a Diverse Glycoside Interaction Repertoire

Rodrigo Araújo Lima Rodrigues,[a] Ludmila Karen dos Santos Silva,[a] Fábio Pio Dornas,[a] Danilo Bretas de Oliveira,[a] Thais Furtado Ferreira Magalhães,[b] Daniel Assis Santos,[b] Adriana Oliveira Costa,[d] Luiz de Macêdo Farias,[c] Paula Prazeres Magalhães,[c] Cláudio Antônio Bonjardim,[a] Erna Geessien Kroon,[a] Bernard La Scola,[f] Juliana Reis Cortines,[e] Jônatas Santos Abrahão[a]

Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais, Brasil (UFMG), Belo Horizonte, Minas Gerais, Brazil[a]; Laboratório de Micologia, Departamento de Microbiologia, Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais, Brasil (UFMG), Belo Horizonte, Minas Gerais, Brazil[b]; Laboratório de Microbiologia Oral e Anaeróbios, Departamento de Microbiologia, Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais, Brasil (UFMG), Belo Horizonte, Minas Gerais, Brazil[c]; Faculdade de Farmácia, Universidade Federal de Minas Gerais, Brasil (UFMG), Belo Horizonte, Minas Gerais, Brazil[d]; Instituto de Microbiologia Paulo de Góes, Departamento de Virologia-Centro de Ciências da Saúde, Cidade Universitária, Ilha do Fundão, Rio de Janeiro, RJ, Brazil[e]; URMITE CNRS UMR 6236-IRD 3R198, Aix Marseille Université, Marseille, France[f]

ABSTRACT

Acanthamoeba polyphaga mimivirus **(APMV) is a giant virus from the *Mimiviridae* family. It has many unusual features, such as a pseudoicosahedral capsid that presents a starfish shape in one of its vertices, through which the ~1.2-Mb double-stranded DNA is released. It also has a dense glycoprotein fibril layer covering the capsid that has not yet been functionally characterized. Here, we verified that although these structures are not essential for viral replication, they are truly necessary for viral adhesion to amoebae, its natural host. In the absence of fibrils, APMV had a significantly lower level of attachment to the *Acanthamoeba castellanii* surface. This adhesion is mediated by glycans, specifically, mannose and *N*-acetylglucosamine (a monomer of chitin and peptidoglycan), both of which are largely distributed in nature as structural components of several organisms. Indeed, APMV was able to attach to different organisms, such as Gram-positive bacteria, fungi, and arthropods, but not to Gram-negative bacteria. This prompted us to predict that (i) arthropods, mainly insects, might act as mimivirus dispersers and (ii) by attaching to other microorganisms, APMV could be ingested by amoebae, leading to the successful production of viral progeny. To date, this mechanism has never been described in the virosphere.**

IMPORTANCE

APMV is a giant virus that is both genetically and structurally complex. Its size is similar to that of small bacteria, and it replicates inside amoebae. The viral capsid is covered by a dense glycoprotein fibril layer, but its function has remained unknown, until now. We found that the fibrils are not essential for mimivirus replication but that they are truly necessary for viral adhesion to the cell surface. This interaction is mediated by glycans, mainly *N*-acetylglucosamine. We also verified that APMV is able to attach to bacteria, fungi, and arthropods. This indicates that insects might act as mimivirus dispersers and that adhesion to other microorganisms could facilitate viral ingestion by amoebae, a mechanism never before described in the virosphere.

Just over a decade ago, some dogmas of virology were put to the test with the discovery of the giant *Acanthamoeba polyphaga* mimivirus (APMV) (1). This was the first member of the family *Mimiviridae*, which comprises one of the most complex viral groups (2). Further studies confirmed its genetic and structural complexity, revealing genes related to the translational apparatus of cell machinery, glycosylation pathways, and a pseudoicosahedral capsid of approximately 750 nm surrounded by a dense layer of fibrils, characteristics never observed before in virology (3–5). Those fibrils have a proteic composition and are covered by a peptidoglycan matrix. Until now, they had no established functions. They surround almost the entire viral particle area, except for the stargate region, the structure responsible for releasing the viral genome inside the host cell (6).

These viruses are hosted by free-living amoebae of the genus *Acanthamoeba* (1). These unicellular organisms go through two distinct life stages: a resistant form (the cyst) and a metabolically active form (the trophozoite), which feeds through phagocytosis (7). By this process, amoebae can incorporate many microorganisms, such as bacteria, fungi, protozoa, and viruses, including giant viruses such as pandoravirus, pithovirus, marseillevirus, and mimivirus (8). It is known that for phagocytosis to occur, the microbial structure and cellular adhesion factors, which are often surface glycans, must interact (9). On the surface of *Acanthamoeba*, glucose, mannose, *N*-acetylgalactosamine (GalNAc), and *N*-acetylglucosamine (GlcNAc) have already been identified to be cellular adhesion factors (10). These carbohydrates serve as the building blocks for several natural polymers, such as peptidoglycan and chitin, both of which are largely distributed in the microbial world as the structural components of bacteria, fungi, and arthropods, among other organisms. Although there are some

Received 5 August 2015 **Accepted** 3 September 2015
Accepted manuscript posted online 16 September 2015
Citation Rodrigues RAL, dos Santos Silva LK, Dornas FP, de Oliveira DB, Magalhães TFF, Santos DA, Costa AO, de Macêdo Farias L, Magalhães PP, Bonjardim CA, Kroon EG, La Scola B, Cortines JR, Abrahão JS. 2015. Mimivirus fibrils are important for viral attachment to the microbial world by a diverse glycoside interaction repertoire. J Virol 89:11812–11819. doi:10.1128/JVI.01976-15.
Editor: G. McFadden
Address correspondence to Jônatas Santos Abrahão, jonatas.abrahao@gmail.com.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

