UNIVERSIDADE FEDERAL DE MINAS GERAIS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

DEPARTAMENTO DE MICROBIOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM MICROBIOLOGIA

DISSERTAÇÃO DE MESTRADO

ISOLAMENTO E CARACTERIZAÇÃO GENÉTICA DE AMOSTRAS DE MIMIVÍRUS OBTIDAS A PARTIR DE ÁGUA DE LAGOAS URBANAS EM MINAS GERAIS, BRASIL

PAULO VICTOR DE MIRANDA BORATTO

Belo Horizonte

2018

PAULO VICTOR DE MIRANDA BORATTO

ISOLAMENTO E CARACTERIZAÇÃO GENÉTICA DE AMOSTRAS DE MIMIVÍRUS OBTIDAS A PARTIR DE ÁGUA DE LAGOAS URBANAS EM MINAS GERAIS, BRASIL.

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do **Grau de Mestre em Microbiologia.**

Orientador: Dr. Jônatas Santos Abrahão

Belo Horizonte

2018

Sumário

RESUMO

Dissertação de Mestrado

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

ISOLAMENTO E CARACTERIZAÇÃO GENÉTICA DE AMOSTRAS DE MIMIVÍRUS OBTIDAS A PARTIR DE ÁGUA DE LAGOAS URBANAS EM MINAS GERAIS, BRASIL.

PAULO VICTOR DE MIRANDA BORATTO

ORIENTADOR: Jônatas Santos Abrahão

Belo Horizonte, fevereiro de 2017

No ano de 2003 iniciou-se uma das fases mais importantes da virologia moderna, com o descobrimento do primeiro exemplar de vírus gigante pertencente à família *Mimiviridae*, sendo este denominado de *Acanthamoeba polyphaga mimivirus* (APMV). O isolamento do mimivírus, e sua posterior caracterização, impressionou a comunidade cientifica pelas dimensões da partícula e por aquilo que o genoma destes virus é capaz de codificar. Diante desse cenário, muitos genes/proteínas ainda são um campo aberto para estudos envolvendo aspectos evolutivos, funcionais e de caracterização de novos isolados de vírus gigantes. Nesse trabalho, o nosso objetivo foi isolar e caracterizar geneticamente amostras de mimivírus de duas lagoas urbanas localizadas no estado de Minas Gerais, Brasil. Para isso, 80 amostras de água coletadas da Lagoa da Pampulha (Belo Horizonte) e 85 da Lagoa Central (Lagoa Santa) foram processadas e inoculadas em amebas da espécie *Acanthamoeba castellanii*. Dois isolados foram obtidos dessas amostras, sendo denominados de Niemeyer vírus (NYMV) e Kroon vírus (KV), respectivamente. Após esse processo focamos inicialmente em fazer uma análise da presença, distribuição e perfil de relações filogenéticas de aminoacil-tRNA-sintetases apresentadas por diferentes amostras de mimivírus. Essas enzimas denotam importância por sua recente descrição em organismos virais, estando relacionadas com o processo de tradução de proteínas. Nossos dados demonstram que NYMV apresenta duplicações em três (metionil, tirosil e arginil tRNA-sintetases) das suas quatro cópias de aaRs, tendo metionil e tirosil aaRs uma maior expressão de RNA mensageiro (RNAm) quando comparadas com o protótipo APMV. As análises

filogenéticas mostraram que em NYMV, cada uma das cópias duplicadas de aaRs apresenta uma origem distinta, sendo associada a diferentes grupos de mimivirus. Em uma outra parte do trabalho, buscamos obter uma melhor compreensão das diferenças estruturais, de organização e de relações filogenéticas relacionadas a região da proteína principal do capsídeo de nossos isolados, quando comparadas com outras diferentes amostras da família *Mimiviridae*. Os resultados demonstraram que o gene do capsídeo de mimivirus apresenta variações marcantes em termos de posição e tipos de regiões intrônicas e exônicas, até mesmo para vírus pertencentes a uma mesma linhagem. Além disso, o sequenciamento do RNA mensageiro do gene do capsídeo de KV e APMV demonstrou que, mesmo pertencendo à mesma família e linhagem, vírus gigantes relacionados podem apresentar diferentes maneiras de processar os seus transcritos para um mesmo gene. Os resultados apresentados nesse trabalho auxiliam no levantamento de novas questões acerca da origem e de pressões seletivas que envolvem o ganho e a perda de aaRs entre os vírus gigantes, além de contribuir para o entendimento da organização genética e de evolução do gene do capsídeo entre os mimivírus.

Palavras chave: Mimivirus, Kroon virus, Niemeyer virus, aminoacil-tRNA-sintetase, capsídeo

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

µg – micrograma

µm – micrômetros

µl - microlitro

µM – micromolar

aaRs – aminoacil-tRNA-sintetase

APMV – *Acanthamoeba polyphaga mimivirus*

APMOUV – *Acanthamoeba polyphaga moumouvirus*

ATCC - *American Type Culture Collection*

g – grama

ICTV – *International Committee on Taxonomy of Viruses* (Comitê Internacional de Taxonomia de Vírus)

KV – Kroon vírus

M – molar

MHCV – *Megavirus chilensis*

MOI – multiplicidade de infecção

Mb – mega bases

MCP – *major capsid protein* (proteína principal do capsídeo)

MET – microscopia eletrônica de transmissão

mL – mililitros

mM – milimolar

MPAAs – micro-organismos patogênicos associados a amebas

NCLDVs – *nucleocytoplasmic large DNA viruses* (vírus grandes nucleocitoplasmáticos de DNA)

nm – nanômetros

NYMV – Niemeyer vírus

PAS – *Page's amoeba saline*

PCR – reação em cadeia da polimerase

PBS – *phosphate buffer saline* (solução salina fosfato tamponada)

pH – potencial hidrogeniônico

RNAm – RNA mensageiro

SMBV – Samba virus

TGH – transferência gênica horizontal

INTRODUÇÃO

1.1 Vírus grandes nucleocitoplasmáticos de DNA (NCLDVs)

O termo vírus é utilizado como uma maneira de definir um grupo de entidades biológicas unidas sob um conjunto de características genéricas e politéticas (Lwoff, 1957). Dentre essas características, os vírus são classificados de maneira conjunta, baseando-se principalmente em sua dependência da maquinaria biossintética do hospedeiro e também pelo seu caráter parasita intracelular. O fato de não haver a existência de genes capazes de unir todas as espécies virais em um único clado filogenético acaba por determinar em um caráter polifilético desses organismos, no qual diferentes espécies virais teriam surgido de ancestrais distintos (Koonin *et al.*, 2006; Moreira e Lopez-Garcia, 2009; Koonin e Yutin, 2010). No entanto, alguns subgrupos de vírus específicos apresentam um conjunto de genes, chamados de "*core"* ou "*hallmark genes"*, que são compartilhados entre todos os seus membros, sugerindo uma possível ancestralidade comum entre eles (Koonin *et al.*, 2006). Os vírus grandes nucleocitoplasmáticos de DNA (NCLDVs) compreendem um desses subgrupos e incluem até o momento sete diferentes famílias virais: *Ascoviridae, Asfarviridae, Iridoviridae, Marseilleviridae, Phycodnaviridae*, *Poxviridae* e *Mimiviridae* (Iyer *et al.*, 2001; Iyer *et al.*, 2006). Além dessas famílias, os NCLDVs são também formados atualmente por mais outros grupos de vírus, ainda não alocados a nenhuma família taxonômica: os Pandoravirus*,* Pithovirus*,* Faustovirus, Mollivirus, Kaumoebavirus e Orpheovirus (Koonin e Yutin, 2010).

O compartilhamento de "*core genes*" sugere que os membros do NCLDV detém um caráter monofilético, mesmo sabendo-se que esses vírus apresentam a mais variada gama de hospedeiros eucariotos, incluindo de animais complexos até os protistas unicelulares mais simples (Iyer *et al.*, 2001; Iyer *et al.*, 2006). Devido a esse caráter hipoteticamente monofilético, além de outras características como semelhanças em seu processo de multiplicação, no arranjo estrutural de proteínas do capsídeo e no tamanho total da partícula viral, foi proposto por alguns pesquisadores a criação de uma nova ordem viral que englobasse todos os membros desse grupo, sendo essa chamada de ordem

Megavirales (Colson *et al.*, 2012; Colson *et al.*, 2013). No entanto, essa ainda não foi reconhecida pelo Comitê Internacional de Taxonomia de Vírus (ICTV).

Um fator marcante envolvendo os membros do grupo NCLDV é o intenso debate sobre a inclusão desses vírus na árvore da vida por meio da formação de um quarto Domínio, ao lado de *Eukarya, Archaea* e *Bacteria*. Segundo estudos de Boyer e colaboradores, a utilização de um sistema de classificação da vida baseado na presença de ribossomos não é a maneira mais correta de definir se um organismo é ou não vivo, visto que esse processo exclui os vírus deste tipo de classificação. Dessa forma, pela utilização de genes envolvidos com o processamento do DNA, compartilhados tanto entre os NCLDVs e membros de todos os outros Domínios da vida, foi observado a formação de quatro clados distintos, separando os NCLDVs do restante dos membros dos outros três Domínios (Boyer *et al.*, 2010). Pesquisadores que refutam o quarto Domínio argumentam que os genes que suportam essa ideia teriam muito provavelmente sido adquiridos de organismos celulares por meio de transferência gênica horizontal, e que após um processo de evolução gênica acelerada (típico de organismos virais), esses genes teriam se tornado tão diferentes a ponto de perder o sinal filogenético capaz de prover essa informação (Yutin *et al.*, 2014). Uma outra ideia é a de que um vírus ancestral dos NCLDV apresentava como hospedeiro um organismo celular pertencente a um suposto quarto Domínio da vida. Esse vírus teria adquirido desse hospedeiro uma grande quantidade de genes por meio de transferência gênica horizontal e, após a extinção do organismo celular, o ancestral dos vírus gigantes acabou por se tornar um “fóssil vivo” de seu hospedeiro original (Yutin *et al.*, 2014). Muito mais estudos no entanto necessitam ser desenvolvidos nessa área, e o depósito de genomas de novos isolados é um ponto chave para que essas ideias tomem um rumo além do especulativo.

1.2 Os mimivírus

1.2.1 Descoberta e isolamento do primeiro mimivírus

Os mimivírus estão entre os membros do NCLDV recentemente mais estudados pela comunidade científica [Figura 1]. Sua descoberta e caracterização foi um processo importante para reacender novos debates sobre evolução e inclusão dos vírus na árvore da vida (Boyer *et al.*, 2010; Koonin e Yutin, 2010; Colson *et al.*, 2012)

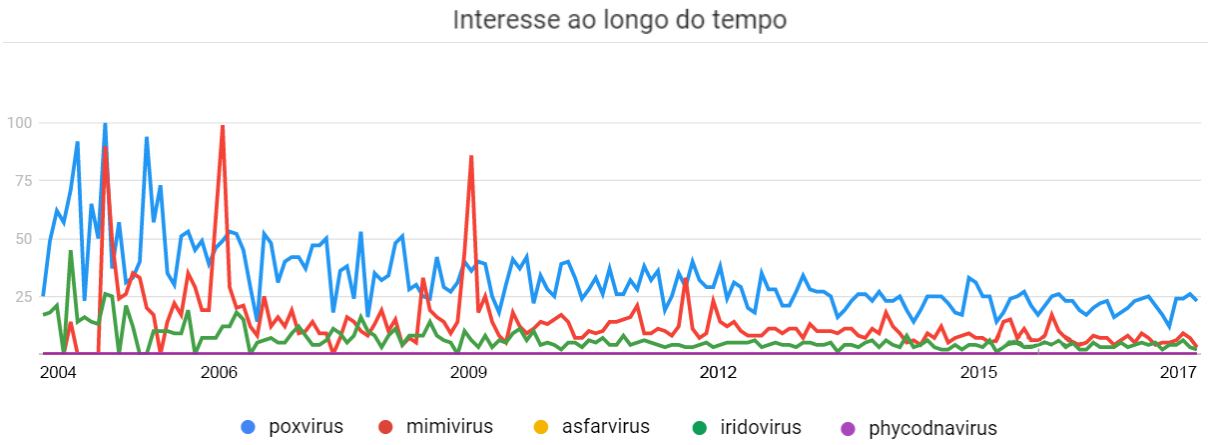


Figura 1: Linha do tempo demonstrando o interesse relativo em relação a alguns membros do NCLDV como termos de pesquisa no Google Trends (www.google.com/trends). Os valores são definidos de acordo com o termo de pesquisa mais procurado, sendo este considerado 100% do valor total. Termos utilizados: “poxvirus”, “mimivirus”, “asfarvirus”, “iridovirus” e “phycodnavirus”. **Fonte:** Modificado de (Abrahao, Dornas*, et al.*, 2014).

O primeiro membro do gênero *Mimivirus*, o *Acanthamoeba polyphaga mimivirus*, foi descoberto a partir de estudos que buscavam identificar e caracterizar micro-organismos patogênicos associados a amebas (MPAAs) em um contexto de um surto de pneumonia que ocorria na cidade de Bradford, Inglaterra, em 1992 (La Scola *et al.*, 2003). Após a coleta de amostras de água provenientes da torre de resfriamento de um hospital da cidade, posteriores caracterizações pela coloração de Gram demonstraram a presença de micro-organismos associados a amebas que se assemelhavam a cocos gram positivos, passando estes então a serem conhecidos como “cocos de Bradford” [Figura 2].

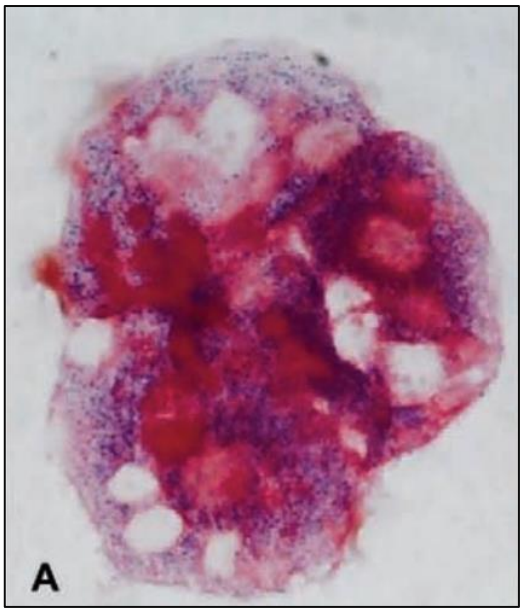


Figura 2: Imagem ilustrativa da coloração de Gram realizada em amostras de água de um hospital na cidade de Bradford, Inglaterra, em 1992. Os micro-organismos gram positivos corados em roxo foram denominados de “cocos de Bradford”. **Fonte:** (Raoult *et al.*, 2007)

A presença de um micro-organismo gram-positivo associado a um surto de pneumonia, ao contrário de típicos organismos gram-negativos como *Legionella sp.*, estimulou a tentativa de uma melhor caracterização microbiológica do possível patógeno. Diversas tentativas de isolamento e cultivo bacteriano foram realizadas, assim como testes de susceptibilidade à antimicrobianos e identificação do micro-organismo por meio da utilização de iniciadores universais para a amplificação do gene 16S de bactérias, porém todas essas tentativas foram frustradas (La Scola *et al.*, 2003; Abrahao, Dornas*, et al.*, 2014; La Scola, 2014). As amostras então foram estocadas e somente 10 anos depois, no início dos anos 2000, um grupo de pesquisadores franceses liderados pelo Dr. Didier Raoult decidiu retomar os estudos e realizar tentativas de identificação do micro-organismo por meio da sua observação através de microscopia eletrônica (La Scola, 2014). As imagens da microscopia revelaram de maneira surpreendente a presença de partículas muito grandes, de aproximadamente 400nm, que apresentavam uma arquitetura e simetria muito semelhantes àquelas já observadas para a maioria das partículas virais. Essas partículas eram ainda circundadas em toda sua extensão por fibrilas de aproximadamente 120-140 nm (La Scola *et al.*, 2003) [Figura 3].

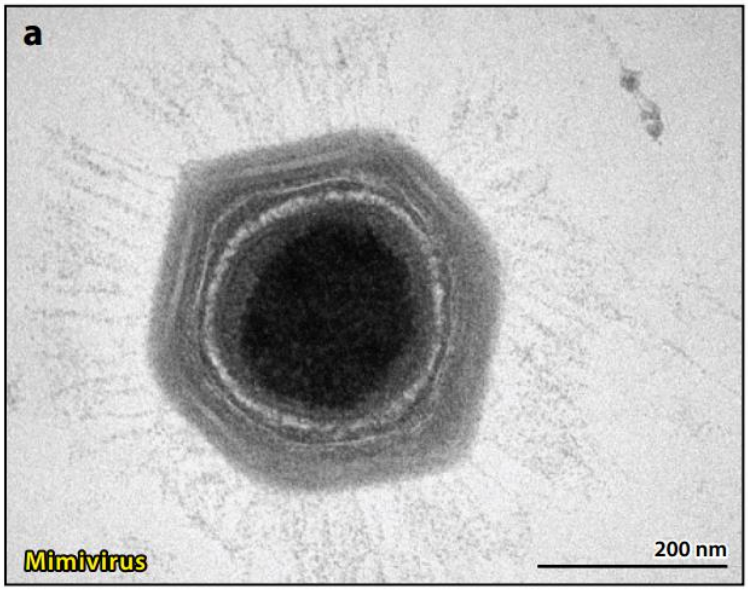


Figura 3: Imagem de uma partícula de mimivirus ao microscópio eletrônico de transmissão. É possível observar o capsídeo composto por muitas camadas proteicas, organizado em um arranjo bem semelhante ao de uma simetria icosaédrica, o que é bastante comum para outros tipos virais. A partícula é ainda circundada em toda sua extensão por fibras de aproximadamente 120-140 nm. **Fonte:** (Colson *et al.*, 2017)

Posteriores caracterizações também demonstraram a presença de um típico ciclo viral, incluindo uma fase de eclipse, o que fez com que os pesquisadores concluíssem que o recém descoberto micro-organismo se tratava na verdade de um vírus, o maior já isolado até aquela época (La Scola *et al.*, 2003; Raoult *et al.*, 2004). O novo vírus ficou então conhecido como *Acanthamoeba polyphaga mimivirus*, sendo este nome criado a partir do fato de inicialmente ter sido confundido com um outro micro-organismo (***mi****micking* ***mi****crobe*).

1.2.2 A família *Mimiviridae* e a sua diversidade

A partir da descoberta e isolamento do APMV, o primeiro vírus gigante, pesquisadores da área passaram a buscar por evidências da presença desses vírus em diversas partes do mundo. Em 2005, dois anos após a descoberta dos mimivírus, um grupo de cientistas realizou uma exaustiva busca de similaridade por genes preditos para o APMV em amostras do mar de Sargasso. Foi observado a presença de sequências muito similares a de mimivírus,

encontradas com relativa abundancia nessas amostras (Ghedin e Claverie, 2005). A constante busca por vírus gigantes relacionados ao APMV acabou por desencadear no isolamento de outras amostras virais durante os anos seguintes, provenientes dos mais diversos países e tipos de ambientes como torres de resfriamento, rios, lagos, solo, lentes de contato, etc [Figura 4]. Em 2008, uma nova amostra de mimivírus foi isolada a partir de água de uma torre de resfriamento em Paris, sendo esta denominada *Acanthamoeba castellanii mamavirus* (La Scola et al., 2008). Além da importância de se demonstrar a presença de vírus relacionados ao APMV em outras partes do mundo, o estudo de La Scola e colegas também se fez importante pela cunhagem do termo virófago, estabelecendo um tipo de relação nunca antes observado na virologia até aquele momento. Isso porque juntamente com o isolamento do mamavírus, foi identificado a presença de um segundo vírus denominado *Sputnik*, muito menor em tamanho, e que só se multiplicava em amebas caso o vírus gigante também estivesse presente sob multiplicação ativa. O vírus *Sputnik*, o primeiro virófago isolado, dependia da formação das fábricas virais do vírus gigante no interior da célula hospedeira para utilizar de sua RNA polimerase, dando continuidade ao seu ciclo de multiplicação (La Scola *et al.*, 2008). As novas amostras virais que vinham sendo isoladas, relacionadas ao APMV, passaram então a ser incluídas dentro de uma família viral conhecida como *Mimiviridae*, presente no gênero *Mimivirus*. Dentro desse gênero, de acordo com análises de similaridade baseadas na sequência da DNA polimerase B viral, esses novos isolados são divididos em três diferentes linhagens: a linhagem A, que contém a maior parte dos mimivírus isolados até o momento, como o mimivírus e o mamavírus; a linhagem B que contém vírus como o *Acanthamoeba polyphaga moumouvirus*, também isolado a partir de amostras de água de uma torre de resfriamento na França (Yoosuf *et al.*, 2012) e a linhagem C que contém vírus como o *Megavirus chilensis*, isolado a partir de amostras de água marinha da costa chilena (Arslan *et al.*, 2011). Um outro gênero de vírus gigante também está presente dentro da família *Mimiviridae*. Esse gênero, denominado *Cafeteriavirus*, inclui apenas um único membro, o *Cafeteria roenbergensis virus*, infectando microflagelados marinhos (Colson, Gimenez*, et al.*, 2011).

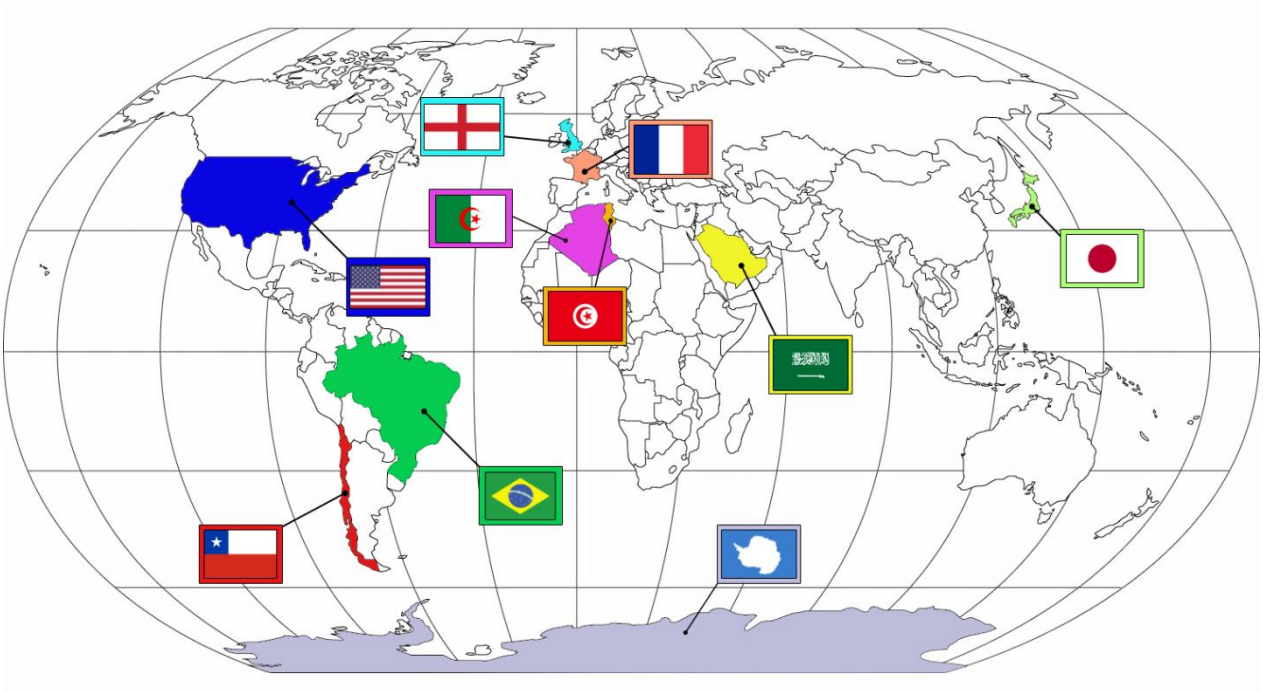


Figura 4: Locais de isolamento dos vírus pertencentes ao gênero *Mimivirus*. A presença desses vírus em diversos tipos de amostras como água de torres de resfriamento, lagos, solo, ambientes marinhos, lentes de contato, entre outros ambientes, denota o caráter ubíquo desses vírus. **Fonte:** Adaptado de Rodrigues, 2015.

1.2.3 O ciclo de multiplicação dos mimivírus

Assim como os outros membros do NCLDV, o ciclo de multiplicação dos mimivírus se estabelece no citoplasma da célula hospedeira, em amebas do gênero *Acanthamoeba* [Figura 5]. Inicialmente as partículas virais são fagocitadas pela célula amebiana e internalizadas por meio da formação de vacúolos (Abrahao, Dornas*, et al.*, 2014). Já totalmente no interior do citoplasma, o vírus inicia o processo de liberação de seu material genético de DNA dupla fita por meio da abertura de um vértice modificado presente em seu capsídeo, também chamado de *stargate* (Zauberman *et al.*, 2008) [Figura 5 e 6]. A presença desse vértice modificado, e o arranjo das subunidades do capsideo acaba estabelecendo à partícula viral a organização em uma simetria pseudo-icosaédrica (Xiao *et al.*, 2009). Após a abertura do *stargate*, a liberação do material genético na forma de uma semente viral se segue por meio da fusão da membrana do fagossomo celular com uma membrana lipídica localizada no

interior do capsídeo do vírus gigante, sendo aí onde está contido o genoma do vírus (Zauberman *et al.*, 2008; Xiao *et al.*, 2009). Durante esse estágio inicial, está caracterizado o começo da fase de eclipse, na qual não é possível se observar nenhuma partícula viral sendo formada. Após a entrega da semente viral ao citoplasma da célula hospedeira, tem início a formação de uma fábrica viral precoce, que posteriormente dará origem a uma fábrica viral madura (Kuznetsov *et al.*, 2013). É na fábrica viral que ocorrerá a replicação e a transcrição do DNA do vírus, sendo este preparado posteriormente para ser encapsidado em uma abertura transiente, localizada no capsídeo viral, em um sítio diferente do *stargate* (Zauberman *et al.*, 2008). A morfogênese das partículas virais se segue na periferia das fábricas, até que as partículas sejam então liberadas por meio da lise da célula amebiana.

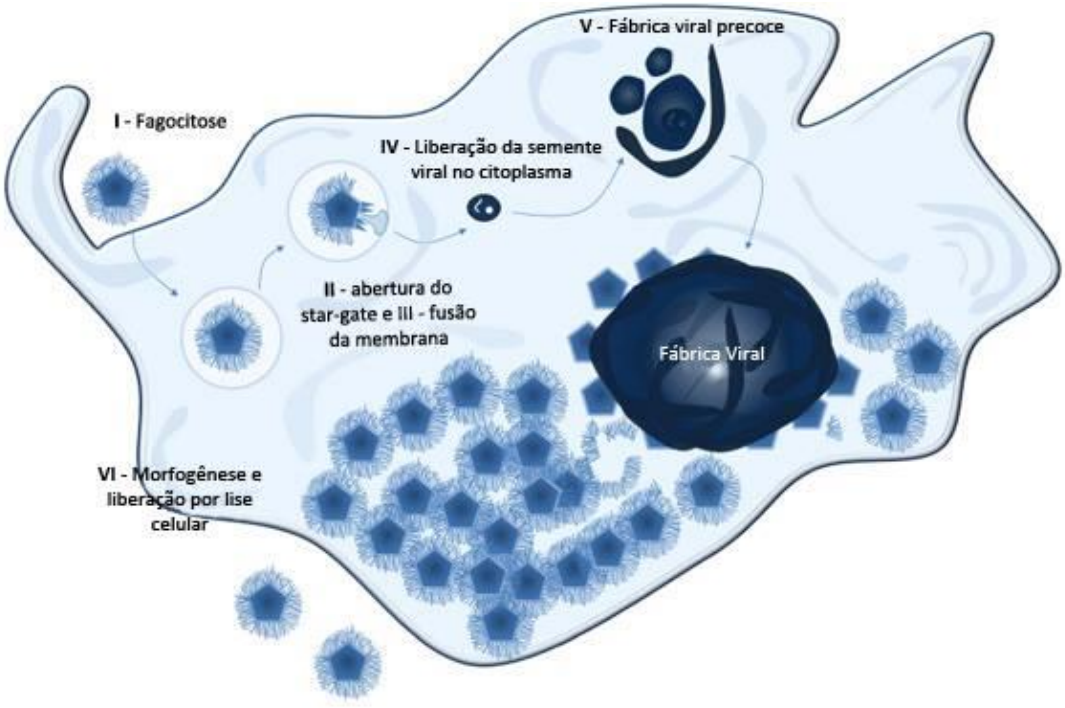


Figura 5: Ciclo de multiplicação do APMV. (I) Inicialmente as partículas do APMV são internalizadas no interior de fagossomos através de eventos de fagocitose. (II, III e IV) Logo após, o genoma do vírus é liberado na forma de uma semente viral através da abertura de um vértice modificado da partícula, também chamado de *stargate*, com posterior fusão da membrana interna viral e da membrana do fagossomo da célula. (V) A semente viral dá origem à formação de uma fábrica viral precoce que posteriormente se transforma em uma fábrica viral

madura. (VI) A morfogênese das partículas se estabelece na periferia da fábrica viral, sendo a liberação destas provocada pela lise da célula. **Fonte:** (Abrahao, Dornas*, et al.*, 2014)

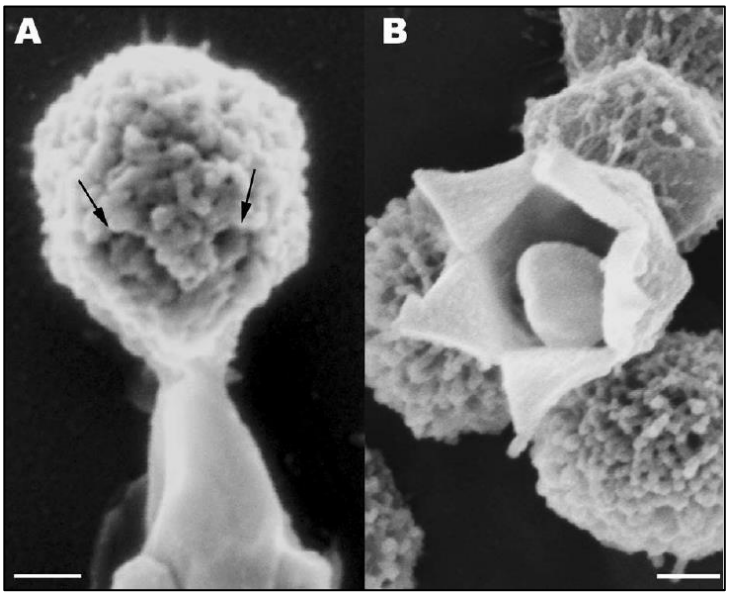


Figura 6: (A e B) Imagem de microscopia eletrônica de varredura evidenciando a extrusão do genoma do APMV através da abertura de um vértice modificado presente no capsídeo, também chamado de stargate. **Fonte:** (Zauberman *et al.*, 2008)

1.2.4 Como os mimivírus expandiram as barreiras da virologia

O processo de caracterização do primeiro mimivirus permitiu com que muitos dogmas dentro da virologia fossem revistos quando nos referimos à maneira como os vírus são classificados. Historicamente, um organismo pode ser denominado como um vírus baseando-se em um conjunto de características que apresentam um caráter puramente excludente (Lwoff, 1957; Abrahao *et al.*, 2017). Muitos exemplos podem ser vistos ao observarmos com cuidado a maioria dos vírus atualmente reconhecidos pelo ICTV. Geralmente, eles acabam se encaixando nesse perfil quando consideramos que em sua grande parte estes são classificados dessa forma por *não* serem agentes filtráveis em membranas de 0,2 µm, por *não* apresentarem genomas muito grandes ou complexos, por *não* serem visíveis à técnica de microscopia óptica, por *não* apresentarem genes relacionados ao processo de tradução, etc. Com os mimivírus, todos esses e muitos outros conceitos passaram a ser quebrados logo assim que foram

demonstradas características que refutavam as ideias explicitadas acima, como por exemplo: os mimivírus apresentam um grande tamanho de partícula viral, com dimensões que chegam em torno dos 750 nanômetros (nm). Eles também apresentam no entorno da partícula uma série de estruturas denominadas fibrilas, importantes para a adesão nos mais diferentes organismos como bactérias, fungos e artrópodes (Rodrigues *et al.*, 2015). Seu genoma detém de um tamanho e complexidade nunca antes vistas em outros vírus, apresentando cerca de 1,2 Mb e codificando para quase 1000 proteínas. Os mimivírus também apresentam um mecanismo próprio de glicosilação, com glicosiltransferases envolvidas tanto na biossíntese de glicanos como também em modificações pós traducionais (Luther *et al.*, 2011). Além disso, há também a característica mais marcante desses vírus com relação ao restante da virosfera, que é a presença de genes que codificam elementos envolvidos diretamente no processo de tradução, as chamadas aminoacil-tRNA-sintetases (aaRs), que são exclusivas para os mimivírus, e outros fatores também importantes mas não exclusivos, como RNAs transportadores (tRNAs) e fatores de tradução [Tabela 1] (La Scola *et al.*, 2003; Piacente *et al.*, 2012; Piacente *et al.*, 2014; Abrahao *et al.*, 2017).

Tabela 1 – Genes envolvidos com a tradução em diferentes linhagens de mimivírus

Grupo/vírus	aaRs	tRNA	Fatores de tradução
Mimivírus linhagem A			
APMV	arginil-RS, cisteinil-RS, metionil-RS, tirosil-RS	leucina (3x), histidina, cisteína, triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1
Mamavirus	arginil-RS, cisteinil-RS, metionil-RS, tirosil-RS	leucina (3x), histidina, cisteína, triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1
Lentille	arginil-RS, cisteinil-RS, metionil-RS, tirosil-RS	leucina (3x), histidina, cisteína, triptofano	IF4A, IF4E, eF-TU, eRF1
Hirudovirus	arginil-RS, cisteinil-RS, metionil-RS, tirosil-RS	leucina (3x), histidina, cisteína, triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1

Samba virus	arginil-RS, cisteinil-RS, metionil-RS, tirosil-RS	leucina (3x), histidina, cisteína, triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1
Oyster virus	arginil-RS (2x), cisteinil-RS, metionil-RS, tirosil-RS	leucina (3x), histidina, cisteína, triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1
Amazonia virus	cisteinil-RS, tirosil-RS	leucina (3x), histidina, cisteína, triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1
Terra2	arginil-RS, cisteinil-RS, metionil-RS, tirosil-RS	leucina (2x), histidina, cisteína, triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1
Bombay	arginil-RS, cisteinil-RS, metionil-RS, tirosil-RS	leucina (3x), histidina, cisteína, triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1
Mimivírus linhagem B			
Moumouvirus	arginil-RS (4x), cisteinil-RS, isoleucil-RS, metionil-RS, tirosil-RS	leucina, histidina, cisteina	IF4E, SUI1, eF-TU, eRF1
Goulette	cisteinil-RS, metionil-RS	leucina (3x), histidina, cisteina	IF4E, SUI1, eF-TU, eRF1
Monve	arginil-RS (2x), asparagil-RS, cisteinil-RS, isoleucil-RS (2x), metionil-RS, tirosil-RS	leucina, histidina, cisteina	IF4A, IF4E (2x), SUI1, eRF1
Mimivírus linhagem C			
Megavirus chilensis	arginil-RS, asparagil-RS, cisteinil-RS, isoleucil-RS, metionil-RS, trpitofanil-RS, tirosil-RS	leucina (2x), triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1
Terra1	arginil-RS, cisteinil-RS, metionil-RS, tirosil-RS	leucina, triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1

LBA111	arginil-RS, asparagil-RS, cisteinil-RS, isoleucil-RS, metionil-RS, trpitofanil-RS, tirosil-RS	leucina (2x), histidina, cisteína, triptofano	IF4A, IF4E, SUI1, eF-TU, eRF1
Courdo7	isoleucil-RS, tirosil-RS	leucina (3x), triptofano	IF4A (2x), IF4E, SUI1, eRF1
Courdo11	arginil-RS, asparagil-RS (2x), cisteinil-RS, isoleucil-RS, metionil-RS, triptofanil-RS, tirosil-RS	leucina (3x), histidina, cisteina, triptofano	IF4A (2x), IF4E, SUI1, eRF1

Fonte: (Abrahao *et al.*, 2017) – **modificado**

1.3 Aminoacil-tRNA-sintetases

As aaRs são enzimas bastante antigas, tendo surgido há cerca de 3,5 bilhões de anos atrás (Ribas De Pouplana e Schimmel, 2001). Essas estão envolvidas no processo de tradução, sendo responsáveis por catalisar uma reação de aminoacilação entre um aminoácido específico e a porção 3’ do seu tRNA correspondente, determinando na formação de uma molécula de aminoacil- tRNA (Woese *et al.*, 2000; Ribas De Pouplana e Schimmel, 2001; Li *et al.*, 2015). Devido à sua ação bioquímica e ao fato de sua evolução poder estar relacionada com desenvolvimento do código genético atual, as aaRs são bastante exploradas em trabalhos envolvendo estudos sobre a origem da vida (Ribas De Pouplana e Schimmel, 2001; Fournier *et al.*, 2011). Com um total de 20 tipos de aaRs conhecidas atualmente, essas enzimas podem ser divididas em dois grupos distintos de acordo com o tipo de dobramento estrutural de seus sítios ativos, as de classe I e as de classe II (Li *et al.*, 2015). As aaRs apresentam ainda uma ampla distribuição dentre membros pertencentes aos três Domínios da vida, tendo sido descritas em vírus apenas recentemente, por meio da caracterização de amostras de vírus gigantes [Tabela 1] (Raoult *et al.*, 2004; Abrahao, Dornas*, et al.*, 2014; Abrahao *et al.*, 2017). Esse fator fez com que,

antes dos mimivírus, essas enzimas fossem consideradas importantes marcadores presentes apenas em organismos celulares (Woese *et al.*, 2000; Fournier *et al.*, 2011).

A caracterização genética do APMV permitiu a observação de um total de quatro cópias de aaRs (arginil-RS, cisteinil-RS, metionil-RS e tirosil-RS) em seu genoma, algo nunca visto até então para outras espécies virais, sendo todas essas pertencentes às enzimas de classe I (Raoult *et al.*, 2004). Outras amostras relacionadas aos mimivírus foram caracterizadas ao longo dos anos, demonstrando a presença de quantidades semelhantes de aaRs, como para os Mamavirus, Samba virus, Hirudovirus e Terra1 virus (Colson, Yutin*, et al.*, 2011; Boughalmi *et al.*, 2013; Campos *et al.*, 2014; Yoosuf *et al.*, 2014; Boratto *et al.*, 2015). No entanto, à medida que novos isolados foram sendo estudados, sua caracterização impressionou pela observação de conjuntos de genes codificantes para aaRs cada vez mais complexos, como é o caso do *Megavirus chilensis*, contendo um total de 6 aaRs, o *Acanthamoeba polyphaga moumouvirus*, contendo um total de 8, além dos Klosneuvirus e Tupanvirus que contém aaRs relacionadas a todos os 20 aminoácidos conhecidos (Arslan *et al.*, 2011; Yoosuf *et al.*, 2012; Schulz *et al.*, 2017) (dados do Laboratório de Vírus já aceitos para publicação).

A presença de aaRs nos vírus gigantes ainda intensificou fortemente os debates a respeito de teorias que se relacionam com a evolução desses vírus. Uma regra geral nos isolados de vírus gigantes atuais é a de que em mimivírus cujos genomas apresentam conjuntos menos ricos de aaRS, é impossível de se observar qualquer aaRs de caráter exclusivo quando comparado com genomas de mimivírus com conjuntos mais ricos (Abrahao *et al.*, 2017). Diante desse cenário, muitos estudos passaram a sugerir que, evolutivamente, as amostras de mimivírus atuais teriam se originado a partir de um ancestral mais complexo, por meio de um processo de perda de genes (Nasir *et al.*, 2012; Nasir e Caetano-Anolles, 2015). Existem porém ainda outros estudos que refutam essa hipótese, com afirmações que sugerem que grande parte das aaRs teriam se originado nos mimivírus a partir de eventos de transferência gênica horizontal (TGH) (Yutin *et al.*, 2014; Koonin *et al.*, 2015). Tomando como base a ideia dos últimos autores, estes defendem o surgimento dos mimivírus a partir de organismos mais simples, se tornando cada vez mais complexos geneticamente através da

aquisição de novos genes incorporados por TGH. Uma última hipótese propõe o gigantismo do genoma desses vírus, excluindo a tendência clássica de se pensar em evolução direcionada apenas ao ganho ou apenas à perda de genes. Nessa ideia, denominada de “hipótese do acordeão”, os vírus gigantes teriam sofrido sucessivos ciclos de expansão e redução do genoma, com respostas adaptativas à diversas modificações que ocorressem tanto nas condições ambientais, como também de novos hospedeiros (Filee, 2015).

1.4 “Core” ou “hallmark genes”

Em um tópico anterior foi mencionado que os membros de um grupo artificial de vírus conhecido como NCLDV compartilham entre si um conjunto de genes chamados de “*core”* ou “*hallmark genes”*, sendo estes responsáveis por conferir a esses vírus um caráter hipoteticamente monofilético (Iyer *et al.*, 2001; Iyer *et al.*, 2006). Dentre os genes pertencentes a esse grupo pode-se observar aqueles que codificam para proteínas envolvidas na replicação do material genético, no processo de transcrição e na morfogênese viral, sendo o fato mais importante de que esses *core genes* apresentam pouca ou nenhuma similaridade com componentes celulares, reforçando a ideia de que os diferentes membros que compõe esse grupo tenham surgido a partir de um único ancestral comum (Iyer *et al.*, 2001). Em um estudo anterior, por meio da análise de genes conservados presentes em diferentes membros do NCLDV, foi possível traçar a presença de pelo menos 47 deles codificando para proteínas no ancestral desse grupo (Yutin *et al.*, 2009). Isso significa que, apesar do tamanho do genoma hipotético desse vírus ancestral ser grande quando comparado a muitos outros grupos de vírus atuais, quando analisamos a relação com os membros do NCLDV este ancestral apresenta uma quantidade muito inferior desses elementos, sugerindo que durante a evolução desses vírus possa ter havido um processo de aquisição em massa de genes. Existe também a possibilidade de que o ancestral dos NCLDVs tenha apresentado muito mais genes do que os membros atuais, mas segundo Iyer e colaboradores, essa hipótese perde força quando devemos considerar que para isso acontecer um grande número de perdas independentes de genes tivesse que ocorrer (Iyer *et al.*, 2006). Dentre os *“core genes”* compartilhados

exclusivamente entre todos os membros do grupo NCLDV podem ser citados cinco: uma helicase-primase, uma DNA polimerase do tipo B, uma DNA-ATPase, um fator de transcrição viral tardio e a proteína principal do capsídeo (MCP) (Colson *et al.*, 2013). Dentre esses o gene para a proteína principal do capsídeo se revela um interessante objeto de estudo, principalmente com relação a termos de estrutura gênica e relações filogenéticas entre vírus gigantes, temas ainda pouco explorados na literatura.

1.5 Proteína principal do capsídeo (MCP)

O gene que codifica para a MCP é o gene estrutural mais importante na composição de partículas maduras dos vírus gigantes. Em um dos estudos que comprovam essa afirmação, por meio de experimentos de anotação genômica, Renesto e colaboradores observaram que na formação da partícula do APMV a proteína principal do capsídeo é a glicoproteína que se apresenta de maneira mais abundante (Renesto *et al.*, 2006). Em um outro estudo, no qual os autores utilizavam da ferramenta tBLASTN em seus experimentos, análises do gene da MCP permitiram observar um pouco da maneira como ele estaria organizado nos vírus gigantes. Foi possível ver que, pelo menos com relação ao APMV (protótipo da família *Mimiviridae*), o gene para a MCP se apresenta organizado na forma de três regiões exônicas, separadas entre si por duas regiões não codificantes ou íntrons (Azza *et al.*, 2009). Desde então, a maneira como o gene da MCP está organizado foi avaliada apenas mais uma única vez, no isolado *Acanthamoeba castellanii mamavirus*. Nesse vírus foi observado que este gene é formado por duas regiões exônicas, homólogas à regiões presentes no APMV, sendo estas separadas entre si por um único íntron, de sequência exclusiva para essa espécie viral (Colson, Yutin*, et al.*, 2011). Devido a esse processo, ainda se mantém bastante escassas as informações disponibilizadas na literatura acerca do modo de organização e do processamento desse gene por *splicing* durante a formação do RNAm em outros isolados de vírus gigantes. Outras informações de caráter incompleto estão relacionadas com a maneira como os diferentes isolados de mimivírus se relacionam filogeneticamente quando utilizamos esse gene como base. Esse fator acaba se tornando importante de maneira a

esclarecer uma potencial utilização do gene da MCP como um marcador genético nos vírus gigantes da família *Mimiviridae*, visto ser esse um dos chamados "*core genes"*.

1.6 Amebas de vida livre

Apesar de não ser o foco deste trabalho, é importante se fazer uma breve introdução sobre os únicos hospedeiros descritos atualmente para os vírus gigantes acima citados. Esses hospedeiros fazem parte de um conjunto de indivíduos conhecidos como amebas de vida livre. Esse termo é utilizado na caracterização de um grupo heterogêneo composto tanto por amebas endoparasitas obrigatórias (como *Entamoeba histolytica*) quanto por amebas parasitas oportunistas (como *Acanthamoeba*), não apresentando nenhuma relevância taxonômica ou filogenética (Scheid, 2014). Os membros que compõe esse grupo de amebas apresentam um caráter ubíquo, tendo sido isolados dos mais diversos ambientes como no solo, ambientes aquáticos, aéreos e inclusive de vertebrados (La Scola, 2014). Esses organismos são conhecidos como fagócitos profissionais, se alimentando principalmente de bactérias, a partir de um processo independente da participação de receptores conhecido como fagocitose (Figura 7). Nesse processo, a internalização da partícula depende exclusivamente do tamanho do elemento a ser fagocitado (partículas acima de 500 nm), ocorrendo então posterior digestão desses elementos em vacúolos denominados de fagossomos (Aderem *et al.* 1999, La Scola 2014). As amebas de vida livre correspondem aos principais consumidores de bactérias no meio ambiente, sendo responsáveis por até 60% de redução da população bacteriana no meio em que estão presentes (Sinclair *et al.*, 1981; Siddiqui e Khan, 2012). O fato de terem bactérias como fonte principal de sua alimentação faz com que as amebas de vida livre apresentem um importante papel ecológico de controle das populações microbianas no ambiente, além do próprio papel na ciclagem de nutrientes, visto esses organismos serem importantes decompositores secundários, liberando no ambiente minerais e nutrientes que muitas vezes estão associados à biomassa dos decompositores primários (como bactérias) (Siddiqui e Khan, 2012). Alguns micro-organismos, no entanto, muitos deles

patogênicos, são conhecidos por serem resistentes ao processo de fagocitose promovido pelas amebas de vida livre, estando muitas vezes adaptados ao ambiente intracelular desses protistas. Isso também faz dessas amebas importantes veículos de transporte de diversos patógenos, além da proteção desses micro-organismos patogênicos contra as condições mais adversas do meio ambiente, como radiação ultravioleta, dessecamento, biocidas, etc (Boratto *et al.*, 2014; Scheid, 2014).

1.7 Gênero *Acanthamoeba* e características gerais

As amebas do gênero *Acanthamoeba* compreende um dos grupos de amebas de vida livre mais estudados nos últimos anos. Esse destaque é derivado de sua importância em diversas áreas de conhecimento como a sua participação como causadora de doenças oportunistas em humanos, sua importância no controle ecológico de populações microbianas e ciclagem de minerais, e mais recentemente, como principais hospedeiros dos vírus gigantes, que foram importantes para a mudança de alguns paradigmas dentro da virologia moderna (Figura 8). Amebas desse gênero foram inicialmente descobertas por Castellanii, em 1930, quando este percebeu a presença de amebas em culturas de *Cryptococcus pararoseus* (Castellanii, 1930). O nome do gênero, no entanto, só foi estabelecido um ano mais tarde, em 1931, por Volkonsky, no qual apenas amebas apresentando projeções em forma de espinho, chamadas de acantopódios, eram incluídas no grupo (Volkonsky, 1931).

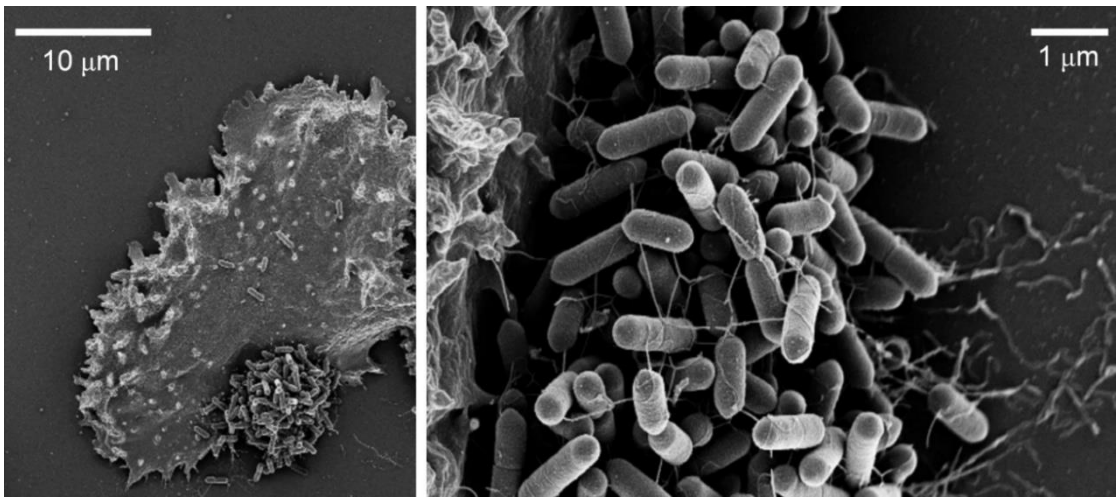


Figura 7: Microscopia eletrônica de varredura demonstrando uma ameba da espécie *Acanthamoeba castellanii* promovendo a fagocitose de um aglomerado de bactérias da espécie *Lysteria monocytogenes*. **Fonte:** Doyscher, 2007.

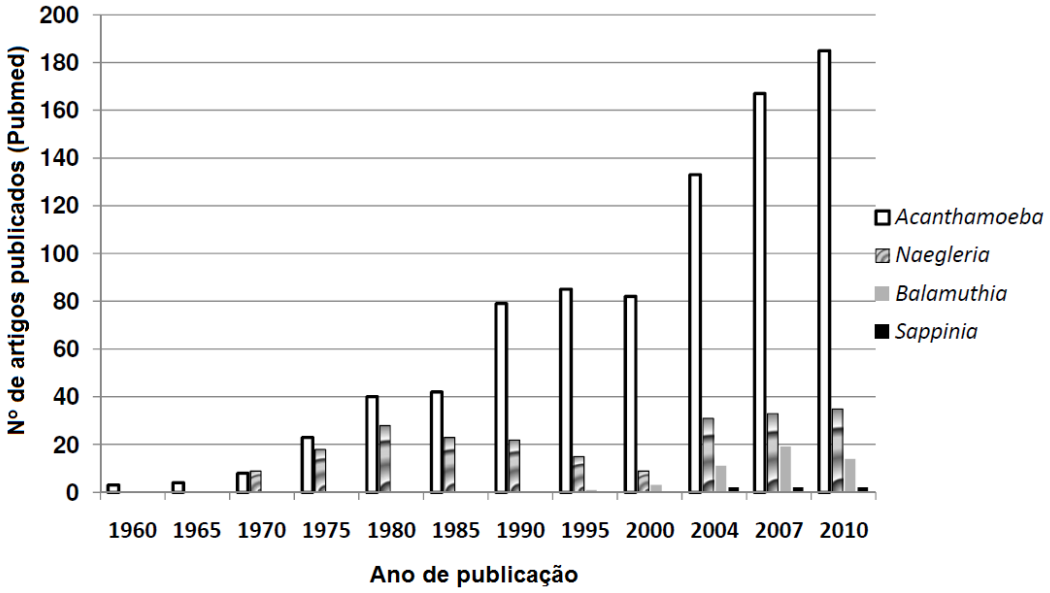


Figura 8: Aumento do interesse na pesquisa cientifica de amebas do gênero *Acanthamoeba*, de acordo com o tempo. Os dados foram levantados de acordo com o número de artigos publicados que são observados na base de dados do Pubmed ao se utilizarem como termos de busca os nomes de conhecidos gêneros de amebas de vida livre: "*Acanthamoeba", "Naegleria", "Balamuthia"* e "*Sappinia"*. **Fonte:** Modificado de Siddiqui e Khan, 2012.

As amebas do gênero *Acanthamoeba* apresentam um ciclo de vida baseado em duas formas celulares: (1) os trofozoítos, que são uma forma celular vegetativa e metabolicamente ativa e (2) os cistos, uma forma celular de resistência, metabolicamente inerte (Figura 9) (Marciano-Cabral e Cabral, 2003). É através da forma trofozoítica que essas amebas realizam suas principais funções no ambiente, como alimentação, locomoção e reprodução. A alimentação ocorre principalmente por meio da fagocitose, sendo responsável pela ingestão de partículas acima de 0,5µm, sem a necessidade do reconhecimento por qualquer tipo de receptor. No entanto, essa alimentação também pode ocorrer por meio de dois outros processos: (1) pela formação de estruturas temporárias utilizadas para a ingestão de bacterias, fungos e outras células, chamadas de *food cups*, ou (2) de forma axênica, através da captação

de nutrientes dissolutos no ambiente, por meio de pinocitose (Bowers, 1977; Bowers e Olszewski, 1983; Marciano-Cabral e Cabral, 2003). A locomoção ocorre pela formação de um pseudópode hialino, que exclui a presença de grânulos citoplasmáticos, e é importante tanto no processo de alimentação quanto de quimiotaxia (Preston e King, 1984; Schuster e Levandowsky, 1996). Em relação à reprodução, pelo que se conhece até hoje as amebas do gênero *Acanthamoeba* se reproduzem exclusivamente de forma assexuada, através de fissão binária, em que uma célula-mãe se divide e promove o aparecimento de duas células filhas, geneticamente semelhantes. No entanto, estudos recentes descrevem nessas amebas a presença de alguns genes relacionados à divisão meiótica, homólogos a genes de mesma função presentes em outros eucariotos, como plantas, fungos e animais, sugerindo também a utilização de um modelo de reprodução sexuada por parte desses micro-organismos (Khan e Siddiqui, 2015). A outra forma celular descrita para amebas desse gênero, chamada de cisto, é metabolicamente inerte, ou seja, nela o organismo se encontra em um estado criptobiótico, fazendo com que grande parte de seus processos metabólicos sejam pausados. O cisto também compreende uma forma de resistência celular no qual há uma massiva transformação do aspecto da célula, fazendo com que essa se transforme de um estágio que apresenta predominantemente características ameboides para um estágio mais arredondado, no qual a célula passa a adquirir uma dupla parede celular (endocisto e ectocisto) responsáveis pelo aumento da resistência do organismo a variadas condições adversas, como grandes variações de temperatura, variações de pH, luz ultravioleta, biocidas químicos, etc. (Marciano-Cabral e Cabral, 2003; Lloyd, 2014).