The Large Marseillevirus Explores Different Entry Pathways by Forming Giant Infectious Vesicles

Thalita Souza Arantes,[a] Rodrigo Araújo Lima Rodrigues,[a] Ludmila Karen dos Santos Silva,[a] Graziele Pereira Oliveira,[a] Helton Luís de Souza,[a] Jacques Y. B. Khalil,[c] Danilo Bretas de Oliveira,[a] Alice Abreu Torres,[a] Luis Lamberti da Silva,[b] Philippe Colson,[c] Erna Geessien Kroon,[a] Flávio Guimarães da Fonseca,[a] Cláudio Antônio Bonjardim,[a] Bernard La Scola,[c] Jônatas Santos Abrahão[a]

Department of Microbiology, Institute of Biological Sciences, Unviversidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil[a]; Department of Cell and Molecular Biology, Ribeirão Preto Medical School, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil[b]; URMITE CNRS UMR 6236–IRD 3R198, Aix Marseille Universite, Marseille, France[c]

ABSTRACT

Triggering the amoebal phagocytosis process is a *sine qua non* condition for most giant viruses to initiate their replication cycle and consequently to promote their progeny formation. It is well known that the amoebal phagocytosis process requires the recognition of particles of >500 nm, and most amoebal giant viruses meet this requirement, such as mimivirus, pandoravirus, pithovirus, and mollivirus. However, in the context of the discovery of amoebal giant viruses in the last decade, *Marseillevirus marseillevirus* (MsV) has drawn our attention, because despite its ability to successfully replicate in *Acanthamoeba*, remarkably it does not fulfill the >500-nm condition, since it presents an ~250-nm icosahedrally shaped capsid. We deeply investigated the MsV cycle by using a set of methods, including virological, molecular, and microscopic (immunofluorescence, scanning electron microscopy, and transmission electron microscopy) assays. Our results revealed that MsV is able to form giant vesicles containing dozens to thousands of viral particles wrapped by membranes derived from amoebal endoplasmic reticulum. Remarkably, our results strongly suggested that these giant vesicles are able to stimulate amoebal phagocytosis and to trigger the MsV replication cycle by an acidification-independent process. Also, we observed that MsV entry may occur by the phagocytosis of grouped particles (without surrounding membranes) and by an endosome-stimulated pathway triggered by single particles. Taken together, not only do our data deeply describe the main features of MsV replication cycle, but this is the first time, to our knowledge, that the formation of giant infective vesicles related to a DNA virus has been described.

IMPORTANCE

Triggering the amoebal phagocytosis process is a *sine qua non* condition required by most giant viruses to initiate their replication cycle. This process requires the recognition of particles of >500 nm, and many giant viruses meet this requirement. However, MsV is unusual, as despite having particles of ~250 nm it is able to replicate in *Acanthamoeba*. Our results revealed that MsV is able to form giant vesicles, containing dozens to thousands of viral particles, wrapped in membranes derived from amoebal endoplasmic reticulum. Remarkably, our results strongly suggest that these giant vesicles are able to stimulate phagocytosis using an acidification-independent process. Our work not only describes the main features of the MsV replication cycle but also describes, for the first time to our knowledge, the formation of huge infective vesicles in a large DNA viruses.

The discovery of the amoebal giant virus *Acanthamoeba polyphaga mimivirus* (APMV) in 2003 (1) raised new and exciting questions regarding the virosphere and boosted the hunt for new giant viruses. Owing to these efforts, an increasing number of remarkable giant viruses have been described (2–6).

Recent data suggest that giant viruses initiate their replication cycles after being phagocytosed by amoebas or other phagocytic cells (1, 7). This conclusion is well supported by classical studies which show that the phagocytosis process is triggered in *Acanthamoeba* by particles of >500 nm (8). Therefore, most of the giant viruses, such as mimivirus, pandoravirus, pithovirus, and mollivirus, meet this requirement (1–3, 5). *Marseillevirus marseillevirus* (MsV) particles, on the other hand, are formed by 250-nm icosahedral capsids that do not reach the 500-nm size threshold (6). Nonetheless, MsV still is able to successfully replicate in *Acanthamoeba*, raising the question of how the virus enters its host cell. After entry, a large and diffuse viral factory is assembled (2 to 4 h), wherein genome replication and virion morphogenesis occurs (6 to 8 h), and the virions are released from the cell within 24 h. Just like the mimivirus, MsV has fibers with globular ends on the surface, although they are shorter in length (~12 nm compared to ~125 nm of mimiviruses), and both have an internal membrane surrounding the nucleocapsid (6). However, the origins of the MsV inner membrane remain unknown (6). The first marseillevirus was isolated from a water sample collected from a cooling

Received 28 January 2016 **Accepted** 10 March 2016

Accepted manuscript posted online 16 March 2016

Citation 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. doi:10.1128/JVI.00177-16.

Editor: G. McFadden

Address correspondence to Bernard La Scola, bernard.la-scola@univ-amu.fr, or Jônatas Santos Abrahão, jonatas.abrahao@gmail.com.

viruses

MDPI

Review

Promoter Motifs in NCLDVs: An Evolutionary Perspective

Graziele Pereira Oliveira [1,†], Ana Cláudia dos Santos Pereira Andrade [1,†], Rodrigo Araújo Lima Rodrigues [1], Thalita Souza Arantes [1], Paulo Victor Miranda Boratto [1], Ludmila Karen dos Santos Silva [1], Fábio Pio Dornas [1], Giliane de Souza Trindade [1], Betânia Paiva Drumond [1], Bernard La Scola [2], Erna Geessien Kroon [1] and Jônatas Santos Abrahão [1,*]

1 Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Minas Gerais, Brazil; graziufmg@yahoo.com.br (G.P.O.); ana.andrade2008@hotmail.com (A.C.d.S.P.A.); rodriguesral07@gmail.com (R.A.L.R.); tsarantes@gmail.com (T.S.A.); pvboratto@gmail.com (P.V.M.B.); ludmilakaren@gmail.com (L.K.d.S.S.); fabiopiod154@gmail.com (F.P.D.); gitrindade@yahoo.com.br (G.d.S.T.); betaniadrumond@gmail.com (B.P.D.); ernagkroon@gmail.com (E.G.K.)