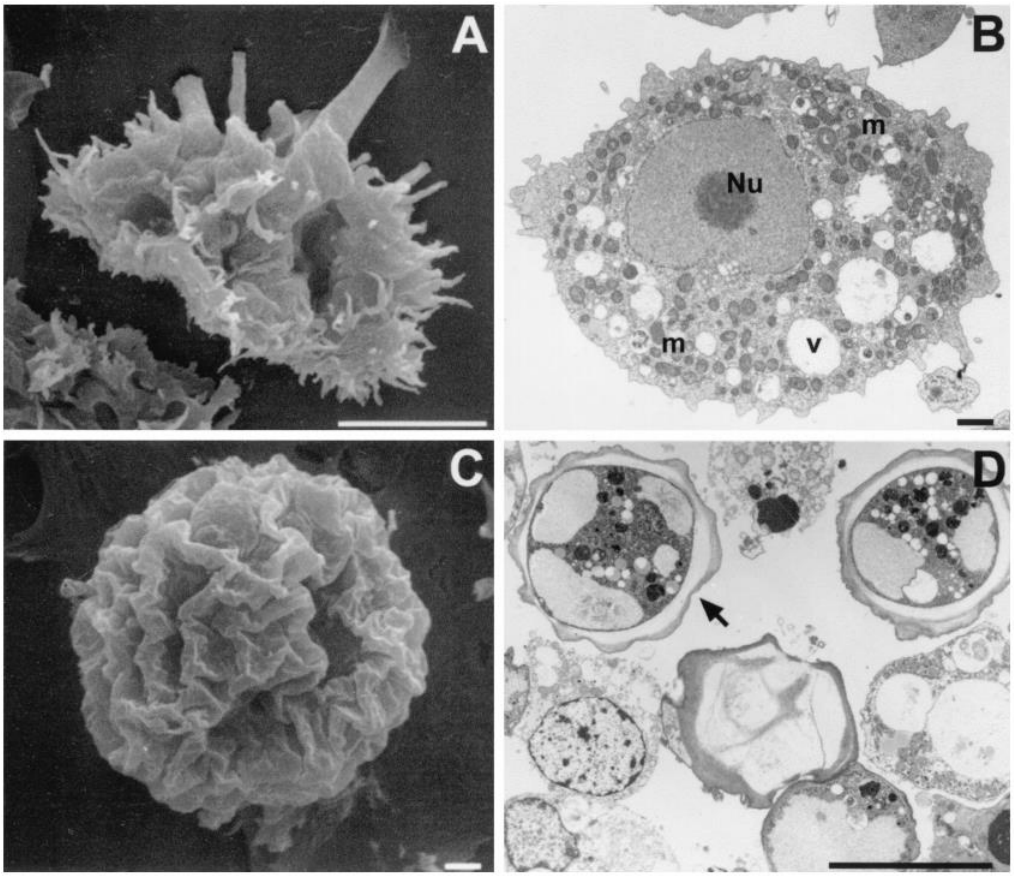


Figura 9: Microscopias eletrônicas de varredura (A e C) e transmissão (B e D) demonstrando as duas formas celulares presentes em amebas do gênero *Acanthamoeba*. Imagens A e B demonstram a forma metabolicamente ativa, chamada de trofozoítos. Na imagem A é possível se observar as projeções em formato de espinho, chamadas de acantopódios. Imagens C e D representam a forma criptobiótica, metabolicamente inerte. Em D, a seta preta indica a presença da dupla parede presente na forma cística. **Fonte:** Marciano-Cabral *et al.* 2003

As amebas do gênero *Acanthamoeba* também são conhecidas por serem agentes causadores de algumas doenças em humanos. Dentre os principais tipos de doença causadas por esses micro-organismos pode-se citar (1) a encefalite granulomatosa amebiana, que apesar de rara, se caracteriza por ser uma doença fatal que afeta o sistema nervoso central e também (2) a ceratite amebiana, que se caracteriza como uma infecção do tecido ocular, associada principalmente a pessoas que utilizam lentes de contato. A primeira sugestão desses organismos como patógenos humanos surgiu em 1958, a partir de testes para o estudo da segurança de vacinas de pólio. O aparecimento de placas em culturas celulares utilizadas para o preparo das vacinas, além da morte por

encefalite causada em camundongos e em macacos inoculados com os fluidos dessa cultura de tecidos foi importante pois não só levou à ligação de amebas do gênero *Acanthamoeba* como causadoras de doenças em humanos, como também permitiu os primeiros estudos que previam o papel de outras amebas de vida livre como agentes patogênicos (Culbertson *et al.*, 1959; Marciano-Cabral e Cabral, 2003).

2. JUSTIFICATIVA

Desde 2003, estudos de prospecção tem promovido o isolamento de vírus gigantes da família *Mimiviridae* a partir dos mais diversos tipos de ambientes, como em amostras de solo, rios, lagoas, oceanos, água de esgoto e até mesmo de amostras clínicas de seres humanos. Por meio de estudos de caracterização desses isolados, muitas características nunca antes vistas na virosfera acabaram sendo descritas, estando estas relacionadas principalmente ao grande tamanho desses novos vírus, às suas capacidades metabólicas e à complexidade do seu genoma. Dois genes de grande importância, mas ainda não tão bem explorados na literatura, correspondem àqueles que codificam para as aminoacil-tRNA-sintetases (aaRs) e ao gene codificante para a proteína principal do capsídeo (MCP). O primeiro tem apresentado uma participação bastante importante em estudos evolutivos que buscam entender a ancestralidade dos vírus gigantes e inclusive a de um possível papel desses organismos na árvore da vida. Isso se deve principalmente ao fato dessas enzimas apresentarem raízes de surgimento muito antigas na história da Terra, além de terem sido consideradas (até a descoberta dos mimivírus) como presentes tipicamente apenas em organismos formados por células. No entanto, para uma enzima cuja presença demarcava os limites entre aqueles organismos considerados celulares e aqueles considerados virais, ainda poucos são os trabalhos que tentam entender os impactos das aaRS no *fitness* dos vírus gigantes e nas pressões seletivas que determinam na conservação dessas enzimas nesse grupo de vírus. Com relação ao gene da MCP, o que se sabe atualmente é a existência de uma grande pressão de conservação desse gene entre todos os membros do grupo de Vírus Grandes Nucleocitoplasmáticos de DNA (NCLDVs), além do curioso fato deste gene se apresentar estruturado por uma mescla de regiões intrônicas e exônicas no protótipo da família *Mimiviridae*. Dessa maneira, torna-se interessante analisar não só como o gene da MCP está estruturado em outros membros dessa família, como também investigar a sua possível utilização como um marcador molecular capaz de separar os mimivírus em suas diferentes linhagens, dado a pressão de conservação desse gene nesses organismos.

3. OBJETIVOS

3.1 Objetivo geral

* Isolar e caracterizar amostras de mimivírus obtidas a partir de duas lagoas urbanas localizadas em Minas Gerais, Brasil.

3.2 Objetivos específicos

* Caracterizar biologicamente o ciclo de multiplicação dos vírus isolados por meio da realização de curvas de ciclo único e de análises de microscopia eletrônica de transmissão.

* Caracterizar morfometricamente o tamanho das partículas virais isoladas e também de seus componentes.

* Avaliar a presença e distribuição dos tipos de aaRs nos vírus isolados, bem como o perfil filogenético dessas enzimas em comparação com outros membros da família *Mimiviridae*.

* Caracterizar estruturalmente o gene que codifica para a MCP em diferentes membros das três linhagens da família *Mimiviridae*, incluindo os vírus isolados nesse trabalho.

* Avaliar o papel do gene da MCP como um possível marcador molecular de classificação filogenética dos membros da família *Mimiviridae*.

5. MATERIAIS E MÉTODOS

5.1 Materiais

5.1.1 Meio de cultura PYG (peptona/ extrato de levedura/ glicose)

Para o cultivo de amebas da espécie *Acanthamoeba castellanii*, o meio de cultura utilizado foi o PYG (protease peptona, extrato de levedura e glicose). Este é o meio comumente usado em cultivo de amebas de vida livre, tendo em sua composição, para cada um litro de meio: 8 µM de sulfato de magnésio hepta hidratado ($MgSO_4.7H_2O$) (Merck, Alemanha), 0,5 µM de cloreto de cálcio ($CaCl_2$) (Merck, Alemanha), 5,0 nM de sulfato de ferro amoniacal hexa hidratado ($Fe(NH_4)2(SO_4).6H_2O$) (Merck, Alemanha), 1,4 mM de fosfato dibásico de sódio hepta hidratado ($Na_2HPO_4.7H_2O$) (Merck, Alemanha), 2,5 mM de fosfato monobásico de potássio (KH_2PO_4) (Merck, Alemanha), 3,4 mM de citrato de sódio dihidratado ($C_6H_5Na_3O_7.2H_2O$) (Merck, Alemanha), 20 g de extrato bactopeptona (Merck, Alemanha,), 2,0 g de extrato de levedura (BD, França), 0,05 M de glicose (Merck, Alemanha) e água destilada. Após a homogeneização dos ingredientes na água, o pH do meio é ajustado para 6,5 e este é então autoclavado por 15 minutos a 121°C. O meio então é resfriado em temperatura ambiente e submetido a uma filtração em membranas de 0,22 µm (Millipore, EUA) para retirada de eventuais cristais que se formam durante a sua produção e causam danos em cultura de amebas. Após esse passo, o meio é armazenado em câmara fria (temperatura de 4 a 8°C) até que este possa ser utilizado. Para isso, de forma paralela, o meio deve passar por um teste de esterilidade no qual 1 mL de PYG é incubado com 9 mL de solução tioglicolato, por 7 dias a 37°C. Não ocorrendo a contaminação do meio durante esse prazo, o PYG está pronto para uso após sua suplementação com Soro Fetal Bovino (SFB) 7%, (Cultilab, Brasil), 200 U/mL de Penicilina (Cristália, Brasil), 50 µg/mL de Estreptomicina (Sigma, EUA) e 2,5 µg/mL de Anfotericina B (Sigma, EUA). O estoque do meio é promovido em câmara fria, a 4^oC.

5.1.2 Meio água-arroz

O meio água-arroz é descrito na literatura como um meio de enriquecimento (Arslan *et al.*, 2011; Campos *et al.*, 2014), sendo este produzido a partir da suplementação de 4% de arroz para cada litro de água destilada (40 grãos de arroz para cada litro de água). Após suplementação o meio é autoclavado durante 15 min a 121ºC, e após resfriamento em temperatura ambiente, apenas o sobrenadante é utilizado em experimentos visando o isolamento de vírus gigantes. O estoque do meio é promovido em câmara fria, a 4ºC.

5.1.3 Meio *Page amoeba saline* (PAS)

A salina PAS é considerada um meio não-nutritivo bastante simples, utilizada na manutenção de amebas do gênero *Acanthamoeba*. Devido à ausência de nutrientes, esse meio também é bastante utilizado em experimentos que visam impedir a multiplicação dessas amebas, e em maior prazo, no estímulo ao seu encistamento (Silva *et al.*, 2015). Para a produção desse meio primeiramente devem ser feitas duas soluções de maneira separada. À primeira solução adiciona-se 0,142g de Na_2HPO_4 (Merck, Alemanha), 0,136g de KH_2PO_4 (Merck, Alemanha) e 500mL de água destilada. À segunda solução são adicionados 4,0mg de $MgSO_4.7H_2O$ (Merck, Alemanha), 4,0mg de $CaCL_2.2H_2O$ (Merck, Alemanha), 0,120g de NaCL (Merck, Alemanha) e 500mL de água destilada. As duas soluções são então autoclavadas a 121ºC por 20 minutos, separadamente, e posteriormente (após esfriamento) essas podem ser misturadas no interior de fluxo laminar. O estoque do meio é promovido em câmara fria, a 4ºC.

5.1.4 Solução tampão fosfato salino (PBS)

Para a produção da solução de PBS, primeiramente é feita uma solução 10x em um balão volumétrico contendo: 1 litro de água deslitilada deionizada q.s.p., 5,68g de Na_2HPO_4 (Merck, Alemanha); 2g de KH_2PO_4 (Merck, Alemanha) e 87,68g de NaCl (Merck, Alemanha). Após a adição desses elementos a solução deve ser colocada sob agitação com o auxílio de uma barra magnética até que todos os reagentes estejam dissolvidos. Posteriormente o pH da solução deve

ser corrigido para pH 7,2. Após, a solução é então diluída para a concentração 1x em outras garrafas de vidro previamente separadas misturando-se 100mL da solução 10x à 900mL de água destilada e deionizada. Essas garrafas são então vedadas e a solução é esterilizada por meio da autoclavação das garrafas a 120ºC pelo tempo de 30 minutos. Após a esterilização espera-se um periodo de tempo até que a solução esteja resfriada e então esta é armazenada em camara fria à 4ºC.

5.1.5 Acanthamoeba castellanii (ATCC 300210)

As células da espécie *Acanthamoeba castellanii* utilizadas neste trabalho são provenientes da *American Type Culture Collection* (ATCC 30010) (Maryland, E.U.A.), tendo sido gentilmente cedidas pelo Laboratório de Amebíases do Instituto de Ciências Biológicas (ICB), da UFMG. As amebas foram utilizadas nesse trabalho como plataforma para isolamento e multiplicação de vírus gigantes, análises de microscopia eletrônica de transmissão (MET) e em ensaios de titulação de amostras virais.

5.1.6 Acanthamoeba polyphaga mimivirus (APMV)

Os estoques iniciais (*pool* semente) de APMV foram gentilmente cedidos pelo Dr. Didier Raoult, da Aix Marseille Université, França. Esses estoques foram expandidos para utilização do vírus gigante em experimentos posteriores, de acordo com a necessidade de uso. A forma como o vírus foi multiplicado para utilização como *pool* trabalho será descrita em um tópico posterior dessa sessão.

5.2 Metodologias gerais

5.2.1 Coleta de amostras e isolamento viral

Para explorar a presença de vírus gigantes em ambientes urbanos, amostras de água foram coletadas em pontos equidistantes ao redor de duas lagoas urbanas localizadas no estado de Minas Gerais, Brasil [Figura 10]. Para a coleta

foram utilizados tubos cônicos do tipo Falcon, estéreis (Kasvi, Brasil). Cerca de 80 amostras foram obtidas da Lagoa da Pampulha (cidade de Belo Horizonte; 19° 51' 0.60" S e 43° 58' 18.90" W) e 85 da Lagoa Central (cidade de Lagoa Santa; 19° 38' 38.53" S e 43° 53' 32.00" W). Após a coleta, essas amostras foram armazenadas em geladeira (4ºC) até posterior processamento. O processamento envolveu inicialmente a junção em *pools* de 5-7 amostras coletadas a partir de pontos próximos, sendo esse processo feito no interior de fluxo laminar. Posteriormente, 500 µL de cada *pool* foram adicionados a 4,5 mL de meio água-arroz e incubados em local escuro, à temperatura ambiente, durante um período de 20 dias. A razão dessa metodologia é diminuir a população de micro-organismos fototróficos enquanto permite um maior crescimento de bactérias heterotróficas. Estas servem então como alimento para amebas presentes nessas amostras, que como consequência, aumentam a sua população e funcionam como uma plataforma para multiplicação de possíveis vírus gigantes. Passado o período de 20 dias, cerca de 5.000 amebas da espécie *Acanthamoeba castellanii* (verificadas para ausência de vírus gigantes) foram ainda inoculadas em cada *pool* de amostras, sendo estes posteriormente incubados sob as mesmas condições descritas anteriormente, durante mais 10 dias. Por fim, cada amostra foi filtrada em membranas de 1,2 µm (Millipore, EUA), de maneira a reter impurezas presentes nessas amostras, e posteriormente em membranas de 0,2 µm (de maneira a reter possíveis vírus gigantes presentes). Essas últimas membranas foram então eluídas em 500 µL de solução salina fosfato tamponada (PBS), e após homogeneização em aparatos do tipo *vórtex,* cerca de 100 µL de cada eluato foram inoculados em monocamada de amebas da espécie *Acanthamoeba castellanii*, contidas em placas estéreis de 96 poços (Corning, NY, EUA). O isolamento de vírus gigantes é caracterizado inicialmente pela observação de efeito citopático (arredondamento e posterior lise das amebas), em uma série de até três passagens. A confirmação se dá pela extração de DNA da monocamada de amebas inoculadas com amostras suspeitas, seguida por tentativa de amplificação do gene da helicase viral por meio e experimentos envolvendo reação em cadeia da polimerase (PCR) quantitativo.

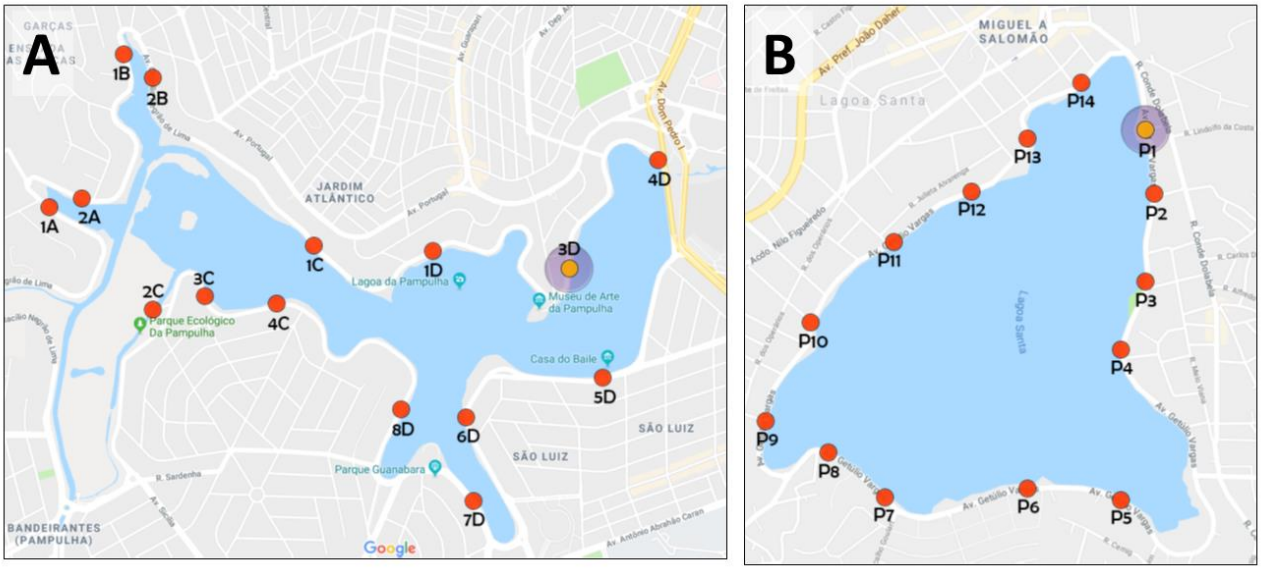


Figura 10 – Mapas representando a localização dos *pools* de amostras de água coletadas ao redor de duas lagoas urbanas em Minas Gerais, Brasil. (A) Lagoa da Pampulha, cidade de Belo Horizonte e (B) Lagoa Central, cidade de Lagoa Santa. Os pontos destacados 3D e P1 são os locais onde foram isolados os vírus desse trabalho.

5.2.2 Extração de DNA pelo método de PCI (fenol-clorofórmio-álcool isoamílico)

Nesse trabalho escolhemos o método de fenol-clorofórmio-álcool isoamílico (proporção 25:24:1, respectivamente) como metodologia padrão para a extração do DNA total de amostras que apresentaram efeito citopático durante tentativas de isolamento de vírus gigantes. Para isso, inicialmente cerca de 100 µl de cada amostra foi mantido a 75ºC durante 60 minutos, de modo a permitir o rompimento do capsídeo de possíveis vírus gigantes presentes nessas amostras, tendo como objetivo a liberação do material genético viral. Em seguida, 400 µl de PBS foram adicionados ao sistema juntamente com mais 500 µl da solução de PCI, de forma a estabelecer uma proporção de 1:1 v/v entre amostra e a solução de extração. Essa mistura foi então homogeneizada em aparatos do tipo *vórtex* e em seguida centrifugada, à temperatura ambiente, a 18000 g (Eppendorf 5430/5430R, rotor F-45-40-11), pelo período de 60 segundos. Posteriormente, a fase superior e mais hialina foi coletada com auxílio de micropipeta e transferida para microtubos estéreis (Eppendorf, USA). De modo a promover a precipitação do material genético, foram adicionados ao volume de sobrenadante coletado cerca de 2,5x

do valor em álcool absoluto (96%) gelado e 0,1x do valor em acetato de sódio 3M, pH 5,5. A amostra foi então homogeneizada por inversão e novamente centrifugada a 18000 g, sendo dessa vez durante o período de 5 minutos e à temperatura de 4ºC. Para finalizar, o sobrenadante gerado foi descartado por inversão e o microtubo deixado aberto em estufa à 37ºC, de maneira que todo o álcool residual evaporasse. O sedimento formado a partir da centrifugação foi ressuspendido em 50 µl de água de injeção e a concentração de material genético foi dosada em espectrofotômetro (Thermo Scientific, Waltham, MA, EUA).

5.2.3 Amplificação do gene da helicase por PCR quantitativo

Para confirmar o isolamento de vírus gigantes através da metodologia descrita nos tópicos anteriores, o DNA extraído de amostras suspeitas foi utilizado como molde visando a amplificação, por meio de PCR quantitativo, do gene da RNA helicase, um gene altamente conservado dentre os membros da família *Mimiviridae*. Para esse processo foi utilizado 1µl de cada amostra (com uma quantidade variando entre 200 – 300 ng de DNA), kit comercial SYBR Green Master Mix (Applied Biosystems, EUA), 4mM de cada iniciador contendo alvo específico para a helicase dos mimivírus (iniciadores 5'-ACCTGATCCACATCCCATAACTAAA-3' e 5'- GGCCTCATCAACAAATGGTTTCT-3) e água de injeção em quantidades adequadas, de maneira a totalizar 10 µl de reação. Todas as reações foram realizadas em placas 48 poços, em duplicata. Os ciclos térmicos utilizados na amplificação das amostras foram estabelecidos de acordo com o padrão do programa da máquina StepOne (Applied Biosystems, EUA), envolvendo uma fase inicial de 95ºC e duração de 10 minutos, seguida por 40 ciclos de 95ºC com duração de 15 segundos, uma nova fase de 60ºC com duração 15 segundos e um último passo de dissociação, com uma temperatura de *melting* específica de 73ºC. A amplificação para o alvo específico foi avaliada de maneira quantitativa, observando-se apenas a presença ou ausência de oligonucleotídeos relacionados ao gene da helicase viral nessas amostras.

5.2.4 Extração de RNA total e transcrição reversa

Alguns ensaios envolveram o processo de extração de RNA total de células de *Acanthamoeba castellanii* infectadas por vírus gigantes. Os ensaios e o modo de infecção serão explicados em um item posterior. Para o processo de extração em si, foi utilizado o kit *Rneasy* (Qiagen, Alemanha) de maneira a estabelecer as seguintes etapas, assim como recomendado pelo fabricante: primeiramente é utilizado um tampão de guanidina com o objetivo de lise celular e inativação de RNases; posteriormente há um processo de desnaturação e precipitação de complexos protéicos através do uso de etanol 70% e passagem das amostras por meio de colunas de afinidade; e finalmente esse processo é seguido por meio de etapas de lavagem com o uso de tampões de lavagem e eluição do RNA extraído com a utilização de água livre de nucleases.

Para o processo de transcrição reversa, o cDNA é produzido tendo como molde 1µg de RNA extraído da cultura celular. As reações posteriores foram então realizadas utilizando a enzima MMLV (Promega, Madison, WI, USA), tampão 5x, dNTPs, oligo Dt e DTT nas concentrações indicadas pelo fabricante, além de água q.s.p utilizada para 20µl de reação. RNA e oligo dT foram então incubados a 70ºC durante um tempo de 5 minutos e posteriormente incubados em gelo pelo mesmo período de tempo. Os outros componentes da reação foram então adicionados com posterior incubação dos tubos a 42ºC, pelo período de 60 minutos e 72ºC pelo tempo de 15 minutos. O cDNA produzido foi então diluído em 40µl de água e armazenado a -20ºC.

5.2.5 Titulação viral

Em ensaios que dependem da estimativa do título viral, este foi obtido através do método de *end-point*, descrito por Reed-Muench em 1938 (Reed e Muench, 1938). Por meio dessa técnica, inicialmente são utilizadas placas de 96 poços (Corning, NY, EUA) contendo 100 µL de meio PYG e aproximadamente 40.000 amebas/poço. Após se atingir uma confluência celular de cerca de 80%, as amostras virais a serem tituladas são diluídas em PBS de maneira seriada, em uma razão de 10 (10^{-1} a 10^{-11}), e em seguida, um total de 100 µL de cada

diluição é adicionado por poço, em quadruplicata. Uma quadruplicata de poços é reservada como controle de viabilidade de amebas, sendo neste adicionados 100 µL de PBS. As placas são vedadas completamente e incubadas a 32 ºC. Cada poço passa então a ser monitorado diariamente para observação de efeito citopático, e após 4 dias de incubação, o cálculo do título viral é realizado com o valor expresso em TCID50/mL.

5.2.6 Multiplicação de vírus gigantes

Para a produção dos vírus utilizados nesse trabalho, as amostras contendo os estoques iniciais de APMV e demais vírus isolados foram utilizadas de acordo com o protocolo proposto por Abrahão e colaboradores (Abrahao, Boratto*, et al.*, 2014). Resumidamente, neste protocolo garrafas T150 (TPP, Suiça) contendo monocamadas de amebas da espécie *A. castellanii*, em uma confluência de cerca de 90%, são inoculadas com os vírus gigantes utilizando-se uma multiplicidade de infecção (MOI) de 0.01. Para que o inoculo seja feito, o meio das garrafas é inicialmente descartado e a suspensão viral, contida PBS, é cuidadosamente adicionada sobre a monocamada. Após uma hora de adsorção, 25 mL de meio PYG suplementado com 7% de SFB são adicionados à monocamada celular. As garrafas são então mantidas a 32ºC, completamente vedadas. Após três dias de incubação, o efeito citopático é avaliado e todo o conteúdo das garrafas é coletado para purificação viral.

5.2.7 Purificação viral

A purificação de vírus gigantes foi baseada no protocolo utilizado por Raoult e colaboradores, apresentando algumas adaptações (Raoult *et al.*, 2004). Primeiramente, foi realizada a multiplicação viral em culturas de *A. castellanii*, conforme descrito no item anterior. Após a observação do efeito citopático e coleta de todo o conteúdo das garrafas, o liquido foi transferido para tubos do tipo Falcon (Kasvi, EUA) de 50 mL e mantidos em banho de gelo. O material coletado foi então submetido a três ciclos de congelamento e descongelamento, com o objetivo de liberar as partículas virais eventualmente aprisionadas em

células que ainda estejam integras, não lisadas. Com o mesmo intuito, após esse processo, o liquido coletado foi submetido a ciclos de lise em homogeneizador do tipo “Douncer” (Wheaton, EUA) por 80 vezes. Em seguida o material sofreu uma filtragem em membranas de 1,2 µm (Millipore, EUA) para a retenção de debris celulares. O conteúdo filtrado foi então vagarosamente gotejado sobre 10 mL de uma solução de sacarose a 24% (Merck, Alemanha), em tubos próprios para Ultracentrífuga *Combi Sorvall*. A amostra foi submetida à ultracentrifugação a 35.000 *g* por 30 minutos, entre 4º C e 8º C, para a sedimentação das partículas virais. Ao final, o sobrenadante foi descartado e o precipitado contendo as partículas virais foi ressuspendido em 500 µL de PBS. Alíquotas da suspensão viral foram feitas, devidamente identificadas e estocadas à -80ºC.

5.2.8 Microscopia eletrônica de transmissão e morfometria

De maneira a observar morfologicamente os vírus gigantes isolados e caracterizar de maneira visual o seu processo de multiplicação, foram também feitos nesse trabalho experimentos envolvendo a microscopia eletrônica de transmissão (MET) de células de *Acanthamoeba castellanii* infectadas com os novos isolados. Para isso, inicialmente essas células foram cultivadas em garrafas T150 (TPP, Suiça) até atingirem cerca de 80-90% de confluência, sendo posteriormente infectadas com os vírus em uma MOI de 0,01. Após 12 horas de infecção, quando cerca de 50-70% das células apresentavam efeito citopático, o meio foi descartado e a monocamada de amebas foi gentilmente lavada com PBS. A monocamada foi fixada pela adição de glutaraldeído 2,5% v/v durante 60 minutos, à temperatura ambiente. As células foram posteriormente coletadas, centrifugadas a 1500g (Eppendorf 5430/5430R, rotor F-45-30-11) por 10 minutos e a 4ºC, tendo em seguida o sobrenadante sido descartado. Ao sedimento foi adicionado 1mL de tampão fosfato (0,1M pH 7,4), sendo este enviado imediatamente ao Centro de Microscopia da UFMG para processamento e cortes em secções ultrafinas. As imagens foram observadas em microscópio eletrônico de transmissão TECNAI G2-20; SuperTwin FEI; 120 kV e utilizadas para análises morfométricas das partículas virais utilizando-se do *software*

ImageJ v1.51t. O protocolo para preparo das soluções utilizadas nesse experimento pode ser encontrado na sessão Anexo I.

5.2.9 Curva de ciclo único

Curvas de ciclo único foram realizadas nesse trabalho com o objetivo de avaliar o perfil de multiplicação dos vírus gigantes aqui isolados. Para isso, inicialmente cerca de 40.000 amebas (total de 100 µl) foram semeadas em cada poço de uma placa estéril contendo 96 poços (Corning, NY, EUA). As amebas foram então infectadas com os diferentes vírus gigantes (40 µl) e incubadas a 28ºC, utilizando-se de uma MOI de 10. Diferentes tempos de coleta (0h, 1h, 2h, 4h, 8h e 24h) foram utilizados de maneira a cobrir os passos mais importantes do ciclo de multiplicação desses vírus. Após um período de 1h de adsorção, o sobrenadante foi gentilmente retirado de cada poço e descartado com auxílio de micropipetas. Cada poço foi então novamente preenchido com 200 µl de meio PYG e a cada tempo de coleta a monocamada de células era raspada, coletada e armazenada a -20ºC. As células coletadas foram submetidas a três ciclos de congelamento/descongelamento, tendo em seguida os vírus gigantes titulados pelo método de $TCID_{50}$/mL, assim como descrito no item 5.2.5.

5.2.10 Sequenciamento e anotação do genoma completo

De forma a iniciar o processo de caracterização genética dos vírus gigantes isolados nesse trabalho, o genoma completo desses organismos foi sequenciado, tendo seus genes sido preditos e funcionalmente anotados. Inicialmente, esses genomas foram sequenciados utilizando-se da plataforma Illumina MiSeq (Illumina Inc., San Diego, CA, EUA) através da execução de *paired-ends*. As sequencias foram montadas *de novo* através da utilização do programa ABySS (Simpson *et al.*, 2009) e os *contigs* resultantes foram ordenados pelo *software* CONTIGuator.py (Galardini *et al.*, 2011). Concomitantemente, o *software* CLC_Bio foi utilizado paralelamente buscando fazer a montagem desses genomas, tendo como base o depósito para o genoma completo do APMV (NC_014649.1), o vírus protótipo da família *Mimiviridae*. Os

genomas completos construídos em ambas estratégias foram utilizados de maneira a verificar a melhor estratégia de montagem de sequências e também para fechar *gaps* presentes durante o processo.

A predição gênica de cada um dos genomas montados foi realizada por meio da utilização das ferramentas fornecidas pelo programa RAST e pelo programa GeneMarkS. Sequencias de RNA de transferência foram identificadas por meio da ferramenta tRNAscan-SE. As anotações funcionais foram inferidas por meio da utilização da ferramenta BLASTp (*e-value* < $1x10^{-3}$), por buscas no grupo de proteínas ortólogas dos NCLDVs (NCVOGs) (Yutin *et al.*, 2013) e também por meio de buscas em bancos de dados especializados utilizando-se do *software* BLAST2GO. A anotação do genoma foi manualmente revisada e curada. As ORFs que apresentavam menos que 100 aminoácidos e não apresentavam correspondentes nos bancos de dados foram retiradas das análises. Aquelas que apresentavam mais de 100 aminoácidos mas não apresentavam correspondentes nos bancos de dados foram consideradas ORFans.

5.2.11 Análises filogenéticas para a DNA polimerase B viral

Os membros da família *Mimiviridae* podem ser divididos atualmente em três grupos (linhagens A, B e C), de acordo com análises filogenéticas feitas para a DNA polimerase B viral, sendo este um gene bastante conservado dentro dessa família de vírus. Nesse trabalho utilizamos esse gene como alvo para a caracterização das linhagens dos novos vírus gigantes aqui isolados. Para isso, sequencias nucleotídicas desse gene, pertencentes a membros das três diferentes linhagens da família *Mimiviridae*, foram coletadas a partir do banco de dados do “National Center for Biotechnology Information” e alinhadas utilizando-se do programa ClustalW. Sequencias pertencentes a vírus gigantes relacionados, mas não pertencentes à família *Mimiviridae*, também foram utilizadas de maneira a incluí-los como grupos externos em nossas análises. Após o alinhamento dessas sequências, a construção de árvores filogenéticas foi gerada utilizando-se do programa MEGA 7.0, pelo método de máxima verossimilhança, e com uma análise de bootstrap tendo como base o valor de 1000 replicatas.

5.3 Metodologias específicas

5.3.1 Padrão de expressão de aaRs pelo vírus NYMV

De forma a avaliar a expressão de aaRs em um dos nossos isolados (Niemeyer virus), quatro desses genes (metionil, tirosil, cisteinil e arginil- tRNA sintetases) foram escolhidos e analisados assim como descrito por Silva e colaboradores, sendo comparados com o perfil de expressão dos mesmos genes presentes no vírus APMV, o protótipo da família *Mimiviridae* (Silva et al., 2015). Para isso, cerca de $1x10^5$ células de *Acanthamoeba castellanii* foram inoculadas em placas de 24 poços (TPP, Suiça), contidas em meio PAS e infectadas com NYMV ou APMV durante um período de 8h, à 32^o C, utilizando-se de uma MOI de 10. O tempo de 8h foi escolhido pelo fato de nesse período a expressão desses genes ser encontrada em nível máximo nos vírus gigantes (Silva *et al.*, 2015). Após o tempo de infecção, as células foram então coletadas com auxílio de uma micropipeta e posteriormente centrifugadas a 1500g (Eppendorf 5430/5430R, rotor F-45-30-11) por 10 minutos. O sobrenadante foi então descartado e o precipitado (*pellet*) foi utilizado para ensaios de extração de RNA e transcrição reversa (descritos no item 5.2.4) e também análise da expressão gênica por PCR quantitativo (assim como descrito no item 5.2.3). Para os ensaios de PCR quantitativo foram utilizados iniciadores desenhados em outros trabalhos, compilados na Tabela 2 a seguir (Silva *et al.*, 2015). Além disso, foram geradas curvas padrão para cada resultado, e estas normalizadas através dos níveis de expressão do rRNA 18S amebiano.

Tabela 2 – sequência de iniciadores usados para ensaio de expressão de aaRs.

Iniciador	**Sequência *Forward***	**Sequência *Reverse***
Metionil RS	TGATTGGCGTGAATGGCTGA	ACCAATCACACTAGCCGGAA
Arginil RS	GTGGGTGATTGGGGAACTCA	TGATACGGTCTCCAATCGGG
Tirosil RS	TTTGGCAAACCAATCGGCAA	TGGTTTTGAACCTAGTGGTCGT
Cisteinil RS	TGCCAACCAGGTACACCAAA	TGCTCTTTGGAAAGGTCGATCA
18S rRNA	TCCAATTTTCTGCCACCGAA	ATCATTACCCTAGTCCTCGCGC

5.3.2 Análises filogenéticas para aaRs presentes nos novos isolados

De forma a se obter um melhor conhecimento das origens das aaRs presentes no genoma de nossos isolados, análises filogenéticas foram realizadas após o sequenciamento e a anotação do genoma completo desses vírus. Foram analisados os genes que codificam para tirosil, cisteinil, metionil e arginil-tRNA-sintetases. O perfil filogenético foi traçado em comparação com diversos membros da família *Mimiviridae,* pertencendo às três diferentes linhagens. A metodologia implementada é a mesma como explicada no item 5.2.11 deste trabalho.

5.3.3 Caracterização do gene codificante para a MCP

Neste trabalho também realizamos um processo de caracterização do gene codificante para a proteína principal do capsídeo em membros das diferentes linhagens da família *Mimiviridae*, inclusive em nossos isolados. Esse processo de caracterização envolveu investigar como esse gene estaria organizado, tanto em sua forma estrutural (tipos de regiões intrônicas e exônicas presentes) como também na sua organização sintênica (como esses diferentes introns e éxons estariam distribuídos espacialmente no gene desses vírus). Para isso, inicialmente utilizamos como referência para nossas análises a sequência do gene da MCP que está presente no vírus protótipo da família *Mimiviridae* (APMV), visto ser esse o vírus com o genoma melhor caracterizado dentre os mimivírus atualmente. A partir de análises feitas utilizando a ferramenta BLASTn para cada região da MCP descrita no APMV (éxons 1, 2 e 3; introns 1 e 2), buscas por regiões homólogas foram realizadas no genoma completo de cada um dos membros restantes analisados nesse trabalho. As regiões que se encaixavam dentro do nosso perfil de homologia (*evalue* < 10^{-3}) foram então separadas, e o esboço final do gene desses vírus foi construído baseado na posição e orientação ocupada pelas diferentes sequencias. Componentes genéticos (introns ou éxons) encontrados de maneira específica dentro de algum membro ou linhagem, foram também procurados nas outras amostras virais.

5.3.4 Sequenciamento do RNAm do gene para MCP em KV

De forma a melhor entender a maneira como o gene do capsídeo é processado durante a etapa de transcrição, também realizamos nesse trabalho o sequenciamento do RNAm desse gene durante o processo de expressão do transcrito em um dos nossos isolados (Kroon virus), comparando com o vírus protótipo da família, APMV. Para isso, cerca de $1x10^5$ células de *Acanthamoeba castellanii* foram incubadas em placas de 24 poços (TPP, Suiça) e infectadas com KV ou APMV em uma MOI de 5, sendo mantidas à 32º C. Após os tempos de 6h, as células foram coletadas e centrifugadas a 1500g (Eppendorf 5430/5430R, rotor F-45-30-11) por 10 minutos. O precipitado foi utilizado para os ensaios de extração de RNA total e de transcrição reversa (assim como descrito no item 5.2.4) sendo posteriormente aproveitado para amplificação do gene da MCP, utilizando a técnica de PCR convencional. Os amplicons foram examinados por meio de eletroforese em um gel de agarose 1% com tampão Tris-borato-EDTA (TBE), a 150V. A banda do amplicon foi então purificada, sequenciada nas duas direções e em triplicata (3730 DNA analyzer; Thermo Fischer Scientific, Waltham, MA). Por fim, as sequências foram alinhadas e analisadas utilizando o programa MEGA 7.

5.3.5 Análises filogenéticas do gene para a MCP em *Mimiviridae*

O caráter conservado apresentado pelo gene do capsídeo em diferentes membros do grupo NCLDV, incluindo os participantes da família *Mimiviridae,* acabou por nos estimular a investigar a utilização deste gene como um provável marcador molecular que auxiliasse na classificação de novos mimivírus isolados ao longo do tempo. Para isso, análises filogenéticas foram realizadas incluindo diversos membros da família *Mimiviridae,* correspondentes às três linhagens existentes atualmente. A metodologia implementada é a mesma como explicado no item 5.2.11 deste trabalho. Uma tabela contendo o número de acesso das sequências utilizadas nessa análise foi anexada ao fim do trabalho.

6. RESULTADOS

6.1 Isolamento de vírus gigantes

Através da utilização do nosso protocolo de isolamento, obtivemos um resultado positivo para a presença de vírus gigantes em dois *pools* de amostra de água (pontos 3D e P1), cada um correspondendo a um dos locais onde foram realizadas as coletas [Figura 10 A e B]. Inicialmente, as amostras foram caracterizadas por meio da observação de efeito citopático em células de *Acanthamoeba castellanii*, que consistiu no arredondamento e lise dessas amebas após quatro dias de infecção. Posteriormente, a confirmação desses novos isolados como membros da família *Mimiviridae* foi estabelecida por meio da amplificação de um gene altamente conservado dentro da família, a RNA helicase. Os novos isolados foram então denominados Niemeyer vírus (NYMV), para a amostra obtida a partir de água da Lagoa da Pampulha, e Kroon virus (KV), para a amostra obtida a partir de água da Lagoa Central.

6.2 Microscopia eletrônica de transmissão e análises morfométricas

Através da observação das imagens de MET reforçamos os nossos dados de isolamento de membros da família *Mimiviridae*. As imagens nos demonstraram a presença de diversas partículas virais com características tipicamente descritas para outros mimivírus, como por exemplo: a visualização de partículas com grandes dimensões, a presença de uma simetria pseudo-icosaédrica albergando a região em formato de estrela conhecida como *stargate*, a formação do capsídeo por uma variação de 3-4 camadas proteicas e a presença de estruturas denominadas de fibrilas em torno de toda essa região [Figura 11]. Por meio da análise dessas imagens pudemos também traçar o perfil morfométrico desses isolados. As partículas de NYMV apresentam uma dimensão em torno dos 696nm, tendo o capsídeo em torno de 463nm e a porção das fibrilas auxiliando com 153nm no diâmetro total da partícula [Figura 11]. A análise das imagens das partículas do KV nos demonstrou um vírus que

apresenta em torno dos 673nm de diâmetro, sendo que a porção do capsídeo contribui em média com 415nm e as fibrilas em torno dos 129nm [Figura 11].

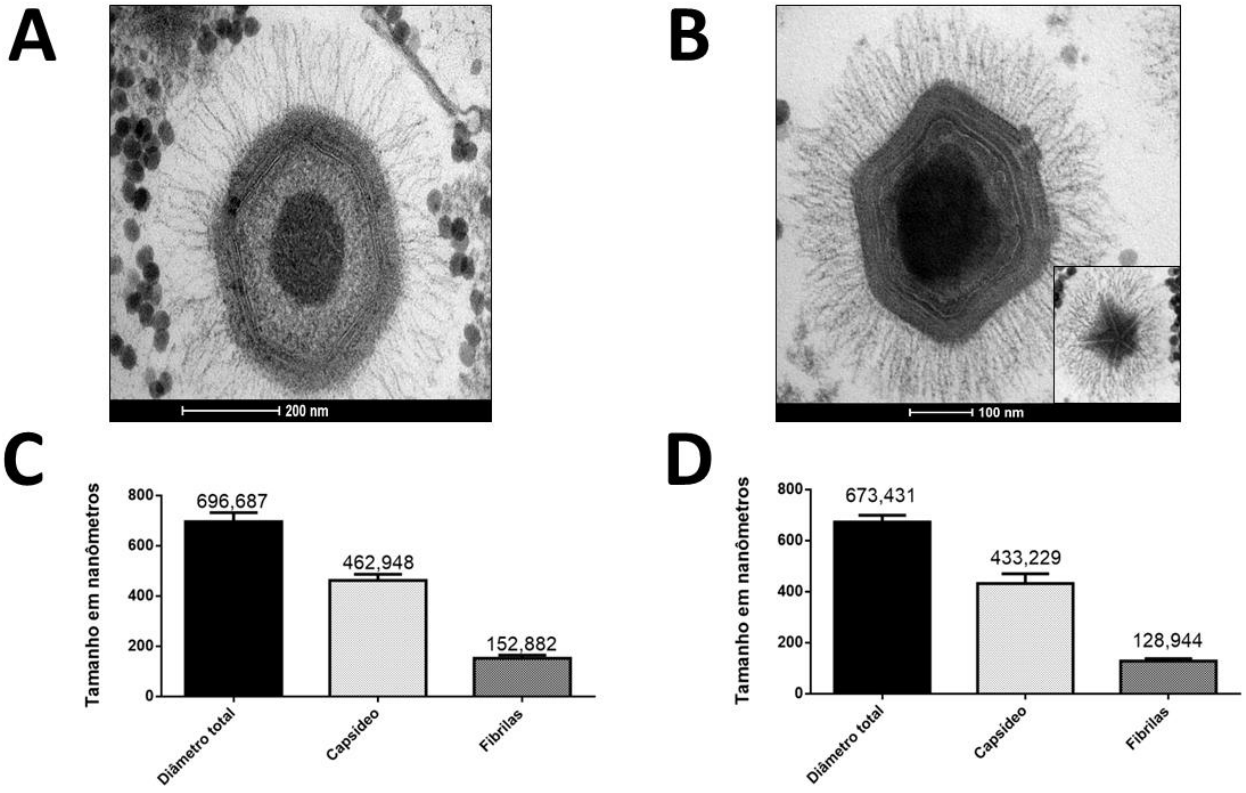


Figura 11 – Microscopia eletrônica de transmissão e análise morfométrica dos vírus isolados neste trabalho. (A) Partícula de NYMV. (B) Partícula de KV com destaque para a região do *stargate* na porção inferior direita da imagem. (C) Análise morfométrica das partículas de NYMV e (D) KV. Para as análises morfométricas foram analisadas cerca de 10 partículas de cada vírus, obtendo-se uma média do tamanho de suas estruturas: tamanho total da partícula, tamanho do capsídeo viral e tamanho das fibrilas.

Os ensaios de MET também nos permitiram observar diversas etapas do ciclo de multiplicação tipicamente descritas para mimivírus. Assim como já observado em outros isolados, o início do ciclo de NYMV e KV se inicia com a penetração das partículas virais por meio de um processo conhecido como fagocitose. Nesse processo os vírus gigantes acabam sendo cercados por estruturas amebianas chamadas de pseudópodes e incorporados no citoplasma da célula hospedeira através da formação de vesículas denominadas de fagossomos [Figura 12 A-B]. Pudemos também observar a etapa de liberação do genoma viral no citoplasma da célula por meio da abertura *stargate* [Figura

12 C]. Posteriormente, assim como caracterizado, também observamos a presença de fábricas virais maduras, levando à formação de diversas partículas que se propagam por todo o citoplasma celular [Figura 12 D-E]. Por fim, a liberação destas é dada por meio de lise celular [Figura 12 F].

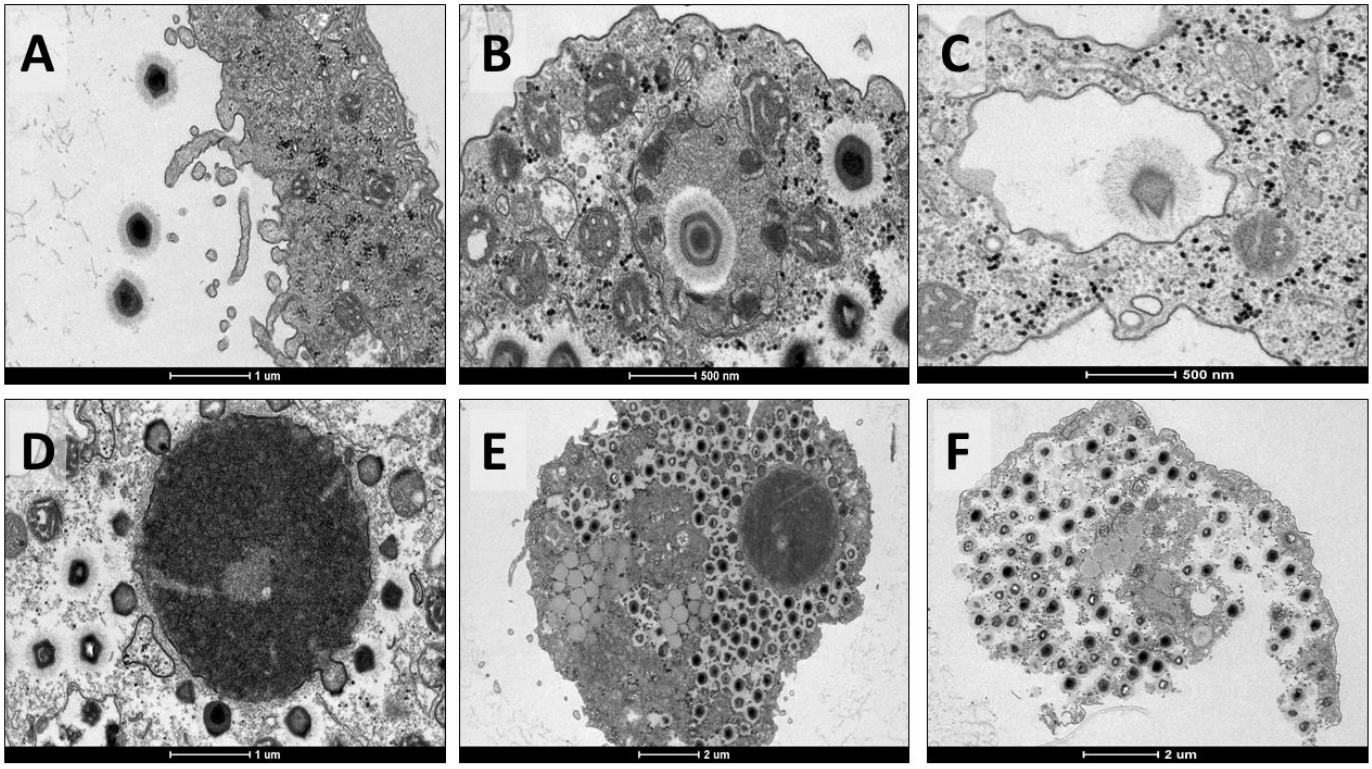


Figura 12 – Esquema do ciclo de multiplicação de NYMV (B-C) e KV (A, D-F) observados por microscopia eletrônica de transmissão. (A) Partículas são primeiramente englobadas por formações chamadas de pseudópodes e (B) incorporadas no citoplasma no interior de fagossomos. (C) As partículas virais liberam então o genoma de DNA dupla-fita através do *stargate*. (D) A formação de uma fábrica viral leva à formação de partículas virais que (E) se propagam por todo o citoplasma da célula. (F) Essas partículas são então finalmente liberadas por meio da lise da célula amebiana.

6.3 Curva de ciclo único

Por meio dos ensaios de curva de ciclo único buscamos traçar um perfil de multiplicação para cada um dos nossos isolados, o que nos auxiliou em experimentos estabelecidos posteriormente. Ambos os vírus iniciam uma produção maciça de partículas virais entre os tempos de 4-6h de infecção, atingindo o pico de produção por volta das 8h pós infecção [Figura 13]. Desse

período de tempo em diante, NYMV estabelece uma ligeira maior produção de partículas do que a observada para o vírus APMV. Enquanto isso KV determina no mesmo período de tempo em uma intensa produção viral ao final de seu ciclo, chegando a uma diferença de quase 3 logs com relação ao protótipo APMV. [Figura 13 B].

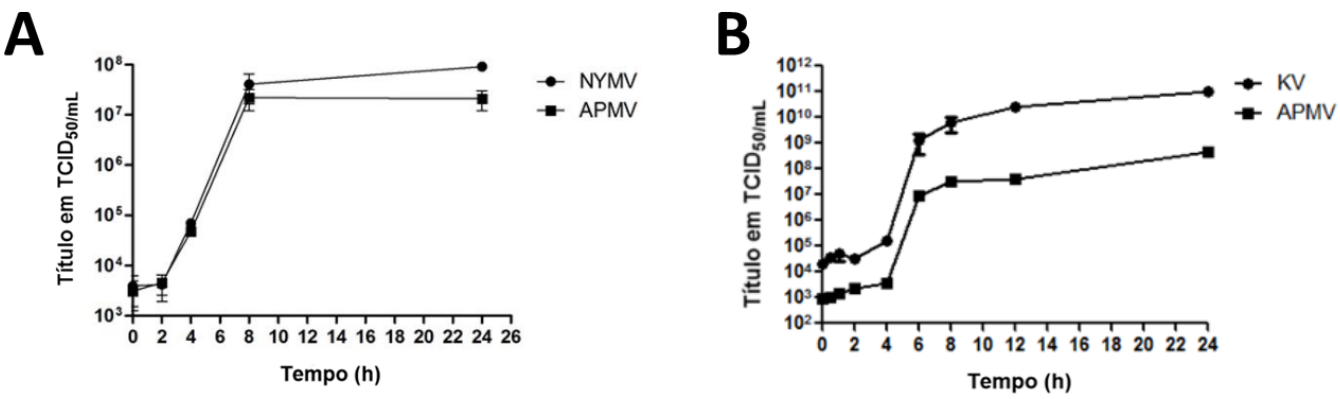


Figura 13 – Curva de ciclo único mostrando o perfil de multiplicação dos vírus gigantes isolados nesse trabalho, (A) NYMV e (B) KV, quando comparados com o vírus protótipo da família *Mimiviridae*, APMV. Ambos apresentam semelhanças no perfil de multiplicação, com a produção máxima de partículas virais sendo atingida em torno do tempo de 8h pós infecção.

6.4 Sequenciamento e análise das aminoacil-tRNA-sintetases

Por meio do sequenciamento e montagem do genoma completo de nossos dois isolados pudemos observar, primeiramente com relação ao genoma de NYMV, que este é caracterizado por uma molécula de DNA de dupla fita composta por aproximadamente 1,3 Mb, apresentando em torno de 27,96% de conteúdo de G-C (bastante similar com outros mimivírus) e codificando para um total de 1003 proteínas. Já para o vírus KV, a análise e sequenciamento do seu genoma havia sido feita em um outro estudo, no qual foram observadas características bastante semelhantes, com o vírus apresentando um genoma de DNA dupla fita de 1,2 Mb, um conteúdo de G-C de 27,5% e a codificação de cerca de 944 proteínas (Assis *et al.*, 2015). Ambos vírus se agruparam como membros da linhagem A da família *Mimiviridae*, de acordo com análises

filogenéticas para o gene da DNA polimerase B, o marcador molecular mais utilizado atualmente nesse tipo de análise para esses vírus gigantes.

Pela anotação funcional do genoma desses dois isolados pudemos observar algumas características interessantes relacionadas à distribuição de aaRS nesses vírus. Enquanto a grande maioria dos mimivírus pertencentes à linhagem A se destaca pela presença de uma única cópia das seguintes aaRS (tirosil, cisteinil, metionil e arginil-tRNA-sintetases), NYMV é marcado por uma distribuição semelhante, mas com a presença de duas cópias para as enzimas tirosil, cisteinil e metionil-tRNA-sintetase [Tabela 3]. Já para o vírus KV, análises do genoma demonstraram a presença de uma distribuição e número de cópias semelhantes às apresentadas para outros membros da linhagem A, da mesma forma como observadas para APMV, Mamavirus, Samba virus (SMBV) e Hirudovirus [Tabela 3].

Tabela 3 – Distribuição e número de cópias de aaRs dentre membros da família *Mimiviridae*

AARS	NYMV	KV	APMV	MAMAVIRUS	SMBV	HIRUDOVIRUS	MHCV	APMOUV
TIROSIL	**2**	1	1	1	1	1	1	1
CISTEINIL	**2**	1	1	1	1	1	1	1
METIONIL	**2**	1	1	1	1	1	1	1
ARGINIL	1	1	1	1	1	1	1	**4**
ISOLEUCIL	-	-	-	-	-	-	1	1
ASPARAGINIL	-	-	-	-	-	-	1	-
TRIPTOFANIL	-	-	-	-	-	-	1	-

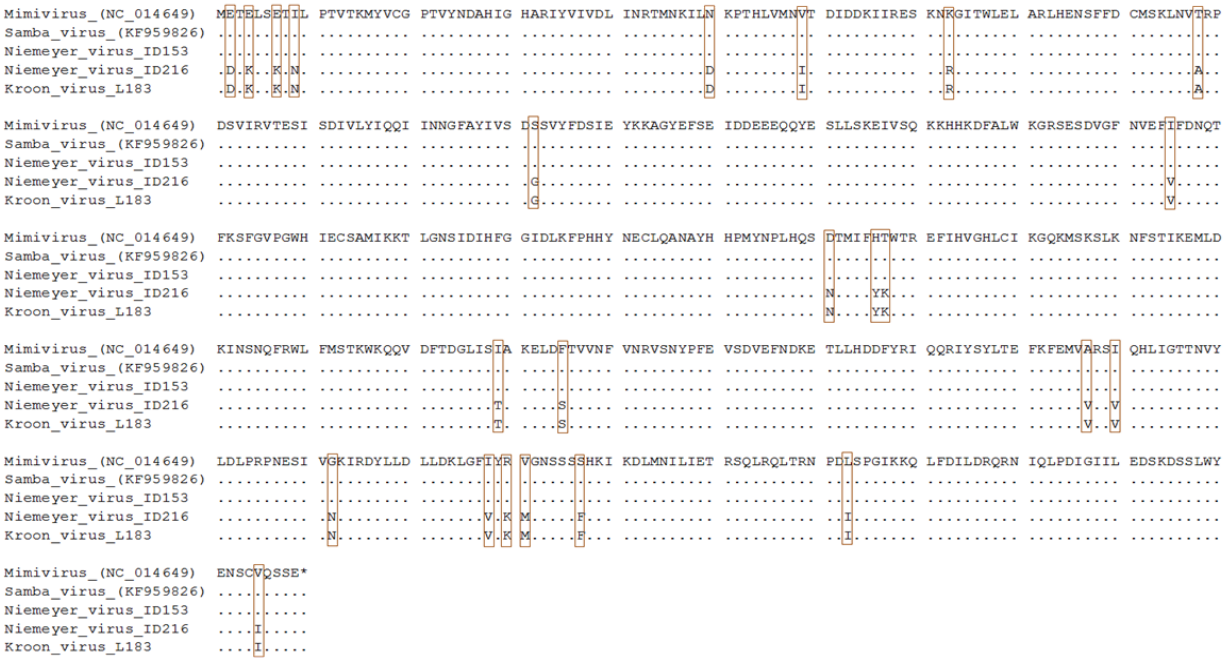
NYMV, Niemeyer virus; KV, Kroon virus; APMV, *Acanthamoeba polyphaga mimivirus;* SMBV, Samba virus; MHCV, *Megavirus chilensis*; APMOUV, *Acanthamoeba polyphaga moumouvirus*

Esse fenômeno de duplicação nos levou a tentar promover um estudo mais refinado sobre esse processo em NYMV. Uma das análises que fizemos foi a avaliação das sequências de aminoácidos das aaRS duplicadas em nosso isolado, e a comparação com as sequências da enzima presentes em outros mimivírus. Através disso pudemos determinar que em NYMV as cópias de aaRS duplicadas apresentam polimorfismos quando comparadas entre si e que além

disso, nessas mesmas enzimas, enquanto uma das cópias apresenta uma sequência que dispõe de 100% de identidade com a de outros vírus da linhagem A, como por exemplo com APMV e SMBV, os polimorfismos encontrados na segunda cópia sempre fazem com que essa se encontre em um nível de identidade semelhante com a sua correspondente em KV [Figura 14].

A

B



C

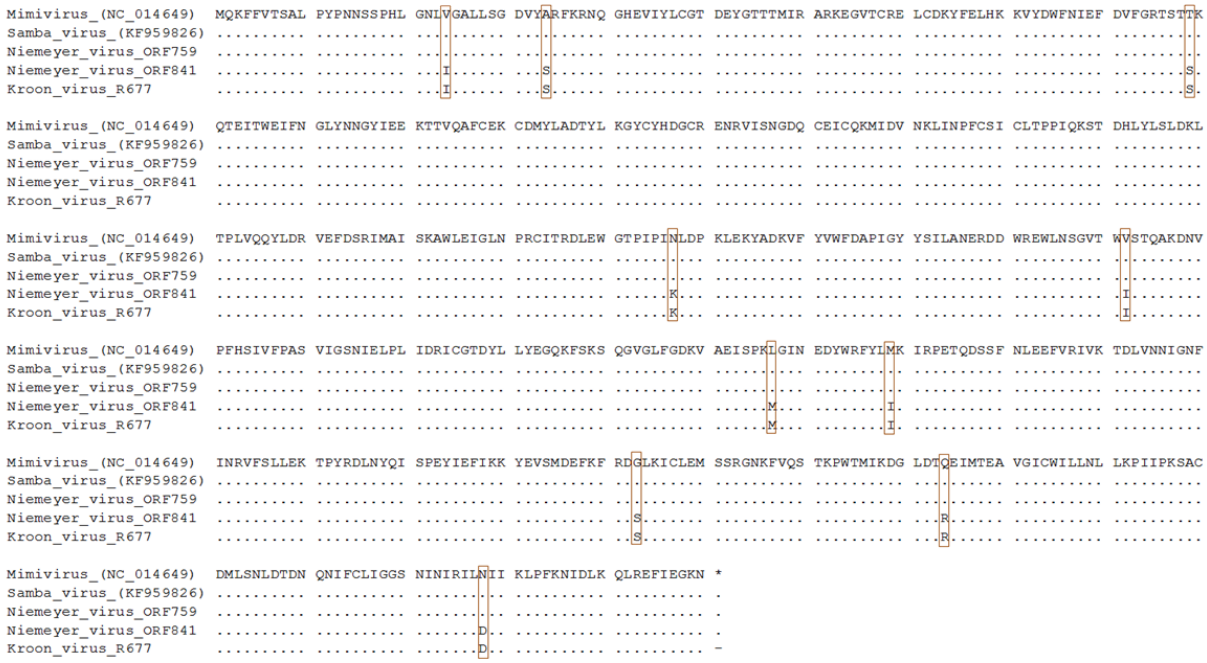


Figura 14 – Sequência de aminoácidos de aaRS presentes em diferentes mimivírus da linhagem A. Foram analisadas aqui sequências de enzimas presentes no genoma dos vírus APMV, SMBV, KV e os ortólogos que se apresentam duplicados no genoma de NYMV. (A) Tirosil-tRNA-sintetase, (B) cisteinil-tRNA-sintetase e (C) metionil-tRNA-sintetase. Das cópias duplicadas, uma sempre se assemelha ao vírus APMV e SMBV, enquanto a segunda cópia sempre tem relação com o gene ortólogos presente em KV. Caixas marrons indicam polimorfismos nas sequências.

Em um outro tipo de análise, buscamos fazer uma observação dos genes que flanqueiam essas aaRs no genoma de NYMV, e no genoma dos outros mimivírus acima analisados, de modo a fazer uma caracterização da sintenia de genes presentes naquela região e também do posicionamento ocupado pelas cópias que estão duplicadas dentro do próprio genoma do NYMV. Por meio dessa analise de vizinhança, pudemos observar que as duplicações apresentadas pelas três enzimas no genoma de NYMV não estão distribuídas *in tandem*, ou seja, as cópias de cada enzima duplicada no genoma desse vírus se localizam em regiões distantes uma das outras, visto que a vizinhança de genes que as flanqueiam são bastante diferentes [Figura 15]. Outra noção que pudemos ter com essas análises foi a de que o gene que codifica para a metionil tRNA sintetase é aquele que se organiza de maneira mais sintênica quando comparado com as outras aaRs dos vírus gigantes analisados. Isso porque, com

relação a esse gene, foi observado que todas as amostras de mimivírus analisadas compartilham pelo menos duas regiões gênicas comuns entre si, tanto à montante quanto à jusante (exceto SMBV na posição do segundo gene na extremidade 3').

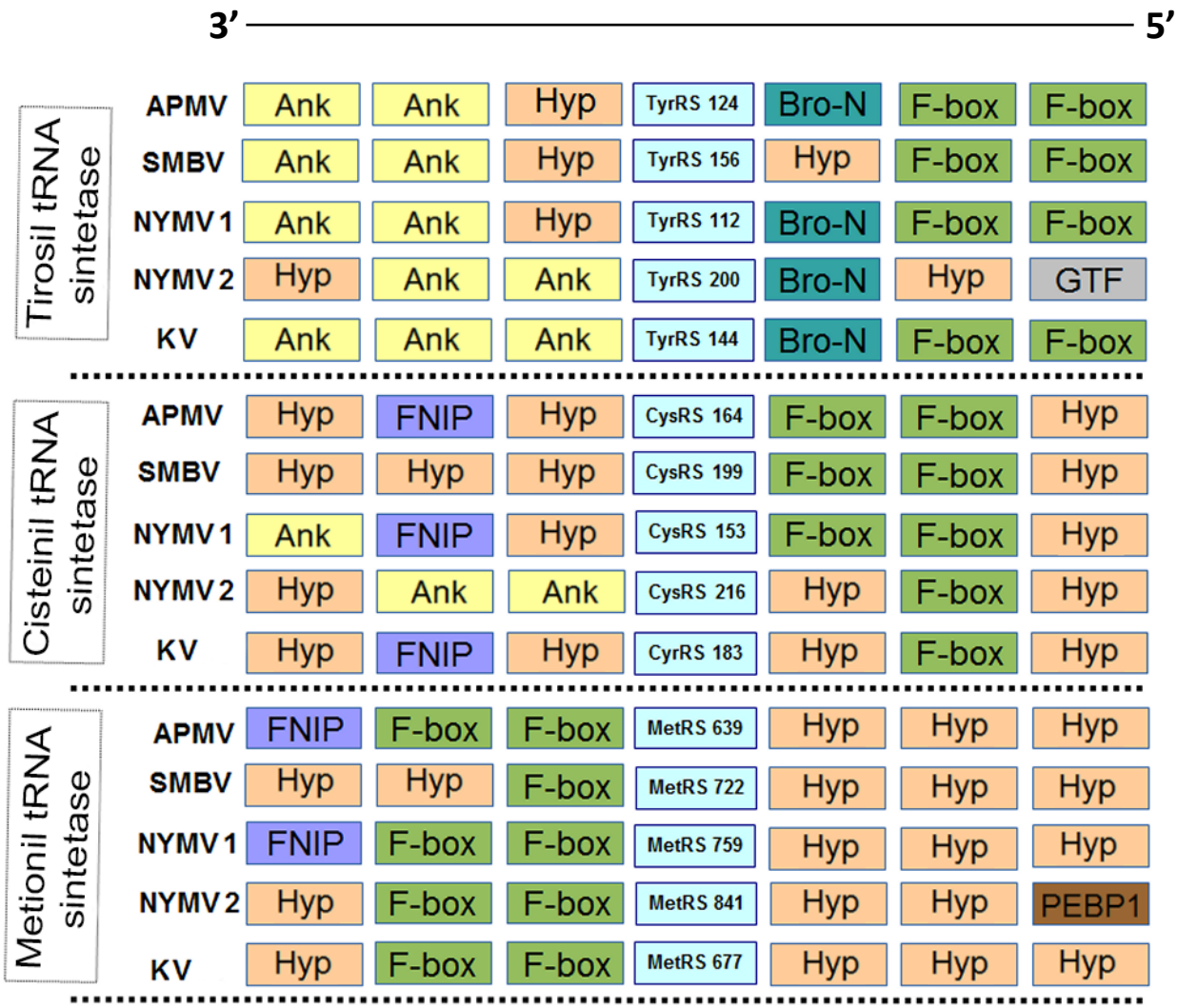
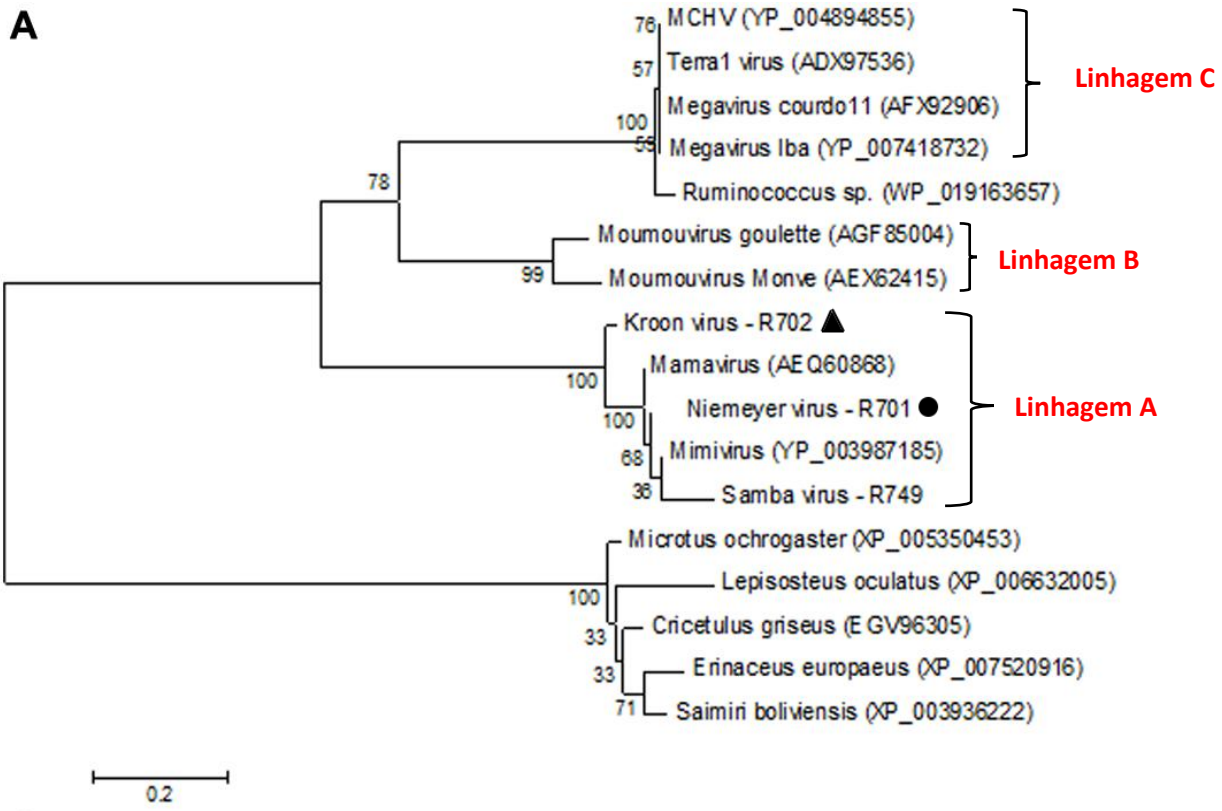
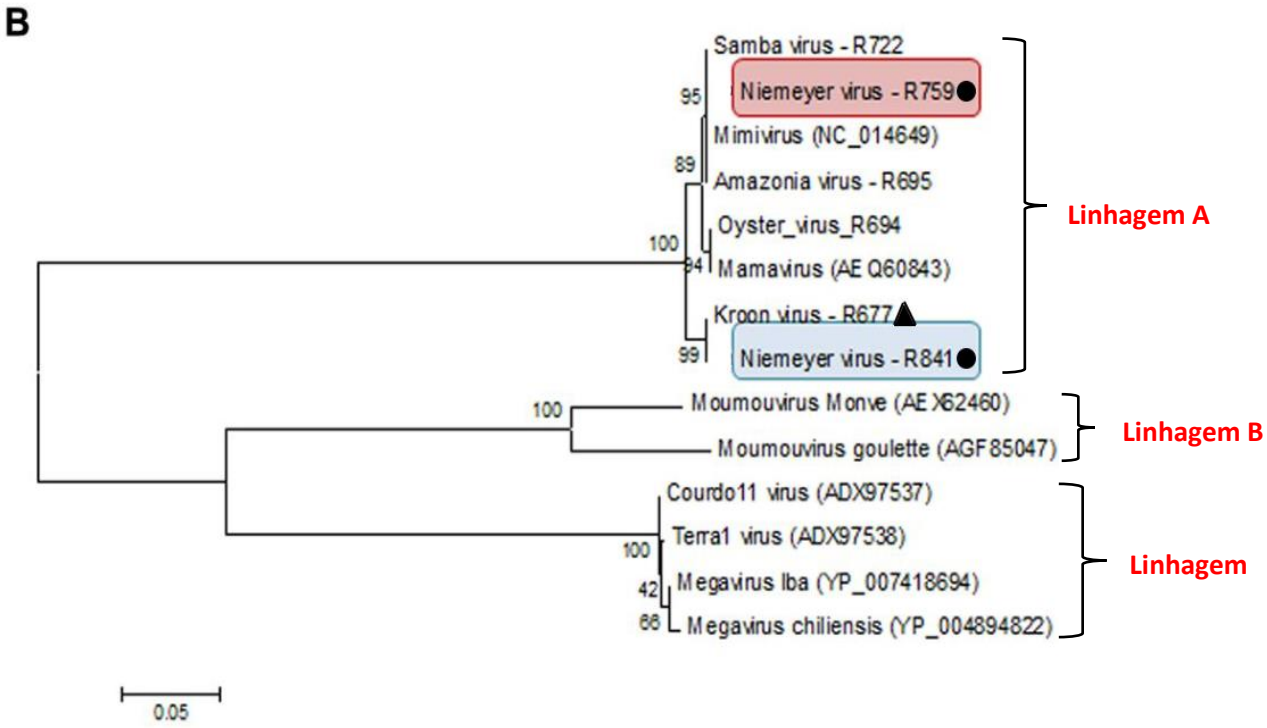
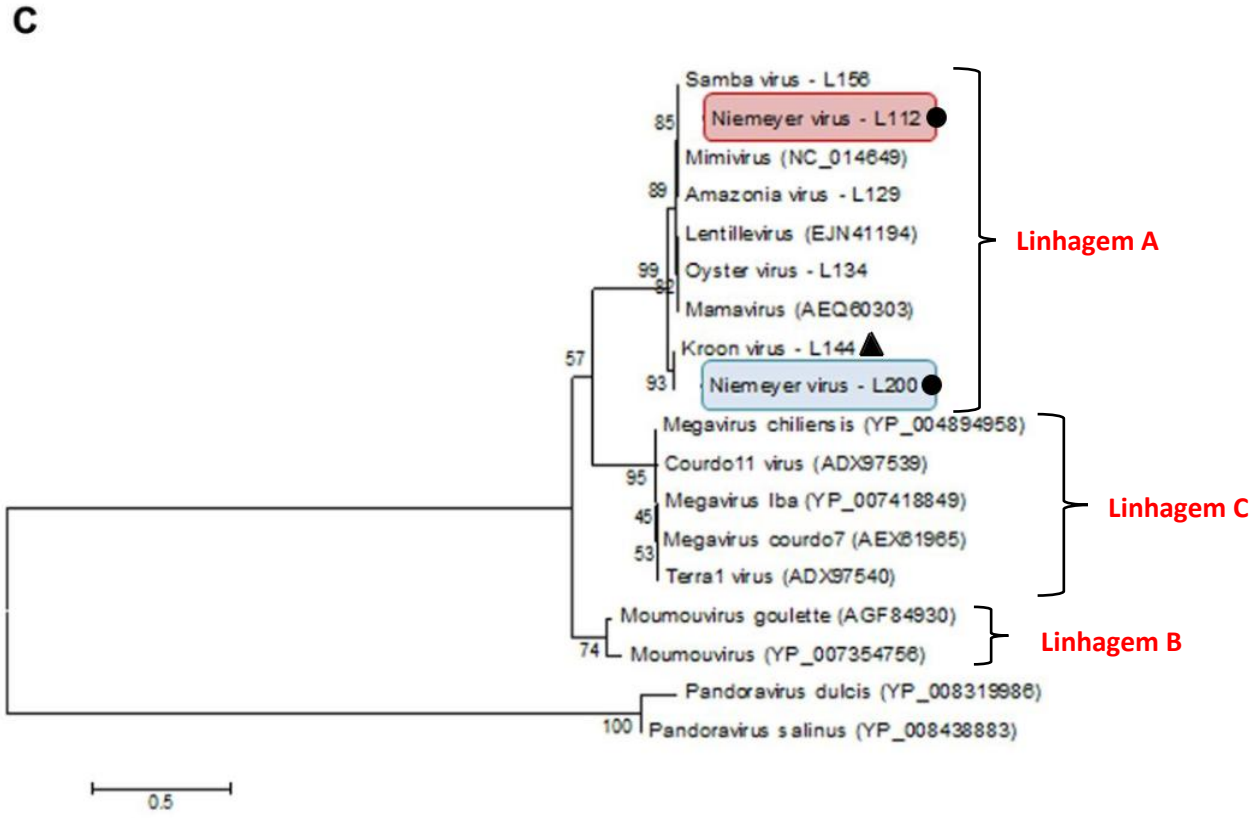


Figura 15 – Análise de vizinhança gênica de aaRs presentes em alguns vírus gigantes. Três genes localizados à montante e à jusante de cada uma das aaRs duplicadas no genoma de NYMV foram analisados, assim como seus correspondentes no genoma de APMV, SMBV e KV. *Ank* – anquirina; *Hyp* – proteína hipotética; *F-box* – proteína de domínio F-box; *Bro-n* – proteína com domínio Bro-n; *GTF* – glicosiltransferase; *FNIP* – proteína que interage com foliculina; *PEBP1* – proteína ligante de fosfatidiletanolamina 1. NYMV 1 e NYMV 2 correspondem às duas cópias de genes duplicados para cada enzima.

Ainda com relação às aaRs, também desenvolvemos construções filogenéticas para as cópias preditas no genoma de NYMV e KV. Assim como já descrito para o gene da polimerase B, análises filogenéticas para os quatro tipos de aaRs presentes em nossos isolados foram capazes de separar os vírus da família *Mimiviridae* nas três linhagens descritas na literatura [Figura 16 A-D]. Um outro resultado bastante interessante se relacionou com as árvores construídas para as aaRs que apresentam duplicações no genoma do isolado NYMV. Nessas árvores, da mesma forma como já havia sido observado para as análises de sequências de aminoácidos dessas enzimas, uma das cópias duplicadas do NYMV sempre se agrupou com um subgrupo (dentro da linhagem A) formado por APMV, SMBV e Amazonia vírus, enquanto a segunda cópia se agrupava com o nosso outro isolado KV [Figura 16 B-D].




B
Samba virus - R722
Niemeyer virus - R759
Mimivirus (NC_014649)
Amazonia virus - R695
Oyster_virus_R694
Mamavirus (AEQ60843)
Kroon virus - R677
Niemeyer virus - R841
Linhagem A
Moumouvirus Monve (AEX62460)
Moumouvirus goulette (AGF85047)
Linhagem B
Courdo11 virus (ADX97537)
Terra1 virus (ADX97538)
Megavirus Iba (YP_007418694)
Megavirus chiliensis (YP_004894822)
Linhagem
95
89
100
94
99
100
100
42
66
0.05


C
Samba virus - L156
Niemeyer virus - L112
Mimivirus (NC_014649)
Amazonia virus - L129
Lentillevirus (EJN41194)
Oyster virus - L134
Mamavirus (AEQ60303)
Kroon virus - L144
Niemeyer virus - L200
Linhagem A
Megavirus chiliensis (YP_004894958)
Courdo11 virus (ADX97539)
Megavirus Iba (YP_007418849)
Megavirus courdo7 (AEX61965)
Terra1 virus (ADX97540)
Linhagem C
Moumouvirus goulette (AGF84930)
Moumouvirus (YP_007354756)
Linhagem B
Pandoravirus dulcis (YP_008319986)
Pandoravirus salinus (YP_008438883)
85
89
99
92
57
93
95
45
53
74
100
0.5

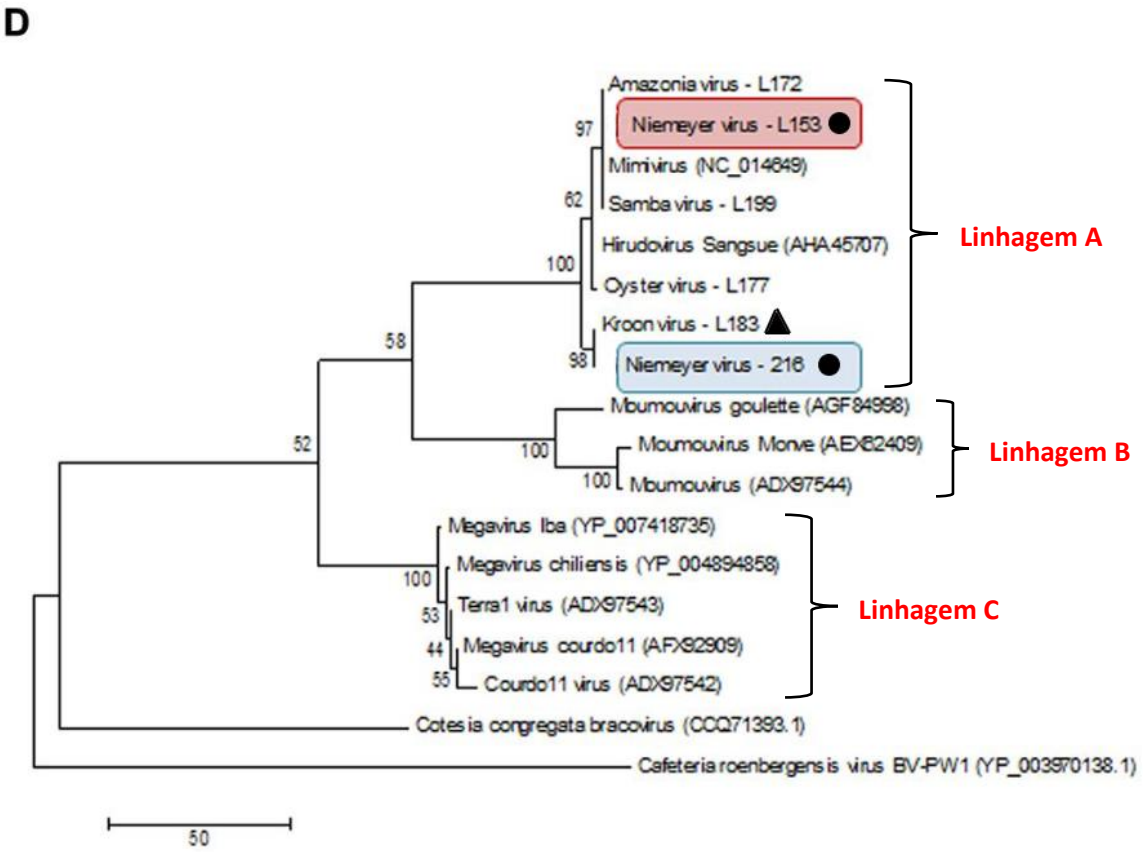


Figura 16 – Análises filogenéticas baseadas nas sequências de (A) arginil-tRNA-sintetase, (B) metionil-tRNA-sintetase, (C) tirosil-tRNA sintetase e (D) cisteinil-tRNA-sintetase, preditas nos genomas de NYMV, KV e outras sequências ortólogas obtidas a partir do banco de dados do NCBI. As árvores foram inferidas utilizando-se do programa MEGA 5 (*neighbor joining* – 1000 replicatas). As aaRs codificadas por NYMV estão representadas com um círculo preto e as codificadas por KV estão representadas por um triangulo. As caixas azul e vermelha representam diferentes cópias de aaRs presentes no genoma de NYMV.

Por fim, como último experimento relacionado às aaRs, resolvemos fazer uma avaliação dos níveis de expressão de RNAm desses genes no genoma de NYMV e comparar com os níveis observados dos genes ortólogos presentes no genoma de APMV. Buscamos através disso observar se a pressão seletiva de manutenção de aaRs duplicadas poderia conferir ao NYMV alguma vantagem com relação ao APMV devido a um maior nível de expressão desses genes. Nossos resultados nos revelaram que a expressão de duas aaRS, metionil e tirosil-tRNA-sintetases (ambas duplicadas em NYMV), foram significantemente maiores em nosso isolado quando comparado com APMV [$p<0,001$ ou $p<0,01$; Figura 17 A e B], enquanto que para os outros dois genes cisteinil (duplicada) e

arginil-tRNA-sintetase (não duplicada), não houve uma diferença significante de expressão entre os vírus [Figura 17 C e D]. Interessante destacar que para todos esses genes, observamos que a sequência responsável pela região promotora permanece a mesma, o que exclui a hipótese de diferenças de expressão devido a modificações nessa região.

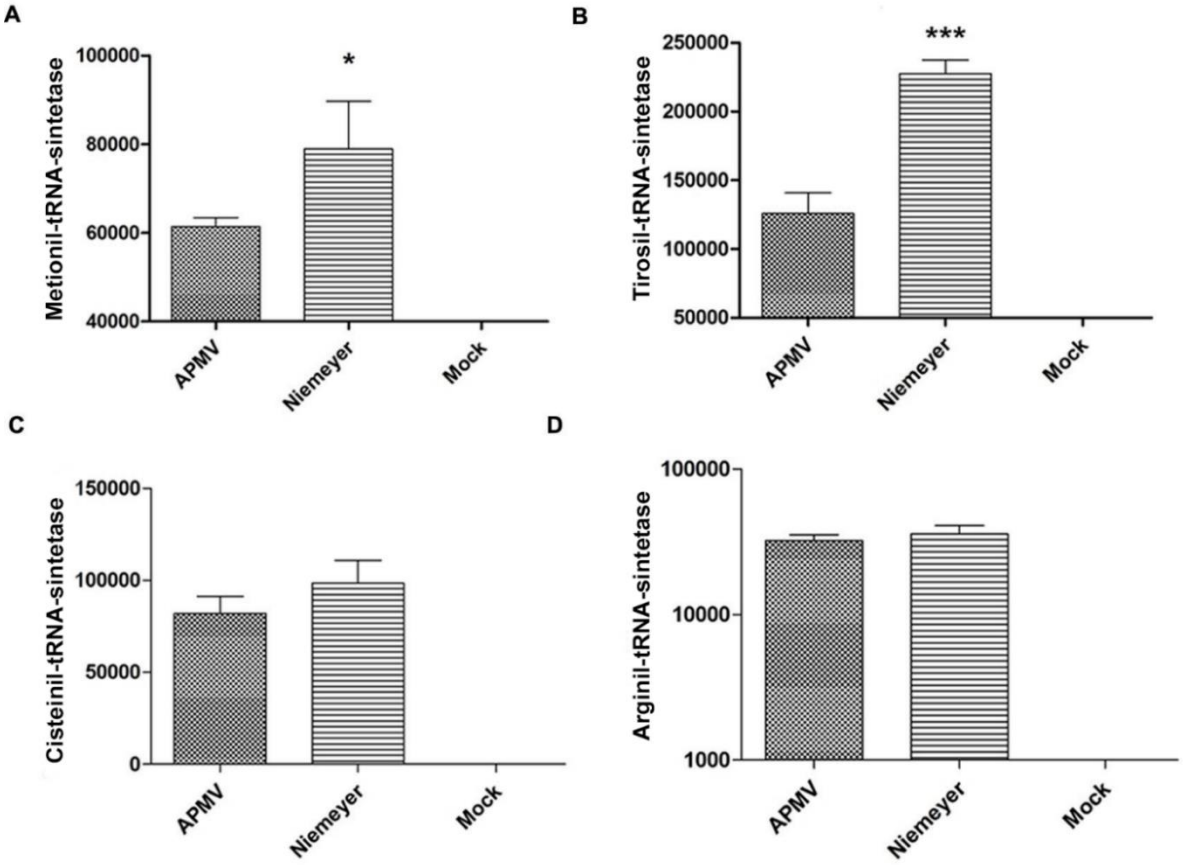
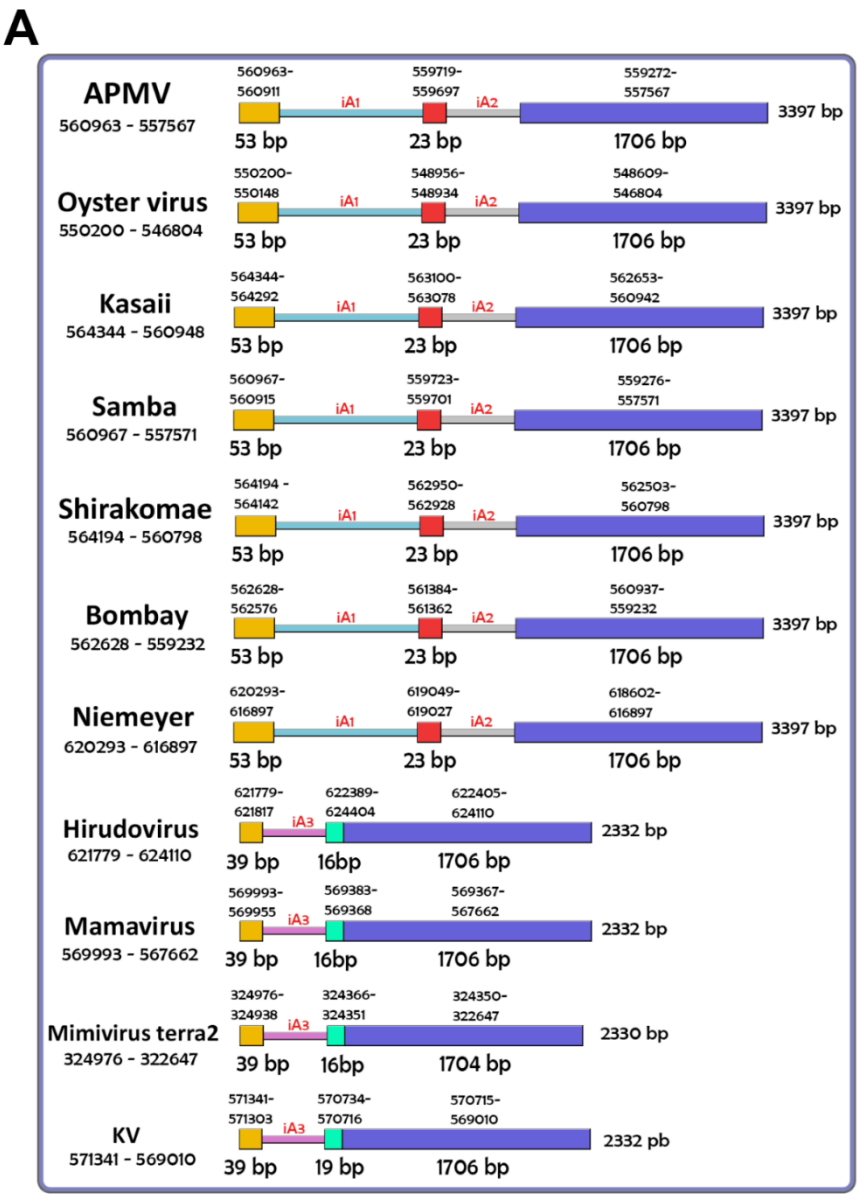


Figura 17 – Expressão do RNAm das aaRs presentes em NYMV. (A) Metionil-tRNA-sintetase, (B) tirosil-tRNA-sintetase, (C) cisteinil-tRNA-sintetase e (D) arginil-tRNA-sintetase. Análises da expressão gênica relativa foram realizadas utilizando-se do método de $\Delta\ \Delta$ Ct e normalizadas por meio da expressão do RNA ribossomal 18S e do RNAm correspondente à helicase viral (calibrados com o menor valor =1). Os valores foram submetidos a diferentes combinações de testes one-way ANOVA e pós teste de Bonferroni (intervalos com 95% de confiança). Diferenças entre grupos foram consideradas quando o *p-valor* fosse menor do que 0,05 (asteriscos). Eixo Y representa abundancia relativa.

6.5 Análises estruturais e de sintenia do gene da MCP

Nós então nos interessamos em estudar um outro gene presente de maneira geral em todos os membros do grupo NCLDV, incluindo os mimivírus. Esse gene codifica para a proteína principal do capsídeo (MCP). No APMV, um trabalho anterior já havia demonstrado que nesse vírus o gene para a MCP apresenta uma característica bastante peculiar, que é ser capaz de produzir um RNAm que sofre o processo de *splicing.* Além disso, nesse mesmo estudo os autores fazem uma caracterização da estrutura e organização sintênica desse gene, no qual este se apresenta na forma de três regiões exônicas, separadas entre si por duas outras regiões não codificantes ou íntrons (Azza et al., 2009). Pouquíssima informação na literatura tem sido compartilhada sobre o tema até então. Após uma análise extensiva dos *layouts* apresentados por diferentes amostras de mimivírus isoladas em todo mundo, incluindo NYMV e KV, nós pudemos observar um tipo de organização separada de três formas principais. O primeiro tipo de organização foi observado por ser um pouco mais complexo do que os outros dois, estando presente em diversos membros da linhagem A. Nesse *layout*, três regiões exônicas (chamadas aqui de *e1, e2* e *e3*) se encontram separadas por duas regiões intrônicas que são específicas para os membros dessa linhagem (chamadas aqui de *iA1* e *iA2*) [Figura 18A]. Nosso isolado NYMV está incluso nesse tipo de organização. No entanto, de maneira interessante, nem todos os membros da linhagem A compartilham dessa mesma estrutura. Nosso isolado KV, juntamente com três outros membros da linhagem A analisados nesse trabalho (hirudovirus, mamavirus e mimivirus Terra2) apresentaram um tipo de organização que se encaixa no segundo tipo de *layout*. Nesses vírus o gene para MCP apresenta uma fração da região correspondente ao *e1* de APMV, seguida por uma região específica e não codificante (chamada aqui de *iA3*), uma nova região específica e codificante (chamada aqui de éxon *eX*) e finalmente por uma sequência homóloga ao éxon *e3* presente em APMV [Figura 18A]. O terceiro *layout* se apresenta melhor relacionado aos membros da linhagem C. Nesses vírus as mesmas regiões compostas pelos éxons *e1* e *e3* estão separadas por duas porções não codificantes específicas para esse grupo, chamadas aqui de *iC1* e *iC2* (a única exceção é o vírus Powai lake Megavirus, que apresenta os éxons *e1* e *e3* separados apenas pela região *iC1*) [Figura 18B].

Já para os mimivírus pertencentes à linhagem B, o pequeno número de vírus isolados e de sequências depositadas nos bancos de dados do NCBI impuseram algumas dificuldades para uma melhor análise. Pelo que pudemos observar, dois membros dessa linhagem (*Moumouvirus goulette* e *Moumouvirus moumouvirus*) se divergiram de maneira bastante acentuada quando a estrutura do gene foi comparada entre os dois. O primeiro apresentou uma organização bastante semelhante aos vírus da linhagem C, enquanto o segundo se assemelhou ao tipo de organização encontrada para nosso isolado KV, com algumas diferenças relacionadas à ausência dos éxons *e1* e *eX*, além da inclusão de duas pequenas regiões intrônicas (*iB1* e *iB2*).



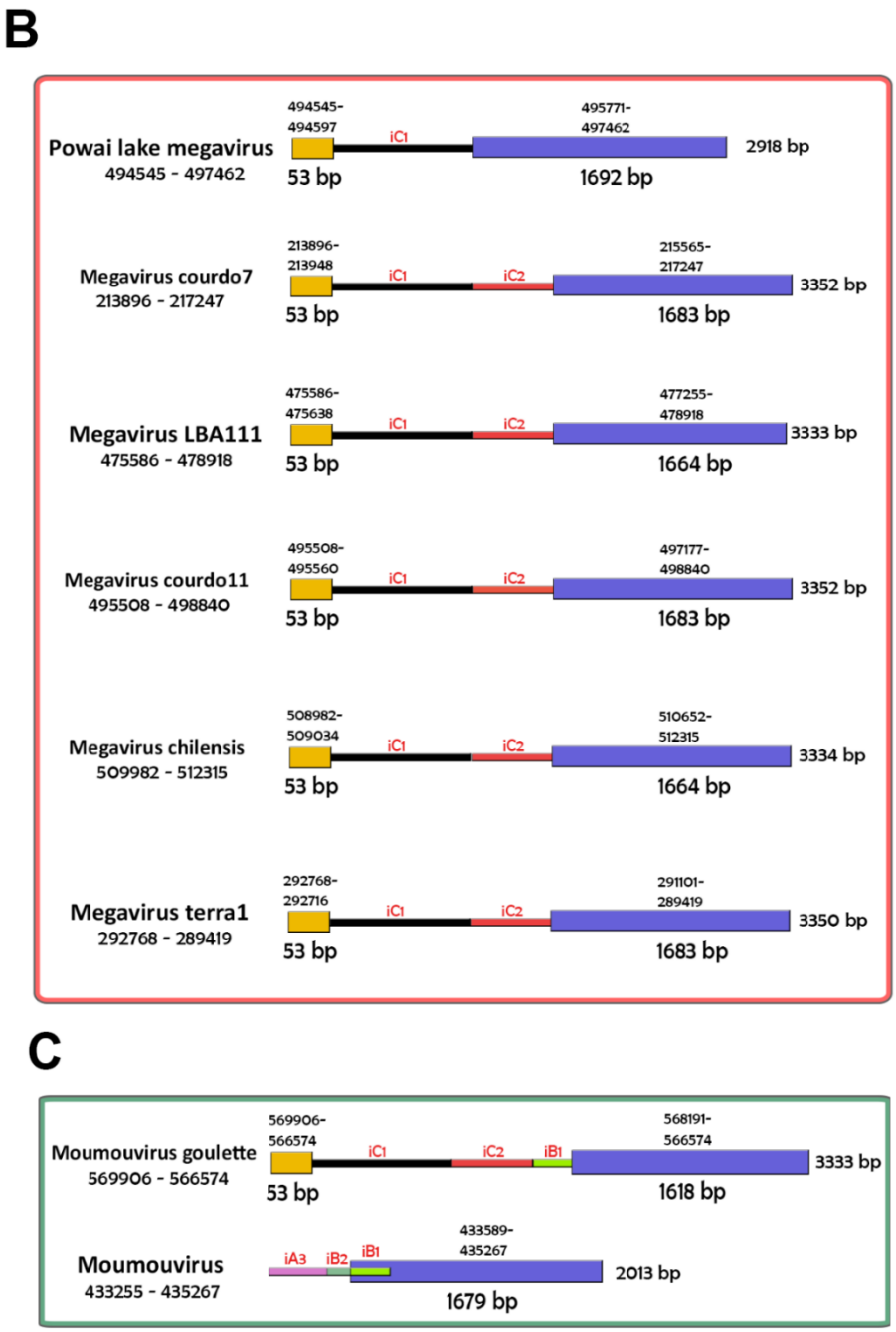
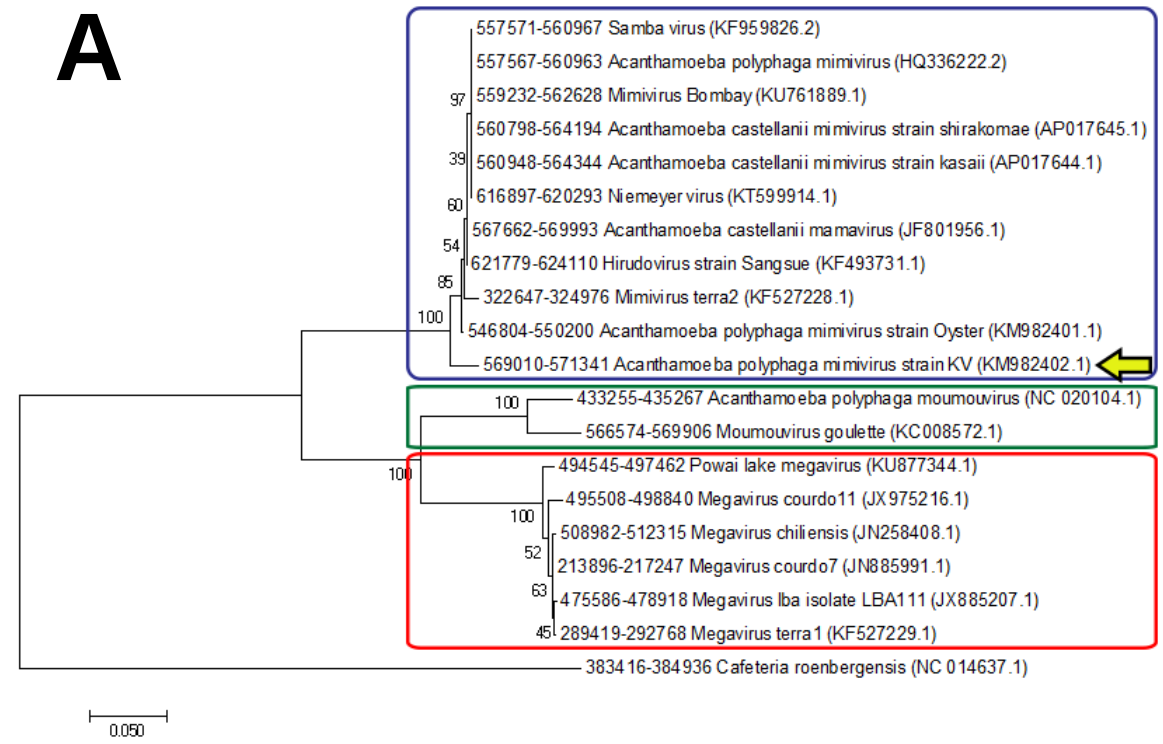


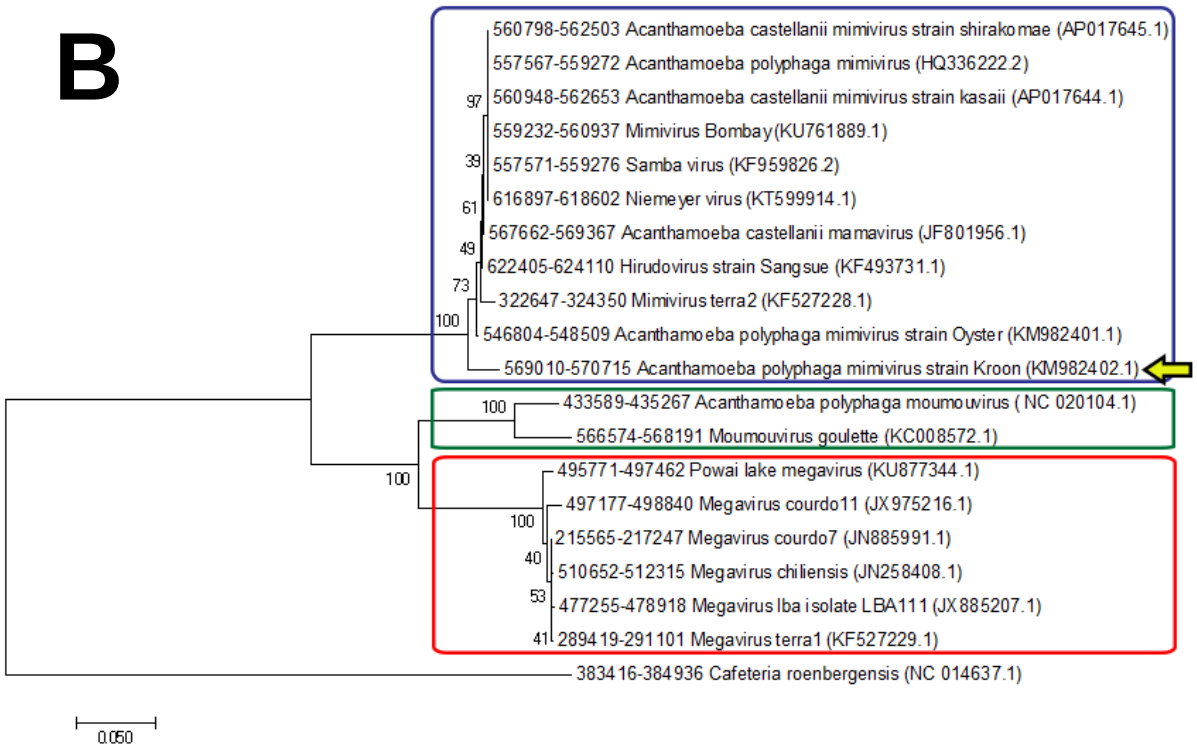
Figura 18 – Esquema geral demonstrando a organização sintênica e estrutural do gene da MCP em diferentes membros da família *Mimiviridae*. As linhagens são representadas aqui por caixas de diferentes cores (linhagem A, azul; linhagem C, vermelha e linhagem B, verde). Representantes dessas linhagens aparentam apresentar diferentes organizações para esse gene. No entanto, o *layout* geral desse gene conservado pode sofrer variações entre seus membros, como observado para KV, mamavirus, hirudovirus e mimivirus Terra 2 quando comparados com outros membros da linhagem A.

6.6 Análises filogenéticas para o gene da MCP

Diferenças estruturais envolvendo o gene do capsídeo, assim como as descritas acima para diversos mimivírus, poderiam estar refletidas no modo como a evolução tem afetado as relações entre membros da família *Mimiviridae*, mesmo para um gene tão conservado. Para termos uma ideia de como essas relações evolutivas são construídas para o gene da MCP, nós também promovemos nesse trabalho a construção de árvores filogenéticas cobrindo sequências tanto relacionadas ao gene completo da MCP nesses vírus gigantes quanto também relacionadas apenas ao éxon *e3* (o maior e única região que está presente em todos os membros avaliados). De forma interessante, as árvores filognéticas refletiram quase que perfeitamente o mesmo perfil observado no esquema de organização do gene descrito anteriormente. Primeiramente, pudemos observar a separação clássica dos mimivírus em três diferentes linhagens, seja considerando a sequência completa do gene ou apenas as relacionadas ao *e3* [Figura 18 A e B].

Outro resultado interessante foi relacionado à recontrução de árvores filogenéticas, da mesma maneira como demonstrado acima, mas considerando apenas organismos que fazem parte da linhagem A (linhagem que contempla nossos isolados). Nessas árvores, KV e os outros três vírus (hirudovirus, mamavirus, mimivirus Terra2) que compartilham um *layout* diferente de outros membros da linhagem, também são representados como os vírus mais filogeneticamente distantes do restante (a única exceção é Oyster virus que também se localizou em um ramo separado na árvore filogenética correspondendo à sequências do *e3*) [Figura 19 C e D]. O uso de sequências do gene da MCP refletiu as relações filogenéticas já descritas para os mimivírus mesmo considerando o vírus *Moumouvirus goulette*. Isso porque, observando atentamente, nem a presença de um *layout* semelhante a vírus pertencentes à linhagem C permitiu com que esse vírus gigante se agrupasse com outro representante que não fosse o *Moumouvirus moumouvirus*, um outro representante da linhagem B [Figura 19 A e B].