2 Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE) UM63 CNRS 7278 IRD 198 INSERM U1095, Aix-Marseille Université., 27 Boulevard Jean Moulin, Faculté de Médecine, 13385 Marseille Cedex 05, France; bernard.la-scola@univ-amu.fr

* Correspondence: jonatas.abrahao@gmail.com; Tel.: +55-031-3409-3002

† These authors contributed equally to this work.

Academic Editor: Eric Freed

Received: 14 November 2016; Accepted: 5 January 2017; Published: 20 January 2017

Abstract: For many years, gene expression in the three cellular domains has been studied in an attempt to discover sequences associated with the regulation of the transcription process. Some specific transcriptional features were described in viruses, although few studies have been devoted to understanding the evolutionary aspects related to the spread of promoter motifs through related viral families. The discovery of giant viruses and the proposition of the new viral order Megavirales that comprise a monophyletic group, named nucleo-cytoplasmic large DNA viruses (NCLDV), raised new questions in the field. Some putative promoter sequences have already been described for some NCLDV members, bringing new insights into the evolutionary history of these complex microorganisms. In this review, we summarize the main aspects of the transcription regulation process in the three domains of life, followed by a systematic description of what is currently known about promoter regions in several NCLDVs. We also discuss how the analysis of the promoter sequences could bring new ideas about the giant viruses' evolution. Finally, considering a possible common ancestor for the NCLDV group, we discussed possible promoters' evolutionary scenarios and propose the term "MEGA-box" to designate an ancestor promoter motif ('TATATAAAATTGA') that could be evolved gradually by nucleotides' gain and loss and point mutations.

Keywords: megavirales; NCLDV; giant viruses; promoter; transcription; evolution; MEGA-box

1. Introduction

For decades, viruses have been strictly considered intracellular parasites, filterable in membranes of 0.22 nm, composed by genomes of DNA or RNA encoding only a few proteins, being entirely dependent on the metabolic machinery of the host cell [1]. However, viruses show a large diversity of genome size and organization, capsid architecture, mechanisms of replication, and interactions with host cells. The extreme diversity of viruses suggests that they must have had multiple evolutionary origins, thus being polyphyletic [2]. In 2001, a supposedly monophyletic

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENOME REPLICATION AND REGULATION OF VIRAL GENE EXPRESSION

The Investigation of Promoter Sequences of Marseilleviruses Highlights a Remarkable Abundance of the AAATATTT Motif in Intergenic Regions

Graziele Pereira Oliveira,[a] Maurício Teixeira Lima,[a] Thalita Souza Arantes,[a] Felipe Lopes Assis,[a] Rodrigo Araújo Lima Rodrigues,[a] Flávio Guimarães da Fonseca,[a] Cláudio Antônio Bonjardim,[a] Erna Geessien Kroon,[a] Philippe Colson,[b] Bernard La Scola,[b] Jônatas Santos Abrahão[a]

Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil[a]; URMITE, Aix Marseille Université, UM63, CNRS 7278, IRD 198, INSERM 1095, IHU–Méditerranée Infection, AP-HM, Marseille, France[b]

ABSTRACT Viruses display a wide range of genomic profiles and, consequently, a variety of gene expression strategies. Specific sequences associated with transcriptional processes have been described in viruses, and putative promoter motifs have been elucidated for some nucleocytoplasmic large DNA viruses (NCLDV). Among NCLDV, the *Marseilleviridae* is a well-recognized family because of its genomic mosaicism. The marseilleviruses have an ability to incorporate foreign genes, especially from sympatric organisms inhabiting *Acanthamoeba*, its main known host. Here, we identified for the first time an eight-nucleotide A/T-rich promoter sequence (AAATATTT) associated with 55% of marseillevirus genes that is conserved in all marseilleviruses lineages, a higher level of conservation than that of any giant virus described to date. We instigated our prediction about the promoter motif by biological assays and by evaluating how single mutations in this octamer can impact gene expression. The investigation of sequences that regulate the expression of genes relative to lateral transfer revealed that the promoter motifs do not appear to be incorporated by marseilleviruses from donor organisms. Indeed, analyses of the intergenic regions that regulate lateral gene transfer-related genes have revealed an independent origin of the marseillevirus intergenic regions that does not match gene-donor organisms. About 50% of AAATATTT motifs spread throughout intergenic regions of the marseilleviruses are present as multiple copies. We believe that such multiple motifs are associated with increased expression of a given gene or are related to incorporation of foreign genes into the mosaic genome of marseilleviruses.

IMPORTANCE The marseilleviruses draw attention because of the peculiar features of their genomes; however, little is known about their gene expression patterns or the factors that regulate those expression patterns. The limited published research on the expression patterns of the marseilleviruses and their unique genomes has led us to study the promoter motif sequences in the intergenic regions of the marseilleviruses. This work is the first to analyze promoter sequences in the genomes of the marseilleviruses. We also suggest a strong capacity to acquire foreign genes and to express those genes mediated by multiple copies of the promoter motifs available in intergenic regions. These findings contribute to an understanding of genomic expansion and plasticity observed in these giant viruses.

KEYWORDS lateral gene transfer, *Marseilleviridae*, gene expression, promoter

Received 29 June 2017 **Accepted** 4 August 2017

Accepted manuscript posted online 9 August 2017

Citation Oliveira GP, Lima MT, Arantes TS, Assis FL, Rodrigues RAL, da Fonseca FG, Bonjardim CA, Kroon EG, Colson P, Scola BL, Abrahão JS. 2017. The investigation of promoter sequences of marseilleviruses highlights a remarkable abundance of the AAATATTT motif in intergenic regions. J Virol 91:e01088-17. https://doi.org/10.1128/JVI.01088-17.

Editor Rozanne M. Sandri-Goldin, University of California, Irvine

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

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

VIRUS-CELL INTERACTIONS

Filling Knowledge Gaps for Mimivirus Entry, Uncoating, and Morphogenesis

Ana Cláudia dos Santos Pereira Andrade,[a] Rodrigo Araújo Lima Rodrigues,[a] Graziele Pereira Oliveira,[a] Kétyllen Reis Andrade,[b] Cláudio Antônio Bonjardim,[a] Bernard La Scola,[c] Erna Geessien Kroon,[a] Jônatas Santos Abrahão[a]

Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil[a]; Laboratório de Imunopatologia de Doenças Virais, Centro de Pesquisas René Rachou—Fiocruz, Minas Gerais, Brazil[b]; URMITE, CNRS UMR 6236, IRD 3R198, Aix Marseille Universite, Marseille, France[c]

ABSTRACT Since the discovery of mimivirus, its unusual structural and genomic features have raised great interest in the study of its biology; however, many aspects concerning its replication cycle remain uncertain. In this study, extensive analyses of electron microscope images, as well as biological assay results, shed light on unclear points concerning the mimivirus replication cycle. We found that treatment with cytochalasin, a phagocytosis inhibitor, negatively impacted the incorporation of mimivirus particles by *Acanthamoeba castellanii*, causing a negative effect on viral growth in amoeba monolayers. Treatment of amoebas with bafilomicin significantly impacted mimivirus uncoating and replication. In conjunction with microscopic analyses, these data suggest that mimiviruses indeed depend on phagocytosis for entry into amoebas, and particle uncoating (and stargate opening) appears to be dependent on phagosome acidification. In-depth analyses of particle morphogenesis suggest that the mimivirus capsids are assembled from growing lamellar structures. Despite proposals from previous studies that genome acquisition occurs before the acquisition of fibrils, our results clearly demonstrate that the genome and fibrils can be acquired simultaneously. Our data suggest the existence of a specific area surrounding the core of the viral factory where particles acquire the surface fibrils. Furthermore, we reinforce the concept that defective particles can be formed even in the absence of virophages. Our work provides new information about unexplored steps in the life cycle of mimivirus.

IMPORTANCE Investigating the viral life cycle is essential to a better understanding of virus biology. The combination of biological assays and microscopic images allows a clear view of the biological features of viruses. Since the discovery of mimivirus, many studies have been conducted to characterize its replication cycle, but many knowledge gaps remain to be filled. In this study, we conducted a new examination of the replication cycle of mimivirus and provide new evidence concerning some stages of the cycle which were previously unclear, mainly entry, uncoating, and morphogenesis. Furthermore, we demonstrate that atypical virion morphologies can occur even in the absence of virophages. Our results, along with previous data, allow us to present an ultimate model for the mimivirus replication cycle.

KEYWORDS mimivirus, electron microscopy, replication cycle, phagocytosis, fibril acquisition area

Received 2 August 2017 **Accepted** 23 August 2017

Accepted manuscript posted online 6 September 2017

Citation Andrade ACDSP, 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.

Editor Rozanne M. Sandri-Goldin, University of California, Irvine

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

A.C.D.S.P.A., R.A.L.R., and G.P.O. contributed equally to this work.

The giant Acanthamoeba polyphaga mimivirus (APMV), which is associated with amoebas of the *Acanthamoeba* genus, was isolated in 2003 and astonished the scientific community with unusual structural and genomic features within the virosphere (1, 2). In subsequent years, several mimivirus-like viruses were uncovered in different parts of the world, thus expanding the *Mimiviridae* family, especially the

Anais da Academia Brasileira de Ciências (2017) 89(3): 1555-1564
(Annals of the Brazilian Academy of Sciences)
Printed version ISSN 0001-3765 / Online version ISSN 1678-2690
http://dx.doi.org/10.1590/0001-3765201720160046
www.scielo.br/aabc | www.fb.com/aabcjournal

Antidiarrheal activity of extracts from *Maytenus gonoclada* and inhibition of *Dengue virus* by lupeol

FERNANDO C. SILVA[1], VANESSA G. RODRIGUES[2], LUCIENIR P. DUARTE[2], IVANA S. LULA[2], RUBEN D. SINISTERRA[2], SIDNEY A. VIEIRA-FILHO[3], RODRIGO A.L. RODRIGUES[4], ERNA G. KROON[4], PATRÍCIA L. OLIVEIRA[4], LUIZ M. FARIAS[4], PAULA P. MAGALHÃES[4] and GRÁCIA D.F. SILVA[2]

[1]Universidade do Estado de Minas Gerais, Departamento das Licenciaturas, Avenida Paraná, 3001, 35501-170 Divinópolis, MG, Brazil
[2]Universidade Federal de Minas Gerais, Departamento de Química, Instituto de Ciências Exatas, Avenida Antônio Carlos, 6627, 31270-901 Belo Horizonte, MG, Brazil
[3]Departamento de Farmácia, Escola de Farmácia, Universidade Federal de Ouro Preto, Campus Morro do Cruzeiro, 35400-000 Ouro Preto, MG, Brazil
[4]Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Avenida Antônio Carlos, 6627, 31270-901 Belo Horizonte, MG, Brazil