A
557571-560967 Samba virus (KF959826.2)
557567-560963 Acanthamoeba polyphaga mimivirus (HQ336222.2)
559232-562628 Mimivirus Bombay (KU761889.1)
560798-564194 Acanthamoeba castellanii mimivirus strain shirakomae (AP017645.1)
560948-564344 Acanthamoeba castellanii mimivirus strain kasaii (AP017644.1)
616897-620293 Niemeyer virus (KT599914.1)
567662-569993 Acanthamoeba castellanii mamavirus (JF801956.1)
621779-624110 Hirudovirus strain Sangsue (KF493731.1)
322647-324976 Mimivirus terra2 (KF527228.1)
546804-550200 Acanthamoeba polyphaga mimivirus strain Oyster (KM982401.1)
569010-571341 Acanthamoeba polyphaga mimivirus strain KV (KM982402.1)
433255-435267 Acanthamoeba polyphaga moumouvirus (NC 020104.1)
566574-569906 Moumouvirus goulette (KC008572.1)
494545-497462 Powai lake megavirus (KU877344.1)
495508-498840 Megavirus courdo11 (JX975216.1)
508982-512315 Megavirus chiliensis (JN258408.1)
213896-217247 Megavirus courdo7 (JN885991.1)
475586-478918 Megavirus lba isolate LBA111 (JX885207.1)
289419-292768 Megavirus terra1 (KF527229.1)
383416-384936 Cafeteria roenbergensis (NC 014637.1)
97
39
60
54
85
100
100
100
100
52
63
45
0.050


B
560798-562503 Acanthamoeba castellanii mimivirus strain shirakomae (AP017645.1)
557567-559272 Acanthamoeba polyphaga mimivirus (HQ336222.2)
560948-562653 Acanthamoeba castellanii mimivirus strain kasaii (AP017644.1)
559232-560937 Mimivirus Bombay (KU761889.1)
557571-559276 Samba virus (KF959826.2)
616897-618602 Niemeyer virus (KT599914.1)
567662-569367 Acanthamoeba castellanii mamavirus (JF801956.1)
622405-624110 Hirudovirus strain Sangsue (KF493731.1)
322647-324350 Mimivirus terra2 (KF527228.1)
546804-548509 Acanthamoeba polyphaga mimivirus strain Oyster (KM982401.1)
569010-570715 Acanthamoeba polyphaga mimivirus strain Kroon (KM982402.1)
433589-435267 Acanthamoeba polyphaga moumouvirus (NC 020104.1)
566574-568191 Moumouvirus goulette (KC008572.1)
495771-497462 Powai lake megavirus (KU877344.1)
497177-498840 Megavirus courdo11 (JX975216.1)
215565-217247 Megavirus courdo7 (JN885991.1)
510652-512315 Megavirus chiliensis (JN258408.1)
477255-478918 Megavirus lba isolate LBA111 (JX885207.1)
289419-291101 Megavirus terra1 (KF527229.1)
383416-384936 Cafeteria roenbergensis (NC 014637.1)
97
39
61
49
73
100
100
100
100
40
53
41
0.050

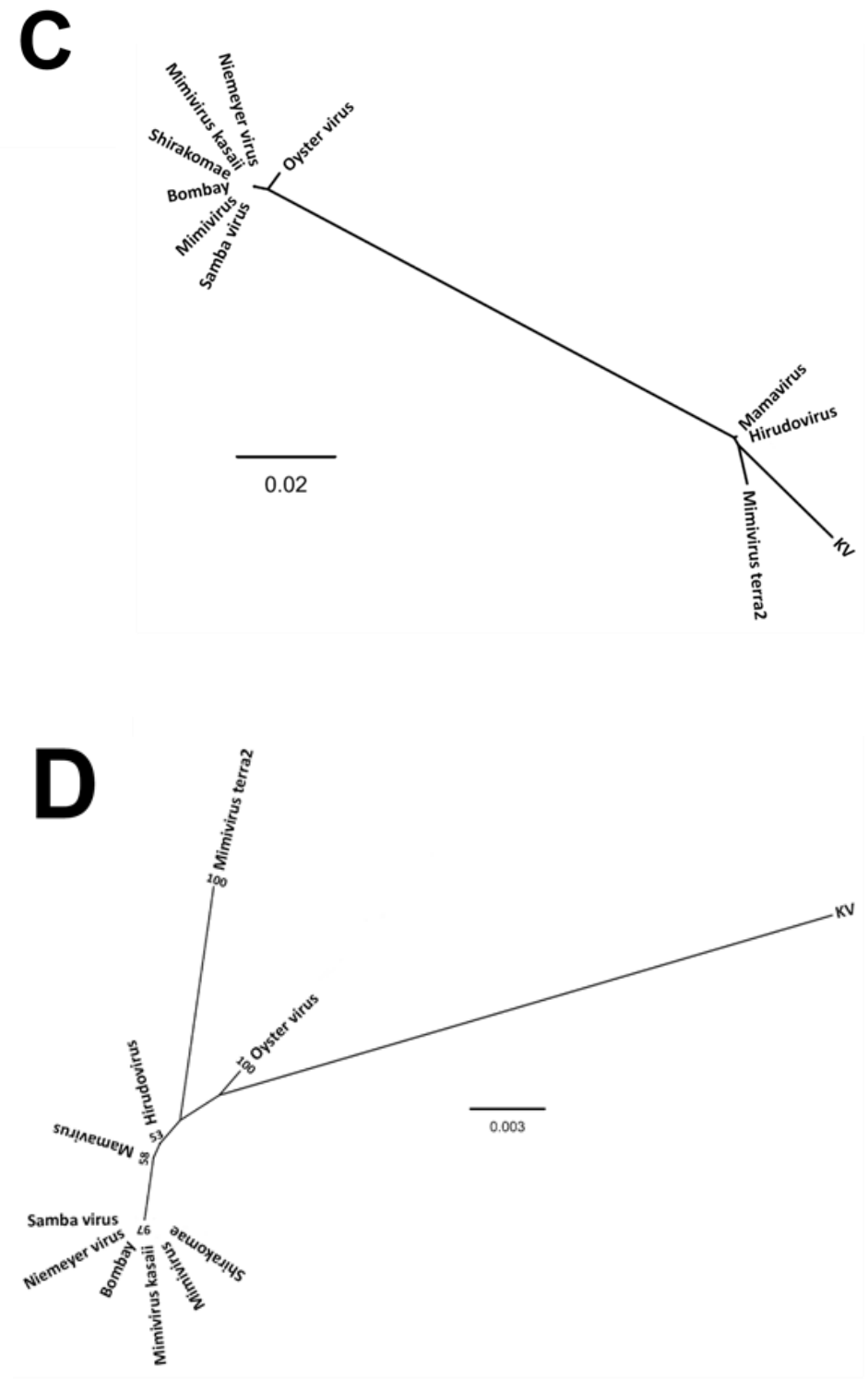


Figura 19 – Filogenia para o gene da MCP (completo ou apenas o éxon *e3*) em diferentes membros da família *Mimiviridae*, incluindo nossos isolados. No painel A está representada uma análise filogenética considerando a sequência completa do gene, contendo membros das três linhagens de mimivírus e representadas por diferentes cores (linhagem A, azul; linhagem B verde

e linhagem C vermelha). No painel B, os mesmos membros são considerados, no entanto as análises são feitas apenas para a porção do éxon *e3*. No painel C está representada uma árvore radial contendo apenas membros da linhagem A, relacionado com a sequência completa do MCP. No painel D as relações filogenéticas de membros da linhagem A estão representadas por outra árvore radial, mas dessa vez com sequências representando apenas o éxon *e3*. Para essas análises, nós utilizamo do programa Mega 7 (método de máxima verossimilhança, 1000 replicatas).

6.6 Análise do transcrito para o gene da MCP

Como as diferenças estruturais no gene da MCP mostraram influenciar o agrupamento dos vírus gigantes nas árvores filogenéticas demonstradas acima, resolvemos então investigar se essas diferenças também poderiam determinar em mudanças relacionadas à produção dos transcritos para esse gene. Nesse teste promovemos o sequenciamento e a análise dos RNAm codificados pelos genes da MCP presentes em APMV e KV [Figura 20 A]. Para o vírus APMV, observamos que as sequências correspondendo aos três éxons do gene são conservadas no transcrito final [Figura 20B]. O gene, inicialmente organizado em uma sequência de 3496 nucleotídeos é reduzido após a maturação do RNAm em uma sequência de 1782 nucleotídeos, apresentando uma cobertura de 51% e identidade de 100% com o gene original. Para KV no entanto, o vírus parece utilizar como RNAm apenas as regiões correspondentes ao éxon *e3* e *eX*, excluindo do transcrito final a região homóloga ao éxon *e1* de APMV e também a porção não codificante *iA3* [Figura 20 B]. Outra característica interessante é que, durante a transcrição e formação do RNAm maduro, as regiões correspondentes às porções 5'UTR e 3'UTR se encontram invertidas quando comparadas à sequência presente originalmente no gene [Figura 20 B]. Em KV, o gene se encontra inicialmente presente como uma sequência de 2332 nucleotídeos, sendo reduzida após a maturação do RNAm a uma sequência de 1716 nucleotídeos. O transcrito apresenta uma cobertura de 73% e identidade de 100% com a sequência do gene original [Figura 20 B].

A

```
#APMV   MAGGLLQLVA YGAQDVYLTG NPMITFFKVV YRRHTNFAVE SIEQFFGGNL GFGKKSSAEI NRSGDLITQV FLKVTLPEVR YCGDFTNFGH VEFAWVRNIG HAIVEETELE IGGSPIDKHY GDWLQIWQDV SSSKDHEKGL AKMLGDVPEL TSISTLSWDV
#KV     ---------- ---------- --MITFFKVV YRRHTNFAVE SIEQFFGGNL GFGKKSSAEI NRSGDLITQV FLKVTLPEVR YCGDFTNFGH VEFAWVRNIG HAIVEETELE IGGSPIDKHY GDWLQIWQDV SSSKDHEKGL AKMLGDVPEL TSISTLSWDV

#APMV   PDNTVLKPSY TLYVPLQFYF NRNNGLALPL IALQYHQVRI YVKFRQADQC YIASDAFKSG CGNLQLDDVS LYVNYVFLDT EERRRFAQVS HEYLIEQLQF TGEESAGSSN SAKYKLNFNH PVKAIYWVTK LGNYQGGKFM TYDPVCWENA RENAAKLLLL
#KV     PDNTVLKPSY TLYVPLQFYF NRNNGLALPL IALQYHQVRI YVKFRQADQC YIASDAFKSG CGNLQLDDVS LYVNYVFLDT EERRRFAQVS HEYLIEQLQY TGEESAGSSN SAKYKLNFNH PVKAIYWVTK LGNYQGGKFM TYDPVCWENA RENAAKLLLL

#APMV   AQYDLDDWGY FQEPGGYECE GNDGRSYVGD CGVQYTAVDP SNPSEEPSYI FNDTTTAEAF DGSLLIGKLA PCVPLLKRNK DVDLKDKVEG IIRIHTDFEN DRMKYPEVEK ITRNDLTLHD LSVPISKYDV DNRVDYIKKF DVTVWQHNNF GLLIDGSGNP
#KV     AQYDLDDWGY FQEPGGYECE GNDGRSYVGD CGVQYTAVDP SNPSEEPSYI FNDTTTAEAF DGSLLIGKLA PCVPLLKRNK DVDLKDKVEG IIRIHTDFEN DRMKYPEVEK ITRNDLTLHD LSVPIAKYDV DNRVDYIKKF DVTVWQHNNF GLLIDGSGNP

#APMV   THEAELQLNG QPRQSKRGGI WYDTVNPTVH HTKSPRDGVN VFSFALNPEE HQPSCTCNFS RIDTAQLNLW FQHFTNHKFA DVFADNDNKV LIFAVNYNVL RMLSGMAGLA YSN*
#KV     THEAELQLNG QPRQSKRGGI WYDTVNPTVH HTKSPRDGVN VFSFALNPEE HQPSCTCNFS RIDTAQLNLW FQHFTNHKFA DVFADNDNKV LIFAVNYNVL RMLSGMAGLA YSN*
```

B

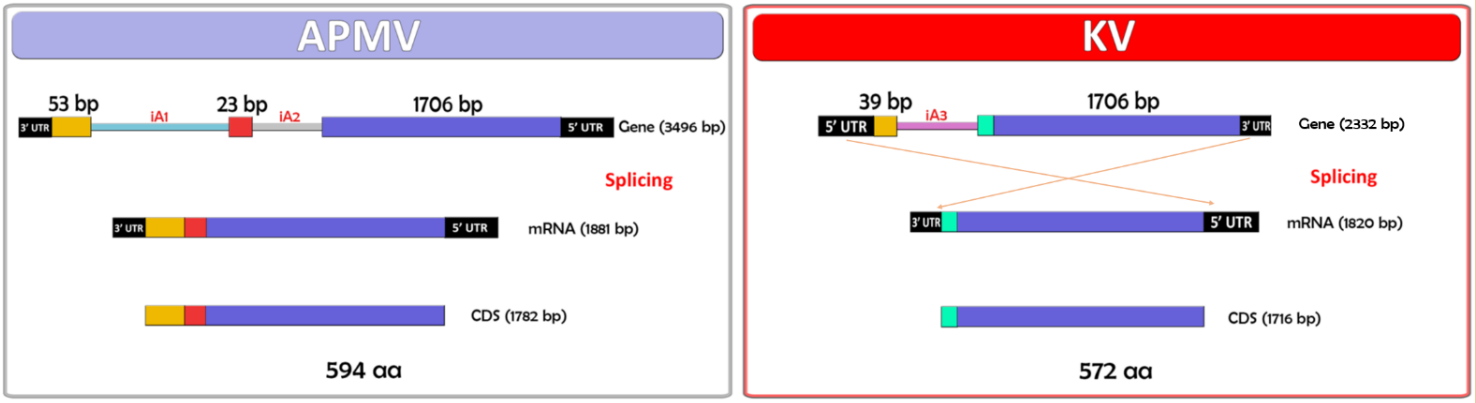


Figura 20 – Esquema geral representando o processo de *splicing* para o gene da MCP nos vírus gigantes APMV e KV. (A) Alinhamento da sequência de aminoácidos produzida pelos dois isolados após a tradução dos seus respectivos genes do capsídeo. Setas vermelhas representam regiões de *mismatches*. (B) O transcrito do gene do capsideo para o APMV sofre um processo de *splicing* no qual sua forma processada é composta pelas sequências dos éxons *e1, e2* e *e3*. KV no entanto, além do seu *layout* diferente, apresenta como transcrito final apenas as regiões compostas pelos éxons *eX* e *e3* (aa = aminoácidos).

7. DISCUSSÃO

Membros da família *Mimiviridae* tem sido isolados mundialmente a partir de uma grande variedade de tipos de amostras (La Scola *et al.*, 2003; Arslan *et al.*, 2011; Colson, Yutin*, et al.*, 2011; Campos *et al.*, 2014). Assim como demonstrado em outros estudos, nosso trabalho também reforça a ideia de que locais marcados por grande concentração de matéria orgânica funcionam como pontos cruciais para o isolamento de uma diversidade de vírus gigantes, o que muito provavelmente pode estar relacionado com a presença de um habitat bastante adequado à sobrevivencia do hospedeiro amebiano desses vírus (Dornas *et al.*, 2015). Nesse trabalho conseguimos realizar o isolamento de duas amostras de mimivírus pertencentes à linhagem A. Até o momento, uma centena desses vírus gigantes (e virus relacionados) tem sido isolados a partir dos mais variados tipos de estratégias envolvendo cultura de amebas, principalmente as que estão incluídas no gênero *Acanthamoeba* (Arslan *et al.*, 2011; Campos *et al.*, 2014; Legendre *et al.*, 2014; Khalil *et al.*, 2016). Interessantemente, apesar do grande número desses vírus já isolados, a grande maioria se encontra dentro da linhagem A (Dornas *et al.*, 2015). Uma hipótese bastante provável é que isso possa ocorrer devido a um fenômeno de seleção desse tipo de amostra, causado pela grande utilização de amebas de um mesmo gênero como suporte para o isolamento viral.

Pelos ensaios de microscopia eletrônica de transmissão, observamos que tanto NYMV quanto KV apresentam um ciclo de multiplicação bastante característico ao de outros mimivírus. No entanto, através de uma observação mais detalhada desse ciclo, os experimentos de curva de ciclo único nos demonstraram que sob as mesmas condições, o isolado KV determina em uma produção de partículas virais muito mais acentuada do que a observada no vírus APMV, indicando que mesmo amostras filogeneticamente relacionadas podem apresentar perfis distintos de multiplicação.

Por meio de análises genômicas e ensaios de expressão gênica também pudemos observar nesse trabalho um fenômeno de duplicação de genes presente em três das quatro cópias de aaRs apresentadas pelo vírus NYMV.

Através desses experimentos é bastante provável sugerir que essas duplicações estejam diretamente associadas com um maior nível de expressão de RNAm das enzimas metionil e tirosil-tRNA-sintetases. O fato das sequências promotoras dos quatro genes apresentarem-se iguais fortalece essa hipótese. A duplicação gênica é um importante meio que guia o processo de evolução molecular permitindo a formação de novos genes com funções biológicas novas ou mesmo redundantes, sendo responsável por afetar a história evolutiva e/ou o *fitness* dos organismos. Esse processo é bem descrito para diversas espécies nos diferentes domínios da vida, especialmente em eucariotos (Zhang, 2003). Simon-Loriere e colaboradores mostraram por meio de análises comparativas entre 55 espécies de vírus de RNA de humanos, animais e plantas (distribuídas entre 19 famílias e 30 diferentes gêneros) que o processo de duplicação de genes apresenta um papel bastante modesto na história evolutiva de vírus com esse tipo de material genético (Simon-Loriere *et al.*, 2013). No entanto, quando observamos os vírus de DNA, esse processo já se torna muito bem descrito (Shackelton e Holmes, 2004). Em membros da família *Mimiviridae,* eventos de duplicação gênica são um campo aberto para estudos. Em outro trabalho, foi demonstrado que esse fenômeno já foi bastante importante na configuração do genoma do APMV durante sua história evolutiva, tendo esse vírus cerca de 1/3 dos seus genes apresentando dentro do mesmo genoma uma outra cópia de gene relacionado (Suhre, 2005). Apesar desse fato, a duplicação de genes relacionados à aaRs não parece ser um evento muito comum em vírus gigantes (assim como demonstrado na Tabela 3), sendo que apenas dois vírus são conhecidos por terem sofrido esse processo, NYMV e o *Acanthamoeba polyphaga moumouvirus* (Yoosuf et al., 2012). Essa característica poderia determinar na apresentação de importantes vantagens adaptativas para ambos, como no auxílio à proteção contra mutações deletérias nesses genes, na emergência de novas amostras de mimivírus e até mesmo auxiliando em uma potencial capacidade infectiva de um espectro de hospedeiros mais amplo. Por exemplo, no caso do gene da metionil-tRNA-sintetase, duplicações para esse gene poderiam ter determinado em uma importante vantagem evolutiva visto que o aminoácido cognato para essa enzima é essencial para o processo de síntese protéica. Para o gene da cisteinil-tRNA-sintetase, um grande nível de conservação é observado dentre os genomas de mimivírus já descritos até hoje.

Isso poderia ter sido importante em uma provável facilitação do processo de duplicação desse gene. Finalmente, para a tirosil-tRNA-sintetase, duplicações nesse gene poderiam representar importantes vantagens no ambiente para o isolado NYMV visto que seus aminoácidos cognatos apresentam uma participação bastante relevante na composição do *codon usage* de diferentes amostras de mimivírus (Silva *et al.*, 2015). Em um outro estudo foi demonstrado também que além dos domínios conservados presentes nas aaRs, relacionados com o processo de aminoacilação, essas enzimas podem também incorporar novos motivos ligados a novas funções biológicas em diferentes organismos eucarióticos, como por exemplo funções relacionadas à atividade angiogênica, atividade angiostática, respostas inflamatórias, etc (Guo *et al.*, 2010). Considerando esse caso, a presença de aaRs duplicadas poderia representar em uma potencial adição de novas e importantes funções nesses vírus gigantes.

Um outro ponto interessante observado para essas enzimas foi que, para cada cópia de aaRs duplicada no gene de NYMV, um deles apresenta 100% de identidade com cópias presentes no genoma de APMV, SMBV e vírus relacionados enquanto a segunda cópia sempre apresenta 100% de identidade com o nosso isolado KV. Esse resultado pode sugerir eventos de transferencia genica entre os ancestrais desses vírus gigantes (KV vs NYMV e/ou APMV-*like* e NYMV), em um hospedeiro comum, durante algum ponto da história evolutiva desses vírus. Acreditamos que a duplicação de genes relacionados ao processo de tradução podem também determinar em nosso isolado uma pequena vantagem durante o seu processo de multiplicação no hospedeiro, estimulando a produção de suas próprias proteínas. A razão para que o gene da cisteinil-tRNA-sintetase não determine em uma expressão gênica significantemente maior ainda precisa ser investigado, mas isso poderia ser o resultado de alguma especificidade gênica relacionada à multiplicação viral em condições intra-amebianas e/ou a presença de outro espectro de hospedeiros (Silva *et al.*, 2015). Os resultados aqui apresentados sugerem a importancia dos fenômenos de duplicação gênica durante a historia evolutiva das aaRs nos mimivírus. Esses genes relacionados à tradução parecem apresentar um papel considerável durante a replicação dos vírus gigantes de ameba. Devido a isso, esse tema se torna interessante para estudos futuros que visem entender a origem,

transferência gênica e evolução das aaRs nos mimivírus, assim como seu papel em processos de aquisição de vantagens evolutivas por esses vírus gigantes.

Também apresentamos nesse trabalho análises sobre a composição estrutural do gene principal do capsídeo em diferentes membros da família *Mimiviridae*. Por meio desses experimentos, observamos que para esse gene, os isolados de vírus gigantes KV, mamavirus, hirudovirus e mimivirus Terra2 apresentam uma estrutura gênica diferente (considerando sequências de regiões intrônicas e exônicas) quando comparados com sequências de regiões homólogas em outros vírus da mesma linhagem, dentre eles NYMV, APMV, mimivirus Shirakomae, mimivirus Kasaii, Oyster virus e SMBV. Esses resultados, juntamente com aqueles observados anteriormente para as sequências das diferentes cópias de aaRs no vírus NYMV, sugere que dentre os vírus gigantes da linhagem A, certos genes podem estar ligados a um fenômeno que determina em um processo de dicotomia entre eles. Essa clara dicotomia nos revela uma importante pista ligada ao gene da MCP, o qual pode ter seguido caminhos evolutivos distintos para cada um dos grupos de isolados de mimivírus. Além disso, esses resultados também demonstraram que mesmo um gene bastante conservado pode carregar diferenças genéticas bastante relevantes entre os isolados. Além dos resultados estruturais, esse fenômeno dicotômico também foi reforçado por nossas análises filogenéticas, que consideraram tanto o gene completo da MCP como apenas a região responsável pelo éxon *e3*. Nessas árvores, os vírus gigantes compostos por KV, mamavirus, hirudovirus e mimivirus Terra2 acabaram se agrupando em posições separadas com relação a outros membros da linhagem A. O fato desse padrão ter se repetido considerando ambas situações implica que se tomarmos em conta análises filogenéticas feitas para o gene da MCP, a região responsável pelo éxon *e3* pode ser associada a um interessante biomarcador capaz de separar os mimivírus no tipo de classificação de três linhagens utilizado atualmente. Outro fator interessante relacionado às analises filogenéticas feitas para o gene da MCP se relacionam com a posição de Oyster virus nas árvores construídas. Como observado, mesmo com uma estrutura gênica que difere da observada para KV, mamavirus, hirudovirus e mimivirus Terra2, em árvores filogenéticas relacionadas à região do éxon *e3*, esse vírus gigante acaba ocupando uma posição filogenética mais

próxima dos vírus mencionados do que a de outros membros da linhagem A. Esse fator reforça a importância da sequência do éxon *e3* na utilização do gene do capsídeo como marcador de classificação de amostras de mimivírus.

Finalmente, pelas análises de sequências que correspondem ao RNAm do gene da MCP, nós demontramos que essa sequência é processada no APMV de uma maneira que ambas regiões intrônicas são retiradas do transcrito final, enquanto para KV, as regiões homólogas ao éxon *e1* e íntron *iA3* são retiradas. Observando esses resultados de maneira conjunta demonstramos que mesmo genes considerados conservados entre diferentes amostras de vírus gigantes podem apresentar diferenças relevantes no que se refere não só à sua porção estrutural mas também como esse elemento é processado após o processo de transcrição.

8. CONCLUSÕES

- ✓ Nesse trabalho apresentamos o isolamento de duas amostras, geneticamente distintas, de mimivírus pertencentes à linhagem A da família *Mimiviridae*.
- ✓ Uma das amostras (NYMV) apresenta um fenômeno de duplicação gênica em três das suas quatro aminoacil-tRNA-sintetases. Esse é um fenômeno raro dentre os mimivírus quando relacionado a esse gene.
- ✓ A duplicação de aminoacil-tRNA-sintetases no genoma de NYMV parece estar ligada a uma maior expressão dos RNA mensageiros de metionil e tirosil-tRNA-sintetases.
- ✓ O gene do capsídeo pode se encontrar organizado (estruturalmente e sintenicamente) de maneiras diferentes nos isolados de mimivírus, mesmo que esses pertençam a membros de uma mesma linhagem.
- ✓ Alguns genes, como as aminoacil-tRNA sintetases e o gene do capsídeo, parecem ter sofrido caminhos evolutivos distintos quando observamos diferentes membros da linhagem A. Esse processo acaba determinando em uma dicotomia para vírus dessa linhagem durante a construção de árvores filogenéticas.
- ✓ A porção do éxon *e3* no gene do capsídeo dos mimivírus funciona como um marcador filogenético para separação nas três linhagens atualmente conhecidas.
- ✓ Mesmo para amostras de mimivírus pertencendo a uma mesma linhagem, o processo de maturação de RNA mensageiro para o gene do capsídeo pode ser diferente durante a transcrição.

9. REFERÊNCIAS BIBLIOGRÁFICAS

1. ABRAHAO, J. S. et al. The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. **PLoS Genet,** v. 13, n. 2, p. e1006532, Feb 2017. ISSN 1553-7404 (Electronic)

1553-7390 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/28207761 >.

2. ABRAHAO, J. S. et al. Growing a giant: evaluation of the virological parameters for mimivirus production. **J Virol Methods,** v. 207, p. 6-11, Oct 2014. ISSN 1879-0984 (Electronic)

0166-0934 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/24972367 >.

3. ABRAHAO, J. S. et al. Acanthamoeba polyphaga mimivirus and other giant viruses: an open field to outstanding discoveries. **Virol J,** v. 11, p. 120, Jun 30 2014. ISSN 1743-422X (Electronic)

1743-422X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/24976356 >.

4. ARSLAN, D. et al. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. **Proc Natl Acad Sci U S A,** v. 108, n. 42, p. 17486-91, Oct 18 2011. ISSN 1091-6490 (Electronic)

0027-8424 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/21987820 >.

5. ASSIS, F. L. et al. Pan-Genome Analysis of Brazilian Lineage A Amoebal Mimiviruses. **Viruses,** v. 7, n. 7, p. 3483-99, Jun 26 2015. ISSN 1999-4915 (Electronic)

1999-4915 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/26131958 >.

6. AZZA, S. et al. Revised Mimivirus major capsid protein sequence reveals intron-containing gene structure and extra domain. **BMC Mol Biol,** v. 10, p. 39, May 11 2009. ISSN 1471-2199 (Electronic)

1471-2199 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/19432951 >.

7. BORATTO, P. V. et al. Niemeyer Virus: A New Mimivirus Group A Isolate Harboring a Set of Duplicated Aminoacyl-tRNA Synthetase Genes. **Front Microbiol,** v. 6, p. 1256, 2015. ISSN 1664-302X (Print)

1664-302X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/26635738 >.

8. BORATTO, P. V. et al. Amoebas as mimivirus bunkers: increased resistance to UV light, heat and chemical biocides when viruses are carried by amoeba hosts. **Arch Virol,** v. 159, n. 5, p. 1039-43, May 2014. ISSN 1432-8798 (Electronic)

0304-8608 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/24271007 >.

9. BOUGHALMI, M. et al. First isolation of a giant virus from wild Hirudo medicinalis leech: Mimiviridae isolation in Hirudo medicinalis. **Viruses,** v. 5, n. 12, p. 2920-30, Nov 27 2013. ISSN 1999-4915 (Electronic)

1999-4915 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/24287596 >.

10. BOWERS, B. Comparison of pinocytosis and phagocytosis in Acanthamoeba castellanii. **Exp Cell Res,** v. 110, n. 2, p. 409-17, Dec 1977. ISSN 0014-4827 (Print)

0014-4827 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/590362 >.

11. BOWERS, B.; OLSZEWSKI, T. E. Acanthamoeba discriminates internally between digestible and indigestible particles. **J Cell Biol,** v. 97, n. 2, p. 317-22, Aug 1983. ISSN 0021-9525 (Print)

0021-9525 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/6350315 >.

12. BOYER, M. et al. Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses. **PLoS One,** v. 5, n. 12, p. e15530, Dec 02 2010. ISSN 1932-6203 (Electronic)

1932-6203 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/21151962 >.

13. CAMPOS, R. K. et al. Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. **Virol J,** v. 11, p. 95, May 14 2014. ISSN 1743-422X (Electronic)

1743-422X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/24886672 >.

14. COLSON, P. et al. Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. **Intervirology,** v. 55, n. 5, p. 321-32, 2012. ISSN 1423-0100 (Electronic)

0300-5526 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/22508375 >.

15. COLSON, P. et al. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. **Arch Virol,** v. 158, n. 12, p. 2517-21, Dec 2013. ISSN 1432-8798 (Electronic)

0304-8608 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/23812617 >.

16. COLSON, P. et al. The giant Cafeteria roenbergensis virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life. **PLoS One,** v. 6, n. 4, p. e18935, Apr 29 2011. ISSN 1932-6203 (Electronic)

1932-6203 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/21559486 >.

17. COLSON, P.; LA SCOLA, B.; RAOULT, D. Giant Viruses of Amoebae: A Journey Through Innovative Research and Paradigm Changes. **Annu Rev Virol,** v. 4, n. 1, p. 61-85, Sep 29 2017. ISSN 2327-0578 (Electronic)

2327-056X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/28759330 >.

18. COLSON, P. et al. Viruses with more than 1,000 genes: Mamavirus, a new Acanthamoeba polyphaga mimivirus strain, and reannotation of Mimivirus genes. **Genome Biol Evol,** v. 3, p. 737-42, 2011. ISSN 1759-6653 (Electronic)

1759-6653 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/21705471 >.

19. CULBERTSON, C. G. et al. Experimental infection of mice and monkeys by Acanthamoeba. **Am J Pathol,** v. 35, n. 1, p. 185-97, Jan-Feb 1959. ISSN 0002-9440 (Print)

0002-9440 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/13617418 >.

20. DORNAS, F. P. et al. Isolation of new Brazilian giant viruses from environmental samples using a panel of protozoa. **Front Microbiol,** v. 6, p. 1086, 2015. ISSN 1664-302X (Print)

1664-302X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/26500630 >.

21. FILEE, J. Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution. **Front Microbiol,** v. 6, p. 593, 2015. ISSN 1664-302X (Print)

1664-302X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/26136734 >.

22. FOURNIER, G. P. et al. Molecular evolution of aminoacyl tRNA synthetase proteins in the early history of life. **Orig Life Evol Biosph,** v. 41, n. 6, p. 621-32, Dec 2011. ISSN 1573-0875 (Electronic)

0169-6149 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/22200905 >.

23. GALARDINI, M. et al. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. **Source Code Biol Med,** v. 6, p. 11, Jun 21 2011. ISSN 1751-0473 (Electronic)

1751-0473 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/21693004 >.

24. GHEDIN, E.; CLAVERIE, J. M. Mimivirus relatives in the Sargasso sea. **Virol J,** v. 2, p. 62, Aug 16 2005. ISSN 1743-422X (Electronic)

1743-422X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/16105173 >.

25. GUO, M.; YANG, X. L.; SCHIMMEL, P. New functions of aminoacyl-tRNA synthetases beyond translation. **Nat Rev Mol Cell Biol,** v. 11, n. 9, p. 668-74, Sep 2010. ISSN 1471-0080 (Electronic)

1471-0072 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/20700144 >.

26. IYER, L. M.; ARAVIND, L.; KOONIN, E. V. Common origin of four diverse families of large eukaryotic DNA viruses. **J Virol,** v. 75, n. 23, p. 11720-34, Dec 2001. ISSN 0022-538X (Print)

0022-538X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/11689653 >.

27. IYER, L. M. et al. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. **Virus Res,** v. 117, n. 1, p. 156-84, Apr 2006. ISSN 0168-1702 (Print)

0168-1702 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/16494962 >.

28. KHALIL, J. Y. et al. High-Throughput Isolation of Giant Viruses in Liquid Medium Using Automated Flow Cytometry and Fluorescence Staining. **Front Microbiol,** v. 7, p. 26, 2016. ISSN 1664-302X (Print)

1664-302X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/26858703 >.

29. KHAN, N. A.; SIDDIQUI, R. Is there evidence of sexual reproduction (meiosis) in Acanthamoeba? **Pathog Glob Health,** v. 109, n. 4, p. 193-5, Jun 2015. ISSN 2047-7732 (Electronic)

2047-7724 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/25800982 >.

30. KOONIN, E. V.; KRUPOVIC, M.; YUTIN, N. Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. **Ann N Y Acad Sci,** v. 1341, p. 10-24, Apr 2015. ISSN 1749-6632 (Electronic)

0077-8923 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/25727355 >.

31. KOONIN, E. V.; SENKEVICH, T. G.; DOLJA, V. V. The ancient Virus World and evolution of cells. **Biol Direct,** v. 1, p. 29, Sep 19 2006. ISSN 1745-6150 (Electronic)

1745-6150 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/16984643 >.

32. KOONIN, E. V.; YUTIN, N. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. **Intervirology,** v. 53, n. 5, p. 284-92, 2010. ISSN 1423-0100 (Electronic)

0300-5526 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/20551680 >.

33. KUZNETSOV, Y. G. et al. Morphogenesis of mimivirus and its viral factories: an atomic force microscopy study of infected cells. **J Virol,** v. 87, n. 20, p. 11200-13, Oct 2013. ISSN 1098-5514 (Electronic)

0022-538X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/23926353 >.

34. LA SCOLA, B. Looking at protists as a source of pathogenic viruses. **Microb Pathog,** v. 77, p. 131-5, Dec 2014. ISSN 1096-1208 (Electronic)

0882-4010 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/25218687 >.

35. LA SCOLA, B. et al. A giant virus in amoebae. **Science,** v. 299, n. 5615, p. 2033, Mar 28 2003. ISSN 1095-9203 (Electronic)

0036-8075 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/12663918 >.

36. LA SCOLA, B. et al. The virophage as a unique parasite of the giant mimivirus. **Nature,** v. 455, n. 7209, p. 100-4, Sep 04 2008. ISSN 1476-4687 (Electronic)

0028-0836 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/18690211 >.

37. LEGENDRE, M. et al. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. **Proc Natl Acad Sci U S A,** v. 111, n. 11, p. 4274-9, Mar 18 2014. ISSN 1091-6490 (Electronic)

0027-8424 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/24591590 >.

38. LI, R. et al. MD Simulations of tRNA and Aminoacyl-tRNA Synthetases: Dynamics, Folding, Binding, and Allostery. **Int J Mol Sci,** v. 16, n. 7, p. 15872-902, Jul 13 2015. ISSN 1422-0067 (Electronic)

1422-0067 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/26184179 >.

39. LLOYD, D. Encystment in Acanthamoeba castellanii: a review. **Exp Parasitol,** v. 145 Suppl, p. S20-7, Nov 2014. ISSN 1090-2449 (Electronic)

0014-4894 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/24726698 >.

40. LUTHER, K. B. et al. Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. **J Biol Chem,** v. 286, n. 51, p. 43701-9, Dec 23 2011. ISSN 1083-351X (Electronic)

0021-9258 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/22045808 >.

41. LWOFF, A. The concept of virus. **J Gen Microbiol,** v. 17, n. 2, p. 239-53, Oct 1957. ISSN 0022-1287 (Print)

0022-1287 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/13481308 >.

42. MARCIANO-CABRAL, F.; CABRAL, G. Acanthamoeba spp. as agents of disease in humans. **Clin Microbiol Rev,** v. 16, n. 2, p. 273-307, Apr 2003. ISSN 0893-8512 (Print)

0893-8512 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/12692099 >.

43. MOREIRA, D.; LOPEZ-GARCIA, P. Ten reasons to exclude viruses from the tree of life. **Nat Rev Microbiol,** v. 7, n. 4, p. 306-11, Apr 2009. ISSN 1740-1534 (Electronic)

1740-1526 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/19270719 >.

44. NASIR, A.; CAETANO-ANOLLES, G. A phylogenomic data-driven exploration of viral origins and evolution. **Sci Adv,** v. 1, n. 8, p. e1500527, Sep 2015. ISSN 2375-2548 (Print)

2375-2548 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/26601271 >.

45. 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 Evol Biol,** v. 12, p. 156, Aug 24 2012. ISSN 1471-2148 (Electronic)

1471-2148 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/22920653 >.

46. PIACENTE, F. et al. Characterization of a UDP-N-acetylglucosamine biosynthetic pathway encoded by the giant DNA virus Mimivirus. **Glycobiology,** v. 24, n. 1, p. 51-61, Jan 2014. ISSN 1460-2423 (Electronic)

0959-6658 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/24107487 >.

47. PIACENTE, F. et al. Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (Viosamine). **J Biol Chem,** v. 287, n. 5, p. 3009-18, Jan 27 2012. ISSN 1083-351X (Electronic)

0021-9258 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/22157758 >.

48. PRESTON, T. M.; KING, C. A. Amoeboid locomotion of Acanthamoeba castellanii with special reference to cell-substratum interactions. **J Gen Microbiol,** v. 130, n. 9, p. 2317-23, Sep 1984. ISSN 0022-1287 (Print)

0022-1287 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/6389765 >.

49. RAOULT, D. et al. The 1.2-megabase genome sequence of Mimivirus. **Science,** v. 306, n. 5700, p. 1344-50, Nov 19 2004. ISSN 1095-9203 (Electronic)

0036-8075 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/15486256 >.

50. RAOULT, D.; LA SCOLA, B.; BIRTLES, R. The discovery and characterization of Mimivirus, the largest known virus and putative pneumonia agent. **Clin Infect Dis,** v. 45, n. 1, p. 95-102, Jul 01 2007. ISSN 1537-6591 (Electronic)

1058-4838 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/17554709 >.

51. REED, L. J.; MUENCH, H. A simple method of estimating fifty percent endpoints. **The American Journal of Hygiene,** v. 27, n. 3, p. 493-497, 1 May 1938 1938.

52. RENESTO, P. et al. Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. **J Virol,** v. 80, n. 23, p. 11678-85, Dec 2006. ISSN 0022-538X (Print)

0022-538X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/16971431 >.

53. RIBAS DE POUPLANA, L.; SCHIMMEL, P. Aminoacyl-tRNA synthetases: potential markers of genetic code development. **Trends Biochem Sci,** v. 26, n. 10, p. 591-6, Oct 2001. ISSN 0968-0004 (Print)

0968-0004 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/11590011 >.

54. RODRIGUES, R. A. et al. Mimivirus Fibrils Are Important for Viral Attachment to the Microbial World by a Diverse Glycoside Interaction Repertoire. **J Virol,** v. 89, n. 23, p. 11812-9, Dec 2015. ISSN 1098-5514 (Electronic)

0022-538X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/26378162 >.

55. SCHEID, P. Relevance of free-living amoebae as hosts for phylogenetically diverse microorganisms. **Parasitol Res,** v. 113, n. 7, p. 2407-14, Jul 2014. ISSN 1432-1955 (Electronic)

0932-0113 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/24828345 >.

56. SCHULZ, F. et al. Giant viruses with an expanded complement of translation system components. **Science,** v. 356, n. 6333, p. 82-85, Apr 07 2017. ISSN 1095-9203 (Electronic)

0036-8075 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/28386012 >.

57. SCHUSTER, F. L.; LEVANDOWSKY, M. Chemosensory responses of Acanthamoeba castellanii: visual analysis of random movement and responses to chemical signals. **J Eukaryot Microbiol,** v. 43, n. 2, p. 150-8, Mar-Apr 1996. ISSN 1066-5234 (Print)

1066-5234 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/8720945 >.

58. SHACKELTON, L. A.; HOLMES, E. C. The evolution of large DNA viruses: combining genomic information of viruses and their hosts. **Trends Microbiol,** v. 12, n. 10, p. 458-65, Oct 2004. ISSN 0966-842X (Print)

0966-842X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/15381195 >.

59. SIDDIQUI, R.; KHAN, N. A. Biology and pathogenesis of Acanthamoeba. **Parasit Vectors,** v. 5, p. 6, 2012. ISSN 1756-3305 (Electronic)

1756-3305 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/22229971 >.

60. SILVA, L. C. et al. Modulation of the expression of mimivirus-encoded translation-related genes in response to nutrient availability during Acanthamoeba castellanii infection. **Front Microbiol,** v. 6, p. 539, 2015. ISSN 1664-302X (Print)

1664-302X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/26082761 >.

61. SIMON-LORIERE, E.; HOLMES, E. C.; PAGAN, I. The effect of gene overlapping on the rate of RNA virus evolution. **Mol Biol Evol,** v. 30, n. 8, p. 1916-28, Aug 2013. ISSN 1537-1719 (Electronic)

0737-4038 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/23686658 >.

62. SIMPSON, J. T. et al. ABySS: a parallel assembler for short read sequence data. **Genome Res,** v. 19, n. 6, p. 1117-23, Jun 2009. ISSN 1088-9051 (Print)

1088-9051 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/19251739 >.

63. SINCLAIR, J. L.; MCCLELLAN, J. F.; COLEMAN, D. C. Nitrogen Mineralization by Acanthamoeba polyphaga in Grazed Pseudomonas paucimobilis Populations. **Appl Environ Microbiol,** v. 42, n. 4, p. 667-71, Oct 1981. ISSN 0099-2240 (Print)

0099-2240 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/16345864 >.

64. SUHRE, K. Gene and genome duplication in Acanthamoeba polyphaga Mimivirus. **J Virol,** v. 79, n. 22, p. 14095-101, Nov 2005. ISSN 0022-538X (Print)

0022-538X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/16254344 >.

65. WOESE, C. R. et al. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. **Microbiol Mol Biol Rev,** v. 64, n. 1, p. 202-36, Mar 2000. ISSN 1092-2172 (Print)

1092-2172 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/10704480 >.

66. XIAO, C. et al. Structural studies of the giant mimivirus. **PLoS Biol,** v. 7, n. 4, p. e92, Apr 28 2009. ISSN 1545-7885 (Electronic)

1544-9173 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/19402750 >.

67. YOOSUF, N. et al. Draft genome sequences of Terra1 and Terra2 viruses, new members of the family Mimiviridae isolated from soil. **Virology,** v. 452-453, p. 125-32, Mar 2014. ISSN 1096-0341 (Electronic)

0042-6822 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/24606689 >.

68. YOOSUF, N. 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,** v. 4, n. 12, p. 1324-30, 2012. ISSN 1759-6653 (Electronic)

1759-6653 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/23221609 >.

69. YUTIN, N. et al. Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. **Virol J,** v. 10, p. 106, Apr 04 2013. ISSN 1743-422X (Electronic)

1743-422X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/23557328 >.

70. 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, Oct 2014. ISSN 1096-0341 (Electronic)

0042-6822 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/25042053 >.

71. YUTIN, N. et al. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. **Virol J,** v. 6, p. 223, Dec 17 2009. ISSN 1743-422X (Electronic)

1743-422X (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/20017929 >.

72. ZAUBERMAN, N. et al. Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. **PLoS Biol,** v. 6, n. 5, p. e114, May 13 2008. ISSN 1545-7885 (Electronic)

1544-9173 (Linking). Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/18479185 >.

73. ZHANG, J. Evolution by gene duplication: an update. **Trends in Ecology & Evolution,** v. 18, n. 6, p. 7, June 2003 2003.

10. ANEXOS

Anexo 1: Componentes utilizados no preparo das soluções da microscopia de transmissão:

Solução Fixadora

- Glutaraldeído 25% ---------- 6mL
- Tampão fosfato 0,2M ---------- 30mL
- Água deionizada (q.s.p. 60 mL) ---------- 24mL

Volume final da solução: 60mL

Solução A: Fosfato de sódio monobásico 0,2M

- NaH2PO4.7H2O ---------- 5,37g
- Água deionizada ---------- 100mL

Solução B: Fosfato de sódio dibásico 0,2M

- NaH2PO4.7H2O ---------- 5,37g
- Água deionizada ---------- 100mL

Solução A+B: tampão fosfato 0,2M, pH 7,4

- Solução A ---------- 19mL
- Solução B ---------- 81mL

Volume final da solução: 100mL

Anexo 2 – Tabela contendo números de acesso e a posição no genoma das sequências utilizadas na montagem das árvores filogenéticas do item 5.3.5

Vírus	Posição no genoma	Número de acesso (Pubmed)
Acanthamoeba polyphaga mimivirus	560963 – 557567	HQ336222.2
Kroon virus	571341 – 569010	KM982402.1
Oyster virus	550200 – 546804	KM982401.1
Mimivirus Shirakomae	564194 – 560798	AP017645.1
Mimivirus kasaii	564344 – 560948	AP017644.1
Mimivirus Bombay	562628 – 559232	KU761889.1
Samba virus	560967 – 557571	KF959826.2
Niemeyer virus	620293 – 616897	KT599914.1
Mimivirus terra 2	324976 – 322647	KF527228.1
Hirudovirus Sangsue	621779 – 624110	KF493731.1
Mamavirus	569993 – 567662	JF801956.1
Powai lake Megavirus	494545 – 497462	KU877344.1
Megavirus terra 1	292768 – 289419	KF527229.1
Megavirus LBA111	475586 – 478918	JX885207.1
Megavirus courdo11	495508 – 498840	JX975216.1
Megavirus courdo7	213896 – 217247	JN885991.1
Megavirus chilensis	508982 – 512315	JN258408.1
Moumouvirus goulette	569906 – 566574	KC008572.1
Acanthamoeba polyphaga moumouvirus	433255 - 435267	NC_020104.1
Cafeteria roenbergensis virus	383416 - 384936	NC_014637.1

11. PRODUÇÕES BIBLIOGRÁFICAS

JVI Accepted Manuscript Posted Online 8 November 2017
J. Virol. doi:10.1128/JVI.01782-17

Title: The analysis of KV mimivirus major capsid gene and its transcript highlights a distinct pattern of gene evolution and splicing among mimiviruses

Running title: KV mimivirus major capsid analysis

Authors and Affiliations

Paulo Victor Miranda Boratto[1], Fábio Pio Dornas[1], Lorena Christine Ferreira da Silva[1], Rodrigo Araújo Lima Rodrigues[1], Graziele Pereira Oliveira[1], Juliana Reis Cortines[2], Betânia Paiva Drumond[1], Jônatas Santos Abrahão[1*]

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

2. Instituto de Microbiologia Paulo de Góes, Departamento de Virologia – Centro de Ciências da Saúde, Cidade Universitária, Ilha do Fundão, Rio de Janeiro, RJ, Brazil

Jônatas Santos Abrahão

jonatas.abrahao@gmail.com

Abstract

The inclusion of *Mimiviridae* members in the putative monophyletic 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 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. In this work, we report the characterization of a new mimivirus isolate, called KV mimivirus. Analysis of the structure, syntheny and phylogenetic relationships of the MCP gene 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 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 being an important part of the so called "hallmark genes", there is little information about the processing and structure of the major capsid protein (MCP) gene in many mimivirus isolates. In this work, we have analyzed the structure, syntheny and phylogenetic relationships of the MCP gene in many mimivirus isolates, showing 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.

Introduction

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 can 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, called as "core/hallmark genes", are present in all the members, suggesting a probable common ancestry between them (4). The nucleocytoplasmic large DNA viruses (NCLDVs) comprise one of those subgroups which is composed of seven different viral families: *Ascoviridae, Asfarviridae, Iridoviridae, Marseilleviridae, Mimiviridae, Phycodnaviridae* and *Poxviridae* (5, 6). These viral families share a common set of core genes that are generally related to the replication of the viral genome and also to the composition of the viral particle structure, encoding the major capsid protein, a primase-helicase, a family B DNA polymerase, a packaging ATPase and a transcription factor (5, 6).

Amoeba giant viruses belonging to the *Mimivirus* genus (*Mimiviridae* family) were first described in 2003, during a characterization study of a pathogenic amoeba-associated microorganism that was related to outbreaks of nosocomial pneumonia in a hospital in Bradford, England (7). Since then a large diversity of related viruses has been isolated from many different places (8-13). As new giant viruses were being described, these organisms started to be phylogenetically classified based on the sequence of the gene related to DNA polymerase B. This gene is conserved between different viral members, but presents sufficient differences to separate them in three different lineages (A, B and C) (14). Other genes that form part of the group of hallmark genes have not been as well explored in the literature when it comes to the phylogenetic relationships between mimiviruses. One example is that related to the synthesis of the major capsid protein (MCP). This gene plays a key role in the formation of the viral particle structure, representing the most abundant glycoprotein in *Acanthamoeba polyphaga mimivirus* (APMV) (15). In a previous study, Azza and colleagues interestingly observed that the APMV MCP gene is structured in five different regions, with three exons separated by two intronic sites (16). The idea of a gene undergoing splicing in giant viruses has also been shown in another study which confirmed that the transcript of the APMV capsid gene is composed of three exonic regions (17). As

observed for DNA polymerase B, the MCP gene also presents high homology rates in other members of the genus *Mimivirus*. However, there is not much information on how this gene is structured in different *Mimivirus* members.

In this work, we report the characterization of Kroon virus (KV), a new mimivirus belonging to lineage A. Here, we analyze the structure, syntheny and phylogenetic relationships of the major capsid gene (a hypothetically conserved gene shared by the *Mimiviridae* members) in different giant virus isolates, including mimivirus KV. By using phylogenetic analyses and comparing the order of the intronic and exonic regions that compose the MCP gene of these giant viruses, we observed a distinct pattern of genetic organization and gene evolution involving KV and other mimiviruses, even among members of lineage A, to which KV belongs. These differences are also reflected by the way in which the MCP mRNA is spliced. By sequencing the mature transcripts of the MCP gene in both APMV (type-species among mimiviruses and member of the lineage A) and KV, we have observed many differences in terms of content and organization, suggesting that this gene is processed in a different form in KV. Taken together, our results enabled us to highlight the MCP gene as a new marker that could also help in future studies involving the characterization of mimiviruses.

Results

Virus isolation

In 2012, in an attempt to isolate novel strains of giant viruses, we decided to collect water samples from several urban lakes located in different cities of Brazil. A total of 14 water samples were collected at equidistant points around an urban lagoon located in the city of Lagoa Santa, Brazil (Figure 1A). Inoculation of these samples in *Acanthamoeba castellanii* monolayers prompted us to isolate a new amoebal-associated microorganism, initially characterized by the observation of cytopathic effects (CPEs) in these cell cultures after 3-4 days after infection. The CPEs were characteristic of the isolation of other giant viruses, including cell rounding, the absence of vacuoles and cell lysis at later times of viral infection.

Virus characterization

With this new microorganism, we started to investigate any resemblance with other mimivirus isolates by using transmission electron microscopy, morphometric analysis and the analysis of genomic amplification by qPCR. First, we have linked our new isolated virus to members of the family *Mimiviridae* since an amplification of the conserved RNA helicase gene was detected.

Electron microscopy and morphometric assays enabled us to observe a great number of viral particles, with a pseudo-icosahedral symmetry, fibrils surrounding the capsid and a size that is comparable to that which has been observed for other studies involving mimivirus isolation (7, 8, 11-13). These mimivirus-like particles presented an average size of 673.43 nm in total, fibers of about 128.9 nm and a capsid size of 433.2, spread along the acanthamoebal cytoplasm [Figure 1B and C]. KV also presents several characteristics that resemble steps of mimivirus replication, as we could observe (i) the initial process of viral penetration, in which the viral particles start to be surrounded by *Acanthamoeba* pseudopods, (ii) the presence of electron-dense regions in the amoebal cytoplasm called viral factories, where the giant virus particles start to be assembled, and (iii) the release of mature particles by cell lysis [Figure 2A-C]. By performing one-step-growth-curve assays, we observed that our isolate reaches the plateau for viral production at around 8-10 hours post-infection (the same as described for other mimiviruses), but with viral titers exceeding the production of APMV particles in about 2.5 logs [Figure 2D]. Finally, the phylogenetic characterization of our isolate, both for the RNA helicase and DNA polymerase genes, have clustered it with members of the lineage A from *Mimiviridae*, that includes *Acanthamoeba castellanii* mamavirus, Mimivirus terra2, Niemeyer virus, Samba virus, and Oyster virus, among others (8, 9, 13, 18) [Figure 3A and 3B].

MCP gene layout

We then started to become more interested in a specific gene that is present among members that compose the family *Mimiviridae*. This gene encodes the major capsid protein (MCP). In APMV, it has been demonstrated that the MCP gene is responsible for producing a spliced mRNA, with a structure that is composed of three exons separated by two untranscribed regions (16, 17). In another giant virus, the *Acanthamoeba castellanii* mamavirus, this gene was found to be lacking the two intronic regions, but presented its own type of untranscribed sequence (10). As far as we know, no other studies have investigated how the capsid gene is arranged and

phylogenetically related in members of the family *Mimiviridae*. After extensive analysis of the mimivirus MCP gene, we have seen that the MCP gene of those viruses can present different layouts. The first layout is slightly more complex than the others, being present in most members of lineage A. It is composed of three exons (called here *e1*, *e2* and *e3*), separated by two intronic regions that are specific for members of this lineage (called here *iA1* and *iA2*) [Figure 4]. Interestingly, not all the members from lineage A share this same structure. KV, as well as three other analyzed members of lineage A (Hirudovirus, Mamavirus and Mimivirus terra2), have a different layout for their MCP genes. In these viruses, the MCP gene has the described structure: (i) a fraction of the region corresponding to exon *e1* in APMV, followed by (ii) a specific untranscribed region (called *iA3* here), and finally by (iii) another specific region (called exon *eX* here), which is followed by a sequence homologue to exon *e3* in APMV [Figure 4]. The third layout is most closely related to members of lineage C. In these viruses, the same transcribed regions composed of exons *e1* and *e3* are separated by two adjacent and specific intronic regions called *iC1* and *iC2* (the exception is Powai lake megavirus, which has the *e1* and *e3* regions separated only by *iC1*). As for the mimiviruses belonging to lineage B, the small number of isolate sequences deposited in NCBI have imposed some difficulties to better analyze the group. From what we can see, two members from this lineage, *Moumouvirus goulette* and *Moumouvirus moumouvirus*, have considerably diverged when we consider how their MCP gene is structured, with the first one resembling the structure of mimiviruses from lineage C and the latter resembling the layout observed for KV virus, with some differences related to the absence of *e1* and *eX*, in addition to the inclusion of two short intronic regions (*iB1* and *iB2*) [Figure 4].

Major capsid gene phylogenetic analysis

Structural differences involving the capsid gene, as described above for several mimiviruses, may reflect the way in which evolution has affected the diverse relationships between members of the family *Mimiviridae* for such a conserved gene. To get an idea of these relationships, we have constructed phylogenetic trees that covered either the entire sequence of the MCP gene or just the sequence related to exon *e3* (the largest and most conserved of the exons described above). Interestingly, the constructed phylogenetic trees almost reflected the profile which was previously described in the structural composition of MCP for the different giant viruses. When the

capsid gene of these viruses was analyzed, considering either the whole gene or just the *e3* fragment, we observed a clear separation that divided the viral members into three groups corresponding to the lineages A, B and C, as expected based on other hallmark genes [Figure 5A-B]. The construction of phylogenetic trees containing members of lineage A alone also matched the disposition of groups from trees that considered the three lineages together [Figure 5C-D]. This is relevant because, in those trees, KV and the other three viruses that share a distinct pattern of MCP organization (Hirudovirus, Mamavirus and Mimivirus terra2) are represented as the viruses that were most phylogenetically separated from the rest of the members in lineage A (the exception is Oyster virus, which is also located on a separate branch in trees corresponding to the analysis of exon 3) [Figure 5D]. The same occurred for Powai lake megavirus, occupying a different branch to the members belonging to the lineage C [Figure 5A-B]. As for the *Moumouvirus goulette*, even with this MCP gene structure resembling organization of the capsid gene in viruses of lineage C, the phylogenetic trees have demonstrated that this virus has been grouped together with *Moumouvirus moumouvirus*, the other member belonging to lineage B of mimiviruses [Figure 5A-B].