Manuscript received on February 2, 2016; accepted for publication on June 20, 2016

ABSTRACT

Diarrhea is an infectious disease caused by bacterial, virus, or protozoan, and dengue is caused by virus, included among the neglected diseases in several underdeveloped and developing countries, with an urgent demand for new drugs. Considering the antidiarrheal potential of species of *Maytenus* genus, a phytochemical investigation followed by antibacterial activity test with extracts of branches and heartwood and bark of roots from *Maytenus gonoclada* were conducted. Moreover, due the frequency of isolation of lupeol from *Maytenus* genus the antiviral activity against *Dengue virus* and cytotoxicity of lupeol and its complex with β-cyclodextrins were also tested. The results indicated the bioactivity of ethyl acetate extract from branches and ethanol extract from heartwood of roots of *M. gonoclada* against diarrheagenic bacteria. The lupeol showed potent activity against *Dengue virus* and low cytotoxicity in LLC-MK_2 cells, but its complex with β-cyclodextrin was inactive. Considering the importance of novel and selective antiviral drug candidates the results seem to be promising.

Key words: Antidiarrheal, Celastraceae, Pentacyclic Triterpene, Antiviral Activity.

INTRODUCTION

Diarrheal disease is a worldwide health problem associated with high morbidity and mortality rates mainly in underdeveloped countries. Most cases of infectious acute diarrhea are self-limited meaning that they resolve on their own. However, there are recommendations relative to antimicrobial treatment for some specific situations. Despite of the low percentage of cases that fulfill this requirement the huge prevalence of the disease makes antimicrobial therapy addressing the etiological agents of the process a relevant matter. Additionally raising drug resistance rates has been

Correspondence to: Fernando César Silva
E-mail: fcsquimico@yahoo.com.br

viruses

MDPI

Review

Poxvirus Host Range Genes and Virus–Host Spectrum: A Critical Review

Graziele Pereira Oliveira †, Rodrigo Araújo Lima Rodrigues †, Maurício Teixeira Lima †, Betânia Paiva Drumond and Jônatas Santos Abrahão *

Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil; graziufmg@yahoo.com.br (G.P.O.); rodriguesral07@gmail.com (R.A.L.R.); maurili15@hotmail.com (M.T.L.); betaniadrumond@gmail.com (B.P.D.)

* Correspondence: jonatas.abrahao@gmail.com; Tel.: +55-31-3409-2766

† These authors contributed equally to this work.

Received: 20 September 2017; Accepted: 6 November 2017; Published: 7 November 2017

Abstract: The *Poxviridae* family is comprised of double-stranded DNA viruses belonging to nucleocytoplasmic large DNA viruses (NCLDV). Among the NCLDV, poxviruses exhibit the widest known host range, which is likely observed because this viral family has been more heavily investigated. However, relative to each member of the *Poxviridae* family, the spectrum of the host is variable, where certain viruses can infect a large range of hosts, while others are restricted to only one host species. It has been suggested that the variability in host spectrum among poxviruses is linked with the presence or absence of some host range genes. Would it be possible to extrapolate the restriction of viral replication in a specific cell lineage to an animal, a far more complex organism? In this study, we compare and discuss the relationship between the host range of poxvirus species and the abundance/diversity of host range genes. We analyzed the sequences of 38 previously identified and putative homologs of poxvirus host range genes, and updated these data with deposited sequences of new poxvirus genomes. Overall, the term host range genes might not be the most appropriate for these genes, since no correlation between them and the viruses' host spectrum was observed, and a change in nomenclature should be considered. Finally, we analyzed the evolutionary history of these genes, and reaffirmed the occurrence of horizontal gene transfer (HGT) for certain elements, as previously suggested. Considering the data presented in this study, it is not possible to associate the diversity of host range factors with the amount of hosts of known poxviruses, and this traditional nomenclature creates misunderstandings.

Keywords: *Poxviridae*; network; host range genes; horizontal gene transfer; evolution

1. Introduction

Poxviruses are among the best known and most feared viruses. The *Poxviridae* family is currently divided in two subfamilies, named *Entomopoxvirinae* (insect-infecting viruses) and *Chordopoxvirinae* (vertebrate-infecting viruses), wherein the first is composed of three genera, and the latter contains 10 genera, in addition to two viral species that have yet to be classified into each subfamily [1]. While the entomopoxviruses have been poorly investigated over the years, the chordopoxviruses are among the most studied groups in virology, due to the medical and veterinary relevance of many of their members. Among the chordopoxviruses, the *Variola virus* (VARV_abbreviations are shown in Supplementary Table S1) is one of the most well-known species. VARV is the agent of smallpox, a disease that has plagued humanity for centuries, until it was considered eradicated by the World Health Organization in 1980 after a successful global vaccination and surveillance campaign [2–4]. Other chordopoxviruses, such as vaccinia virus (VACV), cowpox virus (CPXV) and monkeypox virus

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

Analyses of the Kroon Virus Major Capsid Gene and Its Transcript Highlight a Distinct Pattern of Gene Evolution and Splicing among Mimiviruses

Paulo Victor Miranda Boratto,[a] Fábio Pio Dornas,[a] Lorena Christine Ferreira da Silva,[a] Rodrigo Araújo Lima Rodrigues,[a] Graziele Pereira Oliveira,[a] Juliana Reis Cortines,[b] Betânia Paiva Drumond,[a] Jônatas Santos Abrahão[a]

[a]Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

[b]Instituto de Microbiologia Paulo de Góes, Departamento de Virologia–Centro de Ciências da Saúde, Cidade Universitária, Ilha do Fundão, Rio de Janeiro, RJ, Brazil

ABSTRACT The inclusion of *Mimiviridae* members in the putative monophyletic nucleocytoplasmic large DNA virus (NCLDV) group is based on genomic and phylogenomic patterns. This shows that, along with other viral families, they share a set of genes known as core or "hallmark genes," including the gene for the major capsid protein (MCP). Although previous studies have suggested that the maturation of mimivirus MCP transcripts is dependent on splicing, there is little information about the processing of this transcript in other mimivirus isolates. Here we report the characterization of a new mimivirus isolate, called Kroon virus (KV) mimivirus. Analysis of the structure, synteny, and phylogenetic relationships of the MCP genes in many mimivirus isolates revealed a remarkable variation at position and types of intronic and exonic regions, even for mimiviruses belonging to the same lineage. In addition, sequencing of KV and Acanthamoeba polyphaga mimivirus (APMV) MCP transcripts has shown that inside the family, even related giant viruses may present different ways to process the MCP mRNA. These results contribute to the understanding of the genetic organization and evolution of the MCP gene in mimiviruses.

IMPORTANCE Mimivirus isolates have been obtained by prospecting studies since 2003. Based on genomic and phylogenomic studies of conserved genes, these viruses have been clustered together with members of six other viral families. Although the major capsid protein (MCP) gene is an important member of the so-called "hallmark genes," there is little information about the processing and structure of this gene in many mimivirus isolates. In this work, we have analyzed the structure, synteny, and phylogenetic relationships of the MCP genes in many mimivirus isolates; these genes showed remarkable variation at position and types of intronic and exonic regions, even for mimiviruses belonging to the same lineage. These results contribute to the understanding of the genetic organization and evolution of the MCP gene in mimiviruses.

KEYWORDS KV, mimivirus, *Megavirales*, giant virus, isolation, capsid, splicing

The term virus is used to define a group of biological entities united under a set of generic and polythetic features (1). Among these features, the viruses can be linked together primarily by their dependence on the host biosynthetic machinery and also by their intracellular parasitic character. The fact that there is no a single gene uniting all of the viruses in a separated phylogenetic clade means that these organisms are considered a polyphyletic group, in which the different viral species emerged from distinct ancestors (2–4). However, there are some specific subgroups of viruses in which a set of genes, known as "core/hallmark genes," are present in all the members, suggesting a probable common ancestry between them (4). The nucleocytoplasmic

Received 10 October 2017 **Accepted** 30 October 2017

Accepted manuscript posted online 8 November 2017

Citation Boratto PVM, Dornas FP, da Silva LCF, Rodrigues RAL, Oliveira GP, Cortines JR, Drumond BP, Abrahão JS. 2018. Analyses of the Kroon virus major capsid gene and its transcript highlight a distinct pattern of gene evolution and splicing among mimiviruses. J Virol 92:e01782-17. https://doi.org/10.1128/JVI.01782-17.

Editor Julie K. Pfeiffer, University of Texas Southwestern Medical Center

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

Andrade et al. Virology Journal (2018) 15:22
DOI 10.1186/s12985-018-0930-x

Virology Journal

RESEARCH Open Access

Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica

Ana Cláudia dos S. P. Andrade[1], Thalita S. Arantes[1], Rodrigo A. L. Rodrigues[1], Talita B. Machado[1], Fábio P. Dornas[1], Melissa F. Landell[2], Cinthia Furst[3], Luiz G. A. Borges[4,5], Lara A. L. Dutra[6], Gabriel Almeida[6], Giliane de S. Trindade[1], Ivan Bergier[7], Walter Abrahão[8], Iara A. Borges[1], Juliana R. Cortines[9], Danilo B. de Oliveira[10], Erna G. Kroon[1] and Jônatas S. Abrahão[1*]

Abstract

Background: Since the discovery of giant viruses infecting amoebae in 2003, many dogmas of virology have been revised and the search for these viruses has been intensified. Over the last few years, several new groups of these viruses have been discovered in various types of samples and environments.In this work, we describe the isolation of 68 giant viruses of amoeba obtained from environmental samples from Brazil and Antarctica.

Methods: Isolated viruses were identified by hemacolor staining, PCR assays and electron microscopy (scanning and/or transmission).

Results: A total of 64 viruses belonging to the *Mimiviridae* family were isolated (26 from lineage A, 13 from lineage B, 2 from lineage C and 23 from unidentified lineages) from different types of samples, including marine water from Antarctica, thus being the first mimiviruses isolated in this extreme environment to date. Furthermore, a marseillevirus was isolated from sewage samples along with two pandoraviruses and a cedratvirus (the third to be isolated in the world so far).

Conclusions: Considering the different type of samples, we found a higher number of viral groups in sewage samples. Our results reinforce the importance of prospective studies in different environmental samples, therefore improving our comprehension about the circulation anddiversity of these viruses in nature.

Keywords: Giant viruses, Prospection, Brazil, Antarctica, Pandoravirus, Cedratvirus, Marseillevirus, Mimivirus

Background

The discovery of *Acanthamoeba polyphaga mimivirus* (APMV) in 2003, the first isolated giant virus infecting amoebas, interested the scientific community due to its size and genome content, which culminated in the search for and isolation of new giant viruses [1, 2]. The giant amoebal viruses have many phenotypic and genomic features which had never been seen in other viruses before, like large viral particles presenting up to 1.5 μm in length and large double-stranded DNA genomes ranging from 350 kb in *Marseilleviridae* members to 2500 kb for pandoravirus [3, 4]. These genes encode many hypothetical proteins, uncharacterized, or with functions that have never orrarely been observed before in other viruses, such as those related to translation and DNA repair [5–7]. Common characteristics shared by giant and large DNA viruses permitted their incorporation into a supposedly viral monophyletic group, named nucleocytoplasmic large DNA viruses (NCLDV), created in 2001 [8]. When the NCLDV group was proposed, it was composed of families *Poxviridae* (e.g. *Vaccinia virus, Crocodilepox virus), Asfarviridae* (e.g. *African swine fever virus*) *Iridoviridae* (e.g. *Frog virus 3)* and *Phycodnaviridae* (e.g. *Emiliania huxleyi virus 86, Aureococcus anophagefferens virus)* [8].