Analysis of the MCP gene transcript

As structural differences in the MCP gene have been shown to influence viral grouping in the phylogenetic trees described above, we chose to investigate whether these differences could also determine major changes along the sequence of the mature transcript related to this gene. For this test, we sequenced and analyzed the mRNA encoding the MCP protein belonging to APMV and KV. For APMV, the sequences corresponding to the three exons of MCP gene are conserved in the final transcript [Figure 6A-B]. The gene, initially organized in a sequence of 3496 nucleotides, is reduced after mRNA maturation in a sequence of 1782 nucleotides, with a query cover of 51% and identity of 100% to the original gene. As observed in another study, the transcript is processed as a sequence that excludes the two intronic regions present originally in this gene. For KV however, the virus seems to use as its mature mRNA only the region corresponding to the larger exon *e3* and the adjacent *exon X*, excluding from the processed transcript both the region homologue to the exon *e1* in APMV and the intronic region *iA3* [Figure 6A-B]. Curiously, analysis of the nucleotide sequence from exon *eX* has shown this region as being a small fragment from the 53bp region belonging to exon *e1* in the APMV-like viruses (data not shown). In addition, during

the transcription and formation of the mature mRNA, the regions corresponding to the 5'UTR and 3'UTR seem to be inverted from the observed in the original sequence. In KV, the gene is initially present as a sequence of 2332 nucleotides, being reduced after mRNA maturation to a sequence of 1716 nucleotides. The transcript has a query cover of 73% and an identity of 100% with the original gene.

Discussion

Members of the family *Mimiviridae* have been isolated worldwide from an increasing variety of samples (7, 8, 10, 12, 13). As demonstrated in other studies, environments where there are high concentrations of organic matter seem to work as hot-spots, marking the presence of a great diversity of mimiviruses, which is probably related to the suitable habitat for their amoebal host to live (8, 19). Here, we have managed to obtain another mimivirus isolate belonging to the lineage A of the family *Mimiviridae*. To date, around a hundred mimivirus (and related viruses) isolates have been obtained by many different strategies that use the culture of amoebal cells, especially species that are included in the *Acanthamoeba* genus (12, 13, 20, 21). Despite the considerable number of isolated giant viruses, the vast majority belong to lineage A (19). This could be related to the fact that most of the platforms for mimivirus isolation employ the use of *Acanthamoeba* organisms as the main type of cellular culture. The use of this narrow host spectrum as cell support for mimivirus isolation may be filtering away the discovery of other species of giant viruses that would potentially be present in the samples.

Transmission electron microscopy has shown us that KV has a viral cycle typical of other mimiviruses. Nevertheless, by undertaking a more detailed analysis of the cycle, *one-step-growth curve* assays have demonstrated that this new isolate, under the same conditions, produces a much larger number of mature viral particles than APMV, highlighting that, even phylogenetically related, different mimivirus isolates may present distinct replication profiles in *Acanthamoeba* sp

By analyzing the structural composition of the major capsid protein gene in different members of the family *Mimiviridae,* we have reinforced a previous observation which suggested that, for some genes, there is a certain level of dichotomy separating mimiviruses inside lineage A (8). Here, we have observed that, for the capsid gene, KV, Mamavirus, Hirudovirus and Mimivirus terra2 present a different genetic structure (considering the intronic and exonic regions of the gene) when compared with

homologous regions in other members of the same lineage. This creates inside lineage A an apparent separation with two existent groups, the one corresponding to the four viruses mentioned above and the second group composed of other viruses analyzed in this work: APMV, Mimivirus Shirakomae, Oyster virus, Mimivirus Bombay, Mimivirus Kasaii, Niemeyer virus and Samba virus. This clear dichotomy gives us an important clue, enabling the MCP gene to be considered to follow distinct evolutionary pathways for each group of mimivirus isolates. This has also shown that these many giant viruses harbor different structural variations for such a conserved gene as the MCP.

In addition to the structural results, the dichotomic event was also supported when we considered phylogenetic trees associated with both the whole MCP gene, as well as just for the region of exon 3, in the different mimiviruses. In those trees, the giant viruses composed by KV, Mamavirus, Hirudovirus and Mimivirus terra2 have remained at separate positions to others belonging to lineage A. The fact that this pattern has been the same considering the two situations implies that if we take into account the phylogenetic trees made for the MCP gene, the use of the region associated with *e3* alone gives us an interesting biomarker that also serves to separate mimiviruses in the current classification involving the three lineages. As stated before, interestingly, this dichotomic phenomena has also already been seen for many other genes encoding different aminoacyl-tRNA-synthetases, indicating that KV could represent a divergent lineage A mimivirus (8).

Another interesting factor linked to the phylogenetic analysis that comprehends the MCP gene is related to the position of Oyster virus in the evolutionary trees. As observed, even with a gene structure that differs from that observed for KV, Mamavirus, Hirudovirus and Mimivirus terra2, this giant virus is represented in the phylogenetic trees in a position that is closer to the aforementioned viruses than to other members of lineage A, also taking into consideration analyses that have been performed both for the whole gene or just for the *e3* region. This result reinforces the importance that exonic region 3 can contribute as a phylogenetic biomarker to characterize mimivirus strains.

Finally, by the analysis of sequences belonging to the MCP mRNA, we have shown that while this sequence is processed in APMV in a way that both intronic regions are spliced out, for KV, the MCP transcript is processed differently, before the steps involving translation. In KV, the regions homologous to exon *e1* in APMV and intronic region *iA3* are excluded. Taken together, our results indicate that in giant viruses, even genes that are considered to be conserved among different species, may

present relevant differences that are related not only to the structural part of the gene but also to how this element is processed after transcription.

Material and Methods

Collection of environmental samples

Water samples were collected at the Central lagoon, Lagoa Santa city, Minas Gerais state, Brazil [Figure 1A]. The lagoon has a total area of 1.31 km^2, an approximate perimeter of 6.5 km and maximum depth of between 6 and 7 meters (22). It normally receives household waste water. A total of 14 water samples, 1ml each, were collected along all its extension, at equidistant places, and stored at 4°C until isolation procedures were performed. The samples were collected on the surface of the water, near aquatic plants and areas that indicated the formation of biofilms.

Amoebal culture procedures

Acanthamoeba castellanii (ATCC 30010) were grown in 75 cm^2 cell culture flasks (Nunc, USA), in PYG (peptone-yeast extract-glucose) medium supplemented with 7% fetal calf serum (FCS, Cultilab, Brazil), 25 mg/ml Fungizone (amphotericin B, Cristalia, São Paulo, Brazil), 500 U/ml penicillin and 50 mg/ml gentamicin (Schering-Plough, Brazil).

Enrichment protocol and isolation procedures

Samples were initially submitted to an enrichment protocol adapted from Arslan and collaborators (12, 23). Briefly, 500 µl of each sample was added to 4.5 ml of water-rice medium (40 grains of rice per liter of water) and kept in the dark at room temperature for 20 days. Following this incubation, an input of 5000 amoebas was performed and the samples were incubated at the same conditions for more than 10 days. The samples were then filtered through 1.2 µm membranes (to retain impurities) and then through 0.2 µm membranes (to retain potential giant viruses isolated in these samples). The membranes were eluted in 500 µl of PBS and 100 µl of each eluate was inoculated into amoebal monolayers, contained in 96-well plates in a total series of three passages. Cytopathic effects (CPE) were evaluated daily.

DNA Extraction and PCR assays

Samples presenting CPE in amoebal monolayers were tested for mimivirus by a qPCR targeting the conserved helicase gene (primers: 5'-ACCTGATCCACATCCCATAACTAAA-3' and 5'-GGCCTCATCAACAAATGGTTTCT-3'). The DNA was extracted by PCI and quantified by nanodrop before the PCR assays. The qPCR was performed with a commercial mix Power SYBr Green (Applied Biosystems, USA), primers (4 mM each) and 1 ml sample in reaction of 10 ml final volume. All reactions were performed in a StepOne thermocycler: 95°C (10 min), 40 cycles [95°C (15 s)/60°C (15 s)], followed by a dissociation step (specific Tm = 73°C).

The whole genome of KV was obtained and partially analyzed as described in another study (access number: KM982402.1) (24).

Virus purification procedures

The KV particles were isolated and purified from infected amoebae as previously described (13). Briefly, after reaching confluence, the amoebas were infected with the new isolated giant virus and incubated at 37°C until the appearance of CPE. Virus-rich supernatants from the infected amoeba were collected and filtered through a 0.8-µm (Millipore, USA) filter to remove amoeba debris. The viruses were then purified using a sucrose cushion suspended in PBS, stored at -80°C and titrated using the $TCID_{50}$ methodology (25).

Transmission electron microscopy analyses

For the transmission electron microscopy (TEM), *A. castellanii* cells were infected with KV at an MOI of 0.01. Uninfected amoebae were used as controls. When ~70% of the cells presented cytopathic effects, the amoeba monolayer was fixed with 2.5% glutaraldehyde (Merck, Germany) for 1 hour at room temperature in 0.1 M sodium phosphate buffer. Amoebas were then post-fixed with 2% osmium tetroxide and embedded in Epon resin, and ultrathin sections were examined under a Transmission Electron Microscopy TECNAI G2-20 – SuperTwin FEI – 120 kV at the Center of Microscopy, UFMG, Brazil.

One-step-growth curve

To evaluate the replication profile of the new giant virus isolate, six-well plates containing $1x10^7$ amoebas/well were infected with KV at m.o.i. of 10 and incubated at 32ºC for 0, 0.5, 1, 2, 4, 6, 8, 12 and 24 hours. Infected cells were collected and centrifuged, and the pellet was used for titration in amoebas. The viral titer was determined using the $TCID_{50}$ method calculated with the Reed-Muench method (25).

Synteny analysis of the MCP gene

In this work we have also searched for differences in the layout of the major capsid protein gene of several mimivirus samples. Given that APMV, as the type species of the family *Mimiviridae*, has the best characterized genome among mimiviruses, its nucleotide sequence was used as reference for MCP analyses. To do that, for each gene feature belonging to APMV (*exons 1, 2, 3* and *introns 1*, *2*), we searched for similar sequences in other viruses using the BLAST tool from the NCBI database (e-value of 10^{-3}). Sequences that were found to be homologous to APMV were separated and then the final gene layouts were constructed based in the position and orientation occupied by these different sequences. Specific MCP gene features found in some viruses were also searched for in other mimivirus samples. A scheme of the MCP gene was prepared including isolates from all mimivirus lineages (Figure 4).

MCP mRNA sequencing

Twenty-four-well plates containing $1x10^5$ amoebas/well were infected with KV or APMV at m.o.i. of 5 and incubated at 32ºC. After 30 min and 6 hours, cells were collected and centrifuged, and the pellet was homogenized used for total RNA extraction, reverse transcription and conventional PCR for the MCP gene amplification. The amplicons were examined by electrophoresis in a 1% agarose gel, TBE buffer and run at 150V. The amplicon band was purified and sequenced in both orientations, in triplicate (3730 DNA Analyzer; Thermo Fischer Scientific, Waltham, MA, USA). The sequences were analyzed using the alignment tool available in the software MEGA 7.

Phylogenetic analyses

For the phylogenetic analyses, separated alignments for the major capsid protein, DNA polymerase B subunit, RNA helicase genes and the *exon 3* region of the major capsid gene were used. Strains of several viruses of the three mimivirus lineages (A, B and C) were selected to assemble the dataset. The predicted nucleotide sequences were

obtained from NCBI GenBank and aligned using ClustalW in Mega 7.0 software. Trees were constructed using maximum likelihood method and bootstrap of 1,000.

Conflict of interest statement

The authors declare no conflict of interest

Acknowledgements

We would like to thank colleagues from Gepvig, Laboratório de Vírus, Centro de Microsopia da UFMG and Aix Marseille Université for their excellent support. Also, we would like to thank Pro-reitoria de pesquisa da Universidade Federal de Minas Gerais, CAPES, FAPEMIG, CNPq. JSA is a CNPq researcher.

References

1. **Lwoff A.** 1957. The concept of virus. J Gen Microbiol **17:**239-253.

2. **Koonin EV, Yutin N.** 2010. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. Intervirology **53:**284-292.

3. **Moreira D, Lopez-Garcia P.** 2009. Ten reasons to exclude viruses from the tree of life. Nat Rev Microbiol **7:**306-311.

4. **Koonin EV, Senkevich TG, Dolja VV.** 2006. The ancient Virus World and evolution of cells. Biol Direct **1:**29.

5. **Iyer LM, Aravind L, Koonin EV.** 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J Virol **75:**11720-11734.

6. **Iyer LM, Balaji S, Koonin EV, Aravind L.** 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res **117:**156-184.

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

8. **Boratto PV, Arantes TS, Silva LC, Assis FL, Kroon EG, La Scola B, Abrahao JS.** 2015. Niemeyer Virus: A New Mimivirus Group A Isolate Harboring a Set of Duplicated Aminoacyl-tRNA Synthetase Genes. Front Microbiol **6:**1256.

9. **Andrade KR, Boratto PP, Rodrigues FP, Silva LC, Dornas FP, Pilotto MR, La Scola B, Almeida GM, Kroon EG, Abrahao JS.** 2015. Oysters as hot spots for mimivirus isolation. Arch Virol **160:**477-482.

10. **Colson P, Yutin N, Shabalina SA, Robert C, Fournous G, La Scola B, Raoult D, Koonin EV.** 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.

11. **Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, Robert C, Azza S, Klose T, Wong J, Rossmann MG, La Scola B, Raoult D, Koonin EV.** 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.

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

13. **Campos RK, Boratto PV, Assis FL, Aguiar ER, Silva LC, Albarnaz JD, Dornas FP, Trindade GS, Ferreira PP, Marques JT, Robert C, Raoult D, Kroon EG, La Scola B, Abrahao JS.** 2014. Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. Virol J **11:**95.

14. **Yutin N, Colson P, Raoult D, Koonin EV.** 2013. Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. Virol J **10:**106.

15. **Renesto P, Abergel C, Decloquement P, Moinier D, Azza S, Ogata H, Fourquet P, Gorvel JP, Claverie JM.** 2006. Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. J Virol **80:**11678-11685.

16. **Azza S, Cambillau C, Raoult D, Suzan-Monti M.** 2009. Revised Mimivirus major capsid protein sequence reveals intron-containing gene structure and extra domain. BMC Mol Biol **10:**39.

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

18. **Yoosuf N, Pagnier I, Fournous G, Robert C, Raoult D, La Scola B, Colson P.** 2014. Draft genome sequences of Terra1 and Terra2 viruses, new members of the family Mimiviridae isolated from soil. Virology **452-453:**125-132.

19. **Dornas FP, Khalil JY, Pagnier I, Raoult D, Abrahao J, La Scola B.** 2015. Isolation of new Brazilian giant viruses from environmental samples using a panel of protozoa. Front Microbiol **6:**1086.

20. **Khalil JY, 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.

21. **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 JM.** 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.

22. **Brighenti LS, Pinto-Coelho RM, Bezerra-Neto JF, Gonzaga AV.** 2011. Parâmetros morfométricos da Lagoa Central (Lagoa Santa, Estado de Minas Gerais): comparação de duas metodologias. . Acta Scientiarum Biological Sciences **33:**7.

23. **Dornas FP, Silva LC, de Almeida GM, Campos RK, Boratto PV, Franco-Luiz AP, La Scola B, Ferreira PC, Kroon EG, Abrahao JS.** 2014. Acanthamoeba polyphaga mimivirus stability in environmental and clinical substrates: implications for virus detection and isolation. PLoS One **9:**e87811.

24. **Assis FL, Bajrai L, Abrahao JS, Kroon EG, Dornas FP, Andrade KR, Boratto PV, Pilotto MR, Robert C, Benamar S, Scola BL, Colson P.** 2015. Pan-Genome Analysis of Brazilian Lineage A Amoebal Mimiviruses. Viruses **7:**3483-3499.

25. **Reed LJ, Muench H.** 1938. A simple method of estimating fifty percent endpoints. The American Journal of Hygiene **27:**493 - 497.

Figures

Figure 1. KV mimivirus collection site, electron microscopy and morphometric analysis. (A) A map of the Central Lagoon, from where the samples were collected. (B) An image of KV virus particle, observed from the transmission electron microscopy assays. (C) Shows the size, in nanometers, of different components of KV particle: only the capsid, only the fibers and the particle as a whole.

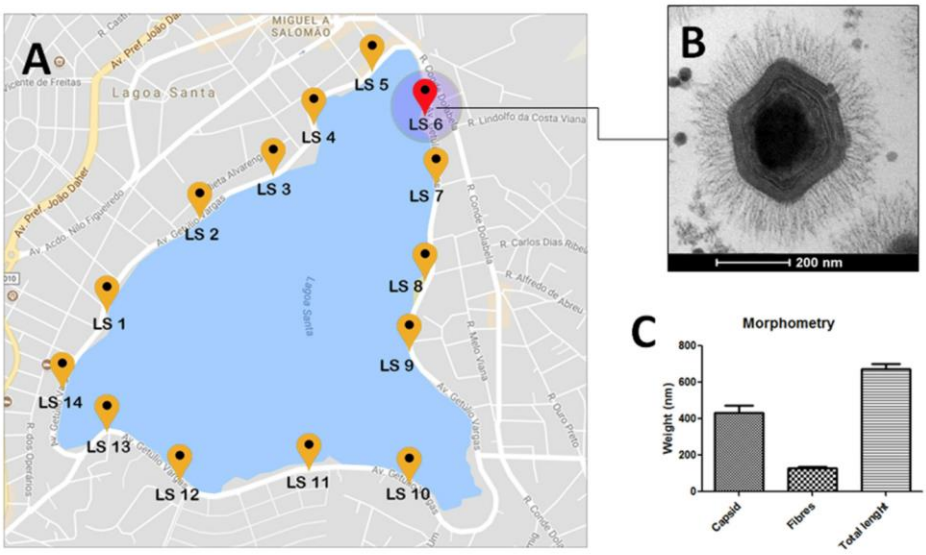
Figure 2. KV replication cycle. The major steps of KV life cycle are highlighted by transmission electron microscopy and one-step-growth curve assays. In (A) giant viral particles initiate the process of viral penetration by being phagocytosed by *Acanthamoeba castellanii* cells. (B) Shows a later step, in which a KV viral factory, located at the cytoplasm of an infected amoebal host, starts to produce and bud several particles. (C) Demonstrates the last step of KV replication, in which the produced mature particles start to be released by lysis of the amoebal cell. (D) Shows the general replication cycle of KV in comparison with APMV by one-step-growth curve assays. After 24h of infection, KV exceeds the production of APMV particles in about 2.5 logs.

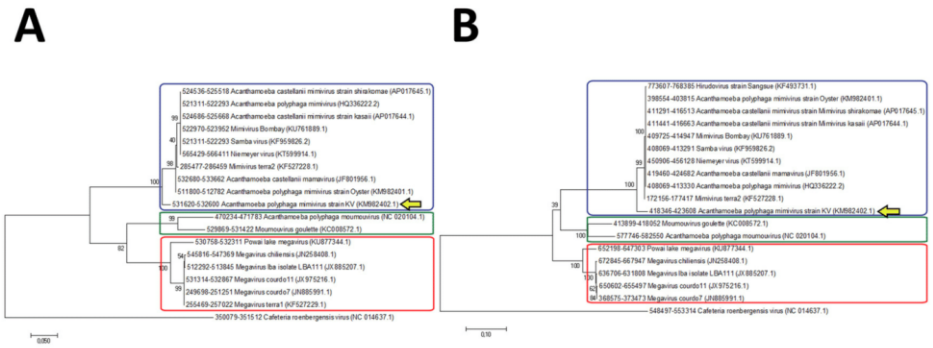
Figure 3. Phylogenetic reconstruction of mimiviruses. Phylogenetic analyses for many representatives of the different *Mimiviridae* lineages (A, B, and C) including KV virus strain, based on the (A) RNA helicase and (B) DNA polymerase-B amino acid sequences. For these analysis, we used MEGA7 software (maximum-likelihood method and 1,000 bootstrap replicates). The different *Mimiviridae* lineages were highlighted using different colored boxes (lineage A: blue, lineage B: green and lineage C: red) and KV is shown by a yellow arrow, being clustered with other mimivirus isolates belonging to lineage A. For each sequence, the gene identification numbers are indicated.

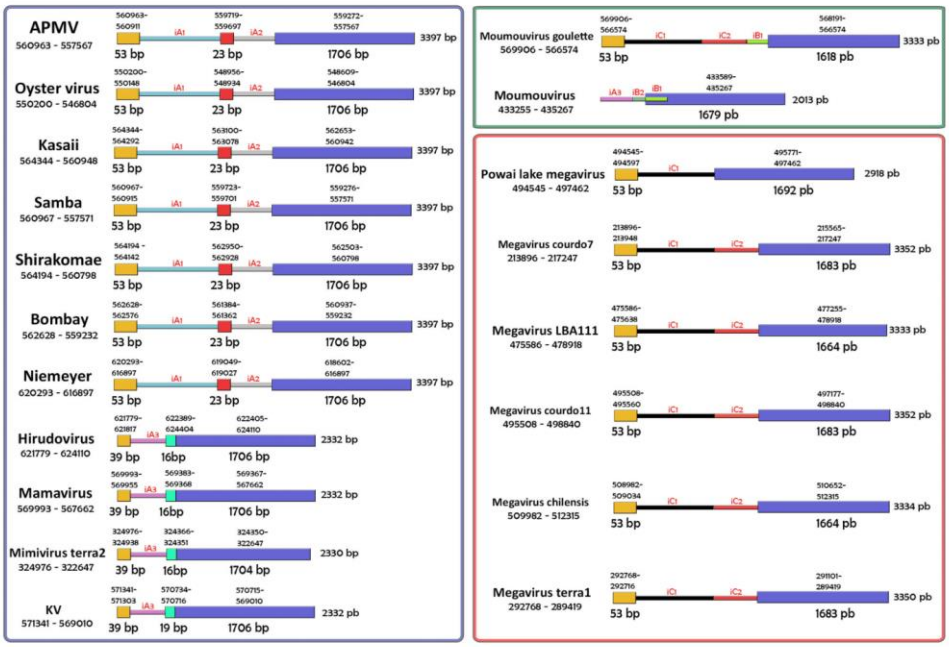
Figure 4. General scheme showing the structure of the MCP gene on different *Mimiviridae* members. This figure shows the structural layout of the MCP gene belonging to several members of the *Mimiviridae* family. The different *Mimiviridae* lineages were highlighted using different colored boxes (lineage A: blue, lineage B: green and lineage C: red). Representatives from these lineages seems to present distinct genetic structures for the gene. However, the general layout for this conserved gene may also suffer variations inside members of the same lineage, as it is observed for KV, Mamavirus, Hirudovirus and Mimivirus terra2 compared to other members of lineage A.

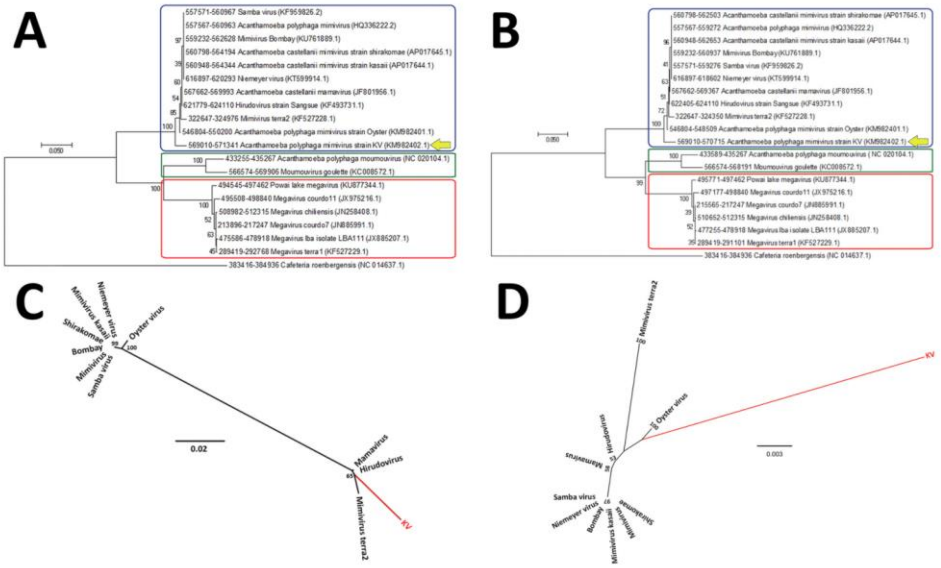
Figure 5. Phylogeny of the MCP gene for different *Mimiviridae* members. Here, a phylogenetic analysis for many representatives of the different *Mimiviridae* lineages is shown (A, B, and C), including for the new Brazilian mimivirus isolate KV strain, based on the nucleotide sequences of the whole MCP gene or for just the sequences corresponding to the *exon 3*. In (A), the rectangular tree layout represents an analysis of the whole gene, considering members of the three different lineages, highlighted by different colored boxes (lineage A: blue, lineage B: green and lineage C: red). In (B), the same members are considered, but the phylogenetic relationships were based on the nucleotide sequences of just the *exon* 3. Image (C) represents a radial tree, corresponding to analysis of the whole MCP gene for members of the lineage A alone. Finally, in (D), the phylogenetic relationships of members of lineage A are represented by another radial tree, but this time considering only sequences belonging to the *exon 3*. For these analysis, we have used MEGA7 software (maximum-likelihood method and

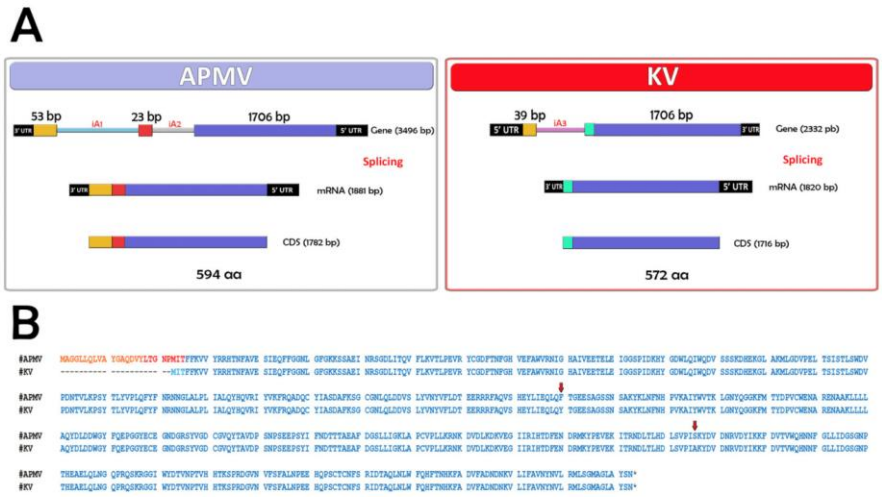
570 1,000 bootstrap replicates). For each sequence, the gene identification numbers are
571 indicated.
572
573 **Figure 6. General scheme representing the process of splicing for the MCP gene in**
574 **APMV and KV isolates.** (A) The APMV capsid transcript suffers a process of splicing
575 in which its processed form is composed by the sequences of *exon 1, exon 2* and *exon 3.*
576 KV, however, beyond its different MCP gene layout, uses as mature mRNA only the
577 regions corresponding to the exons *eX* and *e3.* (B) Alignment for the aminoacid
578 sequence produced by APMV and KV after translation of their respective MCP genes.
579 Arrow indicates mismatched amino acids.
580
581
582
583











frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 November 2015
doi: 10.3389/fmicb.2015.01256

Niemeyer Virus: A New Mimivirus Group A Isolate Harboring a Set of Duplicated Aminoacyl-tRNA Synthetase Genes

Paulo V. M. Boratto[1†], Thalita S. Arantes[1†], Lorena C. F. Silva[1], Felipe L. Assis[1], Erna G. Kroon[1], Bernard La Scola[2] and Jônatas S. Abrahão[1*]*

[1] Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [2] URMITE CNRS UMR 6236 – IRD 3R198, Aix Marseille Université, Marseille, France

It is well recognized that gene duplication/acquisition is a key factor for molecular evolution, being directly related to the emergence of new genetic variants. The importance of such phenomena can also be expanded to the viral world, with impacts on viral fitness and environmental adaptations. In this work we describe the isolation and characterization of Niemeyer virus, a new mimivirus isolate obtained from water samples of an urban lake in Brazil. Genomic data showed that Niemeyer harbors duplicated copies of three of its four aminoacyl-tRNA synthetase genes (cysteinyl, methionyl, and tyrosyl RS). Gene expression analysis showed that such duplications allowed significantly increased expression of methionyl and tyrosyl aaRS mRNA by Niemeyer in comparison to APMV. Remarkably, phylogenetic data revealed that Niemeyer duplicated gene pairs are different, each one clustering with a different group of mimivirus strains. Taken together, our results raise new questions about the origins and selective pressures involving events of aaRS gain and loss among mimiviruses.

Keywords: ***Mimiviridae*, Niemeyer virus, aminoacyl-tRNA synthetase, gene duplication, giant virus isolation**

OPEN ACCESS

Edited by:
Gilbert Greub,
University of Lausanne, Switzerland

Reviewed by:
Hendrik Huthoff,
King's College London, UK
Juliana Cortines,
Universidade Federal do Rio de Janeiro, Brazil

***Correspondence:**
Bernard La Scola
bernard.la-scola@univ-amu.fr;
Jônatas S. Abrahão
jonatas.abrahao@gmail.com

†*These authors have contributed equally to this work.*

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

Received: *03 September 2015*
Accepted: *29 October 2015*
Published: *10 November 2015*

Citation:
Boratto PVM, Arantes TS, Silva LCF, Assis FL, Kroon EG, La Scola B and Abrahão JS (2015) Niemeyer Virus: A New Mimivirus Group A Isolate Harboring a Set of Duplicated Aminoacyl-tRNA Synthetase Genes. Front. Microbiol. 6:1256. doi: 10.3389/fmicb.2015.01256

INTRODUCTION

Since the discovery of the first member of the family *Mimiviridae*, *Acanthamoeba polyphaga mimivirus* (APMV), in 2003, more mimivirus-like viruses are being isolated with increasing frequency from phagotrophic protists (La Scola et al., 2003). Mimivirus-like particles have been detected in the most diverse environments, such as rivers, soil, oceans, hospital and animals, and from different countries, such as France, Tunisia, Chile, Australia among others (La Scola et al., 2010; Arslan et al., 2011; Boughalmi et al., 2013). Recently, Campos et al. (2014) described the discovery of the first giant virus isolated in Brazil, named *Samba virus* (SMBV), which was isolated in 2011 from surface water collected from the Negro River, in the Amazon forest (Campos et al., 2014). SMBV is biologically and molecularly related to other mimiviruses, and was isolated in association with Rio Negro virus (RNV), a novel virophage strain belonging to this new class of viruses that parasitize the viral factory during mimivirus replication (Campos et al., 2014). Currently, the family *Mimiviridae* consists of dozens of mimivirus-like isolates that are able to infect amoeba of the genus *Acanthamoeba*. These viruses have been grouped into three distinct lineages, according to their polymerase B gene sequence and other genetic markers: lineage A (containing APMV), lineage B (containing *Acanthamoeba polyphaga moumouvirus*) and lineage

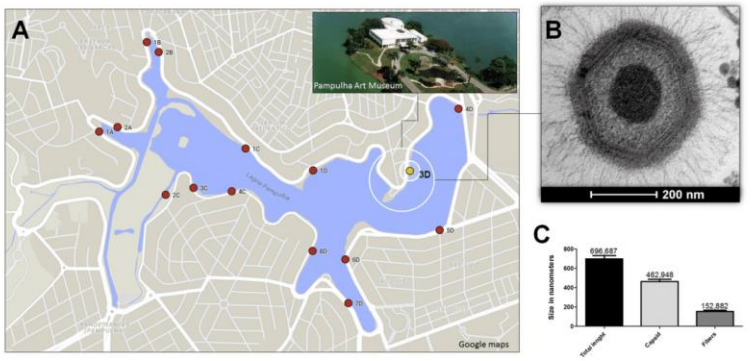


FIGURE 1 | Niemeyer virus (NYMV) collection site, morphometric analysis, and electron microscopy. (A) Shows a map of the Pampulha Lagoon, from where the samples were collected. Red dots: collection points; yellow dot: NYMV isolation site. **(B)** Shows an image of a NYMV particle, observed in electron microscopy. **(C)** Shows the size, in nanometers, of different components of the NYMV particle: only the fibers, only the capsid and the particle as a whole.

C (containing *Megavirus chilensis*) (Gaia et al., 2013). These amoeba-associated viruses have led to a paradigm shift in virus research due to their peculiar features which had never been seen in other viruses until then: large viral particles presenting a diameter of approximately 750 nm, covered by capsid associated fibers, and containing large double stranded DNA genomes of about 1.2 megabases (Mb), and approximately 1000 hypothetical proteins, many of them still uncharacterized or having functions never/rarely seen before in other viruses (La Scola et al., 2003; Raoult et al., 2004). Among the most intriguing predicted proteins in the genome of mimiviruses, it is worth highlighting those related to DNA repair and translation machinery, as well as chaperones related to DNA processing (La Scola et al., 2003; Raoult et al., 2004). Genes that encode translation related proteins, such as aminoacyl tRNA synthetases (aaRS) and translation factors, could hypothetically confer on APMV and other giant viruses a certain degree of autonomy from cellular machinery, and may be under conservative selection pressure (Raoult et al., 2004). Currently, seven aaRS have already been described among mimiviruses genomes: tyrosyl, cysteinyl, methionyl, arginyl, isoleucyl, asparaginyl, and tryptophanyl tRNA-synthetases. Among the aforementioned molecules, the first four enzymes are encoded by the genome of APMV, the prototype of the family *Mimiviridae*. However, no aaRS duplication events in the family *Mimiviridae* have been previously reported, other than in the exceptional case of the *Acanthamoeba polyphaga moumouvirus*, that possesses four orthologs of arginyl-tRNA synthetase in its genome (Yoosuf et al., 2012). In this work we describe the isolation and characterization of Niemeyer mimivirus, a new mimivirus-like virus isolated from water samples from an urban lake in Brazil. Genomic data show that Niemeyer harbors duplicated copies of three of its four aaRS genes (cysteinyl, methionyl, and tyrosyl aaRS), which are associated with increased expression of methionyl and tyrosyl aaRS mRNA by this virus in comparison to APMV. Remarkably, phylogenetic data revealed that Niemeyer duplicated genes are different from each other, each one clustering with a different group of mimivirus strains. Taken together, our results raise new questions about the origins and selective pressures involving aaRS gain and loss events among mimiviruses.

MATERIALS AND METHODS

Sample Collection and Virus Isolation

To explore the presence of giant viruses in an urban lake marked by a high concentration of organic matter, in 2011 we collected about 80 water samples, located at equidistant points, around Pampulha Lagoon (19°51′0.60″S and 43°58′18.90″W), in the city of Belo Horizonte, Brazil (**Figure 1A**). After collection, the samples were stored at 4°C overnight. Then, 500 µl of each sample was added to 4.5 mL of autoclaved rice and water medium made with 40 rice grains in 1 l of water. The samples were stored for 20 days in the dark at room temperature. Afterward, 5×10^3 *Acanthamoeba castellanii* trophozoites (ATCC 30234), kindly provided by the Laboratório de Amebíases (Departamento de Parasitologia, ICB/UFMG) were added, and the samples were re-incubated under the same conditions for 10 days (Dornas et al., 2014). After the enrichment process, samples were pooled in groups of five, and filtered through a 1.2 µm membrane to remove impurities, and a 0.2 µm membrane to retain giant viruses. The samples were then subjected in parallel to real-time PCR, targeting the RNA helicase gene (primers: 5′ACCTGATCCACATCCCATAACTAAA3′ and 5′GGCCTCATCAACAAATGGTTTCT3′) and to viral isolation from *A. castellanii*. As a control for the molecular and biological assays APMV was used. A new virus isolate, Niemeyer

TABLE 1 | Primers used for quantitative PCR.

Gene	Foward primer	Reverse primer
Leucyl-tRNA	GGGATTCGAACCCACGACAT	ATAAGCAAAGGTGGCGGAGT
Histidyl-lRNA	TTAGTGGTAGAACTACTGTTTGTGG	TTTTCAAAAATGACCCGTACAGGAA
Cysteinyl-tRNA	ACAGTCAAaGGATCGTTAGC	AGGATCGTATCAGAATTGAACTGA
Tryptophanyl-tRNA	GTG CAACAATAG ACCTGTTAGTTTA	ACCGGAATCGAACCAGTATCA
Methionyl tRNA synthetase	TGATTGGCGTGAATGGCTGA	ACCAATCACACTAGCCGGAA
Arginyl tRNA synthetase	GTGGGTGATTGGGGAAaCA	TGATACGGTCTCCAATCGGG
Tyrosyl tRNA synthetase	TTTGGCAAACCAATCGGCAA	TGGTTTTGAACCTAGTGGTCGT
Cysteinyl tRNA synthetase	TGCCAACCAGGTACACCAAA	TGCTCTTTGGAAAGGTCGATCA
18S rDNA	TCCAATTTTCTGCCACCGAA	ATCATTACCCTAGTCCTCGCGC
Viral RNA helicase	ACCTGATCCACATCCCATAAaAAA	GGCCTCATCAACAAATGGTTTCT

mimivirus, was grown and purified as described by La Scola et al. (2003) and Abrahão et al. (2014), respectively. The virus was called Niemeyer in tribute to the Brazilian architect who designed important buildings all over the world, including the Pampulha Art Museum near to where the virus was isolated (**Figure 1A**).

NYMV Virus Transmission Electron Microscopy

For the electron microscopy assays, *A. castellanii* cells were cultivated until 80–90% confluence was observed and infected with NYMV in a M.O.I of 0.01. Twelve hours post-infection (hpi), when approximately 50% of the trophozoites were presenting cytopathic effects, the medium was discarded and the monolayer gently washed twice with 0.1 M sodium phosphate buffer. Samples were fixed by adding glutaraldehyde (2.5% v/v) for 1 h at room temperature. The cells were then collected by centrifugation at 1500 *g* for 10 min, the medium was discarded and the cells were stored at 4° C until electron microscopy analysis was performed.

TABLE 2 | Best-hit analysis of proteins predicted in NYMV genome.

Compared strains	Best hits	Identity (%)	SD (%)	proteins
NYMV × APMV	One-way AAI 1	98,65	4,35	975
(group A)	One-way AAI 2	99,94	1,01	841
	Two-way AAI	99,99	0,08	836
NYMV × SMBV	One-way AAI 1	98,62	4,42	956
(group A)	One-way AAI 2	99,94	1,02	825
	Two-way AAI	99,99	0,08	820
NYMV × Moumou	One-way AAI 1	74,98	3,62	61
(group B)	One-way AAI 2	75,3	3,85	58
	Two-way AAI	75,21	3,7	53
NYMV × MCHV	One-way AAI 1	75,82	5,3	55
(group C)	One-way AAI 2	ND	ND	ND
	Two-way AAI	ND	ND	ND

NYMV, Niemeyer virus; APMV, Acanthamoeba polyphaga mimivirus; SMBV, Samba virus; MCHV, Megavirus chilensis; APMOUV, Acanthamoeba polyphaga moumouvirus; SD, standard deviation; AAI, average amino acid identity.

Evaluation of the Replication Profile of NYMV

Briefly, NYMV was inoculated in *A. castellanii* cells until appearance of cytopathic effect and purified by centrifugation on a 25% sucrose cushion as previously described (Abrahão et al., 2014). The titer was obtained by using the Reed–Muench method. To evaluate the replication profile of NYMV, the procedure was performed in 96-well Costar® microplates (Corning, NY, USA) containing 40,000 cells of *A. castellanii* maintained in 100 µl of PAS (Page's amoeba saline, PAS) culture medium per well. The cells were then infected with NYMV at a multiplicity of infection (M.O.I.) of 10. The cells were collected at different time points (0, 1, 2, 4, 8, and 24 hpi) and submitted to cell counting with a Neubauer chamber to evaluate the reduction of cells and the cytopathic effect. As a control for this experiment we used APMV, which was kept under the same conditions as NYMV.

Genome Sequencing and Annotation

The genome of NYMV was sequenced using the Illumina MiSeq instrument (Illumina Inc., San Diego, CA, USA) with the paired-end application. The sequenced reads were imported to CLC_Bio software[1] and assembled into contigs by the *de novo* method. The prediction of open reading frame (ORF) sequences was carried out using the FgenesV tool. ORFs smaller than 100aa were excluded from the annotation. Paralogous groups of genes were predicted by OrthoMCL program. The ORFs were functionally annotated using similarity analyses with sequences in the NCBI database using BLAST tools. In addition, the presence of trademark genes of the family *Mimiviridae* was evaluated, and some of them were analyzed in detail. Genbank number: KT599914.

Similarity Analysis

Viruses of the genus *Mimivirus* are divided into groups A to C. Thereby, the ORFs predicted in NYMV genome were compared to amino acid sequences available in Genebank of APMV (group A), APMOUV (Group B), and MCHV (group C), as well as sequences from SMBV (group A), a Brazilian

[1]http://www.clcbio.com/index.php?id=28

TABLE 3 | Distribution of aminoacyl-tRNA synthetases in mimiviral genomes.

Aminoacyl-tRNA sintetase	Mimivirus strains								
	NYMV	APMV	Mamavirus	SMBV	Hirudovirus	APMV-M4	MCHV	APMOUV	Terra1 virus
Tyrosyl	2	1	1	1	1	0	1	1	1
Cysteinyl	2	1	1	1	1	1	1	1	1
Methionyl	2	1	1	1	1	1	1	1	1
Argiryl	1	1	1	1	1	1	1	4	1
Isoleucyl	0	0	0	0	0	0	1	1	0
Asparaginyl	0	0	0	0	0	0	1	0	0
Tryptophanyl	0	0	0	0	0	0	1	0	0

NYMV, Niemeyer virus; APMV, Acanthamoeba polyphaga mimivirus; SMBV, Samba virus; MCHV, Megavirus chilensis; APMOUV, Acanthamoeba polyphaga moumouvirus.

mimivirus isolate. The AAI calculator program[2] Rodriguez-R and Konstantinos (2014) was used to estimate the average amino acid identity between two protein datasets using both best hits (one-way AAI) and reciprocal best hits (two-way AAI). The similarity and score thresholds for the alignments were 70 and 0%, respectively. A minimum alignment of 50% was considered.

Expression of Translation-related Genes

In order to check the expression of aaRS by NYMV, we selected four genes based on the APMV genome sequence (methionyl, tyrosyl, cysteinyl, and arginyl tRNA synthetases) to evaluate the expression profile of NYMV in comparison with APMV as previously described by Silva et al. (2015). Twenty-four-well plates containing 1×10^5 amoeba per well, kept in PAS medium were infected with NYMV and APMV at M.O.I. 10 and incubated at 32°C for 8 h. Cells were collected, centrifuged and the pellet used for total RNA extraction, reverse transcription and quantitative PCR. Briefly, total RNA was extracted using the RNeasy kit (Qiagen, Germany), and reverse transcription was performed by using the MMLV reverse transcriptase (Promega, USA), as recommended by the manufacturers. The cDNA was used to determine the levels of aaRS mRNA by quantitative PCR (primers in **Table 1**) by using specific primers, SYBR Green Master Mix (Applied Biosystem, USA) and water in 10 μL reactions. Reactions were carried out in a StepOne instrument (Applied Biosystem, USA). All reactions had been previously optimized and presented high efficiency values. Relative gene expression analyses were performed using the ΔΔCt method and normalized to the expression of 18S ribosomal RNA (18S rDNA) and the viral RNA helicase mRNA and calibrated using the lower value (=1). Statistical analysis and primers sequences were as described by Silva et al. (2015).

Phylogeny

The β-DNA polymerase sequence of NYMV was aligned with sequences from other giant viruses, previously deposited in GenBank, using the ClustalW program. After the alignment analysis, phylogeny reconstruction was performed using the maximum likelihood method implemented by the MEGA5 software. Additionally, sequences of aaRS predicted in the genome of NYMV were aligned with sequences from other giant viruses among GenBank sequences as described above, and the phylogeny reconstruction was performed using the neighbor-joining method in the MEGA5 software.

[2]http://enve-omics.ce.gatech.edu/aai/

RESULTS

Niemeyer: A New Mimivirus Group A Isolate

Here we report the isolation of Niemeyer virus (NYMV) from water samples collected in an urban lake in Brazil. The isolation was confirmed by the observation of a cytopathic effect in cells of *A. castellanii* (ATCC 30234) after 4 days of incubation and also by viral RNA helicase gene amplification in qPCR assays, a highly conserved gene amongst mimiviruses.

Electron microscopy and morphometric assays showed virus particles with average size of 616 nm in total, with fibers of about 153 nm and a capsid size of about 463 nm (**Figures 1B,C**), similar to the dimensions described for other mimivirus-like viruses. Large viral factories were observed in the amoebic cytoplasm, and these contained viral particles at distinct steps of morphogenesis. In addition, NYMV demonstrated a similar pattern of replication to APMV in one-step-growth curve assays (**Figure 5A**).

NYMV Genome Analyses and Phylogeny

The final genome assembly of NYMV yielded 69 contigs, consisting of fifteen small contigs (<2,000 bp) and fifty-four large contigs (>2,000 bp), including five contigs larger than 100 kb. The NYMV genome is a double-stranded DNA molecule composed of approximately 1,299,140 base pairs. This genome presented a mean C–G content of 27.96%, which is similar to that of other mimiviruses. A total of 1003 proteins were predicted, ranging in size from 100 to 2156 amino acids, with a mean size of 379 amino acids. Moreover, we identified 970 proteins with high similarity (coverage > 90%; identity > 80%; *e*-value < 10e-5) to mimivirus sequences available in the non-redundant NCBI protein database, as well as 27 proteins with

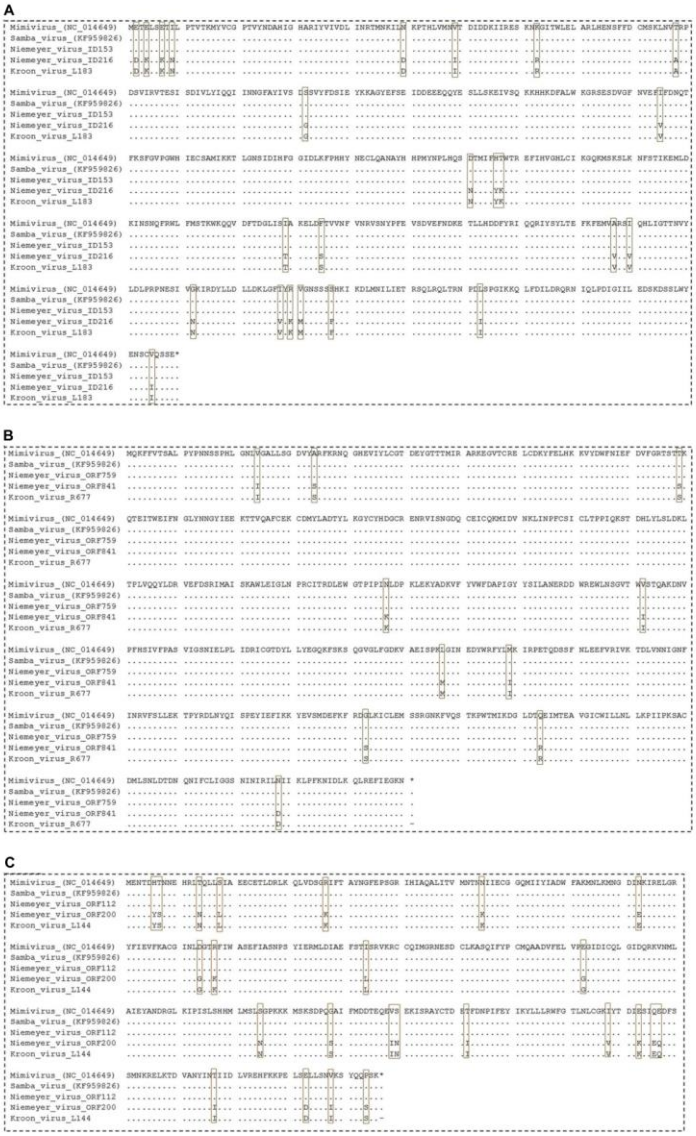


FIGURE 2 | Alignment of NYMV aaRs sequences. (A) Cysteinyl-tRNA synthetase, **(B)** methionyl-tRNA synthetase, and **(C)** tyrosyl-tRNA synthetase sequences. In this analysis sequences from *Acanthamoeba polyphaga mimivirus* (APMV), the prototype of *Mimivirus,* and sequences from Samba virus (SMBV) and Kroon virus (KROV) were used. Brown boxes highlight polymorphisms in the alignments.

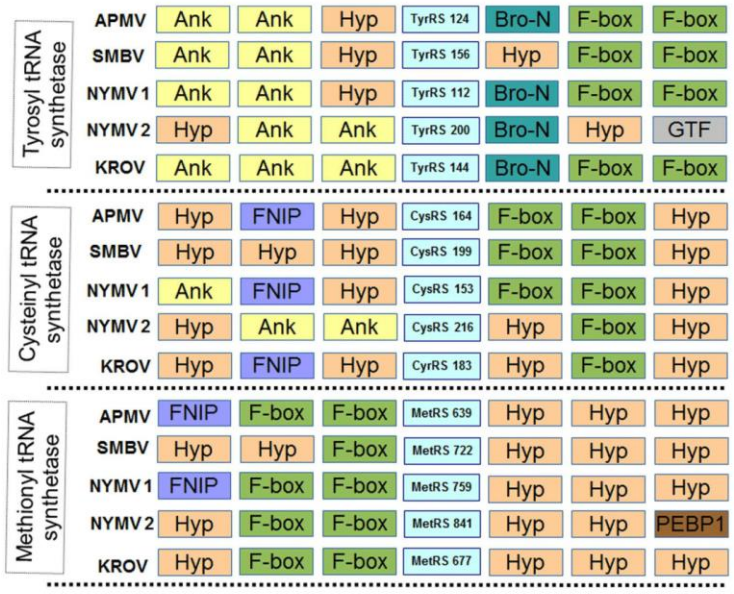


FIGURE 3 | Neighboring gene synteny analyses of aminoacyl-tRNA synthetase. Three upstream and downstream neighbor genes of each aaRS predicted in the genome of NYMV and other mimivirus were checked. Ank, ankyrin repeat family protein; Hyp, Hypothetical Protein; F-box, F-box protein family; Bro-n, bro-n family protein; GTF, glycosyltransferase; FNIP, FNIP repeat-containing protein; PEBP1, phosphatidylethanolamine-binding protein. NYMV 1 and 2 refer to each copy of the aaRS genes.

low similarity (coverage 50%; identity 35–80%; *e*-value < 10e-5) to mimivirus sequences. We identified two proteins with higher similarity to non-viral sequences (ID 80: Ankyrin repeat protein – *Trichomonas vaginalis* [coverage: 83%; identity: 29%; *e*-value: 5e-5]; ID 193: Hypothetical protein – *Volvox carteri* [coverage: 99%; Identity: 36%; *e*-value: 8e-12]). Indeed, four putative proteins of NYMV had no significant hit (*e*-value threshold 1e-2) against the NCBI non-redundant sequence database. A total of 90 clusters consisting of 269 paralogous proteins were identified in the NYMV genome, which is a remarkably higher number than that described for other mimiviruses. It is worth mentioning that neither virophage sequences nor other mobilome elements were detected in the analyzed data set.

A comparative analysis of NYMV gene content with other mimivirus sequences was performed, which showed the highest identity and bit-score distributions against mimivirus group A sequences, such as SMBV (**Supplementary Figure S1A**) and mimivirus (**Supplementary Figure S1B**). Moreover, the similarity decreased toward moumouvirus (**Supplementary Figure S1C**) and *Megavirus chilensis* (**Supplementary Figure S1D**) of groups B and C, respectively. Furthermore, the one-way and two-way best hit analysis (**Table 2**) corroborated the previous observations. This analysis showed a two-way similarity higher than 99% for NYMV with both SMBV and APMV, reinforcing its grouping with other mimiviruses in group A. During functional annotation, we identified important proteins required for virus replication: DNA polymerase, helicases, nucleases, and proteins with DNA polymerase sliding clamp activity related to replication processes; resolvases and topoisomerases related to DNA manipulation and processing; transcription and translation factors; and ATPases for DNA packaging. However, no chaperone molecules were detected in the NYMV genome, as has been described elsewhere (Yutin and Koonin, 2009, Yutin et al., 2009). Furthermore, we identified four regions encoding tRNA molecules for leucine (two sequences), histidine and cysteine amino acids. However, unlike other mimivirus genomes, no tryptophan tRNA gene was detected.

NYMV aaRS Analyses

In addition, we evaluated the presence of the landmark aaRS in NYMV genome. The analysis of such proteins is particularly interesting due to the fact that no virus outside the family *Mimiviridae* has been predicted to encode them. We identified seven aaRS sequences in NYMV genome, being two orthologs

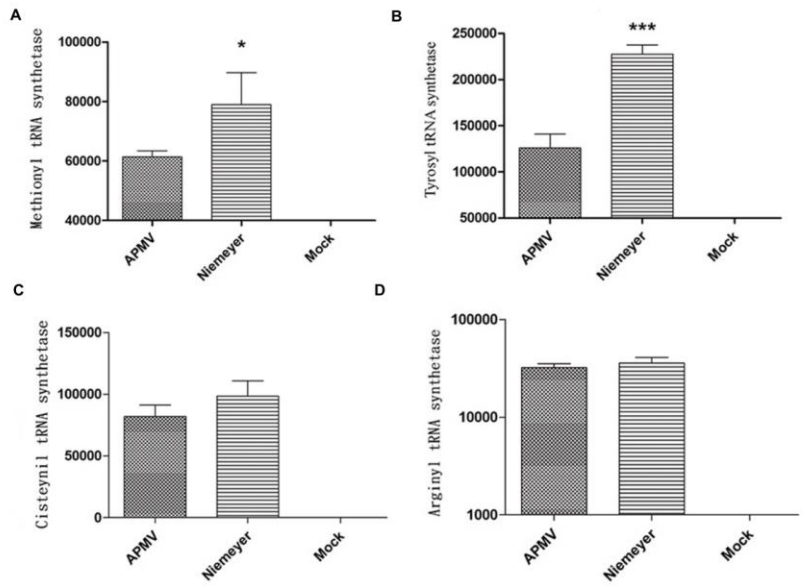


FIGURE 4 | Aminoacyl-tRNA-synthetase (aaRS) mRNA expression by NYMV. (A) Methionyl tRNA synthetase, **(B)** tyrosyl tRNA synthetase, **(C)** cysteinyl tRNA synthetase, and **(D)** arginyl tRNA synthetase. Relative gene expression analyses were performed using the $\Delta\Delta$Ct method and normalized to the expression of 18S ribosomal RNA [18S rDNA) and the viral RNA helicase mRNA (and calibrated with the lower value (=1)]. The values were subjected in different combinations to one-way ANOVA tests and Bonferroni post-tests (95% confidence intervals). Differences between groups were considered significant when the *p*-values were smaller than 0.05 (asterisks). *Y* axis represents relative abundance.

each for tyrosyl, cysteinyl, and methionyl-tRNA synthetases, as well as one sequence for arginyl-tRNA synthetase. The NYMV genome encoded similar aaRS molecules as those detected in the APMV, Mamavirus, SMBV, and Hirudovirus genomes (mimivirus group A), except for Terra1 virus (mimivirus group C). However, no aaRS duplications were observed in such genomes. The aaRS distribution in other mimivirus genomes, including NYMV, can be seen in **Table 3**. The arginyl-tRNA synthetase encoded by NYMV presented 100% identity (amino acid) with the sequences of Mimivirus and SMBV, whereas the duplicated aaRS of NYMV presented some polymorphisms when paralogs were compared with each other (**Figure 2**). Curiously, one copy of each NYMV duplicated aaRS presented 100% identity with APMV and SMBV, while the other copy had 100% identity with Kroon virus (KROV; **Figure 2**), a Brazilian mimivirus-like virus strain isolated from a water sample collected in an urban lake, at Lagoa Santa city, Brazil (approximately 30 km from Pampulha lagoon). The biological and molecular characterization of KROV is in progress, but preliminary results suggest a possible dichotomy among Brazilian mimivirus A isolates. To evaluate the distribution of duplicated aaRS genes within NYMV and other mimivirus genomes, three upstream and downstream genes neighboring aaRS from APMV, SBMV, KROV, and NYMV were analyzed (**Figure 3**). We observed that the duplicated aaRS in the genome of NYMV are not in tandemly duplicated, being located distant from each other (**Figure 3**). The methionyl-tRNA loci presented the best neighbor gene synteny, as all virus strains had two neighboring genes the same on both sides, with the exception of the SMBV strain, which had a distinct neighbor gene at second position toward the 3′ extremity. Altogether, no conservative genomic loci were observed for any aaRS, and none of the analyzed virus strains shared the same neighboring gene for all of the aaRS analyzed, reinforcing the uniqueness of each isolate (**Figure 3**).

Expression of *aaRS* Genes and NYMV Replication Profile

To evaluate the expression profile of NYMV-encoded aaRS during infection, infected *A. castellanii* cells were collected, processed and assayed by Real-time PCR. The results of quantitative PCR were expressed as arbitrary units, fitted to standard curves generated for each target gene and normalized by amoebal 18S rDNA gene levels. Our results

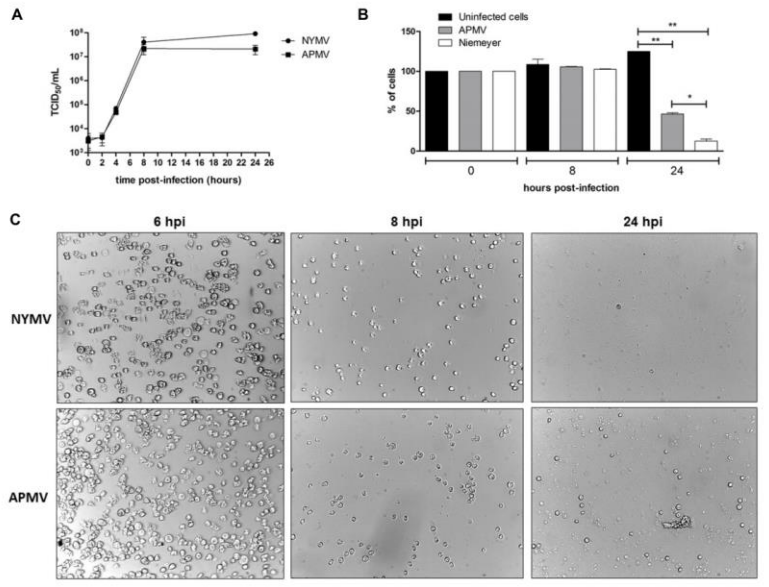


FIGURE 5 | (A) One step growth curves of NYMV and APMV in *Acanthamoeba castellanii*. At 6 hpi cells are similar negative control cell (data not shown). **(B)** *A. castellanii* cells counting at different time points (0, 8, and 24 hpi with APMV or NYMV). **(C)** Images obtained from cells of *A. castellanii* infected with NYMV and APMV at different time points after infection.

revealed that the expression of two aaRS, methionyl-RS, and tyrosyl-RS (both duplicated in NYMV), was significantly distinct between APMV and NYMV ($p < 0.001$ or $p < 0.01$; **Figures 4A,B**), while for the other two analyzed genes, cysteinyl-RS, and arginyl-RS, there was no significant difference in expression between the two viruses (**Figures 4C,D**). Evaluation of the replication profile of NYMV showed that up to 4 hpi, NYMV and APMV showed a similar replication profile. After 8 hpi it was possible to notice an increased lysis of cells infected with NYMV when compared to APMV. After 24 hpi, the lysis induced by NYMV is greater than that induced in the amoebae infected with APMV (**Figure 5**).

NYMV Phylogenetic Analysis

β-DNA polymerase-based phylogenetic analysis corroborated all of the previous observations, clustering the mimivirus NYMV strain with members of *Megavirales* order group A, which includes APMV (the prototype of the family), mamavirus, and other Brazilian mimivirus isolates, such as the SMBV and KROV strains (**Figure 6**). Additionally, we performed aaRS-based phylogenetic analyses of NYMV. The arginyl-tRNA synthetase-based tree (**Figure 7A**) grouped the NYMV within the APMV, SMBV, and Mamavirus branch, with KROV being positioned more distant from the other mimiviruses of group A. The subsequent phylogenetic trees (**Figures 7B–D**), based on duplicated aaRS, presented a peculiar feature, in which one of each doublet grouping within the mimivirus group A branch, and the other copy grouping more distantly with the sequences of the KROV (**Figures 7B–D**).

DISCUSSION

In this work we describe the isolation of NYMV, a new mimivirus group A isolate from an aquatic habitat marked by a high concentration of organic matter, the eutrophicated urban lake Pampulha Lagoon (**Figure 1A**). NYMV presented some features similar to other mimiviruses such as viral particles of approximately 696 nm and the presence of fibers around the capsid; a replication profile similar to APMV; and gene content/similarity resembling viruses belonging to the genus *Mimivirus*. However, as demonstrated by genomic data and gene expression analysis, NYMV harbors duplicated copies of three of its four aaRS genes, which may be associated with an increased expression of methionyl and tyrosyl aaRS mRNA.

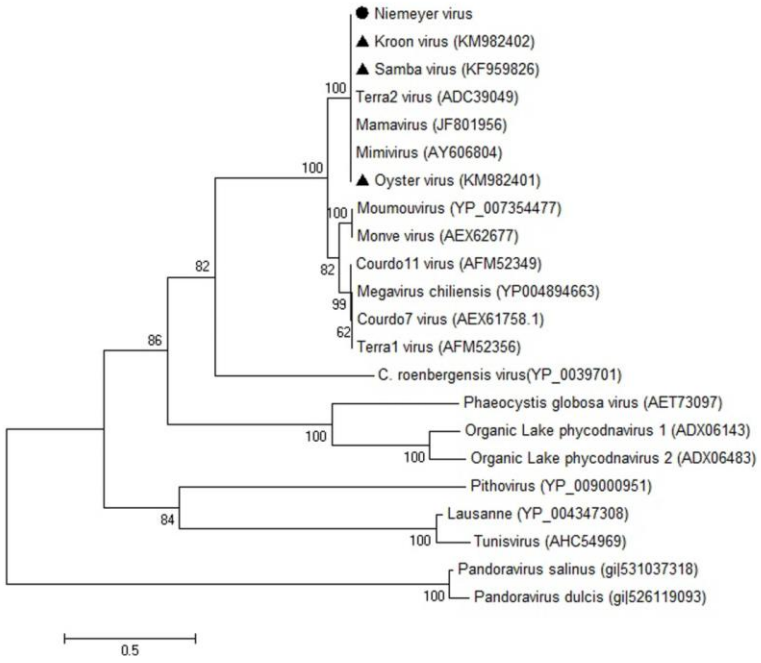


FIGURE 6 | Phylogenetic reconstruction of mimiviruses including the Brazilian Niemeyer mimivirus strain, based on the Polymerase-B amino acid sequences, using MEGA5 software (Maximum-likelihood – 1,000 bootstrap replicates). The NYMV (highlighted by black dot) clustered with other Brazilian mimivirus isolates (triangles), as well as other members of mimivirus lineage A. For each sequence, the gene identification number is indicated.

Gene duplication is an important process driving molecular evolution, allowing the formation of new genes with new or redundant biological functions, affecting the evolutionary history and/or the fitness of the organisms. This process is well described for several species in the different domains of life, especially for the eukaryotes (Zhang, 2003). Simon-Loriere et al. (2013) showed, by using comparative analysis among 55 species of RNA viruses from humans, animals, and plants, distributed across 19 viral families and 30 genera, that genetic duplication seems to have only a modest role in the evolutionary history of RNA viruses (Simon-Loriere et al., 2013). Nonetheless, this process has been described quite commonly for DNA viruses (Shackelton and Holmes, 2004). In viruses belonging to the family *Mimiviridae*, gene duplication events are an open field for new studies. It has already been shown that these processes were very important in shaping the APMV genome during its evolutionary history, with about of one-third of the viral genes having at least one other related gene in the same genome (Suhre, 2005).

Despite this, the duplication of aaRS genes related to protein translation, seems not to be a very frequent event in giant viruses, as demonstrated in **Table 3**, in which only two known viruses present duplications, *Acanthamoeba polyphaga moumouvirus* (Yoosuf et al., 2012) and NYMV. This feature could have brought important adaptive advantages for both viruses, aiding the protection against deleterious mutations in the gene, allowing the emergence of novel mimivirus-like strains during evolutionary history and even potentially endowing the capacity of infecting a larger host range. For example, in the case of the methionyl-tRNA synthetase, duplications of the gene could have brought an important evolutionary advantage since its cognate amino acid is essential for the initiation of protein synthesis. For cysteinyl-tRNA synthetase, a high level of conservation is observed among the mimiviruses in which the genomes have already been described. This could have been important in facilitating the occurrence of an event of genetic duplication. Finally, the gene duplication of tyrosyl-tRNA synthetase could represent an important advantage for NYMV in the environment since its cognate amino acid ranks

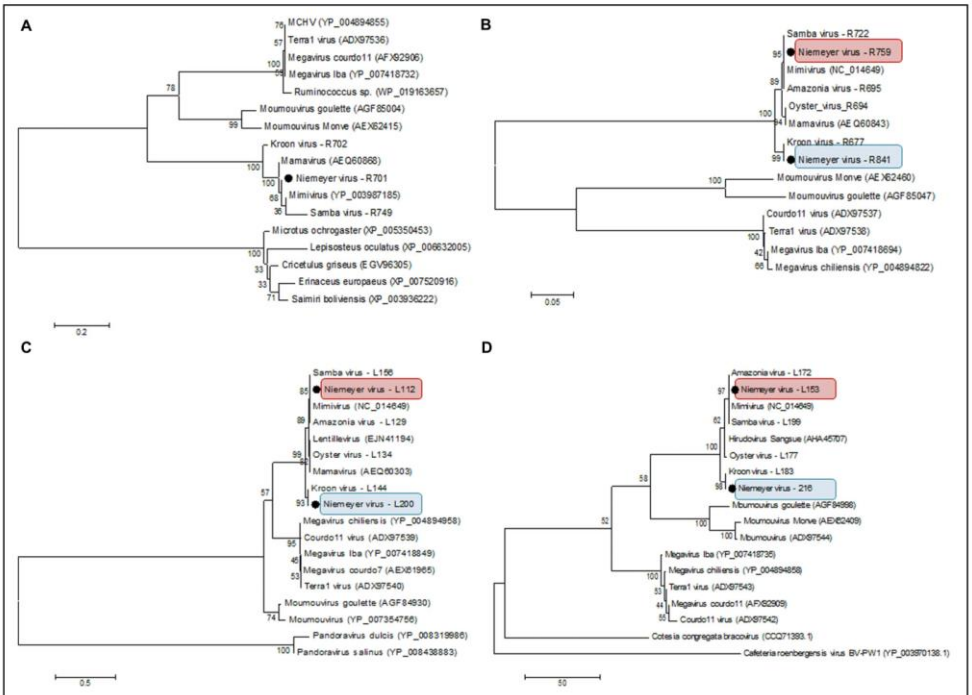


FIGURE 7 | Phylogenetic analysis based on sequences of (A) Arginyl-tRNA synthetase, (B) Methionyl-tRNA synthetase, (C) Tyrosyl-tRNA synthetase, and (D) Cysteinyl-tRNA synthetase predicted in the genome of NYMV, and orthologous sequences obtained from the NCBI database. The trees were inferred by using MEGA5 software (Neighbor-joining – 1,000 bootstrap replicates). The aaRS encoded by NYMV are highlighted by black dots. For each sequence, the gene identification number is indicated. Red and blue boxes indicates the different copies of NYMV aaRS.

highly in the composition of the amino acid usage profile of mimiviruses (Silva et al., 2015). Beyond that, it was demonstrated in another study that apart from the conserved domains presented by aaRSs, which are involved in the process of aminoacylation, these enzymes might also incorporate novel motifs related to new biological functions in different eukaryotic organisms, for example functions related to angiogenic activity, angiostatic activity, and inflammatory response (Guo et al., 2010). Considering this case, the presence of duplicated aaRSs may represent a potential role for the addition of important new biological functions in giant viruses.

Another interesting fact was that from each duplicated copy of aaRS present in NYMV, one gene showed 100% identity with the corresponding APMV and SMBV gene, and the second copy also presented a 100% identity with a giant virus called KROV, isolated by our group from a different urban lake. This result may suggest an event of gene transfer among the ancestors of these viruses (Kroon vs. NYMV and or APMV-like virus and NYMV) in the same host at some moment in their evolutionary history.

Furthermore, analysis of the expression of aaRS genes in NYMV showed that methionyl (duplicated) and tyrosyl-RS (duplicated) mRNA expression was significantly higher in cells infected with NYMV in comparison with APMV (**Figures 4A,B**), while for arginyl-RS (unduplicated) and cysteinyl-RS (duplicated), there was no significant difference in the expression (**Figures 4C,D**). Given the sequence similarity observed for the promoters of all of the aaRs genes (**Supplementary Figure S2**), we believe that this differential expression may be due to the presence of the duplicated genes in NYMV in comparison with APMV. This feature could give NYMV an advantage during its replication cycle into the host, increasing in the production of its own proteins during the process of translation and could be the cause of the faster growth of NYMV in comparison to APMV (**Figure 5**). The reason that cysteinyl RS NYMV duplication does not result in significant gene expression needs to be investigated, but it could be a result of some gene specificity regarding amoebal growth conditions (Silva et al., 2015) and/or virus host range.

The results obtained in this work suggest the importance of gene duplication events during the evolutionary history of the aaRSs in mimiviruses. These translation related genes seems to present a considerable influence during replication of the giant viruses in amoebae. This theme then becomes extremely interesting for future studies trying to understand the origin, genetic exchange and evolution of the aaRSs among mimiviruses, as well as the role of these processes in the acquisition of evolutionary advantages by the giant viruses.

AUTHOR CONTRIBUTIONS

PB, TA, LS, FA performed experiments and wrote the paper.

ACKNOWLEDGMENTS

We would like to thank colleagues from Gepvig, Laboratório de Vírus, and Aix Marseille Université for their excellent technical support. Also would like to thank Pro-reitoria de pesquisa da Universidade Federal de Minas Gerais, CAPES, FAPEMIG, CNPq and Marseille Université.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2015.01256

FIGURE S1 | Comparative analysis of Niemeyer virus (NYMV) gene content with other mimivirus sequences. The analysis showed the highest identity and bit-score distributions of NYMV against mimivirus group A sequences, such as Samba virus (SMBV) **(A)** and *Acanthamoeba polyphaga mimivirus* (APMV) **(B)**. Moreover, the similarity decreased toward moumouvirus **(C)** and Megavirus chilensis **(D)** of groups B and C, respectively. At the top of figures are mean and median of each compared group, considering identity and bit score distribution.

FIGURE S2 | Promoter regions analysis of NYMV AaRS. Red arrows indicate the promoter regions. Blue arrows indicate the start-codon regions.

REFERENCES

Abrahão, J. S., Boratto, P. V. M., Dornas, F. P., Silva, L. C., Campos, R. K., Almeida, G. M., et al. (2014). Growing a giant: evaluation of the virological parameters for mimivirus production. *J. Virol. Methods* 207, 6–11. doi: 10.1016/j.jviromet.2014.06.001

Arslan, D., Legendre, M., Seltzer, V., Abergel, C., and 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. doi: 10.1073/pnas.1110889108

Boughalmi, M., Saadi, H., Pagnier, I., Colson, P., Fournous, G., Raoult, D., et al. (2013). High-throughput isolation of giant viruses of the Mimiviridae and Marseilleviridae families in the Tunisian environment. *Environ. Microbiol.* 15, 2000–2007. doi: 10.1111/1462-2920.12068

Campos, R. K., Boratto, P. V., Assis, F. L., Aguiar, E. R., Silva, L. C., Albarnaz, J. D., et al. (2014). Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. *Virol. J.* 11, 95. doi: 10.1186/1743-422X-11-95

Dornas, F. P., Silva, L. C., de Almeida, G. M., Campos, R. K., Boratto, P. V., Franco-Luiz, A. P., et al. (2014). *Acanthamoeba polyphaga mimivirus* stability in environmental and clinical substrates: implications for virus detection and isolation. *PLoS ONE* 9:e87811. doi: 10.1371/journal.pone.0087811

Gaia, M., Pagnier, I., Campocasso, A., Fournous, G., Raoult, D., and La Scola, B. (2013). Broad spectrum of mimiviridae virophage allows its isolation using a mimivirus reporter. *PLoS ONE* 8:e61912. doi: 10.1371/journal.pone.0061912

Guo, M., Yang, X. L., and Schimmel, P. (2010). New functions of aminoacyl-tRNA synthetases beyond translation. *Nat. Rev. Mol. Cell Biol.* 11, 668–674. doi: 10.1038/nrm2956

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

La Scola, B., Campocasso, A., N'Dong, R., Fournous, G., Barrassi, L., Flaudrops, C., et al. (2010). Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. *Intervirology* 53, 344–353. doi: 10.1159/000312919

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., et al. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344–1350. doi: 10.1126/science.1101485

Rodriguez-R, L. M., and Konstantinos, T. K. (2014). Bypassing cultivation to identify bacterial species – culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species. *Microbe* 9, 111–118.

Shackelton, L. A., and Holmes, E. C. (2004). The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol.* 12, 458–465. doi: 10.1016/j.tim.2004.08.005

Silva, L. C., Almeida, G. M., Assis, F. L., Albarnaz, J. D., Boratto, V. M., Boratto, D., et al. (2015). Modulation of the expression of mimivirus-encoded translation-related genes in response to nutrient availability during *Acanthamoeba castellanii* infection. *Front. Microbiol.* 6:539. doi: 10.3389/fmicb.2015.00539

Simon-Loriere, E., Holmes, E. C., and Pagán, I. (2013). The effect of gene overlapping on the rate of RNA virus evolution. *Mol. Biol. Evol.* 30, 1916–1928. doi: 10.1093/molbev/mst094

Suhre, K. (2005). Gene and genome duplication in *Acanthamoeba polyphaga Mimivirus*. *J. Virol.* 79, 14095–14101. doi: 10.1128/JVI.79.22.14095-14101.2005

Yoosuf, N., Yutin, N., Colson, P., Shabalina, S. A., Pagnier, I., Robert, C., 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. doi: 10.1093/gbe/evs109

Yutin, N., and Koonin, E. V. (2009). Evolution of DNA ligases of nucleo-cytoplasmic large DNA viruses of eukaryotes: a case of hidden complexity. *Biol. Direct.* 4, 51. doi: 10.1186/1745-6150-4-51

Yutin, N., Wolf, Y. I., Raoult, D., and Koonin, E. V. (2009). Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* 6. 223. doi: 10.1186/1743-422X-6-223

Zhang, J. (2003). Evolution by gene duplication: a update. *Trends Ecol. Evol.* 18, 292–298.

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.

Arch Virol
DOI 10.1007/s00705-017-3455-5

RAPID COMMUNICATION

Detection of mimivirus genome and neutralizing antibodies in humans from Brazil

Fábio P. Dornas[1] · Paulo V. M. Boratto[1] · Galileu B. Costa[1] · Lorena C. F. Silva[1] · Erna G. Kroon[1] · Bernard La Scola[2] · Giliane Trindade[1] · Jônatas S. Abrahão[1]

Received: 31 March 2017 / Accepted: 15 May 2017

Abstract In recent years, giant viruses belonging to the family *Mimiviridae* have been proposed to be infectious agents in humans. In this work we provide evidence of mimivirus genome and neutralizing antibodies detection in humans.

The proposed order *Megavirales* comprises large and giant viruses that infect hosts from different taxa [1]. Recently, many species or virus strains from this order infecting free-living amoebas from the genus *Acanthamoeba* (FLAA) have been discovered. Although little is known about infections in other hosts [1] the ubiquitous nature of FLAA and their association with pathogenic microorganisms enhances the idea that diverse environments could shelter these intriguing giant viruses [1, 5, 7–9].

In recent years, giant viruses belonging to the family *Mimiviridae* have been proposed to be infectious agents in humans. This hypothesis has become increasingly strong since related studies have described these viruses being isolated from bronchoalveolar lavage and stool samples from humans with pneumonia, as well as numerous serological, immunological and genome detection human studies [1, 2, 5, 7–9]. In addition, viruses from the family *Marseilleviridae* (a close-related family of the *Mimiviridae*) have been described in the serum of a child with adenitis and also in the blood and feces from apparently healthy humans [1, 2].

We and others have detected mimiviruses in wild and domestic animals, including *Bos Taurus* [4, 6]. Although the veterinary relevance of these isolations needs to be investigated, we believe that humans who handle bovines occupationally may be at risk of mimivirus exposure. Considering this possibility, our group investigated the circulation of mimiviruses in the serum of humans from different rural areas of Minas Gerais State, Brazil (n=285) (Table 1). The first collection of serum samples (Collection 1) (n=240) was originally performed with asymptomatic humans, with the aim being to study the seroprevalence of poxviruses in the city of Serro. These serum samples were collected in 7 sub-areas, covering a total of 43 farms and nearby residences. The second collection of serum samples (Collection 2) (n=45) was performed by State Agrodefense offices that sent serum samples from other cities including Mantena, Jequeri and Rio Pomba, with the aim also being the detection of poxvirus circulation. To streamline the detection of mimiviruses in these human sera the specimens were grouped into pools of 4 to 5 samples (20-25 µL for each sample). Each pool belonged to individuals from the same region and was performed by crossover analysis, i.e. duplicate samples were analyzed in two different pools, establishing a quadruplicate (Table 1). To investigate mimivirus circulation in these sera samples all the pools were subjected to a real-time PCR assay targeting the conserved helicase viral gene following DNA extractions, as reported by Dornas et al. [4].

Thirty of the 117 pools (25.6%) were positive for mimivirus in the PCR assays, including 22 (22.5%) from

Electronic supplementary material The online version of this article (doi:10.1007/s00705-017-3455-5) contains supplementary material, which is available to authorized users.

✉ Jônatas S. Abrahão
jonatas.abrahao@gmail.com

[1] Laboratório de Vírus, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Av. Antonio Carlos, Pampulha, Belo Horizonte, MG 6627, Brazil

[2] URMITE CNRS, UMR 6236, IRD 3R198, Aix Marseille Universite, Marseille, France

Published online: 30 June 2017

Table 1 Real time PCR and virus neutralization test for mimivirus in serum samples taken from rural-dwelling humans

Serum Collection (MG)	Total (serum)	Total (pools)	Real-Time PCR (Helicase gene)		
			Negative pools (%)	Positive pools (%)	VN* > 90%(%)**
Collection 1	240	98	76 (77.5)	22 (22.5)	20 (90.9)
Collection 2	45	19	11 (57.9)	8 (42.1)	6 (75)
TOTAL	285	117	87 (74.4)	30 (25.6)	26 (86.7)

* VN = Virus neutralization test

** Percentage calculated based on PCR-positive samples. Brazilian State: MG: Minas Gerais

collection 1 and 8 (42.1%) from collection 2 (Table 1). A total of 10 positive pooled serum samples were randomly chosen for helicase gene sequencing and analysis (5 from each area), followed by multiple alignment and phylogenetic tree analysis [4]. A neighbor joining phylogenetic tree based on the helicase gene sequences (amplified from pooled sera) revealed the presence of viruses that clustered within mimivirus lineage A isolates (see Supplemental Figure). The sequences were deposited in GenBank (KT595674-KT595683).

Concomitantly with molecular analysis, all of the qPCR-positive pools and some randomly selected qPCR-negative pools were analysed using a virus neutralization test (VN) [3] with a serum dilution of 1:20, our aim being the detection of mimivirus neutralizing antibodies, as has been previously reported by our team [4]. Afterwards, VN-positive pools had their neutralizing antibodies titrated in *A. castellannii* cells using the endpoint method with titers being calculated using the $TCID_{50}$ method. The percentage reduction was calculated, and the cutoff for positive serum pools was defined as a 90% reduction, in comparison with the negative control. The VN results showed that 26 of the 30 PCR-positive pools contained neutralizing antibodies against mimivirus, with 20 (90.9%) being from collection 1 and 6 (75%) from collection 2, while 12 PCR-negative pools were also negative for VN (Table 1).

Here, we have described evidence of mimivirus circulation in humans from rural areas of Brazil. Many studies have attempted to determine whether mimiviruses and marseilleviruses are etiological agents of infection in humans [1–3, 5, 7–10], and indeed a growing number of studies have detected genomic and/or other evidence of the presence of these viruses in humans [1–4, 6–9]. In Brazil, there are no reports of mimivirus detection in pneumonia patients; however, recent data have shown the detection (PCR and viral isolation) of mimivirus group A in respiratory facility areas in a Brazilian hospital [10]. Although the source of the mimivirus described in this work is uncertain, the exposure of these rural populations to dairy cattle, which may be a source of mimiviruses, is noteworthy. In this context, the detection of antibodies against mimivirus as well as its DNA indicate that these patients may be at risk of opportunistic infections, as previously suggested [4]. Such studies could provide new information about the circulation and epidemiology of mimiviruses. In conclusion, our findings provide new biological and epidemiological information about these viruses.

Acknowledgments We thank our colleagues from Gepvig and the Laboratório de Vírus for their technical support. We also thank Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), CAPES-COFECUB and the Pró-Reitoria de Pesquisa da Universidade Federal de Minas Gerais.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

Funding This study was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, the Fundação de Amparo à Pesquisa do Estado de Minas Gerais, and the Ministério da Agricultura, Pecuária e Abastecimento. F.P.D. was supported by a fellowship from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-COFEUB. J.S.A, E.G.K. and G.S.T. are researchers of the Conselho Nacional de Desenvolvimento Científico e Tecnológico.

References

1. Abrahão JS, Dornas FP, Silva LC, Almeida GM, Boratto PV, Colson P, La Scola B, Kroon EG (2014) Acanthamoeba polyphaga mimivirus and other giant viruses: an open field to outstanding discoveries. Virol J 11:120
2. Aherfi S, Colson P, Audoly G, Nappez C, Xerri L, Valensi A, Million M, Lepidi H, Costello R, Raoult D (2016) Marseillevirus in lymphoma: a giant in the lymph node. Lancet Infect Dis 16(10):e225–e234
3. Colson P, Aherfi S, Scola La, Raoult D (2016) The role of giant viruses o amoebas in humans. Curr Opin Microbiol 31:199–208
4. Dornas FP, Rodrigues FP, Boratto PV, Silva LC, Ferreira PC, Bonjardim CA, Trindade GS, Kroon EG, La Scola B, Abrahao JS (2014) Mimivirus circulation among wild and domestic mammals, Amazon Region, Brazil. Emerg Infect Dis 20(3):469–472
5. Dornas FP, Silva LC, de Almeida GM, Campos RK, Boratto PV, Franco-Luiz AP, La Scola B, Kroon EG, Abrahao JS (2014) Acanthamoeba polyphaga mimivirus stability in environmental and clinical substrates: implications for virus detection and isolation. PLoS One 9(2):e87811
6. Hoffmann B, Scheuch M, Höper D, Jungblut R, Holsteg M, Schirrmeier H et al (2012) Novel orthobunyavirus in cattle,

Europe, 2011. Emerg Infect Dis 18:469–472. doi:10.3201/eid1803.111905
7. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie JM, Raoult D (2003) A giant virus in amoebae. Science 299:2033
8. Saadi H, Pagnier I, Colson P, Cherif JK, Beji M, Boughalmi M, Azza S, Armstrong N, Robert C, Foumous G, La Scola B, Raoult D (2013) First isolation of mimivirus in a patient with pneumonia. Clin Infect Dis 57(4):e127–e134
9. Saadi H, Reteno DG, Colson P, Aherfi S, Minodier P, Pagnier I, La Scola B, Raoult D (2013) Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia. Intervirology 56(6):424–429
10. Santos-Silva LK, Arantes TS, Andrade KR, Rodrigues RAL, Boratto PVM, Almeida GMF, Kroon EG, La Scola B, Clemente WT, Abrahão JS (2015) High positivity of mimivirus in inanimate surfaces of a hospital respiratory-isolation facility, Brazil. J Clin Virol 66:62–65

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.

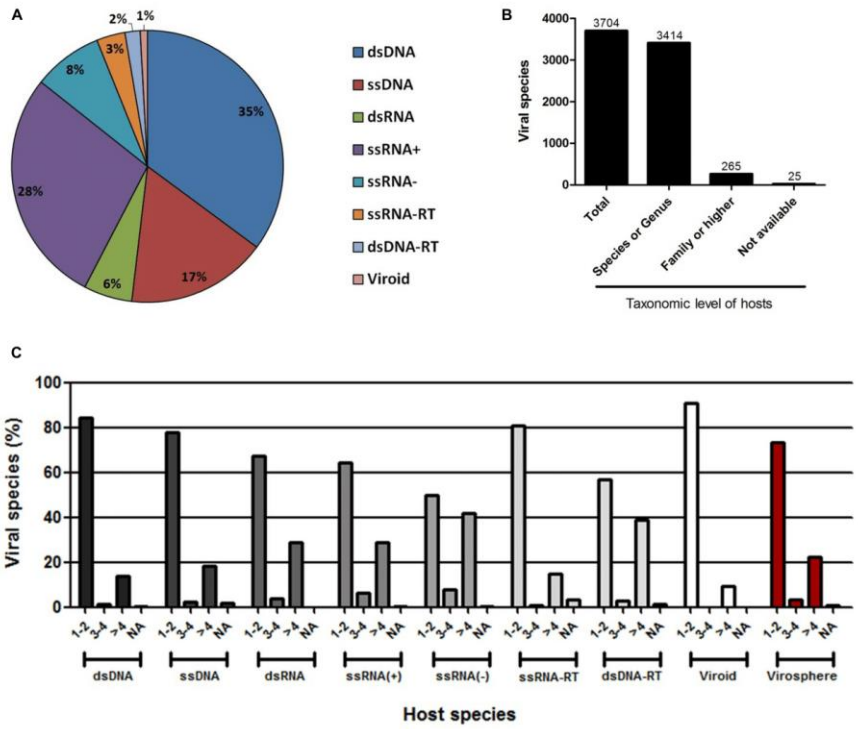


FIGURE 1 | Host range of the known virosphere. **(A)** Pie chart showing the distribution of the viruses recognized by the International Committee on Taxonomy of Viruses (ICTV) according to Baltimore's classification. **(B)** Taxonomic level of the hosts associated to the known viruses. More than 90% 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

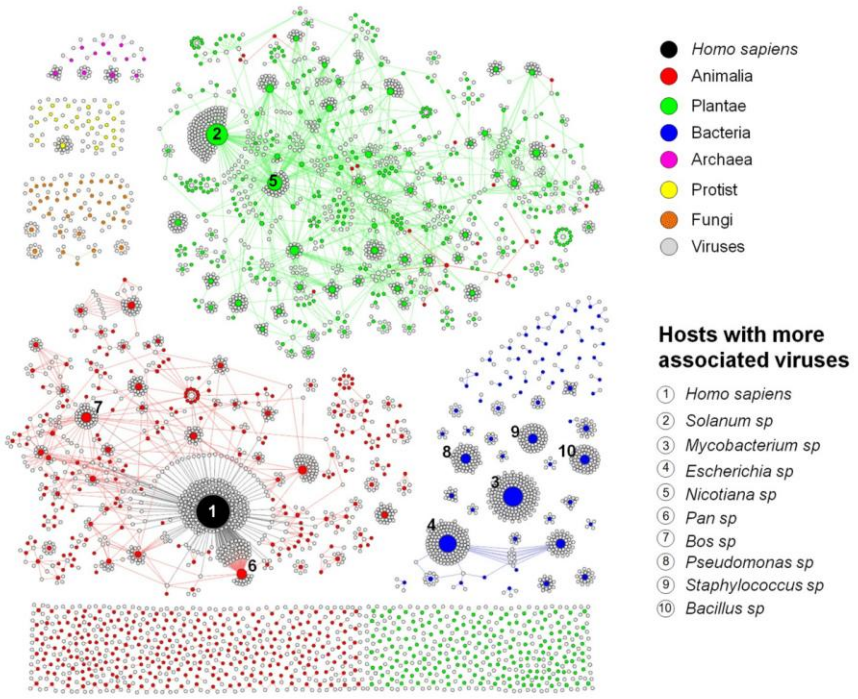


FIGURE 2 | Virus–host interaction network. Bipartite network graph showing a spatially connected network among viruses and hosts, a reflection of our limited knowledge about the viruses and their hosts. Each node represents a virus (gray), or a host genus, classified according to the taxonomic group (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.

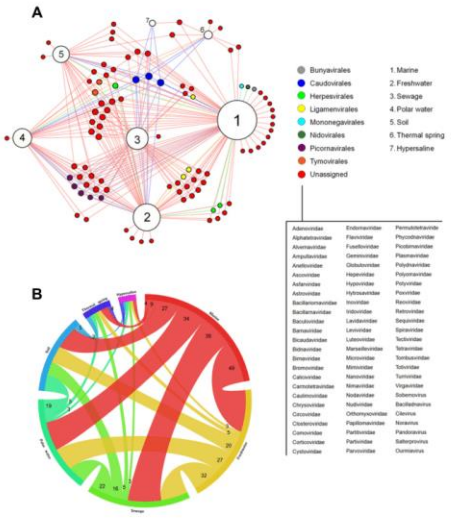


FIGURE 3 | Viral diversity on Earth. **(A)** Network graph showing the viral groups detected in different environments by metaviromic analysis. Each node represents an environment (white) or viral groups (families or known viral genus – colored nodes) classified according to the orders formally recognized 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

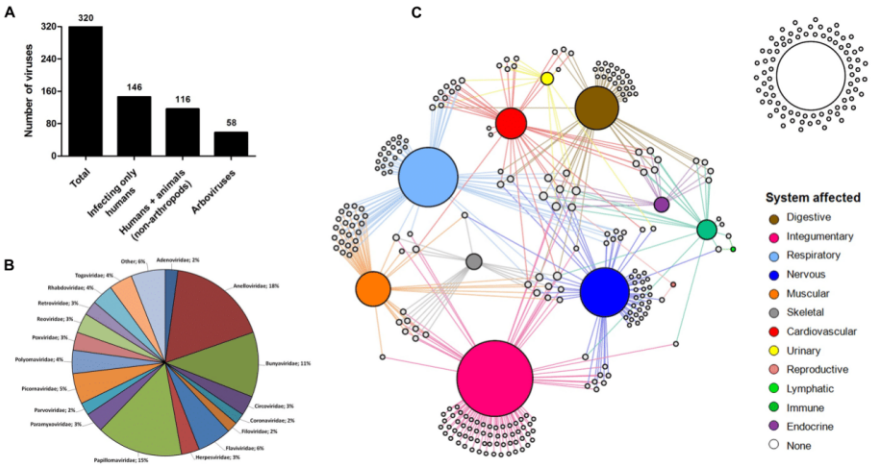


FIGURE 4 | Human viruses and affected systems. **(A)** Human-affecting viruses divided among infecting only humans, infecting humans and other mammals, and arboviruses. **(B)** Pie chart showing the classification of the viruses. A total of 27 groups are represented in the chart. Others: *Deltavirus, 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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 December 2017
doi: 10.3389/fmicb.2017.02562

Genome Characterization of the First Mimiviruses of Lineage C Isolated in Brazil

Felipe L. Assis[1], Ana P. M. Franco-Luiz[1], Raíssa N. dos Santos[2], Fabrício S. Campos[3], Fábio P. Dornas[1], Paulo V. M. Borato[1], Ana C. Franco[2], Jônatas S. Abrahao[1], Philippe Colson[4] and Bernard La Scola[4]**

[1] Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [2] Departamento de Microbiologia, Imunologia e Parasitologia, Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [3] College of Veterinary Medicine and Agronomy, University of Brasília, Brasília, Brazil, [4] CNRS 7278, IRD 198, INSERM 1095, UM63, IHU – Méditerranée Infection, AP-HM, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, Aix-Marseille Université, Marseille, France

The family *Mimiviridae*, comprised by giant DNA viruses, has been increasingly studied since the isolation of the *Acanthamoeba polyphaga mimivirus* (APMV), in 2003. In this work, we describe the genome analysis of two new mimiviruses, each isolated from a distinct Brazilian environment. Furthermore, for the first time, we are reporting the genomic characterization of mimiviruses of group C in Brazil (Br-mimiC), where a predominance of mimiviruses from group A has been previously reported. The genomes of the Br-mimiC isolates Mimivirus gilmour (MVGM) and Mimivirus golden (MVGD) are composed of double-stranded DNA molecules of ∼1.2 Mb, each encoding more than 1,100 open reading frames. Genome functional annotations highlighted the presence of mimivirus group C hallmark genes, such as the set of seven aminoacyl-tRNA synthetases. However, the set of tRNA encoded by the Br-mimiC was distinct from those of other group C mimiviruses. Differences could also be observed in a genome synteny analysis, which demonstrated the presence of inversions and loci translocations at both extremities of Br-mimiC genomes. Both phylogenetic and phyletic analyses corroborate previous results, undoubtedly grouping the new Brazilian isolates into mimivirus group C. Finally, an updated pan-genome analysis of genus *Mimivirus* was performed including all new genomes available until the present moment. This last analysis showed a slight increase in the number of clusters of orthologous groups of proteins among mimiviruses of group A, with a larger increase after addition of sequences from mimiviruses of groups B and C, as well as a plateau tendency after the inclusion of the last four mimiviruses of group C, including the Br-mimiC isolates. Future prospective studies will help us to understand the genetic diversity among mimiviruses.

Keywords: *Mimiviridae*, pan-genome, genomics, giant virus, mimivirus

OPEN ACCESS

Edited by:
Steven M. Short, University of Toronto Mississauga, Canada

Reviewed by:
Masaharu Takemura, Tokyo University of Science, Japan
Yael Mutsafi Benhalevy, National Institutes of Health (NIH), United States

****Correspondence:***
Bernard La Scola
bernard.la-scola@univ-amu.fr
Philippe Colson
philippe.colson@univ-amu.fr; ph.colson@gmail.com

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

Received: *27 October 2017*
Accepted: *11 December 2017*
Published: *22 December 2017*

Citation:
Assis FL, Franco-Luiz APM, dos Santos RN, Campos FS, Dornas FP, Borato PVM, Franco AC, Abrahao JS, Colson P and La Scola B (2017) Genome Characterization of the First Mimiviruses of Lineage C Isolated in Brazil. Front. Microbiol. 8:2562. doi: 10.3389/fmicb.2017.02562

INTRODUCTION

Since the serendipitous discovery of *Acanthamoeba polyphaga mimivirus* (APMV) in 2003, dozens of studies have been conducted to describe how widespread and diverse this new viral family is La Scola et al. (2003, 2010), Raoult et al. (2004), Arslan et al. (2011), Colson et al. (2011a,b), Legendre et al. (2012), Boughalmi et al. (2013b,c), Saadi et al. (2013a), Yoosuf et al. (2014a,b).

Concomitantly, researchers have been working on the biology and molecular characterization of other mimivirus relatives isolated from several human and environmental samples, the latter of which include cooling water tower, freshwater, saltwater, soil, leech, oyster, and other sources collected in many countries in Oceania, Europe, Asia, Africa, and South America (La Scola et al., 2008; Fischer et al., 2010; Arslan et al., 2011; Yoosuf et al., 2012; Boughalmi et al., 2013a,b,c; Pagnier et al., 2013; Saadi et al., 2013a,b; Campos et al., 2014; Bajrai et al., 2016; Takemura et al., 2016). During those studies, notable sets of genes seemingly encoded by the genome of these new viruses were observed. These included genes encoding tRNA sequences, aminoacyl-tRNA synthetases, and peptide synthesis factors. Equally surprising was that mimiviruses can be associated with small viruses, which were named virophages in analogy to bacteriophages that infect bacterial hosts (La Scola et al., 2008). Some mimiviruses were recently predicted to encode a defense system named the MIMIVIRE, which enables them to target specific virophages (Levasseur et al., 2016). However, all these astonishing discoveries could be the "tip of the iceberg" regarding mimivirus features, as ∼50% of the sequences of these viruses encode proteins that are hypothetical, i.e., without a defined function (La Scola et al., 2003; Raoult et al., 2004).

The mimiviruses have a semi-icosahedral 410–550 nm in diameter capsid, with a symmetry breaking at a single vertex of the particle forming a five-branch star structure, called the 'stargate' (Zauberman et al., 2008). The capsid surface is covered, except at the "stargate" vertex, by a 150-nm thick fibril layer, involved in a matrix with a composition initially thought to be similar to peptidoglycan. Although the mimiviruses have been isolated using co-culture on amoebae of the genus *Acanthamoeba*, knowledge about their natural reservoir as well as their host range is still limited. The mimiviruses replicate in the host cytoplasm in a replication factory that is formed after the genome is released (Suzan-Monti et al., 2007; Mutsafi et al., 2010; Colson et al., 2017). The genomes of mimiviruses are comprised by a linear dsDNA molecule that is 0.92–1.22 Mb long and encodes 930–1,178 proteins (Raoult et al., 2004). The genome of the prototype Mimivirus was described to present two inverted repeats of about 900 nucleotides near both extremities, suggesting that the Mimivirus genome might adopt a circular topology during viral replication, as in some other NCLDVs (Raoult et al., 2004).

The family *Mimiviridae* is comprised by two genera, named (1) *Mimivirus*, composed of mimiviruses infecting amoebal species, and (2) *Cafeteriavirus*, a distantly related mimiviruses group comprised by the type species *Cafeteria roenbergensis virus* (CroV; which infects a marine heterotrophic bi-flagellate) International Committee on Taxonomy of Viruses [ICTV], 2017. Other related distant mimiviruses have been associated with CroV, including Organic lake phycodnaviruses and *Phaeocystis globosa* virus (Yutin et al., 2013). The recently described klosneuviruses also seem to be related to *Mimiviridae* members (Schulz et al., 2017). The genus *Mimivirus* can be divided into three lineages A, B, and C, according to phylogenomic data including phylogenies based on conserved core genes, for example family B DNA polymerase and ribonucleotide reductase encoding genes (Boyer et al., 2010; Colson et al., 2012; Legendre et al., 2012; Campos et al., 2014).

We isolated the first Brazilian mimivirus strain, named Samba virus (SBV), from a water sample collected in the Amazon region in 2011. Phylogenomic analyses clustered the SBV into mimivirus lineage A (Campos et al., 2014), which includes the APMV, the prototype species of family *Mimiviridae*. More recently, Brazilian mimivirus strains have been isolated and/or detected from fresh water, oyster, sewage, humans, and both wild and domestic mammals, and their biological and molecular characterization have been reported (Dornas et al., 2014, 2016, 2017; Andrade et al., 2015; Boratto et al., 2015). Curiously, all Brazilian mimivirus strains were classified into mimivirus lineage A, suggesting that this lineage is the most widespread in Brazil (Andrade et al., 2015; Assis et al., 2015; Boratto et al., 2015). In addition, Assis et al. (2015) described the pan-genome of mimivirus lineage A, which was composed of 1129 clusters of orthologous groups (COGs) of proteins encoded by all genomes available at that time. All these data led us to ask more questions about the diversity of mimiviruses circulating in Brazil and resulted in the decision to conduct additional prospective studies. In this way, Dornas et al. (2015), using a panel of protozoa (*Acanthamoeba castellanii* [AC], *Acanthamoeba polyphaga* [AP], *Acanthamoeba griffinii* [AG] and *Vermamoeba vermiformis* [VV]), were able to isolate 62 new mimivirus-like strains from sewage, sludge, water, wet soil, and lake sediment collected from different areas of the Pampulha lagoon in Belo Horizonte, Minas Gerais, Brazil (Dornas et al., 2015). A higher prevalence of lineage A mimiviruses (90.3%) was observed, followed by lineage C mimiviruses (6.4%) and lastly lineage B mimiviruses (3.2%). However, neither further analysis of the biological and molecular features of these viruses nor phylogenies were provided, once the classification of these new isolates into lineages was inferred based on BLAST hits obtained against the NCBI nt database.

In this work, we report the molecular and phylogenetic analysis of two Brazilian mimiviruses from lineage C (Br-mimiC): (1) Mimivirus gilmour (MVGM) – isolated from water collected at Pampulha lagoon by Dornas et al. (2015); (2) Mimivirus golden (MVGD) – isolated from golden mussels (*Limnoperna fortunei*) collected from Guaíba Lake, Rio Grande do Sul, Brazil, in July 2014. Both Br-mimiC viruses were isolated using the protozoa AP as support for co-culture. In addition, we conducted an updated pan-genome analysis of all available genomes of mimiviruses from lineages A to C.

MATERIALS AND METHODS

Sample Collection

A collaborative effort involving the Aix-Marseille University (France), and the Federal Universities of Minas Gerais and Rio Grande do Sul (Brazil) was established aiming to conduct prospective studies of giant viruses in different regions and environments in Brazil. All collection procedures were performed with the authorization of IBAMA-SISBIO (number 34293-2). For

this work, water samples were collected in sterile tubes from Pampulha lagoon in September 2014, and were directly used for inoculation procedures. In addition, golden mussels (*L. fortunei*) were collected from Guaíba Lake in July 2014 (30°01′59″ S, 51°13′48″ W) (Dos Santos et al., 2016). The mussels were collected from the lake bottom at a depth of 2 m, and they were attached to a metal grid that had been submerged for 6 months before the date of collection. Golden mussels were submerged for 15 min in 70% ethanol for superficial shell decontamination. Subsequently, the valves were opened and the inner water was collected and diluted in 1 mL of saline buffer (PBS). The samples were pooled, totaling eight pools. These pools were homogenized with 1 mL of PBS, centrifuged at 10,000 × *g*, filtered through a 0.45-μm pore-size membrane. The resulting filtrate was treated with 10 U/μL of Penicillin-GIBCO by Life Technologies to prevent bacterial contamination. All the samples were stored at 4°C until the inoculation procedures.

Virus Isolation

The MVGM sample was isolated using co-culture of AP strain LINC AP1 previously cultured in a 75-cm^2 cell culture flask with 30 ml of peptone-yeast extract-glucose medium (PYG) at 30°C for 25 h. The culture supernatant was pelleted by centrifugation, suspended in PAS supplemented with an antibiotic mix containing 10 μL of ciprofloxacin (4 μg/mL; Panpharma, Z.I., Clairay, France), 10 μL of vancomycin (4 μg/mL; Mylan, Saint-Priest, France), 10 μL of colimycin (500 IU/mL; Sanofi Aventis, Paris, France), 10 μL of rifampicin (4 μg/mL; Sanofi Aventis), and 10 μL of fungizone (100 μg/mL; Bristol-Myers Squibb, Rueil-Malmaison, France), and dispensed in 0.5 ml amounts to the wells of a 24-well plate with a suspension cell concentration of 10^6 cells/ml. After that, 100 μL of samples were inoculated into wells and incubated at 30°C for 4 days. The sub-cultures were performed twice on fresh amoebae in a 1-10th dilution. A negative amoebal control was used in each microplate (Dornas et al., 2015). For the MVGD strain, the amoeba support for co-culture were AP genotype T4 previously cultured in 10 mL of Peptona-Yeast Extract-Glucose (PYG) medium at 30°C in 25-cmł culture flasks supplemented with 50 μg of gentamicin. After 48 h, the cells were harvested and centrifuged. The pellet was re-suspended in sterile PAS (Page's amoeba saline), and 10^4 amoebas per well were cultured in 24-well microplates. After 24 h, 100 μL of samples were inoculated into wells and incubated at 30°C for 3 days. The sub-cultures were performed as previously mentioned, and amoeba cells were assessed daily for the presence of viruses and for cytopathic effects on the cell monolayer.

DNA Extraction and Genome Sequencing

For MVGM strain, viral DNA was extracted with the automated EZ1 Virus Mini-Kit v.2 kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. DNA quality and concentration were checked, using a nanodrop spectrophotometer (Thermo Scientific, Waltham, MA, United States). For the MVGD strain, the supernatant of the *A. polyphaga*-infected cells was collected, and centrifuged at 5,000 × *g* for 5 min. The cell-free virus particles were pelleted on a 25% sucrose cushion by ultracentrifugation (Sorvall Combi) at 33,000 × *g* for 2 h at 4°C. The pellet was re-suspended in Tris-EDTA-NaCl buffer (TEN). In order to remove the nucleic acids not protected by the capsid, the preparation was treated with 100 U of DNAse I (Roche) and 100 U of RNAse (Invitrogen) at 37°C for 1 h. Next, the virus DNA was extracted using phenol–chloroform (Sambrook and Russell, 2001) and re-suspended in ultrapure water. The quality and amount of virus DNA was analyzed using a NanoSpec and Qubit apparatus (Life Technologies). Both extracted viral DNA were submitted to sequencing performed in a MiSeq (Illumina) apparatus with paired-end applications (2 bp × 150 bp). The pair-end samples were prepared with a Nextera XT DNA sample prep kit.

Genome Assembly and Annotation

After sequencing, reads from MVGM and MVGD were *de novo* assembled using Geneious and SPADES softwares. The gene predictions were performed using RAST (Rapid Annotation using Subsystem Technology) (Aziz et al., 2008) and GeneMarkS (Besemer et al., 2001) tools. Transfer RNA (tRNA) sequences were identified using the tRNAscan-SE tool (Schattner et al., 2005). The functional annotations were inferred by BLAST searches against the GenBank NCBI non-redundant protein sequence database (nr) (e-value $< 1 \times 10^{-3}$), the set of COGs of the NCLDVs (named NCVOGs) (Altschul et al., 1990; Yutin et al., 2009) and by searching specialized databases through the Blast2GO platform (Conesa et al., 2005). The genome annotations were then manually revised and curated. The predicted open reading frames (ORFs) smaller than 50 amino acids (aa) and that had no hit in any database were discarded. The ORFs longer than 50 aa without hits in any database (ORFans) were kept.

Comparative Genomic and Pan-genome Analysis

The synteny among mimiviruses from distinct lineages was checked using MAUVE program (Darling et al., 2010). The OrthoMCL tool (Chen et al., 2006) was used to identify the paralog families from Br-mimiC genomes, while Proteinortho5.pl tool (Lechner et al., 2011) was used to identify orthologous gene sequences shared by Br-mimiC. The average amino acid identity (AAI) calculator tool (Rodriguez-R and Konstantinidis, 2014) was used to compare identity between orthologous genes from Br-mimiC strains and representative mimiviruses of lineages A-C. To estimate the size of the pan-genome of the family *Mimiviridae*, their predicted proteins were clustered using the Proteinortho5.pl program (Lechner et al., 2011), using an aa sequence identity of 30% and a sequence coverage of 50% as thresholds. We also described pan-genome and core genes size variation by stepwise inclusion of each new virus annotation in the pairwise comparisons of the gene contents of all available mimivirus genome sequences.

Phylogeny

The aa sequence alignments and phylogenetic trees were built using the MEGA6 software (Tamura et al., 2013) and the

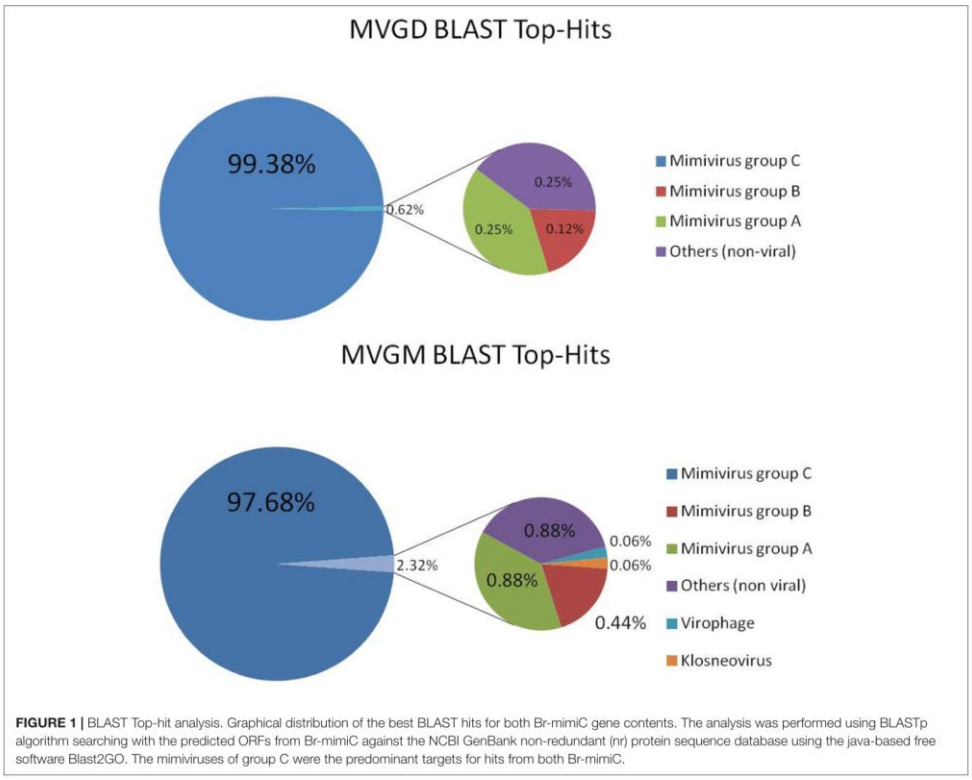


FIGURE 1 | BLAST Top-hit analysis. Graphical distribution of the best BLAST hits for both Br-mimiC gene contents. The analysis was performed using BLASTp algorithm searching with the predicted ORFs from Br-mimiC against the NCBI GenBank non-redundant (nr) protein sequence database using the java-based free software Blast2GO. The mimiviruses of group C were the predominant targets for hits from both Br-mimiC.

maximum likelihood method. Phylogenetic reconstructions were based on individual alignment of the five core genes, namely the family B DNA polymerase, the D6/D11 helicase, the VV A18 helicase, the D5 primase-helicase, and the Major Capsid Protein. In addition, we performed a hierarchical clustering based on the gene presence/absence pattern of 5443 NCVOGs, using the MeV tool (Eisen et al., 1998) with Pearson correlation as distance metric. The phylogenetic tree was visualized using the FigTree v1.4.3 tool (available online: http://tree.bio.ed.ac.uk/software/figtree/).