Subsequently, viruses belonging to the Mimiviridae, Marseilleviridae, Ascoviridae family and also the pandoravirus, faustovirus, pithovirus, mollivirus, kaumoebavirus, cedratvirus and pacmanvirus were also incorporated to NCLDV group [9–17]. Recent prospective studies have

* Correspondence: jonatas.abrahao@gmail.com
[1]Laboratorio de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
Full list of author information is available at the end of the article

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-03168-1 OPEN

Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere

Jônatas Abrahão[1,2], Lorena Silva[1,2], Ludmila Santos Silva[1,2], Jacques Yaacoub Bou Khalil[3], Rodrigo Rodrigues[2], Thalita Arantes[2], Felipe Assis[2], Paulo Boratto[2], Miguel Andrade[4], Erna Geessien Kroon[2], Bergmann Ribeiro[4], Ivan Bergier[5], Herve Seligmann[1], Eric Ghigo[1], Philippe Colson[1], Anthony Levasseur[1], Guido Kroemer[6,7,8,9,10,11,12], Didier Raoult[1] & Bernard La Scola[1]

Here we report the discovery of two Tupanvirus strains, the longest tailed *Mimiviridae* members isolated in amoebae. Their genomes are 1.44–1.51 Mb linear double-strand DNA coding for 1276–1425 predicted proteins. Tupanviruses share the same ancestors with mimivirus lineages and these giant viruses present the largest translational apparatus within the known virosphere, with up to 70 tRNA, 20 aaRS, 11 factors for all translation steps, and factors related to tRNA/mRNA maturation and ribosome protein modification. Moreover, two sequences with significant similarity to intronic regions of 18 S rRNA genes are encoded by the tupanviruses and highly expressed. In this translation-associated gene set, only the ribosome is lacking. At high multiplicity of infections, tupanvirus is also cytotoxic and causes a severe shutdown of ribosomal RNA and a progressive degradation of the nucleus in host and non-host cells. The analysis of tupanviruses constitutes a new step toward understanding the evolution of giant viruses.

[1] MEPHI, APHM, IRD 198, Aix Marseille Univ, IHU-Méditerranee Infection, 19-21 Bd Jean Moulin, 13005 Marseille, France. [2] Laboratório de Vírus, Instituto de Ciências Biológicas, Departamento de Microbiologia, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil. [3] CNRS, 13005 Marseille, France. [4] Laboratório de Microscopia Eletrônica e Virologia, Departamento de Biologia Celular, Instituto de Ciências Biológicas, Universidade de Brasília, Asa Norte, Brasília 70910-900, Brazil. [5] Lab. Biomass Conversion, Embrapa Pantanal, R. 21 de Setembro 1880, 79320-900 Corumbá/MS, Brazil. [6] Cell Biology and Metabolomics Platforms, Gustave Roussy Cancer Campus, Villejuif 94805, France. [7] Equipe 11 labellisée Ligue Nationale contre le Cancer, Centre de Recherche des Cordeliers, Paris 75006, France. [8] Institut National de la Santé et de la Recherche Médicale (INSERM), Paris 75654, France. [9] Université Paris Descartes, Sorbonne Paris Cité, Paris 75015, France. [10] Université Pierre et Marie Curie, Paris 75005, France. [11] Pôle de Biologie, Hôpital Européen Georges Pompidou, AP-HP, Paris 75015, France. [12] Department of Women's and Children's Health, Karolinska University Hospital, Stockholm SE-171 76, Sweden. Jônatas Abrahão, Lorena Silva and Ludmila Santos Silva contributed equally to this work. Correspondence and requests for materials should be addressed to D.R. (email: didier.raoult@gmail.com) or to B.S. (email: bernard.la-scola@univ-amu.fr)

SCIENTIFIC REPORTS

OPEN

Cedratvirus getuliensis replication cycle: an in-depth morphological analysis

Received: 9 January 2018
Accepted: 22 February 2018
Published online: 05 March 2018

Ludmila Karen dos Santos Silva[1], Ana Cláudia dos Santos Pereira Andrade[1], Fábio Pio Dornas[1,2], Rodrigo Araújo Lima Rodrigues[1], Thalita Arantes[1], Erna Geessien Kroon[1], Cláudio Antônio Bonjardim[1] & Jônatas Santos Abrahão[1]

The giant viruses are the largest and most complex viruses in the virosphere. In the last decade, new members have constantly been added to this group. Here, we provide an in-depth descriptive analysis of the replication cycle of Cedratvirus getuliensis, one of the largest viruses known to date. We tracked the virion entry, the early steps of virus factory and particles morphogenesis, and during this phase, we observed a complex and unique sequential organization of immature particle elements, including horseshoe and rectangular compartments, revealed by transverse and longitudinal sections, respectively, until the formation of the final ovoid-shaped striped virion. The genome and virion proteins are incorporated through a longitudinal opening in the immature virion, followed by the incorporation of the second cork and thickening of the capsid well. Moreover, many cell modifications occur during viral infection, including intense membrane trafficking important to viral morphogenesis and release, as evidenced by treatment using brefeldin A. Finally, we observed that Cedratvirus getuliensis particles are released after cellular lysis, although we obtained microscopic evidence that some particles are released by exocytosis. The present study provides new information on the unexplored steps in the life cycle of cedratviruses.

The study of giant viruses has been intensified after the isolation of Acanthamoeba polyphaga mimivirus, a virus of outstanding dimensions, capable of infecting amoebas of the genus *Acanthamoeba*[1]. Since then, the intense prospection and improvement of isolation techniques has made possible the discovery of new viruses[2,3]. The presence of these viruses has been observed in rather diverse environments, such as water, soil, sewage, and clinical samples, as well as in extreme environments, including permafrost and soda lakes, for example[4–6]. These discoveries have revealed a wide diversity and variety of species not previously observed in the virosphere, challenging the concepts and paradigms concerning the canonical definition of viruses[7]. Currently, the International Committee of Taxonomy of Viruses (ICTV) officially recognizes two families of giant virus of amoebas: *Mimiviridae* and *Marseilleviridae*. In addition to these families, other giant viruses (not assigned yet) have been isolated, such as Faustovirus and Kaumoebavirus, the first giant viruses described to replicate in *Vermoamoeba vermiformes*[8,9]. The tupanviruses, recently isolated from Brazilian environments, present a complex virion structure, with a mimivirus-like capsid attached to a long tail, and these viruses replicate in a broad range of protists (unpublished data). Other isolated viruses, such as Pandoravirus, Pithovirus, Mollivirus and Cedratvirus, also have atypical virion morphologies, exhibiting amphora-shaped, spherical or ovoid structures[4,6,10,11].

Among these viruses, the cedratvirus has an ovoid viral particle, morphologically similar to that of pithovirus but presenting two corks, one at each apex[4,10]. The first Cedratvirus, A11, was isolated from environmental samples from Algeria[10]. Then, a second isolate, Cedratvirus lausannensis, was recovered from a water treatment plant in Morsang-sur-Seine, France[12]. Through an extensive prospective study, we isolated the first cedratvirus from Brazil, named Cedratvirus getuliensis. Although studies on the prospection of giant viruses have advanced over the years, enabling the isolation of new viruses, information regarding their biology remains scarce. In the present study, we present an in-depth investigation of the replication cycle of Cedratvirus getuliensis (C. getuliensis). Through transmission electron microscopy and biological assays using different pharmacological

[1]Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. [2]Centro de Microscopia da Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. Correspondence and requests for materials should be addressed to J.S.A. (email: jonatas.abrahao@gmail.com)

Journal of Virology – JVI01942-18

Title: New isolates of pandoraviruses: contribution to the study of replication cycle steps

Running title: Replication cycle of pandoraviruses

Authors: Ana Cláudia dos Santos Pereira Andrade,[1†] Paulo Victor de Miranda Boratto,[1†] Rodrigo Araújo Lima Rodrigues,[1] Talita Machado Bastos,[1] Bruna Luiza Azevedo,[1] Fábio Pio Dornas,[2] Danilo Bretas Oliveira,[2] Betânia Paiva Drumond,[1] Erna Geessien Kroon,[1] Jônatas Santos Abrahão[1*]

Affiliation: [1] Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.
[2] Universidade Federal do dos Vales do Jequitinhonha e Mucuri, Diamantina, Brazil.

Keywords: Pandoravirus, giant virus, virus diversity, replication cycle, viral morphogenesis, viral release

[†]Ana Cláudia dos Santos Pereira Andrade and Paulo Victor de Miranda Boratto contributed equally to this work.

***Corresponding author:** Prof. Jônatas Santos Abrahão, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. Code: 31270-901. jonatas.abrahao@gmail.com.

Abstract

Giant viruses are complex members of the virosphere, exhibiting outstanding structural and genomic features. Among these viruses, the pandoraviruses are one of the most intriguing members, exhibiting giant particles and genomes presenting up to 2.5 Mb, with many genes having no

Current Opinion in Virology - COVIRO_2018_115

Title: The multiple origins of proteins present in tupanvirus particles

Authors: Paulo Victor Miranda Boratto[a]; Ana Cláudia dos Santos Pereira Andrade[a]; Rodrigo Araújo Lima Rodrigues[a]; Bernard La Scola[b]; Jônatas Santos Abrahão[a]

[a]Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil- postal code 31270-901; [b]URMITE, Aix Marseille Université, UM63, CNRS 7278, IRD 198, INSERM 1095, IHU - Méditerranée Infection, AP-HM, 19-21 boulevard Jean Moulin, 13005 Marseille, France

*Corresponding author: Jônatas Santos Abrahão, Departamento de Microbiologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, postal code 31270-901, +55 31 3409-3002, jonatas.abrahao@gmail.com.

Abstract

In the last decades, the isolation of amoebae-infecting giant viruses has challenged established principles related to the definition of virus, their evolution, and their particle structures represented by a variety of shapes and sizes. Tupanviruses are one of the most recently described amoebae-infecting viruses and exhibit a peculiar morphology with a cylindrical tail attached to the capsid. Proteomic analysis of purified viral particles revealed that virions are composed of over one hundred proteins with different functions. The putative origin of these proteins had not yet been investigated. Here we provide evidences for multiple origins of the proteins present in tupanvirus particles, wherein 20% originate from members of the archaea, bacteria and eukarya.

Introduction

Viral particles have a variety of shapes, symmetries and sizes. The large majority of known viruses have extremely small sizes, with dimensions up to 200 nm in length and relatively simple structures, composed by one or few proteins [1]. This characteristic reflects the genomes of these viruses, which have a reduced number of genes that encode only a few proteins. One group that stands out in this scenario is the giant