RESULTS

General Features of Br-mimiC Genomes

The genomes of MVGM (GenBank number: MG602507) and MVGD (GenBank number: MG602508) are double-stranded DNA molecules of 1,258,663 base pairs (bp) (partial sequence) and 1,248,960 base pairs (complete sequence) encoding 1,135 and 1,127 ORFs, respectively. The ORFs length of both Br-mimiC ranged from 37 to 2,907 aa, with an average length of 326 aa. The BLAST analysis (coverage > 90%; identity > 80%; e-value < 10e-5) against the NCBI nr database (updated in October, 2017) identified 1088 and 1090 hits for MVGM and MVGD sequences, respectively. Furthermore, we identified 28 and 19 ORFans into MVGM and MVGD genomes, respectively. In addition, 19 and 18 ORFs without BLAST hit and smaller than 50 aa were not include in the subsequent analysis neither in the final annotation of MVGM and MVGD genomes, respectively.

The comparison between the Br-mimiC viruses genomes showed the presence of 1,042 orthologous proteins, whereas 66 and 61 proteins are unique to MVGM and MVGD, respectively. The set of unique genes of the MVGD included 18 ORFans, besides hypothetical proteins, ankyrins, F-box and FNIP repeat-containing proteins, collagen-like proteins, BTB POZ domain and WD-repeat proteins and a cholinesterase-like protein. With the exception of the cholinesterase-like protein found in the MVGD genome, the set of unique genes of the MVGM was comprised by the same classes of protein, besides a DNA primase and a putative transposase. In addition, the MVGM and MVGD

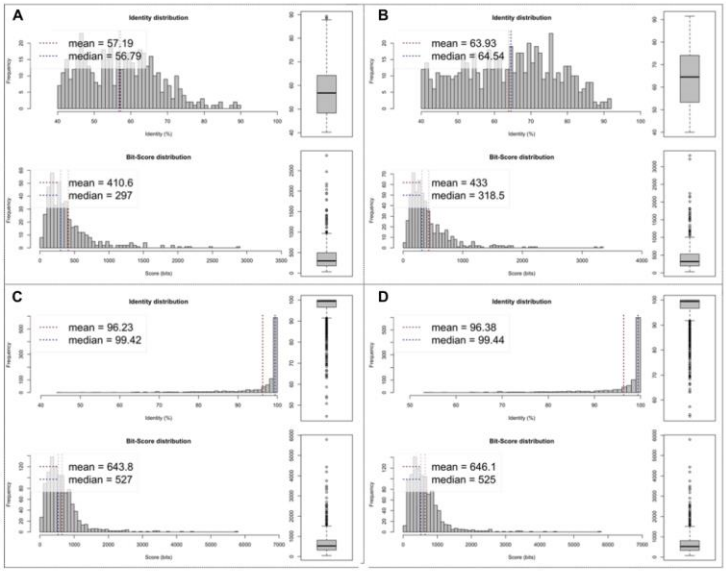


FIGURE 2 | Average amino acid identity (AAI). In this analysis, estimates were reached out using both best hits (one-way AAI) and reciprocal best hits (two-way AAI) between two datasets of proteins from the Br-mimiC isolates and representative strains from mimivirus groups A to C. Plots **(A,B)** demonstrate the AAI between MVGM and mimiviruses from group A to group B, respectively. Plot **(C)** shows the AAI between MVGM and MVGD. Plot **(D)** shows the AAI between MCHV (group C prototype strain) and MVGD.

TABLE 1 | Distribution of aminoacyl-tRNA synthetases encoded by mimivirus from group A to group C, besides Br-mimiC isolates.

	Mimivirus strains			
Aminoacyl-tRNA synthetase	**Mimivirus A**	**Mimivirus B**	**MCHV**	**Br.mimi C**
Tyrosyl	V	V	V	V
Cysteinyl	V	V	V	V
Methionyl	V	V	V	V
Arginyl	V	V	V	V
Isoleucyl	X	V	V	V
Asparaginyl	X	X	V	V
Tryptophanyl	X	X	V	V

MCHV, Megavirus chilensis; V, presence; X, absence.

genomes encoded to 551 and 558 proteins without defined function, respectively.

Both genomes showed a very similar G+C content (∼26%), genome density (∼0.890 genes per kbp), coding percentage (∼88.5%), and average gene length (995 bp). The best hit analysis for the sequences predicted in these Br-mimiC genomes showed the highest percentage (average of 98.5%) of hits against mimivirus group C sequences (**Figure 1**). The average AAI analysis (**Figure 2**) corroborated the best hit analysis showing the greatest AAI value between sequences from Br-mimiC and other mimiviruses of group C (∼96%), followed by mimiviruses of group B (63.9%) and mimiviruses of group A (57.1%). When compared between each other, the Br-mimiC showed an AAI of 96.3% (**Figure 2**). Beyond, the ORFs predicted into Br-mimiC genomes possess orthologs into other mimiviruses, hosts and/or sympatric organisms, beside virophage and other giant viruses. Furthermore, we observed the presence of seven aminoacyl (tyrosyl, cysteinyl, methionyl, arginyl, isoleucyl, asparaginyl, and tryptophanyl) tRNA synthetases (aaRS) in Br-mimiC, which has been described as a signature of mimiviruses of group C (Colson et al., 2013), while mimiviruses of groups A and B encode four and five aaRS, respectively (**Table 1**). Although the best hit analysis has shown a match against a virophage sequence, we did not detect those mimivirus-related virus associated with Br-mimiC.

Even sharing several genetic features, such as a low G+C content and large and similar genome sizes and gene repertories, the Br-mimiC isolates presented singular features which allowed distinguishing them as two distinct isolates. One of the main differences between the Br-mimiC viruses is the presence of six tRNA molecules (2x Leu-TAA, Leu-CAA, Trp-CCA, His-GTG, and Cys-GCA) encoded by MGMV, while the MGDV was

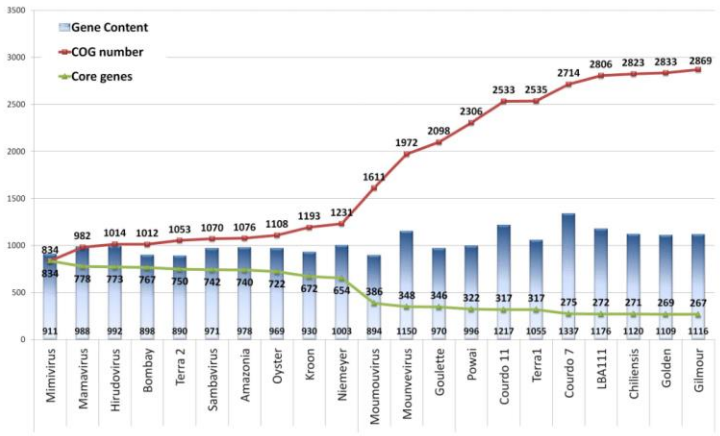


FIGURE 3 | Evolution of the pan-genome (red line) and core genome (green line) size of the family *Mimiviridae*. Numbers at the base of blues bars refer to the number of genes encoded by each virus strain. Numbers at line nodes represent the cumulative COG (red line) and core genes (green line) numbers after the inclusion of a new genome. The COGs definition was performed by using the Proteinortho5 tool, using AAI and coverage of 30 and 50%, respectively.

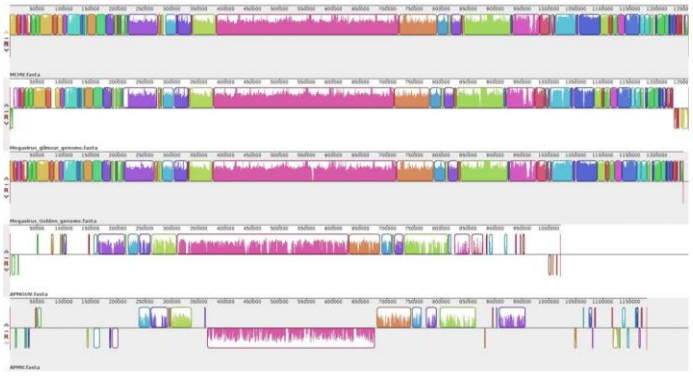


FIGURE 4 | Genome synteny analysis. Schematic genome alignment diagram obtained using the Mauve software package. The analysis was performed using the reference genome of APMV (NC_014649.1), Moumouvirus (NC_020104.1) and MCHV (NC_016072.1), besides the genome sequences of Br-mimiC isolates. The blocks illustrated above X axis are in positive strand (forward sense), while blocks below the X axis are in the negative strand (reverse sense).

predicted to encode only three tRNA molecules (Leu-TAA, Leu-CAA, and Trp-CCA). Taken together, these results confirm the isolation of the first mimiviruses of group C in Brazil.

In order to assess the gene encoding capacity of mimiviruses, we performed an updated pan-genome analysis (**Figure 3**) using all mimivirus genome data available in the NCBI genome database. This analysis will shows us the set of different proteins encoded by all mimiviruses, and will indicate whether the genetic complexity of this group has been fully addressed or not. For this analysis, only complete genome data sets were used, and the result showed a continuous increase in the pan-genome size reaching 2869 COGs, an improvement of 1740 new COGs compared with our previous analysis (Assis et al., 2015) that only considered genomes of mimiviruses of group A. Furthermore, breaks in this rising curve were observed for each new mimivirus representative of the lineages B and C; the number of COGs increased by 380

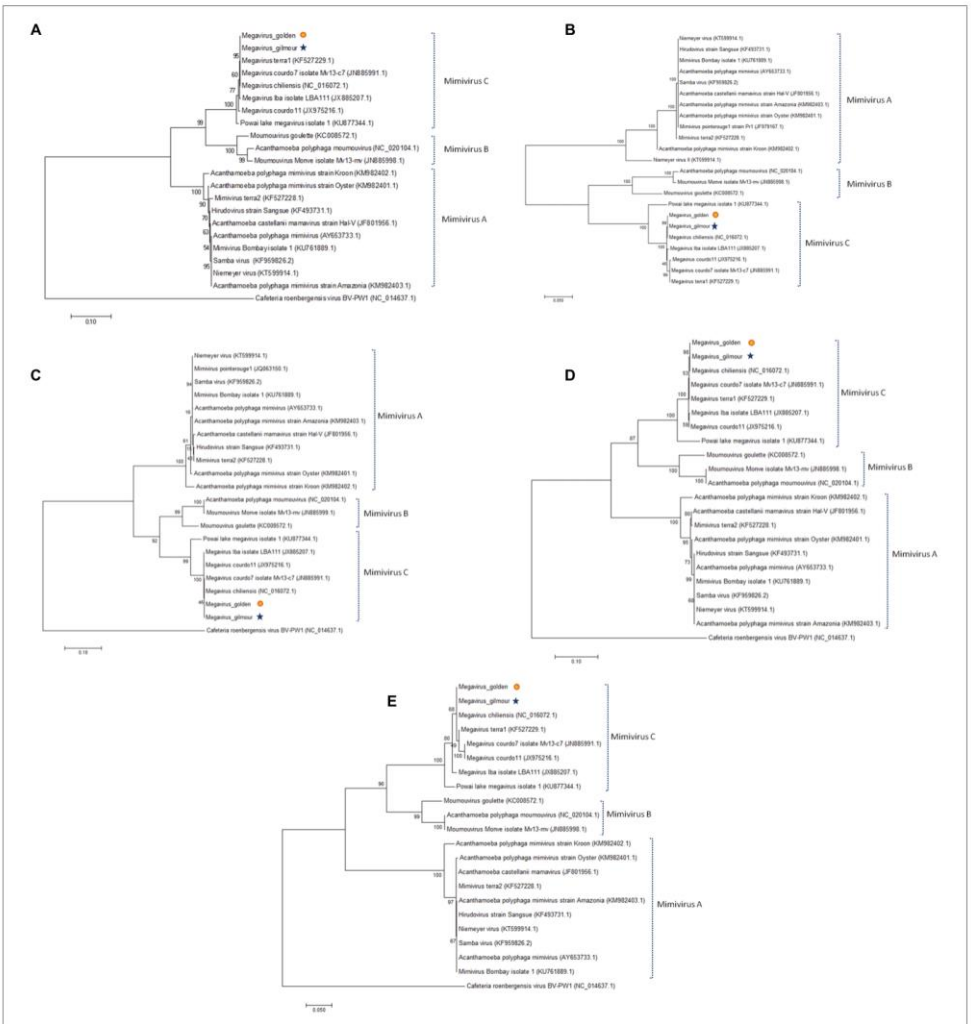


FIGURE 5 | Phylogenetic reconstruction of the NCLDV core genes for different mimiviruses isolates. Here we have done a phylogenetic analysis for many representatives of the different *Mimiviridae* lineages **(A–C)** including for the new Br-mimiC viruses, based on the nucleotide sequences of the five NCLDV core genes, namely the **(A)** major capsid protein, **(B)** the family B DNA polymerase, **(C)** the D6/D11 helicase, the **(D)** VV A18 helicase, and the **(E)** D5 primase-helicase.

from lineages A to B, and an additional increase of 208 COGs from lineages A and B to lineage C were observed. In addition, we observed a stabilization trend after the inclusion of the last four mimiviruses of group C, which included the Br-mimiC isolates.

Conversely, we observed a continuous decrease of the core genome after addition of new representatives. An abrupt reduction was only observed after inclusion of the first mimivirus of group B (268 COGs reduction), while a more discrete reduction was observed when sequences from mimiviruses C

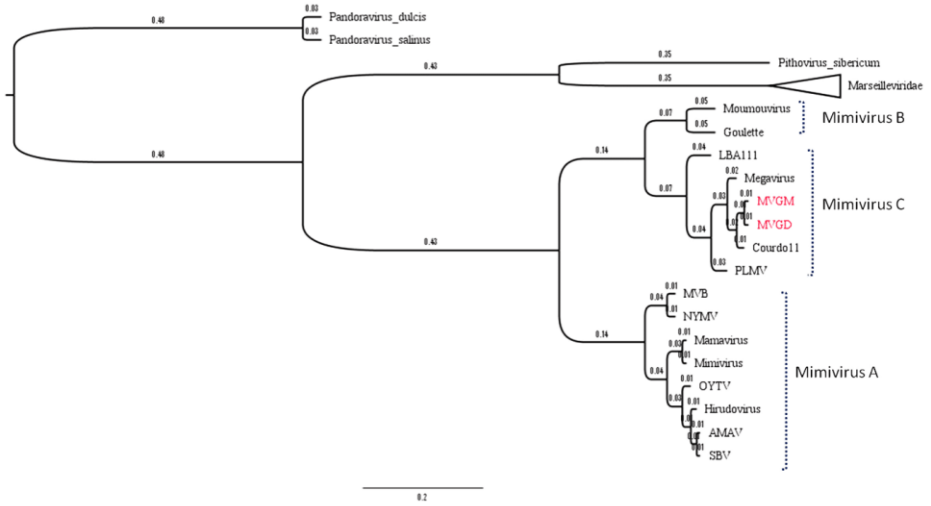


FIGURE 6 | Hierarchical clustering based on the pattern of gene presence/absence for the NCVOGs of different mimivirus isolates. Here we have done a hierarchical clustering tree, based on phyletic patterns, using a presence/absence matrix of 5,443 NCVOG from many representatives of the different *Mimiviridae* lineages (A, B, and C), including the new Br-mimiC viruses, highlighted in red. The distance from horizontal lines connecting the groups represents the time of divergence among the different mimiviruses isolates. The phylogenetic tree was generated using the MeV 4.8.1 tool.

were included (24 COGs reduction). Further, a stabilization trend of the core genome size was observed for the last five mimiviruses C, including Br-mimiC. In addition, we observed an intra-group divergence of 249 COGs among mimiviruses A, 487 COGs among mimiviruses B, and 563 COGs among mimiviruses C. Altogether, these results highlight a stabilization trend in the pan-genome and core genome evolution of amoeba-associated mimiviruses. In Addition our results showed a great divergence even among viruses from the same group (**Figure 3**).

Synteny Analysis

The synteny analysis showed very similar genome architectures for *Megavirus chilensis* (MCHV) and the Br-mimiC viruses (**Figure 4**). However, some divergences were observed among mimiviruses C, such as inversions and translocations at both extremities of the MVGM genome, while the MVGD genome better resembled the MCHV genome architecture than that of others. Furthermore, the genome of mimiviruses C showed a better co-linearity with less block brakes in their central region (from ∼250 to ∼950 kb) compared to what is observed at both extremities, which showed an increased number of shorter homologous regions. Curiously, the central region of mimiviruses C genomes showed an overall smaller similarity when compared to the extremities. In addition, these mimiviruses C presented a distinct genome macrosynteny from mimiviruses A and B.

Phylogeny

To better understand the evolutionary relationship between the Br-mimiC viruses and other mimiviruses, we performed phylogenetic analyses based on NCLDV core genes including the family B DNA polymerase, the VV A18 helicase, the D5 helicase, the D6/D11 helicase and the major capsid protein (**Figures 5A–E**). Furthermore, a hierarchical clustering tree (**Figure 6**), based on the phyletic patterns, was constructed using a presence-absence matrix of 5,443 NCVOG (clusters of orthologous genes shared by NCLDV). The phylogenetic trees recurrently clustered the Br-mimiC viruses into mimivirus group C, corroborating all previous analyses. The core genes-based trees showed the close relationship of Br-mimiC isolates to the MCHV isolate, the mimivirus of group C whose genome was first described, in 2011, and that was obtained from Chile. However, the phyletic tree, which highlights the gene presence/absence pattern, showed a close relationship of Br-mimiC with Courdo11 virus, isolated in 2010 by inoculating *Acanthamoeba* spp. with freshwater collected from a river of southeastern France.

DISCUSSION

In this work, we describe the isolation and genome features of the first two isolates of mimivirus group C from Brazil. Recently, we described the isolation of Samba virus, the first representative of family *Mimiviridae* in Brazil, belonging to

mimivirus group A (Campos et al., 2014). Subsequently, in Brazil, several mimiviruses and other giant viruses have been isolated in several biomes and a hospital respiratory-isolation facility, and mimivirus has been more recently detected in human sera (Dornas et al., 2014, 2015, 2017; Andrade et al., 2015; Boratto et al., 2015; dos Santos Silva et al., 2015). However, this is the first time that a mimivirus of group C is isolated in this country, which highlights the diversity of giant viruses in Brazil and how widespread these viruses are. Although the former member of mimivirus group C, *Mimivirus chilensis*, has been isolated in Chile, the remaining isolates of this group have frequently been isolated from environmental and clinical samples collected in Asia, Africa, and Europe (Pagnier et al., 2013). Thus, we believe that as new prospective studies are performed, new isolates might be discovered.

Even though they share many molecular features, as well as biological ones (data not shown), the Br-mimiC viruses can be recognized as two distinct isolates. The MVGM isolate has a genome ∼10 kb larger than the MVGD genome and encodes eight more ORFs than MVGD. The unique proteins of both Br-mimiC were mainly located at the extremities of both genomes, which have been described as suitable regions for horizontal gene transfers and duplication events in large and giant viruses, including in mimiviruses (Shackelton and Holmes, 2004; Colson et al., 2011a; Filee, 2015).

Even though the Br-mimiC viruses show ORFans in their genomes, which demonstrate the uniqueness of these isolates, there are notwithstanding many family ORFans present, which means that many genes are shared between mimiviruses but have no homolog in other organisms, and the majority of those genes remains functionally unresolved. Furthermore, we could see a still increasing pan-genome of the family *Mimiviridae* after the addition of Br-mimiC viral genome sequences, suggesting that new genes with unpredictable function are out there, yet to be discovered. In addition, the abrupt break in the trend of the core genome evolution after inclusion of lineage B sequences is in line with the fact that mimiviruses of lineages B and C are more related between each other than they are related to mimiviruses of lineage A, as also shown in the phylogeny reconstructions.

A more conserved synteny could be observed in the central region of all the mimiviruses genomes that were analyzed compared to the remaining part of the genomes. In contrast, the central region of mimiviruses C showed a lower mean similarity. The central region possesses the most ancient set of genes (Shackelton and Holmes, 2004), which have been subjected to long-term selective pressure during mimivirus evolutionary history. In contrast, termini regions of the genome more frequently incorporate new genes, and these recently acquired genes still have a more conserved profile. The phylogenetic analysis strongly corroborate all data presented above, indisputably showing the clustering of the new Br-mimiC isolates into mimivirus group C, closely related to *Megavirus chilensis*, the prototype of this group also isolated in South America. However, the phyletic analysis, which is based on gene presence/absence patterns that at least partially result from losses and gains, showed a better grouping of Br-mimiC viruses with the Courdo 11 virus isolate, which was isolated in 2010 from river water samples in France.

CONCLUSION

The discovery of the Br-mimiC viruses contributes to improving the understanding of mimiviral diversity and ubiquity. Nevertheless, the study of giant viruses is still at its beginning. Additional prospective studies must be conducted with the aim of discovering new relatives of these intriguing micro-organisms. Also, this study and others have showed a large number of sequences with unknown function, showing the need of studies focusing in the functional characterization of proteins encoded by the mimiviruses.

AUTHOR CONTRIBUTIONS

FA, PB, and AF-L: data collection and pan-genome analyses. FD, AF, RdS, and FC: samples collection and virus isolation. BLS, PC, and JA: study design. All authors wrote the paper and read its last version.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andrade, K. R., Boratto, P. P., Rodrigues, F. P., Silva, L. C., Dornas, F. P., Pilotto, M. R., et al. (2015). Oysters as hot spots for mimivirus isolation. *Arch. Virol.* 160, 477–482. doi: 10.1007/s00705-014-2257-2

Arslan, D., Legendre, M., Seltzer, V., Abergel, C., and 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. doi: 10.1073/pnas.1110889108

Assis, F. L., Bajrai, L., Abrahao, J. S., Kroon, E. G., Dornas, F. P., Andrade, K. R., et al. (2015). Pan-genome analysis of Brazilian lineage a amoebal Mimiviruses. *Viruses* 7, 3483–3499. doi: 10.3390/v7072782

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Bajrai, L. H., de Assis, F. L., Azhar, E. I., Jardot, P., Robert, C., Abrahao, J., et al. (2016). Saudi moumouvirus, the first group B *Mimivirus* isolated from Asia. *Front. Microbiol.* 7:2029. doi: 10.3389/fmicb.2016.02029

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

Boratto, P. V., Arantes, T. S., Silva, L. C., Assis, F. L., Kroon, E. G., La Scola, B., et al. (2015). Niemeyer virus: a new *Mimivirus* group a isolate harboring a set of duplicated aminoacyl-tRNA synthetase genes. *Front. Microbiol.* 6:1256. doi: 10.3389/fmicb.2015.01256

Boughalmi, M., Pagnier, I., Aherfi, S., Colson, P., Raoult, D., and La Scola, B. (2013a). First isolation of a giant virus from wild *Hirudo medicinalis* leech: mimiviridae isolation in *Hirudo medicinalis*. *Viruses* 5, 2920–2930. doi: 10.3390/v5122920

Boughalmi, M., Pagnier, I., Aherfi, S., Colson, P., Raoult, D., and La Scola, B. (2013b). First isolation of a *Marseillevirus* in the diptera syrphidae *Eristalis tenax*. *Intervirology* 56, 386–394. doi: 10.1159/000354560

Boughalmi, M., Saadi, H., Pagnier, I., Colson, P., Fournous, G., Raoult, D., et al. (2013c). High-throughput isolation of giant viruses of the Mimiviridae and Marseilleviridae families in the Tunisian environment. *Environ. Microbiol.* 15, 2000–2007. doi: 10.1111/1462-2920.12068

Boyer, M., Madoui, M. A., Gimenez, G., La Scola, B., and Raoult, D. (2010). Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses. *PLOS ONE* 5:e15530. doi: 10.1371/journal.pone.0015530

Campos, R. K., Boratto, P. V., Assis, F. L., Aguiar, E. R., Silva, L. C., Albarnaz, J. D., et al. (2014). Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. *Virol. J.* 11:95. doi: 10.1186/1743-422X-11-95

Chen, F., Mackey, A. J., Stoeckert, C. J. Jr., and Roos, D. S. (2006). OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.* 34, D363–D368. doi: 10.1093/nar/gkj123

Colson, P., de Lamballerie, X., Fournous, G., and Raoult, D. (2012). Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. *Intervirology* 55, 321–332. doi: 10.1159/000336562

Colson, P., Fournous, G., Diene, S. M., and Raoult, D. (2013). Codon usage, amino acid usage, transfer RNA and amino-acyl-tRNA synthetases in Mimiviruses. *Intervirology* 56, 364–375. doi: 10.1159/000354557

Colson, P., Gimenez, G., Boyer, M., Fournous, G., and Raoult, D. (2011a). The giant Cafeteria roenbergensis virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of life. *PLOS ONE* 6:e18935. doi: 10.1371/journal.pone.0018935

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610

Colson, P., La Scola, B., Levasseur, A., Caetano-Anolles, G., and Raoult, D. (2017). Mimivirus: leading the way in the discovery of giant viruses of amoebae. *Nat. Rev. Microbiol.* 15, 243–254. doi: 10.1038/nrmicro.2016.197

Colson, P., Yutin, N., Shabalina, S. A., Robert, C., Fournous, G., La Scola, B., et al. (2011b). 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. doi: 10.1093/gbe/evr048

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLOS ONE* 5:e11147. doi: 10.1371/journal.pone.0011147

Dornas, F. P., Assis, F. L., Aherfi, S., Arantes, T., Abrahao, J. S., Colson, P., et al. (2016). A Brazilian marseillevirus is the founding member of a lineage in family marseilleviridae. *Viruses* 8:76. doi: 10.3390/v8030076

Dornas, F. P., Boratto, P. V. M., Costa, G. B., Silva, L. C. F., Kroon, E. G., La Scola, B., et al. (2017). Detection of mimivirus genome and neutralizing antibodies in humans from Brazil. *Arch. Virol.* doi: 10.1007/s00705-017-3455-5 [Epub ahead of print].

Dornas, F. P., Khalil, J. Y., Pagnier, I., Raoult, D., Abrahao, J., and La Scola, B. (2015). Isolation of new Brazilian giant viruses from environmental samples using a panel of protozoa. *Front. Microbiol.* 6:1086. doi: 10.3389/fmicb.2015.01086

Dornas, F. P., Rodrigues, F. P., Boratto, P. V., Silva, L. C., Ferreira, P. C., Bonjardim, C. A., et al. (2014). Mimivirus circulation among wild and domestic mammals, Amazon Region, Brazil. *Emerg. Infect. Dis.* 20, 469–472. doi: 10.3201/eid2003.131050

Dos Santos, R. N., Campos, F. S., Medeiros de Albuquerque, N. R., Finoketti, F., Correa, R. A., Cano-Ortiz, L., et al. (2016). A new marseillevirus isolated in Southern Brazil from *Limnoperna fortunei*. *Sci. Rep.* 6:35237. doi: 10.1038/srep35237

dos Santos Silva, L. K., Arantes, T. S., Andrade, K. R., Lima Rodrigues, R. A., Miranda Boratto, P. V., de Freitas Almeida, G. M., et al. (2015). High positivity of mimivirus in inanimate surfaces of a hospital respiratory-isolation facility, Brazil. *J. Clin. Virol.* 66, 62–65. doi: 10.1016/j.jcv.2015.03.008

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14863–14868. doi: 10.1073/pnas.95.25.14863

Filee, J. (2015). Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution. *Front. Microbiol.* 6:593. doi: 10.3389/fmicb.2015.00593

Fischer, M. G., Allen, M. J., Wilson, W. H., and 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. doi: 10.1073/pnas.1007615107

International Committee on Taxonomy of Viruses [ICTV] (2017). Available at: https://talk.ictvonline.org/ [accessed September, 2017].

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

La Scola, B., Campocasso, A., N'Dong, R., Fournous, G., Barrassi, L., Flaudrops, C., et al. (2010). Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. *Intervirology* 53, 344–353. doi: 10.1159/000312919

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., et al. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104. doi: 10.1038/nature07218

Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011). Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformat.* 12:124. doi: 10.1186/1471-2105-12-124

Legendre, M., Arslan, D., Abergel, C., and Claverie, J. M. (2012). Genomics of Megavirus and the elusive fourth domain of life. *Commun. Integr. Biol.* 5, 102–106. doi: 10.4161/cib.18624

Levasseur, A., Bekliz, M., Chabriere, E., Pontarotti, P., La Scola, B., and Raoult, D. (2016). MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. *Nature* 531, 249–252. doi: 10.1038/nature17146

Mutsafi, Y., Zauberman, N., Sabanay, I., and Minsky, A. (2010). Vaccinia-like cytoplasmic replication of the giant *Mimivirus*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5978–5982. doi: 10.1073/pnas.0912737107

Pagnier, I., Reteno, D. G., Saadi, H., Boughalmi, M., Gaia, M., Slimani, M., et al. (2013). A decade of improvements in Mimiviridae and Marseilleviridae isolation from amoeba. *Intervirology* 56, 354–363. doi: 10.1159/000354556

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., et al. (2004). The 1.2-megabase genome sequence of *Mimivirus*. *Science* 306, 1344–1350. doi: 10.1126/science.1101485

Rodriguez-R, L. M., and Konstantinidis, K. T. (2014). Bypassing cultivation to identify bacterial species. *Microbe* 9, 111–118. doi: 10.1128/microbe.9.111.1

Saadi, H., Pagnier, I., Colson, P., Cherif, J. K., Beji, M., Boughalmi, M., et al. (2013a). First isolation of *Mimivirus* in a patient with pneumonia. *Clin. Infect. Dis.* 57, e127–e134. doi: 10.1093/cid/cit354

Saadi, H., Reteno, D. G., Colson, P., Aherfi, S., Minodier, P., Pagnier, I., et al. (2013b). Shan virus: a new *Mimivirus* isolated from the stool of a Tunisian patient with pneumonia. *Intervirology* 56, 424–429. doi: 10.1159/000354564

Sambrook, J., and Russell, R. W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd Edn. Cold Spring Harbor, NY: Cold spring harbor laboratory press.

Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, W686–W689. doi: 10.1093/nar/gki366

Schulz, F., Yutin, N., Ivanova, N. N., Ortega, D. R., Lee, T. K., Vierheilig, J., et al. (2017). Giant viruses with an expanded complement of translation system components. *Science* 356, 82–85. doi: 10.1126/science.aal4657

Shackelton, L. A., and Holmes, E. C. (2004). The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol.* 12, 458–465. doi: 10.1016/j.tim.2004.08.005

Suzan-Monti, M., La Scola, B., Barrassi, L., Espinosa, L., and Raoult, D. (2007). Ultrastructural characterization of the giant volcano-like virus factory of *Acanthamoeba polyphaga Mimivirus*. *PLOS ONE* 2:e328. doi: 10.1371/journal.pone.0000328

Takemura, M., Mikami, T., and Murono, S. (2016). Nearly complete genome sequences of two *Mimivirus* strains isolated from a Japanese freshwater pond and river mouth. *Genome Announc.* 4:e01378-16. doi: 10.1128/genomeA.01378-16

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Yoosuf, N., Pagnier, I., Fournous, G., Robert, C., La Scola, B., Raoult, D., et al. (2014a). Complete genome sequence of Courdo11 virus, a member of the family Mimiviridae. *Virus Genes* 48, 218–223. doi: 10.1007/s11262-013-1016-x

Yoosuf, N., Pagnier, I., Fournous, G., Robert, C., Raoult, D., La Scola, B., et al. (2014b). Draft genome sequences of Terra1 and Terra2 viruses, new members

of the family Mimiviridae isolated from soil. *Virology* 45, 125–132. doi: 10.1016/j.virol.2013.12.032

Yoosuf, N., Yutin, N., Colson, P., Shabalina, S. A., Pagnier, I., Robert, C., 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. doi: 10.1093/gbe/evs109

Yutin, N., Colson, P., Raoult, D., and Koonin, E. V. (2013). Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. *Virol. J.* 10:106. doi: 10.1186/1743-422X-10-106

Yutin, N., Wolf, Y. I., Raoult, D., and Koonin, E. V. (2009). Eukaryotic large nucleocytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* 6, 01–13. doi: 10.1186/1743-422X-6-223

Zauberman, N., Mutsafi, Y., Halevy, D. B., Shimoni, E., Klein, E., Xiao, C., et al. (2008). Distinct DNA exit and packaging portals in the virus *Acanthamoeba polyphaga* mimivirus. *PLOS Biol.* 6:e114. doi: 10.1371/journal.pbio.0060114

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.

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

group named nucleo-cytoplasmic large DNA viruses (NCLDV) was proposed, composed of families *Poxviridae*, *Asfarviridae*, *Iridoviridae* and *Phycodnaviridae* [3]. This group gained notoriety two years later with the discovery of *Acanthamoeba polyphaga mimivirus* [4] and it is currently composed of the families mentioned above, as well as *Ascoviridae*, and the more recently incorporated *Mimiviridae* and *Marseilleviridae* [5]. Moreover, other recently discovered giant viruses such as pandoraviruses, faustoviruses and pithoviruses were classified as members of the NCLDV group [6–9]. This group has single features such as large genomes and a diverse gene repertoire, which encode some proteins never identified previously in viruses. Therefore, the creation of a new viral order named 'Megavirales', encompassing all families of the NCLDV group was proposed [5].

This proposed order comprises viruses with large double-stranded DNA (dsDNA) genomes, encoding hundreds of proteins and capable of infecting a wide-range of eukaryotic organisms. These viruses replicate completely or partly, in the cytoplasm of eukaryotic cells and some of them are able to synthesize RNA polymerases (RNA pol), helicases and transcription factors involved in the transcription initiation and elongation steps with lower dependence of the host's transcriptional machinery [3]. The presence of a robust transcriptional apparatus in some Megavirales members, along with a quasi-autonomous glycosylation and translational machinery, especially in mimiviruses, boosted the discussion about the origin and evolution of giant viruses and their genome. Recent evolutionary reconstructions mapped about 25–50 genes encoding essential genes for the probable most recent common ancestor [10]. Concerning the origin of such giant genomes, different hypotheses have been proposed. Some authors suggest a "genome degradation hypothesis", wherein the giant viruses are derived from a cellular ancestor through genome simplification linked to the adaptation to some host lineage [11,12]. Other authors argue in favor of a "genome expansion hypothesis", wherein the giant viruses evolved from a smaller viral ancestor and the universal genes have been independently acquired from their eukaryotic hosts by progressive gene accretion and duplication. According to this theory, the genes of giant viruses have several origins and the origin of giant viruses is probably from a simpler ancestor [13,14].

On the other hand, the accordion-like model of evolution proposes that there is no trend of genome expansion or general tendency of genome contraction. Instead, viruses evolving by constant gene gain and loss originated from an ancestor giant virus [10]. All these theories are often contradictory and have stimulated discussion about the establishment of a fourth domain of life where the giant viruses of the proposed order Megavirales were suggested to share a common ancestral origin based on analyses of their sequences and gene repertoires and compose a new domain aside Bacteria, Archaea and Eukarya [14–16].

During the last years, a huge effort has been made to better understand the virus–host interaction on many levels. One of the most interesting research fields is how the viruses can explore host transcriptional machinery to express their genes. Nevertheless, it is important also to look into the transcription process of the cellular organisms. The upstream regions of eukaryotes and prokaryotes genes have been studied in different organisms in an attempt to discover sequences associated with the regulation of the transcription process. The same has been done for viruses, especially considering the proposed Megavirales order, where some putative promoter sequences have already been described. 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 current knowledge of the promoter regions of all members within Megavirales order. Finally, we discuss how the analysis of the promoter sequences found in giant viruses provides new insights into the evolutionary history of these complex and intriguing agents.

2. Gene Expression in Cells

In all cells, thousands of genes encoded in the DNA are transcribed into RNA and for the efficient occurrence of this process, multiple events must be triggered. In eukaryotes, the genome is coupled to histones and other proteins, forming the chromatin compact complex. Since wrapping

DNA around histones blocks the access to the genetic information, decondensation of DNA is required, to allow physical access to the the gene locus and the transcription initiation machinery formation [17–19]. The transcription initiation machinery is formed over a region of the genome, the promoter. The promoter is typically located 40 bp upstream and downstream of the transcription start of a gene, called transcription start sites (TSS). Several transcription factors mediate the transcription machinery assembly on the promoter region. There are thousands of transcription factors involved in the transcription process, such as TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH that recognize and bind the promoter region, called the core promoter, and recruit RNA polymerase (RNA pol) [20]. Eukaryotes have five types of RNA pol (I to V). RNA pol I transcribes ribosomal RNA, whereas the type II is the best characterized one and responsible for transcribing genes encoding proteins, and several noncoding RNA classes [18,21,22]. RNA pol III transcribes genes encoding short, untranslated RNAs, such as tRNAs, 5S ribosomal RNA (rRNA) and the spliceosomal U6 small nuclear RNA (snRNA) [23]. RNA pol IV and V transcribe siRNA in plants [24].

One classical element of the core promoter is the TATA-box, which is a consensus sequence (TATAAAT) located at −25 to −30 bp upstream of the TSS. Although the TATA-box sequence is a well-known promoter core motif, it is present only in a minority of mammalian promoters. This sequence is commonly associated with tissue-specific gene transcription and high conservation within species [25,26]. Other eukariotic promoter elements are Initiator (Inr); Downstream Promoter Element (DPE), Core Element Downstream (CED), TFIIB-Recognition Element (TRE), and Motif Ten Element (MTE) [20,27,28]. Together, these components act synergistically to increase transcription efficiency by providing recognition sites for transcription factors, and indicate the direction of transcription and also the DNA strand to be transcribed [20]. The transcription starts with the binding of the TFIID to the TATA-box region, the Inr sequence and/or other core promoter elements [27]. TFIID is a multiprotein complex comprising the TATA-box binding protein (TBP) and more than 10 different TBP associated factors (TAFs) [22]. After binding TBP to the TATA-box motif, the RNA pol II is recruited, and the transcription is triggered (Figure 1A).

Nevertheless, the transcription in eukaryotes is a much more complex process than previously thought and various strategies are used to increase the diversity of transcripts produced. Among mammals, previous analysis has shown that a large proportion of protein-coding genes (58%) use alternative promoters during transcription [25]. These alternative promoters may have different combinations of core promoter elements to increase the variability of transcripts [20,29,30].

There are many differences between the transcription process of eukaryotic and bacteria cells. The bacterial transcription is much simpler compared to the eukaryotic process since the transcription occurs using a single type of RNA pol and there are no transcription factors [31]. This enzyme is capable of synthesizing RNA from a DNA template, but it is unable to locate the promoter and transcription initiation site. Thus, a key factor to transcription is the free subunit named σ (sigma), which is responsible for recognizing the promoter region (Figure 1B) [32,33]. Although the majority of nucleotides within bacteria promoters vary in sequence, several short motifs are conserved. These include the hexamer (TATAAT), located 10 base pairs (bp) upstream of the TSS and is recognized by domain 2 of RNA pol σ subunit. Another motif is the the hexamer (TTGACA), located 35 base pairs (bp) upstream of the TSS and recognized by domain 4 of the RNA pol σ subunit [31,34,35]. In Archaea, there is a mix of eukarya and bacteria translational apparatus. Just as in eukaryotes, the archaea RNA pol is not able to recognize promoter sequences by itself and at least two transcription factors analogous to TBP and TFIIB are required [36–38]. The archeal TBP also recognizes specifically an AT-rich sequence, homologous to the TATA-box region of eukaryotes [39,40]. Although archaea transcription machinery is similar to that of eukaryotes, the characterization of transcription regulators of some archaeas showed that most of the transcriptional regulation in archaea is done by "bacterial-like" regulators, as two homologues of bacterial leucine-responsive regulatory protein (Lrp)—Lrs14 and Sa-Lrp and metal-dependent repressor 1 (MDR1) homologous to bacterial metal-dependent regulators (Figure 1C) [41–43].

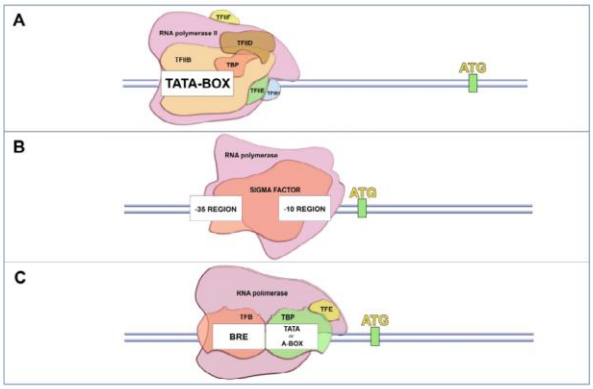


Figure 1. Main features in the transcription initiation machinery presented in the canonical Domains of Life. (**A**) In Eukarya, several components, called general transcription factors (represented as TFIIB, TFIID, TFIIE, TFIIF, TFIIH and TBP), are responsible for assembling over a region called the promoter, where they recruit an RNA polimerase to initiate the transcription process. A classical promoter presented in this group is the TATA-BOX region, located at the positions −25 and −30 from the initial transcription site; (**B**) In Bacteria, the sigma factor recognizes and recruits the RNA polimerase over the promoter regions. These regions are well conserved over the positions −35 and −10 upstream of the initial transcription site; (**C**) Archaea present a mixture of the transcription apparatus of the two other Domains. While the machinery itself is similar to that found in eukaryotes (the general transcription factors, a homologous TATA-BOX region and the RNA polimerase), the archaeal transcription regulators, activators and repressors are homologous to the bacterial ones.

Hypotheses regarding the evolutionary history of translational machinery among the living organisms have been raised during the last years, but the theme is still under debate [44]. Even considering the most recent proposals, the translational process of viruses remains out of the discussion, basically because these organisms are traditionally excluded from the canonical tree of life. However, this scenario has been changing since the discovery of giant viruses [16]. Therefore, it becomes interesting to examine if NCLDV members share similar transcription initiation strategies that could bring insights about how this correlates to giant viruses' evolution.

Gene Expression in NCLDVs

In contrast to cellular genomes, which are formed by dsDNA, viral genomes show a large diversity genome composition, structures, replication and transcription strategies with great implications in virus biology, as virus–host interactions [45]. The majority of the RNA viruses employ virus-coded specific enzymes (RNA-dependent RNA polymerases) to synthesize and modify their mRNA. DNA viruses showing small and intermediate size genomes such as the parvoviruses, papillomaviruses, and adenoviruses, depend on host-cell enzymes for transcription, including the RNA pol [45]. However, viruses with a large genome such as the giant viruses, mostly encode their transcriptional apparatus, which make them relatively independent from their host transcription machinery [15,46].

The transcription of a typical large DNA virus occurs in a temporal pattern in the host cytoplasm (Figure 2). At the start of infection, a subset of immediate early viral proteins is required for DNA replication and host cell manipulation [47,48]. The early mRNAs also encode enzymes and factors needed for transcription of the intermediate genes. Concomitantly with the expression of intermediate genes, the expression of the early genes is often repressed. Finally, late genes are transcribed, directing the synthesis of structural proteins, non-structural proteins and enzymes present in the mature particle

required for viral assembly [45,48]. The efficient transcription of late mRNA usually depends on intermediate gene products, as well as cellular transcription factors that may differ from those used by the early promoters. The products of the late genes include the immediate early transcription factors, which are packaged along with RNA pol and other enzymes within the virus progeny [47–50].

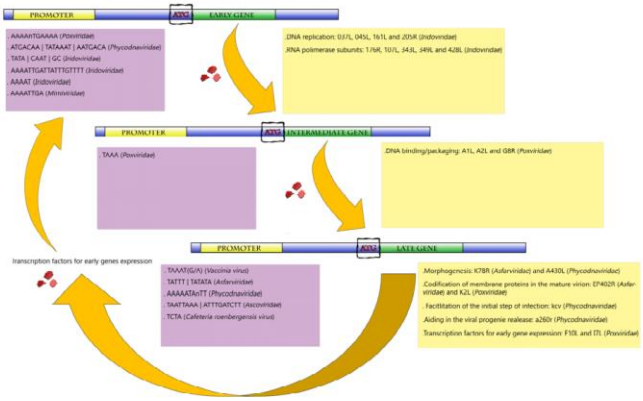


Figure 2. Representative scheme of the temporal gene expression in NCLDVs. During initial times of infection, the expression of genes related to the metabolism of nucleic acids is primarily activated (early and intermediate genes). After DNA replication, the activation of late genes is initiated. Those genes are involved in the production of viral structural proteins, in transcription factors used for early gene expression and also in proteins that facilitate the initial step of infection of the viral progeny in the next round of multiplication. Purple boxes represent the promoters described for giant viruses according to each gene category (early, intermediate and late genes). Yellow boxes exemplify the biological functions involved in each category, with some genes represented inside the parentheses.

This ability to regulate temporally the transcription of genes is characterized as an evolutionary advantage. This strategy is possible due to the presence of promoter codes that dictate when, where, and at what level the classes of early, intermediate, and late genes are transcribed [45,48]. These promoter sequences are different between the three genes classes, but there is a pattern of conservation within the same group. This indicates that during the evolution the gene promoters were selected to ensure the temporal gene expression, and therefore ensure the gene expression in the host cell during its replication [45,47,48,50].

In the following sections, we look closer at how the gene transcription is carried out in each family of the proposed Megavirales order, focusing on the current knowledge about the promoter sequence of these viruses.

3. *Poxviridae* Family

Among NCLDVs, the *Poxviridae* family is one of the most studied. These viruses have enveloped ovoid particles of around 200 nm in diameter and 300 nm in length and present a linear dsDNA genome of approximately 200 kbp coding nearly 200 open reading frames (ORFs). Poxviruses can infect a wide range of hosts, such as insects, birds, and mammals [48,51]. Extensive study of the poxvirus genome and replication cycle allowed a detailed identification of its promoters, as well as important transcription factors. Poxviruses possess their own DNA-dependent RNA polymerase (RNA pol) that is very similar to the eukaryotic protein, regarding size and subunit complexity. In the case of *Vaccinia virus* (VACV), a poxvirus prototype, the enzyme subunits are encoded by eight viral VACV genes

which, in most cases, are homologous to cellular RNApol [52,53]. Gene transcription in poxviruses follows a typical temporal profile regulated by well-conserved promoters of early, intermediate and late genes (Figure 2) [47,48].

The transcription of early genes is characterized by an A/T-rich motif upstream of transcriptional start site with a critical core region located from −13 to −25 to that region. Figure 3 illustrates the promoter motifs described in megavirales members. The representative consensus sequence of the early promoter region is 'AAAANTGAAAA'. Mutagenesis in this promoter region of VACV causes a drastic negative effect on VACV gene transcription [54]. The intermediate genes are transcribed after DNA replication, before the transcription of the late genes. The intermediate core promoter is similar to the early promoter due to the A/T-rich content, but its specific sequence is given by the tetranucleotide 'TAAA'. Furthermore, the intermediate promoter sequence has a bipartite structure presenting a core and an initiator region with similar sequences (TAAA) [55–57]. Three (*A1L*, *A2L*, and *G8R*) of the 53 genes that compose the set of intermediate genes encode transcription factors that are directly related to the late stage of the replication cycle, important to DNA binding/packaging processes and to core-associated proteins [58].

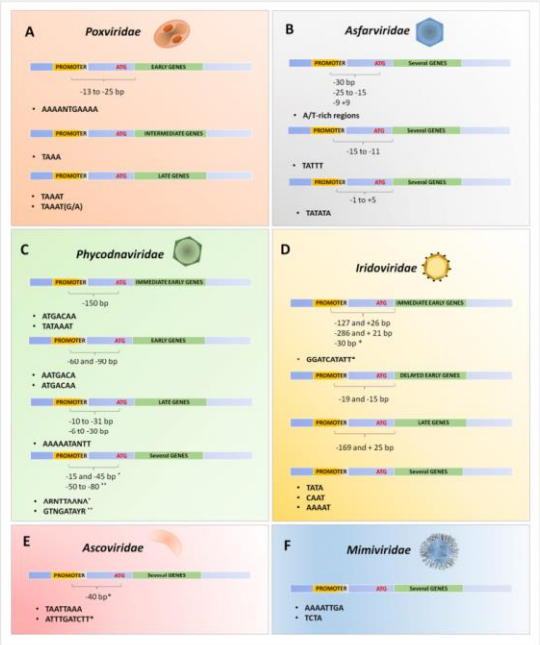


Figure 3. Schematic representation of the promoter's sequences described for different NCLDVs. Compilation of the described promoters for some viral families belonging to the proposed order Megavirales: *Poxviridae* (**A**); *Asfarviridae* (**B**); *Phycodnaviridae* (**C**); *Iridoviridae* (**D**); *Ascoviridae* (**E**) and *Mimiviridae* (**F**). Each promoter was related to the expression of immediate early, early, delayed early, intermediate and late genes, or related to the expression of genes independent of temporal expression (several genes). The distances between the transcription start site or translate start site (ATG) until the promoters are also indicated by brackets.

The transcription of late genes persists until the end of the replication cycle. Around 38 late genes have already been identified, with their main functions related to the codification of membrane proteins in the virion, morphogenesis steps, and also to the production of immediate early transcription factors [57,59]. Most of them are clustered in the central region of the poxviruses genome and also have A/T-rich sequence promoters. These regions consist of a core sequence of about 20 bp with some 'T' residues, separated by a region of about 5–7 bp of a conserved 'TAAAT' motif, which regulates the transcription initiation. Usually, G or A follows the late promoter sequence, performing a 'TAAAT (G/A)' transcription initiation sequence. This sequence is conserved among VACV late promoters, overlapping the site of transcription initiation that is absent in 5' untranslated regions (5'-UTR) [48,54]. Mutations within this conserved element were demonstrated to cause complete inactivation of the promoter, and almost 25% of the 'AAA' sequences are used as transcription initiation sites in VACV. Along with other factors, the viral RNA pol directs the synthesis of late mRNAs, finishing the transcription process [54,60–62].

The presence of complete transcriptional machinery in poxviruses allows a lower dependency of these viruses on their hosts. It permits that the mRNA transcription totally occurs in the host's cytoplasm, right after the virus entry. Addionally, the presence of well conserved promoter regulatory sequences in different poxviruses suggests a conserved evolutionary pattern among them. It is likely that such a complete transcriptional set was already present in their ancestor and was maintained over time. Alternatively, the presence of a robust transcriptional apparatus in all members of the *Poxivirdae* family might be a result of evolutive convergence. Although less parcimonious, the different poxviruses might have had different evolutionary histories regarding the transcription process, including both protein-related elements and promoter sequence regions, but in the course of evolution, they became more similar to each other. It is not yet possible to determine which hypothesis is the correct, or even if other possibilities correspond to the real history of these complex viruses, and this discussion shall continue for a while.

4. *Asfarviridae*

African swine fever virus (ASFV), a large (~200 nm), icosahedral, and enveloped virus is currently the single member of the *Asfarviridae* family, infecting members of the *Suidae* family (pigs, hogs and boars) [63]. The genome is composed of a linear dsDNA molecule of approximately 170 kbp with terminal inverted repeats. It encodes approximately 150 ORFs separated by short intergenic regions [64,65]. ASFV encodes its own RNA pol and all ASFV genes are transcribed by its enzyme [66,67].

Similar to poxviruses, the ASFV gene transcription follows a temporal profile, where immediate early and early genes are expressed before the DNA replication that is followed by the expression of intermediate, late and immediate early genes. Transcription initiation and termination occurs at very precise positions in the genome, encoding a several genes involved in the transcription and modification of viral mRNAs. The transcriptional machinery of ASFV provides an accurate temporal control of gene expression regulated by cis-DNA elements, enhancers, and promoters together with a structural complexity of transcription factors [68]. Analysis of the base composition of the intergenic regions shows that they are rich in A/T sequences, similar to that observed in poxviruses [69–71]. A/T-rich regions located at approximately −30 bp upstream of the ATG translation start site are essential for the expression of the K9L gene, which encodes a protein with similarity to mammalian transcription elongation factor IIS [72]. Furthermore, upstream sequences presented in two intermediate genes exhibit highly conserved sequences at positions −25 to −15, and −9 to +9 to the translational start codon [70]. Experiments involving genetic deletions, linker scan substitutions and point mutations in the promoter sequence of the *p72* gene (major capsid protein) revealed that the replacement of the A/T-rich region by G/C residues strongly reduced the transcription rate, demonstrating the importance of this sequence for efficient late viral transcription [71].

Two other major essential regions for promoter activity are described: one region is located at position −15 to −11 upstream of the transcription start site (TATTT); and the second region at positions −1 to +5 (TATATA) [71]. Mutants presenting the 'TATATA' motif replaced by a G/C-rich sequence had the promoter activity completely abolished, suggesting that ASFV transcription is dependent on such sequence at (or near) the region of transcriptional initiation, similar to what is found in other large viruses [71]. The replacement of the equivalent 'TATATA' sequence on the late genes *K78R*, *EP402R* and *A137R* by the 'GCGC' motif was also demonstrated to be deleterious, suggesting that the A/T-rich sequence could be a motif for late promoter function as well [68,71]. Interestingly, the bipartite structure seen in the late promoter of ASFV is similar to the late and intermediate promoters in poxviruses that contain a core and an initiator region [54,55,62,71]. The similarities found in the transcriptional strategies reinforce the genetic data, indicating a close relationship between poxviruses and asfavirus, pointing to a common ancestor for both viral families.

5. *Phycodnaviridae*

The phycodnaviruses are large and icosahedral viruses (~100–220 nm), with dsDNA genomes ranging from 180 to 560 kbp [73]. Since they infect a diverse group of eukaryotic algae, they are one of the most important groups of organisms regulating the oxygen cycle in the Earth [74,75]. The family *Phycodnaviridae* consists of six genera, named according to the hosts that they infect: *Chlorovirus*, *Coccolithovirus*, *Prasinovirus*, *Prymnesiovirus*, *Phaeovirus*, and *Raphidovirus* [76]. As demonstrated by other giant viruses, the phycodnaviruses exhibit a temporal transcription profile. Early genes are transcribed within 5 to 60 min post-infection (p.i), and transcripts of late genes begin to appear around 60–90 min p.i. However, some early genes can also be detected in later stages of infection [77,78].

The presence of A/T-rich promoters was also observed in phycodnaviruses. Analysis of the *kcv* gene, encoding a potassium ion channel protein in chlorella viruses, revealed a highly conserved 10-nt sequence (AAAAATANTT) in the promoter region of this gene, present in 16 out of 17 chlorellaviruses [77]. This sequence is located at 10–31 nucleotides upstream of the ATG translation start codon in all of the analyzed viruses, and it was associated with late gene transcription, since, apparently, *kcv* transcripts are produced during the late steps of infection. Furthermore, the region that precedes seven genes expressed at later times during the *Paramecium bursaria chlorella virus 1* (PBCV-1) replication cycle (*a85r*, *a237r*, *a248r*, *a260r*, *a292l*, *a430l*, and *a530r*) contain the same sequence or at least a subset of this sequence located at 6–30 nucleotides upstream of the ATG start codon [77]. The study of immediate early genes expressed in chlorovirus infections also revealed A/T-rich sequences as putative promoter regions. Two sequences determined by 'ATGACAA' and 'TATAAAT' (such as the eukaryotic "TATA-box") were located in a 150 bp region from the translation start codon in the upstream regions of almost all immediate early genes (20 of 23 studied) [78]. These elements, especially 'ATGACAA', were absent in all genes so far examined, expressed after 40 min p.i, including *A122R* (Vp260) [79], *A181-182R* (chitinase), *A292L* (chitosanase) [80], *A430L* (major capsid protein) [81], *vAL-1* [82].

Bioinformatics analysis revealed highly conserved nucleotide sequences in putative promoter regions involving three different chlorella viruses: PBCV-1, virus MT325 [83], and *Paramecium bursaria chlorella virus* NY-2A [84]. Three putative AT-rich sequence promoters, comprising seven to nine nucleotides (ARNTTAANA, AATGACA and GTNGATAYR), located at 150-nt upstream of the translation start codon of many ORFs were observed [85]. The 'ARNTTAANA' sequence is found between nucleotides −15 and −45 relative to the ATG translation start codon. This sequence occurs in the promoter region of 25% of PBCV-1 genes, 22% of NY-2A genes and 12% of MT325 genes. Regarding the entire genome, this sequence is present within the 200-nt promoter region during 44% of the time in PBCV-1, 49% of the time in NY-2A, and 37% of the time in MT325. The hotspot for the presence of the 'AATGACA' sequence is located between nucleotides −60 and −90 from the translational start codon. This sequence occurs in the promoter region of 16% of the PBCV-1 genes, 18% of NY-2A genes and 8% of MT325 genes. Regarding the entire genome, this sequence is present within the 200-nt promoter region in 54% of the PBCV-1 genes, 53% of the NY-2A genes, and 25% of the MT325 genes [85].

The 'AATGACA' sequence in PBCV-1 is associated with early genes during the replication cycle [85]. This sequence is very similar to a motif previously identified in some chlorella viruses (ATGACAA), which is also correlated with early transcripts [78]. Finally, the 'GTNGATAYR' sequence is mainly located at nucleotide positions −50 to −80 from the ATG initiation codon, occurring in the promoter region of 13% of PBCV-1 genes, 14% NY-2A genes, and in 11% of MT325 genes. Regarding the entire genome, this sequence is found specifically within the 200-nt promoter region in 28% of the PBCV-1 genes, 22% of the NY-2A genes, and 21% of the MT325 genes [85].

Unlike other members of the NCLDVs, phycodnaviruses do not encode their own RNA pol and need to appropriate the host's RNA pol to properly make their transcripts [86]. However, uniquely for the *Phycodnaviridae* family, *Emiliania huxleyi virus 86* (EhV-86), a coccolithovirus that infects the marine calcifying microalga *Emiliania huxleyi*, contains a total of six RNA pol subunits, which suggests that this virus partially encodes its own transcription machinery [87]. Although these viruses present some important elements for the mRNA synthesis, it is not possible to state that they have their own transcriptional complete apparatus, at least for the majority of them. Therefore, concerning the transcriptional process, the phycodnaviruses seem to present a different evolutionary history.

6. *Iridoviridae*

The *Iridoviridae* family is composed by five genera: *Ranavirus*, *Megalocytivirus* and *Lymphocystivirus* that infect vertebrates; *Iridovirus* and *Chloriridovirus* that infect invertebrates [88]. Iridoviruses have a linear dsDNA genome varying from 105 to 212 kbp, coding between 92 and 211 putative proteins. They present a non-enveloped icosahedral particle of 300 nm in size [89–92]. These large viruses also display a pattern of temporal gene expression regulation, wherein the genes are divided into three classes: immediate-early (IE or α), delayed-early (DE or β), and late (L or γ) genes [93–95]. Iridoviruses are typical nucleo-cytoplasmic viruses. They begin the replication cycle in the nucleus, followed by the second phase of genome replication in the cytoplasm [90].

Gene transcription and promoter sequences studies have been performed for only a few genes in members of the *Iridoviridae* family. The study of promoter sequences in iridovirus is focused mainly in the *Ranavirus* genus (using type species *Frog virus 3* (FV3)) and *Iridovirus* genus (using type species *Invertebrate iridescent virus 6* (IIV-6)), the type species of the *Iridovirus* genus. Notwithstanding, both the gene expression and promoter sequences studies have been performed for only a few genes in the *Iridoviridae* family. The most complex studies were performed with immediate-early *ICR-169* and *ICR-489* genes of FV3 [96,97]. Those studies revealed the importance of a 78 bp sequence before the transcription start site of an IE gene of the FV3 promoter. It was shown that an FV3 protein acts in trans to induce the transcription of the major FV3 IE gene, *ICR-169*, and is dependent on the 78 bp sequence located at the 5′ position from the start site of the transcription of this gene [98]. Two years later, the same group demonstrated that a 23 bp sequence was possibly a critical *cis*-regulatory element for the occurrence of FV3 *trans*-activation, since a significant reduction of transcription occurred after its deletion, located at the 5′ region, showing the sequence 'ATATCTCACAGGGGAATTGAAAC' [96]. Despite the importance of the approximately 23-nt sequence upstream of the transcription start site in the IE *ICR-169* gene of FV3, this sequence had no similarity with the promoter region of the intermediate gene *ICR489*. This lack of similarity indicated that the contemporary regulation of these two promoters is not controlled by sequences upstream of the start point of transcription [97]. It is worthy to note that in the *ICR489* gene, in an upstream region, 'TATA', 'CAAT', and 'GC' motifs were identified, which are similar to those of typical eukaryotic promoters [97].

Another study analyzed three genes—two early (*ICP-18* and *ICP-46*) and a late one [major capsid protein (MCP)] of *Bohle iridovirus* (another *Ranavirus* member)—looking for conserved regions to be considered as regulatory elements [99]. The authors demonstrated that all gene promoters included sequences located 127 to 281 bases upstream of the transcription initiation site (127 pb or ICB-18, 281 pb for ICP46, and 169 pb for MCP), but also sequences located from 21 to 26 bases downstream of this site (26 bases for *ICP-18*, 21 bases for ICP 46 and 25 bases for MCP) [99].

Moreover, a detailed study conducted in the following years identified an essential 'AAAAT' motif in a DE gene of IIV-6 (*Iridovirus*) [100]. The authors described a sequence of 19 bp (AAAATTGATTATTTGTTTT), located between −19 and −2 relative to the mRNA transcription start site, which is the putative region responsible for promoter activity of the DNApol gene. Deletions and point mutations in the DNApol promoter of IIV-6 showed that each of the 5-nt of 'AAAAT' motif located between −19 and −15 were equally essential for promoter activity. Mutations at the downstream side had a lower effect, but the role of individual nucleotides positioned at −14 to −5 was not analyzed in this study [100].

It is noteworthy that the same critical 'AAAAT' motif was found in the 100-nt upstream of the putative translational start codons of several other putative DE IIV-6 genes [91]. In *Invertebrate iridescent virus 3* (IIV-3), many homologues of these genes also presented the 'AAAAT' motif in proximity to their start codon. A great similarity was also found between the region upstream of the DNApol ORF and the corresponding region in 12 iridovirus genomes [101]. Eight of these genomes showed a similar 'AAAAT' motif in the DNApol upstream region and three sequenced ranavirus genomes also shared the related 'TAAAT' motif in their DNA pol promoter region, which may indicate a conserved regulation of DE promoter activity in iridoviruses [101].

A study that targeted a IE gene (012L) of IIV-6 showed that the transcription start site is located 30-nt upstream of the ATG translational start codon. Analyzing mutants (produced by deletion), it was established that the intergenic region located between −21 and −10 (GGATCATATT) upstream of the transcription start site comprised the promoter sequence promoter 012L gene. This type of sequence was not observed in upstream regions of other IE genes of IIV-6, such as 468R, 006L and 010R. The 'TATA' and 'CAAT' sequences were also identified in the intergenic region of this gene, as well as sequences similar to the 'AAAAT' motif described to the DNA pol gene, but this sequence had no promoter activity for the 012L, differently than demonstrated for the DNA pol gene. The 037L and 012L genes of IIV-6, both early genes, do not share conserved key promoter motifs. However, DNA pol is considered a DE gene and 012L an IE gene [100,102].

Despite the presence of homologs of RNA pol subunits in the iridoviruses genome, host RNA pol II is required for the synthesis of *Ranavirus* IE transcripts, and it is likely that the same is true from *Iridovirus* IE genes, contrasting to pox- and asfaviruses [103–106]. It has been proposed that the RNA pol subunits found in members of the *Iridoviridae* family are probably involved in the cytoplasmic phase of transcription in later stages of infection [91,107]. Such a paradox may reflect the long co-evolution period that these viruses had been through. It is possible that the ancestor of iridoviruses presented a complete transcription apparatus, but some elements were lost due to the adaptation to a more parasitic lifestyle. Other possibilities are the occurrence of events of horizontal gene transfer (HGT) between the viruses and their hosts. However, the lack of information about such events involving members of the *Iridoviridae* family prevents further insights into this alternative for the evolution of the transcription apparatus of these viruses.

7. *Ascoviridae*

The *Ascoviridae* family has two genera that include *Ascovirus*, with three species including *Spodoptera frugiperda ascovirus 1a* (SfAV-1a), the prototype of the genus, *Trichoplusia ni ascovirus 2a* (TnAV-2a), and *Heliothis virescens ascovirus 3a* (HvAV-3a), and the *Toursvirus* genus, with only one representative, *Diadromus pulchellus ascovirus 4a* (DpAV-4a) [108,109]. Ascoviruses are enveloped viruses, 300–400 nm long by 100–150 nm in diameter, with a circular dsDNA genome with sizes ranging from 116 to 185 kb, infecting arthropods, mainly lepidopterans [110–112].

The studies regarding the ascoviruses are still in their infancy. Information about the replication and more specifically, the transcription process, are extremely scarce. The current knowledge about transcription in ascoviruses come from the analyses of the *Ascovirus* genus [110,113]. A study performed using a possible variant of HvAV-3, the *Spodoptera exigua ascovirus 5a* (SeAV-5a) showed that the 5′-UTR region of the SeAV-5a MCP gene is composed of 25-nt [114]. The upstream region

of this gene does not present a typical eukaryotic class II promoter motif sequence 'TATAAAT' (TATA box). However, the putative 5′ transcription control region of the SeAV-5a MCP gene shares similarities with other ascoviruses and iridoviruses, containing a conserved TATA-box like motif (TAATTAAA) and an 'ATTTGATCTT' motif within 40-nt upstream of the translation initiation codon ATG [114]. The 'TAATTAAA' and 'ATTTGATCTT' motifs are located downstream and upstream of the transcription initiation site, respectively. Furthermore, the ORF p27 presents a similar 5′ downstream transcription promoter region, suggesting that such a region might be a truly regulatory sequence within ascoviruses [114].

Sequences from the promoter regions of the MCP genes from ascoviruses and IIV-6 (late genes), showed that ascoviruses and iridoviruses are closely related in this aspect, suggesting that the transcription regulation could be maintained during the viral evolution process in closely related viruses [115,116]. Furthermore, phylogenetic studies showed that ascoviruses probably evolved from the iridoviruses [116–118]. It is possible that the same pattern of temporary gene expression exhibited in iridoviruses (and the other members of proposed Megavirales order) was conserved in the ascoviruses lineage, and that such a mechanism might have been present in their common ancestor.

8. *Mimiviridae* and Other Amoebal Giant Viruses

The discovery of mimiviruses in 2003 and the establishment of the *Mimiviridae* family astonished the scientific community, making the term 'giant virus' more appropriated than ever. These viruses have particles visible in light microscopy, with sizes of ~700 nm in diameter. Viral particles have characteristics never described before in the virosphere, such as long proteic fibrils (~125 nm in length) immersed in a peptidoglycan matrix, and a star-shaped face, named stargate, responsible for the releasing of the genome inside the cytoplasm of their host (*Acanthamoeba* genus) [4,119–121]. The genome is a linear dsDNA molecule of about 1.2 Mbp, coding more than 1000 proteins, including a large set of transcriptional elements [15,122].

Similar to other NCLDVs members, mimiviruses genes can be divided into early, intermediate and late categories according to three major temporal classes of transcription determined by mRNA deep sequencing [49]. The analysis of the intergenic regions of *Acanthamoeba polyphaga mimivirus*, the prototype species of *Mimivirus* genus, showed a conserved 'AAAATTGA' motif in nearly 50% of genes [50]. The intergenic regions of the genome of mimiviruses have an average size of 157-nt. In silico analyses showed that the conserved 'AAAATTGA' motifs are present within the 150-nt upstream regions of the translation start codon in 45% of all predicted mimivirus genes [50]. This motif is mainly associated to early (or the late-early) genes during the viral infectious cycle, and it is absent from the upstream regions of mimivirus late genes, such as DNA replication and particle morphogenesis and assembly. It is noteworthy that similar sequences were described regulating the early genes in other giant viruses, such as iridoviruses and phycodnaviruses, as described in the topics above. Besides the early promoter sequence, another A/T-rich motif (two 10-nt informative segments separated by a highly degenerated 4-nt sequence) was identified as a putative late promoter within mimiviruses, which is present in 24.2% of the considered late class genes. To the best of our knowledge, an intermediate promoter sequence has not already been described in mimiviruses [49,50].

In a distant relative, the *Cafeteria roenbergensis virus* [CroV (*Cafeteria* genus)], *Mimiviridae* family; the same early promoter motif was identified in the upstream region of 35% of genes [123]. However, considering the late promoter motif, this virus exhibits a different putative regulatory sequence compared to other mimiviruses, wherein the 'TCTA' tetramer flanked by A/T-rich regions on either side was found in the 5′ upstream of 124 late genes [123]. Moreover, CroV present eight RNA pol II subunits, six transcription factors, several helicases, among others, indicating the presence of nearly complete transcriptional machinery. This feature seems to be a mark to all members of the *Mimiviridae* family, which suggests that such a robust transcriptional apparatus was already present in the last common ancestor.

After the discovery of mimiviruses, other giant viruses infecting amoebae were described, such as marseilleviruses, which is currently classified in the family *Marseilleviridae* [124]. Other viruses have also been isolated but still not properly classified, namely faustoviruses [125], pandoraviruses [8,126], phitoviruses [127,128] and mollivirus [129]. Although these viruses are not yet officially recognized by the ICTV, they are genuine members of the NCLDVs [6,7,9]. In all of these giant viruses, a set of transcriptional elements has already been identified, including many RNA pol subunits, indicating a nearly autonomous process in these viruses. However, analysis of promoters and studies aiming to understand how gene expression is regulated in those newly discovered viruses remain to be performed.

9. MEGA-Box: A Putative Promoter Region in the Common Ancestor of Megavirales

The proposed Megavirales order comprises viral families that exhibit some unique features that allow their clustering into a monophyletic group [5]. In addition to some core genes that are shared among these viruses, they present other similarities, such as a temporal transcription profile. As described above, all viruses present elements to the transcriptional apparatus, most of them reaching up to the independence from their host in this step of the viral life cycle. Also, the presence of an A/T-rich promoter sequence has been described in many representatives of each family, even in those in which the genome presents a high G/C content. More interesting is the fact that some promoter sequences found in one family are very similar to others found in their relatives (Figure 3). This fact suggests that a possible common ancestor of the Megavirales order likely had an A/T-rich promoter sequence. More interesting is the fact that some promoter sequences found in one family are very similar to others found in their giant relatives. This fact suggests that such a common ancestor of Megavirales likely had an A/T-rich promoter sequence.

The origin of the members of the Megavirales order is still under debate, but the evolutionary history of some of its members is already being told, at least concerning genome evolution. The first members to be analyzed were the poxviruses. It has been demonstrated by phylogenetic analysis based on the presence/absence of genes that genomes from this family have been subject to frequent events of gene duplication, deletion, and HGT from their hosts. Many of these genes can interfere with host immune signaling, such as homologues of cytokines receptors which could confer some advantages in the interaction with the hosts [130–132]. By analyzing the poxviruses' closest relative, ASFV, it seems that it has been through the same pattern of evolution, at least considering the multigene and p22 gene families [133,134].

The "accordion-like" pattern of evolution was also identified in different members of the *Iridoviridae* family. It is particularly interesting the fact that iridoviruses infecting the same host-range exhibited a similar pattern of gene gain and loss, but this was slightly different when the viruses infected different hosts (fish vs. insect-infecting viruses), suggesting that such a pattern was driven by host–virus co-evolution [135]. Finally, the same evolutionary model for members of the families *Phycodnaviridae* and *Mimiviridae* has recently been described. The genomic comparisons of closely related viruses belonging to the *Mimiviridae* and *Phycodnaviridae* families show that genomes accumulating genomic mutations occur on successive cycles of genome expansion and reduction. In addition, there is no general tendency of genome expansion or contraction. Each family exhibits a specific pattern for gene acquisition, which might be a reflex of interaction with distinct hosts [10]. Since these viruses seem to exhibit a similar pattern of genome evolution, it is possible that a similar scenario has also happened with their promoter sequences. In the same way, it is reasonable to consider that NCLDVs' common ancestor evolved by the same "accordion-like" pattern, and thus it presented a promoter region that underwent an analogous mechanism.

Considering a common origin for the NCLDVs, a possible scenario is that the Megavirales' common ancestor presented a 'TATATAAAATTGA' promoter motif, which we named here as the "MEGA-box" (an allusion to the conserved TATA-box promoter found in cellular organisms). Over time, with the Megavirales' order radiation, the MEGA-box has been gradually evolved by nucleotides'

gain and loss, analogously to that reported for the entire genome, which evolved through gene gain and loss. The MEGA-box was slightly modified in the poxviruses lineage, at least concerning the early promoter motif. Considering the intermediate and the late promoter motifs of poxviruses, if they truly came from the MEGA-box, this could have happened through a series of nucleotide loss. However, it is also possible that the emergence of other promoters, rather than the early one, have emerged after the establishment of the poxvirus' lineage, thus not originating from the ancestral promoter sequence. The same might be true for mimiviruses, phycodnaviruses and iridoviruses. Considering asfavirus and ascoviruses, their promoter sequences might have originated from the MEGA-box through successive gain and loss of nucleotides. However, another scenario is also possible, wherein their promoter motifs emerged from the poxviruses and iridoviruses lineages respectively (closest evolutionary groups). This scenario is in agreement with the proposition that the Megavirales' ancestor was already a giant virus with a large genome [10]. In this aspect, the giant ancestor also had a large promoter sequence that evolved through constant nucleotide gain and loss, a pattern analogous to the accordion-like model of genome evolution. However, other scenarios are also possible, although less probable, considering the evolutionary data currently available for these viruses. One is that the ancestor had a very short promoter sequence, like a poxvirus intermediate promoter (TAAA), that underwent massive nucleotide gain over time, leading to very large promoter sequences in the majority of the giant viruses. Another one is just the opposite; wherein the ancestor had a very large promoter region that had been losing nucleotides during evolution. A third pathway, equally unlikely, would be the acquisition of promoter sequences by horizontal/lateral transfer. Similar to different genes, the MEGA-box promoter evolutionary pattern during the radiation of NCLDVs members could be related to the co-evolution with different hosts over time.

Whether the NCLDVs came from a simple entity [14,136], or from an already complex organism [10,16,137], is still under debate. Despite this, increasing evidence that they originated from a common ancestor is emerging, and it suggests that such an ancestor evolved through an "accordion-like" pattern. By analyzing the promoter regions currently known for different giant viruses, we provide another piece of evidence to support this statement. Further, we propose how a conserved A/T-rich promoter sequence was present in the possible common ancestor, which might have evolved by continuous gain and loss of nucleotides, in addition to some point mutations in the MEGA-box original sequence. Other scenarios could also be discussed for the evolution of the promoter sequences of the NCLDVs, including selective sweep or convergence. However, these alternatives run off the diffused hypothesis of a common origin for the putative Megavirales order.

10. What Comes Next?

Most of the giant viruses have a powerful genetic arsenal, encoding several proteins necessary for the transcription system which provides a relative independence of their hosts for this process. In addition, the transcription of this high gene content is temporally regulated by promoter regions that exhibit some similarities, indicating a common origin of these regulatory elements. Although many studies have already been done in relation to almost all viral families of the Megavirales order, most of them remain without biological confirmation; i.e., the promoter motifs in many giant viruses were predicted, but not experimentally validated. Therefore, the performance of biological studies to confirm the existence and the effect of all promoter motifs described so far in giant viruses is imperative. This analysis will truly establish the common temporal regulation pattern predicted in these viruses, and will also corroborate (or even refute) the hypothesis of an A/T-rich promoter in the Megavirales common ancestor. Moreover, the deep analysis of the genome of the recently described giant viruses (Marseilleviruses, Pandoraviruses, Pithoviruses, Faustoviruses and Mollivirus), and also the discovery of new complex viruses, will strongly contribute to complete the puzzle of the origin and evolution of Megavirales.

On the other hand, the biotechnology field will also be boosted by the advance in the studies of promoters and gene expression in giant viruses. Among the NCLDVs, the poxviruses are by far the best

characterized group regarding the genome expression, especially the VACV. These viruses have been used as expression vectors for the synthesis of proteins and as vaccine candidates to prevent infectious diseases and treat cancer, mainly due to their high gene expression levels [69,138]. This attribute is clearly shared with other giant viruses that were recently described, and the real comprehension of their gene regulation and expression will bring uncountable possibilities for biotechnology purposes. Finally, the impact of the giant viruses on the basic comprehension of the origin and evolution of life is undeniable, as well as for their ecological, medical and technological importance. The discovery of even more complex viruses associated with the advance of many techniques used for genomic studies will certainly answer those remaining questions around the NCLDVs, and will surely bring new exciting challenges for the whole scientific community.

Acknowledgments: We would like to thank our colleagues from GEPVIG (Grupo de Estudo e Prospecção de Vírus Gigantes) and Laboratório de Vírus of Universidade Federal de Minas Gerais. In addition, 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). G.S.T., E.G.K. and J.S.A. are CNPq researchers. B.L.S., J.S.A. and E.G.K. are members of a CAPES-COFECUB project.

Author Contributions: G.P.O.; A.C.S.P.A.; R.A.L.R.; T.S.A.; P.V.M.B.; L.K.S.S. and F.P.D. wrote the paper. B.P.D., G.S.T., B.L.S., E.G.K. and J.S.A. approved and reviewed the final version.

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

References

1. Lwoff, A. The concept of virus. *J. Gen. Microbiol.* **1957**, *17*, 239–253. [CrossRef] [PubMed]
2. Durzyńska, J.; Goździcka-Józefiak, A. Viruses and cells intertwined since the dawn of evolution. *Virol. J.* **2015**, *16*, 169. [CrossRef] [PubMed]
3. Iyer, L.M.; Aravind, L.; Koonin, E.V. Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.* **2001**, *75*, 11720. [CrossRef] [PubMed]
4. La Scola, B.; Audic, S.; Robert, C.; Jungang, L.; de Lamballerie, X.; Drancourt, M.; Birtles, R.; Claverie, J.M.; Raoult, D. A giant virus in amoebae. *Science* **2003**, *299*. [CrossRef] [PubMed]
5. 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. [CrossRef] [PubMed]
6. Sharma, V.; Colson, P.; Chabrol, O.; Pontarotti, P.; Raoult, D. *Pithovirus sibericum*, a new bona fide member of the "Fourth TRUC" club. *Front. Microbiol.* **2015**, *4*, 722. [CrossRef] [PubMed]
7. Sharma, V.; Colson, P.; Chabrol, O.; Scheid, P.; Pontarotti, P.; Raoult, D. Welcome to pandoraviruses at the "Fourth TRUC" club. *Front. Microbiol.* **2015**, *18*, 423.
8. Scheid, P. A strange endocytobiont revealed as largest virus. *Curr. Opin. Microbiol.* **2016**, *31*, 58–62. [CrossRef] [PubMed]
9. Benamar, S.; Reteno, D.G.; Bandaly, V.; Labas, N.; Raoult, D.; la Scola, B. Faustoviruses: Comparative genomics of new Megavirales family members. *Front. Microbiol.* **2016**, *5*, 3. [CrossRef] [PubMed]
10. Filée, J. Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution. *Front. Microbiol.* **2015**, *16*, 593. [CrossRef] [PubMed]
11. Suzan-Monti, M.; la Scola, B.; Raoult, D. Genomic and evolutionary aspects of Mimivirus. *Virus Res.* **2006**, *117*, 145–155. [CrossRef] [PubMed]
12. Claverie, J.M. Viruses take center stage in cellular evolution. *Genome Biol.* **2006**, *7*, 110. [CrossRef] [PubMed]
13. Moreira, D.; López-García, P. Comment on "The 1.2-megabase genome sequence of Mimivirus". *Science* **2005**, *20*, 1114. [CrossRef] [PubMed]
14. 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* **2014**, *38–52*, 466–467. [CrossRef] [PubMed]
15. Raoult, D.; Audic, S.; Robert, C.; Abergel, C.; Renesto, P.; Ogata, H.; la Scola, B.; Suzan, M.; Claverie, J.M. The 1.2-megabase genome sequence of Mimivirus. *Science* **2004**, *306*, 1344–1350. [CrossRef] [PubMed]
16. Boyer, M.; Madoui, M.A.; 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**, *2*, e15530. [CrossRef] [PubMed]

17. Schones, D.E.; Cui, K.; Cuddapah, S.; Roh, T.-Y.; Barski, A.; Wang, Z.; Wei, G.; Zhao, K. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **2008**, *132*, 887–898. [CrossRef] [PubMed]
18. Fuda, N.J.; Ardehali, M.B.; Lis, J.T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **2009**, *10*, 186–192. [CrossRef] [PubMed]
19. Lacadie, S.A.; Ibrahim, M.M.; Gokhale, S.A.; Ohler, U. Divergent transcription and epigenetic directionality of human promoters. *FEBS J.* **2016**, *283*, 4214–4222. [CrossRef] [PubMed]
20. Haberle, V.; Lenhard, B. Promoter architectures and developmental gene regulation. *Semin. Cell Dev. Biol.* **2016**, *57*, 11–23. [CrossRef] [PubMed]
21. Weiss, S.; Gladstone, L. A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. *J. Am. Chem. Soc.* **1959**, *81*, 4118–4119. [CrossRef]
22. Thomas, M.C.; Chiang, C.M. The general transcription machinery and generalcofactors. *Crit. Rev. Biochem. Mol. Biol.* **2006**, *41*, 105–178. [CrossRef] [PubMed]
23. Sabin, L.R.; Delás, M.J.; Hannon, G.J. Dogma derailed: The many influences of RNA on the genome. *Mol. Cell* **2013**, *49*, 783–794. [CrossRef] [PubMed]
24. Vannini, A.; Ringel, R.; Kusser, A.G.; Berninghausen, O.; Kassavetis, G.A.; Cramer, P. Molecular basis of RNA polymerase III transcription repression by Maf1. *Cell* **2010**, *143*, 59–70. [CrossRef] [PubMed]
25. Carninci, P.; Sandelin, A.; Lenhard, B.; Katayama, S.; Shimokawa, K.; Ponjavic, J.; Semple, C.A.; Taylor, M.S.; Engström, P.G.; Frith, M.C. Genome-wide analysis of mammalian promoter architecture andevolution. *Nat. Genet.* **2006**, *38*, 626–635. [CrossRef] [PubMed]
26. Cooper, S.J.; Trinklein, N.D.; Anton, E.D.; Nguyen, L.; Myers, R.M. Comprehensive analysis of transcriptional promoter structure and functionin 1% of the human genome. *Genome Res.* **2006**, *16*, 1–10. [CrossRef] [PubMed]
27. Smale, S.T.; Kadonaga, J.T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **2003**, *72*, 449–479. [CrossRef] [PubMed]
28. Maston, G.A.; Evans, S.K.; Green, M.R. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genom. Hum. Genet.* **2006**, *7*, 29–59. [CrossRef] [PubMed]
29. Hansen, S.K.; Tjian, R. TAFs and TFIIA mediate differential utilization of thetandem Adh promoters. *Cell* **1995**, *82*, 565–575. [CrossRef]
30. Ren, B.; Maniatis, T. Regulation of Drosophila Adh promoter switching by aninitiator-targeted repression mechanism. *EMBO J.* **1998**, *17*, 1076–1086. [CrossRef] [PubMed]
31. Browning, D.F.; Busby, S.J. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **2016**, *14*, 638–650. [CrossRef] [PubMed]
32. Burgess, R.R.; Travers, A.A.; Dunn, J.J.; Bautz, E.K. Factor stimulating transcription by RNA polymerase. *Nature* **1969**, *221*, 43–46. [CrossRef] [PubMed]
33. Lee, D.J.; Minchin, S.D.; Busby, S.J. Activating transcription in bacteria. *Annu. Rev. Microbiol.* **2012**, *66*, 125–152. [CrossRef] [PubMed]
34. Murakami, K.S.; Masuda, S.; Campbell, E.A.; Muzzin, O.; Darst, S.A. Structural basis of transcription initiation: An RNA polymerase holoenzyme–DNA complex. *Science* **2002**, *296*, 1285–1290. [CrossRef] [PubMed]
35. Browning, D.F.; Busby, S.J. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2004**, *2*, 57–65. [CrossRef] [PubMed]
36. Hausner, W.; Wettach, J.; Hethke, C.; Thomm, M. Two transcription factors related with the eucaryal transcription factors TATA binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *J. Biol. Chem.* **1996**, *271*, 30144–30148. [CrossRef] [PubMed]
37. Bell, S.D.; Jaxel, C.; Nadal, M.; Kosa, P.F.; Jackson, S.P. Temperature template topology, and factor requirements of archaeal transcription. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 15218–15222. [CrossRef] [PubMed]
38. Darcy, T.J.; Hausner, W.; Awery, D.E.; Edwards, A.M.; Thomm, M.; Reeve, J.N. Methanobacterium thermoautotrophicum RNA polymerase and transcription in vitro. *J. Bacteriol.* **1999**, *181*, 4424–4429. [PubMed]
39. Reiter, W.D.; Hudepohl, U.; Zillig, W. Mutational analysis of an archae bacterial promoter—Essential role of a TATA Box for transcription efficiency and start-site selection in vitro. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9509–9513. [CrossRef] [PubMed]

40. Palmer, J.R.; Daniels, C.J. In vivo definition of an archaeal promoter. *J. Bact.* **1995**, *177*, 1844–1849. [CrossRef] [PubMed]
41. Kyrpides, N.C.; Ouzounis, C.A. Transcription in Archaea. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8545–8550. [CrossRef] [PubMed]
42. Aravind, L.; Koonin, E.V. DNA-binding proteins and evolution of transcription regulation in the *Archaea*. *Nucleic Acids Res.* **1999**, *27*, 4658–4670. [CrossRef] [PubMed]
43. Bell, S.D.; Jackson, S.P. Mechanism and regulation of transcription in *Archaea*. *Curr. Opin. Microbiol.* **2001**, *4*, 208–213. [CrossRef]
44. Werner, F.; Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **2011**, *9*, 85–98. [CrossRef] [PubMed]
45. Whelan, S. Viral Replication Strategies. In *Fields Virology*, 6th ed.; Knipe, D.M., Howley, P.M., Eds.; Lippincott, Williams and Wilkins: Philadelphia, PA, USA, 2014; Volume 2, p. 2160.
46. Abergel, C.; Rudinger-Thirion, J.; Giege, R.; Claverie, J.M. Virus-encoded aminoacyl-tRNA synthetases: Structural and functional characterization of mimivirus TyrRS and MetRS. *J. Virol.* **2007**, *81*, 12406–12417. [CrossRef] [PubMed]
47. Broyles, S.S.; Knutson, B.A. Poxvirus transcription. *Future Virol.* **2010**, *5*, 639–650. [CrossRef]
48. Moss, B. Poxviridae. In *Fields Virology*, 6th ed.; Knipe, D.M., Howley, P.M., Eds.; Lippincott, Williams and Wilkins: Philadelphia, PA, USA, 2014; Volume 2, p. 2129.
49. Legendre, M.; Audic, S.; Poirot, O.; Hingamp, P.; Seltzer, V.; Byrne, D.; Lartigue, A.; Lescot, M.; Bernadac, A.; Poulain, J.; et al. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res.* **2010**, *20*, 664–674. [CrossRef] [PubMed]
50. Suhre, K.; Audic, S.; Claverie, J.M. Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14689–14693. [CrossRef] [PubMed]
51. Damon, I.K. Poxviruses. In *Fields Virology*, 6th ed.; Knipe, D.M., Howley, P.M., Eds.; Lippincott, Williams and Wilkins: Philadelphia, PA, USA, 2014; Volume 2, p. 2160.
52. Baroudy, B.M.; Moss, B. Purification and characterization of a DNA dependent RNA polymerase from vaccinia virions. *J. Biol. Chem.* **1980**, *255*, 4372–4380. [PubMed]
53. Knutson, B.A.; Broyles, S.S. Expansion of poxvirus RNA polymerase subunits sharing homology with corresponding subunits of RNA polymerase II. *Virus Genes* **2008**, *36*, 307–311. [CrossRef] [PubMed]
54. Davison, A.J.; Moss, B. The structure of vaccinia virus early promoters. *J. Mol. Biol.* **1989**, *210*, 749–769. [CrossRef]
55. Baldick, C.J.; Keck, J.G.; Moss, B. Mutational analysis of the core, spacer and initiator regions of vaccinia virus intermediate class promoters. *J. Virol.* **1992**, *66*, 4710–4719. [PubMed]
56. Knutson, B.A.; Liu, X.; Oh, J. Vaccinia virus intermediate and late promoter elements are targeted by the TATA-binding protein. *J. Virol.* **2006**, *80*, 6784–6793. [CrossRef] [PubMed]
57. Yang, Z.; Bruno, D.P.; Martens, C.A.; Porcella, S.F.; Moss, B. Genome-wide analysis of the 5′ and 3′ ends of vaccinia virus early mRNAs delineates regulatory sequences of annotated and anomalous transcripts. *J. Virol.* **2011**, *85*, 5897–5909. [CrossRef] [PubMed]
58. Keck, J.G.; Baldick, C.J.; Moss, B. Role of DNA replication in vaccinia virus gene expression: A naked template is required for transcription of three late transactivator genes. *Cell* **1990**, *61*, 801–809. [CrossRef]
59. Yang, Z.; Reynolds, S.E.; Martens, C.A.; Bruno, D.P.; Porcella, S.F.; Moss, B. Expression profiling of the intermediate and late stages of poxvirus replication. *J. Virol.* **2011**, *85*, 9899–9908. [CrossRef] [PubMed]
60. Yang, Z.; Martens, C.A.; Bruno, D.P.; Porcella, S.F.; Moss, B. Pervasive initiation and 3′-end formation of poxvirus postreplicative RNAs. *J. Biol. Chem.* **2012**, *287*, 31050–31060. [CrossRef] [PubMed]
61. Broyles, S.S.; Liu, X.; Zhu, M.; Kremer, M. Transcription factor YY1 is a vaccinia virus late promoter activator. *J. Biol. Chem.* **1999**, *274*, 35662–35667. [CrossRef] [PubMed]
62. Hänggi, M.; Bannwarth, W.; Stunnenberg, H.G. Conserved TAAAT motif in vaccinia virus late promoters: Overlapping TATA box and site of transcription initiation. *EMBO J.* **1986**, *5*, 1071–1076. [PubMed]
63. International Committee on Taxonomy of Viruses. Available online: http://www.ictvonline.org/taxonomyHistory.asp?taxnode_id=20151927&taxa_name=Asfarviridae (accessed on 22 September 2016).
64. Sogo, J.M.; Almendral, J.M.; Talavera, A.; Vinuela, E. Terminal and internal inverted repetitions in African swine fever virus DNA. *Virology* **1984**, *133*, 271–275. [CrossRef]

65. Tulman, E.R.; Delhon, G.A.; Ku, B.K.; Rock, D.L. African swine fever virus. *Curr. Top. Microbiol. Immunol.* **2009**, *328*, 43–87. [PubMed]
66. Kuznar, J.; Salas, M.L.; Vinuela, E. DNA-dependent RNA polymerase in African swine fever virus. *Virology* **1980**, *101*, 169–175. [CrossRef]
67. Salas, J.; Salas, M.L.; Vinuela, E. Effect of inhibitors of the host cell RNA polymerase II on African swine fever virus multiplication. *Virology* **1988**, *164*, 280–283. [CrossRef]
68. Rodríguez, J.M.; Salas, M.L. African swine fever virus transcription. *Virus Res.* **2013**, *173*, 15–28. [CrossRef] [PubMed]
69. Moss, B. Genetically engineered poxviruses for recombinant gene expression, vaccination, and safety. *Proc. Natl. Acad. Sci. USA* **1996**, *15*, 11341–11348. [CrossRef]
70. Rodríguez, J.M.; Salas, M.L.; Vinuela, E. Intermediate class of mRNAs in African swine fever virus. *J. Virol.* **1996**, *70*, 8584–8589. [PubMed]
71. Garcia-Escudero, R.; Vinuela, E. Structure of African swine fever virus late promoters: Requirement of a TATA sequence at the initiation region. *J. Virol.* **2000**, *74*, 8176–8182. [CrossRef] [PubMed]
72. Yates, P.R.; Dixon, L.K.; Turner, P.C. Promoter analysis of an African swine fever virus gene encoding a putative elongation factor. *Biochem. Soc. Trans.* **1995**, *23*, 139S. [CrossRef]
73. Van Etten, J.L.; Graves, M.V.; Müller, D.G.; Boland, W.; Delaroque, N. Phycodnaviridae—Large DNA algal viruses. *Arch. Virol.* **2002**, *147*, 1479–1516. [CrossRef] [PubMed]
74. Van Etten, J.L.; Meints, R.H. Giant viruses infecting algae. *Annu. Rev. Microbiol.* **1999**, *53*, 447–494. [CrossRef] [PubMed]
75. Wilson, W.H.; van Etten, J.L.; Allen, M.J. The *Phycodnaviridae*: The story of how tiny giants rule the world. *Curr. Top. Microbiol. Immunol.* **2009**, *328*, 1–42. [PubMed]
76. International Committee on Taxonomy of Viruses. Available online: http://www.ictvonline.org/taxonomyHistory.asp?taxnode_id=20153552&taxa_name=Phycodnaviridae (accessed on 22 September 2016).
77. Kang, M.; Graves, M.; Mehmel, M.; Moroni, A.; Gazzarrini, S.; Thiel, G.; Gurnon, J.R.; van Etten, J.L. Genetic diversity in chlorella viruses flanking *kcv*, a gene that encodes a potassium ion channel protein. *Virology* **2004**, *326*, 150–159. [CrossRef] [PubMed]
78. Kawasaki, T.; Tanaka, M.; Fujie, M.; Usami, S.; Yamada, T. Immediate early genes expressed in chlorovirus infections. *Virology* **2004**, *318*, 214–223. [CrossRef] [PubMed]
79. Chuchird, N.; Nishida, K.; Kawasaki, T.; Fujie, M.; Usami, S.; Yamada, T. A variable region on the chlorovirus CVK2 genome contains five copies of the gene for Vp260, a viral-surface glycoprotein. *Virology* **2002**, *10*, 289–298. [CrossRef] [PubMed]
80. Yamada, T.; Hiramatsu, S.; Songsri, P.; Fujie, M. Alternative expression of a chitosanase gene produces two different proteins in cells infected with Chlorella virus CVK2. *Virology* **1997**, *14*, 361–368. [CrossRef] [PubMed]
81. Graves, M.V.; Meints, R.H. Characterization of the major capsid protein and cloning of its gene from algal virus PBCV-1. *Virology* **1992**, *188*, 198–207. [CrossRef]
82. Sugimoto, I.; Hiramatsu, S.; Murakami, D.; Fujie, M.; Usami, S.; Yamada, T. Algal-lytic activities encoded by Chlorella virus CVK2. *Virology* **2000**, *10*, 119–126. [CrossRef] [PubMed]
83. Fitzgerald, L.A.; Graves, M.V.; Li, X.; Feldblyum, T.; Hartigan, J.; van Etten, J.L. Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect Chlorella Pbi. *Virology* **2007**, *20*, 459–471. [CrossRef] [PubMed]
84. Fitzgerald, L.A.; Graves, M.V.; Li, X.; Feldblyum, T.; Nierman, W.C.; van Etten, J.L. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect Chlorella NC64A. *Virology* **2007**, *20*, 472–484. [CrossRef] [PubMed]
85. Fitzgerald, L.A.; Boucher, P.T.; Yanai-Balser, G.M.; Suhre, K.; Graves, M.V.; van Etten, J.L. Putative gene promoter sequences in the chlorella viruses. *Virology* **2008**, *25*, 388–393. [CrossRef] [PubMed]
86. Van Etten, J.L. Unusual life style of giant chlorella viruses. *Ann. Rev. Genet.* **2003**, *37*, 153–195. [CrossRef] [PubMed]
87. Wilson, W.H.; Schroeder, D.C.; Allen, M.J.; Holden, M.T.; Parkhill, J.; Barrell, B.G.; Churcher, C.; Hamlin, N.; Mungall, K.; Norbertczak, H.; et al. Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science* **2005**, *12*, 1090–1092. [CrossRef] [PubMed]

88. International Committee on Taxonomy of Viruses. Available online: http://www.ictvonline.org/taxonomyHistory.asp.taxnode_id=20153003&taxa_name=Iridoviridae (accessed on 22 September 2016).
89. Darai, G.; Delius, H.; Clarke, J.; Apfel, H.; Schnitzler, P.; Flugel, R.M. Molecular cloning and physical mapping of the genome of fish lymphocystis disease virus. *Virology* **1985**, *146*, 292–301. [CrossRef]
90. Williams, T.; Barbosa-Solomieu, V.; Chinchar, V.G. A decade of advances in Iridovirus research. *Adv. Virus Res.* **2005**, *65*, 174–248.
91. Jakob, N.J.; Muller, K.; Bahr, U.; Darai, G. Analysis of the first complete DNA sequence of an invertebrate iridovirus: Coding strategy of the genome of Chilo iridescent virus. *Virology* **2001**, *286*, 182–196. [CrossRef] [PubMed]
92. Chinchar, V.G.; Yu, K.H.; Jancovich, J.K. The molecular biology of frog virus 3 and other iridoviruses infecting cold-blooded vertebrates. *Viruses* **2011**, *3*, 1959–1985. [CrossRef] [PubMed]
93. Barray, S.; Devauchelle, G. Protein synthesis in cells infected by Chilo iridescent virus: Evidence for temporal control of three classes of induced polypeptides. *Virology* **1987**, *138*, 253–261.
94. D'Costa, S.M.; Yao, H.; Bilimoria, S.L. Transcription and temporal cascade in Chilo iridescent virus infected cells. *Arch. Virol.* **2001**, *146*, 2165–2178. [CrossRef] [PubMed]
95. D'Costa, S.M.; Yao, H.J.; Bilimoria, S.L. Transcriptional mapping in Chilo iridescent virus infections. *Arch. Virol.* **2004**, *149*, 723–742. [CrossRef] [PubMed]
96. Willis, D.B. DNA sequences required for trans activation of an immediate- early frog virus 3 gene. *Virology* **1987**, *161*, 1–7. [CrossRef]
97. Beckman, W.; Tham, T.N.; Aubertin, A.M.; Willis, D.B. Structure and regulation of the immediate-early frog virus 3 gene that encodes ICR489. *J. Virol.* **1988**, *62*, 1271–1277. [PubMed]
98. Willis, D.B.; Granoff, A. Transactivation of an immediateearly frog virus 3 promoter by a virion protein. *J. Virol.* **1985**, *56*, 495–501. [PubMed]
99. Pallister, J.; Goldie, S.; Coupar, B.; Hyatt, A. Promoter activity in the 59 flanking regions of the Bohle iridovirus ICP 18, ICP 46 and major capsid protein genes. *Arch. Virol.* **2005**, *150*, 1911–1919. [CrossRef] [PubMed]
100. Nalcacioglu, R.; Ince, I.A.; Vlak, J.M.; Demirbag, Z.; van Oers, M.M. The Chilo iridescent virus DNA polymerase promoter contains an essential AAAAT motif. *J. Gen. Virol.* **2007**, *88*, 2488–2494. [CrossRef] [PubMed]
101. Nalcacioglu, R.; Demirbag, Z.; Vlak, J.M.; van Oers, M.M. Promoter analysis of the Chilo iridescent virus DNA polymerase and major capsid protein genes. *Virology* **2003**, *317*, 321–329. [CrossRef] [PubMed]
102. Dizman, Y.A.; Demirbag, Z.; Ince, I.A.; Nalcacioglu, R. Transcriptomic analysis of Chilo iridescent virus immediate early promoter. *Virus Res.* **2012**, *167*, 353–357. [CrossRef] [PubMed]
103. Goorha, R. Frog virus 3 requires RNA polymerase II for its replication. *J. Virol.* **1981**, *37*, 496–499. [PubMed]
104. Goorha, R.; Willis, D.B.; Granoff, A. Macromolecular synthesis in cells infected by frog virus 3. VI. Frog virus 3 replication is dependent on the cell nucleus. *J. Virol.* **1977**, *21*, 802–805. [PubMed]
105. Goorha, R.; Murti, G.; Granoff, A.; Tirey, R. Macromolecular synthesis in cells infected by frog virus 3. VIII. The nucleus is a site of frog virus 3 DNA and RNA synthesis. *Virology* **1978**, *84*, 32–50. [CrossRef]
106. Baroudy, B.M.; Moss, B. Sequence homologies of diverse length tandem repetitions near ends of vaccinia virus genome suggest unequal crossing over. *Nucleic Acids Res.* **1982**, *25*, 5673–5679. [CrossRef]
107. Tidona, C.A.; Darai, G. Molecular anatomy of lymphocystis disease virus. *Arch. Virol. Suppl.* **1997**, *13*, 49–56. [PubMed]
108. Federici, B.A.; Bigot, Y.; Granados, R.R.; Hamm, J.J.; Miller, L.K.; Newton, I.; Stasiak, K.; Vlak, J.M. Family Ascoviridae. In *Virus Taxonomy: 8th Report of the International Committee on Taxonomy of Viruses*; Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A., Eds.; Elsevier Academic Press: San Diego, CA, USA, 2005; pp. 261–265.
109. International Committee on Taxonomy of Viruses. Available online: http://www.ictvonline.org/taxonomyHistory.asp?taxnode_id=20151919&taxa_name=Ascoviridae (accessed on 22 September 2016).
110. Cheng, X.W.; Wang, L.; Carner, G.R.; Arif, B.M. Characterization of three ascovirus isolates from cotton insects. *J. Invertebr. Pathol.* **2005**, *89*, 193–202. [CrossRef] [PubMed]
111. Asgari, S.; Davis, J.; Wood, D.; Wilson, P.; McGrath, A. Sequence and organization of the Heliothis virescens ascovirus genome. *J. Gen. Virol.* **2007**, *88*, 1120–1132. [CrossRef] [PubMed]

112. Bigot, Y.; Rabouille, A.; Sizaret, P.Y.; Hamelin, M.H.; Periquet, G. Particle and genomic characterization of a new member of the Ascoviridae, *Diadromus pulchellus ascovirus*. *J. Gen. Virol.* **1997**, *78*, 1139–1147. [CrossRef] [PubMed]
113. Cheng, X.W.; Carner, G.R.; Arif, B.M. A new ascovirus from *Spodoptera exigua* and its relatedness to the isolate from *Spodoptera frugiperda*. *J. Gen. Virol.* **2000**, *81*, 3083–3092. [CrossRef] [PubMed]
114. Salem, T.Z.; Turney, C.M.; Wang, L.; Xue, J.; Wan, X.-F.; Cheng, X.-W. Transcriptional analysis of a major capsid protein gene from *Spodoptera exigua ascovirus 5a*. *Arch. Virol.* **2008**, *153*, 149–162. [CrossRef] [PubMed]
115. Zhao, K.; Cui, L.W. Molecular characterization of the major virion protein gene from the *Trichoplusia ni ascovirus*. *Virus Genes* **2003**, *27*, 93–102. [CrossRef] [PubMed]
116. Stasiak, K.; Renault, S.; Demattei, M.V.; Bigot, Y.; Federici, B.A. Evidence for the evolution of ascoviruses from iridoviruses. *J. Gen. Virol.* **2003**, *84*, 2999–3009. [CrossRef] [PubMed]
117. Chinchar, V.G.; Hyatt, A.; Miyazaki, T.; Williams, T. Family *Iridoviridae*: Poor viral relations no longer. *Curr. Top. Microbiol. Immunol.* **2009**, *328*, 123–170. [PubMed]
118. Piégu, B.; Asgari, S.; Bideshi, D.; Federici, B.A.; Bigot, Y. Evolutionary relationships of iridoviruses and divergence of ascoviruses from invertebrate iridoviruses in the superfamily Megavirales. *Mol. Phylogenet. Evol.* **2015**, *84*, 44–52. [CrossRef] [PubMed]
119. Zauberman, N.; Mutsafi, Y.; Halevy, D.B.; 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**, *13*, 114. [CrossRef] [PubMed]
120. Xiao, C.; Kuznetsov, Y.G.; Sun, S.; Hafenstein, S.L.; Kostyuchenko, V.A.; Chipman, P.R.; Suzan-Monti, M.; Raoult, D.; McPherson, A.; Rossmann, M.G. Structural studies of the giant mimivirus. *PLoS Biol.* **2009**, *28*, e92. [CrossRef] [PubMed]
121. Rodrigues, R.A.; dos Santos Silva, L.K.; Dornas, F.P.; de Oliveira, D.B.; Magalhães, T.F.; Santos, D.A.; Costa, A.O.; de Macêdo Farias, L.; Magalhães, P.P.; Bonjardim, C.A.; 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. [CrossRef] [PubMed]
122. Legendre, M.; Santini, S.; Rico, A.; Abergel, C.; Claverie, J.M. Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. *Virol. J.* **2011**, *4*, 8–99. [CrossRef] [PubMed]
123. Fischer, M.G.; Allen, M.J.; Wilson, W.H.; Suttle, C.A. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. USA* **2010**, *9*, 19508–19513. [CrossRef] [PubMed]
124. Colson, P.; de Lamballerie, X.; Yutin, N.; Asgari, S.; Bigot, Y.; Bideshi, D.K.; Cheng, X.W.; Federici, B.A.; van Etten, J.L.; Koonin, E.V.; et al. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Arch. Virol.* **2013**, *158*, 2517–2521. [CrossRef] [PubMed]
125. Reteno, D.G.; Benamar, S.; Khalil, J.B.; Andreani, J.; Armstrong, N.; Klose, T.; Rossmann, M.; Colson, P.; Raoult, D.; La Scola, B. Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *J. Virol.* **2015**, *89*, 6585–6594. [CrossRef] [PubMed]
126. 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. [CrossRef] [PubMed]
127. 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. USA* **2014**, *111*, 4274–4279. [CrossRef] [PubMed]
128. Levasseur, A.; Andreani, J.; Delerce, J.; Bou Khalil, J.; Robert, C.; La Scola, B.; Raoult, D. Comparison of a modern and fossil pithovirus reveals its genetic conservation and evolution. *Genome Biol. Evol.* **2016**, *25*, 2333–2339. [CrossRef] [PubMed]
129. Legendre, M.; Lartigue, A.; Bertaux, L.; Jeudy, S.; Bartoli, J.; Lescot, M.; Alempic, J.M.; 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. USA* **2015**, *112*, 5327–5335. [CrossRef] [PubMed]
130. McLysaght, A.; Baldi, P.F.; Gaut, B.S. Extensive gene gain associated with adaptive evolution of poxviruses. *Proc. Natl. Acad. Sci. USA* **2003**, *23*, 15655–15660. [CrossRef] [PubMed]
131. Hughes, A.L.; Friedman, R. Poxvirus genome evolution by gene gain and loss. *Mol. Phylogenet. Evol.* **2005**, *35*, 186–195. [CrossRef] [PubMed]

132. Elde, N.C.; Child, S.J.; Eickbush, M.T.; Kitzman, J.O.; Rogers, K.S.; Shendure, J.; Geballe, A.P.; Malik, H.S. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell* **2012**, *17*, 831–841. [CrossRef] [PubMed]
133. Dixon, L.K.; Chapman, D.A.; Netherton, C.L.; Upton, C. African swine fever virus replication and genomics. *Virus Res.* **2013**, *173*, 3–14. [CrossRef] [PubMed]
134. Portugal, R.; Coelho, J.; Höper, D.; Little, N.S.; Smithson, C.; Upton, C.; Martins, C.; Leitão, A.; Keil, G.M. Related strains of African swine fever virus with different virulence: Genome comparison and analysis. *J. Gen. Virol.* **2015**, *96*, 408–419. [CrossRef] [PubMed]
135. Huang, Y.; Huang, X.; Liu, H.; Gong, J.; Ouyang, Z.; Cui, H.; Cao, J.; Zhao, Y.; Wang, X.; Jiang, Y.; et al. Complete sequence determination of a novel reptile iridovirus isolated from soft-shelled turtle and evolutionary analysis of Iridoviridae. *BMC Genomics* **2009**, *14*, 224. [CrossRef] [PubMed]
136. Koonin, E.V.; 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. [CrossRef] [PubMed]
137. 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 Evol. Biol.* **2012**, *24*. [CrossRef] [PubMed]
138. Moss, B. Reflections on the early development of poxvirus vectors 2013. *Vaccine* **2013**, *6*, 4220–4222. [CrossRef] [PubMed]